Partial coherence and frustration in self-organizing spherical grids

***Federico Stella*[1] ORCID, *Eugenio Urdapilleta*[2,3] ORCID, *Alessandro Treves*[3,4] ORCID**

1 *Institute of Science and Technology Austria, A–3400 Klosterneuburg, Austria*
2 *Centro Atómico Bariloche & Instituto Balseiro, Comisión Nacional de Energía Atómica (CNEA) and Universidad Nacional de Cuyo (UNCUYO), Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), R8402AGP San Carlos de Bariloche, Río Negro, Argentina*
3 *SISSA - Cognitive Neuroscience, 34136 Trieste, Italy*
4 *NTNU - Centre for Neural Computation, 7030 Trondheim, Norway*

Correspondence: Alessandro Treves, Cognitive Neuroscience, SISSA, via Bonomea 265, 34136 Trieste, Italy. Email: ale@sissa.it

bioRxiv version, February 25, 2019

ABSTRACT

Nearby grid cells have been observed to express a remarkable degree of long-range order, which is often idealized as extending potentially to infinity. Yet their strict periodic firing and ensemble coherence are theoretically possible only in flat environments, much unlike the burrows which rodents usually live in. Are the symmetrical, coherent grid maps inferred in the lab relevant to chart their way in their natural habitat?

We consider spheres as simple models of curved environments and, waiting for the appropriate experiments to be performed, we use our adaptation model to predict what grid maps would emerge in a network with the same recurrent connections, which on the plane produce coherence among the units. We find that on the sphere such connections distort the maps that single grid units would express on their own, and aggregate them into clusters. When remapping to a different spherical environment, units in each cluster maintain only partial coherence, similar to what is observed in disordered materials, such as spin glasses.

Keywords: grid cells; ensemble coherence; remapping; continuous attractor; curved environment.

1. INTRODUCTION

What are the defining properties of grid cells? In the 15 years since their discovery in the medial entorhinal cortex (mEC) (Fyhn et al., 2004), two organizational principles appear to have emerged as the cornerstones of the phenomenon. The first one is, of course, the positioning of the fields of each individual cell at the vertices of a regular hexagonal tessellation of the environment (Hafting et al., 2005). The second, a strong propensity of local ensembles of these cells to maintain their co-activation patterns across conditions and environments (Fyhn et al., 2007); in striking contrast to the behavior expressed by neighboring place cells, which make the swapping of activation partners ("remapping") one of their defining features (Bostock et al., 1991). These two properties, the former expressed at the single-cell level, the latter constraining collective states of activity, have come to be regarded as quintessential to grid cells. They are often yoked together when discussing the "grid cell code" (Yoon et al., 2013; Burak, 2014; Stemmler et al., 2015), thus leaving it unclear whether such code is expressed more in the regularity of field arrangements or in the constancy of spatial phase relations, or in a necessary combination of both.

It should not be forgotten, however, that grid cells have been first described and mostly studied in flat, empty, bounded environments. Their entanglement could be possibly related to the Euclidean geometry of this very specific sort of environment, rather than being intrinsic to the cells. The question, then, is to what extent would single-cell and population properties still co-occur, when such a specific setting is abandoned, to reach for more naturalistic settings of complex, curved, partially open environments, such as the burrows where rats live in the wild (Calhoun, 1962).

In two previous modelling studies, we argued that the notion of the hexagonal grid may need to be generalized in order to predict the behavior of such cells in environments of constant positive or negative curvature. With sufficient exposure to these environments, our model indicates how single grid cells may adapt by producing regular tessellations consistent with the underlying curvature (Stella et al., 2013; Urdapilleta et al., 2015): tessellations with 5-fold or lower symmetry for positive curvature; 7-fold or higher symmetry for negative curvature. An analysis based on Calhoun's (1962) study leads to the conclusion that the standard 6-fold grid symmetry would arise at the single-cell level only when the curvature is near zero; while the actual range of curvature values of the natural Norway rat habitat extends further, both at the negative and at the positive ends of the spectrum (Fig.1).

What about the effect of interactions between grid cells? What was shown in (Stella et al., 2013) is only how a population of non-interacting grid-like units can self-organize a representation of the spherical surface, where each unit ends up displaying an independently "oriented", often quasi-regular grid. Other studies of the same model on planar environments have already made clear how the introduction of lateral connectivity in the mEC population can induce an alignment among units, resulting in a common orientation emerging from the self-organization process, while also reinforcing the symmetric arrangement of their fields (Si et al., 2012; Si and Treves, 2013). Urdapilleta

et al. (2015) have observed that also in environments with constant negative curvature the lateral interactions tend to favor a coherent arrangement of the fields across units, but such environments are dominated by their boundaries, which leads to arbitrary modeling choices, that in turn prevent reaching firm conclusions. A complete sphere, on the other hand, has no boundaries, and thus offers a conveniently simple model of a curved environment. Moreover, it has been effective as an experimental set-up, both on the outside (Harvey et al., 2009) and on the inside (Kruge et al., 2013; Kruge, 2016), giving hope that once complemented with the appropriate sensory surround, it will allow developmental studies to investigate the formation of spherical representations in rodents.

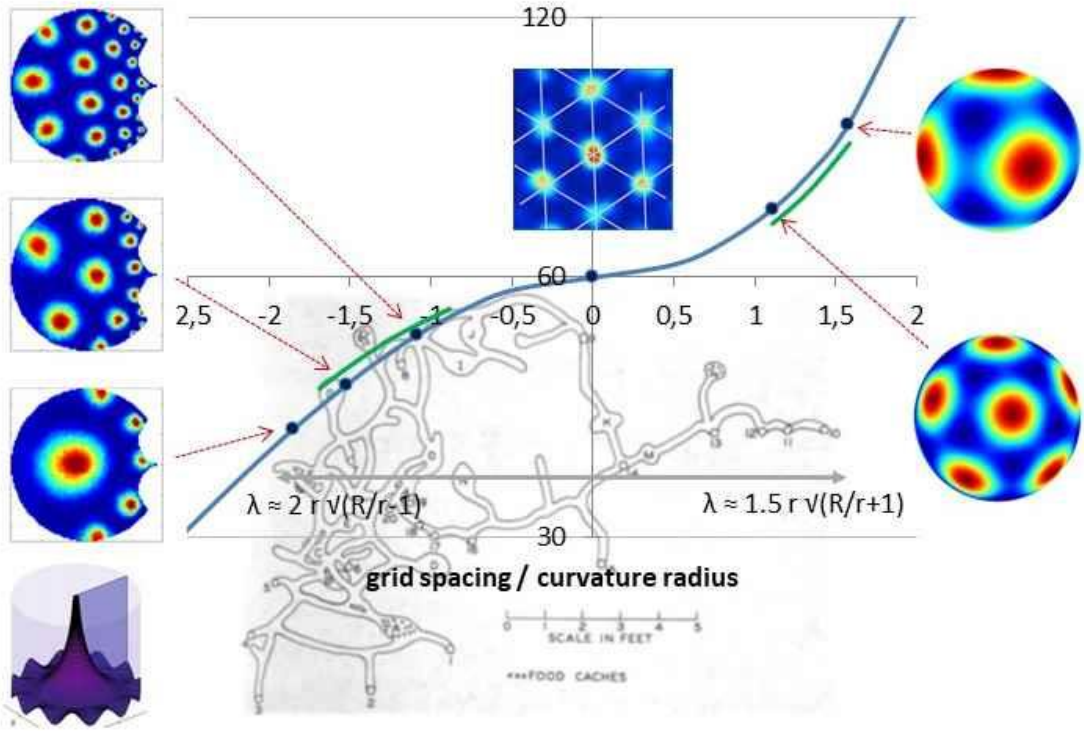

Figure 1. Natural Norway rat environments span regions of positive and negative curvature.
The symmetry expected at the single-unit level for different values of constant Gaussian curvature, measured by the ratio between grid spacing s and the radius of curvature λ. Blue curve: theoretical relation between the angle $\alpha=2\pi/n$ of the n-fold symmetry and the ratio s/λ , $\cosh(s/\lambda)=\cos(\alpha)/[1-\cos(\alpha)]$ (negative curvature) and $\cos(s/\lambda)=\cos(\alpha)/[1-\cos(\alpha)]$ (positive curvature). Green curves are extracted from simulations. The range of curvature radius values λ can be estimated from quantitative descriptions of rat burrows, such as the one superimposed on the diagram, from Calhoun (1962), by estimating the inner r and outer R radii of circularly bent tunnels. The gray arrows indicate a minimal range, considering a small s=40cm grid spacing.

In the meantime, here we use the adaptation model to study the effect of collateral interactions among grid-like units self-organizing on a spherical world, as follows.

2. THE MODEL

The basic details of the model are identical to (Stella et al., 2013), and are described in Appendix 1, with the critical addition of a set of recurrent collaterals connecting units of the EC layer.

Time is discretized in steps of length t = 0.01s. The total length of a simulation is of 100 million steps (nearly 12 days of continuous running, to ensure the self-organization process has approached its asymptote). The virtual rat moves on the surface of a sphere of radius 52.6cm with a constant speed of v = 40cm/s. For simplicity, the change in running direction between two consecutive steps of the virtual rat is sampled from a Gaussian distribution with zero mean and standard deviation h = 0.2 radians. The virtual rat always runs along the great circle determined by its running direction. Our model is comprised of two layers. The input array represents e.g. the CA1 region of the hippocampus and includes N_{hipp}=1400 units with their fields regularly arranged to evenly tile the spherical surface. The output network represents a population of N_{mEC} = 250 would-be grid units in mEC, all with the same adaptation parameters.

Collateral weights between mEC units are set in the following way: each unit is temporarily assigned a preferred position, an auxiliary field at coordinates (φ,θ) on the sphere, and a preferred direction (angle relative to the meridian, with 0 pointing towards the North Pole). The coordinates as well as the angle are randomly chosen. These auxiliary fields are introduced only for the sake of weight definition and do not play any role in other parts of the simulations. The collateral weight between unit i and unit k is then calculated as

$$J_{ik} = \left[f_{\theta_i}(\omega_{ik}) f_{\theta_k}(\omega_{ik}) \exp\left(-\frac{d_{ki}^2}{2\sigma_f^2} \right) - \kappa \right]^+ \qquad (1)$$

where []+ denotes the Heaviside step function, κ=0.05 is an inhibition factor to favor sparse weights, f is a tuning function described in the Appendix and ω_{ik} is the direction, with respect to the North Pole, of the line connecting the auxiliary fields of unit i and k, along the great circle. σ_f = 10cm denotes how broad the connectivity is, and d_{ki} is defined as

$$d_{ki} = \mathrm{R} \cos^{-1}(x_i x_{exp} + y_i y_{exp} + z_i z_{exp}) \qquad (2)$$

i.e., it is the distance between the coordinates of the auxiliary field (x_i, y_i, z_i) and the *expected* position of a virtual rat that had started at the auxiliary field of unit k and had moved 10 cm along the geodesics joining both fields, corresponding to 250 ms of reverberatory delayed activity of movement along this direction. Normalization on this set of connections is performed as usual, setting a unitary L^2 norm. The definition of the weights is such that strong positive interactions are generated between units with similar preferred head direction and activation fields appropriately shifted along the same head direction. The strength of the recurrent input with respect to the feedforward input was set to 0.2 for most of the simulations.

Note that with a radius R=52.6cm and the adaptation parameters we use, most units tend to have 12 fields in the simulations with recurrent connections, but 13 or 14 fields without (see Fig.S1, top). Still, we choose to use the same value of R in the two cases for ease of comparison. Also, when separately varying R in order to optimize the proportion of units evolving 12 fields in each condition, and taking into account all units, the simulation without collaterals yields maps much closer to the "soccer ball" ideal (Fig.S1, bottom).

RESULTS

Simulations produce units with different number of fields, as shown in Fig.S1. In the following, we set the radius to the same value R=52.6cm and consider only units which have developed 12 fields. We compare simulations with and without lateral connections. For each simulation, the same set of grids are left to self-organize in two distinct, independent environments.

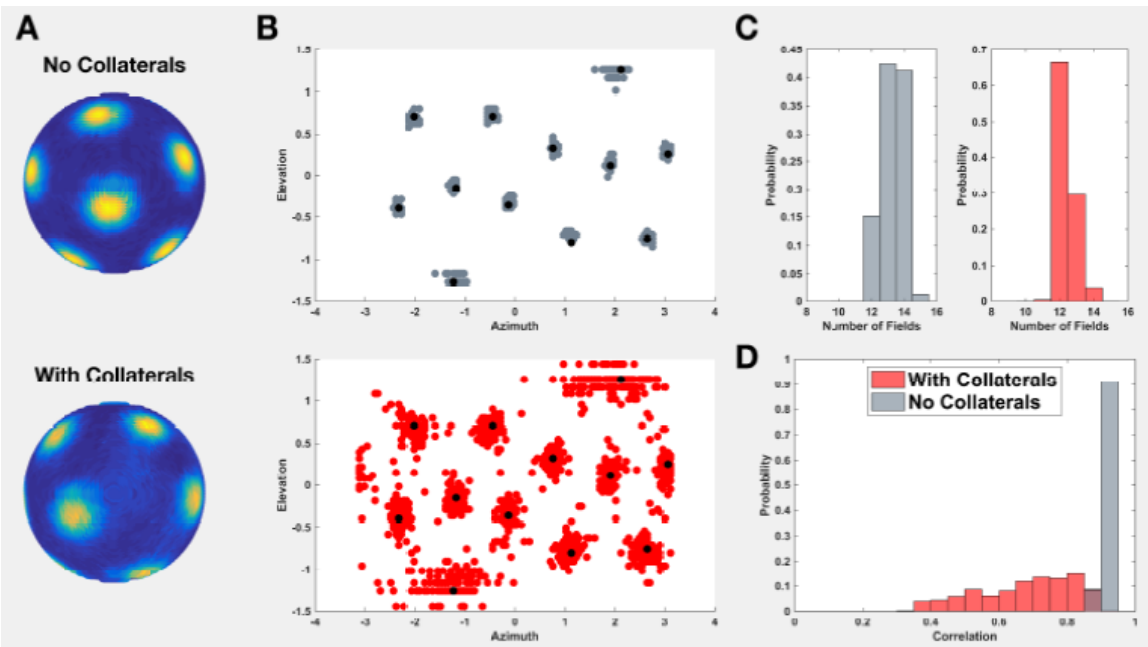

Figure 2. Collateral interaction distort the grid pattern on a sphere.
A Two representative examples of activity developed by grid units on the sphere. Top: unit from a simulation without collateral connectivity. Bottom: unit from a simulation in which units interact through collaterals. **B** Distribution of the position of all the fields from a population after the rate map of each unit has been rotated to maximize its overlap with a common "soccer ball" template (the field centers of this perfect grid are shown in black). Top: fields of a population of non-interacting units (mean distance from perfect center: 3.52º); Bottom: fields of a population of interacting units (mean distance from perfect center: 7.96º). **C** Distribution across the population of the number of fields developed by units, with R=52.6cm. Left: no collaterals; Right: with collaterals. Sample sessions. Values across sessions: mean fraction with 12 fields, for no interaction 0.13, with interactions 0.68. **D** Correlation of all units with 12 fields with a best-rotated 'soccer ball'. Aggregate from all sessions.

Grid Map Distortion

At the single-map level, the introduction of lateral connections has the immediate result of interfering with the development of a regular grid structure. We can quantify this phenomenon by computing the spherical correlation (as the best match over any possible rotation) of the activity map developed by each unit with the template of a perfect 12-field "soccer ball" rate map (Stella et al., 2013). One sees a marked effect of recurrent interactions as a general increase in the distance from the best-matching template, even though the radius of the sphere can be adjusted so that even with the recurrent collaterals most units produce 12 fields (Fig.2C,D, and Fig.S1). The simulations without collaterals produce fairly good exemplars of an ideal spherical grid. This is not the case when recurrent collaterals are introduced: the interactions among EC units lead to a disruption of the regular arrangement of their fields (see the examples in Fig.2A,B; and the quantitative measures in Fig.S2).

This illustrates how quickly the spherical case departs from what is observed on a planar surface. There, it has been shown (Si et al., 2012; Si and Treves, 2013), the presence of lateral connections has the crucial role of inducing a common orientation in the grid population and does not hinder, in fact enhances, the quality of the grids developed by the system. The same process does not occur here, where a similar attempt to induce coordination in grid-evolving units appears at odds with the regularity of the grids.

Grid units tend to cluster

One can then ask, what are the effects of connections on the whole ensemble of grid units, and what does the "common orientation" that they should induce look like, on the sphere? We answer these questions by analyzing the spatial (spherical) similarity in the structure of the maps developed by different units. To do so, for each unit, we computed the activity auto-correlation after 373,248 different rotations. Rotations were randomly drawn to evenly span the space of possible Euler rotations ($2\pi \cdot \pi \cdot 2\pi$, considering also the cosine factor). We then sorted the rotations according to their autocorrelation score, from highest to lowest. We used the first 5000 best rotations to compute their distribution density in the 3-dimensional space of Euler rotations $\Gamma_i(\varphi,\theta,\psi)$ (where i denotes the unit). For each pair of units (i,j) we then computed the overlap between $\Gamma_i(\varphi,\theta,\psi)$ and $\Gamma_j(\varphi,\theta,\psi)$ (as a correlation). This correlation (or effectively, distance) matrix was used to identify clusters of units sharing a similar rotation distribution. The clustering was performed with the "ward" algorithm and the number of clusters was optimized over the range [4-15].

The outcome of this clustering algorithm shows how the population effectively breaks down into sub-groups of segregated units, each developing an internal degree of coherence that is higher than that shared by the entire population. In Fig.3, the spatial position (on a 2D projection of the sphere) of the field centers of all units with 12 fields in the population are shown colored according to their cluster membership. To a large extent, each of these clusters expresses a common orientation, meaning that each of the 12 fields of a unit tends to appear grouped with those of every other unit in the cluster. Different clusters appear instead unrelated, roughly to the same extent that individual units are in the non-connected simulations (Fig.3B).

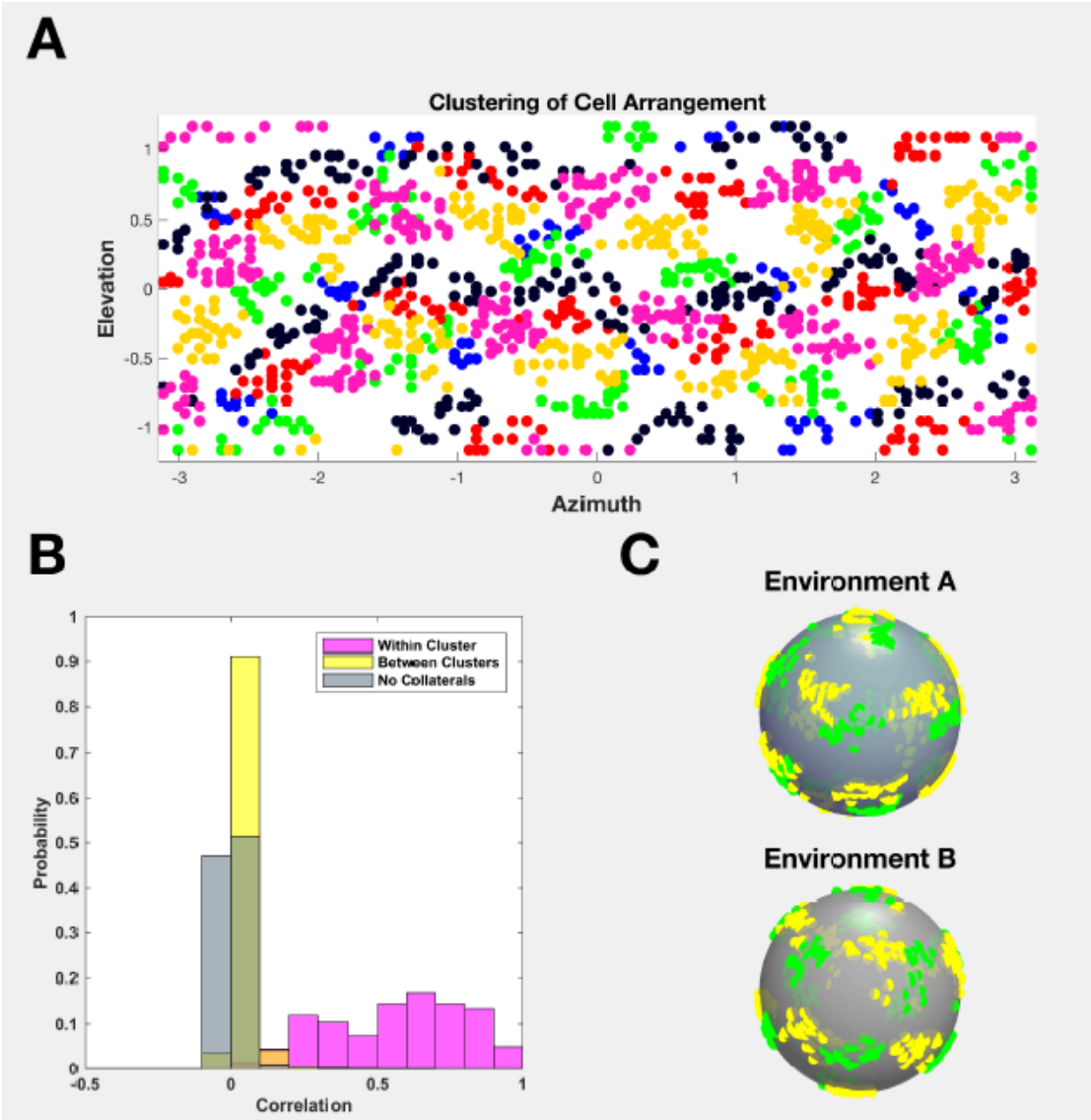

Figure 3. Interacting grid units tend to cluster.
A Spatial distribution of the fields of a population of grid units over the surface of a sphere. Each field center of each unit is assigned a color according to which cluster the unit belongs to. Sample session. Average number of clusters across sessions: 7.6. **B** Measure of clustering quality: distribution of pairwise unit correlations between the 3D density of their 5000 best rotations in the space of Euler angles (the measure used to define clusters). Sample session. **C** Example of the arrangement of the fields from the units of two clusters (green and yellow, as in A). The positions of their fields are shown in two different spherical environments. Clusters were defined solely in the first environment.

Crucially, the membership to one cluster is a feature that is carried over to new environments almost entirely unaltered. In Fig.3C we show a similar plot of the field centers, this time for only two of the clusters, to highlight their correspondence in two different environments. The grouping is conserved, as is the mutual avoidance of the fields in the two clusters.

Thus we observe how on the sphere, the interaction between units results into a break-down of the population, with different sub-sets acquiring a coordinated arrangement, while at the same time it forces each unit to distort its firing pattern away from that of an ideal grid. These features are consistently reproduced across environments.

Remapping

In order to gain a better understanding of the mechanisms underlying grid formation and of the spatial code they can generate, we need to address the properties of remapping. To that end, maps were developed independently in two environments: the set of place cell inputs and the associated feed-forward connectivity was randomly initialized in each of the two environments. Only the strength of the recurrent collaterals (when present), and thus of the grid cell interactions, was maintained after remapping. For each unit, we compare the maps developed in the two environments, maps A and B. Taking map A, we again apply a large random set of rotations spanning the entire Euler rotations range, and for each rotated version of map A we compute the resulting overlap (correlation) with map B. We first consider only the rotation associated with the highest similarity score – the "best" rotation – and its associated rotated version of the map, A'.

Results are shown in Fig.4A: in red one can see the distribution of correlation values between B and A', and for comparison the results for the best rotation of a perfect grid (in light brown). The two distributions lie in the same range of high correlation, although the perfect grid can usually be rotated to achieve a higher correlation with the map in B, suggesting that the distortion observed in A is independent of that in B. We can compare both distributions with that obtained by correlating for each unit the best rotated map A' with the map of each other unit in the same cluster, in B (magenta). In this case correlation values are somewhat lower but, consistent with the partially coherent behavior expressed by units in the same clusters, they are still significantly higher than those obtained using either the maps in B of units in other clusters (yellow) or of all the other units (not shown).

Next, we observe that, if the two maps were perfect soccer balls, there would be $12 \cdot 5=60$ equivalent ways to rotate one into the other. Because of the distortions from the most symmetric configuration, the degeneracy is only approximate, but still massive: there are many different rotations, spanning a diverse set of Euler angle triplets, that lead to almost the same correlation values as the best rotation, for each unit. In fact, if we take for each unit its N_{Best} rotations of the map in A, we see that their average correlation with the corresponding map in B is a smooth function of N_{Best}, which averaged over the population shows a significant decrease only when N_{Best} reaches into the thousands (among an arbitrarily set range of 373,248 randomly chosen triplets (φ,θ,ψ) of Euler angles; Fig.4B).

Remarkably, the clusters of units defined on sphere A maintain a *partial* coherence once remapped onto sphere B, as already suggested by the example in Fig.3C. If we randomly choose 5 sample units per cluster, and consider their N_{Best}=500 individually most correlated rotations, we find that they cluster into distinct 'islands' in Euler space, with each cluster contributing 60 regularly arranged islands, as shown in Fig 4C. Inside each

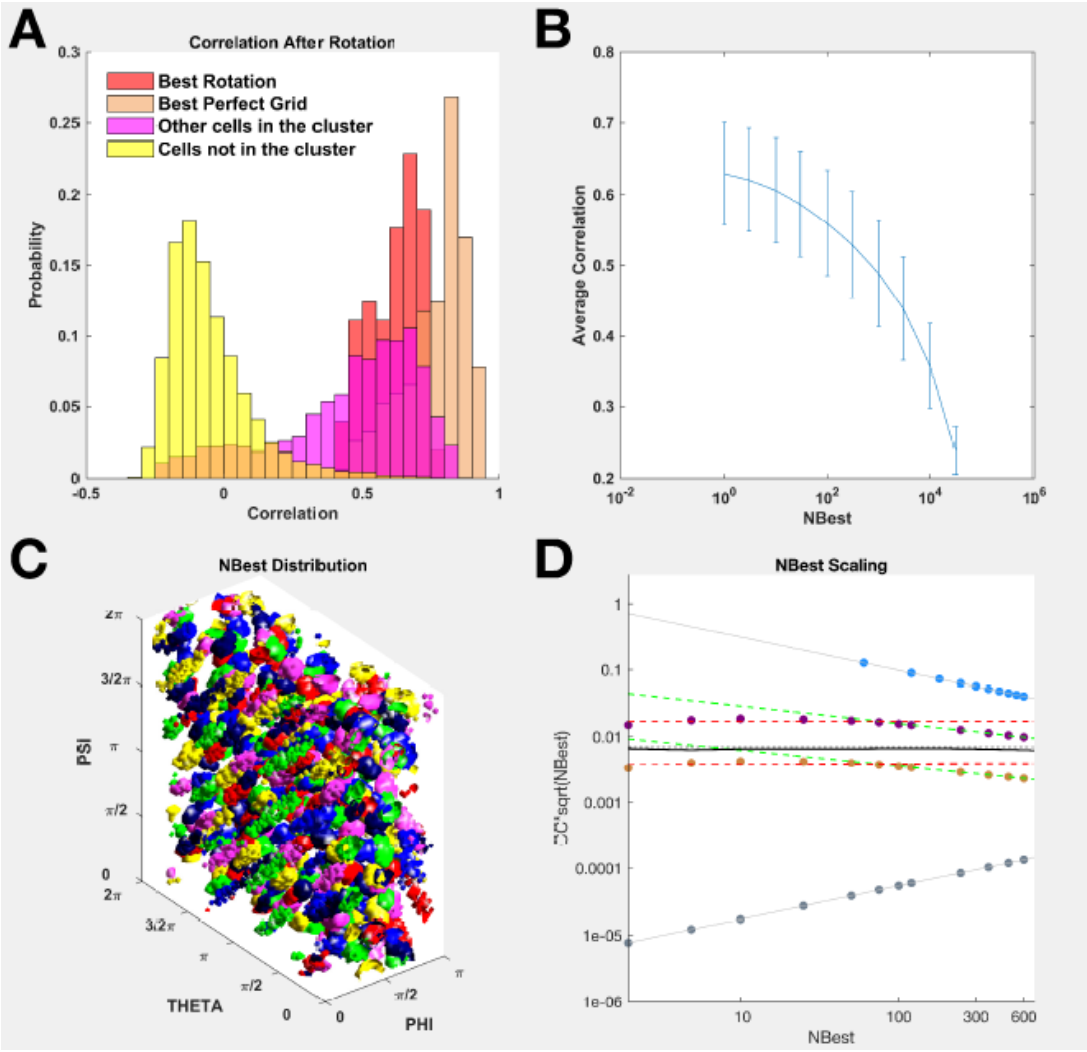

Figure 4. Remapping preserves the clusters.

A Distribution of overlaps after the best rotation (out of 373,248 randomly drawn rotations). All sessions. Mean overlap: with the map in B of the same unit, 0.615; with those of other units in the cluster, 0.455; with those of units not in the same cluster, -0.050; with those of all units, 0.006 (not shown). Mean overlap of the best rotated perfect grid with the maps in B, 0.659. **B** Correlation-N_{Best} dependence. Average value of the correlation between the map in B and the map of the same unit in A rotated N_{Best} times (bars denote SDs). **C** Example of the spatial density distribution of best rotations for cells belonging to different clusters. Each color represents the distribution of the N_{Best}=500 angles for 5 random units for each cluster. Sample session. **D** N_{Best} scaling and partial coherence. Different scaling $CC=1/N_{Best}^{\beta}$ of the Clustering Coefficient (computed over distributions like the one in C) for different types of remapping. Logarithmic scales. For clarity, the quantity on the y-axis is $CC \cdot N_{Best}^{1/2}$. Grey: random remapping of non-interacting units, β=0. Blue: coherent remapping, β=1. Purple: CC computed using interacting grid units from the same cluster. Brown: CC computed using all units in an interacting population. Dashed black: β=0.5; dashed red: CC computed using N_{Best} up to 100 (average over simulations: β=0.53 all units, β=0.52 units in same cluster); dashed green: CC computed using $100 < N_{Best} <= 700$ (average over simulations: β=0.74 all units, β=0.71 units in same cluster); Black: computed using numerical estimation of the analytical formulation (see Appendix 2).

island, however, chaos prevails. Some single units contribute many more of their best rotations to particular islands, and the shape of each island appears randomly distorted.

After failing to observe any further geometrical structure within the islands, we resorted to a quantitative measure of the extent to which the best rotations are concentrated across units. We define a clustering coefficient, CC (Cerasti and Treves, 2013), that measures, starting from the 373,248 randomly chosen Euler triplets and taking the N_{Best} distinct rotations for each of N units, the probability that two such triplets coincide (see Methods for the definition). For a perfectly coherent rotation the N_{Best} rotations would be the very same triplets across units, hence $CC=1/N_{Best}$. For a totally incoherent remapping, triplets would coincide at chance level, hence CC=1/373,248. Fig.4D shows that, whether we take only units within the same cluster or in the entire population, the clustering coefficient has intermediate values, scaling approximately as $1/\sqrt{N_{Best}}$, for N_{Best} small – corresponding to a horizontal line in Fig.4D. For larger N_{Best}, a steeper decrease prevails, presumably because the population remains less than fully coherent also when allowing for looser remapping correspondence. In Appendix 2 we show that an intermediate scaling can be expected from a simple analytical model. A direct numerical evaluation of the mean field formulation for the clustering coefficient (see Equation A2.4 in the Appendix) yields a behavior in very good agreement with the one observed in simulations (Fig 4D, black continuous line). Here the deflection from the $1/\sqrt{N_{Best}}$ scaling appears to happen for larger values of N_{Best} (lying outside the right margin of the plot), presumably as an effect of the mean field approximation. The model does not seem to predict, however, an exact square root scaling, and it remains unclear to us whether the exponent $\beta\approx1/2$ that we find to characterize $CC\approx1/N_{Best}^{\beta}$ (for N_{Best} small) is fundamental or a mere coincidence.

An extensive set of simulations with varying overall interaction strength among the units (through a prefactor) indicates that the absolute value of the clustering coefficient depends mildly on the prefactor, but its scaling exponent $\beta\approx1/2$ for N_{Best} small is the same (not shown), and the clustering coefficient is constant at 1/373,248 only when the prefactor is strictly zero.

DISCUSSION

The simulation of our adaptation model, allowing for collateral interactions among the units, indicates a radically different nature of the grid code on the sphere. The same interactions which on a plane suppress fluctuations and lead to a collapse into a smooth continuous attractor recruiting all units with the same or similar grid spacing (Si et al., 2012; Urdapilleta et al., 2017), on a sphere lead to a hierarchy of effects on single-unit maps that

a) are distorted from the available symmetric ‘soccer ball’ field arrangement
b) are forced into clusters of units with approximately overlapping fields
c) remap to a different sphere with only partial coherence, even within clusters.

We regard these as predictions that could be validated or falsified by experiments which are doable in rodents, even though they may require *ad hoc* arrangements to allow for the

slow emergence, possibly only during a two-week developmental period, in rats (Langston et al., 2010), of a stable set of 2D maps. Alternative models of grid map formation may lead to different predictions, but we would not know how, and are not currently aware of attempts by others, to extend existing models e.g. the oscillator interference model (Burgess et al., 2007) or the continuous attractor model (Burak and Fiete, 2009), to work on a sphere.

A sphere is of course an even more artificial rearing environment than a flat box, but we believe that it may help to capture a more fundamental behavior of grid cell coding by pointing to the ephemeral link being established between those properties of grid cells considered to be universal and the specific use of flat, bounded environments. The qualitative characteristics a)-b)-c) may be general to any curved environment, and they can be contrasted with the character of grid cell activity in rodents reared in standard laboratory conditions. In this sense, a sphere may be closer than a plane to the ecological condition of a Norway rat system of burrows (Calhoun, 1962). Grid cell representations may be presumed to have evolved to be relevant to rodents living in the wild.

In humans, the same fMRI hexagonal signature that has been hypothesized to reflect grid cell activity in a virtual reality navigation task (Doeller et al., 2010) has later been reported when subjects 'move' in a 2D space of drawings (Constantinescu et al., 2016), raising the issue of whether hexagonal symmetry may characterize even abstract conceptual spaces, when described by assigning two dimensions (Bellmund at al., 2018). Our model suggests that this may occur only around locations that are either flat *a priori*, or where curvature has been ironed out, perhaps by extensive training.

From a complex systems point of view, it is remarkable how curvature opens up a scenario different from that of a strictly regular, periodic 2D tessellation (in a separate study, we had argued how the scenario of a 3D crystal, potentially relevant to bats, is in fact implausible; Stella and Treves, 2015). In the new scenario, a network of grid units 'behaves' more like a disordered system than like a crystal. The approximate inverse square root scaling of the clustering coefficient of the rotations, under remapping, reminds us of the partial coherence of a physical system with impurities, where some interactions are perforce 'frustrated' (Mezard et al., 1987). When interactions are short-range, local coherence may survive, avoiding the impurities, somewhat like grid maps away from objects placed in a flat environment (Hoydal et al., 2018; Boccara et al. 2019). A prevailing non-zero curvature is more akin, however, to a system with impurities and long-range interactions, where disorder affects even the shortest organizational scale. Such systems, not unlike human society, can offer only partial coherence.

Partial coherence, together with the realization that a network of grid cells may be endowed with a significant storage capacity (Spalla et al., 2019), in a sense brings back grid cells to the fold of memory systems, with the place cells. Alternative sets of coactivity relations stored on the same synaptic connections, as well as curvature, act on the currently active representation as 'quenched' disorder, and coexisting with such spin-glass-like disorder appears to be the ultimate challenge for memory systems in the brain (Treves, 2009).

Acknowledgements
We are grateful to Matteo Toso, who modeled the habitat of the Norway rat, to YuQiao Gu, who carried out some of the early simulations, and to colleagues in the EU 'Gridmap' and HFSP RGP0057/2016 collaborations, which also partially funded this study. This report was in part drafted during a 2018 visit by AT to the KITP, thereby also supported in part by NSF Grant No. PHY-1748958, NIH Grant No. R25GM067110, and the Gordon and Betty Moore Foundation Grant No. 2919.01.

References

Bellmund, J. L., Gärdenfors, P., Moser, E. I., & Doeller, C. F. (2018). Navigating cognition: Spatial codes for human thinking. *Science*, 362(6415), eaat6766.

Bostock, E., Muller, R. U., & Kubie, J. L. (1991). Experience-dependent modifications of hippocampal place cell firing. *Hippocampus*, 1(2), 193-205.

Burak, Y. (2014). Spatial coding and attractor dynamics of grid cells in the entorhinal cortex. *Current opinion in neurobiology*, 25, 169-175.

Burak, Y., & Fiete, I. R. (2009). Accurate path integration in continuous attractor network models of grid cells. *PLoS computational biology*, 5(2), e1000291.

Burgess, N., Barry, C., & O'keefe, J. (2007). An oscillatory interference model of grid cell firing. *Hippocampus*, 17(9), 801-812.

Calhoun, J. B. (1962). Ecology and sociology of the Norway rat. *US Public Health Service Publication* No. 1008.

Constantinescu, A. O., O'Reilly, J. X., & Behrens, T. E. (2016). Organizing conceptual knowledge in humans with a gridlike code. *Science*, 352(6292), 1464-1468.

Doeller, C. F., Barry, C., & Burgess, N. (2010). Evidence for grid cells in a human memory network. *Nature*, 463(7281), 657.

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.

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

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

Harvey, C. D., Collman, F., Dombeck, D. A., & Tank, D. W. (2009). Intracellular dynamics of hippocampal place cells during virtual navigation. Nature, 461(7266), 941.

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

Kruge, I. U. (2016). Ready, steady, go! Development of spatial representations in the rat. Doctoral Thesis 320 at NTNU. http://hdl.handle.net/11250/2423621

Kruge, I. U., Wernle, T., Moser, E. I., & Moser, M. B. (2013). Grid cells of animals raised in spherical environments. In *Soc. Neurosci. Abstr* (Vol. 39, No. 769.14).

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

Mézard, M., Parisi, G., & Virasoro, M. (1987). *Spin glass theory and beyond: An Introduction to the Replica Method and Its Applications* (Vol. 9). World Scientific Publishing Company.

Hoydal, O. A., Skytoen, E. R., Moser, M. B., & Moser, E. I. (2018). Object-vector coding in the medial entorhinal cortex. *bioRxiv*, 286286.

Si, B., Kropff, E., & Treves, A. (2012). Grid alignment in entorhinal cortex. *Biological cybernetics*, 106(8-9), 483-506.

Si, B., & Treves, A. (2013). A model for the differentiation between grid and conjunctive units in medial entorhinal cortex. Hippocampus, 23(12), 1410-1424.

Spalla, D., Dubreuil, A., Rosay, S., Monasson, R., & Treves, A. (2019). Can grid cell ensembles represent multiple spaces? *BioRxiv*, 527192.

Stella, F., Si, B., Kropff, E., & Treves, A. (2013). Grid cells on the ball. *Journal of Statistical Mechanics: Theory and Experiment*, 2013(03), P03013.

Stella, F., & Treves, A. (2015). The self-organization of grid cells in 3D. *Elife*, 4, e05913.

Stemmler, M., Mathis, A., & Herz, A. V. (2015). Connecting multiple spatial scales to decode the population activity of grid cells. *Science Advances*, 1(11), e1500816.

Treves, A. (2009). Spatial cognition, memory capacity and the evolution of mammalian hippocampal networks. In: Tommasi, L., Peterson, M. A., & Nadel, L. (Eds.). *Cognitive biology: evolutionary and developmental perspectives on mind, brain, and behavior*. Vienna Series in Theoretical Biology, MIT Press. pp. 41-59.

Urdapilleta, E., Troiani, F., Stella, F., & Treves, A. (2015). Can rodents conceive hyperbolic spaces? *Journal of the Royal Society Interface*, 12(107), 20141214.

Urdapilleta, E., Si, B., & Treves, A. (2017). S elforganization of modular activity of grid cells. *Hippocampus*, 27(11), 1204-1213.

Yoon, K., Buice, M. A., Barry, C., Hayman, R., Burgess, N., & Fiete, I. R. (2013). Specific evidence of low-dimensional continuous attractor dynamics in grid cells. *Nature neuroscience*, 16(8), 1077.

Appendix 1: Model

Additional aspects of the model, besides those reported in the main text:

The input to mEC unit i at time t is given by

$$h_i(t) = \sum_j W_{ij}(t) r_j(t) \qquad \text{(A1.1)}$$

The weight W_{ij} connects input unit j to mEC unit i. We assume that at the time the mEC units develop their maps, spatially modulated or place cell-like activity is already present, either in parahippocampal cortex or in the hippocampus. The network model works in the same way with any kind of spatially modulated input, but the place-cell assumption reduces the averaging necessary for learning. Each input unit activity in space is modelled as a Gaussian place field centered at preferred position $\boldsymbol{x}_{j0}$

$$r_j(t) = \exp -\|\boldsymbol{x}(\mathrm{t})-\boldsymbol{x}_{j0}\|^2/2\sigma_p^2 \qquad \text{(A1.2)}$$

where $\boldsymbol{x}(\mathrm{t})$ is the position at time t of the simulated rodent, σ_p =0.05m is the width of the field and $\|\boldsymbol{a}-\boldsymbol{b}\|$ is the great-circle distance on the sphere.

Single Unit Dynamics

The firing rate $\Psi_i(t)$ of mEC unit i is determined by a non-linear transfer function

$$\Psi_i(t) = (\pi/2) \arctan[g(t)(\alpha_i(t)-\mu(t))]\ \Theta(\alpha_i(t)-\mu(t)), \qquad \text{(A1.3)}$$

which is normalized to have maximal firing rate equal to 1 (in arbitrary units), while $\Theta(\cdot)$ is the Heaviside function. The variable $\mu(t)$ is a threshold while $\alpha_i(t)$ represents the adaptation-mediated input to unit i. It is related to $h_i(t)$ as follows:

$$\alpha_i(t) = \alpha_i(t\text{-}1) + b_1[h_i(t\text{-}1) - \beta_i(t\text{-}1) - \alpha_i(t\text{-}1)]$$
$$\beta_i(t) = \beta_i(t\text{-}1) + b_2[h_i(t\text{-}1) - \beta_i(t\text{-}1)], \qquad \text{(A1.4)}$$

where β_i has slower dynamics than α_i, with $b_2=b_1/3$, b_1= 0.1 (in a continuous formulation, the b coefficients become rates, in units of $(\Delta t)^{-1}$. These adaptive dynamics make it more difficult for a neuron to fire for prolonged periods of time, and correspond to the kernel K considered in the analytical treatment (Kropff and Treves, 2008). The gain $g(t)$ and threshold $\mu(t)$ are iteratively adjusted at every time step to fix the mean activity $a = \sum_i \Psi_i(t)/\mathrm{N_{mEC}}$ and the sparsity $s = (\sum_i \Psi_i(t))^2/(\mathrm{N_{mEC}} \sum_i \Psi_i(t)^2)$ within a 10% relative error bound from pre-specified values, a_0=0.1 and s_0=0.3 respectively. If k is indexing the iteration process:

$$\mu_t(k+1) = \mu_t(k) + b_3[a(t) - a_0]$$
$$g_t(k+1) = g_t(k) + b_4[s(t) - s_0], \qquad \text{(A1.5)}$$

b_3=0.01 and b_4=0.1 are also rates, but in terms of intermediate iteration steps. a^k and s^k are the values of mean activity and sparsity determined by $\mu_t(k)$ and $g_t(k)$ in the intermediate iteration steps. The iteration stops once the gain and threshold have been brought within the 10% error range, and the activity of mEC units are determined by the final values of the gain and threshold.

Synaptic Plasticity. The learning process modifies the strength of the feed-forward connections according to a Hebbian rule

$$W'_{ij}(t\text{-}1) = W_{ij}(t\text{-}1) + \varepsilon[\Psi_i(t)\, r_j(t) - \langle\Psi_i(t\text{-}1)\rangle\langle r_j(t\text{-}1)\rangle] \qquad \text{(A1.6)}$$

with a rate ε=0.002. $\langle\Psi_i(t)\rangle$ and $\langle r_j(t)\rangle$ are estimated mean firing rates of mEC unit i and place unit j that are adjusted at each time step of the simulation

$$\langle\Psi_i(t)\rangle = \langle\Psi_i(t\text{-}1)\rangle + \eta[\Psi_i(t) - \langle\Psi_i(t\text{-}1)\rangle]$$
$$\langle r_j(t)\rangle = \langle r_j(t\text{-}1)\rangle + \eta[r_j(t) - \langle r_j(t\text{-}1)\rangle] \qquad \text{(A1.7)}$$

With η=0.05 a time averaging factor. After each learning step, the $W'_{ij}(t)$ weights are normalized into unitary norm

$$\Sigma_j(W_{ij}(t))^2=1. \qquad \text{(A1.8)}$$

HD Input. Head Direction on the sphere is defined as the angle between a vector and the vector pointing towards the north pole. With the addition of HD modulation and collateral connections, the overall input to unit j for the interacting case is:

$$h_i(t) = f_{\theta i}(\omega(t))\left[\Sigma_j W_{ij}(t) r_j(t\text{-}1) + \rho\, \Sigma_k J_{ik}\, \Psi_k\,(t\text{-}\tau)\right] \qquad \text{(A1.9)}$$

with ρ=0.2 a factor setting the relative strength of feed-forward $W_{ij}(t)$ and collateral weights J_{ik}, and τ =25 steps a delay in signal transmission, as discussed by (Si et al. 2012). The multiplicative factor $f_{\theta i}(\omega(t))$ is a tuning function which is maximal when the current direction of the animal movement $\omega(t)$ is along the preferred direction θ_i assigned to unit i

$$f_\theta(\omega) = c + (1\text{-}c)\exp[\nu(\cos(\theta\text{-}\omega)\text{-}1)] \qquad \text{(A1.10)}$$

where c=0.2 and ν=0.8 are parameters determining the minimum value and the width of the cell tuning curve. Preferred head directions are randomly assigned to mEC units and they uniformly span the 2π angle.

Clustering Coefficient (*CC*). Given q cells and taking the first N_{best} Euler rotations from each, the clustering coefficient was defined as:

$$CC = [1/((qN_{best})^2 - qN_{best})]\, \Sigma_{ij}\Sigma_{mn} \exp\text{-}d(\Theta_{i,m},\Theta_{j,n})/\xi \qquad \text{(A1.11)}$$

with $i \neq j$ and where $d(\Theta_{i,m},\Theta_{j,n})$ is the distance between two three-dimensional rotations; and ξ is set to 5 degrees so that only nearly coinciding rotations contribute to the sum.

Grid field definition and properties. Individual fields for each developing unit were identified as continuous portions of the spherical surface where the unit firing rate was above two times the average firing rate computed over the entire environment. Field size was defined as the number of bins passing the threshold in each continuous region, field height as the maximum firing rate within the continuous region, and the field ellipticity as the ratio between the radii of a circle circumscribed to the field and a circle inscribed in the field.

Appendix 2: Statistical Analysis

Let us consider N units that have developed grid representations on a sphere A, and now develop also grid representations on another sphere B. For every triplet of Euler angles (φ,θ,ψ) one can define the overlap (or spatial Pearson correlation) C_i between the representation of unit i in A and that in B rotated by (φ,θ,ψ). Assume that $-1 < C_i < 1$. Then define $C_{mean}(\varphi,\theta,\psi)$ as the mean $C_i(\varphi,\theta,\psi)$ across all units, or the units in a cluster. Of course, $-1 < C_{mean}(\varphi,\theta,\psi) < 1$ as well (in practice its range is much more restricted, if different units do not coincide in their 'best rotations'). Now position all Euler triplets along the x-axis given by their C_{mean} value, and define f(x) as their density (density of angles) along the axis. That is, if one considers a total of N_{angles}, there are $N_{angles}f(x)dx$ of them between x and x+dx. In Fig.S3A we show the C_mean distribution for an entire population (dashed line) and separately for each of 8 grid unit clusters (colored lines).

Assume now that among all N_{angles} angles, we pick for each unit N_{Best} of them, those that have the highest C_i value. How will all the $N_{Best} \times N_{units}$ angles be distributed, on average, in terms of f(x), the C_{mean}–ordered histogram? On average, they will concentrate more at higher C_{mean} values, at the very least because each unit gives a 1/N contribution to C_{mean}; but possibly more concentrated than that. How much they concentrate is critical in order to determine the clustering coefficient *CC*, which measures simply how many of the N_{Best} angles (what fraction) coincide among pairs of distinct units. We assume then that, at least within a cluster,

1) one can write:

$$C_i = C_{mean} + \eta_i \qquad \text{(A2.1)}$$

where η_i is a form of 'noise', i.e. the combined effect of all other factors independent and unrelated to C_{mean}. Note that this decomposition is a strong assumption.

2) this 'noise' is normally distributed, with a width $\sigma(N_{Best},\text{cluster})$ that is the same across units in a cluster.

If we denote with $b(x; N_{Best})$ the average fraction of N_{Best} angles, among the $N_{angles}f(x)$ present at a given C_{mean}–value x, such that, within a cluster,

$$N_{Best} = N_{angles} \int f(x)\, b(x; N_{Best})dx \qquad \text{(A2.2)}$$

and with $x_b(N_{Best}$,cluster) the value of x such that this average fraction is ½, with these assumptions one has that b(x) can be expressed as the complementary error function

$$b(x) \approx (1/2)\ \mathrm{erfc}\ ((x_b-x)/\sigma\sqrt{2}) \qquad \text{(A2.3)}$$

which Fig.S3B shows is not a bad approximation, if one allows x_b and hence b(x) to depend on both the cluster and N_{Best}.

We make now the additional, critical (mean-field) assumption that

3) the clustering coefficient, *CC*, at least within each cluster, is only determined by the average density $b(x; N_{Best})$. Therefore, considering a generic pair of units,

$$CC = [N_{angles} /(N_{Best})^2] \int f(x)b^2(x; N_{Best})dx \qquad \text{(A2.4)}$$

One may observe that CC is given by the extent of the overlap between the two distributions f(x) and b(x) (dropping for ease of notation its argument N_{Best}). Eq.(A2.2) and Fig.S3 show that for small N_{Best} the overlap is limited to the opposing tails of the two distributions. We can evaluate the goodness of the assumptions made so far by comparing the values of the clustering coefficient obtained from the last equation to those obtained from the full analysis of simulation results. Equation A2.4 can be numerically evaluated by making use of the f(x) computed from simulations (Fig S3 A) and of the parameters for b(x) obtained from Gaussian fits (Fig S3 B-D). The resulting mean field approximation curve shows a remarkable similarity with simulation results (Fig 4 D). If we proceed and make the final assumption that

4) f(x) has a quasi-Gaussian upper tail

$$f(x) \approx k\ \exp\text{-}(x-x_0)^2/2\xi^2 \qquad (\text{for } x >> x0) \qquad \text{(A2.5)}$$

with k a suitable factor and ξ the effective width of the tail, we can obtain an analytical estimate of the *CC*. The result is

$$\ln(CC(N_{Best})) \approx -\ln(N_{Best}) - (x_0-x_b)^2\sigma^2/[2(\sigma^2+\xi^2)\ (\sigma^2+2\xi^2)] \qquad \text{(A2.6)}$$

(keeping only the leading exponent). Both x_b and σ may depend on N_{Best}, but Fig.S3 indicates that within each cluster the dependence of σ is weak, while x_b shifts leftward as N_{Best} increases: from Eq.(A2.2) one can derive

$$\ln(N_{Best}) \approx \ln(N_{angles}) - (x_0-x_b)^2/[2(\sigma^2+\xi^2)]. \qquad \text{(A2.7)}$$

This yields a scaling of $\ln(CC(N_{Best})) \approx -\beta \ln(N_{Best})$ with $0 < \beta < 1$, and the particular value $\beta=1/2$ is obtained for $\sigma^2 \approx 2\xi^2$. We have no explanation for why this last relation appears to hold, approximately, for N_{Best} small, as shown in Fig.4D.

Supplementary Figures

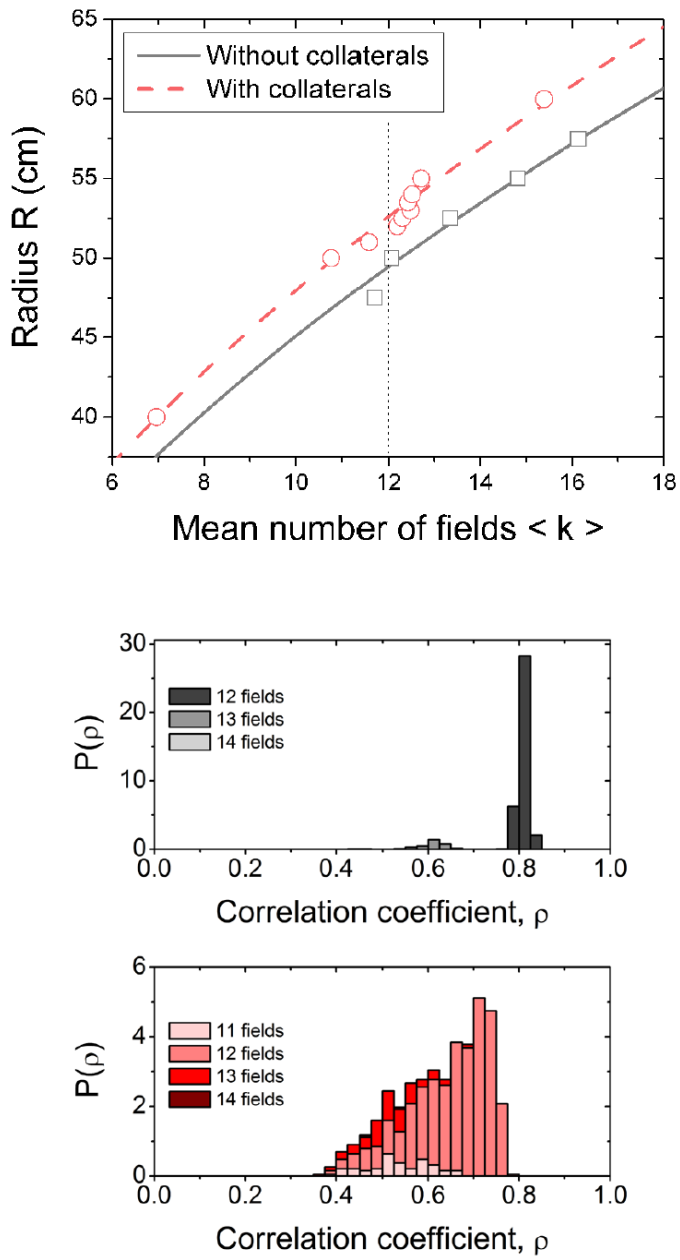

Figure S1. Interactions slightly reduce grid spacing, and distort the grids.
Top. Dependence of the mean number of fields across units in each simulation on the radius R of the sphere, with (red) and without (grey) recurrent collaterals. Bottom. Distribution of the values of the spherical correlation of the map developed by each unit with the appropriately oriented perfect "soccer ball" pentagonal grid, in simulations with R=50cm (without collaterals) and R=52.6cm (with collaterals). Note that in the other "control" simulations without collaterals cited in the text, R was set to 52.6cm, and only the subset of units with 12 fields (a minority) was used in the analyses.

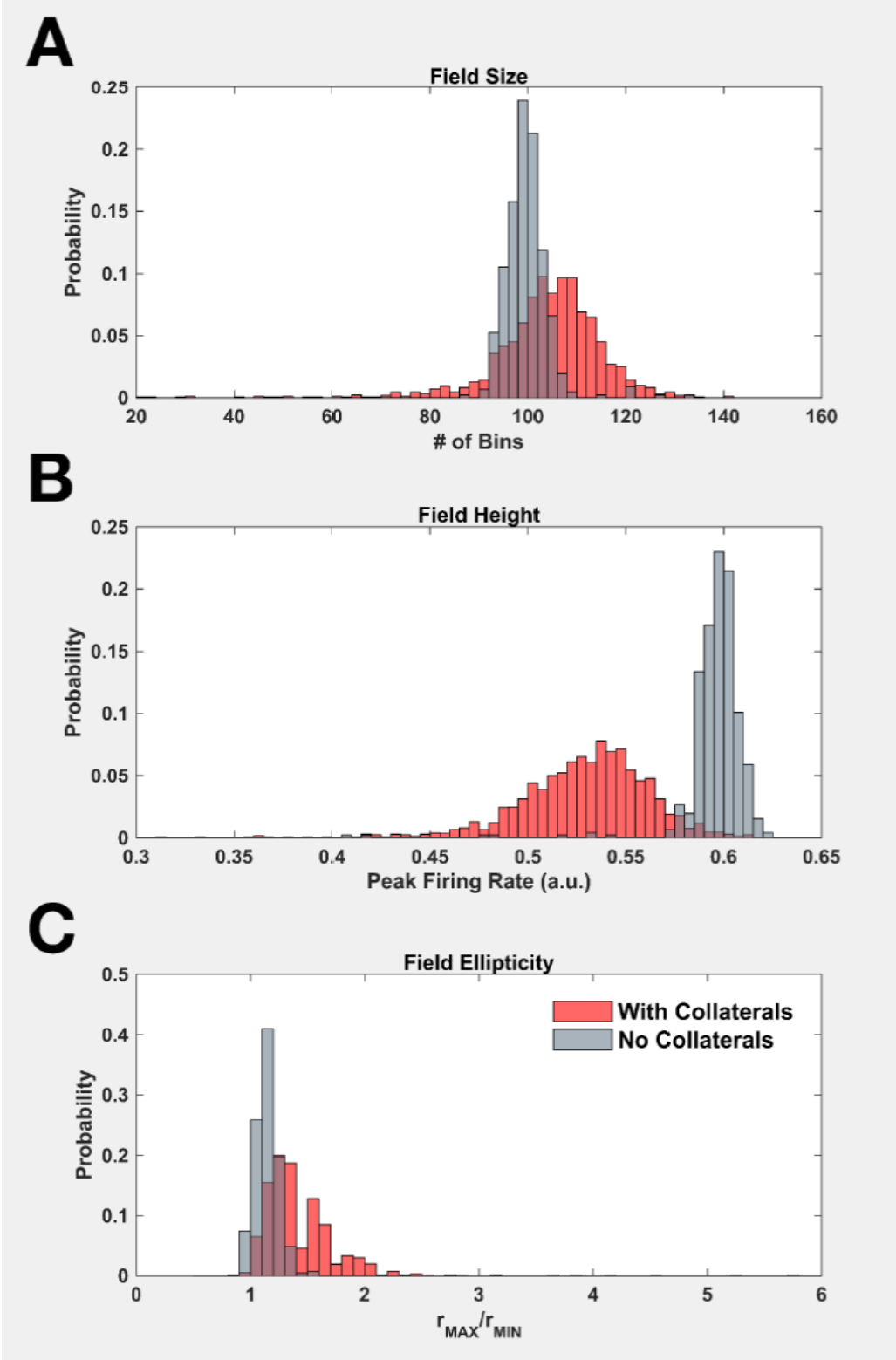

Figure S2. Interactions distort grid field shape.
Distribution of field size (**A**), field maximum height (**B**) and field ellipticity (**C**) across a population of grid cells. Red bars: population with recurrent collaterals; grey bars: population without recurrent collaterals.

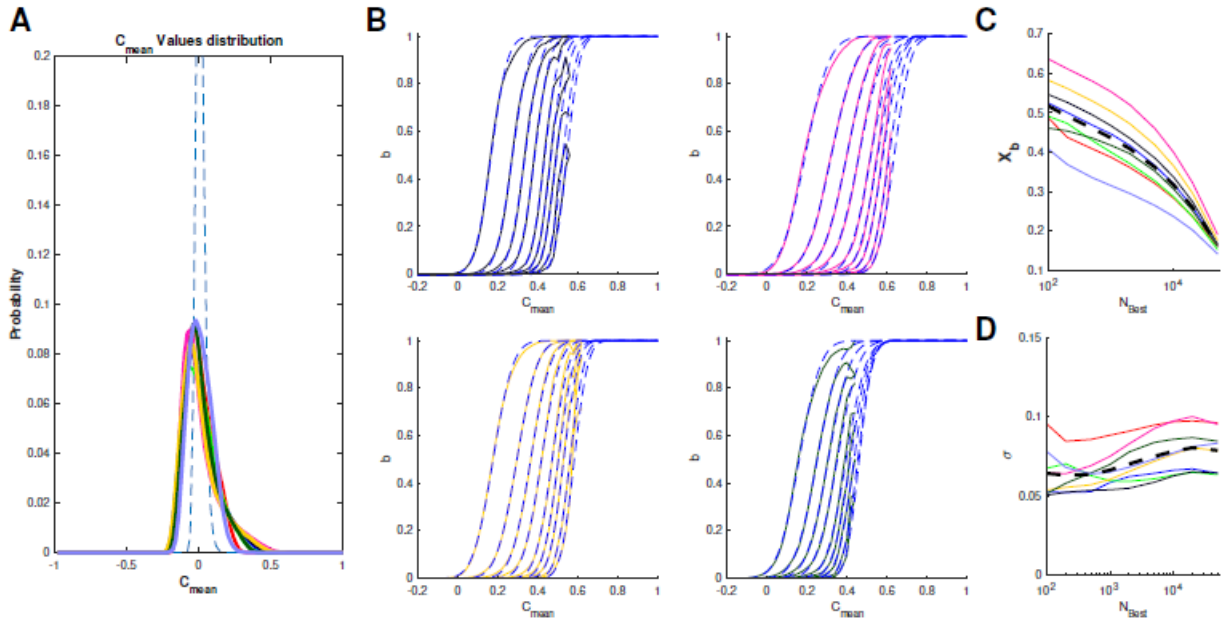

Figure S3. Despite the clustering, remapping appears to be a random process.
A Density of C_{mean} values across the set of Euler rotations. Colored curves are computed within units belonging to the same cluster; the dashed line is the C_{mean} density computed over all units. **B** Four examples (4 different clusters) of the cumulative C_{mean} distribution computed for different values of N_{Best} [=100, 200, 500, 1000, 2000, 5000, 10000, 20000, 50000]. Colored lines: simulation data; dashed lines: Gaussian fits. **C** and **D** Mean and standard deviation, respectively, obtained from the Gaussian fits for different clusters and different N_{Best} values. Black dashed line: mean across clusters.