EFFICIENT, PROBABILISTIC ANALYSIS OF COMBINATORIAL NEURAL CODES

Anonymous authors
Paper under double-blind review

ABSTRACT

Artificial and biological neural networks (ANNs and BNNs) can encode inputs in the form of combinations of individual neurons' activities. These combinatorial neural codes present a computational challenge for direct and efficient analysis due to their high dimensionality and often large volumes of data. Here we improve the computational complexity – from factorial to quadratic time – of direct algebraic methods previously applied to small examples and apply them to large neural codes generated by experiments. These methods provide a novel and efficient way of probing algebraic, geometric, and topological characteristics of combinatorial neural codes and provide insights into how such characteristics are related to learning and experience in neural networks. We introduce a procedure to perform hypothesis testing on the intrinsic features of neural codes using information geometry. We then apply these methods to neural activities from an ANN for image classification and a BNN for 2D navigation to, without observing any inputs or outputs, estimate the structure and dimensionality of the stimulus or task space. Additionally, we demonstrate how an ANN varies its internal representations across network depth and during learning.

1 INTRODUCTION

To understand the world around them, organisms' biological neural networks (BNNs) encode information about their environment in the dynamics of spikes varying over time and space. Artificial neural networks (ANNs) use similar principles, except instead of transmitting spikes they usually transmit a real-valued number in the range of $[0, 1]$ and their dynamics are typically advanced in a step-wise, discrete manner. Both BNNs and ANNs adjust their internal structures, e.g., connection strengths between neurons, to improve their performance in learned tasks. This leads to encoding input data into internal representations, which they then transform into task-relevant outputs, e.g., motor commands. Combinatorial neural coding schemes, i.e., encoding information in the collective activity of neurons (also called 'population coding'), is widespread in BNNs (Averbeck et al., 2006; Osborne et al., 2008; Schneidman et al., 2011; Froudarakis et al., 2014; Bush et al., 2015; Stevens, 2018; Beyeler et al., 2019; Villafranca-Faus et al., 2021; Burns et al., 2022; Hannagan et al., 2021) and long-utilized in ANNs, e.g., in associative memory networks (Little, 1974; Hopfield, 1982; Tsodyks & Feigel'man, 1988; Adachi & Aihara, 1997; Krotov & Hopfield, 2016).

Advances in mathematical neuroscience (Curto & Itskov, 2008; Curto et al., 2019) has led to the development of analyses designed to understand the combinatorial properties of neural codes and their mapping to the stimulus space. Such analyses were initially inspired by the combinatorial coding seen in place cells (Moser et al., 2008), where neurons represent physical space in the form of ensemble and individual activity (Brown & Alex, 2006; Fenton et al., 2008). Place fields, the physical spatial areas encoded by place cells, can be arranged such that they span multiple spatial dimensions, e.g., 3D navigation space in bats (Yartsev & Ulanovsky, 2013). They can also encode for 'social place' (Omer et al., 2018), the location of conspecifics. Just as these spatial and social dimensions of place (external stimuli) may be represented by combinatorial coding, so too may other dimensions in external stimuli, such as in vision (Fujii & Ito, 1996; Panzeri & Schultz, 2001; Averbeck et al., 2006; Froudarakis et al., 2014; Fetz, 1997).

In place cells, the term receptive field (RF) or place field may intuitively be thought of as a physical place. In the context of vision, for example, we may think of RFs less spatially and more abstractly as

representing stimuli features or dimensions along which neurons may respond more or less strongly, e.g., features such as orientation, spatial frequency, or motion (Niell & Stryker, 2008; Juavinett & Callaway, 2015). Two neurons which become activated simultaneously upon visual stimuli moving to the right of the visual field may be said to share the RF of general rightward motion, for example. We may also think of RFs even more abstractly as dimensions in general conceptual spaces, such as the reward–action space of a task (Constantinescu et al., 2016), visual attributes of characters or icons (Aronov et al., 2017), olfactory space (Bao et al., 2019), the relative positions people occupy in a social hierarchy (Park et al., 2021), and even cognition and behaviour more generally (Bellmund et al., 2018).

In the method described in Curto et al. (2019), tools from algebra are used to extract the combinatorial structure of neural codes. The types of neural codes under study are sets of binary vectors $\mathcal{C} \subset \mathbb{F}_2^n$, where there are n neurons in states 0 (off) and 1 (on). The ultimate structure of this method is the *canonical form* of a neural code $CF(\mathcal{C})$. The canonical form may be analysed topologically, geometrically, and algebraically to infer features such as the potential convexity of the receptive fields (RFs) which gave rise to the code, or the minimum number of dimensions those RFs must span in real space. Such analyses are possible because $CF(\mathcal{C})$ captures the minimal essential set of combinatorial descriptions which describe all existing RF relationships implied by $\mathcal{C}$. RF relationships (whether and how RFs intersect or are contained by one-another in stimulus space) are considered to be implied by $\mathcal{C}$ by assuming that if two neurons become activated or spike simultaneously, they likely receive common external input in the form of common stimulus features or common RFs. Given sufficient exploration of the stimulus space, it is possible to infer topological features of the global stimulus space by only observing $\mathcal{C}$ (Curto & Itskov, 2008; Mulas & Tran, 2020). To the best of our knowledge, these methods have only been developed and used for small examples of BNNs. Here we apply them to larger BNNs and to ANNs (by considering the co-activation of neurons during single stimulus trials).

Despite the power and broad applicability of these methods (Curto & Itskov, 2008; Curto et al., 2019; Mulas & Tran, 2020), two major problems impede their usefulness: (1) the computational time complexity of the algorithms to generate $CF(\mathcal{C})$ is factorial in the number of codewords $\mathcal{O}(nm!)$[1], limiting their use in large, real-world datasets; and (2) there is no tolerance for noise in $\mathcal{C}$, nor consideration given towards the stochastic or probabilistic natures of neural firing. We address these problems by: (1) introducing a novel method for improving the time complexity to quadratic in the number of neurons $\mathcal{O}(n^2)$ by computing the generators of $CF(\mathcal{C})$ and using these to answer the same questions; and (2) using information geometry (Nakahara & Amari, 2002; Amari, 2016) to perform hypothesis testing on the presence/absence of inferred geometric or topological properties of the stimulus or task space. As a proof of concept, we apply these new methods to data from a simulated BNN for spatial navigation and a simple ANN for visual classification, both of which may contain thousands of codewords.

2 PRELIMINARIES

Before describing our own technical developments and improvements, we first outline some of the key mathematical concepts and objects which we use and expand upon in later sections. For more detailed information, we recommend referring to Curto & Itskov (2008); Curto et al. (2019).

2.1 COMBINATORIAL NEURAL CODES

Let $\mathbb{F}_2 = \{0, 1\}$, $[n] = \{1, 2, \ldots, n\}$, and $\mathbb{F}_2^n = \{a_1 a_2 \cdots a_n | a_i \in \mathbb{F}_2, \text{ for all } i\}$. A codeword is an element of $\mathbb{F}_2^n$. For a given codeword $\mathbf{c} = c_1 c_2 \cdots c_n,$, we define its support as $supp(\mathbf{c}) = \{i \in [n] | c_i \neq 0\}$, which can be interpreted as the unique set of active neurons in a discrete time bin which correspond to that codeword. A combinatorial neural code, or a code, is a subset of $\mathbb{F}_2^n$. The support of a code C is defined as $supp(C) = \{S \subseteq [n] | S = supp(\mathbf{c}) \text{ for some } \mathbf{c} \in C\}$, which can be interpreted as all sets of active neurons represented by all corresponding codewords in C.

Let Δ be a subset of $2^{[n]}$. The subset Δ is an *abstract simplicial complex* if for any $S \in \Delta$, the condition $S' \subseteq S$ gives $S' \in \Delta$, for any $S' \subseteq S$. In other words, $\Delta \subseteq 2^{[n]}$ is an abstract simplicial

---

[1]n is the number of neurons and m is the number of codewords. In most datasets of interest $n \ll m$.

complex if it is closed under inclusion. So, the simplicial complex for a code C can be defined as

$$\Delta(C) = \{S \subseteq [n] | S \subseteq supp(\mathbf{c}), \text{ for some } \mathbf{c} \in C\}.$$

A set S in a simplicial complex Δ is referred to as an $(|S| - 1)$-simplex. For instance, a set with cardinality 1 is called 0-simplex (geometrically, a point), a set with cardinality 2 is called a 1-simplex (geometrically, an edge), and so on. Let S be an m-simplex in Δ. Any $S' \subseteq S$ is called a face of S.

2.2 SIMPLICIAL COMPLEXES AND TOPOLOGY

Let $C \subseteq \mathbb{F}_2^n$ be a code and $\Delta(C)$ be the corresponding simplicial complex of C. From now on, we will use Δ to denote the corresponding simplicial complex of a code C. Define Δ_m as a set of m-simplices in Δ. Define

$$C_m = \left\{ \sum_{S \in \Delta_m} \alpha_S S \mid \alpha_S \in \mathbb{F}_2, \forall S \in \Delta_m \right\}.$$

The set C_m forms a vector space over $\mathbb{F}_2$ whose basis elements are all the m-simplices in Δ_m. Now, define the chain complex $C_*(\Delta, \mathbb{F}_2)$ to be the sequence $\{C_m\}_{m \geq 0}$. For any $m \geq 1$, define a linear transformation $\partial_m : C_m \to C_{m-1}$, where for any $\sigma \in \Delta_m$, $\partial_m(\sigma) = \sum_{i=0}^{m} \sigma^i$, with $\sigma^i \in \Delta_{m-1}$ as a face of σ, for all $i = 0, \ldots, m$. Moreover, the map ∂_m can be extended linearly to all elements in C_m as follows

$$\partial_m \left(\sum_{S \in \Delta_m} \alpha_S S \right) = \sum_{S \in \Delta_m} \alpha_S \partial_m(S).$$

Define the m-th mod-2 homology group of Δ as

$$H_m(\Delta, \mathbb{F}_2) = \frac{Ker(\partial_m)}{Im(\partial_{m+1})}$$

for all $m \geq 1$ and

$$H_0(\Delta, \mathbb{F}_2) = \frac{C_0}{Im(\partial_1)}.$$

Note that $H_m(\Delta, \mathbb{F}_2)$ is also a vector space over $\mathbb{F}_2$, for all $m \geq 0$. So, the mod-2 m-th Betti number $\beta_m(\Delta)$ of a simplicial complex Δ is the dimension of $H_m(\Delta, \mathbb{F}_2)$. The $\beta_m(\Delta, \mathbb{F}_2)$ gives the number of m-dimensional holes in the geometric realisation of Δ.

2.3 CANONICAL FORM

Let σ and τ be subsets of $[n]$, where $\sigma \cap \tau = \varnothing$. The polynomial of the form $\prod_{i \in \sigma} x_i \prod_{j \in \tau} (1 - x_j) \in \mathbb{F}_2[x_1 m \ldots, x_n]$ is called a *pseudo-monomial*. In a given ideal $\mathcal{J} \subseteq \mathbb{F}_2[x_1, \ldots, x_n]$, a pseudo-monomial f in J is said to be minimal if there is no pseudo-monomial g in $\mathcal{J}$ with $\deg(g) < \deg(f)$ such that $f = gh$ for some $h \in \mathbb{F}_2[x_1, \ldots, x_n]$. For a given code $C \subseteq \mathbb{F}_2^n$, we can define a *neural ideal* related to C as $\mathcal{J}_C = \langle \rho_{\mathbf{c}'} | \mathbf{c}' \mathbb{F}_2^n - C \rangle$, where $\rho_{\mathbf{c}'}$ is a pseudo-monomial of the form $\prod_{i \in supp(\mathbf{c}')} x_i \prod_{j \notin supp(\mathbf{c}')} (1 - x_j)$. A set of all minimal pseudo-monomials in $\mathcal{J}_C$, denoted by $CF(\mathcal{J}_C)$ or simply $CF(C)$, is called the *canonical form* of $\mathcal{J}_C$. Moreover, it can be shown that $\mathcal{J}_C = \langle CF(C) \rangle$. Therefore, the canonical form $CF(C)$ gives a simple way to infer the RF relationships implied by all codewords in C. One way to calculate the $CF(C)$ is by using a recursive algorithm described in Curto et al. (2019). For a code $C = \{\mathbf{c}_1, \ldots, \mathbf{c}_{|C|}\}$, the aforementioned algorithm works by constructing canonical forms $CF(\varnothing), CF(\{\mathbf{c}_1\}), CF(\{\mathbf{c}_1, \mathbf{c}_2\}), \ldots, CF(C)$, respectively. In each stage, the algorithm evaluates polynomials, checks divisibility conditions, and adds or removes polynomials from a related canonical form.

3 METHODS

Our main methodological contributions are: (1) improving the computational complexity of the analyses relying on computing $CF(\mathcal{C})$ (see Algorithm 1); and (2) using information geometry to identify whether identified algebraic or topological features are statistically significant.

3.1 Computing and analysing the canonical form's generators

We may perform the same analyses as in Curto et al. (2019) in quadratic time by using Algorithm 1 to construct the generators of $CF(\mathcal{C})$ rather than constructing $CF(\mathcal{C})$ itself (as in Algorithm 2 of Curto et al. (2019)). Illustrative of this efficiency, representative experimental data with 25 neurons and 46 codewords took < 1 second to analyse on a high-end desktop PC (Intel i7 CPU and 64GB of memory), compared to 2 minutes 57 seconds using Algorithm 2 from Curto et al. (2019).

Algorithm 1 Algorithm for computing generators of $CF(\mathcal{C})$

  Input:
     $M = \mathcal{C} \subset \mathbb{F}_2^n$ as a $patterns \times neurons$ matrix
  Initialize:
     $D \leftarrow$ empty list $\triangleright$ Stores the monomials.
     $P \leftarrow$ empty list $\triangleright$ Stores the mixed monomial constructor tuples (σ, τ).
     $B \leftarrow$ empty list $\triangleright$ Stores the mixed monomial constructor tuples (τ, σ).
  for each column i of M **do**
     for each column j of M **do**
        $s \leftarrow \sum_k (i \cdot j)_k$
        if $s < 1$ **then** $\triangleright$ The pair i, j have disjoint receptive fields.
           append $\{i, j\}$ to D
        else
           $j' \leftarrow j - 1$
           $b \leftarrow \sum_{k=1} (i \cdot j')_k$
           if $b = 0$ **then** $\triangleright$ The receptive field of j is a subset of receptive field of i.
              append (i, j) to P
              append (j, i) to B
           end if
        end if
     end for
  end for

Generating desired elements of $\mathcal{J}_\mathcal{C}$ is then straightforward: monomials are supersets of disjoint pairs (from D) where each pair set has one element shared with at least one other disjoint pair set in the superset; mixed monomials are all possible combinations of first (σ set) and second (τ set) elements in the tuples of P (or vice-versa for B) – we do not allocate all of these elements but instead store the set constructors; and the negative monomial appears if and only if the all 1s codeword exists (which involves a simple summing check on columns of M).

3.2 Information geometry for combinatorial neural codes

Let N be the finite number of time bins for data of the neural activity patterns on n neurons. For any $S \subseteq [n]$, let $v(S) \in \mathbb{F}_2^n$, where $supp(v(S)) = S$ and

$$\mathbf{P}_{v(S)} = \frac{\#\{v(S)\}}{N}.$$

We would like to find the parameters $\theta = \left(\theta_{S_1}, \theta_{S_2}, \ldots, \theta_{S_{2^n-1}}\right)$, where $S_i \subseteq [n]$, $S_i \neq \varnothing$, and $S_{2^n-1} = [n]$, such that the following exponential function

$$\mathbf{P}(\mathbf{x}, \theta) = exp\left(\sum_{S \subseteq [n], S \neq \varnothing} \theta_S x_S - \psi\right),$$

where $x_S = \prod_{i \in S} x_i$ and $\psi = -\log(P_{v(\varnothing)})$, describes a neural activity pattern from the given neural activity data. We can calculate θ_S using the following formula for any $S \subseteq [n]$, where $S \neq \varnothing$,

$$\theta_S = \log\left(\frac{P_{v(S)}}{P_{v(\varnothing)} \prod_{S' \subsetneq S, S' \neq \varnothing} exp(\theta_{S'})}\right),$$

$$\eta_W = \sum_{S \subseteq [n], S \supseteq W} \mathbf{P}_{v(S)}.$$

Given a θ-coordinate, we can calculate the associated $G(\theta) = \left(g_{A,B}^\theta\right)_{A,B \subseteq [n]}$ matrix using the following formula,

$$\begin{aligned}
g_{A,B}^\theta &= E_\theta\left(X_A X_B\right) - \eta_A \eta_B \\
&= \sum_{W \supseteq A \cup B} e^{-\psi} \prod_{\substack{W' \subseteq W \\ W' \neq \varnothing}} e^{\theta_{W'}} - \eta_A \eta_B, \\
&= \sum_{W \supseteq A \cup B} e^{-\psi} e^{\sum_{\substack{W' \subseteq W \\ W' \neq \varnothing}} \theta_{W'}} - \eta_A \eta_B,
\end{aligned}$$

where $\psi = -\log(P_{v(\varnothing)})$.

Example 3.1. Let $n = 4, A = \{1, 2\}$, and $B = \{2, 4\}$, then

$$\begin{aligned}
g_{A,B}^\theta &= \sum_{W \supseteq \{1,2,4\}} e^{-\psi} \prod_{\substack{W' \subseteq W \\ W' \neq \varnothing}} e^{\theta_{W'}} \\
&= e^{-\psi} \prod_{\substack{W' \subseteq \{1,2,4\} \\ W' \neq \varnothing}} e^{\theta_{W'}} + e^{-\psi} \prod_{\substack{W' \subseteq \{1,2,3,4\} \\ W' \neq \varnothing}} e^{\theta_{W'}} \\
&= \left(e^{\theta_{\{1\}} + \theta_{\{2\}} + \theta_{\{4\}} + \theta_{\{1,2\}} + \theta_{\{1,4\}} + \theta_{\{2,4\}} + \theta_{\{1,2,4\}} - \psi} \right. \\
&\qquad \left. + e^{\theta_{\{1\}} + \theta_{\{2\}} + \theta_{\{3\}} + \theta_{\{4\}} + \theta_{\{1,2\}} + \theta_{\{1,3\}} + \theta_{\{1,4\}} + \theta_{\{2,3\}} + \theta_{\{2,4\}} + \theta_{\{3,4\}} + \theta_{\{1,2,3\}} + \theta_{\{1,2,4\}} + \theta_{\{1,2,3,4\}} - \psi} \right) \\
&\quad - \eta_{\{1,2\}} \eta_{\{2,4\}}
\end{aligned}$$

3.3 HYPOTHESIS TESTING FOR ALGEBRAIC AND TOPOLOGICAL FEATURES

Using the previous sections, we can now perform hypothesis testing on specific RF relationships or topological features such as holes.

Given $\mathbf{P}_{v(S)}$ for all $S \subseteq [n]$ as in the previous subsection, we can calculate η_W, for all $W \subseteq [n]$, where η_W is equal to $E\left(\prod_{i \in W} x_i\right) = Prob\{x_i = 1, \forall i \in W\}$, using the following formula.

$$\eta_W = \sum_{S \subseteq [n], S \supseteq W} \mathbf{P}_{v(S)}$$

Given a set of neurons $A \subseteq [n]$, where $|A| = k$, we want to test whether there is a k-th order interaction between neurons in A or not. We can do this by hypothesis testing as follows.

1. Calculate θ_S and η_W, for all $S, W \subseteq [n]$.

2. Specify a coordinate for $\mathbf{P}(\mathbf{x}; \eta, \theta)$ based on A as

$$\zeta_k^A = \left(\eta_{k-}^A; \theta_k^A\right),$$

   where $\eta_{k-}^A = (\eta_H)_{H \subseteq [n], |H| \le k}$ and $\theta_k^A = (\theta_H)_{H \subseteq [n], |H| > k}$.

3. Set the corresponding null hypothesis coordinate as

$$\zeta_k^0 = \left(\eta_{k-}^A; \theta_k^0\right),$$

   where $\eta_A = 0$, η_H is equal to the previous step except for $H = A$, and θ_k^0 is equal to the one in the previous step.

4. Determine the corresponding $G(\theta) = \left(g_{A,B}^\theta\right)_{A,B \subseteq [n]}$ matrix related to θ-coordinate using equation 3.2. Arrange the rows and columns of $G(\theta)$ such that

$$G(\theta) = \begin{pmatrix} A_\theta & B_\theta \\ B_\theta^T & D_\theta \end{pmatrix},$$

where A_θ is the submatrix of $G(\theta)$ with row and column indices from all $H \subseteq [n]$ with $|H| \leq k$ and D_θ is the submatrix of $G(\theta)$ with row and column indices from all $H \subseteq [n]$ with $|H| > k$.

5. Determine the corresponding $G(\eta) = \left(g^\eta_{A,B} \right)_{A,B \subseteq [n]}$ matrix related to η-coordinate using the equation $G(\eta) = G(\theta)^{-1}$. We can write $G(\eta)$ in the form

$$G(\eta) = \begin{pmatrix} A_\eta & B_\eta \\ B_\eta^T & D_\eta \end{pmatrix},$$

where A_η is the submatrix of $G(\eta)$ with row and column indices from all $H \subseteq [n]$ with $|H| \leq k$ and D_η is the submatrix of $G(\eta)$ with row and column indices from all $H \subseteq [n]$ with $|H| > k$.

6. Determine the corresponding $G(\zeta^A_k)$ matrix related to the mixed coordinate ζ^A_k with

$$G(\zeta^A_k) = \begin{pmatrix} A_{\zeta^A_K} & \mathbf{O} \\ \mathbf{O} & D_{\zeta^A_K} \end{pmatrix},$$

where $A_{\zeta^A_K} = A_\theta^{-1}$ and $D_{\zeta^A_K} = D_\eta^{-1}$.

7. Calculate the test statistic as follows

$$
\begin{aligned}
\lambda &= 2 \sum_{i=1}^{N} \log \left(\frac{\mathbf{P}(\mathbf{x}_i; \eta^A_{k-}, \theta^0_k)}{\mathbf{P}(\mathbf{x}_i; \eta^A_{k-}, \theta^A_k)} \right) \\
&\approx 2N\tilde{E} \left(\log \left(\frac{\mathbf{P}(\mathbf{x}; \eta^A_{k-}, \theta^0_k)}{\mathbf{P}(\mathbf{x}; \eta^A_{k-}, \theta^A_k)} \right) \right) \\
&\approx 2ND \left[\mathbf{P}(\mathbf{x}; \eta^A_{k-}, \theta^0_k); \mathbf{P}(\mathbf{x}; \eta^A_{k-}, \theta^A_k) \right] \\
&\approx N g^\zeta_{AA} (\eta^0_A - \eta_A) \\
&\approx N g^{\zeta^A_k}_{AA} (\eta^0_A - \eta_A)
\end{aligned}
$$

where $g^{\zeta^A_k}_{AA}$ is the entry of the $G(\zeta^A_k)$ matrix.

8. Fix a level of significance α and find the value $\chi^2_\alpha(1)$ (chi-square value with significance level α and degree of freedom 1) from the χ^2 look-up table.

9. Compare λ and $\chi^2 = \max\{\chi^2_\alpha(1), 1 - \chi^2_\alpha(1)\}$

   - If $\lambda \geq \chi^2$, there is a significant interaction between neurons in A (reject the null hypothesis)
   - Otherwise, there is no significant interaction between neurons in A (accept the null hypothesis)

Since G scales at 2^n, we use a subset M of all neurons, where $A \subset M$ and $|M| = 10$. We pick a set A relevant to the feature we want to test the significance of, and choose random neurons (without replacement) not already in A for the remaining elements of M, repeating the test until we exhaust all neurons. We then correct for multiple comparisons and use $\alpha = 0.05$ to detect whether there is a significant interaction in A.

The choice of A depends on which feature we wish to analyse. When analysing whether or not two neurons are disjoint or not in their RFs (a monomial relationship in $CF(\mathcal{C})$), we set A as those two neurons. When analysing whether the RF of i is contained within the RF of j (a mixed monomial relationship in $CF(\mathcal{C})$), we first set A as those two neurons, and then set A as i with a random set of neurons (repeating this at least 5 times, with different random sets, and correcting for multiple comparisons). For every dimension m where $\beta_m(\Delta, \mathbb{F}_2) > 0$, when analysing whether a hole is significant, we test for all possible sets $A \subset M \subset [n]$ which close the hole. If any test closes the hole, there is no hole, whereas if no test closes the whole, there is a hole.

4 APPLICATIONS

4.1 SPATIAL NAVIGATION IN BNNS

Using the RatInABox simulation package (George et al., 2022), we created simple 2D navigation environments with 0, 1, 2, or 3 holes in the first dimension. We used a random cover of 40 place cells modelled using Gaussians for the probability of firing and geodesic receptive field geometries. Starting at a random position, we then simulated random walks governed by Ornstein-Uhlenbeck processes for 30m, with parameters based on rat locomotion data in (Sargolini et al., 2006). We constructed a combinatorial neural code $\mathcal{C}$ using a window size of 10ms, allowing for up to 3,000 unique codewords. We constructed $\Delta(\mathcal{C})$ up to dimension 2 and calculated $\beta_1(\Delta, \mathbb{F}_2)$, with the hypothesis that β_1 would be equal to the respective number of holes in the environment. Figure 1 shows an example of a single place cell and part of a simulated trajectory for an environment with $\beta_1 = 1$ and a geometric realisation of $\Delta(\mathcal{C})$ constructed after a 30 minute random walk. Table 1 shows the number of statistically significant holes found after different durations of the trajectories for environments with different topologies. Although after 10 minutes of a random walk some holes were occasionally detected, in all cases after 20 minutes all holes in the environment were detected consistently. There were a large number of of monomials across all conditions (all simulations had > 1000) due to the covering nature of the RF arrangements. There were also a small number (all simulations had < 5) of mixed monomials (RFs found to be subsets, significantly so, of other RFs).
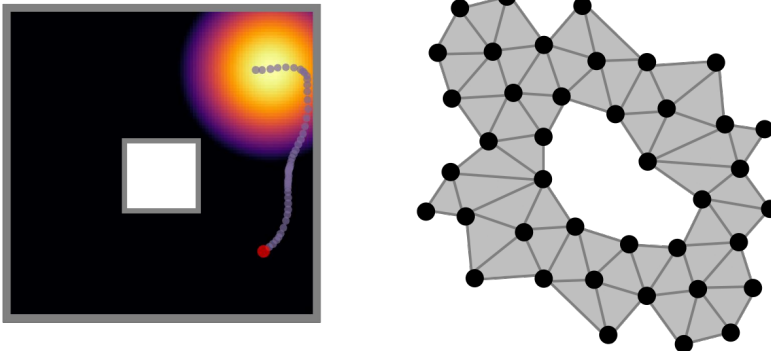

Figure 1: Example of a single place cell's receptive field and part of a simulated trajectory for an environment with $\beta_1 = 1$ (left) and a geometric realisation of $\Delta(\mathcal{C})$ constructed after a 30 minute random walk in that environment (right).

Table 1: Number of statistically significant holes found in 2D environments with 0, 1, 2, or 3 holes after 10, 20, or 30 minutes of simulated time. Quoted are the means $\pm$ S.D.s across 10 simulations, each with different random place cell coverings and trajectories. All simulations had $\beta_0 = 1$.

Time	Number of holes in the 2D environment

Time	0	1	2	3
10 minutes	0 ± 0	0 ± 0	0.8 ± 0.4	1.3 ± 0.46
20 minutes	0 ± 0	1 ± 0	2 ± 0	3 ± 0
30 minutes	0 ± 0	1 ± 0	2 ± 0	3 ± 0

4.2 VISUAL CLASSIFICATION IN ANNS

We trained a multi-layer perceptron (MLP) to classify handwritten digits from the MNIST dataset (LeCun et al., 2010) (see Figure 2, top, for examples). The model consisted of an input layer with 784 neurons (the digit pixel values), followed by two hidden layers, each with 50 neurons using the rectified linear unit activation function and 20% dropout. The final output layer consisted of 10
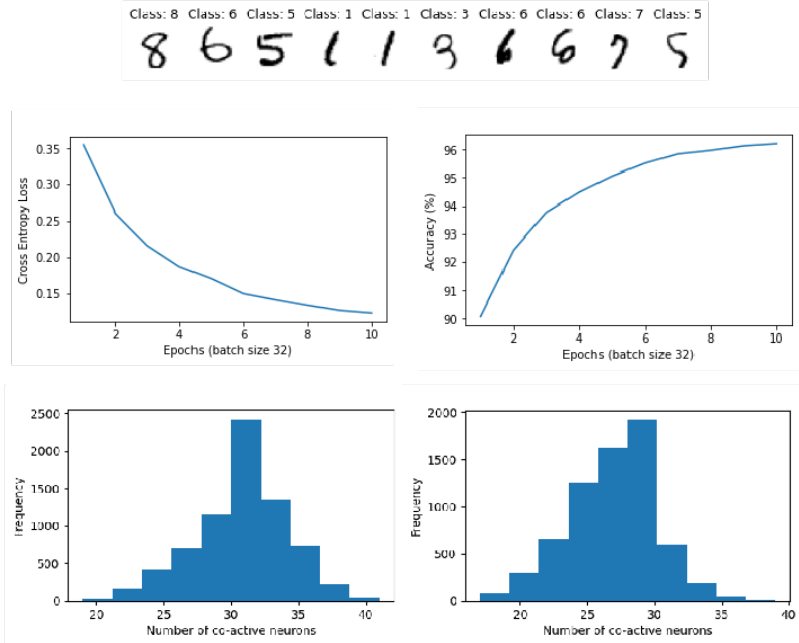
Figure 2: Top: Example MNIST digits and their classes. Middle: Test loss and accuracy over training epochs. Bottom: Histograms of the number of co-active neurons in each codeword, i.e., $|supp(\mathbf{c})|$, for $\mathcal{C}_2$ after 1 (left) and 10 (right) training epochs.

neurons (corresponding to the 10 digit class labels) and used a softmax activation function. The data was split into 50,000 digits for training, 10,000 for validation, and 10,000 for testing, allowing for up to 10,000 unique codewords in our analysis. The network was trained over 10 epochs with a batch size of 32 samples. The optimiser was stochastic gradient descent (with learning rate 0.01 and momentum 0.5) and the criterion was the cross-entropy loss between the one-hot vector of the true class labels and the output layer's activation for each sample. The MLP achieved $> 96\%$ accuracy after 10 epochs (Figure 2, middle).

After each epoch, test samples which the network did not see during training were fed through the network and the activity of all neurons in both hidden layers was recorded. The recorded activities for each hidden layer corresponding to each sample were then binarized about their means (calculated over all samples) to create a code $\mathcal{C}$ of size $10,000 \times 50$ for each layer, which we denote $\mathcal{C}_1$ for layer one and $\mathcal{C}_2$ for layer two.

Table 2: Algebraic and geometric properties of RFs of the codes $\mathcal{C}_1$ and $\mathcal{C}_2$ from the first and second hidden layers, respectively, of the MLP trained on MNIST classification. Numbers indicate counts of polynomials of that type. Monomials are reported as a tuple over orders (first, second, ..., n-th), e.g., '(5,3)' means there were 8 monomials, 5 of order 1 and 3 of order 2.

Epoch	1		5		10	
Layer code	$\mathcal{C}_1$	$\mathcal{C}_2$	$\mathcal{C}_1$	$\mathcal{C}_2$	$\mathcal{C}_1$	$\mathcal{C}_2$
Monomials	(25,9,1)	(12,7)	(5,1)	(2)	(1)	-
Mixed monomials	559,171	266,322	2,530	66	10	3
Lower bound of dimension	4	3	3	2	2	1
Negative monomials	0	0	0	0	0	0
Local obstruction to convexity?	No	No	No	No	No	No
Intersection complete?	No	No	No	No	No	No

The codes $\mathcal{C}_1$ and $\mathcal{C}_2$ showed differences in their algebraic and geometric structures across training epochs, and also differed between themselves (Table 2). In general, $\mathcal{C}_1$ had more overlapping RFs

and spanned a larger number of real dimensions (assuming convexity) than $\mathcal{C}_2$. However, during training, we find both codes lower their dimensionality and gradually spread out their RFs to cover more of the space. This is also shown by the leftward shift between epoch 1 and 10 in the histograms of the number of co-active neurons in $\mathcal{C}_2$ (Figure 2, bottom).

5 DISCUSSION

We have shown it is possible to analyse the intrinsic geometry and topology of combinatorial neural codes from biological and artificial networks in an efficient and probabilistic manner. With these improved methods, we can now comfortably study codes with tens and even hundreds of thousands of codewords. We have shown how these methods can be used to better understand (with some statistically surety) how the internal representations of external inputs within these networks can change through learning, experience, and network depth.

Neuroscientists have shown combinatorial neural codes can occupy low-dimensional subspaces called *neural manifolds* in the covariance of their neural activities (Gallego et al., 2017; Feulner & Clopath, 2021). Trajectories and regions in these subspaces can correspond to task cognition, perceptual classification, and movement (Cohen et al., 2020; Chung & Abbott, 2021). For example, Gardner et al. (2022) show the activity of populations of hundreds of grid cells within single modules of medial entorhinal cortex (a brain area partly responsible for navigation) occupy positions on a toroidal manifold. Positions on this manifold correspond to positions in the 2D space which the animal is navigating in.

These findings might lead us to believe combinatorial neural codes are intrinsically low-dimensional despite being embedded in the high-dimensional combinatorial space of neural activity. However, theoretical (Bartolo et al., 2020) and experimental (Rigotti et al., 2013) studies have shown that the dimensionality of these neural manifolds is influenced and often directly corresponds to the dimensionality of the task or learning under study. Indeed, the low-dimensional embeddings found by Gardner et al. (2022) are predicted by the two-dimensionality of the navigation (the underlying cause of the neural activity). Mathematically-optimal combinatorial neural codes and their RFs are also related to the dimensionality of the inputs those codes are attempting to represent (Wang et al., 2013). In more naturalistic and complex tasks, maintaining high-dimensional representations in the neural code may allow for increased expressibility but lower generalisability, whereas reducing to low-dimensional representations may allow for less expressibility but higher generalisability (Fusi et al., 2016; Badre et al., 2021). High-dimensional codes are often found in recordings from BNNs and are are often found when individual neurons encode for multiple input features, allowing linear read-out of a large number of complex or simple features (Fusi et al., 2016). Such neurons, for example in macaque inferotemporal cortex (Higgins et al., 2021), can also encode for very specific and independent higher-dimensional features.

This implies combinatorial neural codes can include mixtures of coding strategies which are simultaneously low- and high-dimensional. One of the key advantages of the techniques developed and applied in this study is that we can consider these different dimensionalities of coding at the same time. We don't reduce the embedding dimensionality to perform our analysis (which would be equivalent to assuming a low-dimensional code). We also don't try to map individual neuron responses to experimenter-known but network-unknown external, high-dimensional variables (which would be equivalent to assuming a high-dimensional code). Instead, we keep the full, original dimensionality of the data and can identify low- or high-dimensional features and response relationships at local and global levels simultaneously, all without reference to information external to the neural network. We also provide a method for testing the statistical significance of these features and relationships, again while maintaining the original high embedding dimension of the data. This allows us to avoid making any strong assumptions about dimensionality of the task, stimuli, or the corresponding neural code – instead, we let the data speak for themselves.

We do carry over some limitations from prior work, most prominently: (a) we assume joint-activity of neurons corresponds to common inputs or selectivity thereof; and (b) we binarize neural signals into 'on' and 'off' states. We suggest future work now focus on mitigating these limitations by: (a) performing causal inference tests on neural co-activations; and (b) considering polynomials over larger finite fields, e.g., $\mathbb{F}_4$, or extending these methods to more 'continuous' structures, e.g., manifolds.

REFERENCES

Masaharu Adachi and Kazuyuki Aihara. Associative dynamics in a chaotic neural network. *Neural Networks*, 10(1):83–98, 1997. ISSN 0893-6080. doi: https://doi.org/10.1016/S0893-6080(96)00061-5. URL https://www.sciencedirect.com/science/article/pii/S0893608096000615.

Shun-ichi Amari. *Information Geometry and Its Applications*. Springer Publishing Company, Incorporated, 1st edition, 2016. ISBN 4431559779.

Dmitriy Aronov, Rhino Nevers, and David W. Tank. Mapping of a non-spatial dimension by the hippocampal–entorhinal circuit. *Nature*, 543(7647):719–722, Mar 2017. ISSN 1476-4687. doi: 10.1038/nature21692.

Bruno B Averbeck, Peter E Latham, and Alexandre Pouget. Neural correlations, population coding and computation. *Nature Reviews Neuroscience*, 7:358—366, 2006. doi: 10.1038/nrn1888.

David Badre, Apoorva Bhandari, Haley Keglovits, and Atsushi Kikumoto. The dimensionality of neural representations for control. *Current Opinion in Behavioral Sciences*, 38:20–28, 2021. ISSN 2352-1546. doi: https://doi.org/10.1016/j.cobeha.2020.07.002. URL https://www.sciencedirect.com/science/article/pii/S2352154620301042. Computational cognitive neuroscience.

Xiaojun Bao, Eva Gjorgieva, Laura K. Shanahan, James D. Howard, Thorsten Kahnt, and Jay A. Gottfried. Grid-like neural representations support olfactory navigation of a two-dimensional odor space. *Neuron*, 102(5):1066–1075.e5, Jun 2019. ISSN 0896-6273. doi: 10.1016/j.neuron.2019.03.034. URL https://doi.org/10.1016/j.neuron.2019.03.034.

Ramon Bartolo, Richard C. Saunders, Andrew R. Mitz, and Bruno B. Averbeck. Dimensionality, information and learning in prefrontal cortex. *PLOS Computational Biology*, 16:1–26, 04 2020. doi: 10.1371/journal.pcbi.1007514. URL https://doi.org/10.1371/journal.pcbi.1007514.

Jacob L. S. Bellmund, Peter Gärdenfors, Edvard I. Moser, and Christian F. Doeller. Navigating cognition: Spatial codes for human thinking. *Science*, 362(6415):eaat6766, 2018. doi: 10.1126/science.aat6766. URL https://www.science.org/doi/abs/10.1126/science.aat6766.

Michael Beyeler, Emily L. Rounds, Kristofor D. Carlson, Nikil Dutt, and Jeffrey L. Krichmar. Neural correlates of sparse coding and dimensionality reduction. *PLOS Computational Biology*, 15(6):1–33, 06 2019. doi: 10.1371/journal.pcbi.1006908. URL https://doi.org/10.1371/journal.pcbi.1006908.

W Michael Brown and B Alex. Optimal Neuronal Tuning for Finite Stimulus Spaces. *Neural Computation*, 18(1999):1511–1526, 2006.

Thomas F. Burns, Tatsuya Haga, and Tomoki Fukai. Multiscale and extended retrieval of associative memory structures in a cortical model of local-global inhibition balance. *eNeuro*, 9(3), 2022. doi: 10.1523/ENEURO.0023-22.2022. URL https://www.eneuro.org/content/9/3/ENEURO.0023-22.2022.

Daniel Bush, Caswell Barry, Daniel Manson, and Neil Burgess. Using Grid Cells for Navigation. *Neuron*, 87(3):507–520, Aug 2015. ISSN 1097-4199. doi: 10.1016/j.neuron.2015.07.006. URL https://pubmed.ncbi.nlm.nih.gov/26247860.

SueYeon Chung and L.F. Abbott. Neural population geometry: An approach for understanding biological and artificial neural networks. *Current Opinion in Neurobiology*, 70:137–144, 2021. ISSN 0959-4388. doi: https://doi.org/10.1016/j.conb.2021.10.010. URL https://www.sciencedirect.com/science/article/pii/S0959438821001227. Computational Neuroscience.

Uri Cohen, SueYeon Chung, Daniel D. Lee, and Haim Sompolinsky. Separability and geometry of object manifolds in deep neural networks. *Nature Communications*, 11(1):746, Feb 2020.

Alexandra O. Constantinescu, Jill X. O'Reilly, and Timothy E. J. Behrens. Organizing conceptual knowledge in humans with a gridlike code. *Science*, 352(6292):1464–1468, 2016. doi: 10.1126/science.aaf0941. URL https://www.science.org/doi/abs/10.1126/science.aaf0941.

Carina Curto and Vladimir Itskov. Cell groups reveal structure of stimulus space. *PLOS Computational Biology*, 4:1–13, 10 2008. doi: 10.1371/journal.pcbi.1000205. URL https://doi.org/10.1371/journal.pcbi.1000205.

Carina Curto, Alan Veliz-Cuba, and Nora Youngs. Chapter 7 - Analysis of Combinatorial Neural Codes: An Algebraic Approach. In Raina Robeva and Matthew Macauley (eds.), *Algebraic and Combinatorial Computational Biology*, MSE/Mathematics in Science and Engineering, pp. 213–240. Academic Press, 2019. ISBN 978-0-12-814066-6. doi: https://doi.org/10.1016/B978-0-12-814066-6.00007-6. URL https://www.sciencedirect.com/science/article/pii/B9780128140666000076.

A Fenton, Hsin-yi Kao, Samuel A Neymotin, Andrey Olypher, Yevgeniy Vayntrub, William W Lytton, and Nandor Ludvig. Unmasking the CA1 Ensemble Place Code by Exposures to Small and Large Environments: More Place Cells and Multiple, Irregularly Arranged, and Expanded Place Fields in the Larger Space. *Journal of Neuroscience*, 28:11250–11262, 2008. doi: 10.1523/JNEUROSCI.2862-08.2008.

Eberhard E Fetz. Temporal Coding in Neural Populations? *Science*, 278(5345):1901–1902, 1997.

Barbara Feulner and Claudia Clopath. Neural manifold under plasticity in a goal driven learning behaviour. *PLOS Computational Biology*, 17:1–27, 02 2021. doi: 10.1371/journal.pcbi.1008621. URL https://doi.org/10.1371/journal.pcbi.1008621.

Emmanouil Froudarakis, Philipp Berens, Alexander S Ecker, R James Cotton, Fabian H Sinz, Dimitri Yatsenko, Peter Saggau, Matthias Bethge, and Andreas S Tolias. Population code in mouse V1 facilitates readout of natural scenes through increased sparseness. *Nature Neuroscience*, 17: 851—-857, 2014. doi: 10.1038/nn.3707.

Hiroshi Fujii and Hiroyuki Ito. Dynamical Cell Assembly Hypothesis — Theoretical Possibility of Spatio-temporal Coding in the Cortex. *Neural Networks*, 9:1303–1350, 1996.

Stefano Fusi, Earl K Miller, and Mattia Rigotti. Why neurons mix: high dimensionality for higher cognition. *Current Opinion in Neurobiology*, 37:66–74, 2016. ISSN 0959-4388. doi: https://doi.org/10.1016/j.conb.2016.01.010. URL https://www.sciencedirect.com/science/article/pii/S0959438816000118. Neurobiology of cognitive behavior.

Juan A. Gallego, Matthew G. Perich, Lee E. Miller, and Sara A. Solla. Neural manifolds for the control of movement. *Neuron*, 94(5):978–984, 2017. ISSN 0896-6273. doi: https://doi.org/10.1016/j.neuron.2017.05.025. URL https://www.sciencedirect.com/science/article/pii/S0896627317304634.

Richard J. Gardner, Erik Hermansen, Marius Pachitariu, Yoram Burak, Nils A. Baas, Benjamin A. Dunn, May-Britt Moser, and Edvard I. Moser. Toroidal topology of population activity in grid cells. *Nature*, 602(7895):123–128, Feb 2022.

Tom M George, William de Cothi, Claudia Clopath, Kimberly Stachenfeld, and Caswell Barry. RatInABox: A toolkit for modelling locomotion and neuronal activity in continuous environments. *bioRxiv*, 2022. doi: 10.1101/2022.08.10.503541. URL https://www.biorxiv.org/content/early/2022/09/14/2022.08.10.503541.

T. Hannagan, A. Agrawal, L. Cohen, and S. Dehaene. Emergence of a compositional neural code for written words: Recycling of a convolutional neural network for reading. *Proceedings of the National Academy of Sciences*, 118(46), 2021. ISSN 0027-8424. doi: 10.1073/pnas.2104779118. URL https://www.pnas.org/content/118/46/e2104779118.

Irina Higgins, Le Chang, Victoria Langston, Demis Hassabis, Christopher Summerfield, Doris Tsao, and Matthew Botvinick. Unsupervised deep learning identifies semantic disentanglement in single inferotemporal face patch neurons. *Nature Communications*, 12(1):6456, Nov 2021.

ISSN 2041-1723. doi: 10.1038/s41467-021-26751-5. URL https://doi.org/10.1038/s41467-021-26751-5.

J J Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences*, 79(8):2554–2558, 1982. doi: 10.1073/pnas.79.8.2554.

Ashley L. Juavinett and Edward M. Callaway. Pattern and component motion responses in mouse visual cortical areas. *Current Biology*, 25:1759–1764, 2015.

D Krotov and J J Hopfield. Dense associative memory for pattern recognition. In *Neural Information Processing Systems*, volume abs/1606.01164, 2016. URL http://arxiv.org/abs/1606.01164.

Yann LeCun, Corinna Cortes, and Christopher Burges. MNIST handwritten digit database. http://yann.lecun.com/exdb/mnist/, 2010. URL http://yann.lecun.com/exdb/mnist/.

W. A. Little. The existence of persistent states in the brain. *Mathematical Biosciences*, 19(1):101–120, 1974. ISSN 0025-5564. doi: https://doi.org/10.1016/0025-5564(74)90031-5. URL https://www.sciencedirect.com/science/article/pii/0025556474900315.

Edvard I Moser, Emilio Kropff, and May-britt Moser. Place Cells, Grid Cells, and the Brain's Spatial Representation System. *Annual Review of Neuroscience*, 31:69–89, 2008. doi: 10.1146/annurev.neuro.31.061307.090723.

Raffaella Mulas and Ngoc M. Tran. Minimal embedding dimensions of connected neural codes. *Algebraic Statistics*, 11:99—106, 2020. doi: 10.2140/astat.2020.11.99.

Hiroyuki Nakahara and Shun-ichi Amari. Information-Geometric Measure for Neural Spikes. *Neural Computation*, 14(10):2269–2316, 10 2002.

Cristopher Niell and M. P. Stryker. Highly Selective Receptive Fields in Mouse Visual Cortex. *Journal of Neuroscience*, 28:7520–7536, 2008. doi: 10.1523/JNEUROSCI.0623-08.2008.Highly.

David B Omer, Shir R Maimon, and Liora Las. Social place-cells in the bat hippocampus. *Science*, 359:218–224, 2018.

Leslie C. Osborne, Stephanie E. Palmer, Stephen G. Lisberger, and William Bialek. The neural basis for combinatorial coding in a cortical population response. *Journal of Neuroscience*, 28 (50):13522–13531, Dec 2008. ISSN 1529-2401. doi: 10.1523/JNEUROSCI.4390-08.2008. URL https://doi.org/10.1523/JNEUROSCI.4390-08.2008.

Stefano Panzeri and Simon R Schultz. A Unified Approach to the Study of Temporal, Correlational, and Rate Coding. *Neural Computation*, 13:1311–1349, 2001.

Seongmin A. Park, Douglas S. Miller, and Erie D. Boorman. Inferences on a multidimensional social hierarchy use a grid-like code. *Nature Neuroscience*, 24(9):1292–1301, Sep 2021. ISSN 1546-1726. doi: 10.1038/s41593-021-00916-3. URL https://doi.org/10.1038/s41593-021-00916-3.

Mattia Rigotti, Omri Barak, Melissa R. Warden, Xiao-Jing Wang, Nathaniel D. Daw, Earl K. Miller, and Stefano Fusi. The importance of mixed selectivity in complex cognitive tasks. *Nature*, 497 (7451):585–590, May 2013.

Francesca Sargolini, Marianne Fyhn, Torkel Hafting, Bruce L. McNaughton, Menno P. Witter, May-Britt Moser, and Edvard I. Moser. Conjunctive representation of position, direction, and velocity in entorhinal cortex. *Science*, 312(5774):758–762, 2006. doi: 10.1126/science.1125572. URL https://www.science.org/doi/abs/10.1126/science.1125572.

Elad Schneidman, Jason L. Puchalla, Ronen Segev, Robert A. Harris, William Bialek, and 2nd Berry, Michael J. Synergy from silence in a combinatorial neural code. *Journal of Neuroscience*, 31(44):15732–15741, Nov 2011. ISSN 1529-2401. doi: 10.1523/JNEUROSCI.0301-09.2011. URL https://doi.org/10.1523/JNEUROSCI.0301-09.2011.

Charles F. Stevens. Conserved features of the primate face code. *Proceedings of the National Academy of Sciences*, 115(3):584–588, 2018. ISSN 0027-8424. doi: 10.1073/pnas.1716341115. URL `https://www.pnas.org/content/115/3/584`.

M. V Tsodyks and M. V Feigel'man. The enhanced storage capacity in neural networks with low activity level. *Europhysics Letters (EPL)*, 6(2):101–105, may 1988. doi: 10.1209/0295-5075/6/2/002. URL `https://doi.org/10.1209/0295-5075/6/2/002`.

María Villafranca-Faus, Manuel Esteban Vila-Martín, Daniel Esteve, Esteban Merino, Anna Teruel-Sanchis, Ana Cervera-Ferri, Joana Martínez-Ricós, Ana Lloret, Enrique Lanuza, and Vicent Teruel-Martí. Integrating pheromonal and spatial information in the amygdalo-hippocampal network. *Nature Communications*, 12(1):5286, Sep 2021. ISSN 2041-1723. doi: 10.1038/s41467-021-25442-5. URL `https://doi.org/10.1038/s41467-021-25442-5`.

Zhuo Wang, Alan A Stocker, and Daniel D Lee. Optimal neural population codes for high-dimensional stimulus variables. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger (eds.), *Advances in Neural Information Processing Systems*, volume 26. Curran Associates, Inc., 2013. URL `https://proceedings.neurips.cc/paper/2013/file/6da9003b743b65f4c0ccd295cc484e57-Paper.pdf`.

Michael M Yartsev and Nachum Ulanovsky. Representation of Three-Dimensional Space in the Hippocampus of Flying Bats. *Science*, 340:367–373, 2013.