

# Functional Connectivity in the Brains of Zebrafish Larvae

Intern: Ans IMRAN  
Co-Supervisor: Mattéo Dommange-Kott  
Supervisor: Volker Bormuth  
Supervisor: Georges Debrégeas

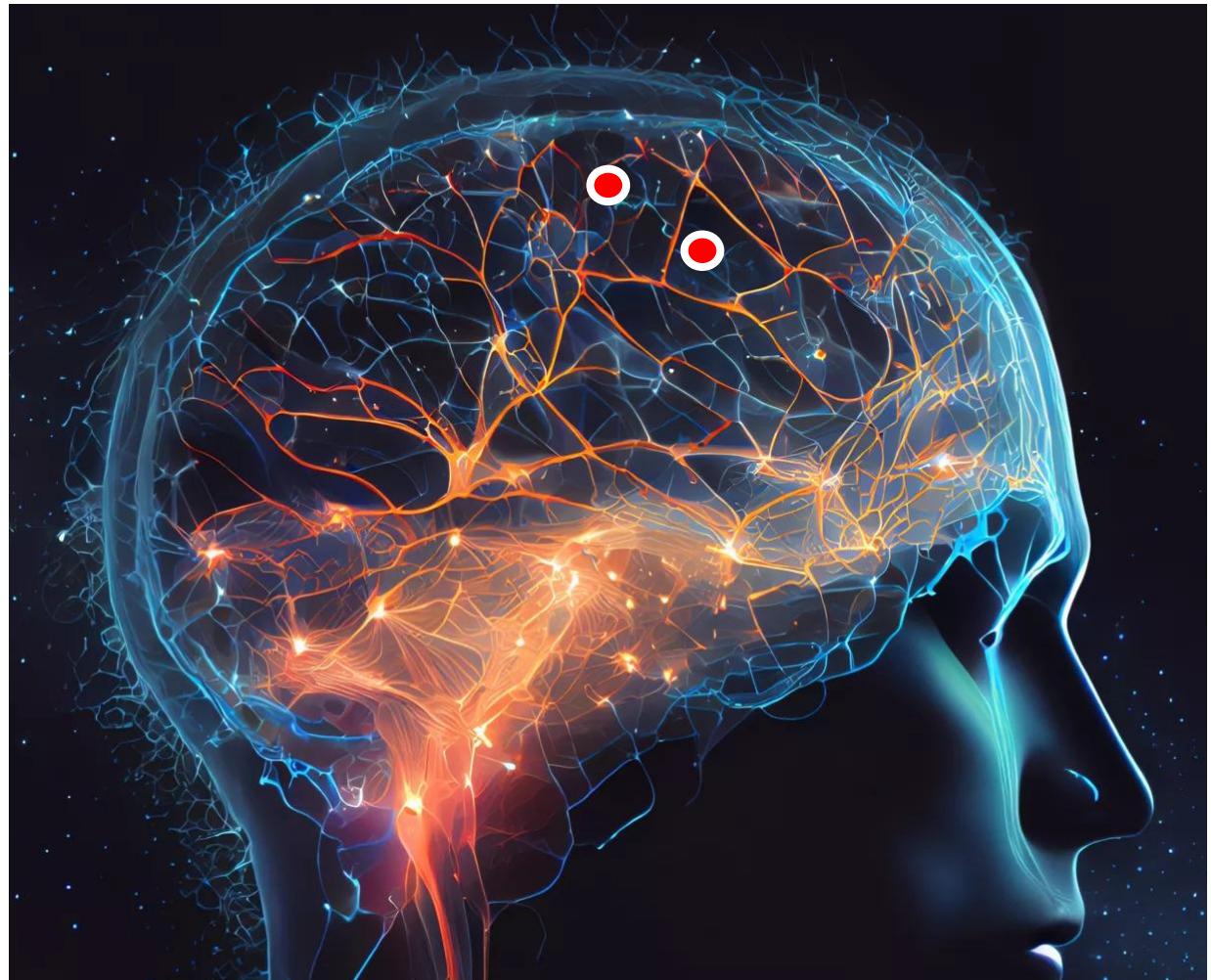
Can we reliably extract functional connectivity?

Is this functional connectivity conserved across individuals?

# Functional connectivity:

The probability that neuron-j is activate given that neuron-i is already active.

The probability that neuron-j is not activate given that neuron-i is not already active.



Credits: Dr. Ian Dobbs



Credit: Shawn Burgess, NHGRI  
Larval zebrafish



DOI

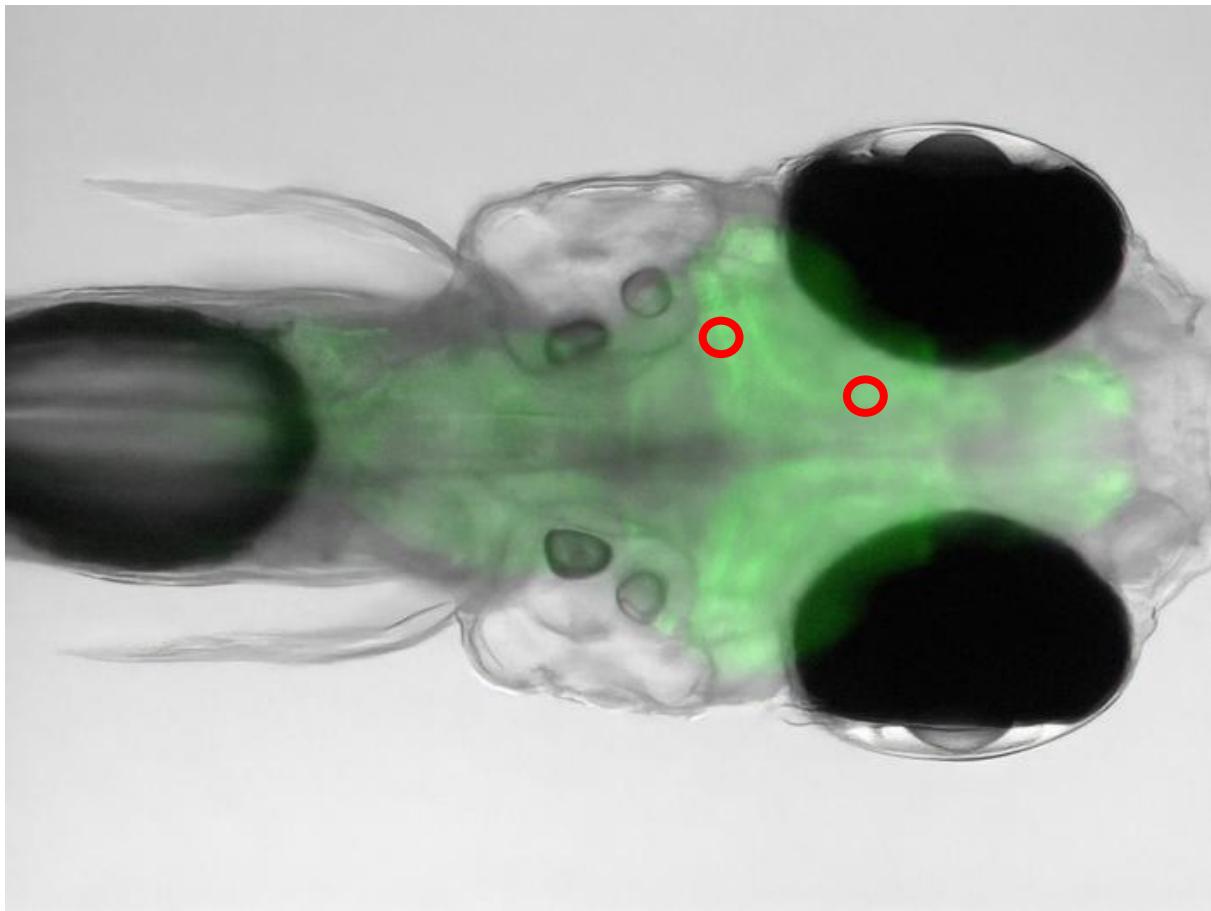
[10.1016/j.cub.2023.12.030](https://doi.org/10.1016/j.cub.2023.12.030)

# Functional connectivity:

The functional connection  $J_{ij}$  between pairs of neurons, where  $J_{ij}$  is quantified by directly perturbing the activity of neuron  $j$  and observing the change in probability to spike of neuron  $i$ .

$J_{ij}$  is defined as the impact of the state of neuron  $j$  on neuron  $i$  in the context of activity pattern  $v$ .

Intuitively, two neurons  $i, j$  are effectively connected if they are connected to the same HUs.

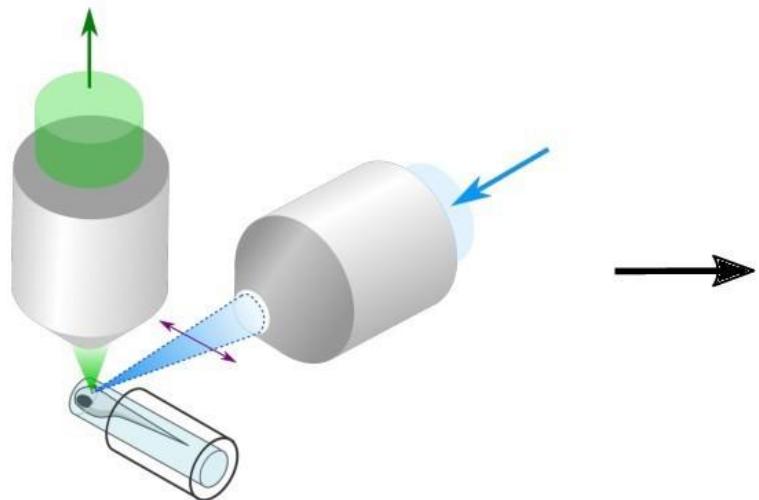


DOI

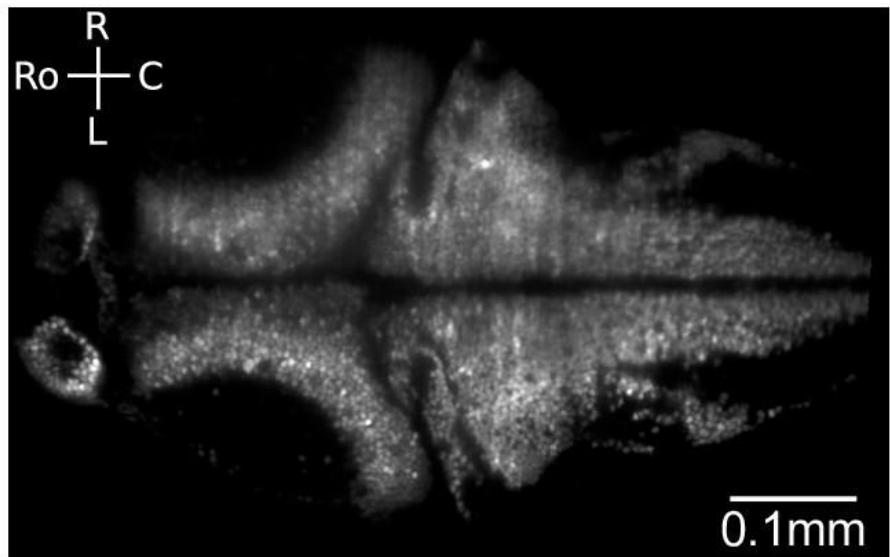
[10.1016/j.cub.2023.12.030](https://doi.org/10.1016/j.cub.2023.12.030)

A

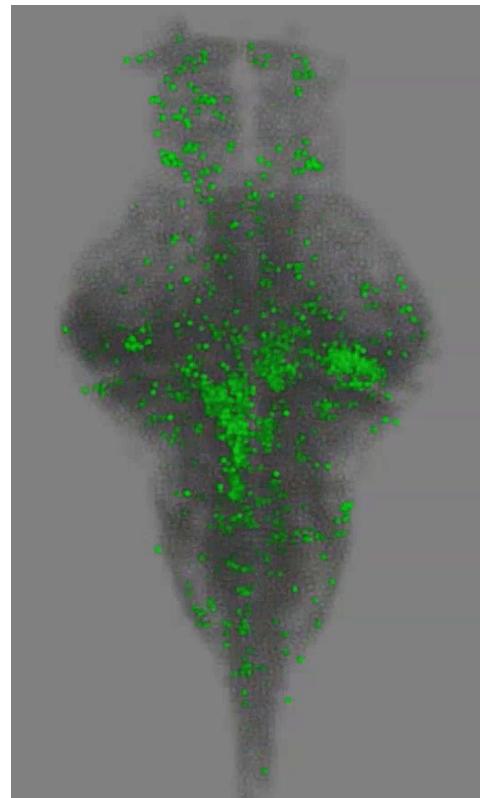
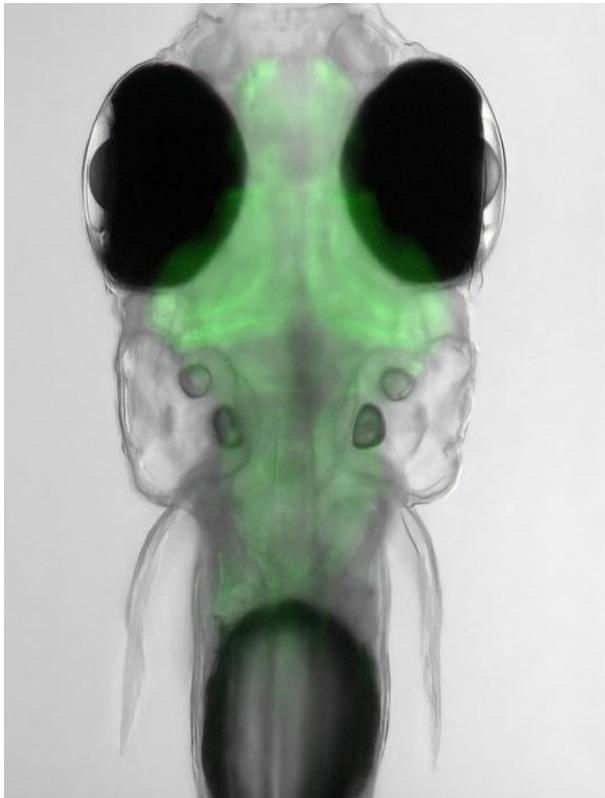
## Zebrafish light-sheet microscopy



## Single-cell resolution whole-brain imaging



# A recording of neuronal activity



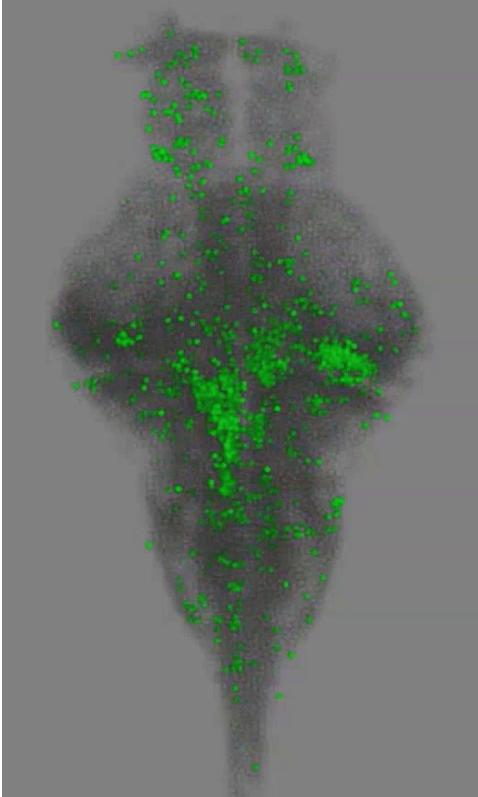
Whole brain size:  
400 micrometers

Time:  
0.4 seconds per frame

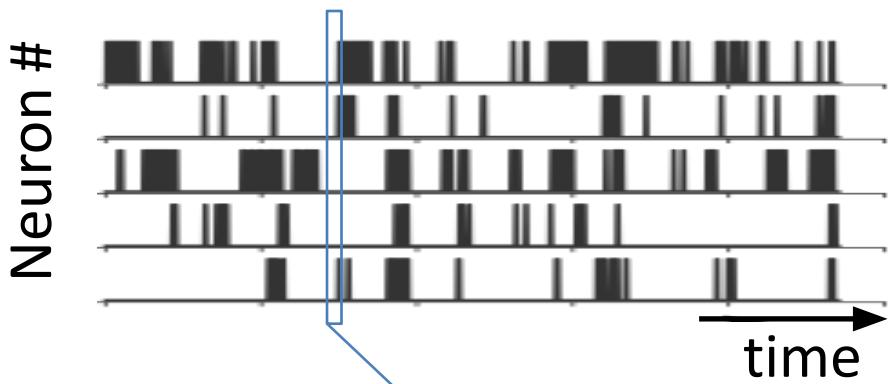
Credits:  
Matteo Dommangé-Kott

# Energy to Probability

Credits:  
Matteo Dommanget-Kott



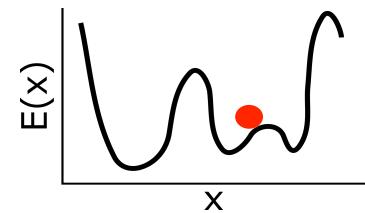
Whole brain size:  
400 micrometers  
  
Time:  
0.4 seconds per frame



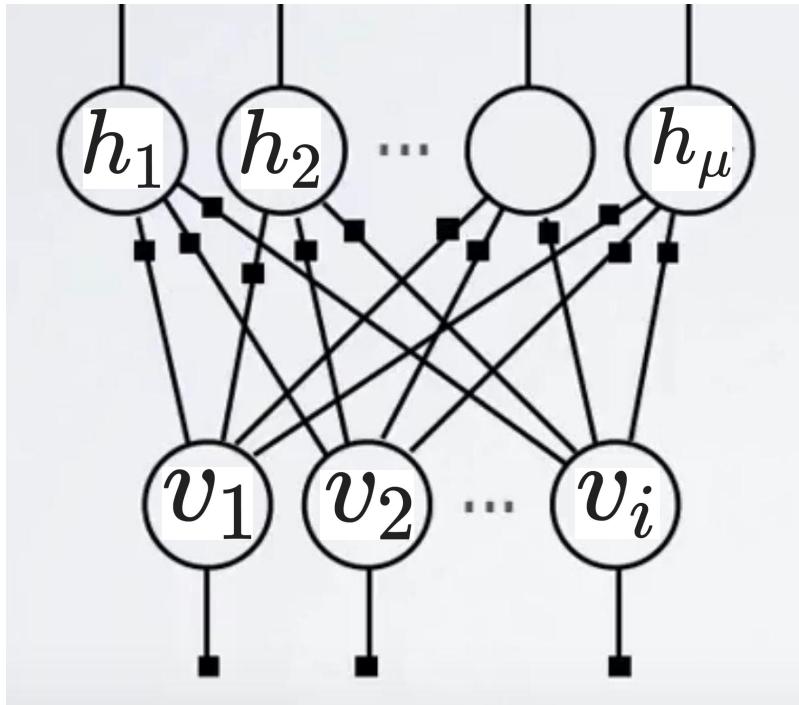
$$\vec{x} \dashrightarrow E(\vec{x})$$

$$E(\vec{x}) = w_1x_1 + w_2x_2 + \dots + w_nx_n$$

$$P(\vec{x}) = \frac{e^{-E(\vec{x})}}{Z}$$



# Restricted Boltzmann Machine

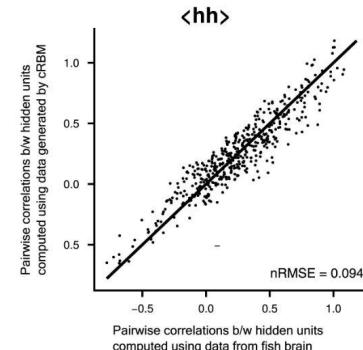
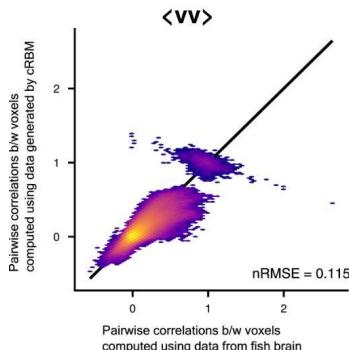
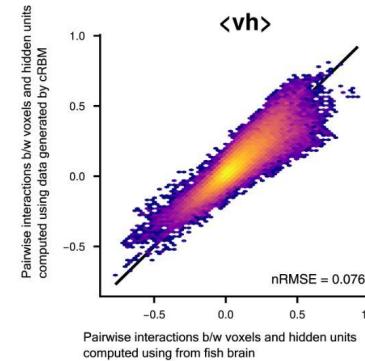
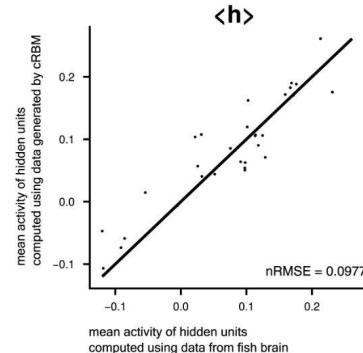
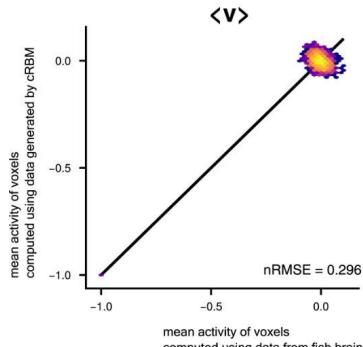


$$E(\mathbf{v}, \mathbf{h}) = - \sum_i g_i v_i + \sum_\mu \mathcal{U}_\mu(h_\mu) - \sum_{i,\mu} w_{i,\mu} v_i h_\mu$$

$$P(\mathbf{v}, \mathbf{h}) = \frac{1}{Z} \exp(-E(\mathbf{v}, \mathbf{h}))$$

# good RBM

id plots of statistics



# L2L1 Regularization

Lab has shown previously in order to get good brain statistics, you need sparsity, we use regularization to ensure sparsity...

$$\text{L2L1 regularization} = \frac{\lambda^2}{2Nq} \sum_{\mu} \left( \sum_{i,v} |w_{i\mu}(v)| \right)^2$$

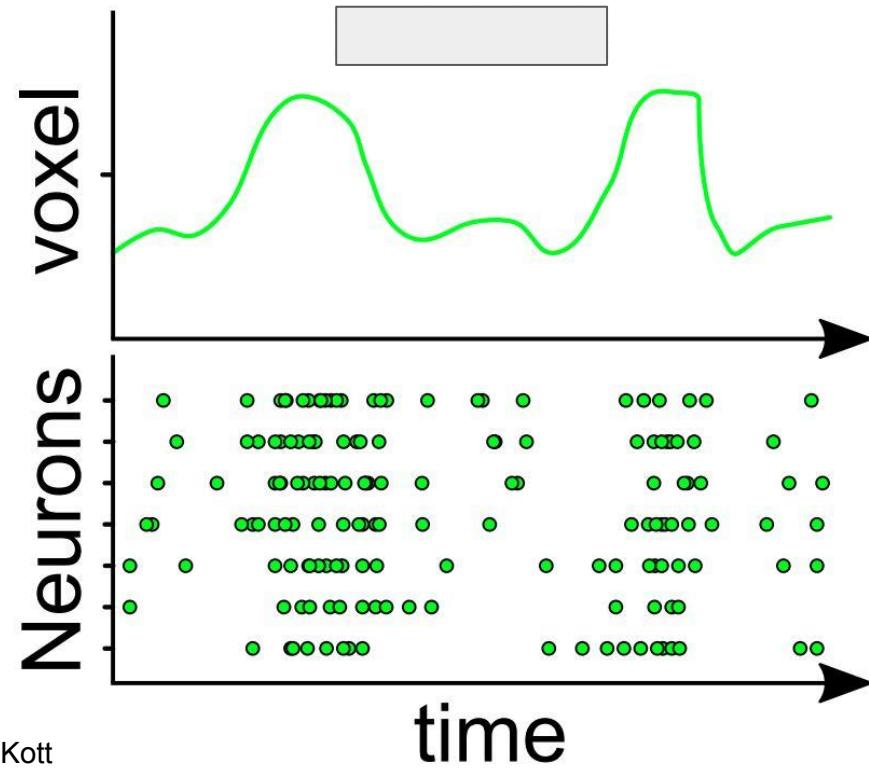
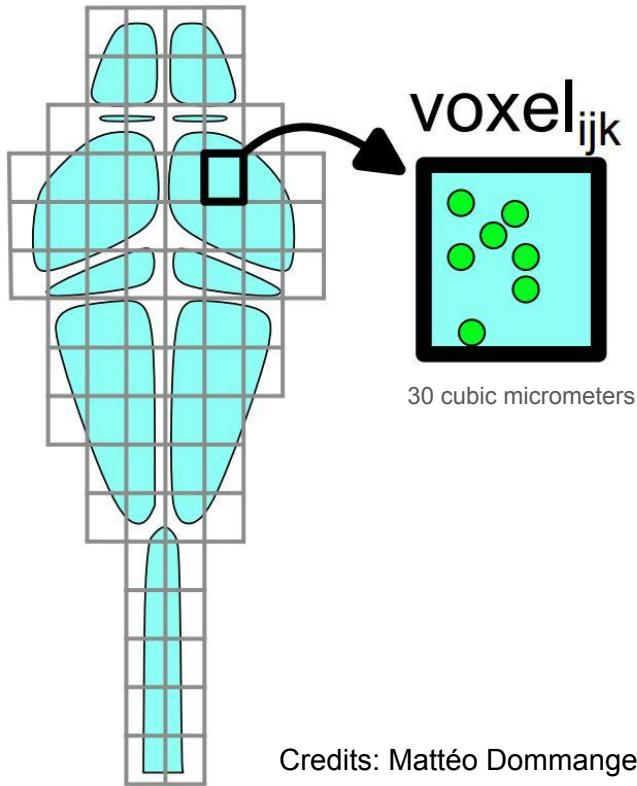
where

$w_{i\mu}$  : The weight of the connection between the visible-unit- $i$  and hidden-unit- $\mu$

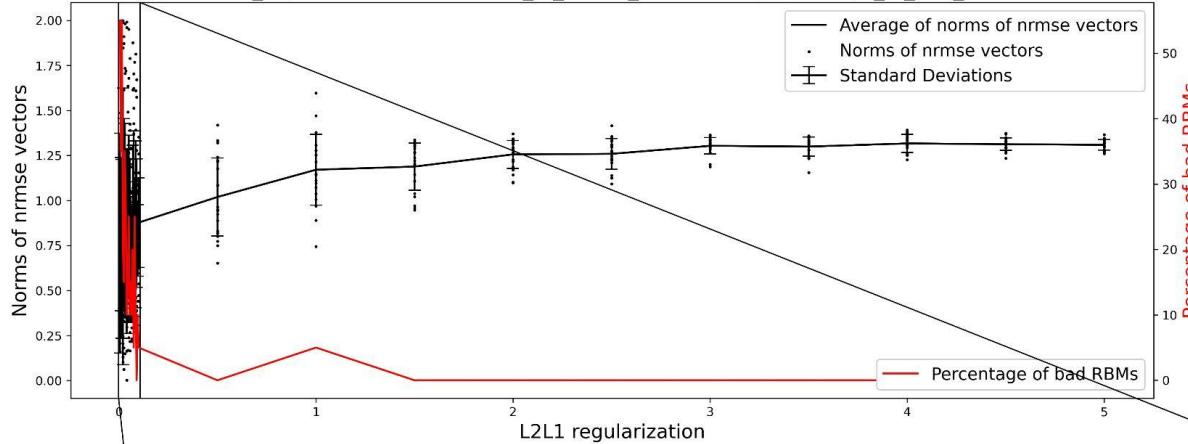
$\lambda$  : The L2L1 regularization parameter

$v$  : Visible configuration of voxels for one time step

# Voxelization



L2L1\_regularization VS norms\_of\_nrmse\_vectors & percentage\_of\_bad\_RBMs

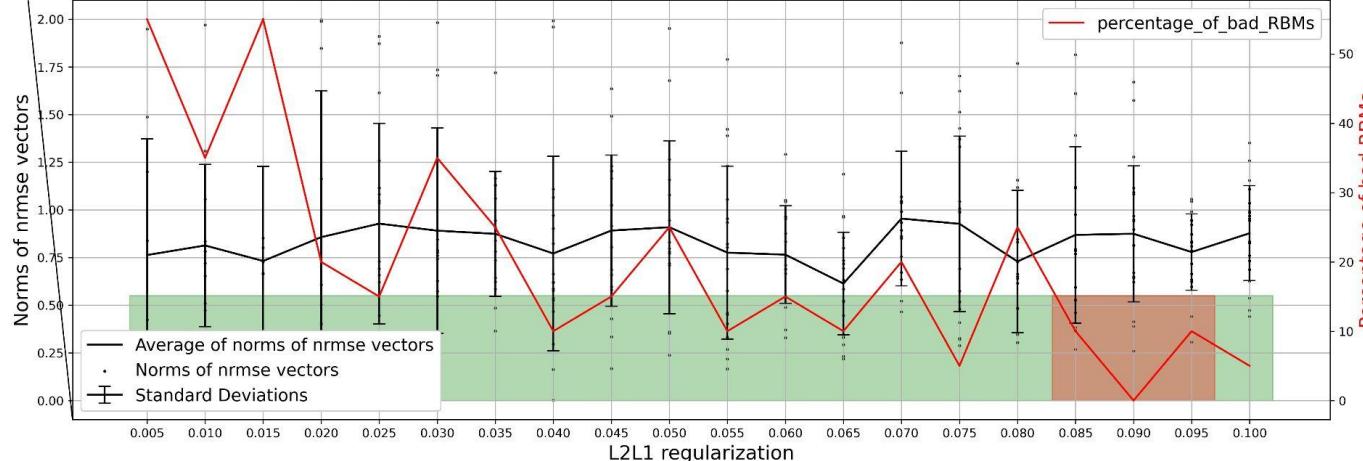


## Choosing the best value of regularization

$400 + 200 = 600$   
RBMs trained

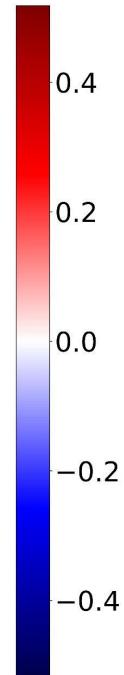
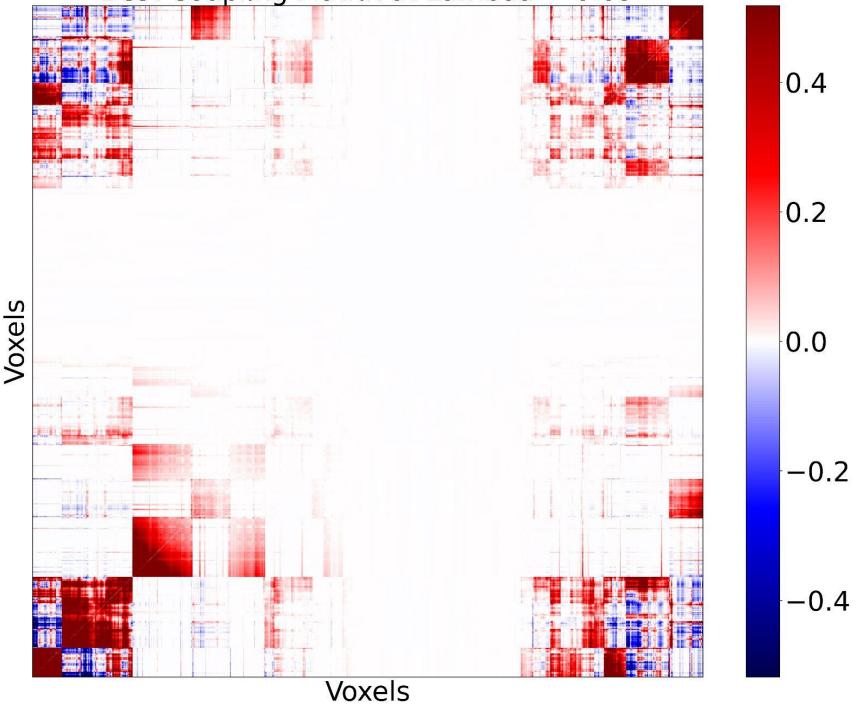
$320 + 199 = 519$   
Non-exploding RBMs

inset\_Plot | L2L1\_regularization VS norms\_of\_nrmse\_vectors & percentage\_of\_bad\_RBMs



**Are all these cRBMs producing same functional connections?**

Best Coupling Matrix at Lambda = 0.09



Question functional  
Positive  
negative

$$J_{ij} = \sum_{\mu=1}^M w_{i\mu} w_{j\mu} \langle \text{Var}(h_\mu | \mathbf{v}) \rangle_{\text{data}}$$

where

$J_{ij}$  : Strength of coupling between voxel- $i$  and voxel- $j$

$w_{i\mu}$  : The weight of the connection between the visible-unit- $i$  and hidden-unit- $\mu$

$M$  : Total number of hidden units

$h_\mu$  : The output of the hidden unit  $\mu$

$\mathbf{v}$  : Visible configuration of voxels for one time step

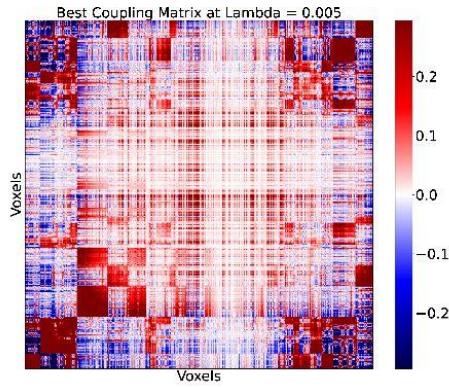
$(h_\mu | \mathbf{v})$  : Output of hidden-unit- $\mu$  given  $\mathbf{v}$

$\langle \text{Var}(h_\mu | \mathbf{v}) \rangle_{\text{data}}$  : Avg of the variations in the output of hidden unit  $\mu$  over all time steps

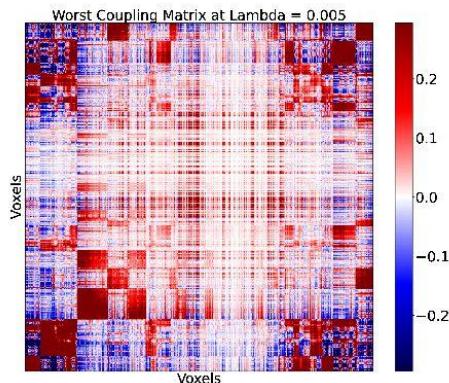
# Coupling matrix is conserved across cRBMs

1/2

Best RBM



Worst RBM

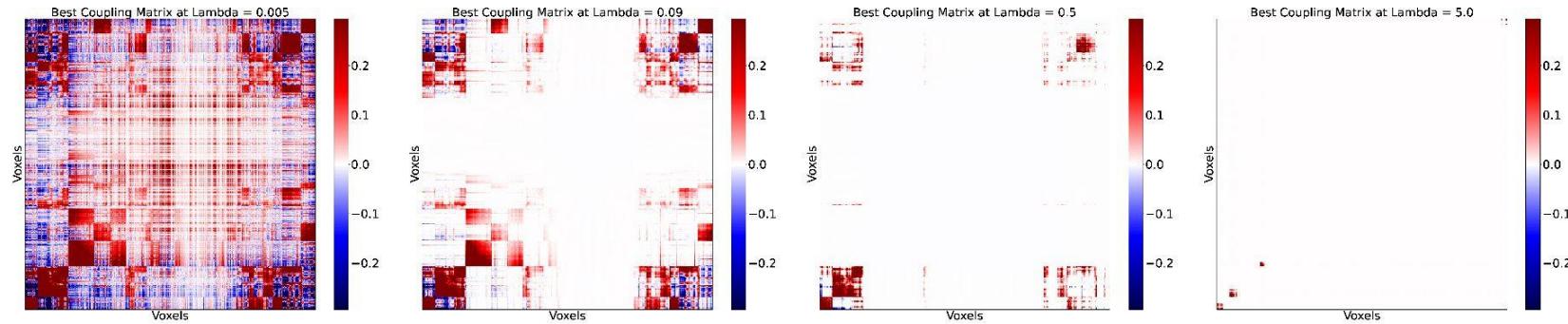


At fixed lambda, no matter how different the underlying RBMs are (performance wise), the coupling matrices are still the same!

# Coupling matrix is conserved across cRBMs

2/2

Best RBMs



A consistent structure across all coupling matrices,  
even at different regularization strengths.

Can we reliably extract functional connectivity?

Is this functional connectivity conserved across individuals?

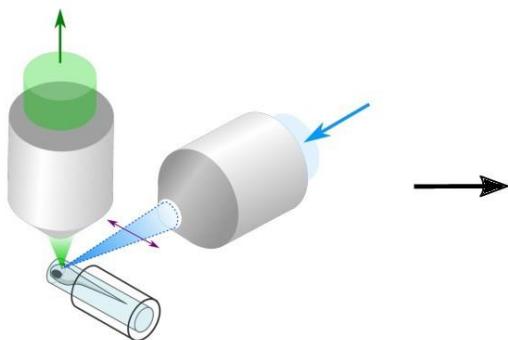
# Conclusion

- We trained Restricted Boltzmann Machines (RBMs).
- A regularization strength of 0.09 was found to be optimal.
- The coupling matrices was conserved across RBMs (good and relatively less good ones).
- This means functional connections are conserved across RBMs.
- Are functional conserved across different fish as well?

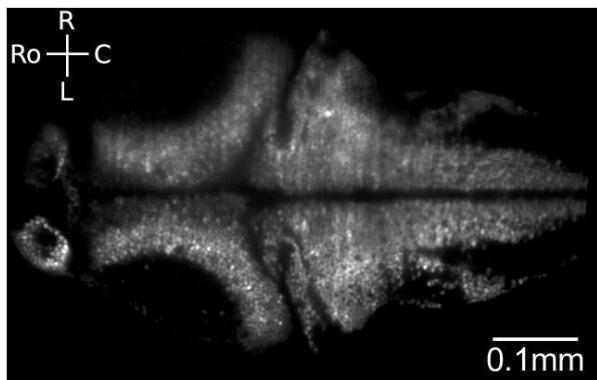
Thanks, Any Questions?

A

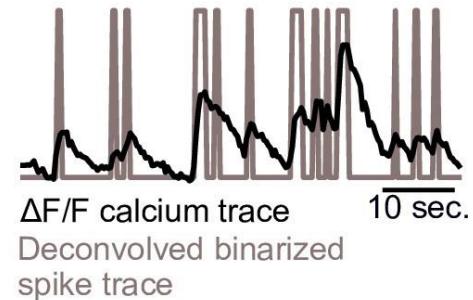
**Zebrafish  
light-sheet microscopy**



**Single-cell resolution  
whole-brain imaging**



**Spike deconvolution**



**Brain Activity Statistics:** In this report, when say Brain activity statistics, we mean the following statistics:

$\langle v_i \rangle$  : The mean activity of neuron-i or voxel-i (a collection of neurons).

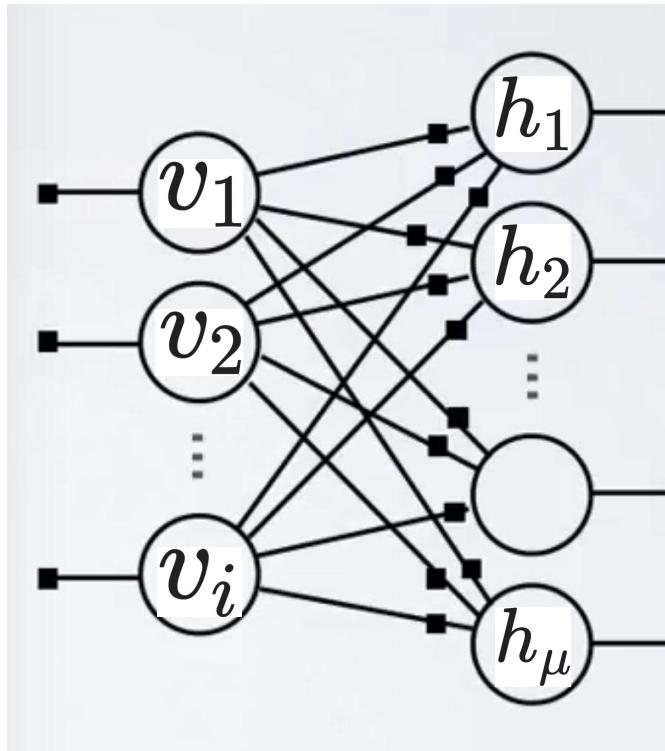
$\langle h_\mu \rangle$  : The mean activity of hidden-unit- $\mu$  (an assembly of voxels).

$\langle v_i h_\mu \rangle$  : The interaction between voxel-i and hidden-unit- $\mu$

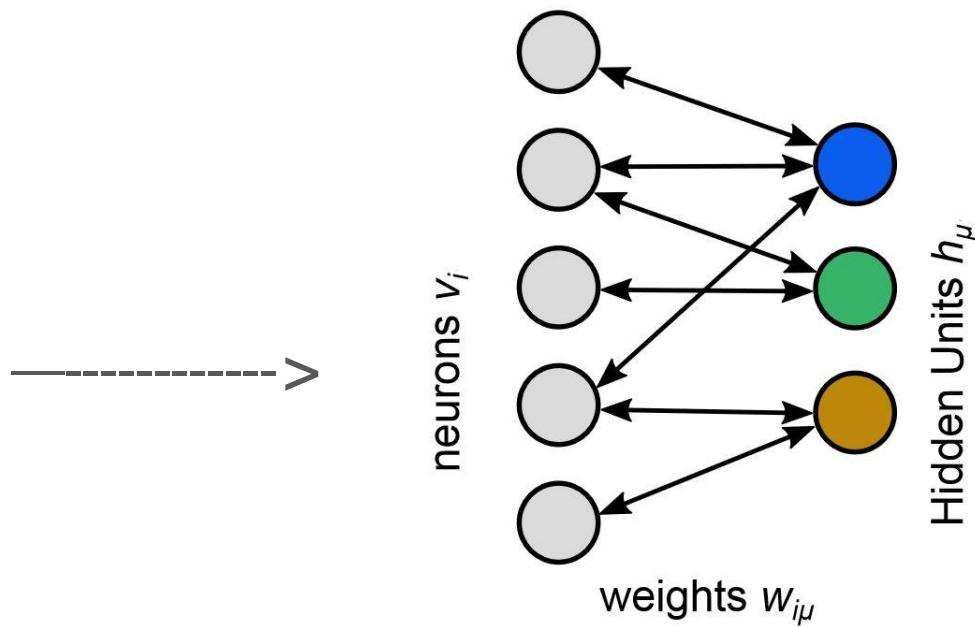
$\langle v_i v_j \rangle$  : The pairwise interaction between voxel-i and voxel-j

$\langle h_\mu h_\nu \rangle$  : The pairwise interaction between hidden-unit- $\mu$  and hidden-unit- $\nu$

Untrained RBM



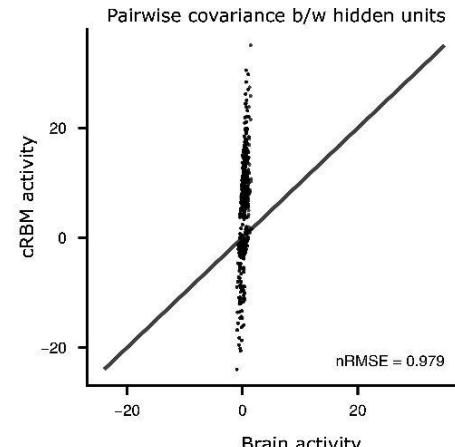
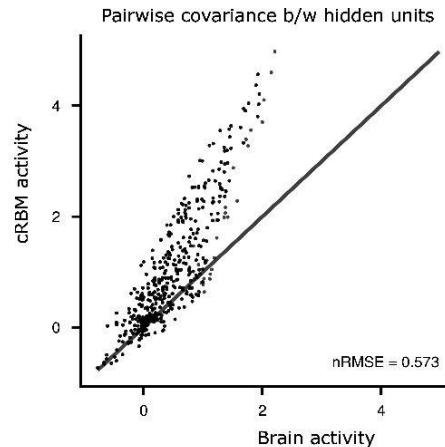
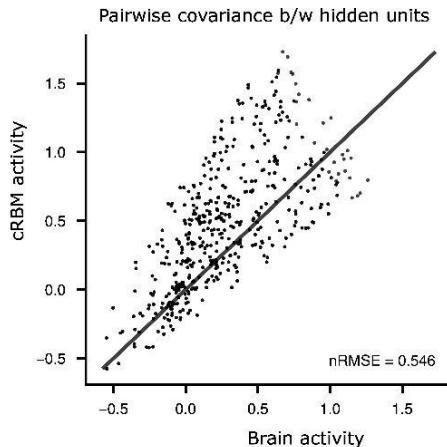
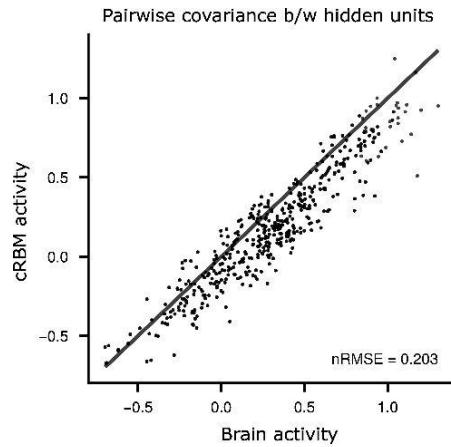
Trained RBM



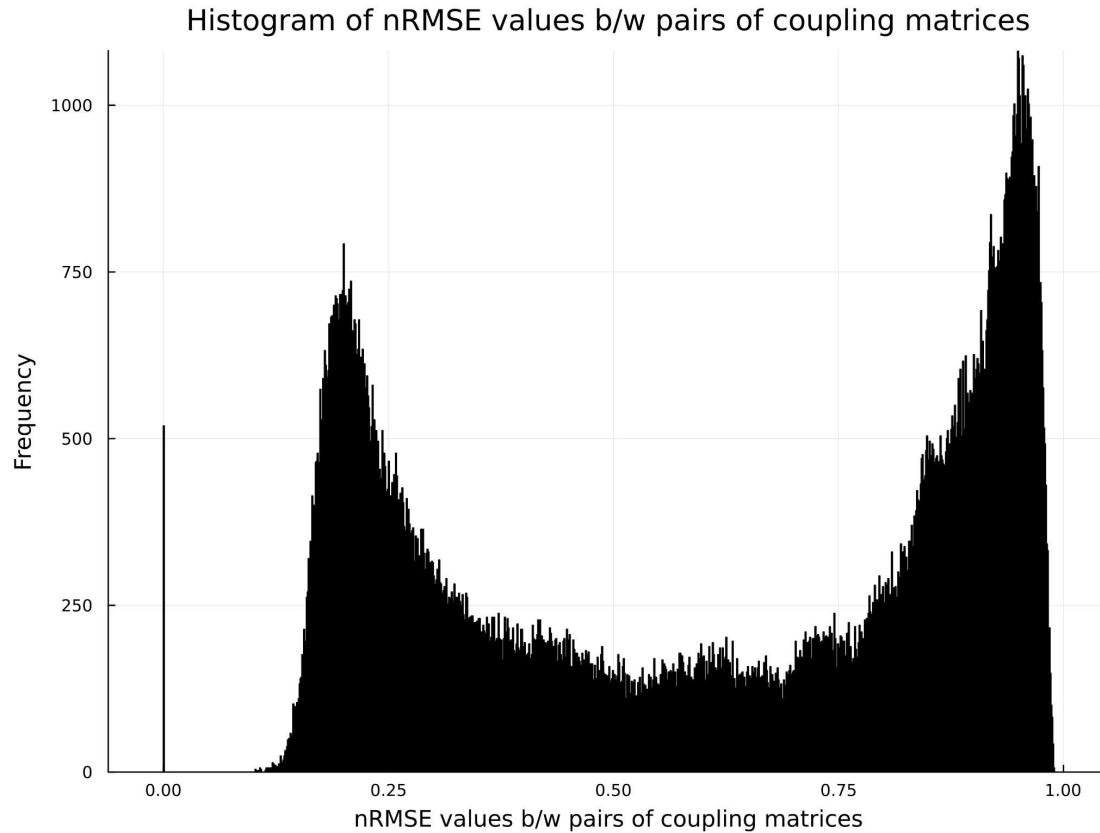
Credits: Hugo Larochelle

Credits:  
van der Plas, Tubiana et al.  
eLife 2023;11:e83139  
DOI: <https://doi.org/10.7554/eLife.83139>

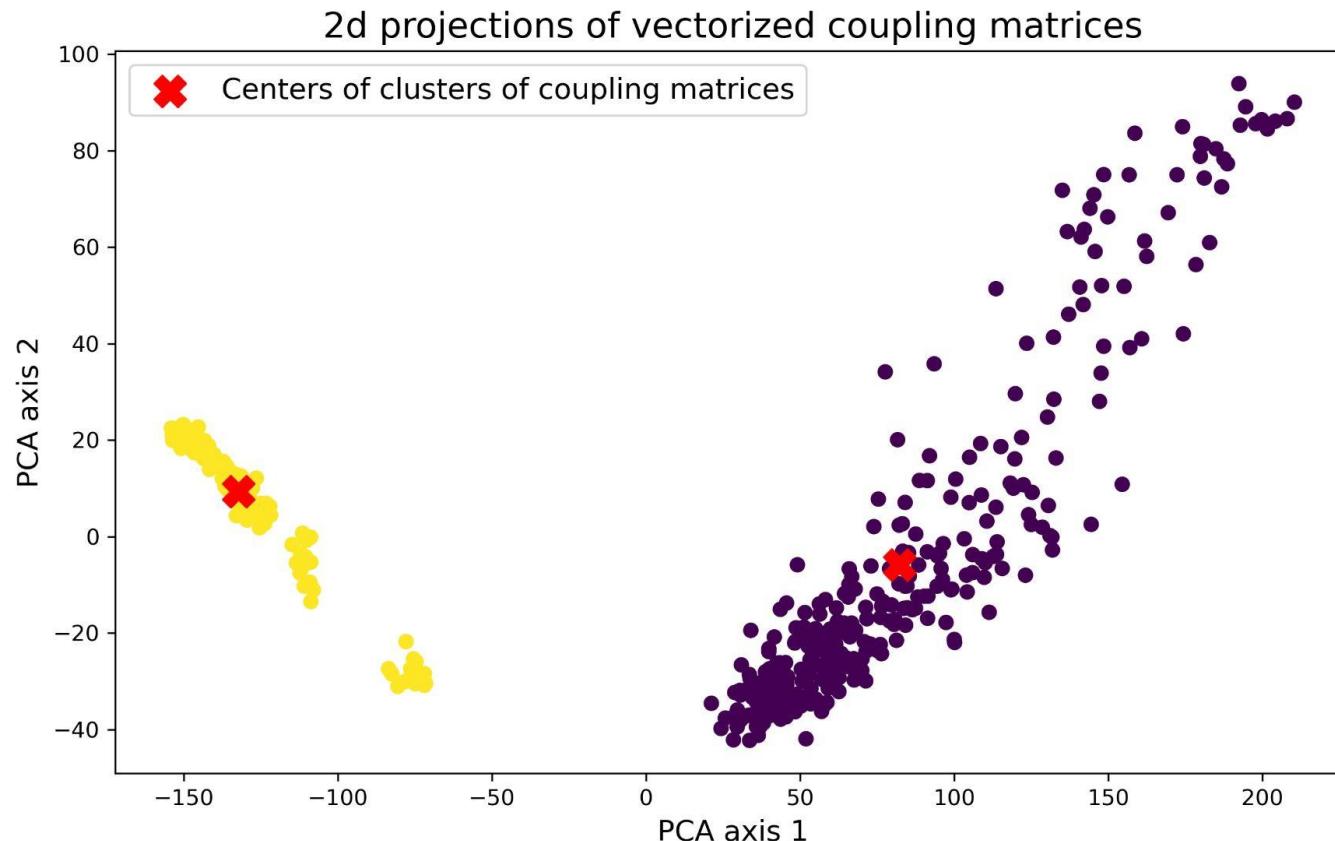
# nRMSE of $\leq 0.55$ Indicates a Good cRBM



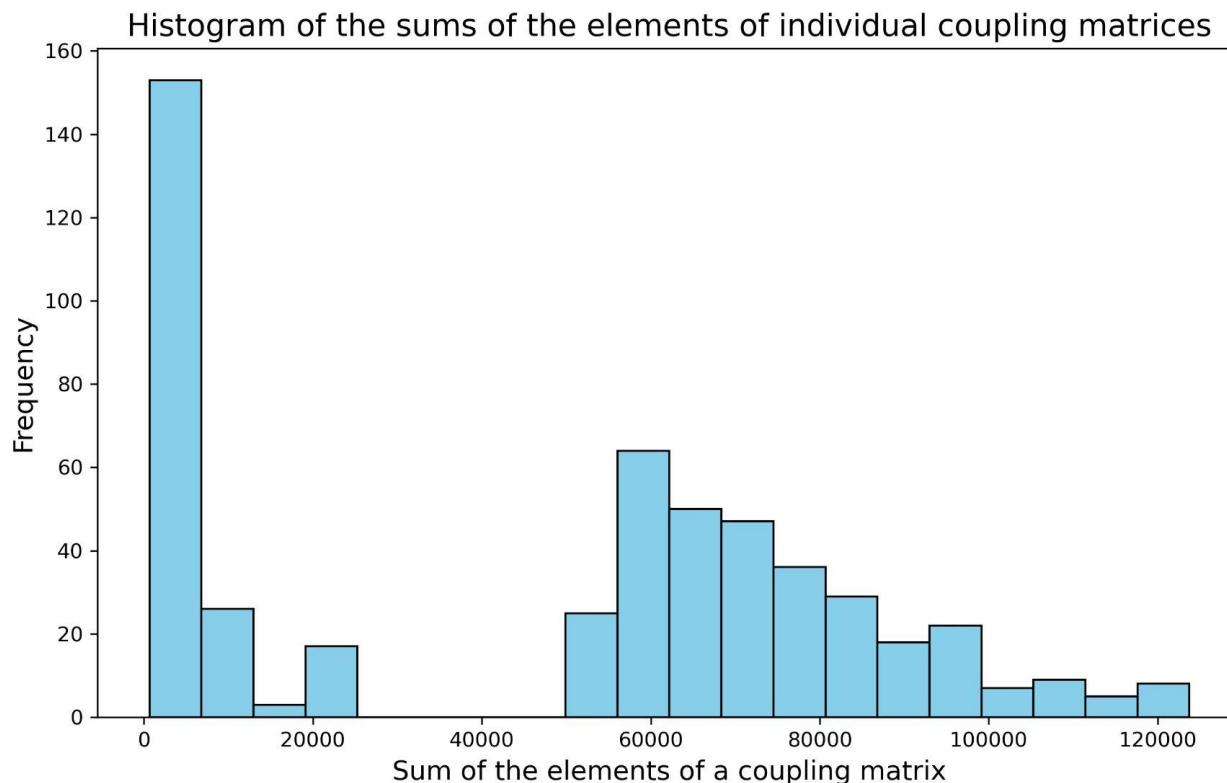
# Two different groups of coupling matrices



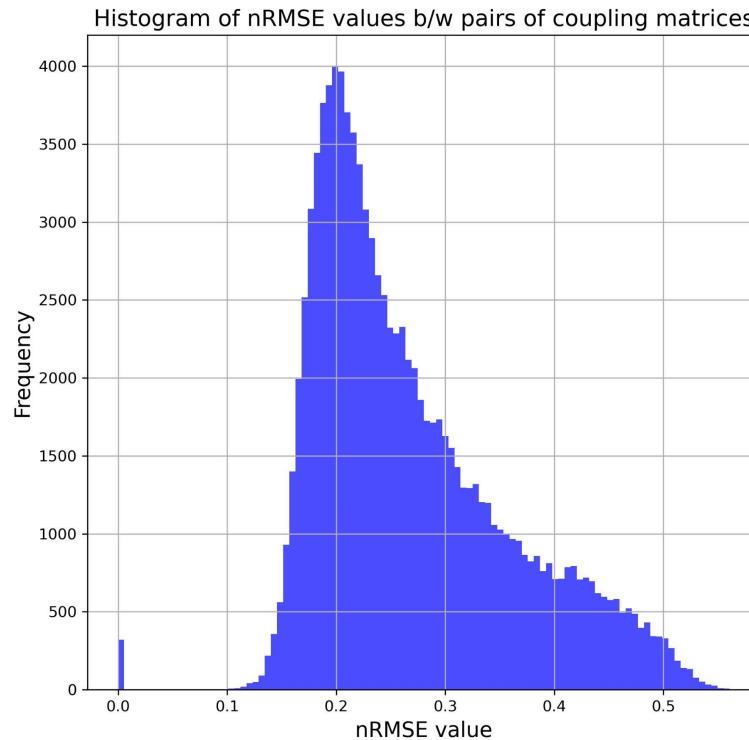
# Two separate clusters of coupling matrices



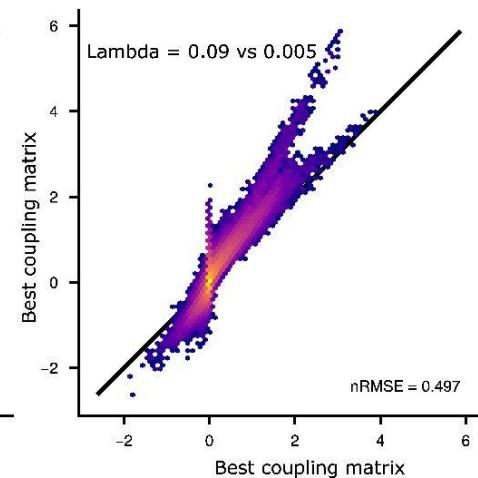
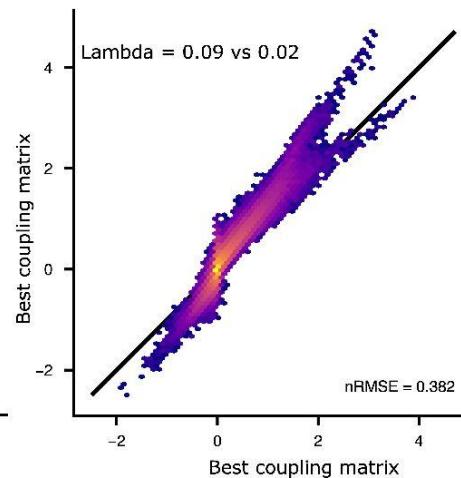
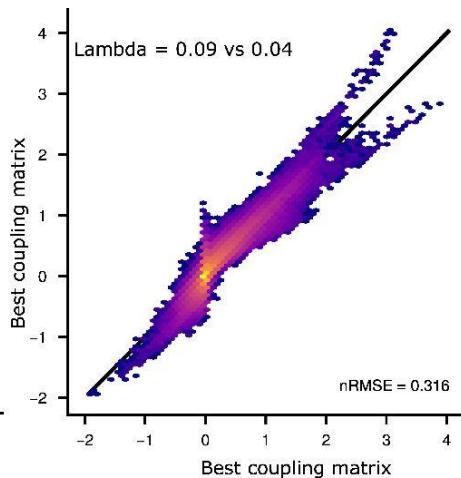
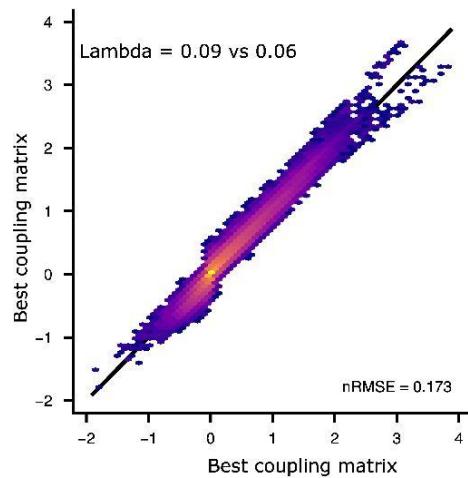
# Two Groups of Weights: High Magnitudes vs. Low Magnitudes



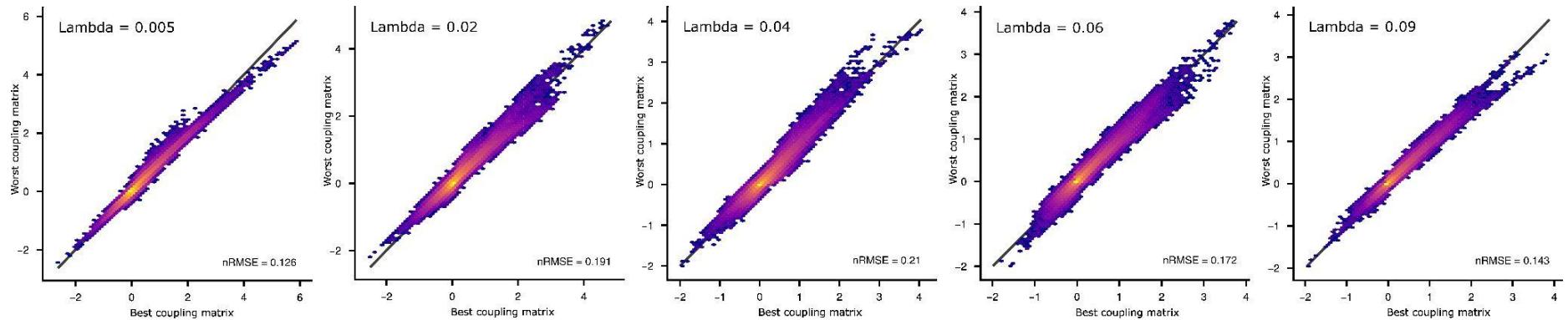
# Coupling matrices are not very different from each other



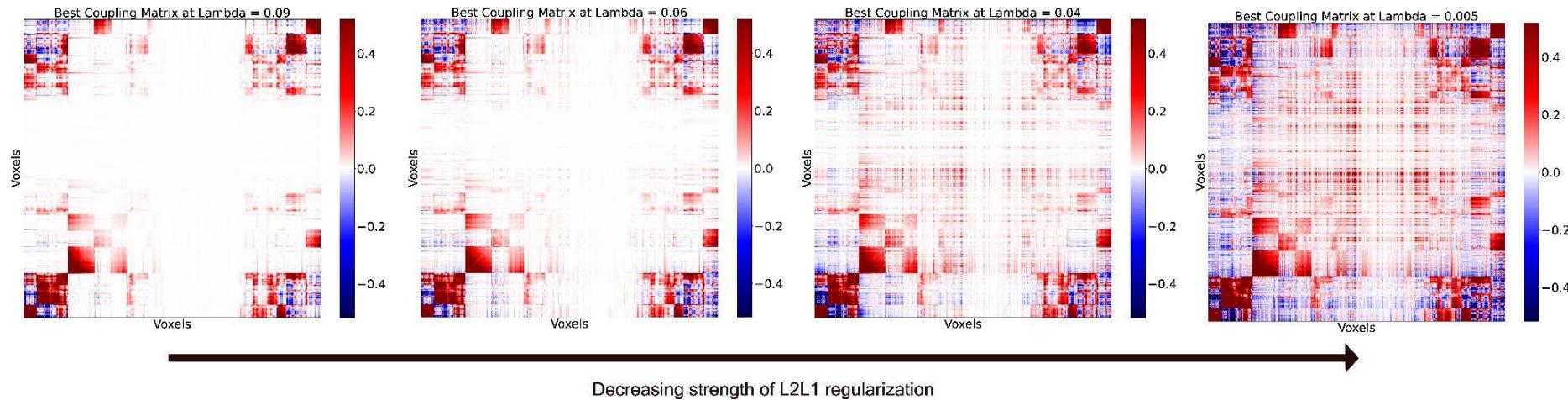
# The greater the $\lambda$ difference, the larger the nRMSE



# nRMSE at fixed $\lambda$ is almost constant

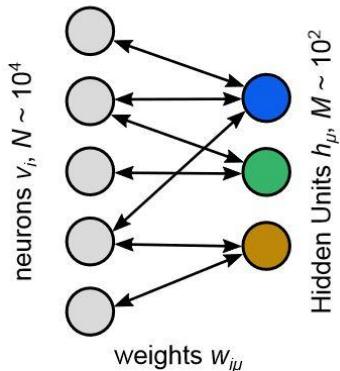


# Coupling matrix is conserved across cRBMs



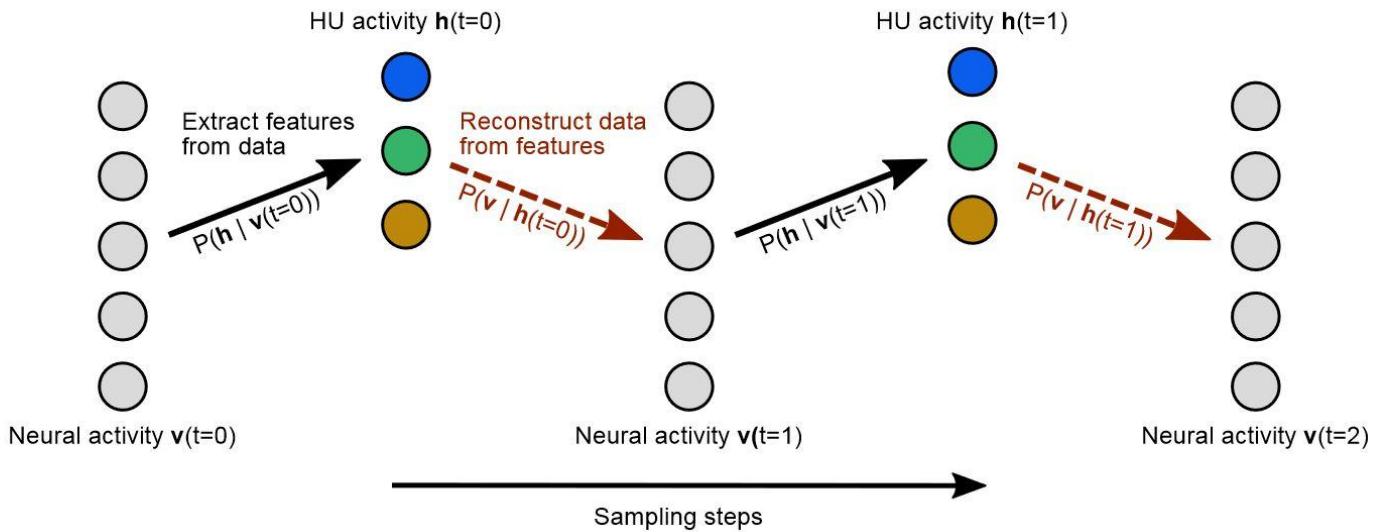
A

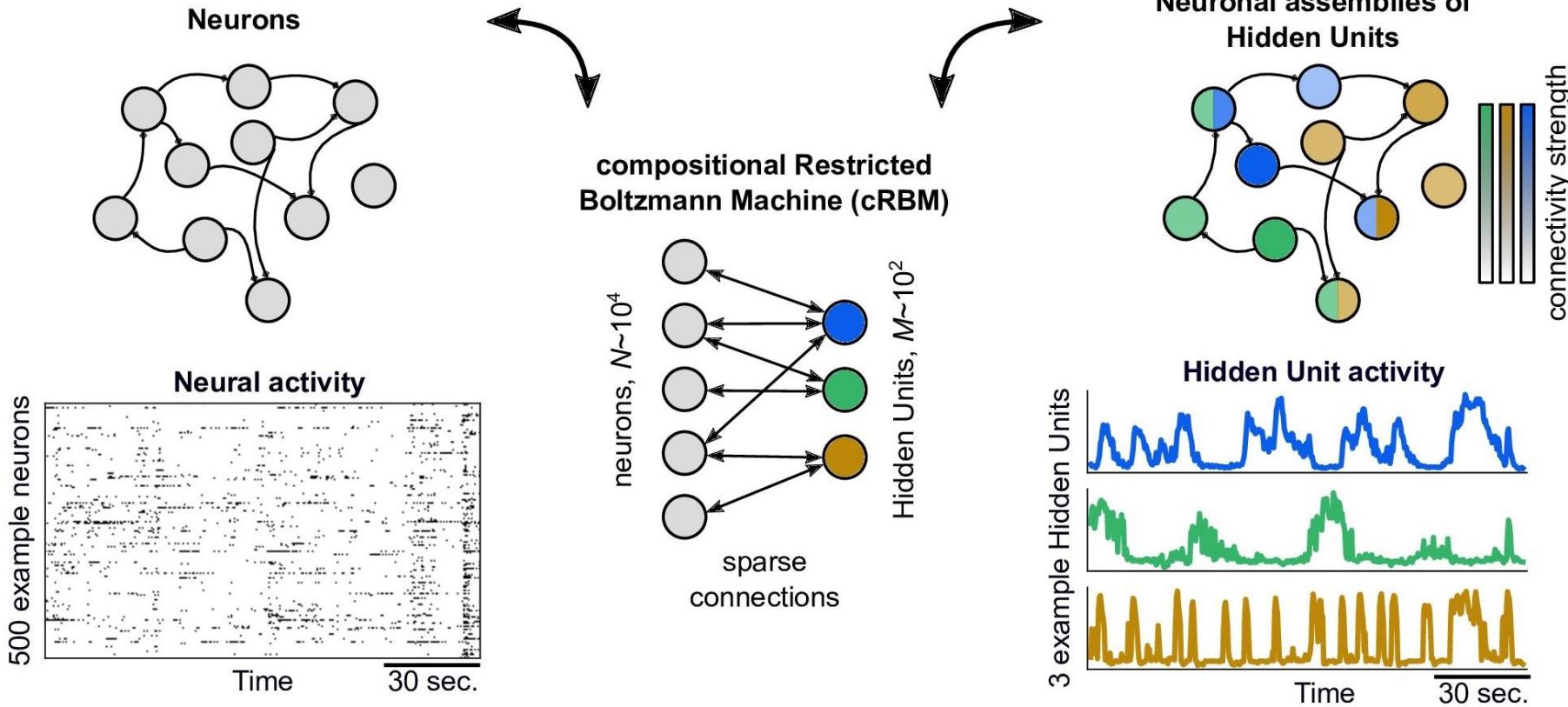
**compositional Restricted Boltzmann Machine (cRBM)**



B

**cRBM data generation schematic**



**B**

**Brain Activity:** In the context of this report, brain activity refers to spikes in neuronal activity. For example, during a specific time frame, if a neuron's activity spikes, we record a 1 for that neuron in that time frame; otherwise, we record a 0. The length of each time frame was 0.4 seconds.

**Brain Activity Statistics:** In this report, when say Brain activity statistics, we mean the following statistics:

$\langle v_i \rangle$  : The mean activity of neuron-i or voxel-i (a collection of neurons).

$\langle h_\mu \rangle$  : The mean activity of hidden-unit- $\mu$  (an assembly of voxels).

$\langle v_i h_\mu \rangle$  : The interaction between voxel-i and hidden-unit- $\mu$

$\langle v_i v_j \rangle$  : The pairwise interaction between voxel-i and voxel-j

$\langle h_\mu h_\nu \rangle$  : The pairwise interaction between hidden-unit- $\mu$  and hidden-unit- $\nu$

## Functional Connectivity Vs Structural Connectivity AND Coupling Matrix

**Functional Connectivity:** If the activities of two voxels or neurons show a high degree of correlation over time, it means that there is a functional connection between these two voxels or neurons.

**Structural Connectivity:** Structural connectivity refers to the anatomical connections between different brain regions, specifically the physical pathways that connect neurons within the brain. Unlike functional connectivity, which measures the temporal correlation of neural or voxel activity, structural connectivity focuses on the physical wiring of the brain's neural network.

**Coupling Matrix:** The coupling matrix is a matrix that gives us the strength of coupling of each voxel with all of the other voxels in the fish brain. Details in section 3.3

### 3.3 Coupling Matrix

The coupling matrix is a matrix that gives us the strength of coupling of each voxel with all of the other voxels in the fish brain. It is computed by using the formula:

$$J_{ij} = \sum_{\mu=1}^M w_{i\mu} w_{j\mu} \langle \text{Var}(h_\mu | \mathbf{v}) \rangle_{\text{data}}$$

where

$J_{ij}$  : Strength of coupling between voxel- $i$  and voxel- $j$

$w_{i\mu}$  : The weight of the connection between the visible-unit- $i$  and hidden-unit- $\mu$

$M$  : Total number of hidden units

$h_\mu$  : The output of the hidden unit  $\mu$

$\mathbf{v}$  : Visible configuration of voxels for one time step

$(h_\mu | \mathbf{v})$  : Output of hidden-unit- $\mu$  given  $\mathbf{v}$

$\langle \text{Var}(h_\mu | \mathbf{v}) \rangle_{\text{data}}$  : Avg of the variations in the output of hidden unit  $\mu$  over all time steps

### 3.4 bad cRBMs

“bad cRBMs” refers to those cRBMs for which the nRMSE value for any of the five brain activity statistics exceeded 1.0 or returned NaN.

### 3.5 normalized Root Mean Square Error - nRMSE

Depending on the availability of an optimal RMSE, two different formulas were used to calculate nRMSE. The optimal RMSE is the RMSE calculated between the statistics derived from the training data and the validation data.

$$nRMSE(X, Y) = \frac{RMSE(X, Y)}{RMSE(X_{\text{shuffled}}, Y_{\text{shuffled}})} \quad \text{without optimal}$$

$$nRMSE(X, Y) = 1 - \frac{RMSE(X, Y) - RMSE(X_{\text{shuffled}}, Y_{\text{shuffled}})}{RMSE(X_{\text{optimal}}, Y_{\text{optimal}}) - RMSE(X_{\text{shuffled}}, Y_{\text{shuffled}})} \quad \text{with optimal}$$

The nRMSE computed with optimal gives us some extra information. E.g, if:

$nRMSE = [-\infty, 0)$  : The nRMSE is better than the optimal RMSE

$nRMSE = 0$  : The nRMSE is as good as optimal RMSE

$nRMSE = 1$  : The nRMSE is as bad as shuffled RMSE

$nRMSE = (1, +\infty]$  : The nRMSE is worse than shuffled RMSE

modified loss = original loss( $w_1, w_2, \dots, w_j$ ) + regularization term( $w_1, w_2, \dots, w_j$ )

This regularization term (for L2 regularization) looks like this

$$\text{regularization term} = A\lambda \sum_j w_j^2$$

where  $A$  is a scaling factor and  $\lambda$  is the regularization parameter.

The details of  $A$  are not important for our current discussion. The regularization parameter  $\lambda$ , however, is important for our current discussion.

Now, notice the full form of regularization term,

$$\text{regularization term} = A(\lambda w_1 + \lambda w_2 + \dots + \lambda w_j)$$

The regularization term contains products of the form

$$\lambda w_j$$

For all these products the value of  $\lambda$  is fixed and is given by us, so for the algorithm this value is fixed until unless we change it by hand. Therefore, the only parameters that the algorithm can change are  $w_j$ .

### 3.7.1 L2L1 Regularization

The regularization that we just discussed is called L2 regularization, however, in the cRBM that I have used in this internship, uses L2L1 regularization. L2L1 regularization is not a standard regularization (in a sense that it is not normally used in cRBMs in the Machine Learning community). The L2L1 regularization term is given by:

$$\text{L2L1 regularization} = \frac{\lambda^2}{2Nq} \sum_{\mu} \left( \sum_{i,v} |w_{i\mu}(v)| \right)^2$$

where

$w_{i\mu}$  : The weight of the connection between the visible-unit- $i$  and hidden-unit- $\mu$

$\lambda$  : The L2L1 regularization parameter

$v$  : Visible configuration of voxels for one time step

The reasons why we use L2L1 regularization and not L1 or L2 regularization are as follows:

- Using only L2 regularization does not result in sparse weights. For more details on why sparsity is important to us, refer to this study: Plas et al. [16].
- On the other hand, using only L1 regularization tends to make the weights too sparse. This can lead to hidden units having connection strengths with all visible units that are very close to zero or zero, making them effectively disconnected from any visible unit. We refer to these as orphan hidden units.

So, we want sparsity, but not too much sparsity. This is where L2L1 regularization comes into play. It provides the right balance of sparsity without creating orphan hidden units. For more details on L2L1 regularization, refer to Tubiana, Cocco, and Monasson [20].

## nRMSE Vector

I treated all the nRMSE values for the 5 brain activity statistics (except for  $\langle v_i \rangle$ ) as elements of a 4-dimensional Euclidean vector. I then computed the L2 norms of these 4d euclidean vectors (nRMSE vectors) for all of the desired cRBMs.

I did not use the  $\langle v_i \rangle_{nRMSE}$  for the following reasons.

The  $\langle v_i \rangle$  statistic is the easiest to reproduce for all visible units. This is because I performed z-score normalization on my input data  $v$  (the voxel activity). Due to the z-score normalization,  $\langle v_i \rangle \sim 0$ . Since  $\langle v_i \rangle \sim 0$ , this statistic is very easy to reproduce for cRBMs.

Even poor cRBMs (those with high nRMSE values for the other four statistics) can reproduce it well. Therefore, it provides little information when trying to differentiate between relatively good and relatively bad cRBMs. So, I wanted to prevent it from unduly influencing the norm of the nRMSE vector.

Now, I had a method to differentiate between relatively good and bad cRBMs in the context of how well they reproduce the brain activity statistics. I simply had to examine the L2 norm of the nRMSE vector associated with each cRBM. If the L2 norm was higher (relative to other cRBMs), it would indicate that the underlying cRBM was not reproducing the brain activity statistics well (relative to other cRBMs). Conversely, if the L2 norm was lower (relative to other cRBMs), it would indicate that the underlying cRBM was reproducing the brain activity statistics well (relative to other cRBMs).

$$\begin{aligned}
J_{ij}(\mathbf{v}) &= \log \left( \frac{P(v_i = 1 | v_1, \dots, v_{i-1}, v_{i+1}, v_j = 1, \dots, v_N)}{P(v_i = 1 | v_1, \dots, v_{i-1}, v_{i+1}, v_j = 0, \dots, v_N)} \right) \\
&\quad - \log \left( \frac{P(v_i = 0 | v_1, \dots, v_{i-1}, v_{i+1}, v_j = 1, \dots, v_N)}{P(v_i = 0 | v_1, \dots, v_{i-1}, v_{i+1}, v_j = 0, \dots, v_N)} \right)
\end{aligned}$$