Hippocampal Spike-Timing Correlations Lead to Hexagonal Grid Fields

Mauro M. Monsalve-Mercado and Christian Leibold

Department Biologie II & Graduate School of Systemic Neurosciences, LMU Munich, Großhadernerstr. 2,
82152 Planegg, Germany, and Bernstein Center for Computational Neuroscience Munich,
Großhadernerstr. 2, 82152 Planegg, Germany

Space is represented in the mammalian brain by the activity of hippocampal place cells, as well as in their spike-timing correlations. Here, we propose a theory for how this temporal code is transformed to spatial firing rate patterns via spike-timing-dependent synaptic plasticity. The resulting dynamics of synaptic weights resembles well-known pattern formation models in which a lateral inhibition mechanism gives rise to a Turing instability. We identify parameter regimes in which hexagonal firing patterns develop as they have been found in medial entorhinal cortex.

The spatial position of an animal can be reliably decoded from the neuronal activity of several cell populations in the hippocampal formation [1–3]. For example, place cells in the hippocampus fire at only a few locations in a spatial environment [4,5], and the position of the animal can be readily read out from single active neurons. Grid cells of the medial entorhinal cortex (MEC) fire at multiple distinct places that are arranged on a hexagonal lattice [6,7]. Although hexagonal patterns are abundant in nature and there exist well-studied physical theories for their emergence, the mechanistic origin of this neuronal grid pattern is still unclear. Initially, it was suggested that they result from continuous attractor dynamics [8,9] or the superposition of plane wave inputs [10] and, based on circuit anatomy, place cells would then result from a superposition of many grid cells [11,12]. More recent experiments, however, reported place cell activity without intact grid cells such that grid cells are not the unique determinants of place field firing [13–17]. Conversely, it would thus be possible that grid fields may arise from place field input as suggested in [18–20]. The biological mechanisms proposed by these latter theories, however, remain hypothetical. In the present Letter, we propose a learning rule for grid cells, based on the individual spike timings of place cells using spike-timing-dependent synaptic plasticity (STDP) [21–23]. The theory thereby predicts that the observed temporal hippocampal firing patterns (phase precession and theta-scale correlations, see below) [24–26] translate the temporal proximity of sequential place field spikes into spatial neighborhood relations, observed in grid-field activity. For our model to work, we only have to assume that the synaptic plasticity rule averages over a sufficiently long time interval.

Model.—We use the classical formulation of pairwise additive STDP [22,27], where the update of a synaptic weight J_n, $n = 1, \ldots, N$ at time t is computed as [22]

$$\frac{d}{dt}J_n = \int_{-\infty}^{\infty} ds\, W(s) C_n(s) + F(J_n). \quad (1)$$

$C_n(s)$ denotes the time-averaged correlation function between the spike train of the presynaptic neuron n and the postsynaptic neuron, the learning window $W(s)$ describes the update of the synaptic weight as a function of the time difference s between a pair of pre- and postsynaptic action potentials, and the function F implements soft bounds for the weight increase. The dynamics is further constrained such that weights cannot become negative.

To be able to treat Eq. (1) analytically, we use a linear Poisson neuron model; i.e., the mean firing rate of the postsynaptic neuron $E(t) = \mathbf{J} \cdot \mathbf{H}(t)$ results from a weighted sum of hippocampal firing rates $\mathbf{H} = (H_1(t), \ldots, H_N(t))^T$. Under these assumptions, $C_n(s)$ can be approximated for large N [22] as $C_n(s) = \sum_{n'} J_{n'} C_{nn'}(s)$, with

$$C_{nn'}(s) := \int_{-\infty}^{\infty} dt\, H_n(t) H_{n'}(t - s). \quad (2)$$

Inserting the correlation functions from Eq. (2) into the weight dynamics from Eq. (1) yields

$$\frac{d}{dt}J_n = \sum_{n'} J_{n'} G_{nn'} + F(J_n), \quad G_{nn'} := \int_{-\infty}^{\infty} ds\, W(s) C_{nn'}(s). \quad (3)$$

Following [28,29], we introduce the quadratic stabilization term $F(J) = F_0 J(K - J)$, $F_0 > 0$ that implements a soft upper bound.

As an input to the postsynaptic neuron, we consider a population of N hippocampal place cells. The firing of these neurons is characterized by a bell-shaped envelope, modulating the spatial path $x_{\mathcal{P}}(t)$ and oscillations in time t [Fig. 1(a)]

$$H_n(t; \mathcal{P}) = a e^{\frac{-[x_{\mathcal{P}}(t) - x_n]^2}{2\sigma^2}} [\cos(\omega t + \phi_n) + 1]/2. \quad (4)$$

The oscillation frequency ω of a neuron is slightly higher than the frequency $\omega_\theta/(2\pi)$ of the theta oscillation ($\sim$8 Hz)
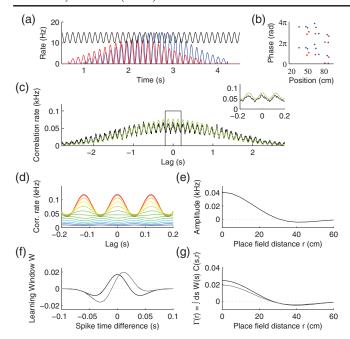
FIG. 1. (a) Poisson model of place field firing for two place cells (red, blue) and slower theta oscillation (black). (b) Phase precession resulting from the path used in (a). (c) Correlation of spike trains for the cells in (a), averaged over 2D trajectories (random walk) (black) and from Eq. (5) (green). Inset magnifies lag 0. (d) Correlation functions for different place field distances (red: autocorrelation, blue: 60 cm). (e) Correlation amplitude as a function of place field distance. (f) Examples of STDP learning windows, see Eq. (9) and below. (g) Γ kernels for correlation function from (c)–(e) and learning windows from (f).

in the local field potential, giving rise to a phenomenon called theta phase precession [Fig. 1(b)]: spikes early in the field come at later phases than spikes late in the field [24]. During the traversal of a place field, phase precession spans a whole theta cycle [30]. Thus, the two frequencies have to relate to each other like $\omega = \omega_\theta + (\pi/R)v$, with R denoting the distance from the place field center, at which, the firing rate has decreased to 10% [i.e., $R = \sigma\sqrt{2\ln(10)}$], and v denoting the running speed, which we fix at 25 cm/s. At each individual entry into the place field, the phase of the cellular oscillation is reset to phase zero with respect to the theta oscillation phase ϕ_θ by fixing $\phi_n = (\phi_\theta - 2\pi)\omega/\omega_\theta$.

To obtain a closed expression for the correlation function $C_{nn'}(s)$ [Fig. 1(c)], the time average in Eq. (2) is performed over all straight paths $\mathcal{P}$ crossing the center of the field overlap. For place fields with identical width R, firing rates a, and at small lags s, we obtain

$$
\begin{aligned}
C_{nn'}(s) &= \int_{\mathcal{P}}\int_{-\infty}^{\infty} dt\, H_n(t;\mathcal{P})H_{n'}(t-s;\mathcal{P}) \\
&= a\frac{\sqrt{\pi}e^{-(r^2+v^2s^2/4\sigma^2)}}{4v/(a\sigma)}\left[1 + \frac{1}{2}J_0\left(\frac{\pi r}{R}\right)\cos(\omega s)\right. \\
&\left. + \frac{vsr}{4\sigma^2}J_1\left(\frac{\pi r}{R}\right)\sin(\omega s)\right],
\end{aligned}
\tag{5}
$$

with place field distance $r = |x_n - x_{n'}|$ and $J_{0,1}$ denoting Bessel functions of the first kind (see Supplemental Material [31]).

In contrast to 1D fields, where the distance of place fields is reflected by the lag of the correlation peak (theta compression) [25], correlation functions in 2D are symmetric because of the symmetry of the path; however, the distance of the place field centers is encoded in the amplitude of the correlation peak at lag 0 [Figs. 1(d) and 1(e)].

Weight Dynamics.—Assuming that the putative grid cells receive inputs from a large number $N \gg 1$ of place cells that sufficiently cover the encoded area, we replace the presynaptic index n by the position of the place field center x; i.e., $G_{nn'} \to \Gamma(|x_n - x_{n'}|)$, and thereby translates the learning Eq. (3) to continuous coordinates

$$
\frac{d}{dt}J(x) = (\Gamma \circ J)(x) + F_0 J(x)(K - J).
\tag{6}
$$

Examples of the convolution kernel $\Gamma(|x|)$ for different learning window functions W are depicted in Figs. 1(f) and 1(g). The development of the weights follows the pattern formation principles of a lateral inhibition system [32]. Indeed, the integro-differential equation (IDE) (6) involves nonlocal interactions effectively implemented through the convolution kernel, inducing a strong close-range potentiation and a weaker long-range depression of neighboring synapses, as observed in the typical shape presented in Fig. 1(g). A general window-dependent kernel can be obtained for the correlation function from Eq. (5) as

$$
\Gamma(r) = c\frac{\sqrt{\pi}e^{-(r^2/4\sigma^2)}}{4v/(a\sigma)}\left[1 + \alpha J_0\left(\frac{\pi r}{R}\right) + \frac{\beta r}{\sigma}J_1\left(\frac{\pi r}{R}\right)\right]
\tag{7}
$$

$$
\begin{aligned}
c[W] &:= a\int_{-\infty}^{\infty}\exp\left[-\frac{v^2s^2}{4\sigma^2}\right]W(s)ds \\
\alpha[W] &:= \frac{a}{2c}\int_{-\infty}^{\infty}\exp\left[-\frac{v^2s^2}{4\sigma^2}\right]\cos(\omega s)W(s)ds \\
\beta[W] &:= \frac{av}{4\sigma c}\int_{-\infty}^{\infty}\exp\left[-\frac{v^2s^2}{4\sigma^2}\right]s\sin(\omega s)W(s)ds,
\end{aligned}
\tag{8}
$$

which can take a Mexican-hat-type shape for qualitatively different learning window functions W due to the symmetry of the correlation function [Figs. 1(f) and 1(g)]. To see this, we can regard Hebbian-like windows to be modeled as the product of a Gaussian and a polynomial of some order m, $W(s) = \exp[-s^2/(2\rho^2\mu^2)]P_m(s/\rho)$. The functionals defined in Eq. (8) then inherit the symmetries from the cross-correlation since all of the odd terms in the polynomial cancel out during integration. Thus, in subsequent numerical investigations, we focus on windows up to second polynomial order

$$
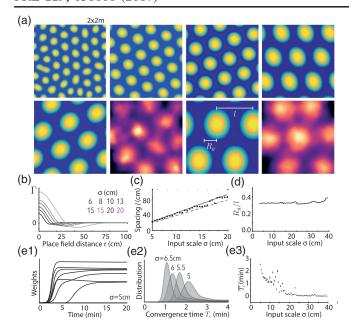W(s) = W_0(2\pi\mu^2\rho^2)^{-\frac{1}{2}}[1 - (s/\rho)^2]e^{-(s^2/2\rho^2\mu^2)},
\tag{9}
$$

FIG. 2. Weight patterns and input scale. (a) Asymptotically stable weight functions $J(x)$ for place field widths σ, as indicated in (b), together with the respective Γ kernels ($\mu = 1.025$, $\rho = 23$ ms, σ indicated by grey level). Firing rates of spiking simulations are colored in red (for numerical details, see Supplemental Material [31]). (c) Grid spacing l [see (a)] scales linearly with place field scale σ. Dots correspond to weight patterns from numerical solutions. Dashed and solid lines indicate theoretical estimates $2\pi/k_m$ and $2\pi/\tilde{k}_m$, see linear theory at early times and positivity constraint. (d) Estimated ratio of field radius R_g [see (a)] to grid spacing l; $R_g/l = [\sqrt{3}\nu/(2\pi)]^{1/2}$ is independent of σ. (e1) Temporal evolution of randomly sampled weights. (e2) Distribution of weight convergence times T_c for different input scales σ as indicated and the respective means (e3).

whose free parameters ρ and μ determine its zeroes $s_0 = \pm\rho$ and negativity $\int W = W_0(1 - \mu^2)$. In Figs. 1(f) and 1(g), we used $\rho = 23$ ms, $\mu = 1.025$ and added a linear term s/ρ to the polynomial to get the asymmetric window (grey lines). The W-dependent functions c, α, and β defining the kernel Γ are given in the Supplemental Material [31].

A numerical evaluation of the learning IDE (6) with periodic boundary conditions reveals that the spatially isotropic kernel Γ can result in hexagonal packing structures [Figs. 2(a) and 2(b)]. Simulations of spiking Poisson neurons confirm these predictions of the meanfield theory [Fig. 2(a)]. As indicated by the kernel function in Eq. (7), the grid spacing only depends on the spatial scale σ of the place fields in the input [Fig. 2(c)].

An experimentally accessible quantity to compare our model results to is the ratio of grid field radius R_g to grid spacing l as indicated in Fig. 2(a). In experiments, the ratio R_g/l for grid cells has been determined to be about 0.3 [7], and for a perfectly hexagonal grid, R_g/l relates to the fraction ν of field size per area as $\nu = \pi R_g^2/[(\sqrt{3}/2)l^2]$. The field fraction ν can be readily accessed from our numerics as the fraction of

nonzero synaptic weights and for the used learning window, fits the experimentally obtained R_g/l [Fig. 2(d)] for all choices of σ. Finally, learning converges faster for large spacing (large σ), consistent with a larger amplitude of Γ [Eq. (7) and Figs. 2(b) and 2(e)].

Linear theory at early times.—Some analytical understanding of the weight dynamics from Eq. (6) can be gained from a neural field theory approach [32–36]. In this framework, we can neglect the effect of the nonlinearities at early times and focus only on the convolution term $\Gamma * J$. The emerging dynamics can be readily understood by looking at the evolution of the weights in Fourier space $\partial_t \hat{J}(k) = \hat{\Gamma}(k)\hat{J}(k)$, which makes evident that the wave number k_m maximizing the kernel Fourier transform $\hat{\Gamma}$ (see Supplemental Material [31]) will exponentially overgrow all other modes [if $\hat{\Gamma}(k_m) > 0$], thus setting the initial periodicity of the pattern.

In a finite region of the parameter space (μ, ρ) of the learning window, the bimodal (Mexican hat) shape of Γ ensures the existence of a Turing instability, i.e., a transition to a single maximum of $\hat{\Gamma}$ [with $\hat{\Gamma}(k_m) > 0$] at a nonzero $k_m > 0$ [Figs. 3(a) and 3(c)]. Similar to previous work on pattern formation in lateral inhibition systems (e.g., [37]), the permitted parameter region (k_m exists and is positive) gives rise to stripelike and hexagonal patterns [Fig. 3(d)]. In the Supplemental Material [31], we also provide a complementary description of the pattern formation process, based on an approximation of Eq. (6) by a partial differential equation (Swift-Hohenberg equation [38]), which corroborates the results from the linear theory.

To connect the resulting patterns to other feed-forward models of grid field formation, we parameterize Γ by the shape factor r_0/r_m [Fig. 3(c)], which is the fraction between the zero and the minimum of Γ. The shape factor r_0/r_m reduces the two-parameter learning window to a single qualitatively descriptive parameter, which can be used to describe the bimodal kernel Γ independently of the hypothesized biological mechanism. If r_0/r_m is large (~ 0.8), Γ shows only little negativity, and the emerging pattern is stripelike [Figs. 3(c) and 3(d)]; if r_0/r_m is small, Γ exhibits strong negativity, the firing fields become dispersed, and the pattern looses hexagonality. Hexagonal patterns arise for r_0/r_m roughly between 0.65 and 0.75 [Fig. 3(d)]. In this region, the shape factor virtually completely determines the geometrical properties of the steady state [Figs. 3(e) and 3(f)]. Values of r_0/r_m that give rise to hexagonal grids can also be identified via the ratio of field width per grid spacing R_g/l. According to our theory, the experimentally observed value 0.3 [7] is achieved with a shape factor of about $r_0/r_m = 0.7$ [Fig. 3(f)]. For higher values of r_0/r_m, R_g/l increases to a point where a periodic pattern cannot further dissociate into disjoint fields, and the stable pattern becomes stripelike. For lower r_0/r_m, R_g/l decreases, and, at some point, the small fields no longer repel each other strongly enough to produce a symmetrical arrangement.
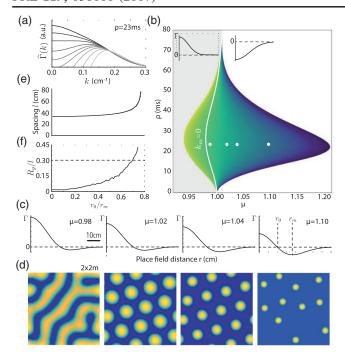
FIG. 3. Permitted learning windows. (a) Fourier transforms of Γ for $\mu = 0.95$ (black) to 1.20 (light grey). (b) Region of structure formation in (μ, ρ) space (for $\sigma = 10$ cm). Nontrivial patterns appear to the right of the white solid line ($k_m = 0$), where the selected wavelength is positive. Colored region indicates bimodal Γ. The color encodes the shape factor r_0/r_m (0: blue, 1: yellow). Grey and white areas correspond to regions where Γ is all positive or all negative. (c) Four examples of Γ kernels, corresponding to the white dots from (b) in the same order. (d) Weight patterns for kernels from (c). (e) Predicted spacing $2\pi/k_m$ as a function of r_0/r_m, obtained for all combinations of μ and ρ from the region right of the solid line in (a). (f) Estimated ratio of field size to grid spacing as a function of r_0/r_m (from simulated patterns). Dashed line indicates experimental value [7].

Positivity constraint.—The grid spacing l predicted by the linear theory, however, consistently underestimates the spacing derived from the numerical solution of the mean field dynamics [Fig. 2(c)]. The reason for this error is that, after the initial growth phase, the synaptic weights are influenced by the nonlinearities, most importantly the constraint that they cannot become negative.

The impact of this positivity constraint can be intuitively understood if we interpret the convolution $\Gamma \circ J$ as an operation that detects the best overlap of an oscillatory pattern $J \propto \cos(kx)$ with a given kernel Γ. However, after the lowest weights reach zero, they stop contributing to the convolution, and a slightly lower wave number $\tilde{k}_m$ maximizing

$$\tilde{\Gamma}(k) := \int_\Omega dx\, \Gamma(x) \cos(kx)\Theta[\cos(kx) - \cos(|k|R_g)] \quad (10)$$

will be favored as the fastest growing mode (Θ denoting the Heaviside function). Similarly, a particular field size R_g

maximizing $\tilde{\Gamma}$ will be selected. In the experimentally relevant case $|k|R_g = 2\pi R_g/l = 2\pi \times 0.3$, numerical maximization of Eq. (10) yielded the predicted wave number $\tilde{k}_m$ [solid line in Fig. 2(c)], which excellently agrees with the numerical solutions of the meanfield dynamics.

Conclusion.—For a large variety of STDP windows, the spike-timing correlations of 2D place cells can account for a feed-forward learning of hexagonal grid patterns. Synaptic plasticity thereby averages over running trajectories of tens of minutes, hence, translating the temporal correlations into a dense code for space. Our model thus predicts that grid cells are generated in the output structures of the hippocampus, e.g., the deep layers of the medial entorhinal cortex [39] or the parasubiculum [40]. While our linear theory provides a good prediction of grid spacing, as well as for conditions that permit structure formation, determining the boundary between hexagonal and stripe-like patterns is less straightforward and has to take into account the nonlinearities. The standard approach, nonlinear bifurcation analysis [37,41,42], is difficult because of the strong nonlinearity introduced via the positivity constraint, which strongly influences the selection of the final pattern. Despite this drawback, our model provides a universal framework in that it encompasses current models of grid field formation that can be mapped to convolutions with Mexican-hat-type kernels that give rise to a Turing instability.

[1] K. D. Harris, J. Csicsvari, H. Hirase, G. Dragoi, and G. Buzsaki, Nature (London) **424**, 552 (2003).
[2] A. Mathis, A. V. Herz, and M. Stemmler, Neural Comput. **24**, 2280 (2012).
[3] A. Kammerer and C. Leibold, PLoS Comput. Biol. **10**, e1003986 (2014).
[4] J. O'Keefe, Exp. Neurol. **51**, 78 (1976).
[5] R. U. Muller, J. L. Kubie, and J. B. Ranck, J. Neurosci. **7**, 1935 (1987).
[6] M. Fyhn, S. Molden, M. P. Witter, E. I. Moser, and M. B. Moser, Science **305**, 1258 (2004).
[7] T. Hafting, M. Fyhn, S. Molden, M. B. Moser, and E. I. Moser, Nature (London) **436**, 801 (2005).
[8] M. C. Fuhs and D. S. Touretzky, J. Neurosci. **26**, 4266 (2006).
[9] Y. Burak and I. R. Fiete, PLoS Comput. Biol. **5**, e1000291 (2009).
[10] N. Burgess, Hippocampus **18**, 1157 (2008).
[11] B. L. McNaughton, F. P. Battaglia, O. Jensen, E. I. Moser, and M.-B. Moser, Nat. Rev. Neurosci. **7**, 663 (2006).
[12] E. I. Moser and M. B. Moser, Hippocampus **18**, 1142 (2008).

[13] J. Koenig, A. N. Linder, J. K. Leutgeb, and S. Leutgeb, Science **332**, 592 (2011).

[14] M. P. Brandon, A. R. Bogaard, C. P. Libby, M. A. Connerney, K. Gupta, and M. E. Hasselmo, Science **332**, 595 (2011).

[15] T. J. Wills, F. Cacucci, N. Burgess, and J. O'Keefe, Science **328**, 1573 (2010).

[16] J. B. Hales, M. I. Schlesiger, J. K. Leutgeb, L. R. Squire, S. Leutgeb, and R. E. Clark, Cell Rep. **9**, 893 (2014).

[17] M. I. Schlesiger, C. C. Cannova, B. L. Boublil, J. B. Hales, E. A. Mankin, M. P. Brandon, J. K. Leutgeb, C. Leibold, and S. Leutgeb, Nat. Neurosci. **18**, 1123 (2015).

[18] E. Kropff and A. Treves, Hippocampus **18**, 1256 (2008).

[19] Y. Dordek, D. Soudry, R. Meir, and D. Derdikman, eLife **5**, e10094 (2016).

[20] A. Stepanyuk, Biol. Inspir. Cogn. Arc. **13**, 48 (2015).

[21] H. Markram, J. Lubke, M. Frotscher, and B. Sakmann, Science **275**, 213 (1997).

[22] R. Kempter, W. Gerstner, and J. L. van Hemmen, Phys. Rev. E **59**, 4498 (1999).

[23] G. Q. Bi and M. M. Poo, J. Neurosci. **18**, 10464 (1998).

[24] J. O'Keefe and M. L. Recce, Hippocampus **3**, 317 (1993).

[25] G. Dragoi and G. Buzsaki, Neuron **50**, 145 (2006).

[26] T. Feng, D. Silva, and D. J. Foster, J. Neurosci. **35**, 4890 (2015).

[27] W. Gerstner, R. Kempter, J. L. van Hemmen, and H. Wagner, Nature (London) **383**, 76 (1996).

[28] M. C. van Rossum, G. Q. Bi, and G. G. Turrigiano, J. Neurosci. **20**, 8812 (2000).

[29] W. M. Kistler and J. L. van Hemmen, Neural Comput. **12**, 385 (2000).

[30] A. P. Maurer, S. L. Cowen, S. N. Burke, C. A. Barnes, and B. L. McNaughton, Hippocampus **16**, 785 (2006).

[31] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.119.038101 for derivation.

[32] N. V. Swindale, Proc. R. Soc. B **208**, 243 (1980).

[33] R. Linsker, Proc. Natl. Acad. Sci. U.S.A. **83**, 7508 (1986).

[34] B. Ermentrout, J. Campbell, and G. Oster, Veliger **28**, 369 (1986).

[35] K. D. Miller, J. B. Keller, and M. P. Stryker, Science **245**, 605 (1989).

[36] H. Shouval, N. Intrator, C. C. Law, and L. N. Cooper, Neural Comput. **8**, 1021 (1996).

[37] G. B. Ermentrout and J. D. Cowan, Biol. Cybern. **34**, 137 (1979).

[38] J. Swift and P. C. Hohenberg, Phys. Rev. A **15**, 319 (1977).

[39] M. P. Witter, Hippocampus **3 S1**, 33 (1993).

[40] T. van Groen and J. M. Wyss, Brain Res. **518**, 227 (1990).

[41] G. B. Ermentrout and J. D. Cowan, SIAM J. Appl. Math. **38**, 1 (1980).

[42] G. B. Ermentrout and J. D. Cowan, SIAM J. Appl. Math. **39**, 323 (1980).