

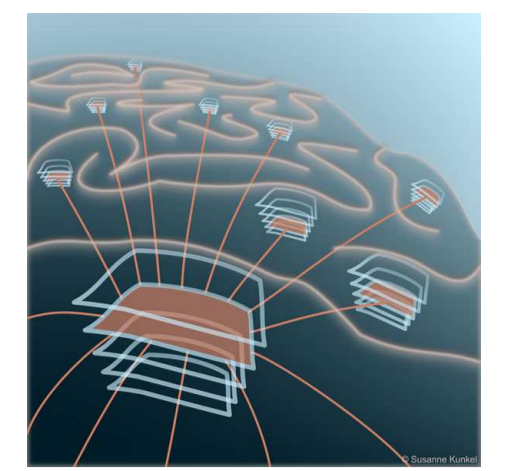
Validation Methods for Neural Network Simulations

Robin Gutzen¹, Sonja Grün^{1,2}*, Michael Denker¹*

¹ Institute of Neuroscience and Medicine (INM-6) and Institute for Advanced Simulation (IAS-6) and JARA-Brain Institute I (INM-10), Jülich Research Centre, Jülich, Germany

² Theoretical Systems Neurobiology, RWTH Aachen University, Aachen, Germany

Contact: r.gutzen@fz-juelich.de *equal contribution



Motivation

► How can we assess if the outcomes of two simulation runs of a network model are the same for different *neuron- or synapse-models, simulators, temporal resolutions, random seeds, parameters, or when performed on different computer architectures?* [1]

► This comparison is not trivial, because the exact spike times are not necessarily identical between the two outcomes.

► Here, we investigate two approaches to use neural correlations as a generic feature to describe the similarity on the level of the coordination of activity.

Approach 1: Compare distributions of correlation coefficients by applying two-sample tests.

Are the pairwise correlation coefficients sampled from the same underlying probability distribution?

Approach 2: Describe the correlation structure by means of an eigenvalue decomposition.

Is it possible to detect a correlation structure beyond pairwise relations?

Data: Validation methods are applied to two different scenarios in parallel: Stochastically generated data and results of a network simulation

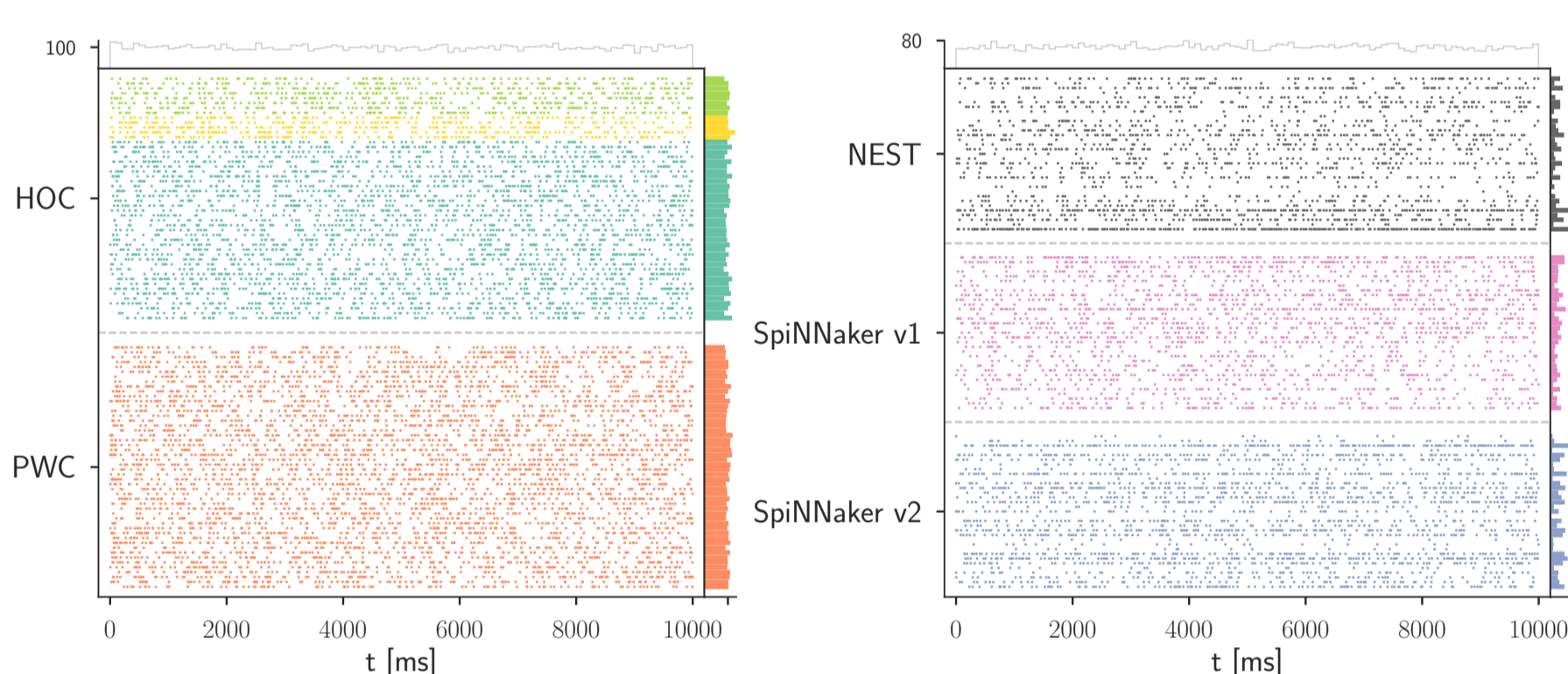
Data Sets

Stochastic Data
HOC vs. PWC

Simulation Data
NEST vs. SpiNNaker v1 vs. SpiNNaker v2

Q: Can different correlation structures be distinguished?

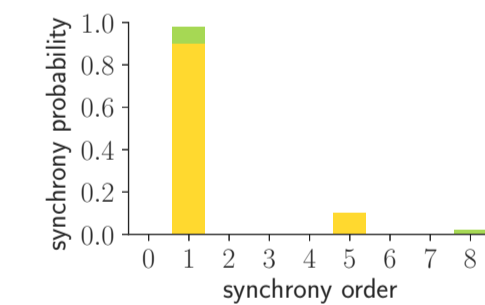
Q: Are the approaches applicable to a real world validation scenario?



► Generate stochastic activity data for 100 neurons

► **HOC**: Higher order correlations of order 8 and 5 via compound Poisson processes + homogeneous Poisson processes

► A compound Poisson process [2][3] samples synchronous events from an amplitude distribution:



► Simulation: 1 mm³ cortical microcircuit model (80k neurons) [5]

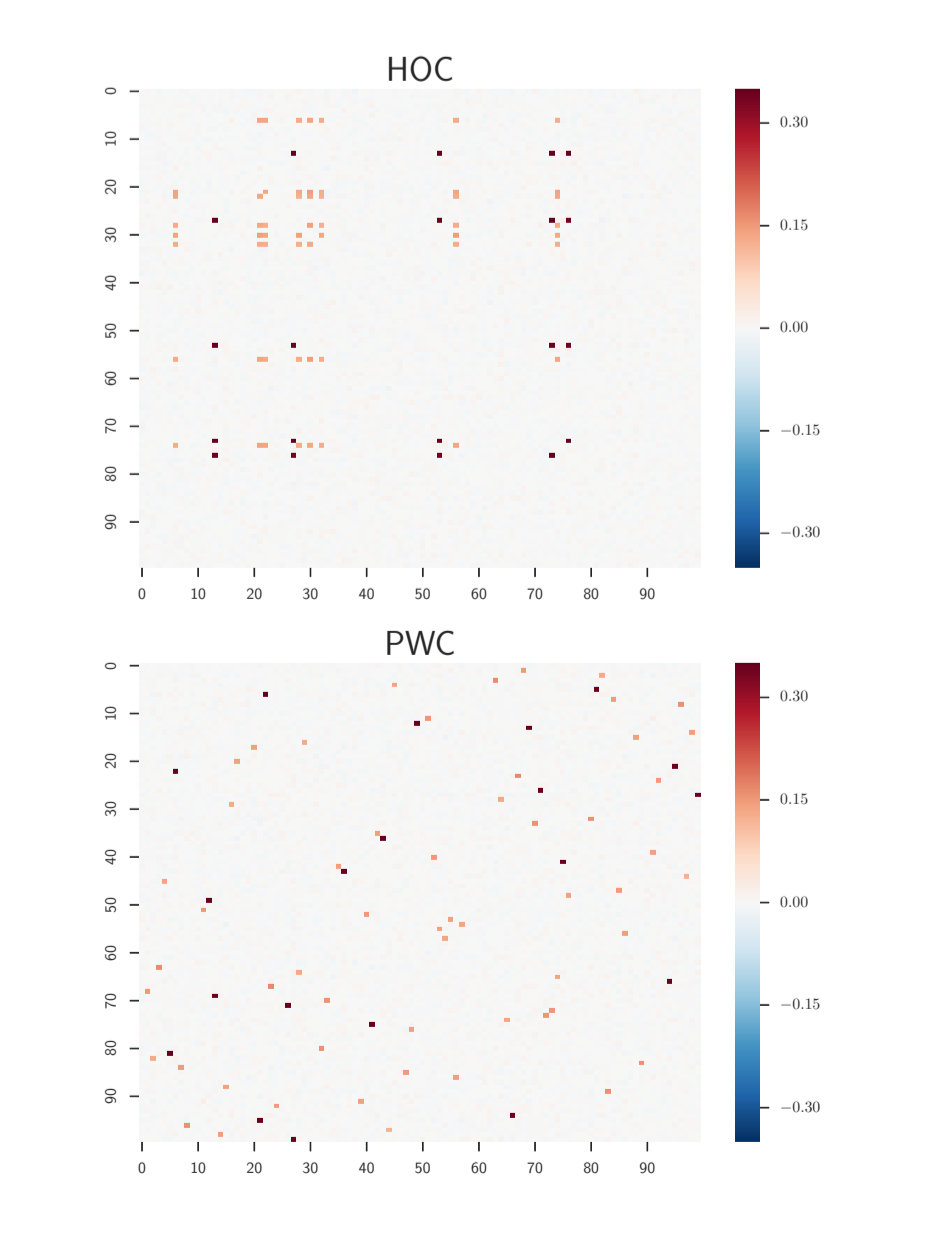
► Analyze 100 inh. neurons subsampled from Layer 4

► **NEST**: Simulated by a conventional simulator (NEST [4]) on an HPC

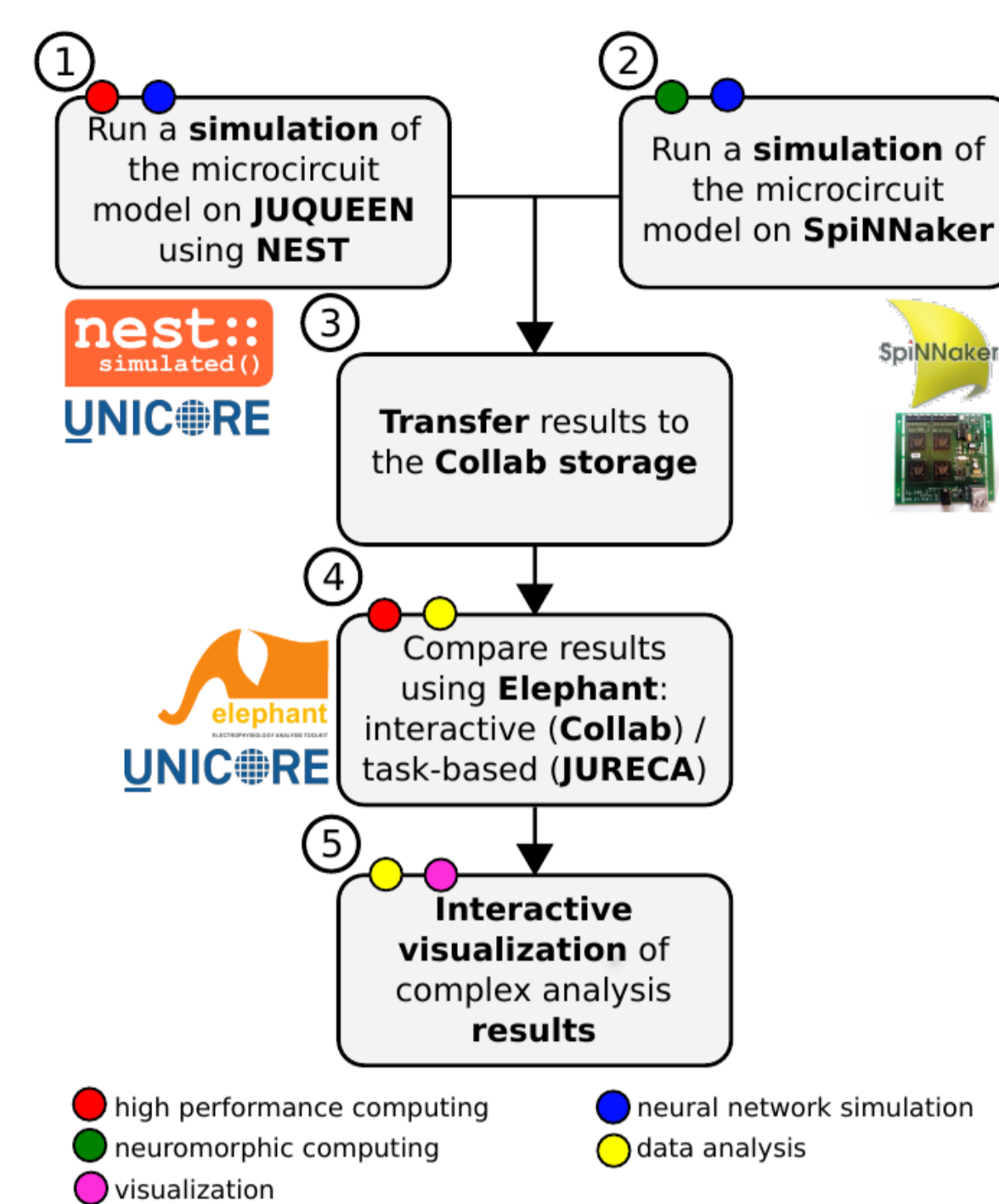
► **SpiNNaker v1**: Simulated on a neuromorphic hardware (SpiNNaker [6])

► **SpiNNaker v2**: Simulated again on SpiNNaker after an implementation bug is fixed (see [1])

► **PWC**: Pairwise correlations via compound Poisson processes of order 2 + homogeneous Poisson processes. Constructed such that the expected distribution of correlation coefficients is identical to the HOC data.



Correlation Matrices: Color-coded correlation coefficient for every neuron pair.



● high performance computing ● neural network simulation
● neuromorphic computing ● data analysis
● visualization

An example validation workflow between simulation runs on the NEST & SpiNNaker simulators implemented in the HBP collaboratory [1]

Conclusion

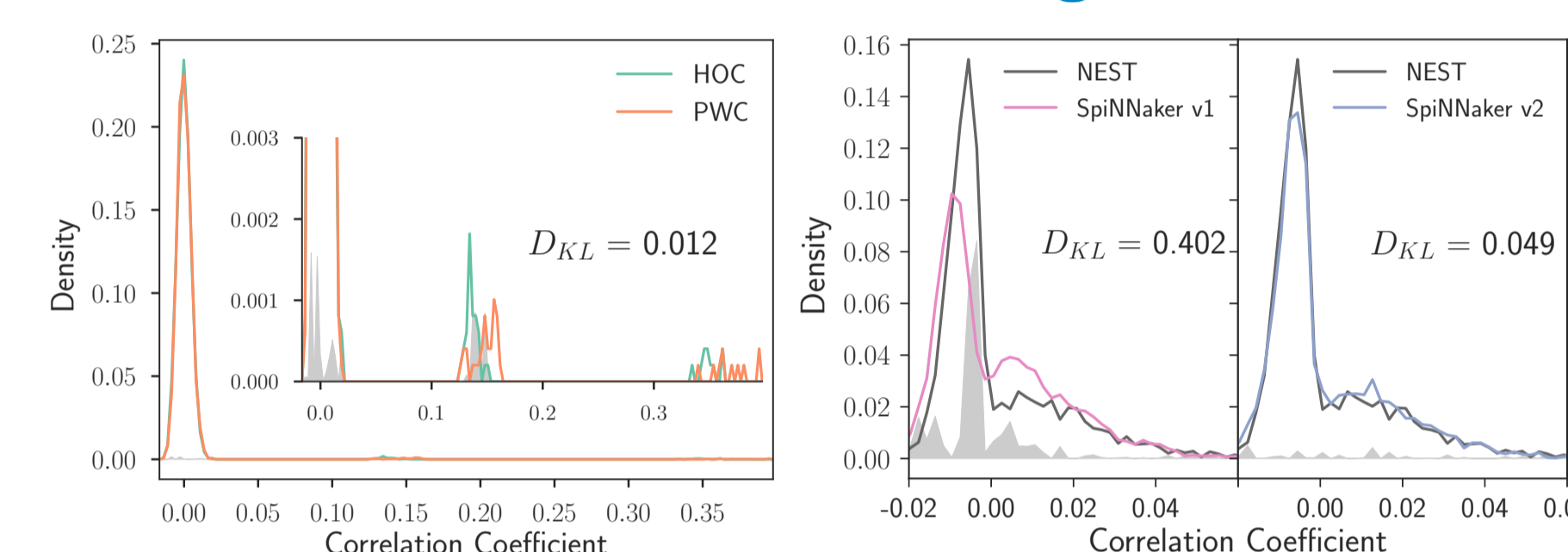
	Stochastic Data	Simulation Data
Two Sample Testing	<ul style="list-style-type: none"> ► The tests correctly conclude the similarity of correlation coefficients. ► Test results vary notably for different runs due to stochasticity. 	<ul style="list-style-type: none"> ► Hypothesis of similarity is rejected for all tests and simulation versions. ► Improvement of v1 → v2 can be quantified by all three measures.
Eigenvalue Decomposition	<ul style="list-style-type: none"> ► The generated correlated groups of neurons can be correctly identified. ► The data sets can be quantitatively and visually distinguished. 	<ul style="list-style-type: none"> ► No salient correlation features can be identified by visual inspection. ► Future work: Quantitative characterization of the correlation structure

► Outlook: Development of a validation methods toolbox and integration into the validation framework of the Human Brain Project [8]

Two Sample Testing

Stochastic Data **Simulation Data**
Pearson pairwise correlation coefficient: $\rho(s_i, s_j)$,
 s_i : binary spike train (2 ms binning)

Kullback-Leibler Divergence

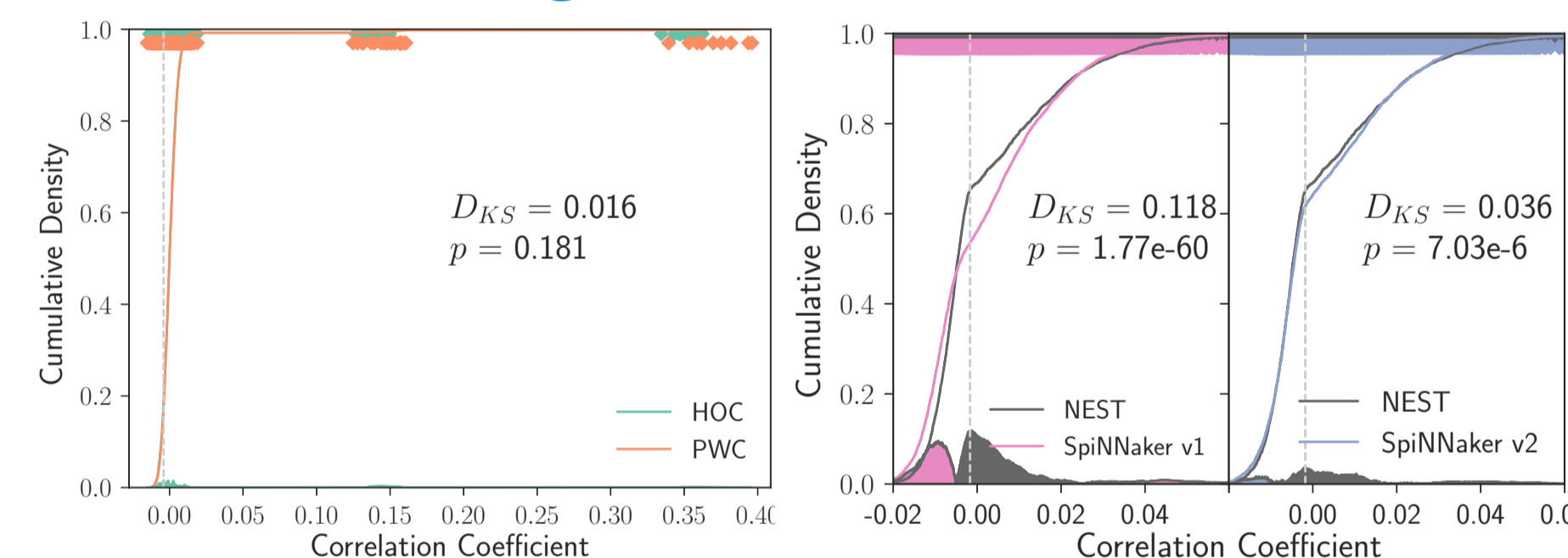


$$D_{KL}(P||Q) = \sum_i P(i) \log_2 \frac{P(i)}{Q(i)} = H(P, Q) - H(P)$$

$$D_{KL} = D_{KL}(P||Q) + D_{KL}(Q||P)$$

- Entropy-based measure of the difference between two distributions
- Interpretation: Information lost when substituting one distribution for the other
- Distributions are similar. D_{KL} is relatively low.
- Divergence was reduced by ~ 90% for v1 → v2.

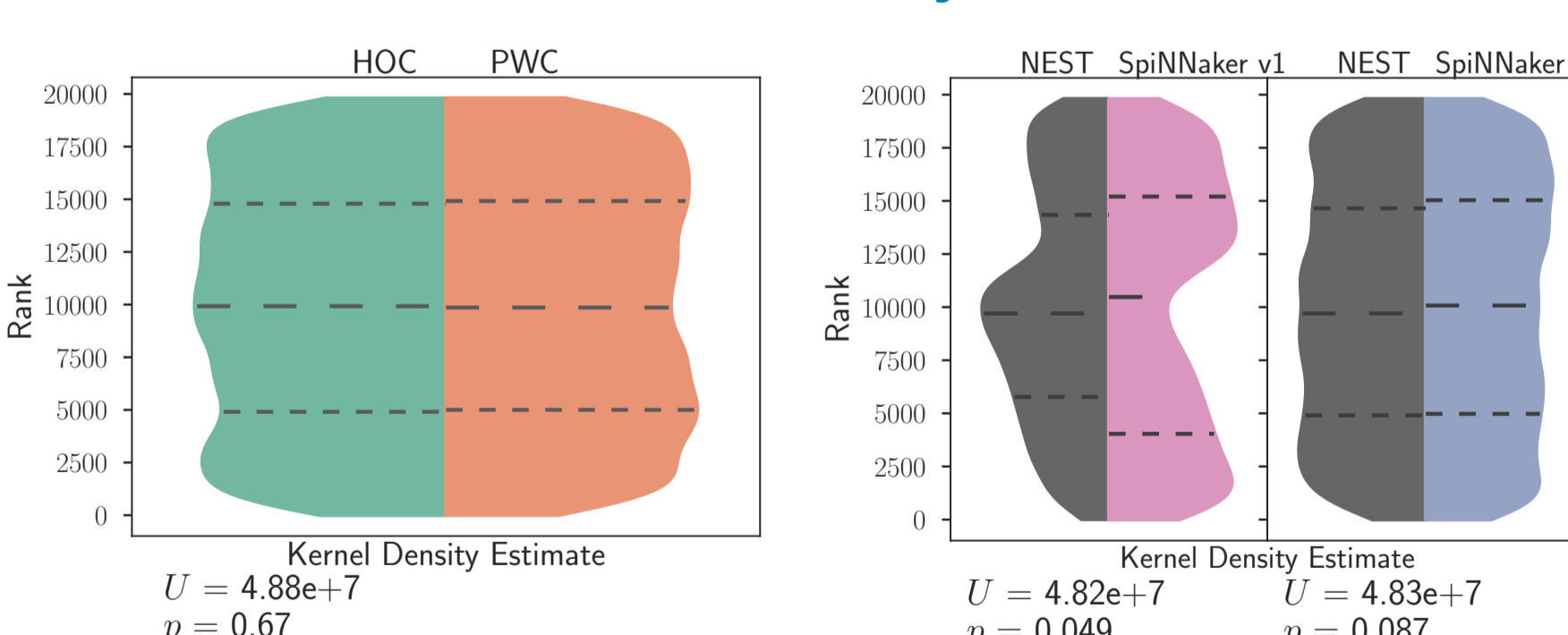
Kolmogorov-Smirnov Distance



$$D_{KS} = \sup |\hat{P}(x) - \hat{Q}(x)|$$

- Distance measure: Maximal vertical difference between the cumulative distributions
- Significant difference is assumed when $p < 0.05$.
- The small D_{KS} can't reject the hypothesis that the underlying distributions are identical.
- The D_{KS} was reduced by ~ 70%. Hypothesis of similarity is still rejected.

Mann-Whitney-U Test



$$U_i = R_i - \frac{n_i(n_i + 1)}{2}, \quad U = \min(U_1, U_2)$$

- R_i : rank sum, n_i : sample size
- $U \in [0, \frac{n_1 n_2}{2}]$ is a rank measure of sameness.
- Significant difference is assumed when $p < 0.05$.
- Rank sums are similar. Hypothesis of similarity is not rejected.
- Mismatch in the rank density is only barely detected in the rank sum statistic.

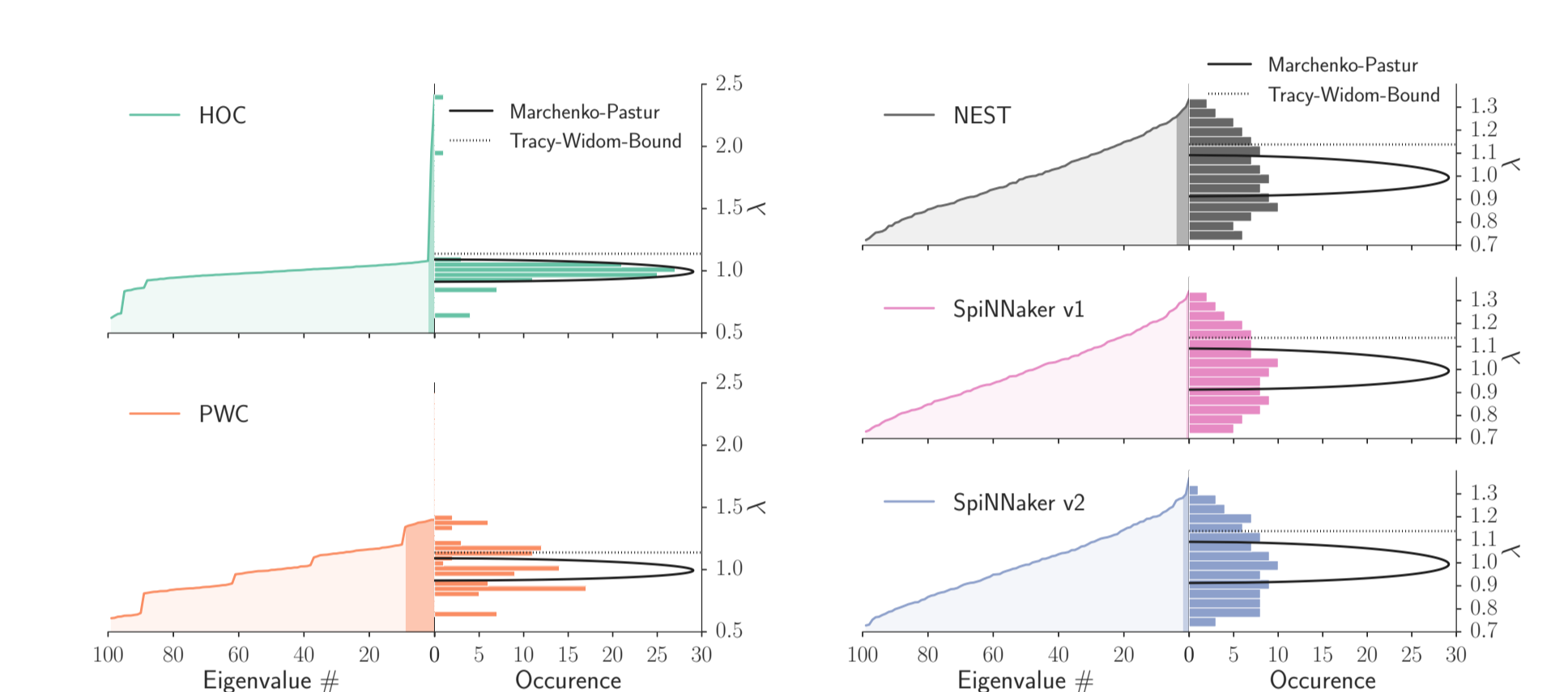
Test Confidence

- The test results may vary notably for a different parameter choice and due to the stochasticity of neurons.

Eigenvalue Decomposition

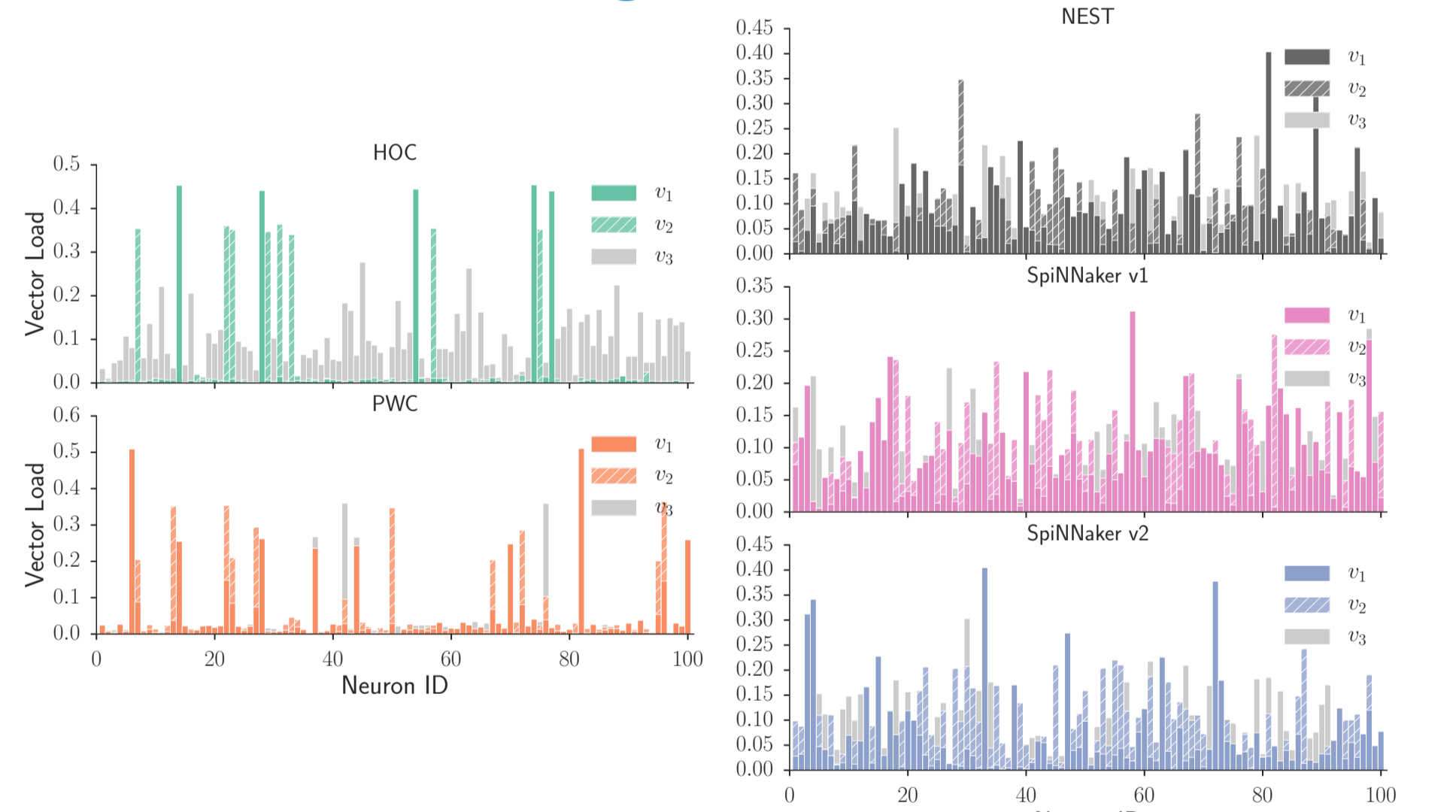
Stochastic Data **Simulation Data**
Performing Principle Component Analysis on the correlation matrices $C : (i, j) \mapsto \rho(s_i, s_j)$
 $C \cdot v_i = \lambda_i v_i$

Eigenvalue Distribution



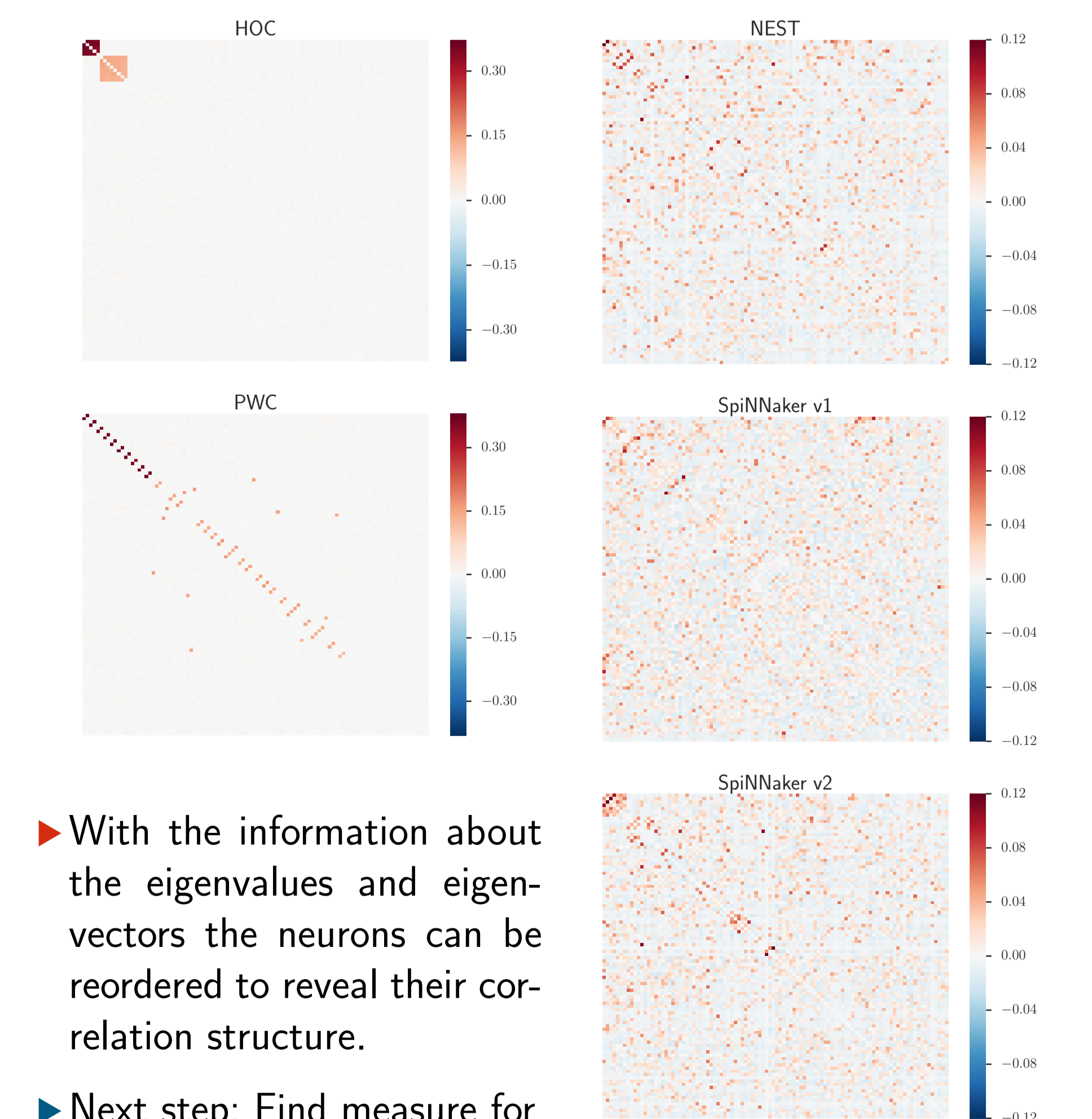
- The *Marchenko-Pastur distribution* describes the distribution of eigenvalues for an infinitely large correlation matrix.
- The *Tracy-Widom bound* accounts for the variance of the bounds for a finitely sized random matrix [7].
- Significantly large eigenvalues indicate dominant correlations in the network.

Eigenvectors



- The vector loads of eigenvectors of significant eigenvalues identify the corresponding groups of correlated neurons.

Ordered Correlation Matrices



- With the information about the eigenvalues and eigenvectors the neurons can be reordered to reveal their correlation structure.

- Next step: Find measure for the quantitative agreement of the correlation structure

References

- [1] Senk, J. et al. (2017). A Collaborative Simulation-Analysis Workflow for Computational Neuroscience Using HPC. In High-Performance Scientific Computing, E. Di Napoli, M.-A. Hermanns, H. Iliiev, A. Lintermann, and A. Peyser, eds. (Cham: Springer International Publishing), pp. 243–256.
- [2] Staude, Benjamin, Stefan Rotter, and Sonja Grün. "CuBIC: cumulant based inference of higher-order correlations in massively parallel spike trains." *Journal of computational neuroscience* 29.1 (2010): 327–350.
- [3] Staude, Benjamin, and Stefan Rotter. "Higher-order correlations in non-stationary parallel spike trains: statistical modeling and inference." *BMC Neuroscience* 10.1 (2009): P108. <http://www.nest-initiative.org/>
- [4] Potjans, Tobias C., Markus Diesmann "The cell-type specific cortical microcircuit: relating structure and activity in a full-scale spiking network model." *Cerebral cortex* 24.3 (2012): 785–806.
- [5] Van Albada, Sacha J., et al. "Full-scale simulation of a cortical microcircuit on SpiNNaker." *Frontiers in Neuroinformatics Conference Abstract: Neuroinformatics 2016*.
- [6] Peyrache, A., Benchenane, K., Khamassi, M., Wiener, S. I. & Battaglia, F. P. (2010) Principal component analysis of ensemble recordings reveals cell assemblies at high temporal resolution. *Journal of computational neuroscience* 29, 309–325.
- [7] NEST - SpiNNaker Workflow: collab.humanbrainproject.eu/#/collab/507/
NEST - SpiNNaker Validation: collab.humanbrainproject.eu/#/collab/2366/