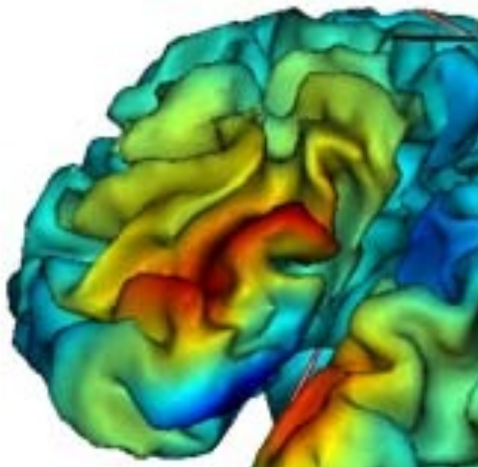
How much stats does it take to look at the brain at a millisecond time scale?

Alexandre Gramfort

alexandre.gramfort@telecom-paristech.fr

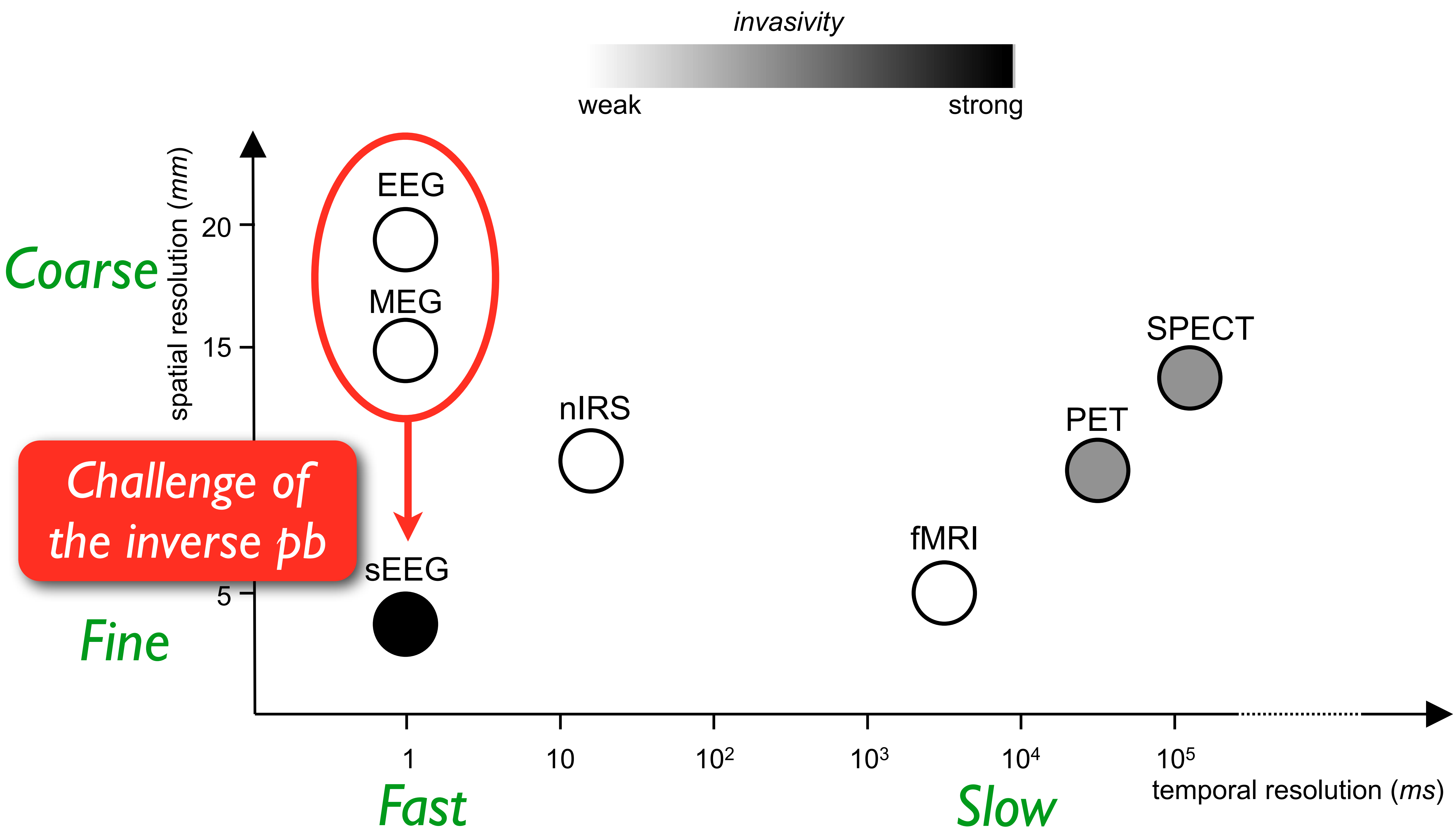
Assistant Prof. Telecom ParisTech

CEA - Neurospin, France

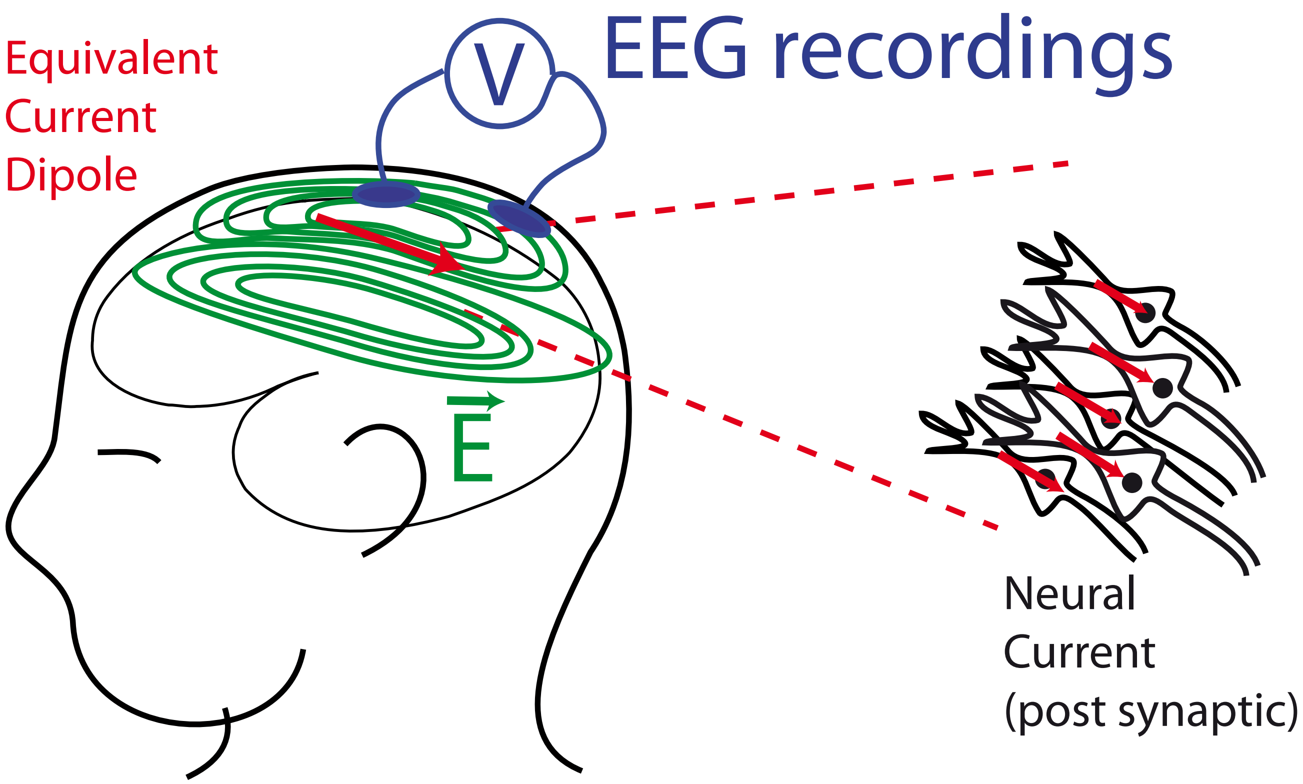


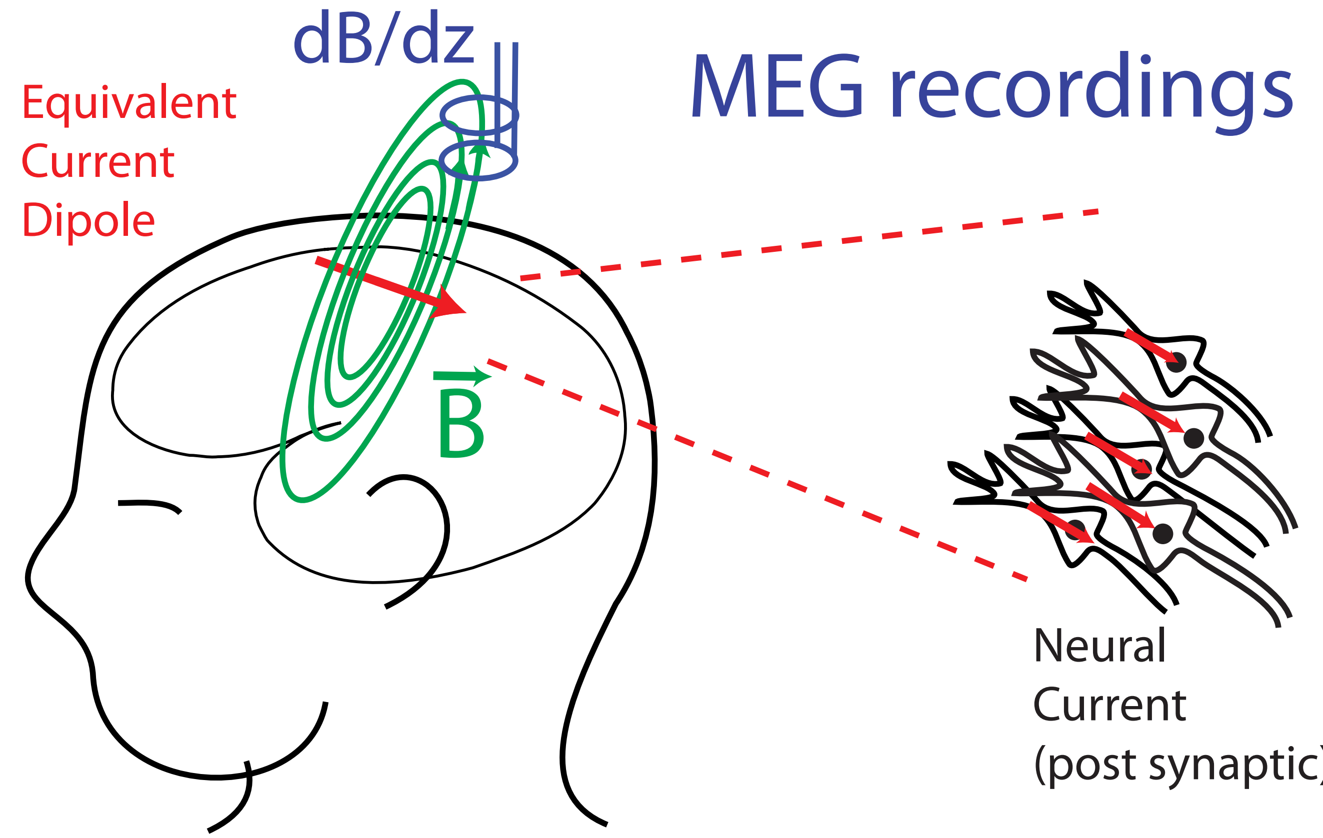
Biomag - Aug. 2014

The overall goal

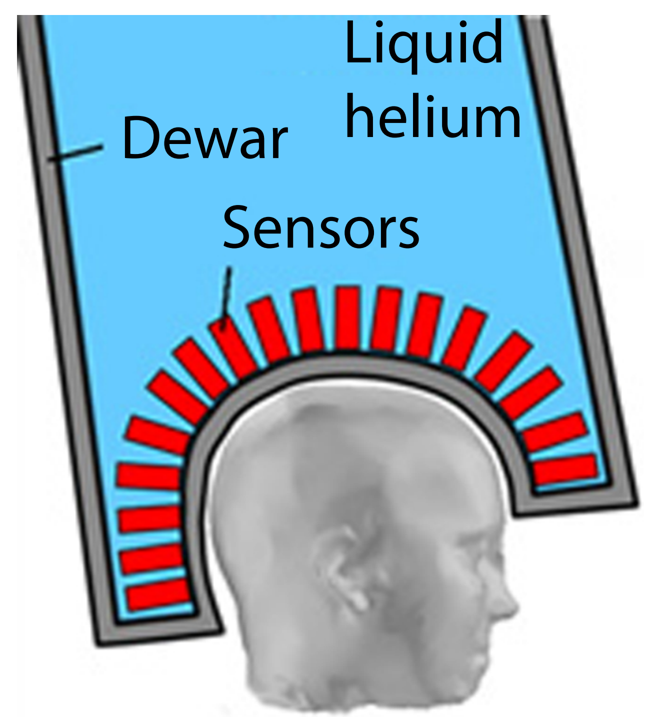


EEG & MEG in a nutshell



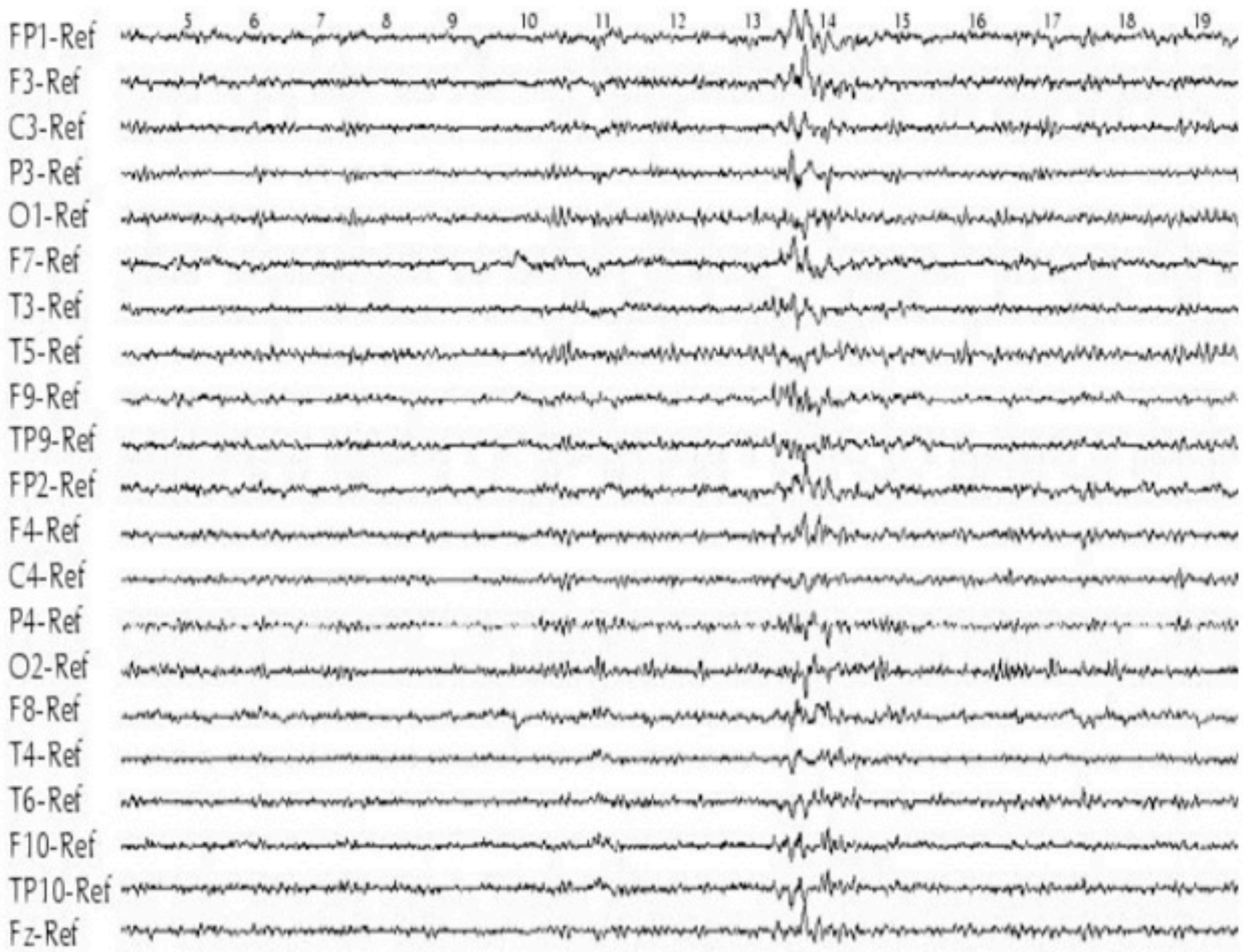


First EEG recordings in 1929 by H. Berger



Hôpital La Timone
Marseille, France

M/EEG Measurements



Sample EEG measurements

EEG :

- ≈ 32 to 100 sensors

MEG :

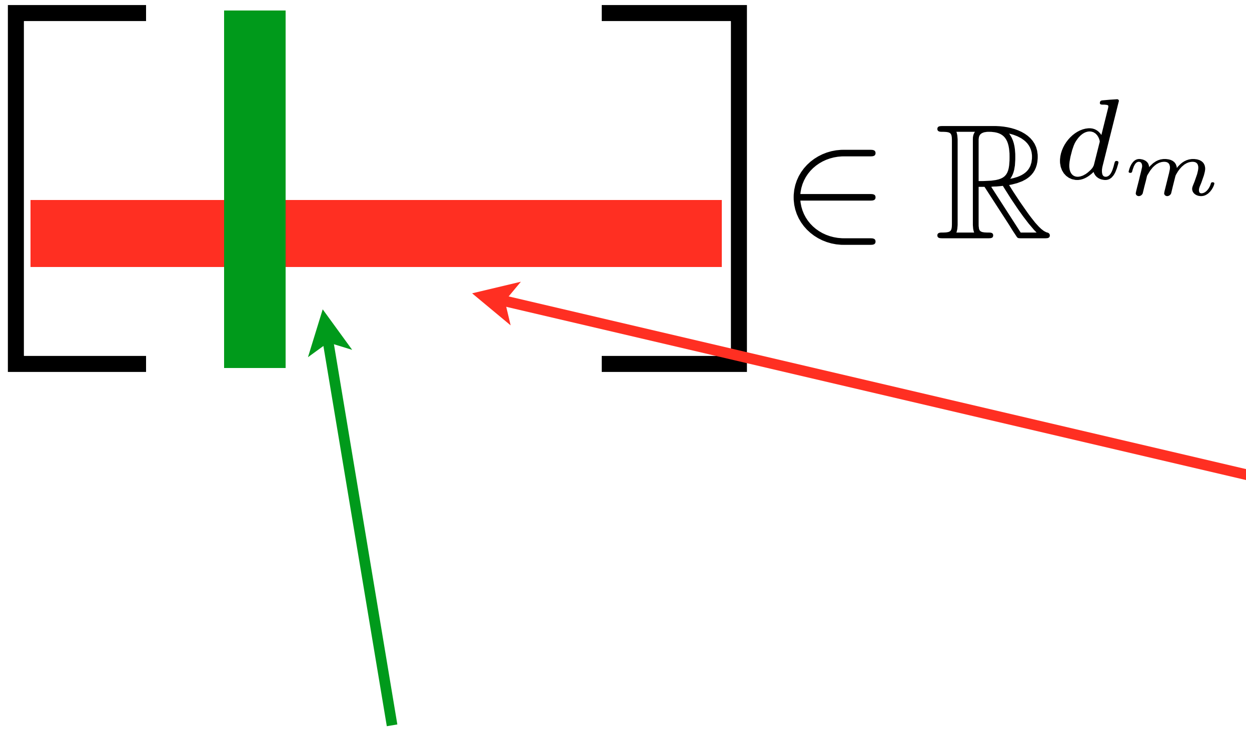
- ≈ 150 to 300 sensors

Sampling between 250 and1000 Hz

THM: High temporal resolution

M/EEG Measurements: Notation

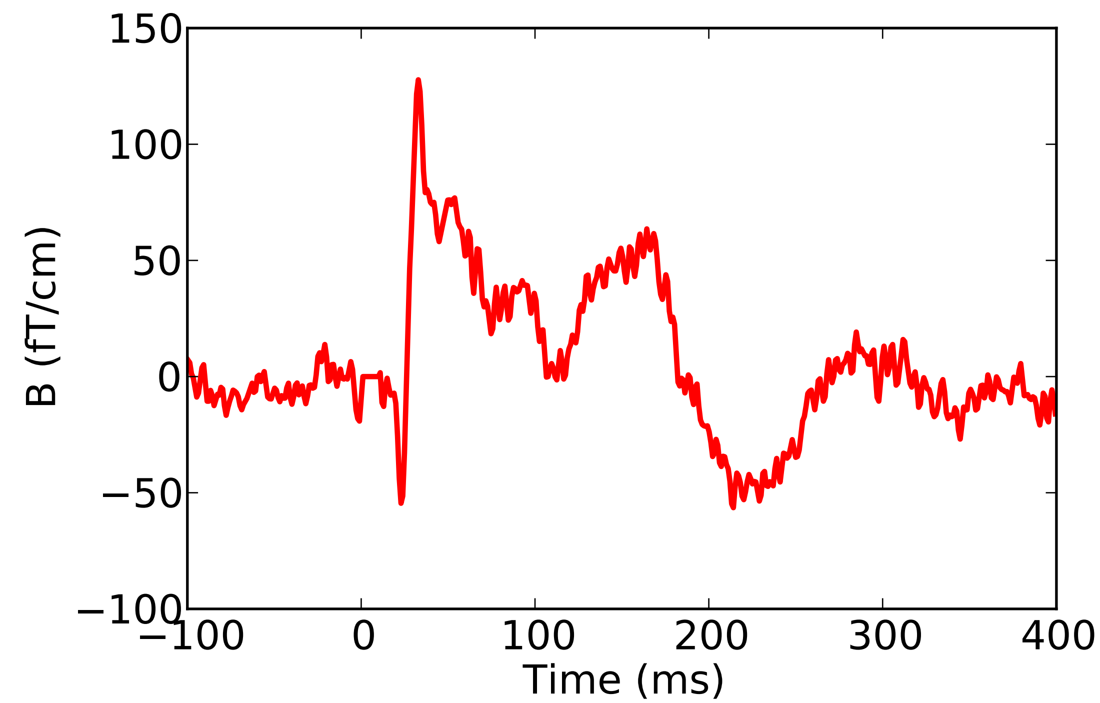
$$M = [\ \ ] \in \mathbb{R}^{d_m \times d_t}$$



MEG and/or EEG

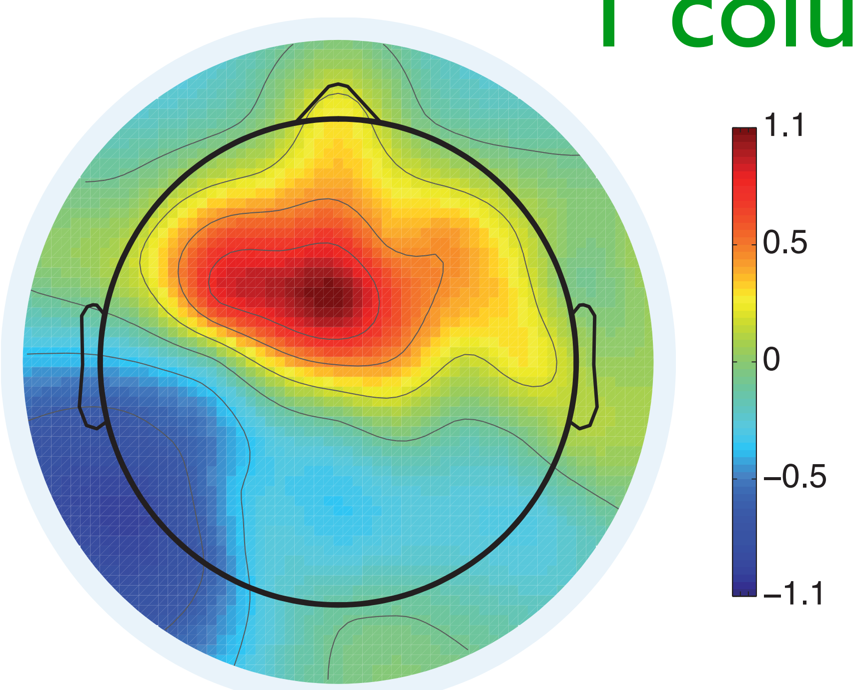
d_m : Nb of sensors

d_t : Nb of time points

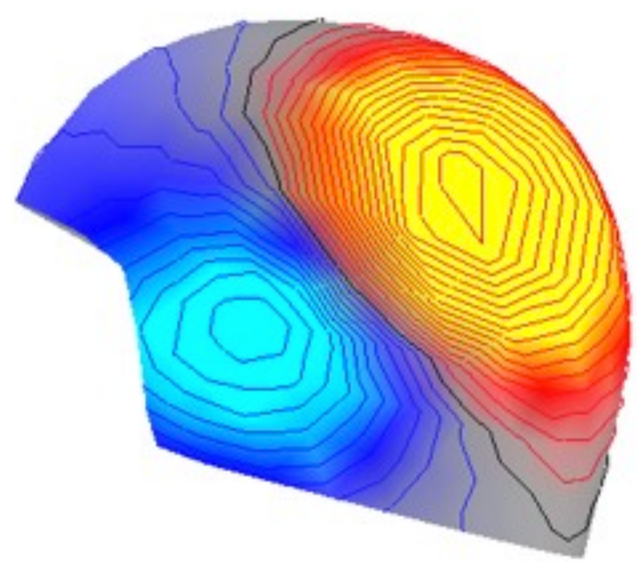


1 column

1 row = 1 time series on 1 sensor

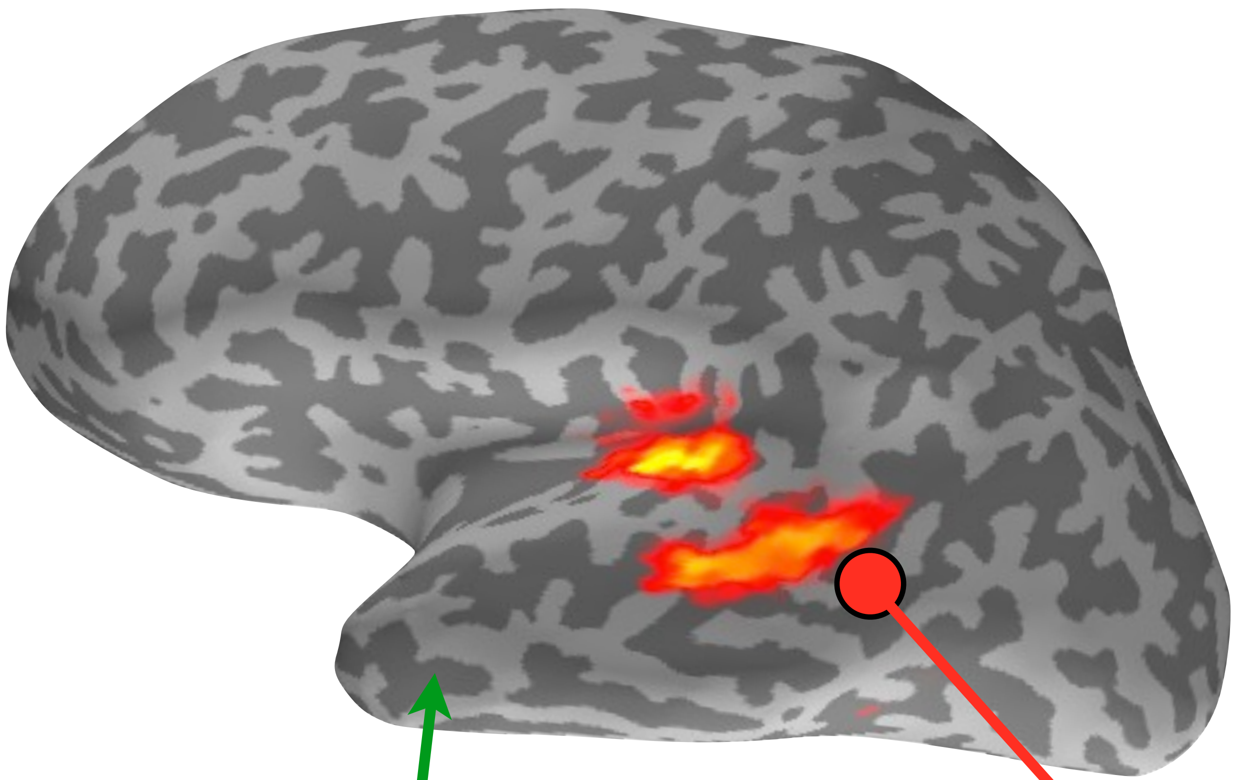


2D topography

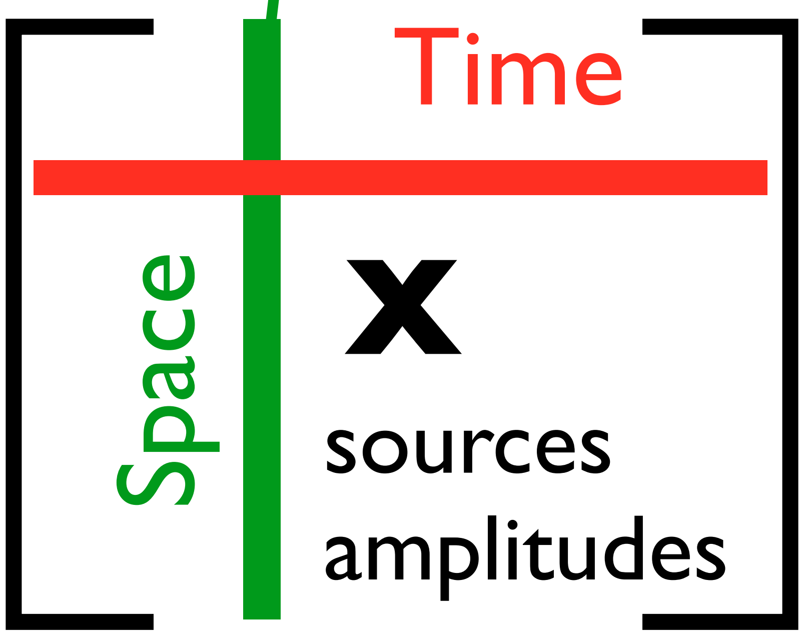


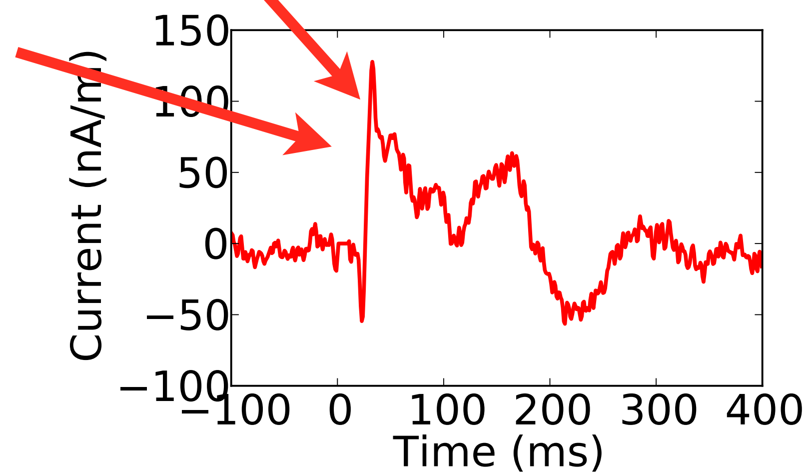
3D topography

Distributed model



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

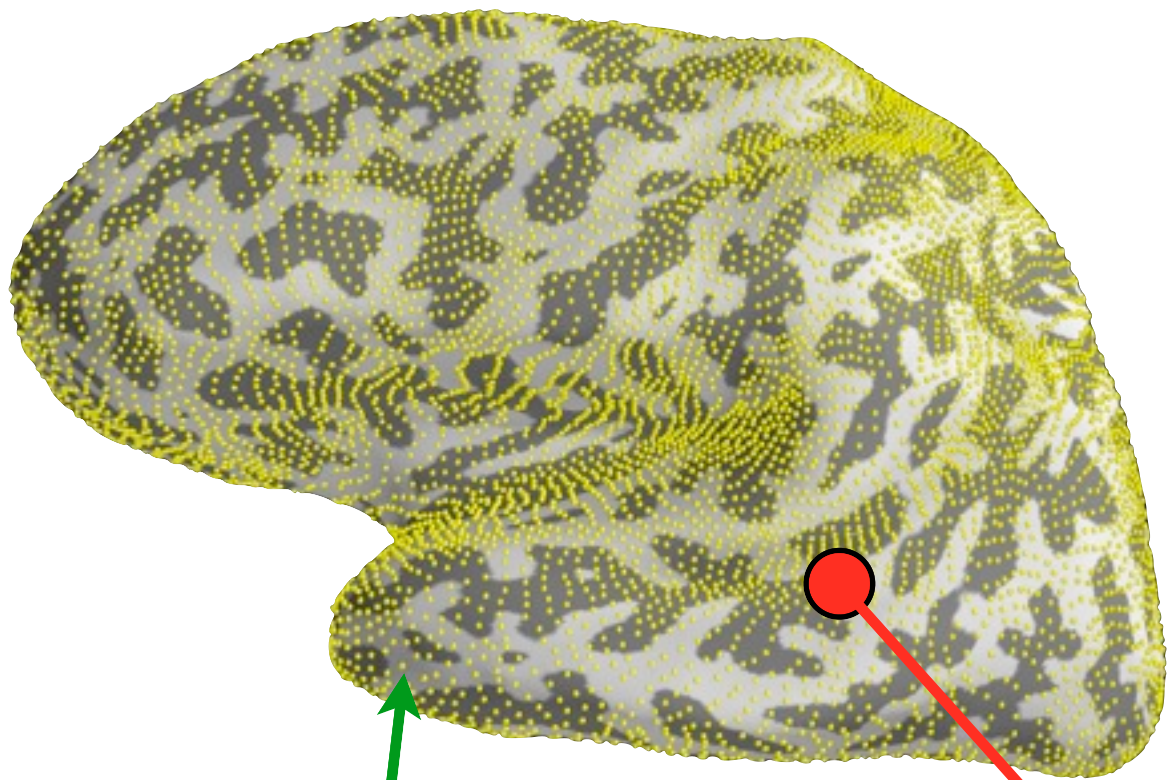




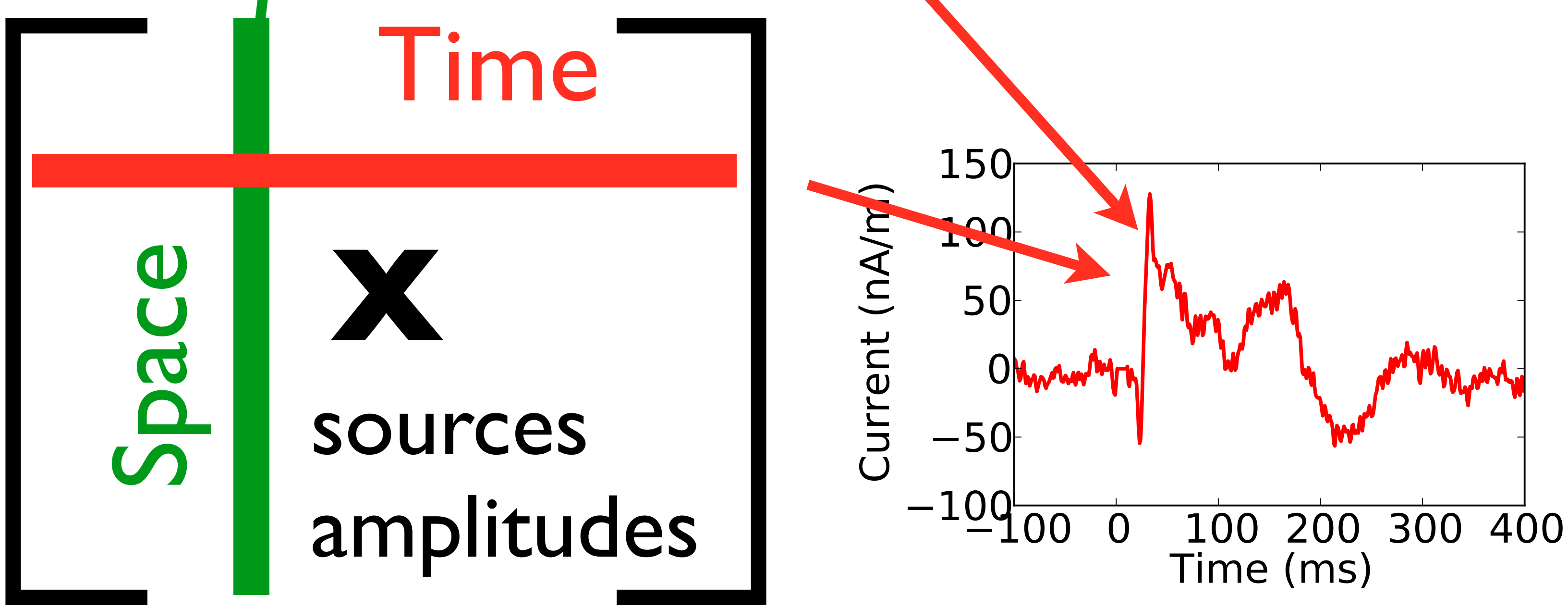
Scalar field defined over time

[Dale and Sereno 93]

Distributed model



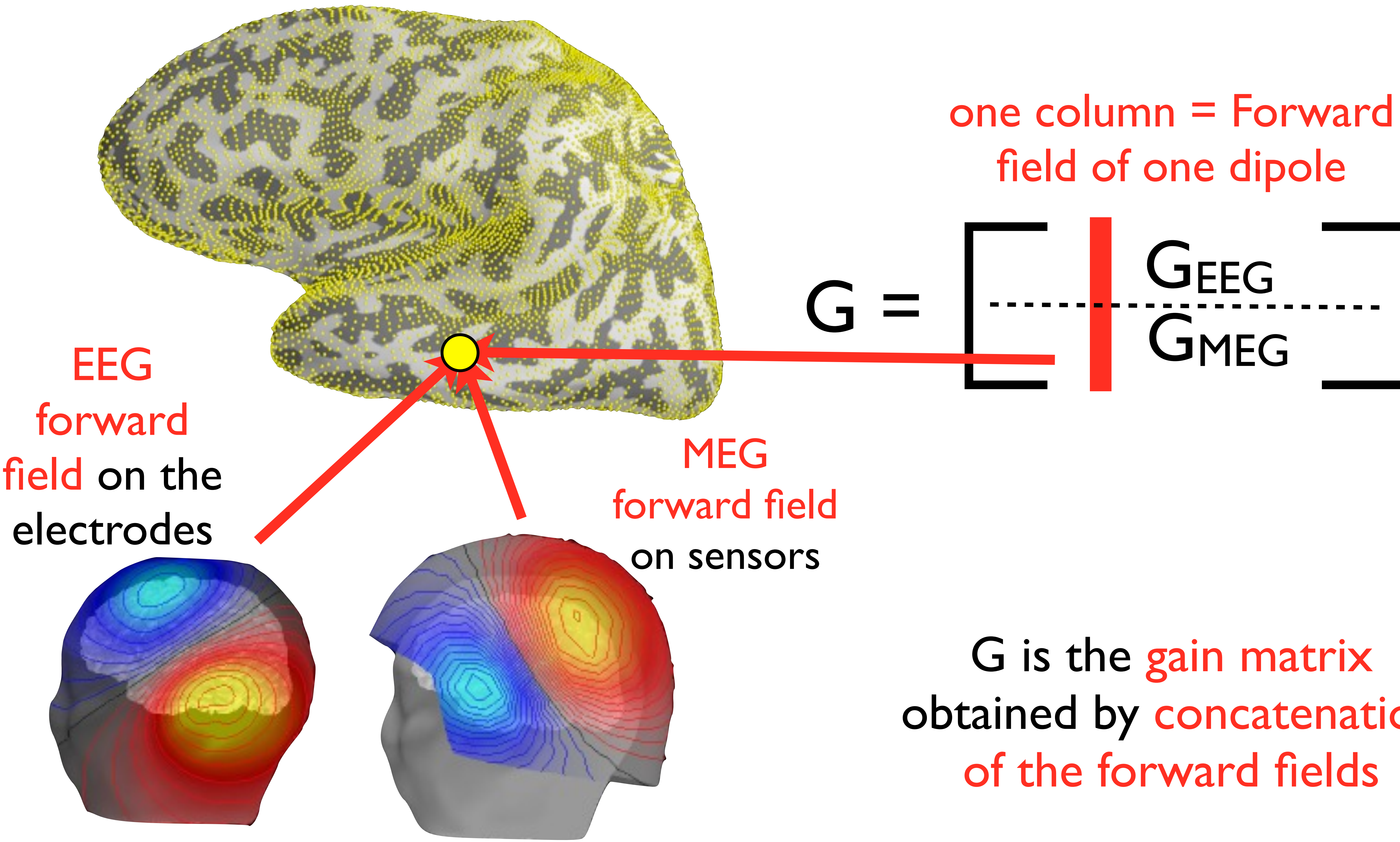
Position 5000 candidate sources over each hemisphere (e.g. every 5mm)



Scalar field defined over time

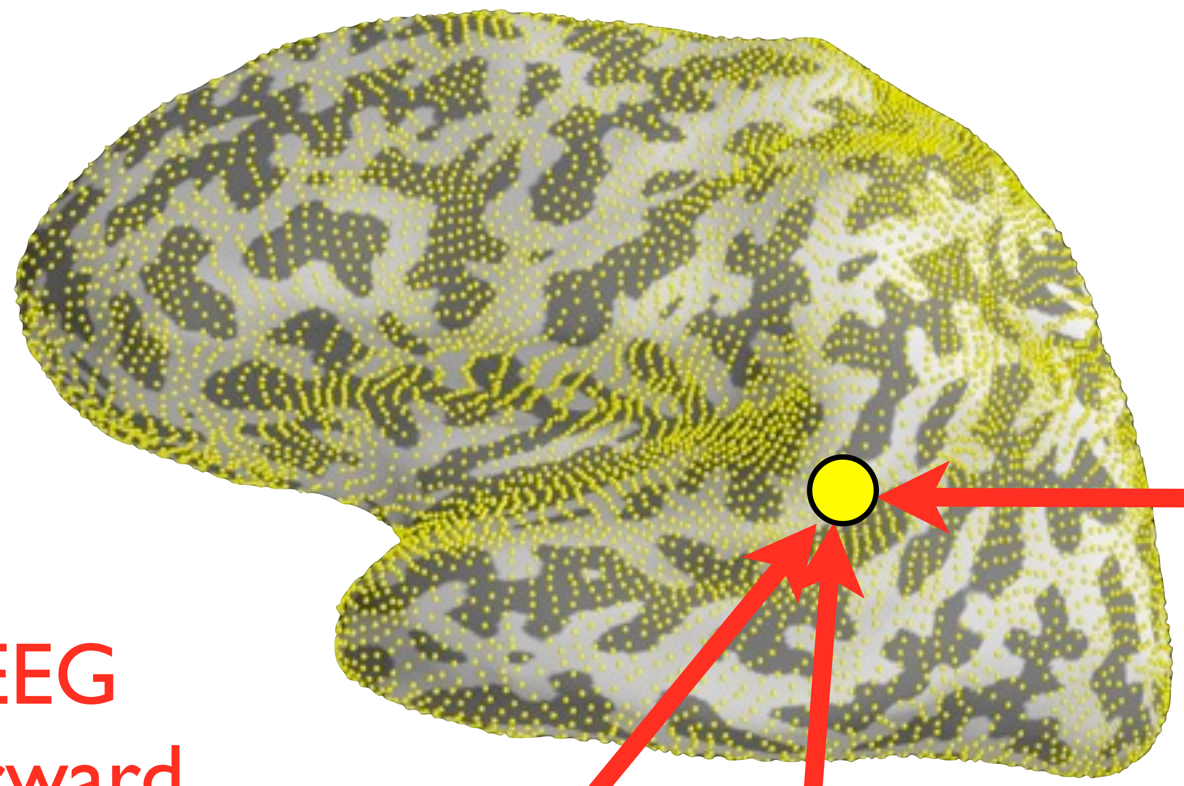
[Dale and Sereno 93]

Distributed model



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

Distributed model

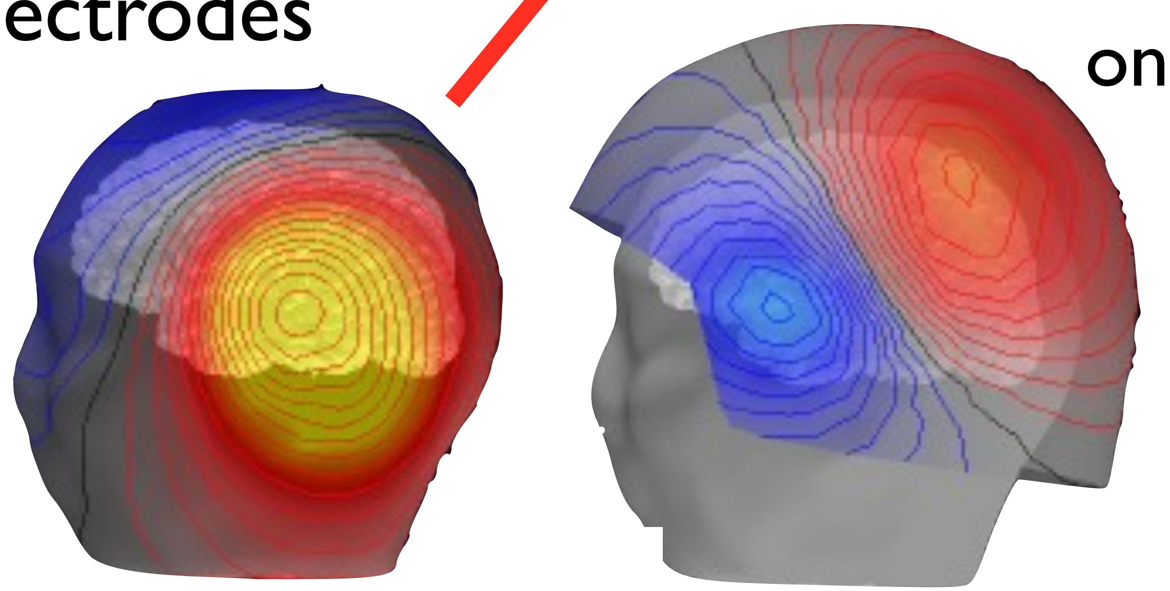


one column = Forward field of one dipole

$$G = \left[\begin{array}{c} G_{EEG} \\ \hline G_{MEG} \end{array} \right]$$

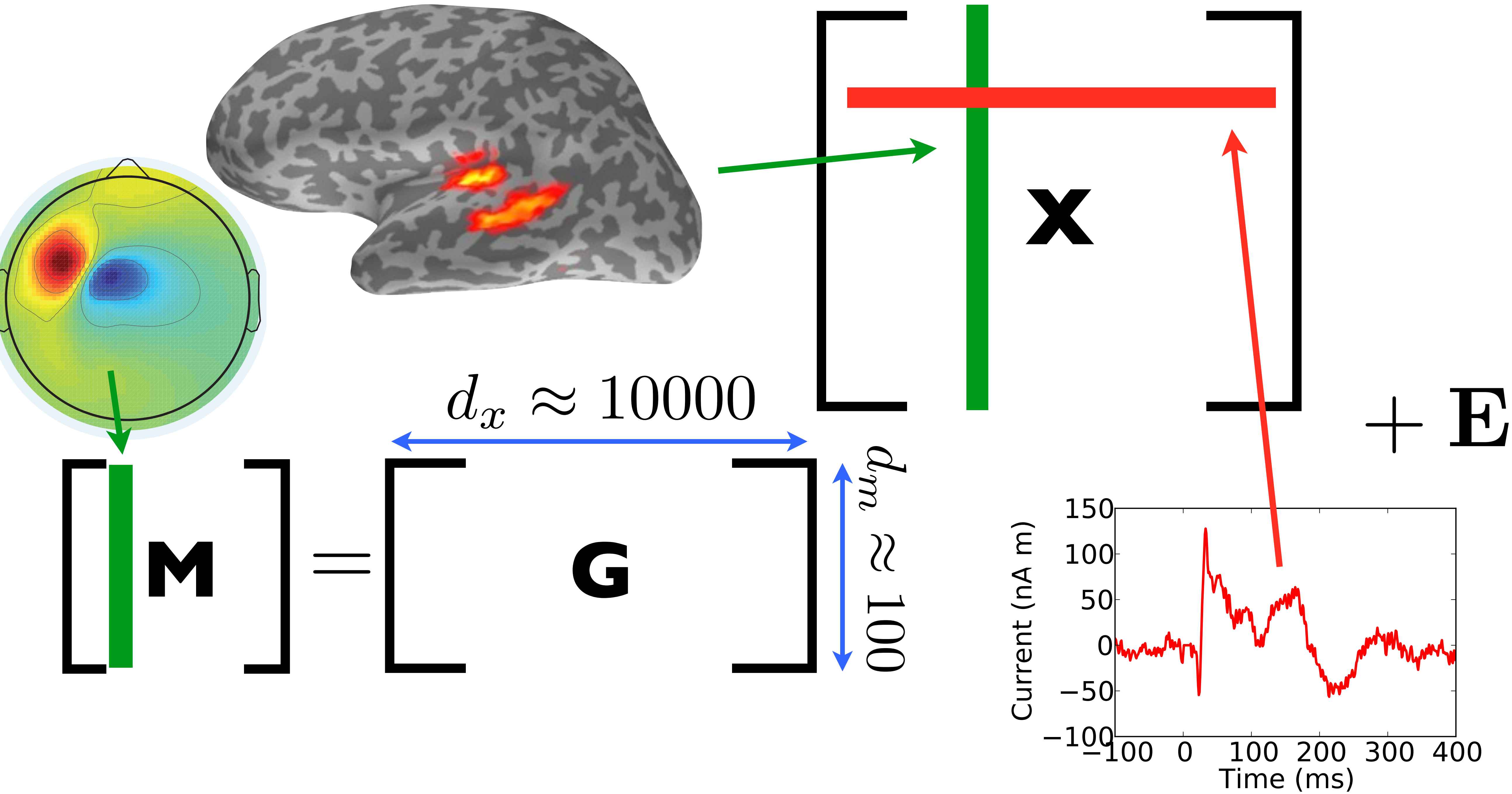
EEG forward field on the electrodes

MEG forward field on sensors



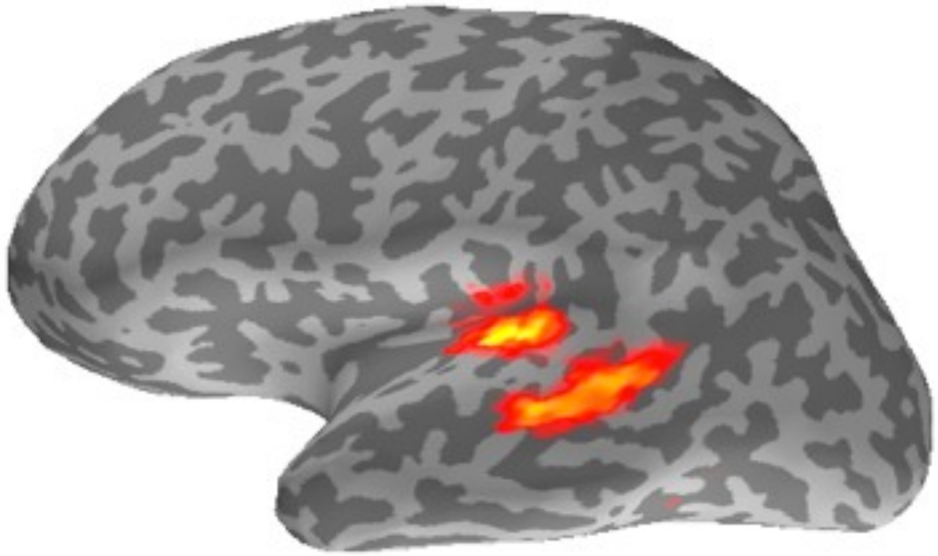
G is the gain matrix obtained by concatenation of the forward fields

M = GX+E : An ill-posed problem



Linear problem with more unknowns than the number of equations: it's ill-posed => Regularize

y = Xw+E : An ill-posed problem



Standard statistics notations

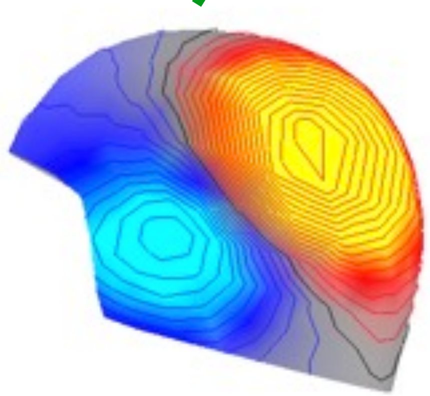
w (or β) regression coefficients

$d_x \approx 10000$

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

X design matrix

$d_m \approx 100$



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

Inverse problem framework

Penalized (variational) formulation (with whitened data):

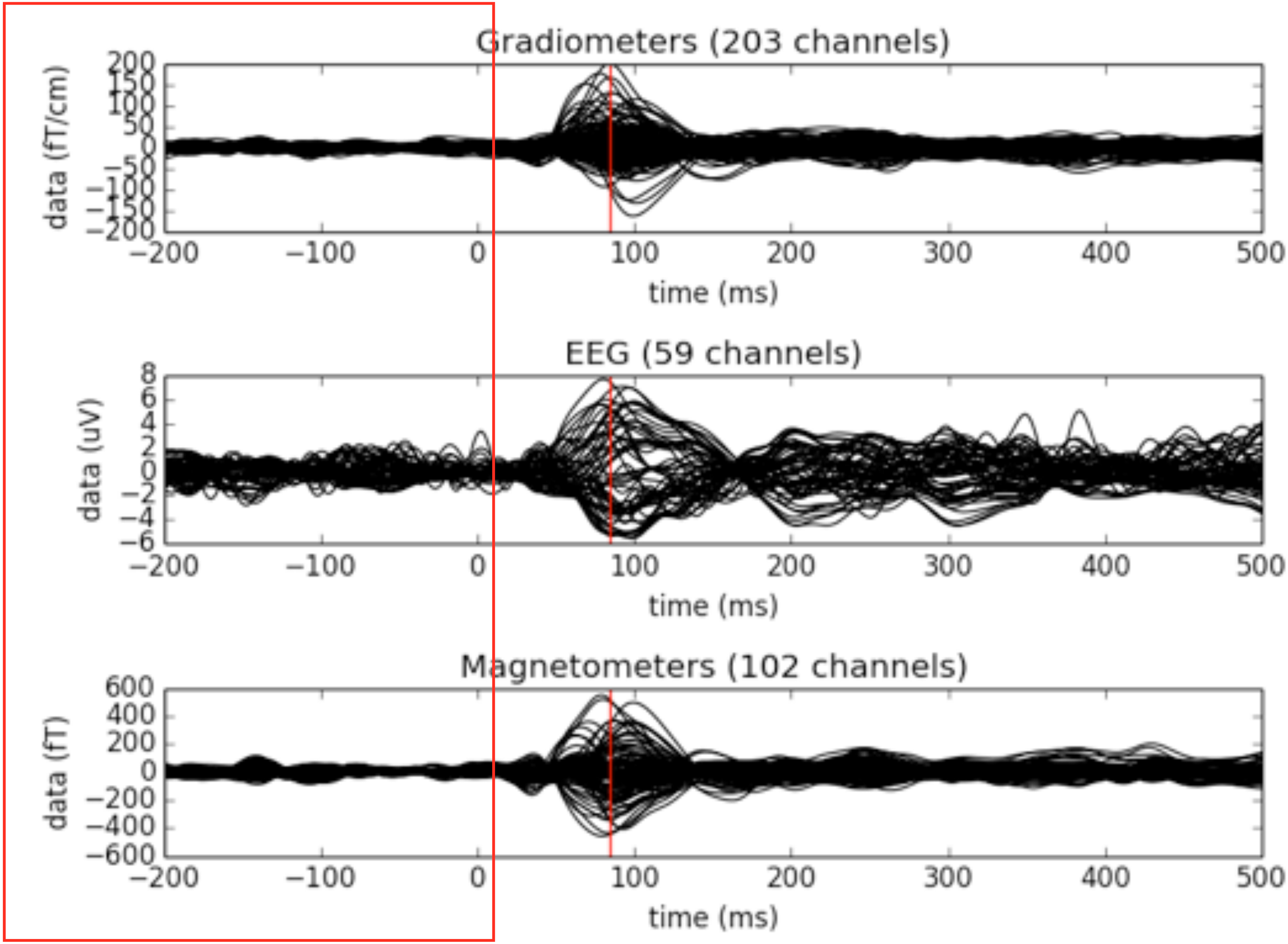
$$\mathbf{X}^* = \arg\min_{\mathbf{X}} \|\mathbf{M} - \mathbf{G}\mathbf{X}\|_F^2 + \lambda\phi(\mathbf{X}), \lambda > 0$$

Data fit Regularization

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

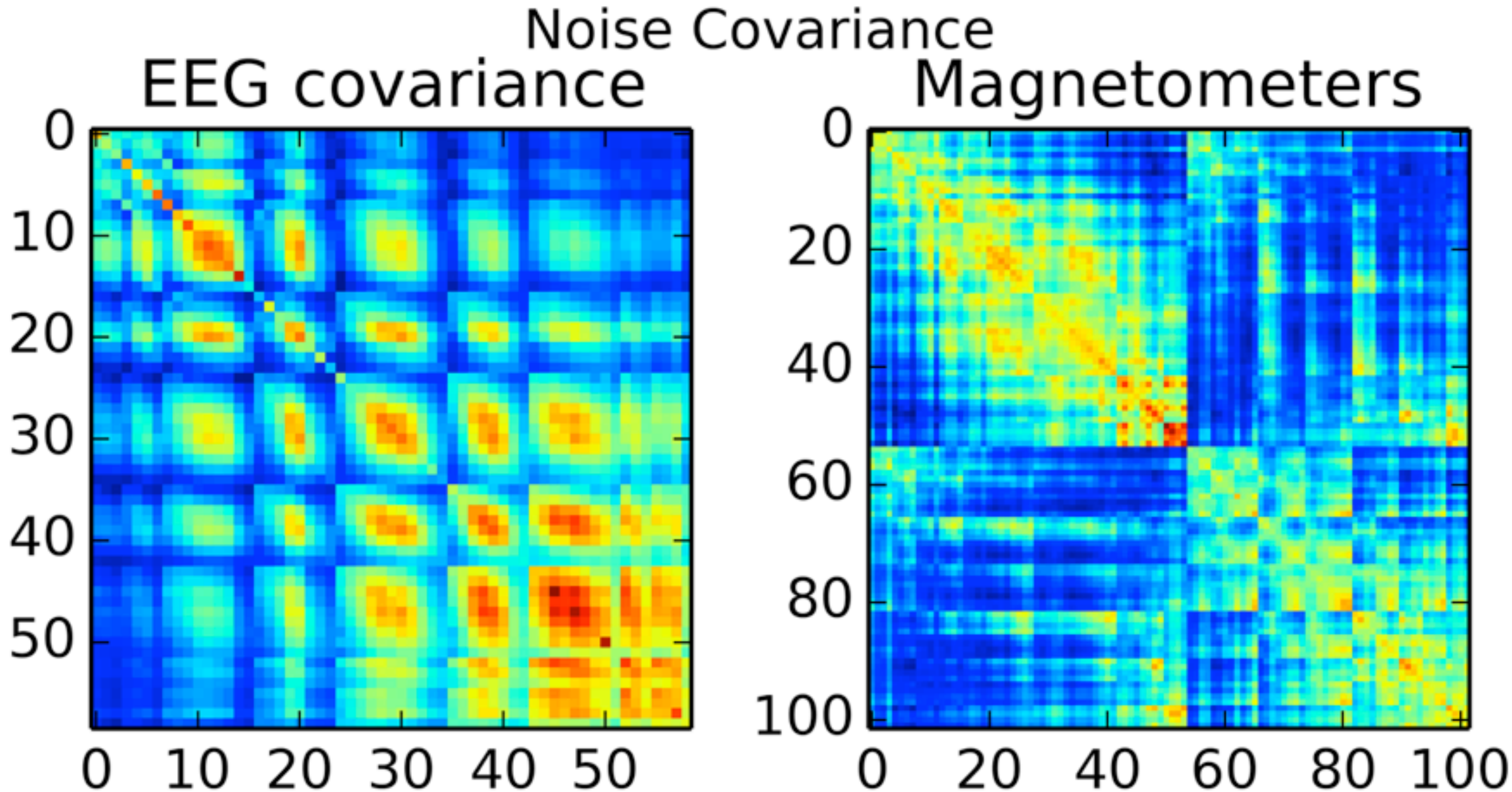
where $\|\mathbf{A}\|_F^2 = \mathbf{tr}(\mathbf{A}^T\mathbf{A})$

How do you whiten data?



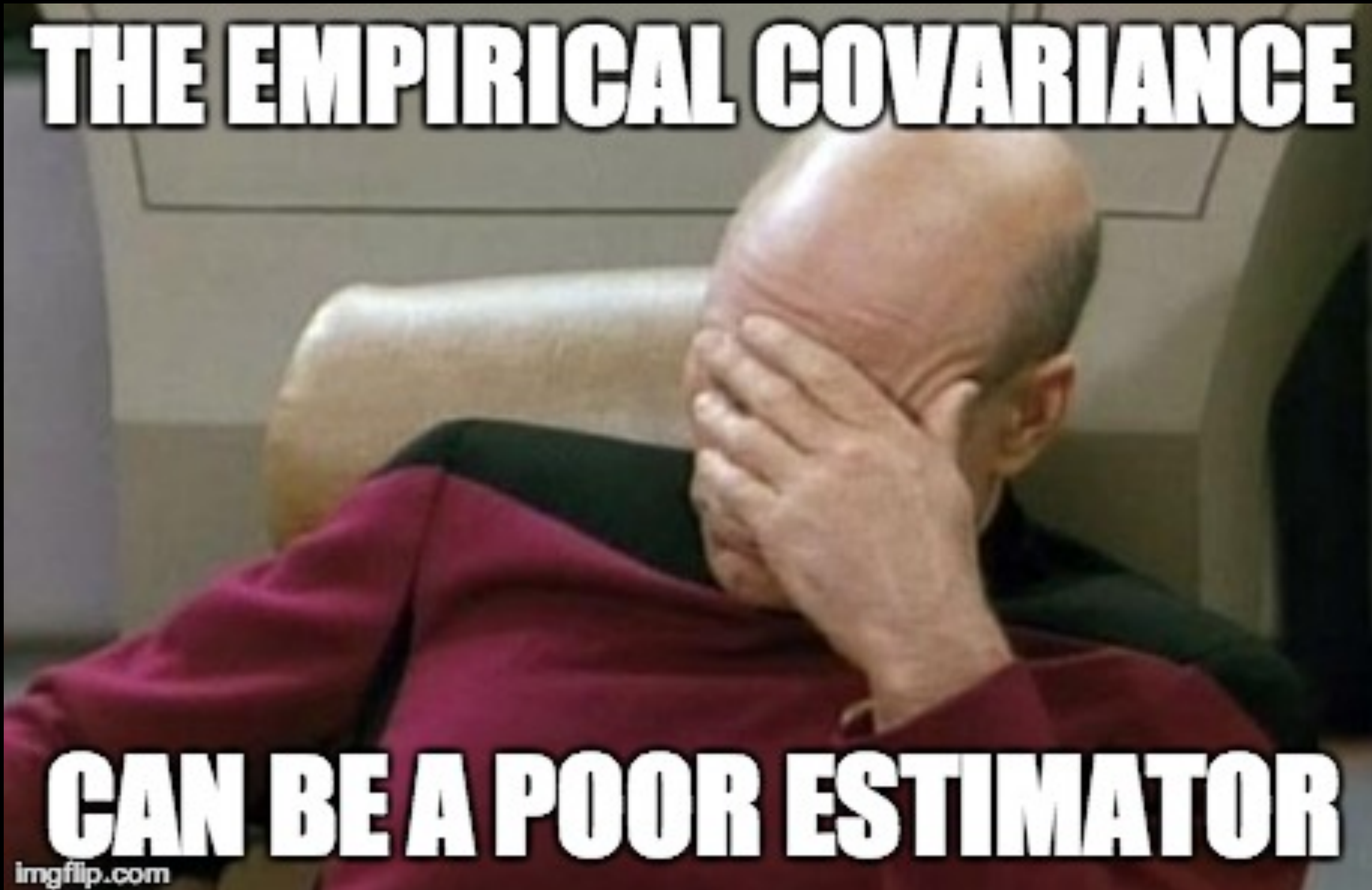
Baseline “noise”

Data from different sensors have to be put on the same scale.



$$C = \frac{1}{T} M M^t$$

With whitened data the covariance would be diagonal


THE EMPIRICAL COVARIANCE
CAN BE A POOR ESTIMATOR
imgflip.com

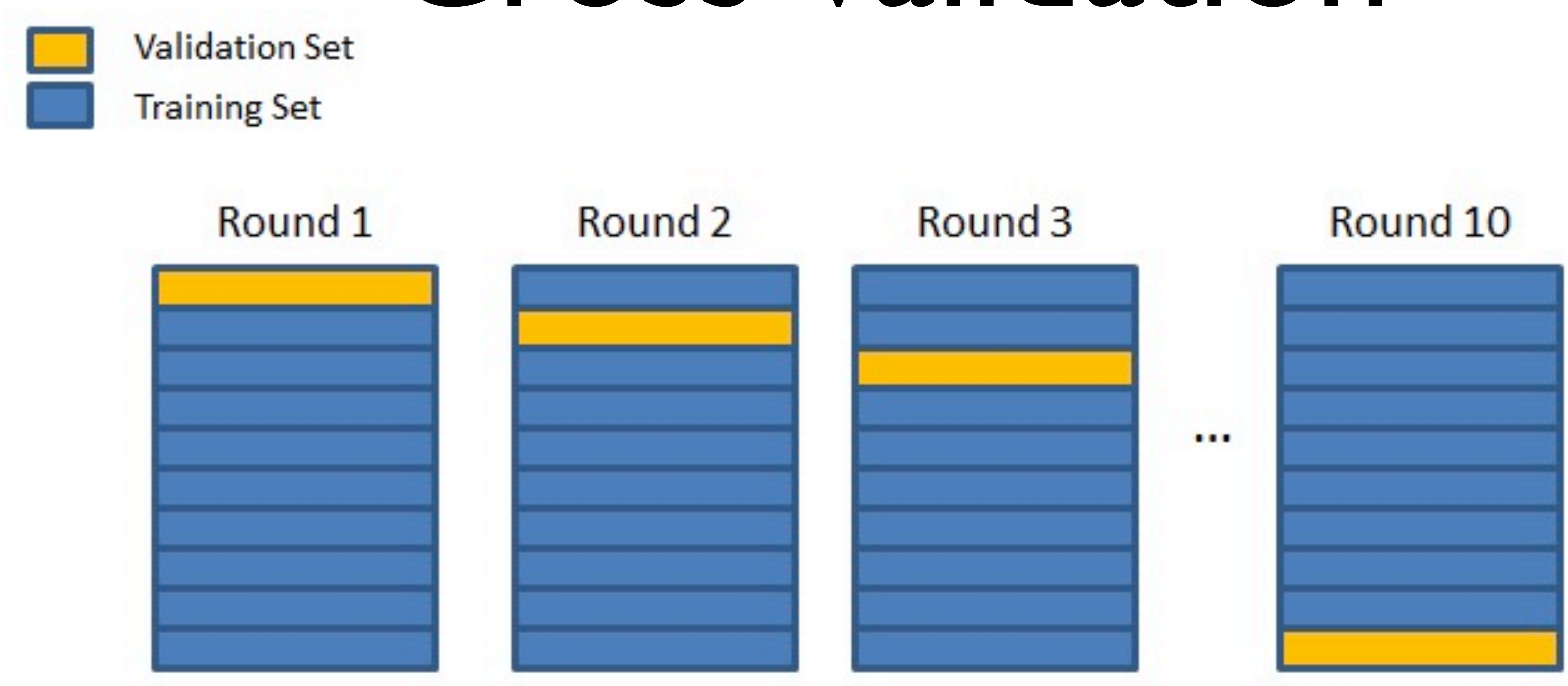
Model selection: Log-likelihood

Given my model C how likely are unseen data Y?

$$\mathcal{L}(Y|C) = -\frac{1}{2T}\mathrm{Trace}(YY^tC^{-1}) - \frac{1}{2}\log((2\pi)^N\det(C))$$

Higher log likelihood = better C & better whitening

Cross-validation



-1234.3 -1324.7 -1467.0 -1178.9

average log likelihood and select the best model

We compared 5 strategies:

simple, fast

1. Hand-set (REG)

$$C' = C + \alpha I, \quad \alpha > 0$$

2. Ledoit-Wolf (LW)

$$C_{LW} = (1-\alpha)C + \alpha\mu I \qquad \mu = \mathrm{mean}(\mathrm{diag}(C))$$

3. Cross-validated shrinkage (SC)

$$C_{SC} = (1-\alpha_{CV})C + \alpha_{CV}\mu I$$

4. Probabilistic PCA (PPCA)

$$C_{PPCA} = HH^t + \sigma^2 I_N$$

5. Factor Analysis (FA)

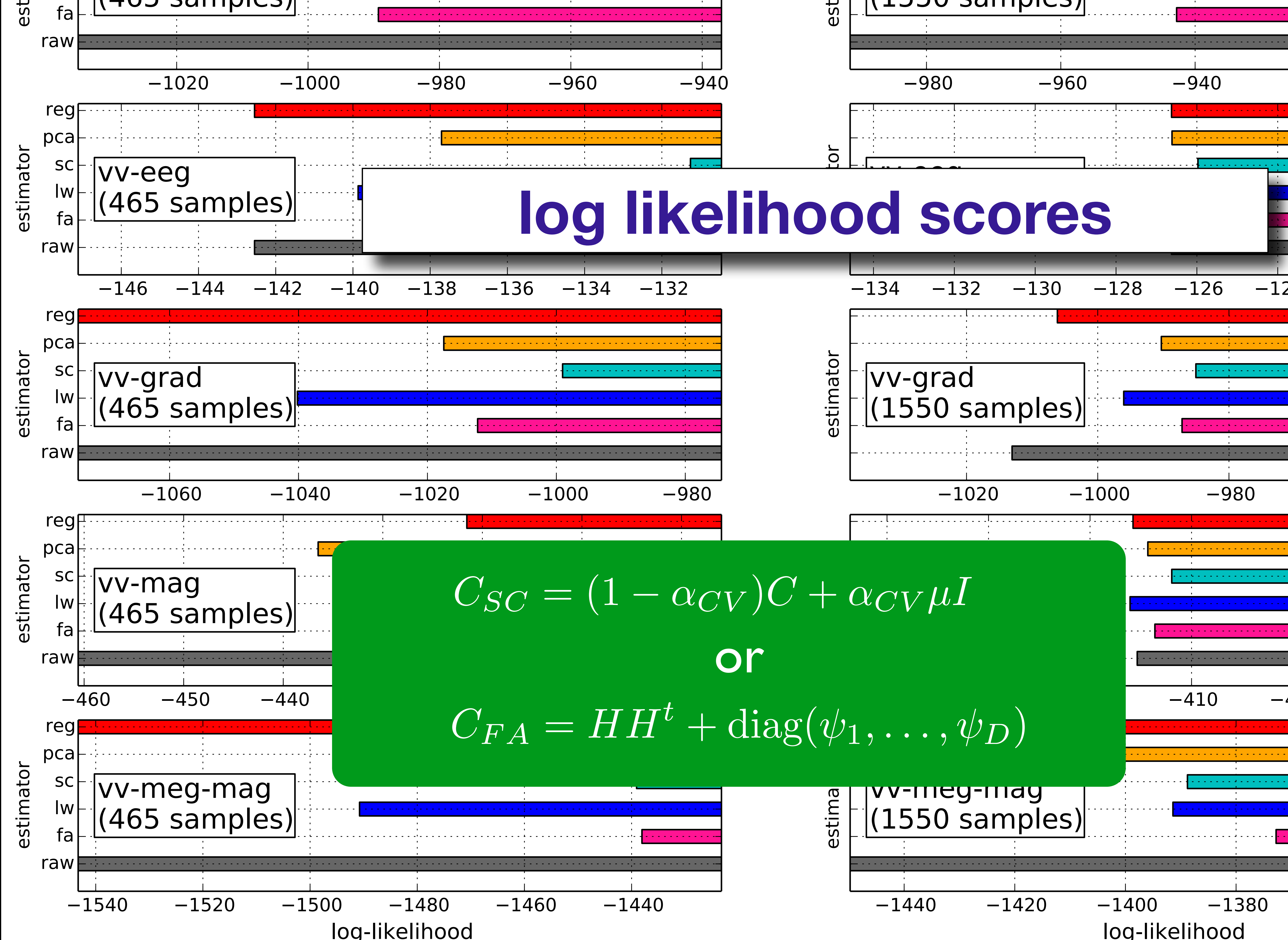
$$C_{FA} = HH^t + \mathrm{diag}(\psi_1, \ldots, \psi_D)$$

complex, slow

MEG and EEG data

key	**dataset and channel type**
bti-mag	4D Magnes 3600 WH magnetometers
ctf-mag	CTF-275 axial gradiometers
vv-eeg	VectorView EEG electrodes
vv-grad	VectorView planar gradiometers
vv-mag	VectorView magnetometers
vv-meg-grad	VectorView planar gradiometers, combined estimation
vv-meg-mag	VectorView magnetometers, combined estimation

log likelihood scores

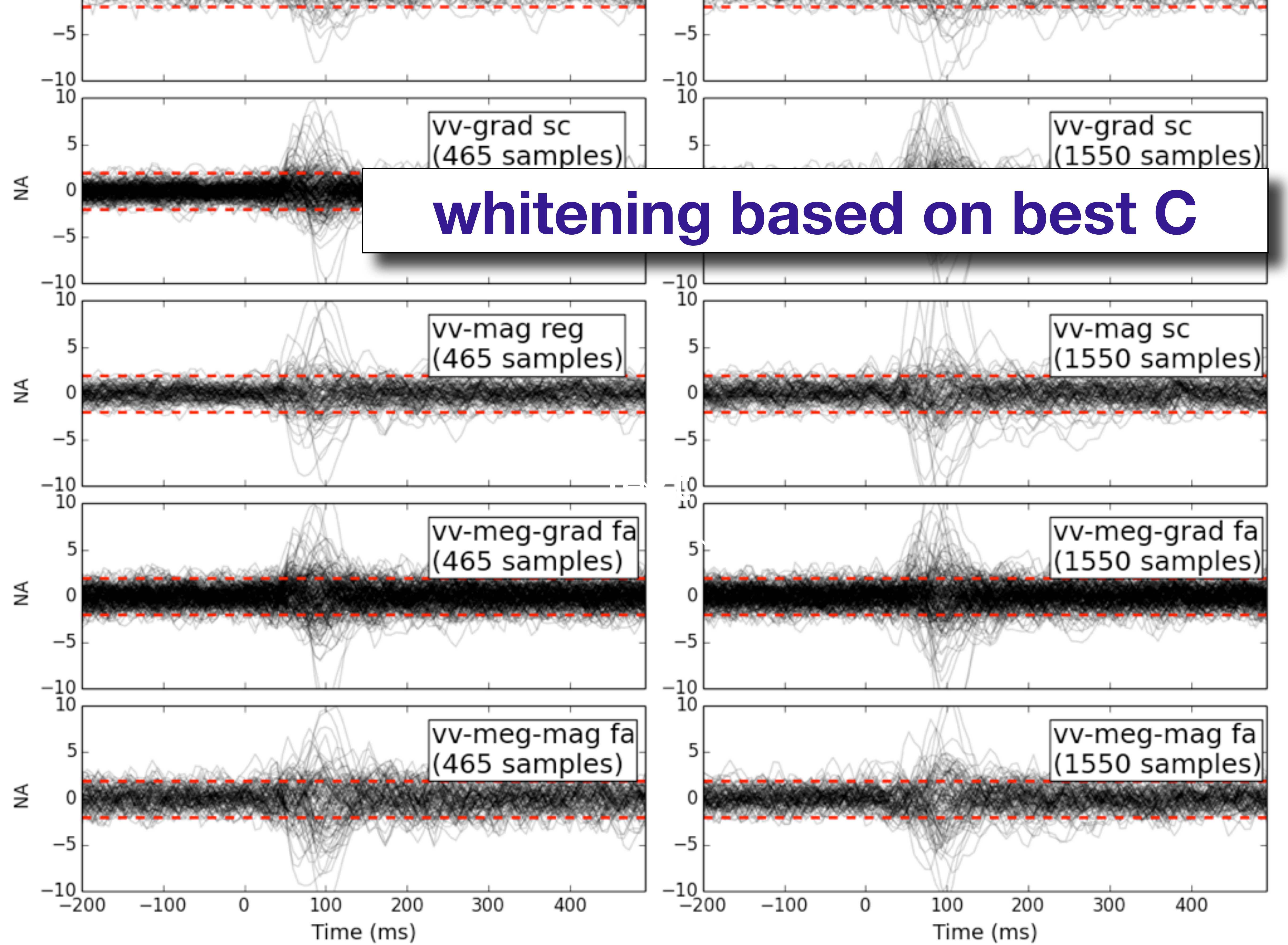


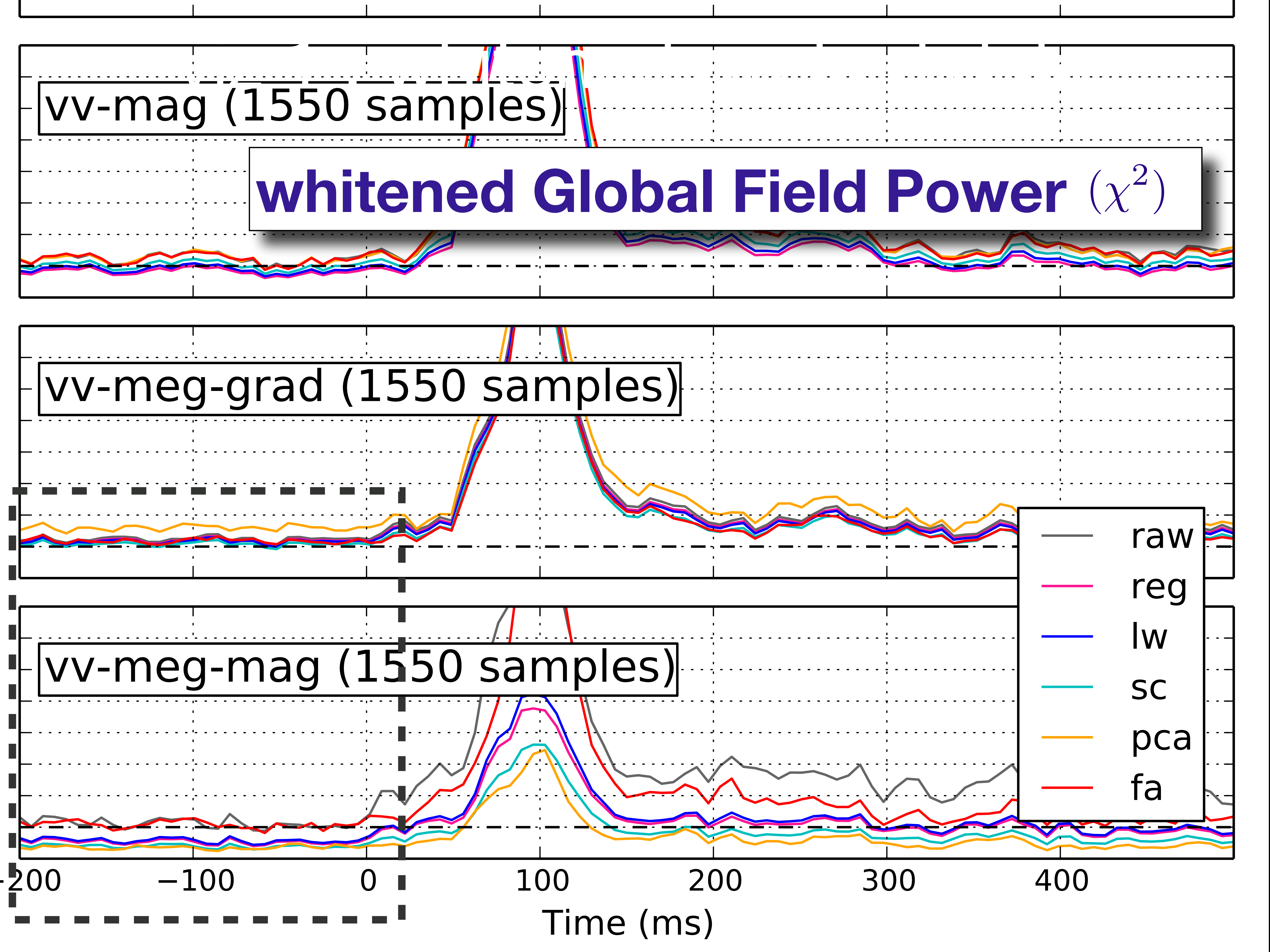
$$C_{SC} = (1 - \alpha_{CV})C + \alpha_{CV}\mu I$$

or

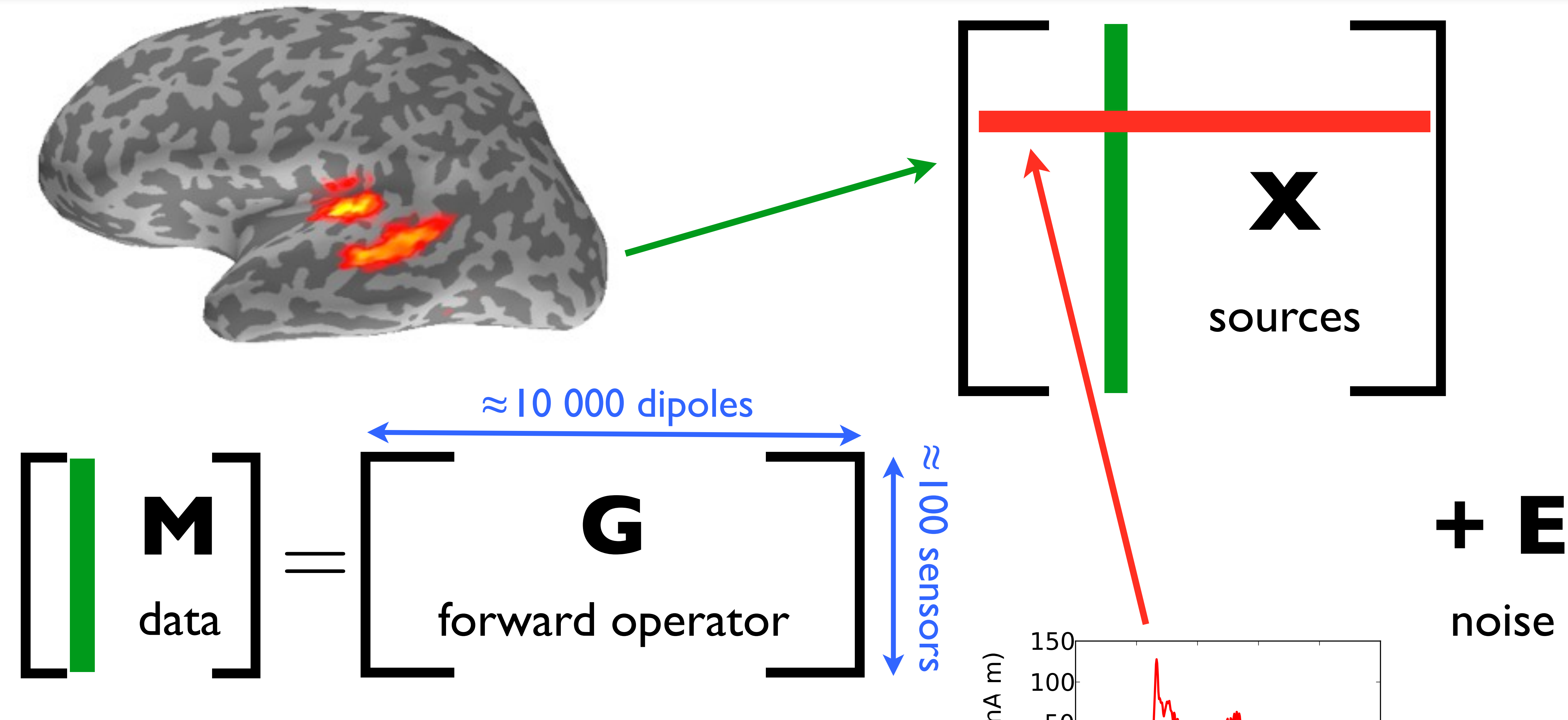
$$C_{FA} = HH^t + \mathrm{diag}(\psi_1, \ldots, \psi_D)$$

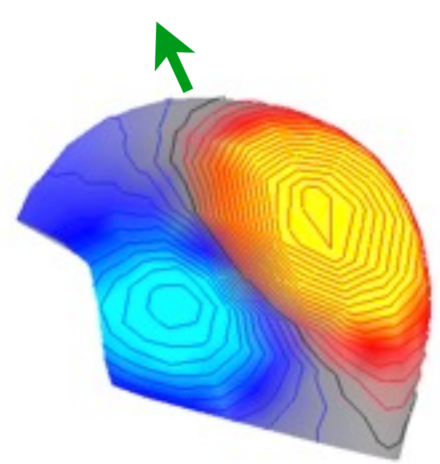
whitening based on best C




vv-mag (1550 samples)
whitened Global Field Power (χ^2)
vv-meg-grad (1550 samples)
vv-meg-mag (1550 samples)
raw
reg
lw
sc
pca
fa
−200
−100
0
100
200
300
400
Time (ms)

back to M = G X + E

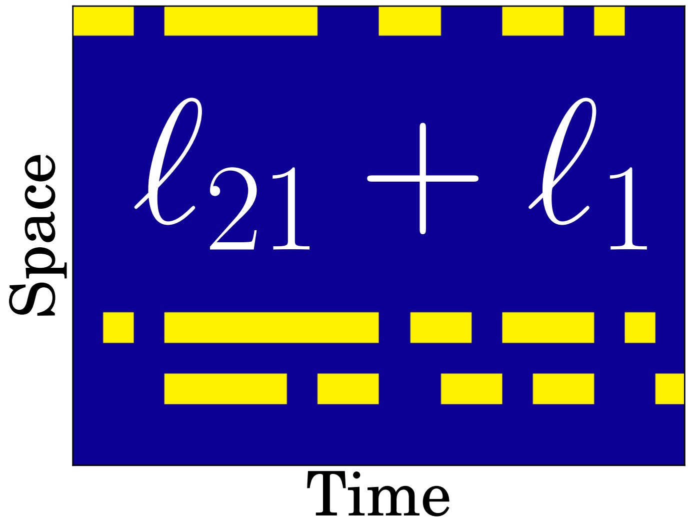




Objective: estimate X given M

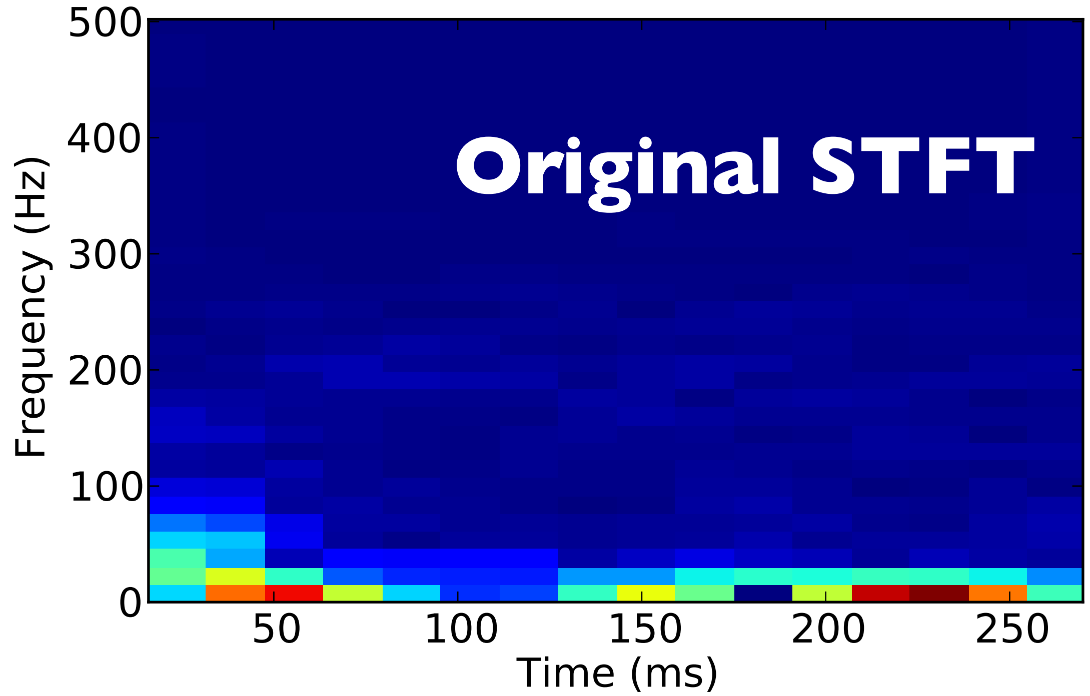
Challenge:

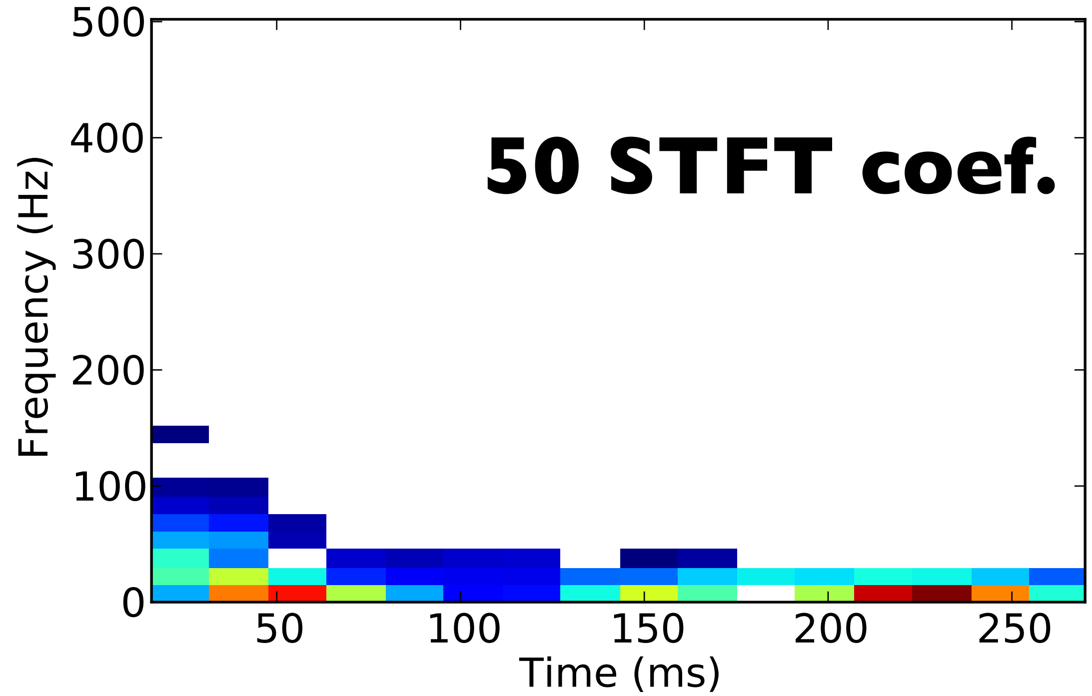
How do you promote sparse solutions with non-stationary sources?

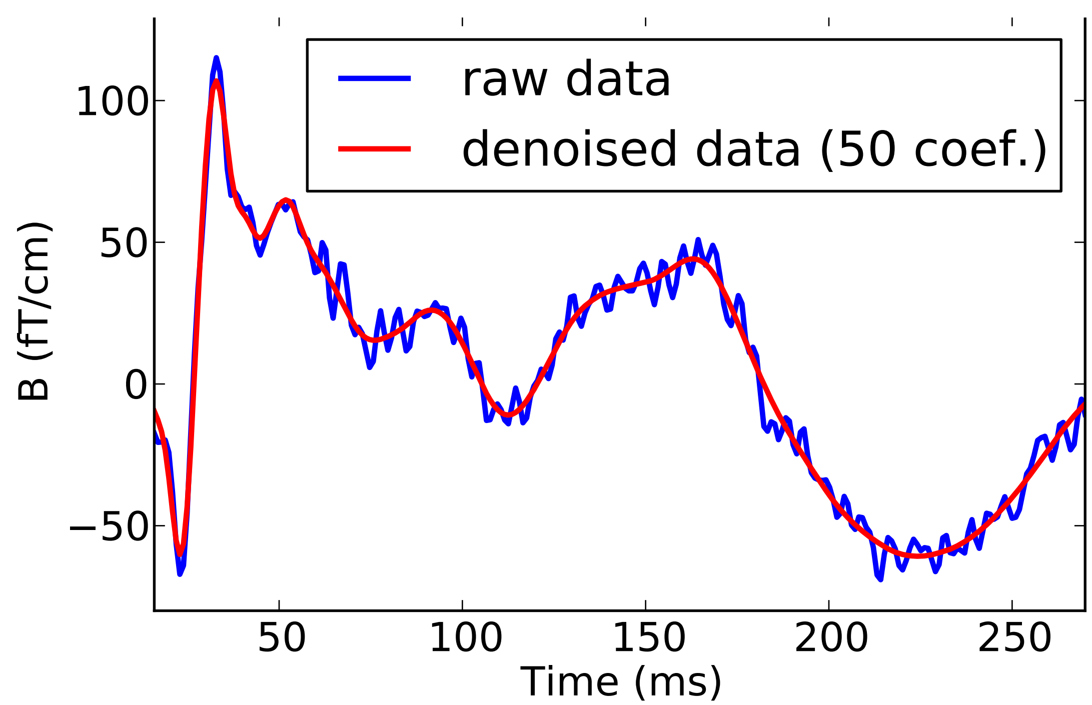


Support of **X**

Change the represensation



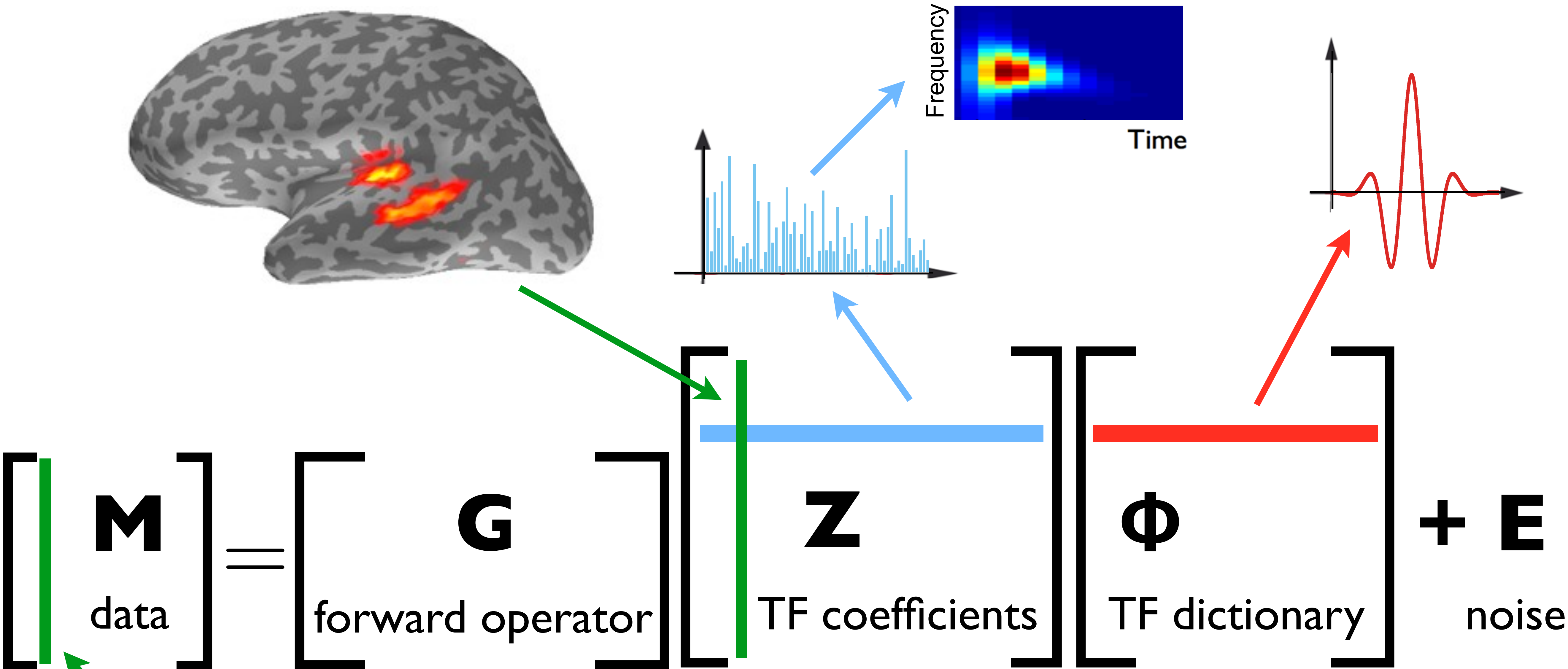


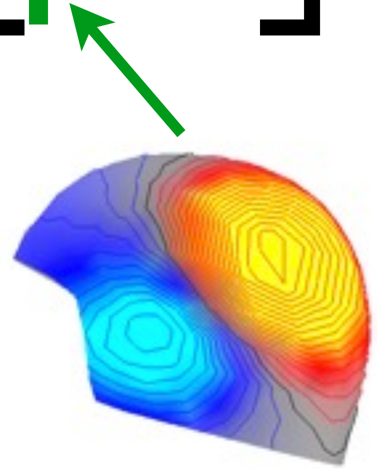


[“Wavelet shrinkage” Donoho & Johnstone 94]
[“Soft thresholding” Donoho 95]
[Application to evoked EEG, O. Bertrand et al. 94]
[Application to ST EEG, Quiroga et al. 03]
etc.

[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]:

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

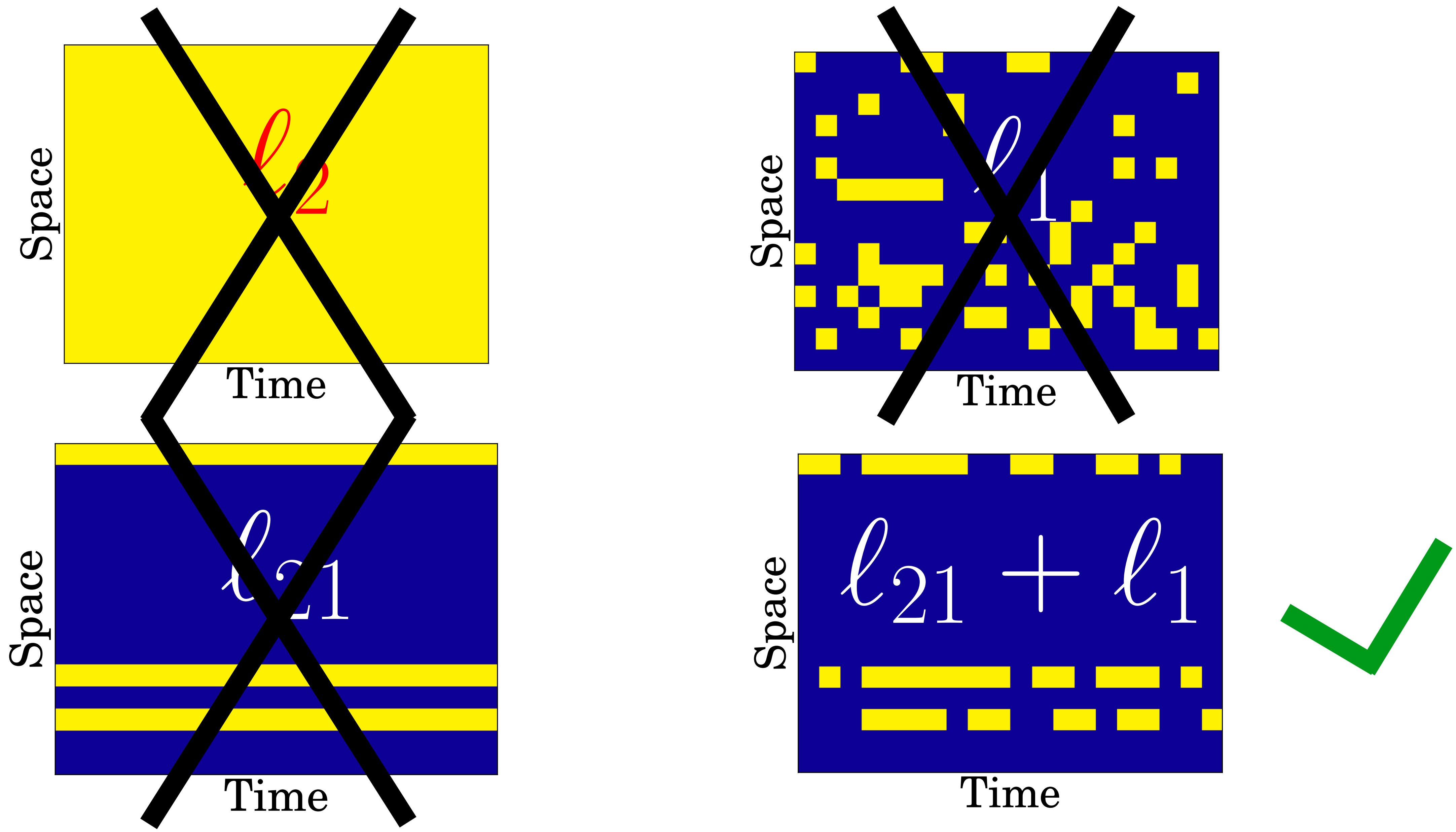
we propose:

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

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

Advantage:
localization in
space, time and frequency
in one step

What regularization?



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

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

Proximal gradient algorithm

Definition 1 (Proximity operator). *Let* $\varphi : \mathbb{R}^M \to \mathbb{R}$ *be a proper convex function. The proximity operator associated to* φ*, denoted by* $\text{prox}_\varphi : \mathbb{R}^M \to \mathbb{R}^M$ *reads:*

$$\text{prox}_\varphi(\mathbf{Z}) = \arg\min_{\mathbf{V}\in\mathbb{R}^M} \frac{1}{2}\|\mathbf{Z}-\mathbf{V}\|_2^2 + \varphi(\mathbf{V}) \ .$$

Lemma 1 (Proximity operator for $\ell_{21}+\ell_1$). *Let* $\mathbf{Y} \in \mathbb{C}^{P\times K}$ *be indexed by a double index* (p,k). $\mathbf{Z} = \text{prox}_{\lambda(\rho\|.\|_1+(1-\rho)\|.\|_{21})}(\mathbf{Y}) \in \mathbb{C}^{P\times K}$ *is given for each coordinates* (p,k) *by*

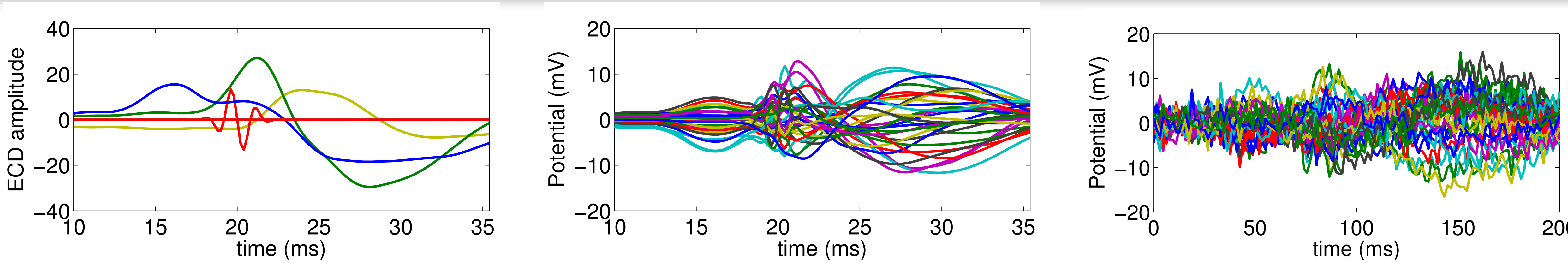
$$Z_{p,k} = \frac{Y_{p,k}}{|Y_{p,k}|}\left(|Y_{p,k}|-\lambda\rho\right)^+ \left(1 - \frac{\lambda(1-\rho)}{\sqrt{\sum_k (|Y_{p,k}|-\lambda\rho)^{+2}}}\right)^+ \ .$$

where for $x \in \mathbb{R}$, $(x)^+ = \max(x,0)$, *and by convention* $\frac{0}{0} = 0$.

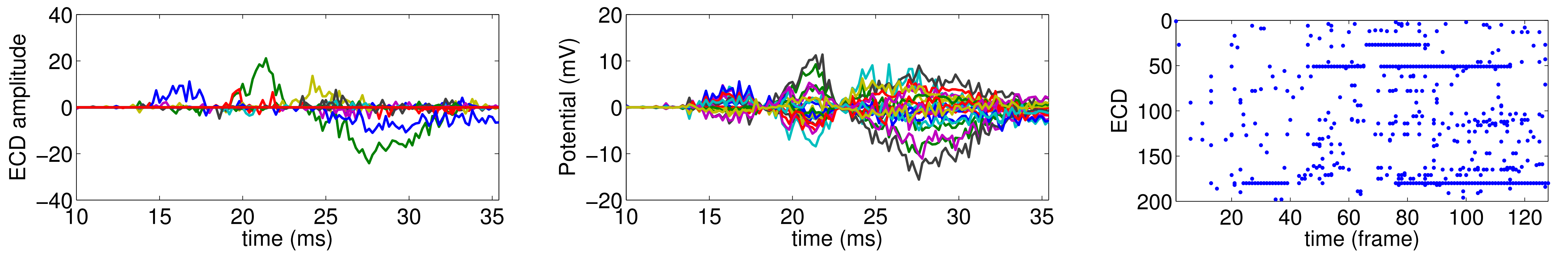
THM: It boils down to 2 successive thresholdings

[Jenatton et al. 2011, Gramfort et al. IPMI 2011]

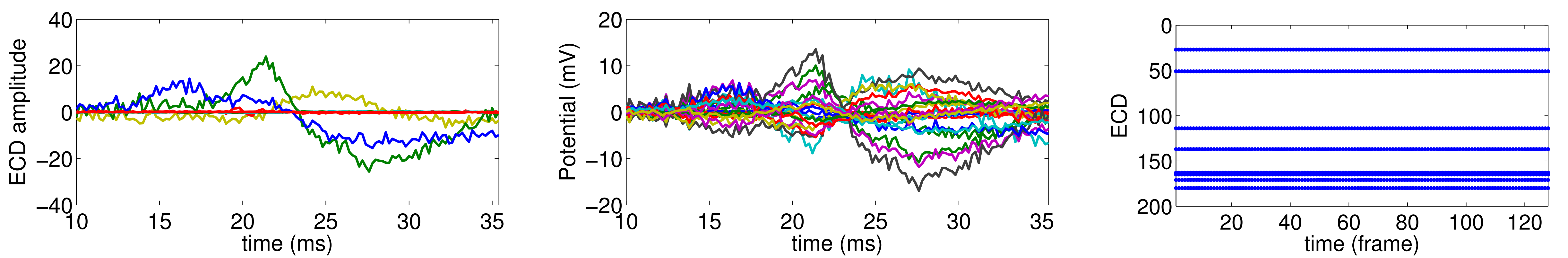
Simulation results (part 1)



(a) $\mathbf{X}$ ground truth (b) $\mathbf{M}$ noiseless (c) $\mathbf{M}$ noisy (SNR=6dB)

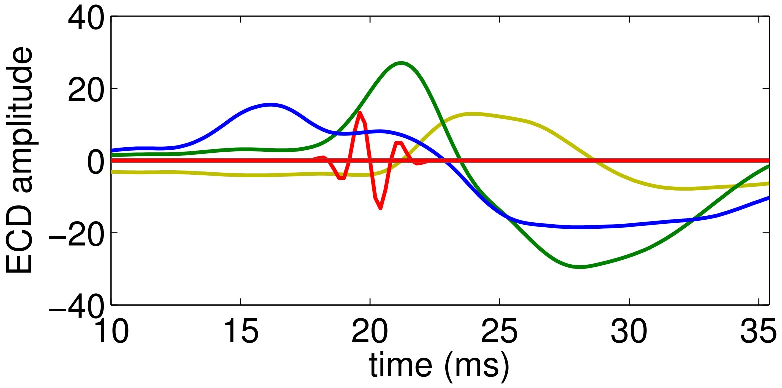


(d) $\mathbf{X}_{\ell_1}$ (e) $\mathbf{M}^\star_{\ell_1} = \mathbf{G}\mathbf{X}^\star_{\ell_{21}}$ (f) Non-zeros of $\mathbf{X}^\star_{\ell_1}$

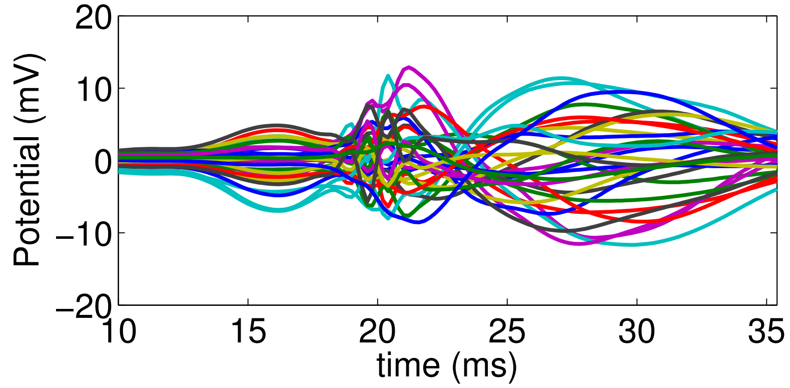


(g) $\mathbf{X}^\star_{\ell_{21}}$ (h) $\mathbf{M}^\star_{\ell_{21}} = \mathbf{G}\mathbf{X}^\star_{\ell_{21}}$ (i) Non-zeros of $\mathbf{X}^\star_{\ell_{21}}$

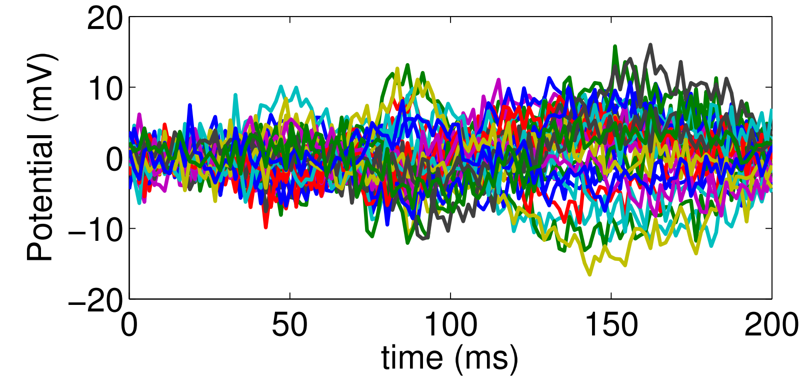
Simulation results (part 2)



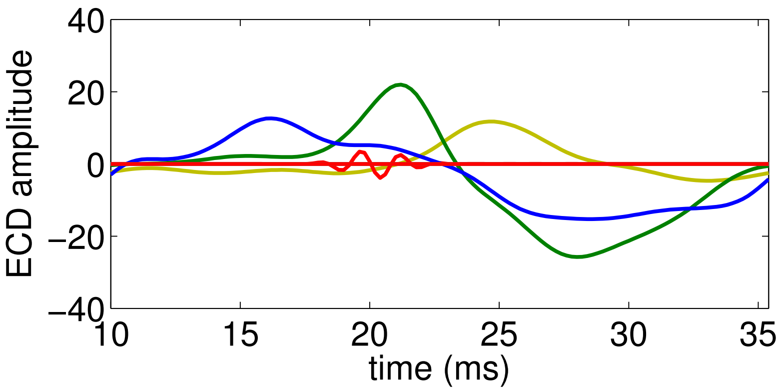
(a) $\mathbf{X}$ ground truth



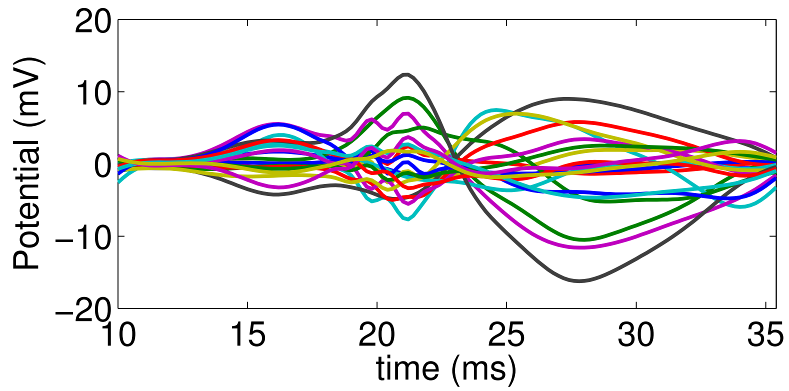
(b) $\mathbf{M}$ noiseless



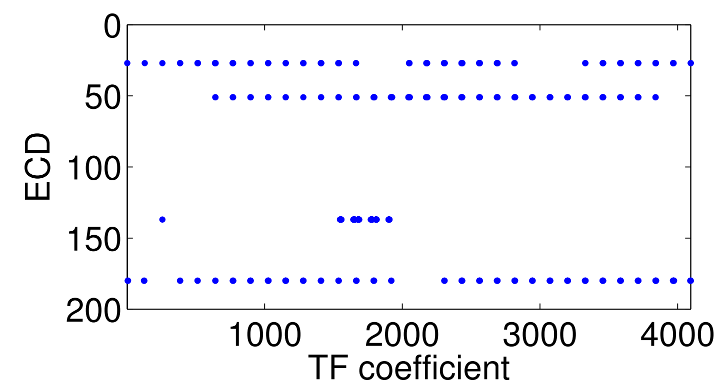
(c) $\mathbf{M}$ noisy (SNR=6dB)



(j) $\mathbf{X}^{\star}_{\mathrm{TF}\ \ell_{21}+\ell_1}$



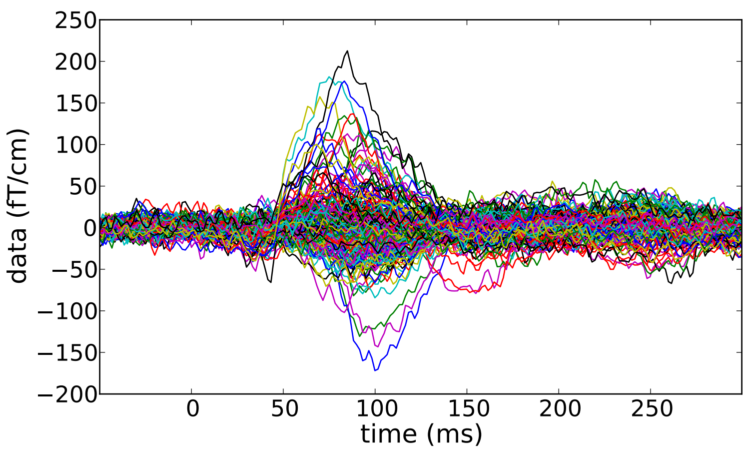
(k) $\mathbf{M}^{\star}_{\mathrm{TF}\ \ell_{21}+\ell_1}$



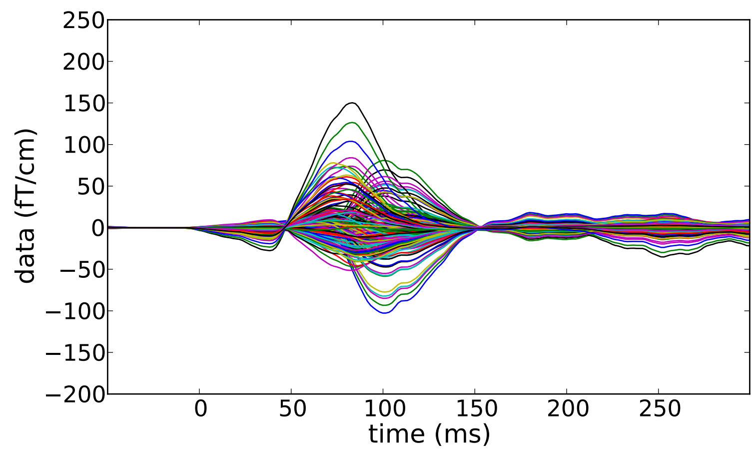
(l) Non-zeros of $\mathbf{Z}^{\star}_{\ell_{21}+\ell_1}$

MEG Auditory data

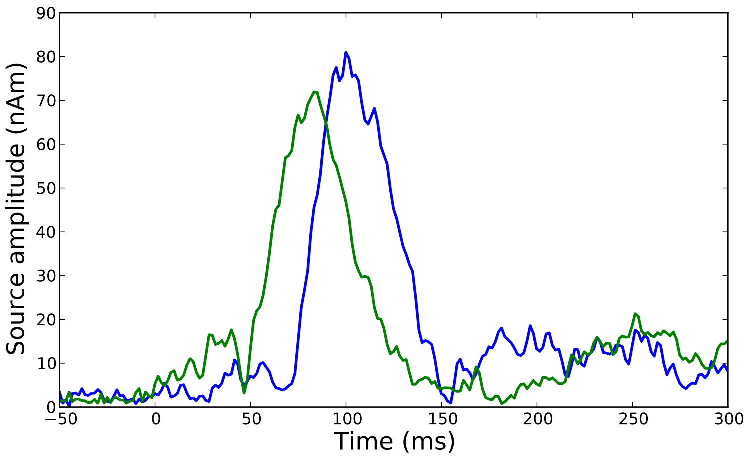
Protocol: 50 epochs of auditory tones in left ear (305 MEG, 59 EEG channels)



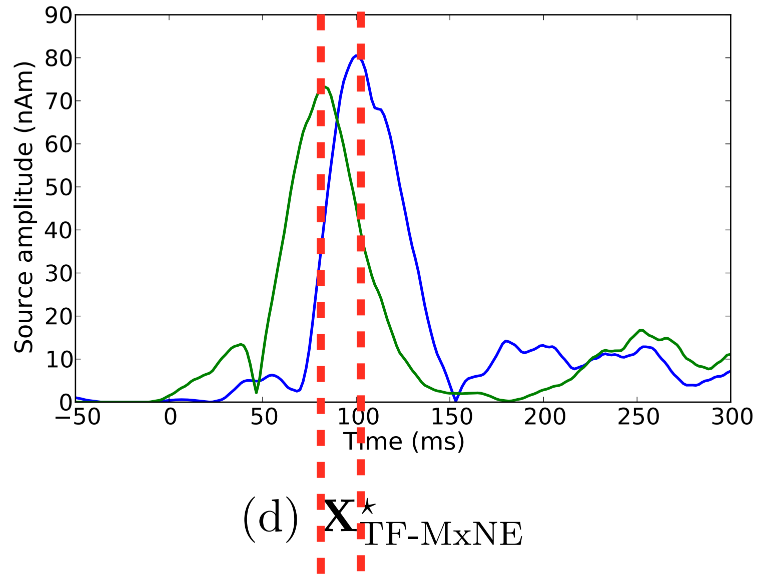
(a) MEG data (Gradiometers only)



(b) $\mathbf{GX}^{\star}_{\text{TF-MxNE}}$ (explained data)

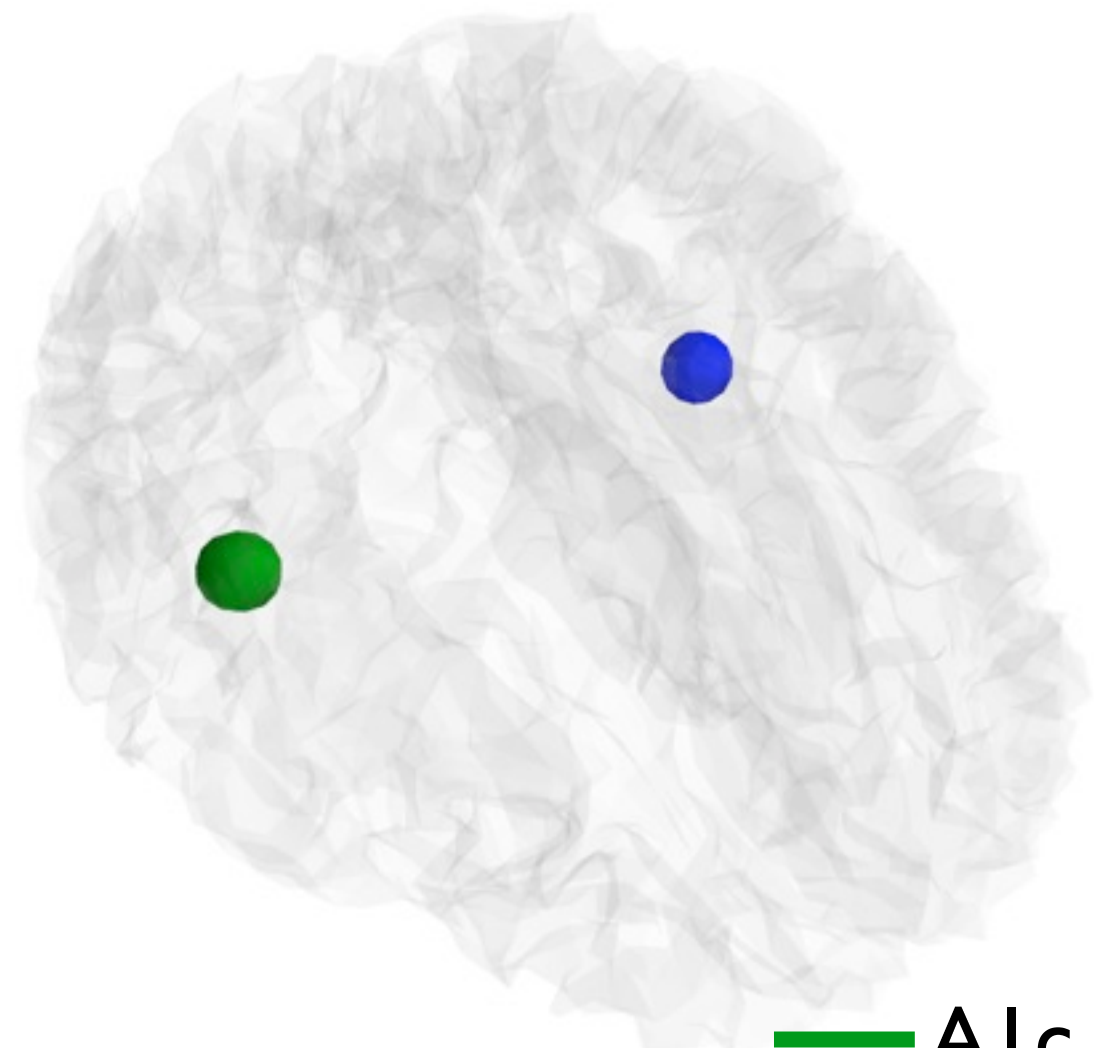


(c) $\mathbf{X}^{\star}_{\text{MxNE}}$



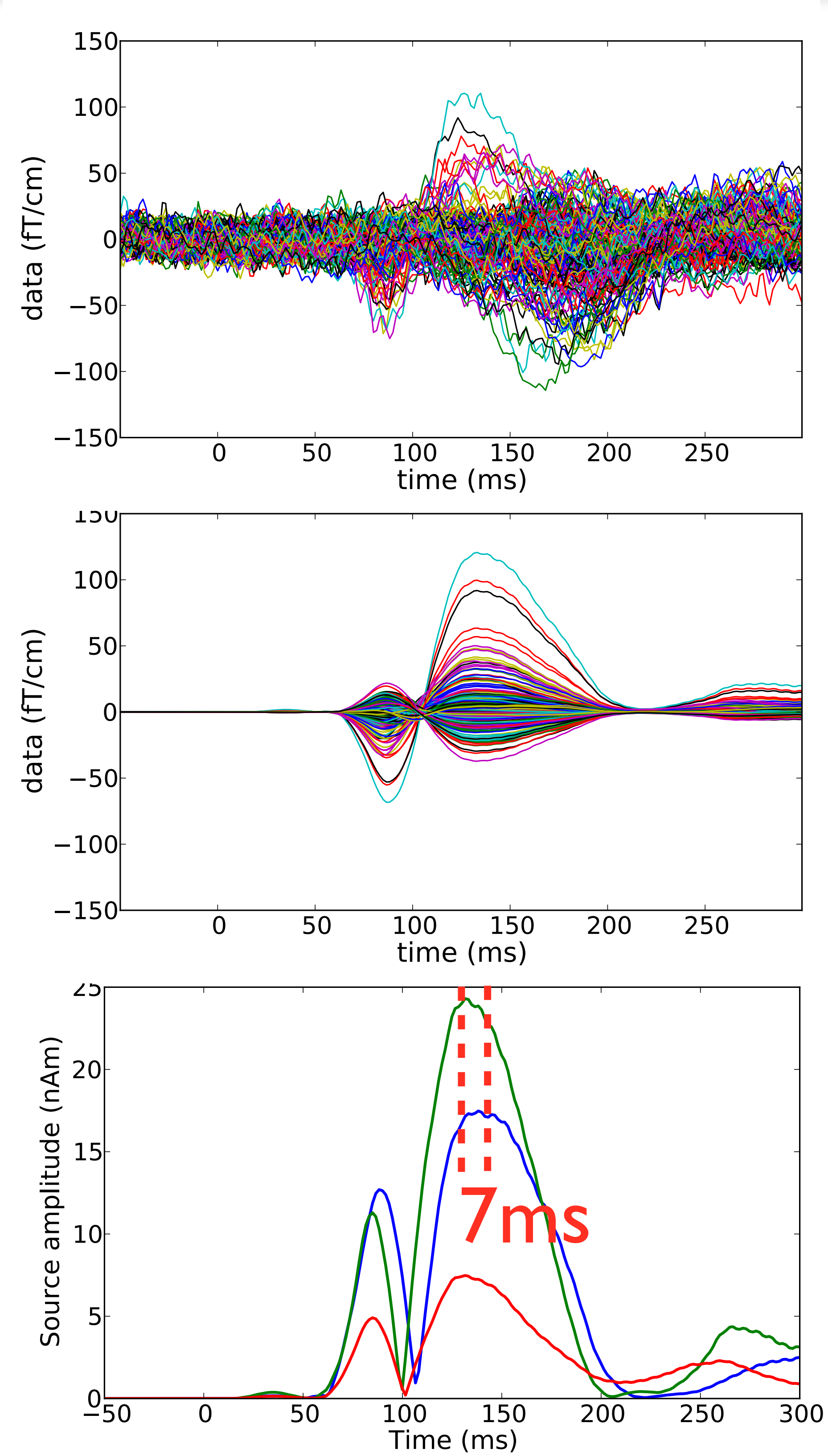
(d) $\mathbf{X}^{\star}_{\text{TF-MxNE}}$

16ms Chronometry

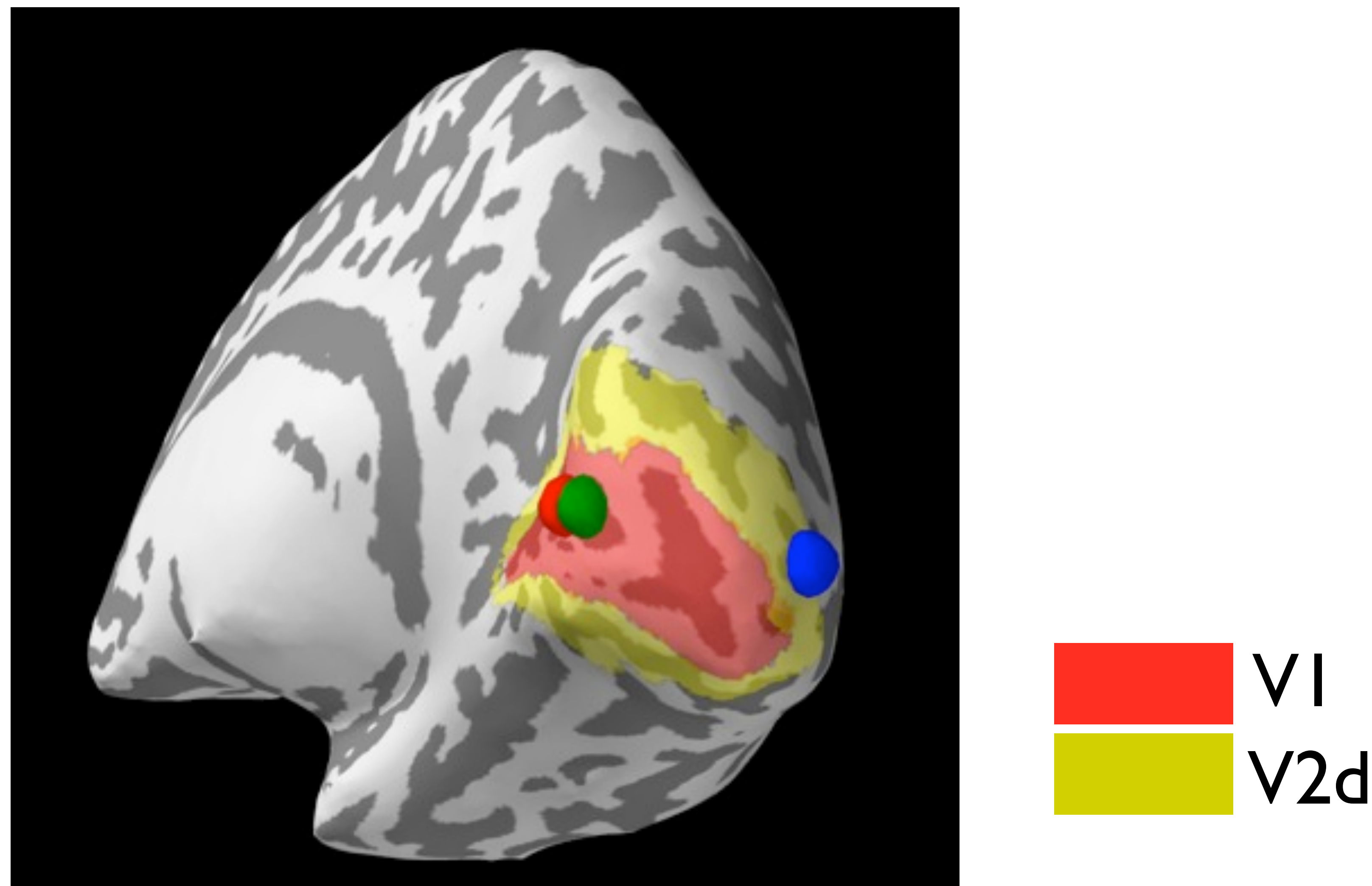


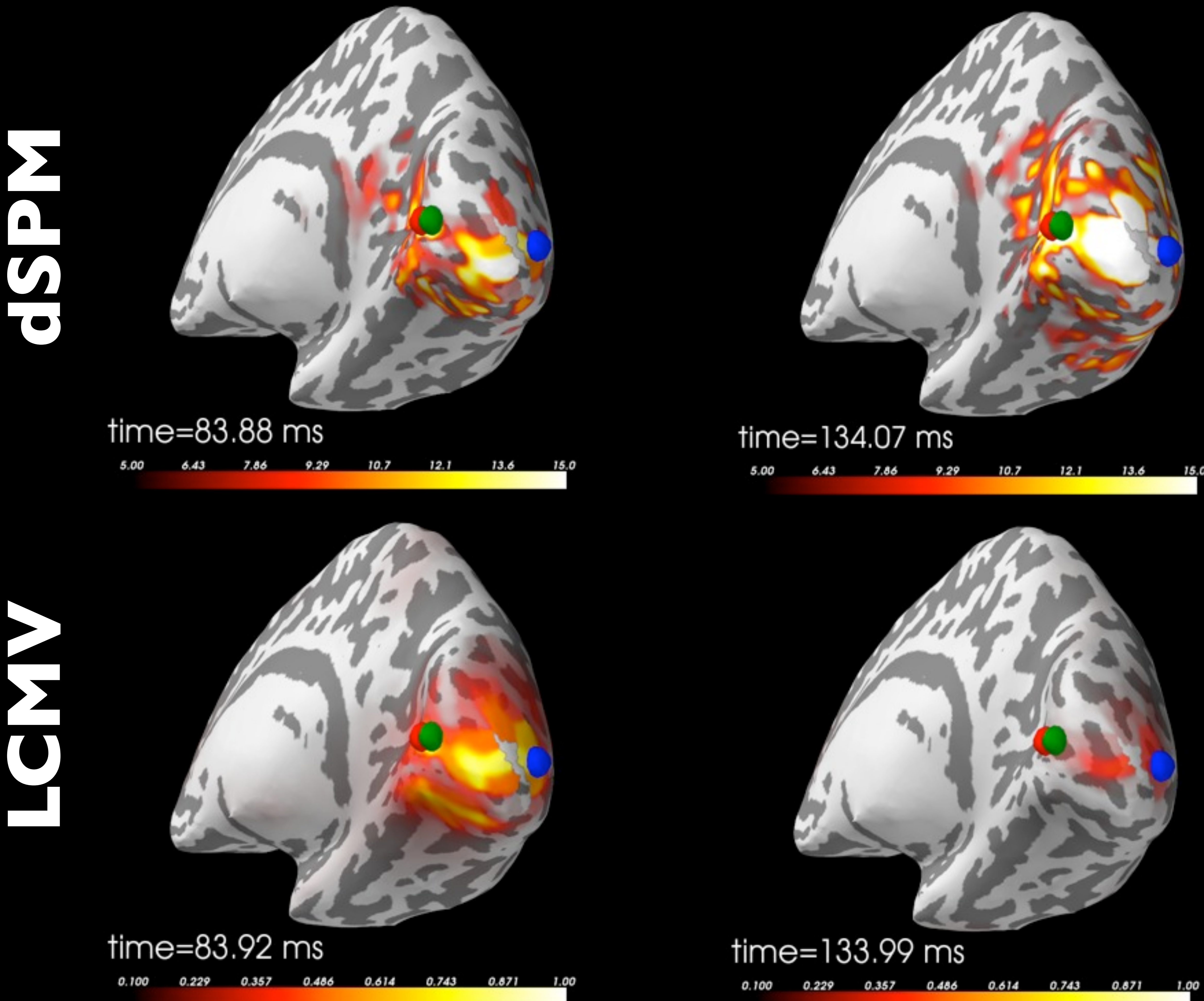
AIc
AIi

MEG Visual data



Protocol: 50 epochs of visual flash in left hemi-field (305 MEG, 59 EEG channels)




dSPM
LCMV
time=83.88 ms
5.00 6.43 7.86 9.29 10.7 12.1 13.6 15.0
time=134.07 ms
5.00 6.43 7.86 9.29 10.7 12.1 13.6 15.0
time=83.92 ms
0.100 0.229 0.357 0.486 0.614 0.743 0.871 1.00
time=133.99 ms
0.100 0.229 0.357 0.486 0.614 0.743 0.871 1.00

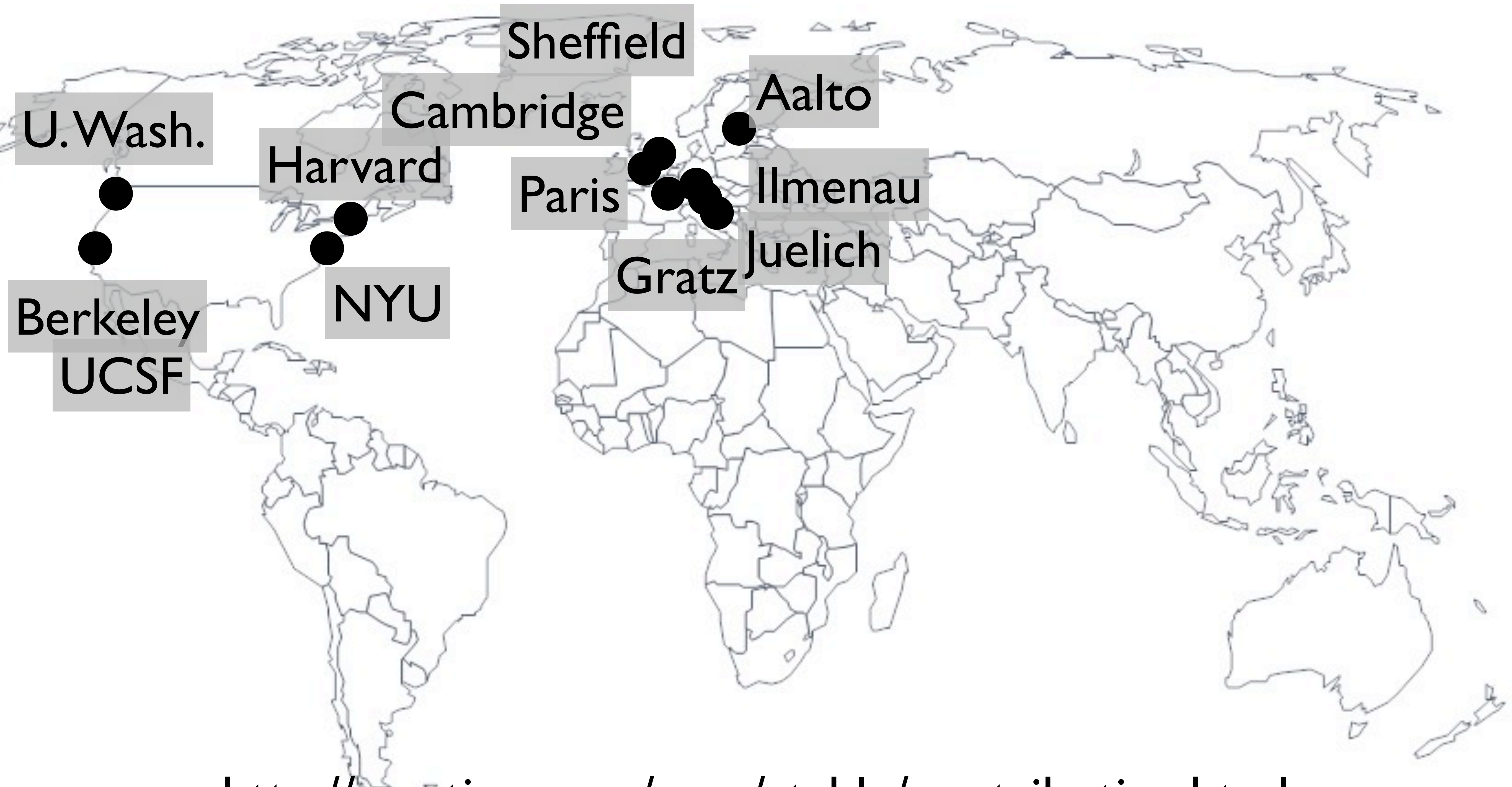
MNE Software for MEG and EEG

http://www.martinos.org/mne

http://www.github.com/mne-tools

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

Development of the MNE software



http://martinos.org/mne/stable/contributing.html

References

1 post-doc position

Gramfort et al., *Mixed-norm estimates for the M/EEG inverse proble[...] 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

Engemann, D.A., Gramfort, A.. Automated model selection in covariance estimation and spatial whitening of MEG and EEG signals. (submitted)

Collaborators:

M. Hämäläinen, MGH / Harvard / MIT, Boston

D. Strohmeier & J. Haueisen, TU Ilmenau, Germany

D.A.Engemann, CEA Neurospin, ICM, Paris, France

M. Kowalski, L2S, Univ. Paris-sud