Raking Echoes in the Time-Domain

Robin Scheibler Ivan Dokmanić Martin Vetterli

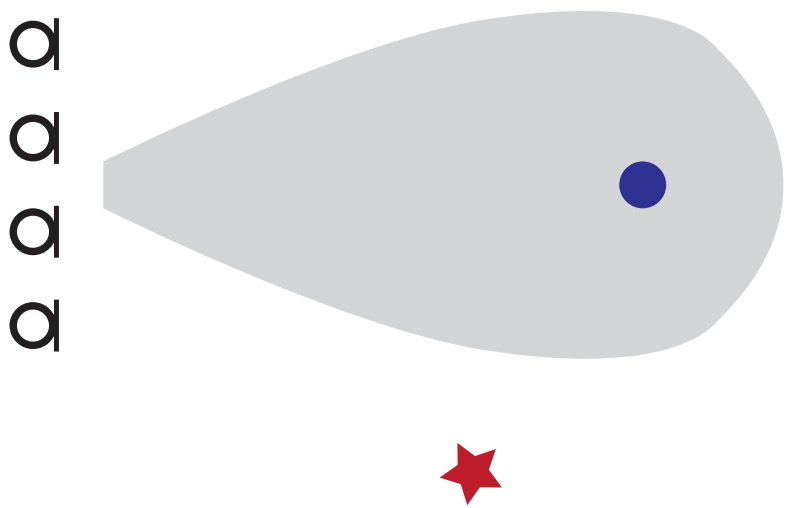
School of Computer and Communication Sciences
École Polytechnique Fédérale de Lausanne, Switzerland

Tmonday Afternoon Meeting
April 1, 2015

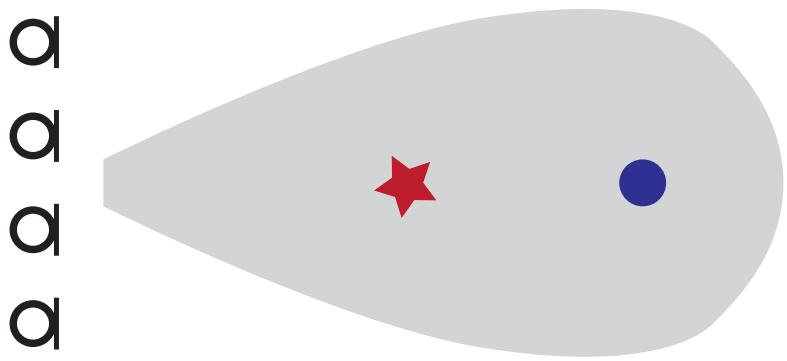
Echoes are good!

a
a
a
a

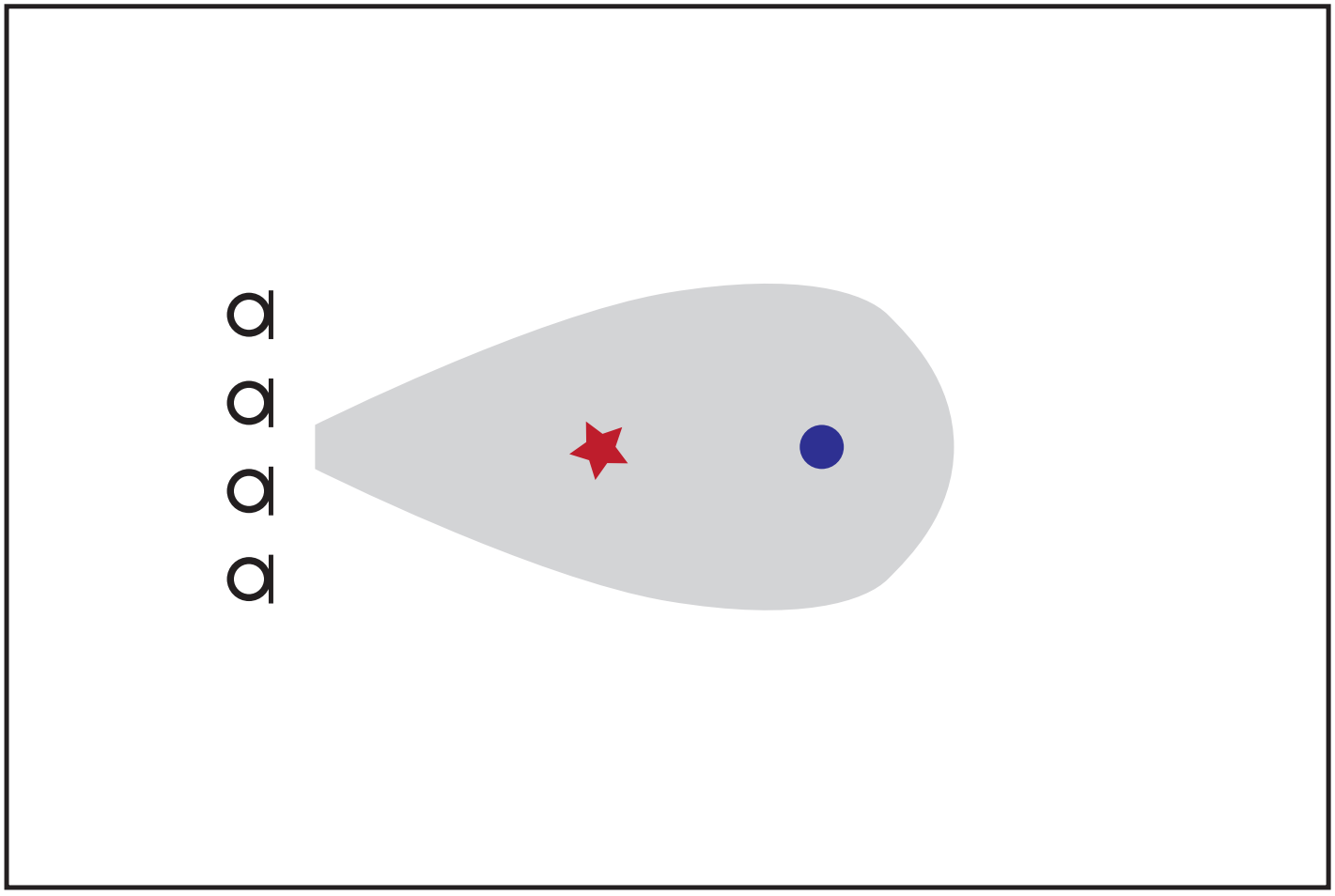
Echoes are good!



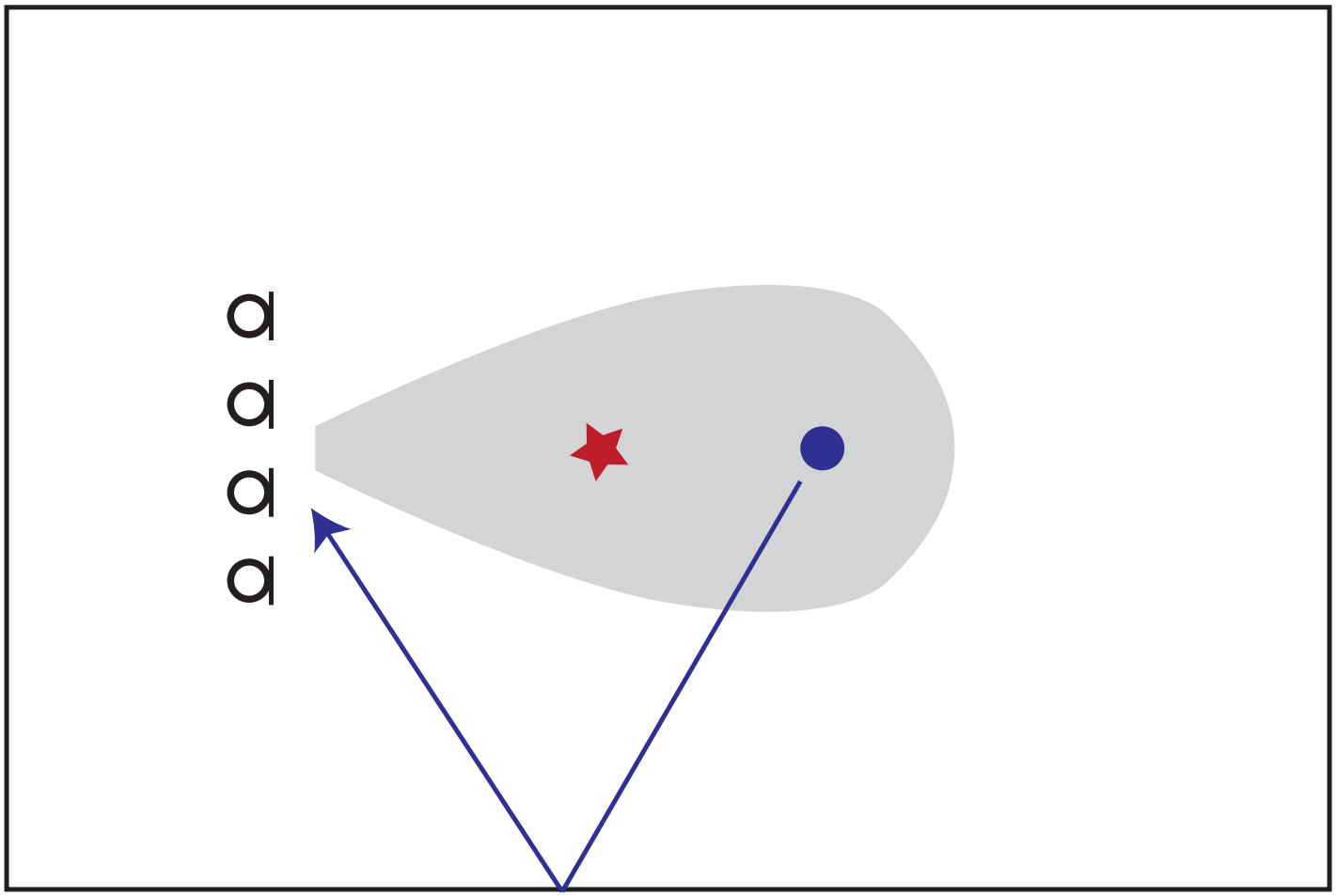
Echoes are good!



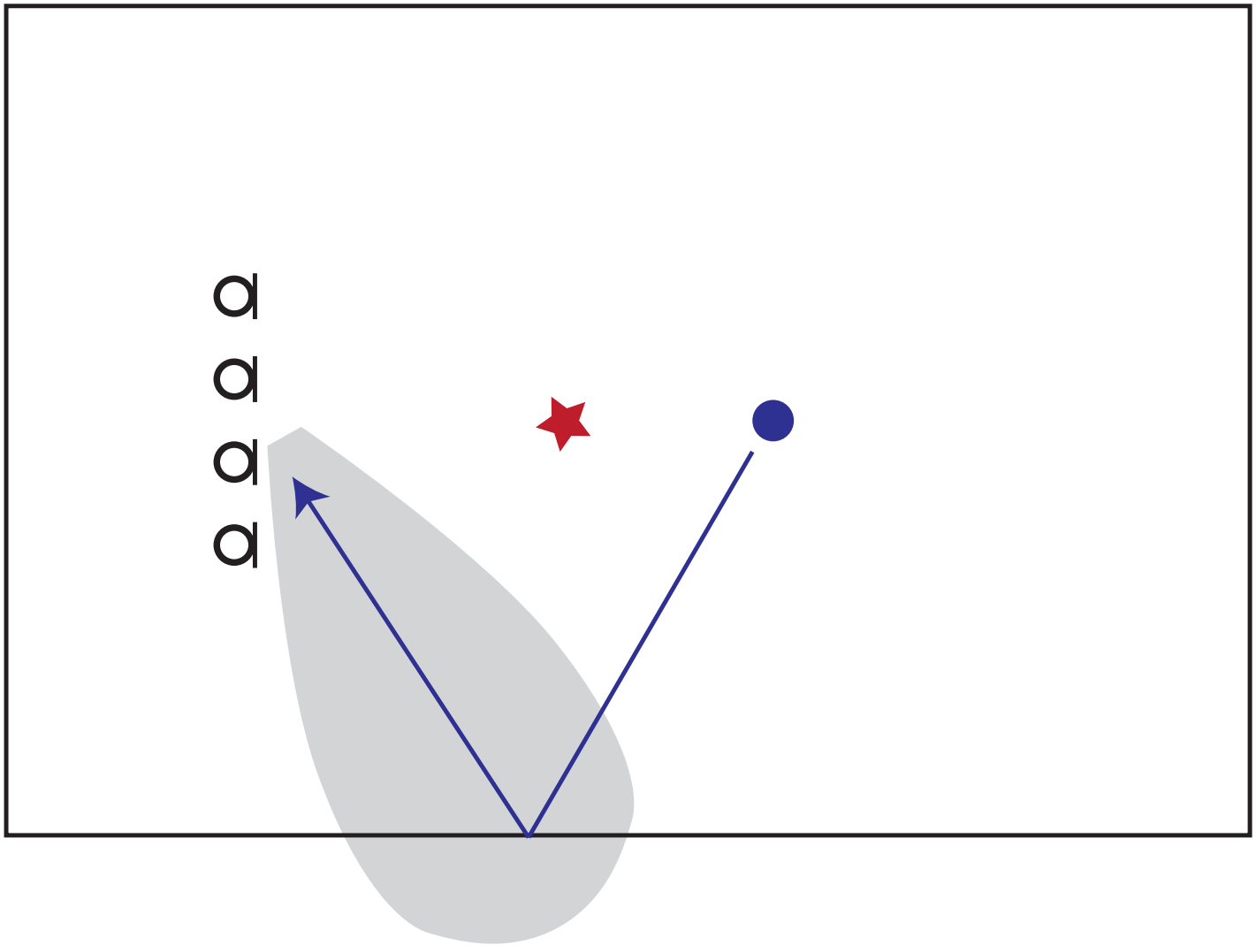
Echoes are good!



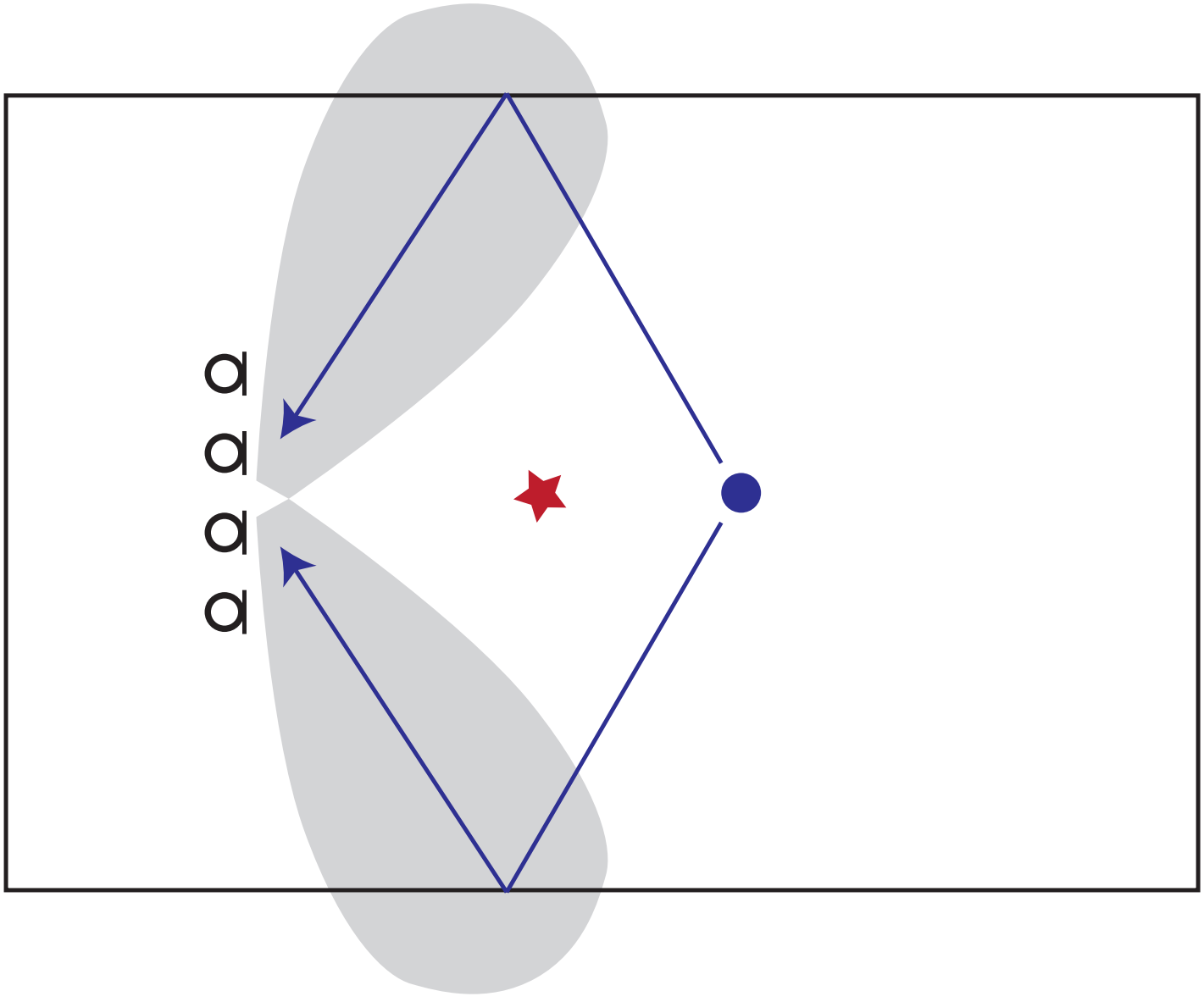
Echoes are good!



Echoes are good!



Echoes are good!



Outline

1. Model and notation
2. Raking beamformers
3. Results

Image source model

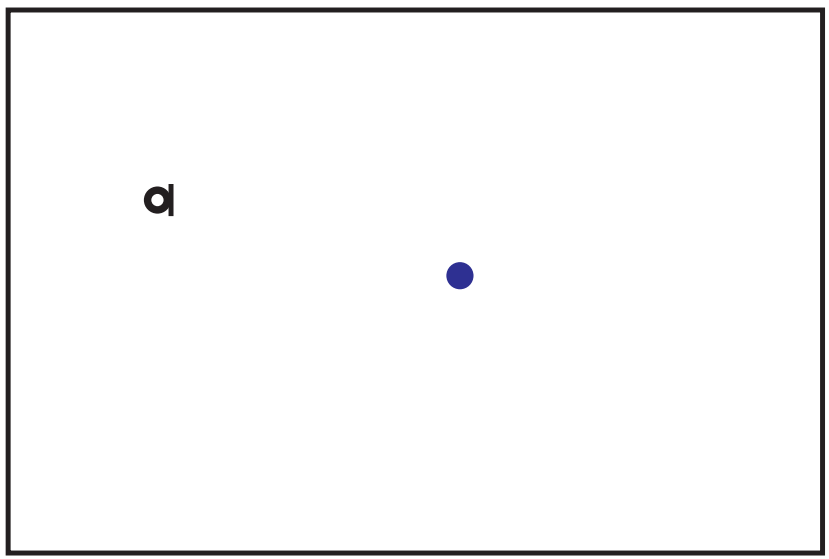


Image source model

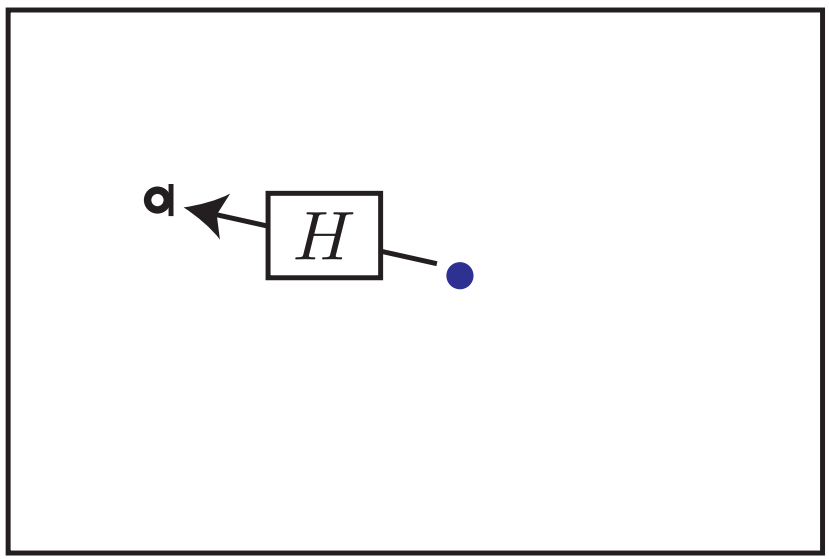


Image source model

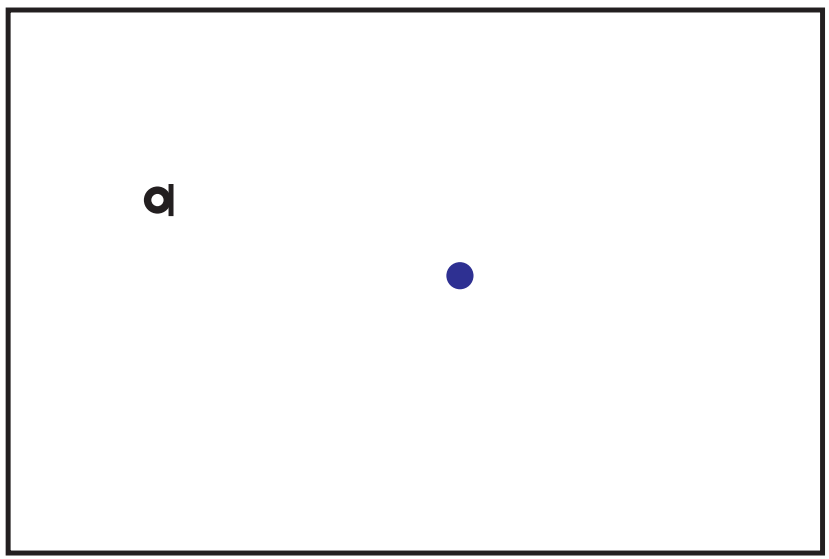


Image source model

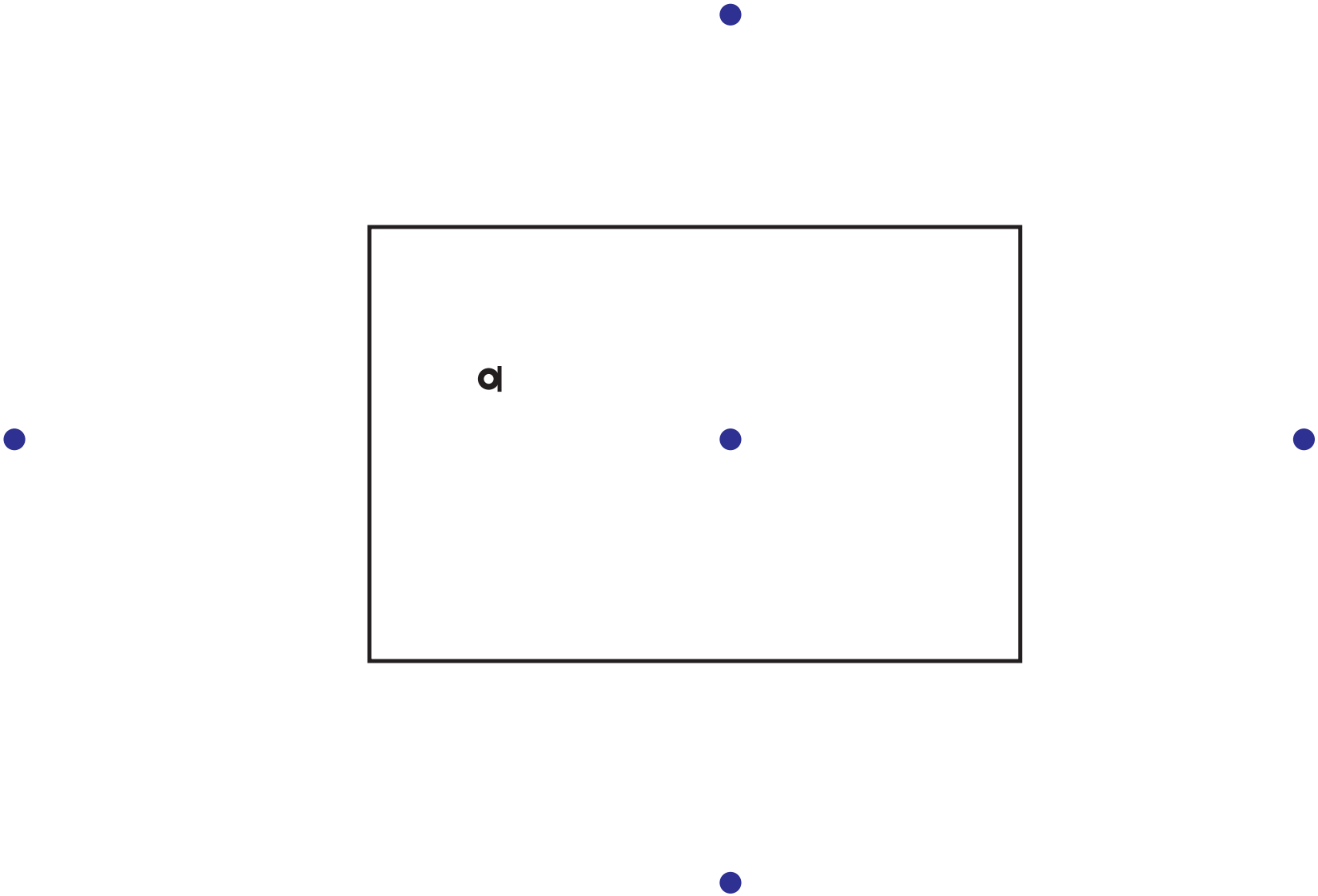


Image source model

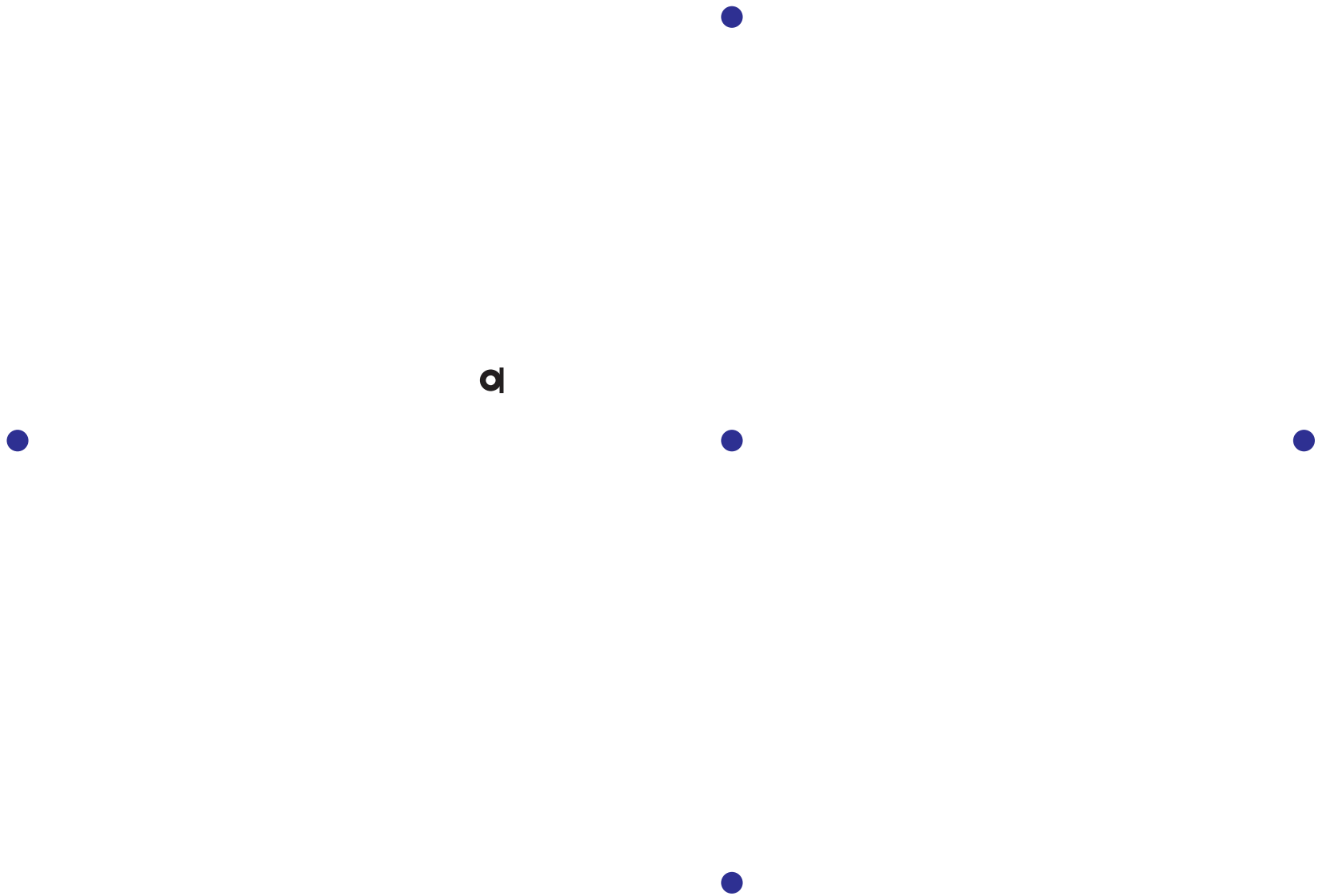
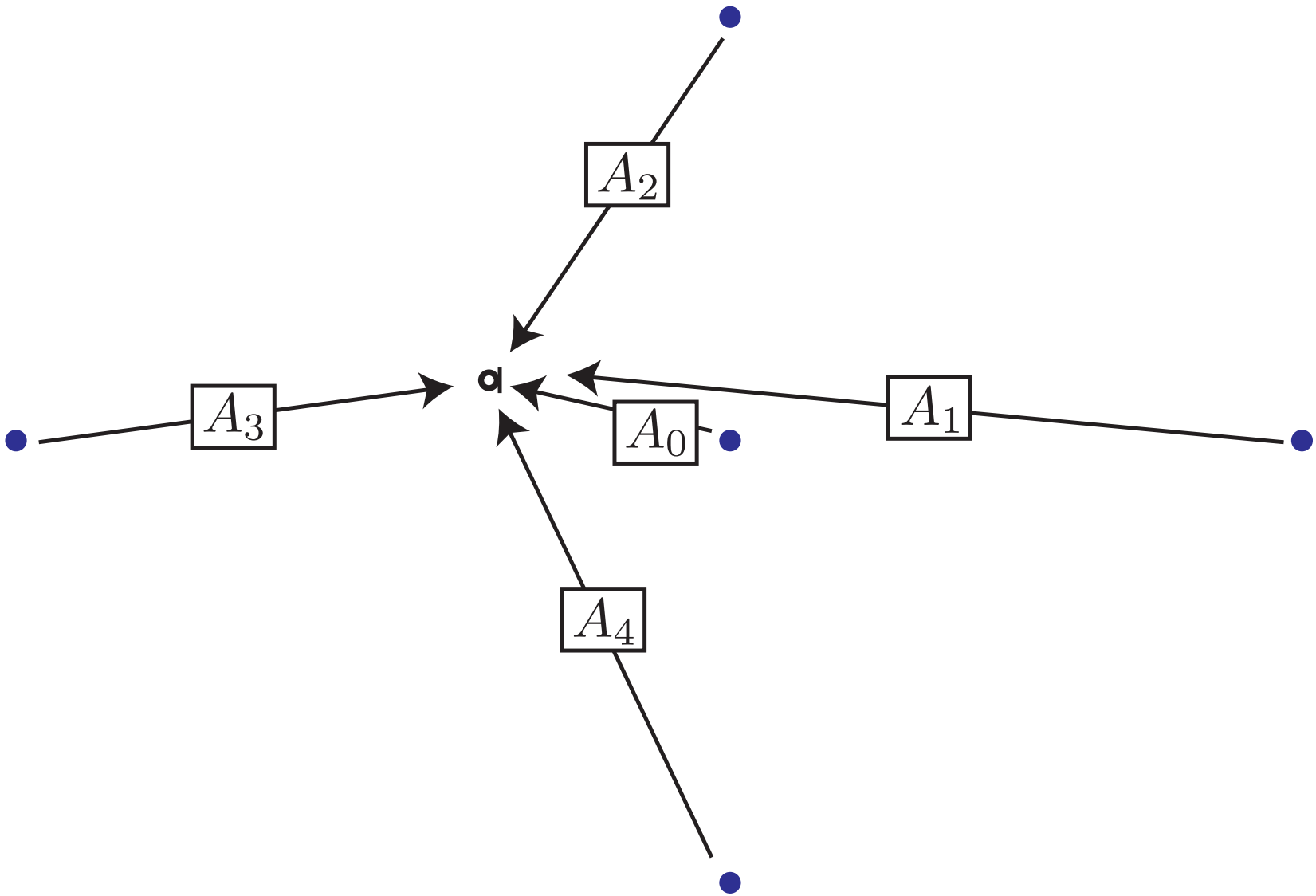
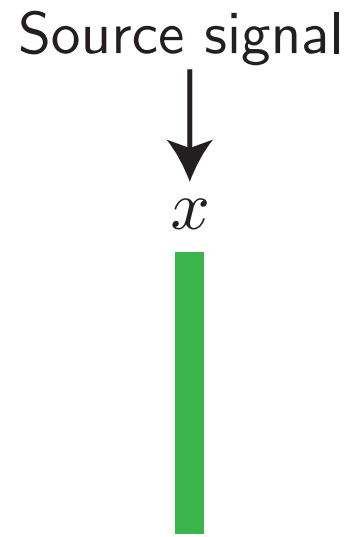


Image source model



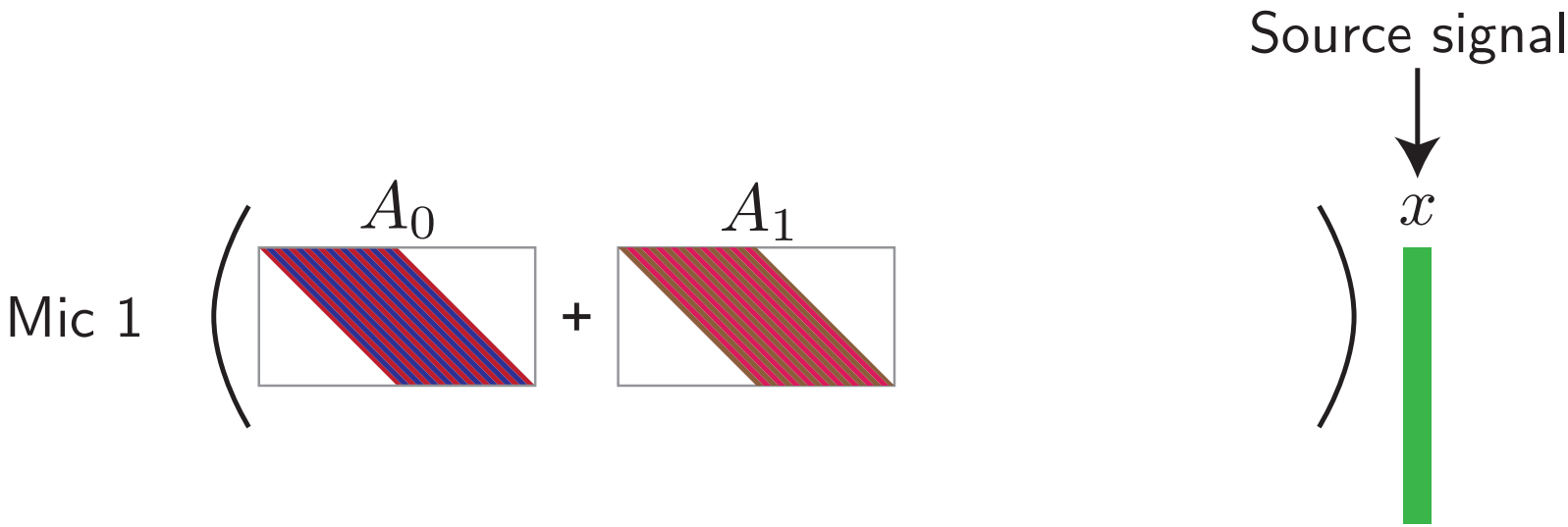
Compact notation



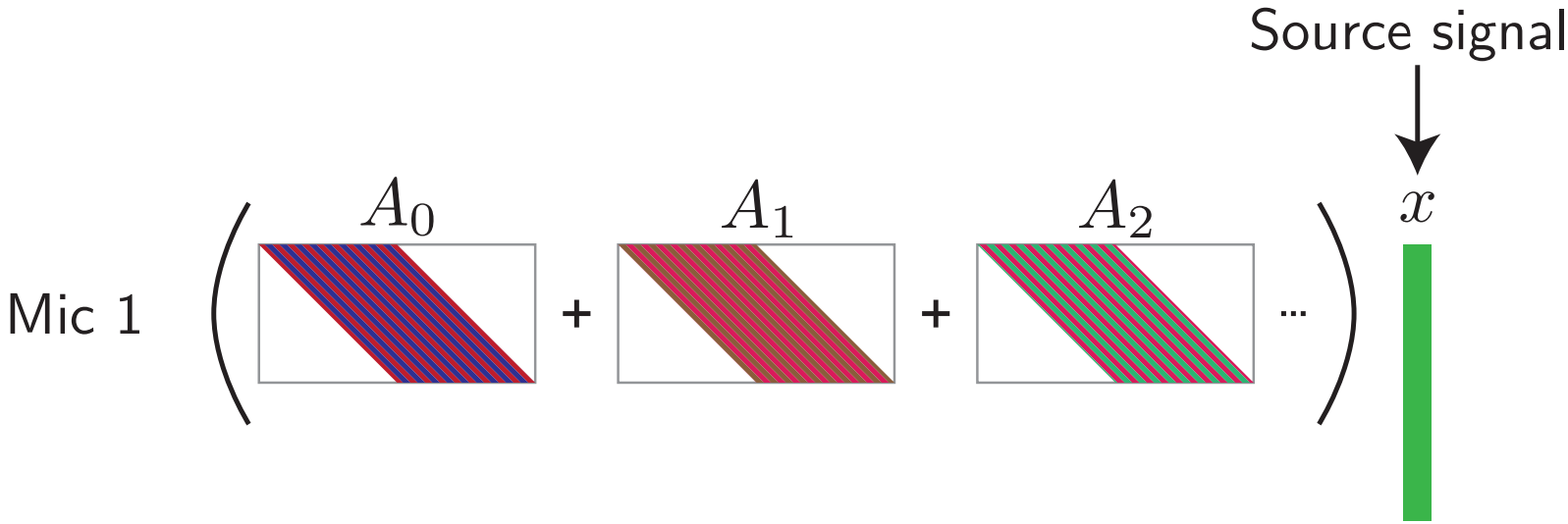
Mic 1

Compact notation

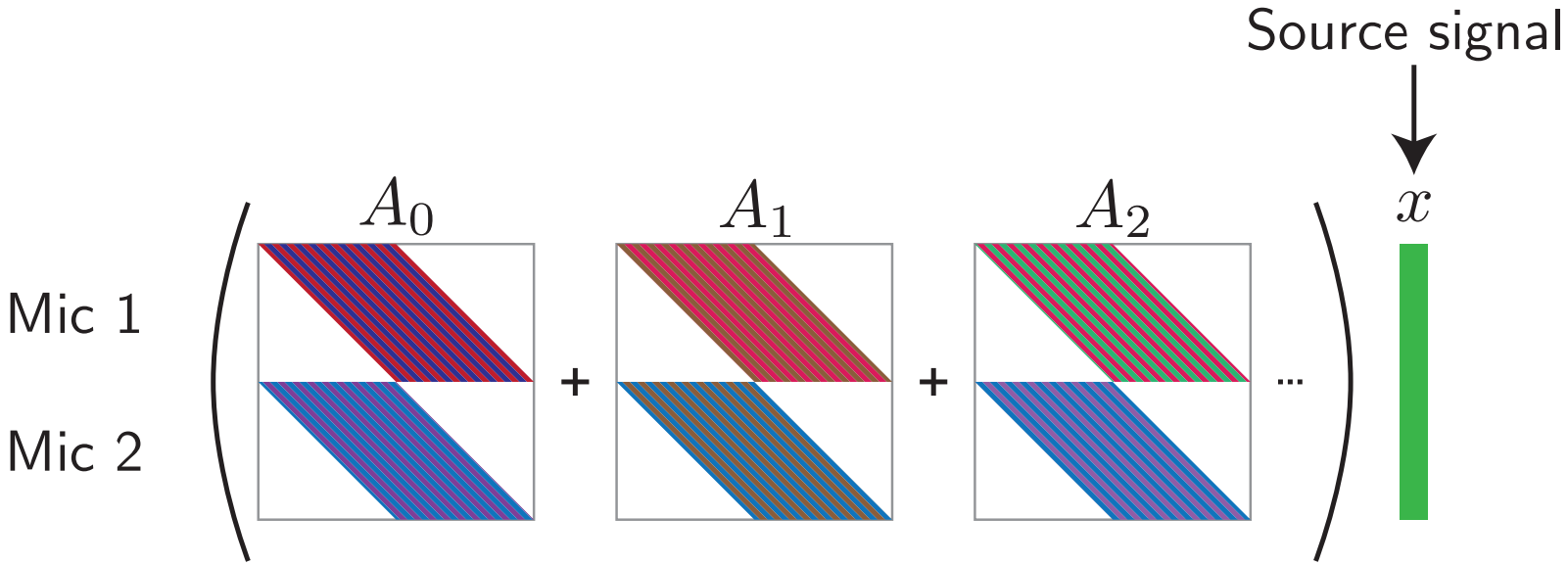
Compact notation



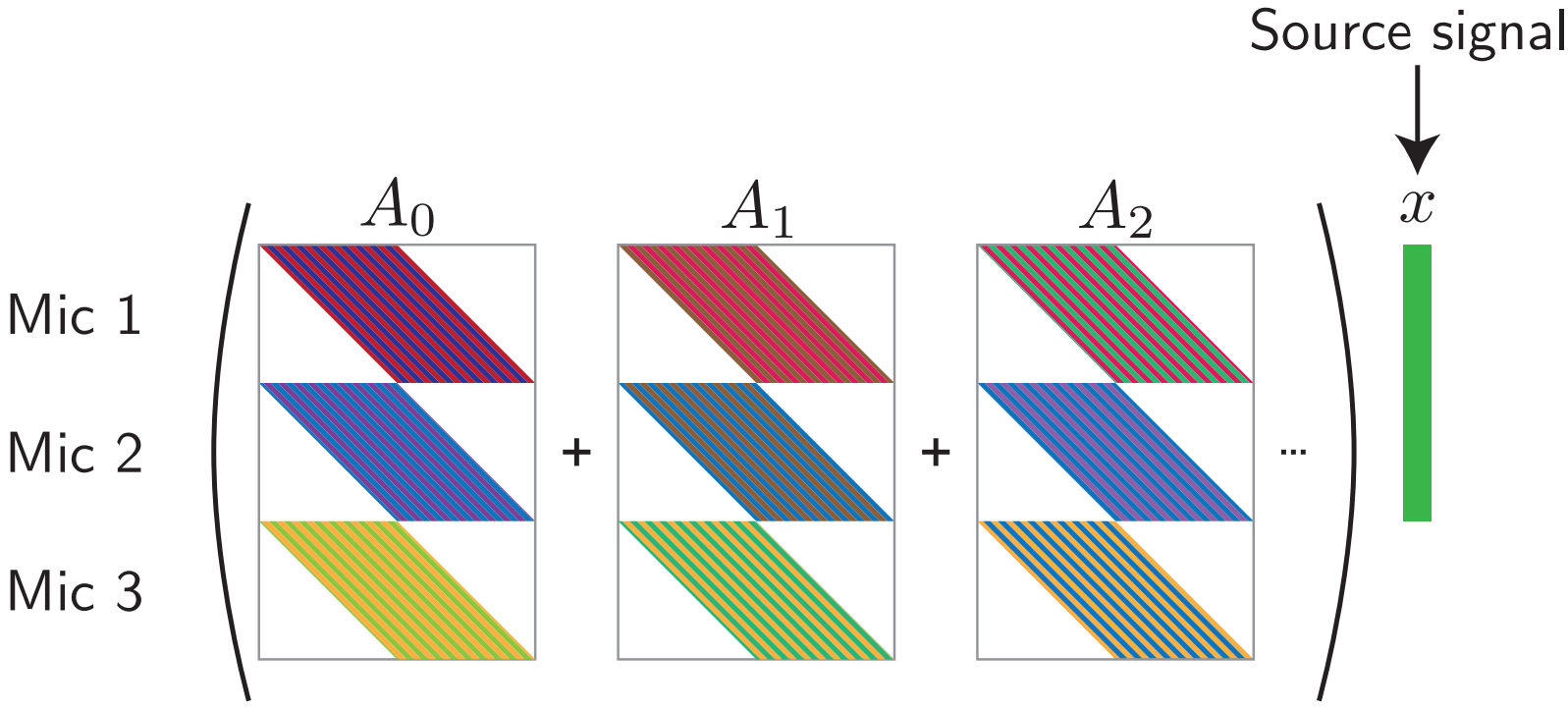
Compact notation



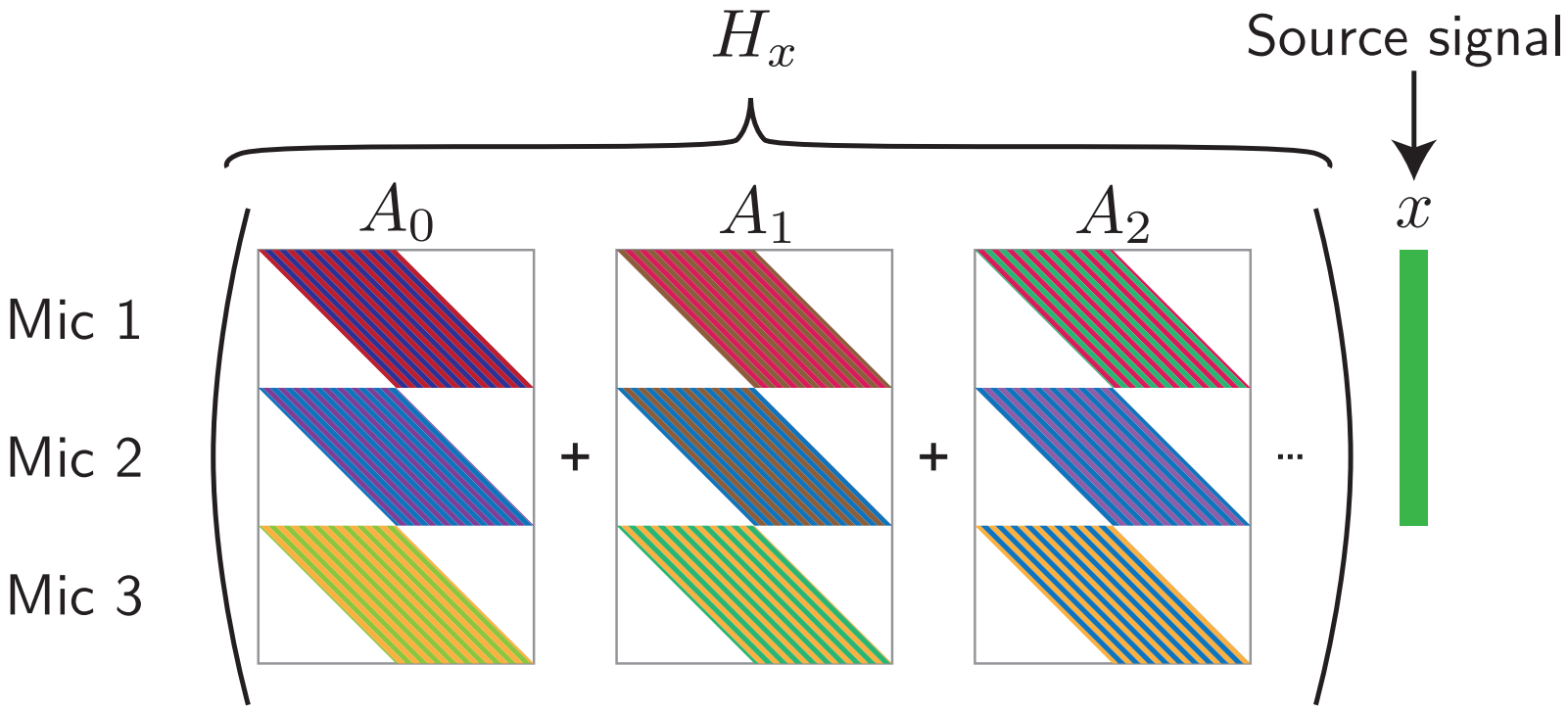
Compact notation



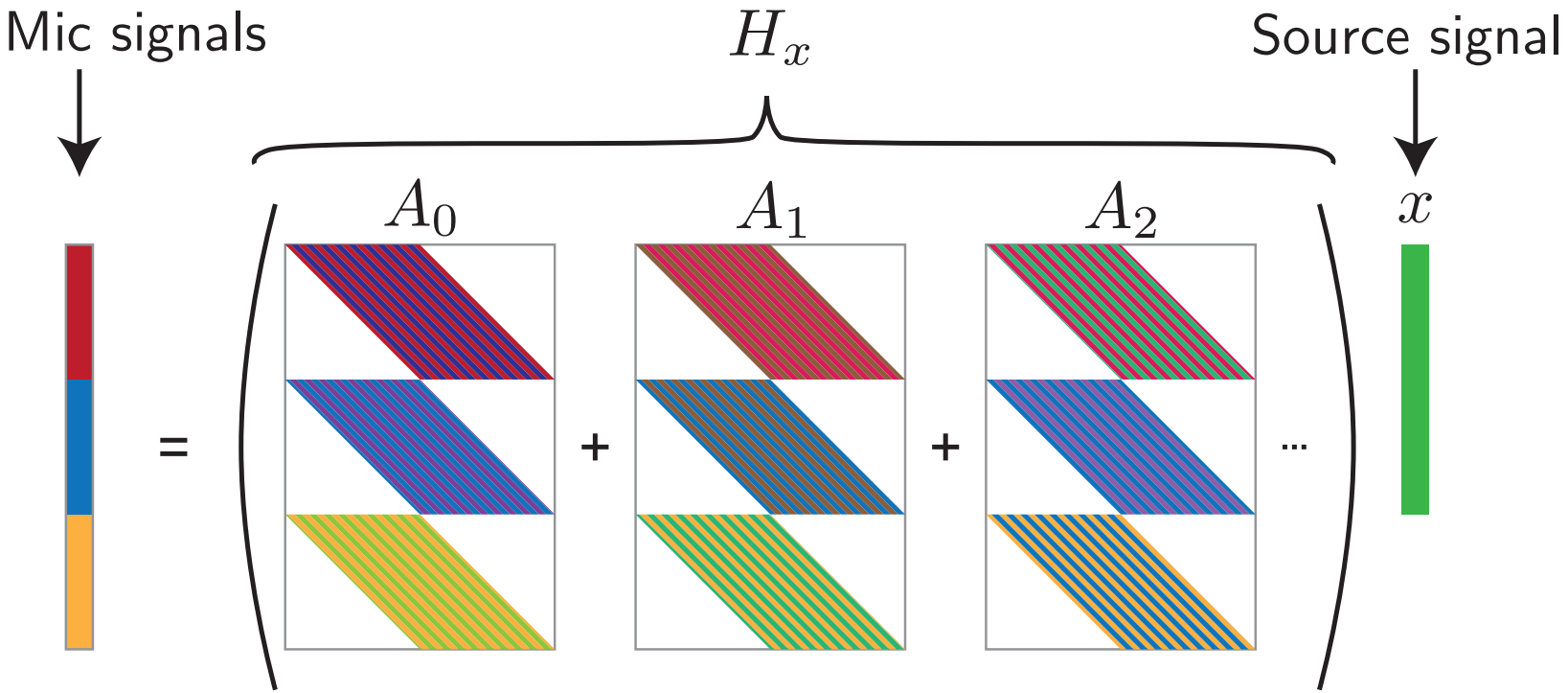
Compact notation



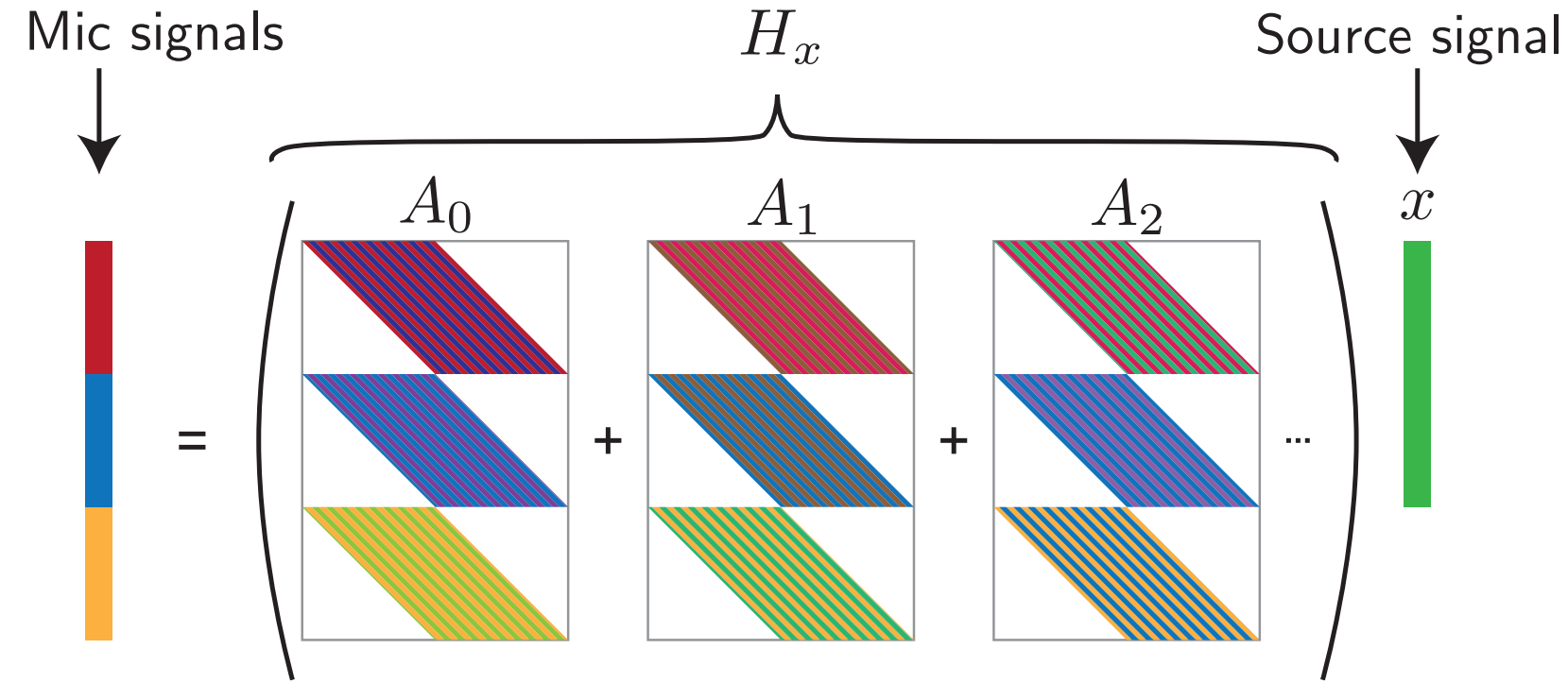
Compact notation



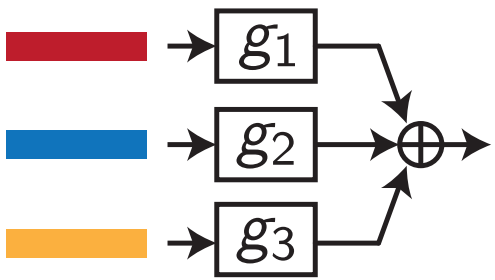
Compact notation



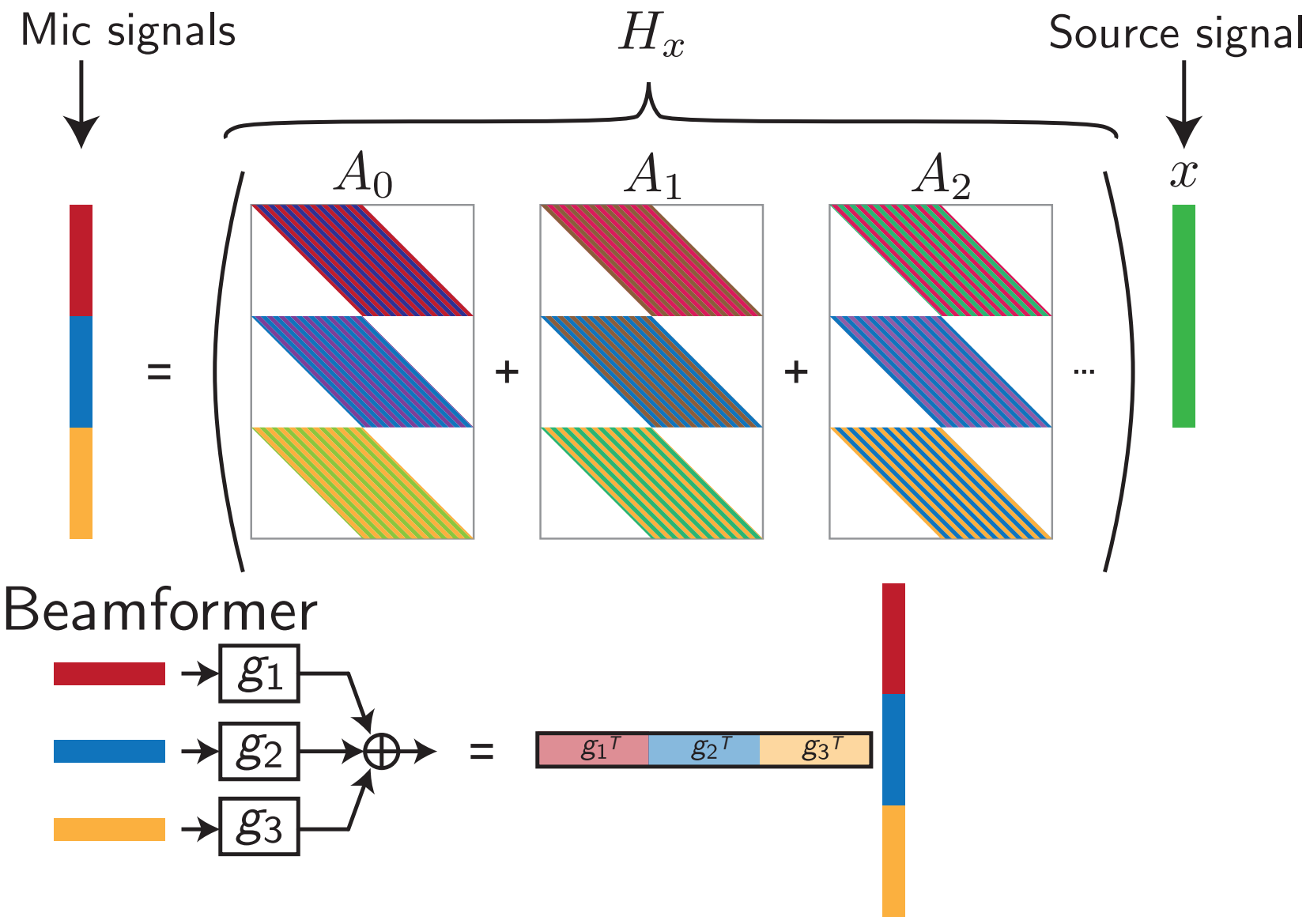
Compact notation



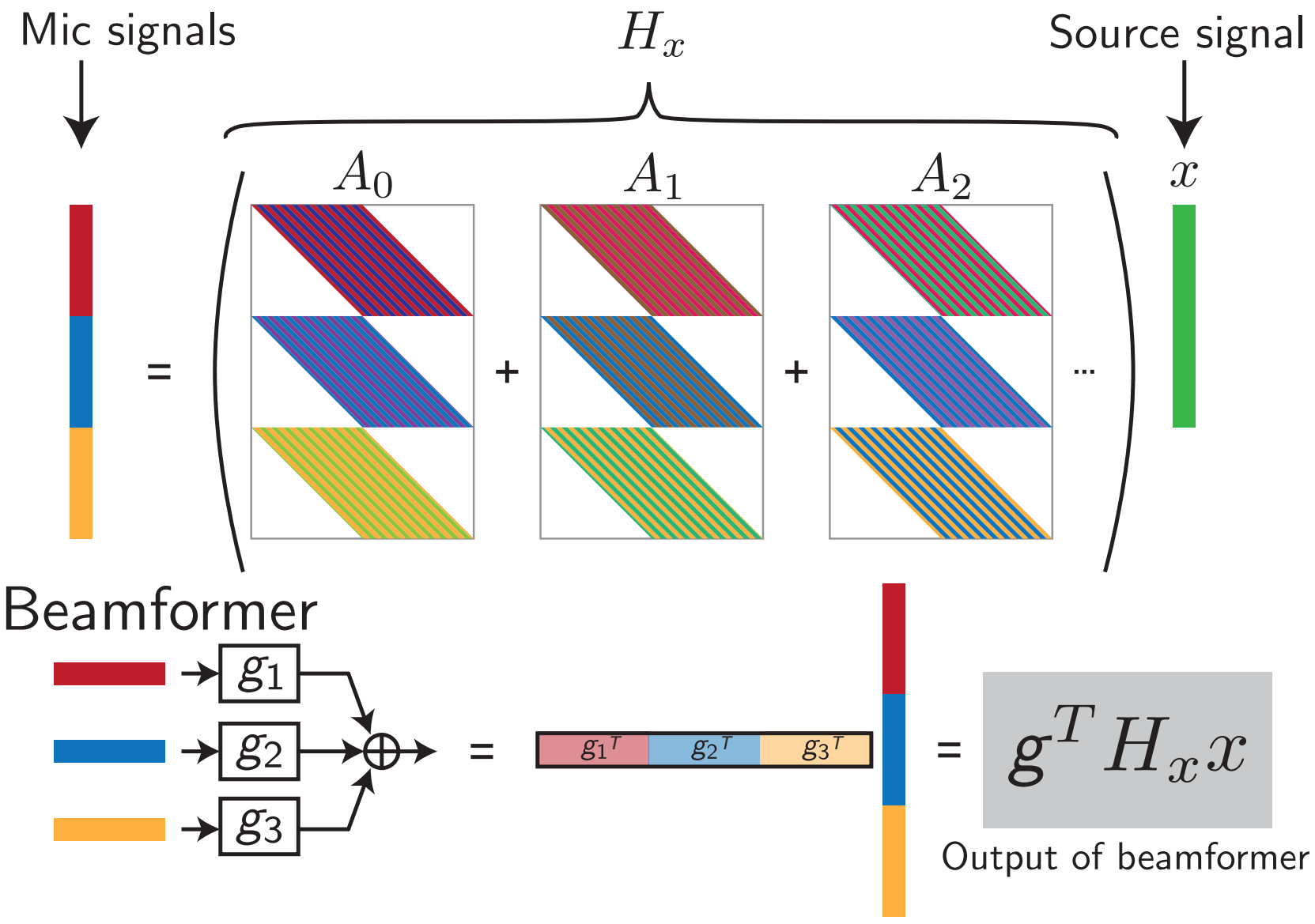
Beamformer



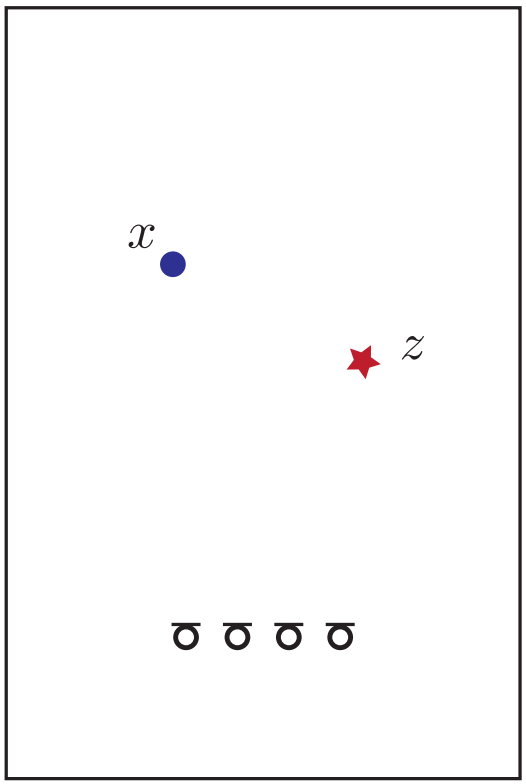
Compact notation



Compact notation



Scenario



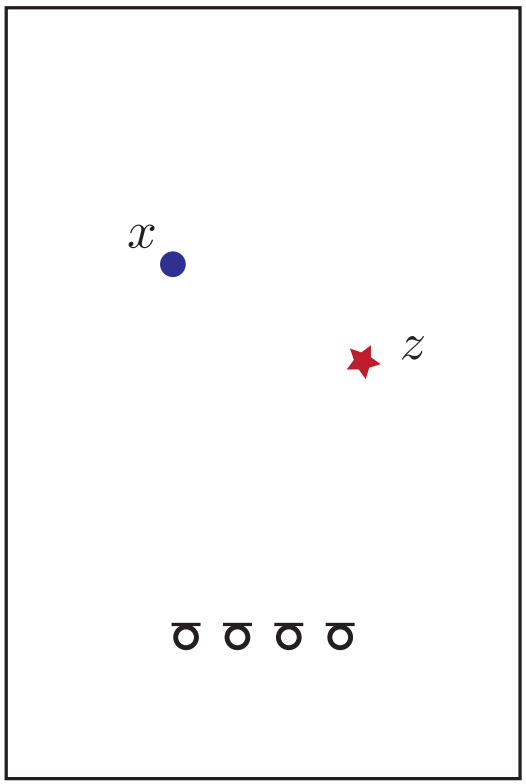
- Signal model:

$$\boldsymbol{y} = \boldsymbol{H}_x \boldsymbol{x} + \boldsymbol{H}_z \boldsymbol{z} + \boldsymbol{n}$$

- $\boldsymbol{H}_x$ and $\boldsymbol{H}_z$ constructed from geometry
- Beamformer response

$$\boldsymbol{u}_x = \boldsymbol{H}_x^T \boldsymbol{g} \qquad \boldsymbol{u}_z = \boldsymbol{H}_z^T \boldsymbol{g}$$

Scenario



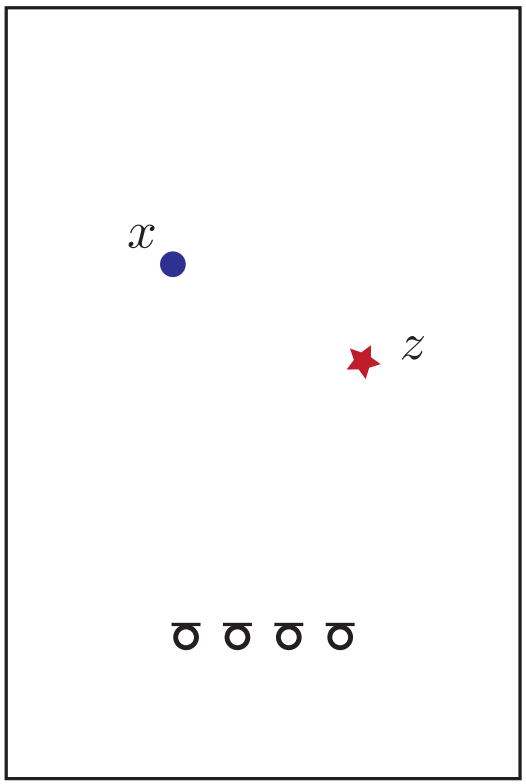
- Signal model:

$$\boldsymbol{y} = \boldsymbol{H}_x\boldsymbol{x} + \boldsymbol{H}_z\boldsymbol{z} + \boldsymbol{n}$$

- $\boldsymbol{H}_x$ and $\boldsymbol{H}_z$ constructed from geometry
- Beamformer response

$$\boldsymbol{u}_x = \boldsymbol{H}_x^T\boldsymbol{g} \qquad \boldsymbol{u}_z = \boldsymbol{H}_z^T\boldsymbol{g}$$

Scenario



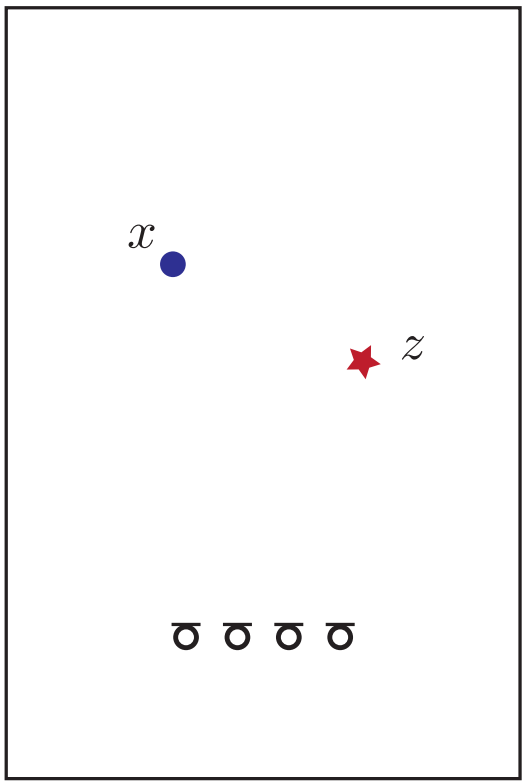
- Signal model:

$$\boldsymbol{y} = \boldsymbol{H}_x \boldsymbol{x} + \boldsymbol{H}_z \boldsymbol{z} + \boldsymbol{n}$$

- $\boldsymbol{H}_x$ and $\boldsymbol{H}_z$ constructed from geometry
- Beamformer response

$$\boldsymbol{u}_x = \boldsymbol{H}_x^T \boldsymbol{g} \qquad \boldsymbol{u}_z = \boldsymbol{H}_z^T \boldsymbol{g}$$

Scenario



- Signal model:

$$\boldsymbol{y} = \boldsymbol{H}_x\boldsymbol{x} + \boldsymbol{H}_z\boldsymbol{z} + \boldsymbol{n}$$

- $\boldsymbol{H}_x$ and $\boldsymbol{H}_z$ constructed from geometry
- Beamformer response

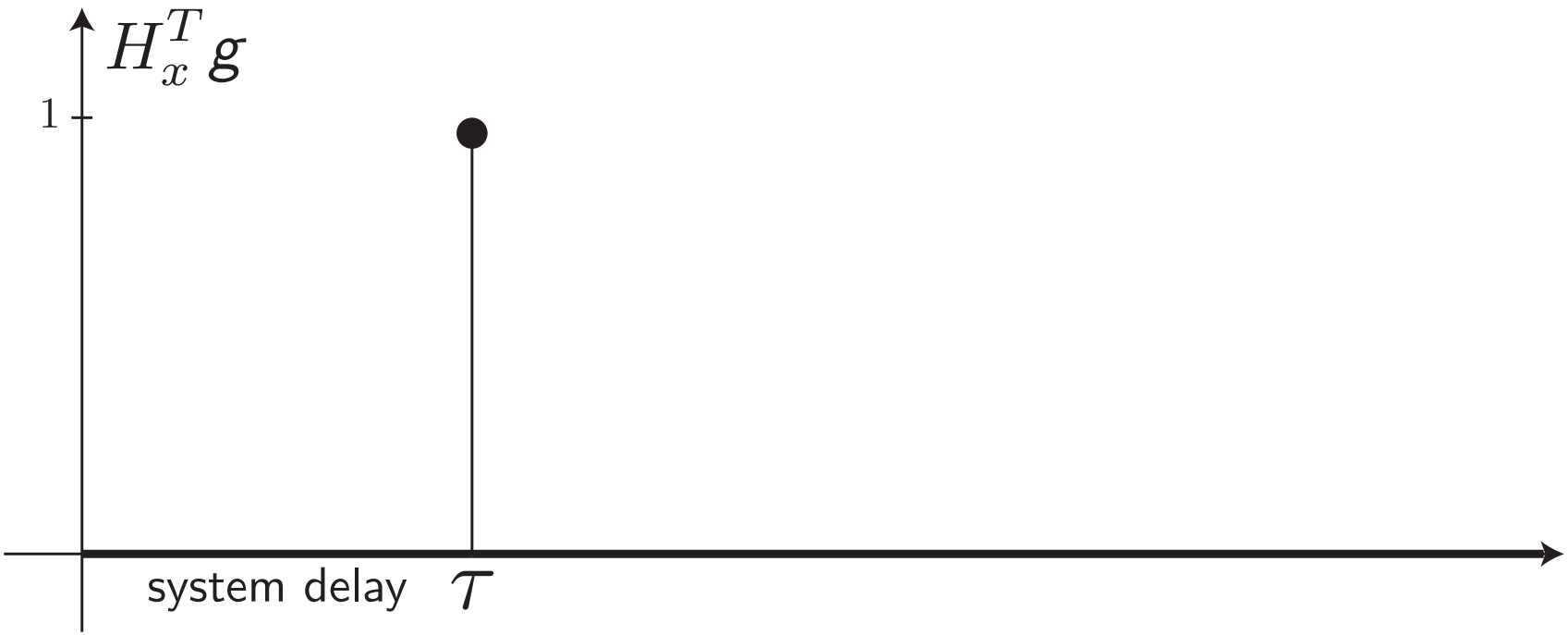
$$\boldsymbol{u}_x = \boldsymbol{H}_x^T\boldsymbol{g} \qquad \boldsymbol{u}_z = \boldsymbol{H}_z^T\boldsymbol{g}$$

Minimum variance distortionless response beamformer

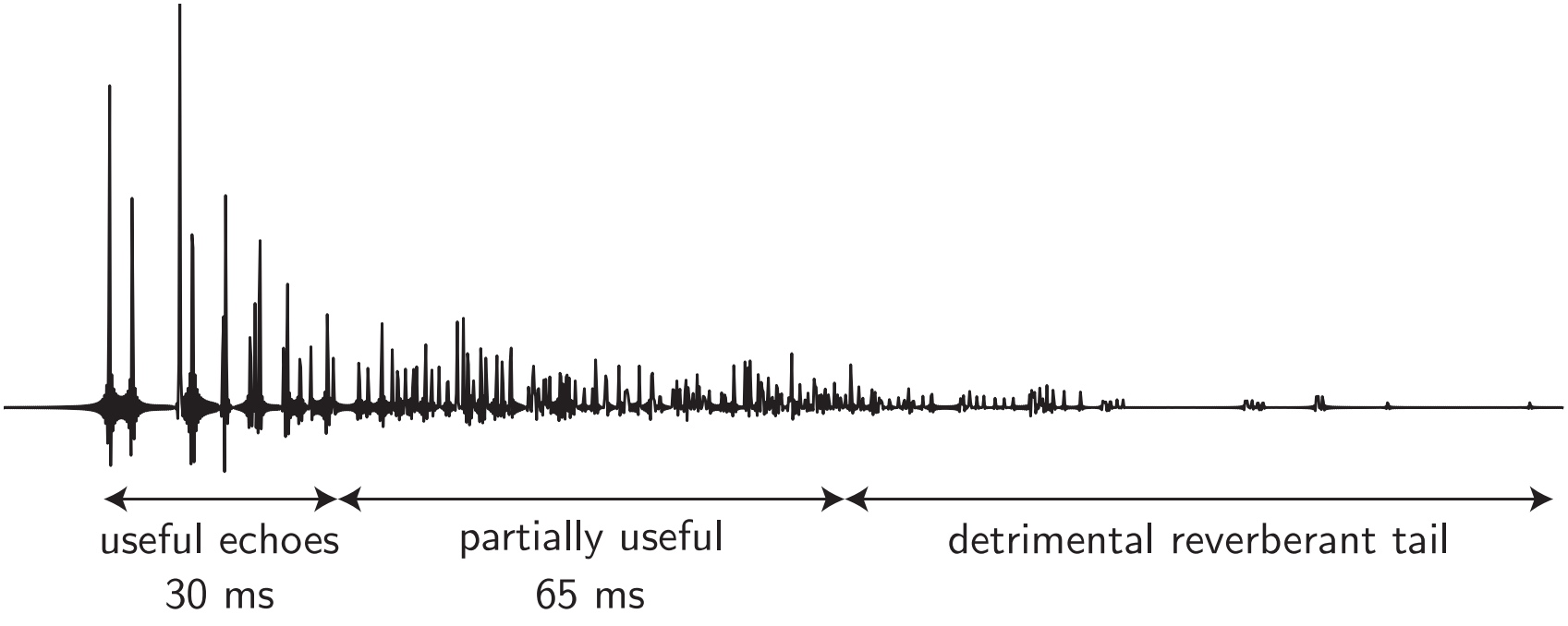
$$\underset{\boldsymbol{g}}{\text{minimize}}\ \mathbb{E}\|\boldsymbol{g}^T\boldsymbol{y}\|^2 \quad \text{subject to } \boldsymbol{H}_x^T\boldsymbol{g} = \delta_\tau$$

Minimum variance distortionless response beamformer

$$\underset{\boldsymbol{g}}{\text{minimize}}\ \mathbb{E}\|\boldsymbol{g}^T\boldsymbol{y}\|^2 \quad \text{subject to } \boldsymbol{H}_x^T\boldsymbol{g} = \delta_\tau$$



Clues from perceptual acoustics



J. Lochner, J.F. Burger, *The Influence of Reflections on Auditorium Acoustics*, 1964.

Relaxing the distortionless constraint

$$\underset{\boldsymbol{g}}{\text{minimize}}\ \mathbb{E}\|\boldsymbol{g}^T(\boldsymbol{H}_z\boldsymbol{z}+\boldsymbol{n})\|^2 \quad \text{subject to } \mathcal{M}\boldsymbol{H}_x^T\boldsymbol{g}=\delta_\tau$$

Relaxing the distortionless constraint

$$\underset{\boldsymbol{g}}{\text{minimize}}\ \mathbb{E}\|\boldsymbol{g}^T(\boldsymbol{H}_z\boldsymbol{z}+\boldsymbol{n})\|^2 \quad \text{subject to } \mathcal{M}\boldsymbol{H}_x^T\boldsymbol{g}=\delta_\tau$$

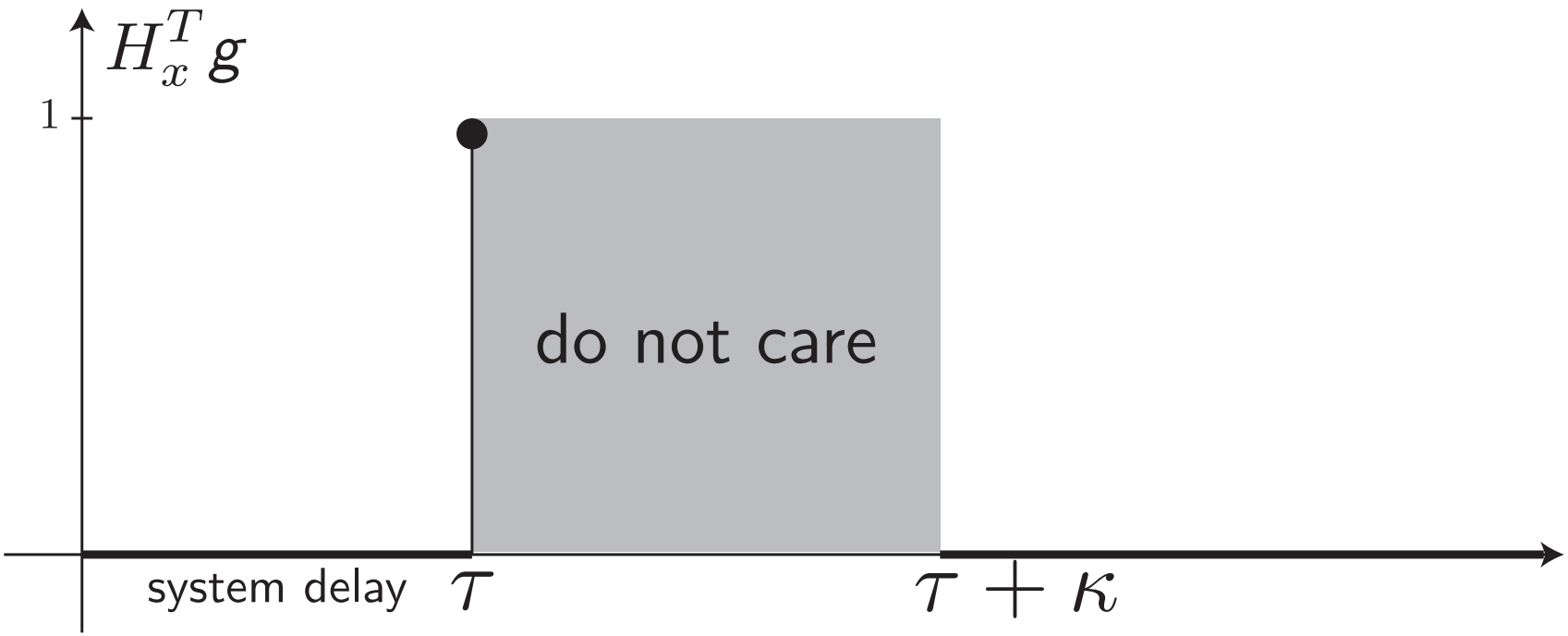


Image source picking

- MVDR: Sources within $c\tau$ contribute energy
- Perceptual: Sources within $c(\tau + \kappa)$ contribute energy

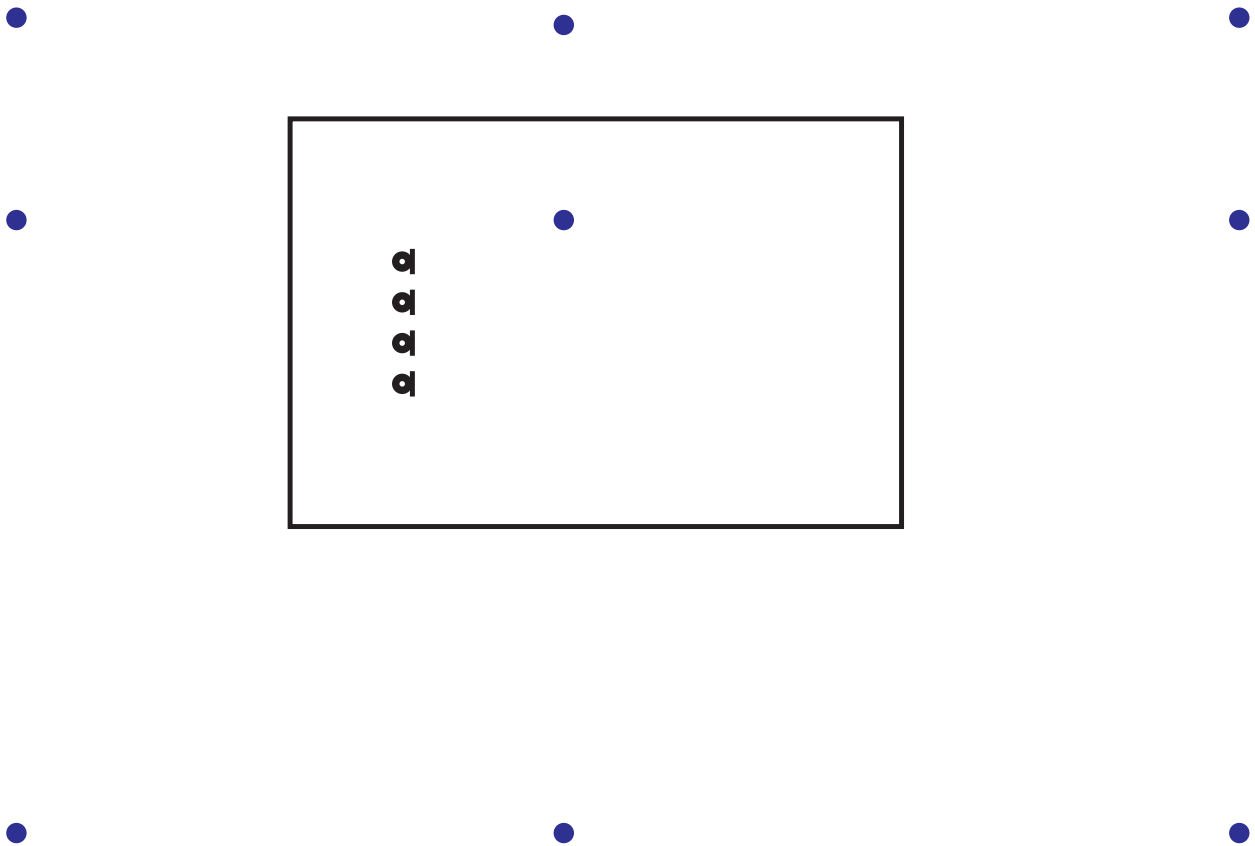


Image source picking

- MVDR: Sources within $c\tau$ contribute energy
- Perceptual: Sources within $c(\tau + \kappa)$ contribute energy

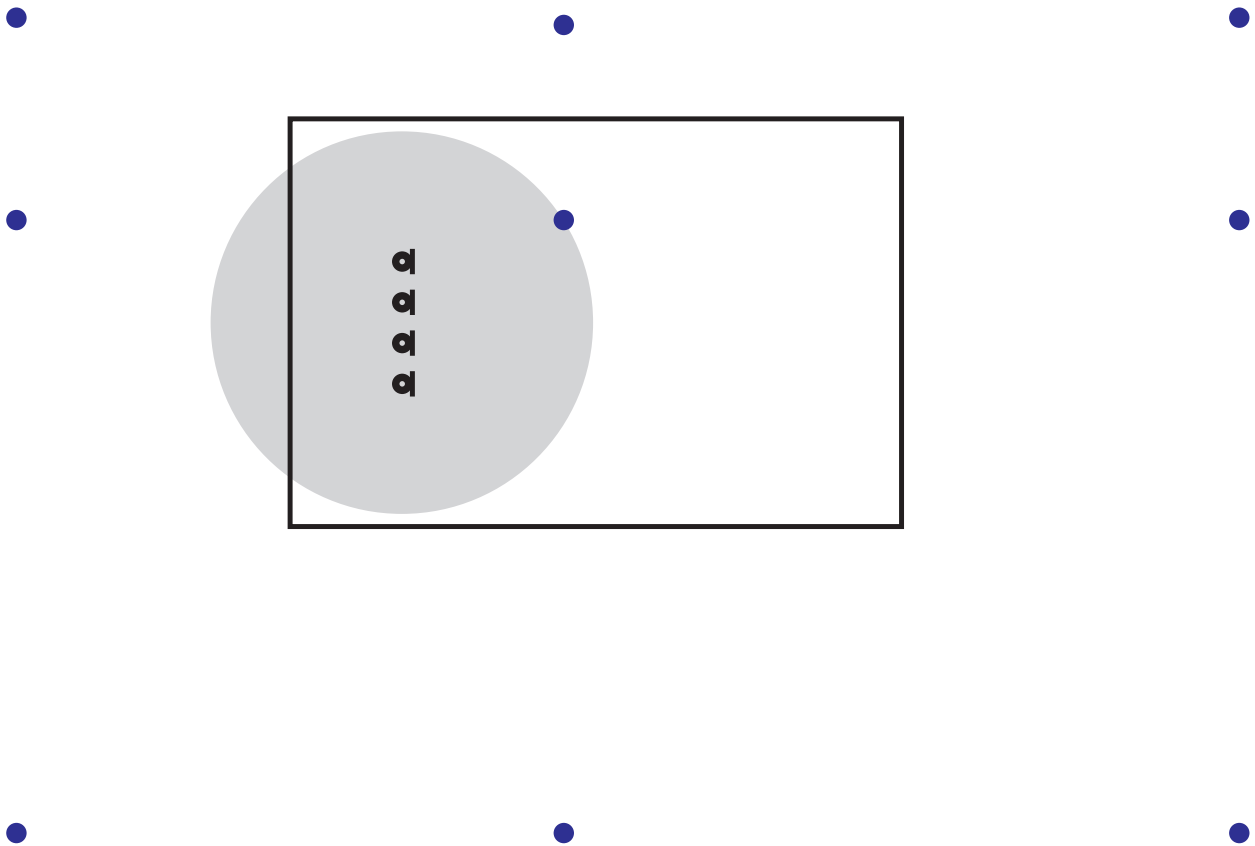


Image source picking

- MVDR: Sources within $c\tau$ contribute energy
- Perceptual: Sources within $c(\tau + \kappa)$ contribute energy

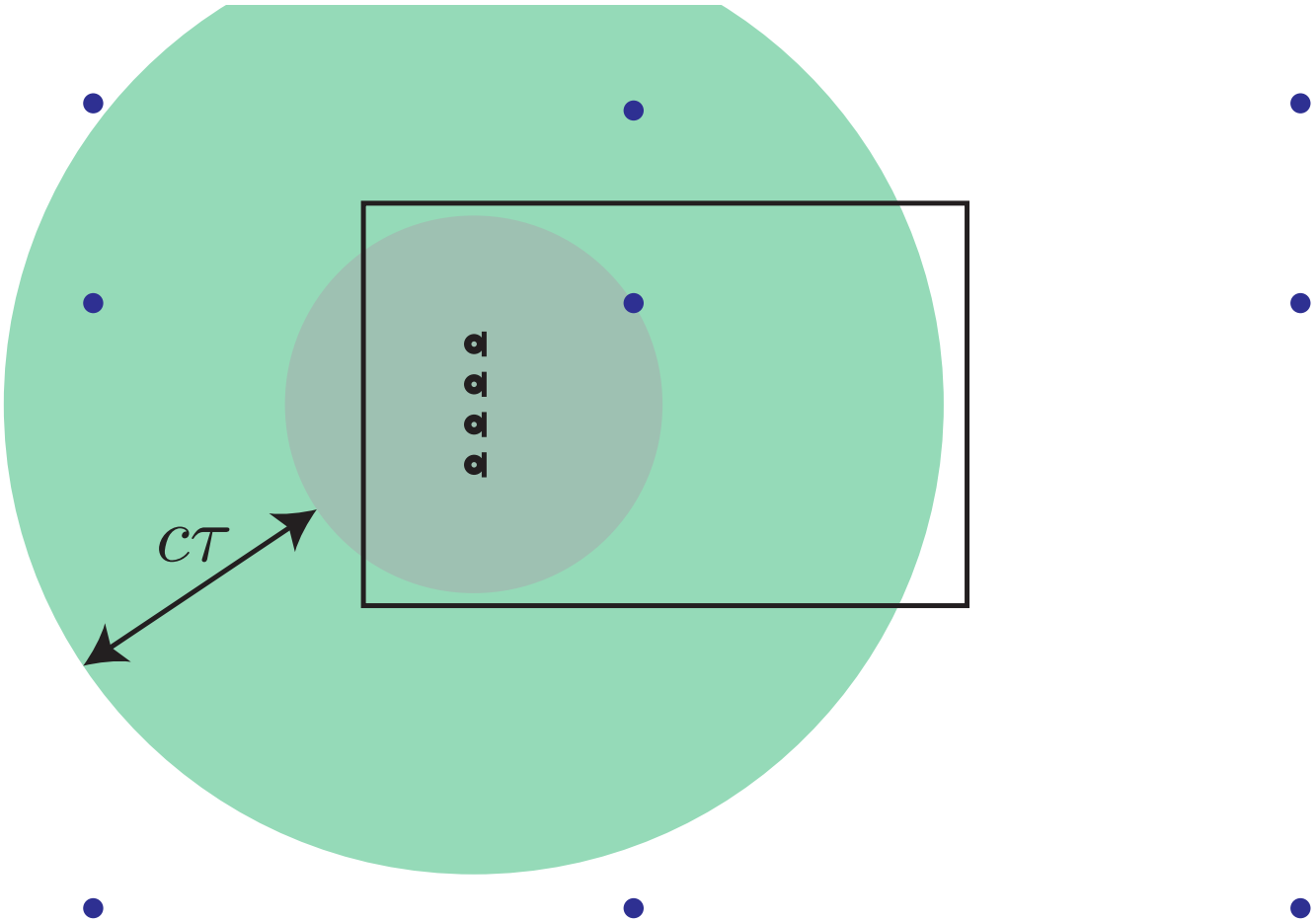
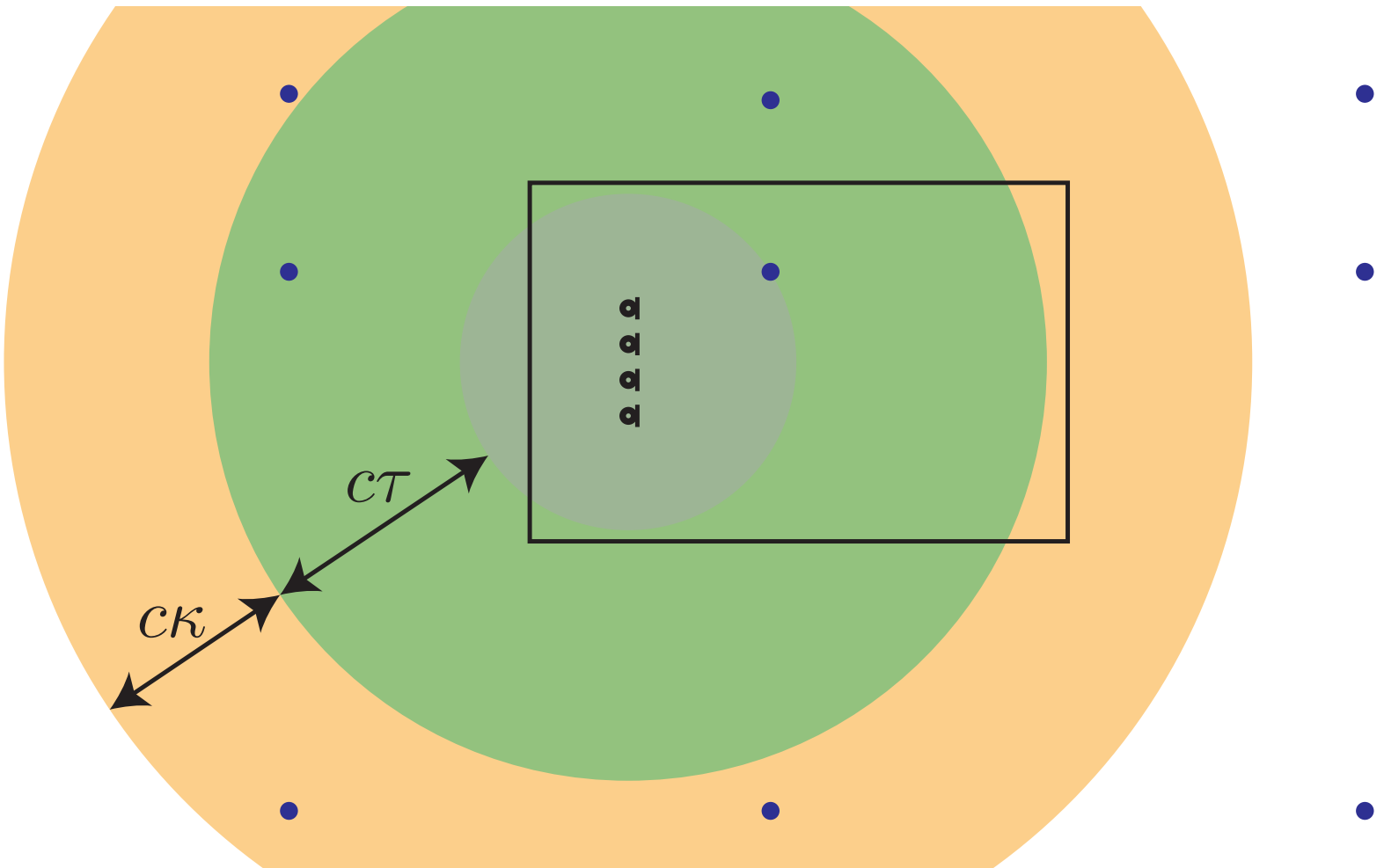


Image source picking

- MVDR: Sources within $c\tau$ contribute energy
- Perceptual: Sources within $c(\tau + \kappa)$ contribute energy



Signal-to-interference-and-noise metric

- Performance metric:

$$\text{SINR} = \mathbb{E}\left[\frac{\|\boldsymbol{g}^T \boldsymbol{H}_x \boldsymbol{x}\|^2}{\|\boldsymbol{g}^T(\boldsymbol{H}_z \boldsymbol{z} + \boldsymbol{n})\|^2}\right]$$

- Optimal beamformer:

$$\underset{\boldsymbol{g}}{\text{maximize}}\ \text{SINR}$$

– Response very distorted : not practical

\+ Upper bound

Signal-to-interference-and-noise metric

- Performance metric:

$$\text{SINR} = \mathbb{E}\left[\frac{\|\boldsymbol{g}^T \boldsymbol{H}_x \boldsymbol{x}\|^2}{\|\boldsymbol{g}^T(\boldsymbol{H}_z \boldsymbol{z} + \boldsymbol{n})\|^2}\right]$$

- Optimal beamformer:

$$\underset{\boldsymbol{g}}{\text{maximize}}\ \text{SINR}$$

– Response very distorted : not practical

\+ Upper bound

Signal-to-interference-and-noise metric

- Performance metric:

$$\text{SINR} = \mathbb{E}\left[\frac{\|\boldsymbol{g}^T \boldsymbol{H}_x \boldsymbol{x}\|^2}{\|\boldsymbol{g}^T(\boldsymbol{H}_z \boldsymbol{z} + \boldsymbol{n})\|^2}\right]$$

- Optimal beamformer:

$$\underset{\boldsymbol{g}}{\text{maximize}}\ \text{SINR}$$

– Response very distorted : not practical

\+ Upper bound

Signal-to-interference-and-noise metric

- Performance metric:

$$\text{SINR} = \mathbb{E}\left[\frac{\|\boldsymbol{g}^T \boldsymbol{H}_x \boldsymbol{x}\|^2}{\|\boldsymbol{g}^T(\boldsymbol{H}_z \boldsymbol{z} + \boldsymbol{n})\|^2}\right]$$

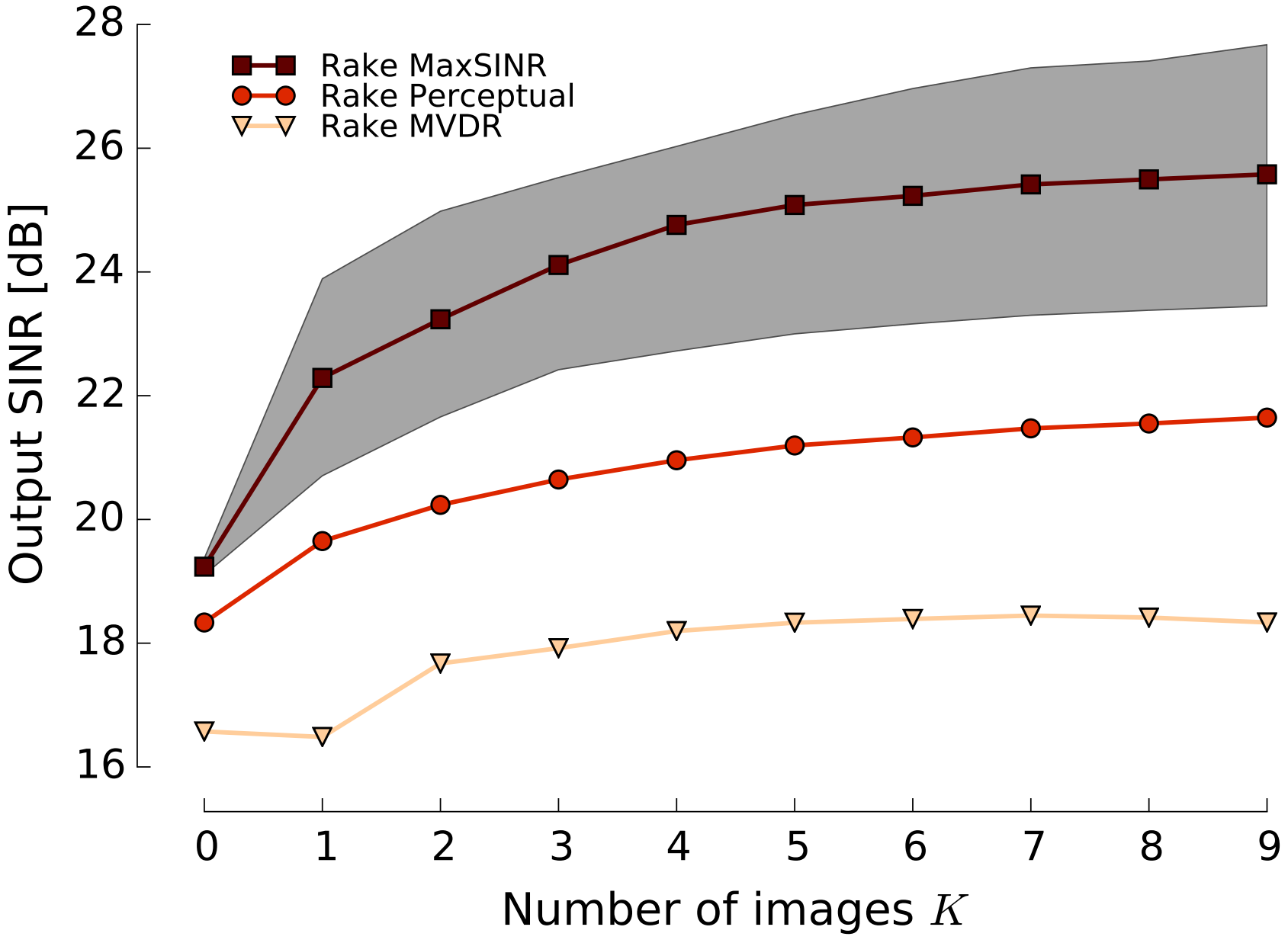
- Optimal beamformer:

$$\underset{\boldsymbol{g}}{\text{maximize}}\ \text{SINR}$$

– Response very distorted : not practical

\+ Upper bound

SINR improvements



Contribution

- A distortionless raking beamformer
- A perceptually motivated raking beamformer
- Time-domain designs allow control on:
  - Delay
  - Pre-echoes
- SINR increases with number of image sources
- Python framework

What's next ?

- Robust formulations
- Experiments

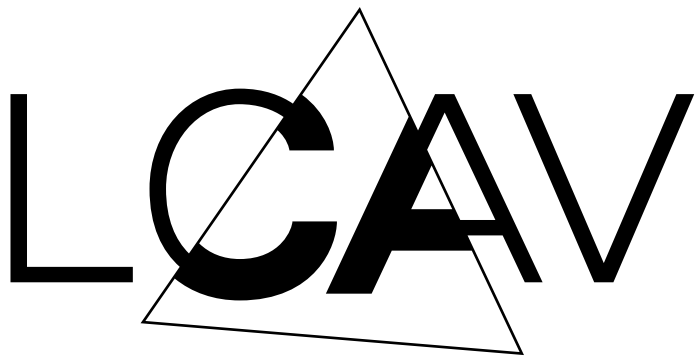
Contribution

- A distortionless raking beamformer
- A perceptually motivated raking beamformer
- Time-domain designs allow control on:
  - Delay
  - Pre-echoes
- SINR increases with number of image sources
- Python framework

What's next ?

- Robust formulations
- Experiments

Thanks for your attention!



Code and figures available at

`http://lcav.github.io/TimeDomainAcousticRakeReceiver/`