Brain Signals & Sparsity

Working Group Review

Update on current state-of-the-art of EEG & fMRI processing with sparsity-aware methods

NKUA/UoA, Athens, February 2013

Harris Georgiou (xgeorgio@di.uoa.gr)

Outline:

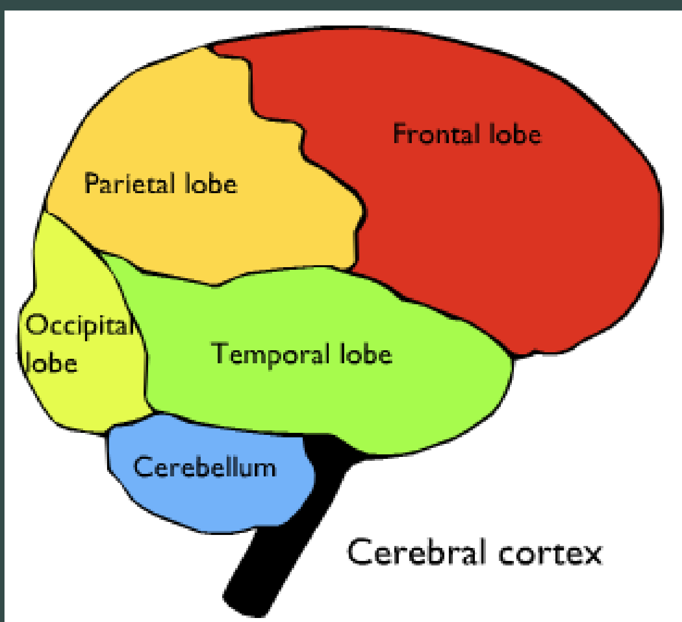
- Part I: Problem overview & modalities
- Part II: Statistical processing of brain signals
- Part III: Brain signals & sparsity
- Part IV: Key directions for future work

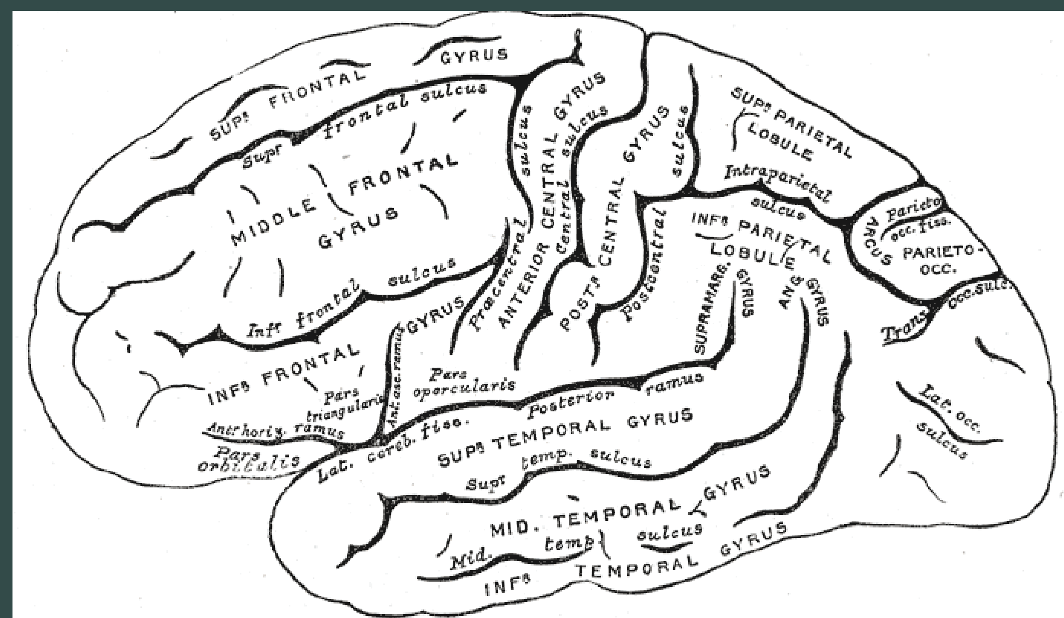
Part I: Problem overview & modalities

- A quick tour of the brain
- Magnetic Resonance Imaging (MRI)
- functional MRI and BOLD
- Electroencephalography (EEG)
- Properties & preprocessing of MRI, fMRI, EEG

A quick tour of the brain

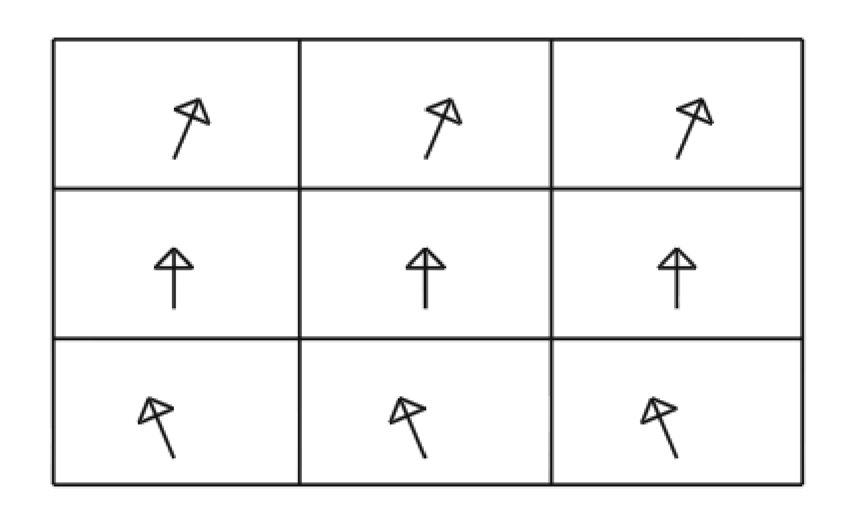
- “flatted” cerebrum has an area of $\approx 2500\ \text{cm}^2$
- **work**: 25% of glucose, 20% of oxygen, 10^{11} neurons x 10^4 synapses
- **core focus**: investigate neurophysiological and cognitive aspects in both normal and pathological cases, understand its structure
- **main problems of interest**: spatial localization of activation areas, temporal correlation of activations, identification of Intrinsic Connectivity Networks (ICNs), <u>sparsity & “pulsed” neural activation</u>





The science of MRI

- the brain is aligned into a 3-D grid of cells (“voxels”)
- each voxel is excited by a strong magnetic field (modulated)
- **k-space**: frequency/phase-matched voxels in brain “slices”
  (distinct freq. per “column” / phase per “row”, or “spiral” trajectories)
- the recorded signal is the DFT of the measured voxel “values”
  (recover the original signal by inverse DFT, inherently complex-valued)
- slices are separated by some gap
  to limit “cross-talk” between them
- some voxels are marked as “null”
  in preprocessing (no tissue)
- result is a 3-D time-varying “hull”



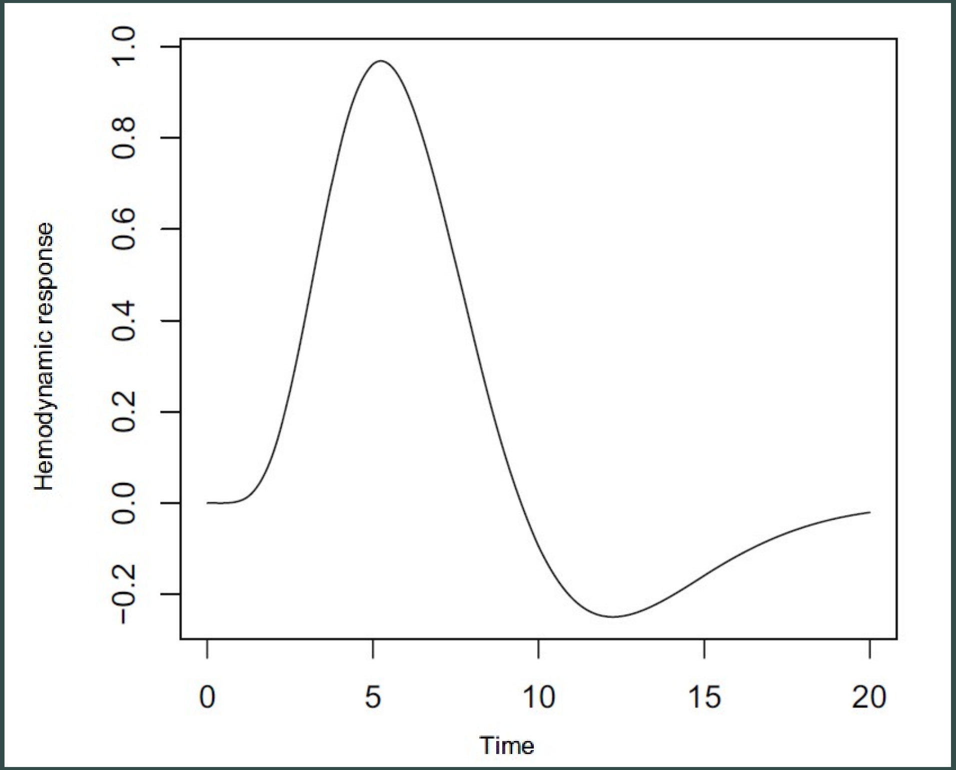
- image size usually 20 cm (diameter)
- 64x64 or 128x128 voxels per slice
  (200 / 64 = 3,125 mm voxel size)
- slice thickness: 3-5 mm (+1 mm gap)

From MRI to functional MRI (fMRI)

- instead of (static) tissue, measure changes in blood flow in voxels
- deoxygenated blood has 20% greater magnetic susceptibility (lower MR)
- BOLD: “Blood Oxygenation Level Dependent” (measure MR differences)
- increased neural activity $\Rightarrow$ increased blood flow (oxygenated)

- HRF: Hemodynamic Response Function (“system” response)
- can be modeled as difference of two gamma distributions
- but not easy to employ as base signal proc. “per voxel”

- about 2 secs delay from sensory input to actual activation
- slowly peaking at 6 secs
- if sensory input persists, activation gradually decreases (“drift”)
- deactivation to baseline is needed
- whole process: 15-20 secs (cycle)



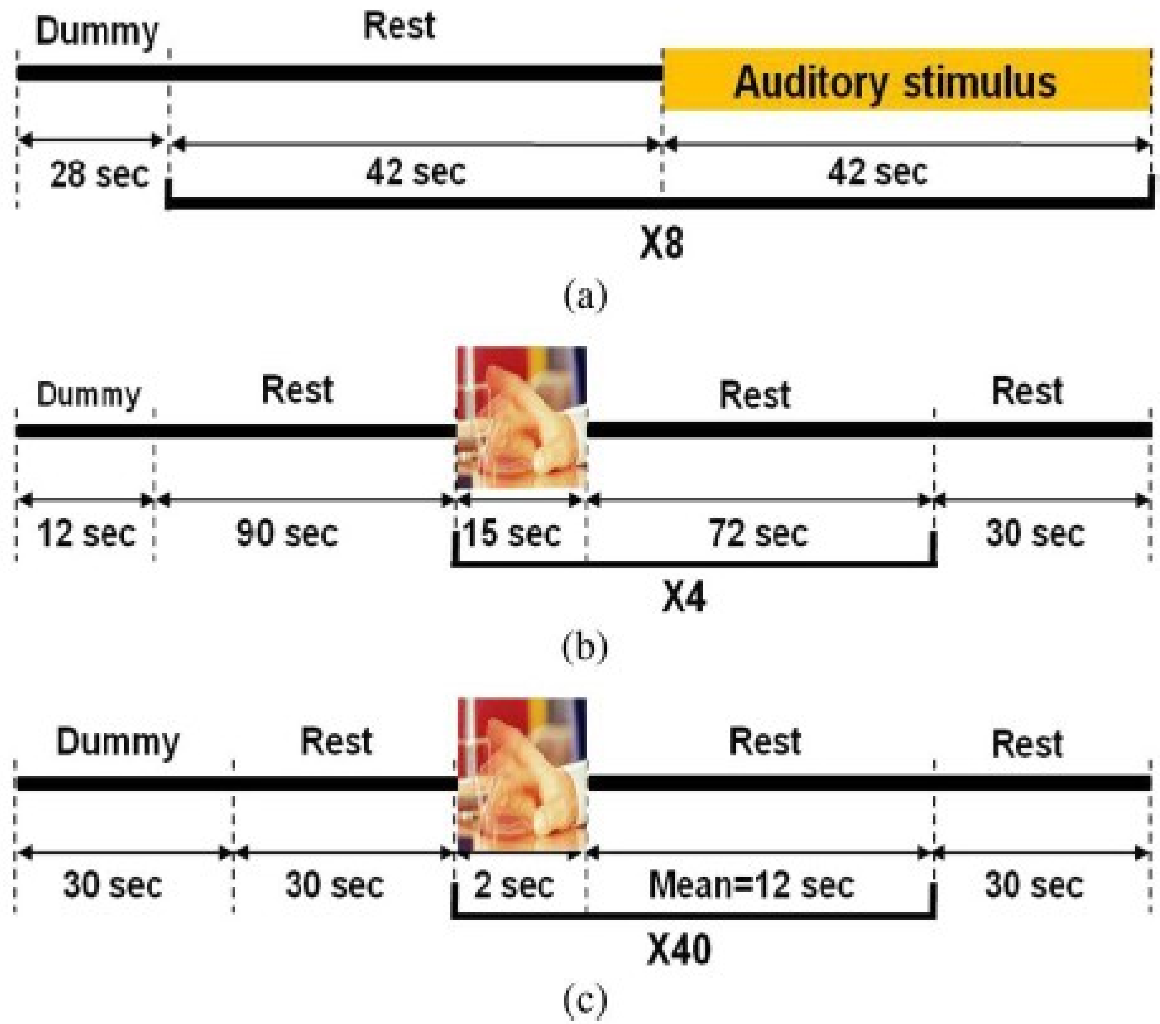
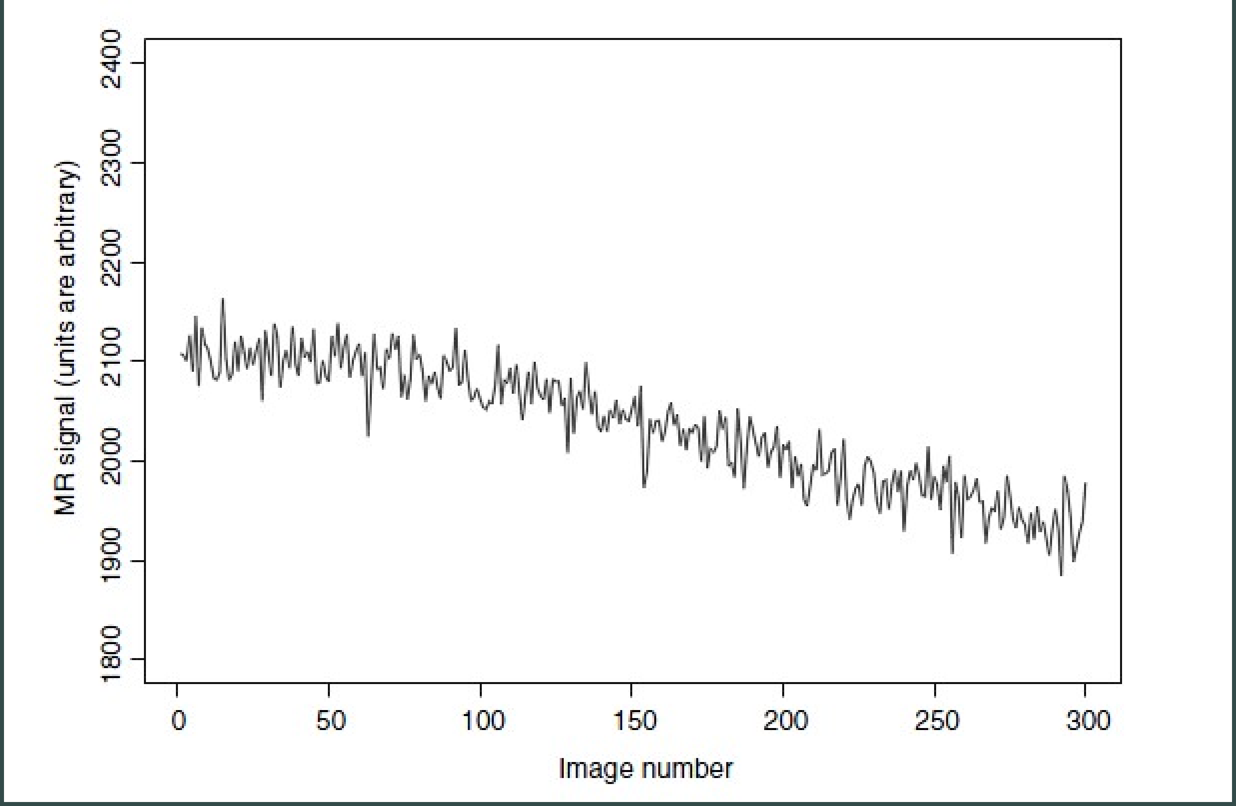


Fig. 3. Experimental paradigms for (a) auditory stimulus tasks, (b) block paradigm right finger tapping tasks, and (c) event-related right finger tapping tasks.

Sources of “noise” in MRI

- machine-related: electronic (internal), sensory, etc.
- subject-related: aliasing (internal), respiration & heartbeat (trends), baseline drift, head motion (MR), volume changes (MR), etc.
- procedure-related: distraction, loss of focus, sensory habituation, etc.



General properties of fMRI signal

- 64x64 = 4096 voxels per slice, 7-10 slices per "snapshot" (low res. mode)
- non-stationary conditions (brain, BOLD), asynchronous properties
- HRF/BOLD/noise different in various brain areas & between subjects
- data are huge in volume, noisy, "multiplexed", highly correlated
- both spatial & temporal resolution is low
- areas of interest ("activated") are usually no more than 3% of total
- spatial/temporal structure of data is not fully understood ("gray box")

Preprocessing in fMRI

✓ slice timing correction (measurement phase)
✓ scanner detrending & equalization (signal stronger at center of MRI)
✓ head motion, brain reshaping, C-R cycles ($\Rightarrow$ affects BOLD accuracy)
✓ spatial & temporal noise is approximately Gamma-distrib.
✓ usually apply a 3-10 mm Gaussian smoothing (usually 3x voxel size)

General properties of EEG signal

- multiples of 1-D time series, 20-100s of electrodes placed on scalp
- “surface” measurements (no 3-D modeling), 10μV-100μV potentials
- modern ADC: 256-512 Hz, most applications require <50 Hz sampling
- data are huge in volume, noisy, “multiplexed”, highly correlated
- spatial resolution is low, temporal resolution is high
- spatial/temporal structure of data is not fully understood (“gray box”),

Components / forms of EEG waves

- “Delta” (up to 4 Hz): when asleep (normal), in lesions (abnormal)
- “Theta” (4-8 Hz): found at locations “activated” for some task
- “Alpha” (8-13 Hz): relaxing/thinking (normal), in coma (abnormal)
- “Beta” (13-30 Hz): focused thought, alert (normal), pharmaceuticals
- “Gamma” (30-100 Hz): cross-sensory processing (normal)
- *“Mu” (8-13 Hz): resting-state in motor neurons*

- **EEG**: diagnostic value has been proven over the years, but only supplementary to the doctor's clinical evaluation (examination)
- doctors can "see" abnormal brain activity in EEG, but the exact information content can not be encoded easily into "features"
- Typical diagnostic applications:
  - ✓ **epileptic seizures ➔ ➔ ➔**
  - ✓ Alzheimer's disease
  - ✓ coma patients
  - ✓ schizophrenia
  - ...

EEG versus fMRI:
(+) much better temporal resolution
(+) less intrusive exam. protocol
(-) "surface" signal, no 3-D modeling
(-) inherently a "black box" approach

Part II: Statistical processing of brain signals

- Statistical processing of fMRI
- Decomposition methods: GLM, PCA, SVD, ICA
- Multi-subject (group) processing of fMRI
- Brain functional models (ICN, FNC, dynamics)

Statistical processing of fMRI

- due to k-space representation, MR signal is inherently complex-valued
- Stat. correlation or General Linear Model (GLM) often used as baseline
- differences between “resting” and “activation” tested statistically
- due to its properties, the signal has to be “demixed” first (decompose)
- instead of GLM, decomposition itself can be used as “model”
- most commonly used: ICA, SVD, sparse (recently)
- in order to acquire useful clinical results, <u>multi-subject</u> methods are used

IEEE REVIEWS IN BIOMEDICAL ENGINEERING, VOL. 5, 2012

Multisubject Independent Component Analysis of fMRI: A Decade of Intrinsic Networks, Default Mode, and Neurodiagnostic Discovery

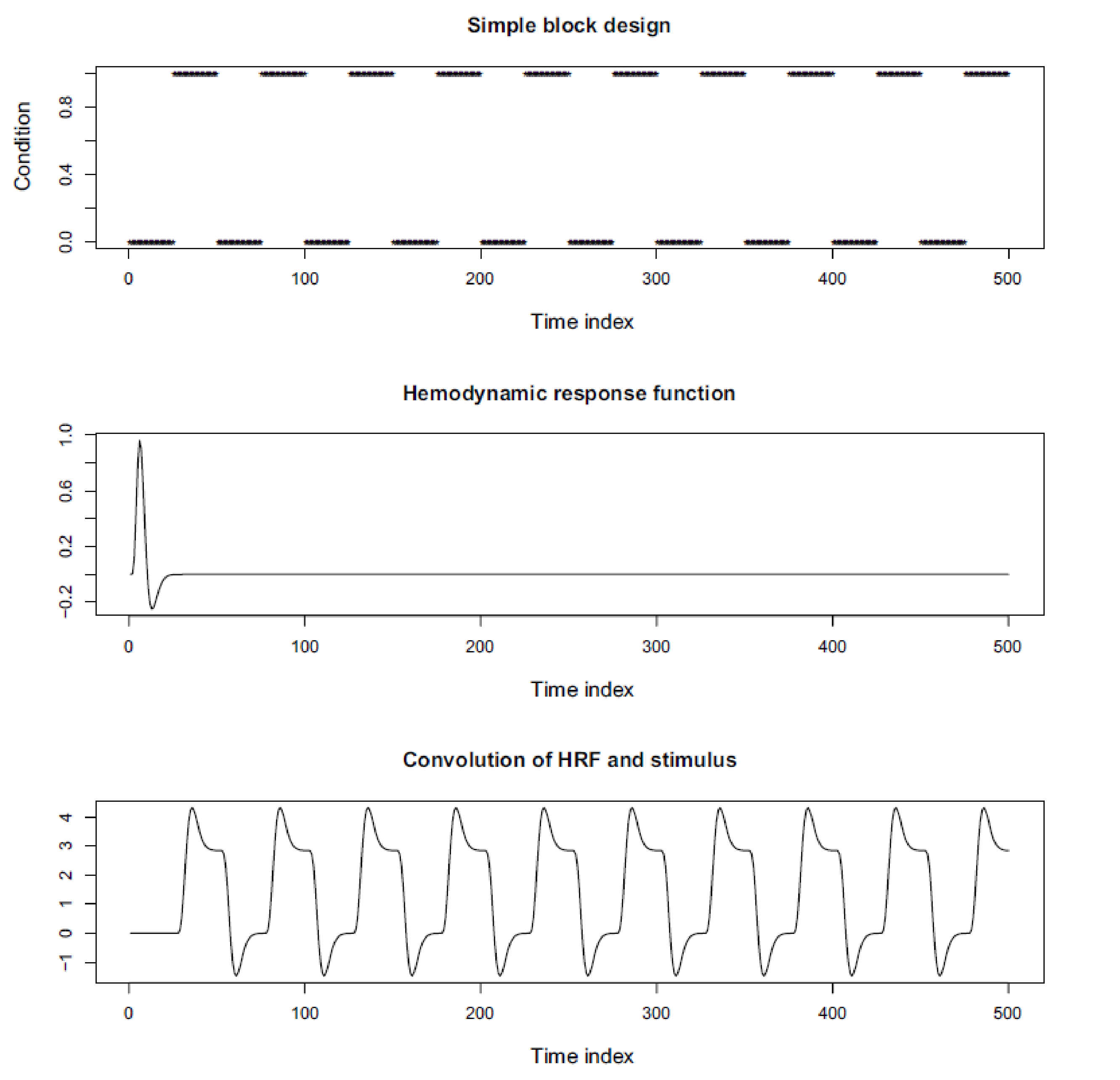
Vince D. Calhoun, *Senior Member, IEEE*, and Tülay Adalı, *Fellow, IEEE*

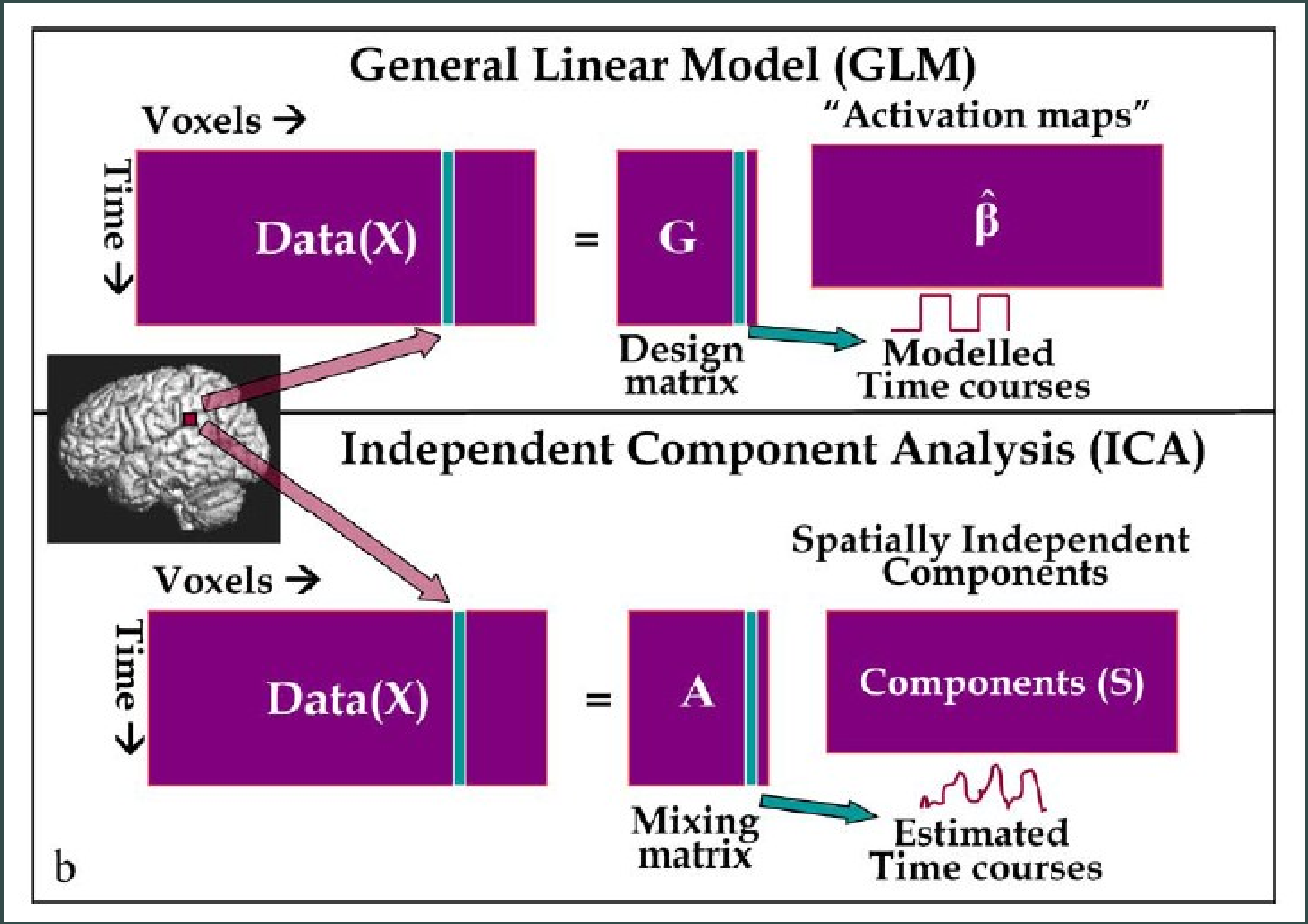
Basic GLM approach on fMRI

- HRF time courses are assumed known, modeled as “design matrix”: binary components (pulses), convolutions with HRF, etc.
- measured values (voxels) are the combined outputs
- a “mixing” (linear) model is calculated: $X = \mathbf{G}*B + \varepsilon$
- very restrictive, assumes independent voxels, times courses, same error variances and same model for all voxels, etc.
- model parameters are compared (stat.signif.) over “resting” and “activating” time courses to locate actual brain activations (e.g. via thresholding)

Basic ICA approach on fMRI

- Similar to GLM, only now “sources” are assumed unknown
- instead, statistical constraints are introduced on them (independency)
- a “mixing” (linear) model is calculated: $X = \mathbf{A}*S$
- measurements are “decomposed” into stat. independent “sources”
- used as a generalization of PCA or SVD decomposition (into mutually orthogonal spatio-temporal components)
- with ICA, the components are *statistically independent*


Simple block design
Condition
0.0
0.4
0.8
0
100
200
300
400
500
Time index
Hemodynamic response function
−0.2
0.2
0.6
1.0
0
100
200
300
400
500
Time index
Convolution of HRF and stimulus
−1
0
1
2
3
4
0
100
200
300
400
500
Time index


General Linear Model (GLM)
Voxels →
Time ↓
Data(X)
=
G
"Activation maps"
$\hat{\beta}$
Design matrix
Modelled Time courses
Independent Component Analysis (ICA)
Voxels →
Time ↓
Data(X)
=
A
Spatially Independent Components
Components (S)
Mixing matrix
Estimated Time courses
b

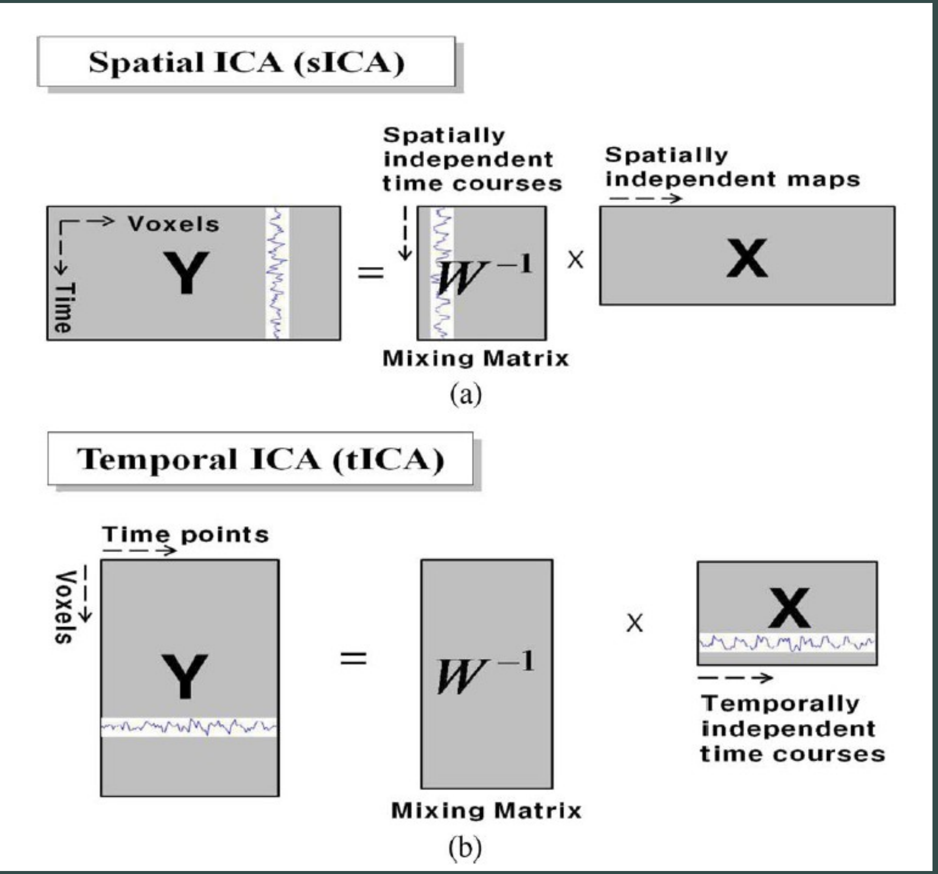
Multi-subject (group) processing of fMRI

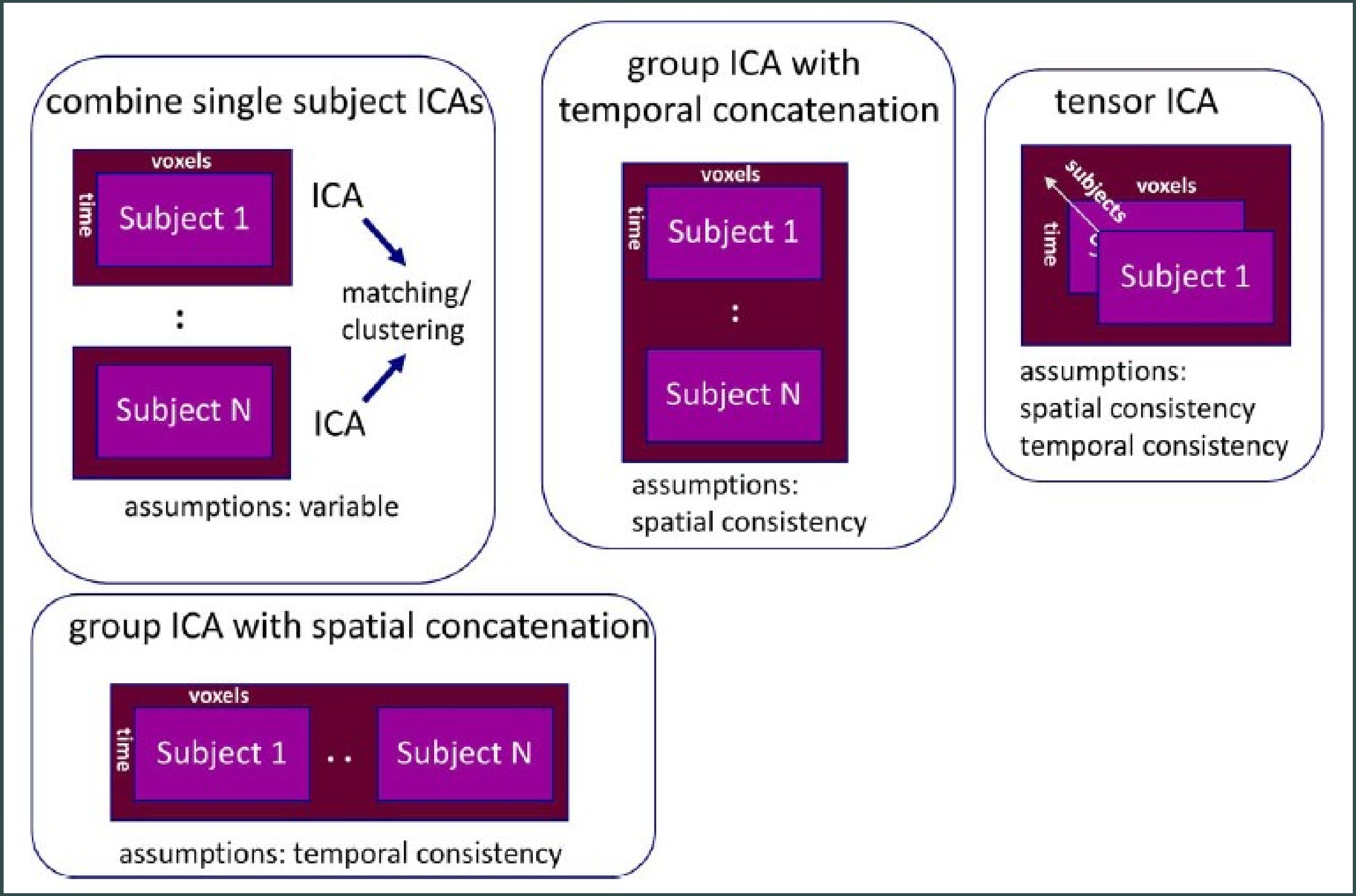
- general functional models & brain networks need multi-subject processing modes
- even on a single subject, fMRI data must be concatenated (spatially or temporally)
- **ICA**: spatial (sICA) or temporal (tICA) concat., **EEG**-based: usually temporal concat.

➔ Similarly, fMRI data from multiple subjects are to be grouped together, it can be:
  - **spatially**
  - **temporally**
  - **combined (tensors)**

➔ For multi-subject, tICA seems to work better than sICA

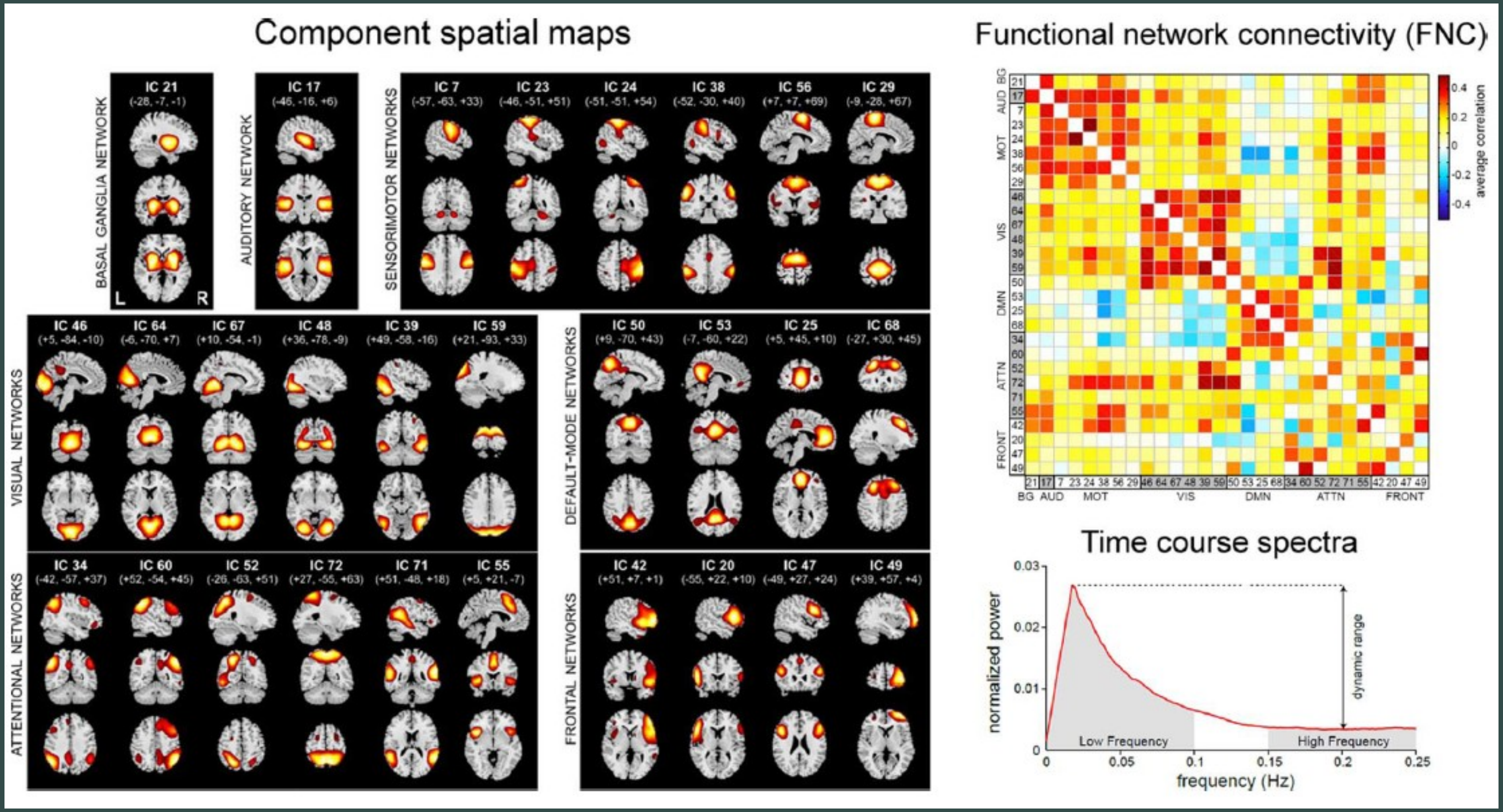
➔ Alternatively, the ICA (or other) model parameters can be investigated on a 2nd stage of processing (e.g. clustering)

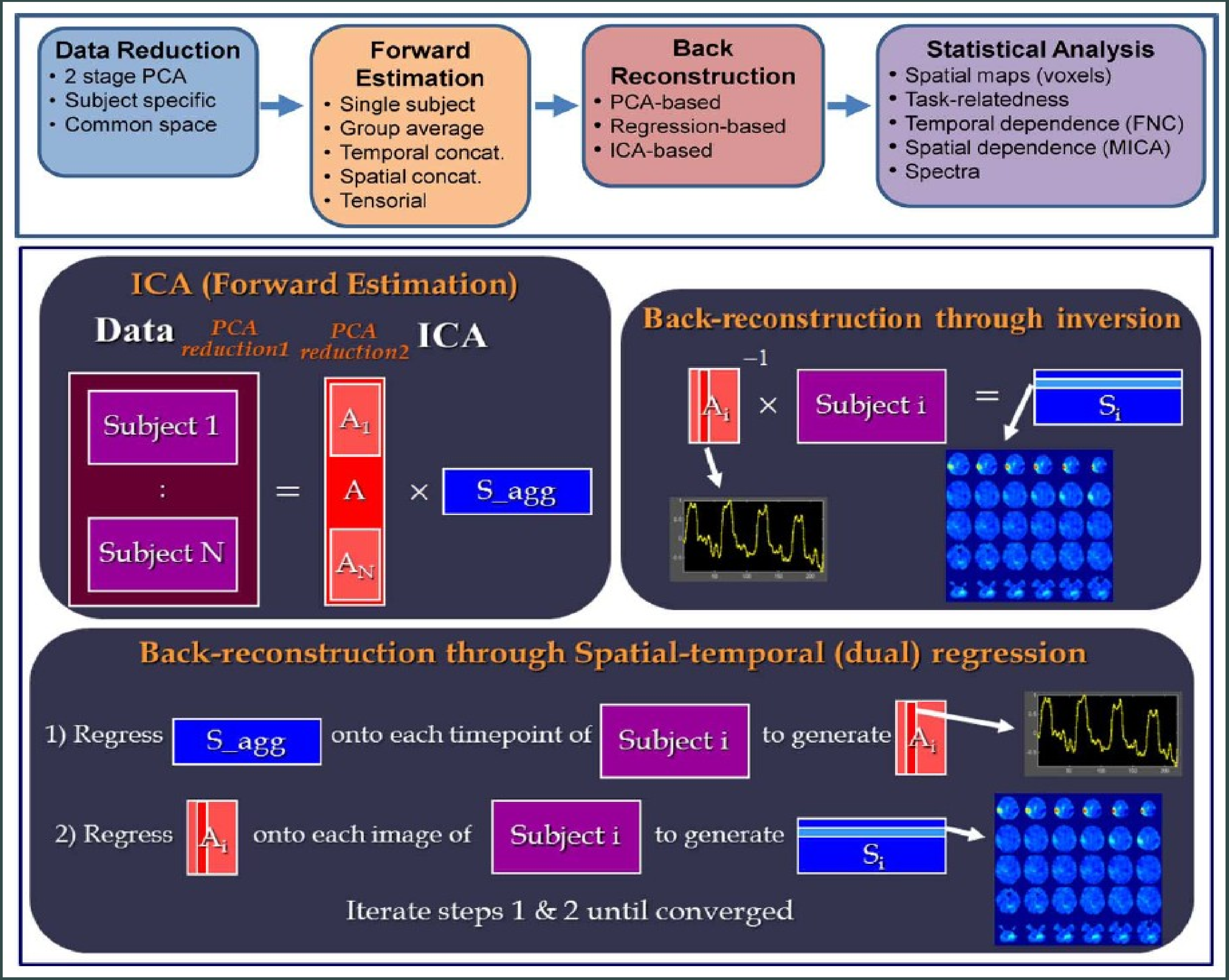


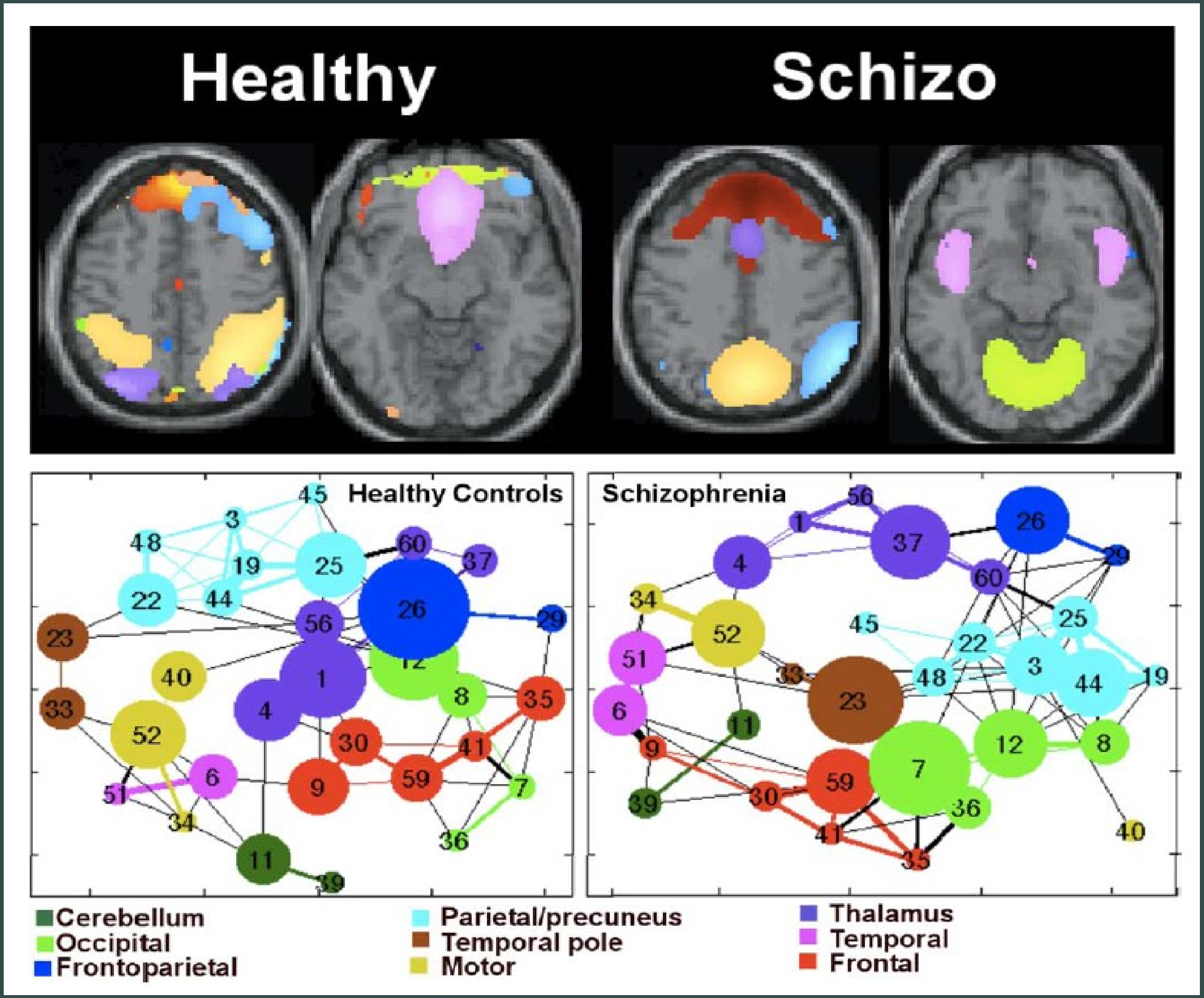

combine single subject ICAs
voxels
time
Subject 1
:
Subject N
ICA
matching/
clustering
ICA
assumptions: variable
group ICA with
temporal concatenation
voxels
time
Subject 1
:
Subject N
assumptions:
spatial consistency
tensor ICA
subjects
voxels
time
Subject 1
assumptions:
spatial consistency
temporal consistency
group ICA with spatial concatenation
voxels
time
Subject 1
. .
Subject N
assumptions: temporal consistency

Typical steps in group ICA processing (fMRI):

1. Data reduction: usually via PCA
2. Forward estimation: grouping of fMRI data (see prev.)
3. Subject back-projection: “translate” components into brain areas
4. Extract functional models: Intrinsic Functional Networks (ICN)




Data Reduction
• 2 stage PCA
• Subject specific
• Common space
Forward Estimation
• Single subject
• Group average
• Temporal concat.
• Spatial concat.
• Tensorial
Back Reconstruction
• PCA-based
• Regression-based
• ICA-based
Statistical Analysis
• Spatial maps (voxels)
• Task-relatedness
• Temporal dependence (FNC)
• Spatial dependence (MICA)
• Spectra
ICA (Forward Estimation)
Data
PCA reduction1
PCA reduction2
ICA
Subject 1
:
Subject N
=
A1
A
AN
×
S_agg
Back-reconstruction through inversion
−1
Ai
×
Subject i
=
Si
Back-reconstruction through Spatial-temporal (dual) regression
1) Regress
S_agg
onto each timepoint of
Subject i
to generate
Ai
2) Regress
Ai
onto each image of
Subject i
to generate
Si
Iterate steps 1 & 2 until converged


Healthy
Schizo
Healthy Controls
Schizophrenia
Cerebellum
Occipital
Frontoparietal
Parietal/precuneus
Temporal pole
Motor
Thalamus
Temporal
Frontal

Recent work: Dynamics of ICNs:

- main idea: capture and study temporal changes in functional connectivity
- brain connectivity: “anatomical”, “*functional”*, “*effective”*
- FNC: *functional* network connectivity (pairwise correlations between ICNs)
- “features”: variability of FNC, graph-based approaches (metrics), etc.
- as a diagnostic tool, differentiate between “normal” and “abnormal” cases

Issues in fMRI data processing:

- *“efficient preprocessing of the data more important than actual prediction method”* (⇒ room for powerful ML algorithms as new test base)
- *“better to start with lower resolution, than with higher + smoothing later”*

Part III: Brain signals & sparsity

- Sparsity in brain signals
- Sparsity and CS in EEG
- Sparsity and CS in fMRI

Sparsity in brain signals

- neurons exhibit “pulsed” activation patterns (thresholded)
- temporal sparsity: activation only when stimulated (sensory data)
- spatial sparsity: localized activation patterns (ICNs)
- usually only 3% of voxels are tagged as “active” in fMRI tests
- similar sparsity observations in EEG data, but <u>not</u> in time domain

Why use sparse algorithms in EEG, fMRI ?

- Sparse Processing (SP): closely linked to Compressed Sensing (CS)
- huge volumes of data can be reduced via sparse repr. (mostly in fMRI)
- CS has been employed for efficient energy management (EEG)
- tests show that SP is <u>more robust</u> than ICA in decomposition (fMRI)

Sparsity and CS in EEG:

- *"EEG is non-sparse in time domain"*
- *"standard compression (e.g. wavelet) not very energy-efficient" (h/w)*
- **main idea**: use a *dictionary* for the *sensing* signal ($Y=\Phi*(\mathbf{D}*z)$) to "make it" sparse
- BSBL: Block Sparse Bayesian Learning (for block-structure signals)
- the block-structure is (assumed) not very strict in practice
- experiments with DCT and DWT dictionary components (D)
- But: *"if energy consumption is not a problem, full DWT compression is better"*

IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING, VOL. 60, NO. 1, JANUARY 2013

Compressed Sensing of EEG for Wireless Telemonitoring With Low Energy Consumption and Inexpensive Hardware

Zhilin Zhang*, Tzyy-Ping Jung, Scott Makeig, and Bhaskar D. Rao

➔ BSBL-BO: see [5] : Z. Zhang, B. D. Rao, *"Extension of SBL algorithms for the recovery of block sparse signals with intra-block correlation"*, IEEE Trans. on Sig. Proc. (submitted)

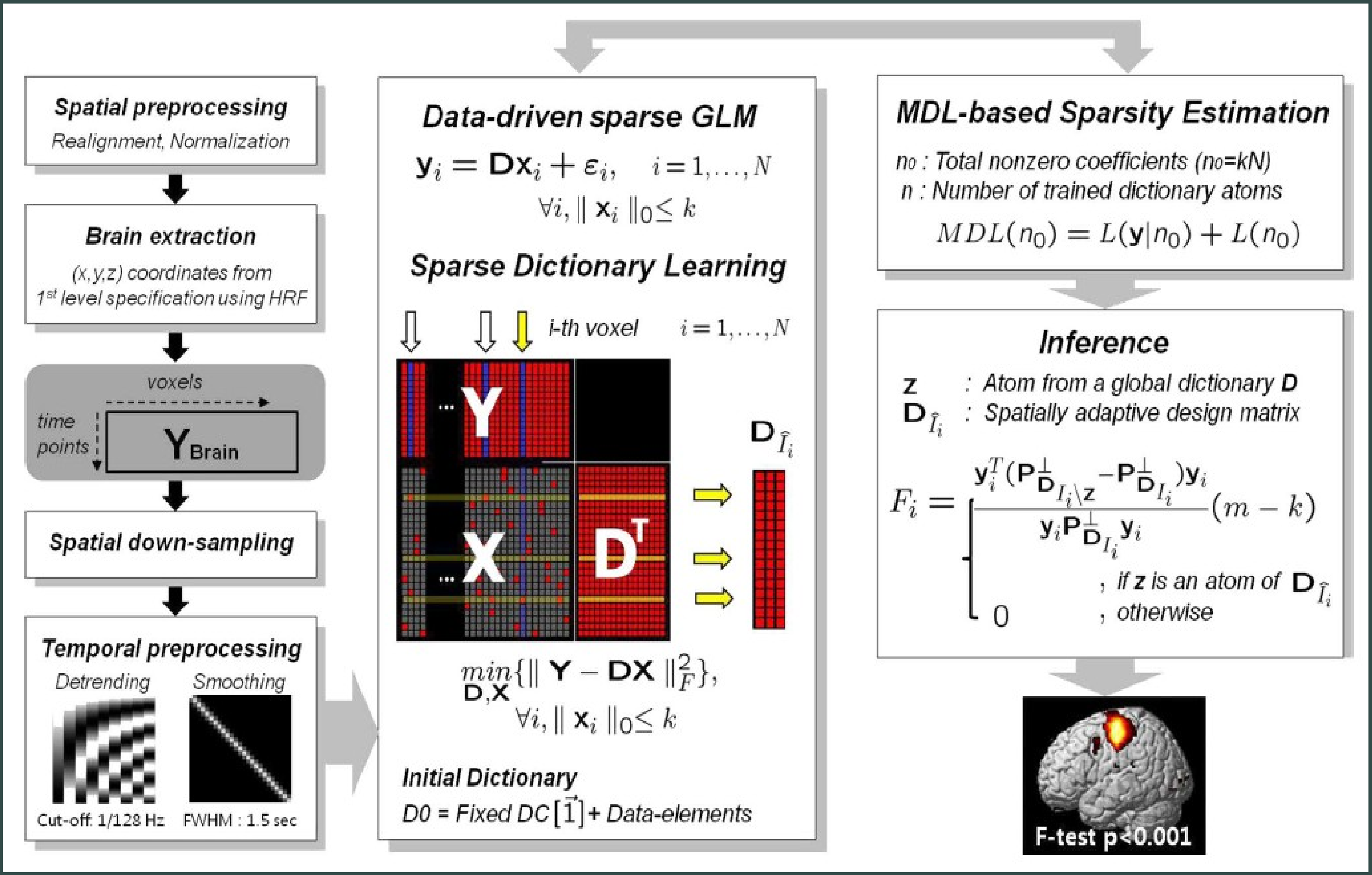
Sparsity and CS in fMRI:

- *"sparsity in (fMRI) signal has been shown to be more promising"*
- *"non-adaptivity of the canonical HRF is a major problem" (in typical GLM)*
- **main idea**: use sparse GLM with "unknown" (data-driven) design matrix
- Minimum Description Length (MDL) criterion for sparsity level estimation
- K-SVD as the main sparse decomposition method is employed

1076 IEEE TRANSACTIONS ON MEDICAL IMAGING, VOL. 30, NO. 5, MAY 2011

A Data-Driven Sparse GLM for fMRI Analysis Using Sparse Dictionary Learning With MDL Criterion

Kangjoo Lee, Sungho Tak, and Jong Chul Ye*, *Member, IEEE*



→ Note: the "Inference" part above (F-maps) refers to the activation detection test (brain areas), see [55]: B. Ardekani, et.al. *"Activation detection in functional MRI using subspace modeling and maximum likelihood estimation"*, IEEE Trans. Med. Imag., vol.18, n.2, 1999.

ICA versus sparse models (fMRI):

- *“independence is non-adaptive for blind source separation in fMRI signals”*
- *“the most influential factor for the success of ICA is sparsity (over independence)”*
- *“preprocessing and hemodynamics make the components inherently correlated”*
- *“performance of the ICA is very sensitive to noise”*
- *“sparsity over indep. is supported by biological findings (sparse coding in neurons)”*
- *“proposed method (GLM-based) is more robust, better localization than standard ICA”*

Independent component analysis for brain fMRI does not select for independence

I. Daubechies[a,b,1], E. Roussos[b], S. Takerkart[a,c], M. Benharrosh[a], C. Golden[b,d], K. D'Ardenne[a,e], W. Richter[a,e], J. D. Cohen[a,f], and J. Haxby[a,f]

ICA versus sparse models (fMRI):

- *"Infomax, FastICA in reality deal with* *<u>sparse</u>* *rather than independent components"*
- *"independence is not the right math. framework, sparsity is more natural to brain sig."*
- *"sparsity promotes independence, hence algorithms should target that directly"*
- *"Infomax gives slightly (but consistently) better results than FastICA"*
- *"fMRI experiments should be designed as sparsity-promoting (spatio-temporal)"*

Part IV: Key directions for future work

- open problems in fMRI
- open problems in EEG
- related problems in SP & CS
- some future issues

Open problems in EEG:

- detection and separation of event-related potentials
- brain connectivity models (EEG only or multi-modal)
- brain connectivity as diagnostic tool (Alzheimer's disease, chronic fatigue, ...)
- brain source localization (very difficult with EEG-only)
- brain-computer interfacing (nervous injuries, "trained" HCI, ...)
- EEG fusion with fMRI, mainly data-driven (based on *prediction* or *constraints*)

Advances in Electroencephalography Signal Processing

Saeid Sanei, Saideh Ferdowsi, Kianoush Nazarpour, and Andrzej Cichocki

IEEE SIGNAL PROCESSING MAGAZINE [170] JANUARY 2013 1053-5888/13/$31.00©2013IEEE

➔ see [5]: B. Cheung, B. Riedner, et. al., *"Estimation of cortical connectivity from EEG using state-space model"*, IEEE Trans. Biomed. Eng., vol.57, n.9, 2010.

Open statistical challenges in fMRI:

- possible application of fractal analysis in 1-D, 2-D, 3-D data
- challenge #1: the classification/prediction task (medical problems, HCI, ...)
- challenge #2: multi-modal techniques (e.g. fMRI/EEG fusion)
- improvements in fMRI data acquisition, resolution, preprocessing
- development of non-linear HRF/BOLD models
- multi-subject (group) studies $\Rightarrow$ fusion of data, models, results

Statistical Science
2008, Vol. 23, No. 4, 439–464
DOI: 10.1214/09-STS282
© Institute of Mathematical Statistics, 2008

The Statistical Analysis of fMRI Data

Martin A. Lindquist

Sparsity in fMRI acquisition:

Magnetic Resonance in Medicine 58:1182–1195 (2007)

Sparse MRI: The Application of Compressed Sensing for Rapid MR Imaging

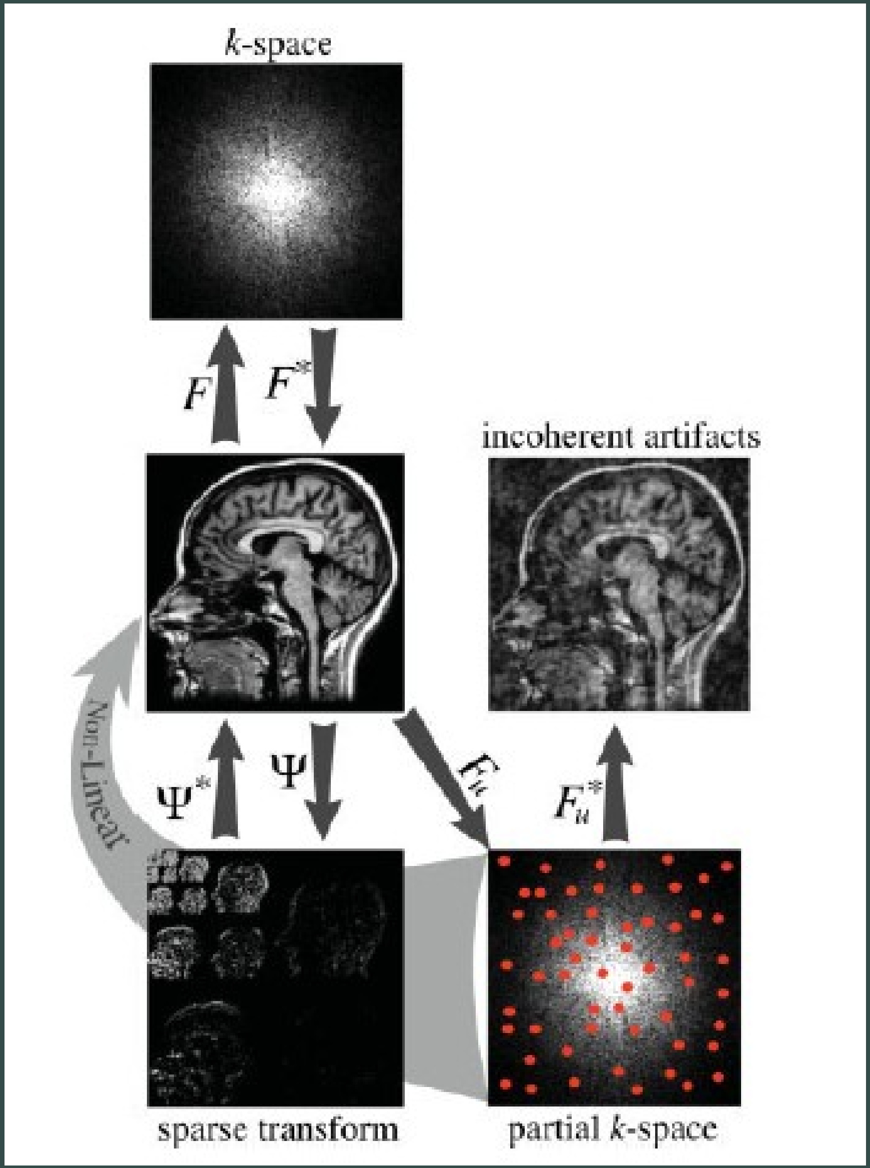
Michael Lustig,[1*] David Donoho,[2] and John M. Pauly[1]

Sparsity in fMRI acquisition:

- Basic idea: undersample the k-space assumed to be sparse in fMRI, then apply effective thresholding & img. quality constraints (preserve only major components)
- *"if the signal/image is to be compressed, is it possible to acquire only a subset of the original sensory data?"* (targeted subsampling)
- undersampling artifacts as "power leaks" between components, can be modeled and removed effectively if they are incoherent (random, not context-relevant)
- <u>random</u> k-space subsampling must be employed, "incoherent sampling trajectories" in k-space are impractical (h/w) but can be mimicked artificially
- example: use full freq. (cols) but subsampled phases (rows) in k-space slices
- extension: subsampling can also employ <u>variable density</u> in k-space

➔see: Donoho, et.al., "*Sparse solution of underdetermined linear equations by stagewise Orthogonal Matching Pursuit" (Donoho, technical report, 2006)*

Sparsity in fMRI acquisition:



Open problems in Sparsity & Compressed Sensing:

➢ *“Theoretical Frontiers, Model Improvements, Applications”*

- better bounds
- targeting general dictionaries
- theory of dictionary learning
- unified theory of simplicity measures

- introducing structure (signal)
- structured dictionaries (components)
- analysis co-sparse models
- model errors (quality/efficiency)

- sparsity in computer graphics
- processing non-conventional signals
  (e.g. connectivity graphs, social networks)

IEEE SIGNAL PROCESSING LETTERS, VOL. 19, NO. 12, DECEMBER 2012

Sparse and Redundant Representation Modeling—What Next?

Michael Elad, *Fellow, IEEE*

→ see [53]: S. Nam, M. Davis, I. Elad, R, Gribonval, *“The cosparse analysis model and algorithms”*, Appl. Comput. Harmon. Anal., *(to be published)*

Other works in Sparsity & CS:

IEEE SIGNAL PROCESSING LETTERS, VOL. 19, NO. 12, DECEMBER 2012

Measure What Should be Measured: Progress and Challenges in Compressive Sensing

IEEE TRANSACTIONS ON SIGNAL PROCESSING, VOL. 59, NO. 9, SEPTEMBER 2011

Structured Compressed Sensing: From Theory to Applications

Marco F. Duarte, *Member, IEEE*, and Yonina C. Eldar, *Senior Member, IEEE*

IEEE TRANSACTIONS ON SIGNAL PROCESSING, VOL. 61, NO. 5, MARCH 1, 2013

Learning Sparsifying Transforms

Saiprasad Ravishankar, *Student Member, IEEE*, and Yoram Bresler, *Fellow, IEEE*

Proceedings of the World Congress on Engineering 2010 Vol I
WCE 2010, June 30 - July 2, 2010, London, U.K.

Sparse Classifier Design Based on the Shapley Value

Prashanth Ravipally and Dinesh Govindaraj *

Other works in Sparsity & CS:

IEEE TRANSACTIONS ON MEDICAL IMAGING, VOL. 31, NO. 12, DECEMBER 2012

Compressed Sensing Based Real-Time Dynamic MRI Reconstruction

Angshul Majumdar*, Rabab K. Ward, and Tyseer Aboulnasr

PERFORMANCE EVALUATION OF ACCELERATED FUNCTIONAL MRI ACQUISITION USING COMPRESSED SENSING

Hong Jung, Jong Chul Ye

Interpolated Compressed Sensing for 2D Multiple Slice Fast MR Imaging

Yong Pang[1¶], Xiaoliang Zhang[1,2,3*¶]

Small-sample brain mapping: sparse recovery on spatially correlated designs with randomization and clustering

Gaël Varoquaux GAEL.VAROQUAUX@INRIA.FR
Alexandre Gramfort ALEXANDRE.GRAMFORT@INRIA.FR
Bertrand Thirion BERTRAND.THIRION@INRIA.FR

Other works in EEG, fMRI, etc:

IEEE TRANSACTIONS ON INFORMATION TECHNOLOGY IN BIOMEDICINE, VOL. 13, NO. 5, SEPTEMBER 2009

Feature-Based Fusion of Medical Imaging Data

Vince D. Calhoun, *Senior Member, IEEE*, and Tülay Adalı, *Fellow, IEEE*

IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING, VOL. 59, NO. 8, AUGUST 2012

Fractional-Order Time Series Models for Extracting the Haemodynamic Response From Functional Magnetic Resonance Imaging Data

Kurt Barbé*, Wendy Van Moer, and Guy Nagels

Medical Image Analysis 16 (2012) 976–990

Structural analysis of fMRI data: A surface-based framework for multi-subject studies

Grégory Operto [a,b], Denis Rivière [b], Bernard Fertil [a], Rémy Bulot [a], Jean-François Mangin [b], Olivier Coulon [a,*]

Other works in classification/prediction:

Neural Networks 37 (2013) 1–47

Adaptive Resonance Theory: How a brain learns to consciously attend, learn, and recognize a changing world☆

Stephen Grossberg*

Pattern Recognition 45 (2012) 2064–2074

Decoding visual brain states from fMRI using an ensemble of classifiers

Carlos Cabral [a], Margarida Silveira [a,b,*], Patricia Figueiredo [a,b]

IEEE TRANSACTIONS ON MEDICAL IMAGING, VOL. 29, NO. 2, FEBRUARY 2010 531

Random Subspace Ensembles for fMRI Classification

Ludmila I. Kuncheva*, *Member, IEEE*, Juan J. Rodríguez, *Member, IEEE*, Catrin O. Plumpton, David E. J. Linden, and Stephen J. Johnston

30 IEEE TRANSACTIONS ON MEDICAL IMAGING, VOL. 29, NO. 1, JANUARY 2010

Comparison of AdaBoost and Support Vector Machines for Detecting Alzheimer's Disease Through Automated Hippocampal Segmentation

Jonathan H. Morra, Zhuowen Tu, Liana G. Apostolova, Amity E. Green, Arthur W. Toga, and Paul M. Thompson*

Future directions:

- Sparsity and CS in fMRI signals (EEG ?) – <u>acquisition</u> & <u>modeling</u>
- Complex-valued fMRI signal ($\Rightarrow$ complex ICA, features)
- Adaptive algorithms (customized per-modality)
- Distributed algorithms (...)
- “deep” analysis of brain signals for training/prediction
- application to real-world cognitive & diagnostic tasks

UbiComp'11 / Beijing, China

The Social fMRI: Measuring, Understanding, and Designing Social Mechanisms in the Real World

Nadav Aharony[1], Wei Pan[1], Cory Ip[1], Inas Khayal[1,2], Alex Pentland[1]
{nadav, panwei, coryip, ikhayal, pentland}@media.mit.edu
[1]The Media Lab, Massachusetts Institute of Technology, Cambridge, MA, USA
[2]Masdar Institute of Science and Technology, Masdar City, Abu Dhabi

???