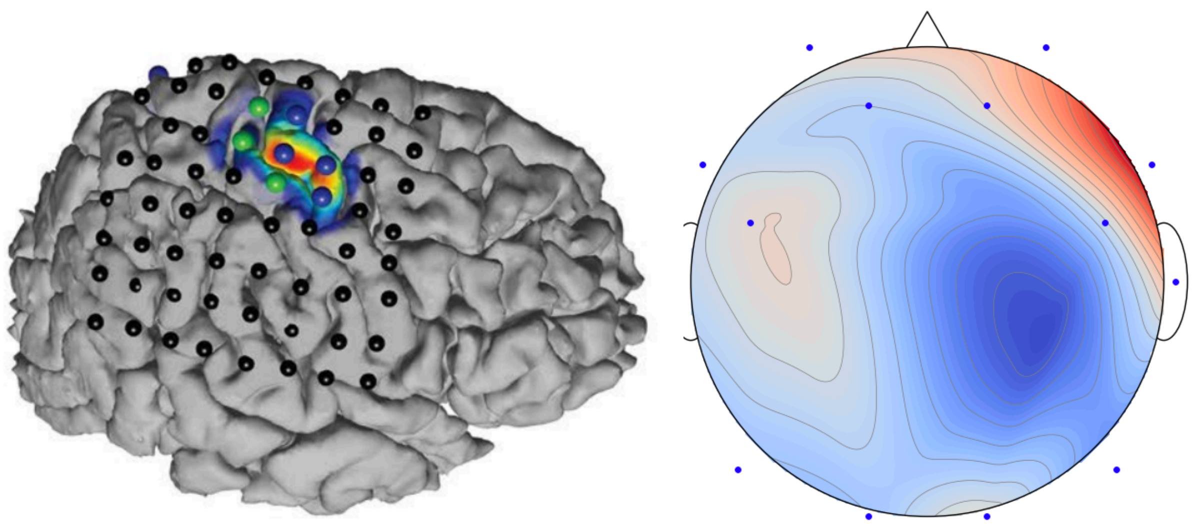
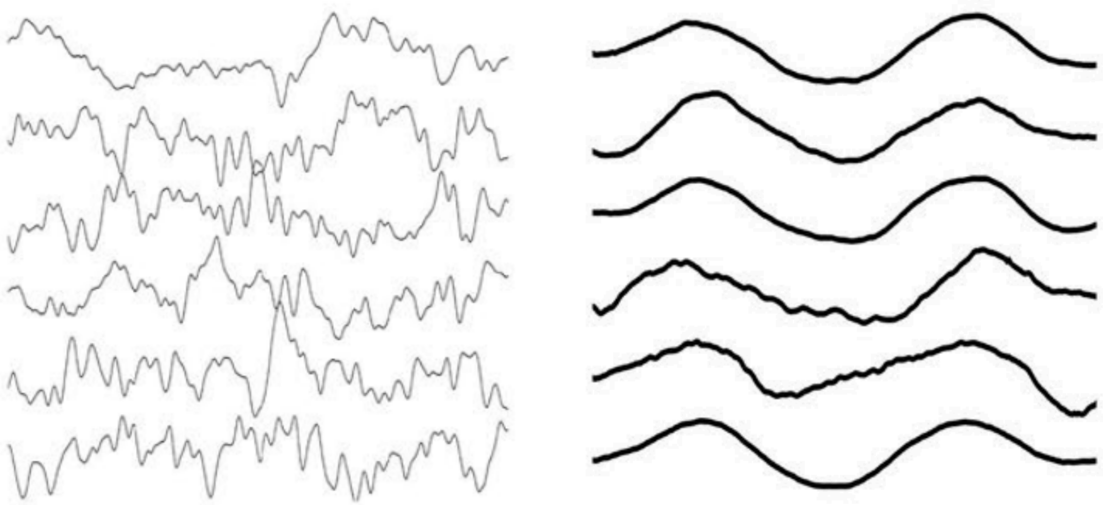
Spatial Filtering for Source Separation in EEG Data

Christian Endl, Felix Wirth, Korbinian Kunst & Taulant Koka

TECHNISCHE UNIVERSITÄT DARMSTADT

SIGNAL PROCESSING GROUP



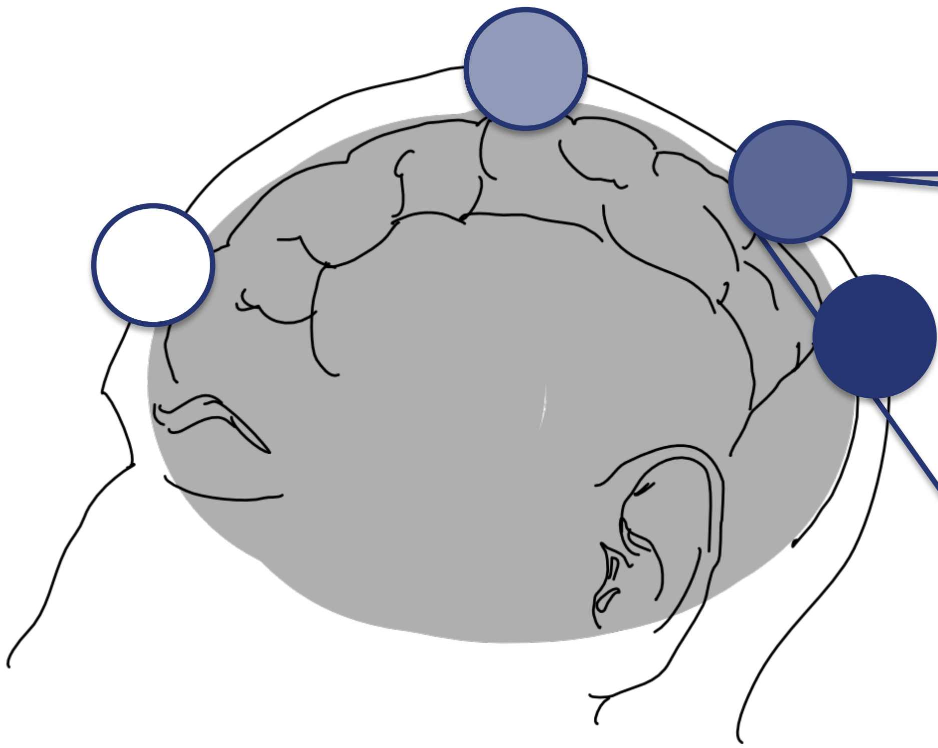


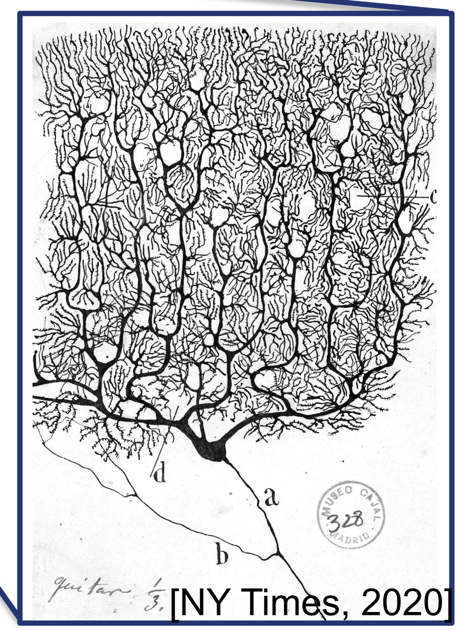
Agenda

1. Understanding the basis of Source Separation via an intuitive example-driven approach
2. Discussion of experiment design
3. Graph Blind Source Separation
4. Graph Blind Source Separation results
5. Summary of major outcomes
6. Problems and future research

Motivation
Challenge of Working with EEG Data

Electroencephalogram (EEG) Concept





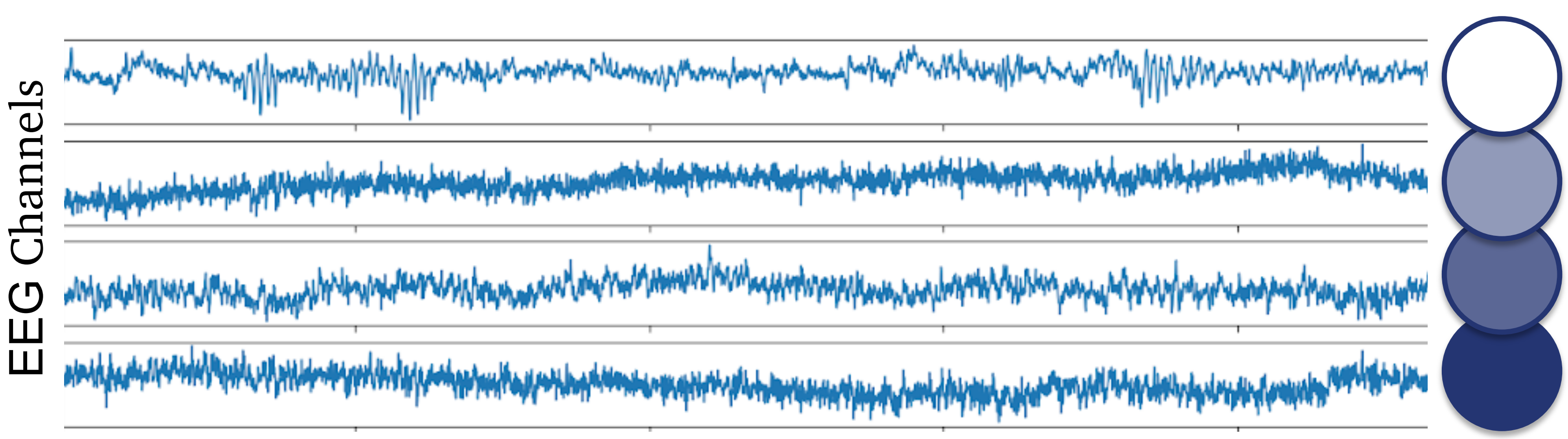
[NY Times, 2020]

What we wish to have?


[BackyardBrains]

- We are observing the "oscillating slow fields" of neurons in the upper layers of the cerebral cortex.
- These changes in electrical potential lead a neuron to be more likely or less likely to fire action potentials and are important in encoding information in the brain.

What do we get in experiments?



> Working with EEG data is therefore the process of source separation

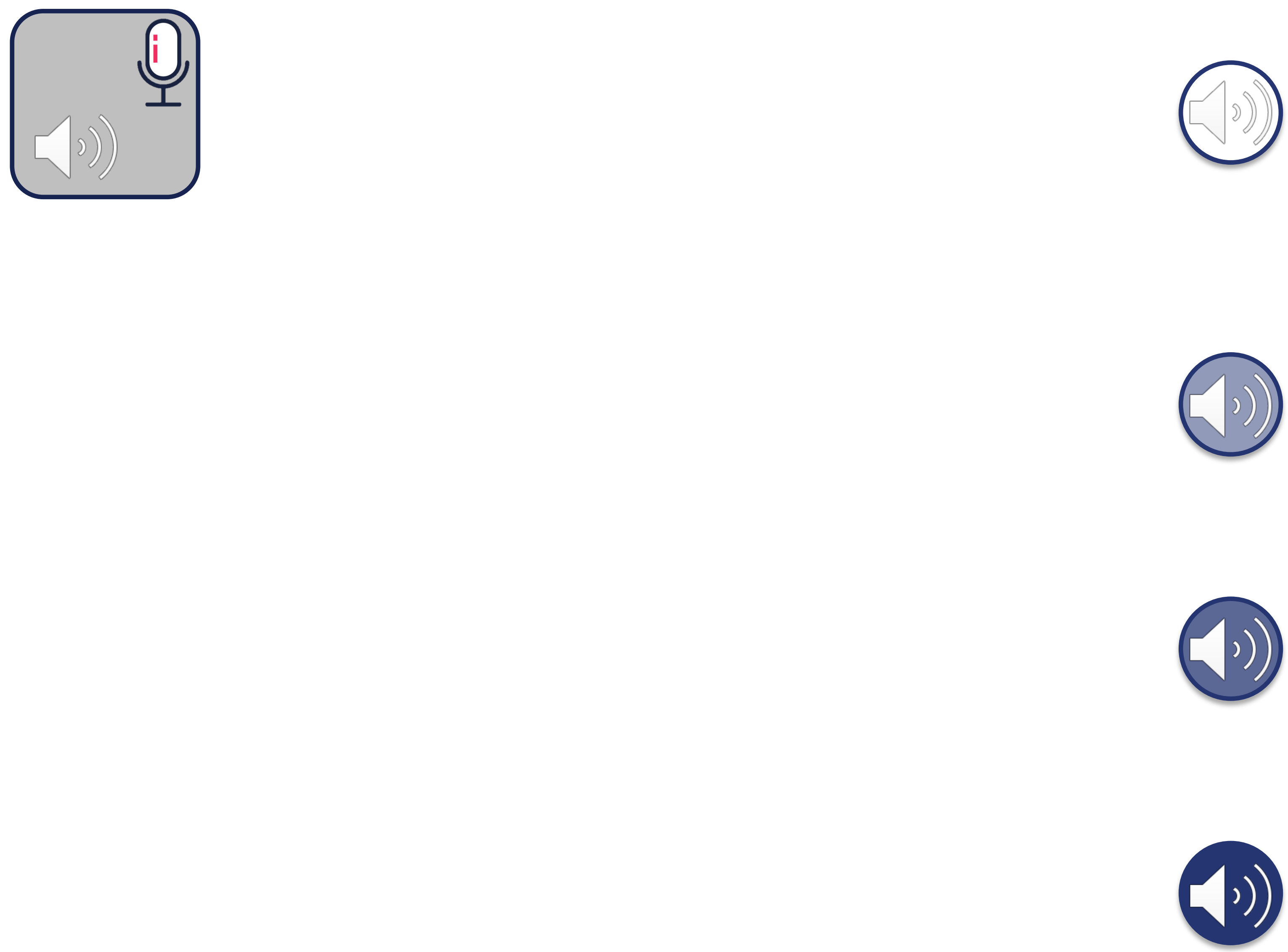
Spatial Filtering for Source Separation
Cocktail Party Problem

Def. I: In contrast to temporal filtering in DSP where a signal is analyzed as time-(in)variant subsequent discrete values of one channel, **spatial filtering** looks at the relation of a given timepoint across multiple channels.

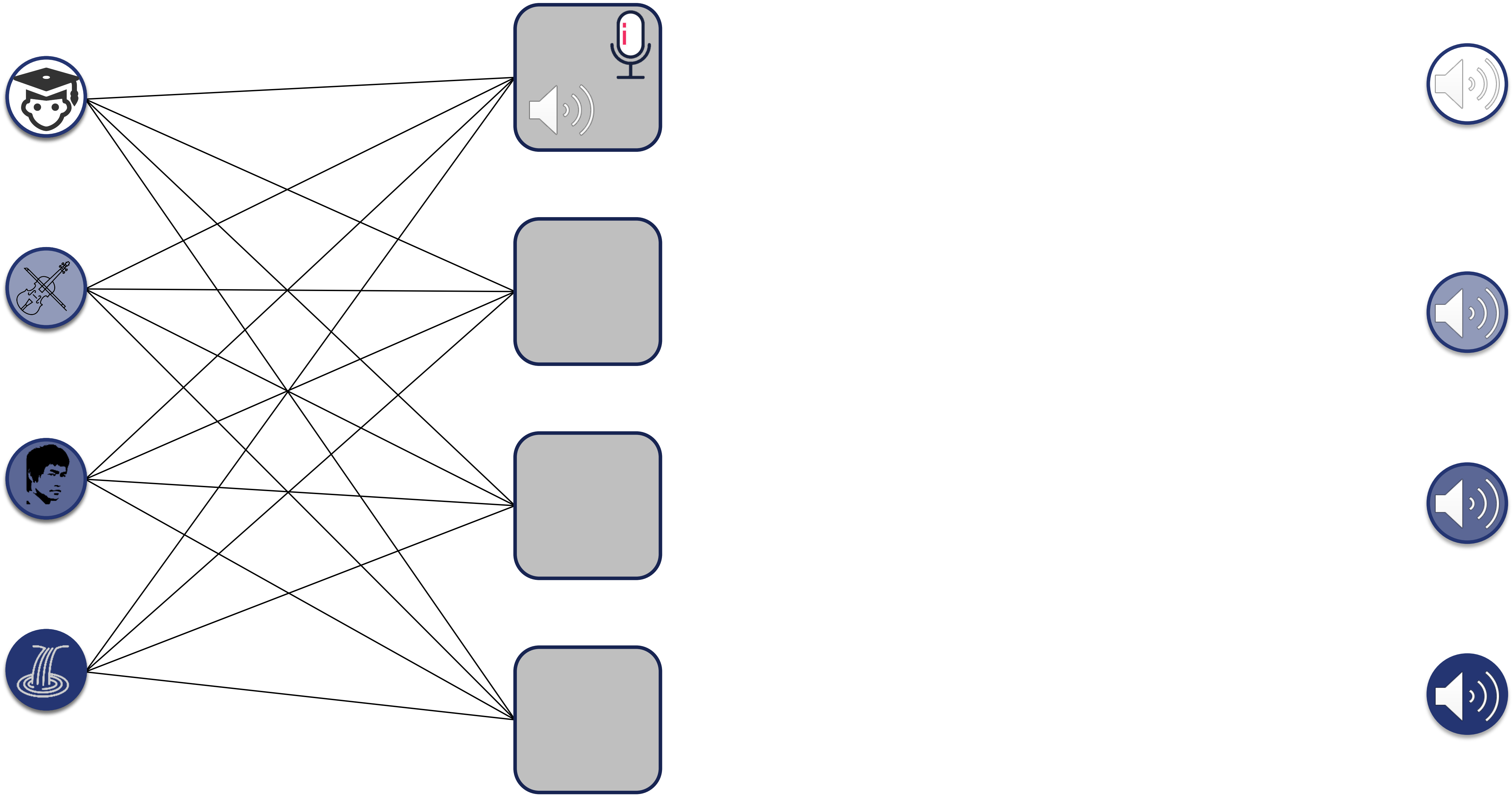
Def. II: The topic of separating mixed sources is called **B**lind **S**ource **S**eparation (BSS) and can be solved by algorithms called **I**ndependent **C**omponent **A**nalysis (ICA).

The Concept of Source Separation
Cocktail Party Problem

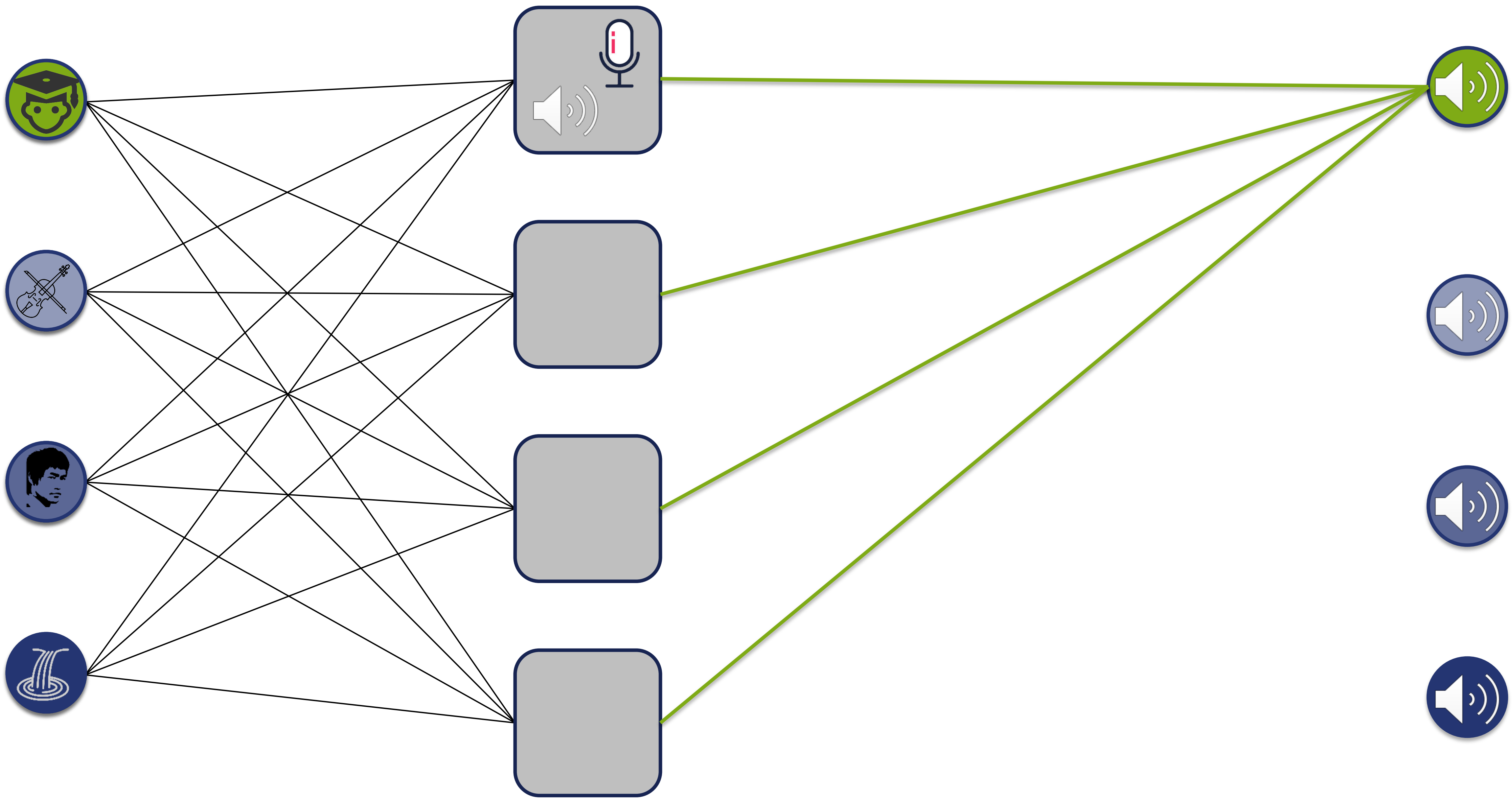
The Concept of Source Separation
Cocktail Party Problem



The Concept of Source Separation
Cocktail Party Problem



The Concept of Source Separation
Cocktail Party Problem

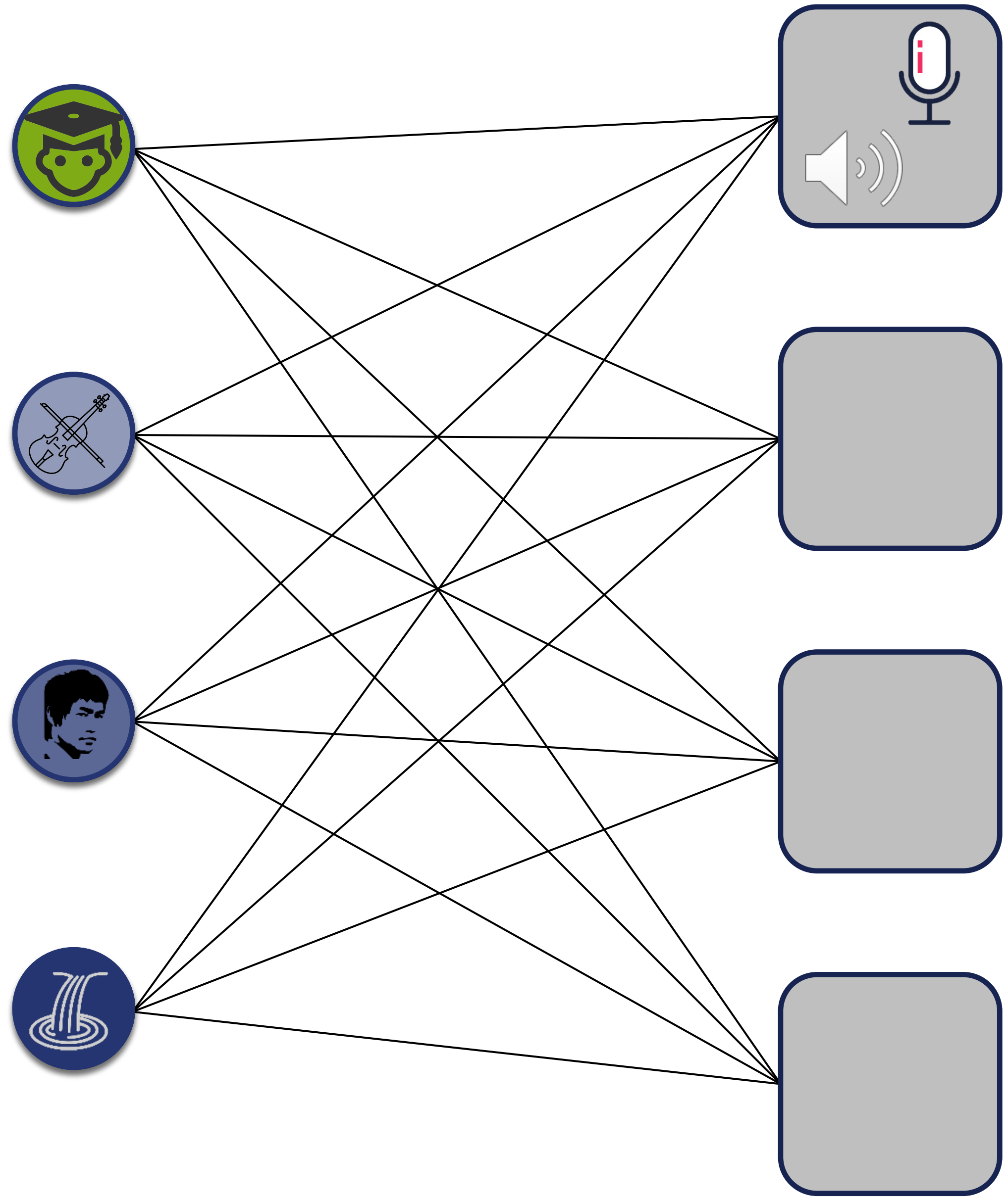


Original Signals (Sources)
Latent Variables

Sensors
Observed Data

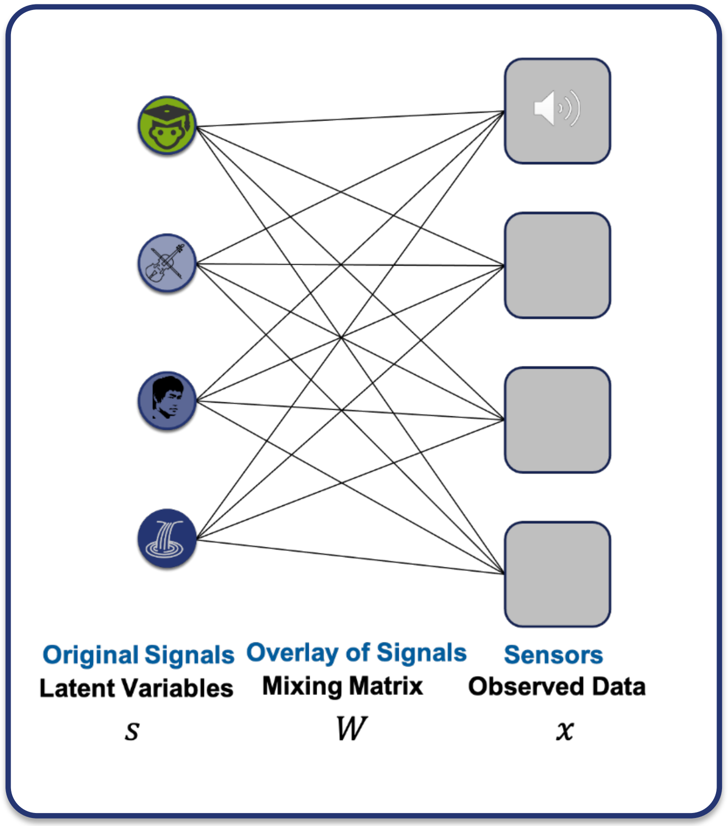
Reconstructed Signals
Result of the ICA

The Concept of Source Separation
Cocktail Party Problem

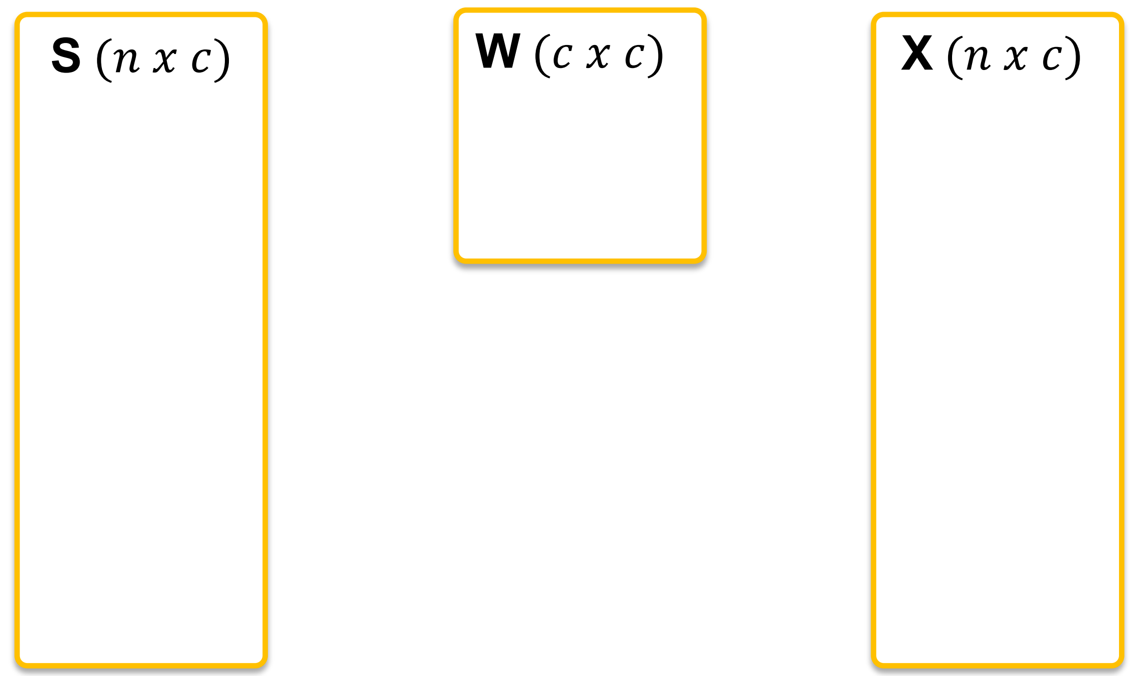


Original Signals (Sources)	Overlay of Signals	Sensors
Latent Variables	Mixing Matrix	Observed Data
S	W	X

The Concept of Source Separation
Math Formulation for ICA (Independent Component Analysis)



$$\mathbf{S}\,\mathbf{W} = \mathrm{X}$$



- $X = (\mathbf{x_1}, \ldots, \mathbf{x_c})^T$ matrix of vector mixtures is generated by $\mathbf{x} = \mathbf{As}$
- $\mathbf{s} = (\mathbf{s_1}, \ldots, \mathbf{s_c})^T$ is the unobserved matrix of statistically independent components (IC)
- $\mathbf{W} \ = (\boldsymbol{w}_1, \cdots, \boldsymbol{w}_c)$ is the unknown $\mathrm{c}{\times}\mathrm{c}$ mixing matrix of full rank, whose coefficient $\mathbf{w_{ij}} = [\mathbf{W}]_{\mathbf{ij}}$ represents the contribution of the j-th source onto i-th mixture.
- $\mathbf{A} = (\mathbf{a}_1 \cdots \mathbf{a}_c)^{\mathrm{T}} \triangleq \mathbf{W}^{-1}$ is the parameter of interest, defined by the inverse of the mixing matrix, called the demixing matrix and its (transposed) i-th row vector $\mathbf{a}_{\mathrm{i}}$ is called the i-th demixing vector.

Independent Component Analysis
Solving blind source seperation using ICA

TECHNISCHE UNIVERSITÄT DARMSTADT

What is the goal of the algorithm?

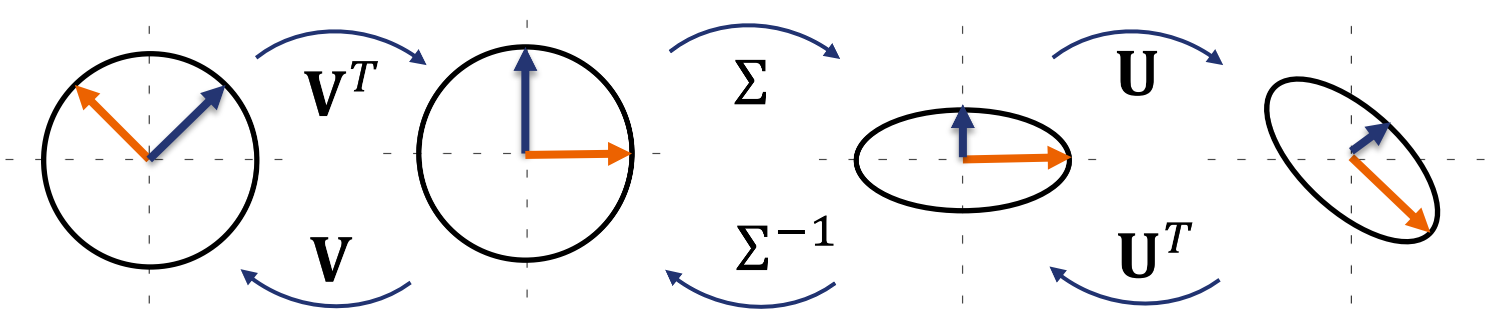
$$\mathbf{S}\,\mathbf{W} = \mathbf{X} \qquad \hat{\mathbf{S}} = \mathbf{X}\mathbf{W}^{-1}$$

Find the demixing matrix $\mathbf{A} \triangleq W^{-1}$ so that the latent source variable **S** can be reconstructed as $\hat{\mathbf{S}}$ by the observed data $\mathbf{X}$

How can we find W based on X?

- Under-constrained problem: unknown $\mathbf{S}$ and $\mathbf{W}$ → focus on finding $\mathbf{W}$ based on underlying statistics of $\mathbf{X}$
- Don't solve for $\mathbf{W}$ all in one: step wise
- Remember Singular Value Decomposition (SVD) ?
  Every symmetric matrix can be decomposed into a rotation $\mathbf{V}$, a stretch along the axes $\mathbf{\Sigma}$, and a second rotation $\mathbf{U}$.

$$\mathbf{W} = \mathbf{U}\Sigma\mathbf{V}^T \qquad \mathbf{A} = \mathbf{V}\Sigma^{-1}\mathbf{U}^T$$



Independent Component Analysis
Solving blind source seperation using ICA

TECHNISCHE UNIVERSITÄT DARMSTADT

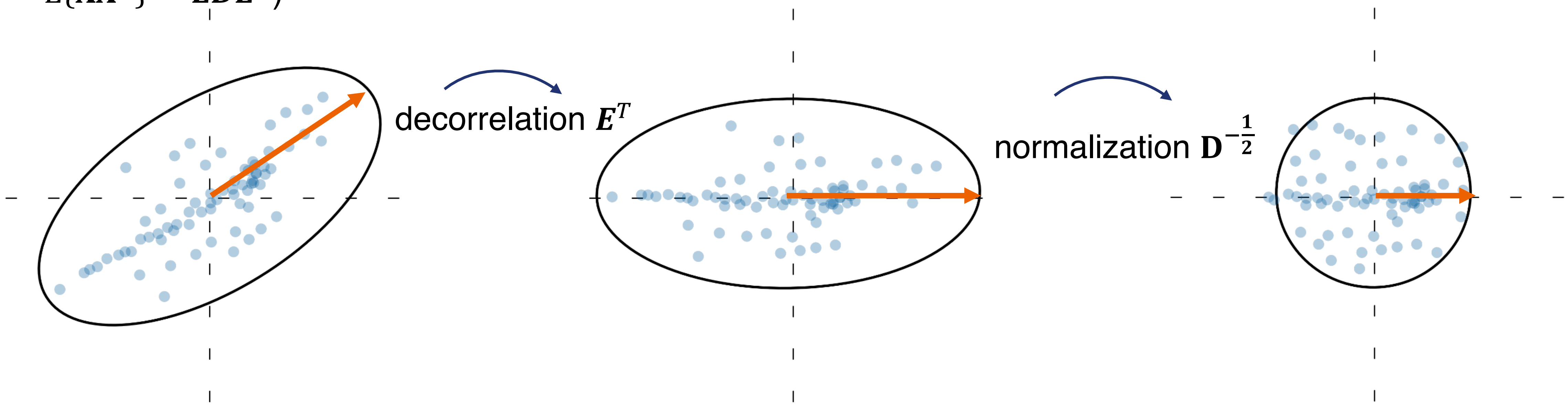
If we can find each of the SVD components we can determine the unmixing matrix A

$$\mathbf{A} = \mathbf{V}\Sigma^{-1}\mathbf{U}^T$$

It can be shown that this can be formulated as an eigenvector problem, where:

$$\boldsymbol{\Sigma}^{-1} = \mathbf{D}^{-1} \quad \& \quad \mathbf{U}^T = \boldsymbol{E}^T$$

D, E are the singular values and eigenvectors computed from the covariance matrix of **X** ($\text{Cov} = \text{E}\{\mathbf{X}\mathbf{X}^\text{T}\} = \mathbf{E}\mathbf{D}\mathbf{E}^\text{T}$)



After this Process $\mathbf{x}_w = \left(\mathbf{D}^{-\frac{1}{2}}\mathbf{E}^T\right)\mathbf{x}$ which is called whitening we ware left with a distribution where all second order correlations are removed [Shlens, 2010]:

→ Problem simplified to finding: $\hat{\boldsymbol{S}} = \mathbf{V}\boldsymbol{X}_w$

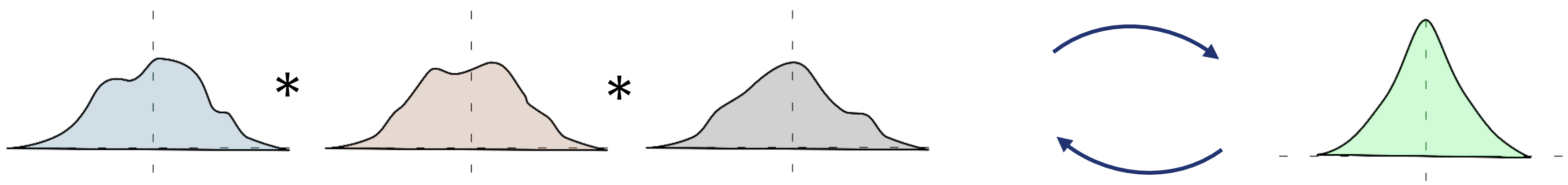
need to exploit the statistics of independence to identify **V**

Notes On ICA
Assumptions

How to measure statistical independence?

- For ICA There are two main approaches [Haykin, 2009] [Hyvarinen, 2000]:
  a) rooted minimization of mutual information
  b) maximization of non-Gaussianity

Why is it possible?



- **Central limit theorem**: A mixture of multiple independent random variables (RV) with non-Gaussian distribution will be more Gaussian than each RV individually
  - Meaning: multiple non-Gaussian distributions tend to form a gaussian distributions
  - We can use this: Gaussian RV can also be decorrelated to yield independent non-Gaussian components
- How can we measure **non-Gaussianity** ?
  - Kurtosis, Entropy and measures derived from this [Hyvarinen, 2000]

What are general assumptions for ICA?

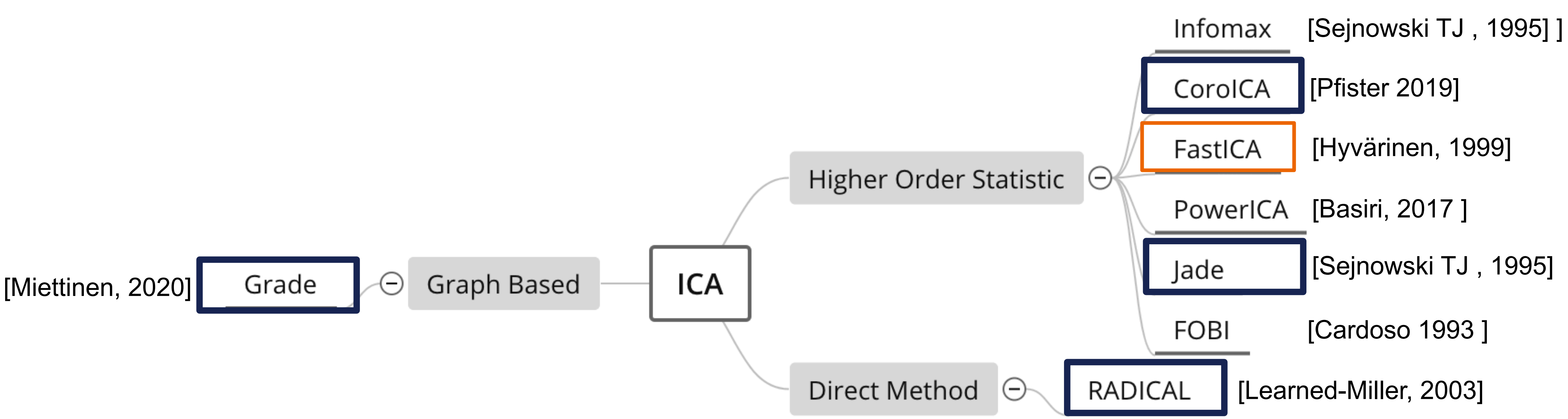
- Sources are statistical independent & non gaussian for identifiability of the model [Comon, 1994]
- Mixture of signals is linear [Shlens, 2010]
- Number of Sensors $\geq$ Number of Sources [C. Jutten, 2010]
- Large sample size of signals
- Sources are stationary [C. Jutten, 2010]
- Observed signals are zero mean [C. Jutten, 2010]

Independent Component Analysis
A generic Algorithm description

Quick Summary of the Algorithm:

- 1. Subtract off the **mean** of the data in each dimension.
- 2. Whiten the data by calculating the **eigenvectors** of the co-variance of the data.
- 3. Identify final rotation matrix that optimizes **statistical independence**
  - Contribution of ICA algo: formulate a contrast function that measures how close the estimated sources are to statistical independence
  - Reach goal by minimizing or maximizing defined objective

ICA Algorithms – a subjective selective overview:



Independent Component Analysis
Example Concept PowerICA [Basiri, 2017]

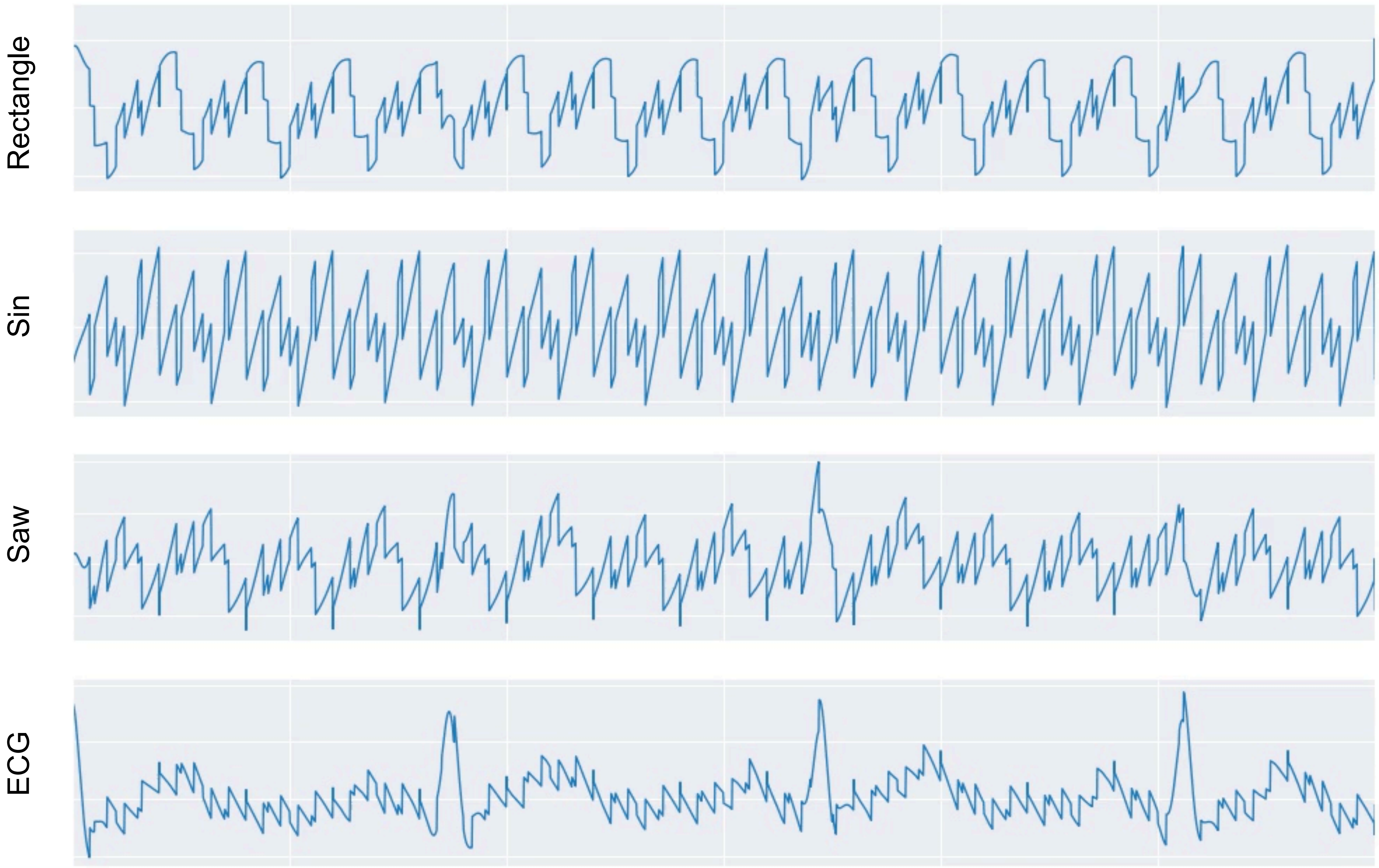
$$\mathcal{L}(\mathbf{w};\lambda) = \underbrace{|\mathbb{E}[G(\mathbf{w}^\top\mathbf{x})]|}_{\text{local maxima}} - \underbrace{\frac{\lambda}{2}(\mathbf{w}^\top\mathbf{w}-1)}_{\|w\| = w^T w = 1}$$

- LaGrangian optimiziation problem for find rotation matrix V
- non-Gaussianity measure
- unit-norm constraint
- G can be any twice continuously differentiable nonlinear and non-quadratic function
  - $g(u) = u\exp(-a_2u^2/2)$ or $g_1(u) = \tanh(a_1u)$ [Hyvärinen, 1999]
- λ is Lagrange Multiplier
- $\mathbf{w}$ itself is found via a stepwise Newton Raphson optimization algorithm

ICA Signal Reconstruction

From Mixed Example Signals to Deconstructed Signals

TECHNISCHE UNIVERSITÄT DARMSTADT



Agenda

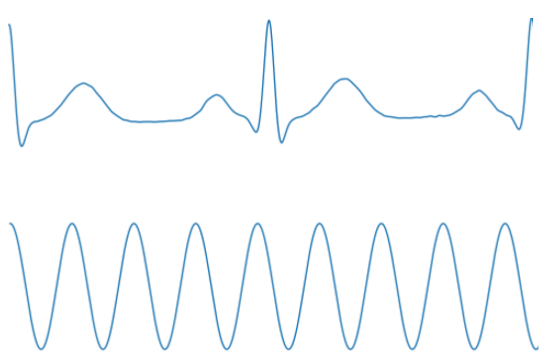
Experiment Design
Goal

Goals of this project:

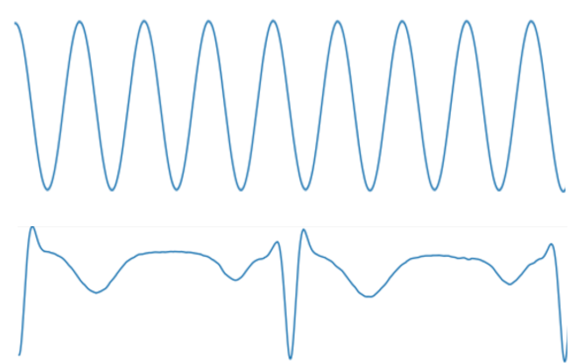
- Evaluate ICA - Algorithms in terms of robustness
- Elaborating new approaches to increase robustness against noise and outliers in EEG - measurements

How can we evaluate the ICA-Algorithms?

- The ICA-Algorithms return the reconstructed signal $\hat{\mathbf{X}}$ and an estimated mixing matrix $\hat{\mathbf{W}}$
- 2 options: compare input $\mathbf{X}$ and reconstructed signals $\hat{\mathbf{X}}$ or compare mixing matrix $\mathbf{W}$ with its estimate $\hat{\mathbf{W}}$



vs.



$$\begin{pmatrix} 0.23 & 0.75 \\ 0.55 & 0.91 \end{pmatrix} \; vs. \; \begin{pmatrix} 0.23 & 0.55 \\ 0.75 & 0.91 \end{pmatrix}$$

$\mathbf{X}$ $\hat{\mathbf{X}}$ $\mathbf{W}$ $\hat{\mathbf{W}}$

Problems:

- The solution of ICA is not unique: $\hat{\mathrm{A}}$ and $\hat{\mathrm{X}}$ can pertubed, scaled or have sign changes
- Find metrics which deal with that

Experiment Design
Metrics: MD, MSE

TECHNISCHE UNIVERSITÄT DARMSTADT

Minimum Distance Index (based on MM)[Ilmonen, 2010]:

- Key idea: The inverse of the Mixing Matrix equals the Unmixing Matrix, so the multiplication is the identity matrix $\mathbf{W} = \mathbf{A}^{-1} \xrightarrow{\text{yields}} \mathbf{WA} = \mathbf{I}$

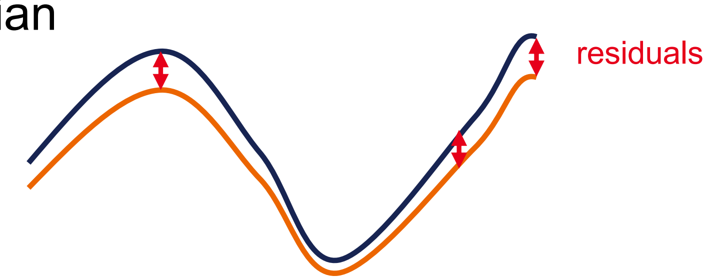
- Therefore: if $\hat{\mathbf{W}}\mathbf{A} = \mathbf{I} \Rightarrow$ The Algorithm performed well

$$\boxed{\mathrm{MD} = \frac{1}{\sqrt{(\mathrm{p}-1)}} \cdot \inf_{\mathrm{c}} \left\| \mathbf{C}\hat{\mathbf{W}}\mathbf{A} - \mathbf{I} \right\|_2}$$

- Where $\mathbf{C} = \mathbf{PJD}$ considers Permutation **P**, Scaling **D** and sign changes **J,** p = number of signals
- $\inf_C$ = Minimize Frobenius Norm
- 0 indicates a good performance, 1 indicates a bad one

Mean squared Error (MSE):

- Taking the Mean of the squared residuals
- Obtained values = number of signals ⇒ take mean or median



Problem:

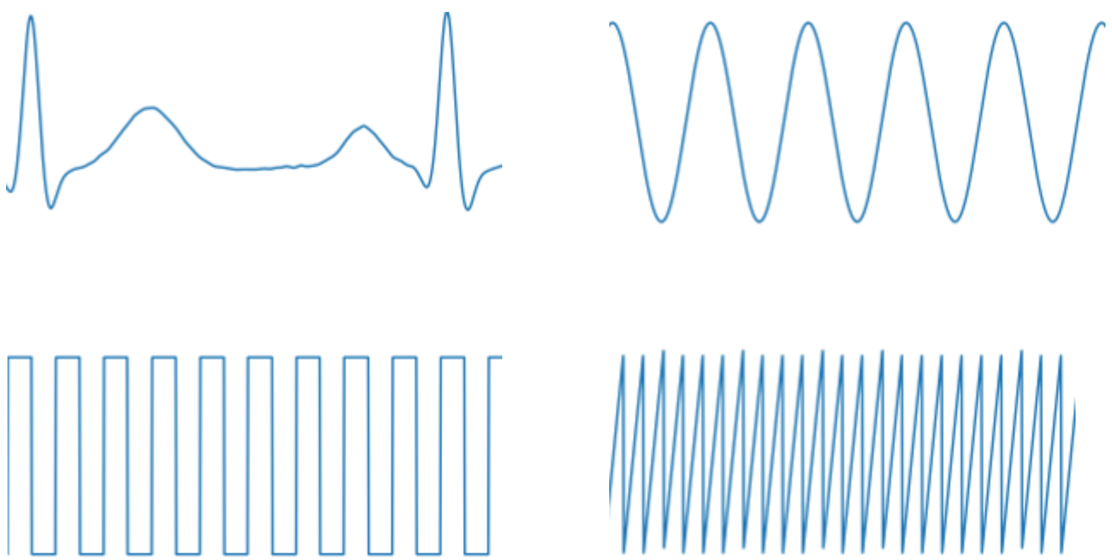
Required groundtruth data for **X** and **W**

Experiment Design
General approach

What we did:

1. We generated ideal Input Signals $\mathbf{X} \in \mathbb{R}^{4xN}$ and a random Mixing Matrix. $\mathbf{W} \in \mathbb{R}^{4x4} \Rightarrow$ groundtruth

From us defined Standard Signals



- ECG Signal: 2Hz
- Sinus: 5Hz
- Rectangle: 10Hz
- Sawtooth: 25Hz
- fs: 1000Hz
- Normalized to [-1,1]

Groundtruth mixing matrix $\mathbf{W}$

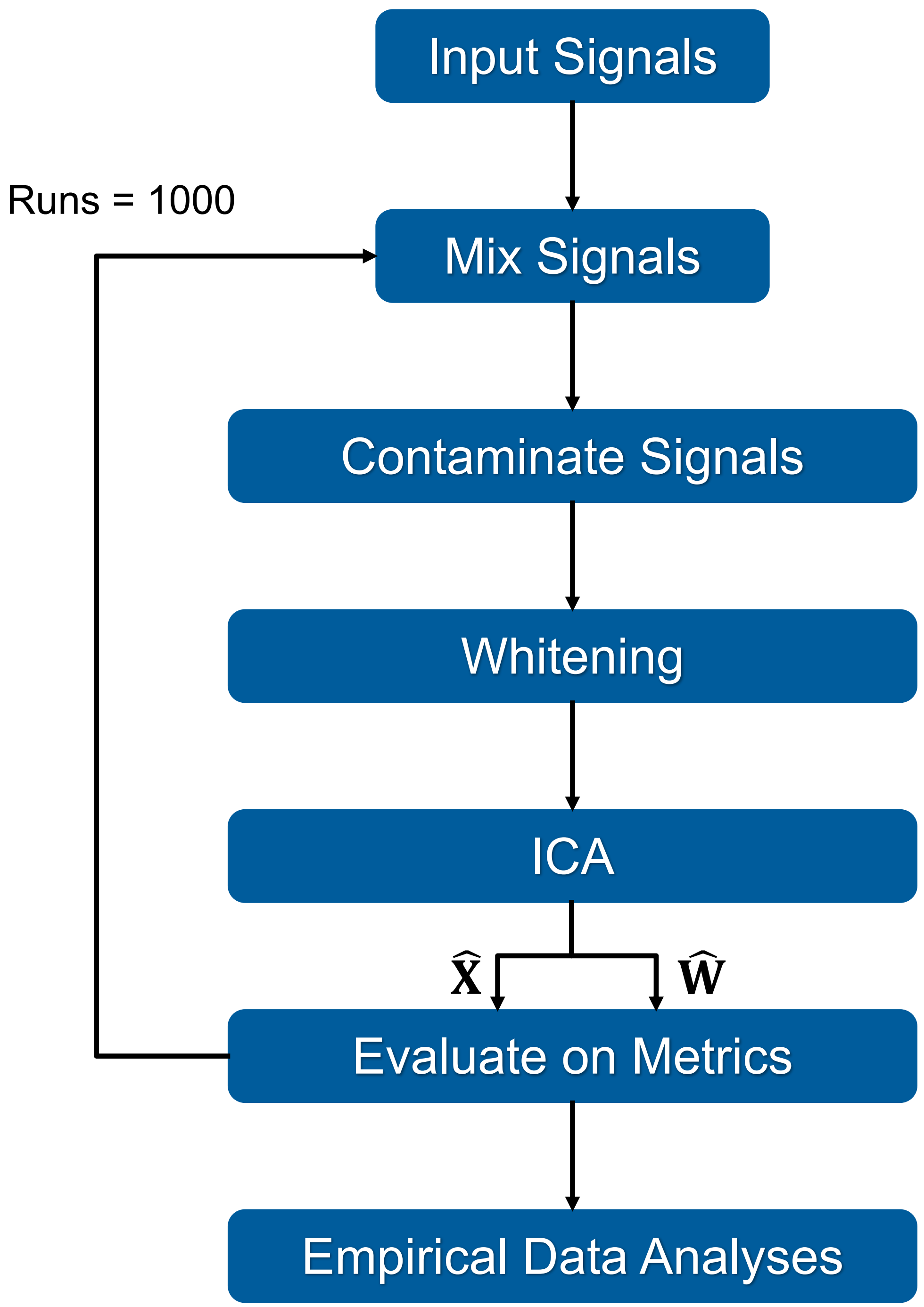
$$\mathbf{W} = \begin{pmatrix} 0.11 & 0.34 & 0.77 & 0.24 \\ 0.41 & 0.91 & 0.63 & 0.49 \\ 0.99 & 0.56 & 0.23 & 0.33 \\ 0.38 & 0.45 & 0.03 & 0.23 \end{pmatrix}$$

Frequencies of Brainwaves [Koudelková2018] :

- $\delta - Waves: 0.5 - 3Hz$
- $\theta - Waves: 3 - 8Hz$
- $\alpha - Waves: 8 - 12Hz$
- $\beta - Waves: 12 - 38Hz$

Experiment Design

Monte Carlo Flow Chart



1. Standard input signals
   - ECG, Sinus, Rectangle, Sawtooth

2. Mix Signals with mixing matrix $\mathbf{W}$ (random)
   - $\mathbf{W} \in \mathbb{R}^{4x4}$

3. Contaminate signals
   - Add single or patchy Outlier (100std) and/or white noise (40dB)

4. Whitening process
   - Decorrelate Signals: meaning that the sample covariance matrix is diagonal and that all the diagonal elements are equal.

5. Perform ICA
   - 4 Algos: JADE, RADICAL, CoroICA, PowerICA

6. Evaluate single run
   - Minimum Distance, MSE

7. Evaluate ICA
   - Get 1000 MDs and MSEs. ► BoxPlot

Experiment Design
Definition of test characteristics

1.) Focus on Sample Size

Initialize standard signals with different Sample Sizes: 1000, 2500, 5000, 10000, 15000 evaluate on 1000runs

Type I: No Noise, No Outlier

Type II: 40dB SNR, No Outlier

Type III: No Noise, 0.1% Outlier(std = 100)

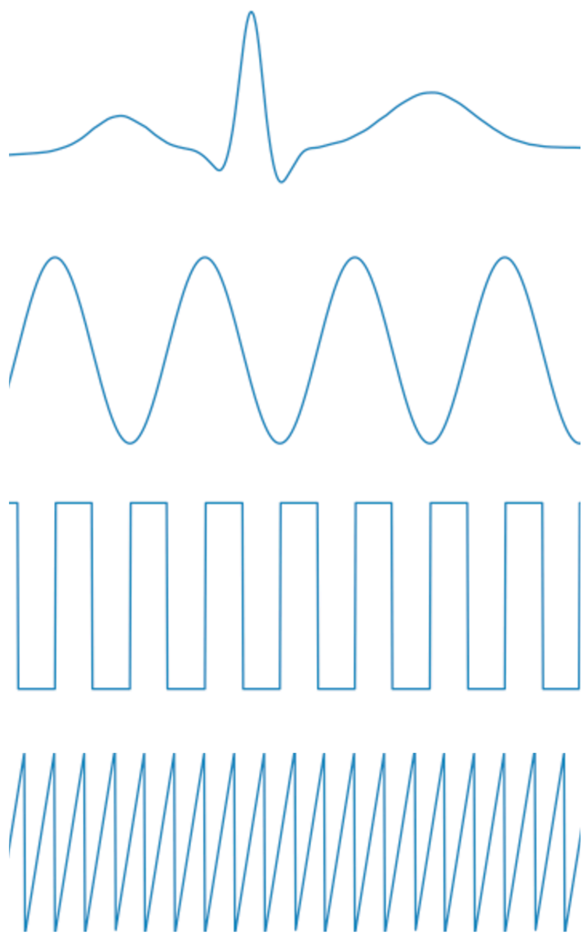
Type IV: 40dB SNR, 0.1% Outlier (std = 100)

2.) Focus on amount of noise – Breakdown Analyses

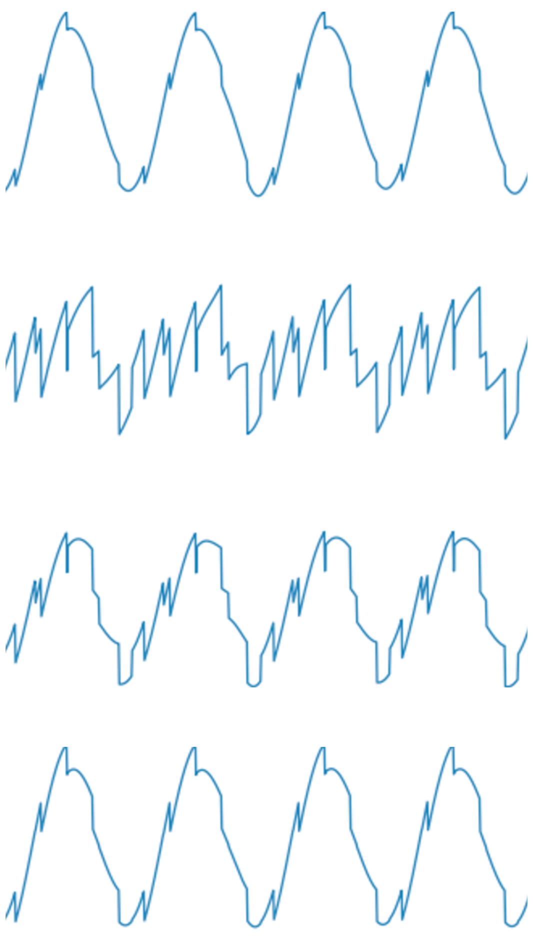
Initialize standard signals with fix Sample Size of 10000

Type V: increase noise, SNR: 40 dB, 30 dB, 20 dB, 10 dB, 6 dB, 3 dB

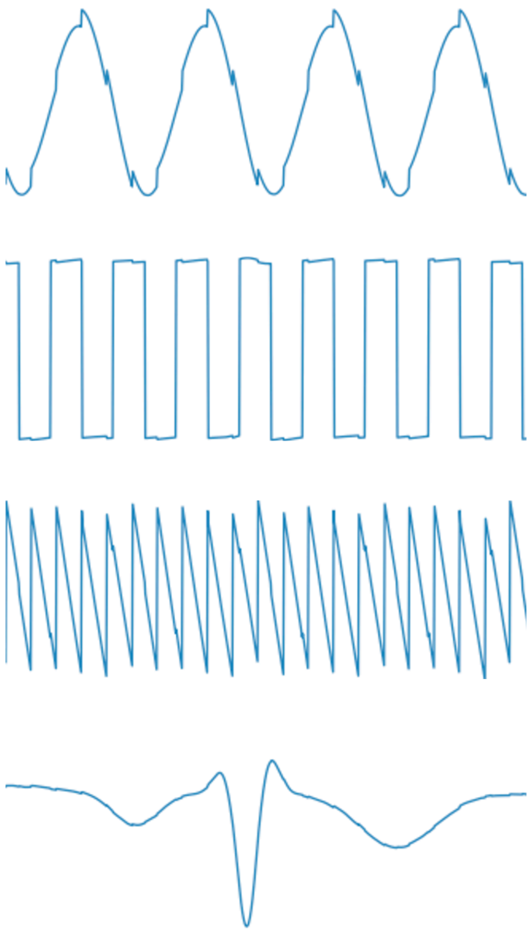
Results
Interpreting the Minimum Distance



$$\cdot\, \mathbf{W} =$$



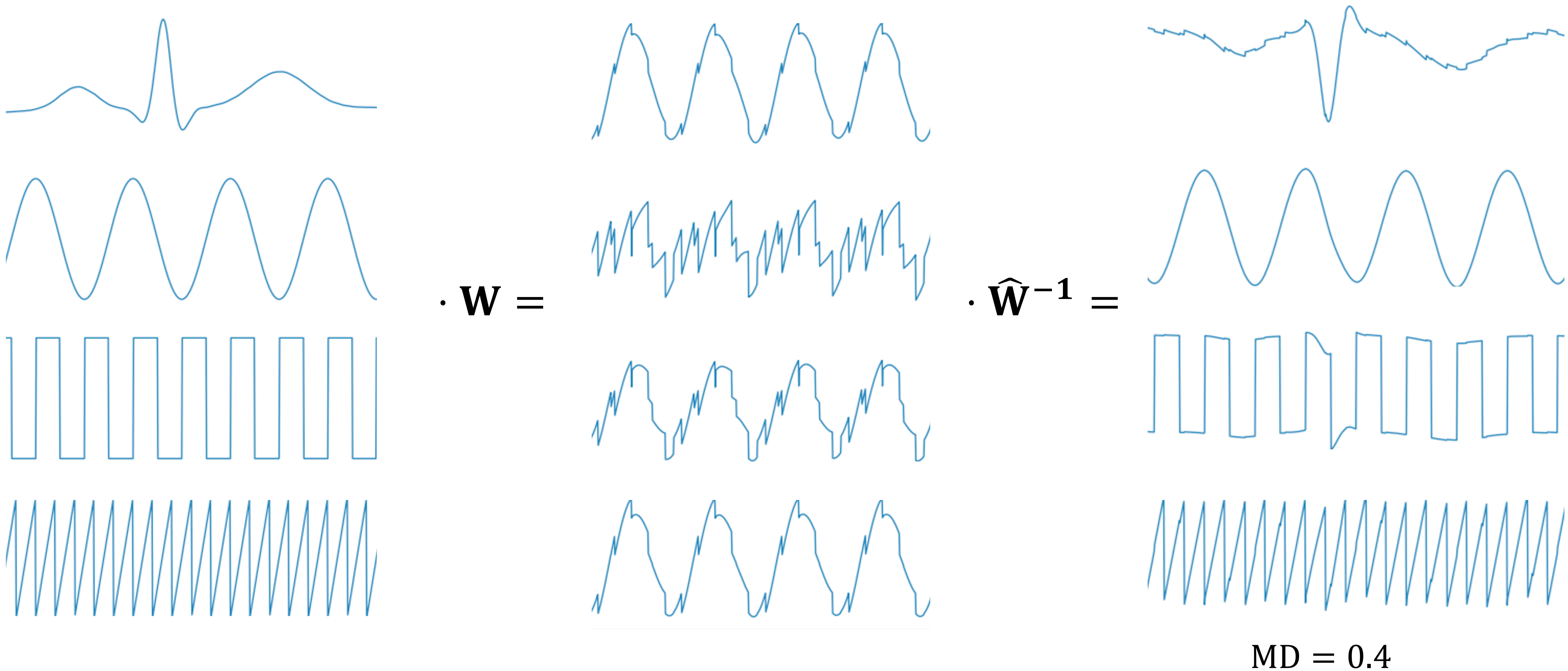
$$\cdot\, \widehat{\mathbf{W}}^{-1} =$$



$\mathrm{MD} = 0.2$

- $\mathrm{MD} = 0.2$: Good results, small spikes, some sign changes

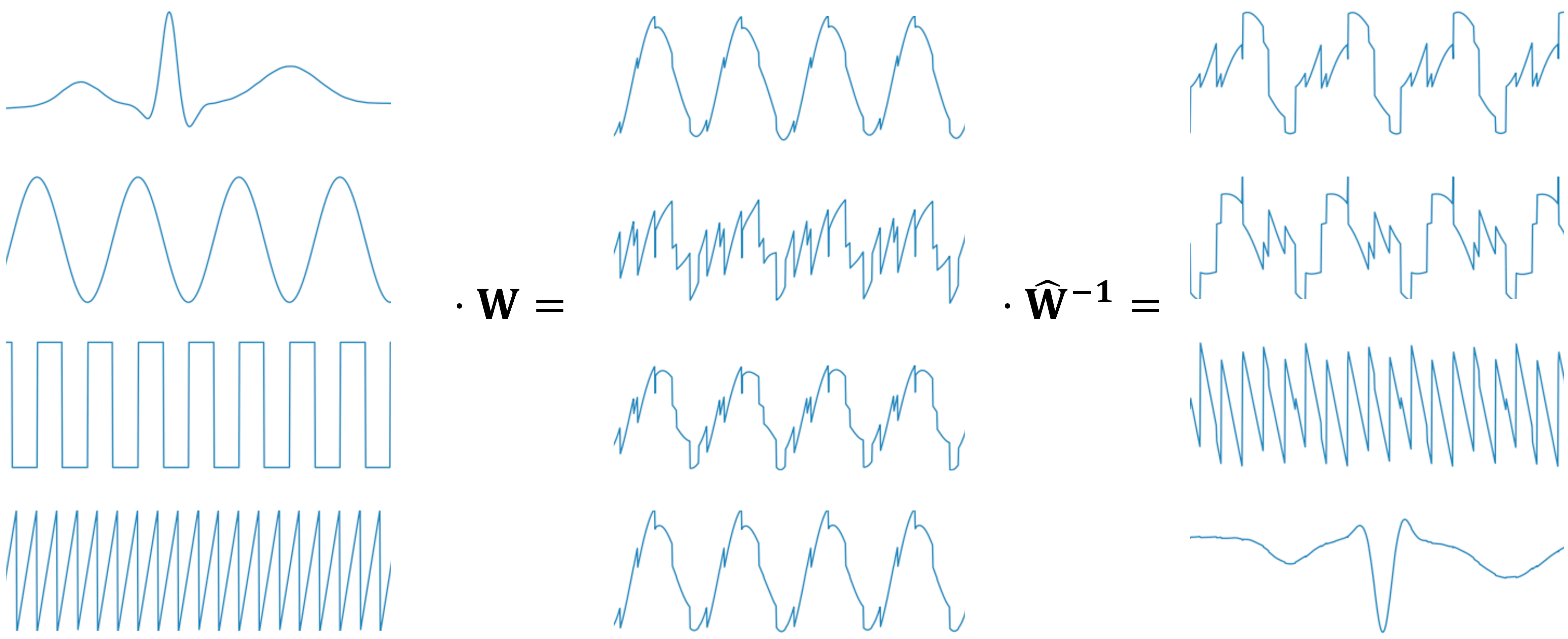
Results
Interpreting the Minimum Distance



- $\mathrm{MD} = 0.2$: Good results, small spikes, some sign changes
- $\mathrm{MD} = 0.4$: signals are recognizable, bigger spikes

Results

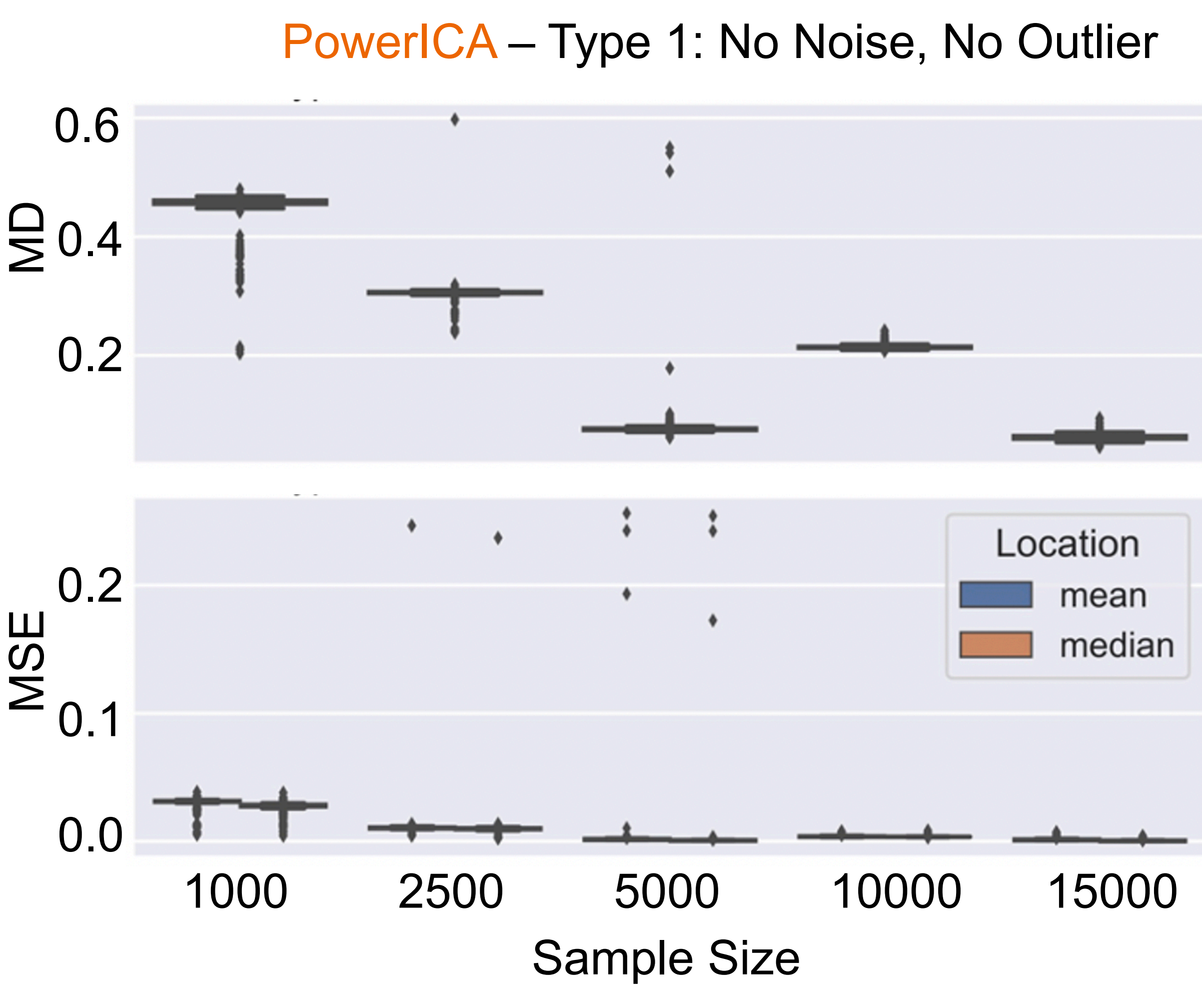
Interpreting the Minimum Distance



$\mathrm{MD} = 0.6$

- $\mathrm{MD} = 0.2$: Good results, small spikes, some sign changes
- $\mathrm{MD} = 0.4$: signals are recognizable, bigger spikes
- $\mathrm{MD} = 0.6$: not all signal reconstructed, bigger spikes

Experiment results
Type 1



Sample Size 10000

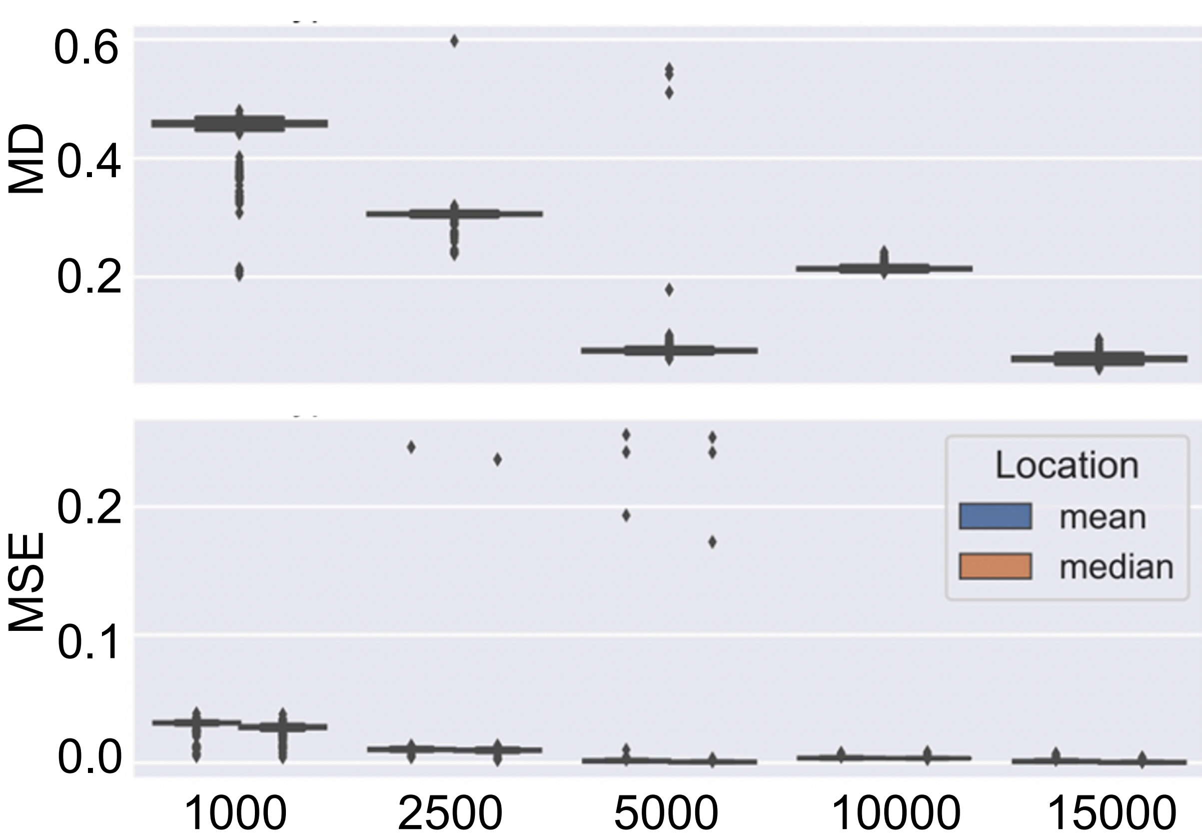
Type 1 (ideal)	Median MD	Median MSE
PowerICA	0.21 ± 0.001	0.0039 ± 0.0002
Jade	0.21 ± 0.03	0.0077 ± 0.002
CoroICA	0.06 ± 0	0.0009 ± 0
Radical	0.03 ± 0.006	0.0001 ± 4.86

Pros:
- All Algorithms are able to reconstruct the source signals

Cons:
- Radical: Long runtime → unusable for real time processing
- CoroICA: Only good results for tested signals by Hyperparameter tuning

Experiment results
Type 1 - 4

PowerICA – Type 1: No Noise, No Outlier



Sample Size 10000

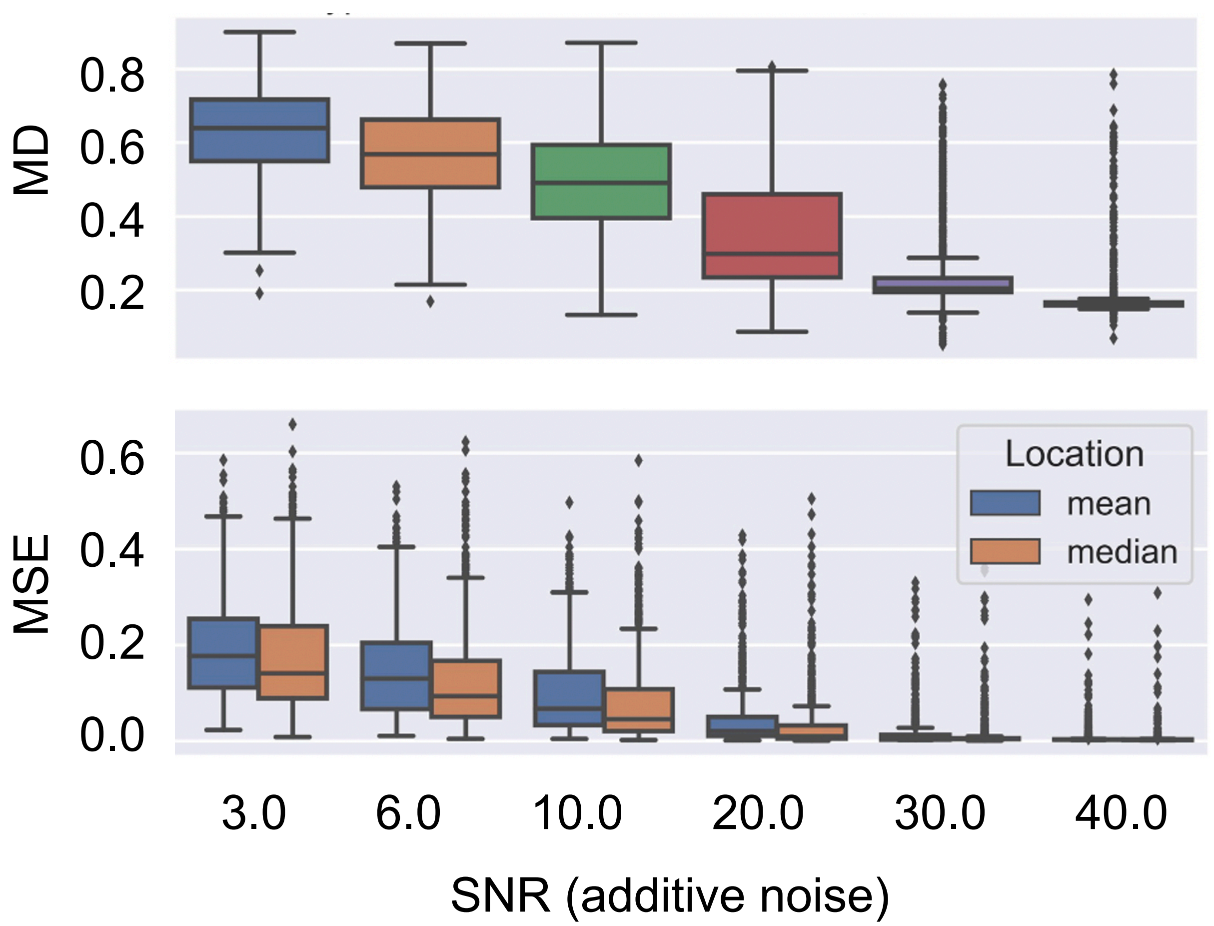
PowerICA	Median MD	Median MSE
Type 1 (ideal)	0.21 ± 0.001	0.0039 ± 0.0002
Type 2 (noise)	0.18 ± 0.006	0.0034 ± 0.0004
Type 3 (outlier)	0.81 ± 0.06	0.36 ± 0.1
Type 4 (noise + outlier)	0.81 ± 0.06	0.35 ± 0.1

- Sample Size of min. n = 5000 is necessary
- For 40dB SNR → variance increases slightly, more outliers
- Even 0.1% Outlier causes failure for every algorithm
- Hop at 10000: Drawback of the MD – metric is not perfect

Next Step:
Breakdown Analyses – How much noise does it take to cause a breakdown?

Experiment results
Type 5

Power ICA – Type 5: 3-40dB Noise, No Outlier



Type 5:
- ▶ Noise < 20dB SNR → breakdown

Did our robustness plugins improve the results ?

Robust Covariance Estimation in Whitening:
- ▶ Could not prevent breakdown
- ▶ Occasionally causes breakdown

Robust objective Function in higher Order statistics:
- ▶ Could not prevent breakdown
- ▶ Occasionally causes breakdown

Intermediate conclusion:
Pre-processing of the data essential! → Outlier removal + smoothing
We need new approaches to improve the performance

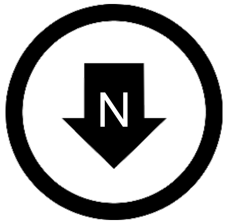
Agenda

TECHNISCHE UNIVERSITÄT DARMSTADT

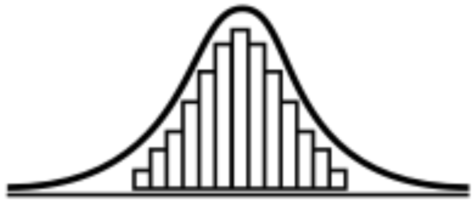
1. Understanding the basis of Source Separation via an intuitive example-driven approach
2. Discussion of experiment design
3. Graph Blind Source Separation
4. Graph Blind Source Separation results
5. Summary of major outcomes
6. Problems and future research

Shortcomings

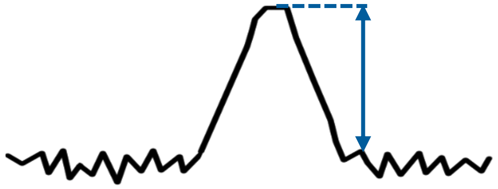
1. Significant performance loss for sample sizes with **N << 5000**
2. Unable to recover more than one **Gaussian signal**
3. For **SNR ≤ 20dB** high variance in performance
   ► Failures become increasingly likely
4. Inherently unrobust against outliers
   ► A **single outlier** will lead to failure of the method



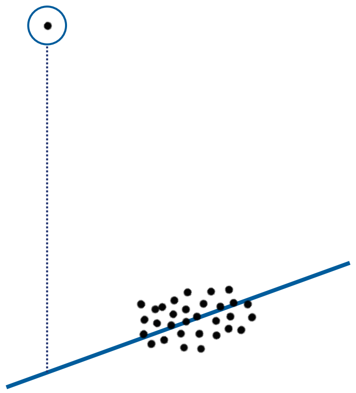
1. Samples



2. Gaussian



3. SNR ≤ 20dB



4. Outlier

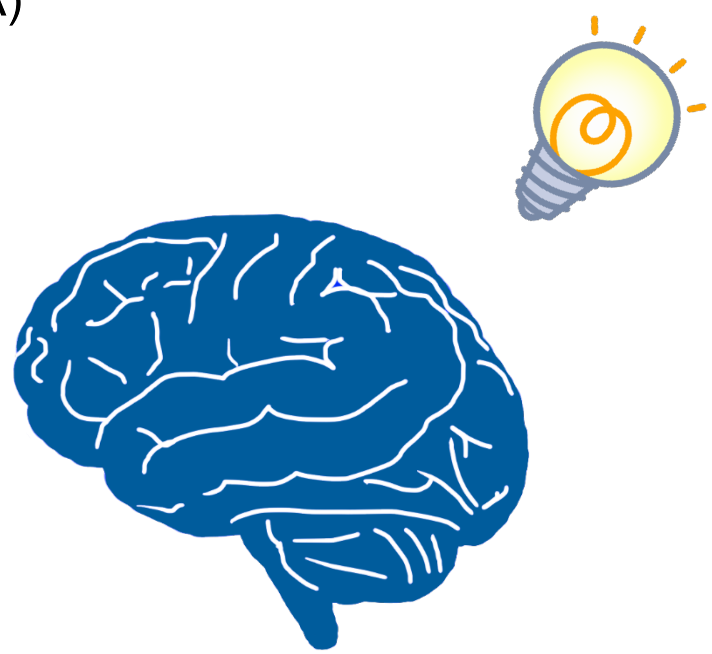
Recent Developments

- In 2010: Blind Source Separation(BSS) using graph theory [Blöchl, 2010]
- In 2020: First approaches to combine non-Gaussianity based BSS with graph based BSS (Jari Miettinen et al.) [Miettinen, 2020]

What is the idea?

- Design a method that uses:

a) Non-Gaussianity measure (classic ICA)

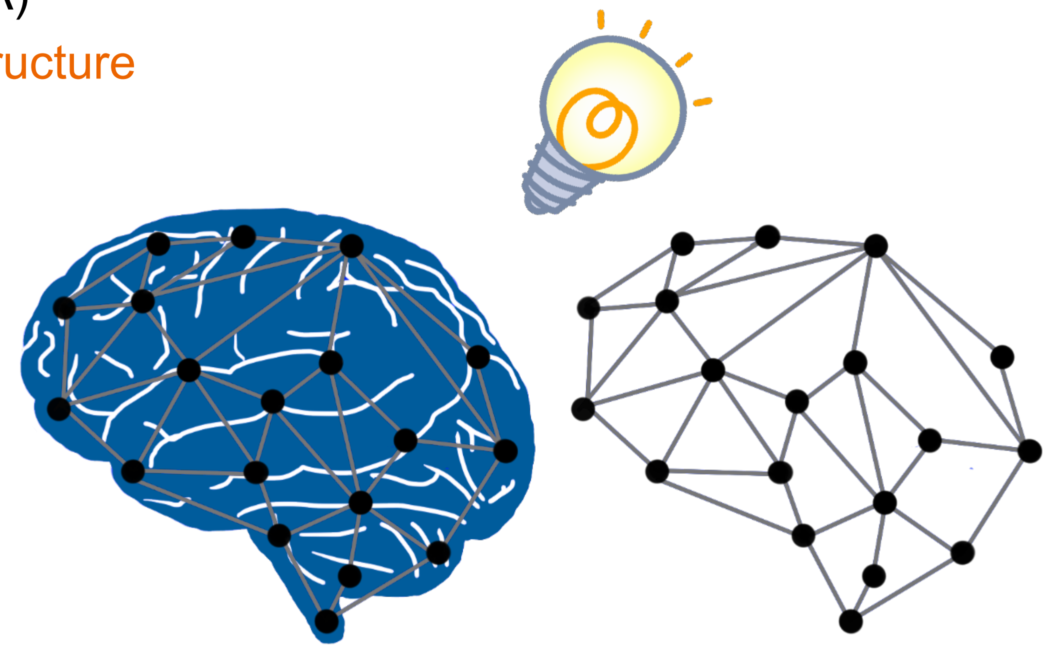


Recent Developments

- In 2010: Blind Source Separation(BSS) using graph theory [Blöchl, 2010]
- In 2020: First approaches to combine non-Gaussianity based BSS with graph based BSS (Jari Miettinen et al.) [Miettinen, 2020]

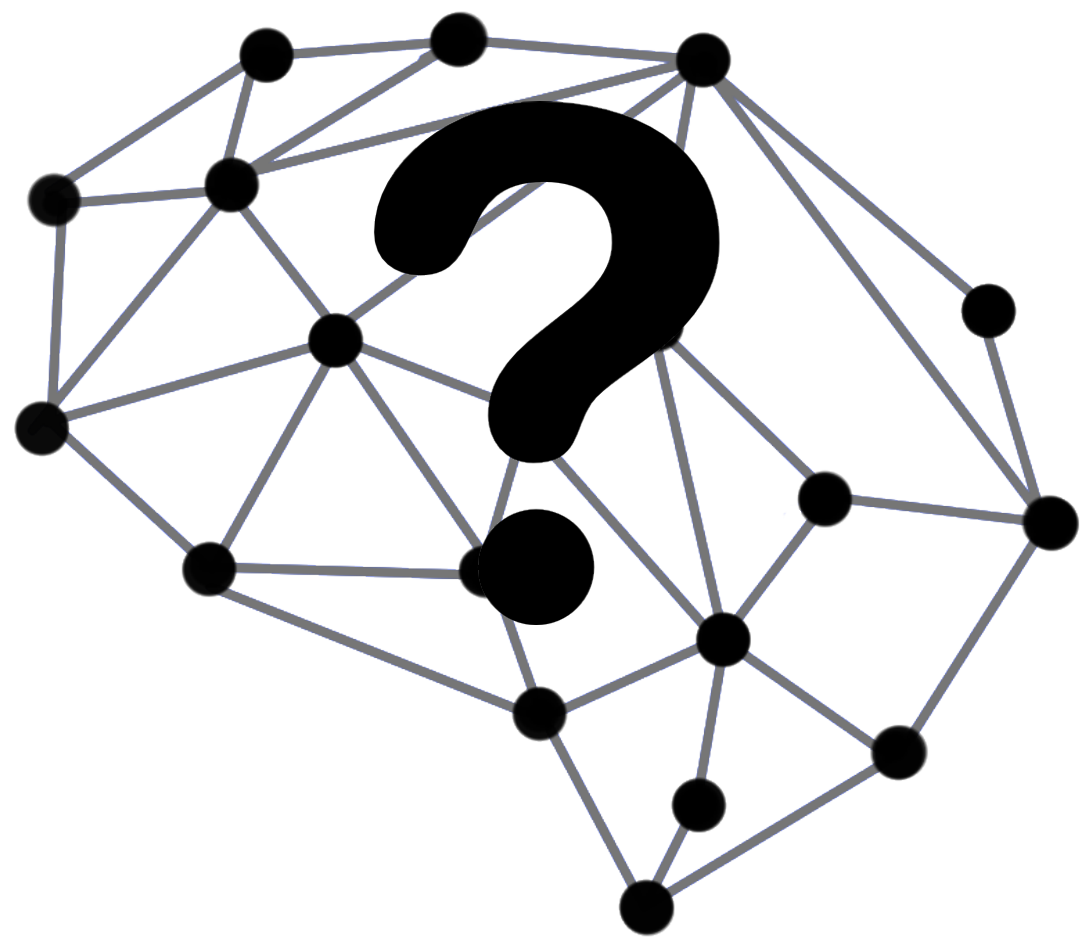
What is the idea?

- Design a method that uses:
  a) Non-Gaussianity measure (classic ICA)
  b) Information about underlying graph structure



- Combine those in a way that they complement each other

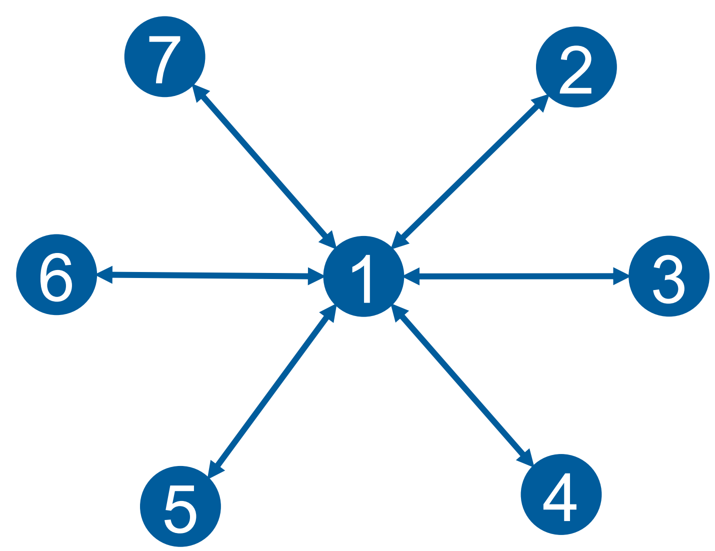
Graph Blind Source Separation
What are Graphs?



Graph Blind Source Separation
What are Graphs?

- Graphs are structures build by vertices that are connected by edges
- Edges represent proximity between two vertices
- Graph structure can be stored in a matrix ⇒ The adjacency matrix **A** [Djuric, 2018]

$$\mathbf{A} = \begin{pmatrix} 0 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$



Undirected unweighted graph

- Examples: Social networks, sensor networks, brain networks, etc.

Graph Blind Source Separation

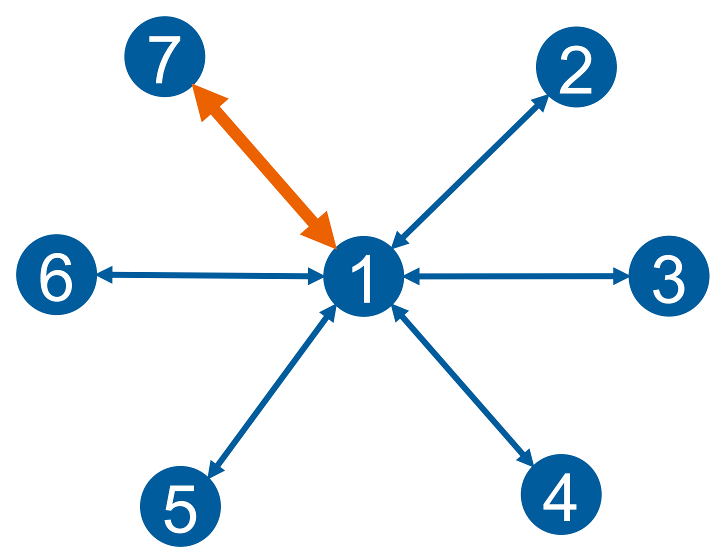
What are Graphs?

- Graphs are structures build by vertices that are connected by edges
- The Entry $\mathbf{A}_{i,j}$ represents the edge status between node i and j

$\mathbf{A}_{i,j}\Big|_{i=1,j=7}$

$$\mathbf{A} = \begin{pmatrix} 0 & 1 & 1 & 1 & 1 & 1 & \boxed{1} \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ \boxed{1} & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

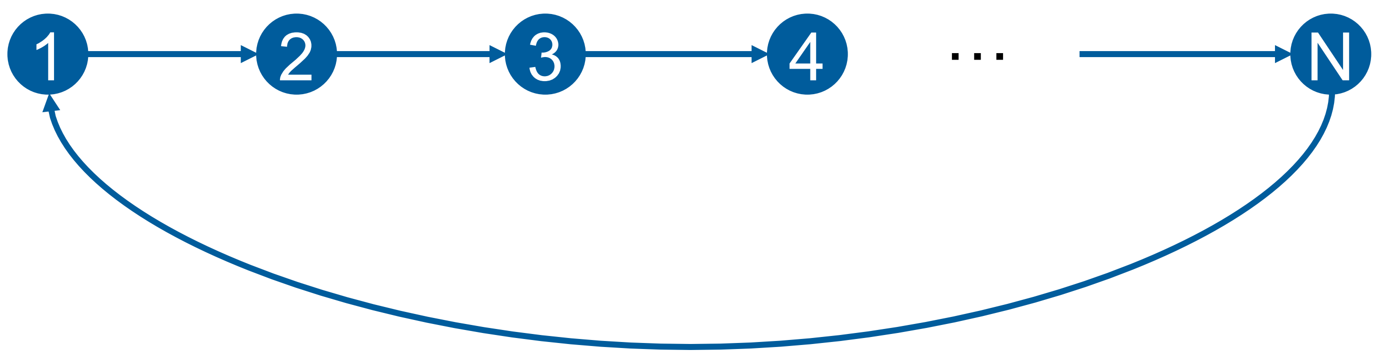
$\mathbf{A}_{i,j}\Big|_{i=7,j=1}$

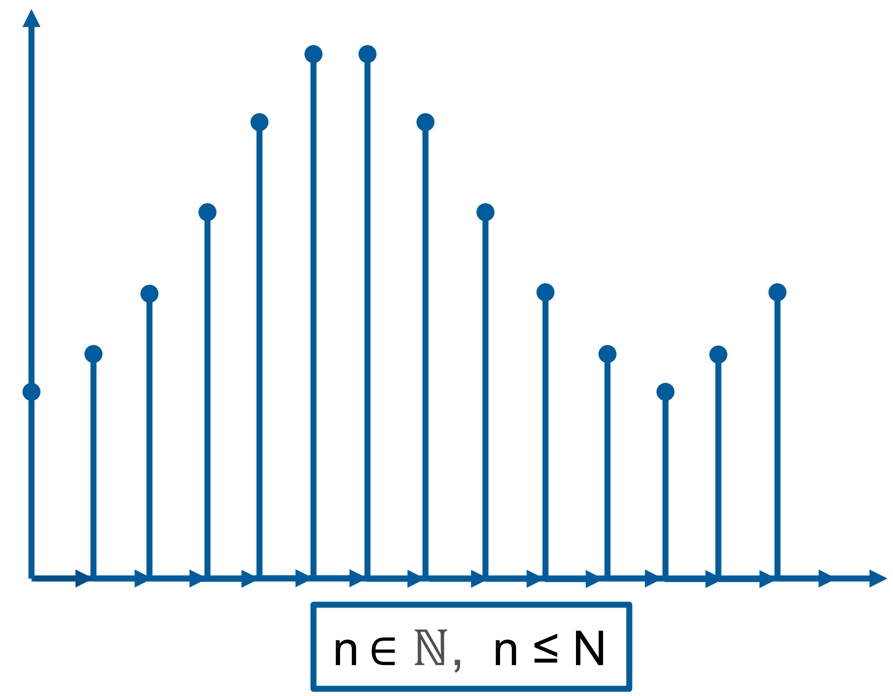


Undirected unweighted graph

Graph Blind Source Separation
Graph Signals

- Analogy to discrete time signals
- Simplest case: Recursive Line Graph



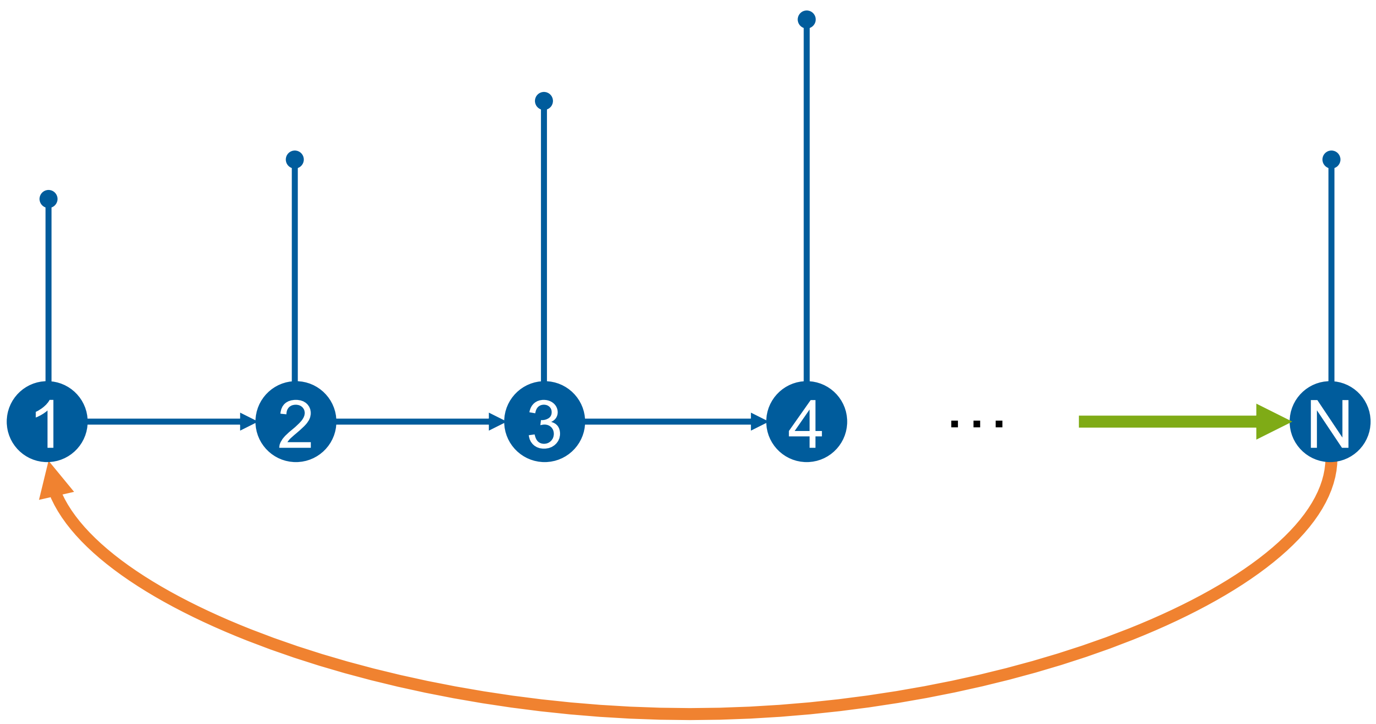


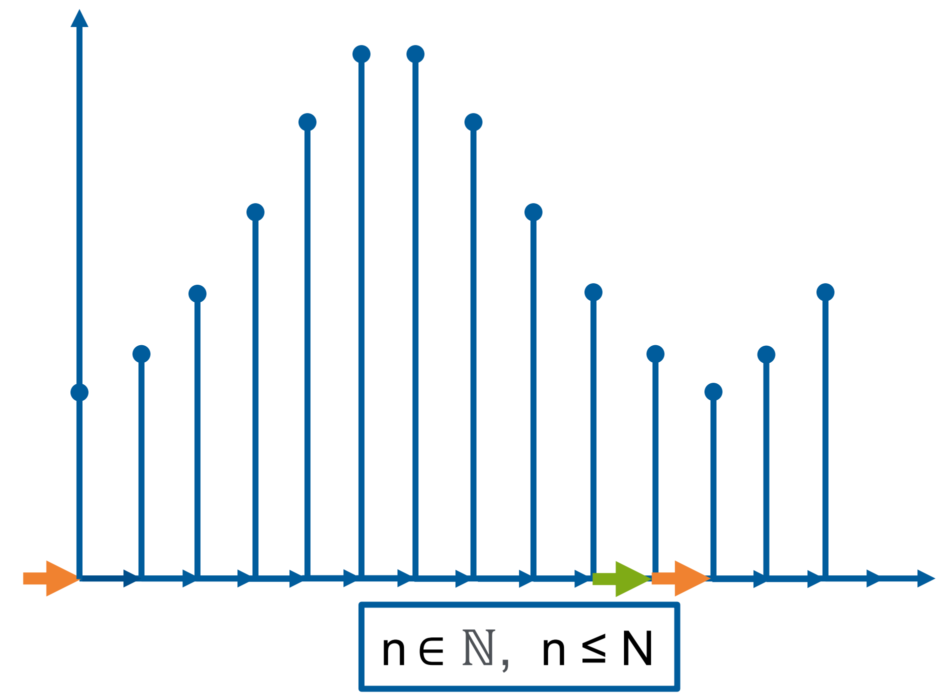
- Each node represents a sample point
- Directed graph ⇒ unsymmetric adjacency [Djuric, 2018]

$$\mathbf{A} = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & \dots & 0 \\ . & . & . & . & . & . & \dots & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \dots & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & \dots & 0 \end{pmatrix}$$

Graph Blind Source Separation
What are Graph Signals?

- Analogy to discrete time signals
- Simplest case: Recursive Line Graph



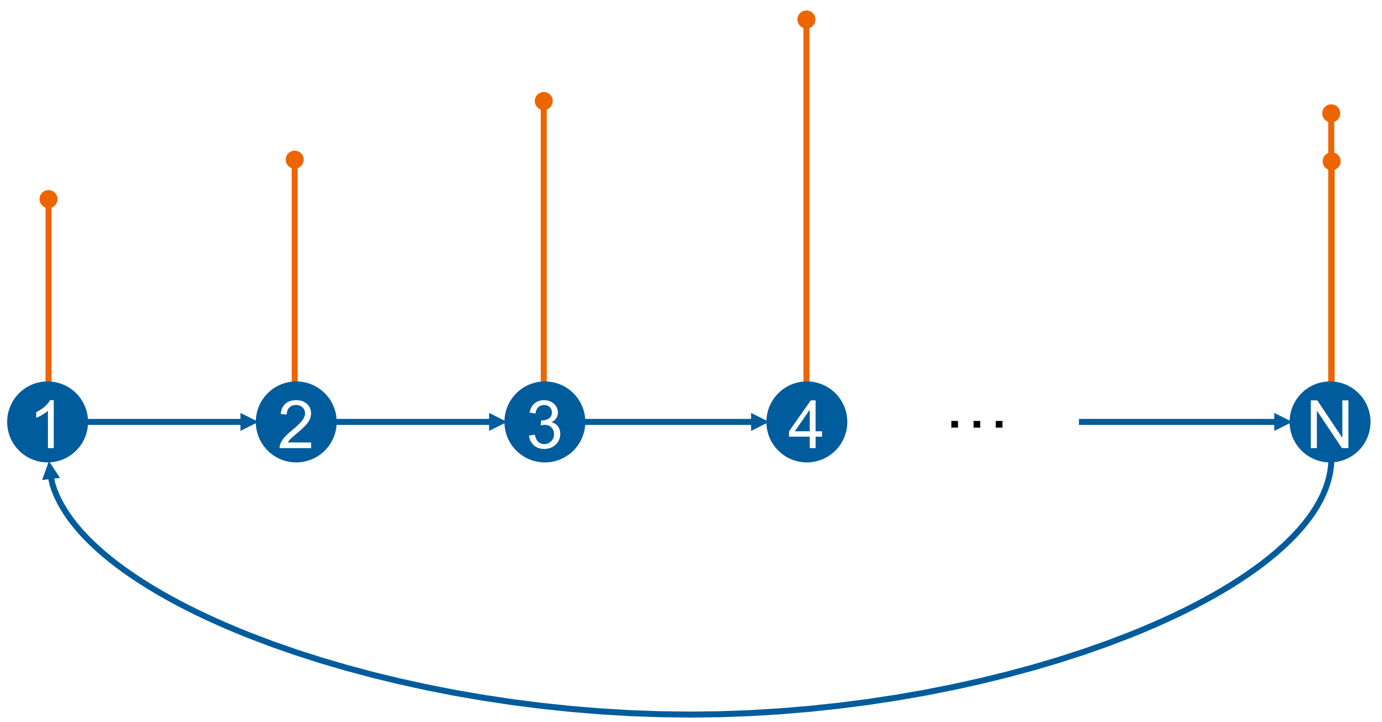


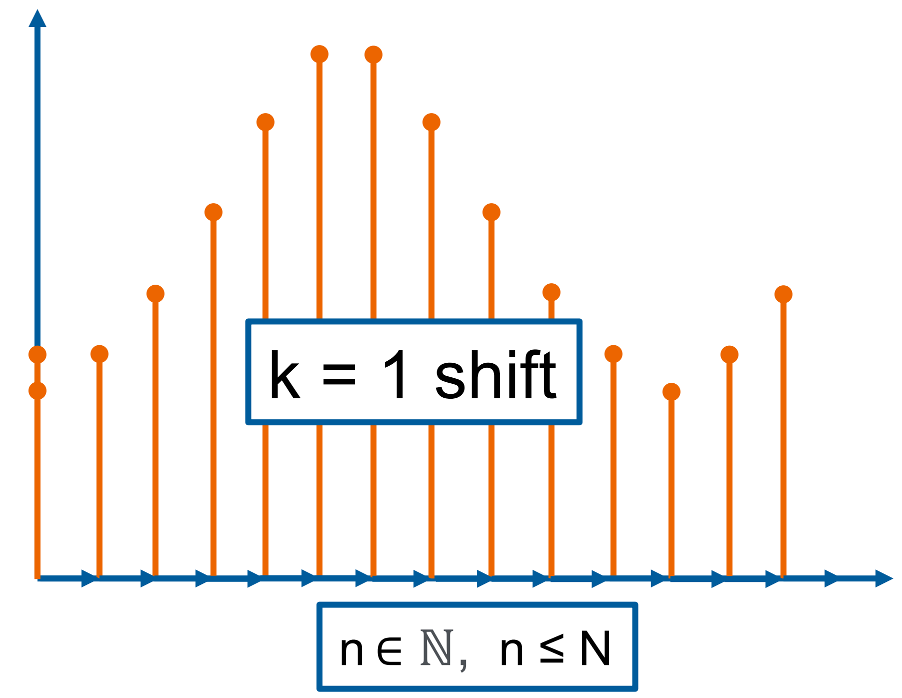
$$\mathbf{A} = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & \ldots & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & \ldots & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & \ldots & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & \ldots & 0 \\ . & . & . & . & . & . & \ldots & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \ldots & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & \ldots & 0 \end{pmatrix}$$

- Multiple Signals can share the same graph
- In this case: The class of all periodic signals with period N

Graph Blind Source Separation
What are Graph Signals?

- Analogy to discrete time signals
- Simplest case: Recursive Line Graph



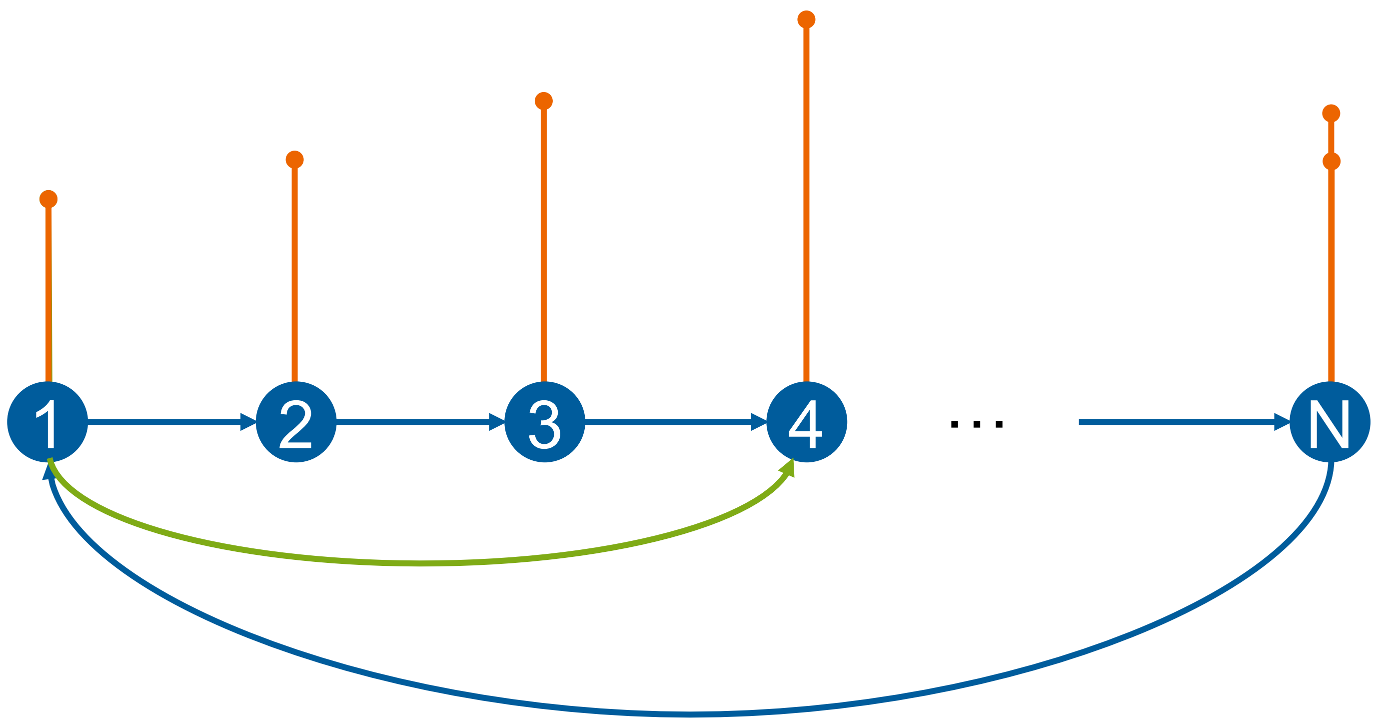


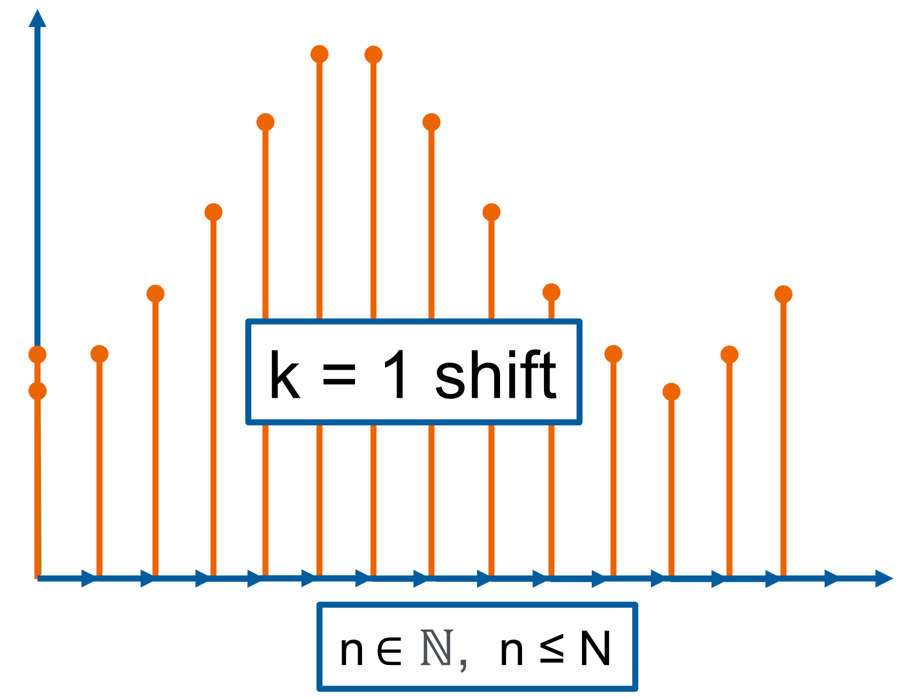
- We call $\mathbf{X}\mathbf{A}^{\mathrm{k}} = \mathbf{X}^{(\mathrm{k})}$ a graph shift by k
  - For Line Graph ⇔ Shift operator as known from Digital Signal Processing
- When a graph shift is applied, the signal will shift along every **connected path** made up of the **k nearest edges**. [Djuric, 2018]

Graph Blind Source Separation

What are Graph Signals?

- Analogy to discrete time signals
- Simplest case: Recursive Line Graph





- We call $\mathbf{X}\mathbf{A}^{\mathrm{k}} = \mathbf{X}^{(\mathrm{k})}$ a graph shift by k
  - For Line Graph ⇔ Shift operator as known from Digital Signal Processing
- When a graph shift is applied, the signal will shift along every **connected path** made up of the **k nearest edges**. [Djuric, 2018]

Graph Blind Source Separation

Graph Autocorrelation

Using the introduced graph shift we define the graph autocorrelation matrix analog to the sample autocorrelation [Blöchl, 2010] **:**

$$\mathbf{C}_X^{\text{Sample}} := \frac{1}{N-1}\mathbf{X}\mathbf{X}^T = \frac{1}{N-1}\mathbf{X}\mathbf{I}\mathbf{X}^T$$

$$\mathbf{C}_X^{\text{Graph}}(k) := \frac{1}{N-1}\mathbf{X}^{(k)}\mathbf{X}^T = \frac{1}{N-1}\mathbf{X}\mathbf{A}^k\mathbf{X}^T$$

$\mathbf{A}^k|_{k=0} = \mathbf{I}$,

$\mathbf{I} \in \mathbb{R}^{N\times N}$ **: Identity Matrix**

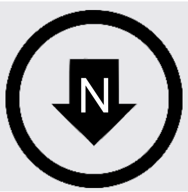
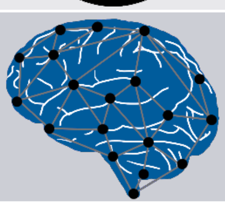
- $\mathbf{C}_X^{\text{Sample}} \in \mathbb{R}^{D\times D}$: D dimensional square matrix of sample autocorrelation
- $\mathbf{C}_X^{\text{Graph}}(k) \in \mathbb{R}^{D\times D}$: D dimensional square matrix of graph autocorrelation for k graph-shifts
- $\mathbf{A}^k \in \mathbb{R}^{N\times N}$: k-th power of adjacency matrix corresponding to k shifts along the graph
- $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots \mathbf{x}_D) \in \mathbb{R}^{N\times D}$: D signals stored as column vectors

- The **k-shift** along paths of an underlying graph can provide **additional information** about the source signals!

Graph Blind Source Seperation
Graph Decorrelation

TECHNISCHE UNIVERSITÄT DARMSTADT

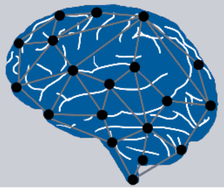
- Find a matrix $\mathbf{W}$ that **diagonalizes** $\mathbf{W}\mathbf{C}_{\mathbf{X}}^{\mathbf{Graph}}\mathbf{W}^{\mathbf{T}}$ as much as possible [Blöchl, 2010]
- The solution for $\mathbf{W}$ decorrelates the signals $\boldsymbol{X}$ in the graph-sense
  - $\mathbf{W}$ is an estimate for the solution of the BSS problem for graph-signals

Advantages		Disadvantages
Independent of data distribution		Graph structure not directly available
Works well with small sample sizes		Highly dependent on accurate graph
Graph structure in many real-world problems		

Graph Blind Source Seperation
Graph Decorrelation

PowerICA shortcomings:

- Significant performance loss for sample sizes with **N << 5000**
- Unable to recover more than one **Gaussian signal**

Advantages	Disadvantages
Independent of data distribution 	Graph structure not directly available
Works well with small sample sizes 	Highly dependent on accurate graph
Graph structure in many real-world problems 	

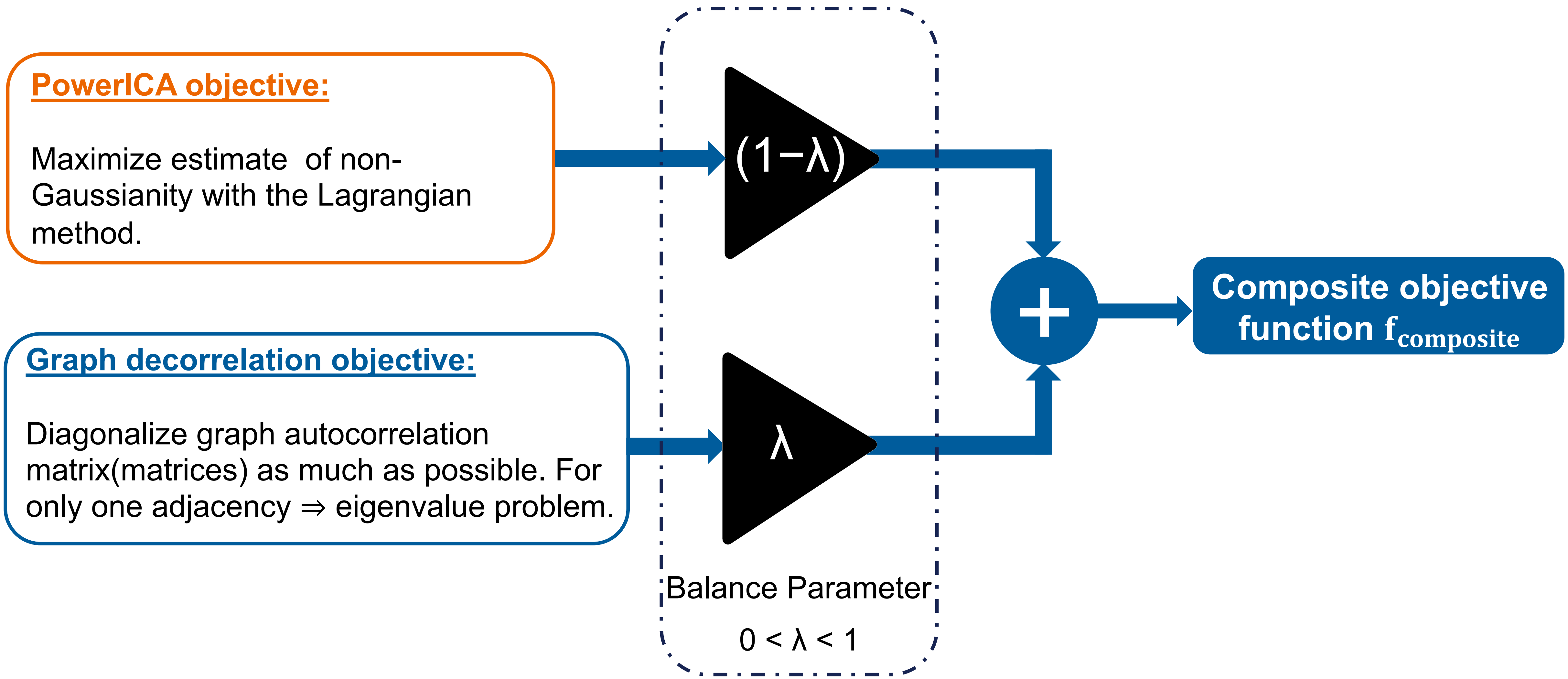
Graph based BSS and non-Gaussianity based BSS seem to make up for each others weak points.

Graph Blind Source Seperation
Composite Objective

TECHNISCHE UNIVERSITÄT DARMSTADT

<u>How can we combine them?</u>

- Approach from Jari Miettinen et al. (Aalto University) [Miettinen, 2020]
- **Define composite objective function $\mathbf{f}_{\text{composite}}$:**



Literature

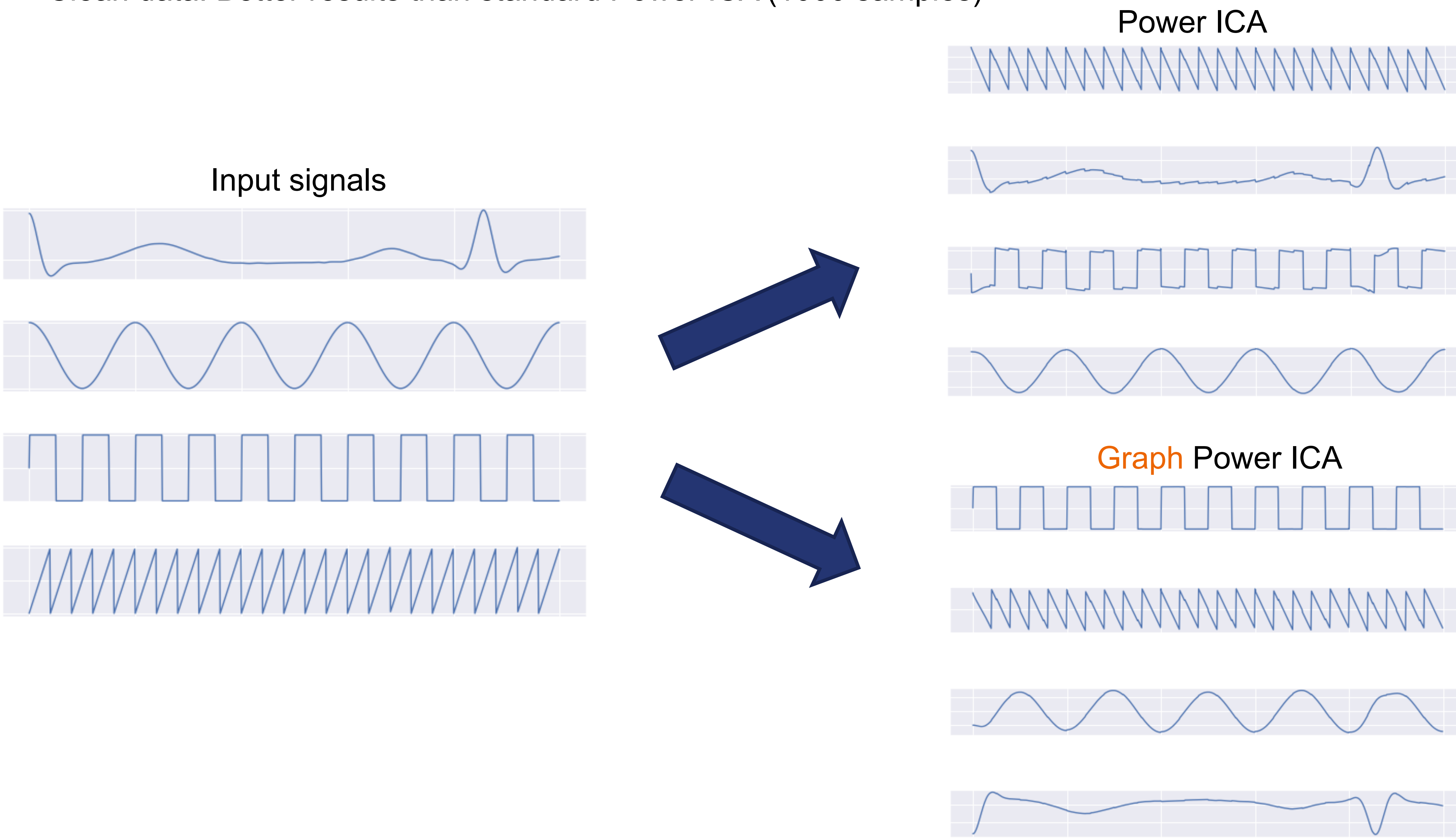
- [Miettinen, 2020] Graph Signal Processing Meets Blind Source Separation, Jari Miettinen, Esa Ollila et al., 2020
- [Blöchl, 2010] Second-Order Source Separation Based on Prior Knowledge Realized in a Graph Model, Florian Blöchl, Andreas Kowarsch, Fabian J. Theis, 2010
- [Djuric, 2018] Cooperative and Graph Signal Processing Principles and Applications, Petar Djuric, Cédric Richard, 2018

Agenda

1. Understanding the basis of Source Separation via an intuitive example-driven approach
2. Discussion of experiment design
3. Graph Blind Source Separation
4. Graph Blind Source Separation results
5. Summary of major outcomes
6. Problems and future research

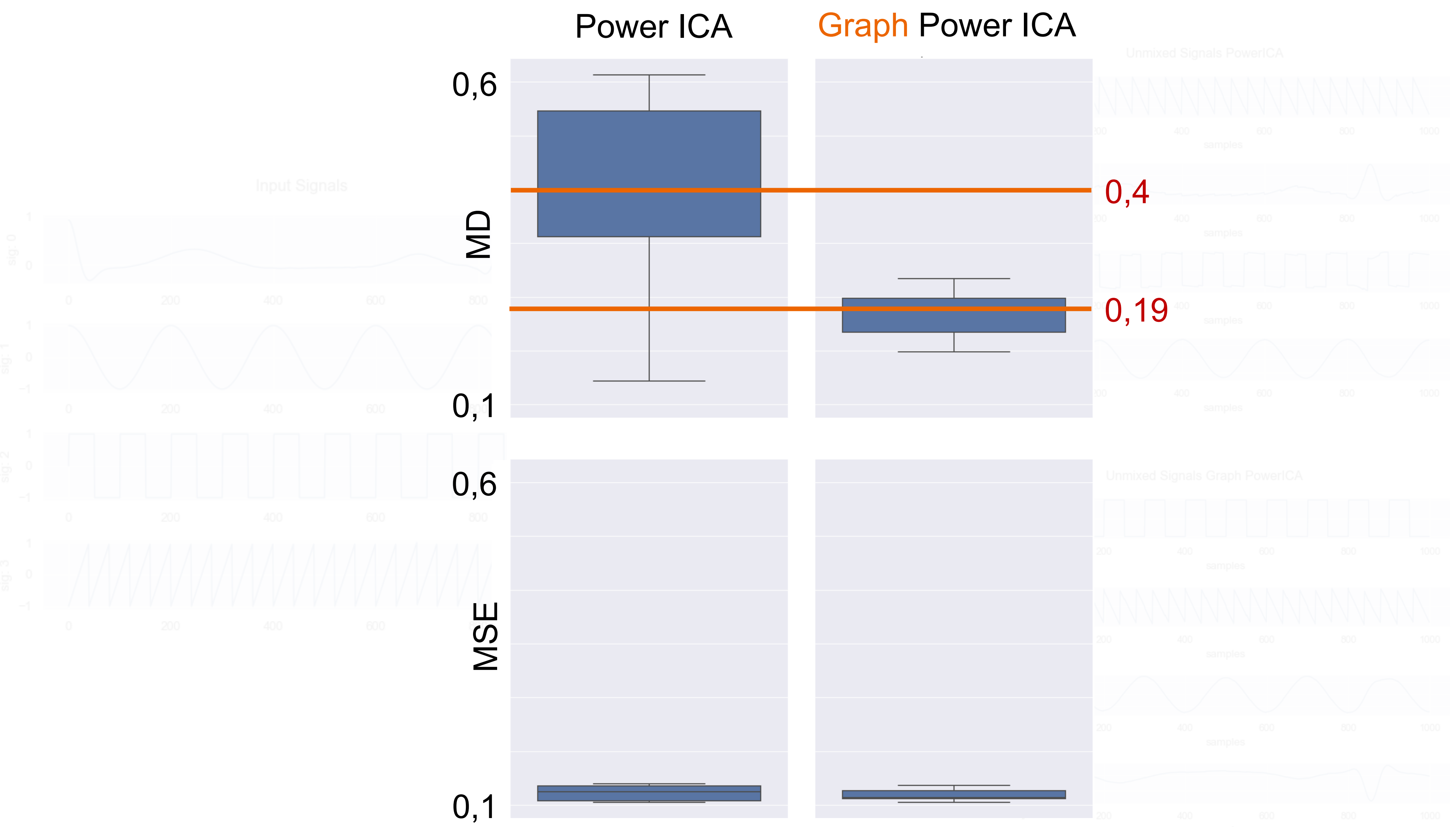
Graph BSS Results
Clean data

- Standard signals (Cos, Rect, ECG, Saw)
- Clean data: Better results than standard Power ICA (1000 samples)



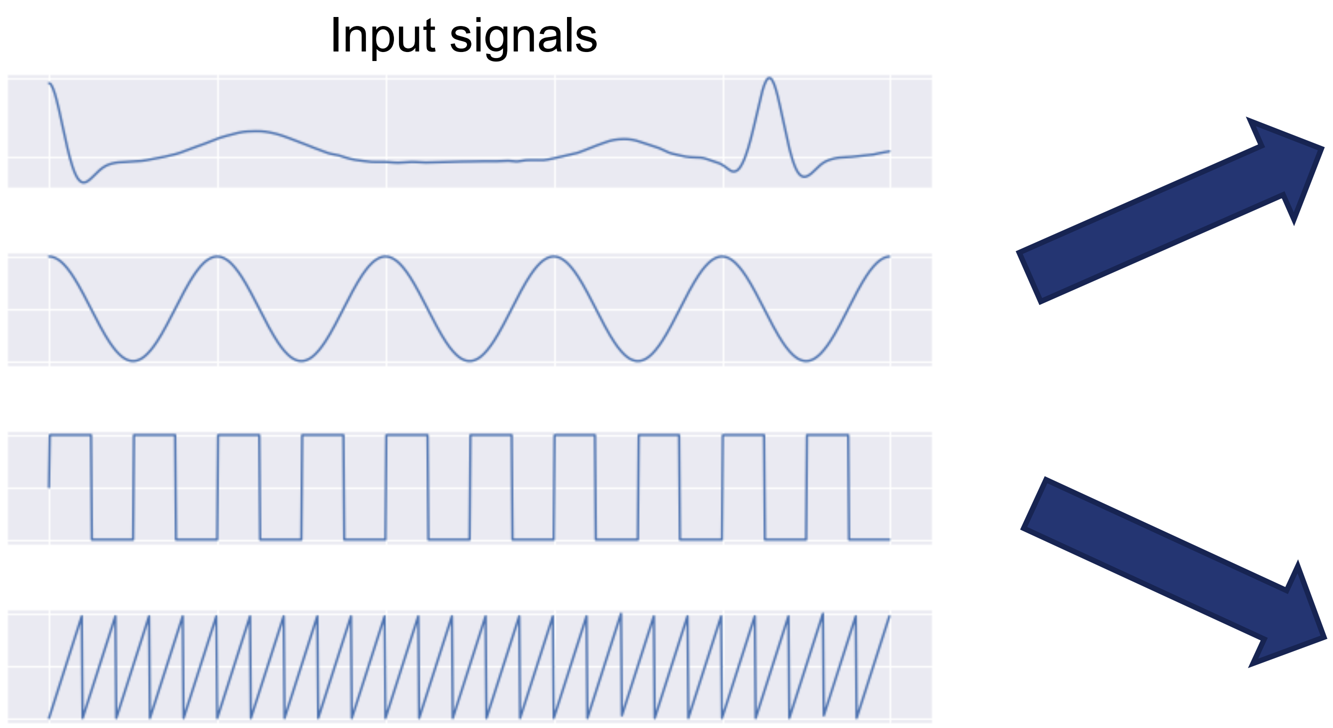
Graph BSS Results
Clean data

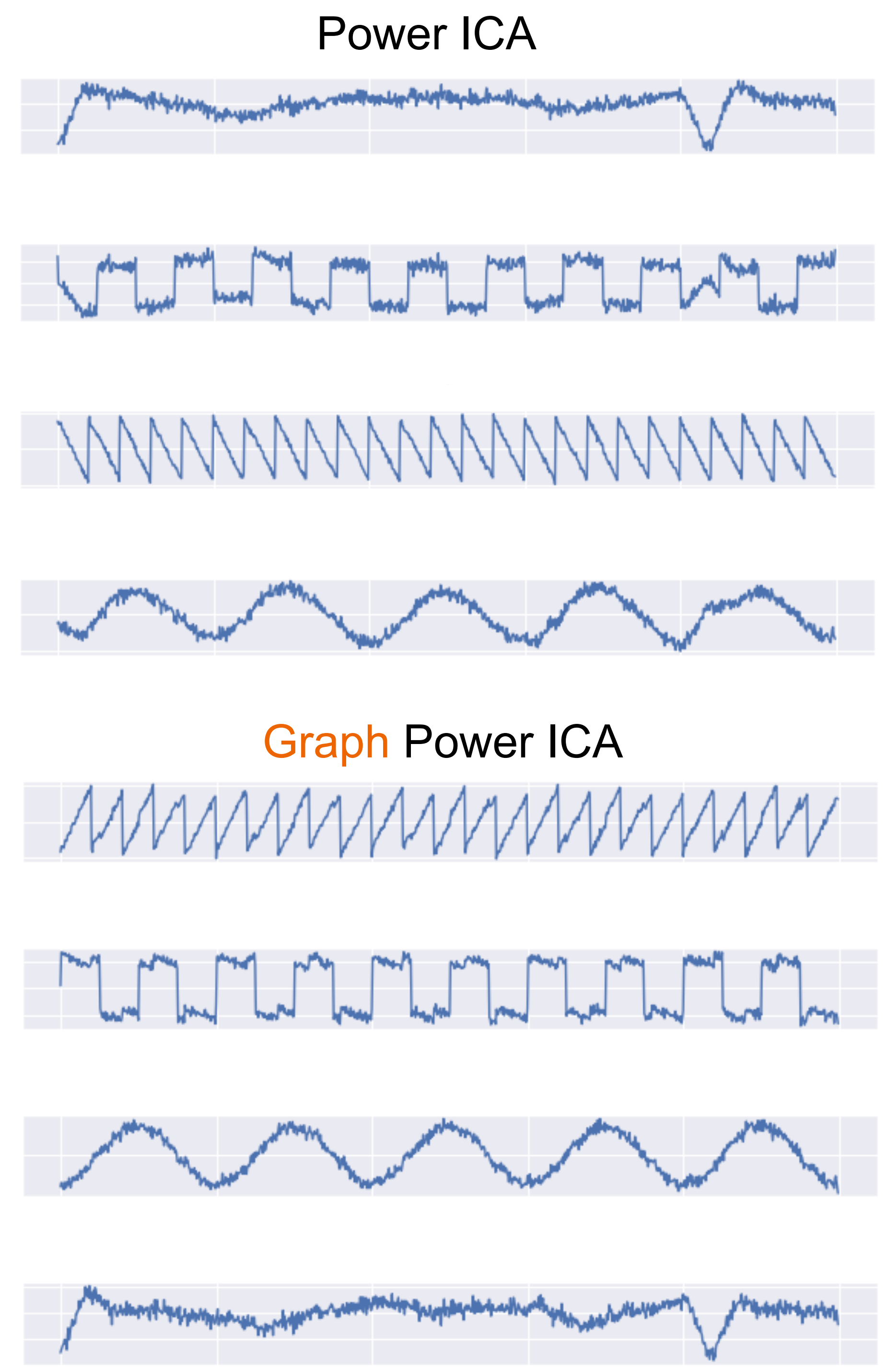
- Standard signals (Cos, Rect, ECG, Saw)
- Clean data: Better results than standard Power ICA (1000 samples)



Graph BSS Results
Noisy data

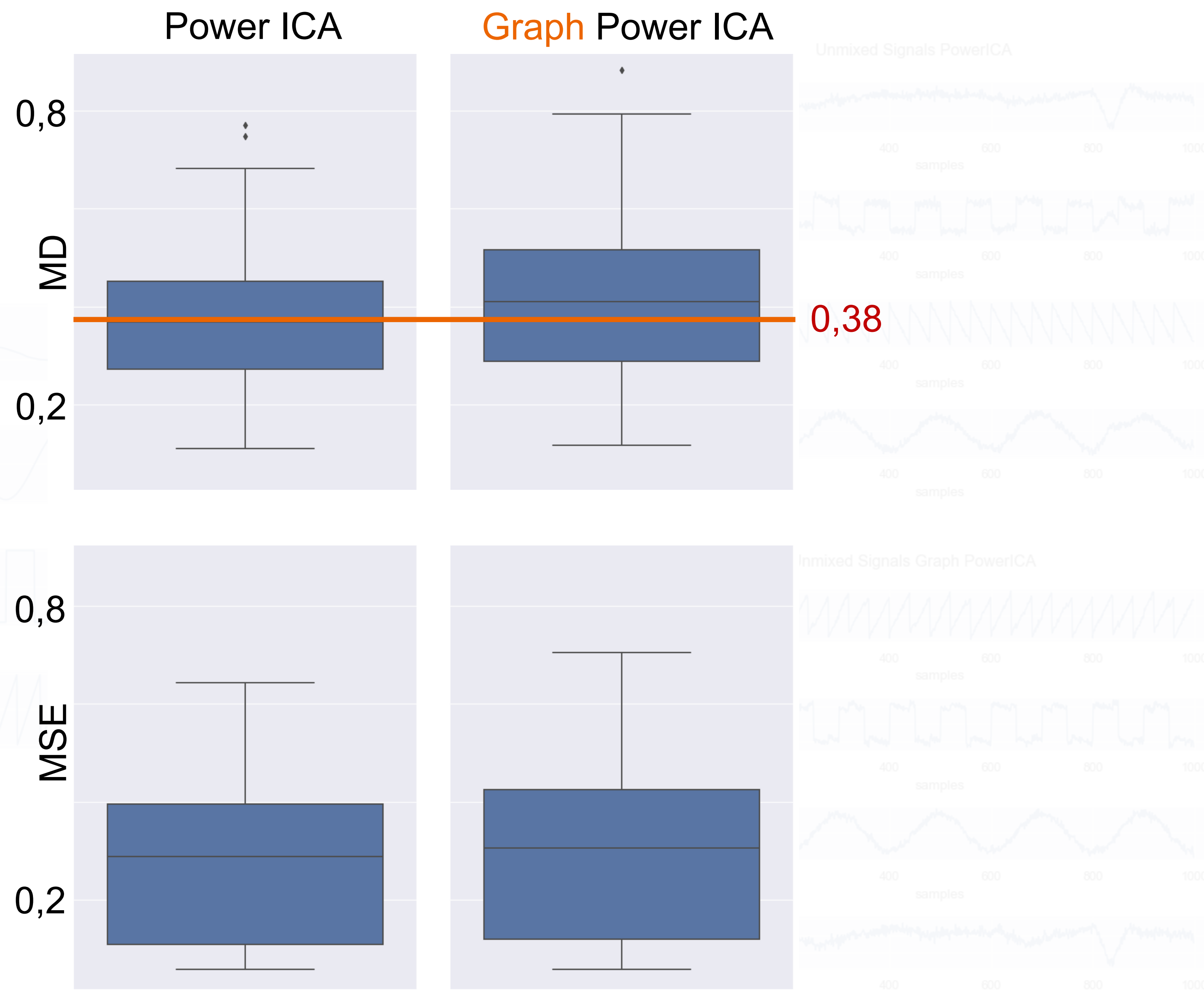
- Comparable for noisy signals (up to 20dB SNR)





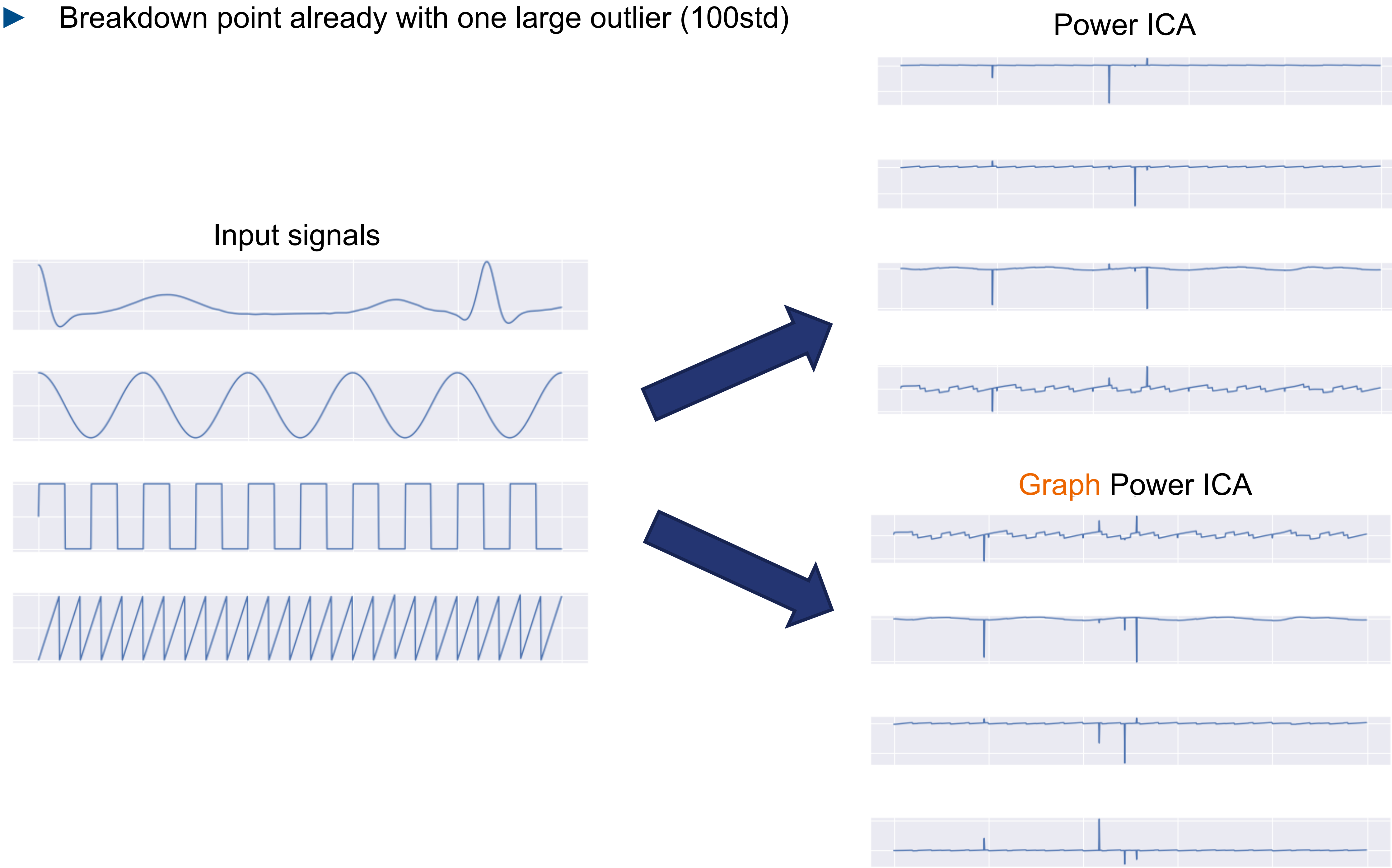
Graph BSS Results
Noisy data

- Comparable for noisy signals (up to 20dB SNR)



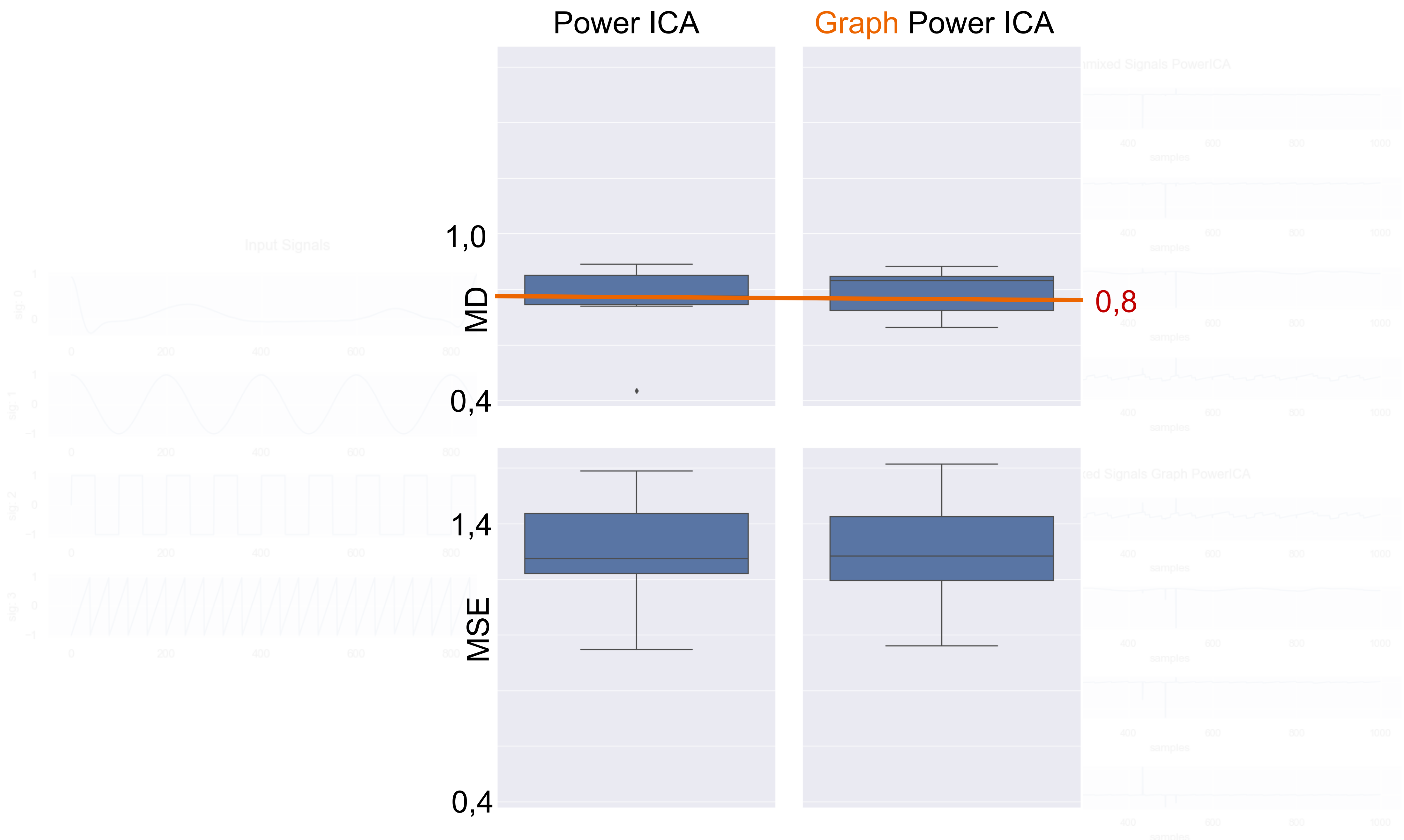
Graph BSS Results
Outlier contamination

- Breakdown point already with one large outlier (100std)



Graph BSS Results
Outlier contamination

- Breakdown point already with one large outlier (100std)

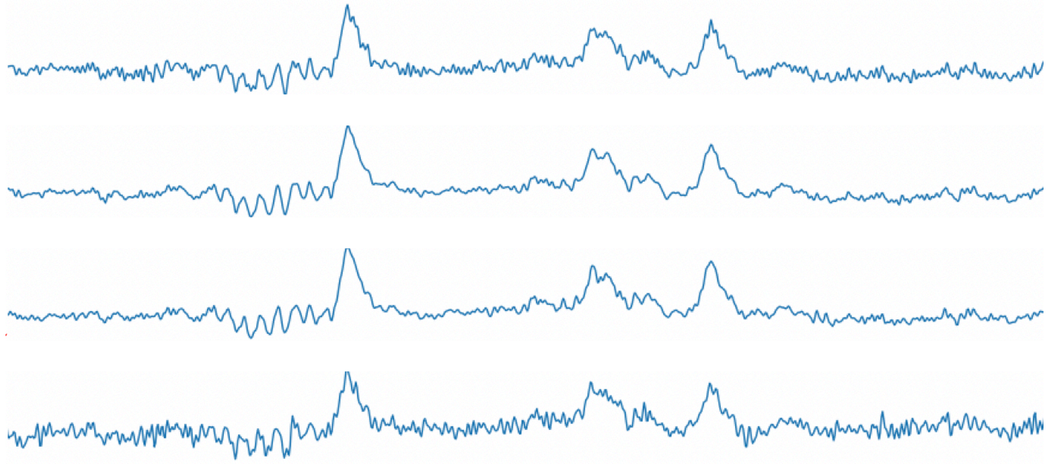


Towards EEG data
Synthetic EEG artifact reconstruction

Semi-synthetic EEG data:

1. Use real EEG data as input signals
2. Mix Data with random mixing matrix

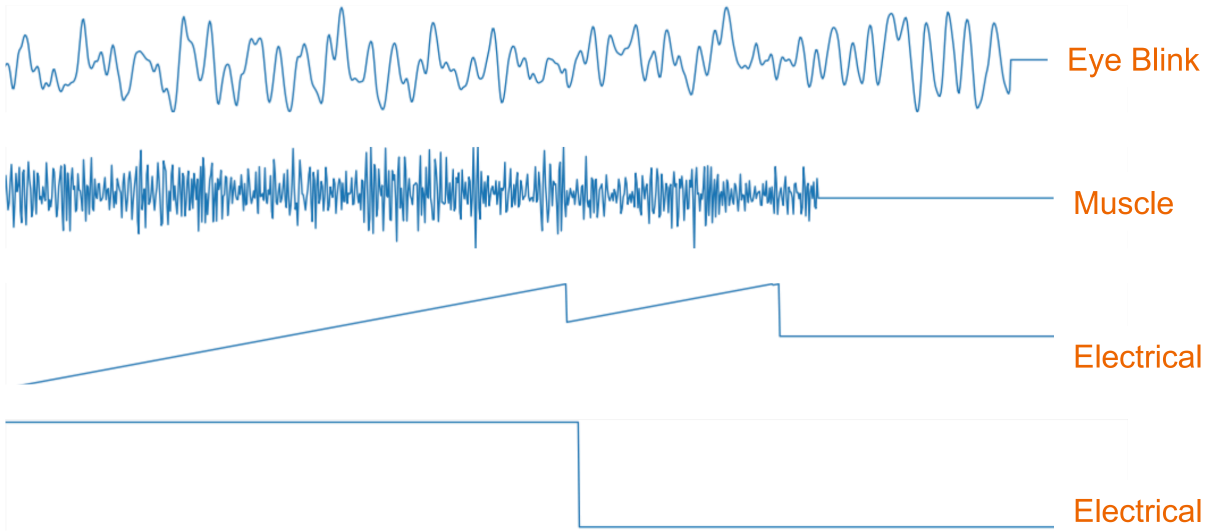
EEG data:



3. We generated realistic artifacts and added one on each channel
   - ► Artifacts which typically come along with EEG data [Delorme2006]
   - ► Optional: Add single or patchy (1000std) Outliers and white noise (3std)

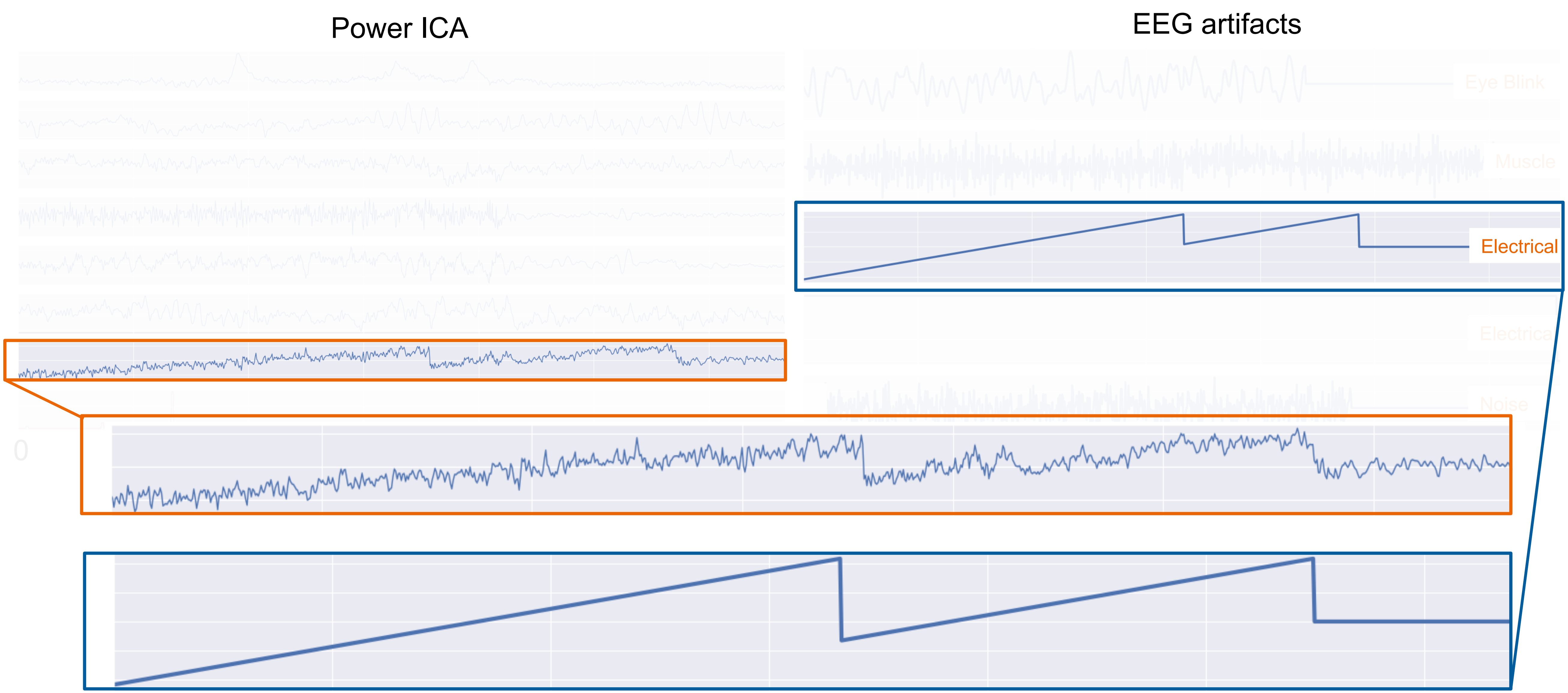
4. Perform Power ICA and Graph Power ICA

Synthetic EEG artifacts:



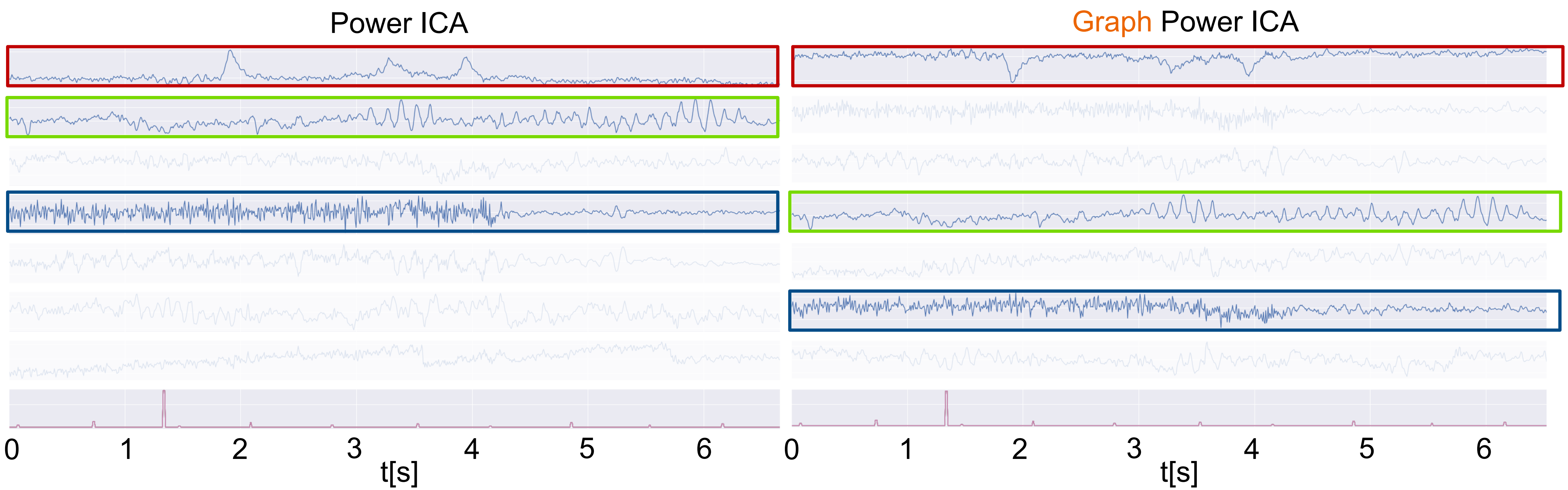
Graph BSS Results
Semi-synthetic EEG data

- Artifacts clearly extractable:



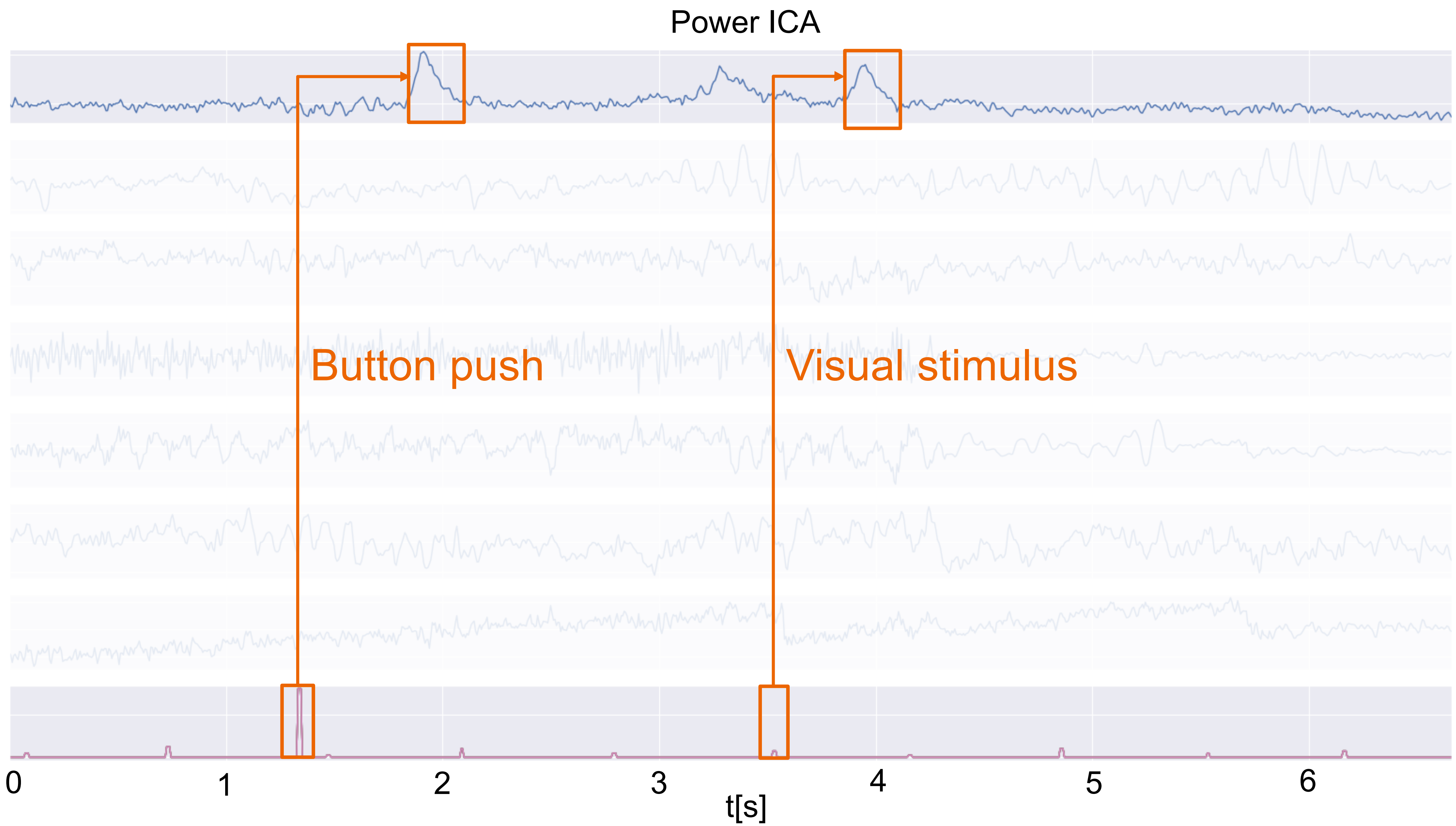
Graph BSS Results
Semi-synthetic EEG data

- Also possible with Graph Power ICA:



Graph BSS Results
Semi-synthetic EEG data

- Events clearly extractable:



Agenda

TECHNISCHE UNIVERSITÄT DARMSTADT

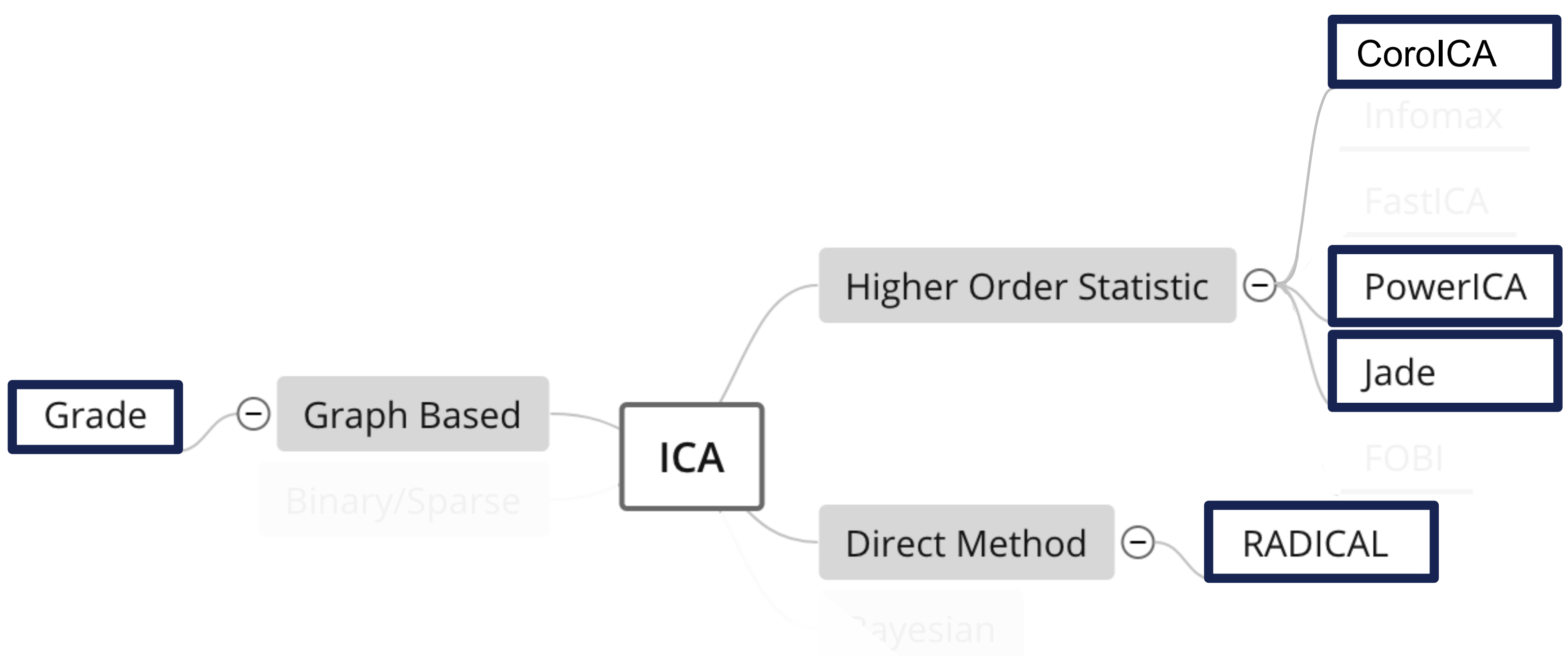
1. Understanding the basis of Source Separation via an intuitive example-driven approach
2. Discussion of experiment design
3. Graph Blind Source Separation
4. Graph Blind Source Separation results
5. Summary of major outcomes
6. Problems and future research

Summary of major outcomes

- Question 1: What is ICA?
- Question 2: How to compare algorithms?
- Question 3: Which algorithm performs best? What happens with noise or outlier?
- Question 4: How to robustify algorithms?
- Question 5: What is Graph Signal processing / Graph BSS?
- Question 6: Outcomes of Graph Signal processing?

Summary of major outcomes
Question 1: What is ICA?

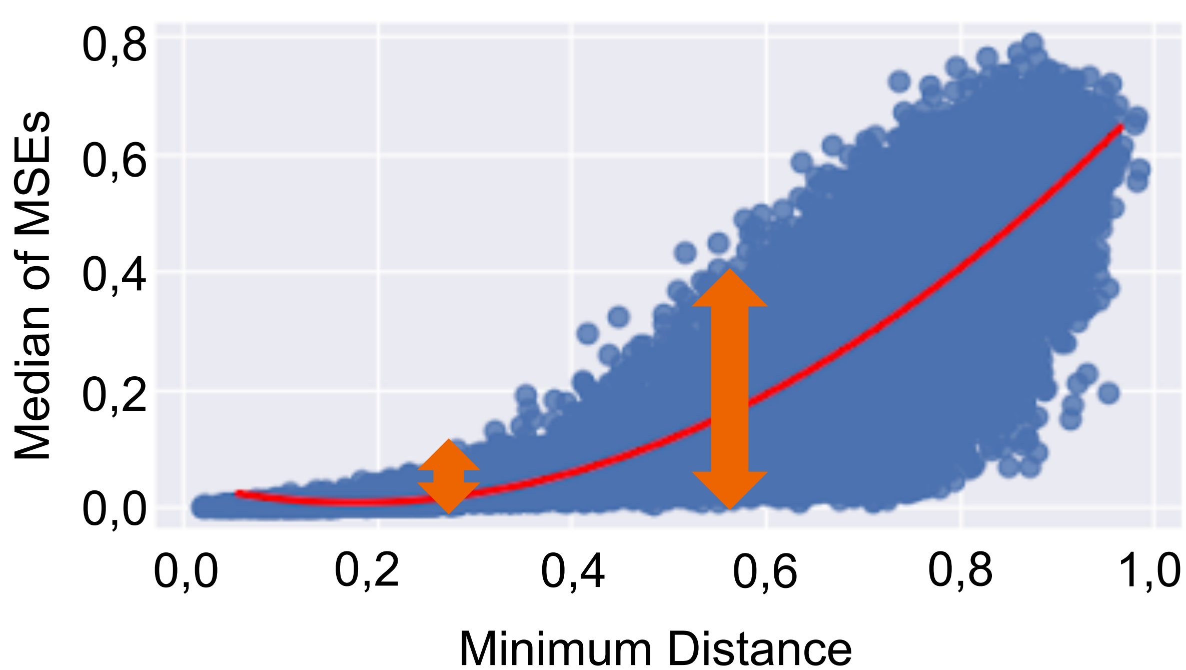
- Mathematical concept to extract (independent = non-gaussian) signals from mixed signals
- Different approaches to obtain the demixing matrix → 4 algorithms
- „ICA tries to decorrelate mixed up signals and find the optimal demixing matrix based on optimizing different measures of independence“

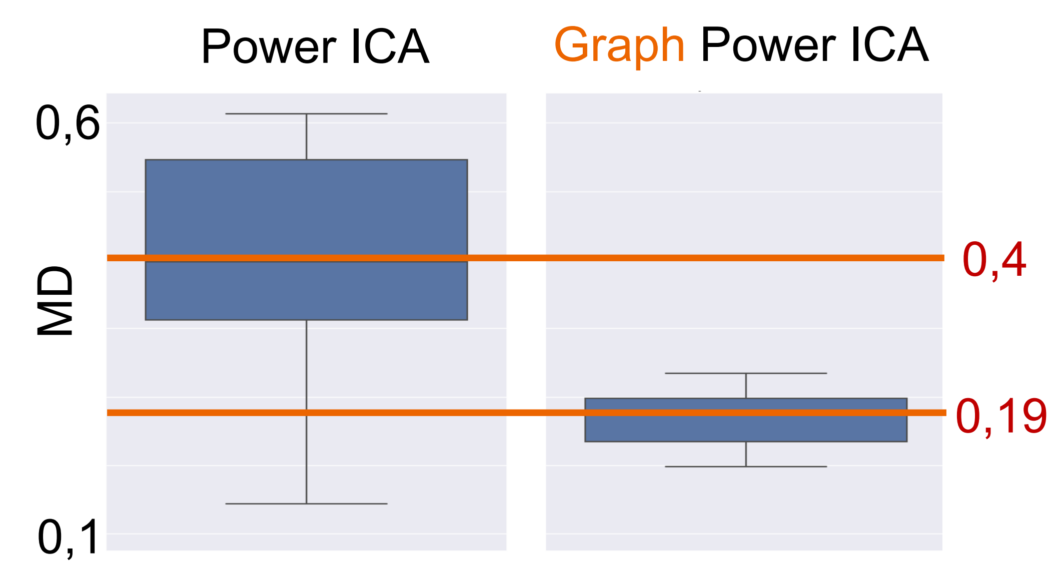


Summary of major outcomes

Question 2: How to compare algorithms?

- Minimum distance, MSE and SNR
- Correlation of different metrics important
- „One Metric sometimes provides not the whole information: We need to find how different metrics correlate“





Summary of major outcomes

Question 3: Which algorithm performs best? What happens with noise or outlier?

- Monte Carlo run with 10.000 runs: Evaluate metrics
  1. More or less similar with good data (simulated and real EEG data with added artifacts)
  2. Power ICA: quite good with reasonable SNR (~ 20dB), fast and with low variance
  3. Low sample sizes are problematic (< 5000 samples)
  4. Higher noise than 20dB SNR: smoothly becoming worse
  5. All reach breakdown point with one large outlier

- „All four algorithms: Good performance on well behaving data but higher noise level or larger outlier cause failure ➔ Violates assumptions“

Summary of major outcomes

Question 4: How to robustify?

- Whitening process crucial for algorithms
- Different objective functions for some algorithms
- Initialization quite important (local maxima/minima)
- Within algorithms: quite complicated → model based

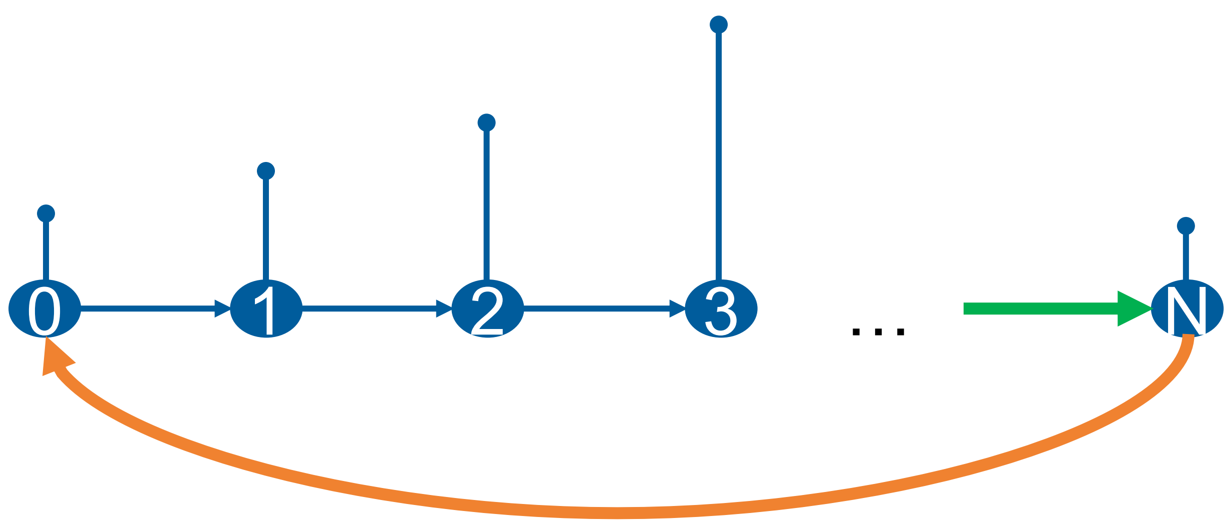
What we tried:

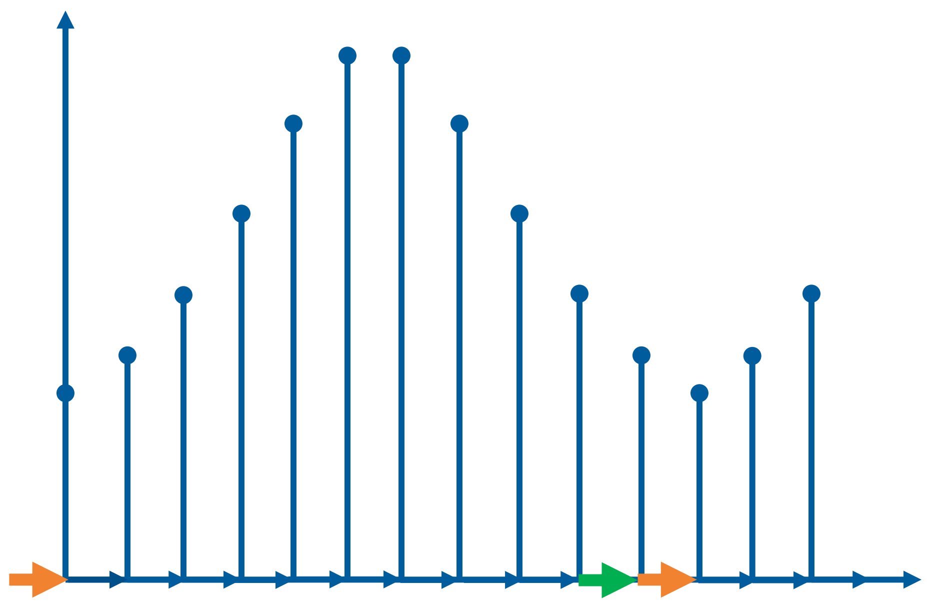
1. Plug-in robustness of covariance estimation in whitening process
2. Trying out different robust objective functions
3. Initialize Power ICA with Radical as a global optimizer

➡ „No real improvement: sometimes with noise a bit better, but still low breakdown point"

Summary of major outcomes
Question 5: What are Graph Signals / Graph BSS?

- Graph signal: Use proximity/similarity between components of the underlying signal
- „Graph BSS tries to combine the ICA approach to maximize the measure of independence and the graph signal approach to decorrelate the nodes of the graph”





Summary of major outcomes

Question 6: Outcomes of Graph BSS?

- Monte Carlo with 1000 runs:
  1. Better Blind Source Separation with good signals (low sample size)
  2. Similar with noisy data
  3. Breakdown point with one outlier as well
  4. Problem with graph structure

- „Graph BSS is good for low sample sizes and if the graph structure is available. Otherwise similar outcomes as standard BSS"
- **Real EEG data with artifacts separable!**

Agenda

TECHNISCHE UNIVERSITÄT DARMSTADT

1. Understanding the basis of Source Separation via an intuitive example-driven approach
2. Discussion of experiment design
3. Graph Blind Source Separation
4. Graph Blind Source Separation results
5. Summary of major outcomes
6. Problems and future research

Problems and future research

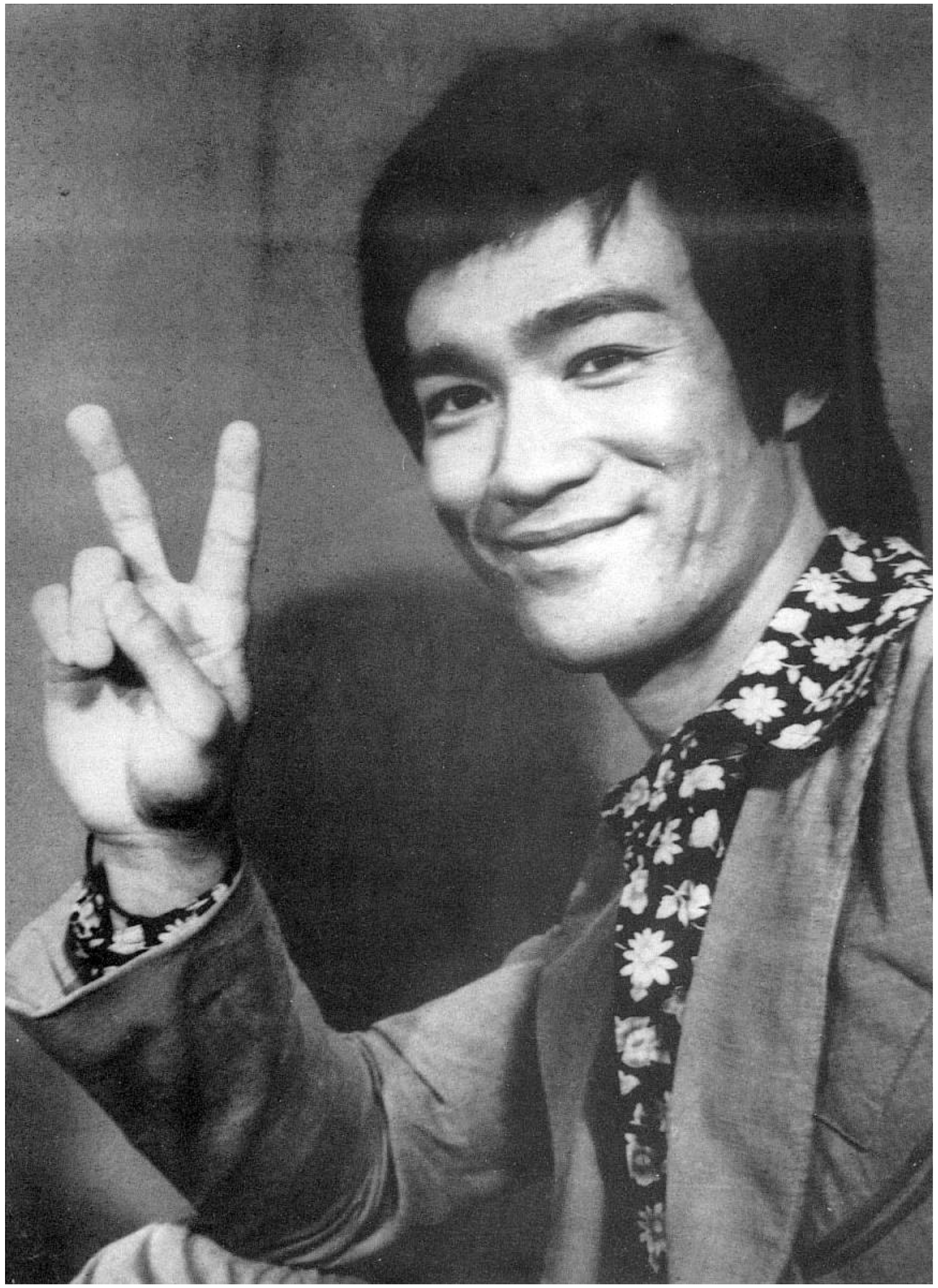
Problems:

1. Low Breakdown point
2. High noise dependency
3. Sample size important for standard ICA
4. Graph structure estimation: How to find correlation of graph signals?
5. Covariance estimation in higher dimensions problematic
6. Time consumption for real time EEG analysis

Future research:

- Robust whitening / preprocessing (Robust covariance estimation in high dimensions)
- Incorporate noise process and outlier in model for ICA algorithm
- Find good graph structure estimation (recent research [Garcia2020])

Thank you for your attention!



Literature
Chapter 1

- [NY Times, 2020] https://www.nytimes.com/2018/01/18/arts/design/brain-neuroscience-santiago-ramon-y-cajal-grey-gallery.html
- [BackyardBrains] https://backyardbrains.com/experiments/EEG#prettyPhoto
- [iStock] https://medium.com/better-humans/how-to-have-more-meaningful-conversations-7b1f9120ff0d
- [Hyvärinen, 1999] "Fast and robust fixed-point algorithm for independent component analysis" IEEE Trans. Neural Netw. vol. 10 no. 3 pp. 626-634 May 1999.
- [Basiri, 2017] S., Ollila, E., & Koivunen, V. (2017). Alternative Derivation of FastICA with Novel Power Iteration Algorithm. IEEE Signal Processing Letters, 24(9), 1378-1382. [7994645]. https://doi.org/10.1109/LSP.2017.2732342
- [Jutten, 2010] Handbook of Blind Source Separation: Independent Component Analysis and Applications
- [Shlens, 2010]: A Tutorial on Independent Component Analysis
- [Haykin, 2009], Neural Networks and Learning Machines
- [Sejnowski TJ , 1995] Bell AJ, Sejnowski TJ (November 1995). "An information-maximization approach to blind separation and blind deconvolutio
- [Learned-Miller, 2003] John W. Fisher III, "ICA Using Spacings Estimates of Entropy", Journal of Machine Learning Research 4 (2003) 1271-1295
- [Pfister 2019], Robustifying Independent Component Analysis by Adjusting for Group-Wise Stationary Noise, Sebastian Weichwald, Peter Bühlmann, Bernhard Schölkopf; 20(147):1−50, 2019.
- [Belouchrani 1997], K. Abed-Meraim, J.-F. Cardoso, and E. Moulines, “A Blind Source Separation Technique Using Second-Order Statistics,” IEEE Transactions on Signal Processing, vol. 45, no. 2, pp. 434–444, 1997.
- [Cardoso 1993], A. Souloumiac, “Blind Beamforming for non Gaussian Signals,” IEEE Proceedings-F, vol. 140, pp. 362–370, 1993.

Literature
Chapter 2

- [Delorme2007] - Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis in NIH 2007
- [Koudelková2018] - Introduction to the identification of brain waves based on their frequency
- [Ilmonen, 2010] - A New Performance Index for ICA: Properties, Computation and Asymptotic Analysis

Literature
Chapter 2

- [Miettinen, 2020] Graph Signal Processing Meets Blind Source Separation, Jari Miettinen, Esa Ollila et al., 2020
- [Blöchl, 2010] Second-Order Source Separation Based on Prior Knowledge Realized in a Graph Model, Florian Blöchl, Andreas Kowarsch, Fabian J. Theis, 2010
- [Djuric, 2018] Cooperative and Graph Signal Processing Principles and Applications, Petar Djuric, Cédric Richard, 2018

Literature
Chapter 3

- Delorme et al.: “*Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis”,* Elsevier 2006
- Garcia et al.: *”Blind Demixing of Diffused Graph Signals”,* arXiv 2020

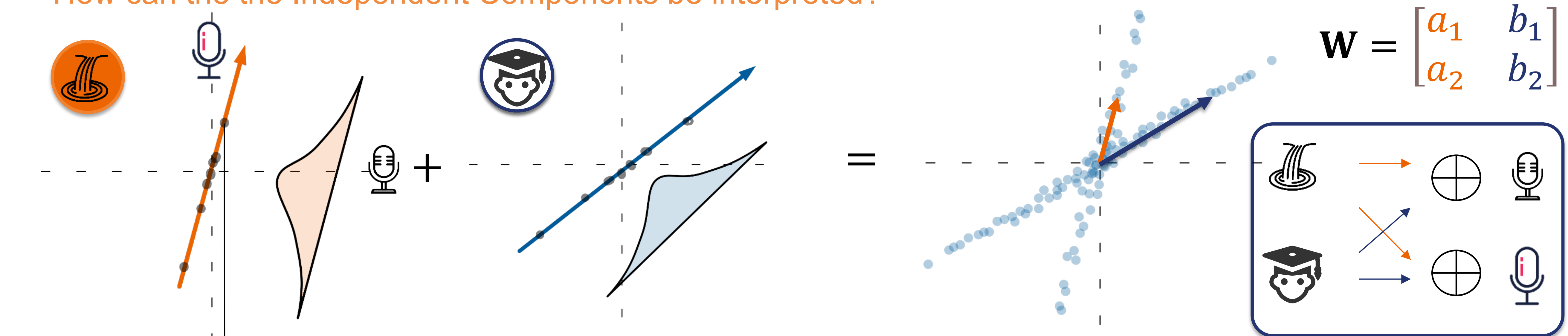
Back Up I:
Alternative Interpretation of ICA

What is the goal of the algorithm?

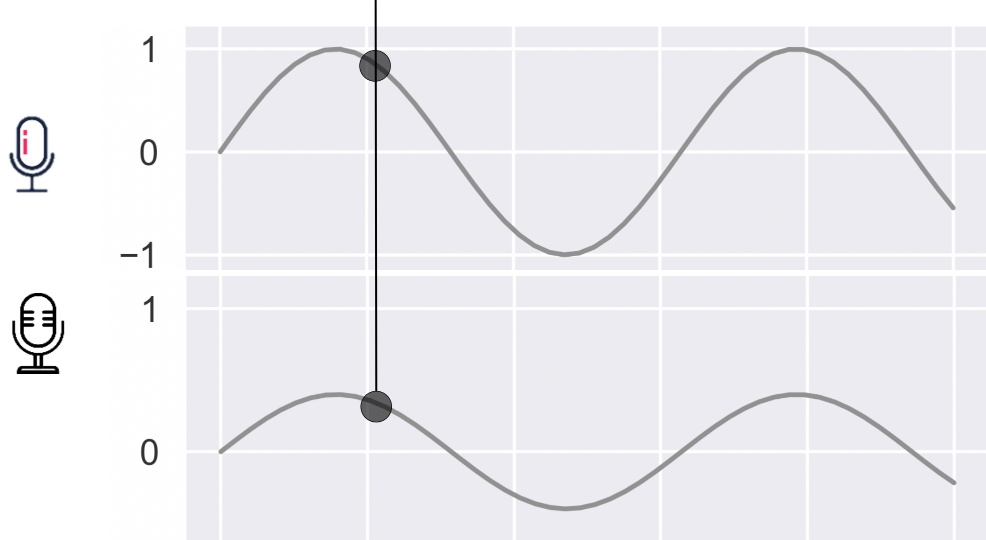
$$\mathbf{SW} = \mathbf{X} \qquad \hat{\mathbf{S}} = \mathbf{XW}^{-1}$$

Find the demixing matrix $\mathbf{A} \triangleq \mathbf{W}^{-1}$ so that the latent source variable s can be reconstructed as $\hat{\boldsymbol{s}}$ by the observed data $\boldsymbol{x}$

How can the the Independent Components be interpreted?



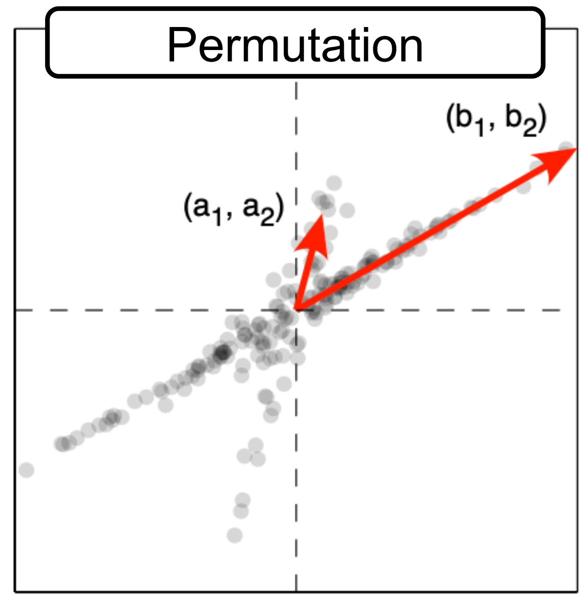
$$\mathbf{W} = \begin{bmatrix} a_1 & b_1 \\ a_2 & b_2 \end{bmatrix}$$



- Amplitudes recorded simultaneously in microphones 1 and 2 are plotted on the x and y axes for each sound (waterfall, student) separately
- Waterfall: samples lie on the vector $\mathbf{a} = (\mathbf{a1}\,, \mathbf{a2})$ reflecting the proximity of the sound to microphones 1 and 2
- Independent components represent to the basis used to generate the data

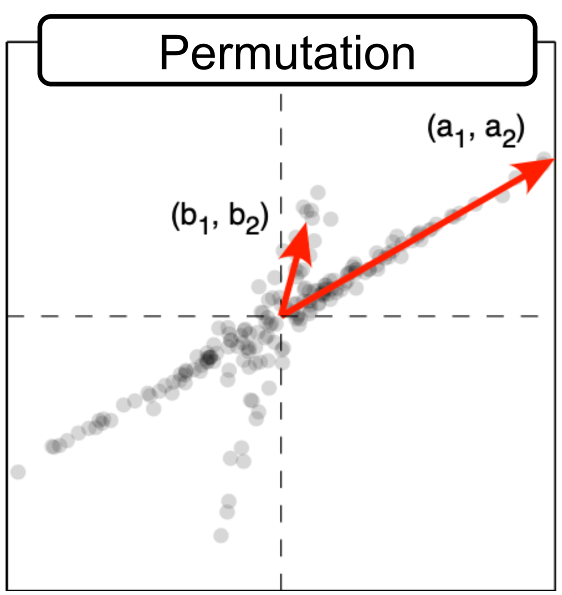
Back Up II:
Notes on peculiarities of ICA

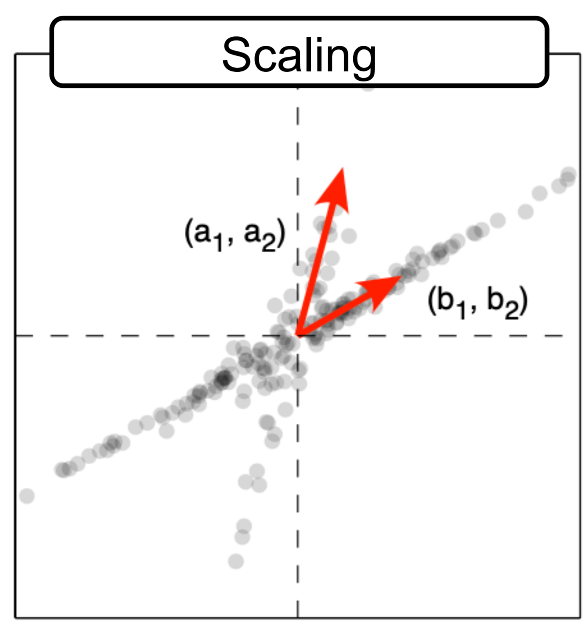
Original Signals

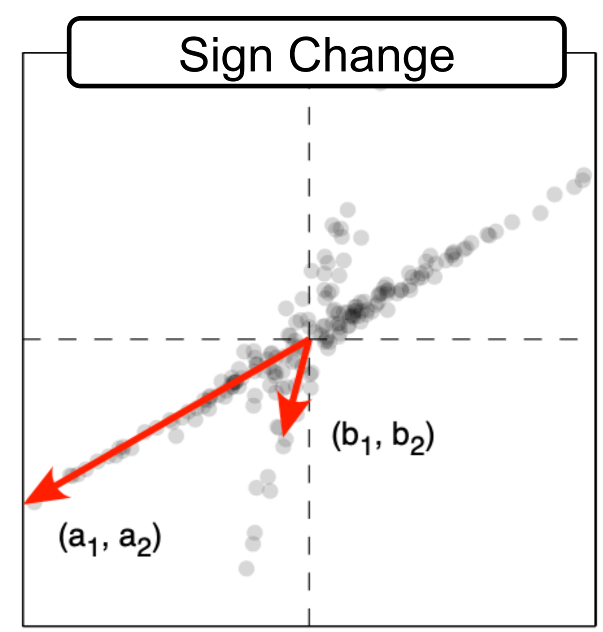


Reconstructred signal Signals

Due to ambiguities of the method the extracting of the unobserved independent source signals is only accurate up to: $d \times d$ permutation matrix $\boldsymbol{P}$, scaling matrix $\boldsymbol{D}$ and sign-change matrix $\boldsymbol{L}$



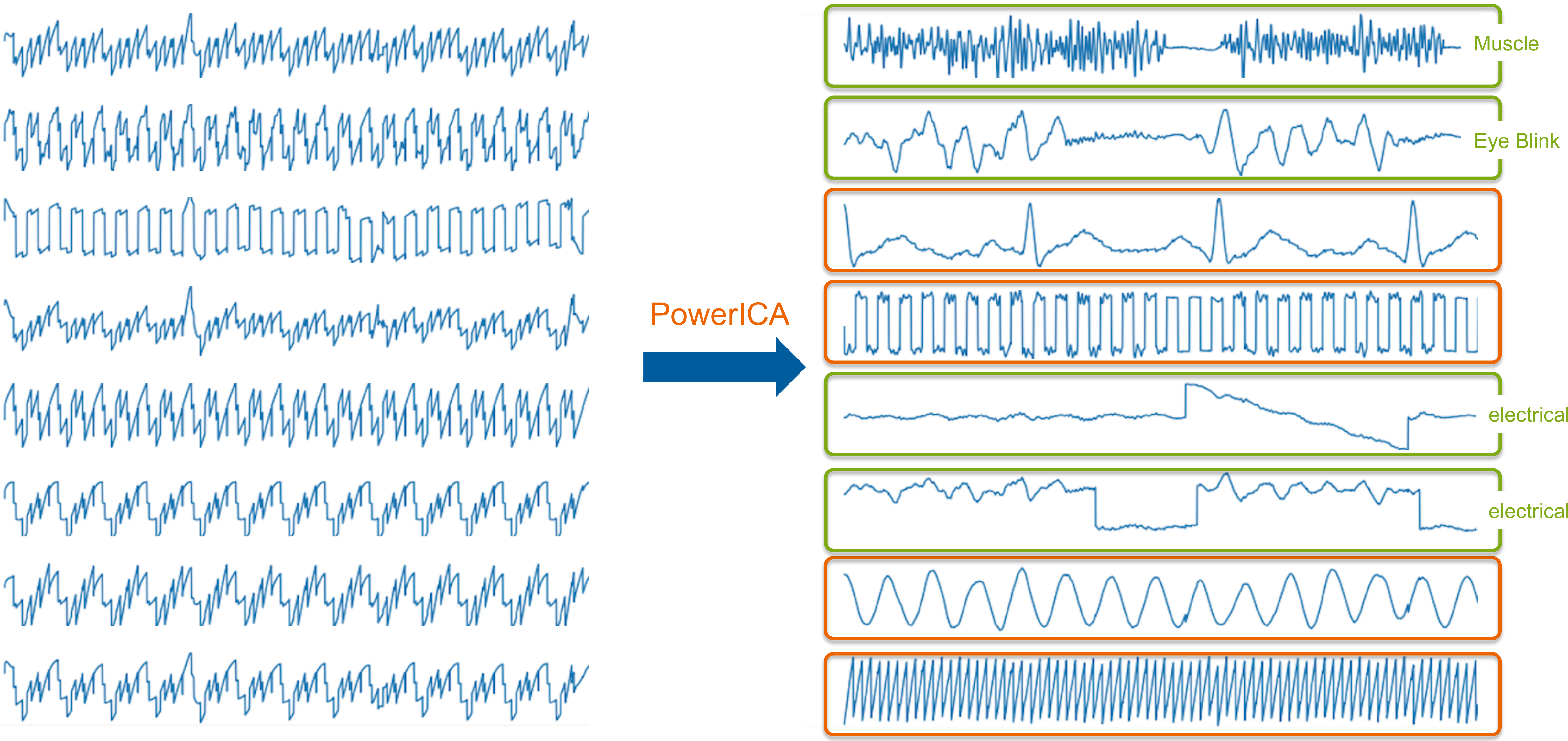




Back Up III: Towards EEG data
Results with synthetic EEG data

Mixed signals + added artifacts

Reconstructed signals + artifacts



Able to reconstruct the standard signals and the EEG artifacts

Back Up IV:
Distribution of Signals