On Conformal Isometry of Grid Cells: Learning Distance-Preserving Position Embedding

Dehong Xu[1], **Ruiqi Gao**[1], **Wen-Hao Zhang**[2], **Xue-Xin Wei**[3], **Ying Nian Wu**[1]
[1]UCLA   [2]UT Southwestern Medical Center   [3]UT Austin
xudehong1996@ucla.edu, ywu@stat.ucla.edu

Abstract

This paper investigates the conformal isometry hypothesis as a potential explanation for the hexagonal periodic patterns in grid cell response maps. We posit that grid cell activities form a high-dimensional vector in neural space, encoding the agent's position in 2D physical space. As the agent moves, this vector rotates within a 2D manifold in the neural space, driven by a recurrent neural network. The conformal hypothesis proposes that this neural manifold is a conformal isometric embedding of 2D physical space, where local physical distance is preserved by the embedding up to a scaling factor (or unit of metric). Such distance-preserving position embedding is indispensable for path planning in navigation, especially planning local straight path segments. We conduct numerical experiments to show that this hypothesis leads to the hexagonal grid firing patterns by learning maximally distance-preserving position embedding, agnostic to the choice of the recurrent neural network. Furthermore, we present a theoretical explanation of why hexagon periodic patterns emerge by minimizing our loss function by showing that hexagon flat torus is maximally distance preserving.

1 Introduction

The mammalian hippocampus formation encodes a "cognitive map" (Tolman, 1948; O'keefe & Nadel, 1979) of the animal's surrounding environment. In the 1970s, it was discovered that place cells in the rodent hippocampus fire at specific locations within the environment (O'Keefe & Dostrovsky, 1971). Later, another type of neurons, called grid cells, was identified in the medial entorhinal cortex (Hafting et al., 2005; Fyhn et al., 2008; Yartsev et al., 2011; Killian et al., 2012; Jacobs et al., 2013; Doeller et al., 2010). Unlike place cells, each grid cell fires at multiple locations that form a hexagonal grid pattern across the environment (Fyhn et al., 2004; Hafting et al., 2005; Fuhs & Touretzky, 2006; Burak & Fiete, 2009; Sreenivasan & Fiete, 2011; Blair et al., 2007; Couey et al., 2013; de Almeida et al., 2009; Pastoll et al., 2013; Agmon & Burak, 2020). Grid cells are thought to interact with place cells and play a crucial role in path integration (Hafting et al., 2005; Fiete et al., 2008; McNaughton et al., 2006; Gil et al., 2018; Ridler et al., 2019; Horner et al., 2016; Ginosar et al., 2023; Boccara et al., 2019), which calculates the agent's self-position by accumulating its self-motion. Thus, this has led to the view that grid cells form an internal Global Positioning System (GPS) in the brain (Moser & Moser, 2016). Although grid cells have primarily been studied in spatial contexts, recent research suggests that grid-like responses may also exist in more abstract, non-spatial cognitive spaces (Constantinescu et al., 2016; Bellmund et al., 2018).

Numerous computational models have been proposed to explain the characteristic firing patterns of grid cells. Early approaches focused on continuous attractor neural networks (CANN) (Amit, 1992; Burak & Fiete, 2009; Couey et al., 2013; Pastoll et al., 2013; Agmon & Burak, 2020). More recently, two papers (Cueva & Wei, 2018; Banino et al., 2018) learned recurrent neural networks (RNNs) on path integration tasks and demonstrated that grid patterns emerge in the learned networks. These results were further extended in subsequent studies (Gao et al., 2019; Sorscher et al., 2019; Cueva et al., 2020; Gao et al., 2021; Whittington et al., 2021; Dorrell et al., 2022; Xu et al., 2022; Sorscher et al., 2023). Besides RNN-based models, basis expansion models based on principal component

Project page: https://github.com/DehongXu/grid-cell-conformal-isometry

analysis (PCA) with non-negativity constraints (Dordek et al., 2016; Sorscher et al., 2023; Stachenfeld et al., 2017) have been proposed to capture interactions between grid and place cells.

While prior models have advanced our understanding of grid cells, the underlying mathematical principles behind the emergence of hexagonal grid patterns remain elusive (Cueva & Wei, 2018; Sorscher et al., 2023; Gao et al., 2021; Nayebi et al., 2021; Schaeffer et al., 2022). Recently, the conformal isometry hypothesis has gained attention and has been studied in various papers on grid cells. This hypothesis was formalized by Xu et al. (2022) with earlier explorations by Gao et al. (2021; 2019), and has been adapted and investigated in recent works (Schaeffer et al., 2023; Schoyen et al., 2024). The conformal isometry hypothesis posits that grid cell activities form a high-dimensional vector in neural space, encoding the agent's position in 2D physical space. As the agent moves, this vector rotates within a 2D neural manifold, guided by a recurrent neural network (RNN). The hypothesis proposes that this manifold is a conformal isometric embedding of the 2D physical space, where local physical distance is preserved by the position embedding up to a scaling factor (or unit of metric). Such distance-preserving position embedding is indispensable for path planning in navigation, including planning local straight-path segments.

This paper investigates the conformal isometry hypothesis as a potential mathematical explanation for the formation of hexagonal periodic patterns in grid cell response maps. Unlike previous studies (Xu et al., 2022; Gao et al., 2021) that examined conformal isometry within models combining both place cells and grid cells, we focus on a minimal setting: a single module of grid cells equipped with an explicit metric. By reducing the system to its essentials, we bring the conformal isometry hypothesis to the forefront and study the grid cell system in isolation with fewer assumptions. This approach allows us to gain a sharper and deeper understanding of the hypothesis.

In this paper, we design numerical experiments in the minimalistic setting demonstrating that the conformal isometry hypothesis underlies the hexagonal periodic patterns in grid cell response maps. Specifically, we show that the hexagonal periodic patterns emerge by learning the maximally distance-preserving position embedding. To further validate this hypothesis, we conduct in-depth mathematical analysis to show that hexagon periodic patterns emerge by minimizing our loss function, due to the fact that the hexagon flat torus exhibits minimal deviation from local conformal isometry, i.e., the hexagon flat torus forms the maximally distance-preserving position embedding.

Contributions. To summarize, our paper investigates the conformal isometry hypothesis as a possible mathematical principle that underlies the grid cell system. Our contributions are as follows. (1) We conduct a systematic numerical investigation of the conformal isometry hypothesis in a minimalistic setting with a single module of grid cells. (2) We provide a general framework that is agnostic to specific forms of transformation models, grid scales, and the number of neurons. (3) We present a theoretical analysis demonstrating that the hexagonal grid patterns emerge by minimizing our conformal and transformation loss function.

2 BACKGROUND

2.1 REPRESENTING SELF-POSITION
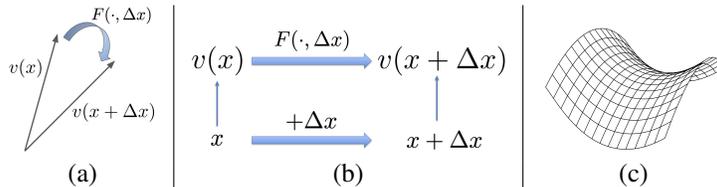
Figure 1: (a) The self-position $\boldsymbol{x} = (x_1, x_2)$ in 2D Euclidean space is represented by a vector $\boldsymbol{v}(\boldsymbol{x})$ in the d-dimensional neural space. When the agent moves by $\Delta\boldsymbol{x}$, the vector is transformed to $\boldsymbol{v}(\boldsymbol{x} + \Delta\boldsymbol{x}) = F(\boldsymbol{v}(\boldsymbol{x}), \Delta\boldsymbol{x})$. (b) $F(\cdot, \Delta\boldsymbol{x})$ is a representation of the self-motion $\Delta\boldsymbol{x}$. (c) $\mathbb{M} = (\boldsymbol{v}(\boldsymbol{x}), \boldsymbol{x} \in D)$ is a 2D manifold in the neural space, and is an embedding of the 2D Euclidean domain $\mathbb{D}$.

Suppose the agent is at the self-position $\boldsymbol{x} = (x_1, x_2) \in \mathbb{R}^2$ within a 2D Euclidean domain $\mathbb{D}$. The activities of the population of d grid cells form a d-dimensional vector $\boldsymbol{v}(\boldsymbol{x}) = (v_i(\boldsymbol{x}), i = 1, ..., d)$, where $v_i(\boldsymbol{x})$ is the activity of the i-th grid cell at position $\boldsymbol{x}$. We call the d-dimensional vector space of $\boldsymbol{v}$ the neural space, and we embed $\boldsymbol{x}$ of the 2D physical space as a vector $\boldsymbol{v}(\boldsymbol{x})$ of the d-dimensional

neural space. We may also call the vector $\boldsymbol{v}(\boldsymbol{x})$ the position embedding, following the deep learning terminology (Vaswani et al., 2017). For each grid cell i, $v_i(\boldsymbol{x})$, as a function of $\boldsymbol{x}$, represents the response map of grid cell i.

2.2 REPRESENTING SELF-MOTION

At self-position $\boldsymbol{x} = (x_1, x_2)$, assume the agent makes a movement $\Delta\boldsymbol{x} = (\Delta x_1, \Delta x_2)$ and moves to $\boldsymbol{x} + \Delta\boldsymbol{x}$. Correspondingly, the vector $\boldsymbol{v}(\boldsymbol{x})$ is transformed to $\boldsymbol{v}(\boldsymbol{x} + \Delta\boldsymbol{x})$. The general form of the transformation can be formulated as:

$$\boldsymbol{v}(\boldsymbol{x} + \Delta\boldsymbol{x}) = F(\boldsymbol{v}(\boldsymbol{x}), \Delta\boldsymbol{x}), \tag{1}$$

where F can be parametrized by an RNN. See Figure 1(a). The input velocity $\Delta\boldsymbol{x}$ can also be represented as $(\Delta r, \theta)$ in polar coordinates, where $\Delta r = \|\Delta\boldsymbol{x}\|$ is the displacement along the heading direction $\theta \in [0, 2\pi]$, so that $\Delta\boldsymbol{x} = (\Delta x_1 = \Delta r \cos\theta, \Delta x_2 = \Delta r \sin\theta)$. $F(\cdot, \Delta\boldsymbol{x})$ is a representation of the self-motion $\Delta\boldsymbol{x}$. See Figure 1(a-b). We can also write $F(\boldsymbol{v}(\boldsymbol{x}), \Delta\boldsymbol{x}) = F(\boldsymbol{v}(\boldsymbol{x}), \Delta r, \theta)$ with slight overloading of notation F. The transformation model is necessary for path integration and path planning.

2.3 CONFORMAL ISOMETRY

For each $\boldsymbol{x} \in \mathbb{D}$, such as a 1m $\times$ 1m square environment, we embed $\boldsymbol{x}$ as a vector $\boldsymbol{v}(\boldsymbol{x})$ in the d-dimensional neural space. Collectively, $\mathbb{M} = (\boldsymbol{v}(\boldsymbol{x}), \forall \boldsymbol{x} \in \mathbb{D})$ is a 2D manifold in the neural space, and $\mathbb{M}$ is an embedding of $\mathbb{D}$. See Figure 1(c) (the shape of $\mathbb{M}$ in the figure is merely illustrative). As the agent moves in $\mathbb{D}$, $\boldsymbol{v}(\boldsymbol{x})$ moves in $\mathbb{M}$.

The conformal isometry hypothesis (Xu et al., 2022; Gao et al., 2021) proposes that $\mathbb{M}$ is a conformal embedding of $\mathbb{D}$. Specifically, at any $\boldsymbol{x} \in \mathbb{D}$, for a local $\Delta\boldsymbol{x}$, we have

$$\|\boldsymbol{v}(\boldsymbol{x} + \Delta\boldsymbol{x}) - \boldsymbol{v}(\boldsymbol{x})\| = s\|\Delta\boldsymbol{x}\| + o(\|\Delta\boldsymbol{x}\|), \tag{2}$$

where s is a constant scaling factor independent of $\boldsymbol{x}$ and $\Delta\boldsymbol{x}$. As the agent moves in $\mathbb{D}$ by $\|\Delta\boldsymbol{x}\|$, the vector $\boldsymbol{v}(\boldsymbol{x})$ moves in $\mathbb{M}$ by $s\|\Delta\boldsymbol{x}\|$.

Unit of metric. The scaling factor s defines the metric unit for distance preserving. For example, if $\|\Delta\boldsymbol{x}\|$ is measured in meters, and we want to work in centimeters, then $s = 100$.

Intrinsic and extrinsic curvatures. For a constant s, the intrinsic geometry of the manifold $\mathbb{M}$ remains Euclidean or flat, i.e., $\mathbb{M}$ is a folding or bending of the flat $\mathbb{D}$ without distorting stretching or squeezing. The intrinsic curvature of $\mathbb{M}$ is zero. But the extrinsic curvature of $\mathbb{M}$ is non-zero, i.e., the $o(\|\Delta\boldsymbol{x}\|)$ term in (equation 2) grows as $\|\Delta\boldsymbol{x}\|$ increases. We want to minimize this term for non-infinitesimal $\|\Delta\boldsymbol{x}\|$ so that distance is maximally preserved within a non-infinitesimal local range.

3 A MINIMALISTIC SETTING

3.1 ASSUMPTIONS

In this section, we seek to study the grid cell system with a minimal number of assumptions. Specifically, we make the following four assumptions:

Assumption 1. Conformal isometry: $\|\boldsymbol{v}(\boldsymbol{x}+\Delta\boldsymbol{x})-\boldsymbol{v}(\boldsymbol{x})\| = s\|\Delta\boldsymbol{x}\|+o(\|\Delta\boldsymbol{x}\|)$. In the minimalistic setting, we specify s explicitly, in order to understand how s affects the learned hexagon patterns, and conversely what the hexagon patterns reveal about the underlying s. This will enable us to gain a deeper geometric understanding of the grid cell patterns. We shall discuss learning s in Appendix I.

Assumption 2. Transformation: $\boldsymbol{v}(\boldsymbol{x}+\Delta\boldsymbol{x}) = F(\boldsymbol{v}(\boldsymbol{x}), \Delta\boldsymbol{x})$, where F is a recurrent neural network. We want to be agnostic about F, and our numerical experiments show that hexagon grid patterns emerge regardless of the form of F. In our experiments, we consider the following simple forms.

(1) Linear model: $\boldsymbol{v}(\boldsymbol{x}+\Delta\boldsymbol{x}) = \boldsymbol{v}(\boldsymbol{x}) + \boldsymbol{B}(\theta)\boldsymbol{v}(\boldsymbol{x})\Delta r$, where $\Delta r = \|\Delta\boldsymbol{x}\|$ is the displacement, and θ is the heading direction of $\Delta\boldsymbol{x}$. $\boldsymbol{B}(\theta)$ is a $d \times d$ square matrix.

(2) Nonlinear model 1: $\boldsymbol{v}(\boldsymbol{x} + \Delta\boldsymbol{x}) = R(\boldsymbol{A}\boldsymbol{v}(\boldsymbol{x}) + \boldsymbol{B}(\theta)\boldsymbol{v}(\boldsymbol{x})\Delta r + \boldsymbol{b})$, where R is elementwise nonlinearity such as ReLU, $\boldsymbol{A}$ and $\boldsymbol{B}(\theta)$ are $d \times d$ square matrices, and $\boldsymbol{b}$ is $d \times 1$ bias vector.

(3) Nonlinear model 2: $\boldsymbol{v}(\boldsymbol{x} + \Delta\boldsymbol{x}) = R(\boldsymbol{A}\boldsymbol{v}(\boldsymbol{x}) + \boldsymbol{B}(\theta)\Delta r + \boldsymbol{b})$, where $\boldsymbol{B}(\theta)$ is a $d \times 1$ vector, $\boldsymbol{b}$ is bias vector, and R is nonlinearity.

Assumption 3. Normalization: $\|\boldsymbol{v}(\boldsymbol{x})\| = 1$ for each $\boldsymbol{x} \in \mathbb{D}$. $\|\boldsymbol{v}(\boldsymbol{x})\|^2 = \sum_i v_i(\boldsymbol{x})^2$ can be interpreted as the total energy of the population of neurons in $\boldsymbol{v}$ at position $\boldsymbol{x}$. This normalization assumption makes the conformal isometry assumption well-defined. Otherwise, we can multiply the vector $\boldsymbol{v}(\boldsymbol{x})$ by an arbitrary constant c, so that the scaling factor s is changed to cs. Such undesirable arbitrariness is eliminated by the normalization assumption.

Assumption 4. Non-negativity: $v_i(\boldsymbol{x}) \geq 0$ for each i and $\boldsymbol{x}$. This is the assumption studied by Dordek et al. (2016); Sorscher et al. (2023); Stachenfeld et al. (2017). This is obviously true for biological neurons. However, our ablation studies show that it is not necessary for the emergence of hexagon grid patterns. On the other hand, this assumption does enable more stable learning of clean patterns, because it greatly constrains the solution space.

The above assumptions form a minimalistic setting for studying grid cells, where place cells are not involved. This enables us to study the grid cells in isolation with an explicit metric s.

Scaling property. Under the above assumptions, for a fixed constant c, let $\tilde{\boldsymbol{v}}(\boldsymbol{x}) = \boldsymbol{v}(c\boldsymbol{x})$, then for Assumption 1, $\|\tilde{\boldsymbol{v}}(\boldsymbol{x} + \Delta\boldsymbol{x}) - \tilde{\boldsymbol{v}}(\boldsymbol{x})\| = \|\boldsymbol{v}(c(\boldsymbol{x} + \Delta\boldsymbol{x})) - \boldsymbol{v}(c\boldsymbol{x})\| = cs\|\Delta\boldsymbol{x}\| + o(\|\Delta\boldsymbol{x}\|)$. For Assumption 2, $\tilde{\boldsymbol{v}}(\boldsymbol{x} + \Delta\boldsymbol{x}) = \boldsymbol{v}(c\boldsymbol{x} + c\Delta\boldsymbol{x}) = F(\boldsymbol{v}(c\boldsymbol{x}), c\Delta\boldsymbol{x}) = F(\tilde{\boldsymbol{v}}(\boldsymbol{x}), c\Delta\boldsymbol{x}) = \tilde{F}(\tilde{\boldsymbol{v}}(\boldsymbol{x}), \Delta\boldsymbol{x})$. Assumptions 3 and 4 continue to hold. Therefore, a system with scaling factor s can be translated to a system with scaling factor cs.

3.2 Learning method

Let $\mathbb{D}$ be a 1m $\times$ 1m Euclidean continuous square environment. We overlay a 40×40 regular lattice on $\mathbb{D}$. We learn $\boldsymbol{v}(\boldsymbol{x})$ on the 40×40 lattice points, but we treat $\boldsymbol{x}$ as continuous, so that for $\boldsymbol{x}$ off the lattice points, we let $\boldsymbol{v}(\boldsymbol{x})$ be the bi-linear interpolation of the $\boldsymbol{v}(\boldsymbol{x})$ on the 4 nearest grid points.

The loss function consists of the following two terms:

$$L_1 = \mathbb{E}_{\boldsymbol{x},\Delta\boldsymbol{x}}[(\|\boldsymbol{v}(\boldsymbol{x} + \Delta\boldsymbol{x}) - \boldsymbol{v}(\boldsymbol{x})\| - s\|\Delta\boldsymbol{x}\|)^2], \tag{3}$$

$$L_2 = \mathbb{E}_{\boldsymbol{x},\Delta\boldsymbol{x}}[\|\boldsymbol{v}(\boldsymbol{x} + \Delta\boldsymbol{x}) - F(\boldsymbol{v}(\boldsymbol{x}), \Delta\boldsymbol{x})\|^2], \tag{4}$$

where L_1 ensures Assumption 1 on conformal isometry, while L_2 satisfies Assumption 2 on transformation. Since L_2 is a one-step transformation loss, back-propagation through time is not required.

In L_2, we assume $\Delta\boldsymbol{x}$ to be infinitesimal for local motion (e.g., $\|\Delta\boldsymbol{x}\| \leq 0.075$). In L_1, however, we assume $s\Delta\boldsymbol{x}$ to be within a non-infinitesimal range, $s\|\Delta\boldsymbol{x}\| \leq D$ (e.g., $D = 1.25$). In defining $\mathbb{E}_{\Delta\boldsymbol{x}}$, we assume $\Delta\boldsymbol{x}$ follows uniform distribution within the ranges specified above. In L_1, we can also assume a more general distribution $p(\Delta\boldsymbol{x})$, so that $L_1 = \mathbb{E}_{\boldsymbol{x}}[\int (\|\boldsymbol{v}(\boldsymbol{x} + \Delta\boldsymbol{x}) - \boldsymbol{v}(\boldsymbol{x})\| - s\|\Delta\boldsymbol{x}\|)^2 p(\Delta\boldsymbol{x}) d\Delta\boldsymbol{x}]$.

We minimize $L = L_1 + \lambda L_2$ over the set of $\boldsymbol{v}(\boldsymbol{x})$ on the 40×40 lattice points as well as the parameters in F, such as $\boldsymbol{B}(\theta)$ for the discrete set of directions θ in the linear model. $\lambda > 0$ is a hyper-parameter that balances L_1 and L_2. We use stochastic gradient descent to minimize L, where in each iteration, the expectations in L are replaced by the averages over Monte Carlo samples of $(\boldsymbol{x}, \Delta\boldsymbol{x})$.

After each iteration, we set any negative elements in $\boldsymbol{v}(\boldsymbol{x})$ to zero to enforce Assumption 4 on non-negativity. Then we normalize $\|\boldsymbol{v}(\boldsymbol{x})\| = 1$ for each $\boldsymbol{x}$, ensuring Assumption 3 on normalization.

3.3 Maximally distance-preserving position embedding

While L_2 can be minimized to close to 0, L_1 has a non-zero minimum due to the non-infinitesimal range D. For small $s\|\Delta\boldsymbol{x}\|$, the deviation is small due to $o(\|\Delta\boldsymbol{x}\|)$ term. In fact, our experiment in Section 3.5 shows that conformal isometry is nearly exact for $s\|\Delta\boldsymbol{x}\| \leq .8$. Beyond that point, the deviation begins to increase due to the extrinsic curvature of the manifold. Minimizing L_1 amounts to finding the maximally distance-preserving position embedding at a given scaling factor s within

the range D. The reason we constrain the range of $s\|\Delta x\|$ is that under the scaling $\tilde{v}(x) = v(cx)$, the deviation from conformal isometry in L_1 satisfies: $\|\tilde{v}(x + \Delta x) - \tilde{v}(x)\| - cs\|\Delta x\| = \|v(cx + c\Delta x) - v(cx)\| - s\|c\Delta x\|$.

3.4 NUMERICAL EXPERIMENTS

We perform numerical experiments in the minimalistic setting, testing the generality of our method across various RNN parameterizations. Additionally, we varied the scaling factor s and the number of neurons. For Δx in L_1, we constrained it locally, ensuring $s\|\Delta x\| \leq 1.25$. For L_2, $\|\Delta x\|$ was restricted to be smaller than 0.075.

The dimensions of $v(x)$, representing the total number of grid cells, were nominally set to 24 for both the linear model and nonlinear model 1, and 1000 for nonlinear model 2. Notably, similar results can be obtained with different numbers of cells, e.g., 500, for both the linear model and nonlinear model 1. All the parameters are updated by Adam optimizer (Kingma & Ba, 2014).

(a) Learned grid cells   (b) Toroidal analysis
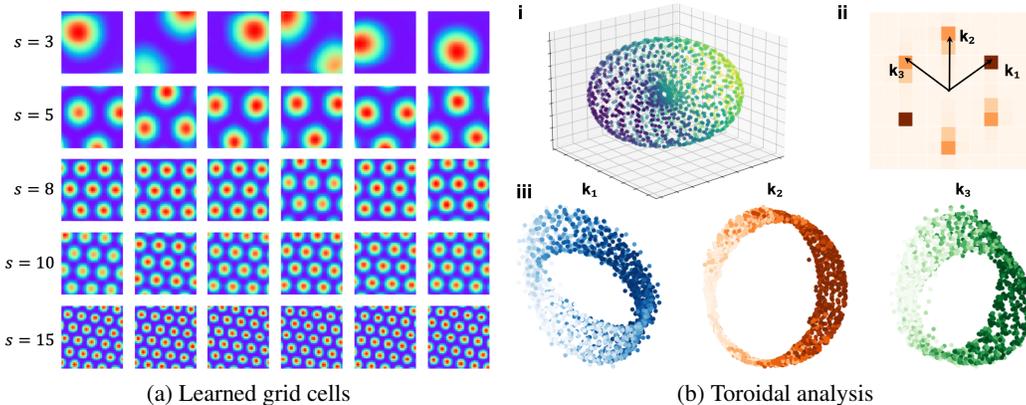
Figure 2: Hexagonal periodic patterns learned in linear models. (a) Learned patterns of linear models with different scales. (b) Toroidal structure spectral analysis of the activities of grid cells.

Hexagonal patterns. We first trained the linear model with manually assigned scaling factor s by minimizing $L = L_1 + \lambda L_2$. Figure 2(a) shows the learned firing patterns of $v(x)$ for linear models with the change of scaling factors s, which controlled the scale or metric. In Figure 2(a), each image represents the response map for a grid cell, with each row displaying 6 randomly selected response maps. The emergence of hexagonal periodic patterns in these response maps is evident. Consistency in scale and orientation is observed within each scale, though variations in phases or spatial shifts are apparent.

We also trained nonlinear models with various activation functions. In Figures 3(a-d), hexagonal periodic patterns still emerge with nonlinear transformations, demonstrating that the grid-like patterns are stable and easily learned regardless of the transformation type, grid scales, or number of neurons.

To evaluate how closely the learned patterns align with regular hexagonal grids, we recruited the most commonly used metric for quantifying grid cells, the gridness score, adopted by the neuroscience literature (Langston et al., 2010; Sargolini et al., 2006). The gridness scores and the valid rate were reported in Table 1. Compared to other existing learning-based approaches, our models exhibit notably high gridness scores (↑) and a high percentage of valid grid cells (↑).

Table 1: Gridness scores and validity rates of grid cells. The last two rows represent the results of our models.

MODEL	GRIDNESS(↑)	VALID RATE(↑)
BANINO ET AL. (2018)	0.18	25.2%
SORSCHER ET AL. (2023)	0.48	56.1%
GAO ET AL. (2021)	0.90	73.1%
OUR LINEAR	1.70	100.0%
OUR NONLINEAR	1.17	100.0%

Table 2: Scaling factor s and estimated scale for learned patterns in single-module linear models.

SCALING FACTOR	ESTIMATED SCALE
$s = 5$	0.82
$s = 10$	0.41
$s = 15$	0.27

We also investigated the relationship between the manually assigned scaling factor s and the estimated scale of the learned patterns following (Langston et al., 2010; Sargolini et al., 2006). As shown in Table 2, the estimated scale of the learned patterns is proportional to $1/s$.

Topological analysis. As discussed in Section 4.1, the joint activities of grid cells should reside on a torus-like manifold, and the positions on the torus correspond to the physical locations of a moving agent. To evaluate whether our empirically learned representations align with the topological properties of theoretical models, we employed a nonlinear dimension reduction method (spectral embedding Saul et al. (2006)) to show that grid cell states fell on a toroidal manifold as depicted in Figure 2b(i). To further investigate periodicity and orientation within the same module, we conducted numerical simulations of pattern forming dynamics. In Figure 2b(ii), we applied 2D Fourier transforms of the learned maps, revealing that the Fourier power is hexagonally distributed along 3 principal directions k_1, k_2, and k_3. Following Schøyen et al. (2022); Schaeffer et al. (2023), projecting the toroidal manifold onto the 3 vectors, we can observe 3 rings in Figure 2b(iii). This indicates the manifold has a 2D twisted torus topology.

(a) Nonlinear Model 1 (Tanh) w/o $v > 0$

(b) Nonlinear Model 1 (Tanh) w/ $v > 0$

(c) Nonlinear Model 1 (ReLU) w/ $v > 0$

(d) Nonlinear Model 2 (ReLU) w/ $v > 0$

(e) Linear model w/o $v > 0$

(f) Linear model w/o $|v| = 1$

(g) Linear model w/o L_2
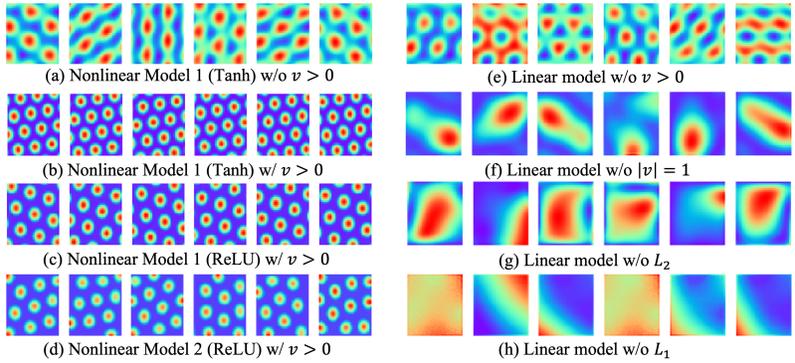
(h) Linear model w/o L_1

Figure 3: *Left*(a-d): Learned patterns for nonlinear models with different rectified functions. *Right*(e-h): Ablation for the linear model.

Ablation study. We show ablation results to investigate the empirical significance of each assumption for the emergence of hexagon grid patterns. First, we highlight the essential role of conformal isometry; in its absence, as shown in Figure 3(h), the response maps display non-hexagon patterns. Next, as shown in Figure 3(a) and (e), we also ablate the non-negative assumption. Without $v(x) \geq 0$, the hexagonal pattern still emerges. For the transformation and normalization assumptions, Figure 3(f) and (g) indicate that they are necessary. See more results for normalization assumption in G.2.

3.5 Numerical evidence for local conformal isometry
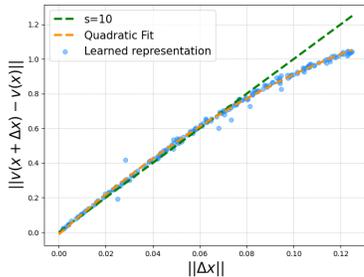
Figure 4: The relationship between $\|v(x + \Delta x) - v(x)\|$ and $\|\Delta x\|$ in the learned representations.

To further assess whether our learned model achieves local conformal isometry, we examine the relationship between $\|v(x + \Delta x) - v(x)\|$ and $\|\Delta x\|$ in our learned model. Specifically, we aim to verify if the learned representation exhibits the expected local isometry for local displacements Δx and how deviations arise as the displacement increases. Here we use the linear model with the metric $s = 10$ for analysis.

As shown in Figure 4, we randomly sample Δx within the range $(0, 0.125)$, and plot $\|v(x + \Delta x) - v(x)\|$ using the learned model. Then, we draw the linear reference line with slope $s = 10$, and also fit a quadratic curve to the data. The results show that the learned representation closely follows a linear relationship for local Δx, as expected from a conformally isometric embedding.

As Δx increases, we observe a quadratic deviation from the linear trend, due to the extrinsic curvature of the manifold $\mathbb{M}$. In our theoretical analysis, we shall analyze this deviation and show that the

hexagon flat torus achieves minimum deviation from local isometry, thus explaining the emergence of hexagonal grid patterns by minimizing our loss term on conformal isometry L_1 in equation 3.

3.6 NEUROSCIENCE EVIDENCE FROM NEURAL RECORDING DATA

To further investigate conformal isometry, we perform analysis on real neural recordings using data from Gardner et al. (2021), which contains simultaneous recording of hundreds of grid cells in rodents using Neuropixels probes. To test the conformal isometry hypothesis, we evaluate whether equation 2 holds in the recorded cell activations. Specifically, we calculate $\|\boldsymbol{v}(\boldsymbol{x}) - \boldsymbol{v}(\boldsymbol{x} + \Delta \boldsymbol{x})\|$ for different locations $\boldsymbol{x}$ and for varying local shifts $\Delta \boldsymbol{x}$. We expect to observe a linear relationship.

(a) Conformal isometry in neural data   (b) Histogram of $\|\boldsymbol{v}\|$ in neural data
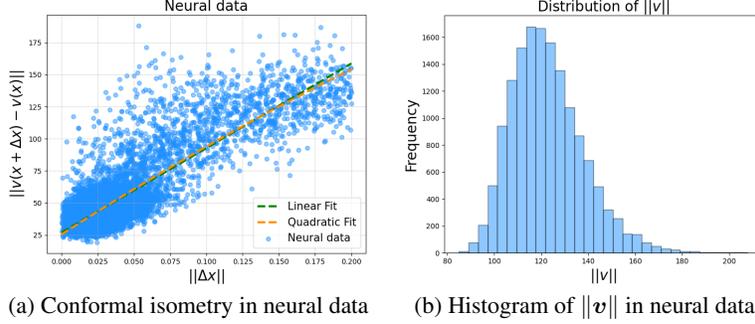
Figure 5: Analysis of real neural data. (a) A clear linear relationship between $\|\boldsymbol{v}(\boldsymbol{x} + \Delta \boldsymbol{x}) - \boldsymbol{v}(\boldsymbol{x})\|$ and $\|\Delta \boldsymbol{x}\|$ is shown in the real neural data. The unit of $\Delta \boldsymbol{x}$ is meter. (b) Distribution of $\|\boldsymbol{v}(\boldsymbol{x})\|$ in neural data.

In the analysis, we used data from one module, which includes 93 pure grid cells. As shown in Figure 5(a), despite the noise in the data, a clear linear relationship emerges between $\|\boldsymbol{v}(\boldsymbol{x}) - \boldsymbol{v}(\boldsymbol{x} + \Delta \boldsymbol{x})\|$ and $\|\Delta \boldsymbol{x}\|$. A quadratic polynomial fit closely matches the linear curve, indicating minimal deviation from linearity for local $\|\Delta \boldsymbol{x}\|$. Furthermore, we analyzed the distribution of $\|\boldsymbol{v}(\boldsymbol{x})\|$ in real neural data, finding a mean of 128.6 with a standard deviation of 15.1. When normalized to a mean of 1, the standard deviation becomes 0.12, indicating that the assumption of constant $\|\boldsymbol{v}(\boldsymbol{x})\|$ holds approximately.

Considering the noisy nature of neural responses and the fact that the data may only contain a subset of grid cells in the module, the results above are consistent with Assumption 1 of local conformal isometry and Assumption 3 of constant $\|\boldsymbol{v}(\boldsymbol{x})\|$ over $\boldsymbol{x}$. See Appendix J for more results.

4 THEORETICAL UNDERSTANDING

This section presents a theoretical understanding of the emergence of hexagon grid patterns when minimizing our loss function. We show that the hexagonal flat torus, which underlies grid cell patterns, achieves minimal deviation from local conformal isometry.

4.1 HEXAGON FLAT TORUS IS MAXIMALLY DISTANCE PRESERVING

We start by showing that the manifold of the grid cell representation $\mathbb{M} = (\boldsymbol{v}(\boldsymbol{x}), \forall \boldsymbol{x})$ forms a 2D torus.

Proposition 1. *The transformations* $(F(\cdot, \Delta \boldsymbol{x}), \forall \Delta \boldsymbol{x})$ *form a group acting on the manifold* $\mathbb{M} = (\boldsymbol{v}(\boldsymbol{x}), \forall \boldsymbol{x})$, *and the 2D manifold* $\mathbb{M}$ *has a torus topology.*

Proof. The group $(F(\cdot, \Delta \boldsymbol{x}), \forall \Delta \boldsymbol{x})$ is a representation of the 2D additive Euclidean group $(\mathbb{R}^2, +)$, i.e., $F(\boldsymbol{v}(\boldsymbol{x}), \Delta \boldsymbol{x}_1 + \Delta \boldsymbol{x}_2) = F(F(\boldsymbol{v}(\boldsymbol{x}), \Delta \boldsymbol{x}_1), \Delta \boldsymbol{x}_2) = F(F(\boldsymbol{v}(\boldsymbol{x}), \Delta \boldsymbol{x}_2), \Delta \boldsymbol{x}_1), \forall \boldsymbol{x}, \Delta \boldsymbol{x}_1, \Delta \boldsymbol{x}_2$, and $F(\boldsymbol{v}(\boldsymbol{x}), 0) = \boldsymbol{v}(\boldsymbol{x}), \forall \boldsymbol{x}$ (Gao et al., 2021). See Figure 1 for an illustration. Since $(\mathbb{R}^2, +)$ is an abelian Lie group, $(F(\cdot, \Delta \boldsymbol{x}), \forall \Delta \boldsymbol{x})$ is also an abelian Lie group. With Assumption 3 on normalization, $\|\boldsymbol{v}(\boldsymbol{x})\| = 1$, the manifold $(\boldsymbol{v}(\boldsymbol{x}), \forall \boldsymbol{x})$ is compact, and $(F(\cdot, \Delta \boldsymbol{x}), \forall \Delta \boldsymbol{x})$ is a compact group. It is also connected because the 2D domain is connected. According to a classical theorem

in Lie group theory (Dwyer & Wilkerson, 1998), a compact and connected abelian Lie group has a topology of a torus. Therefore, the 2D manifold $\mathbb{M}$ is a 2D torus. ∎

The torus topology is supported by neuroscience data (Gardner et al., 2021) as well as our numerical experiments. The torus manifold makes $v(x)$ a 2D periodic function of x. See Appendix B for a detailed explanation of the 2D periodic lattice.

Section 3.5 shows that our learned model achieves local isometry. Without loss of generality, assuming the conformal scaling factor $s = 1$. The 2D manifold $\mathbb{M} = (v(x), \forall x)$ is thus a flat torus with local isometry, and it has zero intrinsic curvature, meaning it has a flat Euclidean geometry intrinsically.

So far we have shown that the geometry of the grid cell representation is a torus and is flat locally. However, viewed from outside the manifold with bigger $\|\Delta x\|$, the extrinsic curvature of the torus within the ambient embedding space is non-zero, and our loss function $\mathbb{E}_{x,\Delta x}[(\|v(x + \Delta x) - v(x)\| - \|\Delta x\|)^2]$ minimizes the deviation from the local flatness for bigger $\|\Delta x\|$.

Let us assume $\|v(x + \Delta x) - v(x)\|^2 = \|\Delta x\|^2 + o(\|\Delta x\|^2)$ for infinitesimal $\|\Delta x\|$, enforced by the loss function and verified numerically in Section 3.5. For larger $\|\Delta x\|$, we further analyze $\|v(x + \Delta x) - v(x)\|^2 - \|\Delta x\|^2$ using a higher-order Taylor expansion, up to $o(\|\Delta x\|^4)$.

Definition 2. *We define notation for the derivatives of $v(x)$: $v^{(1)} = \partial v(x)/\partial x_1$, $v^{(2)} = \partial v(x)/\partial x_2$, $v^{(11)} = \partial^2 v(x)/\partial x_1^2$, $v^{(12)} = \partial^2 v(x)/\partial x_1 \partial x_2$, $v^{(22)} = \partial^2 v(x)/\partial x_2^2$, and their higher-order counterparts. The indices in the superscript indicate the variables with respect to which we take partial derivatives.*

Definition 3. *Assume $\|v(x + \Delta x) - v(x)\|^2 = \|\Delta x\|^2 + o(\|\Delta x\|^2)$. Define high order deviation from local flatness as $\|v(x + \Delta x) - v(x)\|^2 - \|\Delta x\|^2$.*

Proposition 4. *Assuming $\|v(x)\| = 1$ for every x, then the high order deviation*

$$\|v(x + \Delta x) - v(x)\|^2 - \|\Delta x\|^2 = -\frac{1}{12} D(\Delta x) + o(\|\Delta x\|^4). \tag{5}$$

where

$$
\begin{aligned}
D(\Delta x) =& \langle v^{(1111)}, v\rangle \Delta x_1^4 + \langle v^{(2222)}, v\rangle \Delta x_2^4 + 6\langle v^{(1122)}, v\rangle \Delta x_1^2 \Delta x_2^2 \\
&+ 4\langle v^{(1112)}, v\rangle \Delta x_1^3 \Delta x_2 + 4\langle v^{(1222)}, v\rangle \Delta x_1 \Delta x_2^3.
\end{aligned}
\tag{6}
$$

Proof. With $\|v(x)\|^2 = 1$, we have

$$\|v(x + \Delta x) - v(x)\|^2 - \|\Delta x\|^2 = 2 - 2\langle v(x + \Delta x), v(x)\rangle - (\Delta x_1^2 + \Delta x_2^2). \tag{7}$$

We can expand $v(x + \Delta x)$ by 4th-order Taylor expansion, and calculate the inner products between the derivatives and $v(x)$. When the manifold is locally flat and isometric for infinitesimal Δx, the first derivatives $v^{(1)}$ and $v^{(2)}$ form an orthonormal basis, i.e., $\langle v^{(1)}, v^{(2)}\rangle = 0$ and $\|v^{(i)}\|^2 = 1$ for $i = 1, 2$. By repeatedly differentiating these equations, we can derive a series of properties such as $\langle v^{(1)}, v(x)\rangle = \langle v^{(2)}, v(x)\rangle = 0$, $\langle v^{(11)}, v(x)\rangle = \langle v^{(22)}, v(x)\rangle = -1$, $\langle v^{(12)}, v(x)\rangle = 0$. Moreover, the inner products between the third-order derivatives and $v(x)$ all vanish. See Appendix C for detailed proof. Therefore, only the 4-th order terms $D(\Delta x)$ remain. ∎

This fourth-order term $D(\Delta x)$ captures deviation from local flatness or extrinsic curvature for larger $\|\Delta x\|$ (i.e., going beyond second-order $o(\|\Delta x\|^2)$ to up to $o(\|\Delta x\|^4)$).

We shall prove that among all flat tori (such as parallelogram, rectangle, square, and hexagon tori), the hexagonal torus minimizes the deviation derived above, due to its six-fold symmetry that evenly distributes the extrinsic curvature. We first establish the isotropy of hexagon flat torus.

Theorem 5. *If the torus $\mathbb{M} = (v(x), \forall x)$ is a hexagon flat torus, then $D(\Delta x) = c\|\Delta x\|^4$ for a constant coefficient c, i.e., $D(\Delta x)$ is isotropic.*

Proof. A hexagon torus possesses 6-fold rotational symmetry, meaning it is invariant under 60-degree rotations. Under such a rotation, the coordinates transform as:

$$\Delta x_1' = \frac{\sqrt{3}}{2}\Delta x_1 - \frac{1}{2}\Delta x_2; \quad \Delta x_2' = \frac{1}{2}\Delta x_1 + \frac{\sqrt{3}}{2}\Delta x_2. \tag{8}$$

To maintain symmetry, we require that $D(\Delta \boldsymbol{x}') = D(\Delta \boldsymbol{x})$, which leads to the conditions:

$$\langle \boldsymbol{v}^{(1111)}, \boldsymbol{v} \rangle = \langle \boldsymbol{v}^{(2222)}, \boldsymbol{v} \rangle = 3\langle \boldsymbol{v}^{(1122)}, \boldsymbol{v} \rangle = c; \; \langle \boldsymbol{v}^{(1112)}, \boldsymbol{v} \rangle = \langle \boldsymbol{v}^{(1222)}, \boldsymbol{v} \rangle = 0. \tag{9}$$

Thus, the deviation $D(\Delta \boldsymbol{x})$ becomes:

$$D(\Delta \boldsymbol{x}) = c(\Delta x_1^4 + \Delta x_2^4 + 2\Delta x_1^2 \Delta x_2^2) = c(\Delta x_1^2 + \Delta x_2^2)^2 = c\|\Delta \boldsymbol{x}\|^4, \tag{10}$$

which is constant in all directions. This shows that the hexagonal torus is locally isotropic up to the fourth-order term. See Appendix D for a detailed treatment. ∎

Define $\Delta r = \|\Delta \boldsymbol{x}\|$, $\Delta x_1 = \Delta r \cos \theta$, and $\Delta x_2 = \Delta r \sin \theta$. $D(\Delta \boldsymbol{x})$ is constant over direction θ for fixed Δr. As shown in Figure 6 of Appendix, only the hexagon lattice has six-fold symmetry. Other flat tori do not have six-fold symmetry and thus do not have isotropic $D(\Delta \boldsymbol{x})$. For instance, a square torus has 4-fold symmetry, and its $D(\Delta \boldsymbol{x})$ is not isotropic. An example is Clifford torus in 4D, $\boldsymbol{v}(\boldsymbol{x}) = (\cos x_1, \sin x_1, \cos x_2, \sin x_2)$, with $D(\Delta \boldsymbol{x}) \propto (\cos^4 \theta + \sin^4 \theta)\Delta r^4$, which is not isotropic.

Flat tori come with different sizes or surface areas, which are related to their average extrinsic curvatures. When comparing different tori, we should fix the size or average extrinsic curvature for fair comparison. Specifically, for any fixed Δr, we integrate over $\theta \in [0, 2\pi]$ to get expectation with respect to θ.

Theorem 6. *For any fixed average extrinsic curvature $\int D(\Delta \boldsymbol{x})d\theta$, the overall deviation from local flatness*

$$L(\Delta r) = \int (\|\boldsymbol{v}(\boldsymbol{x} + \Delta \boldsymbol{x}) - \boldsymbol{v}(\boldsymbol{x})\|^2 - \|\Delta \boldsymbol{x}\|^2)^2 d\theta = \frac{1}{12^2} \int (D(\Delta \boldsymbol{x}))^2 d\theta \tag{11}$$

is minimized if $D(\Delta \boldsymbol{x})$ is constant over θ.

Proof. This optimality is a result of Cauchy-Schwarz inequality. It can also be proved via the identity $\mathbb{E}_\theta(D(\Delta \boldsymbol{x})^2) = (\mathbb{E}_\theta(D(\Delta \boldsymbol{x}))^2 + \text{Var}_\theta(D(\Delta \boldsymbol{x}))$, which is minimized when $\text{Var}_\theta(D(\Delta \boldsymbol{x})) = 0$. This is achieved by the hexagon torus which is isotropic. Moreover, for the same hexagon torus, we can apply the above analysis and proves its optimality for all Δr. That is, the optimality of a hexagon torus is independent of Δr. ∎

This proves that the hexagon torus minimizes our loss function, i.e., the hexagon torus has local isometry up to $o(\|\Delta \boldsymbol{x}\|^2)$ and it has the minimum deviation from local isometry up to $o(\|\Delta \boldsymbol{x}\|^4)$. To conclude, the hexagon flat torus forms the maximally distance-preserving position embedding.

4.2 MULTIPLE MODULES OF CONFORMAL ISOMETRY

The minimalistic setting studied in the previous sections is about a single module of grid cells with a fixed scaling factor s. Grid cells form multiple modules or blocks (Barry et al., 2007; Stensola et al., 2012), and the response maps of grid cells within each module share the same scale and orientation. Thus, we can learn multiple modules, so that each module satisfies conform isometry with its own scaling factor s.

For a small s, the flat torus is big, and conformal isometry holds within a big range, but the resolution is low, i.e., $\boldsymbol{v}(\boldsymbol{x} + \Delta \boldsymbol{x})$ is close to $\boldsymbol{v}(\boldsymbol{x})$ in the presence of noises in neuron activities. On the contrary, for a big s, the flat torus is small, and conformal isometry holds within a small range, but the resolution is high,i.e., $\boldsymbol{v}(\boldsymbol{x} + \Delta \boldsymbol{x})$ is far from $\boldsymbol{v}(\boldsymbol{x})$ to resist noises in neuron activities.

The agent needs spatial awareness at different scales or resolutions, For instance, in playing golf, the first stroke and the final putt require very different s. Thus it needs multiple modules that satisfy conformal isometry at different scaling factors.

See Appendix H for a conformal modulation scheme to ensure the transformation RNN satisfies conformal isometry automatically. See Appendix I for learning multiple modules with such built-in conformal isometry by incorporating place cells.

5 RELATION TO PAST WORKS

The study of grid cells using machine learning models (Cueva & Wei, 2018; Banino et al., 2018; Sorscher et al., 2019; Whittington et al., 2020; Nayebi et al., 2021) is a topic of interest to both machine learning and computational neuroscience communities. Earlier papers (Cueva & Wei, 2018; Banino et al., 2018) adopted recurrent neural networks and used the objective of path integration to learn grid cells. Their models also involved interactions with place cells at the output layer, and the place cells are often modeled by predefined functions such as Gaussian tuning functions (Cueva & Wei, 2018; Gao et al., 2021; Xu et al., 2022). However, hexagon grid patterns do not consistently emerge from the learned models (Schaeffer et al., 2022). Thus extra assumptions or hypotheses are needed to explain the emergence of hexagon grid patterns.

One hypothesis is to assume the non-negativity of the grid cell responses coupled with Difference-of-Gaussian tuning functions for place cells, as in (Sorscher et al., 2019; Nayebi et al., 2021). However, the biological plausibility of center-surround Difference-of-Gaussian assumption is questioned (Schaeffer et al., 2022).

Another hypothesis is the conformal isometry hypothesis proposed in (Gao et al., 2021; Xu et al., 2022). However, these papers focused on learning multiple modules of grid cells using specific transformation models. Additionally, they introduced extra assumptions regarding place cells and the interactions between grid cells and place cells.

Our paper is built on (Gao et al., 2021; Xu et al., 2022), and we have achieved scientific reductionism by isolating the grid cell system of a single module, without making any assumptions about the place cells as well as the interactions between grid cells and place cells. We have also achieved generality by being agonistic of the form of the transformation model. This enables us to focus solely on the conformal isometry hypothesis and demonstrates its crucial role in the emergence of hexagon grid patterns. Moreover, we also provide neuroscience and theoretical evidences for this hypothesis. In our work, we study the conformal isometry hypothesis with explicit scaling factor s. In contrast, Gao et al. (2021); Xu et al. (2022) incorporate s implicitly within their learned models.

Schaeffer et al. (2023) proposed a self-supervised learning framework to learn grid cells of multiple modules, where conformal isometry is incorporated.

6 DISCUSSION: CONFORMAL ISOMETRY FOR PATH PLANNING

Conformal isometry ensures that the spatial representation in the brain preserves the geometry of the local environment, where distances are preserved locally. In terms of functional benefit, conformal isometry can be crucial for path planning. For instance, planning a straight path in an open field can be easily accomplished by following the steepest descent on the distance between neural representations of the target position and the current position on the path. This geometry-preserving property also facilitates the planning of more complex paths involving obstacles and landmarks. The scaling factor s allows path planning at different spatial scales or resolutions. This rationale suggests planning-centered representation learning in general, i.e., the learning of representations should serve the purpose of planning.

7 CONCLUSION

This paper investigates the conformal isometry hypothesis as a fundamental mathematical principle for the emergence of hexagonal periodic patterns in grid cells, both numerically and theoretically. The simplicity and naturalness of the conformal isometry hypothesis, along with its compelling geometric implications, make it a powerful framework for understanding grid cell patterns. Moreover, the conformal isometry hypothesis is also general enough that it leads to hexagon grid patterns regardless of the specific forms of transformation models or other settings. We believe that this hypothesis will serve as a foundation for further development of normative models of grid cells and enhance our understanding of their interactions with other systems.

ACKNOWLEDGMENT

REFERENCES

Haggai Agmon and Yoram Burak. A theory of joint attractor dynamics in the hippocampus and the entorhinal cortex accounts for artificial remapping and grid cell field-to-field variability. *eLife*, 9: e56894, 2020.

Daniel J Amit. *Modeling brain function: The world of attractor neural networks*. Cambridge university press, 1992.

Neil W Ashcroft, N David Mermin, et al. Solid state physics, 1976.

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

Caswell Barry, Robin Hayman, Neil Burgess, and Kathryn J Jeffery. Experience-dependent rescaling of entorhinal grids. *Nature neuroscience*, 10(6):682–684, 2007.

Jacob LS Bellmund, Peter Gärdenfors, Edvard I Moser, and Christian F Doeller. Navigating cognition: Spatial codes for human thinking. *Science*, 362(6415):eaat6766, 2018.

Hugh T Blair, Adam C Welday, and Kechen Zhang. Scale-invariant memory representations emerge from moire interference between grid fields that produce theta oscillations: a computational model. *Journal of Neuroscience*, 27(12):3211–3229, 2007.

Charlotte N Boccara, Michele Nardin, Federico Stella, Joseph O'Neill, and Jozsef Csicsvari. The entorhinal cognitive map is attracted to goals. *Science*, 363(6434):1443–1447, 2019.

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

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.

Jonathan J Couey, Aree Witoelar, Sheng-Jia Zhang, Kang Zheng, Jing Ye, Benjamin Dunn, Rafal Czajkowski, May-Britt Moser, Edvard I Moser, Yasser Roudi, et al. Recurrent inhibitory circuitry as a mechanism for grid formation. *Nature neuroscience*, 16(3):318–324, 2013.

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.

Christopher J Cueva, Peter Y Wang, Matthew Chin, and Xue-Xin Wei. Emergence of functional and structural properties of the head direction system by optimization of recurrent neural networks. *International Conferences on Learning Representations (ICLR)*, 2020.

Licurgo de Almeida, Marco Idiart, and John E Lisman. The input–output transformation of the hippocampal granule cells: from grid cells to place fields. *Journal of Neuroscience*, 29(23): 7504–7512, 2009.

Christian F Doeller, Caswell Barry, and Neil Burgess. Evidence for grid cells in a human memory network. *Nature*, 463(7281):657, 2010.

Yedidyah Dordek, Daniel Soudry, Ron Meir, and Dori Derdikman. Extracting grid cell characteristics from place cell inputs using non-negative principal component analysis. *Elife*, 5:e10094, 2016.

William Dorrell, Peter E Latham, Timothy EJ Behrens, and James CR Whittington. Actionable neural representations: Grid cells from minimal constraints. *arXiv preprint arXiv:2209.15563*, 2022.

William Gerard Dwyer and CW Wilkerson. The elementary geometric structure of compact lie groups. *Bulletin of the London Mathematical Society*, 30(4):337–364, 1998.

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

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.

Marianne Fyhn, Sturla Molden, Menno P Witter, Edvard I Moser, and May-Britt Moser. Spatial representation in the entorhinal cortex. *Science*, 305(5688):1258–1264, 2004.

Marianne Fyhn, Torkel Hafting, Menno P Witter, Edvard I Moser, and May-Britt Moser. Grid cells in mice. *Hippocampus*, 18(12):1230–1238, 2008.

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.

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. In *Neural Information Processing Systems*, 2021.

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

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, Liora Las, Dori Derdikman, and Nachum Ulanovsky. Are grid cells used for navigation? on local metrics, subjective spaces, and black holes. *Neuron*, 111(12): 1858–1875, 2023.

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, 2005.

Aidan J Horner, James A Bisby, Ewa Zotow, Daniel Bush, and Neil Burgess. Grid-like processing of imagined navigation. *Current Biology*, 26(6):842–847, 2016.

Joshua Jacobs, Christoph T Weidemann, Jonathan F Miller, Alec Solway, John F Burke, Xue-Xin Wei, Nanthia Suthana, Michael R Sperling, Ashwini D Sharan, Itzhak Fried, et al. Direct recordings of grid-like neuronal activity in human spatial navigation. *Nature neuroscience*, 16(9):1188, 2013.

Nathaniel J Killian, Michael J Jutras, and Elizabeth A Buffalo. A map of visual space in the primate entorhinal cortex. *Nature*, 491(7426):761, 2012.

Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

Rosamund F Langston, James A Ainge, Jonathan J Couey, Cathrin B Canto, Tale L Bjerknes, Menno P Witter, Edvard I Moser, and May-Britt Moser. Development of the spatial representation system in the rat. *Science*, 328(5985):1576–1580, 2010.

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, 2006.

May-Britt Moser and Edvard I Moser. Where am i? where am i going? *Scientific American*, 314(1): 26–33, 2016.

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.

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.

John O'keefe and Lynn Nadel. Précis of o'keefe & nadel's the hippocampus as a cognitive map. *Behavioral and Brain Sciences*, 2(4):487–494, 1979.

Hugh Pastoll, Lukas Solanka, Mark CW van Rossum, and Matthew F Nolan. Feedback inhibition enables theta-nested gamma oscillations and grid firing fields. *Neuron*, 77(1):141–154, 2013.

Thomas Ridler, Jonathan Witton, Keith G Phillips, Andrew D Randall, and Jonathan T Brown. Impaired speed encoding is associated with reduced grid cell periodicity in a mouse model of tauopathy. *bioRxiv*, pp. 595652, 2019.

David C Rowland, Horst A Obenhaus, Emilie R Skytøen, Qiangwei Zhang, Cliff G Kentros, Edvard I Moser, and May-Britt Moser. Functional properties of stellate cells in medial entorhinal cortex layer ii. *Elife*, 7:e36664, 2018.

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.

Lawrence K Saul, Kilian Q Weinberger, Fei Sha, Jihun Ham, and Daniel D Lee. Spectral methods for dimensionality reduction. *Semi-supervised learning*, 3, 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, Cristóbal Eyzaguirre, Sanmi Koyejo, and Ila Rani Fiete. Self-supervised learning of representations for space generates multi-modular grid cells. *arXiv preprint arXiv:2311.02316*, 2023.

Vemund Schøyen, Markus Borud Pettersen, Konstantin Holzhausen, Marianne Fyhn, Anders Malthe-Sørenssen, and Mikkel Elle Lepperød. Coherently remapping toroidal cells but not grid cells are responsible for path integration in virtual agents. *bioRxiv*, pp. 2022–08, 2022.

Vemund Sigmundson Schoyen, Kosio Beshkov, Markus Borud Pettersen, Erik Hermansen, Konstantin Holzhausen, Anders Malthe-Sorenssen, Marianne Fyhn, and Mikkel Elle Lepperod. Hexagons all the way down: Grid cells as a conformal isometric map of space. *bioRxiv*, pp. 2024–02, 2024.

Ben Sorscher, Gabriel Mel, Surya Ganguli, and Samuel Ocko. A unified theory for the origin of grid cells through the lens of pattern formation. *Advances in neural information processing systems*, 32, 2019.

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.

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

Kimberly L Stachenfeld, Matthew M Botvinick, and Samuel J Gershman. The hippocampus as a predictive map. *Nature neuroscience*, 20(11):1643, 2017.

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

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, 2012.

Edward C Tolman. Cognitive maps in rats and men. *Psychological review*, 55(4):189, 1948.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

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.

James CR Whittington, Joseph Warren, and Timothy EJ Behrens. Relating transformers to models and neural representations of the hippocampal formation. *arXiv preprint arXiv:2112.04035*, 2021.

Dehong Xu, Ruiqi Gao, Wen-Hao Zhang, Xue-Xin Wei, and Ying Nian Wu. Conformal isometry of lie group representation in recurrent network of grid cells. *arXiv preprint arXiv:2210.02684*, 2022.

Michael M Yartsev, Menno P Witter, and Nachum Ulanovsky. Grid cells without theta oscillations in the entorhinal cortex of bats. *Nature*, 479(7371):103, 2011.

Sheng-Jia Zhang, Jing Ye, Chenglin Miao, Albert Tsao, Ignas Cerniauskas, Debora Ledergerber, May-Britt Moser, and Edvard I Moser. Optogenetic dissection of entorhinal-hippocampal functional connectivity. *Science*, 340(6128), 2013.

Minglu Zhao, Dehong Xu, Deqian Kong, Wen-Hao Zhang, and Ying Nian Wu. A minimalistic representation model for head direction system. *arXiv preprint arXiv:2411.10596*, 2024.

A MORE TRAINING DETAILS

All the models were trained on a single 2080 Ti GPU for $200,000$ iterations with a learning rate of 0.003. For batch size, we generated 4000 simulated data for each iteration. For single-module models, the training time is less than 15 minutes. For multi-module models, the training time is less than 1 hour.

B 2D PERIODIC FUNCTION

Since x and Δx are 2D, the torus formed by $(F(\cdot, \Delta x), \forall \Delta x)$ is 2D, then its topology is $\mathbb{S}_1 \times \mathbb{S}_1$, where each $\mathbb{S}_1$ is a circle. Thus we can find two 2D vectors Δx_1 and Δx_2, so that $F(\cdot, \Delta x_1) = F(\cdot, \Delta x_2) = F(\cdot, 0)$. As a result, $v(x)$ is a 2D periodic function so that $v(x + k_1 \Delta x_1 + k_2 \Delta x_2) = v(x)$ for arbitrary integers k_1 and k_2. We assume Δx_1 and Δx_2 are the shortest vectors that characterize the above 2D periodicity. According to the theory of 2D Bravais lattice (Ashcroft et al., 1976), any 2D periodic lattice can be defined by two primitive vectors $(\Delta x_1, \Delta x_2)$.

If the scaling factor s is constant, as the position x of the agent moves from 0 to Δx_1 in the 2D space, $v(x)$ traces a perfect circle in the neural space due to conformal isometry, i.e., the geometry of the trajectory of $v(x)$ is a perfect circle up to bending or folding but without distortion by stretching. The same with movement from 0 to Δx_2.

Named after Auguste Bravais (1811-1863), the theory of Bravais lattice was developed for the study of crystallography in solid-state physics.

In 2D, a periodic lattice is defined by two primitive vectors $(\Delta x_1, \Delta x_2)$, and there are 5 different types of periodic lattices, as shown in Figure 6.

For Fourier analysis, we need to find the primitive vectors in the reciprocal space, (a_1, a_2), via the relation: $\langle a_i, \Delta x_j \rangle = 2\pi \delta_{ij}$, where $\delta_{ij} = 1$ if $i = j$, and $\delta_{ij} = 0$ otherwise.

For a 2D periodic function $f(x)$ on a lattice whose primitive vectors are (a_1, a_2) in the reciprocal space, define $\omega_k = k_1 a_1 + k_2 a_2$, where $k = (k_1, k_2)$ are a pair of integers (positive, negative, and zero), the Fourier expansion is $f(x) = \sum_k \hat{f}(\omega_k) e^{i \langle \omega_k, x \rangle}$.
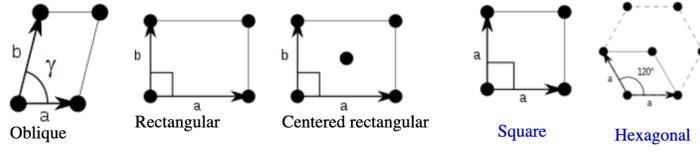
Figure 6: 2D periodic lattice is defined by two primitive vectors.

C   DERIVATIVE PROPERTIES FOR FLAT TORI

Given:

- $\|v(x)\|^2 = 1$ for all x
- The metric tensor $g_{ij} = \langle v^{(i)}, v^{(j)} \rangle = \delta_{ij}$

Let us derive the properties:

Lemma 7. $\langle v^{(1)}, v(x) \rangle = \langle v^{(2)}, v(x) \rangle = 0$

Proof. Differentiate $\|v(x)\|^2 = 1$ with respect to x_1:

$$\frac{\partial}{\partial x_1} \langle v(x), v(x) \rangle = 2\langle v^{(1)}, v(x) \rangle = 0$$

Therefore, $\langle v^{(1)}, v(x) \rangle = 0$. Similarly for x_2. ∎

Lemma 8. $\langle v^{(11)}, v(x) \rangle = \langle v^{(22)}, v(x) \rangle = -1$

Proof. Differentiate $\langle v^{(1)}, v(x) \rangle = 0$ with respect to x_1:

$$\langle v^{(11)}, v(x) \rangle + \langle v^{(1)}, v^{(1)} \rangle = 0$$

But $\langle v^{(1)}, v^{(1)} \rangle = 1$ (from the metric tensor), so $\langle v^{(11)}, v(x) \rangle = -1$. Similarly for $v^{(22)}$. ∎

Lemma 9. $\langle v^{(12)}, v(x) \rangle = 0$

Proof. Differentiate $\langle v^{(1)}, v(x) \rangle = 0$ with respect to x_2:

$$\langle v^{(12)}, v(x) \rangle + \langle v^{(1)}, v^{(2)} \rangle = 0$$

But $\langle v^{(1)}, v^{(2)} \rangle = 0$ (from the metric tensor), so $\langle v^{(12)}, v(x) \rangle = 0$. ∎

Now, let's derive the third-order terms:

Theorem 10. $\langle v^{(111)}, v(x) \rangle = \langle v^{(222)}, v(x) \rangle = \langle v^{(112)}, v(x) \rangle = \langle v^{(122)}, v(x) \rangle = 0$

Proof. 1) For $\langle v^{(111)}, v(x) \rangle$: Differentiate $\langle v^{(11)}, v(x) \rangle = -1$ with respect to x_1:

$$\langle v^{(111)}, v(x) \rangle + \langle v^{(11)}, v^{(1)} \rangle = 0$$

But $\langle v^{(11)}, v^{(1)} \rangle = 0$ (differentiate $\langle v^{(1)}, v^{(1)} \rangle = 1$), so $\langle v^{(111)}, v(x) \rangle = 0$.

2) Similarly, $\langle v^{(222)}, v(x) \rangle = 0$.

3) For $\langle v^{(112)}, v(x) \rangle$: Differentiate $\langle v^{(11)}, v(x) \rangle = -1$ with respect to x_2:

$$\langle v^{(112)}, v(x) \rangle + \langle v^{(11)}, v^{(2)} \rangle = 0$$

But $\langle v^{(11)}, v^{(2)} \rangle = 0$, which we can show by differentiating $\langle v^{(1)}, v^{(2)} \rangle = 0$

$$\langle v^{(11)}, v^{(2)} \rangle + \langle v^{(1)}, v^{(12)} \rangle = 0$$

$\langle v^{(1)}, v^{(12)} \rangle = 0$, which can be derived by differentiating $\langle v^{(1)}, v^{(1)} \rangle = 1$ with respect to x_2. Therefore, $\langle v^{(112)}, v(x) \rangle = 0$.

4) Similarly, $\langle v^{(122)}, v(x) \rangle = 0$. ∎

These derivations show that all third-order inner products $\langle \boldsymbol{v}^{(ijk)}, \boldsymbol{v}(\boldsymbol{x}) \rangle$ are zero for a flat torus with the given conditions.

D Hexagonal symmetry in 4th-order polynomials

Theorem 11. *If a 4th-order polynomial $f(x, y) = ax^4 + bx^3y + cx^2y^2 + dxy^3 + ey^4$ has hexagonal symmetry (invariant under $60°$ rotation), then it must be of the form $f(x, y) = a(x^2 + y^2)^2$.*

This result can be applied to our analysis of hexagon flat torus in the main text, where the 4th-order terms in the Taylor expansion are isotropic under hexagonal symmetry.

Proof. Let us rotate our coordinate system (x, y) by an angle θ to align with the reflection symmetry axes of the hexagon. Let (x', y') be the new coordinate system, so that $f(x, y) = g(x', y')$, which is a 4th-order polynomial of x' and y'. In this new (x', y') coordinate system, we have reflection symmetries:

- $x' \rightarrow -x', y' \rightarrow y'$

- $x' \rightarrow x', y' \rightarrow -y'$

Due to these reflection symmetries, in the (x', y') system, our polynomial must take the form:

$$g(x', y') = Ax'^4 + Bx'^2y'^2 + Cy'^4 \tag{12}$$

Without loss of generality, we can assume $A = 1$. Thus:

$$g(x', y') = x'^4 + Bx'^2y'^2 + Cy'^4 \tag{13}$$

Due to $60°$ rotational symmetry, we must have:

- $g(1, 0) = g(1/2, \sqrt{3}/2)$

- $g(0, 1) = g(\sqrt{3}/2, 1/2)$

These conditions give us two equations:

$$1 = (1/2)^4 + B(1/2)^2(\sqrt{3}/2)^2 + C(\sqrt{3}/2)^4$$
$$1 = 1/16 + 3B/16 + 9C/16$$

$$\tag{14}$$

$$C = (\sqrt{3}/2)^4 + B(\sqrt{3}/2)^2(1/2)^2 + (1/2)^4$$
$$C = 9/16 + 3B/16 + 1/16$$

Solving these equations simultaneously yields:

$$B = 2 \quad \text{and} \quad C = 1 \tag{15}$$

Therefore, in the rotated coordinate system:

$$g(x', y') = x'^4 + 2x'^2y'^2 + y'^4 = (x'^2 + y'^2)^2 \tag{16}$$

Since $(x^2 + y^2)$ is invariant under rotation, we can conclude that in the original coordinate system:

$$f(x, y) = a(x^2 + y^2)^2 \tag{17}$$

where we have reintroduced the coefficient a that we normalized to 1 earlier.

Thus, we have proved that if a 4th-order polynomial has hexagonal symmetry, it must be of the form $f(x, y) = a(x^2 + y^2)^2$.

$\blacksquare$

E PERMUTATION GROUP AND RELATION TO ATTRACTOR NETWORK

The learned response maps of the grid cells in the same module are shifted versions of each other, i.e., there is a discrete set of displacements $\{\Delta x\}$, such as for each Δx in this set, we have $v_i(x + \Delta x) = v_j(x)$, where $j = \sigma(i, \Delta x)$, and σ is a mapping from i to j that depends on Δx. In other words, Δx causes a permutation of the indices of the elements in $v(x)$, and $F(\cdot, \Delta x) \cong \sigma(\cdot, \Delta x)$, that is, the transformation group for the discrete set of Δx is equivalent to a subgroup of the permutation group. This is consistent with hand-designed CANN (Amit, 1992; Burak & Fiete, 2009; Couey et al., 2013; Pastoll et al., 2013; Agmon & Burak, 2020) . A CANN places grid cells on a 2D "neuron sheet" with periodic boundary condition, i.e., a 2D torus, and lets the movement of the "bump" formed by the activities of grid cells mirror the movement of the agent in a conformal way, and the movement of the "bump" amounts to cyclic permutation of the neurons. Our model does not assume such an *a priori* 2D torus neuron sheet, and is much simpler and more generic.

F EIGEN ANALYSIS OF TRANSFORMATION

For the general transformation, $v(x + \Delta x) = F(v(x), \Delta x)$, we have

$$v(x) = F(v(x), 0), \tag{18}$$
$$v(x + \Delta x) = F(v(x + \Delta x), 0), \tag{19}$$

thus

$$\Delta v = v(x + \Delta x) - v(x) = F'_v(v(x))\Delta v + o(\|\Delta v\|), \tag{20}$$

where

$$F'_v(v) = \frac{\partial}{\partial \Delta} F(v + \Delta, 0) \mid_{\Delta=0}. \tag{21}$$

Thus Δv is in the 2D eigen subspace of $F'_v(v(x))$ with eigenvalue 1.

At the same time,

$$v(x + \Delta x) = F(v(x), \Delta x) = v(x) + F'_{\Delta x}(v(x))\Delta x, \tag{22}$$

where

$$F'_{\Delta x}(v) = \frac{\partial}{\partial \Delta x} F(v, \Delta x) \mid_{\Delta x=0}. \tag{23}$$

Thus

$$\Delta v = F'_{\Delta x}(v(x))\Delta x, \tag{24}$$

that is, the two columns of $F'_{\Delta x}(v(x))$ are the two vectors that span the eigen subspace of $F'_v(v(x))$ with eigenvalue 1. If we further assume conformal isometry, then the two column vectors of $F'_{\Delta x}(v(x))$ are orthogonal and of equal length, so that Δv is conformal to Δx.

The above analysis is about $v(x)$ on the manifold. We want the remaining eigenvalues of $F'_v(v(x))$ to be less than 1, so that, off the manifold, $F(v, 0)$ will bring v closer to the manifold, i.e., the manifold consists of attractor points of F, and F is an attractor network.

G MORE EXPERIMENT RESULTS ON MINIMAL SETTING

G.1 MORE ABLATION STUDIES ON ACTIVATION FUNCTIONS

To further examine the impact of the transformation model F, we conducted additional ablation studies using different activation functions in nonlinear models. As illustrated in Figure 7, we tested Leaky ReLU and Swish activations. Both successfully produced hexagonal patterns, demonstrating that the specific form of F does not dictate the emergence of grid patterns. This highlights the robustness of our approach to variations in the transformation model.

(a) Nonlinear Model 1 w/ Leaky ReLU   (b) Nonlinear Model 1 w/ Swish
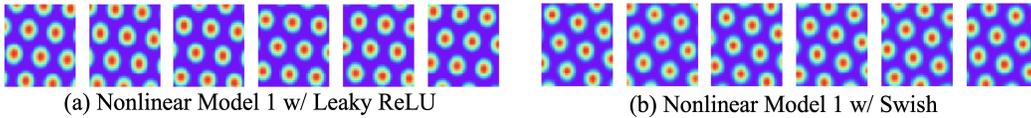
Figure 7: More activation functions: (a) Leaky ReLU and (b) Swish.

G.2 INFLUENCE OF NORMALIZATION ASSUMPTION

Our normalization assumption, $\|\boldsymbol{v}\| = 1$, is critical for the emergence of hexagonal grid patterns, as shown in our ablation study where removing normalization prevents their formation. To further explore this, we conducted experiments with imperfect normalization by introducing noise: $1 - \epsilon \leq \|\boldsymbol{v}(\boldsymbol{x})\| \leq 1 + \epsilon$. During training, in the normalization process, we allow $\|\boldsymbol{v}\|$ to randomly vary within the range $(1 - \epsilon, 1 + \epsilon)$. This enables us to assess the maximum allowable ϵ that still preserves hexagonal patterns.

As shown in Figure 8, we tested four different noise levels. The results indicate that hexagonal patterns are robust to small deviations in normalization but collapse as ϵ becomes large, highlighting the importance of precise normalization for pattern stability.

(a) $\varepsilon = 0.05$ (b) $\varepsilon = 0.1$

(c) $\varepsilon = 0.2$ (d) $\varepsilon = 0.4$
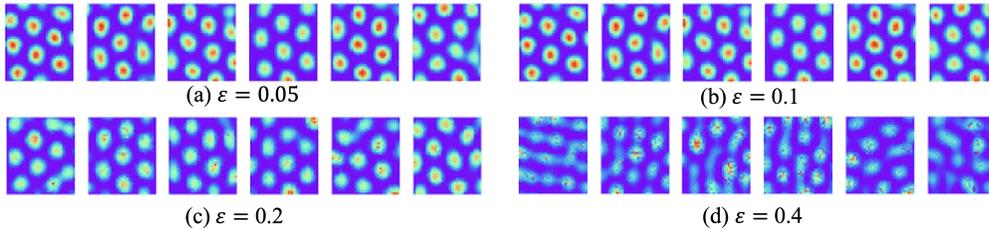
Figure 8: Imperfect normalization with different noise levels.

G.3 INFLUENCE OF THE RANGE OF DISPLACEMENT

As mentioned in the experiments, for L_1, we enforce the constraint $s\|\Delta\boldsymbol{x}\| \leq 1.25$ to ensure local conformal isometry. To further explore the influence of the range of $\Delta\boldsymbol{x}$ on the patterns, we conducted additional experiments by increasing and decreasing the range. As shown in Figure 9, we perturbed the range and found that the emergence of hexagonal patterns remains stable, demonstrating the robustness of the framework to variations in this hyper-parameter.

$s \parallel \Delta x \parallel \leq 1.5$

$s \parallel \Delta x \parallel \leq 1.25$

$s \parallel \Delta x \parallel \leq 1$
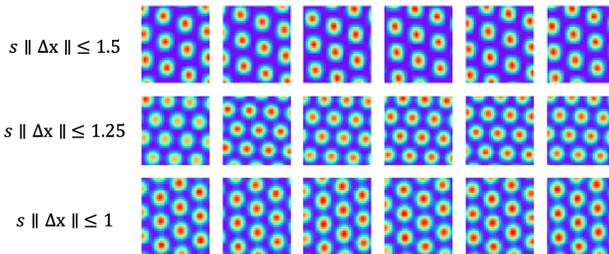
Figure 9: Varying the range of $\|\Delta\boldsymbol{x}\|$ in conformal training.

H CONFORMAL MODULATION FOR BUILT-IN CONFORMAL ISOMETRY

So far we have used a conformal isometry loss to constrain the conformal isometry property (equation 3). Conformal isometry can be directly baked into the transformation model via a mechanism we call conformal modulation of input velocity so that conformal isometry is automatically satisfied.

H.1 Conformal modulation

Consider the general recurrent transformation $v(x+\Delta x) = F(v(x), \Delta x) = F(v(x), \Delta r, \theta)$, where $\Delta x = (\Delta x_1 = \Delta r \cos \theta, \Delta x_2 = \Delta r \sin \theta)$, θ is the heading direction, and Δr is the displacement, the directional derivative of F at (v, θ) is defined as

$$f(v, \theta) = \frac{\partial}{\partial a} F(v, a, \theta) \mid_{a=0} . \tag{25}$$

With the above definition, the first order Taylor expansion of the recurrent transformation at $\Delta r = 0$ gives us

$$v(x + \Delta x) = v(x) + f(v(x), \theta)\Delta r + o(\Delta r). \tag{26}$$

The conformal modulation of Δr is defined as

$$\overline{\Delta r} = \frac{s}{\|f(v(x), \theta)\|} \Delta r, \tag{27}$$

where $\| \cdot \|$ is the ℓ_2 norm, and s is a learnable parameter. With conformal modulation, the transformation is changed to

$$v(x + \Delta x) = \overline{F}(v(x), \Delta r, \theta) = F(v(x), \overline{\Delta r}, \theta). \tag{28}$$

Then the first order Taylor expansion gives

$$v(x + \Delta x) = v(x) + f(v(x), \theta)\overline{\Delta r} + o(\Delta r) = v(x) + s\overline{f}(v(x), \theta)\Delta r + o(\Delta r), \tag{29}$$

where $\overline{f}(v, \theta) = f(v, \theta)/\|f(v, \theta)\|$ is a unit vector with $\|\overline{f}(v, \theta)\| = 1$, which leads to $\|v(x + \Delta x) - v(x)\| = s\|\Delta x\| + o(\|\Delta x\|)$, i.e., conformal isometry. Thus $\overline{F}(v(x), \Delta r, \theta) = \overline{F}(v(x), \Delta x)$ satisfies conformal isometry automatically.

H.2 Implementation of conformal modulation

For the linear model $v(x + \Delta x) = v(x) + B(\theta)v(x)\Delta r$, the directional derivative is $f(v, \theta) = B(\theta)v(x)$. The conformal modulation is

$$\overline{\Delta r} = \frac{s\Delta r}{\|B(\theta)v(x)\|}. \tag{30}$$

The linear transformation model then becomes

$$v(x + \Delta x) = v(x) + s\frac{B(\theta)v(x)}{\|B(\theta)v(x)\|}\Delta r. \tag{31}$$

The above model is similar to the "add + layer norm" operations in the Transformer model (Vaswani et al., 2017).

For the nonlinear model $v(x + \Delta x) = R(Av(x) + B(\theta)v(x)\Delta r + b)$, where A is a learnable matrix, and $R(\cdot)$ is element-wise nonlinear rectification, the directional derivative is $f(v, \theta) = R'(Av) \odot B(\theta)v$, where $R'(\cdot)$ is calculated element-wise, and $\odot$ is element-wise multiplication. Thus, the nonlinear transformation model becomes

$$v(x + \Delta x) = R(Av(x) + s\frac{B(\theta)v(x)}{\|R'(Av) \odot B(\theta)v\|}\Delta r + b). \tag{32}$$

While the linear model is defined for $v(x)$ on the manifold $\mathbb{M}$, the nonlinear model further constrains $v(x) = R(Av(x))$ for $v(x) \in \mathbb{M}$, where $\Delta r = 0$. If $R(Av)$ is furthermore a contraction for v that are off $\mathbb{M}$, then $\mathbb{M}$ consists of the attractors of $R(Av)$ for v around $\mathbb{M}$. The nonlinear model then becomes a continuous attractor neural network (CANN) (Amit, 1992; Burak & Fiete, 2009; Couey et al., 2013; Pastoll et al., 2013; Agmon & Burak, 2020).

In Appendix I, we learn multiple modules of grid cells within a combined model with both place cells and grid cells, where each module undergoes its own transformation with conformal modulation, and each module has a learnable scaling factor s.

I MULTI-MODULE SETTING WITH PLACE CELLS

We can extend the multimodal setting defined in Section 4.2 to further incorporate the interaction with place cells. Define the adjacency kernel $A(\boldsymbol{x}, \boldsymbol{x}') = \exp(-\|\boldsymbol{x} - \boldsymbol{x}'\|^2/2\sigma^2)$ for a certain scale parameter σ. $A(\boldsymbol{x}, \boldsymbol{x}')$ is a model of the response map (as a function of $\boldsymbol{x}$) of a place cell for the place $\boldsymbol{x}'$. We can then use $\boldsymbol{v}(\boldsymbol{x})$ to decode A via $A(\boldsymbol{x}, \boldsymbol{x}') = \langle \boldsymbol{w}(\boldsymbol{x}'), \boldsymbol{v}(\boldsymbol{x}) \rangle$, where $\boldsymbol{w}(\boldsymbol{x}')$ is the decoding vector associated with place cell $\boldsymbol{x}'$. Below we demonstrate the learning of such a system.

I.1 MULTI-MODULE SETUP

The conformal modulation follows the definition in equation 31 and equation 32. For multi-modules, in our model, the vector representations $\boldsymbol{v}(\boldsymbol{x})$ are divided into sub-vectors analogous to modules. Accordingly, we assume that the transformation is also module-wise by assuming $\boldsymbol{B}(\theta)$ is block-diagonal where each block corresponds to a module.

I.2 LEARNING MULTIPLE MODULES

We assume multiple modules (sub-vectors) in $\boldsymbol{v}(\boldsymbol{x})$, each has its own transformation with conformal modulation. The loss consists of two terms:

$$L_0 = \mathbb{E}_{\boldsymbol{x}, \boldsymbol{x}'}[(A(\boldsymbol{x}, \boldsymbol{x}') - \langle \boldsymbol{w}(\boldsymbol{x}'), \boldsymbol{v}(\boldsymbol{x}) \rangle)^2], \tag{33}$$

$$L_2 = \mathbb{E}_{\boldsymbol{x}, \Delta\boldsymbol{x}}[\|\boldsymbol{v}(\boldsymbol{x} + \Delta\boldsymbol{x}) - \overline{F}(\boldsymbol{v}(\boldsymbol{x}), \Delta\boldsymbol{x})\|^2], \tag{34}$$

where $\overline{F}()$ is the transformation after conformal modulation so that we do not need conformal isometry loss term L_1. We assume $\boldsymbol{w}(\boldsymbol{x}') \geq 0$ because the connections from grid cells to place cells are excitatory (Zhang et al., 2013; Rowland et al., 2018). We do not enforce the non-negativity of $\boldsymbol{v}(\boldsymbol{x})$. We continue to enforce $\|\boldsymbol{v}(\boldsymbol{x})\| = 1$.

In the numerical experiments, $A(\boldsymbol{x}, \boldsymbol{x}')$ is the given input, and we jointly learn the position embedding $\boldsymbol{v}(\boldsymbol{x})$, read-out weights $\boldsymbol{w}(\boldsymbol{x}')$, and the transformation model F by minimizing the total loss: $L = L_0 + \lambda_1 L_2$.

I.3 NUMERICAL EXPERIMENTS

We conducted numerical experiments to train multi-module models with linear and nonlinear transformations. Each module consisted of 24 cells, with 15 modules for the linear model and 8 for the nonlinear model. For the response maps of the place cells, we used a Gaussian adjacency kernel with $\sigma = 0.07$. Regarding the transformation, the one-step displacement Δr was restricted to be smaller than 3 grids. Additionally, the scaling factor s was treated as a learnable parameter for each module.

As shown in Figure 10(a-c), each row shows the units belonging to the same module and we randomly present 3 cells in each module. The empirical results illustrate that multi-scale hexagonal grid firing patterns emerge in the learned $\boldsymbol{v}(\boldsymbol{x})$ across all models with conformal modulation. Figures 11 and 12 show more patterns.

Additionally, Figure 10(d) shows the histogram of grid scales of the learned grid cell neurons, which follows a multi-modal distribution. The distribution is best fitted by a mixture of 3 Gaussians with means 0.57, 0.77, and 1.01. The ratios between successive modes are 1.37 and 1.31. The scale relationships in our learned grid patterns across modules align with both theoretical predictions (Stemmler et al., 2015; Wei et al., 2015) and empirical observations from rodent grid cells (Stensola et al., 2012).

I.4 PATH INTEGRATION

Suppose the agent starts from $\boldsymbol{x}_0$, with vector representation $\boldsymbol{v}_0 = \boldsymbol{v}(\boldsymbol{x}_0)$. If the agent makes a sequence of moves $(\Delta\boldsymbol{x}_t, t = 1, ..., T)$, then the vector $\boldsymbol{v}$ is updated by $\boldsymbol{v}_t = F(\boldsymbol{v}_{t-1}, \Delta\boldsymbol{x}_t)$. At time t, the self-position of the agent can be decoded by $\hat{\boldsymbol{x}} = \arg\max_{\boldsymbol{x}'} \langle \boldsymbol{v}_t, \boldsymbol{w}(\boldsymbol{x}') \rangle$, i.e., the place $\boldsymbol{x}'$ that is the most adjacent to the self-position represented by $\boldsymbol{v}_t$. This enables the agent to infer and keep track of its position based on its self-motion even in darkness.

Following Gao et al. (2021); Xu et al. (2022); Zhao et al. (2024), we assess the ability of the learned model to execute accurate path integration in two different scenarios. First of all, for path integration

(a) Linear   (b) Nonlinear(GELU)   (c) Nonlinear(Tanh)   (d) Grid scale distribution
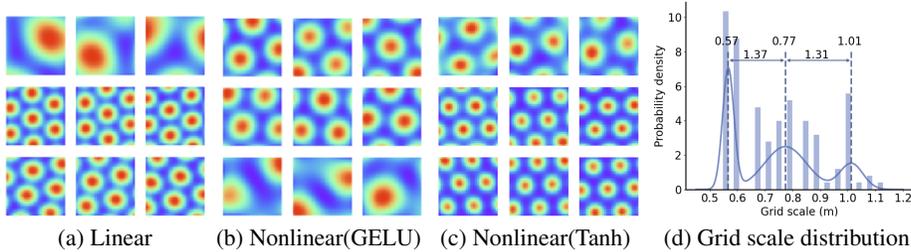
Figure 10: Results of linear and nonlinear models with GELU and Tanh activations and grid scale distribution.
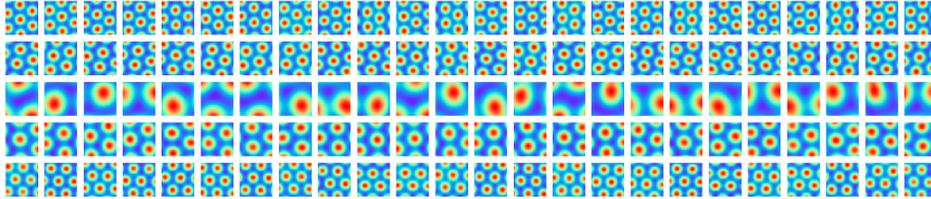
Figure 11: In the linear model, hexagon grid firing patterns are observed in the learned $v(x)$. Each row displays the firing patterns of all the cells within a single module, with each module comprising 24 cells. The units illustrate the neuron activity throughout the entire 2D square environment. The figure presents patterns from five randomly chosen modules.

with re-encoding, we decode $v \to \hat{x}$ to the 2D physical space via $\hat{x} = \arg\max_{x'} \langle v, w(x') \rangle$, and then encode $v \leftarrow v(\hat{x})$ back to the neuron space intermittently. This approach aids in rectifying the errors accumulated in the neural space throughout the transformation. Conversely, in scenarios excluding re-encoding, the transformation is applied exclusively using the neuron vector v. In Figure 13(A), the model adeptly handles path integration up to 30 steps (short distance) without the need for re-encoding. The figure illustrates trajectories with a fixed step size of three grids, enhancing the visibility of discrepancies between predicted and actual paths. It is important to note that the physical space was not discretized in our experiments, allowing the agent to choose any step size flexibly. For path integration with longer distances, we evaluate the learned model for 100 steps over 300 trajectories. As shown in Figure 13(B), with re-encoding, the path integration error for the last step is as small as 0.003, while the average error over the 100-step trajectory is 0.002. Without re-encoding, the error is relatively larger, where the average error over the entire trajectory is approximately 0.024, and it reaches 0.037 for the last step.

J   CONFORMAL ISOMETRY IN NEURAL DATA

J.1   MORE DETAILS AND RESULTS FOR NEURAL DATA

In Figure 5, the real neural recording that we used is rat 'R' day 1, module 1 from Gardner et al. (2021). To provide more results for real neural data, here we choose another module of grid cells (rat 'R' day 1, module 3) to further evaluate conformal isometry.

As shown in Figure 14(a), we can see similar results as in the main paper: a linear relationship between $\|v(x) - v(x + \Delta x)\|$ and $\|\Delta x\|$, and a very small deviation from the linear relationship in local $\|\Delta x\|$. Also, similar distribution shape of $\|v\|$ is shown in Figure 14(b). This provides more empirical evidence of the conformal isometry hypothesis.

J.2   QUANTITATIVE COMPARISON BETWEEN NEURAL DATA AND NUMERICAL SIMULATION

In this section, we present additional quantitative analysis to further compare real neural data with our numerical simulations in the context of conformal isometry.

(a) Tanh activation   (b) GELU activation
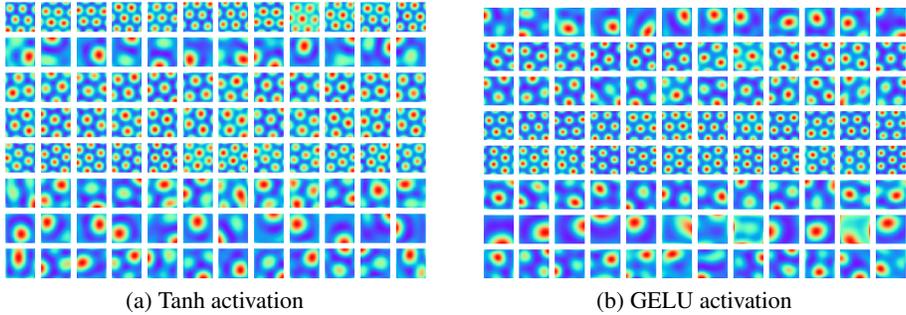
Figure 12: Results of the non-linear models. We randomly chose 8 modules and showed the firing patterns with different rectification functions.
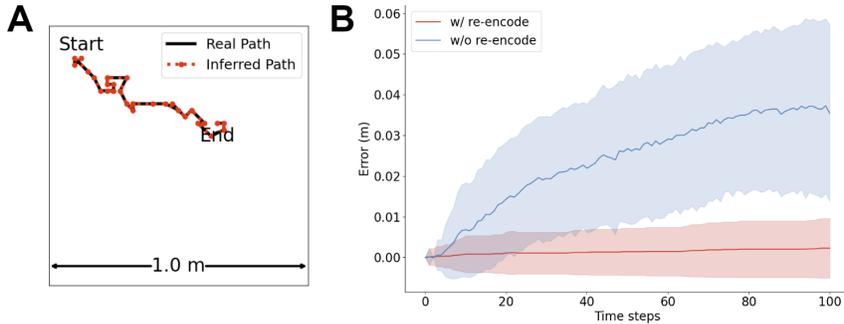
Figure 13: Results for path integration. (A) Path integration for 30 steps without re-encoding. The black line represents the real trajectory and the red one is the predicted trajectory by the learned model. (B) Results for long-distance (100-step) path integration error with and without re-encoding over time by the non-linear model.

To align with the settings in our simulation, we first normalize the side length of the square domain (1.5 m) in the real neural data to a unit length of 1, as is the case with our numerical simulation. Based on the scale of the grid module (rat R, module 1, day 1) reported in Gardner et al. (2021), the spatial grid spacing is 0.5 m, which corresponds to 0.33 spatial grid spacing in our numerical simulation after normalization (0.5m/1.5m). We also normalize $\|v\| = 1$ in the neural data, noting that this pre-processing step was not applied in the original results shown in Figure 5.

To quantitatively examine the scaling factor s in the normalized neural data, we first fit a linear relationship between $\|v(x + \Delta x) - v(x)\|$ (after normalization) and $\|\Delta x\|$ with the intercept fixed at 0. As shown in Figure 15, the fitted scaling factor is $s = 13.4$. In Table 2, we show that the estimated spatial grid spacing of the learned patterns is inversely proportional to s. Specifically, the product of the scaling factor and the estimated spacing in the real neural data ($13.4 \times 0.33 = 4.42$) is comparable to our numerical simulation result ($10 \times 0.41 = 4.1$).

To further examine the local conformal isometry, we fit a quadratic curve to the data, similar to the approach in Section 3.5. The deviation arises as $\|\Delta x\|$ is larger than 0.08, and this result is in agreement with the result in our numerical simulation.

In summary, the real neural data is in agreement with the results in our numerical simulation in terms of the relationship between the scaling factor s and the estimated grid spacing, as well as the range of conformal isometry observed.

(a) Conformal isometry in neural data

(b) Histogram for $\|\boldsymbol{v}\|$ in neural data
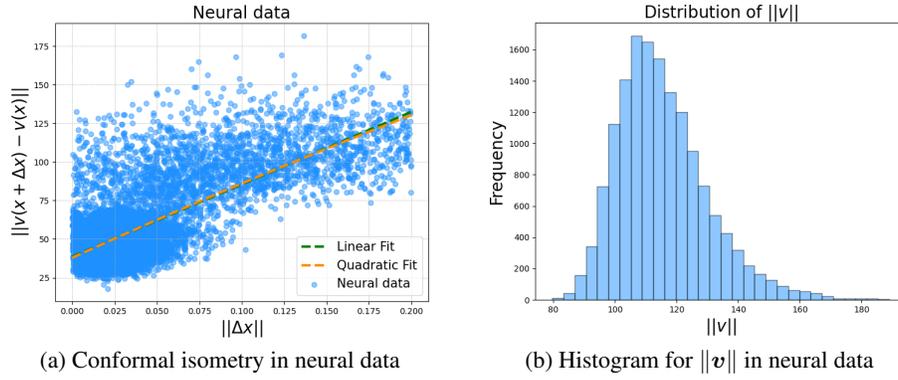
Figure 14: Results for a different module of grid cells.
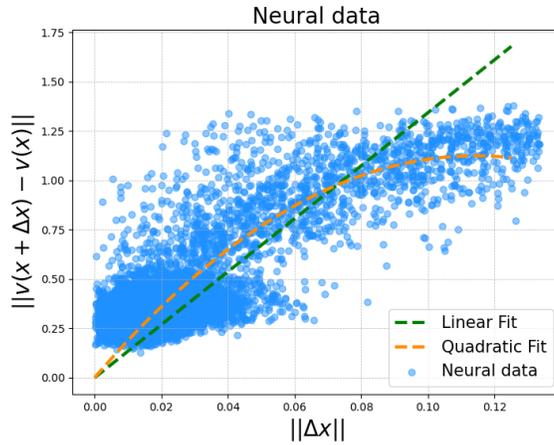
Figure 15: Conformal isometry in normalized neural data.