Neural Coding as a Statistical Testing Problem

Guilherme Ost*   Patricia Reynaud-Bouret†

Abstract

We take the testing perspective to understand what the minimal discrimination time between two stimuli is for different types of rate coding neurons. Our main goal is to describe the testing abilities of two different encoding systems: place cells and grid cells. In particular, we show, through the notion of adaptation, that a fixed place cell system can have a minimum discrimination time that decreases when the stimuli are further away. This could be a considerable advantage for the place cell system that could complement the grid cell system, which is able to discriminate stimuli that are much closer than place cells.

1 Introduction

Grid cells are particular neurons in medial enthorinal cortex [16] that exhibit a periodic spatial firing pattern. Whereas place cells in the hippocampus fire at a given location and appear to be associated with an allocentric representation, grid cells fire at each node of a hexagonal lattice and appear to be involved in a self-localization representation [17]. The grid cell system not only encodes spatial position but also direction and velocity [20] or sounds [1] and could even be used by mammals to encode episodic memories [5]. Organized in modules (one module being dedicated to one lattice scale), the grid cells in each module have a uniform distribution, while the progression of scales between modules appears to be geometric [24].

Since their discovery in 2005 [11], which earned O'Keefe and the Mosers the Nobel Prize, grid cells have been intensively studied from a theoretical point of view in relation to place cells [19]. Some authors (see for example [21]) are interested in how neural networks can generate such patterns, while others try to explain the hexagonal lattice [6] or the exact geometric progression of scales [27, 23].

In particular, one of the main focuses has been the encoding capacity of the system. More specifically, the authors focused on a statistical capacity measure, namely the Fisher information, because of its link with estimation. Indeed in Statistics, Cramér-Rao bound states that the L_2-error (i.e., $\sqrt{\mathbb{E}(|\hat{s}_n - s|^2)}$) of an unbiased estimator $\hat{s}_n$ (i.e. , $\mathbb{E}(\hat{s}_n) = s$)) of a given quantity s is lower bounded by $I_n(s)^{-1/2}$, where $I_n(s)$ is the Fisher information. Moreover, this lower bound is generally achieved by Maximum

---

*Universidade Federal do Rio de Janeiro. E-mail: guilhermeost@im.ufrj.br
†Université Côte d'Azur, CNRS, LJAD, France. E-mail: Patricia.Reynaud-Bouret@univ-cotedazur.fr

Likelihood Estimators (MLE), at least asymptotically in a context of n i.i.d. observations [7]. Informally, the Cramér-Rao bound is interpreted as follows: there is a "best" estimator (which would be the MLE) that would achieve the smallest error $I_n(s)^{-1/2}$. So if a code refers to a model that describes the precise influence of s on the spike trains emitted by the neural cells, Cramèr-Rao bounds paves the path of finding the best code as the code that maximizes the Fisher information.

Initially, [4] worked on the relationship between mutual information and Fisher information for place cells and other types of neurons with receptive fields. They showed in particular that if s represents a position, then $I_n(s)$ grows linearly with n, for a particular code where the n place cells responses to position s are i.i.d. This leads to an error in $n^{-1/2}$. Then Fiete and her co-authors [9, 22] showed that for a given number n of neurons, grid cells can encode many more positions and that $I_n(s)$ grows exponentially with n the number of place cells. In particular, this means that the accuracy of the position estimator that can be done with n grid cells is much more precise than the accuracy that can be done with n place cells (see also these related works [26, 15, 18]).

In the present work, we take a different statistical viewpoint from that of unbiased estimation using Fisher information: we take a testing approach with a minimax viewpoint. Our first argument is that estimation is very complex in fact. For instance, when using Cramér-Rao bound, one has to be aware that it applies only to unbiased estimators and the Stein phenomenon shows that biased estimator might sometimes be faster that $I_n(s)^{-1/2}$ [25]. Minimax theory [25] can help to shed more lights on the right order of magnitude of the error by computing the minimal value of $\max_s \sqrt{\mathbb{E}(|\hat{s}_n - s|^2)}$. Our second argument is that pointwise testing (e.g. testing a point like "$s = 0$" for instance) is easier than estimation: imagine an estimator $\hat{s}_n$ of s whose variance depends on s itself. When testing $s = 0$, we know the variance of $\hat{s}_n$ under this hypothesis and we can therefore use this to describe a rejection region for the test. Going further, the fact that tests are easier to build than estimation, can lead to surprising theoretical effects: depending on the regularity of the s to estimate, it has been shown that typically one makes a minimax estimation error of say Δ_n, with n the number of observations, whereas there exists some test that can detect that $s \neq 0$ as soon as the distance between s and 0 is larger than ρ_n, with ρ_n that is negligible with respect to Δ_n [12, 2]. The difference is more than a mere multiplicative constant: testing rates can be faster than estimation rates. Therefore we want to adopt this testing point of view to see if it can improve our understanding of the place cells/grid cells code.

From this testing point of view, we think that a good encoding system should be able to discriminate quickly between two stimuli or positions s_1 and s_2, as soon as they are sufficiently apart. In particular, the testing procedure can take into account at which distance s_1 is from s_2 (information that cannot be taken into account, at least as explicitly, in an estimation procedure) and one can have a discrimination time between two points that depends on this distance.

Therefore the purpose of the present work is to study the following three theoretical problems. To make things more concrete, imagine a rat in a maze who should learn a certain behavior in position s_1 and another one in position s_2. We give for each problem the "rat" interpretation with respect to this situation.

> **Definition 1.1.** *Given n neurons obeying a certain stochastic model parametrized by a code $f(s)$ in response to a stimulus/position $s \in \mathbb{S}$ presented to the system for a duration T, we define the minimal discrimination time between two locations s_1 and s_2 for the code f, with precision α (denoted $T_{min}(f, s_1, s_2, \alpha)$), as the minimal time the output of the n neurons must be observed in order to distinguish s_1 from s_2 with a probability of error less than $\alpha \in (0, 1)$.*

Problem 1 consists in understanding the behavior of the minimal discrimination time $T_{min}(f, s_1, s_2, \alpha)$. If this minimal time is infinite, it expresses in particular the fact that the coding system cannot distinguish s_1 from s_2. More precisely, this quantity also expresses how the minimal discrimination time decreases when the distance between s_1 and s_2 increases. Note in particular that the test which is able to achieve this minimal time can depend on the precise knowledge of s_1 and s_2 because we define this as a test and this will be the case in our solution. From the "rat" perspective, the discrimination time between the two positions, is a good lower bound for the reaction time of the rat to this given task, because this is only after realising that it is in position s_1 or s_2 that the rat can proceed to the learned associated behavior.

Definition 1.2. *We equip* $\mathbb{S}$ *with a certain metric* d. *We define the minimax discrimination time of a family of codes* $\mathcal{F}$ *at distance* ρ *by*

$$T(\mathcal{F}, \rho, \alpha) = \inf_{f \in \mathcal{F}} \sup_{s_1, s_2 \in \mathbb{S}:\ d(s_1, s_2) \geq \rho} T_{min}(f, s_1, s_2, \alpha). \qquad (1.1)$$

This quantity can be seen as the rate at which the best code f *in a certain family* $\mathcal{F}$ *(for instance place cells or grid cells) can discriminate all stimuli at distance* ρ *or more.*

Problem 2 consists in computing upper and lower bounds for this minimax discrimination time. In particular, this rate is a function of n and ρ and it is not clear whether the best code f depends on ρ or not. From the "rat" perspective, it implies that the important parameter for the distinction between s_1 and s_2 is the distance: if the brain "uses" the best code in the family $\mathcal{F}$ at distance at least ρ", then as soon as $d(s_1, s_2) \geq \rho$, one can guarantee a reaction time that is at least $T(\mathcal{F}, \rho, \alpha)$. From a modeling point of view, this raises a good question: why would the brain "use the best code in the family $\mathcal{F}$ at distance at least ρ" ?. The important part in the previous sentence is in particular: why would the brain focus on one particular ρ? This leads to the third problem: adaptation.

Definition 1.3. *A code* $f \in \mathcal{F}$ *is said to be adaptive if it achieves the rate defined in* (1.1), *up to multiplicative constants, in a given range of values for* ρ, *that is,*

$$\forall \rho \text{ in a given range,} \quad \sup_{s_1, s_2 \in \mathbb{S}:\ d(s_1, s_2) \geq \rho} T_{min}(f, s_1, s_2, \alpha) \simeq T(\mathcal{F}, \rho, \alpha). \qquad (1.2)$$

Problem 3 consists in finding an adaptive code and the corresponding range of ρ for different family of codes. Here the word *adaptation* is meant in the sense of theoretical statistics/minimax theory [12, 2, 25]. From the "rat" perspective, adaptation (in the previous statistical sense) is fundamental. We can indeed pinpoint two scenarios about the learning. *In Scenario 1*, before even learning, the system (place cells or grid cells) can achieve for many ρ the best discrimination time (adaptive code). The only thing that the rat has to learn is the specific response of the cells to position s_1 and s_2 to perform the "best" discrimination test and decide what is the correct behavior. This scenario has the advantage of minimal training time: if a new couple (s_1, s_2) is presented, the learning time should be very quick. In *Scenario 2*, the system is not adaptive and then each time a new couple (s_1, s_2) at a different distance is proposed, the rat either is stuck with a suboptimal code leading to a suboptimal discrimation time, or it has to learn a new representation/new code at the same time as the new couple to react faster. Scenario 2 is of course less "adaptive" because the rat would take a longer time to react/learn.

Solving these three theoretical problems should help us to distinguish between several types of codes and in particular between place cells and grid cells. The minimal

discrimination time computed in Problem 1 is an idealized reference of a certain reaction time, which depends on the coding system. It is already slightly more complex than simple estimation accuracy since it encompasses the idea that if the stimuli are very different, the reaction time should be faster. Second, the minimax time in Problem 2 should in particular tell us which system seems to be more competitive than the other. Finally, the adaptation viewpoint in Problem 3 should give a more specific viewpoint on the practical use of each system to see if it is possible to discriminate stimuli at different scales with the same code.

To push the mathematics as far as possible, we use a very simple stochastic model: the spike trains of the n neurons are homogeneous Poisson processes (i.e. constant firing rates in time, that only depends on the position/stimulus s) with coding performed only via their respective firing rates. We also idealize the rate code f into simple step functions with only two values and used the circle as the stimulus/position set (which is consistent with, for example, a 1D circular maze or with the direction of movement [10]).

We have been able to completely answer all three problems for the place cells code, for which adaptation is possible. All three problems are also solved for grid cell codes when the number of cells per module and scales are fixed, but minimax rate and adaptation for general grid cells code, where even the number of modules is let free is still an open problem. In particular, we have shown that the minimal distance ρ that can be detected by n place cells is of order $1/n$, a distance that is much smaller than the rate $n^{-1/2}$ obtained via the classical use of Fisher information (see [4]). It also appears that grid cells have much better resolution than place cells, up to 2^{-n}, which is consistent with the bounds given by [22]. Grid cells may also be faster for discrimination than place cells, for a fixed ρ, if the system is well calibrated. However, it does not seem clear that there is an adaptive code for general grid cells, and in this sense, place cells might have an advantage in terms of reaction time for sufficiently distant stimuli/positions.

In Section 2, we give the stochastic model, the main notation and compute the minimal discrimination time (Problem 1) in a very general setting, as well as a lower bound in 2^{-n} on the smallest distance ρ that can be discriminated whatever the code. In Section 3, we study more deeply the place cells code, compute minimax rates (Problem 2) in $\lfloor n\rho \rfloor^{-1}$ and prove that even random codes are adaptive in this setting (Problem 3). In Section 4, we study grid cells code and prove that it can reach the resolution 2^{-n} and that another grid code can also achieve the rate $\lfloor n/\log_2(1/\rho) \rfloor^{-1}$. Numerical illustration is provided in Section 5. Conclusion, Discussion and Perspectives are given in Section 6. Auxiliary results are postponed to Section 7.

2 Discrimination time

2.1 Model and notation

Since we are interested in grid cells that have a periodic feature, it is simpler to represent stimuli as a circle than as an interval.

We consider stimulus/position s that belong to $\mathbb{S}^1 = [0, 1)$, that is considered in a periodic way, i.e. $0 \equiv 1$. Equivalently, we can represent this set of stimuli/position as a circle.

The stimulus is encoded by n neurons, which emit spikes as *independent homogeneous Poisson processes*. By homogeneous Poisson process, we mean that if the stimulus s remains constant through time, then the firing rate of the Poisson process is also constant through time. More specifically, for a given stimulus s in $\mathbb{S}^1$, each neuron i has a firing rate $f_i(s)$, which only depends on s, as long as the stimulus s is presented.

In the literature, to model cells with receptor fields, these functions $f_i(s)$ are most of the time centered around a certain favorite stimulus for which the value is very large,
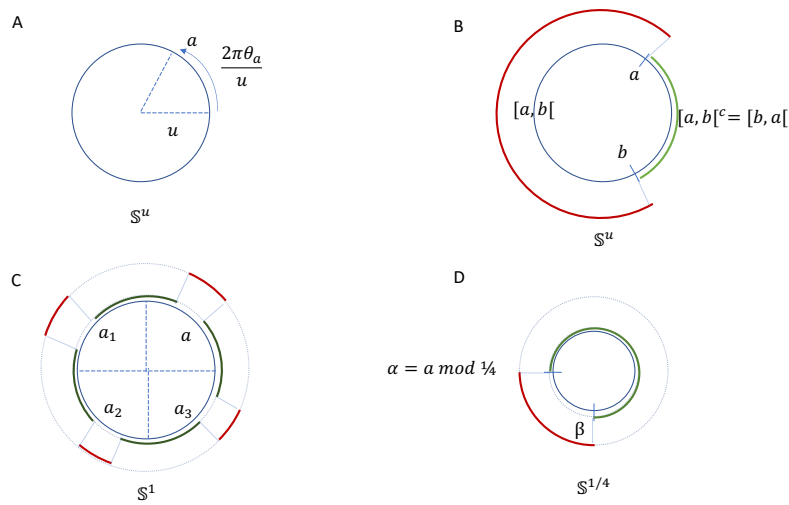
Figure 1: Visual representation of the main notions. In A, $\mathbb{S}^u$, the circle of radius u with a point a and its corresponding argument $\theta_a \in [0, u)$ In B, the visual representation of the intervals. In red and green, the visual representation of the function $s \mapsto g(s) = \mu 1_{[\![a,b[\![}(s) + 1_{[\![a,b[\![^c}(s)$, with the value μ in red and 1 in green. In C and D, visual representation of the action of mod. In these pictures, $\alpha = a \bmod 1/4 = a_1 \bmod 1/4 = a_2 \bmod 1/4 = a_3 \bmod 1/4$. Also in D, the representation of the function $s \mapsto g_{1/4}(s) = \mu 1_{[\![\alpha,\beta[\![}(s) + 1_{[\![\alpha,\beta[\![^c}(s)$ with the same color code as in B. In C, the representation of the periodic function $s \mapsto g(s) = g_{1/4}(s \bmod u) = \mu 1_{[\![\alpha,\beta[\![}(s \bmod 1/4) + 1_{[\![\alpha,\beta[\![^c}(s \bmod 1/4)$

whereas it returns to some very small rate when it is far from this favorite stimulus.

To simplify the mathematical computations, we decide to use piecewise constant functions to model f_i. More precisely, these functions only take two values μ and 1 with μ much larger than 1. We say that neuron i responds to stimulus s or is activated by s when $f_i(s) = \mu$. We denote the code $f = (f_1, ..., f_n)$, the vector of the firing rates and I_s^f is the set of neurons responding to s that is

$$I_s^f = \{i \in [n] \ : \ f_i(s) = \mu\},$$

with $[n]$ the short notation for $\{1, ..., n\}$. In what follows, for any subset $B \subseteq [n]$, we denote $|B|$ the cardinality of the set B.

To define more precise the piecewise constant functions f_i, especially for the grid cells code, we need to introduce further mathematical notation. Let $\mathbb{S}^u$ be the circle of radius u, which can be put in one-to-one map with the segment $[0, u)$, through the argument $a \mapsto \theta_a$ as seen in Figure 1.A. More mathematically, for a given $a \in \mathbb{S}^u$, θ_a is the only value in $[0, u)$ such that a corresponds to the point $u(\cos(2\pi\theta_a/u), \sin(2\pi\theta_a/u))$. The distance we use on $\mathbb{S}^u$ is the geodesic distance on the circle divided by 2π. This can also be viewed as $d_u(a, b) = \min(|\theta_a - \theta_b|, (u - |\theta_a - \theta_b|))$. At most, it is $u/2$. When $u = 1$, we write d instead of d_1 for short. Observe that, in $\mathbb{S}^1$, the largest distance one can have is $1/2$. The interval $[\![a, b[\![$ is defined as the set of all points s such that $\theta_s \in [\theta_a, \theta_b)$ if $\theta_a < \theta_b$, and the set of all points the s such that $\theta_s \notin [\theta_b, \theta_a)$ if $\theta_b < \theta_a$ with the convention that $[\![a, a[\![= \emptyset$ is the empty set. Note that the complementary of $[\![a, b[\![$, satisfies $[\![a, b[\![^c = [\![b, a[\![$. See also Figure 1.B.

The first code that we are interested in, corresponds (via this piecewise constant simplification) to classical neurons having a certain receptor field, or to place cells, as mentioned in the introduction.

Place cells code $\mathcal{P}$. A code f is a place cells code if and only if for all i in $[n] = \{1, ..., n\}$, there exist a_i, b_i in $\mathbb{S}^1$ such that for all s in $\mathbb{S}^1$,

$$f_i(s) = \mu 1_{[\![a_i, b_i[\![}(s) + 1_{[\![a_i, b_i[\![^c}(s). \tag{2.1}$$

With this definition, we can assimilate the receptor field of neuron i with the interval $[\![a_i, b_i[\![$. A typical representation of this place cells code can be seen Figure 1.B

Grid cells have a periodic structure. Therefore, we need to define properly periodic functions as well. To do so, we use the modulus operation. It is defined for $x \in \mathbb{R}$ and $u \in \mathbb{R}_{>0}$ by $x \bmod u = x - \lfloor x/u \rfloor u \in [0, u)$, the remainder of the euclidean division of x by u, where $\lfloor x/u \rfloor$ is the largest integer less than or equal to x/u. Informally, a function g of period u on $\mathbb{S}^1$ is a function that only depends on the value of the stimulus $s \bmod u$, i.e. $g(s) = g(s \bmod u)$ for all $s \in \mathbb{S}^1$. Let us define module on circles more formally: for radii $u, v \in \mathbb{R}_{>0}$ and any $s \in \mathbb{S}^v$, we denote $t = s \bmod u$ the point in $\mathbb{S}^u$ such that $\theta_t^u = \theta_s^v \bmod u$. To give an intuition of what the modulus operation does, we put an example in Figure 1C and D. In this sense, we can formally define periodic functions $s \mapsto g(s)$ on $\mathbb{S}^1$ with period u by saying that $g(s) = g_u(s \bmod u)$, for some function g_u on $\mathbb{S}^u$.

However, if we cannot divide $\mathbb{S}^1$ in an integer number of intervals of length u, the periodic functions are not completely well defined. This remark leads to the restriction on the λ_i's in the following definition of the grid cells code.

Grid cells code $\mathcal{G}((n_i, \lambda_i)_{i=1,...,m})$. [24] showed experimentally that grid cells are grouped by modules. A given module is dedicated to a certain scale (or spatial periodicity of the firing pattern). Once the scale is fixed the exact localisation of the centers of the

grid does not seem to show any particular structure. But the progression of the scales seems to be done in quantized manner. We model this as follows. A code f is a grid cells code with m modules of cardinals $(n_i)_{i=1,\ldots,m}$ and scales $(\lambda_i)_{i=1,\ldots,m}$ if and only if the n neurons are grouped into m modules $M_1, \ldots, M_m$ so that there is $n_i := |M_i|$ cells per module i with $\sum_{i=1}^m n_i = n$ and all these n_i cells have a periodic code of period λ_i. More precisely, for a given neuron $j \in M_i$, there exists $a_{i,j}$ and $b_{i,j}$ in $\mathbb{S}^{\lambda_i}$, the circle of radius λ_i, such that for all $s \in \mathbb{S}^1$

$$f_{i,j}(s) = \mu 1_{[\![a_{i,j}, b_{i,j}[\![}(s \bmod \lambda_i) + 1_{[\![a_{i,j}, b_{i,j}[\![^c}(s \bmod \lambda_i) \qquad (2.2)$$

To ensure coherence of the respective periods in each module, we assume that $1 := \lambda_1 > \lambda_2 > \ldots > \lambda_m$ are real positive numbers satisfying the following relations: $\lambda_i^{-1} \in \mathbb{N}_{>0}$ and $\lambda_i \lambda_{i+1}^{-1} \in \mathbb{N}_{>0}$. In particular $\lambda_{i+1} \leq \lambda_i/2$ for all module i. A typical representation of this grid cells code can be seen Figure 1.C.

Unspecified grid cells code $\bar{\mathcal{G}}$. If we want to consider a grid cell class for which the number of modules, scales and number of cells per module are not specified, one can consider

$$\bar{\mathcal{G}} = \cup_{m=1,\ldots,n} \cup_{n_1+\ldots+n_m=n} \cup_{\lambda_1 > \ldots > \lambda_m} \mathcal{G}((n_i, \lambda_i)_{i=1,\ldots,m}),$$

where the last union is taken only on the λ_i's such that $\lambda_i \lambda_{i+1}^{-1} \in \mathbb{N}_{>0}$ for $i < m$.

Note that grid cells code with only one module are place cells code and this also justifies why it was easier to consider stimulus on the circle in the first place.

Some of the results we are going to prove also hold for more general binary codes.

General binary code A code f is a general binary code if and only if for all i in $[n] = \{1, \ldots, n\}$, $s \mapsto f_i(s)$ is a piecewise constant function on $\mathbb{S}^1$ taking only two possible values, μ and 1. Place cells code and grid cells code are just particular cases of general binary codes. In a general binary code, the set of stimulus to which neuron i responds forms a borelian of $\mathbb{S}^1$.

Notice that the choice of 1 as the smallest of the two possible rates for the code is to simplify computations, but one can always transform the data to be in this case. Indeed, the time-changing theorem [3] allows us to change time in order to fix the smallest rate at 1.

2.2 The statistical testing problem

A stimulus s is applied for a time T and it results from the stimulus s, n spike trains for the n different neurons, i.e. $N^1, \ldots, N^n$, the n independent Poisson processes on $[0, T]$. We consider that the individual (or the agent or the brain) has only these spike trains as source of information on s and that it tries to use the best statistical tool available, based on $N^1, \ldots, N^n$. This philosophy gives us an ideal bound on the performance that the brain can do with one or another encoding system. It is typically used by [4] to say that the inverse of the Fisher Information gives the smallest variance of an estimator of the stimulus (Cramer-Rao bound) and that maximizing the Fisher Information gives the best code. We want to basically adopt the same point of view but for testing instead of estimating.

In the testing problem, given s_1 and s_2 two possible values for s, the individual has to guess that $s = s_1$ or $s = s_2$ based solely on $N^1, \ldots, N^n$. Mathematically, it means that this guess is a test $\Phi = \Phi(N^1, \ldots, N^n)$ that can only take two values s_1 or s_2. The individual can make two mistakes $\mathbb{P}_{s_1}(\Phi = s_2)$ that is the probability that the guess is s_2 whereas the applied stimulus is s_1, and reciprocally $\mathbb{P}_{s_2}(\Phi = s_1)$, that is the probability that the

guess is s_1 whereas the applied stimulus is s_2. There are varieties of possible tests and we are interested only by the ones, which have (up to multiplicative constants) the smallest possible error, that is we want to find Φ such that it achieves

$$p_e(s_1, s_2) = \min_{\Phi \text{ test of } s_1 \text{ vs } s_2} p_{e,\Phi} \quad \text{with} \quad p_{e,\Phi} = \max\left[\mathbb{P}_{s_1}(\Phi = s_2), \mathbb{P}_{s_2}(\Phi = s_1)\right]. \quad (2.3)$$

Order of magnitude of p_e For $s_1, s_2 \in \mathbb{S}^1$, we denote $\Delta^f_{s_1,s_2} = \max(|I^f_{s_1} \setminus I^f_{s_2}|, |I^f_{s_2} \setminus I^f_{s_1}|)$.

Proposition 2.1. *For all $T > 0$, $\mu > 1$ and $s_1, s_2 \in \mathbb{S}^1$*

$$\max\left\{\frac{\exp\left[-T\tilde{C}_\mu \Delta^f_{s_1,s_2}\right]}{4}, \frac{1 - \sqrt{T\tilde{C}_\mu \Delta^f_{s_1,s_2}/2}}{2}\right\} \leq p_e(s_1, s_2) \leq \exp\left[-TC_\mu \Delta^f_{s_1,s_2}\right],$$

with $C_\mu = \frac{(\mu-1)^2}{4} \min\left\{\frac{1}{2\mu}, \frac{3}{5+\mu}\right\}$ and $\tilde{C}_\mu = (\mu - 1)\log(\mu)$.

Proof. We set $I^f_1 = I^f_{s_1}$ (resp. $I^f_2 = I^f_{s_2}$), the set of neurons activated by s_1 (resp. s_2). We also denote $\mathbb{P}_1$ (resp. $\mathbb{P}_2$) the distribution of $N^1, ..., N^n$ and $\mathbb{E}_1$ (resp. $\mathbb{E}_2$) the corresponding expectation, when the applied stimulus is s_1 (resp. s_2). First notice that the Kullback-Leibler distance between both distribution is $K(\mathbb{P}_1, \mathbb{P}_2) = -\mathbb{E}_1(\log(\frac{d\mathbb{P}_2}{d\mathbb{P}_1}))$, where $\frac{d\mathbb{P}_2}{d\mathbb{P}_1}$ the Radon-Nikodym derivative of $\mathbb{P}_2$ with respect to $\mathbb{P}_1$. One can check that

$$\frac{d\mathbb{P}_2}{d\mathbb{P}_1} = \exp\left\{-(\mu-1)T(|I^f_2| - |I^f_1|) + \log(\mu)\left(\sum_{i \in I^f_2} N^i_T - \sum_{i \in I^f_1} N^i_T\right)\right\}$$

$$= \exp\left\{-(\mu-1)T(|I^f_2 \setminus I^f_1| - |I^f_1 \setminus I^f_2|) + \log(\mu)\left(\sum_{i \in I^f_2} N^i_T - \sum_{i \in I^f_1} N^i_T\right)\right\},$$

and

$$\mathbb{E}_1\left(\sum_{i \in I^f_2} N^i_T - \sum_{i \in I^f_1} N^i_T\right) = T(|I^f_2 \setminus I^f_1| - \mu|I^f_1 \setminus I^f_2|),$$

so that

$$K(\mathbb{P}_1, \mathbb{P}_2) = T(\mu-1)(|I^f_2 \setminus I^f_1| - |I^f_1 \setminus I^f_2|) - \log(\mu)T(|I^f_2 \setminus I^f_1| - \mu|I^f_1 \setminus I^f_2|)$$

$$= T(\mu - 1 - \log(\mu))|I^f_2 \setminus I^f_1| + T(\mu\log(\mu) - \mu + 1)|I^f_1 \setminus I^f_2|.$$

As a consequence, it follows that

$$K(\mathbb{P}_1, \mathbb{P}_2) \leq T\Delta^f_{s_1,s_2}(\mu - 1 - \log(\mu) + \mu\log(\mu) - \mu + 1)$$

$$= \tilde{C}_\mu T\Delta^f_{s_1,s_2}.$$

The lower bound on p_e follows from Theorem 2.2 of [25].

Now let us assume that $\Delta^f_{s_1,s_2} = |I^f_1 \setminus I^f_2|$ (if this is not the case we exchange s_1 and s_2). It is now sufficient to consider the test Ψ defined by

$$\Psi(N^1, \ldots, N^n) = \begin{cases} s_1, \text{ if } Z := \sum_{i \in I^f_1 \setminus I^f_2} N^i_T > |I^f_1 \setminus I^f_2|T(\mu+1)/2 \\ s_2, \text{ otherwise} \end{cases}. \quad (2.4)$$

Indeed,

$$\mathbb{P}_2(\Psi = s_1) = \mathbb{P}_2(Z > |I^f_1 \setminus I^f_2|T(\mu+1)/2) = \mathbb{P}_2(Z > T|I^f_1 \setminus I^f_2|(1 + (\mu-1)/2)).$$

Under $\mathbb{P}_2$, we have that $Z \sim \text{Poi}(T|I_1^f \setminus I_2^f|)$. Thus, by applying inequality (7.1) of Lemma 7.1 with $\theta = T|I_1^f \setminus I_2^f|$ and $x = (\mu - 1)/2$, it follows that

$$\mathbb{P}_2(\Psi = s_1) \leq \exp\left\{ -\frac{3T|I_1^f \setminus I_2^f|(\mu-1)^2}{4(5+\mu)} \right\}.$$

Under $\mathbb{P}_1$, we know that $Z \sim \text{Poi}(T\mu|I_1^f \setminus I_2^f|)$, so that by applying inequality (7.2) of Lemma 7.1 with $\theta = T\mu|I_1 \setminus I_2|$ and $x = (\mu - 1)/2)T|I_1 \setminus I_2|$, we deduce that

$$\mathbb{P}_1(\Psi = s_2) = \mathbb{P}_1(Z \leq T\mu|I_1^f \setminus I_2^f| - (\mu-1)/2)T|I_1^f \setminus I_2^f|) \leq \exp\left\{ -\frac{(\mu-1)^2 T|I_1^f \setminus I_2^f|}{8\mu} \right\},$$

which concludes the proof. □

2.3 Minimal discrimination time

From Proposition 2.1 we have that for a given admissible error level $\alpha \in (0, 1)$, if

1. If $T\tilde{C}_\mu \Delta_{s_1,s_2}^f < \log(1/4\alpha) = \log(1/\alpha) - \log(4)$, then $p_e > \alpha$.
2. If $TC_\mu \Delta_{s_1,s_2}^f \geq \log(1/\alpha)$, then $p_e \leq \alpha$.

Hence the minimal discrimination time stated in Problem 1, $T_{min}(f, s_1, s_2, \alpha)$, is of the order of $1/\Delta_{s_1,s_2}^f$, up to positive multiplicative constants in α and μ (see also Section 5 for numerical verification). This turns the other statistical problems into combinatorial problems. In particular, discrimination is not possible if and only if $\Delta_{s_1,s_2}^f = 0$, that is $I_{s_1}^f = I_{s_2}^f$.

The behavior of the quantities in α and μ are quite intuitive. If the level α tends to 0, T_{min} tends to infinity. If μ tends to infinity, T_{min} tends to 0. Considering that μ is some biological parameter that is fixed, as well as the admissible level α of reliability of the system, we are now focusing on the behavior of $1/\Delta_{s_1,s_2}^f$ and from now on, we denote with a slight abuse of language

$$T_{min}(f, s_1, s_2) = \frac{1}{\Delta_{s_1,s_2}^f}. \tag{2.5}$$

Let us now introduce the minimal time for which one can distinguish any pair of stimuli which are at least ρ apart with $\rho \in (0, 1/2]$:

$$T(f, \rho) = \max_{s_1, s_2 \in \mathbb{S}^1 : d(s_1, s_2) \geq \rho} T_{min}(f, s_1, s_2). \tag{2.6}$$

Clearly, the function $\rho \mapsto T(f, \rho)$ is non increasing (the larger the distance between two stimuli the smaller the time one needs to observe the activity of the network to distinguish one from the other).

Before speaking of minimax codes (Problem 2), let us derive an absolute lower bound on the range that can be detected by a general binary code f.

Proposition 2.2. *Let f be a general binary code. Then, for any $0 \leq \rho < 1/2^n$, $T(f, \rho) = \infty$.*

This means that whatever the code, there exists always two stimuli at distance less than 2^{-n} that cannot be discriminated, whatever the observation time.

Proof. For $A \subseteq [n]$, let S_A be the set of all $s \in \mathbb{S}$ such that $I_s^f = A$. Notice that for each $i \in [n]$

$$S_A = \left(\cap_{i \in A} f_i^{-1}(\{\mu\})\right) \cap \left(\cap_{i \in A^c} f_i^{-1}(\{1\})\right),$$

is a borelian of $\mathbb{S}^1$ for each $A \subseteq [n]$.

Notice also that if $s \in S_A \cap S_B$, then $A = I_s^f = B$. Moreover, $\cup_{A \subseteq [n]} S_A = \mathbb{S}^1$. Hence, $\{S_A\}_{A \subseteq [n]}$ forms a partition of $\mathbb{S}^1$. Let us denote Leb_* the pushforward (a measure on $(\mathbb{S}^1, \mathcal{S})$) of the Lebesgue on $[0, 1)$ induced by the map $[0, 1) \in \theta \mapsto (\cos(2\pi\theta), \sin(2\pi\theta)) \in \mathbb{S}$. By definition of Leb_* it holds that $\text{Leb}_*(\llbracket a, b \rrbracket) = 2\pi(\theta_b - \theta_a)$ and $\text{Leb}_*(\mathbb{S}^1) = 2\pi$.

Since $\{S_A\}_{A \subseteq [n]}$ forms a partition of $\mathbb{S}^1$ there exists $A \subseteq [n]$ such that $\text{Leb}_*(S_A) \geq 2\pi/2^n$. Suppose that $\text{diam}(S_A) =: \sup_{s_1, s_2 \in S_A} d(s_1, s_2) \leq \rho$. If this was true, then the set S_A would be contained in an interval of Leb_*-measure $2\pi\rho$ so that $2\pi/2^n \leq \text{Leb}_*(S_A) \leq 2\pi\rho$, implying that $\rho \geq 1/2^n$, a contradiction.

Therefore, if $\rho < 1/2^n$, then we must have that

$$\text{diam}(S_A) > \rho,$$

implying that we can find $s_1, s_2 \in S_A$ such that $d(s_1, s_2) > \rho$. In this case, s_1 and s_2 cannot be distinguished and $T_{min}(f, \rho) = \infty$. $\qquad \square$

3   Results for place cells code

In this section, we focus our analysis on the class of place cells code $\mathcal{P}$ defined in (2.1). Our first result is a lower bound for

$$T(\mathcal{P}, \rho) = \inf_{f \in \mathcal{P}} T(f, \rho), \tag{3.1}$$

which is (up to multiplicative constants in α and μ) the minimax discrimination time for the class of place cells code defined in (1.1).

3.1   Lower bound on the minimax discrimination time

Proposition 3.1. *For any $0 \leq \rho \leq 1/2$, we have that*

$$T(\mathcal{P}, \rho) \geq \frac{1}{\lfloor n/\lfloor (2\rho)^{-1} \rfloor \rfloor} \geq \frac{1}{\lfloor 4n\rho \rfloor}.$$

In particular, for any $0 \leq \rho \leq (2(n + 1))^{-1}$, we have that $T(\mathcal{P}, \rho) = \infty$.

Proof. Let $0 \leq \rho \leq 1/2$ and take $L = \lfloor 1/2\rho \rfloor$ so that $L \leq 1/2\rho < L + 1$ and hence $L > 1/2\rho - 1$. Set $\theta_0 = 0$ and $1 \leq k \leq L$, write $\theta_k = k\rho$ and $\theta_{-k} = 1 - k\rho$. For each $f \in \mathcal{P}$, denote $m_\rho^f = \min\{\Delta_{\rho_k, \rho_{k-1}}^f : -L + 1 \leq k \leq L\}$, where $\rho_k \in \mathbb{S}^1$ is identified with its argument θ_k.

Note that

$$I_{\rho_k}^f \setminus I_{\rho_{k-1}}^f = \{i \in [n] : a_i \in \rrbracket\rho_{k-1}, \rho_k\rrbracket \text{ and } b_i \notin \rrbracket\rho_{k-1}, \rho_k\rrbracket\} \subset \{i \in [n] : a_i \in \rrbracket\rho_{k-1}, \rho_k\rrbracket\}$$

and

$$I_{\rho_{k-1}}^f \setminus I_{\rho_k}^f = \{i \in [n] : b_i \in \rrbracket\rho_{k-1}, \rho_k\rrbracket \text{ and } a_i \notin \rrbracket\rho_{k-1}, \rho_k\rrbracket\} \subset \{i \in [n] : b_i \in \rrbracket\rho_{k-1}, \rho_k\rrbracket\}.$$

Hence, it follows that

$$\sum_{k=-L+1}^{L} |I_{\rho_k}^f \setminus I_{\rho_{k-1}}^f| \leq \sum_{k=-L+1}^{L} |\{i \in [n] : a_i \in \rrbracket\rho_{k-1}, \rho_k\rrbracket\}| \leq n,$$

and

$$\sum_{k=-L+1}^{L} |I_{\rho_{k-1}}^f \setminus I_{\rho_k}^f| \leq \sum_{k=-L+1}^{L} |\{i \in [n] : b_i \in \rrbracket\rho_{k-1}, \rho_k\rrbracket\}| \leq n.$$

Therefore, it follows that

$$\sum_{k=-L+1}^{L} \Delta_{\rho_k, \rho_{k-1}}^f = \sum_{k=-L+1}^{L} \max\left\{|I_{\rho_k}^f \setminus I_{\rho_{k-1}}^f|, |I_{\rho_{k-1}}^f \setminus I_{\rho_k}^f|\right\} \leq 2n,$$

implying that $2Lm_\rho^f \leq 2n$, that is, $m_\rho^f \leq n/L$. Since m_ρ^f is an integer, we must have $m_\rho^f \leq \lfloor n/L \rfloor = \lfloor n/\lfloor (2\rho)^{-1} \rfloor \rfloor$ so that

$$\min_{s_1, s_2 \in \mathbb{S}^1 : d(s_1, s_2) \geq \rho} \Delta_{s_1, s_2}^f \leq m_\rho^f \leq \lfloor n/\lfloor (2\rho)^{-1} \rfloor \rfloor.$$

Since $f \in \mathcal{P}$ is arbitrary, the above inequality ensures that

$$T^{-1}(\mathcal{P}, \rho) = \sup_{f \in \mathcal{P}} \min_{s_1, s_2 \in \mathbb{S}^1 : d(s_1, s_2) \geq \rho} \Delta_{s_1, s_2}^f \leq \lfloor n/\lfloor (2\rho)^{-1} \rfloor \rfloor.$$

Since $(2\rho)^{-1} \geq 1$ which implies that $\lfloor (2\rho)^{-1} \rfloor \geq 1$ we have that $4\lfloor (2\rho)^{-1} \rfloor \geq 2(\lfloor (2\rho)^{-1} \rfloor + 1) \geq \rho^{-1}$, so that the first part of the proof follows from inequality above.

To conclude the proof, observe that if $\rho \leq (2(n+1))^{-1}$ then $\lfloor (2\rho)^{-1} \rfloor \geq n+1 > n$, implying that $\lfloor n/\lfloor (2\rho)^{-1} \rfloor \rfloor = 0$ and in this case $T(\mathcal{P}, \rho) = \infty$. □

To better understand the lower bound provided by Proposition 3.1, let us adopt an asymptotic point of view in which the number of observed neurons $n \to \infty$ and $\rho = \rho_n$ is a function of n. For example, when ρ_n is a constant function of n the lower bound decreases as $1/n$, whereas in the regime $\rho_n \to 0$ such that $n\rho_n \to \infty$ the lower bound behaves as $(2n\rho_n)^{-1}$. Moreover, as long as $\rho_n \leq (2(n+1))^{-1} \approx (2n)^{-1}$, whatever the code $f \in \mathcal{P}$, there exists always two stimulus at distance ρ_n that cannot be distinguished, whatever the observation time.

3.2 Upper bounds and minimax codes

Next we obtain an upper bound for $T(\mathcal{P}, \rho)$ matching the lower bound above, up to multiplicative constants. To that end, we study the behavior of $T(f, \rho)$ for some examples of place cell codes and use the fact that $T(\mathcal{P}, \rho) \leq T(f, \rho)$ for any $f \in \mathcal{P}$.

Example 3.2 (1-Uniform code). For each $i \in \{0, 1, \ldots, n\}$, let p_i be the point in $\mathbb{S}^1$ associated with $\theta_i = \frac{i}{n}$. The 1-uniform code is defined as $f^{n,n} = (h_1, \ldots, h_n)$ (the superscript n, n stands for n neurons divided into n groups of size 1) where each h_i is given by (2.1) with $a_i = p_{i-1}$ and $b_i = p_i$. It is straightforward to check that $T(f^{1,1}, \rho) = \infty$ for all $0 \leq \rho < 1/2$ and $T(f^{1,1}, 1/2) = 1$. Let us analyse the case $n \geq 2$. In this case,

$$T(f^{n,n}, \rho) = \begin{cases} \infty, & \text{if } 0 \leq \rho < 1/n, \\ 1, & \text{otherwise} \end{cases}.$$

To check that, suppose first that $\rho < 1/n$. In this case, take θ_ρ such that $\rho \leq \theta_\rho < 1/n$, denote p_ρ the point in $\mathbb{S}^1$ associated to θ_ρ and observe that $p_\rho \in [\![p_0, p_1[\![$, $I_{p_\rho}^{f^{n,n}} = I_{p_0}^{f^{n,n}}$ and $d(p_\rho, p_0) = \theta_\rho \geq \rho$. Hence, $T(f^{n,n}, \rho) = \infty$. Now, suppose $1/n \leq \rho \leq 1/2$. In this case, it is not difficult to check that $I_{s_1}^{f^{n,n}} \neq I_{s_2}^{f^{n,n}}$ for all $s_1, s_2 \in \mathbb{S}^1$ such that $d(s_1, s_2) \geq \rho$ which implies that

$$\Delta_{s_1, s_2}^{f^{n,n}} \geq 1.$$

But since $|I_s^{f^{n,n}}| = 1$ for all $s \in \mathbb{S}^1$, we deduce that

$$\min_{s_1, s_2 \in \mathbb{S}^1 : d(s_1, s_2) \geq \rho} \Delta_{s_1, s_2}^{f^{n,n}} = 1,$$

and the result follows.

Example 3.3 (d-Uniform code)**.** For each $k \in \{0, 1, \ldots, d\}$, let p_k denote the point in $\mathbb{S}^1$ defined by $\theta_k = \frac{k}{d}$ and let h_k be given by (2.1) with $a_k = p_{k-1}$ and $b_k = p_k$. Also, write $L = \lfloor n/d \rfloor$. The d-uniform code is defined as $f^{n,d} = (f_1, \ldots, f_n)$ (the superscript n, d stands for n neurons divided into d groups) where $f_i = h_k$ for $(k-1)L + 1 \leq i \leq kL$ with $k \in \{1, \ldots, d-1\}$, and $f_i = h_d$ for $n \geq i \geq (d-1)L + 1$. One can check that for any $n \geq 2$,

$$T(f^{n,d}, \rho) = \begin{cases} \infty, & \text{if } 0 \leq \rho < 1/d, \\ 1/\lfloor n/d \rfloor, & \text{otherwise} \end{cases} .$$

To verify that, consider first the case that $\rho < 1/d$. In this case, take θ_ρ such that $\rho \leq \theta_\rho < 1/d$, denote p_ρ the point in $\mathbb{S}^1$ given by θ_ρ and observe that $p_\rho \in [\![p_0, p_1[\![$, $I_{p_\rho}^{f^{n,d}} = I_{p_0}^{f^{n,d}}$ and $d(p_\rho, p_0) = \theta_\rho \geq \rho$. Thus, we have that $T(f^{n,d}, \rho) = \infty$. Suppose now that $1/d \leq \rho \leq 1/2$. In this case, one can check that $I_{s_1}^{f^{n,d}} \cap I_{s_2}^{f^{n,d}} = \emptyset$ for all $s_1, s_2 \in \mathbb{S}^1$ such that $d(s_1, s_2) \geq \rho$ so that

$$\Delta_{s_1, s_2}^{f^{n,d}} = \max \left\{ |I_{s_1}^{f^{n,d}} \setminus I_{s_2}^{f^{n,d}}|, |I_{s_2}^{f^{n,d}} \setminus I_{s_1}^{f^{n,d}}| \right\} = \max \left\{ |I_{s_1}^{f^{n,d}}|, |I_{s_2}^{f^{n,d}}| \right\} \geq L.$$

Since the lower bound L is attained for some pair $s_1, s_2 \in \mathbb{S}^1$ (for example, take s_1 and s_2 given by $\theta_{s_1} = 0$ and $\theta_{s_2} = 1/2$ respectively), the results follows.

This codes allows us to prove the following result.

Corollary 3.4. *For all $\rho \in (1/n, 1/2]$, let $d = \lceil 1/\rho \rceil$. Then the d-Uniform code satisfies*

$$\frac{1}{\lfloor 4n\rho \rfloor} \leq T(\mathcal{P}, \rho) \leq T(f^{n,d}, \rho) \leq \frac{1}{\lfloor n/d \rfloor} \leq \frac{1}{\lfloor 3n\rho/2 \rfloor}$$

and is therefore minimax up to multiplicative constant.

3.3 A particular adaptive code

Now that we have found a minimax code for given ρ up to multiplicative constant, one could wonder if there exists a given place cells code, which for every ρ in a given range of values, achieves the minimax rate up to a constant. We refer to this problem as Problem 3, adaptation to the distance. The following code is a particular example of such a code.

Example 3.5 (An adaptive place cells code)**.** Let $g = (g_1, \ldots, g_n)$ be the code in $\mathcal{P}$ with f_i defined by (2.1) where the points $a_i, b_i \in \mathbb{S}^1$ are associated to $\theta_i = i/2n$ and $\theta_i + 1/2$ respectively. A visualisation is given in Figure 2. For the code g, one can show the following.

Proposition 3.6. *For any $n \geq 1$ and $0 \leq \rho \leq 1/2$,*

$$\frac{1}{\lfloor 2n\rho \rfloor} \leq T(g, \rho) \leq \frac{2}{\lfloor 2n\rho \rfloor}.$$

In particular the g code cannot discriminate at distance $\rho < 1/(2n)$ and it is adaptive to the distance in the class $\mathcal{P}$, up to some absolute multiplicative constant, in the range $\rho \in [1/(2n), 1/2]$.

Proof. One can easily check that for any $0 \leq s < 1$,

$$I_s^g = \begin{cases} \emptyset, & \text{if } 0 \leq \theta_s < 1/2n \\ \{1, \ldots, k\}, & \text{if } k/2n \leq \theta_s < (k+1)/2n \text{ with } 1 \leq k \leq n \\ \{k+1, \ldots, n\}, & \text{if } (n+k)/2n \leq \theta_s < (n+k+1)/2n \text{ with } 1 \leq k \leq n-1 \end{cases} .$$
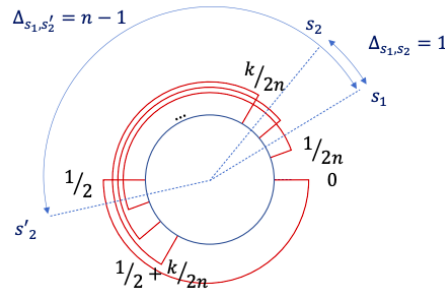
G. Ost and P. Reynaud-Bouret

Figure 2: Visual representation of Examples 3. For an easier visualisation, the different half red circles have a different radius. Nevertheless, each of them correspond a certain cell i of the code, and more precisely to the locations s such that $f_i(s) = \mu$. Two couples are considered (s_1, s_2) (standing for two positions that are very close) and (s_1, s_2') (standing for positions that are very far). We see that $\Delta_{s_1, s_2'} = n - 1 >> \Delta_{s_1, s_2} = 1$.

Thus, for any $0 \le \rho < 1/2n$, it follows that if one takes $s_1, s_2 \in \mathbb{S}^1$ with $\theta_{s_1} = 0$ and $\rho \le \theta_{s_2} < 1/2n$, then $I_{s_1}^g = I_{s_2}^g = \emptyset$ and $d(s_2, s_1) = |\theta_{s_2} - \theta_{s_1}| = \theta_{s_2} \ge \rho$, implying that $T_{min}(g, \rho) = \infty$, and the result holds since $\lfloor 2n\rho \rfloor = 0$.

Now, take $1/2n < \rho \le 1/2$ and $s_1 \in \mathbb{S}^1$. Suppose that $0 \le \theta_{s_1} < 1/2$. Let $k = \lfloor 2n\rho \rfloor$ and $q \in \{0, \ldots, n-1\}$ be such that $q/2n \le \theta_{s_1} < (q+1)/2n$.

Let us consider first the case $\theta_{s_1} + \rho < (n+1)/2n$. In this case, for all $\theta_{s_1} + \rho \le \theta_{s_2} < (n+1)/2n$, we have that $I_{s_1}^g \setminus I_{s_2}^g = \emptyset$ and $\{q+1, \ldots, q+k\} \subseteq I_{s_2}^g \setminus I_{s_1}^g$ so that $\Delta_{s_1, s_2}^g = \max\{|I_{s_1}^g \setminus I_{s_2}^g|, |I_{s_2}^g \setminus I_{s_1}^g|\} \ge q$ (and the equality can be achieved). For $(n+1)/2n \le \theta_{s_2} \le \theta_{s_1} + 1/2$, one can check that $I_{s_2}^g \setminus I_{s_1}^g = \{q+1, \ldots, n\}$ so that $\Delta_{s_1, s_2}^g = \max\{|I_{s_1}^g \setminus I_{s_2}^g|, |I_{s_2}^g \setminus I_{s_1}^g|\} \ge n - q \ge k$.

Hence, we have shown that for any $s_1 \in \mathbb{S}^1$ such that $0 \le \theta_{s_1} + \rho < 1/2$,

$$\min_{s_2 \in \mathbb{S}^1 : d(s_2, s_1) \ge \rho, \theta_{s_2} \le \theta_{s_1} + 1/2} \Delta_{s_1, s_2}^g \ge k. \tag{3.2}$$

Let us now consider the case $\theta_{s_1} + \rho \ge (n+1)/2n$. In this case, one can check that $n + 1 \le \lfloor 2n(\theta_{s_1} + \rho) \rfloor \le n + q$. Let $1 \le r \le q$ be such that $n + r = \lfloor 2n(\theta_{s_1} + \rho) \rfloor$. Since $2n(\theta_{s_1} + \rho) \ge q + k$, it follows that $n + r \ge q + k$, implying that $k \le (n - q) + r$. Now, note that for $\theta_{s_1} + \rho \le \theta_{s_2} \le \theta_{s_1} + 1/2$, we have that $I_{s_2}^g \setminus I_{s_1}^g = \{q+1, \ldots, n\}$ and $I_{s_1}^g \setminus I_{s_2}^g = \{1, \ldots, \ell\}$ with $r \le \ell \le q$. Hence, we have that

$$\Delta_{s_1, s_2}^g = \max\{|I_{s_1}^g \setminus I_{s_2}^g|, |I_{s_2}^g \setminus I_{s_1}^g|\} \ge \max\{(n-q), r\} \ge \frac{1}{2}(n - q + r) \ge k/2,$$

so that for any $s_1 \in \mathbb{S}^1$ such that $\theta_{s_1} + \rho \ge (n+1)/2n$,

$$\min_{s_2 \in \mathbb{S}^1 : d(s_2, s_1) \ge \rho, \theta_{s_2} \le \theta_{s_1} + 1/2} \Delta_{s_1, s_2}^g \ge k/2. \tag{3.3}$$

Therefore, combining inequalities (3.2) and (3.3) we deduce that for any $s_1 \in \mathbb{S}^1$ with $0 \le \theta_{s_1} < 1/2$,

$$\min_{s_2 \in \mathbb{S}^1 : d(s_2, s_1) \ge \rho, \theta_{s_2} \le \theta_{s_1} + 1/2} \Delta_{s_1, s_2}^g \ge k/2.$$

By similar arguments, one can also show that for any $s_1 \in \mathbb{S}^1$ with $0 \le \theta_{s_1} < 1/2$

$$\min_{s_2 \in \mathbb{S}^1 : d(s_2, s_1) \ge \rho, \theta_{s_2} \ge \theta_{s_1} + 1/2} \Delta_{s_1, s_2}^g \ge k/2 = \lfloor 2n\rho \rfloor /2.$$

The case $1/2 \leq \theta_{s_1} \leq 1$ follows along the same lines of reasoning by observing that if $\theta_{\tilde{s}_1} = \theta_{s_1} - 1/2$ then $0 \leq \theta_{\tilde{s}_1} < 1/2$ and $\theta_{\tilde{s}_1} + 1/2 = \theta_{s_1}$.

To conclude the proof, observe that

$$\min_{s_1 \in \mathbb{S}^1} \min_{s_2 \in \mathbb{S}^1 : d(s_2, s_1) \geq \rho} \Delta_{s_1, s_2}^g \leq \Delta_{s_0, a_k}^g = |I_{k/2n}^g| = k = \lfloor 2n\rho \rfloor,$$

where s_0 is the point $\mathbb{S}^1$ defined by $\theta_0 = 0$ and the result follows. $\qquad \square$

3.4 Random codes are adaptive as well

In the next example, we show that even random codes are adaptive (in the range $\rho \geq n^{-1/2}$, up to multiplicative constants) with large probability.

Example 3.7 (Random code). Let $A_1, \ldots, A_n$ and $B_1, \ldots, B_n$ be independent and uniformly distributed points on $\mathbb{S}^1$. The random code f^r is defined as $f^r = (f_1^r, \ldots, f_n^r)$, where for each $i \in [n]$, f_i is of the form (2.1) with $a_i = A_i$ and $b_i = B_i$. For the random code f^r, we can prove the following result.

Proposition 3.8. *There exist constants $K_1, K_2 > 0$ such that for each $n \geq K_1$ the following property holds: for each $x \in \mathbb{R}_{>0}$ and $0 \leq \rho \leq 1/2$,*

$$T(f^r, \rho) \leq \frac{1}{\left\lceil \frac{n}{2} \left(\rho - K_2 n^{-1/2} (1 + x^{1/2}(1 + (x/n)^{1/2})) \right) \right\rceil}, \tag{3.4}$$

with probability at least $1 - \exp(-x)$. In particular, for any $\delta \in (0, 1/4)$ and each n satisfying

$$n \geq \max\{K_1, (1 - 4\delta)^2 / (4K_2^2 \delta^2 (1 + x^{1/2})^2)\} := K_3,$$

we have the following: if $1/2 \geq \rho \geq n^{-1/2} K_2 (1 + x^{1/2} + x)/(1 - 4\delta)$ then

$$T(f^r, \rho) \leq \frac{1}{\lceil \delta n \rho \rceil}, \tag{3.5}$$

with probability at most $1 - \exp(-x)$.

Proof. First observe that to prove inequality in (3.4), it suffices to show that

$$\min_{s_1, s_2 \in \mathbb{S}^1 : d(s_1, s_2) \geq \rho} \Delta_{s_1, s_2}^{f^r} \geq \frac{n}{2} \left(\rho - K_2 n^{-1/2}(1 + x^{1/2}(1 + (x/n)^{1/2})) \right) \tag{3.6}$$

with probability at least $1 - \exp(-x)$. To that end, first observe that

$$|I_{s_1}^{f^r} \setminus I_{s_2}^{f^r}| + |I_{s_2}^{f^r} \setminus I_{s_1}^{f^r}| = \sum_{i=1}^n \mathbf{1}_{s_1 \in [\![A_i, B_i[\![, s_2 \in [\![B_i, A_i[\![} + \mathbf{1}_{s_2 \in [\![A_i, B_i[\![, s_1 \in [\![B_i, A_i[\![}$$

$$= n \left(1 - \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{s_1, s_2 \in [\![A_i, B_i[\![} + \mathbf{1}_{s_1, s_2 \in [\![B_i, A_i[\![} \right)$$

which implies that

$$\Delta_{s_1, s_2}^{f^r} \geq \frac{1}{2} \left(|I_{s_1}^{f^r} \setminus I_{s_2}^{f^r}| + |I_{s_2}^{f^r} \setminus I_{s_1}^{f^r}| \right)$$

$$= \frac{n}{2} \left(1 - \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{s_1, s_2 \in [\![A_i, B_i[\![} + \mathbf{1}_{s_1, s_2 \in [\![B_i, A_i[\![} \right).$$

Write $\xi_i = (A_i, B_i) \in \mathbb{S}^1 \times \mathbb{S}^1$ for $i \in [n]$, let $C_{s_1, s_2} = \{(a, b) \in \mathbb{S}^1 \times \mathbb{S}^1 : s_1, s_2 \in [\![a, b]\!]$ or $s_1, s_2 \in [\![b, a, [\![\}$ for each $s_1, s_2 \in \mathbb{S}^1$. With this notation, we deduce from the previous inequality that

$$\min_{s_1, s_2 \in \mathbb{S}^1 : d(s_1, s_2) \geq \rho} \Delta_{s_1, s_2}^{f^r} \geq \frac{n}{2} \left(1 - \max_{s_1, s_2 \in \mathbb{S}^1 : d(s_1, s_2) \geq \rho} \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{C_{s_1, s_2}}(\xi_i) \right).$$

Now, by Lemma 7.5,

$$\mathbb{E}\left(\mathbf{1}_{C_{s_1,s_2}}(\xi_i)\right) = 2\mathbb{P}(s_1, s_2 \in [\![A_i, B_i[\![) = 2\left(\frac{1}{2} - |\theta_{s_2} - \theta_{s_1}|\,(1 - |\theta_{s_2} - \theta_{s_1}|)\right).$$

Hence, it follows that for $s_1, s_2 \in \mathbb{S}^1$ such that $d(s_1, s_2) \geq \rho$ (that is $\rho \leq |\theta_{s_2} - \theta_{s_1}| \leq 1 - \rho$), we have that

$$\mathbb{E}\left(\mathbf{1}_{C_{s_1,s_2}}(\xi_i)\right) \leq 1 - 2\rho(1-\rho) \leq 1 - \rho,$$

where in the last inequality we have used that $0 \leq \rho \leq 1/2$. From the above inequality, we deduce that

$$\max_{s_1,s_2\in\mathbb{S}^1:d(s_1,s_2)\geq\rho} \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{C_{s_1,s_2}}(\xi_i) \leq (1-\rho) + W_{\mathcal{C}},$$

where $W_{\mathcal{C}}$ is the random variable defined as

$$W_{\mathcal{C}} = \max_{C\in\mathcal{C}} \frac{1}{n}\sum_{i=1}^{n} \left(\mathbf{1}_C(\xi_i) - \mathbb{E}\left(\mathbf{1}_C(\xi_i)\right)\right),$$

with the set $\mathcal{C} = \left\{C_{s_1,s_2} : s_1, s_2 \in \mathbb{S}^1 : d(s_1, s_2) \geq \rho\right\}$. Combining the above inequalities, we deduce that

$$\min_{s_1,s_2\in\mathbb{S}^1:d(s_1,s_2)\geq\rho} \Delta^{f^r}_{s_1,s_2} \geq \frac{n}{2}\left(\rho - W_{\mathcal{C}}\right). \tag{3.7}$$

By applying the Bousquet inequality (e.g., see inequality (5.49) on page 170 of [14]), we deduce that for any $x \in \mathbb{R}_{>0}$,

$$\mathbb{P}\left(W_{\mathcal{C}} \leq \mathbb{E}(W_{\mathcal{C}}) + \sqrt{\frac{2x}{n}(1 - \rho + \mathbb{E}(W_{\mathcal{C}}))} + \frac{x}{3n}\right) \geq 1 - \exp(-x). \tag{3.8}$$

From now on, we denote K a positive constant which can change from line to line. Let us denote $V(\mathcal{C})$ the VC-dimension of the set $\mathcal{C}$. By Lemma 6.4 of [14], there exists an absolute constant K such that, for all n for which $1 - \rho \geq K^2 V(\mathcal{C})(1 - \frac{1}{2}\log(1-\rho))/n$, it holds that

$$\mathbb{E}(W_{\mathcal{C}}) \leq \frac{K}{2}\sqrt{\frac{(1-\rho)}{n}V(\mathcal{C})(1 - \frac{1}{2}\log(1-\rho))} \leq \frac{K}{2}\sqrt{\frac{V(\mathcal{C})}{n}}(1 + \frac{1}{2}\log(2)),$$

where in the last inequality we have used that $0 \leq \rho \leq 1/2$. Combining the above inequality with (3.8) and by using again that $0 \leq \rho \leq 1/2$, it follows that for all n for which $n \geq 2K^2 V(\mathcal{C})(1 + \log(2)/2)$ and any $x \in \mathbb{R}_{>0}$,

$$W_{\mathcal{C}} \leq n^{-1/2}\left(\frac{K}{2}\sqrt{V(\mathcal{C})(1+\log(2))} + x^{1/2}\sqrt{2\left(1 + \frac{K}{2}\sqrt{V(\mathcal{C})(1+\log(2))}\right)} + x/3n^{1/2}\right)$$

with probability at least $1 - \exp(-x)$. Let us assume that the VC-dimension $V(\mathcal{C})$ is bounded by some absolute constant. In this case, it follows from the above inequality and (3.7) that for all $n \geq K^2$ and any $x \in \mathbb{R}_{>0}$,

$$\min_{s_1,s_2\in\mathbb{S}^1:d(s_1,s_2)\geq\rho} \Delta^{f^r}_{s_1,s_2} \geq \frac{n}{2}\left(\rho - n^{-1/2}K\left(1 + x^{1/2}\left(1 + (x/n)^{1/2}\right)\right)\right)$$

with probability at least $1 - \exp(-x)$, proving (3.6). For the VC-dimension $V(\mathcal{C})$, there are various ways to see that it is finite. One way is to say that $\xi = (A, B) \in C_{s_1,s_2}$ is equivalent to saying that (θ_A, θ_B) belongs to the union of

- $[0, \min(\theta_{s_1}, \theta_{s_2})] \times (\max(\theta_{s_1}, \theta_{s_2}), 1]$

- $(\max(\theta_{s_1}, \theta_{s_2}), 1] \times [0, \min(\theta_{s_1}, \theta_{s_2})]$
- $[\min(\theta_{s_1}, \theta_{s_2}), \max(\theta_{s_1}, \theta_{s_2})) \times [\min(\theta_{s_1}, \theta_{s_2}), \max(\theta_{s_1}, \theta_{s_2}))$
- $[0, \min(\theta_{s_1}, \theta_{s_2})] \times [0, \min(\theta_{s_1}, \theta_{s_2})]$
- $(\max(\theta_{s_1}, \theta_{s_2}), 1] \times (\max(\theta_{s_1}, \theta_{s_2}), 1]$

Hence it is included in the union of 5 rectangles. It is well known that the VC dimension of the family of rectangles is 4. Class $\mathcal{C}$ is included in the 5-fold unions of rectangles, whose VC dimension is bounded by $4 * 5 \log(5)$ up to multiplicative constants, thanks to [8]. Therefore Class $\mathcal{C}$ is of finite VC-dimension, which concludes the proof.

Note that with a closer look at all possibilities, one can show that class $\mathcal{C}$ cannot shatter samples of ξ_i's of size 5 and that in fact its VC dimension is therefore 4. However, the proof with all the possibilities is much longer than using the bound given by [8].

Let us now prove (3.5). Denote $\eta_{n,x} = (1 + x^{1/2} + x/n^{1/2})$. Clearly $(1 + x^{1/2} + x) \geq \eta_{n,x} \geq (1 + x^{1/2})$. Moreover, one can check that if $n \geq K_3$, then $2\delta(K_2 \eta_{n,x})/(1 - 4\delta) \geq n^{-1/2}$ which implies that $(K_2 \eta_{n,x})/(1 - 4\delta) \geq 2n^{-1/2} + K_2 \eta_{n,x}$. Hence, if $\rho \geq n^{-1/2} K_2(1 + x^{1/2} + x)/(1 - 4\delta) \geq n^{-1/2} K_2 \eta_{n,x}/(1 - 4\delta)$, then

$$n(\rho - K_2 \eta_{n,x}/n^{1/2})/2 \geq 1$$

so that

$$\lfloor n(\rho - K_2 \eta_{n,x}/n^{1/2})/2 \rfloor \geq n(\rho - K_2 \eta_{n,x}/n^{1/2})/4 \geq n\rho\delta,$$

and the result follows from (3.4).

$\square$

Summary of the results on place cells codes

1. For any given place cells code f of size n, and any $0 \leq \rho \leq 1/(2(n+1))$, there exists a pair (s_1, s_2) at distance at least ρ that cannot be discriminated.
2. For any $\rho \geq 1/(2n)$, the minimax discrimination rate for place cell codes is of order $1/(n\rho)$.
3. The code g defined in Example 3.5 is adaptive, meaning that it reaches this rate for all $\rho \geq 1/(2n)$, up to multiplicative constants.
4. Random codes f^r defined in Example 3.7 are adaptive in the range $\rho \geq c/n^{1/2}$, for some constant $c > 0$, up to multiplicative constants, with large probability.

---

4 Results for grid cells code

In this section, we discuss some results for the class of grid cells code defined in (2.2) $\mathcal{G}((n_i, \lambda_i)_{i=1,\dots,m})$ and the more general $\bar{\mathcal{G}}$. In what follows, for $f \in \bar{\mathcal{G}}$, $s_1, s_2 \in \mathbb{S}^1$ and $i \in [m]$, let us denote

$$\Delta_{s_1,s_2}^{f,i} = \max \left\{ |I_{s_1}^{f,i} \setminus I_{s_2}^{f,i}|, |I_{s_2}^{f,i} \setminus I_{s_1}^{f,i}| \right\},$$

where $I_s^{f,i} = \{j \in M_i : f_{ji}(s) = \mu\}$ is the set of cells activated by stimulus $s \in \mathbb{S}^1$ in the i-th module. The first result establishes a useful link between grid cells code and place cells code.

Proposition 4.1. *For any $f \in \bar{\mathcal{G}}$ and $s_1, s_2 \in \mathbb{S}^1$, the following inequality holds*

$$\sum_{i=1}^{m} \Delta_{s_1,s_2}^{f,i} \geq \Delta_{s_1,s_2}^{f} \geq \frac{1}{2} \sum_{i=1}^{m} \Delta_{s_1,s_2}^{f,i}. \tag{4.1}$$

Proof. Suppose that $\Delta^f_{s_1,s_2} = |I^f_{s_1} \setminus I^f_{s_2}|$ (if not exchange s_1 and s_2). On one hand, observe that since the modules are disjoint we have that

$$|I^f_{s_1} \setminus I^f_{s_2}| = \sum_{i=1}^m |I^{f,i}_{s_1} \setminus I^{f,i}_{s_2}| \leq \sum_{i=1}^m \max\left\{|I^{f,i}_{s_1} \setminus I^{f,i}_{s_2}|, |I^{f,i}_{s_2} \setminus I^{f,i}_{s_1}|\right\} = \sum_{i=1}^m \Delta^{f,i}_{s_1,s_2}.$$

On the other hand, using that

$$|I^f_{s_1} \setminus I^f_{s_2}| + |I^f_{s_2} \setminus I^f_{s_1}| = \sum_{i=1}^m |I^{f,i}_{s_1} \setminus I^{f,i}_{s_2}| + |I^{f,i}_{s_2} \setminus I^{f,i}_{s_1}| \geq \sum_{i=1}^m \max\left\{|I^{f,i}_{s_1} \setminus I^{f,i}_{s_2}|, |I^{f,i}_{s_2} \setminus I^{f,i}_{s_1}|\right\},$$

we obtain that

$$2\Delta^f_{s_1,s_2} \geq |I^f_{s_1} \setminus I^f_{s_2}| + |I^f_{s_2} \setminus I^f_{s_1}| \geq \sum_{i=1}^m \Delta^{f,i}_{s_1,s_2},$$

and the result follows. $\qquad\square$

4.1 Lower bound on the minimax discrimination time for $\mathcal{G}((n_i, \lambda_i)_{i=1,\ldots,m})$

The following result provides a lower bound on the minimax discrimination time (up to constants depending on α and μ) defined in (1.1), when restricted to the class of grid cells code:

$$T(\mathcal{G}((n_i, \lambda_i)_{i=1,\ldots,m}), \rho) = \inf_{f \in \mathcal{G}((n_i, \lambda_i)_{i=1,\ldots,m})} T(f, \rho). \tag{4.2}$$

Theorem 4.2. *If $0 \leq \rho \leq 1/2$ and λ_1/λ_2 is even or if $0 \leq \rho \leq 1/2 - \lambda_2/2$ and λ_1/λ_2 is odd, we have that*

$$T(\mathcal{G}((n_i, \lambda_i)_{i=1,\ldots,m}), \rho) \geq \frac{1}{\left\lceil 6\min_{1\leq k\leq m}\left\{\sum_{i=1}^k \frac{n_i}{\lambda_i}\max\{\rho, \lambda_{k+1}\}\right\}\right\rceil},$$

with the convention that $\lambda_{m+1} = 0$. In particular, if we define $j_\rho = \max\{1 \leq k \leq m : \lambda_k \geq \rho\}$, then $T(\mathcal{G}((n_i, \lambda_i)_{i=1,\ldots,m}), \rho) = \infty$ whenever one of the following conditions hold:

- *either there exists $2 \leq k \leq j_\rho$ such that*

$$(k-1)\lambda_k < \min_{1 \leq i < k}\left\lceil \frac{\lambda_i}{6n_i}\right\rceil,$$

- *or $j_\rho \rho < \min_{1 \leq i \leq j_\rho}\left\lceil \frac{\lambda_i}{6n_i}\right\rceil$.*

Proof. We need to compute an upper bound for $\min_{s_1,s_2 \in \mathbb{S}^1 : d(s_1,s_2) \geq \rho} \Delta^f_{s_1,s_2}$ uniformly on $f \in \mathcal{F}_2$. By the definition of j_ρ, if $j_\rho \geq 2$ then we have that

$$\lambda_k \geq \rho > \lambda_{j_\rho+1}, \text{ for all } 2 \leq k \leq j_\rho \text{ and } \rho \leq \lambda_1/2.$$

When $j_\rho = 1$ it holds that $\lambda_2 < \rho \leq 1/2 = \lambda_1/2$.

Let us consider first the case $j_\rho \geq 2$. In this case, let $2 \leq k \leq j_\rho$ and take $s_1, s_2 \in \mathbb{S}^1$ such that $d(s_1, s_2) = \lambda_k \geq \rho$. By Lemma 7.3 it follows that $s_1 \mod \lambda_i = s_2 \mod \lambda_i$ for all $k \leq i \leq m$ so that for all $j \in M_i$,

$$f_{ij}(s_1) = f_{ij}(s_1 \mod \lambda_i) = f_{ij}(s_2 \mod \lambda_i) = f_{ij}(s_2),$$

implying that $\Delta^{f,i}_{s_1,s_2} = 0$.

Hence, by Proposition 4.1 we have that

$$\Delta^f_{s_1,s_2} \leq \sum_{i=1}^{k-1} \Delta^{f,i}_{s_1,s_2}.$$

For each $1 \le i \le k-1$, let ρ_ℓ^i in S^{λ_i} defined by $\theta_{\rho_\ell^i} = \ell \lambda_k$ with $0 \le \ell \le D_{i+1:k}$, where $D_{i+1:k} := \lambda_i / \lambda_k$ is an integer larger than 2 for all $i < k$. For later use, let us also define $D_{2:1} = 1$. By arguing as in the proof of Proposition 3.1, one can show that

$$\sum_{\ell=1}^{D_{i+1:k}} \Delta_{\rho_\ell^i, \rho_{\ell-1}^i}^{f,i} \le 2n_i.$$

Now, take $0 \le \ell \le D_{2:k}, 1 \le i \le k-1$ and $j \in M_i$, and observe that $f_{ij}(\rho_\ell^1) = f_{ij}(\rho_\ell^1 \bmod \lambda_i)$.

Claim. It holds true that $\rho_\ell^1 \bmod \lambda_i = \rho_{\ell \bmod D_{i+1:k}}^i$.

Proof of the Claim. Write $\ell = rD_{i+1:k} + \ell \bmod D_{i+1:k}$ for some positive integer r. Since $\lambda_i = D_{i+1:k}\lambda_k$, we deduce that

$$\ell \lambda_k = r\lambda_i + (\ell \bmod D_{i+1:k})\lambda_k$$

with $(\ell \bmod D_{i+1:k})\lambda_k < \lambda_i$. As a consequence, we obtain that

$$\theta_{\rho_\ell^1 \bmod \lambda_i} = \theta_{\rho_\ell^1} \bmod \lambda_i = \ell \lambda_k \bmod \lambda_i = (\ell \bmod D_{i+1:k})\lambda_k = \theta_{\rho_{\ell \bmod D_{i+1:k}}^i},$$

implying that $\rho_\ell^1 \bmod \lambda_i = \rho_{\ell \bmod D_{i+1:k}}^i$ (by the injectivity of the function $\mathbb{S}^1 \ni a \mapsto \theta_a$.) This concludes the proof of the Claim.

The claim implies that $f_{ij}(\rho_\ell^1) = f_{ij}(\rho_{\ell \bmod D_{i+1:k}}^i)$ for each $0 \le \ell \le D_{2:k}, 1 \le i \le k-1$ and $j \in M_i$, so that (notice that $D_{2:k} = D_{2:i}D_{i+1:k}$),

$$\sum_{\ell=1}^{D_{2:k}} \Delta_{\rho_\ell^1, \rho_{\ell-1}^1}^{f,i} = D_{2:i} \sum_{\ell=1}^{D_{i+1:k}} \Delta_{\rho_\ell^i, \rho_{\ell-1}^i}^{f,i} \le D_{2:i} 2n_i.$$

Therefore, the above inequality implies that

$$\sum_{\ell=1}^{D_{2:k}} \sum_{i=1}^{k-1} \Delta_{\rho_\ell^1, \rho_{\ell-1}^1}^{f,i} \le 2(D_{2:1}n_1 + D_{2:2}n_2 + \ldots + D_{2:k-1}n_{k-1})$$

which, in turn, ensures that

$$D_{2:k} \min_{s_1,s_2 \in \mathbb{S}^1: d(s_1,s_2)=\lambda_k} \left\{ \sum_{i=1}^{k-1} \Delta_{s_1,s_2}^{f,i} \right\} \le \sum_{\ell=1}^{D_{2:k}} \sum_{i=1}^{k-1} \Delta_{\rho_\ell^1, \rho_{\ell-1}^1}^{f,i}$$
$$\le 2(D_{2:1}n_1 + D_{2:2}n_2 + \ldots + D_{2:k-1}n_{k-1})$$

or equivalently,

$$\min_{s_1,s_2 \in \mathbb{S}^1: d(s_1,s_2)=\lambda_k} \left\{ \sum_{i=1}^{k-1} \Delta_{s_1,s_2}^{f,i} \right\} \le \frac{2}{D_{2:k}}(D_{2:1}n_1 + D_{2:2}n_2 + \ldots + D_{2:k-1}n_{k-1}) = 2\lambda_k \sum_{i=1}^{k-1} \frac{n_i}{\lambda_i},$$

where in the equality we have used that $D_{2:k} = D_{2:i}D_{i+1:k}$ and $D_{i+1:k} = \lambda_i \lambda_k^{-1}$.

Hence, we have shown that for each $2 \le k \le j_\rho$,

$$\min_{s_1,s_2 \in \mathbb{S}^1: d(s_1,s_2) \ge \rho} \Delta_{s_1,s_2}^f \le 2\lambda_k \sum_{i=1}^{k-1} \frac{n_i}{\lambda_i} = \sum_{i=1}^{k-1} \frac{2n_i}{\lambda_i} \max\{\lambda_k, \rho\}.$$

If $\rho > \lambda_{j_\rho}/2 - \lambda_{j_\rho+1}/2$, then

$$\lambda_{j_\rho} < 2\rho + \lambda_{j_\rho+1} < 3\rho.$$

Hence

$$\sum_{i=1}^{j_\rho-1} \frac{2n_i}{\lambda_i} \max\{\lambda_{j_\rho}, \rho\} = \sum_{i=1}^{j_\rho-1} \frac{2n_i}{\lambda_i} \lambda_{j_\rho} < 3 \sum_{i=1}^{j_\rho-1} \frac{2n_i}{\lambda_i} \rho$$

and for all $k \geq j_\rho$,

$$3 \sum_{i=1}^{k} \frac{2n_i}{\lambda_i} \max\{\rho, \lambda_{k+1}\} \geq 3 \sum_{i=1}^{j_\rho} \frac{2n_i}{\lambda_i} \rho > 3 \sum_{i=1}^{j_\rho-1} \frac{2n_i}{\lambda_i} \rho,$$

so that

$$\min_{s_1,s_2 \in \mathbb{S}^1 : d(s_1,s_2) \geq \rho} \Delta^f_{s_1,s_2} \leq \min_{1 \leq k \leq j_\rho - 1} \left\{ \sum_{i=1}^{k} \frac{2n_i}{\lambda_i} \max\{\lambda_{k+1}, \rho\} \right\}$$

$$= \min \left\{ \min_{1 \leq k \leq j_\rho - 1} \left\{ \sum_{i=1}^{k} \frac{2n_i}{\lambda_i} \max\{\lambda_{k+1}, \rho\} \right\}, \min_{j_\rho \leq k \leq m} \left\{ 3 \sum_{i=1}^{k} \frac{2n_i}{\lambda_i} \max\{\lambda_{k+1}, \rho\} \right\} \right\}.$$

Therefore, we deduce that when $j_\rho \geq 2$ and $\rho > \lambda_{j_\rho}/2 - \lambda_{j_\rho+1}/2$, it holds that

$$\min_{s_1,s_2 \in \mathbb{S}^1 : d(s_1,s_2) \geq \rho} \Delta^f_{s_1,s_2} \leq 3 \min_{1 \leq k \leq m} \left\{ \sum_{i=1}^{k} \frac{2n_i}{\lambda_i} \max\{\lambda_{k+1}, \rho\} \right\}. \qquad (4.3)$$

Now, suppose that $\rho \leq \lambda_{j_\rho}/2 - \lambda_{j_\rho+1}/2$. Let $r = \lceil \rho/\lambda_{j_\rho+1} \rceil$. If $D_{j_\rho+1:j_\rho+1}$ is even then $r < D_{j_\rho+1:j_\rho+1}/2$ because $r < \lambda_{j_\rho}/2$. If $D_{j_\rho+1:j_\rho+1}$ is odd, then $\lambda_{j_\rho}/2 - \lambda_{j_\rho+1}/2 = \lambda_{j_\rho+1}(D_{j_\rho+1:j_\rho+1} - 1)/2$ is a multiple of $\lambda_{j_\rho+1}$. So in both cases, $r < D_{j_\rho+1:j_\rho+1}/2$. Therefore $L = \lfloor D_{j_\rho+1:j_\rho+1}/r \rfloor$ is always larger than 2.

In module j_ρ, we construct $L + 1$ different points $\rho_\ell^{j_\rho}$ of argument $\ell r \lambda_{j_\rho+1}$, for $\ell = 0, ..., L$ which leads to L distinct pairs at distance $\bar{\rho} = r \lambda_{j_\rho+1}$.

By periodicity, we deploy these points on module 1, there are $D_{2:j_\rho} L$ of those. Taking their modulo, they correspond on module $i = 1, ..., m$ to $D_{i+1:j_\rho} L$ pairs at distance $\bar{\rho}$. Each of these pairs are at distance a multiple of λ_k for $k > j_\rho$ and therefore cannot be detected by the modules for $k > j_\rho$.

Therefore a similar argument as above lead us to

$$\min_{s_1,s_2 \in \mathbb{S}^1 : d(s_1,s_2) = \bar{\rho}} \Delta^f_{s_1,s_2} \leq \frac{2}{D_{2:j_\rho} L}(D_{2:1} n_1 + ... + D_{2:j_\rho} n_{j_\rho}).$$

Since $L \geq 2$, $L \geq 2/3(D_{j_\rho+1:j_\rho+1}/r)$, hence

$$\min_{s_1,s_2 \in \mathbb{S}^1 : d(s_1,s_2) = \bar{\rho}} \Delta^f_{s_1,s_2} \leq 3r \sum_{i=1}^{j_\rho} \frac{n_i}{\lambda_i} \lambda_{j_\rho+1}.$$

But $r \leq 2\rho/\lambda_{j_\rho+1}$. Hence

$$\min_{s_1,s_2 \in \mathbb{S}^1 : d(s_1,s_2) = \bar{\rho}} \Delta^f_{s_1,s_2} \leq 3 \sum_{i=1}^{j_\rho} \frac{2n_i}{\lambda_i} \rho$$

Also, as before

$$3 \sum_{i=1}^{j_\rho} \frac{2n_i}{\lambda_i} \rho \leq 3 \sum_{i=1}^{k} \frac{2n_i}{\lambda_i} \max\{\lambda_{k+1}, \rho\},$$

for all $k \geq j_\rho$. Finally since $\bar{\rho} \geq \rho$, we get that as long as $j_\rho \geq 2$, (4.3) holds.

The last case to treat is when $j_\rho = 1$. If $\rho \leq \lambda_1/2 - \lambda_2/2$, we do as above with $r = \lceil \rho/\lambda_2 \rceil$ and $\bar{\rho} = r\lambda_2 \leq 1/2$, constructing pairs that can be distinguished only by module 1 and the same bound applies.

In the case $\rho \geq \lambda_1/2 - \lambda_2/2$ and λ_1/λ_2 is even, we can always take $s_1 = 0$ and $s_2 = 1/2$. Since λ_1/λ_2 is even, $d(s_1, s_2)$ is a multiple of λ_2 and therefore of all the other λ_i for $i \geq 2$. Hence it can only be detected by the first module and

$$\min_{s_1, s_2 \in \mathbb{S}^1 : d(s_1, s_2) \geq \rho} \Delta^f_{s_1, s_2} \leq n_1 \leq 3n_1\rho \leq 3 \sum_{i=1}^{k} \frac{2n_i}{\lambda_i} \max\{\lambda_{k+1}, \rho\},$$

for all $k \geq 1$ and we conclude as for the other cases.

To conclude, let us just remark that if there exists $2 \leq k \leq j_\rho$ such that

$$(k-1)\lambda_k < \min_{i<k}\left[\frac{\lambda_i}{6n_i}\right] = \frac{1}{6\max_{1 \leq i < k}\{n_i/\lambda_i\}},$$

then $\sum_{i=1}^{k-1} \frac{n_i}{\lambda_i}\lambda_k \leq (k-1)\lambda_k \max_{1 \leq i < k}\{\frac{n_i}{\lambda_i}\} < 1/6$, so that the integer part in the lower bound is null. The same phenomenon appears at $k = j_\rho$ if $j_\rho\rho < \min_{i \leq j_\rho}\left[\frac{\lambda_i}{6n_i}\right]$. $\qquad \square$

To understand better the bound given by Theorem 4.2, note that $1/\lfloor 2n_i\rho/\lambda_i \rfloor$ is in fact the rate for the place cells code in module i. So, roughly up to multiplicative constant, to distinguish at distance ρ, the code needs to be such that the modules of the grid cells have to be coherent (i.e. the period $k\lambda_{k+1}$ needs to be detected by at least one of the modules before, for all k such that $\lambda_{k+1} \geq \rho$) and that $j_\rho\rho$ also needs to be in the detection range of at least one of modules $i \leq j_\rho$. It is not clear whether the factor k should be present or not. Also, for the rate itself note that $\sum_{i=1}^{k} \frac{n_i}{\lambda_i}\lambda_{k+1}$ is not necessarily monotonous in k. Hence we could be in a situation where even if we are interested by small ρ that cannot be detected by the first module, the rate of detection of λ_2 by the first module acts as a threshold that impacts the rates at ρ.

4.2 Upper bounds and a particular adaptive code for $\mathcal{G}((n_i, \lambda_i)_{i=1,\ldots,m})$

Example 4.3. (An adaptive grid cell code) This code is made of adaptive place cells code as in Example 3 for each of the modules. More precisely, the module i is the set of neurons $M_i = \{n_{i-1}+1, \ldots, n_{i-1}+n_i\}$ and $f_{n_i+j,i}$ is associated with the points $a_{ij}, b_{ij} \in \mathbb{S}^{\lambda_i}$ corresponding to the angles $\theta_{a_{ij}} = \frac{j\lambda_i}{2n_i}$ and $\theta_{b_{ij}} = \frac{(j+n_i)\lambda_i}{2n_i}$. Let us denote this code by g_{gc}.

For the code g_{gc}, one can prove the following result.

Proposition 4.4. *For any $0 \leq \rho \leq 1/2$,*

$$T(g_{gc}, \rho) \leq \frac{4}{\min_{1 \leq k \leq m} \sum_{j=1}^{k}\left[\frac{n_j}{\lambda_j}\max\{\lambda_{k+1}, \rho\}\right]},$$

with the convention $\lambda_{m+1} = 0$. In particular, if we define $j_\rho = \max\{1 \leq k \leq m : \lambda_k \geq \rho\}$, then for all ρ large enough such that

- *for all $k \leq j_\rho$, there exists $j < k$ such that $\lambda_k \geq \frac{\lambda_j}{n_j}$*

- *and there exists $j \leq j_\rho$ with $\rho \geq \frac{\lambda_j}{n_j}$, we have that*

$$\frac{1}{6\min_{1\le k\le m}\sum_{j=1}^{k}\frac{n_j}{\lambda_j}\max\{\lambda_{k+1},\rho\}} \le T(\mathcal{G}((n_i,\lambda_i)_{i=1,\dots,m}),\rho)$$

$$\le T(g_{gc},\rho) \le \frac{16\log_2(1/\rho)}{\min_{1\le k\le m}\sum_{j=1}^{k}\frac{n_j}{\lambda_j}\max\{\lambda_{k+1},\rho\}},$$

meaning that g_{gc} is adaptive in this range of ρ up to a factor $\log_2(1/\rho)$.

Proof. Take a pair $s_1, s_2 \in \mathbb{S}^1$ such that $d(s_1,s_2) \ge \rho$ and define

$$j_{s_1,s_2} = \max\left\{i\in[n]:\frac{\lambda_i}{2}\ge d(s_1,s_2)\right\}.$$

By using Lemma 7.2, we then conclude that for $1\le j\le j_{s_1,s_2}$,

$$d(s_1 \bmod \lambda_j, s_2 \bmod \lambda_j) = d(s_1,s_2) \ge \rho.$$

Hence, proceeding as in the proof of Proposition 3.6, we have then that

$$\sum_{i=1}^{m}\Delta_{s_1,s_2}^{g_{gc},i} \ge \sum_{i=1}^{j_{s_1,s_2}}\Delta_{s_1,s_2}^{g_{gc},i}$$

$$\ge \sum_{i=1}^{j_{s_1,s_2}}\min_{u_1,u_2\in\mathbb{S}^{\lambda_i}:d_{\lambda_i}(u_1,u_2)\ge\rho}\Delta_{u_1,u_2}^{g_{gc},i}$$

$$\ge \frac{1}{2}\sum_{j=1}^{j_{s_1,s_2}}\left\lfloor\frac{2n_j}{\lambda_j}\rho\right\rfloor = \frac{1}{2}\sum_{j=1}^{j_{s_1,s_2}}\left\lfloor\frac{2n_j}{\lambda_j}\max\left\{\frac{\lambda_{j_{s_1,s_2}+1}}{2},\rho\right\}\right\rfloor.$$

Combining the above inequality and Proposition 4.1, we deduce that

$$\min_{s_1,s_2\in\mathbb{S}^1:d(s_1,s_2)\ge\rho}\Delta_{s_1,s_2}^{g_{gc}} \ge \frac{1}{2}\min_{s_1,s_2\in\mathbb{S}^1:d(s_1,s_2)\ge\rho}\sum_{i=1}^{m}\Delta_{s_1,s_2}^{g_{gc},i}$$

$$\ge \frac{1}{4}\min_{1\le k\le m}\sum_{j=1}^{k}\left\lfloor\frac{2n_j}{\lambda_j}\max\left\{\frac{\lambda_{k+1}}{2},\rho\right\}\right\rfloor.$$

$$\ge \frac{1}{4}\min_{1\le k\le m}\sum_{j=1}^{k}\left\lfloor\frac{n_j}{\lambda_j}\max\{\lambda_{k+1},\rho\}\right\rfloor.$$

$\square$

To conclude the proof we need to show that

$$\min_{1\le k\le m}\sum_{j=1}^{k}\left\lfloor\frac{n_j}{\lambda_j}\max\{\lambda_{k+1},\rho\}\right\rfloor \ge 4\log(\rho^{-1})\min_{1\le k\le m}\sum_{j=1}^{k}\frac{n_j}{\lambda_j}\max\{\lambda_{k+1},\rho\}. \quad (4.4)$$

To that end, first observe that

$$\min_{1\le k\le m}\sum_{j=1}^{k}\left\lfloor\frac{n_j}{\lambda_j}\max\{\lambda_{k+1},\rho\}\right\rfloor = \min\left\{\min_{2\le k\le j_\rho}\sum_{j=1}^{k-1}\left\lfloor\frac{n_j}{\lambda_j}\lambda_k\right\rfloor,\sum_{j=1}^{j_\rho}\left\lfloor\frac{n_j}{\lambda_j}\rho\right\rfloor\right\}, \quad (4.5)$$

and also that for each $k\le j_\rho$

$$\sum_{j=1}^{k-1}\left\lfloor\frac{n_j}{\lambda_j}\lambda_k\right\rfloor \ge \max_{1\le j<k}\left\lfloor\frac{n_j}{\lambda_j}\lambda_k\right\rfloor$$

$$\ge \frac{1}{2}\max_{1\le j<k}\frac{n_j}{\lambda_j}\lambda_k \ge \frac{1}{2k}\sum_{j=1}^{k-1}\frac{n_j}{\lambda_j}\lambda_k.$$

Since $\rho \leq \lambda_{j_\rho} \leq 2^{-(j_\rho-1)}$ it follows that $2\log_2(\rho^{-1}) \geq \log(\rho^{-1}) + 1 \geq j_\rho \geq k$, so that from the above inequality we obtain that

$$\sum_{j=1}^{k-1} \left\lfloor \frac{n_j}{\lambda_j} \lambda_k \right\rfloor \geq \frac{1}{4\log_2(\rho^{-1})} \sum_{j=1}^{k-1} \frac{n_j}{\lambda_j} \lambda_k.$$

By similar arguments, one can also show that

$$\sum_{j=1}^{j_\rho} \left\lfloor \frac{n_j}{\lambda_j} \rho \right\rfloor \geq \frac{1}{4\log_2(\rho^{-1})} \sum_{j=1}^{j_\rho} \frac{n_j}{\lambda_j} \rho.$$

Combining the last two inequalities above with (4.5), we then deduce that

$$\min_{1 \leq k \leq m} \sum_{j=1}^{k} \left\lfloor \frac{n_j}{\lambda_j} \max\{\lambda_{k+1}, \rho\} \right\rfloor \geq \frac{1}{4\log_2(\rho^{-1})} \min\left\{ \min_{2 \leq k \leq j_\rho} \sum_{j=1}^{k-1} \frac{n_j}{\lambda_j} \lambda_k, \sum_{j=1}^{j_\rho} \frac{n_j}{\lambda_j} \rho \right\}$$

$$= \frac{1}{4\log_2(\rho^{-1})} \min_{1 \leq k \leq m} \sum_{j=1}^{k} \frac{n_j}{\lambda_j} \max\{\lambda_{k+1}, \rho\},$$

proving (4.4), and the result follows.

4.3 Random codes are also adaptive in $\mathcal{G}((n_i, \lambda_i)_{i=1,\ldots,m})$

Example 4.5 (Random code for grid cells)**.** Again we use random codes on each of the modules. More precisely, for each $1 \leq i \leq m$, let $M_i = \{n_{i-1} + 1, \ldots, n_{i-1} + n_i\}$ denote the set of neurons in module i. Let $(A_{ij})_{j \in M_i}$ and $(B_{ij})_{j \in M_i}$ be independent and uniformly distributed points on $\mathbb{S}^{\lambda_i}$. The random code f_{gc}^r for the grid cells is defined as $f_{gc}^r = (f_{gc,11}^r, \ldots, f_{gc,ij}^r, \ldots, f_{gc,Mn}^r)$, where $f_{gc,ij}^r$ is given by (2.2) with $a_{ij} = A_{ij}$ and $b_{ij} = B_{ij}$ for each $i \in [m]$ and $j \in M_i$.

Proposition 4.6. *There exist constants $K_1, K_2 > 0$ such that if $\min_{1 \leq i \leq m}\{n_i\} \geq K_1$ the following property holds: for $x \in \mathbb{R}_{>0}$ and $0 \leq \rho \leq 1/2$, $T(f_{gc}^r, \rho)$ is upper bound by the inverse of the ceiling function of*

$$\min_{1 \leq k \leq m} \sum_{i=1}^{k} \frac{n_i}{4} \left(\frac{\max\left\{\rho, \frac{\lambda_{k+1}}{2}\right\}}{\lambda_i} - \frac{K_2}{n_i^{1/2}} \left[1 + (x + \log(m))^{1/2} \left[1 + \left(\frac{(x + \log(m))}{n_i} \right)^{1/2} \right] \right] \right),$$

with probability $1 - e^{-x}$, with the convention $\lambda_{m+1} = 0$. In particular, denoting $j_\rho = \max\{1 \leq k \leq m : \lambda_k \geq \rho\}$ we have that

$$T(f_{gc}^r, \rho) \leq \frac{16}{\left\lceil \min_{1 \leq k \leq m} \left\{ \sum_{i=1}^{k} \frac{n_i}{\lambda_i} \right\} \right\rceil} \tag{4.6}$$

with probability $1 - e^{-x}$, that is f_{gc}^r is adaptive in the range of ρ where the following extra conditions holds:

- *for all $k \leq j_\rho$,*

$$\lambda_k \geq \frac{K_3}{\sqrt{\sum_{i<k} \frac{n_i}{\lambda_i}}},$$

- *and also that*

$$\rho \geq \frac{K_3}{\sqrt{\sum_{i \leq j_\rho} \frac{n_i}{\lambda_i}}},$$

where $K_3 := 4\sqrt{2}K_2(1 + x + \log(m) + (x + \log(m)^{1/2})$.

Proof. Fix $0 \le \rho \le 1/2$, take $s_1, s_2 \in \mathbb{S}^1$ with $d(s_1, s_2) \ge \rho$ and define, as in the proof of Proposition 4.4,

$$j_{s_1, s_2} = \max \left\{ i \in [n] : \frac{\lambda_i}{2} \ge d(s_1, s_2) \right\}.$$

Lemma 7.2 ensures that for $1 \le j \le j_{s_1, s_2}$,

$$d(s_1 \bmod \lambda_j, s_2 \bmod \lambda_j) = d(s_1, s_2) \ge \rho.$$

Hence, by Proposition 4.1 we have that

$$\Delta_{s_1, s_2}^{f_{gc}^r} \ge \frac{1}{2} \sum_{i=1}^{j_{s_1, s_2}} \Delta_{s_1, s_2}^{f_{gc}^r, i} \ge \frac{1}{2} \sum_{i=1}^{j_{s_1, s_2}} \min_{u_1, u_2 \in \mathbb{S}^{\lambda_i} : d(u_1, u_2) \ge \rho} \Delta_{u_1, u_2}^{f_{gc}^r, i}. \tag{4.7}$$

Next, proceeding as in the proof of Proposition 3.8 one can show that

$$\min_{u_1, u_2 \in \mathbb{S}^{\lambda_i} : d(u_1, u_2) \ge \rho} \Delta_{u_1, u_2}^{f_{gc}^r, i} \ge \frac{n_i}{2} \left(\frac{\rho}{\lambda_i} - W_{\mathcal{C}_i} \right), \tag{4.8}$$

where $W_{\mathcal{C}_i}$ is the random variable given by

$$W_{\mathcal{C}_i} = \max_{C \in \mathcal{C}_i} \frac{1}{n_i} \sum_{j \in M_i} \left(\mathbf{1}_{C_i}(\xi_{ij}) - \mathbb{E}\left(\mathbf{1}_{C_i}(\xi_{ij}) \right) \right),$$

with the set $\mathcal{C}_i = \left\{ C_{u_1, u_2} : u_1, u_2 \in \mathbb{S}^{\lambda_i} : d(u_1, u_2) \ge \rho \right\}$ and $\xi_{ij} = (A_{ij}, B_{ij}) \in \mathbb{S}^{\lambda_i}$. Arguing for each $W_{\mathcal{C}_i}$ as it was done for $W_{\mathcal{C}}$ in the proof of Proposition 3.8, one deduces that if $\min_{1 \le i \le m}\{n_i\} \ge K_1$, then for all $x_1, \ldots, x_m \in \mathbb{R}_{>0}$,

$$\min_{u_1, u_2 \in \mathbb{S}^{\lambda_i} : d(u_1, u_2) \ge \rho} \Delta_{u_1, u_2}^{f_{gc}^r, i} \ge \frac{n_i}{2} \left(\frac{\rho}{\lambda_i} - n_i^{-1/2} K_2 \left(1 + x_i^{1/2} \left(1 + (x_i/n_i)^{1/2} \right) \right) \right) \text{ for all } i \le m, \tag{4.9}$$

with probability at least $1 - \sum_{i=1}^m \exp(-x_i)$. Therefore, combining inequalities (4.7), (4.8) and (4.9) it follows if $\min_{1 \le i \le m}\{n_i\} \ge K^2$ then for all $1 \le k \le m$ and all $x_1, \ldots, x_m \in \mathbb{R}_{>0}$

$$\min_{s_1, s_2 \in \mathbb{S}^1 : d(s_1, s_2) \ge \rho, j_{s_1, s_2} = k} \Delta_{s_1, s_2}^{f_{gc}^r}$$

$$\ge \frac{1}{2} \sum_{i=1}^k \frac{n_i}{2} \left(\frac{\max\{\rho, (\lambda_{k+1}/2)\}}{\lambda_i} - n_i^{-1/2} K \left(1 + x_i^{1/2} \left(1 + (x_i/n_i)^{1/2} \right) \right) \right)$$

$$\ge \frac{1}{2} \min_{1 \le k \le m} \sum_{i=1}^k \frac{n_i}{2} \left(\frac{\max\{\rho, (\lambda_{k+1}/2)\}}{\lambda_i} - n_i^{-1/2} K \left(1 + x_i^{1/2} \left(1 + (x_i/n_i)^{1/2} \right) \right) \right),$$

with probability larger or equal than $1 - \sum_{i=1}^m \exp(-x_i)$. Finally, since

$$\min_{s_1, s_2 \in \mathbb{S}^1 : d(s_1, s_2) \ge \rho} \Delta_{s_1, s_2}^{f_{gc}^r} = \min_{1 \le k \le M} \min_{s_1, s_2 \in \mathbb{S}^1 : d(s_1, s_2) \ge \rho, j_{s_1, s_2} = k} \Delta_{s_1, s_2}^{f_{gc}^r},$$

the result follows from the above inequality with $x_i = x + \log(m)$.

It remains to prove (4.6). Denote $\eta_x = (1 + x + \log(m) + (x + \log(m)^{1/2})$ and $\eta_{x,i} = (1 + (x + \log(m)^{1/2} + (x + \log(m)/n_i^{1/2})$, and observe that $\eta_x \ge \eta_{x,i}$ which ensures that

$$\sum_{i=1}^{k-1} \frac{n_i}{4} \left(\frac{\lambda_k}{2\lambda_i} - n_i^{-1/2} K_2 \eta_{x,i} \right) \ge \frac{1}{8} \sum_{i=1}^{k-1} \frac{n_i \lambda_k}{2\lambda_i} - \frac{K_2 \eta_x}{4} \sum_{i=1}^{k-1} \frac{n_i^{1/2}}{\lambda_i^{1/2}} \lambda_i^{1/2}.$$

By Cauchy-Schwarz inequality, we know that

$$\sum_{i=1}^{k-1} \frac{n_i^{1/2}}{\lambda_i^{1/2}} \lambda_i^{1/2} \leq \sqrt{\sum_{i=1}^{k-1} \frac{n_i}{\lambda_i}} \sqrt{\sum_{i=1}^{k-1} \lambda_i} \leq \sqrt{2} \sum_{i=1}^{k-1} \frac{n_i}{\lambda_i},$$

where in the second inequality we have used that $\lambda_i \leq 2^{-(i-1)}$ for each $1 \leq i \leq m$. As a consequence, we deduce that

$$\sum_{i=1}^{k-1} \frac{n_i}{4} \left(\frac{\lambda_k}{2\lambda_i} - n_i^{-1/2} K_2 \eta_{x,i} \right) \geq \frac{1}{8} \sum_{i=1}^{k-1} \frac{n_i}{\lambda_i} \left(\lambda_k - \frac{2\sqrt{2} K_2 \eta_x}{\sqrt{\sum_{i=1}^{k-1} n_i/\lambda_i}} \right) \geq \frac{1}{16} \sum_{i=1}^{k-1} \frac{n_i}{\lambda_i} \lambda_k,$$

Similarly, one can show that if the second extra condition holds then we also have that

$$\sum_{i=1}^{j_\rho} \frac{n_i}{4} \left(\frac{\rho}{2\lambda_i} - n_i^{-1/2} K_2 \eta_{x,i} \right) \geq \frac{1}{16} \sum_{i=1}^{j_\rho} \left(\frac{n_i \rho}{\lambda_i} \right). \tag{4.10}$$

From the last to inequalities above it is easy to deduce (4.6), concluding the proof of the result. □

4.4 What can we say about the general class $\bar{\mathcal{G}}$?

The rate

$$\frac{1}{\min_{1 \leq k \leq m} \sum_{j=1}^{k} \left\lfloor \frac{2n_j}{\lambda_j} \max\left\{ \frac{\lambda_{k+1}}{2}, \rho \right\} \right\rfloor},$$

is very difficult to understand. What would be the best choice of n_i's and λ_i's ?

Minimal range of detection

Example 4.7 (Extreme Dyadic code)**.** In the extreme dyadic code f_d, we have $m = n$ modules $M_1, \ldots, M_n$ where the i-th module $M_i = \{i\}$ has period $\lambda_i = (1/2)^{i-1}$ and whose points $a_{ii} := a_i, b_{ii} := b_i \in \mathbb{S}^{\lambda_i}$ are such that $\theta_{a_i} = 0$ and $\theta_{b_i} = (1/2)^i$. In the sequel, let us denote $f_{d,ii} = f_{d,i}$ so that $f_d = (f_{d1}, \ldots, f_{dn})$.

Proposition 4.8. *Let f_d be the extreme dyadic code. For any $\rho \geq 1/2^n$, we have $T(f_d, \rho) = 1$.*

Proof. First, we will show that $T(f_d, \rho, \alpha) \leq 1$. To that end, we need to show that for all $s_1, s_2 \in [0, 1)$ such that $d(s_1, s_2) \geq \rho$, we have $\max\left\{ |I_{s_1}^{f_d} \setminus I_{s_2}^{f_d}|, |I_{s_2}^{f_d} \setminus I_{s_1}^{f_d}| \right\} \geq 1$. Notice that by Lemma 7.4, we can write $(\theta_{s_i}) = \sum_{j=1}^{\infty} (\theta_{s_i})_j / 2^j, i \in \{1, 2\}$.

Claim. There exists $j \in [n]$ such that $(\theta_{s_1})_j \neq (\theta_{s_2})_j$.

Assuming that the claim is true, we can then use again Lemma 7.4 to conclude that $f_{dj}(s_1) \neq f_{dj}(s_2)$ implying that $\max\left\{ |I_{s_1}^{f_d} \setminus I_{s_2}^{f_d}|, |I_{s_2}^{f_d} \setminus I_{s_1}^{f_d}| \right\} \geq 1$.

We now prove the Claim. We argue by contradiction. Suppose that $(\theta_{s_1})_j = (\theta_{s_2})_j$ for all $j \in [n]$. Suppose also that $\theta_{s_1} \leq \theta_{s_2} \leq \theta_{s_1} + 1/2$ (if this is not the case we change s_2 by s_1). In this case,

$$1/2^n \leq \rho \leq d(s_1, s_2) = |\theta_{s_1} - \theta_{s_2}| = \sum_{j=n+1}^{\infty} |((\theta_{s_1})_j - (\theta_{s_2})_j)|/2^j \leq 1/2^n = \lambda_n/2,$$

and Lemma 7.2 implies that

$$d(s_1 \bmod \lambda_n, s_2 \bmod \lambda_n) = \frac{1}{2^n}.$$

This implies that $f_{dn}(s_1) = f_{dn}(s_1 \bmod \lambda_n) \neq f_{dn}(s_2 \bmod \lambda_n) = f_{dn}(s_2)$ and we can apply Lemma 7.4 to deduce that $(\theta_{s_1})_n \neq (\theta_{s_2})_n$, which is absurd. This concludes the proof of the Claim.

To conclude the proof, observe that by taking $\theta_{s_1} = 0$ and $\theta_{s_2} = 1/2$, we have $(\theta_{s_1})_j = 0$ for all $j \geq 1$, $(\theta_{s_2})_1 = 1$ and $(\theta_{s_2})_j = 0$ for all $j \geq 2$. By Lemma 7.4 we have that $f_{d1}(s_1) \neq f_{d1}(s_2)$, $f_{dj}(s_1) = f_{dj}(s_2)$ for all $2 \leq j \leq n$ and $d(s_1, s_2) = |\theta_{s_1} - \theta_{s_2}| = 1/2 \geq \rho$. Hence, $T(f_d, \rho) = 1$ and the result follows.

$\square$

This result needs to be put in perspective with Proposition 2.2 : the extreme dyadic code is able to reach the best precision that one can have, that is $1/2^n$, but it is unable to reach a faster minimax discrimination time when ρ increases. So either (whatever the code in $\bar{\mathcal{G}}$) one cannot achieve a faster rate than 1 or it means that this extreme dyadic code cannot be adaptive. The following example shows a faster rate for larger ρ, which means that we are in the second case.

Example 4.9. (Grid cells - adaptive balanced) For a given $1 \leq m \leq n$, let $\lambda_i = 2^{-(i-1)}$ and $n_i = \lfloor n/m \rfloor$ for $1 \leq i \leq m-1$, and $\lambda_m = 2^{-(m-1)}$ and $n_m = \lfloor n/m \rfloor + n \bmod m$. We call balanced grid cells code the class $\mathcal{G}_{b,m} := \mathcal{G}((n_i, \lambda_i)_{i=1,\ldots,m})$ corresponding to these choices of λ_i's and n_i's. Note that each code in this class has the same number of neurons per module (except the last one). In what follows, let us denote $g_{gc}^{b,m}$ the code g_{gc} defined in Example 4.3 belonging to the class of balanced grid cells code $\mathcal{G}_{b,m}$.

Proposition 4.10. *Let $1 \leq m \leq n$ such that $n \geq 2m$. For any ρ such that $1/2^m \leq \rho \leq 1/2$,*

$$\frac{1}{3\lfloor n/m \rfloor} \leq T(\mathcal{G}_{b,m}, \rho) \leq T_{min}(g_{gc}^{b,m}, \rho) \leq \frac{16}{\lfloor n/m \rfloor},$$

that is, the code $g_{gc}^{b,m}$ is adaptive in the range $\rho \geq 1/2^m$ in the class $\mathcal{G}_{b,m}$.

Proof. Let us first compute the corresponding lower bound for $T(\mathcal{G}_{b,m})$ provided by Theorem 4.2. Notice that for all $2 \leq k \leq j_\rho$ we have that

$$\sum_{i=1}^{k-1} \frac{n_i \max\{\rho, \lambda_k\}}{\lambda_i} = \sum_{i=1}^{k-1} \frac{n_i \lambda_k}{\lambda_i} = \lfloor n/m \rfloor \sum_{i=1}^{k-1} 2^{-i} = \lfloor n/m \rfloor \left(1 - 2^{-(k-1)}\right) := x_{k-1}.$$

Moreover, one can check that

$$\sum_{i=1}^{j_\rho} \frac{n_i \max\{\rho, \lambda_{j_\rho+1}\}}{\lambda_i} \geq \rho \lfloor n/m \rfloor \sum_{i=1}^{j_\rho} \frac{1}{\lambda_i}$$

$$= \rho \lfloor n/m \rfloor (2^{j_\rho} - 1) := x_{j_\rho}.$$

Since $x_{k+1} - x_k = 2\lfloor n/m \rfloor 2^{-(k-2)} > 0$ for $1 \leq k \leq j_\rho - 2$, it follows that

$$\min\left\{ \min_{2 \leq k \leq j_\rho} \sum_{i=1}^{k-1} \frac{n_i \max\{\rho, \lambda_k\}}{\lambda_i}, \sum_{i=1}^{j_\rho} \frac{n_i \max\{\rho, \lambda_{j_\rho+1}\}}{\lambda_i} \right\} = \min\{x_1, x_{j_\rho}\}.$$

Now, for any $1/2^m \leq \rho \leq 1/2$ it holds that $2^{-j_\rho} = \lambda_{j_\rho+1} \leq \rho \leq \lambda_{j_\rho}$ so that $2\rho(2^{j_\rho} - 1) \geq 2(1 - \rho) \geq 2(1/2) = 1$, implying that $x_1 \leq x_{j_\rho}$, that is $\min\{x_1, x_{j_\rho}\} = x_1 = (\lfloor n/m \rfloor)/2$. Hence Theorem 4.2 implies that

$$T(\mathcal{G}_{b,m}, \rho) \geq \frac{1}{3\lfloor n/m \rfloor}. \tag{4.11}$$

Next, we will compute the upper bound for $T_{min}(g_{gc}^{m,b}, \rho)$ using Proposition 4.4. First, observe that for all $2 \leq k \leq j_\rho$,

$$y_{k-1} := \sum_{j=1}^{k-1} \left\lfloor \frac{n_j}{\lambda_j} \max\{\lambda_k, \rho\} \right\rfloor = \left\lfloor \lfloor n/m \rfloor \frac{1}{2} \right\rfloor + \ldots + \left\lfloor \lfloor n/m \rfloor \frac{1}{2^{k-1}} \right\rfloor$$

so that $y_1 \leq y_2 < \ldots \leq y_{j_\rho - 1}$. Let,

$$y_{j_\rho} := \sum_{j=1}^{j_\rho} \left\lfloor \frac{n_j}{\lambda_j} \max\{\lambda_{j_\rho + 1}, \rho\} \right\rfloor = \sum_{j=0}^{j_\rho - 1} \left\lfloor \lfloor n/m \rfloor \rho 2^j \right\rfloor.$$

implying that

$$\min_{1 \leq k \leq m} \sum_{j=1}^{k} \left\lfloor \frac{n_j}{\lambda_j} \max\{\lambda_{k+1}, \rho\} \right\rfloor = \min\{y_1, y_{j_\rho}\}.$$

Since, $\lfloor n/m \rfloor / 2 \geq 1$ we have that $y_1 \geq \lfloor n/m \rfloor / 4$ (here we use that $\lfloor x \rfloor \geq x/2$ when $x \geq 1$). Moreover,

$$y_{j_\rho} = \sum_{j=0}^{j_\rho - 1} \left\lfloor \lfloor n/m \rfloor \rho 2^j \right\rfloor$$

$$= \sum_{j=0: \lfloor n/m \rfloor \rho 2^j \geq 1}^{j_\rho - 1} \left\lfloor \lfloor n/m \rfloor \rho 2^j \right\rfloor > (\lfloor n/m \rfloor / 2) \sum_{j=0: \lfloor n/m \rfloor \rho 2^j \geq 1}^{j_\rho - 1} \rho 2^j,$$

so that if we denote $j_{min} = \min\{0 \leq j \leq j_\rho - 1 :: \lfloor n/m \rfloor \rho 2^j \geq 1\}$, then

$$y_{j_\rho} > (\lfloor n/m \rfloor / 2) \sum_{j=j_{min}}^{j_\rho - 1} \rho 2^j = (\lfloor n/m \rfloor / 2) 2^{j_{min}} \rho (2^{j_\rho - j_{min}} - 1).$$

Finally, since $\rho 2^{j_\rho} \geq 2$ and since $\lfloor n/m \rfloor \geq 2$, one can check that $j_{min} \leq j_\rho - 2$. As a consequence,

$$y_{j_\rho} > (\lfloor n/m \rfloor / 2) 2^{j_{min}} \rho (2^{j_\rho - j_{min}} - 1)$$
$$\geq (\lfloor n/m \rfloor / 2)(2^{j_\rho - j_{min} + 1} / 2^{j_\rho} - \rho) 2^{j_{min}} = (\lfloor n/m \rfloor / 2)(2 - \rho 2^{j_{min}}) \geq \lfloor n/m \rfloor (1 - 1/4)$$
$$> \lfloor n/m \rfloor / 4.$$

Hence, we have shown that if $\lfloor n/m \rfloor \geq 2$, then $\min\{y_1, y_{j_\rho}\} \geq (\lfloor n/m \rfloor / 4)$ and Proposition 4.4 ensures that

$$T(\mathcal{G}_{b,m}, \rho) \leq T_{min}(g_{gc}^{b,m}, \rho) \leq \frac{16}{\lfloor n/m \rfloor}.$$

The result follows from (4.11) and the inequality above. $\qquad\square$

This result about adaptivity is very interesting. Indeed it tells us that whatever the balanced adaptive code that we use with m modules, the minimax discrimination time does not vary as a function of ρ and in particular does not decrease when the distance between two stimuli increases. It stays constant at m/n up to multiplicative constants. In the extreme case with $m = n$ modules, we recover the extreme dyadic code whose minimax discrimination time is 1 whatever the distance. On the other hand, and as a corollary of Proposition (4.10), we have the following result.

Corollary 4.11. *For all $2^{-n/2} \leq \rho \leq 1/2$, let $m = \lfloor \log_2(\rho^{-1}) \rfloor$. Then the code $g_{gc}^{b,m}$ satisfies*

$$\frac{1}{6} \left\lfloor \frac{\log_2(\rho^{-1})}{n} \right\rfloor \leq \frac{1}{3 \lfloor n/m \rfloor} \leq T(\mathcal{G}_{b,m}, \rho) \leq T_{min}(g_{gc}^{b,m}, \rho) \leq 32 \left\lceil \frac{\log_2(\rho^{-1})}{n} \right\rceil.$$

The proof of Corollary 4.11 is straightforward. Note that this result shows that if we know in advance at which distance ρ one needs to detect, one can use a balanced adaptive code with $m = \lfloor \log_2(\rho^{-1}) \rfloor$ modules to reach a rate much faster than the place cells codes. Indeed, $\log_2(\rho^{-1}) \ll \rho^{-1}$ when ρ is small. On the other hand, once m is fixed Proposition 4.10 tells us that the minimax discrimination time is then constant and cannot decreases when ρ increases as $\log_2(\rho^{-1})/n$. In this sense, balanced grid cell codes cannot be adaptive. The question remains open for unbalanced grid cells codes.

Summary of the results on grid cells codes

1. One can compute the minimax discrimination rate on $\mathcal{G}((n_i, \lambda_i)_{i=1,\ldots,m})$.

2. In general, grid cells code obtained with the adaptive code of Example 3 on each module are able to reach this minimax discrimination rate, for every ρ up to a $\log_2(\rho^{-1})$ factor but on a restricted range of ρ.

3. The random codes do not lose this extra $\log_2(\rho^{-1})$ factor but they have a different range of ρ.

4. Extreme dyadic codes can reach the best possible precision in 2^{-n}. However, their discrimination rate cannot be faster for larger ρ.

5. Balanced grid cells code obtained with the adaptive code of Example 3 on each module with number of modules $m = \log_2(\rho^{-1})$ are able to reach the rate $\log_2(\rho^{-1})/n$, which is faster than the corresponding minimax rate for place cells code. However, balanced grid cells code cannot go faster when the distance between the stimuli increases.

6. In particular, we do not know the minimax discrimination rate over the general class $\bar{\mathcal{G}}$ and we do not know if adaptivity is possible there. However, we do know that the extreme dyadic code which achieves the precision 2^{-n} or the balanced grid cells code cannot be adaptive.

---

5   Numerical illustration

We have simulated what happens for $n = 100$ cells with firing rate $\mu = 30$. We used 5 different configurations, as indicated below.

- Place cells - adaptive. That is the code f is given by Example 3
- Place cells - random. The code f is picked at random as in Example 4
- Grid cells - adaptive balanced. It consists in 20 modules with $\lambda_i = 2^{-(i-1)}$. All the 20 modules have the same $n_i = 5$. Inside each module, the code is taken as Example 3.
- Grid cells - adaptive decreasing. It consists in 20 modules with $\lambda_i = 2^{-(i-1)}$, but the vector of n_i is given by $(15, 13, 11, 10, 8, 7, 6, 5, 4, 4, 3, 3, 2, 2, 1, 1, 1, 1, 1)$. Inside each module, the code is taken as Example 3.
- Grid cells - random balanced. It consists in 20 modules with $\lambda_i = 2^{-(i-1)}$. All the 20 modules have the same $n_i = 5$. Inside each module, the code is taken at random as Example 4.

We picked $s = 1/3$, which is not in any of the periods of the different modules. We also picked $s' = s + \rho$ with ρ in a grid between 2^{-21} and 0.5.

For various possible T in a grid from 0.001 to 20, we simulated 5000 times the test Ψ given by (2.4) where we choose $s_1 = s$ and $s_2 = s'$ if $|I_s^f \setminus I_{s'}^f| \geq |I_{s'}^f \setminus I_s^f|$ and the reverse if this is not the case. In each case, we found $T_{min}(f, s, s', \alpha)$ the smallest time in the grid that gives an error less than α (the error is evaluated by Monte Carlo method on the 5000 simulations). In Figure 3 on the left, we see $T_{min}(f, s, s', \alpha)$ as a function of
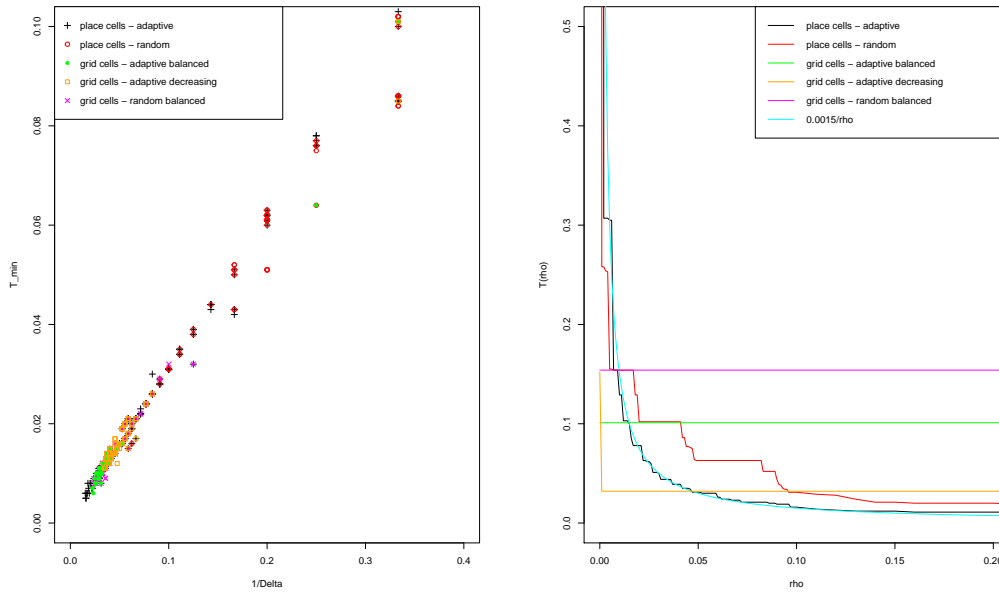
Figure 3: Discrimination time as a function of Δ_{s_1,s_2}^f and ρ. On the left, $T_{min}(f,s_1,s_2,\alpha)$ as a function of $1/\Delta_{s,s'}^f$ for the 5 different codes. On the right, $T_{min}(f,\rho,\alpha)$ as a function of ρ for the 5 different codes.

$1/\Delta_{s,s'}^f$, with $\Delta_{s,s'}^f = \max\left\{|I_s^f \setminus I_{s'}^f|, |I_{s'}^f \setminus I_s^f|\right\}$. As derived in (2.5), we also see in the simulations that the same constant of proportionality holds whatever the code (this constant depending only on μ and α) and that it was legitimate to study directly $1/\Delta_{s,s'}^f$.

It was not possible to compute $T(f,\rho)$ as defined by (2.6), so we used the following quantity as a proxy:

$$T(f,s,\rho,\alpha) = \max_{\rho' \geq \rho \text{ in the grid}} T_{min}(f,s,s+\rho',\alpha),$$

for $s = 1/3$. In Figure 3 on the right, we plotted $T(f,s,\rho,\alpha)$ as a function of ρ. We see that place cells - adaptive are indeed following a curve in $1/\rho$ as expected and that place cells - random have a similar behavior. In particular, with $n = 100$ they cannot detect a ρ of order $1/(2n) = 0.005$. On the other hand, all grid cells can reach this range, because they can reach a much smaller range, at least 2^{-20}. Besides, grid cells - (random or adaptive) balanced cannot have a time of detection which decreases when ρ increases. Hence there is a point where the place cells system is quicker in discrimination time than the grid cells system.

We tried a non balanced version of the grid cells (grid cells - adaptive decreasing), to make the discrimination time decreasing in ρ. However, as one can see on Figure 3 on the right, even if the discrimination time becomes decreasing in ρ, it reaches a plateau like behavior (up to logarithmic behavior that cannot be seen on the curve). Therefore, even this non balanced version of the grid cells is still slower than the place cells system.

6 Conclusion, discussion and Perspectives

We have adopted a new point of view on rate coding : the testing point of view and use it to enhance differences between a place cell system and a grid cell system with the same number of neurons.

On the testing point of view. The testing point of view is complementary to the estimation/information theory point of view developed originally by [4] for place cells or other cells with simple receptor field. Indeed, for place cells, Brunel and Nadal proved in a similar framework as the random code of the present work that the Fisher information is proportional to n, the number of cells. The only main difference is that they used triangular codes f instead of step functions with two parameters. If the Fisher information is proportional to n, it means that the standard deviation of the best estimator of the stimulus s is in $n^{-1/2}$. We have proved that a similar random place cell system can discriminate between two positions as long as they are at least at distance of order $n^{-1/2}$. In this sense, both frameworks seem consistent. However, we have also proved (see Example 3) that another place cells code achieves a much smaller precision in $1/n$. Moreover the testing set-up allows us to capture an interesting phenomenon, which should take place in practice and might be tested via experiments. When two stimuli are very different, the discrimination time between both should be smaller.

On the place cell system. We have shown that adaptivity in terms of ρ (the distance between two stimuli) is possible. It means that not only certain codes have the ability to discriminate faster when the stimuli are further away but it also means that the rate $1/(n\rho)$ is up to constant not improvable by a place cell system. This is reached not only by very particular codes such as the one of Example 3 but also by random codes, for $\rho \geq c/\sqrt{n}$, for some positive constant c.

On the grid cell system. As shown already by the study of the Fisher information [9, 22], we can prove that grid cells can reach a precision that is much smaller than the one of place cells and which is exponentially decreasing with n. We have also shown that the rate 2^{-n} for the discrimination time is in fact an absolute lower bound for all kind of codes with only two values and that the grid cells in this sense are the ones able to achieve the smallest precision not only with respect to place cells, but with respect to any code. We have also been able to derive upper and lower bound on the minimal discrimination time for grid cells with a given number of cells per module and given period for each module. In particular, the distribution of the cells inside a module is derived from their place cell equivalent and a random uniform distribution inside a module leads to the best rate up to constant as soon as ρ is large enough. This is coherent with the experiment of [24] that has found no structure in the distribution of the center of the receptor field inside a given module. Finally, for a fix ρ, particular balanced grid cells code are able to achieve the discrimination rate of $\log(1/\rho)/n$ which is much faster than place cells, but we have not been able to find a fix grid cells code whose discrimination rate would decrease when ρ increases in an interesting range (that is past the detection range of the first module). Informally once $\rho < 1/(2n_1)$, it seems that all codes are limited by what happens in the first module, that is all discrimination rates seem to be equal to $1/(2n_1)$. However, we have not been able to rigorously prove this fact.

Place cells versus grid cells. As a consequence, and contrary to place cells, we have not been able to find a general grid cell adaptive code that would achieve a minimax discrimination rate that can only be smaller or equal to $\log(1/\rho)/n$ thanks to our results. Our tentative non balanced codes as shown in the simulation were not satisfying. We do not know if adaptation is possible for general grid cells. If it is not possible, then there is definitely an advantage to have both systems (grid cells and place cells) at the same time. The simulations show a compromise that can be made in using both systems at

once (see Figure 3 on the right). Indeed, grid cells can go at much smaller precision that place cells but are much slower when ρ increase than place cells. In this sense, having a combination of both systems would allow a fast reaction time when stimuli are far away and a good reaction time even if the stimuli are very close.

Limitations and Open problems. We restricted ourselves to one dimensional, periodic stimuli (or circular maze). We do not know at the moment how to generalize it to larger dimension, especially if we want grid cells to have an hexagonal pattern. Moreover, [24] have proved that the periods progress in a geometric manner (see also [27] for similar geometric progressions in 1d) but that the ratio is not an integer and so far our method is relying too strongly on periodicity to allow that. To go beyond this, we would need to consider stimuli with boundaries effect and if the adaptation to boundaries for grid cells have been described [13] modeling precisely the boundary effect from a mathematical point of view is not straightforward. Also one might want to add more realistic rate functions $f(s)$ that step functions, but we do not think this would massively change the rates we found as long as the shape of $f(s)$ is not authorized to vary much. Finally, even in the 1d circular case, the problem of adaptivity and even the computation of the minimax discrimination rate of the general grid cells codes remains open. It means that we do not know what is the best choice in terms of discrimination rate for the number of cells per module, the scales or even the number of modules.

7 Auxiliary Results

In this section, we collect all auxiliary results used throughout the paper. We start with the following proposition.

Lemma 7.1. *Let X be a Poisson random variable with parameter $\theta > 0$. For any $x \geq 0$,*

$$\mathbb{P}\left(X \geq \theta(1+x)\right) \leq \exp\left\{-\frac{\theta x^2}{2(1+x/3)}\right\}, \tag{7.1}$$

and for any $0 \leq x \leq \theta$,

$$\mathbb{P}\left(X \leq \theta - x\right) \leq \exp\left\{-\frac{x^2}{2\theta}\right\}. \tag{7.2}$$

Proof. Let $\psi(\lambda) = e^\lambda - \lambda - 1$ for $\lambda \in \mathbb{R}$ and notice that $\mathbb{E}(e^{\lambda(X-\theta)}) = e^{\theta\psi(\lambda)}$. One can check that for all $0 \leq \lambda < 3$, we have

$$\psi(\lambda) \leq \frac{\lambda^2}{2(1-\lambda/3)}.$$

so that for any $t > 0$,

$$\mathbb{P}(X \geq \theta + t) \leq \exp\left\{-\sup_{\lambda \in (0,3)}\left(t\lambda - \frac{\lambda^2\theta}{2(1-\lambda/3)}\right)\right\} = \exp\{-9\theta h_1(t/3\theta)\},$$

where $h_1(x) = 1 + x - \sqrt{1+2x}, x > 0$. Since $h_1(x) \geq x^2/(2(1+x))$ it follows that for any $t > 0$,

$$\mathbb{P}(X \geq \theta + t) \leq \exp\left\{-\theta\frac{(t/\theta)^2}{2(1+(t/3\theta))}\right\}.$$

Taking $t = \theta x$ we obtain (7.1).

To prove inequality (7.2), one can use the fact that $\psi(\lambda) \leq \lambda^2/2$ for $\lambda \leq 0$, to deduce for any $0 \leq x \leq \theta$,

$$\mathbb{P}(X \leq \theta - x) \leq \exp\{-\sup_{\lambda > 0}(\lambda x - \theta\lambda^2/2)\} = \exp\{-x^2/2\theta\},$$

concluding the proof. □

Lemma 7.2. *Let* $0 < \lambda \leq 1$. *For all* $s_1, s_2 \in \mathbb{S}^1$ *such that* $d(s_1, s_2) \leq \lambda/2$, *it holds that* $d(s_1, s_2) = d(s_1 \bmod \lambda, s_2 \bmod \lambda)$.

Proof. It is sufficient to analyse the case $s_1, s_2 \in \mathbb{S}^1$ are such that $0 \leq \theta_{s_1} \leq 1/2$, $\theta_{s_1} \leq \theta_{s_2} \leq \theta_{s_1} + 1/2$ and $d(s_1, s_2) \leq \lambda/2$. Write $\theta_{s_1} = k\lambda + r_1$ where $0 \leq r_1 = \theta_{s_1} \bmod \lambda < \lambda$.

Suppose that $r_1 + \lambda/2 < \lambda$. In this case, we must that $r_1 < \lambda/2$ and $\theta_{s_2} = k\lambda + r_2$ with $r_1 \leq r_2 = \theta_{s_2} \bmod \lambda \leq r_1 + \lambda/2$. Hence

$$\theta_{s_1 \bmod \lambda} = \theta_{s_1} \bmod \lambda = r_1 \leq r_2 = \theta_{s_2} \bmod \lambda = \theta_{s_2 \bmod \lambda} \leq r_1 + \lambda/2$$

so that

$$d(s_1 \bmod \lambda, s_2 \bmod \lambda) = (\theta_{s_2 \bmod \lambda} - \theta_{s_1 \bmod \lambda}) = (r_2 - r_1) = d(s_1, s_2).$$

Now, it remains to consider the case $r_1 + \lambda/2 \geq \lambda$. In this case, we have that $r_1 \geq \lambda/2$. If $\theta_{s_2} = k\lambda + r_2$ then we must have $r_1 \leq r_2 < \lambda$ so that $d(s_1 \bmod \lambda, s_2 \bmod \lambda) = (\theta_{s_2 \bmod \lambda} - \theta_{s_1 \bmod \lambda}) = r_2 - r_1 = d(s_1, s_2)$. Otherwise, $\theta_{s_2} = (k+1)\lambda + r_2$ where $r_2 \leq r_1 + \lambda/2 - \lambda = r_1 - \lambda/2$ so that $d(s_1 \bmod \lambda, s_2 \bmod \lambda) = r_2 + \lambda - r_1 = d(s_1, s_2)$ and the result follows. □

Lemma 7.3. *Let* $0 < \sigma_1 \leq \sigma_2 \leq 1/2$ *such that* $\sigma_2 = p\sigma_1$ *and* $1 = q\sigma_1$ *where* $p, q \in \mathbb{N}$ *are such that* $q \geq p$. *For all* $s_1, s_2 \in \mathbb{S}^1$ *such that* $d(s_1, s_2) = \sigma_2$, *it holds that* $s_1 \bmod \sigma_1 = s_2 \bmod \sigma_1$.

Proof. Suppose that $\theta_1 \leq \theta_2$ and write $\theta_{s_1} = k\sigma_1 + r$ where $0 \leq r = \theta_{s_1} \bmod \sigma_1 < \sigma_1$. If $0 \leq \theta_{s_2} \leq \theta_{s_1} + 1/2$ and $d(s_2, s_1) = \sigma_2$, then $\theta_{s_2} - \theta_{s_1} = d(s_2, s_1) = \sigma_2 = p\sigma_1$ so that $\theta_{s_2} = (p+k)\sigma_1 + r$, implying that $\theta_{s_2} \bmod \sigma_1 = r = \theta_{s_1} \bmod \sigma_1$. Since $\theta_{s_2 \bmod \sigma_1} = \theta_{s_2} \bmod \sigma_1 = \theta_{s_1} \bmod \sigma_1 = \theta_{s_1 \bmod \sigma_1}$ it follows that $s_1 \bmod \sigma_1 = s_2 \bmod \sigma_1$.

Now, if $1/2 + \theta_{s_1} \leq \theta_{s_2} < 1$ and $d(s_2, s_1) = \sigma_2$, then $1 - (\theta_{s_2} - \theta_{s_1}) = d(s_2, s_1) = \sigma_2 = p\sigma_1$ so that $\theta_{s_2} = 1 + (k-p)\sigma_1 + r = (q+k-p)\sigma_1 + r$ and once again $\theta_{s_2} \bmod \sigma_1 = r = \theta_{s_1} \bmod \sigma_1$, and the result follows. □

Lemma 7.4. *For each* $x \in [0,1)$, *let* $x_1 = \lfloor 2x \rfloor$ *and* $x_j = \left\lfloor 2^j \left(x - \sum_{k=1}^{j-1} x_k/2^k \right) \right\rfloor$ *for* $j \geq 2$. *Then* $x_j \in \{0,1\}$ *for all* $j \in \mathbb{N}_{>0}$, $x = \sum_{j=1}^{\infty} x_j/2^j$ *and* $0 \leq \sum_{k=j}^{\infty} x_k/x^k < 1/2^{j-1}$ *for any* $j \in \mathbb{N}_{>0}$. *In particular, if* $f_d = (f_{d,1}, \ldots, f_{d,n})$ *denotes the dyadic code defined in Example 4.7, then for any* $i \in [n]$, *we have* $f_{di}(s) = \mu$ *if and only if* $(\theta_s)_i = 0$.

Proof. It is clear that $x_1 \in \{0,1\}$. To show that $x_j \in \{0,1\}$, first observe that denoting $u := x - \sum_{k=1}^{j-1} x_k/2^k$, we have $u - \lfloor 2^j u \rfloor/2^j \in [0, 1/2^j)$. Since $\lfloor 2^j u \rfloor = x_j$, it follows then that $x - \sum_{k=1}^{j} x_k/2^k = x - \lfloor 2^j u \rfloor/2^j \in [0, 1/2^j)$. This implies that $x = \sum_{j=1}^{\infty} x_k/2^k$ and also that $0 \leq 2^{j+1}(x - \sum_{k=1}^{j} x_k/2^k) = x - \lfloor 2^j u \rfloor/2^j) < 2$ which ensures that $x_{j+1} \in \{0,1\}$ and also that $\sum_{k=j+1}^{\infty} x_k/2^k < 2^{-j}$.

It remains to show that $f_{di}(s) = \lambda$ if and only if $s_i = 0$. To that end, first observe that $f_{di}(s) = f_{di}(s \bmod \lambda_i) = \mu$ if and only if $\theta_s - \lfloor 2^{i-1}\theta_s \rfloor/2^{i-1} = \theta_s \bmod \lambda_i = \theta_{s \bmod \lambda_i} < 1/2^i$. Now, since $\theta_s - \lfloor 2^{i-1}\theta_s \rfloor/2^{i-1} = \sum_{j=i+1}^{\infty} (\theta_s)_j/2^j + (\theta_s)_i/2^i$, it follows that $f_{di}(s) = \mu$ implies that $\sum_{j=i+1}^{\infty}(\theta_s)_j/2^j + (\theta_s)_i/2^i < 1/2^i$, forcing $(\theta_s)_i = 0$. On the other hand, if $(\theta_s)_i = 0$ then $\sum_{j=i+1}^{\infty}(\theta_s)_j/2^j + (\theta_s)_i/2^i = \sum_{j=i+1}^{\infty}(\theta_s)_j/2^j < 1/2^i$, where the inequality holds by the first part of the proof. Recalling that $\sum_{j=i+1}^{\infty}(\theta_s)_j/2^j + (\theta_s)_i/2^i = \theta \bmod \lambda_i = \theta_{s \bmod \lambda_i}$, it follows that $(\theta_s)_i = 0$ implies that $f_{di}(s) = \mu$, concluding the proof. □

Lemma 7.5. *Let A, B be independent and uniformly distributed points in $\mathbb{S}^\lambda$ with $\lambda \in \mathbb{R}_{>0}$. For any pair $s_1, s_2 \in \mathbb{S}^\lambda$, we have that*

$$\mathbb{P}\left(s_1, s_2 \in [\![A, B[\![\right) = \lambda^{-2}\left(\frac{\lambda^2}{2} - |\theta_{s_2} - \theta_{s_1}|\left(\lambda - |\theta_{s_2} - \theta_{s_1}|\right)\right).$$

Proof. Suppose that $\theta_{s_1} \leq \theta_{s_2}$ and note that

$$\mathbb{P}\left(s_1, s_2 \in [\![A, B[\![\right) = \mathbb{P}\left(\theta_A \leq \theta_{s_1}, \theta_{s_2} < \theta_B < \lambda\right) + \mathbb{P}\left(\theta_{s_1} \leq \theta_A < \theta_{s_2}, \theta_{s_1} < \theta_B < \theta_A\right)$$
$$+ \mathbb{P}\left(\theta_B < \theta_A \leq \theta_{s_1}\right) + \mathbb{P}\left(\theta_{s_2} \leq \theta_A < \lambda, \theta_{s_2} < \theta_B < \theta_A\right).$$

We now compute each one of the probabilities appearing in the right-hand side of the equality above. Since θ_A and θ_B are independent and uniformly distributed on $[0, \lambda)$, one can check that

$$\begin{cases} \mathbb{P}\left(\theta_A \leq \theta_{s_1}, \theta_{s_2} < \theta_B < \lambda\right) = \frac{\theta_{s_1}}{\lambda}\frac{(\lambda - \theta_{s_2})}{\lambda}, \\ \mathbb{P}\left(\theta_B < \theta_A \leq \theta_{s_1}\right) = \frac{1}{\lambda^2}\int_0^{\theta_{s_1}}(\theta_{s_1} - \theta_b)d\theta_b = \theta_{s_1}^2/(2\lambda^2), \\ \mathbb{P}\left(\theta_{s_1} < \theta_B < \theta_A < \theta_{s_2}\right) = \frac{1}{\lambda^2}\int_{\theta_{s_1}}^{\theta_{s_2}}(\theta_{s_2} - \theta_b)d\theta_b = (\theta_{s_2} - \theta_{s_1})^2/(2\lambda^2), \\ \mathbb{P}\left(\theta_{s_2} < \theta_B < \theta_A < \lambda,\right) = \frac{1}{\lambda^2}\int_{\theta_{s_2}}^{\lambda}(\lambda - \theta_b)d\theta_b = (\lambda - \theta_{s_2})^2/(2\lambda^2.) \end{cases}$$

Summing these fours terms, we obtain that

$$\mathbb{P}\left(s_1, s_2 \in [\![A, B[\![\right) = \lambda^{-2}\left(\theta_{s_1}(\lambda - \theta_{s_2}) + \theta_{s_1}^2/2 + (\theta_{s_2} - \theta_{s_1})/2 + (\lambda - \theta_{s_2})^2/2\right)$$
$$= \lambda^{-2}\left(\lambda^2/2 - \lambda(\theta_{s_2} - \theta_{s_1}) + (\theta_{s_2} - \theta_{s_1})^2\right)$$
$$= \lambda^{-2}\left(\lambda^2/2 - |\theta_{s_2} - \theta_{s_1}|(\lambda - |\theta_{s_2} - \theta_{s_1}|)\right),$$

and the result follows. $\square$

References

[1] D. Aronov, R. Nevers, and D.W. Tank, *Mapping of a non-spatial dimension by the hippocampal–entorhinal circuit*, Nature **543** (2017), 719–722, doi.

[2] Y. Baraud, S. Huet, and B. Laurent, *Adaptive tests of linear hypotheses by model selection*, Ann. Stat. **31** (2003), no. 1, 225–251, doi. MR-1962505

[3] E. N. Brown, R. Barbieri, V. Ventura, R. E. Kass, and L.M. Frank, *The time-rescaling theorem and its application to neural spike train data analysis*, Neural Comput. **14** (2002), no. 2, 325–346, doi.

[4] N. Brunel and J.-P. Nadal, *Mutual information, Fisher information and Population Coding*, Neural Comput. **10** (1998), no. 7, 1731–1757, doi.

[5] G. Buzsáki and E.I. Moser, *Memory, navigation and theta rhythm in the hippocampal-entorhinal system*, Nat. Neurosci. **16** (2013), no. 2, 130–138, doi.

[6] D. Chen, K.-J. Sun, L. Wang, Z. Zhang, Y.-C. Lai, N. Axmacher, and W.-X. Wang, *Coding advantage of grid cell orientation under noisy conditions*, https://www.biorxiv.org/content/10.1101/507061v1.abstract, 2018, doi.

[7] D. R. Cox and D. V. Hinkley, *Theoretical statistics*, Chapman & Hall, 1974. MR-0370837

[8] M. Csikós, N. H. Mustafa, and A. Kupavskii, *Tight Lower Bounds on the VC-dimension of Geometric Set Systems*, J. Mach. Learn. Res. **20** (2019), no. 81, 1–8, http://jmlr.org/papers/v20/18-719.html.

[9] I.R. Fiete, Y. Burak, and T. Brookings, *What Grid Cells Convey about Rat Location*, J. Neurosci. **28** (2008), no. 27, 6858–6871, doi.

[10] A.P. Georgopoulos, A. B. Schwartz, and R.E. Kettner, *Neuronal population coding of movement direction*, Science **233** (1986), no. 4771, 1416–1419, doi.

[11] T. Hafting, M. Fyhn, S. Molden, M.-B. Moser, and E.I. Moser, *Microstructure of a spatial map in the entorhinal cortex*, Nature **436** (2005), 801–806, doi.

[12] Y. I. Ingster, *Asymptotically minimax hypothesis testing for nonparametric alternatives. I*, Math. Methods Statist. **2** (1993), no. 2, 85–114. MR-1257978

[13] J. Krupic, M. Bauza, S. Burton, C. Barry, and J. O'Keefe, *Grid cell symmetry is shaped by environmental geometry*, Nature **518** (2015), 232–235, doi.

[14] P. Massart, *Concentration inequalities and model selection*, Lecture Notes in Mathematics, vol. 1896, Springer, Berlin, 2007, doi. MR-2319879

[15] A. Mathis, A. V. M. Herz, and M. Stemmler, *Optimal population codes for space: grid cells outperform place cells*, Neural Comput. **24** (2012), no. 9, 2280–2317, doi. MR-2986772

[16] B. L. McNaughton, F.P. Battaglia, O. Jensen, E.I. Moser, and M.B. Moser, *Path integration and the neural basis of the 'cognitive map'*, Nat. Rev. Neurosci. **7** (2006), no. 8, 663–678, doi.

[17] E. I. Moser, E. Kropff, and M.-B. Moser, *Place cells, grid cells, and the brain's spatial representation system*, Annu. Rev. Neurosci. **31** (2008), no. 1, 69–89, PMID: 18284371, doi.

[18] N. Mosheiff, H. Agmon, A. Moriel, and Y. Burak, *An efficient coding theory for a dynamic trajectory predicts non-uniform allocation of entorhinal grid cells to modules*, PLoS Comput. Biol. **13** (2017), no. 6, 1–19, e1005597, doi.

[19] J. O'Keefe and J. Dostrovsky, *The hippocampus as a spatial map. preliminary evidence from unit activity in the freely-moving rat*, Brain Res. **34** (1971), no. 1, 171–175, doi.

[20] F. Sargolini, M. Fyhn, T. Hafting, B. L. McNaughton, M. P. Witter, M.-B. Moser, and E. I. Moser, *Conjunctive Representation of Position, Direction, and Velocity in Entorhinal Cortex*, Science **312** (2006), no. 5774, 758–762, doi.

[21] D. M. Schwartz and O. O. Koyluoglu, *On the organization of grid and place cells: neural denoising via subspace learning*, Neural Comput. **31** (2019), no. 8, 1519–1550, doi. MR-4002804

[22] S. Sreenivasan and I. Fiete, *What grid cells convey about rat location*, Nat. Neurosci. **14** (2011), no. 10, 1330–1337, doi.

[23] M. Stemmler, A. Mathis, and A.V.M. Herz, *Connecting multiple spatial scales to decode the population activity of grid cells*, Sci. Adv. **1** (2015), no. e1500816, 1–11, doi.

[24] H. Stensola, T. Stensola, T. Solstad, K. Frøland, M.B. Moser, and E.I. Moser, *The entorhinal grid map is discretized*, Nature **492** (2012), 72–78, doi.

[25] A. B. Tsybakov, *Introduction to nonparametric estimation*, Springer Series in Statistics, Springer, New York, 2009, doi. MR-2724359

[26] L. Vágó and B.B. Ujfalussy, *Robust and efficient coding with grid cells*, PLoS Comput. Biol. **14** (2018), no. 1, 1–28, e1005922, doi.

[27] X.-X. Wei, J. Prentice, and V. Balasubramanian, *A principle of economy predicts the functional architecture of grid cells*, eLife **4** (2015), no. e08362, 1–29, doi.