Learning an efficient place cell map from grid cells using non-negative sparse coding

Yanbo Lian[1,*], Anthony N. Burkitt[1,*]

[1] Department of Biomedical Engineering, The University of Melbourne, Melbourne, VIC, Australia

* {yanbo.lian, aburkitt}@unimelb.edu.au

Abstract

Experimental studies of grid cells in the Medial Entorhinal Cortex (MEC) have shown that they are selective to an array of spatial locations in the environment that form a hexagonal grid. However, in a small environment, place cells in the hippocampus are only selective to a single-location of the environment while granule cells in the dentate gyrus of the hippocampus have multiple discrete firing locations, but lack spatial periodicity. Given the anatomical connection from MEC to the hippocampus, previous feedforward models of grid-to-place have been proposed. Here, we propose a unified learning model that can describe the spatial tuning properties of both hippocampal place cells and dentate gyrus granule cells based on non-negative sparse coding. Sparse coding plays an important role in many cortical areas and is proposed here to have a key role in the navigational system of the brain in the hippocampus. Our results show that the hexagonal patterns of grid cells with various orientations, grid spacings and phases are necessary for model cells to learn a single spatial field that efficiently tile the entire spatial environment. However, if there is a lack of diversity in any grid parameters or a lack of cells in the network, this will lead to the emergence of place cells that have multiple firing locations. More surprisingly, the model shows that place cells can also emerge even when non-negative sparse coding is used with weakly-tuned MEC cells, instead of MEC grid cells, as the input to place cells. This work suggests that sparse coding may be one of the underlying organizing principles for the navigational system of the brain.

1 Introduction

The brain can perform extremely complex spatial navigation tasks, but how the brain does this remains unclear. Since the Nobel-prize-winning discovery of place cells in the hippocampus (Moser et al., 2008; O'Keefe, 1976; O'Keefe and Dostrovsky, 1971) and grid cells in the Medial Entorhinal Cortex (MEC) (Hafting et al., 2005; Rowland et al., 2016), brain regions involved in spatial awareness and navigation have attracted much attention from both experimental and computational neuroscientists.

Experimental rat studies show that hippocampal place cells have a single specific location in the environment at which they have an elevated firing rate (O'Keefe and Dostrovsky, 1971) and neighboring cells have firing fields at different locations of the environment, such that the local cell population in the hippocampus can represent the whole spatial environment (O'Keefe, 1976). In contrast, granule cells in the dentate

gyrus of the hippocampal formation have multiple discrete firing locations without spatial periodicity (Jung and McNaughton, 1993; Leutgeb et al., 2007). However, Park et al. (2011) also showed that rat place cells can have multiple firing locations in large environments.

MEC grid cells are also spatially tuned to the locations of the environment. However, unlike hippocampal place cells, firing fields of grid cells form a hexagonal grid that evenly tile the entire environment (Hafting et al., 2005). The hexagonal grid of each grid cell is characterised by spacing (distance between fields on the grid), orientation (the degree of rotation relative to an external reference), and phase (offset relative to an external reference). The spacing of the grid increases step-wise monotonically along the dorsal-ventral axis (Hafting et al., 2005). Moreover, the progression in grid spacing along the dorsal-ventral axis is geometric, with ratio around 1.42, such that grid cells are organised into discrete modules according to their spacing (Stensola et al., 2012). Additionally, grid cells in each module also have similar orientation but random phases (Stensola et al., 2012).

Experimental evidence indicates that MEC grid cells are the main projecting neurons to the dentate gyrus and CA3 of the hippocampus (Leutgeb et al., 2007; Steward and Scoville, 1976; Tamamaki and Nojyo, 1993; Zhang et al., 2013). Consequently a variety of models have been proposed to explain the emergence of the firing fields of hippocampal place cells based on the feedforward connection from MEC grid cells, from mathematical models that have no learning (de Almeida et al., 2009; Solstad et al., 2006) to models with plasticity (Franzius et al., 2007a,b; Rolls et al., 2006; Savelli and Knierim, 2010).

For the learning models of grid-to-place formation, Rolls et al. (2006) used a competitive learning procedure to learn place cells from grid cell input. However, only approximately 10% of model cells were found to have a single-location place field. Furthermore, the competition in the model was introduced by manually setting the population activation to a small specified value that indicates a sparse network. Similarly, Franzius et al. (2007b) applied independent component analysis (ICA) (Hyvarinen, 1999) to maximising the sparseness of the model place cells. However, the examples of model place cells in these studies are mostly located at the border of the environment (Figure 1G in Franzius et al. (2007b) and Figure 3C in Franzius et al. (2007a)). Additionally, in their model the connection strength between grid and place cells can be positive or negative and the place cell responses were manually shifted by the addition of a constant term to ensure that they were non-negative, which puts into question the biological realization of the model. Furthermore, previous models do not investigate how well the learned place map represents the spatial environment.

Sparse coding, proposed by Olshausen and Field (1996), provides a compelling explanation of many experimental findings of brain network structures. One particular variant of sparse coding, non-negative sparse coding (Hoyer, 2003), has recently been shown to account for a wide range of neuronal responses in areas including the retina, primary visual cortex, inferotemporal cortex, auditory cortex, olfactory cortex and retrosplenial cortex (see Beyeler et al. (2019) for a review). However, whether sparse coding can account for the formation of hippocampal place cells has not previously been investigated in detail.

Here we applied sparse coding with non-negative constraints, where neuronal responses and connection weights are restricted to be non-negative, to building a learning model of place cells using grid cell responses as the input. Our results show that single-location place fields can be learnt that tile the entire environment, given a sufficient diversity in grid spacings, orientations and phases of the input grid cells. However, if there is a lack of diversity in any of these grid parameters, the learning of the place cells is impeded; instead, the learning results in more place cells with multiple

firing locations. Furthermore, a lower number of grid cell inputs results in learning multiple place cell firing locations. The competition generated by the principle of sparse coding in the model naturally provides a global inhibition such that the place cells display discrete firing fields, suggesting that the proposed model can be implemented by biologically based neural mechanisms and circuits. Moreover, the model can still learn place cells even when the inputs to the place cells are replaced by the responses of weakly-tuned MEC cells. This suggests a plausible explanation of why place cells emerge earlier than grid cells during development (Langston et al., 2010; Wills et al., 2010).

2 Materials and Methods

2.1 Sparse coding with non-negative constraints

Sparse coding was originally proposed by Olshausen and Field (1996) to demonstrate that simple cells in the primary visual cortex represent their sensory input using an efficient neuronal representation, namely that their firing rates in response to natural images tend to be sparse (rarely attain large values) and statistically independent. In addition, sparse coding finds a reconstruction of the sensory input through a linear representation of features with minimal error, which can be understood as minimizing the following cost function

$$E(\mathbf{A}, \mathbf{s}) = \frac{1}{2}\|\mathbf{I} - \mathbf{A}\mathbf{s}\|_2^2 + \beta \sum_i Q(s_i), \tag{1}$$

where the matrix $\mathbf{I}$ is the input, columns of $\mathbf{A}$ are basis vectors (universal features) from which any input can be constructed from a weighted sum, the vector $\mathbf{s}$ represents the neural responses and each element, s_i, is the coefficient for the corresponding basis vector, the function $Q(\cdot)$ is a function that penalises high activity of model units, and β is a sparsity constant that scales the penalty function (Olshausen and Field, 1996, 1997). The term $\mathbf{A}\mathbf{s}$ in Eq. (1) represents the model reconstruction of the input, so this cost function represents the sum of squared reconstruction error and response penalty. Therefore, the model finds a sparse representation for the input by solving this minimization problem. By taking the partial derivatives of Eq. (1) in terms of the elements of $\mathbf{A}$ and $\mathbf{s}$, and then applying gradient descent, the dynamic equations and the learning rule are given by

$$\begin{aligned} \dot{\mathbf{s}} &= \mathbf{A}^T(\mathbf{I} - \mathbf{A}\mathbf{s}) - \beta Q'(\mathbf{s}) \\ \Delta\mathbf{A} &\propto \langle(\mathbf{I} - \mathbf{A}\mathbf{s})\mathbf{s}^T\rangle, \end{aligned} \tag{2}$$

where $\langle\cdot\rangle$ is the average operation, $Q'(\cdot)$ is the derivative of $Q(\cdot)$, and the dot notation represents differentiation with regard to time.

Non-negative sparse coding is simply sparse coding with non-negative constraints, i.e., the connection weights $\mathbf{A}$ and model responses $\mathbf{s}$ are restricted to non-negative values in the cost function Eq. (1). Note that, when β in Eq. (1) is set to zero, the cost function of non-negative sparse coding reduces to the cost function of non-negative matrix factorization (Lee and Seung, 1999).

2.2 The environment

The 2D spatial environment used in this study is a 1m×1m square box. A 32×32 grid with 1024 points is used to represent the entire environment. Therefore, a 1024×1 vector, denoted by $\mathbf{r}$, with only one non-zero element can be used to represent the location of a virtual rat.

2.3 Grid cell model

2.3.1 The mathematical model

The hexagonal firing fields of grid cells are represented in this study by the sum of three sinusoidal gratings (de Almeida et al., 2009; Kropff and Treves, 2008; Solstad et al., 2006), as described by

$$G(\vec{r}) = \frac{2}{3}\left(\frac{1}{3}\sum_{j=1}^{3}\cos\left(\frac{4\pi}{\sqrt{3}\lambda}\vec{u}_j\cdot(\vec{r}-\vec{r}_0)\right)+\frac{1}{2}\right), \tag{3}$$

where $G(\vec{r})$ is the grid cell response at the spatial location $\vec{r} = (x, y)$, λ is the grid spacing, θ is the grid orientation, $\vec{r}_0 = (x_0, y_0)$ represents the phase offset, and $\vec{u}_j = (\cos(2\pi j/3+\theta)), \sin(2\pi j/3+\theta))$ is the unit vector with direction $2\pi j/3+\theta$. $G(\cdot)$ described in Eq. (3) is normalised to have a maximum value of 1 and minimum of 0. Because of the periodicity of the hexagonal pattern, the grid orientation, θ, lies in the interval of $[0, \pi/3)$, and the phase offsets in both x and y axes are smaller than the grid spacing, i.e., $0 \le x_0, y_0 < \lambda$.

Since grid cells have different spacings, orientations and phases, Eq. (3) is used to generate diverse grid fields. The value of the grid spacing, λ, ranges in value from 28 cm (Hafting et al., 2005; Solstad et al., 2006) and increases by a geometric ratio 1.42 that is consistent with experimental results (Stensola et al., 2012), and the optimal grid scale is derived by a mathematical study (Wei et al., 2015). For example, if there are N_λ different grid spacings, the spacings will be 28 cm, $28 \times 1.42 = 39.76$ cm, $\cdots$, $28 \times 1.42^{N_\lambda - 1}$ cm. For each grid spacing, different values of grid orientation, θ, are uniformly taken from the interval $[0, 60°)$. For example, if there are 3 different grid orientations, the values will be 0, 20° and 40°. The number of different orientations for each grid spacing is denoted as N_θ. Furthermore, it is assumed here that there are N_x and N_y phase offsets along x-axis and y-axis for each specific grid spacing and orientation. Similar to grid orientation, the value of the phase is taken uniformly from $[0, \lambda)$. For example, if there are 2 phases along the x-axis, they will have the values $x_0 = 0$ and $\lambda/2$. The resulting total number grid cells, denoted as N_g, will be the product of numbers of spacings, orientations and phases:

$$N_\mathrm{g} = N_\lambda N_\theta N_x N_y \tag{4}$$

Some examples of grid fields described by Eq. (3) are shown in Figure 1. These grid fields have diverse grid spacings, orientations and phases.

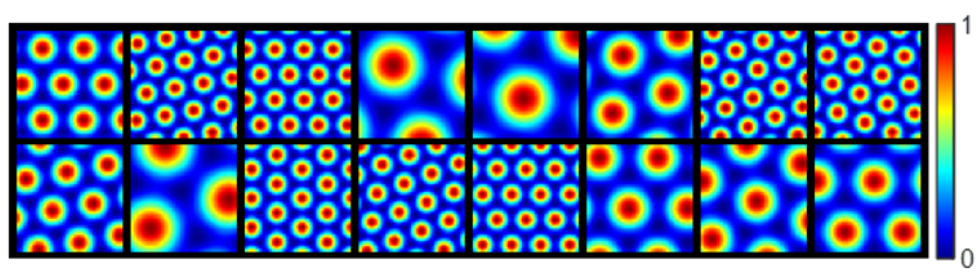


Figure 1. Examples of grid fields, Eq. (3), with different grid spacings, orientations and phases. Each block represents the hexagonal firing field of a grid cell in a 1 m×1 m environment. Values in each block are normalised to [0 1] in this plot.

Since the environment is represented by a 32×32 grid, a 1024×1 vector, denoted by $\mathbf{g}$, can be used to represent the firing field of the grid cell over the entire environment. For a given position $\mathbf{r}$ in the environment, the response of a grid cell is simply $\mathbf{g}^T\mathbf{r}$.

The number of different grid parameters (N_λ, N_θ, N_x and N_y) of grid cells defined above are assigned to different values to investigate the effect of the diversity of grid cells on the formation of place cells. Next we define grid cells with parameters that better capture the biologically observed variability, which will be used to investigate the robustness of the model.

2.3.2 The more realistic model

Similarly, grid cells are separated into discrete modules where grid cells are anatomically adjacent and have similar grid spacings, which is supported by experimental evidence (Stensola et al., 2012). However, values of the grid spacing in each module are randomly sampled from normal distributions: there are four discrete modules for grid spacings (λ) with mean of 38.8 cm, 48.4 cm, 65 cm and 98.4 cm and a same standard deviation of 8 cm. For grid orientation (θ), since grid cells in the same spacing module tend to have similar orientations (Stensola et al., 2012), grid cells in the four discrete modules also have mean orientations 15°, 30°, 45° and 0°, and a common standard deviation of 3°. Because grid phases, (x_0, y_0), are random in each module (Stensola et al., 2012), grid phase is randomly sampled from a uniform distribution. Stensola et al. (2012) showed that 87% of grid cells belong to the two modules with small spacings. Therefore, when using realistic grid fields in the study, we have 43.5%, 43.5%, 6.5% and 6.5% of grid cells in the modules with mean spacings 38.8 cm, 48.4 cm, 65 cm and 98.4 cm, respectively (unless otherwise noted).

The firing field of the grid cells is taken to be the sum of the spatial firing pattern at every vertex on the hexagonal pattern. The spatial firing pattern with vertex (x_v, y_v) is described by a function with the following form (Neher et al., 2017)

$$Q_v(x, y) = \gamma_v e^{-\ln(5)\frac{(x-x_v)^2+(y-y_v)^2}{\sigma^2}}, \tag{5}$$

where γ_v is the amplitude, σ determines the radius of the firing field, and the response will be $1/5\gamma_v$ at a distance σ away from the center. The amplitude at every vertex of the hexagonal pattern, γ_v, is chosen from a normal distribution with mean 1 and standard deviation 0.1, and σ is determined by the grid spacing, λ, with $\sigma = 0.32\lambda$ (Neher et al., 2017). The grid field is then the sum of firing fields (described by Eq. 5) at all vertices of the hexagonal pattern. The locations of vertices of the hexagonal pattern are determined by grid spacing, λ, grid orientation, θ, and grid phase, (x_0, y_0).

2.4 Structure of the model

In this study a two-layer network is proposed to model the activities of grid cells (first layer) and place cells (second layer), respectively. Given a spatial location in the environment, grid cells respond according to their firing fields. Grid cell responses then feed into place cells and the grid-place network implements a sparse coding model with non-negative constraints. The model structure is shown in Figure 2.

Denote $\mathbf{G}$ as a $1024 \times N_\mathrm{g}$ matrix that represents the firing fields for N_g grid cells in the network; i.e. each column of $\mathbf{G}$, $\mathbf{g}_i$ $(i = 1,\ 2, ..., N_\mathrm{g})$ is a 1024×1 vector that represents the firing field of grid cell i. For a spatial location $\mathbf{r}$ in the environment, grid cell responses (firing rates), $\mathbf{s}_\mathrm{g}$, are given by $\mathbf{s}_\mathrm{g} = \mathbf{G}^T\mathbf{r}$. Place cell responses (firing rates), $\mathbf{s}_\mathrm{p}$, are computed by a sparse coding model for the grid-place network with non-negative connection $\mathbf{A}$. Assume there are N_p place cells in the network. Then $\mathbf{A}$ is a $N_\mathrm{g} \times N_\mathrm{p}$ matrix and $\mathbf{s}_\mathrm{p}$ is a $N_\mathrm{p} \times 1$ vector. Denote $\mathbf{u}_\mathrm{p}$ as a $N_\mathrm{p} \times 1$ vector that

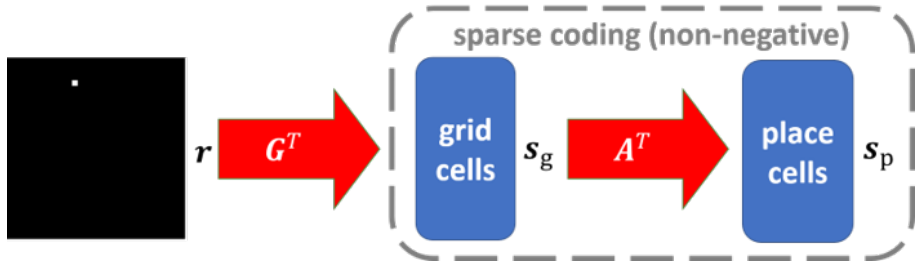


Figure 2. Graphical representation of the model. Red arrows represent non-negative connections. Notation is defined in the main text.

represents membrane potentials of place cells. The model dynamics is given by

$$\begin{aligned} \mathbf{s}_{\mathrm{g}} &= \mathbf{G}^T \mathbf{r} \\ \tau \dot{\mathbf{u}}_{\mathrm{p}} &= -\mathbf{u}_{\mathrm{p}} + \mathbf{A}^T \mathbf{s}_{\mathrm{g}} - \mathbf{W} \mathbf{s}_{\mathrm{p}} \\ \mathbf{s}_{\mathrm{p}} &= \max(\mathbf{u}_{\mathrm{p}} - \beta, 0), \end{aligned} \tag{6}$$

where τ is the time constant for place cells, β is the threshold of the rectifying function of firing rates, and $\mathbf{W}$ can be interpreted as the matrix of recurrent connections between place cells. In this paper, we take $\mathbf{W} = \mathbf{A}^T\mathbf{A} - \mathbb{1}$, where $\mathbb{1}$ is a $N_{\mathrm{p}} \times N_{\mathrm{p}}$ identity matrix. The dynamics of place cells described in Eq. (6) is derived from the local competitive algorithm (LCA) proposed by Rozell et al. (2008) that solves sparse coding efficiently. However, place cell responses, $\mathbf{s}_{\mathrm{p}}$, and connection matrix, $\mathbf{A}$, are taken to be non-negative in this study.

The code to run the model is available online (`https://github.com/lianyunke/Learning-Place-Cells-from-Grid-Cells-Using-Nonnegative-Sparse-Coding`).

2.4.1 Learning rule

The learning rule for updating the connection strength matrix $\mathbf{A}$ is similar to that in previous studies of sparse coding (Olshausen and Field, 1997; Zhu and Rozell, 2013), as given by

$$\Delta \mathbf{A} = \eta(\mathbf{s}_{\mathrm{g}} - \mathbf{A}\mathbf{s}_{\mathrm{p}}){\mathbf{s}_{\mathrm{p}}}^T, \tag{7}$$

where η is the learning rate. Elements of $\mathbf{A}$ are kept non-negative during training, i.e., the element will be set to 0 if it becomes negative after applying the learning rule described in Eq. (7). Then each column of $\mathbf{A}$ is normalised to unit length, similar to previous studies (Lian et al., 2019; Olshausen and Field, 1997; Rolls et al., 2006; Zhu and Rozell, 2013).

The model dynamics and learning rule described in Eqs. (6) and (7) can be implemented in a biologically realistic network (Lian et al., 2019). Here we simply use the equations described above to demonstrate that the principle of non-negative sparse coding can learn both place cells in the hippocampus and granule cells in the dentate gyrus.

2.5 Training

Since the environment used in this study is 1m×1m, the maximal grid spacing is taken to be smaller than 1m, i.e., $1 \leq N_\lambda \leq 4$. All possible grid spacings are 28 cm, 39.76 cm, 56.46 cm and 80.17 cm. For grid orientation, we have $1 \leq N_\theta \leq 7$. For grid phase, we have the same number of phases in each direction and the maximal number is 5, i.e. $1 \leq N_x = N_y \leq 5$.

There are 100 model cells at the second layer in our simulations, i.e., $N_{\mathrm{p}} = 100$. The dynamical system described by Eq. (6) is implemented by the first-order Euler method, where the membrane time constant is $\tau = 10$ms, consistent with the physiological value (Dayan and Abbott, 2001), the threshold is $\beta = 0.3$, and there are 200 integration time steps with a time step of 0.8ms which we found to provide numerically stable solutions. We use $20,000$ epochs in our training. In each epoch, a random location, $\mathbf{r}$, is presented to the grid cells and the model responses are computed using Eq. (6) and the matrix of connection strengths, $\mathbf{A}$, is updated by Eq. (7). The learning rate, η, is chosen to be 0.03. The parameters above were chosen to ensure a stable solution in a reasonable time scale, but the results were found to be robust to moderate changes of these parameters.

2.6 Recovering the firing fields of model cells

After training, we use the method of reverse correlation to recover the firing fields, denoted as $\mathbf{F}$, of model cells. We present K uniformly sampled random locations, $\mathbf{r}_1, \cdots, \mathbf{r}_K$, to the model, compute according to Eq. (6) the neural responses of a model cell, $s_1, \cdots, s_K$, and then compute the firing field, $\mathbf{F}$, of this model cell by

$$\mathbf{F} = \frac{s_1\mathbf{r}_1 + \cdots + s_K\mathbf{r}_K}{s_1 + \cdots + s_K}. \tag{8}$$

$K = 10^5$ is used in this paper.

2.6.1 Fitting firing fields to functions

The recovered firing field, $\mathbf{F}$ (recovered by Eq 8), is fitted by a function $Q(x, y)$ of the form

$$Q(x, y) = \gamma e^{-\ln(5)\frac{(x-x_c)^2+(y-y_c)^2}{\sigma^2}}, \tag{9}$$

where γ is the amplitude, σ is the breadth of the firing field, and (x_c, y_c) represents the center of the 2D Gaussian function. The built-in MATLAB (version R2020a) function, *lsqcurvefit*, is used to fit these parameters. The fitting error is defined as the square of the ratio between the fitting residual and firing field.

2.6.2 Selecting single-location firing field

Some firing fields of model cells have multiple firing locations and noise in the background. A firing field is categorised as a place field if the following two criteria are satisfied: (1) the fitting error is smaller than 15% (2) the breadth, σ, is larger than 5 cm. These two rules exclude any cells with no obvious firing field or with multiple-location firing field. A model cell is called a *place cell* if its firing field meets the two criteria of a place field.

2.7 Measuring the uniformity of place cell representation

For place cells with a single-location firing field, the field center (x_c, y_c) fitted by Eq. (9) indicates the spatial location that the place cell responds to. We measured how well these place cells represent the entire environment using two measures.

The first measure is *distance to place field*, d_{PF}, which indicates the Euclidean distance between each spatial location (p_x, p_y) in the environment and the nearest place field, described as

$$d_{\mathrm{PF}} = \min_i \left(\sqrt{(x_i - p_x)^2 + (y_i - p_y)^2} \right). \tag{10}$$

If the distance to a place field is large for a location, it means that there are no place fields near this location. Therefore, the distribution of this measure can tell us how well

place fields tile the entire spatial environment. When all spatial locations have small values of this distance to place field, d_{PF}, the entire environment is tiled by the place cells.

The second measure is the *nearest distance*, d_{ND}. For all the centers of place cells with single-location firing field, we define nearest distance as the maximal Euclidean distance of 2 nearest centers of each center (x_j, y_j), described as

$$d_{\mathrm{ND}}(j) = \max\left(\min_{i,i\neq j}^{2}\left(\sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}\right)\right), \tag{11}$$

where $\min^2$ returns a set of 2 smallest values. The distribution of d_{ND} for all centers shows the uniformity of place cells in the environment. However, this measure alone provides little information about the coverage of all place cells because place cells with small nearest distance might only lie in a small sub-region of the entire environment, which would give a small value of this measure but would not represent a good tiling of the entire environment.

The distance to place field, d_{PF}, together with nearest distance, d_{ND}, provides quantitative measures of how well the place cells code for the spatial environment. Small values of both measures indicate that place cells can tile the entire environment fairly evenly. For example, if 100 place cells are organised on a 10×10 grid that evenly tile the 1m $\times$1m environment, the nearest distance will be $100/(10-1) \approx 11.11$ cm for each place cell and the distance to place field for every location is smaller than $11.11/2 \approx 5.56$ cm.

3 Results

3.1 The model can learn single-field place-like cells if grid cells are diverse

Our simulations show that the non-negative sparse coding model proposed here can learn single-location place cells given diverse grid cells as the input. Grid cells with 4 different spacings ($N_\lambda = 4$), 6 different orientations ($N_\theta = 6$) and 5 different phases in both x and y axes ($N_x = N_y = 5$) are used here, i.e., there are 600 grid cells in total ($N_{\mathrm{g}} = N_\lambda N_\theta N_x N_y = 600$).

3.1.1 Learned place cells are localised and tile the whole environment

The place field, $\mathbf{F}$ defined by Eq.(8), for 100 model cells is shown in Figure 3A. All model cells shown here have a single-location firing field. Furthermore, different cells have spatially different firing fields. After learning, the place field of each cell is fitted to a 2D Gaussian function. All cells have a small fitting error ($< 3\%$) and meet the criteria of single-location place field (as defined in Materials and Methods 2.6.2). The centers of all the place cells are displayed together in the 1m$\times$1m spatial environment represented by a 32×32 pixel-like image, Figure 3B, which shows that the centers of the 100 place cells tile the entire environment without any overlap. In addition, the box plot in Figure 3C shows that any location within the space is within a distance of no more than 8.2 cm from the nearest place fields. The histogram of nearest distance of all 100 place cells is displayed in Figure 3D, which shows that the distribution is centered around a mean value of 10.70 cm and standard deviation 0.75 cm. Given that the learned place cells have mean radius 8.69 cm, Figure 3B,C and D illustrate that the learned place cells tile the whole environment rather evenly, i.e, the model learned by non-negative sparse coding can give an accurate neural representation of spatial locations in the environment.

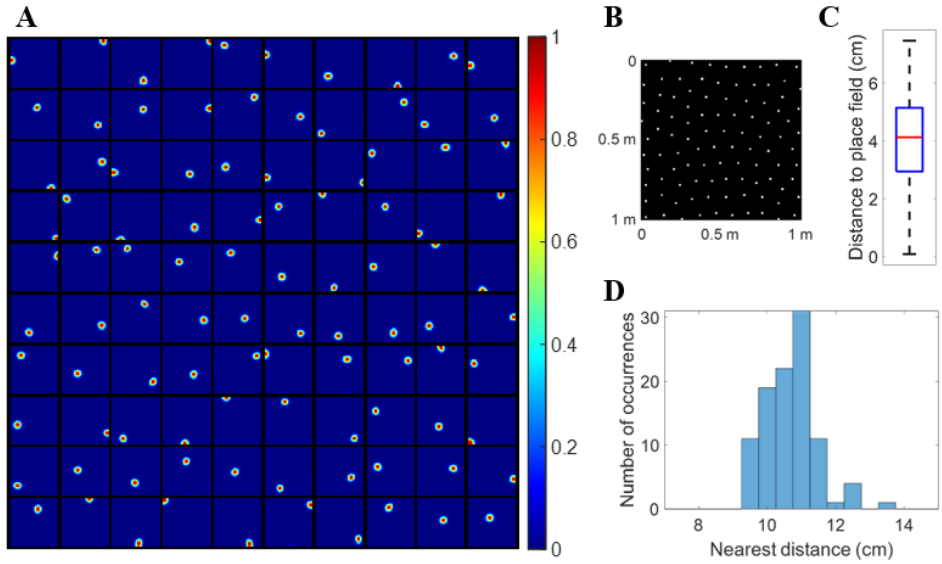


Figure 3. Place fields and their centers learned by the model with $N_\lambda = 4$, $N_\theta = 6$ and $N_x = N_y = 5$. (A) Place fields, **F** (defined by Eq 8), for 100 model cells. All 100 cells learn a place field at different spatial locations. Each block represents the recovered firing field of a model cell in a 1m×1m environment. Values in each block are normalised to [0 1]. The mean radius of place cells is 8.69 cm. **(B)** Centers of the place cells plotted together in the environment, showing that they evenly tile the whole spatial environment. **(C)** Box plot of distance to place field for all spatial locations in the spatial environment. The black lines at the bottom and top indicate the minimum and maximum, and the bottom edge of the blue box, red line inside the blue box and top edge of the blue box represent 25%, 50% (median) and 75% percentile of the data. In this box plot there are no outliers. **(D)** Histogram of nearest distance for all place cells.

3.1.2 The competition introduced by sparse coding provides the inhibition for place cells

The connectivity profile between 600 grid cells and 100 place cells is plotted in Figure 4A, which shows that each place cell selects a group of particular grid cells with different weights. As a result, the overall feedforward connection from the spatial environment to the place cells, namely the matrix product **GA**, has the spatial structure plotted in Figure 4B, which shows that each place cell is selective to one spatial location similar to the recovered firing fields (Figure 3A). However, **GA** has strong average offsets, which can be seen from the grey background in Figure 4B. The model of place cells proposed by Solstad et al. (2006) has an inhibition term to balance the excitation so that the place fields are responsive to a single location. As for the model of place cells proposed by Franzius et al. (2007b), an offset constant is added and signs of model units are adjusted in order to achieve single location place fields. Nevertheless, comparing Figure 3A and Figure 4B, we can conclude that the network implemented by sparse coding naturally introduce the competition to inhibit place cells such that they have firing fields similar to those found in experiments. As stated earlier in Materials and Methods, the sparse coding model used in this paper can be implemented by a biologically realistic network (Lian et al., 2019), suggesting that principle used here can be a potential mechanism used in the navigational system of the brain. The results presented in this study are not sensitive to different parameter values as long as there are diversities in spacing, orientation and phase. Even 81 grid cells are

sufficient for the model to learn place cells that tile the whole environment (see S1 Fig).

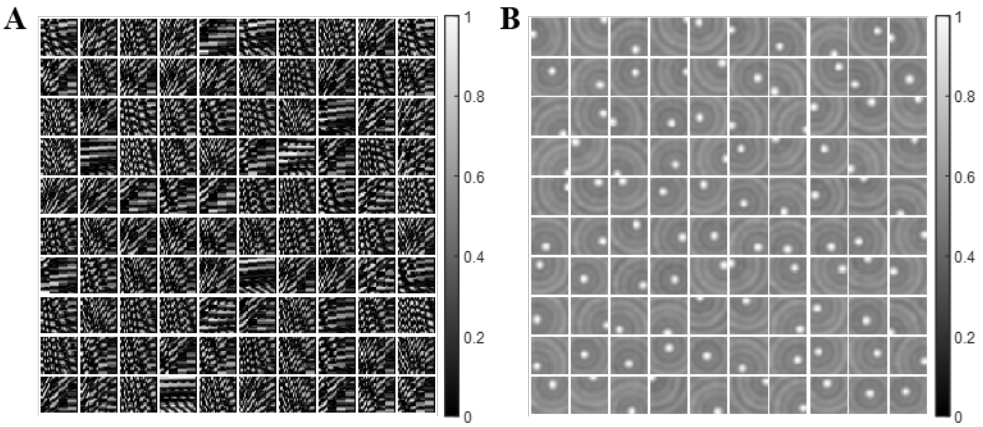


Figure 4. Feedforward connections of the model. (A) Connection strengths between grid and place cells, **A**. Each block represents the connections between 600 grid cells and a place cell on a 24×25 matrix. **(B)** Feedforward connection strengths from the spatial environment to place cells, **GA**. Each block shows the selective spatial structure of a place cell. For both (A) and (B), values in each block are normalised to $[0, 1]$ for plotting.

In addition, the principle of sparse coding forces the model to learn an efficient representation of the grid cell input. The average percentage of active model cells in response to a spatial location is 5.59%. The sparse population activity is consistent with the experimental study that shows sparse ensemble activities in the macaque hippocampus (Skaggs et al., 2007).

3.2 The model can learn cells with multiple firing locations

In this section, we show that the lack of diversity in any grid parameters will prevent the model from learning place cells and cells with multiple firing locations start to emerge, i.e., the same model can learn cells similar to dentate gyrus cells that have multiple firing locations.

3.2.1 When grid cells are less diverse

A lack of diversity in grid spacing results in the emergence of multiple firing locations of the model cells, as illustrated in Figure 5A compared with Figure 3A. Similarly, compared with Figure 3B and C, the lack of diversity in grid orientation or grid phase will also cause the model to learn more cells with multiple firing locations (Figure 5B and C). These model cells are similar to dentate gyrus cells that are found to have multiple firing locations in experimental studies (Jung and McNaughton, 1993; Leutgeb et al., 2007).

Though the model cells learn multiple firing locations using grid cells that lack diversity, limited diversity is sufficient for the model to learn place cells (see S2 Fig).

Recall that the principle of sparse coding finds a linear representation of the input, namely the grid cell responses. Our results suggest that grid cells with less diversity in grid parameters are not sufficient to well represent the whole environment, so that the system gives an ambiguous representation of the spatial location. Therefore, the diverse grid cells found in the MEC are crucial to the emergence of hippocampal place cells. Similar to MEC-hippocampus connections, there are also feedforward connections from

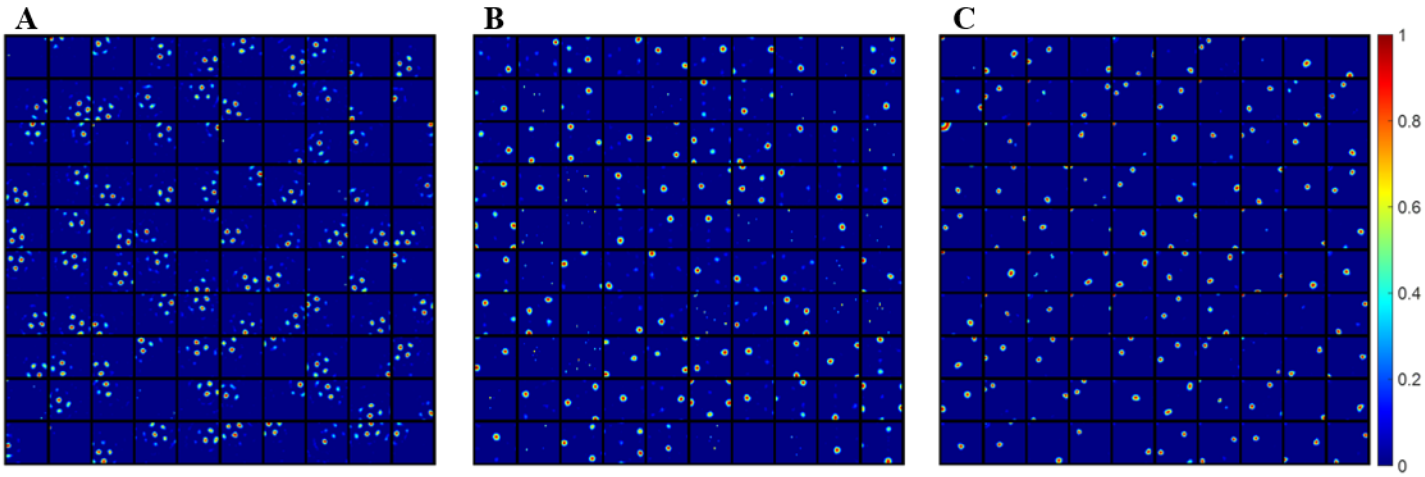


Figure 5. Cells with multiple firing locations start to emerge due to the lack of diversity in any of the grid parameters. Each block represents the recovered firing field, **F** (Eq. 8), of a model cell in a 1m×1m environment. Values in each block are normalised to [0 1]. **(A)** Lack of diverse grid spacings: 1 grid spacing, 6 orientations and 25 phases ($N_\lambda = 1$, $N_\theta = 6$ and $N_x = N_y = 5$). **(B)** Lack of diverse grid orientations: 4 grid spacing, 1 orientation and 25 phases ($N_\lambda = 4$, $N_\theta = 1$ and $N_x = N_y = 5$). **(C)** Lack of diverse grid phases: 4 grid spacing, 6 orientations and 1 phase ($N_\lambda = 4$, $N_\theta = 6$ and $N_x = N_y = 1$).

MEC to the dentate gyrus, so the lack of diversity in afferent grid cells may be one possible factor explaining how cells with multiple firing locations emerge in the dentate gyrus.

3.2.2 When there are fewer model cells

Simulations also show that a smaller number of model cells, N_p, can also cause the model to learn cells with multiple firing locations, even though the grid cells are diverse. This is illustrated in Figure 6, which shows the firing fields of the model when there are different numbers of model cells in the network. The values of the remaining parameters are exactly the same as ones used in Figure 5 which shows well-learned place cells, except the number of model cells, N_p. Figure 6 demonstrates that as the number of model cells, N_p, decreases, cells with more firing-locations start to emerge. The less cells, the larger the proportion of cells with multiple firing locations that emerge. When $N_\mathrm{p} = 10$ all model cells have more than one firing location. When $N_\mathrm{p} = 20$ there are five cells that are categorised as place cells. When N_p is larger than 30, almost all cells are found to have single-location firing fields.

Consequently, though a network with diverse grid cells can represent the spatial environment well, having less model cells does not result in an unique representation of the spatial location with single-location place cells and the animal's location might need to be encoded using more than 1 cell. This suggests more generally that cells with multiple firing locations may be generated by having a small number of cells in the population that implements the sparse coding.

3.3 The spatial resolution of the model increases as more place cells are utilised to represent the environment

As discussed above, when N_p is larger than 30, almost all cells have a single-location firing field. In addition, the learned place cells tile the whole environment rather well

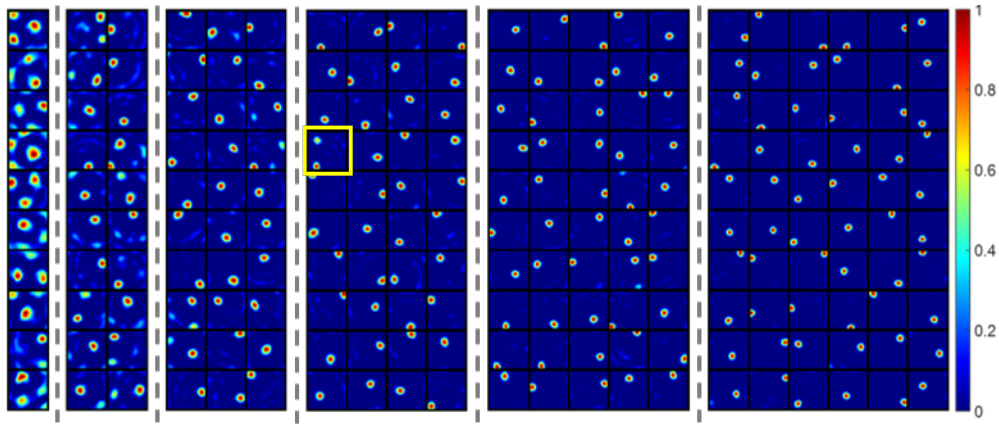


Figure 6. The firing fields, F (recovered by Eq. 8), as the number of cells (N_p) increases from 10 to 60 (separated by dash lines). A smaller number of cells, N_p, leads to more cells with multiple firing fields after learning. Each block represents the recovered firing field of a cell in a 1m×1m environment. Values in each block are normalised to [0 1]. Results shown in this figure have the same parameters as Figure 3 except the number of cells, N_p.

with small values of the nearest distance, Eq.(11). Furthermore, as N_p increases, the mean nearest distance and field breadth decreases (Figure 7), indicating that the spatial resolution of the neural representation by place cells improves.

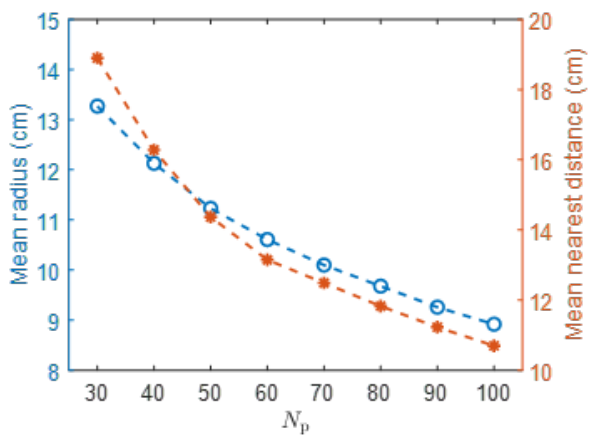


Figure 7. N_p vs. mean radius and N_p vs. mean nearest distance. As N_p increases, the mean radius and mean nearest distance decrease.

3.4 Model results are robust to realistic grid fields

When more realistic grid fields (Materials and Methods, 2.3) are used that incorporate the observed biological variability, the model can still learn a robust representation of the spatial location of the entire environment, as shown in Figure 8.

Figure 8A shows that when grid fields are diverse in grid spacing, orientation and phase, each model cell learns a single-location firing field such that centers of all place fields tile the entire spatial environment rather evenly. The box plot of distance to place field shows fairly small values and indicates the whole environment is covered well. The distribution of nearest distance has mean 10.71 cm and standard deviation 0.72 cm,

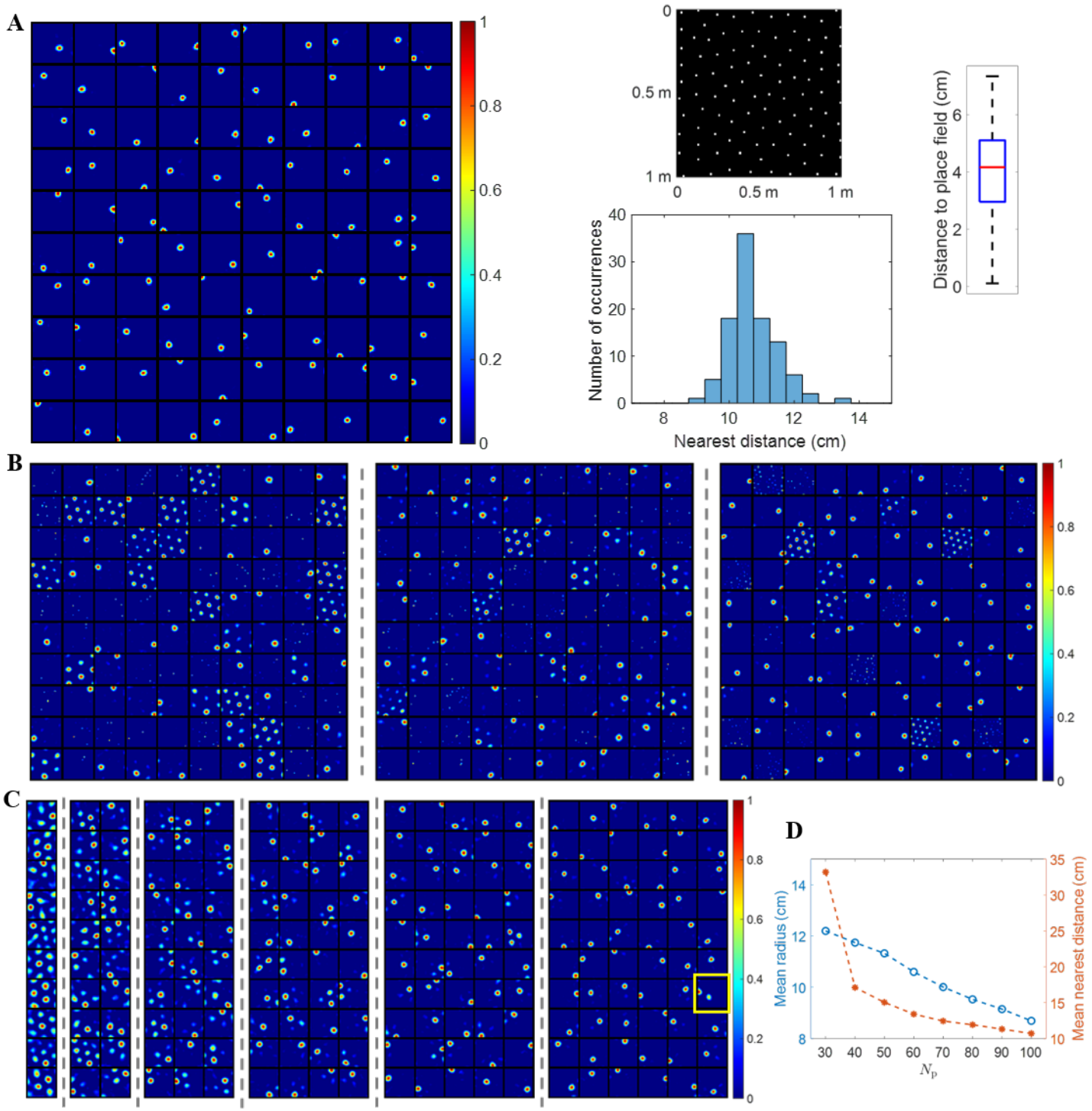


Figure 8. Results are robust when realistic grid fields are used. (A) Learned place cells evenly tile the entire spatial environment (similar to Figure 3). **(B)** The lack of diversity in grid spacing, orientation or phase (left, middle and right plot) leads to cells with multiple firing locations (similar to Figure 5). **(C)** Fewer model cells leads to cells with multiple firing locations (similar to Figure 6). **(D)** Spatial resolution of the neural representation increases as N_p increases (similar to Figure 7).

qualitatively consistent with results shown in Figure 3C (mean 10.70 cm and standard deviation 0.75 cm). Therefore, the learned place cells evenly tile the entire environment.

Figure 8B shows that realistic grid fields with less diversity will learn model cells with multiple firing locations. The left plot displays the firing fields of 100 model cells with less diversity in grid spacing. The standard deviation of spacing in four modules is set to 0 cm instead of 8 cm while the standard deviation of orientation is still 3° and the phase is random. The middle plot shows firing fields when there is less diversity in grid orientation, where the standard deviation of spacing is 4 cm, the standard deviation of orientation is 0° and phase is random. The right plot is for the case of less diversity in grid phase, where the standard deviation of spacing is 8 cm, the standard deviation of orientation is 3° and phase is $(0, 0)$ for all grid cells. Though the model learn cells with multiple locations when there is less diversity in grid parameters, the place cells still tile the entire environment (see S3 Fig).

Figure 8C shows that having fewer model cells makes the model learn cells with multiple firing locations. The six plots separated by dashed lines in Figure 8C represent the firing fields of model cells when the number of model cells, N_{p}, is 10, 20, 30, 40, 50, and 60, respectively. When $N_{\mathrm{p}} = 10$ and 20, there is no place cell. When $N_{\mathrm{p}} \geq 40$, almost all model cells are place cells.

Similar to Figure 7, the neural representation of the spatial environment has better resolution (smaller radius and smaller nearest distance) as N_{p} increases, as seen from Figure 8D.

3.5 The model can generate large place fields

As discussed in a previous study (Neher et al., 2017), most existing model of place cells cannot produce large place fields, such as CA3 place cells with size around 1225 cm^2. The model proposed here can generate large place fields by simply having grid cells with large grid spacings as the input to model cells.

In this part of the study, only grid cells with grid spacings in the fourth module are used; i.e., the grid spacing is sampled from the normal distribution with mean 98.4 cm and standard deviation 8 cm, grid orientation is sampled from the distribution with mean 0° and standard deviation 3°, and grid phases are randomly chosen from a uniform distribution. Similarly, 600 grid cells are used. The number of model cells, N_{p}, is set to 20.

After learning, each model cell pools a particular group of grid cells, similar to Figure 4A. 18 out of 20 model cells satisfy the definition of place cells. Figure 9A shows that large place fields can emerge after learning. Figure 9C shows that these place cells have radii from 18.71 cm to 21.22 cm (mean 19.68 cm and standard deviation 0.75 cm). Therefore, the size of place fields range from 1099.76 cm^2 to 1414.62 cm^2. Figure 9B,D and E show that these 18 place cells with large size cover the entire environment rather evenly.

Above all, the model can learn place cells if the afferent grid cells have large grid spacings, consistent with experimental evidence that the sizes of grid cells and place cells increase along the dorsal-ventral axis (Fyhn et al., 2007; Kjelstrup et al., 2008) and with topographic entorhinal-hippocampal projections along the dorsal-ventral axis (Dolorfo and Amaral, 1998).

3.6 Weakly-tuned MEC cells are sufficient for place cells to emerge

Recent experimental evidence shows that the emergence of hippocampal place cells happens earlier in development than grid cells (Langston et al., 2010; Wills et al., 2010). Here we show that even weakly-tuned MEC cells can provide sufficient spatial

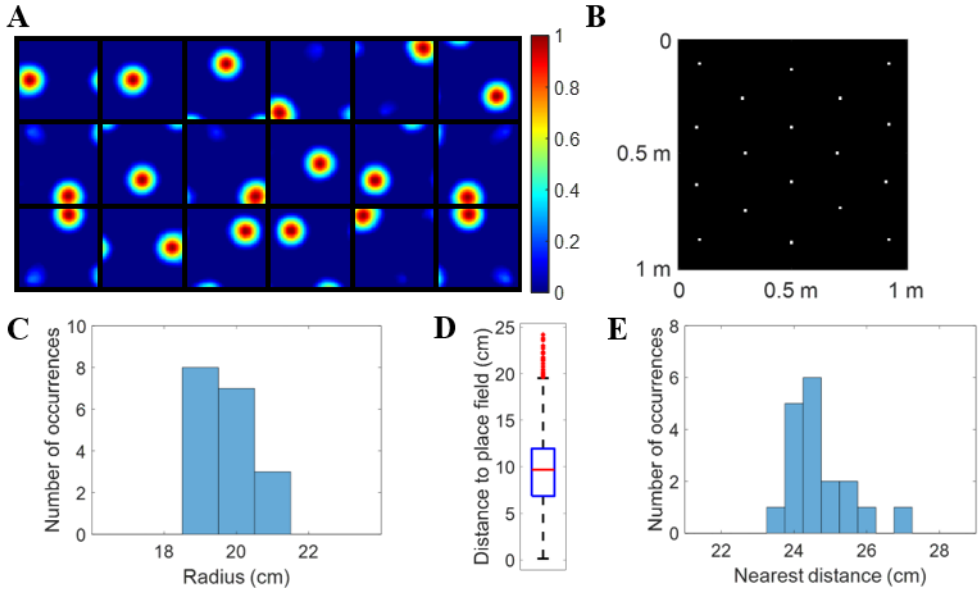


Figure 9. Large place fields emerge. (A) Place fields, **F** (recovered by Eq 8), for 100 model cells. All 100 model cells learn a place field at different spatial locations. Each block represents the recovered firing field of a model cell in a 1m×1m environment. Values in each block are normalised to [0 1]. **(B)** Centers of place cells plotted together in the environment. Place cells evenly tile the whole spatial environment. **(C)** Histogram of radii for all place cells. **(D)** Box plot of distance to place field for all spatial locations in the spatial environment. The black lines at the bottom and top indicate the minimum and maximum, and the bottom edge of the blue box, red line inside the blue box and top edge of the blue box represent 25%, 50% (median) and 75% percentile of the data while outliers are excluded and represented as red dots. **(E)** Histogram of nearest distance for all place cells.

information for the emergence of place cells which have an accurate representation of spatial location. This suggests that place cells can emerge throughout the development of MEC grid cells, from the initial weakly-tuned spatial pattern to the fully developed hexagonal grid pattern.

Weakly-tuned cells are observed to be abundant in the MEC (Zhang et al., 2013). The weakly-tuned field is generated in the simulation by first assigning a random activation, sampled from a uniform distribution between 0 and 1, to each location, then smoothing the map with a Gaussian kernel with standard deviation 6 cm, and normalizing the map such that the values are between 0 and 1 (Neher et al., 2017). Similar to Figure 3 and 8A, 600 weakly-tuned MEC cells and 100 model cells are used. 10 examples of weakly-tuned field are shown in Figure 10A.

The fields of weakly-tuned MEC cells are very different from the periodic pattern of grid cells. Surprisingly, they can nevertheless provide sufficient spatial information such that the model based on sparse coding can decode MEC cell responses and give an accurate representation of the spatial location. Figure 10B shows the firing field of learned place cells. Figure 10C, D and E shows that the centers of place cells evenly tile the entire spatial environment.

Compared with Figure 3 and 8A, using weakly-tuned MEC cells instead of grid cells results in learning a hippocampal place-map with less resolution. The mean radius of place fields in Figure 10B (11.45 cm) is larger than Figure 3 and 8A (8.92 cm and 8.69 cm, respectively). Furthermore, the nearest distance in Figure 10 (mean 12.32 cm and

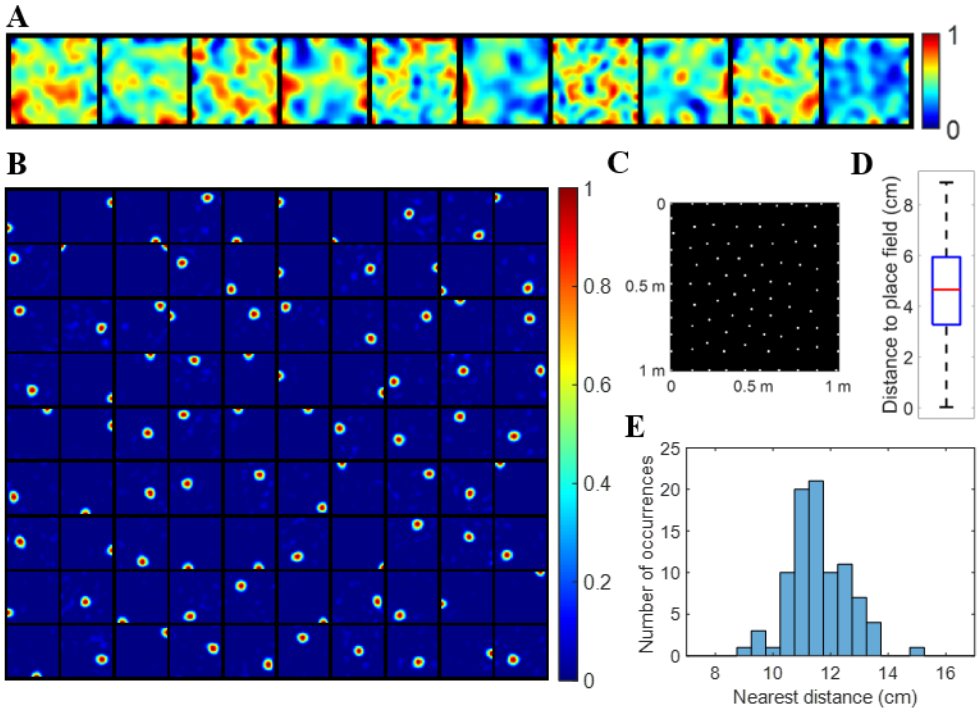


Figure 10. Weakly tuned MEC cells are sufficient for place cells to emerge. **(A)** Examples of weakly-tuned MEC cells. Each block represents the spatial field of a cell in a 1 m×1 m environment. **(B)** Place fields, **F** (recovered by Eq 8), for 90 model cells. 90 out of 100 model cells learn a place field at different spatial locations. Each block represents the recovered firing field of a model cell in a 1 m × 1 m environment. Values in each block are normalised to [0 1]. **(C)** Centers of 90 place cells plotted together in the environment. Place cells evenly tile the whole spatial environment.
(D) Box plot of distance to place field for all spatial locations in the spatial environment. The black lines at the bottom and top indicate the minimum and maximum, and the bottom edge of the blue box, red line inside the blue box and top edge of the blue box represent 25%, 50% (median) and 75% percentile of the data. In this box plot, there are no outliers. **(E)** Histogram of nearest distance for place cells.

standard deviation 7.10 cm) is also larger, compared with Figure 3 (mean 10.70 cm and standard deviation 0.75 cm) and 8A (mean 10.71 cm and standard deviation 0.72 cm) when grid cells are used. The large standard deviation in Figure 10D suggests that the irregular fields of weakly-tuned MEC cells lead to the less even tiling of place cells.

Compared with the model with grid cell input (Figure 3),The average active rate for model cells is 30.68%, much larger than the percentage when grid cells are used (5.59%). Therefore, the place cell map using grid cell input is more efficient and energy-saving.

Furthermore, the model is quite robust to noise and an efficient place map can still be learned, even though a relatively strong noise is added to the MEC cell responses in Eq. (6):

$$\mathbf{s}_{\mathrm{g}} = \mathbf{G}^T\mathbf{r} + \gamma_n \mathbf{n}, \tag{12}$$

where $\mathbf{n}$ is the Gaussian noise with mean 0 and variance 1, and γ_n is the amplitude of the noise. Note that the maximal value of $\mathbf{G}^T\mathbf{r}$ is 1 because $\mathbf{G}$ is normalised to have the maximum 1. We find that the model can still learn an efficient map when γ_n is 0.3 (Figure S4 Fig).

Therefore, the model suggests that place cells emerge earlier than grid cells during

development, in part because the neural system can learn a hippocampal map even when the hexagonal spatial field is not well developed.

Sparse coding can learn place cells even though the input cells (MEC cells in this paper) are weakly tuned to the spatial environment. Thus, input cells with stronger spatial selectivity can provide more spatial information so that unique place field can be decoded by sparse coding. Barry and Burgess (2007) used a learning model to learn place cells from responses of boundary vector cells that are selective to boundaries of the environment at particular angles and distances. Their result can be regarded as a special case of the results presented in this paper, where boundary vector cells are simply input cells with stronger tuning of the spatial environment.

Our results also suggest that these weakly-tuned MEC cells can arise from any form of sensory inputs, such as visual input and auditory input, that encode spatial information. For example, the visual input at different locations of the environment actually carries information about spatial locations and consequently the afferent visual information to MEC can lead to spatially tuned MEC cells. Moreover, the principle of sparse coding can cause the MEC cells to generate a place map. The conjecture proposed here can explain a recent experimental study that shows that place cell firings mainly reflect visual inputs (Chen et al., 2019) and another experimental study that suggests homing abilities of mice even in darkness may not need accurate grid cell firing (Chen et al., 2016).

4 Discussion

In this paper, we applied sparse coding with non-negative constraints to a hierarchical model of grid-to-place cell formation. Our results show that sparse coding can learn an efficient place code that represents the entire environment when grid cells are diverse in grid spacing, orientation and phase. However, lack of diversity in grid cells and fewer model cells leads to the emergence of cells with multiple firing locations, like those cells found in the dentate gyrus. In addition, weakly-tuned Medial Entorhinal Cortex (MEC) cells are sufficient for sparse coding to learn place cells, suggesting that place cells can emerge even when grid cells have not been fully developed.

4.1 Comparison with other learning models

Our work differs from them significantly from previous studies on learning place cells from grid cell input (Franzius et al., 2007b; Neher et al., 2017; Rolls et al., 2006). First, we systematically investigate the influence of the diversity in grid cells upon the formation for place cells. Second, we demonstrate that learned place cells can represent the entire spatial environment well. Third, the same model can produce cells with one firing location, multiple firing locations and large place field size, which can account for the emergence of a range of different observed hippocampal cell types. Fourth, we demonstrate that weakly-tuned MEC cells can also provide sufficient spatial information for the emergence of place cells after learning and the model is very robust to noise. Most importantly, all the results presented in this paper are generated by the same model, namely sparse coding with a non-negative constraint.

Though the principle (ICA) used by Franzius et al. (2007a,b) to learn place cells from grid cells is similar to the principle of sparse coding used here, the place cell examples in their paper are mostly near the boundary of the environment. Our model can generate much better results, learn an efficient place map that covers the entire environment, and automatically provide the needed inhibition by the dynamics of sparse coding. One reason is that place cells in a population are not necessarily independent because nearby cells do overlap, so ICA might put a too strong assumption. Also, the

non-negativity introduced in this paper makes the model more similar to the real neural system, which might help the model uncover important biological properties.

4.2 Properties of grid cells that are necessary for the emergence of place cells

Though the model based on sparse coding can learn place cells when only weakly-tuned MEC cells are used, it does not imply that grid cells are not necessary for their formation. The active firing rate for model place cells when weakly-tuned MEC cells are used as input (30.68%) is much larger than the rate when grid cells are used (5.59%), suggesting that grid cells are more efficient and thereby reduce the energy required by the neural system. Fiete et al. (2008) proposed that grid cells with different spacings and phases altogether form a residual system that efficiently encode the spatial location. In addition, the triangular lattice of the grid pattern is known to be the solution to the optimal circle packing problem (Thue, 1892) and the geometric scale of grid spacings can represent the spatial environment efficiently (Wei et al., 2015).

4.3 Underlying neural circuits

Our study examines the extent to which sparse coding is as an underlying principle in the navigational system of the brain. However, the current model implies no specific neural circuits for the implementation of the sparse coding, rather it is one of the principles that underlies the formation of the neural circuits. Neurophysiological and anatomical studies suggest that the entorhinal cortex and the hippocampus interact via a loop (Tamamaki, 1997; Tamamaki and Nojyo, 1995; Witter et al., 2014). Therefore, feedforward connections from the entorhinal cortex to the hippocampus, recurrent connections within the hippocampus, and feedback connections from the hippocampus to the entorhinal cortex all play an important role, though their specific contributions to the overall function of the network have not been fully uncovered yet. The proposed model based on sparse coding in this study does not rule out any of the network structures mentioned above, as sparse coding can be implemented in neural circuits either in a feedforward network with recurrent connections (Zylberberg et al., 2011) or a network with feedforward-feedback loops (Lian et al., 2019).

4.4 Future work

The current study does not propose a specific biological neural circuit for implementing sparse coding in the entorhinal-hippocampal region, which is the study of ongoing work. Such a model of these neural circuits would need to take into account the experimentally known networks in this area. Also, other properties of place cells such as phase precession (O'Keefe and Recce, 1993), multiple place fields in large environments (Park et al., 2011) and place map in 3D environments (Grieves et al., 2020) will be investigated in the future. In addition, the model here used prefixed grid cells. We did not attempt here to provide a description for how grid cells emerge, but rather the grid cells are assumed to provide an efficient representation of the environment. It would be interesting to also investigate the role of sparse coding in how grid cells themselves emerge. It is hoped that such future work, which incorporates these aspects of the development process of both grid cells and place cells, will provide further insights into how the navigational system of the brain works. Sparse coding represents just one of a number of possible mechanisms that shape network structures, and much remains to be explored to incorporate other mechanisms, such as those associated with the complexities of metabotropic receptor effects, as discussed in Hasselmo et al. (2020).

5 Conclusion

In this study we examined the role of non-negative sparse coding upon hippocampal place cells that receive input from MEC grid cells. The model showed that both place fields and cells with multiple locations can be learned, depending upon specific network parameters. In addition, the learned place cells give an accurate representation of spatial location. Furthermore, weakly-tuned MEC cells are sufficient to drive hippocampal cells to learn place fields. This study elucidates the role of sparse coding as an important mechanism in the navigational system of the brain.

6 Supporting information

S1 Fig. A place map can be learned with a small number grid cells. Place fields and their centers learned by the model with $N_\lambda = 3$, $N_\theta = 3$ and $N_x = N_y = 3$. Only 81 grid cells with diversity in spacing, orientation and phase are sufficient for the model to learn place cells that tile the entire environment. **(A)** Place fields, **F** (recovered by Eq 8), for 100 model cells. All 100 model cells learn a place field at different spatial locations. Each block represents the recovered firing field of a model cell in a 1m×1m environment. Values in each block are normalised to [0 1]. **(B)** Centers of place cells plotted together in the environment. Place cells evenly tile the whole spatial environment. **(C)** Box plot of distance to place field for all spatial locations in the spatial environment. **D** Histogram of nearest distance for all place cells.

S2 Fig. Model cells can learn place cells even with limited diversity of grid parameters. Compared with Fig 5, the model will learn place cells well if the grid parameter has more than 1 different values. Firing fields, **F** (recovered by Eq 8), are show below. Each block represents the recovered firing field of a model cell in a 1m×1m environment. Values in each block are normalised to [0 1]. **(A)** Two different grid spacings are sufficient compared with Fig 5A: 2 grid spacing, 6 orientations and 25 phases ($N_\lambda = 2$, $N_\theta = 6$ and $N_x = N_y = 5$). **(B)** Two different orientations are sufficient compared with Fig 5B: 4 grid spacing, 2 orientations and 25 phases ($N_\lambda = 4$, $N_\theta = 2$ and $N_x = N_y = 5$). Note that 2 out of 100 model cells do not have obvious firing fields. **(C)** Two different phases in $x-$ and $y-$ axes are sufficient compared with Fig 5C: 4 grid spacing, 6 orientations and 4 phases ($N_\lambda = 4$, $N_\theta = 6$ and $N_x = N_y = 2$).

S3 Fig. Effect of loss of diversity in grid cells. The model learns cells with multiple firing locations when the realistic grid fields are less diverse: The place cells still tile the entire environment, but with larger nearest distance. The top row of plots shows the centers of all place cells and the bottom row of histogram shows the distribution of nearest distance. **(A)** When there is less diversity in grid spacings (corresponding to the left plot in Figure 8B). There are 35 place cells out of 100 model cells. The distribution of nearest distance has mean 15.91 cm and standard deviation 2.51 cm. **(B)** When there is less diversity in grid orientation (corresponding to the middle plot in Figure 8B). There are 45 place cells out of 100 model cells. The distribution of nearest distance has mean 14.78 cm and standard deviation 1.95 cm. **(C)** When there is less diversity in grid phase (corresponding to the right plot in Figure 8B). There are 81 place cells out of 100 model cells. The distribution of nearest distance has mean 11.67 cm and standard deviation 1.05 cm.

S4 Fig. The model is robust to noise even when weakly tuned MEC cells are used. When a relatively strong noise is added ($\gamma_n = 0.3$ in Eq. (12)), the model can still learn an efficient place map. **(A)** Place fields, **F** (recovered by Eq 8), for 80

model cells. 80 out of 100 model cells learn a place field at different spatial locations. Each block represents the recovered firing field of a model cell in a 1 m $\times$ 1 m environment. Values in each block are normalised to [0 1]. **(B)** Centers of 80 place cells plotted together in the environment. Place cells tile the whole spatial environment. **(C)** Box plot of distance to place field for all spatial locations in the spatial environment. The black lines at the bottom and top indicate the minimum and maximum, and the bottom edge of the blue box, red line inside the blue box and top edge of the blue box represent 25%, 50% (median) and 75% percentile of the data. In this box plot, there are no outliers. **(D)** Histogram of nearest distance for place cells.

7 Acknowledgments

This work received funding from the Australian Government, via grant AUSMURIB000001 associated with ONR MURI grant N00014-19-1-2571.

References

Barry, C. and Burgess, N. (2007). Learning in a geometric model of place cell firing. *Hippocampus*, 17(9):786–800.

Beyeler, M., Rounds, E. L., Carlson, K. D., Dutt, N., and Krichmar, J. L. (2019). Neural correlates of sparse coding and dimensionality reduction. *PLoS Comput. Biol.*, 15(6):e1006908.

Chen, G., Lu, Y., King, J. A., Cacucci, F., and Burgess, N. (2019). Differential influences of environment and self-motion on place and grid cell firing. *Nat. Commun.*, 10(1):1–11.

Chen, G., Manson, D., Cacucci, F., and Wills, T. J. (2016). Absence of visual input results in the disruption of grid cell firing in the mouse. *Curr. Biol.*, 26(17):2335–2342.

Dayan, P. and Abbott, L. (2001). *Theoretical neuroscience: computational and mathematical modeling of neural systems*. MIT Press Cambridge, MA.

de Almeida, L., Idiart, M., and Lisman, J. E. (2009). The input–output transformation of the hippocampal granule cells: from grid cells to place fields. *J. Neurosci.*, 29(23):7504–7512.

Dolorfo, C. L. and Amaral, D. G. (1998). Entorhinal cortex of the rat: topographic organization of the cells of origin of the perforant path projection to the dentate gyrus. *J. Comp. Neurol.*, 398(1):25–48.

Fiete, I. R., Burak, Y., and Brookings, T. (2008). What grid cells convey about rat location. *J. Neurosci.*, 28(27):6858–6871.

Franzius, M., Sprekeler, H., and Wiskott, L. (2007a). Slowness and sparseness lead to place, head-direction, and spatial-view cells. *PLoS Comput. Biol.*, 3(8):e166.

Franzius, M., Vollgraf, R., and Wiskott, L. (2007b). From grids to places. *J. Comput. Neurosci.*, 22(3):297–299.

Fyhn, M., Hafting, T., Treves, A., Moser, M.-B., and Moser, E. I. (2007). Hippocampal remapping and grid realignment in entorhinal cortex. *Nature*, 446(7132):190–194.

Grieves, R. M., Jedidi-Ayoub, S., Mishchanchuk, K., Liu, A., Renaudineau, S., and Jeffery, K. J. (2020). The place-cell representation of volumetric space in rats. *Nat. Commun.*, 11(1):1–13.

Hafting, T., Fyhn, M., Molden, S., Moser, M.-B., and Moser, E. I. (2005). Microstructure of a spatial map in the entorhinal cortex. *Nature*, 436(7052):801–806.

Hasselmo, M. E., Alexander, A. S., Hoyland, A., Robinson, J. C., Bezaire, M. J., Chapman, G. W., Saudargiene, A., Carstensen, L. C., and Dannenberg, H. (2020). The unexplored territory of neural models: Potential guides for exploring the function of metabotropic neuromodulation. *Neurosci. (In press).*

Hoyer, P. O. (2003). Modeling receptive fields with non-negative sparse coding. *Neurocomputing*, 52:547–552.

Hyvarinen, A. (1999). Survey on independent component analysis. *Neural Comput. Surv.*, 2(4):94–128.

Jung, M. and McNaughton, B. (1993). Spatial selectivity of unit activity in the hippocampal granular layer. *Hippocampus*, 3(2):165–182.

Kjelstrup, K. B., Solstad, T., Brun, V. H., Hafting, T., Leutgeb, S., Witter, M. P., Moser, E. I., and Moser, M.-B. (2008). Finite scale of spatial representation in the hippocampus. *Science*, 321(5885):140–143.

Kropff, E. and Treves, A. (2008). The emergence of grid cells: Intelligent design or just adaptation? *Hippocampus*, 18(12):1256–1269.

Langston, R. F., Ainge, J. A., Couey, J. J., Canto, C. B., Bjerknes, T. L., Witter, M. P., Moser, E. I., and Moser, M.-B. (2010). Development of the spatial representation system in the rat. *Science*, 328(5985):1576–1580.

Lee, D. D. and Seung, H. S. (1999). Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755):788–791.

Leutgeb, J. K., Leutgeb, S., Moser, M.-B., and Moser, E. I. (2007). Pattern separation in the dentate gyrus and CA3 of the hippocampus. *Science*, 315(5814):961–966.

Lian, Y., Grayden, D. B., Kameneva, T., Meffin, H., and Burkitt, A. N. (2019). Toward a biologically plausible model of LGN-V1 pathways based on efficient coding. *Front. Neural Circuits*, 13:13.

Moser, E. I., Kropff, E., and Moser, M.-B. (2008). Place cells, grid cells, and the brain's spatial representation system. *Annu. Rev. Neurosci.*, 31:69–89.

Neher, T., Azizi, A. H., and Cheng, S. (2017). From grid cells to place cells with realistic field sizes. *PLoS ONE*, 12(7):1–27.

O'Keefe, J. (1976). Place units in the hippocampus of the freely moving rat. *Exp. Neurol.*, 51(1):78–109.

O'Keefe, J. and Dostrovsky, J. (1971). The hippocampus as a spatial map: preliminary evidence from unit activity in the freely-moving rat. *Brain Res.*, 34:171–175.

O'Keefe, J. and Recce, M. L. (1993). Phase relationship between hippocampal place units and the EEG theta rhythm. *Hippocampus*, 3(3):317–330.

Olshausen, B. A. and Field, D. J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 381(6583):607–609.

Olshausen, B. A. and Field, D. J. (1997). Sparse coding with an overcomplete basis set: A strategy employed by V1? *Vision Res.*, 37(23):3311–3325.

Park, E., Dvorak, D., and Fenton, A. A. (2011). Ensemble place codes in hippocampus: CA1, CA3, and dentate gyrus place cells have multiple place fields in large environments. *PLoS ONE*, 6(7):e22349.

Rolls, E. T., Stringer, S. M., and Elliot, T. (2006). Entorhinal cortex grid cells can map to hippocampal place cells by competitive learning. *Network*, 17(4):447–465.

Rowland, D. C., Roudi, Y., Moser, M.-B., and Moser, E. I. (2016). Ten years of grid cells. *Annu. Rev. Neurosci.*, 39:19–40.

Rozell, C. J., Johnson, D. H., Baraniuk, R. G., and Olshausen, B. A. (2008). Sparse coding via thresholding and local competition in neural circuits. *Neural Comput.*, 20(10):2526–2563.

Savelli, F. and Knierim, J. J. (2010). Hebbian analysis of the transformation of medial entorhinal grid-cell inputs to hippocampal place fields. *J. Neurophysiol.*, 103(6):3167–3183.

Skaggs, W. E., McNaughton, B. L., Permenter, M., Archibeque, M., Vogt, J., Amaral, D. G., and Barnes, C. A. (2007). EEG sharp waves and sparse ensemble unit activity in the macaque hippocampus. *J. Neurophysiol.*, 98(2):898–910.

Solstad, T., Moser, E. I., and Einevoll, G. T. (2006). From grid cells to place cells: a mathematical model. *Hippocampus*, 16(12):1026–1031.

Stensola, H., Stensola, T., Solstad, T., Frøland, K., Moser, M.-B., and Moser, E. I. (2012). The entorhinal grid map is discretized. *Nature*, 492(7427):72–78.

Steward, O. and Scoville, S. A. (1976). Cells of origin of entorhinal cortical afferents to the hippocampus and fascia dentata of the rat. *J. Comp. Neurol.*, 169(3):347–370.

Tamamaki, N. (1997). Organization of the entorhinal projection to the rat dentate gyrus revealed by Dil anterograde labeling. *Exp. Brain Res.*, 116(2):250–258.

Tamamaki, N. and Nojyo, Y. (1993). Projection of the entorhinal layer II neurons in the rat as revealed by intracellular pressure-injection of neurobiotin. *Hippocampus*, 3(4):471–480.

Tamamaki, N. and Nojyo, Y. (1995). Preservation of topography in the connections between the subiculum, field CA1, and the entorhinal cortex in rats. *J Comp. Neurol.*, 353(3):379–390.

Thue, A. (1892). Om nogle geometrisk talteoretiske Theoremer. *Fordhl. Skand. Naturforsk.*, 14:352–353.

Wei, X. X., Prentice, J., and Balasubramanian, V. (2015). A principle of economy predicts the functional architecture of grid cells. *eLife*, 4:e08362.

Wills, T. J., Cacucci, F., Burgess, N., and O'Keefe, J. (2010). Development of the hippocampal cognitive map in preweanling rats. *Science*, 328(5985):1573–1576.

Witter, M. P., Canto, C. B., Couey, J. J., Koganezawa, N., and O'Reilly, K. C. (2014). Architecture of spatial circuits in the hippocampal region. *Philos. Trans. R. Soc. B: Biol. Sci.*, 369(1635):20120515.

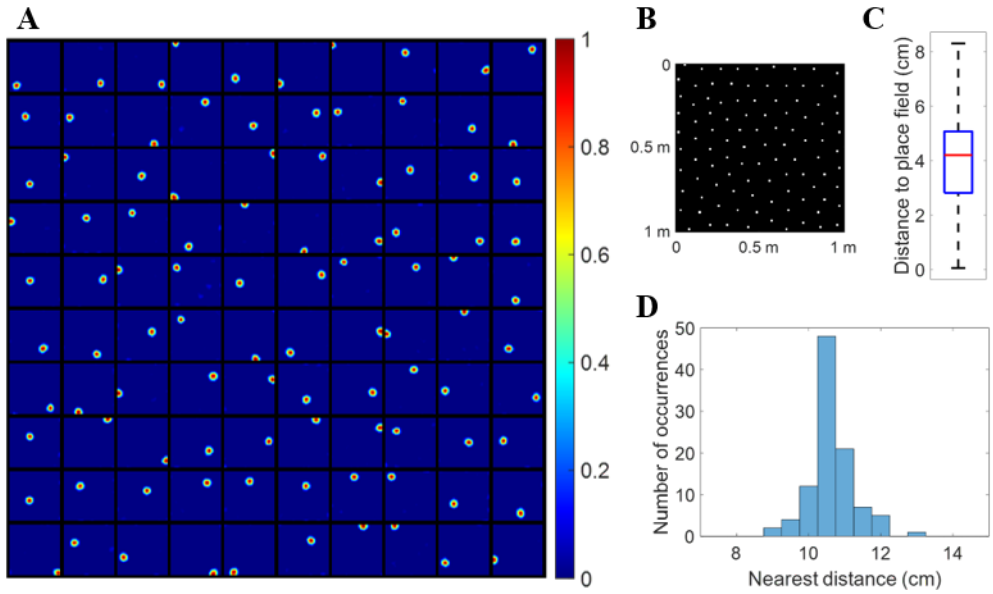
Zhang, S.-J., Ye, J., Miao, C., Tsao, A., Cerniauskas, I., Ledergerber, D., Moser, M.-B., and Moser, E. I. (2013). Optogenetic dissection of entorhinal-hippocampal functional connectivity. *Science*, 340(6128):1232627.

Zhu, M. and Rozell, C. J. (2013). Visual nonclassical receptive field effects emerge from sparse coding in a dynamical system. *PLoS Comput. Biol.*, 9(8):e1003191.

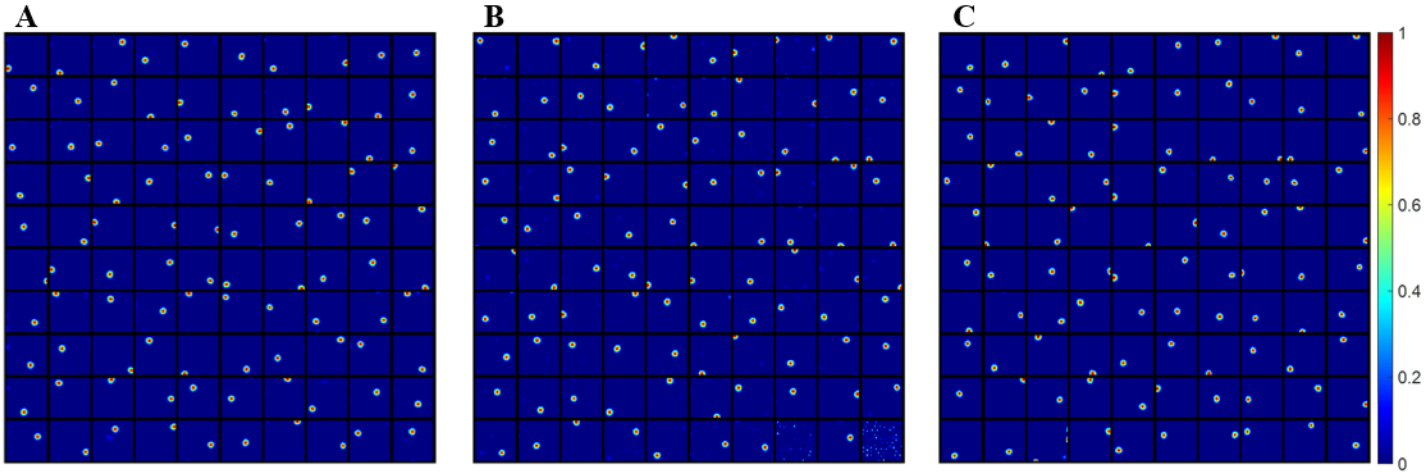
Zylberberg, J., Murphy, J., and DeWeese, M. (2011). A sparse coding model with synaptically local plasticity and spiking neurons can account for the diverse shapes of V1 simple cell receptive fields. *PLoS Comput. Biol.*, 7(10):e1002250.

8 Supporting materials

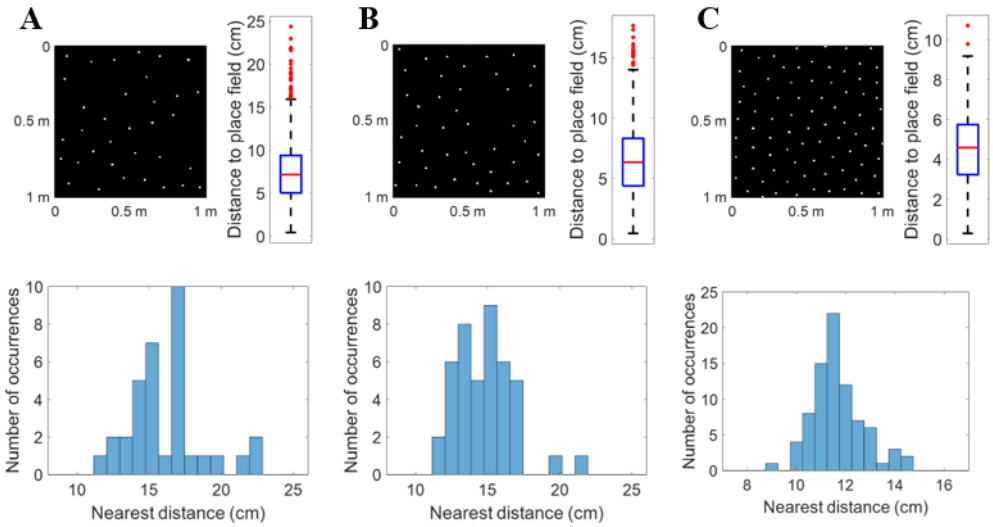
S1 Fig:



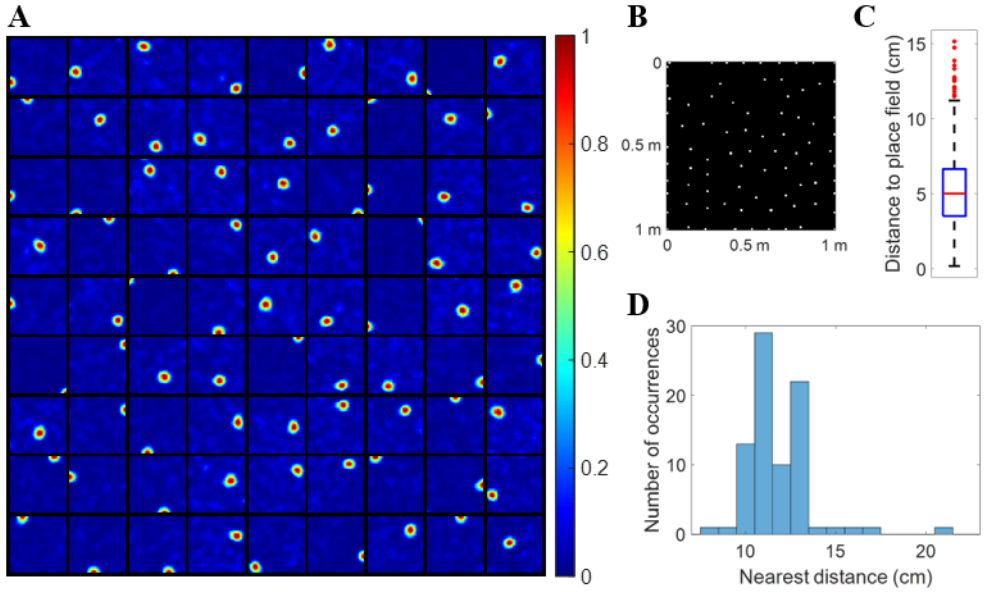
S2 Fig:



S3 Fig:


A
B
C
0
0.5 m
1 m
Distance to place field (cm)
Number of occurrences
Nearest distance (cm)

S4 Fig:


A
B
C
D
0
0.5 m
1 m
Distance to place field (cm)
Number of occurrences
Nearest distance (cm)