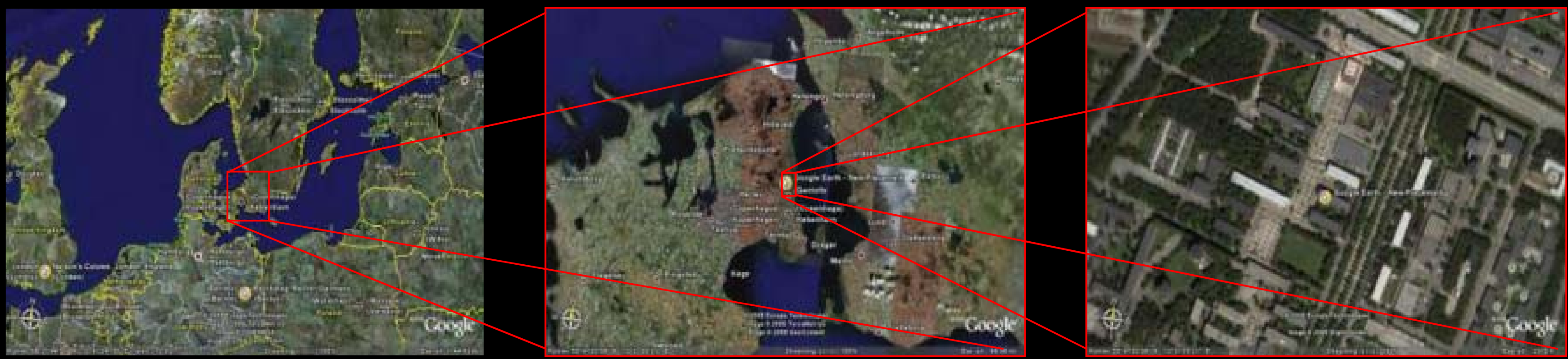
Multi-linear decomposition of event related EEG data

Morten Mørup

Section for Cognitive Systems

DTU Informatics

Technical University of Denmark



The material presented joint work with

Kristoffer Hougaard Madsen

Anne-Marie Dogonowski

Lek-Heng Lim

Sidse M. Arnfred

Lars Kai Hansen

Hartwig Siebner

Christoph S. Herrmann and Josef Parnas

Univariate statistical analysis in NeuroImaging

Problems:

1)Multiple comparisons, i.e. many voxels/channels tested.

2)What is the true number of independent tests, i.e. voxels/ channels are highly correlated

3)Data extremely noisy, i.e. low SNR rendering tests insignificant.

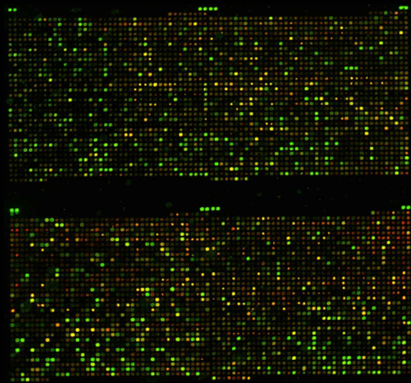
Need for advanced multivariate methods that can efficiently extract the underlying structures and how they interact.

This problem is no different than the problems encountered in general in Modern Massive Datasets (MMDS)

$\mathbf{X}^{Space \times Time}$

NeuroInformatics

$\mathbf{X}^{Gene\ \ seq. \times Samples}$

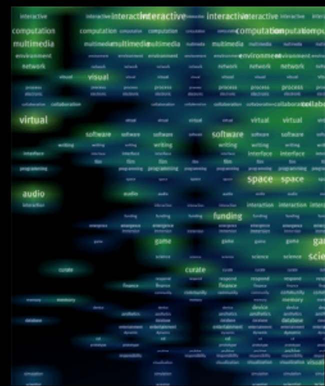


BioInformatics

$\mathbf{X}^{Webpages \times Webpages}$

ComplexNetworks

$\mathbf{X}^{Term \times Document}$



WebDataMining

Unsupervised Learning attempts to find the hidden causes and underlying dynamics in the data.

(Multivariate exploratory analysis – driving hypotheses)

Goal of unsupervised Learning

(Ghahramani & Roweis, 1999)

- Perform dimensionality reduction
- Build topographic maps
- Find the hidden causes or sources of the data
- Model the data density
- Cluster data

Purpose of unsupervised learning

(Hinton and Sejnowski, 1999)

- Extract an efficient internal representation of the statistical structure implicit in the inputs

2008

WIRED MAGAZINE: 16.07

SCIENCE : DISCOVERIES

The End of Theory: The Data Deluge Makes the Scientific Method Obsolete

By Chris Anderson 06.23.08

Illustration: Marian Bantjes

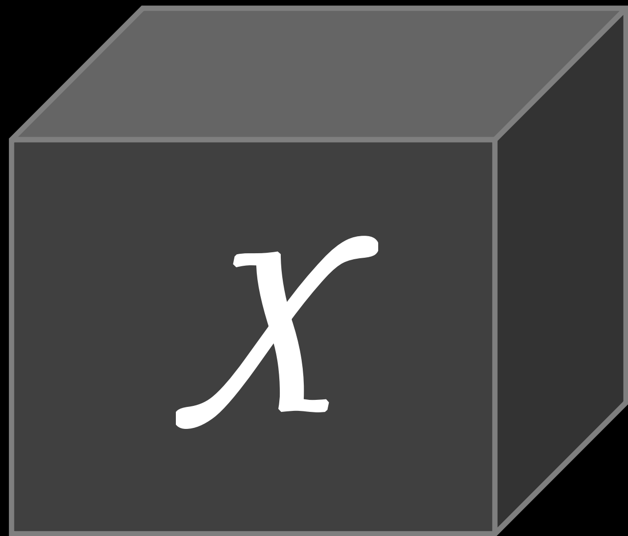
THE PETABYTE AGE:
Sensors everywhere. Infinite storage. Clouds of processors. Our ability to capture, warehouse, and understand massive amounts of data is changing science, medicine, business, and technology. As our collection of facts and figures grows, so will the opportunity to find answers to fundamental questions. Because in the

"All models are wrong, but some are useful."

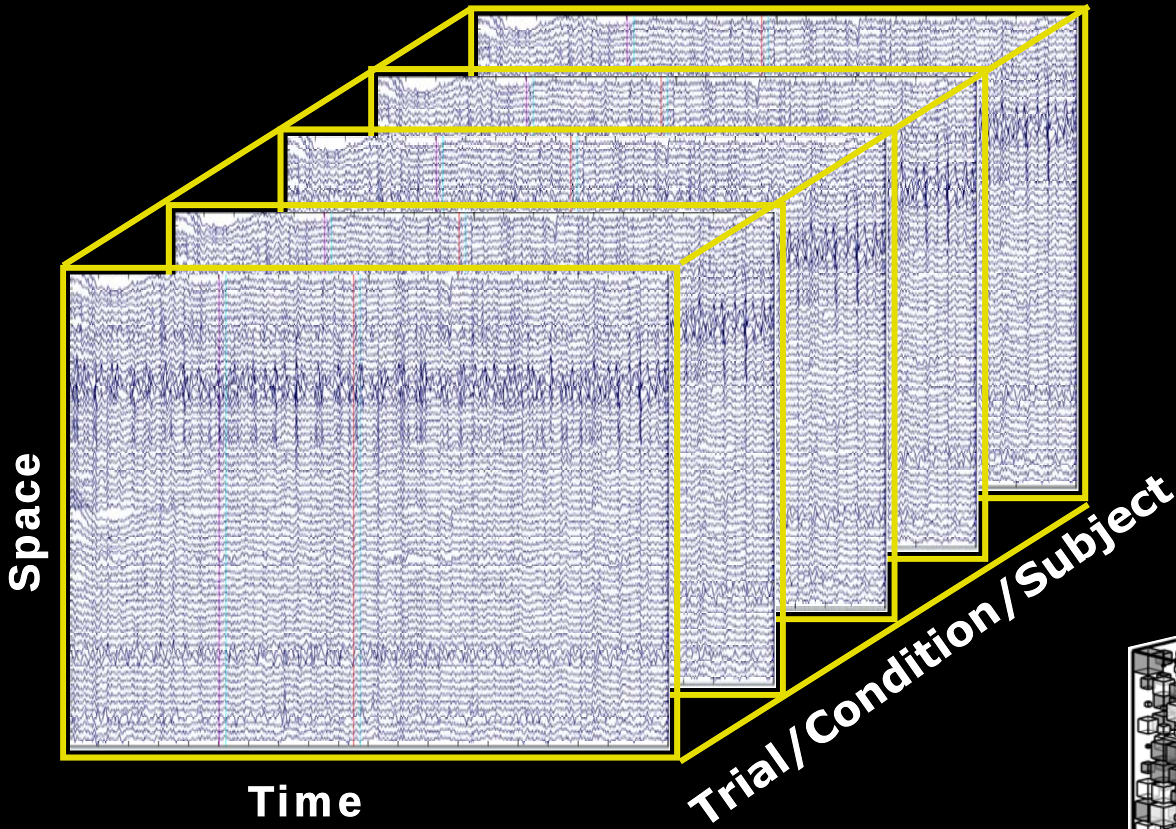
So proclaimed statistician George Box 30 years ago, and he was right. But what choice did we have? Only models, from cosmological equations to theories of human behavior, seemed to be able to consistently, if imperfectly, explain the world around us. Until now. Today companies like Google, which have grown up in an era of massively abundant data, don't

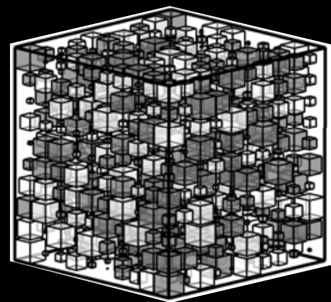
Analysis of massive amounts of data will be the main driving force of all sciences in the future!!

Unfortunately, multi-way/tensor structure has been widely ignored in many fields of research!



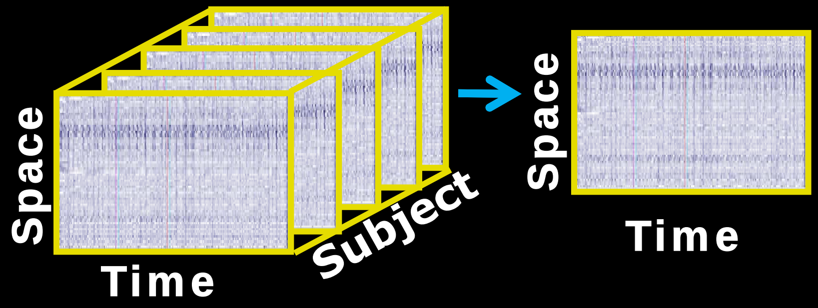
Multi-way arrays naturally emerge in NeuroImaing



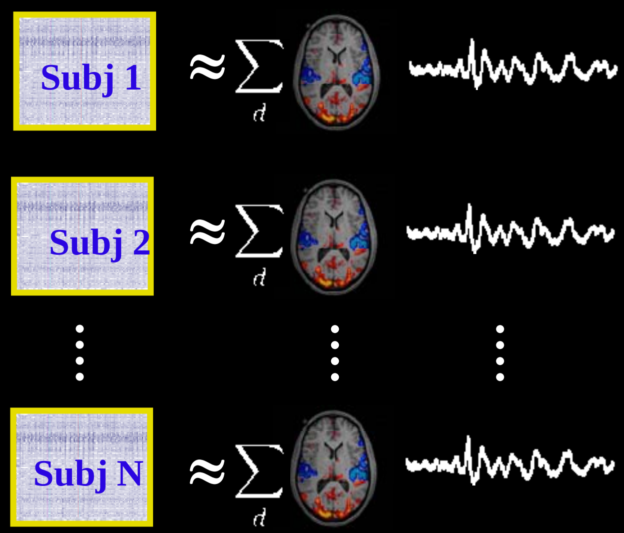


3 common ways of avoiding tensors

Preaverage

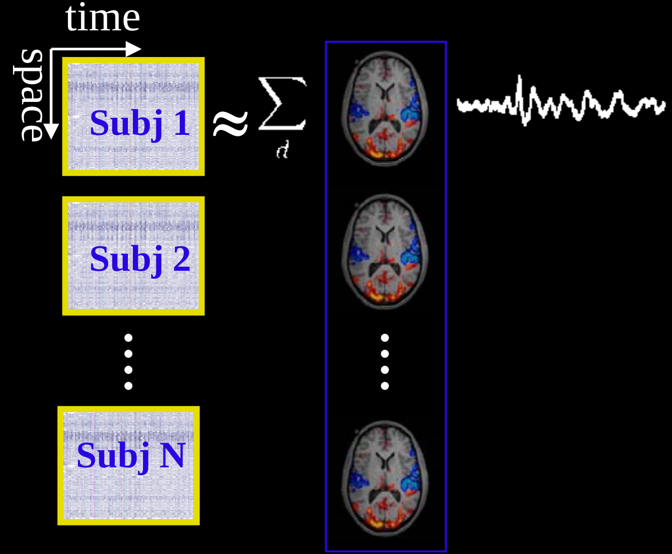


Separate Analysis

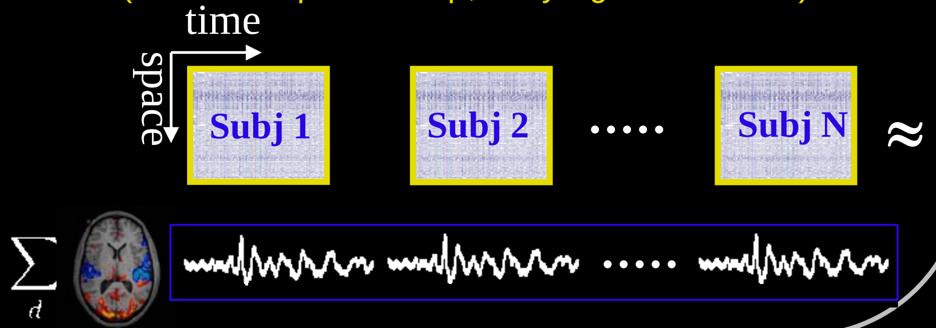


Concatenation

(identical time series varying spatial maps)

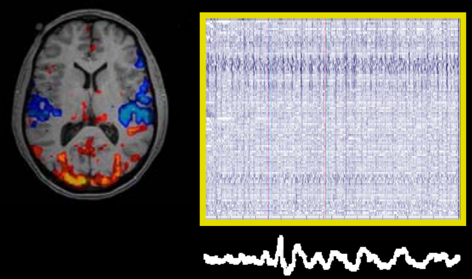


(identical spatial map, varying time series)

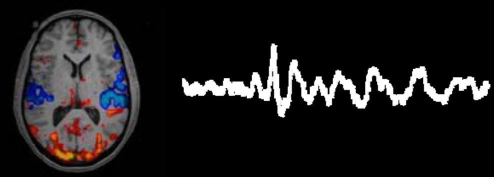


Bilinear Model:

$$\mathbf{X}^{\mathrm{Channel}\times\mathrm{Time}} \approx \sum_d \mathbf{a}_d^{Channel}\mathbf{b}_d^{\mathrm{Time}}$$



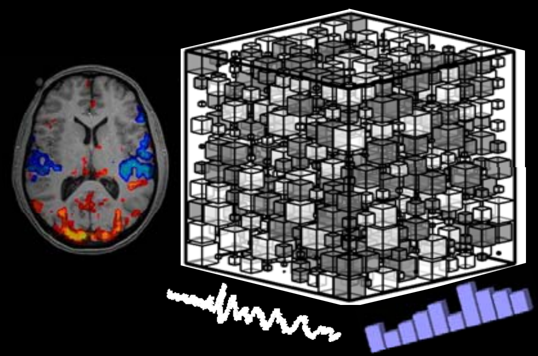
$$\sum_d$$



Assumption: Data **instantaneous** mixture of temporal signatures.
(PCA/ICA/NMF)

Trilinear Model:

$$\mathbf{X}^{\mathrm{Channel}\times\mathrm{Time}\times Trial} \approx \sum_d \mathbf{a}_d^{Channel}\mathbf{b}_d^{\mathrm{Time}}\mathbf{c}_d^{Trial}$$



$$\sum_d$$

Assumption: Data **instantaneous** mixture of temporal signatures that are expressed to **various degree** over the trials.

(Hitchcock 1927: Canonical Polyadic Form)
(Carrol and Chang; Harshman, 1970: Canonical Decomposition, Parallel Factor (CP))
(Möcks, 1988: Topographic Component Analysis TCA
(first to analyze ERP of *(channel x time x subject)*)

Bilinear decomposition not unique

$$\mathbf{X} \approx \mathbf{A}\mathbf{B}^\top = \mathbf{A}\mathbf{Q}\mathbf{Q}^{-1}\mathbf{B}^\top = \widetilde{\mathbf{A}}\widetilde{\mathbf{B}}^\top$$

Multi-linear decomposition is in general unique!!

$$\begin{aligned} \boldsymbol{X}_{(:,:,k)} \approx \boldsymbol{A}\,\mathrm{diag}(\boldsymbol{C}_{k,:})\boldsymbol{B}^T &= (\boldsymbol{A}\boldsymbol{T})\underbrace{(\boldsymbol{T}^{-1}\,\mathrm{diag}(\boldsymbol{C}_{k,:})\boldsymbol{Q})}(\boldsymbol{Q}^{-1}\boldsymbol{B}^T) \\ &= \widehat{\boldsymbol{A}}\,\mathrm{diag}(\widehat{\boldsymbol{C}}_{k,:})\widehat{\boldsymbol{B}}^T. \end{aligned}$$

Kruskal (1976, 1977) derived the following uniqueness criterion generalized to N-ways arrays in (Sidiropoulos and Bro, 2000):

$$\textit{3-way array:} \quad k_{\mathbf{A}} + k_{\mathbf{B}} + k_{\mathbf{C}} \geq 2D + 2$$

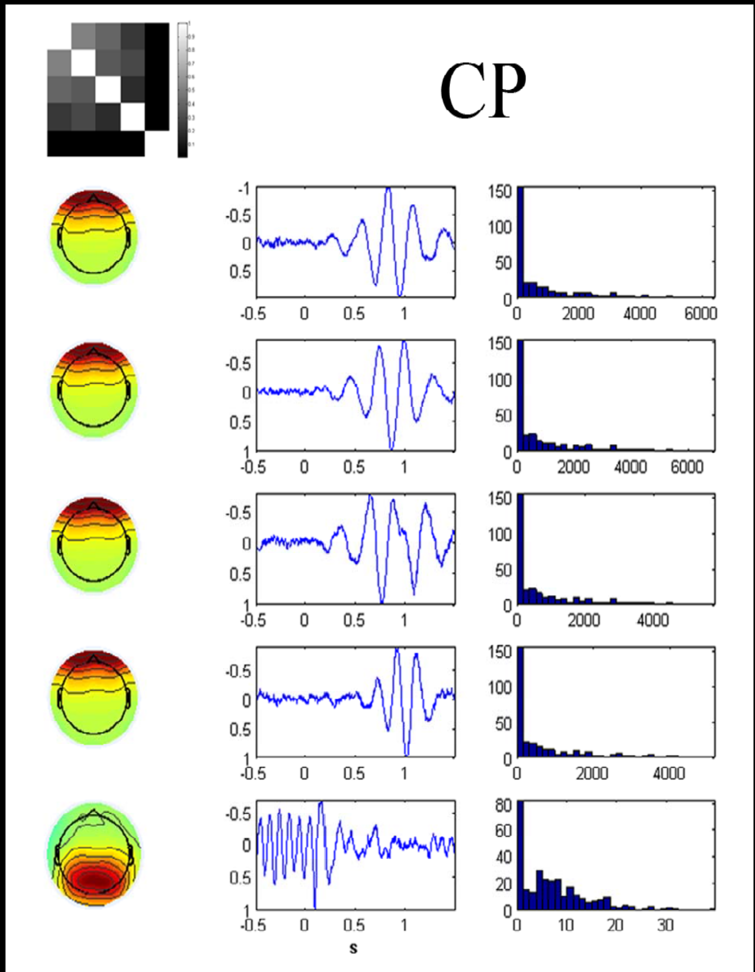
$$\textit{N-way array:} \quad \sum_n k_{\mathbf{A}^{(n)}} \geq 2D + N - 1$$

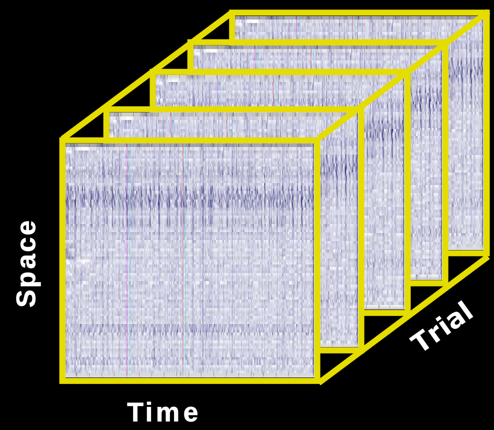
where $k_{\mathbf{A}}$ is the k-rank denoting the smallest subset of columns of $\mathbf{A}$ that is guaranteed to be linearly independent. Thus, $k_{\mathbf{A}} \leq rank(\mathbf{A})$.

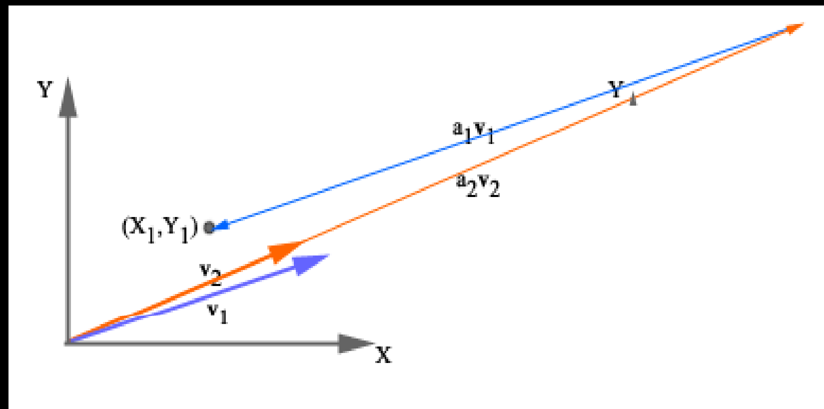
"A surprising fact is that the nonrotatability characteristic can hold even when the number of factors extracted is greater than every dimension of the three-way array." - Kruskal 1976

DTU

However, violation of multi-linearity causes degeneracy

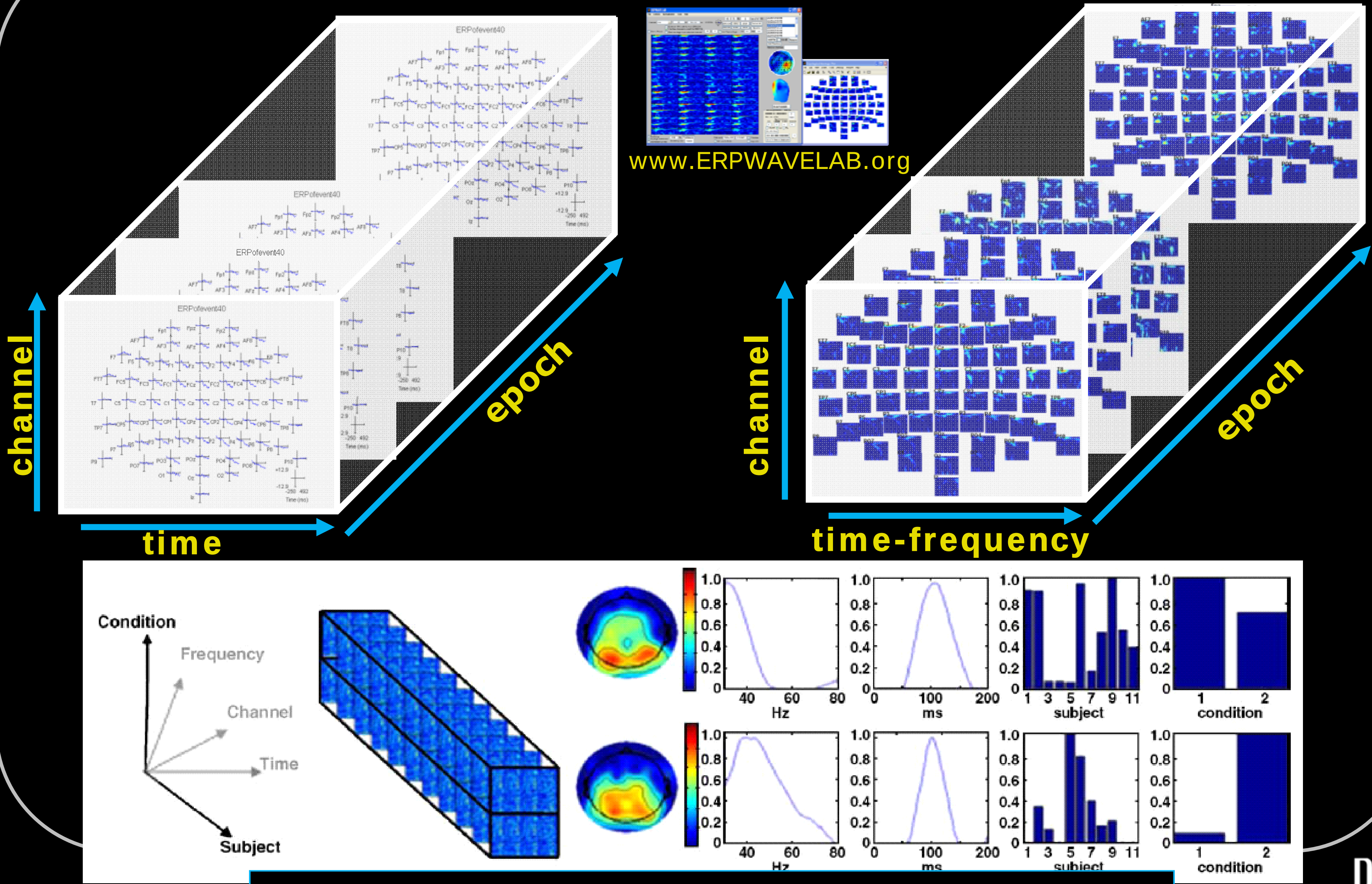






Common fixes: Impose orthogonality , regularization or consider the data in the frequency domain based on non-negative decompsoitions that are quaranteed not to degenerate.

Wavelet transformed data



Mørup, Hansen, Herman, Parnas, Arnfred, NeuroImage 2006
Mørup, Hansen, Arnfred, Journal of Neuroscience Methods, 2007

Multi-linear models are often too restrictive for the modeling of EEG/fMRI

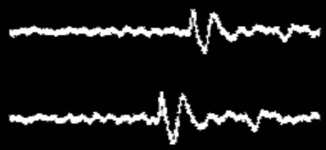
Trilinear model can encompass:

- Variability in strength over repeats

However, other common causes of variation are:

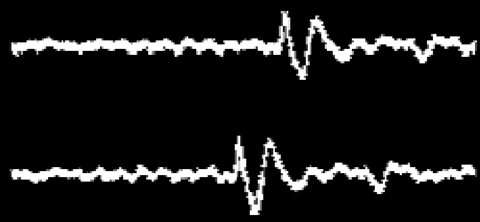
- Delay Variability
- Shape Variability

Trial 1
Trial 2



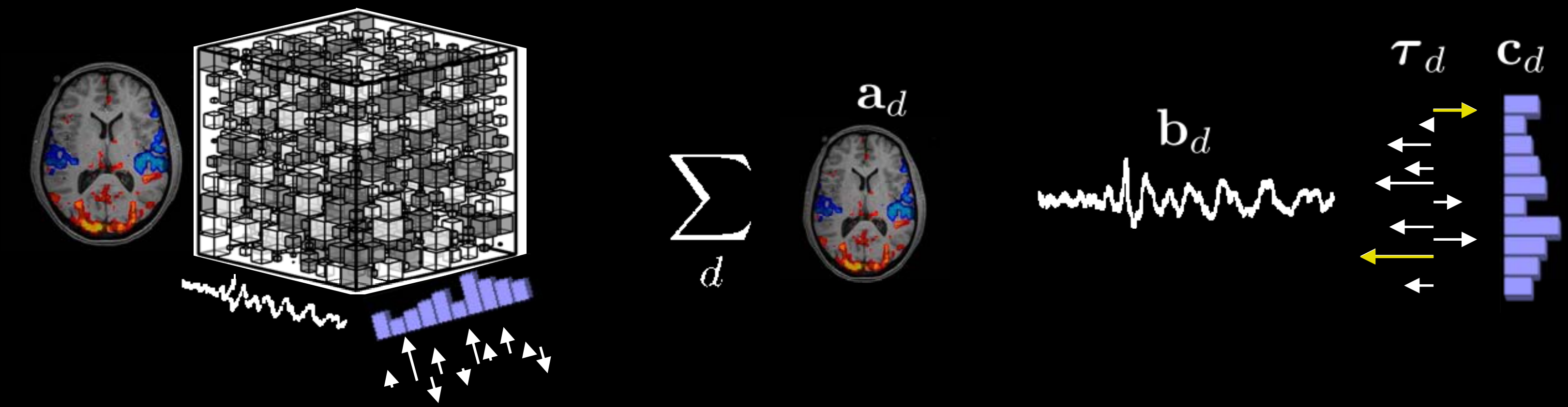
Trial 1
Trial 2

Modelling Delay Variability

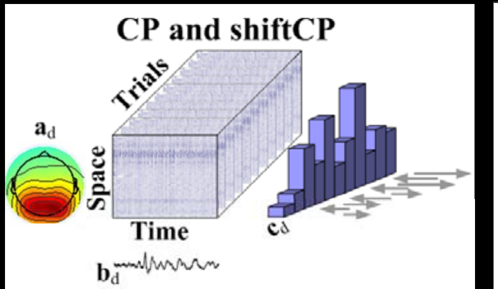


Shifted CP:

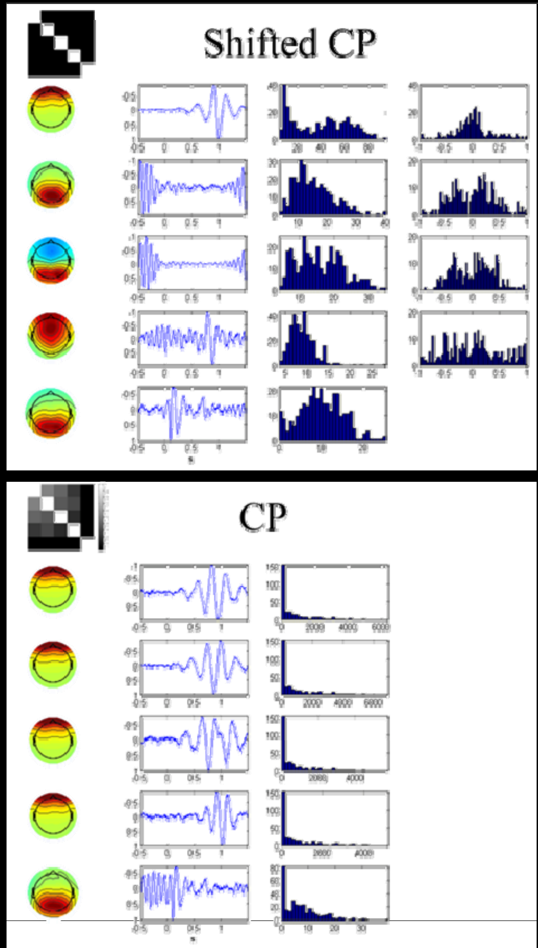
$$x_{i,k}(t) \approx \sum_d a_{i,d} b_d(t - \boldsymbol{\tau}_{k,d}) c_{k,d}$$

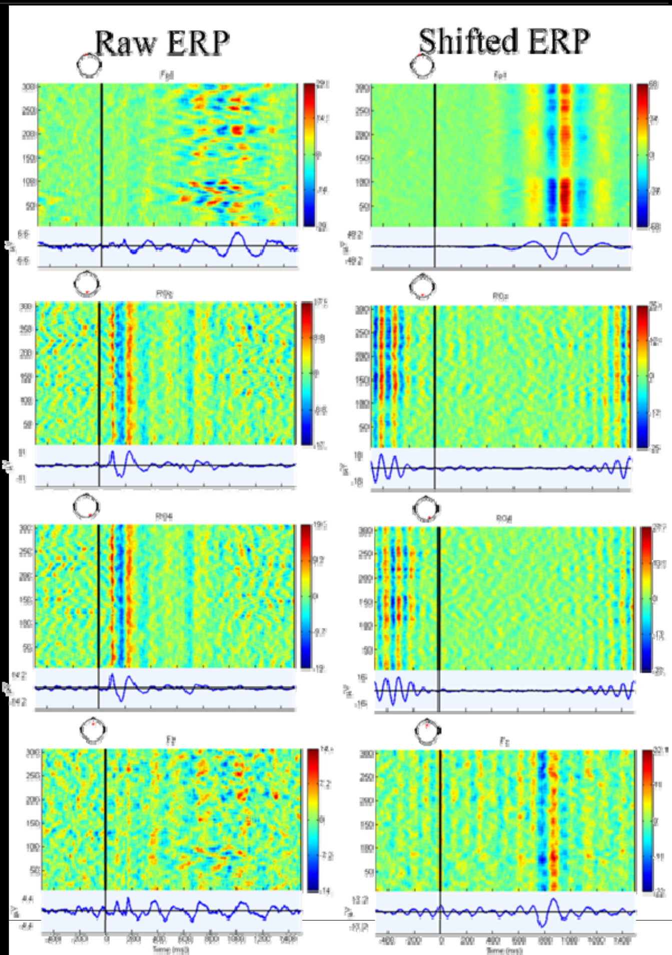


(shiftCP: Harshman, Hong and Lundy 2003)



$$x_{i,j,k} \approx \sum_{d} a_{i,d} b_{j-\tau_{k,d},d} c_{k,d}$$

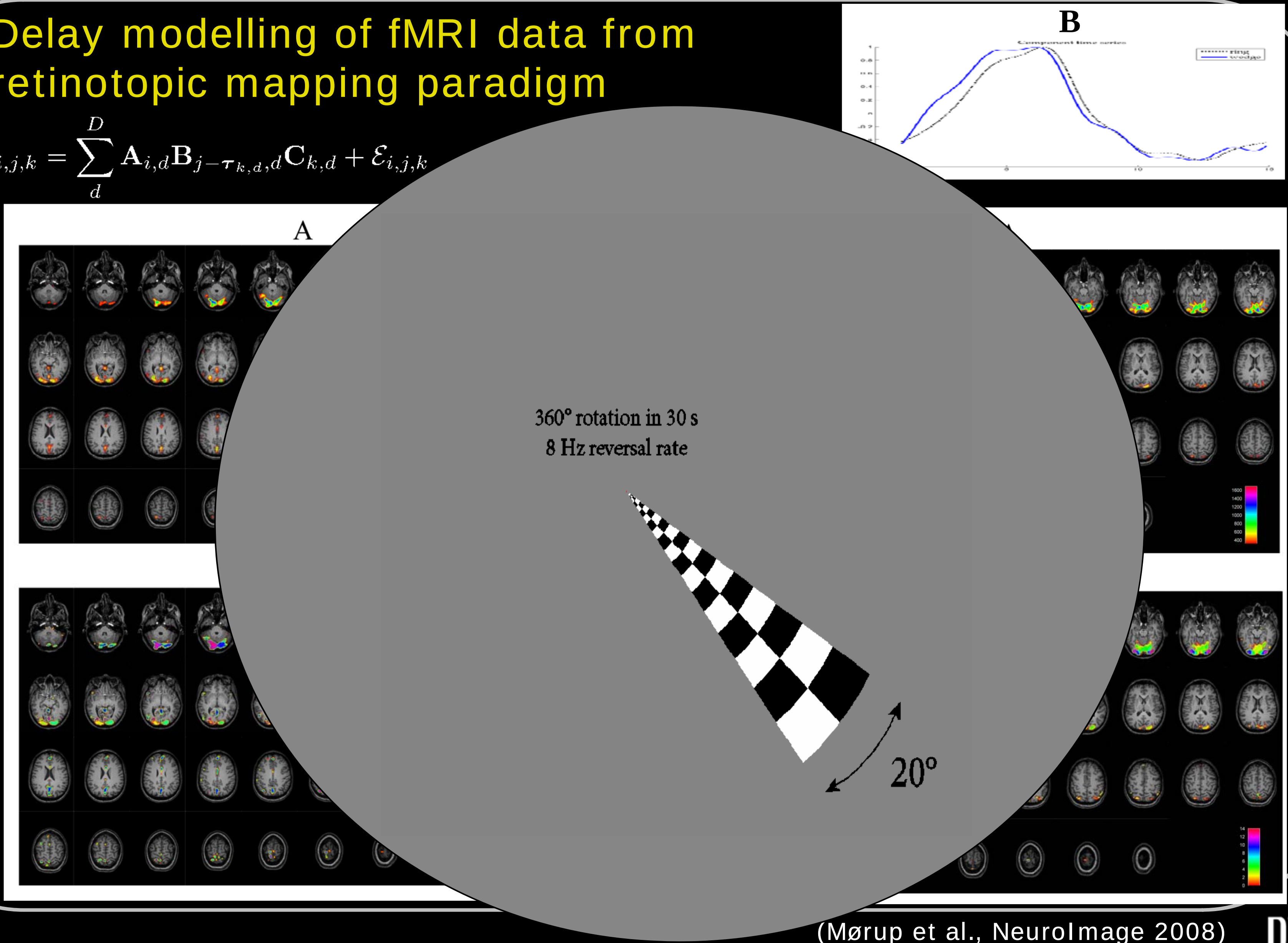




(Mørup et al.,
NeuroImage 2008)

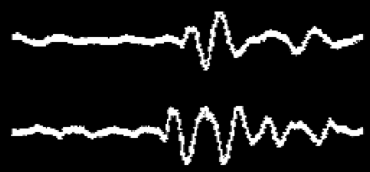
Delay modelling of fMRI data from retinotopic mapping paradigm

$$\mathcal{X}_{i,j,k} = \sum_{d}^{D} \mathbf{A}_{i,d} \mathbf{B}_{j-\tau_{k,d},d} \mathbf{C}_{k,d} + \mathcal{E}_{i,j,k}$$



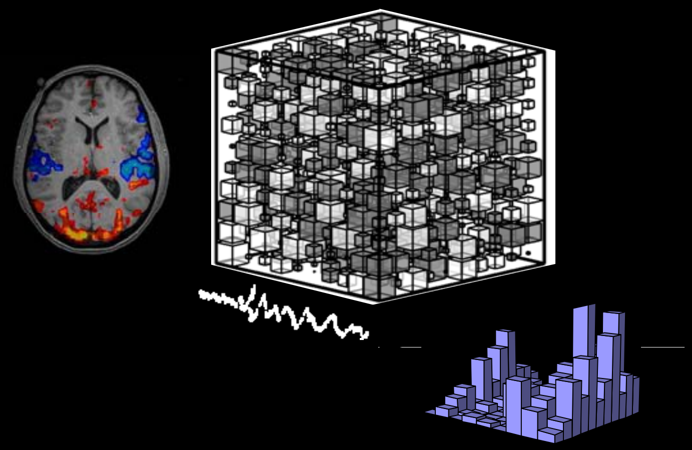
(Mørup et al., NeuroImage 2008)

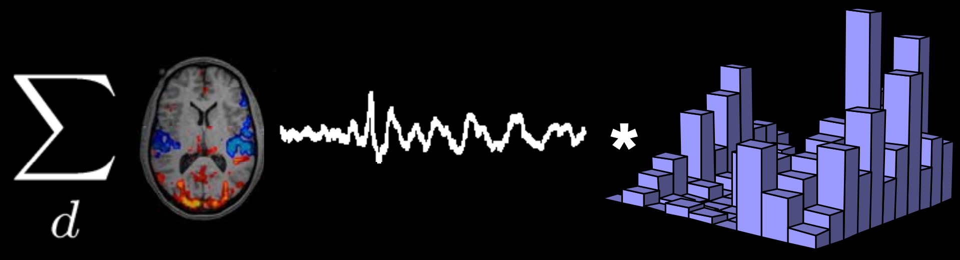
Modeling Shape (and delay) Variability



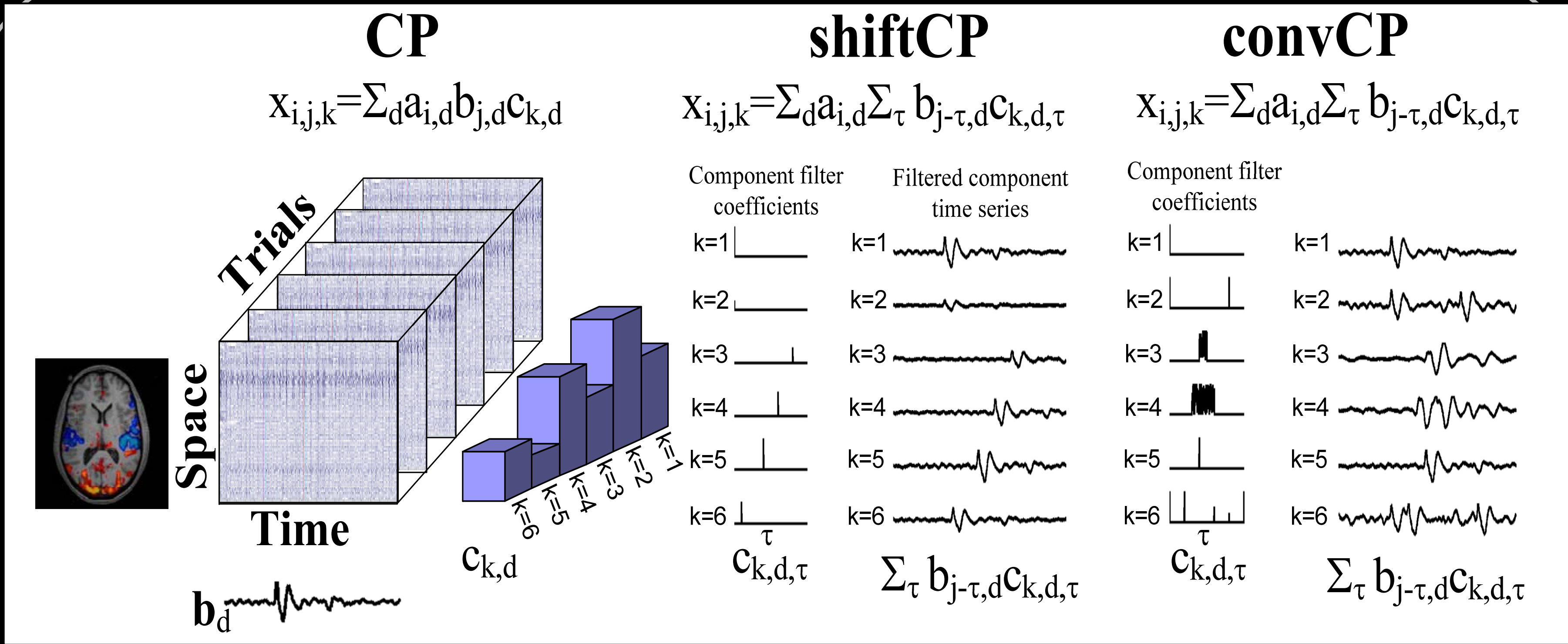
convolutive CP:

$$x_{i,k}(t) \approx \sum_{d,\tau} a_{i,d} b_d(t-\tau) c_{k,d}(\tau)$$





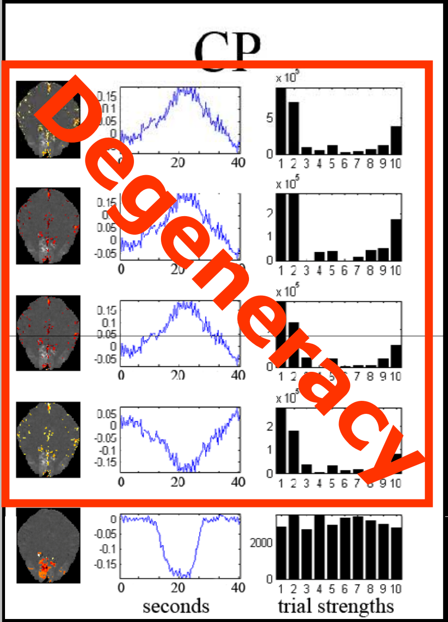
Mørup, Hougard, Hansen, Nips workshop on New Directions in Statistical Learning for Meaningful and Reproducible fMRI Analysis 2008



ConvCP: **Can model arbitrary number of component delays within trial and account for shape variation within convolutional form.**
To resolve ambiguity sparsity imposed on filter coefficients.
Number of components and degree of sparsity learned through Bayesian inference using a framework called automatic relevance determination (ARD)

Mørup et al, Journal of Chemometrics 2010
Mørup et al, Wiley Interscience 2011
Mørup et al., 2008 NIPS 2008
workshops on New Directions in Statistical Learning for Meaningful and Reproducible fMRI Analysis

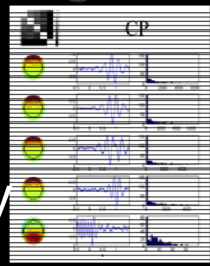
Analysis of fMRI data



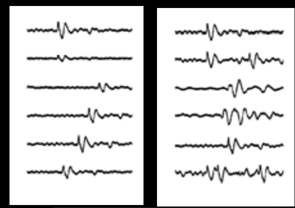
Each trial consists of a visual stimulus delivered as an annular full-field checkerboard reversing at 8 Hz.

Mørup et al., 2008 NIPS 2008 workshops on
New Directions in Statistical Learning for Meaningful and Reproducible fMRI Analysis

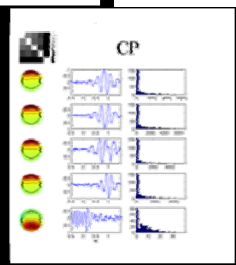
Conclusion and outlook



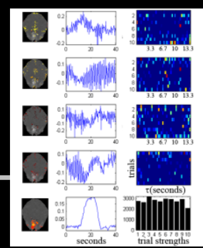
- Multi-linear modeling offers the ability to extract the consistent activity of neuroimaging data over repeats/subjects/conditions etc.
- However, violation of multi-linearity due to variability causes degeneracy



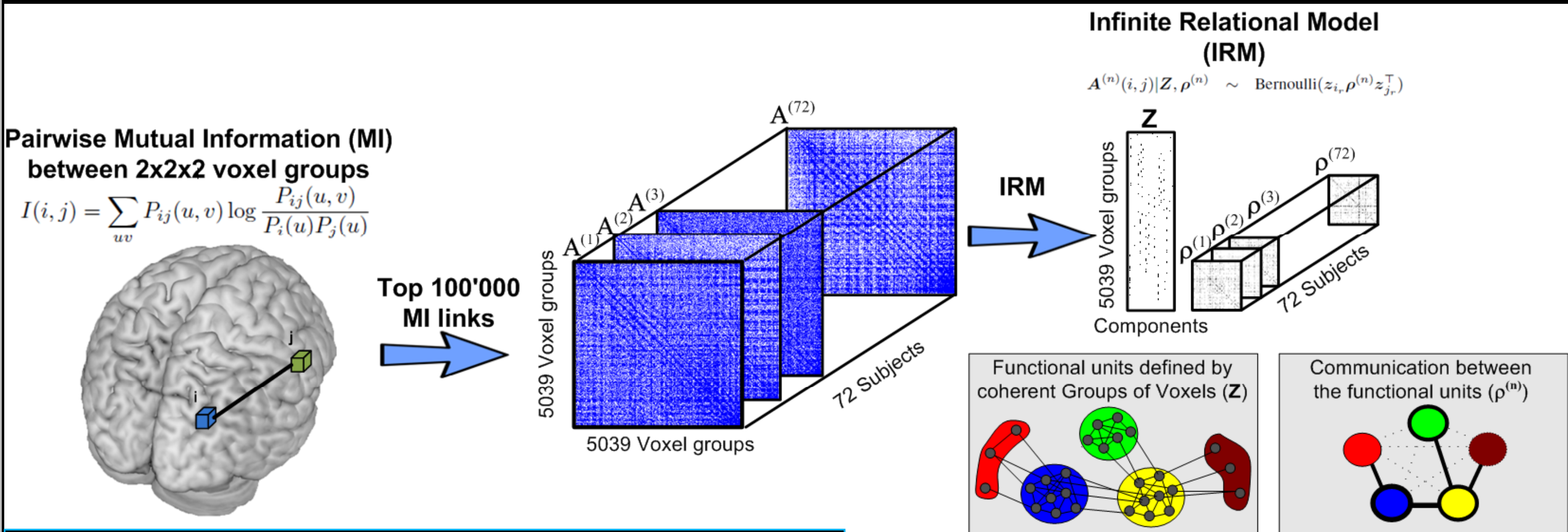
- Common causes of variability are delay and shape variation
- **Advancing the CP model to ShiftCP and ConvCP enables to address these types of variability for improved identification of the consistent patterns of activation.**



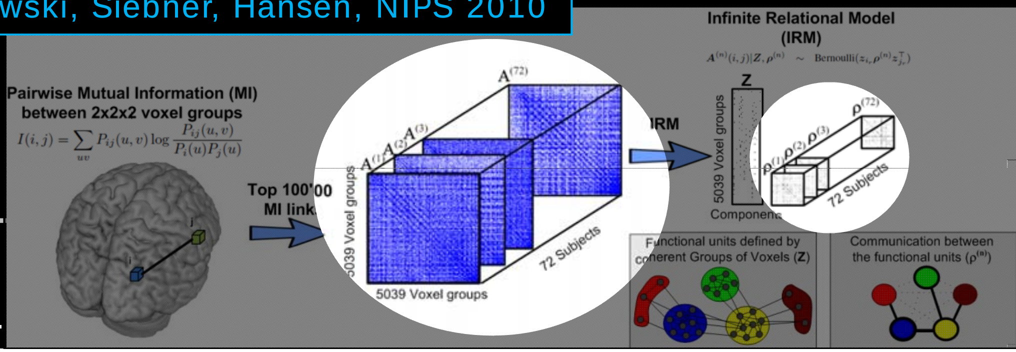
- **Shape and latency changes can carry important neurophysiological information, this is of current research.**



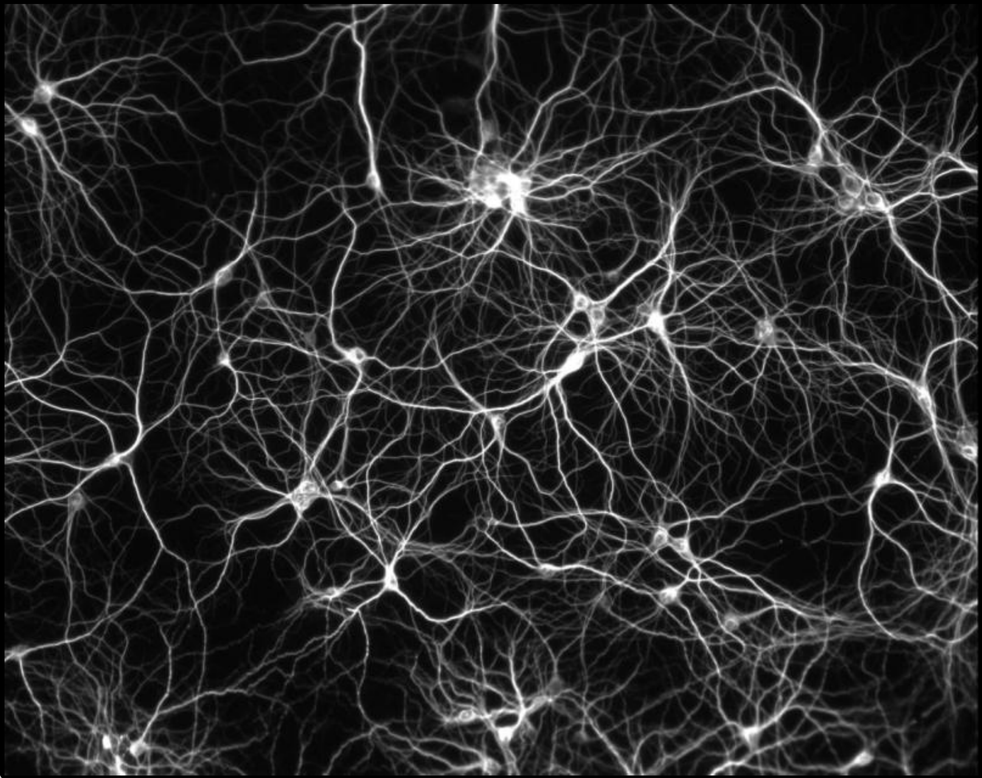
If time allows it I will give some remarks on my current research in tensor modeling of functional and structural connectivity changes in the brain.



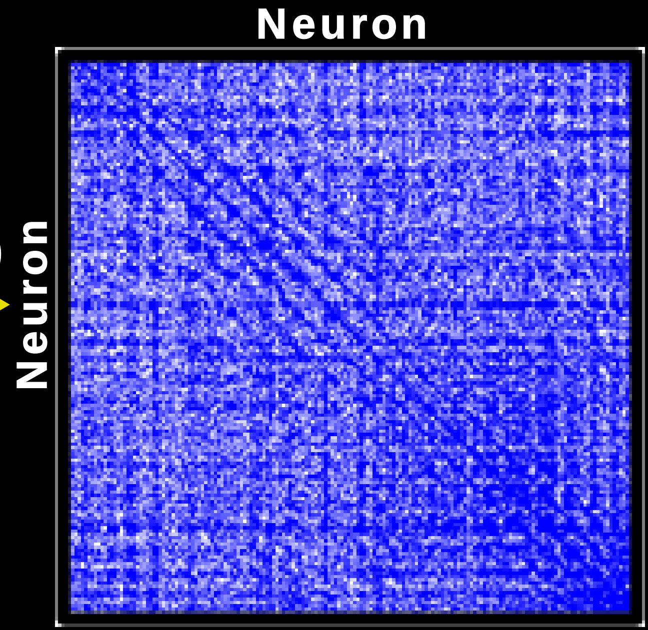
Mørup, Madsen, Dogonowski, Siebner, Hansen, NIPS 2010



The perhaps most formidable and fascinating of all networks is the complex network of neurons constituting our mind, i.e. our connectome!



G(E,V)



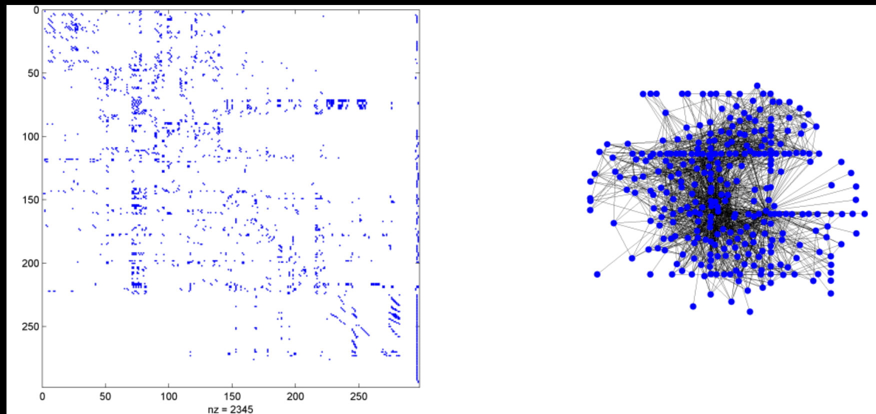
This can be represented as a complex network
$\sim 10^{11}$ neurons,
$\sim 10^{15}$ connections

(Sporns et al. 2005, Murre et. al. 1995, Braitenberg et al. 1991)

Mørup, Madsen, Dogonowski, Siebner, Hansen, NIPS 2010

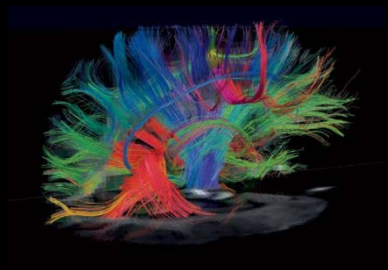
- Micro-scale (units: Single neurons)

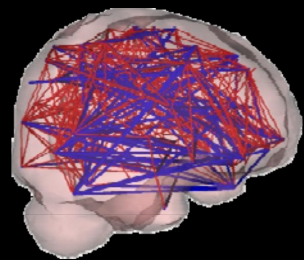
Connectome of the worm C. Elegans
(White et al. 1986)



- Macro scale (units: Voxels [mm^3])

Structural connectivity: Diffusion Weighted Imaging (DWI)
Functional connectivity: functional Magnetic Resonance Imaging (fMRI)

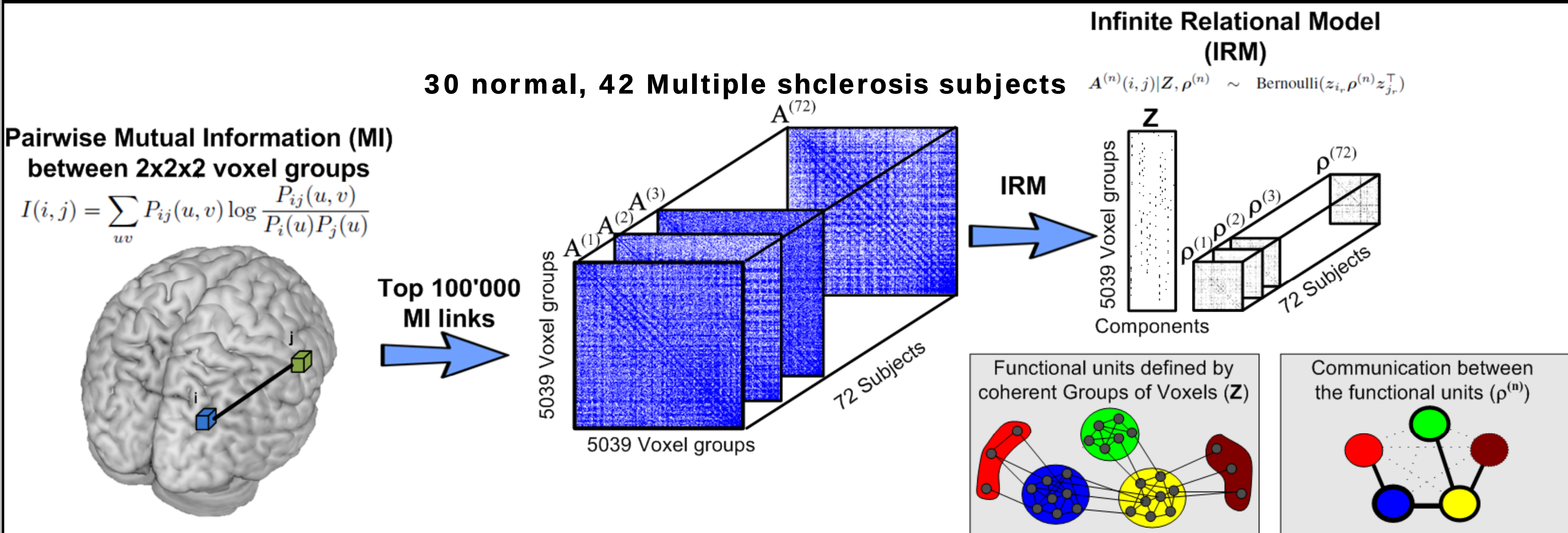




Mørup, Madsen, Dogonowski, Siebner, Hansen, NIPS 2010

Outline of the analysis

Tucker2 model representation

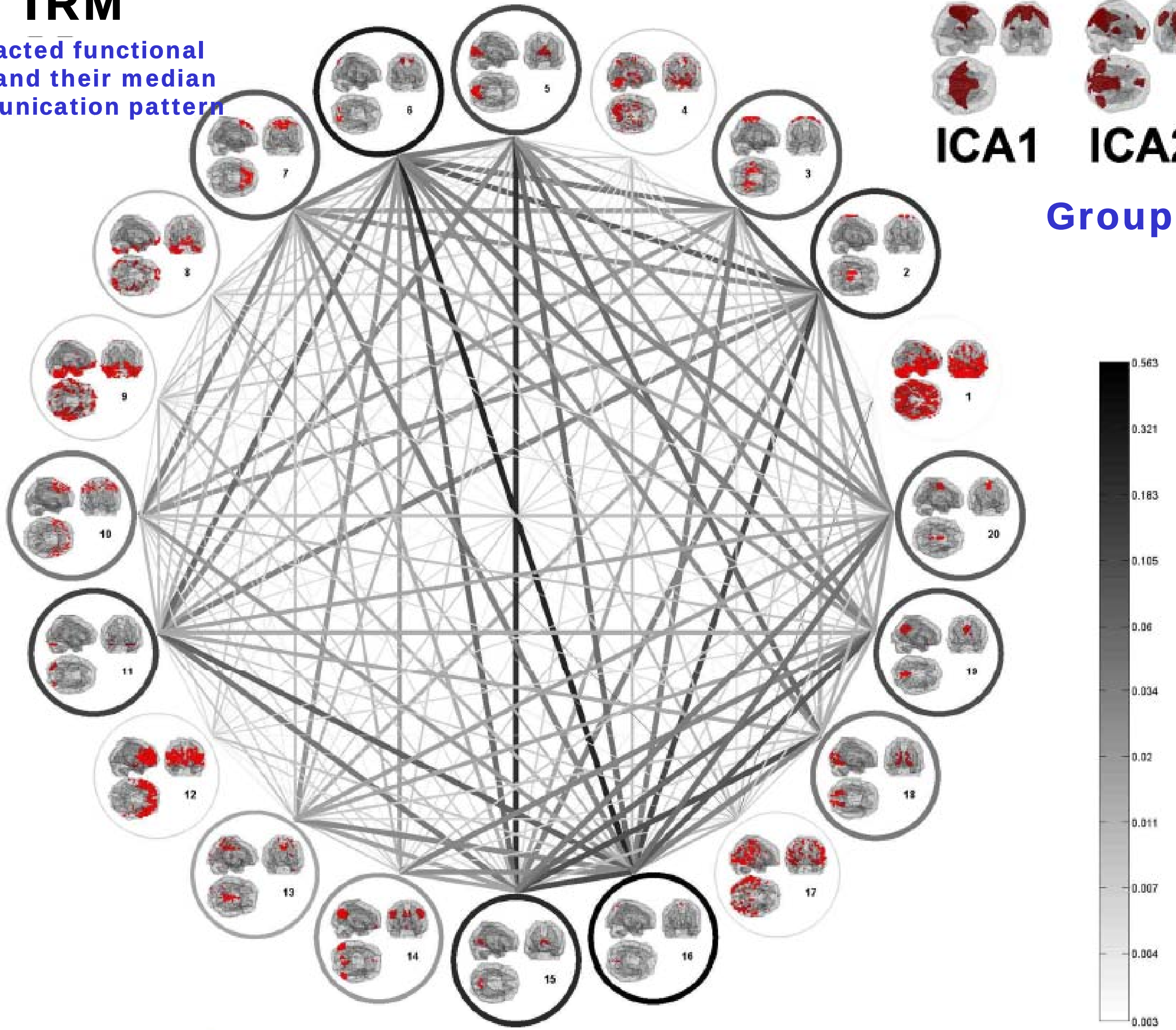


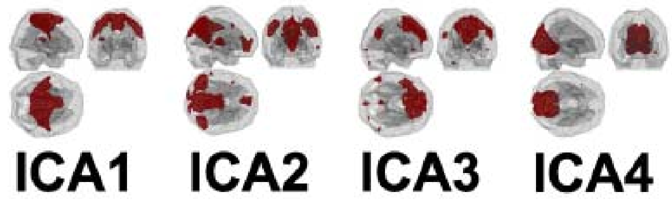
IRM is a Bayesian statistical model of networks that automatically (unsupervised) infers the functional units as well as their subject specific interactions at a relatively low computational complexity.

Mørup, Madsen, Dogonowski, Siebner, Hansen, NIPS 2010

IRM

Extracted functional units and their median communication pattern





Group ICA analysis

Mørup, Madsen, Dogonowski, Siebner, Hansen, NIPS 2010

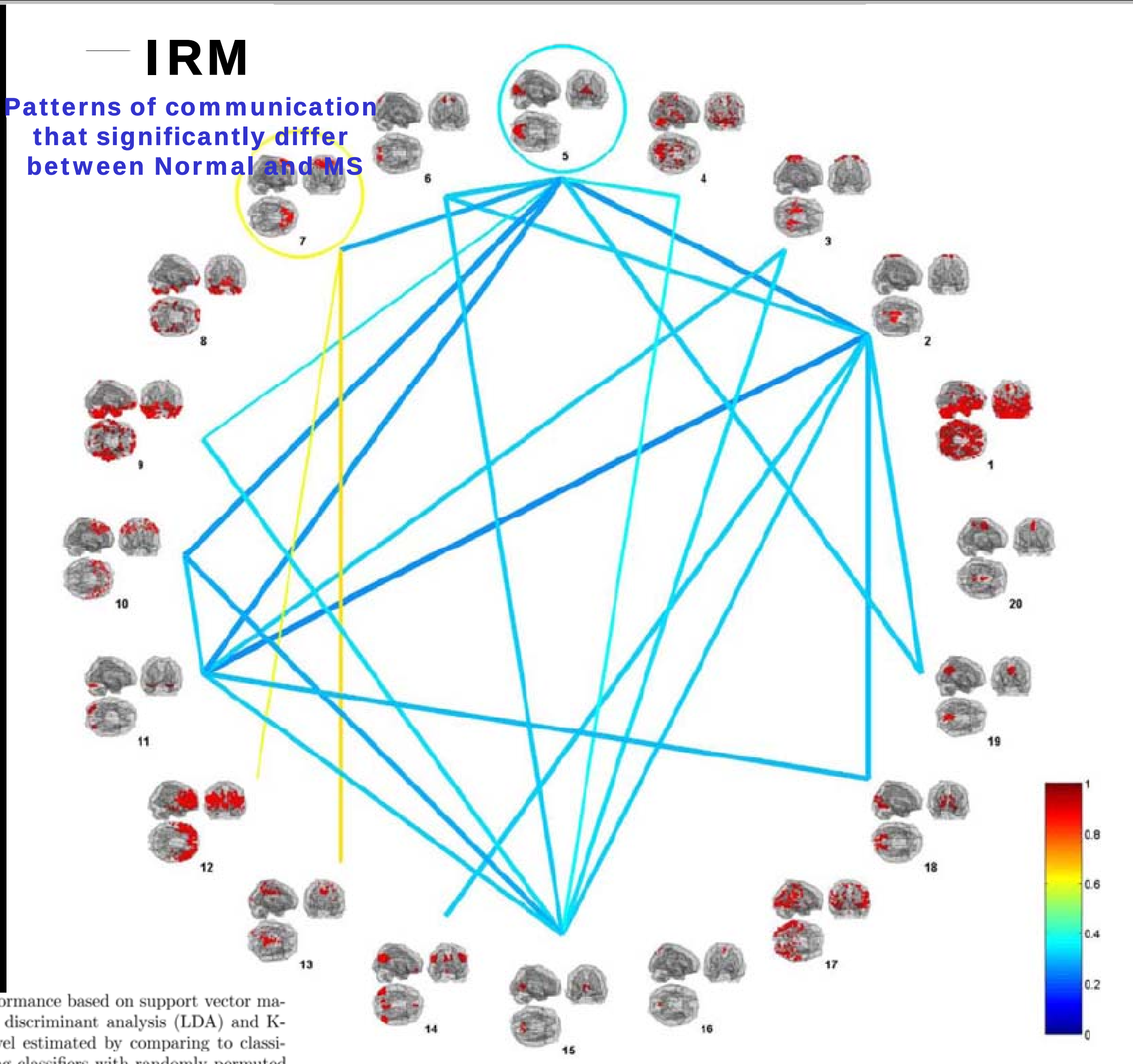


Table 1: Leave one out classification performance based on support vector machine (SVM) with a linear kernel, linear discriminant analysis (LDA) and K-nearest neighbor (KNN). Significance level estimated by comparing to classification performance for the corresponding classifiers with randomly permuted class labels, bold indicates significant classification at a $p \leq 0.05$.

	Raw data	PCA	ICA	Degree	IRM
SVM	51.39	55.56	**63.89** ($p \leq 0.04$)	59.72	**72.22**($p \leq 0.002$)
LDA	59.72	51.39	**63.89** ($p \leq 0.05$)	51.39	**75.00**($p \leq 0.001$)
KNN	38.89	58.33	56.94	51.39	**66.67**($p \leq 0.01$)

Mørup, Madsen, Dogonowski, Siebner, Hansen, NIPS 2010