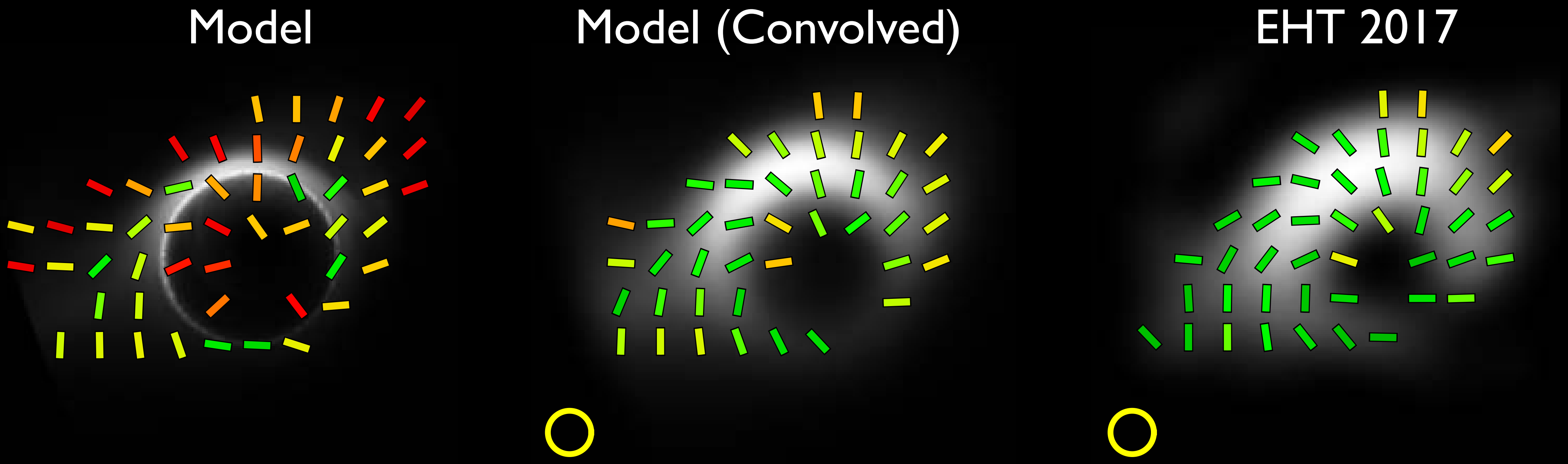
Imaging Supermassive Black Holes with the Event Horizon Telescope



Kazu Akiyama

(MIT Haystack Observatory / JSPS Fellow)

On Behalf of the EHT Imaging Working Group:

Michael Johnson, Andrew Chael, Ramesh Narayan (CfA/SAO),
Katie Bouman (MIT CSAIL), Mareki Honma (NAOJ) et al.

Radio Interferometry: Sampling Fourier Components of the Images

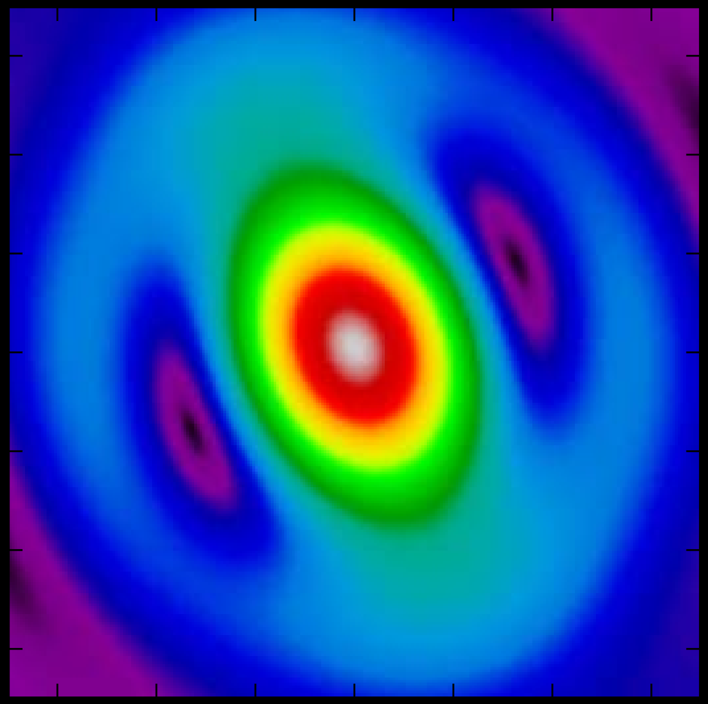
Radio Interferometry: Sampling Fourier Components of the Images

Image

Radio Interferometry:
Sampling Fourier Components of the Images

Image

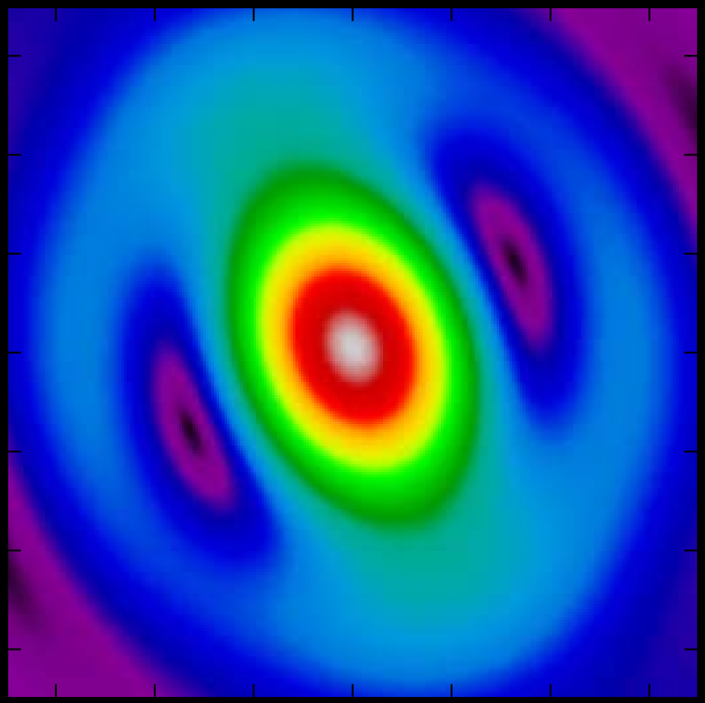
Fourier Domain
(*Visibility*)



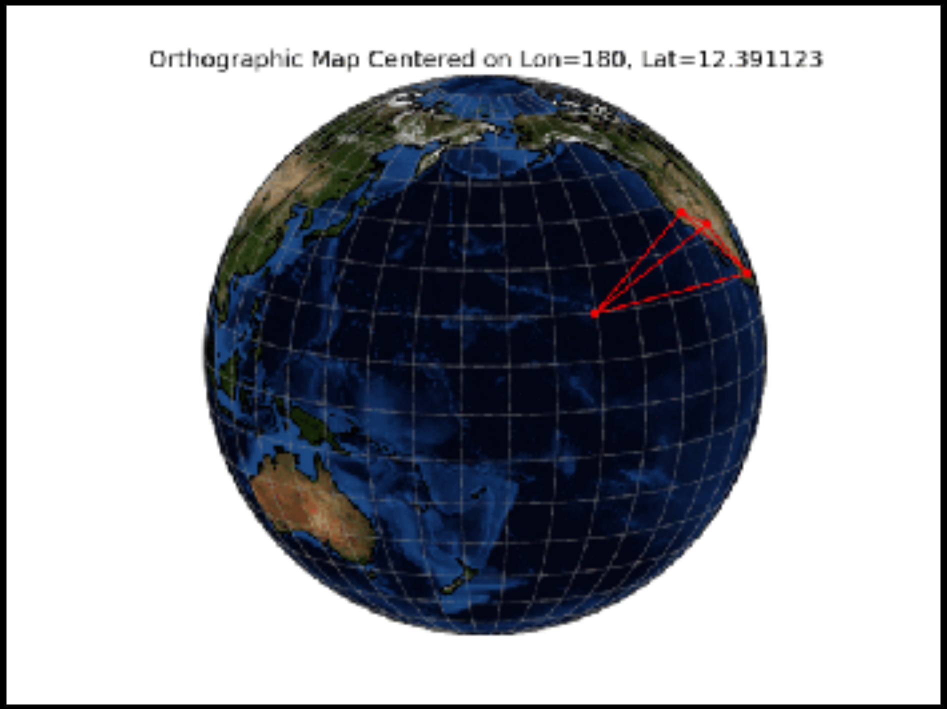
Radio Interferometry: Sampling Fourier Components of the Images

Image

Fourier Domain
(Visibility)



Sampling Process
(Projected Baseline = Spatial Frequenc

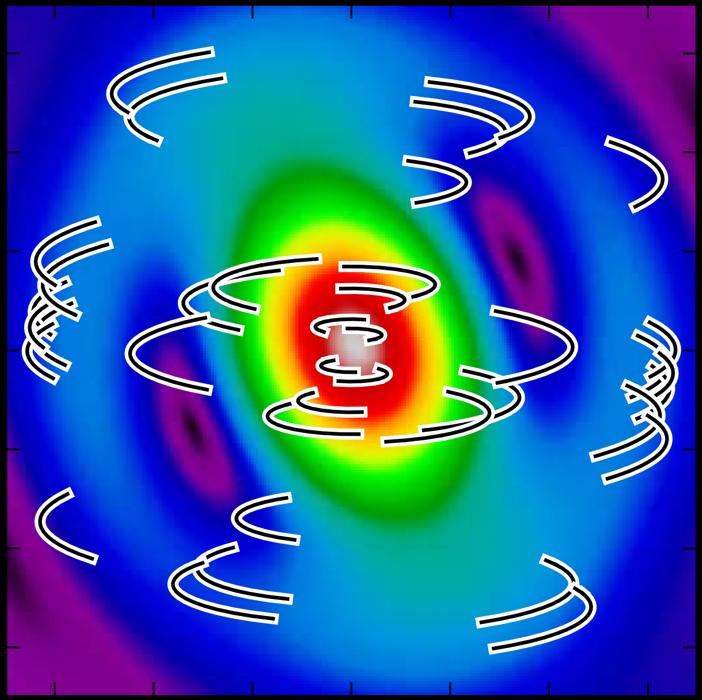


(Images: adapted from *Akiyama* et al. 2015, ApJ ; Movie: Laura Vertatschitsch)

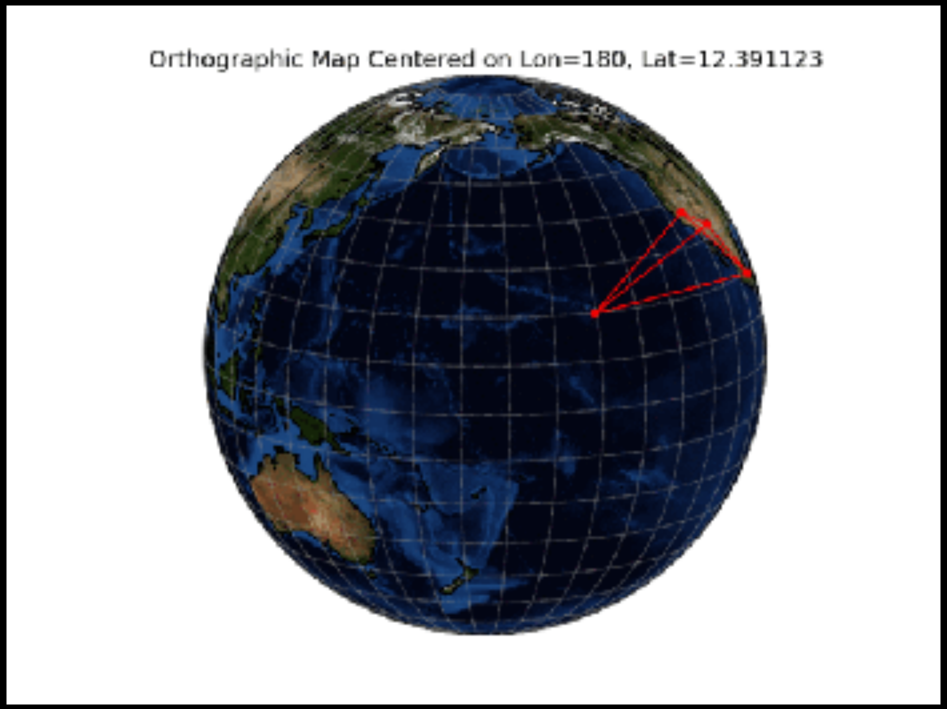
Radio Interferometry: Sampling Fourier Components of the Images

Image

Fourier Domain
(*Visibility*)



Sampling Process
(Projected Baseline = Spatial Frequenc

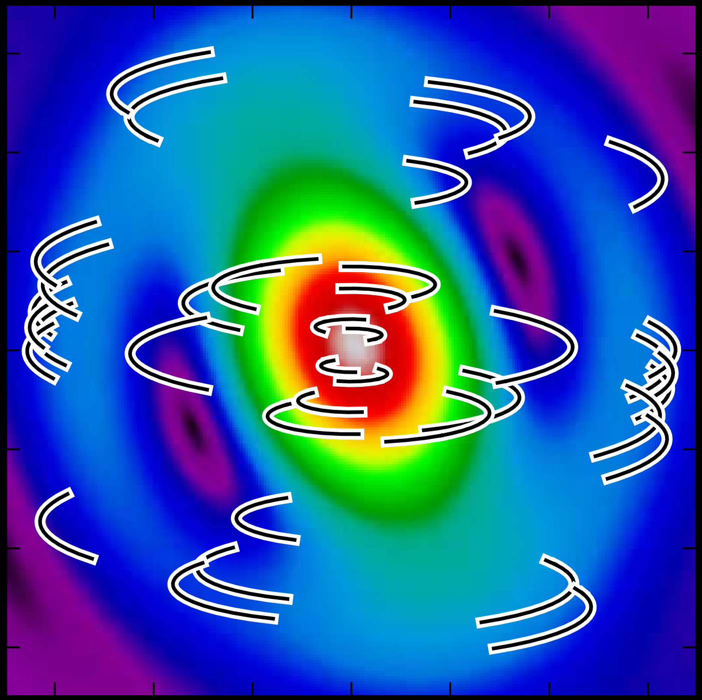


(Images: adapted from *Akiyama* et al. 2015, ApJ ; Movie: Laura Vertatschitsch)

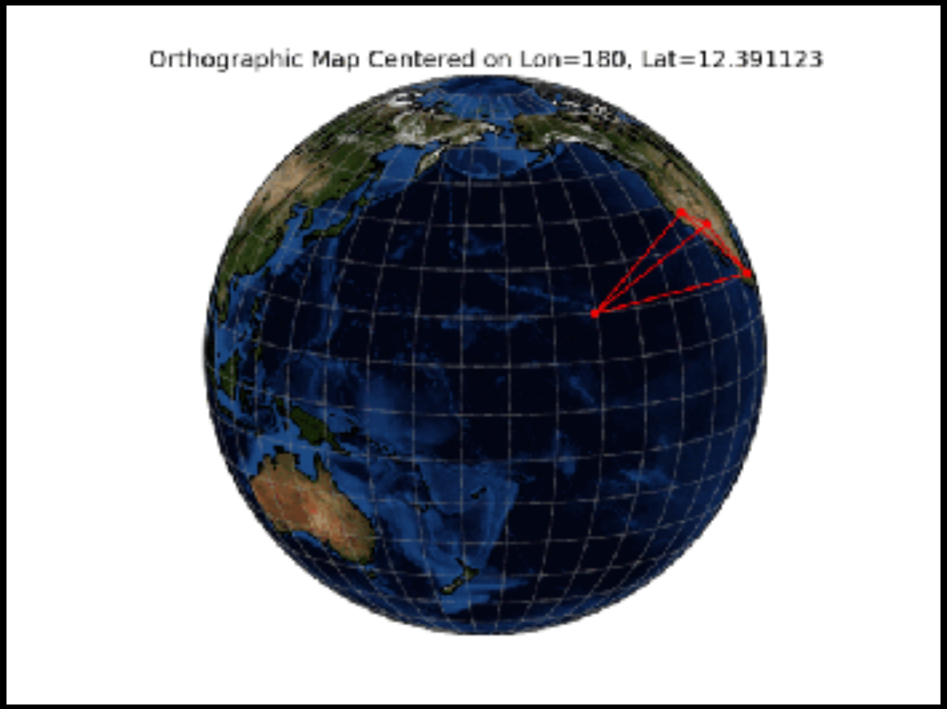
Radio Interferometry: Sampling Fourier Components of the Images

Image

Fourier Domain
(Visibility)



Sampling Process
(Projected Baseline = Spatial Frequenc



(Images: adapted from *Akiyama* et al. 2015, ApJ ; Movie: Laura Vertatschitsch)

Sampling is NOT perfect

Interferometry Imaging: Observational equation is *ill-posed*

$$\underset{\text{(Data)}}{Y} = \underset{\text{(Fourier Matrix)}}{A} \; \underset{\text{(Image)}}{X}$$

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_M \end{pmatrix} = \begin{pmatrix} \exp(i2\pi u_1 x_1) & \exp(i2\pi u_1 x_2) & \dots & \exp(i2\pi u_1 x_N) \\ \exp(i2\pi u_2 x_1) & \exp(i2\pi u_2 x_2) & \dots & \exp(i2\pi u_2 x_N) \\ \exp(i2\pi u_3 x_1) & \exp(i2\pi u_3 x_2) & \dots & \exp(i2\pi u_3 x_N) \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_N \end{pmatrix}$$

Interferometry Imaging: Observational equation is *ill-posed*

$$\begin{array}{cccc} Y & = & A & X \\ \text{(Data)} & & \text{(Fourier Matrix)} & \text{(Image)} \\ \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_M \end{pmatrix} & = & \begin{pmatrix} \exp(i2\pi u_1 x_1) \;\; \exp(i2\pi u_1 x_2) \;\ldots\; \exp(i2\pi u_1 x_N) \\ \exp(i2\pi u_2 x_1) \;\; \exp(i2\pi u_2 x_2) \;\ldots\; \exp(i2\pi u_2 x_N) \\ \exp(i2\pi u_3 x_1) \;\; \exp(i2\pi u_3 x_2) \;\ldots\; \exp(i2\pi u_3 x_N) \end{pmatrix} & \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_N \end{pmatrix} \end{array}$$

- Sampling is NOT perfect

  Number of data ***M*** < Number of image pixels ***N***

Interferometry Imaging: Observational equation is *ill-posed*

$$\underset{\text{(Data)}}{Y} = \underset{\text{(Fourier Matrix)}}{A} \; \underset{\text{(Image)}}{X}$$

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_M \end{pmatrix} = \begin{pmatrix} \exp(i2\pi u_1 x_1) & \exp(i2\pi u_1 x_2) & \ldots & \exp(i2\pi u_1 x_N) \\ \exp(i2\pi u_2 x_1) & \exp(i2\pi u_2 x_2) & \ldots & \exp(i2\pi u_2 x_N) \\ \exp(i2\pi u_3 x_1) & \exp(i2\pi u_3 x_2) & \ldots & \exp(i2\pi u_3 x_N) \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_N \end{pmatrix}$$

- Sampling is NOT perfect
  Number of data ***M*** < Number of image pixels ***N***
- Equation is *ill-posed*: infinite numbers of solutions

Interferometry Imaging: Observational equation is *ill-posed*

$$\begin{array}{cccc} Y & = & A & X \\ \text{(Data)} & & \text{(Fourier Matrix)} & \text{(Image)} \end{array}$$

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_M \end{pmatrix} = \begin{pmatrix} \exp(i2\pi u_1 x_1) & \exp(i2\pi u_1 x_2) & \dots & \exp(i2\pi u_1 x_N) \\ \exp(i2\pi u_2 x_1) & \exp(i2\pi u_2 x_2) & \dots & \exp(i2\pi u_2 x_N) \\ \exp(i2\pi u_3 x_1) & \exp(i2\pi u_3 x_2) & \dots & \exp(i2\pi u_3 x_N) \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_N \end{pmatrix}$$

- Sampling is NOT perfect
  Number of data ***M*** < Number of image pixels ***N***
- Equation is *ill-posed*: infinite numbers of solutions
- Interferometric Imaging:
  Picking a reasonable solution based on a prior assumption

Approach 1: Sparse Reconstruction

Approach 1: Sparse Reconstruction

Philosophy: **Reconstructing images with the smallest number of point sources within a given residual error**

Approach 1: Sparse Reconstruction

Philosophy: **Reconstructing images with the smallest number of point sources within a given residual error**

$$\min_{\mathbf{x}} ||\mathbf{x}||_0 \ \text{subject to} \ ||\mathbf{y} - \mathbf{A}\mathbf{x}||_2^2 < \varepsilon$$

Approach 1: Sparse Reconstruction

Philosophy: **Reconstructing images with the smallest number of point sources within a given residual error**

$$\min_{\mathbf{x}} ||\mathbf{x}||_0 \ \text{subject to} \ ||\mathbf{y} - \mathbf{A}\mathbf{x}||_2^2 < \varepsilon$$

L_p-norm:

$$||\mathbf{x}||_p = \left(\sum_i |x_i|^p\right)^{\frac{1}{p}} \quad (p>0)$$

$||x||_0 =$ number of non-zero pixels in the image

Approach 1: Sparse Reconstruction

Philosophy: Reconstructing images with the smallest number of point sources within a given residual error

$$\min_{\mathbf{x}} ||\mathbf{x}||_0 \ \text{subject to} \ ||\mathbf{y} - \mathbf{A}\mathbf{x}||_2^2 < \varepsilon$$

Number of non-zero pixels (point sources)

L$_p$-norm:

$$||\mathbf{x}||_p = \left(\sum_i |x_i|^p\right)^{\frac{1}{p}} \quad (p>0)$$

$||x||_0$ = number of non-zero pixels in the image

Approach 1: Sparse Reconstruction

Philosophy: **Reconstructing images with the smallest number of point sources within a given residual error**

$$\min_{\mathbf{x}} ||\mathbf{x}||_0 \text{ subject to } ||\mathbf{y} - \mathbf{A}\mathbf{x}||_2^2 < \varepsilon$$

Number of non-zero pixels (point sources)

Data

Obs. Matrix

Image

Chi-square: Consistency between data and the image

L_p-norm:

$$||\mathbf{x}||_p = \left(\sum_i |x_i|^p\right)^{\frac{1}{p}} \quad (p>0)$$

$||x||_0$ = number of non-zero pixels in the image

Approach 1: Sparse Reconstruction

Philosophy: **Reconstructing images with the smallest number of point sources within a given residual error**

$$\min_{\mathbf{x}} ||\mathbf{x}||_0 \text{ subject to } ||\mathbf{y} - \mathbf{A}\mathbf{x}||_2^2 < \varepsilon$$

Number of non-zero pixels

Data

Obs. Matrix

Image

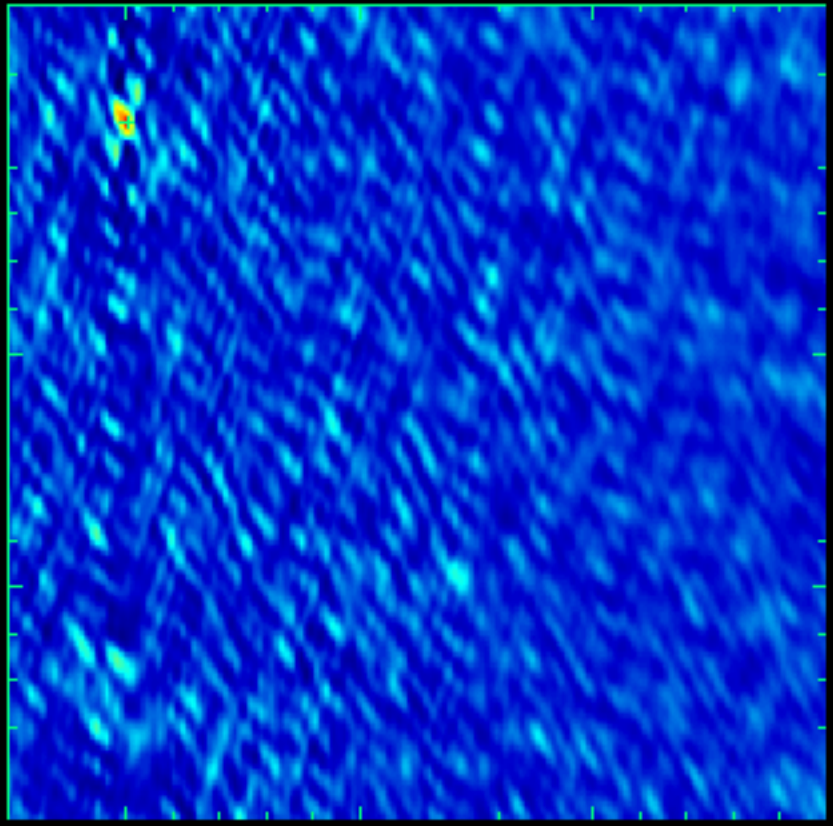
Computationally very expensive!!
(It can be solved for N < ~100)

- L_0 norm is not continuous, nondifferentiable
- Combinational Optimization

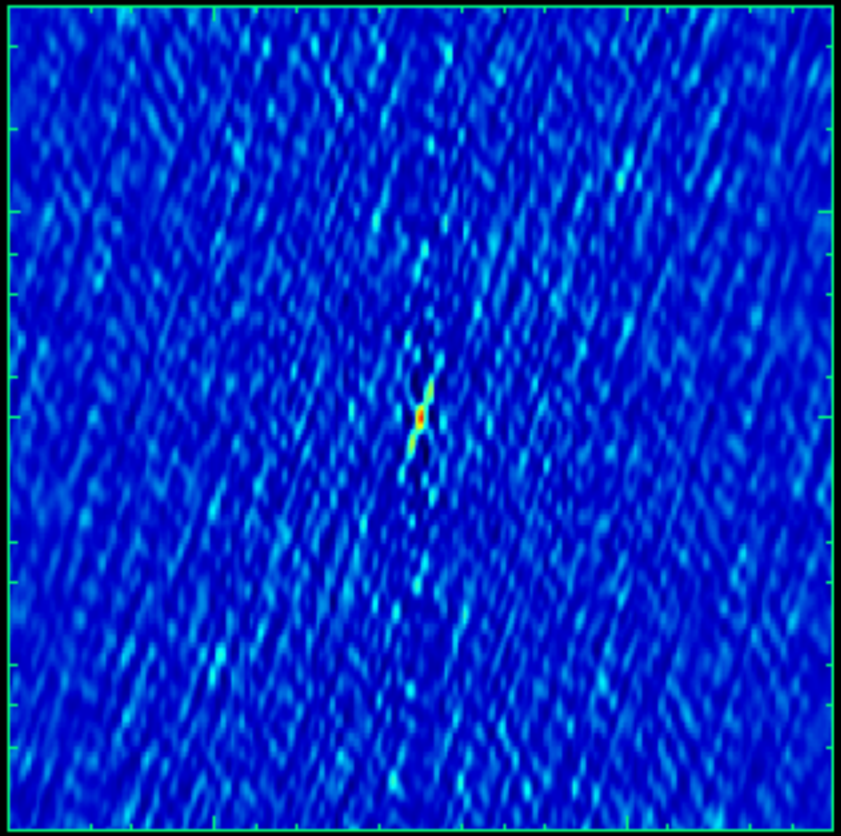
Approach 1: Sparse Reconstruction

CLEAN (Hobgom 1974) = Matching Pursuit (Mallet & Zhang 1993)

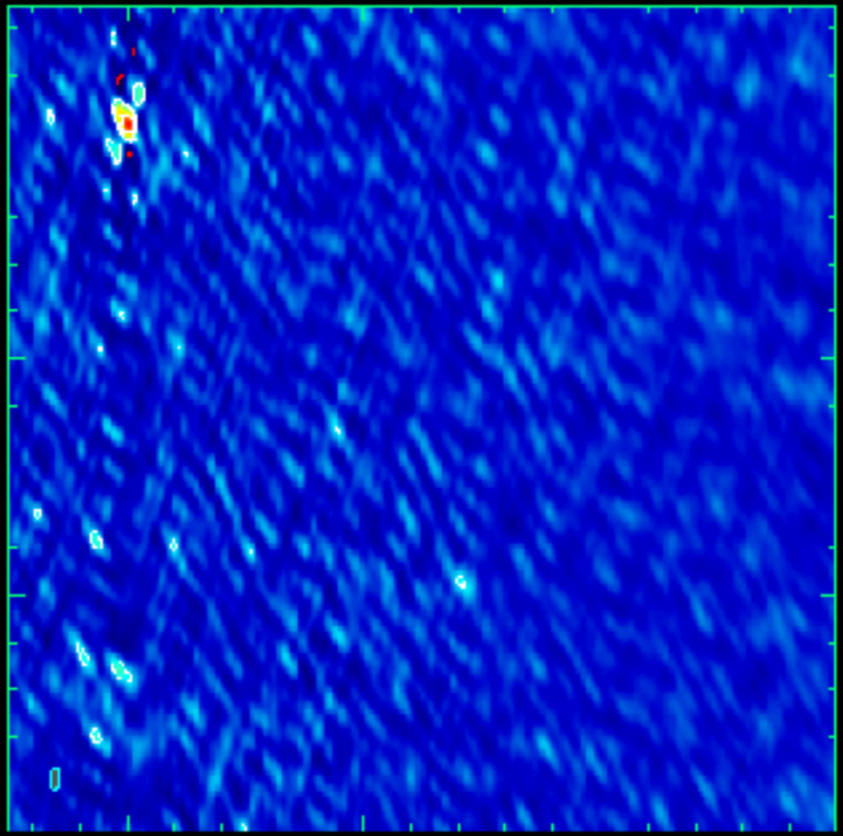
Computationally very cheap, but highly affected by the Point Spread Function



Dirty map:
FT of *zero-filled*
Visibility



Point Spread Function:
Dirty map
for the point source



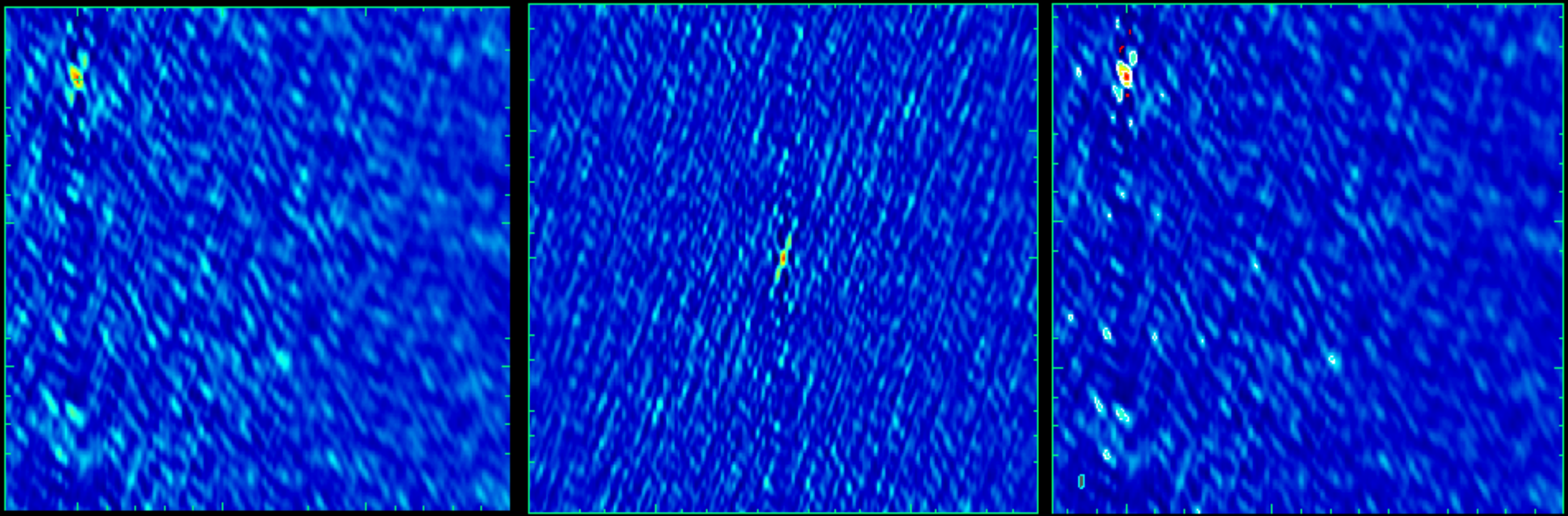
Solution:
Point sources
+ Residual Map

(3C 273, VLBA-MOJAVE data at 15 GHz)

Approach 1: Sparse Reconstruction

CLEAN (Hobgom 1974) = Matching Pursuit (Mallet & Zhang 1993)

Computationally very cheap, but highly affected by the Point Spread Function



Dirty map:
FT of *zero-filled* Visibility

Point Spread Function:
Dirty map
for the point source

Solution:
Point sources
+ Residual Map

(3C 273, VLBA-MOJAVE data at 15 GHz)

Approach 1: Sparse Reconstruction

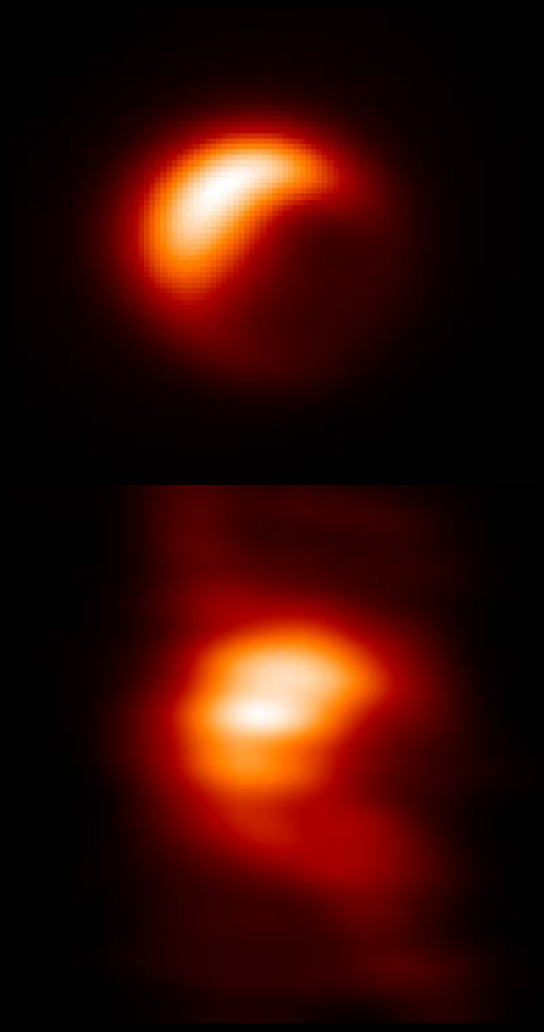
CLEAN (Hobgom 1974) = Matching Pursuit (Mallet & Zhang 1993)

Computationally very cheap, but highly affected by the Point Spread Function

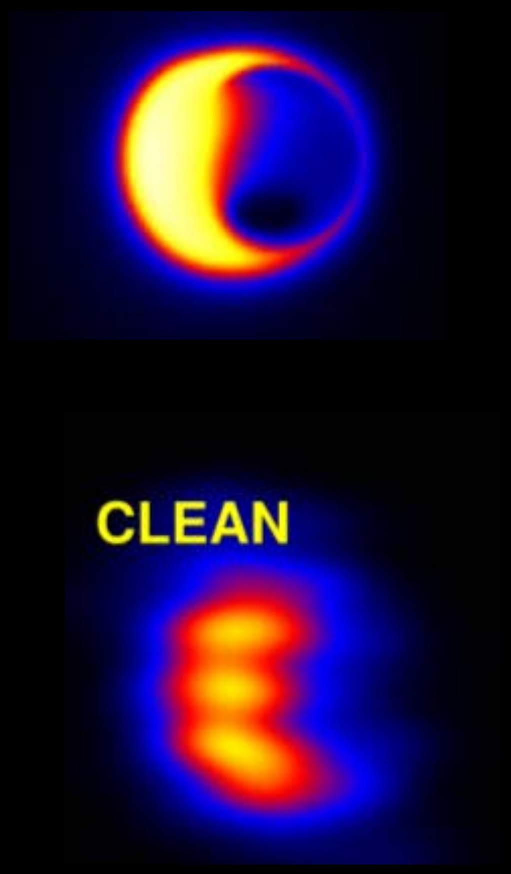
CLEAN is problematic for the black hole shadows?

Ground Truth

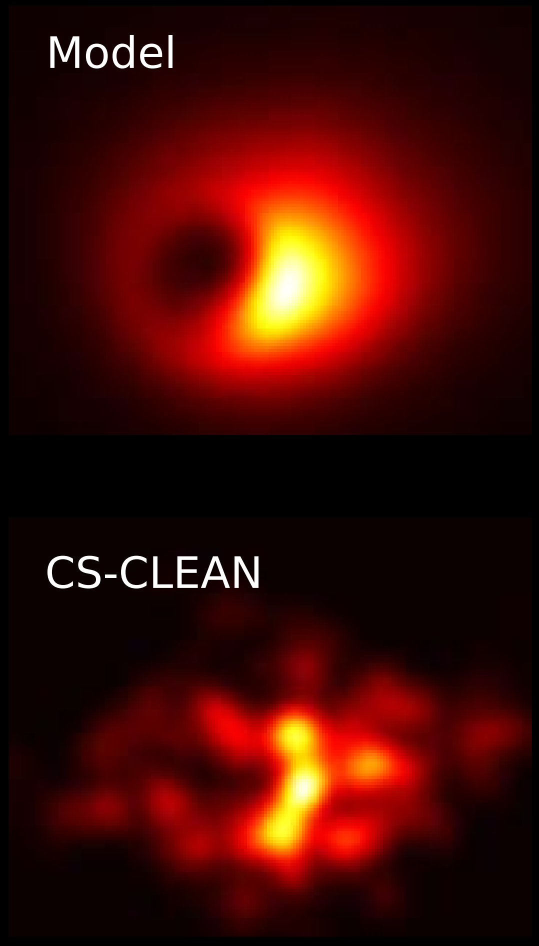
CLEAN



Fabian Baron+



Chael+2016 ApJ



Akiyama+2016b,c, submitted to ApJ

Approach 1: Sparse Reconstruction

L1 regularization (LASSO, Tibishirani 1996)

Approach 1: Sparse Reconstruction

L1 regularization (LASSO, Tibishirani 1996)

***Convex* Relaxation**: Relaxing L0-norm to a convex, continuous, and differentiable function

Approach 1: Sparse Reconstruction

L1 regularization (LASSO, Tibishirani 1996)

Convex Relaxation: Relaxing L0-norm to a convex, continuous, and differentiable function

$$\min_{\mathbf{x}} ||\mathbf{x}||_1 \quad \text{subject to } ||\mathbf{y} - \mathbf{A}\mathbf{x}||_2^2 < \varepsilon$$

Approach 1: Sparse Reconstruction

L1 regularization (LASSO, Tibishirani 1996)

Convex Relaxation: Relaxing L0-norm to a convex, continuous, and differentiable function

$$\min_{\mathbf{x}} ||\mathbf{x}||_1 \quad \text{subject to } ||\mathbf{y} - \mathbf{A}\mathbf{x}||_2^2 < \varepsilon$$

⟷

equivalent

$$\min_{\mathbf{x}} \left(\underbrace{||\mathbf{y} - \mathbf{A}\mathbf{x}||_2^2}_{\text{Chi-square}} + \underbrace{\Lambda_l ||\mathbf{x}||_1}_{\text{Regularization on } \textit{sparsity}} \right).$$

Approach 1: Sparse Reconstruction

L1 regularization (LASSO, Tibishirani 1996)

Convex Relaxation: Relaxing L0-norm to a convex, continuous, and differentiable function

$$\min_{\mathbf{x}} ||\mathbf{x}||_1 \quad \text{subject to} \quad ||\mathbf{y} - \mathbf{A}\mathbf{x}||_2^2 < \varepsilon$$

⟷ equivalent

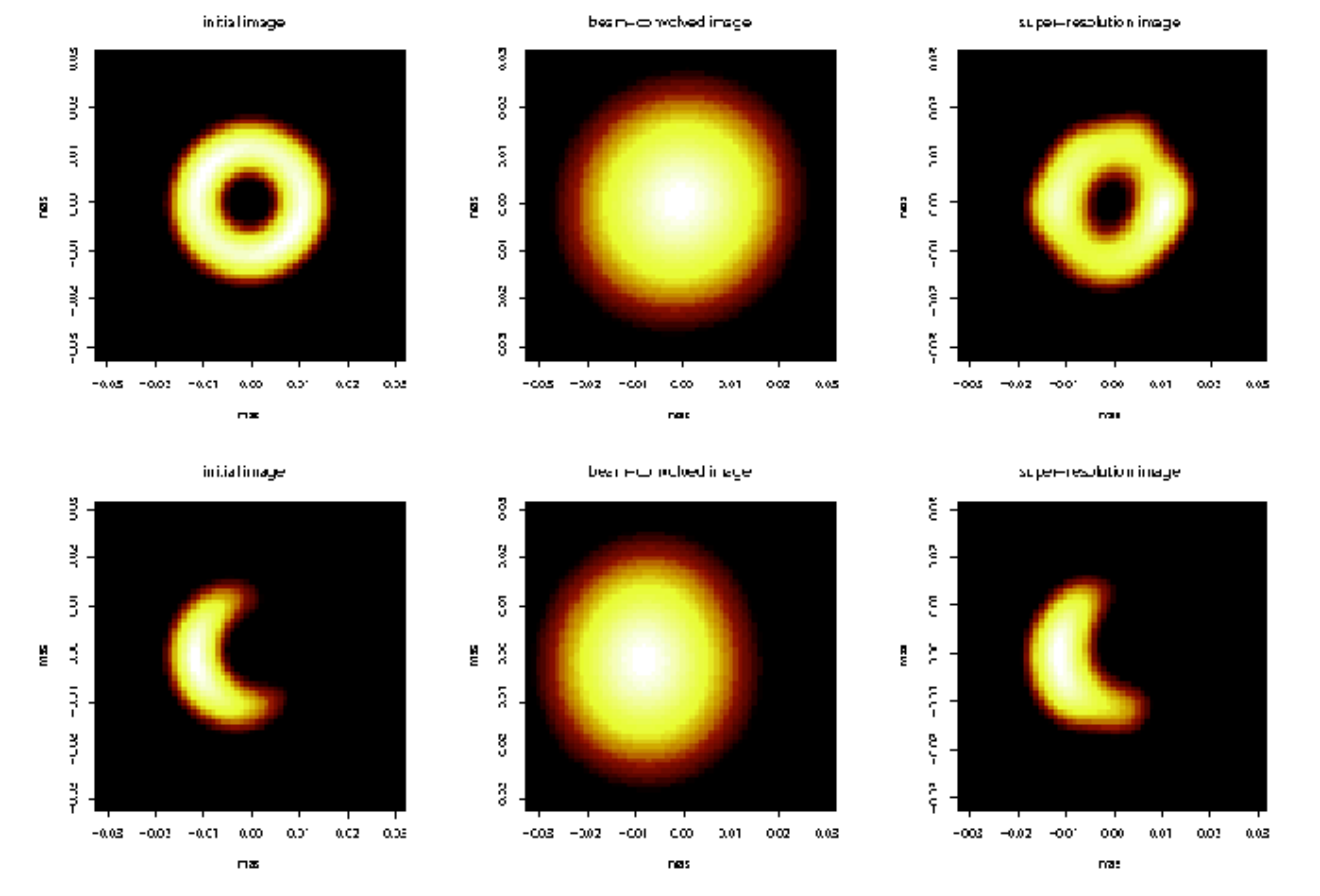
$$\min_{\mathbf{x}} \left(\underbrace{||\mathbf{y} - \mathbf{A}\mathbf{x}||_2^2}_{\text{Chi-square}} + \underbrace{\Lambda_l ||\mathbf{x}||_1}_{\text{Regularization on } \textit{sparsity}} \right) .$$

- Reconstruction purely in the visibility domain:
  Not affected by de-convolution beam (point spread function)

- Many applications after appearance of *Compressed Sensing* (Donoho, Candes+)

Approach 1: Sparse Reconstruction

Application of LASSO (Honma et al. 2014)



(Honma, ***Akiyama***, Uemura & Ikeda 2014, PASJ)

Approach 1: Sparse Reconstruction

For Smoother Image: sparsity on gradient domain

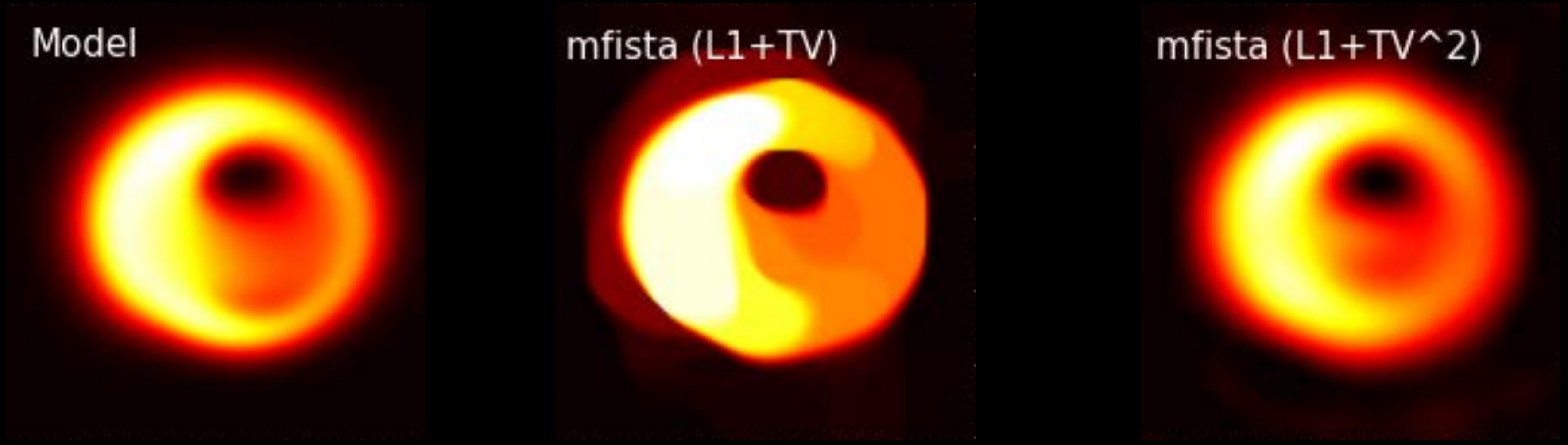
Total Variation: Sparse regularizer of the image in *its gradient domain*

$$||\mathbf{x}||_{\rm tv} = \sum_i \sum_j \sqrt{|x_{i+1,j} - x_{i,j}|^2 + |x_{i,j+1} - x_{i,j}|^2}.$$

$$||\mathbf{x}||_{\rm tv} = \sum_i \sum_j \left(|x_{i+1,j} - x_{i,j}|^2 + |x_{i,j+1} - x_{i,j}|^2\right).$$

L_l + TV regularization (Akiyama et al. 2016b,c, Kuramochi+ in prep.)

$$\min_{\mathbf{x}} \left(||\mathbf{y} - \mathbf{A}\mathbf{x}||_2^2 + \Lambda_l ||\mathbf{x}||_1 + \Lambda_t ||\mathbf{x}||_{\rm tv}\right)$$



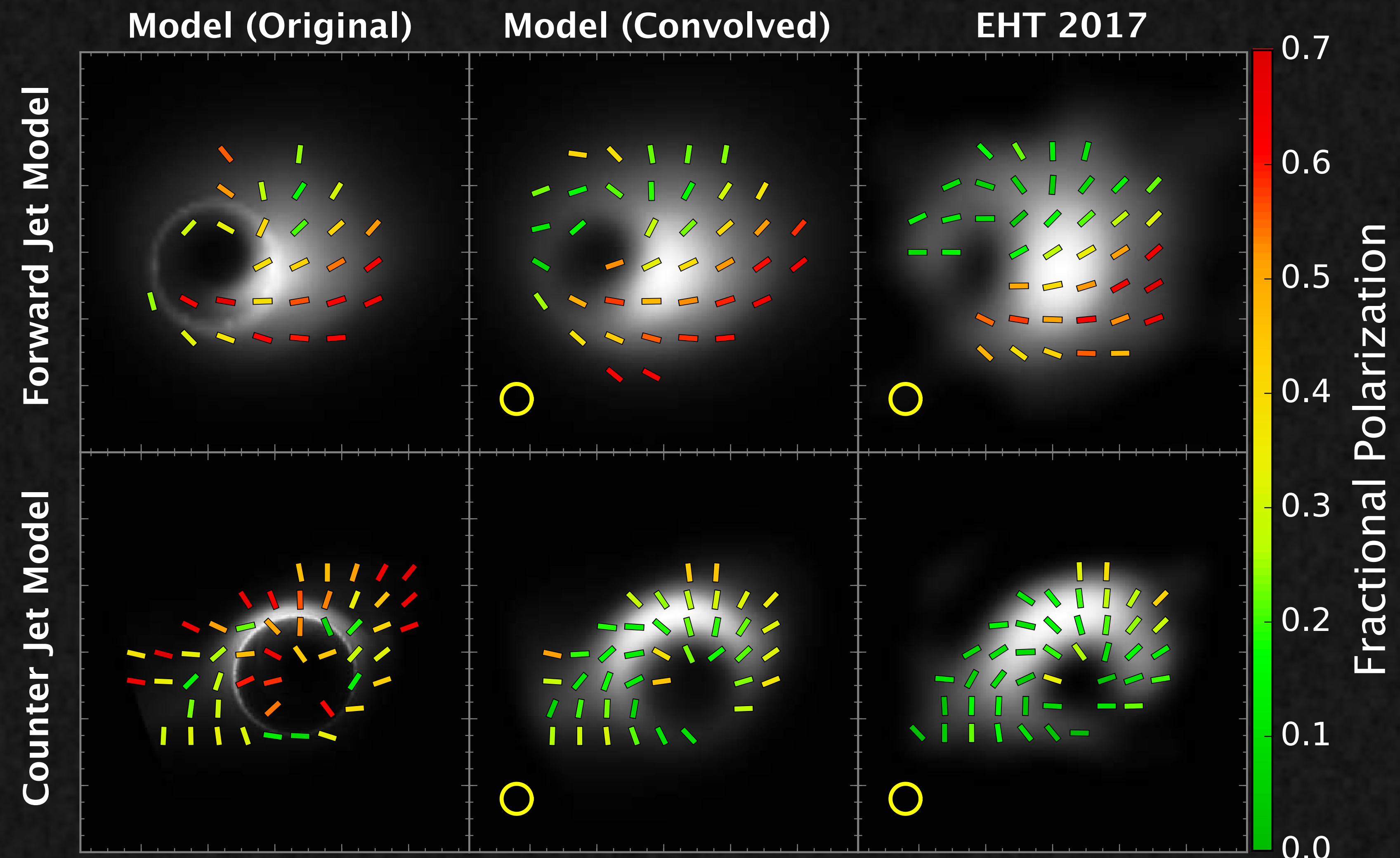
(Sgr A*; Kuramochi, *Akiyama*, et al. in prep.)

Approach I: Sparse Reconstruction

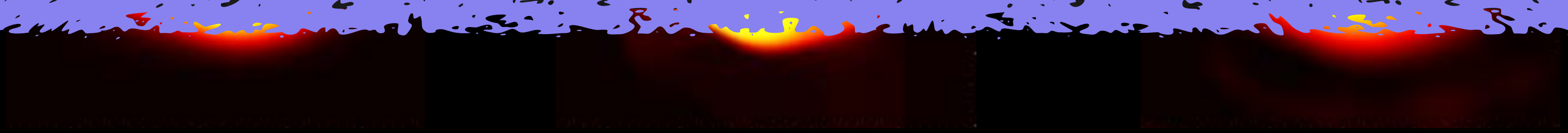
For Smoother Images: sparsity on gradient domain

Total V

L_l + T



(Akiyama et al. 2016c subm. to ApJ)



(Sgr A*; Kuramochi, *Akiyama*, et al. in prep.)

Approach 2: Maximize the Information Entropy

Maximum Entropy Methods (MEM; Frieden 1972; Gull & Daniell 1978)

Approach 2: Maximize the Information Entropy

Maximum Entropy Methods (MEM; Frieden 1972; Gull & Daniell 1978)

$$\min_{\mathbf{x}} \left(||\mathbf{y} - \mathbf{A}\mathbf{x}||_2^2 - \Lambda f_{\mathrm{entropy}}(\mathbf{x}) \right)$$

$$f_{\mathrm{entropy}}(\mathbf{x}) = - \sum_i x_i \log \left(\frac{x_i}{m_i} \right)$$

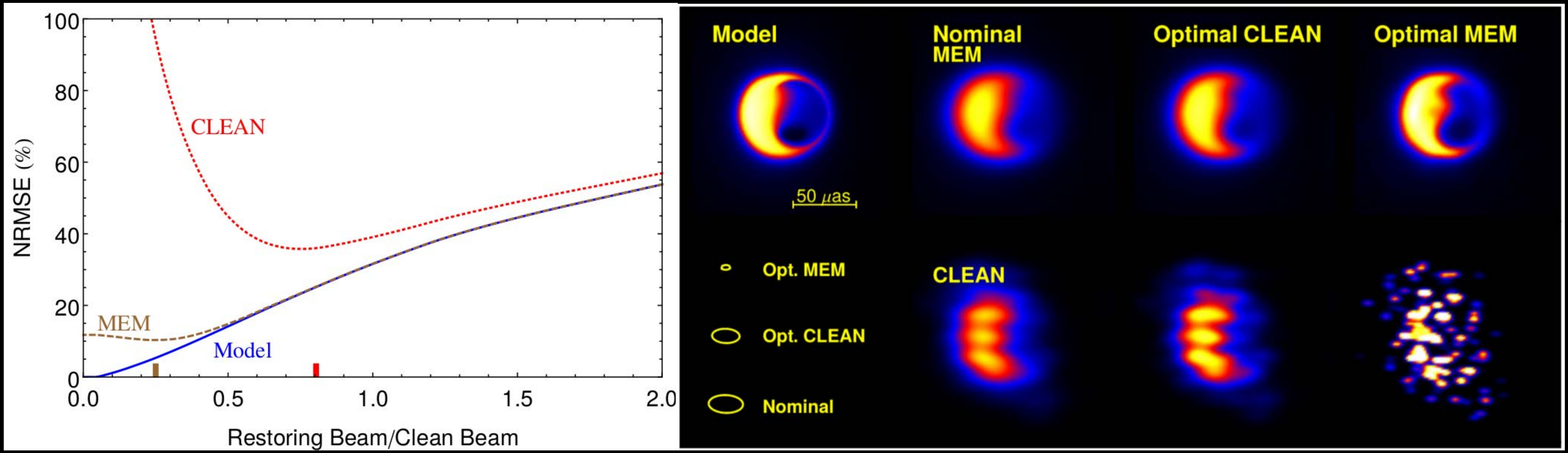
Approach 2: Maximize the Information Entropy

Maximum Entropy Methods (MEM; Frieden 1972; Gull & Daniell 1978)

$$\min_{\mathbf{x}} \left(||\mathbf{y} - \mathbf{A}\mathbf{x}||_2^2 - \Lambda f_{\text{entropy}}(\mathbf{x}) \right)$$

$$f_{\text{entropy}}(\mathbf{x}) = -\sum_i x_i \log\left(\frac{x_i}{m_i}\right)$$

- **Compared with CLEAN:**
  (1) Better fidelity for Smooth Structure (2) Better optimal resolution

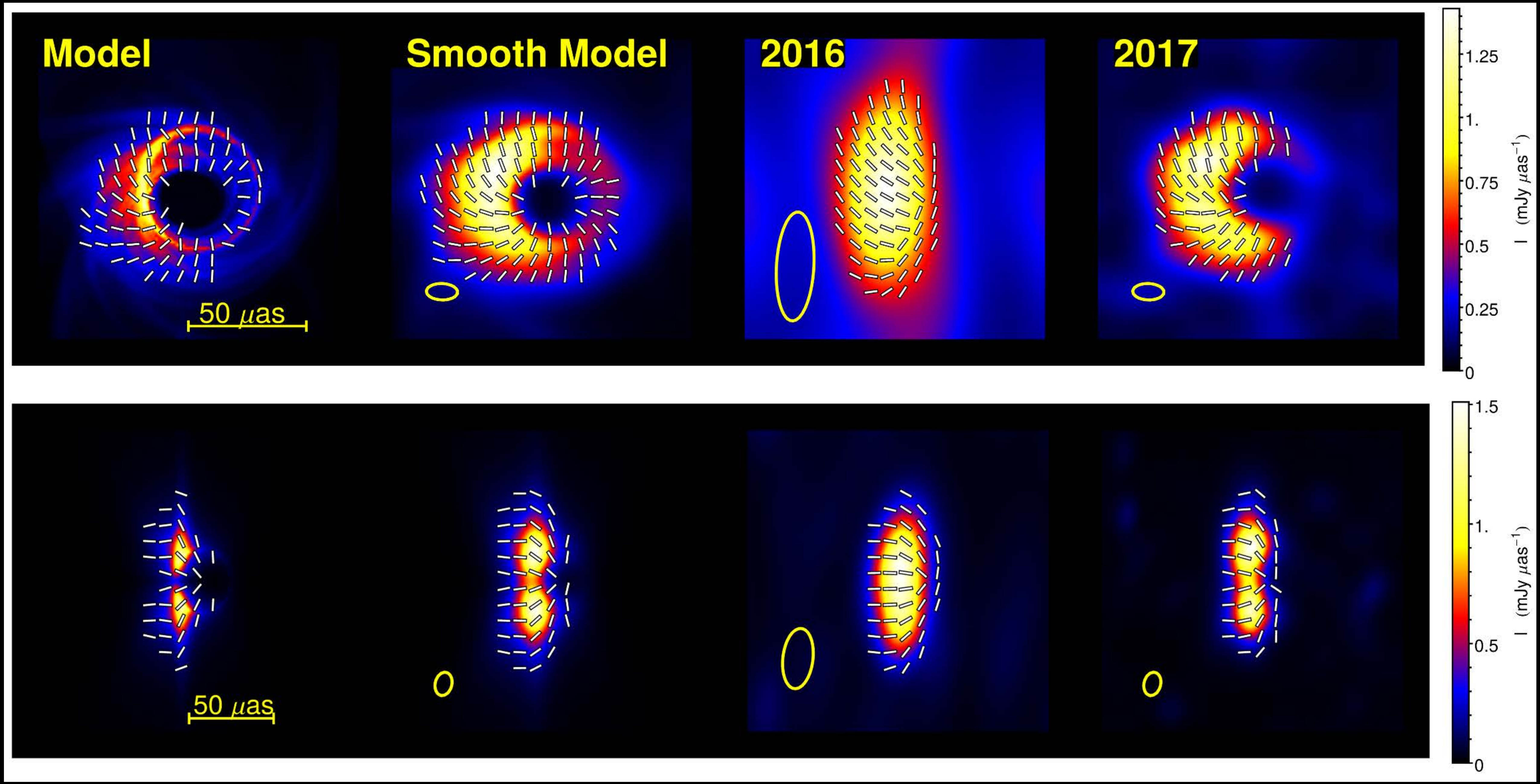


(Chael et al. 2016, ApJ)

Approach 2: Maximize the Information Entropy

Maximum Entropy Methods (MEM; Frieden 1972; Gull & Daniell 1978)

- **PolMEM: Extension of MEM to full-polarimetric Imaging (Chael+16)**



(Chael et al. 2016, ApJ)

Approach 3: Machine-learn Distribution of Image Patches

A patch prior (CHIRP; Bouman et al. 2015)

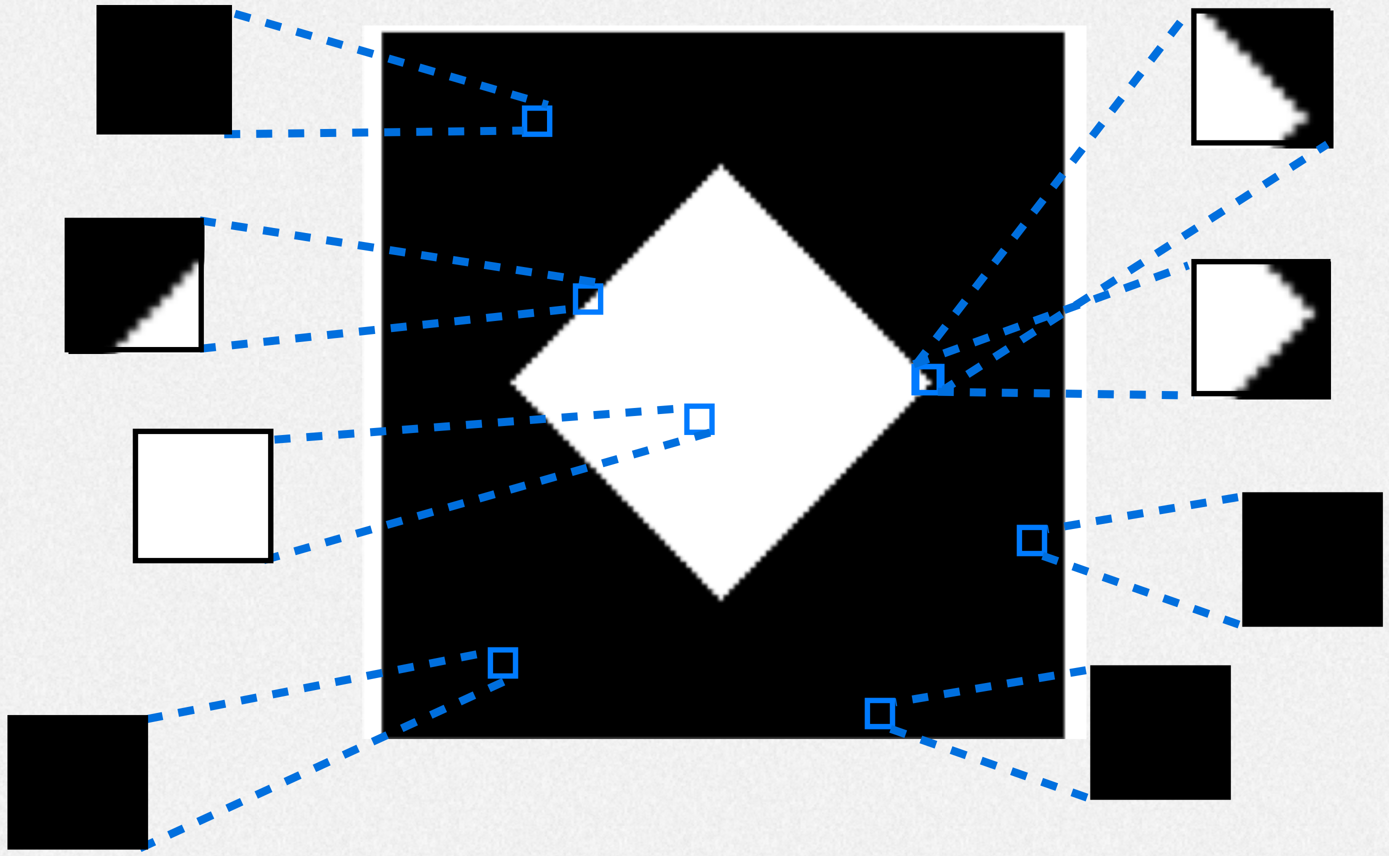
Simple Example

(courtesy of Katie Bouman)

Approach 3: Machine-learn Distribution of Image Patches

A patch prior (CHIRP; Bouman et al. 2015)

Simple Example

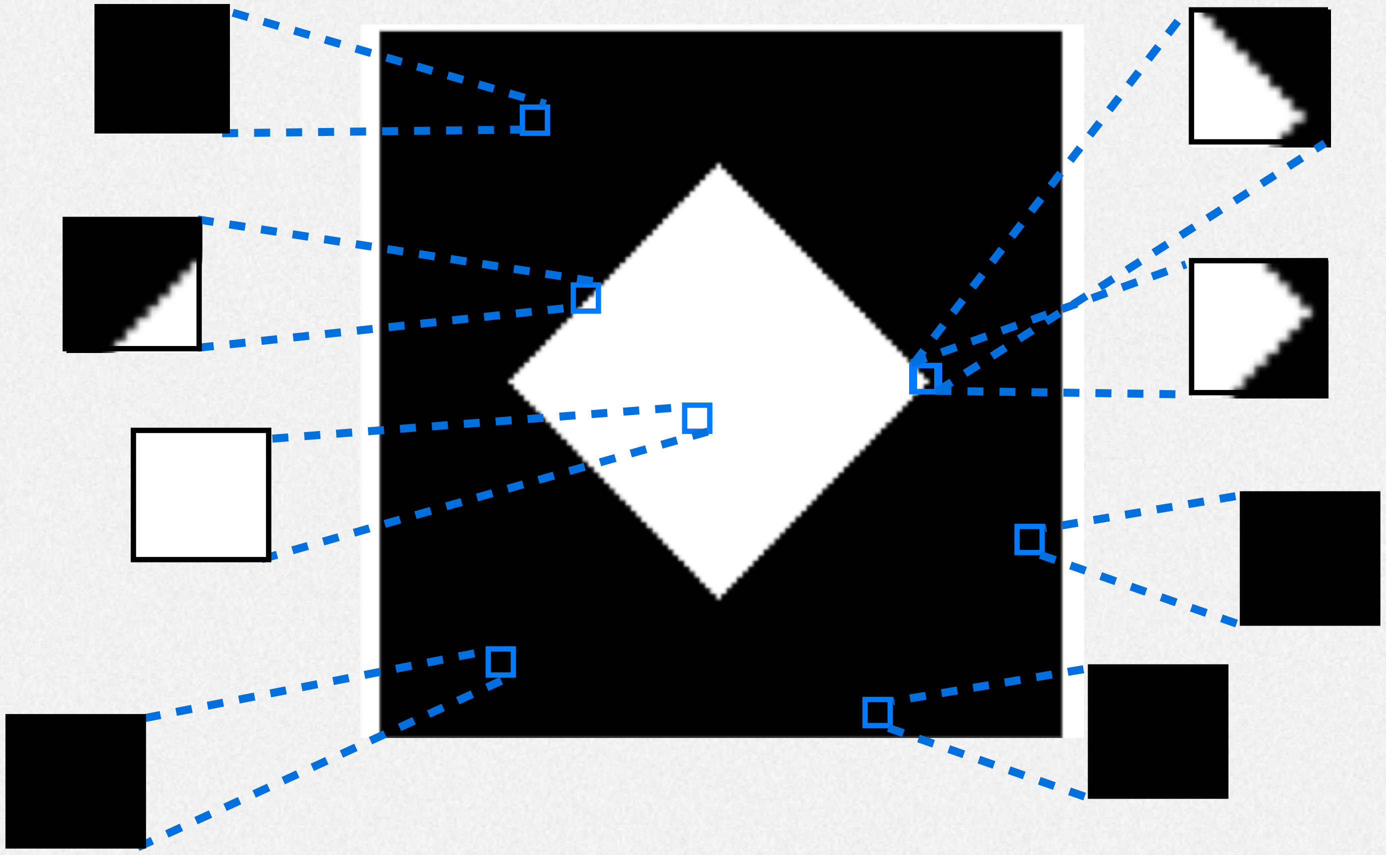


(courtesy of Katie Bouman)

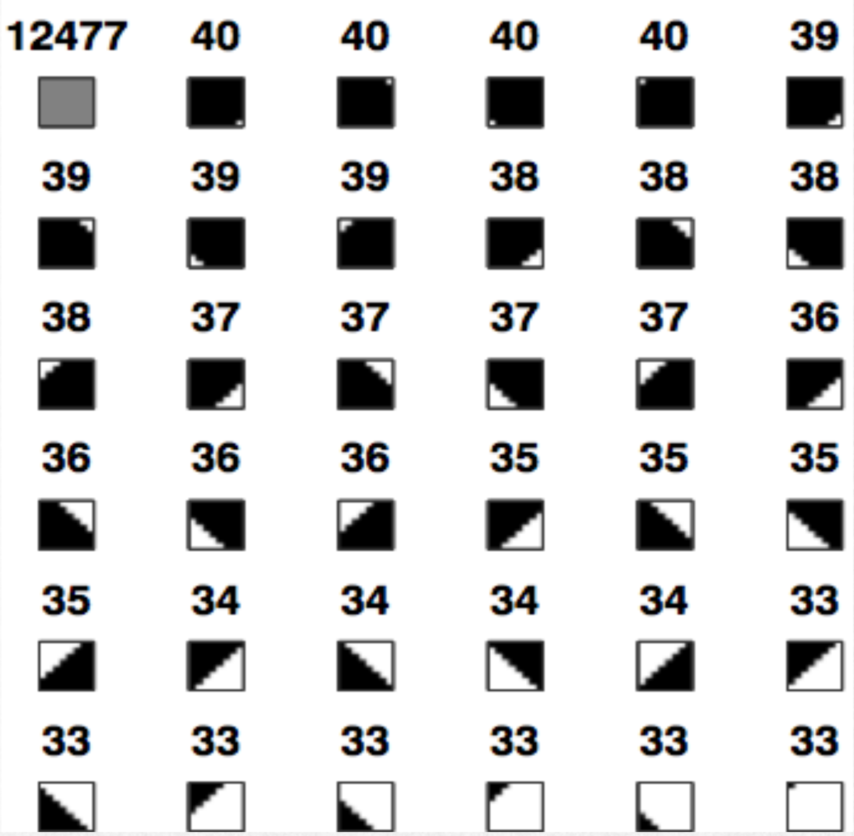
Approach 3: Machine-learn Distribution of Image Patches

A patch prior (CHIRP; Bouman et al. 2015)

Simple Example



Probability Distribution of "Multi-scale Patches"



Can be used as "A Prior Knowledge"

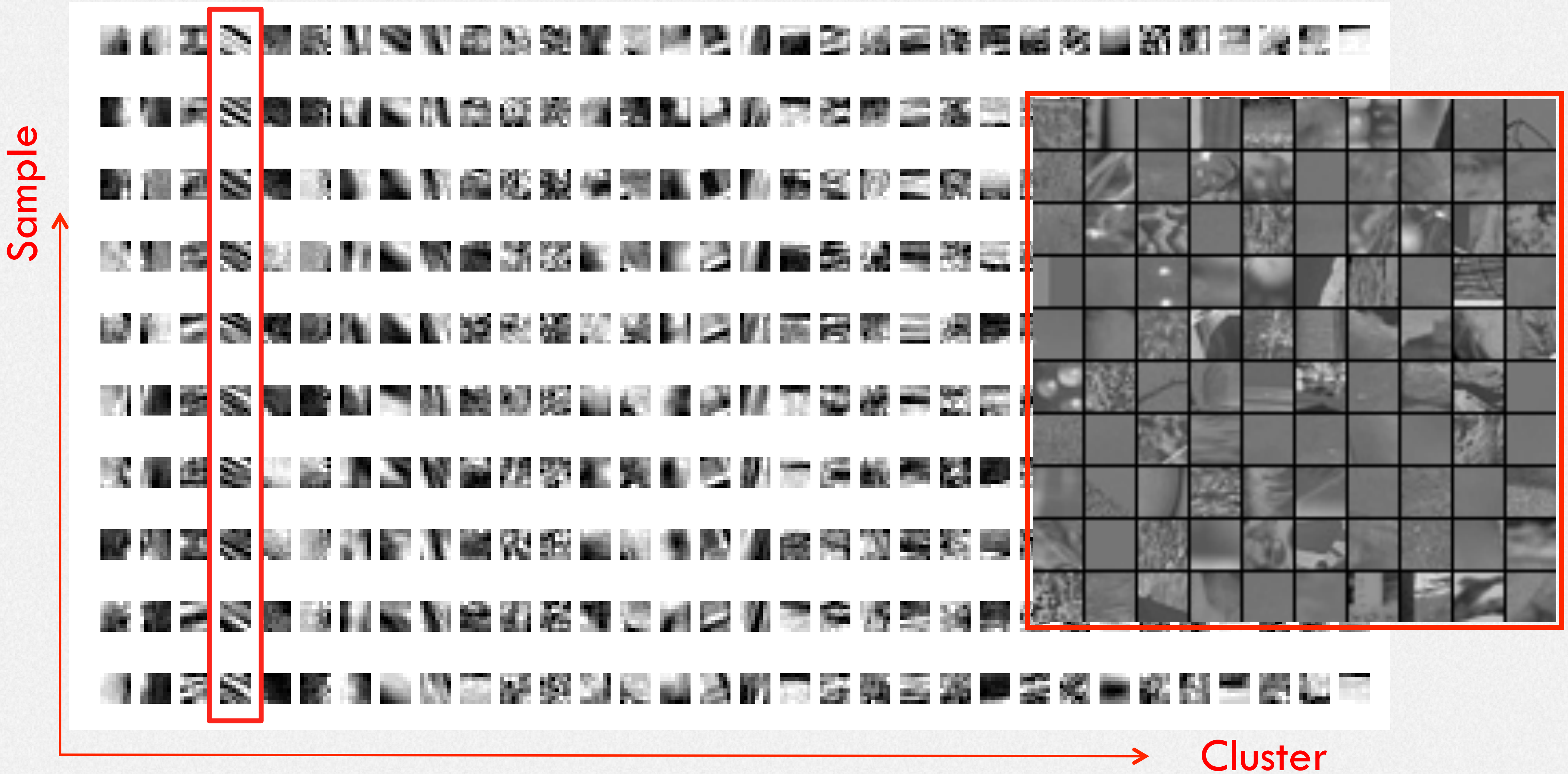
(courtesy of Katie Bouman)

Approach 3: Machine-learn Distribution of Image Patches

A patch prior (CHIRP; Bouman et al. 2015)

CHIRP: Continuous High Image Resolution using Patch priors

Reconstruct the image so that it maximizes consistency with a machine-leaned patch prior distribution



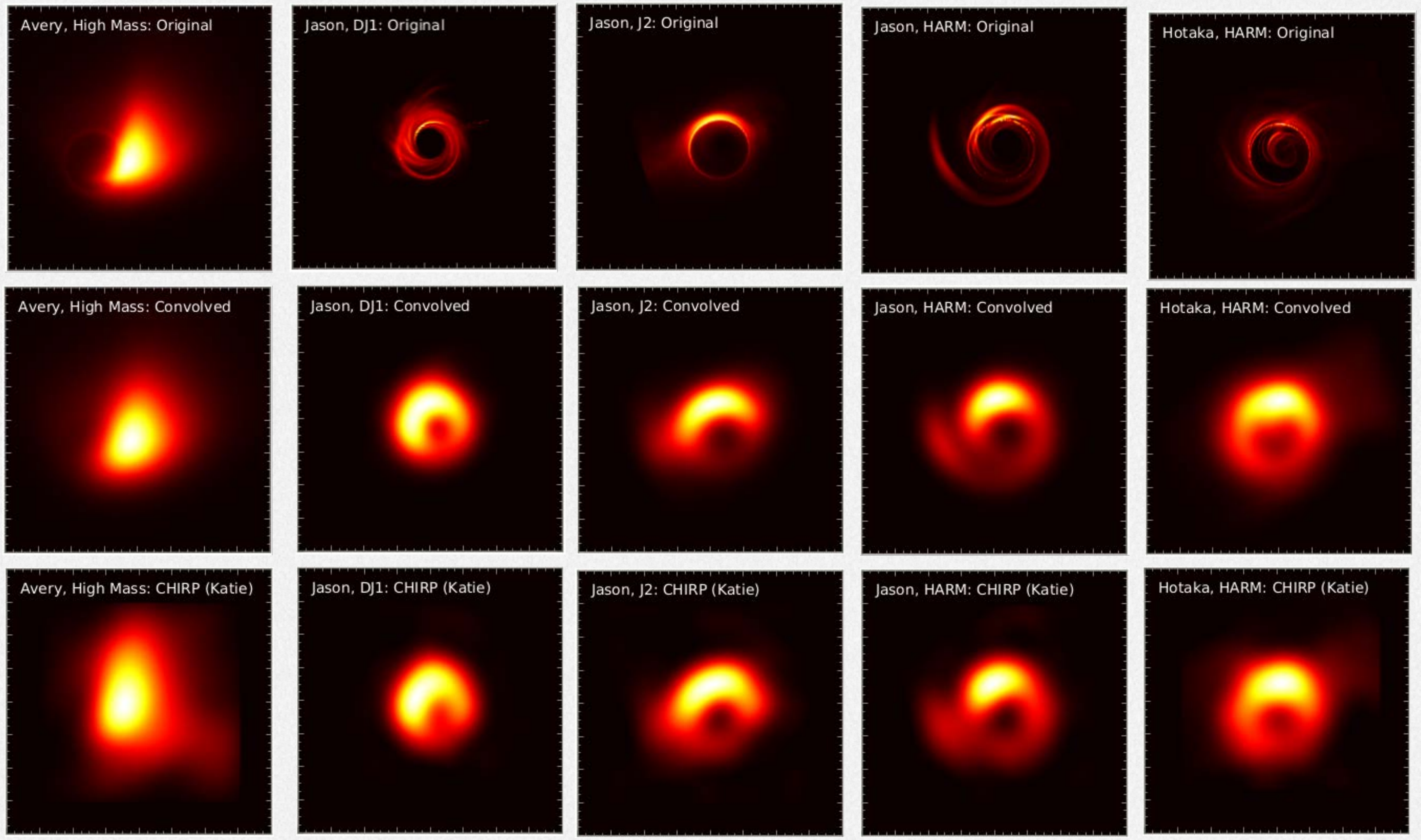
(courtesy of Katie Bouman)

Approach 3: Machine-learn Distribution of Image Patches

A patch prior (CHIRP; Bouman et al. 2015)

CHIRP: Continuous High Image Resolution using Patch priors

Reconstruct the image so that it maximizes consistency with a machine-leaned patch prior distribution



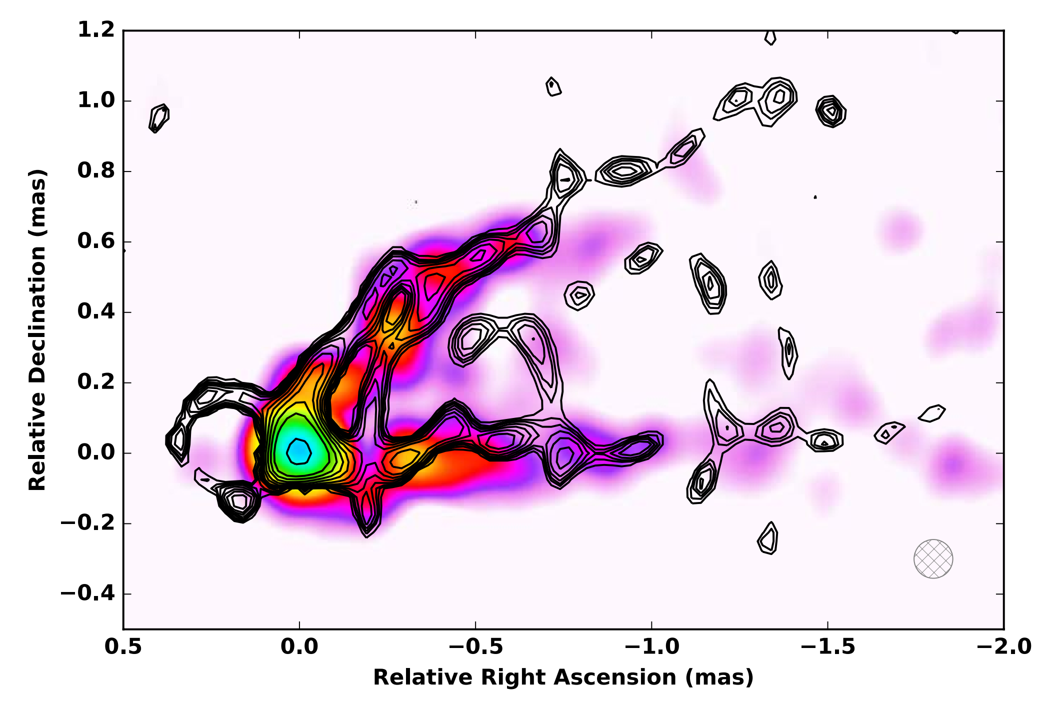
(courtesy of Katie Bouman)

Summary

- All state-of-the-art imaging techniques developed for the EHT have shown much better performance than the traditional CLEAN.
- These techniques can be applied to any existing interferometers
- These techniques would be applicable to similar Fourier-inverse problems

  (e.g.) Faraday Tomography (RM Synthesis)

  Mostly equivalent to linear polarimetric imaging



M87 jets (Application to VLBA data)

- Color: CLEAN (3mm)

- Lines: Sparse Modeling (7mm)