Information theoretic properties of colluding populations of grid, place, and time cells

David Schwartz*

1 Abstract

A model of interactions between mammalian brain regions Hippocampal CA1, CA3, and Medial Entorhinal Cortex (MEC) is considered and shown to be effective in exploiting each population's activity to ameliorate the effects of neural noise on the other in [19, 20, 21]. Here, this model is recosidered in an information theoretic framework. Results of the evaluation of collective information processing properties of the network's de-noising process through the lenses of interaction information are presented and discussed. Preliminary observations on transfer entropies during de-noising are discussed. Further exploration from this new vantage point, involving validation of the model presented here against experimental observations, as well as the investigation of directed information during de-noising are proposed.

2 Introduction

Place cells are spatially modulated neurons with bivariate Gaussian tuning curves centered on particular locations in the environment, and have been identified in the hippocampus [4]. Grid cells are spatially modulated neurons that exhibit a peak firing rate at a periodic and hexagonally symmetric distribution of locations in the envirionment, and are found in the Entorhinal Cortex (EC) of rats, mice, bats, and humans [29, 5, 6, 8]. Grid cells are clustered in discrete modules wherein cells share grid scale. Anatomically, both cell types share a dorsoventral organization, with cells possessing wider receptive fields distributed towards the ventral end [24, 23]. It is known that the rat grid cell network requires communication from the hippocampus to maintain grid-like activity [1], and that a significant improvement in accuracy of the rodent place cell representation is tightly correlated with the emergence of the grid cell network [12]. However, the mechanisms by which these networks communicate and how each may bolster the other's accuracy are unknown.

To the best of the author's knowledge, information theoretical investigation of these cells activities and their potential for collusion has yet to be published or investigated rigorously. In this work, we seek to characterize and study limits on the information processing in colluding populations of grid, place, and time cells, modeled as in our earlier approaches (i.e. as in [19, 20, 21]). Such a theoretical characterization may serve to provide experimental neuroscientists with hunches and predictions to guide their own interrogations, while simultaneously enhancing neuroscience's understanding of information and coding theoretic properties of the mammalian navigational neural codes.

This paper is organized sectionally: We begin by with a short primer on navigational neural codes, and their experimentally observed behavior. Section 4 develops the theoretical framework supporting all numerical simulation experiments, results, and analysis thereof. In section 5, we state and illustrate observations on distributions of interaction informations across populations. In section 6, we analyze the aforementioned results, analyze said results, discuss limiations of these analyses, and suggest avenues for further investigation.

*The author is with the University of Arizona. Email:dmschwar@email.arizona.edu.

3 Background

3.1 Place cells

Place cells are pyramidal cells in the mammalian hippocampus (HC) whose firing rates depend on the location of the organism [15]. Their receptive fields (mappings of a space of percepts - e.g. position, time, head direction, angle of a bar of light, etc. - to firing rate of a cell) appear to be approximately bivariate Gaussian centered at the cell's preferred location. Anatomically, place cells are organized dorso-ventrally by increasing area of coverage of their place fields [24].

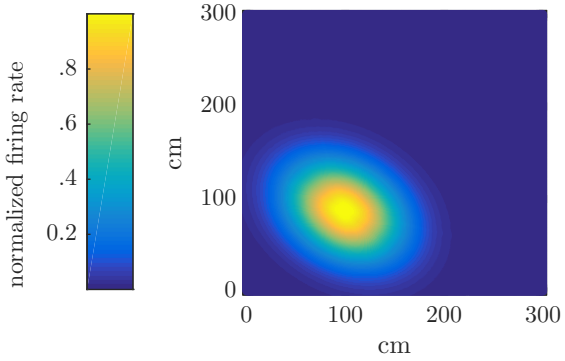

Figure 1: Receptive field of a typical ideal place cell

3.2 Grid cells

Grid cells are spatially modulated neurons, found in the meidal entorhinal cortex (MEC), that exhibit a peak firing rate at a periodic and hexagonally symmetric distribution of locations in the envirionment, and are found in the Entorhinal Cortex (EC) of rats, mice, bats, and humans [5, 6, 29, 8]. Grid cells are clustered in discrete modules wherein cells share grid scale. Like place cells, grid cells are also organized dorso-ventrally by width the period of the grid field, with wider receptive fields distributed towards the ventral end [23].

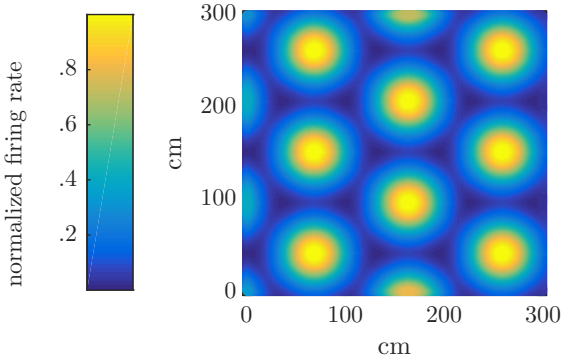

Figure 2: Receptive field of a typical ideal grid cell

3.3 Time cells

Time cells, found in HC and MEC, fire at successive moments in a temporally ordered sequence of events [17]. Here, they are modeled phenomenologically with a Gaussian receptive field centered on the cells preferred instant. It has been reported that there is a significant positive correlation between a time cell's preferred moment (i.e. the center of the time field) and the apparent width of the time field - that is, cells that prefer to fire later in the sequence tend to spike for a greater duration [17].

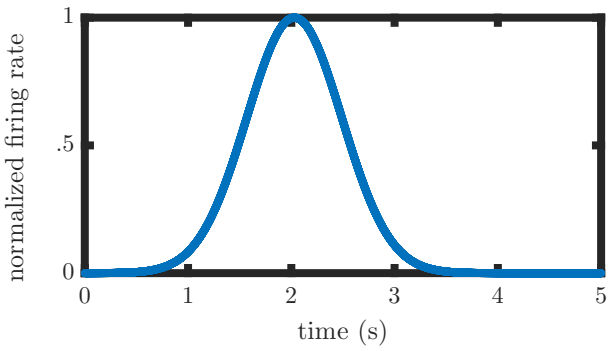

Figure 3: Receptive field of a typical ideal time cell

3.4 Decoding population activity

In order to quantify the information content of the population, we estimated the location encoded by the population using a brute force maximum a posteriori approach, in which we compute

$$\hat{\mathbf{x}} = \underset{\mathbf{x}\in\mathcal{E}}{\operatorname{argmax}}\, P(\mathbf{k}|\mathbf{x}),$$

where $\mathcal{E}$ is the set of all possible quantized positions in space-time and $\mathbf{k}$ is a vector representing the population's firing rate response to the position and time we wish to estimate. Since we assume that each cell fires as a poisson process with rate parameter equal to the cell's instantaneous firing rate, if we further assume that cells fire independently and that simulated cell firing rate is sampled uniformly at a sampling period of $\tau = 1$ second, and that cell i has receptive field given by $f_i(\mathbf{x})$, it is easy to decompose $P(\mathbf{k}|\mathbf{x})$ as

$$P(\mathbf{k}|\mathbf{x}) = \prod_{i=1}^{n} P(k_i|\mathbf{x}) = \prod_{i=1}^{n} \frac{(f_i(\mathbf{x}))^{k_i}\exp(-f_i(\mathbf{x}))}{k_i!}$$

Then, $\log(P(\mathbf{k}|\mathbf{x})) = \sum_{i=1}^{n} -k_i f_i(\mathbf{x})\log(f_i(\mathbf{x})) - \log(k_i!)$ and we can safely further reduce our decoding complexity by ignoring the constant term $(\sum_{i=1}^{n} \log(k_i!))$. Unfortunately, if one wishes to make no assumptions about the continuity of the rat's motion, one must search the complete set of quantized positions for maximizing $\mathbf{x}$. This maximimum likelihood approach optimally estimates location from population activity when coupled with strong and true assumptions about the cells' receptive fields (i.e. trustworthy knowledge of f_i, for $i \in \{1, ..., n\}$). There are several other decoding mechanisms popularly employed in decoding activity from real and simulated place cells including basis function decoders and kalman filters [30]. Consider a naive linear estimator, implemented in Algorithm 1, which produces a firing-rate-weighted average of the place cells' centers. As discussed in [30], this linear estimator is equivalent to an instance of the more general basis function decoder for the appropriate choice of basis and template function.

Algorithm 1 Linear Decoding

Input: Vector of firing rates, $\mathbf{f}$, Matrix whose rows are coordinates of centers of place cell RFs $[\mathbf{x}, \mathbf{y}]$
Output: Position estimate, $(\hat{x}, \hat{y})$

1: Normalization: $\mathbf{f}_N \leftarrow \frac{\mathbf{f}}{\sum_i \mathbf{f}_i}$
2: $\hat{x} \leftarrow \mathbf{f}_N \cdot \mathbf{x}$
3: $\hat{y} \leftarrow \mathbf{f}_N \cdot \mathbf{y}$

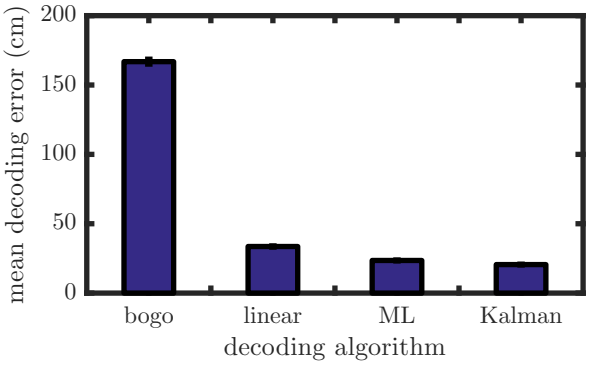

Figure 4: Mean error of position decoding from 100 networks of 120 place cells with realistically parameterized place field sizes and place field centers chosen uniformly randomly from a 300cm wide square environment

We also consider an iterated extended Kalman filter, described in detail in [2, 11] whose internal models assume Poisson-like spiking governed by rate parameters equal to the cells' firing rates. This Kalman filter decoder was considered in order to acquire a benchmark for comparison of the aforementioned decoders to state of the art. In figure 4 we show mean error of position decoding in (cm) from 100 networks of 120 place cells each, with a realistic variety of place field widths, and place field centers chosen uniformly randomly from quantized positions in a 300cm wide square environment. Surprisingly, the naive linear decoder is only barely outperformed by the Kalman filter and the maximum likelihood (ML) approach. Error bars indicate standard error of the mean (i.e. ratio of standard deviation to square root of number of trials). Unsurprisingly, all of the decoding methods considered dramatically and significantly outperform bogo decoder, which guesses locations uniformly randomly from the environment. Also unsurprisingly, each systematic decoder performs close to the empirically observed optimum (achieved by the Kalman filter).

4 Theoretical framework

4.1 Population parameterization

To develop a code for spatio-temporal position, we extend the code described in [21] and [19] with instantiation of a population of n_G grid, n_P place, and n_T time cells (a total of N neurons), modeled phenomenologically: The time cells are defined by univariate gaussian tuning curves with mean $\nu_t \in [0, \tau]$ and variance $\sigma_t^2 \in [0.1, \frac{\tau}{2}]$, where τ is the duration of the path to encode. Denote the 'time field' of time cell k by T_k. In electrophysiologcal single cell recordings, time cells in observed in both hippocampal regions CA1 and CA3 exhibit a strong postive correlation between their preferred temporal position (ν_t) and the width of their time field, (σ_t^2) [17]. We consider both this deliberately parameterized scheme, as well as one in which we impose no correlation between ν_t and σ_t^2. We parameterize populations of grid and place cells similarly. That is, each grid cell's tuning curve follows a two-dimensional distribution resembling a von-Mises density function,

$$g_{m,j}(\mathbf{s}) = \frac{f_{\max}}{Z} \exp\left[\sum_{k=1}^{3} \cos\left(\frac{4}{\lambda_m\sqrt{3}}\mathbf{u}(\theta_k - \theta_{m,j}) \cdot (\mathbf{s} - \mathbf{c}_{m,j}) + \frac{3}{2}\right) - 1\right],$$

where $\mathbf{u}(\theta_k - \theta_{m,j})$ is a unit vector in the direction of $\theta_k - \theta_{m,j}$, $\mathbf{s} \in [0, L] \times [0, L]$ is the position stimulus, $\mathbf{c}_{m,j}$, $\theta_{m,j}$, and λ_m are the grid cell's spatial phase offset, orientation offset, and scaling ratio (L is defined as the length of a sqaure arena). The grid cells' scaling ratio $\lambda^* = \frac{\lambda_{m+1}}{\lambda_m}$, for the jth grid cell in module m. Grid cell orientations were considered as ideal values about which the measurements presented in [23] fluctuate. Place fields are modeled as thresholded bivariate Gaussian distributions over $[0, L] \times [0, L]$. Denote the place field belonging to place cell i as p_i. Upon instantiation, place field centers are chosen

uniformly pseudorandomly from $[0, L] \times [0, L]$, and place field widths are chosen uniformly pseudorandomly from $[0.9\lambda_1, 1.1\lambda_M]$, where M is the largest grid module index.

4.2 A hybrid code for spatio-temporal position

We compute neural codes for space and time as functions of real paths recorded from a rat's traversal of a square arena of length 300cm during a spatial navigation task. By using real paths, we ensure that our results depend on biologically real spatiotemporal stimulus statistics, strengthening biological predictions. Specifically, we stimulate each grid, place, and time cell with $(x_i, y_i, t_i)_i$, the recorded sequence of positions of an led tracker mounted on the animal's head ($\{(x_i, y_i)\}$) and times ($\{t_i\}$) at which the rat was visited the ith position.

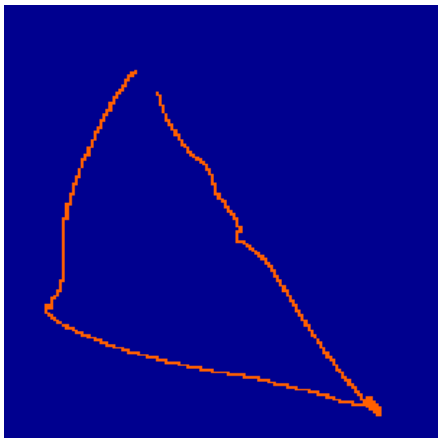

Figure 5: An example of a short path traversed by Zenith, a Brown Norway

4.3 De-noising network

Our earlier results, discussed in [21], indicate that a clustered de-noising network oupterforms all unclustered variants considered. With this in mind, we implement a de-noising network with the topology illustrated in figure 6, custering interneurons to associate with corresponding grid modules, as well as any place or time cell. This way, information may only flow accross grid modules through interneurons mutually coupled to place or time cells.

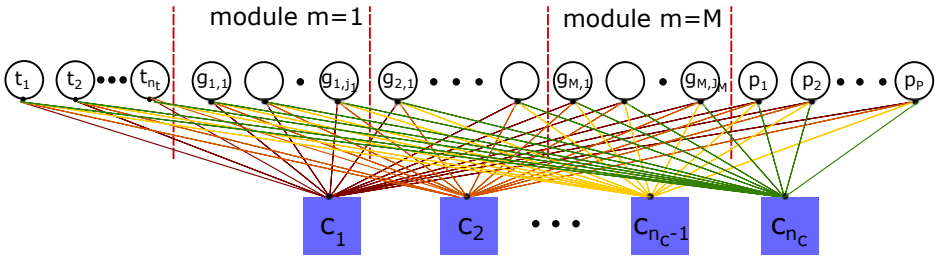

Figure 6: The clustered de-noising network implmented in [21], augmented with time cells

4.4 Path space learning

Before we can use the de-noising system to correct corrupted codewords, it must learn (i.e. adapt its weights for) the code. This process is complete when the interneurons may be read to determine if the states of the pattern neurons map to a valid codeword. Formally, this amounts to finding a connectivity matrix, W ($W_{i,j}$ is the synaptic weight between interneuron i and pattern neuron j), whose rows are approximately perpendicular to the code space. A procedure to procure such a matrix is outlined in [14], and improved in [16]. These algorithms begin with a random set of vectors, and for each, seeks a nearby

vector orthogonal to $\mathcal{C}$ (i.e. a vector onto which each element of $\mathcal{C}$ has minimal projection). We implement this in Algorithm 4.4 (a derivation of this algorithm can be found in the appendix in section 7.1) by applying it to each cluster's local connectivity matrix. Note that here, all arithmetic on the synaptic weights, $W_{i,j}$ is performed in $\mathbb{R}$, while arithmetic on states of neurons (i.e. their firing rates), is quantized to the nearest integer in $[0, Q-1]$. The maximum firing rate, $Q-1$, is identical for all neurons. With each update,

$$\mathbf{w} \leftarrow \mathbf{w} - \alpha_t(y(\mathbf{x} - \frac{y\mathbf{w}}{\|\mathbf{w}\|^2}) + \eta\Gamma(\mathbf{w}, \theta)),$$

where θ is a sparsity threshold, η is a penalty coefficient, $y = \mathbf{x}^T\mathbf{w}$ is the scalar projection of $\mathbf{x}$ onto $\mathbf{w}$, and α_t is the learning rate at iteration t. Γ is a sparsity enforcing function, approximating the gradient of a penalty function, $g(\mathbf{w}) = \sum_{k=1}^{m} \tanh(\sigma {\mathbf{w}_k}^2)$, which, for appropriate choices of σ, penalizes non-sparse solutions early in the learning procedure [16].

As in [16], to speed up learning, we approximate $\Gamma = \nabla g$ with

$$\Gamma(w_t, \theta_t) = \begin{cases} w_t & : |w_t| \leq \theta_t \\ 0 & : \text{otherwise} \end{cases}$$

This is a an improved approximation to Oja's Hebbian learning algorithm [14], with advantages in both biological plausibility and computational complexity. For connections of fixed type (i.e. inhibitory vs excitatory), Oja's rule alone is biologically dubious without the inclusion of many interneurons to manage this change in type. Dale's Principle states that real synaptic connections change type rarely, if ever [3]. In accordance with this principle, our update rule does not allow weights to change sign. This is accomplished after the updated weights are determined: If the sign has changed after applying the update, set the new weight to a value just above (resp. just below) zero if the previous weight was positive (resp. negative). Thus, when learning is complete, these weights will be small in magnitude and are thresholded to zero.

In Algorithm 4.4, line 11 terminates the learning process if the sum of the projections of the current weight vector on each pattern is no more than ϵ away from zero, that is, if the current weight vector is approximately orthogonal to the code space. Lines 17-21 perform a thresholding operation that maps to zero any weight sufficiently small in magnitude. This is primarily to suppress numerical errors and promote consistency, as in Line 11, we use ϵ as a small positive constant. Note that since the weights processed on each iteration are independent of those in other iterations, this algorithm can be readily parallelized so that each constraint neuron learns its weights simultaneously.

Algorithm 2 Neural learning

Input: set of C patterns, $\mathcal{C}$, stopping point, ϵ
Output: learned weights matrix, W

for rows, $\mathbf{w}$, of W **do**
  for $t \in \{1, ..., T_{\max}\}$ **do**
    $\alpha_t \leftarrow \max\{\frac{50 \cdot \alpha_0}{50 + \log_{10}(t)}, 0.005\}$
    $\theta_t \leftarrow \frac{\theta_0}{t}$
    for $\mathbf{c} \in \mathcal{C}$ **do**
      $y \leftarrow \mathbf{c} \cdot \mathbf{w}$
      if $\|\mathbf{c}\| > \epsilon$ **then**
        $\alpha_t \leftarrow \frac{\alpha_0}{\|\mathbf{c}\|^2}$
      end if
      $\mathbf{w} \leftarrow \text{Dale}(\text{update}(\mathbf{c}, \mathbf{w}, \alpha_t, \theta_t, \eta))$
    end for
    if $\|\underline{C}\mathbf{w}'\| < \epsilon$ **then**
      break
    end if
    $t \leftarrow t + 1$
  end for
  for components, w_i of $\mathbf{w}$ **do**
    if $|w_i| \leq \epsilon$ **then**
      $\mathbf{w}_i \leftarrow 0$
    end if
  end for
end for

4.5 De-noising

We implemented a Bit Flipping style neural de-noising process. In all configurations considered here (and in [19, 20, 21]), the bit flipping algorithm has proven to perform no worse that winner-take-all. Moreover, since it requires only the implementation of additional parallel thresholding operations for each pattern neuron, its implementation is no less biologically feasible. The goal of this algorithm is to recover the correct activity pattern, $\mathbf{x}$, which has been corrupted by noise, and as such, is represented by a noisy version, $\mathbf{x}_n = \mathbf{x} + \mathbf{n}$, where $\mathbf{n}$ is this noise pattern. Since each weight vector is nearly perpendicular to every pattern, for a matrix of weights, W , $\mathbf{x}_n W'$ reveals inconsistencies in $\mathbf{x}_n$, which the de-noising algorithm seeks to correct in the feedback stage [1]. The clustered de-noising process begins with algorithm 4.5, in which each cluster attempts to detect errant pattern neurons. If no errors are detected, the process is complete. Otherwise, Algorithm 4.5 is invoked for each cluster that detected errant neurons. Note that this de-noising mechanism differs from error correction methods presented in [5] and [22] in that information contributed by place cells only reaches grid cells through interneurons, and place information contributed by grid cells at module i only reaches other modules through interneurons if connectivity (determined by correlations in the neural code) allows.

[1] To see this, consider that $\mathbf{x}_n W' = (\mathbf{x} + \mathbf{n})W' = \mathbf{x}W' + \mathbf{n}W' \approx 0 + \mathbf{n}W'$

Algorithm 3 Sequential de-noising

Input: local weights, W_i, for each cluster, $i \in \{1, ..., M\}$, noisy pattern, $\mathbf{x}_n$, stopping threshold, ϵ
Output: denoised pattern, $\mathbf{x}_d$

```
x_d ← x_n
while t < T_max or a cluster has an unsatisfied constraint do
    for each cluster, i ∈ {1, ..., M} do
        x ← subpattern corresponding to cluster i
        d ← Modular_Recall(x, W_i)
        if |dW_i| ≤ ε then
            x_d(cluster i's subpattern indices) ← d
        end if
    end for
    t ← t + 1
end while
```

Algorithm 4 Modular recall

Input: local weights for this cluster, W, maximum number of iterations, $T_{\max}$, noisy subpattern, $\mathbf{x}$, feedback threshold, ϕ
Output: denoised subpattern, $\mathbf{d}$

```
d ← p
while t < T_max do
    y ← xW'
    r ← y'W
    if ||y|| < ε then
        break;
    end if
    f ← (|y'|·|W|) / (Σ_{i=1}^{m} |W|)
    for each pattern neuron,j do
        if f_j ≥ φ then f_j = sign(x_j)
        else f_j = 0
        end if
    end for
     d ← d + f
end while
```

4.6 Interaction information

Here we will consider activities (i.e. firing rates) of members of populations of cells as random variables. We will use here the traditional definitions of shannon entropy, Kullback-Leibler divergence, and mutual informations. Following the examples of Mcgill and others (as discussed in [27]), we will define interaction information between random variables X, Y, and Z as

$$\begin{aligned} II(X;Y;Z) &= I(X;Y|Z) - I(X;Y) \\ &= I(X;Z|Y) - I(X;Z) \\ &= I(Z;Y|X) - I(Z;Y) \end{aligned}$$

When $II(X;Y;Z) > 0$, we say that the variables interact synergistically (as this implies that knowedge X enhances quantity of information transmitted by Y about Z and by Z about Y). Otherwise, we say

that the variables interact redundantly. This is also intuitive since this inequlity implies that knowledge of X does not improve quantity of information about Z carried by Y, nor does it improve quantity of information carried by Z about Y.

4.7 Transfer entropy and directed information

Transfer entropy is measure of quantity of information transferred between two stochastic processes originally developed to quantify statistical coherence of time varying systems. For random processes denoted by X_t and Y_t, transfer entropy from X_t to Y_t is defined as

$$T_{X\to Y} = H(Y_t|Y_{(t-1):(t-L)}) - H(Y_t|Y_{(t-1):(t-L)}, X_{(t-1):(t-L)})$$

, where L defines the length of recorded history of X_t and Y_t. It is established that transfer entropy better captures information transfer between two stochastic processes described by time series than time lagged mutual information [18] and [27].

Another information theoretic measurement that quantifies the extent to which information appears to flow from a random process, $X_{1:n}$, to $Y_{1:n}$, given as $I(X_{1:n} \to Y_{1:n}) = \sum_{i=1}^{n} I(X_{1:n}; Y_i|Y_{i-1})$. Directed information can be thought of as the amount of information that flows between two random processes. Moreover, Massey demonstrated that in a channel with feedback, diredcted information provides a more useful characterization of a channel than mutual information. Specifically, he showed that when the channel in question (i.e. the distributions linking $X_{1:n}$ and $Y_{1:n}$) experiences feedback, directed information gives a tighter upper bound on information $Y_{1:n}$ encodes about $X_{1:n}$ [10].

5 Results

5.1 Interaction information results

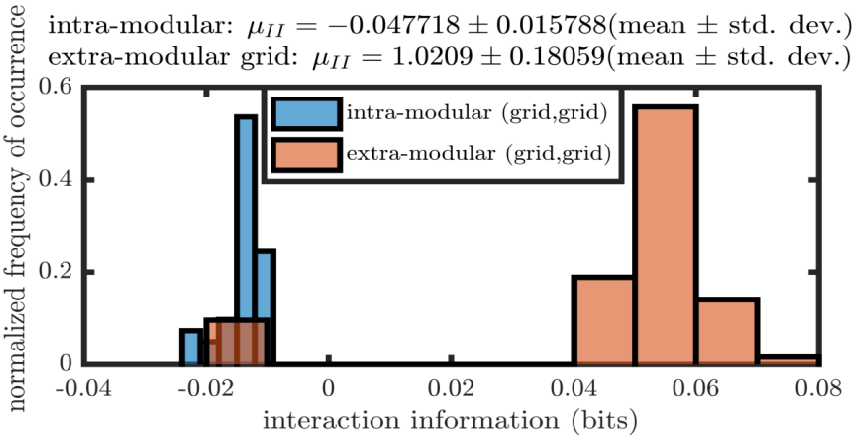

Figure 7: Shown here is the observed distribution of interaction informations between pairs of grid cells whose members belong to the same module, and pairs whose members belong to different modules; results here computed from 10 model networks consisting of 4 modules, with 20 grid cells per module

We compute interaction information between pairs of grid cells residing in the same module, and pairs whose members reside in different modules, over ten networks of randomly initialized grid cells, following the deliberate biological scheme, in which grid cell phase and orientation offests, and the scales are chosen so as to mimic the empirical observations of [23]. In order to assess a network's synergy, we compute the average interaction information across all cell pairs, μ_{II}. Interestingly, we found that intra-modular grid cell pairs are mostly redundant ($\mu_{II} = 0.040 \pm 0.016$ bits - mean ± standard deviation). On the other hand, extra-modular grid cell pairs tended strongly towards synergy ($\mu_{II} = 1.021 \pm 0.181$). Most interestingly, if we neglect to enforce the deliberate biological parameterization of the grid cell network, we observe no

significant difference in average interaction information between intra - and extra - modular pairs. From this, we predict that de-noising and de-coding will be impaired if synergistic populations of grid cells are unable to communicate across modules, or if redundant populations are unable to support eachother. We

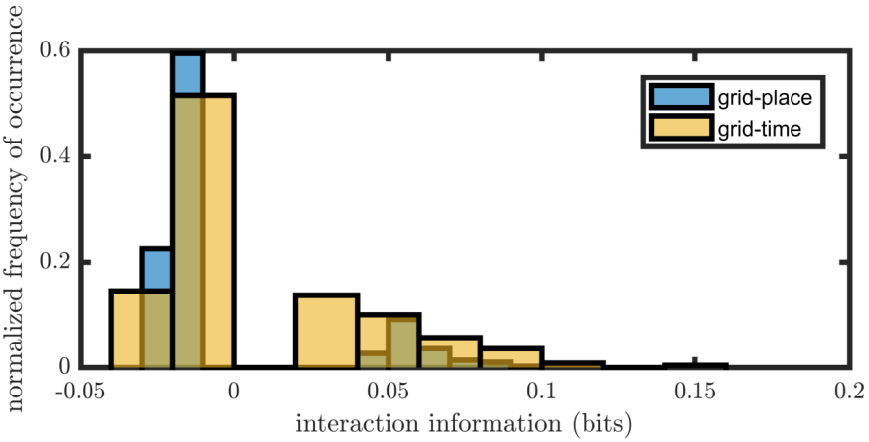

Figure 8: Shown here are observed interaction informations between pairs of cells consisting of one grid cell and one place cell, as well as those consisting of one grid cell and one time cell; results here computed from the same networks as in figure 7, augmented with a population of 20 place cells and 20 time cells

also computed interaction information of pairs consisting of one grid cell and one place cell, as well as those consisting of one grid cell and one time cell. Interestingly, the two follow qualitatively similar trends, and have no significant preference for synergy or redundancy. We observed similar results in pairs consisting of two distinct place cells, two distinct time cells, as well as those consisting of a place cell and a time cell. In each of these cases, the standard deviation of the interaction informations in quetion was at least one order of magnitude larger than $|\mu_{II}|$.

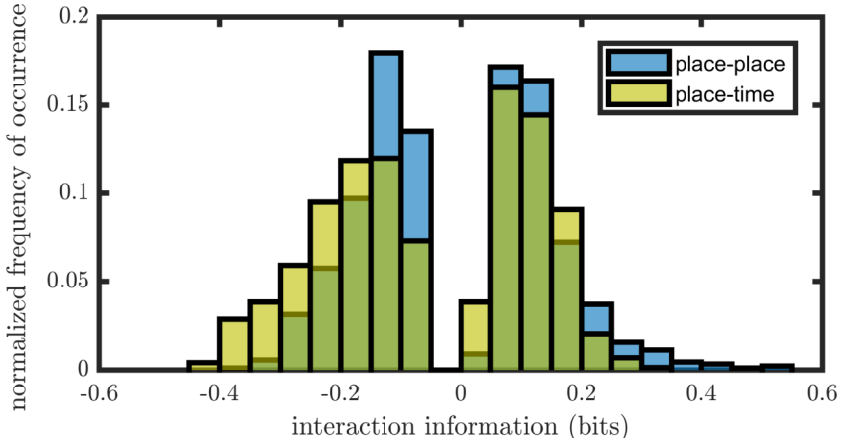

Figure 9: Shown here are observed interaction informations between pairs of place cells, and pairs consisting of one place cell and one time cell; results here computed from the same networks as in figure 8

5.2 Denoising results

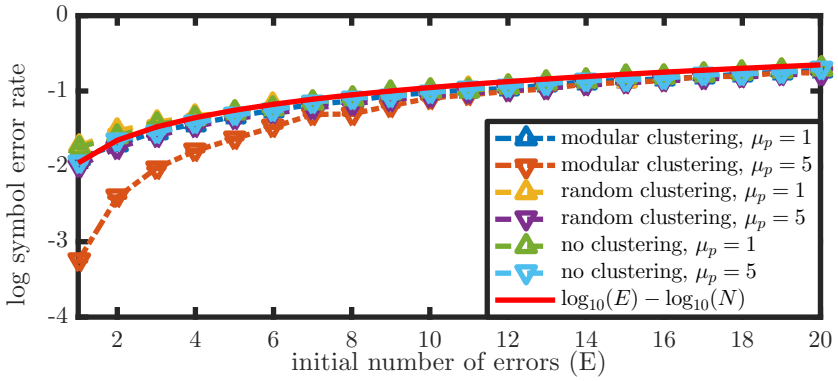

Figure 10: Log symbol error rate vs. initial number of errors (E) for clustered and non-clustered hybrid codes; here, each code utilizes a uniform distribution of grid cells to modules, and deliberately chosen spatial phases and orientations

Define μ_p, a code's spatial phase multiplicity, which we define to be the number of grid cells with the same phase in the same module. Figure 10 shows symbol error rates of hybrid codes for several configurations with deliberately chosen grid cell phases and orientations. This demonstrates that generally, clustered de-noising networks do not offer improved symbol error rate, $\mathrm{P_{se}}$, compared to their un-clustered counterparts. However, for a small initial number of errors (E), when the grid cells exhibit sufficient redundancy in their phases, a randomly clustered de-noising network is only outperformed by a modularly clustered network. Plotted in figure 10 is a red, solid curve, $\log_{10}(\frac{\mathrm{E}}{N})$. This curve is a threshold between regions of desirable and unacceptable P_{se} (i.e. $\log_{10}(P_{\mathrm{se}})$ for a network that performs no de-noising). To see this, consider a de-noising network that does not correct the E initial errors. For this network, $P_{\mathrm{se}} = \frac{E}{N}$, so $\log_{10}(P_{\mathrm{se}}) = \log_{10}(E) - \log_{10}(N)$. Surprisingly, figure 10 shows that for a small E, configurations with $\mu_p = 1$ have $\log_{10}(P_{\mathrm{se}})$ above this threshold, that is, they increase the number of symbol errors! From this, we conclude that the respecting the organization of grid cells into modules, grouped by spatial period, provides the de-noising network with better information (about the stimulus, in this case, space and time) than not. Thise confirms our earlier prediction that if redundant populations of pairs of grid cells (i.e. populations of intra-modular pairs) are unable to support eachother, de-noising and de-coding will be impaired, as the randomly clustered network, which leaves grid cells from the same module disorganized across clusters, performs significanly worse than the deliberately organized de-noising network, even when the grid cell network has tremendous redundancy ($\mu_p = 5$).

5.3 Transfer entropy results

Preliminary transfer entropy calculations are inconclusive. More precisely, none of the de-noising networks I have tested so far generate sufficient de-noising history that the transfer entropy estimation techniques (discussed in further detail in [27]) converge and declare acceptable significance. This estimation of transfer entropy performs a validation step to compute significance by shuffling the values of the random processes in time, and computing expected transfer entropy from chance. More simulations are required before conclusions can be made about transfer entropy during de-noising, for every network configuration described here.

6 Discussion

In our investigation of interaction information between pairs of navigational cells in the mammalian brain, we observed several intuitively explained trends. Firstly, it feels quite intuitive that pairs of grid cells residing in the same module should tend to fire redundantly, while pairs of cells residing in different modules should tend to fire synergistically. These observations of simulated cell activities predict that

we should see the same trend in real pairs of grid cells. Work to confirm or refute this hypothesis is already underway with analysis of a portion of the cell activities analyzed in [23], acquired with permission from the authors. Additionally, simultaneous single cell recordings of place cell activity were collected and curated by the computational and experimental neuroscience lab, in the psychology department at our own University of Arizona. Unfortunately, given the difficulty of recording single cell activity from MEC and HC simultaneously, procuring data from which we can compute interaction information of heterogenously typed cell pairs in vivo in the traditional manner (i.e. with the insertion of tetrodes and the recording of individual spikes on wires) remains fantasy. However, with recent advances in optogenetics, which enable single cell resolution of network activity for a population of inoculated cells (e.g. a collection of grid cells, as in [25]), we may be able to obtain these desired data, despite the technical difficulty of imaging simultaneous activity of grid and place cells at high temporal precision [7]. After inferring spikes from these images, simultaneous firing rates of all innoculated and imaged cells may be estimated, and these interaction informations computed [26]. We observed no significant tendency towards synergy or redundancy in homogenous pairs of distinct place cells and distinct time cells. These observations, upon extension to the empirically observed behavior of these cells, become hypotheses. A first step towards validating this model should invovle testing these hypotheses, as well as those outlined in [20, 19, 21]. Additional investigation also involves the use of interaction informations and possibly other information theoretic measurements, to analytically develop fundamental bounds on performance of the de-noising network discussed here.

Unfortunately, at the time of submission, no significant transfer entropies were observed in any of the de-noising epochs tested. This stems both from the computational complexity of transfer entropy estimation, and the lack of long de-noising epochs. It appears that the transfer entropy estimation techniques developed in [27] may be inappropriate for the random processes considered in this work (i.e. time-dependent activities of the cells through the de-noising process). However, further research, and correspondence with authors of [27] are necessary before this conclusion may be solidified. An alternative information theoretic measurement that may be employed to characterize flow of information during de-noising is directed information, disucssed in section 4. This quantity should prove to demand less computational overhead as compared to transfer entropy. Additionally, given the feedback inherent in the de-noising process, directed information may provide a more powerful understanding of network information flow during de-noising[10]. This work is already under way: most recently passing the milestone of adapting the matlab toolbox provided by the authors of [9] to ensure safe interface with the simulations discussed here. At the time of submission, these simulations have yet to complete. We hope that directed information during de-noising will help explain the stochastic resonance observed and described in [20], whose pathology remains a mystery. A more complete understanding of HC-MEC communication is highly sought after in certain spheres of computational neuroscience. By testing the hypotheses presented here (when possible as technology and the availability of data allows) against empirical observations, the completion of this work via analysis of the recently procured datasets, and the evaluation of directed information flow during de-noising may move us towards accomplishing this goal. Completion of the aforementioned work, as well as submission of the results to a journal of the highest esteem that will publish it, is expected by the end of the summer.

7 Appendix

7.1 Subspace learning

In [13], the authors propose an algorithm that is capable of computing a basis for the null space of a random matrix, A, which is assumed to be the expected value of sample matrices, A_t. The update rule for the matrix whose columns are the resulting basis vectors is

$$\tilde{W}_t = W_{t-1} + A_{t-1} W_{t-1} \boldsymbol{\alpha}_{t-1} \tag{1}$$

$$W_t = \tilde{W}_t R_t^{-1}, \tag{2}$$

where $\boldsymbol{\alpha}_t$ is a diagonal (and compatible) matrix of gain factors. As in [13], equations 1 and 2 may be re-written as operations on column vectors, $\mathbf{w}_t$.

$$\tilde{\mathbf{w}}_t = \mathbf{w}_{t-1} + \alpha_{t-1} A_{t-1} \mathbf{w}_{t-1} \tag{3}$$

$$\mathbf{w}_t = \frac{\tilde{\mathbf{w}}_t}{\|\tilde{\mathbf{w}}_t\|}, \tag{4}$$

in which α_t is the gain factor corresponding to the current column. This number may be equivalently understood as a learning rate. Indeed in [28], the authors show that for appropriate choices of A_t, the update rule is a form of anti-Hebbian learning. In [13] the authors prove convergence of this algorithm to the eigenvectors of A corresponding to the largest eigenvalues. Further, when A_t is replaced by $-A_t$, $\mathbf{w}_t$ converges to the eigenvectors of A corresponding to the smallest eigenvalues. In [13], it is demonstrated that by combining equations 3 and 4, expanding as a power series in α_t, and ignoring second (and higher) order terms, we arrive at

$$\mathbf{w}_t = \mathbf{w}_{t-1} + \alpha_{t-1}(A_{t-1}\mathbf{w}_{t-1} - \frac{\mathbf{w}_t^T A_{t-1} \mathbf{w}_{t-1}}{\mathbf{w}_{t-1}^T \mathbf{w}_{t-1}} \mathbf{w}_{t-1}). \tag{5}$$

The authors of [16] choose $A_t = (\mathbf{x}_t^T \mathbf{x}_t) P_{\mathbf{x}_t} = \mathbf{x}_t \mathbf{x}_t^T$, the product of projections onto the space spanned by $\mathbf{x}_t$, and define $y_t = \mathbf{x}_t^T \mathbf{w}_t = \mathbf{w}_t^T \mathbf{x}_t$. In [13], it is mentioned that this update rule finds eigenvectors corresponding to the largest eigenvalue of A_t, or those corresponding to the smallest eigenvalues of $-A_t$, when this matrix is used instead. Since A_t is a projection matrix, it has rank 1. Thus it has one eigenvector with non-zero eigenvalue, $\mathbf{x}_t$, and $\dim(\mathbf{x}) - 1$ eigenvectors with eigenvalue 0. Each of these eigenvectors, $\mathbf{v}$, is guaranteed to be perpendicular to $\mathbf{x}$ because $A_t \mathbf{v} = 0\mathbf{v} = \mathbf{0}$, that is, the $\mathbf{v}$'s projection onto $\mathbf{x}$ has magnitude 0. By choosing $\mathbf{x}_t \in \mathcal{C}$, with the aforementioned choice for A_t, this algorithm should compute vectors approximately perpendicular to the code space.
Now, we may rewrite equation 5 as

$$\begin{aligned}
\mathbf{w}_t &= \mathbf{w}_{t-1} - \alpha_{t-1} \mathbf{x}_{t-1} \mathbf{x}_{t-1}^T \mathbf{w}_{t-1} + \alpha_{t-1} \frac{\mathbf{w}_{t-1}^T \mathbf{x}_{t-1} \mathbf{x}_{t-1}^T \mathbf{w}_{t-1}}{\|\mathbf{w}_{t-1}\|^2} \mathbf{w}_{t-1} \\
&= \mathbf{w}_{t-1} - \alpha_{t-1} y_{t-1} \mathbf{x}_{t-1} + \alpha_{t-1} \frac{y_{t-1}^2}{\|\mathbf{w}_{t-1}\|^2} \mathbf{w}_{t-1}.
\end{aligned} \tag{6}$$

To obtain a sparse basis for null($\underline{\mathbb{C}}$), one may add to equation 6 a regularizing term that penalizes non-sparse solutions. In particular, using $\eta\Gamma(\mathbf{w}_{t-1}, \theta_{t-1})$, as considered in [16], to arrive at

$$\mathbf{w}_t = \mathbf{w}_{t-1} - \alpha_{t-1}(y_{t-1}(\mathbf{x}_{t-1} - \frac{y_{t-1}\mathbf{w}_{t-1}}{\|\mathbf{w}_{t-1}\|^2})) - \alpha_{t-1}\eta\Gamma(\mathbf{w}_{t-1}, \theta_{t-1}). \tag{7}$$

7.2 Choices of parameters

In learning, normalized weights are initialized randomly with degree $\lceil 4 \log_e(n) \rceil$, where n is the length of the weight vector. We used, $\theta_0 = 0.031$, $\epsilon = C10^{-3}$, $\eta = 0.075$, and $\alpha_0 = 0.95$. In de-noising, we set $\phi = 0.95$.

8 Acknowledgements

I would like to thank my thesis advisor, Dr. Koyluoglu, for his support and advice in the development of [19, 20, 21]. I would also like to thank Dr. Tandon for approving this work for the network information theory project, as without this approval, the investigations described herein would have suffered delay until the completion of my thesis, possibly until the completion of the Spring 2017 semester.

References

[1] Tora Bonnevie, Benjamin Dunn, Marianne Fyhn, Torkel Hafting, Dori Derdikman, John L Kubie, Yasser Roudi, Edvard I Moser, and May-Britt Moser. Grid cells require excitatory drive from the hippocampus. *Nature neuroscience*, 16(3):309–317, 2013.

[2] Emery N Brown, Loren M Frank, Dengda Tang, Michael C Quirk, and Matthew A Wilson. A statistical paradigm for neural spike train decoding applied to position prediction from ensemble firing patterns of rat hippocampal place cells. *Journal of Neuroscience*, 18(18):7411–7425, 1998.

[3] John Eccles. From electrical to chemical transmission in the central nervous system. *Notes and records of the Royal Society of London*, 30(2):219–230, 1976.

[4] Arne D Ekstrom, Michael J Kahana, Jeremy B Caplan, Tony A Fields, Eve A Isham, Ehren L Newman, and Itzhak Fried. Cellular networks underlying human spatial navigation. *Nature*, 425(6954):184–188, 2003.

[5] Ila R Fiete, Yoram Burak, and Ted Brookings. What grid cells convey about rat location. *The Journal of Neuroscience*, 28(27):6858–6871, 2008.

[6] Marianne Fyhn, Torkel Hafting, Menno P Witter, Edvard I Moser, and May-Britt Moser. Grid cells in mice. *Hippocampus*, 18(12):1230–1238, 2008.

[7] Benjamin F Grewe, Dominik Langer, Hansjörg Kasper, Björn M Kampa, and Fritjof Helmchen. High-speed in vivo calcium imaging reveals neuronal network activity with near-millisecond precision. *Nature methods*, 7(5):399–405, 2010.

[8] Joshua Jacobs, Christoph T Weidemann, Jonathan F Miller, Alec Solway, John F Burke, Xue-Xin Wei, Nanthia Suthana, Michael R Sperling, Ashwini D Sharan, Itzhak Fried, et al. Direct recordings of grid-like neuronal activity in human spatial navigation. *Nature neuroscience*, 16(9):1188–1190, 2013.

[9] Jiantao Jiao, Haim H Permuter, Lei Zhao, Young-Han Kim, and Tsachy Weissman. Universal estimation of directed information. *IEEE Transactions on Information Theory*, 59(10):6220–6242, 2013.

[10] James Massey. Causality, feedback and directed information. In *Proc. Int. Symp. Inf. Theory Applic.(ISITA-90)*, pages 303–305. Citeseer, 1990.

[11] Patrick Mineault. Using an iterated extended kalman filter to decode place cells. *xcorr: comp neuro - computational neuroscience, machine learning*, 2012.

[12] Laurenz Muessig, Jonas Hauser, Thomas Joseph Wills, and Francesca Cacucci. A developmental switch in place cell accuracy coincides with grid cell maturation. *Neuron*, 86(5):1167–1173, 2015.

[13] Erkki Oja and Juha Karhunen. On stochastic approximation of the eigenvectors and eigenvalues of the expectation of a random matrix. *Journal of mathematical analysis and applications*, 106(1):69–84, 1985.

[14] Erkki Oja and Teuvo Kohonen. The subspace learning algorithm as a formalism for pattern recognition and neural networks. In *Neural Networks, 1988., IEEE International Conference on*, pages 277–284. IEEE, 1988.

[15] John O'Keefe. Place units in the hippocampus of the freely moving rat. *Experimental neurology*, 51(1):78–109, 1976.

[16] Amir Hesam Salavati, K Raj Kumar, and Amin Shokrollahi. Nonbinary associative memory with exponential pattern retrieval capacity and iterative learning. *Neural Networks and Learning Systems, IEEE Transactions on*, 25(3):557–570, 2014.

[17] Daniel M Salz, Zoran Tiganj, Srijesa Khasnabish, Annalyse Kohley, Daniel Sheehan, Marc W Howard, and Howard Eichenbaum. Time cells in hippocampal area ca3. *Journal of Neuroscience*, 36(28):7476–7484, 2016.

[18] Thomas Schreiber. Measuring information transfer. *Physical review letters*, 85(2):461, 2000.

[19] David Schwartz and O. Ozan Koyluoglu. A hybrid code from grid and place cells. In *Neuroscience 2016*, number 183.26. Neuroscience, 2016.

[20] David Schwartz and O. Ozan Koyluoglu. Neural noise improves path representation in a simulated network of grid, place, and time cells. Number III-42 in Cosyne abstracts, 2017.

[21] David Schwartz and O. Ozan Koyluoglu. On the organization of grid and place cells: Neural de-noising via subspace learning. submitted.

[22] Martin Stemmler, Alexander Mathis, and Andreas VM Herz. Connecting multiple spatial scales to decode the population activity of grid cells. *Science Advances*, 1(11):e1500816, 2015.

[23] Hanne Stensola, Tor Stensola, Trygve Solstad, Kristian Frøland, May-Britt Moser, and Edvard I Moser. The entorhinal grid map is discretized. *Nature*, 492(7427):72–78, 2012.

[24] Bryan A. Strange, Menno P. Witter, Ed S. Lein, and Edvard I. Moser. Functional organization of the hippocampal longitudinal axis. *Nat Rev Neurosci*, 15(10):655–669, October 2014.

[25] Chen Sun, Takashi Kitamura, Jun Yamamoto, Jared Martin, Michele Pignatelli, Lacey J. Kitch, Mark J. Schnitzer, and Susumu Tonegawa. Distinct speed dependence of entorhinal island and ocean cells, including respective grid cells. *Proceedings of the National Academy of Sciences*, 112(30):9466–9471, 2015.

[26] Lucas Theis, Philipp Berens, Emmanouil Froudarakis, Jacob Reimer, Miroslav Román Rosón, Tom Baden, Thomas Euler, Andreas S Tolias, and Matthias Bethge. Benchmarking spike rate inference in population calcium imaging. *Neuron*, 90(3):471–482, 2016.

[27] Nicholas Timme, Wesley Alford, Benjamin Flecker, and John M Beggs. Synergy, redundancy, and multivariate information measures: an experimentalists perspective. *Journal of computational neuroscience*, 36(2):119–140, 2014.

[28] Lei Xu, Adam Krzyzak, and Erkki Oja. Neural nets for dual subspace pattern recognition method. *International Journal of Neural Systems*, 2(03):169–184, 1991.

[29] Michael M Yartsev, Menno P Witter, and Nachum Ulanovsky. Grid cells without theta oscillations in the entorhinal cortex of bats. *Nature*, 479(7371):103–107, 2011.

[30] Kechen Zhang, Iris Ginzburg, Bruce L McNaughton, and Terrence J Sejnowski. Interpreting neuronal population activity by reconstruction: unified framework with application to hippocampal place cells. *Journal of neurophysiology*, 79(2):1017–1044, 1998.