Advantages and problems with covariance-based source-reconstruction methods

With an emphasis on their behavior toward noise

Kensuke Sekihara

Tokyo Metropolitan Institute of Technology, Tokyo, Japan

Least-squares-based methods

Least-squares principle in which the errors between the estimates and the measurements are minimized.

Covariance-based methods

Low-rank signal assumption and the orthogonality principle.

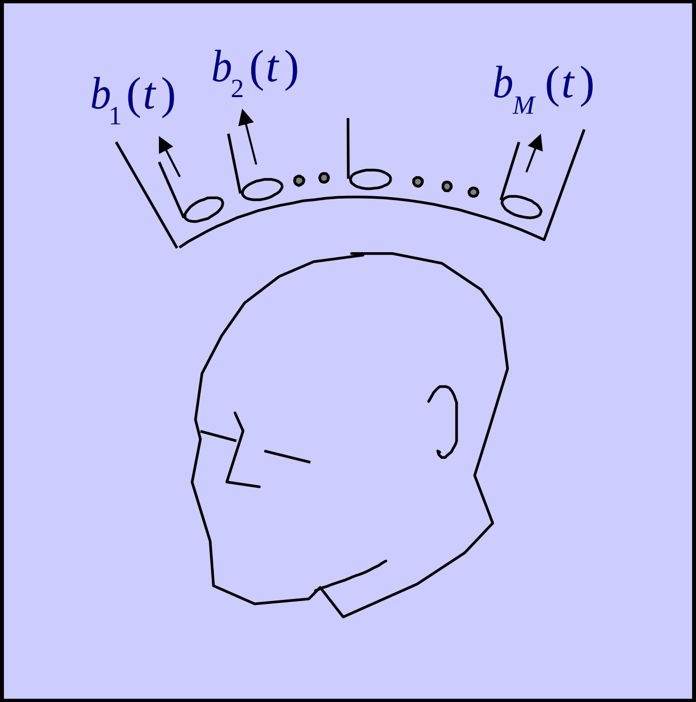
- The MUSIC Algorithm
- Adaptive spatial filter (adaptive beamformer)

Organization of my talk

- Give a brief review on the covariance-based methods.

- Discuss their behavior toward noise.

Definitions

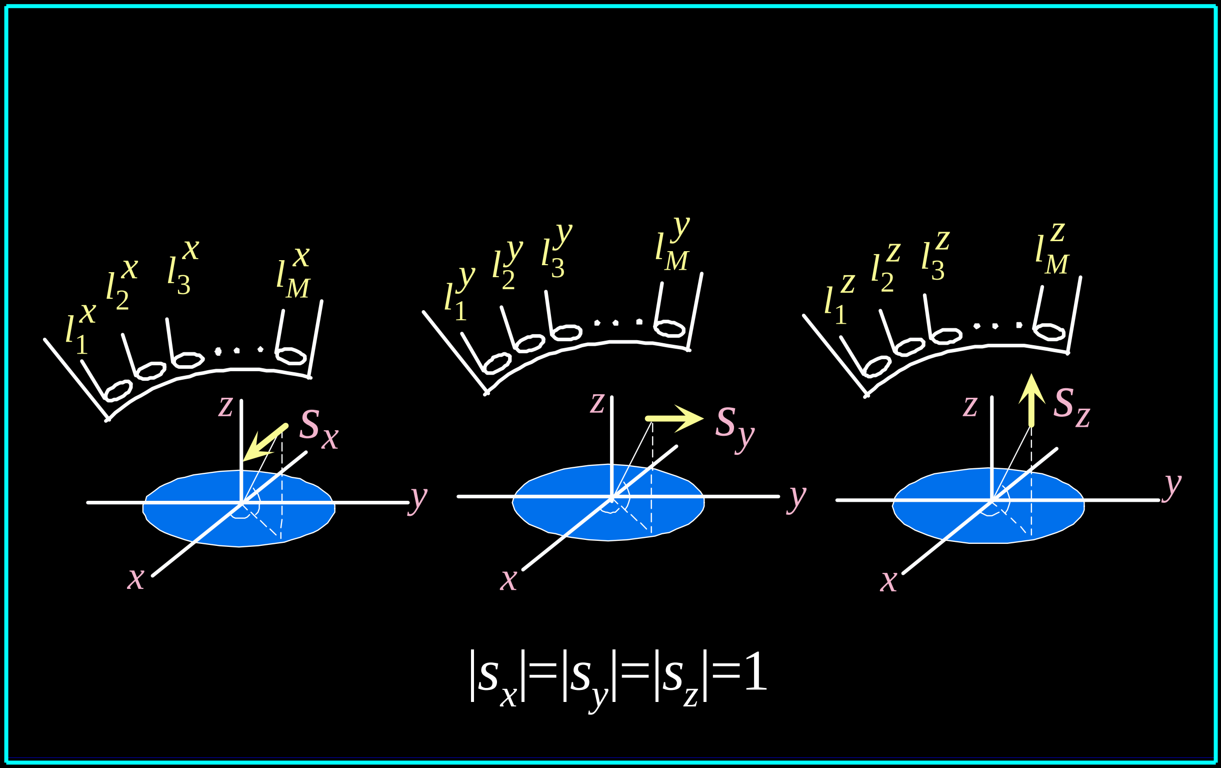


- data vector: $\boldsymbol{b}(\mathrm{t}) = \begin{bmatrix} b_1(t) \\ b_2(t) \\ \vdots \\ b_M(t) \end{bmatrix}$

- data covariance matrix: $\boldsymbol{R} = \left\langle \boldsymbol{b}(t)\boldsymbol{b}^T(t) \right\rangle$

- source magnitude: $s(\boldsymbol{r},t)$

- source orientation: $\boldsymbol{\eta}(\boldsymbol{r},t) = [\eta_x(\boldsymbol{r},t), \eta_y(\boldsymbol{r},t), \eta_z(\boldsymbol{r},t)]$



Lead field vector for the source orientation $\boldsymbol{\eta}(\boldsymbol{r})$

$$\boldsymbol{L}(\boldsymbol{r}) = \begin{bmatrix} l_1^x(\boldsymbol{r}) & l_1^y(\boldsymbol{r}) & l_1^z(\boldsymbol{r}) \\ l_2^x(\boldsymbol{r}) & l_2^y(\boldsymbol{r}) & l_2^z(\boldsymbol{r}) \\ \vdots & \vdots & \vdots \\ l_M^x(\boldsymbol{r}) & l_M^y(\boldsymbol{r}) & l_M^z(\boldsymbol{r}) \end{bmatrix}, \quad \boldsymbol{l}(\boldsymbol{r}) = \boldsymbol{L}(\boldsymbol{r}) \begin{bmatrix} \eta_x(\boldsymbol{r}) \\ \eta_y(\boldsymbol{r}) \\ \eta_z(\boldsymbol{r}) \end{bmatrix}$$

Low-rank signal assumption

Number of sensors M > Number of sources Q

$$\boldsymbol{R} = \boldsymbol{U}\left[\begin{array}{ccc|cc} \lambda_1 & 0 & \cdots & \cdot & 0 \\ 0 & \ddots & & 0 & \cdot \\ \vdots & & \lambda_Q & & \vdots \\ \hline \cdot & 0 & & \ddots & 0 \\ 0 & \cdot & \cdots & 0 & \lambda_M \end{array}\right]\boldsymbol{U}^T = \boldsymbol{U}\begin{bmatrix} \boldsymbol{\Lambda}_S & 0 \\ 0 & \boldsymbol{\Lambda}_N \end{bmatrix}\boldsymbol{U}^T$$

$$\boldsymbol{U} = [\underbrace{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_Q}_{\boldsymbol{E}_S} \mid \underbrace{\boldsymbol{e}_{Q+1}, \ldots, \boldsymbol{e}_M}_{\boldsymbol{E}_N}] = [\boldsymbol{E}_S \mid \boldsymbol{E}_N]$$

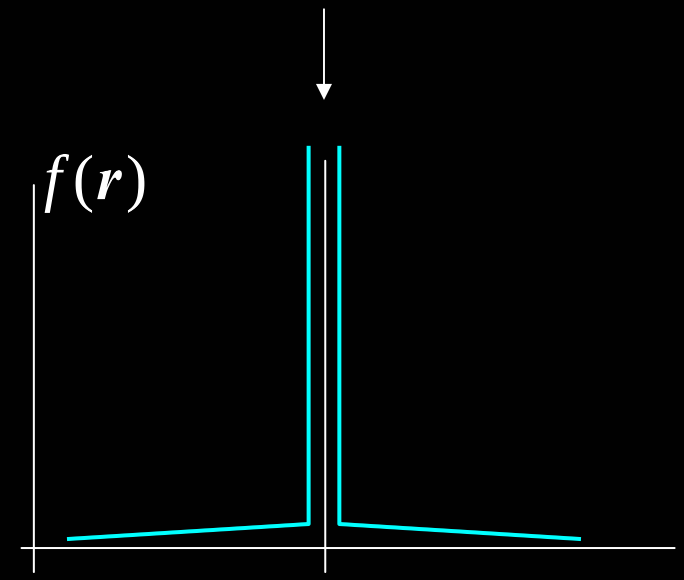
$$\boxed{\boldsymbol{\Gamma}_S^{-1} = \boldsymbol{E}_S \boldsymbol{\Lambda}_S^{-1} \boldsymbol{E}_S^T \text{ and } \boldsymbol{\Gamma}_N^{-1} = \boldsymbol{E}_N \boldsymbol{\Lambda}_N^{-1} \boldsymbol{E}_N^T \Rightarrow \boldsymbol{R}^{-1} = \boldsymbol{\Gamma}_S^{-1} + \boldsymbol{\Gamma}_N^{-1}}$$

Orthogonality principle

$$\boxed{\boldsymbol{E}_N^T \boldsymbol{l}(\boldsymbol{r}_q) = (\boldsymbol{\Gamma}_N^{-1})^T \boldsymbol{l}(\boldsymbol{r}_q) = 0 \text{ at any source location } \boldsymbol{r}_q}$$

The MUSIC algorithm

source location



$$f(\boldsymbol{r}) \approx \frac{1}{\boldsymbol{l}^T(\boldsymbol{r})\boldsymbol{E}_N \boldsymbol{E}_N^T \boldsymbol{l}(\boldsymbol{r})}$$

This term becomes zero at source locations

MUSIC algorithm chooses each location where the localizer has a very large value as one source location.

Spatial filter

$$\hat{s}(\boldsymbol{r},t) = \boldsymbol{w}^T(\boldsymbol{r})\boldsymbol{b}(t) = [w_1(\boldsymbol{r}),\ldots,w_M(\boldsymbol{r})]\begin{bmatrix} b_1(t) \\ \vdots \\ b_M(t) \end{bmatrix} = \sum_{m=1}^{M} w_m(r)b_m(t)$$

$\Uparrow$ estimate of $\boldsymbol{s}(\boldsymbol{r},t)$

$\Uparrow$ weight vector

Minimum-variance spatial filter

$$\min_{\boldsymbol{w}} \boldsymbol{w}^T \boldsymbol{R}\boldsymbol{w} \text{ subject to } \boldsymbol{w}^T\boldsymbol{l}(\boldsymbol{r}) = 1 \quad \Rightarrow \quad \boldsymbol{w}^T(\boldsymbol{r}) = \frac{\boldsymbol{l}^T(\boldsymbol{r})\boldsymbol{R}^{-1}}{\boldsymbol{l}^T(\boldsymbol{r})\boldsymbol{R}^{-1}\boldsymbol{l}(\boldsymbol{r})}$$

$$\left\langle \hat{s}(\boldsymbol{r},t)^2 \right\rangle = \frac{1}{\boldsymbol{l}^T(\boldsymbol{r})\boldsymbol{R}^{-1}\boldsymbol{l}(\boldsymbol{r})}$$

With constraint: $\boldsymbol{w}^T(\boldsymbol{r}_p)\boldsymbol{l}(\boldsymbol{r}_p) = 1,$

$$\boldsymbol{w}^T(\boldsymbol{r}_p)\boldsymbol{R}\boldsymbol{w}(\boldsymbol{r}_p) = \left\langle s(\boldsymbol{r}_p,t)^2\right\rangle + \sum_{q\neq p}\left\langle s(\boldsymbol{r}_q,t)^2\right\rangle\left\|\boldsymbol{w}^T(\boldsymbol{r}_p)\boldsymbol{l}(\boldsymbol{r}_q)\right\|$$

$$\min_{\boldsymbol{w}}\left[\boldsymbol{w}^T(\boldsymbol{r}_p)\boldsymbol{R}\boldsymbol{w}(\boldsymbol{r}_p)\right] \Rightarrow \boldsymbol{w}^T(\boldsymbol{r}_p)\boldsymbol{l}(\boldsymbol{r}_q)=0,\ q\neq p$$

Therefore, this minimization gives the weight satisfying

$$\begin{aligned}\boldsymbol{w}^T(\boldsymbol{r}_p)\boldsymbol{l}(\boldsymbol{r}_q) &= 1 \quad \text{for } p = q \\ &= 0 \quad \text{for } p \neq q\end{aligned}$$

$\boldsymbol{w}(\boldsymbol{r}_p)$ only passes the signal from a source at $\boldsymbol{r}_p$ and blocks all other signals.

Necessity of the low-rank signal assumption for minimum-variance spatial filter

Consider a easiest case where we know locations and orientations of all sources

weight $\boldsymbol{w}(\boldsymbol{r}_1)$ (containing M unknowns) can be obtained by solving a set of Q linear equations:

$$\boldsymbol{w}^T(\boldsymbol{r}_1)\boldsymbol{l}(\boldsymbol{r}_1) = w_1(\boldsymbol{r}_1)l_1(\boldsymbol{r}_1) + \ldots + w_M(\boldsymbol{r}_1)l_M(\boldsymbol{r}_1) = 1$$

$$\boldsymbol{w}^T(\boldsymbol{r}_1)\boldsymbol{l}(\boldsymbol{r}_2) = w_1(\boldsymbol{r}_1)l_1(\boldsymbol{r}_2) + \ldots + w_M(\boldsymbol{r}_1)l_M(\boldsymbol{r}_2) = 0$$

$$\vdots$$

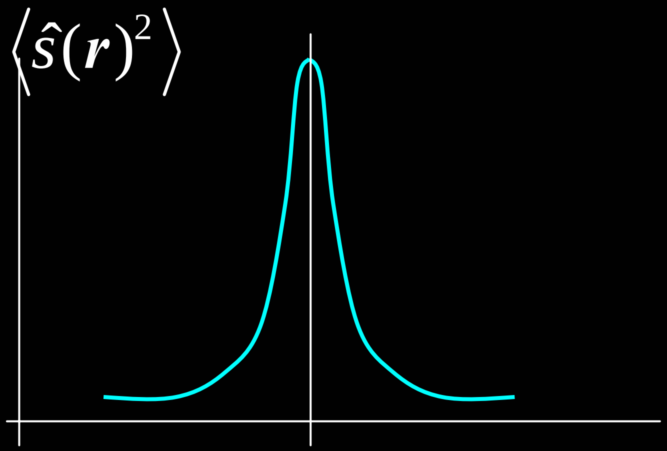
$$\boldsymbol{w}^T(\boldsymbol{r}_1)\boldsymbol{l}(\boldsymbol{r}_Q) = w_1(\boldsymbol{r}_1)l_1(\boldsymbol{r}_Q) + \ldots + w_M(\boldsymbol{r}_1)l_M(\boldsymbol{r}_Q) = 0$$

when $Q > M$, there is no solution for $\boldsymbol{w}^T(\boldsymbol{r}_1)$

Minimum variance spatial filter output:

$$\left\langle \hat{s}(\boldsymbol{r})^2 \right\rangle = \frac{1}{\boldsymbol{l}^T(\boldsymbol{r})\boldsymbol{R}^{-1}\boldsymbol{l}(\boldsymbol{r})}$$

$$= \frac{1}{\boldsymbol{l}^T(\boldsymbol{r})\boldsymbol{\Gamma}_S^{-1}\boldsymbol{l}(\boldsymbol{r}) + \boldsymbol{l}^T(\boldsymbol{r})\boldsymbol{\Gamma}_N^{-1}\boldsymbol{l}(\boldsymbol{r})}$$



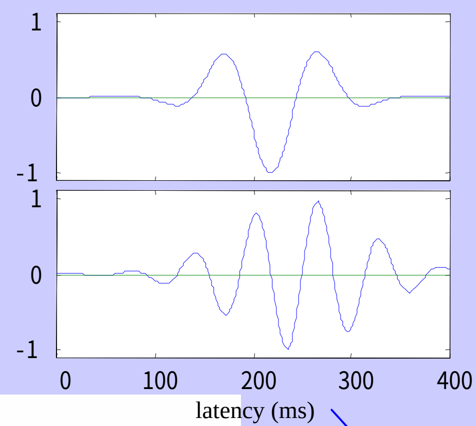
This term becomes zero at source locations

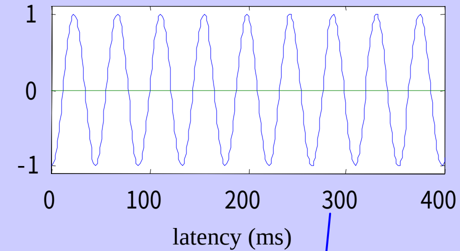
Eigen decomposition of $\boldsymbol{R}$:

$$\boldsymbol{R}^{-1} = \sum_{j=1}^{M} \frac{1}{\lambda_j}\boldsymbol{e}_j\boldsymbol{e}_j^T = \sum_{j=1}^{Q} \frac{1}{\lambda_j}\boldsymbol{e}_j\boldsymbol{e}_j^T + \frac{1}{\lambda_{\text{noise}}}\sum_{j=Q+1}^{M}\boldsymbol{e}_j\boldsymbol{e}_j^T \approx \boldsymbol{E}_N\boldsymbol{E}_N^T \quad (\lambda_j \gg \lambda_{\text{noise}})$$

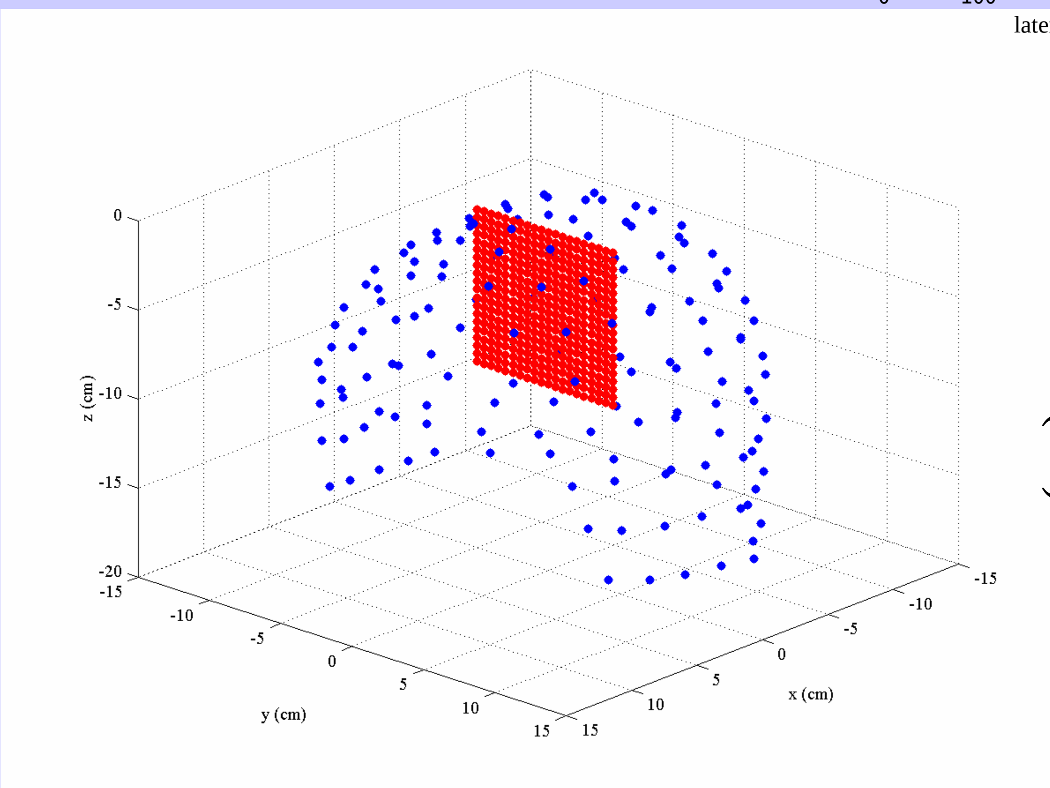
$$\left\langle s(\boldsymbol{r})^2 \right\rangle = \frac{1}{\boldsymbol{l}^T(\boldsymbol{r})\boldsymbol{R}^{-1}\boldsymbol{l}(\boldsymbol{r})} \approx \frac{1}{\boldsymbol{l}^T(\boldsymbol{r})\boldsymbol{E}_N\boldsymbol{E}_N^T\boldsymbol{l}(\boldsymbol{r})} \quad \text{(MUSIC localizer)}$$

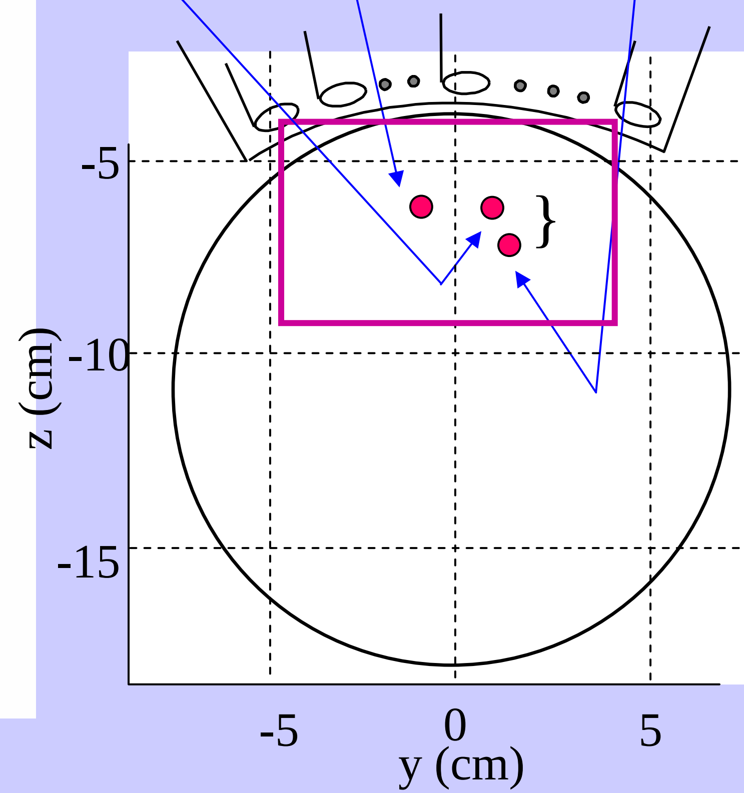
assumed source waveform



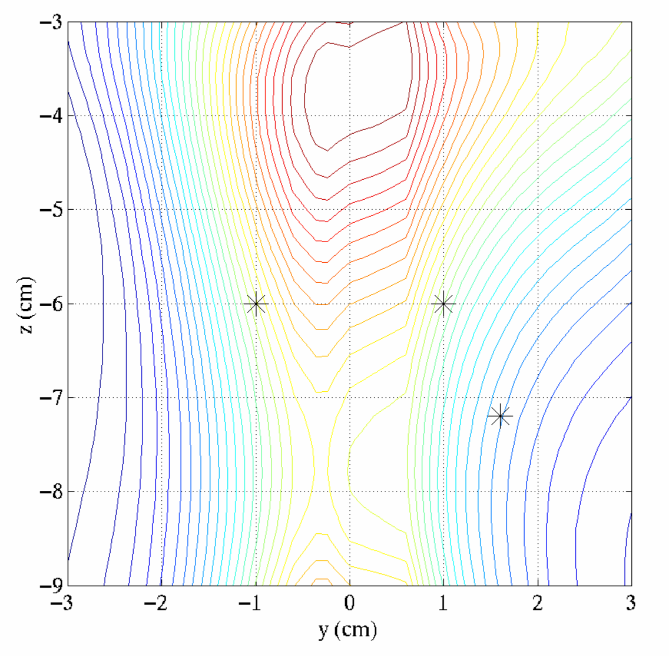


148-channel sensor array

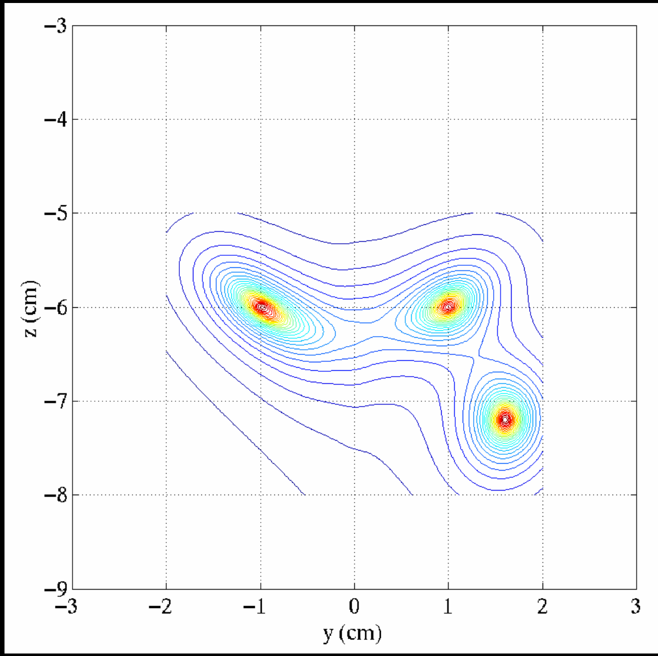




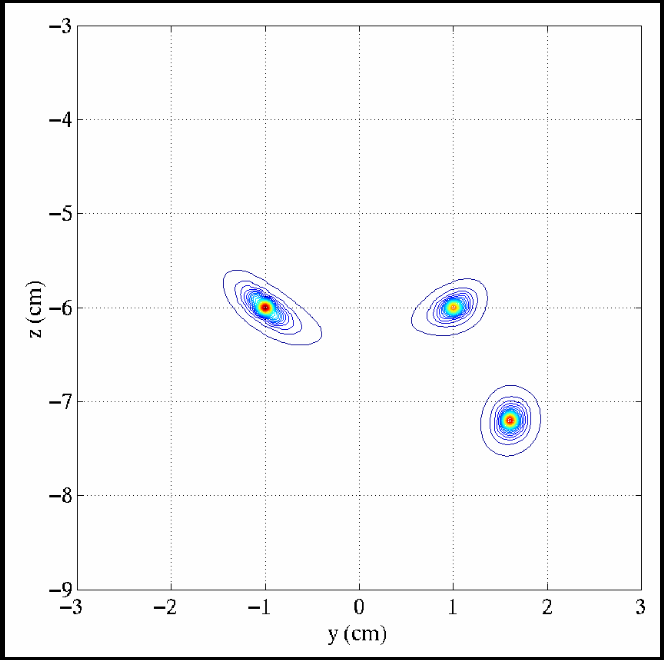
Minimum-norm reconstruction



Minimum-variance spatial filter reconstruction



MUSIC reconstruction



The behavior of covariance-based methods toward various types of noise are significantly different from that of the least-squares-based methods.

My talk introduces several interesting behavior of the covariance-based methods toward noise.

- My arguments use the minimum-variance spatial filter as a representative method among covariance-based methods.
- Due to the similarity, most of my arguments can be applied to the MUSIC algorithm.

Noise in measurements

signal source of interest

sensor noise

$$\boldsymbol{b}(t) = \boldsymbol{L}\sum_{q=1}^{Q}\boldsymbol{s}(\boldsymbol{r}_q, t) + \boldsymbol{L}\sum_{k=1}^{K}\boldsymbol{\xi}(\boldsymbol{r}_k, t) + \boldsymbol{d}(t) + \boldsymbol{n}(t)$$

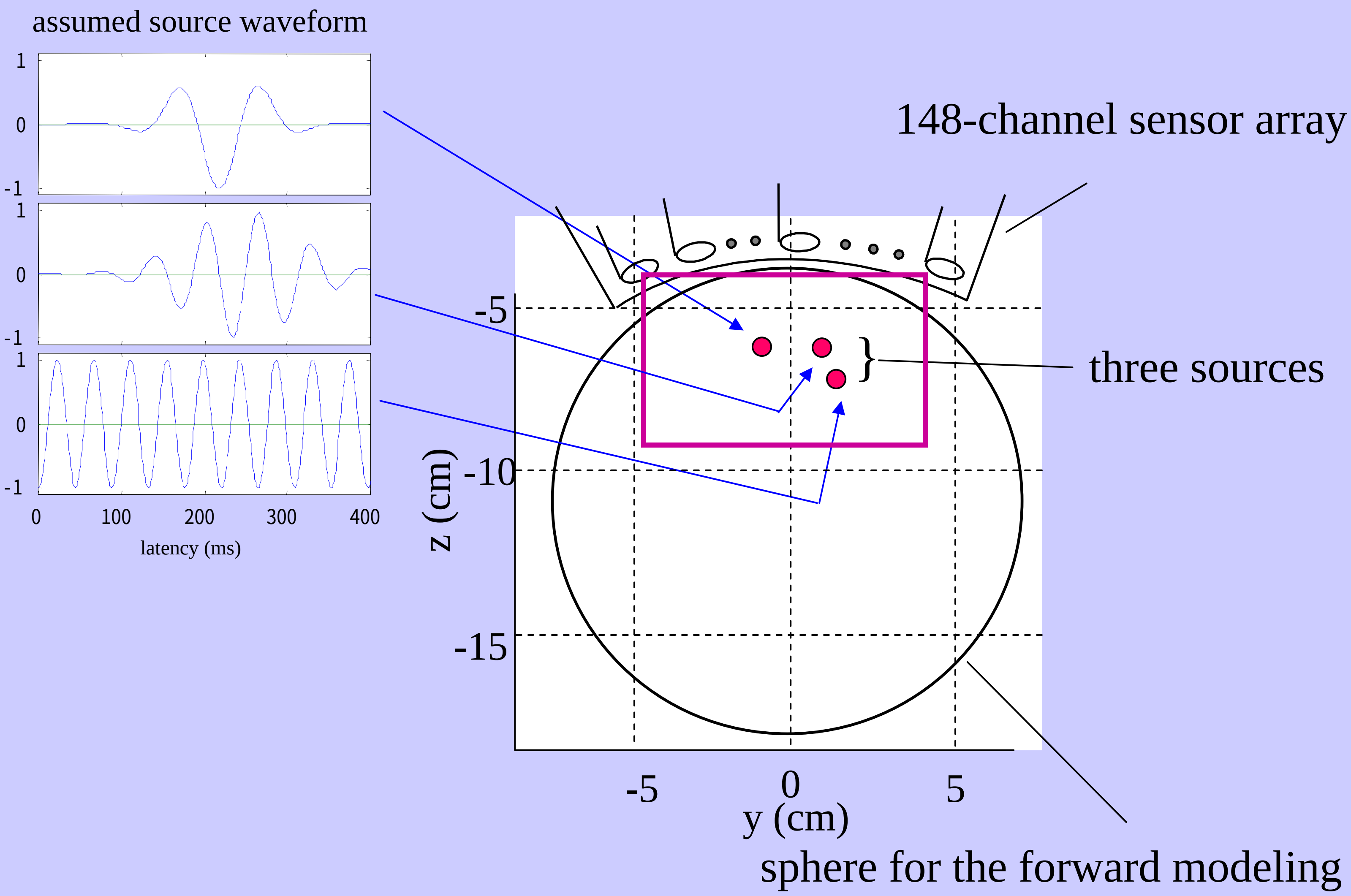
neurophysiological noise

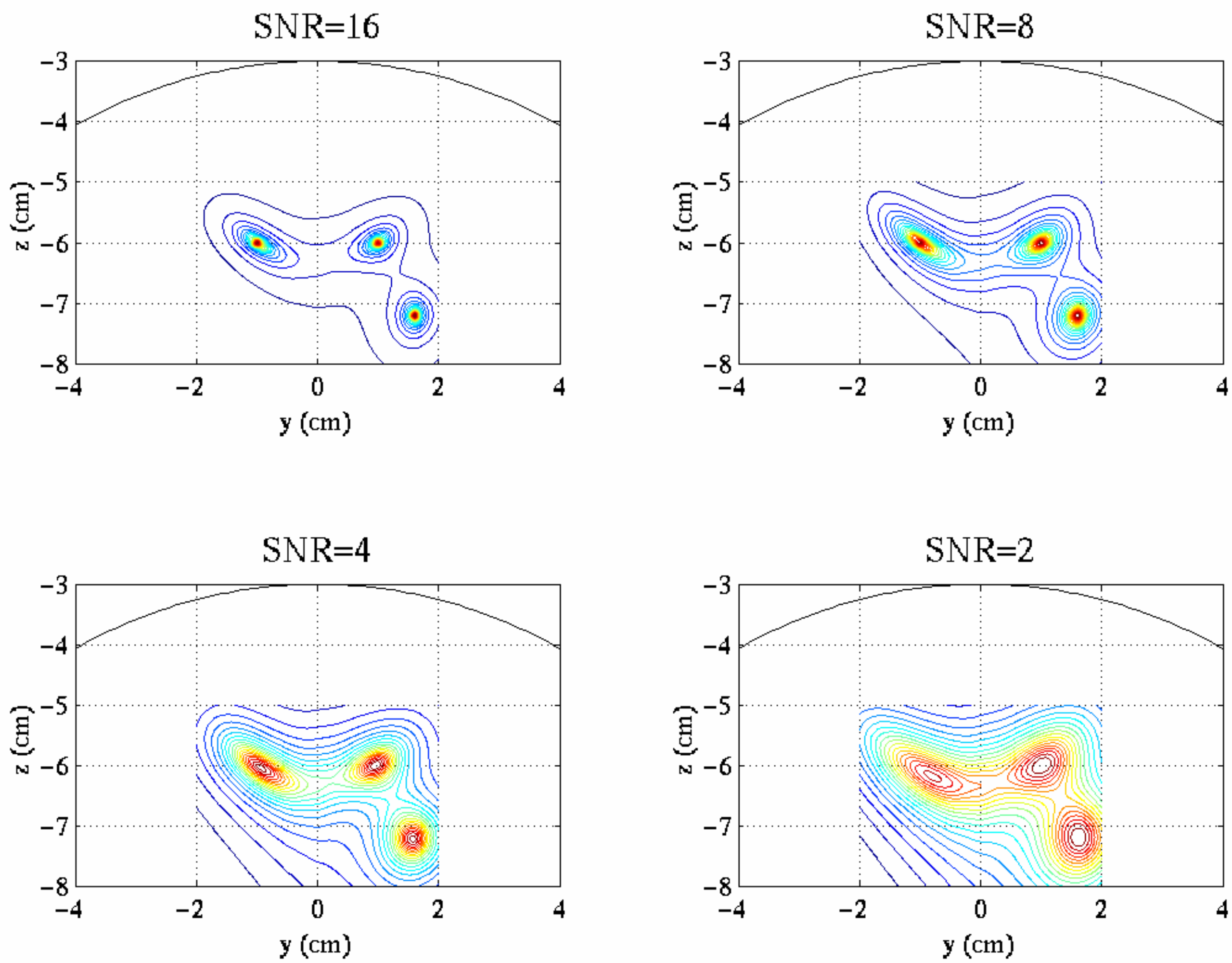
external disturbances

Sensor noise

- can be modeled by white Gaussian noise
- uncorrelated among sensor channels

Sensor noise causes the spatial resolution degradation


assumed source waveform
1
0
-1
1
0
-1
1
0
-1
0
100
200
300
400
latency (ms)
148-channel sensor array
three sources
-5
-10
-15
z (cm)
-5
0
5
y (cm)
sphere for the forward modeling


SNR=16
SNR=8
SNR=4
SNR=2
z (cm)
y (cm)
−3
−4
−5
−6
−7
−8
−4
−2
0
2
4

Resolution kernel: $\hat{s}(\boldsymbol{r}) = \int K(\boldsymbol{r}, \boldsymbol{r}')s(\boldsymbol{r}')d\boldsymbol{r}'$

When a single source exists at $\boldsymbol{r}_1$,

$$K(\boldsymbol{r}, \boldsymbol{r}_1) = \frac{\|\boldsymbol{l}(\boldsymbol{r}_1)\|}{\|\boldsymbol{l}(\boldsymbol{r})\|} \frac{\cos[\boldsymbol{l}(\boldsymbol{r}), \boldsymbol{l}(\boldsymbol{r}_1)]}{[1 + (\mathrm{SNR})\sin^2[\boldsymbol{l}(\boldsymbol{r}), \boldsymbol{l}(\boldsymbol{r}_1)]]},$$

$\Uparrow$

Input power SNR

where $\cos^2(\boldsymbol{a}, \boldsymbol{b}) = \left|\boldsymbol{a}^T \boldsymbol{b}\right|^2 \Big/ [(\boldsymbol{a}^T \boldsymbol{a})(\boldsymbol{b}^T \boldsymbol{b})],$

$$\sin^2(\boldsymbol{a}, \boldsymbol{b}) = 1 - \cos^2(\boldsymbol{a}, \boldsymbol{b})$$

The detail of the analysis on the spatial resolution degradation and some other problems related to this type of noise has been presented in our poster:

K. Sekihara *et al*. “**Spatial resolution, leakage, and signal-to-noise ratio in adaptive-beamformer source reconstruction techniques**”

External disturbances

$$\boldsymbol{b}(t) = \boldsymbol{L}\sum_{q=1}^{Q} \boldsymbol{s}(\boldsymbol{r}_q, t) + \boldsymbol{L}\sum_{k=1}^{K} \boldsymbol{\xi}(\boldsymbol{r}_k, t) + \boldsymbol{d}(t) + \boldsymbol{n}(t)$$

$\boldsymbol{d}(t)$ may includes:

- power-line interferences
- Base-line drift
- MCG artifacts

$\Downarrow$

Low rank

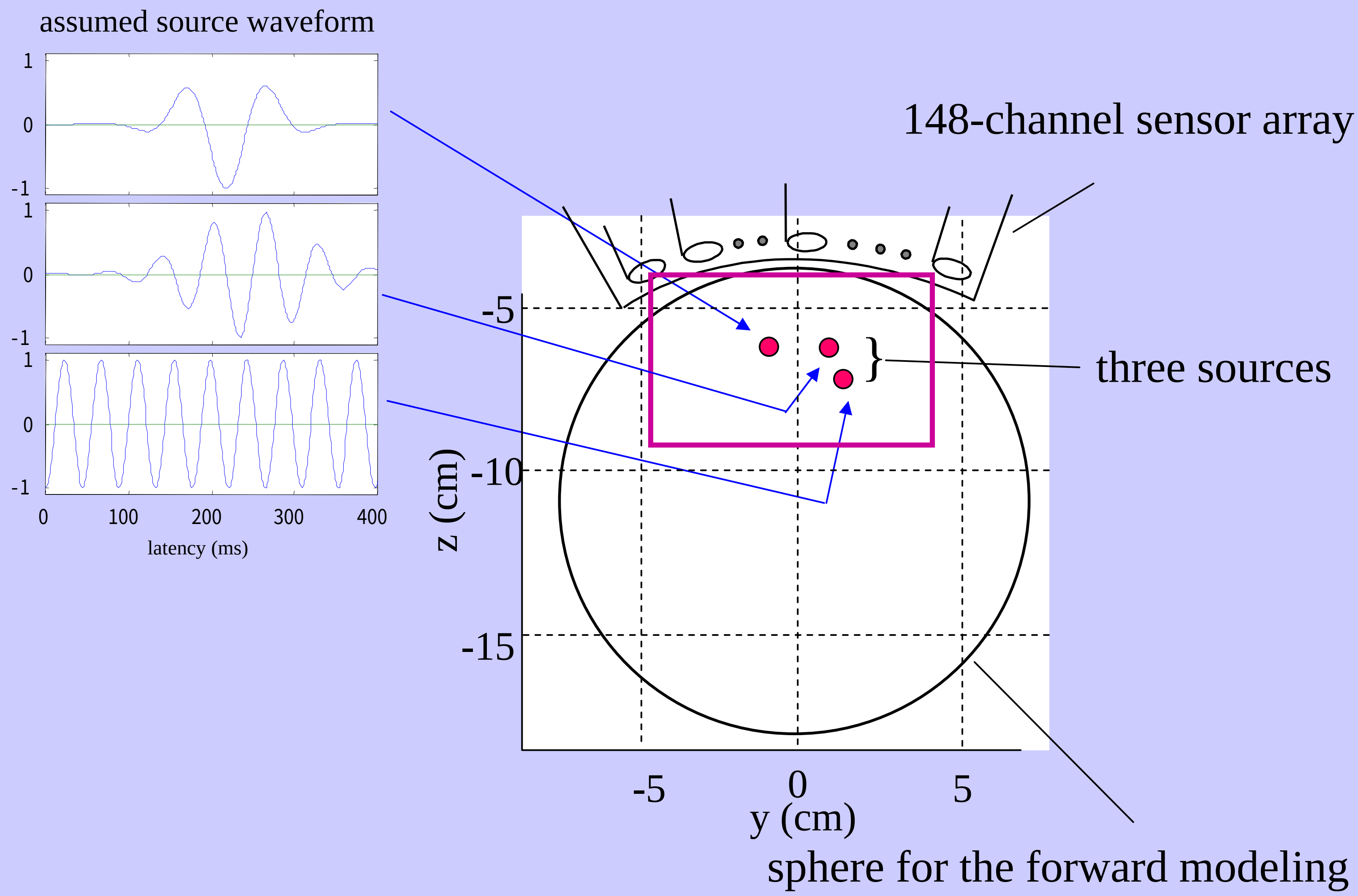
Their spatio-temporal activities have small number of significantly large eigenvalues.

Simulated disturbances

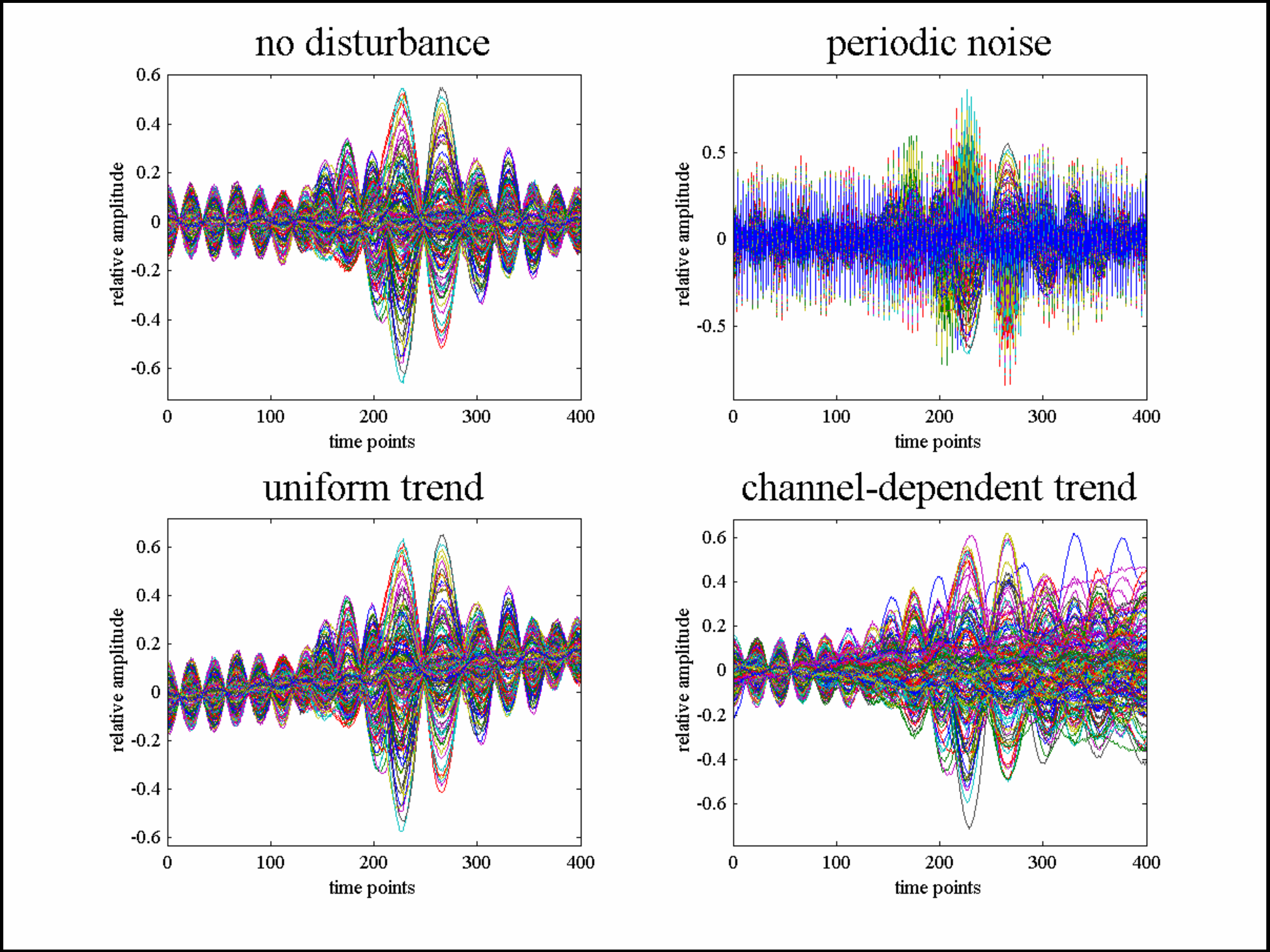
- Case1: Recordings from right hemisphere channels (total 60 channels) contain the same periodic noise.

- Case2: All channel recordings have uniform linear trends

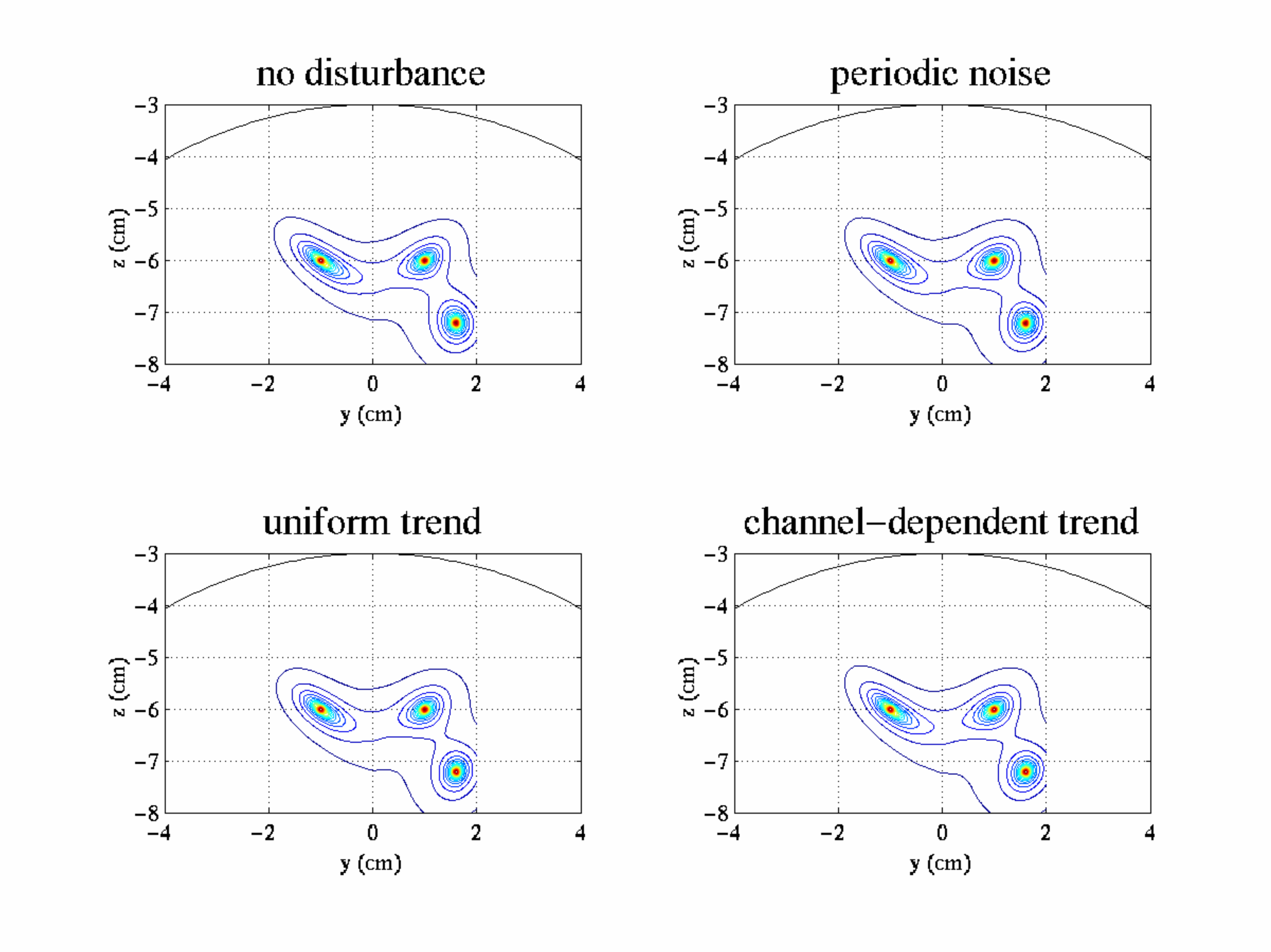
- Case3: Each channel has its own linear trend different to each other


assumed source waveform
1
0
-1
1
0
-1
1
0
-1
0
100
200
300
400
latency (ms)
148-channel sensor array
three sources
-5
-10
-15
z (cm)
-5
0
5
y (cm)
sphere for the forward modeling

Simulated recordings



Minimum-variance spatial filter reconstruction results
Signal to sensor noise ratio: 16



Low-rank disturbance

signal source activities $\downarrow$ disturbance $\downarrow$

$$\boldsymbol{b}(t) = \boldsymbol{L}\sum_{q=1}^{Q} \boldsymbol{s}(\boldsymbol{r}_q, t) + \boldsymbol{n}(t) + \boldsymbol{d}(t)$$

- Assume no correlation between $\boldsymbol{s}(\boldsymbol{r}_q, t)$ and $\boldsymbol{d}(t)$

Covariance matrices

from $\boldsymbol{L}\sum_{q=1}^{Q} \boldsymbol{s}(\boldsymbol{r}_q, t) + \boldsymbol{n}(t)$ $\downarrow$

$$\boldsymbol{R} = \boldsymbol{R}_B + \boldsymbol{R}_D$$

$\uparrow$ from $\boldsymbol{b}(t)$

$\uparrow$ from $\boldsymbol{d}(t)$: $\boldsymbol{R}_D = \left\langle \boldsymbol{d}(t)\boldsymbol{d}^T(t) \right\rangle$

When $\boldsymbol{R}_D$ is a rank one matrix,

$$\boldsymbol{R}_D = \lambda \boldsymbol{u}\boldsymbol{u}^T$$

$$\Downarrow$$

We can derive,

$$\boldsymbol{l}^T(\boldsymbol{r})\boldsymbol{R}^{-1}\boldsymbol{l}(\boldsymbol{r}) = \boldsymbol{l}^T(\boldsymbol{r})\boldsymbol{R}_B^{-1}\boldsymbol{l}(\boldsymbol{r})\left[1 - \cos^2(\boldsymbol{l},\boldsymbol{u} \mid \boldsymbol{R}_B^{-1})\right]$$

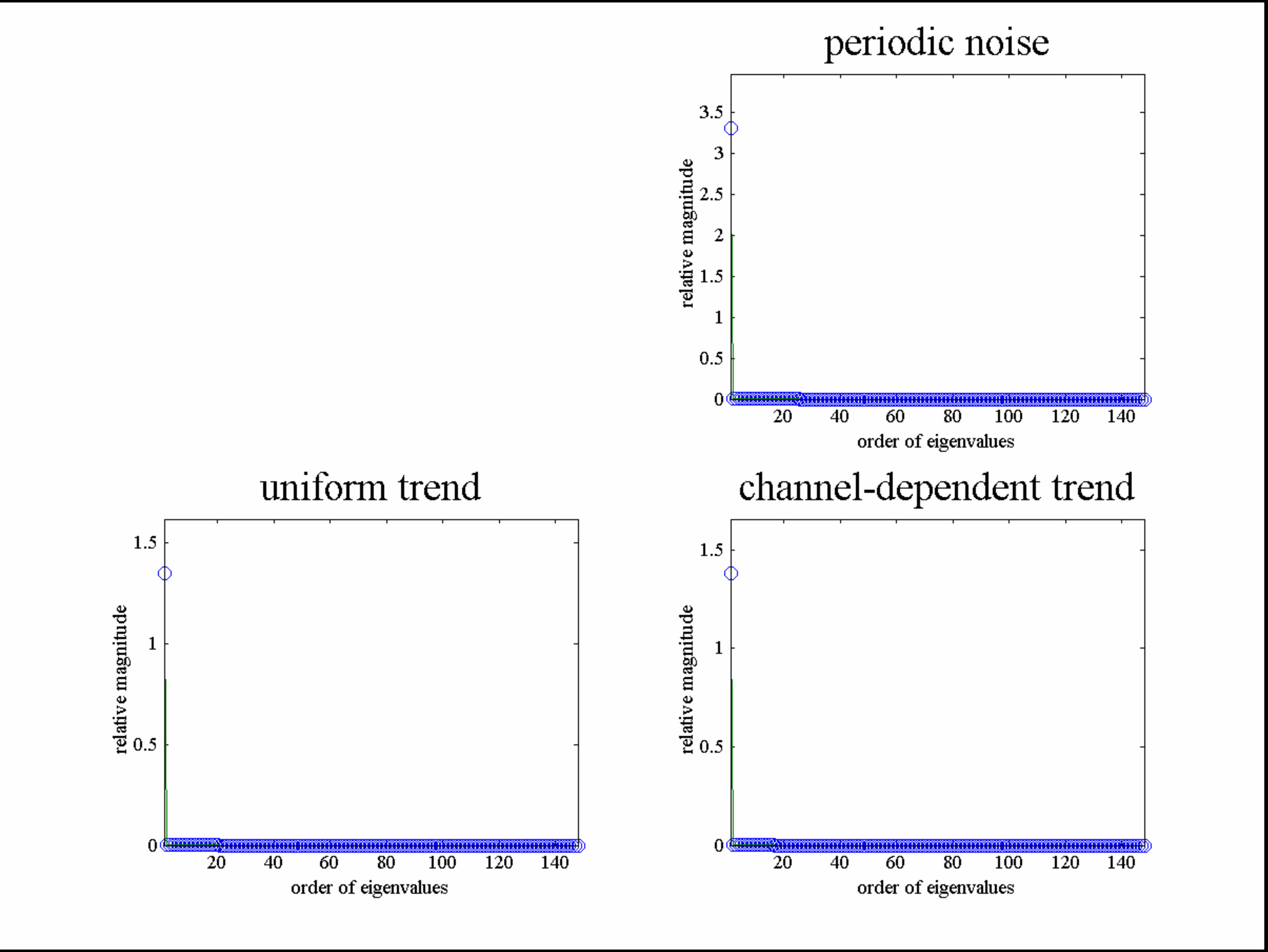
$$\Uparrow$$

generalized cosine between $\boldsymbol{l}$ and $\boldsymbol{u}$ with the metric $\boldsymbol{R}_B^{-1}$

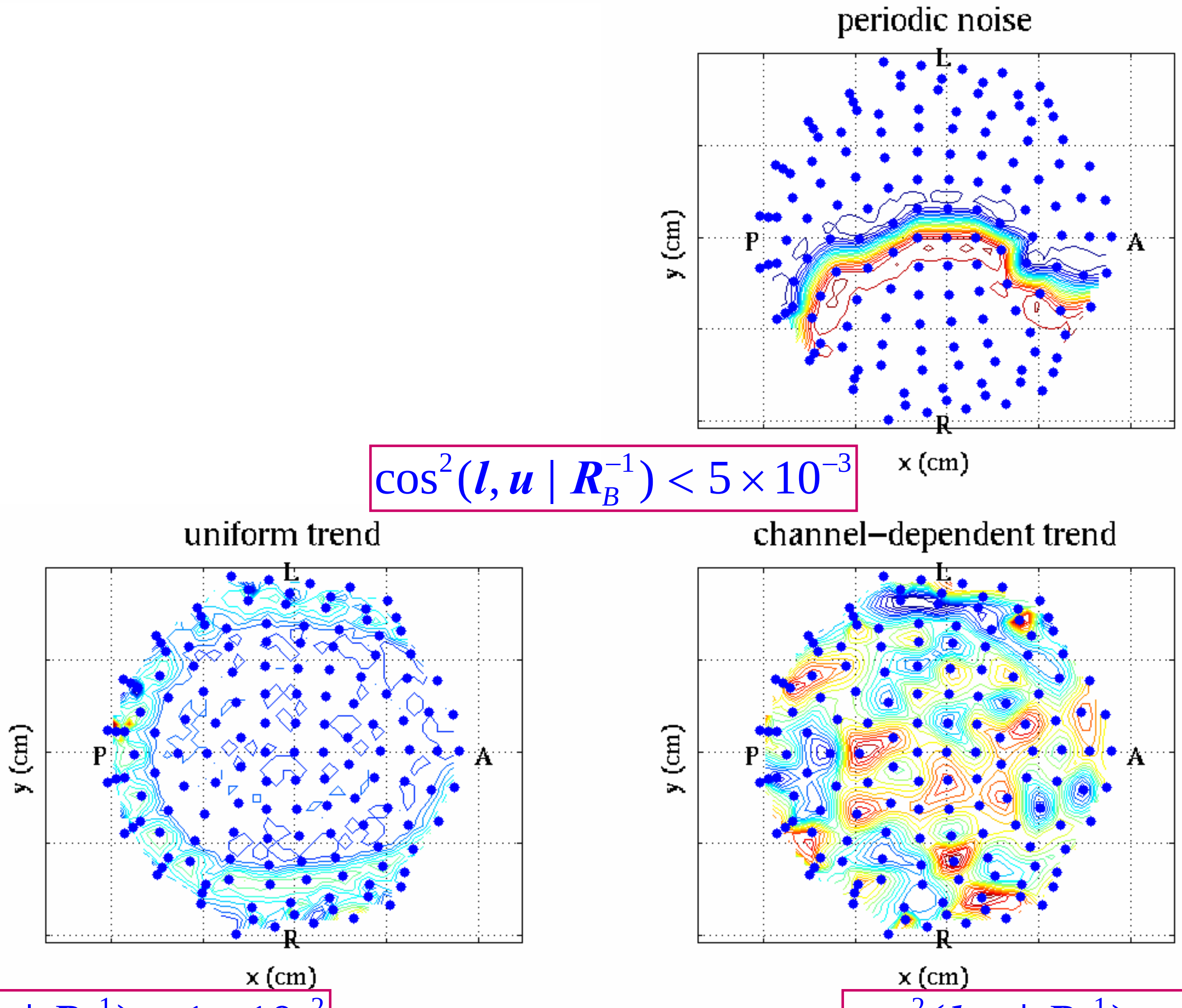
When $\boldsymbol{l}$ and $\boldsymbol{u}$ are very different, $\cos^2(\boldsymbol{l},\boldsymbol{u} \mid \boldsymbol{R}_B^{-1}) \ll 1$, and

$$\boxed{\left\langle s(\boldsymbol{r})^2 \right\rangle = \frac{1}{\boldsymbol{l}^T(\boldsymbol{r})\boldsymbol{R}^{-1}\boldsymbol{l}(\boldsymbol{r})} \approx \frac{1}{\boldsymbol{l}^T(\boldsymbol{r})\boldsymbol{R}_B^{-1}\boldsymbol{l}(\boldsymbol{r})}}$$

Eigenspectrum of $\boldsymbol{R}_D$



Visualization of the first eigenvector of the disturbances



$\cos^2(\boldsymbol{l}, \boldsymbol{u} \mid \boldsymbol{R}_B^{-1}) < 5 \times 10^{-3}$

$\cos^2(\boldsymbol{l}, \boldsymbol{u} \mid \boldsymbol{R}_B^{-1}) < 1 \times 10^{-2}$

$\cos^2(\boldsymbol{l}, \boldsymbol{u} \mid \boldsymbol{R}_B^{-1}) < 2 \times 10^{-3}$

Neurophysiological noise

$$\boldsymbol{b}(t) = \boldsymbol{L}\sum_{q=1}^{Q} \boldsymbol{s}(\boldsymbol{r}_q, t) + \boldsymbol{L}\sum_{k=1}^{K} \boldsymbol{\xi}(\boldsymbol{r}_k, t) + \boldsymbol{n}(t)$$

closely related to the resting state of the brain or the default mode of brain activities.

Neurophysiological noise can be modeled by randomly distributed incoherent dipoles.

•de Munck et al., IEEE Trans. Biomed. Eng., 39, 791-804, 1992.
•Valdes et al., Brain Topography, 4, 309-319, 1992.
•Lutkenhoner, J. Appl. Phys., 75, 7204-7210, 1994.

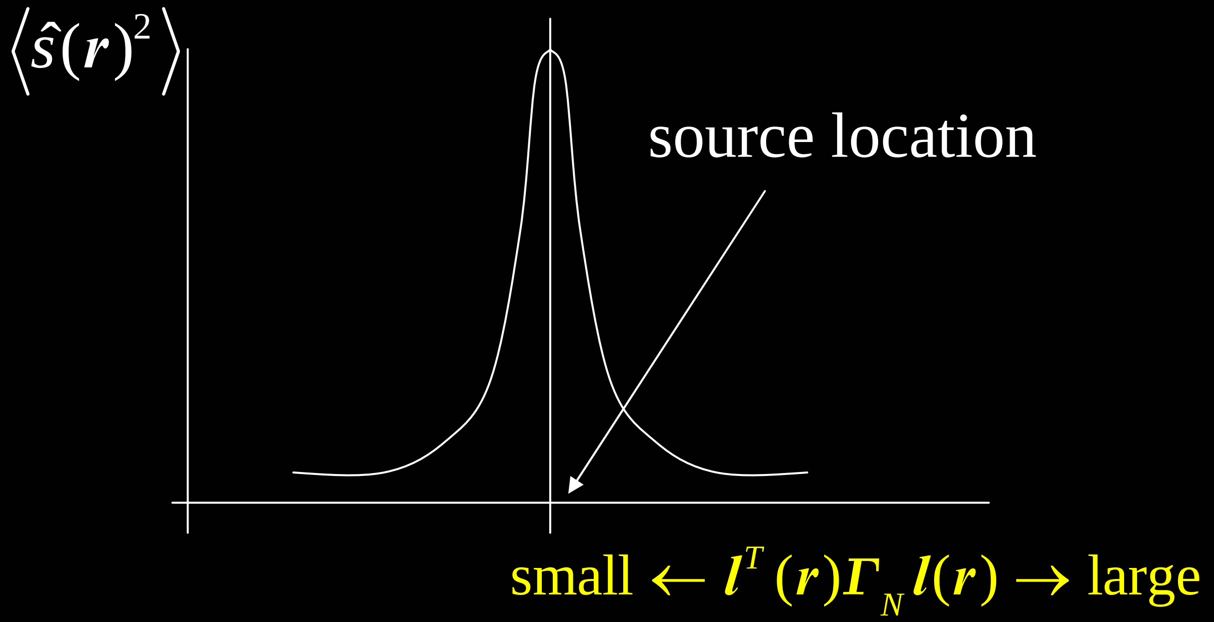
This type of noise may invalidate the low-rank signal assumption;

Number of sensors M > Number of sources P

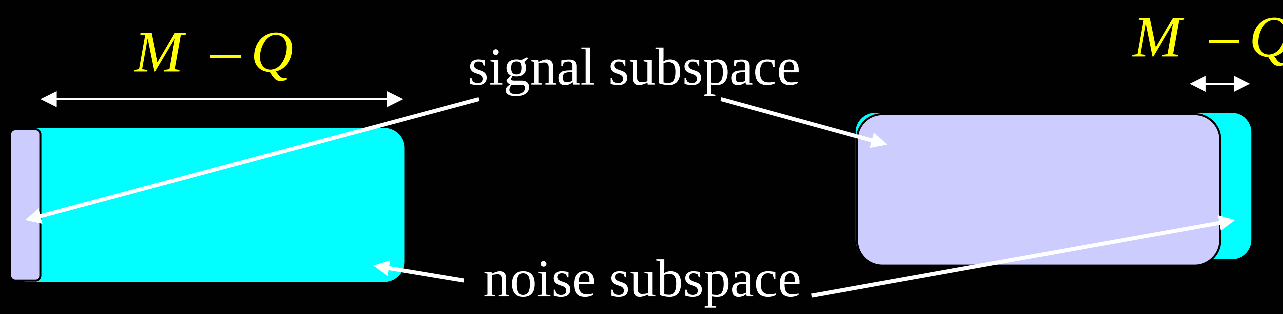
Even when the low-rank signal assumption is satisfied, the size of noise subspace affects the spatial resolution of the reconstructed results.

Minimum-variance spatial filter output:

$$\left\langle \hat{s}(\boldsymbol{r})^2 \right\rangle = \frac{1}{\boldsymbol{l}^T(\boldsymbol{r})\boldsymbol{R}^{-1}\boldsymbol{l}(\boldsymbol{r})} = \frac{1}{\boldsymbol{l}^T(\boldsymbol{r})\boldsymbol{\Gamma}_S^{-1}\boldsymbol{l}(\boldsymbol{r}) + \boldsymbol{l}^T(\boldsymbol{r})\boldsymbol{\Gamma}_N^{-1}\boldsymbol{l}(\boldsymbol{r})}$$



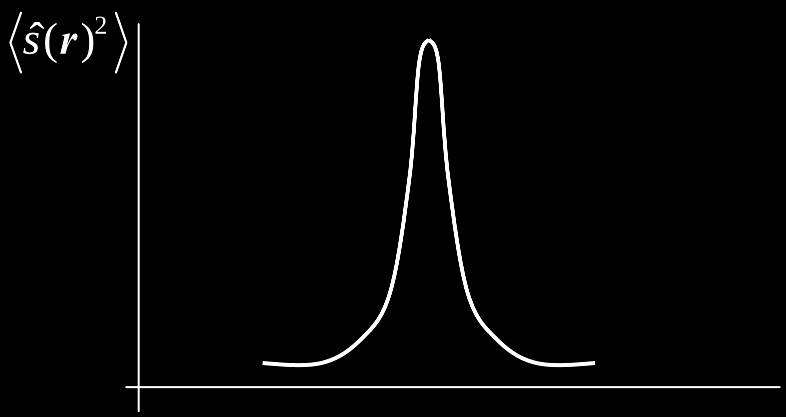
$$\boldsymbol{l}^T(\boldsymbol{r})\boldsymbol{\Gamma}_N\boldsymbol{l}(\boldsymbol{r}) \approx \sum_{j=Q+1}^{M} \left\| \boldsymbol{l}^T(\boldsymbol{r})\boldsymbol{e}_j \right\|^2$$



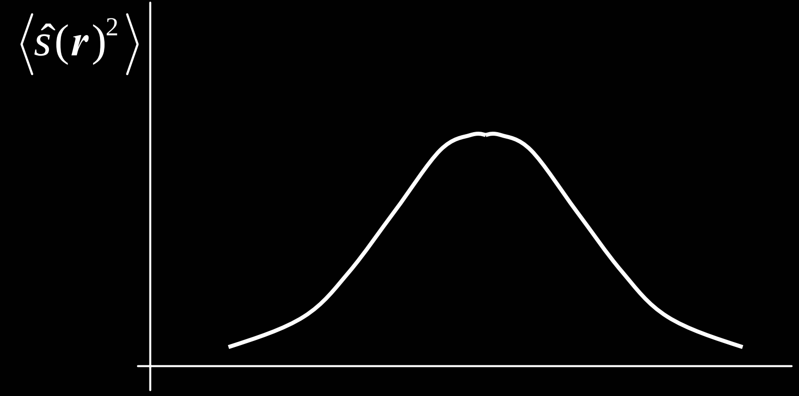
When $\boldsymbol{r} \neq$ source location,

$$\boldsymbol{l}^T(\boldsymbol{r})\boldsymbol{\Gamma}_N\boldsymbol{l}(\boldsymbol{r}) \quad > \quad \boldsymbol{l}^T(\boldsymbol{r})\boldsymbol{\Gamma}_N\boldsymbol{l}(\boldsymbol{r})$$

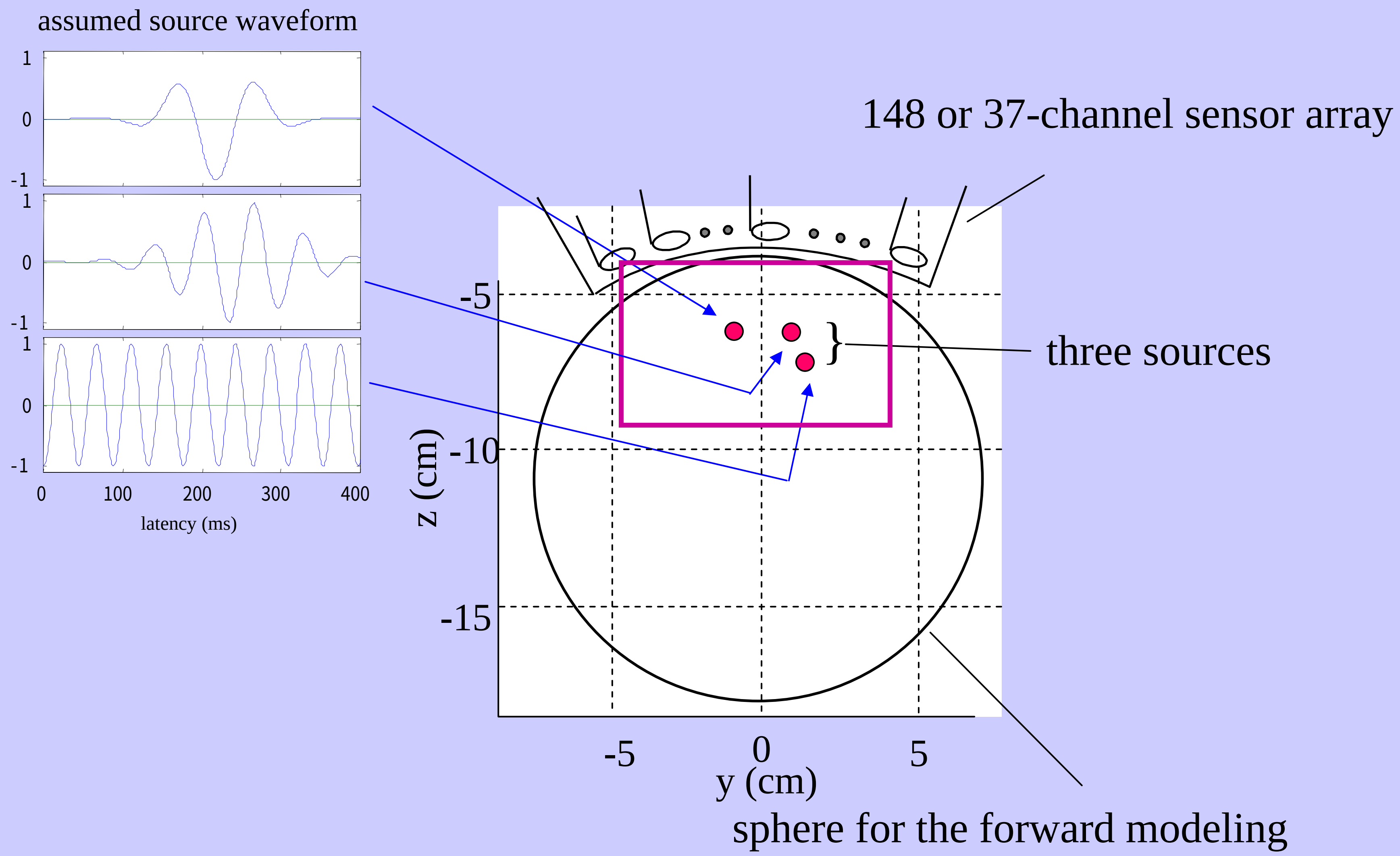
high spatial resolution



low spatial resolution

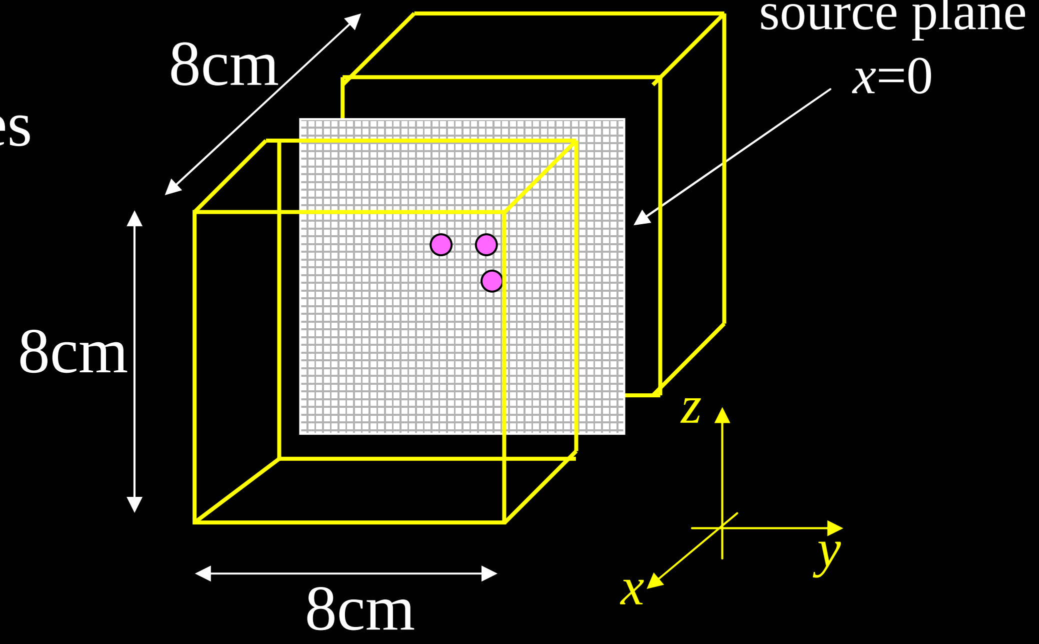


The spatial resolution depends on the size of the noise subspace


assumed source waveform
1
0
-1
1
0
-1
1
0
-1
0
100
200
300
400
latency (ms)
148 or 37-channel sensor array
three sources
-5
-10
-15
z (cm)
-5
0
5
y (cm)
sphere for the forward modeling

Generate many random dipoles in a volume:

$$-4 < x < -1, \quad 1 < x < 4$$
$$-4 < y < 4$$
$$-10 < z < -2$$



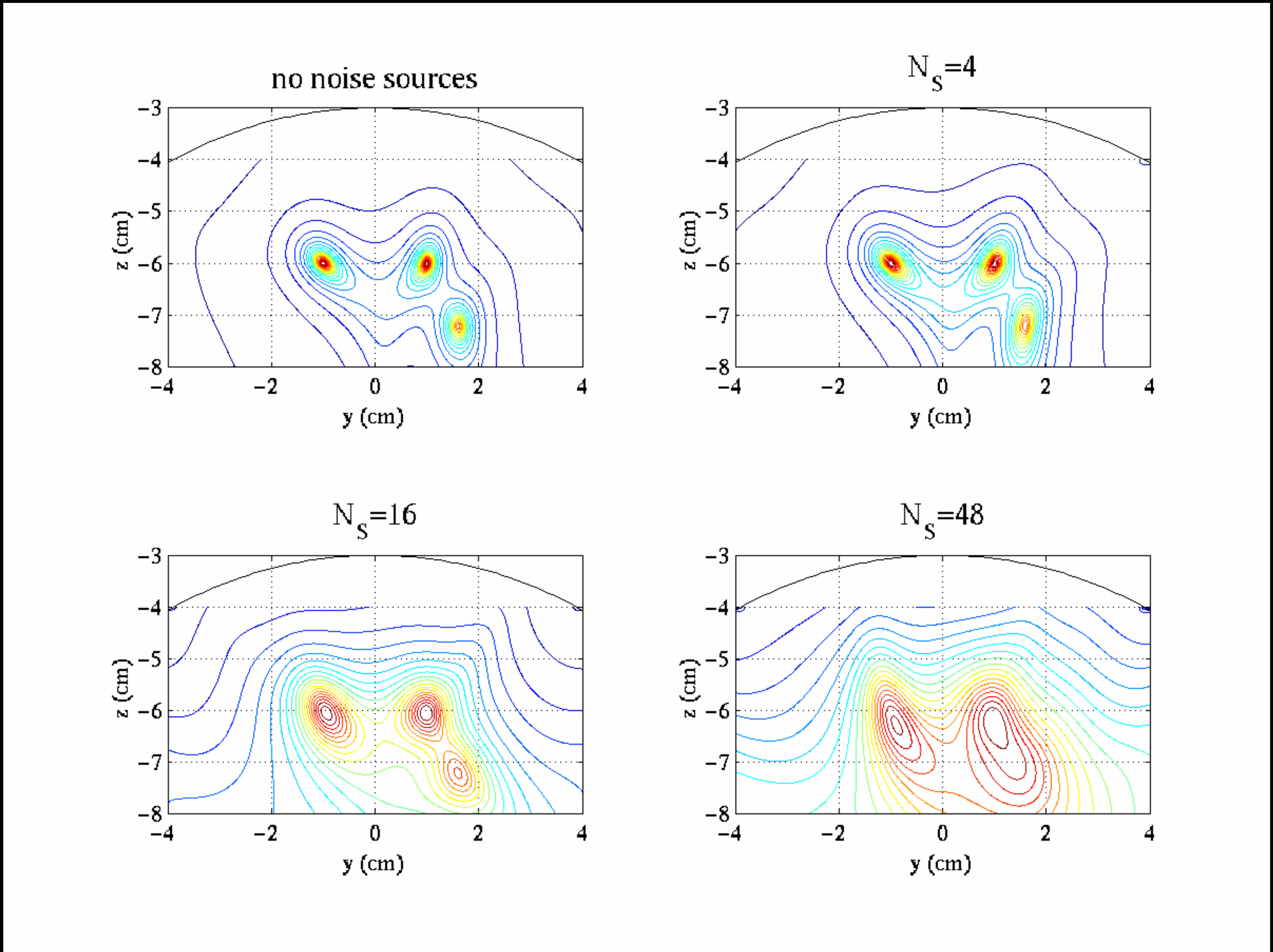
N_S : Number of noise random dipoles

P_N : The power of noise dipoles is fixed at

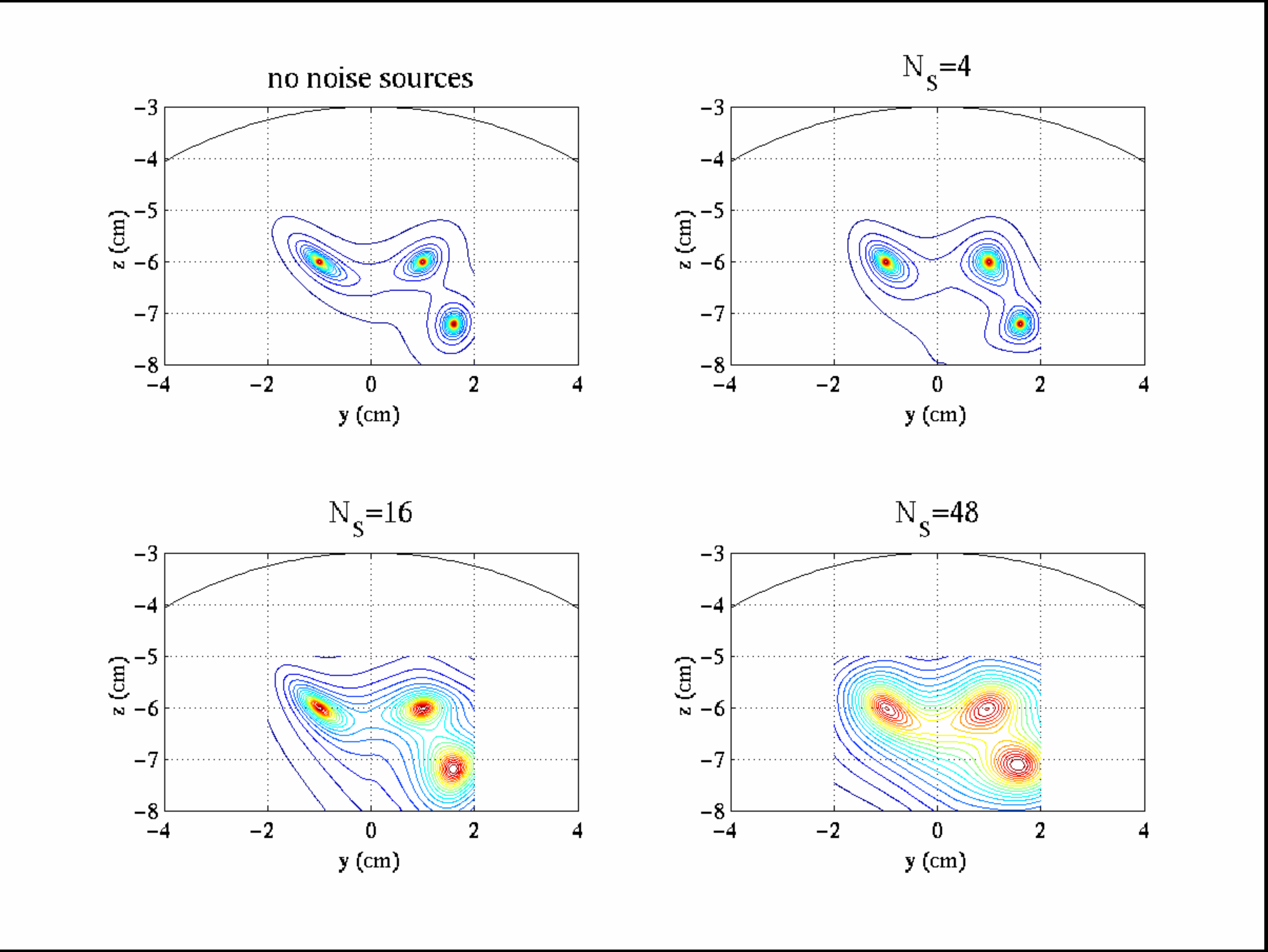
$P_N = 0.1P_3$ where P_3 is the power of the third source ($P_1 = P_2 = 1.2P_3$).

Time courses of the noise sources are incoherent to each other.

37-channel sensors used (*M*=37)

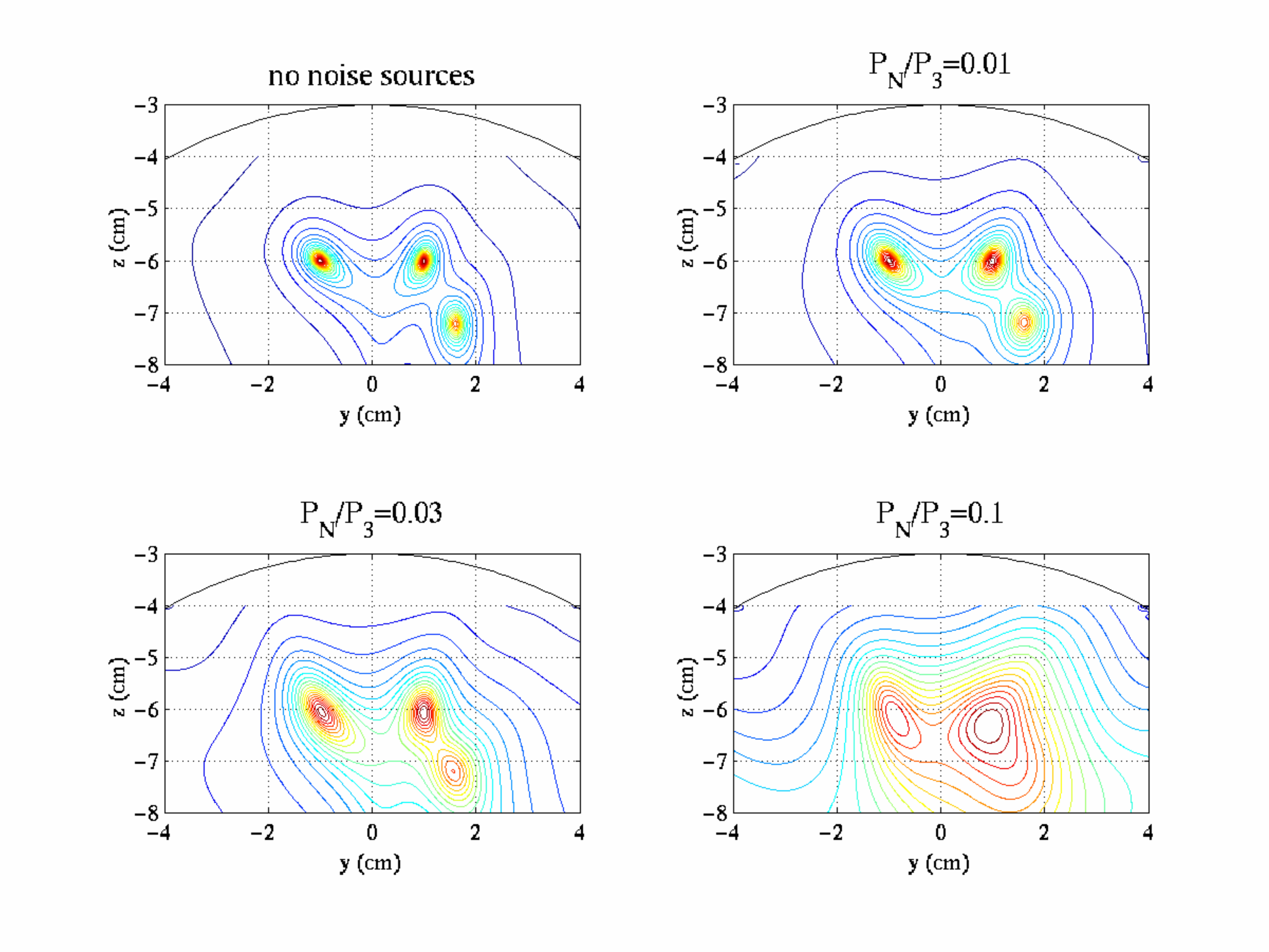


148-channel sensors used (M=148)



Experiments for changing P_N while N_S is fixed at 48

M=37



Questions

A large number of randomly distributed dipoles as a model of spontaneous neural activity!!

Do such noise sources really exist?

If yes, how large is the power of each dipole?

Summary

Sensor noise

The spatial resolution is affected by this type of noise.
Some other effects are described in our poster.

External disturbances

Their effects on the reconstruction is negligible, if their eigenvectors are very different from lead field vectors in the source space.

Neurophysiology noise

It can seriously affect the quality of source reconstruction, if a large number of incoherent dipoles are an appropriate model for it.

Collaborators

University of California, San Francisco
Biomagnetic Imaging Laboratory
Dr. Srikantan S. Nagarajan

University of Maryland
Linguistics and Cognitive Neuroscience Laboratory
Dr. David Poeppel

Massachusetts Institute of Technology,
Department of Linguistics and Philosophy
Dr. Alec Marantz

Kanazawa Institute of Technology
Human Science Laboratory
Dr. Isao Hashimoto