GIGA-Lens:

A Fast Differentiable Bayesian Lens Modeling Framework

Andi Gu
UC Berkeley Department of Physics, LBNL Physics Division

Collaborators

X. Huang (Univ. San Francisco, LBNL Physics Division)

D. Schlegel (LNBL Physics Division)

G. Aldering (LBNL Physics Division)

S. Perlmutter (LBNL Physics Division)

W. Sheu (UC Berkeley Department of Physics)

A. Bolton (NSF NOIRLab)

A. Dey (NSF NOIRLab)

E. Schlafly (LLNL)

K. Boone (University of Washington DiRAC Institute)

A. Fillip (Max Planck Institute for Astrophysics)

Y. Shu (Max Planck Institute for Astrophysics)

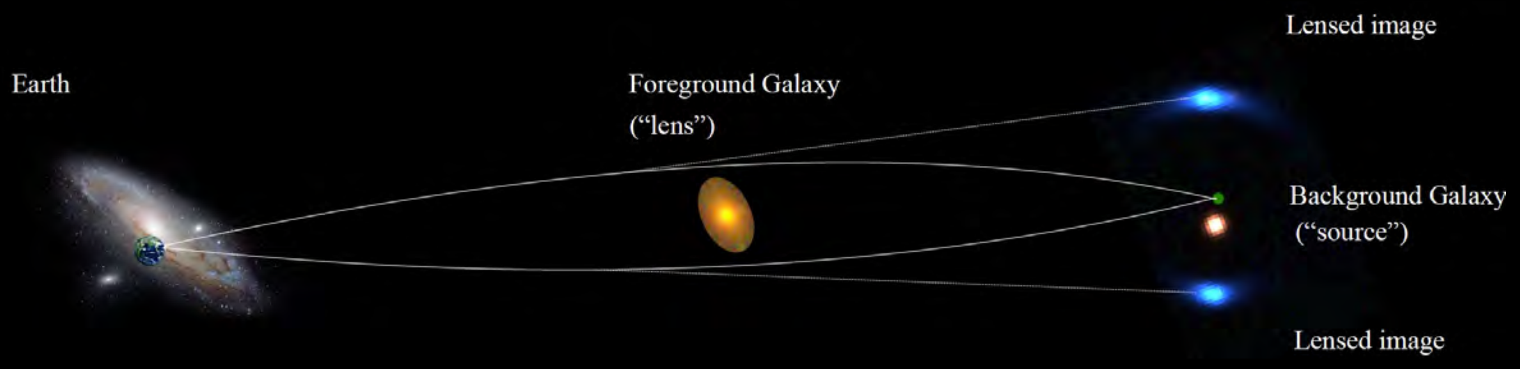
S. Suyu (Max Planck Institute for Astrophysics)

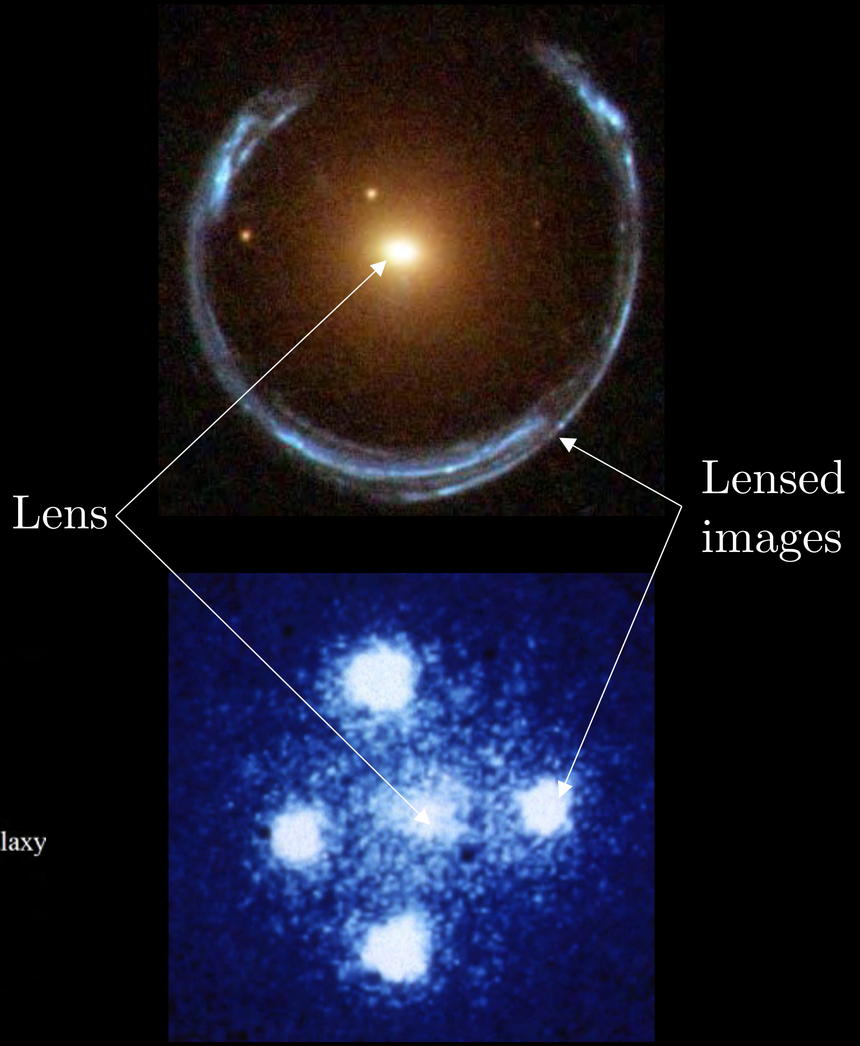
E. Jullo (The French National Centre for Scientific Research)

D. Rubin (University of Hawaii Department of Physics & Astronomy)

Strong Gravitational Lensing

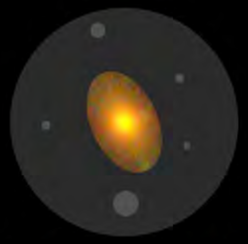
- Chance alignment (1 in 10000)
- Warping of space-time by the mass of the foreground galaxy
- Large arcs & multiple images



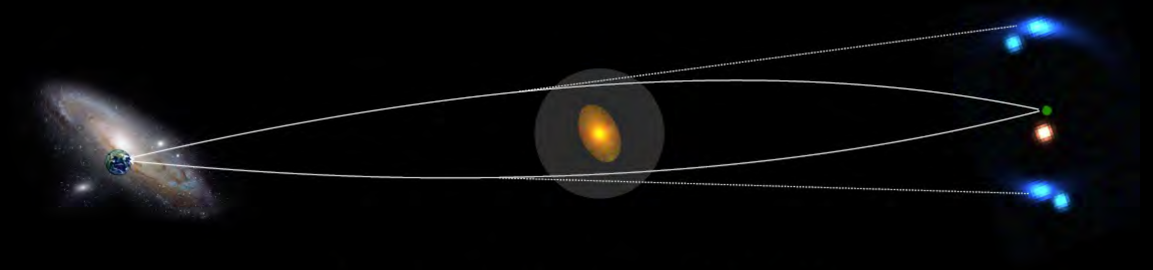


Cosmology with Strong Lenses

1. Test predictions of cold dark matter model (Nadler+ 2021)
2. Infer Hubble constant, H_0 (Wong+ 2019, Birrer 2021)
3. Dark energy (e.g., Sharma & Linder 2022)



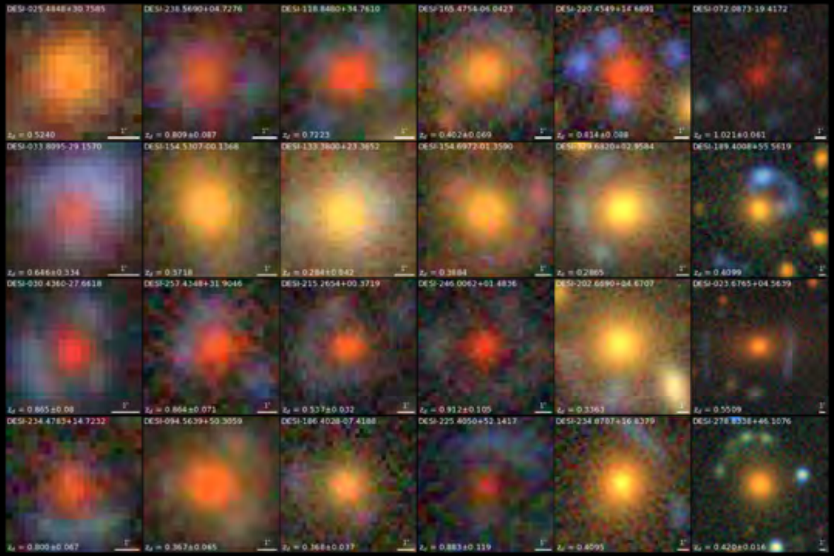
Sub-galactic dark matter halo



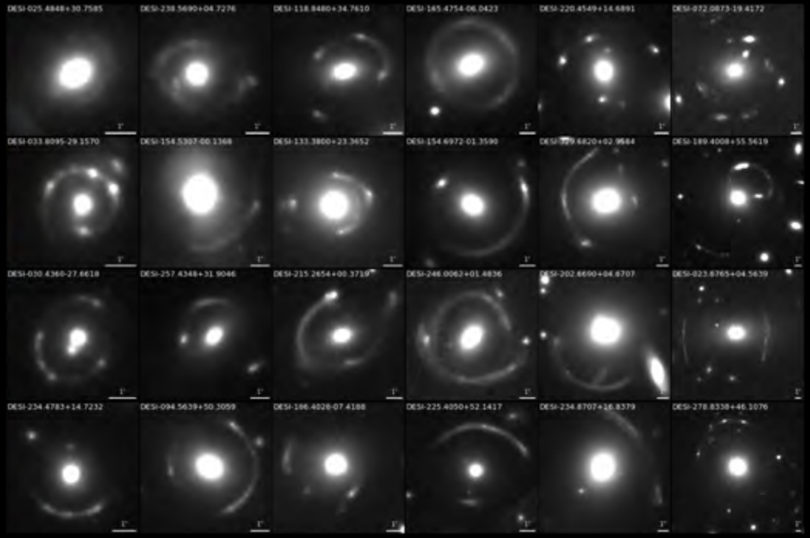
Supernova appears to explode twice:
Time Delay ~ $1/H_0$

The Dark Energy Spectroscopic Instrument (DESI)

- DESI Legacy Imaging Surveys (Dey, Schlegel+ 2019) – 14,000 deg^2
- 30 million elliptical galaxies
- We found >1500 strong lens candidates (Huang+ 2020, 2021) using ResNets
- 51 of our candidates have been observed by the HST – all confirmed to be actual lenses
- Many more systems confirmed by the ongoing DESI experiment



DESI Legacy Surveys



Hubble Space Telescope

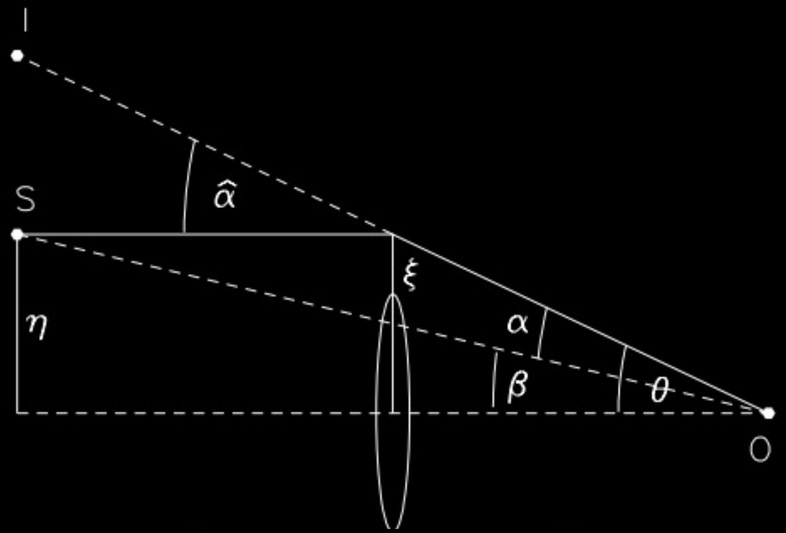
Gravitational lensing

- Fundamental "lensing equation":

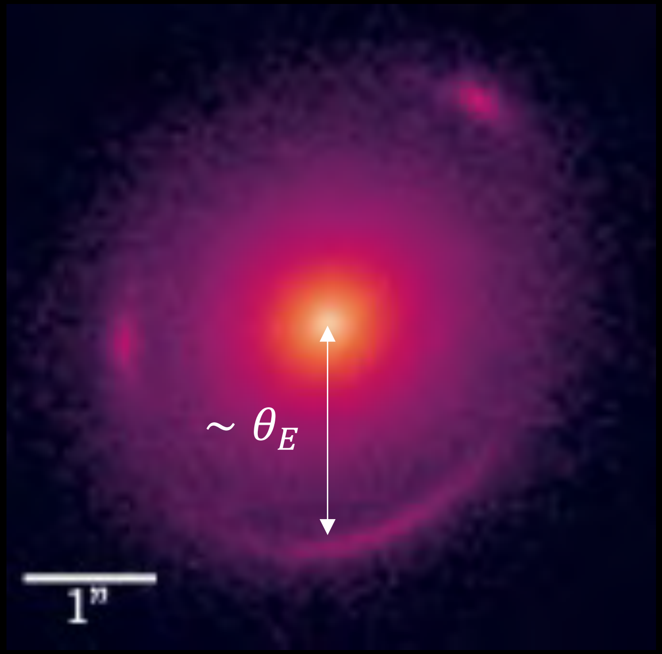
$$\beta = \theta - \alpha(\theta)$$

$$\alpha(\theta) = \theta_E \quad \text{(singular isothermal sphere)}$$

- We know how to simulate: if we know the source and lens, we can compute the image. This is the forward problem.



Narayan & Bartelmann 1996



Our Goal

- We have the inverse problem. What can we learn about:
  - Light profile of the source galaxy
  - Distribution of matter in the lensing galaxy
  - Sub-galactic dark matter halos near the lens (subhalos) or along the line of sight
- In the language of statistics, we have an *inference* problem

$$P(\Theta \mid X) = \frac{P(X \mid \Theta) \cdot P(\Theta)}{P(X)}$$

Strategy

1. Write down a parameterized model for the:
   - Source light
   - Lens mass
   - Lens light (nuisance parameters)
2. Define a prior $p(\Theta)$ for the parameters Θ and a likelihood $p(X \mid \Theta)$
   - Simulate the expected outcome if Θ are the true parameters, and compare with observed image X

$$\log p(X \mid \Theta) = -\sum \frac{\left(X_{obs} - X_{sim}(\Theta)\right)^2}{2\sigma^2} + C$$

3. Sample from the posterior $p(\Theta \mid X)$

How do we sample?

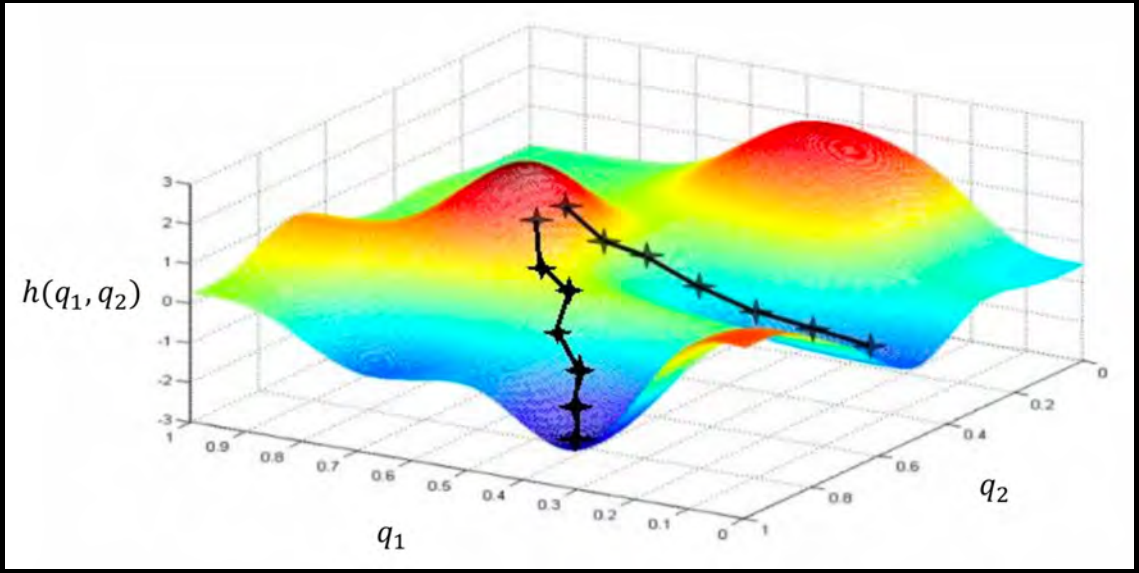
Most regions of the parameter space have vanishing likelihood:

<u>sampling is hard</u>

Look for the global maximum of the posterior density first

The parameter space is highly non-convex:

<u>search is expensive</u>



Desideratum	Solution
Fast simulation	Lens models are linear algebraic: **use GPUs**
A guide through high dimensional parameter spaces	Lens models are differentiable: **use the gradient** (with autodiff)
Robust	Use many candidate solutions: **use parallel computation**

Enter: GIGA-Lens

- We developed a gradient-informed, GPU-accelerated lens (GIGA-Lens) modeling framework
- Written in TensorFlow and JAX
  - Fully vectorized
  - Fully differentiable via autodiff
  - Distributed compute (multi-GPU)

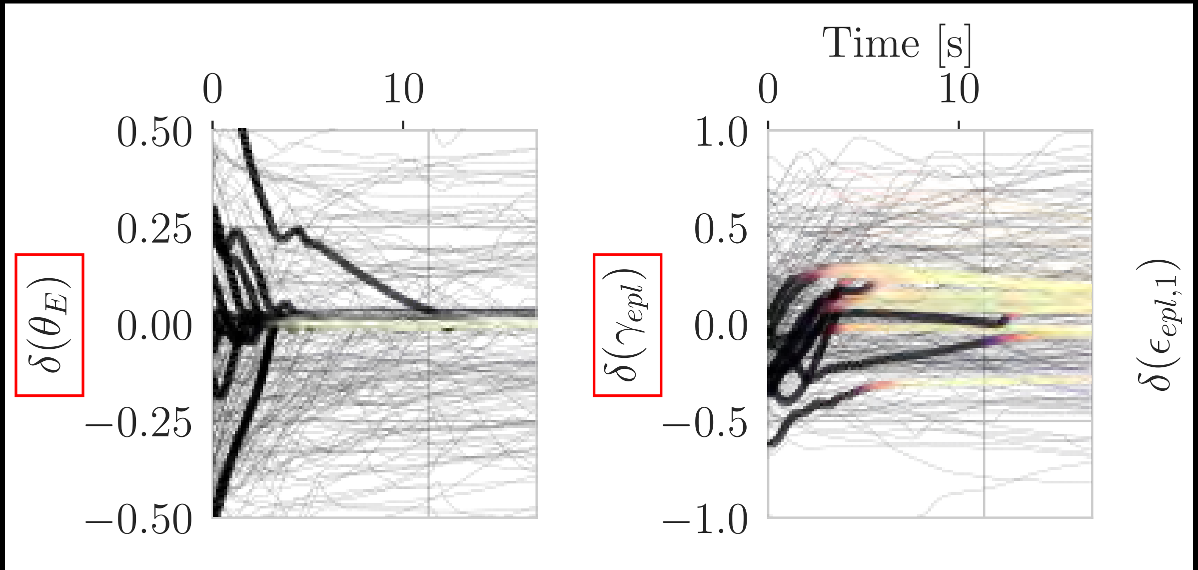
Modelling Procedure

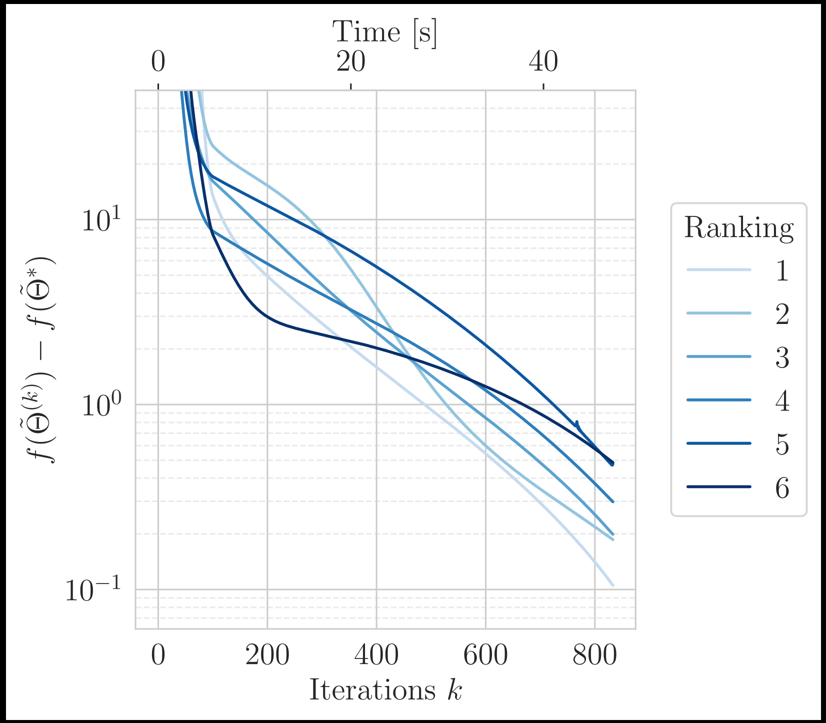
I. Multi-Starts Gradient Descent

Goal: identify global maximum of posterior density

1. Initialize many candidate solutions by sampling from the prior
2. Use a bijector to unconstrain parameters $\theta_E \to \log \theta_E$, or $\gamma \to S^{-1}(\gamma)$
3. Locally optimize with gradient descent the posterior density of each solution
4. Select the maximum a posteriori (MAP) estimate

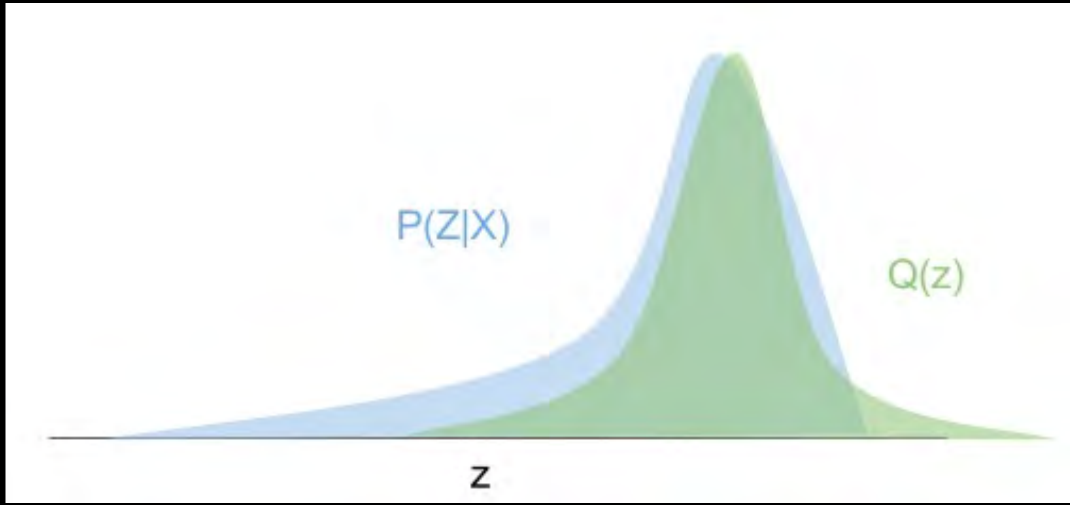
I. Multi-Starts Gradient Descent





II. Covariance Estimation

- Use variational inference (Hoffman 2013), with a surrogate $Q(\Theta; \mu, \Sigma) \sim \mathcal{N}(\mu, \Sigma)$



$$min_{\{\mu,\Sigma\}} KL(Q(\Theta; \mu, \Sigma) \,||\, p(\,\Theta \mid X\,))$$

II. Covariance Estimation

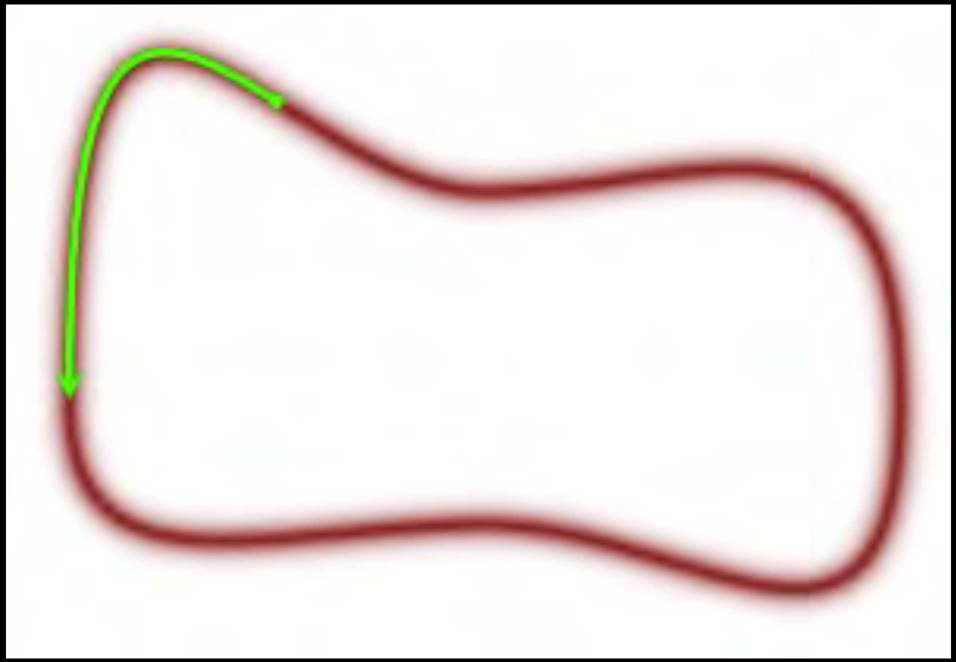
The KL can be written as an expectation known as evidence lower bound ('ELBO')

$$ELBO(\mu, \Sigma) = E_{\Theta \sim \mathcal{N}(\mu,\Sigma)}[\log Q(\Theta; \mu, \Sigma) - \log p(\Theta \mid X)]$$
$$\nabla ELBO = E_{\Theta}[(\log Q(\Theta; \mu, \Sigma) - \log p(\Theta \mid X))\, \nabla(\log Q(\Theta; \mu, \Sigma))]$$

1. Initialize μ at the MAP and set a small (10^{-6}) diagonal covariance.
2. Minimize the ELBO with stochastic gradient descent.
3. The optimized Σ is an estimate for the true covariance

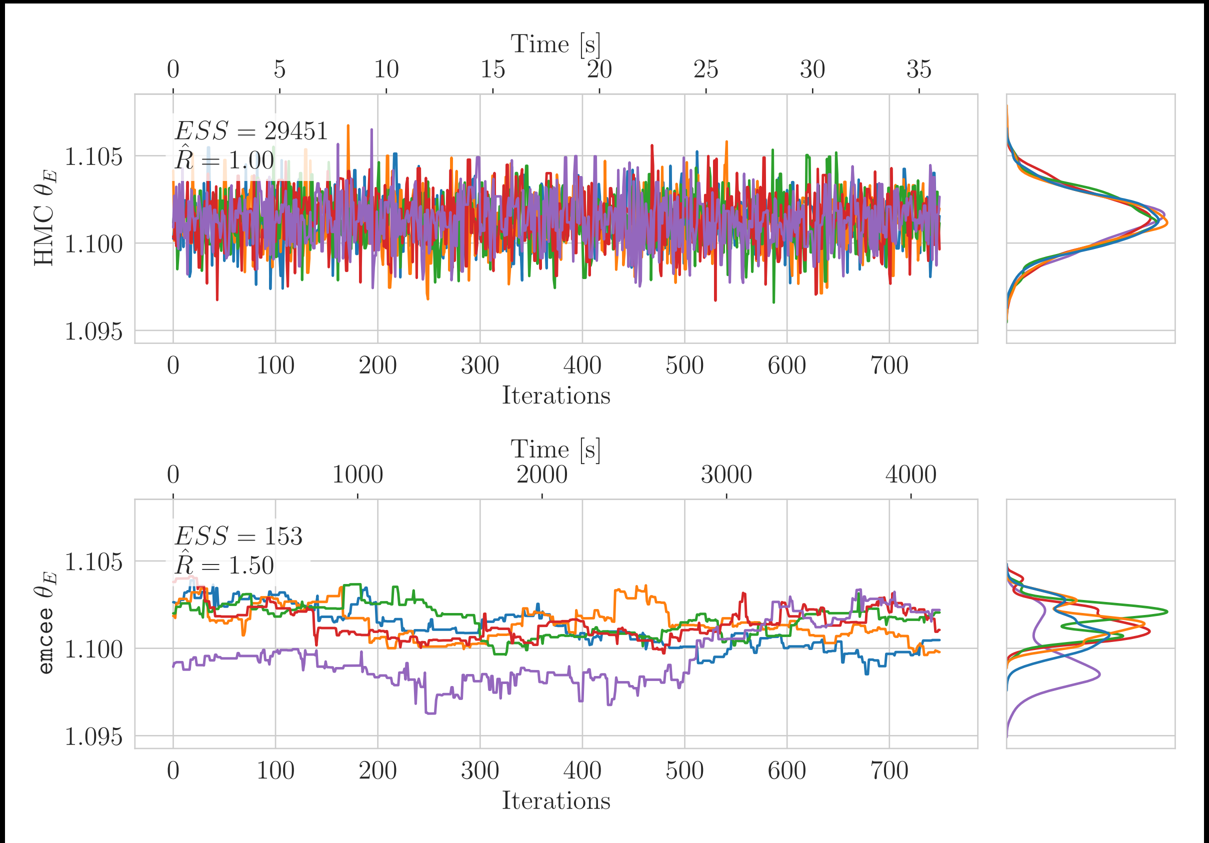
III. Hamiltonian Monte Carlo

- We take advantage of gradient information for Hamiltonian Monte Carlo (Betancourt 2021)
- We initialize HMC with samples drawn from the VI posterior, and the mass matrix $M = \Sigma^{-1}$
- Other parameters (step size ϵ, length L) adaptively tuned

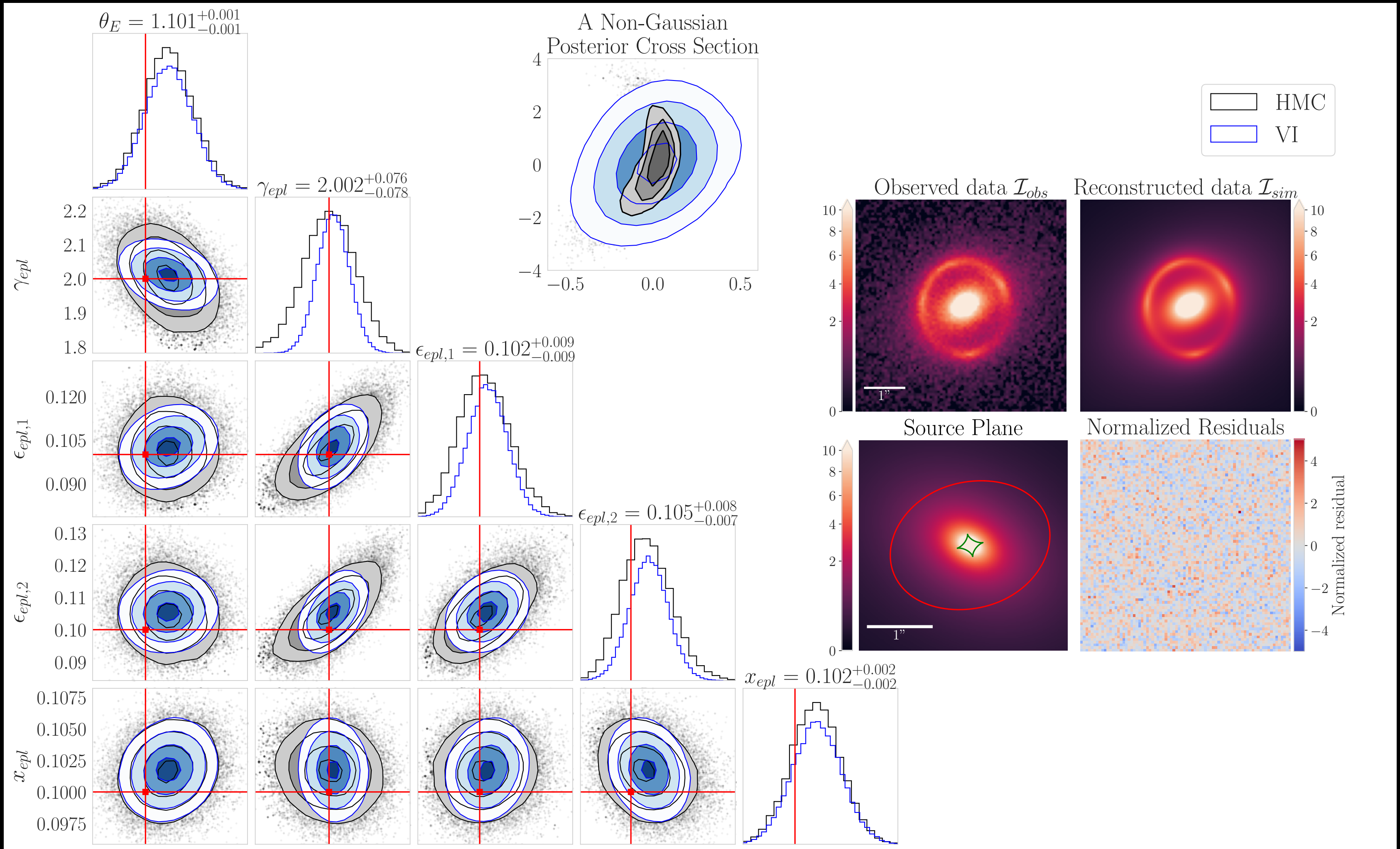


Betancourt 2021

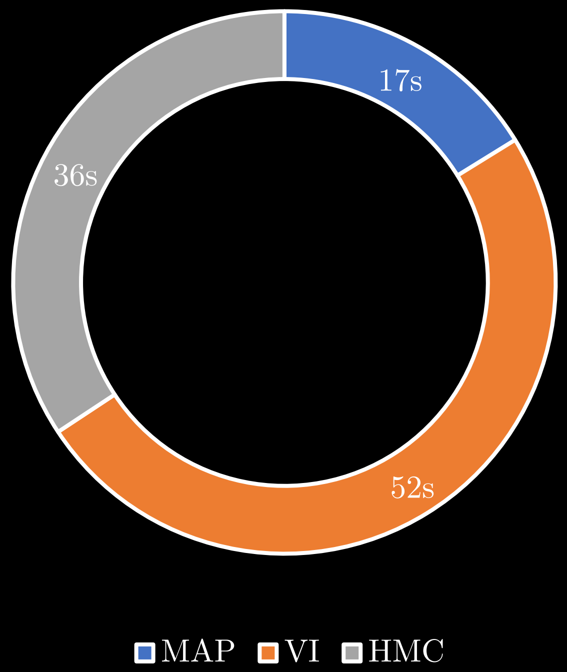
III. Hamiltonian Monte Carlo

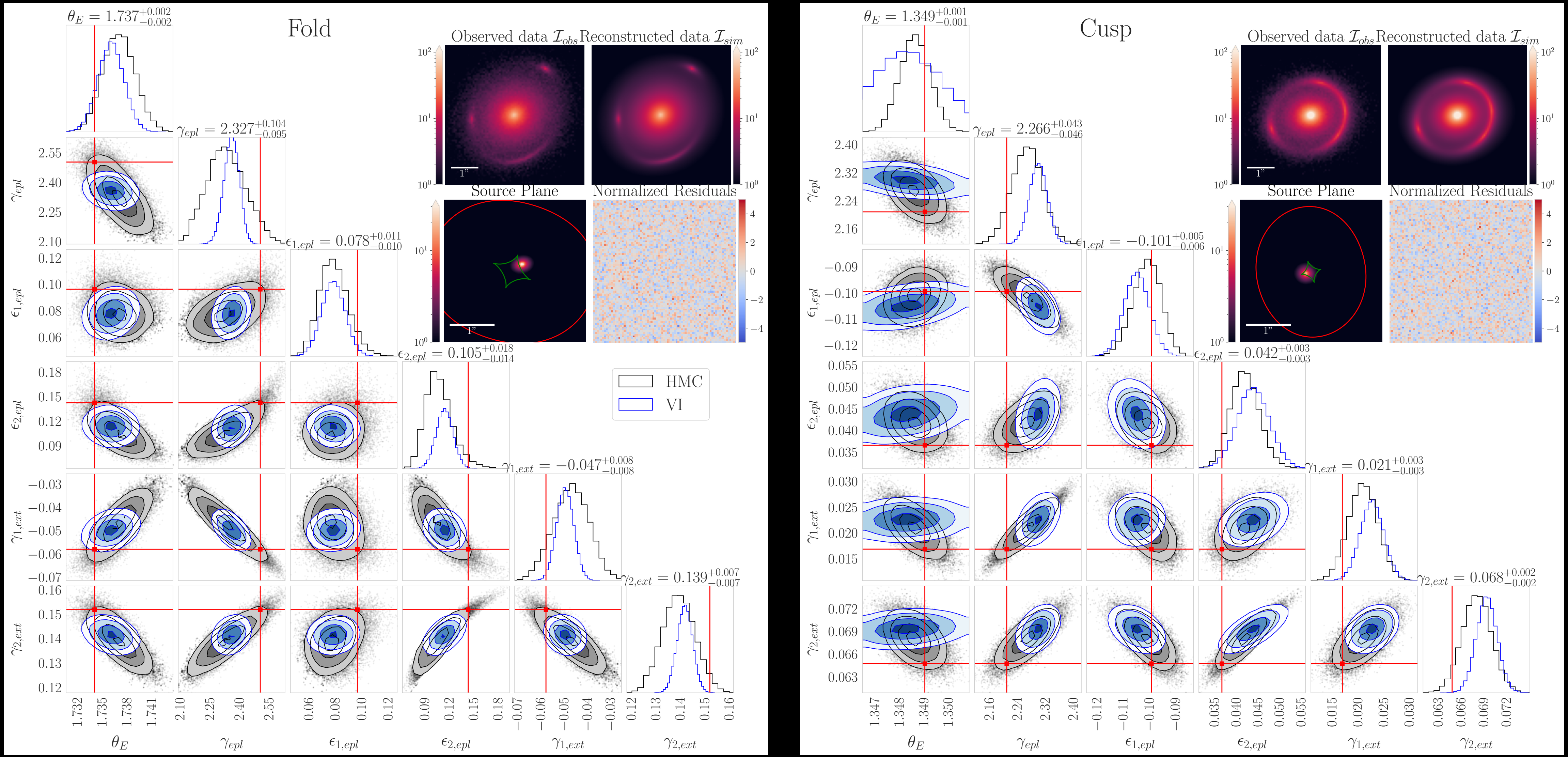


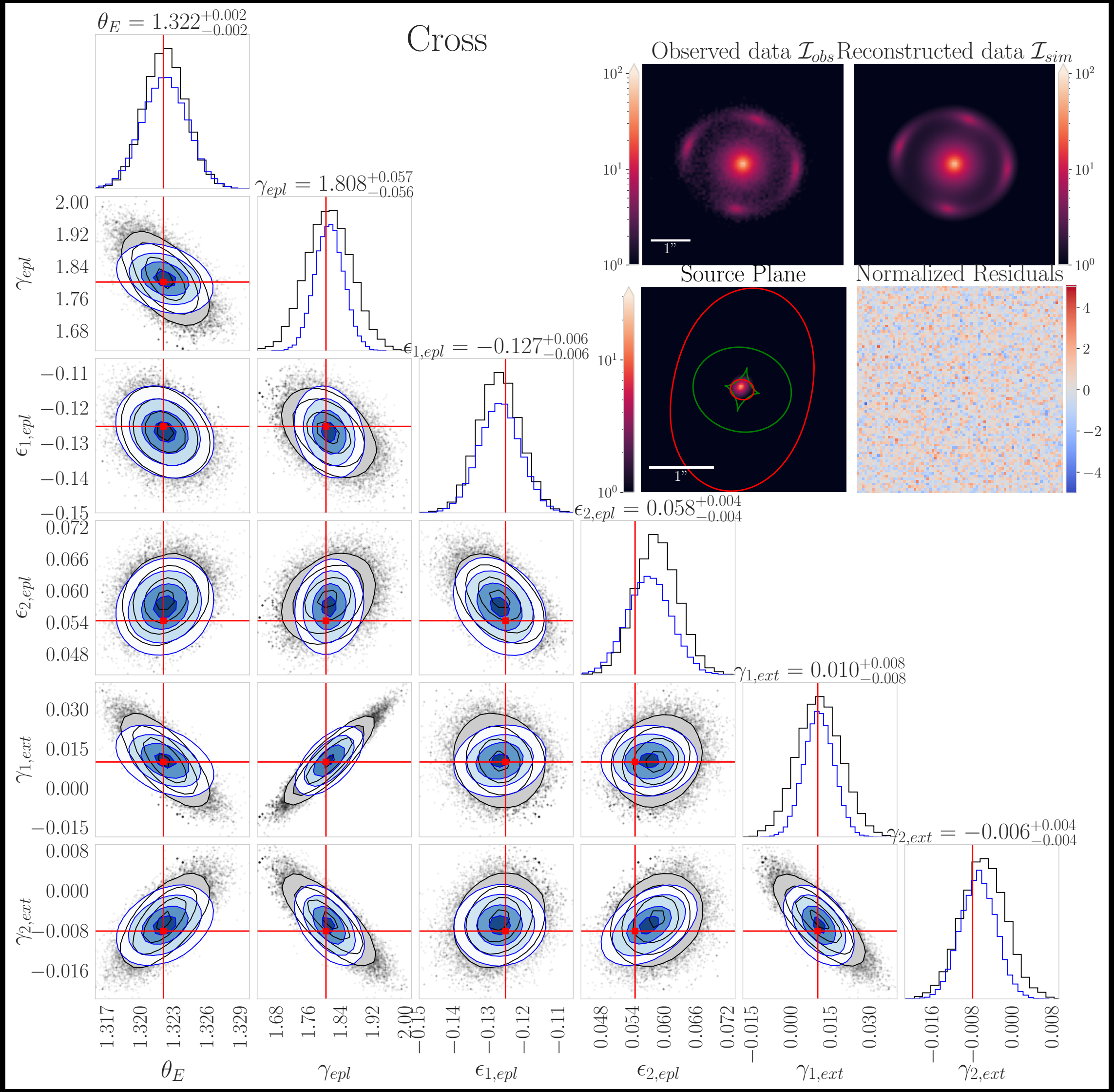
Modeling Results on a Simulated System



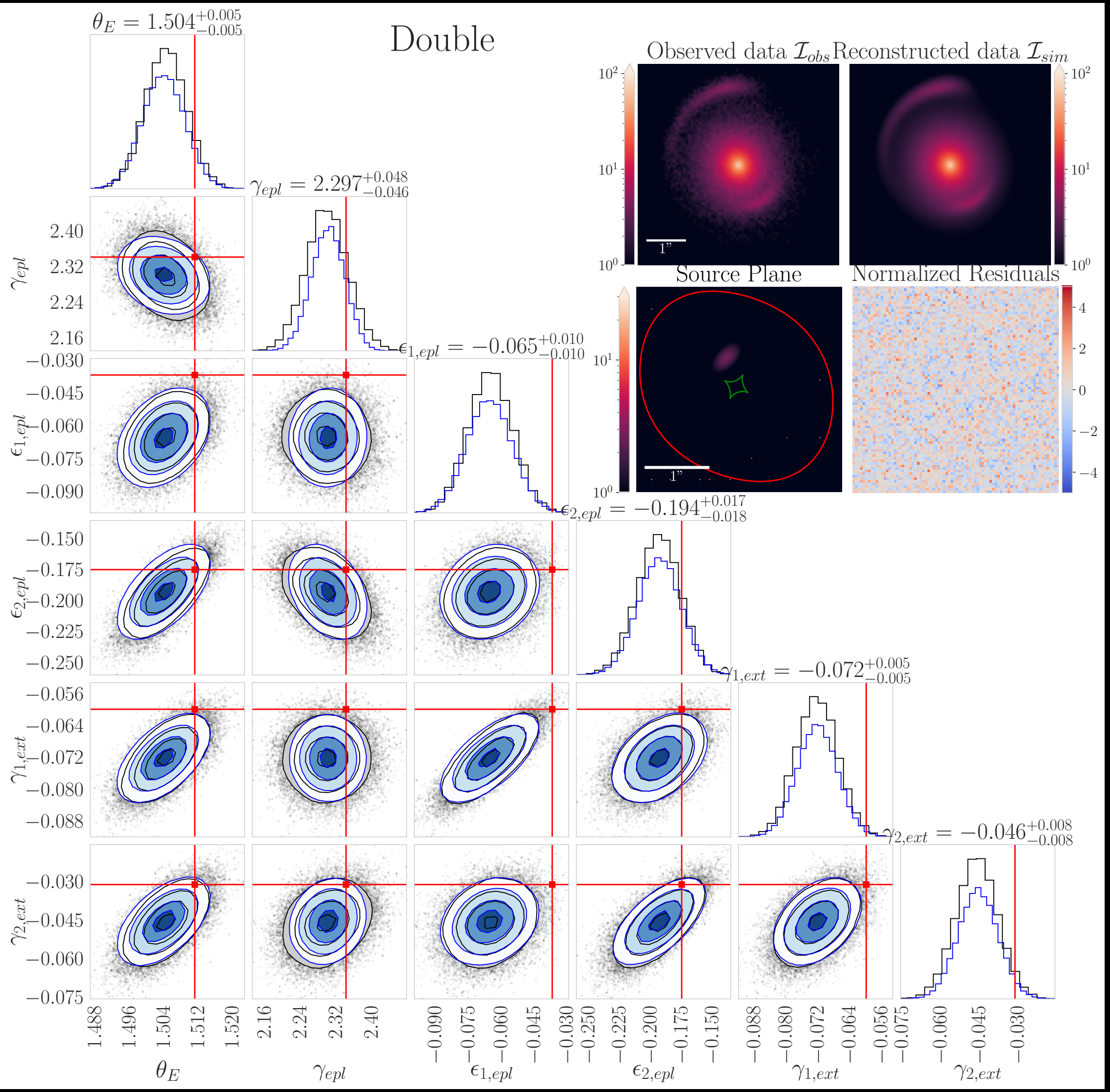
Execution time on 4 A100s (Perlmutter early access)




Fold
$\theta_E = 1.737^{+0.002}_{-0.002}$
$\gamma_{epl} = 2.327^{+0.104}_{-0.095}$
$\epsilon_{1,epl} = 0.078^{+0.011}_{-0.010}$
$\epsilon_{2,epl} = 0.105^{+0.018}_{-0.014}$
$\gamma_{1,ext} = -0.047^{+0.008}_{-0.008}$
$\gamma_{2,ext} = 0.139^{+0.007}_{-0.007}$
Observed data $\mathcal{I}_{obs}$
Reconstructed data $\mathcal{I}_{sim}$
Source Plane
Normalized Residuals
1"
HMC
VI
θ_E
γ_{epl}
$\epsilon_{1,epl}$
$\epsilon_{2,epl}$
$\gamma_{1,ext}$
$\gamma_{2,ext}$
Cusp
$\theta_E = 1.349^{+0.001}_{-0.001}$
$\gamma_{epl} = 2.266^{+0.043}_{-0.046}$
$\epsilon_{1,epl} = -0.101^{+0.005}_{-0.006}$
$\epsilon_{2,epl} = 0.042^{+0.003}_{-0.003}$
$\gamma_{1,ext} = 0.021^{+0.003}_{-0.003}$
$\gamma_{2,ext} = 0.068^{+0.002}_{-0.002}$
Observed data $\mathcal{I}_{obs}$
Reconstructed data $\mathcal{I}_{sim}$
Source Plane
Normalized Residuals
1"
θ_E
γ_{epl}
$\epsilon_{1,epl}$
$\epsilon_{2,epl}$
$\gamma_{1,ext}$
$\gamma_{2,ext}$

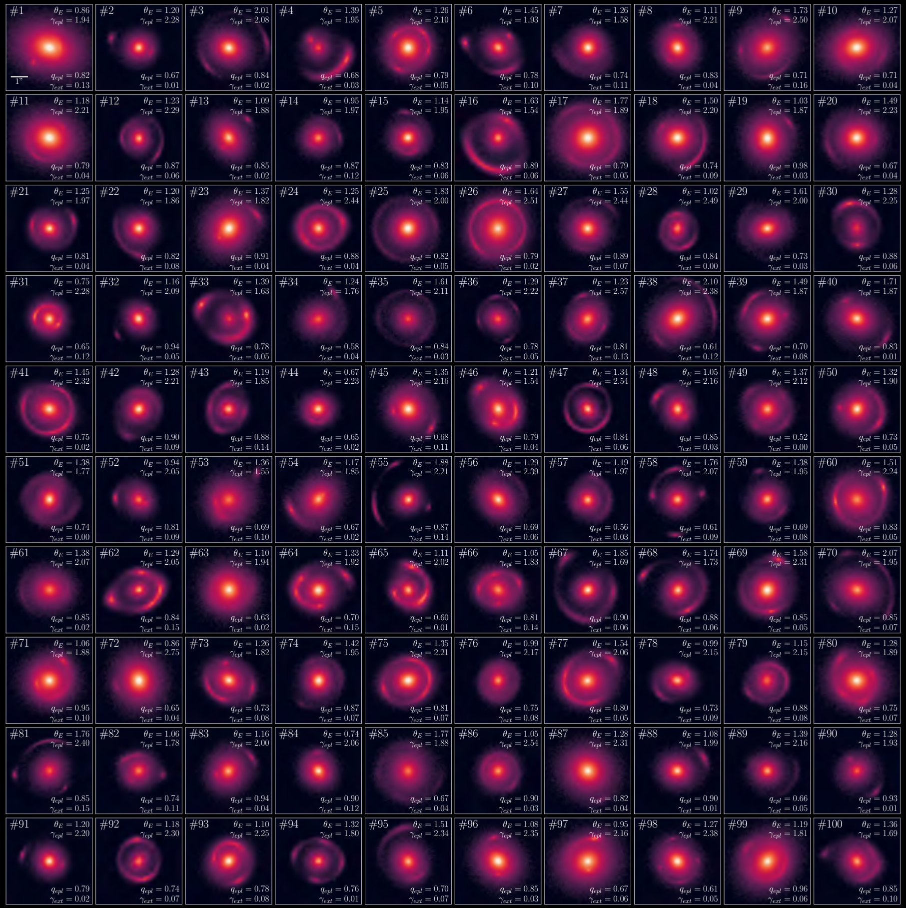


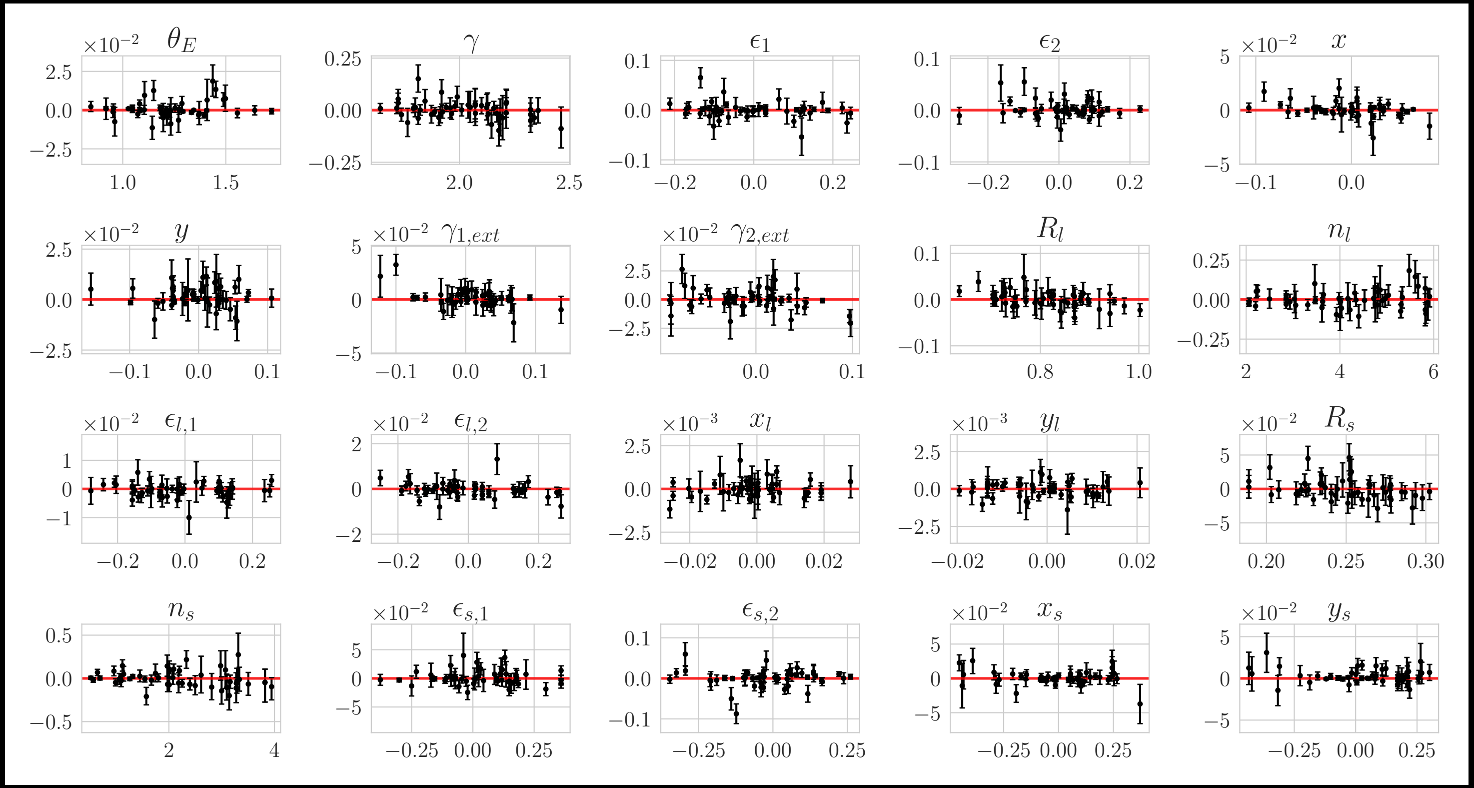
Cross
$\theta_E = 1.322^{+0.002}_{-0.002}$
$\gamma_{epl} = 1.808^{+0.057}_{-0.056}$
$\epsilon_{1,epl} = -0.127^{+0.006}_{-0.006}$
$\epsilon_{2,epl} = 0.058^{+0.004}_{-0.004}$
$\gamma_{1,ext} = 0.010^{+0.008}_{-0.008}$
$\gamma_{2,ext} = -0.006^{+0.004}_{-0.004}$
Observed data $\mathcal{I}_{obs}$
Reconstructed data $\mathcal{I}_{sim}$
Source Plane
Normalized Residuals
1"
θ_E
γ_{epl}
$\epsilon_{1,epl}$
$\epsilon_{2,epl}$
$\gamma_{1,ext}$
$\gamma_{2,ext}$

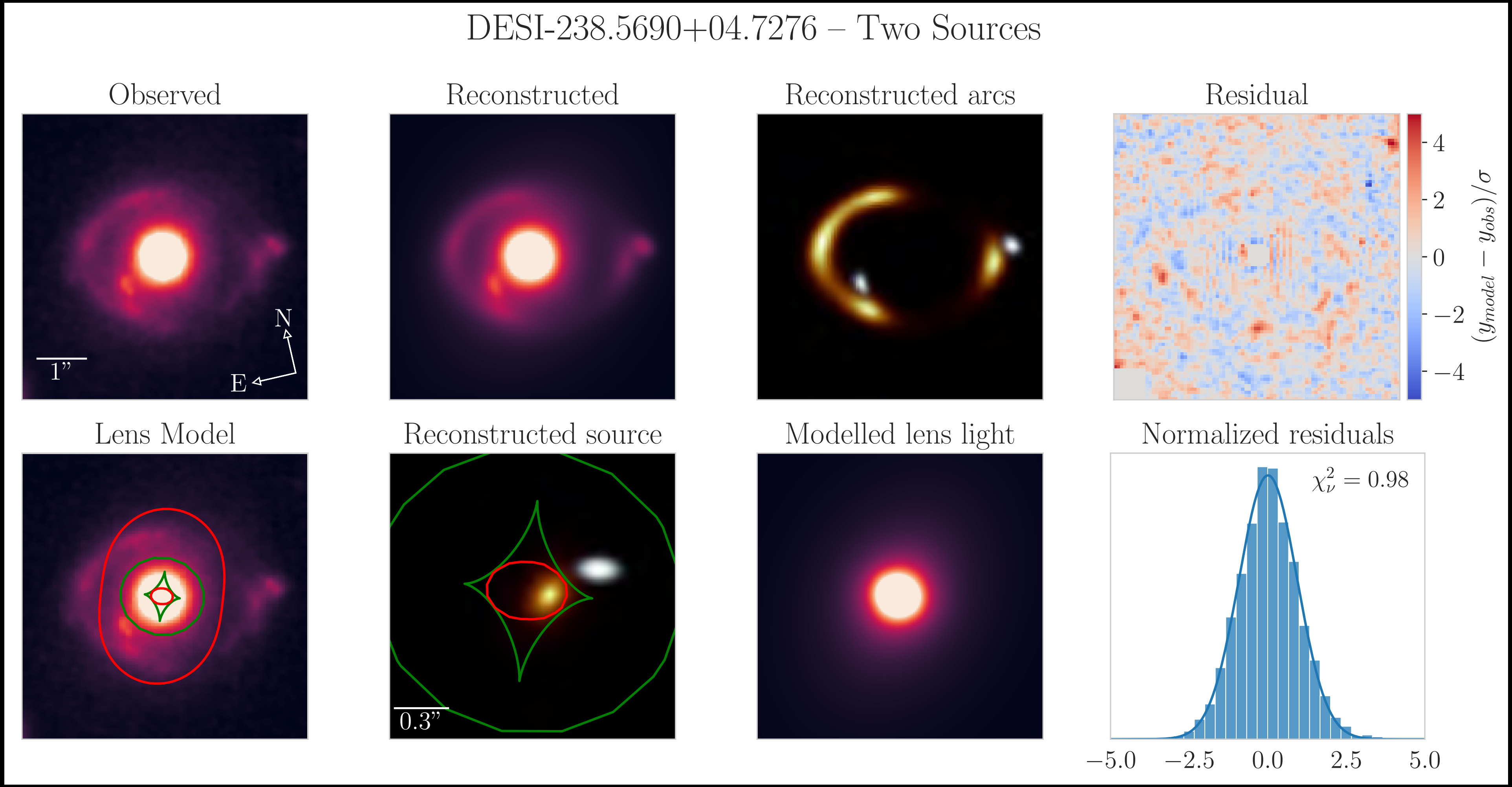


Double
$\theta_E = 1.504^{+0.005}_{-0.005}$
$\gamma_{epl} = 2.297^{+0.048}_{-0.046}$
$\epsilon_{1,epl} = -0.065^{+0.010}_{-0.010}$
$\epsilon_{2,epl} = -0.194^{+0.017}_{-0.018}$
$\gamma_{1,ext} = -0.072^{+0.005}_{-0.005}$
$\gamma_{2,ext} = -0.046^{+0.008}_{-0.008}$
Observed data $\mathcal{I}_{obs}$
Reconstructed data $\mathcal{I}_{sim}$
Source Plane
Normalized Residuals
1"
θ_E
γ_{epl}
$\epsilon_{1,epl}$
$\epsilon_{2,epl}$
$\gamma_{1,ext}$
$\gamma_{2,ext}$

Testing Results







DESI-238.5690+04.7276 – Two Sources
Observed
Reconstructed
Reconstructed arcs
Residual
N
E
1”
4
2
0
−2
−4
$(y_{model} - y_{obs})/\sigma$
Lens Model
Reconstructed source
Modelled lens light
Normalized residuals
0.3”
$\chi^2_\nu = 0.98$
−5.0
−2.5
0.0
2.5
5.0

Next Steps

Our paper reporting this work (arXiv: 2202.07663) and has been accepted for publication in The Astrophysical Journal. What next?

1. Develop a suite of comprehensive, automated tests + improve documentation
2. Public code release soon
3. Improved speed with 2 GPU nodes (8 GPUs: configurable on Perlmutter)
   - Can we eliminate the VI step using mass matrix adaptation for HMC?
4. Apply our code to observed systems from our Hubble Space Telescope program (ID: 15867)
   - Sub-galactic dark matter halo detection
   - Using ground-based data, can model systems spectroscopically confirmed by DESI