

Acoustic source localization and acquisition based on a sparse grid-shift approach

Christos Tzagkarakis, W. Bastiaan Kleijn, Jorge Martinez, Richard Heusdens

Faculty of Electrical Engineering, Mathematics and Computer Science
Delft University of Technology

24 August 2016, Mountain View, CA

Outline

Basics of Compressed Sensing

Acoustic source localization

Acoustic source localization:
sparsity-based

Simulations

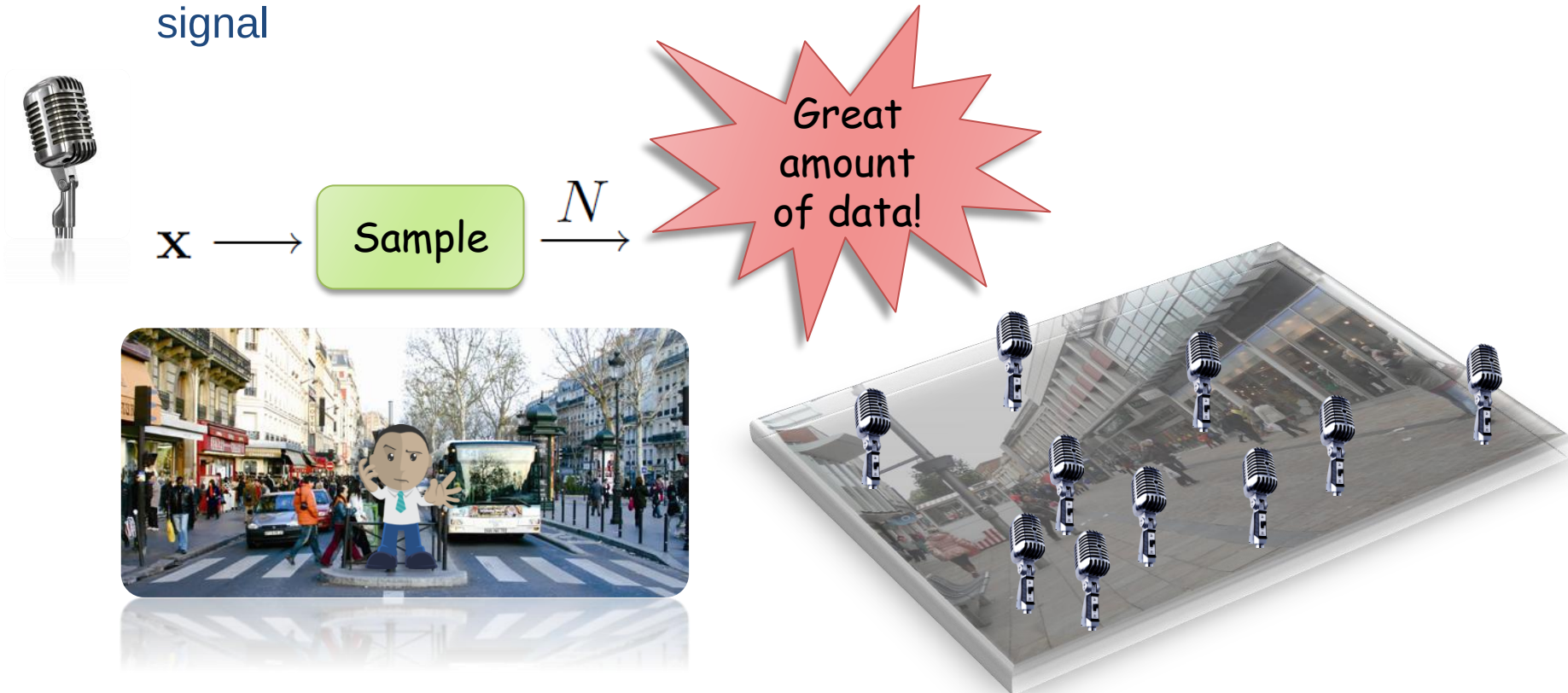
Grid-shift method

Simulations

Basics of CS

Sensing by sampling

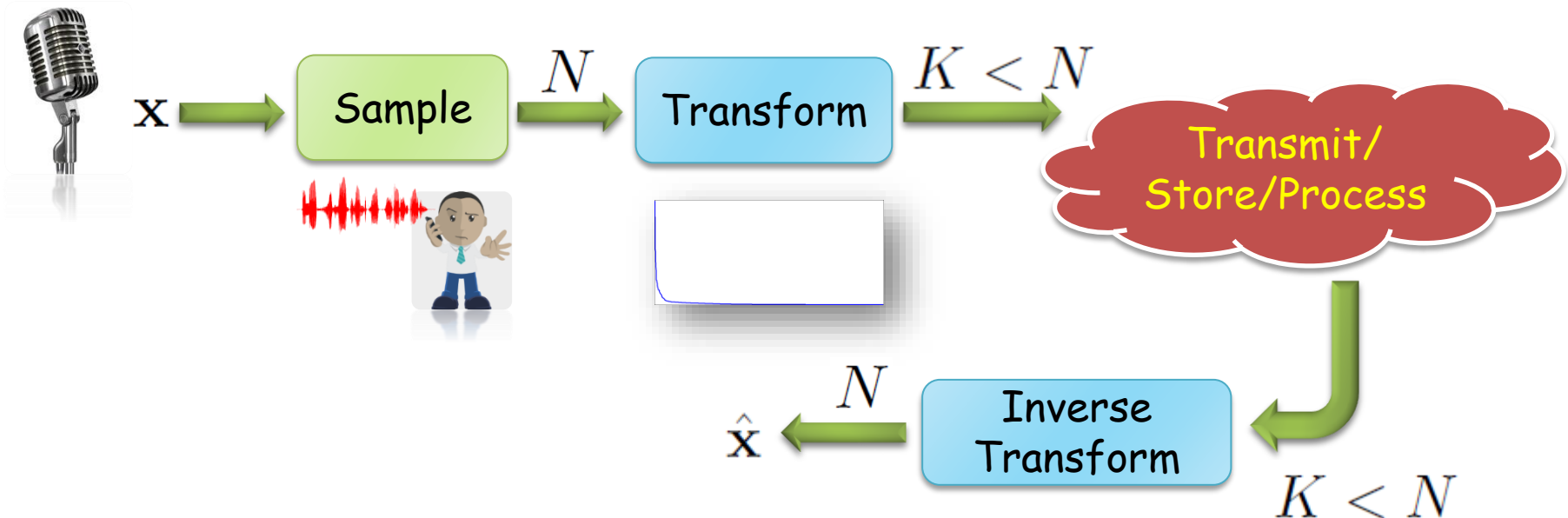
- ▶ Conventional trend: **Shannon-Nyquist** sampling theorem
 - ▶ Sample densely enough ($2 \times BW$) to perfectly reconstruct the original signal



Basics of CS

Alternative approach

- ▶ Conventional trend: **Shannon-Nyquist** sampling theorem
 - ▶ Sample densely enough ($2 \times \text{BW}$) to perfectly reconstruct the original signal



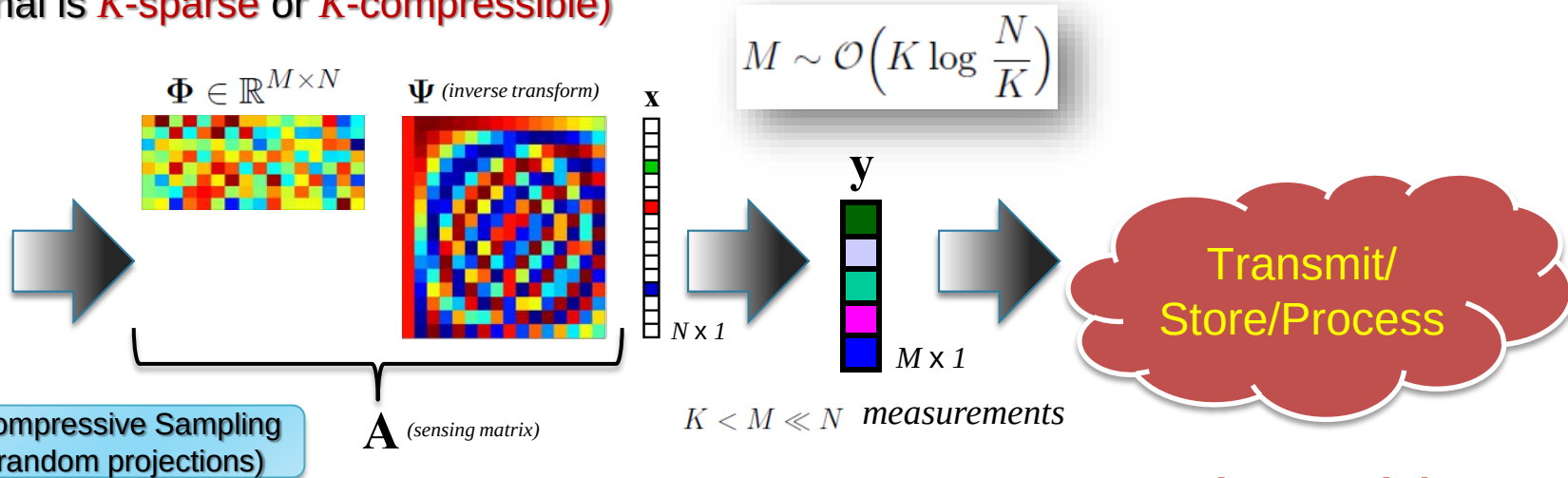
- ▶ Key assumption: **sparsity** or **compressibility** in a transform domain

Basics of CS

CS approach

(signal is K -sparse or K -compressible)

sampling

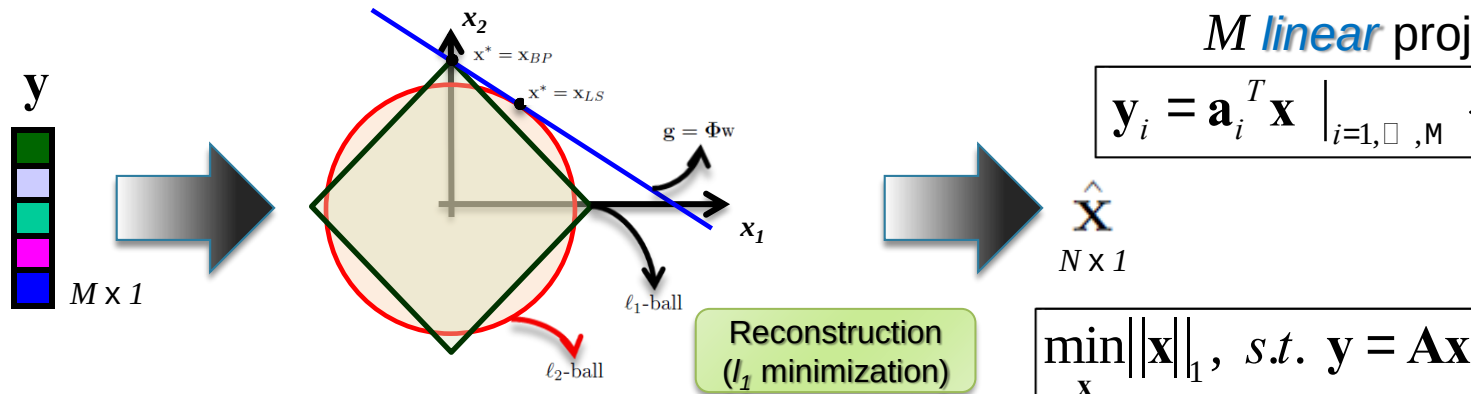


Sensing model

M linear projections

$$\mathbf{y}_i = \mathbf{a}_i^T \mathbf{x} \mid_{i=1, \dots, M} \Leftrightarrow \mathbf{y} = \mathbf{A} \mathbf{x}$$

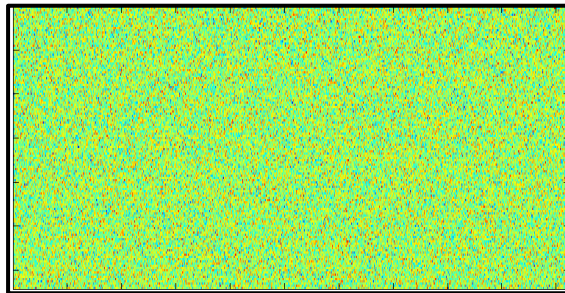
recovery



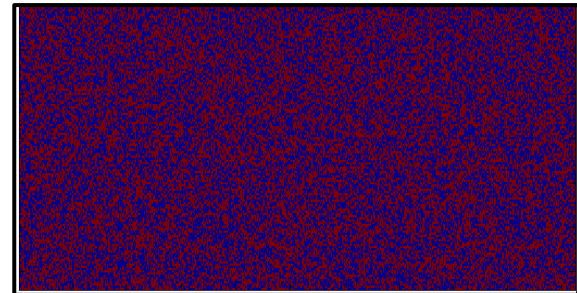
Basics of CS

CS performance guarantee

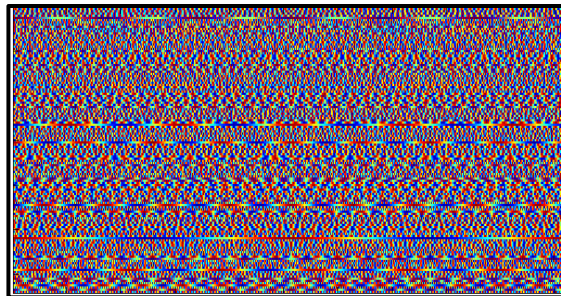
- ▶ Sensing matrix columns should be sufficiently uncorrelated (incoherent)
- ▶ Examples of incoherent matrices



Gaussian



Bernoulli



random Fourier

Basics of CS

CS performance guarantee

Restricted Isometry Property (RIP)

$\mathbf{A}$ satisfies the RIP of order K whenever

$$(1 - \epsilon_K) \|\mathbf{x}\|_2^2 \leq \|\mathbf{A}\mathbf{x}\|_2^2 \leq (1 + \epsilon_K) \|\mathbf{x}\|_2^2$$

holds simultaneously for all K -sparse vectors $\mathbf{x} \in \mathbb{R}^N$ for sufficiently small values of ϵ_K

Mutual incoherence

$$\mu(\mathbf{A}) = \max_{1 \leq i, j \leq N, i \neq j} \frac{|\mathbf{a}_i^H \mathbf{a}_j|}{\|\mathbf{a}_i\|_2 \|\mathbf{a}_j\|_2}$$

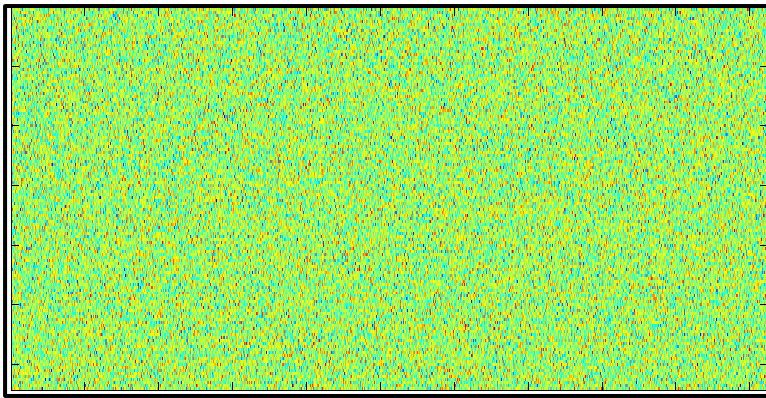
Basics of CS

CS key-points:

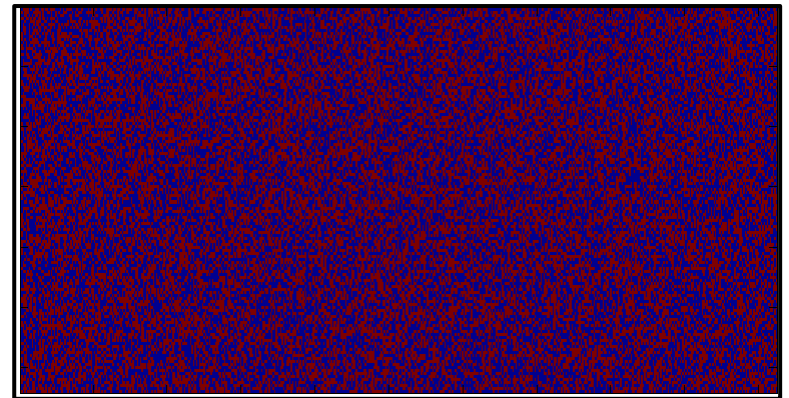
- ▶ **Universality:** the same random projections can be used with *any* given transformation (guaranteed incoherence with high prob.)
- ▶ **Asymmetric nature:** light encoder or sensing process, burden at decoder or reconstruction process
- ▶ **Democracy:**
 - ▶ Each measurement contains the same amount of information
 - ▶ Robustness to measurement loss & quantization
- ▶ **Decoupling property:** recovery is independent of acquisition process

RIP:

- ▶ Examples of matrices known to satisfy the RIP:

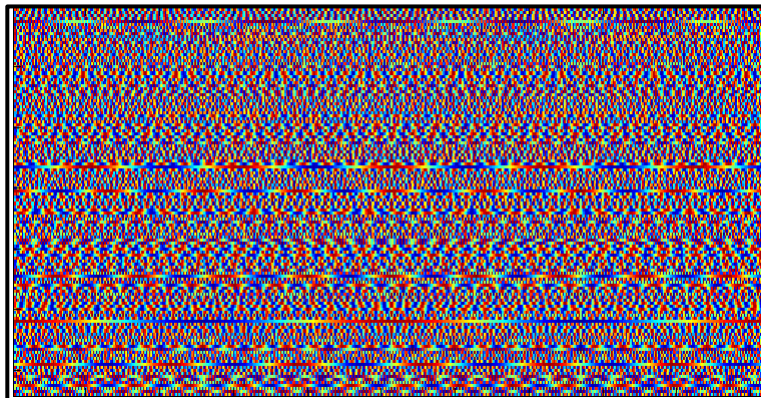


Gaussian



Bernoulli

random Fourier



- ▶ Reconstruction is guaranteed via the following fundamental theorem:

Theorem [Candes, Romberg, Tao, Donoho,...]:

If Φ satisfies the RIP of order $2K$, then any K -sparse signal $\mathbf{x}$, observed by $\mathbf{g} = \Phi\mathbf{x}$, is reconstructed *perfectly* by solving:

$$\mathbf{w}^* = \arg \min_{\mathbf{w}} \|\mathbf{w}\|_0 \quad s.t. \quad \mathbf{g} = \Phi\Psi\mathbf{w}$$

- ▶ **(-) *NP-complete*** problem (combinatorial complexity)!
- ▶ **Relaxed problem:** $\mathbf{w}^* = \arg \min_{\mathbf{w}} \|\mathbf{w}\|_1 \quad s.t. \quad \mathbf{g} = \Phi\Psi\mathbf{w}$

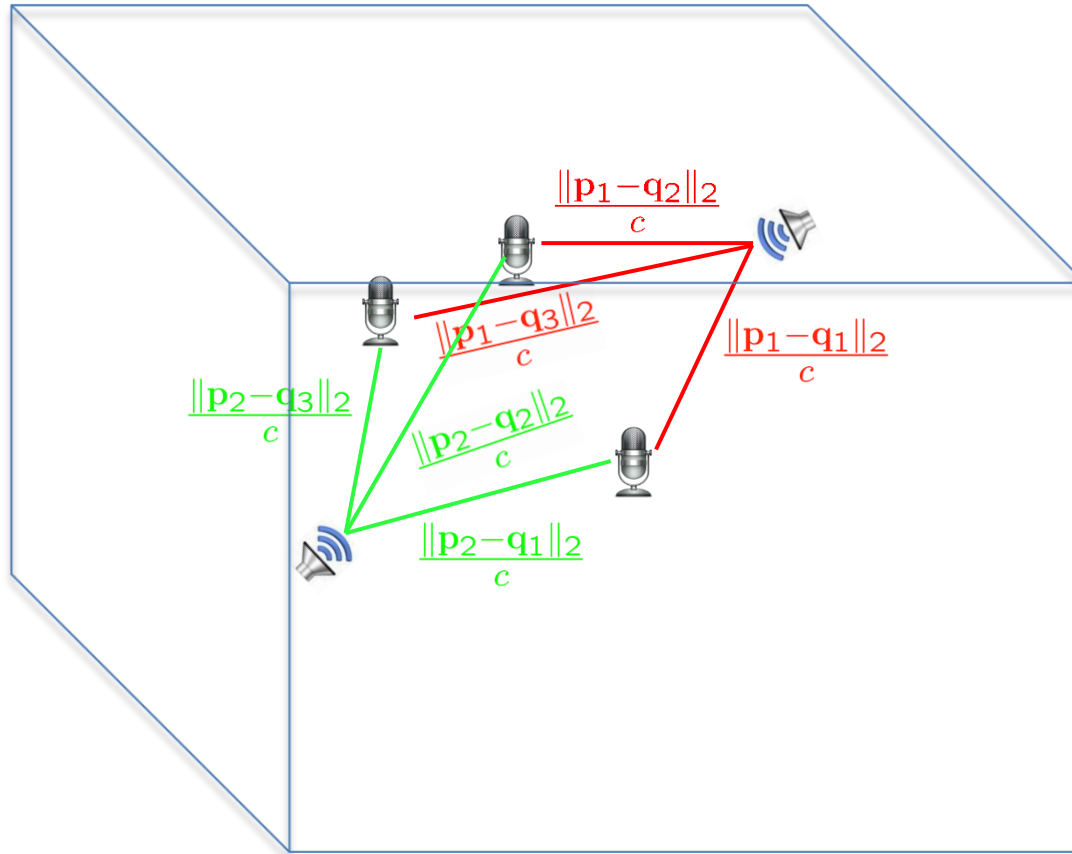
Acoustic source localization

Goal  *estimate the acoustic source locations*
acoustic source separation

- ▶ Available information
 - ▶ Known microphone positions (placed randomly)
 - ▶ Parameters of the medium
 - ▶ Observations on the microphones' side → measurement vector $\mathbf{y}$
 - ▶ Acoustic scene assumed sparse → sparse vector $\mathbf{x}$
- } sensing matrix $\mathbf{A}$
- ▶ Far-field scenario using a microphone array
 - ▶ K sources
 - ▶ M microphones

Acoustic source localization

Acoustic scene



Time of arrival
source @ $p_i \rightarrow$ mic @ q_m

$$\frac{\|p_i - q_m\|_2}{c}$$

Propagation function
free-field Green's function
source @ $p_i \rightarrow$ mic @ q_m

$$A_{m,i}(\omega_f) = \frac{e^{j\omega_f \|p_i - q_m\|_2 / c}}{4\pi \|p_i - q_m\|_2}$$

- ω_f : angular frequency
- c : speed of sound

Acoustic source localization

- ▶ Acoustic field dependent time-frequency representation

$$Y_m(t, \omega_f) = \sum_{i=1}^K A_{m,i}(\omega_f) X_i(t, \omega_f), \quad \forall m = 1 \dots, M$$

- time frame index $\rightarrow t = 1, \dots, T$
- angular frequency index $\rightarrow f = 1, \dots, F$

Propagation vector from the i -th position

$$\mathbf{a}(\mathbf{p}_i) = \begin{bmatrix} \frac{e^{j\omega_f \|\mathbf{p}_i - \mathbf{q}_1\|_2 / c}}{4\pi \|\mathbf{p}_i - \mathbf{q}_1\|_2} \\ \frac{e^{j\omega_f \|\mathbf{p}_i - \mathbf{q}_2\|_2 / c}}{4\pi \|\mathbf{p}_i - \mathbf{q}_2\|_2} \\ \vdots \\ \frac{e^{j\omega_f \|\mathbf{p}_i - \mathbf{q}_M\|_2 / c}}{4\pi \|\mathbf{p}_i - \mathbf{q}_M\|_2} \end{bmatrix} \in \mathbb{C}^{M \times 1}$$

$$\mathbf{A} = [\mathbf{a}(\mathbf{p}_1) | \mathbf{a}(\mathbf{p}_2) | \dots | \mathbf{a}(\mathbf{p}_K)] \in \mathbb{C}^{M \times K}$$

Array response matrix

Acoustic source localization: sparsity-based

- Matrix form for a fixed time frame t and angular frequency ω_f

$$\begin{bmatrix} Y_1(t, \omega_f) \\ \vdots \\ Y_M(t, \omega_f) \end{bmatrix} = \begin{bmatrix} A_{1,1}(\omega_f) & \dots & A_{1,K}(\omega_f) \\ \vdots & \ddots & \vdots \\ A_{M,1}(\omega_f) & \dots & A_{M,K}(\omega_f) \end{bmatrix} \cdot \begin{bmatrix} X_1(t, \omega_f) \\ \vdots \\ X_K(t, \omega_f) \end{bmatrix}$$

$$\mathbf{Y}(t, \omega_f) = \mathbf{A}(\omega_f) \mathbf{X}(t, \omega_f)$$

GIVEN mic's observations
&
array response matrix

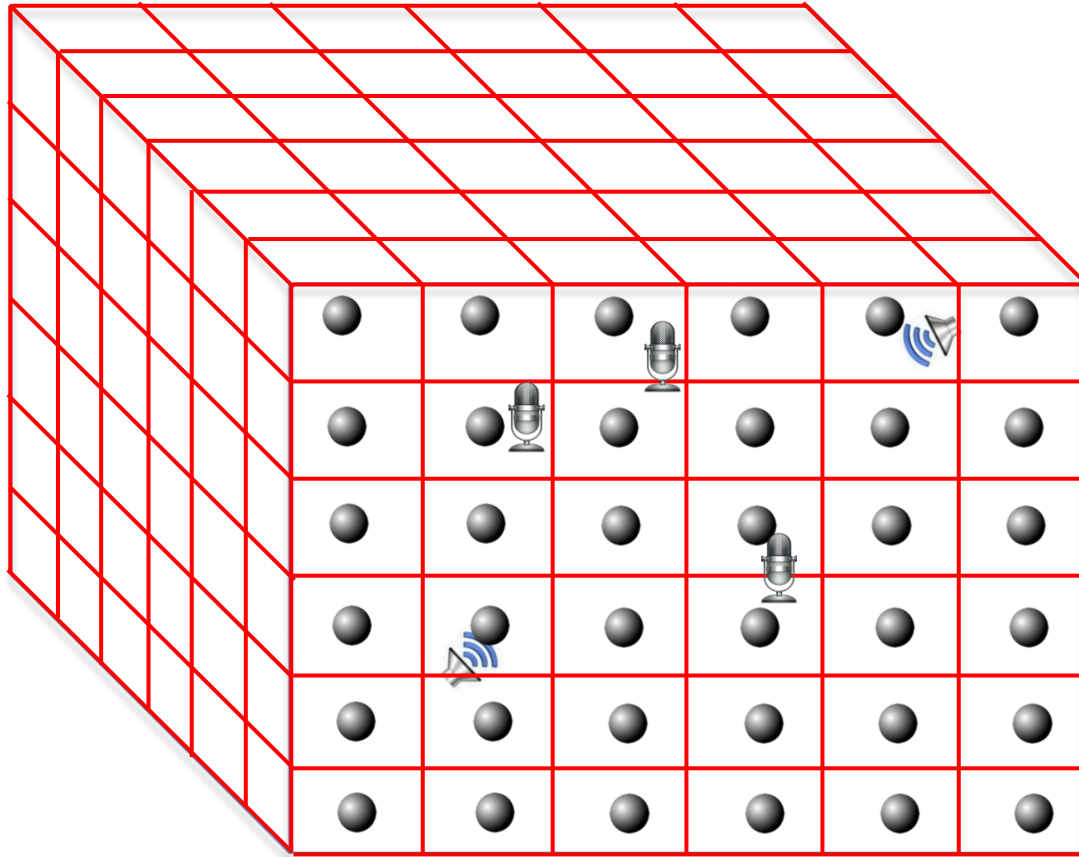
Recast as a sparse
recovery problem

Discretization of the
acoustic scene

Estimate source
locations

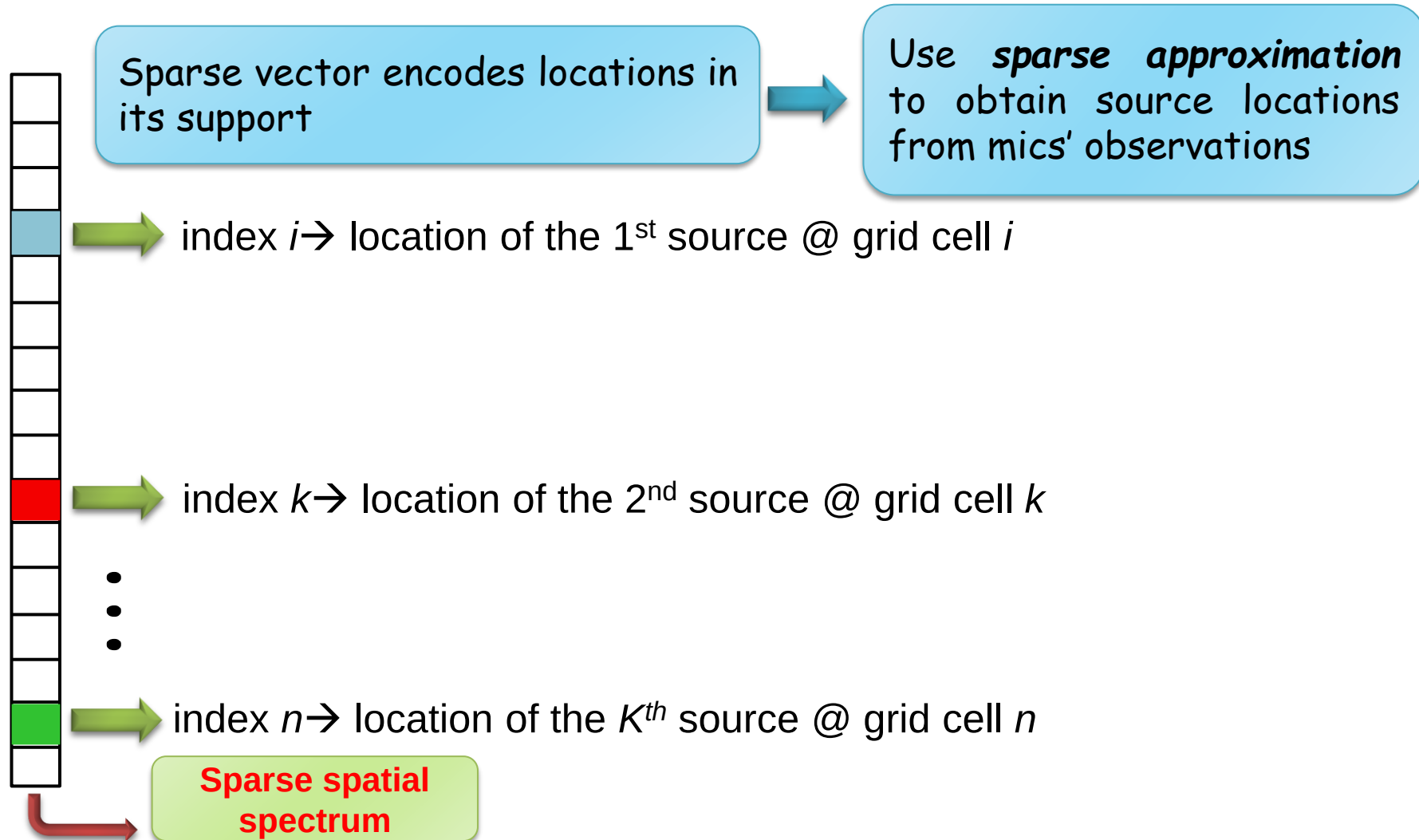
Find the number of the sources K
and their locations $\mathbf{p}_1, \dots, \mathbf{p}_K$

Acoustic source localization: sparsity-based



- Enforce sparsity $\rightarrow$ discretize the acoustic scene into N grid cells
- # sources $K \ll$ # grid points N

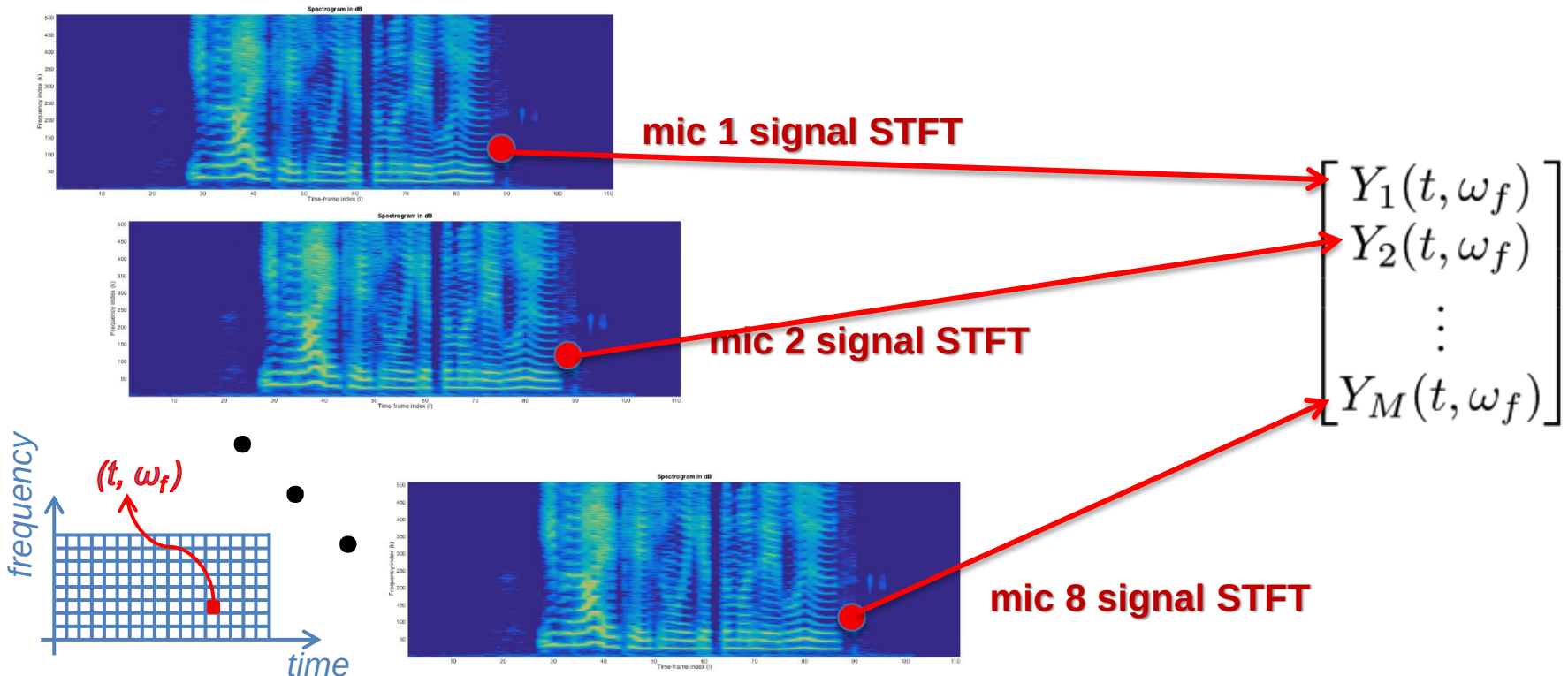
Acoustic source localization: sparsity-based



Acoustic source localization: sparsity-based

- Sparse approximation solving the L1-norm optimization problem

$$\hat{\mathbf{X}}(t, \omega_f) = \arg \min_{\mathbf{X}(t, \omega_f)} \|\mathbf{X}(t, \omega_f)\|_1, \text{ s.t. } \|\mathbf{Y}(t, \omega_f) - \mathbf{A}(\omega_f)\mathbf{X}(t, \omega_f)\|_2 \leq \epsilon$$



Acoustic source localization: sparsity-based

- Sparse approximation solving the L1-norm optimization problem

$$\hat{\mathbf{X}}(t, \omega_f) = \arg \min_{\mathbf{X}(t, \omega_f)} \|\mathbf{X}(t, \omega_f)\|_1, \text{ s.t. } \|\mathbf{Y}(t, \omega_f) - \mathbf{A}(\omega_f)\mathbf{X}(t, \omega_f)\|_2 \leq \epsilon$$



$$\begin{bmatrix} Y_1(t, \omega_f) \\ Y_2(t, \omega_f) \\ \vdots \\ Y_M(t, \omega_f) \end{bmatrix}$$

Acoustic source localization: sparsity-based

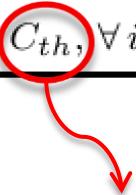
► Acoustic source localization

Input: $\mathbf{Y} \in \mathbb{C}^{M \times T \times F}$, $\mathbf{A} \in \mathbb{C}^{M \times N \times F}$, tolerance ϵ , decision threshold C_{th}

Output: estimated locations, estimated number of sources

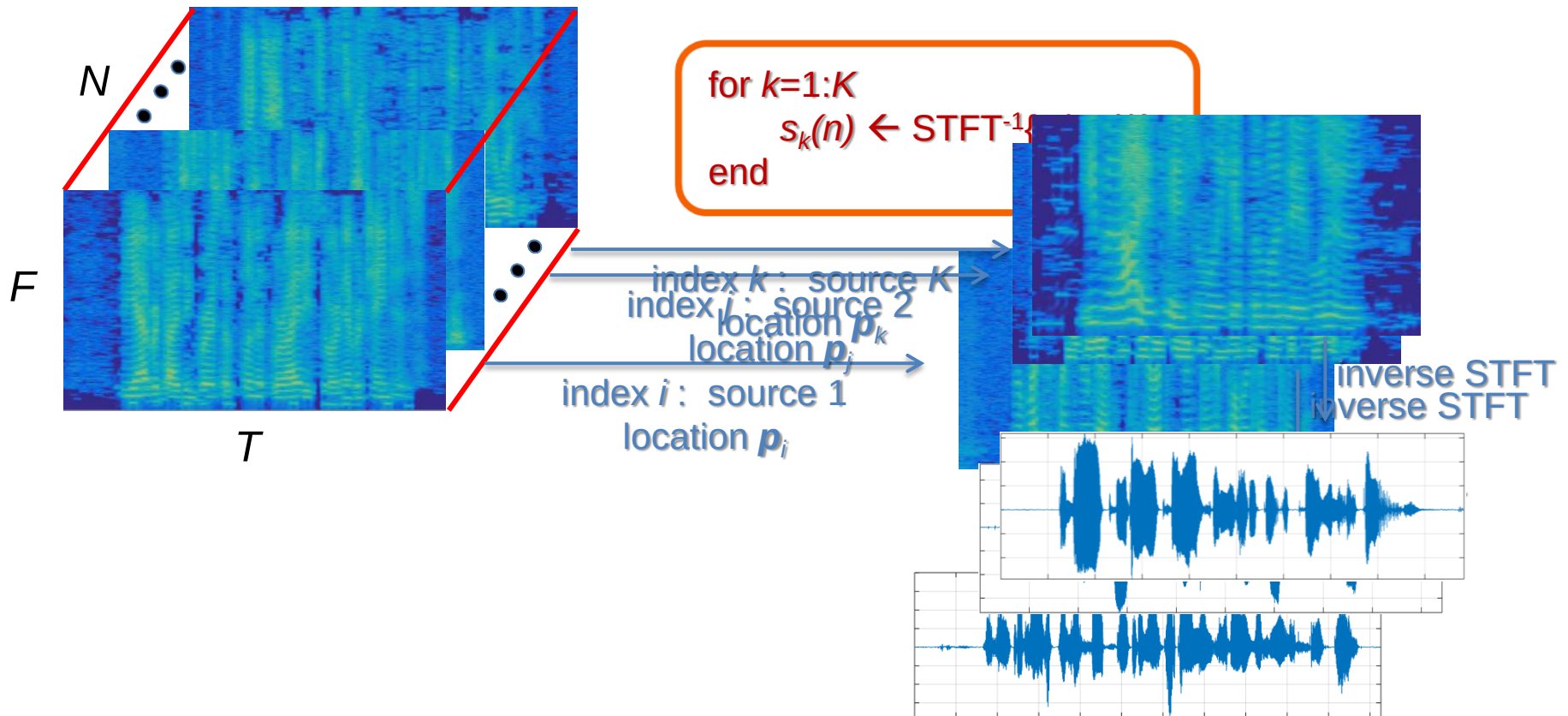
```
1 for  $t = 1$  to  $T$  do
2   for  $f = 1, \dots, F$  do
3      $\hat{\mathbf{X}}(t, \omega_f) = \arg \min_{\mathbf{X}(t, \omega_f)} \|\mathbf{X}(t, \omega_f)\|_1$ , s.t.  $\|\mathbf{Y}(t, \omega_f) - \mathbf{A}(\omega_f)\mathbf{X}(t, \omega_f)\|_2 \leq \epsilon$ 
4      $\mathbf{L}(t, \omega_f, :) \leftarrow \hat{\mathbf{X}}(t, \omega_f)^T$ 
5   end
6 end
7  $\mathbf{L} \in \mathbb{C}^{T \times F \times N}$ 
8 compute the absolute mean estimated spatial spectrum across time and frequency :  $\tilde{\mathbf{L}} \in \mathbb{C}^{1 \times N}$ 
9 return  $\mathcal{I} = \{i : \tilde{L}(i) > C_{th}, \forall i = 1, \dots, N\}$ 
```

- M : # mics
- N : # grid points
- T : # time frames
- F : # freq. bins
- D : # time samples


$$C_{th} = \frac{\frac{1}{M} \sum_{m=1}^M \left(\frac{1}{D} \sum_{d=1}^D |y_m(d)| \right)}{\epsilon M}$$

Acoustic source acquisition

- Acoustic source localization based on L1-norm optimization estimates a 3-dimensional matrix



Acoustic source localization: optimization algorithm

- ▶ Complex-valued orthogonal matching pursuit (OMP) for source localization

Input: $\mathbf{Y}$, $\mathbf{A}$, number of iterations $k_{\max}$, tolerance ϵ

Output: estimated sparse vector $\hat{\mathbf{X}}$

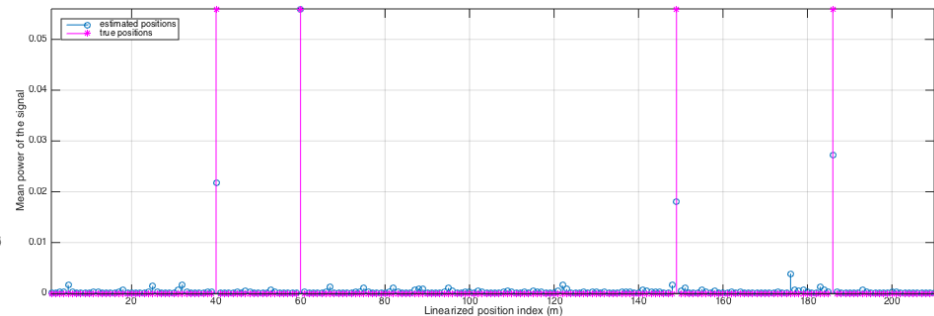
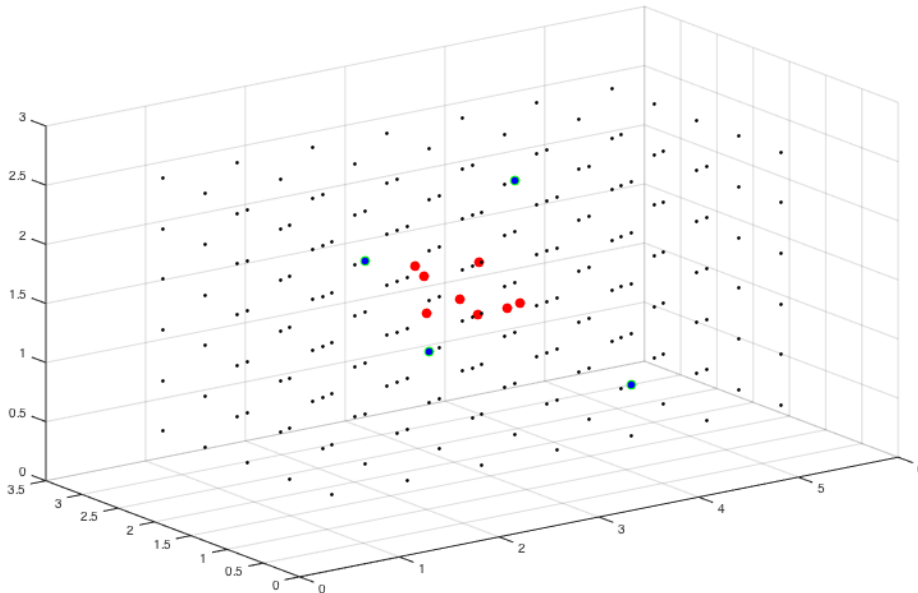
```
1 Initialization:  $k = 1$ ,  $\mathbf{r}_{k-1} = \mathbf{Y}$ ,  $\Lambda_{k-1} = \emptyset$ 
2 while  $k \leq k_{\max}$  or  $\|\mathbf{r}_{k-1}\|_2 \leq \epsilon$  do
3    $i^* = \arg \max_{1 \leq i \leq N} |\mathbf{a}(\mathbf{p}_i)^H \mathbf{r}_{k-1}| / \|\mathbf{a}(\mathbf{p}_i)\|_2$ 
4    $\Lambda_k = \Lambda_{k-1} \cup i^*$ 
5    $\mathbf{X}_{\Lambda_k} = \mathbf{A}_{\Lambda_k}^\dagger \mathbf{Y}$ 
6    $\mathbf{r}_k = \mathbf{Y} - \mathbf{A}_{\Lambda_k} \mathbf{X}_{\Lambda_k}$ 
7    $k = k + 1$ 
8 end
```

Simulations

- ▶ Box-shaped room: 6 x 3.5 x 3 meters
- ▶ 3-dimensional grid: 7 x 5 x 6=210 grid points
- ▶ Speech excerpts: VOICES corpus
 - ▶ 4.5 sec
 - ▶ 8 kHz
- ▶ Evaluation metrics
 - ▶ Segmental SNR (SNR_{seg})
 - ▶ Global SNR (SNR)
 - ▶ Short-time objective intelligibility measure ($STOI$)

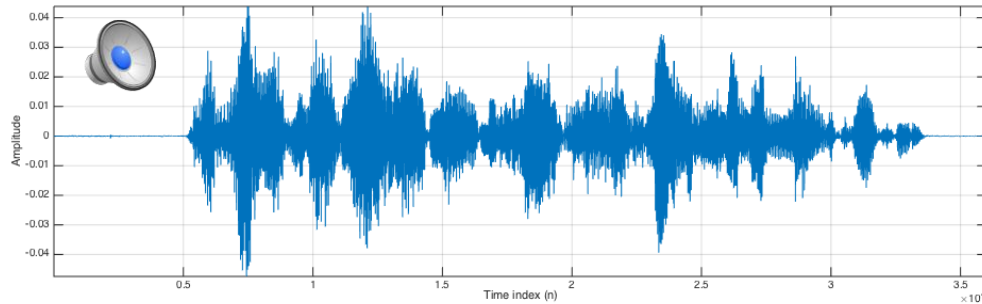
Simulations

- ▶ On-grid scenario: sources lying on the grid points
- ▶ 8 microphones in random locations (red dots)
- ▶ 4 active sources (blue dots)
- ▶ Free-field assumption
- ▶ No reverberation & no noise

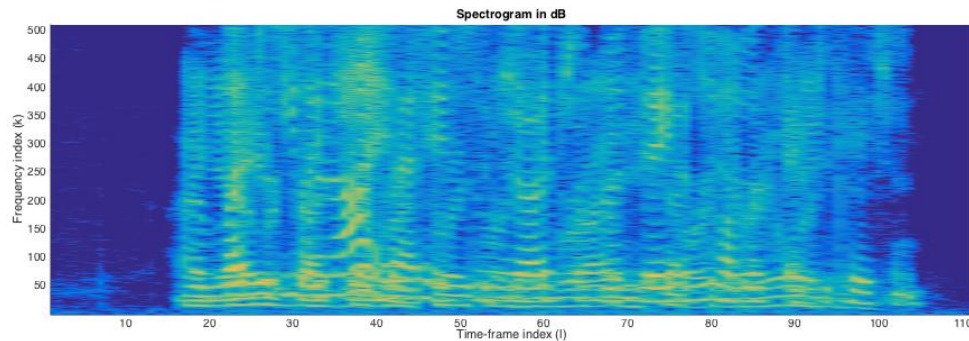


Simulations

► Mixture signal in time-domain



► Mixture signal STFT



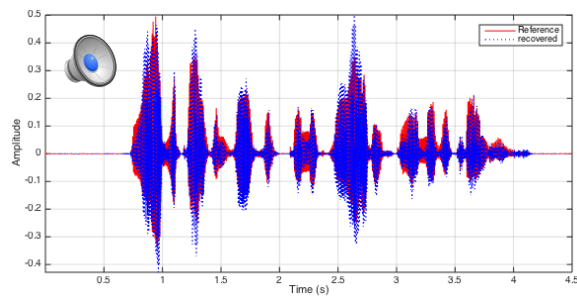
Simulations

Recovered source 1

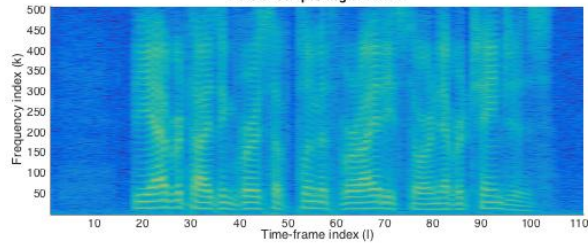
SNR_{seg} : 4.99

SNR: 6.26

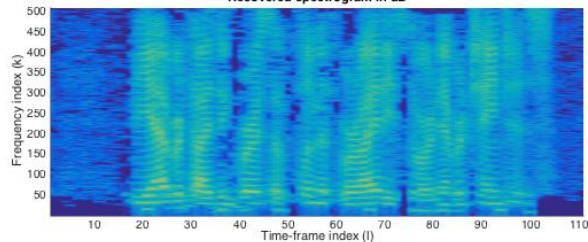
STOI: 0.83



Reference spectrogram in dB



Recovered spectrogram in dB

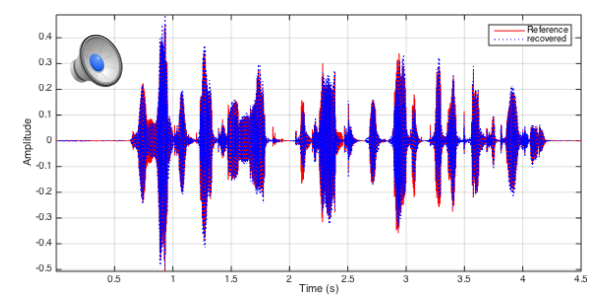


Recovered source 2

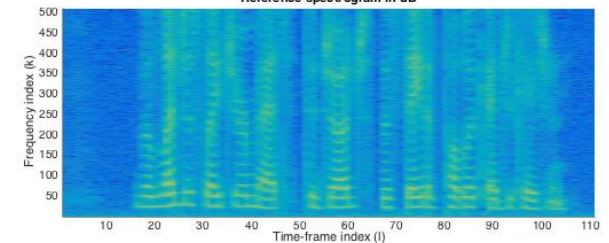
SNR_{seg} : 10.35

SNR: 11.30

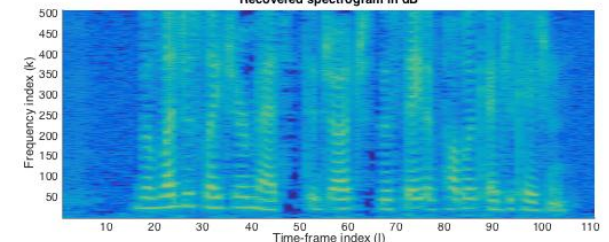
STOI: 0.93



Reference spectrogram in dB



Recovered spectrogram in dB



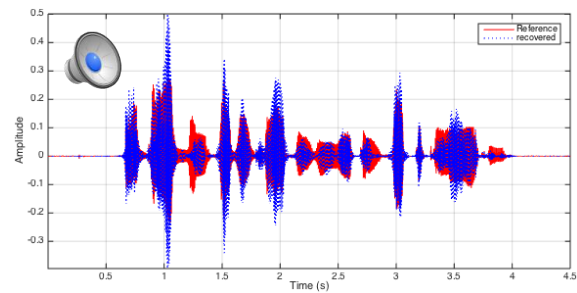
Simulations

Recovered source 3

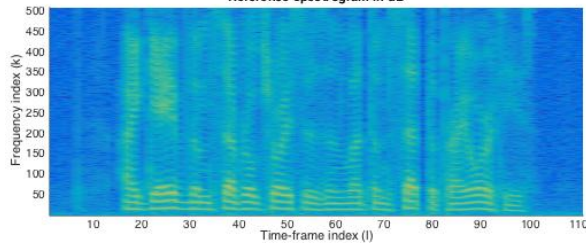
SNR_{seg}: 2.68

SNR: 3.41

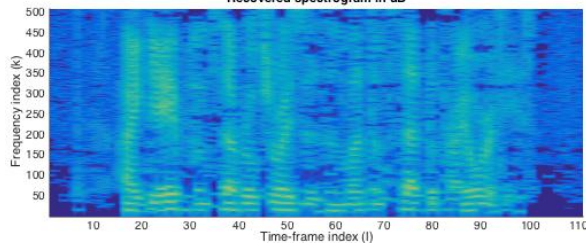
STOI: 0.82



Reference spectrogram in dB



Recovered spectrogram in dB

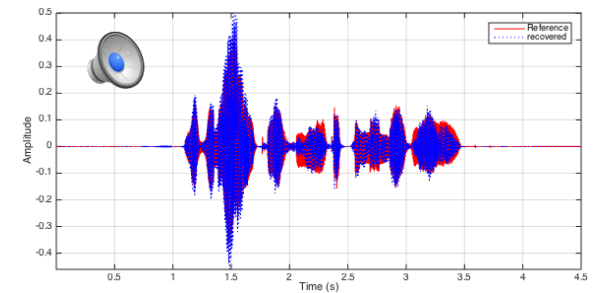


Recovered source 4

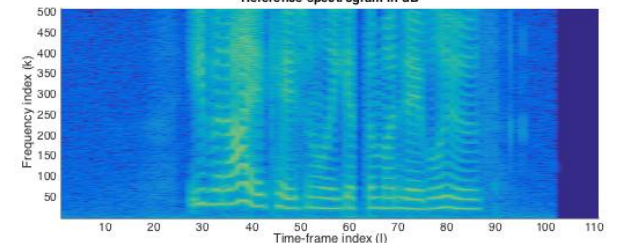
SNR_{seg}: 6.29

SNR: 8.11

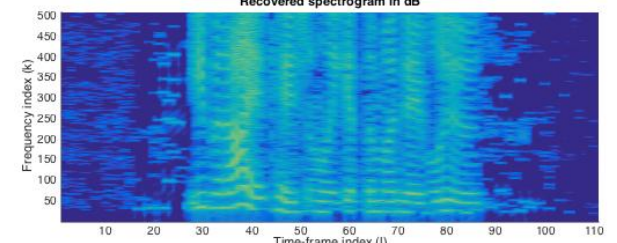
STOI: 0.82



Reference spectrogram in dB

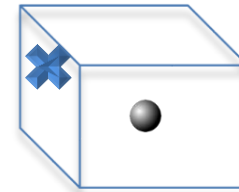


Recovered spectrogram in dB



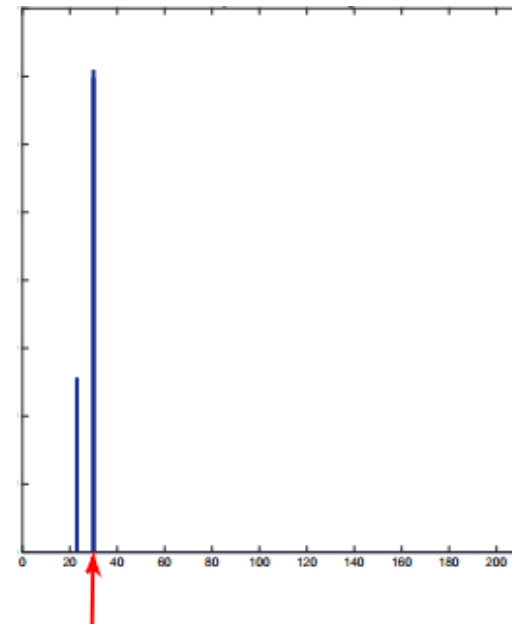
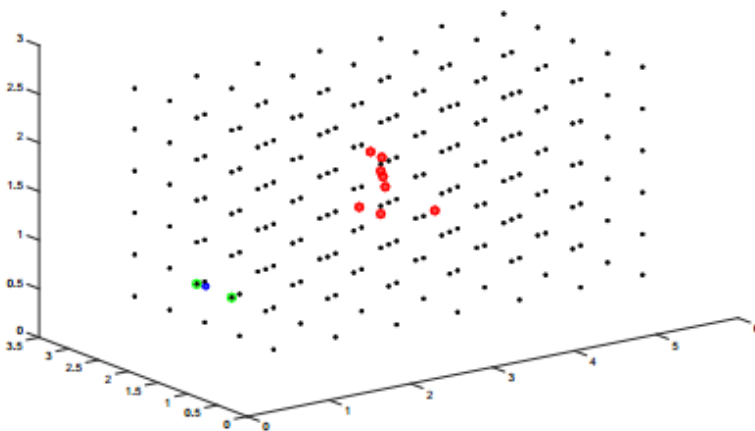
Grid-shift method

- ▶ Acoustic source localization: assumptions so far
 - ▶ Acoustic scene forms a continuous space → discretize into finite grid points
 - ▶ Physical model assumption → generate sensing matrix A
 - ▶ Acoustic sources do not exactly lie on the grid points → mismatch between the predefined sensing matrix and the actual one
- ▶ Off-grid localization
 - ▶ Sources are located away from the grid points
 - ▶ Cannot arbitrarily increase make the grid denser → RIP violation
 - ▶ Introduce the concept of “grid-shift” → generate a grid which is shifted across the entire acoustic scene
 - ▶ High probability that a point on one of the grids is close to the source



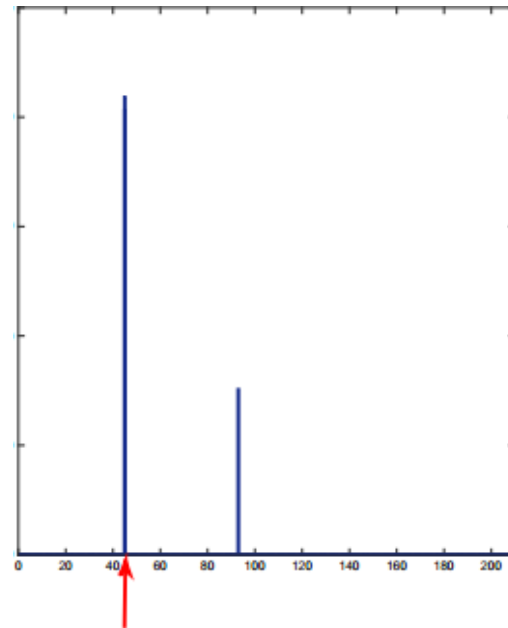
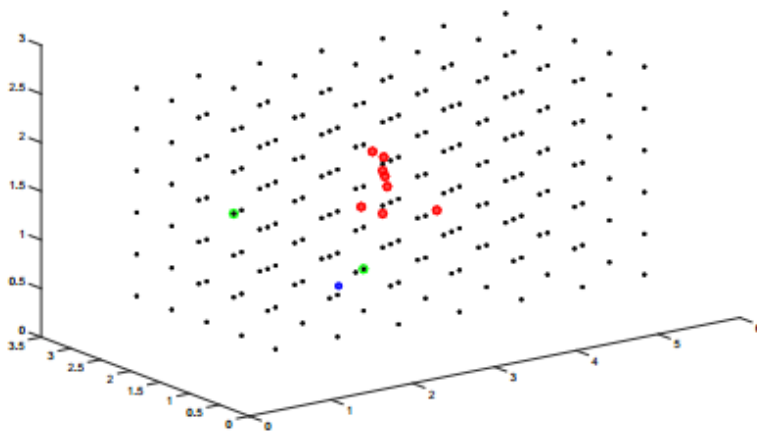
Grid-shift method

- ▶ Mismatch between grid points vs. source locations
 - ▶ How localization performance is affected?
 - ▶ “energy leakage”: (from) source location $\rightarrow$ (to) neighboring grid nodes
- ▶ Assume a moving source



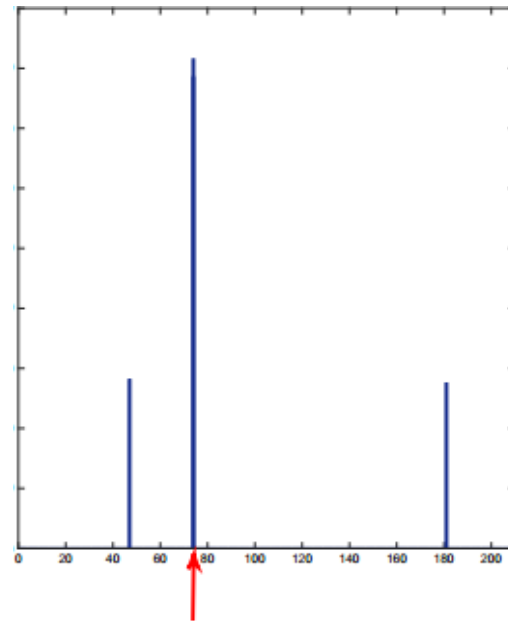
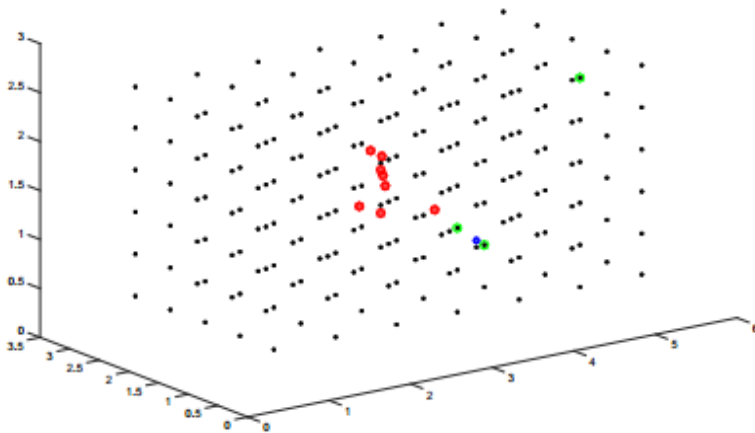
Grid-shift method

- ▶ Mismatch between grid points vs. source locations
 - ▶ How localization performance is affected?
 - ▶ “energy leakage”: (from) source location $\rightarrow$ (to) neighboring grid nodes
- ▶ Assume a moving source

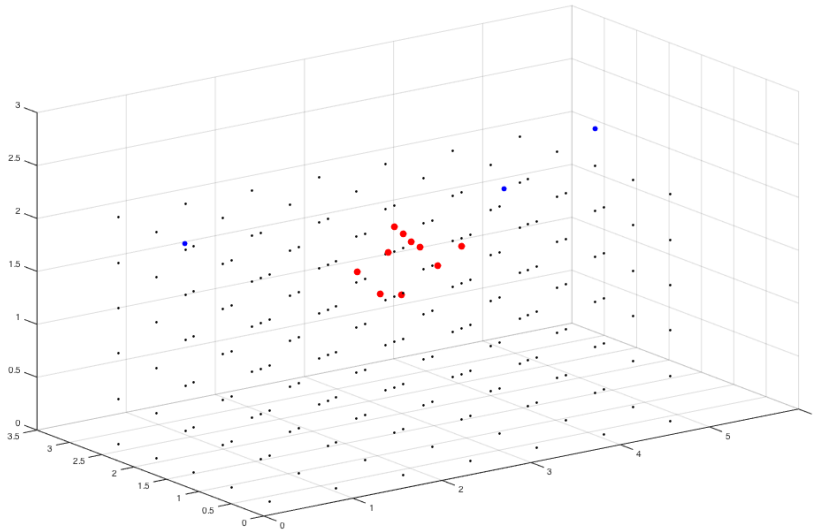


Grid-shift method

- ▶ Mismatch between grid points vs. source locations
 - ▶ How localization performance is affected?
 - ▶ “energy leakage”: (from) source location $\rightarrow$ (to) neighboring grid nodes
- ▶ Assume a moving source



Grid-shift method



1-st grid $\rightarrow \mathbf{G}_1 = \begin{bmatrix} \mathbf{p}_1^1 \\ \mathbf{p}_2^1 \\ \vdots \\ \mathbf{p}_N^1 \end{bmatrix} \in \mathbb{R}^{N \times 3}$

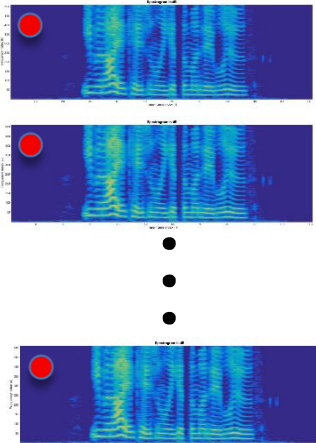
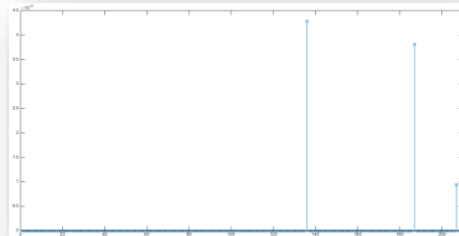


1-st array response matrix

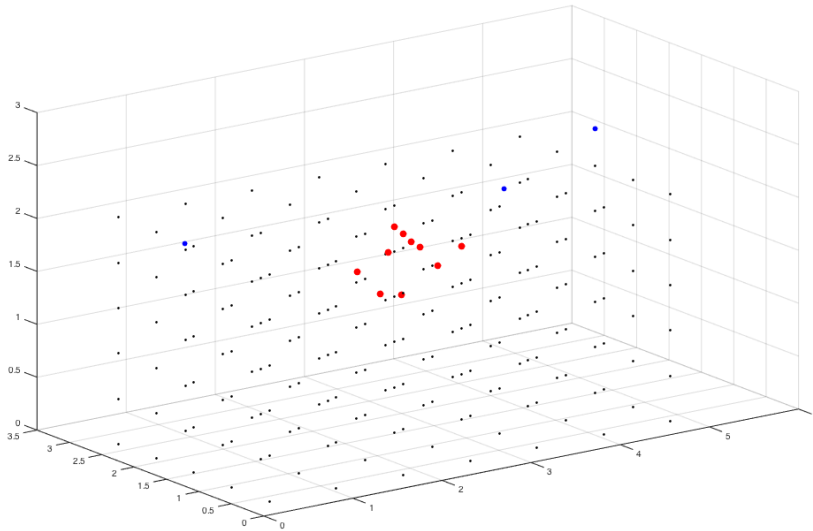
$$\mathbf{A}_1 = [\mathbf{a}(\mathbf{p}_1^1) | \mathbf{a}(\mathbf{p}_2^1) | \dots | \mathbf{a}(\mathbf{p}_N^1)] \in \mathbb{C}^{M \times N}$$



$$\hat{\mathbf{X}}(t, \omega_f) = \arg \min_{\mathbf{X}(t, \omega_f)} \|\mathbf{X}(t, \omega_f)\|_1, \text{ s.t. } \|\mathbf{Y}(t, \omega_f) - \mathbf{A}_1(\omega_f)\mathbf{X}(t, \omega_f)\|_2 \leq \epsilon$$



Grid-shift method



1-st grid $\rightarrow \mathbf{G}_1 = \begin{bmatrix} \mathbf{p}_1^1 \\ \mathbf{p}_2^1 \\ \vdots \\ \mathbf{p}_N^1 \end{bmatrix} \in \mathbb{R}^{N \times 3}$

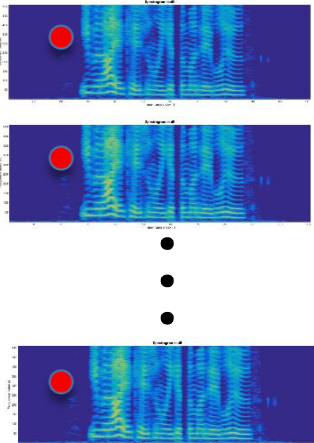
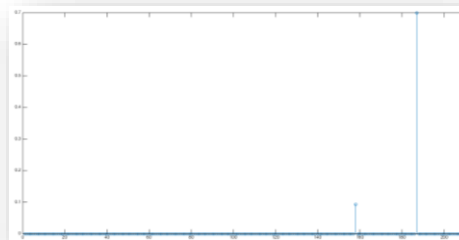


1-st array response matrix

$$\mathbf{A}_1 = [\mathbf{a}(\mathbf{p}_1^1) | \mathbf{a}(\mathbf{p}_2^1) | \dots | \mathbf{a}(\mathbf{p}_N^1)] \in \mathbb{C}^{M \times N}$$

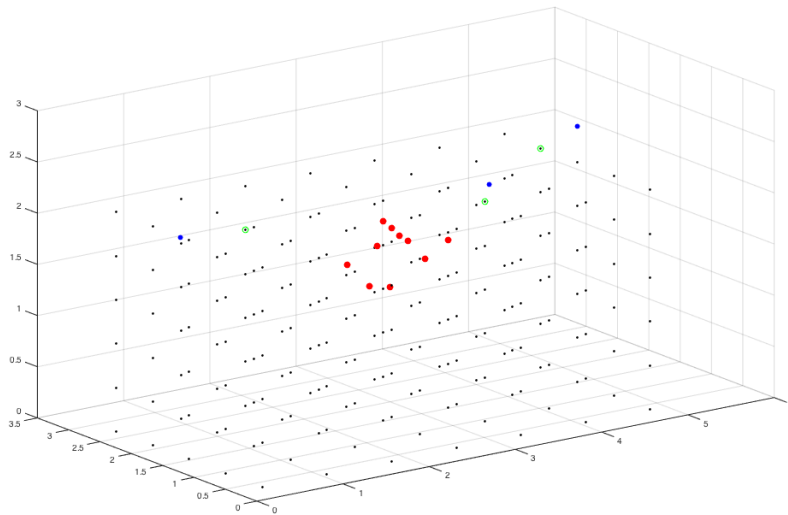


$$\hat{\mathbf{X}}(t, \omega_f) = \arg \min_{\mathbf{X}(t, \omega_f)} \|\mathbf{X}(t, \omega_f)\|_1, \text{ s.t. } \|\mathbf{Y}(t, \omega_f) - \mathbf{A}_1(\omega_f)\mathbf{X}(t, \omega_f)\|_2 \leq \epsilon$$

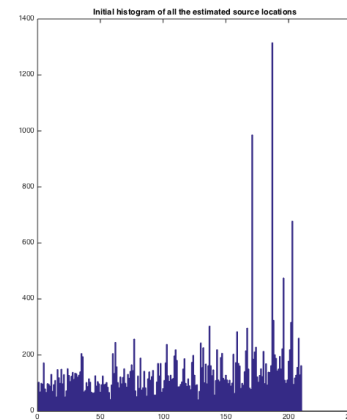


Grid-shift method

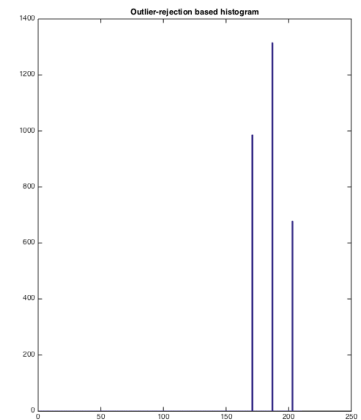
1st grid $\rightarrow$ TxF sparse solutions



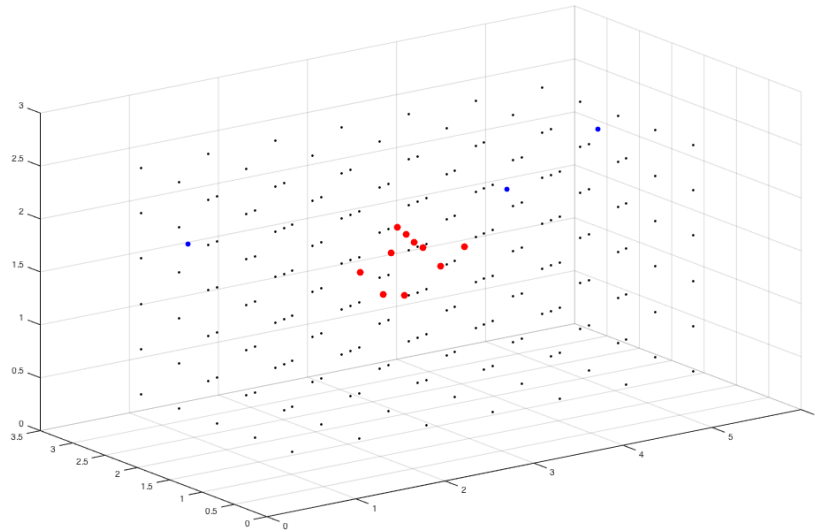
Histogram of all
the estimated TxF
sparse solutions



Outlier rejection:
reject values $\leq$
 $0.4 * \max(\text{histogram})$



Grid-shift method



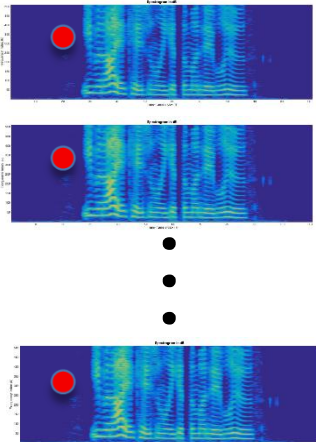
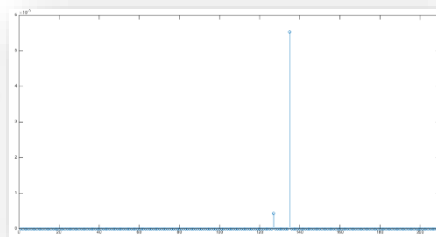
2-nd grid $\rightarrow \mathbf{G}_2 = \begin{bmatrix} \mathbf{p}_1^2 \\ \mathbf{p}_2^2 \\ \vdots \\ \mathbf{p}_N^2 \end{bmatrix} \in \mathbb{R}^{N \times 3}$



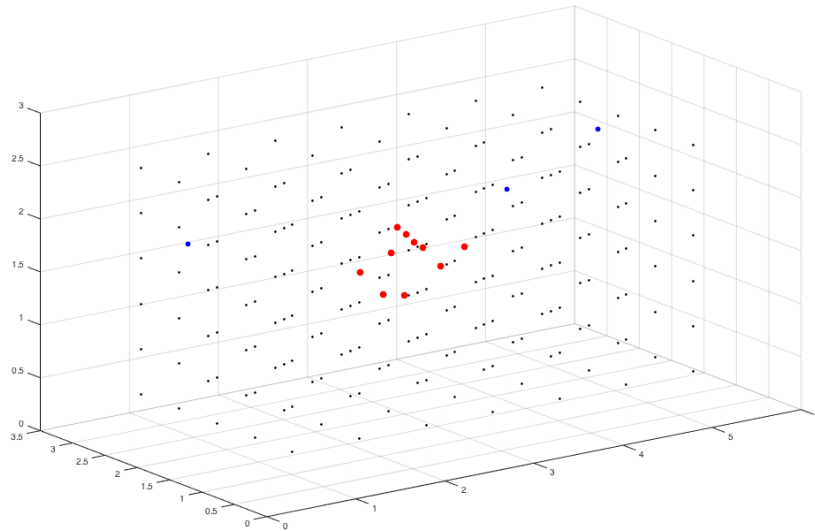
2-nd array response matrix
 $\mathbf{A}_2 = [\mathbf{a}(\mathbf{p}_1^2) | \mathbf{a}(\mathbf{p}_2^2) | \dots | \mathbf{a}(\mathbf{p}_N^2)] \in \mathbb{C}^{M \times N}$



$$\hat{\mathbf{X}}(t, \omega_f) = \arg \min_{\mathbf{X}(t, \omega_f)} \|\mathbf{X}(t, \omega_f)\|_1, \text{ s.t. } \|\mathbf{Y}(t, \omega_f) - \mathbf{A}_2(\omega_f)\mathbf{X}(t, \omega_f)\|_2 \leq \epsilon$$



Grid-shift method



2-nd grid $\rightarrow \mathbf{G}_2 = \begin{bmatrix} \mathbf{p}_1^2 \\ \mathbf{p}_2^2 \\ \vdots \\ \mathbf{p}_N^2 \end{bmatrix} \in \mathbb{R}^{N \times 3}$

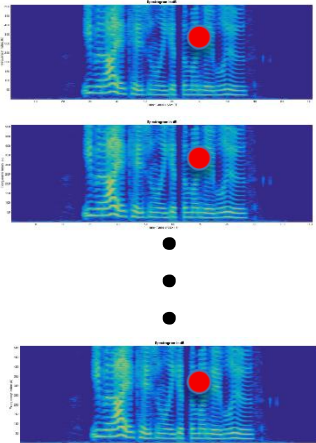
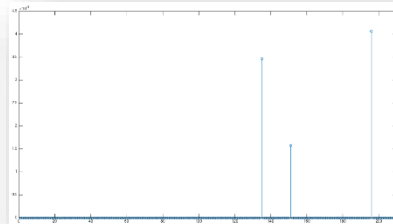


2-nd array response matrix

$$\mathbf{A}_2 = [\mathbf{a}(\mathbf{p}_1^2) | \mathbf{a}(\mathbf{p}_2^2) | \dots | \mathbf{a}(\mathbf{p}_N^2)] \in \mathbb{C}^{M \times N}$$

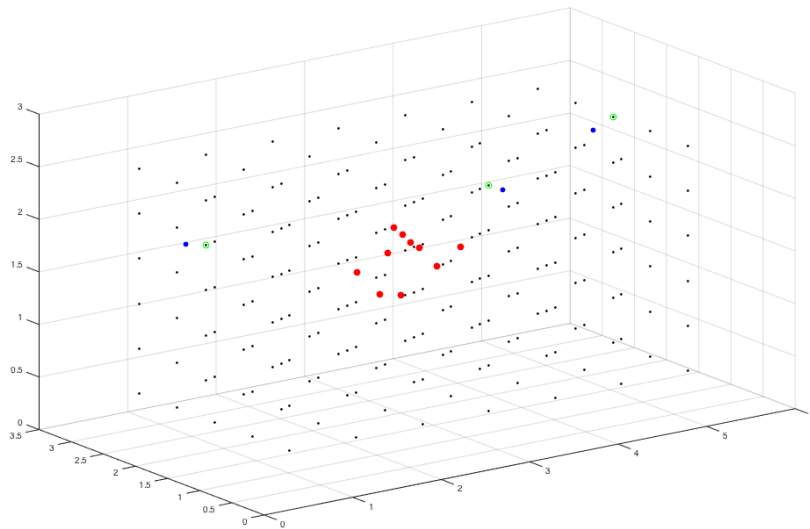


$$\hat{\mathbf{X}}(t, \omega_f) = \arg \min_{\mathbf{X}(t, \omega_f)} \|\mathbf{X}(t, \omega_f)\|_1, \text{ s.t. } \|\mathbf{Y}(t, \omega_f) - \mathbf{A}_2(\omega_f)\mathbf{X}(t, \omega_f)\|_2 \leq \epsilon$$

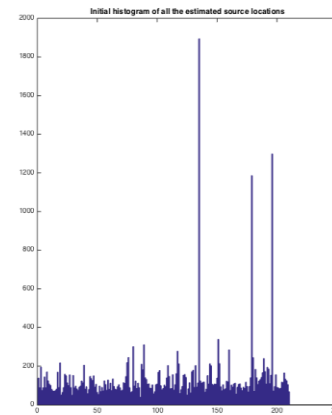


Grid-shift method

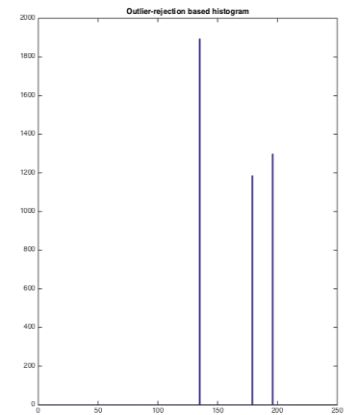
2nd grid → TxF sparse solutions



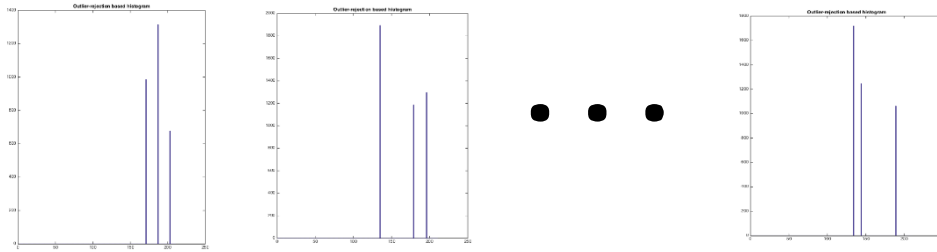
Histogram over
all the estimated
 TxF sparse
solutions



Outlier rejection:
reject values $\leq$
 $0.4 \cdot \max(\text{histogram})$

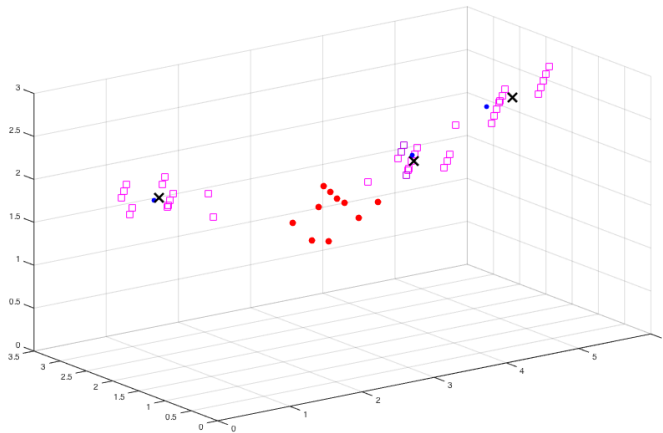


Grid-shift method



Majority voting on thresholded histograms → find #sources

Estimated number of sources



Apply K-means clustering to Estimate the source locations

Grid-shift method

► Acoustic source localization using “grid-shift”

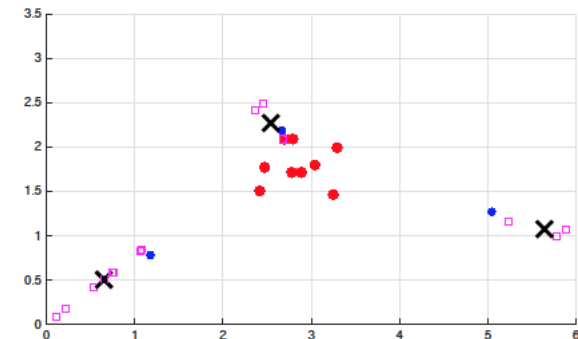
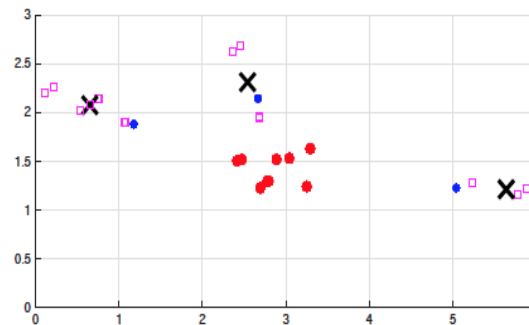
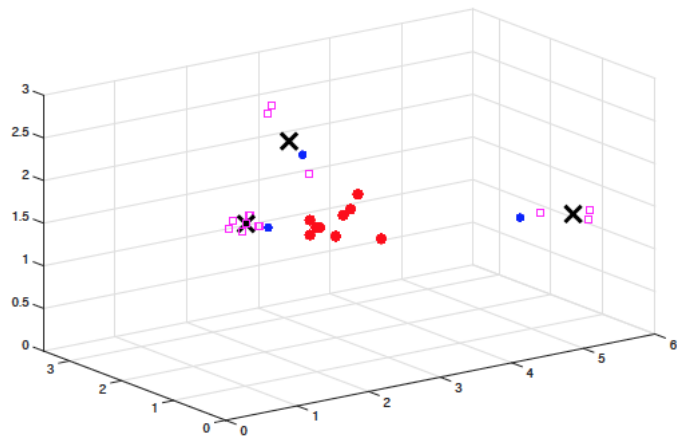
Input: $\mathbf{Y} \in \mathbb{C}^{M \times T \times F}$, maximum number of grid shifts $g_{\max}$, tolerance ϵ , decision threshold C_{th}

Output: estimated locations, estimated number of sources

```
1 for  $r = 1$  to  $g_{\max}$  do
2   built matrix  $\mathbf{A}_r$  using the current grid  $\mathbf{G}_r$  according to  $\mathbf{A}_r = [\mathbf{a}(\mathbf{p}_1^r) | \mathbf{a}(\mathbf{p}_2^r) | \dots | \mathbf{a}(\mathbf{p}_N^r)] \in \mathbb{C}^{M \times N}$ 
3   for  $t = 1$  to  $T$  do
4     for  $f = 1, \dots, F$  do
5        $\hat{\mathbf{X}}(t, \omega_f) = \arg \min_{\mathbf{X}(t, \omega_f)} \|\mathbf{X}(t, \omega_f)\|_1$ , s.t.  $\|\mathbf{Y}(t, \omega_f) - \mathbf{A}_r(\omega_f)\mathbf{X}(t, \omega_f)\|_2 \leq \epsilon$ 
6        $\mathbf{L}_r(t, \omega_f, :) \leftarrow \hat{\mathbf{X}}(t, \omega_f)^T$ 
7     end
8   end
9    $\mathbf{L}_r \in \mathbb{C}^{T \times F \times N}$ 
10  compute histogram  $\mathbf{h}_r \in \mathbb{R}^{N \times 1}$  corresponding to all  $T \times F$  sparse solutions
11   $h_{\max}^r = \max\{\mathbf{h}_r\}$ 
12  save the index of the maximum histogram value:  $v(r) = \arg \max\{\mathbf{h}_r\}$ 
13  estimate the 3D coordinates of the maximum histogram peak:  $\mathbf{G}_r(v(r), :)$ 
14  for  $n = 1$  to  $N$  do
15    if  $h_r(n) < C_{\text{th}} h_{\max}^r$  then
16       $h_r(n) = 0$  (thresholded histogram)
17    end
18  end
19 end
20 count the number of non-zero values for all thresholded histograms
21 apply majority voting to estimate the number of acoustic sources  $\hat{K}$ 
22 use k-means clustering to estimate the source locations with  $k = \hat{K}$ 
```

Simulations

- ▶ Off-grid scenario: sources lying away from the grid points
- ▶ 9 microphones in random locations (red dots)
- ▶ 3 active sources (blue dots)
- ▶ Free-field assumption
- ▶ No reverberation & no noise
- ▶ #grid-shifts=15, $C_{th}=0.35$, $\varepsilon=0.01$ (64 equally spaced frequencies selected)



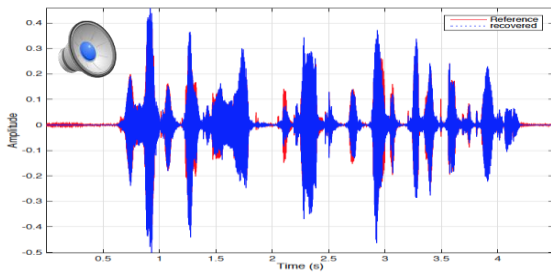
Simulations

Recovered source 1

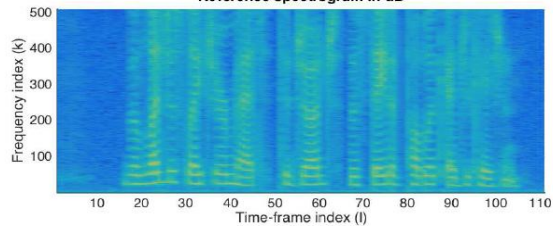
SNR_{seg} : 4.64

SNR: 5.96

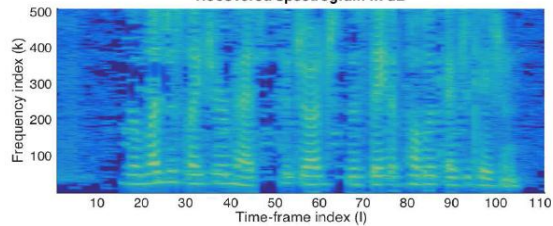
STOI: 0.91



Reference spectrogram in dB



Recovered spectrogram in dB

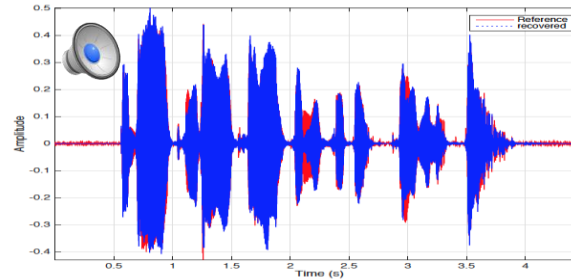


Recovered source 2

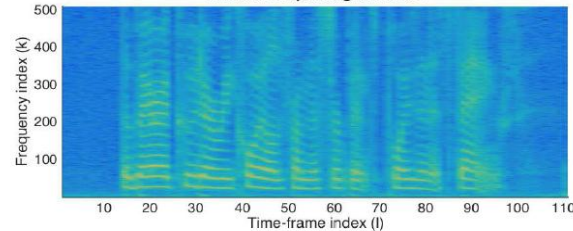
SNR_{seg} : 8.70

SNR: 11.33

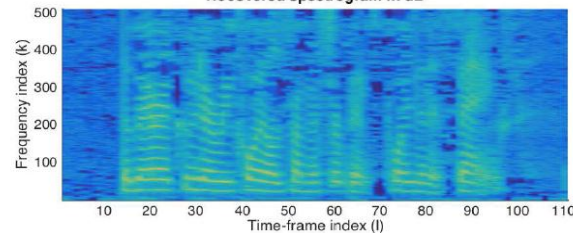
STOI: 0.89



Reference spectrogram in dB



Recovered spectrogram in dB

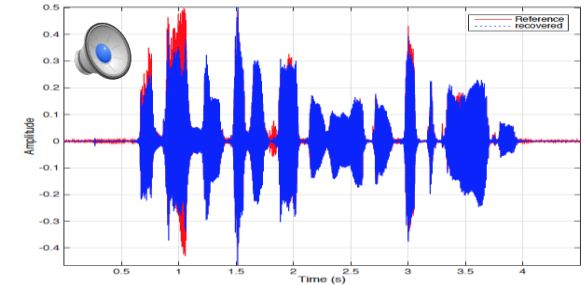


Recovered source 3

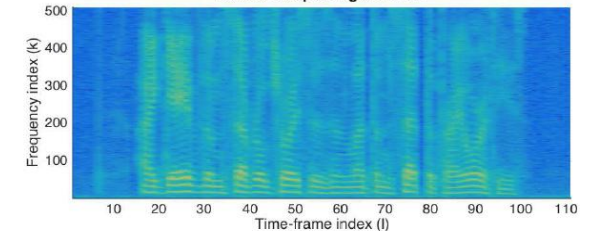
SNR_{seg} : 10.77

SNR: 8.84

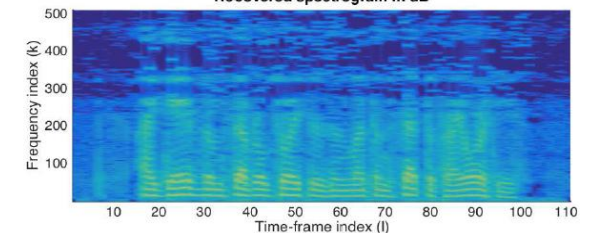
STOI: 0.90



Reference spectrogram in dB

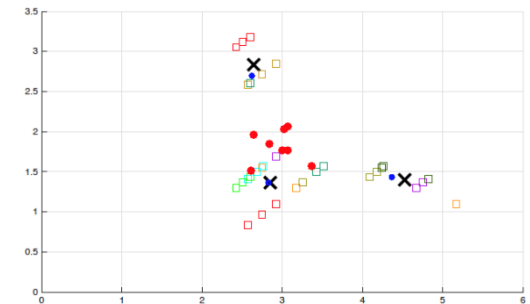
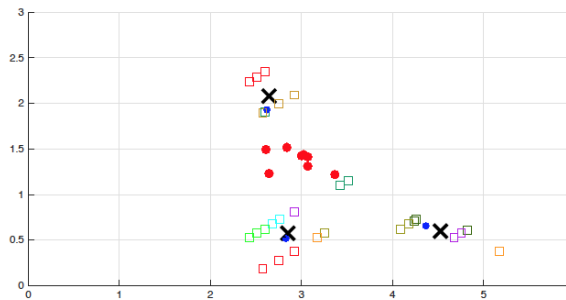
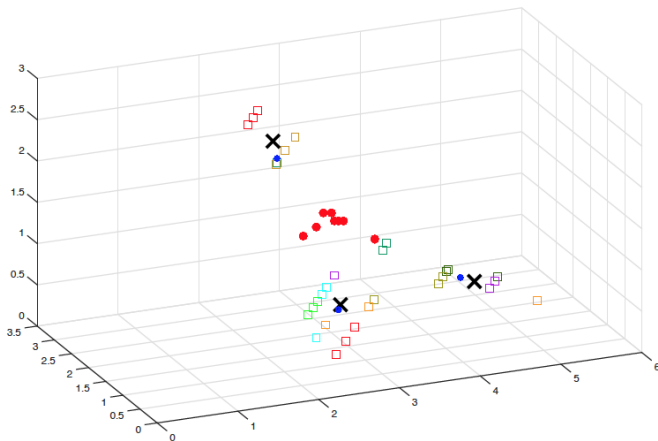


Recovered spectrogram in dB



Simulations

- ▶ Off-grid scenario: sources lying away from the grid points
- ▶ 8 microphones in random locations (red dots)
- ▶ 3 active sources (blue dots)
- ▶ Free-field assumption
- ▶ #grid-shifts=15 , $C_{th}=0.35$, $\varepsilon=0.01$
- ▶ 10 highest peaks (frequencies) of each spectrogram selected
- ▶ No reverberation but added diffuse noise @ 5dB SNR

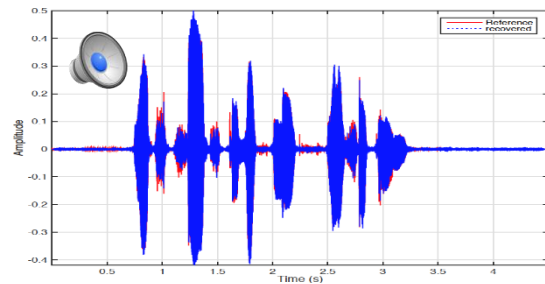


Simulations

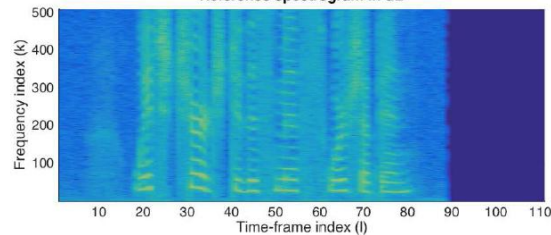
Part II

Recovered source 1

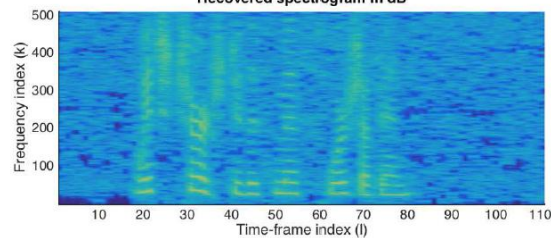
SNR_{seg} : 7.05
SNR: 11.51
STOI: 0.96



Reference spectrogram in dB

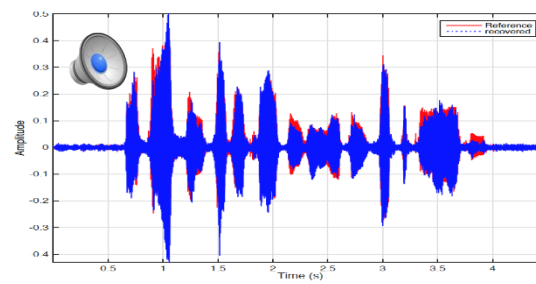


Recovered spectrogram in dB

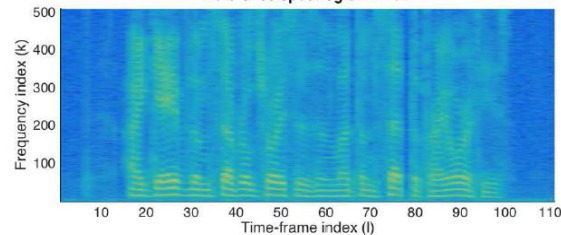


Recovered source 2

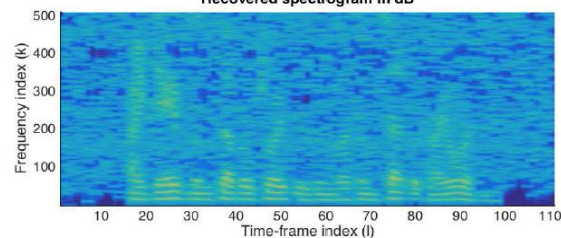
SNR_{seg} : 5.71
SNR: 8.75
STOI: 0.77



Reference spectrogram in dB

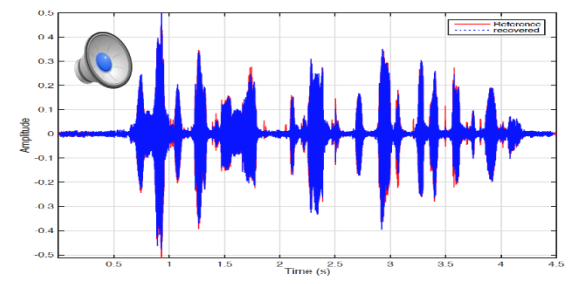


Recovered spectrogram in dB

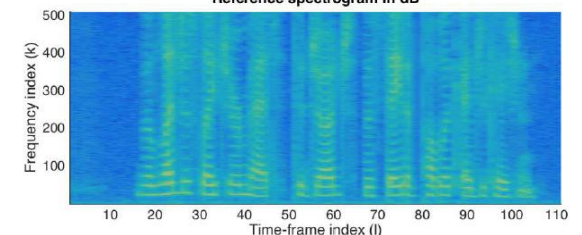


Recovered source 3

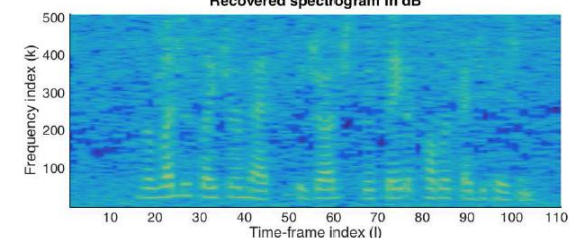
SNR_{seg} : 7.85
SNR: 12.20
STOI: 0.92



Reference spectrogram in dB



Recovered spectrogram in dB



Conclusions

- ▶ Wideband source localization and acquisition
- ▶ Spatial sparsity
 - ▶ Discretize the acoustic scene
 - ▶ Solve an L1-norm minimization problem
- ▶ Use grid-shift process to compensate for the off-grid scenario
- ▶ Tested under clean and noisy data conditions
 - ▶ Effective source localization and acquisition
- ▶ Next steps
 - ▶ Finer grid can be used on acoustic scene regions containing most energy
 - ▶ For static sources e.g. teleconference, joint-sparsity could be exploited to model the same sparsity pattern for all frequencies

**Thank You
Questions!**