SAGECal and the reduction of LOFAR data

Sarod Yatawatta and Sanaz Kazemi

ASTRON and Kapteyn Institute

The Netherlands.

Calibration

For K discrete sources, we observe

$$\mathbf{y} = \sum_{i=1}^{K} \mathbf{s}_i(\boldsymbol{\theta}) + \mathbf{n}, \quad \mathbf{n} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Pi})$$

Maximum Likelihood (ML) estimate, under White Gaussian Noise

$$\widehat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{\arg\min}\ \phi(\boldsymbol{\theta}) = \underset{\boldsymbol{\theta}}{\arg\min}\ \|\mathbf{y} - \sum_{i=1}^{K} \mathbf{s}_i(\boldsymbol{\theta})\|^2$$

Traditional calibration: using Levenberg-Marquardt (LM) algorithm

$$\boldsymbol{\theta}^{k+1} = \boldsymbol{\theta}^{k} - (\boldsymbol{\nabla}_{\boldsymbol{\theta}} \boldsymbol{\nabla}^{T}_{\boldsymbol{\theta}} \phi(\boldsymbol{\theta}) + \lambda \mathbf{H})^{-1} \boldsymbol{\nabla}_{\boldsymbol{\theta}} \phi(\boldsymbol{\theta})|_{\boldsymbol{\theta}^{k}}$$

where $\mathbf{H} \overset{\triangle}{=} \mathrm{diag}(\boldsymbol{\nabla}_{\boldsymbol{\theta}} \boldsymbol{\nabla}^{T}_{\boldsymbol{\theta}} \phi(\boldsymbol{\theta}))$.
Much faster methods are available [Kazemi et al., 2012]

university of groningen ASTRON

EM: Formal Description

[Dempster, Laird, Rubin, 77]

$$\mathbf{y} = \sum_{i=1}^{K} \mathbf{s}_i(\boldsymbol{\theta}) + \mathbf{n}$$

☐ ML estimate: $\hat{\boldsymbol{\theta}}_{ML} = \arg\max_{\boldsymbol{\theta}} \log f(\mathbf{y}|\boldsymbol{\theta})$

☐ Auxiliary random variable $\mathbf{x}$: hidden data, $\mathbf{y} = \mathbf{F}(\mathbf{x})$

☐ The *E Step*: compute conditional expectation
$Q(\boldsymbol{\theta}|\boldsymbol{\theta}^k) = E\{\log f(\mathbf{x}|\boldsymbol{\theta})|\mathbf{y}, \boldsymbol{\theta}^k\}$

☐ The *M Step*: Maximize $\boldsymbol{\theta}^{k+1} = \arg\max_{\boldsymbol{\theta}} Q(\boldsymbol{\theta}|\boldsymbol{\theta}^k)$

☐ Can be simplified for exponential family distributions.

☐ Can be even more simplified for Gaussian distributions.

Classic EM

- Auxiliary random variables

$$\tilde{\mathbf{x}}_i = \mathbf{s}_i(\boldsymbol{\theta}_i) + \tilde{\mathbf{n}}_i$$

- Noise (Gaussian)

$$\mathbf{n} = \sum_{i=1}^{K} \tilde{\mathbf{n}}_i, \quad E\{\tilde{\mathbf{n}}_i \tilde{\mathbf{n}}_j^H\} = \beta_i \delta_{ij} \mathbf{\Pi}, \quad \sum_{i=1}^{K} \beta_i = 1$$

- *E Step*: (conditional mean)

$$\widehat{\tilde{\mathbf{x}}}_i = \mathbf{s}_i(\boldsymbol{\theta}_i^k) + \beta_i(\mathbf{y} - \sum_{l=1}^{K} \mathbf{s}_l(\boldsymbol{\theta}_l^k))$$

- *M Step*: (LM iteration)

$$\boldsymbol{\theta}_i^{k+1} = \boldsymbol{\theta}_i^k - (\boldsymbol{\nabla}_{\boldsymbol{\theta}_i} \boldsymbol{\nabla}_{\boldsymbol{\theta}_i}^T \phi_i(\boldsymbol{\theta}_i) + \lambda \mathbf{H}_i)^{-1} \boldsymbol{\nabla}_{\boldsymbol{\theta}_i} \phi_i(\boldsymbol{\theta}_i)|_{\boldsymbol{\theta}_i^k}$$

SAGE

SAGE: Space Alternating Generalized Expectation Maximization [Fessler and Hero, 94] [Kazemi et al., 2011]

☐ Auxiliary random variable (all noise associated)

$$\mathbf{x}^S = \mathbf{s}_i(\boldsymbol{\theta}_i) + \mathbf{n}$$

☐ *E Step*: (conditional mean)

$$\widehat{\mathbf{x}^S} = \mathbf{s}_i(\boldsymbol{\theta}_i^k) + (\mathbf{y} - \sum_{l=1}^{K} \mathbf{s}_l(\boldsymbol{\theta}_l^k)) = \mathbf{y} - \sum_{l=1, l\neq i}^{K} \mathbf{s}_l(\boldsymbol{\theta}_l^k)$$

☐ *M Step*: (LM iteration)

$$\boldsymbol{\theta}_i^{k+1} = \boldsymbol{\theta}_i^k - (\boldsymbol{\nabla}_{\boldsymbol{\theta}_i} \boldsymbol{\nabla}_{\boldsymbol{\theta}_i}^T \phi_i(\boldsymbol{\theta}_i) + \lambda \mathbf{H}_i)^{-1} \boldsymbol{\nabla}_{\boldsymbol{\theta}_i} \phi_i(\boldsymbol{\theta}_i)|_{\boldsymbol{\theta}_i^k}$$

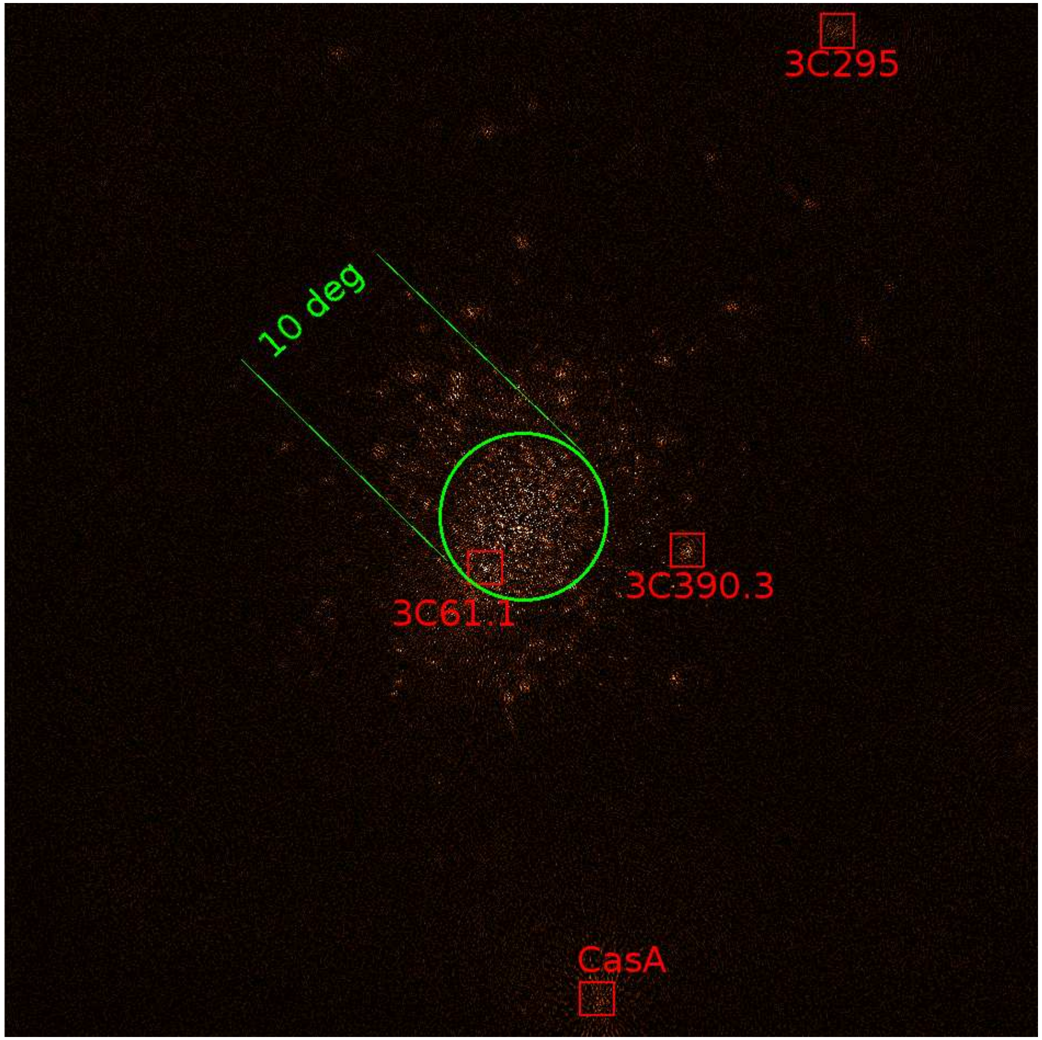
☐ Caveat: $f(\mathbf{y}, \mathbf{x}^S|\boldsymbol{\theta}) = f(\mathbf{y}|\mathbf{x}^S, \boldsymbol{\theta}_{\tilde{S}}) f(\mathbf{x}^S|\boldsymbol{\theta})$

☐ Faster convergence than the classic EM.

SAGECal

- The fastest multisource calibration program ($50\times$ to $100\times$ faster than BBS or meqtrees).
- Complexity: directions $\times$ stations^2.
- Very modest memory usage: (1 million data points, 60 000 parameters, < 6 GB RAM).
- Highly parallelized and vectorized. Uses GPU acceleration when available (> 8 speedup).
- Pure C code with only standard libraries used. Not linked against casacore etc.
- Data I/O done using binary files. Easy conversion to binary format using pyrap.
- Supports all source models: points, Gaussians, disks, rings, (widefield) shapelets (prolate spheroidal wave functions).

LOFAR NCP Window

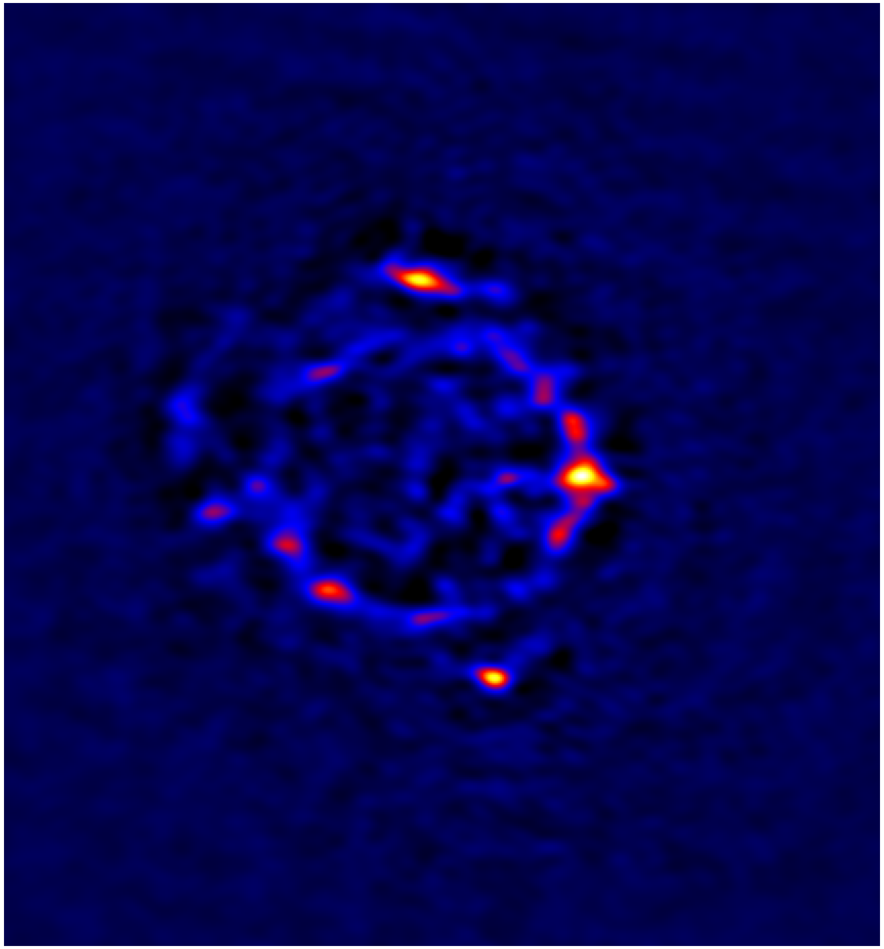


Core baselines < 3 km, 130 MHz, 62×62 sq. deg. image, noise 0.7 mJy

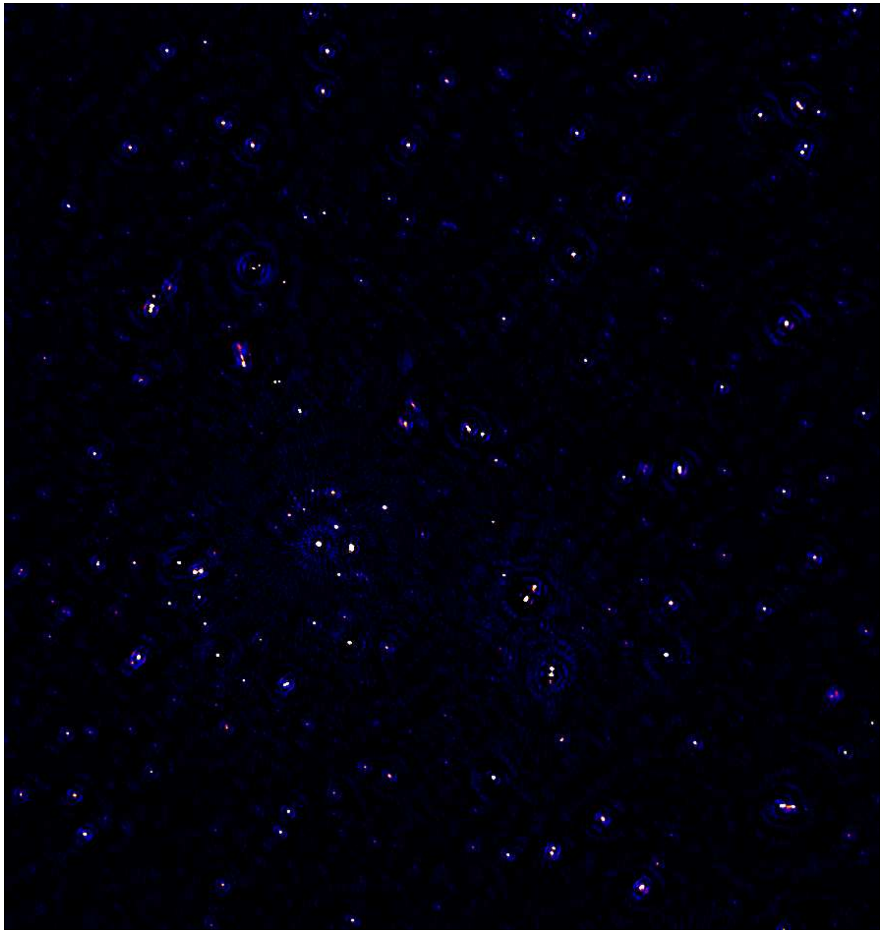
Challenges in LOFAR Calibration

- A Few Complex sources: Use points, shapelets, Prolate Spheroidal Wave Functions,...
- Many more point/double/triple... sources: Careful sky model construction.
- How many directions in the sky to calibrate: Use source clustering to reduce the number of directions. [Kazemi et al., 2011]
- How to calibrate along multiple directions in an accurate and an efficient way: Use SAGECal.
- What is the limit in number of directions? [Kazemi et al., 2012]
- Current noise limits for LOFAR NCP: I $100\ \mu$Jy (3 nights), Polarization $110\ \mu$Jy (1 night).

Complex Sources

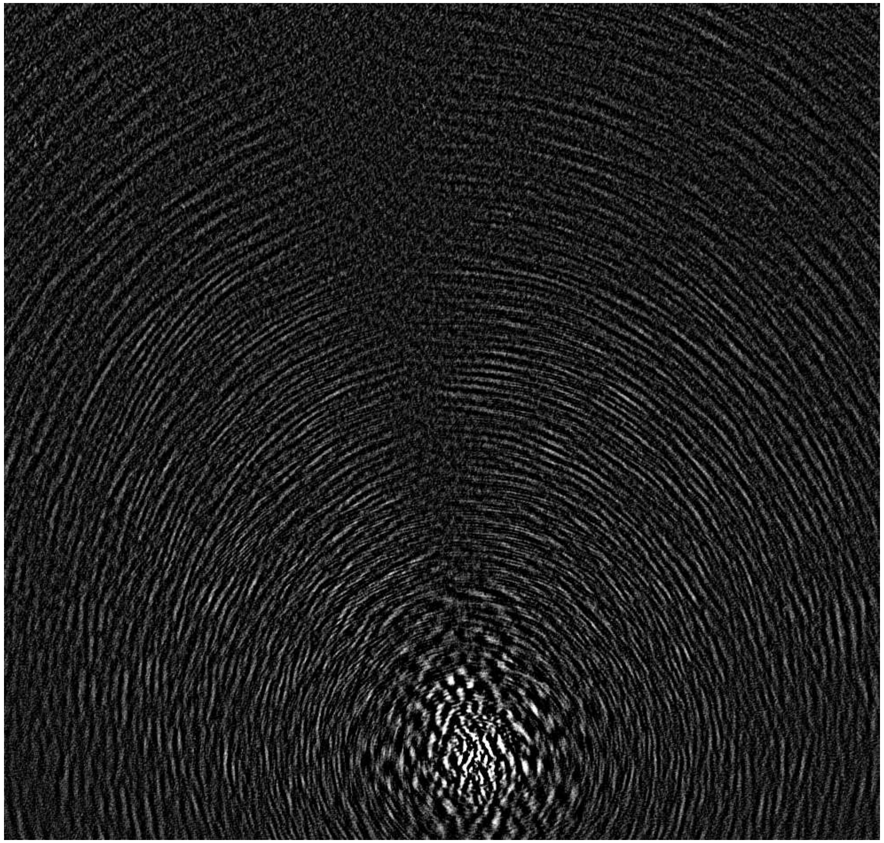


Cassiopeia A, 120 MHz

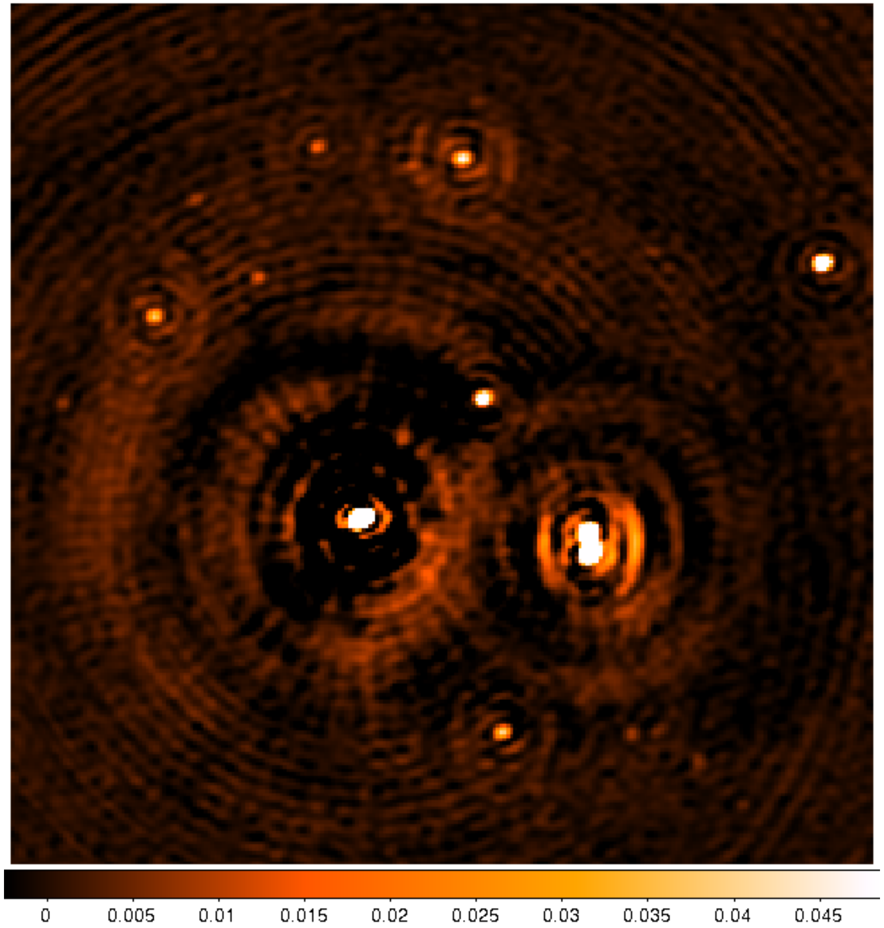


NCP, 130 MHz, 1×1 sq. deg.

Before SAGECal

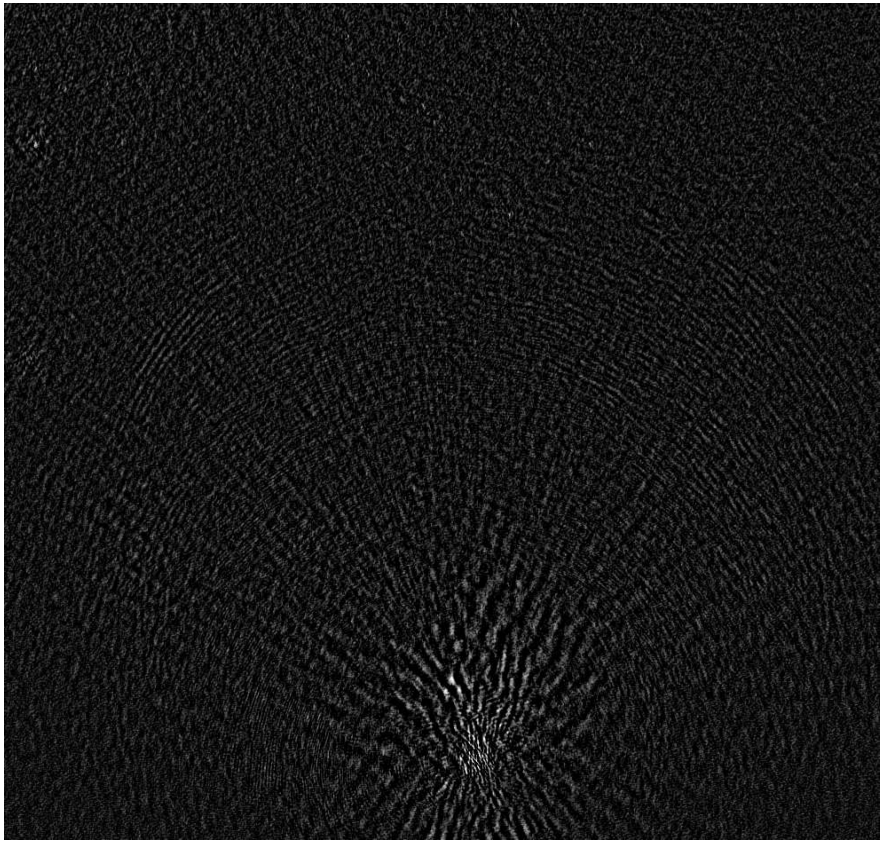


Cas A, 30 deg. away

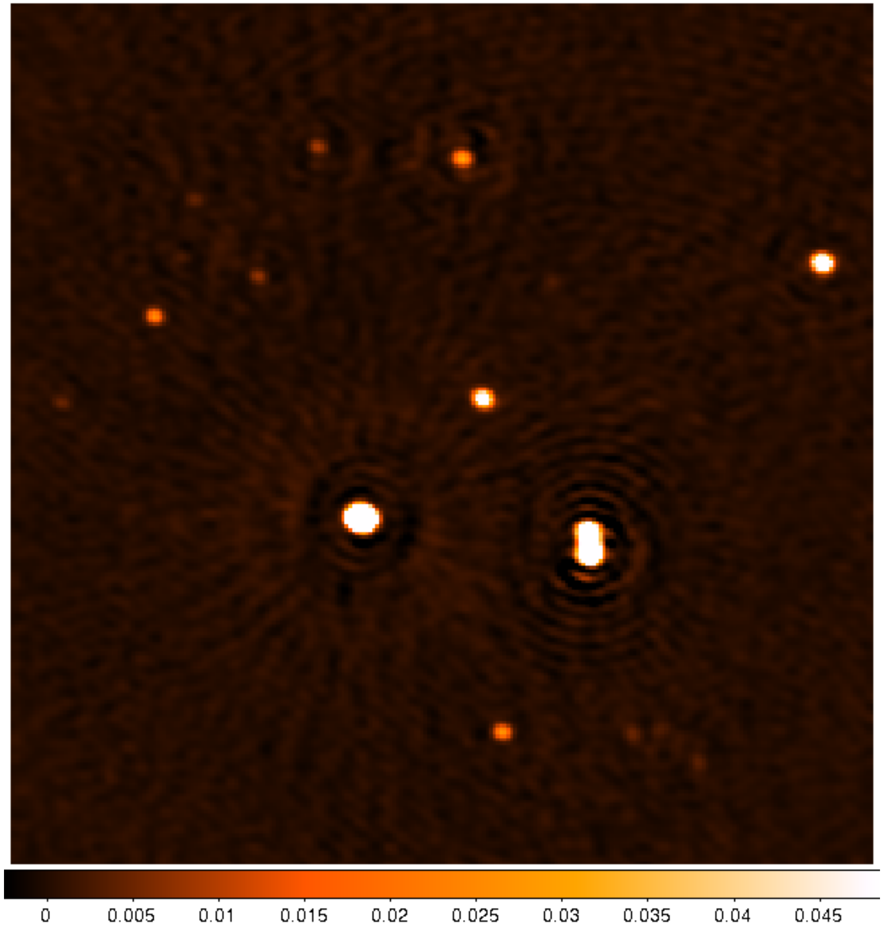


Center

After SAGECal

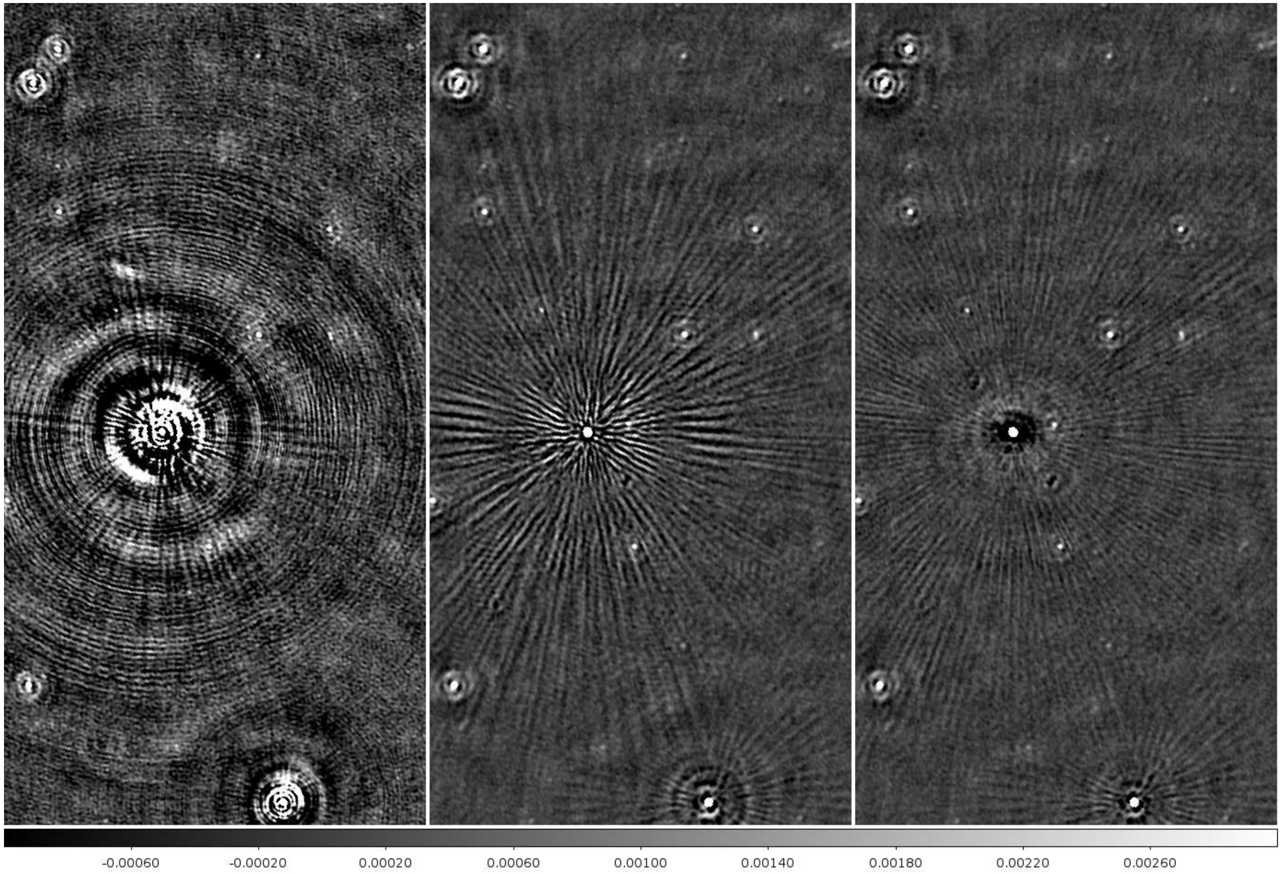


Cas A, 30 deg. away



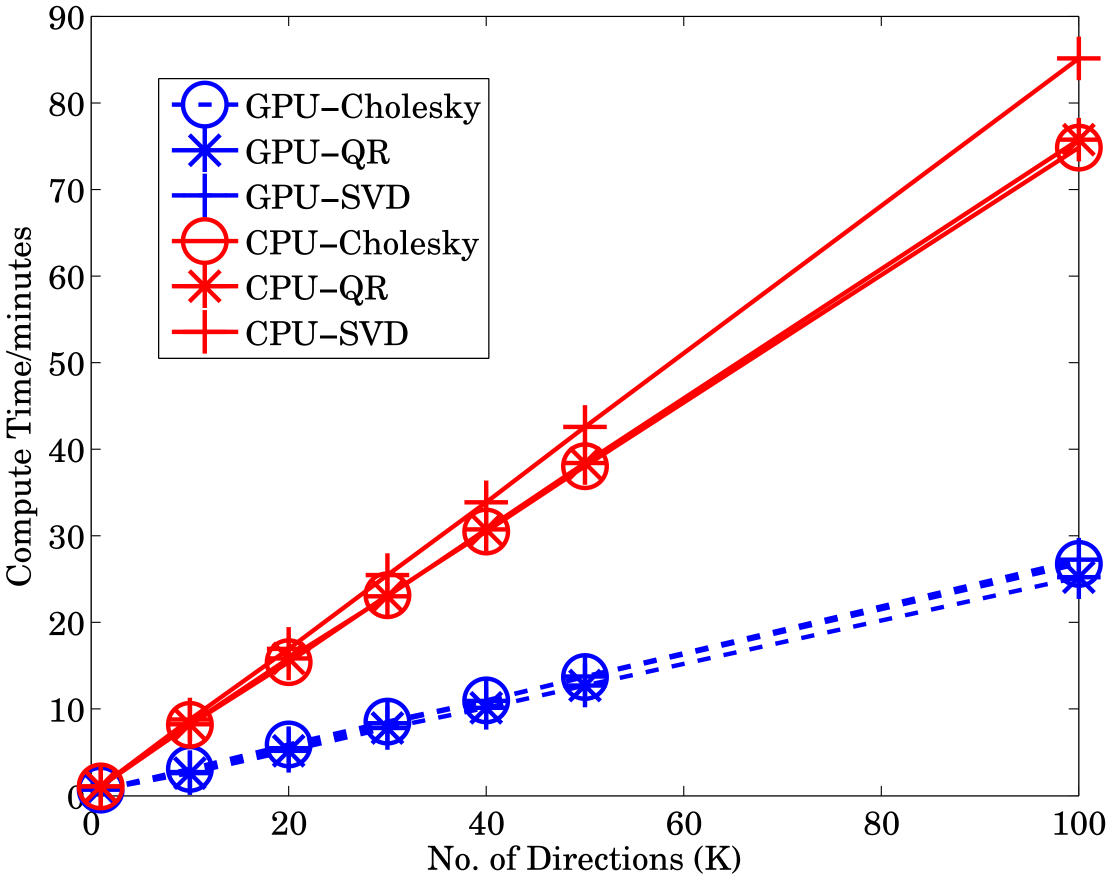
Center

Ionosphere/Beam



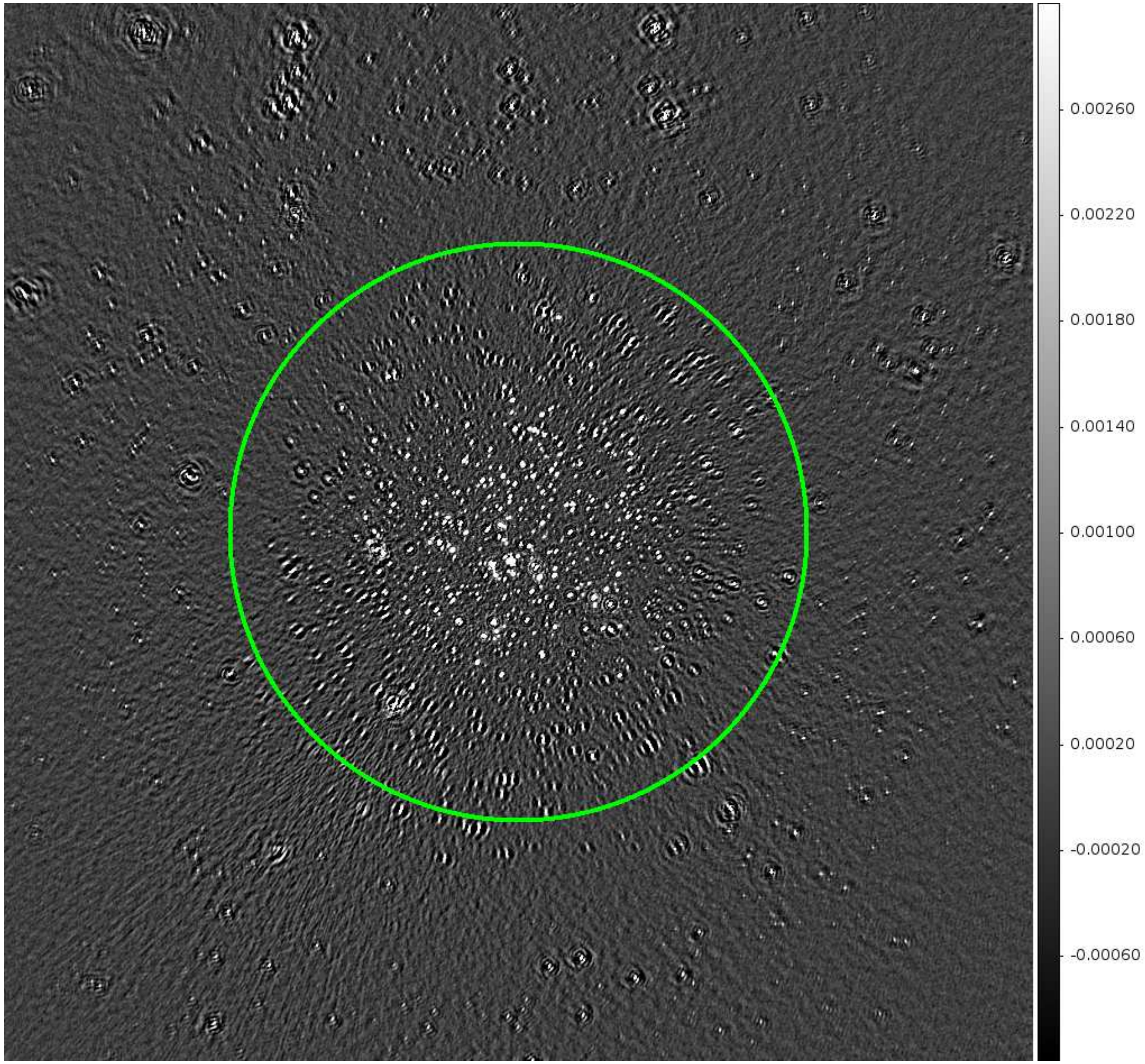
(left) original, (middle) normal (right) hybrid SAGECal

GPU Acceleration



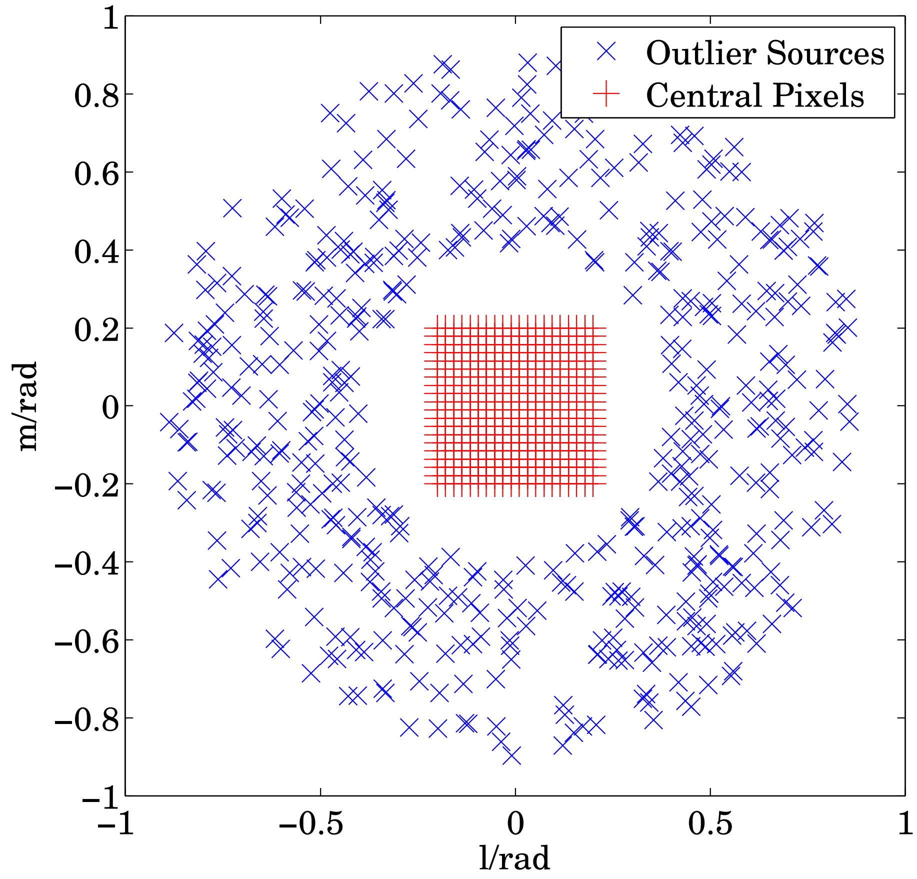
Compute time with no of directions (Tesla M1060)

Outlier Sources



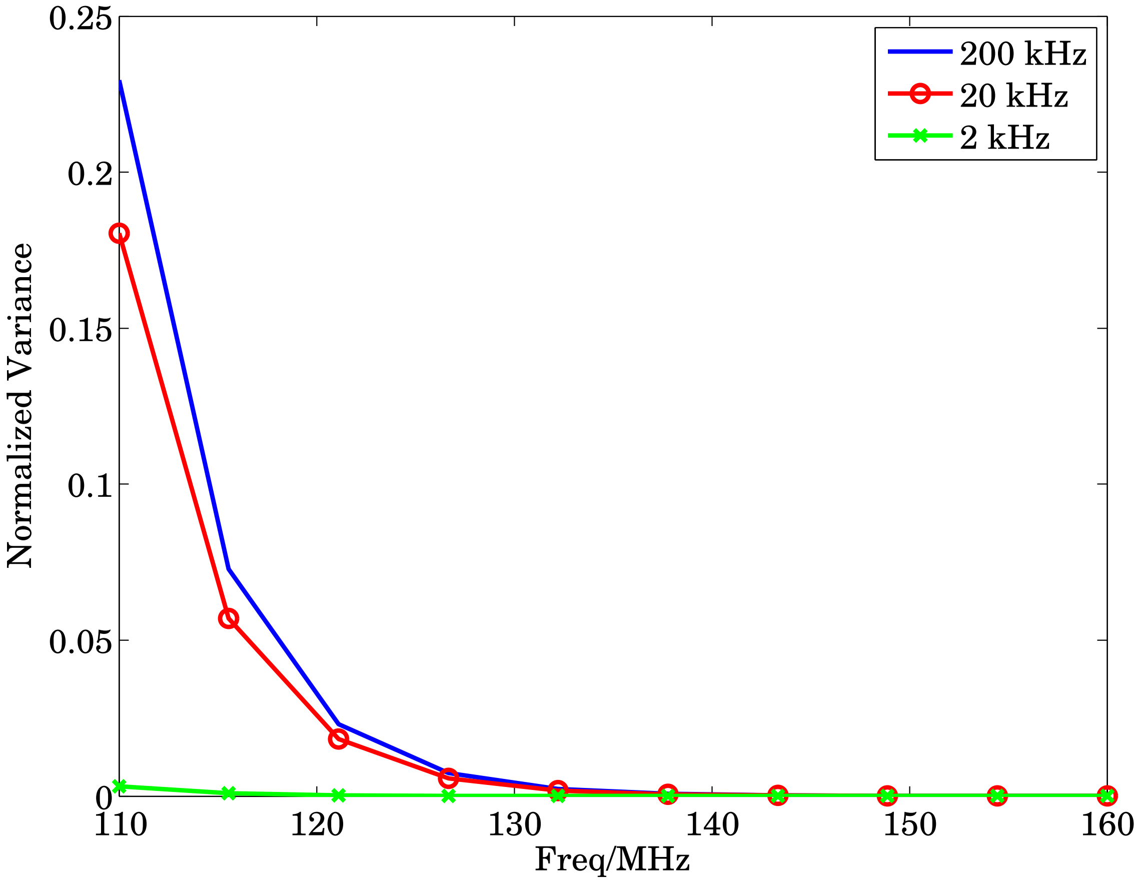
outlier sources outside the 10deg FOV

Excess Noise



outlier source positions X (most of them below noise) and pixels +

Excess Noise



Need to subtract/supress all outlier sources to reduce excess noise.
The wider the beam⇒ the narrower freq. resolution

Beam Estimation

[IEEE SAM 2012]
Ideally

$$\mathbf{C}_{pqm}\gamma_{pm}\gamma_{qm}^{\star} = \mathbf{J}_{pm}\widetilde{\mathbf{C}}_{pqm}\mathbf{J}_{qm}^{H}$$

where $\gamma_{pm} = \mathbf{e}_p^T\mathbf{B}\mathbf{b}_m$ gives the beam model.

Sky coherency (intrinsic) $\mathbf{C}_{pqm}$ (model) $\widetilde{\mathbf{C}}_{pqm}$ ($\in \mathbb{C}^{2\times 2}$).

Calibration solutions are $\mathbf{J}_{pm}$,$\mathbf{J}_{qm}$ ($\in \mathbb{C}^{2\times 2}$) and have a unitary ambiguity.

Beam model (unknown) is $\mathbf{B}$ ($\in \mathbb{C}^{N\times D}$).

Minimize the cost function

$$f(\mathbf{B}) = \sum_{p,q,m} \|\mathbf{C}_{pqm}\gamma_{pm}\gamma_{qm}^{\star} - \mathbf{J}_{pm}\widetilde{\mathbf{C}}_{pqm}\mathbf{J}_{qm}^{H}\|^2$$

to estimate $\mathbf{B}$. Ill conditioned.
Enforce power constraint

$$trace(\mathbf{B}^H\mathbf{B}) = \alpha$$

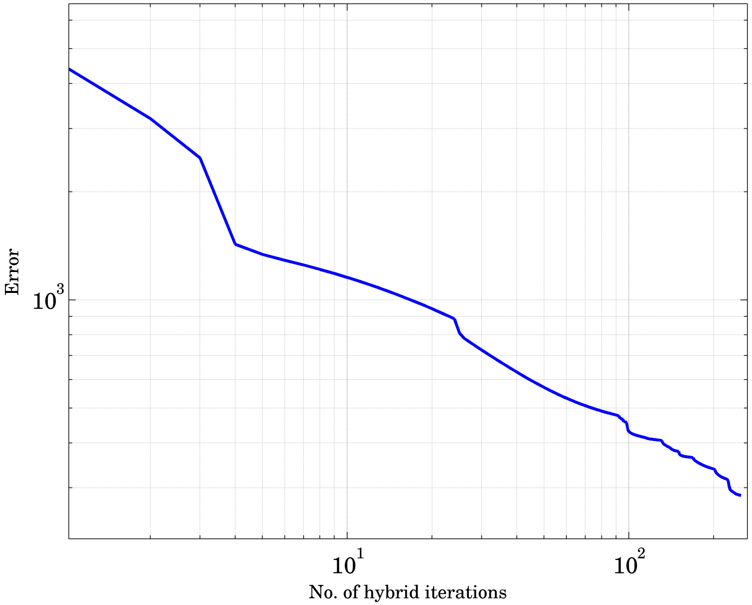
which makes $\mathbf{B}$ restricted to a manifold.

Riemannian Optimization

We use two algorithms

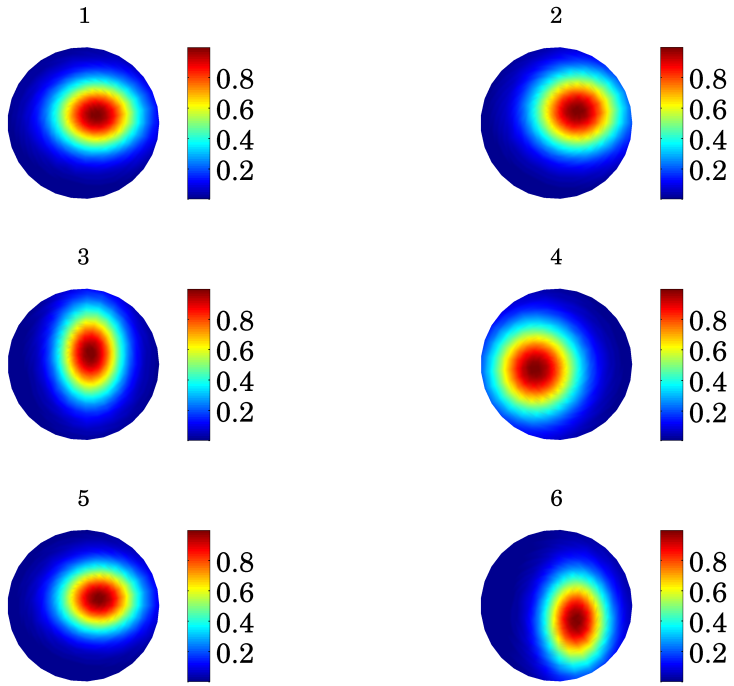
- ☐ Riemannian Steepest Descent [Fiori S. (2011)], on the manifold $trace(\mathbf{B}^H\mathbf{B}) = \alpha$.
- ☐ Riemannian Broyden Fletcher Goldfarb Shanno [Qi C., Gallivan K. and Absil P.A., (2010)] on the $2ND$ unit sphere (Stiefel).

Hybrid use of RSD and RBFGS gives faster convergence. The only requirements are the cost function $f(\mathbf{B})$ and its gradient $\frac{\partial f}{\partial \mathbf{B}}$.

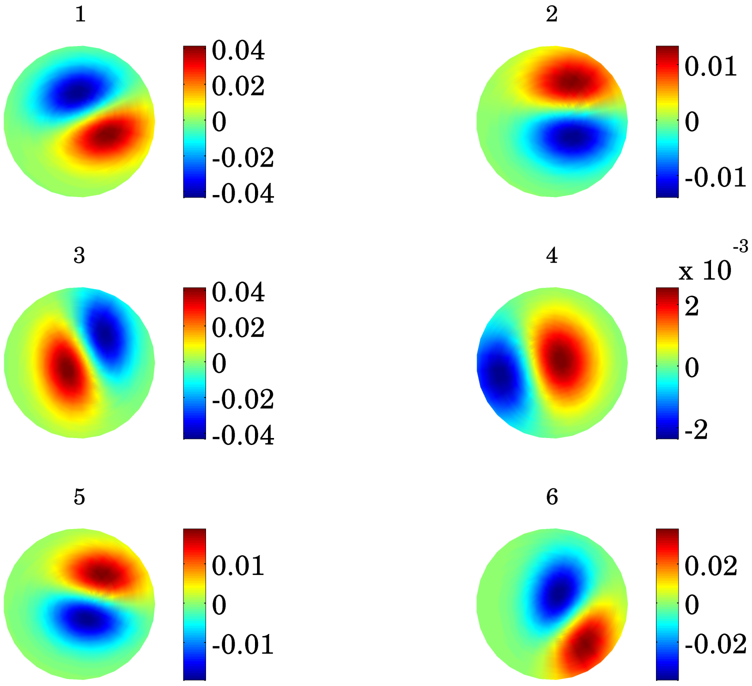


convergence

Beam Model

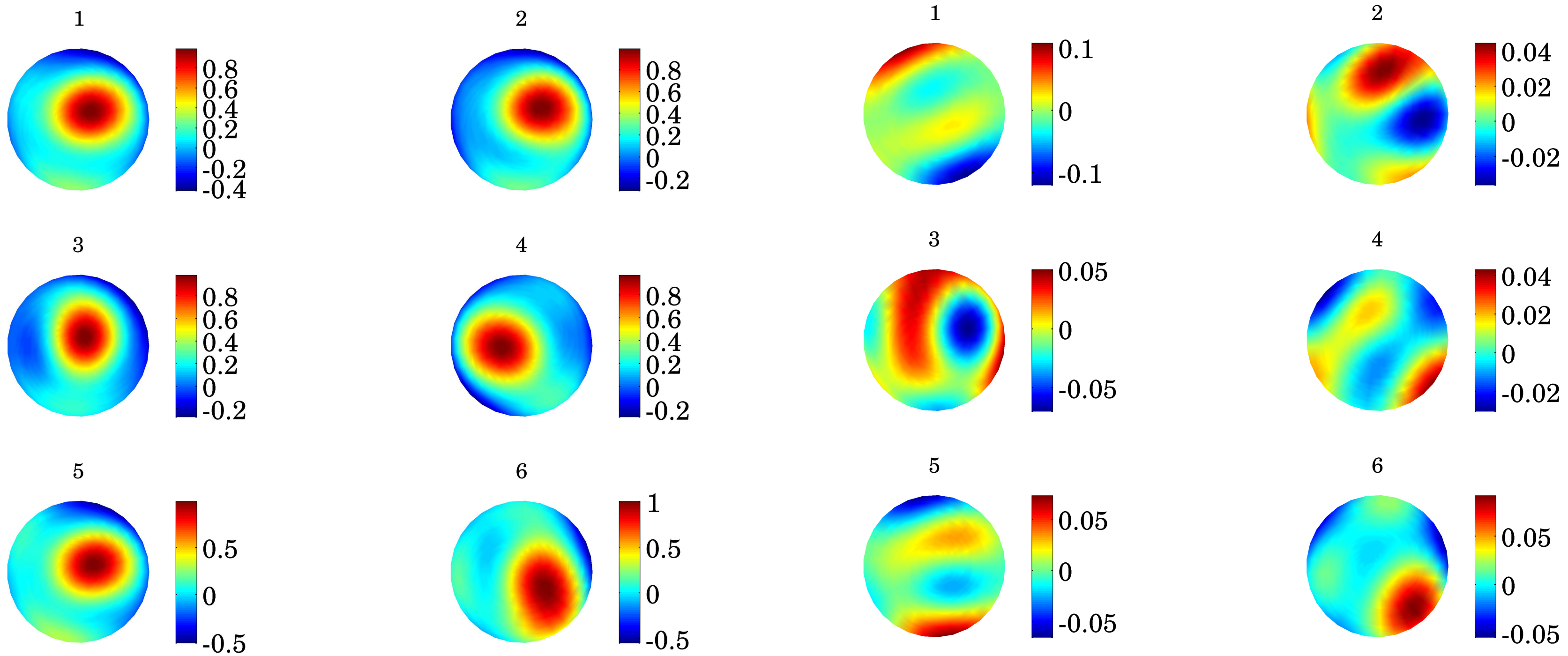


real



imaginary

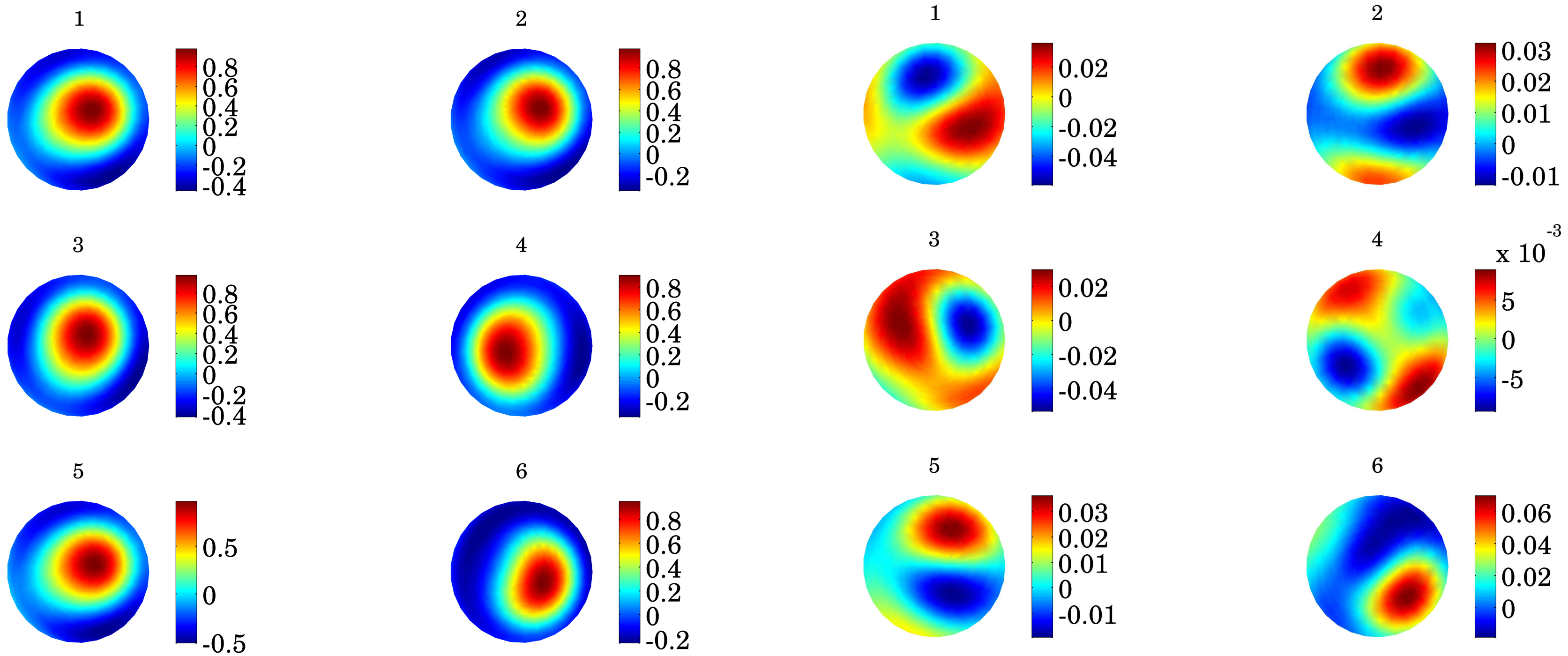
Unconstrained Estimate



real

imaginary

Constrained Estimate



real

imaginary

Conclusions

- LOFAR calibration to reach the noise limit requires subtraction of several thousand sources along several hundred directions.
- SAGECal does this fast and accurately.
- Many more astronomers are using SAGECal to process LOFAR HBA/LBA data (and get good results).
- Current LOFAR limits ≈ 1 million in dynamic range [Labropoulus] and $100\ \mu$Jy in I and polarization.
- Sources outside the FOV play a role almost as important as sources inside the FOV in reaching the noise limit (for any interferometer).
- SKA designers need to keep this in mind.