

RESEARCH ARTICLE

# Robust and efficient coding with grid cells

Lajos Vágó\*, Balázs B. Ujfalussy<sup>‡\*</sup>

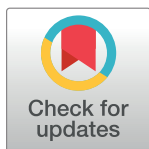
NAP-B PATTERN Group, MTA Wigner Research Center for Physics, Budapest, Hungary

<sup>‡</sup> Current address: Laboratory of Neuronal Signalling, MTA Research Institute of Experimental Medicine, Budapest, Hungary

\* [vagolala@gmail.com](mailto:vagolala@gmail.com) (LV); [balazs.ujfalussy@gmail.com](mailto:balazs.ujfalussy@gmail.com) (BBU)

## Abstract

The neuronal code arising from the coordinated activity of grid cells in the rodent entorhinal cortex can uniquely represent space across a large range of distances, but the precise conditions for optimal coding capacity are known only for environments with finite size. Here we consider a coding scheme that is suitable for unbounded environments, and present a novel, number theoretic approach to derive the grid parameters that maximise the coding range in the presence of noise. We derive an analytic upper bound on the coding range and provide examples for grid scales that achieve this bound and hence are optimal for encoding in unbounded environments. We show that in the absence of neuronal noise, the capacity of the system is extremely sensitive to the choice of the grid periods. However, when the accuracy of the representation is limited by neuronal noise, the capacity quickly becomes more robust against the choice of grid scales as the number of modules increases. Importantly, we found that the capacity of the system is near optimal even for random scale choices already for a realistic number of grid modules. Our study demonstrates that robust and efficient coding can be achieved without parameter tuning in the case of grid cell representation and provides a solid theoretical explanation for the large diversity of the grid scales observed in experimental studies. Moreover, we suggest that having multiple grid modules in the entorhinal cortex is not only required for the exponentially large coding capacity, but is also a prerequisite for the robustness of the system.



## OPEN ACCESS

**Citation:** Vágó L, Ujfalussy BB (2018) Robust and efficient coding with grid cells. *PLoS Comput Biol* 14(1): e1005922. <https://doi.org/10.1371/journal.pcbi.1005922>

**Editor:** Jakob H Macke, Stiftung caesar, GERMANY

**Received:** March 13, 2017

**Accepted:** December 8, 2017

**Published:** January 8, 2018

**Copyright:** © 2018 Vágó, Ujfalussy. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the paper.

**Funding:** B.B.U. and L.V. were supported by the National Brain Research Program of Hungary (KTI-NAP-12-2-201). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

## Author summary

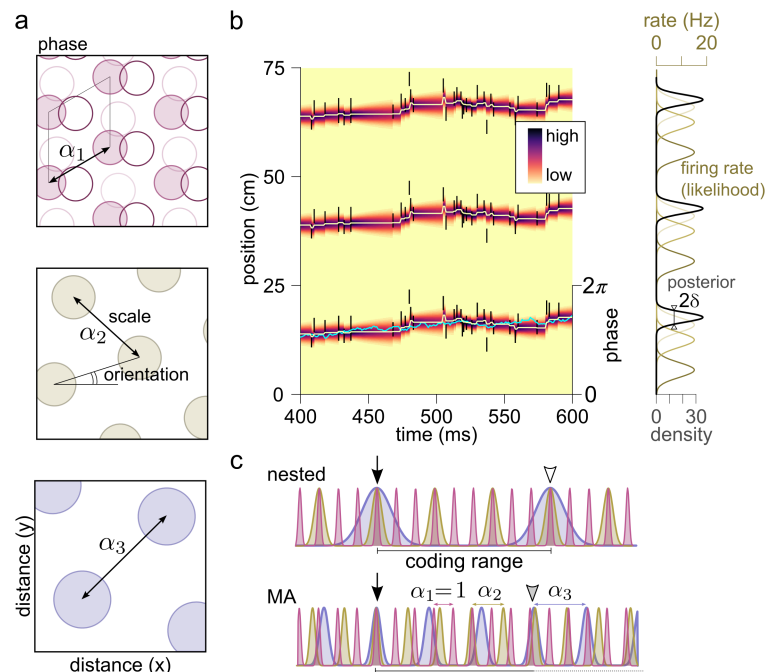
Navigation in natural, open environments poses serious challenges to animals as the distances to be represented may span several orders of magnitudes and are potentially unbounded. The recently discovered grid cells in the rodent brain are thought to play a crucial role in generating unique representations for a large number of spatial locations. However, it is unknown how to choose the parameters of the grid cells to achieve maximal capacity, i.e., to uniquely encode the utmost locations in an open environment. In our manuscript, we demonstrate the surprising robustness of the grid cell coding system: The population code realised by grid cells is close to optimal for unique space representation irrespective of the choices of grid parameters. Thus, our study reveals a remarkable

robustness of the grid cell coding scheme and provides a solid theoretical explanation for the large diversity of the grid scales observed in experimental studies.

## Introduction

Optimising neuronal systems for efficient processing and representation of information is a key principle for both understanding and designing neuronal circuits [1], but deciding whether a particular neuronal phenomenon reflects an optimisation process is often difficult. Grid cells in the medial entorhinal cortex have been suggested to efficiently represent spatial location of the animal by their spatially periodic firing fields near optimally [2, 3, 4, 5]. However, it remained controversial whether the efficiency of the grid cell code is the result of the precise tuning of the grid parameters [6, 7, 8] or the performance of the system is relatively insensitive to the actual parameter settings [4, 5, 9].

Grid cells are spatially tuned neurons with multiple firing fields organised along the vertices of a triangular grid (Fig 1a; [10, 11]). Grid cells of any particular animal are organised into functional modules [12, 13] cells within a module share the same grid scale and orientation, but differ in the location of their firing fields, i.e., their preferred firing phase within the grid



**Fig 1. Coding with grid cells.** (a) Schematic firing fields (circles) of two-dimensional grid cells as function of spatial position. Grid cells are organised into modules: Cells from the same module share the orientation and scale parameter but differ in their spatial phase (top, shades of purple). Different modules have different scale and orientation (top to bottom). (b) Grid cell spikes encode the phase of a module. Spiking of grid cells (black ticks, each spike is shown three times, at the maxima of the cells' firing rate) from a single module represents the movement of the animal (light-blue line) in a 1 dimensional environment. Since the firing rate of the cells (right, olive) is periodic, the position (left: colormap, right: black) which is represented by the phase of the module is also periodic. The uncertainty of the representation fluctuates over time around a typical value,  $\delta$  (right). (c) Grid cell coding schemes. The location of the animal (origin, filled arrow) is jointly encoded by the phases of the different modules in both nested (top) and modulo arithmetic (bottom) codes. Grey and empty arrowheads indicate locations with large or catastrophic interference between the modules, respectively.

<https://doi.org/10.1371/journal.pcbi.1005922.g001>

period (Fig 1a). Modules form the functional units of the grid representation: The joint activity of all (possibly hundreds of) cells within each module is captured by the (two dimensional) phase of the given module (Fig 1b; [14, 15]) and the relationship between different cells from the same module remains stable across different environments [16], during sleep [17, 18] or after environmental distortions [13]. A given spatial location is represented by the phases of the different modules ('phase vector'). The representations are unique up to a critical distance above which the coding becomes ambiguous: the phase vectors, and hence the firing rates of all grid cells, become (nearly) identical at two separate physical locations (Fig 1c).

Depending on the magnitude of the critical distance compared to the largest grid scale, two complementary coding schemes have been proposed for grid cells (Fig 1c): In nested coding [4, 6, 8] smaller grid modules iteratively refine the position coding of larger modules and the modules span a wide range of scales. The capacity of nested codes, defined as the ratio of the coding range and the resolution, is exponential in the number of modules. Maximal capacity can be achieved by setting the coding range equal to the maximal grid period and then optimising the resolution by a geometric progression of the grid scales [6]. When the total capacity is utilised to encode locations within the maximal grid scale, catastrophic interference will cause ambiguity in the grid code beyond this distance (Fig 1c).

When the coding is not optimised for a fixed range, the unique combination of the activity of grid modules can encode a potentially unbounded range that can be substantially larger than the scale of the largest module using a modulo arithmetic (MA) code [2, 3, 14] (Fig 1c). In this case the grid periods can be similar in magnitude (e.g., co-prime integers or a geometric progression with a relatively small ratio). However, it is not known under what conditions the MA coding system can achieve exponential capacity [3, 14], and how robust is the capacity against the choice of the grid periods or neuronal noise.

Here we develop a novel approach to study the capacity of the grid coding system that is based on Diophantine approximations, i.e., approximation of real numbers by rational numbers. First, we apply the technique to study coding with two grid modules. We show that the capacity of the system is extremely sensitive to the number theoretic properties of the scale ratio between the modules. Next, we generalise our approach to the case of multiple modules, and show both analytically and numerically that the exponential capacity of the grid cell coding system can be achieved using the MA coding scheme. Finally, we demonstrate that when the coding range is constrained by neuronal noise, the capacity of the system is extremely robust against the choices of the scaling of the modules.

## Results

In the first section of the Results we briefly introduce the terminology and define the concepts used throughout the paper. We include this section for completeness, although several ideas presented in this section have been described before, e.g., in [3, 4].

We investigate grid cell population codes along a linear trajectory as the one dimensional results extend to two (or higher) dimensions without difficulty, at least for axis aligned grid modules (Methods) [3, 4]. The periodic population activity of module  $i$  can be summarised by its spatial phase

$$\psi_i(x) := (x \bmod \alpha_i) / \alpha_i \in [0, 1),$$

which depends on the position ( $x$ ) and the scale of the module ( $\alpha_i$ , Methods) with  $\alpha_0 = 1$ , which means that distances are expressed in the unit of the smallest grid period. We assume, that, without loss of generality, at the spatial origin all modules are in their 0 phase. Spikes of the neurons in module  $i$ , represent the spatial location of the animal with a maximum error

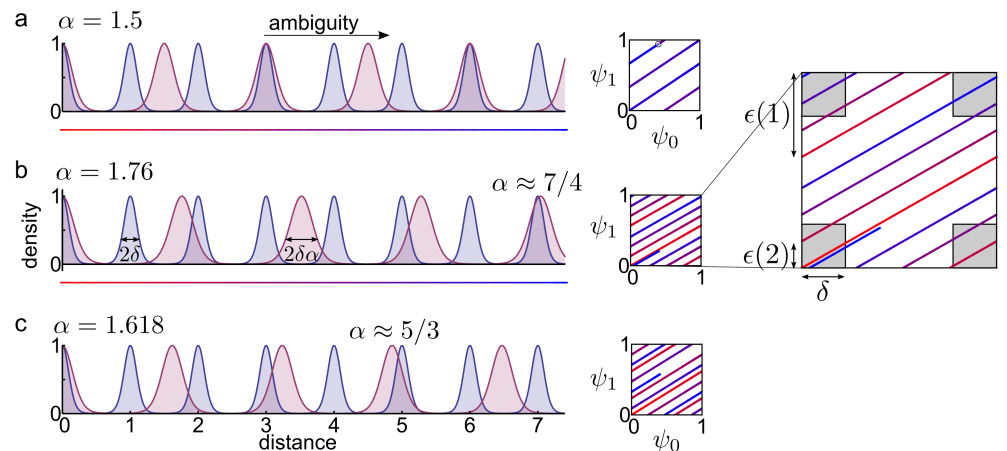
$\delta \alpha_i$  (i.e., with the same  $0.01 \leq \delta \leq 0.2$  phase error for all modules; [Methods](#); [8]). Ambiguity occurs if the phase difference  $\epsilon$  between the modules is smaller than  $\delta$  at integer distance  $\ell$  from the origin (Figs 1c and 2b, inset, [Methods](#)):

$$\epsilon(\ell) = \|\psi_1(\ell) - \psi_0(\ell)\| = \|\psi_1(\ell)\|, \tag{1}$$

where  $\|\psi\|$  means distance from the nearest integer.

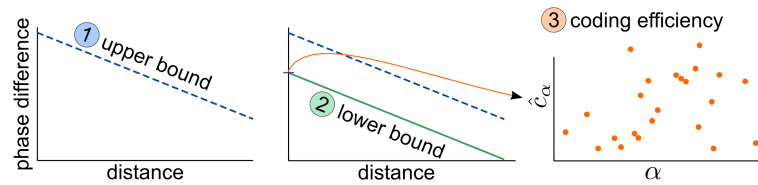
To analyse the coding properties of the grid cell system, we follow the same three logical steps both in the two module and in the multi-module case (Fig 3). First, we show the existence of an upper bound on how the maximal phase difference  $\epsilon(\ell)$  between the modules decreases with the distance. Intuitively, this upper bound expresses the fact that interference between the modules necessarily becomes stronger at larger distances. Second, we demonstrate that for appropriately chosen scale ratios a lower bound on the phase difference also exists and is parallel with the upper bound (Fig 3). For these scale choices catastrophic interference is avoided until a critical distance, that depends on the noise level in the system. Importantly, the slope of the bounds depends only on the number of modules, but not on the choice of the scale parameters. Therefore the efficiency of the scale choices (the magnitude of the critical distance) can be characterised by the offset parameter,  $c_\alpha$  (defined below), associated with the lower bound. Thus, our third step is to estimate  $c_\alpha$  for various choices of the scale parameter  $\alpha$ .

Our analytic derivations provide an estimate for the asymptotic performance of the system that is valid in the low-noise limit. The main advantage of our approach is that it provides strict bounds on the achievable coding efficiency that can be used as a metric to evaluate the efficiency of different scale choices at realistic noise levels, two or more modules alike. As we found using numerical simulations, these bounds can be approached with random scale choices at realistic levels of noise and number of modules.



**Fig 2. Interference depends on the choice of the scales.** (a) Interference with rational scale ratio. Left: Representative posteriors  $P(x|s)$  for two modules with scale 1 and  $\alpha = 3/2$ . Encoding becomes ambiguous at distance 3 from the origin where perfect interference occurs ( $3 = 2\alpha$ ). Right: Phase plot of the two modules, with the colour (red to blue) encoding the distance from the origin (see the coloured line below the left panel). Perfect interference occurs when the phase-curve overlaps with itself. (b) Interference with  $\alpha = 1.76 \dots$ , which is close to  $7/4$  and therefore leads to strong interference at distance 7. Right: Interference occurs when the distance between two neighbouring segments of the phase curve becomes smaller than the limit set by the neuronal noise (grey squares of side  $\delta$  around the origin, see inset). Note, that both grids are around phase 0.3 at the distance 2.3 without interference. (c) Interference with  $\alpha = \sigma \approx 1.618$ , which is the golden ratio. Interference still becomes stronger at larger distances, (e.g. at distance 5, since  $\sigma \approx \frac{5}{3}$ ). Interference in grid codes is related to the approximation of irrationals with rational numbers having small denominators (see text for further details). Right: Interference is inevitable since the phase space has a limited volume.

<https://doi.org/10.1371/journal.pcbi.1005922.g002>



**Fig 3. Logical steps of the argument.** First, we provide an absolute upper bound on the phase difference in the function of the distance, that is linear in log-log scale (left). Second, we show, that for certain scales a lower bound also exists (middle). Third, we characterise the efficiency of the scales ( $\alpha$ ) by their offset,  $c_\alpha$  (right).

<https://doi.org/10.1371/journal.pcbi.1005922.g003>

### Coding is extremely sensitive to the scale ratio with two modules

We can formalise the problem of interference between two modules as having a pair of integers  $k$  and  $\ell$  with  $\ell \approx k\alpha$ , meaning that module 2 (with scale  $\alpha$ ) is close to being in phase 0 at distance  $\ell$ , which would cause ambiguity between the coding of the spatial point  $\ell$  and the origin. This is formally identical to the number theoretic question of the approximability of the scale  $\alpha \approx \ell/k$  with rationals having numerator  $\ell$ , also known as Diophantine approximations (Fig 2b and 2c).

Hurwitz’s theorem [19, 20] states that for all irrational numbers  $\alpha > 1$  there are infinitely many relative primes  $k, \ell$  such that the error of the approximation, defined as

$$\epsilon(\ell) = |k - \ell/\alpha|, \tag{2}$$

is smaller than the upper bound:

$$\epsilon(\ell) < \frac{1}{\sqrt{5}} \frac{1}{\ell}. \tag{3}$$

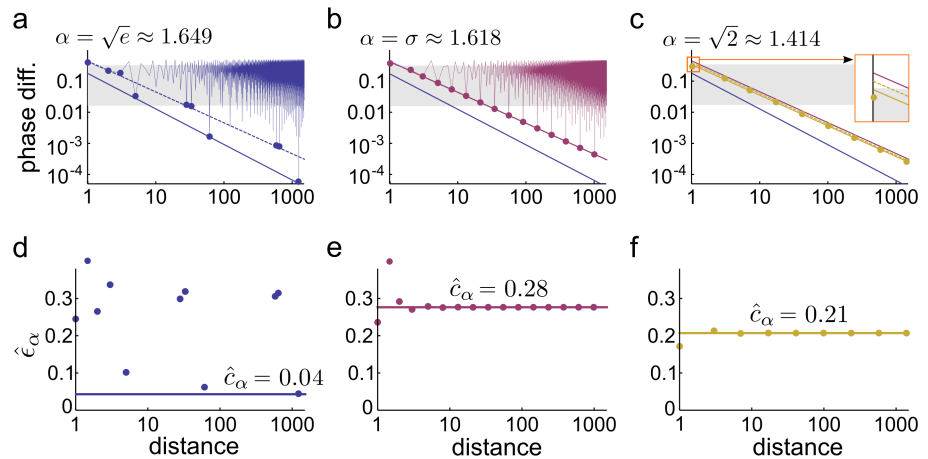
Note that the approximation error  $\epsilon(\ell)$  (Eq 2) is the same as the phase difference between the modules, defined in Eq 1, since  $\psi_2(\ell) = [\ell/\alpha] \bmod 1 = |k - \ell/\alpha|$  for an appropriately chosen integer  $k$  (Fig 2b and 2c). We call  $\epsilon(\ell)$  ‘approximation error’ only when we are talking about approximating irrationals with integer ratios while in the context of grid cells we will call  $\epsilon(\ell)$  the ‘phase difference’.

Applied to the grid cells, Hurwitz’s theorem provides an upper bound on how the phase difference between the modules shrinks with the distance. Specifically, the theorem states that there are infinitely many integer distances  $\ell$ , where the phase difference is smaller than  $\epsilon(\ell) \propto 1/\ell$  (Fig 4a–4c, dashed lines), implying that on the long run interference can not be avoided no matter how carefully we choose  $\alpha$ . This is a fundamental upper bound on the efficiency of coding with grid cells.

The critical distance where the phase difference  $\epsilon(\ell)$  leads to interference, and the representation of the position becomes ambiguous, depends on the noisiness of the two modules,  $\delta$  and  $\alpha\delta$ . Interference occurs if there is a spatial point  $x$  for which both  $|x - k\alpha| < \alpha\delta$  and  $|x - \ell| < \delta$  for integers  $k, \ell$ , or equivalently, if  $|k\alpha - \ell| < (1 + \alpha)\delta$ . Hence, by the definition of  $\epsilon(\ell)$  (Eq 2) the coding is ambiguous near  $\ell$  if and only if

$$\epsilon(\ell) < \frac{1}{\alpha} (1 + \alpha)\delta. \tag{4}$$

Therefore, no matter how we chose  $\alpha$ , we can expect ambiguity at distances  $\ell$  from the origin if



**Fig 4. Coding efficiency in two modules.** (a-c) The phase difference in the function of distance for different values of  $\alpha$ . The zigzag line indicates the phase difference (PD) at all integer distances, circles indicate record low PD. Dashed line shows the theoretical upper bound of the PD, solid line shows the numerical fit on the lower bound (allowing finitely few exceptions at low  $\ell$ ). Note, that the lower and the upper bound coincides in b. Also note the  $1/\ell$  scaling of PD for algebraic scale ratios (b-c). Grey shading indicates the range of PD smaller than the neuronal noise,  $(1 + \alpha)\delta$ . (d)-(f) The scaled phase difference ( $\hat{\epsilon}_\alpha(\ell)$ , Eq 7) for different distances and scale ratios. The highest constant under which there are only finitely few values of  $\hat{\epsilon}_\alpha(\ell)$  at small distances estimates  $\hat{c}_\alpha$ . The value of  $\hat{c}_\alpha$  is slightly higher for the golden ratio (e) than for  $\sqrt{2}$  (f), and much larger than for non-algebraic numbers (d,  $\alpha = \sqrt{e}$ ,  $\hat{c}_\alpha = 0.04$ ).

<https://doi.org/10.1371/journal.pcbi.1005922.g004>

the noise in the system is larger than the upper bound on efficiency provided by Eq 3, i.e.

$$\ell > \frac{\alpha}{\sqrt{5}(1 + \alpha)} \frac{1}{\delta}, \quad (5)$$

that is, at distance of order  $1/\delta$ . Consequently, it is impossible to code position with two modules better than this bound.

The question arises then whether the above theoretical bound is achievable, at least for some appropriately chosen  $\alpha$ . The answer is yes, namely the upper bound in Eq 3 is sharp for the golden ratio  $\alpha = \sigma := \frac{\sqrt{5}+1}{2} \approx 1.618$ . Practically, this also introduces a limit on  $\epsilon(\ell)$ , saying that the phase difference between the modules remains always larger than a specific lower bound:

$$\epsilon(\ell) > \left( \frac{1}{\sqrt{5}} - \epsilon \right) \frac{1}{\ell} \quad (6)$$

except for a couple of small distances, even for arbitrary small  $\epsilon > 0$  [19] (Fig 4b).

It may sound strange that there are finitely many exceptions, but in our simulations we found only a few instances with  $\ell$  being small (Fig 4e). Therefore, if the ratio of the two grid modules equals the golden ratio then the phase difference between the two modules is guaranteed to be larger than the lower bound defined by Eq 6. Since  $\epsilon$  can be arbitrarily small, the lower bound for the golden ratio approaches the theoretical upper bound (Eq 3) and  $\sigma$  is an optimal choice for the scale ratio to avoid interference in case of two modules. To give a geometric picture, the golden ratio guarantees approximately uniform coverage of the phase space for both short and arbitrarily long distances (Fig 2c, right).

However, it turns out that there are many good choices [20]: for any algebraic integer  $\alpha$  of order 2 (i.e. irrational which is a root of a polynomial of degree 2 with integer coefficients, see

(Methods) there exists a maximal positive constant  $c_\alpha > 0$  such that

$$\epsilon(\ell) \geq c_\alpha \frac{1 + \alpha}{\ell} \tag{7}$$

holds except for a couple of small distances (Fig 4c–4f, [21]). Hence, from Eqs 4 and 7 we see that the representation is unambiguous whenever  $c_\alpha(1 + \alpha)/(\alpha \ell) > \delta(1 + \alpha)/\alpha$ , that is up to

$$\ell \leq L_{\max} := \frac{c_\alpha}{\delta} \tag{8}$$

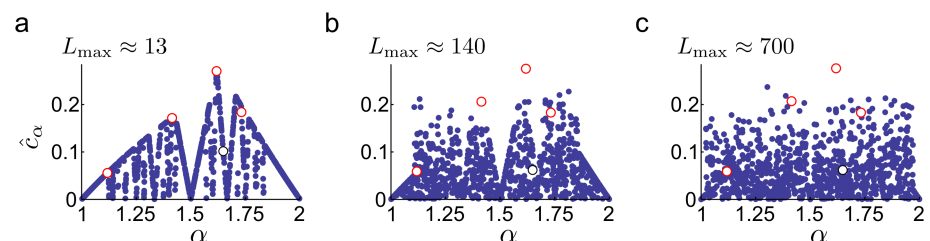
for all  $\delta$  which is small enough. This last condition on the magnitude of the noise is only needed to exclude the possible exceptionally small  $\ell$  distances in Eq 6, which in practice is not a crucial condition (Fig 4d and 4e).

The constant  $c_\alpha$  is the single parameter that determines the critical distance up to which encoding is unique (coding range) as well as the information rate of the system (Methods). Therefore, we use  $c_\alpha$  to compare the efficiency of different choices of  $\alpha$  (Fig 4d–4f). We have already noted that for the golden ratio the lower and the upper bounds (Eqs 3 and 7) coincide (Fig 4b), but the critical distance may be larger for some  $\alpha$  even if the corresponding lower bound on the phase difference is weaker, since the upper bound also depends on  $\alpha$  (Eq 5).

We estimated the value of  $c_\alpha$  for various scale ratios at different noise levels (Methods). Unlike for algebraic numbers,  $\hat{c}_\alpha$  of real numbers depends on the distance range used for the estimation, which we controlled by setting different intervals for  $\delta$  in the simulations.

Our simulations confirmed that  $\sigma$  is the best scale ratio choice in case of two modules with  $\hat{c}_\sigma = 0.28 \approx \frac{\sigma}{\sqrt{5}(1+\sigma)}$ , but also showed that, on both short and long run,  $\hat{c}_\alpha$  is extremely sensitive to the choice of  $\alpha$  (Fig 5): in case of a small error in the tuning of  $\alpha$ , the efficiency can drop substantially and  $\hat{c}_\alpha$  becomes practically 0, implying that in the immediate neighbourhood of the optimal  $\alpha$ , there are close to pessimal grid cell configurations. This is because the lower bound on the phase difference (Eq 7) requires  $\alpha$  to be an algebraic number, and in an arbitrary small neighbourhood of any algebraic number there are (infinitely) many non-algebraic numbers, i.e., transcendental numbers ( $\alpha = \sqrt{e}$ , Fig 4a) or rational numbers ( $\alpha = 3/2$ ,  $c_\alpha \rightarrow 0$ , Fig 2a). As non-algebraic irrational numbers can be much better approximated with rationals than algebraic numbers, non-algebraic grid scale ratios will lead to much stronger interference between the two modules, but only at distances moderately large compared to the scale of the modules (Fig 5).

The extremely rough landscape of  $c_\alpha$  renders optimisation for  $\alpha$  an especially difficult problem: it is very unlikely that a biological system would be able to find the global optimum for



**Fig 5. Approximate  $c_\alpha$  values as a function of  $\alpha$ .** The values are shown for 1000  $\alpha$  randomly selected from the interval (1, 2). The  $c_\alpha$  of algebraic ( $\sqrt{2}$ ,  $\sqrt{3}$ ,  $\sigma$ ,  $\sigma - 1/2$ ) and non-algebraic ( $\sqrt{e}$ ) irrationals are also shown in red and black, respectively. We estimated  $c_\alpha$ , based on Eq 8, as  $c_\alpha \approx \inf\{L_{\max} \delta \mid \delta_{\min} < \delta < \delta_{\max}\}$ , for different noise ranges  $[\delta_{\min}, \delta_{\max}]$  in the three panels: (a)  $0.05 < \delta < 0.2$ , (b)  $0.005 < \delta < 0.05$ , (c)  $0.001 < \delta < 0.01$ . Approximate  $L_{\max}$  values are indicated on the top of the panels for  $\alpha = \sigma$ .

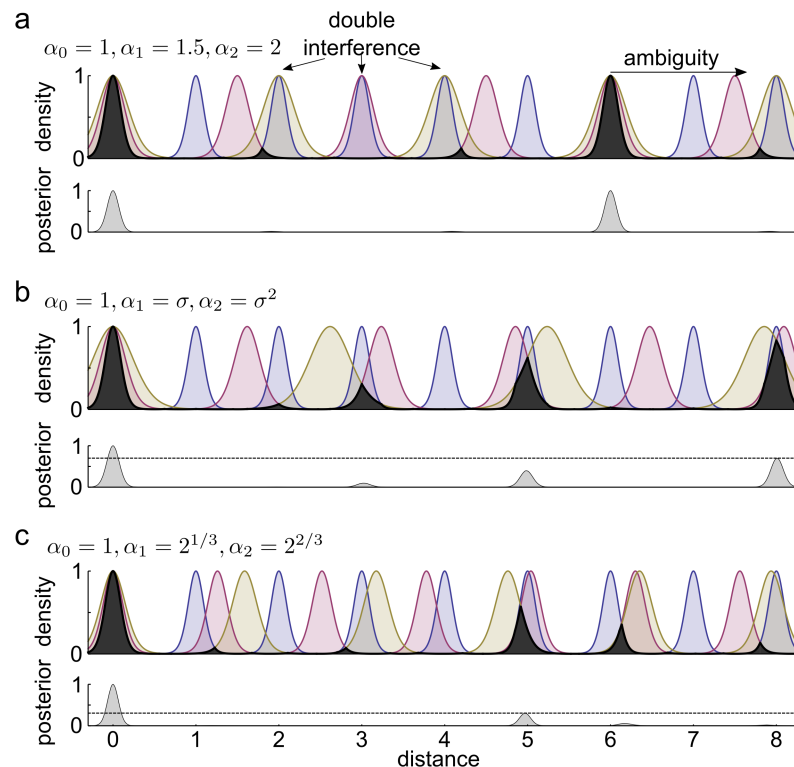
<https://doi.org/10.1371/journal.pcbi.1005922.g005>

the scale ratio of two grid modules and a relatively small mistuning from a local optimum could significantly deteriorate the efficiency of the system. Therefore, at least in the case of two modules, it seems to be impossible to achieve asymptotically optimal scale ratio for the grid cells.

### Generalisation to multiple modules

To derive the general solution for  $M$  grid modules, we focus on a set of 1-dimensional grids with scales  $\alpha_0 = 1 < \alpha_1 < \dots < \alpha_{M-1}$ . Spatial representation is unambiguous up to a distance  $L$  from the origin if there is at least one module for which the phase is significantly different from 0 (Fig 6). Interestingly, avoiding interference between adjacent modules (giving  $\alpha = \sigma$ ) is not a good solution, since it leads to interference between the distant modules wherever the adjacent modules are in close apposition (Methods). The logic of the general solution for multiple modules is the same as in the case of two modules. Here we only state the main results and the technical details of the analysis can be found in the Methods.

First, we show that a similar upper bound exists for the maximal phase difference between the modules. Compared to the two-module case, the bound is weaker when  $M > 2$  as the



**Fig 6. Interference of  $M = 3$  modules with different choice of scales.** (a) Top: Posterior densities for three modules with rational scale ratios. The overlap between the modules is shown in black, its height indicates the interference of the three modules as a function of distance from the origin. The representation becomes ambiguous only if all 3 modules interfere, as at distance 6. Bottom: Ambiguity in position coding quantified by the multi-modality of the combined posterior. (b) Posterior densities for three modules with pairwise optimal scale ratios. The scales are 1 (blue),  $\sigma$  (red), and  $\sigma^2$  (olive), where  $\sigma$  is the golden ratio. As we have more modules (3) than the order of  $\sigma$  (2), wherever any two modules interfere with each other, then they interfere with the third as well: at distance 8 the three peaks almost coincide. (c) The same as in (b) for scales 1,  $2^{1/3}$ ,  $2^{2/3}$ , powers of a third order algebraic number. Although pairwise interference can be very strong between any pairs (e.g. at distances 5, 6.2 and 8), the total interference is substantially lower than in panel b (bottom).

<https://doi.org/10.1371/journal.pcbi.1005922.g006>



phase difference scales only with  $1/\ell^{1/(M-1)} \gg 1/\ell$  meaning that it ensures simultaneous interference between all modules only at much larger distances.

Second, we found that the upper bound can be satisfied, up to a constant multiplier,  $c_{\mathcal{A}}$ , for algebraic scale ratios (Methods). Specifically, if the scales of  $M$  modules form a geometric series with common ratio  $\alpha$  being an algebraic number of degree  $M$ , the upper bound is tight, meaning that the phase difference does not shrink faster than  $1/\ell^{1/(M-1)}$ . Intuitively, this scaling indicates that there is always at least one pair of modules for which the phase difference at the integer distance  $\ell$  from the origin is larger than the lower bound.

The critical distance  $L_{\max}$  up to which coding is unambiguous can be expressed as (cf. Eq 8):

$$L_{\max} := \left(\frac{c_{\mathcal{A}}}{\delta}\right)^{M-1}, \tag{9}$$

for all  $\delta$  which is small enough, where  $c_{\mathcal{A}}$  and its estimate  $\hat{c}_{\mathcal{A}}$  are defined analogously as in the two modules case (see Methods for the definition). Intuitively, Eq 9 expresses an exponential scaling of the maximal distance uniquely represented by a population of grid cells with the number of grid modules,  $M$ . The coding range of a particular set of the grid scales,  $\mathcal{A} = (\alpha_1, \dots, \alpha_{M-1})$ , depends on both the noise in the system and on the basis of the exponential  $c_{\mathcal{A}}$ .

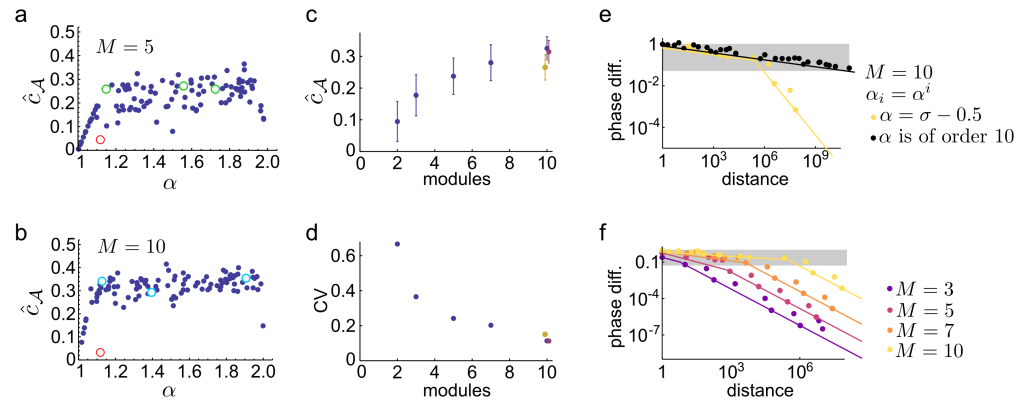
Interestingly, the above described geometric sequence of algebraic numbers are the only known explicit examples of badly approximable vectors (to the best of our knowledge). However, it is known that there are much more such vectors which do not form geometric sequences [22], therefore the scale ratio of a well-tuned MA grid cell system does not have to be constant.

The expression about the exponential scaling (Eq 9) is similar to the capacity estimates of Fiete et al. [3] (see their Eq 6) obtained using a combinatorial upper bound and numerical simulations. Importantly, our analytic derivation also provides insight about why certain grid systems are more efficient than others and give examples for efficient grid scales. Moreover, when  $c_{\mathcal{A}} = 0.5$  our formula for the capacity of the grid code becomes identical to the theoretically maximum capacity found in the case of nested coding [4, 6].

In the next sections we first numerically estimate the value of  $c_{\mathcal{A}}$  for various choices of the grid scales  $\mathcal{A}$  and then we show that with sufficiently large number of modules  $c_{\mathcal{A}}$  is guaranteed to approach its theoretical maximum  $c_{\mathcal{A}} = 0.5$  for randomly chosen grid periods.

### Numerical estimation of the $c_{\mathcal{A}}$

We developed an efficient method to numerically estimate the value of  $c_{\mathcal{A}}$  for various parameter settings that is based on the simultaneous Diophantine approximations of a set of irrational numbers (Methods). Using realistic noise levels we found that, in contrast to the case of two modules, the sensitivity of the coding efficiency to the choice of  $\alpha$  gradually vanishes when the number of grid modules is increased (Fig 7a and 7b). In particular, with  $M = 10$  modules  $\hat{c}_{\mathcal{A}} \in [0.2, 0.4]$  for almost all choices of the grid scales (Fig 7b), both when the scales follow a geometric series with a common scale ratio  $\alpha$  (Fig 7b) and when all the  $M$  scales are chosen from the bounded interval  $(1, 2)$ . We also found that  $\hat{c}_{\mathcal{A}}$  vanishes only for pathological examples such as rational numbers or powers of the second order algebraic number  $\alpha = \sigma - 1/2 \approx 1.118$  (Fig 7a and 7b, red). The only random scale choice that significantly degrades the performance is when  $\alpha \approx 1$  (Fig 7a and 7b) in which case all grid modules have nearly identical spatial scale.



**Fig 7. Robustness of the grid code with multiple modules.** (a)-(b): Values of  $\hat{c}_A$  estimated for 100  $\alpha$  randomly selected from the interval (1, 2) with  $M = 5$  (a) and  $M = 10$  (b) (see also Fig 5a for  $M = 2$ ). The scales form a geometric series, i.e.,  $\mathcal{A} = \{\alpha^i\}_{i=1}^M$ . Red circle indicates  $\hat{c}_A$  for a second order algebraic number  $\alpha = \sigma - 1/2 \approx 1.118$ . Green (cyan) shows  $\hat{c}_A$  for 5th (10th) order algebraic numbers, respectively. Noise level is the same as in Fig 5a ( $0.05 < \delta < 0.2$ ). (c)-(d): Mean (c) and coefficient of variation (d) of  $\hat{c}_A$  evaluated on the range  $\alpha = \{1.1, 1.9\}$ . For  $M = 10$  the  $\hat{c}_A$  is also shown for two alternative selection of the scales: if all 10 scales are selected randomly from the interval [1, 2] (olive) and when  $\alpha$ s form a geometric series perturbed as  $\mathcal{A} = \{(1 + \epsilon_i)\alpha^i\}_{i=1}^{10}$ , where  $\epsilon_i$  are i.i.d. uniform random variables on the range  $[-0.01, 0.01]$  (purple). (e)-(f) Phase difference ( $\hat{\epsilon}_A(\ell)$ , Eq 25) in the function of the distance,  $\ell$ . (e) Effect of number theoretic properties with  $M = 10$ . When  $\alpha$  is the root of the 10th order polynomial  $x^{10} - x^7 - 1$ ,  $\alpha \approx 1.12725$ ,  $\epsilon(\ell)$  decays as  $1/\ell^{1/9}$  (black). When  $\alpha$  is second order,  $\alpha = \sigma - 1/2 \approx 1.11803$ , the initial decay is similar, but after a critical distance at  $\ell \approx 10^6$  the decay becomes  $1/\ell$  (yellow). (f) The critical distance grows with the number of modules ( $\alpha_i = (\sigma - 1/2)^i$ ). Grey shading in (e-f) indicates the range of phase difference smaller than noise (Eq 26).

<https://doi.org/10.1371/journal.pcbi.1005922.g007>

To quantify the sensitivity of the grid system against the choice of the scale parameters we calculated the mean and the coefficient of variation of  $\hat{c}_A$  with random choices of  $\alpha$  (Fig 7c and 7d). We found, that the average  $\hat{c}_A$  increased monotonically with the number of grid modules indicating that the system’s performance becomes closer to the ideal  $c_A = 0.5$  value as the number of modules increased (Fig 7c). Moreover, the variability of  $\hat{c}_A$  consistently decreased with the number of modules reflecting the improved robustness of the system to the choice of grid periods (Fig 7d). Therefore not only the maximal coding distance increases exponentially with the number of modules, but the basis of the exponential,  $c_A$ , also increases.

To further investigate the mechanisms responsible for the robustness of the system, we numerically evaluated the minimal phase difference between the modules,  $\epsilon(\ell)$ , in the function of the distance (Fig 7e and 7f). In line with the predictions of the theory (Eq 25), we found that the phase difference decreased with  $\ell^{-1/(M-1)}$ , i.e., with a small negative power of the distance for  $\alpha$  being an order  $M$  algebraic number (Fig 7e, black). For suboptimal  $\alpha$ -s, the scaling of the phase difference was nearly optimal up to a critical point beyond which the scaling followed the algebraic rank of  $\alpha$  (i.e., second order  $\alpha$  scales with  $1/\ell$ , Fig 7e, yellow). Importantly, this critical point, where the transition occurs between ideal and number theoretical scaling is located at increasingly larger distances when the number of modules is increased (Fig 7f). Therefore, the asymptotic, number theoretical properties of the grid periods have a gradually lower impact on the performance of the system in the distance range limited by the intrinsic variability on neuronal spiking (Fig 7e and 7f, background shading).

These observations suggest that even random scale choices might achieve optimal performance as the number of modules grow. In the next section we make this statement mathematically precise and demonstrate that indeed,  $\hat{c}_A$  approaches its theoretical maximum, 0.5, when the number of modules grow and the scales are chosen uniformly at random from a bounded interval.

## Capacity of non-geometric grid scales

Our number theoretic argument (Eq 9) alone does not imply exponential capacity, since it does not exclude the possibility that the base of the exponential,  $c_A$ , converges to 0 as  $M$  increases (although we observed the opposite trend, see Fig 7c). In this section we investigate the asymptotic properties of the grid code when the number of modules increases and the relative uncertainty  $\delta$  of the modules remains fixed. Here we only state these results informally, and leave the precise statements and the slightly technical mathematical proof to the Methods.

The main idea behind the proof is that the phase of a given module at particular distance  $x$  from the origin depends only on the scale of that module,  $\alpha$ . If the scale is randomly chosen from a bounded interval  $[1, \alpha_{\max}]$ , then the phase is also a random variable with probability distribution approaching the uniform distribution as the distance increases. Then, the probability of simultaneous interference between  $M$  modules, that is, the probability of all modules being near phase 0 at some distance  $x$ , is proportional to the volume of an  $M$ -dimensional hypercube, which is  $V = (2\delta)^M$ , where the side of the cube is  $2\delta$ . The ratio of the volume of the hypercube and the unit cube (the number of distinguishable phases) diminishes exponentially with  $M$ , and the total distance (expressed in units of  $\alpha^0 = 1$ ) covered without ambiguity is  $\frac{2\delta}{V} \propto \left(\frac{1}{2\delta}\right)^{M-1}$ . Specifically, our statement is, roughly speaking, that if  $0 < \delta < 1/2$  is fixed,  $M$  is large enough, and the module scales are drawn uniformly at random from a not too narrow bounded interval, e.g. from (1, 2), then the representation is unambiguous up to the exponential distance

$$\ell < \left(\frac{c_A}{\delta}\right)^{M-1} \tag{10}$$

with probability approaching 1, and  $c_A$  approaching 1/2. Although the above statement applies only for  $M \rightarrow \infty$ , and it does not provide examples for efficient scale choices for finite  $M$ , we emphasise that this result is stronger than our previous derivation (Eq 9) in four aspects: First, our previous derivation (Eq 9) allowed  $c_A$  to tend to 0 as  $M$  increased. Now we showed that this does not happen for random scale choices, rather the value of the constant tends to its theoretical maximum,  $c_A = 0.5$  [6] for large  $M$  with high probability, confirming our previous numerical results (Fig 7c). Second, one can achieve this nearly optimal performance without increasing the scales exponentially, with the scales chosen from a bounded interval. Third, this almost optimal efficiency is not only reached for some appropriately chosen scales, but for almost all choices. Fourth, near-optimal performance is guaranteed for 2 or higher dimensional grid codes even if the modules are randomly rotated relative to each other or in the absence of long-range coherence within the modules.

Thus, our results demonstrate that no meticulous tuning of the grid scales is required for close to optimal grid system performance.

## Discussion

In this paper we developed a novel analytic technique to investigate the coding properties of grid cells. Using this technique, which is based on Diophantine approximation of real numbers by fractions of integers, we were able to derive several novel and non-trivial properties of the grid cell code. First, we demonstrated that on the long run, the capacity of the system depends heavily and chaotically on the number theoretic properties of the scale ratio between the successive modules. To achieve optimal performance in a system with  $M$  modules the scale ratio has to be an algebraic number of order  $M$ . Second, we showed that in the presence of neuronal noise the capacity of the grid code becomes increasingly more robust to the choice of the scale parameters when the number of modules is increased: when  $M > 2$ , randomly chosen scales

perform nearly as well as the optimal scales. Finally, we demonstrated that the capacity of MA and nested grid codes are asymptotically identical (in the large  $M$  limit), even for randomly chosen scale parameters for the MA codes.

### Exponential coding range

Previous works used specific assumptions to derive exponential coding range for the grid cell coding system: they assumed either a nested coding scheme [5, 6] or presumed that the phase space is covered evenly and that the readout noise in a given module decreases when the number of modules increases [3, 14]. Here we generalised these findings and demonstrated that nested and MA codes have asymptotically equal capacity.

When we studied the capacity of MA codes we realised that achieving uniform coverage of the phase space is not trivial in the case of two modules, but can only be attained with appropriately chosen scales. Specifically, we recognised that approximately uniform coverage of the phase space by the phase curve at arbitrary distances is guaranteed if the scale ratio between the two modules is an algebraic number of order 2. Using our formalism allowed us to generalise this intuition for arbitrary number of grid modules and to demonstrate that even a random choice of grid scales guarantees uniform coverage of the phase space when the number of modules is high.

We also relaxed the assumption of an earlier study [14] that the total amount of the noise remains constant in the grid system even when the number of modules is increased, i.e., the readout error of each module decreases with  $M$ . Here we derived these results using the more general assumption that the coding precision of each module is independent of  $M$  and proportional to the scale of the module.

We confirmed our analytical results by extensive numerical simulations regarding the simultaneous interference between grid systems with various choices of the scale parameters. In line with previous results [3, 9], our simulations supported that the grid system is robust to the choice of the scale parameter and that the coding range is exponential in the number of modules.

### Nested coding versus MA code

Although the efficiency of the coding investigated in this paper is slightly worse than that of the optimal nested coding [5, 6], MA codes also have several advantages. First it uses orders of magnitude smaller scale lengths than the maximal distance up to which the coding works properly. The largest grid scales measured experimentally are  $\sim 3$  m [23, 24] and extrapolations based on the dorso-ventral location of the recording electrodes within the entorhinal cortex extend to  $\sim 10$  m [13], a period still substantially smaller than the typical distances travelled daily by rodents (several hundreds of meters [25]) or bats (several kilometres, [26]; see also [27]).

Second, while the consequence of a module failure simply decreases the capacity of the system in the case of MA coding, it can have more dramatic effect in nested codes: Although malfunction of the largest or smallest module reduces either the capacity or the resolution of nested codes, respectively, the lack of intermediate modules functionally breaks the interaction between the remaining modules decreasing both the resolution and the capacity of the system in a disproportionate manner.

Third, once the scales are optimised for a given noise level, the coding range of nested grid codes does not depend on  $\delta$ . Therefore, contrary to MA codes, it is not possible to increase the capacity by inserting more neurons into the same modules or by observing more grid cells from the same set of modules. Conversely, the functioning of the nested codes critically

depends on accurate decoding of each module: If the readout neuron does not have access to enough presynaptic neurons from a given module, then the corresponding posterior becomes too wide leading to interference between the modules. This has similar consequences as the absence of the given module in nested codes. In contrast, in MA codes the coding properties remain similar for postsynaptic neurons receiving different number of synapses from different modules, although the coding range is the function of the precision available for the observer (Eq 9).

When encoding dynamic trajectories instead of static locations, the number of neurons required to participate in a given module decreases quadratically with the scale of the module, i.e.,  $n_i \sim 1/\alpha_i^2$  [8]. For example, if representing the position in the 2D space with some fixed accuracy with  $\alpha_i = 0.2$  m requires  $\sim 4000$  neurons then  $\alpha_j = 2$  m needs only  $\sim 40$  neurons. This scaling implies that the coding range of the nested grid system can be easily and parsimoniously extended by adding a new module with larger scale but containing only relatively few neurons. Although the relationship between the number of neurons in a module and its scale holds also for MA codes, the total number of neurons required to achieve similar coding range can be substantially smaller in nested codes.

Another consequence of dynamical coding is that the time constant of the readout has to be matched to the scale of the grid modules [8]. As the grid scale varies over a large range in the case of nested codes, the postsynaptic neuron has to integrate inputs from different grid cells with time constants ranging from 1 ms to 1 second [8]. In MA codes, the modules have similar scales and their outputs can be integrated with similar time constants.

Finally we note that nested coding and MA coding are not mutually exclusive: although they imply fundamentally different way of decoding the same positional information [7, 14], but both can be present in the same system. The MA code has a larger coding range if  $c_A > \alpha\delta$  so it is favoured by small  $\alpha$  (small differences between scales) and small  $\delta$  (high accuracy). Even in this case locations within the largest grid scale can be decoded as in nested coding, while MA decoder is required beyond this distance.

## Planar grid cells

In the Methods we show that the coding capacity of two or higher dimensional grid cells depend on the same number theoretic properties, and therefore the results obtained in dimension one extend to planar or cubic grid cells as well [28], provided that the main axes of the different modules remain aligned with each other.

If the two dimensional grid modules are rotated compared to each other, then the scale choices which perform well will be different from the scales that are optimal for axis aligned modules. Consider for example that  $\alpha = \sqrt{3}$ , which is a relatively good choice for  $M = 2$  (right-most red circle in Fig 5a–5c), leads to catastrophic interference at  $\ell = \alpha$  when the grids are rotated by  $30^\circ$ . Consequently, the incoherent reorientation of the grid modules during global remapping [13] renders the optimisation of the grid scales unfeasible. However, the main point of this paper is that we have shown analytically that almost all scale choices perform near optimally if the number of modules is high enough, which also applies for grid cells rotated uniformly at random relative to each other (Methods).

Moreover, the 2D grids does not need to show perfect triangular symmetry to achieve high capacity: environmental boundaries [29, 24] or non-euclidean geometry [30, 31] can distort the grid pattern, but as long as the distortion is coherent among modules, our theory applies unchanged. If the scales slightly vary on the long range, then our derivation based on the Diophantine approximations does not apply. However, our derivation stating exponential capacity for grid systems with many random scales (Methods) remains still valid.

## Optimization and robustness

The highly organised, regular patterns formed by the firing fields of grid cells suggest that the characteristics of the grids must be closely related to the computational function of these neurons: optimally representing and processing information about the spatial location of the animal [32, 33, 14, 4, 11]. Besides the general optimality of triangular grid-like firing fields for representing unbounded 2D space [28], recent theoretical work derived optimal scale ratio of successive grid modules in the case of nested coding [6, 7, 8].

These studies, using different assumptions, arrived at slightly different conclusions regarding the optimal value of  $\alpha$ . Stemmler et al., [7] fixed both the coding range to  $L_{\max} = 3$  for a pair of grid modules with scales  $\{1, \alpha\}$  and found that  $\alpha = 3/2$  minimises the ambiguity errors within that range. Mathis et al., [4] and Wei et al., [6] also fixed the coding range and minimised the number of neurons required to achieve a given resolution and provided both estimates for the maximal capacity of the grid cell coding system and a specific architecture (i.e., optimised nested codes) that achieves maximal efficiency. The optimal scale ratio for nested codes was found to depend both on the magnitude of the noise in the system and on the type of decoder [4, 6]. Rather than fixing the coding range, we were interested in grid codes that work for potentially unbounded environments and found a similar asymptotic capacity for MA codes using random grid scales. Although predictions derived from nested coding roughly agree with the average scale ratio observed in the entorhinal cortex [12, 13, 29], they do not explain the substantial amount of variability which characterises the data.

In our derivations we assumed that the decoding error of a given grid module is larger than  $\delta$  with some small probability. Inaccurate decoding of a single grid module can lead to disproportionately high error in the position representation if the subsequent time frames are decoded independently [14, 9]. However, the chance of catastrophic ambiguity errors can be substantially reduced if a dynamical decoder combines prior information representing the predicted spatial position with the location encoded by the incoming grid cell spikes [34, 14, 8].

Our results based on the Diophantine approximations requires that the scale of the modules are set precisely, so that the phase of the different modules does not drift relative to each other (i.e.,  $\alpha_i \frac{d\phi_i}{dx} = \alpha_j \frac{d\phi_j}{dx}$ ). Although theoretical considerations suggest that drift can not be completely suppressed in a noisy neuronal system [35, 36], whether different grid modules respond coherently to distortions caused by environmental manipulations is not known [15, 29, 24]. The remarkable robustness of the grid system's efficiency against the choice of the scale ratio suggests that grids with loosely set scale parameters could also obtain a similar performance. Indeed, our derivation using randomly selected grid scales does not require precisely set scale parameters yet it provides the asymptotically exponential capacity for the grid system.

The optimization principle assumes that substantial improvement in the performance of the system can be achieved with precise tuning of its parameters. In the present study we demonstrated that this is indeed the case in the absence of noise. However, even in this case, optimization would be almost unfeasible for three reasons. First, the coding range is an extremely irregular, discontinuous function of the scale parameter, making optimisation essentially a trial and error game. Second, a scale parameter that is optimal for a given number of modules is guaranteed to be inefficient when the number of modules is increased precluding the possibility of pairwise or modular optimization. Finally, the optimal grid scales depend on the rotation of the modules relative to each other, which can change independently during changes in the environment [13].

However, taking the variability of neuronal firing into account changes the picture dramatically. We demonstrated that when the coding accuracy of grid modules is limited by neuronal noise, the capacity of the system becomes surprisingly robust to the choice of the scale

parameters making its optimization unnecessary. Note, that even if the grid periods are not optimized across modules, generating the regular, periodic firing fields of grid cells demands accurate integration of velocity inputs [37, 36] and repeated error correction [38, 35], both requiring the precise tuning of single neuron and network parameters within a given module. In conclusion, our study demonstrates that the capacity of the grid cell system is nearly optimal with randomly chosen grid scales, and, instead of accurate parameter tuning, the experimentally observed scales could reflect the combined effect of random fluctuations and a gradient in the cellular properties along the dorso-ventral axis of the entorhinal cortex [39, 40].

## Predictions

Our finding, that grid cells have an exponentially large coding range even with randomly chosen grid scales of similar magnitudes makes several important predictions. First, MA coding predicts that the coding range is substantially larger than the largest grid period. Since grid cells are likely to be involved in path integration [32, 41] this prediction could be tested by probing path integration abilities of rodents beyond distances of the largest grid period [42].

Second, in the case of MA coding, different modules have similar contributions to the coding range of the system. Therefore, the effect of targeted dMEC lesion (inactivating a single module, as in [43]) on the rat's navigation behaviour would be largely independent of the actual location of the lesion (i.e., which module is inactivated).

Third, since the performance of the system is independent of the precise choice of the grid scales, we expect a large variability in the scale ratio of successive grid modules both within and across animals. This prediction is consistent with the experimental data available [12, 13, 29], although further statistical analysis would be required to specifically determine the distribution of scale ratios.

Finally, we predict that the performance of the system is not particularly sensitive to incoherent changes in the scale parameter of a subset of modules during e.g., global remapping induced by environmental changes [16]. It has been shown that under certain conditions simultaneously recorded grid cells respond coherently within a module and independently across modules to environmental distortions [13]. To test the prediction of our theory, the behavioural consequences of incoherent realignment across modules should be assessed and compared with the effects of environmental manipulations inducing coherent realignment [29] or coherent distortion in the shape of the grid pattern [24, 29, 30, 31].

## Methods

### Grid cells in the 2D plane

Consider a system  $G^2$  of planar grid cells with a set of scales  $\mathcal{A}$ . Suppose that the axis of all modules are aligned and use the coordinate system

$$\mathcal{B} = \{[1, 0], [\cos(60^\circ), \sin(60^\circ)]\}, \quad (11)$$

which is naturally generated by the triangular lattice. To compare with consider the one dimensional grid cell system  $G^1$  which has the same number of modules with the same set of scales, and for which each module represents the position of the animal with the same relative precision. To achieve this, the two dimensional modules need squared as many cells, nevertheless they also able to distinguish between squared as many spatial positions within one period of the scale.

If  $G^2$  represents a planar position  $(x, y)_B$  ambiguously, i.e.,  $\psi(x, y)_B \approx \psi(0, 0)_B = 0$ , then clearly planar positions  $(x, 0)_B$  and  $(0, y)_B$  are also represented ambiguously. Therefore, the

corresponding one dimensional positions  $x$  and  $y$  are represented ambiguously by  $G^1$  as well. Conversely, if  $z$  is represented ambiguously by  $G^1$ , then  $(z, 0)_B, (0, z)_B$  will be ambiguous in  $G^2$ . Therefore, an ambiguity of position at a given distance from the origin in case of planar cells can be matched to an ambiguity at the same order of magnitude of distance in the one dimensional grid system, and vica versa. The above argument also shows that the same scale choices perform best for both one dimensional grid cells and two or higher dimensional ones when the axes are aligned with each other.

### Estimating the precision of a single module

We chose  $\alpha_0 = 1$  and fix the resolution of the system to  $\delta < \alpha_0$  (defined below) and investigate its coding range. A formally identical system with a fixed coding range and optimised resolution can be achieved by appropriately rescaling the grid scales.

We numerically estimated the precision of position coding by a single module by first simulating the motion of the animal as a one dimensional Gaussian random walk:

$$P(x_{t+1}|x_t) = \mathcal{N}(x_t, \Delta t D) \tag{12}$$

with  $\Delta t = 1$  ms temporal resolution and  $D = 0.005$  m<sup>2</sup>/s, which gives  $\approx 5$  cm displacement in 0.5 s [8]. We simulated the activity of  $N = [10, 300]$  grid cells from a single module. Grid cells had a circular tuning curve:

$$r^k(x) = r_{\max} \left( \sin \left( \frac{2\pi x}{2\lambda} - \phi^k \right) \right)^n + r_0 \tag{13}$$

with the following parameters:  $r_{\max} = 15$  Hz,  $r_0 = 0.1$  Hz,  $\lambda = 0.25$  m and  $\phi^k$  chosen to uniformly cover the interval  $[0, 2\pi]$ . The power  $n = 22$  was set to match the mean firing rate of the grid cells,  $\langle r(x) \rangle = 2.5$  Hz, to experimental data [16]. Larger ( $\lambda = 2.5$  m) grid spacing was modelled by decreasing the speed of the animal by a factor of 10 ( $D = 0.00005$  m<sup>2</sup>/s). The firing rate is shown in Fig 1b, right (olive).

Spike trains were generated as an inhomogeneous Poisson process with neurons conditionally independent given the simulated location:

$$P(s_t^k|x_t) = \text{Poisson}(\Delta t r^k(x_t)) \tag{14}$$

Spikes of the neurons in module  $i$ ,  $\mathbf{s}_{0:t,i}$  represent the spatial location of the animal with error  $\delta \alpha_i$  (i.e., with the same  $\delta$  phase error for all modules) which can be interpreted as the width of the (periodic) posterior probability distribution  $P(x|\mathbf{s}_{0:t,i})$ . For an ideal observer this posterior distribution quantifies how much a given spatial location is consistent with the observed spike pattern. The posterior distribution of the position was numerically calculated by recursive Bayesian filtering:

$$P(x_t|\mathbf{s}_{0:t}) \propto \prod_k P(s_t^k|x_t) \int P(x_{t-1}|\mathbf{s}_{0:t-1}) P(x_t|x_{t-1}) dx_{t-1} \tag{15}$$

The colormap in Fig 1b shows this posterior distribution with  $N = 50$  cells and  $\lambda = 0.25$  m.

Naturally, the width of the posterior depends on several factors, most importantly on the number of neurons observed in a given module and on the scale of the modules relative to the typical speed of the animal [8]. At each timestep the posterior distribution was fitted with a von Mises distribution with a location  $\mu_t$  and a concentration parameter  $\kappa_t$ . The width of the



posterior relative to the grid scale was estimated as:

$$\delta_t = \frac{\lambda}{2\pi\sqrt{\kappa_t}} \tag{16}$$

For analytic tractability, we use a bounded noise model in the derivations assuming that the location decoded from the spikes of a module is within  $\delta\alpha_i$  distance from the true location. To be conservative, we chose  $\delta$  to be the 99% of the empirical CDF of  $\delta_t$ . The largest  $\delta = 0.12$  was found with  $\lambda = 0.25$  m and  $N = 10$  cells. The smallest  $\delta = 0.01$  corresponds to the parameters  $\lambda = 2.5$  m and  $N = 300$  cells.

We assume that the modules are conditionally independent given the location of the animal, and hence position decoding, or representation, can be implemented by an ideal observer independently reading out the spikes,  $\mathbf{s}_i$ , emitted by the different modules:  $P(x|\mathbf{s}) = \prod_i P(x|\mathbf{s}_i)$ . When loosely talking about interference between the grid modules at a spatial point we refer to the interference between these periodic posterior distributions  $P(x|\mathbf{s}_i)$ , i.e., all module posteriors being larger than 0 at a location different from the origin (Fig 1c).

### Interference at integer distances

Since we measure the distance in units of the smallest grid scale ( $\alpha_0 = 1$ ), avoiding interference at integer distances from the origin also guarantees the absence of interference elsewhere, i.e., all positions in the interval  $[0, L]$  will be distinguishable by the grid code. Hence we loosely call  $\epsilon(\ell)$  defined in Eq 1 the phase difference, but note that it is the phase difference at integer distance  $\ell$ . Indeed, if the grid code was ambiguous confusing spatial locations  $x_1$  and  $x_2$ , then it would also confuse the origin with  $|x_1 - x_2|$  as well, since the phase differences of each module are the same between 0 and  $|x_1 - x_2|$  and between  $x_1$  and  $x_2$  (Fig 2b and 2c, right). But  $|x_1 - x_2|$  can be confused with the origin only if  $|x_1 - x_2|$  is an integer, that is a multiple of the smallest scale, 1. Note that this argument is correct only if the phase representation ambiguity of each module is independent of the actual position, which holds if we suppose that firing fields of cells from the same module are spaced evenly, which we do assume.

Graphically, interference between locations occurs when two segments of the phase curve come close to each other. Since the segments of the phase curve are parallel (Fig 2), and we started the phase curve in the origin, interference first occurs in the origin. Avoiding interference at the origin as much as possible at arbitrary distances thus also guarantees that the segments of the phase curve are separated from each other as much as possible, leading to a uniform coverage of the phase space [14].

### Definition of algebraic numbers

We call a real number  $\alpha$  algebraic of order  $n$  (positive integer), if  $n$  is the least integer such that  $\alpha$  is the root of a polynomial of degree  $n$  with integer coefficients. Algebraic numbers of order one are exactly the rational numbers. Another example is the golden ratio,  $\sigma$ , which is irrational, and is the root of  $x^2 - x - 1$ , a integer polynomial of degree two. Therefore,  $\sigma$  is an algebraic number of order two.

### Information rate

Since we fixed the resolution, the capacity of the code is proportional to the coding range. Moreover, as the coding precision of the modules was the same, we assume that the population size of each module is approximately  $N$  for grid scales chosen randomly from a bounded interval. The information rate of the grid system, defined as the ratio of the logarithm of the

capacity and the total number of conveyed bits [14] is

$$\rho \propto \frac{1}{\bar{r}NM} \log\left(\frac{c_x}{\delta}\right)^M \tag{17}$$

$$\propto \frac{1}{\bar{r}N} \log \frac{c_x}{\delta} \tag{18}$$

$$\propto \frac{1}{\bar{r}N} \log \frac{c_x \log N}{k} \tag{19}$$

where  $\bar{r}$  is the average firing rate of a grid cell and in the third line we used that  $\delta = k/\log(N)$  [14]. Thus, the information rate is independent of the number of modules and increases with  $\log c_x$ .

For a geometric code with scale ratio  $\alpha$  the optimal population size for dynamical decoding and constant  $\delta$  decreases as  $n_i = n_0/\lambda_i^2 = n_0/\alpha^{2i}$  where  $\lambda_i = \alpha^i$  is the scale of module  $i$  and  $n_0$  is the number of neurons in the first module [8]. In this case the total number of neurons in the population is

$$\mathbf{N} = \sum_{i=0}^M n_0/\alpha^{2i} = \frac{n_0 \alpha^2}{\alpha^2 - 1} \tag{20}$$

Since the total number of neurons does not grow linearly with the number of modules, the information rate becomes proportional to  $M$ :

$$\rho \propto M \frac{\alpha^2 - 1}{n_0 \alpha^2 \bar{r}} \log \frac{c_x}{\delta} \tag{21}$$

Although a constraint on the minimal number of cells per module will limit the finite information rate to remain finite, Eq 21 emphasises that adding further modules with larger periods increases the efficiency of the grid system if the number of cells per module is set optimally for dynamical decoding [8]. Although a geometric progression of scales is consistent with both nested and MA codes, the information rate is higher for optimal nested codes since they maximise  $\alpha$ .

### Interference with $M$ modules I: Golden ratio is suboptimal

In this section we demonstrate that a set of grid cells with scale ratio ( $\alpha$ ) optimally chosen between pairs of successive grid modules is close to being pessimal for efficient space representation for  $M > 2$ . Such pairwise optimisation leads to a set of scales showing geometric progression with the scale ratio being  $\alpha$ , i.e.,  $[1, \alpha, \alpha^2, \dots]$ , which is consistent with the experimental data [10, 12, 23, 13]. The representation of the position becomes ambiguous if all modules show interference at the same location, i.e., the phase of all modules are very close to 0 at distance  $\ell$  from the origin.

Consider for example the golden ratio  $\alpha = \sigma$ , which is a second order algebraic number, i.e., it is the root of the integer coefficient polynomial  $x^2 - x - 1$ . Therefore, the phase  $\psi_2(x) = (x \bmod \sigma^2)/\sigma^2$  of any spatial point  $x$  according to the third module can be simply expressed with the phase of the first two modules as

$$\psi_2(x) = [\psi_0(x) - \psi_1(x)] \bmod 1. \tag{22}$$

To see this, consider that by the definition of the phases  $\psi_i(x)$  when the animal is at distance  $x$

from the origin there are some integers  $\ell, k_1, k_2$  so that

$$x = \ell + \psi_0(x) = \sigma(k_1 + \psi_1(x)) = \sigma^2(k_2 + \psi_2(x)).$$

Using that  $\sigma^2 - \sigma - 1 = 0$  we get that

$$\sigma^2(\ell + \psi_0(x)) - \sigma^2(k_1 + \psi_1(x)) - \sigma^2(k_2 + \psi_2(x)) = 0.$$

Rearranging terms yields

$$\begin{aligned} \psi_2(x) &= \ell - k_1 - k_2 + \psi_0(x) - \psi_1(x) \\ &= [\psi_0(x) - \psi_1(x)] \pmod{1}. \end{aligned}$$

In other words, the phase of the third module provides no additional information given the phase of the other two modules. In particular, if both  $\psi_0(x)$  and  $\psi_1(x)$  are close to 0 (Fig 6b), then so is  $\psi_2(x)$  and hence the third module fails to resolve the ambiguity when the two first modules interfere. Similarly, if we have  $n$  grid cell modules with scales  $1, \alpha, \dots, \alpha^{n-1}$  with  $\alpha$  being an algebraic number of order  $k < n$ , then all of the  $n$  phases can be expressed by any  $k$  of them, leading to redundant and inefficient representation.

Clearly the same argument works not only for the powers of the golden ratio, but for powers of any algebraic number of order lower than the number of modules.

### Interference with $M$ modules II

To derive the general solution for  $M$  grid modules, we consider a set of 1-dimensional grids with scales  $\alpha_0 = 1 < \alpha_1 < \dots < \alpha_{M-1}$ . Again, the interference between the modules can be expressed by the simultaneous Diophantine approximation of the vector  $\mathcal{A} = (\alpha_1, \dots, \alpha_{M-1})$  using fractions of integers with the common numerator  $\ell$ , i.e.,  $\alpha_i \approx \ell/k_i$ . Importantly, a theorem by Dirichlet provides an upper bound on the efficiency of the approximation. Namely, for all  $(M - 1)$ -tuple of irrational numbers  $\alpha_1, \dots, \alpha_{M-1}$  we have infinitely many collections of integers  $k_0, k_1, \dots, k_{M-1}$  (with  $k_0 = \ell$ ), such that the approximation error defined as

$$\tilde{\epsilon}_{ij}(\ell) = |k_i \alpha_i - k_j \alpha_j| \tag{23}$$

is simultaneously smaller than the upper bound for all items in the tuple:

$$\tilde{\epsilon}_{ij}(\ell) < \frac{\alpha_i + \alpha_j}{\ell^{1/(M-1)}} \quad (\forall i, j = 0, \dots, M - 1, i \neq j). \tag{24}$$

Note, that  $\tilde{\epsilon}_{ij}$  differs from  $\epsilon$  defined for two modules (Eq 2) as it is not normalised with  $\alpha$ .

*Proof of Eq 24.* First we prove that any vector of irrationals can be approximated to the claimed order with rationals having the same denominator. Let  $\mathcal{A} = (\alpha_1, \dots, \alpha_{n-1})$ . To approximate  $\mathcal{A}$  with rationals of denominator at most  $Q$  let us define the vectors  $\mathbf{a}_j = j\mathcal{A} - [j\mathcal{A}]$ ,  $j = 0, \dots, Q$ , where floor is understood coordinate-wise. Let us partition the unit cube  $[0, 1]^{n-1}$  into small cubes of side length  $Q^{-1/(n-1)}$ , so that altogether we have  $Q$  of them. Since we have  $Q + 1$  many  $\mathbf{a}_j$ -s each falling into  $[0, 1]^{n-1}$ , hence there will be (at least) 2 of them falling into the same small cube,  $\mathbf{a}_k$  and  $\mathbf{a}_l$ , say. Then

$$\|k - l\mathcal{A} - [k\mathcal{A}] - [l\mathcal{A}]\| \leq |\mathbf{a}_k - \mathbf{a}_l| \leq Q^{-1/(n-1)},$$

with the inequalities holding coordinate-wise. Therefore, because of  $|k - l| \leq Q$ ,  $\mathcal{A}$  is approximable with denominator  $|k - l|$  and numerator (vector)  $\|[k\mathcal{A}] - [l\mathcal{A}]\|$  with error not exceeding  $|k - l|^{-(1+1/(n-1))}$ . The desired statement follows then by simultaneously approximating the

numbers  $1/\alpha_i$  with common denominator, which is also a simultaneous approximation of  $\alpha_i$  with common numerator, which completes the proof.

For a set of grid scales  $\alpha_i = \alpha^i$  ( $i = 0, \dots, M - 1$ ) where  $\alpha$  is an algebraic number of degree  $M$ , there exists a maximal positive constant  $c_A$ , such that

$$\tilde{\epsilon}(\ell) = \max_{i,j} \left\{ \frac{1}{\alpha_i + \alpha_j} \tilde{\epsilon}_{ij}(\ell) \right\} > \frac{c_A}{\ell^{1/(M-1)}} \tag{25}$$

holds, except for at most finitely many integers  $\ell$ .

To see that Eq 25 holds, we start from the work of [44] (see also [45]) stating that powers of an algebraic number are badly simultaneously approximable with common denominator in the following sense. Let  $\beta$  be an algebraic number of order  $M$ . There exists  $c_\beta > 0$  such that for all integer  $\ell$ ,  $k_i$  there is  $i \in \{1, \dots, M - 1\}$  for which

$$|\beta^i \ell - k_i| > \frac{c_\beta}{\ell^{1/(M-1)}}.$$

*Derivation of Eq 25.* Our goal is to give a lower bound on  $|\alpha^i k_i - \alpha^j k_j|$ , where  $\alpha$  is algebraic of order  $M$ ,  $0 \leq i, j \leq M - 1$ . Without loss of generality suppose that  $i < j$ .

$$|\alpha^i k_i - \alpha^j k_j| = |\alpha^{i-j} k_i - k_j| \alpha^j > \alpha^j \frac{c_A}{k_i^{1/(M-1)}}.$$

Now the fact that  $k_i \sim \ell/\alpha^i$  implies Eq 25 if  $c_A > 0$  is chosen appropriately.

The position representation is unambiguous if there is at least one pair of modules for which the phase difference is larger than the threshold set by the noise, i.e.,  $\tilde{\epsilon}_{ij}(\ell) > \delta(\alpha_i + \alpha_j)$  which holds if

$$\delta < \frac{c_A}{\ell^{1/(M-1)}} \tag{26}$$

From here, the critical distance  $L_{\max}$  up to which coding is unambiguous can be expressed as (cf. Eq 9):

$$L_{\max} := \left( \frac{c_A}{\delta} \right)^{M-1}, \tag{27}$$

for all  $\delta$  which is small enough.

To directly compare the capacity of the MA grid cell system derived in Eq 9 with previous estimates for nested coding [4, 6], we also calculate  $N_{\max}$ , the number of distinguishable spatial phases:

$$N_{\max} := \frac{L_{\max}}{2\delta} = \frac{1}{2c_A} \left( \frac{c_A}{\delta} \right)^M \tag{28}$$

Efficient coding with nested modules requires that  $\alpha_i = r^i$  with  $0 \leq i \leq M - 1$  and  $r$  being the scale ratio with fixed relative uncertainty of modules  $2\delta = 1/r$  [6]. The position of the animal can be determined at precision  $1/r$  without ambiguity if the animal is restricted to move in an environment with the size identical to the scale of the largest module,  $r^{M-1}$ . In this case the number of distinguishable spatial phases is  $r^M = \left(\frac{1/2}{\delta}\right)^M$ , which is identical to the capacity we found for non-nested coding when  $c_A = 0.5$  (Eq 28).

### Coding is unambiguous up to exponential distance in the number of modules

To derive Eq 27 we first show that interference of the grid representation is equivalent to pairwise interference between all pairs of modules. To test unambiguity of coding note that the place at distance  $x$  from the origin is confusable with 0 if for all  $i = 0, \dots, M - 1$  there exists an integer  $k_i$  such that

$$|k_i \alpha_i - x| < \alpha_i \delta, \tag{29}$$

where  $\delta$  is the relative uncertainty of modules. It turns out that, as for  $M = 2$ , there is no need to consider all  $x \in [0, L_{\max}]$ , it is enough to care with integers:

**Claim.** *There exists  $x \in [0, L_{\max}]$  for which Eq 29 holds for all  $i$  exactly when the following pairwise interference occurs between all modules:*

$$|k_i \alpha_i - k_j \alpha_j| < (\alpha_i + \alpha_j) \delta \tag{30}$$

for all  $i, j$  with some integers  $k_i$  ( $i = 0, \dots, M - 1$ ) such that  $0 < k_i \alpha_i \leq L_{\max}$ .

*Proof.* Let us fix  $k_i$ ,  $i = 0, \dots, M - 1$ . Pairwise interference means that there is a point  $x_{i,j}$  in the intersection of  $(k_i \alpha_i - \alpha_i \delta, k_i \alpha_i + \alpha_i \delta) = (a_i, b_i)$  and  $(k_j \alpha_j - \alpha_j \delta, k_j \alpha_j + \alpha_j \delta) = (a_j, b_j)$ . Due to the topology of the line, it is easy to see by induction that the intersection of all such intervals is nonempty and hence one can choose  $x_{i,j} = x$ . The statement is obvious for  $M = 2$ . Now suppose that the intersection  $\cap_{i=0}^n (a_i, b_i) \neq \emptyset$ . Then it is the interval  $(a, b)$  with

$$a = \max_{i=0, \dots, n} a_i \quad \text{and} \quad b = \min_{i=0, \dots, n} b_i.$$

If  $(a_{n+1}, b_{n+1})$  intersects  $(a_n, b_n)$ , then both  $a_{n+1} < b_n$  and  $b_{n+1} > a_n$ , and therefore  $a_{n+1} < b$  and  $b_{n+1} > a$ , which completes the induction. Therefore Eq 29 implies Eq 30. The other direction is immediate.

Now using the above Claim Equation Eq 27 easily follows by rearranging Eq 25.

### Asymptotic capacity of the random grid cell system

Let us fix the relative uncertainty of modules  $\delta < 1/2$  and a number  $\alpha_{\max} > (1 + \delta)/(1 - \delta)$ . We show that if scales  $\alpha_1, \alpha_2, \dots$  are drawn uniformly at random from  $[1, \alpha_{\max}]$ , independently of each other, then for any  $\delta < \zeta < 1/2$  the representation with  $M$  modules having scales  $\alpha_1, \alpha_2, \dots, \alpha_M$  is unambiguous in every spatial position  $x > 0$  up to

$$X_{\max} := \left(\frac{\zeta}{\delta}\right)^{M-1} \tag{31}$$

with probability of order  $1 - (2\zeta)^M$  as  $M \rightarrow \infty$ .

Here  $\zeta$  is the analog of  $c_A$  which characterises the capacity of a particular grid cell system. As we will see, the convergence holds for any  $\zeta < 1/2$ , but the speed of the convergence depends on  $\zeta$ : higher efficiency is guaranteed to be achieved only for larger number of modules.

*Proof:* Let  $\alpha_1, \alpha_2, \dots$  be independent random variables distributed uniformly on  $[1, \alpha_{\max}]$ . Let  $x$  be a spatial point and let  $\tilde{\psi}_i = \tilde{\psi}_i(x, \alpha_i)$  denote the phase of module  $i$  (with scale  $\alpha_i$ ) at  $x$ , that is

$$\tilde{\psi}_i := (x \bmod \alpha_i) / \alpha_i = x / \alpha_i \bmod 1 = x / \alpha_i - \left\lfloor \frac{x}{\alpha_i} \right\rfloor \in [0, 1].$$

Note that for fixed  $x$  the distribution of phases  $\tilde{\psi}_i$  are independent of each other since the  $\alpha$ -s are independent. We also use the notation  $p_1(x)$  for the probability that the phase  $\tilde{\psi}_i$  is (almost) indistinguishable from 0, defined in the following way:

$$p_1(x) = \mathbb{P}_{\alpha_i}(\tilde{\psi}_i \in [0, (1 + \varepsilon)\delta] \cup [(1 - (1 + \varepsilon)\delta), 1] \mid x).$$

where  $\varepsilon > 0$  is determined later. It is easy to see that  $p_1(x)$  does not depend on  $i$ , i.e., it is the same for all modules. Moreover, the distribution of  $\tilde{\psi}_i$  converges to uniform as the distance increases, in particular  $\lim_{x \rightarrow \infty} p_1(x) = 2(1 + \varepsilon)\delta$ . The convergence of this distribution to the uniform is a key observation that remains true even in higher dimensions with uniform random rotations or in case of slight variation of the grid scales on the long range. Hence there exists a critical distance,  $x_0 = x_0(\delta, \varepsilon)$  for which all  $x > x_0$  we have  $|p_1(x) - 2(1 + \varepsilon)\delta| \leq \delta\varepsilon$ . Therefore, for  $x > x_0$  we have

$$p_1(x) \leq (2 + 3\varepsilon)\delta. \tag{32}$$

It also implies a bound on the probability of interference of many modules at a given point  $x$ . If we consider  $M$  modules with scales drawn uniformly at random from  $[1, \alpha_{\max}]$  and independently of each other, then by Eq 32 for  $x > x_0$  the probability of all phases being close to 0 is

$$\begin{aligned} p_M(x) &= \mathbb{P}((\forall i \leq M) \tilde{\psi}_i \in (0, (1 + \varepsilon)\delta) \cup ((1 - (1 + \varepsilon)\delta), 1) \mid x) \\ &= p_1(x)^M \leq ((2 + 3\varepsilon)\delta)^M, \end{aligned} \tag{33}$$

that is,  $p_M(x)$  is exponentially small in  $M$ .

There remains to estimate the probability of interference of many modules anywhere up to a maximally allowed spatial distance. Our goal is to show that

$$\mathbb{P}((\exists x < X_{\max})(\forall i \leq M) \tilde{\psi}_i(x) \in (0, \delta) \cup ((1 - \delta), 1)) \rightarrow 0 \tag{34}$$

as  $M \rightarrow \infty$ , where  $X_{\max} = (\frac{\alpha_{\max}}{\delta})^{M-1}$ , as in Eq 31. Note, that satisfying Eq 34 is not trivial, since  $X_{\max}$  increases exponentially with  $M$ .

There is no need to investigate all  $x < X_{\max}$ , it is enough to show, that there is no interference on a set which is dense enough in  $[0, X_{\max}]$  in the stronger sense of Eq 33. Indeed, let  $Y$  be an  $\varepsilon$  dense set in  $[0, X_{\max}]$  with at most  $2X_{\max}/\varepsilon$  elements. Then

$$\begin{aligned} &\{(\exists x < X_{\max})(\forall i \leq M) \tilde{\psi}_i(x) \in (0, \delta) \cup ((1 - \delta), 1)\} \\ &\Rightarrow \{(\exists x \in Y)(\forall i \leq M) \tilde{\psi}_i(x) \in (0, (1 + \varepsilon)\delta) \cup ((1 - (1 + \varepsilon)\delta), 1)\} \end{aligned}$$

where we used the fact that the  $\frac{d}{dx} \tilde{\psi}_i = \frac{1}{\alpha_i} < 1$  since  $\alpha_i$  was chosen from the interval  $[1, \alpha_{\max}]$ . The corresponding inequality for the probabilities of these events is

$$\begin{aligned} &\mathbb{P}((\exists x < X_{\max})(\forall i \leq M) \tilde{\psi}_i(x) \in (0, \delta) \cup ((1 - \delta), 1)) \\ &\leq \mathbb{P}((\exists x \in Y)(\forall i \leq M) \tilde{\psi}_i(x) \in (0, (1 + \varepsilon)\delta) \cup ((1 - (1 + \varepsilon)\delta), 1)). \end{aligned}$$

Now for these finitely many points  $x \in Y$  we can use Eq 33 one by one, if  $x > x_0$ :

$$\begin{aligned} \mathbb{P}((\exists x_0 < x \in Y)(\forall i \leq M) \tilde{\psi}_i(x) \in (0, (1 + \epsilon)\delta) \cup ((1 - (1 + \epsilon)\delta), 1)) \\ \leq ((2 + 3\epsilon)\delta)^M 2X_{\max}/\epsilon \\ = \frac{2}{\epsilon} ((2 + 3\epsilon)\delta)^M \left(\frac{\zeta}{\delta}\right)^{M-1} \\ < \frac{1}{\epsilon\zeta} (\zeta(2 + 3\epsilon))^M \rightarrow 0 \end{aligned}$$

if  $\epsilon < \frac{1-2\zeta}{3\zeta}$ , which we assume, where in the first inequality we used Eq 33 and union bound, and then in the second one that  $X_{\max} = \left(\frac{\zeta}{\delta}\right)^{M-1}$ . We have to remark that interference in different spatial points is not independent of each other, but union bound works even in that case.

There remains to show that the grid cell representation works up to  $x_0$ . Clearly there is no ambiguity up to  $x = 1 + \delta$ . To estimate the probability

$$\mathbb{P}((\exists x_0 \geq x \in Y)(\forall i \leq M) \tilde{\psi}_i(x) \in (0, (1 + \epsilon)\delta) \cup ((1 - (1 + \epsilon)\delta), 1)) \tag{35}$$

we first have to observe that the cardinality of  $Y \cap [1 + \delta, x_0]$  is independent of  $M$ . Therefore to guarantee that the probability in Eq 35 goes to 0 we need to show that for all  $1 + \delta \leq x \leq x_0$  there is a scale  $\alpha \in [1, A]$  which is able to distinguish  $x$  from the origin, that is  $\alpha$  such that

$$\tilde{\psi}(\alpha) = (x/\alpha \bmod 1) \in [\delta, 1 - \delta].$$

This is so because  $x/\alpha$  is monotonically decreasing in  $\alpha$  and because

$$x/1 - x/A \geq x(2\delta/(1 + \delta)) \geq 2\delta,$$

where we used that  $\alpha_{\max} > (1 + \delta)/(1 - \delta)$  and  $x > 1 + \delta$ . Therefore  $\tilde{\psi}(\alpha)$  can not lay in  $[0, \delta] \cup [1 - \delta, 1]$  for all  $\alpha \in [1, \alpha_{\max}]$ .

### Numerical estimation of the $c_{\mathcal{A}}$ with $M$ modules

The constant  $c_{\mathcal{A}}$  (and  $c_{\alpha}$ ) is well defined only for algebraic numbers, but can also be estimated for real numbers from the scaling of the phase difference with distance using numerical simulations. As  $c_{\mathcal{A}}$  is defined asymptotically (Eq 25), in order to estimate it numerically we need an approximation of it for finite distances. An alternative definition of  $c_{\mathcal{A}}$  (equivalent with Eq 25) is

$$c_{\mathcal{A}} = \liminf_{\ell \rightarrow \infty} \hat{c}_{\mathcal{A}}(\ell), \tag{36}$$

where  $\hat{c}_{\mathcal{A}}(\ell)$  is defined by

$$\hat{c}_{\mathcal{A}}(\ell) = \min_{\mathcal{K}, k_0 = \ell} \max_{i,j} \{|\alpha_i k_i - \alpha_j k_j| / (\alpha_i + \alpha_j)\} \ell^{1/(M-1)}, \tag{37}$$

where  $\mathcal{K} = (k_1, \dots, k_{M-1})$ . Intuitively, to find the magnitude of interference at location  $\ell$ , for all possible values of  $\mathcal{K}$  we first select the maximum phase difference in the set and then choose the set with the smallest maximum. From the plots Figs 4 and 7 it is clear that the naive way of approximating  $c_{\mathcal{A}}$  with  $c_{\mathcal{A}}(\ell)$  for some large  $\ell$  is not a good idea, as  $c_{\mathcal{A}}(\ell)$  may vary heavily with  $\ell$ , especially for non-algebraic scale ratios. Note, that the calculation of  $\hat{c}_{\mathcal{A}}$  is a special case of  $\hat{c}_{\mathcal{A}}$  with  $M = 2$ .

To estimate coding efficiency in the presence of noise we are mostly interested in the above infimum when  $\ell$  is such that the phase difference  $\hat{c}_{\mathcal{A}}(\ell)/\ell^{1/(M-1)}$  is close to the precision  $\delta$  of

the modules. It motivates to investigate the (numerically computable) minimum

$$\hat{c}_{\mathcal{A}}(\delta_1, \delta_2) = \min \{ \hat{c}_{\mathcal{A}}(\ell) \mid \ell_2(\delta_2) \leq \ell \leq \ell_1(\delta_1) \}$$

for some pair  $\delta_1 < \delta_2$ , where  $\ell_2$  is so that for all  $\ell \geq \ell_2$  we have  $\hat{c}_{\mathcal{A}}(\ell)/\ell^{1/(M-1)} < \delta_2$  and  $\ell_1$  is the smallest  $\ell$  so that  $\hat{c}_{\mathcal{A}}(\ell)/\ell^{1/(M-1)} < \delta_1$ .

### Tools for the numerical investigation of Diophantine approximation

A common and natural way to numerically investigate Diophantine approximation is using lattice reduction [46]. By lattice we mean a subset  $\mathbb{L}$  of  $\mathbb{R}^d$  defined by some vectors  $v_1, \dots, v_m \in \mathbb{R}^d, m \leq d$  so that

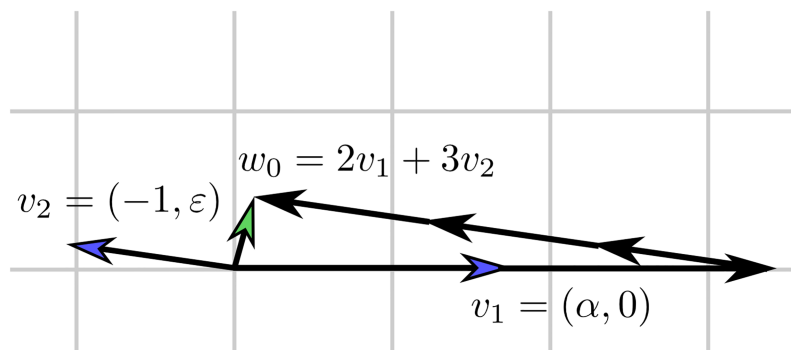
$$\mathbb{L} = \left\{ w = \sum_{i=1}^m b_i v_i \mid b_i \in \mathbb{Z} \right\}.$$

Given a lattice  $\mathbb{L}$ , a classical computational problem is to find the shortest non-zero vector of it (Fig 8). In the followings we show how Diophantine approximation of a vector  $(\alpha_1, \dots, \alpha_n)$  can be investigated with the help of finding shortest vectors of appropriately chosen lattices.

Let us first consider a simple example. Let the lattice  $\mathbb{L}$  be defined by the rows of the matrix

$$V = \begin{bmatrix} v_1 \\ \vdots \\ v_{n+1} \end{bmatrix} := \begin{bmatrix} \alpha_1 & 0 & \dots & 0 & 0 \\ 0 & \alpha_2 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & \alpha_n & 0 \\ -1 & -1 & \dots & -1 & \varepsilon \end{bmatrix},$$

where  $\varepsilon > 0$ . For all  $\varepsilon$  which is small enough the shortest vector  $w_0 = \sum_{i=1}^{n+1} b_i v_i$  of  $\mathbb{L}$  corresponds to a simultaneous Diophantine approximation of  $(\alpha_1, \dots, \alpha_n)$  with the common numerator  $b_{n+1}$  and denominators  $b_i, i = 1, \dots, n$ . The parameter  $\varepsilon$  can be considered as a penalty term: the smaller this term the bigger the numerator can be.



**Fig 8.** Which element of the lattice generated by the above two blue headed vectors is closest to the origin? Or in other words, what is the shortest nonzero vector which can be obtained as an integer coefficient linear combination of the above vectors?

<https://doi.org/10.1371/journal.pcbi.1005922.g008>



When speaking about shortest vectors we need to specify the norm with respect to which vectors are compared. Here we are looking for the largest phase difference between the modules so we use supremum norm (Eq 25). The shortest vector in supremum norm of the lattice defined by  $V$  is an approximation so that

$$\max \{b_{n+1}\epsilon, \max_i \{|b_{n+1} - b_i\alpha_i|\}\}$$

is as small as possible. By this we can compute what is the maximal phase difference between the module with scale 1 and all other modules up to distance  $b_{n+1}$ .

Remember that according to Eq 25 we are searching for an approximation minimizing

$$\max \{b_{n+1}\epsilon, \max_{ij} \{|\alpha_i k_i - \alpha_j k_j| / (\alpha_i + \alpha_j)\} \ell^{1/(M-1)}\}. \tag{38}$$

Similarly to the previous example, it can be done simply by dividing columns  $i, i = 1, \dots, n$  of  $V$  by  $(1 + \alpha_i)$ , and by adding some more columns of similar form which refer to interference between modules  $i$  and  $j$ . For example, for  $n = 3$  the shortest (in sup norm) element of the lattice generated by the rows of the following matrix gives an approximation minimizing Eq 38:

$$\begin{bmatrix} \frac{\alpha_1}{1 + \alpha_1} & 0 & 0 & \frac{\alpha_1/\alpha_3}{\alpha_1 + \alpha_3} & 0 & \frac{\alpha_1/\alpha_2}{\alpha_1 + \alpha_2} & 0 \\ 0 & \frac{\alpha_2}{1 + \alpha_2} & 0 & 0 & \frac{\alpha_2/\alpha_3}{\alpha_2 + \alpha_3} & \frac{-1}{\alpha_1 + \alpha_2} & 0 \\ 0 & 0 & \frac{\alpha_3}{1 + \alpha_3} & \frac{-1}{\alpha_1 + \alpha_3} & \frac{-1}{\alpha_2 + \alpha_3} & 0 & 0 \\ \frac{-1}{1 + \alpha_1} & \frac{-1}{1 + \alpha_2} & \frac{-1}{1 + \alpha_3} & 0 & 0 & 0 & \epsilon \end{bmatrix}.$$

In this way maximal interference in the grid cell system can be computed numerically as shortest vectors of some lattices in supremum norm. Finding this shortest vector is an integer linear programming (ILP) problem, which in general is an NP-hard computational problem, and can be solved by e.g. a branch and bound algorithm [47]. There are also efficient methods which find approximation solutions in polynomial time, such as the LLL algorithm due to Lenstra, Lenstra and Lovász [46].

The LLL algorithm finds not only a short vector of a lattice, but also another basis of it which consists of short and nearly orthogonal vectors in the  $L^2$  norm, a so called LLL reduced basis. The error made by the LLL algorithm is too high to precisely compute the constant terms in Eq 27, and therefore we could not rely only on this algorithm. Nevertheless, compared to the ILP solution, we could significantly speed up our computations by first applying the LLL algorithm to find an approximate solution (and a reduced lattice), and then an ILP solver on this LLL reduced basis, which could find nontrivial optimal solutions very efficiently if started from this input.

### Acknowledgments

We thank Gergő Orbán and András Telcs for useful discussions.

### Author Contributions

**Conceptualization:** Lajos Vágó, Balázs B. Ujfalussy.

**Formal analysis:** Lajos Vágó.

**Funding acquisition:** Balázs B. Ujfalussy.

**Supervision:** Balázs B. Ujfalussy.

**Visualization:** Lajos Vágó, Balázs B. Ujfalussy.

**Writing – original draft:** Lajos Vágó, Balázs B. Ujfalussy.

**Writing – review & editing:** Lajos Vágó, Balázs B. Ujfalussy.

## References

1. Sterling P, Laughlin S. Principles of Neural Design. MIT Press; 2015.
2. Burak Y, Brookings T, Fiete I. Triangular lattice neurons may implement an advanced numeral system to precisely encode rat position over large ranges; 2006.
3. Fiete IR, Burak Y, Brookings T. What Grid Cells Convey about Rat Location. *The Journal of Neuroscience*. 2008; 28(27):6858–6871. <https://doi.org/10.1523/JNEUROSCI.5684-07.2008> PMID: 18596161
4. Mathis A, Herz AVM, Stemmler M. Optimal population codes for space: grid cells outperform place cells. *Neural Comput*. 2012; 24(9):2280–317. [https://doi.org/10.1162/NECO\\_a\\_00319](https://doi.org/10.1162/NECO_a_00319) PMID: 22594833
5. Mathis A, Herz AVM, Stemmler MB. Resolution of nested neuronal representations can be exponential in the number of neurons. *Phys Rev Lett*. 2012; 109(1):018103. <https://doi.org/10.1103/PhysRevLett.109.018103> PMID: 23031134
6. Wei XX, Prentice J, Balasubramanian V. A principle of economy predicts the functional architecture of grid cells. *Elife*. 2015; 4:e08362. <https://doi.org/10.7554/eLife.08362> PMID: 26335200
7. Stemmler M, Mathis A, Herz AVM. Connecting multiple spatial scales to decode the population activity of grid cells. *Sci Adv*. 2015; 1(11):e1500816. <https://doi.org/10.1126/science.1500816> PMID: 26824061
8. Mosheiff N, Agmon H, Moriel A, Burak Y. An efficient coding theory for a dynamic trajectory predicts non-uniform allocation of entorhinal grid cells to modules. *PLoS Comput Biol*. 2017; 13(6):e1005597. <https://doi.org/10.1371/journal.pcbi.1005597> PMID: 28628647
9. Towse BW, Barry C, Bush D, Burgess N. Optimal configurations of spatial scale for grid cell firing under noise and uncertainty. *Philos Trans R Soc Lond B Biol Sci*. 2014; 369(1635):20130290. <https://doi.org/10.1098/rstb.2013.0290> PMID: 24366144
10. Hafting T, Fyhn M, Molden S, Moser MB, Moser EI. Microstructure of a spatial map in the entorhinal cortex. *Nature*. 2005; 436:801–806. <https://doi.org/10.1038/nature03721> PMID: 15965463
11. Moser EI, Roudi Y, Witter MP, Kentros C, Bonhoeffer T, Moser MB. Grid cells and cortical representation. *Nat Rev Neurosci*. 2014; 15(7):466–81. <https://doi.org/10.1038/nrn3766> PMID: 24917300
12. Barry C, Hayman R, Burgess N, Jeffery KJ. Experience-dependent rescaling of entorhinal grids. *Nature Neuroscience*. 2007; 10:682–684. <https://doi.org/10.1038/nn1905> PMID: 17486102
13. Stensola H, Stensola T, Solstad T, Frøland K, Moser MB, Moser EI. The entorhinal grid map is discretized. *Nature*. 2012; 492(7427):72–78. <https://doi.org/10.1038/nature11649> PMID: 23222610
14. Sreenivasan S, Fiete I. Grid cells generate an analog error-correcting code for singularly precise neural computation. *Nat Neurosci*. 2011; 14(10):1330–7. <https://doi.org/10.1038/nn.2901> PMID: 21909090
15. Yoon K, Buice MA, Barry C, Hayman R, Burgess N, Fiete IR. Specific evidence of low-dimensional continuous attractor dynamics in grid cells. *Nat Neurosci*. 2013; 16(8):1077–84. <https://doi.org/10.1038/nn.3450> PMID: 23852111
16. Fyhn M, Hafting T, Treves A, Moser MB, Moser EI. Hippocampal remapping and grid realignment in entorhinal cortex. *Nature*. 2007; 446:190–194. <https://doi.org/10.1038/nature05601> PMID: 17322902
17. Gardner RJ, Lu L, Wernle T, Moser MB, Moser EI. Correlation structure of grid cells is preserved during sleep. *bioRxiv*. 2017; p. <http://dx.doi.org/10.1101/198499>.
18. Trettel SG, Trimper JB, Hwaun E, Fiete IR, Colgin LL. Grid cell co-activity patterns during sleep reflect spatial overlap of grid fields during active behaviors. *bioRxiv*. 2017; p. <https://doi.org/10.1101/198671>.
19. Hurwitz A. Ueber die angenaherte Darstellung der Irrationalzahlen durch rationale Brüche (On the approximate representation of irrational numbers by rational fractions). *Mathematische Annalen*. 1891; 39 (2):279–284. <https://doi.org/10.1007/BF01206656>
20. Perron O. Die Lehre von den Kettenbrüchen [The Theory of Continued Fractions] (in German), Chapter 2. Leipzig: B. G. Teubner; 1913.
21. Oxtoby JC. Measure and Category. New York-Berlin: Springer-Verlag; 1980.

22. Broderick R, Fishman L, Kleinbock D, Reich A, Weiss B. The set of badly approximable vectors is strongly C1 incompressible. *Mathematical Proceedings of the Cambridge Philosophical Society*. 2012; 153(2):319–339. <https://doi.org/10.1017/S0305004112000242>
23. Brun VH, Solstad T, Kjelstrup KB, Fyhn M, Witter MP, Moser EI, et al. Progressive increase in grid scale from dorsal to ventral medial entorhinal cortex. *Hippocampus*. 2008; 18(12):1200–12. <https://doi.org/10.1002/hipo.20504> PMID: 19021257
24. Stensola T, Stensola H, Moser MB, Moser EI. Shearing-induced asymmetry in entorhinal grid cells. *Nature*. 2015; 518(7538):207–12. <https://doi.org/10.1038/nature14151> PMID: 25673414
25. Taylor KD. Range of Movement and Activity of Common Rats (*Rattus norvegicus*) on Agricultural Land. *Journal of Applied Ecology*. 1978; 15(3):663–677. <https://doi.org/10.2307/2402767>
26. Tsoar A, Nathan R, Bartan Y, Vyssotski A, Dell’Omo G, Ulanovsky N. Large-scale navigational map in a mammal. *Proc Natl Acad Sci U S A*. 2011; 108(37):E718–24. <https://doi.org/10.1073/pnas.1107365108> PMID: 21844350
27. Geva-Sagiv M, Las L, Yovel Y, Ulanovsky N. Spatial cognition in bats and rats: from sensory acquisition to multiscale maps and navigation. *Nat Rev Neurosci*. 2015; 16(2):94–108. <https://doi.org/10.1038/nrn3888> PMID: 25601780
28. Mathis A, Stemmler MB, Herz AV. Probable nature of higher-dimensional symmetries underlying mammalian grid-cell activity patterns. *Elife*. 2015; 4. <https://doi.org/10.7554/eLife.05979> PMID: 25910055
29. Krupic J, Bauza M, Burton S, Barry C, O’Keefe J. Grid cell symmetry is shaped by environmental geometry. *Nature*. 2015; 518(7538):232–5. <https://doi.org/10.1038/nature14153> PMID: 25673417
30. Stella F, Si B, Kropff E, Treves A. Grid cells on the ball. *Journal of Statistical Mechanics: Theory and Experiment*. 2013; 2013(03):P03013. <https://doi.org/10.1088/1742-5468/2013/03/P03013>
31. Urdapilleta E, Troiani F, Stella F, Treves A. Can rodents conceive hyperbolic spaces? *J R Soc Interface*. 2015; 12(107). <https://doi.org/10.1098/rsif.2014.1214>
32. McNaughton BL, Battaglia FP, Jensen O, Moser EI, Moser MB. Path integration and the neural basis of the “cognitive map”. *Nature Reviews Neurosci*. 2006; 7(8):663–678. <https://doi.org/10.1038/nrn1932>
33. Kropff E, Treves A. The emergence of grid cells: Intelligent design or just adaptation? *Hippocampus*. 2008; 18(12):1256–69. <https://doi.org/10.1002/hipo.20520> PMID: 19021261
34. Brown EN, Frank LM, Tang D, Quirk MC, Wilson MA. A statistical paradigm for neural spike train decoding applied to position prediction from ensemble firing patterns of rat hippocampal place cells. *J Neurosci*. 1998; 18(18):7411–25. PMID: 9736661
35. Samu D, Eros P, Ujfalussy B, Kiss T. Robust path integration in the entorhinal grid cell system with hippocampal feed-back. *Biol Cybern*. 2009; 101(1):19–34. <https://doi.org/10.1007/s00422-009-0311-z> PMID: 19381679
36. Burak Y, Fiete IR. Fundamental limits on persistent activity in networks of noisy neurons. *Proc Natl Acad Sci U S A*. 2012; 109(43):17645–50. <https://doi.org/10.1073/pnas.1117386109> PMID: 23047704
37. Issa JB, Zhang K. Universal conditions for exact path integration in neural systems. *Proc Natl Acad Sci U S A*. 2012; 109(17):6716–20. <https://doi.org/10.1073/pnas.1119880109> PMID: 22493275
38. Burak Y, Fiete IR. Accurate path integration in continuous attractor network models of grid cells. *PLoS Comput Biol*. 2009; 5(2):e1000291. <https://doi.org/10.1371/journal.pcbi.1000291> PMID: 19229307
39. Giocomo LM, Hussaini SA, Zheng F, Kandel ER, Moser MB, Moser EI. Grid cells use HCN1 channels for spatial scaling. *Cell*. 2011; 147(5):1159–70. <https://doi.org/10.1016/j.cell.2011.08.051> PMID: 22100643
40. Urdapilleta E, Si B, Treves A. Selforganization of modular activity of grid cells. *Hippocampus*. 2017; 27(11):1204–1213. <https://doi.org/10.1002/hipo.22765> PMID: 28768062
41. Moser EI, Kropff E, Moser MB. Place cells, grid cells, and the brain’s spatial representation system. *Annu Rev Neurosci*. 2008; 31:69–89. <https://doi.org/10.1146/annurev.neuro.31.061307.090723> PMID: 18284371
42. Etienne AS, Jeffery KJ. Path integration in mammals. *Hippocampus*. 2004; 14:180–192. <https://doi.org/10.1002/hipo.10173> PMID: 15098724
43. Ormond J, McNaughton BL. Place field expansion after focal MEC inactivations is consistent with loss of Fourier components and path integrator gain reduction. *Proc Natl Acad Sci U S A*. 2015; 112(13):4116–21. <https://doi.org/10.1073/pnas.1421963112> PMID: 25733884
44. Drmota M, Tichy RF. Sequences, Discrepancies and Applications. *Lecture Notes in Math*, Springer Verlag, Berlin. 1997; no. 1651.

45. Cassels J. Simultaneous Diophantine Approximation II. *Proc London Math Soc.* 1955;(3), 5:435–448. <https://doi.org/10.1112/plms/s3-5.4.435>
46. Lenstra AK, Lenstra HW, Lovász L. Factoring polynomials with rational coefficients. *Mathematische Annalen.* 1982; 261(4):515–534. <https://doi.org/10.1007/BF01457454>
47. Land AH, Doig AG. An Automatic Method of Solving Discrete Programming Problems. *Econometrica.* 1960; 28(3):497–520. <https://doi.org/10.2307/1910129>