Dimensionality of population-level latent mechanisms encoding spatial representations

Nur Efsan Delikkaya*, **Alperen Cimen***, **Francisco Acosta**, **Adele Myers**, **Andy Alexander**, **Fatih Dinc**†, **Nina Miolane**†
University of California, Santa Barbara, CA, USA

Abstract

How does the brain efficiently encode space, and can this be achieved with low-dimensional neural codes? In this work, we show that the answer depends critically on whether spatial information is encoded continuously or discretely. We trained recurrent neural networks (RNNs) on two path integration tasks: one with continuous spatial outputs and another with discrete, place-cell-like responses. RNNs encoding continuous spatial coordinates developed low-dimensional latent dynamics, whereas in the latter case, the spatial resolution demanded by the place fields, rather than their number, determined the dimensionality of the neural code. Overall, by shifting focus from individual neuron tuning to population-level representations, our work identifies a fundamental constraint on the computational resources required for different neural coding strategies in spatial navigation.

Keywords: Spatial navigation, path-integrating RNNs

The question of how the brain compactly represents space has been closely tied to the discovery of grid cells. These neurons, first identified in the medial entorhinal cortex, exhibit a striking hexagonal lattice of firing fields across the environment (Gardner et al., 2022). Several theoretical frameworks have attempted to explain their emergence. A particularly influential line of work treated the firing patterns of individual neurons as the fundamental "basis functions" of spatial representation (Sorscher et al., 2019, 2023). Under certain assumptions, interference between sinusoidal components was shown to generate hexagonal lattice patterns resembling grid cell activity. However, these models typically produced only a single module of grid cells and focused on their tuning properties rather than the required dimensionality of the basis functions and their rich characteristics. Later work argued that explaining the coexistence of multiple modules of grid cells requires a self-supervised framework, and even questioned whether grid cells directly support path integration, the process of computing position by integrating velocity over time, via outputs to place cells (Schaeffer et al., 2022).

In parallel, computational neuroscience, primarily focusing on short-term memory and visual tasks, has developed a complementary perspective. Rather than emphasizing the tuning of individual neurons, this approach highlights how populations of neurons collectively encode task variables in low-dimensional latent representations (Dinc et al., 2025; Dubreuil et al., 2022; Beiran et al., 2021; Valente et al., 2022; Mastrogiuseppe and Ostojic, 2018). From this viewpoint, path integration can be studied as an important problem in its own right and arguably a more central concern for the brain than producing grid cells. Specifically, grid-like responses may or may not arise from efficient coding principles (Schaeffer et al., 2022) or from biological constraints such as the non-negativity of neural

* Equal contribution
† Co-supervision

© 2025 N.E. Delikkaya,, A. Cimen,, F. Acosta,, A. Myers,, A. Alexander,, F. Dinc, & N. Miolane.

responses (Sorscher et al., 2023), but biological networks must solve the path integration problem to self-center and navigate in the real world. This motivates the perspective shift we take in this work: Can spatial encoding be carried out within low-dimensional latent codes, or does it require high-dimensional dynamics?

To date, low-dimensional latent dynamical systems have been sufficient to explain many cognitive tasks within the population-code perspective (Valente et al., 2022; Dubreuil et al., 2022). This success has fostered the view that low-dimensional solutions may be a general feature of neural computation (Perich et al., 2025). As we show in this work, spatial navigation, however, provides a decisive test case that spans both regimes. When space is represented discretely, navigation cannot be achieved with a low-dimensional latent code and instead requires training an inherently high-dimensional dynamical system to handle the numerous place cell outputs. By contrast, when space is represented continuously, the low-dimensionality of the output enables path integration to be solved with a low-dimensional dynamical system.

Results. In this work, we ask: What is the optimal way to represent space in a latent population code? To address this question, we treat the basis functions of spatial representation as latent variables, implemented by neural populations through linear combinations (Dinc et al., 2025). Our focus is not on how these representations are learned, but on the consequences once they are in place.

Studying compact representations of space: Building on Sorscher et al. (2023), we formulate spatial encoding as an optimization problem and extend this framework to examine the dimensionality and properties of the basis functions when outputs are either discrete (place-cell-like) or continuous. The key distinction in our formulation is that we do not assume individual neurons themselves constitute the basis functions. Instead, we treat the basis functions as latent variables and ask: How many are truly needed? For clarity, our analytical treatment in this extended abstract concerns a one-dimensional environment. The generalization to two dimensions is straightforward, which is also what we use to perform our simulations.

To start with, we focus on representing the space with discrete, place-like, outputs, which will cover the continuous case trivially. Let $\{f_\mu(x)\}$ denote the family of tuning functions associated with place cells indexed by μ. Here, μ denotes the location of the place field, whereas $f_\mu(x)$ is simply translated across the field based on μ. For instance, $f_\mu(x) = \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$ for some spatial resolution σ, where $x \in \Omega$ denotes position in the environment. Our goal is to study the joint approximation of each function by a set of orthonormal basis functions $\{\phi_k(x)\}_{k=1}^K$:

$$f_\mu(x) \approx \sum_{k=1}^{K} w_k(\mu)\, \phi_k(x), \tag{1}$$

where $w_k(\mu)$ specifies the output weight from the kth basis function to produce μth place cell. Since $\phi_k(t)$ are thought of as basis functions, we further enforce an orthonormality constraint: $\int_\Omega \phi_j(x)\phi_k(x)\, dx = \delta_{jk}$, where Ω represents the spatial domain to be covered (*e.g.*, $\Omega = \{x : x \in [0, L]\}$ for some length L). Then, following (Sorscher et al., 2019, 2023), the optimal set of basis functions is defined as the one that minimizes the mean squared reconstruction error $\mathbb{E}_\mu \int_\Omega \left[f_\mu(x) - \sum_{k=1}^K w_k(\mu)\phi_k(x)\right]^2 dx$. This problem reduces

to the classical Karhunen–Loève (KL) expansion, or functional principal component analysis (Ramsay and Silverman, 2005). The solution is determined by the correlation kernel $C(x, x') = \mathbb{E}_\mu [f_\mu(x) f_\mu(x')]$, whose eigenfunctions provide the optimal basis:

$$\int_\Omega C(x, x')\, \phi_k(x')\, dx' = \lambda_k \phi_k(x). \tag{2}$$

Compact spatial representations are obtained by using the leading eigenfunctions of $C(x, x')$ as the basis functions, whereas the error in the reconstruction is obtained by summing over the eigenvalues λ_k that belong to the eigenfunctions not picked as the basis function. For instance, using the top K components, the minimized loss function becomes $\mathcal{L}_K = \sum_{j>K} \lambda_j$.

Due to space constraints (pun not intended), we focus on the dimensionality of the learned representations in this extended abstract. In future work, we plan to expand on our preliminary results: These basis functions turn out to be Fourier series under uniform coverage of space by place cells, and their perturbations could explain recent empirical evidence about distortions to tuning properties of individual cells under specific landmark and/or reward locations.

Dimensionality of the learned representations: Having established the concept of basis functions, we now ask: how many such modes are required to represent space at a given resolution? For a continuous representation in a general K-dimensional space, the answer is straightforward. We can replace $f_\mu(x)$ with $\{f_1(x) = x_1, f_2(x) = x_2, \ldots, f_K(x) = x_K\}$, where x_i is the ith coordinate of the current location. In this case, the number of basis functions is K, *e.g.*, two basis functions are sufficient for representing $\mathbb{R}^2$.

The discrete case, however, is more subtle. To see why, suppose place fields have characteristic width σ, so that each tuning function can be expressed as a rescaled template $f_\mu(x) = g\left(\frac{x-\mu}{\sigma}\right)$, where g specifies the field shape (*e.g.*, Gaussian). Substituting this form into the kernel gives $C(x, x') = \tilde{C}\left(\frac{x}{\sigma}, \frac{x'}{\sigma}\right)$ for some scale-invariant coordinates $\tilde{x} = x/\sigma$ and a corresponding kernel function $\tilde{C}$. Then, we can consider the eigenvalue problem in Eq. (2). Using $C(x, x') = \tilde{C}\left(\frac{x}{\sigma}, \frac{x'}{\sigma}\right)$ and defining rescaled variables $\tilde{x} = x/\sigma$ and $\tilde{x}' = x'/\sigma$, we obtain $\sigma \int_{\tilde{\Omega}} \tilde{C}(\tilde{x}, \tilde{x}') \phi_k(\sigma \tilde{x}') d\tilde{x}' = \lambda_k \phi_k(\sigma \tilde{x})$, so that $\lambda_k(\sigma) = \sigma \tilde{\lambda}_k(\sigma)$, where $\tilde{\Omega} = \Omega/\sigma$ and $\tilde{\lambda}_k(\sigma)$ are the eigenvalues of the rescaled kernel. This shows two effects: an explicit prefactor σ from the Jacobian, and an implicit dependence through the effective domain size $|\tilde{\Omega}| \sim L/\sigma$ for the rescaled kernel $\tilde{C}(\tilde{x}, \tilde{x}')$. Notably, the latter effect is often stronger than a polynomial decay, hence tends to dominate, in practice (see below for an example).

Since the optimal loss function value, $\mathcal{L}(K) = \sum_{j>K} \lambda_j$, is controlled by the first neglected eigenvalue λ_{K+1} that decays fast with j, the remaining loss value after using K basis functions is determined by $O(\sigma \tilde{\lambda}_{K+1}(\sigma))$. To see why narrower fields require more modes K to achieve the same accuracy, consider the translation-invariant Gaussian kernel with scale σ, $C(x, x') = \exp(-\frac{(x-x')^2}{\sigma^2})$. In this case, the eigenfunctions are Fourier modes (Sorscher et al., 2023) with eigenvalues $\lambda(q; \sigma) \propto \sigma \exp(-\sigma^2 q^2/4)$, which decay exponentially once $|q| \gtrsim 1/\sigma$. Thus the number of significant modes scales as $K \sim L/\sigma$: finer resolution (smaller σ) demands higher-dimensional latent codes, while broader fields reduce dimensionality at the expense of precision. This is rather intuitive, since a larger σ means that outputs are correlated with each other and therefore the effective output dimensionality decreases even if there are many place cells.

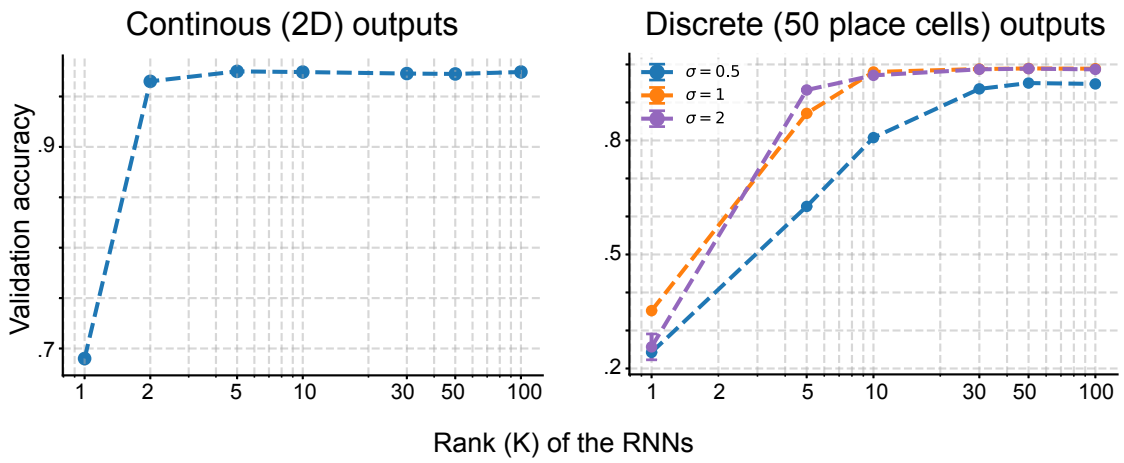

Figure 1: **Path-integration is a low- or high-dimensional process when the output is continuous or discrete, respectively.** We trained low-rank RNNs to perform path integration in a rectangular 2D domain. In the continuous output case, the RNNs were instructed to output continuously coded spatial locations. In the discrete output case, RNN outputs were linear projections from the neural activities to 50 uniformly sampled Gaussian-shaped place cells with s.d. σ. Accuracy refers to the average (discrete-case) or aggregate (continuous-case) correlations (Pearson's r) between the target and predicted the outputs.

Tests with path-integrating RNNs: We tested the dimensionality of learned solutions in RNNs trained to perform path integration, in which the velocity inputs and an initial location needed to be transformed into a location estimate throughout a trial. We compared two output formats: (i) a discrete representation with many place-cell–like outputs and (ii) a continuous representation with two outputs corresponding to spatial coordinates in 2D. In this setting, the rank of the RNN constrains the dimensionality of its latent representation, and thus the number of basis functions available for spatial coding. The results matched theoretical expectations (Fig. 1). For continuous outputs, rank-two networks were sufficient to achieve near-perfect performance, since only two latent variables were needed to represent the two coordinates. In contrast, for discrete outputs, performance improved gradually with increasing rank, and accurate path integration required higher-dimensional networks. In this case, the number of required ranks depending on the resolution σ that set the width of the place-cell tuning curves.

Conclusion. In this work, we showed that spatial navigation can rely on low-dimensional dynamics in the continuous case or high-dimensional dynamics in the discrete case, with dimensionality determined by spatial resolution for the latter. In other words, place-cell responses require high-dimensional latent representations resulting from the effective high-dimensionality of the required outputs. This suggests that their function is not necessarily efficiency but likely other computational goals, such as robustness to noise, flexible remapping, or compatibility with downstream readouts that rely on sparse, localized responses.

References

Manuel Beiran, Alexis Dubreuil, Adrian Valente, Francesca Mastrogiuseppe, and Srdjan Ostojic. Shaping dynamics with multiple populations in low-rank recurrent networks. *Neural Computation*, 33(6):1572–1615, 2021.

Fatih Dinc, Marta Blanco-Pozo, David Klindt, Francisco Acosta, Yiqi Jiang, Sadegh Ebrahimi, Adam Shai, Hidenori Tanaka, Peng Yuan, Mark J Schnitzer, et al. Latent computing by biological neural networks: A dynamical systems framework. *arXiv preprint arXiv:2502.14337*, 2025.

Alexis Dubreuil, Adrian Valente, Manuel Beiran, Francesca Mastrogiuseppe, and Srdjan Ostojic. The role of population structure in computations through neural dynamics. *Nature Neuroscience*, pages 1–12, 2022.

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.

Francesca Mastrogiuseppe and Srdjan Ostojic. Linking connectivity, dynamics, and computations in low-rank recurrent neural networks. *Neuron*, 99(3):609–623, 2018.

Matthew G Perich, Devika Narain, and Juan A Gallego. A neural manifold view of the brain. *Nature Neuroscience*, pages 1–16, 2025.

James O Ramsay and Bernard W Silverman. *Functional data analysis*. Springer, 2005.

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.

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.

Adrian Valente, Jonathan W Pillow, and Srdjan Ostojic. Extracting computational mechanisms from neural data using low-rank rnns. *Advances in Neural Information Processing Systems*, 35:24072–24086, 2022.