HIPPOCAMPUS 18:1256–1269 (2008)

The Emergence of Grid Cells: Intelligent Design or Just Adaptation?

Emilio Kropff[1,2] and Alessandro Treves[1,2*]

ABSTRACT: Individual medial entorhinal cortex (mEC) 'grid' cells provide a representation of space that appears to be essentially invariant across environments, modulo simple transformations, in contrast to multiple, rapidly acquired hippocampal maps; it may therefore be established gradually during rodent development. We explore with a simplified mathematical model the possibility that the self-organization of multiple grid fields into a triangular grid pattern may be a single-cell process, driven by firing rate adaptation and slowly varying spatial inputs. A simple analytical derivation indicates that triangular grids are favored asymptotic states of the self-organizing system, and computer simulations confirm that such states are indeed reached during a model learning process, provided it is sufficiently slow to effectively average out fluctuations. The interactions among local ensembles of grid units serve solely to stabilize a common grid orientation. Spatial information, in the real mEC network, may be provided by any combination of feedforward cortical afferents and feedback hippocampal projections from place cells, since either input alone is likely sufficient to yield grid fields. © **2008 Wiley-Liss, Inc.**

KEY WORDS: hippocampus; entorhinal cortex; firing rate adaptation; attractor network; memory

DO GRIDS STEM FROM ATTRACTORS OR FROM OSCILLATIONS?

Among the complex memory processes operating within the medial temporal lobe (see e.g., Eichenbaum and Lipton, 2008), the core contribution of the hippocampus may be its capacity to retrieve multiple arbitrary representations, a capacity that has been associated to the 'collateral effect' (Marr, 1971). McNaughton and Morris (1987) and Rolls (1989) proposed that the collateral effect is implemented by the recurrent connections of the CA3 region, which led to the insight that the CA3 network may 'compute' just by following attractor dynamics (Amit, 1989), as described by the simplified Hopfield (1982) model. The striking demonstration of abrupt global remapping, indicative of attractor dynamics, in rat place cells (Wills et al., 2005) has reinforced the notion that attractors are a key to understanding hippocampal memory computation. The concurrent discovery of grid cells in neighboring medial entorhinal cortex (mEC; Fyhn et al., 2004; Hafting et al., 2005) has led to the attractor idea reverberating into mEC networks: grid cells have been interpreted as the stable attractor states of spin-glass-like interactions among pyramidal cells, mediated by recurrent connections (Fuhs and Touretzky, 2006). Unlike the multiple representations in CA3, however, which require global remapping transitions from one to the other (Leutgeb et al., 2005), local ensembles of mEC grid cells seem to demonstrate a single representation, which shifts and rotates coherently in different environments (Fyhn et al., 2007). If so, and if attractor computation were its core design principle, what the recurrent network in mEC would produce is merely the recovery of this *single* representation, e.g., when disorganized, a somewhat meager yield for a network likely employing thousands of synapses per cell (cf. Battaglia and Treves, 1998). In contrast, those synapses may be effectively utilized for the accurate coding of position in an abstract, context-independent frame of reference—reflecting a computation along the dimensions of physical space, exactly *orthogonal* to the dimensions of convergence to the single putative attractor state. Interestingly, grid cells whose soma are physically close in the tissue do not present similar spatial phases, not even in mice (Fyhn et al., 2008), and in fact accurate position coding does not require such similarity (Fiete et al., 2008). Similar phases for nearby cells would instead be expected if the grid-structuring connectivity matrix, in the recurrent attractor network scenario, were a simple function of distance between cells in the adult tissue, as noted by McNaughton et al. (2006), who proposed a revised version of the mEC attractor idea. In the revised version, the connectivity depends on distance in the developing tissue, before cells acquire their final position in the adult tissue.

The theta rhythm and its associated phase precession (O'Keefe and Recce, 1993) have been another powerful source of inspiration for approaching hippocampal computation and, by extension, the emergence of grid units. Since the distance typically covered by a rat within a theta period is much shorter than the minimum grid spacing observed, grids have been hypothesized to emerge from the interference patterns among oscillations with slightly different frequencies close to theta (Blair et al., 2007). Although mathematically attractive, the mechanism requires the somewhat implausible combination of precisely two 'theta grid' units to produce an interference unit with larger spacing, and the theta grids must already present the 2D grid pattern themselves, on a finer scale. Alterna-

[1] **Kavli Institute for Systems Neuroscience and Centre for the Biology of Memory, NTNU—Norwegian University of Science and Technology, 7489 Trondheim, Norway;** [2] **Cognitive Neuroscience Sector, SISSA—International School for Advanced Studies, Trieste, Italy**

Grant sponsor: EU Spacebrain.

*Correspondence to: Alessandro Treves, SISSA, via Beirut 2, 34014 Trieste, Italy. E-mail: ale@sissa.it

Accepted for publication 8 September 2008
DOI 10.1002/hipo.20520
Published online 19 November 2008 in Wiley InterScience (www.interscience.wiley.com).

tively, one may think of a similar interference produced by the common rhythm and by intrinsic oscillations of individual cells (Burgess et al., 2007; Giocomo et al., 2007). In this case, the full-fledged spatial grid arises from the purely temporal oscillations only when combining three unidimensional waves at 60 degrees to each other, which might perhaps emerge from a yet-to-be-detailed self-organizing process. Assuming the self-organizing process to adjust the three waves at 60 degrees and their relative phases, the model successfully recreates an exact triangular grid pattern, though it is not clear, to us, what it predicts in terms of phase precession, the very phenomenon it was originally inspired by (but see Burgess, 2008; Hasselmo, 2008). The finding of phase precession in Layers II and V but not in Layer III grid cells (Hafting et al., 2006, 2008) has recently added a new degree of complexity to the analysis.

All these approaches to understanding the grid cell phenomenon, which are elaborated further in several of the contributions to this special issue (Blair et al., 2008; Burgess, 2008; Giocomo and Hasselmo, 2008; Jeewajee et al., 2008), rely on a common hypothesis: the grid expresses path integration mechanisms based on the rat's own perception of speed and direction (Barlow, 1964), while sensory information serves as an anchor that makes the map reproducible across sessions. This somehow secondary role assigned to sensory information has been recently challenged by the finding that the grid can expand and contract following gradual variations in the size of the environment (Barry et al., 2007) and is in general affected by manipulating sensory information, e.g., removing the boundaries of the recording enclosure (Savelli et al., 2008). If path integration has to be corrected constantly by sensory information, so as to generate a new map with different origin or even grid spacing, more attention should perhaps be devoted to sensory information itself as a source of grid fields.

In this line, the perspective that arises from consideration of the slowness principle (Wiskott, 2003) is that spatially modulated patterns of response may result from the extraction of the slowly varying components of the afferent sensory inputs, with no special role of either recurrent processing or oscillations (Franzius et al., 2007). This perspective suggests that similar spatial modulation (including grid cells, place cells, head-direction cells) may be expected to be observed in species with similar behavioral patterns to rodents, irrespective of whether they present similar neural circuitry (e.g., bats, Ulanovsky and Moss, 2007) or quite different circuitry (e.g., birds, Bingman and Sharp, 2006; see the pattern cells in Kahn et al., 2008). Such modulation would not necessarily be expected, instead, in species with similar circuitry but very different spatial behavior, e.g., large primates (Rolls, 1999). Although the slowness principle by itself does not seem to yield the triangular grid patterns, it suggests reconsidering the possibility that single-cell processes (and afferent inputs) may play the crucial role in determining the appearance of grid cells.

We had considered such a possibility early on, by focusing on a simplified model where the critical single-cell process is firing rate adaptation (Treves et al., 2005; Cerasti and Treves, 2006). Although the mathematical analysis indicated that firing rate adaptation should lead to triangular grids, the spatial modulation emerging from simulations was often triangular locally, but irregular at a larger spatial scale, and rather unstable (Cerasti and Treves, 2006). We reasoned that the lack of long-range order might result from the casual assortment of spatial variables, which had been taken, in our simulations, to model processed sensory and proprioceptive inputs to mEC. Recently, one of us (EK) wondered whether more finely balanced inputs, such as those arising from an orderly array of single place field units would produce 'better grids' than our previous simulations. We decided then to use, as inputs to our model mEC array, an array of 'place units,' but solely to have precise control over the regularity of their spatial code, without any conceptual commitment to the hypothesis that actual hippocampal place cells may participate in setting up the grid representation in entorhinal cortex. The converse hypothesis that grid units may participate in determining place fields (Brun et al., 2002) has of course been considered in several modeling studies (Rolls et al., 2006; Solstad et al., 2006; Hayman and Jeffery, 2008; Molter and Yamaguchi, 2008). Since CA fields and mEC are reciprocally connected by very substantial synaptic systems, it is evident that each structure will strongly influence the determination of spatial correlates in the other. The developmental time course of place fields and grid fields is still somewhat controversial (Martin and Berthoz, 2002; Ainge et al., 2008; Langston et al., 2008; Wills et al., 2008). To utilize the clarity that a computational model can offer, however, we prefer to 'cut' the reciprocal connections, isolate the phenomenon of pattern formation in the grid units, and simulate its emergence under the influence of regularly arranged, place-cell-like inputs, leaving open the possibility that the latter may model actual CA place units, or else regular spatial inputs from upstream cortical areas, or a combination of both. We report here the result of such a study, and we discuss below the effect of relaxing the regularity requirement, by using less regular, nonplace-field-like spatial inputs. Finally, we also sketch with preliminary simulations a possible auxiliary role of the collateral connections in mEC in the organization of a common orientation across grid units.

A MODEL BURDENED WITH NEURONAL FATIGUE

The network architecture is the simplest possible, with an input layer with N_I neurons projecting to a mEC layer composed of N_{mEC} threshold-linear neurons with saturation. At any given time t, a neuron i in the mEC layer receives inputs $\{r_j^t\}$ (denoting firing rate; but we shall use the notation $\psi_i(\mathbf{x})$ for firing rates in the mEC layer, as a function of the animal position $\mathbf{x}$, to point at the spatial map expressed by the activity of unit i) and computes the total synaptic activation

$$h_i^t = \frac{1}{N_I}\sum_{j=1}^{N_I} J_{ij} r_j^t$$

where J_{ij} is the weight of the synapse going from neuron j in the input layer to neuron i in the mEC layer. This activation would ordinarily be transformed directly into the output firing rate at time $t + 1$, ψ_i^{t+1}, e.g., by the transfer function

$$\psi(h) = \psi_{\text{sat}} \frac{2}{\pi} \arctan[g(h - \theta)]\Theta[h - \theta],$$

where the parameters θ and g stand for the threshold and the gain, assumed for simplicity to take common values for all mEC neurons, and the Heaviside function $\Theta(h - \theta)$ ensures that $\psi(x)$ is always positive, and zero when $h < \theta$. If g is small, the transfer function operates in the linear regime, away from its saturation value $\pi/2$ (which is then rescaled to make the saturation rate equal to ψ_{sat}). Since we want however to add a mechanism for adaptation, we do not apply the transfer function directly to the activation h_i^t, but rather introduce intermediate activation variables with fatigue dynamics

$$r_{\text{act}}^{t+1} = r_{\text{act}}^t + b_1\left(h^t - r_{\text{inact}}^t - r_{\text{act}}^t\right)$$
$$r_{\text{inact}}^{t+1} = r_{\text{inact}}^t + b_2\left(h^t - r_{\text{inact}}^t\right)$$

such that

$$\psi_i^t = \psi(r_{\text{act}}^t).$$

The parameters b_1 and b_2 are associated to the speed of rise and fall of the activity of a neuron that receives strong input. Note that since we set $b_1 > b_2$, when the neuron starts receiving a strong excitation, the variable r_{act} rises from 0 (its initial resting value) toward h, but as r_{inact} approaches h with a slower dynamics, the ascent of r_{act} is eventually reversed and its value comes back to 0.

Competition in the mEC layer is implemented by fixing the mean activity $a = \frac{1}{N_{\text{mEC}}} \sum \psi_k$ and the sparseness $s = (\sum \psi_k)^2/(N_{\text{mEC}} \sum \psi_k^2)$. To achieve this normalization, the parameters θ and g are modified at every time step after updating all neurons. In general terms, θ has a strong (inverse) influence on the mean activity, while g can be used to control the sparseness. After updating all neurons, we apply one or several times

$$\theta^{t+1} = \theta^t + b_3(a - a_0)$$
$$g^{t+1} = g^t + b_4 g^t (s - s_0)$$

(where a_0 and s_0 are the target mean activity and sparseness and b_3 and b_4 are parameters controlling the speed and smoothness of the changes) until a and s meet criteria of similarity to a_0 and s_0.

We use a hebbian learning rule to update the weights J_{ij}

$$J_{ij}^{t+1} = J_{ij}^t + \varepsilon\left(\psi_i r_j - \langle \psi_i \rangle \langle r_j \rangle\right)$$

where ε is the learning rate and $\langle r \rangle$ is a temporal average of the activity of the presynaptic unit. Similar to the BCM rule (Bienenstock et al., 1982), this provision has the advantage of correcting for correlations in the input (Kropff and Treves, 2007). In addition, weights are normalized in such a way that the total input weight to any neuron is constant. The main parameters have been given the values indicated in Table 1, unless stated differently.

TABLE 1.

Default Parameters Used in Simulation, Unless Otherwise Noted

N_{mEC}	N_{I}	b_1	b_2 (= $b_1/3$)	ψ_{sat}	a_0	s_0	ε
100	200	0.1	0.033	30	$0.1 \times \psi_{\text{sat}}$	0.3	0.001

The virtual rat moves following a smooth random walk in a square box with rigid walls. Figure 1a shows an example of such a trajectory. The simulations shown in this paper include trajectories at constant speed unless stated differently. Figure 1b shows a scheme of the model including only two mEC and four input neurons. As the rat explores the environment, different input neurons (in the particular scheme of Fig. 1b and of most simulations, model hippocampal place cells; otherwise, in the control simulations of Fig. 6 more generic spatially modulated units) get activated. A fully connected network of feedforward synapses transmits this information to the mEC neurons, which compete to get activated and strengthen those input synapses that have excited them enough to win the competition. Figure 1c shows the map and autocorrelogram of a single mEC unit during the first few steps of this learning process. Multiple peaks appear and, as learning proceeds, they become stronger and slightly move around, as if seeking a stable configuration.

A MATHEMATICAL ANALYSIS OF THE ASYMPTOTIC STATES OF THE MODEL

The way units in the model develop their response profiles is investigated through computer simulations. A mathematical analysis of a very abstract version of the model, however, indicates which are the asymptotic states that it may reach by adjusting its connection weights. The mathematical analysis does not even explicitly incorporate the individual parameters of the model, but is focused on the end result of its general learning dynamics, as expressed in the response profile of an individual unit. Of course, for this individual 'map' to be useful, different mEC units would have to be active when the animal is in different positions in space—to be ensured by competition, i.e., by the recurrent network—and also any environment would have to be bound to the existing universal map—to be ensured by the network of feedforward connections. We eschew these aspects, which we leave for the simulations to address, and in the mathematical analysis, we consider only the requirements on the single-cell map $\psi_i(x)$ that models the firing rate of cell i at position $\mathbf{x}$.

The analysis leads to two conclusions, as derived in the Appendix.

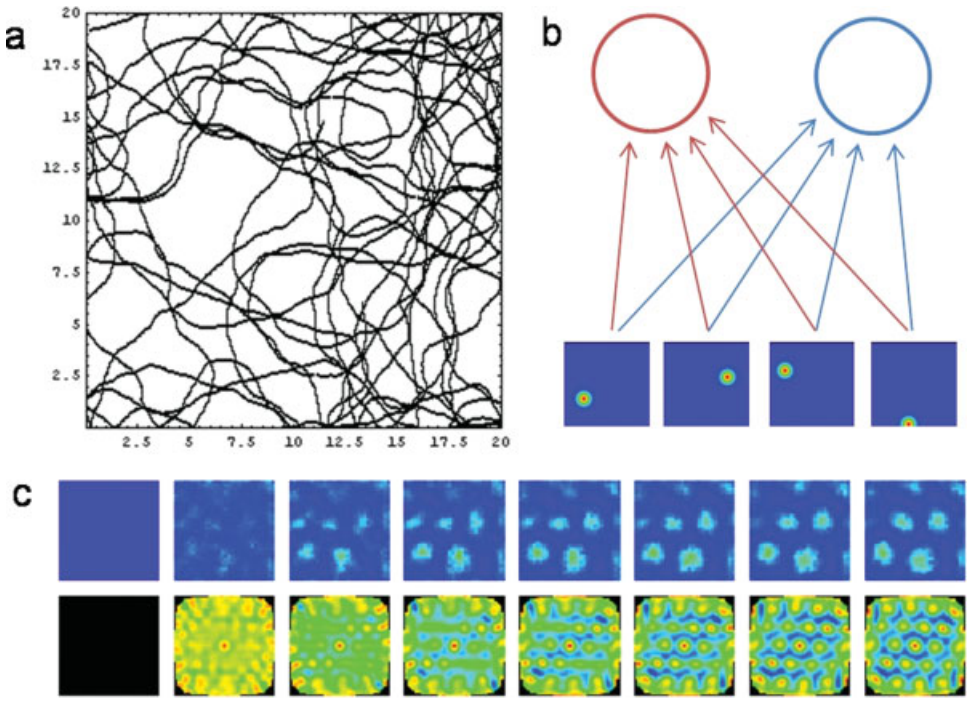

FIGURE 1. Scheme of the simulations and an example of the learning dynamics of the system. All color plots range from blue (minimum = 0 for maps and −1 for autocorrelograms) to red (maximum = ψ_{sat} for rate maps and 1 for autocorrelograms). Autocorrelation is computed only if the area of overlap between the map and its displaced version has more than 20 pixels, and drawn black otherwise. (a) An example of the trajectory of the virtual rat, lasting 10,000 time steps. At each step, the rat moves forward at a constant speed and chooses a new direction of movement close to the previous one. The only restriction in choosing a new direction is that the rat cannot get out of the environment. If so, the selection of direction is repeated until an appropriate one comes out of the draw. (b) The feedforward network, including here only two mEC neurons (the real simulations include 100). In this first version of the simulations, the would-be grid units receive inputs from 200 place-cell-like units (only four illustrated in the scheme). The place units fire whenever the rat passes through their field. If any of them is successful in stimulating some of the mEC neurons (which compete to get activated), the weight of the corresponding feedforward synapse is increased. All input synapses to a neuron (in this case either the red or the blue group) are normalized in such a way that their sum is constant. (c) Initial evolution of the map and autocorrelogram of a single unit, through the first 10^6 time steps. The weights are initially random and they evolve through hebbian learning, generating a grid-like map. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

Conclusion 1

The wavelength or spacing associated to any solution $\psi(\mathbf{x})$ is expected to vary almost linearly with the time scale τ_L of firing rate decay due to adaptation, in the corresponding units. The wavelength is observed to vary with recording location in the entorhinal cortex (Brun et al., 2008) and the model predicts (or rather, had earlier predicted, Treves et al., 2005) an almost proportional variation in the average adaptation time constant τ_L (Giocomo et al., 2007).

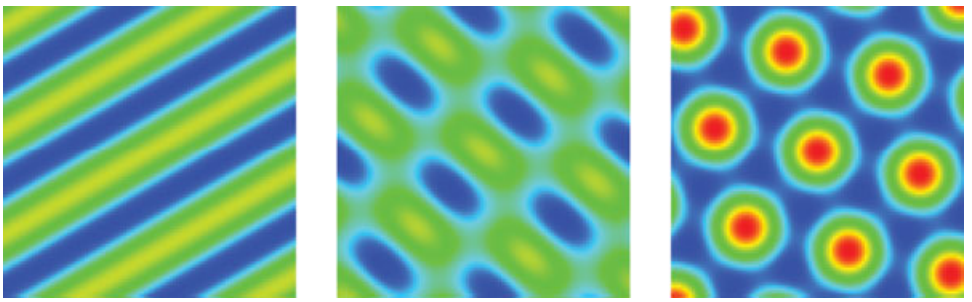

FIGURE 2. The three solutions with simple periodicity that minimize the cost function discussed in the Appendix: unidimensional, rhomboid, and triangular. Note that the color scale is the same for all graphs, but ψ_3 has higher maxima. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

Conclusion 2

There are three possible solutions $\psi(\mathbf{x})$ with simple spatial periodicity in two dimensions, either linear, rhomboid, or triangular, as shown in Figure 2. The triangular solution

$$\psi_3(\mathbf{x}) = (2/3)\sum_{i=1}^{3}\cos(\mathbf{k}_i \cdot \mathbf{x}) + 1$$

with $\mathbf{k}_i = k^*\{\cos(2\pi i/3 + \phi), \sin(2\pi i/3) + \phi\}$ (where the phase ϕ defines the orientation of the grid) is the one which is favored, in terms of reaching the optimal compromise between representing the continuity of spatial inputs and allowing for the lull in firing imposed by adaptation. This solution also contributes the most information, relative to metabolic costs, about the position of the animal in the environment (see Appendix).

NUMERICAL SIMULATIONS OF mEC UNITS WITH PLACE UNITS AS INPUTS

Does the analysis above point at the type of solution that prevails in actual simulations, as indicated by the unit in Figure 1c, or is that unit an aptly chosen rare example? Figure 3 shows the firing fields of a larger, randomly selected sample of mEC units, as they stabilize after running the virtual rat for 10^7 time steps, roughly corresponding to 10^4 lengths of the square environment. We do not claim this to be the optimal running speed or learning rate, and indeed understanding the time scales necessary for learning is an interesting issue that we leave for further study. The gridness score (Sargolini et al., 2006) of the stabilized fields is calculated from a ring-shaped cropping of the autocorrelogram including the six maxima that are closer to the center. The ring is rotated 30°, 60°, 90°, 120°, and 150°, and for every case the Pearson's correlation with the unrotated map is obtained. If C_α is this result for the rotation angle α, the gridness index is $1/2(C_{60} + C_{120}) - 1/3(C_{30} + C_{90} + C_{150})$. The histogram with the distribution of gridness scores across the mEC population (Fig. 3c) has values in line with those corresponding to the different layers of the real entorhinal cortex, though presenting a higher average gridness. Based on the same ring extracted from the autocorrelogram, the position of the three maxima above the horizontal axis is obtained and their position with respect to the center plotted for the whole population in Figure 3d, such that $3 \times N_{\mathrm{mEC}} = 300$ points are included in the plot. Note that the environment is 20×20 arbitrary units in size, and therefore, the autocorrelogram is related to displacements that span a 40×40 space. The figure shows that the grid fields of *all* different units spontaneously acquire triangular symmetry, extending fairly long-range, and that they present similar spacing, but random orientation.

The gridness of the fields observed in simulations thus validates Conclusion 2 drawn from the theoretical analysis. To test Conclusion 1, we have run several simulations with the exact same network but varying the parameters b_1 and b_2, keeping the proportion $b_2 = b_1/3$. The inverse of these parameters roughly controls the rise and decay time of firing rate adaptation, similarly to what is modeled by the parameters τ_S and τ_L in the Appendix. Figure 4 shows a monotonic increase of grid spacing with the inverse of b_1, in line with the increase in τ_S and τ_L described by Eq. (A4) and following ones.

Finally, we have run simulations to test whether the triangular grid pattern depends significantly on providing a homogeneous distribution of input fields, or else on running trajectories at fixed speed. We have run a first simulation (Fig. 5a–c) where the place fields used as inputs were concentrated near the south wall of the environment. Strikingly, we observed no difference in the individual rate maps or in the overall distribution of grid fields in the environment. In another simulation, the speed of the rat was continuously changed during training, using a random walk acceleration that produced smooth changes (Fig. 5d,e). Since we wanted the same field to be traversed with different speeds every time, we made the speed variation slow enough not to be concentrated in a single pass through a field. The typical distance of variation was rather comparable with the size of the environment, i.e., the acceleration was of the order of the mean square speed divided by the size of the environment. Again, we observed no difference. To compare the maps resulting from different conditions (Fig. 5e, Columns 1–4), we obtained the Pearson correlation coefficient between maps over the whole population of mEC cells. For the training versus testing with variable speed condition, the maps were correlated with a coefficient of $R = 0.98$, for training versus running slow $R = 0.94$ and for training versus running fast $R = 0.98$.

MORE GENERAL SPATIALLY MODULATED UNITS AS INPUTS

It appears from the previous sections that grid maps may result from a learning process with place-cell-like inputs. Although this could be construed as a model of the formation of grid maps in layers V and VI of mEC, which receive their main input from CA1, we had expected our adaptation model to lead to similar grid patterns also when fed with more general kinds of spatial input (e.g., visual, as in Franzius et al., 2007). Given the lack of experimental evidence for any definite form of spatial map in the cortical inputs to mEC, we used for simulations a fairly generic construction, including units whose 'maps' are composed by the sum of 20 Gaussian bumps of unitary standard deviation, randomly dispersed over the environment, each map normalized so as to have a maximum firing equal to ψ_{sat}. Some examples of such input units are shown in Figure 6.

The poor results obtained in our first attempts confirmed that place-cell-like fields are better suited as input for our network. By including many more inputs to the network ($N_I = 2{,}500$), however, we were able to achieve an acceptable distri-

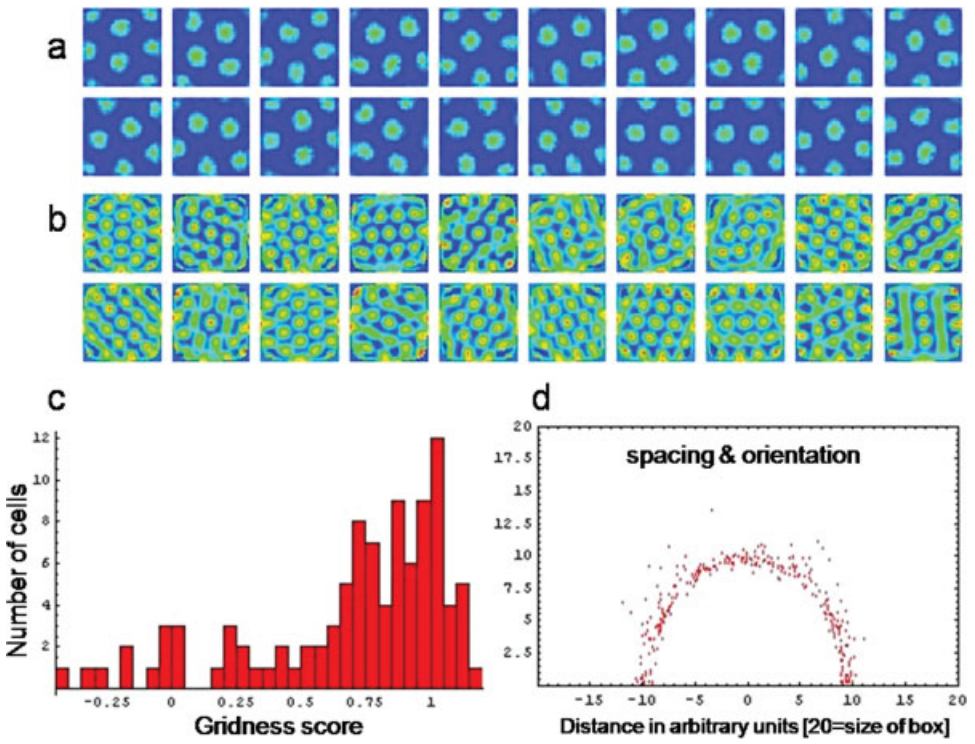

FIGURE 3. Numerical simulations where mEC units receive inputs from place units and self-organize feedforward weights based on adaptation dynamics alone, leading to grids broadly similar to experimentally observed ones. (a) Fields of a random set of 20 out of 100 mEC neurons; (b) the corresponding autocorrelograms; (c) histogram with the distribution of the gridness score across the entire mEC population; (d) orientation and spacing of the maxima (the position with respect to the center of the autocorrelogram of each of the first three maxima is plotted as an individual point), showing similar spacing, but random orientation of the grids across the population. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

bution of gridness score after learning, as shown in Figure 6. Note that while the inputs have irregular fields, the multiple fields that result from learning are roundish in shape, even for those maps with a very poor gridness score.

THE ALIGNMENT OF GRID FIELDS

The simulations shown above result in grid fields with a common spacing and random orientation. Experiments suggest, however, that grid cells that belong to the same local network should share spacing and orientation (Hafting et al., 2005). We hypothesize that the latter kind of coherence is provided by excitatory collaterals in mEC. The role of these connections might also be crucial to yield the exact same map (modulo translations and rotations) for every environment (Fyhn et al., 2007), since in the absence of a common orientation whether individual cells exhibit the same map or not remains ill-defined.

Let us assume that a strong collateral connection exists going from neuron A to neuron B in mEC. This connection favors the two neurons to have close-by fields, in any environment, but in a *sequence*, where firing by A will always be followed by B. In a two-dimensional environment, however, there is an intrinsic ambiguity (Fig. 7a): if the connectivity favors a sequence where B follows after A, the favored position of the

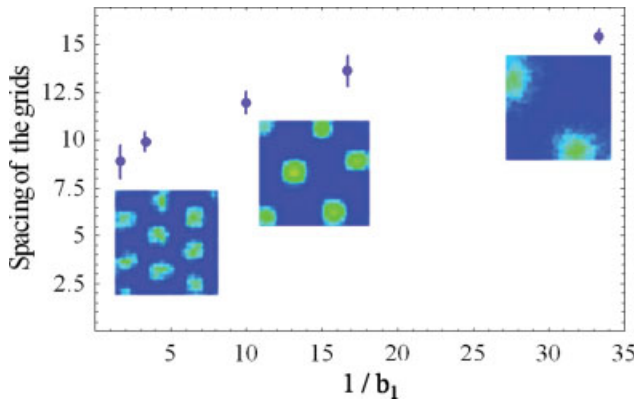

FIGURE 4. For the exact same network and inputs, we have varied the parameter b_1 keeping $b_2 = b_1/3$. These parameters control the adaptation rate and can be roughly associated to the inverse of τ_S and τ_L, introduced in the Appendix. In a qualitative agreement with the analytical results, the spacing of simulated grid units increases monotonically with the inverse of b_1. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

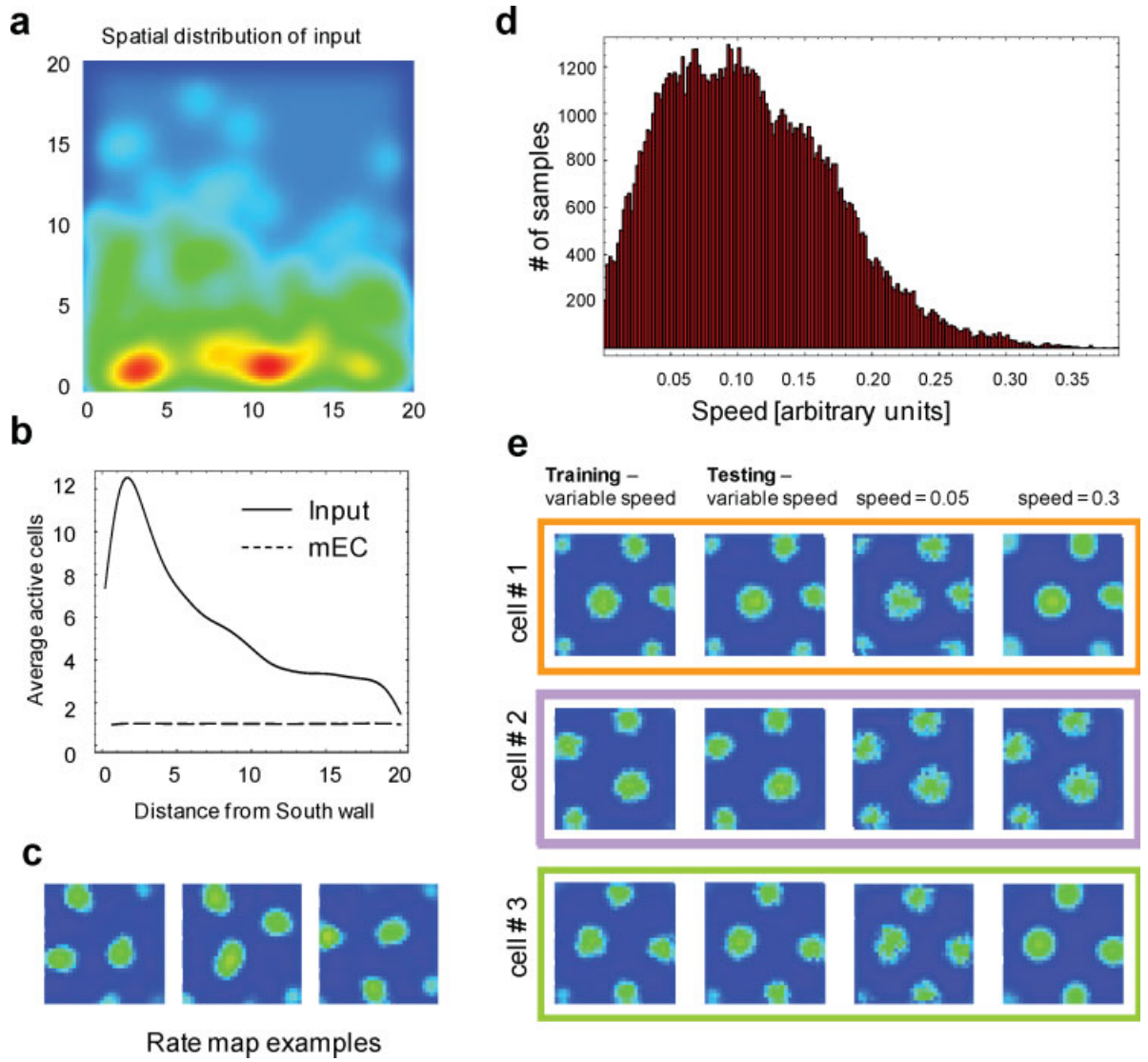

FIGURE 5. Grid units are not sensibly affected by either an inhomogeneous spatial distribution of inputs or changes in the speed of the simulated rat. (a) Spatial distribution of place-cell-like inputs used for this simulation (Panels a–c). The density of input fields close to the South wall is approximately four times higher than that near the North wall. (b) Average profile going from South (left) to North (right) of the figure in (a). In addition, the activity profile averaged over all mEC cells after learning (dashed line; the result is multiplied by a factor of 10 to make the two curves comparable). Unlike the input distribution, that of grid fields is almost flat. (c) Three examples of rate maps corresponding to mEC cells after learning. No sensible difference in the peaks is observed as a function of their position with respect to the South wall. (d) In a different simulation, histogram of a sample of speeds used to train the virtual rat. The speed was modified through a smooth random walk. The acceleration was controlled so that the change in velocity from one extreme to the other of the distribution could not occur within a single field (see text). (e) Rate maps of three cells during training and testing, at different speeds. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

field B, though always close to the field A, obviously depends on the direction of the rat while passing through it. Summing all possible contributions would result in a field for neuron B that is a ring around the field of neuron A.

One possible way to disambiguate this situation is the following. If head direction modulation is introduced, either in the firing of cells or in the efficiency of synapses, the connection from A to B is approximately associated with a specific direction in the two-dimensional environment, as shown in Figure 7b. Interestingly, this kind of modulation has been reported in Layers III, V, and VI of mEC (Sargolini et al., 2006). In addition, it has been reported that when the hippocampus is inactivated, grid cells in Layer III loose their characteristic field and a substantial number of them gain a strong head directional preference, while the firing of close-by head direction cells remains virtually unchanged (Bonnevie et al., 2006).

To test whether or not collateral connections in mEC may be able to align the grid fields, we performed simulations of a

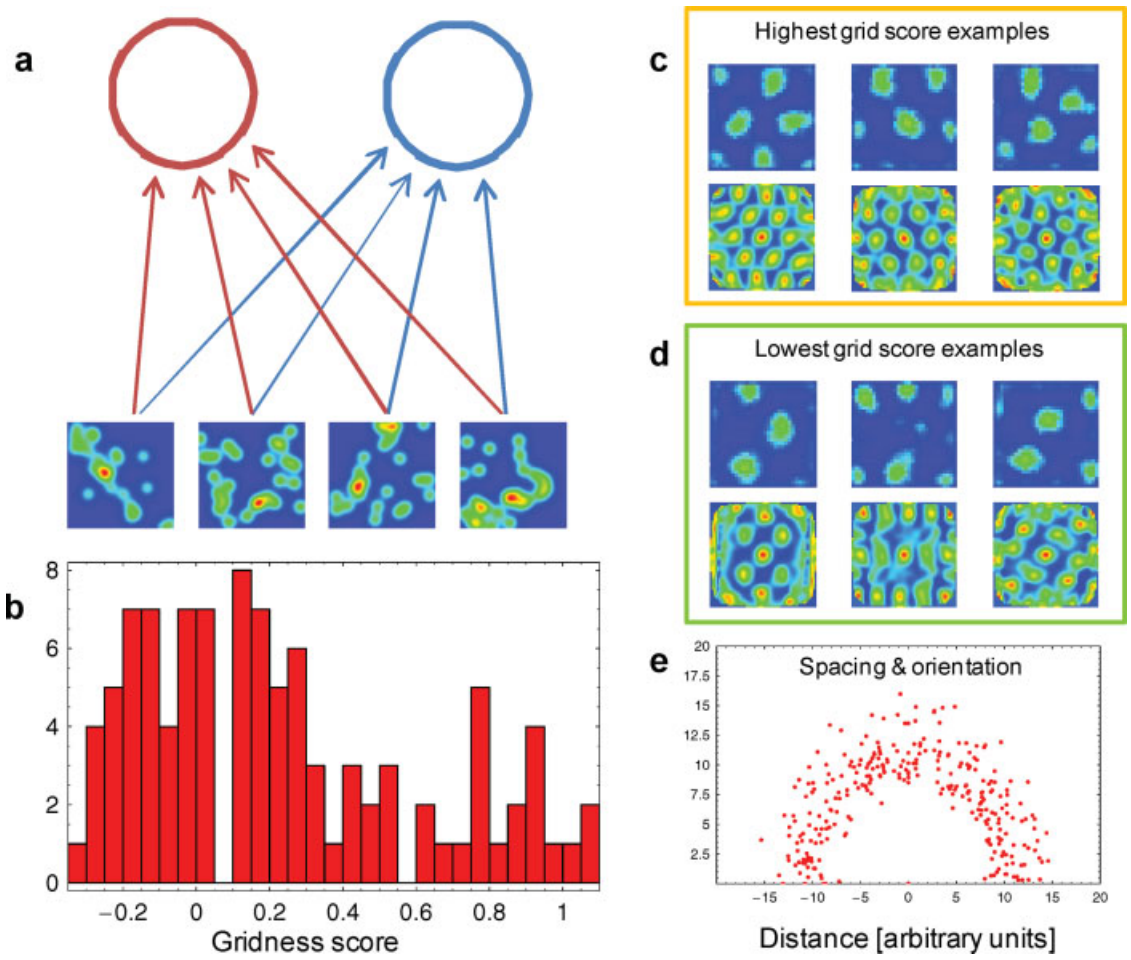

FIGURE 6. Simulations of a network with 100 mEC neurons and 2,500 spatially modulated input neurons. (a) The spatial modulation of the inputs is a sum of 20 Gaussians widespread across the environment. (b) Gridness score histogram for the asymptotic learning states. Sixteen out of 100 cells pass the criteria of a score higher than 0.75. (c, d) Examples of maps (top) and the corresponding autocorrelograms (bottom) of cells at both extremes of the histogram in (b). (e) The spacing and orientation plot shows a less well defined typical spacing. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

network with fixed, ad hoc-assigned collateral weights. One could conceive of these weights as having been adjusted through hebbian plasticity mechanisms during an extended training period, in one or several environments. To model the result of such a process, we do the following:

1. Assign to each collateral connection a common initial weight J_0.
2. Assign to each neuron in mEC an imaginary place field in a small environment. Such a field is only used for the purpose of assigning weights and then discarded.
3. Multiply the weight of the synapse going from neuron A to neuron B by a factor, exponentially decaying with the distance between the corresponding imaginary fields, for all pairs {A,B}.
4. Assign to each neuron a preferred head direction and an angular variance, resulting in a tuning function such as the one shown in Figure 7b. During simulations, this tuning function will modulate the collateral input to the corresponding neuron.
5. Multiply the weight of the synapse going from neuron A to neuron B by the overlap between the previously assigned head direction tuning functions, in such a way that its value will be high only if the difference between the two preferred head directions is small.

The resulting set of fixed collateral weights has a structure similar to the one that could be expected to result from a long hebbian learning process. We performed simulations similar to those described in the previous sections with the only addition of the collateral network. Figure 7c shows a histogram of the gridness of resulting fields in mEC. Though the average gridness is much lower than that in the distribution showed in Figure 3c—indicating that recurrent connections in our network slightly impair gridness rather than promoting it—there are still 30 out of 100 neurons with gridness score higher than 0.75,

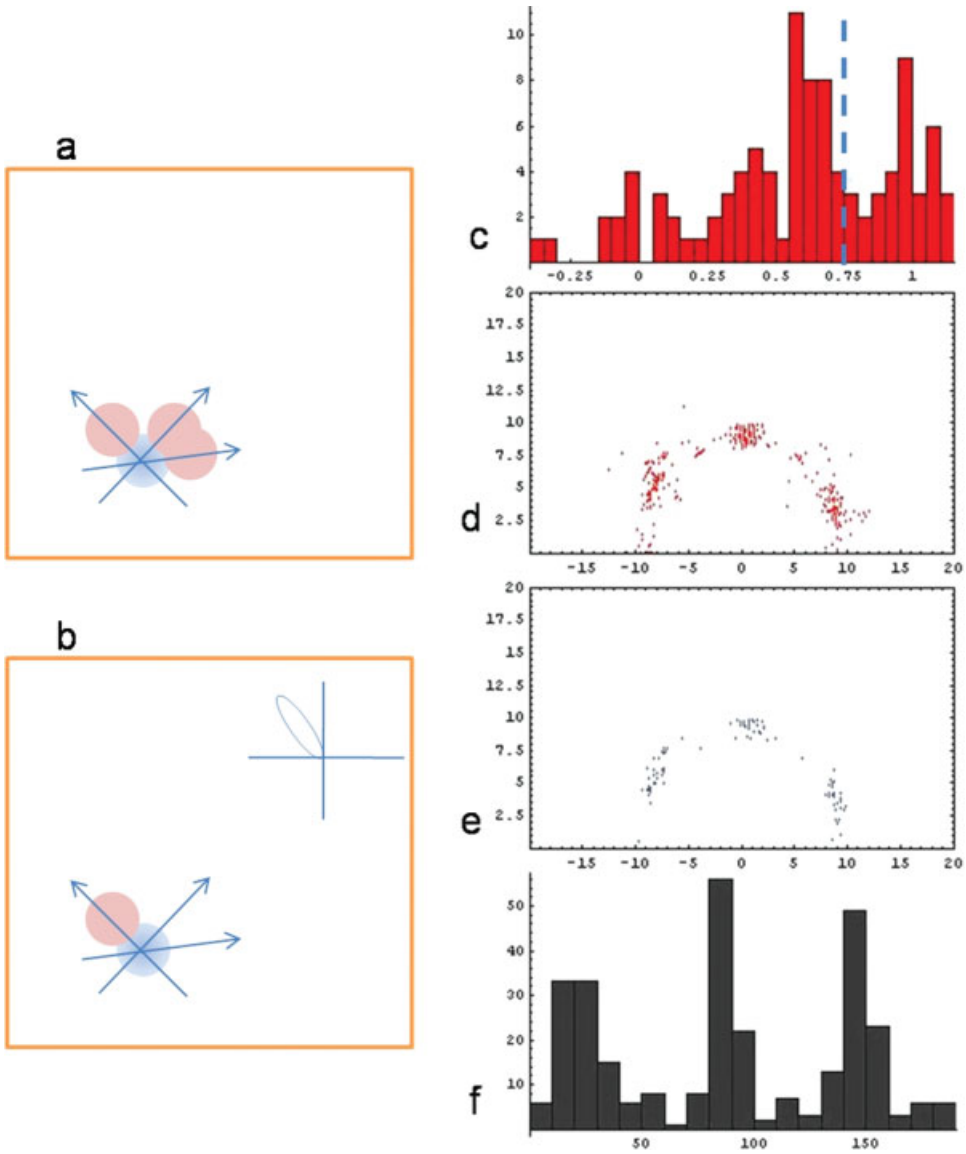

FIGURE 7. Simulations with excitatory collateral connections. (a, b) Head direction can disambiguate the sequence of grid fields of units interacting through collateral connections. We assume an established field for neuron A (in blue) and favor through head direction modulation a preferred location for the establishment of the closest field of a neuron B (in pink) that receives a strong input from A. (c) The distribution of gridness score in the population of 100 mEC neurons has a lower mean than the one shown in Figure 3, but is still comparable to experimental ones (Sargolini et al., 2006). (d, e) A representation of the spacing and orientation of grid fields obtained by plotting for each neuron the position of the three peaks close to the center of the autocorrelogram that stand above the horizontal axis, for all neurons in (d) and for those with high gridness score in (e). A clustering into three preferred orientations is observed, corresponding to the alignment of the resulting grids, in agreement with experimental findings. (f) Histogram of the angles plotted in (e). [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

while the whole distribution is still comparable to experimentally obtained ones. The plot of spacing and orientation for this data is shown in Figures 7d (for the whole population of mEC neurons) and 7e (for the 30 neurons with the highest gridness score), showing clusters corresponding to the three maxima in the upper half of each autocorrelogram that are closest to the center. A second way to visualize the clustering in the distribution of angles in Figure 7e is the histogram shown in Figure 7f.

CONCLUDING REMARKS

Grid cells are such a beautiful and unique phenomenon in the nervous system that it is tempting to regard them as a crucial element of its design, and to try to understand the organization of at least the portion of the nervous system where they are found, the rodent entorhino-hippocampal complex, by focusing on its capacity to express and utilize grid cells. A similar

enthusiasm arose earlier in connection with the discovery of place cells (O'Keefe and Dostrovsky, 1971; O'Keefe, 1976). For place cells, it is yet unclear whether they have anything to do with the organization of the mammalian hippocampus (see e.g., Treves et al., 1992). The mammalian hippocampus has preserved a strikingly clear and self-similar design, and several mammals present hippocampal cells with strikingly clear place fields, but the two phenomena may well be unrelated. The possibility should be entertained, therefore, that also in the case of grid cells the way they structure their activity in space may be unrelated to network design.

We have presented a model of grid field formation that is alternative to the ones found in the literature. The main difference is that in our model the grid field emerges from the contrast between the continuity of space, as expressed in slowly varying sensory inputs, and the fatigue rapidly decreasing neuronal firing, rather than out of the integration of proprioceptive or other self-motion-based measures of velocity. The gradual formation of the grid field is governed by hebbian learning in the feedforward connection weights in our competitive network model.

May the brain still utilize grid cells for path integration, or express path integration through grid cells? These, especially the second one, remain likely possibilities, irrespective of whether triangular grids are in any sense optimal or strictly associated with path integration. Path integration might involve a system of networks in which grid cells participate, just like Olympic games are reported extensively through the media, even though individual BBC journalists are not necessarily in optimal athletic form themselves. Sophisticated models of path integration had in fact been developed before the discovery of grid cells (Samsonovich and McNaughton, 1997), while neural fatigue and synaptic plasticity have been shown in networks models without grid cells to produce simple but effective forms of path integration (Mehta, 2001), which lead a virtual rat to predict its future position, even in a 2D environment (Treves, 2004).

Nevertheless, some aspects of the network may be closely related to some properties of grid cells. From the results presented in Figure 7, it can be concluded that it is possible to align the grids using collateral connections. Furthermore, it may well be that the network of collateral connections, if extensive, may be crucial in ensuring the smooth continuity of the putative single attractor state generated by a local network of grid cells, rather than a multiplicity of attractor states as in the hippocampus (Battaglia and Treves, 1998). The departure of attractor states generated by networks of finite size from the ideal notion of a continuous attractor has been noted early (Tsodyks and Sejnowski, 1995), but it has only recently emerged as a key issue in computational neuroscience (Hamaguchi and Hatchett, 2006; Papp et al., 2007; Roudi and Treves, 2008).

We leave for future reports simulations that use learned collateral weights rather than fixed ones, in line with the notion of a slow collateral learning that is independent of the environment, as opposed to a rapid feedforward learning that relates grid fields (the universal map) to each particular environment. Other possible follow-ups include the analysis of global remapping in our model, as investigated in the rat (Fyhn et al., 2007); and the association of mEC cells with path integration cues, as suggested by several authors (Burak and Fiete, 2006; Fuhs and Touretzky, 2006; McNaughton et al., 2006; Guanella and Verschure, 2007), in such a way that the grid can be activated in the absence of strong or familiar sensory input, as shown in experiments with relative sensory deprivation or novelty (Hafting et al., 2005).

Acknowledgments

We are grateful for many and most helpful discussions with all colleagues at the Kavli Institute, including the advance discussion of their data before publication (an opportunity missed by AT, who failed to run the model correctly for over 3 years). Arindam Biswas and Erika Cerasti participated in the early phases of the project.

REFERENCES

Ainge JA, Langston RF, Moser M-B, Moser EI. 2008. Ontogeny of spatial representations in CA1 pyramidal neurons of the rat. Poster presented at the 6th FENS Forum of European Neuroscience, Geneva, Switzerland, July 12–16, 2008, FENS Abstr. Vol 4, 128.17.

Amit DJ. 1989. Modelling Brain Function: The World of Attractor Networks. New York: Cambridge University Press.

Barlow JS. 1964. Inertial navigation as a basis for animal navigation. J Theor Biol 6:76–117.

Barry C, Hayman R, Burgess N, Jeffery KJ. 2007. Experience-dependent rescaling of entorhinal grids. Nat Neurosci 10:682–684.

Battaglia FP, Treves A. 1998. Attractor neural networks storing multiple space representations: A model for hippocampal place fields. Phys Rev E 58:7738–7753.

Bienenstock EL, Cooper LN, Munro PW. 1982. Theory for the development of neuron selectivity: Orientation specificity and binocular interaction in visual cortex. J Neurosci 2:32–48.

Bingman VP, Sharp PE. 2006. Neuronal implementation of hippocampal-mediated spatial behavior: A comparative evolutionary perspective. Behav Cogn Neurosci Rev 5:80–91.

Blair HT, Welday AC, Zhang K. 2007. Scale-invariant memory representations emerge from Moire interference between grid fields that produce theta oscillations: A computational model. J Neurosci 27:3211–3229.

Blair HT, Gupta K, Zhang K. 2008. Conversion of a phase- to a rate-coded position signal by a three-stage model of theta cells, grid cells, and place cells. Hippocampus 18:1239–1255.

Bonnevie T, Fyhn M, Hafting T, Moser E, Moser M-B. 2006. Misalignment of entorhinal grid fields after hippocampal inactivation. Soc Neurosci Abstr 32:68.1/BB9.

Brun VH, Otnass MK, Molden S, Steffenach HA, Witter MP, Moser M-B, Moser EI. 2002. Place cells and place recognition maintained by direct entorhinal–hippocampal circuitry. Science 296:2243–2246.

Brun VH, Solstad T, Kjelstrup KB, Fyhn M, Witter MP, Moser EI, Moser M-B. 2008. Progressive increase in grid scale from dorsal to ventral medial entorhinal cortex. Hippocampus 18:1200–1212.

Burak Y, Fiete I. 2006. Do we understand the emergent dynamics of grid cell activity? J Neurosci 26:9352–9354.

Burgess N. 2008. Grid cells and theta as oscillatory interference: Theory and predictions. Hippocampus 18:1157–1174.

Burgess N, Barry C, O'Keefe J. 2007. An oscillatory interference model of grid cell firing. Hippocampus 17:801–812.

Cerasti E, Treves A. 2006. Coherent population dynamics in a grid cell network model. Soc Neurosci Abstr 32:68.12.

Eichenbaum H, Lipton PA. 2008. Towards a functional organization of the medial temporal lobe memory system: Role of the parahippocampal and medial entorhinal cortical areas. Hippocampus 18: 1314–1324.

Fiete IR, Burak Y, Brookings T. 2008. What grid cells convey about rat location. J Neurosci 28:6858–6871.

Franzius M, Sprekeler H, Wiskott L. 2007. Slowness and sparseness lead to place, head-direction, and spatial-view cells. PLoS Comput Biol 3:e166. doi:10.1371/journal.pcbi.0030166.

Fuhs MC, Touretzky DS. 2006. A spin glass model of path integration in rat medial entorhinal cortex. J Neurosci 26:4266–4276.

Fyhn M, Molden S, Witter MP, Moser EI, Moser M-B. 2004. Spatial representation in the entorhinal cortex. Science 305:1258–1264.

Fyhn M, Hafting T, Treves A, Moser M-B, Moser EI. 2007. Hippocampal remapping and grid realignment in entorhinal cortex. Nature 446:190–194.

Fyhn M, Hafting T, Witter MP, Moser EI, Moser M-B. 2008. Grid cells in mice. Hippocampus 18:1230–1238.

Giocomo LM, Hasselmo ME. 2008. Computation by oscillations: Implications of experimental data for theoretical models of grid cells. Hippocampus 18:1186–1199.

Giocomo LM, Zilli EA, Fransen E, Hasselmo ME. 2007. Temporal frequency of subthreshold oscillations scales with entorhinal grid cell field spacing. Science 315:1719–1722.

Guanella A, Verschure PF. 2007. Prediction of the position of an animal based on populations of grid and place cells: A comparative simulation study. J Integr Neurosci 6:433–446.

Hafting T, Fyhn M, Molden S, Moser M-B, Moser EI. 2005. Microstructure of a spatial map in the entorhinal cortex. Nature 436: 801–806.

Hafting T, Fyhn M, Moser M-B, Moser EI. 2006. Phase precession and phase locking in entorhinal grid cells. Soc Neurosci Abstr 32:68.8.

Hafting T, Fyhn M, Bonnevie T, Moser M-B, Moser EI. 2008. Hippocampus-independent phase precession in entorhinal grid cells. Nature 453:1248–1252.

Hamaguchi K, Hatchett JPL. 2006. Analytic solution of neural network with disordered lateral inhibition. Phys Rev E 73:051104.

Hasselmo ME. 2008. Grid cell mechanisms and function: Contributions of entorhinal persistent spiking and phase resetting. Hippocampus 18:1213–1229.

Hayman R, Jeffery KJ. 2008. How heterogeneous place cells responding arises from homogeneous grids - a contextual gating hypothesis. Hippocampus 18:1301–1313.

Hopfield JJ. 1982. Neural networks and physical systems with emergent collective computational abilities. Proc Natl Acad Sci USA 79:2554–2558.

Jeewajee A, Barry C, O'Keefe J, Burgess N. 2008. Grid cells and theta as oscillatory interference: electrophysiological data from freely moving rats. Hippocampus 18:1175–1185.

Kahn MC, Siegel JJ, Jechura TJ, Bingman VP. 2008. Response properties of avian hippocampal formation cells in an environment with unstable goal locations. Behav Brain Res 191:153–163.

Kropff E, Treves A. 2007. Uninformative memories will prevail: the storage of correlated representations and its consequences. HFSP J 1:249–262. http://hfspj.aip.org/doi/10.2976/1.2793335.

Langston RF, Ainge JA, Moser EI, Moser M-B. 2008.Ontogeny of spatial representations in the dorsomedial entorhinal cortex of the rat. Poster presented at the 6th FENS Forum of European Neuroscience, Geneva, Switzerland, July 12–16, 2008, FENS Abstr. Vol 4, 128.18.

Leutgeb S, Leutgeb JK, Barnes CA, Moser EI, McNaughton BL, Moser M-B. 2005. Independent codes for spatial and episodic memory in hippocampal neuronal ensembles. Science 309:619–623.

Marr D. 1971. Simple memory: A theory for archicortex. Philos Trans R Soc Lond B Biol Sci 262:23–81.

Martin PD, Berthoz A. 2002. Development of spatial firing in the hippocampus of young rats. Hippocampus 12:465–480.

McNaughton BL, Morris RGM. 1987. Hippocampal synaptic enhancement and information storage within a distributed memory system. Trends Neurosci 10:408–415.

McNaughton BL, Battaglia FP, Jensen O, Moser EI, Moser M-B. 2006. Path integration and the neural basis of the 'cognitive map'. Nat Rev Neurosci 7:663–678.

Mehta MR. 2001. Neuronal dynamics of predictive coding. The Neuroscientist 7:490–495.

Molter C, Yamaguchi Y. 2008. Entorhinal theta phase precession sculpts dentate gyrus place fields. Hippocampus 18:919–930.

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

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

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

Papp G, Witter MP, Treves A. 2007. The CA3 network as a memory store for spatial representations. Learn Mem 14:732–744.

Rolls ET. 1989. Functions of neuronal networks in the hippocampus and cerebral cortex in memory. In: Cotterill R, editor. Models of Brain Function. Cambridge, UK: Cambridge University Press. pp 15–33.

Rolls ET. 1999. Spatial view cells and the representation of place in the primate hippocampus. Hippocampus 9:467–480.

Rolls ET, Stringer SM, Elliot T. 2006. Entorhinal cortex grid cells can map to hippocampal place cells by competitive learning. Network 15:447–465.

Roudi Y, Treves A. 2008. Representing where along with what information in a model of a cortical patch. PLoS Comput Biol 4:e1000012. doi:10.1371/journal.pcbi.1000012.

Samsonovich A, McNaughton BL. 1997. Path integration and cognitive mapping in a continuous attractor neural network model. J Neurosci 17:272–275.

Sargolini F, Fyhn M, Hafting T, McNaughton BL, Witter MP, Moser M-B, Moser EI. 2006. Conjunctive representation of position, direction and velocity in entorhinal cortex. Science 312:754–758.

Savelli F, Yoganarasimha D, Knierim JJ. 2008. Influence of boundary removal on the spatial representations of the medial entorhinal cortex. Hippocampus 18:1270–1282.

Skaggs WE, McNaughton BL, Gothard K, Markus E. 1993. An information theoretic approach to deciphering the hippocampal code. Adv Neural Inf Process Syst 5:1030–1037.

Solstad T, Moser EI, Einevoll GT. 2006. From grid cells to place cells: A mathematical model. Hippocampus 16:1026–1031.

Treves A. 2004. Computational constraints between retrieving the past and predicting the future, and the CA3–CA1 differentiation. Hippocampus 14:535–556.

Treves A, Miglio O, Parisi D. 1992. Rats, nets, maps and the emergence of place cells. Psychobiology 20:1–8.

Treves A, Kropff E, Biswas A. 2005. On the triangular grid of entorhinal place fields. Soc Neurosci Abstr 31:198.11.

Tsodyks M, Sejnowski T. 1995. Associative memory and hippocampal place cells. Int J Neural Syst 6 (Suppl):81–86.

Ulanovsky N, Moss CF. 2007. Hippocampal cellular and network activity in freely moving echolocating bats. Nat Neurosci 10:224–233.

Wills TJ, Lever C, Cacucci F, Burgess N, O'Keefe J. 2005. Attractor dynamics in the hippocampal representation of the local environment. Science 308:873–876.
Wills TJ, Cacucci F, O'Keefe J. 2008. Ontogenetic of the spatial properties of hippocampal place cells and entorhinal cortex grid cells in the rat. Poster presented at the 6th FENS Forum of European Neuroscience, Geneva, Switzerland, July 12–16, 2008, FENS Abstr. Vol 4, 225.29.
Wiskott L. 2003. Slow feature analysis: A theoretical analysis of optimal free responses. Neural Comput 15:2147–2177.

APPENDIX: ASYMPTOTIC STATES

To analyze the possible asymptotic states reached by the map expressed by a single unit, we assume the following:

1. The single-unit map is indefinitely *periodic* in space, to be able to represent, with the same population of units, environments of arbitrary shape and size (we expect a periodic solution to be favored anyway by collateral connections, such as those introduced in the last simulation of this paper).
2. While units physically located close to each other in the cortex express maps $\psi_i(\mathbf{x})$ with similar characteristics, remote units express maps with the same shape, generated by the same developmental process, but allowing for a size rescaling, reflecting different biophysical parameters prevailing at different locations in the tissue [e.g., the gradient in the dynamical properties of stellate cells along the dorsoventral axis (Giocomo et al., 2007)].
3. The shape of the single unit map $\psi_i(\mathbf{x})$ optimizes the representation of continuous space by *minimizing the square gradient*, $(\overrightarrow{\nabla}\psi_i(\mathbf{x}))^2$. The continuity in the representation stems, in our simulations, from the smoothly varying inputs to grid units, further smoothed by hebbian associative learning.
4. The shape of the single-unit firing map also reflects firing rate *adaptation* in the sense of minimizing the integral $\langle\int dt'\psi(\mathbf{x}(t))K(t - t')\psi(\mathbf{x}(t'))\rangle$, where $K(\Delta t)$ is a kernel, of appropriate strength, quantifying the reluctance to fire a spike at time t if one has been fired at time $t - \Delta t$, and the average $\langle\ldots\rangle$ is over all trajectories and speeds experienced during training. Note that if adaptation is negligible, continuity would tend to make a cell fire all along the environment. If, in contrast, adaptation is very strong, the emerging fields are expected to be small, and not to stabilize easily.

We can first find maps that minimize constraints (3) and (4) and then analyze whether there is one among them which also optimizes the information it conveys, e.g., by *maximizing* its variance, $\langle\psi^2(\mathbf{x}) - \langle\psi(\mathbf{x})\rangle^2\rangle$, or its information rate $\langle\psi(\mathbf{x})\log[\psi(\mathbf{x})/\overline{\psi}]\rangle$, or some other quantifier of information content.

Minimization of a Cost Function

We aim to find single-unit maps $\psi(\mathbf{x})$ that minimize the cost function

$$L = \int d\mathbf{x}[\nabla\psi(\mathbf{x})]^2 + \gamma\int d\mathbf{x}\int dt'\psi(\mathbf{x}(t))K(t - t')\psi(\mathbf{x}(t'))$$

where γ parameterizes the relative importance of adaptation over representational continuity. In the second term, the average over all trajectories can be taken to transform the time-difference-dependent kernel $K(\Delta t)$ into an effective position-difference-dependent one $K(\mathbf{x} - \mathbf{x}')$. We will consider different models for this spatial version of the kernel.

To derive the expected form of the asymptotic states of the training process, it is useful to remind ourselves of some basic mathematical facts. First, the integral of the function $f(x) = \sin^2(x)$ is half the one of the constant $g(x) = 1$ if integrated over the same domain, plus some 'border' corrections that depend on the particular choice of domain, and generally scale as its perimeter. Thus, if we choose a two-dimensional domain with a large area $A \gg 1$, we can consider

$$\frac{1}{A}\int_A \sin^2(\mathbf{k}\cdot\mathbf{x} + \phi)d\mathbf{x} = \frac{1}{2} + o\left(\frac{1}{\sqrt{A}}\right) \approx \frac{1}{2}$$

and for similar reasons

$$\frac{1}{A}\int_A \sin(\mathbf{k}\cdot\mathbf{x} + \phi)d\mathbf{x} = o\left(\frac{1}{\sqrt{A}}\right) \approx 0$$

neglecting all border contributions.

It is also useful to remember the orthogonal products between basis functions that give rise to the Fourier formalism. If we consider a set of different wave vectors $\{\mathbf{k}_i\}$ and constant phases $\{\phi_i\}$,

$$\frac{1}{A}\int_A \sin(\mathbf{k}_i\cdot\mathbf{x} + \phi_i)\sin(\mathbf{k}_j\cdot\mathbf{x} + \phi_j)d\mathbf{x} = \frac{1}{2}\delta_{ij} + o\left(\frac{1}{\sqrt{A}}\right) \approx \frac{1}{2}\delta_{ij}.$$

Let us now consider our cost function, normalized by the area of the integration domain (which we will consider to tend to infinity)

$$L = \frac{1}{A}\int_A d\mathbf{x}[\nabla\psi(\mathbf{x})]^2 + \frac{\gamma}{A}\int_A d\mathbf{x}\psi(\mathbf{x})\int_A d\mathbf{x}'\psi(\mathbf{x}')K(|\mathbf{x}' - \mathbf{x}|) \tag{A1}$$

and analyze a general form for the solution, decomposed into two-dimensional Fourier modes

$$\psi(\mathbf{x}) = a_0 + \sum_i a_i\cos(\mathbf{k}_i\cdot\mathbf{x} + \phi_i). \tag{A2}$$

Inserting the decomposition into the cost function, and applying the properties described above, we obtain for the first term

$$L_1 \approx \frac{1}{2}\sum_i a_i^2 k_i^2.$$

For the second term of the cost function, we use the change of variables $\mathbf{q} = \mathbf{x}' - \mathbf{x}$ and also the trigonometric property

$$\cos(\mathbf{k}_i \cdot \mathbf{q} + \mathbf{k}_i \cdot \mathbf{x} + \phi_i) = \cos(\mathbf{k}_i \cdot \mathbf{q})\cos(\mathbf{k}_i \cdot \mathbf{x} + \phi_i) - \sin(\mathbf{k}_i \cdot \mathbf{q})\sin(\mathbf{k}_i \cdot \mathbf{x} + \phi_i).$$

By construction, the integration domain is symmetric around $\mathbf{q} = 0$; so, the terms with $\sin(\mathbf{k} \cdot \mathbf{q})$ do not survive the integration over $d\mathbf{q}$, since $\sin(\mathbf{k} \cdot \mathbf{q})K(q)$ is an odd function [adaptation has obviously no preferred spatial direction and thus $K(q)$ has radial symmetry]. The second term of the cost function is simply

$$L_2 \approx \gamma \cdot a_0^2 \widetilde{K}(0) + (\gamma/2)\sum_i a_i^2 \widetilde{K}(k_i)$$

where we have introduced the two-dimensional Fourier transform of $K(q)$

$$\widetilde{K}(k_i) = \int_A d\mathbf{q} K(q)\cos(\mathbf{k}_i \cdot \mathbf{q}).$$

Taking the derivative of our cost function

$$L = L_1 + L_2 = \gamma \cdot a_0^2 \widetilde{K}(0) + \frac{1}{2}\sum_i a_i^2 [k_i^2 + \gamma \cdot \widetilde{K}(k_i)]$$

with respect to the norm of each of the basis vectors, and setting it equal to zero, yields the set of conditions

$$k_i = -(\gamma/2)\partial_{k_i}\widetilde{K}(k_i) \qquad \text{(A3)}$$

Examples of Adaptation Kernels

The first explicit model we consider is a difference of radially symmetric Gaussians

$$K(q) = \frac{1}{v\tau_L\sqrt{2\pi}}\exp\left[-\frac{q^2}{2(v\tau_L)^2}\right] - \rho\frac{1}{v\tau_S\sqrt{2\pi}}\exp\left[-\frac{q^2}{2(v\tau_S)^2}\right]$$

which expresses the hypothesis that, after averaging over trajectories and speeds, adaptation effects, which in real time become significant over time-differences τ_S and decay away after time-differences τ_L, are strongest within a ring of radius $v\tau_L$, with v an average speed parameter, but are also felt, reduced by a factor $\rho(\tau_L/\tau_S)$ (constrained to be <1), within a distance $v\tau_S$ of the current position, because the animal may sometimes stay still or move only in its immediate surrounding. If a neuron fires at a given time, the spatial region of marked adaptation is thus a ring between radii $v\tau_S$ and $v\tau_L$ around the current position, while adaptation decreases by a factor $\rho(\tau_L/\tau_S)$ inside the ring, and it decreases to zero toward the outside.

The Fourier transform of the Gaussian kernel is

$$\widetilde{K}(k_i) = \exp\left[-\frac{1}{2}(k_i v\tau_L)^2\right] - \rho\exp\left[-\frac{1}{2}(k_i v\tau_S)^2\right]$$

and Eq. (A3) becomes

$$1 = \frac{\gamma v^2}{2}\left\{\tau_L^2\exp\left[-\frac{(k_i v\tau_L)^2}{2}\right] - \rho\tau_S^2\exp\left[-\frac{(k_i v\tau_S)^2}{2}\right]\right\} \qquad \text{(A4)}$$

Since $0 < \tau_S < \tau_L$, we can analyze the two limit cases of τ_S. If $\tau_S \ll \tau_L$, the solution of Eq. (A4) is approximately

$$k^* = \frac{1}{v\tau_L}\sqrt{2\ln\left[\frac{\gamma v^2\tau_L^2}{2}\right]}$$

while if $\tau_S \approx \tau_L$

$$k^* = \frac{1}{v\tau_L}\sqrt{2\ln\left[\frac{(1-\rho)\gamma v^2\tau_L^2}{2}\right]}$$

Both situations show that the typical spacing of the solution (the inverse of k^*) is proportional to $v\tau_L$ with some logarithmic correction.

One may also consider a second example of kernel with an exponential rather than Gaussian decay:

$$K(q) = \frac{1}{v\tau_L}\exp\left[-\frac{q}{v\tau_L}\right] - \frac{\rho}{v\tau_S}\exp\left[-\frac{q}{v\tau_S}\right]$$

with transform

$$\widetilde{K}(q) = [1 + (v\tau_L)^2]^{-\frac{3}{2}} - \rho[1 + (v\tau_S)^2]^{-\frac{3}{2}}$$

The solution in this case is very similar. The two extreme cases yield $k^* = \frac{1}{v\tau_L}\sqrt{\left[\frac{3}{2}\gamma(v\tau_L)^2\right]^{\frac{2}{5}} - 1}$ and $k^* = \frac{1}{v\tau_L} \times \sqrt{\left[\frac{3}{2}\gamma(1-\rho)(v\tau_L)^2\right]^{\frac{2}{5}} - 1}$, again showing a linear dependence of the typical spacing with $v\tau_L$, modulated by a slower correction [tough in this case, for very high values of $v\tau_L$, the correction becomes significant and the scaling of the spacing is $(v\tau_L)^{\frac{3}{5}}$, a flattening that is also observed in the simulations of Fig. 4].

These equations point at the first conclusion mentioned in the main text: the wavelength of the solution should vary roughly linearly with the timescales for firing rate adaptation. Simplistic as such an analysis may be, this result is qualitatively confirmed in simulations (Fig. 4), where the parameters τ_L and τ_S are associated to the inverse of the adaptation rates b_1 and b_2.

Two-Dimensional Periodic Solutions

Taking into account that Eq. (A3) constrains all Fourier contributions to have wave vectors of norm k^*, if we further constrain the minimum of the firing rate map to be 0, and set the average activity (to unity per unit area, for simplicity), there are three possible solutions $\psi(\mathbf{x})$ that show simple spatial periodicity, either linear, rhomboid, or triangular. The first solution is effectively unidimensional, while the second and third correspond to rhomboid and triangular tessellation of the plane, respectively

$$\begin{aligned}\psi_1(\mathbf{x}) &= \cos(\mathbf{k}\cdot\mathbf{x}) + 1\\ \psi_2(\mathbf{x}) &= (1/2)\cos(\mathbf{k}'\cdot\mathbf{x}) + (1/2)\cos(\mathbf{k}''\cdot\mathbf{x}) + 1\\ \psi_3(\mathbf{x}) &= (2/3)\sum_{i=1}^{3}\cos(\mathbf{k}_i\cdot\mathbf{x}) + 1\end{aligned} \tag{A5}$$

These solutions have the form of Eq. (A2), with $a_0 = 1$ to force unitary average activity and a_i adjusted such that $\psi(\mathbf{x}) = 0$ at its minima. All wave vectors have norm k^*, but while $\mathbf{k}$, $\mathbf{k}'$, and $\mathbf{k}''$ can have any orientation, the triangular symmetry requires vectors to be $\mathbf{k}_i = k^*\{\cos(2\pi i/3 + \phi), \sin(2\pi i/3) + \phi\}$ for some constant phase ϕ that defines the orientation of the triangular grid. The three types of solution are represented in Figure 2.

The cost function associated to each solution is

$$\psi_1\colon \gamma\widetilde{K}(0) + \frac{1}{2}[(k^*)^2 + \gamma\widetilde{K}(k^*)]$$

$$\psi_2\colon \gamma\widetilde{K}(0) + \frac{1}{4}[(k^*)^2 + \gamma\widetilde{K}(k^*)]$$

$$\psi_3\colon \gamma\widetilde{K}(0) + \frac{2}{3}[(k^*)^2 + \gamma\widetilde{K}(k^*)].$$

In the region of parameters where $[(K^*)^2 + \gamma\widetilde{K}(k^*)] > 0$ any of these solutions is worse than the trivial constant solution $\psi_0 = 1$, with an associated cost function $\gamma\widetilde{K}(0)$. This region of parameters is not interesting, since it reflects no competition between the opposing factors of spatial continuity and adaptation. In the complementary and thus interesting region of parameters, where $[(k^*)^2 + \gamma\widetilde{K}(k^*)] < 0$, the solution with the lowest cost is ψ_3.

Moreover, $\psi_3(\mathbf{x})$ is also the most informative of the solutions, again because of the largest summed amplitude of the cosine terms, allowed by the fact that the sum reaches its minimum when each of the cosines takes the value $\frac{1}{2}$. This results in more variance [with the help of Eqs. (A5), it is easy to see that $\mathrm{var}(\psi_1) = 1/2$, $\mathrm{var}(\psi_2) = 1/4$ and $\mathrm{var}(\psi_3) = 2/3$], more information content as quantified by the information rate, or higher values of any index of information that is superlinear in $\psi(\mathbf{x})$ (such as the bits per spike measure first developed by Skaggs et al., 1993). Therefore, we reach the second conclusion that the triangular grid contributes the most information, and 'costs' less in terms of our cost function, among the three periodic solutions. Note that this conclusion does not depend on the particular choice of adaptation kernel $K(q)$ as long as it has an interesting region of parameters such that continuity does not prevail over adaptation, and rather the competition between these two driving forces determines a typical spacing for $\psi(\mathbf{x})$ given by $(k^*)^{-1}$.