Tutorial on EEG/MEG inverse source reconstruction

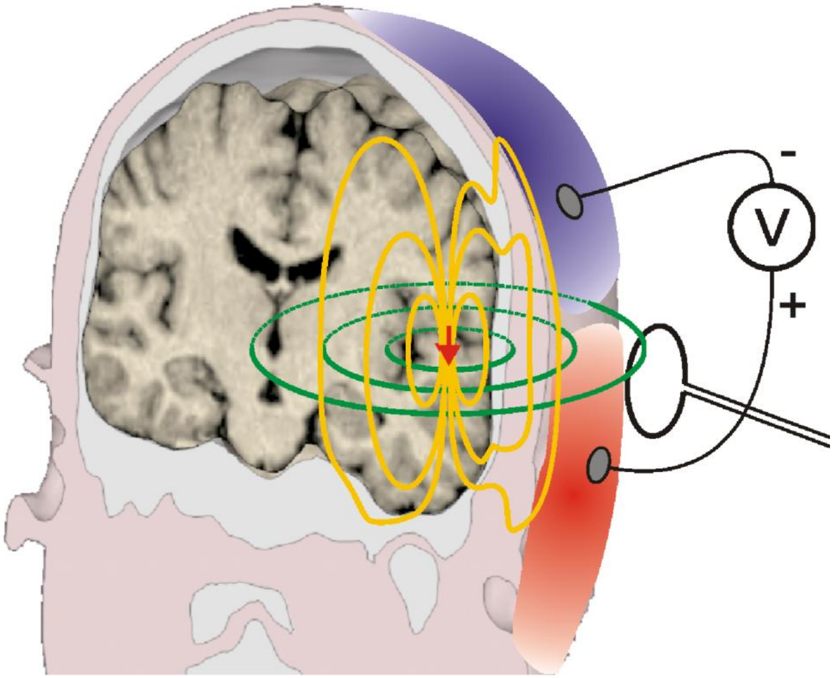
Guido Nolte and Stefan Haufe

Part 1: dipole fits and distributed inverse solutions

Part 2: beamformers and subspace methods

BBCI Summer School 2012, Berlin

EEG and MEG



Cellular currents due to synchronous firing of a large population of equally spatially aligned neurons give rise to extracranial electric potentials and magnetic fields.

Electroencephalography (EEG)
= measuring the potential differences on the scalp

Magnetoencephalography (MEG)
= measuring the extracranial magnetic fields

MEG and EEG are different views of the same neural sources.

Slide by Lauri Parkkonen

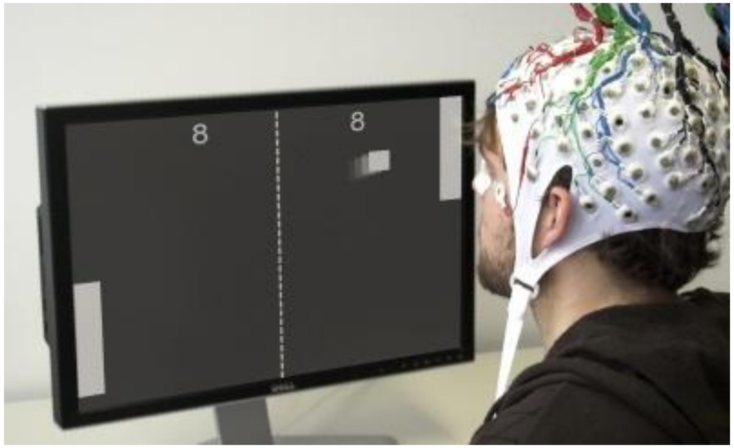
EEG and MEG

Advantages:

- Noninvasive
- High temporal resolution
- Portable (EEG)
- Low cost (EEG)

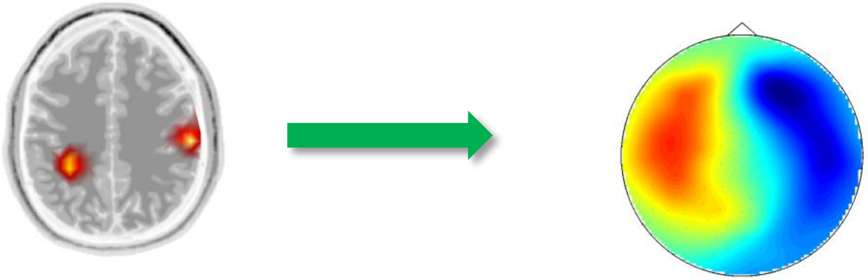
Disadvantage: only indirect measurements

- Low signal-to-noise ratio
- Difficult to interpret



Volume conduction

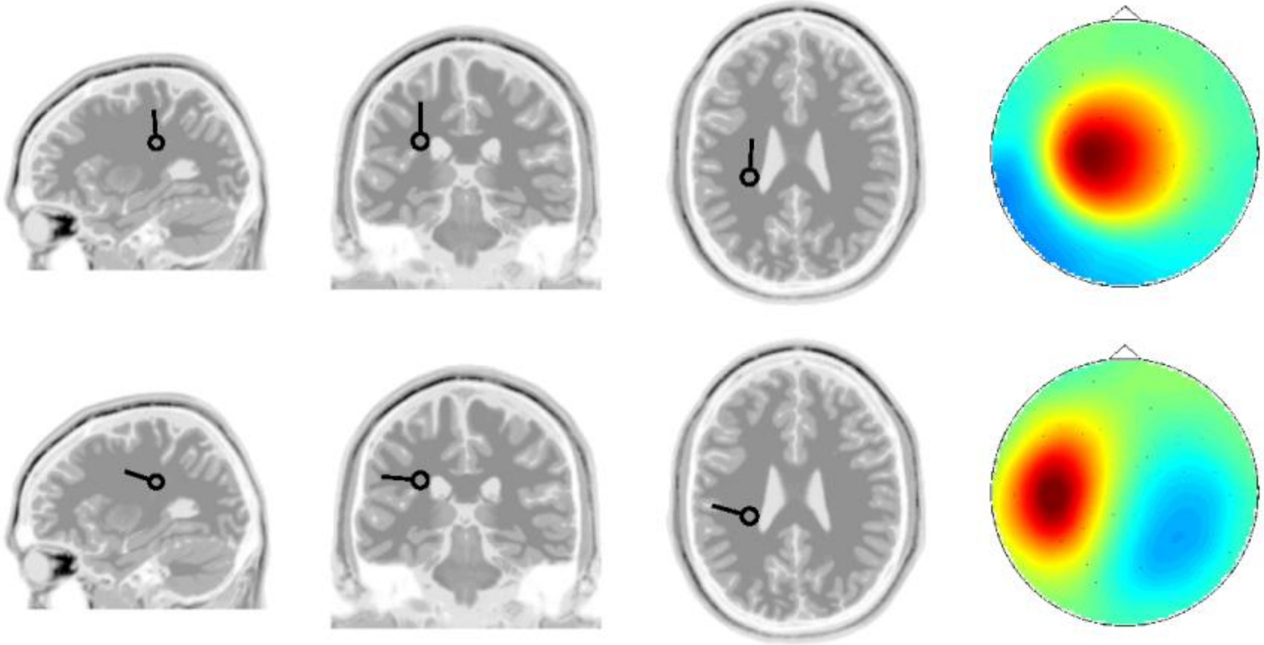
Problem: Spatial smearing of the source activity in the sensors due to the propagation of el. currents/magnetic fields in the head tissue.



Volume conduction should explicitly be modeled in order to

- Improve signal-to-noise ratio (e.g., in BCIs)
- Interpret the results, localize the sources

Interpretability

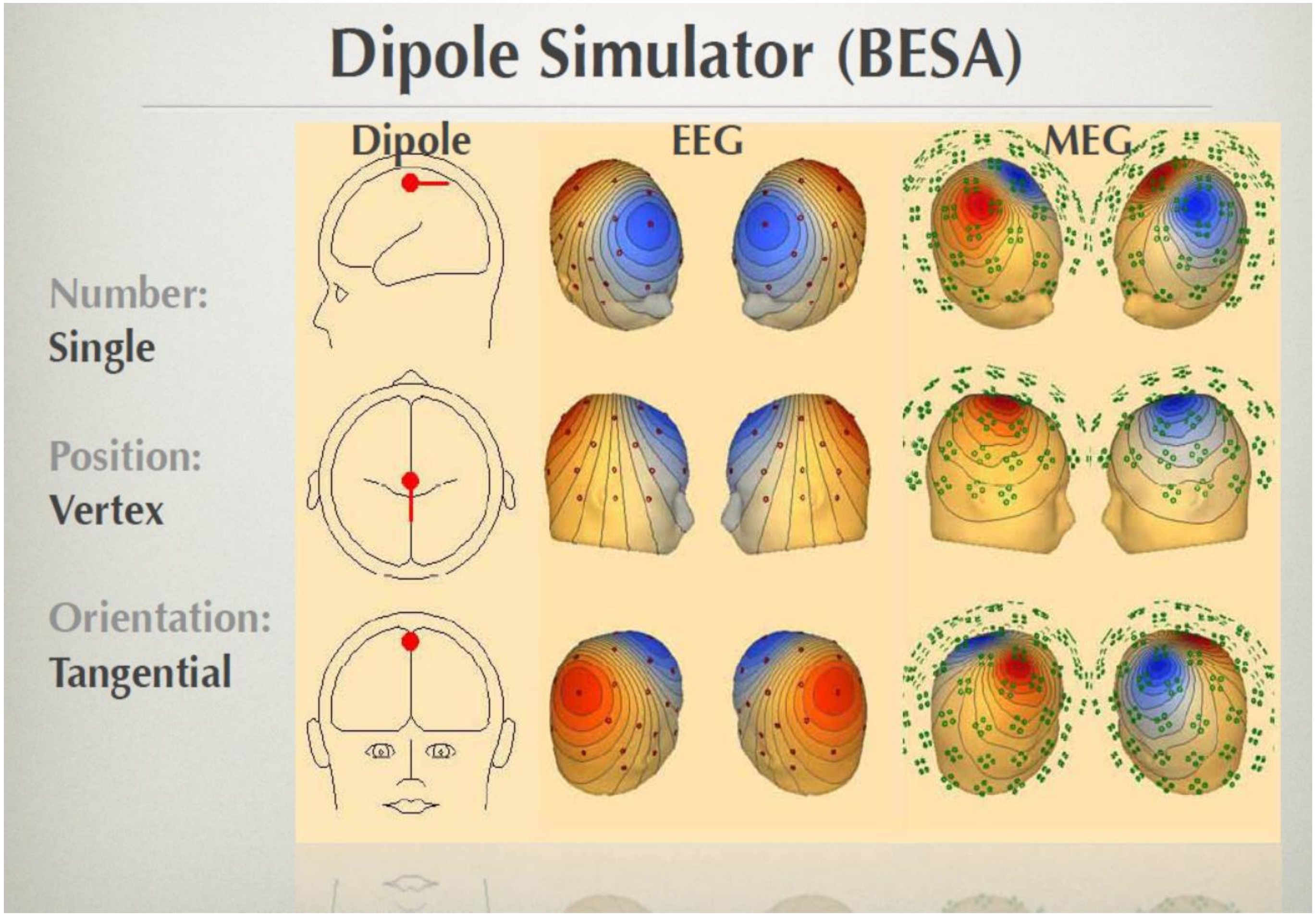


source electrical dipole

EEG scalp potential

Direction of dipole current (determined by the local curvature of the cortex) has more influence than location.

Characteristics of the EEG and MEG generation



Dipole Simulator (BESA)

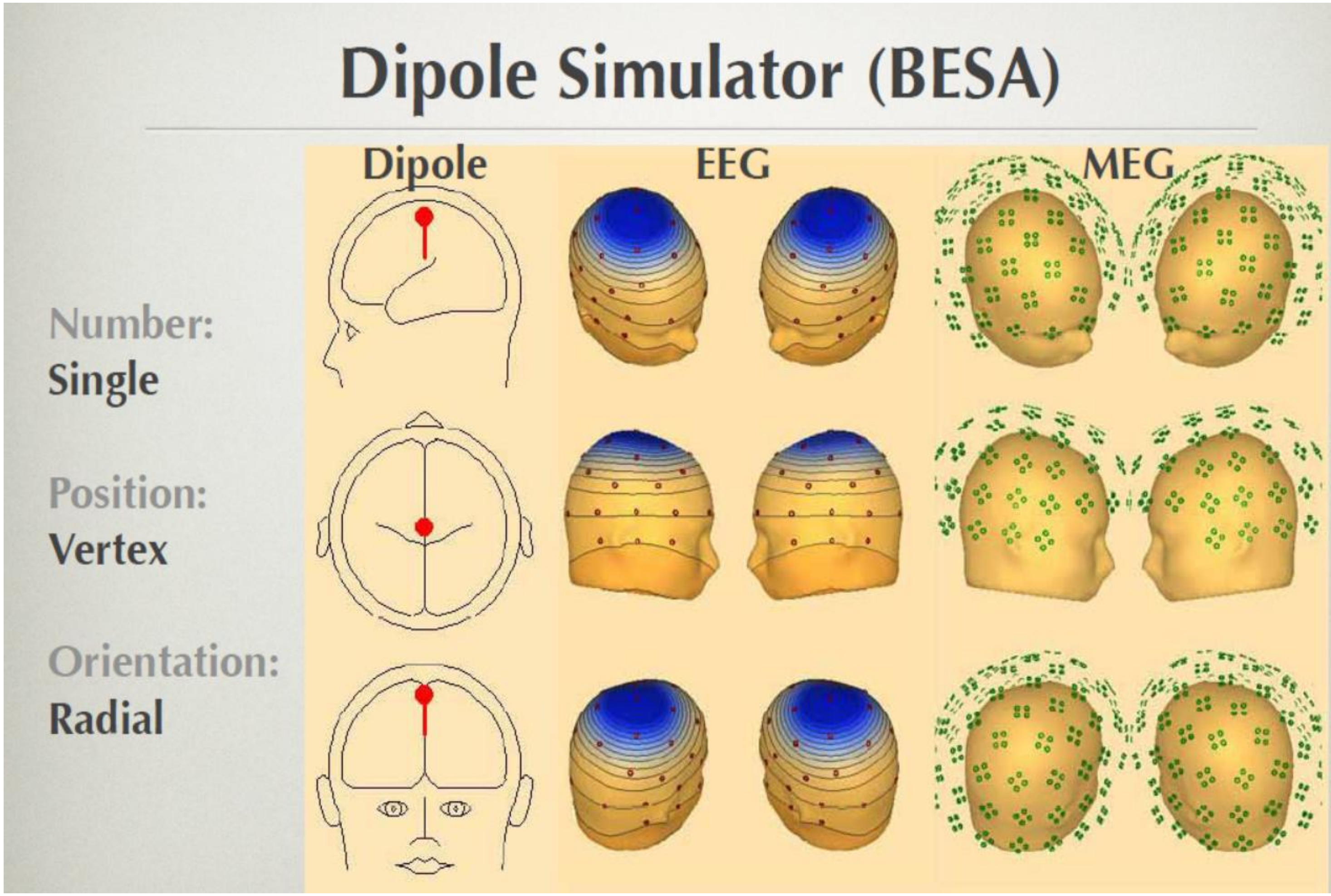
Number: Single

Position: Vertex

Orientation: Tangential

slide from: Nobukazu Nakasato, 2009

Characteristics of the EEG and MEG generation



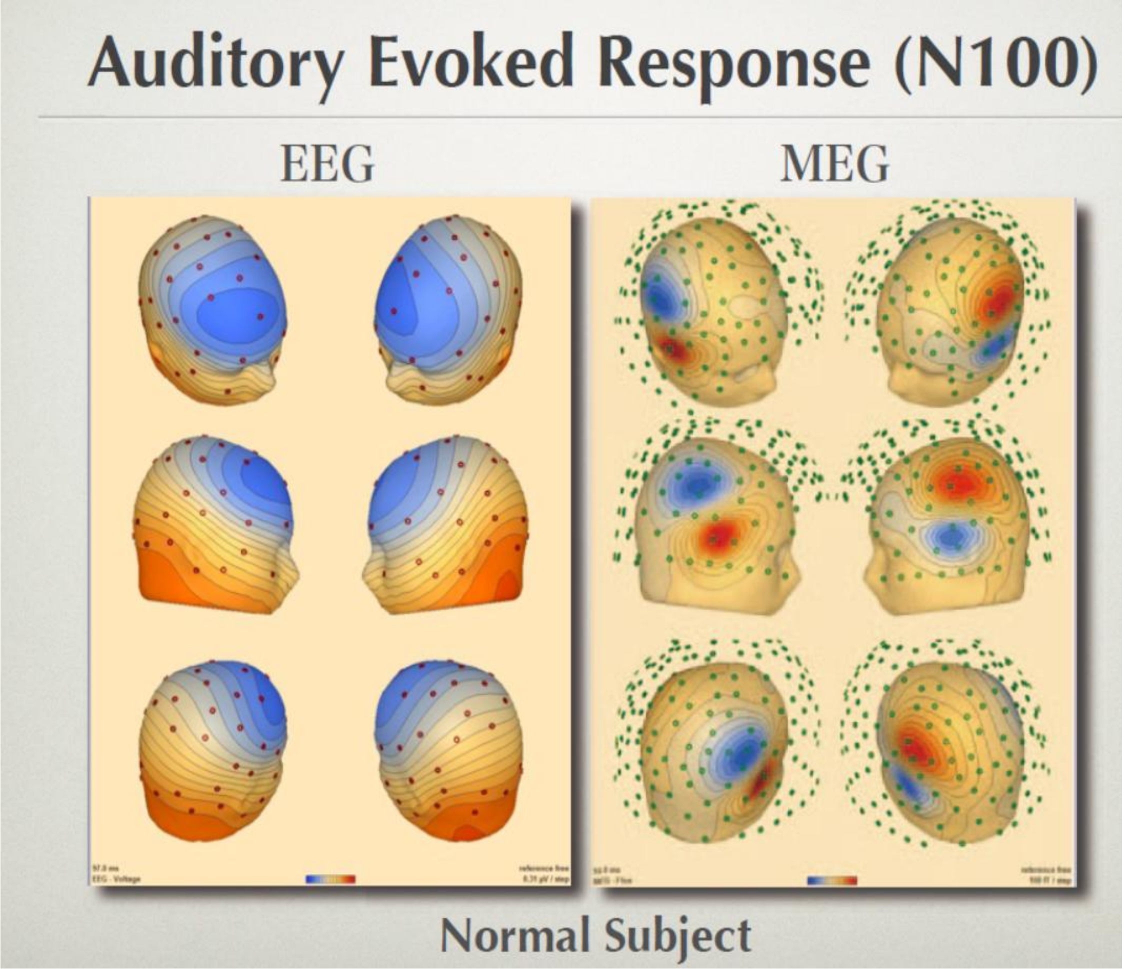
Dipole Simulator (BESA)

Number:
Single

Position:
Vertex

Orientation:
Radial

slide from: Nobukazu Nakasato, 2009



Auditory Evoked Response (N100)

EEG

MEG

Normal Subject

slide from: Nobukazu Nakasato, 2009

Generative model of the EEG

$$\mathbf{x}(t) = \int_{\mathbf{u}\in\mathcal{B}} F(\mathbf{u})\mathbf{s}(\mathbf{u},t)\mathrm{d}\mathbf{u} + \boldsymbol{\epsilon}(t)$$

$\mathcal{B}$: brain volume

Scalp potential $\mathbf{x}(t) \in \mathbb{R}^M$ depends on

$\mathbf{s}(\mathbf{u},t): \ \mathbb{R}^3 \times \mathbb{R} \rightarrow \mathbb{R}^3$: sources of interest (el. activity in the brain)

$F(\mathbf{u}): \ \mathcal{B} \rightarrow \mathbb{R}^M \times \mathbb{R}^3$: forward mapping describing the propagation of currents from sources to sensors within volume conductor

$\boldsymbol{\epsilon}(t) \in \mathbb{R}^M$: sources of no interest (e.g., sensor noise, artifacts)

Forward modeling

- $F(\mathbf{u})$ can be computed given
  - The geometry of the brain/skull/skin compartments
  - The conductivities of the different tissue types
  - The electrode positions

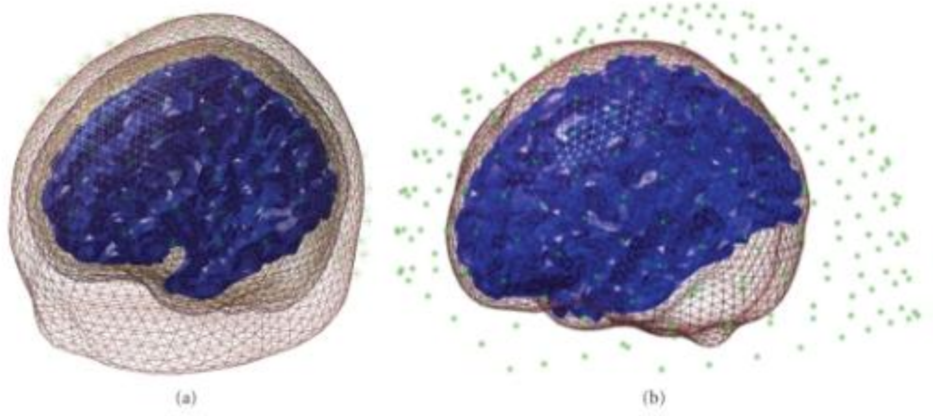
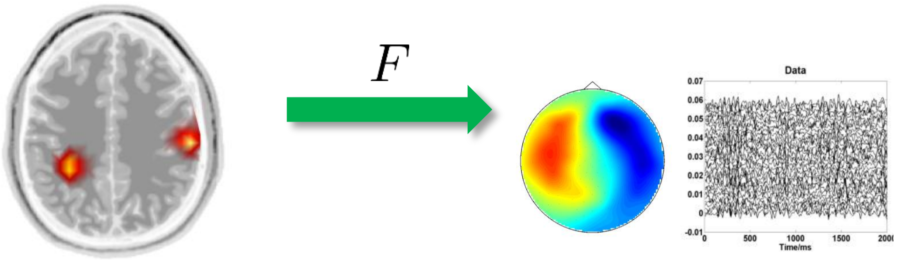


Figure from Litvak et al., 2011

- Slightly simpler for MEG than for EEG

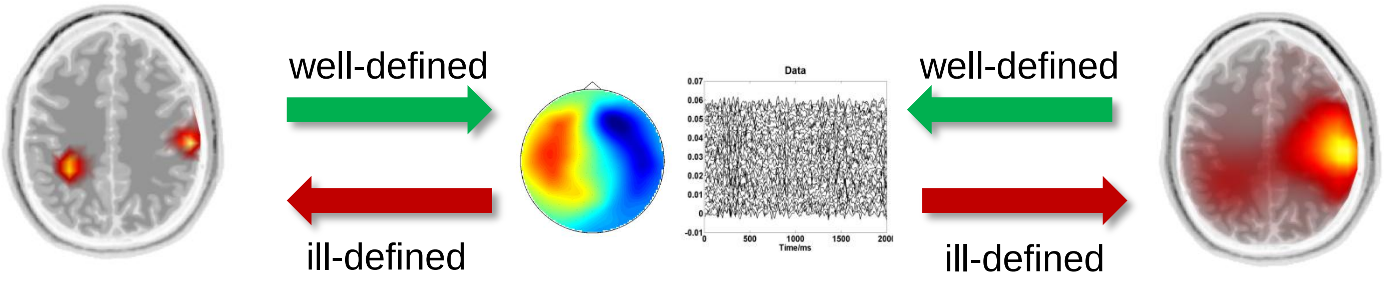
The Inverse Problem

Source reconstruction is an ill-posed problem.



The Inverse Problem

Source reconstruction is an ill-posed problem.



well-defined

ill-defined

well-defined

ill-defined

Which is the correct solution?

Source reconstruction = selecting the sources that best match prior expectations (assumptions), while explaining the data.

Source Reconstruction Paradigms

EEG model:

$$\mathbf{x}(t) = \int_{\mathbf{u}\in\mathcal{B}} F(\mathbf{u})\mathbf{s}(\mathbf{u},t)\mathrm{d}\mathbf{u} + \boldsymbol{\epsilon}(t)$$

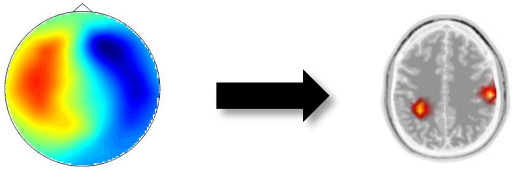
Blind source separation:

F and $\mathbf{s}$ unknown,
estimate both

e.g., ICA, CSP, xDAWN

Inverse source reconstruction:

F given by physical model,
estimate $\mathbf{s}$



Inverse methods

MNE
MCE
WMNE
Loreta
sLORETA
eLORETA
Laura
Electra
WROP
DICS
LCMV-Beamformer
Nulling Beamformer
FOCUSS
Champagne
Minimum Entropy
Dipole Modeling
Multipole Modeling
MUSICRAP-MUSIC
S-FLEX
DCM

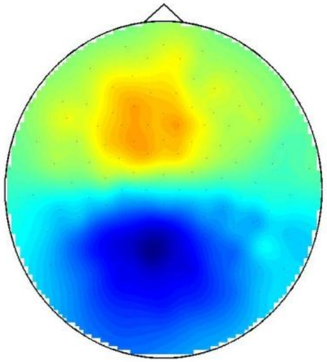
- Every inverse methods makes specific assumptions.
- (Ideally) performs well if assumptions are met.
- There can be no method that performs well in general.

Dipole modeling

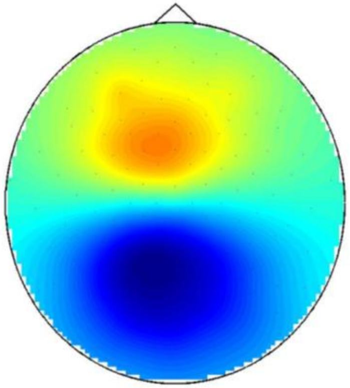
Dipole modeling

Assumption: few number of point sources

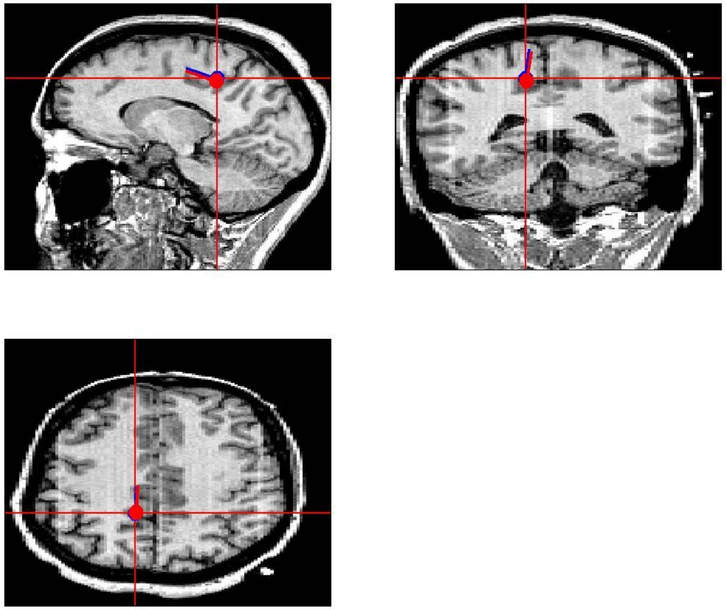
Estimate their locations and orientation.



Observation



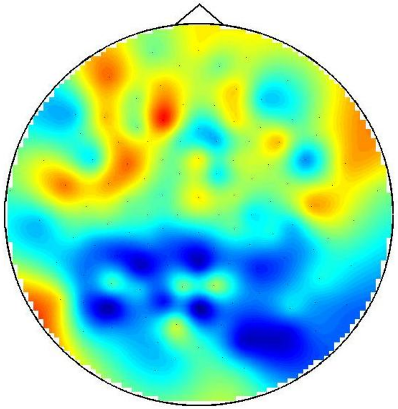
Model potential



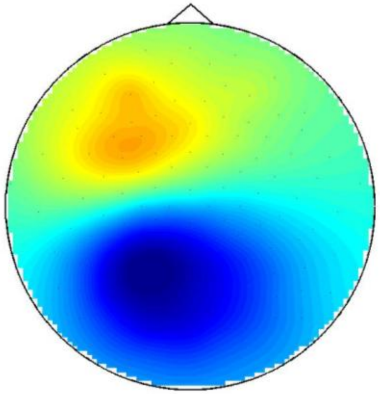
Estimated source (blue)
True source (red)

Dipole Modeling, high noise

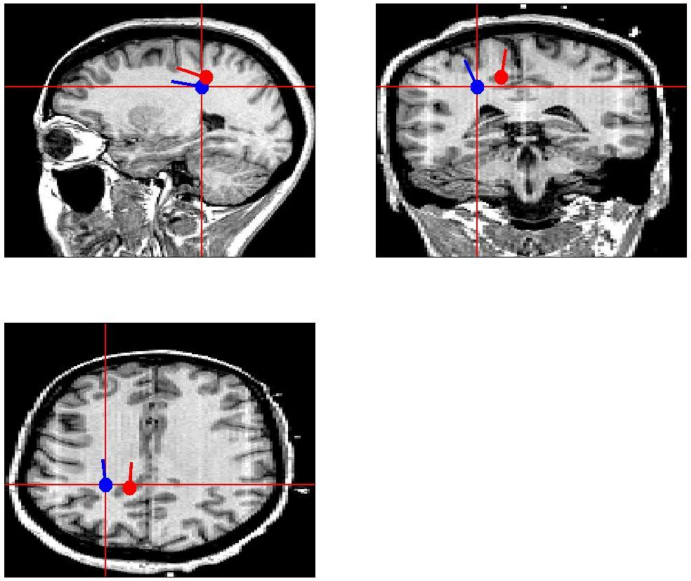
- White noise is fairly harmless for dipole fits.
- Magnitude of error is not very informative.



Observation

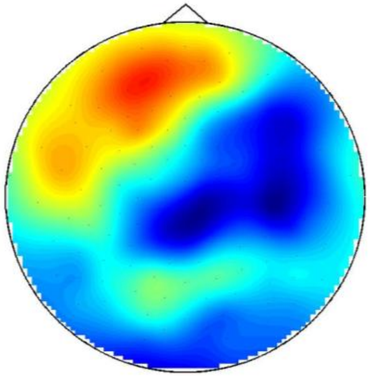


Model potential
(explains 35%)



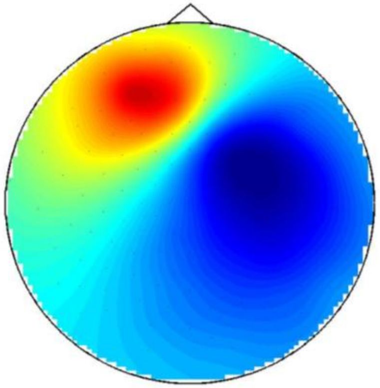
Estimated source (blue)
True source (red)

Dipole Modeling, brain noise

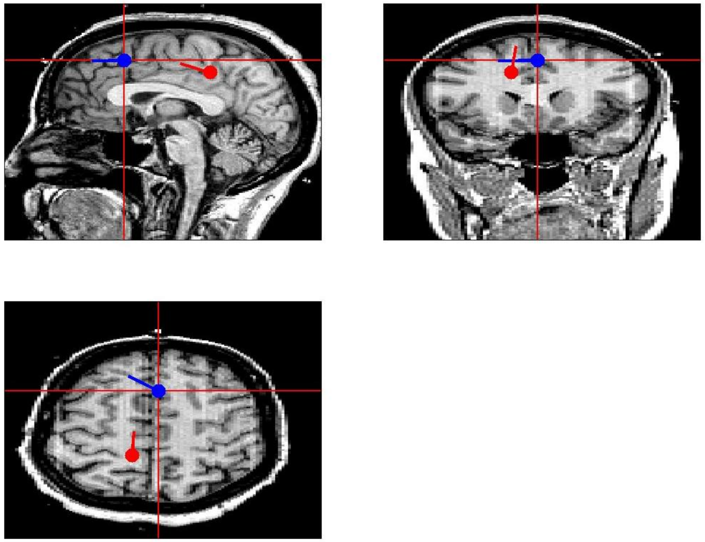


Observation

- Brain noise looks harmless but isn‘t.
- Again, magnitude of error is not very informative.



Model potential
(explains 65%)



Estimated source (blue)
True source (red)

What is minimized?

Measured potential : $\mathbf{x}$

Model potential : $\tilde{\mathbf{x}}(\{\mathbf{u}_k, \mathbf{s}_k\}) = \sum_{k=1}^{K} F(\mathbf{u}_k)\mathbf{s}_k$

Minimize a cost function $L(\{\mathbf{u}_k, \mathbf{s}_k\})$:

1. **Least-squares error**

$$L(\{\mathbf{u}_k, \mathbf{s}_k\}) = \sum_{i=1}^{M} \left(\mathbf{x}_i - \tilde{\mathbf{x}}(\{\mathbf{u}_k, \mathbf{s}_k\})_i\right)^2$$

(assumes independent and equal channel noise)

[e.g., Scherg, 1992]

What is minimized?

2. Weighted least squares error

$$L(\{\mathbf{u}_k, \mathbf{s}_k\}) = \sum_{i=1}^{M} \frac{(\mathbf{x}_i - \tilde{\mathbf{x}}(\{\mathbf{u}_k, \mathbf{s}_k\})_i)^2}{\sigma_i^2}$$

(for independent and unequal channel noise)

noise level in channel i

3. Full Maximum Likelihood

$$L(\{\mathbf{u}_k, \mathbf{s}_k\}) = \sum_{j=1}^{M} \left(\sum_{i=1}^{M} W_{j,i} \left(\mathbf{x}_i - \tilde{\mathbf{x}}(\{\mathbf{u}_k, \mathbf{s}_k\})_i \right) \right)^2$$

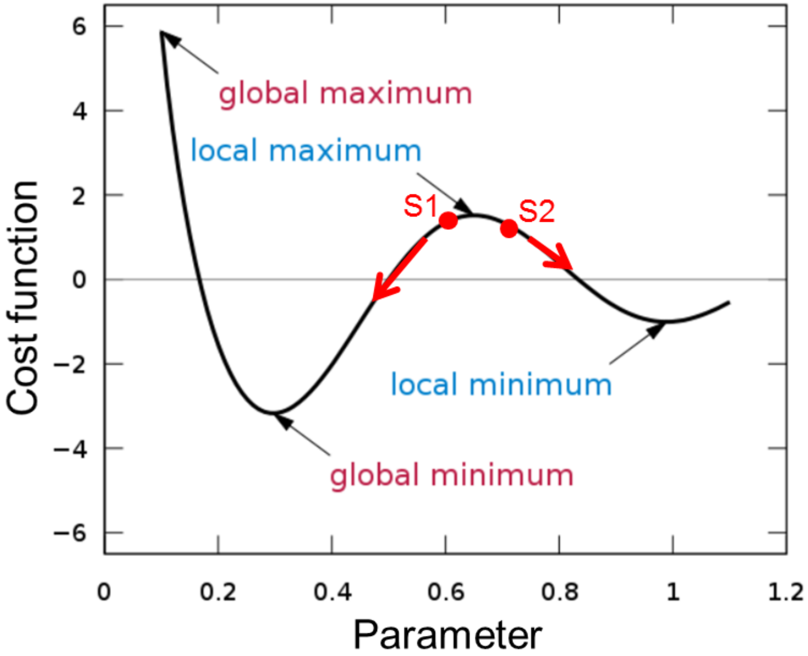
(for arbitrary noise with covariance matrix C)

Weights: $W = C^{-\frac{1}{2}}$

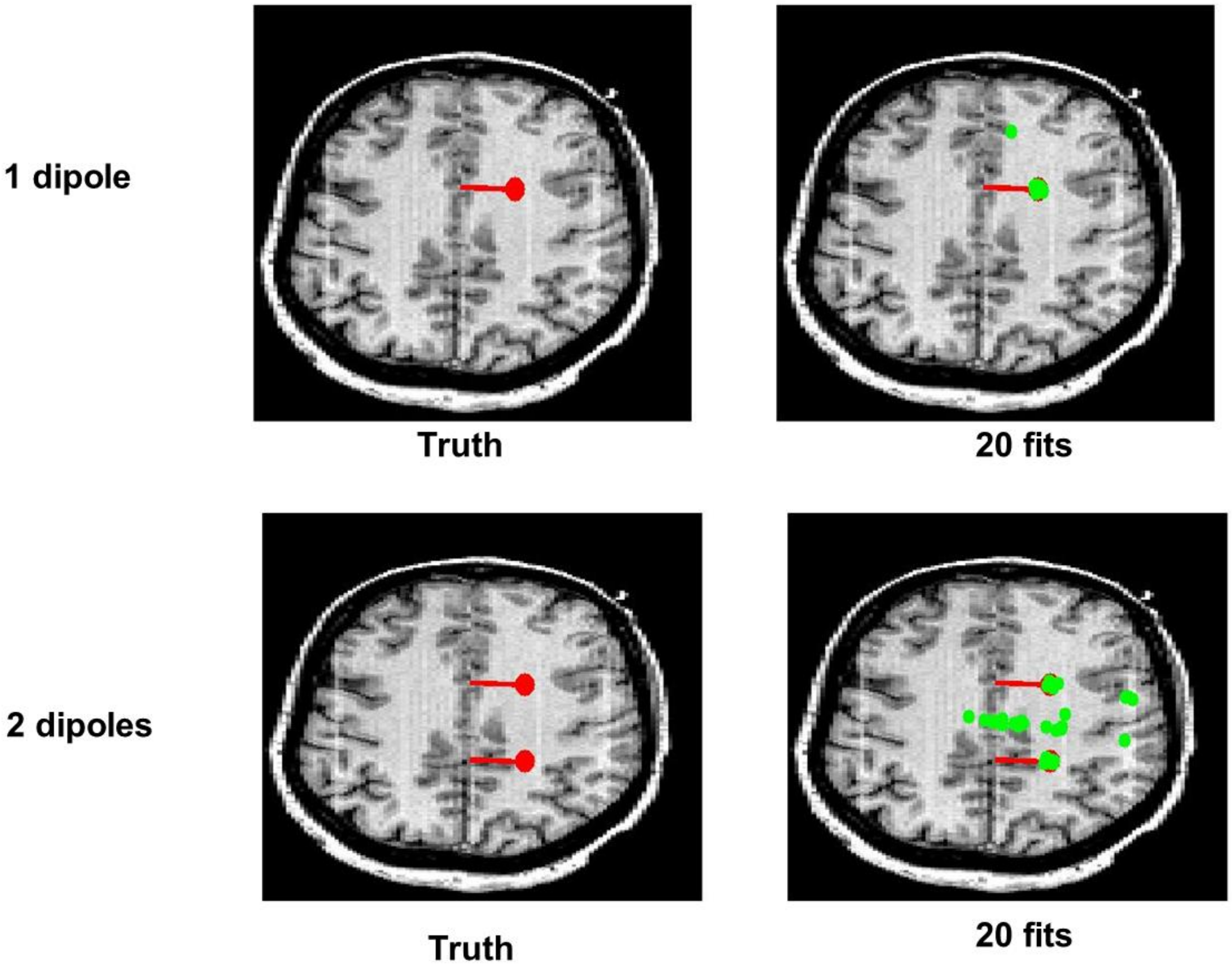
The Problem of Local Minima

In general, $F(\mathbf{u})$ is nonlinear in the location $\mathbf{u}$.

- The cost function $L(\{\mathbf{u}_k, \mathbf{s}_k\})$ has local minima (is „nonconvex“) .



The Problem of Local Minima

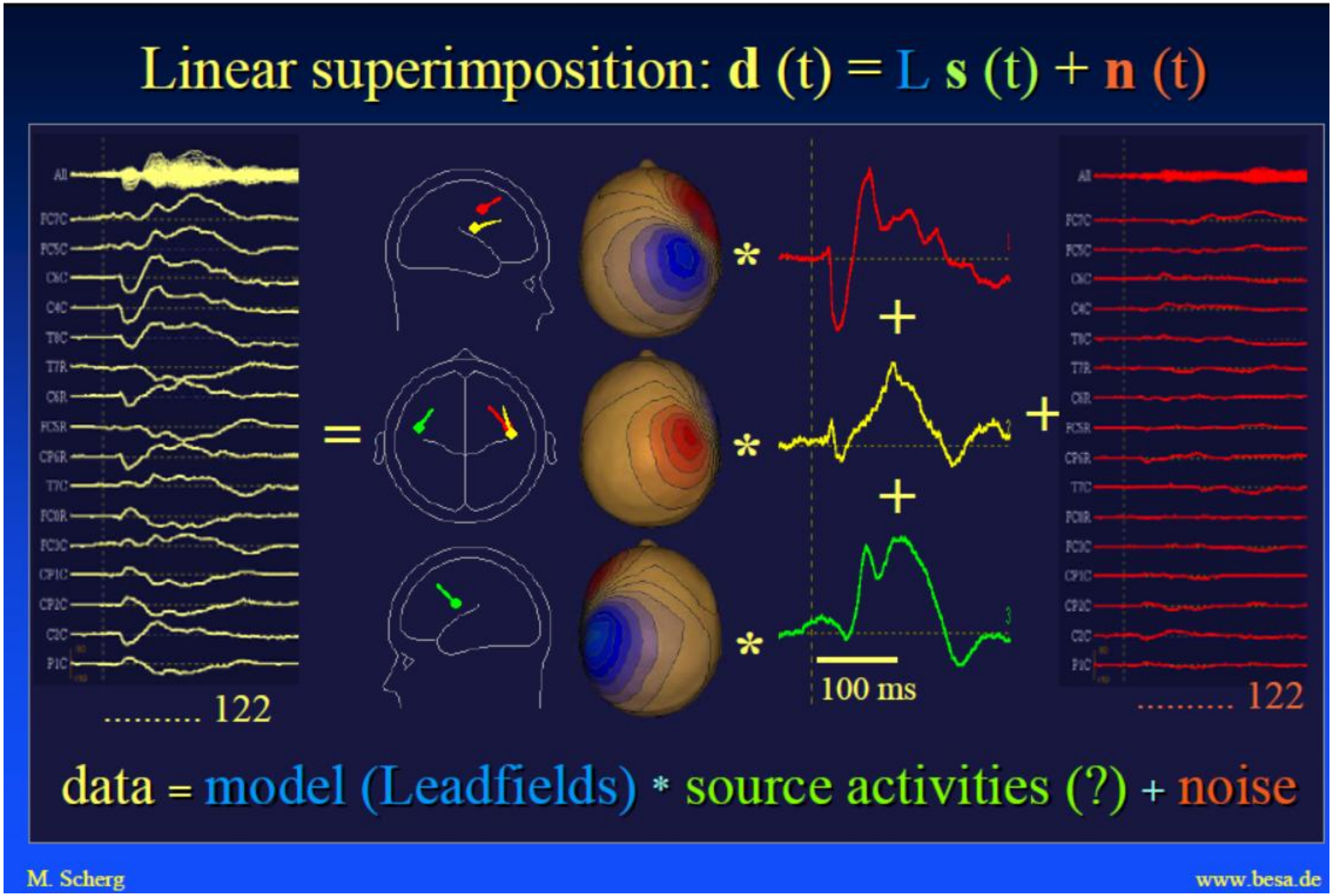


Including time

1. **„Moving dipoles“:** treat each time point separately
2. **„Fixed dipoles“:** location and orientation fixed over time, amplitude varies
3. **„Rotating dipoles“:** location is fixed, orientation and amplitude varies (models multiple sources too close to be resolved spatially)

Generally: including time stabilizes inverse solution

Example: Event-related Potentials (ERP)



Linear superimposition: $\mathbf{d}(t) = L\,\mathbf{s}(t) + \mathbf{n}(t)$

data = model (Leadfields) * source activities (?) + noise

M. Scherg

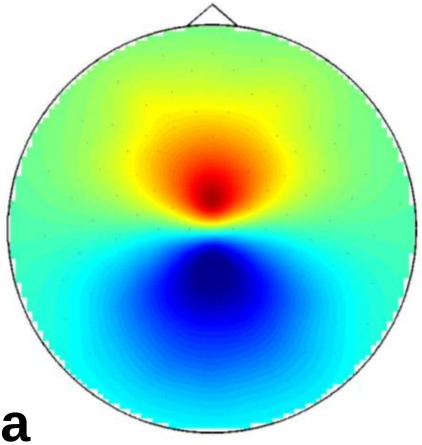
www.besa.de

Distributed Inverse Imaging

Distributed Inverse Imaging

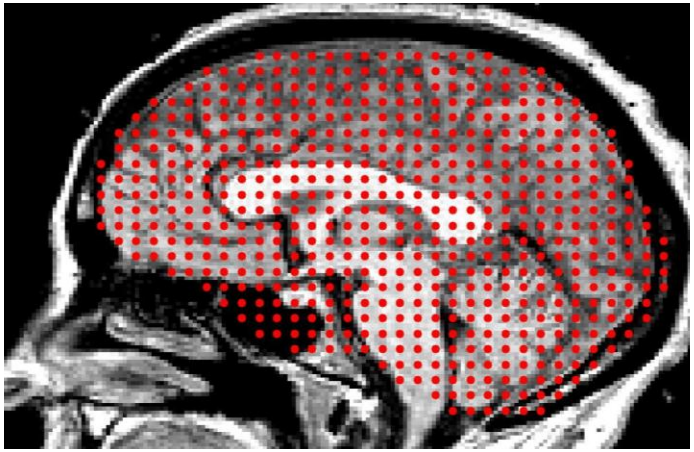
Recall: Dipole modeling tries to explain data with few sources as good as possible.

Here: Explain data ‚exactly‘ with many sources + additional constraint (regularizer/penalty).



Data

Inverse →



Voxels in brain

Model N dipoles with fixed locations, optimize over orientations.

Cost Function

- Model potential is **linear** in the orientation parameters

$$\tilde{\mathbf{x}} = \sum_{i=1}^{N} F(\mathbf{u}_i)\mathbf{s}(\mathbf{u}_i) = \sum_{i=1}^{N} F_i\mathbf{s}_i = A\mathbf{s}$$

➢ Solve underdetermined linear system for $\mathbf{s}$

Infinitely many solutions

➢ Additional constraint $g(\mathbf{s})$ needed to achieve uniqueness

1. **No noise setting:** minimize $g(\mathbf{s})$ subject to $\mathbf{x} = A\mathbf{s}$
2. **Noise setting :** minimize $L(\mathbf{s}) = \|\mathbf{x} - A\mathbf{s}\|_2^2 + \lambda g(\mathbf{s})$

Likelihood

Regularizer/penalty

Constraints

Typically: $M \ll N$ e.g., $M = 100, N = 10\,000$

- $\mathrm{rank}(A) = 100, \quad \dim\left(\mathrm{null}(A)\right) = 9\,900$
- For $\mathbf{s}_0 \in \mathrm{null}(A):\quad \|\mathbf{x} - A\mathbf{s}\|_2^2 = \|\mathbf{x} - A(\mathbf{s} + \mathbf{s}_0)\|_2^2$

- The constraint on $\mathbf{s}$ influences the solution more than the Likelihood, must be chosen wisely.

The regularizer should reflect prior knowledge on the sources.

Technical aspect: choose $g(\mathbf{s})$ to be a convex function (e.g., norm)

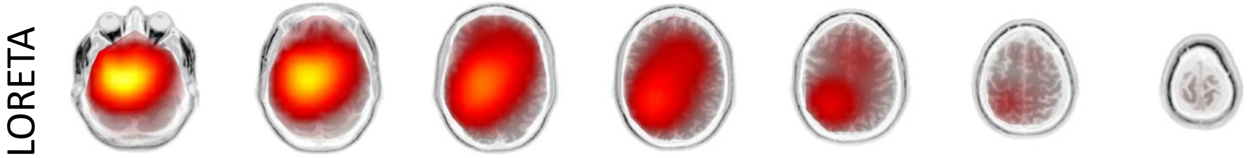
- Overall cost function $L(\mathbf{s})$ is convex, has only one global minimum

Spatial smoothness

- Assumption: neighboring voxels show similar activity

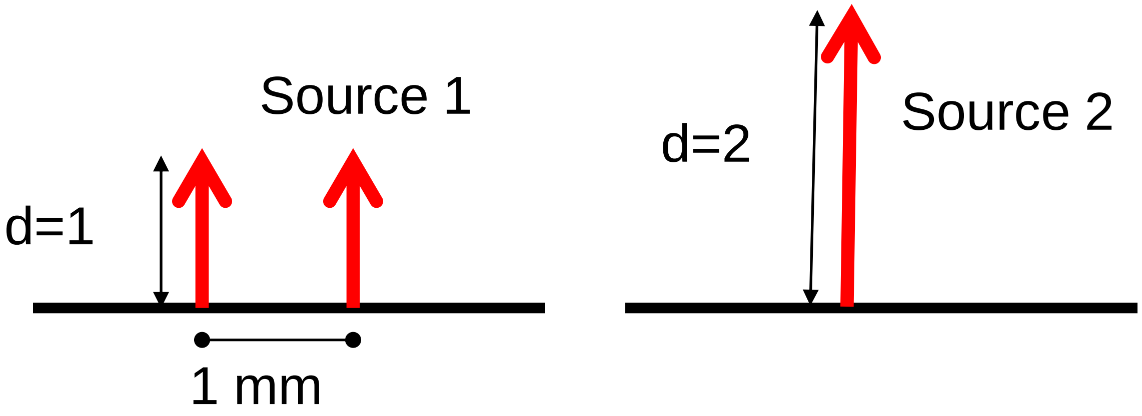
- E.g., weighted minimum norm estimate (WMN), LORETA

[Jeffs et al., 1987; Pascual-Marqui et al., 1994]



- Technically: L_2-norm $g(\mathbf{s}) = \|\Gamma \mathbf{s}\|_2^2$ leads to smoothness
  - ➢ Convex optimization
  - ➢ Solution linear in data: $\hat{\mathbf{s}} = \underbrace{\left(A^\top A + \lambda \Gamma^\top \Gamma\right)^{-1} A^\top}_{B} \mathbf{x}$
  - ➢ B are precomputable **spatial filters** → very efficient

Origin of blurring

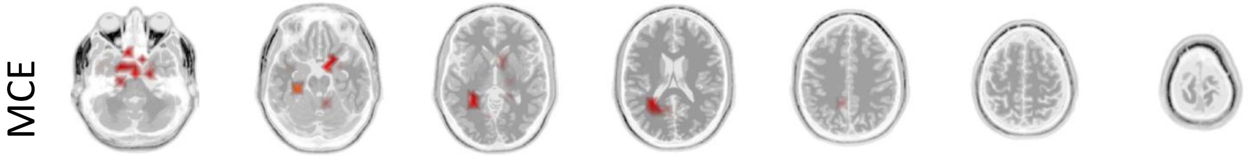


- Both sources explain data equally well
- Source 1 has L_2-norm: $\sqrt{1^2 + 1^2} = \sqrt{2}$
- Source 2 has L_2-norm: $\sqrt{2^2} = 2$

Spatial sparsity

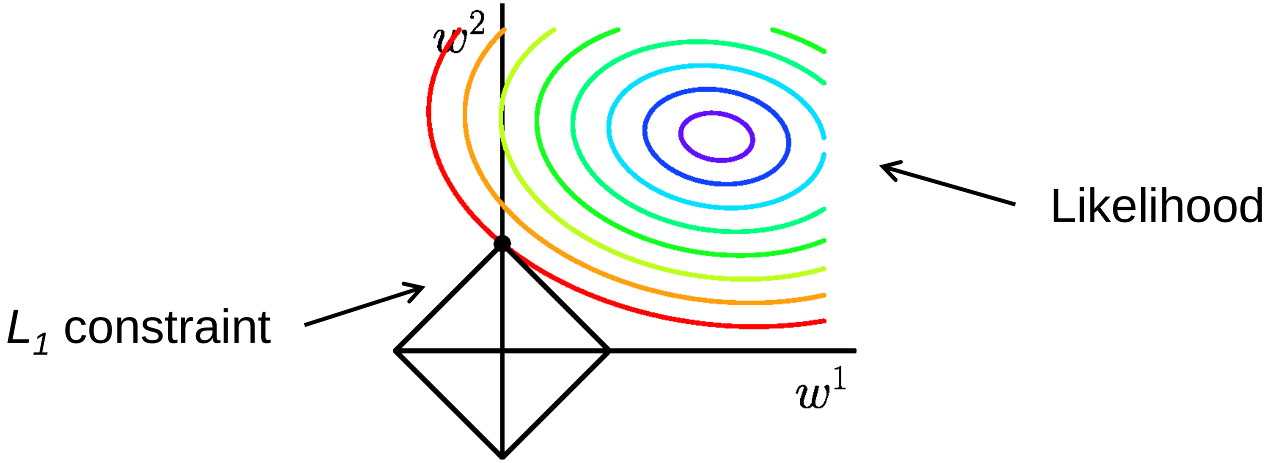
- Assumption: only a small part of the brain is active during task
- E.g., minimum current estimate (MCE), FOCUSS

[Matsuura et al., 1995; Gorodnitsky et al., 1995]



- Technically: L_1-norm $g(\mathbf{s}) = \|\mathbf{s}\|_1$ leads to sparsity
  - ➢ Convex optimization
  - ➢ Solution $\hat{\mathbf{s}}$ nonlinear in data, iterative optimization required

Origin of sparsity

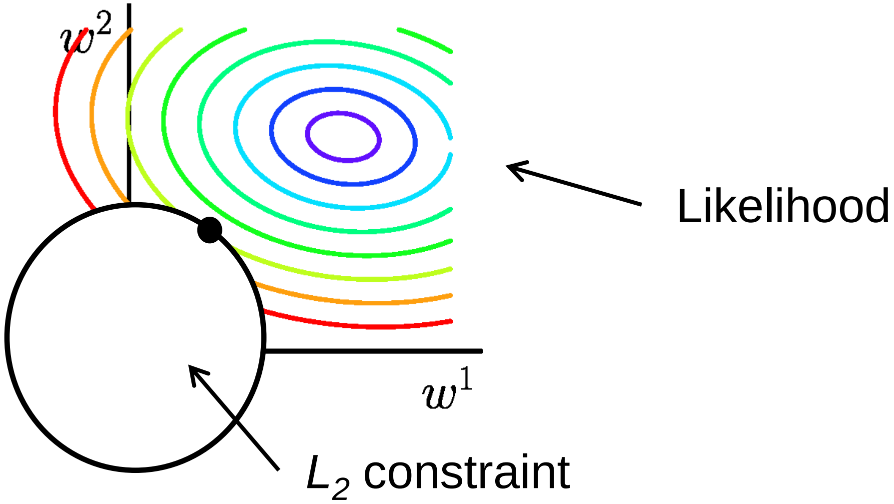


Likelihood

L_1 constraint

The level sets of Likelihood and constraint **almost always** intersect at the coordinate axes.

No sparsity using L_2-norm



The level sets of Likelihood and constraint **almost never** intersect at the coordinate axes.

Limitations of smooth (linear) and sparse inverses

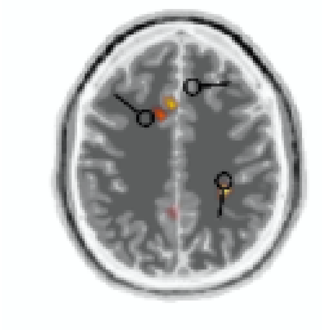
Smooth inverses

- Difficulty to distinguish sources

- „Ghost sources“

Sparse inverses

- Scattered sources in the presence of noise



Alternative constraints

Goal: combine strengths of smooth and sparse approaches

1. **Mixed-norm penalties**, e.g., $g(\mathbf{s}) = \|\mathbf{s}\|_1 + \gamma\|\mathbf{s}\|_2$

→ Solution is sparse, but still smooth

[Haufe et al., 2008; Vega-Hernández et al., 2008]

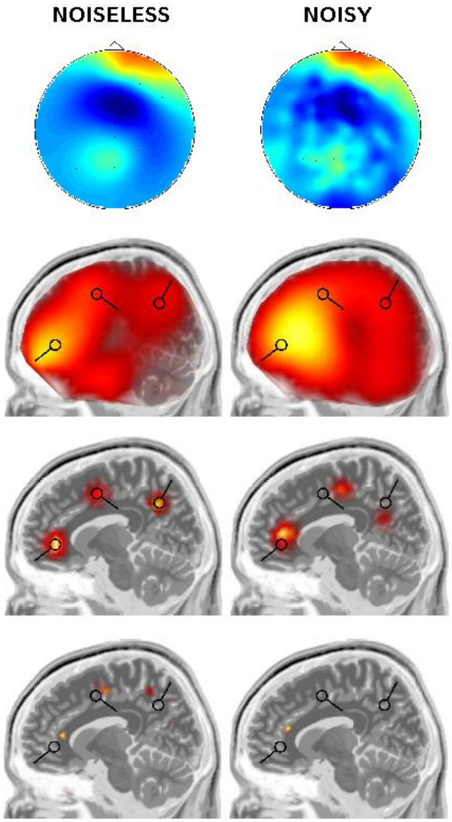
2**. Sparsity after transform**, **sparsity in different basis**

E.g. $g(\mathbf{s}) = \|\Gamma\mathbf{s}\|_1$, or $g(\mathbf{s}) = \|\tilde{\mathbf{s}}\|_1$, with $s = \|\Pi\tilde{\mathbf{s}}\|_1$

→ Solution has simple („low-dimensional“) structure

[Haufe et al., 2008; Haufe et al., 2011]

Alternative constraints



Smooth

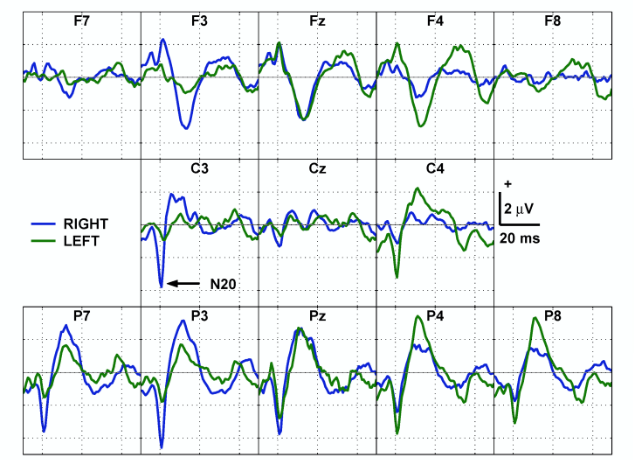
Both

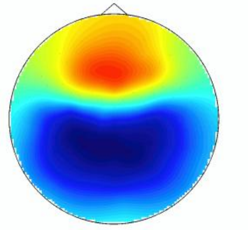
Sparse

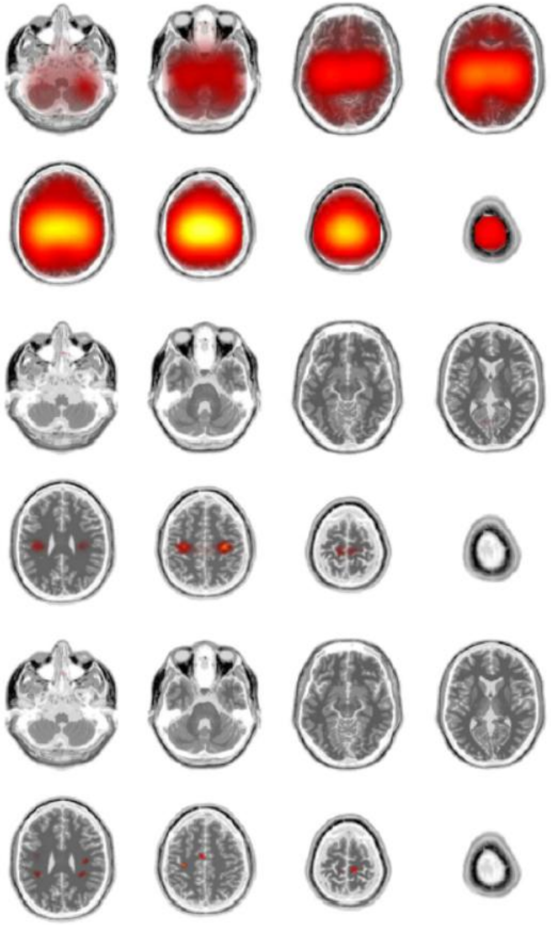
Real-world Example

- Electrical stimulation at both thumbs (Median nerves)

➢ N20 potential in the EEG







- Localization (should) reveal two symmetric sources in somatosensory cortex

[Haufe et al., 2008]

Depth compensation

- Superficial sources contribute more to the EEG than deep ones

➢ many superficial sources „cost less“ than one deep source.

➢ Location bias towards superficial sources.

Countermeasure: minimize norm of *weighted* sources

$$g(\mathbf{s}) = \|W\mathbf{s}\|_p$$

with diagonal or blockdiagonal W encoding a voxel-specific penalty

Depth compensation

1. Norm of the columns of the forward matrix *A*

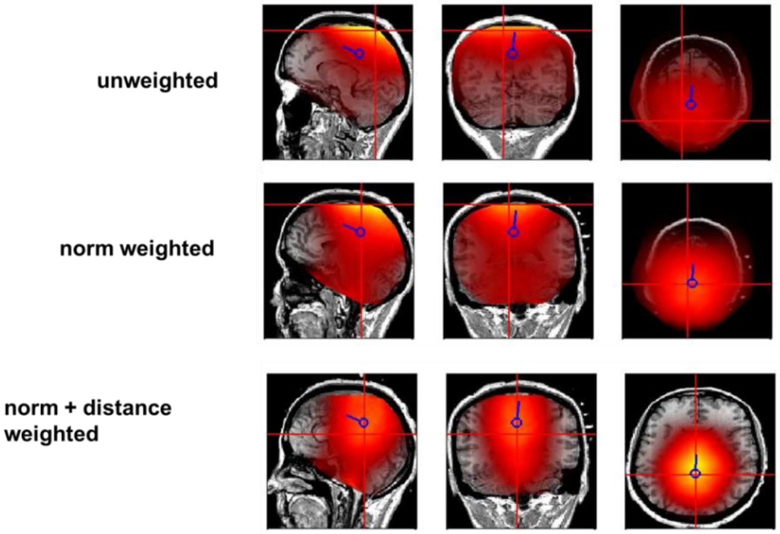
[Jeffs et al., 1987]

2. Voxel-wise (co-) variance of the minimum-norm solution

[Pascual-Marqui, 2002; Haufe et al., 2008]

3. Norm + distance from EEG sensors

[Marzetti et al., 2008]

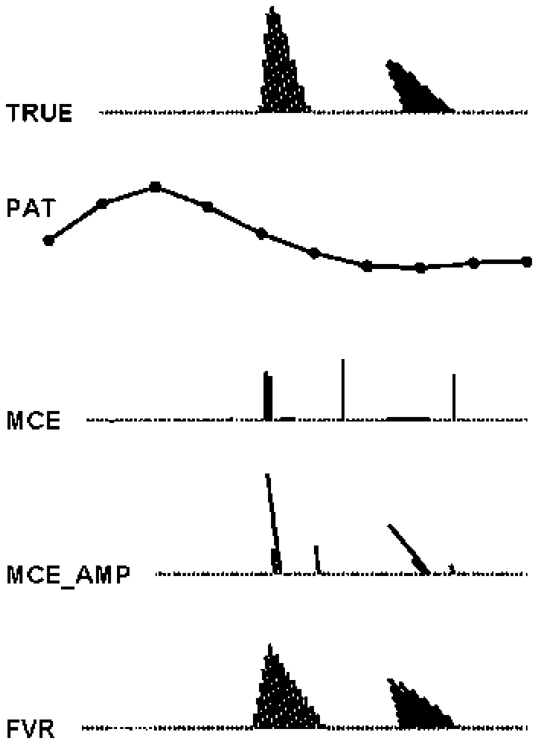


Choice of W is crucial!

Optimality results exist for 2.

Sparsity of Vector Fields

Dipole orientations are 3D vectors,
source distributions are 3D vectorfields



Technical problem: sparsification using the L_1-norm sets single dimensions to zero

- ➢ Estimated sources are not physiologically plausible (parallel to coordinate axes)

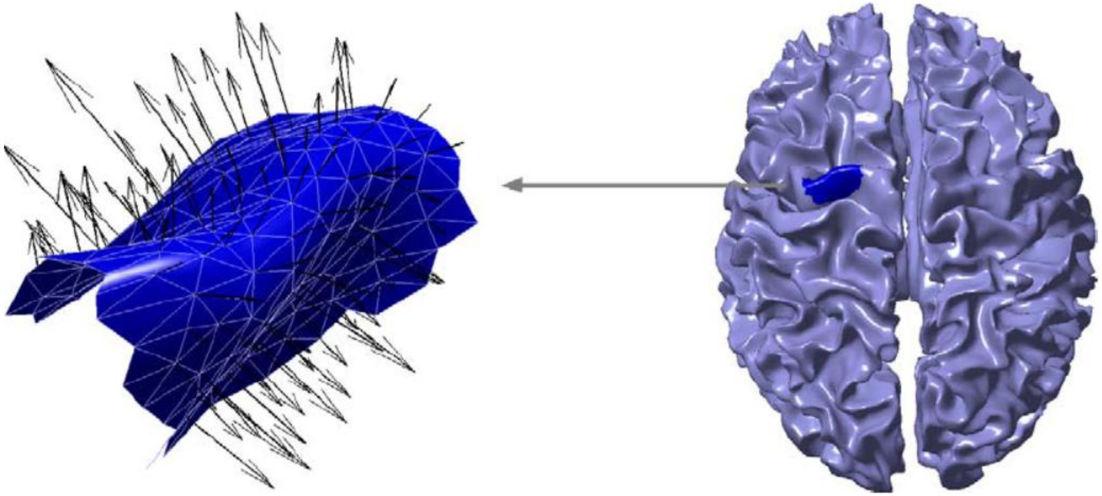
Solution: $L_{1,2}$-norm penalty $\sum_i \|\mathbf{s}_i\|_2$

- ➢ Dipole dimensions can only be pruned jointly

[Haufe et al., 2008; Ding et al., 2008; Ou et al., 2009]

More „physiological“ constraints

K. Jerbi et al. / NeuroImage 22 (2004) 779–793



1. Sources on cortex, arbitrary orientation
2. Sources on cortex, orientation normal to surface (dangerous!)
3. Regions of interest
4. Symmetric configurations

Summary (1. part)

- Inverse problem is ill-posed, constraints needed to „solve“ it
- Correct solution always relies on correctness of assumptions

- Dipole fits: few number of sources

- Distributed inverse imaging: constraints on the spatial distribution of the sources

- Both approaches (try) to explain the data completely

Inverse Methods for EEG/MEG

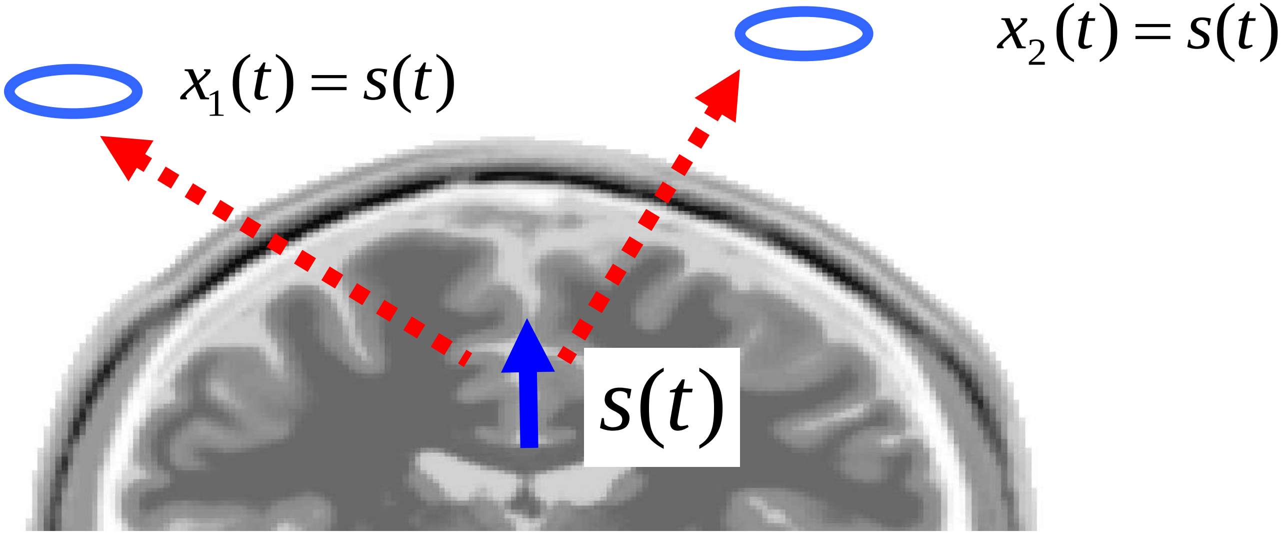
Part II

Tutorial

1. Beamformer
2. MUSIC

Beamformers

Dipole amplitude?



Task: Reconstruct signal s(t) from sensor data x_1(t) and x_2(t)!

There are many perfect solutions:

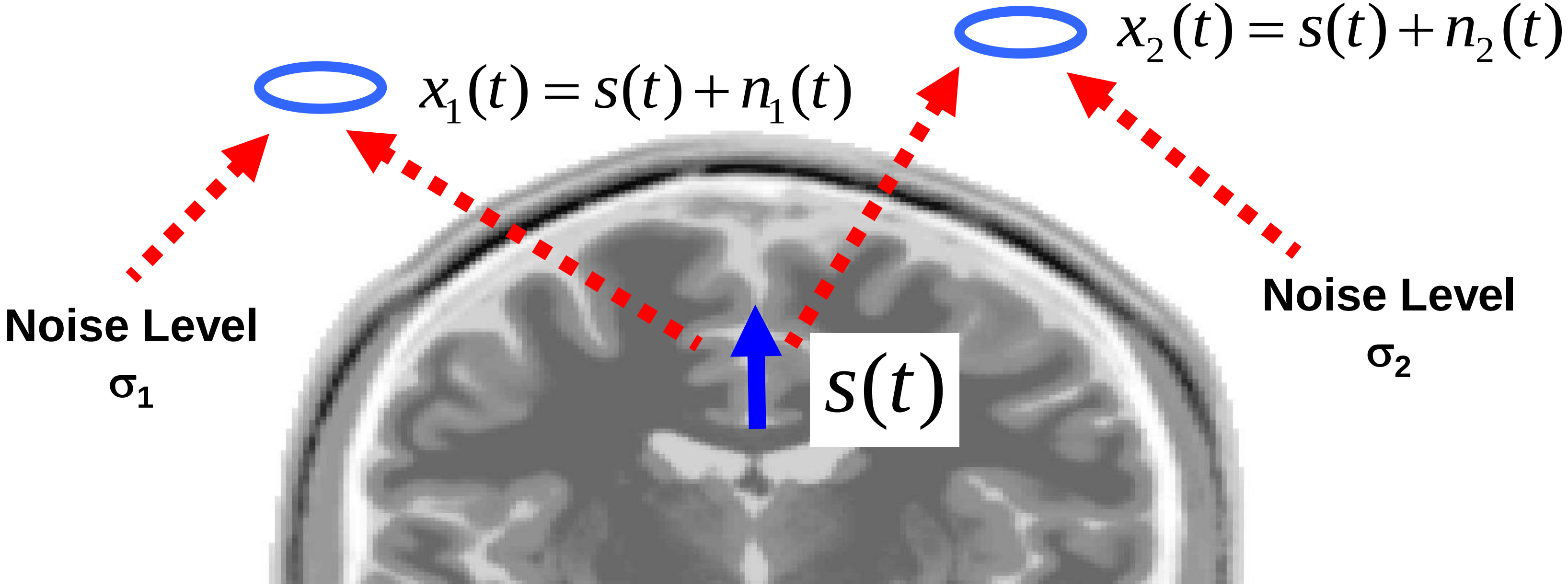
$$s(t) = x_1(t)$$

$$s(t) = x_2(t)$$

$$s(t) = \frac{x_1(t)}{2} + \frac{x_2(t)}{2}$$

$$s(t) = ax_1(t) + (1-a)x_2(t)$$

Dipole amplitude?



$$x_1(t) = s(t) + n_1(t)$$

$$x_2(t) = s(t) + n_2(t)$$

Noise Level
σ_1

Noise Level
σ_2

$s(t)$

Beamformer: Use the freedom to also maximize signal to noise ratio!

If σ_1 =0 and σ_2 >0 then choose $$s(t) = x_1(t)$$

If σ_1 =σ_2 then choose $$s(t) = \frac{x_1(t)}{2} + \frac{x_2(t)}{2}$$

If σ_1 =2σ_2 then choose $$s(t) = \frac{1}{5} x_1(t) + \frac{4}{5} x_2(t)$$

Coefficients=
„spatial filter“

We measure source + noise
How do we know the noise level?

Theorem:
If source and noise are independent, then

$$\underset{\substack{\uparrow \\ \text{can be measured}}}{\text{Var(sensor data)}}=\underset{\substack{\uparrow \\ \text{is fixed} \\ \text{„linearly constrained“}}}{\text{Var(source)}}+\text{Var(noise)}$$

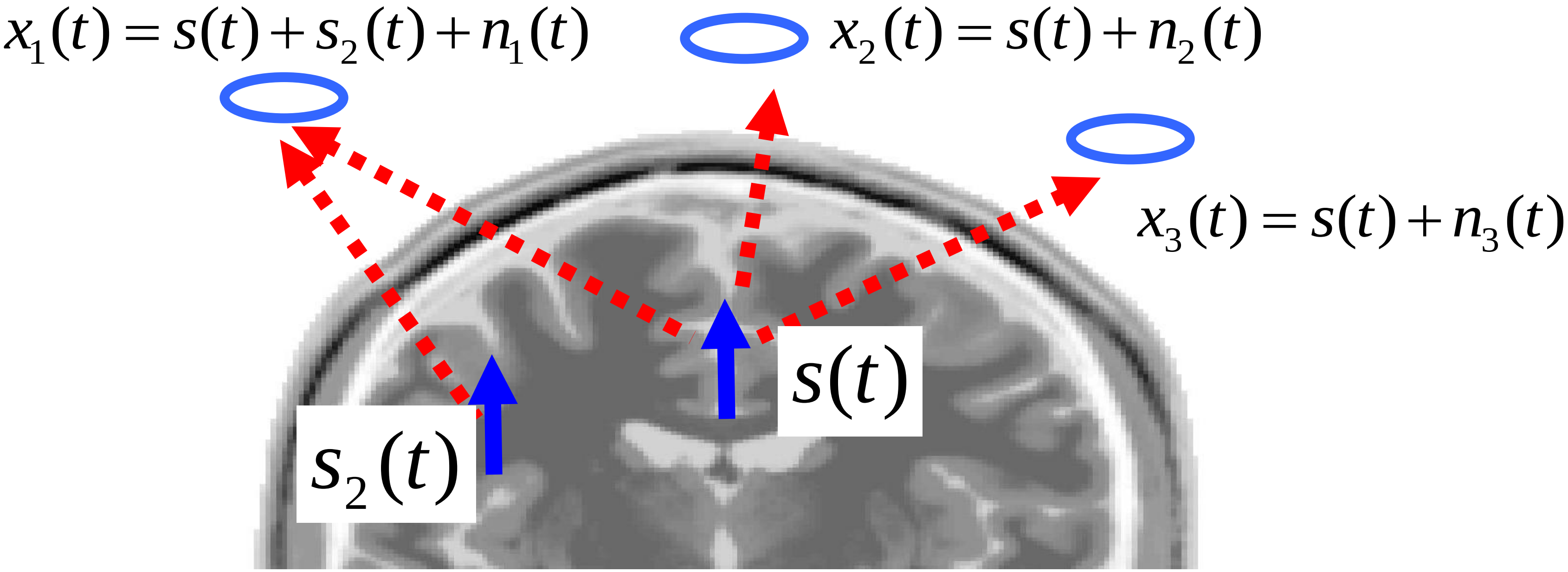
$\Rightarrow$

Minimizing Var(sensor data)
=Minimizing noise level
= Maximizing Signal to Noise raio

LCMV = Linearly Constrained Minimum Variance

Problem if „noise“ is not independent

Nulling beamformer



$$x_1(t) = s(t) + s_2(t) + n_1(t)$$

$$x_2(t) = s(t) + n_2(t)$$

$$x_3(t) = s(t) + n_3(t)$$

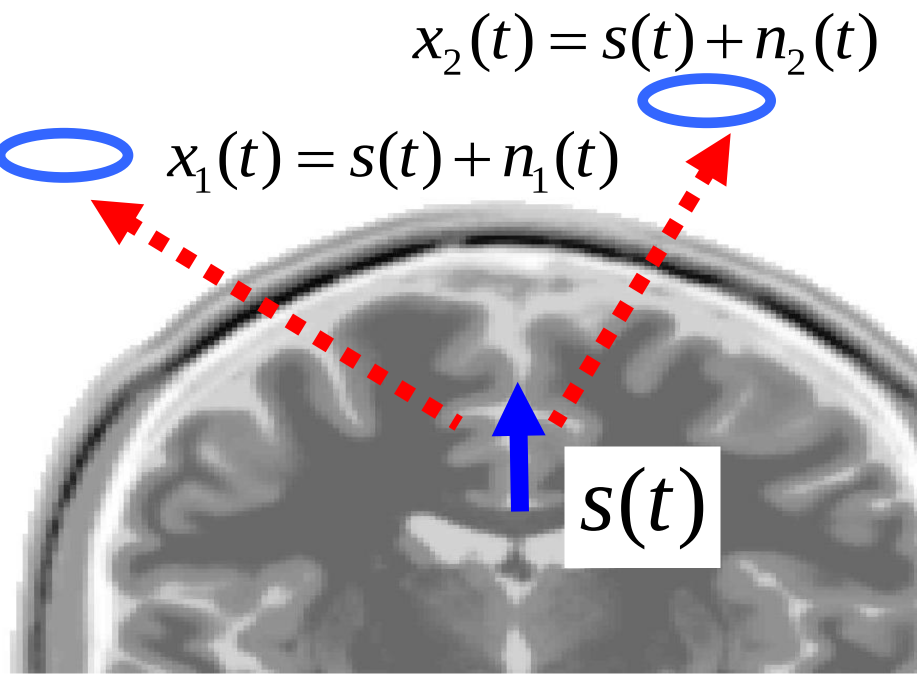
Three goals

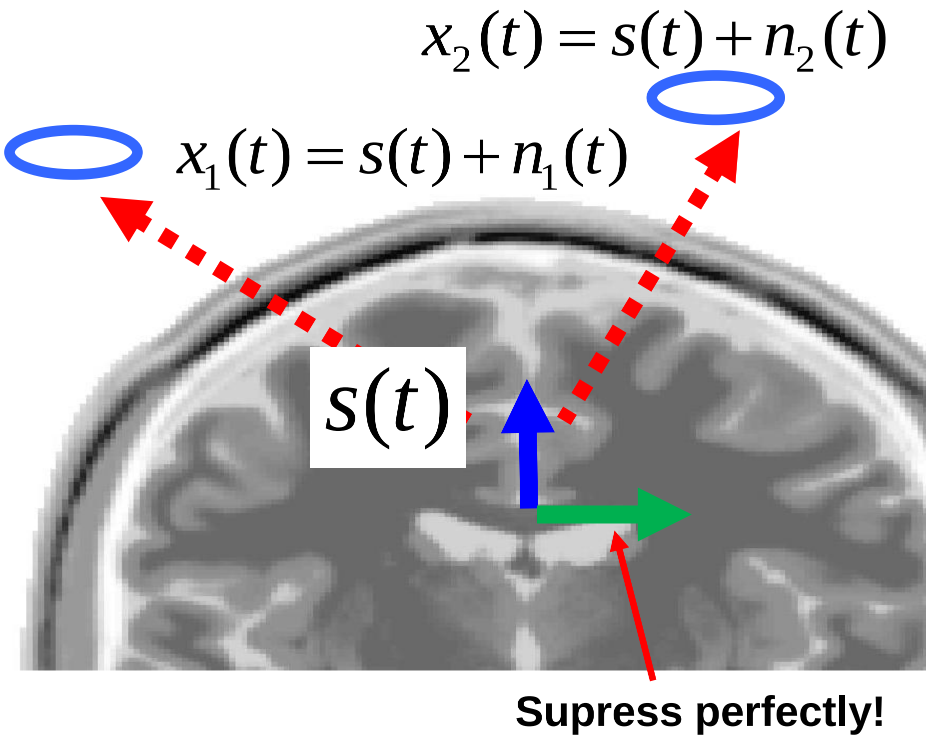
1. **See s(t) perfectly**
2. **suppress s_2(t) perfectly**
3. **suppress noise as good as possible**

Solution here: just ignore first sensor and apply beamformer on the remaining channels

Can be solved in the general case

SAM and LCMV beamformer





SAM = Synthetic Aperture Magnetometry

1. **See s(t)**
2. **Minimize noise**

LCMV= Linearly Constrained Minimum Variance

1. **See s(t)**
2. **Suppress orthogonal dipoles perfectly**
3. **Minimize noise**

For each direction seperarately, then fix direction with maximal power

DICS=LCMV beamformer in frequency domain

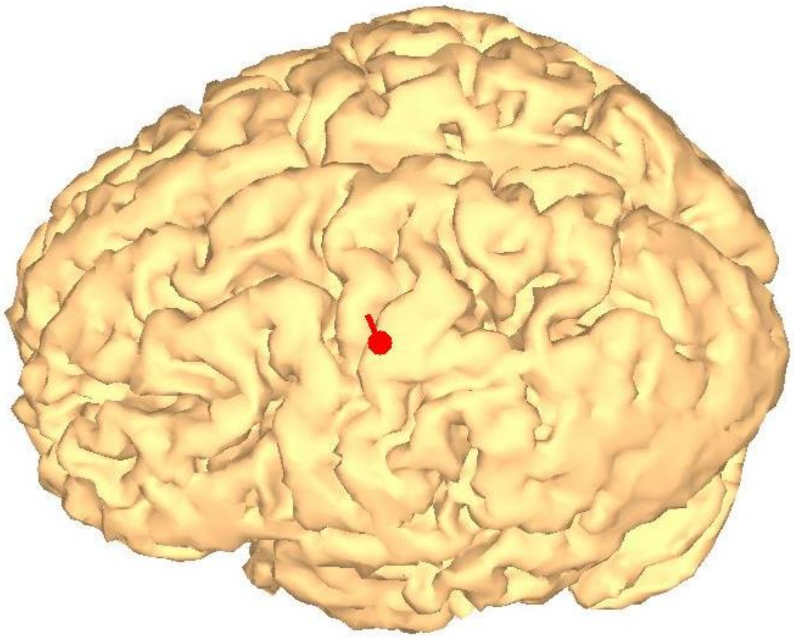
Variance → Power at frequency f
Covariance matrix → cross spectrum at frequency f

EEG-simulation of ERD (1 source)

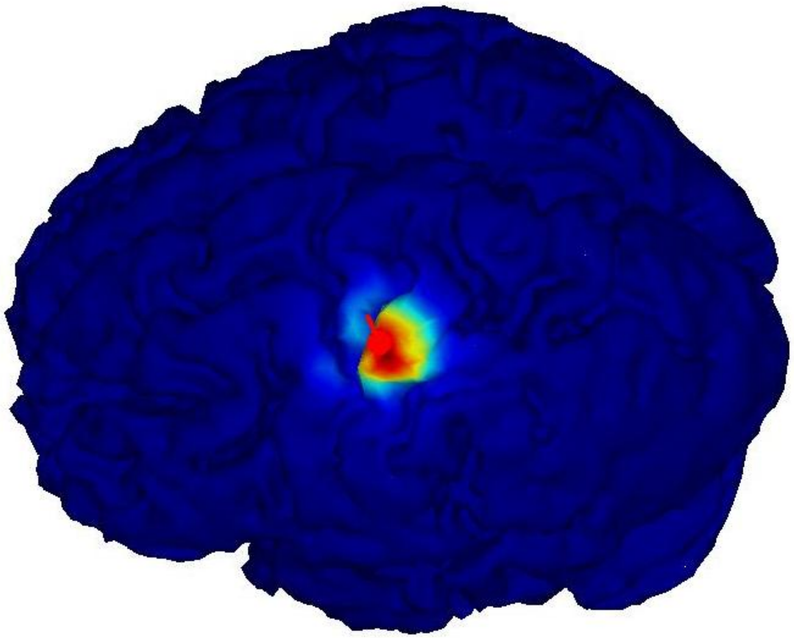
Rest: Real background + simulated dipole
Task: Real background

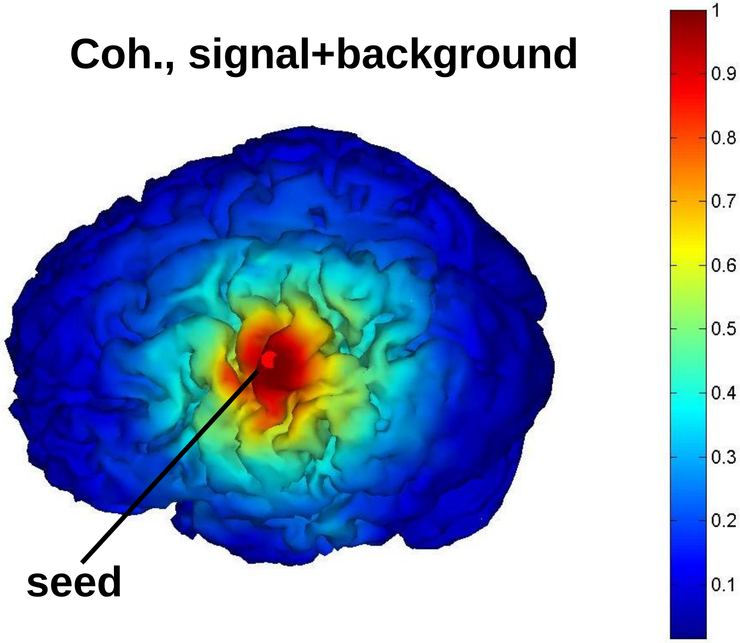
Inverse using beamformer (DICS) on cortex

Simulated dipole

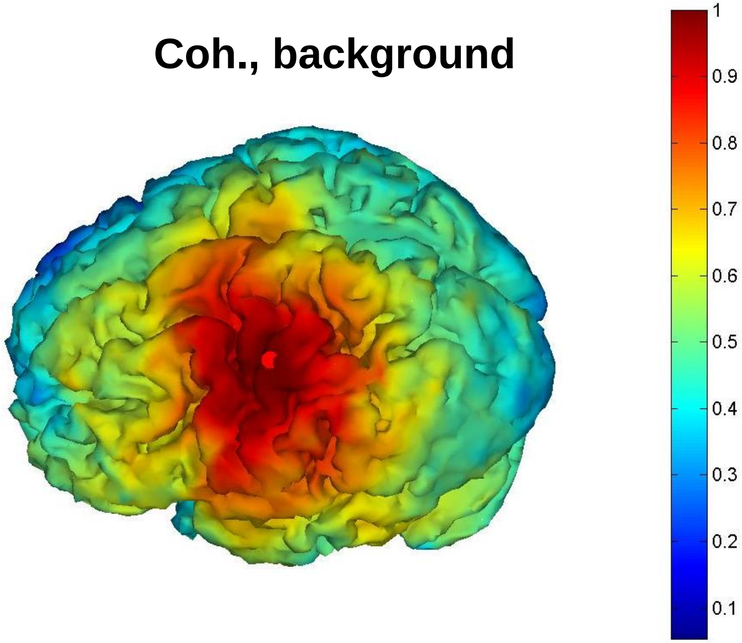


Estimated power ratio: Rest/Task

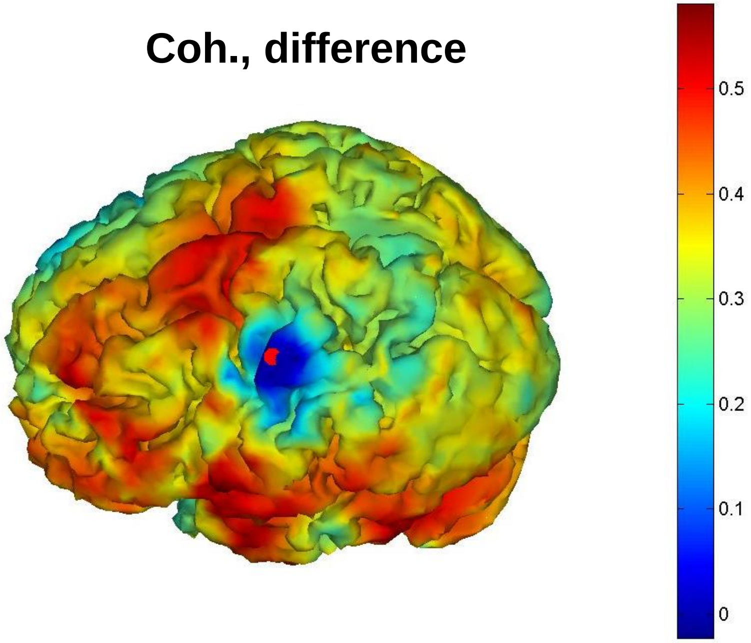




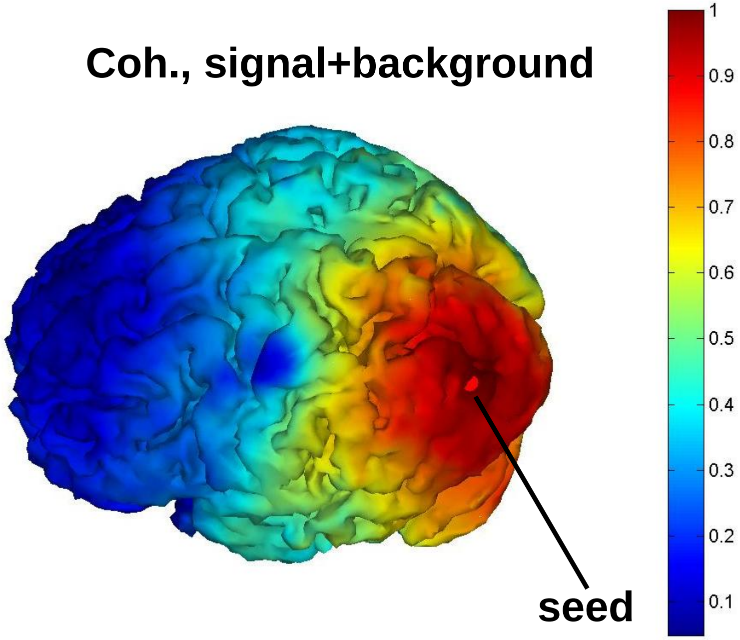
Coh., signal+background



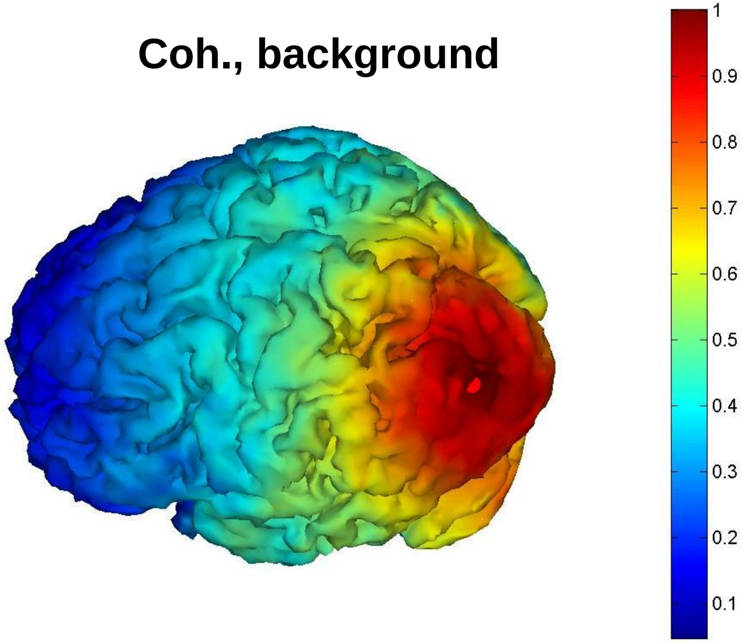
Coh., background



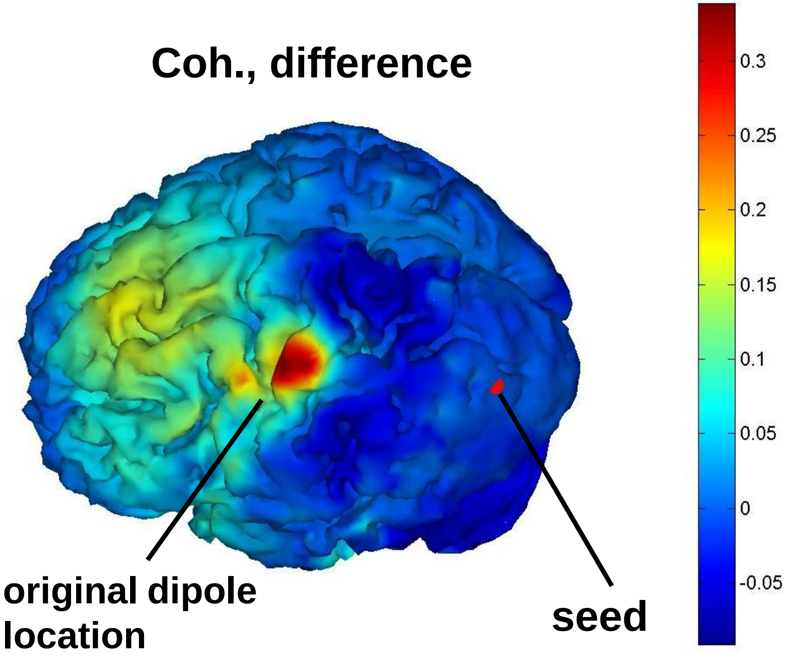
Coh., difference



Coh., signal+background

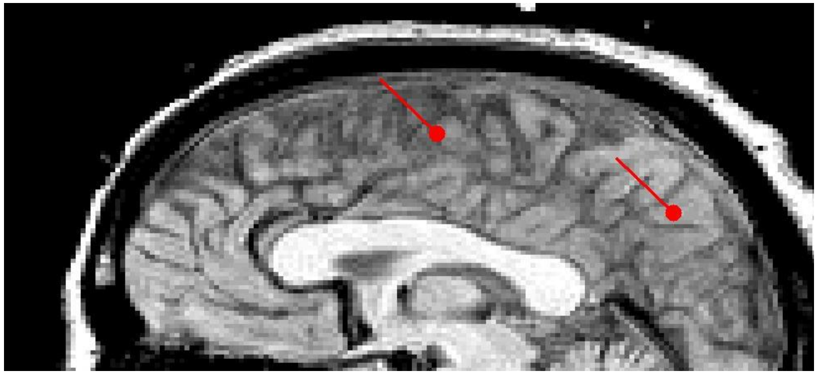


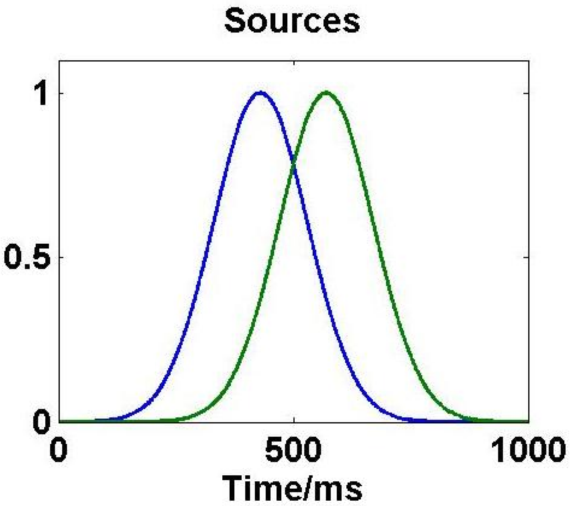
Coh., background

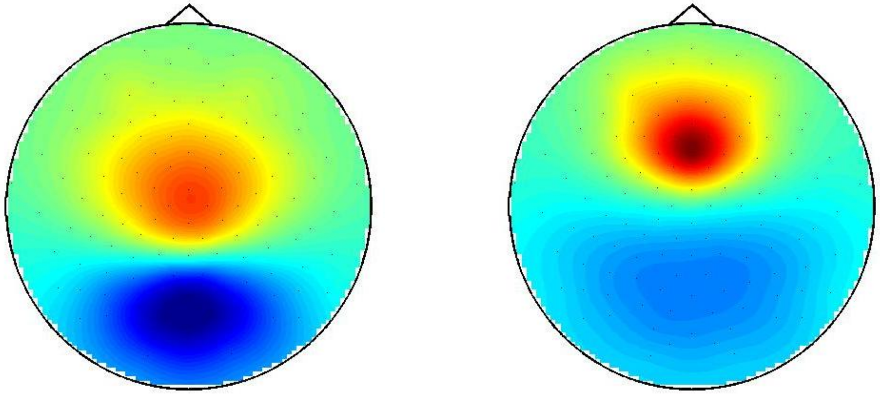


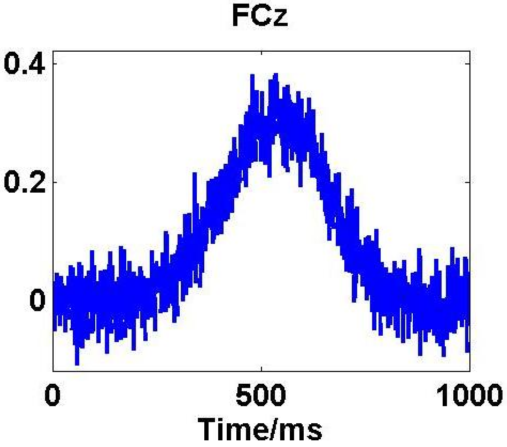
Coh., difference

MUSIC (Multiple Signal Classification)







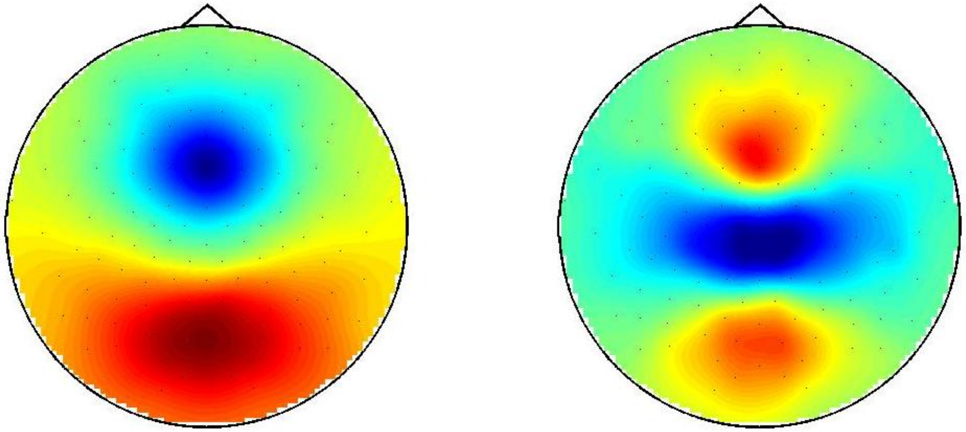


Noise

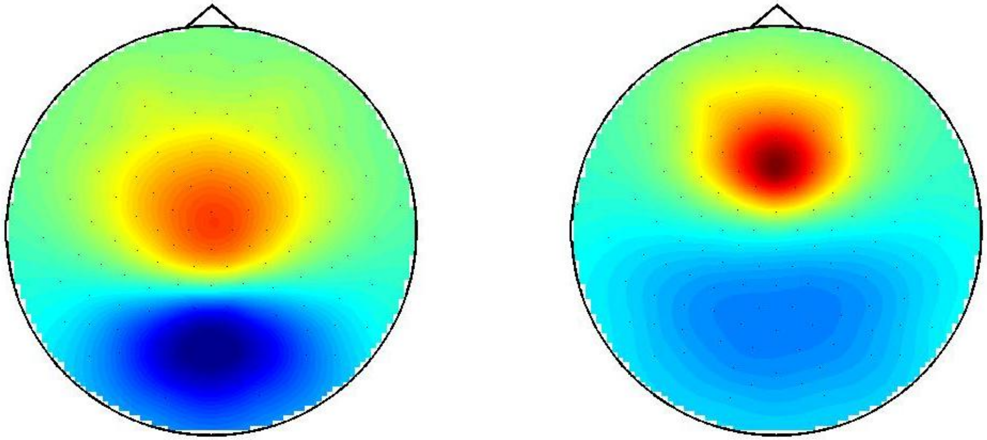
Data

1. Find important patterns in data: PCA of covariance matrix

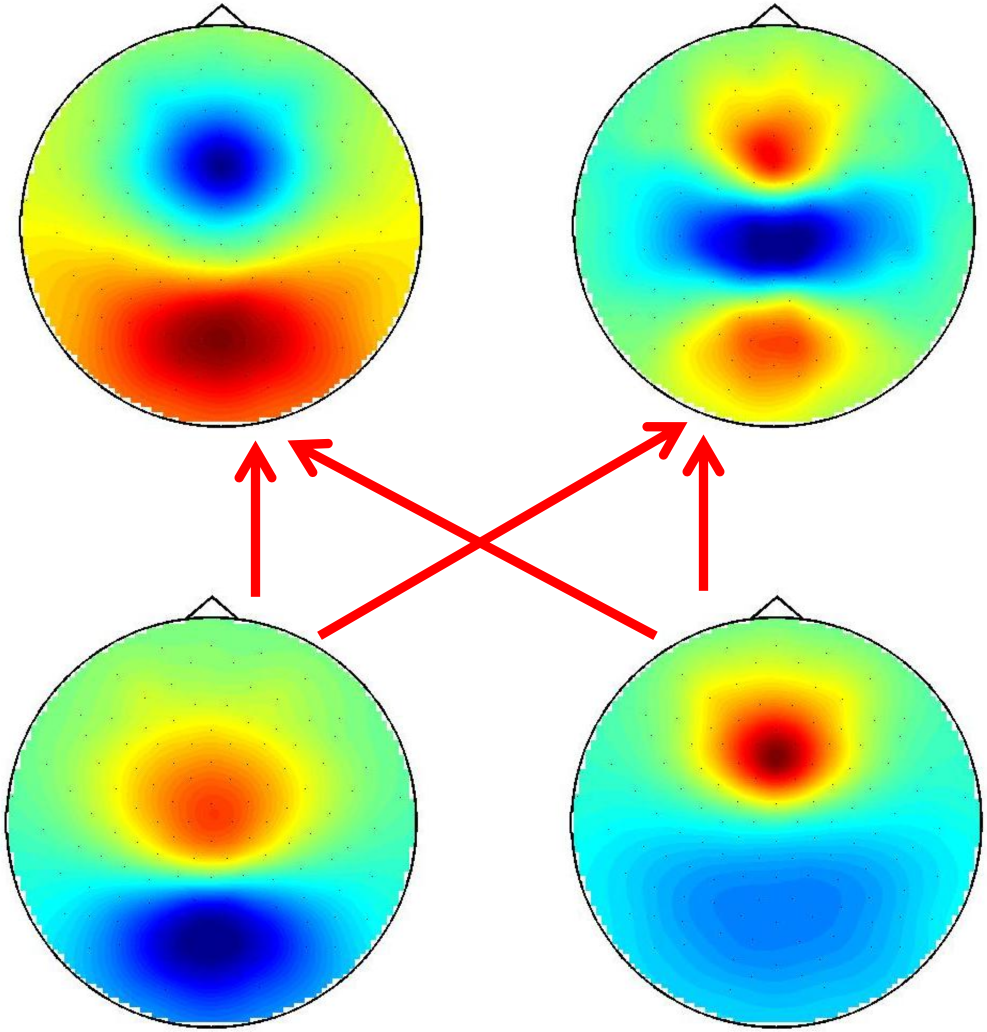
First two eigenvectors:



True patterns:



1. Find important patterns in data: PCA of covariance matrix

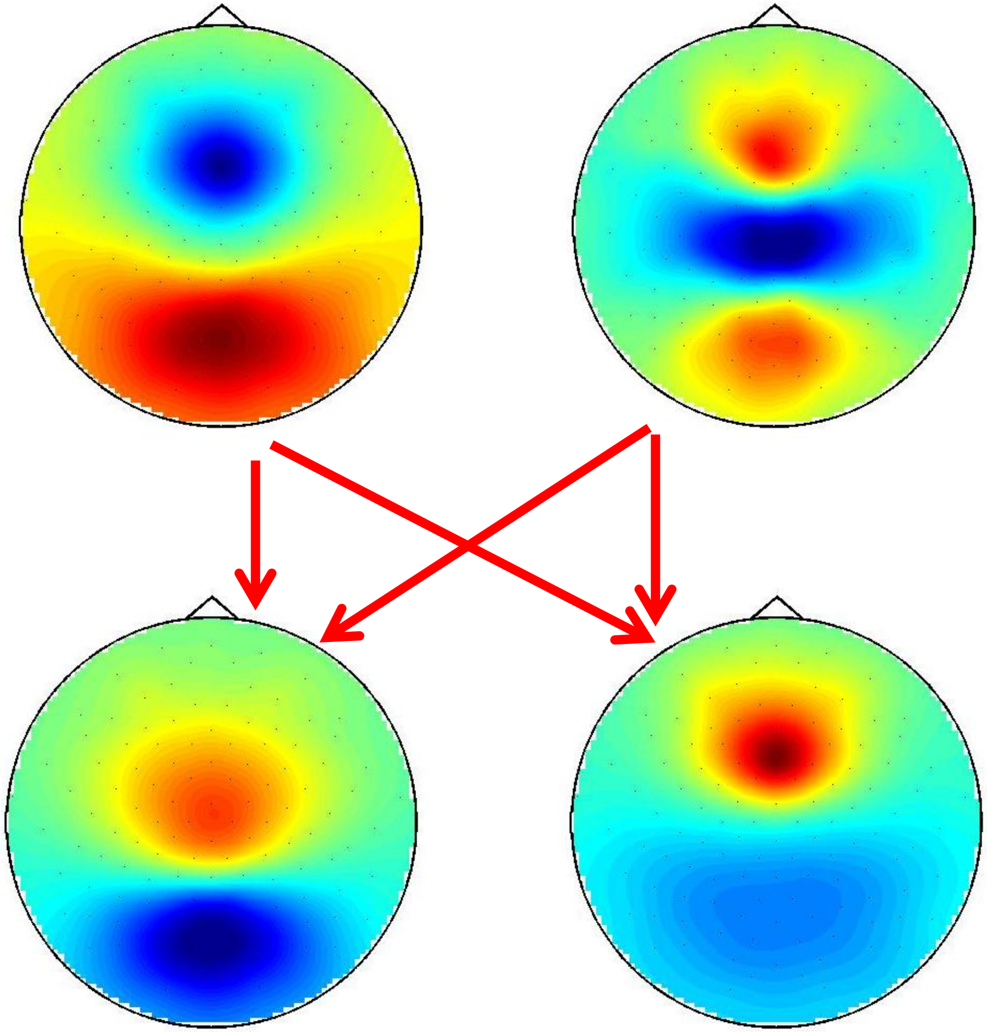


First two eigenvectors:

True patterns:

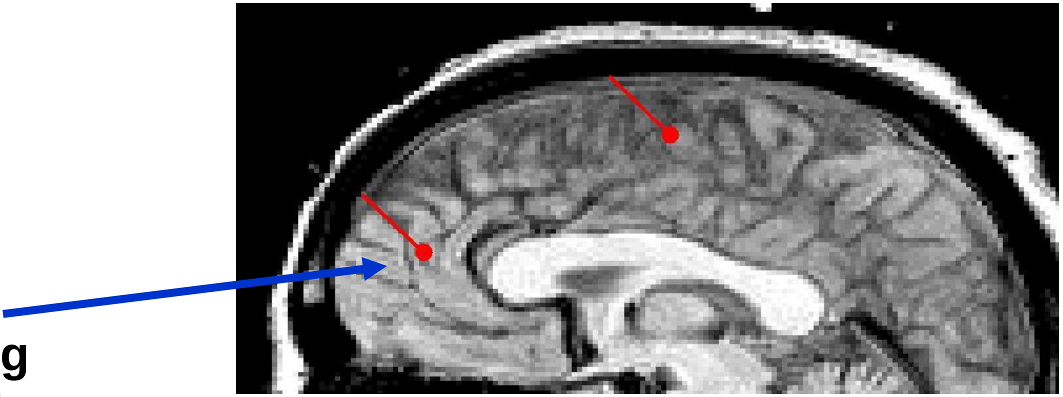
1. Find important patterns in data: PCA of covariance matrix

First two eigenvectors:

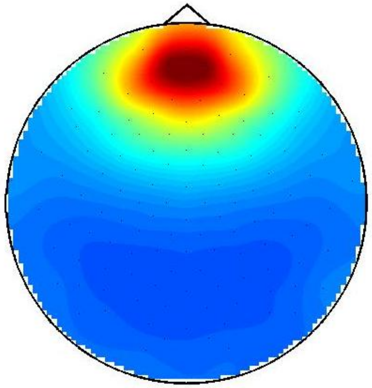


True patterns:

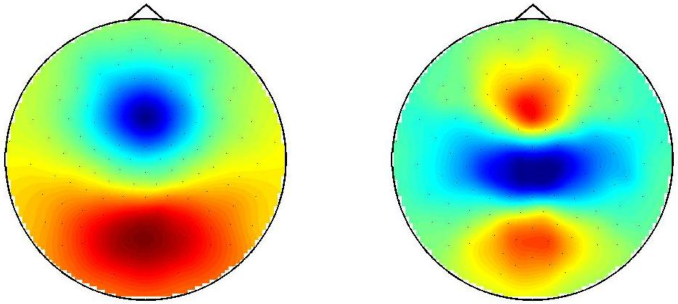
2. Does a combination of eigenvectors look like a dipole at a some location?



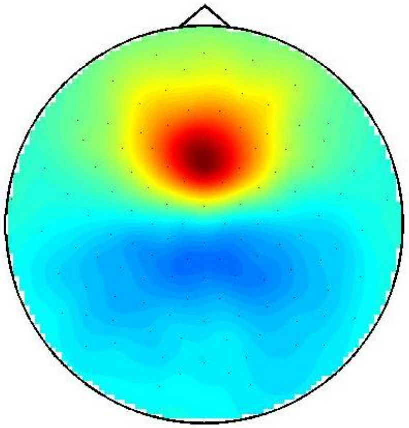
Wrong location



Topography of wrong dipole

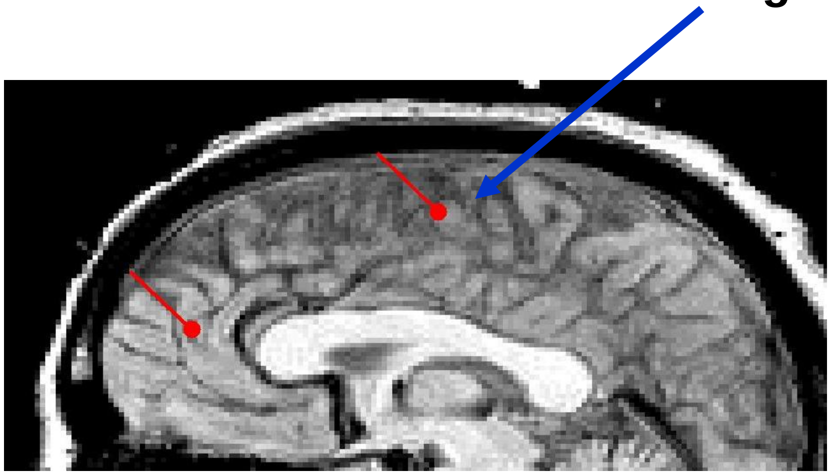


First two PCA comp.

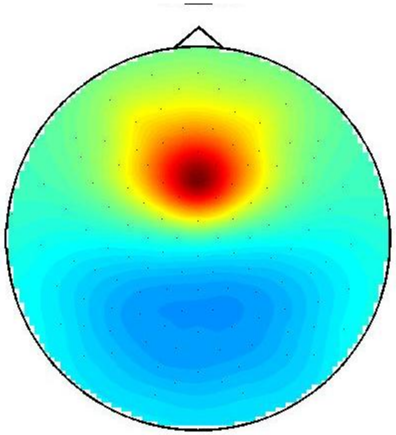


Best fit (is bad)

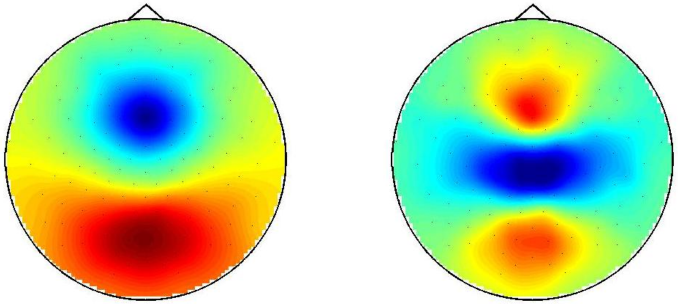
2. Does a combination of eigenvectors look like a dipole at a some location?



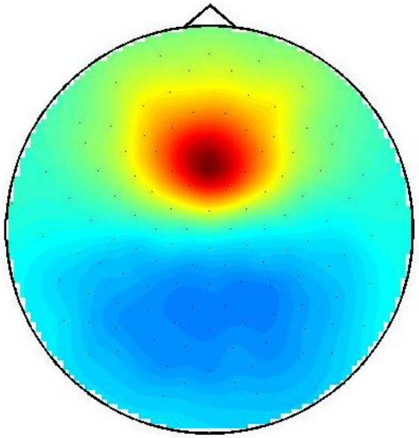
Right location



Topography of right dipole

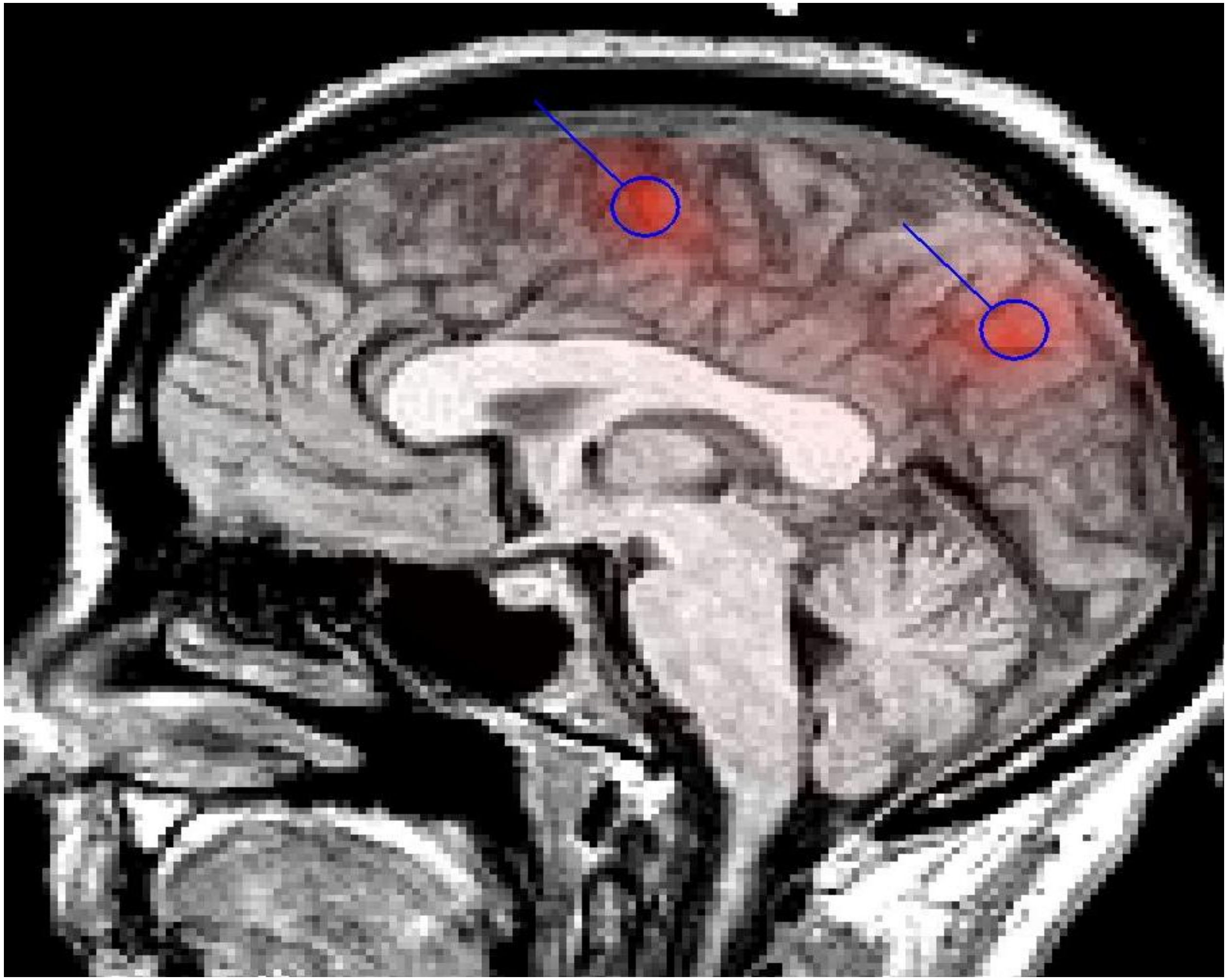


First two PCA comp.



Best fit (is good)

Scan: one slice



Scan: whole brain

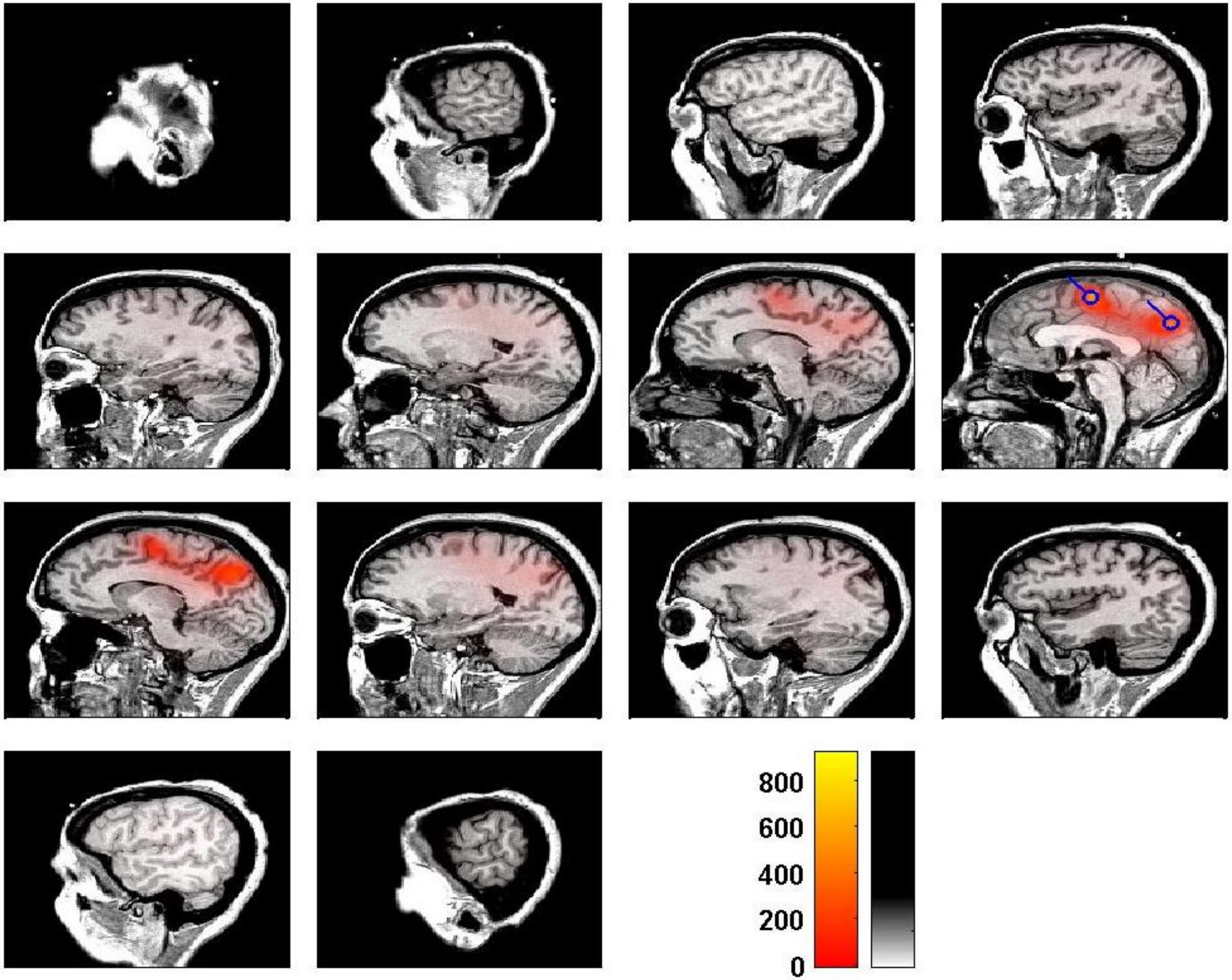
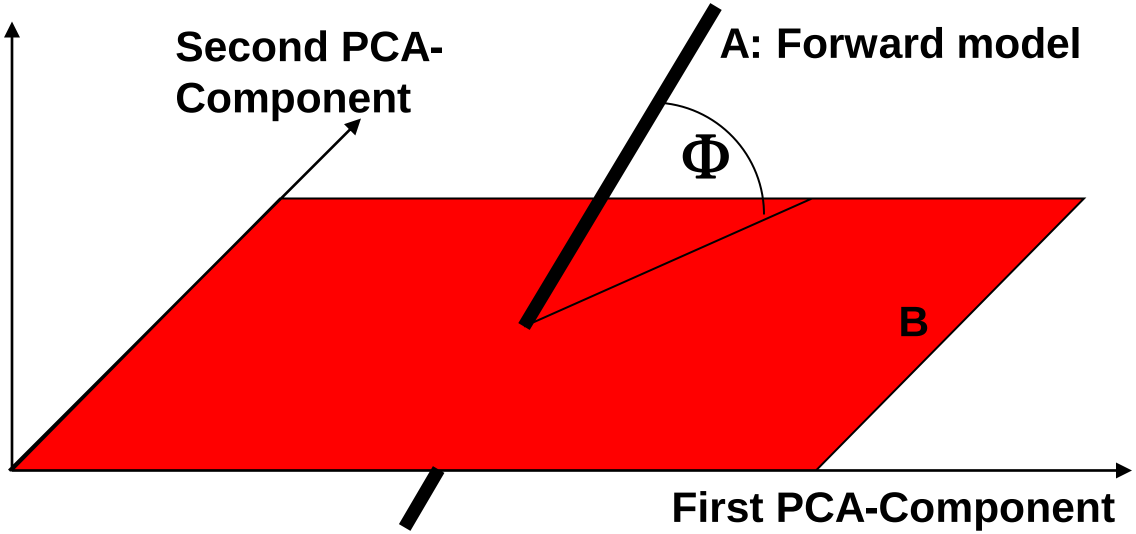


Illustration for P=2
For each voxel in the brain:



Second PCA-Component

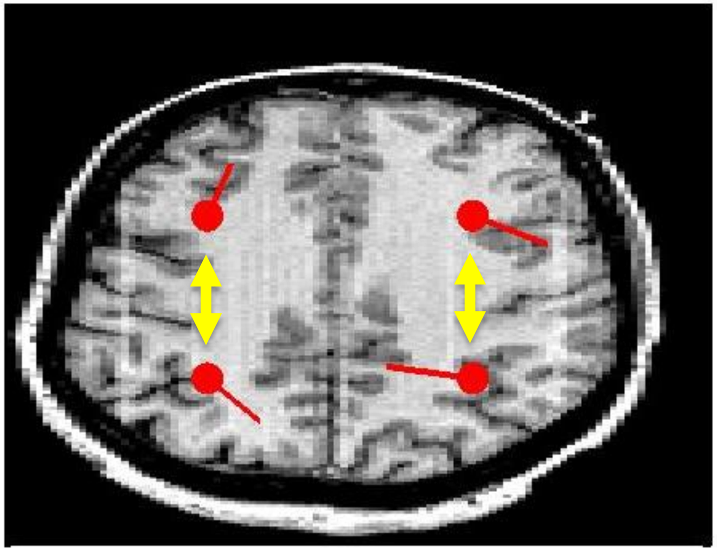
A: Forward model

Φ

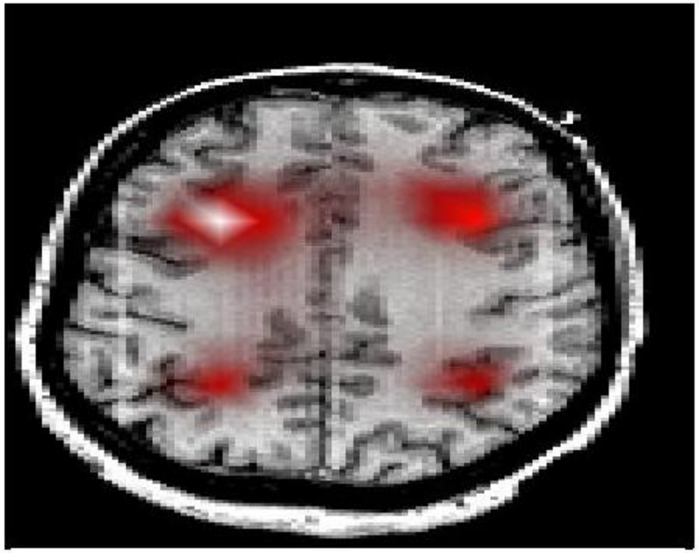
B

First PCA-Component

Display: 1/error=1/sin(Φ)



Truth



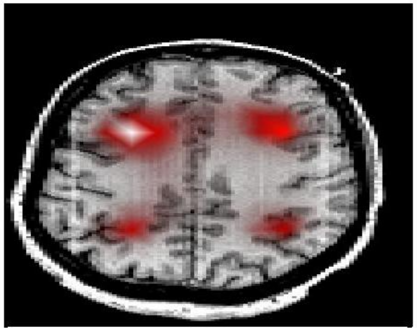
MUSIC

RAP-MUSIC

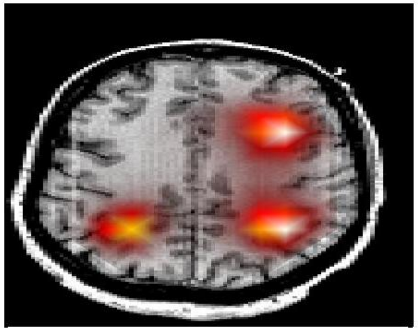
Recursively applied MUSIC

Project out maxima iteratively

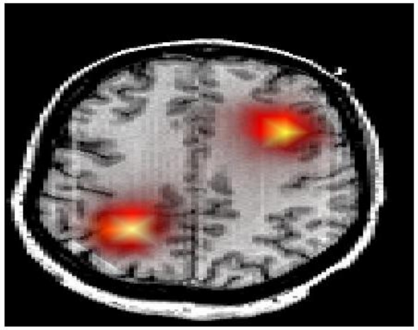
1



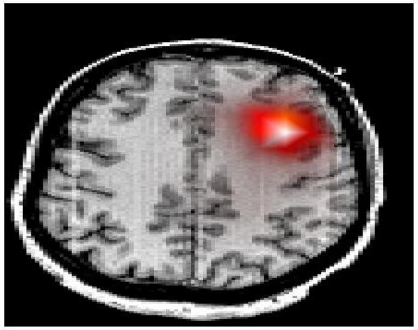
2



3



4



Source explains data

1. Dipole model
2. Minimum norm solutions

Source doesn’t explain data

1. Beamformer
2. MUSIC

The End