HHS Public Access

Author manuscript

Cell Rep. Author manuscript; available in PMC 2022 December 31.

Published in final edited form as:

Cell Rep. 2022 December 13; 41(11): 111777. doi:10.1016/j.celrep.2022.111777.

Excitatory-inhibitory recurrent dynamics produce robust visual grids and stable attractors

Xiaohan Zhang[1], **Xiaoyang Long**[2], **Sheng-Jia Zhang**[2], **Zhe Sage Chen**[1,2,3,4,*]

[1]Department of Psychiatry, New York University Grossman School of Medicine, New York, NY, USA

[2]Department of Neurosurgery, Xinqiao Hospital, Chongqing, China

[3]Neuroscience Institute, New York University Grossman School of Medicine, New York, NY, USA

[4]Lead contact

SUMMARY

Spatially modulated grid cells have been recently found in the rat secondary visual cortex (V2) during active navigation. However, the computational mechanism and functional significance of V2 grid cells remain unknown. To address the knowledge gap, we train a biologically inspired excitatory-inhibitory recurrent neural network to perform a two-dimensional spatial navigation task with multisensory input. We find grid-like responses in both excitatory and inhibitory RNN units, which are robust with respect to spatial cues, dimensionality of visual input, and activation function. Population responses reveal a low-dimensional, torus-like manifold and attractor. We find a link between functional grid clusters with similar receptive fields and structured excitatory-to-excitatory connections. Additionally, multistable torus-like attractors emerged with increasing sparsity in inter- and intra-subnetwork connectivity. Finally, irregular grid patterns are found in recurrent neural network (RNN) units during a visual sequence recognition task. Together, our results suggest common computational mechanisms of V2 grid cells for spatial and non-spatial tasks.

In brief

Zhang et al. train biologically realistic neural network models to perform a spatial navigation task with multisensory input and discover emergent grid-like responses and a low-dimensional, multistable torus-like attractor with imposed network connectivity and sparsity constraints.

*Correspondence: zhe.chen@nyulangone.org.

AUTHOR CONTRIBUTIONS

Z.S.C. conceived the study; Z.S.C. and X.Z. designed the experiment; X.L. and S.-J.Z. provided animal data recordings that motivated the computational modeling study; X.Z. performed all computer experiments and analyses; Z.S.C. wrote the initial draft of the manuscript; Z.S.C., S.-J.Z., and X.Z. edited and reviewed the final manuscript; Z.S.C. acquired the funding; Z.S.C. supervised the project.

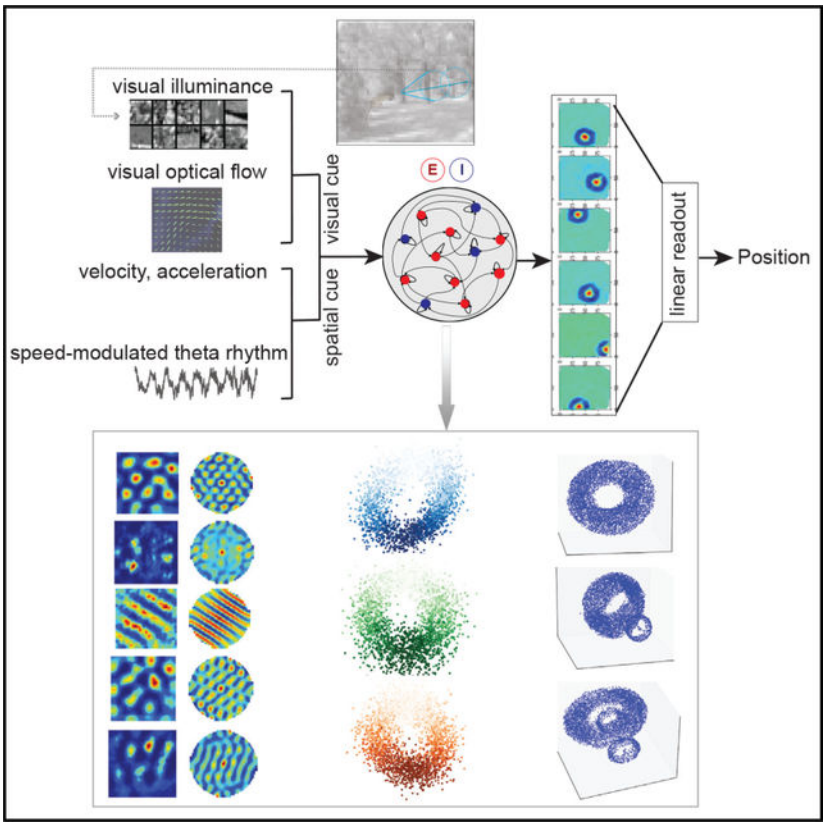
DECLARATION OF INTERESTS

The authors declare no competing interests.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.celrep.2022.111777.

Graphical Abstract



INTRODUCTION

The discoveries of periodic grid cells or grid-like responses have been reported in the rat, mouse, bat, and human brains during various spatial and non-spatial tasks.[1–14] One of important roles of grid cells is to integrate self-motion information that provides a path integrative input to identify spatial location even when external sensory inputs are lacking or noisy.[15,16] Grid patterns were first found in single neurons of the rat medial entorhinal cortex (mEC),[1,2] and recently reported in the rat primary somatosensory cortex (S1)[17] and the rat secondary visual cortex (V2).[18] These S1 and V2 grid cells share some common features as the mEC grid cells, such as conjunctive grid-head direction tunings and theta-modulated firing; furthermore, these grid-like responses are not disrupted by the absence of vibrissae or visual input.[17,18]

Attractor dynamics have been suggested in the hippocampal and entorhinal representations of the local environment.[19–21] To date, many computational models of mEC grid cells have been proposed (for reviews, see Giocomo et al.,[22] Zilli,[23] and Rowland et al.[24]), such as the continuous attractor models,[25,26] oscillator interference models,[27,28] feedforward neural network with excitatory and inhibitory synaptic plasticity,[29] and other hybrid models.[30–32] Recent work has shown that grid cells emerge from trained recurrent neural networks (RNNs) that predict spatial location based on a pure velocity input,[33–35] which supports the

hypothesis of recurrent attractor dynamics and path integrator in the cognitive map,[36] such that the attractor state may encode a stable representation of a variable (such as position) in the absence of external input.

Vision plays an important role in spatial navigation, and various visual cues can be integrated with spatial cues to guide movement. Spatial tunings have been reported in the dorsal lateral geniculate nucleus (dLGN), V1, and other visual cortical areas from head-fixed or freely foraging animals during spatial navigation tasks.[37–45] Vision and movement jointly contribute to hippocampal place codes.[46] Rodent V1 and hippocampal cornu ammonis (CA1) have also shown coherent coding of spatial signals.[41,47,48] However, it remains unclear whether similar grid patterns can emerge from visually cued navigation or motion.

To understand this question and further delineate the impact of multisensory input on neural representations, we developed a biologically constrained RNN to model experimentally observed grid cells in the rat V2. We adapted our computational models to incorporate both visual and spatial cues into the RNN for a spatial navigation task. We investigated the impact of various spatial (velocity or acceleration) and visual cues (illuminance or optical flow) on the grid responses of excitatory and inhibitory units. At the population level, we employed dimensionality reduction to reveal low-dimensional ring attractor dynamics and investigated the stability of grid responses with respect to visual and spatial inputs, synaptic connectivity, and excitatory-inhibitory (E/I) balance. In parallel to the spatial navigation task, we trained a combined convolutional neural network (CNN)-RNN model to perform a visual sequence recognition task and investigated the emergent grid patterns. Together, these simulation results reveal unexplored computational mechanisms of grid cells in the visual cortex and produce experimentally testable hypotheses for future investigation.

RESULTS

One of the potential roles of grid codes in sensory systems is to provide relative spatial mapping and localization within a reference system associated with the behavior. Accordingly, it is reasonable to hypothesize that the stability of grid patterns may predict the stability of behavioral output. Our investigation was centered on three essential questions: first, when and how do grid-like patterns emerge from visually cued navigation or motion? Second, how robust are visual grid patterns with respect to the input? Third, what is the functional role of these grid codes in relation to behavior?

Trained RNNs produced robust grid patterns with various spatial and visual cues

We trained biologically constrained E/I RNNs[49–51] to perform a spatial navigation task in a two-dimensional (2D) environmental enclosure. We envisioned that the RNN received various forms of visual and spatial cues in the input (Table S1) and predicted the position in the output. The network consisted of both excitatory and inhibitory units according to a 4:1 ratio and employed a non-negative rectified linear unit (ReLu) in the activation function (Figure 1A). We adopted a similar computation simulation setup to train the standard RNN[33–35] but with additionally imposed biological constraints.

RNN input configuration—First, to replicate previous computational simulation results,[33–35] we employed the E/I-RNN with a pure velocity (i.e., speed and direction) input (setup #1). We assumed that the run speed followed a normal distribution, and the run direction was uniformly distributed between 0 and 360° (Figure 1B). Next, we added additional visual illuminance input with varying dimensionality (setup #2). Upon reaching convergence, we projected the hidden unit activations of trained network onto the 2D space to obtain the position-modulated, speed-modulated, direction-modulated, and illumination-based tunings (Figures 1C and S1). We have witnessed a wide range of heterogeneity in spatially tuning the RNN units under different input configurations (Figure S1). Depending on specific configurations, subsets (20%–50%) of excitatory and inhibitory units showed clear grid-like responses. The range of the grid score (GS) varied depending on the input configurations or cell types. These grid-like units displayed various spatial frequencies as shown in their autocorrelograms and also displayed conjunctive coding for the head direction and speed (Figure 1C). In addition to grid-like excitatory and inhibitory units, we also found some periodic band-like excitatory units (Figures 1C and S1), which appeared as a combination of multiple grid units.[53,54] In setup #2, we systematically varied the dimensionality of principal components of visual illumination features and found consistent grid patterns in the RNN units (Figure 1D). To examine the stability of the learned RNN, we calculated the eigenvalues of the recurrent weight matrix and found that a large majority of complex (or real) eigenvalues were within the unit circle (Figure 1E), whereas a very small percentage of eigenvalues were slightly greater than 1. This result may suggest the chaotic spontaneous activity present in the trained RNN.[55] Furthermore, similar grid responses were observed when we replaced velocity (V_x, V_y) with acceleration $(A_x, A_y) \equiv (\dot{V}_x, \dot{V}_y)$ in the spatial input (setup #3; Figure 1F). Motivated by the recent experimental data of theta-modulated firing in V2 grid cells,[18] as well as the finding that the frequency or amplitude of theta oscillations increased proportionally to animal's run speed,[52,56] we relaxed the assumption of direct speed access and used the frequency of theta oscillations as the input (setup #4; Figure 1G). Consequently, we still observed robust grid responses (Figure 1F). The results of setups #3 and #4 were not really surprising as these substituted variables involved only approximately linear operations from velocity or speed. Notably, the direction input to the RNN was crucial to the formation of grid patterns, and grid units did not emerge if we removed the direction input in setup #4.

Finally, we computed the optical flow cue from the consecutive visual scenes and used that vector fields as the input (see STAR Methods; setup #5) to train the RNN. With pure visual cues, the trained RNN still preserved spatially modulated grid patterns (Figure 1F). This result was also easy to interpret since the optical flow offered an indirect source of motion cue (i.e., direction and speed information). Overall, varying the input configuration in our computer simulations yielded robust grid patterns with comparable GS statistics (Figures 1H and 1I). In the rest of this article, we will focus the analyses on two configurations (setups #2 and #5).

Impact of the sequence length—The E/I-RNN was trained by batches of simulated trajectories with a fixed length. We found that a wide range of sequence lengths, ℓ, produced grid patterns from the trained RNN units (Figure 2A). Specifically, we varied ℓ from 5 to

50 (corresponding to 100–1,000 ms for a 20-ms temporal bin size) and found that the GS statistics were robust with respect to the sequence length (Figure 2B). Our simulation results from multiple independently trained RNNs showed that the minimum ℓthat the achieved good GS statistic was 100 ms, roughly matching the timescale of one theta (5–10 Hz) cycle.

Activation function—We further tested whether the relaxation of default ReLu (non-negativity) activation function to an unconstrained linear unit might yield similar results (setups #2 and #5). Interestingly, the trained linear E/I-RNN could still produce grid-like responses (Figure 2A), but the overall GSs were lower, and the grid patterns were sensitive to the sequence length (Figure 2B). In this special case, the linear attractor network (without the non-negativity constraint) is a linear path integrator, integrating velocity or acceleration information in time to predict the future position (see STAR Methods).

Symmetry breaking is a critical condition for complex pattern formation.[36] It has been shown that the non-negativity of the activation function suffices to generate the grid patterns;[35] our results showed that even in the absence of non-negativity, Dale's principle alone may be sufficient for symmetry breaking. As a sanity check, we also trained an RNN without the Dale's principle constraint. In fact, fewer grid patterns emerged, but there was no clustered structure in the 2D embedding space (Figure S2C).

Mixed selectivity, paired unit correlation, and emerged functional clusters—For each recurrent unit, we empirically set the unit with GS <0.3 as grid units. We then computed the percentage of grid cells from both excitatory and inhibitory populations. Among the identified grid units, we plotted their tuning curves with respect to direction, speed, acceleration, or theta frequency (Figures 1B and S1). In setup #2, many of identified grid cells showed strong speed (71%) or directional (79%) tuning or both (65%). Because of the high dimensionality of visual features, we only plotted the tunings with respect to the dominant principal components (PCs) (e.g., Figure 1C). Speed tuning or directional tuning could also be observed for non-grid or band-like units (Figure S1). Interestingly, we found that a subset of band-like patterns had cosine-shaped direction tunings, and together, the preferred direction covered uniformly between 0° and 360° (Figure S3).

In setup #2, we examined the impact of changes in speed and visual input on the grid field (Figure 3A). To examine the co-dependency of spatial and visual tunings for the j -th unit, we correlated the time-averaged unit firing rates contributed by spatial or visual input alone: $\bar{r}_j^{\text{spatial}}$ (by setting the visual input to zero) and $\bar{r}_j^{\text{visual}}$ (by setting the velocity input to zero). We found statistically significant correlation between them (Figure 3B, left panel; Pearson's correlation, $p < 10^{-5}$), suggesting that the main driving factor of mean firing rate was the internal recurrent dynamics instead of external visual or speed input in the standard setting. However, increasing the speed out of the normal range (e.g., 10-fold) would substantially increase the mean firing rate (Figure 3B, right panel). Additionally, by setting the velocity to zero, we plotted the temporal firing rates of grid cells and measured the pairwise firing rate correlation when receiving a series of visual input (analogous to watching image sequences in a head-fixed setting). In some examples, we found that units with spatially similar grid fields showed temporally correlated visual responses (e.g., E_1

versus E_2 units in Figure 3C), whereas in other examples, units with correlated grid fields showed uncorrelated visual responses (e.g., E_3 versus E_4 units in Figure 3C), suggesting the independence between spatial and visual tunings. In a general setting, the excitatory grid units that had mutually strong excitatory synaptic connections tended to have similar grid fields; however, the converse was not necessarily true for the inhibitory grid units (for illustrated examples, see Figure 3D). Therefore, these functionally clustered grid units emerged as a result of strong synaptic connections, whereas weakly coupled grid units tended to be functionally decoupled.

To identify the functionally similar clusters in the grid cell subnetwork, we embedded the grid fields onto a 2D space for visualization (STAR Methods) and found that grid-like patterns formed many distinct clusters, especially among the excitatory grid-like units; this observation was robust regardless of the chosen activation function or input setup (Figures 3E and S4).

Emergent low-dimensional ring manifolds and attractors

Next, we examined the population representation of E/I-RNN units. According to the percentage of the explained variance (Figure 4A), we visualized two of the first three dominant PCs onto the latent space and found an emergent 2D ring-shaped manifold (Figure 4B). In setup #2, when the simulated sequence length was short, the PC_1 - PC_2 plane formed a ring attractor that could primarily be explained by the dominant spatial components; in contrast, when the sequence length was very long, the ring attractor was occupied on the PC_2 - PC_3 plane. This was possibly because the visual input of longer visuospatial sequences contributed more variance to the RNN's hidden unit activations. In both cases, the population activity was confined to lie close to a 2D manifold. Alternatively, we constructed the 3D embedding of N-dimensional population activity using a two-step hybrid dimensionality reduction procedure:[57] linear PC analysis (PCA; with the first 6 PCs) followed by a non-linear dimensionality reduction method known as uniform manifold approximation and projection (UMAP). This visualization step revealed a 3D torus-like structure (Figure 4C for setup #2; see Figure S5 for more results). Notably, the ring structure did not emerge in the linear E/I-RNN (Figure S5F).

In light of Fourier analysis (STAR Methods), we projected the 3D manifold onto three pre-determined pairs of axes (Figure 4D), each revealing the ring structure along different unit vector spaces (e.g., 0°, 60°, 120°). Similar ring structures were also found in other axis spaces (e.g., 30°, 90°, 150°; results not shown). Furthermore, we identified fixed points of the attractor using numerical simulations (STAR Methods). Figure 4E shows a torus attractor, indicating that the dynamic system had a stable activity pattern and was able to maintain memory. Furthermore, the fixed points also showed grid activity (Figure 4F). Interestingly, these fixed-point grid patterns resembled the activity of some hidden units in the PCA subspace. Together, these results suggested that these fixed-point patterns generated attractor fields, and the population activity of recurrent dynamic systems converged to them and formed stable grid patterns.

Robustness of grid patterns to visual and spatial inputs

The network state and recurrent dynamics could be biased by new external inputs. We further tested how the grid responses of the trained E/I-RNN would change with respect to an unseen visuospatial input. First, to emulate the darkness (light off) condition (setup #2), we set the visual input to zeros and found only small changes in the visual grid patterns (Figure 5A). The overall ring attractor remained similar other than being rotated (Figure 5B). Second, to emulate a new navigation environment, we switched to a new visual scene such that an identical run trajectory would have the same velocity input but different visual input. Again, we observed a stable grid pattern for a wide range of visual input (Figures 5C and 5D for setup #2 and S6 for setup #5). These results suggest that the trained E/I-RNNs and underlying recurrent attractors were robust to input perturbation.

The stability of grid patterns also predicted the stability of behavioral output. In testing, we simulated run trajectories (sequence length of 50) longer than the ones (sequence length of 10) used in the training phase. In the output space, with the same initial conditions, the noisy long trajectories either remained stable or first perturbed then converged to the simulated trajectories. In contrast, when the ring attractor was absent (as in the linear E/I-RNN; see Figure S5F), the trajectory output was unstable. Some snapshot examples under various testing conditions are shown in Figure 5E. Additionally, we tested how the change in behavioral output would affect grid cell representations. As a demonstration, in setup #2, we varied the run speed and found that the grid patterns were relatively stable for various speed inputs (Figure 5F). However, grid patterns and the ring manifold disappeared when speed was out of the normal range.

Robustness of grid patterns to recurrent network connectivity

In the literature, it has been suggested that the grid pattern formation can be generated through attractor dynamics in a recurrent network with a specific local E/I connectivity.[36,58] The recurrent attractor dynamics of the E/I-RNN was determined by a generally fully connected matrix $\mathbf{W}^{\mathrm{rec}}$. However, sensory cortices are known to have columnar organization: excitatory neurons within a columnar structure are densely connected and form functionally similar cortical maps,[59] whereas between-columnar neurons are sparsely connected. To make the E/I-RNN more biologically realistic, we further modified the structural synaptic connectivity and sparsity (i.e., the percentage of zeros in connection weights $\{W_{ij}^{rec}\}$).

Clustered E-E connectivity—The recurrent network connectivity was motivated by the anatomical evidence for non-uniform or clustered connections between cortical pyramidal neurons.[60] In the rodent mEC, grid cells tend to cluster anatomically.[45] Previous studies have shown clustered or columnar structures in the primary and secondary areas of visual cortex.[61,62] Therefore, we first split the E-population into two subnetworks, E_1 and E_2, and kept such clustered connectivity during the course of RNN training (type 1; Figure 6A, top panel). Upon the completion of training, we again observed robust grid responses in both E- and I-neuronal populations. Interestingly, both grid-like and band-like patterns emerged from the clustered subnetworks. This emergent continuous spectrum of spatially periodic units reflect different combinations of a small set of elemental periodic bands.[53,54] Additionally, the periodic grid or band-like units within the same excitatory subnetwork

tended to group together (i.e., with similar spatial frequency or orientation), whereas inhibitory grid units tended not to group together (Figure 6B). This was conceptually consistent with the observation seen in the non-structured networks (Figures 3C and 3D). Such clustered grouping based on similar spatial receptive fields (i.e., spatial frequency and orientation) was reminiscent of the minicolumns in the visual cortex. These periodic non-localized, band-like units were different from the spatially localized on/off visual receptive fields.

Furthermore, in training, we systematically changed the sparsity of inter-subnetwork connectivity and investigated the impact of sparsity on grid patterns. The sparsity level varied between 0 and 1, with 0 meaning fully connected (i.e., original setup) and 1 meaning completely disconnected between two subnetworks. Our results showed that the population GS statistics (Figures 6A, bottom panel) and 3D ring-like manifold (Figure 6C) remained stable with increasing sparsity in inter-subnetwork connectivity. Notably, an increasingly higher degree of sparsity in inter-subnetwork connectivity would produce more isolated clusters within E_1 and E_2 groups, many of which showed band-like patterns (Figures 6B, 6E, and S7).

Clustered I-I connectivity—Inhibitory projections were usually not clustered in our RNN model, consistent with a prior study showing that inhibitory neurons connect densely and non-specifically to pyramidal neurons.[60] However, inhibitory-to-inhibitory connections underlying a disinhibitory microcircuit may play an important role in reshaping the recurrent dynamics.[63] To test the influence of inhibitory connectivity, we further split the inhibitory population into two subnetworks, I_1 and I_2, with each projecting to specific excitatory subpopulations. We assumed that two inhibitory subnetworks were fully, yet weakly, coupled, and the excitatory-to-inhibitory and inhibitory-to-excitatory connections were strongly coupled (type 2). Again, we varied the sparsity in inter-subnetwork connectivity and further compared their GS statistics, grid field embedding, and ring manifold structure (Figures 6D–6F). Interestingly, we observed a mixture of band-like and grid-like patterns, and similar band-like patterns tended to group together within the same excitatory subnetwork. Similar observations were also found for other types of structured network connectivity (see Figure S7).

E/I balance—The E/I balance of trained E/I-RNNs was controlled by the relative degree of excitation and inhibition. To investigate the role of inhibition in reshaping spatial tunings of hidden units, we randomly selected some excitatory grid-like units and modified their relative inhibition by gradually decreasing or increasing the inhibitory-to-excitatory input or connection strengths (STAR Methods). Consequently, the grid pattern and GS changed (Figure 6G). Specifically, the grid-like patterns of excitatory units had a tendency to evolve into band-like patterns with decreasing inhibition strength, yet the grid patterns became weak or diminished with increasing inhibition strength. These results may explain the clustered band-like firing patterns under various structured network connectivity and sparsity levels.

Grid remapping—Firing patterns of mEC grid cells may change following environmental changes, such as translation and/or rotation of fields.[1,2] To emulate this remapping

condition, we conducted manipulations in multiple testing conditions. First, we rotated the environment by 90° and changed all place fields accordingly by 90° rotation. Second, we proportionally increased or decreased the size of place fields in the output (to emulate a larger or smaller environment size). Third, we permuted the trained RNN output order (or, equivalently, permuting the columns of $\mathbf{W}^{out}$) while keeping the place cell coverage unchanged. In the first condition, the grid patterns also rotated 90°, but the GS statistics remained unchanged (Figure S8A). In the second condition, the grid units displayed either zoom-in or zoom-out grid patterns within the original environment, but the overall grid patterns were preserved (Figure S8B). In the third condition, most grid patterns changed or remapped (Figure S8C). At the population representation level, the ring manifold structures remained close to the original one.

Multistable attractors emerged from sparsely connected RNNs

The structural connectivity density of inter- and intra-subnetwork jointly determines the overall sparsity in network connectivity. Next, from a pre-trained E/I RNN, in testing, we randomly set a small percentage of inter- or intra-subnetwork excitatory-to-excitatory connection weights to zeros. This would give us an opportunity to test the "corrupt" versions of the trained RNN with disrupted pre-wired synaptic connectivity. Interestingly, we found that with an increasing level of sparsity in the inter- and intra-subnetwork connectivity of the pre-trained E/I-RNN, two or three isolated 3D torus structures emerged (Figures 7 and S9), although individual grid patterns remained relatively stable. The co-existence of stable multiple-loop torus patterns implied a bifurcation phenomenon and the possibility of multistable attractor states.[64,65] Together, the results from Figures 6 and 7 suggest that functionally distinct grid patterns and multistable attractors may emerge from weakly coupled excitatory subnetworks.

To get insight into this phenomenon, we extracted statistics of $\mathbf{W}^{\text{rec}}$ with increasing sparsity using one trained E/I-RNN (setup #2). Specifically, we applied Schur decomposition to the non-normal matrix $\mathbf{W}^{\text{rec}}$ and quantified the strength of functionally feedforward connections (FFCs; denoted as κ) as well as the eigenvalue statistics (STAR Methods). The maximum eigenvalue of $\frac{\mathbf{W}^{\text{rec}}+\mathbf{W}^{\text{rec}\top}}{2}$ that characterizes the short-term behavior monotonically decreased with increasing sparsity in both inter- and intra-subnetwork connectivity (Figure S10A). In contrast, the maximum (real-part) eigenvalue of $\mathbf{W}^{\text{rec}}$ that characterizes the long-term behavior decreased with increasing sparsity in inter-subnetwork but decreased in intra-subnetwork connectivity. Additionally, κ reached a maximum with an intermediate sparsity level of intra-subnetwork but monotonically decreased with increasing sparsity in inter-subnetwork connectivity (Figure S10B). Since the largest (real-part) eigenvalue was greater than 1, it suggested that the chaotic state or bifurcation might be present in the RNN, explaining the phase transition of multiple loops in the torus manifold.

Grid-like patterns emerged from a trained RNN performing a non-spatial task

Can grid patterns emerge from the E/I-RNN that performs a non-spatial task? To answer this question, we considered a modified MNIST handwritten digit sequence recognition task. We first embedded the high-dimensional visual stimuli into a 2D space. In the output

space, the images were grouped together in the embedded space according to their similarity. Navigating between points in the embedded space can be viewed as a sequence recognition task. We employed a pre-trained CNN that emulated early-stage V1 pre-processing (Figure S11A) and subsequently took the flattened features from the CNN to feed to the E/I-RNN input (STAR Methods). Only the RNN parameters were trained in the CNN-RNN architecture, with the goal of mapping the image sequence to the 2D coordinate in the embedded space. We envisioned that when a set of digital image sequences (each with sequence length 10) was presented to the CNN-RNN, the output produced a trajectory in the embedded space. However, unlike our previous settings, the embedding coordinate space was not uniformly sampled by the trajectories; additionally, the step size of trajectory was non-even, with random speed and direction at every single step. After RNN training, we repeated the same analysis as in the spatial task and found a wide range of spatially tuned units: some displayed non-periodic irregular grid-like patterns, and others showed random patterns (Figure S11B). In this case, the overall GS statistics were lower (Figure S11C), and the low-dimensional population response formed a 10-cluster manifold structure in the PCA subspace (Figure S11D).

DISCUSSION

Accumulating evidence has pointed to rich spatial modulation phenomena in the visual thalamus and cortex.[18,38–41,43,44,47,48] A preliminary finding has suggested that a compact spatial map consisting of place cells, grid cells, head-direction cells, and border cells exists in the rat V2 visual cortex.[18] However, the sources and functional role of these spatially modulated signals in the visual cortex remain a puzzle. Theta oscillations and theta-modulated firing have also been found in the rat V2,[18] providing a possible source of speed signaling. Anatomical connections between the V2 and V1, and additional projections from the secondary motor (M2) and retrosplenial (RSC) cortices to visual cortices may provide additional self-motion, place-modulating, and directional inputs. Spatial modulation of place cells and grid cells in the rat V2 persisted in the darkness,[18] suggesting the robustness of these spatially modulated neurons and the existence of generalized cognitive map.[66] These results were consistent with previous experimental findings in rodent mEC grid cells.[1,52,67]

The receptive fields of V2 neurons have been identified with pure visual stimuli. For instance, V2 neurons in monkeys could be broadly classified as V1-like (typical Gabor-shaped subunits), with ultralong (subunits with high aspect ratios) or complex-shaped (subunits with multiple oriented components) subunits.[68] Rodent extrastriate areas may also process information related to other sensory modalities.[69,70] This may indicate fewer hierarchical stages in the rodent, delivering visual information more readily to multimodal interactions in naturalistic behaviors.[71] Among the emergent visual receptive fields, we found mutual independence between spatial and visual tunings of the grid units, suggesting task-dependent mixed selectivity in generalized grid codes.[72] Structural network connectivity is linked to functional clustering. It has been known that V1 neurons with similar non-classical extra-receptive fields (ERFs) tend to group into clusters;[73] these clusters are randomly distributed in all cortical layers, with no detectable relationship with orientation and ocular dominance columns. Our results of clustered grid-/band-like patterns emerged from sparse structural RNN connectivity seem to support this intuition. However,

experimental verification of this model prediction would require large-scale recordings from the rat V2. To date, a complete understanding of the relationship between V2 grid cells and mEC grid cells remains unknown; future simultaneous recordings of mEC grid cells and V2 grid cells can provide new insight. We speculate that multiple interconnected brain regions may be coordinated to perform local computation for tasks such as path integration or spatial localization. It is not completely impossible that grid responses are universal in the brain for a wide range of cognitive tasks that involve locomotion or mental navigation.[74,]

We envisioned that the output layer of the RNN model that encodes the place is the V1, which serves as a teaching signal to the V2. This is not impossible because rodent V1 neurons have place tunings.[41,47,48] Additionally, the back-propagating error from output V1 units to recurrent V2 units represents an information flow in the visual pathway. The V2 also provides a modulation input to the V1 in the feedback pathway.[75] Our V2 grid cell model is different from other mEC grid cell models in several ways. First, we generalized the standard RNN models[34,35] by incorporating Dale's principle and structured intra-cortical connectivity, making the model more biologically realistic in relation to the V2 visual cortex. Furthermore, many other models of mEC grid cells cannot be simply transferred to the V2 finding. Second, we showed that the visual optical flow could be an alternative input as the velocity to the V2, producing similar robust grid patterns and torus-like attractors. The optical flow can be implemented in the early stage of the visual system,[76,77] and the heading direction can be estimated from optical flow in the visual cortex[78] or sensorimotor circuit.[79] Third, our model reveals torus-like manifolds and attractors, consistent with the other experimental and computational findings.[35,57] Specifically, multistable torus-like manifolds emerge from the E/I-RNN with increasing inter- or intra-subnetwork connectivity, suggesting that sparsely connected biological networks may use multiple ring attractors to store independent information. Overall, our work provides a biologically plausible model to produce robust visual grid patterns; it also supports the hypothesis that the stability of grid patterns predicts the stability of behavioral output. The emerged multistable attractors may imply multiple attractor states, but detailed theoretical analyses of attractor multistability will be the subject of our future study.

Predictive maps serve as the common computational principles for generating "place codes" and "grid codes".[80,81] Specifically, the successor representation (SR)[82] is the product of the inherent state-action transition dynamics that characterizes the predictive dynamics. In the context of spatial navigation, the SR for a given state (i.e., a spatial location) is radially symmetric over space, and the columns or eigenvectors of the SR matrix correspond to the place fields and grid fields, respectively.[80,83] The state-action transition matrix bears a functional resemblance with the recurrent weight matrix in the RNN that characterizes the network state transition. Our work is also conceptually in line with a recent work that trained RNNs to predict a future action based on the current state and action, and the trained recurrent weight matrix that characterizes a predictive representation showed emergent place-like response patterns.[81] Unlike other continuous attractor models, the network connectivity matrix of our E/I-RNN is asymmetric and non-normal, and such a recurrent network structure produces stable grid patterns and torus-like attractors under various configurations of input and network connectivity. Notably, RNNs are capable of constructing a ring manifold that constrains a set of discrete fixed points.[84]

Path integration relies on an egocentric coding process and allows animals to integrate information (e.g., speed of movement, travel time, and directional change) generated by self-movements to update the position. Grid-like firing patterns provide a mechanism for dynamic computation of self-position based on continuously updated information about position and direction. Visual landmarks and motion cues are also critical for visuospatial integration. In virtual reality experiments, internal vestibular and proprioceptive cues are disrupted; however, some idiothetic cues that are not internally generated can still be used for path integration. Such cues may include visual optic flow, airstream detection (e.g., by a rat's whiskers), or other sensory reafference inputs produced by locomotion.[85] Can path-integration-like computations be used in non-spatial domains, for example, for constructing non-spatial representations such as time intervals or trajectories defined in a sensory stimulus space? Some evidence seems to suggest a confirmative answer. For instance, self-organized-domain general learning algorithms may explain the emergence of grid cells in both spatial and conceptual domains[86] so that grid representations provide efficient similarity search strategies in the generalized and continuous cognitive space. Additionally, a computational model of visual grid cells has been proposed for visual recognition memory where a sequence of memory-guided saccades can encode salient stimuli.[87] Our demonstration of visual grid patterns produced from a CNN-RNN model suggests that the computational principle of grid computation may be beyond the velocity-driven path-integration task. To date, RNNs have provided a dynamical systems viewpoint for motor movement,[88,89] path integration,[34,35] information integration,[90] and predictive representations.[81] We envision that the recurrent dynamics of RNNs can be implemented by a neural substrate in a wide range of cortical networks outside the traditional hippocampus-entorhinal system.

Limitations of the study

Several limitations are noticed in our RNN models. First, our simulated visual input (either illumination or optical flow) to the RNN was oversimplified and could not capture the complexity of natural vision. However, we speculate that the change in visual stimuli by adding an additional level of complexity will not affect the finding since the emergent grid responses appear robust to different visual input configurations. Second, the E/I-RNN did not explicitly consider the cortico-cortical input relevant to the task. For instance, there is strong cortico-cortical connectivity between the visual cortex and the parietal cortex as well as the RSC; these brain areas may carry additional spatial information essential for spatial and mental navigation. Third, our CNN-RNN model architecture to simulate early visual processing pathway (such as V1-V2) was oversimplified as the biological visual processing is more complex; therefore, incorporation of an architecture that aligns stages of visual processing along the ventral stream may reveal additional insight.[91–93]

Despite these limitations, our computational modeling work may produce experimentally testable predictions. First, our results suggest that the intra-cortical connectivity and E/I balance have an impact on the grid responses. This can be tested using intra-cortical microstimulation[94,95] or optogenetic stimulation.[96] Second, it remains unknown how visual grid patterns will change upon modifying the upstream input of the V2 (such as the V1 and visual thalamus). This can be investigated by selective inactivation of these FFCs.

Third, the mixed selectivity of V2 neurons can be tested via a series of experiments using well-controlled spatial and visual stimuli. Finally, it would be alluring to search for grid patterns in rodent visual and other sensory cortices during non-spatial tasks.

STAR★METHODS

RESOURCE AVAILABILITY

Lead contact—Further information and requests for data should be directed to and will be fulfilled by the lead contact, Zhe S. Chen (zhe.chen@nyulangone.org).

Materials availability—All materials are contained in the paper or available from the lead contact.

Data and code availability—This paper does not report original or raw data. Computer simulated data and custom Python scripts and codes have been deposited at Zenodo (https://doi.org/10.5281/zenodo.7275282) and are publicly available as of the data of publication. Any additional information required to reanalyzed the data reported in this paper is available from the lead contact author upon request.

EXPERIMENTAL MATERIALS AND DATA

MNIST dataset—The MNIST (Modified National Institute of Standards and Technology) dataset consists of 70000 28 × 28 pixel grayscale images of handwritten single digits between 0 and 9. The digits have been size-normalized and centered in a fixed-size image. The dataset is publicly available.

Rat V2 experimental recordings—The experimental recordings of V2 grid cells from freely behaving rats were described in details elsewhere.[18] We adapted our computer simulation setup to match animal's behavioral statistics (such as the run speed and direction). The local field potential (LFP) theta (8–10 Hz) rhythms from the rat V2 visual cortex were also used to guide the computer simulation (Setup #4). Specifically, we used the animal's real speed and interpolated the LFP theta frequency based on a 3-rd polynomial mapping. The resultant theta frequency was ranged between 8 Hz and 9 Hz (Figure 1G).

METHOD DETAILS

Input stimuli and output encoding for a spatial navigation task—In computer simulations, we generated 5000 random trajectories within a two-dimensional (2D) environment (2.2 m × 2.2 m) to simulate the animal's trajectories based on random speed and head direction. Each trajectory had a fixed sequence length (ℓ = 5–30, equivalent to 100–600 ms). The initial speed and head direction were randomized. To encode a 2D position, we assumed that the encoding of individual place unit had a 2D isotropic Gaussian shape as defined below[33]:

$$p_i(\mathbf{z}) = \frac{\exp\left(-\frac{\|\mathbf{z} - \mathbf{c}_i\|^2}{2\sigma^2}\right)}{\sum_{j=1}^{M} \exp\left(-\frac{\|\mathbf{z} - \mathbf{c}_j\|^2}{2\sigma^2}\right)} \quad \text{(Equation 1)}$$

where $\mathbf{c}_i \in \mathbb{R}^2$ denotes the center of place receptive field, and $\sigma > 0$ denotes the place cell scale. We used $M = 1024$ units to uniformly cover the 2D environment (Figure 1B). From the population activity of these place-modulating units, we could recover the 2D position $\mathbf{z} \equiv (x, y)$. Head-direction activations were sampled from a Gaussian distribution with zero mean and variance of 11.5 radians.

For Setup #1, the RNN only received the 2D velocity input at each time bin. For Setup #2, we added the additional visual input illuminance. Based on the heading direction, we defined a viewing region of interest (ROI) defined by an image of 8×8 pixels (i.e., dimensionality 64), which corresponded to the raw visual input (such as the luminance of pixels). To reduce the dimensionality, we further applied principal component analysis (PCA) and projected the vectorized image onto the dominant PC subspace. In our experiment, we tried varying numbers of PCs (2–20) that explained up to 93.4% variance in the visual stimuli. For Setup #3, the RNN received the 2D acceleration and visual input illuminance at each time bin. For Setup #4, we used the actual animal's run speed in the 2D environment to generate the trajectories, and further simulated the theta frequency based on a previously reported relationship between the run speed and theta frequency.[56]

RNN structure and training for a spatial navigation task—We trained an excitatory-inhibitory (E/I) RNN to perform a simulated spatial navigation task in the 2D open field enclosure (Figure 2A). We assumed the N-dimensional neural state dynamics, $\mathbf{x}(t)$, was driven by the following recurrent dynamics plus an N_{in} -dimensional input $\mathbf{u}(t)$:

$$\tau \dot{\mathbf{x}} = -\mathbf{x} + \mathbf{W}^{\text{rec}}\mathbf{r} + \mathbf{W}^{\text{in}}\mathbf{u} + \sigma\xi \quad \text{(Equation 2)}$$

where τ denotes the time constant, ξ denotes additive N-dimensional Gaussian noise, each independently drawn from a standard normal distribution, σ^2 defines the scale of the noise variance; $\mathbf{W}^{\text{rec}}$ is an $N \times N$ matrix of recurrent connection weights, and $\mathbf{w}^{\text{in}} \in \mathbb{R}^{N \times N_{\text{in}}}$ denotes the matrix of connection weights from the inputs to network units. The network produced an N_{out} -dimensional output $\mathbf{z}(t) = \mathbf{W}^{\text{out}}\mathbf{r}$, where the neuronal firing rate vector $\mathbf{r}$ is defined by an activation function $\varphi(\mathbf{x})$, which by default is a nonnegative rectified linear unit (ReLu)

$$\mathbf{r} = [\mathbf{x}]_+ = \max\{\mathbf{x}, \mathbf{0}\} \quad \text{(Equation 3)}$$

The ReLu is scale-invariant and favors sparse activation. In a special case when the activation function is a linear unit, the firing rate dynamics will be characterized by a linear dynamical system[50]

$$\tau \dot{\mathbf{r}} = -\mathbf{r} + \mathbf{W}^{\mathrm{rec}}\mathbf{r} + \mathbf{W}^{\mathrm{in}}\mathbf{u} + \sigma\xi \quad \text{(Equation 4)}$$

The E/I RNN was designed to satisfy the Dale's rule such that the ratio of excitatory to inhibitory units was 4:1[49]. We used $N = 512$ and $dt = 20$ ms in all numerical simulations. Depending on specific assumed input configurations (Table S1), the dimensionality varied in the RNN input, and the RNN output consisted of $N_{\mathrm{out}} = 1024$ place-modulated units, whose linear readout produced the 2D spatial position.

The recurrent weight matrix consisted of four functional submatrices according to the cell types: excitatory-to-excitatory (EE), inhibitory-to-excitatory (IE), excitatory-to-inhibitory (EI), and inhibitory-to-inhibitory (II) connections. Generally, $\mathbf{W}^{\mathrm{rec}}$ is non-normal (unless all submatrices $\{\mathbf{W}_{\mathrm{EE}}, \mathbf{W}_{\mathrm{EI}}, \mathbf{W}_{\mathrm{IE}}, \mathbf{W}_{\mathrm{II}}\}$ are symmetric and EI and IE connections are identical); as a result, its eigenvectors are not mutually orthogonal.[100,101] For individual postsynaptic excitatory or inhibitory units, the net excitatory and inhibitory currents were summed by the presynaptic input as follows

$$I_i^{\mathrm{exc}} = \sum_{j \in \mathrm{exc}} w_{ij,\mathrm{EE}}^{\mathrm{rec}} r_j, \; I_i^{\mathrm{inh}} = \sum_{j \in \mathrm{inh}} w_{ij,\mathrm{IE}}^{\mathrm{rec}} r_j$$

where $r_j = [x_j]_+ = \max(x_j, 0)$ denotes the neuronal firing rate of the j-th presynaptic neuron, $w_{ij,\mathrm{EE}}^{\mathrm{rec}}$ and $w_{ij,\mathrm{IE}}^{\mathrm{rec}}$ represent the EE and IE weights within $\mathbf{W}^{\mathrm{rec}}$, respectively.

We used the mean squared error as the cost function. The RNN was trained by back-propagation through time (BPTT) using the Adam algorithm with the default configuration of hyperparameters. In batch training, we used a learning rate of 0.0005 and a batch size of 256 randomly generated run trajectories. In each input configuration, we trained at least 10 RNNs with independent initializations in parallel via GPU. In the paper, we reported the representative results from one or ten trained networks.

CNN-RNN architecture for a non-spatial task—In the non-spatial task, we applied stochastic neighbor embedding (SNE) to the MINIST handwritten digit images and projected the images of 28×28 pixels onto a low-dimensional (2D or 3D) space. We used the t-distributed SNE algorithm and color coded different classes of digits in visualization.[98]

We used a pre-trained convolutional neural network (CNN) to emulate the early visual processing in the V1 visual cortex. The simple CNN architecture consisted of two convolution layers, each followed by max pooling operations. Each unit used a ReLu activation function. In our setting, we discarded fully connected layer in the pre-trained CNN (for classification); instead, we used the flattened input (dimensionality: 128) and fed that into the E/I-RNN to perform a visual navigation task in the embedded space (Figure S11A). In the new task, the pre-trained CNN parameters were fixed, and the E/I-RNN parameters were modified using a similar optimization procedure as in the spatial navigation task. The visual sequence consisted of 10 randomly permuted handwritten digit images (0–9) in the feature space; a batch size of 1024 sequences was used in training. It is noteworthy

that embedding the model's input into the 2D space was simply for better visualization of 2D grid-like representations. Generally, this concept can be extended to 3D or higher dimensional space.

Analysis of ring manifold and attractor—In order to identify the low-dimensional structure of population activity in the trained E/I-RNN, let $\mathbf{r}_i(x, y)$ denote the i-th unit firing rate map with respect to the 2D location (x, y); we computed its 2D (spatial) Fourier transform

$$\begin{aligned}\mathscr{R}_i(u, v) &= \iint \mathbf{r}_i(x, y)\exp(-j2\pi(ux + vy))dxdy \\ &= \iint \mathbf{r}_i(x, y)\exp(-j2\pi[x, y]^\top[u, v])dxdy\end{aligned}$$

where (u, v) denote the spatial frequencies. The function $\mathscr{R}_i(u, v)$ is constant when $[x, y]^\top[u, v] = (ux + vy)$ is constant; the magnitude of the vector (u, v) produces a frequency, and its phase gives an orientation. The function is a sinusoid with this frequency along the direction, and constant perpendicular to the orientation. The maxima and minima of real-valued sinusoidal basis $\cos 2\pi(ux + vy)$ occur when the inner product $2\pi[x, y]^\top[u, v] = n\pi$ corresponds to a set of equally spaced parallel lines, which have wave-length $1/\sqrt{u^2 + v^2}$ and are perpendicular to vector $[u, v]$.

We further computed three spatial phases for each unit's rate map as follows[35]

$$\varphi_i^s = \arg\left[\iint \mathbf{r}_i(x, y)\exp(-[x, y]^\top \mathbf{k}^s)dxdy\right],\ s = 0, 60, 120$$

where $\mathbf{k}^0$, $\mathbf{k}^{60}$, $\mathbf{k}^{120}$ represent the 0°, 60° and 120° rotation unit vectors, respectively. The inner product between $[x, y]$ and the rotation vector yields a new set of coordinate (x_{new}, y_{new}):

$$x_{new} = x\cos\psi - y\sin\psi$$

$$y_{new} = x\sin\psi + y\cos\psi$$

where ψ denotes the rotation angle. Finally, we projected the population activity onto the three orthogonal pairs of axes: $\{u^0 \equiv \cos(\varphi^0),\ v^0 \equiv \sin(\varphi^0)\}$, $\{u^{60} \equiv \cos(\varphi^{60}),\ v^{60} \equiv \sin(\varphi^{60})\}$ and $\{u^{120} \equiv \cos(\varphi^{120}),\ v^0 \equiv \sin(\varphi^{120})\}$. Note that the choice of (0°, 60°, 120°) was based on the assumption of a perfect hexagonal grid; for the quadrilateral grid, the choice would be (0°, 90°, 180°). In practice, we found that a wide range of rotation angles produced qualitatively similar results.

To identify the attractor of the low-dimensional dynamics, let $F(\mathbf{x}) = \dot{\mathbf{x}}$ in Equation 2; the Jacobian of the RNN was computed as

$$\frac{\partial F}{\partial \mathbf{x}} = \frac{1}{\tau}\left(-\mathbf{I} + \mathbf{W}^{\mathrm{rec}}\Phi\right)$$

where Φ represents an N-by-N diagonal matrix containing the derivative of $\varphi(\mathbf{x})$ with respect to $\mathbf{x}$. Applying eigenvalue decomposition to the Jacobian matrix, we obtained N eigenmodes (eigenvectors) in the state space, and the associated N complex-valued eigenvalues that quantified the rate and direction along individual dimension.[84] At the fixed points, we set $F(\mathbf{x}^*) = 0$, or equivalently

$$\mathbf{x}* = \mathbf{W}^{\mathrm{rec}}\,\varphi(\mathbf{x}*) + \mathbf{W}^{\mathrm{in}}\,\mathbf{u}$$

The analytic solution was not available because of the ReLu nonlinearity $\varphi(\mathbf{x})$. In the special case of linear RNN, the fixed point was given by $\mathbf{x}^* = (\mathbf{I} - \mathbf{W})-1\mathbf{W}^{\mathrm{in}}\mathbf{u}$. When $(\mathbf{I} - \mathbf{W})$ has a full rank, then the fixed point is unique; otherwise, the linear RNN has more than one fixed point.[102] In our analysis, we identified fixed-points or slow points by numerically solving the optimization problem[97]:

$$\min_{\mathbf{x}} q(\mathbf{x}), \quad \text{where} \quad q(\mathbf{x}) = \left\| -\mathbf{x} + \mathbf{W}^{\mathrm{rec}}\varphi(\mathbf{x}) + \mathbf{W}^{\mathrm{rec}}\mathbf{u}\right\|^2$$

We collected a set of fixed-points by randomly initializing the network on a grid of 100×100 spatial locations in the 2D environment. Specifically, the dimensionality of fixed points was the same as dim($\mathbf{x}$); once the numerical optimization was completed, we applied PCA to visualize the fixed points in the three- or two-dimensional PC subspace.

Linear RNN as a path integrator—Let's consider a continuous-time vector differential equation that describes the dynamics of the linear RNN with N recurrent connected neurons:

$$\begin{aligned}\tau_{\mathbf{x}}\dot{x}(t) = & -\mathbf{x}(t) + \mathbf{W}\mathbf{x}(t) + \mathbf{W}^{\mathrm{in}}\mathbf{u}(t)\\ & = (\mathbf{W} - \mathbf{I})\mathbf{x}(t) + \mathbf{W}^{\mathrm{in}}\mathbf{u}(t)\end{aligned}$$

For simplicity, let the time constant $\tau_x = 1$ and set $\mathbf{A} = \mathbf{W} - \mathbf{I}$ (where $\mathbf{I}$ is an N×N identity matrix), then

$$\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{W}^{\mathrm{in}}\,\mathbf{u}(t)$$

Given the linear dynamical system, we define a matrix function $\Phi(t, \tau)$ that has the following two properties

$$\dot{\Phi}(t,\tau) = \mathbf{A}\Phi(t,\tau)$$

$$\Phi(\tau,\tau) = \quad \mathbf{I}$$

where the matrix function is referred to as the state transition matrix. If **A** is non-singular, then the state transition matrix is unique. Solving the linear vector differential equation yields

$$\mathbf{x}(t) = \Phi(t, t_0)\mathbf{x}(t_0) + \int_{t_0}^{t} \Phi(t, \tau)\mathbf{W}^{\text{in}}\,\mathbf{u}(\tau)d\tau$$

When the input $\mathbf{u}(t)$ is completely absent, the dynamical system is purely driven by the recurrent dynamics governed by the eigen-functions of **A**:

$$\mathbf{x}(t) = \mathbf{x}(t_0) + c_1 e^{\lambda 1 t}\mathbf{v}_1 + c_2 e^{\lambda 2 t}\mathbf{v}_2 + \ldots + c_N e^{\lambda N t}\mathbf{v}_N$$

where $\{\lambda_1, \ldots, \lambda_N\}$ a are the eigenvalues, and $\{\mathbf{v}_1, \ldots, \mathbf{v}_N\}$ are the eigenvectors of the matrix **A**.

The linear readout of the RNN output is given as

$$\begin{aligned}\mathbf{y}(t) &= \mathbf{W}^{\text{out}}\mathbf{x}(t) \\ &= \mathbf{W}^{\text{out}}\left(\Phi(t, t_0)\mathbf{x}(t_0)() + \int_{t_0}^{t} \Phi(t, \tau)\mathbf{W}^{\text{in}}\,\mathbf{u}(\tau)d\tau\right)\end{aligned}$$

Non-normal connectivity and dynamics—We call a matrix **A** *normal* if it it satisfies $\mathbf{A}\mathbf{A}^{\top} = \mathbf{A}^{\top}\mathbf{A}$, and a stable linear normal system is contractive.[101] On the other hand, a matrix A is non-normal if it satisfies $\mathrm{A}\mathrm{A}^{\top} \neq \mathrm{A}^{\top}\mathrm{A}$. The recurrent weight matrix $\mathbf{W}^{\text{rec}}$ of the E/I-RNN is asymmetric and non-normal, unless all submatrices $\{\mathbf{W}_{\text{EE}}, \mathbf{W}_{\text{EI}}, \mathbf{W}_{\text{IE}}, \mathbf{W}_{\text{II}}\}$ are symmetric and the EI and IE connections are identical. As a result, the eigenvectors of $\mathbf{W}^{\text{rec}}$ do not form the orthonormal bases.[101]

Specifically, an arbitrary recurrent weight matrix $\mathbf{W}^{\text{rec}}$ can be rewritten in the following form

$$\mathbf{W}^{\text{rec}} = \mathbf{U}^{\top}(\mathbf{\Lambda} + \mathbf{T})\mathbf{U}$$

where $\mathbf{U} = \{u_{ij}\}$ is unitary, and $\mathbf{\Lambda}$ is a diagonal matrix that contains the eigenvalues $\{\lambda_k\}$ of $\mathbf{W}^{\text{rec}}$, and **T** is a lower-diagonal matrix. The vectors of **U** are called the Schur vectors (or Schur modes) and are mutually orthogonal. In the linear E/I-RNN, let $H\left(\mathbf{W}^{\text{rec}}\right) = \frac{\mathbf{W}^{\text{rec}} + \mathbf{W}^{\text{rec}\,\top}}{2}$ be the Hermittian part of the recurrent weight matrix, then the maximum of the eigenvalue of $H(\mathbf{W}^{\text{rec}})$ characterizes the short-term behavior of the network undergoing a transient growth before asymptotically converging to zeros.[103] Additionally, the maximum of the real (or real-part) eigenvalues of $\mathbf{W}^{\text{rec}}$ characterizes the long-term behavior for the speed of network's steady state decaying to zero. Highly non-symmetric interactions of simulated neurons may create non-normal dynamics, such as large transients (G. Kerg et al., 2019, NuerIPS, conference).

QUANTIFICATION AND STATISTICAL ANALYSIS

Computation of grid score, grid field and autocorrelogram—The grid score (GS) was previously established to quantify the grid-like responses of neurons.[1,33] To be consistent with all setup conditions, we set an empirical GS threshold of ≥ 0.3 for the grid unit. The empirical threshold was guided by a random shuffle distribution using a Monte Carlo $P< 0{:}01$. If a unit was categorized as a grid cell, the grid field was defined as the spatial rate map (50×50 bins) normalized by the behavioral occupancy. Using a similar method,[33] we calculated the spatial autocorrelation with smoothed rate maps. Let $r(x, y)$ denote the unit's mean firing rate at a two-dimensional Cartesian coordinate (x,y), the autocorrelation of the spatial firing field was calculated as[17]:

$$\rho(\tau_x, \tau_y) = \frac{n\sum r(x, y)r(x - \tau_x, y - \tau_y) - \sum r(x, y)\sum r(x - \tau_x, y - \tau_y)}{\sqrt{n\sum r(x, y)^2 - [\sum r(x, y)]^2}\sqrt{n\sum r(x - \tau_x, y - \tau_y)^2 - [\sum r(x - \tau_x, y - \tau_y]^2}}$$

where the summation was over n pixels for both $r(x, y)$ and $r(x - \tau_x, y - \tau_y)$ (where τ_x and τ_y denote the spatial lags).

Computation of optical flow—Optical flow is commonly referred to as the pattern of apparent motion of objects in a visual scene. In computer vision, the optical flow methods try to calculate the motion between two image frames which are taken at times t and $t + \Delta t$. Based on local Taylor series approximations of the frame images, these methods use partial derivatives with respect to the spatial and temporal coordinates. For a 2D + t dimensional case, if a voxel at location (x, y, t) with visual illuminance $I(x, y, t)$ is moved by $(\Delta x, \Delta y, \Delta t)$ between two image frames, then the following "brightness constancy constraint" needs to be satisfied

$$I(x, y, t) = I(x + \Delta x, y + \Delta y, t + \Delta t)$$

Based on a first-order Taylor series expansion, the following equation can be derived

$$\frac{\partial I}{\partial x}\Delta x + \frac{\partial I}{\partial y}\Delta y + \frac{\partial I}{\partial t}\Delta t = 0$$

or

$$\frac{\partial I}{\partial x}V_x + \frac{\partial I}{\partial y}V_y + \frac{\partial I}{\partial t} = 0$$

where $V_x = \frac{\Delta x}{\Delta t}$, $V_y = \frac{\Delta y}{\Delta t}$. We then used the Horn-Schunck estimation method to estimate the optical flow based on neighboring frames.[104,105] The optical flow was represented and visualized by a 2D vector field, with arrows indicating the direction, and the size of arrow proportional to the scale. In our experiment, we used a 16×16 visual frame to compute the optical flow, resulting in a 512-dimensional vectorized feature.

Dimensionality reduction—To visualize the low-dimensional recurrent dynamic attractor, we applied linear principal component analysis (PCA) and projected the N -dimensional latent vector onto a 2D subspace (PC_1 - PC_2 or PC_2 - PC_3) according to the percentage of explained variance. In addition, we used a nonlinear dimensionality reduction technique known as UMAP (Uniform Manifold Approximation and Projection).[99,106] The algorithm is designed to find an embedding by searching for a low dimensional projection of the data that has the closest possible equivalent fuzzy topological structure.

To quantify the similarity of grid fields, we embedded the 50×50 grid maps into a 2D space using the t-SNE algorithm,[98] with a default perplexity parameter of 30.

Schur decomposition—We applied Schur decomposition to the trained recurrent weight matrix: $\mathbf{W}^{\text{rec}} = \mathbf{QTQ}^{-1}$, where $\mathbf{Q}$ is a unitary matrix whose columns contain the orthogonal Schur mode, and $\mathbf{T}$ is a lower triangular matrix that contains the eigenvalues along the diagonal.[50,100,101] The triangular structure of $\mathbf{T}$ can be interpreted as transforming an RNN into a feedforward neural network, and the recurrent weight matrix $\mathbf{W}^{\text{rec}}$ corresponds to a rotated version of the effective feedforward matrix $\mathbf{T}$, which defines self-connections and functionally feedforward connections (FFCs) of the neural network. Unlike eigenvalue decomposition, the Schur decomposition produces the simplest (yet non-unique) orthonormal basis for a non-normal matrix. The Schur decomposition of the non-normal matrix $\mathbf{W}^{\text{rec}}$ naturally provides a separation of “diagonal” (recurrent) and “non-diagonal” (feedforward) parts. To quantify the strength of FFCs (denoted by κ), we computed the sum of absolute squares of the off-diagonal elements of $\mathbf{T}$, and further normalized it by the sum of absolute squares of all the elements of $\mathbf{T}$ [101]:

$$\kappa = \frac{\text{Trace}\left(\mathbf{TT}^{\top}\right) - \sum_{j=1}^{N} |\lambda_j|^2}{\text{Trace}\left(\mathbf{TT}^{\top}\right)}$$

where $\{\lambda_j\}_{j=1}^{N}$ denote the eigenvalues of the lower diagonal matrix $\mathbf{T}$. The value κ is interpreted as the proportion of dynamics driven by FFCs, whereas $1 - \kappa$ is interpreted as the proportion of dynamics driven by functionally recurrent connections.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

Z.S.C. and X.Z. were partly supported by the US National Institutes of Health (MH118928, NS121776, and DA056394). Z.S.C. also received cloud computing resources supported by the Oracle for Research Award.

REFERENCES

1. Hafting T, Fyhn M, Molden S, Moser MB, and Moser EI (2005). Microstructure of a spatial map in the entorhinal cortex. Nature 436, 801–806. [PubMed: 15965463]
2. Fyhn M, Hafting T, Treves A, Moser MB, and Moser EI (2007). Hippocampal remapping and grid realignment in entorhinal cortex. Nature 446, 190–194. [PubMed: 17322902]

3. Yartsev MM, Witter MP, and Ulanovsky N (2011). Grid cells without theta oscillations in the entorhinal cortex of bats. Nature 479, 103–107. [PubMed: 22051680]
4. Jacobs J, Weidemann CT, Miller JF, Solway A, Burke JF, Wei XX, Suthana N, Sperling MR, Sharan AD, Fried I, and Kahana MJ (2013). Direct recordings of grid-like neuronal activity in human spatial navigation. Nat. Neurosci. 16, 1188–1190. [PubMed: 23912946]
5. Doeller CF, Barry C, and Burgess N (2010). Evidence for grid cells in a human memory network. Nature 463, 657–661. [PubMed: 20090680]
6. Constantinescu AO, O'Reilly JX, and Behrens TEJ (2016). Organizing conceptual knowledge in humans with a gridlike code. Science 352, 1464–1468. [PubMed: 27313047]
7. Bellmund JL, Deuker L, Doeller CF, and Doeller CF (2019). Grid-cell representations in mental stimulation. Elife 8, e17089.
8. Bellmund JLS, Gärdenfors P, Moser EI, and Doeller CF (2018b). Navigating cognition: spatial codes for human thinking. Science 362, eaat6766. [PubMed: 30409861]
9. Nau M, Navarro Schröder T, Bellmund JLS, and Doeller CF (2018). Hexadirectional coding of visual space in human entorhinal cortex. Nat. Neurosci. 21, 188–190. [PubMed: 29311746]
10. Bao X, Gjorgieva E, Shanahan LK, Howard JD, Kahnt T, and Gottfried JA (2019). Grid-like neural representations support olfactory navigation of a two-dimensional odor space. Neuron 102, 1066–1075.e5. [PubMed: 31023509]
11. Shilnikov AL, and Maurer AP (2016). The art of grid fields: geometry of neuronal time. Front. Neural Circuits 10, 12. [PubMed: 27013981]
12. Rueckemann JW, Sosa M, Giocomo LM, and Buffalo EA (2021). The grid code for ordered experience. Nat. Rev. Neurosci. 22, 637–649. [PubMed: 34453151]
13. Ginosar G, Aljadeff J, Burak Y, Sompolinsky H, Las L, and Ulanovsky N (2021). Locally ordered representation of 3D space in the entorhinal cortex. Nature 596, 404–409. [PubMed: 34381211]
14. Grieves RM, Jedidi-Ayoub S, Mishchanchuk K, Liu A, Renaudineau S, Duvelle É, and Jeffery KJ (2021). Irregular distribution of grid cell firing fields in rats exploring a 3D volumetric space. Nat. Neurosci. 24, 1567–1573. [PubMed: 34381241]
15. Bush D, Barry C, and Burgess N (2014). What do grid cells contribute to place cell firing? Trends Neurosci. 37, 136–145. [PubMed: 24485517]
16. Bush D, Barry C, Manson D, and Burgess N (2015). Using grid cells for navigation. Neuron 87, 507–520. [PubMed: 26247860]
17. Long X, and Zhang S-J (2021). A novel somatosensory spatial navigation system outside the hippocampal formation. Cell Res. 31, 649–663. [PubMed: 33462427]
18. Long X, Deng B, Cai J, Chen ZS, and Zhang S-J (2021). A compact spatial map in V2 visual cortex. Preprint at bioRxiv. 10.1101/2021.02.11.430687v1.
19. Wills TJ, Lever C, Cacucci F, Burgess N, and O'Keefe J (2005). Attractor dynamics in the hippocampal representation of the local environment. Science 308, 873–876. [PubMed: 15879220]
20. Burak Y (2014). Spatial coding and attractor dynamics of grid cells in the entorhinal cortex. Curr. Opin. Neurobiol. 25, 169–175. [PubMed: 24561907]
21. Agmon H, and Burak Y (2020). A theory of joint attractor dynamics in the hippocampus and the entorhinal cortex accounts for artificial remapping and grid cell field-to-field variability. Elife 9, e56894. [PubMed: 32779570]
22. Giocomo LM, Moser M-B, and Moser EI (2011). Computational models of grid cells. Neuron 71, 589–603. [PubMed: 21867877]
23. Zilli EA (2012). Models of grid cell spatial firing published 2005–2011. Front. Neural Circuits 6, 16. [PubMed: 22529780]
24. Rowland DC, Roudi Y, Moser MB, and Moser EI (2016). Ten years of grid cells. Annu. Rev. Neurosci. 39, 19–40. [PubMed: 27023731]
25. Fuhs MC, and Touretzky DS (2006). A spin glass model of path integration in rat medial entorhinal cortex. J. Neurosci. 26, 4266–4276. [PubMed: 16624947]
26. Burak Y, and Fiete IR (2009). Accurate path integration in continuous attractor network models of grid cells. PLoS Comput. Biol. 5. e1000291. [PubMed: 19229307]

27. Burgess N, Barry C, and O'Keefe J (2007). An oscillatory interference model of grid cell firing. Hippocampus 17, 801–812. [PubMed: 17598147]
28. Burgess N (2008). Grid cells and theta as oscillatory interference: theory and predictions. Hippocampus 18, 1157–1174. [PubMed: 19021256]
29. Weber SN, and Sprekeler H (2018). Learning place cells, grid cells and invariances with excitatory and inhibitory plasticity. Elife 7, e34560. [PubMed: 29465399]
30. Bush D, and Burgess N (2014). A hybrid oscillatory interference/continuous attractor network model of grid cell firing. J. Neurosci. 34, 5065–5079. [PubMed: 24695724]
31. Kang L, and Balasubramanian V (2019). A geometric attractor mechanism for self-organization of entorhinal grid modules. Elife 8, e46687. [PubMed: 31373556]
32. Rosay S, Weber S, and Mulas M (2019). Modeling grid fields instead of modeling grid cells. J. Comput. Neurosci. 47, 43–60. [PubMed: 31286380]
33. Banino A, Barry C, Uria B, Blundell C, Lillicrap T, Mirowski P, Pritzel A, Chadwick MJ, Degris T, Modayil J, et al. (2018). Vector-based navigation using grid-like representations in artificial agents. Nature 557, 429–433. [PubMed: 29743670]
34. Cueva CJ, and Wei XX (2018). Emergence of grid-like representations by training recurrent neural networks to perform spatial localization. Preprint at arXiv. https://arxiv.org/abs/1803.07770.
35. Sorscher B, Mel GC, Ocko SA, Giocomo LM, and Ganguli S (2022). A unified theory for the computational and mechanistic origins of grid cells. Neuron. 10.1016/j.neuron.2022.10.003.
36. McNaughton BL, Battaglia FP, Jensen O, Moser EI, and Moser M-B (2006). Path integration and the neural basis of the "cognitive map". Nat. Rev. Neurosci. 7, 663–678. [PubMed: 16858394]
37. Fiser A, Mahringer D, Oyibo HK, Petersen AV, Leinweber M, and Keller GB (2016). Experience-dependent spatial expectations in mouse visual cortex. Nat. Neurosci. 19, 1658–1664. [PubMed: 27618309]
38. Hok V, Jacob P-Y, Bordiga P, Truchet B, Poucet B, and Save E (2018). A spatial code in the dorsal lateral geniculate nucleus. Preprint at bioRxiv, 473520.
39. Saleem AB, Diamanti EM, Fournier J, Harris KD, and Carandini M (2018). Coherent encoding of subjective spatial position in visual cortex and hippocampus. Nature 562, 124–127. [PubMed: 30202092]
40. Campbell MG, and Giocomo LM (2018). Self-motion processing in visual and entorhinal cortices: inputs, integration, and implications for position coding. J. Neurophysiol. 120, 2091–2106. [PubMed: 30089025]
41. Fournier J, Saleem AB, Diamanti EM, Wells MJ, Harris KD, and Carandini M (2020). Mouse visual cortex is modulated by distance traveled and by theta oscillations. Curr. Biol. 30, 3811–3817.e6. [PubMed: 32763173]
42. Diamanti EM, Reddy CB, Schröder S, Muzzu T, Harris KD, Saleem AB, and Carandini M (2021). Spatial modulation of visual responses arises in cortex with active navigation. Elife 10, e63705. [PubMed: 33538692]
43. Flossmann T, and Rochefort NL (2021). Spatial navigation signals in rodent visual cortex. Curr. Opin. Neurobiol. 67, 163–173. [PubMed: 33360769]
44. Zong W, Obenhaus HA, Skytøen ER, Eneqvist H, de Jong NL, Vale R, Jorge MR, Moser MB, and Moser EI (2022). Large-scale two-photon calcium imaging in freely moving mice. Cell 185, 1240–1256.e30. [PubMed: 35305313]
45. Obenhaus HA, Zong W, Jacobsen RI, Rose T, Donato F, Chen L, Cheng H, Bonhoeffer T, Moser MB, and Moser EI (2022). Functional network topography of the medial entorhinal cortex. Proc. Natl. Acad. Sci. USA 119. e2121655119. [PubMed: 35135885]
46. Chen G, King JA, Burgess N, and O'Keefe J (2013). How vision and movement combine in the hippocampal place code. Proc. Natl. Acad. Sci. USA 110, 378–383. [PubMed: 23256159]
47. Ji D, and Wilson MA (2007). Coordinated memory replay in the visual cortex and hippocampus during sleep. Nat. Neurosci. 10, 100–107. [PubMed: 17173043]
48. Haggerty DC, and Ji D (2015). Activities of visual cortical and hippocampal neurons co-fluctuate in freely moving rats during spatial behaviors. Elife 4, e08902. [PubMed: 26349031]

49. Song HF, Yang GR, and Wang X-J (2016). Training excitatory-inhibitory recurrent neural networks for cognitive tasks: a simple and flexible framework. PLoS Comput. Biol. 12, e1004792. [PubMed: 26928718]
50. Rajakumar A, Rinzel J, and Chen ZS (2021). Stimulus-driven and spontaneous dynamics in excitatory-inhibitory recurrent neural networks for sequence representation. Neural Comput. 33, 2603–2645. [PubMed: 34530451]
51. Xue X, Wimmer RD, Halassa MM, and Chen ZS (2022). Spiking recurrent neural networks represent task-relevant neural sequences in rule-dependent computation. Cognit. Comput. 14. 10.1007/s12559-022-09994-2.
52. Dannenberg H, Lazaro H, Nambiar P, Hoyland A, and Hasselmo ME (2020). Effects of visual inputs on neural dynamics for coding of location and running speed in medial entorhinal cortex. Elife 9, e62500. [PubMed: 33300873]
53. Krupic J, Burgess N, and O'Keefe J (2012). Neural representations of location composed of spatially periodic bands. Science 337, 853–857. [PubMed: 22904012]
54. Narvatilova Z, Godfrey KB, and McNaughton BL (2016). Grids from bands, or bands from grids? An examination of the effects of single unit contamination on grid firing patterns. J. Neurophysiol. 115, 992–1002. [PubMed: 26683071]
55. Sussillo D, and Abbott LF (2009). Generating coherent patterns of activity from chaotic neural networks. Neuron 63, 544–557. [PubMed: 19709635]
56. Chen G, Manson D, Cacucci F, and Wills TJ (2016). Absence of visual input in the disruption of grid cell firing in the mouse. Curr. Biol. 26, 2335–2342. [PubMed: 27498565]
57. Gardner RJ, Hermansen E, Pachitariu M, Burak Y, Baas NA, Dunn BA, Moser MB, and Moser EI (2022). Toroidal topology of population activity in grid cells. Nature 602, 123–128. [PubMed: 35022611]
58. Couey JJ, Witoelar A, Zhang S-J, Zheng K, Ye J, Dunn B, Czajkowski R, Moser MB, Moser EI, Roudi Y, and Witter MP (2013). Recurrent inhibitory circuitry as a mechanism for grid formation. Nat. Neurosci. 16, 318–324. [PubMed: 23334580]
59. Buxhoeveden DP, and Casanova MF (2002). The minicolumn hypothesis in neuroscience. Brain 125, 935–951. [PubMed: 11960884]
60. Litwin-Kumar A, and Doiron B (2012). Slow dynamics and high variability in balanced cortical networks with clustered connections. Nat. Neurosci. 15, 1498–1505. [PubMed: 23001062]
61. Horton JC, and Adams DL (2005). The cortical column: a structure without a function. Philos. Trans. R. Soc. Lond. B Biol. Sci. 360, 837–862. [PubMed: 15937015]
62. Laramée ME, Rockland KS, Prince S, Bronchti G, and Boire D (2013). Principal component and cluster analysis of Layer V pyramidal cells in visual and non-visual cortical areas projecting to the primary visual cortex of the mouse. Cereb. Cortex 23, 714–728. [PubMed: 22426333]
63. Kim R, and Sejnowski TJ (2021). Strong inhibitory signaling underlies stable temporal dynamics and working memory in spiking neural networks. Nat. Neurosci. 24, 129–139. [PubMed: 33288909]
64. Patra M (2018). Multiple attractor bifurcation in three-dimensional piecewise linear maps. Int. J. Bifurcation Chaos 28. 1830032.
65. Song ZG, Xu J, and Zhen B (2019). Mixed-coexistence of periodic orbits and chaotic attractors in an inertial neural system with a nonmonotonic activation function. Math. Biosci. Eng. 16, 6406–6425. [PubMed: 31698569]
66. Cheung A, Ball D, Milford M, Wyeth G, and Wiles J (2012). Maintaining a cognitive map in darkness: the need to fuse boundary knowledge with path integration. PLoS Comput. Biol. 8. e1002651. [PubMed: 22916006]
67. Barry C, Ginzberg LL, O'Keefe J, and Burgess N (2012). Grid cell firing patterns signal environmental novelty by expansion. Proc. Natl. Acad. Sci. USA 109, 17687–17692. [PubMed: 23045662]
68. Liu L, She L, Chen M, Liu T, Lu HD, Dan Y, and Poo M.m. (2016). Spatial structure of neuronal receptive field in awake monkey secondary visual cortex (V2). Proc. Natl. Acad. Sci. USA 113, 1913–1918. [PubMed: 26839410]

69. Miller MW, and Vogt BA (1984). Direct connections of rat visual cortex with sensory, motor, and association cortices. J. Comp. Neurol. 226, 184–202. [PubMed: 6736299]
70. Sanderson KJ, Dreher B, and Gayer N (1991). Prosencephalic connections of striate and extrastriate areas of rat visual cortex. Exp. Brain Res. 85, 324–334. [PubMed: 1716594]
71. Marshel JH, Garrett ME, Nauhaus I, and Callaway EM (2011). Functional specialization of seven mouse visual cortical areas. Neuron 72, 1040–1054. [PubMed: 22196338]
72. Klukas M, Lewis M, and Fiete I (2020). Efficient and flexible representation of higher-dimensional cognitive variables with grid cells. PLoS Comput. Biol. 16. e1007796. [PubMed: 32343687]
73. Yao H, and Li C-Y (2002). Clustered organization of neurons with similar extra-receptive field properties in the primary visual cortex. Neuron 35, 547–553. [PubMed: 12165475]
74. Chen ZS, Zhang X, Long X, and Zhang S-J (2022). Are grid-like representations a component of all perception and cognition? Front. Neural Circuits 16, 924016. [PubMed: 35911570]
75. De Pasquale R, and Sherman SM (2013). A modulatory effect of the feedback from higher visual areas to V1 in the mouse. J. Neurophysiol. 109, 2618–2631. [PubMed: 23446698]
76. Wang HT, Mathur B, and Koch C (1989). Computing optical flow in the primate visual system. Neural Comput. 1, 92–103.
77. Wurtz RH (1998). Optic flow: a brain region devoted to optic flow analysis? Curr. Biol. 8, 554–556. [PubMed: 9601639]
78. Lappe M, and Rauschecker J (1992). Computation of heading direction from optical flow in visual cortex. In Advances in Neural Information Processing Systems 5 (NIPS), Hanson S, Cowan J, and Giles C, eds. (MIT Press).
79. Leinweber M, Ward DR, Sobczak JM, Attinger A, and Keller GB (2017). A sensorimotor circuit in mouse cortex for visual flow predictions. Neuron 95, 1420–1432.e1425. [PubMed: 28910624]
80. Stachenfeld KL, Botvinick MM, and Gershman SJ (2017). The hippocampus as a predictive map. Nat. Neurosci. 20, 1643–1653. [PubMed: 28967910]
81. Recanatesi S, Farrell M, Lajoie G, Deneve S, Rigotti M, and Shea-Brown E (2021). Predictive learning as a network mechanism for extracting low-dimensional latent space representations. Nat. Commun. 12, 1417. [PubMed: 33658520]
82. Dayan P (1993). Improving generalization for temporal difference learning: the successor representation. Neural Comput. 5, 613–624.
83. Gershman SJ (2018). The successor representation: its computational logic and neural substrates. J. Neurosci. 38, 7193–7200. [PubMed: 30006364]
84. Pollock E, and Jazayeri M (2020). Engineering recurrent neural networks from task-relevant manifolds and dynamics. PLoS Comput. Biol. 16. e1008128. [PubMed: 32785228]
85. Savelli F, and Knierim JJ (2019). Origin and role of path integration in the cognitive representations of the hippocampus: computational insights into open questions. J. Exp. Biol. 222. jeb188912. [PubMed: 30728236]
86. Mok RM, and Love BC (2019). A non-spatial account of place and grid cells based on clustering models of concept learning. Nat. Commun. 10, 5685. [PubMed: 31831749]
87. Bicanski A, and Burgess N (2019). A Computational model of visual recognition memory via grid cells. Curr. Biol. 29, 979–990.e4. [PubMed: 30853437]
88. Sussillo D, Churchland MM, Kaufman MT, and Shenoy KV (2015). A neural network that finds a naturalistic solution for the production of muscle activity. Nat. Neurosci. 18, 1025–1033. [PubMed: 26075643]
89. Michaels JA, Dann B, and Scherberger H (2016). Neural population dynamics during reaching are better explained by a dynamical system than representational tuning. PLoS Comput. Biol. 12. e1005175. [PubMed: 27814352]
90. Zhang X, Liu S, and Chen ZS (2021). A geometric framework for understanding dynamic information integration in context-dependent computation. iScience 24, 102919. [PubMed: 34430809]
91. Yamins DLK, Hong H, Cadieu CF, Solomon EA, Seibert D, and DiCarlo JJ (2014). Performance-optimized hierarchical models predict neural responses in higher visual cortex. Proc. Natl. Acad. Sci. USA 111, 8619–8624. [PubMed: 24812127]

92. Sexton NJ, and Love BC (2022). Reassessing hierarchical correspondences between brain and deep networks through direct interface. Sci. Adv. 8. eabm2219. [PubMed: 35857493]

93. Kriegeskorte N (2015). Deep neural networks: a new framework for modeling biological vision and brain information Processing. Annu. Rev. Vis. Sci. 1, 417–446. [PubMed: 28532370]

94. Averna A, Pasquale V, Murphy MD, Rogantin MP, Van Acker GM, Nudo RJ, Chiappalone M, and Guggenmos DJ (2020). Differential effects of open- and closed-Loop intracortical microstimulation on firing patterns of neurons in distant cortical areas. Cereb. Cortex 30, 2879–2896. [PubMed: 31832642]

95. Averna A, Hayley P, Murphy MD, Barban F, Nguyen J, Buccelli S, Nudo RJ, Chiappalone M, and Guggenmos DJ (2021). Entrainment of network activity by closed-loop microstimulation in healthy ambulatory rats. Cereb. Cortex 31, 5042–5055. [PubMed: 34165137]

96. Bridi MCD, Zong FJ, Min X, Luo N, Tran T, Qiu J, Severin D, Zhang XT, Wang G, Zhu ZJ, et al. (2020). Daily oscillation of the excitation-inhibition balance in visual cortical circuits. Neuron 105, 621–629.e4. [PubMed: 31831331]

97. Sussillo D, and Barak O (2013). Opening the black box: low-dimensional dynamics in high-dimensional recurrent neural networks. Neural Comput. 25, 626–649. [PubMed: 23272922]

98. van der Maaten LJP, and Hinton GE (2008). Visualizing data using t-SNE. J. Mach. Learn. Res. 9, 2579–2605.

99. McInnes L, and Healy J (2018). UMAP: uniform manifold approximation and projection for dimension reduction. Preprint at ArXiv. 1802. 03426.

100. Goldman MS (2009). Memory without feedback in a neural network. Neuron 61, 621–634. [PubMed: 19249281]

101. Murphy BK, and Miller KD (2009). Balanced amplification: a new mechanism of selective amplification of neural activity patterns. Neuron 61, 635–648. [PubMed: 19249282]

102. Seung HS (1998). Continuous attractors and oculomotor control. Neural Netw. 11, 1253–1258. [PubMed: 12662748]

103. Asllani M, Lambiotte R, and Carletti T (2018). Structure and dynamical behavior of non-normal networks. Sci. Adv. 4, eaau9403. [PubMed: 30547090]

104. Horn BK, and Schunck BG (1981). Determining optical flow. Artif. Intell. 17, 185–203.

105. Cao L, Varga V, and Chen ZS (2021). Uncovering spatial representations from spatiotemporal patterns of rodent hippocampal field potentials. Cell Rep. Methods 1, 100101. [PubMed: 34888543]

106. Becht E, McInnes L, Healy J, Dutertre C-A, Kwok IWH, Ng LG, Ginhoux F, and Newell EW (2019). Dimensionality reduction for visualizing single-cell data using UMAP. Nat. Biotechnol. 37, 38–44.

Highlights

- Grid patterns emerge in trained RNNs with multisensory inputs
- Grid patterns are robust to the RNN input and network connectivity
- Population responses show emergent ring-like manifolds and attractors
- Grid-like patterns persist in RNNs while performing a non-spatial task

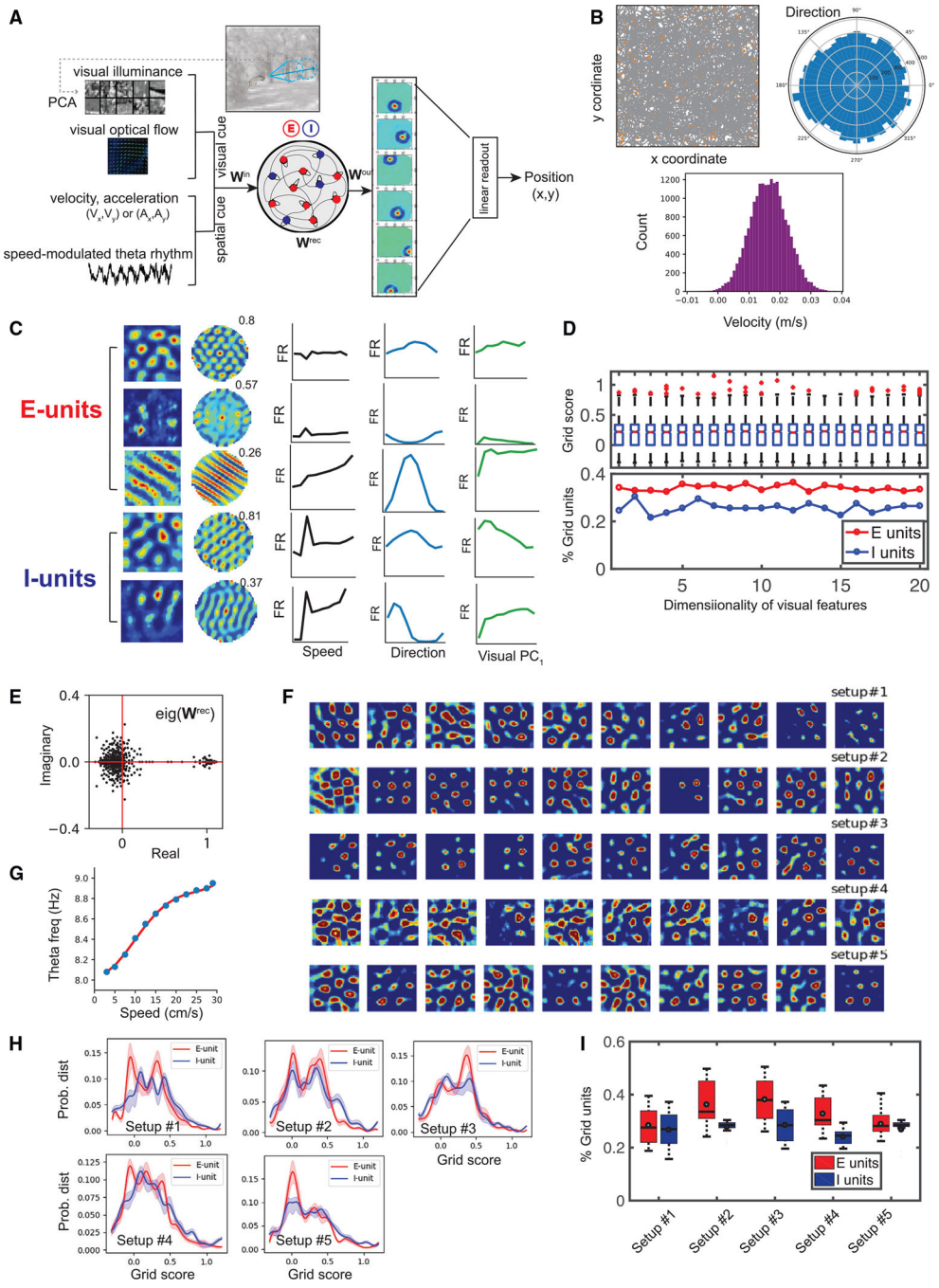


Figure 1. Multisensory input and recurrent dynamics of E/I-RNN produce robust grid patterns

(A) Schematic of the excitatory-inhibitory (E/I)-RNN to perform a 2D spatial navigation task with multisensory cues. The visual cue may appear in the form of principal components (PCs) for the 8 × 8 image patch along the heading angle or the form of visual optical flow. The spatial cue may appear in the form of velocity-(speed and direction), acceleration-, or speed-modulated theta rhythms.

(B) Top left: simulated trajectories (gray curve). Orange dots denote the place cell centers uniformly distributed within the 2D enclosure. Top right and bottom: distributions of simulated run speed and direction statistics.
(C) Examples of emerged grid-like and band-like patterns from excitatory and inhibitory units of the trained RNN (setup #2). First column: firing rate (FR) heatmap; second column: spatial autocorrelogram (the numbers indicate the grid score); third column: speed tuning curve; fourth column: direction tuning curve; fifth and sixth columns: tuning curve with respect to visual illumination PC (PC_1). All tuning curves are in the same scale (a.u.).
(D) Statistics of grid units were relatively stable with respect to the dimensionality of visual features in PCA subspace (1–20).
(E) Complex eigenspectrum of $\mathbf{W}^{rec}$ from a trained E/I-RNN.
(F) Emergent grid-like RNN units with highest grid scores under different input configurations (rows 1–5 corresponded to results from setups #1-#5).
(G) The relationship between theta frequency and run speed, redrawn from Dannenberg et al., 2020.[52]
(H) Distributions of grid scores from the excitatory and inhibitory units under different input configurations. Statistics were generated from 10 trained RNNs in each setup. The shaded area along the curve represents ± SD.
(I) Percentages of grid units in the trained RNNs under different input configurations. Boxplot statistics were generated from 10 trained RNNs in each setup.
See also Figure S1.

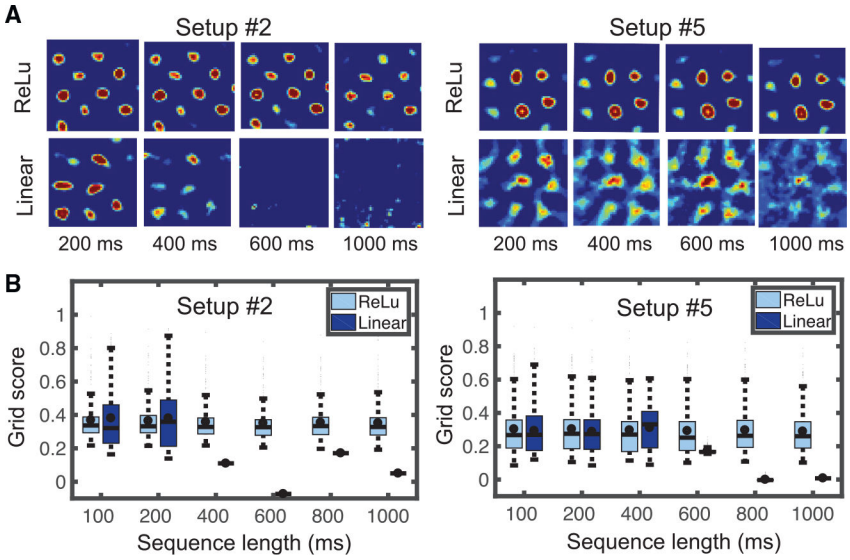


Figure 2. Impact of sequence length on grid patterns in trained RNNs

(A) Grid-like patterns were robust with respect to a wide range of sequence lengths (200, 400, 600, and 1,000 ms) and activation functions in the E/I-RNNs.

(B) Grid score statistics with different sequence lengths. In each condition, the top 50% of grid scores were used for better visualization.

See also Figure S2.

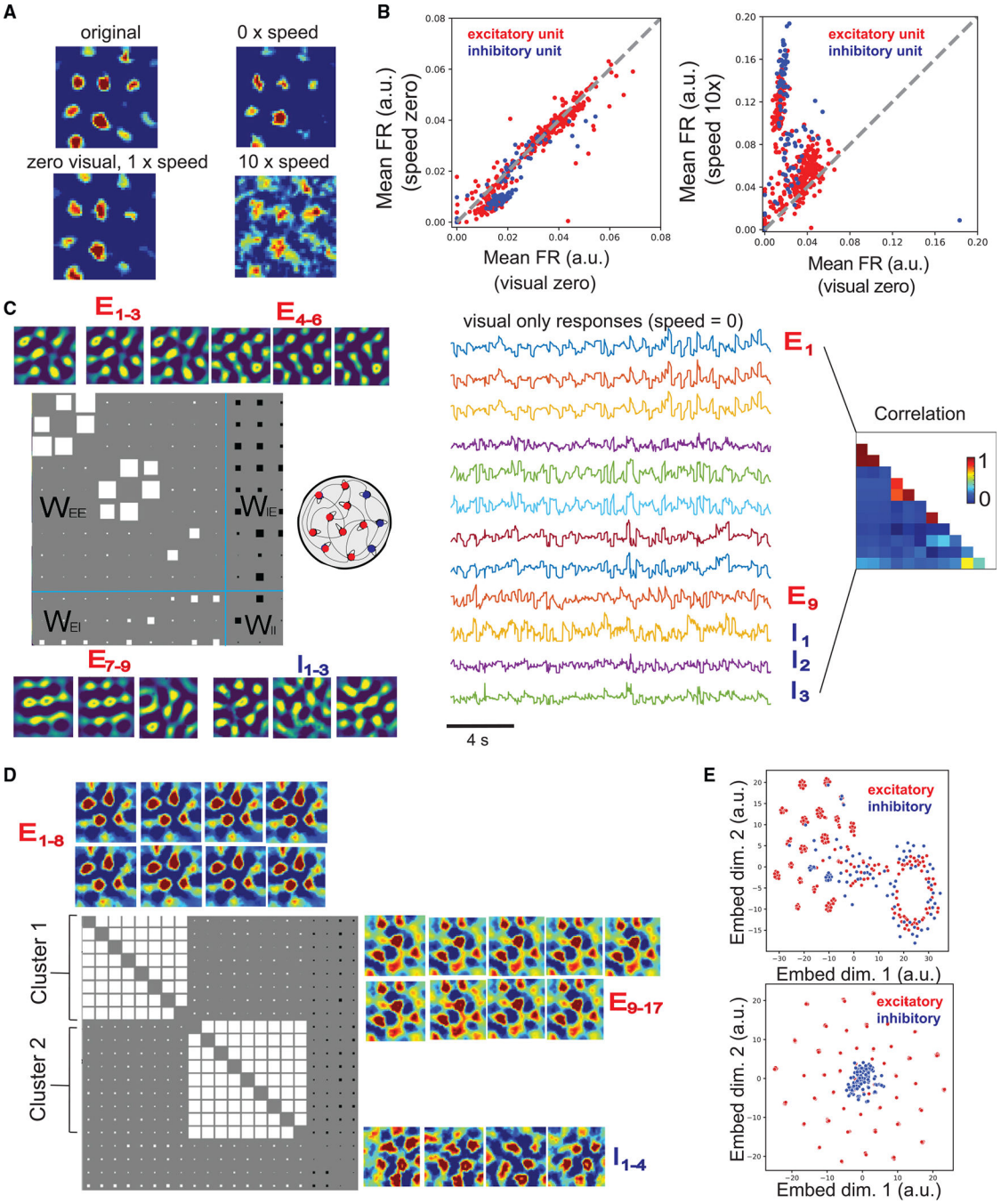


Figure 3. Mixed selectivity of RNN units and emerged functional clusters

(A) Examples of paired grid-like units and their co-activated firing with respect to visual input.

(B) Scatterplots of $\bar{r}_j^{\text{spatial}}$ and $\bar{r}_j^{\text{visual}}$ among the E/I-RNN units (setup #2). Red and blue denote excitatory and inhibitory units, respectively.

(C) From a trained E/I RNN (setup #2), selected 12 (9 excitatory plus 3 inhibitory) grid units and their weight connectivity (white/black square denotes positive/negative synaptic weight; the size of square is proportion to the strength). Temporal traces represent the firing of these

12 units with zero velocity input. The lower triangular matrix denotes pairwise correlation between 12 FR traces.

(D) From a trained linear E/I-RNN (setup #2), selected 21 (17 excitatory plus 4 inhibitory) grid units and their weight connectivity.

(E) 2D embedding of RNN grid fields that were associated with top 60% grid score (GS; setup #2, top: ReLu activation function; bottom: linear activation function). Red and blue denote excitatory and inhibitory grid units, respectively.

See also Figures S3 and S4.

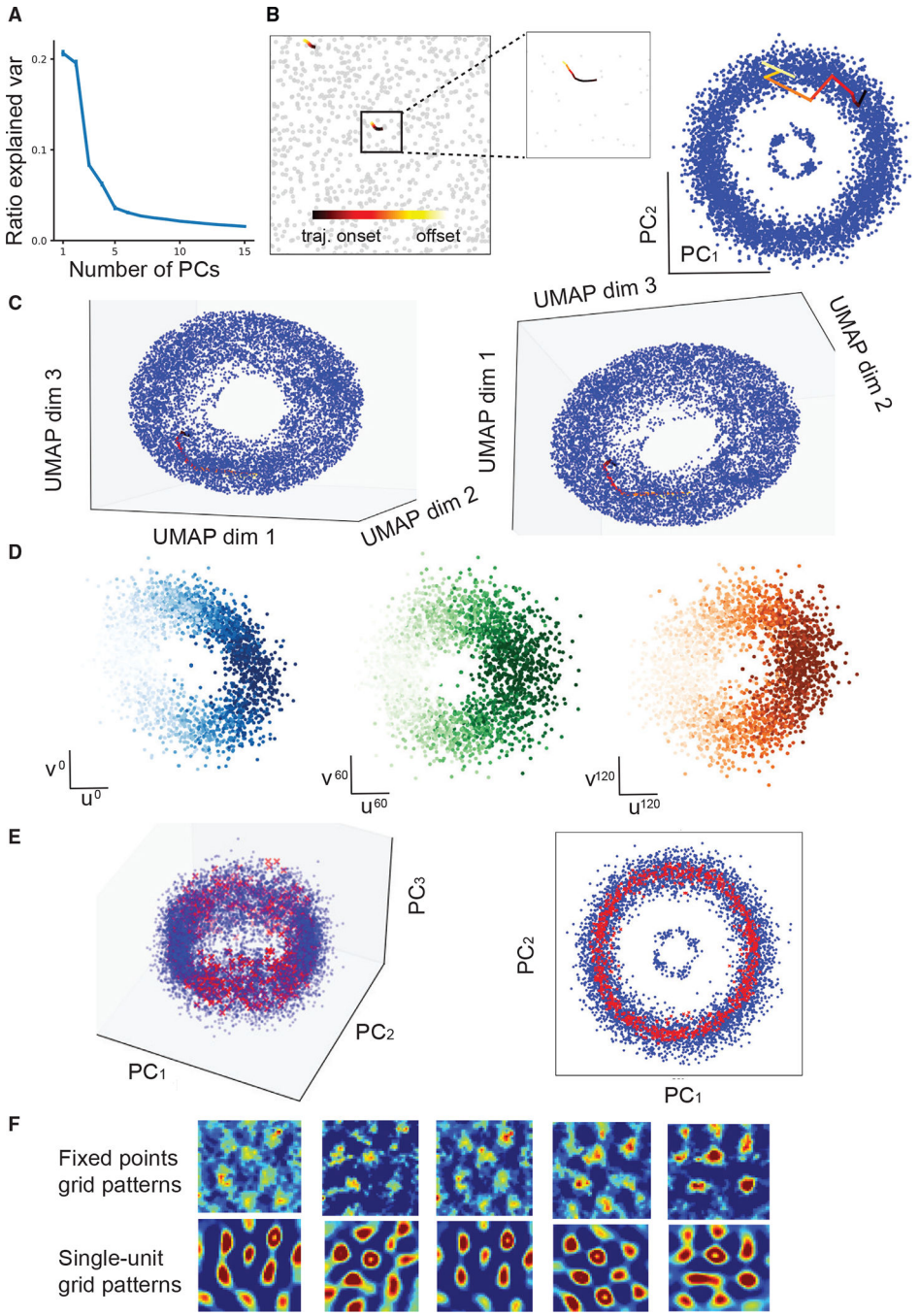


Figure 4. Emerged low-dimensional ring manifold and attractor

(A) PCA revealed the explained variance ratio of trained RNN units. Error bar denotes the SEM from n = 10 trained networks.

(B) Left: 200-ms color-coded simulated trajectory in the 2D enclosure. Right: 2D ring manifold; projection of the high-dimensional RNN population activity onto the first two dominant PCs (PC_1 - PC_2 plane). Each blue dot represents a temporal sample in the simulated trajectory. Neural trajectory was color coded according to the simulated trajectory in the left.

(C) The 3D torus manifold shown from two different angles. The 3D manifold was produced by PCA followed by UMAP (see STAR Methods).
(D) Projections of the 3D manifold onto three pairs of axes.
(E) Visualization of the identified torus-shaped attractor (red crosses represent the fixed points) in 3D (left) and 2D (right) PCA subspaces.
(F) Comparison between five fixed-point grid patterns and selected single-unit grid patterns from the trained E/I-RNN. Notice the close resemblance. Pearson's correlation statistics between top and bottom five panels were (from left to right) 0.828, 0.781, 0.861, 0.841, and 0.827.
See also Figure S5 and Video S1.

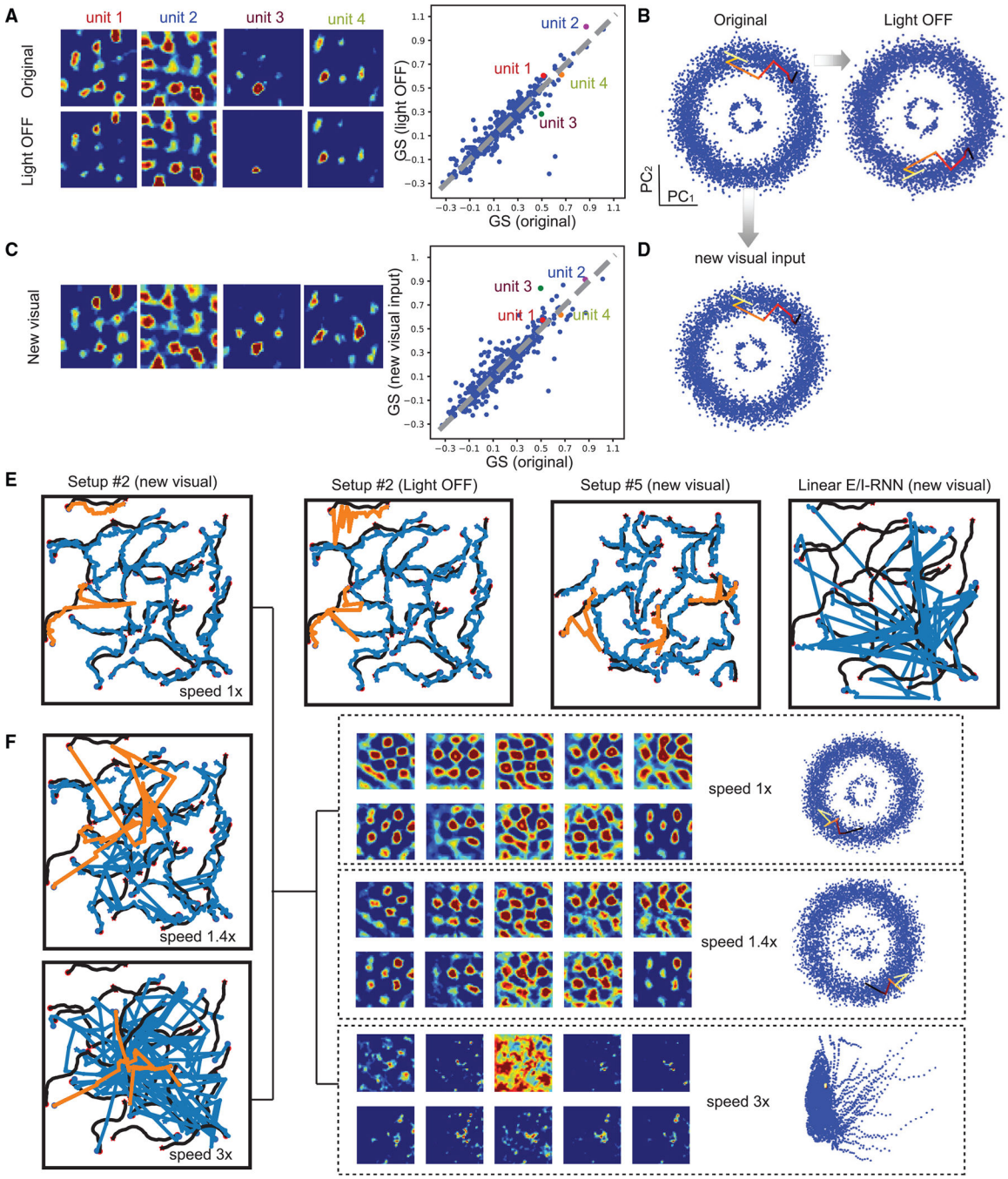


Figure 5. Stability of grid patterns and ring attractor

(A) Comparison of grid unit patterns and GS statistics when the visual input was set to zero (light off) during testing (setup #2). The GS statistics showed no statistical difference (two-sided paired signed-rank test, non-significant [n.s.]).

(B) Comparison of ring attractor and neural trajectories between the original and light-off conditions.

(C) Comparison of grid unit patterns and GS statistics when a new visual input was used in testing while keeping the velocity unchanged (setup #2). The GS statistics showed no statistical difference (two-sided paired signed-rank test, n.s.).
(D) Ring attractor and neural trajectory while using a new visual input.
(E) In testing, the noisy long (sequence length: 50) trajectories either remained stable (blue) or first perturbed then converged (orange) to the simulated paths (black) in the E/I-RNN (ReLu), suggesting the stability of ring attractor. In contrast, the noisy long trajectory tended to deviate from the simulated trajectory in the linear E/I-RNN. Four snapshot examples under different testing conditions are shown. The black curve denotes the simulated path, the overlaid blue curve denotes the predicted position, and the red/blue circle and red/blue star represent the simulated/predicted start and end position, respectively.
(F) Changes in behavioral speed led to changes in ring attractor and grid cell representations. The attractor and grid cells appeared relatively stable until speed was out of the normal range.
See also Figure S6.

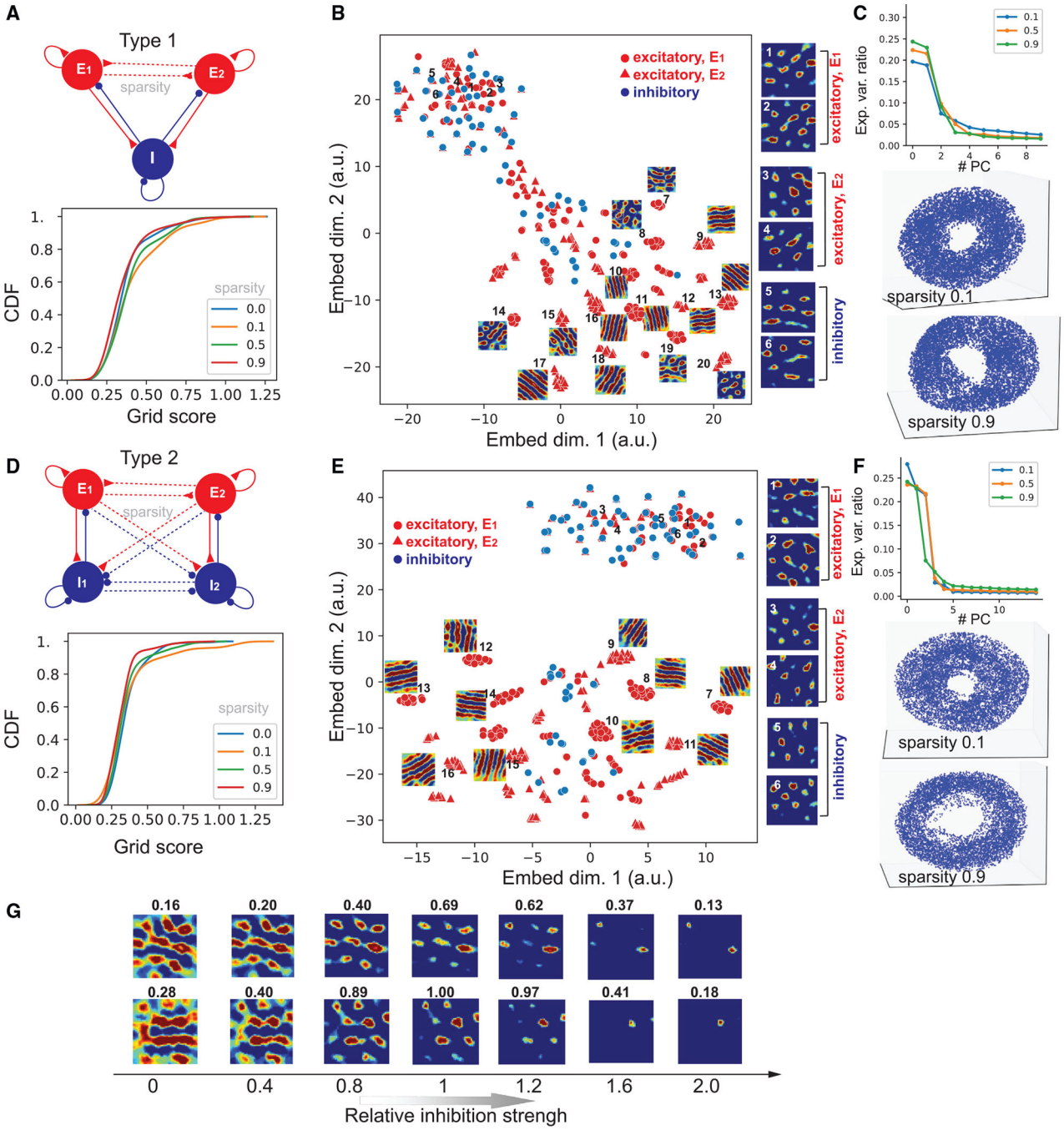


Figure 6. Robust grid patterns with respect to subnetwork connectivity

(A) Distribution of grid scores in type 1 subnetwork connectivity. The dashed line denotes weak connections with various sparsity levels (0, 0.1, 0.5, 0.9). Sparsity level 0 implies full connections (original setting). Excitatory neurons were divided into two subnetworks: E_1 and E_2.

(B) 2D embedding of RNN population responses (units with top 50% grid scores) (sparsity level: 0.5). Representative grid fields are shown from individual functional groups.

(C) PCA explained variance ratio (top) and 3D ring manifold of RNN population responses for two sparsity levels 0.1 and 0.9 in type 1 subnetwork.
(D) Distribution of grid scores in type 2 subnetwork connectivity. Inhibitory neurons were further divided into two subnetworks: I_1 and I_2. The E-to-I, I-to-I, and I-to-E connections were weakly coupled.
(E) Similar to (B) except for type 2 subnetwork (sparsity level: 0.9).
(F) Similar to (C) except for type 2 subnetwork.
(G) Changes of grid-like responses with increasing or decreasing E/I balance. Number on the top of each grid field denotes the grid score.
See also Figures S7 and S8 and Video S2.

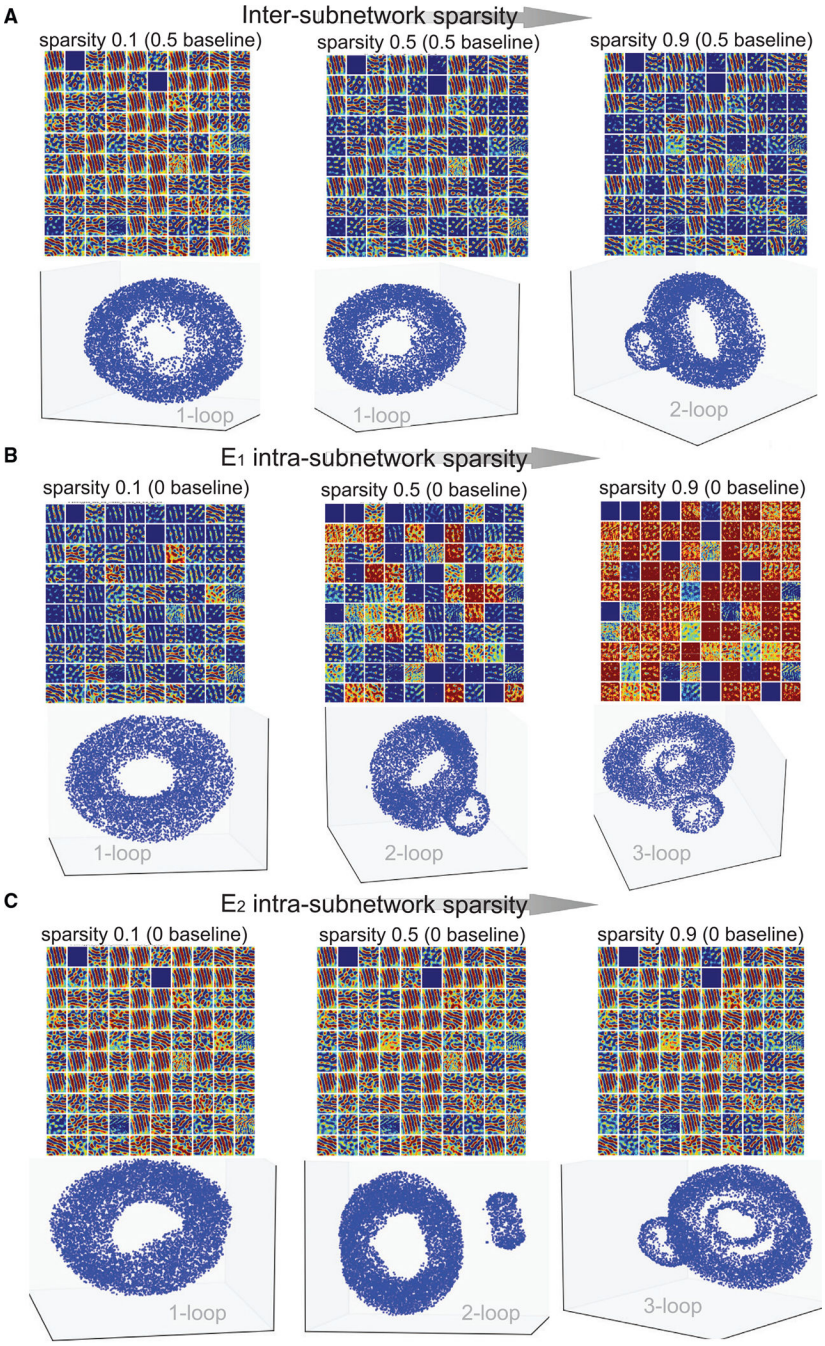


Figure 7. Emerged multistable ring attractors with increasing sparsity in network connectivity

(A) Changes in 100 unit firing patterns of E_1 subnetwork (top) and evolution of 3D manifold structure (bottom) with increasing inter-subnetwork connectivity (type 1, setup #2, sparsity baseline: 0.5).

(B) Changes in 100 unit firing patterns of E_1 subnetwork (top) and evolution of 3D manifold structure (bottom) with increasing E_1 intra-subnetwork connectivity, whereas the E_2 intra-subnetwork connectivity remained unchanged (type 1, setup #2, sparsity baseline: 0).

(C) Changes in 100 unit firing patterns of E_1 subnetwork (top) and evolution of 3D manifold structure (bottom) with increasing E_2 intra-subnetwork connectivity, whereas the E_1 intra-subnetwork connectivity remained unchanged (type 1, setup #2, sparsity baseline: 0).
See also Figures S9 and S10 and Video S3.

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software and algorithm		
Grid score computation	(Banino et al.)[33]	https://github.com/deepmind/grid-cells
E/I RNN	(Song et al.)[49]	https://github.com/frsong/pycog
fixed-point finder	(Sussillo and Barak)[97]	https://github.com/mattgolub/fixed-point-finder
t-SNE	(van der Matten and Hinton)[98]	https://lvdmaaten.github.io/tsne/
UMAP	(McInnes and Healy)[99]	https://github.com/lmcinnes/umap
CNN	Open source	https://github.com/iamkrut/MNIST_handwriting_classification
Deposited code	This study	https://doi.org/10.5281/zenodo.7275282
Other		
MNIST dataset	Open source	http://yann.lecun.com/exdb/mnist/