A GROUP-THEORETIC PERSPECTIVE ON PATH INTEGRATION AND THE ORIGIN OF GRID CELLS

Anonymous authors
Paper under double-blind review

ABSTRACT

We develop a group-theoretic framework predicting that the neural representations supporting path integration reflect the mathematical structure of the space being navigated. Path integration is the process through which an animal estimates its allocentric position by integrating egocentric movements and is thought to be a core computation of the entorhinal–hippocampal system. We formalize path integration as *sequential group composition* over finite groups, with different groups describing different spaces. Using group Fourier analysis we analytically construct recurrent neural networks that perform the computation exactly. These circuits decompose into modules indexed by irreducible representations (irreps) and achieve asymptotically optimal hidden width. For a group coupling translations and rotations in 2D space, our construction reproduces key features of the entorhinal grid-cell code: hexagonal spatial tuning, modules with distinct spatial scales, conjunctive position–heading tuning, and toroidal population geometry. Extending the construction to 3D navigation predicts a grid-cell code with irregular, yet locally structured, spatial tuning built from irreps of a 3D navigation group. Lastly, we introduce *irrep selectivity*, a group-theoretic generalization of gridness that identifies new units aligned to path integration tasks without requiring strict hexagonal symmetry. Taken together, our results provide simple accounts of detailed grid cell phenomenology and new ways for analyzing grid cell data.

1 INTRODUCTION

Navigation is essential for survival: animals must move through the physical world to forage, avoid predators, locate mates, find shelter, and gather resources. While these behaviors are diverse and complex, they rely in part on a core navigational ability: *path integration*.

Path integration is the process by which an animal estimates its allocentric spatial position, without using external landmarks, by integrating egocentric self-motion signals, such as vestibular and proprioceptive input (Mittelstaedt & Mittelstaedt, 1980; McNaughton et al., 2006; Buzsáki & Moser, 2013). Path integration is thought to help an animal track its position and heading as it moves across a 2D surface (Zhang, 1996; Burak & Fiete, 2009), or its position and orientation as it moves through a 3D volume (Finkelstein et al., 2015; Ginosar et al., 2021). Path integration has also been proposed as a mechanism for navigating abstract or conceptual spaces (Constantinescu et al., 2016; Aronov et al., 2017; Theves et al., 2019), where "position" corresponds to a latent cognitive variable and "movement" corresponds to transformations in that space (Whittington et al., 2020; Iyer et al., 2024; Hwang et al., 2026). While the spaces differ, the computation underlying path integration remains the same: composing local, egocentric updates to track a global, allocentric state.

To understand how a neural circuit might implement this computation, it is useful to identify the mathematical properties of these updates. There is always a null update, remaining still, that leaves the state unchanged. Every movement can be reversed by another movement that returns the animal to its previous state. Two movements can be combined into a new movement of the same kind: traveling from $\mathbf{a}$ to $\mathbf{b}$ and then from $\mathbf{b}$ to $\mathbf{c}$ produces the same transformation as traveling from $\mathbf{a}$ to $\mathbf{c}$. Lastly, the cumulative effect of a sequence of movements is independent of how they are associated, although it may depend on their order. A collection of transformations that is closed under composition, contains an identity and inverses, and whose composition is associative, but not necessarily commutative, is precisely the mathematical structure called a *group*.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

In this work, we develop a group-theoretic perspective on path integration. The group plays a dual role: it describes both allocentric states (positions) and the egocentric updates between them (movements). Path integration can then be formalized as sequential group composition, in which successive egocentric group elements update the current allocentric group element. Different spaces correspond to different groups: the 2D rotation group for head direction, the 2D rigid motion group for planar navigation, and the 3D rigid motion group for volumetric navigation. This perspective reveals a unifying algebraic structure underlying path integration across different spaces.

Crucially, the group-theoretic perspective on path integration also suggests a natural organization for the neural circuit that performs it. Every navigation group we consider admits a *group Fourier transform* that decomposes functions on the group into a fundamental basis determined by its *irreducible representations (irreps)*. Using this decomposition, we make several contributions:

1. We construct an exact circuit for path integration with asymptotically optimal hidden width, organized into irrep modules that share a universal recurrent motif across groups (Section 3).
2. For planar navigation, this circuit produces grid-like and conjunctive tuning alongside toroidal population geometry, linking 2D spatial symmetries to grid-cell phenomenology (Section 4).
3. We extend the construction to 3D navigation, predicting a grid-cell code with irregular, yet locally structured, spatial tuning built from irreps of a 3D navigation group (Section 5).
4. We introduce *irrep selectivity*, a group-theoretic generalization of gridness for identifying candidate grid cells that support path integration across physical and conceptual spaces (Section 6).

Overall, our group-theoretic perspective predicts that the neural representations supporting path integration reflect the mathematical structure of the space being navigated.

2 Path Integration as Sequential Group Composition (SGC)

A group G is a set equipped with a binary operation $(g, h) \mapsto gh$ that is associative, has an identity element, and gives every element an inverse (a primer is provided in Section A). As motivated in the introduction, the transformations involved in path integration satisfy these axioms. In our framework, we represent each update (movement) and state (position) as an element of the same group, with different groups describing different spaces.

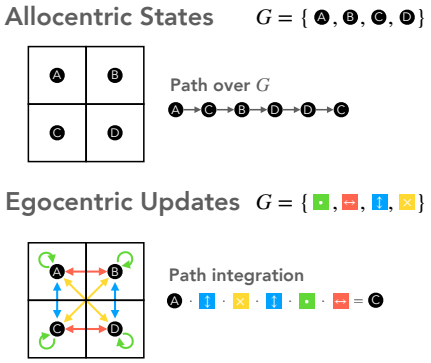


Figure 1: **A four-state arena.** The elements of the group $G = C_2^2$ index both allocentric states and egocentric updates. A path over allocentric states can be described by its initial state and a sequence of egocentric updates, whose composition recovers the final state.

Consider the four-state arena in Fig. 1. The group $G = C_2^2$ describes four possible transformations: remain still, move horizontally, move vertically, and move diagonally. After choosing one corner as an allocentric reference state—for example, the top-left corner—the same four group elements also label the four positions in the arena indexed by the transformation that carries the reference state there: top-left, top-right, bottom-left, and bottom-right. Notice, how we can generalize this example to an $n \times n$ periodic arena using the group $G = C_n^2$.

An element $g \in G$ therefore has a dual interpretation as either an *allocentric state* or an *egocentric update*. Path integration then amounts to composing the group element $g_0 \in G$ indexing the initial allocentric state with a sequence of group elements $g_1, \ldots, g_T \in G$ indexing egocentric updates. The resulting product, under right multiplication, is the element indexing the final allocentric state:

$$(g_0; g_1, \ldots, g_T) \longmapsto g_0 g_1 \cdots g_T \in G. \tag{1}$$

Allocentric and egocentric firing maps. To realize this group-theoretic formulation in neural activity, we consider two neural populations: one representing allocentric states and the other representing egocentric updates. This separation mirrors the organization of hippocampal–entorhinal navigation circuits (Wang et al., 2020) and has precedent in studies of the Drosophila central complex (Lyu et al., 2022; Lu et al., 2022; Mussells Pires et al., 2024).

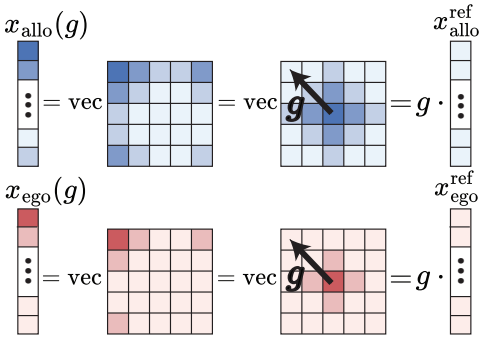


Figure 2: **Firing rates.** For the group $G = C_5^2$ of 2D discrete translations, with $|G| = 25$ neurons in each population. The regular action simply cyclically permutes, vertically and horizontally, the entries of the reference activity vectors of firing rates $x^{\text{ref}}_{\text{allo}}$ and $x^{\text{ref}}_{\text{ego}}$.

We restrict our attention to finite groups and, for analytic simplicity, assume that each population contains $|G|$ neurons, where $|G|$ denotes the number of elements in G. Let $x_{\text{allo}}(g) \in \mathbb{R}^{|G|}$ denote the firing rates of the allocentric population when the animal is at position g. Similarly, let $x_{\text{ego}}(g) \in \mathbb{R}^{|G|}$ represent the firing rates of the egocentric population when the animal makes movement g. Rather than assigning unrelated firing patterns to each group element, we assume that these firing rates are generated from reference activity patterns $x^{\text{ref}}_{\text{allo}}, x^{\text{ref}}_{\text{ego}} \in \mathbb{R}^{|G|}$. Concretely, for each $g \in G$, we define

$$x_{\text{allo}}(g) \coloneqq g \cdot x^{\text{ref}}_{\text{allo}}, \quad x_{\text{ego}}(g) \coloneqq g \cdot x^{\text{ref}}_{\text{ego}}, \tag{2}$$

where $\cdot$ denotes the regular action of G, which permutes a vector's components according to multiplication by g, as illustrated in Fig. 2. This construction captures how localized firing fields are distributed across the space represented by an allocentric cell population, such as for head-direction or place cells (Taube et al., 1990; O'Keefe & Dostrovsky, 1971). However, it idealizes tuning maps as group-transformed copies of a common profile, omitting both biological variation in shape and amplitude (Muller et al., 1987; Schaeffer et al., 2022) and magnitude-dependent gain modulation (Kropff et al., 2015; Spalla et al., 2022). We revisit these modeling choices in Section 7 and discuss how they might be relaxed.

Sequential group composition (SGC). We can now state the algebraic abstraction of path integration studied in this work. Given a finite group G, we cast path integration as *sequential group composition*: from an initial allocentric activity pattern $x_{\text{allo}}(g_0)$ encoding the initial state and a stream of egocentric activity patterns $x_{\text{ego}}(g_1), \ldots, x_{\text{ego}}(g_T)$ encoding successive updates, path integration amounts to tracking the allocentric activity pattern corresponding to the cumulative state,

$$(x_{\text{allo}}(g_0); x_{\text{ego}}(g_1), \ldots, x_{\text{ego}}(g_T)) \longmapsto x_{\text{allo}}(g_0 g_1 \cdots g_T). \tag{3}$$

The group G determines how states and updates compose, while the reference activity patterns $x^{\text{ref}}_{\text{allo}}$ and $x^{\text{ref}}_{\text{ego}}$ determine how the group elements are represented in the neural activity of two populations.

3 An Efficient Recurrent Neural Network that Performs SGC

In this section, we seek a recurrent neural network that can perform sequential group composition, using as few hidden neurons as possible. We consider an RNN given by the update equations

$$\begin{aligned} h^{(1)} &= \sigma\big(\mathbf{W}_{\text{in}}\, x_{\text{allo}}(g_0) + \mathbf{W}_{\text{drive}}\, x_{\text{ego}}(g_1)\big), \\ h^{(t)} &= \sigma\big(\mathbf{W}_{\text{mix}}\, h^{(t-1)} + \mathbf{W}_{\text{drive}}\, x_{\text{ego}}(g_t)\big), \\ y^{(t)} &= \mathbf{W}_{\text{out}}\, h^{(t)}, \end{aligned} \tag{4}$$

where $\sigma(z) = (\max\{0, z\})^2$ is applied elementwise. The parameters $\mathbf{W}_{\text{in}} \in \mathbb{R}^{H \times |G|}$ initializes the hidden state, $\mathbf{W}_{\text{drive}} \in \mathbb{R}^{H \times |G|}$ embeds self-motion updates, $\mathbf{W}_{\text{mix}} \in \mathbb{R}^{H \times H}$ defines the recurrence, and $\mathbf{W}_{\text{out}} \in \mathbb{R}^{|G| \times H}$ decodes the allocentric state. H denotes the number of recurrent neurons. This architecture, depicted in Fig. 3, admits a natural interpretation in entorhinal–hippocampal circuitry: $x^{\text{ref}}_{\text{allo}}$ represents an allocentric place-cell code, $x^{\text{ref}}_{\text{ego}}$ represents egocentric self-motion input, and $h^{(t)}$ represents recurrent activity in medial entorhinal cortex (MEC). Accordingly, the four weight matrices describe inputs to, outputs from, and recurrent connections within MEC. Using the following two ideas we can construct an explicit family of parameters that can perform SGC in an interpretable and optimal manner:

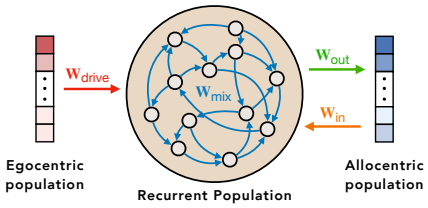


Figure 3: **RNN model.** We study the RNN model depicted here and described in Eq. 4. An initial allocentric state and a stream of egocentric updates drive a recurrent population, whose activity can be decoded to track the cumulative allocentric state.

1. The group Fourier transform. The Fourier transform decomposes a signal into elementary patterns adapted to the symmetries of the space on which it is defined. For example, signals on a circle decompose into sine and cosine waves, or equivalently the harmonics of a cyclic group. The same principle extends to arbitrary finite groups: the elementary symmetry modes are given by the group's *irreducible representations*, or irreps. We denote the set of irreps of a finite group G by $\mathcal{I}(G)$ and the dimension of an irrep $\rho \in \mathcal{I}(G)$ by n_ρ. For a signal $x \in \mathbb{C}^{|G|}$, its Fourier coefficient at ρ is

$$\widehat{x}[\rho] = \sum_{g \in G} \rho(g)^\dagger x[g] \in \mathbb{C}^{n_\rho \times n_\rho}, \qquad \text{where } \rho(g) \in \mathbb{C}^{n_\rho \times n_\rho}. \tag{5}$$

This is the group-theoretic generalization of the classical discrete Fourier transform, recovered when G is cyclic. The coefficients $\widehat{x}[\rho]$ describe the contribution of each symmetry mode to the signal.

2. Associativity. Group composition is associative, which implies the SGC task can be carried out recursively. Rather than directly computing the full product $g_0 g_1 \cdots g_T$, an RNN can maintain a representation of the partial product $g_0 g_1 \cdots g_{t-1}$ and update it with the next input g_t. Thus, each step composes a representation of the current allocentric state with the new egocentric update.

Leveraging the group Fourier transform and associativity, we analytically construct an RNN that can perform SGC exactly. The Fourier transform decomposes the allocentric state and egocentric update into irrep-aligned modes, where composition reduces to matrix multiplication. Our construction assigns hidden neurons to these modes and uses the nonlinearity to carry out the products recursively. The resulting RNN exactly tracks the activity pattern of the cumulative group element:

Theorem 3.1. *Let G be finite, let $\mathcal{S} = \{\rho \in \mathcal{I}(G) \colon \widehat{x}^{\text{ref}}_{\text{allo}} \neq 0\}$, and suppose $\widehat{x}^{\text{ref}}_{\text{ego}}[\rho] \in \mathbb{C}^{n_\rho \times n_\rho}$ is invertible for all $\rho \in \mathcal{S}$. Choose integers $q_\rho \geq 3$ for each $\rho \in \mathcal{S}$, and construct an RNN of the form Eq. (4) with hidden width $H = 4\sum_{\rho \in \mathcal{S}} q_\rho\, n_\rho^3$. Index its hidden neurons by tuples $i = (\rho, \epsilon_1, \epsilon_2, \delta, k_0, k_1, k_2)$, where $\epsilon_1, \epsilon_2 \in \{\pm 1\}$, $\delta \in [q_\rho]$, and $k_0, k_1, k_2 \in [n_\rho]$. Define $\mathbf{W}_{\text{in}}, \mathbf{W}_{\text{drive}}$ and $\mathbf{W}_{\text{out}}$ with rows and columns respectively given by the vectors $u_i, v_i, w_i \in \mathbb{R}^{|G|}$ with elements,*

$$u_i[g] = \epsilon_1 A \operatorname{Re}\left(e^{\mathrm{i}\pi\delta/q_\rho} \rho(g)[k_2, k_0]\right), \tag{6}$$

$$v_i[g] = \epsilon_1 \epsilon_2 A \operatorname{Re}\left(e^{\mathrm{i}\pi\delta/q_\rho} \left(\rho(g) \widehat{x}^{\text{ref}}_{\text{ego}}[\rho]^{-\dagger}\right)[k_1, k_2]\right), \tag{7}$$

$$w_i[g] = \epsilon_2 A \operatorname{Re}\left(e^{\mathrm{i}2\pi\delta/q_\rho} \rho(g)[k_1, k_0]\right). \tag{8}$$

where $A = \sqrt[3]{n_\rho/(q_\rho |G|)}$ and $\mathrm{i} = \sqrt{-1}$ is the imaginary unit. Define $\mathbf{W}_{\text{mix}} = \mathbf{W}_{\text{in}} \mathbf{W}_{\text{out}}$. With these parameters the RNN from Eq. (4) can perform SGC exactly.

See Section C for a proof of Theorem 3.1 and intuition on the 7-tuple indexing of hidden neurons.

Irrep-modular organization. The construction in Theorem 3.1 is organized by the irreducible representations of the navigation group. Each hidden neuron receives allocentric and egocentric inputs through features associated with a single irrep and contributes to the output through that same irrep. Thus, $\mathbf{W}_{\text{in}}$ and $\mathbf{W}_{\text{drive}}$ act as group-specific Fourier encoders, while $\mathbf{W}_{\text{out}}$ decodes these features back into the original population code. Furthermore, because of the orthogonality of inequivalent irreps, $\mathbf{W}_{\text{mix}} = \mathbf{W}_{\text{in}} \mathbf{W}_{\text{out}}$ has recurrent connectivity restricted to neurons associated with the same irrep or its complex conjugate. Specifically, if hidden neurons i and j are associated with irreps ρ_i and ρ_j, respectively, then

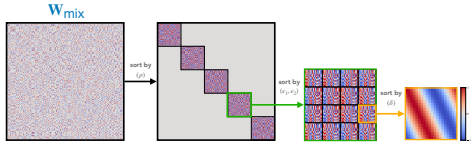


Figure 4: **A universal recurrent motif.** Grouping neurons by conjugate-irrep class $\{\rho, \overline{\rho}\}$ reveals a block-diagonal structure in $\mathbf{W}_{\text{mix}}$. Ordering neurons within a block by sign ϵ_1, ϵ_2 and phase δ exposes a near-circulant cosine motif. Only block dimensions and multiplicities depend on the group. Here we show for a group with 1D irreps.

$$\mathbf{W}_{\text{mix}}[i, j] = 0 \qquad \text{whenever} \qquad \rho_j \notin \{\rho_i, \overline{\rho_i}\}. \tag{9}$$

After grouping complex-conjugate pairs, $\mathbf{W}_{\text{mix}}$ is block diagonal across irreps, with sparse, near-circulant connectivity within each block governed by a phase-dependent cosine kernel, as shown in Fig. 4. We give its exact form and derivation in Corollary C.3. Notably, apart from group-dependent dimensions and multiplicities, this recurrent motif is *universal* for all groups.

Why this solution? At first glance, the construction in Theorem 3.1 may appear highly engineered for analytic tractability, but in fact its structure is very natural from two complementary perspectives.

1. Task-driven perspective. A similar irrep-modular organization emerges when the RNN is optimized to perform SGC. Rather than hand designing the RNN parameters, we could instead initialize them randomly and optimize them to minimize a reconstruction loss over sampled SGC trajectories. Prior work has shown that networks trained in this way by gradient descent converge to closely related Fourier-aligned solutions, with irrep modules emerging sequentially over the course of learning (Chughtai et al., 2023; Kunin et al., 2026; Marchetti et al., 2026).

2. Normative perspective. Our construction is highly neuron-efficient. The SGC task admits many possible exact RNN solutions that need not exploit the irreducible representations of the group. For example, a circuit could represent the full regular representation and directly implement the action of each group element. However, this construction would require $O(|G|^2)$ neurons. By exploiting the irreducible representations, our construction uses $O(|G|)$ neurons for the navigation groups we consider, achieving optimal scaling in the number of neurons. See Section C.3 for a derivation.

Taken together, these perspectives suggest that the solution is not merely analytically convenient: it is mechanistically explicit, naturally learned, and asymptotically optimal in the hidden width. We also discuss how our construction relates to continuous attractor networks in Section C.4.

4 Explaining Grid-Cell Phenomenology in 2D Navigation

Having established an explicit RNN that can perform SGC, we now specialize this construction to 2D navigation and show how it captures the characteristic phenomenology of entorhinal grid cells. Since their discovery, a growing body of evidence has established that grid cells in the MEC subserve path integration in 2D space (Gil et al., 2018). This has shifted attention from *what* grid cells do to *why* the grid-cell code has its particular structure (Dorrell & Whittington, 2026). Grid cells exhibit striking phenomena: cells fire on hexagonal spatial lattices, cells cluster into modules that span distinct spatial scales, cells in the same module share tuning curves up to a translation, some cells are conjunctively tuned to position and heading, and population activity within a module traces a low-dimensional toroidal manifold (Hafting et al., 2005; Sargolini et al., 2006; Stensola et al., 2012; Gardner et al., 2022). Here, we show how these phenomena emerge naturally from our construction.

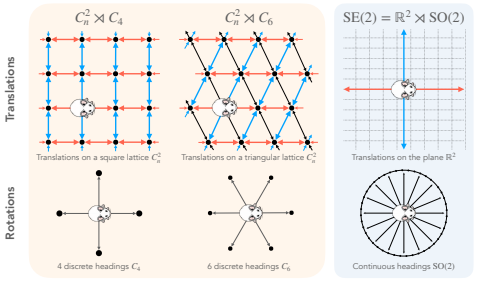


Figure 5: **Semidirect product groups.** Motion in two-dimensional space combines translations with rotations, where the effect of a translation depends on the current heading. This structure is captured by a semidirect product of the translation and rotation groups. On square and triangular lattices this gives $C_n^2 \rtimes C_4$ and $C_n^2 \rtimes C_6$, respectively. In continuous space it gives the special Euclidean group $\mathrm{SE}(2) = \mathbb{R}^2 \rtimes \mathrm{SO}(2)$.

Planar navigation group. For 2D navigation, a natural choice for the navigation group would be C_n^2, discussed in Section 2, which describes translations on a periodic lattice. However, translations alone cannot capture how movement in 2D space depends on heading direction: moving forward while facing east produces a different displacement than moving forward while facing north. We therefore model two-dimensional path integration with the semidirect product group $G = C_n^2 \rtimes C_m$, where C_n^2 represents translations, C_m represents heading, and the semidirect product $\rtimes$ captures how headings act on translations through the composition rule

$$(\tau_1, \theta_1) \cdot (\tau_2, \theta_2) = \big(\tau_1 + R^{\theta_1}\tau_2,\ [\theta_1 + \theta_2]_m\big). \quad (10)$$

Here R is a rotation associated with C_m, and $[\cdot]_m$ denotes modulo m. While n can be any natural number determining the lattice resolution, m is restricted to $m \in \{2, 3, 4, 6\}$ such that R preserves the lattice. This group is the finite analogue of $\mathrm{SE}(2) = \mathbb{R}^2 \rtimes \mathrm{SO}(2)$, the continuous group of 2D rigid motions, as illustrated in Fig. 5. As we will see, the coupling between translations and rotations is essential for our model to reproduce the spatial tuning and population geometry of grid cells.

Simulating a random walk. With the navigation group $G = C_n^2 \rtimes C_m$, we can now place our RNN in a simulated arena and study its activity in the same way an experimental neuroscientist would analyze neural recordings from a navigating animal. We set $m = 6$, corresponding to a triangular lattice, and $n = 21$ setting the resolution. We choose $x_{\text{allo}}^{\text{ref}}$ as a localized Gaussian bump, whose orbit under the group produces tuning resembling that of place cells across the arena. We choose $x_{\text{ego}}^{\text{ref}}$ as a generic reference population code with support across the same irreps as $x_{\text{allo}}^{\text{ref}}$. We generate a random walk by repeatedly sampling and composing local egocentric movements from a small subset of translations and rotations $\mathcal{M} \subseteq G$, as shown in Fig. 6. With G, $x_{\text{allo}}^{\text{ref}}$, and $x_{\text{ego}}^{\text{ref}}$, we instantiate the exact RNN construction of Theorem 3.1. Along the trajectory, we track the RNN's predicted allocentric state and record the activity of its recurrent neurons. This allows us both to verify accurate path integration and to ask whether the resulting neural activity exhibits the characteristics of grid cells. This simulated-experiment perspective follows prior task-driven RNN studies of path integration (Banino et al., 2018; Cueva & Wei, 2018; Sorscher et al., 2023; Schaeffer et al., 2023), and when possible we adopt similar hyperparameter choices (see Section E.1).

Single-cell tuning. In spatial neuroscience, neurons such as place cells, grid cells, and head-direction cells, are characterized by their *tuning curves*: average firing rate as a function of a spatial variable (Moser et al., 2017). We can analyze the recurrent neurons in our model in the same way. For neuron i, we define its tuning curve $r_i(g)$ as its average activity when the allocentric state is $g \in G$, estimated over repeated visits to that state along the simulated trajectory. For sufficiently long trajectories, these empirical averages converge to one-step transition averages that can be computed directly (see Section D). For planar navigation, because $g \in C_n^2 \rtimes C_m$ specifies both position and heading, positional tuning is obtained by averaging $r_i(g)$ over heading, while heading tuning is obtained by averaging over position. These tuning curves inherit the structure of each neuron's associated irrep of G, giving rise to characteristics of grid-cell phenomenology, as shown in Fig. 6.

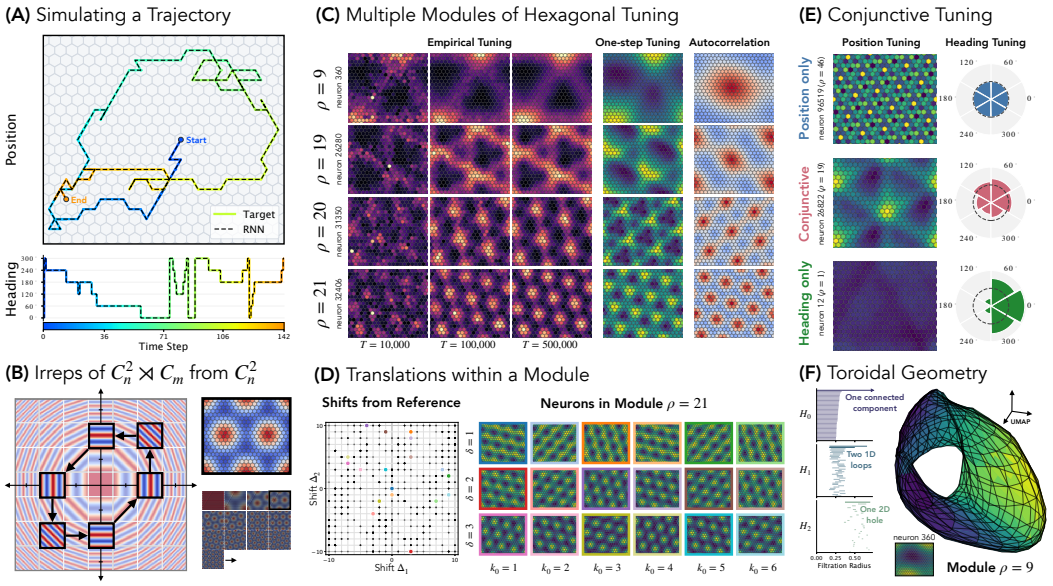


Figure 6: **Grid-cell phenomenology from group structure.** The RNN constructed for $G = C_n^2 \rtimes C_6$ captures key features of grid-cell phenomenology. **(A)** The RNN tracks position (top) and heading (bottom) along a simulated trajectory over long sequence lengths. **(B)** Hexagonal firing patterns reflect the geometry of the irreps of $C_n^2 \rtimes C_6$, which are built from Fourier modes of C_n^2 related by C_6 rotations. **(C)** As trajectory length increases, empirical spatial tuning converges toward directly computable one-step transition averages. Different irreps (rows) exhibit distinct spatial scales. Spatial autocorrelations are shown at right. **(D)** Neurons within an irrep module share spatial scale and lattice orientation but differ in spatial offset. Shifts relative to a reference neuron (left) correspond to translated spatial tuning (right). **(E)** Spatial tuning (left) and heading-tuning curves (right) show three classes of cells: those selective for position alone (top), both position and heading (middle), or heading alone (bottom). **(F)** A UMAP embedding of population activity (right) and persistent homology (left) show toroidal population geometry within an irrep module.

1. *Hexagonal spatial tuning.* Grid cells have positional tuning curves arranged on an approximately hexagonal lattice (Hafting et al., 2005). For $C_n^2 \rtimes C_m$, the irreps are built from Fourier modes of C_n^2 related by rotations in C_m, producing superpositions of rotated spatial waves (see Section B). When $m \in \{3, 6\}$, these modes inherit the triangular symmetry of the lattice, naturally producing hexagonal tuning, as illustrated in Fig. 6. By contrast, each irrep of C_n^2 consists of a single spatial wave, so the same construction yields periodic bands rather than hexagonal lattices. Prior works connecting group theory and grid cells used translation-only groups like C_n^2, and thus needed additional constraints to obtain hexagonal tuning (Gao et al., 2019; 2021; Dorrell et al., 2023).
2. *Multiple modules with distinct spatial scales.* The entorhinal grid code contains discrete modules with different characteristic spatial scales (Stensola et al., 2012). In our construction, different irreps correspond to different spatial frequencies and therefore different spatial scales. Which modules are expressed is determined by the Fourier content of $x_{\text{allo}}^{\text{ref}}$, while the block-diagonal recurrent connectivity across irreps keeps these modules dynamically distinct. This separation of the recurrent dynamics parallels the anatomical clustering of modules in MEC (Gu et al., 2018).
3. *Translated tuning curves within a module.* Grid cells within a module share spatial scale and lattice orientation while differing primarily in spatial phase (Hafting et al., 2005; Yoon et al., 2013). The same structure appears in our RNN: neurons associated with a common irrep share the same underlying spatial pattern, while the phase parameters of the construction translate that pattern across the arena. A single irrep therefore gives rise to a module of neurons with matched lattice geometry but different spatial offsets, as illustrated in Fig. 6D.
4. *Conjunctive position–heading tuning.* Many medial entorhinal neurons are jointly selective for position and heading direction (e.g., Sargolini et al. (2006); Gerlei et al. (2020)). This dependence appears naturally in the full tuning curve $r_i(g)$. Because the irreps of the non-Abelian group $C_n^2 \rtimes C_m$ jointly encode translation and rotation, a neuron's response to position can depend on heading. Conjunctive position–heading tuning therefore emerges directly from the coupled representation of position and orientation in the planar navigation group.
5. *Toroidal population geometry.* The same irrep structure that organizes single-cell tuning also shapes the joint activity of the population. Within a grid-cell module, population activity traces a low-dimensional toroidal manifold as the animal moves through space (Gardner et al., 2022). In our construction, neurons within an irrep module differ in spatial phase while sharing the same underlying periodic structure. As position varies over the two periodic coordinates of C_n^2, their joint activity traces a two-dimensional periodic orbit that approaches a torus, $\mathbb{T}^2$, in the continuous limit. This geometry is illustrated in Fig. 6F and confirmed with persistent homology.

Overall, we have shown how a single recurrent construction for composing translations and rotations in 2D space accounts for many of the single-cell and population phenomena of grid cells.

5 Predictions for the Grid-Cell Code in 3D Navigation

The grid-cell code for 3D navigation remains substantially less understood than in two dimensions. Early theories based on efficient coding predicted that 3D grid fields would form regular close-packed lattices (Wei et al., 2015; Mathis et al., 2015). Experimental findings have instead revealed a complex picture. In rats exploring a volumetric lattice, grid cells exhibit stable 3D firing fields, but these fields are larger, sparser, and more irregularly arranged than in two dimensions (Grieves et al., 2021). Recordings from freely flying bats similarly found multifield neurons with characteristic local spacing but no global 3D lattice (Ginosar et al., 2021). Recently, large-scale recordings from freely flying bats found that co-modular grid-cell populations retain a 2D toroidal organization during 3D navigation, with activity aligned to approximately planar flight trajectories (Qi & Yartsev, 2026). Together, these findings

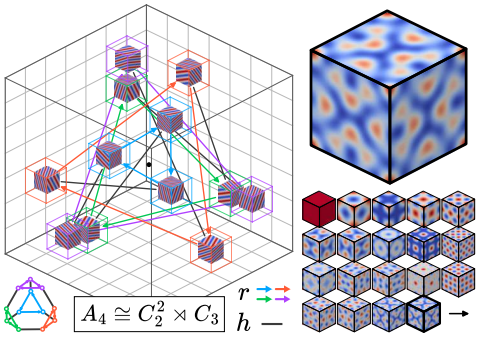


Figure 7: **Irrep of $C_n^3 \rtimes A_4$.** An A_4-orbit of 3D plane waves generated by rotation $r \in C_3$ and half turn $h \in C_2^2$, with Cayley graph as reference (left) gives a 3D spatial pattern through superposition (top right). Different orbits yield distinct patterns (bottom right). See Section B for mathematical details.

suggest that the grid-cell code remains highly structured in 3D navigation despite lacking a regular lattice of firing fields. This raises a natural question: what determines the spatial organization of these fields? Our framework predicts it should reflect the spatial symmetries of 3D space.

Volumetric navigation is associated with the rigid motion group $\mathrm{SE}(3) = \mathbb{R}^3 \rtimes \mathrm{SO}(3)$. We consider finite analogues $C_n^3 \rtimes A_4$ and $C_n^3 \rtimes S_4$, which couple discrete translations to the rotational symmetries of the tetrahedron and octahedron, respectively. As in two dimensions, the structure of each irrep is built from an orbit of plane waves under the rotation group. Superposing these waves produces spatial patterns whose geometry depends on their wavevectors, amplitudes, and phases, as shown in Fig. 7. Our framework predicts that the grid-cell code should reflect these patterns. While some superpositions produce lattice-like patterns, others yield more irregular firing fields resembling those observed in rats and bats (Grieves et al., 2021; Ginosar et al., 2021). Whether this construction can also capture the 2D toroidal population geometry observed in bats remains open (Qi & Yartsev, 2026). Nevertheless, it provides a concrete mathematical target for experimental analysis: spatial tuning of 3D grid cells are organized around the irreducible representations of finite analogues of $\mathrm{SE}(3)$. In the next section, we introduce a measure to detect this structure in neural activity.

6 Irrep Selectivity: Measuring Navigational Tuning

Neurons are commonly identified by their selectivity: place cells by position, head-direction cells by heading, grid cells by periodic spatial location. Yet many neurons in MEC exhibit reliable but heterogeneous tuning that resists these categories (Hardcastle et al., 2017; Mallory et al., 2021; Nayebi et al., 2021). Population-level approaches reveal coherent navigational structure in collective activity, even when single-cell tuning is irregular (Gardner et al., 2022; Qi & Yartsev, 2026). Here we pursue a complementary single-cell approach that remains applicable when few neurons are recorded simultaneously. Our framework motivates *irrep selectivity*, a score measuring how strongly a neuron's tuning concentrates on a few irreps of the navigation group. We can then use irrep selectivity to identify cells supporting path integration from their neural activity.

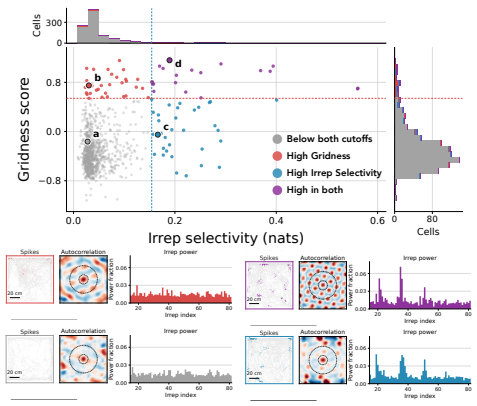


Figure 8: **Irrep selectivity versus gridness in freely moving MEC recordings.** Each point represents one neuron, plotted by its irrep selectivity (horizontal axis) and gridness score (vertical axis). Dashed lines mark the 95th percentiles, defining descriptive groups: below both cutoffs (grey), high gridness only (red), high irrep selectivity only (blue), and high on both (purple). Outlined points (a), (b), (c), (d) identify example neurons, arranged to match the scatter quadrants. For each example, we show the trajectory (grey), spike locations (matching group color), spatial autocorrelogram, and normalized power over the included irreps. See Section E.2 for experimental details.

To compute the score, we treat each neuron's tuning as a signal on the navigation group and use the group Fourier transform to decompose its power across irreps. For example, for 2D navigation, we estimate mean activity in $n \times n$ position bins and m heading bins, obtaining a response pattern in $\mathbb{R}^{n^2 \times m}$ that we interpret as a signal on $C_n^2 \rtimes C_m$. Normalizing the irrep power spectrum gives a probability distribution p, from which we define *irrep selectivity* as,

$$\mathrm{IS} = H_{\max} - H(p). \tag{11}$$

Here $H(p)$ is the entropy and $H_{\max}$ its maximum over the included irreps. The score is high when power concentrates in a few irreps and low when it is broadly distributed. The same procedure extends to other navigation groups once the relevant states and transformations are specified.

Canonical grid cells should exhibit both high gridness and high irrep selectivity, reflecting their concentration on rotationally related Fourier modes. However, band-like (Krupic et al., 2012), conjunctive, or apparently irregular tuning may also exhibit high irrep selectivity despite low gridness. Such cells are candidates for navigation-related representations that conventional geometric classifications miss. Irrep selectivity thus provides a group-theoretic generalization of gridness, extending the search from hexagonal firing patterns to neural tuning aligned with the symmetries of the physical or conceptual spaces being navigated.

7 DISCUSSION

In this work, we used a *group-theoretic perspective on path integration* to construct an RNN that solves sequential group composition with asymptotically optimal hidden width (Section 3), show how this circuit captures key features of grid-cell phenomenology (Section 4), predict properties of the grid-cell code in three-dimensional navigation (Section 5), and introduce irrep selectivity to detect spatially periodic tuning (Section 6). In this section, we situate our framework among existing theoretical models of path integration, discuss its limitations, and outline directions for future work.

Relationship to existing models. Existing models of grid cells and path integration broadly emphasize one of three complementary perspectives: mechanistic, normative, or task-driven. *Mechanistic models* specify connectivity and dynamics that support path integration and grid-like activity. For example, continuous attractor networks use recurrent dynamics to maintain stable estimates of heading direction or allocentric position (Fuhs & Touretzky, 2006; Burak & Fiete, 2009; Yoon et al., 2013; Khona & Fiete, 2022). However, because their connectivity is typically hand-designed, these models do not by themselves explain why a particular circuit should arise or if it is optimal in some way. *Normative models* explain grid-cell representations as optimal solutions to spatial coding and path integration tasks under biological constraints. They provide general requirements for path-integrating systems (Issa & Zhang, 2012), relate grid-cell codes to algebraic structure (Gao et al., 2019; Dorrell et al., 2023; Xu et al., 2025a;b), and analyze their spatial coding efficiency (Fiete et al., 2008; Mathis et al., 2012; Sreenivasan & Fiete, 2011; Kymn et al., 2024). These models explain the advantages of certain representations, but often leave their circuit dynamics unspecified. *Task-driven models* train RNNs on simulated path integration tasks and study when grid-like representations emerge (Banino et al., 2018; Cueva & Wei, 2018; Sorscher et al., 2023; Schaeffer et al., 2023; Chu et al., 2025; Redman et al., 2024; Acosta et al., 2024). These models link emergent representations to behavioral objectives and create candidate circuits. However, their learned solutions are often difficult to study analytically, obscuring which features reflect the task versus modeling choices (Schaeffer et al., 2022). Our model combines aspects of each: it provides an explicit circuit construction (mechanistic), establishes its asymptotically optimal hidden width (normative), and connects its irrep-modular organization to networks found via gradient-based training (task-driven). By unifying these perspectives, our framework avoids key limitations of each approach in isolation.

Limitations and extensions of our model. Our framework has three main limitations, each suggesting a natural direction for future work. *Continuous navigation.* Our analysis focuses on finite navigation groups, whereas natural navigation unfolds over continuous transformation groups such as $\mathrm{SE}(2)$ and $\mathrm{SE}(3)$. One extension is to retain the continuous group while replacing its regular representation with a finite spectral representation supported on selected irreps. *Neural heterogeneity and gain modulation.* We model allocentric tuning maps as group-transformed copies of a shared profile and egocentric self-motion as discrete group elements. This omits observed heterogeneity in tuning shape and amplitude (Muller et al., 1987; Schaeffer et al., 2022), as well as magnitude-dependent gain modulation by speed or angular velocity (Kropff et al., 2015; Spalla et al., 2022). Prior work has shown that heterogeneity can augment neuronal computation (Dahmen et al., 2026) and be productively related to idealized models (Clark et al., 2025; Hulse et al., 2026). Therefore, extending our theory to incorporate these neural phenomena remains an exciting area for future work. *Naturalistic trajectories.* Our construction solves path integration for any arbitrary sequence of group elements, whereas natural trajectories are structured, nonuniform, and constrained by the animal's body and environment. Thus, our construction may require greater representational capacity than animals need for path integration. Indeed, prior work suggests that trajectory statistics can predict the allocation of grid cells to modules (Mosheiff et al., 2017). Thus, training networks on the statistics of empirically observed trajectories (e.g., as in George et al. (2024)) could reveal whether fewer irrep modules suffice and how experiential biases reshape neural tuning and connectivity.

Conclusion. Overall, our work points toward a general organizing principle for understanding neural navigation. Path integration is fundamentally a problem of composing transformations, and the structure of those transformations provides a principled starting point for understanding the neural representations and circuits that support it. Rather than treating grid-like, modular, and conjunctive responses as specializations of the brain's navigation system, our group-theoretic perspective predicts that they simply reflect the mathematical structure of the spaces being navigated.

AI USE STATEMENT

In this work, we used generative AI tools for the following tasks with required disclosure: help developing theoretical models or conceptual frameworks, assist in the writing of proofs, implement methods, clean and reformat datasets. We have not used generative AI tools for these tasks with required disclosure: formulate mathematical claims, provide critical ingredients for proving mathematical claims, propose or refine hypotheses, design or provide feedback on research methodology or experiments, support qualitative and thematic data analysis, interpret results, and the tasks of generating synthetic datasets, and assisting with translation are not applicable to this work. Additionally, we used generative AI tools for these tasks with recommended disclosure: create or modify scientific figures or images, suggest experimental parameters, create or edit software code, create artifacts, summarize or analyse existing literature, and draft parts of the research paper. We have reviewed all AI-assisted work. We have reviewed every LLM-generated sentence and performed several rounds of human text editing. All LLM-generated code was reviewed by at least one researcher. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

REPRODUCIBILITY STATEMENT

The code for reproducing all experiments and figures is available in an anonymized repository at `https://anonymous.4open.science/r/grids-and-groups/`. Details on theory including proofs are given in Appendix.

REFERENCES

Francisco Acosta, Fatih Dinc, William T Redman, Manu Madhav, David Klindt, and Nina Miolane. Global distortions from local rewards: Neural coding strategies in path-integrating neural systems. *Advances in Neural Information Processing Systems*, 37:100241–100269, 2024.

Edward H Adelson and James R Bergen. Spatiotemporal energy models for the perception of motion. *Journal of the optical society of america A*, 2(2):284–299, 1985.

Dmitriy Aronov, Rhino Nevers, and David W Tank. Mapping of a non-spatial dimension by the hippocampal–entorhinal circuit. *Nature*, 543(7647):719–722, 2017.

Andrea Banino, Caswell Barry, Benigno Uria, Charles Blundell, Timothy Lillicrap, Piotr Mirowski, Alexander Pritzel, Martin J Chadwick, Thomas Degris, Joseph Modayil, et al. Vector-based navigation using grid-like representations in artificial agents. *Nature*, 557(7705):429–433, 2018.

Yoram Burak and Ila R Fiete. Accurate path integration in continuous attractor network models of grid cells. *PLoS computational biology*, 5(2):e1000291, 2009.

György Buzsáki and Edvard I Moser. Memory, navigation and theta rhythm in the hippocampal-entorhinal system. *Nature neuroscience*, 16(2):130–138, 2013.

Tianhao Chu, Yuling Wu, Neil Burgess, Zilong Ji, and Si Wu. Unfolding the black box of recurrent neural networks for path integration. *bioRxiv*, pp. 2025–10, 2025.

Bilal Chughtai, Lawrence Chan, and Neel Nanda. A toy model of universality: Reverse engineering how networks learn group operations. In *International Conference on Machine Learning*, pp. 6243–6267. PMLR, 2023.

David G Clark, LF Abbott, and Haim Sompolinsky. Symmetries and continuous attractors in disordered neural circuits. *bioRxiv*, pp. 2025–01, 2025.

Alexandra O Constantinescu, Jill X O'Reilly, and Timothy EJ Behrens. Organizing conceptual knowledge in humans with a gridlike code. *Science*, 352(6292):1464–1468, 2016.

Madison Cotteret, Christopher J Kymn, Hugh Greatorex, Martin Ziegler, Elisabetta Chicca, and Friedrich T Sommer. High-resolution spatial memory requires grid-cell-like neural codes. *arXiv preprint arXiv:2507.00598*, 2025.

Christopher J Cueva and Xue-Xin Wei. Emergence of grid-like representations by training recurrent neural networks to perform spatial localization. *arXiv preprint arXiv:1803.07770*, 2018.

David Dahmen, Axel Hutt, Giacomo Indiveri, Ann Kennedy, Jeremie Lefebvre, Luca Mazzucato, Adilson E Motter, Rishikesh Narayanan, Melika Payvand, Henrike Planert, et al. How heterogeneity shapes dynamics and computation in the brain. *Neuron*, 114(5):804–819, 2026.

Will Dorrell, Peter E. Latham, Timothy E. J. Behrens, and James C. R. Whittington. Actionable neural representations: Grid cells from minimal constraints. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=xfqDe72zh41.

William Dorrell and James CR Whittington. If grid cells are the answer, what is the question? a review of normative grid cell theory. *arXiv preprint arXiv:2601.12424*, 2026.

Gidon Felsen, Jon Touryan, Feng Han, and Yang Dan. Cortical sensitivity to visual features in natural scenes. *PLoS biology*, 3(10):e342, 2005.

Ila R Fiete, Yoram Burak, and Ted Brookings. What grid cells convey about rat location. *Journal of Neuroscience*, 28(27):6858–6871, 2008.

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

Gerald B Folland. *A course in abstract harmonic analysis*, volume 29. CRC press, 2016.

Mark C Fuhs and David S Touretzky. A spin glass model of path integration in rat medial entorhinal cortex. *Journal of Neuroscience*, 26(16):4266–4276, 2006.

Ruiqi Gao, Jianwen Xie, Song-Chun Zhu, and Ying Nian Wu. Learning grid cells as vector representation of self-position coupled with matrix representation of self-motion. In *International Conference on Learning Representations*, 2019. URL https://openreview.net/forum?id=Syx0Mh05YQ.

Ruiqi Gao, Jianwen Xie, Xue-Xin Wei, Song-Chun Zhu, and Ying Nian Wu. On path integration of grid cells: group representation and isotropic scaling. *Advances in Neural Information Processing Systems*, 34:28623–28635, 2021.

Richard J Gardner, Erik Hermansen, Marius Pachitariu, Yoram Burak, Nils A Baas, Benjamin A Dunn, May-Britt Moser, and Edvard I Moser. Toroidal topology of population activity in grid cells. *Nature*, 602(7895):123–128, 2022.

Tom M George, Mehul Rastogi, William de Cothi, Claudia Clopath, Kimberly Stachenfeld, and Caswell Barry. Ratinabox, a toolkit for modelling locomotion and neuronal activity in continuous environments. *Elife*, 13:e85274, 2024.

Klara Gerlei, Jessica Passlack, Ian Hawes, Brianna Vandrey, Holly Stevens, Ioannis Papastathopoulos, and Matthew F Nolan. Grid cells are modulated by local head direction. *Nature communications*, 11(1):4228, 2020.

Mariana Gil, Mihai Ancau, Magdalene I Schlesiger, Angela Neitz, Kevin Allen, Rodrigo J De Marco, and Hannah Monyer. Impaired path integration in mice with disrupted grid cell firing. *Nature neuroscience*, 21(1):81–91, 2018.

Gily Ginosar, Johnatan Aljadeff, Yoram Burak, Haim Sompolinsky, Liora Las, and Nachum Ulanovsky. Locally ordered representation of 3d space in the entorhinal cortex. *Nature*, 596 (7872):404–409, 2021.

Roddy M Grieves, Selim Jedidi-Ayoub, Karyna Mishchanchuk, Anyi Liu, Sophie Renaudineau, Éléonore Duvelle, and Kate J Jeffery. Irregular distribution of grid cell firing fields in rats exploring a 3d volumetric space. *Nature neuroscience*, 24(11):1567–1573, 2021.

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

Torkel Hafting, Marianne Fyhn, Sturla Molden, May-Britt Moser, and Edvard I Moser. Microstructure of a spatial map in the entorhinal cortex. *Nature*, 436(7052):801–806, 2005.

Kiah Hardcastle, Niru Maheswaranathan, Surya Ganguli, and Lisa M Giocomo. A multiplexed, heterogeneous, and adaptive code for navigation in medial entorhinal cortex. *Neuron*, 94(2): 375–387, 2017.

Brad K Hulse, P.B. Aneesh, Sandro Romani, Vivek Jayaraman, and Ann M Hermundstad. Hidden symmetries in network connectivity support ring attractor dynamics in the fly's neural compass. *bioRxiv*, 2026. doi: 10.64898/2026.05.18.725766. URL https://www.biorxiv.org/content/early/2026/05/20/2026.05.18.725766.

Jaedong Hwang, Sujaya Neupane, Mehrdad Jazayeri, and Ila Fiete. Structural generalization and continual learning enabled by factorized entorhinal-hippocampal memory and entorhinal-parietal action circuits. *bioRxiv*, pp. 2026–08, 2026.

Aapo Hyvärinen and Patrik Hoyer. Emergence of phase-and shift-invariant features by decomposition of natural images into independent feature subspaces. *Neural computation*, 12(7):1705–1720, 2000.

John B Issa and Kechen Zhang. Universal conditions for exact path integration in neural systems. *Proceedings of the National Academy of Sciences*, 109(17):6716–6720, 2012.

Abhiram Iyer, Sarthak Chandra, Sugandha Sharma, and Ila Fiete. Flexible mapping of abstract domains by grid cells via self-supervised extraction and projection of generalized velocity signals. *Advances in Neural Information Processing Systems*, 37:85441–85466, 2024.

Mikail Khona and Ila R Fiete. Attractor and integrator networks in the brain. *Nature Reviews Neuroscience*, 23(12):744–766, 2022.

Emilio Kropff, James E Carmichael, May-Britt Moser, and Edvard I Moser. Speed cells in the medial entorhinal cortex. *Nature*, 523(7561):419–424, 2015.

Julija Krupic, Neil Burgess, and John O'Keefe. Neural representations of location composed of spatially periodic bands. *Science*, 337(6096):853–857, 2012.

Daniel Kunin, Giovanni Luca Marchetti, Feng Chen, Dhruva Karkada, James Simon, Michael Deweese, Surya Ganguli, and Nina Miolane. Alternating gradient flows: A theory of feature learning in two-layer neural networks. *Advances in Neural Information Processing Systems*, 38:4377–4424, 2026.

Christopher J Kymn, Sonia Mazelet, Anthony Thomas, Denis Kleyko, Edward Paxon Frady, Friedrich T Sommer, and Bruno A Olshausen. Binding in hippocampal-entorhinal circuits enables compositionality in cognitive maps. *Advances in neural information processing systems*, 37:39128–39157, 2024.

Jenny Lu, Amir H Behbahani, Lydia Hamburg, Elena A Westeinde, Paul M Dawson, Cheng Lyu, Gaby Maimon, Michael H Dickinson, Shaul Druckmann, and Rachel I Wilson. Transforming representations of movement from body-to world-centric space. *Nature*, 601(7891):98–104, 2022.

Cheng Lyu, LF Abbott, and Gaby Maimon. Building an allocentric travelling direction signal via vector computation. *Nature*, 601(7891):92–97, 2022.

Caitlin S Mallory, Kiah Hardcastle, Malcolm G Campbell, Alexander Attinger, Isabel IC Low, Jennifer L Raymond, and Lisa M Giocomo. Mouse entorhinal cortex encodes a diverse repertoire of self-motion signals. *Nature communications*, 12(1):671, 2021.

Giovanni Luca Marchetti, Daniel Kunin, Adele Myers, Francisco Acosta, and Nina Miolane. Sequential group composition: A window into the mechanics of deep learning. *arXiv preprint arXiv:2602.03655*, 2026.

Alexander Mathis, Andreas VM Herz, and Martin B Stemmler. Resolution of nested neuronal representations can be exponential in the number of neurons. *Physical review letters*, 109(1): 018103, 2012.

Alexander Mathis, Martin B Stemmler, and Andreas VM Herz. Probable nature of higher-dimensional symmetries underlying mammalian grid-cell activity patterns. *Elife*, 4:e05979, 2015.

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

M-L Mittelstaedt and Horst Mittelstaedt. Homing by path integration in a mammal. *Naturwissenschaften*, 67(11):566–567, 1980.

Edvard I Moser, May-Britt Moser, and Bruce L McNaughton. Spatial representation in the hippocampal formation: a history. *Nature neuroscience*, 20(11):1448–1464, 2017.

Noga Mosheiff, Haggai Agmon, Avraham Moriel, and Yoram Burak. An efficient coding theory for a dynamic trajectory predicts non-uniform allocation of entorhinal grid cells to modules. *PLoS computational biology*, 13(6):e1005597, 2017.

Robert U Muller, John L Kubie, and James B Ranck Jr. Spatial firing patterns of hippocampal complex-spike cells in a fixed environment. *The Journal of neuroscience*, 7(7):1935–1950, 1987.

Peter Mussells Pires, Lingwei Zhang, Victoria Parache, LF Abbott, and Gaby Maimon. Converting an allocentric goal into an egocentric steering signal. *Nature*, 626(8000):808–818, 2024.

Aran Nayebi, Alexander Attinger, Malcolm Campbell, Kiah Hardcastle, Isabel Low, Caitlin S Mallory, Gabriel Mel, Ben Sorscher, Alex H Williams, Surya Ganguli, et al. Explaining heterogeneity in medial entorhinal cortex with task-driven neural networks. *Advances in Neural Information Processing Systems*, 34:12167–12179, 2021.

Marcella Noorman, Brad K Hulse, Vivek Jayaraman, Sandro Romani, and Ann M Hermundstad. Maintaining and updating accurate internal representations of continuous variables with a handful of neurons. *Nature neuroscience*, 27(11):2207–2217, 2024.

John O'Keefe and Jonathan Dostrovsky. The hippocampus as a spatial map: preliminary evidence from unit activity in the freely-moving rat. *Brain research*, 1971.

Kevin K Qi and Michael M Yartsev. A two-dimensional grid-cell code for three-dimensional navigation in freely flying bats. *bioRxiv*, pp. 2026–06, 2026.

William T Redman, Francisco Acosta, Santiago Acosta-Mendoza, and Nina Miolane. Not so griddy: Internal representations of rnns path integrating more than one agent. *Advances in Neural Information Processing Systems*, 37:22657–22689, 2024.

Daniel B Rubin, Stephen D Van Hooser, and Kenneth D Miller. The stabilized supralinear network: a unifying circuit motif underlying multi-input integration in sensory cortex. *Neuron*, 85(2):402–417, 2015.

Francesca Sargolini, Marianne Fyhn, Torkel Hafting, Bruce L McNaughton, Menno P Witter, May-Britt Moser, and Edvard I Moser. Conjunctive representation of position, direction, and velocity in entorhinal cortex. *Science*, 312(5774):758–762, 2006.

Rylan Schaeffer, Mikail Khona, and Ila Fiete. No free lunch from deep learning in neuroscience: A case study through models of the entorhinal-hippocampal circuit. *Advances in neural information processing systems*, 35:16052–16067, 2022.

Rylan Schaeffer, Mikail Khona, Tzuhsuan Ma, Cristobal Eyzaguirre, Sanmi Koyejo, and Ila Fiete. Self-supervised learning of representations for space generates multi-modular grid cells. *Advances in Neural Information Processing Systems*, 36:23140–23157, 2023.

Jean-Pierre Serre et al. *Linear representations of finite groups*, volume 42. Springer, 1977.

H Sebastian Seung. How the brain keeps the eyes still. *Proceedings of the National Academy of Sciences*, 93(23):13339–13344, 1996.

Ben Sorscher, Gabriel C Mel, Samuel A Ocko, Lisa M Giocomo, and Surya Ganguli. A unified theory for the computational and mechanistic origins of grid cells. *Neuron*, 111(1):121–137, 2023.

Davide Spalla, Alessandro Treves, and Charlotte N Boccara. Angular and linear speed cells in the parahippocampal circuits. *Nature communications*, 13(1):1907, 2022.

Sameet Sreenivasan and Ila Fiete. Grid cells generate an analog error-correcting code for singularly precise neural computation. *Nature neuroscience*, 14(10):1330–1337, 2011.

Hanne Stensola, Tor Stensola, Trygve Solstad, Kristian Frøland, May-Britt Moser, and Edvard I Moser. The entorhinal grid map is discretized. *Nature*, 492(7427):72–78, 2012.

Jeffrey S Taube, Robert U Muller, and James B Ranck. Head-direction cells recorded from the postsubiculum in freely moving rats. i. description and quantitative analysis. *Journal of Neuroscience*, 10(2):420–435, 1990.

Stephanie Theves, Guillén Fernandez, and Christian F Doeller. The hippocampus encodes distances in multidimensional feature space. *Current Biology*, 29(7):1226–1231, 2019.

Cheng Wang, Xiaojing Chen, and James J Knierim. Egocentric and allocentric representations of space in the rodent brain. *Current opinion in neurobiology*, 60:12–20, 2020.

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

James CR Whittington, Timothy H Muller, Shirley Mark, Guifen Chen, Caswell Barry, Neil Burgess, and Timothy EJ Behrens. The tolman-eichenbaum machine: unifying space and relational memory through generalization in the hippocampal formation. *Cell*, 183(5):1249–1263, 2020.

Dehong Xu, Ruiqi Gao, Wenhao Zhang, Xue-Xin Wei, and Ying Nian Wu. On conformal isometry of grid cells: Learning distance-preserving position embedding. In *The Thirteenth International Conference on Learning Representations*, 2025a. URL https://openreview.net/forum?id=Xo0Q1N7CGk.

Maoshen Xu, Fei Song, Yuxiu Shao, Bailu Si, and Shanshan Qin. The principle of isomorphism: A theory of population activity in grid cells and beyond. *arXiv preprint arXiv:2510.02853*, 2025b.

KiJung Yoon, Michael A Buice, Caswell Barry, Robin Hayman, Neil Burgess, and Ila R Fiete. Specific evidence of low-dimensional continuous attractor dynamics in grid cells. *Nature neuroscience*, 16(8):1077–1084, 2013.

Kechen Zhang. Representation of spatial orientation by the intrinsic dynamics of the head-direction cell ensemble: a theory. *Journal of Neuroscience*, 16(6):2112–2126, 1996.

A GROUPS AND THEIR FOURIER TRANSFORM

Groups formalize the idea of a set of (invertible) transformations or symmetries, that can be composed.

Definition A.1. *A* group *is a set G equipped with a binary operation $G \times G \to G$ denoted by $(g, h) \mapsto gh$, with an* inverse element $g^{-1} \in G$ *for each $g \in G$ and an* identity element $1 \in G$ *such that for all $g, h, k \in G$:*

Associativity	Inversion	Identity
$g(hk) = (gh)k$	$g^{-1}g = gg^{-1} = 1$	$g1 = 1g = g$

A group is *Abelian* if its elements commute ($gh = hg$ for all $g, h \in G$) and otherwise it is *non-Abelian*. Abelian groups model order-insensitive transformations, such as the *cyclic group* $C_p = \mathbb{Z}/p\mathbb{Z}$, which consists of integers modulo p with addition modulo p as the group operation. Non-Abelian groups capture order-sensitive transformations, such as the *dihedral group* D_p, which consists of all rotations and reflections of a regular p-gon.

Elements of any group can be *represented* concretely as invertible matrices, where composition corresponds to matrix multiplication. This allows group operations to be analyzed through linear algebra. We focus on representations with n-dimensional *unitary* matrices, which form the unitary group $\mathrm{U}(n) = \{A \in \mathbb{C}^{n\times n} \mid A^\dagger A = I\}$, where $\dagger$ denotes the conjugate transpose.

Definition A.2. *An n-dimensional* unitary representation *of G is a map $\rho\colon G \to \mathrm{U}(n)$ such that $\rho(gh) = \rho(g)\rho(h)$ for all $g, h \in G$, i.e., a homomorphism between G and $\mathrm{U}(n)$.*

Representations that can not be decomposed as a direct sum of other non-trivial ones are called irreducible (*irreps*, for short) and serve as the fundamental building blocks of every other unitary representation. For Abelian groups, the irreps are one-dimensional, while non-Abelian groups necessarily include higher-dimensional irreps that capture their order-sensitive structure. Every group has a one-dimensional *trivial* irrep, which maps each $g \in G$ to the scalar 1. For a finite group G, there exists a finite number of irreps up to isomorphism. We denote by $\mathcal{I}(G)$ the set of irreps up to isomorphism, and by n_ρ the dimension of $\rho \in \mathcal{I}(G)$.

The traditional discrete Fourier transform (DFT) can be generalized over arbitrary finite groups. The idea is that irreps play the role of harmonics. These harmonics are (unitary) matrix-valued, especially for non-Abelian groups. The Fourier transform then assigns to an arbitrary (complex) signal over G – i.e., an element of $\mathbb{C}^{|G|}$ – coefficients indexed by $\mathcal{I}(G)$.

Definition A.3. *The* Fourier transform *over a finite group G is the map $\mathbb{C}^{|G|} \to \bigoplus_{\rho\in\mathcal{I}(G)} \mathbb{C}^{n_\rho\times n_\rho}$, $x \mapsto \widehat{x}$, defined as:*

$$\widehat{x}[\rho] = \sum_{g\in G} \rho(g)^\dagger x[g] \quad \in \mathbb{C}^{n_\rho\times n_\rho}, \tag{12}$$

where $x[g] \in \mathbb{C}$ indexes the g-th element of x.

The classical DFT can be recovered for the cyclic group C_p. The irreps of C_p are one-dimensional and correspond to the p roots of unity, $\rho_k(g) = \exp(2\pi \mathrm{i} gk/p)$ for $k \in \{0, \dots, p-1\}$. Substituting these irreps into Definition A.3 yields exactly the standard DFT.

The Fourier transform enjoys several properties. It is an invertible change of coordinates that block-diagonalizes the regular actions, turning it into matrix multiplication of Fourier coefficients. *Fourier inversion* reconstructs the original signal, while *Plancherel's Theorem* states that the Fourier transform is an isometry (i.e., a unitary operator) with respect to appropriate inner products. For details, we refer the reader to the introduction of Marchetti et al. (2026), or to the monography Folland (2016).

B Representation Theory of Semidirect Products

Here, we discuss the representation theory of semidirect products. In particular, we argue how the real part of matrix entries of some irreps exhibit grid patterns.

Let A and B be groups and let $\varphi\colon B \to \mathrm{Aut}(A)$ be a homomorphism, where $\mathrm{Aut}(A)$ is the automorphism group of A.

Definition B.1. *The* semidirect product $A \rtimes B$ *is the group with underlying set* $A \times B$ *and multiplication*

$$(a, b)(a', b') = \big(a\,\varphi(b)(a'),\, bb'\big). \tag{13}$$

Intuitively, taking A as translations and B as rotations captures spatial navigation because the current orientation determines how a subsequent local displacement changes position.

Consider a semidirect product $G = A \rtimes B$ between a finite Abelian group A and another finite group B (induced by some homomorphism φ). Note that B also acts on the set of (one-dimensional) irreps $\chi\colon A \to \mathrm{U}(1)$ of A via $(b \cdot \chi)(a) := \chi(b^{-1} \cdot a)$. The irreps of G can be constructed as follows. Let χ be a (one-dimensional) irrep of A, and consider its stabilizer $K = \{b \in B\colon b \cdot \chi = \chi\} \subseteq B$. Given an irrep $\tau\colon K \to \mathrm{U}(m)$ of K (of some dimension m), the tensor product $\chi \otimes \tau\colon (a, k) \mapsto \chi(a)\tau(k)$ defines an irrep of $A \rtimes K \subset G$. The induced representation $\rho = \mathrm{Ind}_{A\rtimes K}^{G}(\chi \otimes \tau)$ on G is irreducible, and has dimension $m|B|/|K|$, i.e., the index of K in B times the dimension of τ. These irreps are non-isomorphic for χ's in distinct B-orbits, and any irrep of G is isomorphic to an irrep of this form – see (Serre et al., 1977, Chapter 8.2).

More explicitly, this representation takes the following form. Given a set of representatives $\{c_1, \ldots, c_t\} \subseteq B$ of cosets of K in B, where $t = |B|/|K|$, the unitary matrix associated to the irrep can be constructed via $m \times m$ blocks $\rho_{i,j}$, $i, j = 1, \ldots, t$, where:

$$\rho_{i,j}(a, b) = \begin{cases} \chi(c_i^{-1} \cdot a)\ \tau(c_i^{-1} b c_j) & \text{if } c_i^{-1} b c_j \in K, \\ 0 & \text{otherwise.} \end{cases} \tag{14}$$

Thus, ρ is a block-monomial matrix.

An interesting case is where the auxiliary irrep τ is chosen to be trivial. Then the above construction specializes to a monomial (i.e., generalized permutation) matrix. Now, if we express the entries of ρ in another (orthonormal) basis basis of $\mathbb{C}^t$ containing two vectors $z_1, z_2 \in \mathbb{C}^t$, they will take form:

$$z_1^\top \rho(a, b)\ z_2 = \sum_{i,j\colon c_i^{-1} b c_j \in K}^{t} z_1[i]\ \chi(c_i^{-1} a)\ z_2[j] = \sum_{i=1}^{t} z_1[i] z_2[j_i]\ \chi(c_i^{-1} a), \tag{15}$$

where j_i is the unique coset index such that $c_i^{-1} b c_{j_i} \in K$. The right-hand side of equation 15 can be interpreted as a linear combination of (one-dimensional) irreps of A, all related by the action of B, with coefficients given by $z_1[1] z_2[j_i]$. For almost all bases, all the coefficients will be non-vanishing. The real part of this expression is a sum of cosine waves (with respect to $a \in A$), whose frequencies are related by B, and whose phases are determined by the coefficients. This superposition of cosine waves gives rise to the grid patterns.

B.1 Specialization to Navigation Groups

In this section, we specialize the previous discussion to discrete rigid motion groups $G = C_n^d \rtimes C_m$, $d = 2, 3$.

Two dimensions. We start with the two-dimensional case $G = C_n^2 \rtimes C_m$. Assuming, without loss of generality, that C_m acts faithfully on C_n^2, this action is defined by a matrix $A \in \mathrm{GL}(2, \mathbb{Z}_n)$ of order m, meaning that $A^m = I$, and $A^i \neq I$ for $1 \leq i < m$. By the classical two-dimensional crystallographic theory, the only possible finite orders of elements in the arithmetic general linear group $\mathrm{GL}(2, \mathbb{Z})$ are $1, 2, 3, 4, 6$. Thus, only these values allow for an action by C_m that is valid for all position resolutions n. For specific values of n, other values of m are possible depending on the arithmetic properties of n.

Recall that the (one-dimensional) irreps of C_n^2 take form $\chi_\xi(a) = \exp(2\pi \mathrm{i}\, \xi \cdot a\, /n)$, where $\xi \in C_n^2$ represents a two-dimensional frequency. The action by C_m on such irreps corresponds to applying (powers of) the matrix $A^{-\top}$ to ξ. Then the real part is equation 15 a (real) linear combination of cosines waves in the form:

$$\cos\left(\frac{2\pi}{n}\, \xi \cdot (A^{-i}a) + \eta_i\right), \tag{16}$$

where η_i is an appropriate phase shift. When ξ is chosen so that the stabilizer K of χ_ξ is trivial (i.e., the orbit $\{(A^{-i})^\top \xi\}_{i=1}^m$ consists of m distinct elements), and for almost all bases, equation 15 is a superposition of m cosine waves, whose frequencies are rotated cyclically. This results in a grid pattern. Specifically, the maxima of equation 16 are located among a triangular ($m = 3, 6$) or square ($m = 4$) lattice.

Even more concretely, for $m = 3$ set

$$A := \begin{pmatrix} 0 & -1 \\ 1 & -1 \end{pmatrix} \in \mathrm{GL}(2, \mathbb{Z}_n). \tag{17}$$

Let $\xi \in C_n^2$ such that $\xi_1 \neq \xi_2$ or $3\xi_1 \neq 0 \pmod n$, which guarantees trivial stabilizer. Then the cosine waves take form equation 16:

$$\cos\frac{2\pi}{n}(a_1\xi_1 + a_2\xi_2), \qquad \cos\frac{2\pi}{n}((-a_1 - a_2)\xi_1 + a_1\xi_2), \qquad \cos\frac{2\pi}{n}(a_2\xi_1 + (-a_1 - a_2)\xi_2). \tag{18}$$

The corresponding image gives rise to the triangular pattern reminiscent of grid cells – see Figure 9.

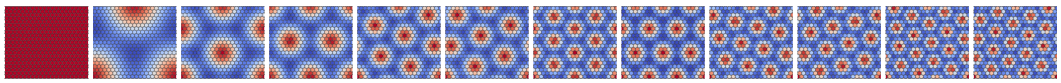


Figure 9: Predicted firing maps of 2D grid cells formed by $C_n^2 \rtimes C_3$ irreps for $n = 21$.

Three dimensions. We now consider the three-dimensional case. We consider the group $C_n^3 \rtimes A_4$, where the alternating group A_4 corresponds to rotations of the tetrahedron. Concretely, the embedding of A_4 into $\mathrm{GL}(3, \mathbb{Z})$ is generated by two rotation matrices h, r of order 2 and 3 respectively, for example:

$$h = \mathrm{diag}(1, -1, -1), \qquad r = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}. \tag{19}$$

Since these matrices are orthogonal, their dual action on (one-dimensional) irreps amounts to applying them to frequencies $\xi \in C_n^3$. Thus, the cosine waves equation 15 specialize to $\cos(2\pi\, \xi \cdot (Aa)/n + \eta_A)$ as $A \in A_4$ varies.

The corresponding 3D image gives rise to a 3D firing fields that resembles those of some "3D grid cells" observed in practice – see Figure 10

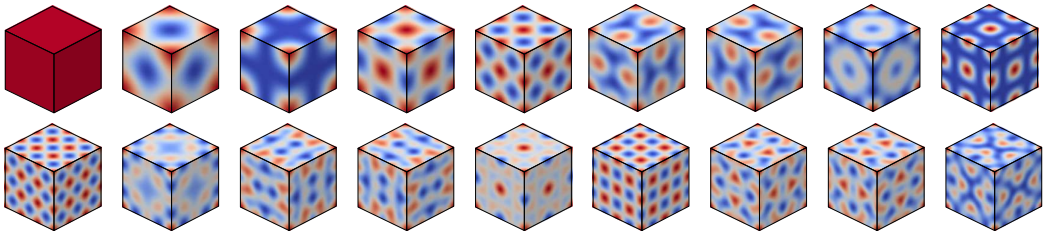


Figure 10: Predicted firing fields of 3D grid cells formed by $C_n^3 \rtimes A_4$ irreps for $n = 60$.

C An Explicit Solution for the RNN

In this section, we prove Theorem 3.1. Specifically, we construct an explicit RNN solution for the sequential group-composition task over arbitrary finite groups. We first state a binary composition solution using a two-layer MLP with squared ReLU nonlinearities[1] that is based on the quadratic computations studied in Kunin et al. (2026). We then lift this solution to an RNN by reusing the same Fourier-feature circuit, based on a construction first presented in Marchetti et al. (2026).

C.1 Binary composition with a two-layer MLP

Let H denote the hidden width and $|G|$ denote the order of the finite group G. The two-layer network has parameters $\Theta = (U, V, W)$ consisting of input weights $U, V \in \mathbb{R}^{H\times|G|}$ and output weights $W \in \mathbb{R}^{|G|\times H}$. Given $g, h \in G$, it predicts

$$f(x_{\text{allo}}(g), x_{\text{ego}}(h); \Theta) = W\,\sigma(Ux_{\text{allo}}(g) + Vx_{\text{ego}}(h)), \tag{20}$$

where $\sigma(z) = (\max\{0, z\})^2$ is applied elementwise. We describe the solution neuronwise. Each of the H hidden neurons are specified by vectors $u, v \in \mathbb{R}^{|G|}$ (rows of U, V) and $w \in \mathbb{R}^{|G|}$ (a column of W). Each neuron is assigned to a single irrep ρ of some dimension n_ρ. Then the vectors u, v, w take the following form for $g \in G$:

$$u[g] = \text{Re Tr}(\rho(g)s_u), \quad v[g] = \text{Re Tr}(\rho(g)s_v), \quad w[g] = \text{Re Tr}(\rho(g)s_w), \tag{21}$$

where $s_u, s_v, s_w \in \mathbb{C}^{n_\rho\times n_\rho}$ are complex matrices. Put simply, the weight vectors of the neuron are real parts of complex linear combinations of matrix entries of ρ.

We will now explicitly describe the matrices s_u, s_v, s_w. Each nonzero irrep ρ is implemented by a circuit containing $4q_\rho n_\rho^3$ neurons, indexed by $\{\pm1\}^2 \times \{0, \dots, q_\rho - 1\} \times \{1, \dots, n_\rho\}^3$. Given $a, b \in \{1, \dots, n_\rho\}$, denote by $E_{a,b}$ the $n_\rho \times n_\rho$ matrix with 1 in the entry (a, b), and 0 elsewhere. Then the neuron indexed by $i := (\epsilon_1, \epsilon_2, \delta, k_0, k_1, k_2)$, with $\epsilon_1, \epsilon_2 \in \{\pm1\}, \delta \in \{0, \dots, q_\rho - 1\}, k_0, k_1, k_2 \in \{1, \dots, n_\rho\}$, satisfies:

$$\begin{aligned} s_u^{(i)} &= \epsilon_1 A_u\, e^{\pi\mathrm{i}\delta/q_\rho}\, E_{k_0,k_2}, \\ s_v^{(i)} &= \epsilon_1\epsilon_2 A_v\, e^{\pi\mathrm{i}\delta/q_\rho}\, \widehat{x}_{\text{ego}}^{\text{ref}}[\rho]^{-\dagger} E_{k_2,k_1}, \\ s_w^{(i)} &= \epsilon_2 A_w\, e^{2\pi\mathrm{i}\delta/q_\rho}\, E_{k_0,k_1}, \end{aligned} \tag{22}$$

where $\mathrm{i} = \sqrt{-1} \in \mathbb{C}$ is the imaginary unit, and $A_u, A_v, A_w \in \mathbb{R}$ are coefficients such that $A_u A_v A_w = n_\rho/(q_\rho|G|)$. The above construction assumes that $\widehat{x}_{\text{ego}}^{\text{ref}}[\rho]$ is an invertible matrix.

Note that the matrices $s_u^{(i)}$ and $s_w^{(i)}$ have a single non-vanishing entry. Thus, by plugging equation 22 into equation 21, we see that the weight vectors of u and w correspond to the real part of a specific matrix entry of an irrep ρ, up to a phase shift and an amplitude. More precisely, if $\theta_{a,b}(g)$ is the phase of $\rho(g)[a, b]$, then $u[g]$ coincides, up to multiplicative constant, with $\cos(\theta_{k_2,k_0}(g) + \pi\delta/q_\rho)$, and similarly for $w[g]$. For $v[g]$, the expression is more complicated, due to the Fourier coefficient $\widehat{x}_{\text{ego}}[\rho]$, resulting in a superposition of multiple matrix entries.

Theorem C.1. *Denote $\mathcal{S} = \{\rho \in \mathcal{I}(G) \colon \widehat{x}_{\text{allo}}^{\text{ref}} \neq 0\}$, and assume that $\widehat{x}_{\text{ego}}^{\text{ref}}[\rho]$ is an invertible matrix for all $\rho \in \mathcal{S}$. Consider the weights Θ with neurons in the form of equation 21 and equation 22 for every irrep $\rho \in \mathcal{S}$. Then the corresponding two-layer MLP with activation $\sigma(z) = (\max\{0, z\})^2$ solves exactly the binary group composition task, i.e.*

$$f(x_{\text{allo}}(g), x_{\text{ego}}(h); \Theta) = x_{\text{allo}}(gh) \tag{23}$$

for all $g, h \in G$.

Proof. Given $r \in G$, express the corresponding output of the MLP equation 20 neuron-wise:

$$f(x_{\text{allo}}(g), x_{\text{ego}}(h); \Theta)[r] = \sum_{i=1}^{H} w_i[r]\,\sigma\left(u_i^\top x_{\text{allo}}(g) + v_i^\top x_{\text{ego}}(h)\right) \tag{24}$$

[1] Squared ReLU combines a firing threshold with expansive gain, consistent with standard firing-rate models Rubin et al. (2015). Quadratic nonlinearities also appear in models of V1 complex cells Adelson & Bergen (1985); Hyvärinen & Hoyer (2000), and quadratic contrast responses have been observed in V1 simple cells Felsen et al. (2005).

where each u_i, v_i, w_i is in the form of equation 21 and equation 22 for the corresponding irrep $\rho \in \mathcal{S}$ and index $(\epsilon_1, \epsilon_2, \delta, k_0, k_1, k_2) \in \{\pm 1\} \times \{0, \dots, q_\rho - 1\} \times \{1, \dots, n_\rho\}^3$. First, note that for a neuron in the form of equation 22, Plancherel's Theorem implies that

$$\begin{aligned} u^\top x_{\text{allo}}(g) &= \text{Re Tr}\left(\rho(g) s_u^\dagger \widehat{x}^{\text{ref}}_{\text{allo}}[\rho]\right) = \epsilon_1 A_u \text{ Re}\left(e^{-\pi \mathrm{i}\delta/q_\rho} (\widehat{x}^{\text{ref}}_{\text{allo}}[\rho]\rho(g))[k_0, k_2]\right), \\ v^\top x_{\text{ego}}(h) &= \text{Re Tr}\left(\rho(h) s_v^\dagger \widehat{x}^{\text{ref}}_{\text{ego}}[\rho]\right) = \epsilon_1 \epsilon_2 A_v \text{ Re}\left(e^{-\pi \mathrm{i}\delta/q_\rho} \rho(h)[k_2, k_1]\right). \end{aligned} \tag{25}$$

By using the identity $z^2 = \sigma(z) + \sigma(-z)$ and the polarization identity $4\alpha\beta = (\alpha+\beta)^2 - (\alpha-\beta)^2$, the choice of the signs ϵ_1, ϵ_2 in equation 22 implies that, after summing over $(\epsilon_1, \epsilon_2) \in \{\pm 1\}^2$, equation 24 reduces to a sum of expressions in the form:

$$\frac{4n_\rho}{q_\rho |G|} \text{ Re}\left(e^{2\pi \mathrm{i}\delta/q_\rho} \rho(r)[k_1, k_0]\right) \text{ Re}\left(e^{-\pi \mathrm{i}\delta/q_\rho} (\widehat{x}^{\text{ref}}_{\text{allo}}[\rho]\rho(g))[k_0, k_2]\right) \text{ Re}\left(e^{-\pi \mathrm{i}\delta/q_\rho} \rho(h)[k_2, k_1]\right). \tag{26}$$

Recall that for complex numbers $\alpha, \beta, \gamma \in \mathbb{C}$, we have $4\text{Re}(\alpha)\text{Re}(\beta)\text{Re}(\gamma) = \text{Re}(\alpha\beta\gamma + \alpha\beta\overline{\gamma} + \alpha\overline{\beta}\gamma + \overline{\alpha}\beta\gamma)$. After applying this identity, and summing over the index δ, the imaginary exponentials in equation 26 will give rise to four cyclic sums of roots of unity. Since these cyclic sums vanish for non-trivial roots of unities, the only surviving term is:

$$\frac{n_\rho}{|G|} \text{ Re}\left(\rho(r)[k_1, k_0] \; (\widehat{x}^{\text{ref}}_{\text{allo}}[\rho]\rho(g))[k_0, k_2] \; \rho(h)[k_2, k_1]\right). \tag{27}$$

By summing the above expression over the indices (k_0, k_1, k_2), we obtain:

$$\frac{n_\rho}{|G|} \text{ Re Tr}\left(\rho(r)\widehat{x}^{\text{ref}}_{\text{allo}}[\rho]\rho(g)\rho(h)\right) = \frac{n_\rho}{|G|} \text{ Re Tr}\left(\widehat{x}^{\text{ref}}_{\text{allo}}[\rho] \; \rho(ghr)\right) \tag{28}$$

Finally, the properties of the Fourier transform imply that the sum over $\rho \in \mathcal{S}$ of the right-hand side of the above expression equals $x^{\text{ref}}_{\text{allo}}[ghr] = x_{\text{allo}}(gh)[r]$, as desired. □

Intuition on the 7-tuple indexing. Each element in the tuple $(\rho, \epsilon_1, \epsilon_2, \delta, k_0, k_1, k_2)$ indexing hidden neurons in Theorem 3.1 plays a concrete computational role in implementing group composition. Here we explain the main idea behind each element:

- ρ *separates composition into independent blocks.* The Fourier transform diagonalizes group composition: Each irrep provides a separate matrix-valued Fourier component, within which group composition becomes matrix multiplication. The circuit computes these components independently and combines them through Fourier inversion to recover the output, as shown in equation 28 and the concluding step of the proof.
- k_0, k_1, k_2 *implement matrix multiplication.* The indices k_0, k_1 select an entry of the matrix product, while k_2 selects an intermediate coordinate. Summing over k_2 forms the matrix product; summing over k_0, k_1 combines it with the output weights to form the trace in equation 28.
- ϵ_1, ϵ_2 *implement real multiplication.* The four sign choices allow squared-ReLU neurons to multiply two real inputs. Summing over ϵ_1 pairs opposite preactivations to recover a square, while summing over ϵ_2 subtracts the squared difference from the squared sum via the polarization identity, leading to equation 25.
- δ *implements complex multiplication.* The phase index selects different real projections of the complex inputs. Averaging their products over these phases cancels unwanted conjugated terms and retains the desired complex product in the real-valued output. The condition $q_\rho \geq 3$ ensures this cancellation, yielding equation 27.

C.2 Sequential composition with an RNN

We now lift the MLP solution for the binary task constructed in the previous section to an RNN solution for the sequential task. Consider weights $\Theta = (U, V, W)$ as in Theorem C.1. Set the weights of the RNN as:

$$\mathbf{W}_{\text{in}} = U, \qquad \mathbf{W}_{\text{drive}} = V, \qquad \mathbf{W}_{\text{out}} = W, \qquad \mathbf{W}_{\text{mix}} = UW. \tag{29}$$

Theorem C.2. *Assume that $\widehat{x}^{\mathrm{ref}}_{\mathrm{ego}}[\rho]$ is an invertible matrix for all $\rho \in \mathcal{S}$. Consider the weights Θ with neurons in the form of equation 29. Then the corresponding RNN with activation $\sigma(z) = (\max\{0, z\})^2$ solves exactly the sequential group composition task, i.e., in the notation of equation 4,*

$$y^{(t)} = x_{\mathrm{allo}}(g_0 \cdots g_t) \tag{30}$$

for all $g_0, \ldots, g_t \in G$.

Proof. Recall that group composition is associative, implying that

$$g_0 \cdots g_t = ((((g_0 \cdot g_1) \cdot g_2) \cdots g_{t-1}) \cdot g_t). \tag{31}$$

Since the RNN weights equation 29 are constructed from the ones of the MLP solving the binary task, the associativity property equation 31 implies inductively that, after t iterations, the RNN outputs $y^{(t)} = x_{\mathrm{allo}}(g_0 \cdots g_t)$, as desired. □

This concludes the proof of Theorem 3.1. As anticipated in Section 3, the mixing matrix exhibits an interesting block structure.

Corollary C.3. *Consider the construction in Theorem 3.1. For hidden neurons $i = (\rho, \epsilon_1, \epsilon_2, \delta, k_0, k_1, k_2)$ and $j = (\rho', \epsilon_1', \epsilon_2', \delta', k_0', k_1', k_2')$, the recurrent weight $\mathbf{W}_{\mathrm{mix}}[i, j]$ is given by*

$$\mathbf{W}_{\mathrm{mix}}[i,j] \propto \begin{cases} \epsilon_1 \epsilon_2' \cos\left(\dfrac{\pi}{q_\rho}(\delta - 2\delta')\right) \mathbf{1}\{k_0 = k_0'\}\mathbf{1}\{k_2 = k_1'\}, & \rho' \in \{\rho, \overline{\rho}\}, \\ 0, & \text{otherwise.} \end{cases} \tag{32}$$

Consequently, grouping neurons by conjugate-irrep class makes $\mathbf{W}_{\mathrm{mix}}$ block diagonal.

Proof. By construction, $W_{\mathrm{mix}}[i, j] = \langle u_i, w_j \rangle$, where u_i is row i of U and w_j is column j of W. If u_i and v_j are neurons associated to irreps ρ and ρ' respectively, via computations similar to Marchetti et al. (2026) we obtain:

$$W_{\mathrm{mix}}[i,j] = \begin{cases} \frac{|G|}{2n_\rho} \operatorname{Re} \operatorname{Tr}\left(\left(s_u^{(i)}\right)^\dagger s_w^{(j)}\right) & \text{if } \rho' \in \{\rho, \overline{\rho}\}, \\ 0 & \text{otherwise.} \end{cases} \tag{33}$$

This means that the mixing matrix is block-diagonal, with a block for each irrep (up to conjugate). In the notation of the previous section, if $i = (\epsilon, \delta, k_0, k_1, k_2)$ and $j = (\epsilon', \delta', k_0', k_1', k_2')$, then using equation 22 we can further expand the above expression as:

$$\operatorname{Re} \operatorname{Tr}\left(\left(s_u^{(i)}\right)^\dagger s_w^{(j)}\right) \propto \begin{cases} \cos\left(\frac{\pi}{q_\rho}(\delta - 2\delta')\right) & \text{if } k_0 = k_0' \text{ and } k_2 = k_1', \\ 0 & \text{otherwise.} \end{cases} \tag{34}$$

Thus, each block of the mixing matrix is sparse, with non-vanishing entries corresponding to a cosine kernel in the index δ. □

C.3 On the Optimality of the Solution

We now ask to what extent our construction from Theorem 3.1 is optimal in terms of hidden width H. To this end, we define the dimension of the orbit span of the allocentric code as,

$$d_{\mathrm{allo}} := \dim\left(\operatorname*{span}_{g \in G} x_{\mathrm{allo}}(g)\right) = \sum_{\rho \in \mathcal{S}} n_\rho \operatorname{rank} \widehat{x}^{\mathrm{ref}}_{\mathrm{allo}}[\rho]. \tag{35}$$

The second equality follows from the decomposition of the regular representation into irreducible representations. The following simple observation provides a lower bound for any two-layer MLP solving binary group composition.

Proposition C.4. *Any two-layer MLP of the form in equation 20 that exactly solves binary group composition has hidden width*

$$H \geq d_{\mathrm{allo}}. \tag{36}$$

Proof. The output of the network lies in the column space of $W \in \mathbb{R}^{|G|\times H}$, which has dimension at most H. On the other hand, exactness implies that, as $g, h \in G$ vary, the outputs contain every vector $x_{\text{allo}}(k)$ for $k \in G$. Hence

$$\operatorname*{span}_{k\in G} x_{\text{allo}}(k) \subseteq \operatorname{Im}(W), \tag{37}$$

and therefore $d_{\text{allo}} \le \operatorname{rank}(W) \le H$. □

If every nonzero Fourier coefficient $\widehat{x}^{\text{ref}}_{\text{allo}}[\rho]$ is invertible, then $d_{\text{allo}} = \sum_{\rho\in\mathcal{S}} n_\rho^2$. Thus any exact MLP must satisfy $H \ge \sum_{\rho\in\mathcal{S}} n_\rho^2$. By comparison, choosing $q_\rho = 3$ in the construction from Theorem 3.1 requires a width of $12\sum_{\rho\in\mathcal{S}} n_\rho^3 \le 12 n_{\max}\sum_{\rho\in\mathcal{S}} n_\rho^2$, where $n_{\max} = \max_{\rho\in\mathcal{S}} n_\rho$. Hence the construction is $O(n_{\max})$-optimal in width. In particular, whenever the dimensions of the relevant irreps remain uniformly bounded, the construction is asymptotically optimal up to a constant factor.

This is precisely the case for the semidirect-product groups considered in Section B. For $G = A \rtimes B$ with A Abelian, every irrep is induced from a character of A and an irrep τ of a stabilizer K, and has dimension

$$n_\rho = \frac{|B|}{|K|}\dim\tau \le |B|. \tag{38}$$

Thus $n_{\max}$ depends only on the rotational factor B, and not on the size of the Abelian translation group A. For the discrete isometry groups considered in Section B.1, B is fixed while the spatial resolution n grows. Consequently, the width of our construction has asymptotically optimal scaling in the spatial resolution, both for the two-dimensional groups $C_n^2 \rtimes C_m$ and for the three-dimensional group $C_n^3 \rtimes A_4$.

C.4 Relation to Continuous Attractor Networks

We show that our construction shares the persistence and attraction of continuous attractor networks (CANs), while differing in the topology of its fixed-point family.

CANs are widely used to model path integration (e.g., Seung (1996); Zhang (1996); Burak & Fiete (2009); Noorman et al. (2024); Cotteret et al. (2025)). They support a continuous family of stable activity patterns representing variables such as head direction or spatial position, often with ring or toroidal topology. Recurrent dynamics restore activity toward this family following perturbations, while self-motion inputs move activity along it to update the represented state. Typically, CANs with ring topology implement this mechanism with a circulant connectivity matrix. More specifically, neurons are arranged along a ring, where connection weight is only a function of distance. Nearby neurons form excitatory connections, while distant neurons inhibit each other. Prior works (e.g., Zhang (1996); Noorman et al. (2024)) make mathematically precise how to choose weights that faithfully implement attractor dynamics. In contrast, our construction takes the form of a discrete-time RNN organized into irrep modules, without explicitly imposing an attractor manifold. To clarify its relation to CANs, we examine a continuous-time relaxation of our circuit. We show that, under identity drive, it preserves the decoded allocentric state while attracting hidden activity toward a continuous family of fixed points. This family, however, is parameterized by all representable allocentric activity patterns and thus has the topology of a vector space, rather than a ring or torus reflecting the navigational variable.

Consider the continuous-time dynamics

$$\tau\dot{h} = -h + \sigma(\mathbf{W}_{\text{mix}}h + b)\,, \qquad h(0) = h_0, \tag{39}$$

where $\mathbf{W}_{\text{mix}} \in \mathbb{R}^{H\times H}$ and $b \in \mathbb{R}^H$ is a fixed drive. A forward-Euler discretization with step-size τ recovers the RNN update equations in Eq. (4) with the drive clamped to b, making Eq. (39) a continuous-time relaxation of the RNN. We use this system as a probe of the recurrent circuit to ask which hidden states are self-consistent and in what sense are these states stable to perturbations under fixed drive.

Let $\mathbf{W}_{\text{mix}}$ be defined as in Theorem 3.1 and fix the drive to the identity update, $b = \mathbf{W}_{\text{drive}}\, x^{\text{ref}}_{\text{ego}}$. Under this drive, a hidden state representing an allocentric code should remain unchanged, because

composing with the identity element leaves the allocentric state fixed. Indeed, if we define the map $\Phi : \mathbb{R}^{|G|} \to \mathbb{R}^H$ by

$$\Phi(x) = \sigma\big(\mathbf{W}_{\text{in}} x + \mathbf{W}_{\text{drive}}\, x_{\text{ego}}^{\text{ref}}\big)\,, \tag{40}$$

then by construction $\mathbf{W}_{\text{out}}\Phi(x) = x$ for all $x \in \mathbb{R}^{|G|}$. This implies that the image $\text{Im}(\Phi) = \big\{\Phi(x) : x \in \mathbb{R}^{|G|}\big\}$ forms a family of fixed points of the identity-drive recurrence. Moreover, this family is attracting: from any initial hidden state, the dynamics converge exponentially to the unique point in $\text{Im}(\Phi)$ with the same decoded allocentric state. In other words, under identity drive, the recurrent circuit behaves like an attracting memory scaffold: the decoded state $\mathbf{W}_{\text{out}} h_0$ is conserved, while the hidden activity contracts to a canonical representation of that decoded state $\Phi(\mathbf{W}_{\text{out}} h_0)$.

Proposition C.5 (Exponential attraction under identity drive)**.** *Consider the RNN constructed in Theorem 3.1 and the continuous-time dynamics Eq.* (39) *with identity drive* $b = \mathbf{W}_{\text{drive}}\, x_{\text{ego}}^{\text{ref}}$*. Then* $\mathbf{W}_{\text{out}} h(t) = \mathbf{W}_{\text{out}} h_0$*, and*

$$h(t) = h^* + e^{-t/\tau}\,[h_0 - h^*]\,, \qquad h^* = \Phi(\mathbf{W}_{\text{out}} h_0) \in \text{Im}(\Phi). \tag{41}$$

Proof. Since $\mathbf{W}_{\text{mix}} = \mathbf{W}_{\text{in}}\mathbf{W}_{\text{out}}$, the continuous-time dynamics in Eq. (39) can be written as

$$\tau \dot{h} = -h + \Phi(\mathbf{W}_{\text{out}} h)\,. \tag{42}$$

Because the construction solves the identity-update task for every allocentric input, $\mathbf{W}_{\text{out}}\Phi(x) = x$ for all $x \in \mathbb{R}^{|G|}$. Applying $\mathbf{W}_{\text{out}}$ to both sides of the dynamics gives

$$\tau \frac{d}{dt}\,(\mathbf{W}_{\text{out}} h) = -\mathbf{W}_{\text{out}} h + \mathbf{W}_{\text{out}}\Phi(\mathbf{W}_{\text{out}} h) = 0. \tag{43}$$

Thus $\mathbf{W}_{\text{out}} h(t) = \mathbf{W}_{\text{out}} h(0)$ for all $t \geq 0$. Substituting this conserved quantity back into the hidden dynamics yields $\tau \dot{h} = -h + \Phi(\mathbf{W}_{\text{out}} h(0))$, a linear differential equation with constant forcing. Solving gives

$$h(t) = \Phi(\mathbf{W}_{\text{out}} h(0)) + e^{-t/\tau}\,[h(0) - \Phi(\mathbf{W}_{\text{out}} h(0))]\,, \tag{44}$$

as claimed. □

D TUNING CURVES

In neuroscience, the *tuning curve* $r_i(s)$ of neuron i is its average firing rate as a function of a stimulus s. Tuning curves are used to identify neurons *selective* for particular variables or behaviors. For example, in spatial neuroscience, place cells and grid cells are characterized by the structure of their tuning curves over the animal's position in space Moser et al. (2017). We use the same idea to characterize hidden neurons in our RNN construction. Because different trajectories can produce different hidden activities at the same cumulative state, these tuning curves depend on the statistics generating the trajectories. We make this dependence explicit in the following definition.

Definition D.1 (RNN tuning curve). *Let $\mathbb{P}$ be a distribution over trajectories $\gamma = (g_0, g_1, \dots, g_k) \in G^{k+1}$, where g_0 is the initial state and $g_1, \dots, g_k$ are the updates. Let $c_t(\gamma) = g_0 g_1 \cdots g_t$ denote the cumulative group state and $h_i^{(t)}(\gamma)$ the activation of hidden neuron i at time t. The* tuning curve *of neuron i under $\mathbb{P}$ is*

$$r_i(g; \mathbb{P}) = \mathbb{E}_{\gamma \sim \mathbb{P},\, t}\left[h_i^{(t)}(\gamma) \,\middle|\, c_t(\gamma) = g \right], \tag{45}$$

where t is sampled independently and uniformly from $\{1, \dots, k\}$. The curve is defined for states with positive sampling probability.

Thus, $r_i(g; \mathbb{P})$ is the average activation of neuron i over paths and time points at which the network represents cumulative state g, weighted by their sampling probabilities. Unless otherwise specified, we sample the initial state uniformly from G and each update independently and uniformly from an allowed movement set $\mathcal{M} \subseteq G$, so that $\mathbb{P}$ is uniform on $G \times \mathcal{M}^k$. We abbreviate $r_i(g; \mathbb{P})$ as $r_i(g)$ when the trajectory distribution is clear. When all states have positive sampling probability, the tuning curve can equivalently be viewed as a vector $r_i(G) \in \mathbb{R}^{|G|}$ indexed by group elements.

D.1 COMPUTING TUNING CURVES

We can compute tuning curves for our construction either by averaging activity along sampled trajectories or by averaging over one-step transitions.

Empirical estimation. For trajectories sampled from $\mathbb{P}$, we estimate the tuning curve by the occupancy-normalized average

$$\widehat{r}_i(g; \mathbb{P}) = \frac{\sum_t h_i^{(t)} \mathbf{1}\{c_t = g\}}{\sum_t \mathbf{1}\{c_t = g\}}, \tag{46}$$

where the sums run over all sampled updates across trajectories. States with no observations are left undefined.

Exact one-step computation. In our construction, exact decoding and $\mathbf{W}_{\text{mix}} = \mathbf{W}_{\text{in}} \mathbf{W}_{\text{out}}$ imply that hidden activity depends only on the predecessor state and the current movement. We can therefore compute tuning without simulating trajectories. Under the sampling distribution defined above—a uniform initial state and independent uniform updates from $\mathcal{M}$—each movement $j \in \mathcal{M}$ is equally likely conditional on reaching a target state g. Its unique predecessor is gj^{-1}, giving

$$r_i(g; \mathbb{P}) = \frac{1}{|\mathcal{M}|} \sum_{j \in \mathcal{M}} \left[\sigma\left(\mathbf{W}_{\text{in}} x_{\text{allo}}(gj^{-1}) + \mathbf{W}_{\text{drive}} x_{\text{ego}}(j) \right) \right]_i . \tag{47}$$

This expression can be evaluated for selected neurons without constructing the full hidden activity vector. For other trajectory distributions, the transitions must instead be weighted by their conditional probabilities given the target state.

D.2 PROVABLE IRREP SELECTIVITY

Each hidden neuron in our construction is associated with an irrep of the navigation group. This organization suggests that its tuning should reflect the same irrep. We make this connection exact under two conditions: trajectories are sampled uniformly from G^{k+1}, including the initial state, and tuning curves are summed over neurons differing only in the sign ϵ_1. The first condition gives uniform averaging over all transitions into a state, the second removes rectification through the identity $\mathrm{ReLU}(z)^2 + \mathrm{ReLU}(-z)^2 = z^2$.

Corollary D.2 (Irrep-aligned paired tuning curves)**.** *Consider the explicit RNN construction in Theorem 3.1, with trajectory distribution $\mathbb{P}$ uniform on G^{k+1}. Let i_+ and i_- be two neurons whose labels agree except for $\epsilon_1 = +1$ and $\epsilon_1 = -1$, respectively, and let ρ_i be their shared irrep. Define their paired tuning curve by $s_i(g) := r_{i_+}(g;\mathbb{P}) + r_{i_-}(g;\mathbb{P})$. Then, there exist $C_i \in \mathbb{R}$ and $M_i \in \mathbb{C}^{n_{\rho_i}\times n_{\rho_i}}$ such that*

$$s_i(g) = C_i + \operatorname{Re}\operatorname{Tr}(\rho_i(g)M_i)\,. \tag{48}$$

Consequently, the centered paired tuning curve $\bar{s}_i(g) = s_i(g) - \frac{1}{|G|}\sum_{g'\in G} s_i(g')$ has Fourier support contained only in $\{\rho_i, \overline{\rho_i}\}$.

Proof. Fix the pair $i_\pm = (\rho, \pm 1, \epsilon_2, \delta, k_0, k_1, k_2)$. By Eq. (25), the two contributions to the preactivation of i_+ are

$$U_i(a) := {u_{i_+}}^\top x_{\text{allo}}(a) = A_u \operatorname{Re}\Big(e^{-\pi \mathrm{i}\delta/q_\rho}\big(\widehat{x}^{\text{ref}}_{\text{allo}}[\rho]\rho(a)\big)[k_0,k_2]\Big)\,, \tag{49}$$

$$V_i(b) := {v_{i_+}}^\top x_{\text{ego}}(b) = \epsilon_2 A_v \operatorname{Re}\Big(e^{-\pi \mathrm{i}\delta/q_\rho}\rho(b)[k_2,k_1]\Big)\,. \tag{50}$$

The preactivation of i_- is the negative of that of i_+. Applying the exact one-step computation with $\mathcal{M} = G$ and the identity $\mathrm{ReLU}(z)^2 + \mathrm{ReLU}(-z)^2 = z^2$ gives

$$s_i(g) = \frac{1}{|G|}\sum_{a\in G}\big[U_i(a) + V_i(a^{-1}g)\big]^2 = C_i + \frac{2}{|G|}\sum_{a\in G} U_i(a)V_i(a^{-1}g), \tag{51}$$

where $C_i = \frac{1}{|G|}\sum_{a\in G}\big[U_i(a)^2 + V_i(a)^2\big]$ is independent of g. Let $E_{k_1k_2}$ denote the matrix with a single unit entry at (k_1,k_2). Using $\rho(a^{-1}g) = \rho(a^{-1})\rho(g)$ and cyclicity of the trace, the cross term becomes

$$\frac{2}{|G|}\sum_{a\in G} U_i(a)V_i(a^{-1}g) = \operatorname{Re}\operatorname{Tr}(\rho(g)M_i)\,, \tag{52}$$

with

$$M_i := \frac{2\epsilon_2 A_v e^{-\pi \mathrm{i}\delta/q_\rho}}{|G|}\sum_{a\in G} U_i(a)\, E_{k_1k_2}\rho(a^{-1}). \tag{53}$$

Hence $s_i(g) = C_i + \operatorname{Re}\operatorname{Tr}(\rho(g)M_i)$. Centering removes the constant component, leaving Fourier support contained in $\{\rho, \overline{\rho}\}$. □

This corollary connects our RNN's irrep organization directly to the Fourier structure of its tuning curves. It does not guarantee exact irrep support for individual squared-ReLU neurons or for nonuniform transition sampling. Nevertheless, it motivates irrep selectivity as a quantitative test of whether the algebraic structure remains concentrated in neural tuning.

E EXPERIMENTAL DETAILS

We detail the experimental setups and analyses for planar navigation (Section E.1) and irrep selectivity (Section E.2). An anonymized repository containing the code used to run these experiments is available at `https://anonymous.4open.science/r/grids-and-groups/`.

E.1 PLANAR NAVIGATION

In the two-dimensional path-integration experiments shown in Fig. 6, we use

$$G = C_n^2 \rtimes C_6, \qquad n = 21,$$

with initial pose $g_0 = (2, 2, 0)$. The resulting group contains $6n^2 = 2{,}646$ position–heading states. We use the exact, fixed-weight RNN construction with $q_\rho = 3$, a squared-ReLU nonlinearity, and all irreps retained.

Figure 11 summarizes the geometric constraints on the heading order and spatial tiling. Because headings act by lattice-preserving rotations, the crystallographic restriction limits the rotation order in two dimensions to $m \in \{1, 2, 3, 4, 6\}$. The three regular tessellations of the Euclidean plane are the triangular, square, and hexagonal tessellations. Square and hexagonal cells are centrally symmetric: every edge has a parallel opposite edge, so persistent straight-ahead motion is well-defined. Triangular cells are not centrally symmetric. Their upward- and downward-facing states have edge-normal directions offset by $60°$, so an edge crossing cannot preserve both the heading and a one-step forward transition. Representing triangular faces would therefore require an additional sublattice-dependent motion rule, which is absent from our construction.

Square tilings are compatible with $m = 2, 4$, while hexagonal tilings are compatible with $m = 2, 3, 6$. Although $m = 2$ is permitted by the crystallographic restriction, the rotation orbit of a forward translation is only $\{v, -v\}$ and spans one spatial axis. Forward motion and rotation therefore fail to generate the full two-dimensional translation group. We consequently consider $m = 3$ on the hexagonal tiling, $m = 4$ on the square tiling, and $m = 6$ on the hexagonal tiling. For $m = 3$, the three headings select alternating directions among the six neighboring cells but still generate the full translation lattice.

E.1.1 REFERENCE POPULATION CODES

We construct $x^{\mathrm{ref}}_{\mathrm{allo}}$ as a periodic Gaussian bump on the triangular spatial lattice, centered at $(2, 2)$ with width $\sigma = 1$. Its amplitudes across the six headings are $(1.0,\ 0.8,\ 0.4,\ 0.2,\ 0.4,\ 0.8)$. Thus, the allocentric reference code is localized in position and nonuniform in heading.

We construct $x^{\mathrm{ref}}_{\mathrm{ego}}$ as a random signal using seed 10. Each candidate is sampled as the sum of a Gaussian vector and one half of a squared Gaussian vector. We resample until its Fourier coefficient $\widehat{x}^{\mathrm{ref}}_{\mathrm{ego}}[\rho]$ is invertible for every retained irrep, using a minimum singular-value threshold of 10^{-5}. Invertibility is required by the closed-form construction of the drive weights. No spatial or directional interpretation is assigned to this reference vector.

E.1.2 SIMULATING TRAJECTORIES

We use two trajectory distributions for different purposes. The illustrative trajectory is forward-biased and uses nonperiodic boundaries. At each step, the model samples a turn and then a body-relative translation. The translation probabilities are 0.05 for staying still, 0.625 for moving forward, 0.1175 for each forward-oblique direction, 0.0125 for each backward-oblique direction, and 0.065 for moving backward. Left and right turns each have probability 0.12, with turn persistence 0.275. Candidate movements that leave the arena are excluded, and movements toward nearby walls are downweighted. The displayed trajectory uses 143 updates and random seed 3.

For the empirical tuning calculation, we instead use periodic boundaries and sample uniformly from the 21 local movements. Each trajectory contains 110 updates, of which the first 10 are discarded. Independent trajectories are accumulated until the requested numbers of retained samples,

$$T \in \{10{,}000,\ 100{,}000,\ 500{,}000\},$$

have been obtained. These empirical estimates are compared with the exact local one-step averages defined above.

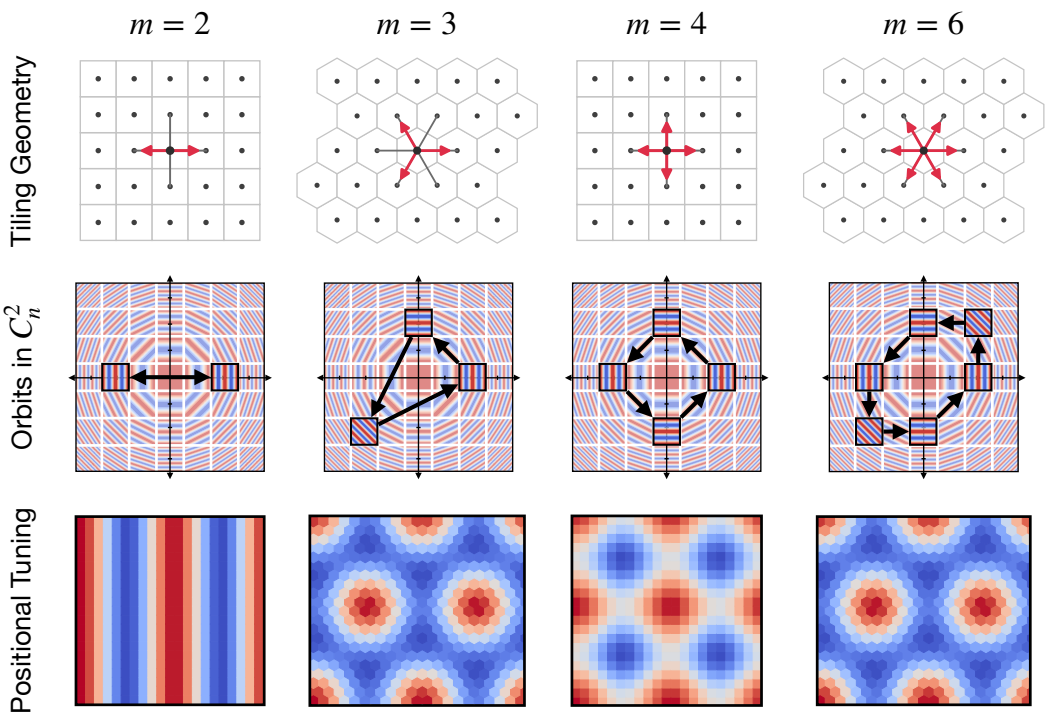


Figure 11: **Tiling geometry, heading order, and spatial tuning.** Columns compare $G = C_n^2 \rtimes C_m$ for $m = 2, 3, 4, 6$. **Top row:** Local neighborhoods and forward directions. Dots mark position states, gray segments connect neighboring states, and red arrows show one-step forward translations. The cases $m = 2, 4$ share a square tiling, with two opposite or four forward directions, respectively. The cases $m = 3, 6$ share a hexagonal tiling, with three alternating or six forward directions, respectively. **Middle row:** Orbits of Fourier modes of C_n^2 under C_m rotations, which underlie the irreps of $C_n^2 \rtimes C_m$. Boxes highlight the modes in each selected orbit, and arrows indicate the action of the rotation generator. **Bottom row:** Corresponding spatial patterns, showing bands for $m = 2$, square periodicity for $m = 4$, and hexagonal periodicity for $m = 3, 6$.

E.1.3 Tuning Curves

Local versus global averaging. We use two movement sets. The local set contains the Cartesian product of seven body-frame translations—staying still and moving to one of the six neighboring lattice sites—with three heading changes, $-60°$, $0°$, and $+60°$. It therefore contains 21 movements. The global set is $\mathcal{M} = G$ and contains all 2,646 possible movements.

Local one-step tuning is the appropriate limit for trajectories whose updates are sampled uniformly from the 21 local movements. Global tuning instead averages uniformly over every possible factorization of a target pose. It is a different transition distribution, not simply a longer local trajectory. In the current Figure 6 notebook, the empirical and single-cell panels use local tuning, while the population-geometry calculation uses global tuning.

How m affects spatial tuning. Figure 12 compares local one-step tuning while holding the spatial irrep and neuron labels fixed. The heading order determines the rotation orbit of the selected spatial frequency and therefore the symmetry of the tuning pattern: $m = 4$ produces square periodicity, whereas $m = 3$ and $m = 6$ produce hexagonal periodicity. Changing δ changes the phase and placement of the fields without changing this symmetry. Unless otherwise stated, our experiments use $m = 6$.

Inverted spatial maps. Some hidden units have spatial maps that appear inverted relative to another unit in the same module. These are not negative firing rates: the squared-ReLU activity is nonnegative. Rather, the construction includes sign-paired hidden features, indexed by ϵ_1 and ϵ_2,

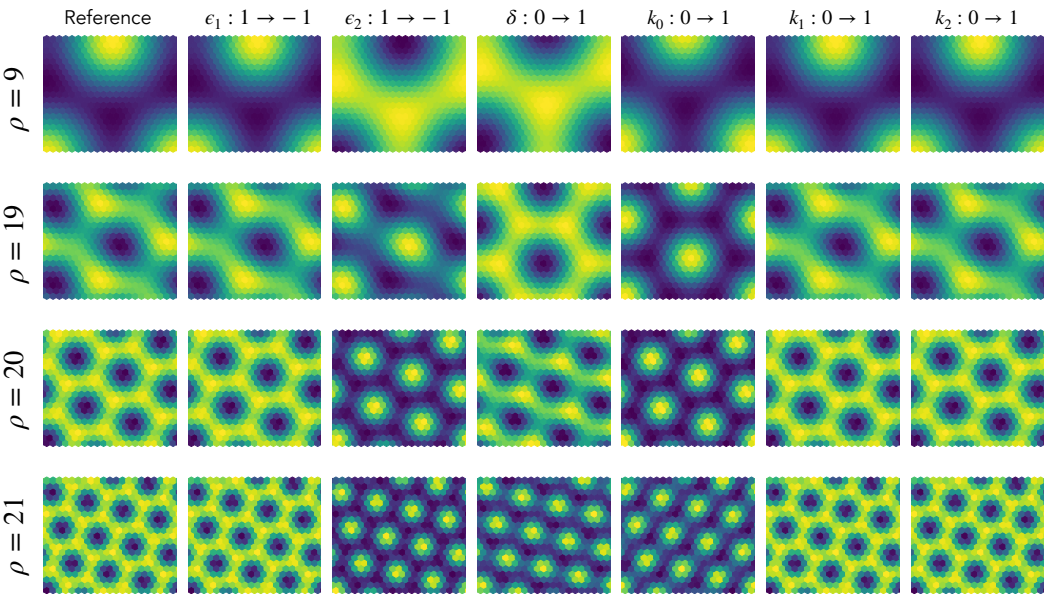


Figure 12: **Effect of neuron labels on spatial tuning.** Exact local one-step tuning curves, averaged over heading, for $G = C_n^2 \rtimes C_6$ with $n = 21$. Rows show four irrep modules, labeled $\rho = 9, 19, 20, 21$. The first column shows a reference neuron with $\epsilon_1 = \epsilon_2 = +1$ and $\delta = k_0 = k_1 = k_2 = 0$; subsequent columns change one label at a time as indicated. In these examples, flipping ϵ_2 exchanges peaks and troughs, while changing δ or k_0 alters the spatial pattern. The displayed changes to ϵ_1, k_1, and k_2 leave the normalized maps unchanged. Each tuning curve is normalized independently for visualization.

that are required to reconstruct the bilinear group-composition operation. Reversing the sign of a preactivation changes which half of an oscillatory feature passes through the ReLU. After conditional averaging and separate display normalization, the resulting map can exchange peaks and troughs with a reference map. We therefore treat these inverted maps as complementary computational features of the construction, not as a separate predicted biological cell class.

Effects of the neuron labels. Each hidden neuron is indexed by $(\rho, \epsilon_1, \epsilon_2, \delta, k_0, k_1, k_2)$. The irrep ρ determines the underlying spatial-frequency orbit and therefore the module's spatial scale and lattice orientation. The signs ϵ_1, ϵ_2 select the sign-paired features described above. The phase index $\delta \in \{0, 1, 2\}$ and the matrix index k_0 change the phase of the spatial map and generally translate it within the periodic arena.

These labels do not provide a universal Cartesian coordinate system for translation. In particular, the relation between (δ, k_0) and the observed x- and y-shifts depends on the irrep. Our sweeps did not identify a single one-to-one rule that applies across modules.

The remaining indices k_1 and k_2 affect the internal matrix coefficients used by the construction but do not change the normalized positional marginal at fixed $(\rho, \epsilon_1, \epsilon_2, \delta, k_0)$. Across the four six-dimensional irreps tested, the largest normalized shape discrepancy produced by varying k_1 and k_2 was below 3×10^{-14}. We therefore fix $k_1 = k_2 = 0$ when displaying the distinct spatial maps.

Defining conjunctive tuning. For each neuron, we reshape the pose tuning curve as $T_i(\theta, x, y)$. Its positional and heading marginals are

$$T_i^{\mathrm{pos}}(x, y) = \frac{1}{6} \sum_{\theta} T_i(\theta, x, y), \qquad T_i^{\mathrm{head}}(\theta) = \frac{1}{n^2} \sum_{x,y} T_i(\theta, x, y).$$

For either marginal v, we define its relative modulation as

$$M(v) = \frac{\max v - \min v}{\max |v|}.$$

Using threshold 10^{-2}, a neuron is classified as spatial-only if only $M(T_i^{\mathrm{pos}})$ exceeds threshold, orientation-only if only $M(T_i^{\mathrm{head}})$ exceeds threshold, conjunctive if both exceed threshold.

The current analysis evaluates the 7,938 neurons satisfying $\epsilon_1 = \epsilon_2 = +1, \qquad k_1 = 0$, and reports the four class counts separately for every irrep. This fixed-label slice contains $3d_\rho^2$ neurons from an irrep of dimension d_ρ, so raw counts across irreps reflect both tuning type and irrep dimension. Proportions within each irrep should be used when comparing selectivity across modules. Conjunctive tuning is concentrated in irreps whose matrix coefficients retain both positional and heading dependence; it is not distributed uniformly across irreps.

E.1.4 Population Geometry

For the population-geometry analysis, we use all neurons in irrep module $\rho = 9$ and compute their global positional tuning at each of the $n^2 = 441$ positions. Each neuron's activity is centered across positions, after which the full population is divided by one shared scale. Euclidean distances are then computed between all pairs of population vectors.

The current configuration embeds this distance matrix in three dimensions using UMAP with 30 neighbors, minimum distance 0.15, and random seed 101. Persistent homology is computed through dimension two using all 441 positions. The periodic lattice connectivity used to draw the surface and wireframe affects only the visualization; it is not used to compute the persistent homology.

E.2 Irrep selectivity in freely moving MEC recordings

Data and preprocessing. We analyzed freely moving mouse MEC recordings from Mallory et al. (2021). Spikes were assigned to the nearest original tracking sample, excluding invalid positions, out-of-recording events, and tracking gaps exceeding three times the median frame interval. Occupancy was estimated from valid tracking intervals. For irrep selectivity, we discretized position into 21×21 bins and movement direction, estimated from position rather than head orientation, into six equal angular bins. We included samples with speed $\geq 2\,\mathrm{cm\,s^{-1}}$ and applied the same movement criterion to spikes and occupancy. Firing rates were computed as spike counts divided by occupancy, with bins containing less than $0.1\,\mathrm{s}$ of occupancy treated as unobserved. Observed rates were centered by their unweighted mean, after which unobserved bins were set to zero. No smoothing was applied to these tuning estimates.

Irrep selectivity. We represented the resulting tuning function f on $G = C_{21}^2 \rtimes C_6$, using $n = 21$ and $m = 6$. The angular generator acts on spatial indices through $A = \left(\begin{smallmatrix} 0 & -1 \\ 1 & 1 \end{smallmatrix}\right)$ modulo 21. Thus, rectangular spatial-bin indices are interpreted as triangular-lattice group coordinates; positions are not transformed into a physical triangular grid. For each irrep ρ of dimension d_ρ, we computed

$$\widehat{f}(\rho) = \sum_{g \in G} f(g)\rho(g)^\dagger, \qquad P_\rho = \frac{d_\rho}{|G|} \left\| \widehat{f}(\rho) \right\|_F^2 .$$

We retained the $K = 67$ six-dimensional irreps with zero-based indices 15–81 in the implementation's ordering. Power was normalized over this retained set $\mathcal{I}$, giving

$$p_\rho = \frac{P_\rho}{\sum_{\sigma \in \mathcal{I}} P_\sigma}, \qquad \mathrm{IS} = \log K + \sum_{\rho \in \mathcal{I}} p_\rho \log p_\rho .$$

Natural logarithms express selectivity in nats, with $0 \log 0 = 0$. Higher values indicate greater concentration of retained irrep power.

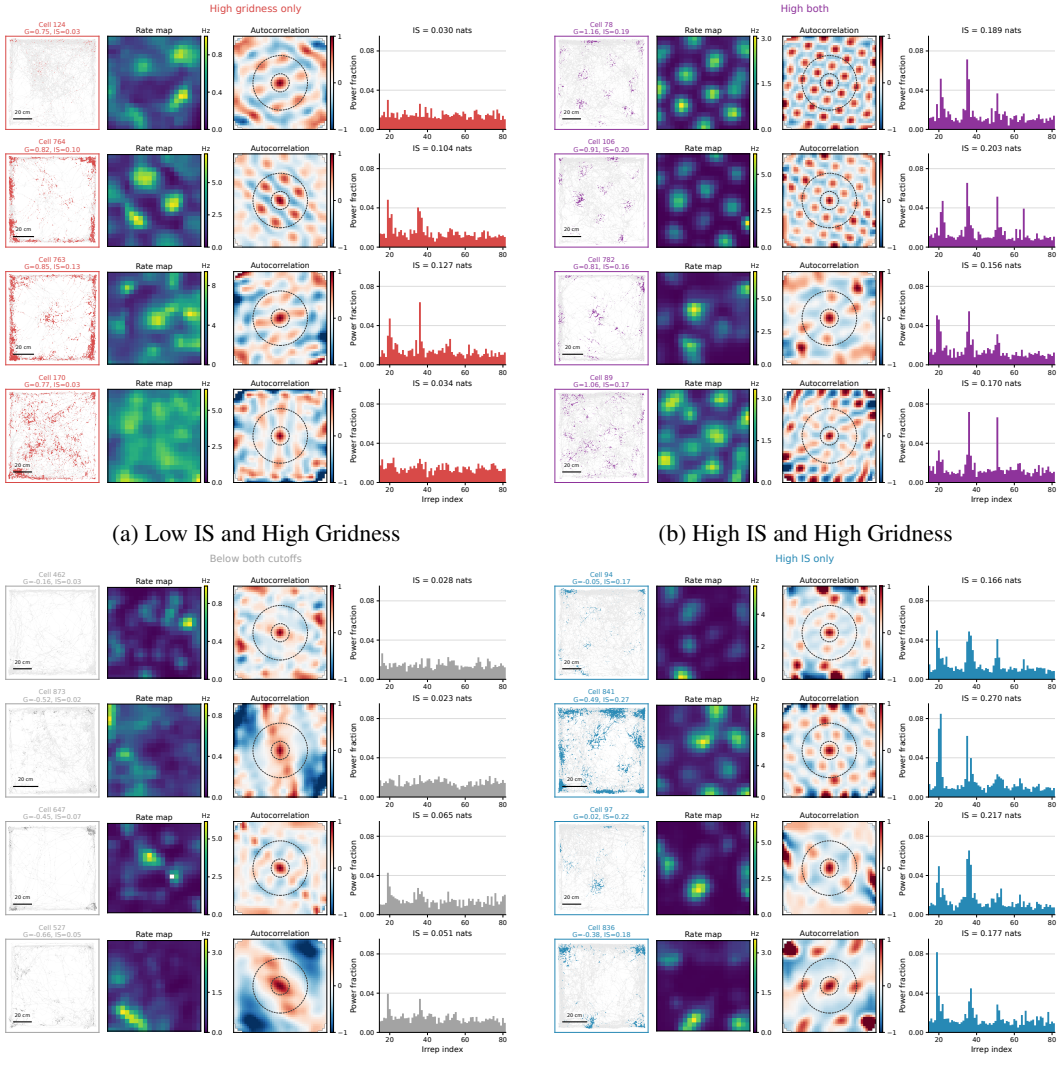
Gridness. Gridness used all valid positions, without the movement-speed cutoff. We constructed a 21×21 spatial rate map by separately smoothing spike counts and occupancy with a Gaussian kernel ($\sigma = 1$ bin, reflecting boundaries), then dividing and masking smoothed occupancy below $0.1\,\mathrm{s}$. We computed the linear autocorrelation of the mean-centered map, divided by the number of overlapping valid bins, and normalized at zero lag. Within a fixed annulus spanning 0.2–0.6 of the autocorrelogram radius, Pearson correlations c_θ with bilinearly rotated copies gave

$$\mathrm{Gridness} = \min(c_{60^\circ}, c_{120^\circ}) - \max(c_{30^\circ}, c_{90^\circ}, c_{150^\circ}).$$

Lags required at least five overlapping valid bins, and each rotation correlation required at least 20 valid pixels.

Comparison and visualization. We included neurons with finite values for both scores and used their empirical 95th percentiles to define four descriptive groups. These cutoffs are not shuffle-based significance thresholds. Highlighted points show the exact scores of the corresponding example neurons. Additional appendix examples were sampled reproducibly within groups (seed 0), displaying up to four cells per group. Irrep power distributions share a common vertical scale across the main examples and, separately, across the appendix panels.

Potential use with limited data. Because irrep selectivity is computed per neuron in a prescribed group basis, it can be applied when too few neurons are recorded simultaneously for reliable toroidal population reconstruction (Gardner et al., 2022). It may also complement gridness when limited sampling obscures spatial periodicity. However, the present estimator still requires adequate position–direction coverage: low occupancy or few spikes can bias the estimated power distribution. Improved robustness in these regimes remains to be established through controlled subsampling and occupancy-matched null models.



(a) Low IS and High Gridness

(b) High IS and High Gridness

(c) Low IS and Low Gridness

(d) High IS and Low Gridness

Figure 13: **Example cells across gridness and irrep selectivity.** Neurons are grouped into four categories according to their gridness and irrep selectivity (IS), corresponding to the four quadrants of the gridness–IS scatterplot. Rows show high (top) and low (bottom) gridness. Columns show low (left) and high (right) irrep selectivity. This comparison distinguishes hexagonal spatial tuning from concentration of tuning power within a small number of irreps.