An Integrated Deep Learning Based Model of Hippocampal Spatial Cells That Combines Self-motion With Sensory Information

Azra Aziz  
Indian Institute of Technology Madras

S S Sree Harsha Peesapati  
Indian Institute of Technology Madras

N Rohan  
Indian Institute of Technology BHU

V Srinivasa Chakravarthy (schakra@ee.iitm.ac.in)  
Indian Institute of Technology Madras

Research Article

Keywords: Grid cells, Place cells, PCA, Deep Networks, Bessel function

Posted Date: April 5th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-368836/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License
An integrated deep learning based model of hippocampal spatial cells that combines self-motion with sensory information

Azra Aziz¹, Peesapati S S Sreeharsha¹, Rohan N², *V Srinivasa Chakravarthy¹

¹Computational Neuroscience Lab, Indian Institute of Technology Madras, Chennai, Tamil Nadu, India, 600036
²Department of Electronics Engineering, Indian Institute of Technology (BHU), Varanasi, Uttar Pradesh, 221005
*V Srinivasa Chakravarthy, 506, Department of Biotechnology, Indian Institute of Technology Madras, Chennai, Tamil Nadu, India, 600036, Phone no. 9500069322, schakra@ee.iitm.ac.in

Keywords: Grid cells, Place cells, PCA, Deep Networks, Bessel function
Abstract

A special class of hippocampal neurons broadly known as the spatial cells, whose subcategories include place cells, grid cells and head direction cells, are considered to be the building blocks of the brain’s map of the spatial world. We present a general, deep learning-based modeling framework that describes the emergence of the spatial cell responses and can also explain behavioral responses that involve a combination of path integration and vision. The first layer of the model consists of Head Direction (HD) cells that code for preferred direction of the agent. The second layer is the path integration (PI) layer with oscillatory neurons: displacement of the agent in a given direction modulates the frequency of these oscillators. Principal Component Analysis (PCA) of the PI cell responses showed emergence of cells with grid-like spatial periodicity. We show that the response of these cells could be described by Bessel functions. The output of PI layer is used to train stack of autoencoders. Neurons of both the layers exhibit responses resembling grid cells and place cells. The paper concludes by suggesting a wider applicability of the proposed modeling framework beyond the two simulated behavioral studies.

Introduction

The realization that Deep neural networks (DNNs) match human performance in certain perceptual classification tasks has prompted researchers to consider DNNs, not just as a tool for solving artificial pattern recognition problems, but as a potential model of human behavior. In the visual domain, for example, DNNs trained on visual object recognition matched even human error patterns across object classes, variation of viewpoint shape, and judgement of object similarity. However, the similarity between DNN and human error pattern stops at the level of object class and does not carry over to individual images. In the auditory domain also, DNNs trained on speech and music recognition closely match human performance.

Encouraged by the above successes of DNNs in emulating human behavior, researchers explored the possibility of DNNs being able to predict neural responses in sensory areas of the brain. Although DNN hidden layer responses do not resemble responses of brain’s sensory cortical areas in themselves, it was shown that they can predict neural responses in both visual and auditory domains. Lower layers of DNNs are able to better predict activity of primary sensory cortices while deeper layers could predict activities of higher cortical areas in both the sensory domains. These relatively recent developments open the possibility that DNN models, despite drawbacks like use of a biologically unrealistic learning algorithm, when appropriately constrained and interpreted by knowledge from neurobiology, can serve as suitable and effective models of brain function.

Unrelated to deep learning approach, there is a long line of theoretical studies that sought to model hippocampal spatial cells. Classical models of grid cells fall under two broad categories: oscillatory interference models and attractor-based models. Solstad et al describe a model that can produce place cells from grid cells; the
model, however, is not trainable and depends on special constraints on the weights connecting the grid cell layer to the place cell layer.

There are, however, trainable networks with a small number of stages (1 or 2), capable of reproducing hippocampal spatial cell responses. Dordek et al.\textsuperscript{16} describe a trainable model that can generate grid cells from place cell responses. But the model depends on a crucial and unrealistic periodic boundary condition on the input space. Cueva and Wei\textsuperscript{17} showed that a recurrent neural network trained to merely perform path integration of velocity inputs to estimate position can exhibit neural responses analogous to hippocampal spatial cells like grid cells, border cells and band cells. Kanitscheider and Fiete\textsuperscript{18} show that a recurrent neural network with Long Short-Term Memory (LSTM) blocks in the hidden layer can be trained to solve the problem of Simultaneous Localization and Mapping (SLAM). However, the last model does not claim to reproduce some of the common hippocampal spatial cells like grid or place cells.

Soman et al.\textsuperscript{19,20} describe a hierarchical model that performs path integration by integrating velocity into the phases of a layer of oscillators, followed by a cascade of layers in which the weights are trained by Hebbian (feedforward) and anti-Hebbian (lateral) learning. This model in 2 dimensions was able to generate predominantly grid cells in the lower layer and place cells in the higher layer. The model, when extended to 3D navigation in bats, was able to reproduce 3D place cells and a novel cell type dubbed the plane cell\textsuperscript{19}. However, the last two modelling approaches were able to explain single cell responses but do not model behavioral studies. There is a need to develop integrated models that successfully account for single neuron responses and at the same time be able to explain navigational behavior.

In this study, we present two approaches namely, a PCA based model and a deep learning-based model of spatial navigation that exhibit neural responses of hippocampal spatial cells like grid cells and place cells, and also explain behavioral experiments involving rats navigating two-dimensional mazes. We begin with the oscillator model of Soman et al.\textsuperscript{20} for path integration, but eliminate the time-dependency using an averaging process. We also use scale parameter ($\beta$) as an additional parameter in path integration (PI). Four different variations of path integration are considered. As an effort to establish a theory behind the emergence of the grid cell behavior, we propose the PCA model which generates grid cell-like patterns while providing insights into the possible mathematical nature of these encoded representations, revealing the relation of these patterns to Bessel functions. When the responses of path integration layer are used to train a hierarchical deep network, both layers show mixed response of grid cell-like firing and place cell-like firing. However, different PI variations gave different responses for grid cell-like firing as well as place cell-like firing and is discussed in results section. The network thus developed is used to model two experimental studies\textsuperscript{21,22}. Hippocampus combines self-motion information with sensory information to generate spatial representations of the world. In order to show that the proposed model
can work with such combinations, we simulate an experimental study in which rats explore a multi-compartmental environment with inhomogeneous lighting conditions. Our model is able to reproduce the changes in the place cell firing patterns in response to environmental lighting conditions.

**Methods**

In this study, we propose two models for representing spatial cell responses – 1) PCA-based model, and 2) a deep learning approach. The model inherently consists of one major assumption: the animal is assumed to be a point animal eliminating the effects of head rotation. So, the head direction is always along the direction of movement of the animal.

**Trajectory**: A virtual point animal is allowed to explore a square shaped environment of size 2 units. From a modelling viewpoint, the points of the trajectory have been generated using a uniform distribution of points within the region of the square environment. Moreover, there are some constraints imposed on preventing the animal from taking sudden turns and collisions with the walls. These points are then interpolated using “splines” algorithm to get a smooth trajectory. To simulate the trajectories for modelling experimental studies, the movement constraints are introduced based on the experimental conditions. This trajectory is the input to the proposed PCA based model and Autoencoder based model.

Both the deep neural network model and the PCA model mentioned above start with the two layers described below.

**Head Direction (HD) layer**: This layer consists of 100 neurons each having its preferred direction spanning 360 degrees (Fig: 1a). The responses of head direction cells are modelled by the equation 1:

\[ HD_i = v \cdot u_i \]  

where, \( v \) is the velocity of the animal and \( u_i \), a unit vector, is the preferred direction of the \( i^{th} \) HD neuron.

**Path Integration (PI) layer**: This layer consists of oscillatory neurons that have one-to-one connection with neurons from the HD layer (Soman et al 2018a, 2018b). The process involved in this layer can be divided into two major stages. The first stage is frequency modulation of the HD responses followed by low pass filtering to eliminate time-dependent high frequency oscillations as the second stage which gives the path integration output (Fig: 1a).

The first stage of frequency modulation (FM) of PI is described by the following equations:
\[ PI_{fm} = \sin[\omega t + \beta \int (v \cdot u_i) dt] \]  
\[ = \sin[\omega t + \beta \int (v \cdot u_i) dt] \]  
\[ = \sin[\omega t + \beta z \cdot u_i] \]  
\[ \text{(3)} \]

where, \( \omega \) is the base angular frequency of the oscillators which lies in theta frequency, \( z \) is the displacement of the animal from its initial position and \( \beta \) is the scaling factor.

This was the model of path integration used in (Soman et al 2018a, 2018b). Note that the PI cell described in eqn. (3) is a spatio-temporal model dependent explicitly on space, \( z \), and on time, \( t \). Such a model was successful in explaining certain temporal phenomena like phase precession (Soman et al 2018a). However, most hippocampal spatial cells are described in purely spatial terms (e.g., place cells and grid cells), depicting their responses as exclusive functions of space. In the present study, we are interested in describing these purely spatial responses. Therefore, we convert the spatio-temporal model of eqn. (3) into a purely spatial model by an averaging process described below.

In order to eliminate temporal variation, a \( \sin(\omega t) \) term is multiplied with the \( PI_{fm} \) shown in eqn. 3, and passed through a low pass filter which blocks the high frequency signals as shown in eqn 4 and eqn 5.

\[ PI_i = \sin(\omega t) \sin[\omega t + \beta z \cdot u_i] \]  
\[ \text{(4)} \]

\[ PI_i = \cos(\beta z \cdot u_i) - \cos(2\omega t + \beta z \cdot u_i) \]

After eliminating the high frequency term, we have,

\[ PI_i = \cos(\beta z \cdot u_i) \]  
\[ \text{(5)} \]

This formulation of Path Integration can also be extended to a complex version, shown in eqn. 6:

\[ PI_i = e^{j(\beta z \cdot u_i)} \]  
\[ \text{(6)} \]

where \( j \) is the complex number \( \sqrt{-1} \).

For implementation purposes, an alternative of this form of the above is an ordered pair of \( \cos(\beta z \cdot u_i) \) and \( \sin(\beta z \cdot u_i) \) which is given as a concatenation of cosine and sine terms. The eigenvectors and firing fields obtained in this concatenated and complex forms are identical making this formulation valid (see Supplementary material 3).

Solstad et al \(^{15}\) suggest that the place cells are generated as a result of a linear summation of grid patterns of different spatial scales determined by \( \beta \) parameter. To implement this, two other cases of PI are considered to account for information coming from different scales of grid cells. For such cases, multiple \( \beta \)s are considered. These formulations are used in the autoencoder model. For the current study, we use a normal distribution \(^{15}\) of \( \beta \) with mean = 9, standard deviation (s.d) = 2 and use 7 samples from the distribution for the simulation. Two different variations of PI with multiple \( \beta \)s are shown in eqn 7 and eqn 8:
\[ PL_i = \cos(\beta_k z. u_i) \]  
\[ PL_i = e^{j\beta z. u_i} \]  

Table 1 summarizes all the four variations of path integration considered for the simulations.

**Table 1: Four variations of PI considered for the simulations.**

| S.No. | PI Theoretical expression | PI Implementation | 
|-------|--------------------------|-------------------| 
| 1     | \( \cos(\beta z. u_i) \) | \( \cos(\beta z. u_i) \) | Type I 
| 2     | \( e^{j\beta z. u_i} \) | Concatenation (\( \cos(\beta z. u_i) \), \( \sin(\beta z. u_i) \)) | Type II 
| 3     | \( \cos(\beta_k z. u_i) \) | For all \( k \) {Concatenation (\( \cos(\beta_k z. u_i) \))} | Type III 
| 4     | \( e^{j\beta_k z. u_i} \) | For all \( k \) {Concatenation (Concatenation (\( \cos(\beta_k z. u_i) \), \( \sin(\beta_k z. u_i) \)))} | Type IV 

The above formulations of PI will be referred to as Type I, Type II, Type III and Type IV respectively from here on. We now take the PI vectors described by table 1 and generate representations using two unsupervised learning approaches: 1) PCA, and 2) Autoencoder network.

**Principal Component Analysis (PCA) Model:** The input for this step is the path integration data in two formats, Type I and Type II as described in the previous section (Table 1). PCA is performed on this input data by calculating the covariance matrix (Fig: 1b). The eigenvectors of the covariance matrix give us the principal components which form a basis for the encoded representation of the spatial information. We will now show that the responses obtained by the PCA process result grid cell-like responses with spatial periodicity. It can be shown that the covariance matrix of the path integration data is a circulant matrix (see supplementary material: section 1). It is a well-known fact that the eigenvectors of a circulant matrix are sinusoidal. Hence, we get sinusoidal eigenvectors from the PCA steps. We can show that these eigenvectors will lead to the final output resembling the modes of vibration of a circular membrane written in terms of the Bessel functions of first kind (See Supplementary material: section 2). The comparison between the output calculated from this derivation and the output from the model can be seen in the results section.

**Auto Encoder model:** The autoencoder network (Fig: 1c) is built in stages using the autoencoder stack approach (Aurelio et al., 2007)(Bengio et al., 2007). The layer 1 in the autoencoder is fed with the input from one of the variants of PI layer as discussed above (Table 1). The output of layer 1 is given as input to the second layer of the autoencoder stack. Layer 2 is also considered for spatial cells analysis. Both layer 1 and layer 2 contain 50 neurons each. Hyperparameters in the network are adjusted in order to accommodate some physical conditions while replicating the experimental results.
**Simulation of Experimental Studies:**

**Disto code Model:** An experimental study\(^{21}\) reports *disto code* in the CA1 region of the hippocampus. The disto code behavior is observed when a neuron fires at a relatively same distance between starting point to the end point on various edges of a triangular track. The experiments were performed on trajectories with random initial and end points as well as on systematic pillar tasks. The overlapping of normalized random trajectories shows that the neuron fires at approximately the same distance. Same kind of results are observed with systematic triangular trajectory. The animal is made to forage on a triangular trajectory in a systematic manner i.e., clockwise or anticlockwise by giving visual cues at the vertices of an equilateral triangle. As the animal reaches the vertex the visual cue disappears and appears at the next vertex. The neurons which fired at relatively the same distance from all the vertices of the triangle were categorized as disto codes, whereas the neurons which have a single firing field in any one arm of the triangular trajectory are the place cells.

To reproduce the results of disto codes by \(^{21}\), the autoencoder model is fed with input from Type II PI layer and the visual vector of 100 dimensions \( (V_i) \) (Supplementary material: section 5, Fig 5a). The details of construction of triangular trajectory and visual vector is explained in the (Supplementary material: section 4). The input vector \( X \) to the autoencoder model for this case is \( X = [PI(\text{Type II}) \ V_i] \) (Fig: 1d). The results are described in results section.
Multi compartment model: The second experimental study simulated is by \(^{25}\). In this experiment the rat is allowed to randomly forage in four compartments environment and the activity of place cells in the CA1 region of hippocampus is recorded. Initially, all four compartments have the same visual cues and lighting conditions. In such conditions the place cell fires at a relatively same location in each compartment. As the lighting condition of one of the compartments is changed, the place cell firing location inside that compartment dislocates or disappears, thereby proving that there is local remapping of place cells activity.

To reproduce “the local remapping of place cell activity” results by \(^{25}\), a multi-compartmental environment is created and the animal is allowed to randomly forage in all the four compartments. The autoencoder model is fed with the input from Type II PI layer. Lighting condition (Lc) is represented in binary format i.e \([1\ 0]\) for bright room and \([0\ 1]\) for dark room, and Head direction (HDi) cell response of the animal is given by \(\cos(\theta - \theta_i)\) where, \(\theta\) is current heading direction and \(\theta_i\) is the preferred direction of the HD neurons (Supplementary material: section 5, Fig 5b). The heading direction information is given in order to provide boundary conditions for the animal since the animal, when it is close to a wall can either move along the wall or in normal direction away from the wall; movement into the wall is obviously forbidden. There are 100 neurons in the Heading Direction layer which span 360°. The input vector X to the autoencoder model for this case is \(X = [PI(\text{Type II})\ \ Lc\ \ HDi]\) (Fig: 1d). The results are described in the results section.

For the different simulation studies described so far, the choice of \(\beta\) is crucial to generate the corresponding results. In the oscillatory interference (OI) model, \(^{10}\) have characterized the wave length of the interference pattern, which represents the spatial scaling of the grid cells as a function of \(\beta\), as shown below.

\[
L = \frac{2\pi}{\beta}
\]  

(9)

where \(L\) is the wave length of the interference pattern, and is equal to the distance between two adjacent peaks in the grid cell firing field. Therefore, the choice of \(\beta\) can be made based on \(L\) (or the number of square or hexagonal fields in the environment) using the following equation.

\[
\beta = \frac{2\pi}{L}
\]  

(10)

For obtaining grid cells in a 2 by 2 square environment, the value \(L = \frac{2}{3}\) will give one complete grid field. However, the experimental studies are focused on the place cell responses and hence, we choose \(L = 1\) such that there is one place field in the environment.

Analysis of Results:

For each of the studies performed, various methods have been used to analyze the results. We characterize the single neuron responses in the model in terms of Firing rate maps, Autocorrelation \(^{26}\) and grid scores \(^{20}\). To
characterize place cells. Connected component analysis has been used on the firing rate maps. The detailed procedure for each of the methods can be found in Supplementary Material Section 6.

**Results**

**Grid Cells from PCA model:**

The PCA model described previously is simulated with two kinds of path integration inputs, the Type I and Type II forms of PI. The final outputs from this network depict different trends in spatial behavior depending on the nature of path integration. The spatial cell responses are suitably thresholded and plotted as a red dot against the trajectory of the animal shown in blue (see Fig. 2). This analysis maps the firing data to the location of its firing hence revealing the spatial behavior. Fig.2 (a) depicts the firing fields when the network is simulated with Type I form of PI. The second column depicts the firing rate maps and the third column depicts the autocorrelation maps for the corresponding firing field (see methods section for rate maps and autocorrelation maps). These behaviors could be successfully understood theoretically using the derivations in the methods section. This has been validated by the results shown in Fig. 2 (b) where the first and second column are the raw outputs from the simulation. They can be compared to the third column which shows the output function from the derivation (Supplementary material: section 2). The following equation shows the derived theoretical output for PI Type I:

\[
f_k(x, y) = J_k(\beta r)\cos(-k\alpha) + (-1)^k J_k(\beta r)\cos(k\alpha)
\]

where, \(r\) and \(\alpha\) correspond to the polar notation of \((x, y)\). \(J_k\) is the \(k^{th}\) order Bessel function of first kind and \(f_k(x, y)\) is firing field of neuron corresponding to \(k^{th}\) principal component.

This comparison confirms the theoretical understanding and provides insights into the mathematical basis of grid cell activity.
Fig 2: PCA model results with PI Type I. (a) Firing fields of the first few neurons with PI Type I as input to the PCA model. The columns show firing fields, firing rate maps and autocorrelation maps of the corresponding firing fields. (b) A comparison of the raw outputs from the model (columns 1 and 2) and the theoretical values of the derived output function (column 3).

Similar simulations are performed with the PI Type II and the firing fields from this simulation are plotted in Fig 3. Fig 3 (a) shows the firing fields, firing rate maps and autocorrelation maps of the output for this model. Fig 3 (b) depicts the similarity between the raw output of the simulations and the output function from the derivation for the Type II path integration case (See Supplementary material: section 2) evaluated in the environment. The following equation shows the derived theoretical output for PI Type II:

$$f_k(x, y) = (-1)^k J_k(\beta r) \cos(k\alpha)$$  

where, \(r\) and \(\alpha\) correspond to the polar notation of \((x, y)\), \(J_k\) is the \(k^{th}\) order Bessel function of first kind and \(f_k(x, y)\) is related to the firing fields as given below:

Firing field of neuron number 2k = real \((f_k(x, y))\) and firing field of neuron number 2k + 1 = imaginary \((f_k(x, y))\).
Fig 3: PCA model results with PI Type II. (a) Firing fields of the first few neurons with PI Type II as input to the PCA model. The columns show firing fields, firing rate maps and autocorrelation maps of the corresponding firing fields. (b) A comparison of the outputs from the model (columns 1, 2, 4 and 5) and the theoretical values of the derived output function (column 3, 6)

It is observed from Fig. 2 (b) that the firing fields show only even order grid fields while the outputs from Fig 3 (b) show both even and odd order grid fields. This once again confirms that the theoretical framework proposed holds true for both the cases and successfully describes the simulation results.

1. Spatial cells from the autoencoder model:
   The neurons in layer 1 and layer 2 of the autoencoder model are analyzed for the spatial responses using all four different cases of PI as discussed in the methods section (Table 1). The activity of the neurons (red dots) is plotted over the trajectory (blue) followed by the virtual animal (see Fig 4). It is observed from the firing responses that these activities are similar to those seen in hippocampal formation in terms of place cells and grid cells. In this section, we also present a comparative analysis of place cells and grid cells in layer 1 and layer 2 of the four different formulations of path integration.

   Grid cells response: The firing pattern of some neurons in layer 1 and layer 2 resembles the firing of grid cells observed in Entorhinal Cortex (EC) in all four variants of PI described in the method section (Table 1). Two types of grid patterns are observed in the model: hexagonal grid pattern and square grid pattern. The square grid patterns are previously reported in experimental studies as well as computational studies\textsuperscript{16,28}. In the model with PI Type I as the input, 56% of neurons are hexagonal grid cells and 18% of neurons are square grid cells out of
50 neurons in layer 1 (fig 4: 1.a, 1.b and 1.c). Layer 2 contains 30% of hexagonal grid cells and 14% of square grid cells out of 50 neurons (fig 4: 1.d, 1.e and 1.f). Similarly, when the input is PI Type III, the Layer 1 contains 8% of hexagonal grid cells and 34% of square grid cells (Supplementary fig 7a). Layer 2 contains 8% hexagonal grid cells and 44% square grid cells (Supplementary fig 7b).

In the model, with input as PI Type II to the autoencoders, 46% of neurons are hexagonal grid cells and 9% of neurons are square grid cells in Layer 1 (fig 4: 2.a, 2.b and 2.c). In Layer 2, 40% of neurons are hexagonal grid cells and 14% of neurons are square grid cells (fig 4: 2.d, 2.e and 2.f). Similarly, with PI Type IV, Layer 1 contains 28% of hexagonal grid cells and 24% of square grid cells (Supplementary fig 7c). Layer 2 contains 22% of hexagonal grid cells and 24% of square grid cells (Supplementary fig 7d).

Fig 4: Output of neurons from layer 1 and layer 2 with PI Type I and PI Type II as input to the autoencoder. (1.a) Firing field, firing rate map and autocorrelation map of an example hexagonal grid neuron from layer 1 of the model. (1.b) Firing field, firing rate map and autocorrelation map of an example square grid neuron from layer 1 of the model. (1.c) Classification of neurons into Hexagonal grid cells, Square grid cells and non-grid cells in layer 1 of the model. (1.d) Similar to (1.a), (1.e) Similar to (1.b) and (1.f) Similar to (1.c) in layer 2 of the model. (2.a) Firing field, firing rate map and autocorrelation map of an example hexagonal grid neuron from layer 1 of the model. (2.b) Similar plot for square grid neurons in layer 1 of the model. (2.c) Classification of neurons into Hexagonal grid cells, Square grid cells and non-grid cells in layer 1 of the model. (2.d) Similar to (2.a), (2.e) similar to (2.b) and (2.f) similar to (2.c) for layer 2 of the model.
**Place cell responses:** A neuron which fires at a confined region in the environment is called a place cell. With PI Type I and PI Type III as input to the autoencoder model, place cell emergence is not observed. The emergence of place cells is observed when input to the autoencoder is PI Type II and PI Type IV. In PI type II case, layer 1 of the model consists of 17 place cells and layer 2 consists of 18 place cells whereas when input to the autoencoder model is PI Type IV, the number of place cells in layer 1 are 32 and place cells in layer 2 are 19 (Table 2).

**Stability of place cells with different trajectories:** An important way to characterize a neural response as a place cell is by checking for the stability of the firing field, i.e., the firing field must be in the same area of the environment irrespective of the trajectory the animal takes. The place cells from the autoencoder model have been analyzed for stability by testing against multiple trajectories and are found to be stable. The models with PI Type II and PI Type IV have been tested as shown in the Fig 5.

Fig 5: Place cell activity and stability in autoencoder model. (a) Layer 1 place cell activity for autoencoders trained with PI Type II. Left and right columns of (a) are the response of the same place cell on different trajectories. (b) Place cell response for a layer 2 neuron plotted similar to (a). (c) Layer 1 place cell activity for autoencoders trained with PI Type IV. Left and right columns of (c) are the response of the same place cell on different trajectories. (d) Place cell response for a layer 2 neuron plotted similar to (c).
Table 2 summarizes the number of place cells and number of grid cells in layer 1 and layer 2 in all four cases of PI considered.

**Table 2: Summary of Place cells and grid cells for different combinations in PI in both the layers of model.**

| PI formulation | Number of place cells | Number of Hexagonal Grid cells | Number of Square Grid cells | Ratio of place cells to Hexagonal Grid cells |
|----------------|-----------------------|-------------------------------|----------------------------|---------------------------------------------|
| Layer 1        |                       |                               |                            |                                             |
| PI Type I      | 0                     | 28                            | 9                          | 0                                           |
| PI Type II     | 17                    | 38                            | 3                          | 0.44                                        |
| PI Type III    | 0                     | 4                             | 17                         | 0                                           |
| PI Type IV     | 32                    | 14                            | 12                         | 2.28                                        |
| Layer 2        |                       |                               |                            |                                             |
| PI Type I      | 0                     | 15                            | 7                          | 0                                           |
| PI Type II     | 18                    | 20                            | 7                          | 0.9                                         |
| PI Type III    | 0                     | 4                             | 22                         | 0                                           |
| PI Type IV     | 19                    | 11                            | 12                         | 1.72                                        |

**Simulation of experimental studies:** Using the autoencoder model proposed, two experimental studies are simulated and the results are validated against the experimental findings.

**Disto code Model:** To model this experiment, a triangular trajectory is created in clockwise direction and visual input is provided by a visual vector of size 100 along with PI Type II (see Methods section). The magnitude of this vector increases as the animal approaches the vertex of a triangle (see supplementary material: section IV). Both Disto codes and place cells are observed in layer 1 and layer 2 of the model (fig: 6).
Multi compartment Model: The second experimental study simulated is by Spiers et.al \textsuperscript{25}. In this experiment the rat is allowed to randomly forage in the four compartments of the environment and the activity of place cells in the CA1 region of hippocampus is recorded. Initially, all four compartments have the same visual cues and lighting conditions. In such conditions the place cell fires at relatively same location in each compartment. As the lighting condition of one of the compartments is changed, the firing location of that compartment dislocates or disappears, suggestive of local remapping of place cells activity.

The input vector to the autoencoder model is described in the ‘Methods’ section. The model is trained with all compartments lit and the firing of neurons is observed in each compartment from layer 2 of the model. In fig 6, the left column shows two example neurons and their firing responses when the model is trained with all compartments lit. It can be observed that the firing location is relatively the same in each compartment. The model is tested after making the third compartment dark. It can be observed in the right column of fig 6, that the firing in only the third compartment is distorted and is intact in rest of the compartments hence verifying the local remapping of the place cells.

Fig 6: Results from the simulation of experimental study by Aghajan et.al\textsuperscript{21} and Spiers et.al \textsuperscript{25}. (a) and (c) Disto code like behavior in Layer 1 and Layer 2 respectively. (b) and (d) place cell like behavior in Layer 1 and Layer 2 respectively. (e1) Firing response of a neuron showing place cell fire at same location in all 4 compartments when all the compartments are lit. (e2) Firing fields of same neuron showing the remapping of firing field in compartment 3 alone, when lighting condition of that compartment is changed i.e., made dark. (f1) and (f2) Place cell response of different neuron similar to (e1) and (e2).

Discussion

There are several models of hippocampal spatial cells in a network structure and yet very little can be explained about how these representations are generated and what they contribute to the larger challenge of spatial
navigation. Most of the existing grid cell models can be broadly classified into two categories namely: Continuous attractor neural networks (CANN) and Oscillatory Interference (OI) models. The CANN has a 2D sheet of neurons with special symmetrical lateral connections to the nearby neurons whose weights are inversely proportional to the distance. Also, in order to manage the problem of distorted connections at the boundary of a 2D sheet, the model makes an assumption of connections between last and first boundary neurons to make a toroid kind of structure. This toroid pushes the network to produce periodic responses. However, the DNN model proposed here produces the periodic response without setting an in-built cyclic connection between the neurons by encoding the path integration data. The other category of model which is OI model, assumes interference of three inputs from dendrites with the soma oscillating at theta frequency. The assumption made here is the pre-tuned directionality of these inputs which differ by 60° making it again unrealistic.

In an effort to answer the questions of spatial cell modeling while avoiding unrealistic assumptions about head direction selectivity, the PCA and Autoencoder model have been proposed. The PCA approach is taken up mainly to establish a mathematical framework for the encoding of spatial information. This helps us understand the underlying principles and subsequently apply these principles for spatial reconstruction and versatile navigation as observed in animals. However, the autoencoder model, being a multi-layer architecture, it is more nonlinear and is capable of modelling spatial cells in a hierarchical fashion. In addition to modelling spatial cells like grid cells, place cells, in the present study we also use it to model behavioural studies like behaviour of place cells in multicompartment and emergence of disto codes in a triangular trajectory which include multisensory integration.

All these results follow from the original ansatz where PI is implemented by integrating the velocity into the phase of the theta oscillations given by eqn 2. This ansatz can be justified by invoking biological evidence that theta frequency varies with velocity nearly linearly. For example, in 1D navigation, the equation is $\sin(\omega t + \beta v \cdot t) = \sin[(\omega + \beta v) t]$. Therefore, the model is consistent with experimental data. In other words, velocity is used to frequency modulate theta oscillations. The eqn. 2 after solving becomes $\sin[(\omega t + \beta(ax + by))]$. Therefore, the oscillatory terms are dependent both on position and time. The spatial cell responses are often depicted as exclusive functions of positions by performing averaging that eliminates the time dependence. By performing the averaging step described in eqns. (2,3), we are able to substitute $\sin[(\omega t + \beta(ax + by))]$ by $\sin[(\beta(ax + by))]$. This last generated representation of PI becomes the starting point of the hierarchy of spatial representation that can be used to study emergence of various kinds of spatial cells. This hierarchical framework can also be adapted to various experimental paradigms involving spatial navigation.

To understand which form of PI gives more realistic spatial cells in the proposed models, we considered two different variations of PI (Type I and Type II) in the PCA model and four different variations of PI (Type I, Type II, Type III, Type IV) in the autoencoder model. Grid cell-like responses that we see with PCA are modelled by Bessel functions (eqn 14 and eqn 15). The PCA on PI Type I shows output with only even order grid fields.
and PCA on PI Type II shows output with both even and odd order grid fields which are in accordance with the theoretical framework (Supplementary material: section 2). These output functions obtained are also seen in modes of vibration of a circular membrane. It is a known fact that any type of vibration can be decomposed into an infinite series of the vibrational modes. This process is like how a time signal can be decomposed into a Fourier series. The model being able to produce this kind of a basis in the encoded representation points to the existence of a possible basis for storing spatial information in the spatial frequency domain of the brain.

In autoencoder model, out of all the four cases, a greater number of grid cells appears when input to the model is PI Type I and PI Type II. Since there is no phase offset in PI Type I and PI Type III, we do not observe any place cell in this case. The phase offset in the input is because of the complex terms in PI Type II and PI Type IV and we observe a greater number of place cells in this case which is in accordance with other studies like 10. There are existing models which suggest emergence of place cells from linear summation of variably scaled grid cells 15. The same approach is taken when when PI Type IV is presented as input to the autoencoder model, in which case we see the emergence of more place cells (refer Table 2).

The proposed deep model is not only capable of modelling spatial cells in a simple environment but can also model the experimental paradigms involving more complex environmental conditions. In this study, we have modelled two experimental paradigms which in our knowledge are not modelled before. The model is able to reproduce the experimental results efficiently. The experiments utilize different sensory inputs as cues to the animal while the model is able to generate the same results with much simplified cues which substitute for vision, lighting conditions, etc.

The multilayer autoencoder model is capable of modelling more complex non-linear representations, which suggest future direction of combining audition, olfaction and more realistic vision along with PI to study spatial navigation as a whole. Since this approach of modelling cognitive space in the hippocampus provides a robust framework, it supports multiple modelling studies like modelling object responsive neurons found in lateral entorhinal cortex (LEC) 31, object vector cells 32 etc. Muller et.al 33 have shown that place cells exhibit direction sensitivity which varies with environments and is more prominent in polarized environments. This behaviour can also be further explored using the DNN model. The DNN approach can be used for modelling theta sequences (which involve rapid sequential firing of place cells within a few theta cycles) and Sharp Wave Ripples (SWR) which are high frequency synchronous bursts of neurons when the animal is immobile 34. The continuous integration of speed in the PI will lead to an accumulation of error. As a result, the position estimates are not accurate. This issue is addressed by resetting the integration limits in PI using visual input 35. The current model does not account for phase resetting in PI which is an important phenomenon to correct accumulated error during navigation and needs to be addressed in the future. Furthermore, both the DNN and PCA models can be extended for studying 3D navigation to explore the extensions of the spatial cell behaviours in 3D environments and also
observe the emergence of new spatial cell patterns like plane cells and FCC lattice type firing fields. Moreover, we can also investigate a suitable basis in 3D using the PCA model similar to the 2D case.

**Data accessibility**

The simulation code is made available at the ModelDB database (Model URL: [http://modeldb.yale.edu/266930](http://modeldb.yale.edu/266930)).

Access code: cnslab).

**References**

1. R. M. Cichy, A. Khosla, D. Pantazis, A. Torralba, A. Oliva, Comparison of deep neural networks to spatio-temporal cortical dynamics of human visual object recognition reveals hierarchical correspondence. *Sci. Rep.* **6**, 1–13 (2016).

2. R. Geirhos, *et al.*, “Comparing deep neural networks against humans: object recognition when the signal gets weaker *” (2018) (January 6, 2021).

3. S. G. Wardle, C. Baker, Recent advances in understanding object recognition in the human brain: Deep neural networks, temporal dynamics, and context. *F1000Research* **9** (2020).

4. H. A. Atabay, BINARY SHAPE CLASSIFICATION USING CONVOLUTIONAL NEURAL NETWORKS (2017).

5. J. Wang, *et al.*, “Learning Fine-grained Image Similarity with Deep Ranking” (2014) (February 3, 2021).

6. G. E. Dahl, D. Yu, L. Deng, A. Acero, Context-dependent pre-trained deep neural networks for large-vocabulary speech recognition. *IEEE Trans. Audio, Speech Lang. Process.* **20**, 30–42 (2012).

7. B. Y. Jang, W. H. Heo, J. H. Kim, O. W. Kwon, Music detection from broadcast contents using convolutional neural networks with a Mel-scale kernel. *Eurasip J. Audio, Speech, Music Process.* **2019**, 11 (2019).

8. B. Lau, G. B. Stanley, Y. Dan, Computational subunits of visual cortical neurons revealed by artificial neural networks. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 8974–8979 (2002).

9. A. J. E. Kell, D. L. K. Yamins, E. N. Shook, S. V. Norman-Haignere, J. H. McDermott, A Task-Optimized Neural Network Replicates Human Auditory Behavior, Predicts Brain Responses, and Reveals a Cortical Processing Hierarchy. *Neuron* **98**, 630–644.e16 (2018).

10. N. Burgess, C. Barry, J. O’Keefe, An oscillatory interference model of grid cell firing. *Hippocampus* **17**, 801–812 (2007).
11. M. C. Fuhs, D. S. Touretzky, A spin glass model of path integration in rat medial entorhinal cortex. *J. Neurosci.* **26** (2006).

12. B. L. McNaughton, F. P. Battaglia, O. Jensen, E. I. Moser, M. B. Moser, Path integration and the neural basis of the “cognitive map.” *Nat. Rev. Neurosci.* **7** (2006).

13. A. Guanella, P. F. M. J. Verschure, Prediction of the position of an animal based on populations of grid and place cells: A comparative simulation study. *J. Integr. Neurosci.* **6** (2007).

14. Y. Burak, I. R. Fiete, Accurate path integration in continuous attractor network models of grid cells. *PLoS Comput. Biol.* **5**, 1000291 (2009).

15. T. Solstad, E. I. Moser, G. T. Einevoll, From grid cells to place cells: A mathematical model. *Hippocampus* **16**, 1026–1031 (2006).

16. Y. Dordek, D. Soudry, R. Meir, D. Derdikman, Extracting grid cell characteristics from place cell inputs using non-negative principal component analysis. *Elife* **5** (2016).

17. C. J. Cueva, X.-X. Wei, Emergence of grid-like representations by training recurrent neural networks to perform spatial localization. *arXiv* (2018).

18. I. Kanitscheider, I. Fiete, Training recurrent networks to generate hypotheses about how the brain solves hard navigation problems. *Adv. Neural Inf. Process. Syst.* **2017-December**, 4530–4539 (2016).

19. K. Soman, S. Chakravarthy, M. M. Yartsev, A hierarchical anti-Hebbian network model for the formation of spatial cells in three-dimensional space. *Nat. Commun.* **9**, 1–15 (2018).

20. K. Soman, V. Muralidharan, V. S. Chakravarthy, A unified hierarchical oscillatory network model of head direction cells, spatially periodic cells, and place cells. *Eur. J. Neurosci.* **47**, 1266–1281 (2018).

21. Z. M. Aghajan, *et al.*, Impaired spatial selectivity and intact phase precession in two-dimensional virtual reality. *Nat. Neurosci.* **18**, 121–128 (2015).

22. H. J. Spiers, R. M. A. Hayman, A. Jovalekic, E. Marozzi, K. J. Jeffery, Place Field Repetition and Purely Local Remapping in a Multicompartment Environment. *Cereb. Cortex* **25**, 10–25 (2015).

23. M. ’Aurelio, R. C. Poulney, S. Chopra, Y. Lecun, “Efficient Learning of Sparse Representations with an Energy-Based Model.”

24. Y. Bengio, P. Lamblin, D. Popovici, H. Larochelle, “Greedy Layer-Wise Training of Deep Networks.”

25. H. J. Spiers, R. M. A. Hayman, A. Jovalekic, E. Marozzi, K. J. Jeffery, Place field repetition and purely local remapping in a multicompartment environment. *Cereb. Cortex* **25**, 10–25 (2015).
26. T. Hafting, M. Fyhn, S. Molden, M.-B. Moser, E. I. Moser, Microstructure of a spatial map in the entorhinal cortex. *Nature* **436**, 801–806 (2005).

27. H. Samet, M. K. Tamminen, Efficient Component Labeling of Images of Arbitrary Dimension Represented by Linear Bintrees. *IEEE Trans. Pattern Anal. Mach. Intell.* **10**, 579–586 (1988).

28. J. Krupic, M. Bauza, S. Burton, C. Barry, J. O’Keefe, Grid cell symmetry is shaped by environmental geometry. *Nature* **518**, 232–235 (2015).

29. A. Samsonovich, B. L. McNaughton, Path integration and cognitive mapping in a continuous attractor neural network model. *J. Neurosci.* **17**, 5900–5920 (1997).

30. C. H. Vanderwolf, Hippocampal electrical activity and voluntary movement in the rat. *Electroencephalogr. Clin. Neurophysiol.* **26**, 407–418 (1969).

31. S. S. Deshmukh, J. J. Knierim, Representation of non-spatial and spatial information in the lateral entorhinal cortex. *Front. Behav. Neurosci.* (2011) https://doi.org/10.3389/fnbeh.2011.00069.

32. Ø. A. Høydal, E. R. Skytøen, S. O. Andersson, M.-B. Moser, E. I. Moser, Object-vector coding in the medial entorhinal cortex. *Nature* **568**, 400–404 (2019).

33. R. U. Muller, E. Bostock, J. S. Taube, J. L. Kubie2, “On the Directional Firing Properties of Hippocampal Place Cells” (1994).

34. C. Drieu, M. Zugaro, Hippocampal sequences during exploration: Mechanisms and functions. *Front. Cell. Neurosci.* **13**, 232 (2019).

35. T. Li, A. Arleo, D. Sheynikhovich, Modeling place cells and grid cells in multi-compartment environments: Hippocampal-entorhinal loop as a multisensory integration circuit. *bioRxiv*, 602235 (2019).

**Acknowledgements**

We acknowledge the suggestions of Dr. Ayan Mukhopadhyay regarding the calculation of Bessel function integrals.

**Author Contributions**

A Aziz and P Sree Harsha contributed in designing, coding, analysis of the model and preparation of the manuscript. Rohan N contributed in coding. S Chakravarthy contributed in designing the model and preparation of the manuscript.

**Competing Interests**

The authors declare that no competing interests exists.
Figures

Outline of the model architecture. (a) Common architecture up to PI showing averaging process to remove time dependency. (b) PCA on final PI after averaging process. (c) Autoencoder model on final PI after averaging process. (d) Concatenating other sensory inputs along with final PI to model experimental paradigms.

| Firing Fields | Rate maps | Autocorrelation | Model Outputs | Theoretical Outputs |
|---------------|-----------|-----------------|---------------|---------------------|
| neuron 2      |           |                 | neuron 2      | \( J_2 \cos(2\phi) \) |
| neuron 4      |           |                 | neuron 3      | \( J_4 \cos(4\phi) \) |
| neuron 6      |           |                 | neuron 4      | \( J_6 \cos(6\phi) \) |
Figure 2

PCA model results with PI Type I. (a) Firing fields of the first few neurons with PI Type I as input to the PCA model. The columns show firing fields, firing rate maps and autocorrelation maps of the corresponding firing fields. (b) A comparison of the raw outputs from the model (columns 1 and 2) and the theoretical values of the derived output function (column 3).

Figure 3
PCA model results with PI Type II. (a) Firing fields of the first few neurons with PI Type II as input to the PCA model. The columns show firing fields, firing rate maps and autocorrelation maps of the corresponding firing fields. (b) A comparison of the outputs from the model (columns 1, 2, 4 and 5) and the theoretical values of the derived output function (column 3,6)

Figure 4

Output of neurons from layer 1 and layer 2 with PI Type I and PI Type II as input to the autoencoder. (1.a) Firing field, firing rate map and autocorrelation map of an example hexagonal grid neuron from layer 1 of the model. (1.b) Firing field, firing rate map and autocorrelation map of an example square grid neuron from layer 1 of the model. (1.c) Classification of neurons into Hexagonal grid cells, Square grid cells and non-grid cells in layer 1 of the model. (1.d) Similar to (1.a), (1.e) Similar to (1.b) and (1.f) Similar to (1.c) in layer 2 of the model. (2.a) Firing field, firing rate map and autocorrelation map of an example hexagonal grid neuron from layer 1 of the model. (2.b) Similar plot for square grid neurons in layer 1 of the model. (2.c) Classification of neurons into Hexagonal grid cells, Square grid cells and non-grid cells in layer 1 of the model. (2.d) Similar to (2.a), (2.e) similar to (2.b) and (2.f) similar to (2.c) for layer 2 of the model.
Figure 5

Place cell activity and stability in autoencoder model. (a) Layer 1 place cell activity for autoencoders trained with PI Type II. Left and right columns of (a) are the response of the same place cell on different trajectories. (b) Place cell response for a layer 2 neuron plotted similar to (a). (c) Layer 1 place cell activity for autoencoders trained with PI Type IV. Left and right columns of (c) are the response of the same place cell on different trajectories. (d) Place cell response for a layer 2 neuron plotted similar to (c).
Figure 6

Results from the simulation of experimental study by Aghajan et al. and Spiers et al. (a) and (c) Disto code like behavior in Layer 1 and Layer 2 respectively. (b) and (d) place cell like behavior in Layer 1 and Layer 2 respectively. (e1) Firing response of a neuron showing place cell fire at same location in all 4 compartments when all the compartments are lit. (e2) Firing fields of same neuron showing the remapping of firing field in compartment 3 alone, when lighting condition of that compartment is changed i.e., made dark. (f1) and (f2) Place cell response of different neuron similar to (e1) and (e2).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryMaterial.pdf