

## Understanding encoding and redundancy in grid cells using partial information decomposition

**Summary:** The brain is capable of performing reliable computation using neural circuitry that has high internal variability: this can only be possible if the brain controls for this variability through redundancy. Understanding such mechanisms will likely require quantifying the degree of uniqueness, redundancy and synergy that exists in the neural code. Recently, the literature in information theory has developed a structured framework called Partial Information Decomposition (PID) [Bertschinger et al. 2014; Lizier et al., 2018], which defines the unique, redundant and synergistic interactions between two or more variables in conveying information about a particular message. While the idea of decomposing mutual information into unique, redundant and synergistic components has roots in neuroscience [Schneidman et al., 2003], it has not received significant attention therein since these recent developments. In this abstract, we show two instances in which the PID provides insights into neural encoding, using computational models of grid cells based on prior work by Sreenivasan and Fiete [2011]. (1) We identify redundant information in the grid cell code, which we associate with error correction capability, and identify redundant *and* unique information when error correction is in effect. (2) We find that determining the animal's location at a coarse spatial resolution requires the aggregation of synergistic information from *just as many* grid cell networks, as required to determine location at a much finer resolution: this might have implications for downstream networks that utilize grid cell outputs [Fiete et al., 2008]. More generally, our results suggest that computing the PID in different settings might provide fine-grained insights into neural coding and function that are harder to attain with simpler correlation-based analyses. For instance, recent work by Venkatesh et al. [2019] explicitly connects the PID with a formal definition for information flow in the brain.

**Methods:** We first provide an overview of the Partial Information Decomposition framework. The bivariate PID describes how the mutual information between a “message” ( $M$ ) and two other random variables ( $X$  and  $Y$ ) can be decomposed into four non-negative components: information about  $M$  that is (i) *unique to  $X$* ; (ii) *unique to  $Y$* ; (iii) *redundant*, which can be recovered from either  $X$  or  $Y$ ; and (iv) “*synergistic*”, which can only be recovered when  $X$  and  $Y$  are taken together:

$$I(M; (X, Y)) = UI(M: X \setminus Y) + UI(M: Y \setminus X) + RI(M: X; Y) + SI(M: X; Y) \quad (A)$$

Bertschinger et al. [2014] provided a mathematical definition for unique information which satisfies many intuitively desirable properties, and also determines the values of the redundant and synergistic components:

$$UI(M: X \setminus Y) = \min_{q \in \Delta(p)} I_q(M; X|Y) \quad \text{where } \Delta(p) = \{q: q(m, x) = p(m, x), q(m, y) = p(m, y)\} \quad (B)$$

Here,  $p$  is the true joint distribution of  $(M, X, Y)$ , and  $I_q(M; X|Y)$  is the conditional mutual information between  $M$  and  $X$ , given  $Y$ , under the joint distribution  $q$ .  $\Delta(p)$  is the set of all distributions that have the same  $(M, X)$  and  $(M, Y)$  marginals as  $p$ . It can be shown that the above is a convex optimization problem in the space of distributions, and is hence easily computed for discrete distributions with modest support.

Next, we provide a brief overview of the grid cell code. Fiete et al. [2008], and later, Sreenivasan and Fiete [2011], described the distinctive “modulo code” employed by grid cells, and explained how this neural encoding scheme forms a robust error-correcting code. Grid cells are organized into multiple “grid networks”: each network forms a population code for the *phase* of the animal's location within some fixed wavelength. Different networks have different wavelengths, which are “coprime” with respect to each other (see Fig. 1a). For a fixed precision, the number of locations these networks jointly encode can be as large as the product of their wavelengths (e.g., three grid networks with wavelengths 3, 4 and 5 units can encode up to 60 locations with 1-unit precision in 1D). However, by choosing to use the same number of networks to encode a smaller range of locations (e.g., 12), the brain may introduce redundancy in encoding, and hence error resilience. For example, such a code may be robust to “readout error”, the error seen by a downstream neuron when decoding population activity. Based on the work of Sreenivasan and Fiete, we create a computational model of three grid networks, with wavelengths of  $\lambda = (3, 4, 5)$  units. We specify the joint distribution of the erroneous phase readout from these networks, given the animal's location (see Fig. 1b), and compute the PID for this code, as described above in equations (A) and (B).

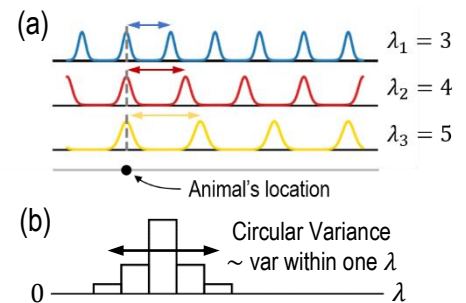


Fig 1. (a) Probability distributions of location, given readout phases of three grid networks with wavelengths  $\lambda_1=3$  (blue),  $\lambda_2=4$  (red),  $\lambda_3=5$  (yellow). Maximum total range is achieved for “coprime” wavelengths (here, the maximum range is 60 units). All three distributions peak *simultaneously* only at the animal's true location. (b) Discretized von Mises distribution used to model the noisy grid network readout. The “circular variance” of this distribution parameterizes the amount of variability in the grid network's readout (the maximum circular variance is 1, for the uniform distribution)

**Results:** For intuition, consider a simple classical error correcting code on boolean variables: we encode a 2-bit message  $M = (X_1, X_2)$  as  $[X_1, X_2, X_3]$ , where  $X_3 = X_1 \oplus X_2$  and “ $\oplus$ ” represents “XOR”. Any two encoded bits suffice to decode  $M$ , thus any  $X_i$  carries *redundant* information about  $M$  with respect to the other two variables. However, if we erase one of these three bits, the remaining two bits convey *unique* information about  $M$  (e.g., if  $X_2$  is erased,  $X_1$  carries information about  $M$ ’s first bit, while  $X_3$  independently carries information about whether  $X_1$  and  $X_2$  are equal, which is not present in  $X_1$ ).

We now examine the PID of the grid cell code in two separate instances. (1) First, we show how PID quantifies redundancy by comparing information content when the grid networks encode the whole range of possible locations,  $R = 60$  units, and when they encode only a limited range,  $R = 12$  units, in an error-resilient manner: (i) When the networks encode location in a range of  $R = 60$  units, the message  $M$  takes one of sixty values. Each network carries purely unique information about  $M$  with respect to the other two: in particular, the  $i^{th}$  network ( $i = 1, 2, 3$ ) carries  $\log_2(\lambda_i)$  bits of unique information, and the total mutual information is  $\log_2(60) \approx 5.9$  bits (Fig. 2a). As readout error increases, total mutual information drops significantly, losing  $>2$  bits within an error variance of 0.15 squared-units. (ii) When the networks encode a limited range of  $R = 12$  units, the  $i^{th}$  network carries *zero* unique information with respect to the others in the absence of noise, and instead carries  $\log_2(\lambda_i)$  bits of *redundant* information, as we saw with classical error correcting codes. As readout error increases, the total mutual information drops relatively gradually, losing only  $\sim 0.7$  bits for the same variance (Fig. 2b). The redundant component of information drops more sharply however, while the unique component rises, indicating that error correction is in play. In particular, the presence of redundancy across networks prevents a sudden drop in total information.

(2) Next, consider how the grid networks encode information about location at different spatial resolutions  $r$ , i.e., we divide  $R = 60$  units into bins of size  $r$ , and  $M$  now indicates *which bin* the animal is located in. Having examined the PID at a fine resolution of  $r = 1$  unit in (i), we ask how the PID behaves at a very coarse resolution of  $r = 30$  units, i.e., in determining whether the animal is in the left or right half of the domain. We find that all three networks *synergistically* encode information about left vs. right. No two networks fully specify which half of the space the animal is in (see Fig. 3a). Thus, one needs the same number of networks (here, 3) to say whether the animal is in a given 1-unit region, or within a 30-unit region. This can be shown to be true, no matter what integer resolution we choose between 1 and 30. Further, without redundancy as in (ii), even information about coarse spatial location is highly susceptible to readout error, and small amounts of noise result in a sharp drop in total mutual information (see Fig. 3b).

**Conclusions:** We examined two instances where the PID is able to provide a more nuanced understanding of how information is encoded by grid cells. In practice, estimates of the PID can be obtained for neural firing rates, however this could require significant amounts of data and computation, as it involves optimization over the high-dimensional space of distributions. Nevertheless, we expect this could be ameliorated by appropriate discretization, based on desired accuracy.

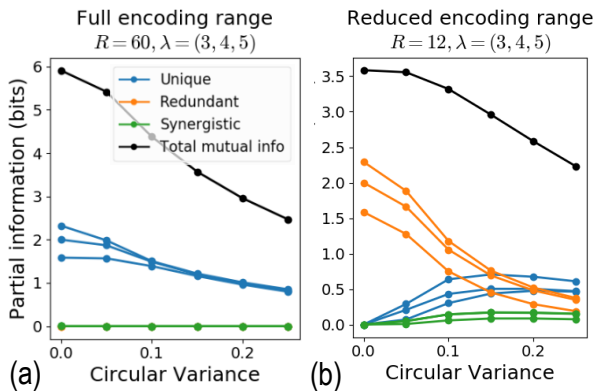


Fig 2. PID of location between each grid network and the two others. (a) PID for the full encoding range,  $R=60$  units. Note that all information is unique; information drops steadily with increase in error. (b) PID for a limited encoding range,  $R=12$  units. Information is initially redundant, but becomes unique as error increases. The drop in mutual information is slower than the drop in redundant information, indicating error correction.

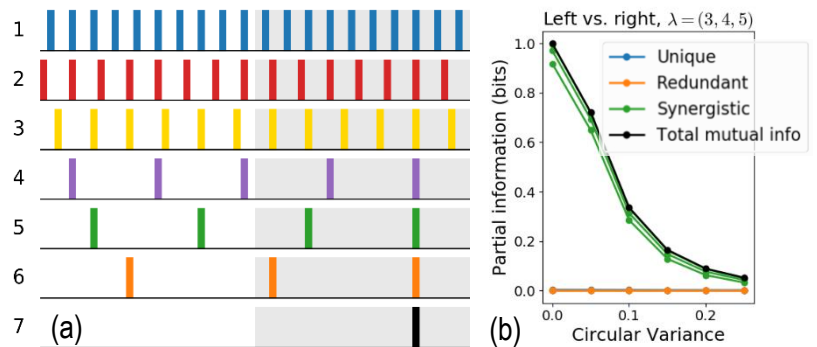


Fig 3. (a) Schematic depicting how no two networks  $\lambda = (3, 4, 5)$  fully specify even coarse spatial information, i.e. left vs. right (shaded area). Each row shows uncertainty in location given one or more grid networks’ readouts: 1- $\lambda_1$ , 2- $\lambda_2$ , 3- $\lambda_3$ , 4- $(\lambda_1, \lambda_2)$ , 5- $(\lambda_1, \lambda_3)$ , 6- $(\lambda_2, \lambda_3)$ , 7- $(\lambda_1, \lambda_2, \lambda_3)$ . For larger wavelengths, probability of being inside or outside of the shaded region would be roughly equal, unless given all networks’ readouts. (b) PID of information about left vs. right: only synergistic information is present, quantitatively validating the schematic in (a).