~~When machines look at neurons: learning from neuroscience time series~~

Why neuroscience needs ML / Stats and CS

Alexandre Gramfort

alexandre.gramfort@telecom-paristech.fr

Assistant Professor

CNRS LTCI, Télécom ParisTech, Université Paris-Saclay

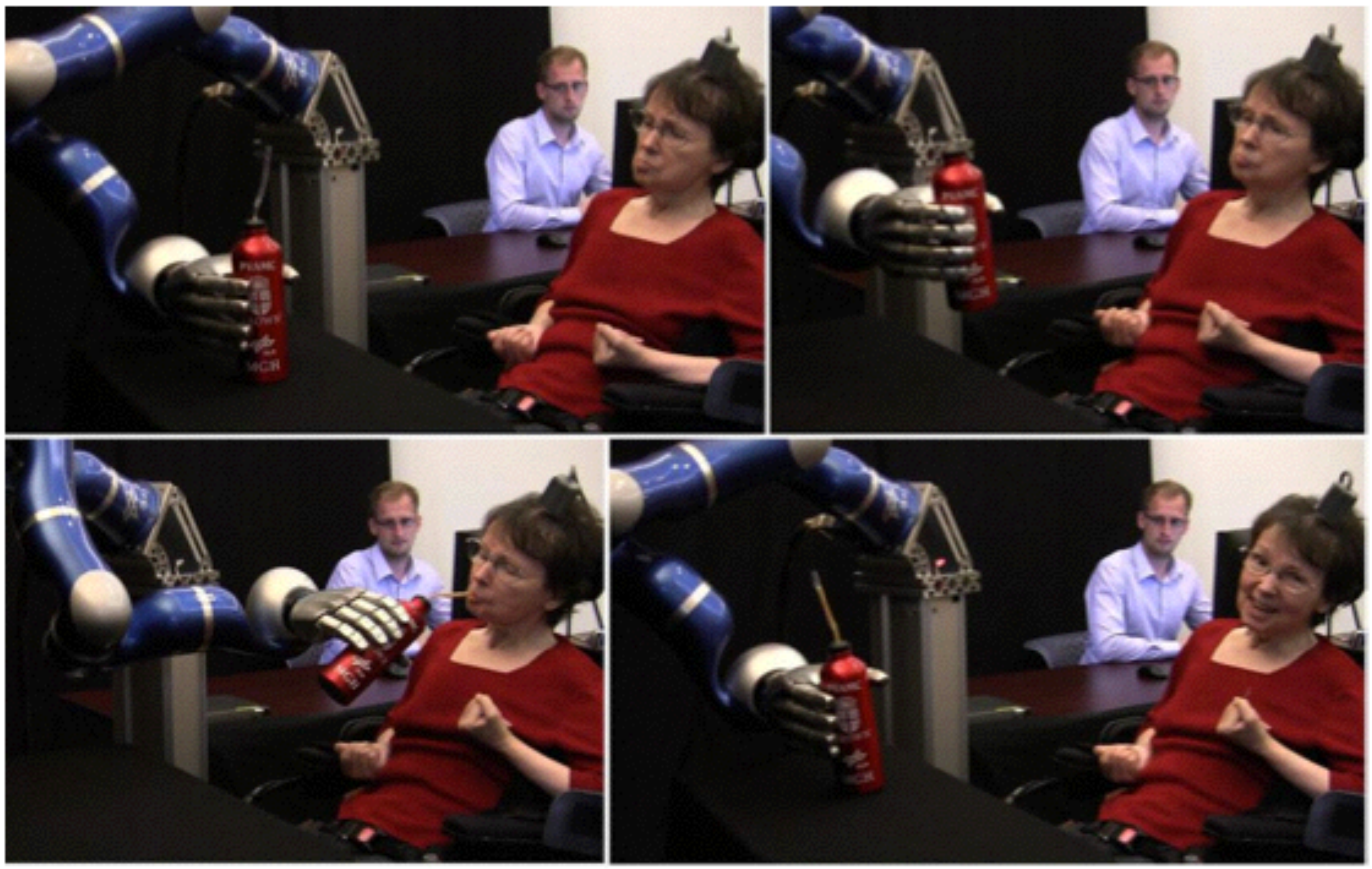
Paris-Saclay Center for Data Science

JSTAR - 20 Oct. 2016

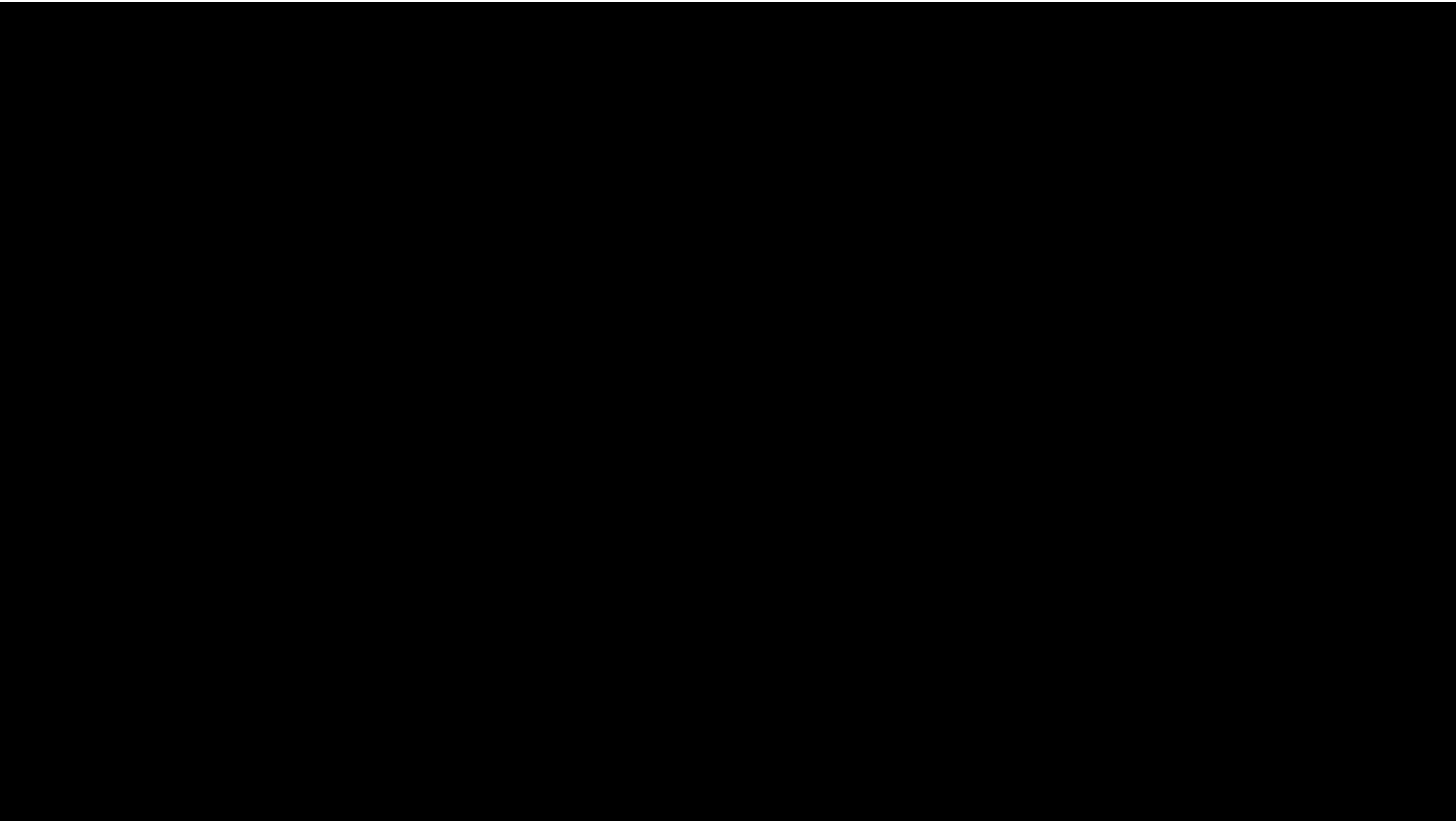
nature

Reach and grasp by people with tetraplegia using a neurally controlled robotic arm. 2012.

Leigh R. Hochberg1,2,4,5, **Daniel Bacher**2, **Beata Jarosiewicz**1,3, **Nicolas Y. Masse**3, **John D. Simeral**1,2,4, **Joern Vogel**6, **Sami Haddadin**6, **Jie Liu**1,2, **Sydney S. Cash**4,5, **Patrick van der Smagt**6, and **John P. Donoghue**1,2,3



https://www.youtube.com/watch?v=QRt8QCx3BCo

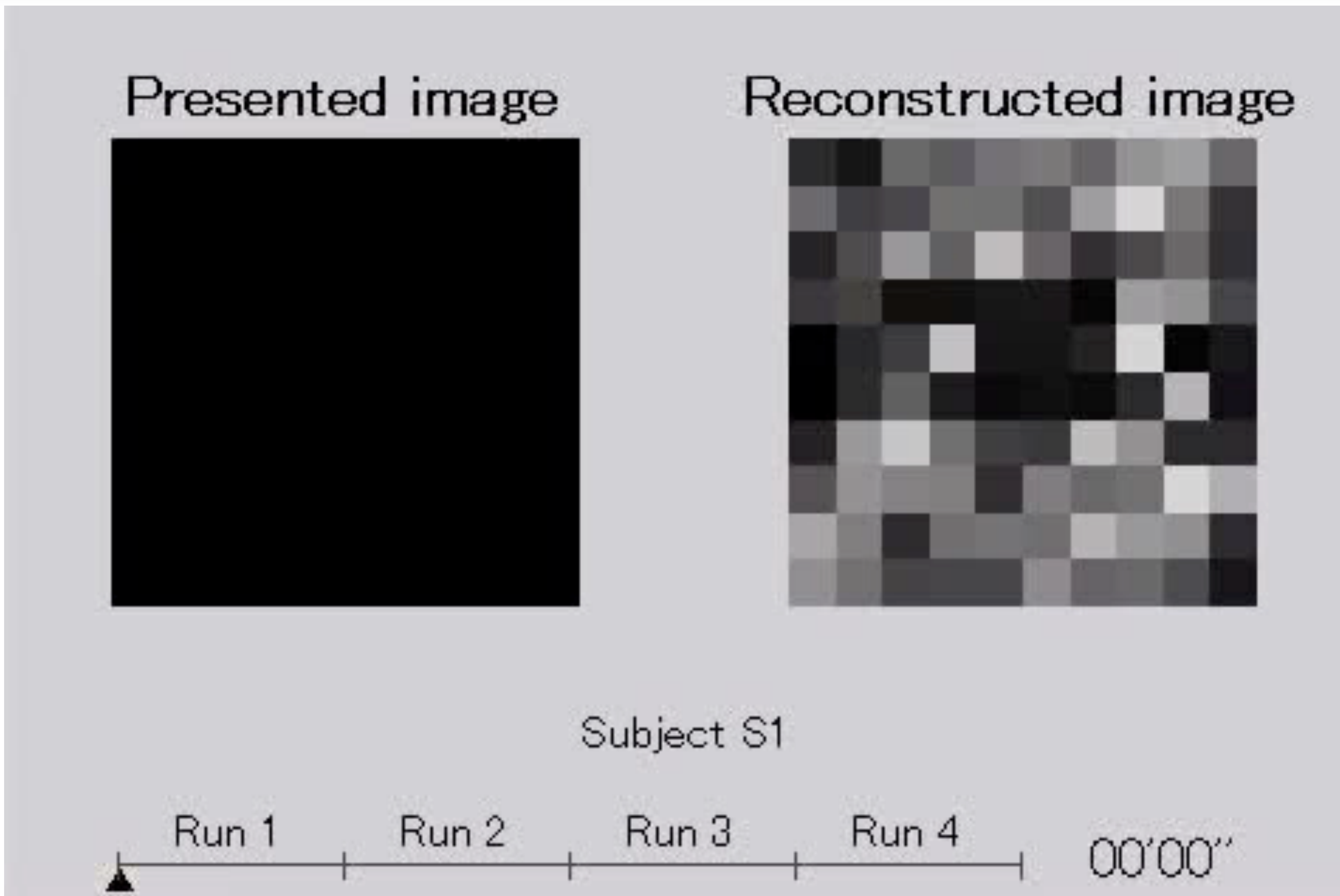


https://www.youtube.com/watch?v=QRt8QCx3BCo

Neuron
Article
Cell PRESS

Visual Image Reconstruction from Human Brain Activity using a Combination of Multiscale Local Image Decoders

Yoichi Miyawaki,[1,2,6] Hajime Uchida,[2,3,6] Okito Yamashita,[2] Masa-aki Sato,[2] Yusuke Morito,[4,5] Hiroki C. Tanabe,[4,5] Norihiro Sadato,[4,5] and Yukiyasu Kamitani[2,3,*]



http://www.youtube.com/watch?v=h1Gu1YSoDaY

Current Biology 21, 1641–1646, October 11, 2011 ©2011 Elsevier Ltd All rights reserved DOI 10.1016/j.cub.2011.08.031

Report

Reconstructing Visual Experiences from Brain Activity Evoked by Natural Movies

Shinji Nishimoto,[1] An T. Vu,[2] Thomas Naselaris,[1]
Yuval Benjamini,[3] Bin Yu,[3] and Jack L. Gallant[1,2,4,*]

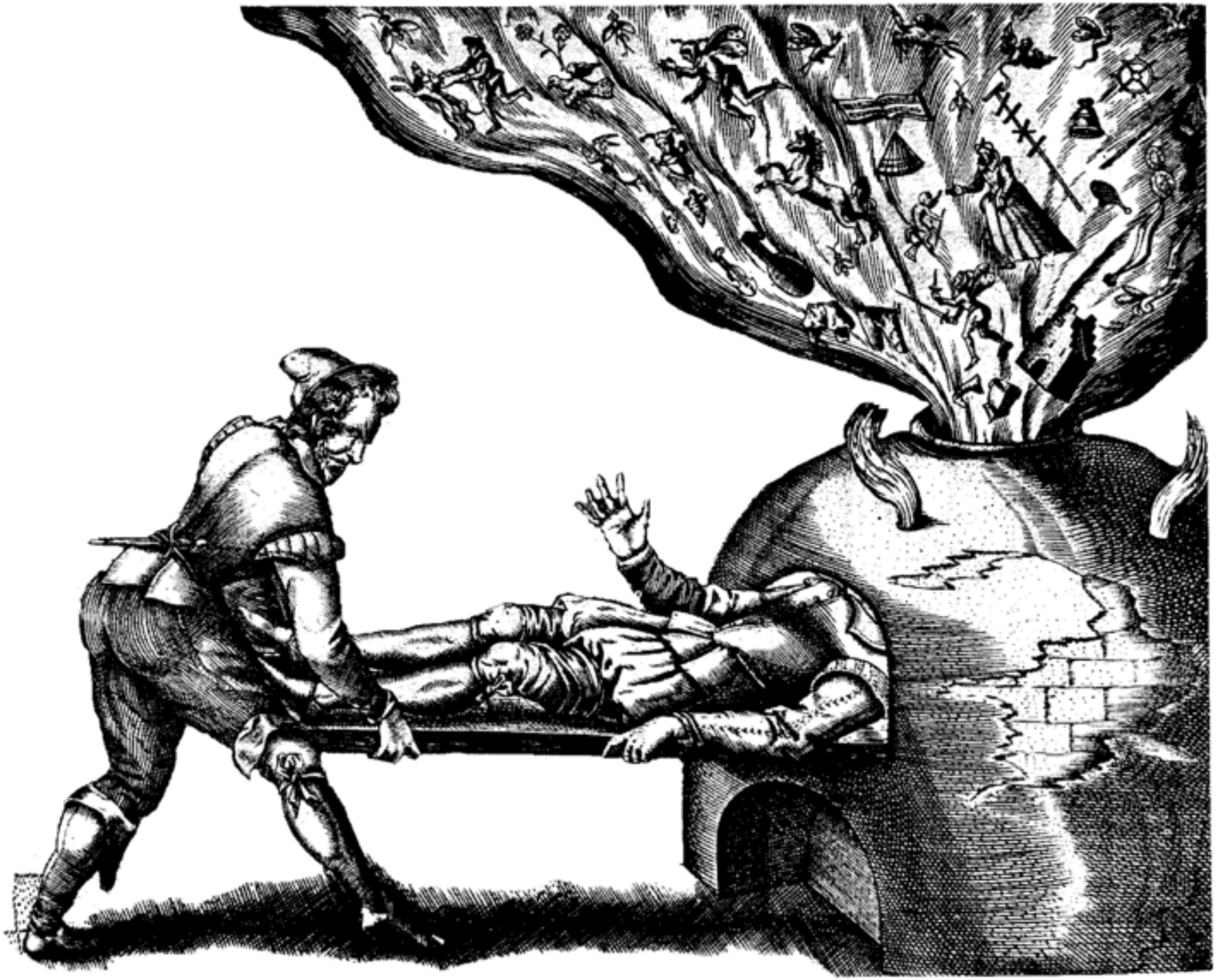
mental processes. It has therefore been assumed that fMRI data would not be useful for modeling brain activity evoked

Presented clip

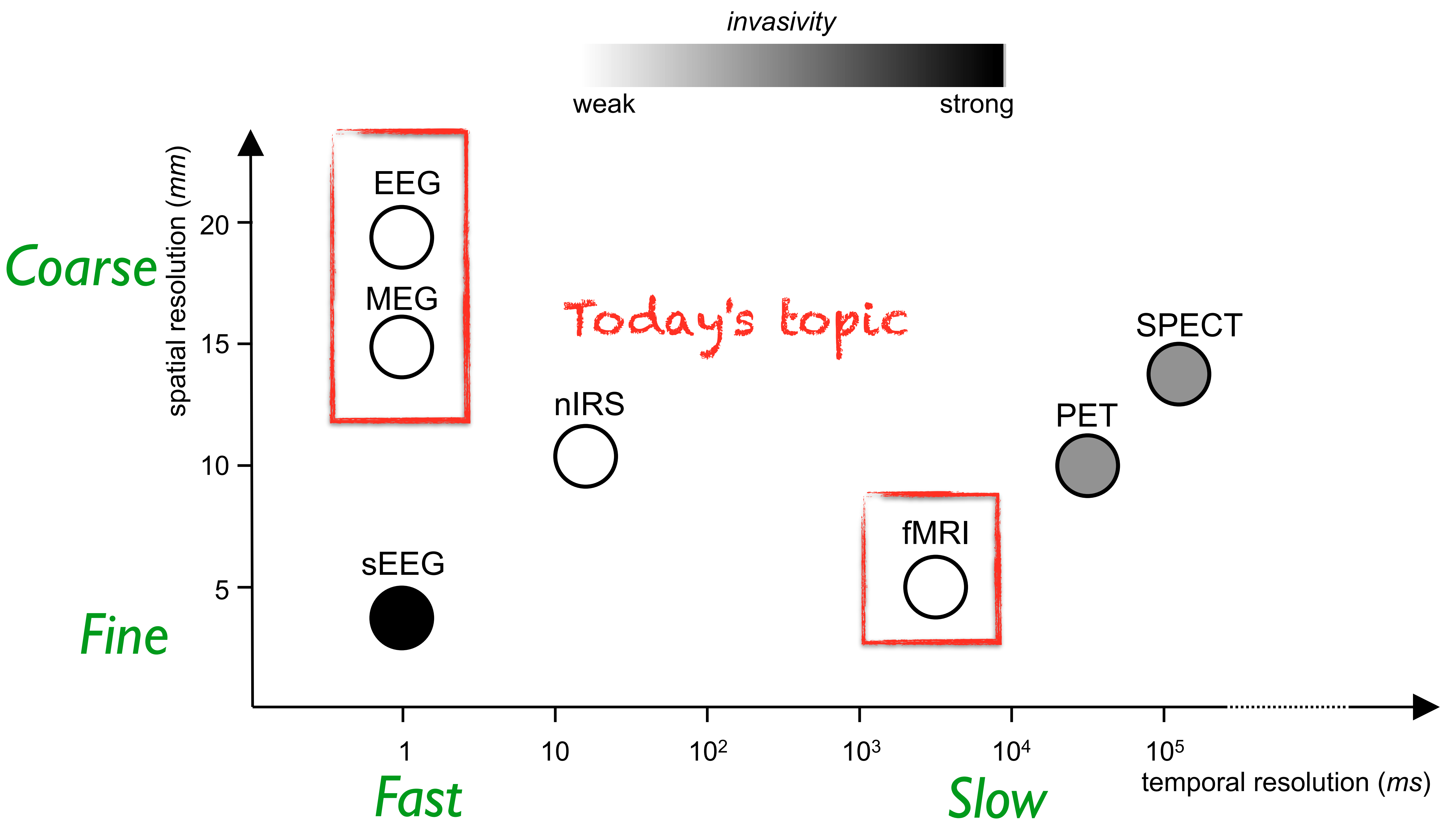
Clip reconstructed from brain activity

http://www.youtube.com/watch?v=nsjDnYxJ0bo

What is functional brain imaging?



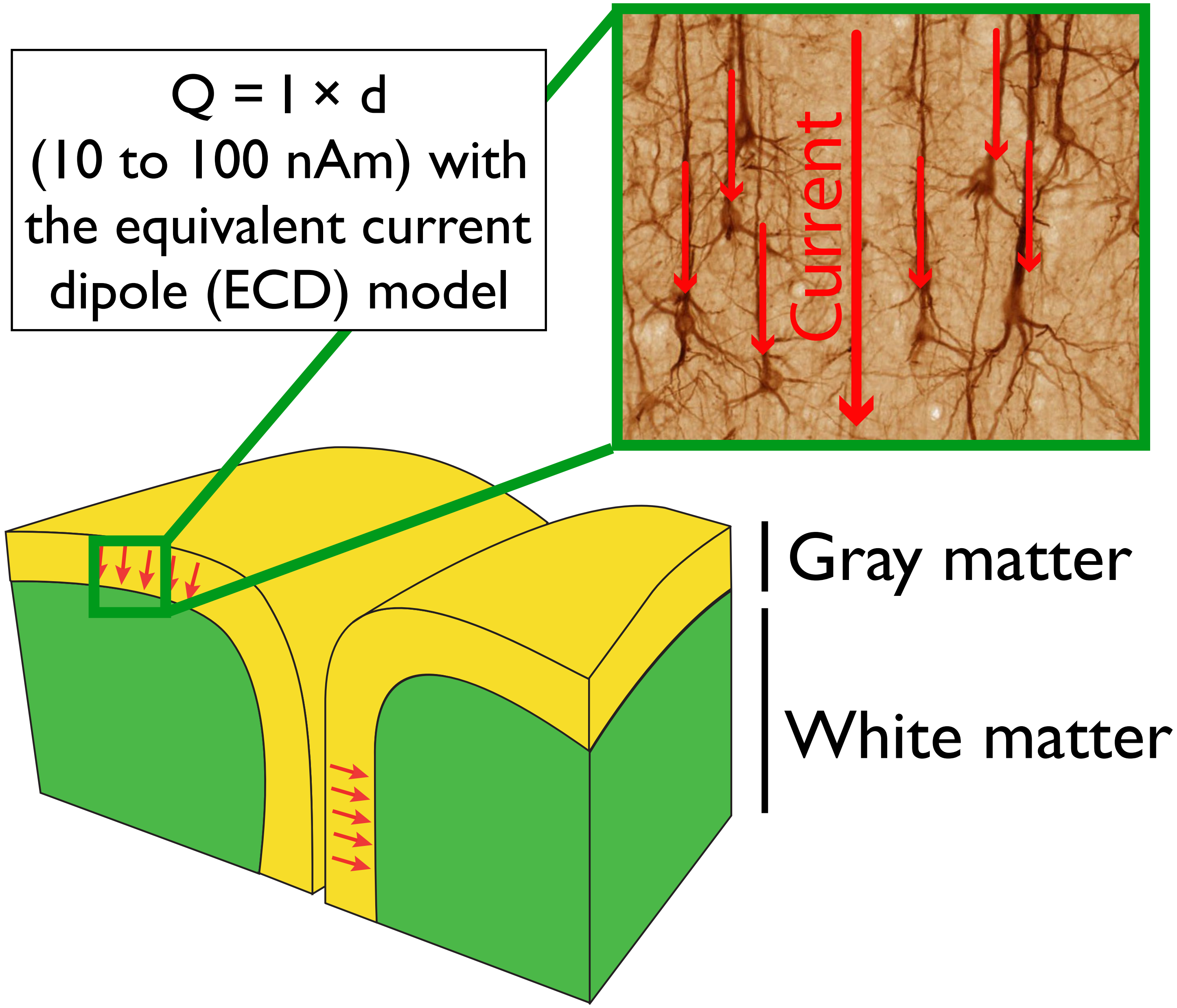
What is functional brain imaging?

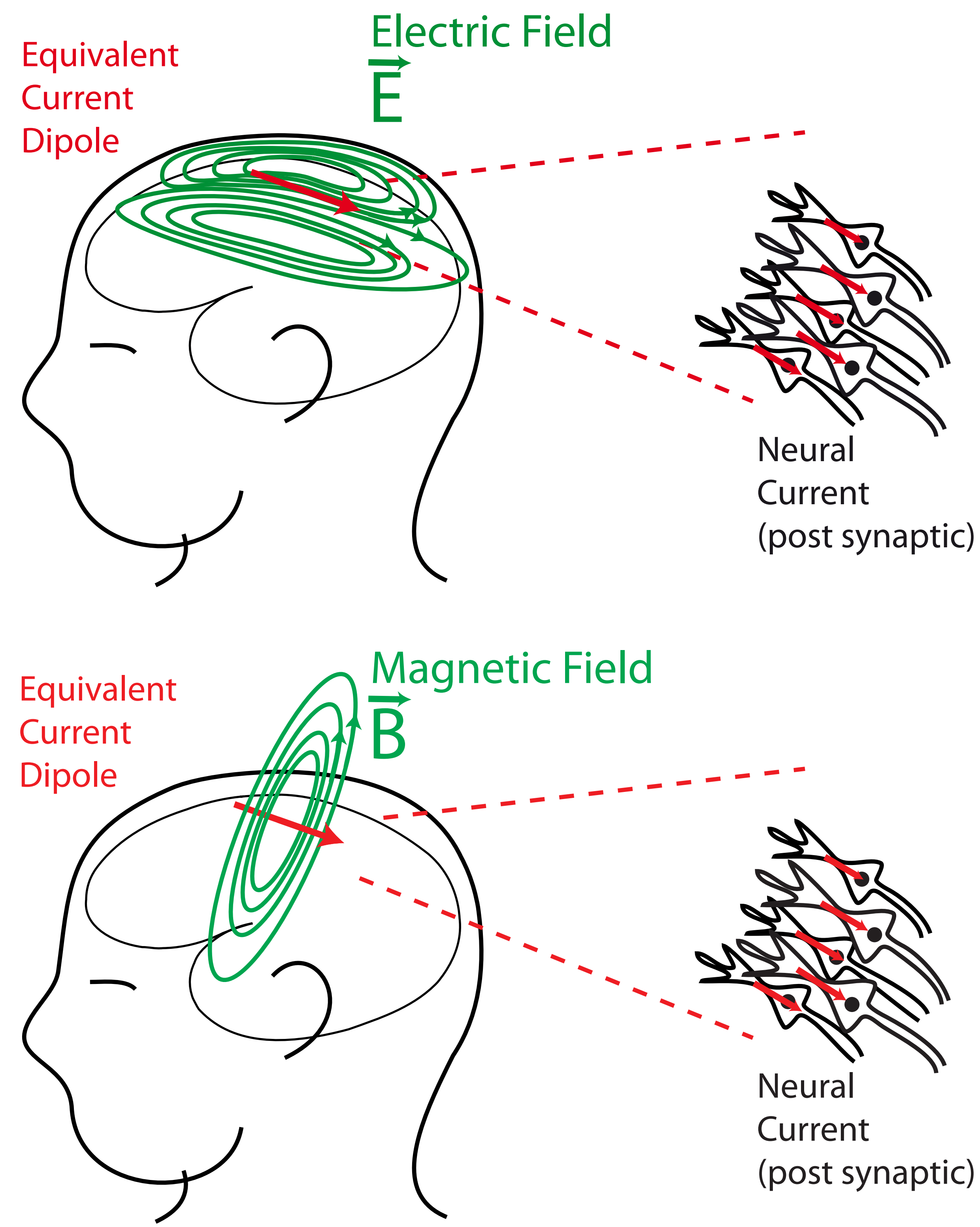


Neurons as current generators

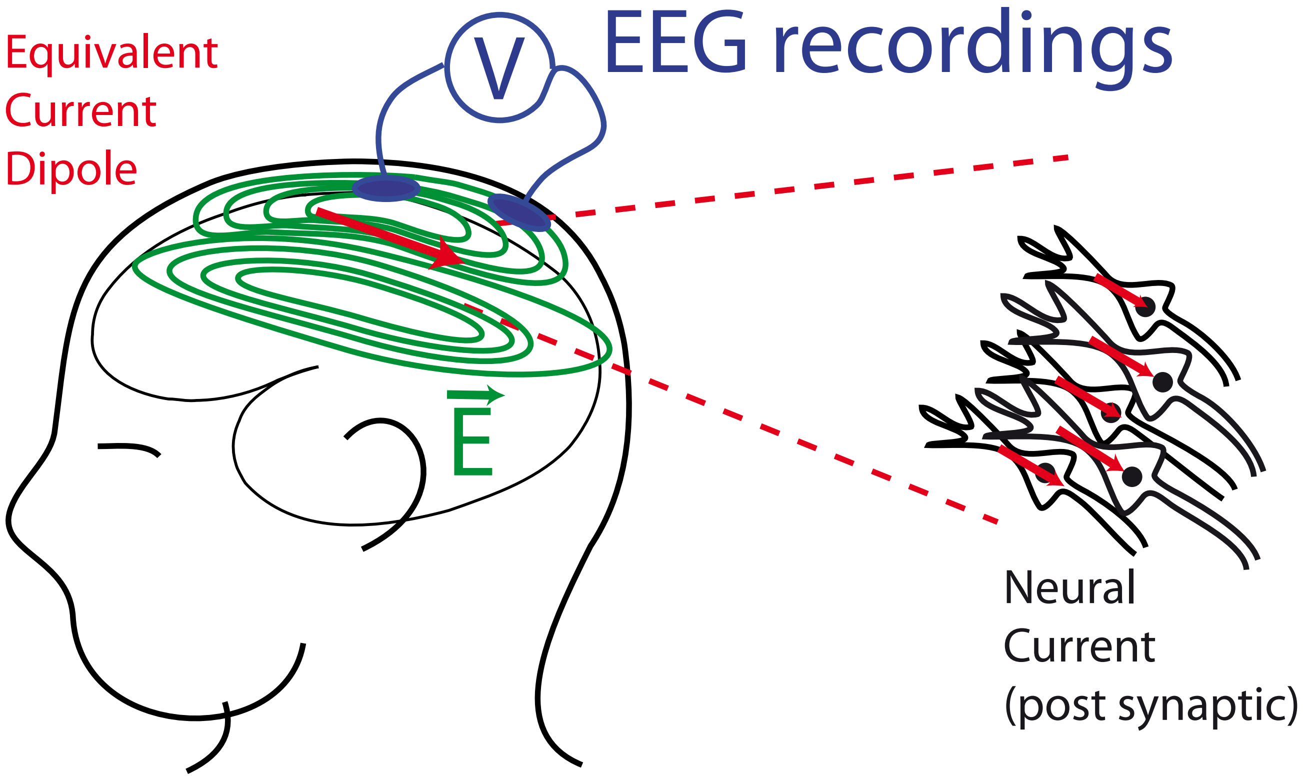
Large cortical pyramidal cells organized in macro-assemblies with their **dendrites normally oriented to the local cortical surface**

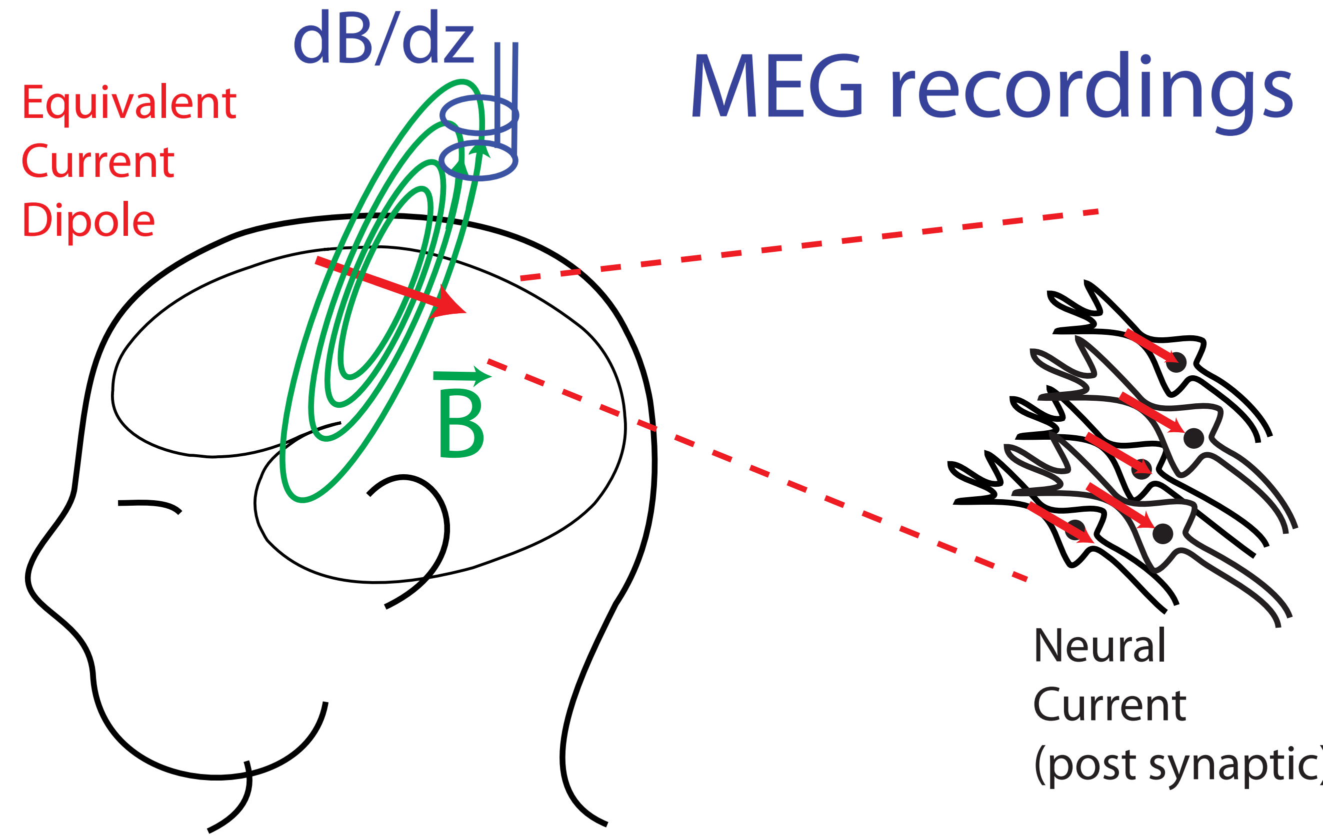
Q = I × d
(10 to 100 nAm) with the equivalent current dipole (ECD) model



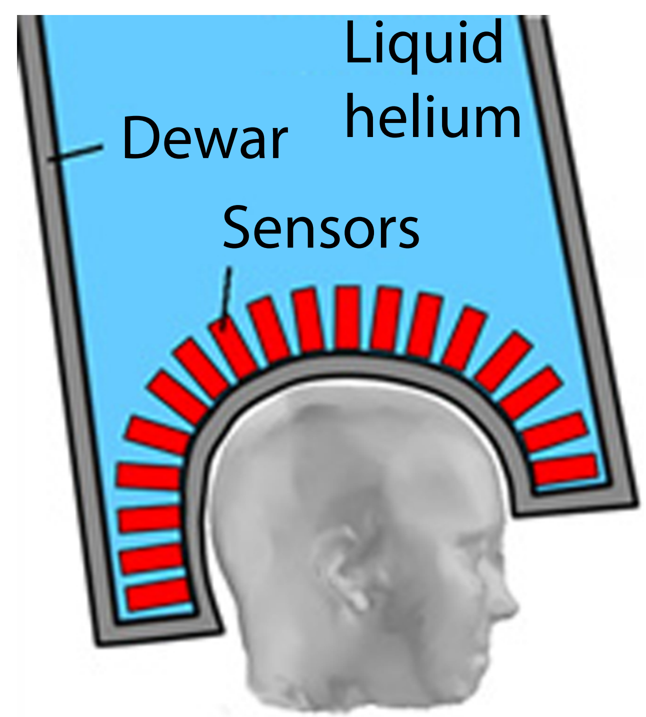


Electro- & Magneto-encephalography





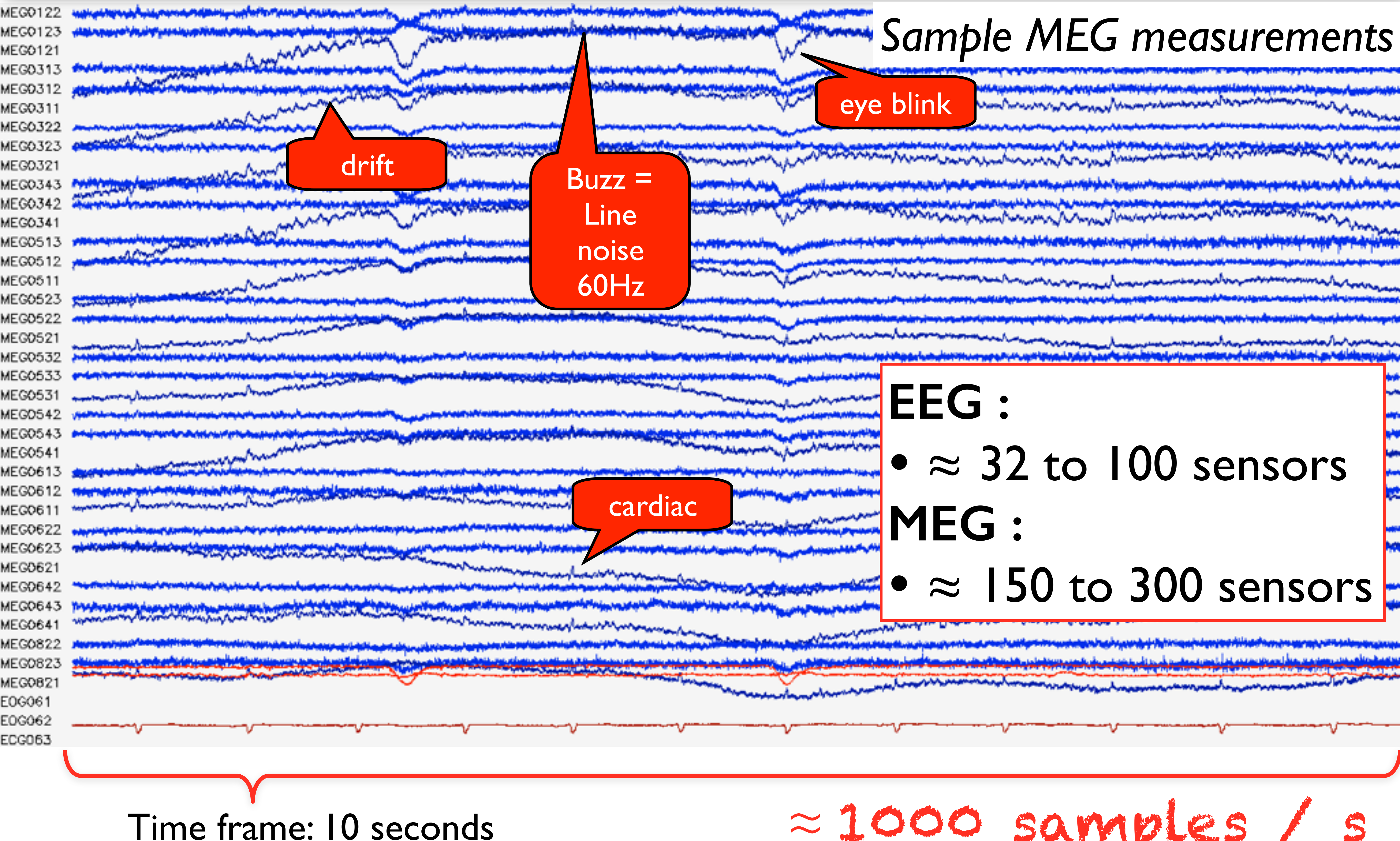
First EEG recordings in 1929 by H. Berger



Hôpital La Timone
Marseille, France

M/EEG Measurements

Sample MEG measurements



EEG :

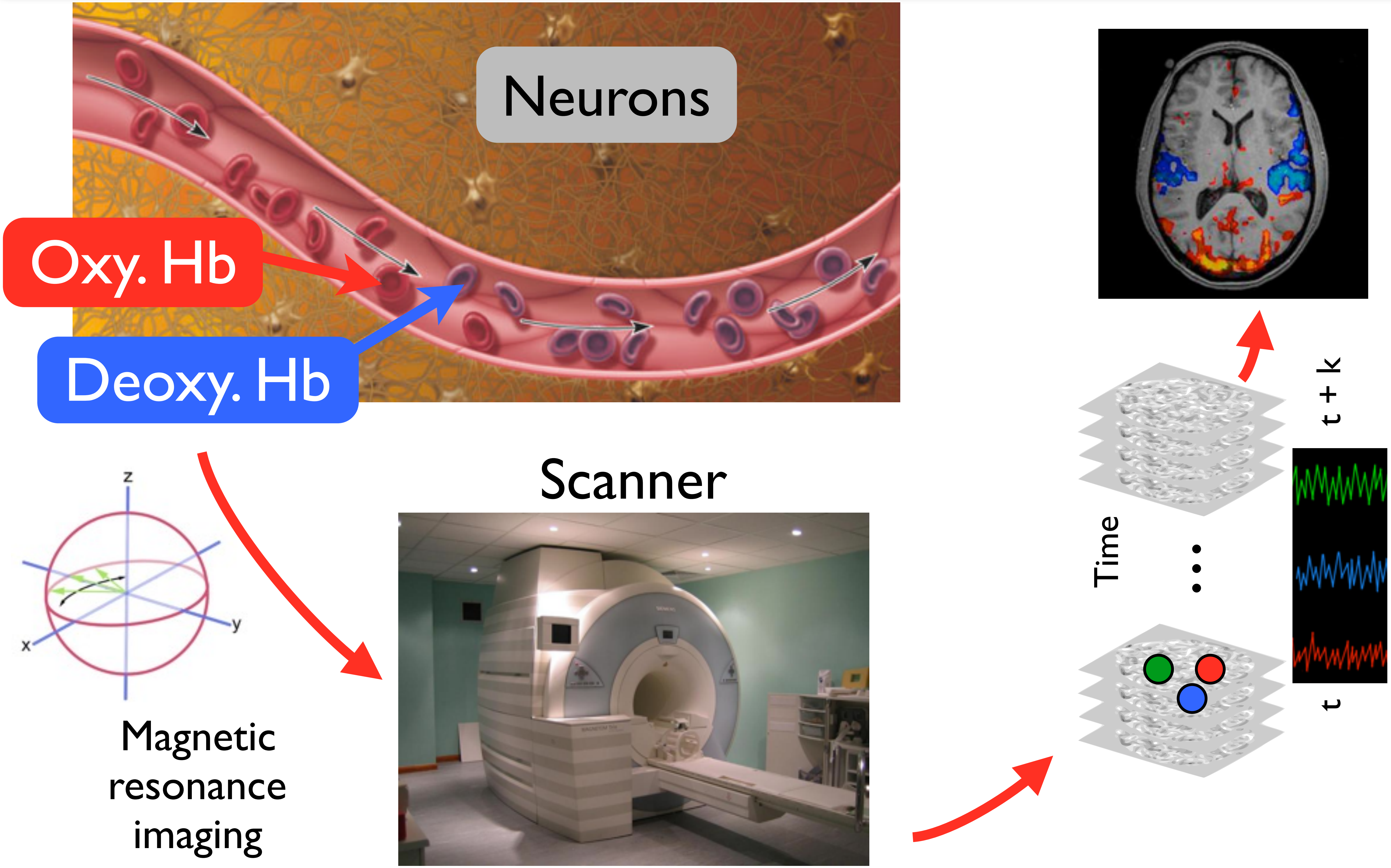
- ≈ 32 to 100 sensors

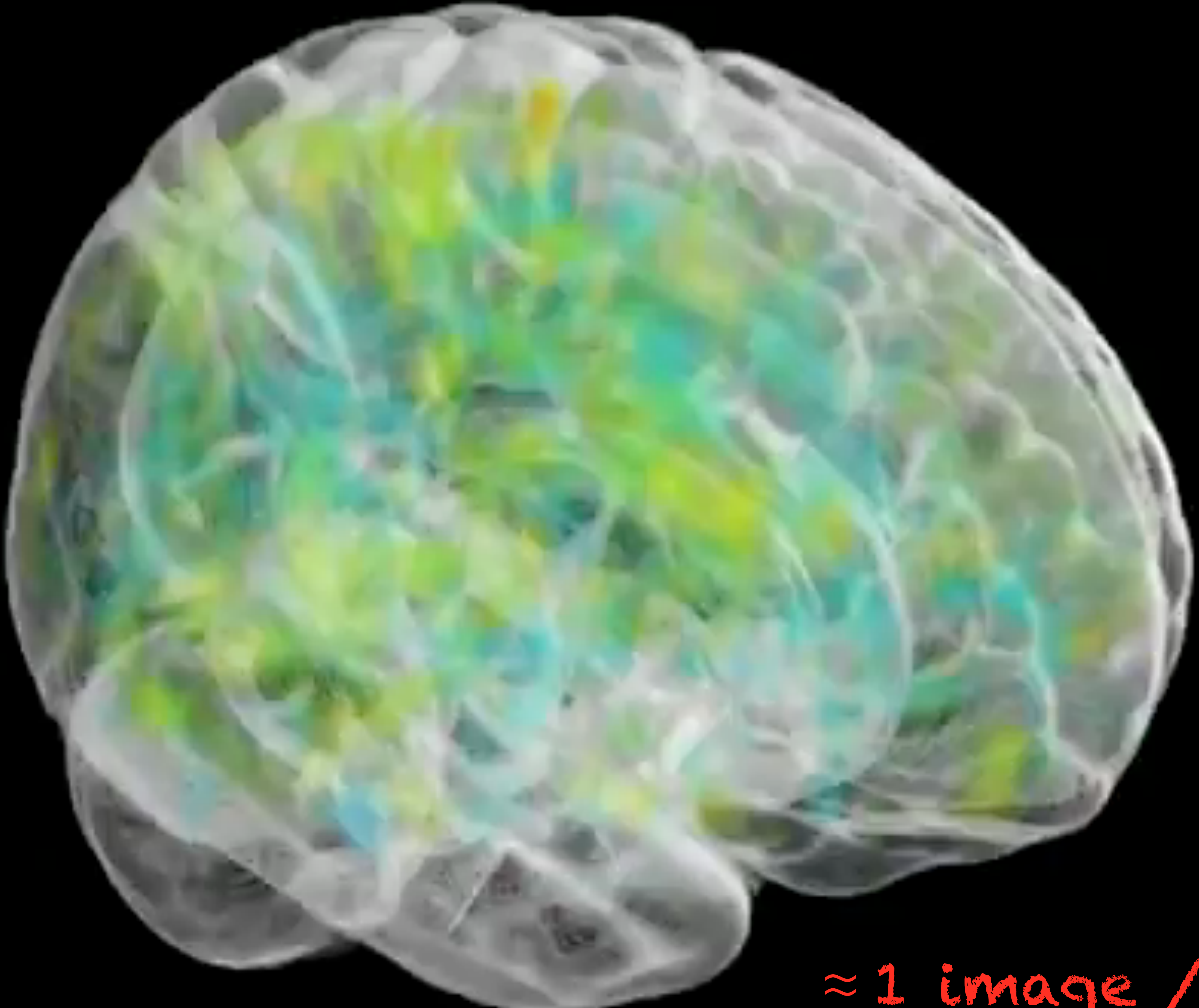
MEG :

- ≈ 150 to 300 sensors

≈ 1000 samples / s

Functional MRI (fMRI)





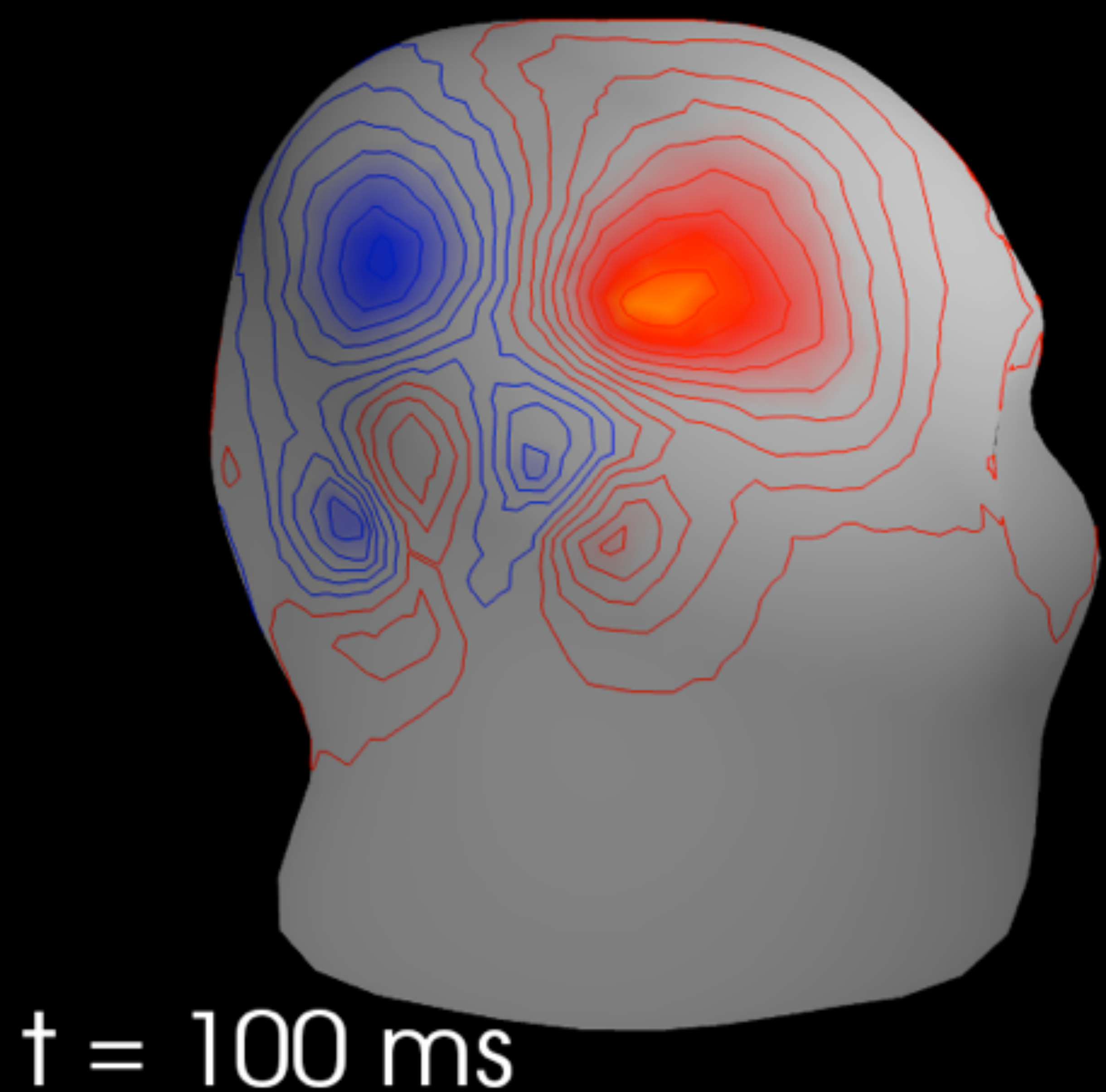
≈1 image / 2s

http://www.youtube.com/watch?v=uhCF-zlk0jY

courtesy of Gael Varoquaux

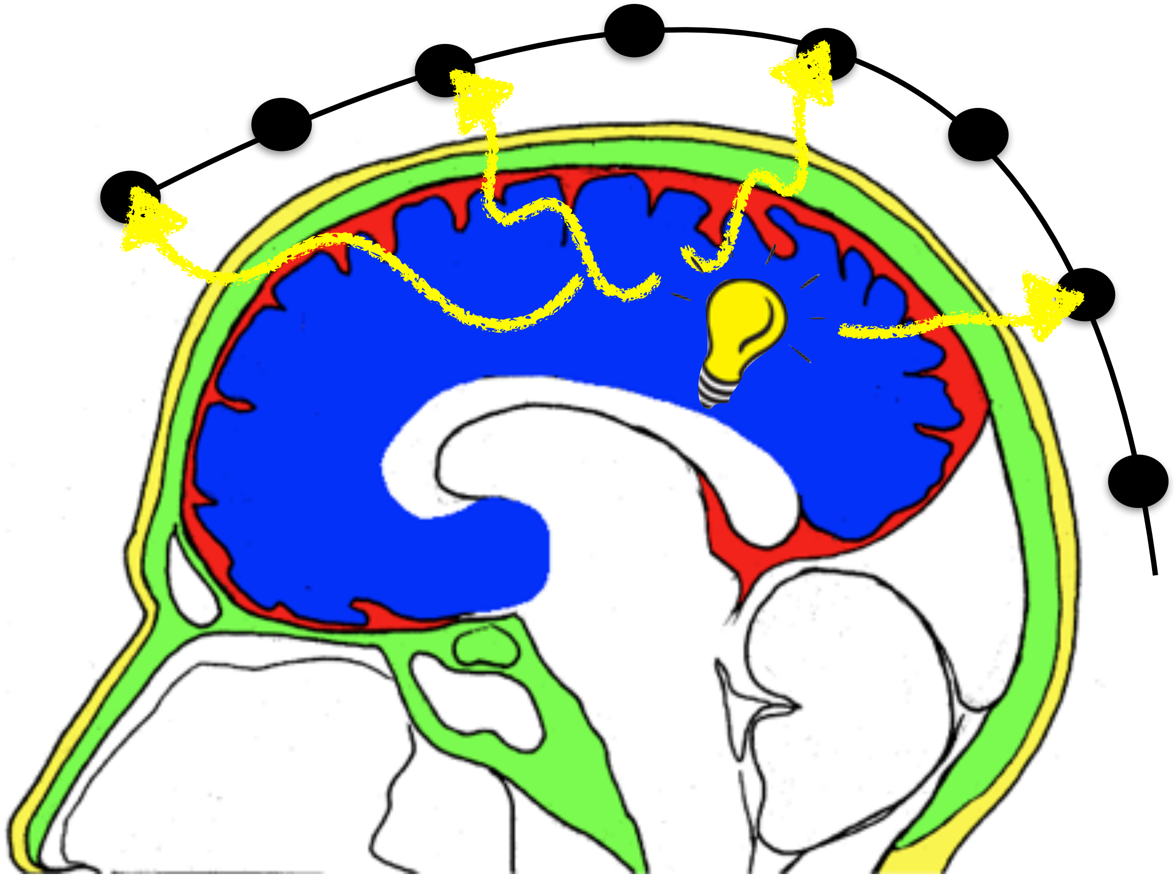
Imaging the brain at a millisecond time scale with MEG and EEG
and stats and optimization

Find the current generators that produced the MEG measurements



What do we measure?

$$\nabla \times \vec{B} = \mu_0 \vec{J}$$

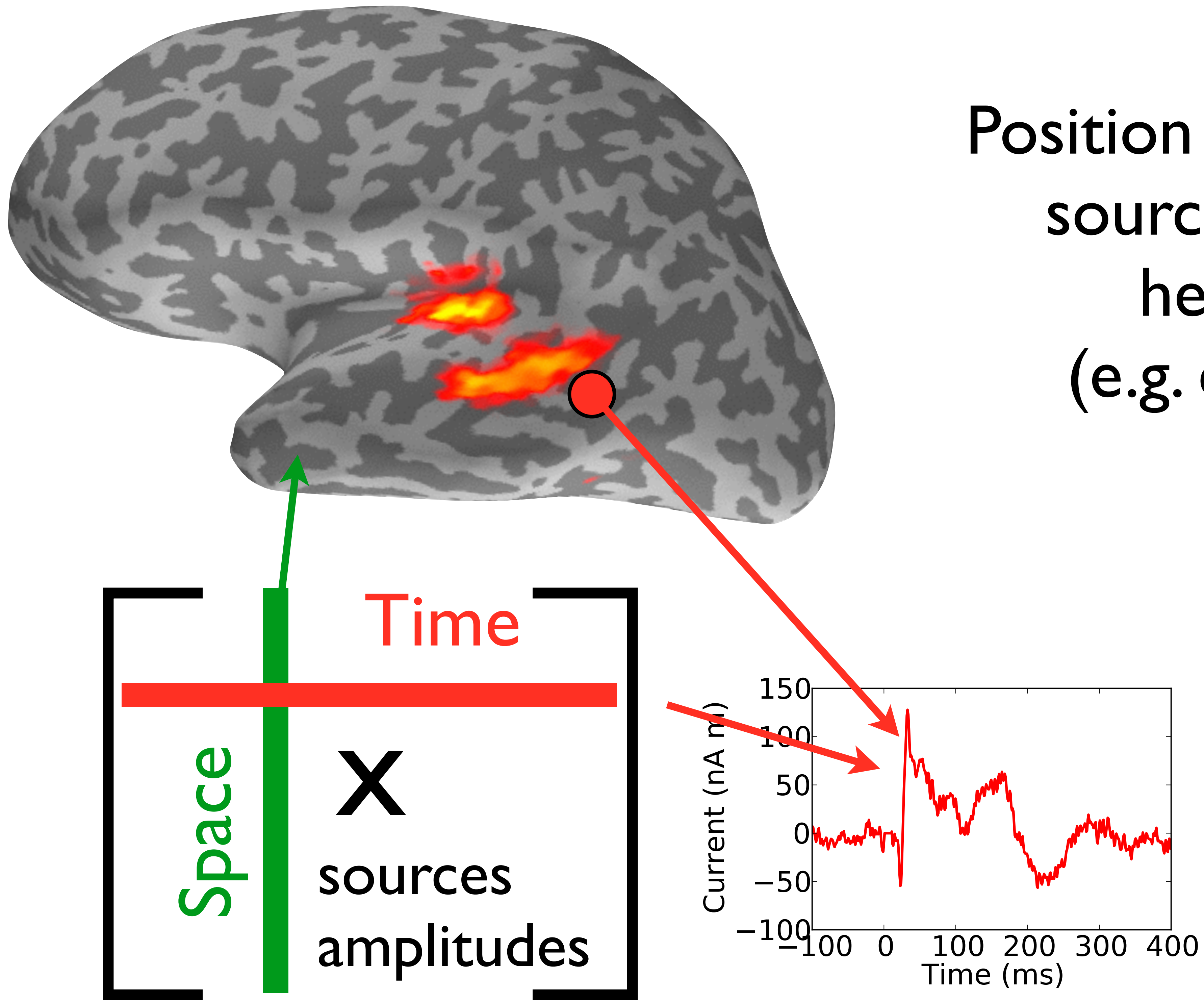


Boundary element method (BEM), i.e., numerical solver with approximate solution.

Linear PDE -> Linear forward problem / Fixed design

[Geselowitz 67, De Munck 92, Kybic et al. 2005, Gramfort et al. 2010]

The source model

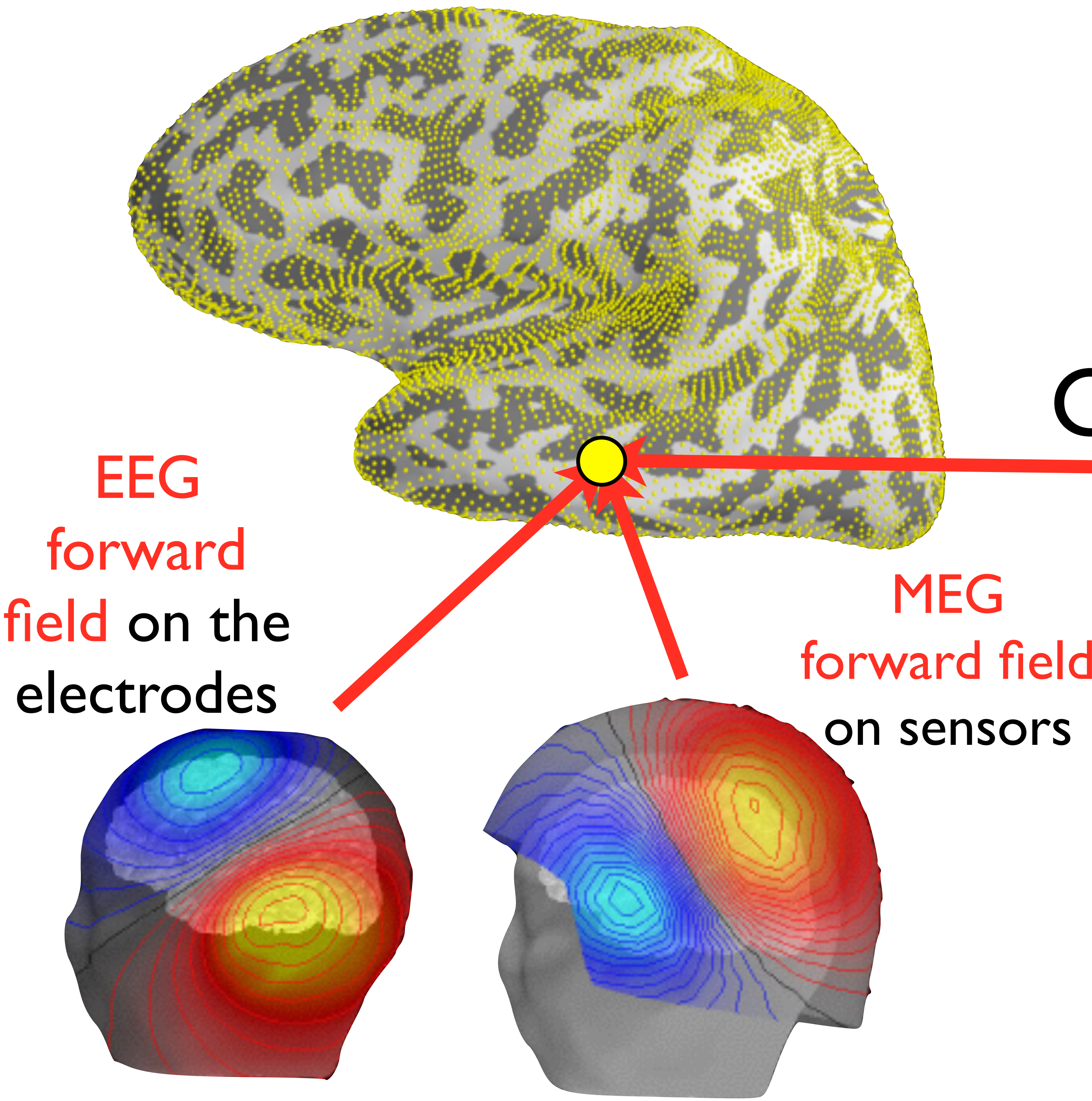


Position 5000 candidate sources over each hemisphere (e.g. every 5mm)

Scalar field defined over time

[Dale and Sereno 93]

The source model

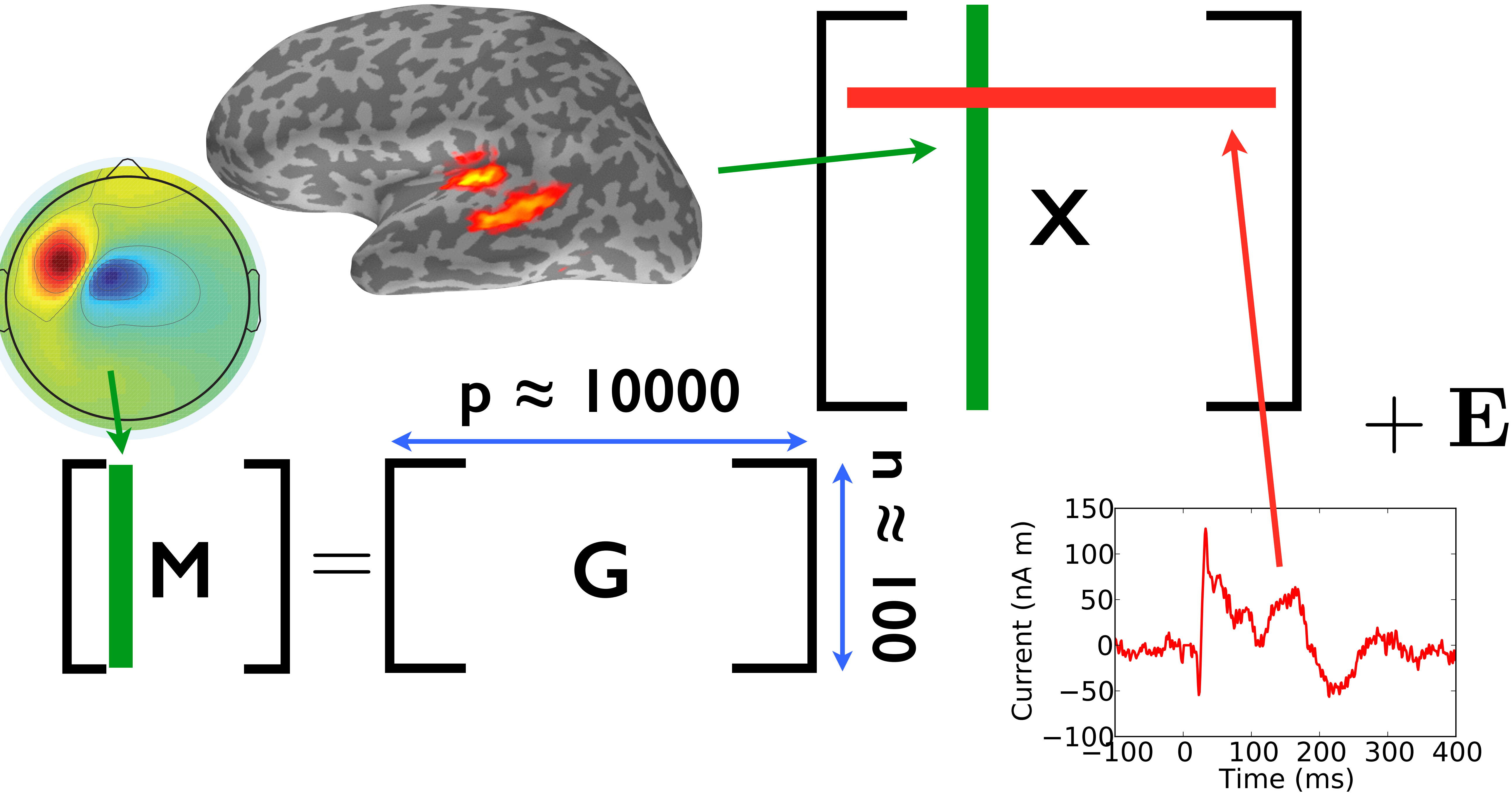


one column = Forward field of one dipole

$$G = \begin{bmatrix} G_{EEG} \\ \hdashline G_{MEG} \end{bmatrix}$$

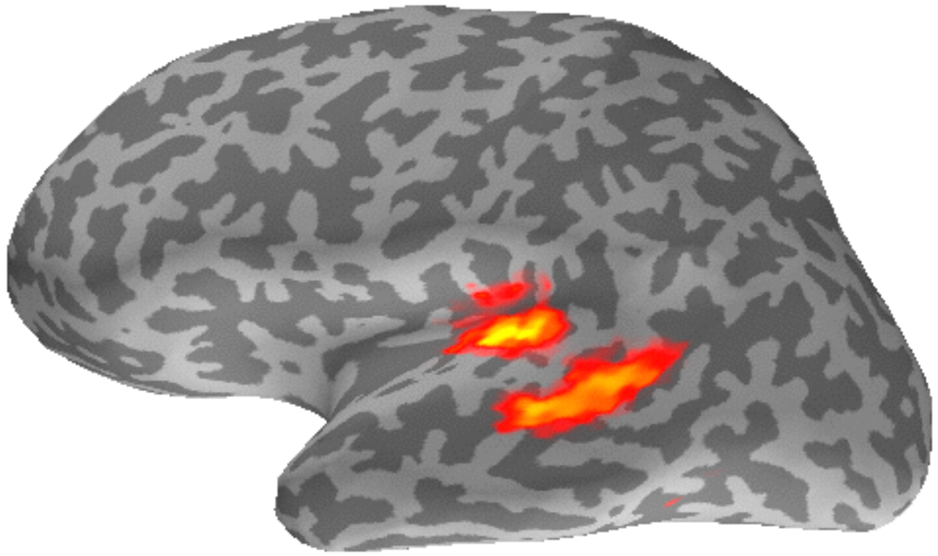
G is the gain matrix / forward operator obtained by concatenation of the forward fields

M = GX+E : An ill-posed problem



Small "n" large "p" problem

y = Xβ+E : An ill-posed problem



Standard statistics notations

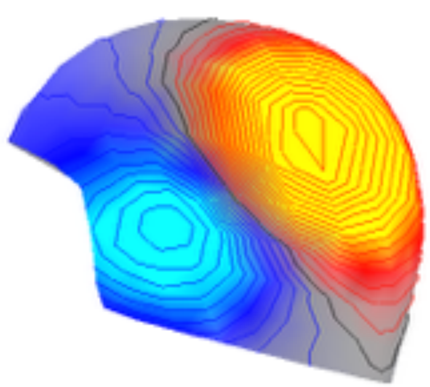
β regression coefficients

$p \approx 10000$

$$\mathbf{y} = \mathbf{X}\,\beta + \mathbf{E}$$

X design matrix

$n \approx 100$



At **each time instant** the M/EEG inverse problem **IS** a **regression** with more variables than observations

Variational formulation

$$\mathbf{X}^* = \arg\min_{\mathbf{X}} \|\mathbf{M} - \mathbf{GX}\|_F^2 + \lambda\phi(\mathbf{X}), \lambda > 0$$

Data fit Regularization

λ : Trade-off between the data fit and the regularization

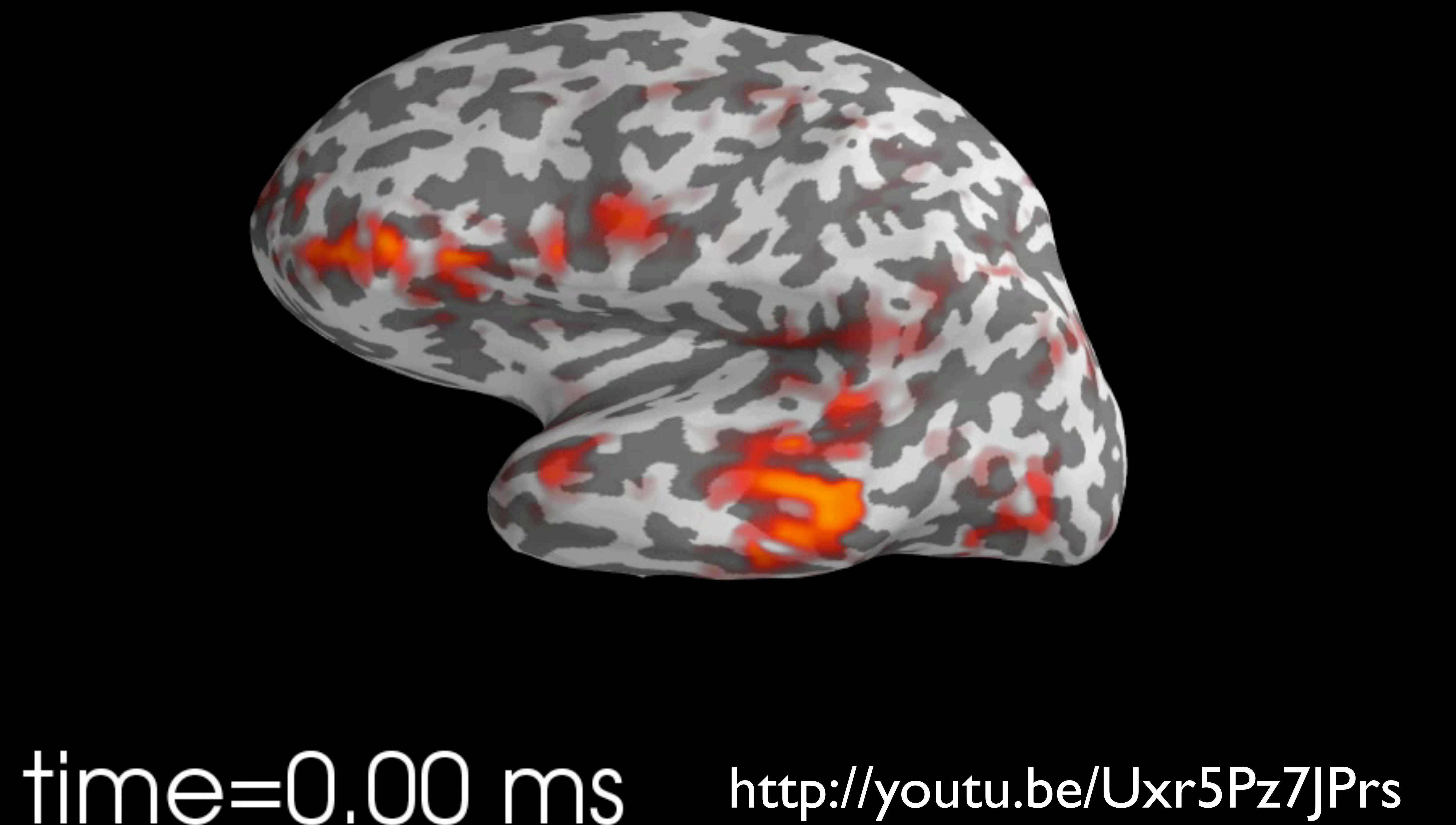
where $\|\mathbf{A}\|_F^2 = \mathbf{tr}(\mathbf{A}^T\mathbf{A})$

Remark: Assumes Gaussian i.i.d. homoscedastic noise…

In practice *heteroscedastic*, *autocorrelated*, :(

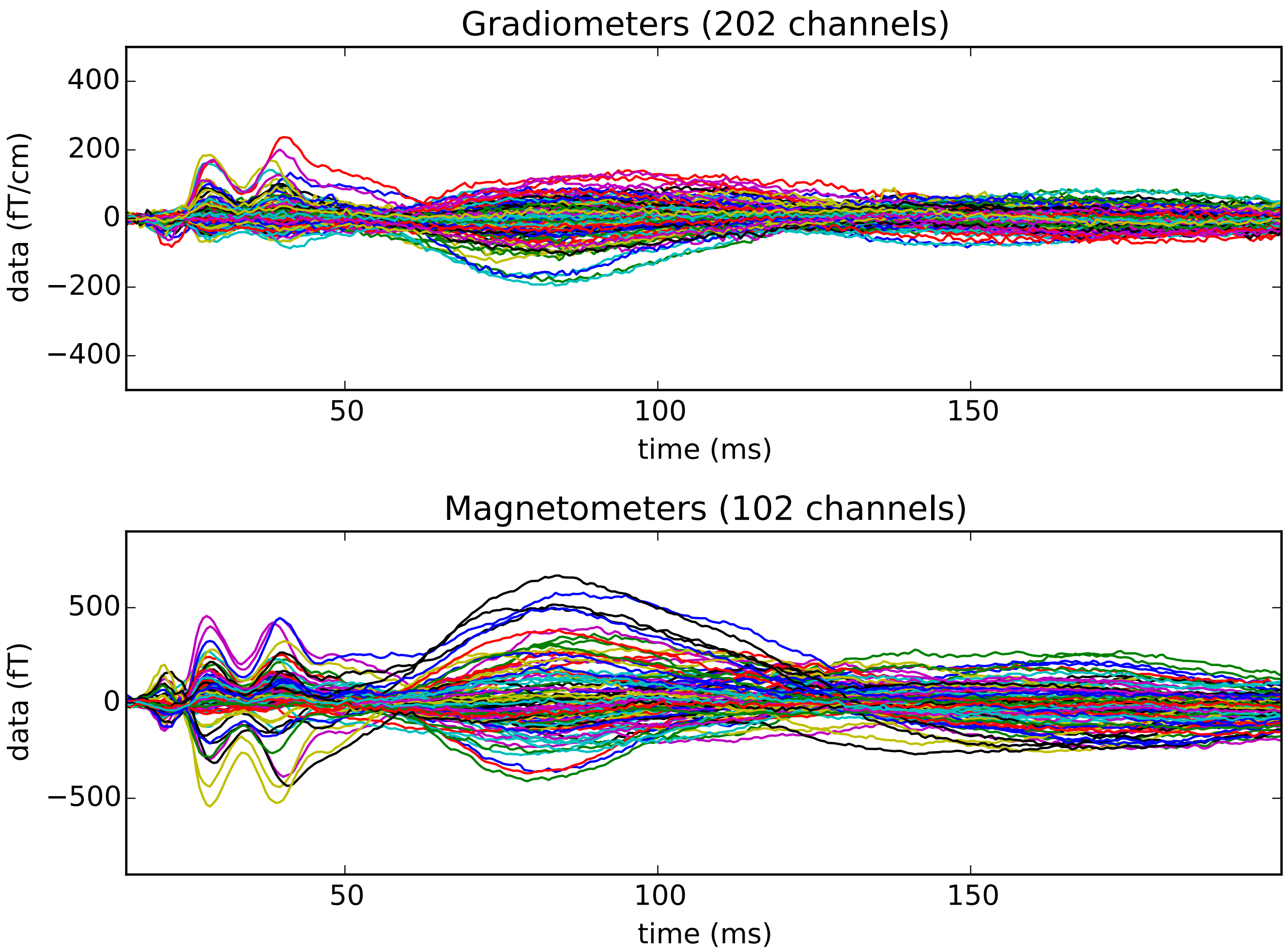
[Engemann & Gramfort NI 2015]

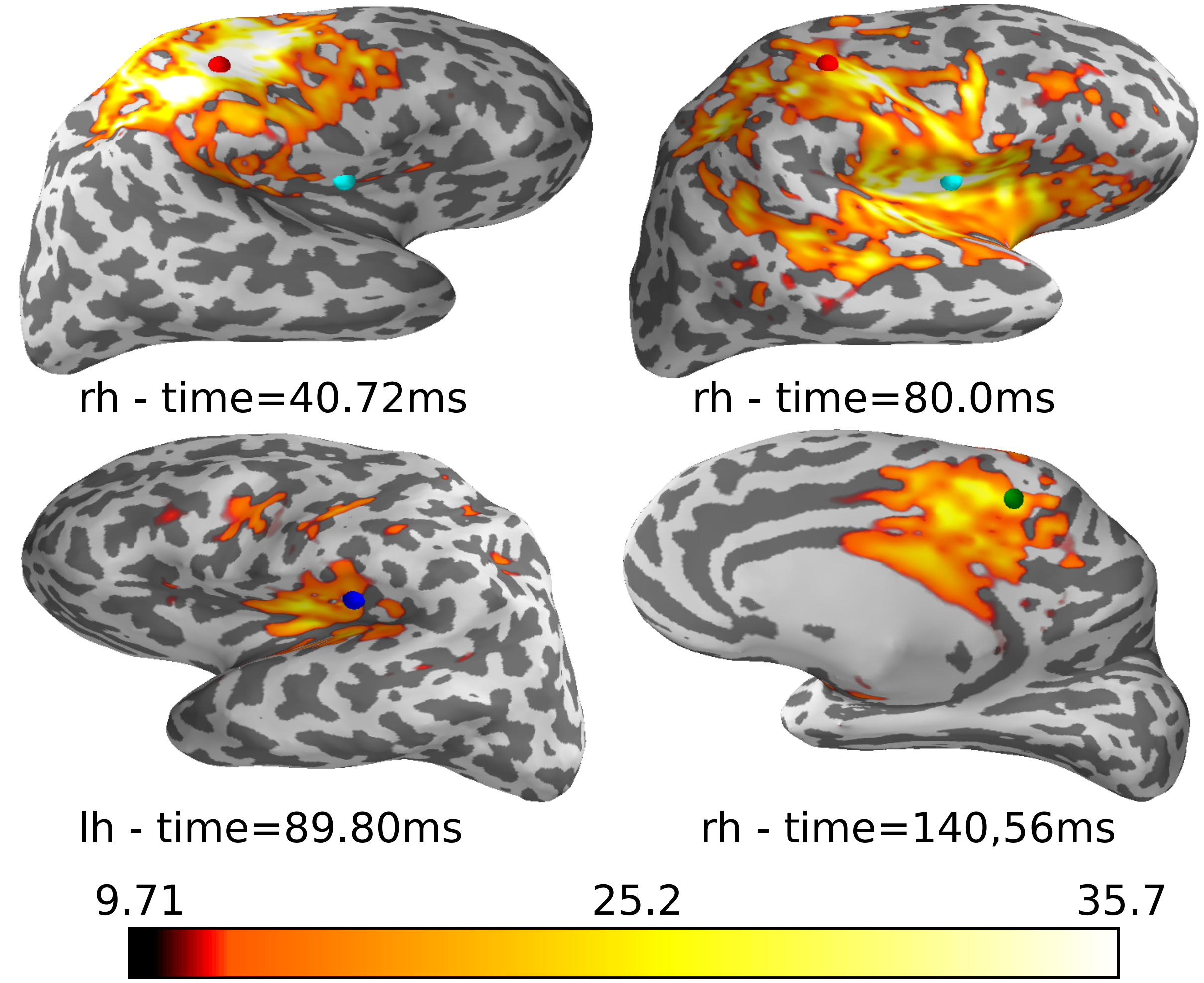
Result obtained with L2 regularization: $\phi(\mathbf{X}) = \|\mathbf{X}\|_F^2$

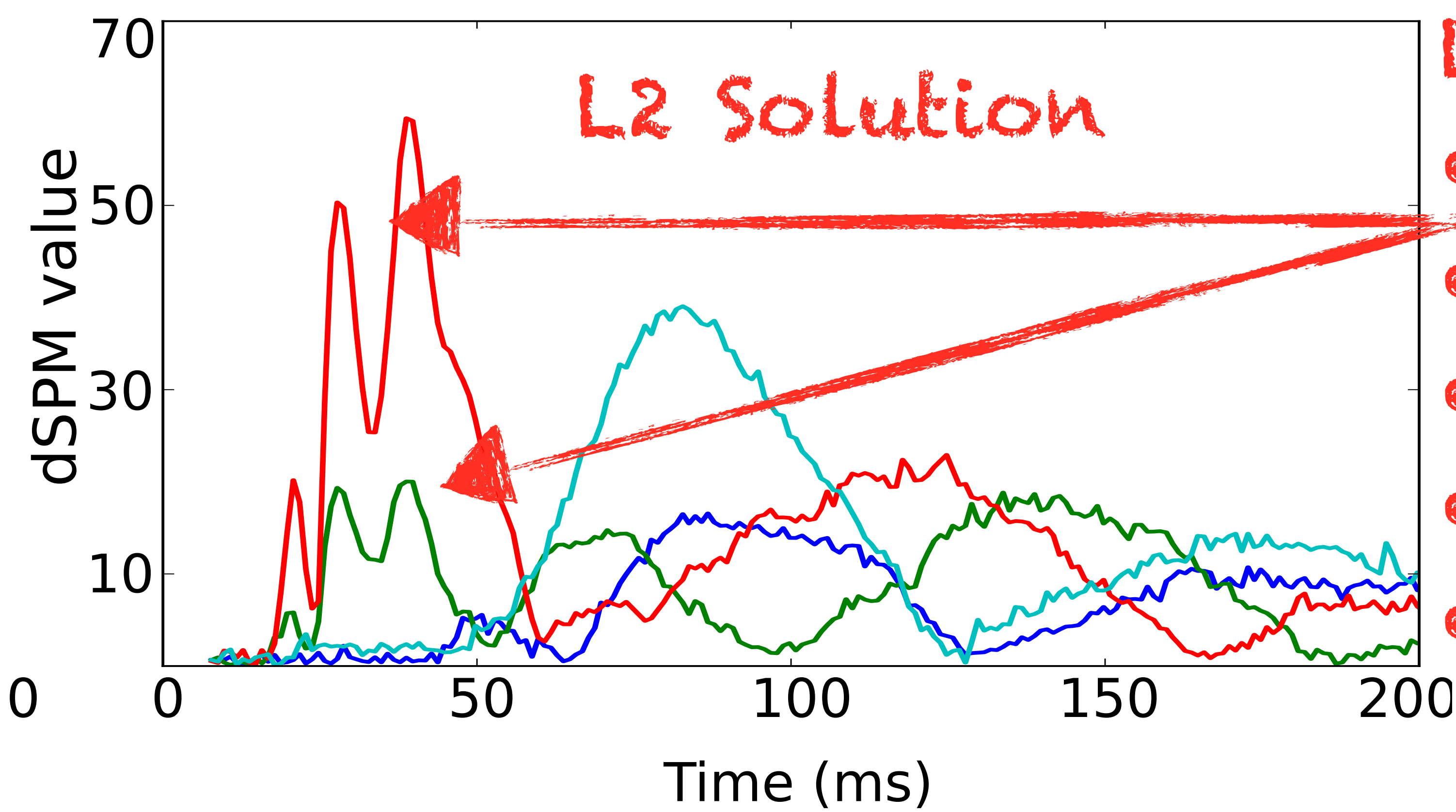


http://youtu.be/Uxr5Pz7JPrs

$$\phi(\mathbf{X}) = \|\mathbf{X}\|_F^2$$



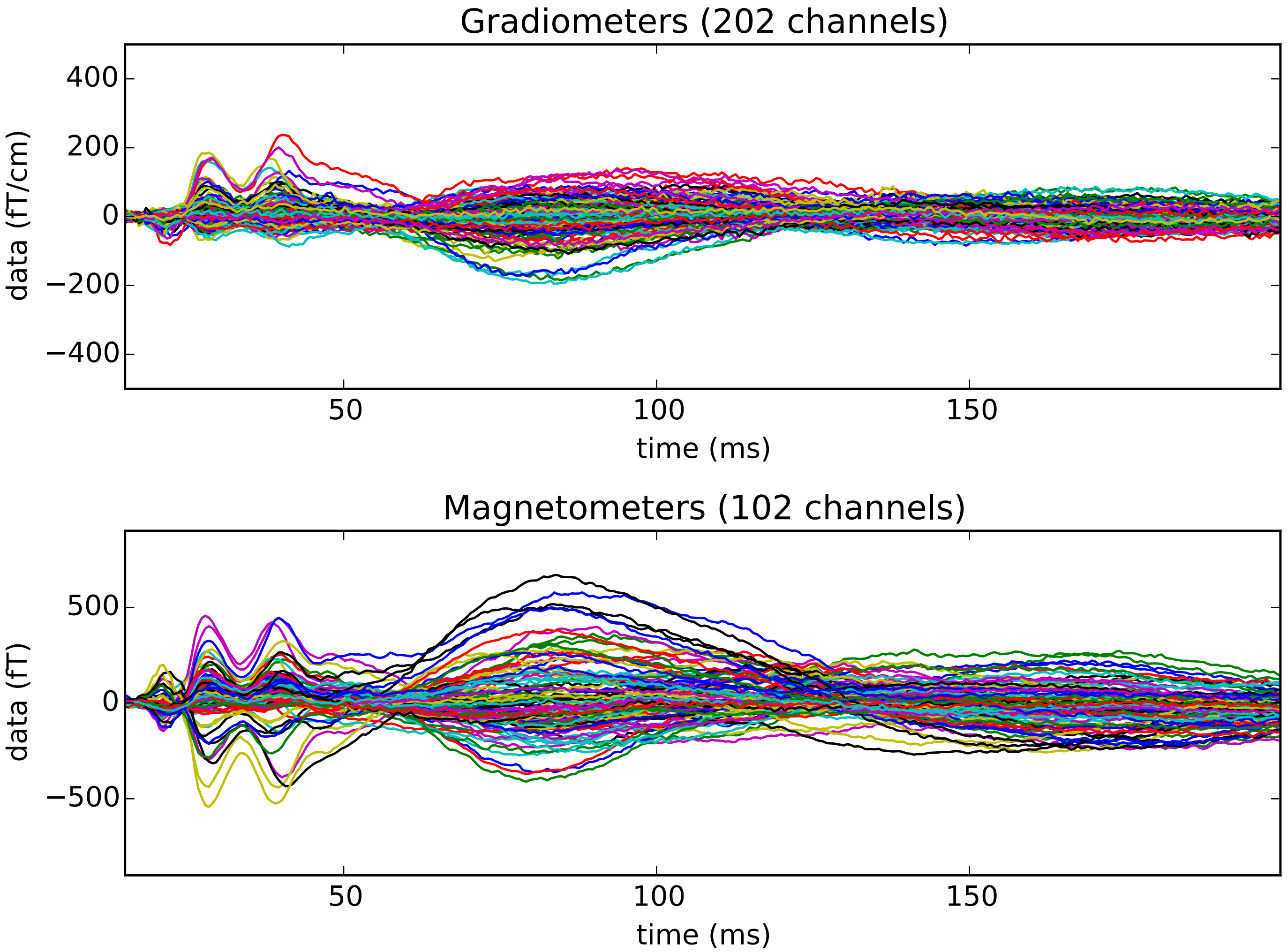


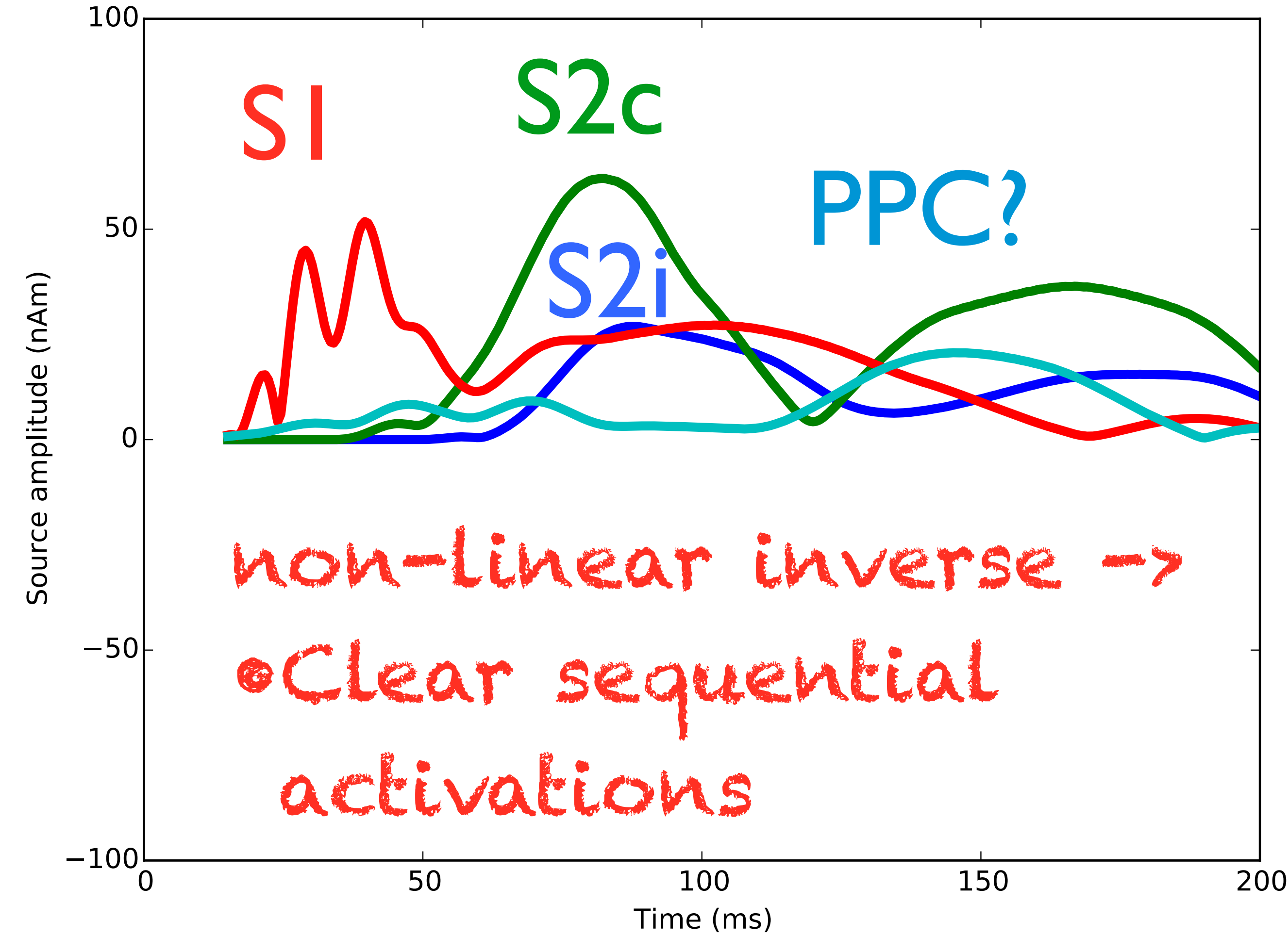


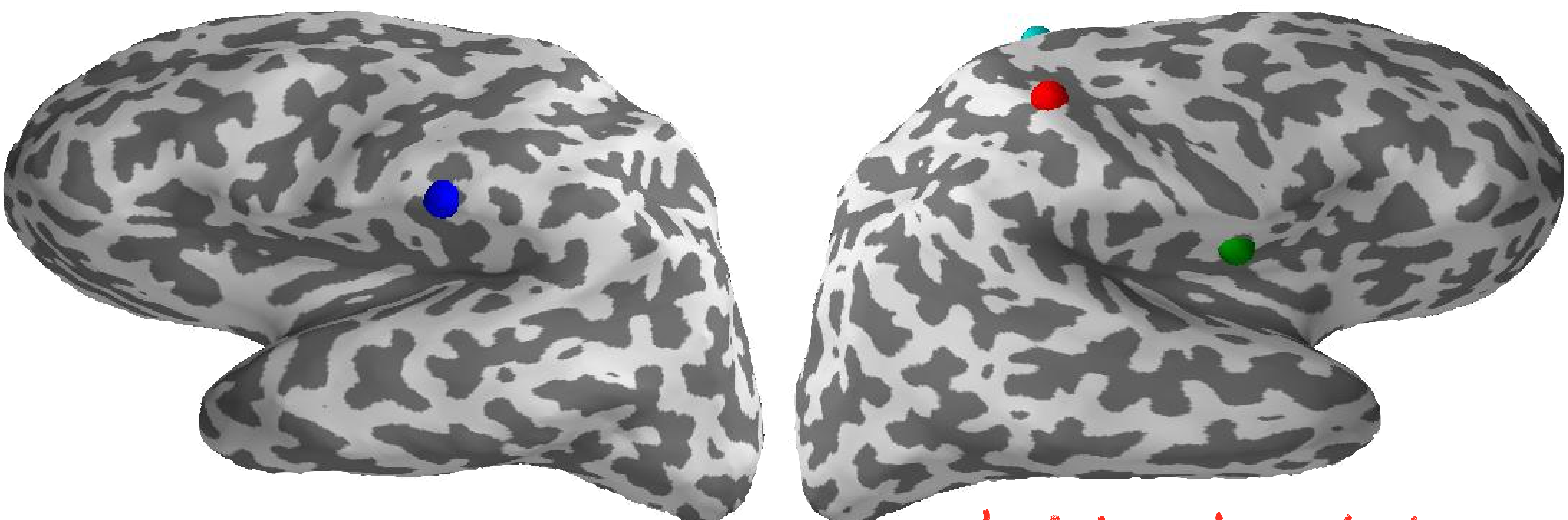
Linear inverse ->

- Imperfect deconvolution
- spatial leakage
- smeared activations
- no temporal smoothing
- but really fast...

$\phi(\mathbf{X})$ sparse / non-smooth







but harder / slower ....

Source localization under sparsity assumptions

Let's start with the Lasso

Warning: with the stats notations !

$$\hat{\beta}^{(\lambda)} \in \arg\min_{\beta \in \mathbb{R}^p} \Bigg(\underbrace{\frac{1}{2}\|y - X\beta\|_2^2}_{\text{data fitting term}} + \underbrace{\lambda\|\beta\|_1}_{\text{sparsity-inducing penalty}} \Bigg)$$

- Compute $\hat{\beta}^{(\lambda)}$ for **many** λ's: *e.g.*, T values from $\lambda_{\max} := \|X^\top y\|_\infty$ to $\epsilon\lambda_{\max}$ on log-scale ($T = 100, \epsilon = 0.001$)

Denoising case

Suppose the design is simple: $n = p$ and $X = \mathrm{Id_n}$, meaning the atoms are canonical elements: $\mathbf{x}_j = (0, \cdots, 0, \underset{\underset{j}{\uparrow}}{1}, 0, \cdots, 1)^\top$

$$\hat{\beta}^{(\lambda)} \in \arg\min_{\beta \in \mathbb{R}^p} \left(\frac{1}{2} \|y - \beta\|^2 + \lambda \|\beta\|_1 \right)$$

$$\hat{\beta}^{(\lambda)} = \arg\min_{\beta \in \mathbb{R}^p} \left(\frac{1}{2} \|y - \beta\|^2 + \lambda \|\beta\|_1 \right) \qquad \text{(strictly convex)}$$

$$\hat{\beta}_j^{(\lambda)} = \arg\min_{\beta_j \in \mathbb{R}} \left(\frac{1}{2} (y_i - \beta_j)^2 + \lambda |\beta_j| \right), \forall j \in [n] \qquad \text{(separable)}$$

This reduces to a 1D problem.

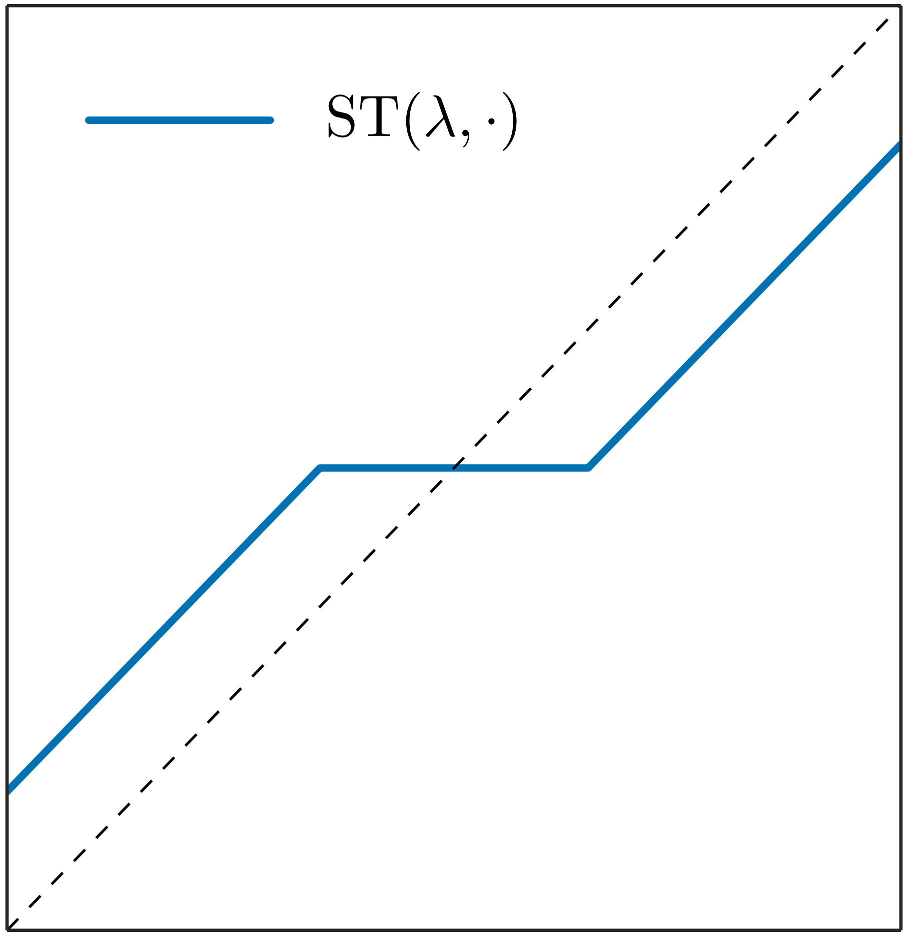
Rem: The solution is called the **proximal** operator of $\lambda \| \cdot \|_1$

Soft Thresholding

The 1D problem has a closed form solution: **Soft-Thresholding**:

$$\mathrm{ST}(\lambda, y) = \arg\min_{\beta \in \mathbb{R}} \left(\frac{1}{2}(y - \beta)^2 + \lambda|\beta| \right)$$
$$= \mathrm{sign}(y) \cdot (|y| - \lambda)_+$$

with the notation $(\cdot)_+ = \max(0, \cdot)$



Proof: easy with sub-gradients and Fermat condition

Soft Thresholding

Possible algorithms for solving this **convex** program:

- Homotopy method / LARS : very efficient for small p Osborne *et al.* (2000), Efron *et al.* (2004) and full path
- Forward - Backward / proximal algorithm: useful in signal/image for case where $r \rightarrow \mathbf{x}_j^\top r$ is cheap to compute (*e.g.*, with FFT, Fast Wavelet Transform, etc.) Beck and Teboulle (2009)
- Coordinate Descent: very useful for large p and potentially sparse matrix X (*e.g.*, from text encoding) Friedman *et al.* (2007)

Also better for badly conditioned problems

Dual problem

Primal function : $$P_\lambda(\beta) = \frac{1}{2}\|y - X\beta\|^2 + \lambda\|\beta\|_1$$

Dual feasible set : $$\Delta_X = \left\{\theta \in \mathbb{R}^n \ : \ |\mathbf{x}_j^\top \theta| \leqslant 1, \forall j \in [p]\right\}$$

Dual solution : $$\hat{\theta}^{(\lambda)} = \underset{\theta \in \Delta_X \subset \mathbb{R}^n}{\arg\max} \underbrace{\frac{1}{2}\|y\|^2 - \frac{\lambda^2}{2}\left\|\theta - \frac{y}{\lambda}\right\|^2}_{=D_\lambda(\theta)}$$

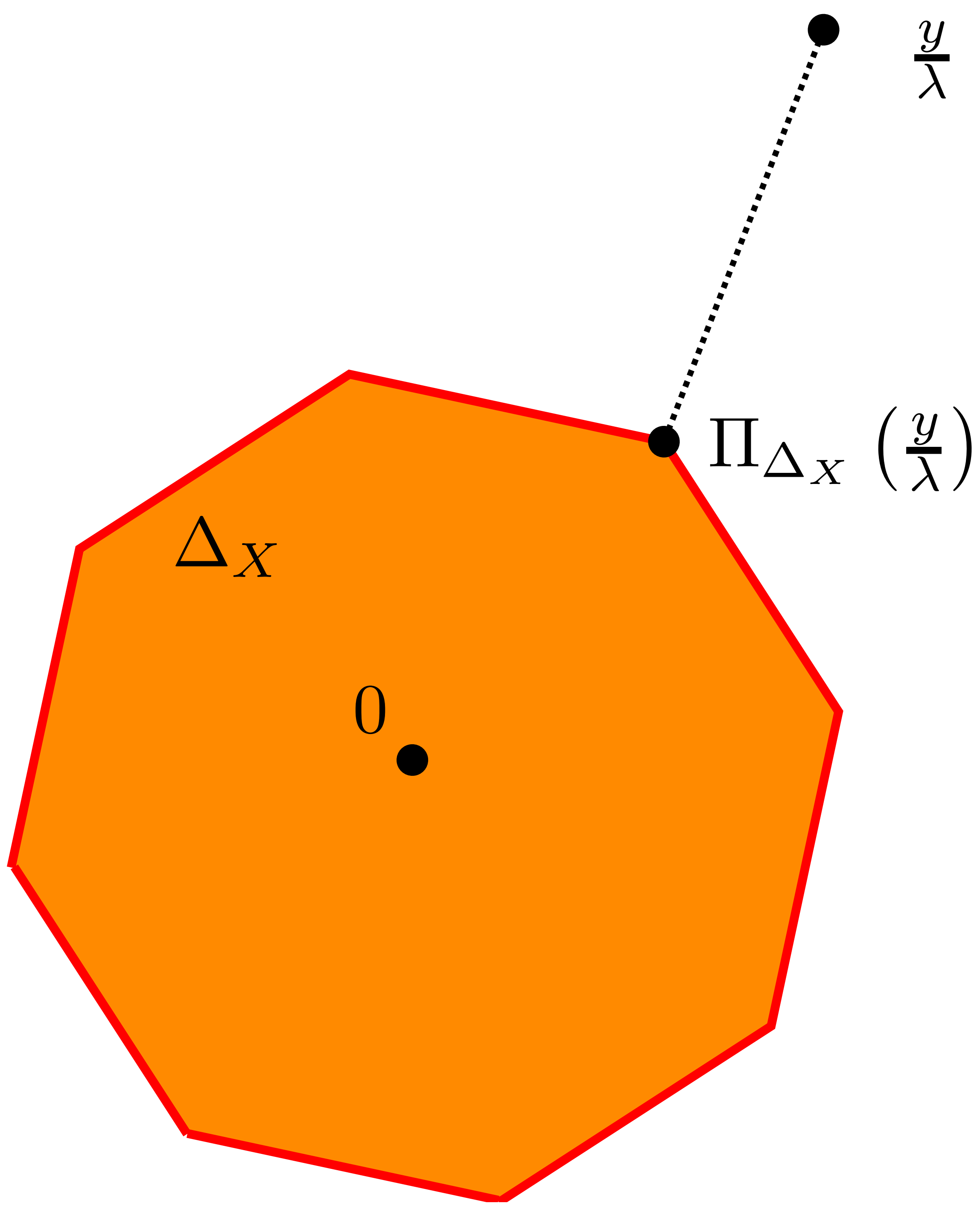
Rem: The dual feasible set is a polytope

$$\Delta_X = \bigcap_{j=1}^{p} \left\{\theta \in \mathbb{R}^n : |\mathbf{x}_j^\top \theta| \leqslant 1\right\} = \left\{\theta \in \mathbb{R}^n : \|X^\top \theta\|_\infty \leqslant 1\right\}$$

Rem: the dual formulation is obtained using an additional variable $z = (y - X\beta)/\lambda$ and considering the Lagrangian, *cf.* Kim *et al.* (2007)

Geometric interpretation

The dual optimal solution is the projection of y/λ over the dual feasible set $\Delta_X = \{\theta \in \mathbb{R}^n : \|X^\top \theta\|_\infty \leqslant 1\}$: $\hat{\theta}^{(\lambda)} = \Pi_{\Delta_X}(y/\lambda)$



Duality gap properties

- **Primal objective:** P_λ, **Primal solution:** $\hat{\beta}^{(\lambda)} \in \mathbb{R}^p$
- **Dual objective:** D_λ, **Primal solution:** $\hat{\theta}^{(\lambda)} \in \Delta_X \subset \mathbb{R}^n$,

Duality gap: for any $\beta \in \mathbb{R}^p$ and any $\theta \in \Delta_X$,

$$G_\lambda(\beta, \theta) = P_\lambda(\beta) - D_\lambda(\theta)$$
$$= \frac{1}{2}\|X\beta - y\|^2 + \lambda\|\beta\|_1 - \left(\frac{1}{2}\|y\|^2 - \frac{\lambda^2}{2}\left\|\theta - \frac{y}{\lambda}\right\|^2\right)$$

Rem: For all $\beta \in \mathbb{R}^p, \theta \in \Delta_X$,

$$D_\lambda(\theta) \leqslant D_\lambda(\hat{\theta}^{(\lambda)}) = P_\lambda(\hat{\beta}^{(\lambda)}) \leqslant P_\lambda(\beta) \quad \textbf{(Strong duality)}$$

Consequences:

- $G_\lambda(\beta, \theta) \geqslant 0$
- $G_\lambda(\beta, \theta) \leqslant \epsilon \implies P_\lambda(\beta) - P_\lambda(\hat{\beta}^{(\lambda)}) \leqslant \epsilon$ (stopping criterion!)

KKT Optimality conditions

- **Primal solution :** $\hat{\beta}^{(\lambda)} \in \mathbb{R}^p$
- **Dual solution :** $\hat{\theta}^{(\lambda)} \in \Delta_X \subset \mathbb{R}^n$

Primal/Dual link: $\boxed{y = X\hat{\beta}^{(\lambda)} + \lambda\hat{\theta}^{(\lambda)}}$

Necessary and sufficient optimality conditions:

KKT/Fermat: $$\boxed{\forall j \in [p],\ x_j^\top \hat{\theta}^{(\lambda)} \in \begin{cases} \{\mathrm{sign}(\hat{\beta}_j^{(\lambda)})\} & \text{if} \quad \hat{\beta}_j^{(\lambda)} \neq 0, \\ [-1, 1] & \text{if} \quad \hat{\beta}_j^{(\lambda)} = 0. \end{cases}}$$

Rem: the KKT implies that $\forall \lambda \geqslant \lambda_{\max} = \|X^\top y\|_\infty$, $0 \in \mathbb{R}^p$ is the (unique here) primal solution for P_λ

Safe rules [El Ghaoui et al. 2012]

Screening thanks to the KKT is possible:

$$\boxed{\text{If } |\mathbf{x}_j^\top \hat{\theta}^{(\lambda)}| < 1 \text{ then, } \hat{\beta}_j^{(\lambda)} = 0}$$

Beware: $\hat{\theta}^{(\lambda)}$ is unknown, so one need to consider a **safe region** $\mathcal{C}$ containing $\hat{\theta}^{(\lambda)}$, *i.e.*, $\hat{\theta}^{(\lambda)} \in \mathcal{C}$, leading to :

safe rule :
$$\boxed{\text{If } \sup_{\theta \in \mathcal{C}} |\mathbf{x}_j^\top \theta| < 1 \text{ then } \hat{\beta}_j^{(\lambda)} = 0} \qquad (\star)$$

The new goal is simple, find a region $\mathcal{C}$:

- as narrow as possible containing $\hat{\theta}^{(\lambda)}$
- such that $\mu_{\mathcal{C}} : \begin{cases} \mathbb{R}^n & \mapsto \mathbb{R}^+ \\ \mathbf{x} & \to \sup_{\theta \in \mathcal{C}} |\mathbf{x}^\top \theta| \end{cases}$ is easy to compute

Safe sphere rules

Let $\mathcal{C} = B(c, r)$ be a ball of center $c \in \mathbb{R}^n$ and radius $r > 0$. Then simple computation provide:

$$\mu_{\mathcal{C}}(\mathbf{x}) = |\mathbf{x}^\top c| + r\|\mathbf{x}\|$$

so the safe rule becomes

$$\boxed{\text{If } |\mathbf{x}_j^\top c| + r\|\mathbf{x}_j\| < 1 \text{ then } \hat{\beta}_j^{(\lambda)} = 0} \tag{1}$$

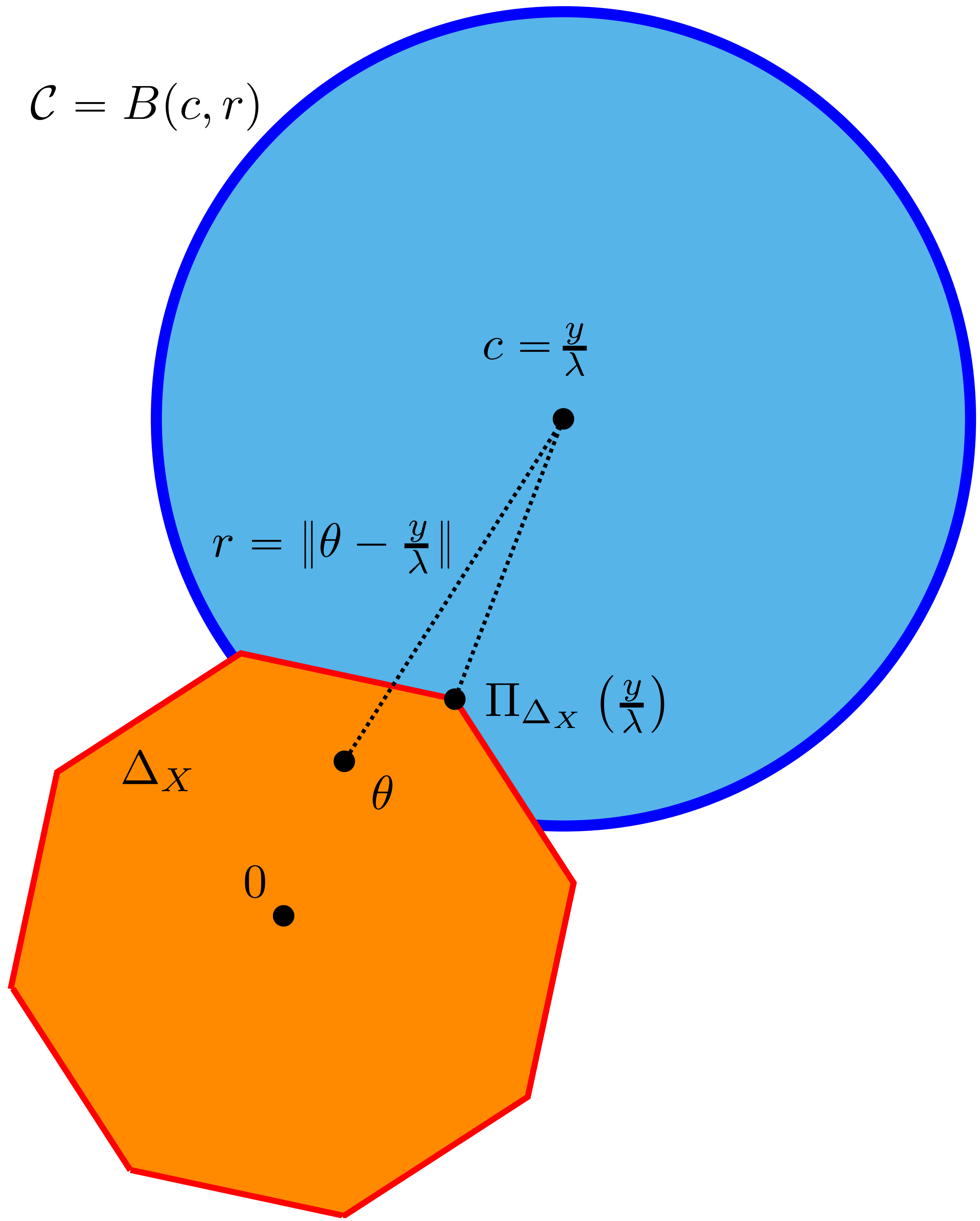
We say we screen-out the variables $\mathbf{x}_j$ satisfying (1)

Active set : $A^{(\lambda)}(\mathcal{C}) = \{j \in [p] : \mu_{\mathcal{C}}(\mathbf{x}_j) \geqslant 1\}$

New objective:

- find r as small as possible
- find c as close to $\hat{\theta}^{(\lambda)}$ as possible.

Creating safe sphere



Gap safe sphere

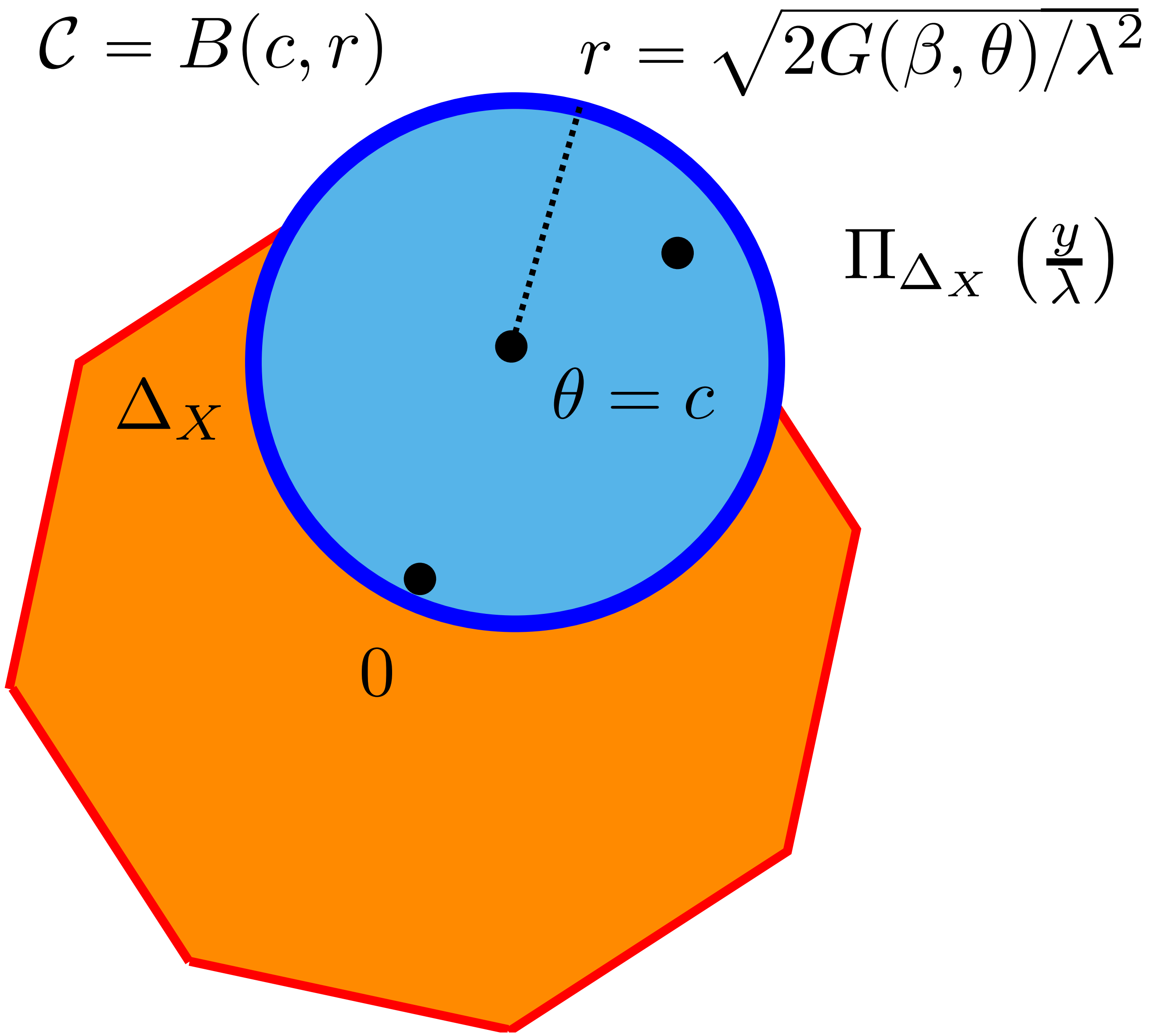
For any $\beta \in \mathbb{R}^p, \theta \in \Delta_X$

$$G_\lambda(\beta, \theta) = \frac{1}{2}\|X\beta - y\|^2 + \lambda\|\beta\|_1 - \left(\frac{1}{2}\|y\|^2 - \frac{\lambda^2}{2}\left\|\theta - \frac{y}{\lambda}\right\|^2\right)$$

Gap Safe ball: $\boxed{B(\theta, r_\lambda(\beta, \theta)), \text{ where } r_\lambda(\beta, \theta) = \sqrt{2G_\lambda(\beta, \theta)/\lambda^2}}$

Rem: If $\beta_k \to \hat{\beta}^{(\lambda)}$ and $\theta_k \to \hat{\theta}^{(\lambda)}$ then $G_\lambda(\beta_k, \theta_k) \to 0$: a converging solver leads to converging safe rule!

Gap safe sphere is safe !



Algorithm 1 Coordinate descent (Lasso)

Input: $X, y, \epsilon, K, f, (\lambda_t)_{t \in [T-1]}$

Initialization: $\lambda_0 = \lambda_{\max}, \quad \beta^{\lambda_0} = 0$

for $t \in [T-1]$ **do** ▷ Loop over λ's

    $\beta \leftarrow \beta^{\lambda_{t-1}}$ ▷ previous ϵ-solution

    for $k \in [K]$ **do**

        if $k \mod f = 1$ **then**

            Construct $\theta \in \Delta_X$

            if $G_{\lambda_t}(\beta, \theta) \leqslant \epsilon$ **then** ▷ Stop if duality gap small

                $\beta^{\lambda_t} \leftarrow \beta$

                break

            end if

        end if

        for $j \in [p]$ **do** ▷ Soft-Threshold coordinates

            $\beta_j \leftarrow \mathrm{ST}\left(\frac{\lambda_t}{\|\mathbf{x}_j\|^2}, \beta_j - \frac{\mathbf{x}_j^\top (X\beta - y)}{\|\mathbf{x}_j\|^2}\right)$

        end for

    end for

end for

Algorithm 2 Coordinate descent (Lasso) with GAP Safe screening

Input: $X, y, \epsilon, K, f, (\lambda_t)_{t \in [T-1]}$

Initialization: $\lambda_0 = \lambda_{\max}, \quad \beta^{\lambda_0} = 0$

for $t \in [T-1]$ **do** ▷ Loop over λ's

  $\beta \leftarrow \beta^{\lambda_{t-1}}$ ▷ previous ϵ-solution

  for $k \in [K]$ **do**

    if $k \mod f = 1$ **then**

      Construct $\theta \in \Delta_X$, $A^{\lambda_t}(\mathcal{C}) = \{j \in [p] : \mu_{\mathcal{C}}(\mathbf{x}_j) \geqslant 1\}$

      if $G_{\lambda_t}(\beta, \theta) \leqslant \epsilon$ **then** ▷ Stop if duality gap small

        $\beta^{\lambda_t} \leftarrow \beta$

        break

      end if

    end if

    for $j \in A^{\lambda_t}(\mathcal{C})$ **do** ▷ Soft-Threshold coordinates

      $\beta_j \leftarrow \mathrm{ST}\left(\frac{\lambda_t}{\|\mathbf{x}_j\|^2}, \beta_j - \frac{\mathbf{x}_j^\top (X\beta - y)}{\|\mathbf{x}_j\|^2}\right)$

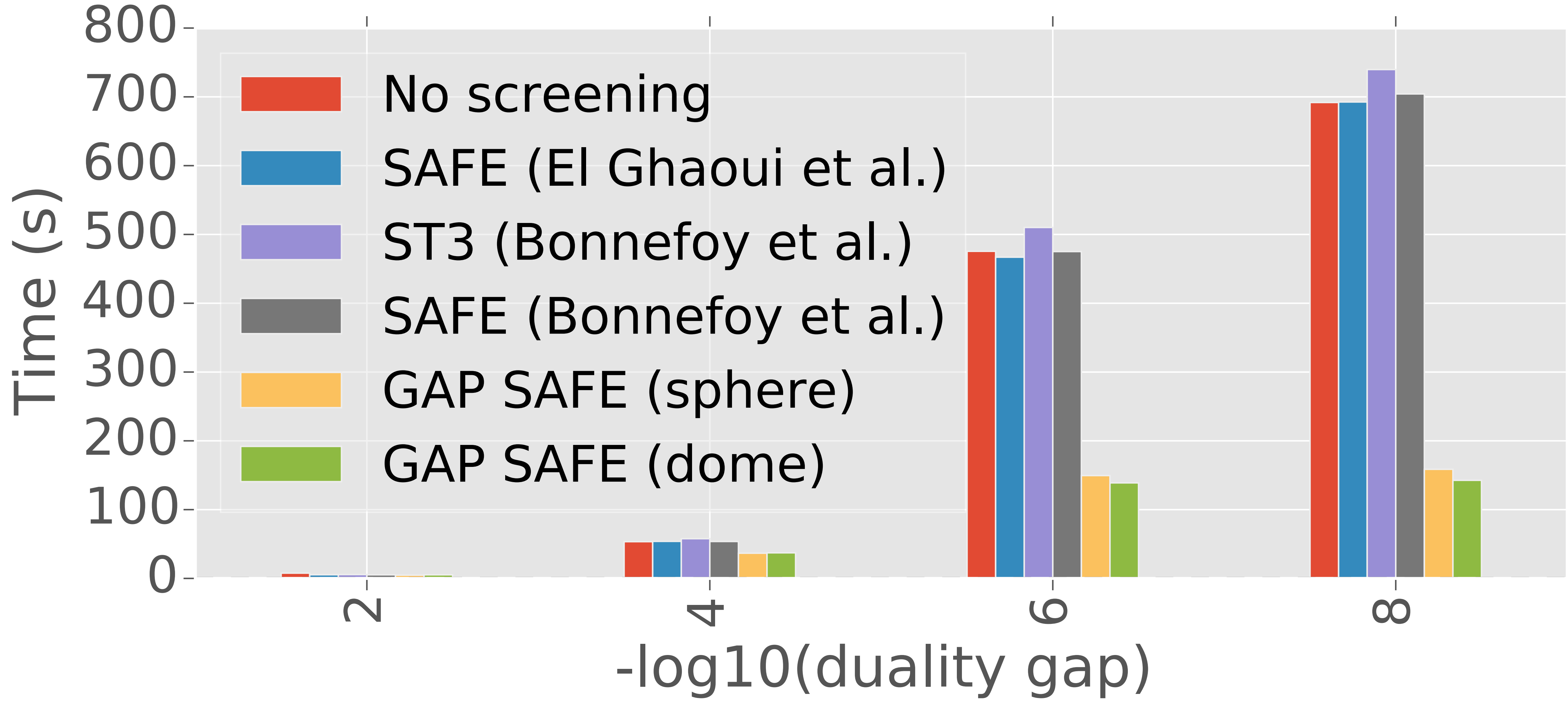
    end for

  end for

end for

Results

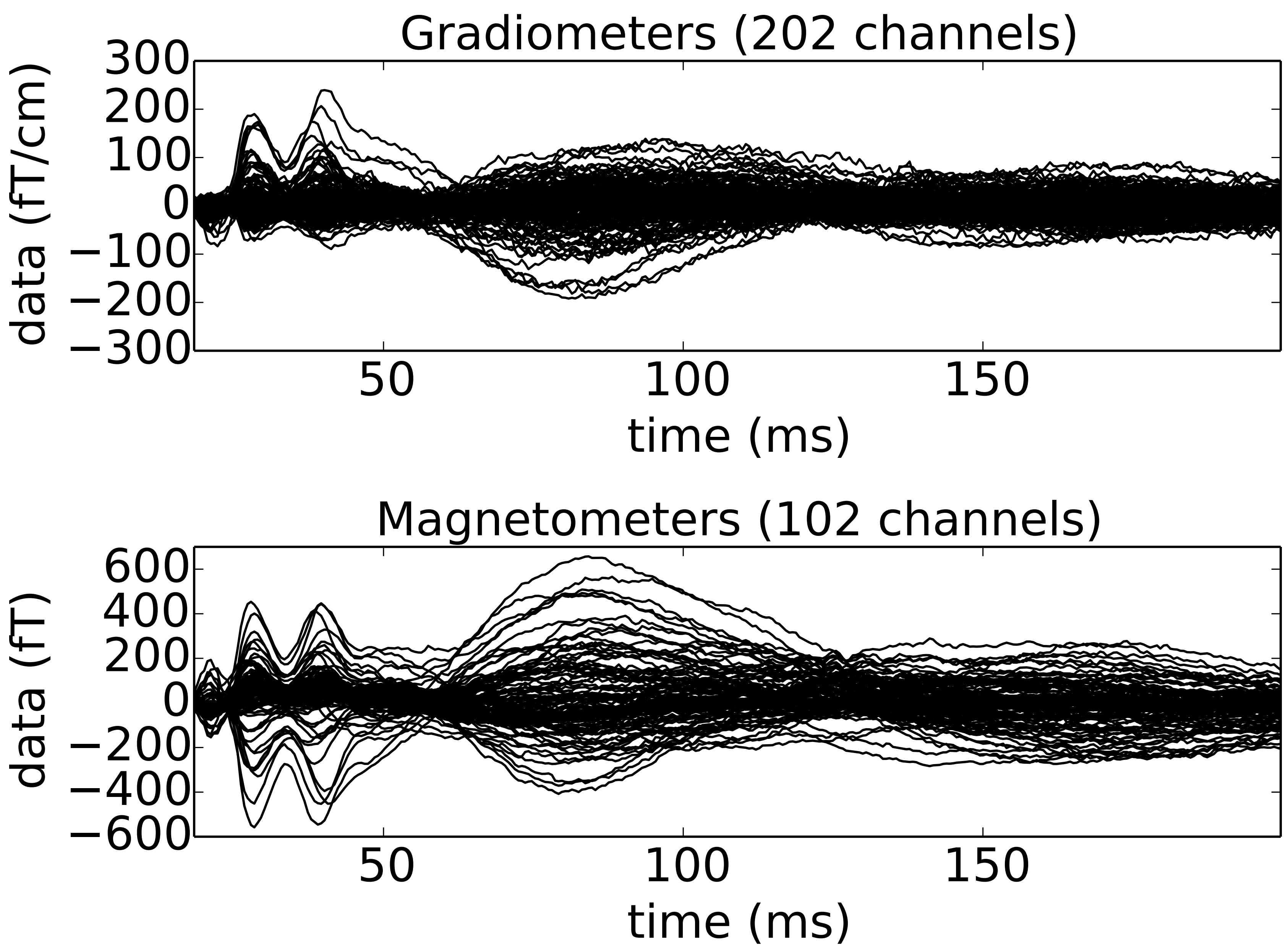
Results



Time to reach convergence using various screening rules.
Full path with 100 values of λ on logarithmic grid from λ_{max} to $\lambda_{max}/1000$

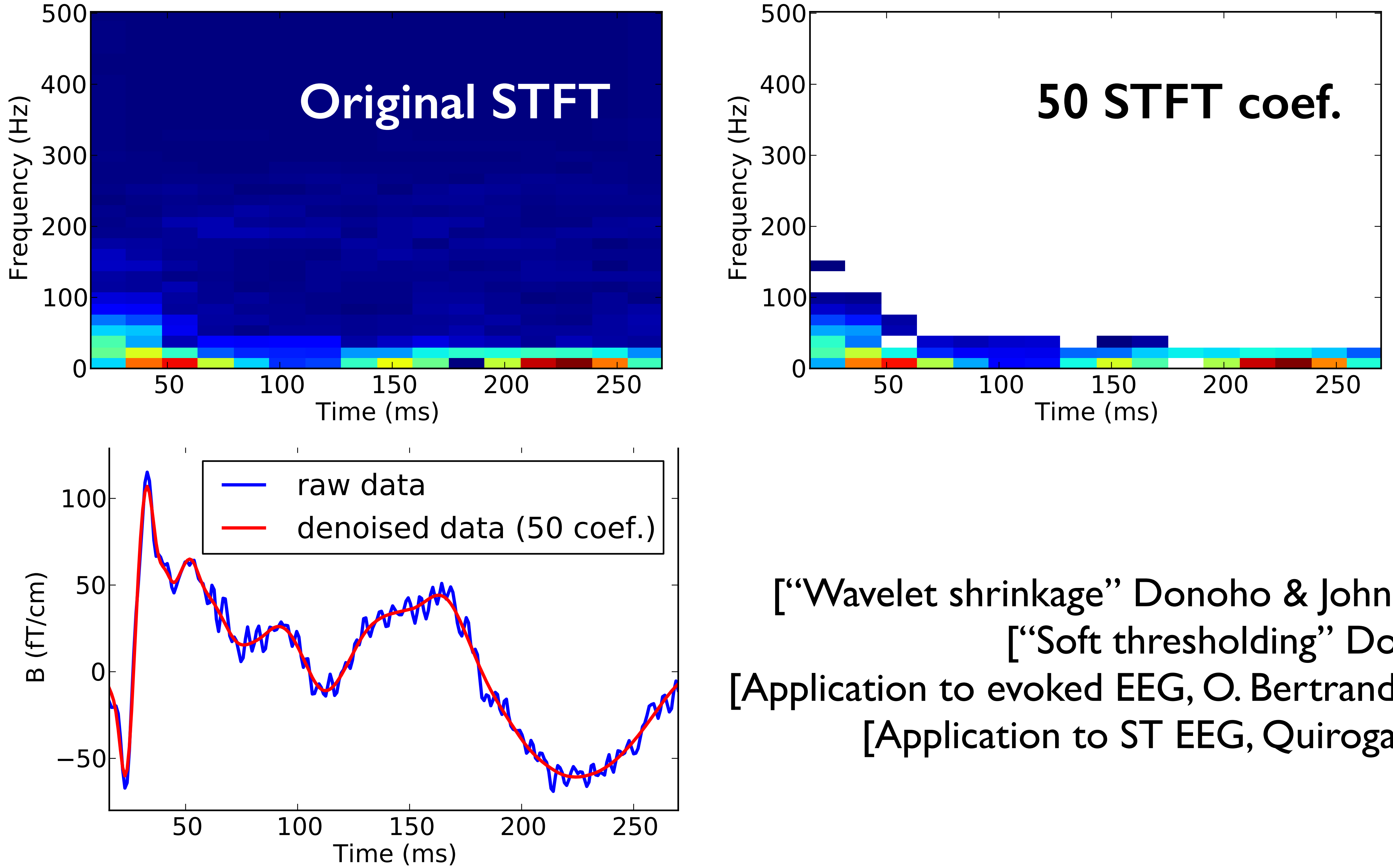
[Fercoq et al., *Mind the duality gap: Safer screeing rules for the Lasso*, ICML 2015]

Beyond Lasso with time



Challenge: How do you promote sparse solutions with non-stationary sources?

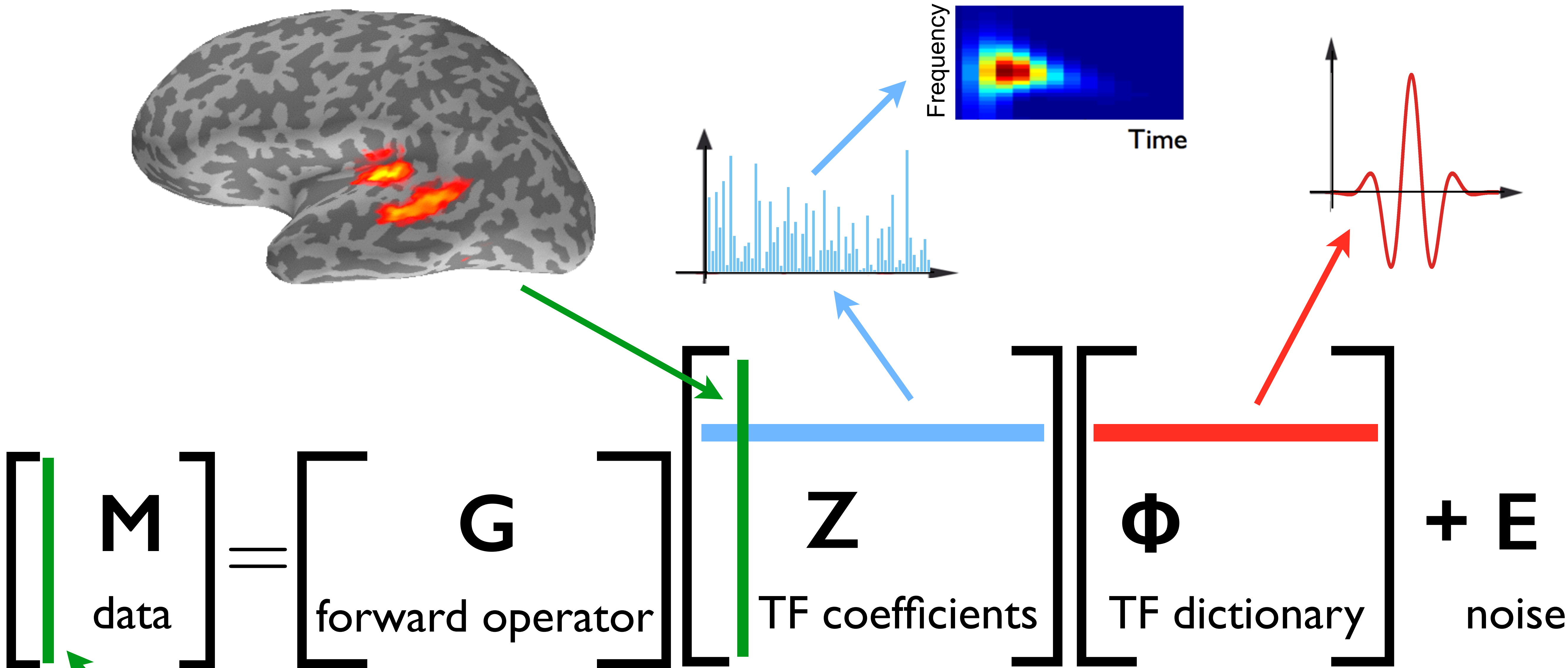
Change the represensation

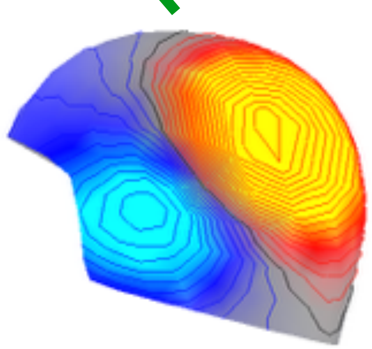


[“Wavelet shrinkage” Donoho & Johnstone 94]
[“Soft thresholding” Donoho 95]
[Application to evoked EEG, O. Bertrand et al. 94]
[Application to ST EEG, Quiroga et al. 03]
etc.

[Moussallam, Gramfort, Richard, Daudet, Signal Processing Letters 2014]

$M = GZ\Phi + E$





Objective: estimate Z given M

[Gramfort et al., *Time-Frequency Mixed-Norm Estimates: Sparse M/EEG imaging with non-stationary source activations*, Neuroimage 2013]

Time-frequency (TF) regularization

The classical approach [MNE, dSPM, sLORETA]:

$$\hat{\mathbf{X}} = \arg\min_{\mathbf{X}} \underbrace{\|\mathbf{M} - \mathbf{G}\mathbf{X}\|_F^2}_{\text{data fit}} + \underbrace{\lambda\phi(\mathbf{X})}_{\text{regularization}}, \;\; \lambda > 0$$

we propose:

$$\hat{\mathbf{Z}} = \arg\min_{\mathbf{Z}} \|\mathbf{M} - \mathbf{G}\mathbf{Z}\mathbf{\Phi}^{\mathcal{H}}\|_F^2 + \lambda\phi(\mathbf{Z}), \;\; \text{then } \hat{\mathbf{X}} = \hat{\mathbf{Z}}\mathbf{\Phi}^{\mathcal{H}}$$

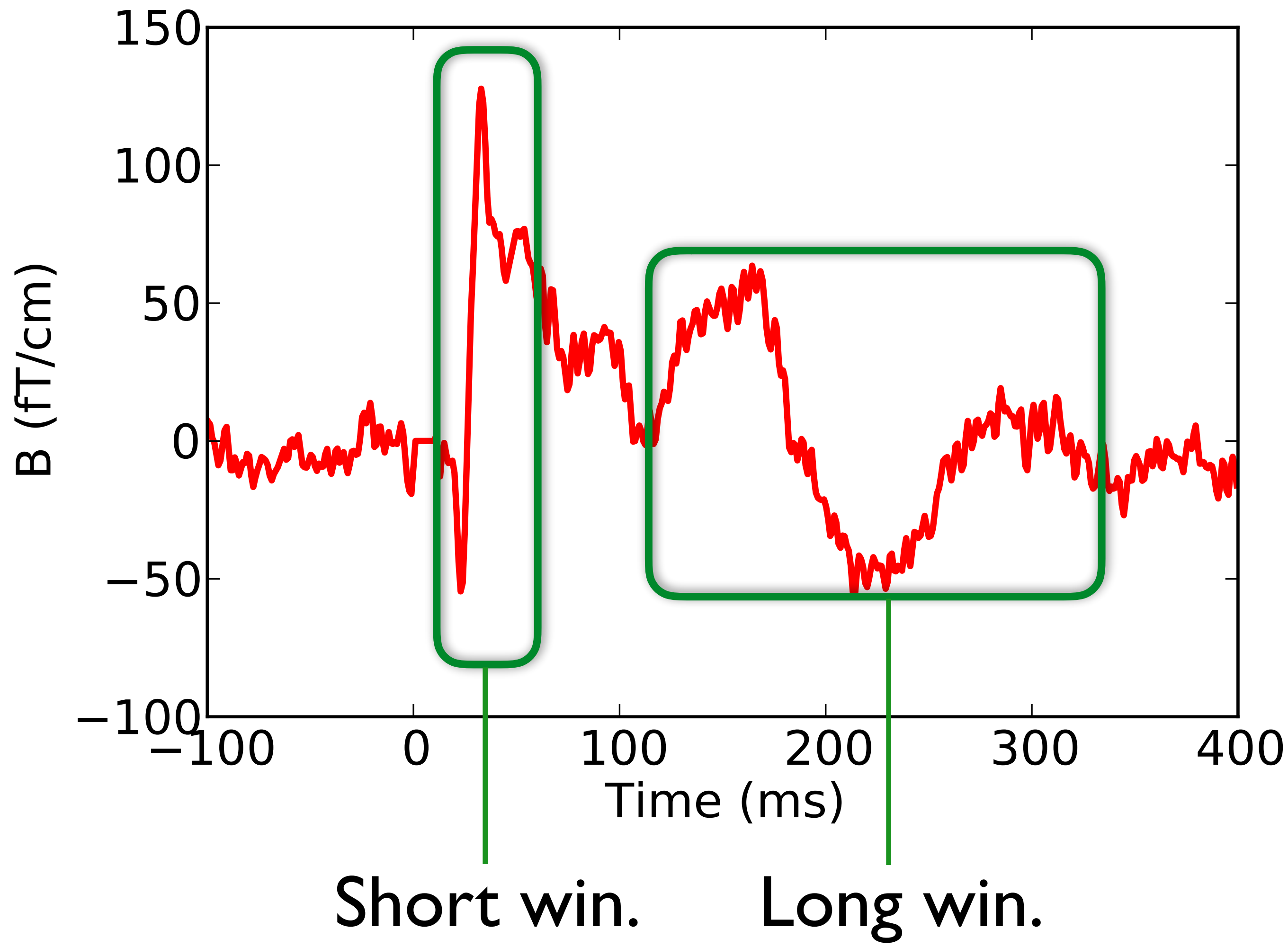
- $\mathbf{\Phi}$: is a **TF dictionary**
- $\mathbf{Z}$: **coefficients** of the **TF transform** of the sources

localization in space, time
and frequency in one step

Multi-scale dictionary

$$\hat{\mathbf{Z}} = \arg\min_{\mathbf{Z}} \|\mathbf{M} - \mathbf{G}\mathbf{Z}\mathbf{\Phi}^{\mathcal{H}}\|_F^2 + \lambda\phi(\mathbf{Z}), \text{ then } \hat{\mathbf{X}} = \hat{\mathbf{Z}}\mathbf{\Phi}^{\mathcal{H}}$$

- $\mathbf{\Phi}$: union of n **STFT dict. with diff. window lengths**
- $\mathbf{Z}$: is the combination of **coefficients** of the diff. **TF transforms** of the sources

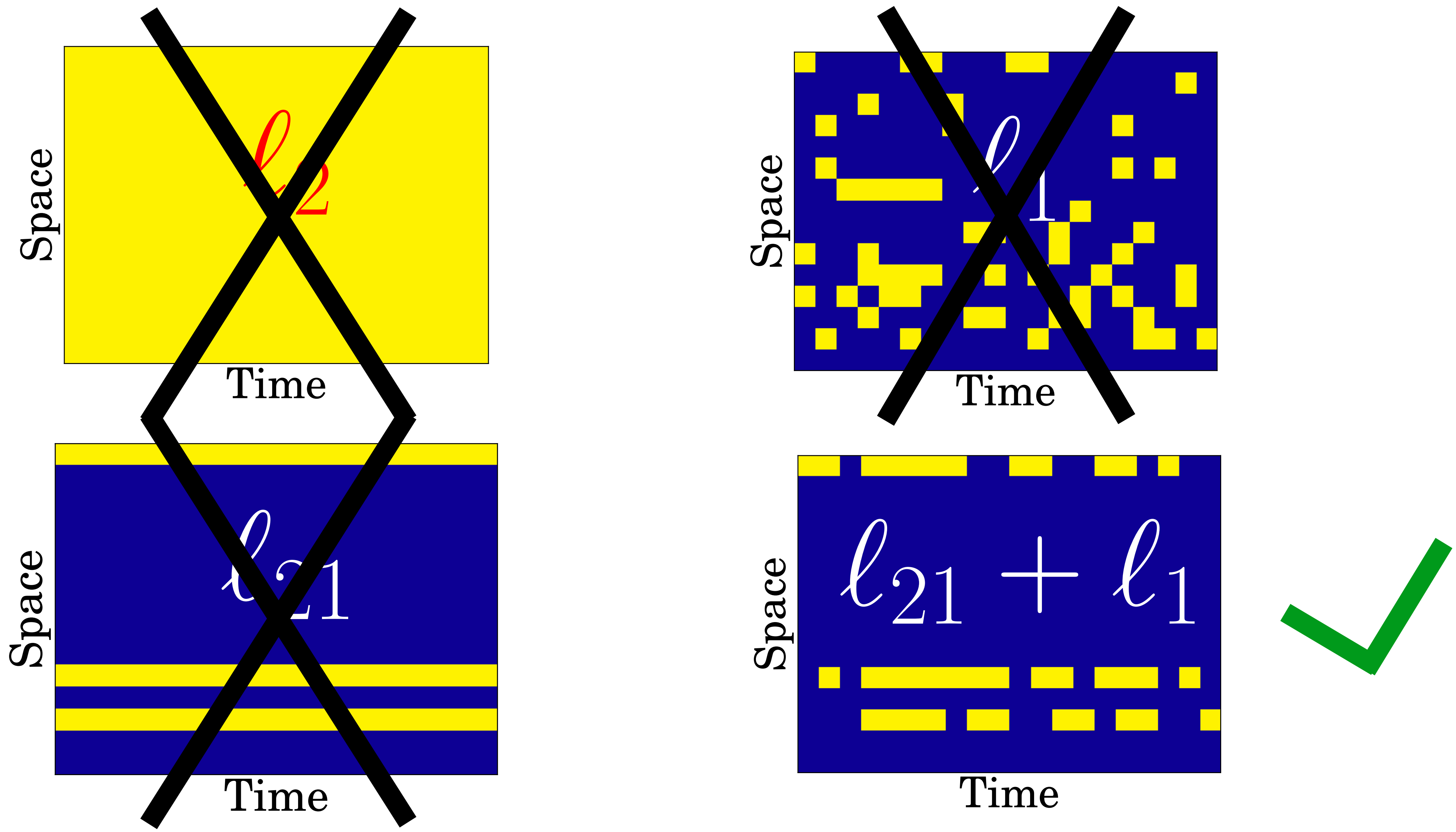


[Bekhti et al. 2016]

cf. [Kowalski et al. 2008]

cf. [Starck et al. 2005]

What regularization?



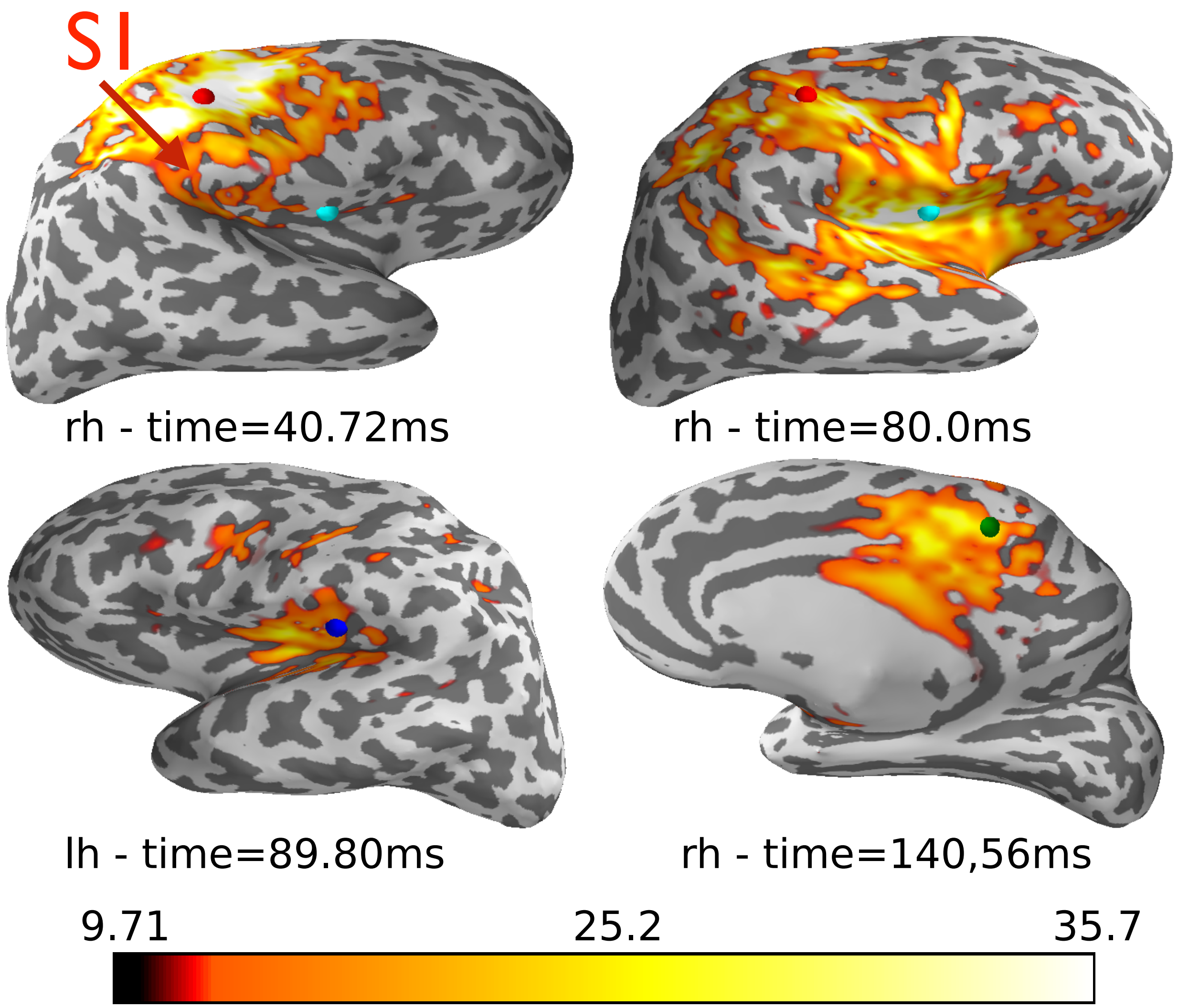
$$\phi(Z) = \lambda(\rho\|Z\|_1 + (1-\rho)\|Z\|_{21})$$

$$\|\mathbf{X}\|_{21} = \sum_i \sqrt{\sum_t |x_{i,t}|^2}$$

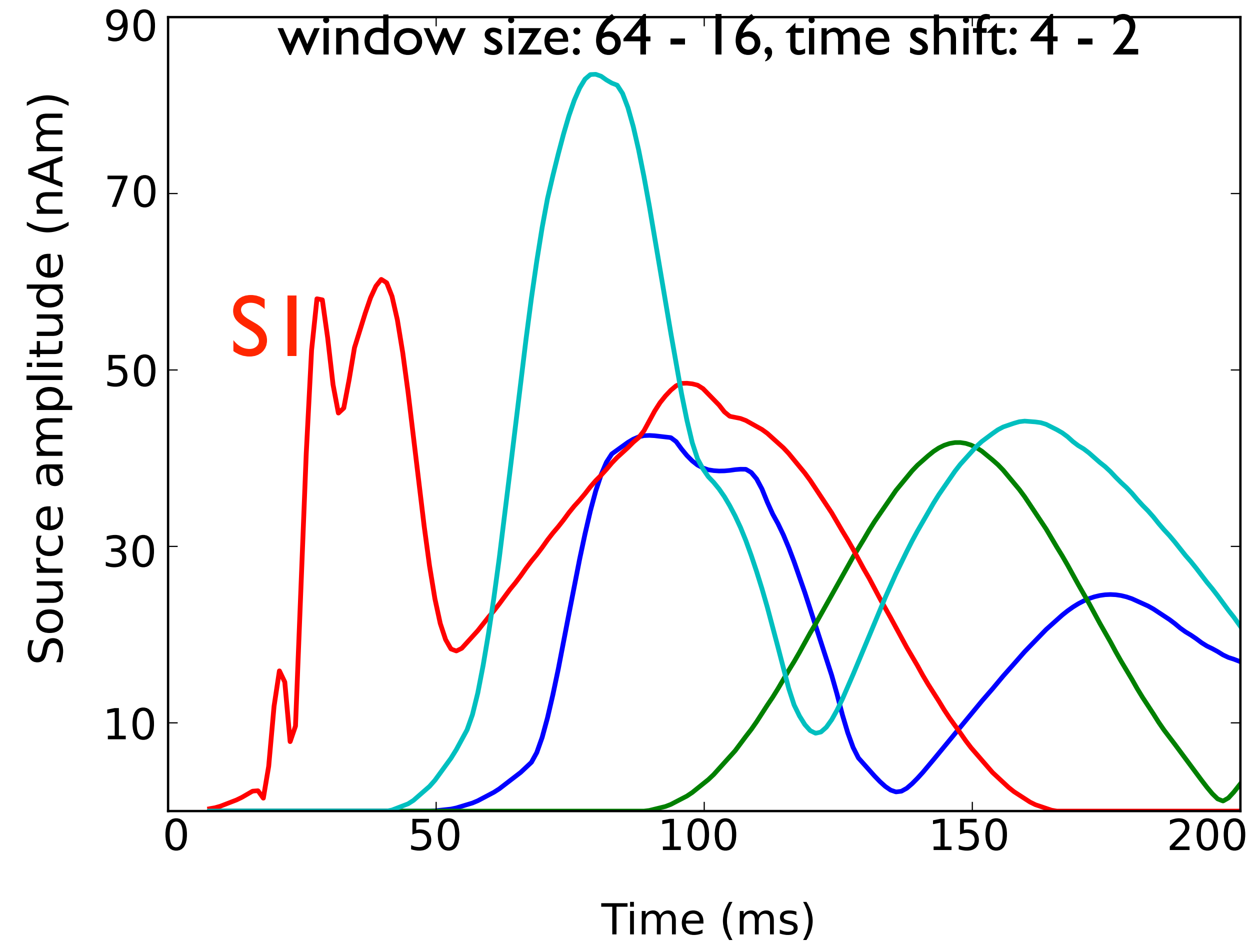
Gap safe screening for this model in [Ndiaye Fercoq Gramfort Salmon NIPS 2016]

Results

Somatosensory - MIND dataset



dSPM



- Clear sequential activations
- No spatial leakage

[Bekhti et al. PRNI 2016]

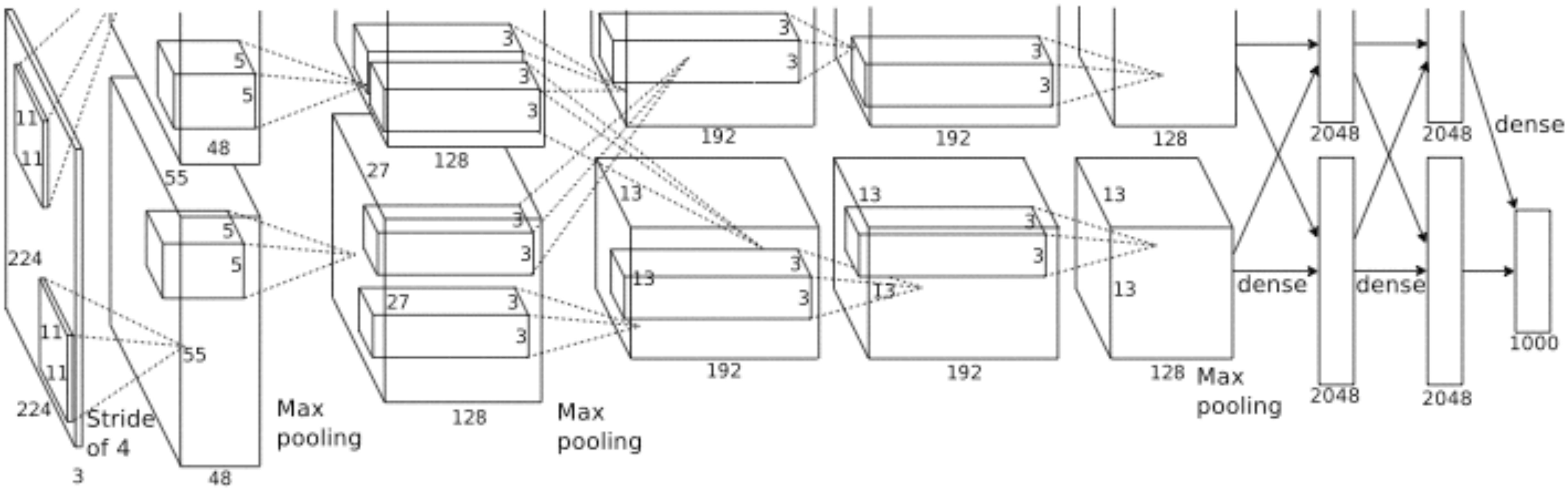
Convolutional Networks Map the Architecture of the Human Visual System

work of Michael Eickenberg

joint work with Bertrand Thirion and Gaël Varoquaux

"Seeing it all: Convolutional network layers map the function of the human visual system"
Michael Eickenberg, Alexandre Gramfort, Gaël Varoquaux, Bertrand Thirion, Neuroimage (to appear)

Convolutional Nets for Computer Vision

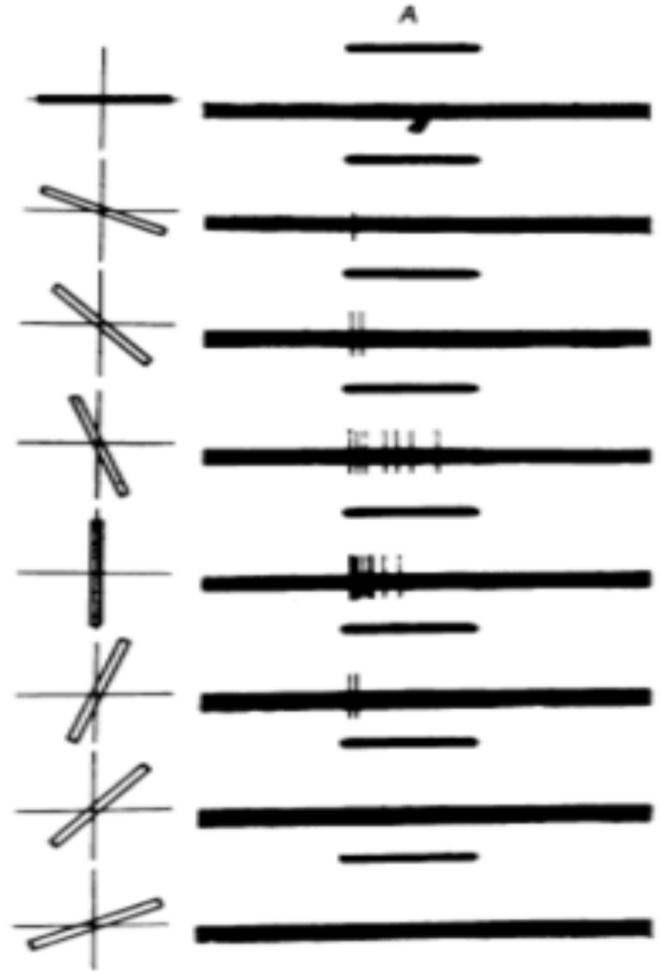


[Krizhevski et al, 2012]

Relating biological and computer vision

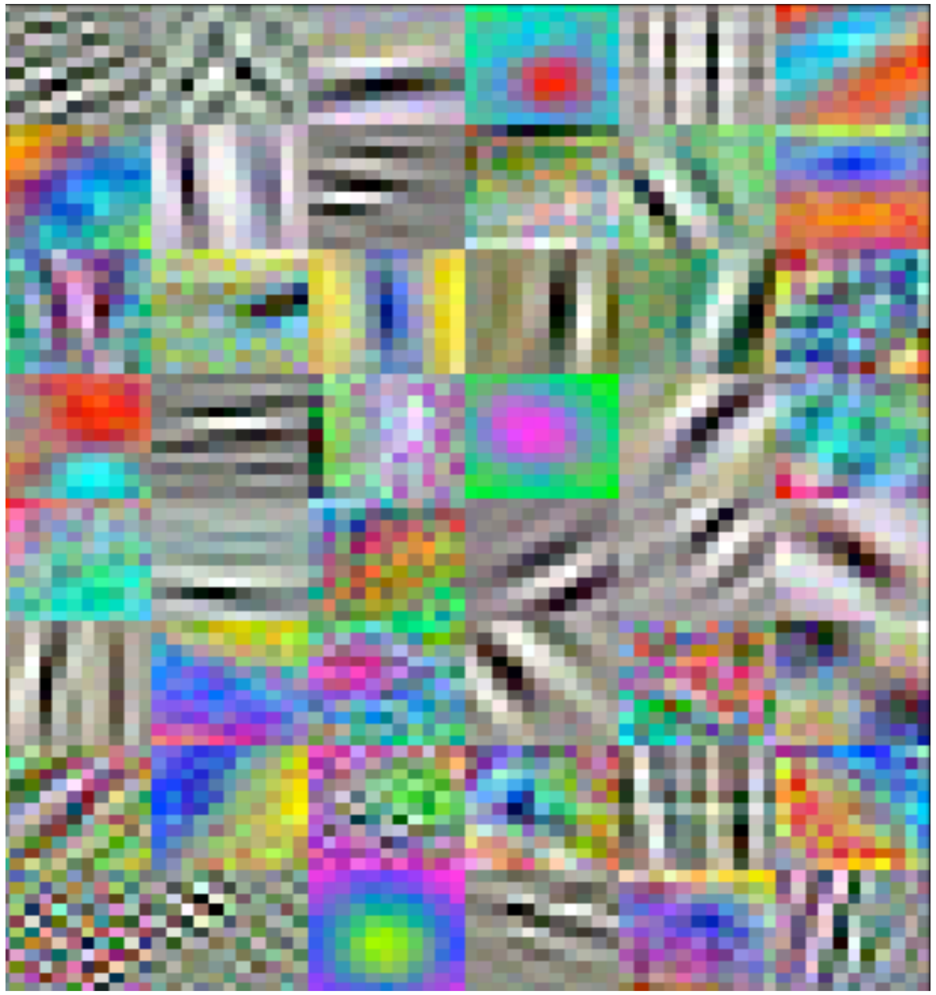
Low Level

Cat V1
orientation selectivity



[Hubel & Wiesel, 1959]

ConvNet Layer 1

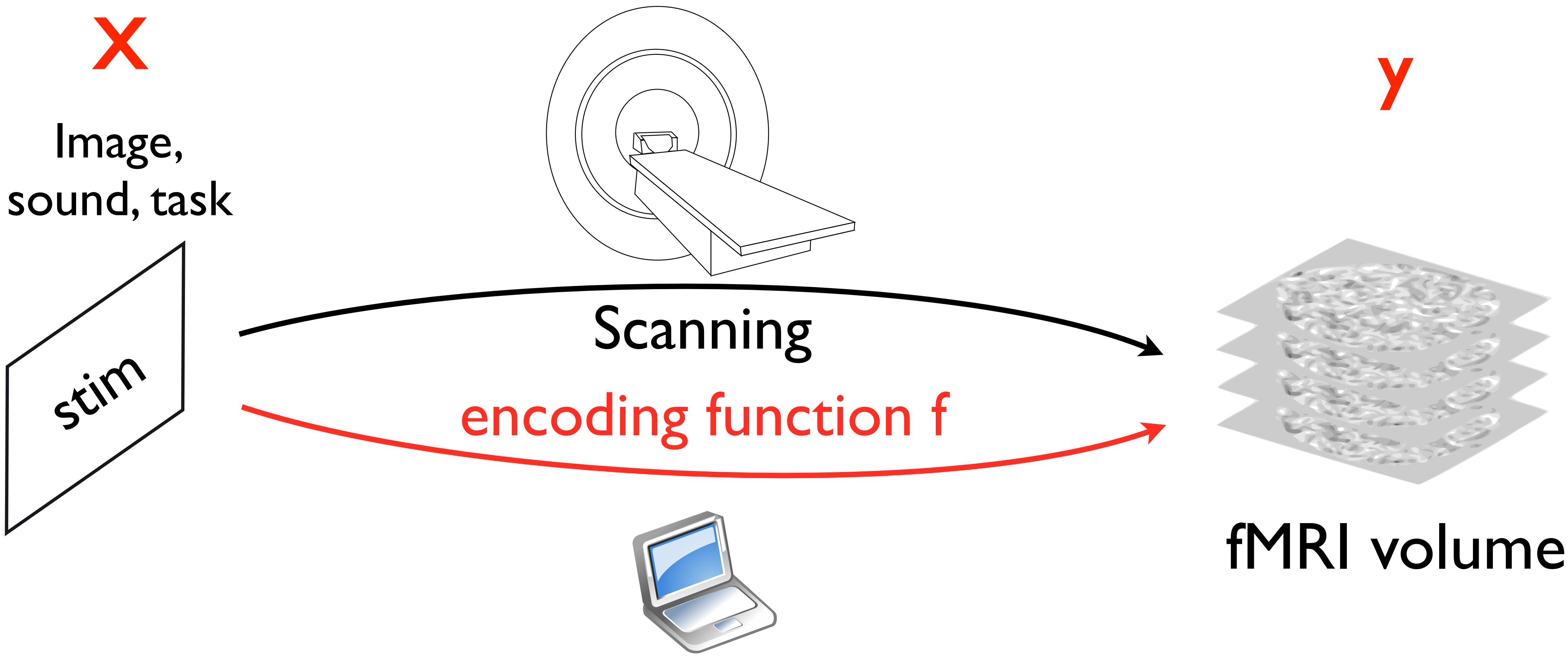


[Sermanet 2013]

- V1 functionality comprises edge detection
- Convolutional nets learn edge detectors, color boundary detectors and blob detectors

Can we use computer vision models and a large fMRI data to better understand human vision?

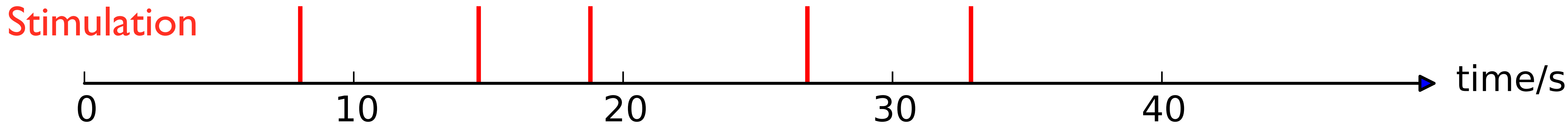
Learning the fMRI encoding function



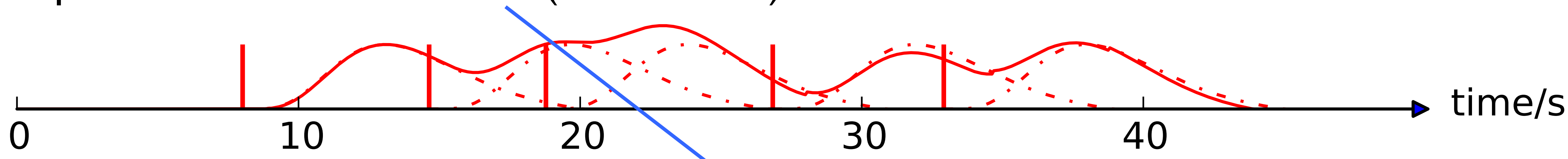
Challenge: Predict y given X or learn a function $f : X \rightarrow y$

[Thirion et al. 06, Kay et al. 08, Naselaris et al. 11, Nishimoto et al. 2011, Huth 2012 et al., Schoenmakers et al. 13, Güçlü et al. 2015, Cichy, et al. 2016, Huth et al. 2016 ...]

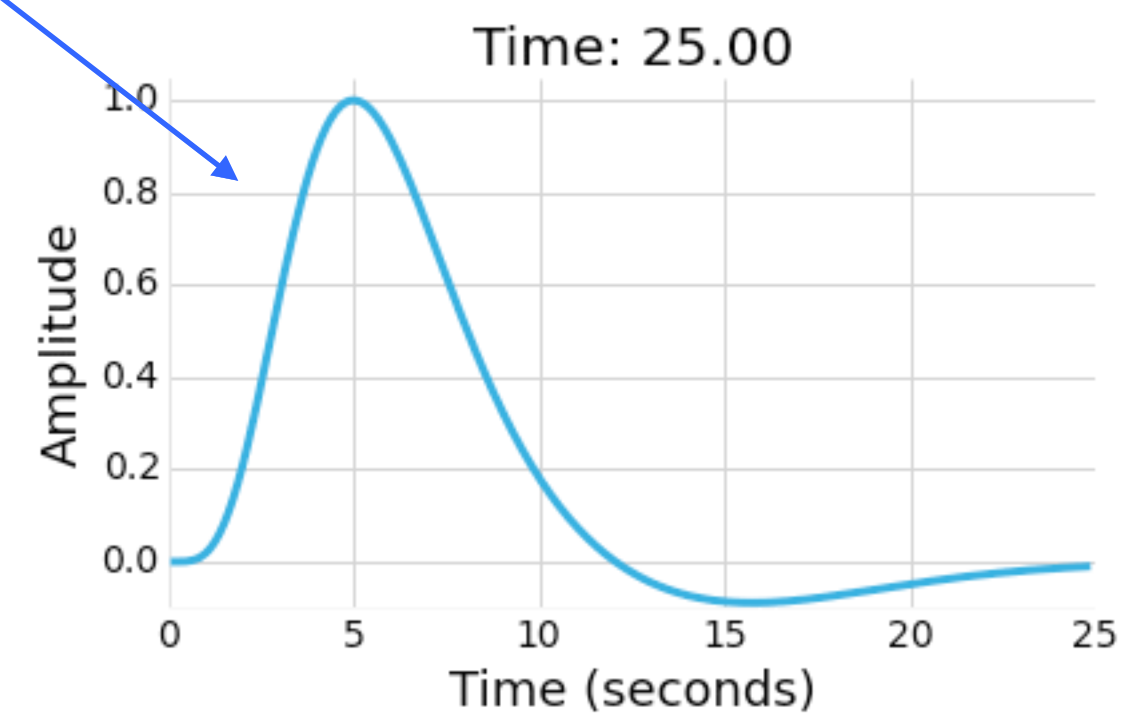
fMRI paradigm and HRF



Model pred. = stimulation ⊛ filter (convolution)



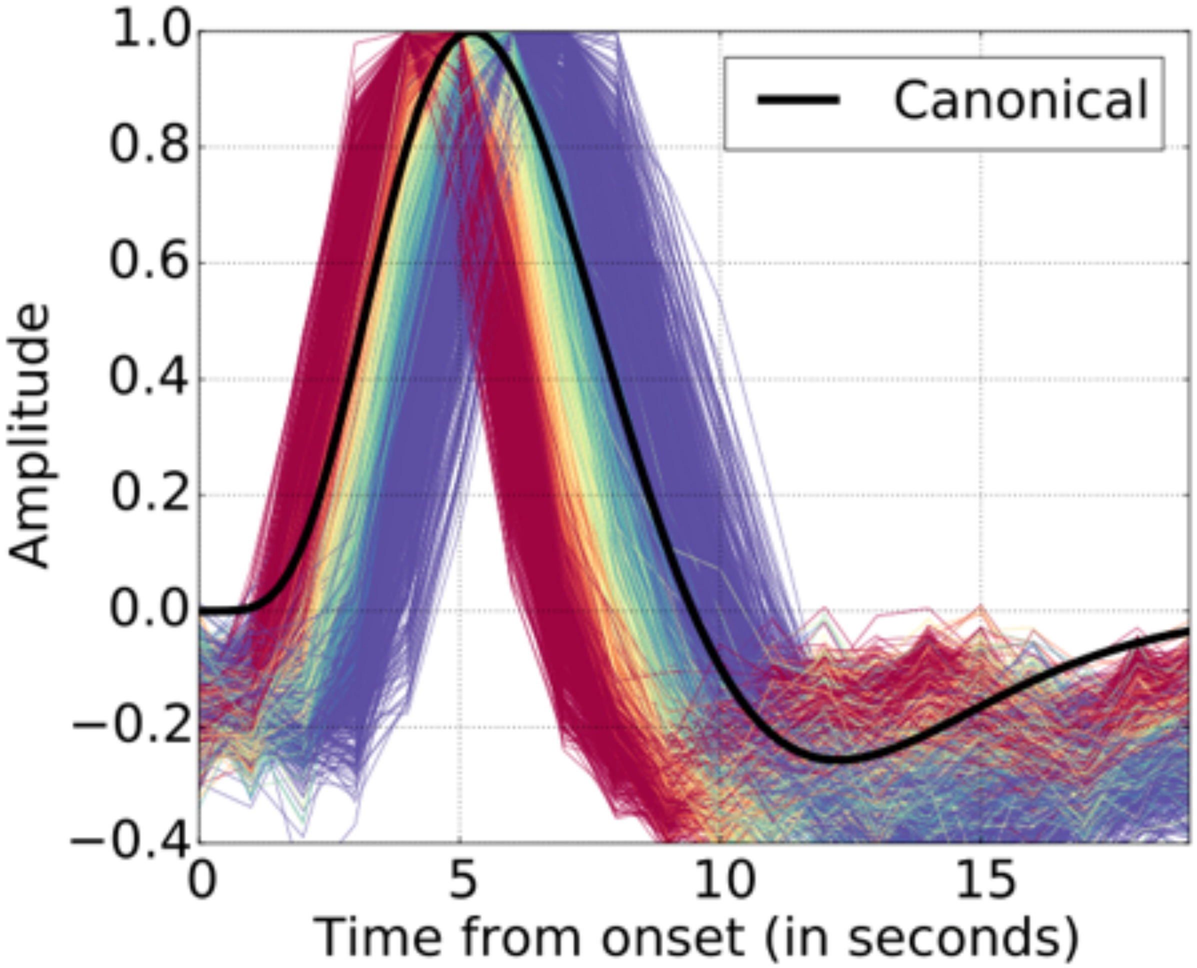
HRF: Hemodynamic response function

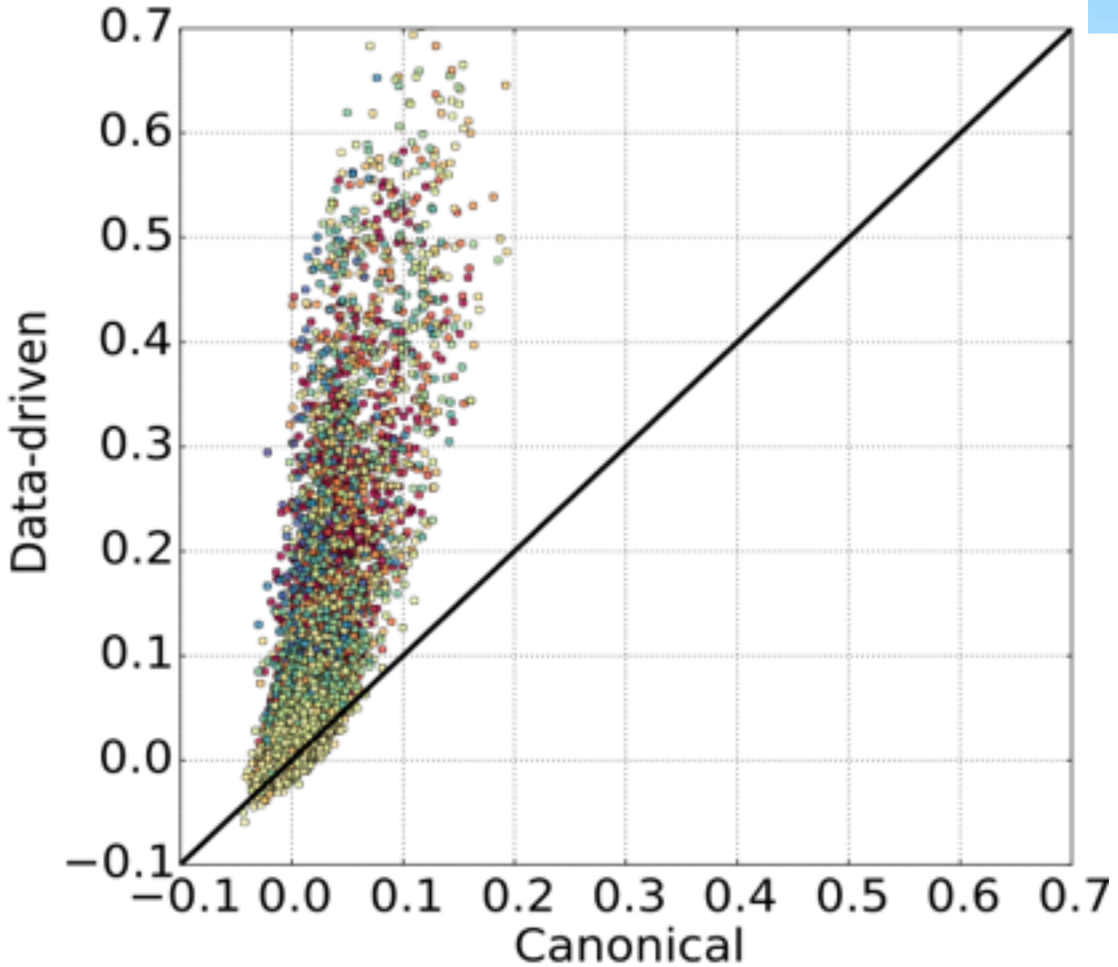


Feature extraction by deconvolution

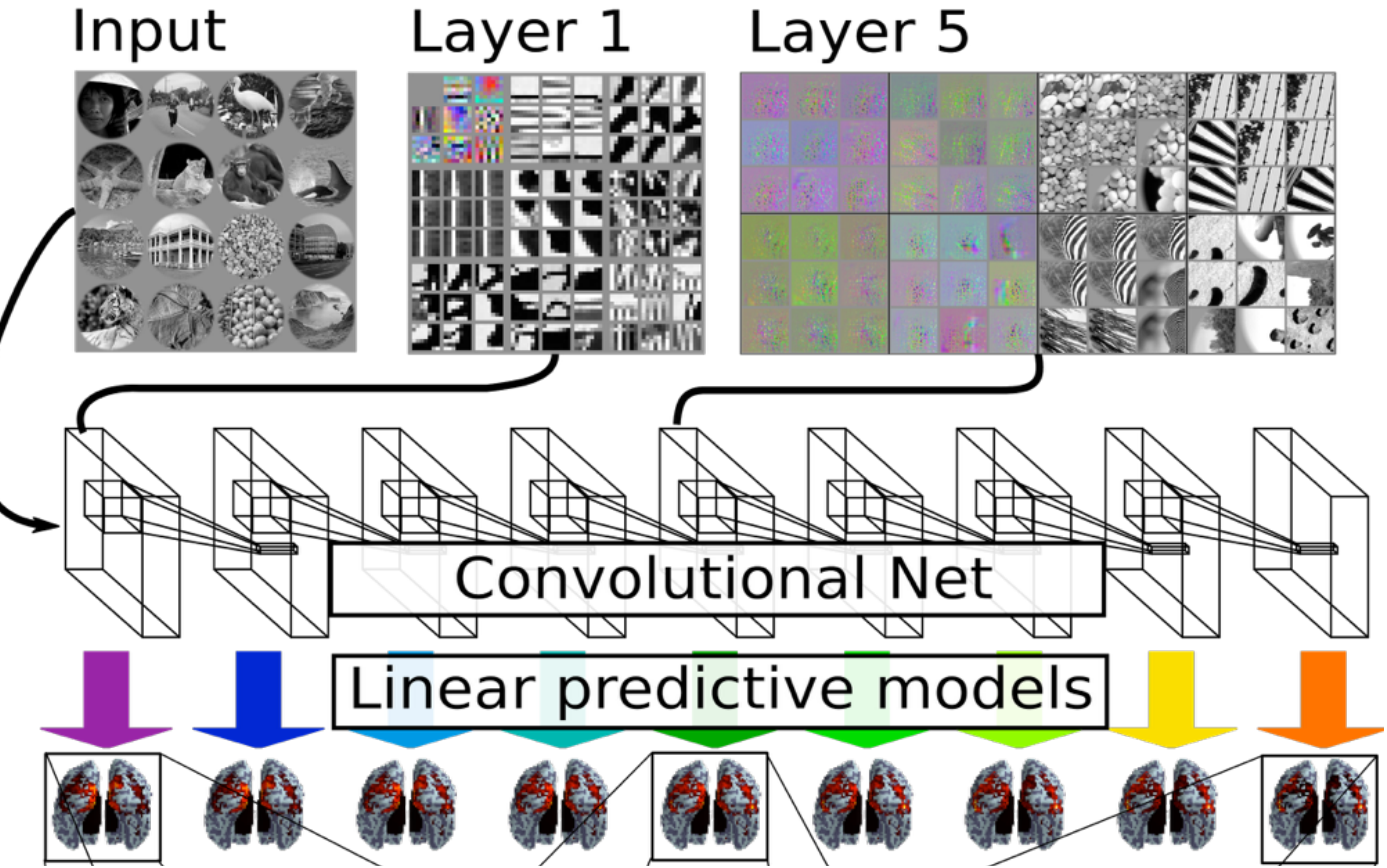
Fabian Pedregosa

One estimates a filter for each voxel using rank constrained optimization





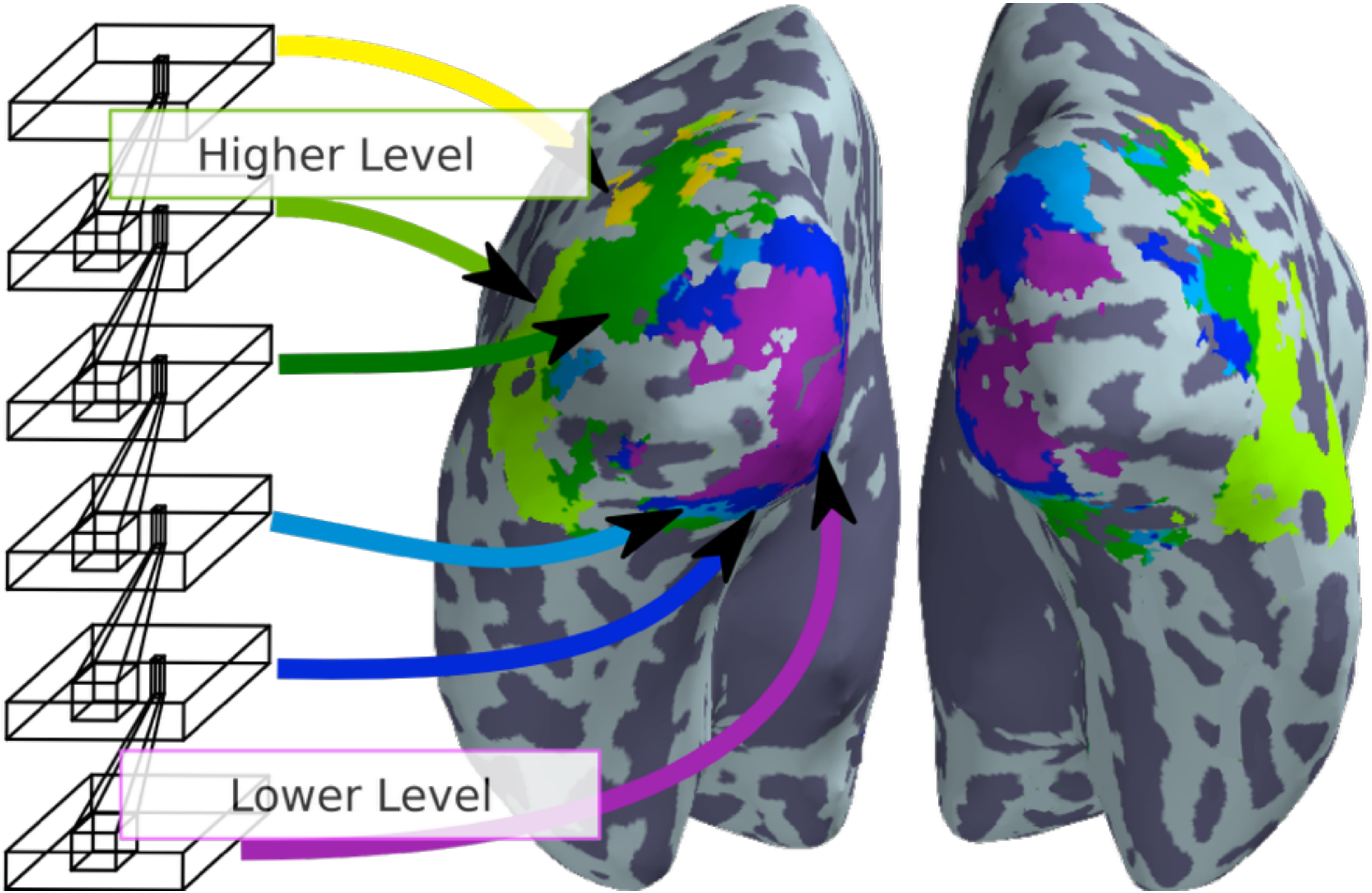
Data-driven HRF estimation for encoding and decoding models, Fabian Pedregosa, Michael Eickenberg, Philippe Ciuciu, Bertrand Thirion and Alexandre Gramfort, Neuroimage 2015


Input
Layer 1
Layer 5
Convolutional Net
Linear predictive models

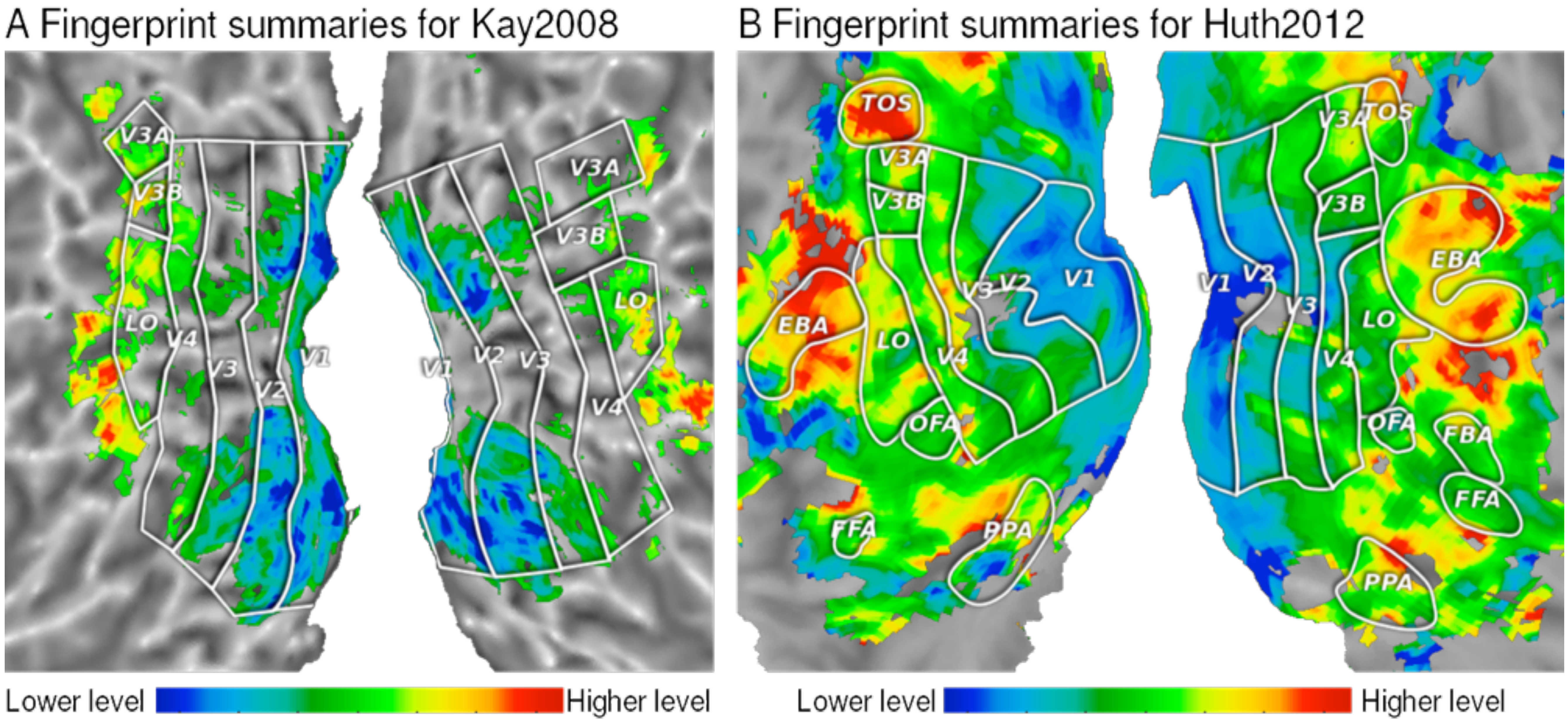
Some details on the data

- Some details about the data:
  - 30GB of stimuli (15 frames/s in .png for 3h)
  - about 4,000 volumes
  - about 10GB of raw data
  - 30,000 “good” voxels
  - > 3h in the scanner

Best Predicting Layers per Voxel



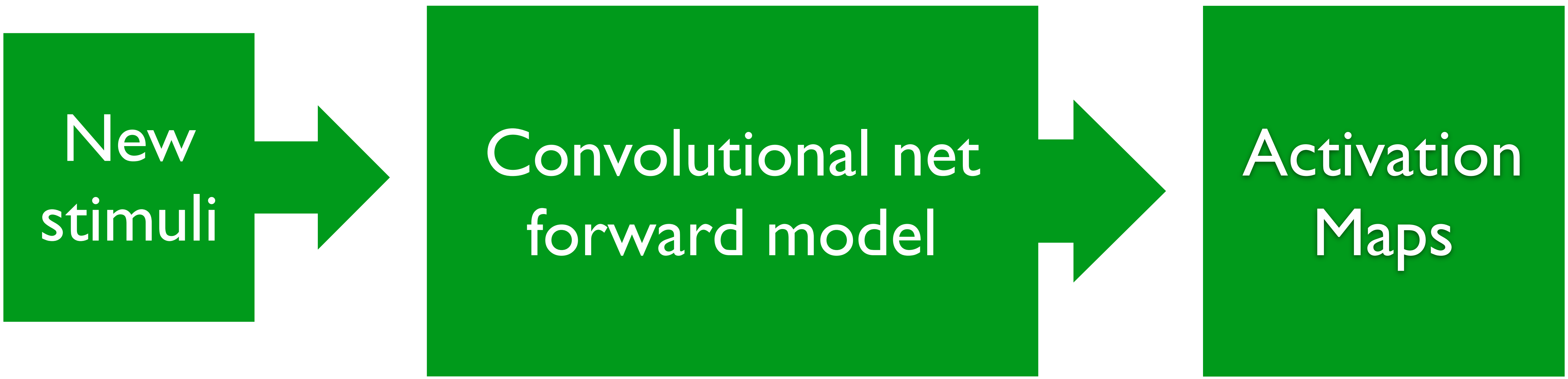
Fingerprints summary statistic



Photos Videos

2 public datasets from UC Berkeley

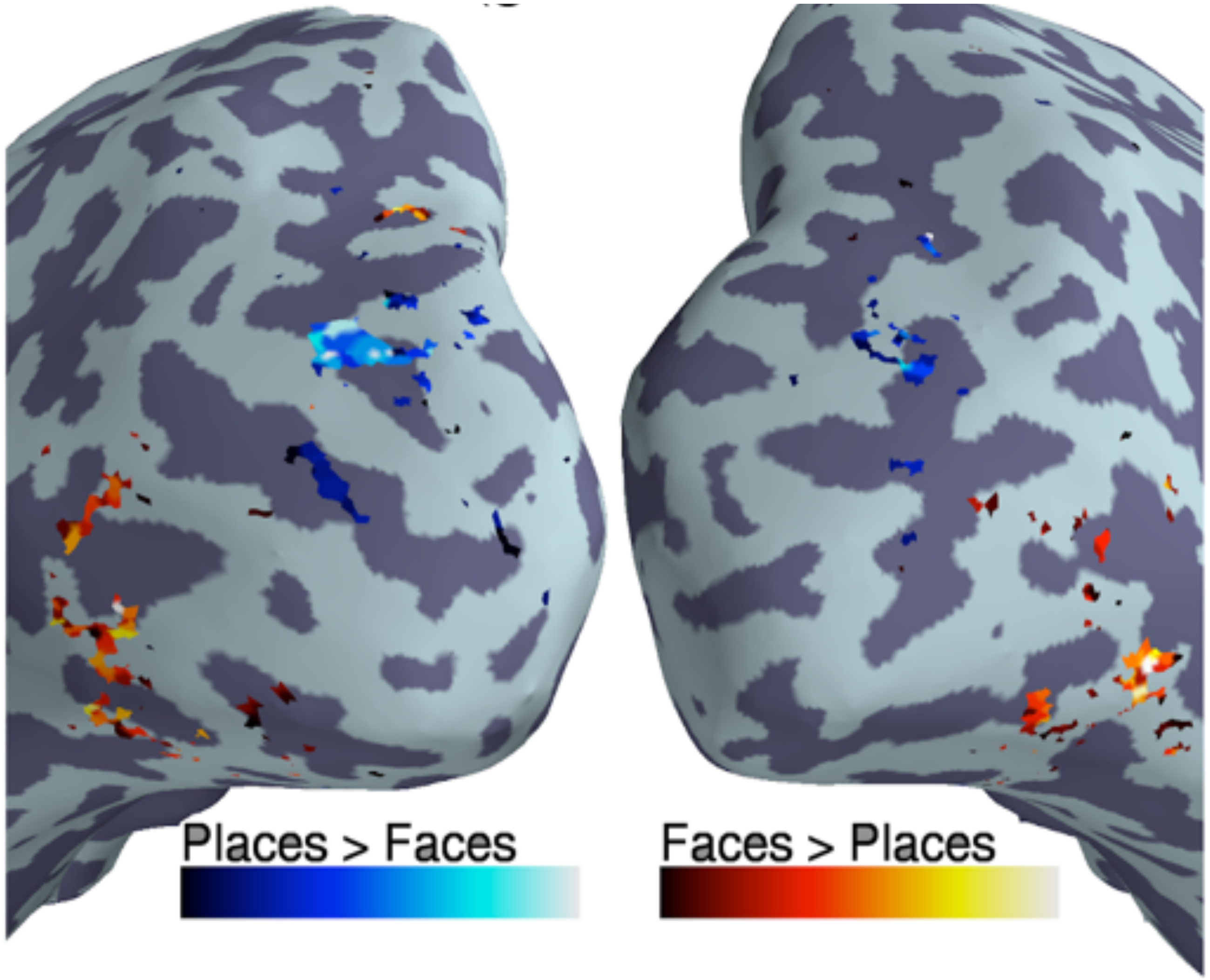
Synthesizing Brain activation maps



Did we learn a good forward model of brain activation as seen with fMRI?

Faces vs Places: Ground Truth

Stimuli from [Kay 2008]
Close-up faces and scenes



Contrast of
stimuli from [Kay 2008]
Close-up faces and scenes

Let’s take a step back…

What is changing?

Volume (Computational issues)

- Standard MEG Study (25 subjects, 10 GB per subject)
- Human Connectome Project (18GB x1000 subjects), USA with first MEG data released in March 2015 (100 subjects)
- Human Brain Project, EU

Data variability (Computational issues)

- 7000 fMRI pipelines lead to different neuroscience findings [Carp 2012]

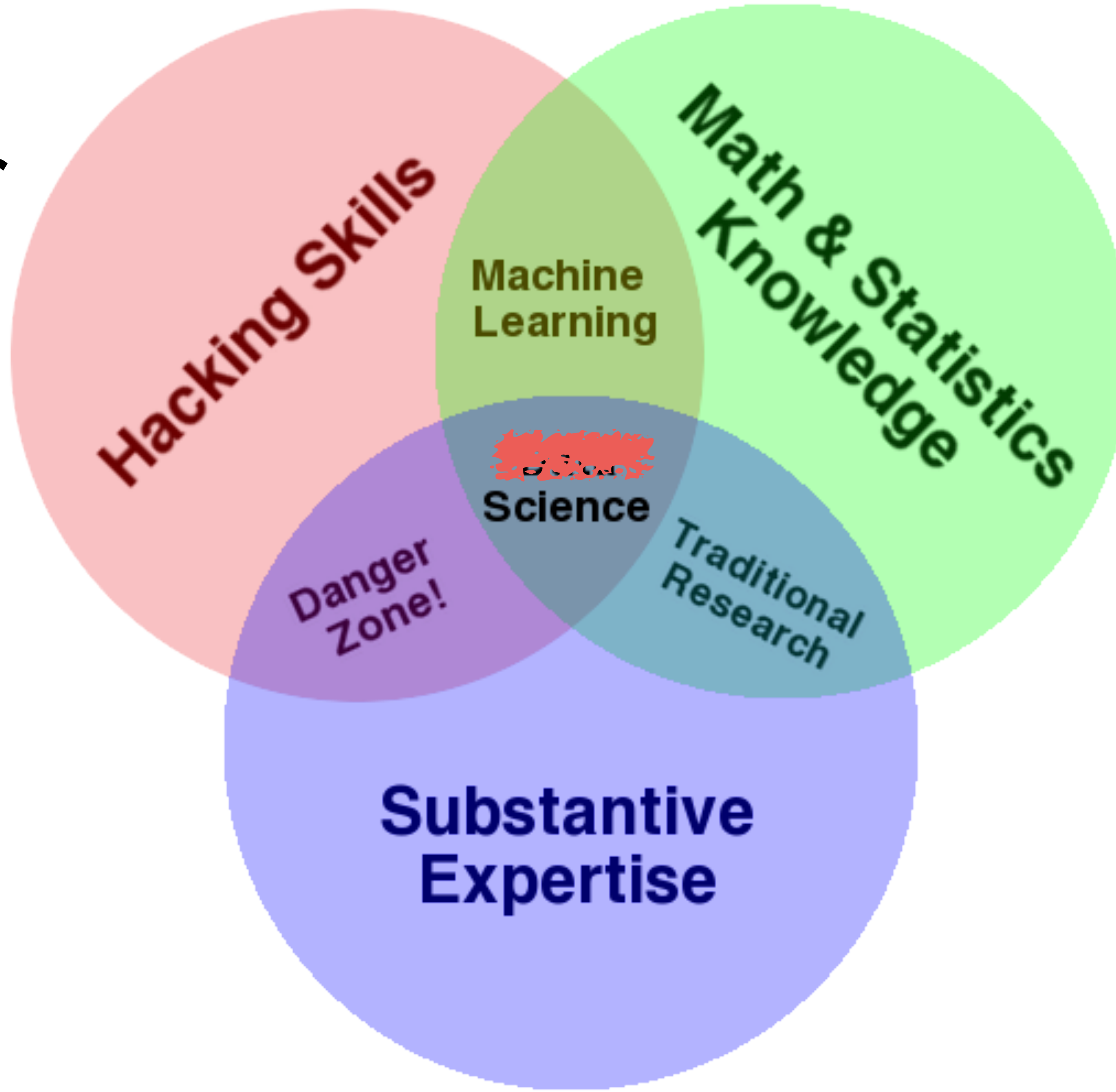
Conclusion

- The world of neuroimaging is full of challenging stats and optimization problems ...
- … look at the data to find the relevant ones

"An approximate answer to the right problem is worth a good deal more than an exact answer to an approximate problem. ~ John Tukey"

Computer Science

Statistics



neuro ~~Domain~~ Science

Drew Conway

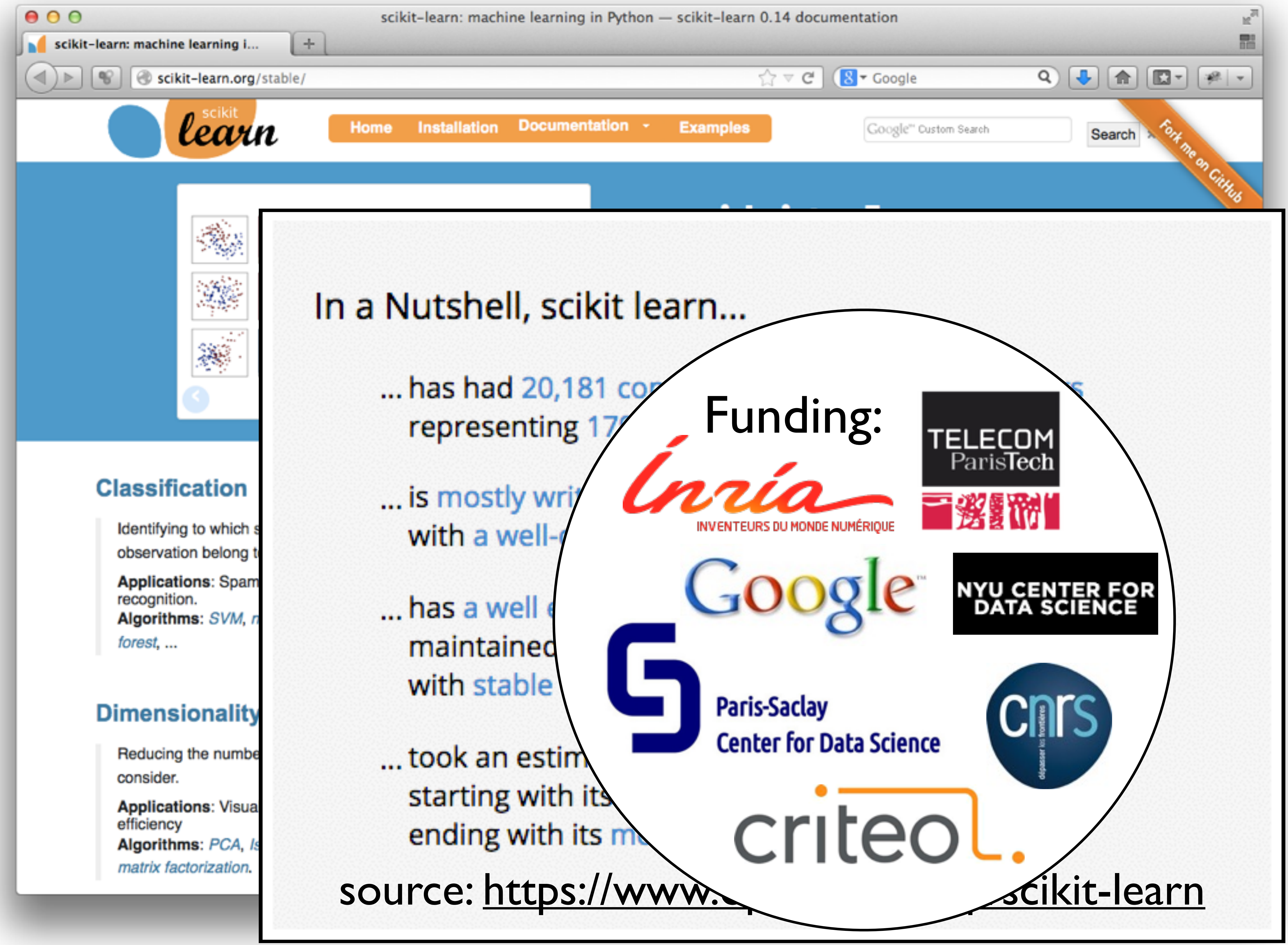
The human inverse problem

Observations

Sparse, Convex optimization, STFT, proximal operator, neural networks, etc...

brain imaging people

How do you solve this ill-posed problem?



scikit-learn: machine learning in Python — scikit-learn 0.14 documentation
scikit-learn.org/stable/
scikit learn
Home
Installation
Documentation
Examples
Search
Fork me on GitHub
Classification
Identifying to which
observation belong
Applications: Spam
recognition.
Algorithms: SVM,
forest, ...
Dimensionality
Reducing the numbe
consider.
Applications: Visua
efficiency
Algorithms: PCA,
matrix factorization.
In a Nutshell, scikit learn...
... has had 20,181 co
representing 17
... is mostly wri
with a well-
... has a well
maintained
with stable
... took an estim
starting with its
ending with its m
Funding:
Inria
INVENTEURS DU MONDE NUMÉRIQUE
TELECOM ParisTech
Google
NYU CENTER FOR DATA SCIENCE
Paris-Saclay Center for Data Science
cnrs
criteo
source: https://www. scikit-learn

MNE Tutorials Gallery Manual API FAQ Cite Site Page Search

MNE

MEG + EEG ANALYSIS & VISUALIZATION

http://www.martinos.org/mne

MNE is a community-driven software package designed for for **processing electroencephalography (EEG) and magnetoencephalography (MEG) data** providing comprehensive tools and workflows for:

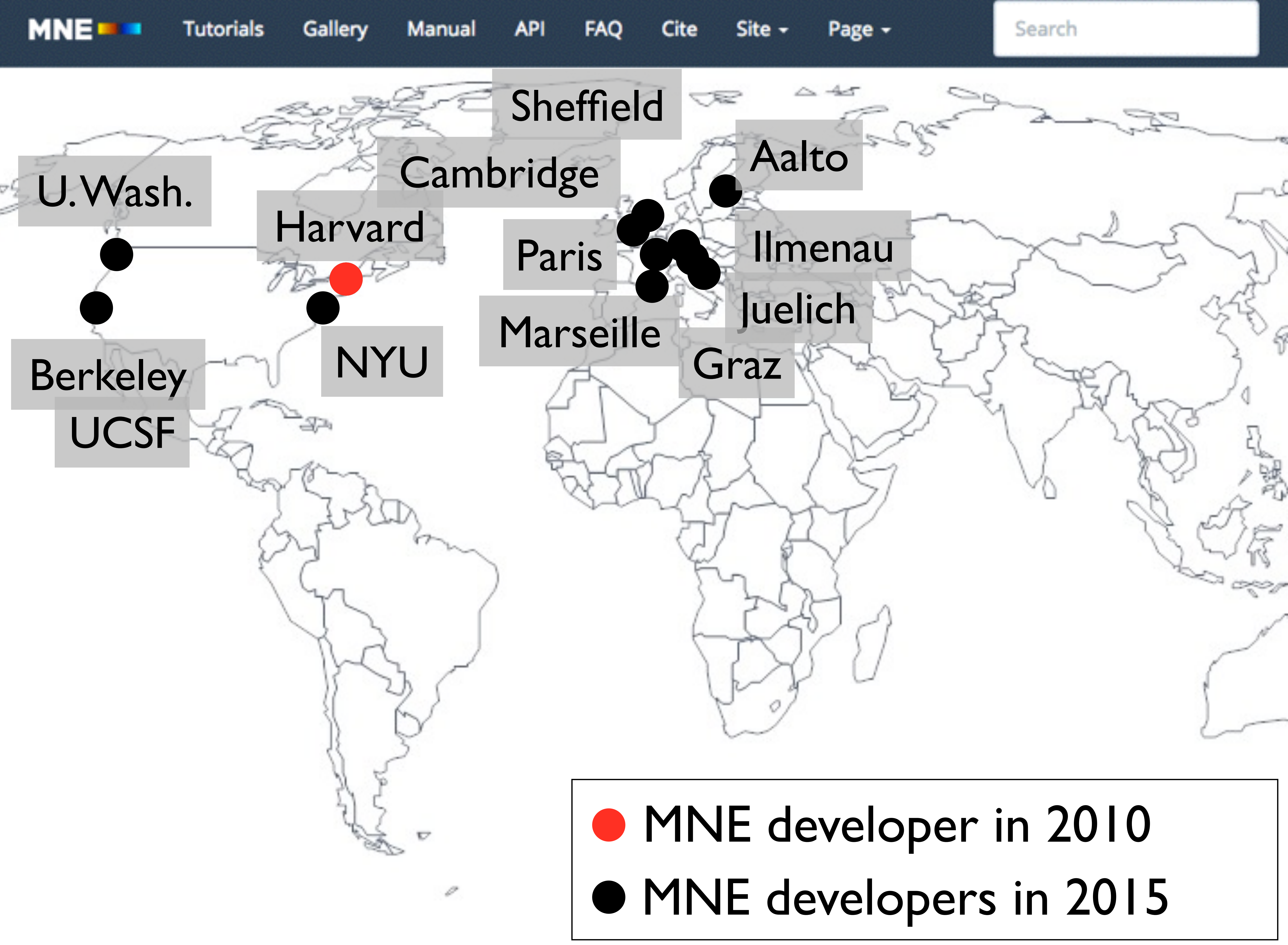
1. Preprocessing
2. Source estimation
3. Time-frequency analysis
4. Statistical testing
5. Estimation of functional connectivity
6. Applying machine learning algorithms
7. Visualization of sensor- and source-space data

MNE includes a comprehensive Python package (provided under the simplified BSD license), supplemented by tools compiled from C code for the LINUX and Mac OSX operating systems, as well as a MATLAB toolbox.

Documentation

- Getting Started
- What's new
- Cite MNE
- Related publications
- Tutorials
- Examples Gallery
- Manual
- API Reference
- Frequently Asked Questions
- Advanced installation and setup
- MNE with CPP

MNE software for processing MEG and EEG data, A. Gramfort, M. Luessi, E. Larson, D. Engemann, D. Strohmeier, C. Brodbeck, L. Parkkonen, M. Hämäläinen, Neuroimage 2013


MNE
Tutorials
Gallery
Manual
API
FAQ
Cite
Site
Page
Search
Sheffield
Aalto
Cambridge
U.Wash.
Harvard
Paris
Ilmenau
Juelich
Marseille
Graz
NYU
Berkeley
UCSF
MNE developer in 2010
MNE developers in 2015

Thanks !

- M. Hämäläinen
- M. Kowalski
- D. Strohmeier
- J. Haueisen
- Y. Bekhti
- M. Jas
- G. Varoquaux
- B. Thirion
- M. Eickenberg
- F. Pedregosa
- J. Salmon
- O. Fercoq
- E. Ndiaye
- … the scikit-learn contributors
- … the MNE contributors

Post-docs positions available !

Contact http://alexandre.gramfort.net

GitHub : @agramfort *Twitter : @agramfort*

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

References

Gramfort et al., *Mixed-norm estimates for the M/EEG inverse problem using accelerated gradient methods*, Physics in Medicine and Biology, 2012

Gramfort et al. *Time-frequency mixed-norm estimates: Sparse M/EEG imaging with non-stationary source activations*, NeuroImage, 2013

Strohmeier et al., *Improved MEG/EEG source localization with reweighted mixed-norms,* International Workshop on Pattern Recognition in Neuroimaging (PRNI), 2014

Strohmeier et al., *MEG/EEG source imaging with a non-convex penalty in the time-frequency domain,* Int. Workshop on Pattern Recognition in Neuroimaging (PRNI), 2015

Fercoq et al., *Mind the duality gap: Safer rules for the Lasso,* ICML, 2015

Ndiaye et al., *GAP Safe screening rules for sparse multi-task and multi-class models,* NIPS, 2015

Strohmeier et al. *The iterative reweighted Mixed-Norm Estimate for spatio-temporal MEG/EEG source localization.* IEEE Transactions on Medical Imaging (TMI), 2016.

Bekhti et al. *M/EEG source localization with multi-scale time-frequency dictionaries,* Int. Workshop on Pattern Recognition in Neuroimaging (PRNI), 2015

Fabian Pedregosa, Michael Eickenberg, Philippe Ciuciu, Bertrand Thirion and Alexandre Gramfort, *Data-driven HRF estimation for encoding and decoding models,* Neuroimage 2015

Michael Eickenberg, Alexandre Gramfort, Gaël Varoquaux, Bertrand Thirion, Seeing it all: Convolutional network layers map the function of the human visual system, (submitted)