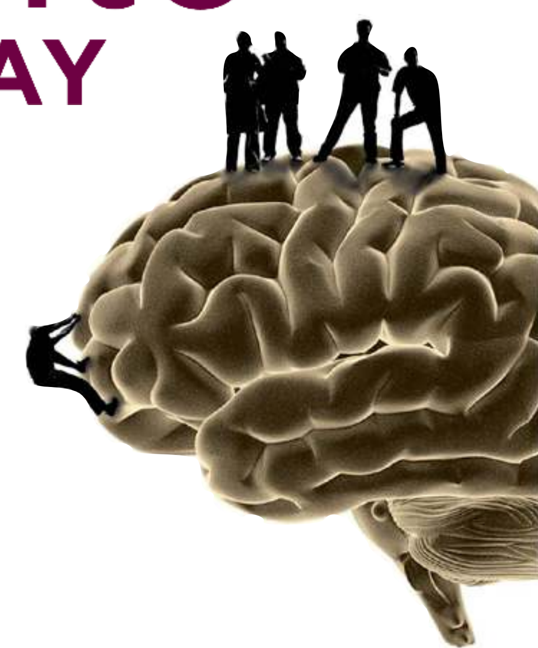
Learning representations from neural signals

Alexandre Gramfort
alexandre.gramfort@inria.fr

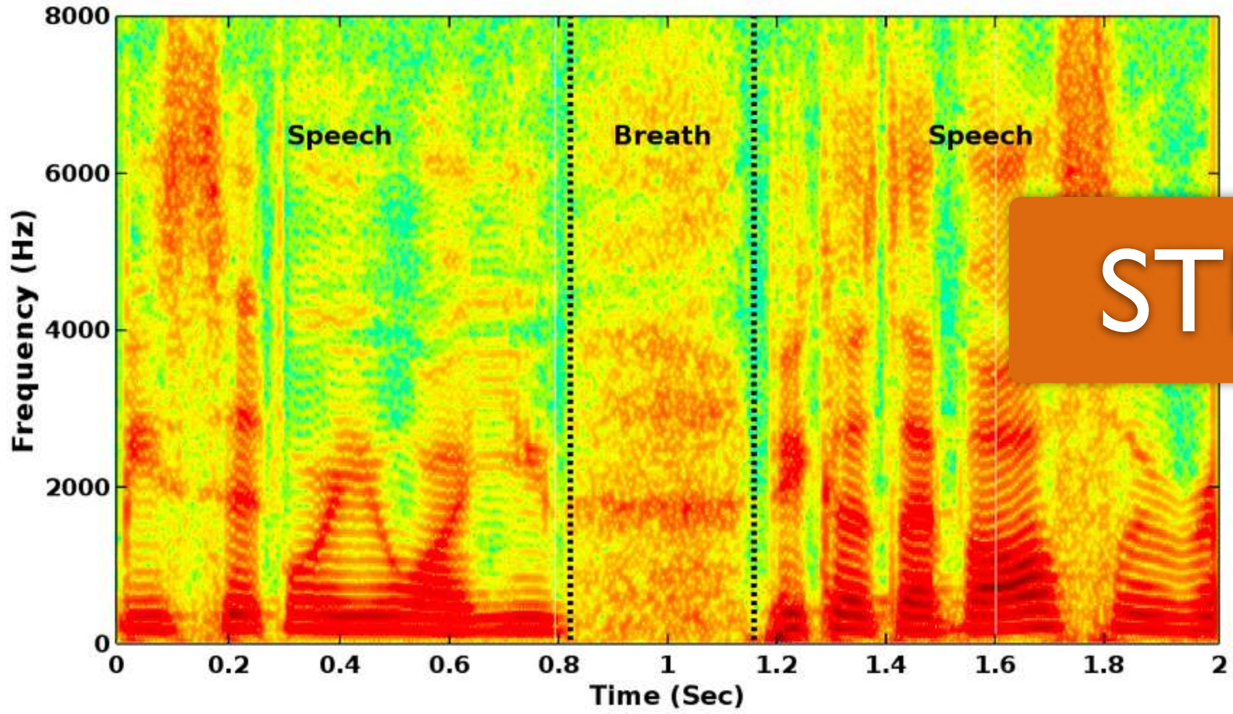
informatics mathematics

université PARIS-SACLAY



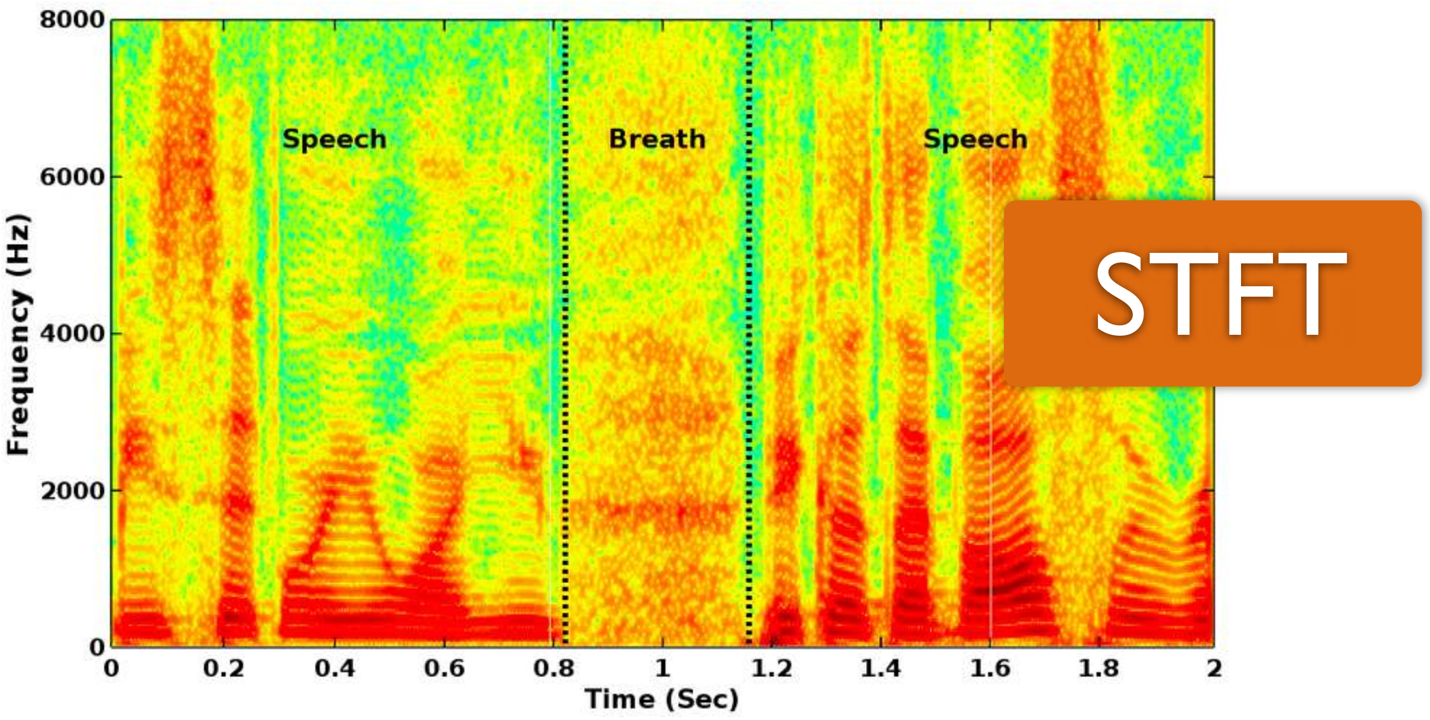
PalAIsien Seminar - Nov. 2020

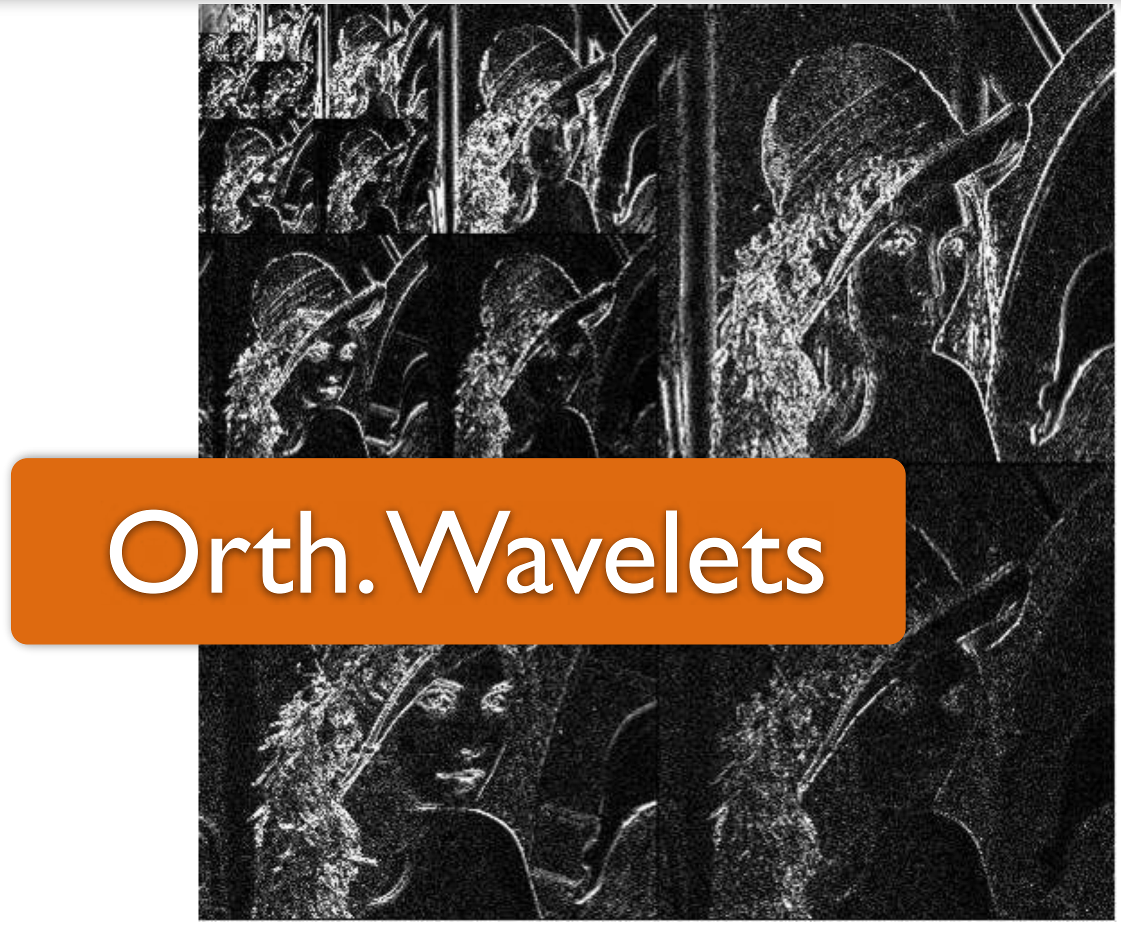
What are representations?



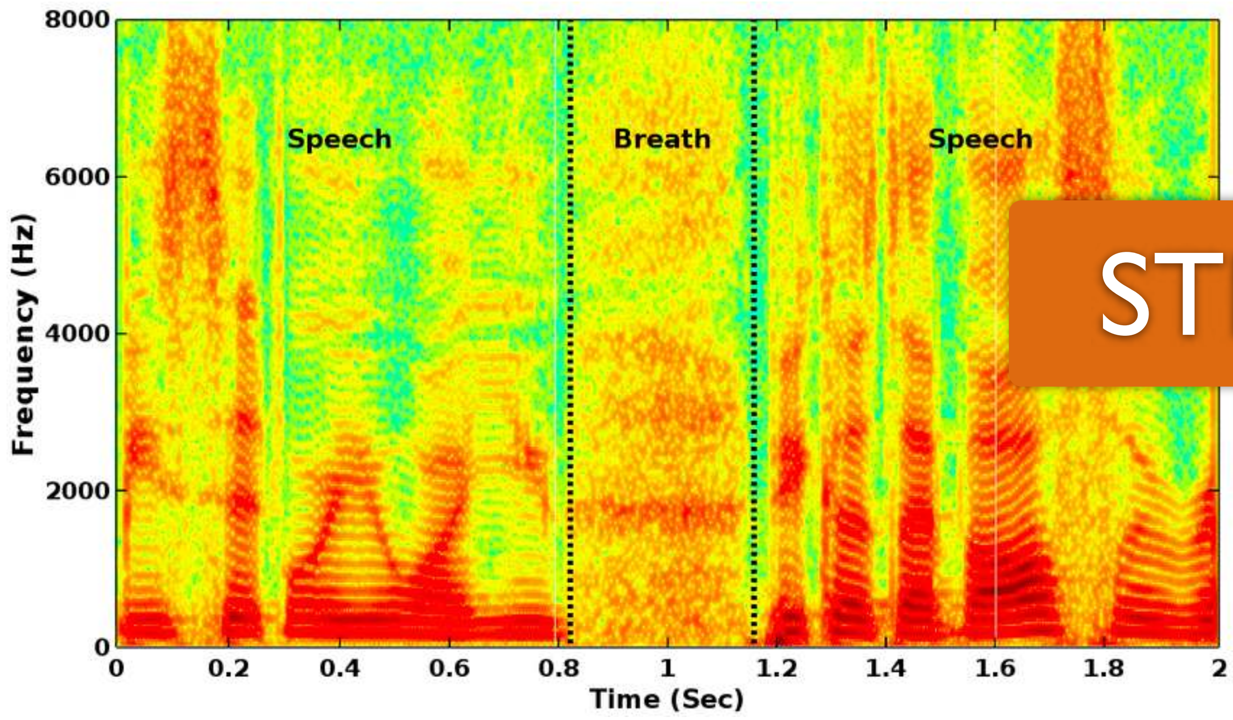
STFT

What are representations?

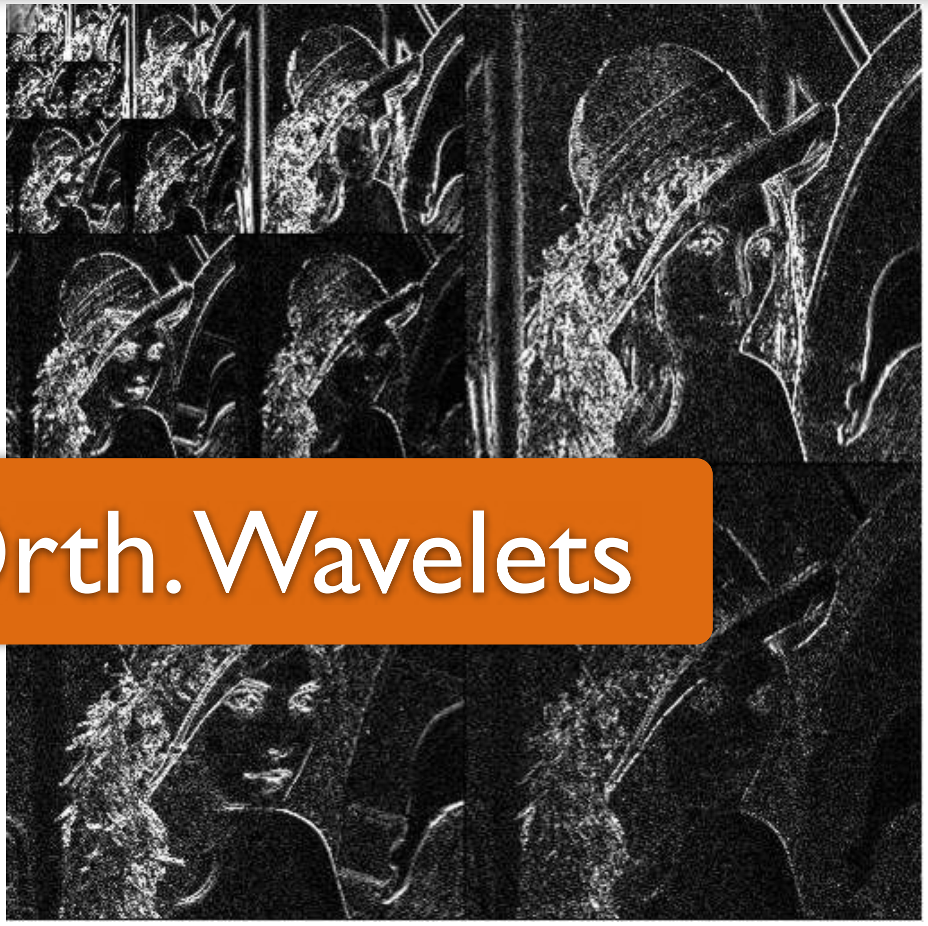




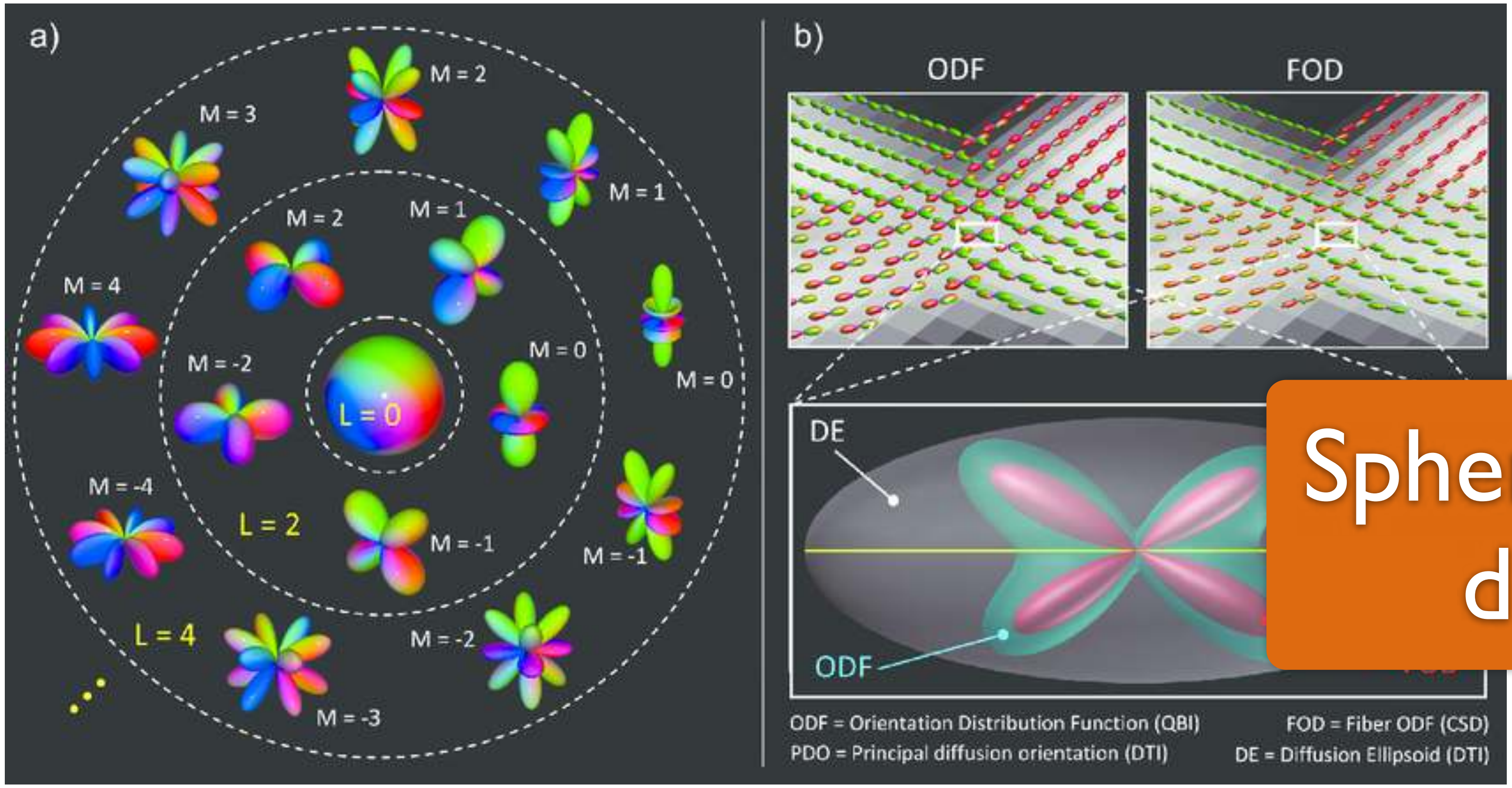
What are representations?



STFT

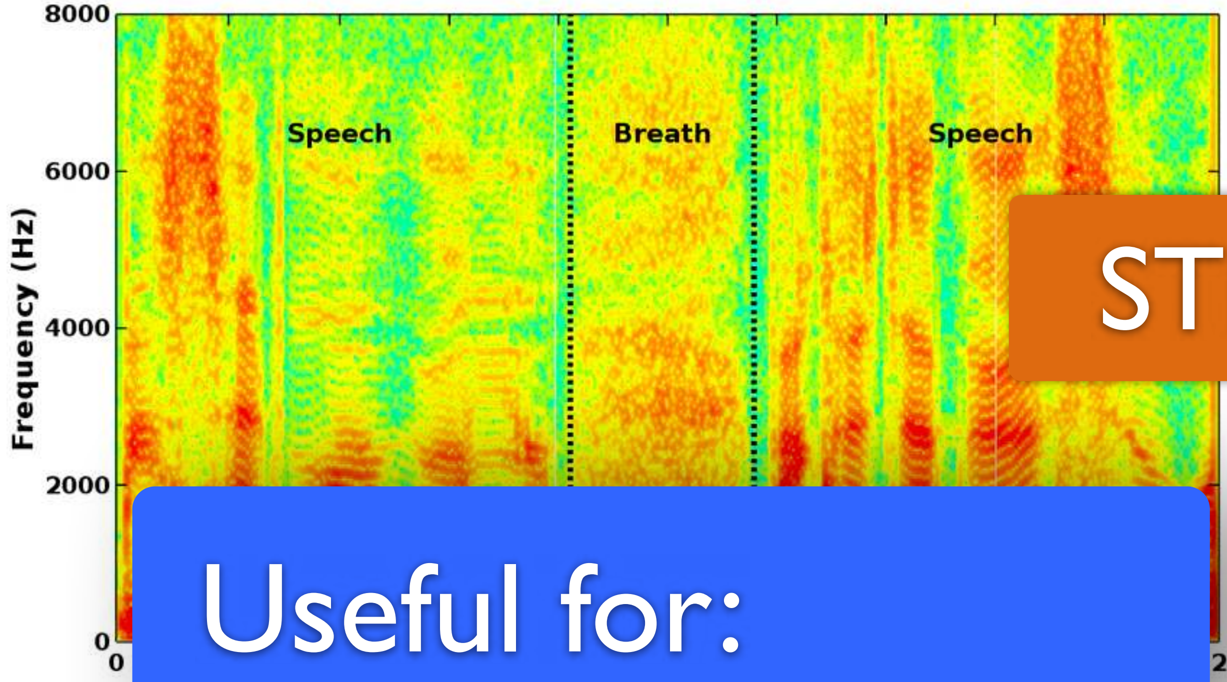


Orth. Wavelets

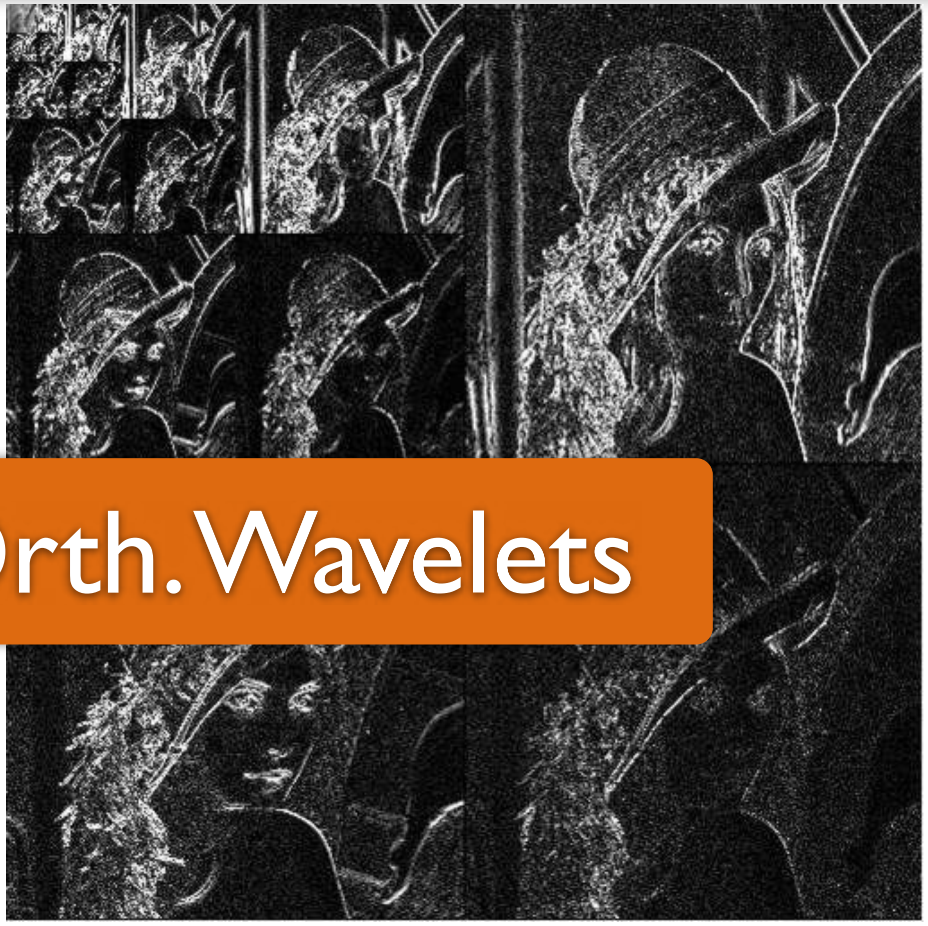


Spherical Harmonics
diffusion MRI

What are representations?



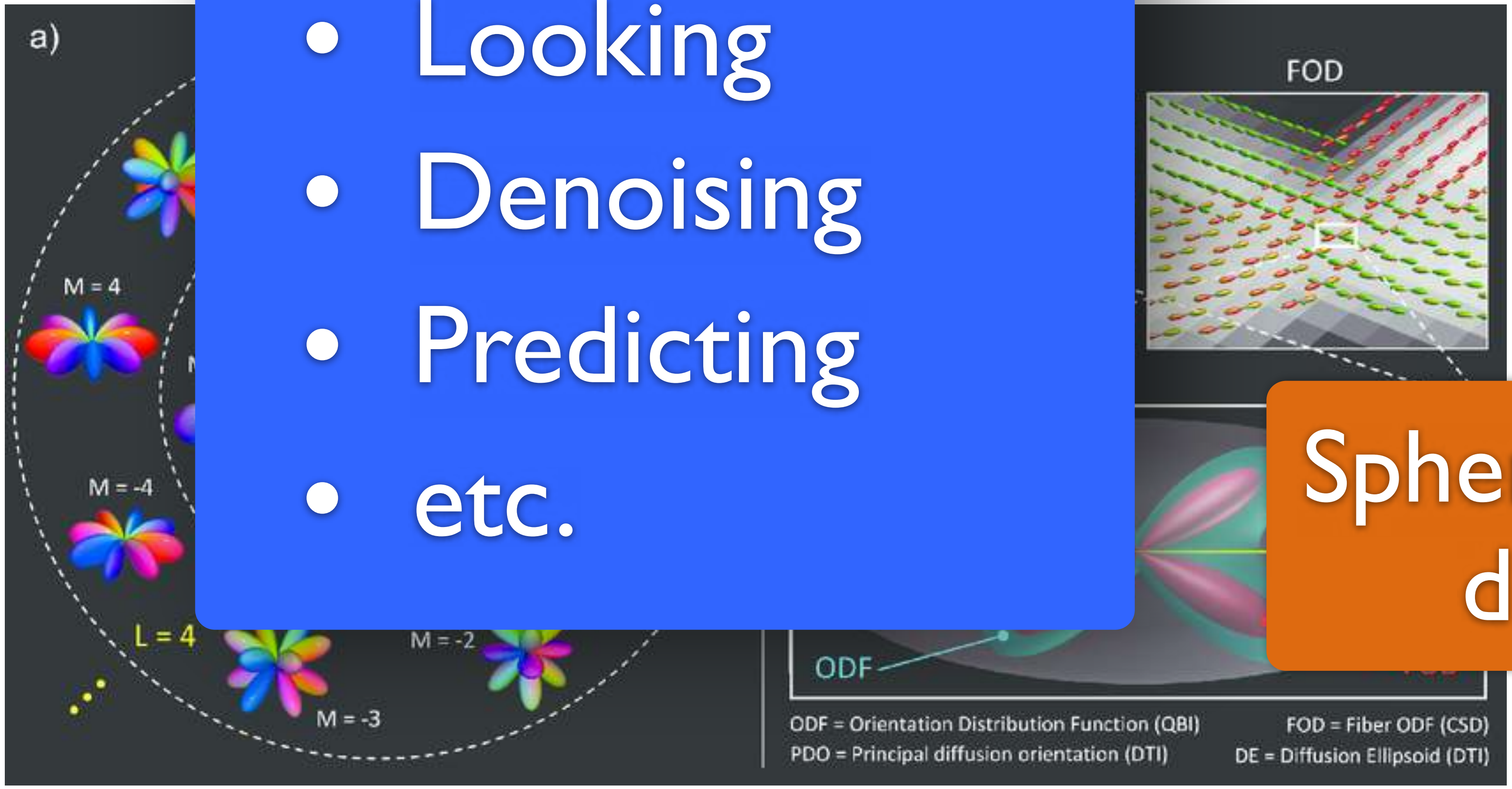
STFT



Orth. Wavelets

Useful for:

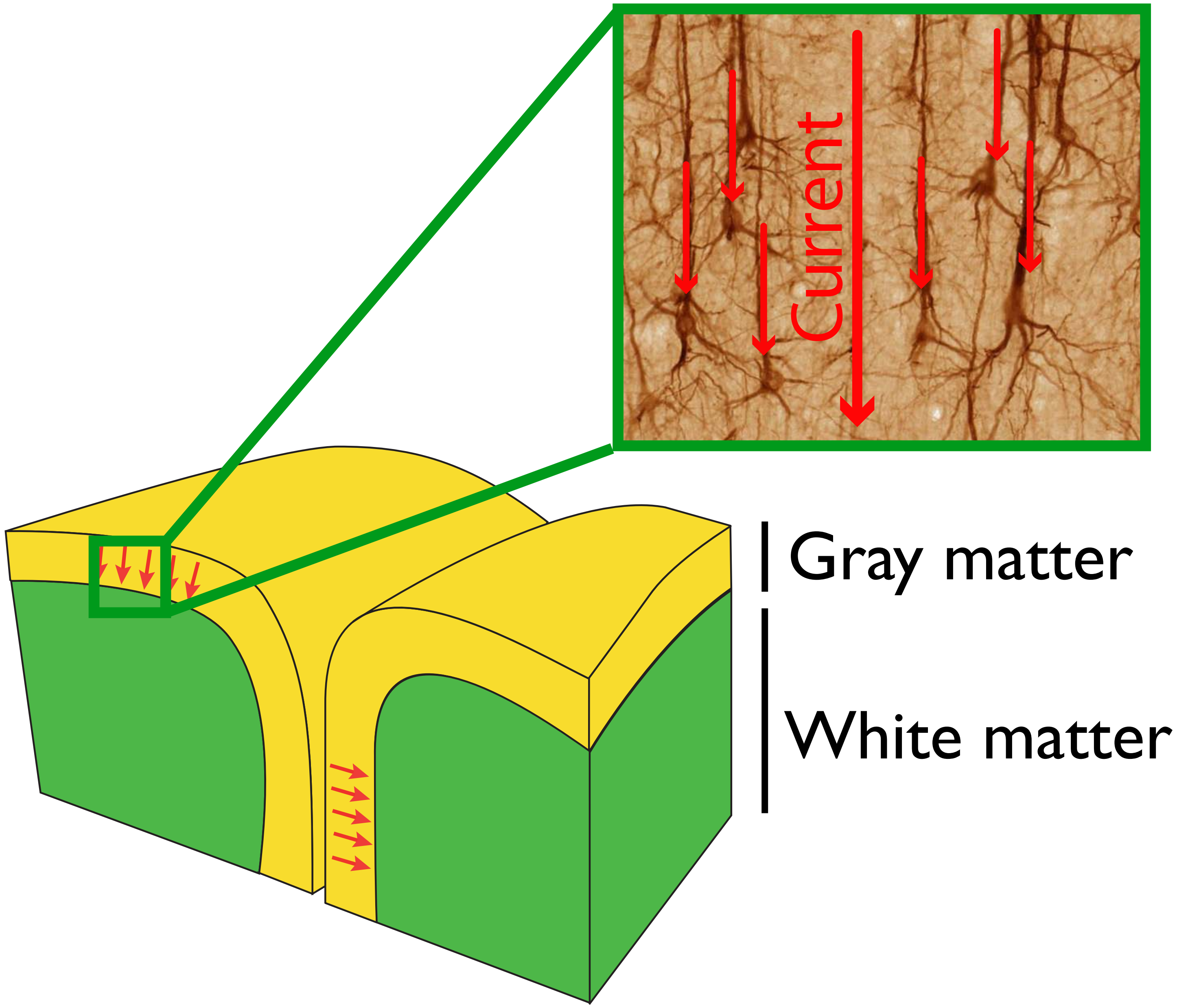
- Looking
- Denoising
- Predicting
- etc.



Spherical Harmonics diffusion MRI

Neurons as current generators

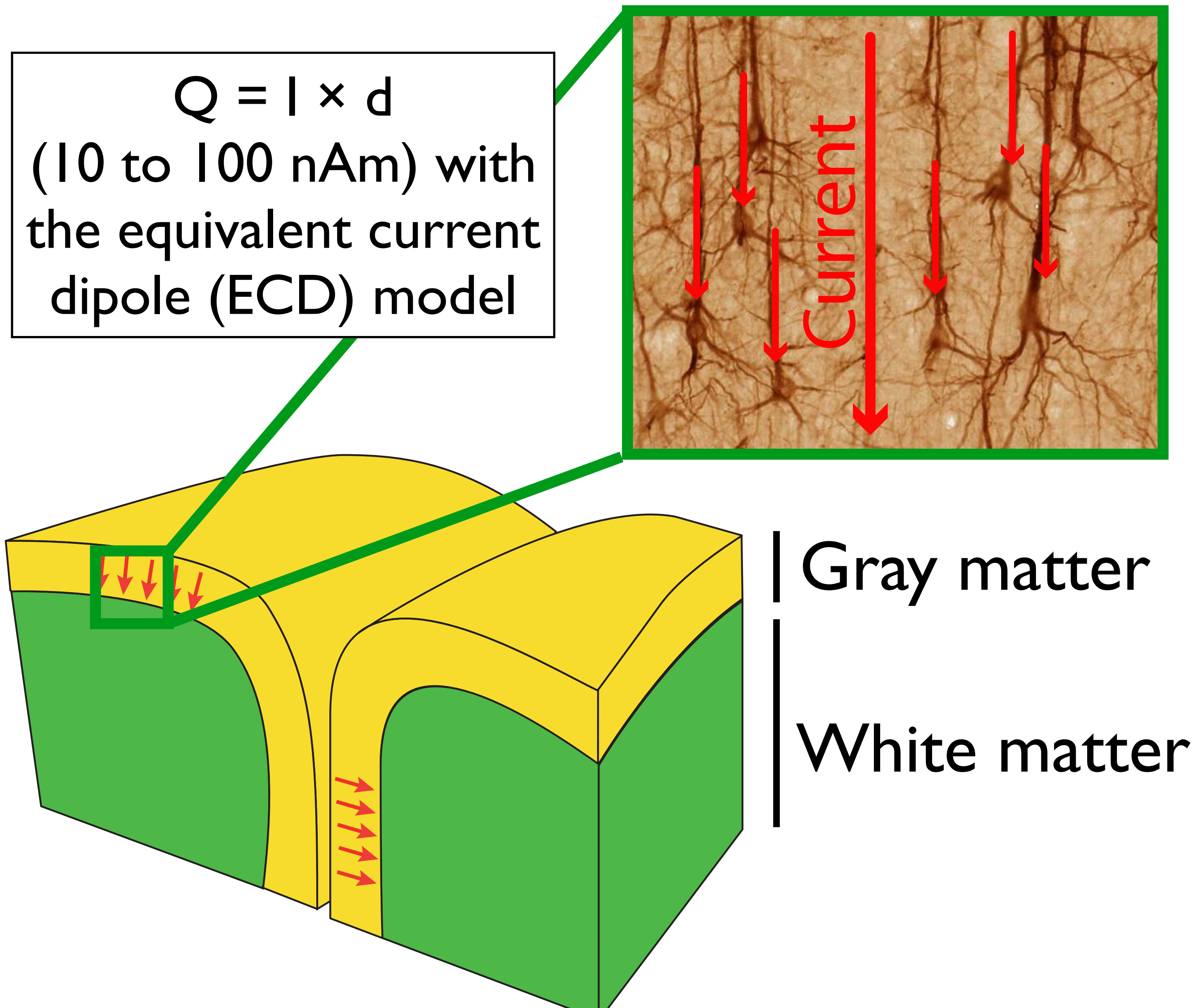
Large cortical pyramidal cells organized in macro-assemblies with their **dendrites normally oriented to the local cortical surface**

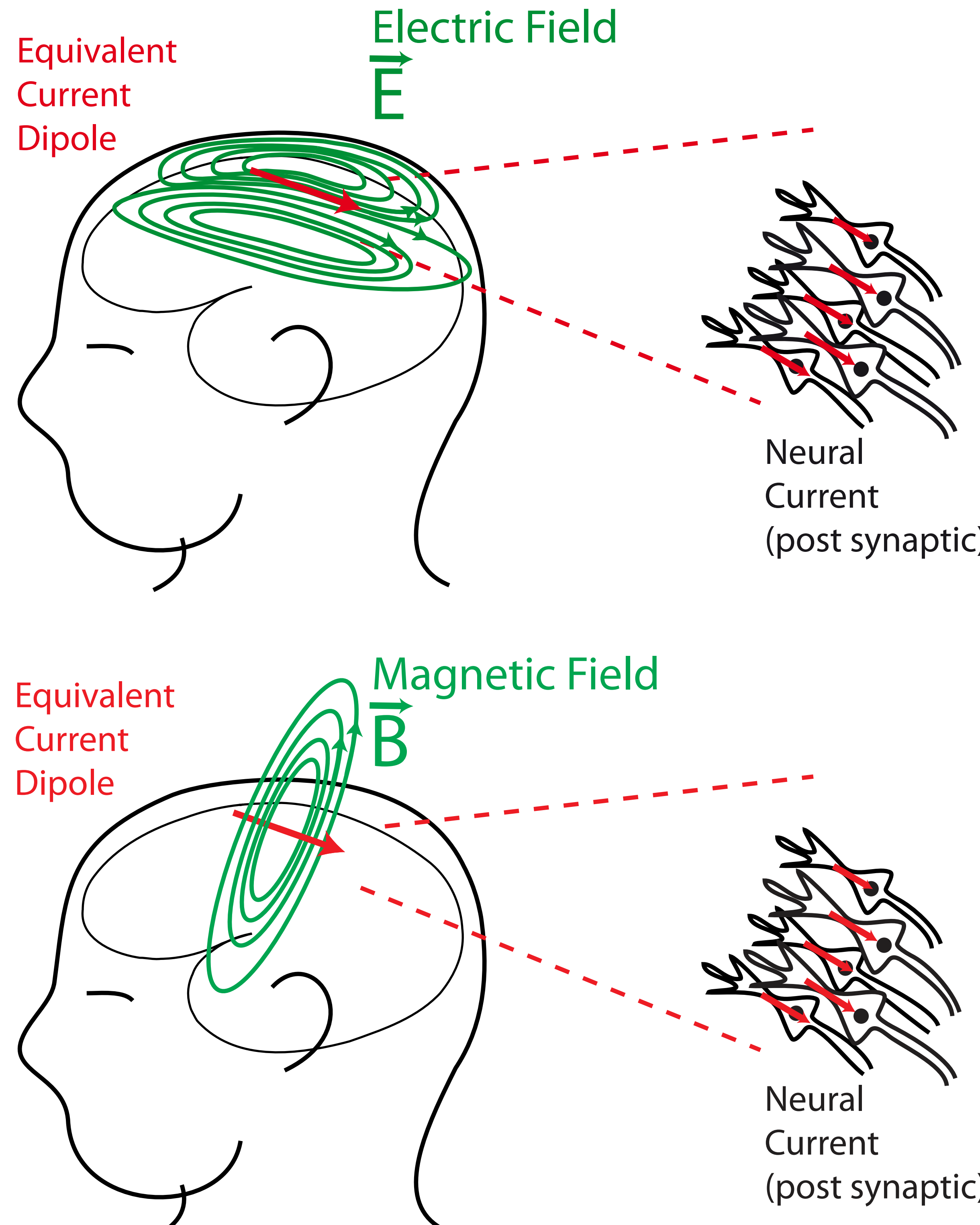


Neurons as current generators

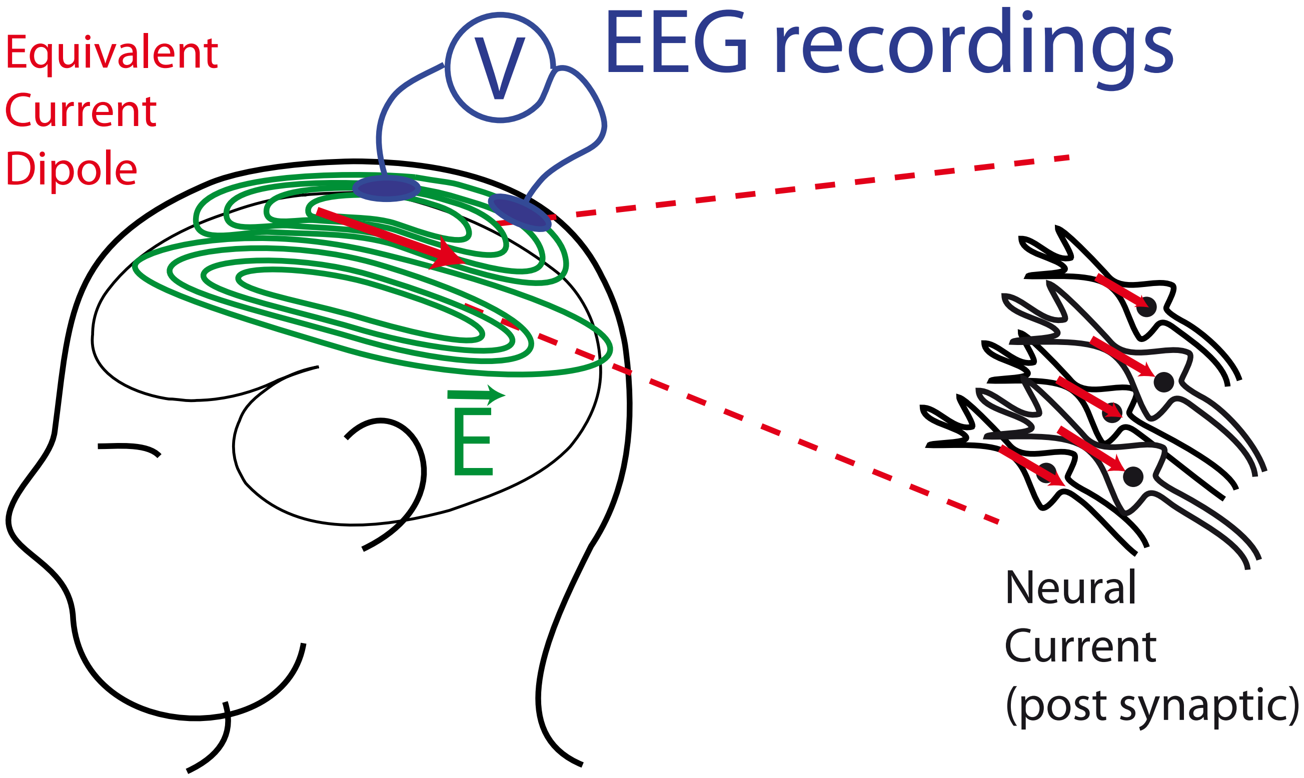
Large cortical pyramidal cells organized in macro-assemblies with their **dendrites normally oriented to the local cortical surface**

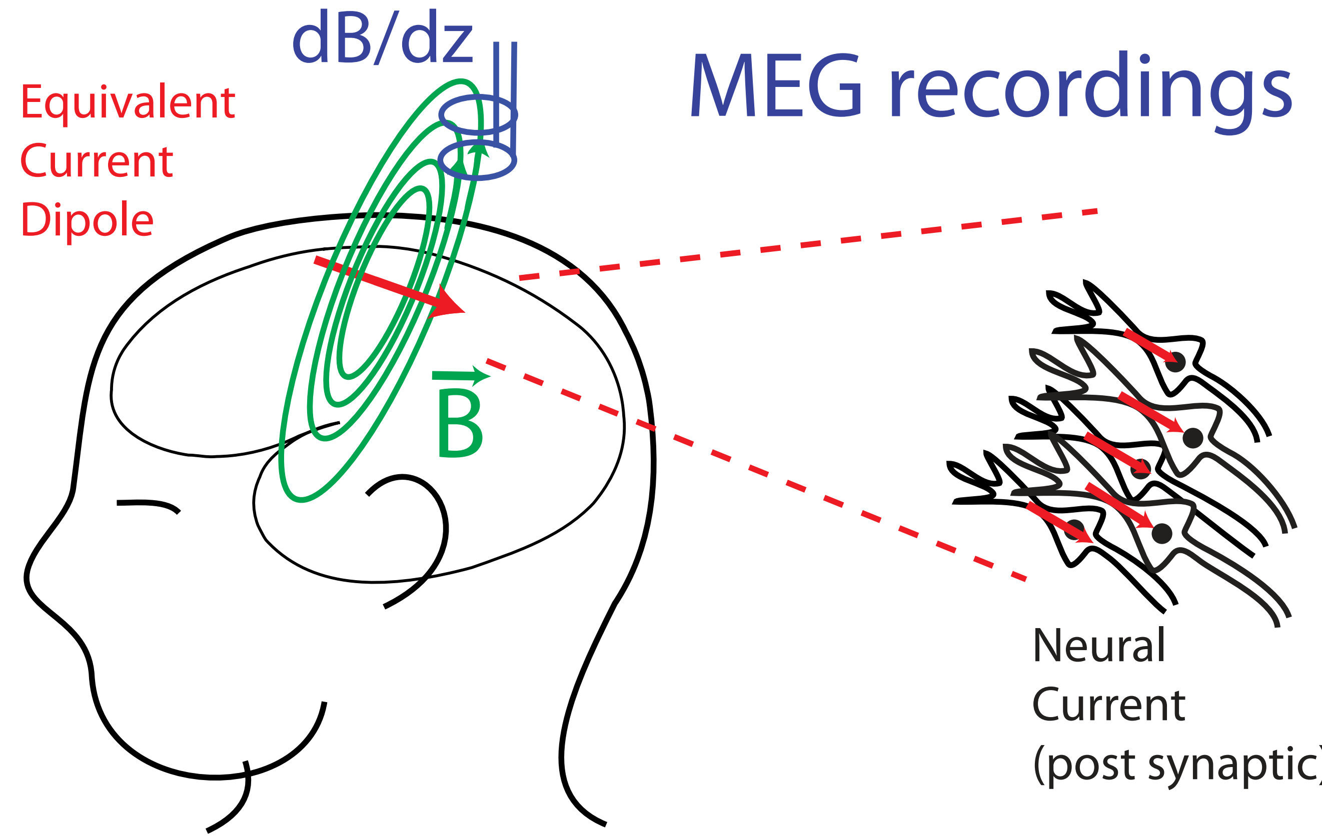
$Q = I \times d$
(10 to 100 nAm) with the equivalent current dipole (ECD) model



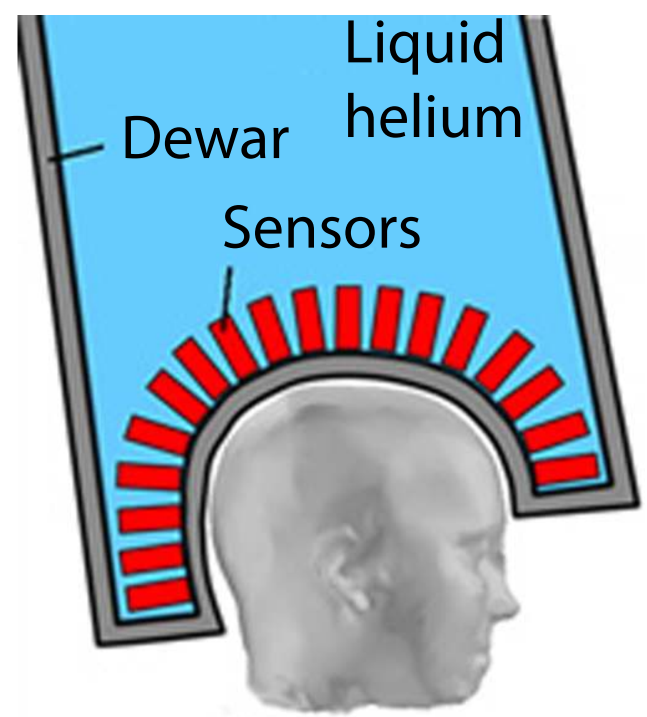


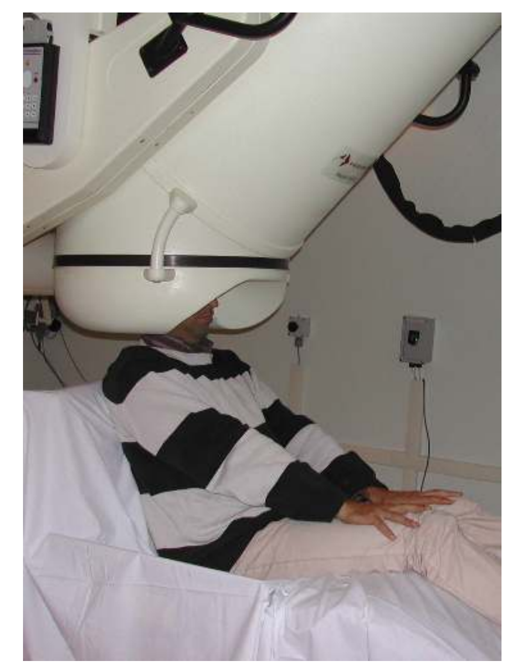
Electro- & Magneto-encephalography





First EEG recordings in 1929 by H. Berger

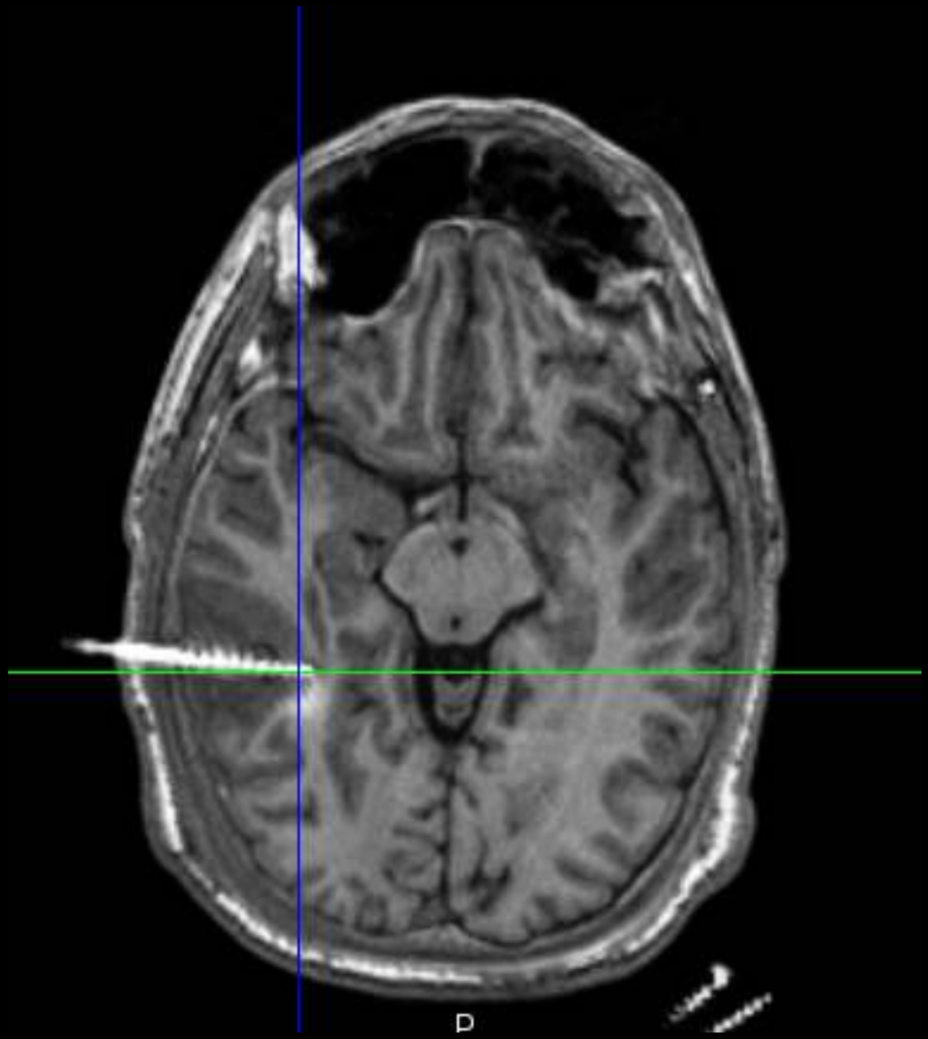




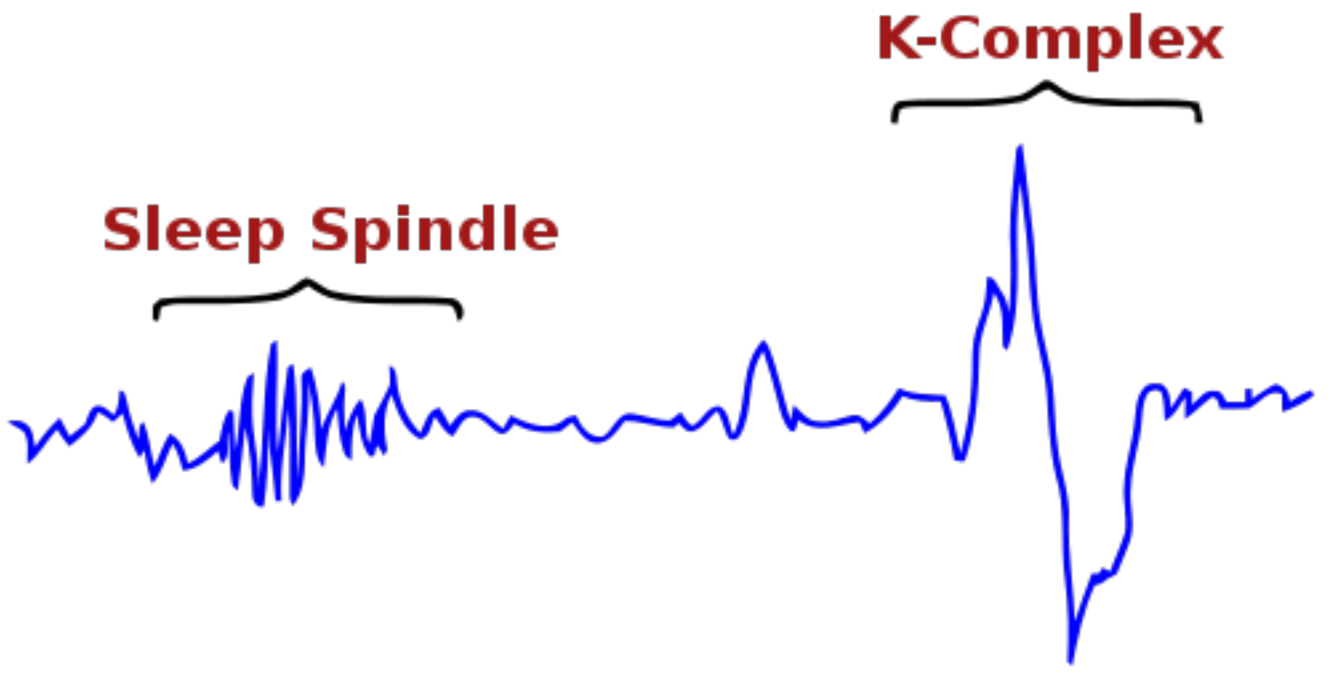
Hôpital La Timone
Marseille, France

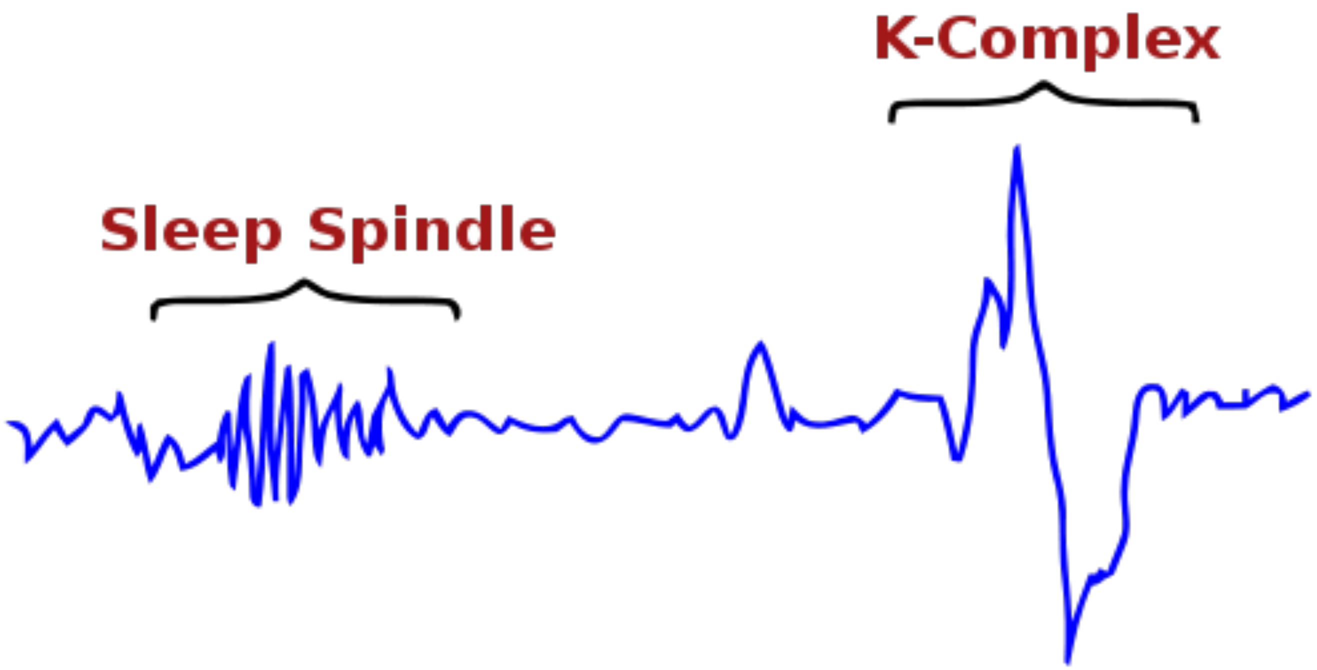
Stereotaxic EEG (sEEG)

Intracranial electrodes;
5 to 15 contacts per electrode
Around 10 electrodes are implanted

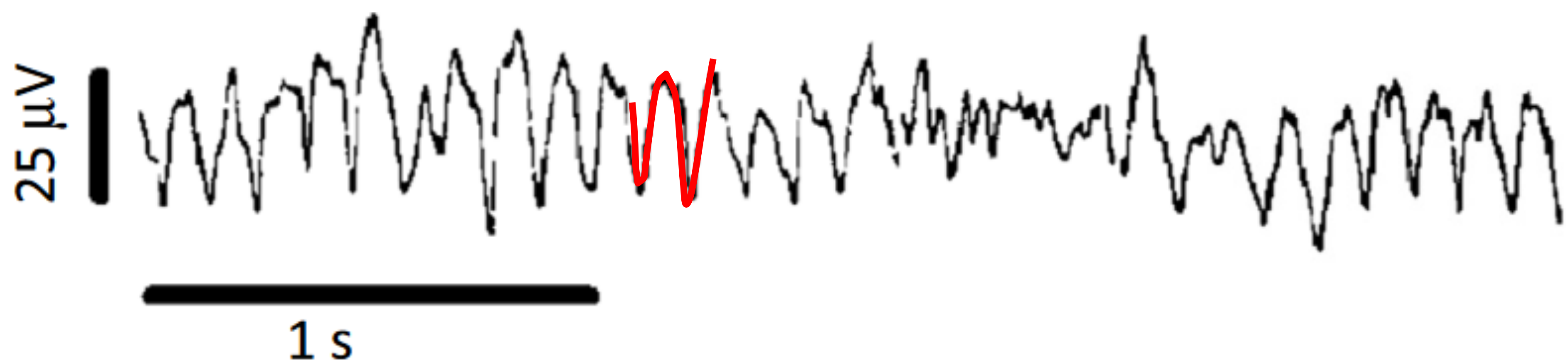


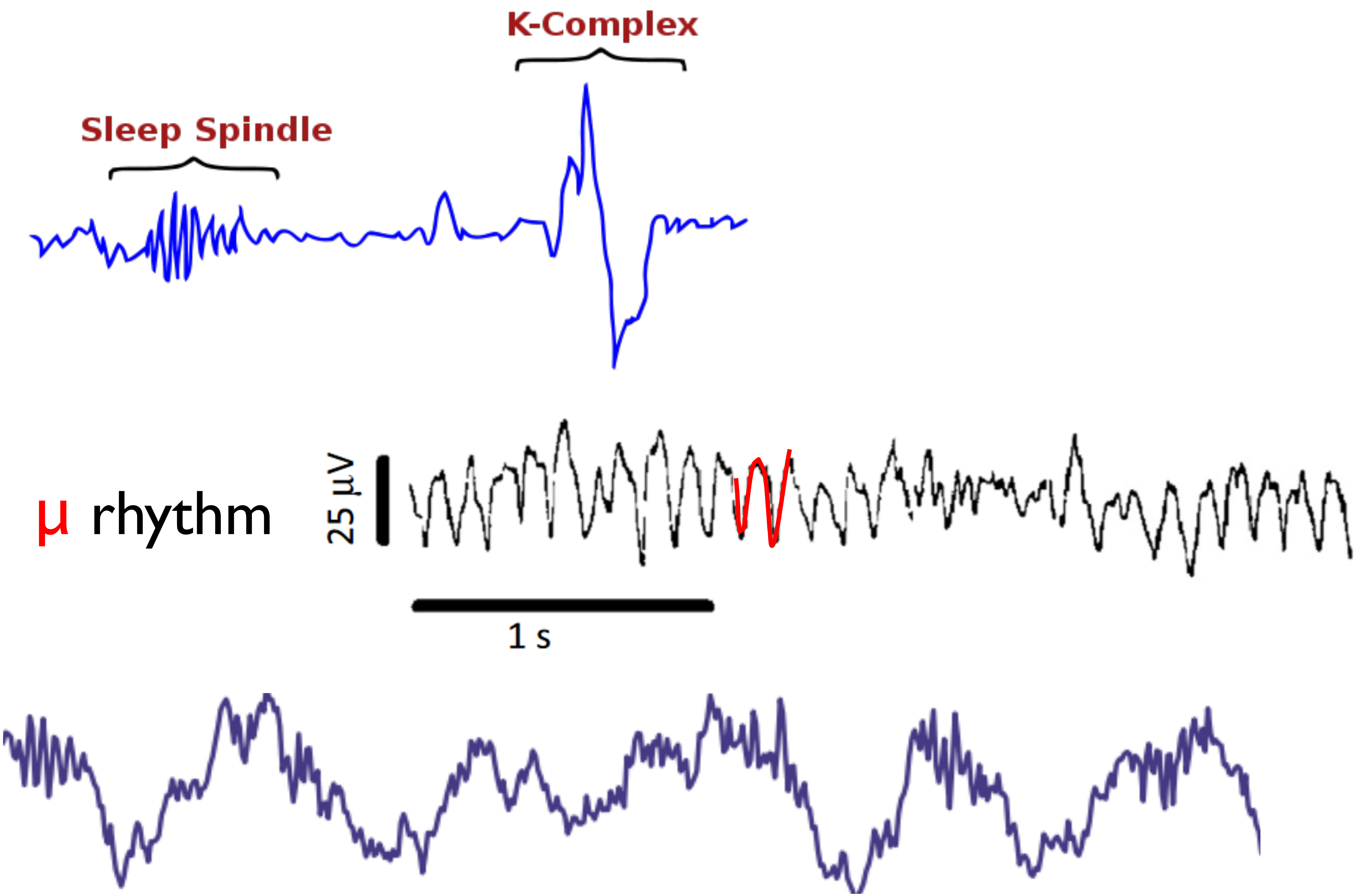
Stereotaxic Implantation


K-Complex
Sleep Spindle

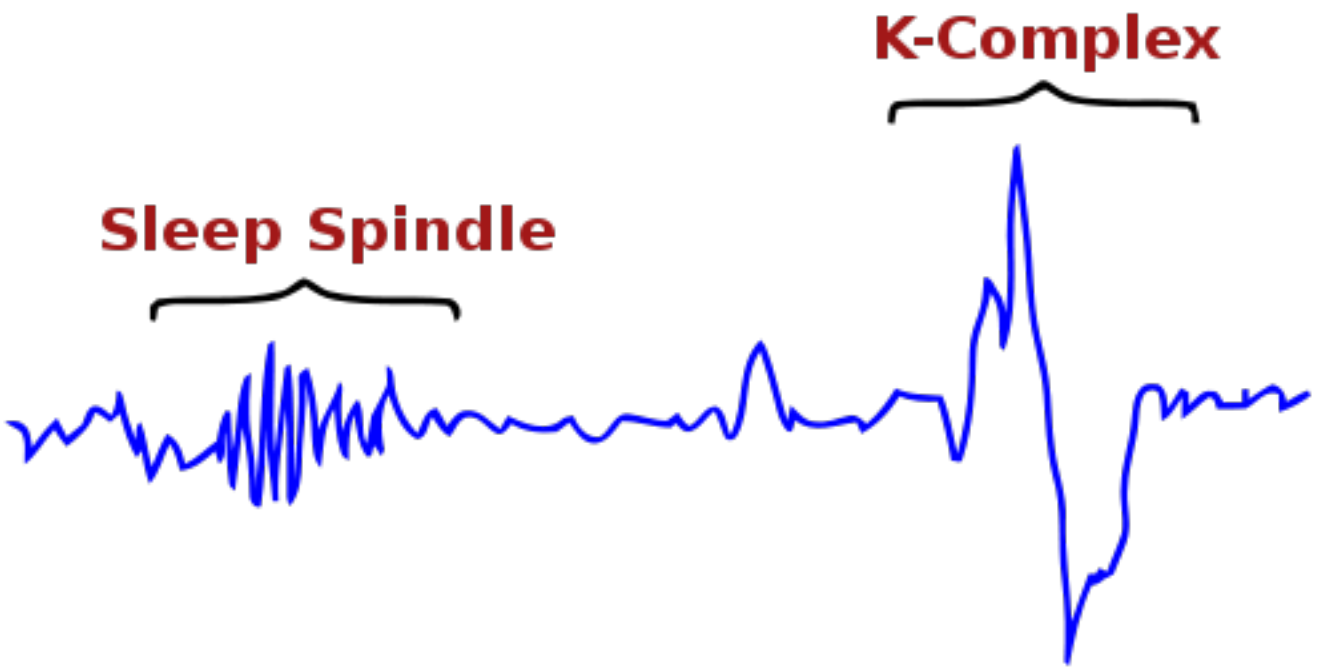


μ rhythm

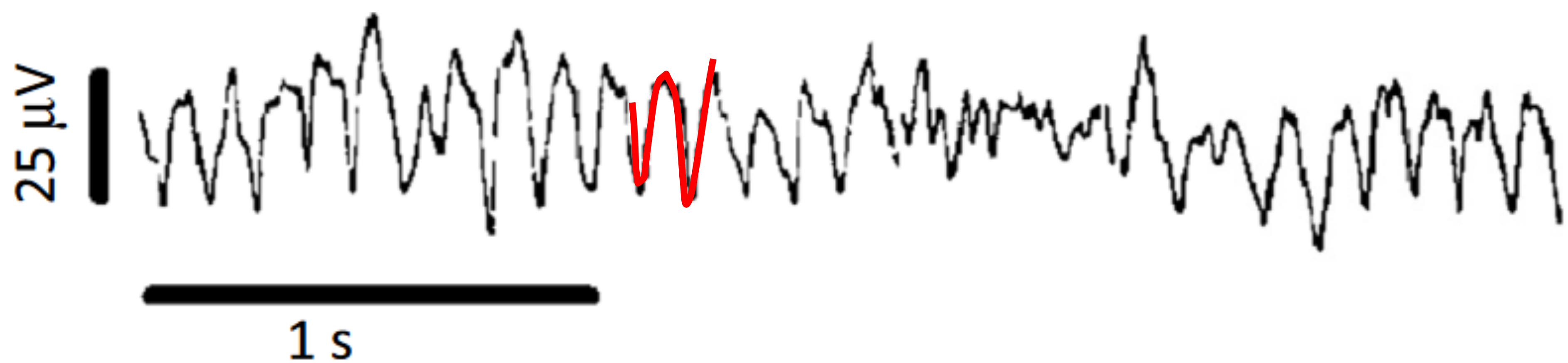




[T. Dupré la Tour, L. Tallot, L. Grabot, V. Doyère, V. van Wassenhove, Y. Grenier, A. Gramfort, (2017) PLOS Computational biology]

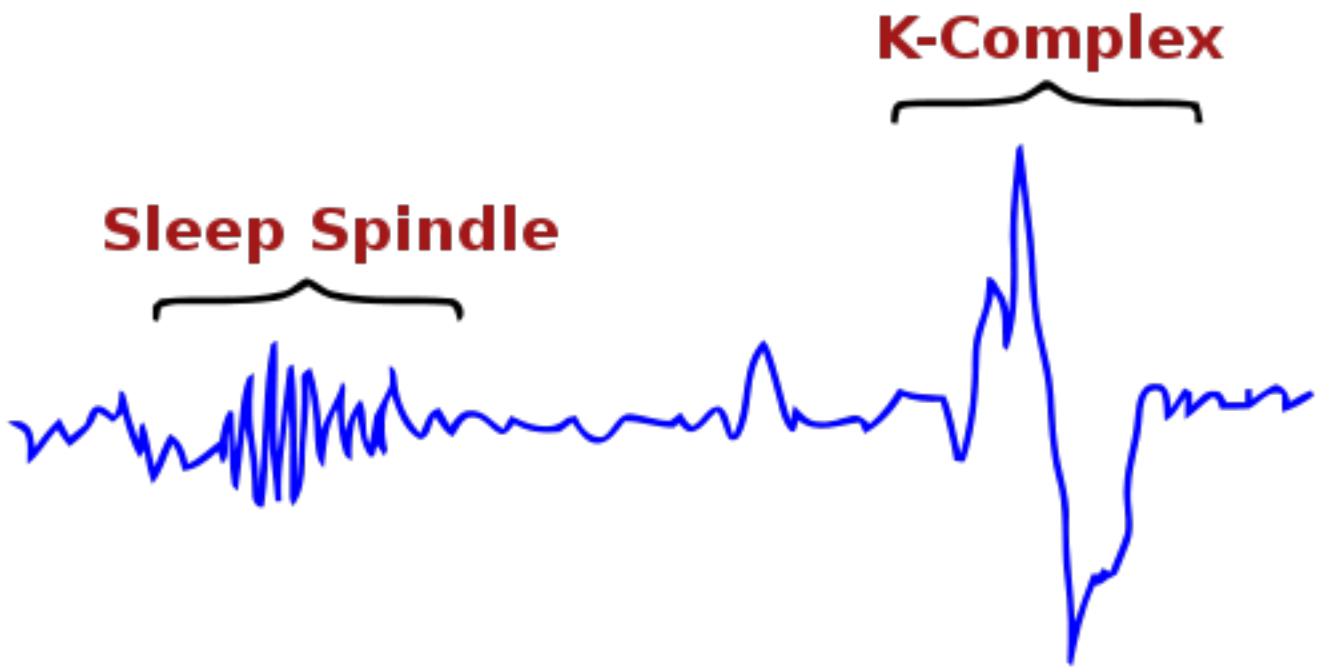


μ rhythm



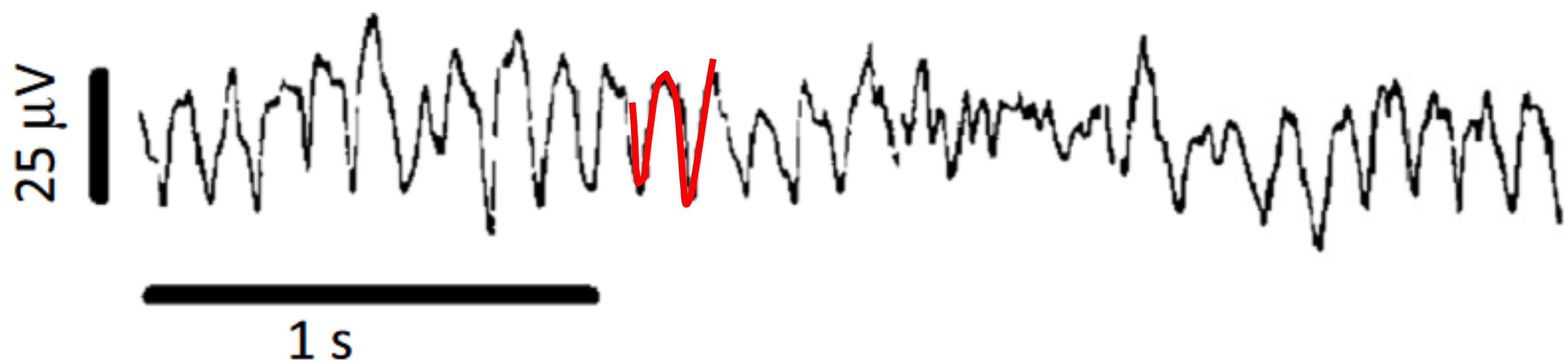
CFC: High frequency bursts coupled with slow waves

[T. Dupré la Tour, L. Tallot, L. Grabot, V. Doyère, V. van Wassenhove, Y. Grenier, A. Gramfort, (2017) PLOS Computational biology]



Neural signals exhibit diverse and complex morphologies

μ rhythm



CFC: High frequency bursts coupled with slow waves

[T. Dupré la Tour, L. Tallot, L. Grabot, V. Doyère, V. van Wassenhove, Y. Grenier, A. Gramfort, (2017) PLOS Computational biology]

Outline

3 routes to Representation Learning

Outline

3 routes to Representation Learning

The supervised

Outline

3 routes to Representation Learning

The supervised

The unsupervised

Outline

3 routes to Representation Learning

The supervised	The unsupervised	The semi-supervised

1

Deep supervised learning on sleep EEG data

A deep learning architecture for temporal sleep stage classification using multivariate and multimodal time series, S. Chambon, M. Galtier, P. Arnal, G. Wainrib, A. Gramfort (2018), IEEE Trans. Neural Systems and Rehabilitation Engineering

DOSED: a deep learning approach to detect multiple sleep micro-events in EEG signal, S. Chambon, V. Thorey, P. J. Arnal, E. Mignot, A. Gramfort. (2018), J. of Neuroscience Methods

The supervised way…

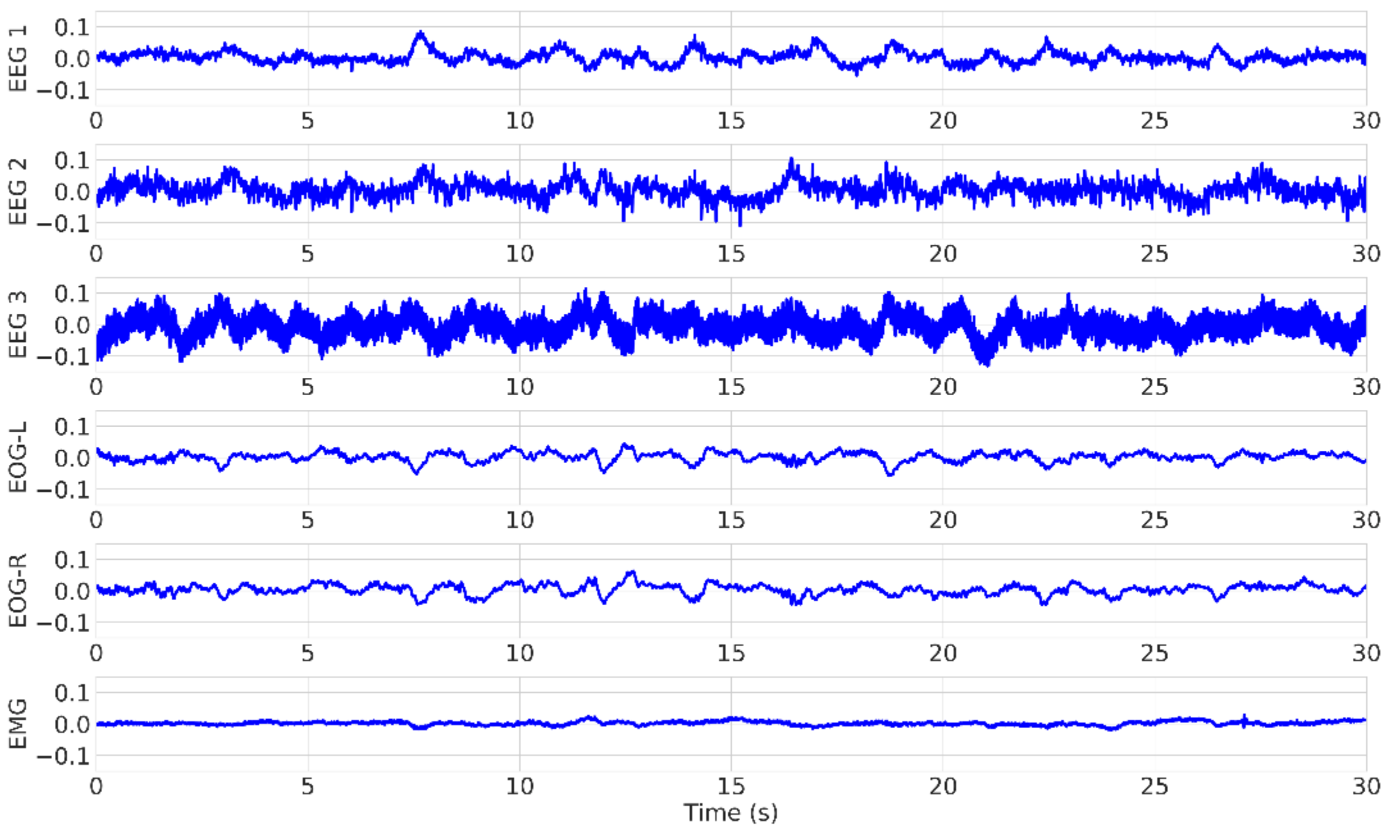
Polysomnography (PSG)

- Clinical exam
- Electrophysiological signals

Electro-encephalography (EEG)

Electro-oculography (EOG)

Electro-myography (EMG)



Polysomnography (PSG)

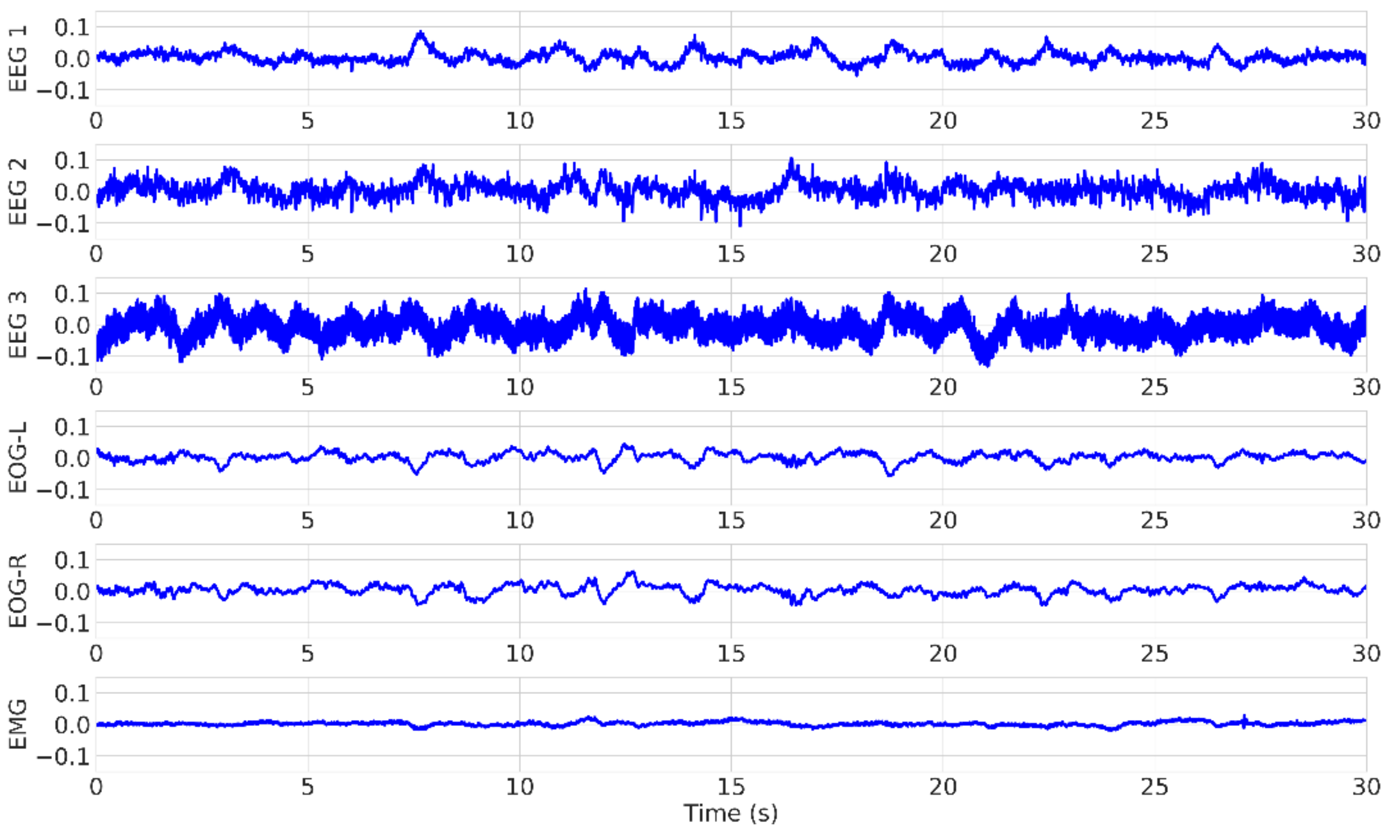
- Clinical exam
- Electrophysiological signals

Routinely annotated by sleep experts

Electro-encephalography (EEG)

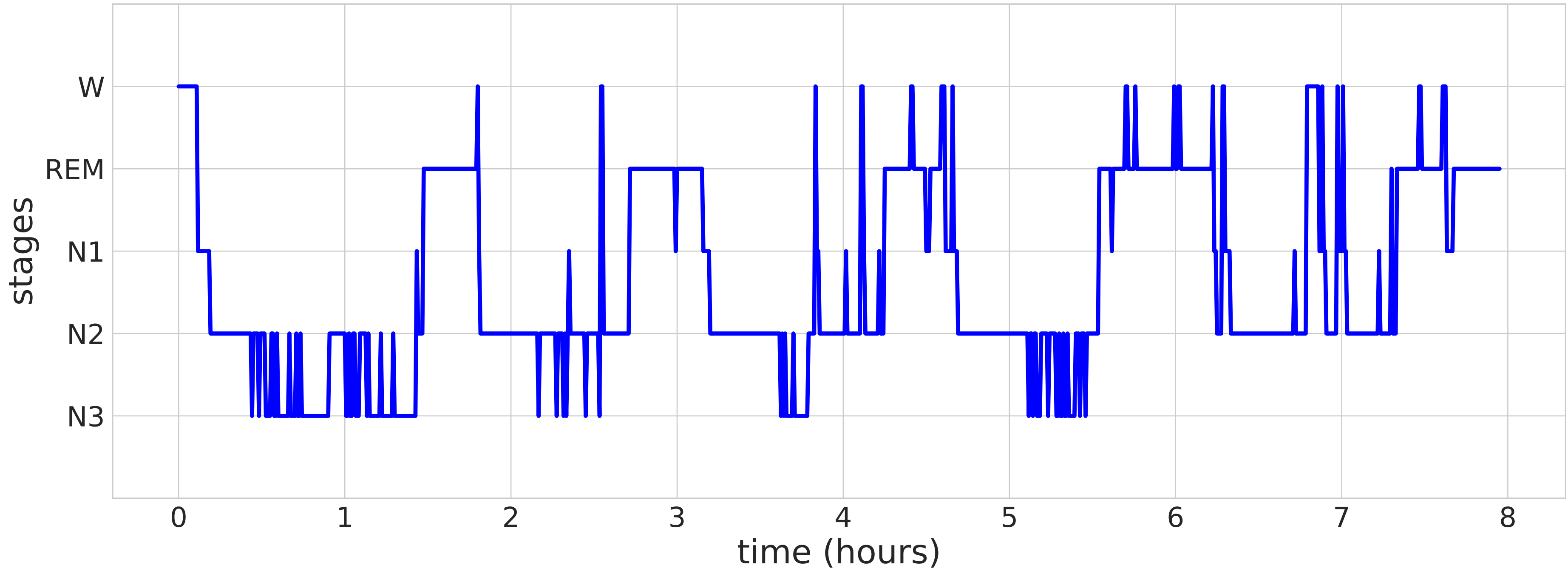
Electro-oculography (EOG)

Electro-myography (EMG)

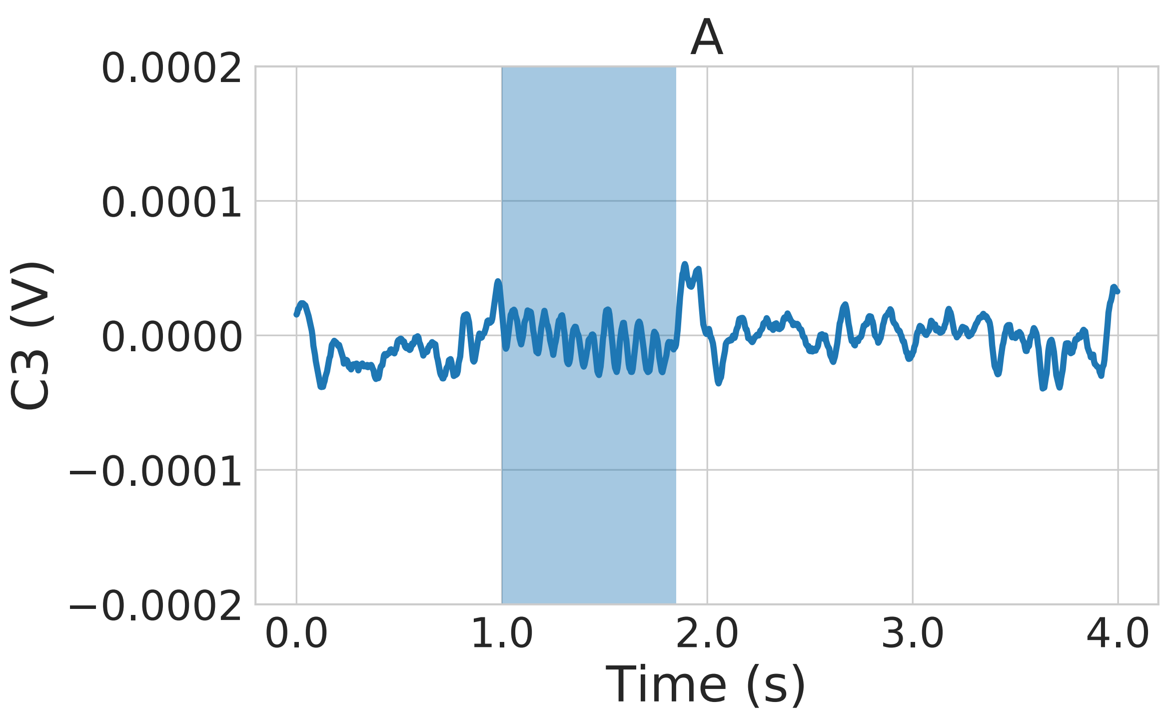


2 types of annotations

Hypnogram of sleep stages

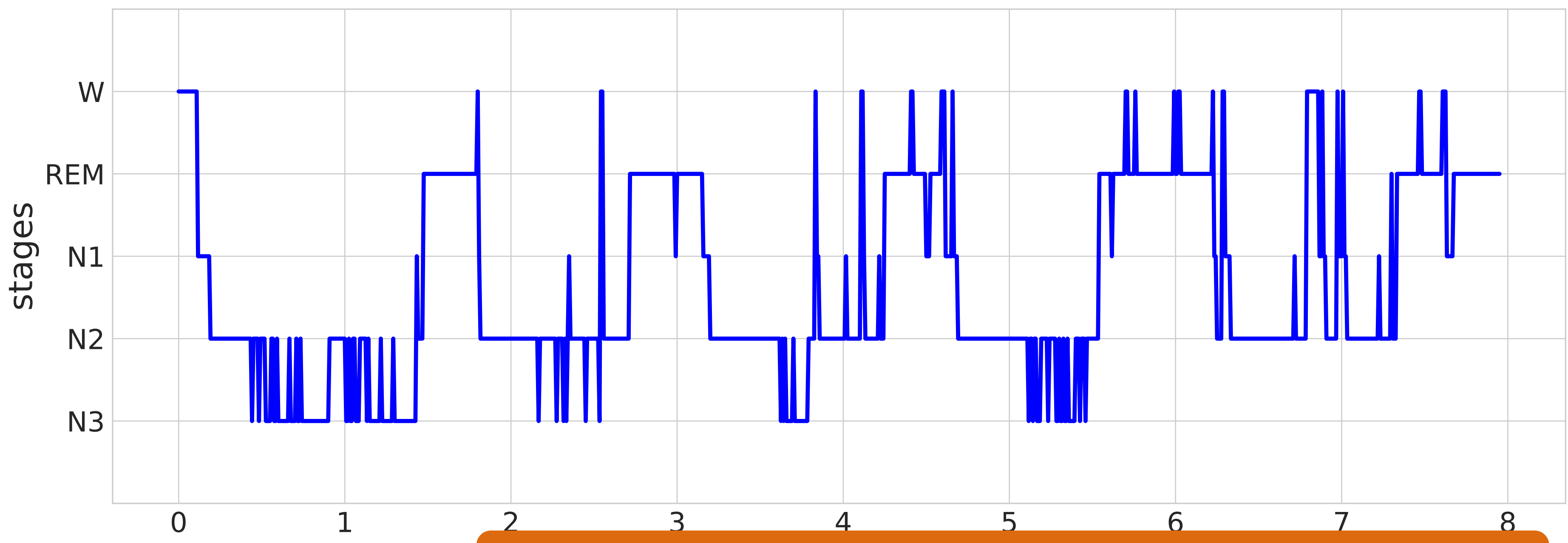


Micro-events: Spindles, K-complex etc.



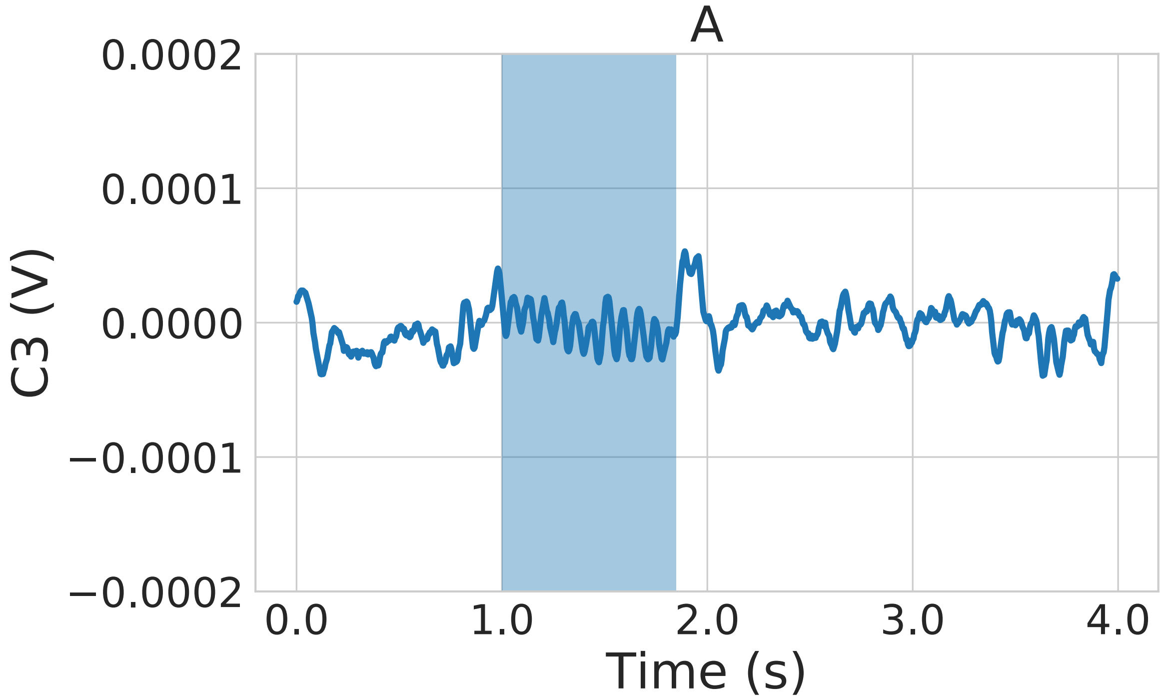
2 types of annotations

Hypnogram of sleep stages



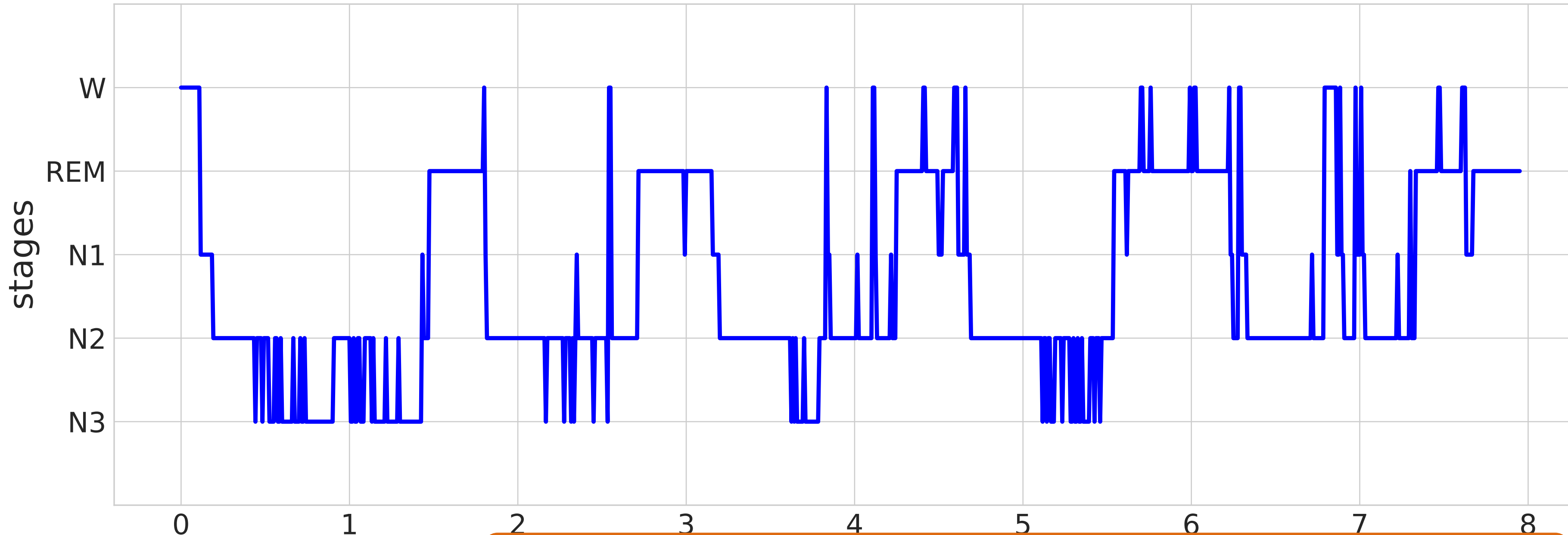
Classification problem

Micro-events: Spindles, K-complex etc.



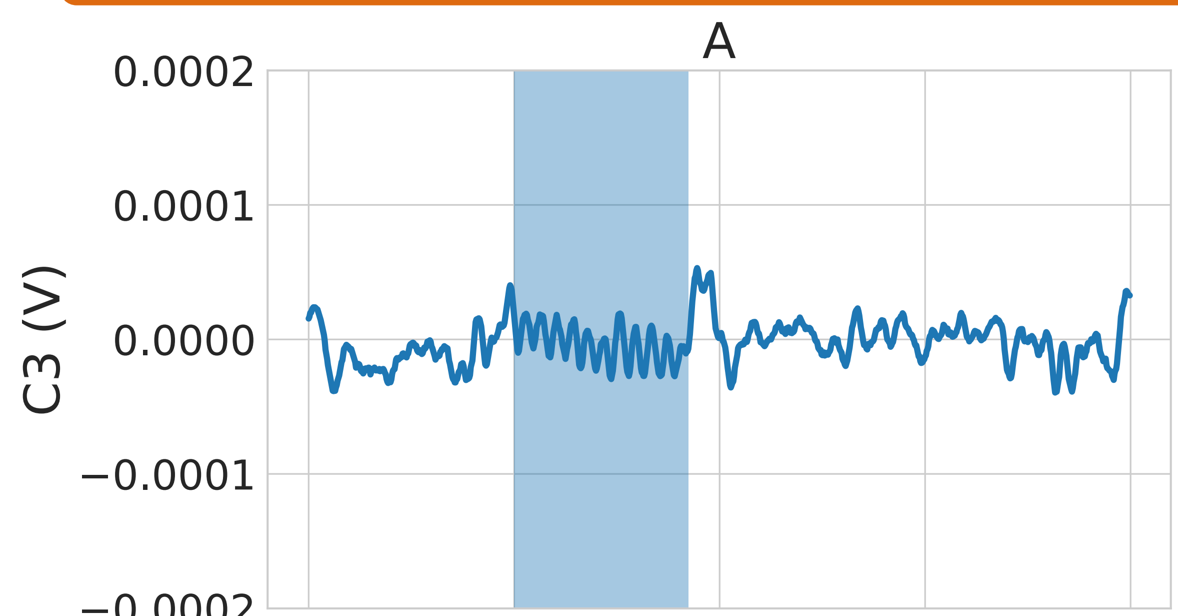
2 types of annotations

Hypnogram of sleep stages



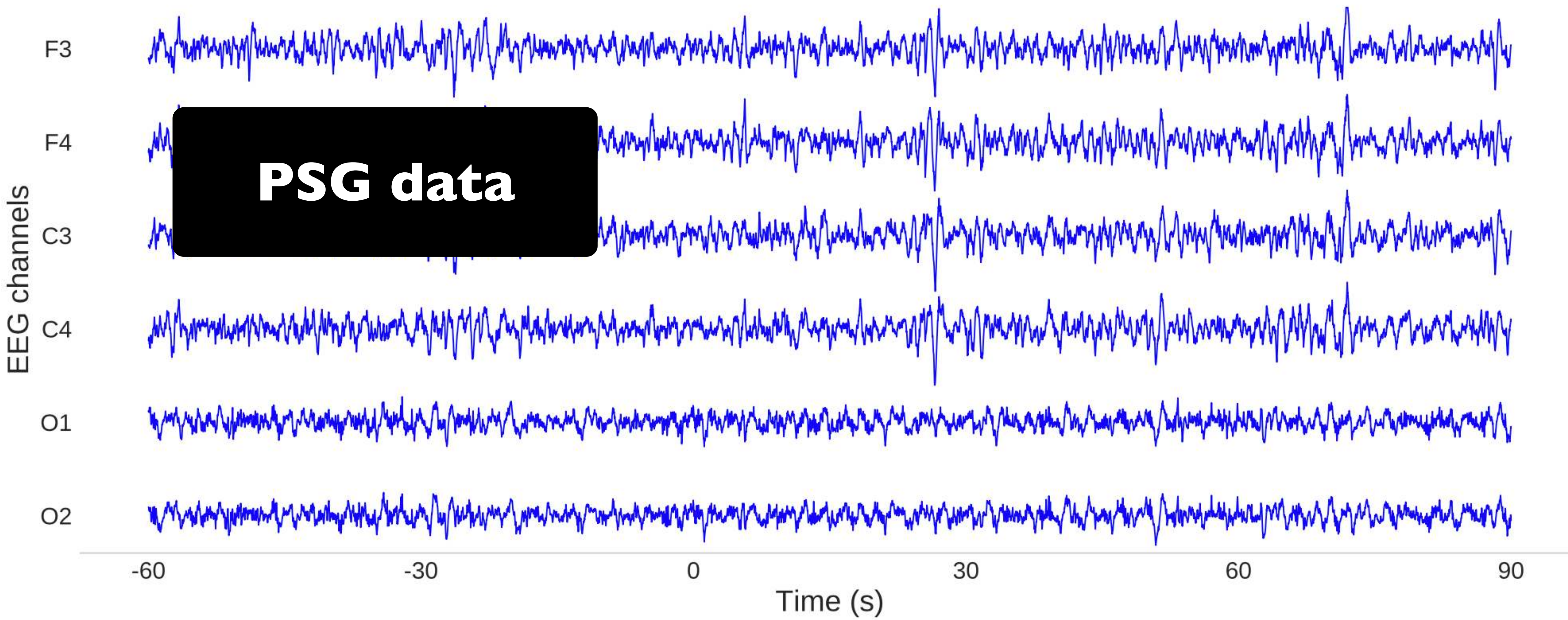
Classification problem

Micro-events: Spindles, K-complex etc.

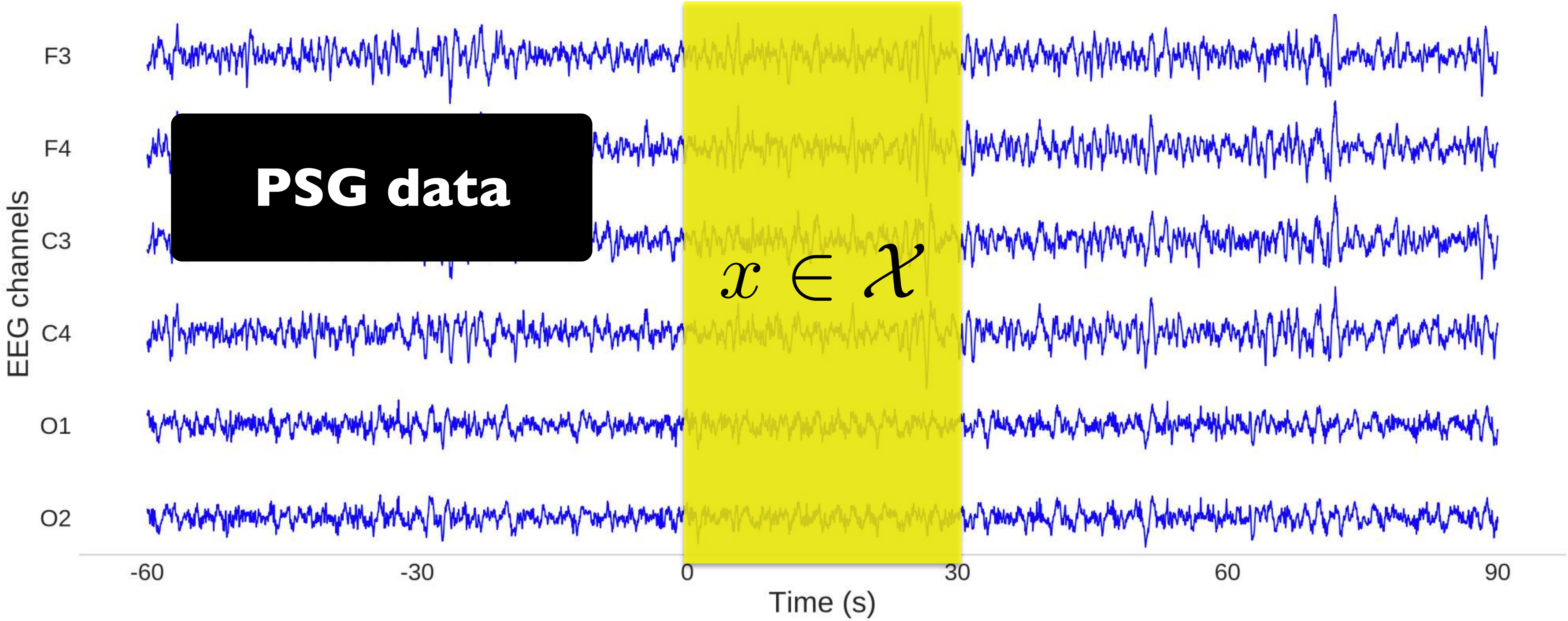


Joint detection and classification problem

Objective

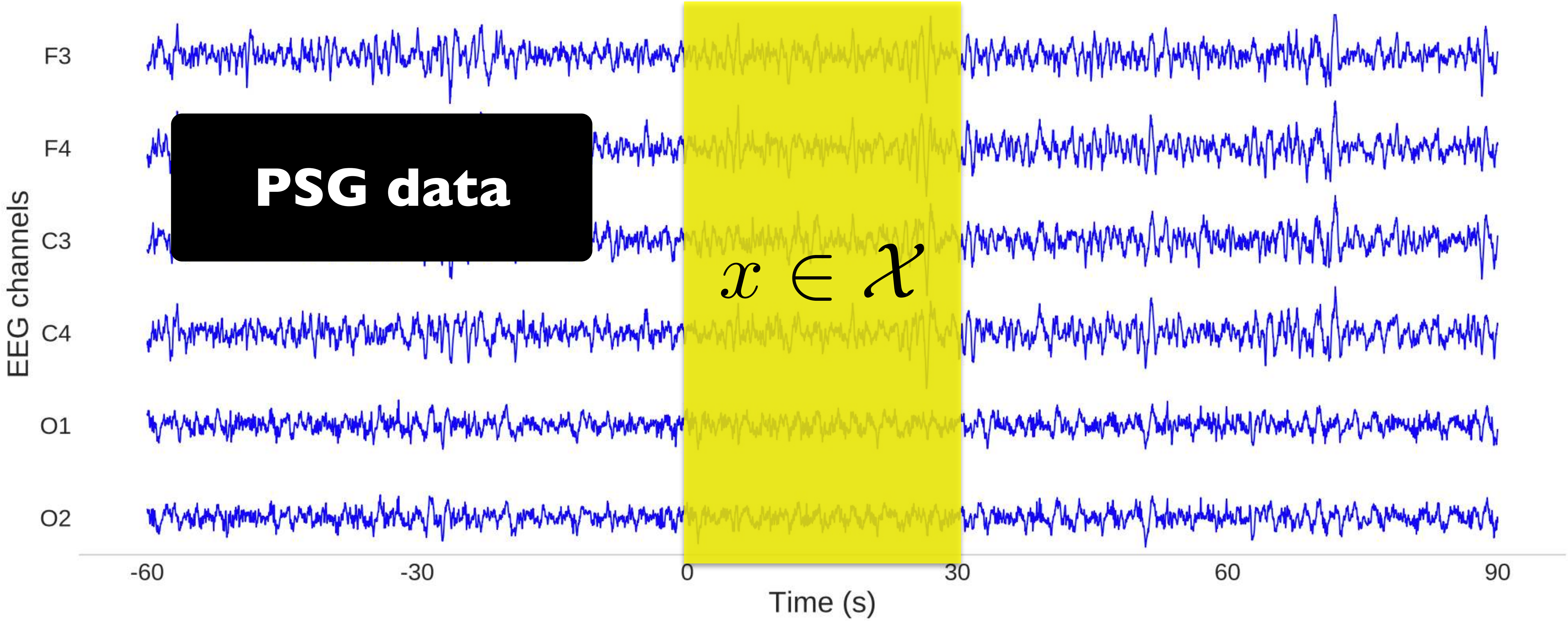


Objective



PSG data
$x \in \mathcal{X}$
F3
F4
C3
C4
O1
O2
EEG channels
-60
-30
0
30
60
90
Time (s)

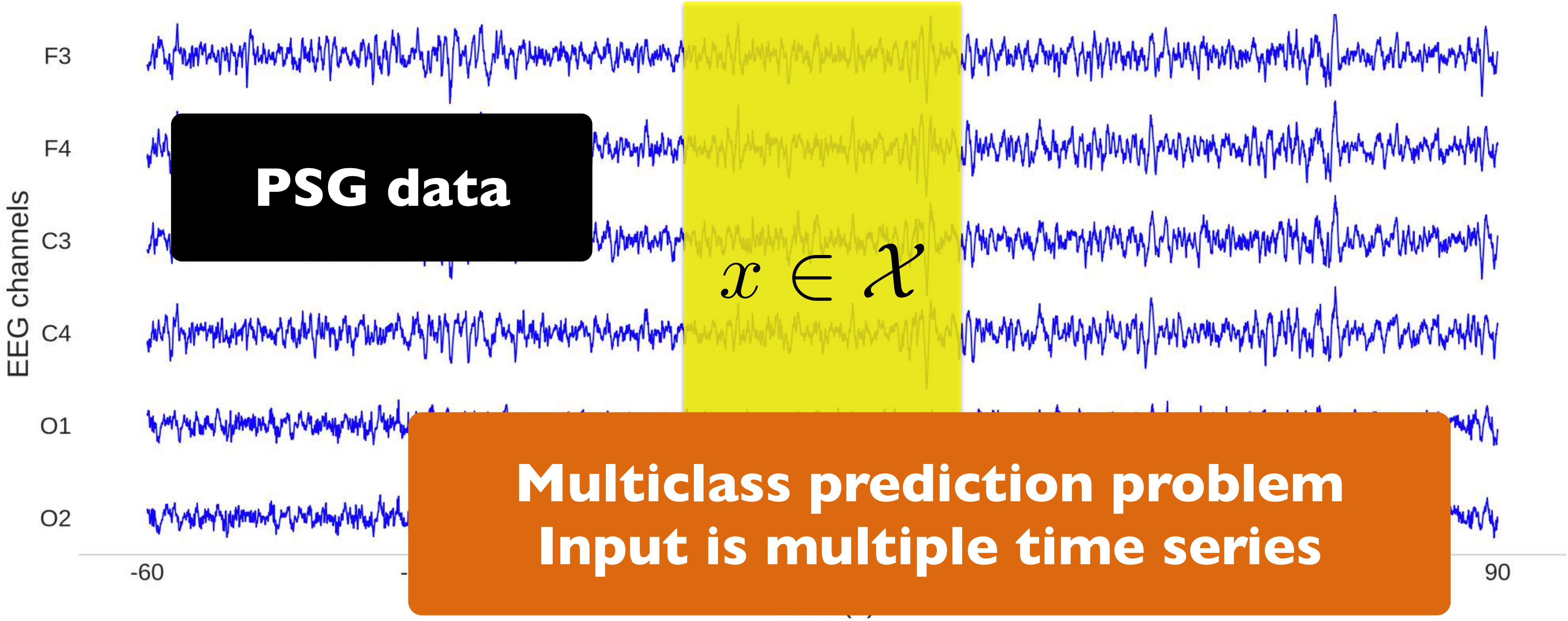
Objective



Learn: $\hat{f} : \mathcal{X} \rightarrow \mathcal{Y}$

$$\mathcal{Y} = \{\text{Awake, REM, Stage 1, Stage 2, etc.}\}$$

Objective



PSG data

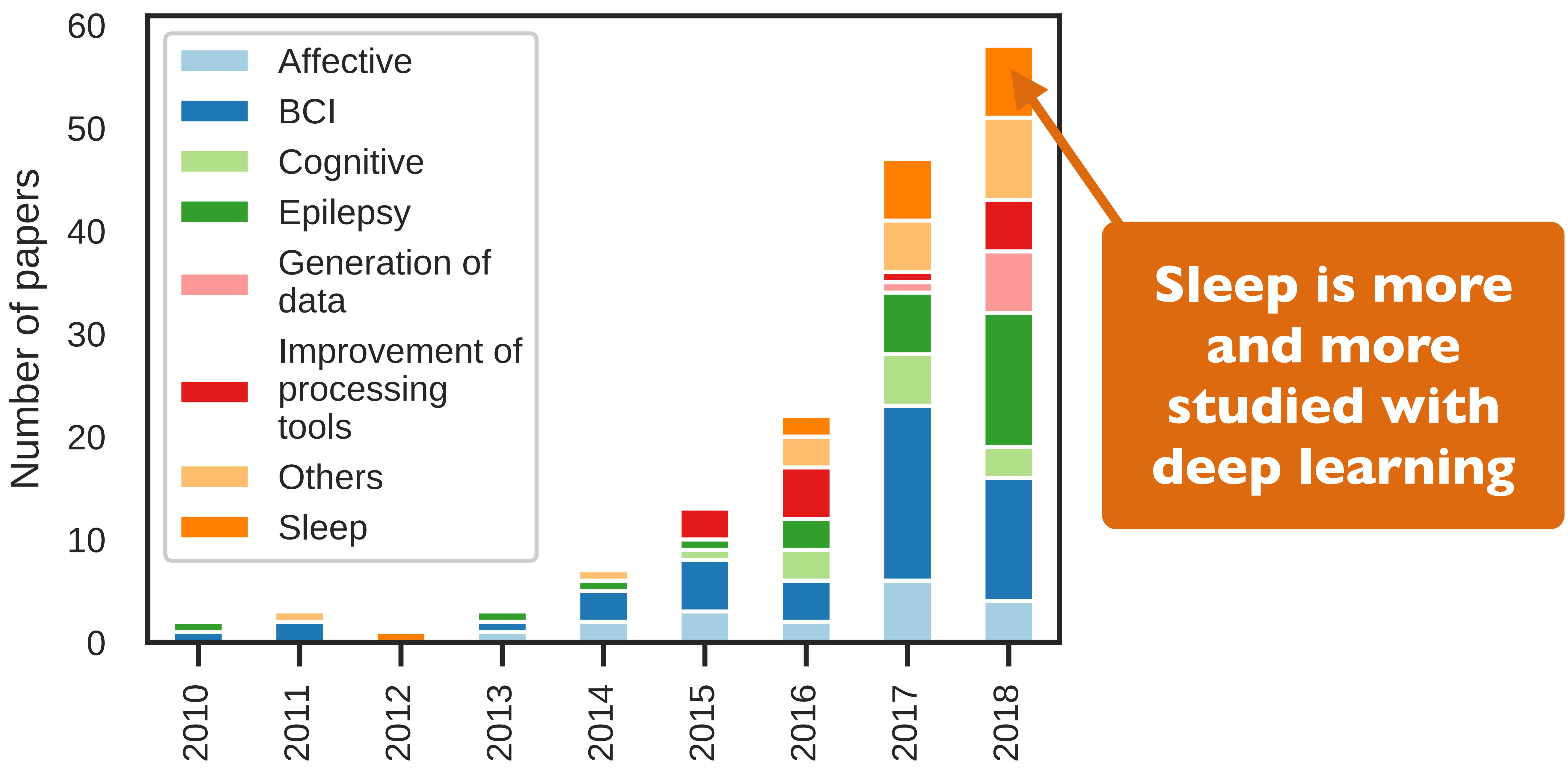
$x \in \mathcal{X}$

Multiclass prediction problem
Input is multiple time series

Learn: $\hat{f} : \mathcal{X} \rightarrow \mathcal{Y}$

$$\mathcal{Y} = \{\text{Awake, REM, Stage 1, Stage 2, etc.}\}$$

An old yet timely problem



[Deep learning-based electroencephalography analysis: a systematic review
Roy, Y., Banville, H., Albuquerque, I., Gramfort, A., Falk, T. and Faubert, J. (2019)
Journal of Neural Engineering 16: (051001).]

Learn from the multimodal and multivariate signals

Insight on time processing

- Convolutional Neural Net (CNN)

Spectral content is discriminant

[*A deep learning architecture for temporal sleep stage classification using multivariate and multimodal time series*. S. Chambon, M. N. Galtier, P. J. Arnal, G. Wainrib, A. Gramfort. IEEE TSRNE 2018]

Learn from the multimodal and multivariate signals

Insight on time processing

- Convolutional Neural Net (CNN)

Spectral content is discriminant

Insight on spatial structure

Denoising

- Multivariate signals: spatial filtering (like ICA would do)
- Multimodal inputs: different pipelines for EEG, EOG and EMG

Spectral content differs between modalities

[*A deep learning architecture for temporal sleep stage classification using multivariate and multimodal time series*. S. Chambon, M. N. Galtier, P. J. Arnal, G. Wainrib, A. Gramfort. IEEE TSRNE 2018]

Learn from the multimodal and multivariate signals

Insight on time processing

- Convolutional Neural Net (CNN)

Spectral content is discriminant

Insight on spatial structure

- Multivariate signals: spatial filtering (like ICA would do)
- Multimodal inputs: different pipelines for EEG, EOG and EMG

Denoising

Spectral content differs between modalities

Insight on sequence of states

- Concatenate features of neighboring samples

non i.i.d. samples

[*A deep learning architecture for temporal sleep stage classification using multivariate and multimodal time series*. S. Chambon, M. N. Galtier, P. J. Arnal, G. Wainrib, A. Gramfort. IEEE TSRNE 2018]

Problem formulation

Risk minimization:

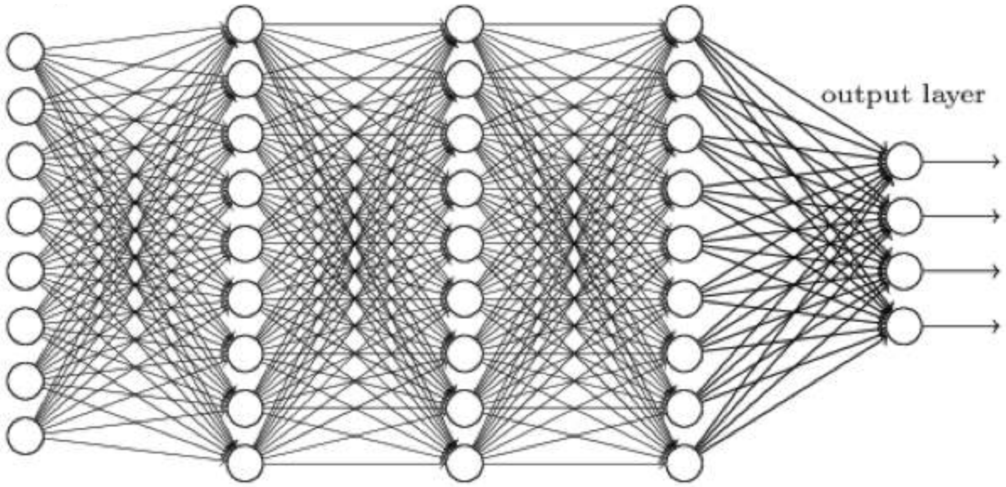
$$\hat{f} \in \operatorname{argmin}_{f \in \mathcal{F}} \mathbb{E}_{x,y \in \mathcal{X} \times \mathcal{Y}} \left[\ell(y, f(x)) \right]$$

Problem formulation

Risk minimization:

$$\hat{f} \in \operatorname{argmin}_{f \in \mathcal{F}} \mathbb{E}_{x,y \in \mathcal{X} \times \mathcal{Y}} \left[\ell(y, f(x)) \right]$$

- $\mathcal{F}$: a class of models (neural networks architecture).

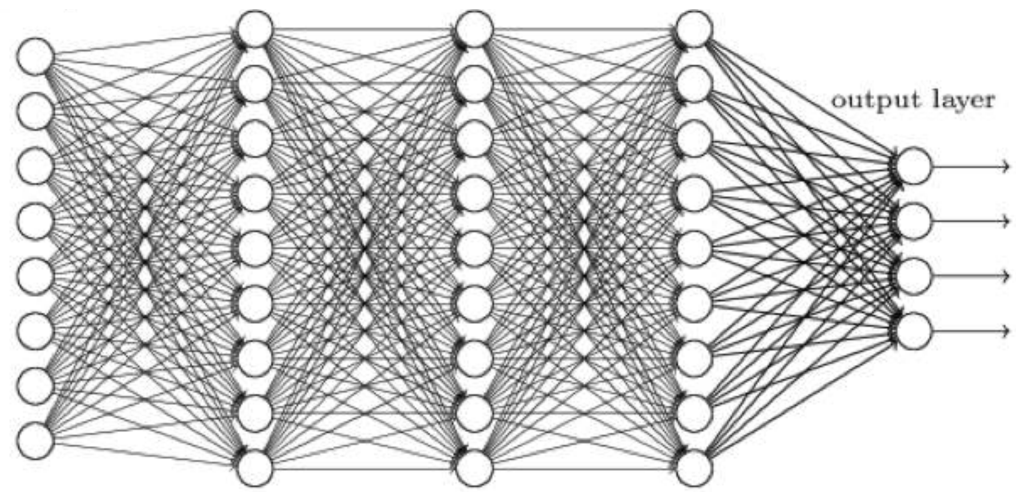


Problem formulation

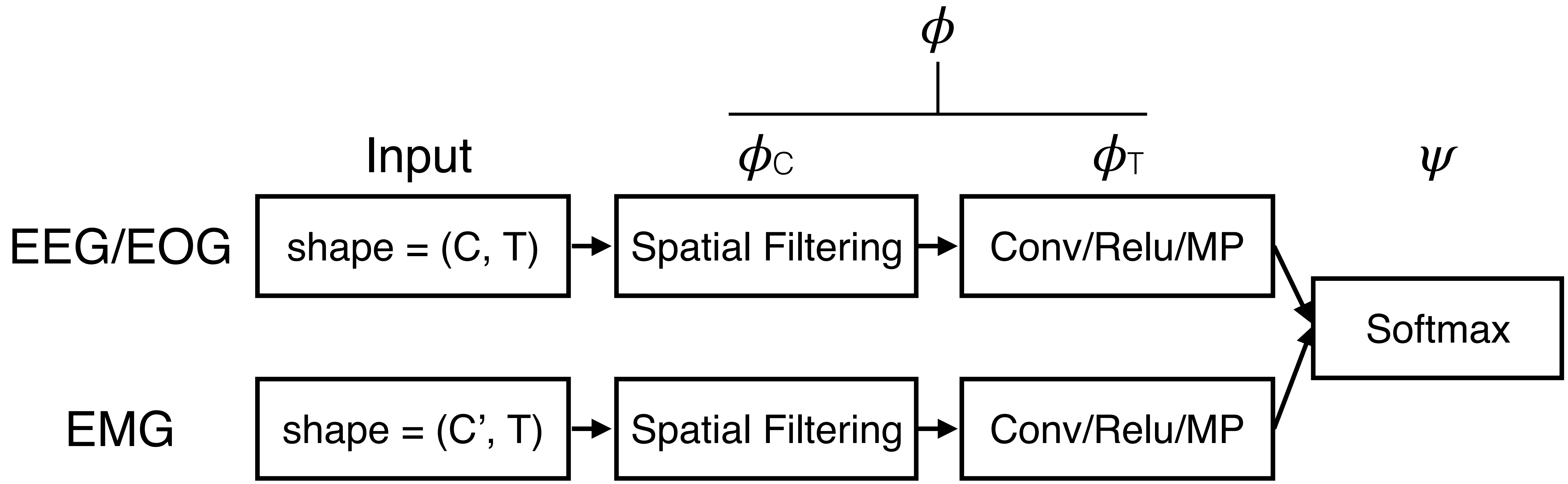
Risk minimization:

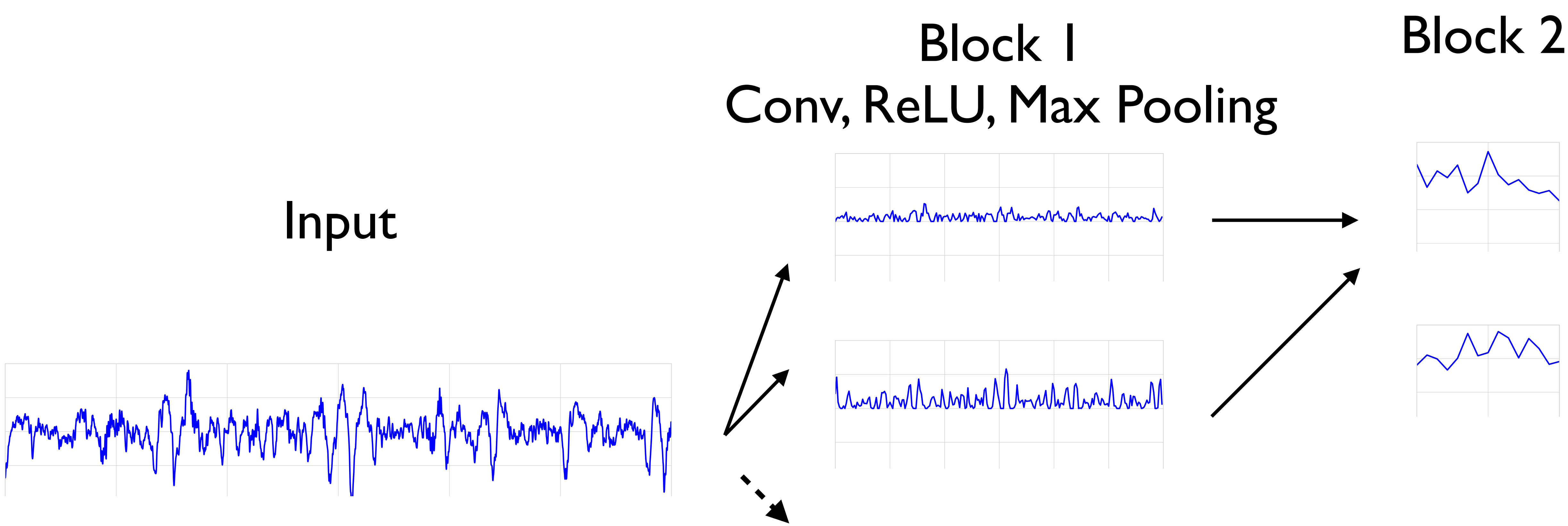
$$\hat{f} \in \operatorname{argmin}_{f \in \mathcal{F}} \mathbb{E}_{x,y \in \mathcal{X} \times \mathcal{Y}} [\ell(y, f(x))]$$

- $\mathcal{F}$: a class of models (neural networks architecture).

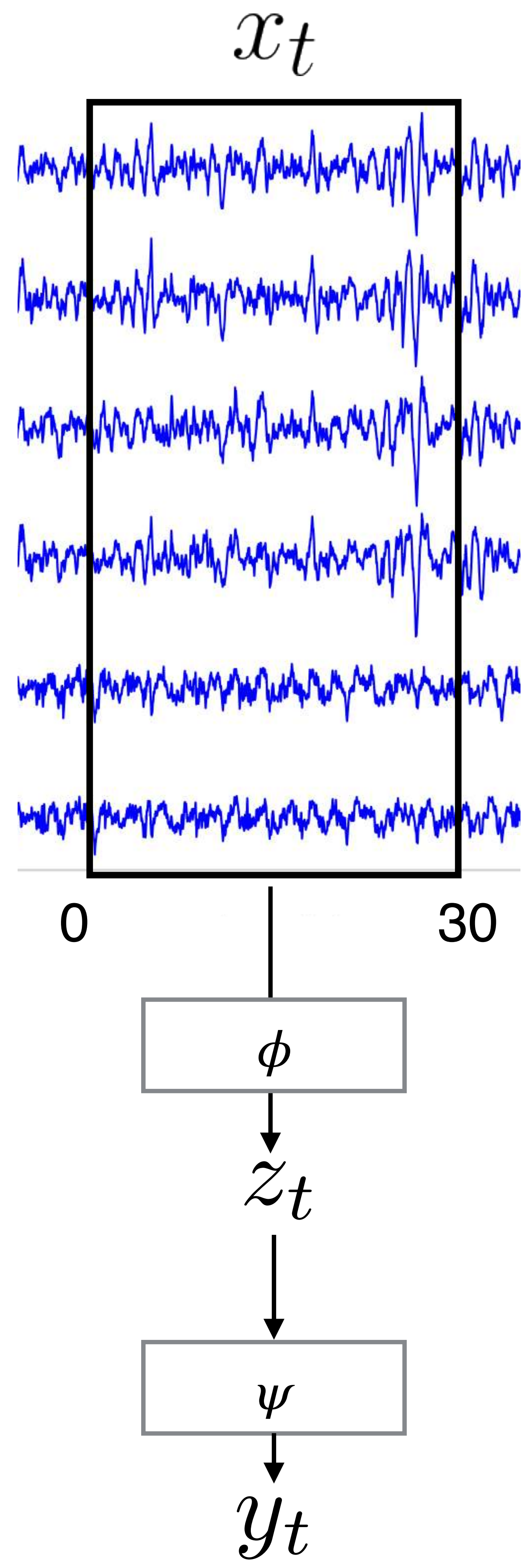


- ℓ : loss. We use the categorical cross-entropy.
- Minimized on a training set using backpropagation and online stochastic gradient descent.

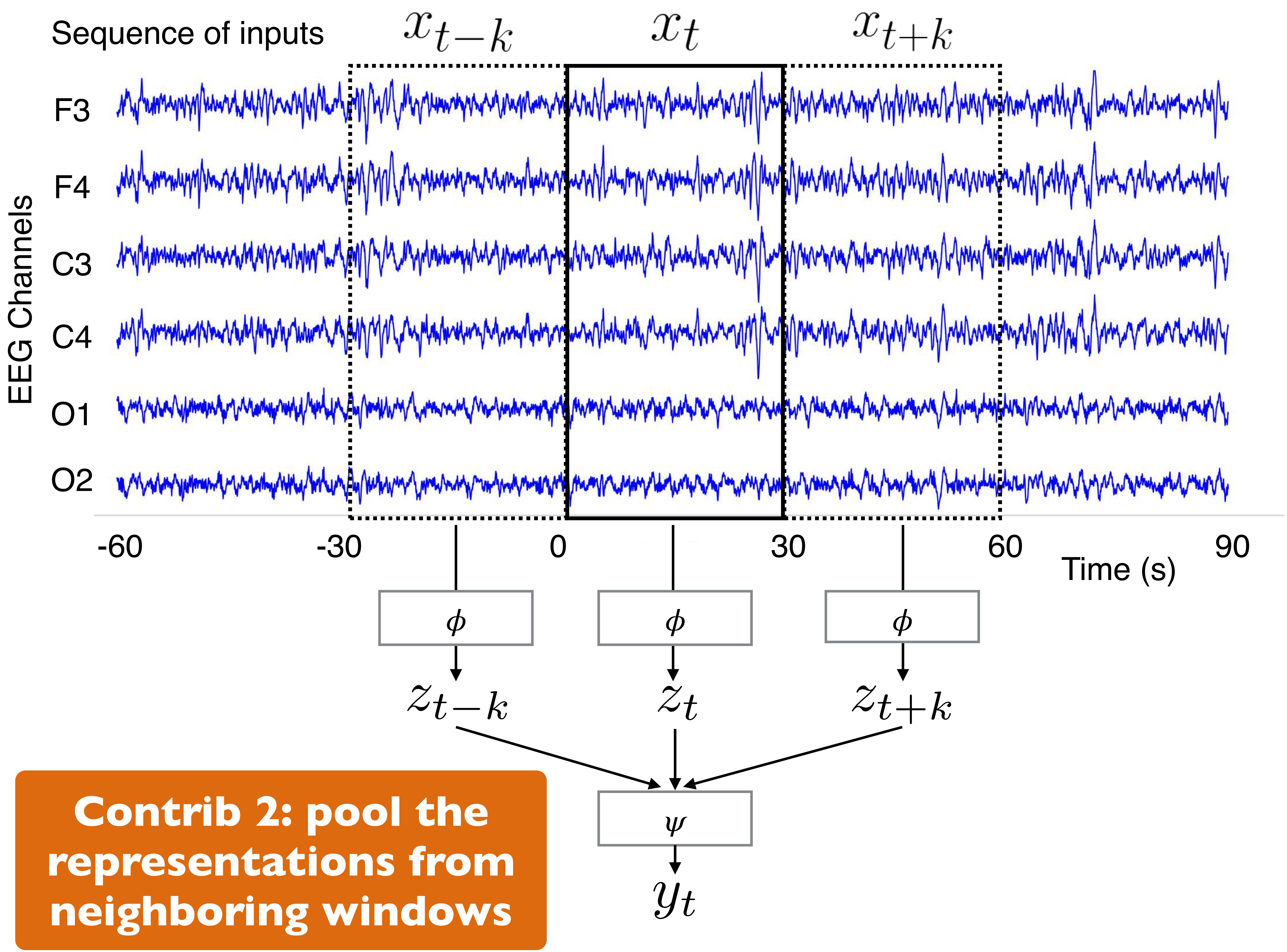




Contrib 1: one trunk per modality before pooling



x_t
0
30
ϕ
z_t
ψ
y_t



Sequence of inputs
x_{t-k}
x_t
x_{t+k}
F3
F4
C3
C4
O1
O2
EEG Channels
-60
-30
0
30
60
90
Time (s)
ϕ
ϕ
ϕ
z_{t-k}
z_t
z_{t+k}
ψ
y_t
Contrib 2: pool the representations from neighboring windows

Experiments…

Experimental setup

- **Data**
  - MASS - session 3: *O'Reilly et al. 2014*
  - http://www.ceams-carsm.ca/en/MASS
  - 61 records, ~10h / record
  - Total number of samples about 60,000 (like for MNIST)

- **Preprocessing**
  - lowpass filtering (30Hz)
  - downsampling 128Hz

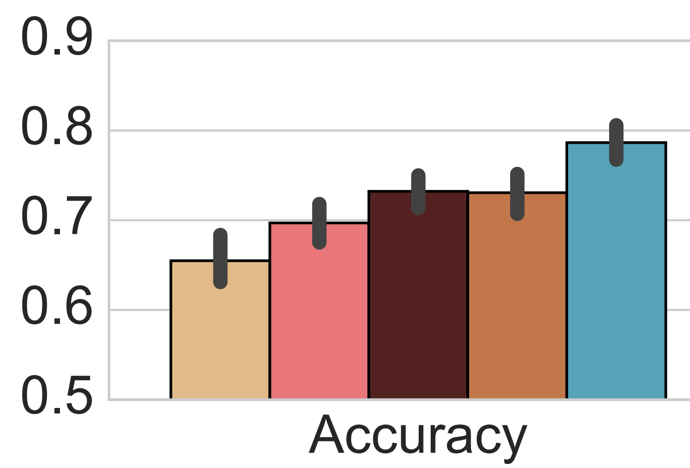
Experimental setup

- **Cross validation**
  - 5 splits
  - 41 records for training, 10 for validation, 10 for testing

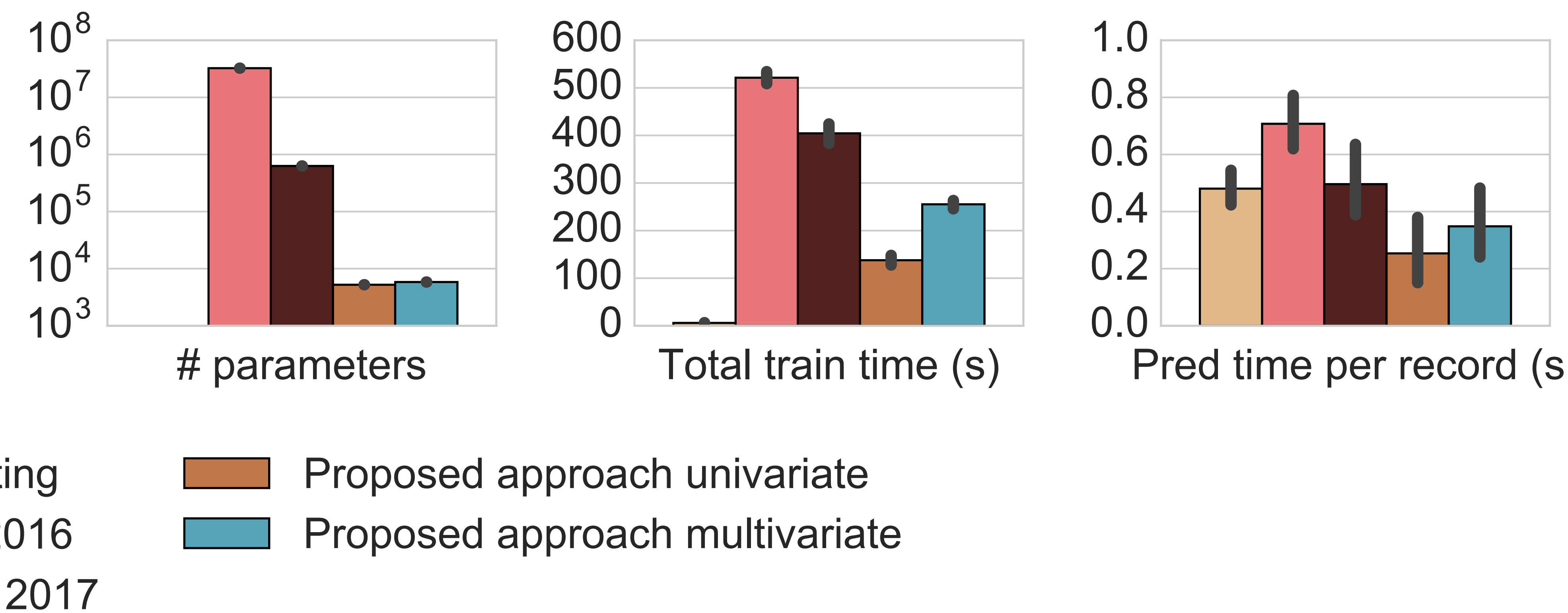
- **Baselines**
  - Gradient boosting: Hand crafted features (mean, variance… relative power in frequency bands) *[Lajnev et al. 2015]*
  - *[Tsinalis et al. 2016]*: deep convolutional network
  - *[Supratak et al. 2017]*: deep convolutional network processing low and high frequency contents specifically

Performance vs SoTA

Performance:

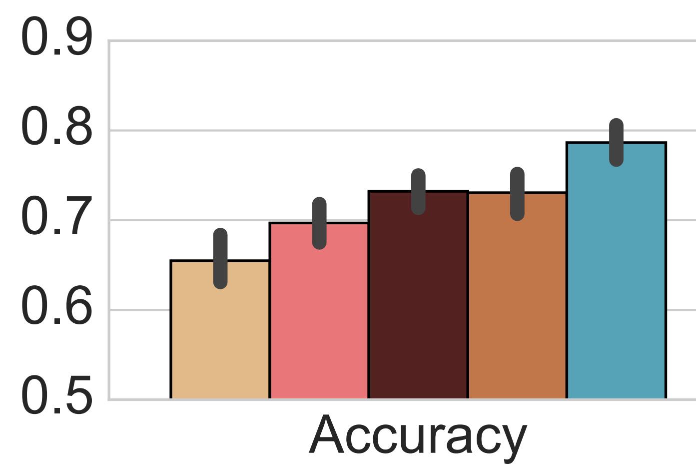


Computational efficiency:

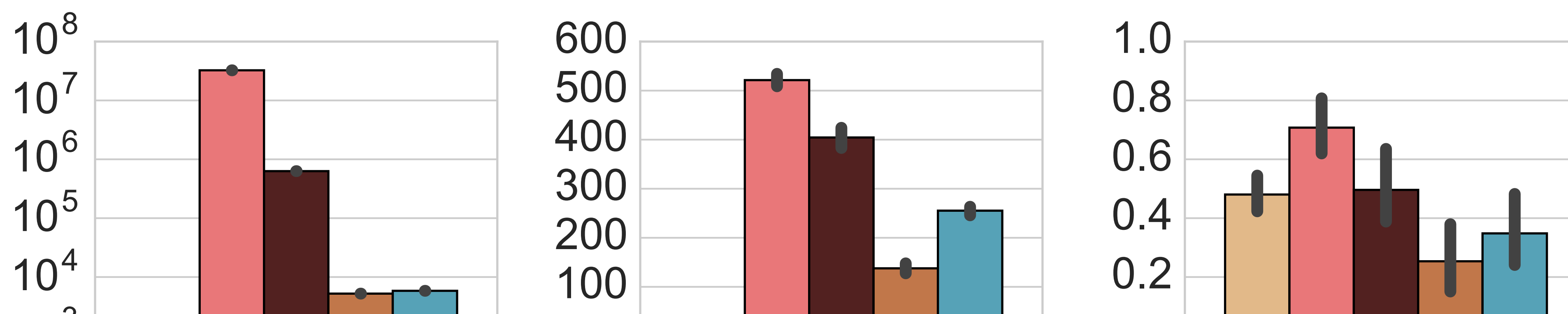


Performance vs SoTA

Performance:



Computational efficiency:



- **Networks outperform hand crafted features with gradient boosting**
- **Combining all modalities helps (multivariate)**
- **SoTA is obtained with less parameters (fast)**

“Opening” the black box

Confusion matrix:

True labels \ Predicted Labels	W	N1	N2	N3	REM
W	0.85	0.11	0.01	0.00	0.03
N1	0.11	0.52	0.10	0.00	0.27
N2	0.01	0.10	0.77	0.09	0.04
N3	0.00	0.00	0.09	0.91	0.00
REM	0.02	0.14	0.01	0.00	0.83

N1 is hard

"Opening" the black box

Confusion matrix:

True labels \ Predicted Labels

	W	N1	N2	N3	REM
W	0.85	0.11	0.01	0.00	0.03
N1	0.11	0.52	0.10	0.00	0.27
N2	0.01	0.10	0.77	0.09	0.04
N3	0.00	0.00	0.09	0.91	0.00
REM	0.02	0.14	0.01	0.00	0.83

N1 is hard

"Hiding" frequencies from the network:

Unfiltered (True labels \ Predicted labels)

	W	N1	N2	N3	REM
W	0.83	0.14	0.01	0.00	0.02
N1	0.08	0.62	0.08	0.00	0.22
N2	0.00	0.14	0.75	0.07	0.04
N3	0.00	0.00	0.11	0.89	0.00
REM	0.01	0.14	0.00	0.00	0.84

δ (0.5, 4.5) Hz (True labels \ Predicted labels)

	W	N1	N2	N3	REM
W	0.04	0.00	0.11	0.84	0.00
N1	0.00	0.00	0.37	0.63	0.00
N2	0.00	0.00	0.16	0.84	0.00
N3	0.00	0.00	0.00	1.00	0.00
REM	0.00	0.00	0.43	0.57	0.00

θ (4.5, 8.5) Hz (True labels \ Predicted labels)

	W	N1	N2	N3	REM
W	0.69	0.18	0.01	0.00	0.11
N1	0.07	0.51	0.17	0.00	0.25
N2	0.01	0.37	0.54	0.00	0.08
N3	0.00	0.13	0.87	0.00	0.00
REM	0.02	0.27	0.02	0.00	0.69

α (8.5, 11.5) Hz (True labels \ Predicted labels)

	W	N1	N2	N3	REM
W	1.00	0.00	0.00	0.00	0.00
N1	1.00	0.00	0.00	0.00	0.00
N2	0.99	0.00	0.01	0.00	0.00
N3	1.00	0.00	0.00	0.00	0.00
REM	1.00	0.00	0.00	0.00	0.00

σ (11.5, 15.5) Hz (True labels \ Predicted labels)

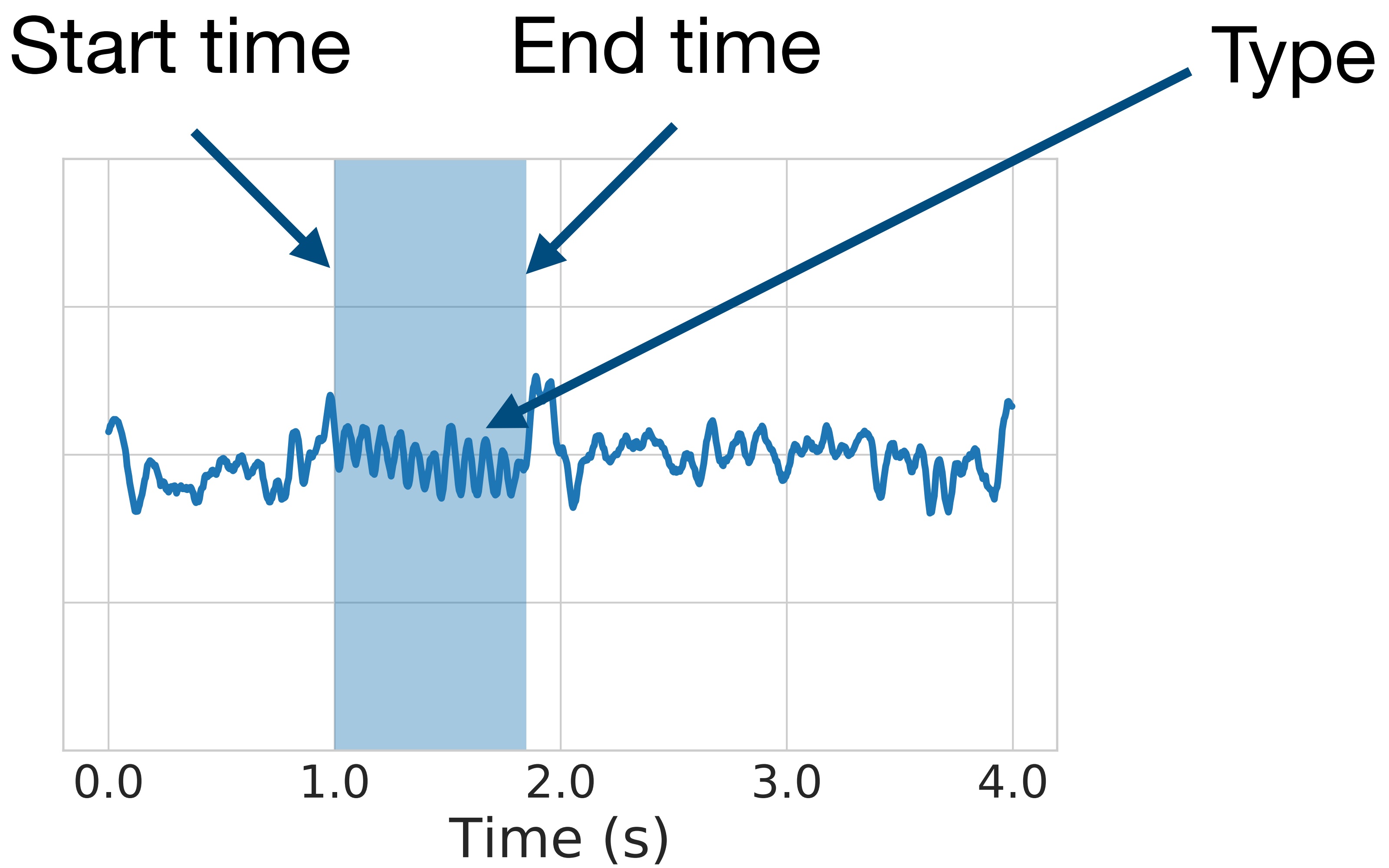
	W	N1	N2	N3	REM
W	0.76	0.09	0.15	0.00	0.00
N1	0.51	0.29	0.20	0.00	0.00
N2	0.23	0.16	0.61	0.01	0.00
N3	0.66	0.06	0.28	0.00	0.00
REM	0.75	0.21	0.04	0.00	0.00

β (15.5, 30) Hz (True labels \ Predicted labels)

	W	N1	N2	N3	REM
W	1.00	0.00	0.00	0.00	0.00
N1	1.00	0.00	0.00	0.00	0.00
N2	1.00	0.00	0.00	0.00	0.00
N3	1.00	0.00	0.00	0.00	0.00
REM	1.00	0.00	0.00	0.00	0.00

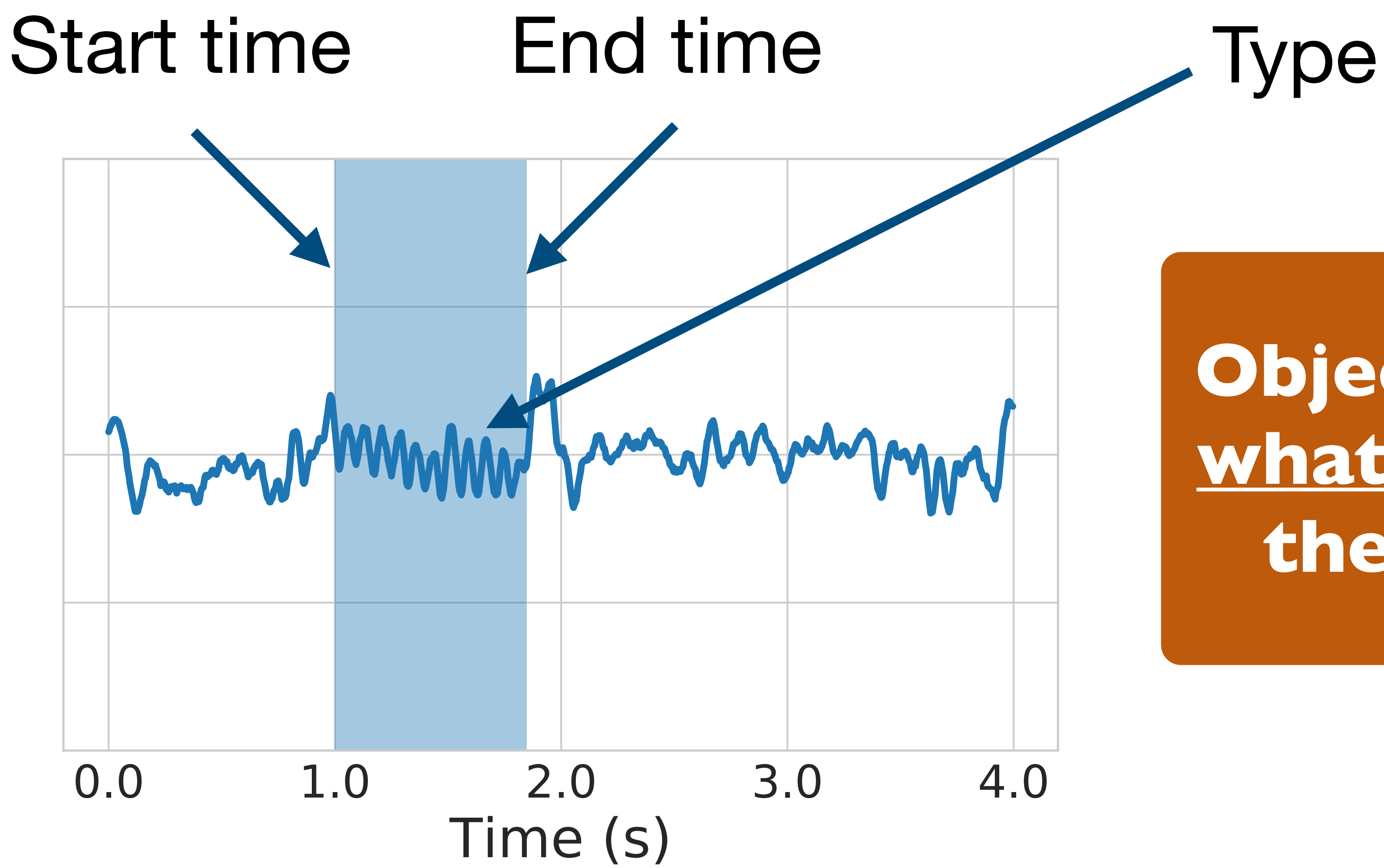
network has learnt scoring rules

Joint detection and classification of micro-events



[*A deep learning architecture to detect events in EEG signals during sleep.* S. Chambon, V. Thorey, P. J. Arnal, E. Mignot, A. Gramfort. MLSP 2018]
[*DOSED: a deep learning approach to detect multiple sleep micro-events in EEG signal,.* S. Chambon, V. Thorey, P. J. Arnal, E. Mignot, A. Gramfort. *Arxiv 2500163*]

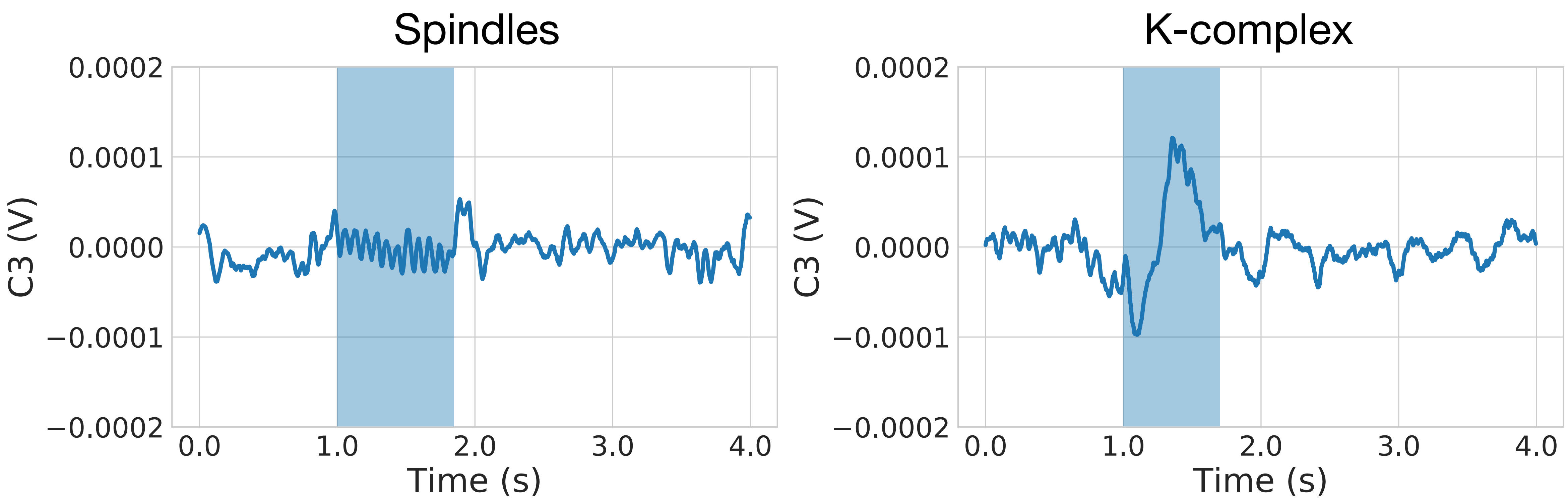
Joint detection and classification of micro-events



Objective: Predict <u>what</u> and <u>when</u> at the same time

[*A deep learning architecture to detect events in EEG signals during sleep.* S. Chambon, V. Thorey, P. J. Arnal, E. Mignot, A. Gramfort. MLSP 2018]
[*DOSED: a deep learning approach to detect multiple sleep micro-events in EEG signal,.* S. Chambon, V. Thorey, P. J. Arnal, E. Mignot, A. Gramfort. *Arxiv 2500163*]

Micro-events



Spindles and K-complexes in N2 stage

Why is it hard?

- Different types, durations, scales

Why should you care?

- Associated to certain sleep stages and sleep disorders

State of the art

- Signal processing:
  - Band-pass filtering + thresholding — [Ray et al. 2015, Wamsley et al. 2012, Wendt et al. 2012, Mölle et al. 2011, Nir et al. 2011, Ferrarelli et al. 2007]
  - Decomposition into components — [Parekh et al. 2017, Lajnef et al. 2017]
- Shallow learning algorithms:
  - Clustering — [Patti et al. 2017]
  - Binary classifier — [Patti et al. 2015, Lachner-Piza et al. 2018]

None of these works use deep learning

DOSED algorithm

1. Predict: type, start & end
2. Allows for multiple types of events
3. Detection and classification in one pass
4. Convolutional network

Inspired by: YOLO and SSD

[Redmon et al. 2016,
Liu et al. 2015]

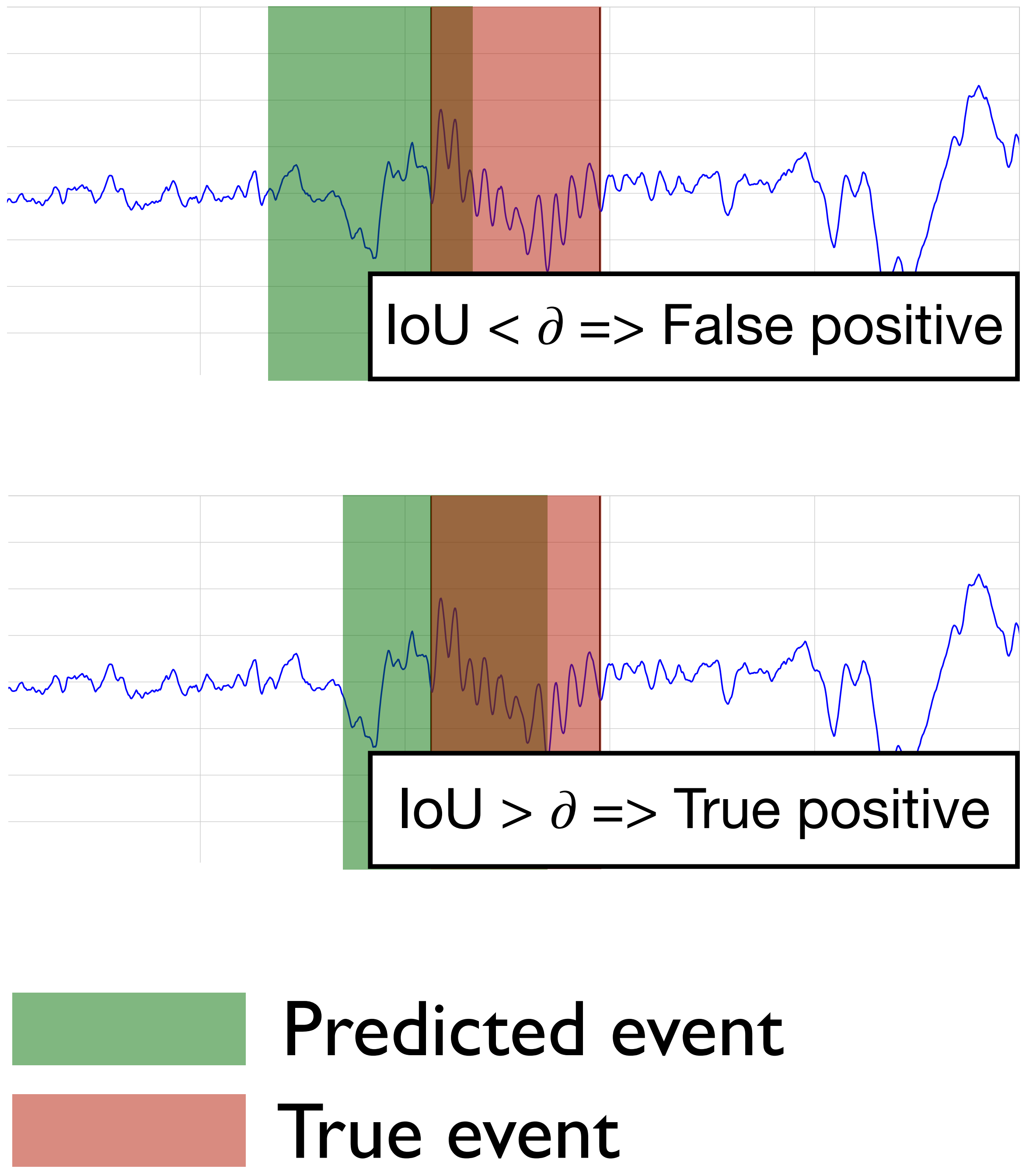
Evaluation

- **By event** metrics

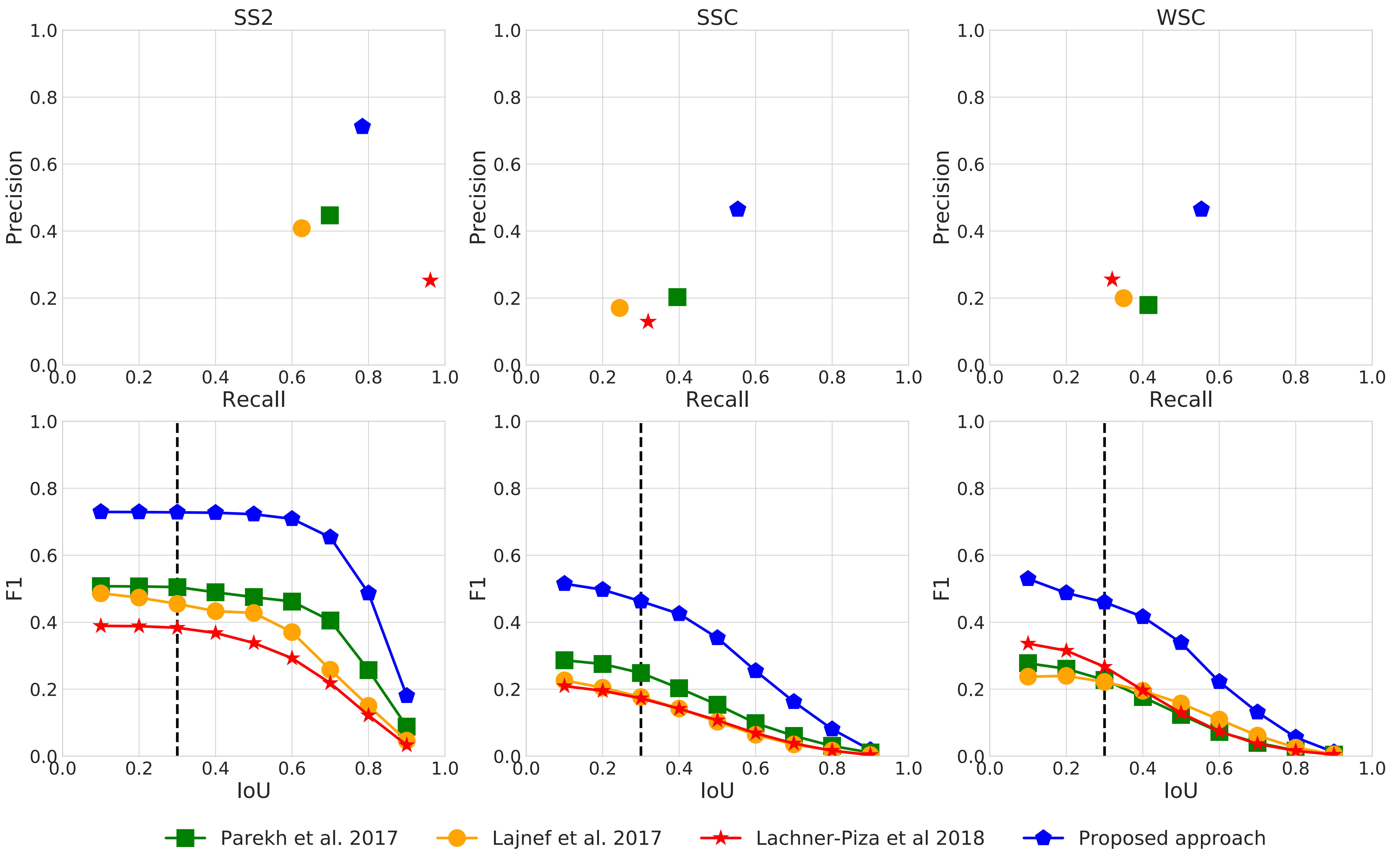
  [Warby et al. 2014]

- Based on Intersection over Union (IoU)

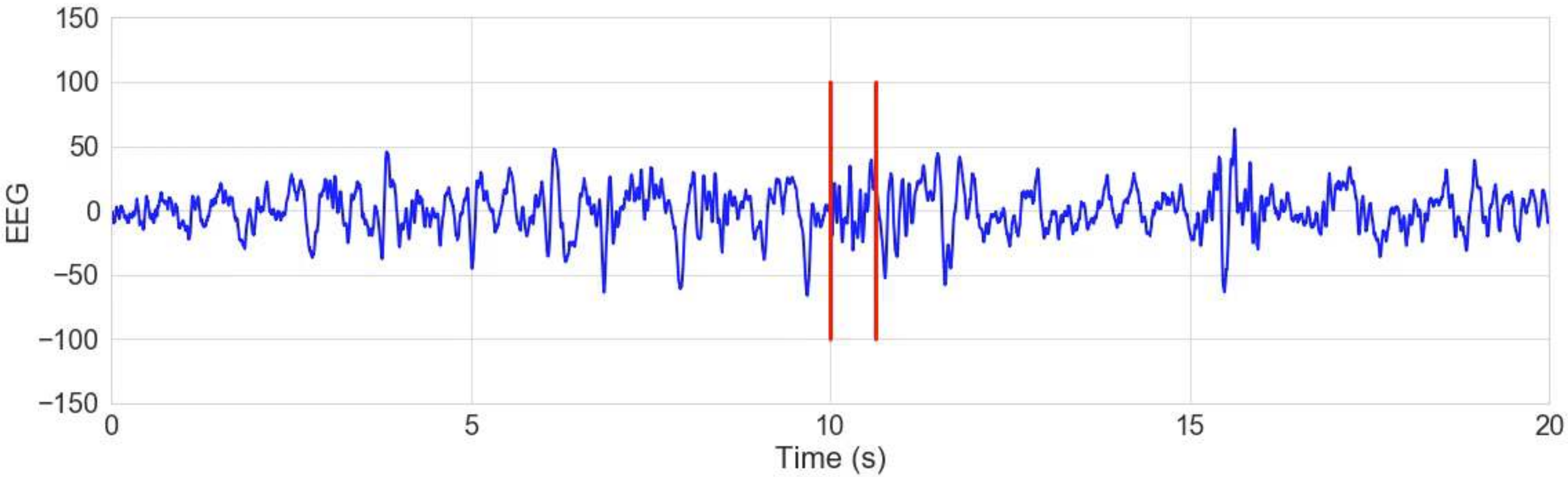
- Precision, Recall, F1



General benchmark



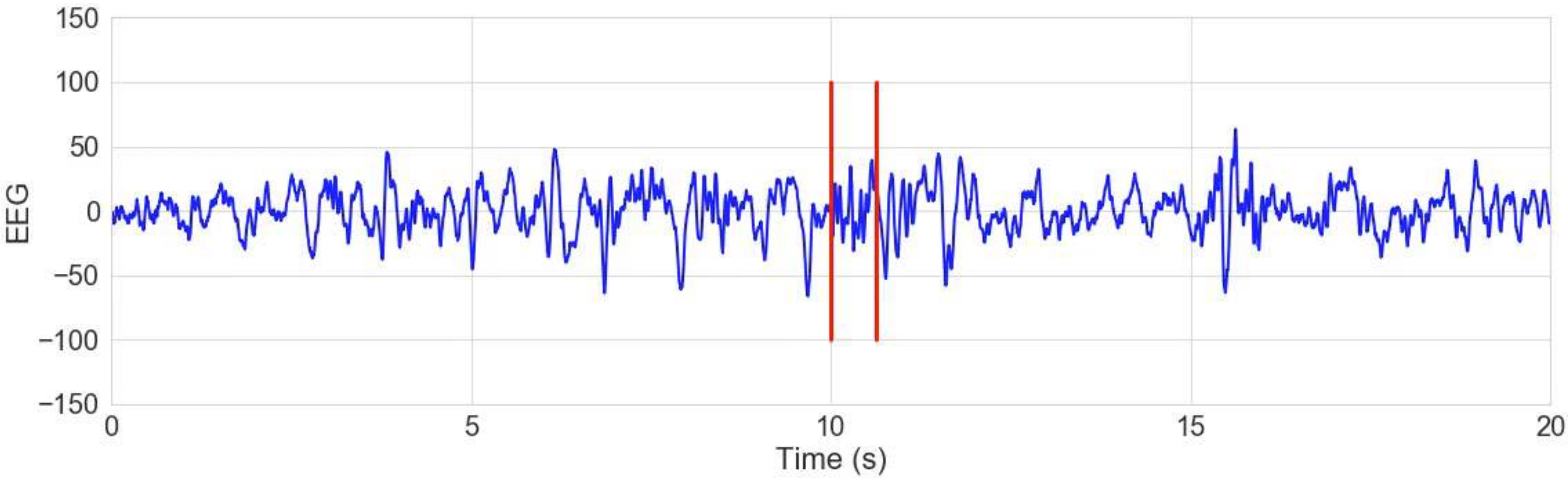
Illustration



True event

Predicted event

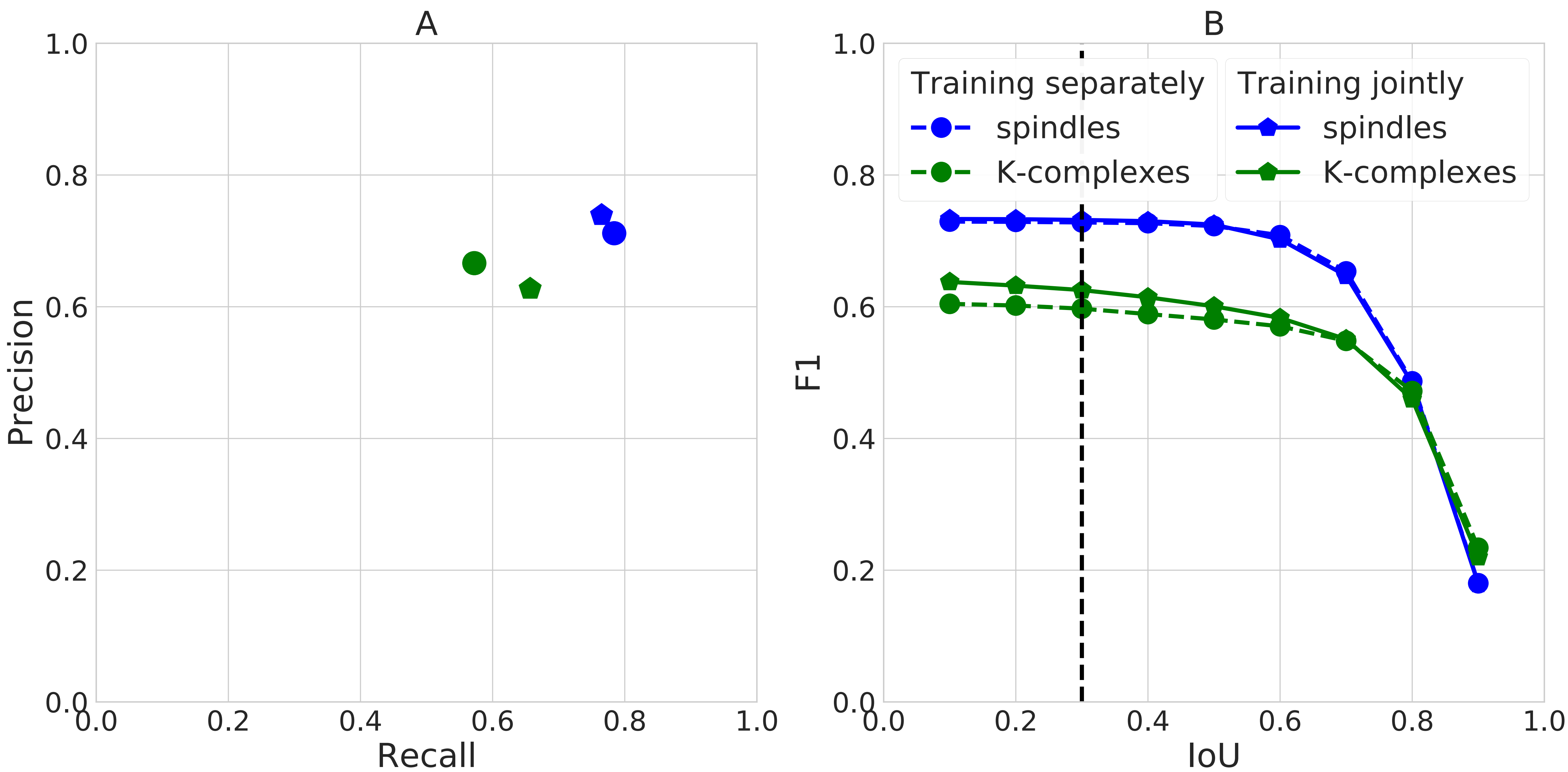
Illustration



True event

Predicted event

Joint detection



Joint detection does not deteriorate performance

2

Convolutional Sparse Coding (CSC) for learning the morphology of neural signals

Multivariate Convolutional Sparse Coding for Electromagnetic Brain Signals, (2018), T. Dupré la Tour, T. Moreau, M. Jas, A. Gramfort, Proc. NeurIPS Conf.

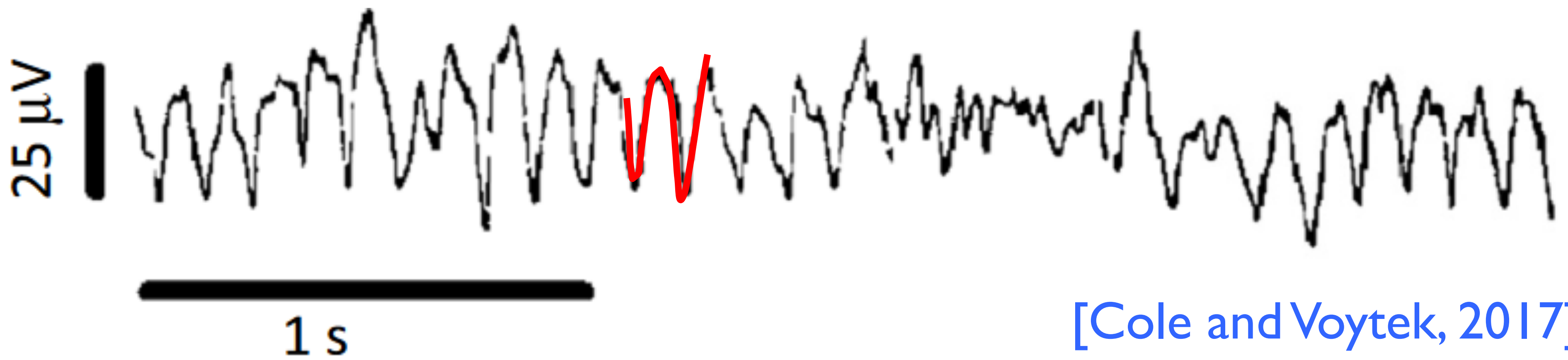
Learning the Morphology of Brain Signals Using Alpha-Stable Convolutional Sparse Coding, (2017), M. Jas, T. Dupré la Tour, U. Simsekli, A. Gramfort, Proc. NeurIPS Conf.

Code: https://alphacsc.github.io

The unsupervised way…

Shape of brain rhythms matter

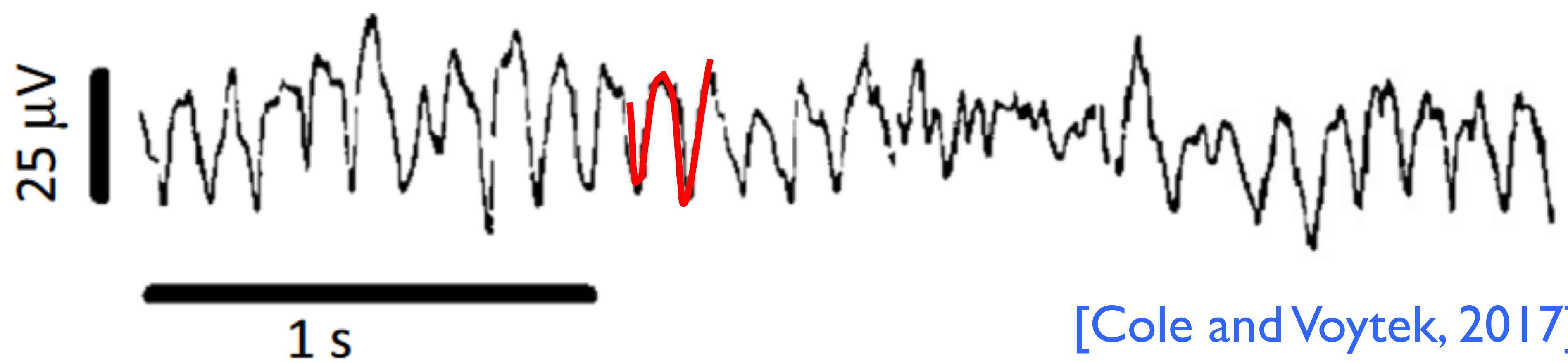
μ rhythm



[Cole and Voytek, 2017]

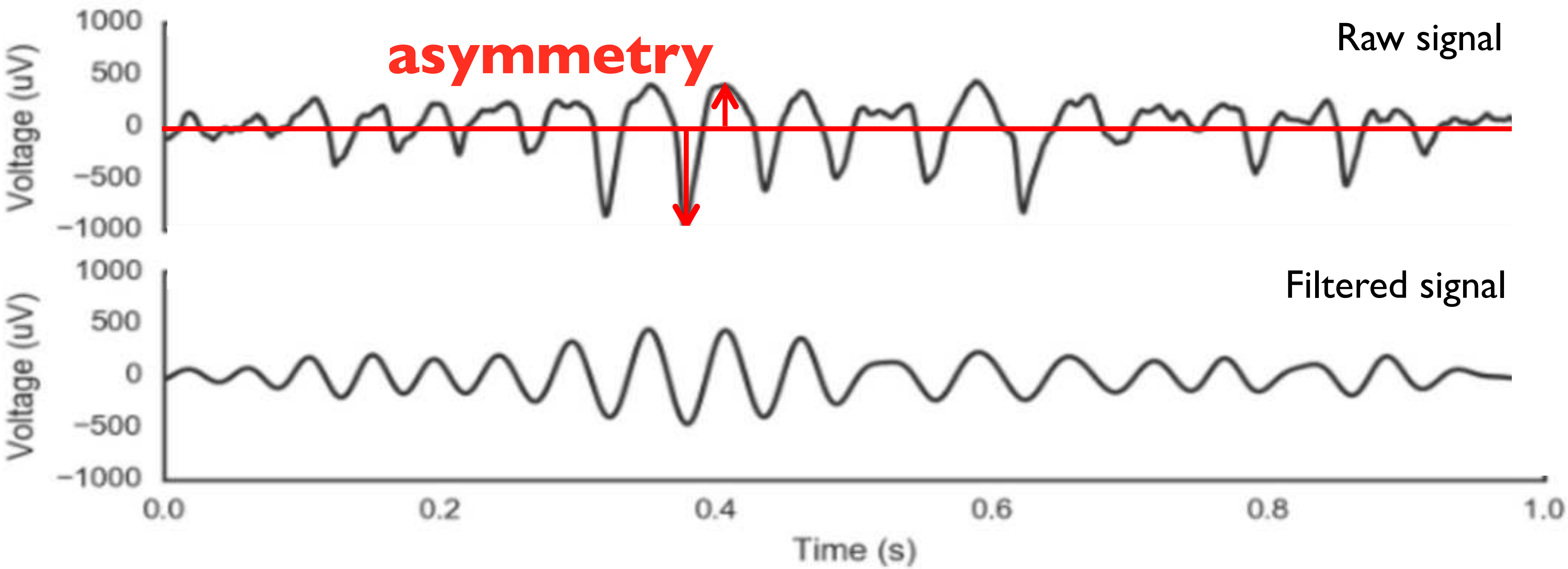
Shape of brain rhythms matter

μ rhythm



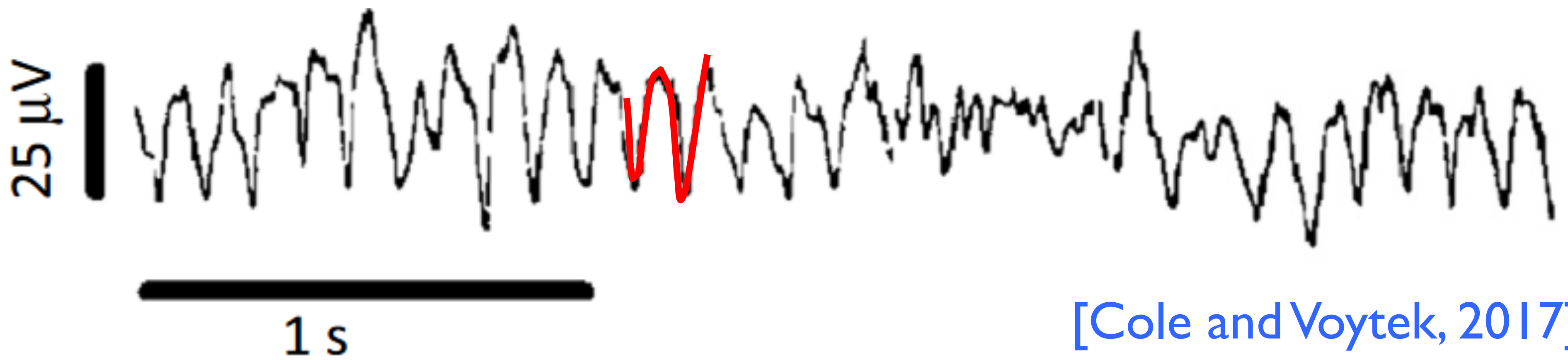
[Cole and Voytek, 2017]

Problem of linear filtering:



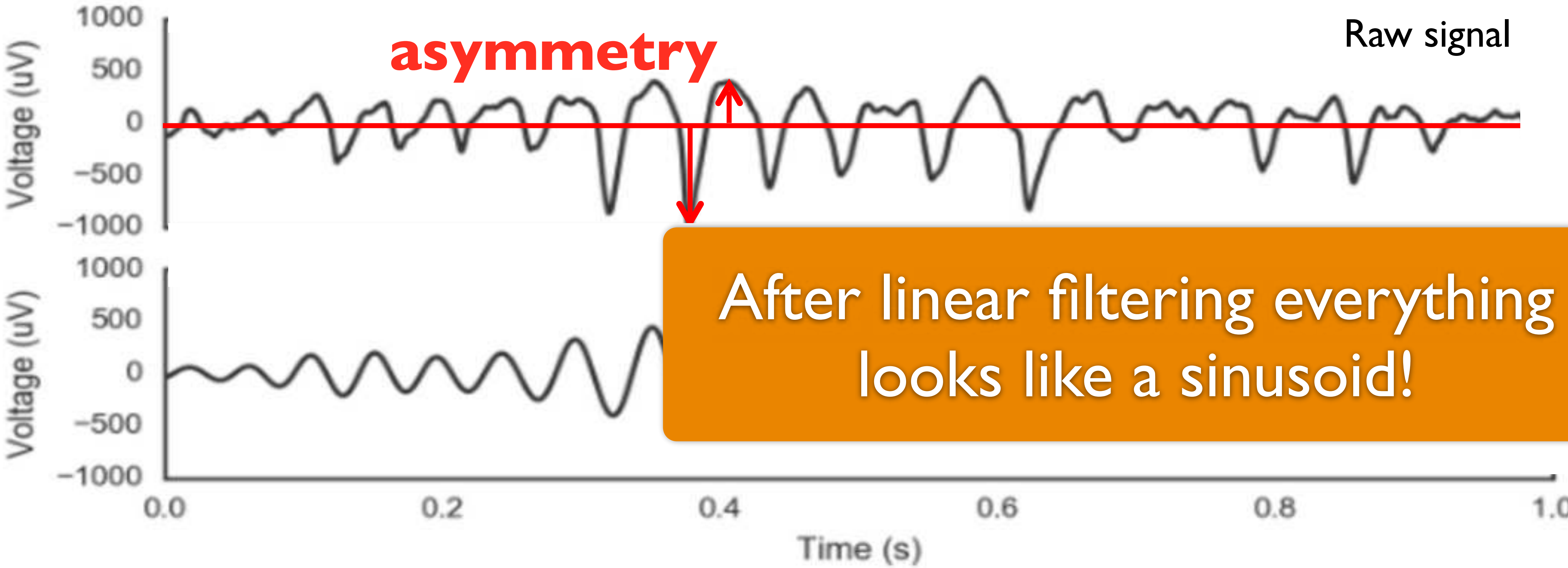
Shape of brain rhythms matter

μ rhythm



[Cole and Voytek, 2017]

Problem of linear filtering:



After linear filtering everything looks like a sinusoid!

Signal representations

- Sparse representations: wavelet basis

  [Morlet 70', Meyer 80', Mallat 90' etc.]

- Sparse coding / dictionary learning

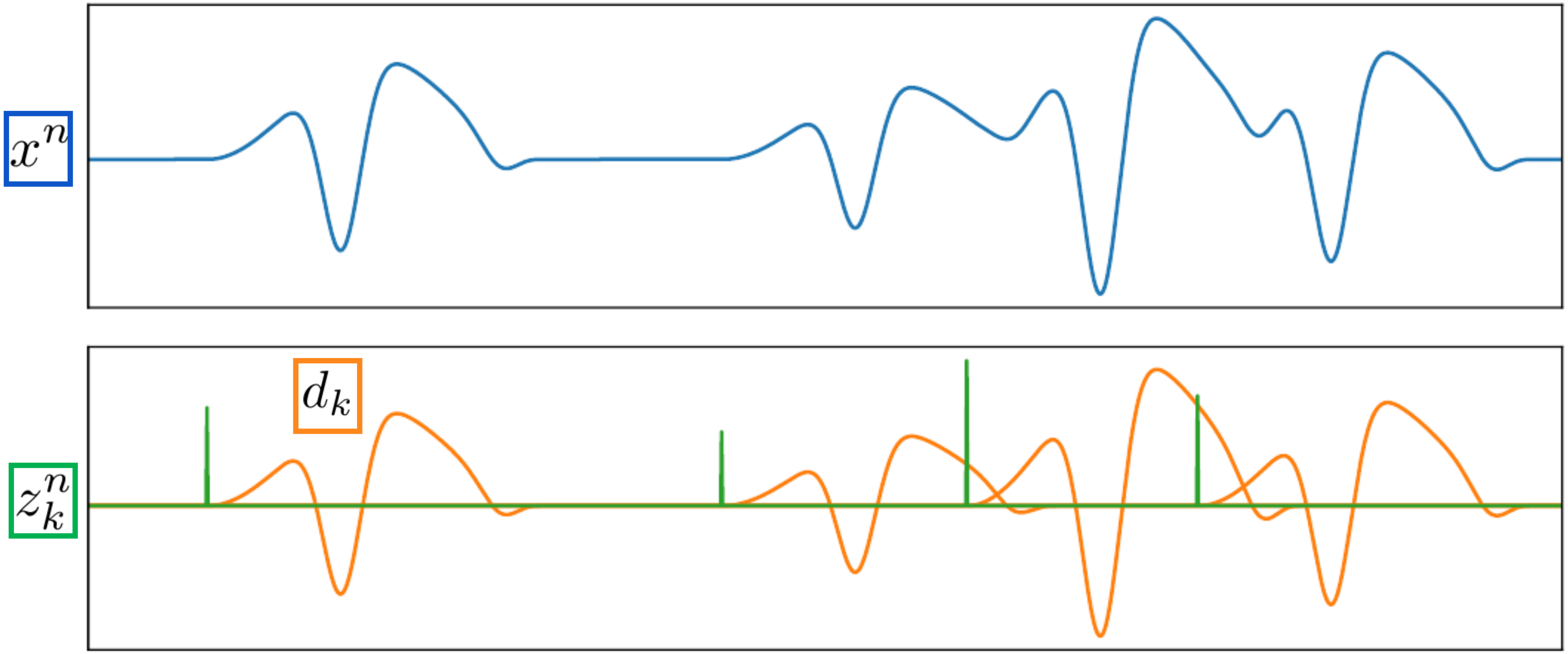
  [Olshausen and Field, 1996, Elad and Aharon, 2006]

- Shift-invariant representations

  [Lewicki and Sejnowski, 1999, Grosse et al, 2007]

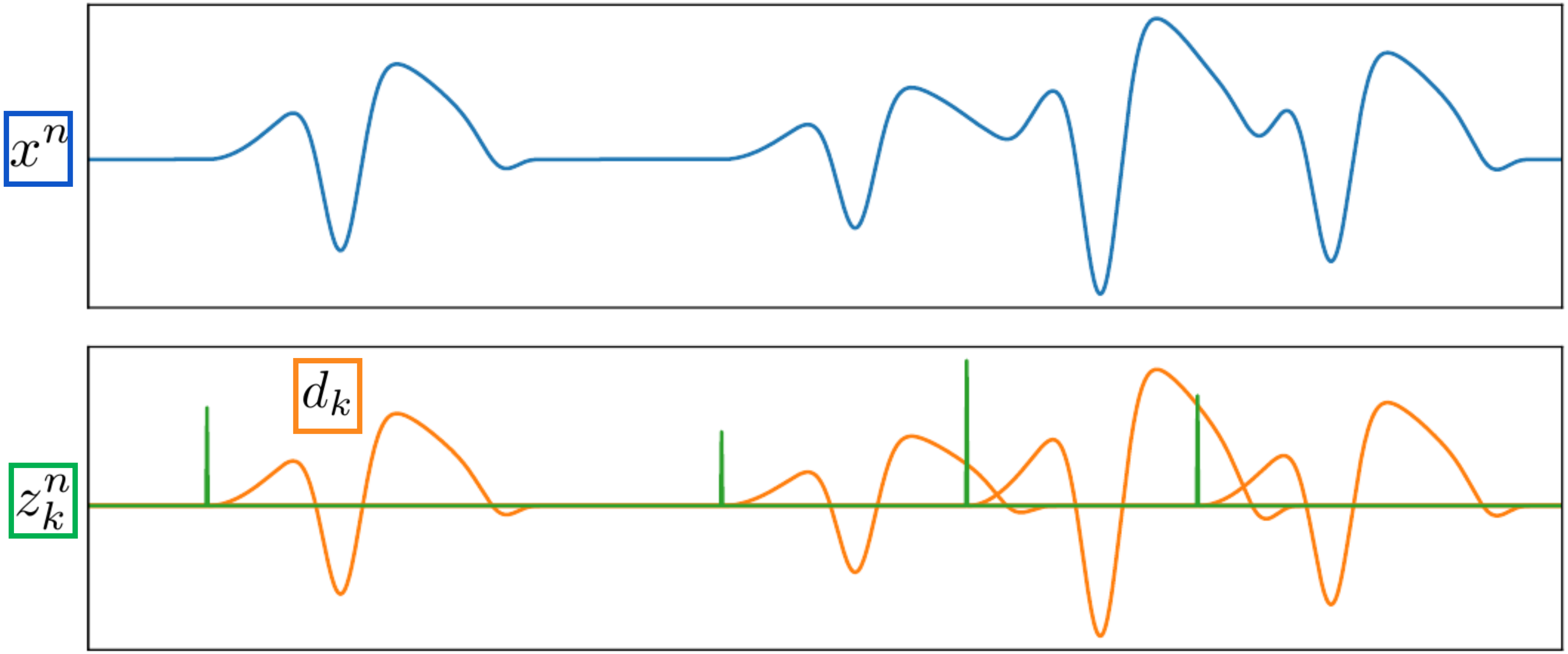
- In neurophysiology:
  - Matching of time-invariant filters (MOTIF) [Jost et al, 2006]
  - Multivariate orthogonal matching pursuit [Barthélemy et al, 2012]
  - Matching pursuit and heuristics [Brokmeier and Principe, 2016]
  - Sliding window machine [Gips et al, 2017]
  - Adaptive waveform learning [Hitziger et al, 2017]

Convolutional sparse coding



[Grosse et al, 2007]

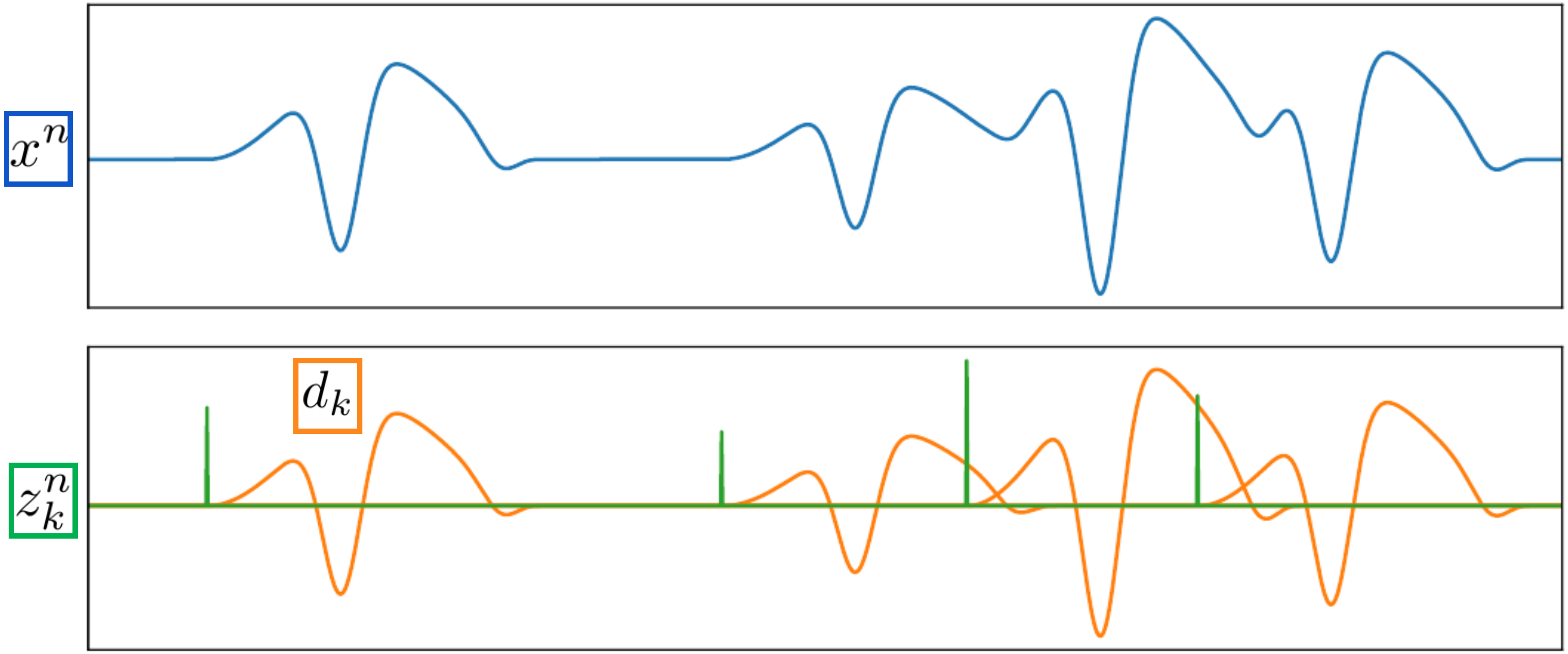
Convolutional sparse coding



$$\boxed{x^n}[t] = \sum_{k=1}^{K} (\boxed{z_k^n} * \boxed{d_k})[t] + \varepsilon[t]$$

[Grosse et al, 2007]

Convolutional sparse coding



$$\min_{d,z} \sum_{n=1}^{N} \frac{1}{2} \left\| \boxed{x^n} - \sum_{k=1}^{K} \boxed{z_k^n} * \boxed{d_k} \right\|_2^2 + \lambda \sum_{k=1}^{K} \|\boxed{z_k^n}\|_1 ,$$

$$\text{s.t.} \quad \|\boxed{d_k}\|_2^2 \leq 1 \text{ and } \boxed{z_k^n} \geq 0.$$

[Grosse et al, 2007]

Optimization strategy

$$\min_{d,z} \sum_{n=1}^{N} \frac{1}{2} \left\| x^n - \sum_{k=1}^{K} z_k^n * d_k \right\|_2^2 + \lambda \sum_{k=1}^{K} \|z_k^n\|_1,$$
$$\text{s.t.} \quad \|d_k\|_2^2 \leq 1 \text{ and } z_k^n \geq 0.$$

Block-coordinate descent:

Optimization strategy

$$\min_{d,z} \sum_{n=1}^{N} \frac{1}{2} \left\| x^n - \sum_{k=1}^{K} z_k^n * d_k \right\|_2^2 + \lambda \sum_{k=1}^{K} \|z_k^n\|_1 ,$$
$$\text{s.t.} \quad \|d_k\|_2^2 \leq 1 \text{ and } z_k^n \geq 0.$$

Block-coordinate descent:

- Z-step
  - GCD [Kavukcuoglu et al, 2010]
  - FISTA [Chalasani et al, 2013]
  - ADMM [Bristow et al, 2013]
  - ADMM + FFT [Wohlberg, 2016]
  - L-BFGS [Jas et al, 2017]
  - LGCD [Dupré la Tour et al, 2018] [Moreau et al, 2018]

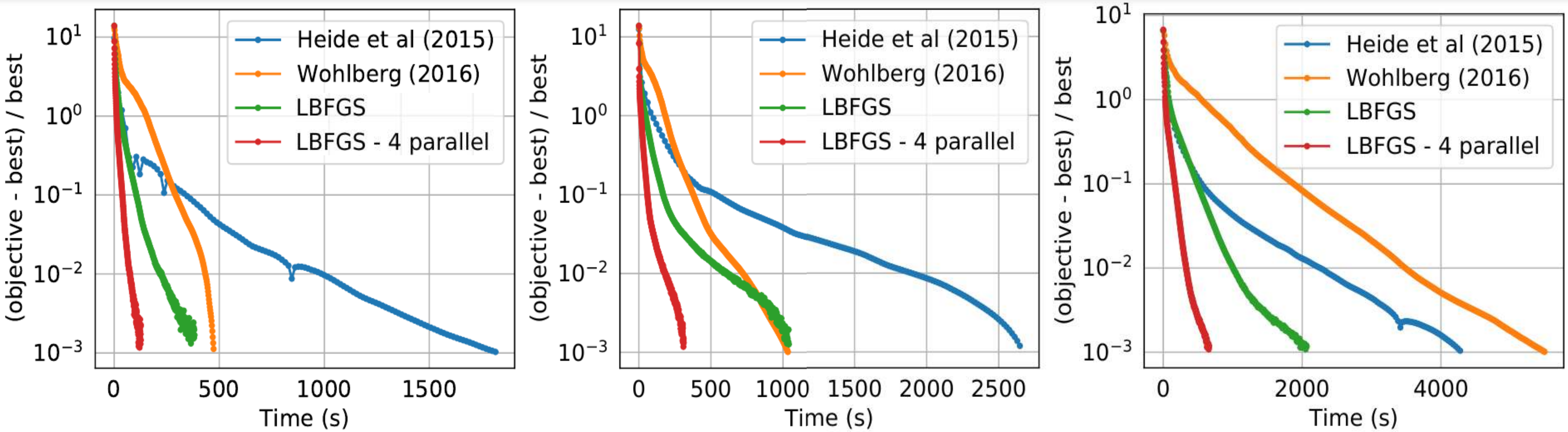
Optimization strategy

$$\min_{d,z} \sum_{n=1}^{N} \frac{1}{2} \left\| x^n - \sum_{k=1}^{K} z_k^n * d_k \right\|_2^2 + \lambda \sum_{k=1}^{K} \|z_k^n\|_1,$$
$$\text{s.t.} \quad \|d_k\|_2^2 \leq 1 \text{ and } z_k^n \geq 0.$$

Block-coordinate descent:

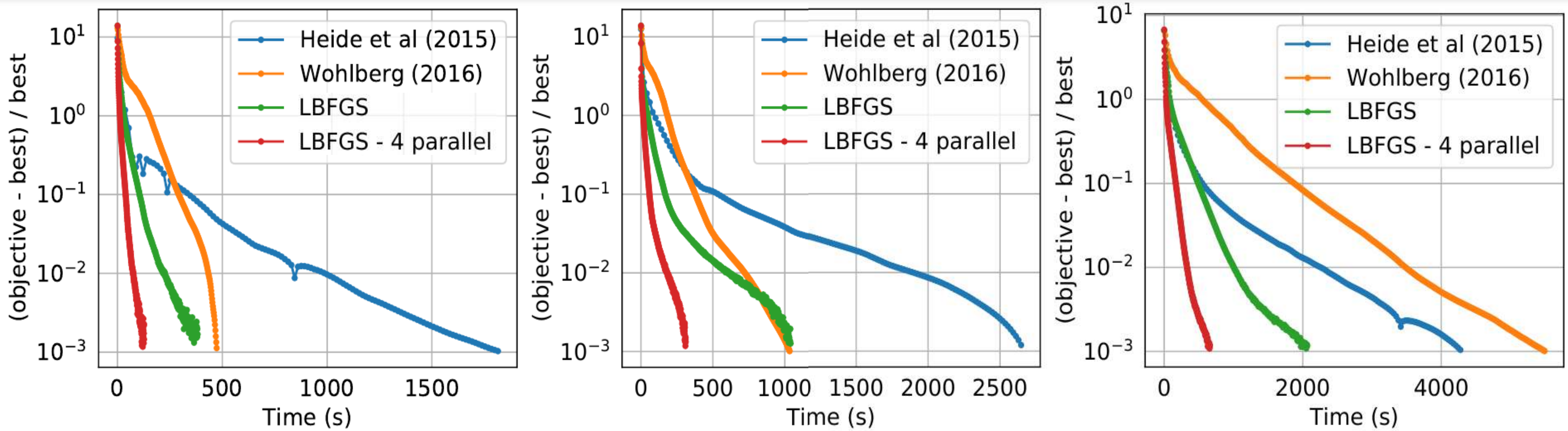
- Z-step
  - GCD [Kavukcuoglu et al, 2010]
  - FISTA [Chalasani et al, 2013]
  - ADMM [Bristow et al, 2013]
  - ADMM + FFT [Wohlberg, 2016]
  - L-BFGS [Jas et al, 2017]
  - LGCD [Dupré la Tour et al, 2018] [Moreau et al, 2018]
- D-step
  - FFT [Grosse et al, 2007]
  - ADMM + FFT [Heide et al, 2015]
  - ADMM + FFT [Wohlberg, 2016]
  - L-BFGS (dual) [Jas et al, 2017]
  - PGD [Dupré la Tour et al, 2018)

Speed benchmarks

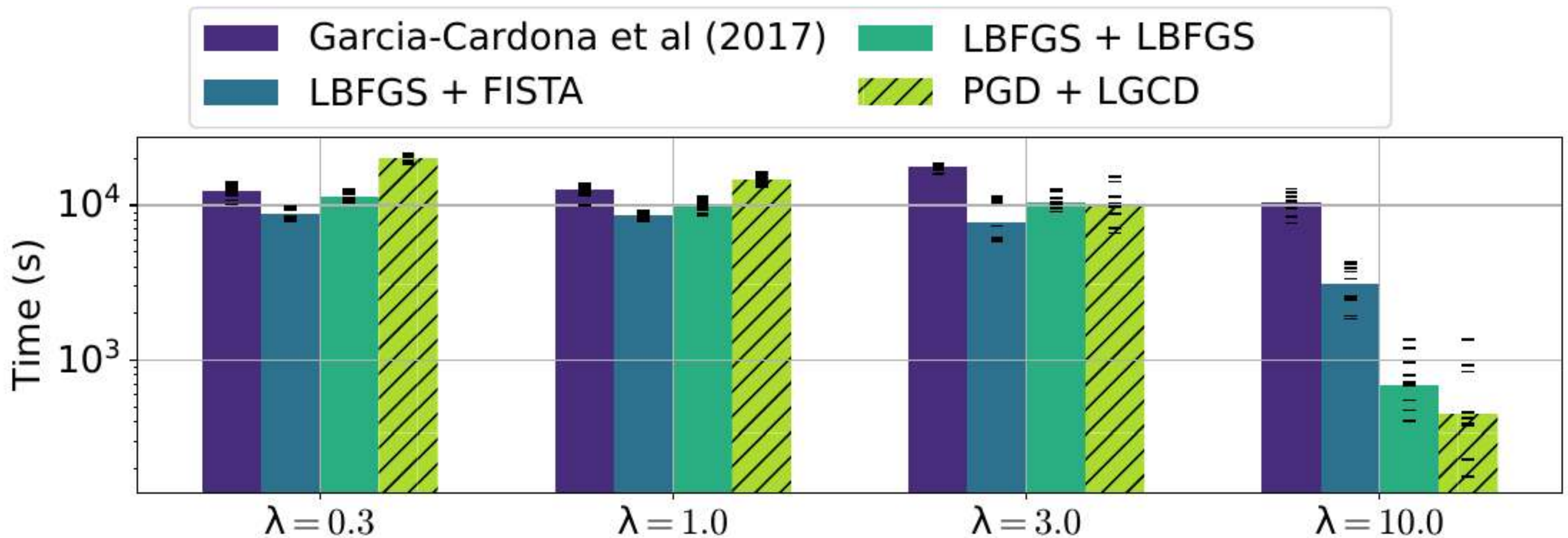


(a) $K = 2,\ L = 32.$ (b) $K = 2,\ L = 128.$ (c) $K = 10,\ L = 32.$

Speed benchmarks

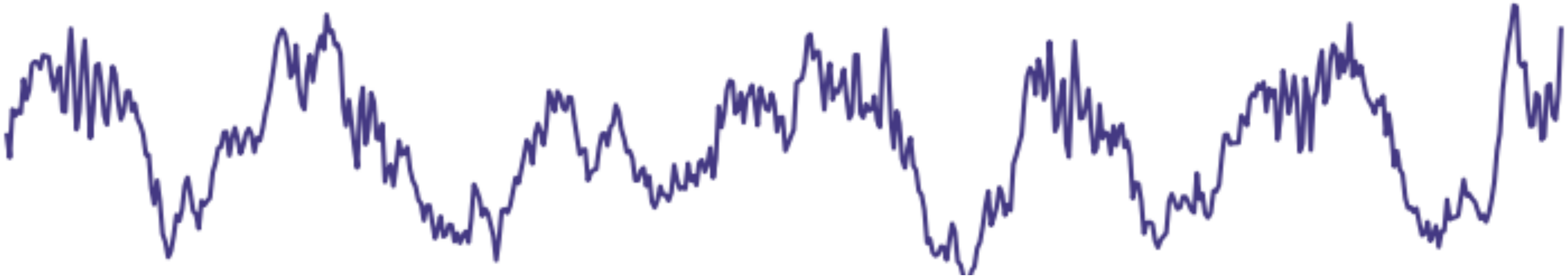


(a) $K = 2,\ L = 32.$ (b) $K = 2,\ L = 128.$ (c) $K = 10,\ L = 32.$

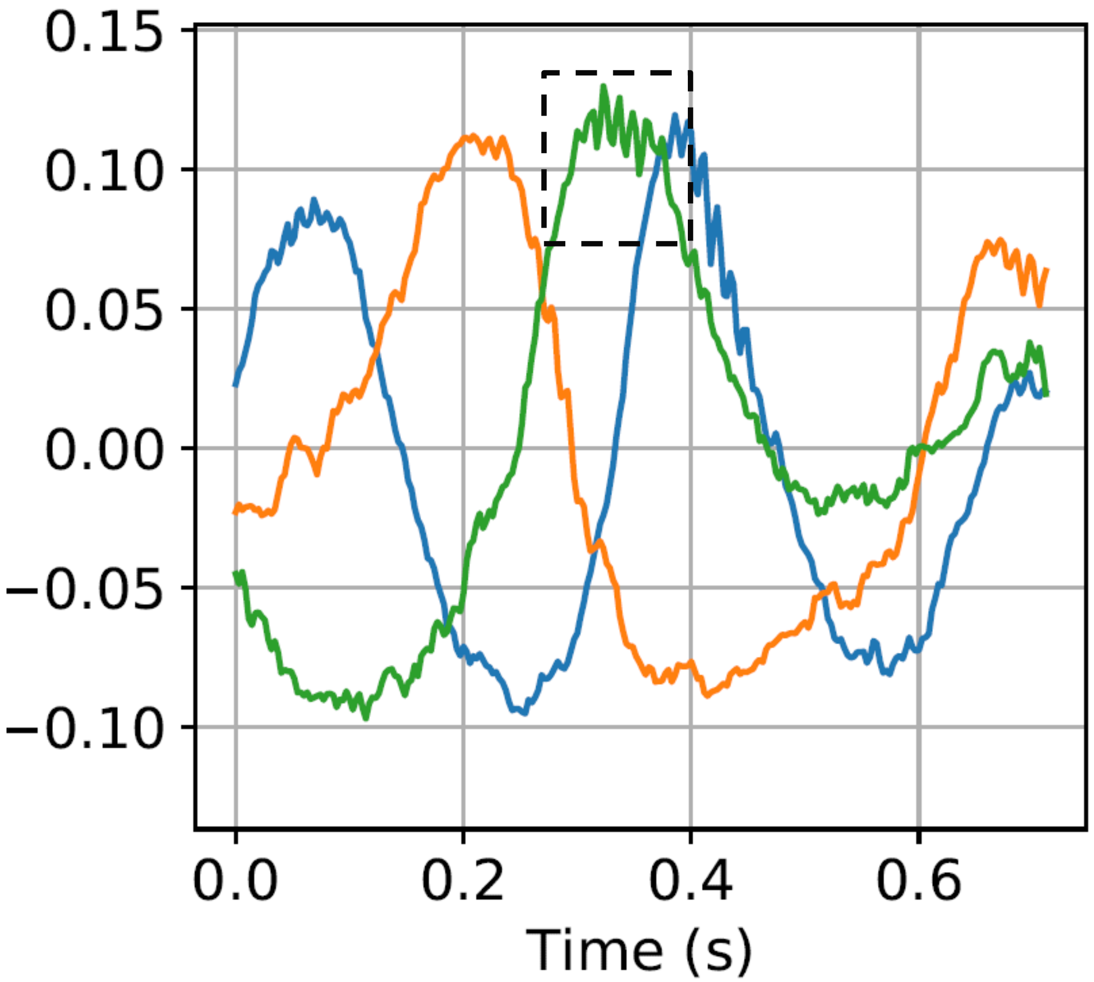


Learned atoms

Data:

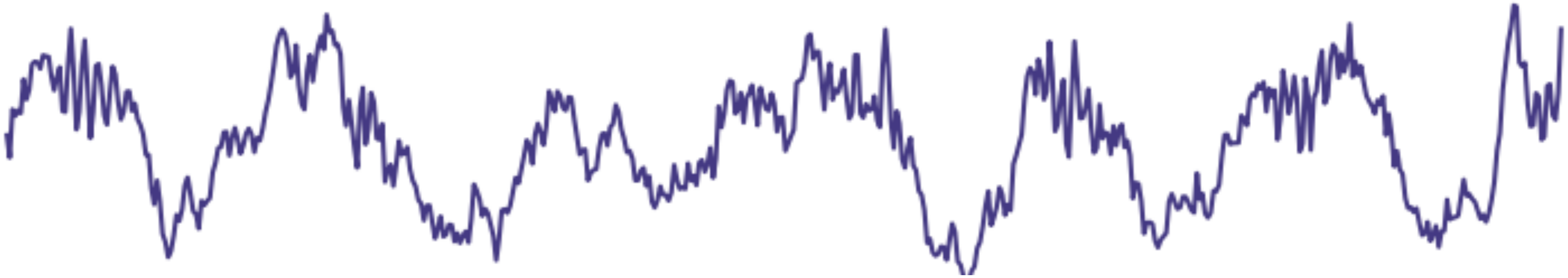


~80 Hz

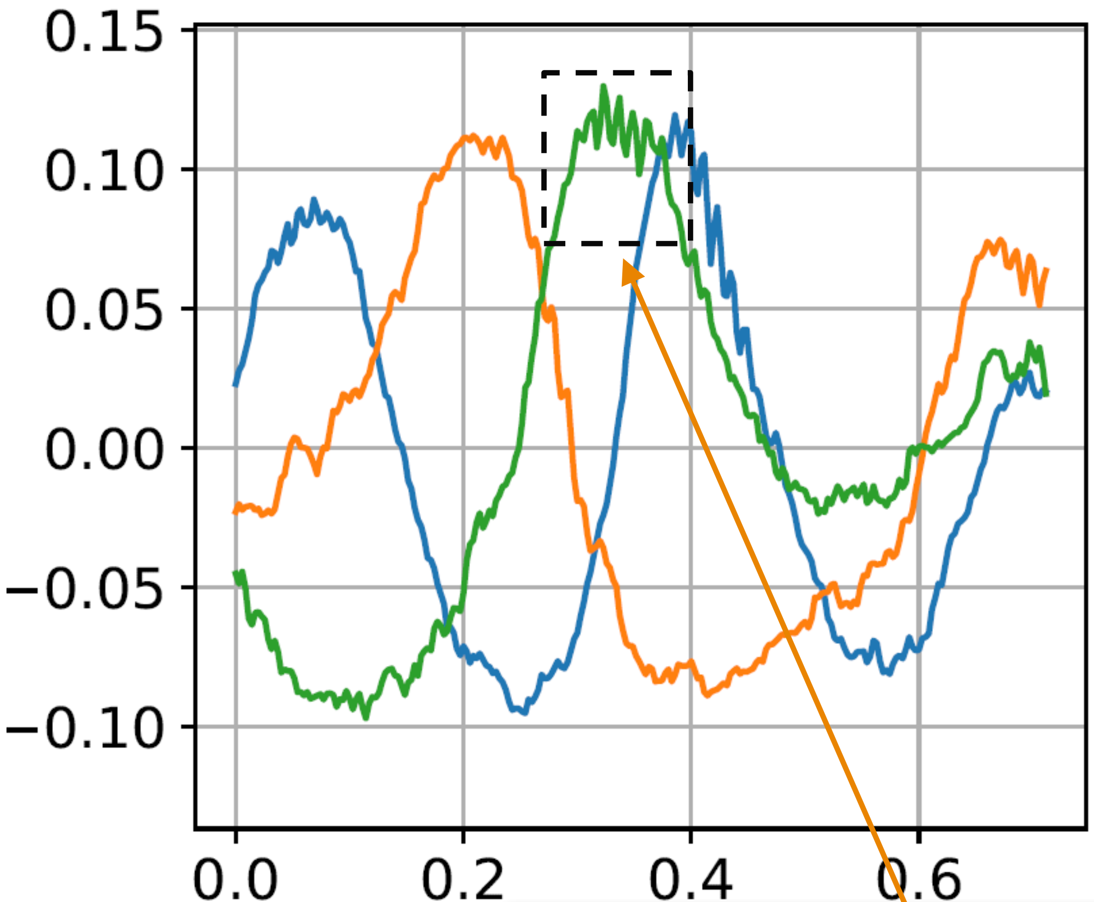


Learned atoms

Data:



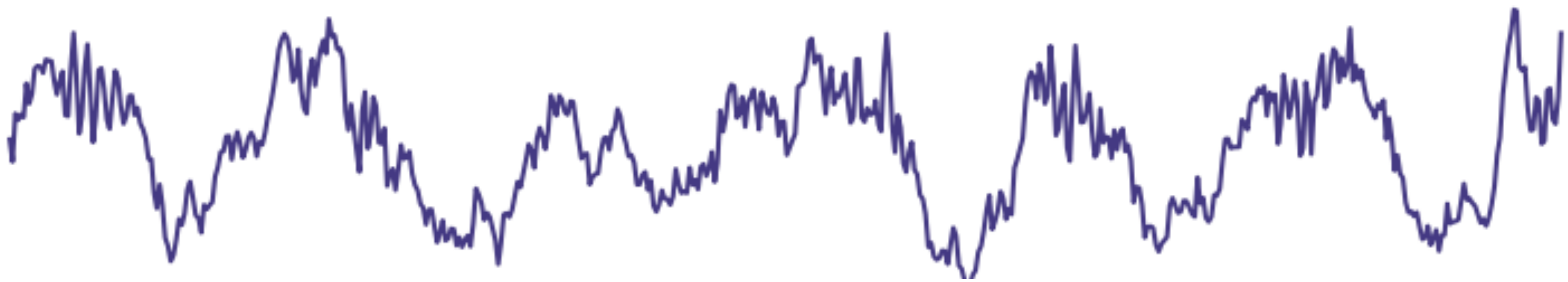
~80 Hz



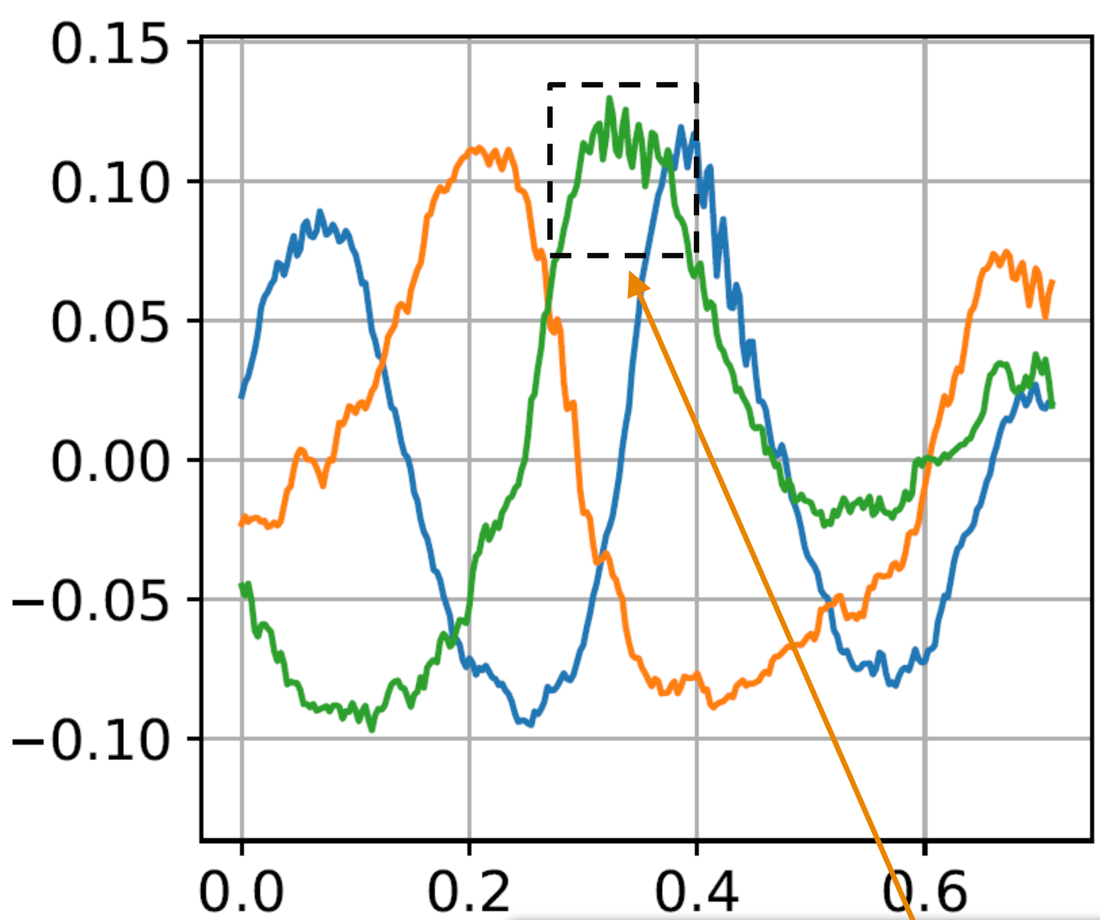
CSC reveals CFC

Learned atoms

Data:



~80 Hz

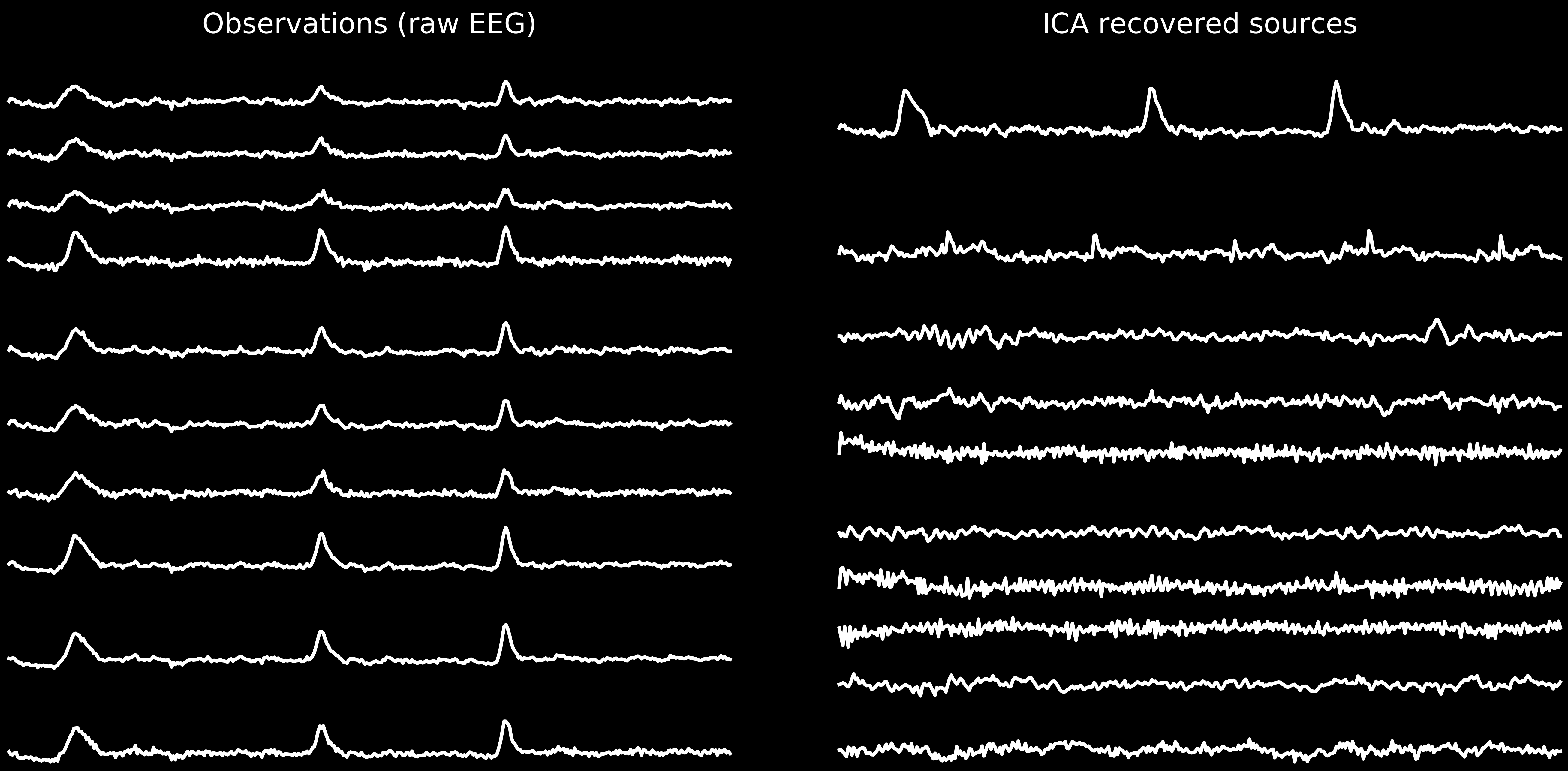


How about if I have many channels?

CSC reveals CFC

From ICA to CSC

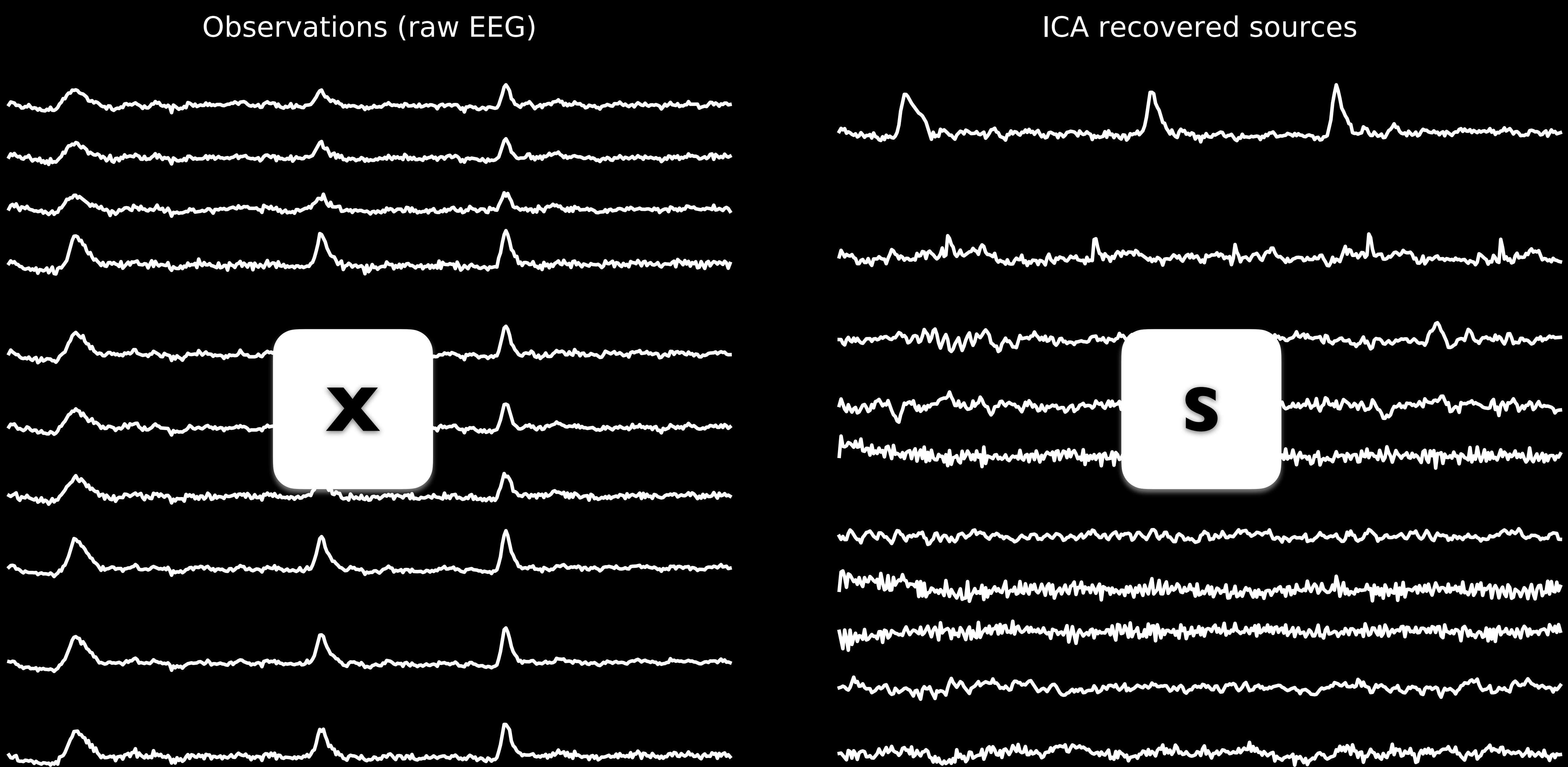
Independent Component Analysis (ICA)



https://pypi.python.org/pypi/python-picard/0.1

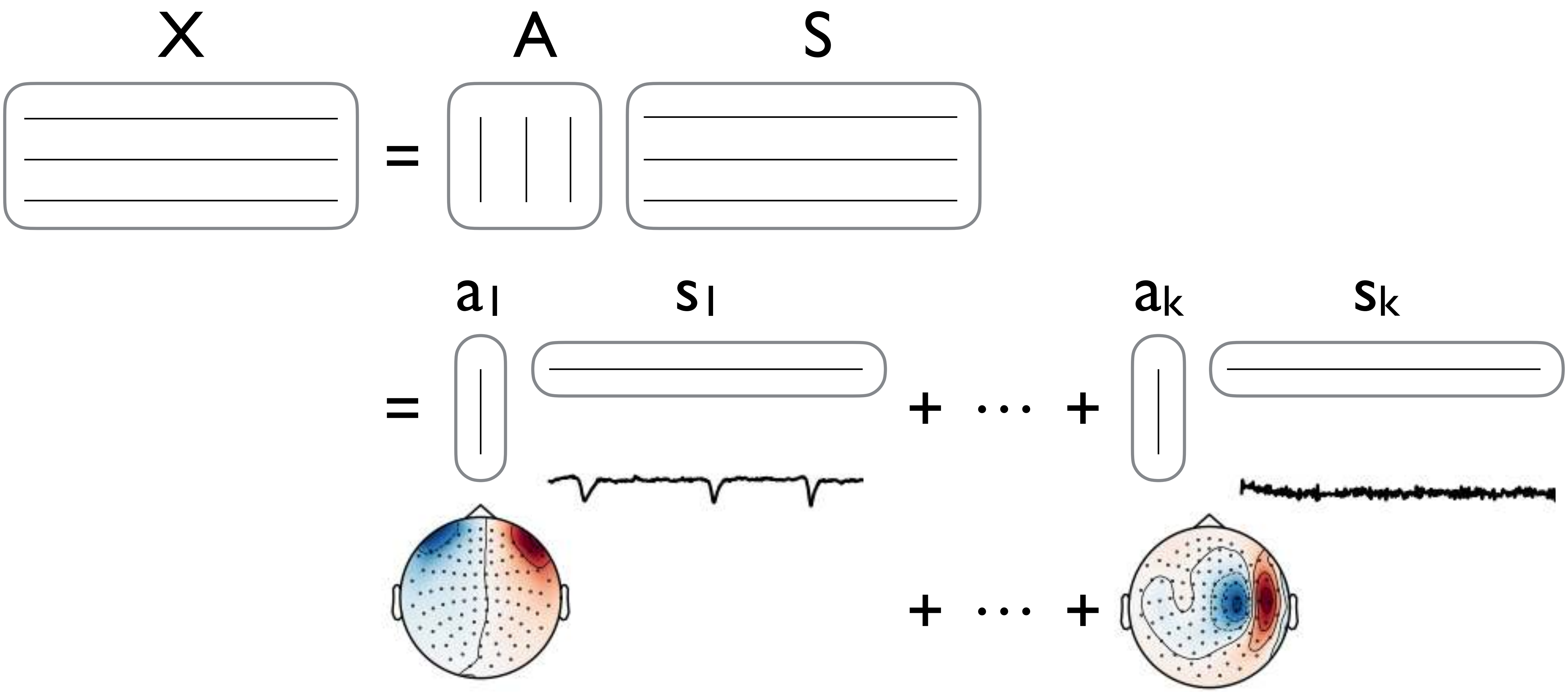
From ICA to CSC

Independent Component Analysis (ICA)



https://pypi.python.org/pypi/python-picard/0.1

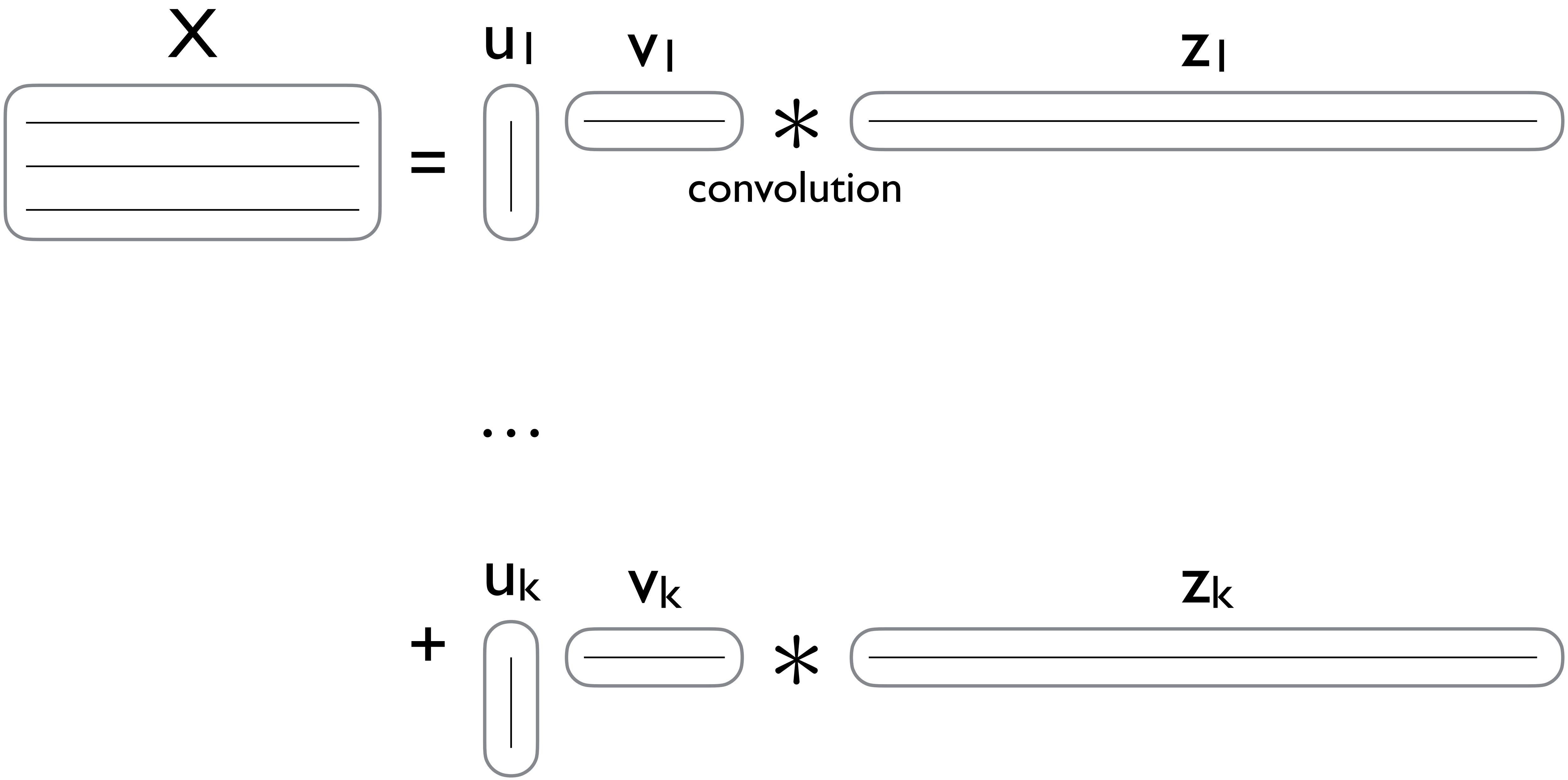
From ICA...



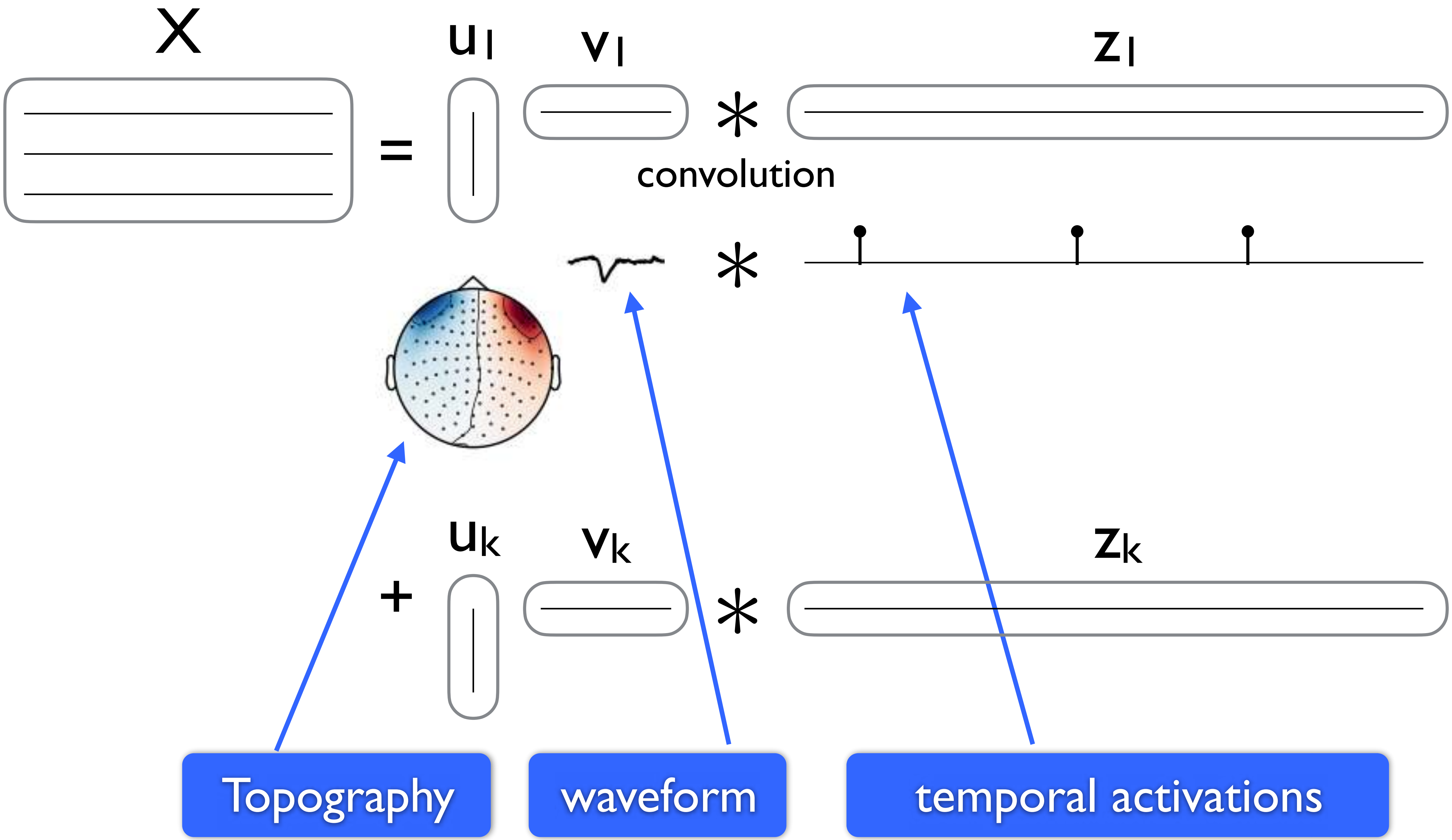
https://www.martinos.org/mne/stable/auto_tutorials/plot_artifacts_correction_ica.html

https://pierreablin.github.io/picard/auto_examples/plot_ica_eeg.html

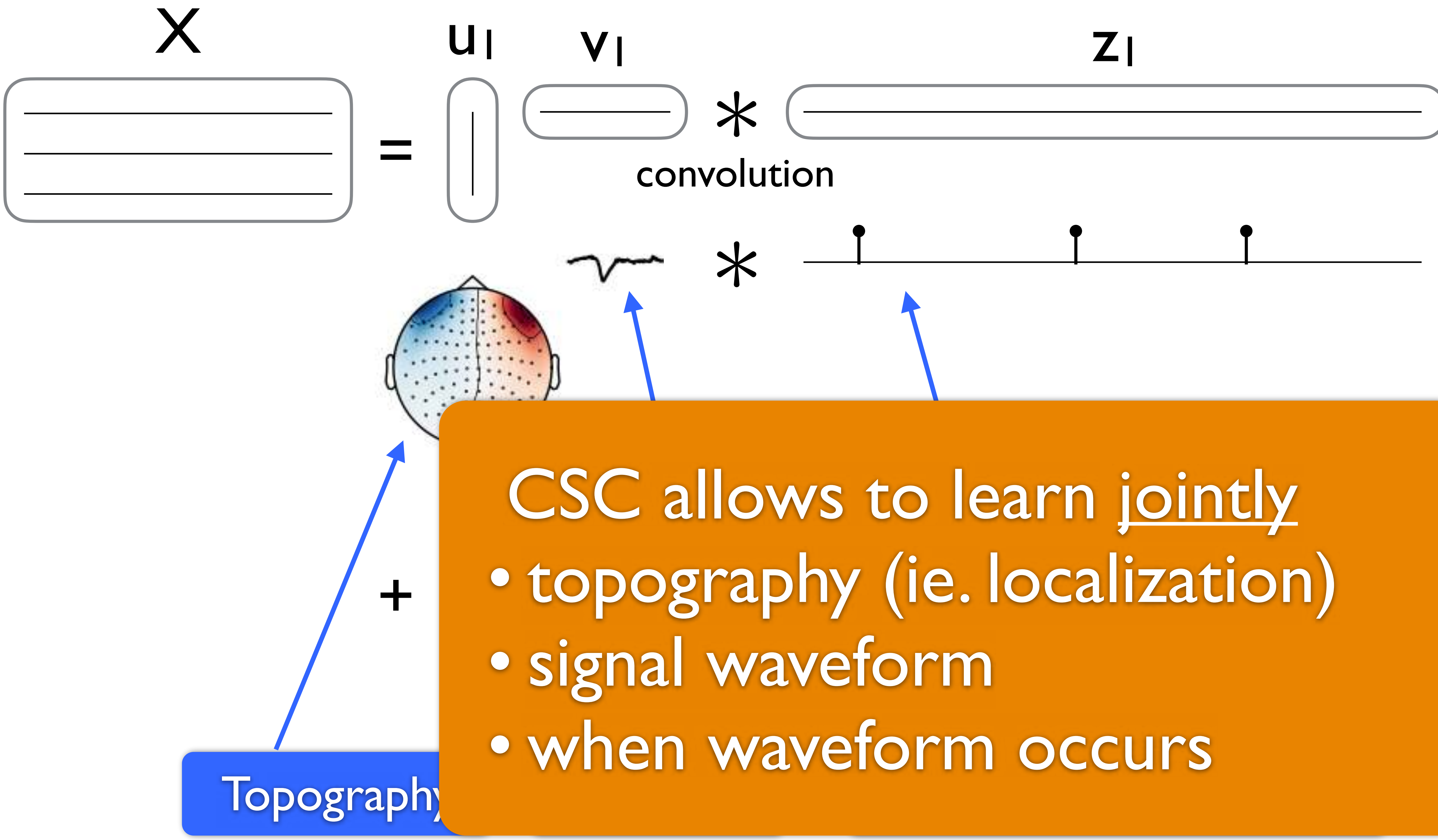
... to CSC



… to CSC


X
=
u_1
v_1
z_1
*
convolution
+
u_k
v_k
z_k
Topography
waveform
temporal activations

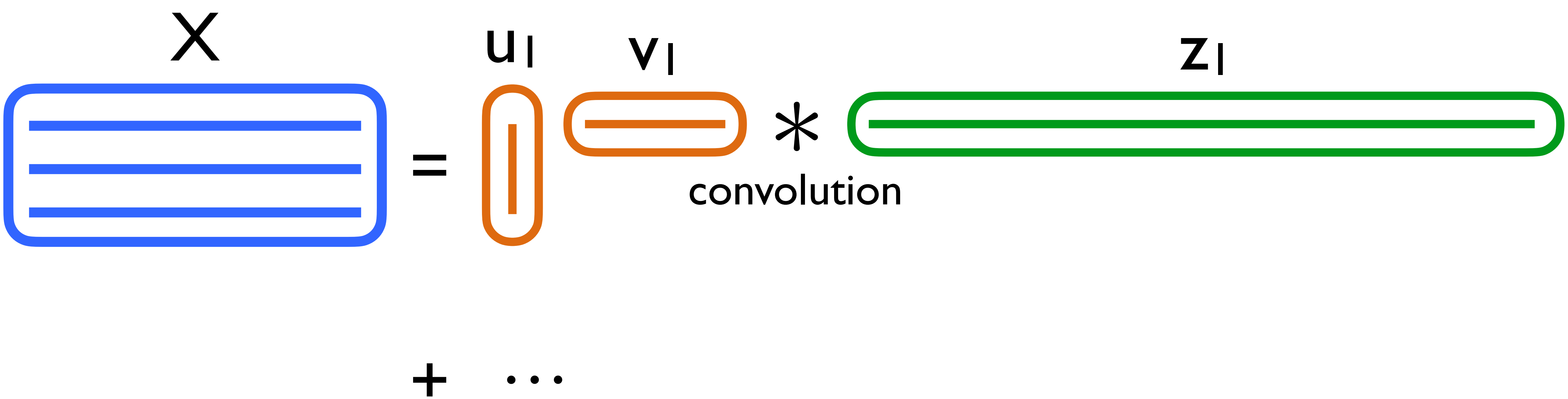
... to CSC



CSC allows to learn jointly

- topography (ie. localization)
- signal waveform
- when waveform occurs

Rank 1 Multivariate CSC

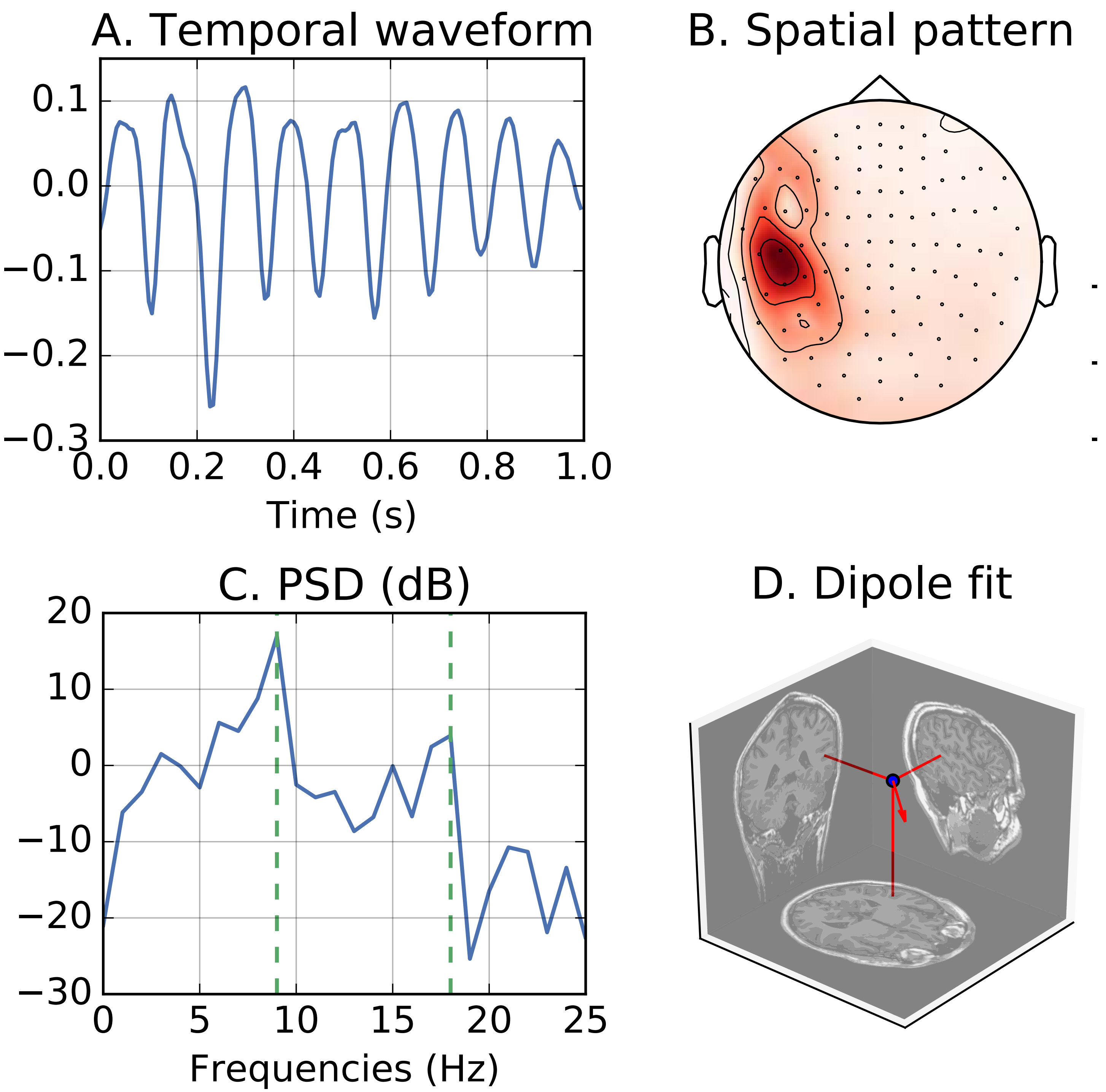


$$\min_{u,v,z} \sum_{n=1}^{N} \frac{1}{2} \left\| \boxed{X^n} - \sum_{k=1}^{K} \boxed{z_k^n} * \boxed{(u_k v_k^\top)} \right\|_2^2 + \lambda \sum_{k=1}^{K} \|z_k^n\|_1 ,$$

$$\text{s.t.} \quad \|u_k\|_2^2 \leq 1 \ , \ \|v_k\|_2^2 \leq 1 \text{ and } z_k^n \geq 0.$$

[Multivariate Convolutional Sparse Coding for Electromagnetic Brain Signals, (2018), T. Dupré la Tour, T. Moreau, M. Jas, A. Gramfort, Proc. NeurIPS Conf.]

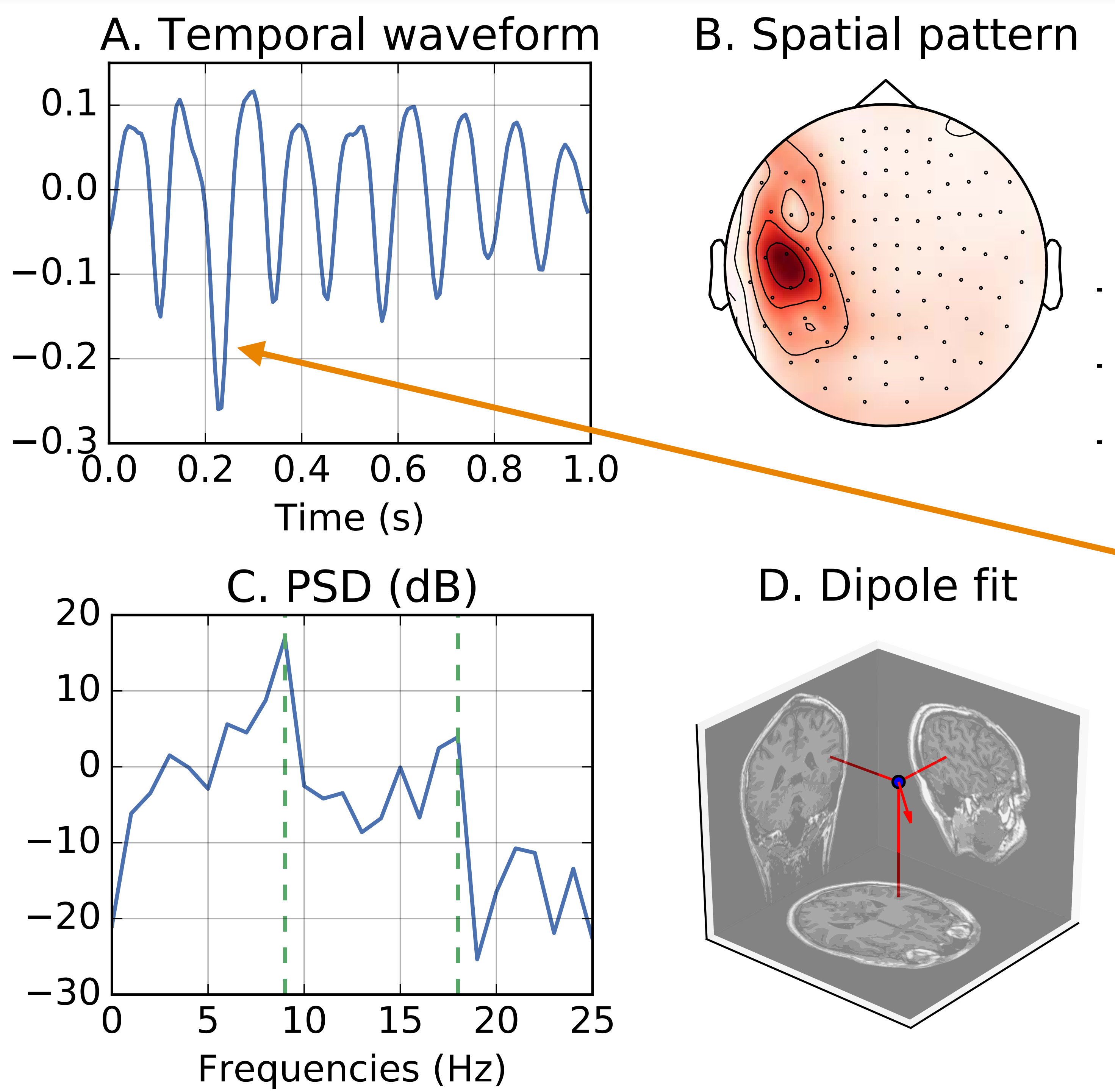
CSC on MEG



- MEG vectorview
- Median nerve stim.

[Multivariate Convolutional Sparse Coding for Electromagnetic Brain Signals, (2018), T. Dupré la Tour, T. Moreau, M. Jas, A. Gramfort, Proc. NeurIPS Conf.]

CSC on MEG

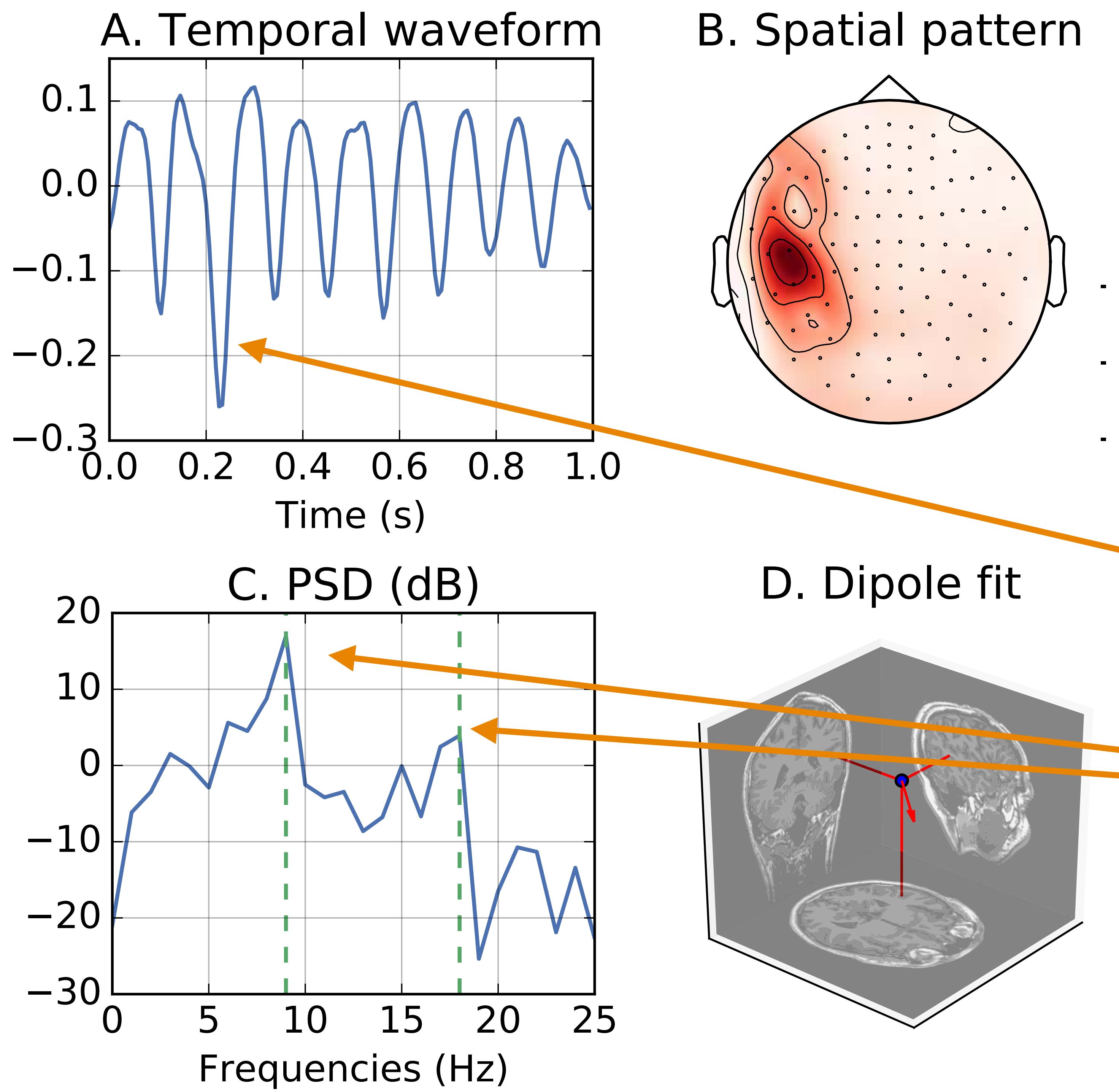


- MEG vectorview
- Median nerve stim.

CSC reveals mu-shaped waveforms

[Multivariate Convolutional Sparse Coding for Electromagnetic Brain Signals, (2018), T. Dupré la Tour, T. Moreau, M. Jas, A. Gramfort, Proc. NeurIPS Conf.]

CSC on MEG



- MEG vectorview
- Median nerve stim.

CSC reveals mu-shaped waveforms

See the frequency harmonics

[Multivariate Convolutional Sparse Coding for Electromagnetic Brain Signals, (2018), T. Dupré la Tour, T. Moreau, M. Jas, A. Gramfort, Proc. NeurIPS Conf.]

alphaCSC: Convolution sparse coding for time-series

build passing codecov 81%

This is a library to perform shift-invariant sparse dictionary learning, also known as convolutional sparse coding (CSC), on time-series data. It includes a number of different models:

1. univariate CSC
2. multivariate CSC
3. multivariate CSC with a rank-1 constraint [1]
4. univariate CSC with an alpha-stable distribution [2]

A mathematical descriptions of these models is available in the documentation.

Installation

To install this package, the easiest way is using `pip`. It will install this package and its dependencies. The `setup.py` depends on `numpy` and `cython` for the installation so it is advised to install them beforehand. To install this package, please run

```
pip install numpy cython
pip install git+https://github.com/alphacsc/alphacsc.git#egg=alphacsc
```

If you do not have admin privileges on the computer, use the `--user` flag with `pip`. To upgrade, use the `--upgrade` flag provided by `pip`.

To check if everything worked fine, you can run:

```
python -c 'import alphacsc'
```

and it should not give any error messages.

https://alphacsc.github.io

Quickstart

Here is an example to present briefly the API:

```
import numpy as np
```

Extracting artifact and evoked response atoms from the sample dataset

This example illustrates how to learn rank1 atoms on the sample dataset from `mne`. We display a selection of atoms, featuring heartbeat and eyeblink artifacts, three atoms of evoked responses, and a non-sinusoidal oscillation.

```
# Authors: Thomas Moreau <thomas.moreau@inria.fr>
#          Mainak Jas <mainak.jas@telecom-paristech.fr>
#          Tom Dupre La Tour <tom.duprelatour@telecom-paristech.fr>
#          Alexandre Gramfort <alexandre.gramfort@telecom-paristech.fr>
#
# License: BSD (3-clause)
```

Let us first define the parameters of our model.

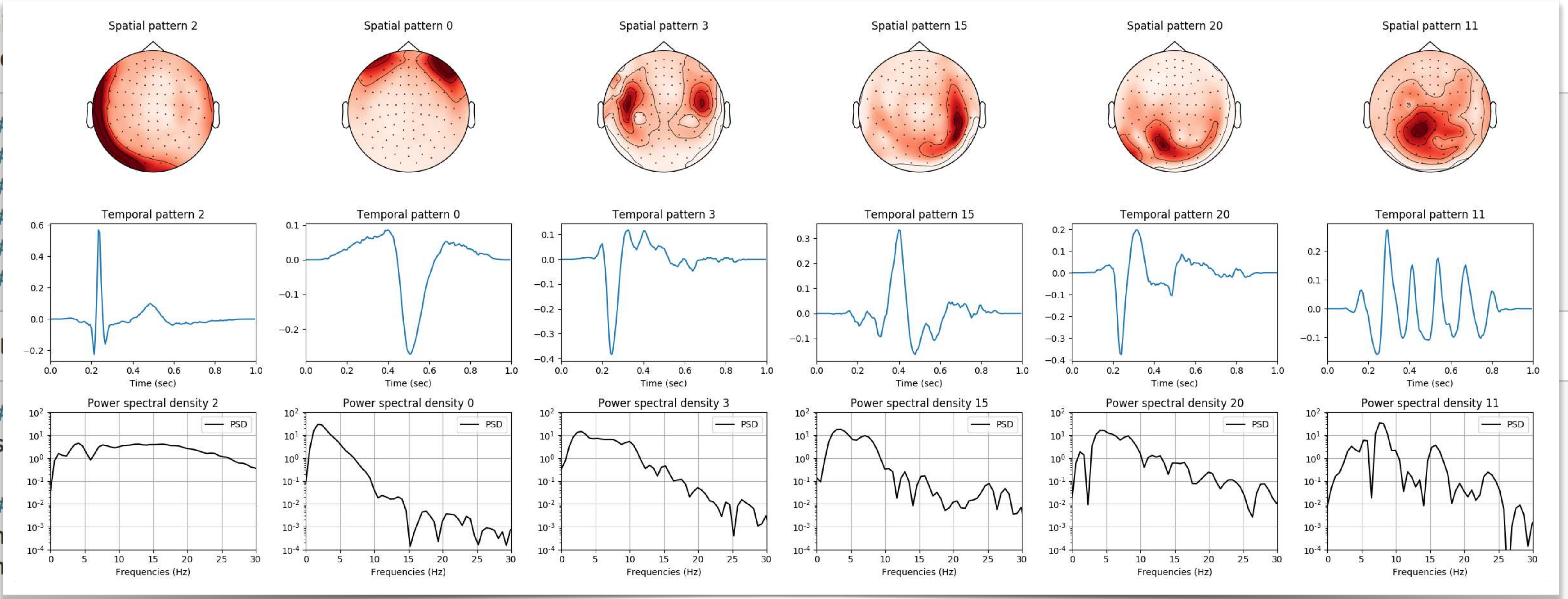
```
# sample frequency
sfreq = 150.

# Define the shape of the dictionary
n_atoms = 40
n_times_atom = int(round(sfreq * 1.0))  # 1000. ms

# Regularization parameter which control sparsity
reg = 0.1

# number of processors for parallel computing
n_jobs = 5
```

Next, we define the parameters for multivariate CSC

https://alphacsc.github.io/auto_examples/multicsc/plot_sample_evoked_response.html

Extracting artifact and evoked response atoms from the sample dataset

Th
ey



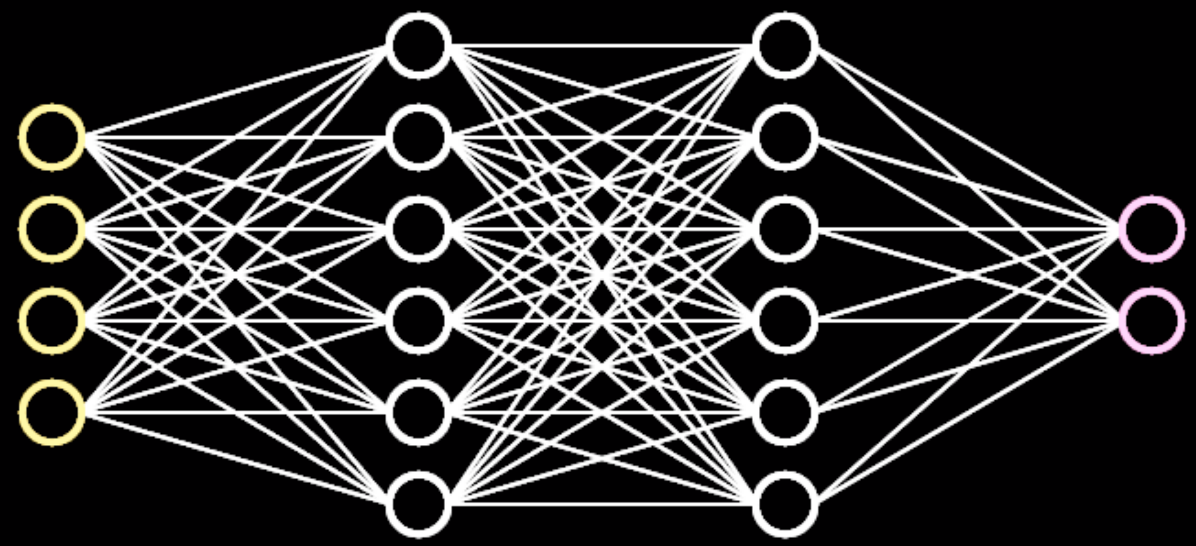

```
# Regularization parameter which control sparsity
reg = 0.1

# number of processors for parallel computing
n_jobs = 5
```

Next, we define the parameters for multivariate CSC

https://alphacsc.github.io/auto_examples/multicsc/plot_sample_evoked_response.html

3

Self-supervised learning on EEG



Uncovering the structure of clinical EEG signals with self-supervised learning
Banville, H., Chehab, O., Hyvärinen, A., Engemann, D. and Gramfort, A. (2020)
Journal of Neural Engineering & ArXiv abs/2007.16104
Self-supervised representation learning from electroencephalography signals
Banville, H., Albuquerque, I., Moffat, G., Engemann, D. and Gramfort, A. (2019)
Proc. Machine Learning for Signal Processing (MLSP) .

The self-supervised way…

Self-supervision

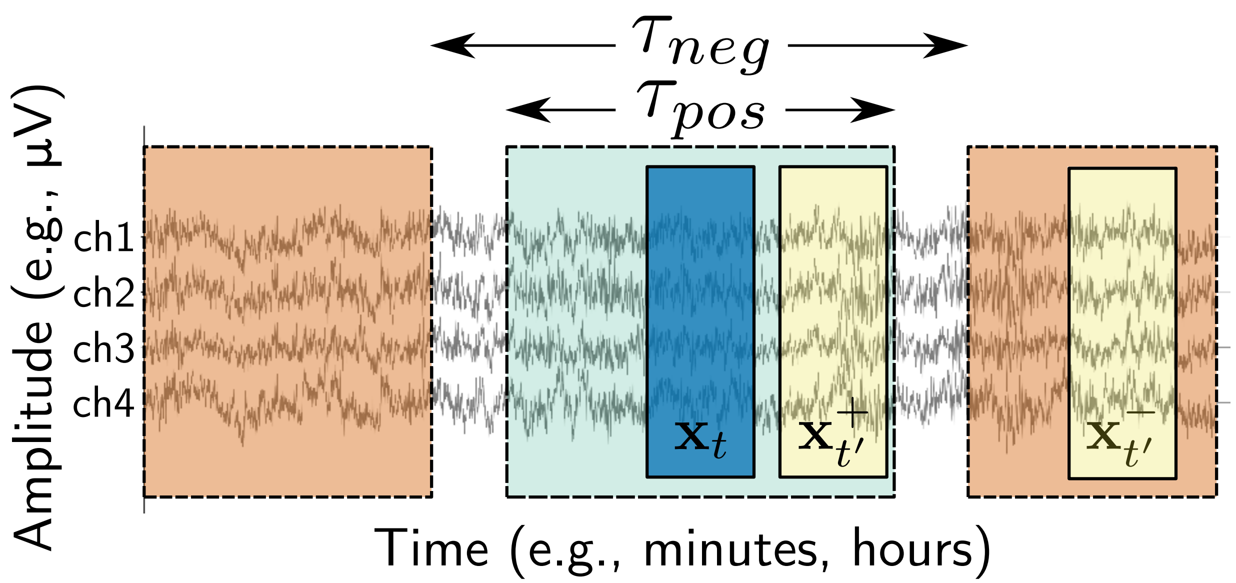
[Noroozi & Favaro 2016] use a deep neural network to solve the Jigsaw puzzle

Use the **structure** of the data to pretrain a **feature extractor** with a supervised *pretext task* – then use the features on a *downstream task*.

Pretext Task

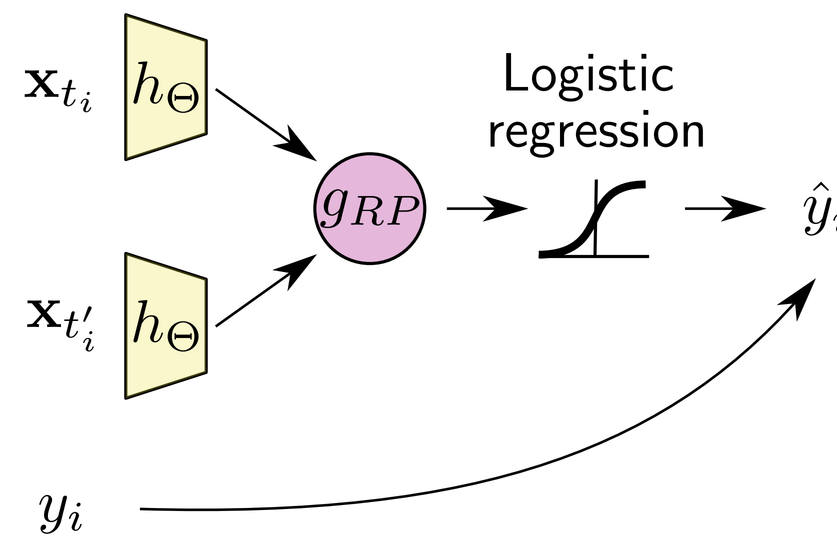
Relative positioning (RP)

1 Sampling



$$y_i = \begin{cases} 1, & \text{if } |t_i - t_i'| \leq \tau_{pos} \\ \text{-}1, & \text{if } |t_i - t_i'| > \tau_{neg} \end{cases}$$

2 Training



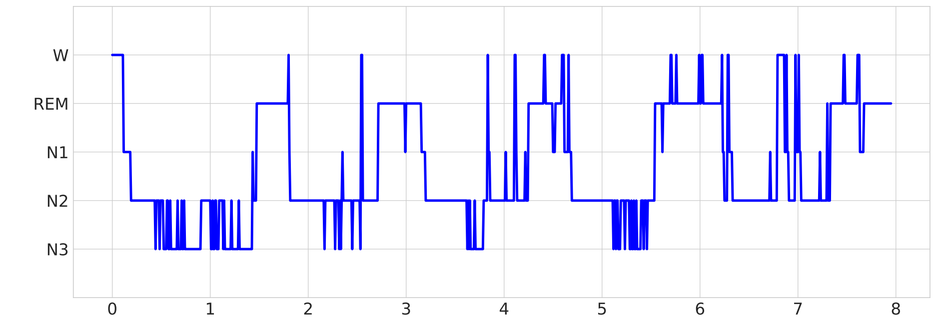
Predict if 2 windows of data are close in time

Other approaches: CPC [Oord et al. 2018], PCL [Hyvärinen et al. 2017] etc.

Downstream tasks on clinical EEG

Sleep staging:

Predict sleep stage from EEG
(5-class: W, N1, N2, N3, R)



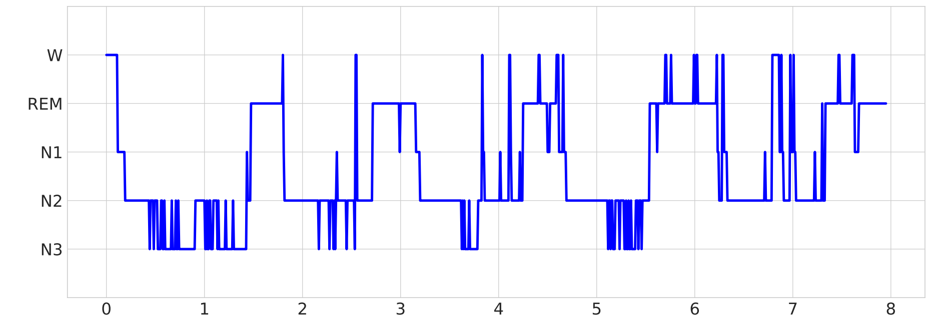
Hypnogram

- **Dataset:** Physionet Challenge 2018 (PC18) [Ghassemi et al. 2018]

Downstream tasks on clinical EEG

Sleep staging:

Predict sleep stage from EEG
(5-class: W, N1, N2, N3, R)



Hypnogram

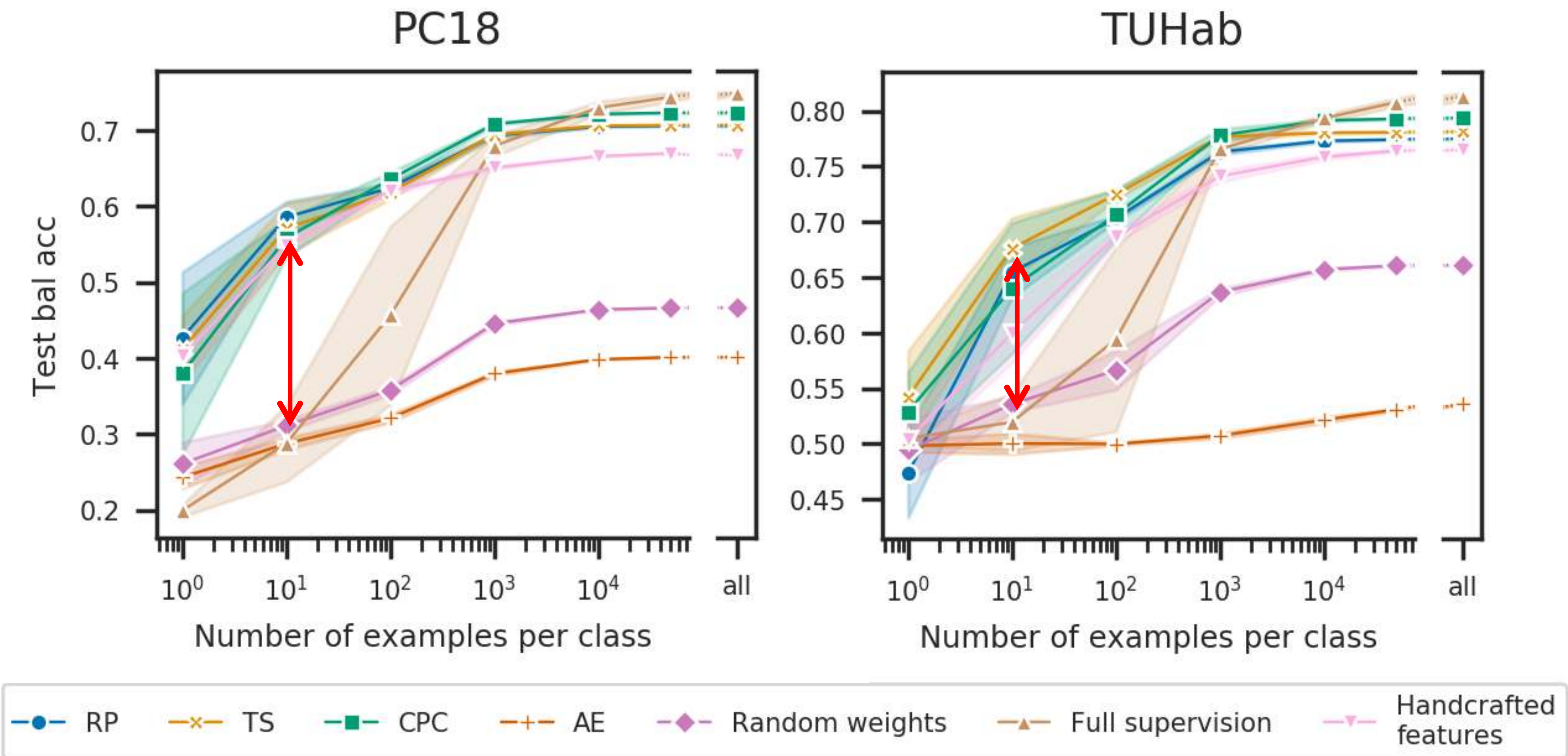
- **Dataset:** Physionet Challenge 2018 (PC18) [Ghassemi et al. 2018]

Pathology detection:

Is someone's EEG pathological?
(2-class: normal, abnormal)

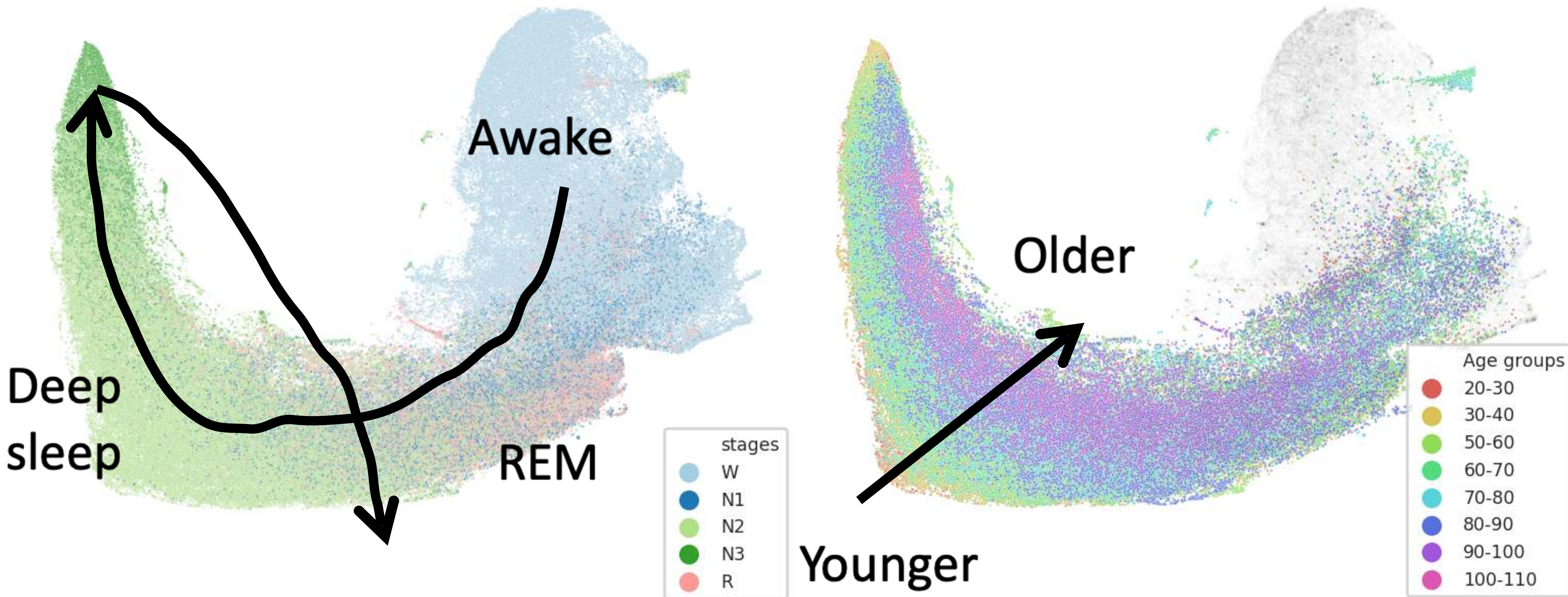
- **Dataset:** TUH Abnormal EEG (TUHab) [López 2017]

Results: Prediction accuracy



SSL is **better than full supervision** when limited data is available, and **competitive** when all data is available.

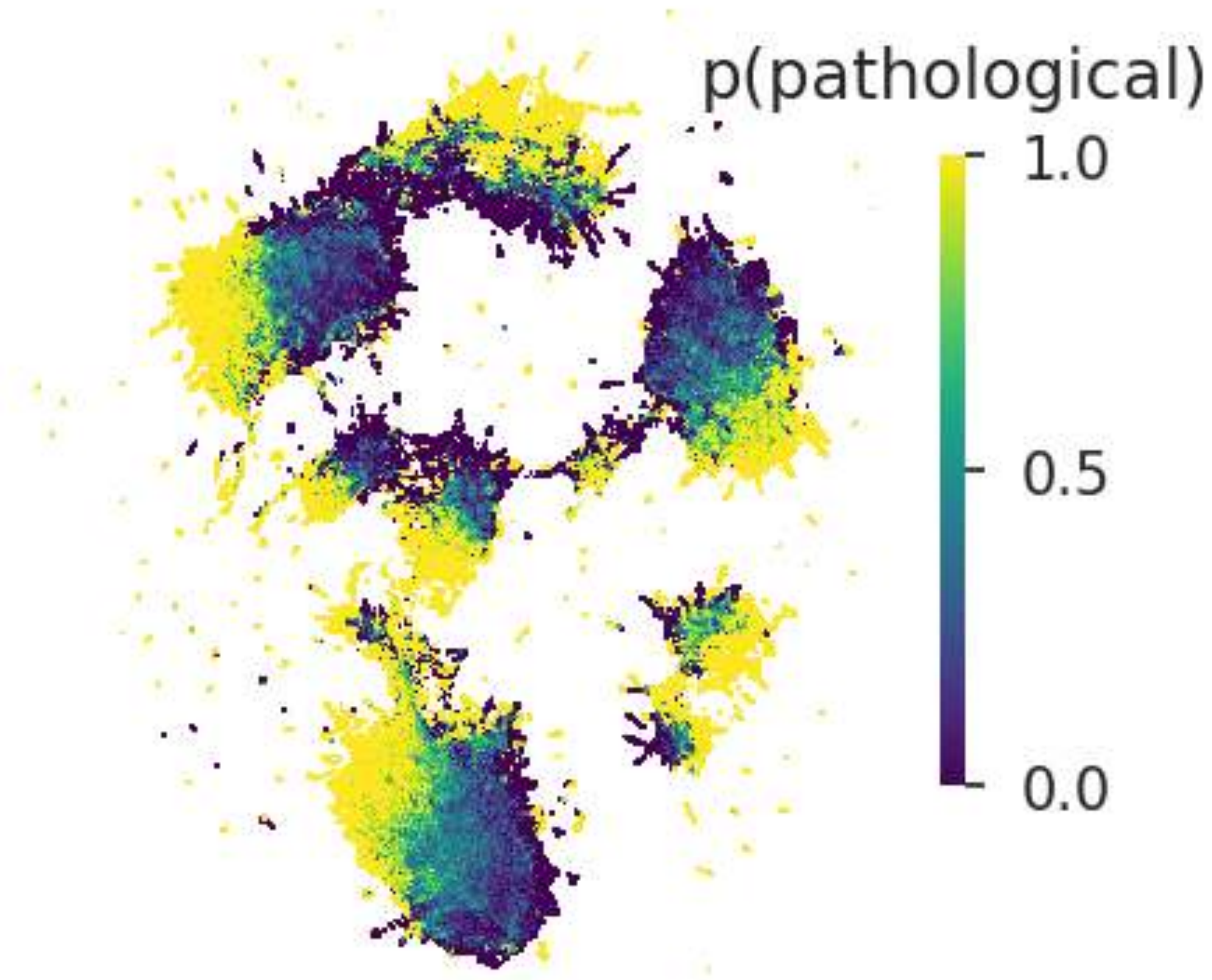
Results on sleep EEG

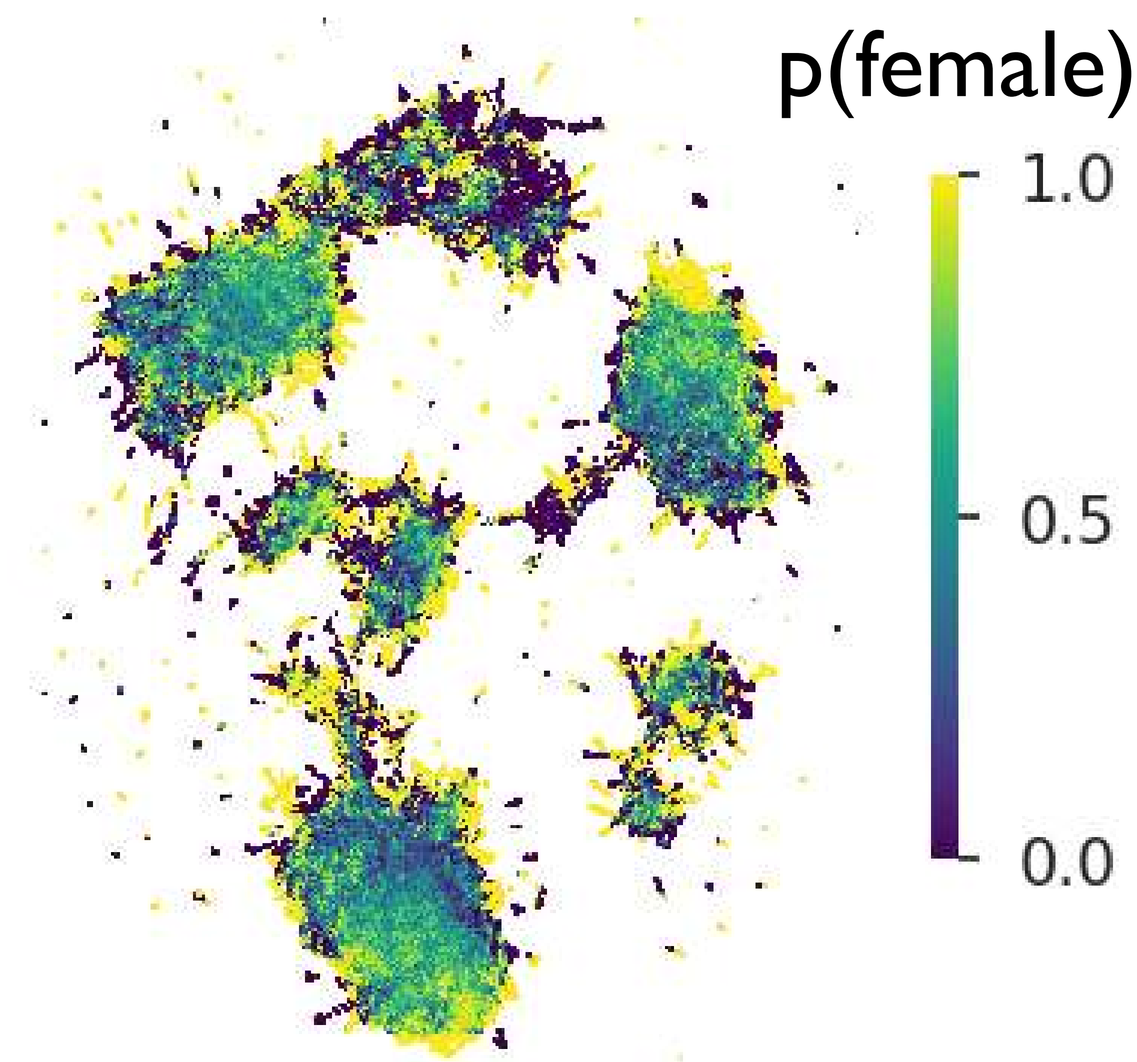


SSL can **uncover structure** without human supervision

[Banville et al. MLSP 2019]

Results on TUH data





SSL can **uncover clinically-relevant structure**
without human supervision

Conclusion

- Neuroscience signals are under exploited
- Need for better models and tools
- Need more interdisciplinary work (CS, ML, stats, neuro, physics…)
- If you want the maths look at papers…

Thanks !

S. Chambon, V. Thorey, P. J. Arnal, E. Mignot, A. Gramfort. (2018), DOSED: a deep learning approach to detect multiple sleep micro-events in EEG signal, J. Neuroscience Methods

S. Chambon, M. Galtier, P. Arnal, G. Wainrib, A. Gramfort (2018), A deep learning architecture for temporal sleep stage classification using multivariate and multimodal time series, IEEE Trans. Neural Systems and Rehabilitation Engineering

T. Dupré la Tour, T. Moreau, M. Jas, A. Gramfort, Multivariate Convolutional Sparse Coding for Electromagnetic Brain Signals, (2018), Proc. NeurIPS Conf.

M. Jas, T. Dupré la Tour, U. Simsekli, A. Gramfort, Learning the Morphology of Brain Signals Using Alpha-Stable Convolutional Sparse Coding, (2017), Proc. NeurIPS Conf.

Banville, H., Chehab, O., Hyvärinen, A., Engemann, D. and Gramfort, A. (2020), Uncovering the structure of clinical EEG signals with self-supervised learning, J. Neural Engineering

Contact http://alexandre.gramfort.net

GitHub : @agramfort *Twitter : @agramfort*

Support ERC SLAB, ANR THALAMEEG ANR-14-NEUC-0002-01
NIH R01 MH106174, DFG HA 2899/21-1.