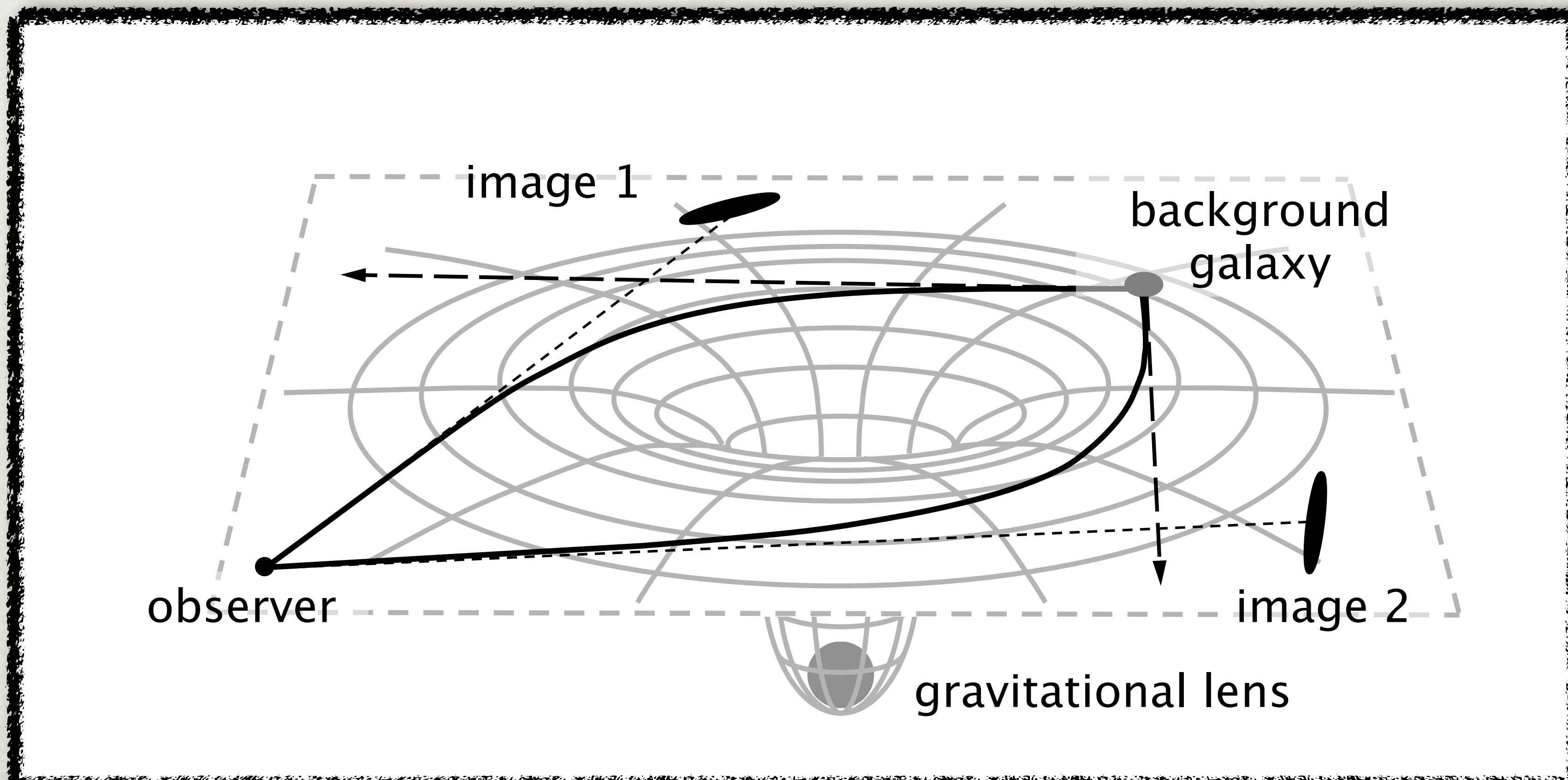
Bayesian modeling of regularized and gravitationally lensed sources

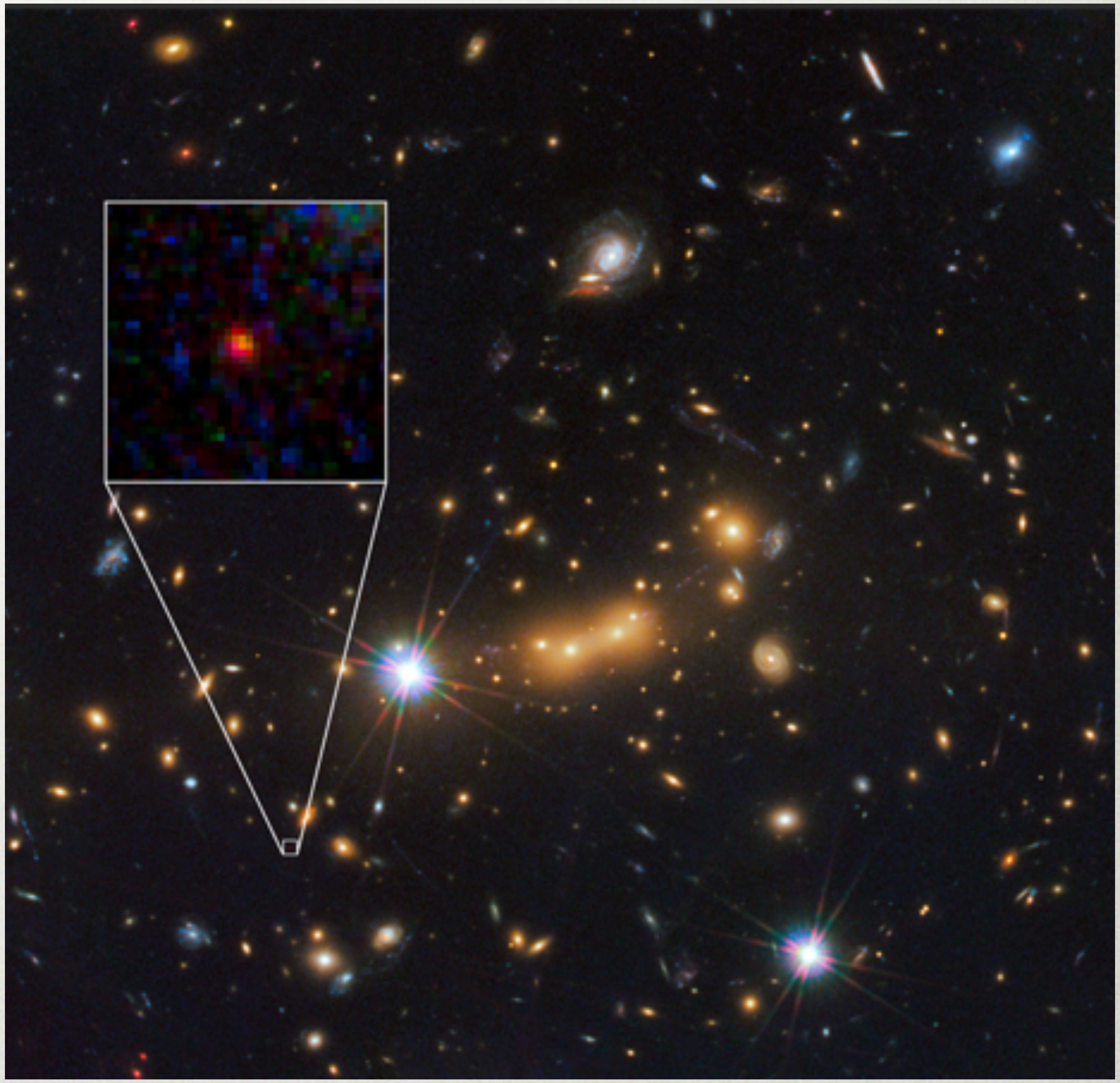
Simona Vegetti - MPA

Gravitational Lensing

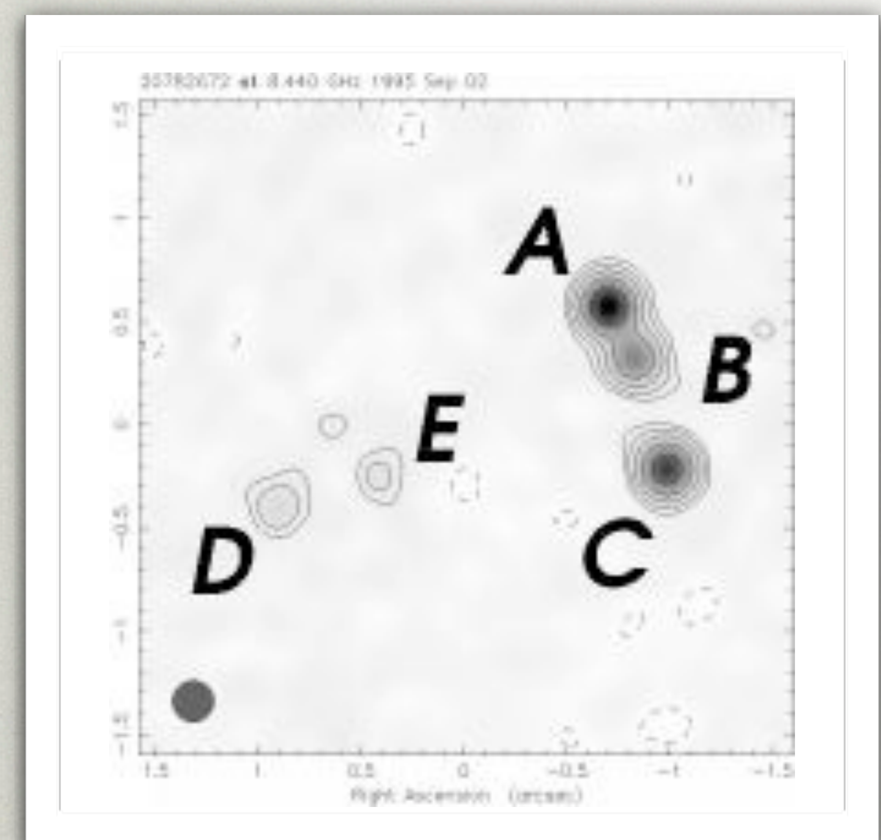


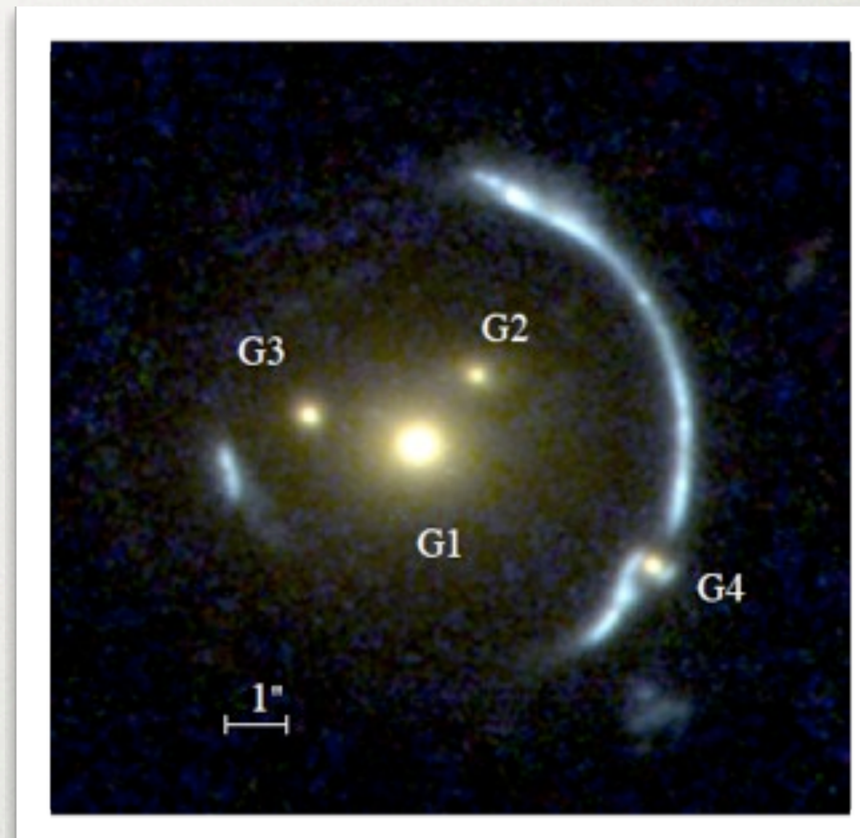
$$\mathbf{y} = \mathbf{x} - \alpha(\mathbf{x}) \qquad \alpha(\mathbf{x}) \propto \int d\mathbf{x}' \int dz \rho(\mathbf{x}', z) \frac{\mathbf{x} - \mathbf{x}'}{|\mathbf{x} - \mathbf{x}'|^2}$$

Gravitational Lensing, Why?

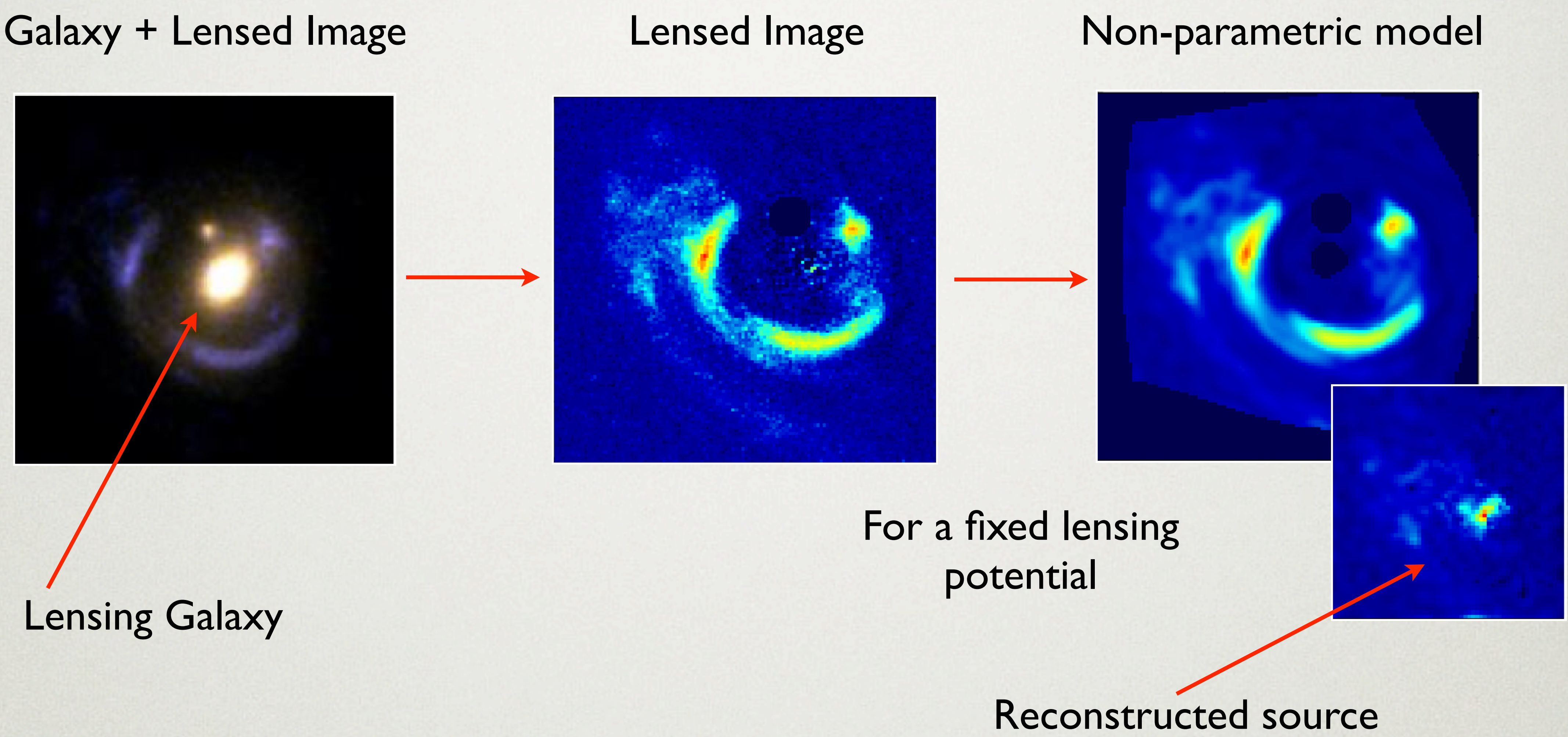


Gravitational Lensing

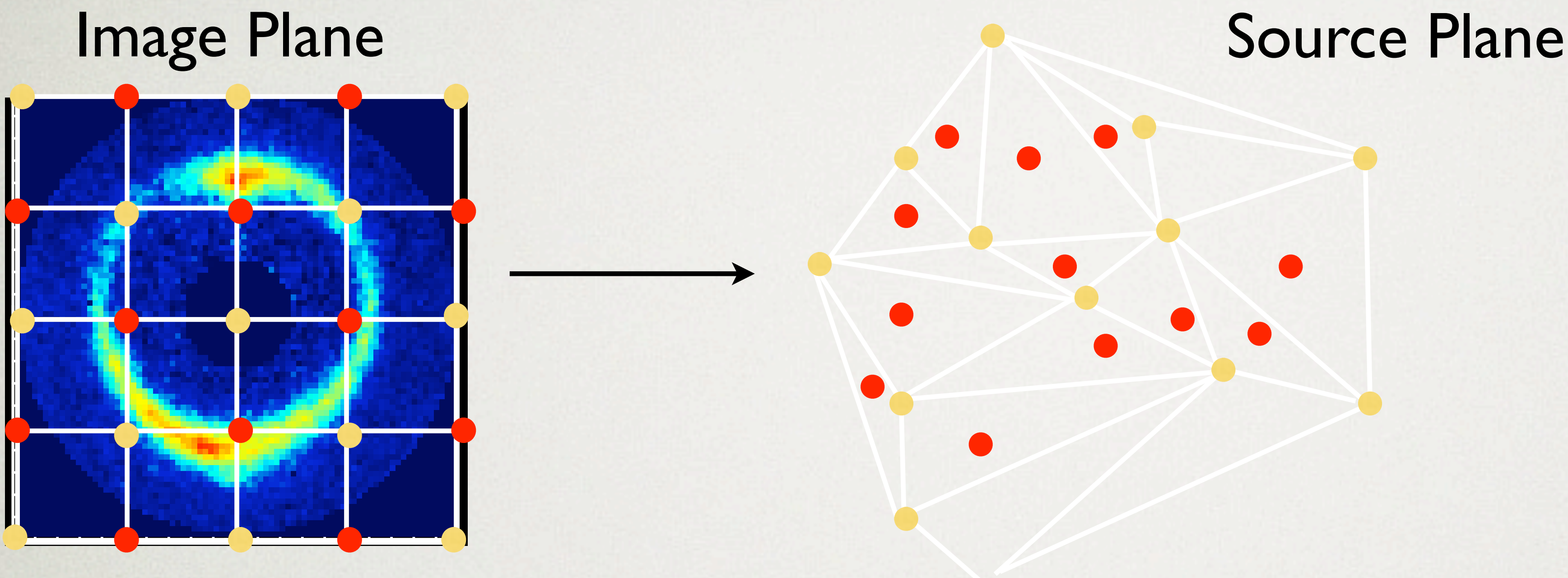


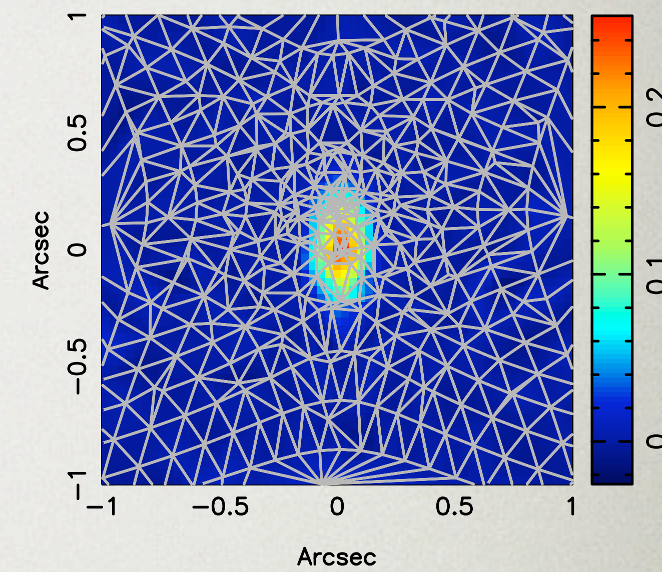


- makes very little to no assumptions about the structure of the source (and possibly the lens potential as well; see end of this lecture), although the solutions often require regularization because the number of free parameters can be large.
- makes use of all (or most) information available in the lensed images or even absence of information; i.e. the prediction of lensed images that are not present in the data are penalized.
- allows structure of the source to be separated from structure in the lens potential, in a statistical (i.e. Bayesian) sense.



- Pixelated source defined on adaptive grid+ analytic main lens mass model
- Pixelated potential corrections
- All embedded in a Bayesian framework



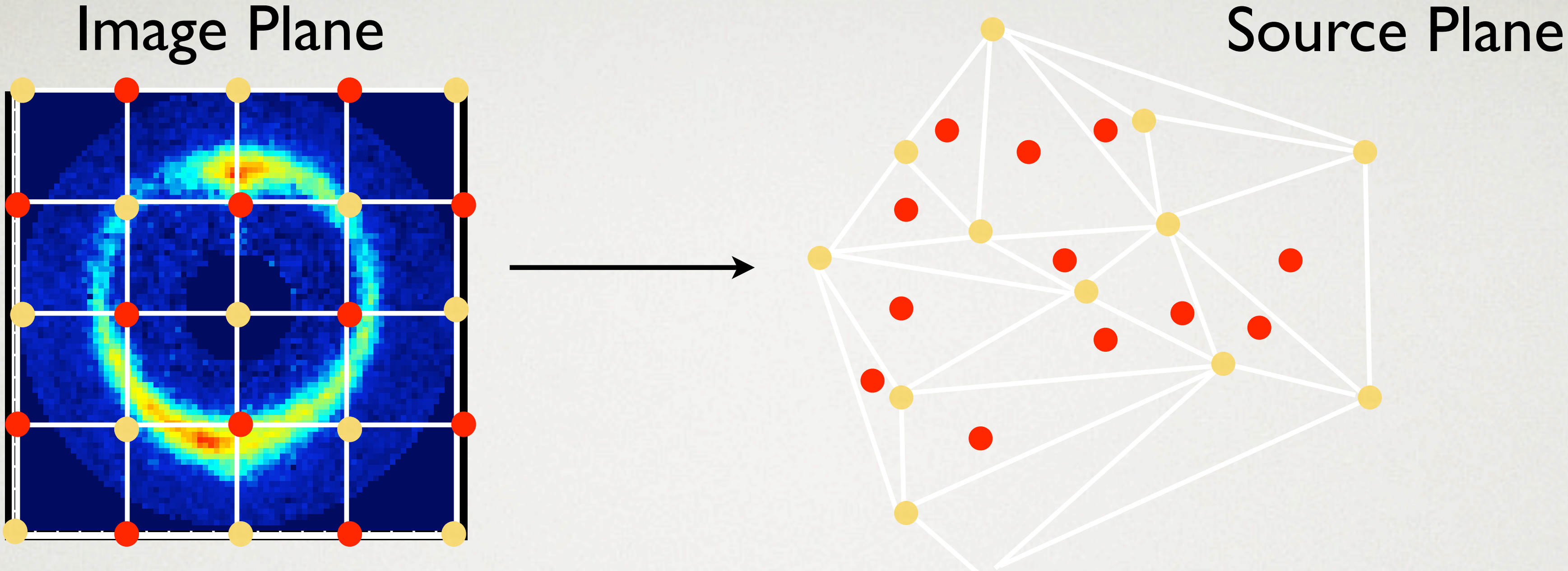


- Choose a parametric model for the total mass lens distribution: elliptical power-law defined by the free parameters p (include external shear)

- Each point $x_{i,j}$ with its corresponding surface brightness distribution $d_{i,j}$ maps to a point on the source plane $y_{l,m}$ via the lens equation

$$\vec{y}'_{i,j} = \vec{x}_{i,j} - \vec{\nabla}\psi(\vec{x}_{i,j};\vec{p})$$

- Assign to the point $y_{l,m}$ a surface brightness value $s_{i,j}$ using surface brightness conservation



$$s'_{i,j} = \sum_{\mu=0}^{1} \sum_{\nu=0}^{1} w_{k'+\mu,l'+\nu} s_{k'+\mu,l'+\nu}$$

$$\begin{cases} w_{k',l'} & = & (1-t)(1-u) \\ w_{k'+1,l'} & = & t(1-u) \\ w_{k'+1,l'+1} & = & tu \\ w_{k',l'+1} & = & (1-t)u \end{cases}$$

$$s'_{i,j} = \vec{l}^{\,T}_{i,j}\,\vec{s}$$

$$\boldsymbol{L}(\psi) \equiv \begin{pmatrix} \vec{l}^{\,T}_{0,0} \\ \vdots \\ \vec{l}^{\,T}_{i,j} \\ \vdots \\ \vec{l}^{\,T}_{\mathrm{ndim}_i,\mathrm{ndim}_j} \end{pmatrix}$$

$$\vec{d}\,' = \boldsymbol{B}\boldsymbol{L}(\psi)\vec{s}$$

$$\vec{s}\,' = \overbrace{[\boldsymbol{B}\boldsymbol{L}(\psi)]^{+}}^{\mathrm{Pseudo-inverse}}\ \vec{d}\,.$$

In *practical* terms

$$P = \chi^2(\vec{s}\,, \psi) + \lambda\, \mathrm{Reg}(\vec{s}\,)$$

$$P_{\chi^2} = [\boldsymbol{B}\boldsymbol{L}(\psi)\vec{s} - \vec{d}\,]^T \boldsymbol{C}^{-1} [\boldsymbol{B}\boldsymbol{L}(\psi)\vec{s} - \vec{d}\,] + \lambda_s\, [\vec{s}^{\,T}(\boldsymbol{H}^T\boldsymbol{H})\vec{s}\,]$$

$$(\boldsymbol{M}^T\boldsymbol{C}^{-1}\boldsymbol{M}^T + \lambda\, \boldsymbol{H}^T\boldsymbol{H})\vec{s} = \boldsymbol{M}^T\boldsymbol{C}^{-1}\vec{d}\,, \qquad \boldsymbol{M} \equiv \boldsymbol{B}\boldsymbol{L}(\psi)$$

In Bayesian terms

$$\overbrace{P(\vec{s}\,|\,\vec{d}\,, \lambda, \boldsymbol{L}(\vec{p}), \boldsymbol{H})}^{\text{Posterior}} = \frac{\overbrace{P(\vec{d}\,|\,\vec{s}\,, \boldsymbol{L}(\vec{p}))}^{\text{Likelihood}} \times \overbrace{P(\vec{s}\,|\lambda, \boldsymbol{H})}^{\text{Prior}}}{\underbrace{P(\vec{d}\,|\lambda, \boldsymbol{L}(\vec{p}), \boldsymbol{H})}_{\text{Evidence}}}$$

$$\mathcal{P} \propto e^{-\frac{1}{2}(\chi^2 + \lambda\vec{s}^{\,T}\boldsymbol{H}^T\boldsymbol{H}\vec{s}\,)} = \overbrace{e^{-\frac{1}{2}\chi^2}}^{\text{Likelihood}} \times \overbrace{e^{-\frac{1}{2}\lambda\vec{s}^{\,T}\boldsymbol{H}^T\boldsymbol{H}\vec{s}}}^{\text{Regularisation Prior}}$$

In *practical* terms

$$P = \chi^2(\vec{s}\,, \psi) + \lambda\,\mathrm{Reg}(\vec{s}\,)$$

The best non-linear parameter of the lensing potential are derived by minimizing the penalty function with a Simplex downhill method

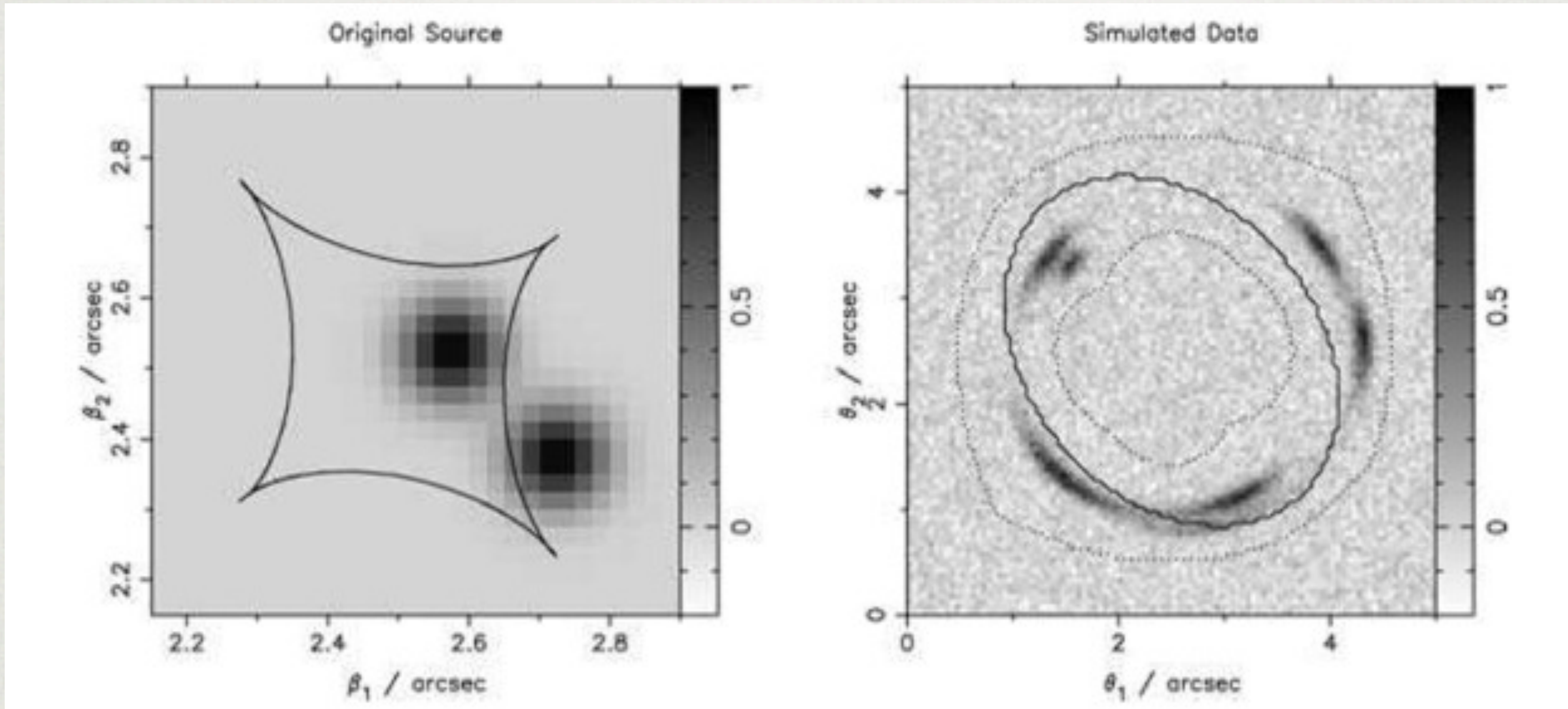
---

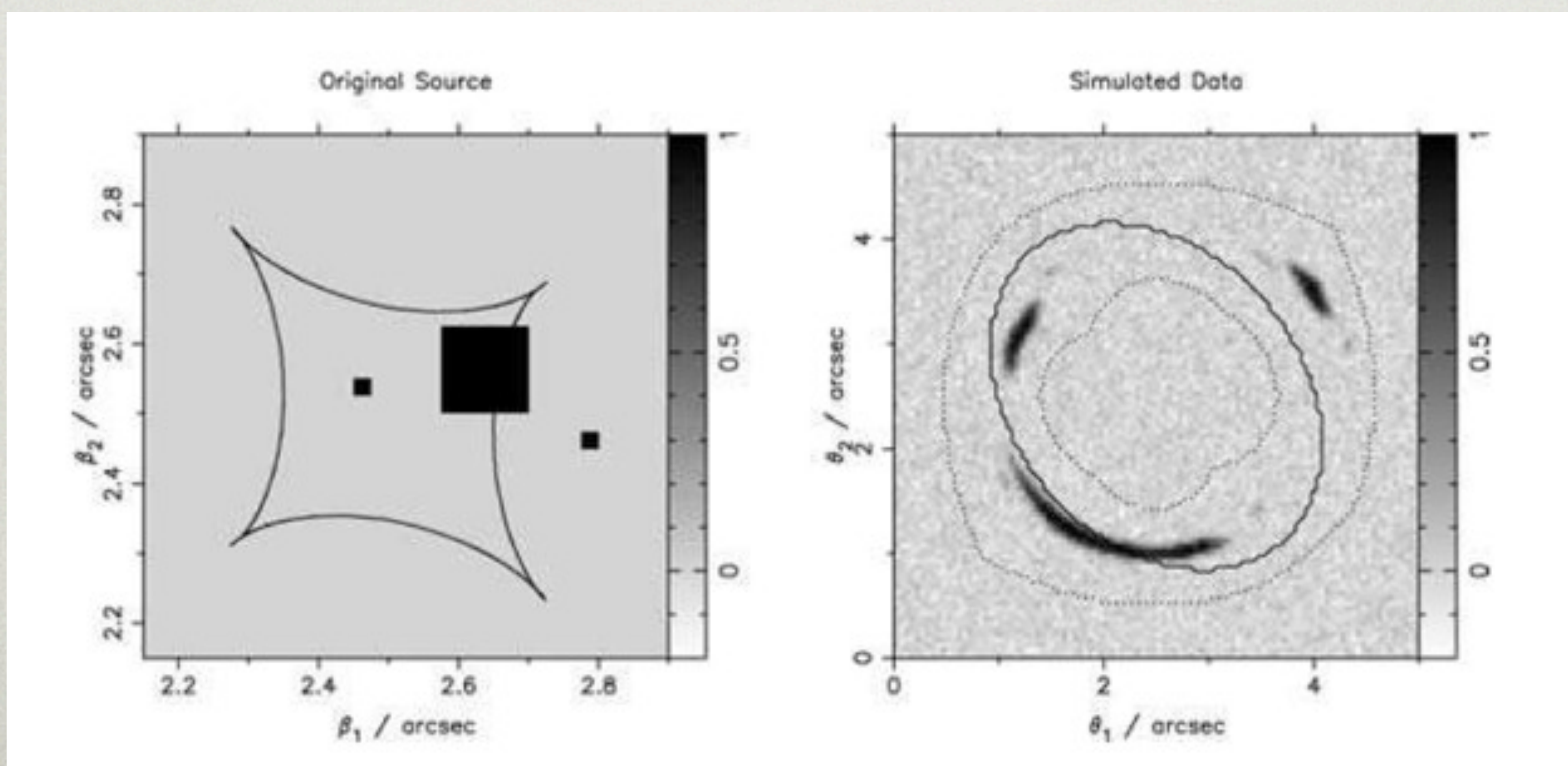
In Bayesian terms

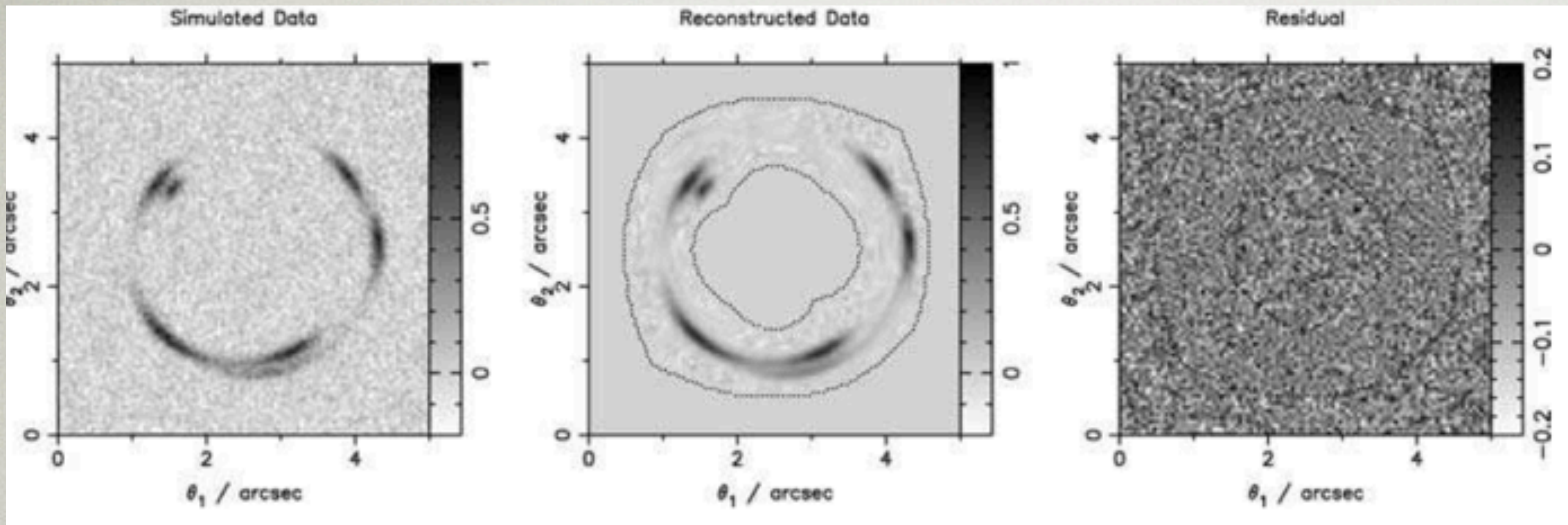
$$P(\mathbf{p}, \lambda | \mathbf{d}, \mathbf{L}, H) = \frac{P(\mathbf{d} | \mathbf{p}, \lambda, \mathbf{L}, H) P(\mathbf{p}, \lambda)}{P(\mathbf{d} | \mathbf{L}, H)}$$

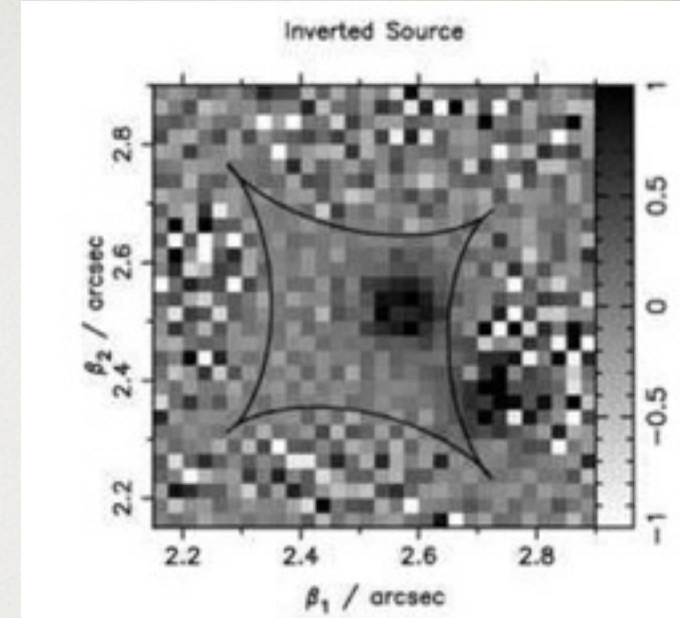
$$P(\mathbf{d} | \mathbf{p}, \lambda, \mathbf{L}, H) = \int P(\mathbf{d} | \mathbf{p}, s, \mathbf{L}) P(s | H, \lambda) ds$$

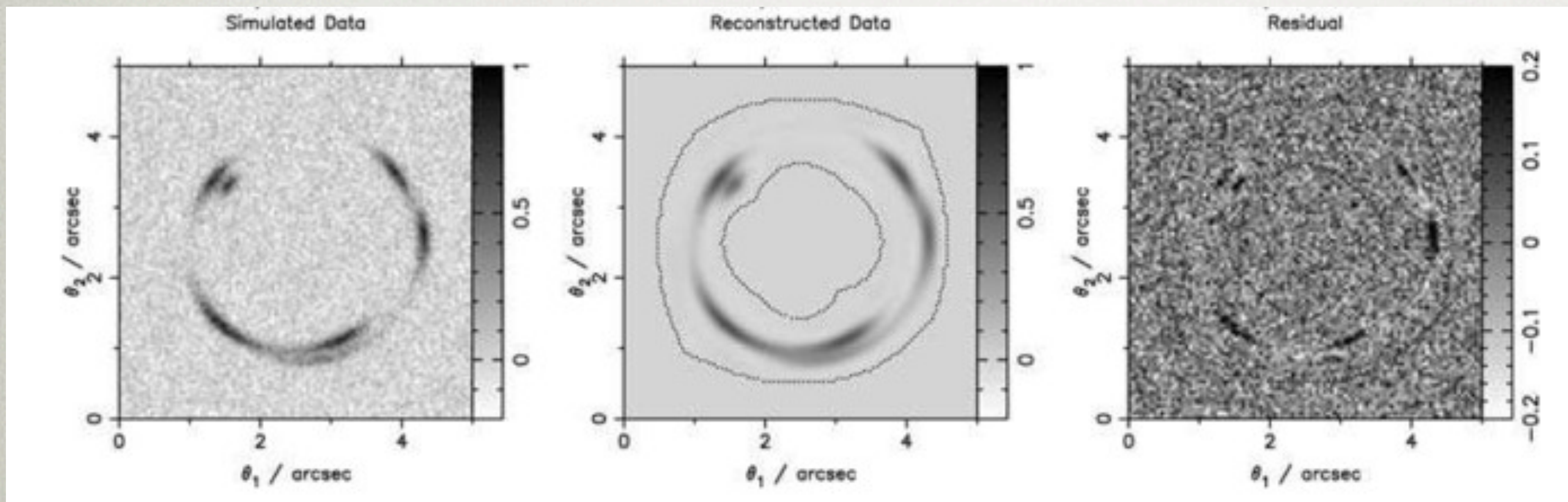
$$\begin{aligned}\log\left(P(\vec{d}\,|\,\lambda, \boldsymbol{M} \equiv \boldsymbol{B}\,\boldsymbol{L}, \boldsymbol{H})\right) \;=\; & -\frac{1}{2}[\chi^2 + \lambda \parallel \boldsymbol{H}\,\vec{s}\parallel^2] - \frac{1}{2}\log\left[\det\left(\boldsymbol{M}^T\boldsymbol{C}^{-1}\boldsymbol{M} + \lambda\,\boldsymbol{H}^T\boldsymbol{H}\right)\right] + \frac{N_s}{2}\log(\lambda) \\ & + \frac{1}{2}\log[\det\left(\boldsymbol{H}^T\boldsymbol{H}\right)] - \frac{N_d}{2}\log(2\pi) + \frac{1}{2}\log[\det\left(\boldsymbol{C}^{-1}\right)].\end{aligned}$$

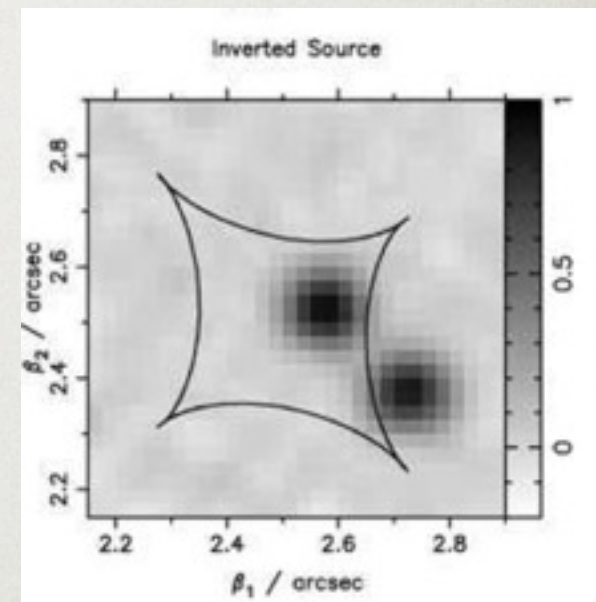

Original Source
Simulated Data
β2 / arcsec
β1 / arcsec
θ2 / arcsec
θ1 / arcsec

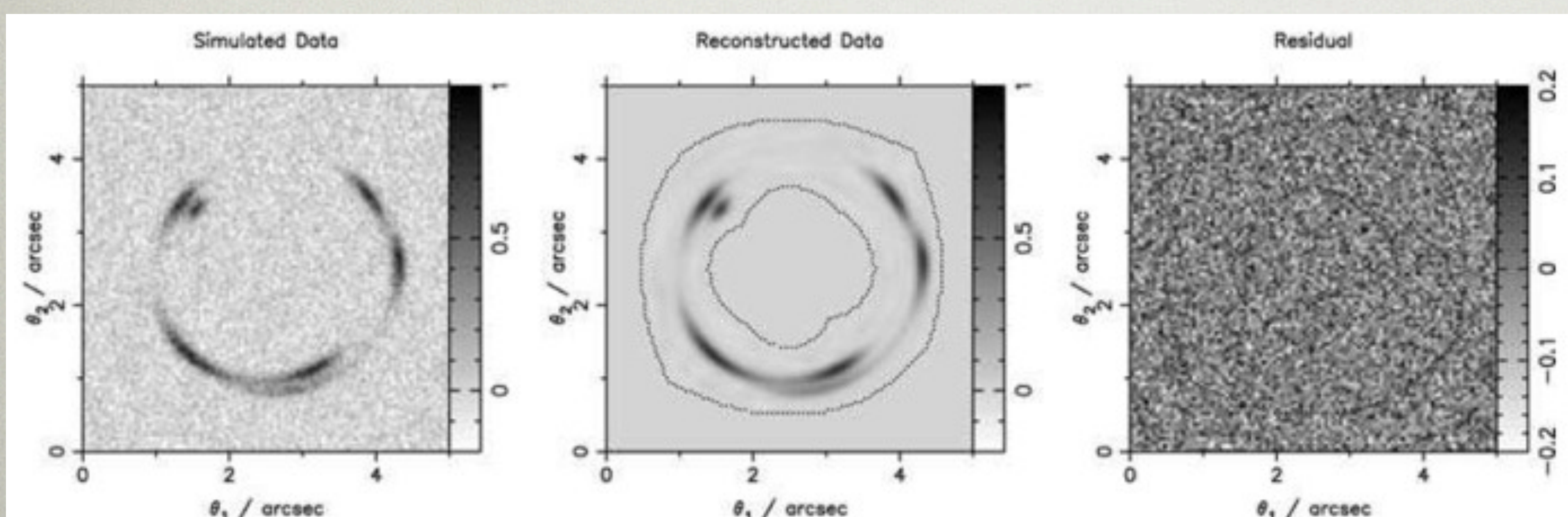

Original Source
Simulated Data
β2 / arcsec
β1 / arcsec
θ2 / arcsec
θ1 / arcsec

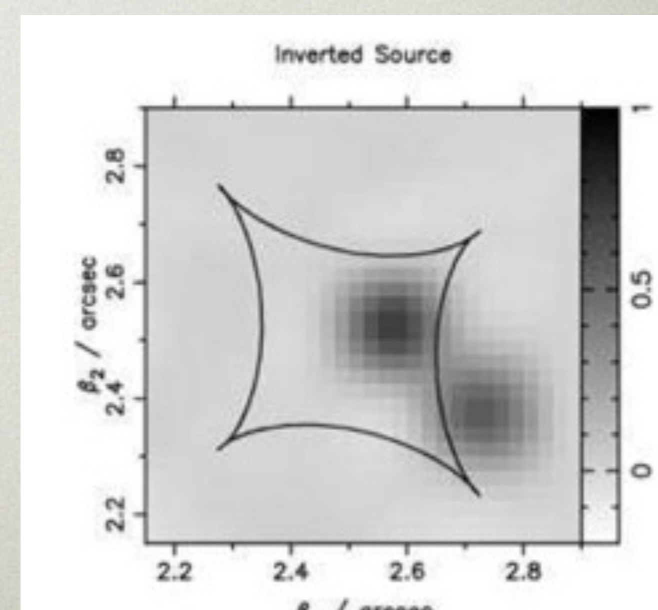

Simulated Data
Reconstructed Data
Residual
θ_1 / arcsec
θ_2 / arcsec

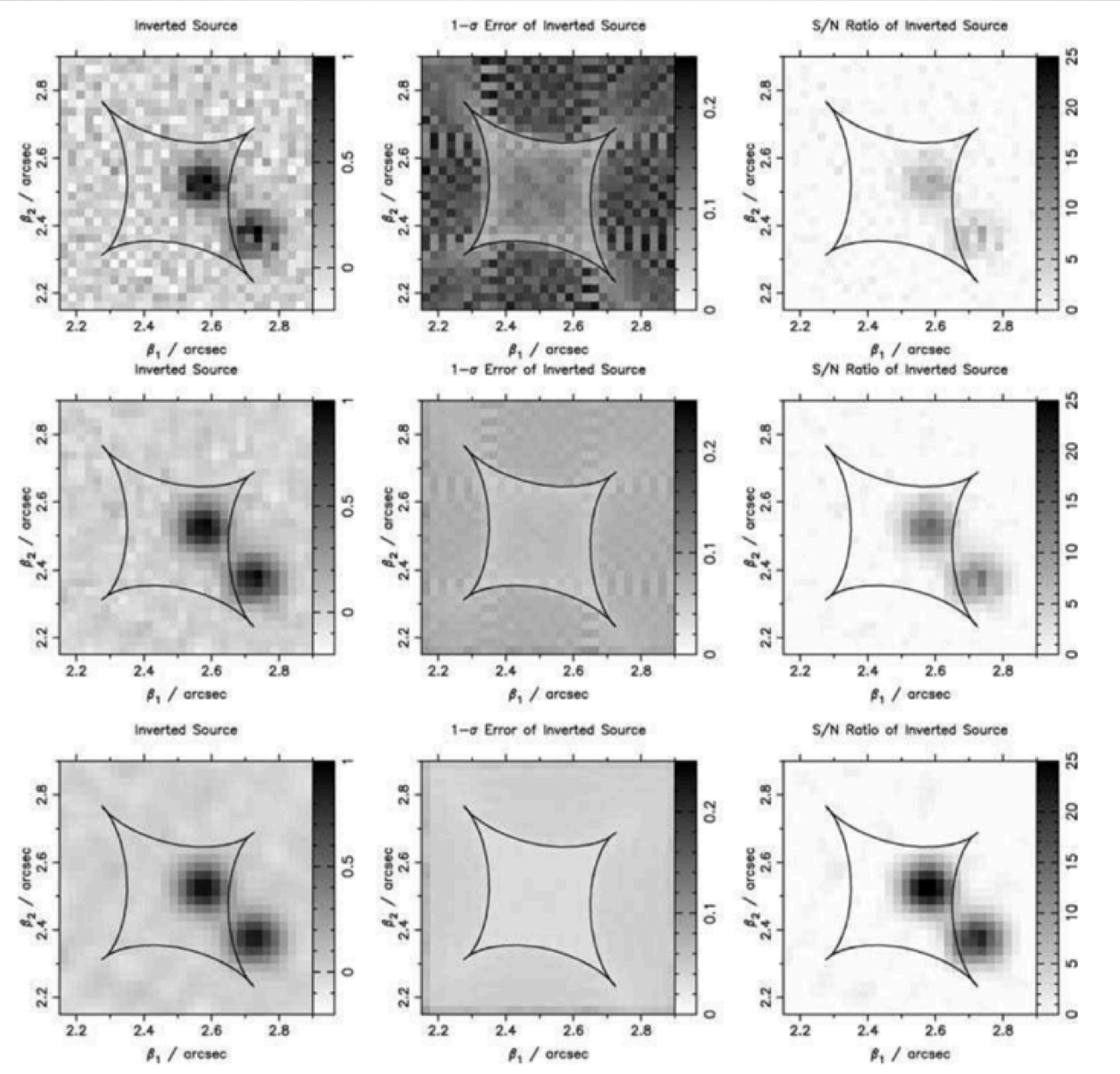

Inverted Source
β_1 / arcsec
β_2 / arcsec

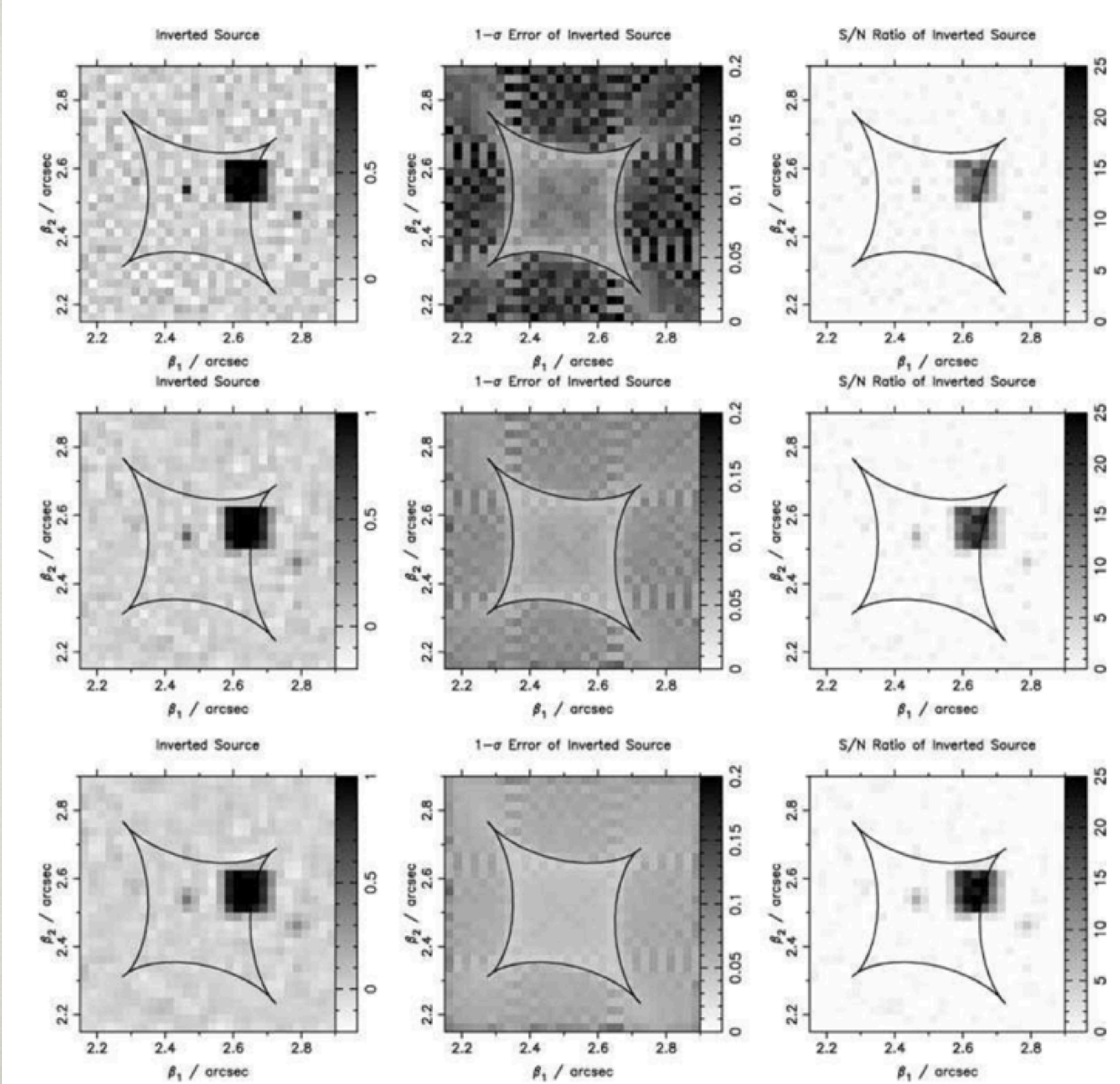

Simulated Data
Reconstructed Data
Residual
θ_1 / arcsec
θ_2 / arcsec


Inverted Source
β_1 / arcsec
β_2 / arcsec


Simulated Data
Reconstructed Data
Residual
θ_1 / arcsec
θ_2 / arcsec


Inverted Source
β_1 / arcsec
β_2 / arcsec


Inverted Source
1-σ Error of Inverted Source
S/N Ratio of Inverted Source
β₂ / arcsec
β₁ / arcsec


Inverted Source
1−σ Error of Inverted Source
S/N Ratio of Inverted Source
β₂ / arcsec
β₁ / arcsec
2.2
2.4
2.6
2.8
0
0.5
1
0.05
0.1
0.15
0.2
5
10
15
20
25

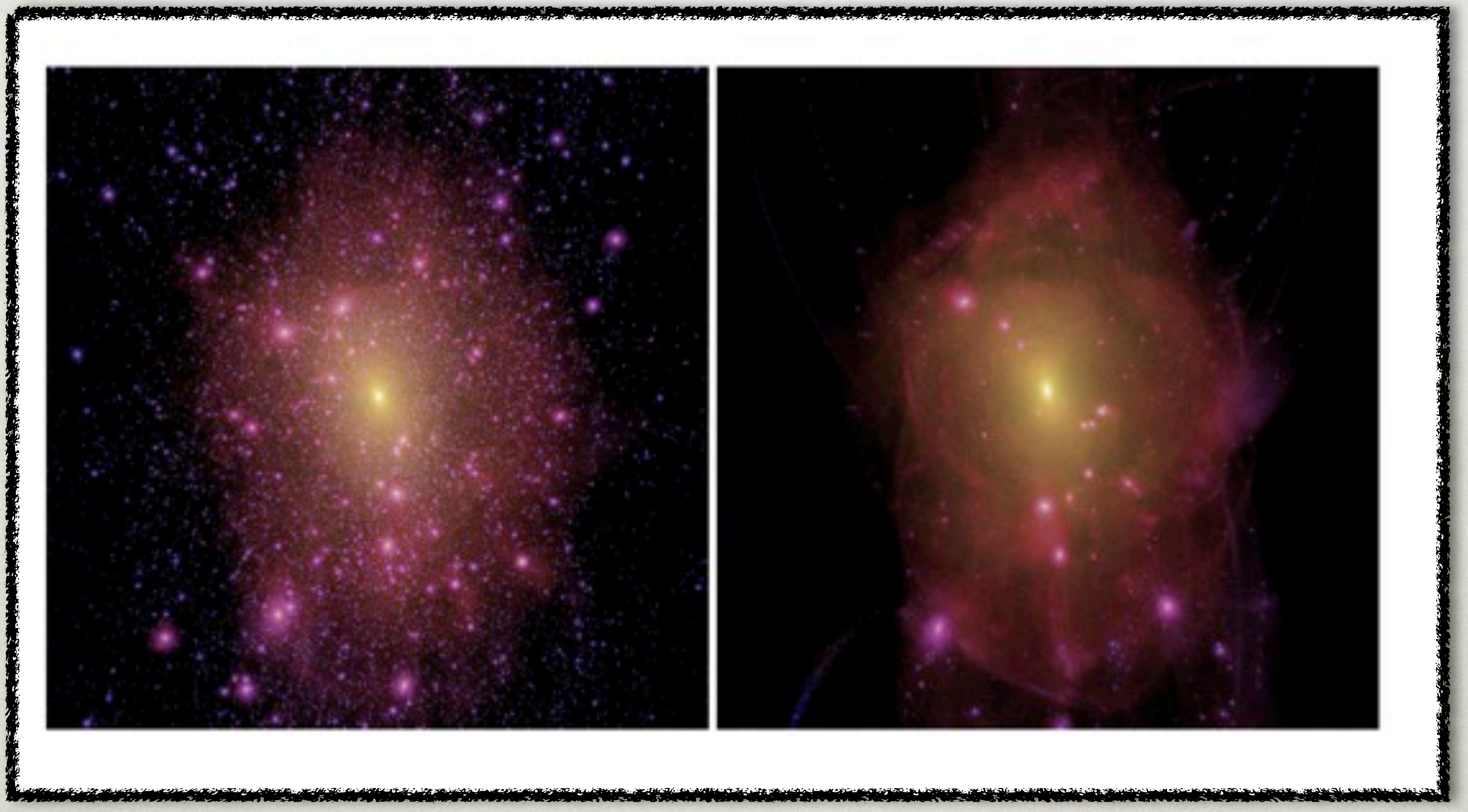
Model Comparison

$$P(\mathbf{L}, H|\mathbf{d}) \propto P(\mathbf{d}|\mathbf{L}, H)P(\mathbf{L}, H)$$

$$P(\mathbf{p}, \lambda|\mathbf{d}, \mathbf{L}, H) = \frac{P(\mathbf{d}|\mathbf{p}, \lambda, \mathbf{L}, H)P(\mathbf{p}, \lambda)}{P(\mathbf{d}|\mathbf{L}, H)}$$

$$P(\mathbf{d}|\mathbf{L}, H) = \int P(\mathbf{d}|\mathbf{p}, \lambda, \mathbf{L}, H)P(\mathbf{p}, \lambda)d\lambda d\mathbf{p}$$

CDM vs WDM



How do we probe the small scales beyond the Local Universe and independently from baryons?

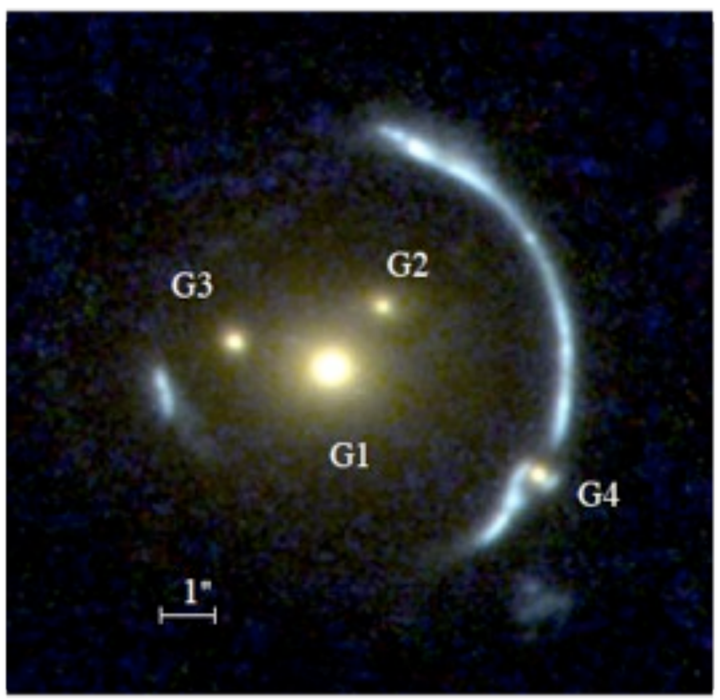
Using strong gravitational lensing!

- Independent of the baryonic content
- Independent of the dynamical state of the system
- Only way to probe small satellites at high redshift

Koopmans 2005

Vegetti & Koopmans 2009

Gravitational Imaging



$$\psi(\mathbf{x},\eta)_{tot} = \psi(\mathbf{x},\eta) + \delta\psi(\mathbf{x})$$

$\psi(\mathbf{x},\eta)$ Smooth analytic power-law model

$\delta\psi(\mathbf{x})$ pixellated potential correction

$$s(\vec{y}) = d(\vec{x}) \ \text{ with } \vec{y} = \vec{x} - \vec{\nabla}\psi(\vec{x})$$

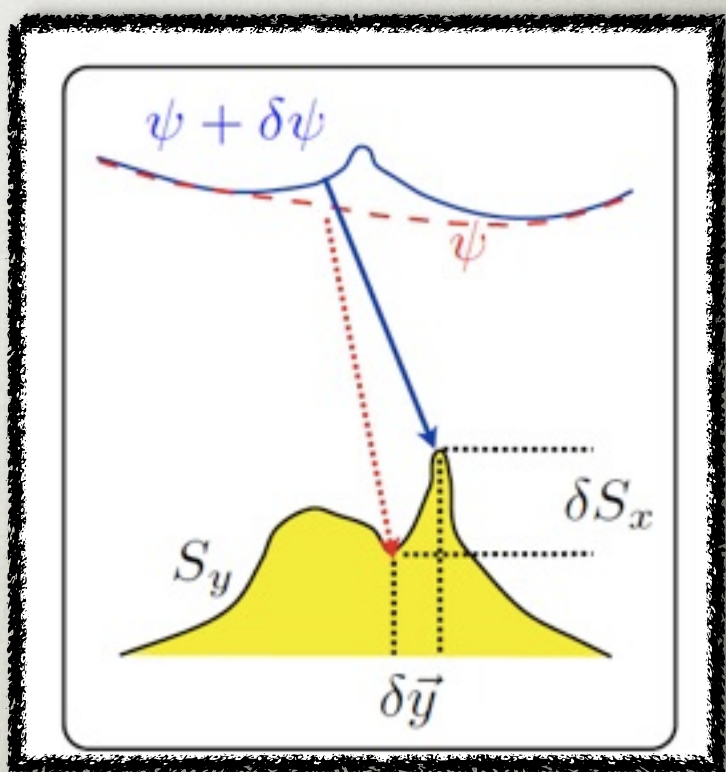
$$s(\vec{x} - \vec{\nabla}[\psi(\vec{x}) + \delta\psi(\vec{x})]) = d(\vec{x})$$
$$s(\vec{x} - \vec{\nabla}\psi(\vec{x})) = d(\vec{x}) + \delta d(\vec{x}).$$

$$\delta d(\vec{x}) = -\frac{\partial s(\vec{y})}{\partial \vec{y}} \cdot \frac{\partial \delta\psi(\vec{x})}{\partial \vec{x}} = -\vec{\nabla}_y s(\vec{y}) \cdot \vec{\nabla}_x \delta\psi(\vec{x}).$$

$$\delta\vec{y} = \delta\vec{\nabla}_x \psi(\vec{x}) = \vec{\nabla}_x \delta\psi(\vec{x})$$

$$\vec{\nabla}_y s(\vec{y}) \cdot \vec{\nabla}_x \delta\psi(\vec{x}) = \frac{\partial s(\vec{y})}{\partial \vec{y}} \cdot \delta\vec{y} \approx \delta s(\vec{y})$$

$$\delta\vec{s} = -\boldsymbol{D}_s(\vec{s})\, \boldsymbol{D}_x\, \delta\vec{\psi},$$



$$\overbrace{\boldsymbol{B}[\boldsymbol{L}(\psi)\,|-\boldsymbol{D_s}(\vec{\boldsymbol{s}})\;\boldsymbol{D_x}]}^{\equiv \boldsymbol{K}\ \text{(block matrix)}}\overbrace{\begin{pmatrix}\vec{s}\\ \delta\vec{\psi}\end{pmatrix}}^{\equiv\vec{r}}=\boldsymbol{K}\,\vec{r}=\vec{d}\,'$$

$$(\boldsymbol{K}^T\boldsymbol{C}^{-1}\boldsymbol{K}+\boldsymbol{R}^T\boldsymbol{R})\,\vec{r}=\boldsymbol{K}^T\boldsymbol{C}^{-1}\vec{d}\,,$$

$$\boldsymbol{R}^T\boldsymbol{R}\equiv\begin{pmatrix}\lambda_s\boldsymbol{H}_s^T\boldsymbol{H}_s & \boldsymbol{0}\\ \boldsymbol{0} & \lambda_{\delta\psi}\boldsymbol{H}_{\delta\psi}^T\boldsymbol{H}_{\delta\psi}\end{pmatrix},$$

In Bayesian terms

$$\overbrace{P(\boldsymbol{\delta\psi}|\boldsymbol{d},\mathbf{f},\boldsymbol{s},\mathbf{t},\mu,\mathbf{g}_{\delta\psi})}^{\text{posterior}}=\frac{\overbrace{P(\boldsymbol{d}|\boldsymbol{\delta\psi},\mathbf{t},\mathbf{f},\boldsymbol{s})}^{\text{likelihood}}\;\overbrace{P(\boldsymbol{\delta\psi}|\mu,\mathbf{g}_{\delta\psi})}^{\text{prior}}}{\underbrace{P(\boldsymbol{d}|\mathbf{f},\boldsymbol{s},\mathbf{t},\mu,\mathbf{g}_{\delta\psi})}_{\text{evidence}}},$$

$$\overbrace{P(\delta\boldsymbol{\psi}|\boldsymbol{d}, \mathbf{f}, s, \mathbf{t}, \mu, \mathbf{g}_{\delta\psi})}^{\text{posterior}} = \frac{\overbrace{P(\boldsymbol{d}|\delta\boldsymbol{\psi}, \mathbf{t}, \mathbf{f}, s)}^{\text{likelihood}}\,\overbrace{P(\delta\boldsymbol{\psi}|\mu, \mathbf{g}_{\delta\psi})}^{\text{prior}}}{\underbrace{P(\boldsymbol{d}|\mathbf{f}, s, \mathbf{t}, \mu, \mathbf{g}_{\delta\psi})}_{\text{evidence}}},$$

$$P(\boldsymbol{d}|\delta\boldsymbol{\psi}, \mathbf{t}, \mathbf{f}, s) = \frac{\exp(-E_{\mathrm{D}}(\boldsymbol{d}|\delta\boldsymbol{\psi}, \mathbf{t}, \mathbf{f}, s))}{Z_{\mathrm{D}}},$$

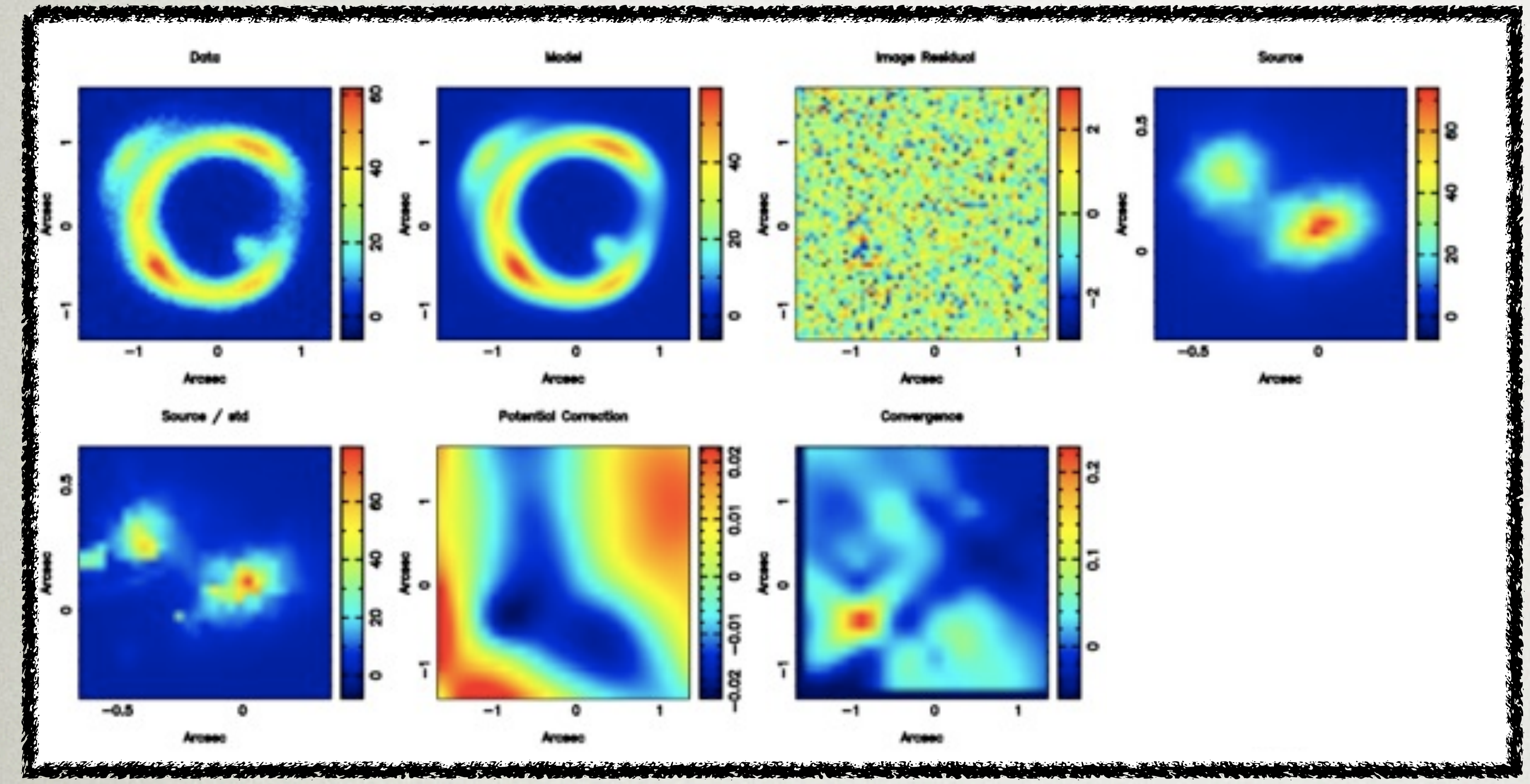
where

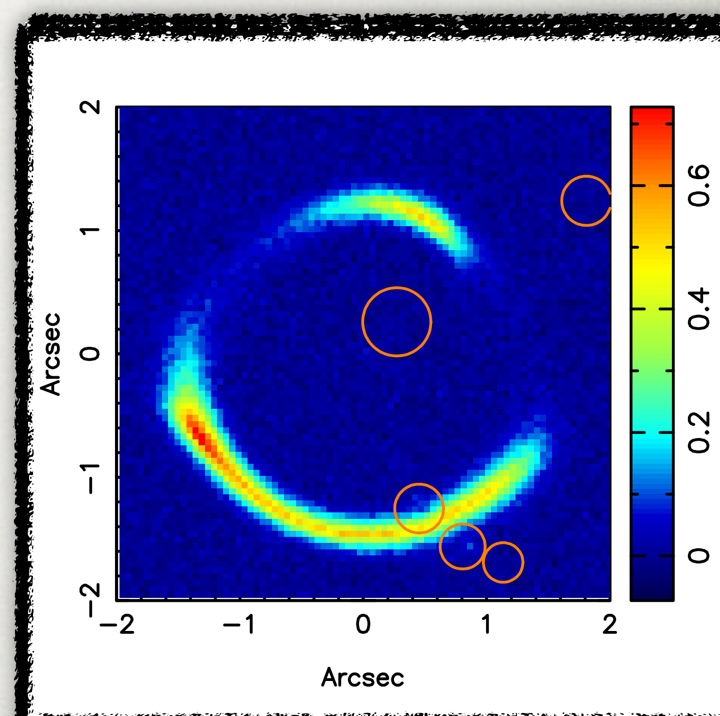
$$E_{\mathrm{D}}(\boldsymbol{d}|\delta\boldsymbol{\psi}, \mathbf{t}, \mathbf{f}, s) = \frac{1}{2}(\boldsymbol{d} - \mathbf{f}s - \mathbf{t}\delta\boldsymbol{\psi})^{\mathrm{T}}\mathbf{C}_{\mathrm{D}}^{-1}(\boldsymbol{d} - \mathbf{f}s - \mathbf{t}\delta\boldsymbol{\psi})$$

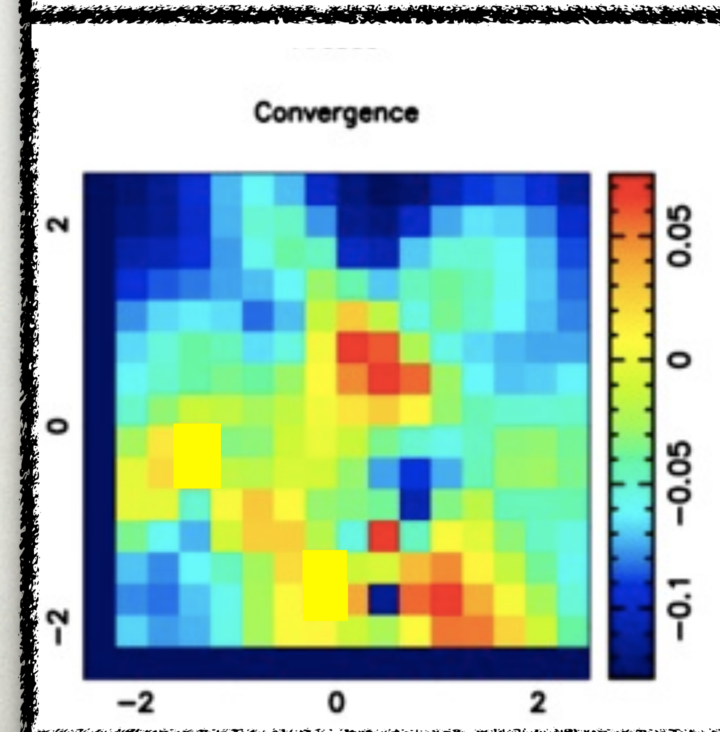
$$= \frac{1}{2}\chi^2,$$

$$P(\delta\boldsymbol{\psi}|\mu, \mathbf{g}_{\delta\psi}) = \frac{\exp(-\mu E_{\delta\psi}(\delta\boldsymbol{\psi}|\mathbf{g}_{\delta\psi}))}{Z_{\delta\psi}(\mu)}$$

Gravitational Imaging







- substructures are responsible of localised surface brightness perturbations and are detected as localised potential corrections

- Any substructure can be detected provided it is mass enough and/or close enough to the Einstein ring

- For each substructure detected its mass can be measured by assuming a mass model or directly from the pixelated corrections in a model independent way

Vegetti & Koopmans 2009

Statistics of detections

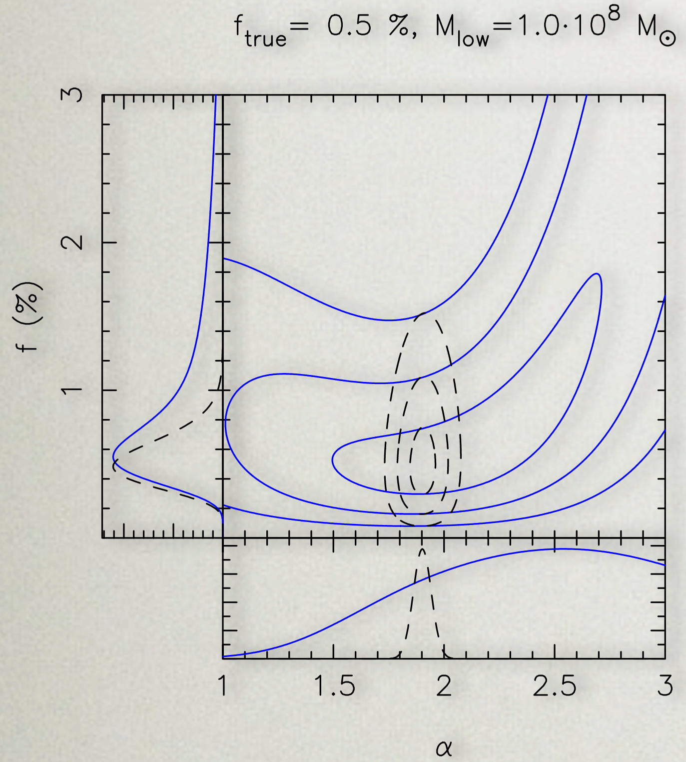
$$P\,(\alpha, f \mid \{n_s, \mathbf{m}\}, \mathbf{p}) = \frac{\mathcal{L}\,(\{n_s, \mathbf{m}\} \mid \alpha, f, \mathbf{p})\, P\,(\alpha, f \mid \mathbf{p})}{P\,(\{n_s, \mathbf{m}\} \mid \mathbf{p})}$$

$$L\,(\{n_s, \mathbf{m_s}, \mathbf{R_s}\} \mid \alpha, f(< R), \mathbf{p}) = \frac{e^{-\mu(\alpha, f, \mathbf{p})}\;\; \mu(\alpha, f, \mathbf{p})^{n_s}}{n_s!} \prod_{k=1}^{n_s} P\,(m_k, R_k \mid \mathbf{p}, \alpha)$$

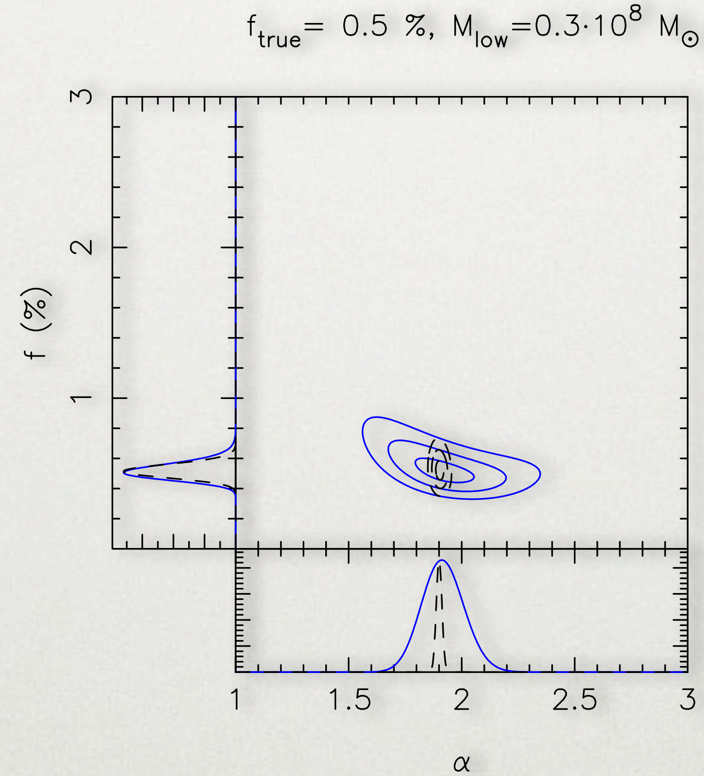
$$\mu_j(\alpha, f, \mathbf{p}) = \mu_{0,j}(\alpha, f, \mathbf{p}) \int_{M_{\mathrm{low,j}}}^{M_{\mathrm{max}}} P\,(m, R_j \mid \mathbf{p}, \alpha)\, dm$$

$$= \mu_{0,j}(\alpha, f, \mathbf{p}) \int_{M_{\mathrm{low,j}}}^{M_{\mathrm{max}}} \frac{dP}{dm}\; dm$$

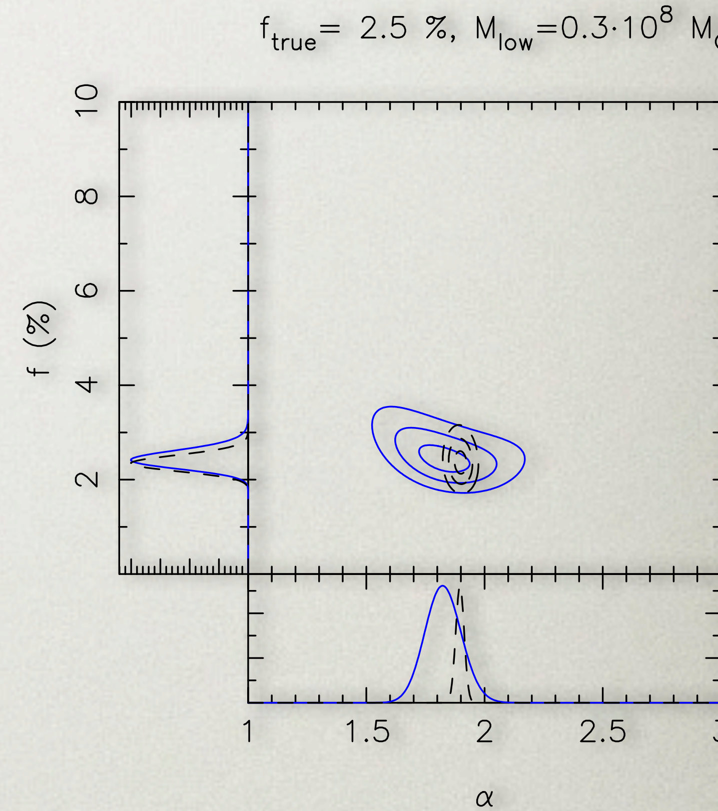
Statistics of detections



10 not-very sensitive lenses cannot constrain the slope of the mass function

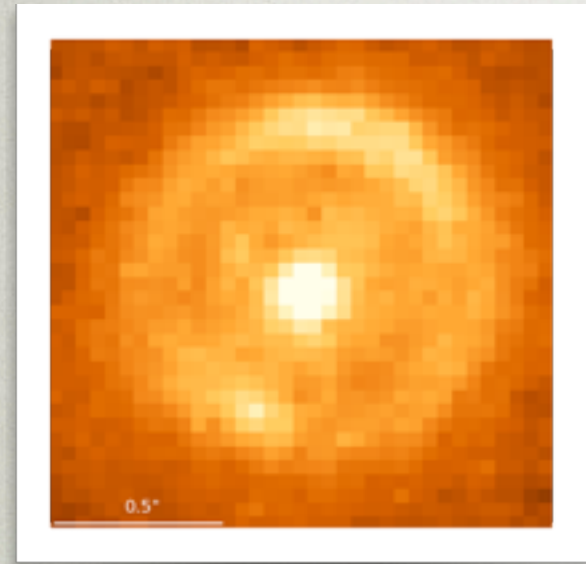


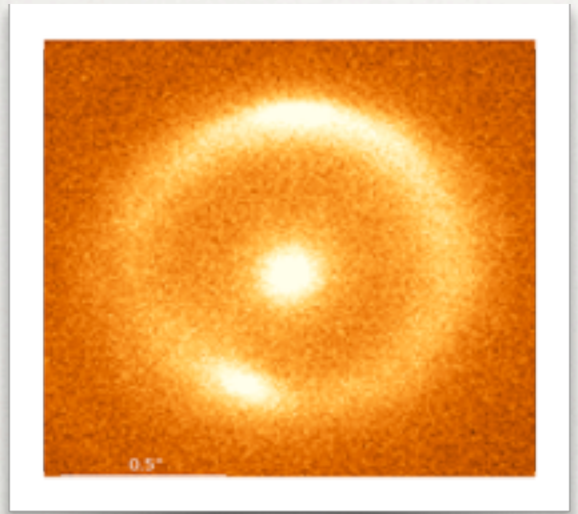
200 very sensitive lenses can constrain the mass function at the few percent level

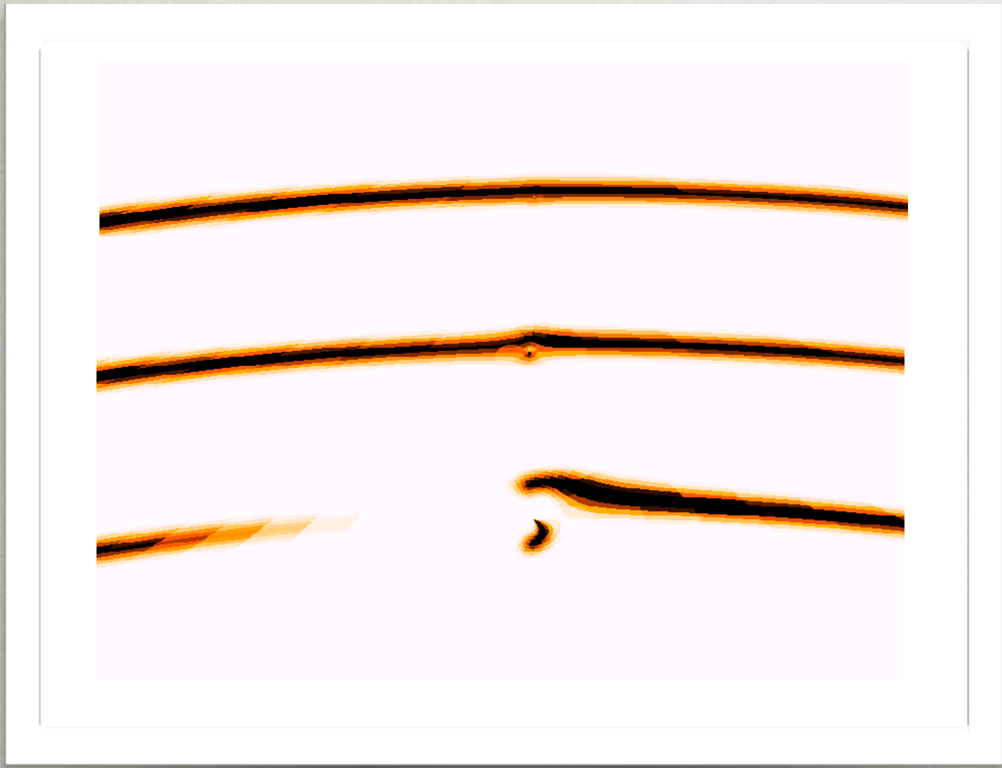


but 10 may be just enough

Sensitivity Function



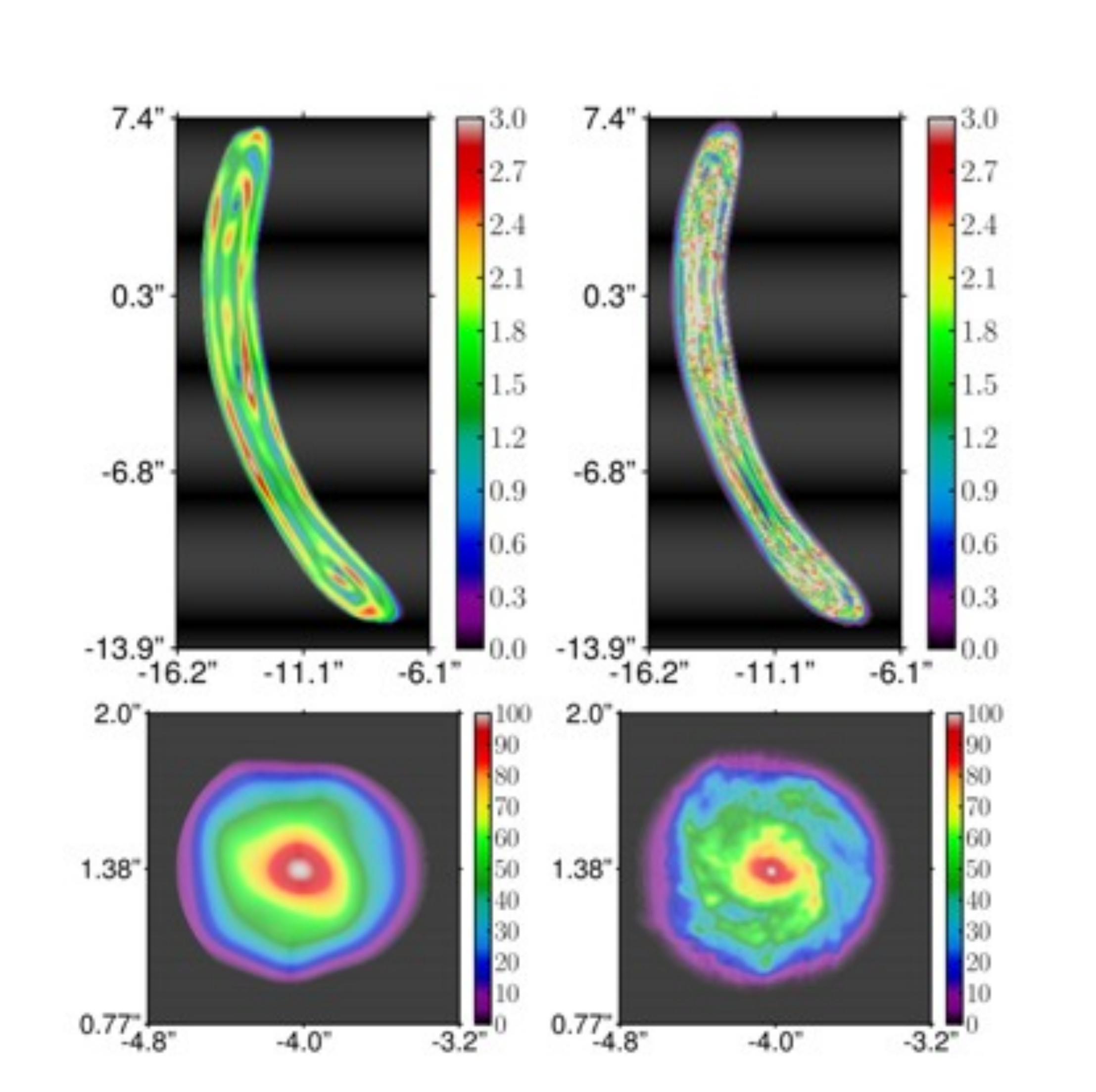




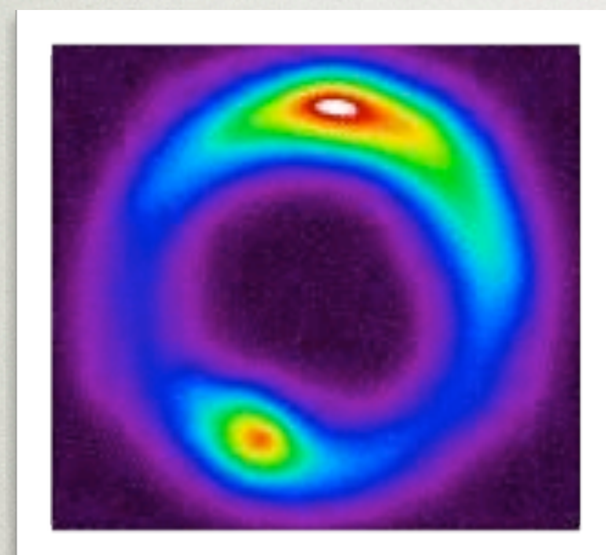
$$\delta\theta \approx \mu \, \theta_E$$

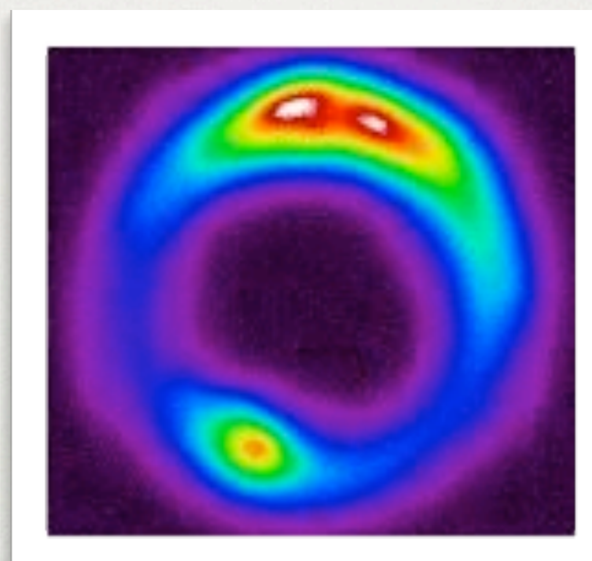
Sensitivity Function

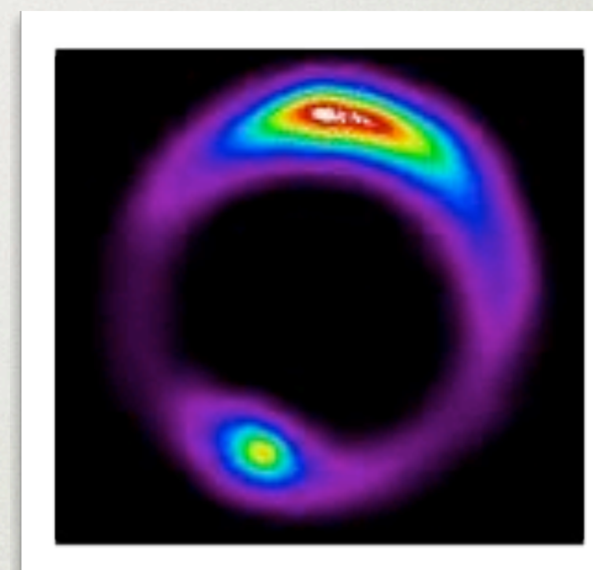
$$\delta I \approx \nabla S \cdot \nabla \delta\psi = \nabla S \cdot \alpha_{sub}$$



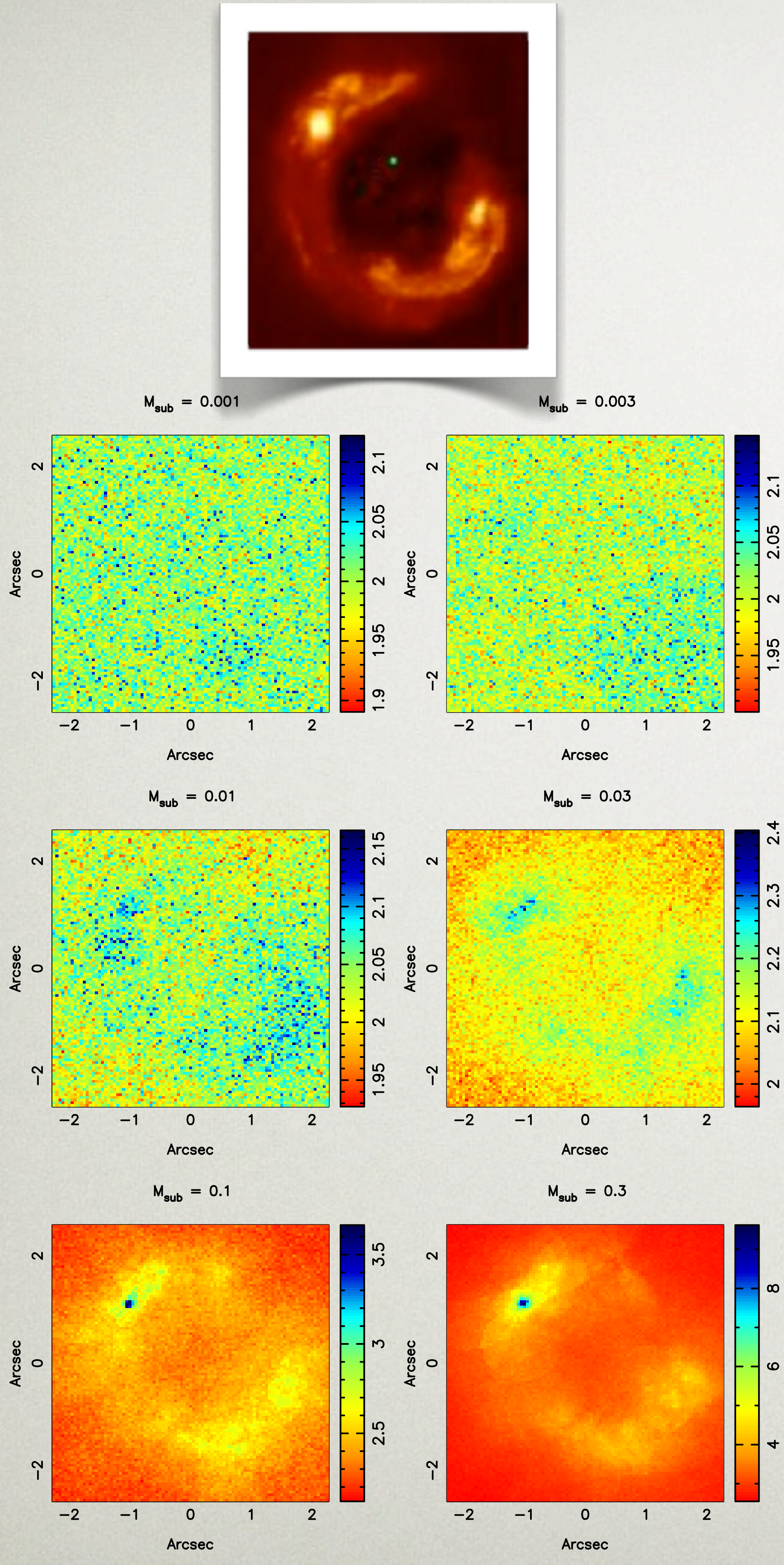
Sensitivity Function



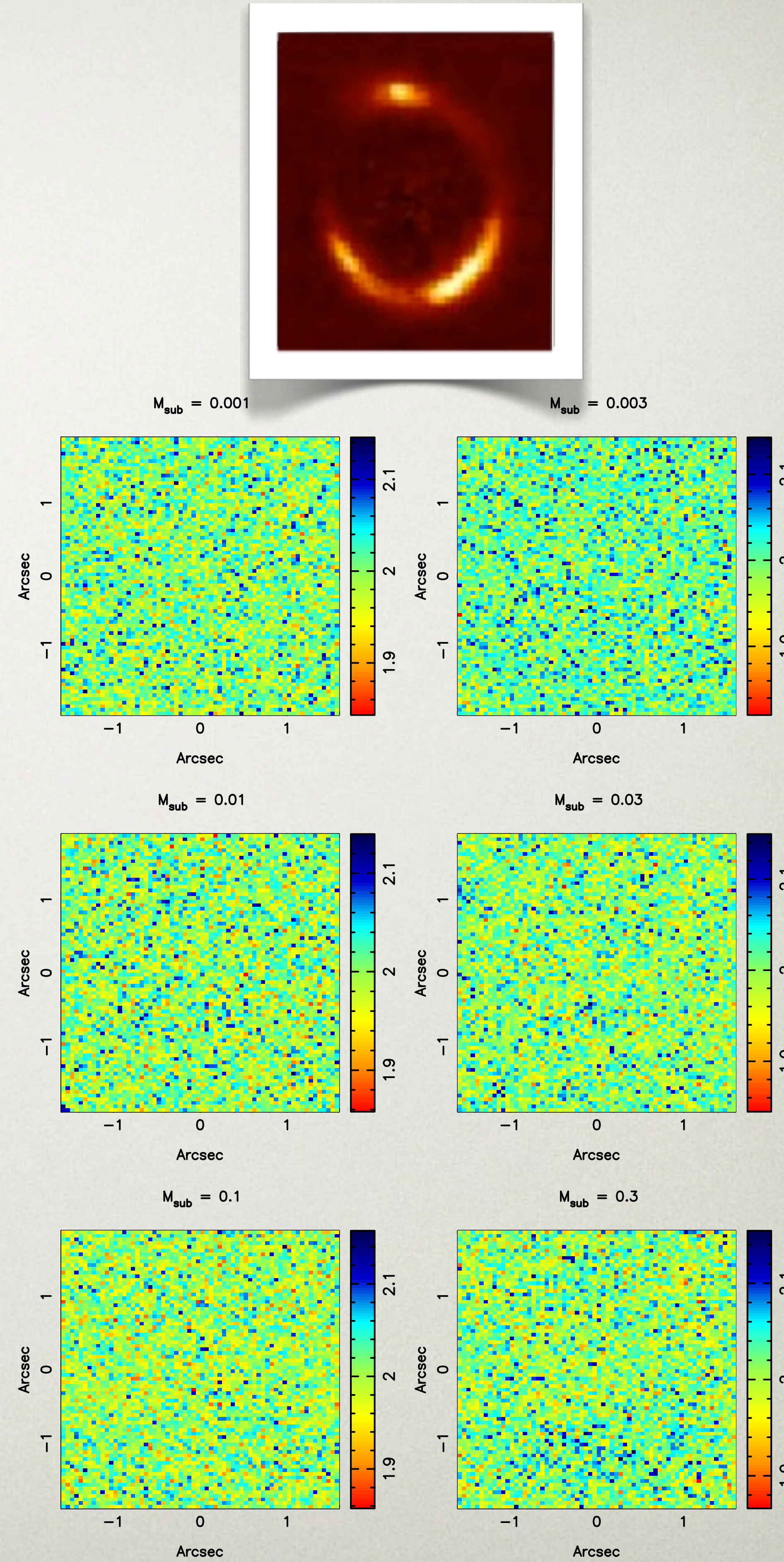




$$L = \sum_{i}^{n_{pix}} \left(\frac{\delta I}{\sigma} \right)^2$$



M_sub = 0.001
M_sub = 0.003
M_sub = 0.01
M_sub = 0.03
M_sub = 0.1
M_sub = 0.3
Arcsec



M_sub = 0.001
M_sub = 0.003
M_sub = 0.01
M_sub = 0.03
M_sub = 0.1
M_sub = 0.3
Arcsec

Conclusions

- I developed a fully-Bayesian adaptive lens modeling technique for the analysis of extended sources
- Sources are regularizes to unsure smoothness and avoid noise fitting
- The regularization level is a free parameter of the model and the regularization form depends on the source structure
- The method allows for regularized linear and local potential corrections, that could be the signature of mass substructure
- I developed a statistical interpretation to turns substructure detections into constraints on the substructure population
- The sensitivity to substructure depends on the substructure mass, the data angular resolution and S/N, the source structure
- Computing the sensitivity function properly is computationally very expensive and better way is need