Characterizing information stored in synaptic connections rather than firing activities: An analytical information-theoretic framework

Xinhao Fan[1], Shreesh P. Mysore[1,2,3]

[1]Department of Neuroscience, [2]Department of Psychological and Brain Sciences, [3]Kavli Neuroscience Discovery Institute, Johns Hopkins University, Baltimore, MD; | xfan20@jhu.edu

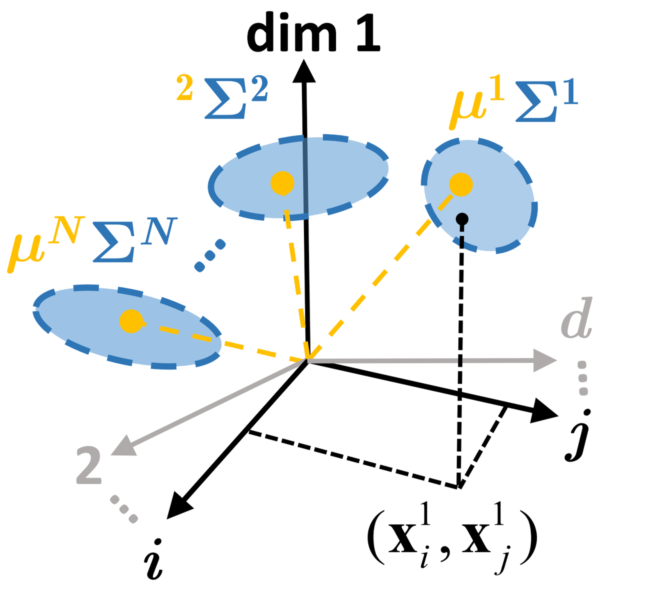
Background

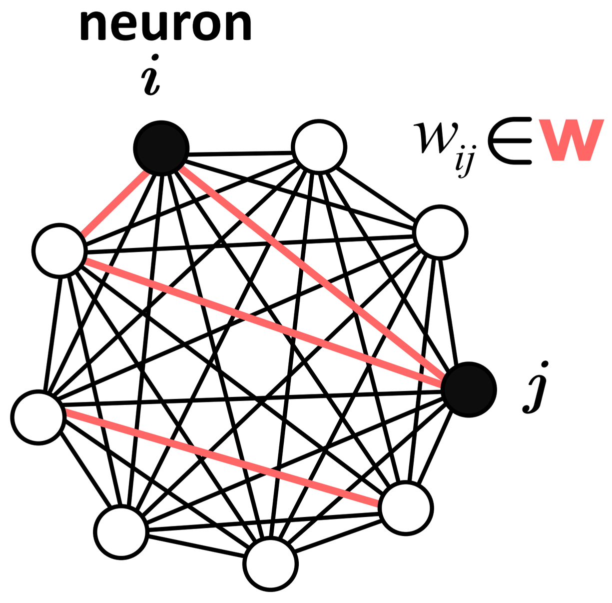
Studying **information encoded in the brain** has a long history, with most, if not all, research focusing on **neural firing activity**[1,2,3].
An alternative view is that information is stored in **synaptic connections**, and spiking activity only reflects changes in these connections[4]. It has been much less explored.

Goals

1. **An analytical solution** for the **information stored in synaptic connections**.
2. **Neuroscientific insights** through this dual perspective of connections.

Model





Input data are modeled as N **data patterns**, each represented as a d-dimensional random variable:

$$\{\mathbf{x}^1, \dots, \mathbf{x}^N\}$$

Each pattern is hypothesized to follow a **multivariate log-normal distribution**

$$\ln(\mathbf{x}^k) \sim \mathcal{N}(\mu^k, \Sigma^k)$$

A **continuous Hopfield network**[5] is used to model the brain, characterized by the connectivity:

$$w_{ij} = \sum_{k=1}^{N} \mathbf{x}_i^k \mathbf{x}_j^k$$

An **ensemble** of n synaptic connections is denoted as:

$$\mathbf{w} = (w_{ij(1)}, \dots, w_{ij(n)})^T \qquad ij(1): i(1) \leftarrow j(1)$$

The goal is to derive the **information stored in an ensemble** about a pattern:

$$MI(\mathbf{w}; \mathbf{x}^l)$$

Analytical solutions

To get $MI(\mathbf{w}; \mathbf{x}^l)$, we need to derive $p(\mathbf{w})$ and $p(\mathbf{w}|\mathbf{x}^l)$:

$$MI(\mathbf{w}; \mathbf{x}^l) = h(\mathbf{w}) - h(\mathbf{w} \mid \mathbf{x}^l) = -\int p(\mathbf{w}) \log p(\mathbf{w}) d\mathbf{w} + \iint p(\mathbf{w}, \mathbf{x}^l) \log p(\mathbf{w} \mid \mathbf{x}^l) d\mathbf{w} d\mathbf{x}^l$$

- For $p(\mathbf{w})$, we proved that it can be approximated by a n-dimensional log-normal distribution:

$$\ln(\mathbf{w}) = \begin{pmatrix} \ln w_{ij(1)} \\ \vdots \\ \ln w_{ij(n)} \end{pmatrix} \sim \mathcal{N}\left(\mu_{\mathbf{w}} = \begin{pmatrix} \mu_{w_{ij(1)}} \\ \vdots \\ \mu_{w_{ij(n)}} \end{pmatrix}, \Sigma_{\mathbf{w}} = \begin{pmatrix} \sigma^2_{w_{ij(1)}} & \sigma_{w_{ij(1),ij(2)}} & \cdots & \sigma_{w_{ij(1),ij(n)}} \\ \sigma_{w_{ij(2),ij(1)}} & \sigma^2_{w_{ij(2)}} & \cdots & \sigma_{w_{ij(2),ij(n)}} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{w_{ij(n),ij(1)}} & \sigma_{w_{ij(n),ij(2)}} & \cdots & \sigma^2_{w_{ij(n)}} \end{pmatrix}\right)$$

The mean parameters $\mu_{\mathbf{w}}$ and diagonal values for $\Sigma_{\mathbf{w}}$ are:

$$\mu_{w_{ij}} = \ln M_{1,ij} - \frac{1}{2}\ln\left(1 + \frac{M_{2,ij}}{M_{1,ij}^2}\right) \qquad \sigma^2_{w_{ij}} = \ln\left(1 + \frac{M_{2,ij}}{M_{1,ij}^2}\right)$$

$$M_{1,ij} = \sum_{k=1}^{N} \exp\left(\mu_i^k + \mu_j^k + \frac{1}{2}\left[(\sigma_i^k)^2 + (\sigma_j^k)^2 + 2\sigma_{ij}^k\right]\right)$$

$$M_{2,ij} = \sum_{k=1}^{N} \left[\exp\left((\sigma_i^k)^2 + (\sigma_j^k)^2 + 2\sigma_{ij}^k\right) - 1\right] \exp\left(2(\mu_i^k + \mu_j^k) + (\sigma_i^k)^2 + (\sigma_j^k)^2 + 2\sigma_{ij}^k\right)$$

The off-diagonal elements for $\Sigma_{\mathbf{w}}$ are:

$$\sigma_{w_{ij(r),ij(s)}} = \ln\left(1 + \frac{\sum_{k=1}^{N} \exp\left(\sum_{m}^{\{r,s\}}\left(\sum_{\phi}^{\{i(m),j(m)\}} \mu_\phi^k + \frac{1}{2}\left(\sigma_\phi^k\right)^2\right) + \sigma_{ij(m)}^k\right)\left(\exp\left(\sum_{\phi}^{\{i(r),j(r)\}\times\{i(s),j(s)\}} \sigma_\phi^k\right) - 1\right)}{\sum_{k,l=1}^{N} \exp\left(\left(\sum_{\phi}^{\{i(r),j(r)\}} \mu_\phi^k + \frac{1}{2}\left(\sigma_\phi^k\right)^2\right) + \left(\sum_{\phi}^{\{i(s),j(s)\}} \mu_\phi^l + \frac{1}{2}\left(\sigma_\phi^l\right)^2\right) + \sigma_{ij(r)}^k + \sigma_{ij(s)}^l\right)}\right)$$

- For $p(\mathbf{w}|\mathbf{x}^l)$, we proved that it is isomorphic to $p(\mathbf{w}_{/l})$, in the sense one is a translated version of the other. The distribution $p(\mathbf{w}_{/l})$ follows:

$$\ln \mathbf{w}_{/l} = \begin{pmatrix} \ln w_{ij(1)/l} \\ \vdots \\ \ln w_{ij(n)/l} \end{pmatrix} \sim \mathcal{N}\left(\mu_{\mathbf{w}/l} = \begin{pmatrix} \mu_{w_{ij(1)}/l} \\ \vdots \\ \mu_{w_{ij(n)}/l} \end{pmatrix}, \Sigma_{\mathbf{w}/l} = \begin{pmatrix} \sigma^2_{w_{ij(1)}/l} & \cdots & \sigma_{w_{ij(1),ij(n)}/l} \\ \vdots & \ddots & \vdots \\ \sigma_{w_{ij(n),ij(1)}/l} & \cdots & \sigma^2_{w_{ij(n)}/l} \end{pmatrix}\right)$$

As a result, we have an analytical expression for the $MI(\mathbf{w}; \mathbf{x}^l)$:

Proposition. *The mutual information between the joint activity of n synaptic connections* $\mathbf{w}$ *and a data pattern* $\mathbf{x}^l$ *is:*

$$MI\left(\mathbf{w}; \mathbf{x}^l\right) = \sum_{k=1}^{n}\left(\mu_{w_{ij(k)}} - \mu_{w_{ij(k)}/l}\right) + \frac{1}{2}\ln\left|\Sigma_{\mathbf{w}}\right| - \frac{1}{2}\ln\left|\Sigma_{\mathbf{w}/l}\right|$$

Simulations for analyses

Idea: Exploring insights into neuroscience by studying how MI changes with parameter tuning. Two approaches are taken:

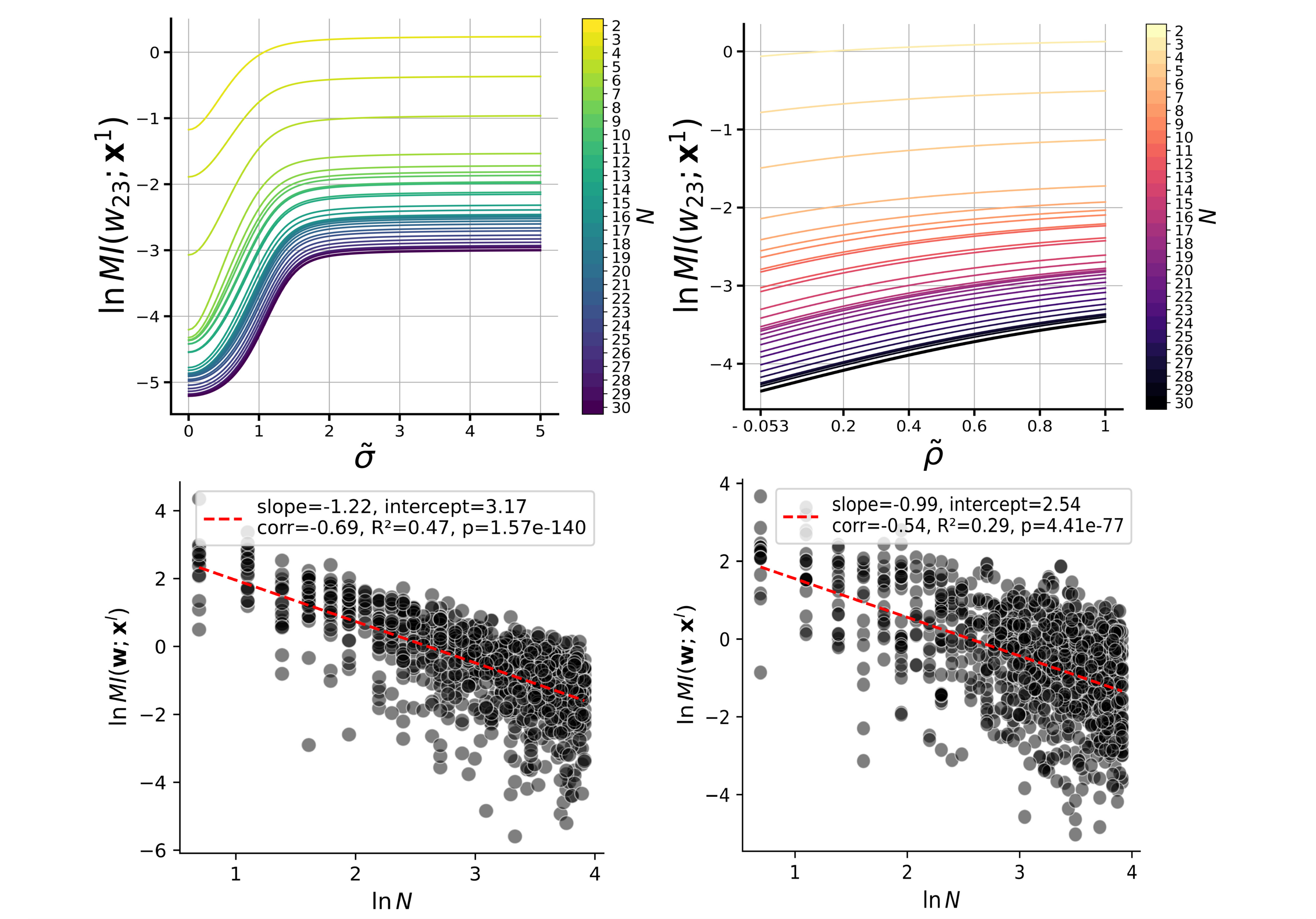
1. Systematically varying.

$$\Sigma^1 = \Sigma^2 = \dots = \Sigma^N = \begin{pmatrix} \tilde{\sigma}^2 & \tilde{\rho}\tilde{\sigma}^2 & \cdots & \tilde{\rho}\tilde{\sigma}^2 \\ \tilde{\rho}\tilde{\sigma}^2 & \tilde{\sigma}^2 & \cdots & \tilde{\rho}\tilde{\sigma}^2 \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{\rho}\tilde{\sigma}^2 & \tilde{\rho}\tilde{\sigma}^2 & \cdots & \tilde{\sigma}^2 \end{pmatrix}$$

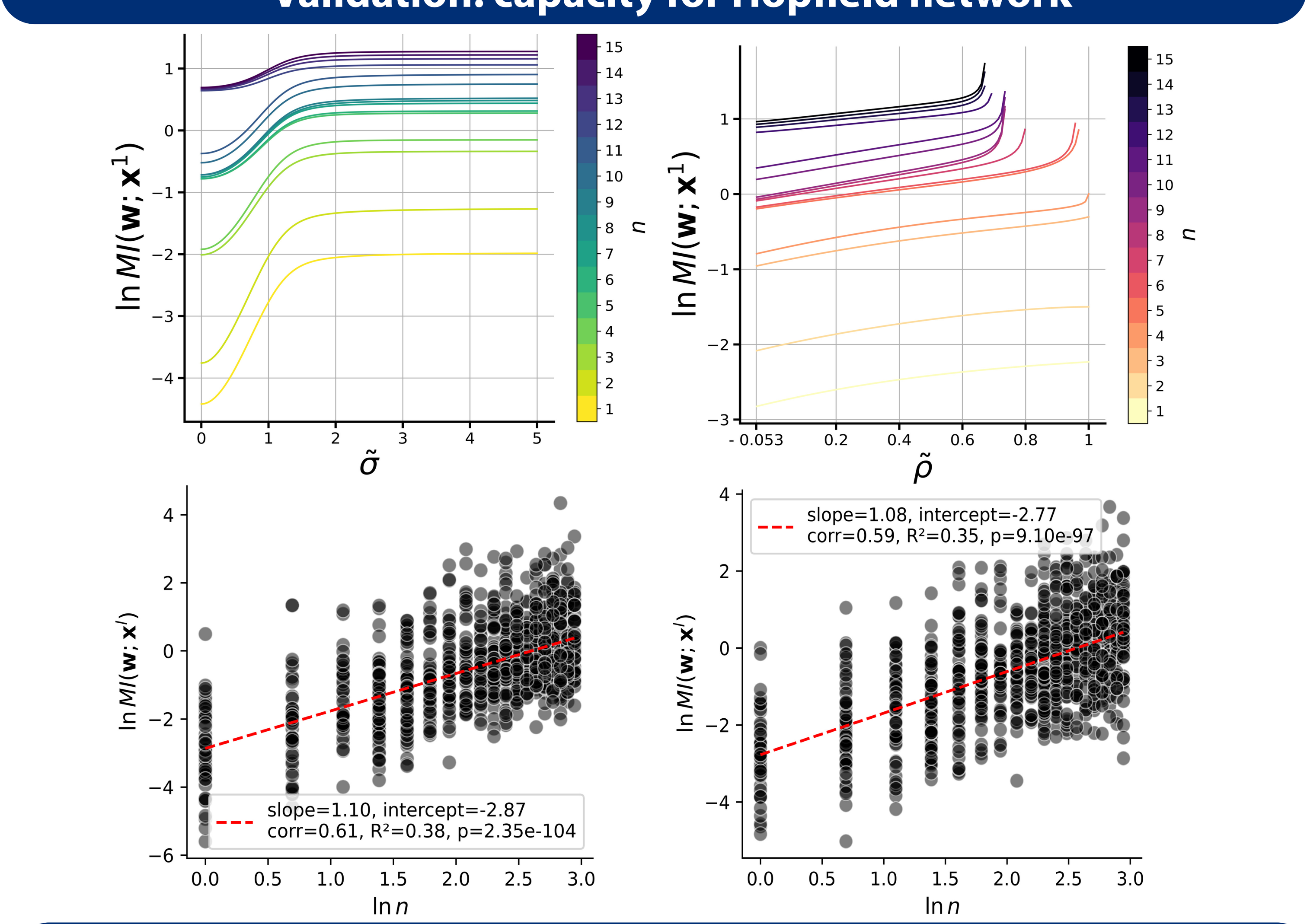
2. Randomly sample parameters.

$$\left(\{\mathbf{x}^1, \dots, \mathbf{x}^N\}, \mathbf{w}, MI(\mathbf{w}; \mathbf{x}^l)\right)$$

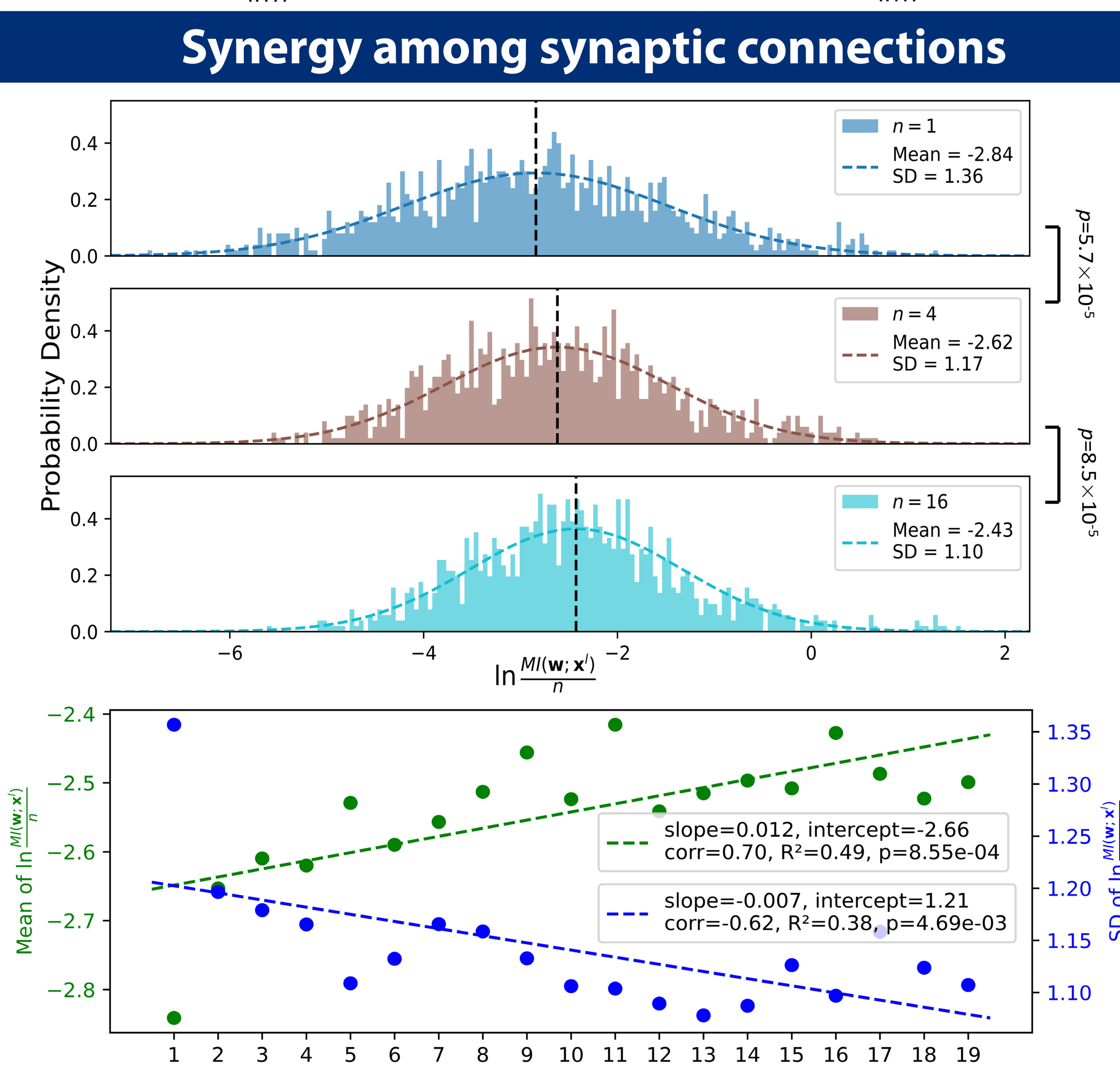
Validation: distributed coding in synaptic weights



Validation: capacity for Hopfield network



Synergy among synaptic connections



Conclusions & Discussions

1. Our analysis aligns with the **distributed coding paradigm** in neural networks.
2. We find that the storage **capacity of the Hopfield network** scales as $P \sim d^{1.2}$, consistent with classical results.
3. Synergistic coding is observed among synaptic connections.

* Assumption of log-normal data distributions.
* Errors for Fenton-Wilkinson approximation.
* Implication for modern deep learning. (neighboring layers, pruning)
* Retrieval dynamics.

[1] Borst, A., & Theunissen, F. E. (1999). Information theory and neural coding. Nature neuroscience, 2(11), 947-957.
[2] Quian Quiroga, R., & Panzeri, S. (2009). Extracting information from neuronal populations: information theory and decoding approaches. Nature Reviews Neuroscience, 10(3), 173-185.
[3] Dimitrov, A. G., Lazar, A. A., & Victor, J. D. (2011). Information theory in neuroscience. Journal of computational neuroscience, 30, 1-5.
[4] Hebb, D. O. (2005). The organization of behavior: A neuropsychological theory. Psychology press.
[5] Hopfield, J. J. (1984). Neurons with graded response have collective computational properties like those of two-state neurons. Proceedings of the national academy of sciences, 81(10), 3088-3092.