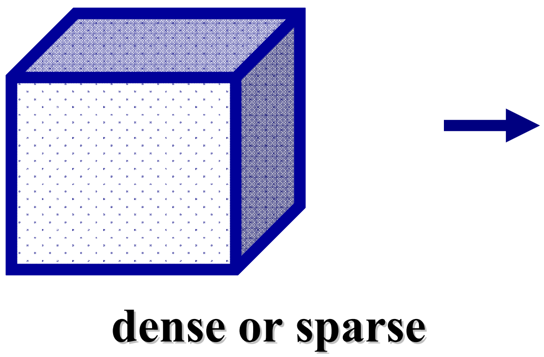
(Some) Challenges in Tensor Mining

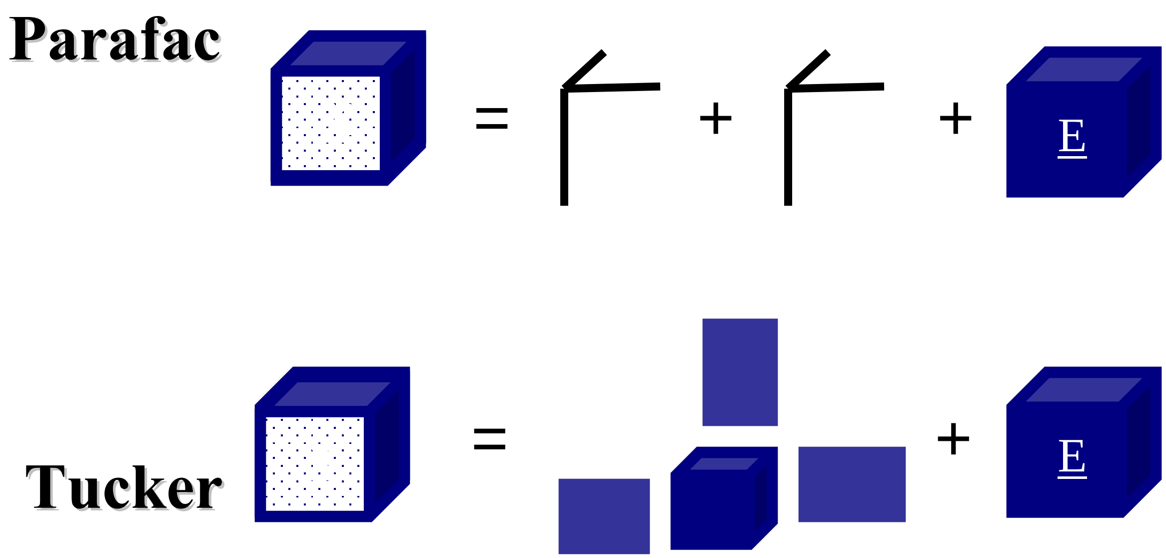
Evrim Acar
Sandia National Labs., Livermore, CA

Sandia National Laboratories

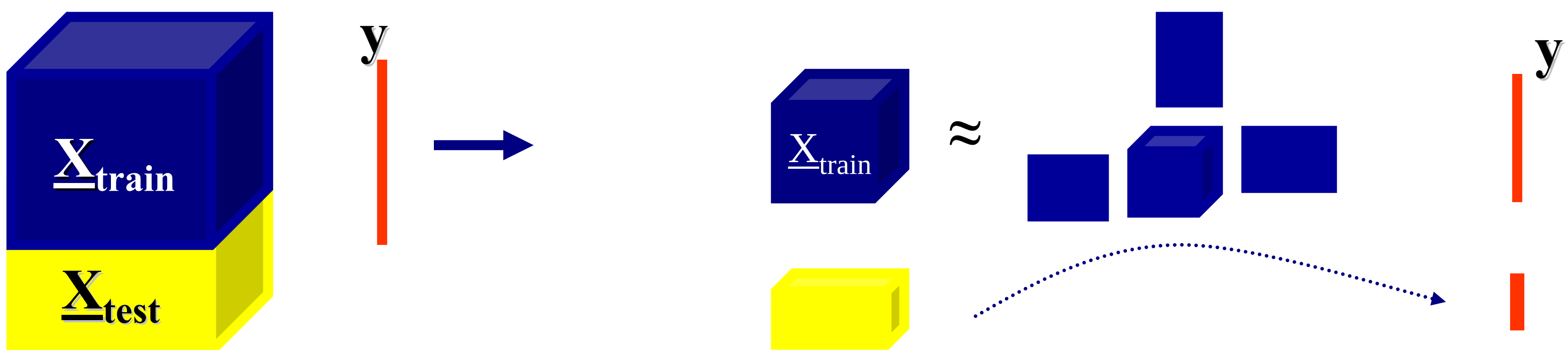
Tensor Mining

unsupervised





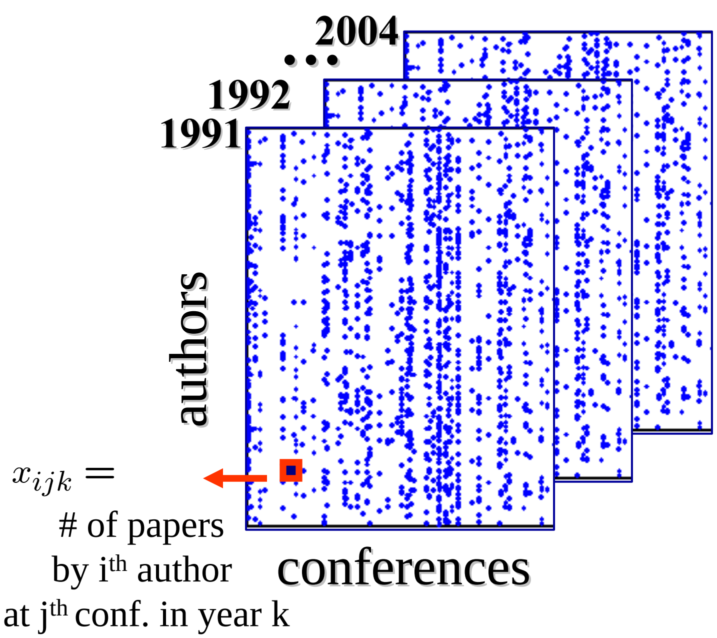
supervised



App I: Social Networks Analysis

Joint work with
T.G. Kolda and D. M. Dunlavy

- In social networks, we are interested in modeling relationships (links) evolving over time.
- Example:
  - DBLP dataset: Authors x Conferences x Years (10K x 2K x 14: ~0.1% dense)



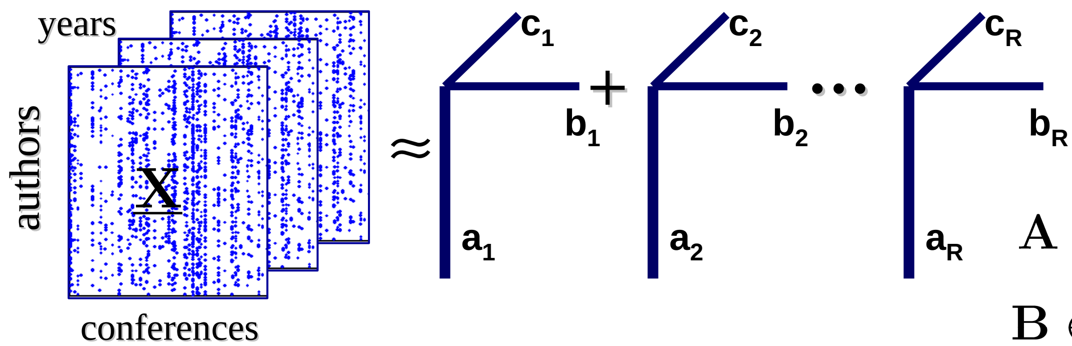
Q1: Can we use tensor decompositions to model the data and extract meaningful underlying factors?

Q2: Can we predict who is going to publish at which conferences in future?

(Link Prediction in time)

Modeling DBLP using PARAFAC

$\underline{\mathbf{X}} \in \mathbb{R}^{I \times J \times K}$



$$\mathbf{A} \in \mathbb{R}^{I \times R} = \left[\mathbf{a}_1 \cdots \mathbf{a}_R\right]$$
$$\mathbf{B} \in \mathbb{R}^{J \times R} = \left[\mathbf{b}_1 \cdots \mathbf{b}_R\right]$$
$$\mathbf{C} \in \mathbb{R}^{K \times R} = \left[\mathbf{c}_1 \cdots \mathbf{c}_R\right]$$

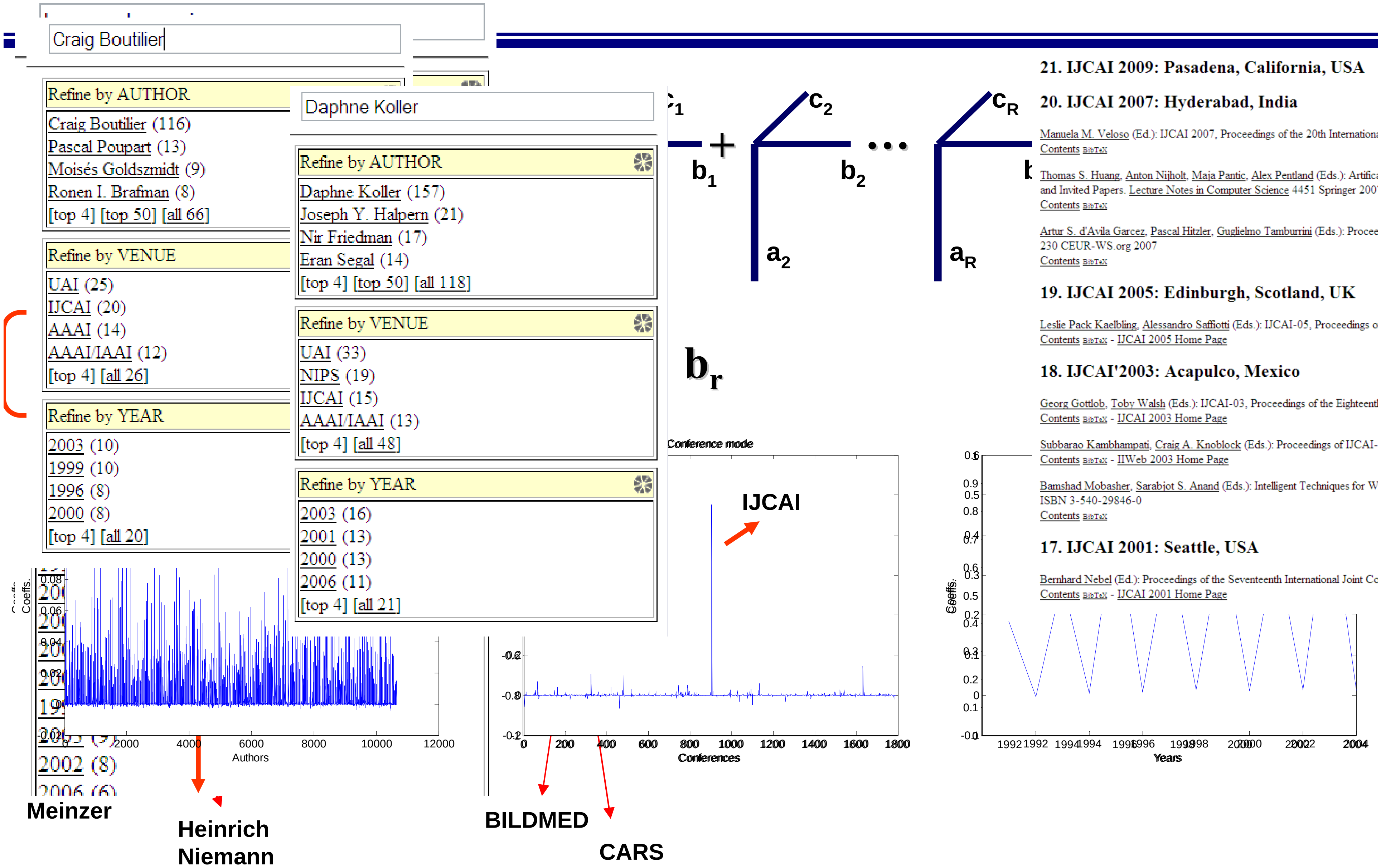
$$\underline{\mathbf{X}} \approx \sum_{r=1}^{R} \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r$$

$$\min_{\mathbf{A},\mathbf{B},\mathbf{C}} \left\| \underline{\mathbf{X}} - \left(\sum_{r=1}^{R} \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r \right) \right\|^2$$

$$\| \underline{\mathbf{X}} \| = \sum_{k=1}^{K} \sum_{j=1}^{J} \sum_{i=1}^{I} x_{ijk}^2$$

- Solve using a gradient-based optimization approach
- Initialization:
  - first two modes using svd, $R \leq I, J$
  - last mode: random, $R > K$

Components make sense!



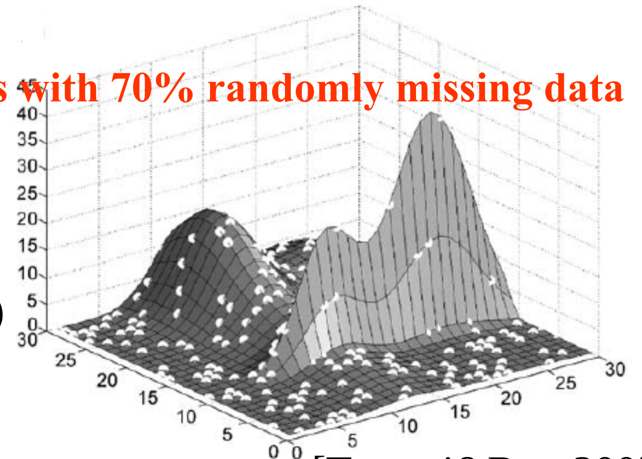
What if data is a Sparse tensor with Missing entries?

- Sparse Data: $$\min_{\mathbf{A},\mathbf{B},\mathbf{C}} \left\| \underline{\mathbf{X}} - \left(\sum_{r=1}^{R} \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r \right) \right\|^2$$

$$\mathbf{A} \in \mathbb{R}^{I \times R} = \left[\mathbf{a}_1 \cdots \mathbf{a}_R \right]$$

$$\mathbf{B} \in \mathbb{R}^{J \times R} = \left[\mathbf{b}_1 \cdots \mathbf{b}_R \right]$$

Success with 70% randomly missing data



[Tomasi&Bro, 2005]

- Missing Data [Kiers, 1997; Tomasi & Bro, 2005] :

$$\min_{\mathbf{A},\mathbf{B},\mathbf{C}} \left\| \underline{\mathbf{W}} * \left(\underline{\mathbf{X}} - \left(\sum_{r=1}^{R} \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r \right) \right) \right\|$$

$$w_{ijk} = \begin{cases} 1, & \text{if } x_{ijk} \text{ not missing,} \\ 0, & \text{if } x_{ijk} \text{ missing.} \end{cases}$$

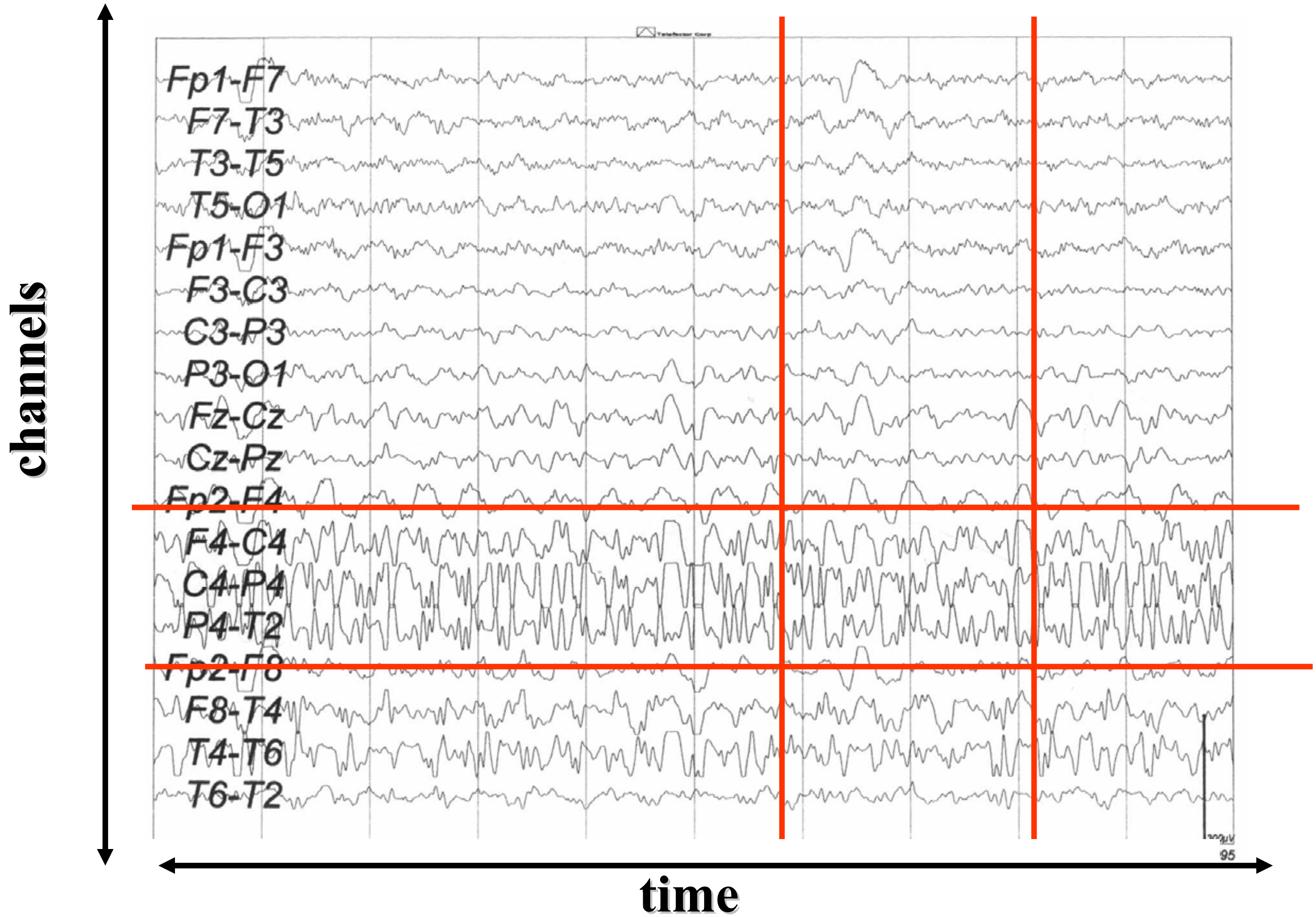
- Sparse & Missing:

$$\min_{\mathbf{A},\mathbf{B},\mathbf{C}} \left\| \underline{\mathbf{W}} * \left(\underline{\mathbf{X}} - \left(\sum_{r=1}^{R} \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r \right) \right) \right\|^2 + ???$$

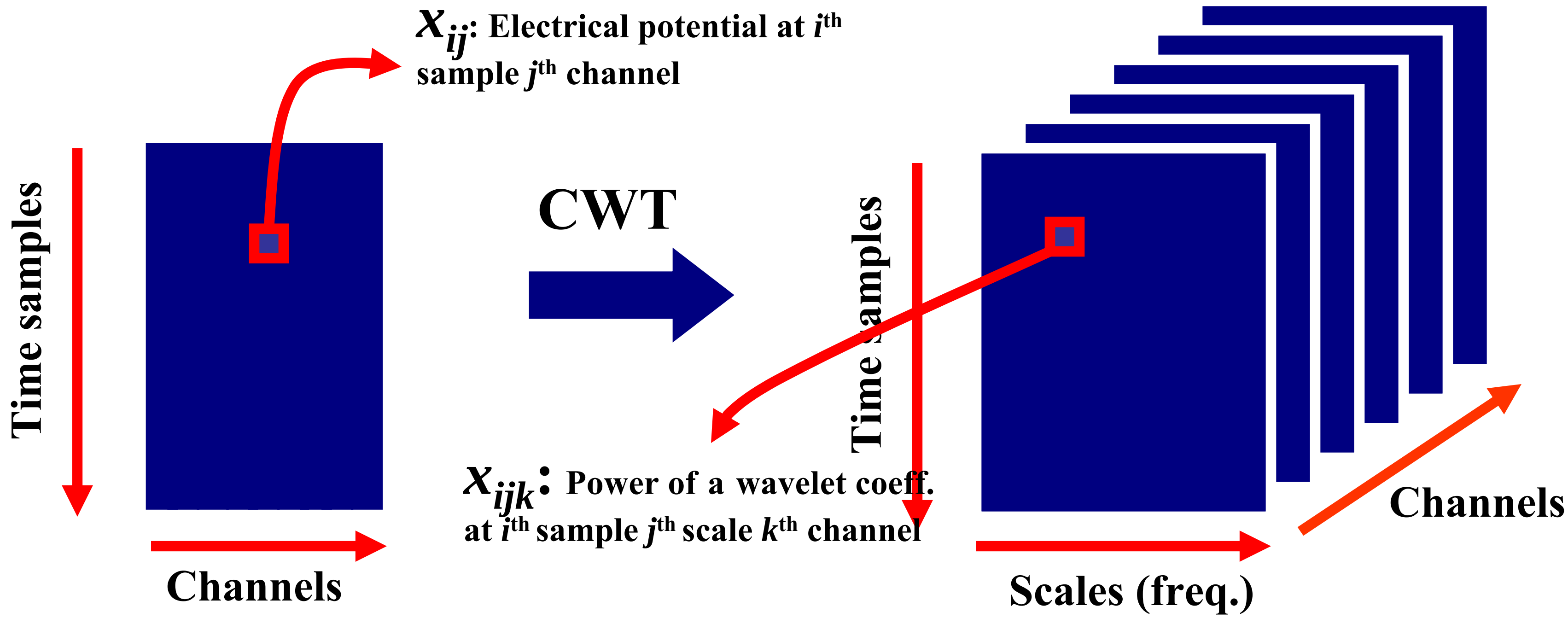
$$w_{ijk} = \begin{cases} 1, & \text{if } x_{ijk} \text{ not missing,} \\ 0, & \text{if } x_{ijk} \text{ missing.} \end{cases}$$

App II: Understanding Epileptic Seizures

Joint work with
R. Bro, B. Yener, C. A. Bingol. H. Bingol



Epilepsy Tensors

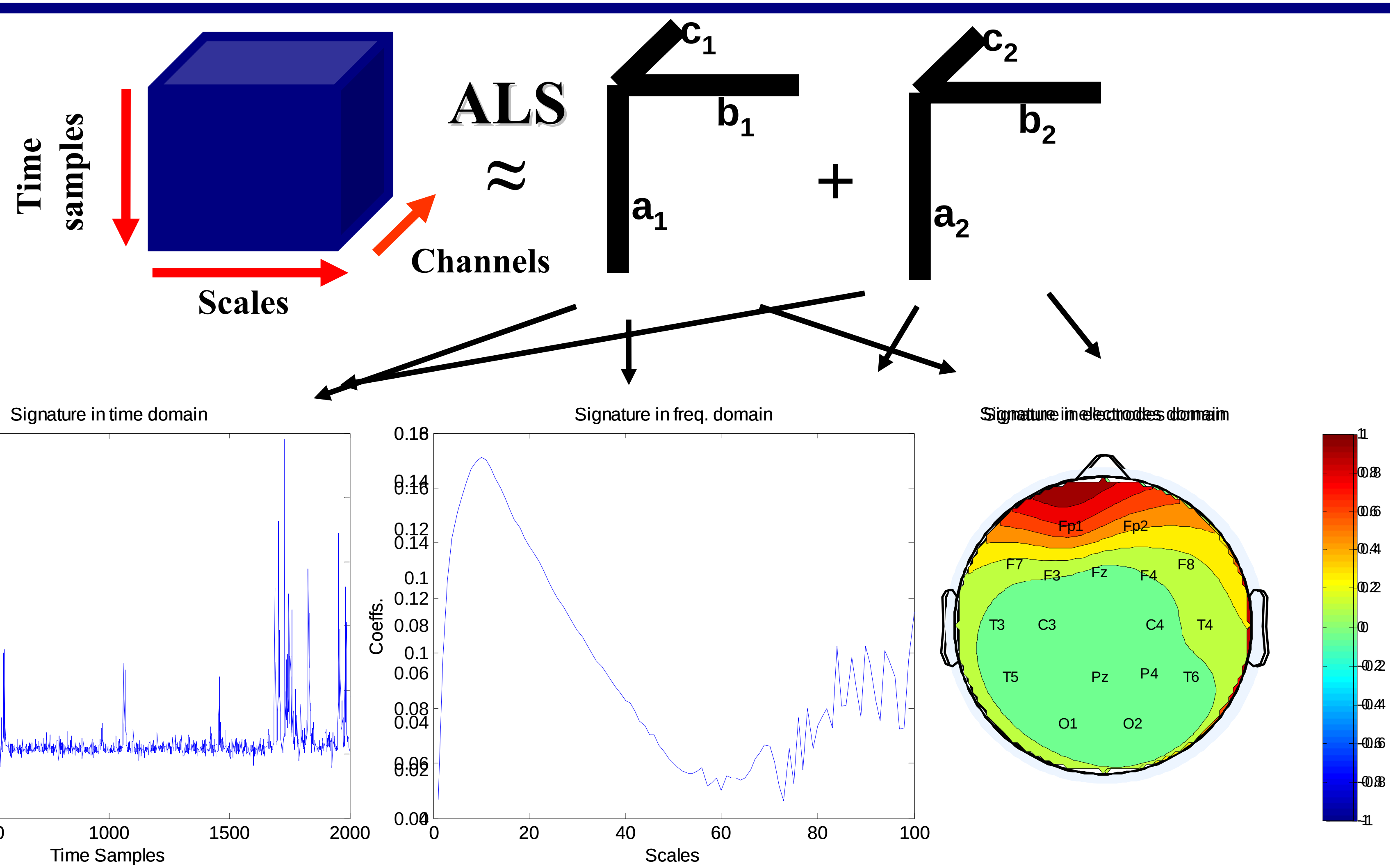


• Data rearranged as a three-way array using continuous wavelet transform (CWT):

- Let c_{ijk} be the wavelet coefficient at time sample *i* at scale *j* for the k^{th} channel.
- An Epilepsy Tensor is a three-way array, $\underline{X}$, where each entry x_{ijk} is computed as:

$$x_{ijk} = |c_{ijk}|^2$$

Epilepsy Focus Localization

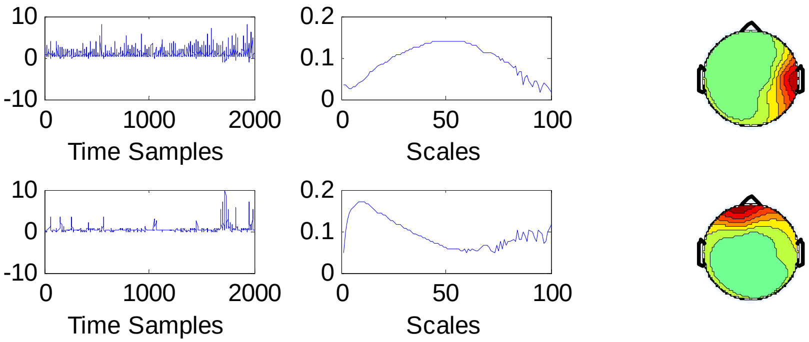


Acar et al’07, De Vos et al’07

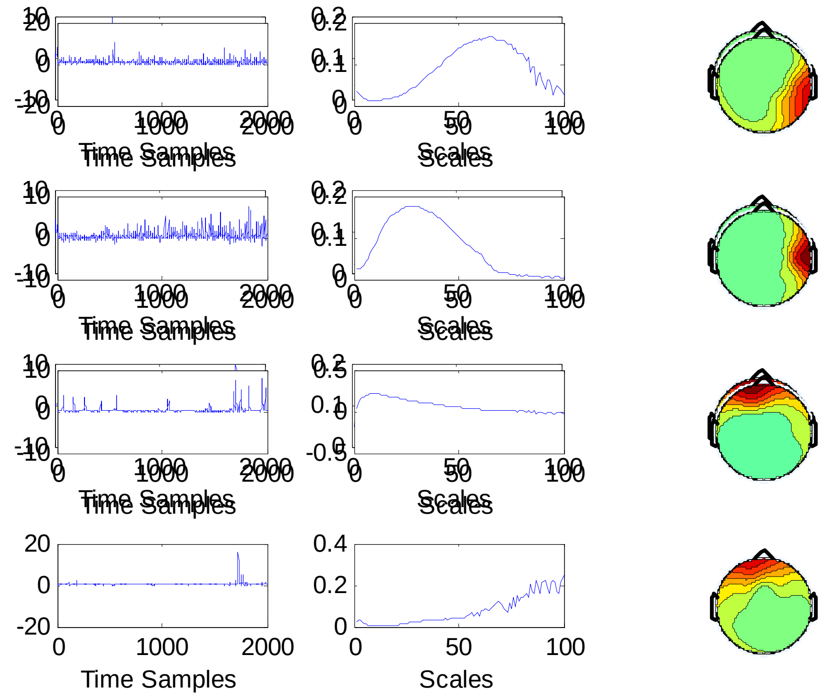
How many components?

$$\underline{\mathbf{X}} \approx \sum_{r=1}^{R} \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r$$

$R = 2$



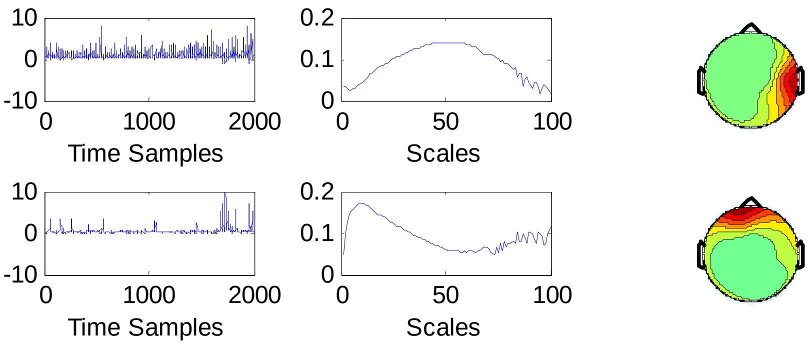
$R = 4$



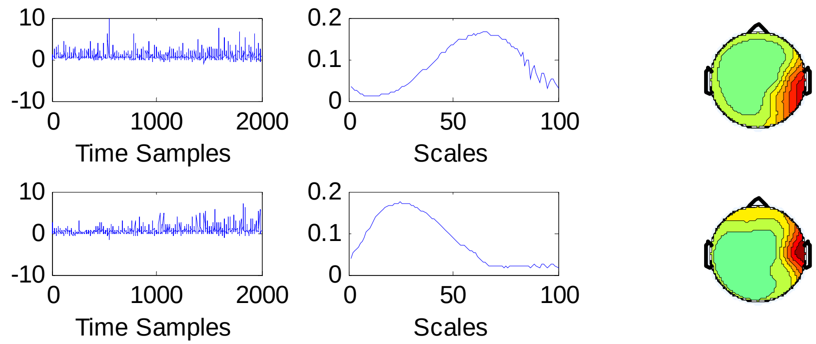
How to initialize?

$$\underline{\mathbf{X}} \approx \sum_{r=1}^{2} \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r$$

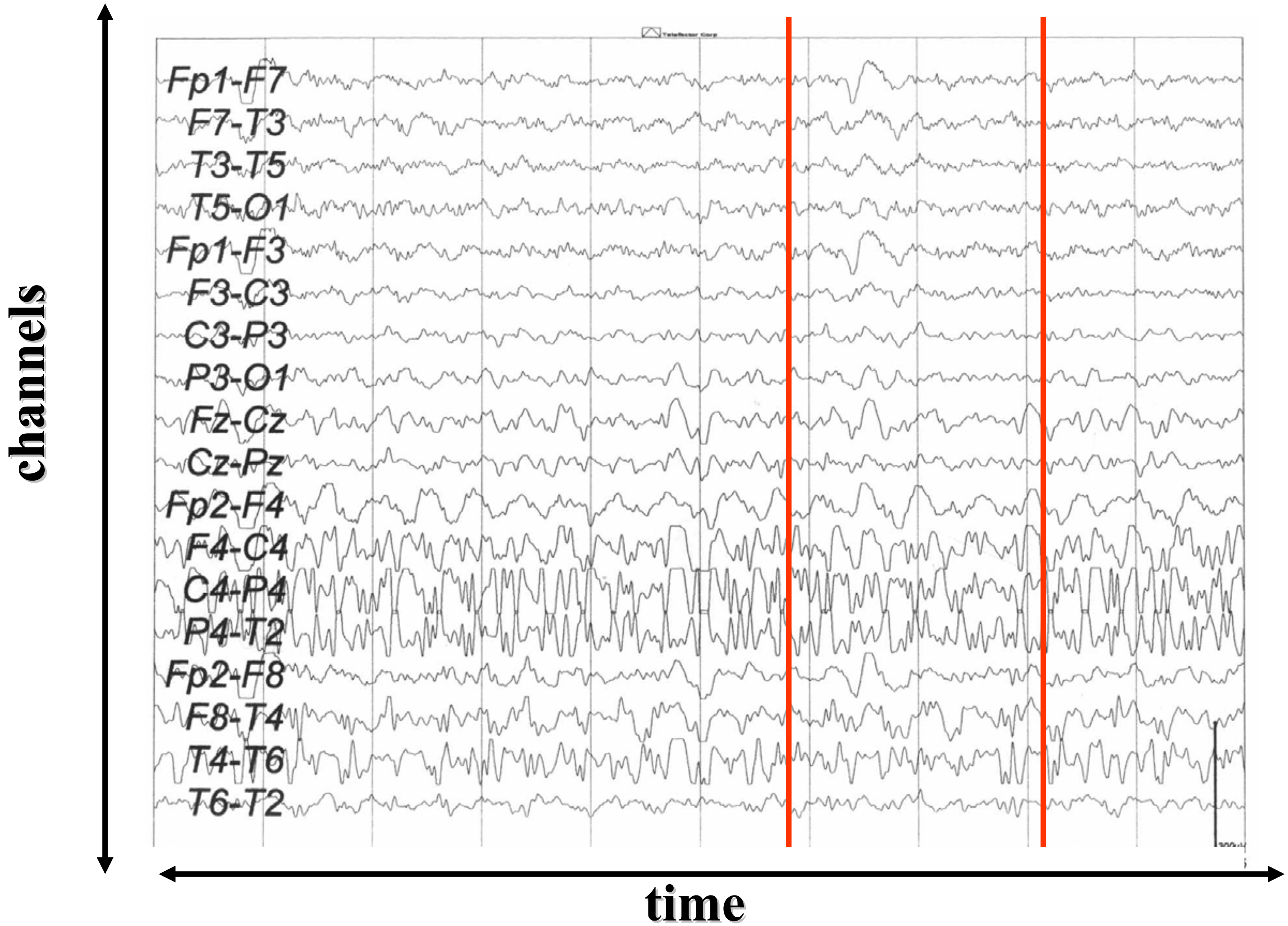
HOSVD



RANDOM

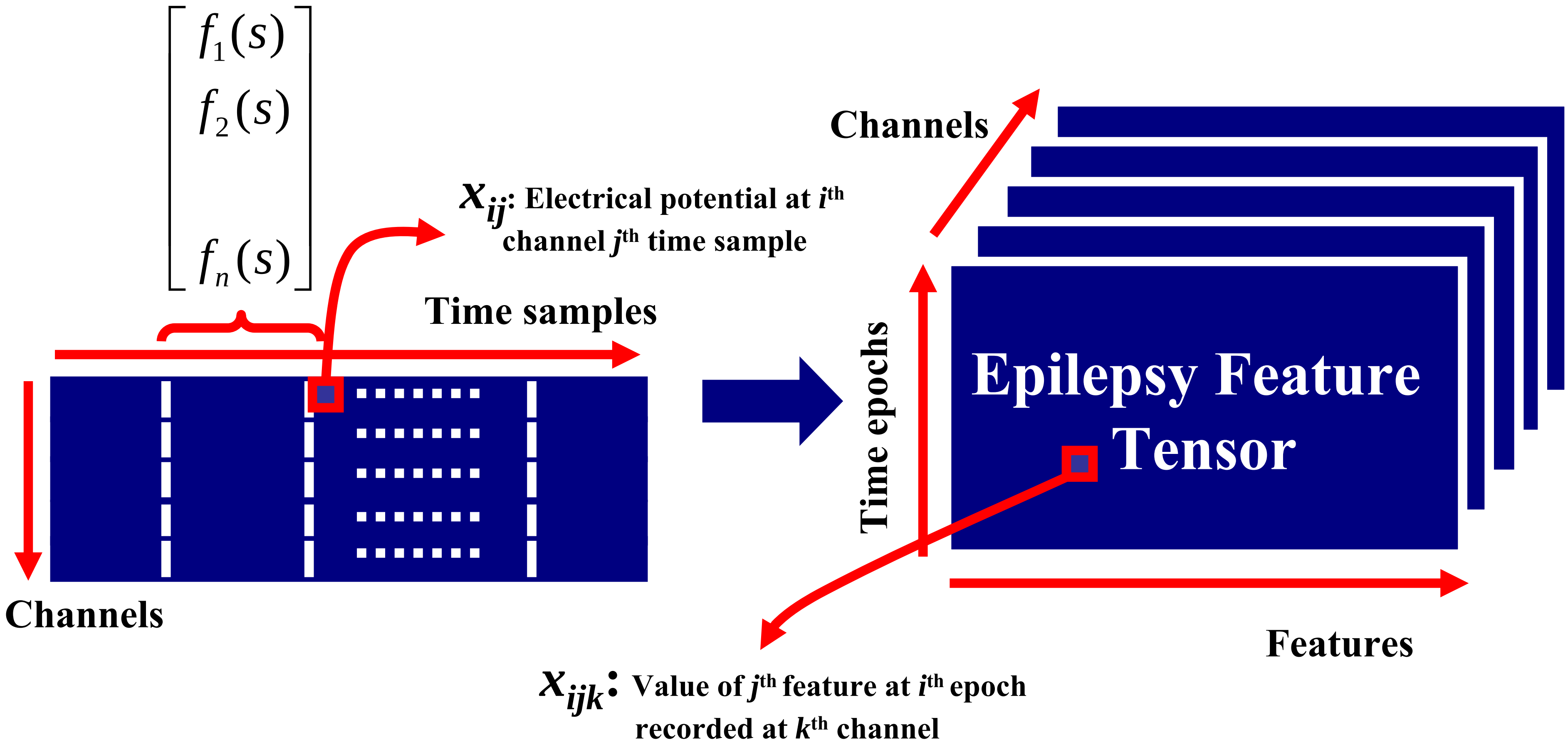


Understanding Epileptic Seizures



Epilepsy Feature Tensor

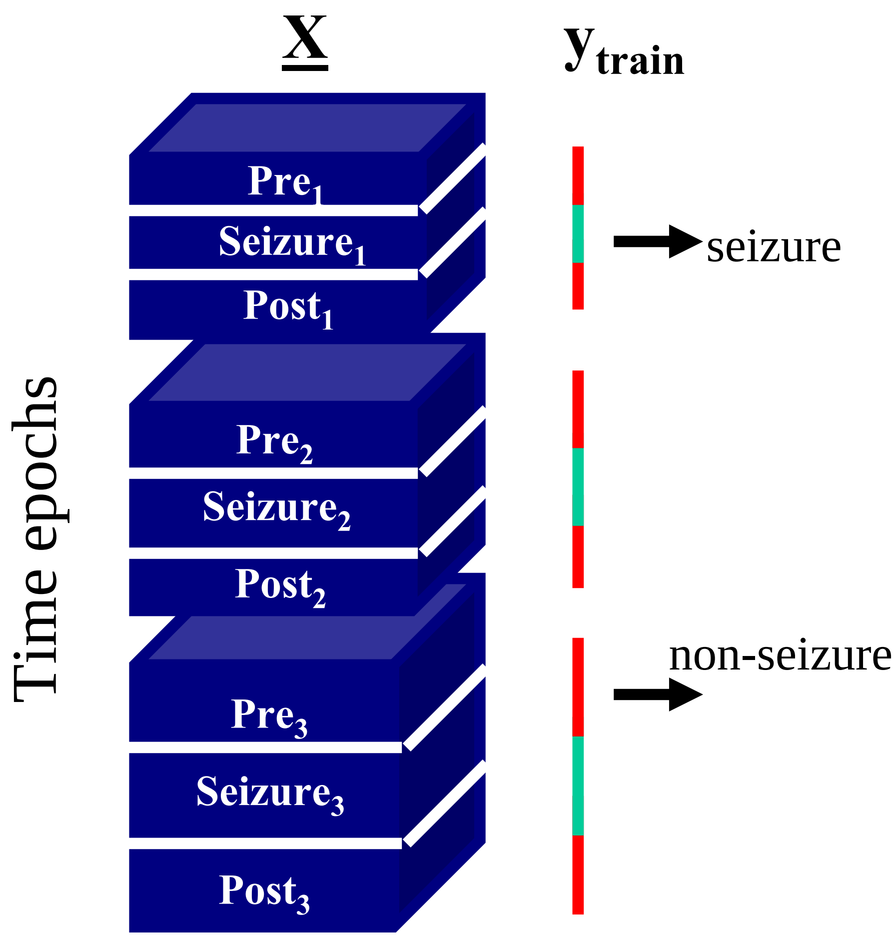
- Construction of an Epilepsy Feature Tensor from multi-channel EEG



Seizure Recognition

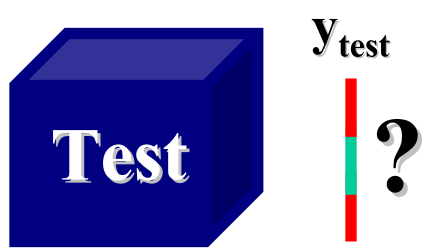
Training Set

- Build a model using the training set $\underline{\mathbf{X}}$ and the labels **y**.



Test Set

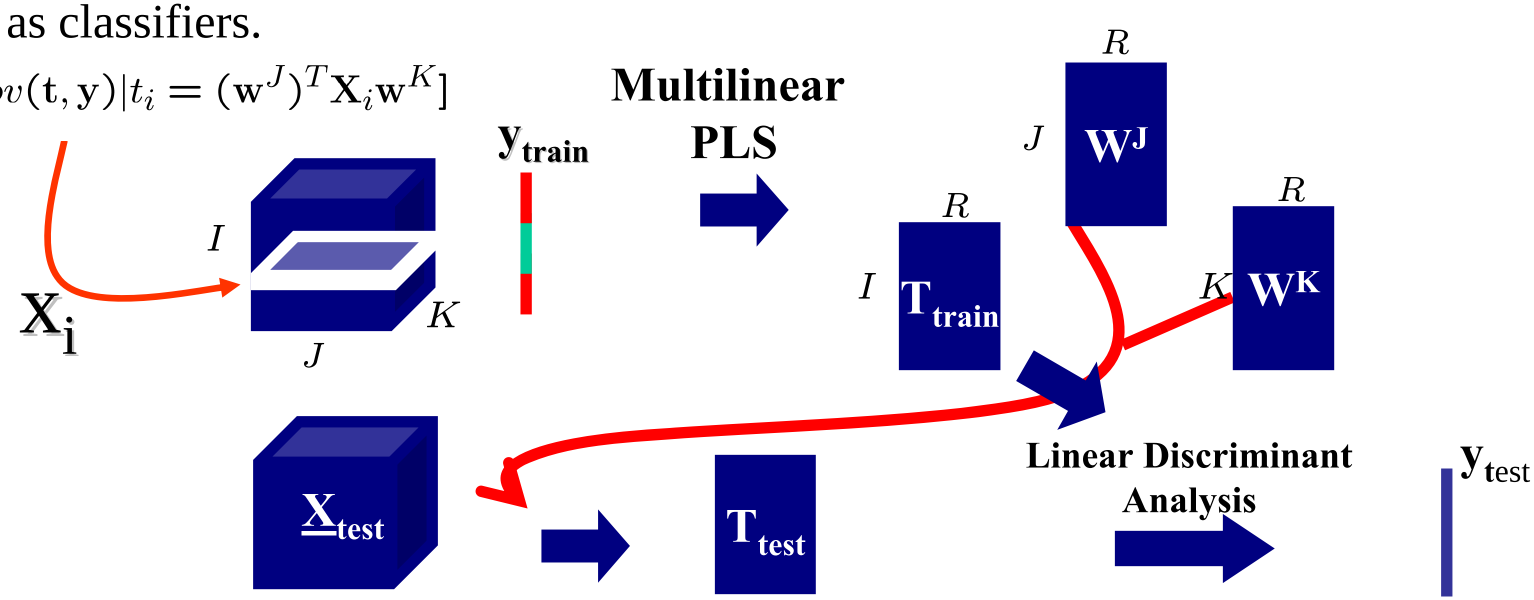
- Predict the labels of new recordings.



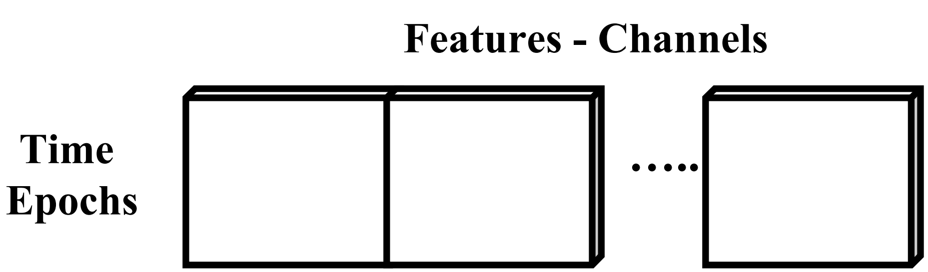
Multiway Classification(?)

- Potential Approaches
  - Modify multiway regression models, e.g., multilinear PLS [Bro, 1996; Bro et al., 2001], as classifiers.

$$\max_{\mathbf{w}^J,\mathbf{w}^K} [cov(\mathbf{t},\mathbf{y}) | t_i = (\mathbf{w}^J)^T \mathbf{X}_i \mathbf{w}^K]$$



  - Unfold the data and apply two-way classification, e.g., SVM.



Some challenges are …

- **Handling Sparse Data with Missing Entries:**
  - We need models to capture the underlying sparse factors in sparse tensors with missing entries.
- **Determining the Rank:**
  - Important also in practice.
- **Initialization:**
  - Algorithms suffer from the local minima problem. In practice, we may end up interpreting our results differently.
- **Supervised learning on tensors:**
  - We need classification models for tensors as good as the state-of-the-art two-way classification approaches such as SVMs.

Thank you!

- **References:**
  - **Social Networks Analysis**: [Tensor toolbox & Poblano toolbox (by Sandia)]
    - Acar, Kolda and Dunlavy, An Optimization Approach for Fitting Canonical Tensor Decompositions, SAND2009-0857, Feb. 2009.
  - **Understanding Epileptic Seizures:** [PLS toolbox (by Eigenvector Research)]
    - Acar, Bingol, Bingol, Bro and Yener, Multiway Analysis of Epilepsy Tensors, *Bioinformatics*, 23(13): i10-i18, 2007.
    - Acar, Bingol, Bingol, Bro and Yener, Seizure Recognition on Epilepsy Feature Tensor, *Proc. 29th Int. Conf. IEEE Engineering in Medicine and Biology Society*, 2007.
  - **Survey:**
    - Acar and Yener, Unsupervised Multiway Data Analysis: A Literature Survey, *IEEE Transactions on Knowledge and Data Engineering*, 21(1): 6-20, 2009.
- **Contact:**

Evrim Acar, Sandia National Laboratories, **eacarat@sandia.gov**