

# Multi-Transition Systems: A theory for neural spatial navigation

Nicolai Waniek

**Abstract**—Spatial navigation is fundamental for animals and is attributed to place and grid cells in the rodent brain. Believed to perform path integration or localization, the true objective of grid cells, their hexagonal fields, and especially their discrete scales remain insufficiently understood. Here it is proposed that grid cells encode spatial transitions. First, a model for dendritic computation in grid cells is presented. A network of competitive cells quickly develops positive gridness and realigns the orientation of all cells over time. Next, a scale-space model is introduced for which the optimal scale-increment is shown to be  $\sqrt{2}$ . It improves behaviorally questionable run-times of a single scale significantly by transition look-ahead. Then, a novel formal theory for sequences and transitions is provided to proof optimality of hexagonal transition encoders in Euclidean space. Finally the suggested purpose and results are discussed, testable predictions stated, and relevant connections to computer science noted.

**Index Terms**—grid cells, place cells, spatial navigation, transition systems, scale-space

## I. INTRODUCTION

Decades of research uncovered neurons which are relevant for navigation and express correlation with spatial information. The probably most prominent ones are Head Direction (HD) cells, Place Cells (PCs), and Grid Cells (GCs) [1], all of which can be found in the Hippocampal Formation (HF) [2]. Their combined representations are thought to form a map of the surrounding environment [1, 3], just as anticipated by Edward Tolman in his proposal of a *cognitive map* in 1948 [4].

HD cells show preferential tuning towards directions [5, 6, 7]. The firing rate of an HD cell is maximal when the animal faces the neuron's preferred direction [8, 9]. Thus, a network of HD cells can be considered to resemble a compass [10].

A PC is active only when an animal is in one particular or a few randomly distributed locations of the environment, called place fields [11, 12, 13, 14]. PCs are pyramidal neurons and were discovered in the areas Cornu Ammonis 3 (CA3) and Cornu Ammonis 1 (CA1) of the Hippocampus [15]. Changes of place fields after environmental modifications during experiments suggest that they are influenced by visually perceived geometrical information [16, 17, 18]. Furthermore, PCs were found to integrate non-visual afferents [19, 20]. Due to the stability of place fields over time, PCs are considered to memorize spatial locations [21]. In addition, they were found to be crucial for goal-directed navigation [22, 23, 24, 25, 26].

PCs are subject to certain temporal processes and effects. Consolidation of spatial and episodic memories appears to

happen during Sharp Waves and Ripples (SPW-R) [27, 28, 29, 30]. These are events in which temporally compressed sequences of PCs are re-played in the order in which they were perceived. In addition, a recent study found pre-play activity of PCs in awake but stationary animals [31]. The study reported that sequences of PCs from the current location of an animal to a target location were pre-played in the order in which they were likely to be walked along afterwards, hinting to a path planning operation. Other studies reported PC activity in running animals which is temporally relative to Theta, an oscillation in the Local Field Potential (LFP) at 4 – 10 Hz [32], termed Theta Phase Precession (TPP) [33, 34]. Theta is believed to either provide or be the observable effect of synchronization of neural activity within and beyond the Hippocampus, thereby improving memory consolidation [35]. Synchronization with extra-hippocampal areas such as the Pre-Frontal Cortex (PFC) is necessary, for instance, for decision processes and spatial navigation [36, 37]. During TPP, the PC which corresponds to the current location of the animal is active at the peak of Theta. In addition, several future and past cells spike in order of the spatial location of their place fields during the downward and upward slope, respectively, of the oscillation. Some authors suggest that TPP provides a mechanism for temporal buffering [38, 39].

GCs are stellate cells of the rodent Medial Entorhinal Cortex (mEC) and one synapse upstream of PCs [40]. After their initial discovery in 2005, they were since also discovered in other species, e.g. mice [41] and bats [42, 43]. Opposed to PCs, their spatial correlate is expressed as multiple fields of activity with respect to an environment [44, 45]. Curiously, the grid fields of a single GC arrange in a near-perfect hexagonal tessellation of the environment [40]. Grid fields can be characterized according to their relative phase, orientation, and size within an environment [45]. Commonly, the quality of grid fields is numerically assessed by the gridness score [46], a value which is determined by auto-correlating all responses of a GC within an environment and subsequently measuring how well peaks in the auto-correlation map are distributed hexagonally. It was discovered that grid field sizes increase in discrete steps along the dorsoventral axis, thereby forming several scales of representation [47]. Remarkably, the scales were reported to increase approximately by the factor  $\sqrt{2}$  within and even across animals. Cells of the same scale are said to be part of one grid module, and the responses of one module densely cover the entire environment [40, 47]. Apparently, grid fields are influenced by the geometry of the surrounding environment [48, 49, 50, 51, 52], and a shearing effect was observed in regularly bounded environments [53]. Although GCs need excitatory afferents from the Hippocampus [54]

N. Waniek is with the Neuroscientific System Theory Group (NST), Technische Universität München, Arcisstraße 21, 80333 München, Germany. E-mail: nicolai.waniek@tum.de. This work was partially supported by EU FET project GRIDMAP 600725.

and reportedly depend on Theta in the rodent mEC [55], they were also found to require converging inputs from visual pathways [56]. Furthermore, a somewhat overlooked statistical analysis reported that GCs tend to fire in correlation with an animal's head direction and not with its movement direction [57]. The function of the hexagonal pattern in one and multiple scales is believed to provide a metric representation of space [44].

Conclusively, the HF is important for spatial information processing [11, 22, 58, 59]. In addition, it takes a critical role in the formation and retrieval of episodic memories [59, 60, 61, 62, 63, 64, 65]. It is structurally organized to form several loops of information processing, shows intricate local synaptic circuits of both excitatory and inhibitory neurons, and expresses significant amounts of recurrent connectivity [64, 66, 67, 68]. In particular, the inter-connectivity suggests that areas CA3 and CA1 form an auto- and hetero-associative memory, respectively [69, 70, 71, 72, 73, 74]. While auto-association is suitable for storage and retrieval of patterns even when presented only with partial information, hetero-association allows storage of mappings from inputs to target states [75, 76, 77]. In combination, they can be used to store sequences of data [78].

#### A. Overview of existing models and theoretical investigations

The origin of the hexagonal grid fields as well as their discretized field sizes are discussed controversially. Several models were developed to explain the peculiar arrangement and produce phenomenologically similar responses to real cells [79, 80, 81]. In Continuous Attractor Neural Network (CAN) models, the hexagonal fields appear due to temporal dynamics and recurrent lateral connectivity [82, 83, 84]. Recently, indirect evidence in favor of CAN models was presented by [85]. In Oscillatory Interference (OI) models, the hexagonal patterns emerge due to integration of oscillatory afferents which depend on the Theta rhythm [86]. Although evidence suggests that Theta is required for the stable formation of GCs [87], the necessary accuracy of the oscillation in OI models was not reported so far. Other models use spatially modulated input, e.g. in form of PC activity, to drive a self-organizing process for the hexagonal arrangement [23, 88, 89]. They are supported by the observation that GCs require driving input from the Hippocampus [54].

Theoretical analyses found that Bayesian inference can be used to decode GC activity of multiple scales for the purpose of localization [90]. Although this encoding outperforms PCs [91], it is often assumed that the combination of several scales of GCs converge to form PCs [41, 92]. Another study demonstrated that a scale increment of  $\sqrt{2}$  ideally covers a two dimensional input space when used for localization [93]. Other investigations used the hexagonal pattern for path integration [83, 94]. Furthermore, it was shown that the hexagonal encoding lattice can perform error correction during localization [95].

#### B. Motivating questions

The theoretical investigations and existing models provoke several concerns. First, a redundant encoding for localization

in both GCs and PC is unlikely due to energetically expensive maintenance of neural networks [96], especially when localization in GCs outperforms PCs [91]. Numerous evidence is in favor that PCs perform localization [3, 12, 22, 97], rendering the purpose of GCs under-determined. Second, the hexagonal encoding yields ambiguities when used as a path integration mechanism, and requires stabilization to prevent or reduce drift due to noise in real-world scenarios [98]. Other integration schemes such as a homing vector or a pedometer appear more likely. Third, only few GCs models address temporal aspects of spatial information [99]. However, the Hippocampus is known to form a basis for episodic memory [60, 62, 63, 100], and GCs appear to require episodic data [101, 102, 103].

The goal of this work is to address these concerns. Given that the Hippocampus and Entorhinal Cortex (EC) are required for goal-directed navigation [104, 105], it is proposed that GCs efficiently encode spatial transitions in sequences of locations.

#### C. Organization and brief summary

The paper is organized as follows. First, a model for self-organizing GCs is presented in which each GC learns as many valid transitions as possible via dendritic computations. The model is evaluated in simulations of competitive, recurrently coupled GCs. Grid fields emerge quickly and remain stable over long simulation times.

Next, transition encoding is examined in the light of computational performance and behavioral relevance. Here, a novel scale-space model of GCs is proposed to achieve behaviorally significant run-times. In particular, it is first shown theoretically that the optimal increment across scales is  $\sqrt{2}$ . Then, it is demonstrated that multiple scales improve run-times exponentially in a simplified model. To form the scale-space representation, temporal buffering of spatial locations is required.

Subsequently, the formal theory for transition encoding on which the self-organizing model is based is presented. Here, goal-directed navigation is examined from a logical perspective in which sequences of locations and transitions are formalized mathematically. Several constraints are introduced which are necessary to form consistent sequences. To logically examine transition encoding, inspiration was taken from the analysis of time in distributed computing systems [106], and from approaches to understand and model causality in theoretical computer science [107]. Furthermore, a notation for transitions is used which is based on Communicating Sequential Processes (CSP) [108]. The block concludes with proving that a hexagonal arrangement of transition encoders is optimal for two dimensional Euclidean space with tools and methods from graph theory.

Finally, the results are discussed in detail. They are examined with respect to influential related work and biological findings, testable predictions are made, and links to other areas of research are established.

## II. SELF-ORGANIZING GRID CELLS FOR TRANSITION ENCODING

A formal theory for sequences and transition encoding is introduced further below. Briefly summarized, the results are as

follows. The theory defines transition bundles which encode as many transitions from one symbol to another. Thereby symbols and bundles can be used to generate sequences, for instance for goal-directed navigation. To generate valid sequences, a bundle cannot associate to starting points if it already encodes a transition leading to this point. Given the assumption of a suitable sensory input, it is shown that the optimal distribution of bundles to minimize their number is hexagonal in Euclidean space, and that the minimal number is three.

The results of the formal theory are now used to derive a biologically plausible model of self-organizing GCs. Each cell corresponds to a transition bundle that encodes transitions between spatial locations. The necessary behavior of each cell, i.e. associating with as many presynaptic states as possible, is modeled by dendritic computation in form of multiple dendritic spines, or short *dendrites*.

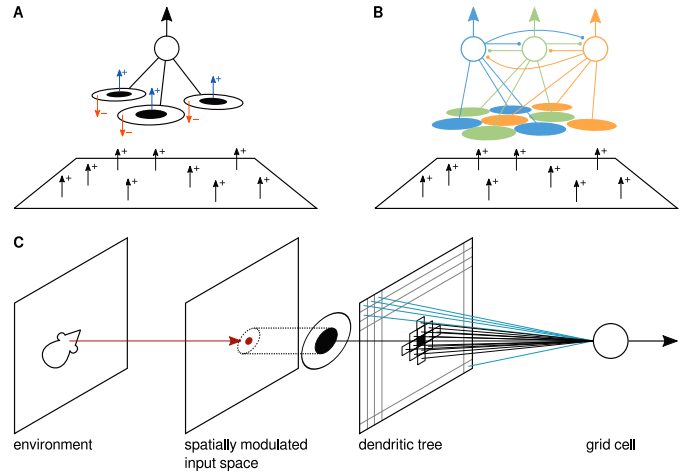
However, a single cell must decorrelate from locations to which one of its memorized transitions leads. In Euclidean space, the target region of a transition is a volume around the current location. Conceptually this is modeled as receptive field dynamics with on-center and off-surround regions, the first corresponding to the start location of a transition and the latter to the target region. The on-center off-surround receptive field is expressed by each dendrite individually. **Figure 1a** illustrates the dendritic tree of a single cell as well as the preference of cells to correlate with multiple inputs in two dimensions.

Due to the decorrelation, a single cell is unable to capture transitions from arbitrary locations in the environment. Any transition which starts at a location from which the cell decorrelated has to be covered by another cell. In concordance with the theoretical results presented further below, the system is modeled as a minimal network of  $N_g = 3$  cells, illustrated in **Figure 1b**. To preserve uniqueness of transitions and their corresponding starting locations, the interactions between cells have to be competitive.

### A. Assumptions and model specification

Each GC samples an input space with  $N_d$  dendrites. Interactions between a GC and presynaptic afferents are modeled as follows, based on several assumption. For instance, it is assumed that any location in the experimental environment induces a unique sensory representation. A likely candidate for this purpose is the Boundary Vector (BV) space [18]. However, coordinates are used to reduce the complexity of the model.

The location  $\mathbf{x}$  of the animal is thought to evoke spikes from several presynaptic neurons with overlapping tuning curves. The spikes are integrated by dendrites  $i$  which encode locations not further apart than  $\sigma_1$ , i.e.  $|\mathbf{x}_i - \mathbf{x}| \leq \sigma_1$ . The totality of these dendrites, denoted as the set  $B_{n,t}$  for cell  $n$  at simulation time  $t$ , is assumed to be sufficient to drive the GC to its spiking threshold. In addition, dendrites  $j$  which encode nearby locations are expected to generate an excitatory post-synaptic potential shortly after the GC fired, i.e.  $\sigma_1 < |\mathbf{x}_j - \mathbf{x}| \leq \sigma_2$  and denoted  $C_{n,t}$ . In the simulations, the temporal integration of presynaptic spikes relative to the spike time of a GC is collapsed into one singular time-step. The simplification can be understood as a binarization of Spike-Timing Dependent Plasticity (STDP) dynamics. Consequently,



**Figure 1:** Conceptual overview of the model and detailed view of the dendritic organization. **(a)** A single cell shows preference to associate with as many inputs as possible, indicated by black arrows with a + sign. Transition constraints are expressed as dendritic computation in form of on-center off-surround fields, indicated by small circles and arrows with - and + signs. **(b)** To cover an entire input space, at least three cells are required which organize their fields due to competitive dynamics. **(c)** The location of the animal induces a singular response (red circle) in a spatially modulated input space. Specifically, coordinates are used to during simulations. The dendritic tree of each grid cell covers the entire input space, and dendrites are organized on a regular lattice according to the image (indicated by little boxes and gray lines). Each dendrite expresses a receptive field of given certain size (circles with black center), which may overlap with neighboring dendrites. Therefore, multiple dendrites may get activated (black small boxes) for a single stimulus. All other dendrites remain silent (blue lines).

the set of all dendrites which receive stimulation at time step  $t$  for cell  $n$  is the set  $D_{n,t} = B_{n,t} \cup C_{n,t}$ , and corresponds to the formation of an on-center and off-surround receptive field. The on-center represents the location at which a transition starts, and the off-center region the target area to which a transition leads.

Due to the correspondence of the input space with coordinates, each dendrite  $i$  can be identified by a coordinate  $\mathbf{x}_i$ . The dendrites are organized on a regular lattice, and each dendrite  $i$  of cell  $n$  stores a weight  $w_{n,i} \in [0, 1]$  that indicates the dendrite's probability to associate with the corresponding input. The entire vector of weights is denoted as  $\mathbf{w}_n$ . For the simulations presented below, the dendritic tree reduces to a two dimensional sheet of weights which covers the entire input space. The processing of presynaptic afferents, dendritic computation, and dendritic organization of a single cell are illustrated in **Figure 1c**.

Activation  $a_n$  of a single cell  $n$  is computed according to **Equation 1** as the weighted sum over all dendrites  $B_{n,t}$  which receive stimulation at time-step  $t$ . Weighting only appears on the dendritic side, i.e. presynaptic activity is considered to be



binary.

$$a_n(t) = \frac{\sum_{i \in B_{n,t}} w_{n,i,t}}{|B_{n,t}|}. \quad (1)$$

Each cell maximizes the number of locations to which it is associated. This is formalized as a function  $L_n(w_n)$ , specified in Equation 2, which tells the *dendritic load* of a cell. It can be understood as a diffusive *eagerness* of each dendrite to associate with arbitrary presynaptic inputs.

$$L_n(t) = \frac{-1}{N_g N_d} \sum_{i=1}^{N_d} (1 - w_{n,i,t})^2. \quad (2)$$

Furthermore, a network of  $N_g$  cells has to minimize co-activation  $K$ , modeled according to Equation 3.

$$K_n(t) = \frac{1}{2(N_g - 1)} \sum_{m \neq n} a_m(B_{m,t}) a_n^2(B_{n,t}). \quad (3)$$

In addition to diffusive dendritic load, each cell has to express specificity for transitions. A cell which associates with a certain presynaptic state has to dissociate from nearby or similar states. Clearly this information is present in the on-center and off-surround dynamics of the dendritic computation described above. The contributed errors for the on- and off-ports are denoted  $E^+(t)$  and  $E^-(t)$ , respectively. Intuitively, the error terms can be understood as follows. Consider the term  $E^-(t)$ . Any dendrite  $i$  which synapses with erroneous presynaptic input, i.e. a state to which a transition that is encoded by the cell leads to, contributes an error proportional to  $w_{n,i,t}$ . The errors are minimized according to the derivatives specified in Equations 4 and 5, respectively.

$$\frac{\partial}{\partial w_{n,k,t}} E^+(t) = \frac{-4(N_g - 1)}{N_g N_d} |B_{n,t}| w_{n,k,t} \quad (4)$$

$$\frac{\partial}{\partial w_{n,k,t}} E^-(t) = \frac{4(N_g - 1)}{N_g N_d} |C_{n,t}| w_{n,k,t} \quad (5)$$

The derivatives are inspired by the instar learning rule [109]. Hereby, minimization depends only on  $w_{n,k,t}$  and the total amount of stimulated dendrites  $|B_{n,t}|$  and  $|C_{n,t}|$ . Specific cross-talk with potentially remote dendrites  $w_{n,i,t}$ ,  $i \neq k$ , perceived to be biologically unlikely, is thereby avoided.

Summarized, a GC is characterized by the error function  $F_n$  of Equation 6.

$$F_n(t) = L_n(t) + K_n(t) + E_n^+(t) + E_n^-(t). \quad (6)$$

The error function can be derived with respect to a weight  $w_{n,k,t}$  to determine a gradient descent learning rule to minimize the error over time, shown in Equation 7.

$$\begin{aligned} \frac{\partial}{\partial w_{n,k,t}} F(t) = & -\frac{2}{N_g N_d} (1 - w_{n,k}) \\ & + \frac{1}{N_g - 1} \sum_{m \neq n} a_m(B_{m,t}) a_n(B_{n,t}) \\ & + \frac{\partial}{\partial w_{n,k,t}} E^+(t) + \frac{\partial}{\partial w_{n,k,t}} E^-(t) \quad (7) \end{aligned}$$

Changes to dendritic weights according to Equation 7 are subject to global winner-take-all and local activation mechanisms. All cells perform weight updates with respect to  $L_n(t)$  and  $K_n(t)$ , however limited to dendrites  $D_t$ . Furthermore, only the most active GC, i.e. if  $n = \arg \max_m a_m(t)$ , receives non-zero gradients  $E^+$  and  $E^-$ . The gradients  $E^+$  and  $E^-$  are only applied to corresponding dendritic weights, i.e. only dendrites of  $B_{m,t}$  receive a gradient for  $E^+$  and only dendrites of  $C_{m,t}$  a gradient for  $E^-$ . Thereby, changes are local to dendrites which perceived presynaptic activity and proportional to their individual weights. Subsequently, the update of a weight  $w_{n,k,t}$  is modulated non-linearly. Finally, all weights are rectified, i.e. weights are clamped from below to zero, indicated by the bracket notation  $[\cdot]^+$ . Conclusively, a weight update follows according to Equations 8, 9 and 10.

$$w_{n,k,t+1} = \left[ \tanh \left( w_{n,k,t} - G(t) \right) \right]^+ \quad (8)$$

$$G(t) = \begin{cases} \eta(t) \frac{\partial}{\partial w_{n,k,t}} F(t), & \text{if } w_{n,k,t} \in D_{n,t} \\ 0, & \text{otherwise} \end{cases} \quad (9)$$

$$\eta(t) = \frac{1}{3} \exp \left( -\frac{1}{s_{\text{avg}}} s_t^2 \right) \quad (10)$$

The learning rate  $\eta(t)$  depends on the current speed  $s_t$  in  $m/s$  of the animal as well as the average speed  $s_{\text{avg}}$ . This dependence models increased uncertainty of perception. It is assumed that an animal which moves at a high speed is less likely to identify locations perfectly. Correspondingly, the accuracy of learning transitions declines with increased speed, expressed by a lowered learning rate. In the simulations presented below, the average speed  $s_{\text{avg}}$  was extracted from pre-computed trajectories. However, a moving average is expected to yield qualitatively similar results.

## B. Simulation setup and results

A dendritic weight is initialized to 1.0 with a probability of 0.1, or set to zero otherwise. Input to the model was presented in form of the speed of the simulated animal to adjust  $\eta(t)$ , as well as the location to drive the presynaptic state. The simulated animal moved throughout the environment with movement statistics close to real recordings, extracted from recordings by [40]. An example of a simulated trajectory as well as statistics of the distribution of speeds and angular velocities are given in Figure 2. The trajectories were pre-computed to extract  $s_{\text{avg}}$ . In each simulation, the animal started in the middle of the environment. The network model was simulated 400 times for the minimal number of  $N_g = 3$  cells per network, each cell exhibiting  $N_d = 48$  dendritic weights. Dendritic stimulation due to presynaptic activation was governed by the parameters  $\sigma_1 = 0.10$ , and  $\sigma_2 = 2\sigma_1$ .

The gridness score of a cell was calculated without generating spikes. The value was extracted directly from the dendritic weights. All weights  $w_{n,k} \in [0, 1]$ , thus they encode a probability to spike given a certain input. Consequently, the distribution of spikes precisely follows the dendritic weights. The gridness score was computed according to Sargolini et al. [46], however without previously smoothing the weight

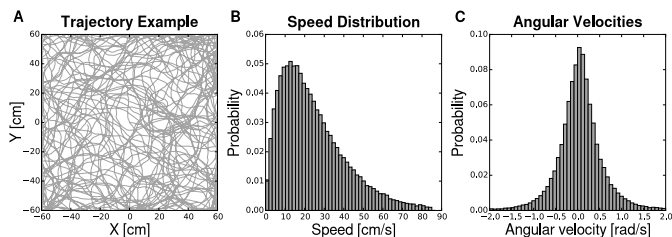


Figure 2: Realistic trajectories (a) were generated by tuning the distribution of speeds (b) as well as angular velocities (c) to resemble real recordings.

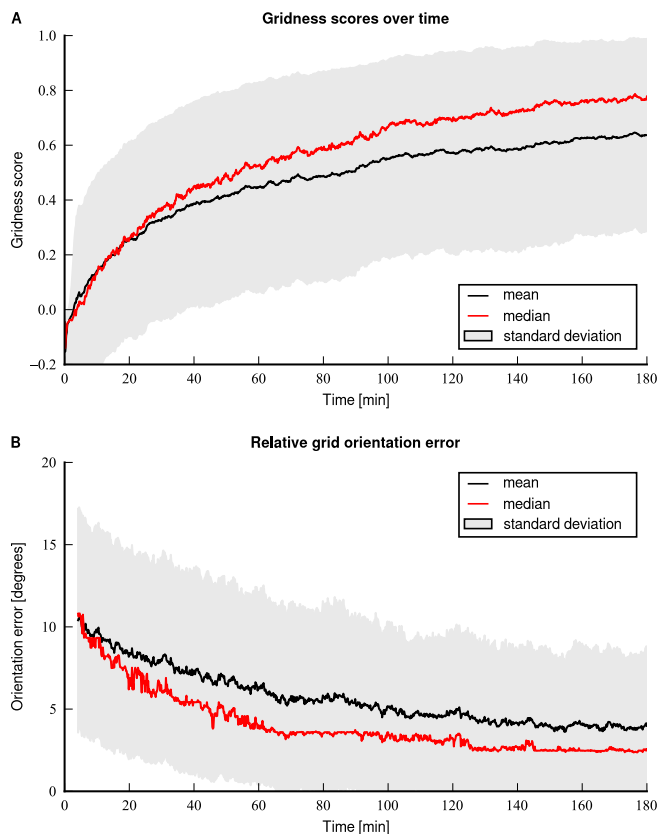


Figure 3: Gridness scores and orientation over time for 400 simulations. (a) Gridness increases over time, the median stays above zero already after approximately 2.5 minutes of simulated time. The mean remains positive after about 4 minutes. (b) Relative grid orientation errors decline over time. The first four minutes of data were cut off to allow the cells to form detectable peaks in their weight distributions.

maps. Furthermore, the orientation of the dendritic weights was computed following the methods described by [40]. The relative grid orientation error of the network was computed by accumulating the mutual differences of orientations between all cells.

Both, gridness scores and relative orientation errors, are presented in Figure 3 for 400 simulations with a total simulated duration of 180 minutes. The top row of the figure shows that cells with a gridness score above zero appeared early in most of the simulations. Furthermore, the bottom row demonstrates

that relative grid orientation errors declined over time.

Examples of three simulations, one with a high final average gridness score, one with mean score, and one with low score, are depicted in Figure 4. The figure displays the dendritic weights of each of the neurons at several points in time during the simulation in form of heat maps. The first column shows the dendritic weights after a few iterations of the system and not immediately after initialization, visible in form of slightly increased values in the central weights, and that the weight maps are presented as-is and were not post-processed with a smoothing filter. Each panel contains a small inlay displaying the gridness score of the presented weight distribution. The emerging weights are non-binary and smooth despite the binary presynaptic activation of the dendritic tree. In addition, the weights at time  $t = 180$  min indicate effects similar to recently reported distortions such as skewing, or misalignments with walls [53]. Analysis of these effects is left for future work, though. The last column of Figure 4 shows the auto-correlogram of the dendritic weights at the end of each simulation. The detected orientation of the arrangement of weights is displayed in form of a black bar from the central peak to the first peak above the horizon in counter-clockwise direction.

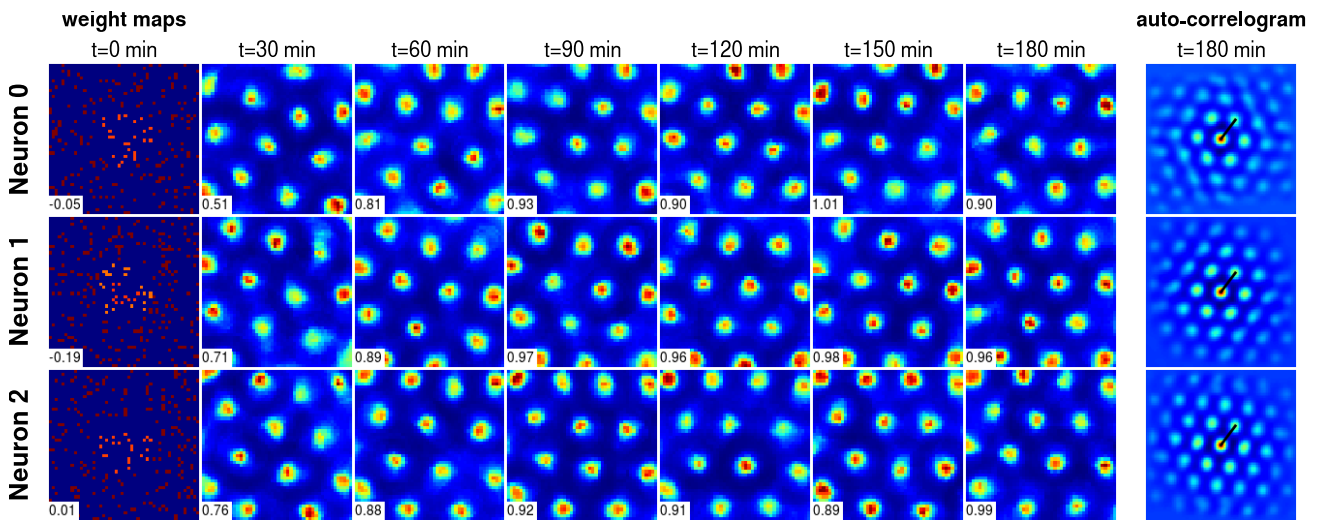
### III. A SCALE-SPACE MODEL OF GRID CELLS

The entorhinal-hippocampal loop is proposed to form a Multi-Transition System (MTS) for both episodic sequences as well as goal-directed navigation, and stores symbols and transition bundles as follows. A neuron representing a spatial symbol can be understood as a PC, hence the terms symbol and PC are used interchangeably. Novel symbols are acquired in a neural associative memory  $M_{\Sigma}$ , episodic and actually performed transitions are recorded as bundles in  $M_{\Pi}$ , while purely spatial bundles are subject of  $M_{\Gamma}$ . Spatially modulated afferents, required for the recruitment of new symbols and transitions, are represented by neural memory  $M_{\Delta}$ . The interactions between these memories are depicted in Figure 5d and focus of the remainder of this section. The model is inspired by previous work by [110, 111] and [112].

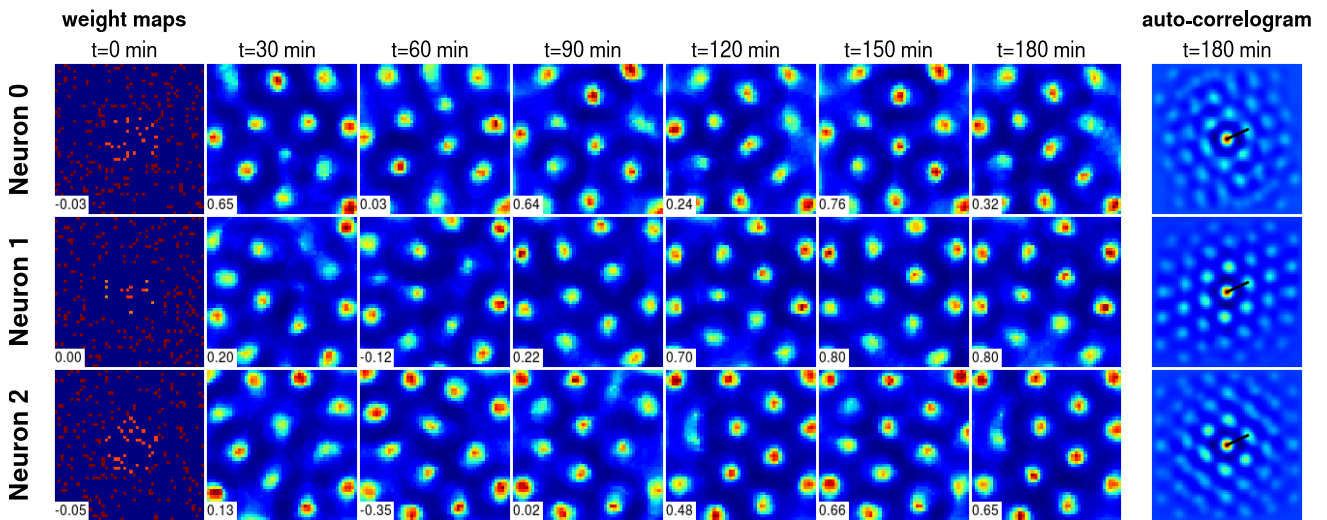
The system can be queried to expand the path  $\sigma_s \rightsquigarrow \sigma_t$  by recursion. After activation of a start symbol  $\sigma_s$  in  $M_{\Sigma}$  all corresponding transitions activate in  $M_{\Pi}$  and  $M_{\Gamma}$ . Recursively this leads to activation of subsequent symbols in  $M_{\Sigma}$  due to the bi-directional connection between  $M_{\Sigma}$  and the two transition memories. Simultaneous co-activation of multiple symbols or transitions is allowed. Hereby the system is able to expand  $\sigma_s \rightsquigarrow \sigma_t$  into known paths in parallel until the target state  $\sigma_t$  is activated in  $M_{\Sigma}$ , or until a maximal number of recursions is reached.

#### A. Multi-scale transitions and the number $\sqrt{2}$

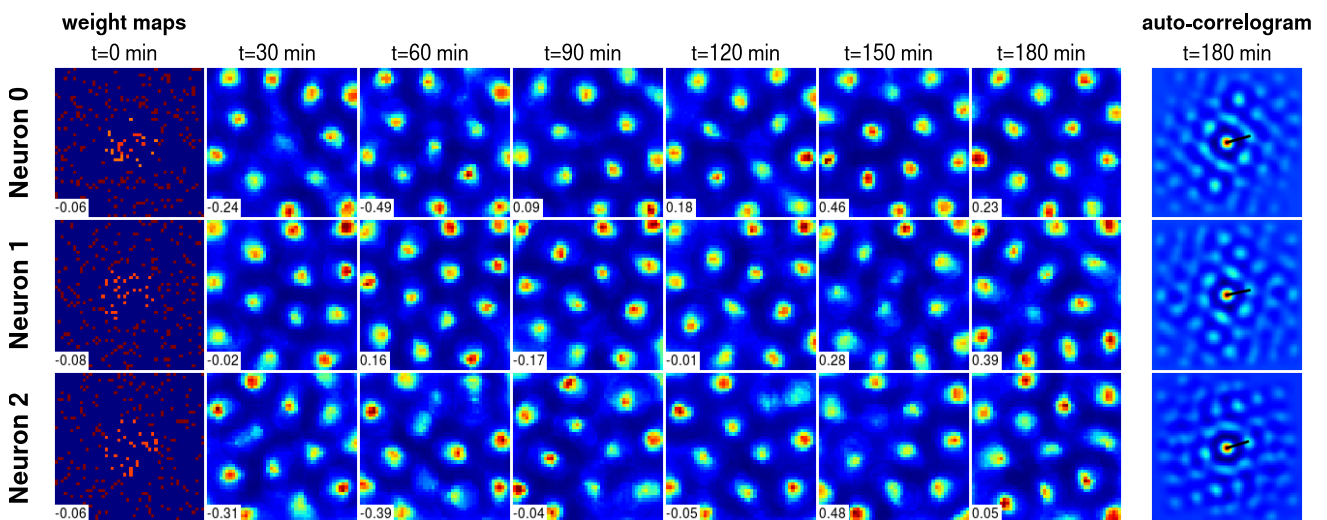
Both the outlined model above as well as the ones published by [110, 111] or [112] have a biologically significant issue. Consider the computational and thus behavioral implications of the recursive retrieval in the transition system for an animal which has to travel 200m in a straight line from a feeding site back to its home location. The animal is equipped with an entorhinal-hippocampal transition system for path planning



Simulation #311. Final average gridness score = 0.949. Final relative orientation error = 0.000°.



Simulation #60. Final average gridness score = 0.589. Final relative orientation error = 0.000°.



Simulation #19. Final average gridness score = 0.223. Final relative orientation error = 3.339°.

Figure 4: Examples of grid cells at several points in time during the simulation and final auto-correlograms with indicated grid orientation. White inlays with numerical values show gridness score of the respective weight map.



which acquires novel spatial symbols every 20 cm. Thus, 1000 spatial symbols cover the path from start to goal, ignoring overlapping place fields. Furthermore, assume that temporal dynamics and delays for spike generation and propagation consume 10 ms of time per memory. After activation of a symbol in  $M_\Sigma$ , recursive retrieval with the help of  $M_\Pi$  and  $M_\Sigma$  therefore requires 20 ms of time to activate a subsequent symbol in  $M_\Sigma$ . In total, this accumulates to 20 s of time to query the existence of an expansion – if it exists – in which the animal may fall victim to a roaming predator. Clearly an acceleration technique is required.

From the perspective of computer science, symbols and transitions stored in a transition system form an ordered list of data. Given random access to elements, binary search is known to demonstrate the asymptotically optimal runtime [113]. For lists without random access, skip lists were developed [114]. The data structure builds a hierarchy of look-ahead links, often called *fast lanes*, by which search is accelerated exponentially. Given uniformly sampled elements of data, skip lists express an immediate duality to binary search [115]. In particular, the hierarchically structured fast-lanes exponentially decrease the search time for a range query on discrete and equidistant data. The look-ahead distance is governed by a factor of two, depicted in Figure 5a. Two consecutive locations have to be compared and combined as fast lane to form the next level of transitions in this hierarchical approach.

How does this relate to the entorhinal-hippocampal loop? The value represented or detected by a single real neuron depends on its tuning curve, which is usually bell-shaped [116, 117]. Furthermore, the perception of two consecutive locations is assumed to depend on optimally sampling an input space. Due to a finite number of neurons to represent sensory stimuli, it can be assumed that there is a minimal resolution to discriminate between two (sensorily) adjacent positions, called *eigenresolution* and denoted  $\sigma_{\text{eigen}}$  and  $\Sigma_{\text{eigen}}$  in the one and multidimensional case, respectively. In other words, two locations can be distinguished if they are at least  $\sigma_{\text{eigen}}$  apart in the sensory space. For ease of comprehension,  $\sigma_{\text{eigen}}$  can be understood to correspond to a distance in Euclidean metric space, similar to the parameter  $\sigma_1$  used for the self-organizing grid cell model above. Conclusively, neural afferents to GCs are proposed to be described by an  $n$ -dimensionally normally distributed Probability Density Function (pdf) according to Equation 11.

$$G(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{D} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right) \quad (11)$$

The pdf is parametrized by the preferred stimulus  $\boldsymbol{\mu}$  and tuning width expressed by the covariance matrix  $\boldsymbol{\Sigma}$ . The normalizer  $D$  is specified according to  $D = (2\pi)^{n/2} |\boldsymbol{\Sigma}|^{1/2}$ .

Combining two locations to construct a transition look-ahead requires to convolve two consecutive spatial samples, parametrized by  $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2$ , and denoted as  $\boldsymbol{\Sigma}_1$ , and  $\boldsymbol{\Sigma}_2$ . Combining distant locations would violate the coherency constraint of Multi-Transition Theory (MTT), presented further below. It can be shown that convolving two normally distributed pdfs results in yet another normally distributed pdf with combined mean and co-variances (see supplementary material to [118]

for details). In particular, the resulting pdf is parametrized according to  $\boldsymbol{\mu} = \boldsymbol{\mu}_1 + \boldsymbol{\mu}_2$  and  $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2$ . Given uniform input sampling and symmetry, i.e.  $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}_{\text{eigen}}$  and  $\boldsymbol{\Sigma}_{\text{eigen}} = \sigma_{\text{eigen}}^2 \mathbf{I}$ , it follows that the variance  $\sigma^2$  of a sampling process for spatial look-ahead is  $\sigma^2 = 2\sigma_{\text{eigen}}^2$ . Thereby the integration area of a spatial sampling process *doubles* during the construction of an additional scale to perform transition look-ahead, regardless of the dimensionality of the pdf. The convolution process and three consecutive scales are depicted for the one-dimensional case in Figure 5b and c, respectively.

Conclusively, the radius of the spatial sampling process ideally increases by a factor of  $\sqrt{2}$  from the first scale to the next. In concordance, the spatial period of transition encoders which perform spatial sampling as part of their dendritic computation increases by  $\sqrt{2}$ . Repeatedly applied, the technique forms an entire stack of probabilistic look-ahead transition encoders with a discrete scale increment by a factor of  $\sqrt{2}$ . This is similar to the *fast lanes* of skip lists in the asymptotically optimal case, however with the difference of using probabilistic instead of discrete data.

### B. Implementation details of a proof-of-principle model

The theoretical results were simulated in the following deterministic manner as proof-of-principle. The model consists of the four memories  $M_\Delta$  to represent input,  $M_\Sigma$  to store symbols,  $M_\Pi$  to record temporal transitions, and  $M_\Gamma$  for the storage of spatial transitions, interacting according to Figure 5d. Each memory sacrifices biological accuracy to focus on the effect and requirements of using and learning a scale-space model of transition encoding.

Likewise the input to the self-organizing grid cell model above, the memory  $M_\Delta$  reports the coordinate of the animal in two dimensional space to reduce the complexity of the model. The memories  $M_\Sigma$  and  $M_\Pi$  are implemented as artificial neural networks similar to growing neural gas [119].  $M_\Sigma$  associates with presynaptic states of  $M_\Delta$  and thereby corresponds to PCs, whereas  $M_\Pi$  records actually performed transitions from one neuron of  $M_\Sigma$  to another. Thus, any neuron in  $M_\Pi$  associates presynaptically with a single symbol neuron of  $M_\Sigma$ , and recurrently connects back to all symbols in  $M_\Sigma$  to which transitions were detected. A novel transition neuron is recruited in  $M_\Pi$  only when the corresponding symbol of  $M_\Sigma$  is not yet known to  $M_\Pi$ . Recruitment of novel neurons in  $M_\Sigma$  is determined by the difference between the afferent from  $M_\Delta$  and the value represented by the best matching unit. Given the use of Euclidean coordinates in  $M_\Delta$ , this is simply the distance between the respective coordinates. In the results presented below, recruitment of novel PCs was triggered when the best matching unit represented a value that was displaced by at least 10 cm from the current location of the animal.

The distribution of optimal transition bundles in memory  $M_\Gamma$  is predetermined to simplify the simulation. Instead of learning the optimal distribution of the bundles similar to the self-organizing model above, hexagonally distributed Voronoi cells with a radius of 20 cm are used on top of the coordinate space provided by  $M_\Delta$ , and the scale-increment from one scale to the next is determined by the factor  $\sqrt{2}$ . Although

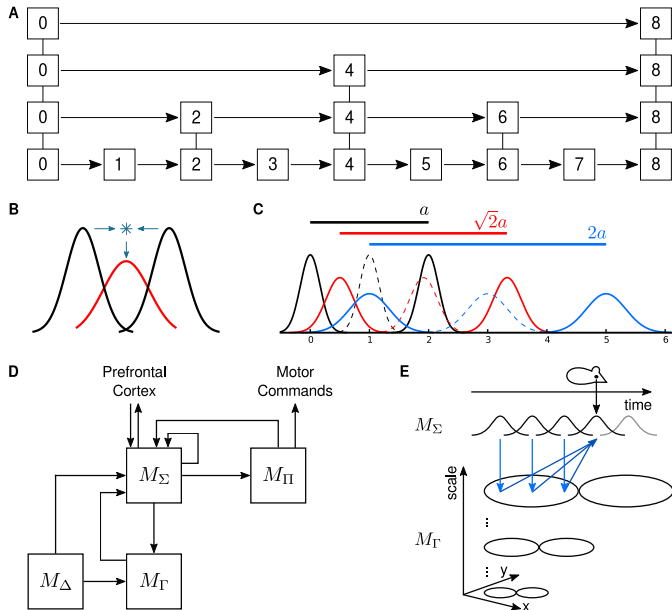


Figure 5: Skip list, scale-space model construction, overview, and learning. **(a)** A skip list for discrete data is a data structure which uses a hierarchy of *fast lanes* to accelerate search [114]. In the ideal case illustrated here, the hierarchy guarantees logarithmic time, i.e. exponential speed up of retrieval, and corresponds to search in a binary tree. For fast lane traversal, comparison of items is required. **(b)** For probabilistic data represented in a neural network, convolution of two consecutive normally distributed probability density functions (pdf) is required. This yields another pdf with variance increased by factor  $\sqrt{2}$ . **(c)** Recursively applied this constructs a scale-space representation. The scale space allows to compare items across larger distances of the represented space, similar to the comparison on fast lanes in a discrete skip list. Dashed lines indicate off-surround areas. **(d)** The final model consists of the four memories  $M_\Delta$ , providing spatially unique information,  $M_\Sigma$ , storing spatial symbols or place cells,  $M_\Gamma$ , which learns spatial transitions between perceived locations, and  $M_\Pi$ , storing actually performed or *temporal* transitions between symbols. The memories  $M_\Sigma$  and  $M_\Gamma$  directly correspond to the items and transitions of a discrete skip list. The model was inspired by previous work by [110, 111]. **(e)** During learning, acquisition of large-scale transitions requires to access locations and symbols which were perceived in the past to learn the association with the currently active symbol. Input from  $M_\Delta$ , which is co-active with  $M_\Sigma$  during learning, is omitted in the figure for clarity, and that only the on-center regions of transition encoders of  $M_\Gamma$  are shown for the same reason.

absolute areas in form of Voronoi cells neglect the probabilistic nature of neural encodings, used to derive the optimal scale increment of  $\sqrt{2}$ , it is believed to suffice for demonstration of proof-of-principle.

Learning look-ahead transitions requires to buffer symbols temporally. Identification of locations and recruitment of symbols in  $M_\Sigma$  depends on afferents from  $M_\Delta$ , similar to the smallest scale of GCs in  $M_\Gamma$ . Without additional information, a symbol is thus only active in a place field that is expected to depend in size on the resolution  $\sigma_{\text{eigen}}$  of the sensor representation. However, any PC which falls into the on-center region of a large-scale GC indicates the start of a look-ahead transition. Similarly, each PC which is active immediately after the large-scale GC has to be considered a target of a look-ahead transition. Consequently, a large-scale GC has to associate with several previously experienced PCs before spiking and dissociate from prospective PCs to learn look-ahead transitions. Hence, all symbols of  $M_\Sigma$  are buffered in consecutive order to allow learning, illustrated in Figure 5e. Conclusively, learning of transitions in  $M_\Gamma$  is gated by co-activation of afferents from  $M_\Delta$  and temporally buffered information from  $M_\Sigma$ . The biological feasibility of buffering as well as co-activation learning will be discussed further below.

The sensor representation  $M_\Delta$  is not required for retrieval, only the memories  $M_\Sigma$ ,  $M_\Pi$ , and  $M_\Gamma$  are recursively iterated. While co-activation of  $M_\Sigma$  and  $M_\Delta$  during look-ahead learning is implemented as a *logical and* operation, retrieval requires to toggle this modality to *logical or* to allow  $M_\Sigma$  to drive  $M_\Gamma$  by itself. Recursion is performed until the neuron which corresponds to the target symbol is activated in  $M_\Sigma$ , or until a maximal number of recursive invocations is reached. Furthermore, querying the existence of a trajectory performs retrieval across all scales concurrently. For instance, assume that a transition between two locations was learned in the largest scale. Certainly this means that a trajectory is known which links the two locations on the smallest scale across several intermediate places. However, these small-scale details of the trajectory are not necessary to merely determine the existence of a viable path, or can be retrieved while the animal is already walking in the general direction of the goal. Thus, a large scale transition encoder activates all corresponding distal symbols concurrently to any other symbols that are activated due to smaller scales of transitions.

### C. Simulation results

The deterministic model was evaluated on an S-shaped trajectory in a square environment, depicted in Figure 6a. To observe the impact of temporal buffering, two slightly different versions of the model were examined. In the first variant, temporal buffering was ignored and random access to all symbols was granted during learning. For instance, learning a feasible spatial transition only depended on the spatial vicinity between locations. The second variant required that previous locations were in the temporal buffer in a suitable time window. The S-shaped trajectory was designed such that any effects of the temporal buffer are recognizable when using at least one additional transition scale, illustrated in Figure 6a-c.



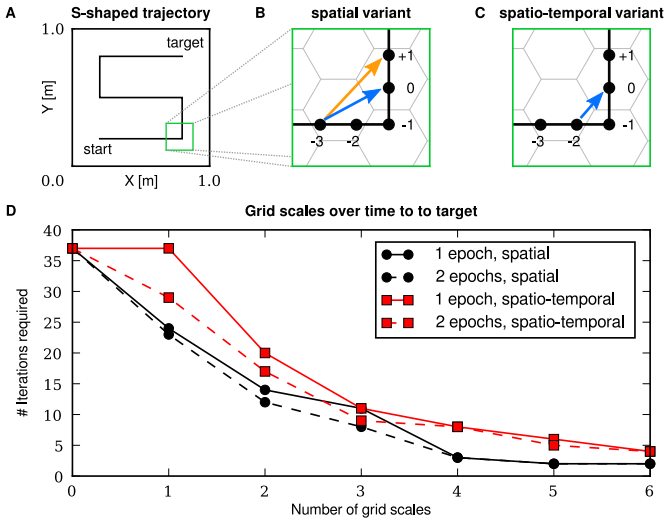


Figure 6: Expected behavior and results of the scale-space model. **(a)** The S-shaped trajectory which was used during evaluation. **(b + c)** The zoom in on one of the corners of the trajectory shows the difference in behavior between the spatial and spatio-temporal variant when one additional transition scale is used. Black dots indicate place field centers, the numerical value closest to each dot the relative time of perception with negative values going to the past. In the spatial variant, transition learning is only gated by spatial vicinity. During the initial exploration, a transition is thus learned from place cell at time  $-3$  to  $0$  (blue arrow). During the second exploration, a transition from  $-3$  to  $+1$  is learned due to prefetching (orange arrow). In the spatio-temporal variant, learning is restricted to items in a suitable temporal buffer. Consequently, only the transition from the place at  $-2$  to  $0$  is learned. The hexagonal Voronoi cells only depict the on-center of transition encoders. **(d)** Using multiple scales of grid cells improves the run-time, measured in number of iterations, exponentially to determine if a viable path from start to target exists.

Each variant was evaluated after only one trial during which learning was disabled. Furthermore, each variant was examined in a second trial with learning enabled during which the temporal buffer pre-fetched future symbols from  $M_\Sigma$  that were acquired during the first trial. After the second trial, learning was disabled again and replay of the trajectory was recorded.

Results are depicted in Figure 6d, which shows the number of recursive iterations required until the target location was found for increasing numbers of additional scales. The results show that the number of necessary iterations declines exponentially with an increased number of scales, thus improving the run-time of the model. In addition, the impact of temporal buffering is discernible, i.e. the spatio-temporal model performs slightly worse than the purely spatial model. However, this is considered to be an artifact of the simplified model of GCs that was used. Furthermore, mental travel during the second exploration of the environment improves the runtime of the model. The reason is prefetching of future locations in the temporal buffer structure combined with learning.

#### IV. A FORMAL THEORY FOR MULTI-TRANSITION SYSTEMS

Spatial navigation can be viewed as a process in which sequences of symbols are learned and produced. For this purpose, knowledge about feasible transitions from one symbol to another is required. Goal-directed navigation is therefore examined using the following axiomatic system, called Multi-Transition Theory (MTT), in which spatial locations are represented by symbols and movement from one location to another by transitions.

##### A. Symbols, alphabets, and sequences

Consider an animal which moves through three rooms. The trajectory of the animal can be described by the sequence of symbols  $A, B, C$ , e.g. each representing one room. The meaning of a symbol is not pre-determined, for instance the symbols could also represent the *event of perception* of each corresponding room. The entirety of symbols forms an alphabet and their consecutive ordering a sequence, both of which are captured in the following definition.

**Definition 1** (Alphabet and sequence). *An alphabet  $\Sigma$  is a finite set of symbols. A sequence (or word) is an ordered tuple of symbols  $\sigma_i \in \Sigma$ , i.e.  $(\sigma_0, \sigma_1, \dots) = w \in \Sigma^+$ , where  $^+$  is the Kleene plus operator.*

A trajectory of an animal is thus viewed as a sequence, and moves along several symbols. However, repetition of a single symbol is disallowed.

**Axiom 1** (Non-stationarity). *A sequence is non-stationary if any two successive symbols  $\sigma_i$  and  $\sigma_{i+1}$  are distinguishable, i.e.  $\sigma_i \neq \sigma_{i+1}$ .*

This does not limit general capabilities. Two consecutive but distinct symbols of a sequence can have the same associated meaning, for instance the perception of a certain room.

The directional ordering of a sequence is expressed using the arrow notation  $\rightarrow$ . For example,  $A \rightarrow B$  means that the symbol  $B$  causally follows after symbol  $A$ . However, *time* is not immediately given in the definition and needs to be stated explicitly. Thus, symbols and transitions can be viewed as terms of propositional logic. For instance,  $A \rightarrow B$  means that if  $A$  is *true*, it follows that  $B$  is also true. Thereby they form a *chain of causality*. In addition to  $\rightarrow$ , the arrow  $\rightsquigarrow$  exists, e.g.  $A \rightsquigarrow C$  means that there exists a *path* from  $A$  to  $C$  which bridges  $n \geq 0$  intermediate symbols. The negations of the notation are  $\nrightarrow$  and  $\nrightsquigarrow$ .

**Axiom 2** (Coherency). *Let  $w = \sigma_i, i \in \{0, \dots, N\}$  be a sequence of  $N$  symbols.  $w$  is coherent if and only if  $\sigma_i \rightarrow \sigma_{i+1}, \forall i \leq N - 1$ .*

Coherency is required in goal-direction navigation. Consider an animal which travels from a starting point to a target location. The animal has to construct a trajectory without any gaps to reach its destination. Otherwise, it may express undefined behavior or displacement activity as it does not know how to proceed to its goal. Without a distinct goal, the animal performs explorative movement in which novel symbols are acquired.

**Axiom 3** (Validity). A sequence  $w$  is valid or acceptable if it is both non-stationary and coherent.

Using these notations and axioms, goal-directed navigation from a start  $A$  to a goal  $C$  can be expressed as a *program* which expands the path  $A \rightsquigarrow C$  into any valid sequence  $A \rightarrow \sigma_1 \rightarrow \dots \rightarrow \sigma_{N-1} \rightarrow C$ , if it exists. The next section will discuss how the expansion can be denoted and performed for arbitrary symbols.

### B. On Universal Multi-Transition Systems

The arrow notation specifies relations between symbols. Consider the example  $A \rightarrow B$  which contains the transition from  $A$  to  $B$ . The transition is a tuple  $(A, B)$  which maps one symbol to another and known, for instance, from Reinforcement Learning (RL). There it is denoted as a transition function mapping from a set of states and a set of actions  $R$  to the next state, i.e.  $\tau : \Sigma \times R \rightarrow \Sigma$  [120]. The notation is now extended for the purpose of encoding multiple feasible transitions.

**Definition 2** (Transition system, set, bundle, and point). A *Multi-Transition System (MTS)*  $\mathcal{M}$  is the pair

$$\mathcal{M} = (\mathcal{P}(\Sigma), \Pi) \quad (12)$$

where  $\mathcal{P}(\Sigma)$  is the power set of  $\Sigma$ . A set  $\Omega \in \mathcal{P}(\Sigma)$  is called a configuration of  $\mathcal{M}$ . All symbols  $\sigma_i \in \Omega$  are considered to be true.

The set  $\Pi$  is called transition set and contains sets  $\pi_i$ , called transition bundles. In turn, a transition bundle  $\pi_i$  is a set of transitions  $\tau_i^k : \Sigma \rightarrow \Sigma$ , called  $k$ -th transition point of  $\pi_i$ .

An MTS  $\mathcal{M}$  can be interpreted as a state machine in which multiple states can be active simultaneously. Indices will be dropped if they are clear from context. Figure 7a illustrates a sequence, transition points, and bundles. The following additional terminology and notation will be used.

- 1) A transition  $\tau$  from  $A \in \Sigma$  to  $B \in \Sigma$  can be written  $(A, B)$  or  $(A \rightarrow B)$ .
- 2)  $\tau = (A \rightarrow B)$  is defined for  $A$  and leads to  $B$ , written  $A \prec \tau$  and  $\tau \succ B$ , respectively. The notation is transitive to bundles and sets, i.e.  $A \prec \pi \Leftrightarrow \exists \tau \in \pi, A \prec \tau$ , and  $\pi \succ B \Leftrightarrow \exists \tau \in \pi, \tau \succ B$ , respectively.
- 3) A bundle  $\pi$  forms a tuple  $(S, T)$  with start and target symbols  $S = \{\sigma | \sigma \prec \tau, \tau \in \pi\}$  and  $T = \{\sigma | \tau \succ \sigma, \tau \in \pi\}$ , respectively.
- 4) If a transition bundle  $\pi_i$  is true, then so are all contained transitions  $\tau_i^k \in \pi_i$ .

Transitions form terms which are independent of symbols. Consider the symbol  $A$  and the transition  $(A \rightarrow B)$ , both of which are propositional terms, in the term  $A \wedge (A \rightarrow B)$ . If  $A$  is true, then it can be deduced logically that  $B$  is also true, written  $A \wedge (A \rightarrow B) \Rightarrow B$ . Here,  $\wedge$  is the logical and operator.  $A$  forms a *precondition* for the transition  $(A \rightarrow B)$ . As  $A$  is true, the precondition is met and thus also the transition is true.  $B$  is the *conclusion* of the entire term. Order of evaluation is not specified during logical deduction. Therefore, sequential evaluation of transitions is made explicit as follows.

**Definition 3** (Transition evaluation). A configuration  $\Omega \in \mathcal{P}(\Sigma)$  of an MTS  $\mathcal{M}$  is evaluated according to the functions

$$F_{\mathcal{M}} : \Omega, \Pi \mapsto \cup_i f_{\mathcal{M}}(\Omega, \pi_i \in \Pi) \quad (13)$$

$$f_{\mathcal{M}} : \Omega, \pi \mapsto \{\sigma_l | \sigma_k \in \Omega, \sigma_k \prec \pi, \sigma_l \text{ is true in } \pi\}. \quad (14)$$

The union in Equation 13 goes over all transition bundles contained in  $\Pi$ . Hence, evaluation yields all symbols which are true given an initial configuration. The function  $F_{\mathcal{M}}(\Omega, \Pi)$  allows recursive usage for evaluation until a target symbol is reached. Consider the following example with four symbols  $A, B, C, D$ , where  $A$  is the start and  $B$  the target. Hence the initial configuration is  $\Omega = \{A\}$ . Furthermore, the following transition set, bundles, and points are defined.

$$\Pi = \{\pi_0 = \{\tau_0^0, \tau_0^1\}, \pi_1 = \{\tau_1^0, \tau_1^1\}\} \quad (15)$$

$$\tau_0^0 = (A, B) \quad \tau_0^1 = (A, D) \quad (16)$$

$$\tau_1^0 = (B, C) \quad \tau_1^1 = (D, C) \quad (17)$$

It requires two recursive evaluations, i.e.  $F_{\mathcal{M}}(F_{\mathcal{M}}(\Omega, \Pi), \Pi)$ , until the target symbol is found. This is similar to the evaluation of transition functions in RL [120]. However, the set notation allows the superposition of several symbols at the same time.

### C. Encoding capacity of the Universal MTS

The definition of  $\pi$  introduced a *bundling trick* which provides several benefits to analyze the computational logic and storage requirements of an MTS, especially when viewed in the light of neural encodings. Consider the following thought-experiment. Suppose that the generation of a bundle (e.g. a neuron) is energetically expensive, however the addition of a transition point (e.g. a dendritic spine) to an existing bundle is comparably cheap. To avoid evolutionary pressure [96], the goal is thus to minimize the overall cost. Optimizing this cost corresponds to maximizing the number of transition points while minimizing the number of bundles. As will be shown now, it is not possible to merge arbitrary transition points in one bundle without violating the constraints of MTT.

**Theorem 1.** Let  $\sigma \in \Sigma$ ,  $\mathcal{M}$  an MTS on the alphabet  $\Sigma$ ,  $\Pi$  the corresponding transition set, and  $\pi = (S, T)$  a transition bundle.  $\mathcal{M}$  generates valid sequences if and only if the following conditions hold.

- 1)  $\sigma_k \prec \pi \Rightarrow \pi_k \not\prec \sigma_i$ .
- 2)  $\pi \succ \sigma_l \Rightarrow \sigma_l \not\prec \pi$ .

*Proof.* 1) From Axiom 1 it follows immediately that any transition  $\pi$  which is defined for  $\sigma_k$  and leads to  $\sigma_k$  violates the non-stationarity condition. 2) Without loss of generality, consider the three symbols  $\sigma_0, \sigma_1, \sigma_2 \in \Sigma$  and  $\sigma_0 \rightarrow \sigma_1 \rightarrow \sigma_2$  but  $\sigma_0 \not\rightarrow \sigma_2$ . This yields the transition points  $\tau_0 = (\sigma_0, \sigma_1)$  and  $\tau_1 = (\sigma_1, \sigma_2)$ . Assume further that  $\tau_0$  and  $\tau_1$  are bundled in  $\pi$ , and that  $\sigma_0$  and  $\pi$  are true. It follows that  $\sigma_0 \wedge \tau_0 \Rightarrow \sigma_1$ . However,  $\sigma_1 \wedge \tau_1 \Rightarrow \sigma_2$  and thus  $\sigma_0 \wedge \pi \Rightarrow \sigma_2$ . This contradicts the assumption and violates the coherency constraint.  $\square$

Therefore, the sets of symbols  $S, T$  for a  $\pi = (S, T)$  have to be mutually exclusive, i.e.  $S \cap T = \emptyset$ .

**Definition 4** (Minimality, universality). An MTS  $\mathcal{M}$  is minimal if there exists only one  $\pi_i$  for any  $\sigma_k$ , i.e.  $\sigma_k \prec \pi_i \Rightarrow \sigma_k \not\prec \pi_j$

for any  $j \neq i$ . In a universal  $\mathcal{M}$ , any arbitrary transition between two symbols  $\sigma_k, \sigma_l$  is possible.

**Corollary 1.** The input set  $S_i$  of a transition bundle  $\pi_i$  is singleton for a minimal and universal  $\mathcal{M}$ .

*Proof.*  $\sigma_k < \pi_i$  and  $\pi_i > \sigma_l, \forall l \neq k$ . According to **Theorem 1**,  $\sigma_l \not< \pi_i, \forall l \neq k$ .  $\square$

**Corollary 2.** Let  $\Sigma$  be an alphabet of size  $M$ ,  $\Pi$  a transition set of  $N$  transition bundles  $\pi_i = \{S_i, T_i\}$  for a minimal universal  $\mathcal{M}$ . When all possible transitions are realized, then  $M = N$ .

*Proof.* Construct the graph  $\mathcal{G}$  of  $\mathcal{M}$  in which each transition is represented by a node, and any symbol  $\sigma_0, \sigma_1, \sigma_2$  which are connected by the transitions  $\tau_0 = (\sigma_0, \sigma_1)$  and  $\tau_1 = (\sigma_1, \sigma_2)$ . The corresponding graph has only two nodes for  $\tau_0$  and  $\tau_1$  which are connected by a single directed edge representing  $\sigma_1$ . The directed graph for a complete *universal minimal*  $\mathcal{M}$  in which all transitions are realized is fully connected. Reducing any pair of directed edges to a single undirected edge yields a fully connected undirected graph. According to **Theorem 1**,  $S_i \cap T_i = \emptyset$  for any  $\pi_i$ . Therefore, only those transitions can be bundled which are not connected by an edge in  $\mathcal{G}$ . Furthermore, the number of independent nodes in  $\mathcal{G}$  is equivalent to the chromatic number of the graph, i.e. the minimal number of colors which can be assigned to nodes of a graph, known as graph-coloring problem [121]. The chromatic number of a fully-connected graph equals the number of nodes. Conclusively,  $M = N$ .  $\square$

An example of the reduced graph for a sequence of four symbols  $\sigma_1, \dots, \sigma_4$  is depicted in **Figure 7b**. The figure shows transitions as vertices, and edges correspond to symbols. Removing the directedness of the graph by fusing two symbols means that each edge is associated with two symbols.

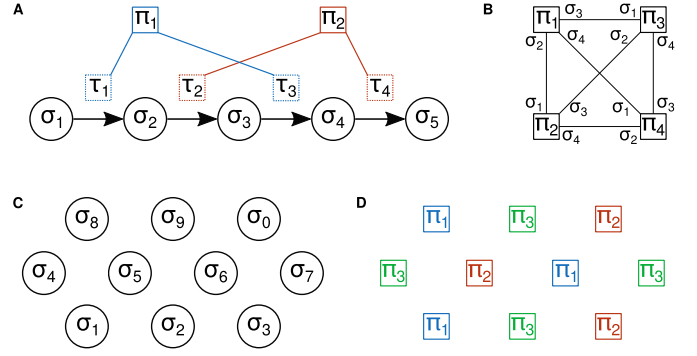
#### D. Brief discussion of the Universal MTS with implications for neural networks

Following **Corollary 2**, an implementation of a universal minimal  $\mathcal{M}$  requires to have as many entities to store transition bundles as it has symbols. Furthermore, these encoders have to decorrelate from their target symbols to fulfill **Theorem 1**.

Consider a neural implementation in which a neuron represents a transition bundle. To make efficient use of the neuron, it has to represent multiple transitions and thereby expose several receptive fields. However, each transition also dictates that the neuron has to dissociate from target symbols. Combined, each receptive field of such a transition neuron is suggested to consist of an on-region in which it associates to the symbol for which it is defined, and an off-region in which it decorrelates. In addition, the transition system requires the possibility of representing a *logical and* operation, feasible in neural networks [123]. Finally, each neural transition encoder will co-activate with any symbol for which it is defined during recursive retrieval.

#### E. Multi-Transition Systems in Euclidean space

The space which is constructed by symbols  $\delta_i$  and transitions  $\tau_j$  above is the discrete topological space with the induced



**Figure 7: Transition graph examples. (a)** A sequence of five consecutive symbols  $\sigma_1, \dots, \sigma_5$ . The transitions  $\tau_1, \dots, \tau_4$  can be bundled into  $\pi_1$  and  $\pi_2$  without violating any constraints. **(b)** In a universal transition system, the transition graph is fully connected. By construction, each edge is associated with two symbols. **(c)** In the two dimensional case, the optimal distribution of symbols is a hexagonal arrangement [122]. **(d)** The optimal distribution of transition bundles in the two dimensional case follows the distribution of the symbols. However, bundles can be repeated periodically in a hexagonal fashion. Edges in the graphs of (c) and (d) were omitted to improve clarity.

discrete metric. However, the world in which animals reside is not discrete and arbitrary jumps between two locations are infeasible. In particular, the perceived environment corresponds to a complete metric space, i.e. the Euclidean, from now on simply called *metric space*. Hence, an MTS  $\mathcal{L}$  which encodes transitions in a metric space has different constraints than a universal MTS  $\mathcal{M}$ .

Encoding transitions between locations in a metric space depends on the detection of two consecutive positions. The following analysis is based on the assumption that there exists a continuous signal which depends on and uniquely identifies each possible location of the animal. In terms of the Euclidean space  $M$ , this corresponds to locations  $\mathbf{x} \in M$ . Certainly an animal does not have access to coordinates. However, other stimuli are likely to provide the necessary information. For instance, geometrical information combined with head direction signals is sufficient to represent singular locations, which was demonstrated in the BV cell model presented by [18].

According to **Definition 1**, an alphabet is finite. This can be understood to correspond to a finite number of neurons which have to represent locations. However, the alphabet  $\Delta$  of *spatial symbols*  $\delta_i$  has to represent the continuous signals  $\mathbf{x}$  of the input space  $M$ . This corresponds to the well-known *sampling theorem*.

**Definition 5** (Spatial symbol, enablement, and assignment). Let  $\delta_i \in \Delta$  be spatial symbols according to a sampling process of a complete metric space  $\mathcal{D} = (M, d)$ . Each  $\delta_i$  is centered at a  $\mathbf{x}_i \in M$ .

A point  $\mathbf{p} \in M$  enables  $\delta_i$  if it is within the support of  $\delta_i$  given by the open ball  $B_{i,s}$  of radius  $r_s$ , i.e.  $B_{i,s} = \{\mathbf{p} \in M | d(\mathbf{x}_i, \mathbf{p}) < r_s\}$ .

The point  $\mathbf{p}$  is assigned to the closest  $\delta_i$ , i.e.  $\delta_i$  for which



$d(\mathbf{x}_i, \mathbf{p})$  is minimal. Given two adjacent symbols  $\delta_i, \delta_j$ , then  $r_w = \|\|\mathbf{x}_i - \mathbf{x}_j\|\|/2$ , describing a ball  $B_{i,w}$  of radius  $r_w$ .

The definitions of enablement and assignment can be interpreted in the following way. The region in which a spatial symbol is enabled can be understood as its receptive field, and according to the definition, multiple spatial symbols can have overlapping receptive fields. In contrast, assignment identifies the closest symbol, for instance as a result of a winner-take-all mechanism.

According to the Petersen-Middleton theorem [122], the ideal sampling strategy for two-dimensional continuous signals and therefore placement of spatial symbols  $\delta_i$  is a hexagonal arrangement. The sampling process can also be understood as a solution to the problem of packing spheres with diameter  $r_w$  as densely as possible. The sphere packing problem also yields a hexagonal lattice in the two dimensional case [124, 125].

Assuming an ideal sampling process, the question remains about the optimal distribution of transition bundles.

**Theorem 2.** *Let  $\mathcal{D} = (M, d)$  be an Euclidean space. Let  $\mathcal{L} = (\mathcal{P}(\Delta), \Gamma)$  be a minimal transition system on  $\mathcal{D}$  such that the countably finite alphabet  $\Delta$  corresponds to the densest optimal covering with respect to  $r_w$ .*

- 1) *The number of transition bundles  $\gamma_i \in \Gamma$  is constant.*
- 2) *The occurrence of any transition bundle  $\gamma_i$  is periodic.*

The theorem is proved by its corresponding graph-coloring problem which was introduced above.

*Proof.* The densest arrangement of spatial symbols according to the Petersen-Middleton theorem is a hexagonal lattice [122]. Furthermore, transitions between symbols are only possible between adjacent symbols. Consequently, the corresponding transition graph is not complete, i.e. only neighboring transitions are connected. The chromatic number of the resulting graph is 3 and the occurrence of colors is periodic.  $\square$

An example for the two dimensional arrangement of symbols is depicted in Figure 7c, and one solution of the graph coloring problem in Figure 7d.

As mentioned above, likely candidates to provide an input space which yields unique signatures for arbitrary locations in confined environments are BV and HD cells. However, an analysis which uses biologically plausible afferents is postponed to future work.

## V. DISCUSSION

A biologically plausible self-organizing model of GCs, proposed to encode transitions, was presented. Hexagonal grid fields emerged due to recurrent dynamics and to optimally encode transitions. Subsequently, a scale-space model for GCs was presented to remedy the behaviorally problematic runtime of spatial transition systems during path planning. Finally, MTT was introduced to mathematically examine the storage requirements both for episodic as well as spatial transitions.

Learning sequences and transitions in the Hippocampus was explored previously [126, 127]. However, these studies ignored spatial information or GCs. Spatially modulated inputs and association of motor commands and rewards to spatial

transitions were already suggested by [110, 111]. In their model, places and transitions are stored separately. In addition, their model contains biologically plausible interactions between the motor cortex, and was demonstrated in a real-world example using a robot. Hirel et al. extended the model to incorporate reinforcement learning for the acquisition of goal-directed transitions in a biologically plausible manner [112]. However, these models did not differentiate between spatial and temporal transition systems. Furthermore, sequences were not defined rigorously and optimality of transition encoding was not subject of the studies.

### A. On dendritic computation

The presented work used dendritic computation for the self-organization of GCs. A similar model was suggested previously by [128]. However, the authors proposed that GCs perform sampling of a circular input space for the purpose of localization. Thereby, the cells perform Voronoi clustering, which yields a hexagonal arrangement of grid fields in the ideal case. In addition, the self-organizing grid cell model is similar to the work by [129] and [128] in that it suggests that GCs of one module properly align to each other due to recurrent dynamics.

Recent findings suggest that dendritic computation in general is more elaborate than previously thought. Dendritic spines were found to express individual structural plasticity [130], as well as local synaptic plasticity [131, 132, 133]. Furthermore dendrites were found to be capable to encode multiple sensory stimuli [134]. In combination with the complex intrinsic organization of the EC [67, 135], it therefore appears likely that GCs perform multiple distinct computations in, and self-organization of their dendritic tree akin to the proposed multiple receptive fields for transition encoding.

### B. Biological interpretation of MTT

For arbitrary sequences and transitions from one symbol to another, the mathematical proof shows that as many transition encoders are required as there are symbols. In Euclidean metric space, the result differs in that only a finite number of transition encoders is required given the assumption of unique spatial input. Widloski et al. recently proposed a model of GC learning which exhibits similarity with respect to the expected behavior of the cells [136]. The authors demonstrated that a spiking neural network converges to a hexagonal arrangement of responses given spatially unique inputs.

Retrieval of sequences from an MTS requires both symbols and transition encoders. Neurons which represent transitions between spatial symbols will thus co-activate with and express similar place field behavior as the neurons encoding the symbols. Likely candidates for the storage of symbols and transitions appear to be PCs of CA1 and CA3. However, it is unclear if episodic transitions are stored within recurrent collaterals, plenty of which were found for CA3 [73, 137], or if there exists a suitable recurrence from CA1 to CA3. Another possibility is an embedding in the trisynaptic loop, which spans CA1, CA3, and EC. For the first case, disruption of activity

within CA1 or CA3 should lead to a decline of performance in tasks in which temporally accurate sequences are required.

It is expected that the deliberately abstract MTT can be applied to other domains which require transitions between symbols.

### C. A note on spatially unique input

All three introduced aspects, i.e. MTT, the self-organizing GC model, as well as the scale-space model, were developed under the assumption of spatially unique sensory inputs. Although the models used coordinates, it is proposed that the BV space forms a suitable and biologically plausible sensory representation. This is exemplified by the BV model by [18], who demonstrated that afferents from BV cells are sufficient to plausibly generate and correctly predict the firing of PCs with respect to the environment. Consequently, it is assumed that the coordinates can be replaced by a similar presynaptic sensory state. In fact, it is expected that coupling GCs and BV cells in the scale-space model will be able to explain recent findings in which GCs appeared to be influenced by the geometrical layout of the environment [138, 139].

### D. Remarks on the self-organizing model of grid cells

The self-organizing grid cell model quickly develops cells with a gridness score above zero for realistic trajectories of rodents. Furthermore, the dynamics align the orientation of all participating cells of one grid module. Competitive dynamics were required to achieve co-orientation of several cells, which is supported by Couey et al. who found that recurrent connectivity in the mEC is primarily inhibitory [84].

The model uses an animal's speed to adapt learning of spatial transitions. In particular, the rule defined in Equation 10 modulates the learning rate inversely proportional to the running speed of the animal. Hence, knowledge about the linear speed of the animal is required, which was recently discovered to be represented by speed cells in the mEC [140]. The effect is expected to be realized in either of the two ways. On the one hand, inhibitory interneurons could suppress the activity of the currently spiking GCs more strongly the faster the animal runs. On the other hand, a purely excitatory mechanism could be based on influencing GCs which sample future sensory states. Due to the findings of Couey et al. [84], the first mechanism appears to be more likely.

The presented model is independent of movement direction. On the contrary, the model depends on the instantaneous perception of the environment and identification of locations, similar to other previously proposed self-organizing models [88, 89, 128]. However, it is expected that activity of simulated GCs will slightly follow the head direction of an animal when the visual system is modeled accurately. There are two reasons for this expectation. First, the visual system of rodents is slightly forward facing, making it more likely that locations are recognized that are in the viewing direction. Second, one of the main purposes of the transition system is to plan trajectories towards goals. It appears to be more relevant to plan towards where an animal is facing, rather than where it is moving. In principle, this could explain a recent statistical analysis by [57],

who showed that GCs activity correlates more strongly with head direction than with movement direction.

The presented self-organizing model is limited due to negligence of temporal dynamics on the synaptic level. For instance, the receptive fields with on-center and off-surround dynamics are expected to be a result of temporal interactions between presynaptic afferents in combination with an asymmetric STDP learning rule. Furthermore, the model was only shown for three cells which are likely to organize themselves hexagonally and evenly distribute across the input. However, stochastic input in combination with a suitable STDP rule is expected to generate hexagonal arrangements and GCs with partially overlapping receptive fields. Finally, the programmatic global and local winner-take-all mechanisms are expected to also be remedied by a detailed model with accurate temporal dynamics.

### E. Interpretation of the scale-space model and related areas

The scale-space structure, used to improve run-times significantly, can be interpreted in the following way. The finest resolution of the spatial transition system reacts to transitions on the perceivable level, represented by presynaptic activity. In other words, detection of spatial locations during the dendritic computations corresponds to an *identity* function. The target region of any transition which is detected on this scale corresponds to any location with which an animal can interact directly. Larger scales allow the animal to assess viability of movement towards locations which are further apart. To allow comparison of such displaced locations, the corresponding sensory representations are low-pass filtered in discrete steps. In turn, these comparisons allow to perform approximate look-ahead. Look-ahead was already proposed to be performed in the entorhinal-hippocampal loop by Kubie et al. [141]. However, the authors did not address optimality of the encoding.

During learning, the simplified scale-space model requires co-activation of both spatially modulated afferents, as well as already acquired spatial symbols. The reason is that on the one hand, GCs are assumed to detect spatial transitions directly on the input space due to dendritic computations. On the other hand, these transitions must be related to already learned PCs. However, sensory information may not be available during retrieval, especially mental travel. Hence, retrieval operates only due to activated PCs, which requires to drive GC activity only on behalf of PCs. It is therefore expected that these operations are implemented hetero-synaptically in such a way that there is a clear distinction between learning and retrieval in the mEC.

The described recursive technique to construct multi-scale look-ahead transition encoders generates a Gaussian pyramid, or scale-space, well known in the computer vision and signal processing communities [142, 143, 144]. Among others, scale-spaces were used to describe biologically plausible retinal and visuocortical receptive fields [145, 146, 147], and it was proved that a scale increment of  $\sqrt{2}$  is optimal for Gaussian functions [144]. In image processing, application of a Gaussian pyramid corresponds to consecutively smoothing an image, i.e. an input gets low-pass filtered by which fine-scale information is removed. Finally, this allows to detect features in the

input across several scales [148, 149], thereby making feature detection scale invariant. The same principle is used in the scale-space model for GCs to search for feasible transitions on larger distances. Essentially, comparison of two distal locations for look-ahead requires to ignore fine structures.

#### F. Temporal buffering and a relation to Theta phase precession

Lindeberg points out that access to temporally buffered information is necessary to construct scale-space structures [150]. Certainly, this was also required for the construction of the scale-space for transition look-ahead. The temporal buffer was required to bind large-scale transition encoders to consecutive places across long distances. Hence, GCs of larger scales require excitation not only from the current sensory perception, but also from PCs which encode past and future locations on a linear track. This observation is supported by findings which report that excitatory drive from the Hippocampus is required for grid cells [54]. The observation predicts that the temporal integration window of GCs from larger scales needs to be increased when compared to cells from smaller scales. Or, that some other mechanism exists which prevents GCs of larger scales from firing before the necessary afferents from PCs were accumulated. Most importantly, it requires a biologically plausible mechanism to buffer data temporally.

The most likely candidate for a temporal buffer is Theta phase precession [34, 35]. When an animal is running, several PCs which correspond to past locations, the PCs of the current location, as well as PCs which encode future locations spike *in order* of their traversal within one Theta cycle. The reported maximal compression ratio is of the order of 10 : 1 [33], i.e. at most ten consecutive places are represented within one theta cycle. One iteration of the proposed transition system requires at least two neural memories. Given approximate numbers of neural activation and axonal delay of 5 – 10 ms, one iteration requires 15 ms on average. In concordance with the reported compression ratio, at most 10 iterations of the transition system fit into one Theta cycle which oscillates between 6–10 Hz [151]. Conclusively, Theta is proposed to form a main-loop with nested sub-loops that iterate the transition system. Without other mechanisms, it is estimated that on average 5-7 stable grid scales can be formed on top of this temporal buffer, depending on the frequency of Theta. It is expected that temporal compression during SPW-R contributes to the formation of grid scales, which may lead to larger scales.

#### G. Links to computer science and robotics

Skip lists [114], used as inspiration for the look-ahead in the scale-space model, operate on one-dimensional lists of data. Several related data structures which form hierarchies for two dimensional data are known in computer science. For instance, quad- and octrees subdivide dimensions of an input space to accelerate queries exponentially [152, 153]. On topological data without global coordinates, contraction hierarchies were shown to significantly improve search time [154, 155]. The scale-space model of GCs demonstrates how these data structures can be extended to probabilistic data.

An MTS creates a topological representation of perceived locations. Dabaghian et al. already suggested that the rodent Hippocampus forms a topological map [25, 156]. However, the novel scale-space model not only links immediately neighboring locations but allows to extract approximate relations between remote places. Retrieval of a path according to the scale-space model corresponds to the well known Dijkstra’s algorithm [157], or more specifically to A\* [158]. The latter algorithm was developed for robotic navigation and is an extension of the first. A\* uses heuristic information to accelerate search of viable trajectories, similarly to look-ahead transitions in the scale-space model. The benefit of a topological map, formed on spatially modulated sensor data and augmented with look-ahead, is the avoidance of a global coordinate system. It is thus expected to be useful in indoor robotics systems without access to global coordinates.

#### H. Future work

Several simplifications and abstractions were necessary to reduce the complexity of the introduced models. An implementation which is currently in development will use a spiking neural network to address the issue of missing temporal dynamics, absent in the self-organizing grid cell model. Early results show that on-center and off-surround fields for dendritic computation form as expected, and that several cells exhibit partially overlapping grid-like fields.

Furthermore, an extended scale-space transition model is in development which includes RL as a mechanism for trajectory selection. Currently, the scale-space transition model lacks a quality measure to narrow down multiple possible trajectories to a singular best fit and only reports the existence of any viable path. It is expected that the previous work by [112] can be adapted and reproduced. Subsequently, the RL mechanism will be combined with the spiking neuron model. It is expected that the observation of pre-play of trajectories, reported in [31], can be reproduced in the combined model.

## VI. CONCLUSION

This work proposed an entirely novel function of grid cells, i.e. optimal encoding of transitions across multiple scales in form of a multi-transition system. First a biologically plausible model of grid cells in a competitive network was developed. Then, the spatial multi-transition system was extended into a scale-space model for transition encoding to solve behaviorally problematic run-times. Given the assumption of probabilistic detection of locations, it was shown that the optimal scale increment for transition encoding in the scale-space model is  $\sqrt{2}$ . Finally, Multi-Transition Theory was introduced to formally analyze transition systems. It was shown that a hexagonal arrangement of transition encoders minimizes the number of required encoders in two-dimensional Euclidean space. The results give rise to future developments. Extended models with spiking dynamics are in development to address biologically plausible reinforcement learning for trajectory selection and input space representation. Furthermore, the results are currently used for the development of novel mobile



swarm-robotic systems which operate on and communicate with probabilistic data similar to the principles of neural systems.

Conclusively, this work addressed spatial navigation on all three levels of analysis suggested by David Marr [159]. 1) The purpose of the computation, which is twofold. On the one hand an animal requires to record locations. On the other hand, the recorded locations have to be recalled in such a way that trajectory planning can be achieved in a behaviorally acceptable run-time. 2) The algorithmic description and especially the scale-space model were a result of the computational purpose and its behavioral relevance. 3) A biologically plausible implementation for GCs as well as the scale-space multi-transition system were demonstrated.

## VII. ACKNOWLEDGMENTS

The author would like to thank Christoph Richter and Jörg Conradt for invaluable discussions and feedback during the research, as well as their support and suggestions on the manuscript. This work was partially funded by EU FET project GRIDMAP 600725.

## VIII. COMPETING INTERESTS

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## REFERENCES

- [1] E. I. Moser, E. Kropff, and M. B. Moser, "Place cells, grid cells, and the brain's spatial representation system," *Annu. Rev. Neurosci.*, vol. 31, pp. 69–89, 2008.
- [2] R. M. Grieves and K. J. Jeffery, "The representation of space in the brain," *Behavioural Processes*, vol. 135, pp. 113 – 131, 2017.
- [3] M. B. Moser, D. C. Rowland, and E. I. Moser, "Place cells, grid cells, and memory," *Cold Spring Harb Perspect Biol*, vol. 7, p. a021808, Feb 2015.
- [4] E. C. Tolman, "Cognitive maps in rats and men," *Psychological Review*, vol. 55, no. 4, pp. 189–208, 1948.
- [5] L. L. Chen, L.-H. Lin, E. J. Green, C. A. Barnes, and B. L. McNaughton, "Head-direction cells in the rat posterior cortex," *Experimental Brain Research*, vol. 101, no. 1, pp. 8–23, 1994.
- [6] J. B. Ranck, "Head-direction cells in the deep cell layers of dorsal presubiculum in freely moving rats," *Society for Neuroscience Abstracts*, vol. 10, 1984.
- [7] P. Preston-Ferrer, S. Coletta, M. Frey, and A. Burgalossi, "Anatomical organization of presubicular head-direction circuits," *eLife*, vol. 5, p. e14592, jun 2016.
- [8] J. Taube, R. Muller, and J. Ranck, "Head-direction cells recorded from the postsubiculum in freely moving rats. i. description and quantitative analysis," *Journal of Neuroscience*, vol. 10, no. 2, pp. 420–435, 1990.
- [9] J. Taube, R. Muller, and J. Ranck, "Head-direction cells recorded from the postsubiculum in freely moving rats. ii. effects of environmental manipulations," *Journal of Neuroscience*, vol. 10, no. 2, pp. 436–447, 1990.
- [10] J. S. Taube, "The head direction signal: origins and sensory-motor integration," *Annu. Rev. Neurosci.*, vol. 30, pp. 181–207, 2007.
- [11] J. O'Keefe and J. Dostrovsky, "The hippocampus as a spatial map. preliminary evidence from unit activity in the freely-moving rat," *Brain Research*, vol. 34, no. 1, pp. 171 – 175, 1971.
- [12] J. O'Keefe, "A review of the hippocampal place cells," *Prog. Neurobiol.*, vol. 13, no. 4, pp. 419–439, 1979.
- [13] A. A. Fenton, H. Y. Kao, S. A. Neymotin, A. Olypher, Y. Vayntrub, W. W. Lytton, and N. Ludvig, "Unmasking the CA1 ensemble place code by exposures to small and large environments: more place cells and multiple, irregularly arranged, and expanded place fields in the larger space," *J. Neurosci.*, vol. 28, pp. 11250–11262, Oct 2008.
- [14] E. Park, D. Dvorak, and A. A. Fenton, "Ensemble place codes in hippocampus: Ca1, ca3, and dentate gyrus place cells have multiple place fields in large environments," *PLOS ONE*, vol. 6, pp. 1–9, 07 2011.
- [15] I. Lee, D. Yoganarasimha, G. Rao, and J. J. Knierim, "Comparison of population coherence of place cells in hippocampal subfields ca1 and ca3," *Nature*, vol. 430, pp. 456–459, Jul 2004.
- [16] A. Treves, O. Miglino, and D. Parisi, "Rats, nets, maps, and the emergence of place cells," *Psychobiology*, vol. 20, no. 1, pp. 1–8, 1992.
- [17] P. E. Sharp, "Computer simulation of hippocampal place cells," *Psychobiology*, vol. 19, no. 2, pp. 103–115, 1991.
- [18] C. Barry, C. Lever, R. Hayman, T. Hartley, S. Burton, J. O'Keefe, K. Jeffery, and N. Burgess, "The boundary vector cell model of place cell firing and spatial memory," *Rev Neurosci*, vol. 17, no. 1-2, pp. 71–97, 2006.
- [19] E. Save, A. Cressant, C. Thinus-Blanc, and B. Poucet, "Spatial firing of hippocampal place cells in blind rats," *J. Neurosci.*, vol. 18, pp. 1818–1826, Mar 1998.
- [20] S. Zhang, F. Schonfeld, L. Wiskott, and D. Manahan-Vaughan, "Spatial representations of place cells in darkness are supported by path integration and border information," *Front Behav Neurosci*, vol. 8, p. 222, 2014.
- [21] L. T. Thompson and P. J. Best, "Long-term stability of the place-field activity of single units recorded from the dorsal hippocampus of freely behaving rats," *Brain Res.*, vol. 509, pp. 299–308, Feb 1990.
- [22] R. G. Morris, P. Garrud, J. N. Rawlins, and J. O'Keefe, "Place navigation impaired in rats with hippocampal lesions," *Nature*, vol. 297, pp. 681–683, Jun 1982.
- [23] A. Gorchetchnikov and S. Grossberg, "Space, time and learning in the hippocampus: how fine spatial and temporal scales are expanded into population codes for behavioral control," *Neural Netw*, vol. 20, pp. 182–193, Mar 2007.
- [24] T. J. Wills, L. Muessig, and F. Cacucci, "The development of spatial behaviour and the hippocampal neural representation of space," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 369, pp. 20130409–20130409, dec 2013.
- [25] Y. Dabaghian, V. L. Brandt, and L. M. Frank, "Recon-

- ceiving the hippocampal map as a topological template,” *eLife*, vol. 3, p. e03476, aug 2014.
- [26] A. M. Wikenheiser and A. D. Redish, “Hippocampal theta sequences reflect current goals,” *Nat Neurosci*, vol. 18, pp. 289–294, Feb 2015. Article.
- [27] J. Csicsvari and D. Dupret, “Sharp wave/ripple network oscillations and learning-associated hippocampal maps,” *Philos. Trans. R. Soc. Lond., B, Biol. Sci.*, vol. 369, p. 20120528, Feb 2014.
- [28] S. P. Jadhav, C. Kemere, P. W. German, and L. M. Frank, “Awake hippocampal sharp-wave ripples support spatial memory,” *Science*, vol. 336, no. 6087, pp. 1454–1458, 2012.
- [29] N. Maier, A. Draguhn, D. Schmitz, and M. Both, “Fast network oscillations in the hippocampus,” *e-Neuroforum*, vol. 4, no. 1, pp. 1–10, 2013.
- [30] G. Buzsáki, “Hippocampal sharp wave-ripple: A cognitive biomarker for episodic memory and planning,” *Hippocampus*, vol. 25, pp. 1073–1188, Oct 2015.
- [31] B. E. Pfeiffer and D. J. Foster, “Hippocampal place-cell sequences depict future paths to remembered goals,” *Nature*, vol. 497, pp. 74–79, May 2013.
- [32] G. Buzsáki, “Theta oscillations in the hippocampus,” *Neuron*, vol. 33, no. 3, pp. 325 – 340, 2002.
- [33] W. E. Skaggs, B. L. McNaughton, M. A. Wilson, and C. A. Barnes, “Theta phase precession in hippocampal neuronal populations and the compression of temporal sequences,” *Hippocampus*, vol. 6, no. 2, pp. 149–172, 1996.
- [34] A. Jeewajee, C. Barry, V. Douchamps, D. Manson, C. Lever, and N. Burgess, “Theta phase precession of grid and place cell firing in open environments,” *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, vol. 369, no. 1635, 2013.
- [35] M. E. Hasselmo, “What is the function of hippocampal theta rhythm?—Linking behavioral data to phasic properties of field potential and unit recording data,” *Hippocampus*, vol. 15, no. 7, pp. 936–949, 2005.
- [36] M. W. Jones and M. A. Wilson, “Theta rhythms coordinate hippocampal-prefrontal interactions in a spatial memory task,” *PLOS Biology*, vol. 3, 11 2005.
- [37] H. T. Ito, “Prefrontal–hippocampal interactions for spatial navigation,” *Neuroscience Research*, pp. –, 2017.
- [38] R. A. Koene and M. E. Hasselmo, “Reversed and forward buffering of behavioral spike sequences enables retrospective and prospective retrieval in hippocampal regions CA3 and CA1,” *Neural Netw*, vol. 21, no. 2-3, pp. 276–288, 2008.
- [39] K. Mizuseki, A. Sirota, E. Pastalkova, and G. Buzsáki, “Theta oscillations provide temporal windows for local circuit computation in the entorhinal-hippocampal loop,” *Neuron*, vol. 64, pp. 267–280, Oct 2009.
- [40] T. Hafting, M. Fyhn, S. Molden, M.-B. Moser, and E. I. Moser, “Microstructure of a spatial map in the entorhinal cortex,” *Nature*, vol. 436, pp. 801–806, Aug 2005.
- [41] M. Fyhn, T. Hafting, A. Treves, M.-B. Moser, and E. I. Moser, “Hippocampal remapping and grid realignment in entorhinal cortex,” *Nature*, vol. 446, pp. 190–194, Mar 2007.
- [42] N. Ulanovsky and C. F. Moss, “Hippocampal cellular and network activity in freely moving echolocating bats,” *Nat. Neurosci.*, vol. 10, pp. 224–233, Feb 2007.
- [43] M. M. Yartsev, M. P. Witter, and N. Ulanovsky, “Grid cells without theta oscillations in the entorhinal cortex of bats,” *Nature*, vol. 479, pp. 103–107, Nov 2011.
- [44] E. I. Moser and M. B. Moser, “A metric for space,” *Hippocampus*, vol. 18, no. 12, pp. 1142–1156, 2008.
- [45] E. I. Moser, Y. Roudi, M. P. Witter, C. Kentros, T. Bonhoeffer, and M.-B. Moser, “Grid cells and cortical representation,” *Nat Rev Neurosci*, vol. 15, pp. 466–481, Jul 2014. Review.
- [46] F. Sargolini, M. Fyhn, T. Hafting, B. L. McNaughton, M. P. Witter, M. B. Moser, and E. I. Moser, “Conjunctive representation of position, direction, and velocity in entorhinal cortex,” *Science*, vol. 312, pp. 758–762, May 2006.
- [47] H. Stensola, T. Stensola, T. Solstad, K. Froland, M.-B. Moser, and E. I. Moser, “The entorhinal grid map is discretized,” *Nature*, vol. 492, pp. 72–78, Dec. 2012.
- [48] F. Savelli, D. Yoganarasimha, and J. J. Knierim, “Influence of boundary removal on the spatial representations of the medial entorhinal cortex,” *Hippocampus*, vol. 18, no. 12, pp. 1270–1282, 2008.
- [49] J. Krupic, M. Bauza, S. Burton, C. Lever, and J. O’Keefe, “How environment geometry affects grid cell symmetry and what we can learn from it,” *Philos. Trans. R. Soc. Lond., B, Biol. Sci.*, vol. 369, p. 20130188, Feb 2014.
- [50] E. Marozzi and K. J. Jeffery, “Place, space and memory cells,” *Current Biology*, vol. 22, no. 22, pp. R939 – R942, 2012.
- [51] J. A. Pérez-Escobar, O. Kornienko, P. Latuske, L. Kohler, and K. Allen, “Visual landmarks sharpen grid cell metric and confer context specificity to neurons of the medial entorhinal cortex,” *eLife*, vol. 5, p. e16937, jul 2016.
- [52] D. Derdikman, J. R. Whitlock, A. Tsao, M. Fyhn, T. Hafting, M.-B. Moser, and E. I. Moser, “Fragmentation of grid cell maps in a multicompartment environment,” *Nat Neurosci*, vol. 12, pp. 1325–1332, Oct 2009.
- [53] T. Stensola, H. Stensola, M.-B. Moser, and E. I. Moser, “Shearing-induced asymmetry in entorhinal grid cells,” *Nature*, vol. 518, pp. 207–212, Feb 2015. Article.
- [54] T. Bonnevie, B. Dunn, M. Fyhn, T. Hafting, D. Derdikman, J. L. Kubie, Y. Roudi, E. I. Moser, and M.-B. Moser, “Grid cells require excitatory drive from the hippocampus,” *Nat Neurosci*, vol. 16, pp. 309–317, Mar 2013.
- [55] C. Schmidt-Hieber and M. Hausser, “How to build a grid cell,” *Philos. Trans. R. Soc. Lond., B, Biol. Sci.*, vol. 369, p. 20120520, Feb 2014.
- [56] G. Chen, D. Manson, F. Cacucci, and T. J. Wills, “Absence of visual input results in the disruption of grid cell firing in the mouse,” *Current Biology*, vol. 26, no. 17, pp. 2335 – 2342, 2016.
- [57] F. Raudies, M. P. Brandon, G. W. Chapman, and M. E. Hasselmo, “Head direction is coded more strongly than movement direction in a population of entorhinal

- neurons,” *Brain Research*, vol. 1621, pp. 355 – 367, 2015. Brain and Memory: Old Arguments and New Perspectives.
- [58] R. Morris, “Developments of a water-maze procedure for studying spatial learning in the rat,” *J. Neurosci. Methods*, vol. 11, pp. 47–60, May 1984.
- [59] E. T. Rolls, “Functions of the primate hippocampus in spatial and nonspatial memory,” *Hippocampus*, vol. 1, no. 3, pp. 258–261, 1991.
- [60] W. B. Scoville and B. Milner, “Loss of recent memory after bilateral hippocampal lesions,” *J Neurol Neurosurg Psychiatry*, vol. 20, pp. 11–21, Feb 1957. 13406589[pmid].
- [61] D. Marr, “Simple memory: a theory for archicortex,” *Philos. Trans. R. Soc. Lond., B, Biol. Sci.*, vol. 262, pp. 23–81, Jul 1971.
- [62] E. Tulving, “Episodic and semantic memory 1,” *Organization of Memory. London: Academic*, vol. 381, no. 4, pp. 382–404, 1972.
- [63] L. E. Jarrard, “On the role of the hippocampus in learning and memory in the rat,” *Behav. Neural Biol.*, vol. 60, pp. 9–26, Jul 1993.
- [64] N. M. van Strien, N. L. Cappaert, and M. P. Witter, “The anatomy of memory: an interactive overview of the parahippocampal-hippocampal network,” *Nat. Rev. Neurosci.*, vol. 10, pp. 272–282, Apr 2009.
- [65] S. Cheng, “The crisp theory of hippocampal function in episodic memory,” *Frontiers in Neural Circuits*, vol. 7, p. 88, 2013.
- [66] M. P. Witter, “Intrinsic and extrinsic wiring of CA3: indications for connectional heterogeneity,” *Learn. Mem.*, vol. 14, pp. 705–713, Nov 2007.
- [67] C. B. Canto, F. G. Wouterlood, and M. P. Witter, “What does the anatomical organization of the entorhinal cortex tell us?,” *Neural Plasticity*, vol. 2008, pp. 1–18, 2008.
- [68] E. Fuchs, A. Neitz, R. Pinna, S. Melzer, A. Caputi, and H. Monyer, “Local and distant input controlling excitation in layer ii of the medial entorhinal cortex,” *Neuron*, vol. 89, pp. 194–208, 12 2015.
- [69] B. L. McNaughton and R. G. M. Morris, “Hippocampal synaptic enhancement and information storage within a distributed memory system,” *Trends in Neurosciences*, vol. 10, pp. 408–415, 2017/03/06 1987.
- [70] A. Treves and E. T. Rolls, “Computational analysis of the role of the hippocampus in memory,” *Hippocampus*, vol. 4, pp. 374–391, Jun 1994.
- [71] S. Káli and P. Dayan, “The involvement of recurrent connections in area ca3 in establishing the properties of place fields: a model,” *Journal of Neuroscience*, vol. 20, no. 19, pp. 7463–7477, 2000.
- [72] G. Papp, M. P. Witter, and A. Treves, “The CA3 network as a memory store for spatial representations,” *Learn. Mem.*, vol. 14, pp. 732–744, Nov 2007.
- [73] C. Le Duiou, J. Simonnet, M. T. Telenczuk, D. Fricker, and R. Miles, “Recurrent synapses and circuits in the CA3 region of the hippocampus: an associative network,” *Front Cell Neurosci*, vol. 7, p. 262, Jan 2014.
- [74] R. K. Mishra, S. Kim, S. J. Guzman, and P. Jonas, “Symmetric spike timing-dependent plasticity at ca3-ca3 synapses optimizes storage and recall in autoassociative networks,” *Nature Communications*, vol. 7, pp. 11552 EP –, May 2016. Article.
- [75] G. Palm, “On associative memory,” *Biological Cybernetics*, vol. 36, no. 1, pp. 19–31, 1980.
- [76] G. Palm, F. Schwenker, F. T. Sommer, and A. Strey, “Neural associative memories,” *Biological Cybernetics*, vol. 36, pp. 36–19, 1993.
- [77] A. Knoblauch, “Neural associative memory for brain modeling and information retrieval,” *Information Processing Letters*, vol. 95, no. 6, pp. 537 – 544, 2005. Applications of Spiking Neural Networks.
- [78] T. Wennekers and G. Palm, “Syntactic sequencing in hebbian cell assemblies,” *Cogn Neurodyn*, vol. 3, pp. 429–441, Dec 2009. 9095[PII].
- [79] L. M. Giocomo, M.-B. Moser, and E. I. Moser, “Computational models of grid cells,” *Neuron*, vol. 71, no. 4, pp. 589 – 603, 2011.
- [80] E. Zilli, “Models of grid cell spatial firing published 2005-2011,” *Frontiers in Neural Circuits*, vol. 6, p. 16, 2012.
- [81] O. Shipston-Sharman, L. Solanka, and M. F. Nolan, “Continuous attractor network models of grid cell firing based on excitatory-inhibitory interactions,” *J. Physiol. (Lond.)*, vol. 594, pp. 6547–6557, Nov 2016.
- [82] M. C. Fuhs and D. S. Touretzky, “A spin glass model of path integration in rat medial entorhinal cortex,” *J. Neurosci.*, vol. 26, pp. 4266–4276, Apr 2006.
- [83] Y. Burak and I. R. Fiete, “Accurate path integration in continuous attractor network models of grid cells,” *PLOS Computational Biology*, vol. 5, pp. 1–16, 02 2009.
- [84] J. J. Couey, A. Witoelar, S. J. Zhang, K. Zheng, J. Ye, B. Dunn, R. Czajkowski, M. B. Moser, E. I. Moser, Y. Roudi, and M. P. Witter, “Recurrent inhibitory circuitry as a mechanism for grid formation,” *Nat. Neurosci.*, vol. 16, pp. 318–324, Mar 2013.
- [85] K. Yoon, M. A. Buice, C. Barry, R. Hayman, N. Burgess, and I. R. Fiete, “Specific evidence of low-dimensional continuous attractor dynamics in grid cells,” *Nat Neurosci*, vol. 16, pp. 1077–1084, Aug 2013. Article.
- [86] N. Burgess, C. Barry, and J. O’Keefe, “An oscillatory interference model of grid cell firing,” *Hippocampus*, vol. 17, no. 9, pp. 801–812, 2007.
- [87] J. Koenig, A. N. Linder, J. K. Leutgeb, and S. Leutgeb, “The spatial periodicity of grid cells is not sustained during reduced theta oscillations,” *Science*, vol. 332, no. 6029, pp. 592–595, 2011.
- [88] E. Kropff and A. Treves, “The emergence of grid cells: Intelligent design or just adaptation,” *Hippocampus*, vol. 18, no. 12, pp. 1256–1269, 2008.
- [89] A. Stepanyuk, “Self-organization of grid fields under supervision of place cells in a neuron model with associative plasticity,” *Biologically Inspired Cognitive Architectures*, vol. 13, pp. 48 – 62, 2015.
- [90] M. Stemmler, A. Mathis, and A. V. M. Herz, “Connecting multiple spatial scales to decode the population activity of grid cells,” *Science Advances*, vol. 1, no. 11,



- 2015.
- [91] A. Mathis, A. V. M. Herz, and M. Stemmler, “Optimal population codes for space: Grid cells outperform place cells,” *Neural Computation*, vol. 24, pp. 2280–2317, May 2012.
- [92] T. Solstad, E. I. Moser, and G. T. Einevoll, “From grid cells to place cells: a mathematical model,” *Hippocampus*, vol. 16, no. 12, pp. 1026–1031, 2006.
- [93] X.-x. Wei, J. Prentice, and V. Balasubramanian, “A principle of economy predicts the functional architecture of grid cells,” *eLife*, 2015.
- [94] B. L. McNaughton, F. P. Battaglia, O. Jensen, E. I. Moser, and M.-B. Moser, “Path integration and the neural basis of the ‘cognitive map’,” *Nat Rev Neurosci*, vol. 7, pp. 663–678, Aug 2006.
- [95] S. Sreenivasan and I. Fiete, “Grid cells generate an analog error-correcting code for singularly precise neural computation,” *Nat Neurosci*, vol. 14, pp. 1330–1337, Oct 2011. Article.
- [96] J. E. Niven and S. B. Laughlin, “Energy limitation as a selective pressure on the evolution of sensory systems,” *J. Exp. Biol.*, vol. 211, pp. 1792–1804, Jun 2008.
- [97] H. Eichenbaum, P. Dudchenko, E. Wood, M. Shapiro, and H. Tanila, “The hippocampus, memory, and place cells: Is it spatial memory or a memory space?,” *Neuron*, vol. 23, no. 2, pp. 209 – 226, 1999.
- [98] M. Mulas, N. Waniek, and J. Conradt, “Hebbian plasticity realigns grid cell activity with external sensory cues in continuous attractor models,” *Front Comput Neurosci*, vol. 10, p. 13, Feb 2016. 26924979[pmid].
- [99] S. Grossberg and P. K. Pilly, “Coordinated learning of grid cell and place cell spatial and temporal properties: multiple scales, attention and oscillations,” *Philos. Trans. R. Soc. Lond., B, Biol. Sci.*, vol. 369, p. 20120524, Feb 2014.
- [100] D. M. Finch, N. L. Nowlin, and T. L. Babb, “Demonstration of axonal projections of neurons in the rat hippocampus and subiculum by intracellular injection of HRP,” *Brain Res.*, vol. 271, pp. 201–216, Jul 1983.
- [101] H. Eichenbaum, “Time cells in the hippocampus: a new dimension for mapping memories,” *Nat. Rev. Neurosci.*, vol. 15, pp. 732–744, Nov 2014.
- [102] B. J. Kraus, M. P. Brandon, R. J. I. Robinson, M. A. Connerney, M. E. Hasselmo, and H. Eichenbaum, “During running in place, grid cells integrate elapsed time and distance run,” *Neuron*, vol. 88, pp. 578–589, 2017/02/10 2015.
- [103] D. M. Salz, Z. Tiganj, S. Khasnabish, A. Kohley, D. Sheehan, M. W. Howard, and H. Eichenbaum, “Time Cells in Hippocampal Area CA3,” *J. Neurosci.*, vol. 36, pp. 7476–7484, Jul 2016.
- [104] C. A. Barnes, “Spatial learning and memory processes: the search for their neurobiological mechanisms in the rat,” *Trends Neurosci.*, vol. 11, pp. 163–169, Apr 1988.
- [105] M. Fyhn, T. Solstad, and T. Hafting, “Entorhinal grid cells and the neural basis of navigation,” in *Hippocampal Place Fields*, pp. 237–252, Oxford University Press (OUP), mar 2008.
- [106] L. Lamport, “Time, clocks, and the ordering of events in a distributed system,” *Commun. ACM*, vol. 21, pp. 558–565, July 1978.
- [107] J. Y. Halpern, “A modification of the halpern-pearl definition of causality,” *CoRR*, vol. abs/1505.00162, 2015.
- [108] C. A. R. Hoare, “Communicating sequential processes,” *Commun. ACM*, vol. 21, pp. 666–677, Aug. 1978.
- [109] S. Grossberg, “Embedding fields: A theory of learning with physiological implications,” *Journal of Mathematical Psychology*, vol. 6, pp. 209–239, jun 1969.
- [110] N. Cuperlier, P. Laroque, P. Gaussier, and M. Quoy, “Planning and navigation strategies using transition cells and neural fields,” in *Proc. of ASC/IASTED 2004 (Artificial intelligence and Soft Computing / International Association of Science and TEchnology for Development)*, (Marbella), sept 2004.
- [111] N. Cuperlier, M. Quoy, C. Giovannangeli, P. Gaussier, and P. Laroque, *Transition Cells for Navigation and Planning in an Unknown Environment*, pp. 286–297. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006.
- [112] J. Hirel, P. Gaussier, M. Quoy, and J.-P. Banquet, *Why and How Hippocampal Transition Cells Can Be Used in Reinforcement Learning*, pp. 359–369. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010.
- [113] D. E. Knuth, *The Art of Computer Programming, Volume 3: (2Nd Ed.) Sorting and Searching*. Redwood City, CA, USA: Addison Wesley Longman Publishing Co., Inc., 1998.
- [114] W. Pugh, “Skip lists: A probabilistic alternative to balanced trees,” *Commun. ACM*, vol. 33, pp. 668–676, June 1990.
- [115] B. C. Dean and Z. H. Jones, “Exploring the duality between skip lists and binary search trees,” in *Proceedings of the 45th Annual Southeast Regional Conference, ACM-SE 45*, (New York, NY, USA), pp. 395–399, ACM, 2007.
- [116] D. A. Butts and M. S. Goldman, “Tuning curves, neuronal variability, and sensory coding,” *PLoS Biol.*, vol. 4, p. e92, Apr 2006.
- [117] M. Jazayeri and J. A. Movshon, “Optimal representation of sensory information by neural populations,” *Nat Neurosci*, vol. 9, pp. 690–696, May 2006.
- [118] S. Vinga and J. S. Almeida, “Rényi continuous entropy of DNA sequences,” *J. Theor. Biol.*, vol. 231, pp. 377–388, Dec 2004.
- [119] B. Fritzsche, “A growing neural gas network learns topologies,” in *Advances in Neural Information Processing Systems 7*, pp. 625–632, MIT Press, 1995.
- [120] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction (Adaptive Computation and Machine Learning series)*. A Bradford Book, 1998.
- [121] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to Algorithms, Third Edition*. The MIT Press, 3rd ed., 2009.
- [122] D. P. Petersen and D. Middleton, “Sampling and reconstruction of wave-number-limited functions in n-dimensional euclidean spaces,” *Information and Control*,

- vol. 5, no. 4, pp. 279 – 323, 1962.
- [123] C. Koch, *Biophysics of Computation: Information Processing in Single Neurons (Computational Neuroscience Series)*. New York, NY, USA: Oxford University Press, Inc., 2004.
- [124] J. H. Conway, N. J. A. Sloane, and E. Bannai, *Sphere-packings, Lattices, and Groups*. New York, NY, USA: Springer-Verlag New York, Inc., 1987.
- [125] J. Leech and N. J. A. Sloane, *Sphere Packing and Error-Correcting Codes*, pp. 136–156. New York, NY: Springer New York, 1999.
- [126] H. Hayashi and J. Igarashi, “LTD windows of the STDP learning rule and synaptic connections having a large transmission delay enable robust sequence learning amid background noise,” *Cogn Neurodyn*, vol. 3, pp. 119–130, Jun 2009.
- [127] M. Hattori and Y. Kobayashi, “A hippocampal model for episodic memory using neurogenesis and asymmetric stdp,” in *2016 International Joint Conference on Neural Networks (IJCNN)*, pp. 5189–5193, July 2016.
- [128] J. Kerdels and G. Peters, “A computational model of grid cells based on dendritic self-organized learning,” in *Proceedings of the 5th International Joint Conference on Computational Intelligence (IJCCI 2013)*, pp. 420–429, 2013.
- [129] B. Si, E. Kropff, and A. Treves, “Grid alignment in entorhinal cortex,” *Biological Cybernetics*, vol. 106, no. 8, pp. 483–506, 2012.
- [130] M. Bosch and Y. Hayashi, “Structural plasticity of dendritic spines,” *Curr. Opin. Neurobiol.*, vol. 22, pp. 383–388, Jun 2012.
- [131] M. Segal, “Dendritic spines and long-term plasticity,” *Nat Rev Neurosci*, vol. 6, pp. 277–284, Apr 2005.
- [132] J. Cichon and W.-B. Gan, “Branch-specific dendritic ca2+ spikes cause persistent synaptic plasticity,” *Nature*, vol. 520, pp. 180–185, Apr 2015. Article.
- [133] J. P. Weber, B. K. Andrasfalvy, M. Polito, A. Mago, B. B. Ujfalussy, and J. K. Makara, “Location-dependent synaptic plasticity rules by dendritic spine cooperativity,” *Nat Commun*, vol. 7, p. 11380, Apr 2016.
- [134] Z. Varga, H. Jia, B. Sakmann, and A. Konnerth, “Dendritic coding of multiple sensory inputs in single cortical neurons in vivo,” *Proceedings of the National Academy of Sciences*, vol. 108, pp. 15420–15425, aug 2011.
- [135] M. P. Witter, T. P. Doan, B. Jacobsen, E. S. Nilssen, and S. Ohara, “Architecture of the Entorhinal Cortex A Review of Entorhinal Anatomy in Rodents with Some Comparative Notes,” *Front Syst Neurosci*, vol. 11, p. 46, 2017.
- [136] J. Widloski and I. R. Fiete, “A model of grid cell development through spatial exploration and spike time-dependent plasticity,” *Neuron*, vol. 83, pp. 481–495, Jul 2014.
- [137] J. S. Bains, J. M. Longacher, and K. J. Staley, “Reciprocal interactions between ca3 network activity and strength of recurrent collateral synapses,” *Nat Neurosci*, vol. 2, pp. 720–726, Aug 1999.
- [138] J. Krupic, M. Bauza, S. Burton, C. Barry, and J. O’Keefe, “Grid cell symmetry is shaped by environmental geometry,” *Nature*, vol. 518, pp. 232–235, Feb 2015. Letter.
- [139] J. Krupic, M. Bauza, S. Burton, and J. O’Keefe, “Framing the grid: effect of boundaries on grid cells and navigation,” *The Journal of Physiology*, vol. 594, no. 22, pp. 6489–6499, 2016.
- [140] E. Kropff, J. E. Carmichael, M.-B. Moser, and E. I. Moser, “Speed cells in the medial entorhinal cortex,” *Nature*, vol. 523, pp. 419–424, Jul 2015. Article.
- [141] J. L. Kubie and A. A. Fenton, “Linear look-ahead in conjunctive cells: an entorhinal mechanism for vector-based navigation,” *Front Neural Circuits*, vol. 6, p. 20, 2012.
- [142] A. P. Witkin, “Scale-space filtering,” in *Proceedings of the Eighth International Joint Conference on Artificial Intelligence - Volume 2, IJCAI’83*, (San Francisco, CA, USA), pp. 1019–1022, Morgan Kaufmann Publishers Inc., 1983.
- [143] T. Lindeberg, “Scale-space theory: A basic tool for analyzing structures at different scales,” *Journal of Applied Statistics*, vol. 21, pp. 225–270, 1994.
- [144] T. Lindeberg, *Scale-Space Theory in Computer Vision (The Springer International Series in Engineering and Computer Science)*. Springer, 2010.
- [145] S. Behnke and R. Rojas, “Neural abstraction pyramid: a hierarchical image understanding architecture,” in *1998 IEEE International Joint Conference on Neural Networks Proceedings. IEEE World Congress on Computational Intelligence (Cat. No.98CH36227)*, vol. 2, pp. 820–825 vol.2, May 1998.
- [146] R. A. Young, “The Gaussian derivative model for spatial vision: I. Retinal mechanisms,” *Spat Vis*, vol. 2, no. 4, pp. 273–293, 1987.
- [147] M. A. Georgeson, K. A. May, T. C. A. Freeman, and G. S. Hesse, “From filters to features: Scale–space analysis of edge and blur coding in human vision,” *Journal of Vision*, vol. 7, no. 13, p. 7, 2007.
- [148] D. G. Lowe, “Distinctive image features from scale-invariant keypoints,” *Int. J. Comput. Vision*, vol. 60, pp. 91–110, Nov. 2004.
- [149] T. Lindeberg, “Image matching using generalized scale-space interest points,” *Journal of Mathematical Imaging and Vision*, vol. 52, no. 1, pp. 3–36, 2015.
- [150] T. Lindeberg, “A computational theory of visual receptive fields,” *Biol Cybern*, vol. 107, pp. 589–635, Dec 2013.
- [151] E. V. Lubenov and A. G. Siapas, “Hippocampal theta oscillations are travelling waves,” *Nature*, vol. 459, pp. 534–539, May 2009.
- [152] R. A. Finkel and J. L. Bentley, “Quad trees a data structure for retrieval on composite keys,” *Acta Informatica*, vol. 4, no. 1, pp. 1–9, 1974.
- [153] J. A. Orenstein, “Multidimensional tries used for associative searching,” *Information Processing Letters*, vol. 14, no. 4, pp. 150 – 157, 1982.
- [154] J. Maue, P. Sanders, and D. Matijevec, “Goal-directed shortest-path queries using precomputed cluster dis-

- tances,” *Journal of Experimental Algorithmics (JEA)*, vol. 14, p. 2, 2009.
- [155] D. Delling, M. Holzer, K. Müller, F. Schulz, and D. Wagner, “High-performance multi-level graphs,” in *In: 9th DIMACS Implementation Challenge*, pp. 52–65, 2006.
- [156] Y. Dabaghian, F. Memoli, L. Frank, and G. Carlsson, “A topological paradigm for hippocampal spatial map formation using persistent homology,” *PLoS Comput. Biol.*, vol. 8, no. 8, p. e1002581, 2012.
- [157] E. W. Dijkstra, “A note on two problems in connexion with graphs,” *Numerische Mathematik*, vol. 1, pp. 269–271, dec 1959.
- [158] P. E. Hart, N. J. Nilsson, and B. Raphael, “A formal basis for the heuristic determination of minimum cost paths,” *IEEE Transactions on Systems, Science, and Cybernetics*, vol. SSC-4, no. 2, pp. 100–107, 1968.
- [159] D. Marr, *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. New York, NY, USA: Henry Holt and Co., Inc., 1982.