

Emergent Symmetry in a Finite Model of Navigational Neurons

Vadim K. Weinstein (vadim.weinstein@oulu.fi), Filip Georgiev (filip.georgiev@oulu.fi)
Kalle G. Timperi (kalle.timperi@oulu.fi), Nicoletta Prencipe (nicoletta.prencipe@oulu.fi),
Steven M. LaValle (steven.lavalle@oulu.fi),

Center for Ubiquitous Computing, Faculty of Information Technology and Electrical Engineering, University of Oulu
Erkki Koiso-Kanttilan katu 3, 90570 Oulu, Finland

Abstract

Consider an agent moving in the n -dimensional integer lattice. We model a neuron of this agent as a finite transition system and show, using a pumping lemma type of argument, that a grid cell-like behavior emerges under very mild assumptions. If there is one location in the environment which is reliably recognized by the neuron, then there is a “grid” of locations which are indistinguishable from the point of view of that neuron. By a grid we mean a finite union of cosets of a full-rank subgroup of the lattice. We propose that this may shed light on the possible origins of grid cells.

Keywords: grid cells; finite automata; information transition systems; emergence

A minimal requirement for a neural structure which helps navigation is that its functionality depends in a reliable way on the spatial location of the agent. The ability to return to the same spot after either a long or a short time is one of the fundamental needs of most mobile living creatures. Grid cells offer an example of neurons whose firing depends on the organism’s location. They have been found in rats’ and other mammals’ medial entorhinal cortices (Hafting, Fyhn, Molden, Moser, & Moser, 2005; Fyhn, Hafting, Witter, Moser, & Moser, 2008; Yartsev, Witter, & Ulanovsky, 2011; Killian, Jutras, & Buffalo, 2012). Other navigation cells include border¹ (Solstad, Boccara, Kropff, Moser, & Moser, 2008), head direction (Taube, Muller, & Ranck, 1990), speed (Kropff, Carmichael, Moser, & Moser, 2015), object-vector (Høydal, Skytøen, Andersson, Moser, & Moser, 2019), and place cells (O’Keefe & Nadel, 1978). If the locations of the organism navigating a 2D-space are recorded when a given grid cell fires, a periodic triangular pattern covering the available environment typically emerges (Fyhn, Molden, Witter, Moser, & Moser, 2004).

A lot of attention has been devoted to the question of how grid cells “know” when to fire. Proposed models include attractor dynamics (Yoon et al., 2013), path integration (Burak & Fiete, 2009; Fuhs & Touretzky, 2006), superposition of wave inputs (Burgess, 2008), Fourier analysis (Rodríguez-Domínguez & Caplan, 2018), spike-timing-dependent synaptic plasticity (Monsalve-Mercado & Leibold, 2017), and many more (Giocomo, Moser, & Moser, 2011; Moser et al., 2014).

In contrast to this line of research, in this paper we propose that a grid cell-like behaviour will under certain assumptions

inevitably emerge from any attempt of a neuron to correlate with any spatial location. The intuition is that if a neuron has a limited capacity, its behaviour will inevitably be repeated; eventually its state will circle back to the same state as before and from that point on its behaviour will be repeated. Since the environment is “larger” than the neuron’s state space, this circling back is bound to happen even if the agent has not returned to the same location. This intuition is the same as the one behind the so-called *pumping lemma* in the theory of deterministic finite automata and transition systems, (Sipser, 2013). We assume that the neuron is modelled by a finite state deterministic strongly connected deterministic transition system which receives the agent’s motion primitives as an input (the “efferent copy”), and that the environment is the n -dimensional integer lattice.

Basic Definitions

A *deterministic transition system* (DTS) is a tuple (X, x_0, U, τ) where X and U are sets, $\tau: X \times U \rightarrow X$ is a transition function, and $x_0 \in X$ is the initial state. By convention, we denote xu instead of $\tau(x, u)$. Denote by U^* the set of finite strings of elements of U including the empty sequence denoted by λ . Given $x \in X$ and $\mathbf{u} \in U^*$, denote by $x\mathbf{u} = xu_1 \cdots u_n$ where $\mathbf{u} = (u_1, \dots, u_n)$. Given $\mathbf{u}_1, \mathbf{u}_2 \in U^*$, denote their concatenation by $\mathbf{u}_1\mathbf{u}_2$. We now have $(x\mathbf{u}_1)\mathbf{u}_2 = x(\mathbf{u}_1\mathbf{u}_2)$ for all $x \in X$, $\mathbf{u}_1, \mathbf{u}_2 \in U^*$ and by convention $x\lambda = x$. This makes $x \mapsto x\mathbf{u}$ an action of the monoid U^* under concatenation on X . A DTS is *strongly connected*, if for all $x_1, x_2 \in X$ there is $\mathbf{u} \in U^*$ with $x_1\mathbf{u} = x_2$.

By \mathbb{N} we denote the set of natural numbers including zero, \mathbb{Z} is the set of integers. Let $d \in \mathbb{N}$ be the dimension of the environment. Let U be the standard basis of \mathbb{R}^d , $U = \{\mathbf{e}_1, \dots, \mathbf{e}_d\}$ where \mathbf{e}_k is the unit vector along the k :th coordinate. The *environment* is the DTS $(\mathbb{Z}^d, \mathbf{0}, U, \tau)$ where $\mathbf{0}$ is the null vector, and $\tau(\mathbf{z}, u) = \mathbf{z} + u$, for all $\mathbf{z} \in \mathbb{Z}^d, u \in U$. This means that for all $\mathbf{u} \in U^*$, we have $\mathbf{0}\mathbf{u} = \sum_{k=0}^{n-1} u_k$ where $n = |\mathbf{u}|$ is the length of \mathbf{u} . Note that now for all $\mathbf{u}, \mathbf{u}' \in U^*$ we have $\mathbf{0}\mathbf{u}\mathbf{u}' = \mathbf{0}\mathbf{u} + \mathbf{0}\mathbf{u}'$.

The *information transition system* (or the *neuron*) is a tuple $\mathcal{J} = (I, i_0, U, \varphi, F)$ where (I, i_0, U, φ) is a DTS and $F \subseteq I$ is the set of *firing states*. Note that the set U is the same as that of the environment. This reflects the modelling assumption that the motion primitives (elements of U) which govern the motion of the agent are also the only inputs to this neuron. These can be thought of as the efferent copies. As before, if $u \in U$ and $i \in I$, denote by $iu = \varphi(i, u)$, and by $i\mathbf{u} = iu_1 \cdots u_n$ where $\mathbf{u} = (u_1, \dots, u_n)$. When the agent is moving around, it is executing some motion $\mathbf{u} \in U^*$. The resulting location of the agent is $\mathbf{z} = \mathbf{0}\mathbf{u}$ and the resulting internal state is $i = i_0\mathbf{u}$. Note that

¹Also known as boundary cells.

by our notational convention the first one is evaluated using τ and the second using φ . If $i \in F$, we say that the *neuron fires* at \mathbf{z} , otherwise it does not fire. If the neuron reliably fires at \mathbf{z} , i.e. independently of \mathbf{u} which led there, we say that \mathbf{z} is an *always firing location*. More formally, let

$$\mathbf{A}(\mathcal{J}) = \{\mathbf{z} \in \mathbb{Z}^d \mid \forall \mathbf{u} \in U^* (\mathbf{0}\mathbf{u} = \mathbf{z} \rightarrow i_0\mathbf{u} \in F)\}.$$

For our purposes we will define a relatively general concept of a *grid*. A *full-rank sublattice of \mathbb{Z}^d* is a set of the form

$$\{\mathbf{c} + n_1\mathbf{z}_1 + \dots + n_d\mathbf{z}_d \mid n_1, \dots, n_d \in \mathbb{Z}\}$$

where $\mathbf{c} \in \mathbb{Z}^d$ is a constant, and $\mathbf{z}_1, \dots, \mathbf{z}_d \in \mathbb{Z}^d$ is a basis for \mathbb{R}^d , or equivalently, $\mathbf{z}_1, \dots, \mathbf{z}_d$ are linearly independent as vectors in \mathbb{R}^d . Using algebraic terminology this is the same as a coset of a full-rank subgroup of \mathbb{Z}^d . A *grid* in \mathbb{Z}^d is a finite union of full-rank sublattices of \mathbb{Z}^d .

Main Theorem

We are now in the position to formulate our main theorem using the notation from the previous section:

Theorem 1. *If \mathcal{J} is finite and strongly connected, then $\mathbf{A}(\mathcal{J})$ is a grid.*

Proof. For each $i \in I$, let $Z(i) = \{\mathbf{z} \in \mathbb{Z}^d \mid \exists \mathbf{u} \in U^* (\mathbf{z} = \mathbf{0}\mathbf{u} \wedge i_0\mathbf{u} = i)\}$. First observe that $Z(i_0)$ is closed under addition: given $\mathbf{z}, \mathbf{z}' \in Z(i_0)$ there are \mathbf{u} and \mathbf{u}' such that $\mathbf{z} = \mathbf{0}\mathbf{u}$, $\mathbf{z}' = \mathbf{0}\mathbf{u}'$, and $i_0\mathbf{u} = i_0\mathbf{u}' = i_0$. Now $\mathbf{z} + \mathbf{z}' = \mathbf{0}\mathbf{u} + \mathbf{0}\mathbf{u}' = \mathbf{0}\mathbf{u}\mathbf{u}'$; on the other hand we have $i_0\mathbf{u}\mathbf{u}' = i_0\mathbf{u}' = i_0$, so $\mathbf{u}\mathbf{u}'$ witnesses that $\mathbf{z} + \mathbf{z}' \in Z(i_0)$. A subset $Z \subset \mathbb{Z}^d$ is *k-dense*, if for all $\mathbf{z} \in \mathbb{Z}^d$ there is $\mathbf{z}' \in Z$ with $|\mathbf{z} - \mathbf{z}'| \leq k$. We observe then that $Z(i_0)$ is *k-dense*: Since \mathcal{J} is finite and strongly connected, there is $k \in \mathbb{N}$ such that every state $i \in I$ can be reached from any other state in less than k steps. Then given $\mathbf{z} \in \mathbb{Z}^d$, find \mathbf{u} such that $\mathbf{0}\mathbf{u} = \mathbf{z}$ and let \mathbf{u}' be the shortest sequence which takes $i_0\mathbf{u}$ to i_0 . Then $\mathbf{z}' = \mathbf{z} + \mathbf{0}\mathbf{u}' \in Z(i_0)$ and $|\mathbf{z} - \mathbf{z}'| = |\mathbf{0}\mathbf{u}'| < k$. We use the following without proof due to space limitations:

Fact. *Every k-dense subset of \mathbb{Z}^d which is closed under addition is a full-rank subgroup of \mathbb{Z}^d .*

Thus, $Z(i_0)$ is a full-rank subgroup of \mathbb{Z}^d . From this it is straightforward to see that for each \mathbf{u} , the coset $\mathbf{0}\mathbf{u} + Z(i_0)$ equals $Z(i_0\mathbf{u})$. In particular, if $Z(i_0\mathbf{u}) \cap Z(i_0\mathbf{u}')$ is non-empty, then $Z(i_0\mathbf{u}) = Z(i_0\mathbf{u}')$. Define $\mathbf{u} \sim \mathbf{u}'$ iff $Z(i_0\mathbf{u}) = Z(i_0\mathbf{u}')$. Then \sim is an equivalence relation on U^* . Denote the \sim -equivalence class of $\mathbf{u} \in U^*$ by $[\mathbf{u}]$. Let $F' \subset F$ be defined by

$$F' = \{i_0\mathbf{u} \mid \forall \mathbf{u}' \in [\mathbf{u}] (i_0\mathbf{u}' \in F)\}.$$

The theorem will follow once we prove that

$$\mathbf{A}(\mathcal{J}) = \bigcup_{i \in F'} Z(i). \quad (1)$$

Suppose $\mathbf{z} \in Z(i)$ and $i \in F'$. By the definition of $Z(i)$, there is some \mathbf{u} such that $\mathbf{z} = \mathbf{0}\mathbf{u}$ and $i_0\mathbf{u} = i$. Suppose \mathbf{u}' is any (other) sequence such that $\mathbf{z} = \mathbf{0}\mathbf{u}'$. Then $\mathbf{z} \in Z(i) \cap Z(i_0\mathbf{u}')$

and so $\mathbf{u} \sim \mathbf{u}'$. By the assumption that $i \in F'$, we have $i_0\mathbf{u}' \in F$. By arbitrariness of \mathbf{u}' it follows that $\mathbf{z} \in \mathbf{A}(\mathcal{J})$. The inclusion from right to left in (1) follows. Suppose now that $\mathbf{z} \notin Z(i)$ for all $i \in F'$. Let \mathbf{u} be some string such that $\mathbf{z} = \mathbf{0}\mathbf{u}$. Then $\mathbf{z} \in Z(i_0\mathbf{u})$ and so $i_0\mathbf{u} \notin F'$. This means that there is $\mathbf{u}' \in [\mathbf{u}]$ such that $i_0\mathbf{u}' \notin F$. Thus $\mathbf{z} \in Z(i_0\mathbf{u}')$, and there is $\mathbf{u}'' \in U^*$ such that $\mathbf{z} = i_0\mathbf{u}''$ and $i_0\mathbf{u}'' = i_0\mathbf{u}'$. Since the latter is not in F , the string \mathbf{u}'' witnesses that it is possible to arrive to \mathbf{z} without firing, so $\mathbf{z} \notin \mathbf{A}(\mathcal{J})$ which proves the inclusion from left to right in (1) and concludes the proof. \square

In particular, $\mathbf{A}(\mathcal{J})$ can be empty (corresponding to the case $F' = \emptyset$ in the proof). But, if it is non-empty, then it automatically contains a full-rank sublattice of \mathbb{Z}^d and is in fact a finite union of such lattices. This means that if the neuron learns to recognize a location (fire consistently when the agent arrives to that location), then it will recognize a grid of other locations and they will all be indistinguishable from the perspective of that neuron.

Discussion

We have shown that if the neuron is modelled as a finite deterministic transition system and the environment in which the agent moves is \mathbb{Z}^d , then the set \mathbf{A} of those points $\mathbf{z} \in \mathbb{Z}^d$ in which the neuron is reliably in a firing state (no matter which action sequence took it there), is a grid. By our own definition a grid is a finite union of full-rank sublattices. This means that it is a period set in a strong sense: for each unit vector $\mathbf{e} \in \mathbb{Z}^d$ there is n such that $\mathbf{A} = n\mathbf{e} + \mathbf{A}$.

Weaknesses

Our approach does not cover cases where the environment is more complex than the infinite lattice. Most experiments on grid cells are done in rooms with four walls. We also do not assume any sensory input to the neuron (only efferent copy of the performed actions). Finally, even though our framework explains the emergence of periodic activity, it does not explain the emergence of a hexagonal pattern. Our assumption of discreteness might be crucial. Can we reach a similar conclusion with a continuous action on \mathbb{R}^d with a compact internal state space? A suitable continuous framework for motion planning developed by Yershov and LaValle (2010) and further by Weinstein and LaValle (2024).

Future work

An obvious path forward is to try to account for the weaknesses described above. The emergence of the hexagonal pattern might be explained with the current approach, if it is combined with certain minimality considerations such as those presented by Weinstein, Sakcak, and LaValle (2022); Sakcak, Timperi, Weinstein, and LaValle (2023). A non-deterministic version of Theorem 1 should be possible to prove: one where the dynamics of the neuron are non-deterministic. For this case define the set \mathbf{A} as the set of all those locations where the neuron will always end up in a firing state notwithstanding the indeterminacy. It has already

been shown by Boccara, Nardin, Stella, O'Neill, and Csicsvari (2019) that the grid is distorted in the presence of goals. Thus, introducing information about objectives or other sensory data to the picture as well as other navigation cells would also be interesting.

Origins of Grid Cells?

Does our Theorem 1 suggest a potential new theory of the evolutionary emergence of grid cells? One can argue that there is an evolutionary pressure for even the smallest organisms to learn to recognize when they have come back to a particular location. For example, if the organism finds a comfortable spot, but has to leave it to find nutrition, it benefits from knowing its way back. Natal homing is an extreme example of this which has developed in many species to large scale navigation patterns, but likely has evolutionary origins in more local strategies. In accordance with Theorem 1 this might lead to a grid-like recognition pattern as a side effect. This, in turn, may lead to the emergence of other “grid cells” to disambiguate the indistinguishability raised by the first grid cell and so on.

Acknowledgments

This work was supported by a European Research Council Advanced Grant (ERC AdG, ILLUSIVE: Foundations of Perception Engineering, 101020977), and Academy of Finland (PERCEPT 322637).

References

- Boccara, C. N., Nardin, M., Stella, F., O'Neill, J., & Csicsvari, J. (2019, March). The entorhinal cognitive map is attracted to goals. *Science*, *363*(6434), 1443–1447. doi: 10.1126/science.aav4837
- Burak, Y., & Fiete, I. R. (2009, February). Accurate path integration in continuous attractor network models of grid cells. *PLoS Computational Biology*, *5*(2), e1000291. doi: 10.1371/journal.pcbi.1000291
- Burgess, N. (2008, November). Grid cells and theta as oscillatory interference: Theory and predictions. *Hippocampus*, *18*(12), 1157–1174. doi: 10.1002/hipo.20518
- Fuhs, M. C., & Touretzky, D. S. (2006, April). A spin glass model of path integration in rat medial entorhinal cortex. *The Journal of Neuroscience*, *26*(16), 4266–4276. doi: 10.1523/jneurosci.4353-05.2006
- Fyhn, M., Hafting, T., Witter, M. P., Moser, E. I., & Moser, M. B. (2008, August). Grid cells in mice. *Hippocampus*, *18*(12), 1230–1238. doi: 10.1002/hipo.20472
- Fyhn, M., Molden, S., Witter, M. P., Moser, E. I., & Moser, M. B. (2004). Spatial representation in the entorhinal cortex. *Science*, *305*(5688), 1258–1264. doi: 10.1126/science.1099901
- Giocomo, L. M., Moser, M.-B., & Moser, E. I. (2011, August). Computational models of grid cells. *Neuron*, *71*(4), 589–603. doi: 10.1016/j.neuron.2011.07.023
- Hafting, T., Fyhn, M., Molden, S., Moser, M.-B., & Moser, E. I. (2005, June). Microstructure of a spatial map in the entorhinal cortex. *Nature*, *436*(7052), 801–806. doi: 10.1038/nature03721
- Høydal, Ø. A., Skytøen, E. R., Andersson, S. O., Moser, M.-B., & Moser, E. I. (2019, April). Object-vector coding in the medial entorhinal cortex. *Nature*, *568*(7752), 400–404. doi: 10.1038/s41586-019-1077-7
- Killian, N. J., Jutras, M. J., & Buffalo, E. A. (2012, October). A map of visual space in the primate entorhinal cortex. *Nature*, *491*(7426), 761–764. doi: 10.1038/nature11587
- Kropff, E., Carmichael, J. E., Moser, M.-B., & Moser, E. I. (2015, July). Speed cells in the medial entorhinal cortex. *Nature*, *523*(7561), 419–424. doi: 10.1038/nature14622
- Monsalve-Mercado, M. M., & Leibold, C. (2017, July). Hippocampal spike-timing correlations lead to hexagonal grid fields. *Physical Review Letters*, *119*(3). doi: 10.1103/physrevlett.119.038101
- Moser, E. I., Roudi, Y., Witter, M. P., Kentros, C., Bonhoeffer, T., & Moser, M.-B. (2014, June). Grid cells and cortical representation. *Nature Reviews Neuroscience*, *15*(7), 466–481. doi: 10.1038/nrn3766
- O'Keefe, J., & Nadel, L. (1978). *The hippocampus as a cognitive map*. Oxford University Press.
- Rodríguez-Domínguez, U., & Caplan, J. B. (2018, November). A hexagonal fourier model of grid cells. *Hippocampus*, *29*(1), 37–45. doi: 10.1002/hipo.23028
- Sakcak, B., Timperi, K. G., Weinstein, V. K., & LaValle, S. M. (2023, September). A mathematical characterization of minimally sufficient robot brains. *The International Journal of Robotics Research*. doi: 10.1177/02783649231198898
- Sipser, M. (2013). *Introduction to the theory of computation* (Third ed.). Boston, MA: Course Technology.
- Solstad, T., Boccara, C. N., Kropff, E., Moser, M.-B., & Moser, E. I. (2008, December). Representation of geometric borders in the entorhinal cortex. *Science*, *322*(5909), 1865–1868. doi: 10.1126/science.1166466
- Taube, J., Muller, R., & Ranck, J. (1990, February). Head-direction cells recorded from the postsubiculum in freely moving rats. i. description and quantitative analysis. *The Journal of Neuroscience*, *10*(2), 420–435. doi: 10.1523/jneurosci.10-02-00420.1990
- Weinstein, V. K., & LaValle, S. M. (2024). Equivalent environments and covering spaces for robots. In F. Michael & G. Jesus (Eds.), *Geometry, topology and AI*. EMS.
- Weinstein, V. K., Sakcak, B., & LaValle, S. M. (2022, September). An enactivist-inspired mathematical model of cognition. *Frontiers in Neurorobotics*, *16*. doi: 10.3389/fnbot.2022.846982
- Yartsev, M. M., Witter, M. P., & Ulanovsky, N. (2011, November). Grid cells without theta oscillations in the entorhinal cortex of bats. *Nature*, *479*(7371), 103–107. doi: 10.1038/nature10583
- Yershov, D. S., & LaValle, S. M. (2010). Sufficient conditions for the existence of resolution complete planning algorithms. In *Algorithmic foundations of robotics ix* (p. 303–320). Springer Berlin Heidelberg.

Yoon, K., Buice, M. A., Barry, C., Hayman, R., Burgess, N., & Fiete, I. R. (2013, July). Specific evidence of low-dimensional continuous attractor dynamics in grid cells. *Nature Neuroscience*, *16*(8), 1077–1084. doi: 10.1038/nn.3450