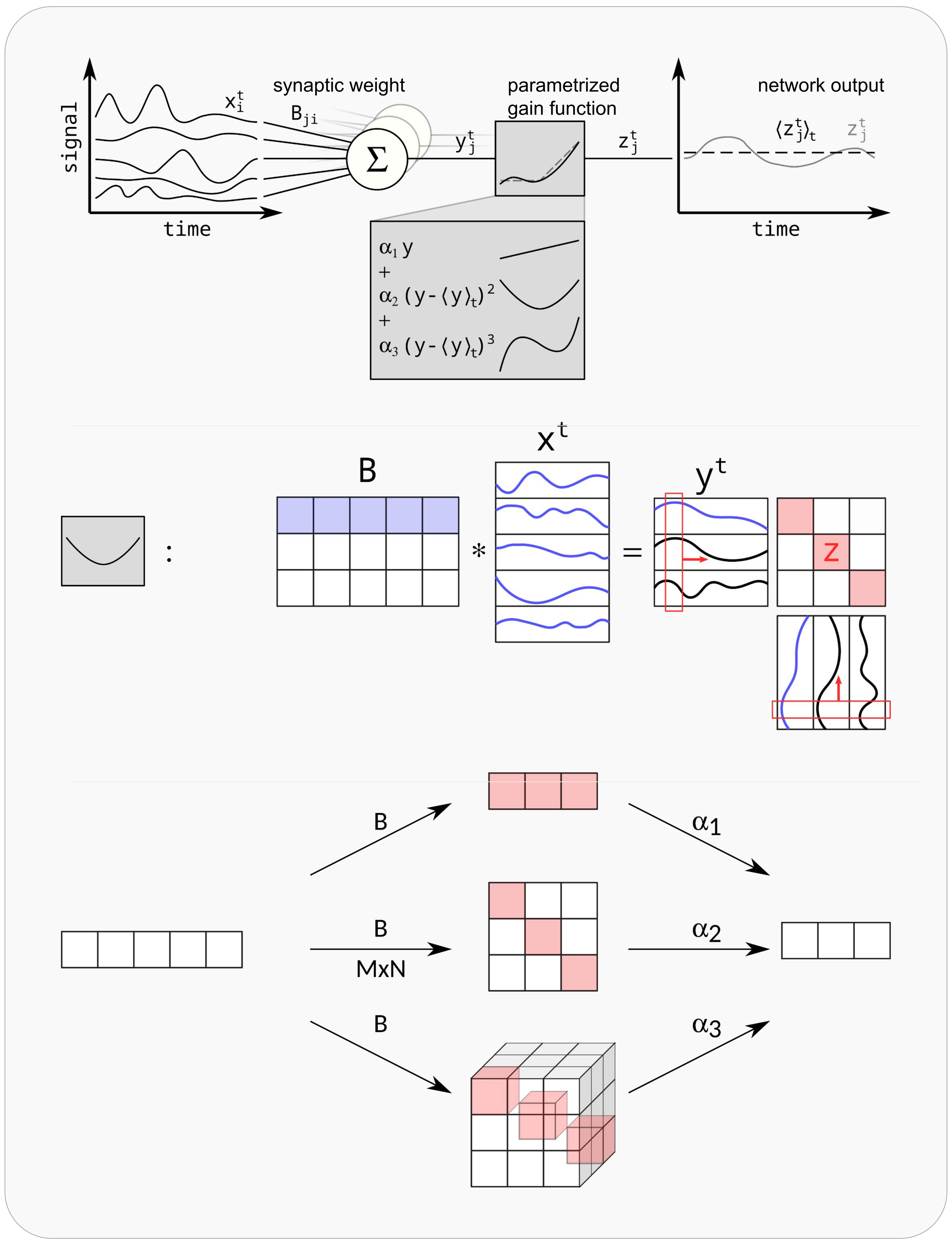
# Cumulant selective perceptron: Propagation of statistical information through a trainable non-linearity

Sandra Nestler[1,2], Moritz Helias[1,3], Matthieu Gilson[1,4]

[1] Institute of Neuroscience and Medicine (INM-6) and Institute for Advanced Simulation (IAS-6) and JARA-Institute Brain Structure-Function Relationships (INM-10), Jülich Research Centre, Jülich, Germany
[2] RWTH Aachen University, Aachen, Germany
[3] Department of Physics, Faculty 1, RWTH Aachen University, Aachen, Germany
[4] Institut de Neurosciences des Systèmes (INS, UMR1106), INSERM-AMU, Marseille (France)

## Setup

- Perceptron: The minimal network to extract information
- Linear operation: Sensitive only to mean of the data
- Covariance perceptron[1,2]: Sensitive to correlation structure
  - Corresponds to quadratic non-linearity applied to the inputs
- Combine arbitrary polynomials to obtain a cumulant selective perceptron
- Trainable weights indicate importance of corresponding cumulant




## Propagation of cumulants

- Initial perceptron layer followed by non-linearities

$$y^t_m = \sum_n B_{mn} x^t_n \qquad Y^{(i)}_m = \langle\langle (y^t_m)^i \rangle\rangle_t$$

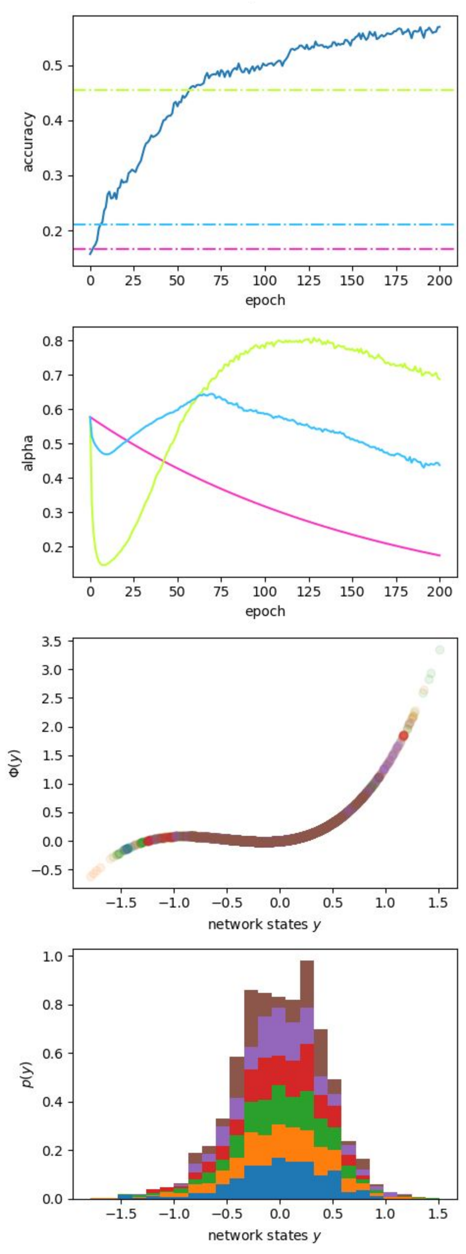
- Combine statistical orders linearly for readout

$$Z_m = \sum_i \alpha_i Y^{(i)}_m$$

- Train cumulant selective perceptron alongside fully connected model given the same statistical information

$$Y^{\mathrm{ML}\,(i)}_m = \sum_{n_1\dots n_i} B^{\mathrm{ML}\,(i)}_{mn_1\dots n_i} \langle\langle x^t_{n_1}\cdots x^t_{n_i}\rangle\rangle_t$$

- Joint weights of cumulant selective perceptron force to project only the most significant information to diagonals $Y^{(i)}_m$
- Linear readout parameters $\alpha_i$ represent importance of statistical orders
- Classify by the maximum entry of $Z_m$

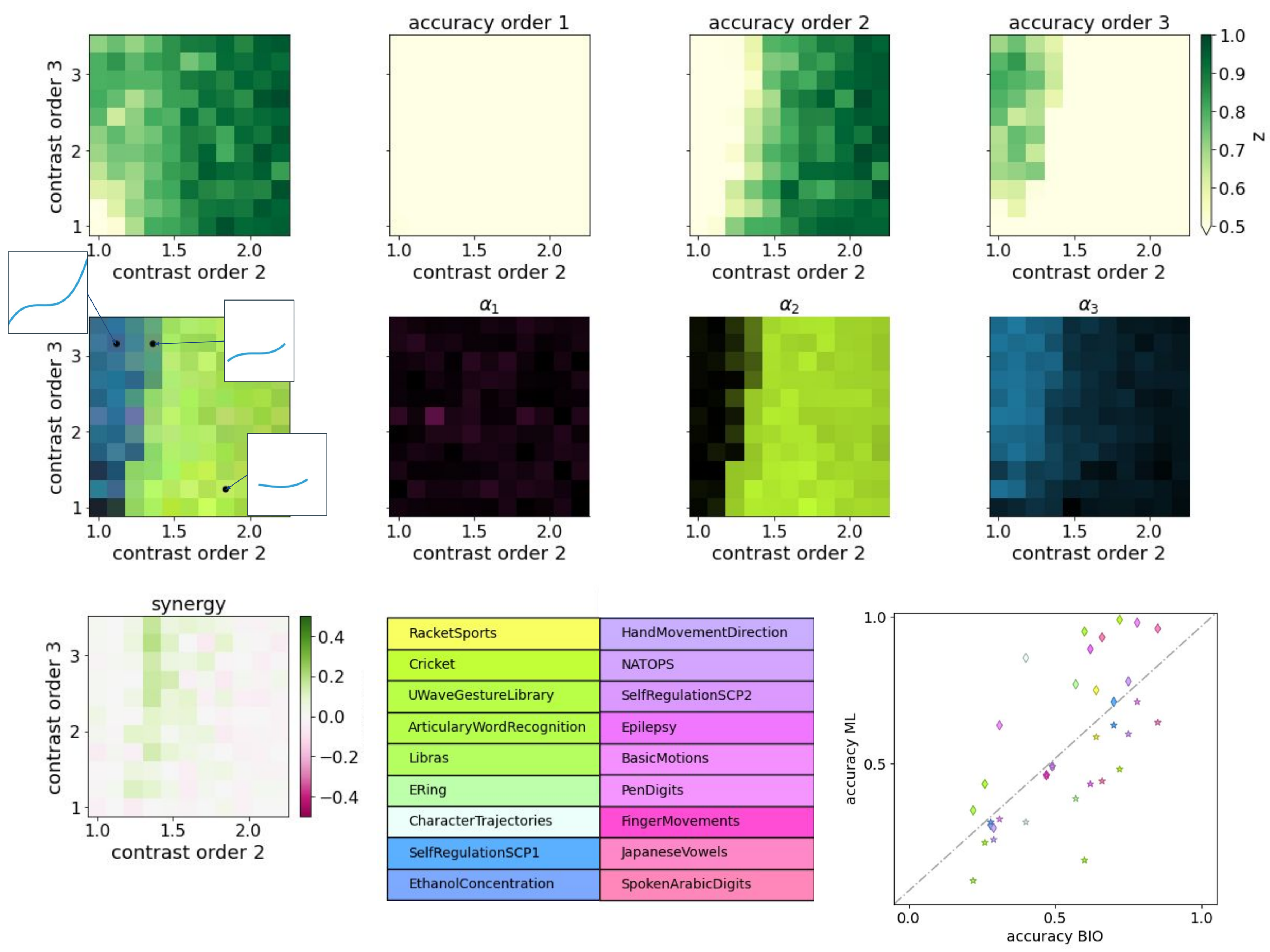



## Finding underlying structure in data

- Create a controlled setting with known statistical properties
- Use a stochastic differential equation with carefully chosen Lagrangian

$$\frac{\partial x_i(t)}{\partial t} = -\Gamma \frac{\delta L[x]}{\delta x_i} + \xi_i(t) \qquad L[x] = j^T x + \frac{1}{2} x^T J x + \frac{1}{3!}\sum_{ijk} K_{ijk} x_i x_j x_k$$

- Select parameters to obtain distribution with desired cumulants in large-time limit
- Cumulant selective perceptron trains to high accuracy, shows synergy effect of combining individual orders, and detects significant orders successfully
- Performance is comparable to fully connected model at full scale and better when restricted to the same number of trainable parameters
- Applied to benchmark datasets, can extract underlying information about statistics as well as approximately optimal gain function




## Summary

- Non-linear components of the activation function filter for different statistical information in the data
- Perceptron-like model with trainable non-linearity efficiently collects information for classification despite low parameter complexity
- Detection of underlying statistical structure verified using artificial data based on stochastic differential equation
- Combination of statistical orders gives synergy effect on classification performance
- Optimal activation function depends on the cumulants of the data

## References

[1] Gilson et al. (2020). The covariance perceptron: A new paradigm for classification and processing of time series in recurrent neuronal networks. PLoS computational biology, 16(10), e1008127.
[2] Dahmen et al. (2020). Capacity of the covariance perceptron. Journal of Physics A: Mathematical and Theoretical, 53(35), 354002.