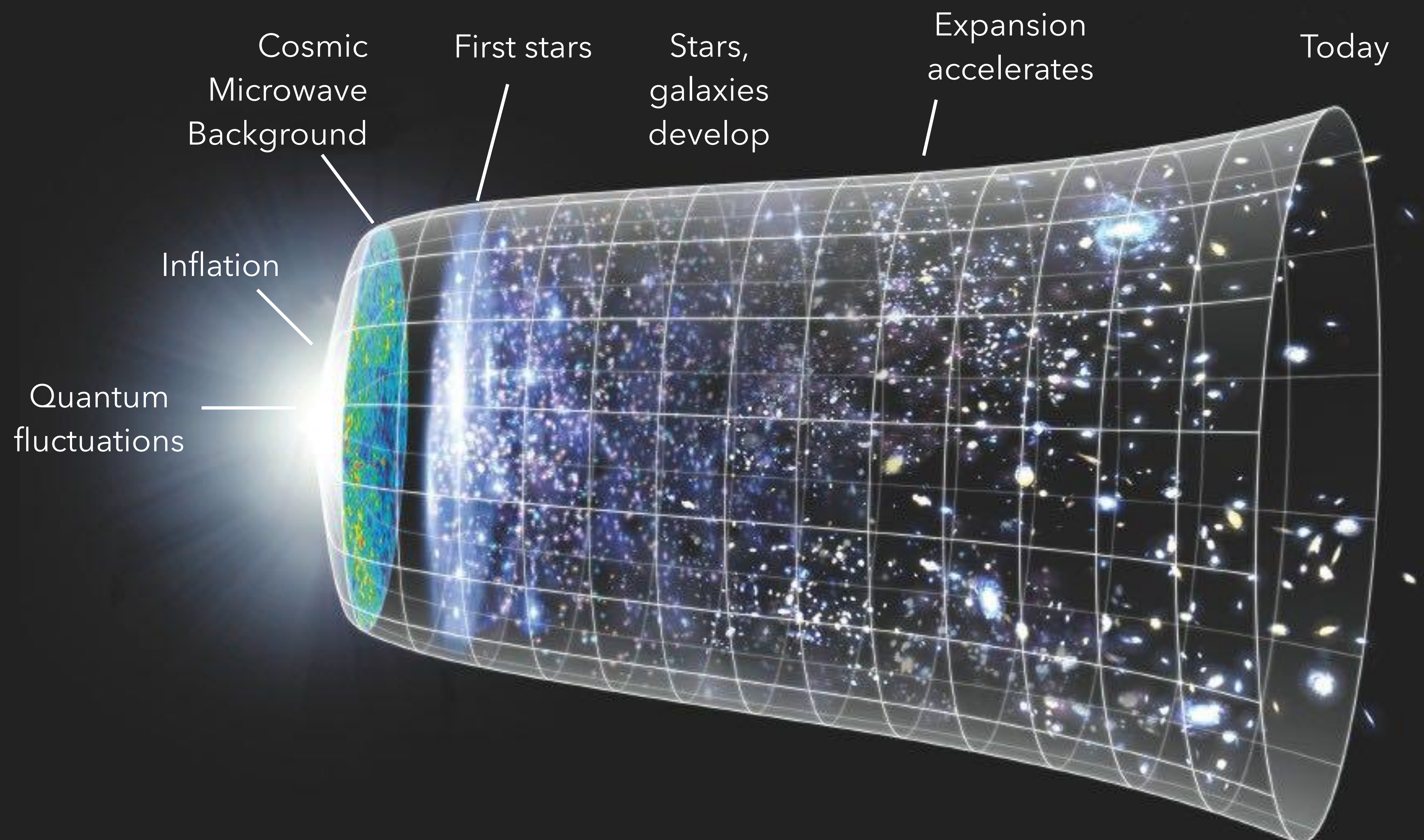
Data-Driven Strong Gravitational Lensing Analysis in the Era of Large Sky Surveys

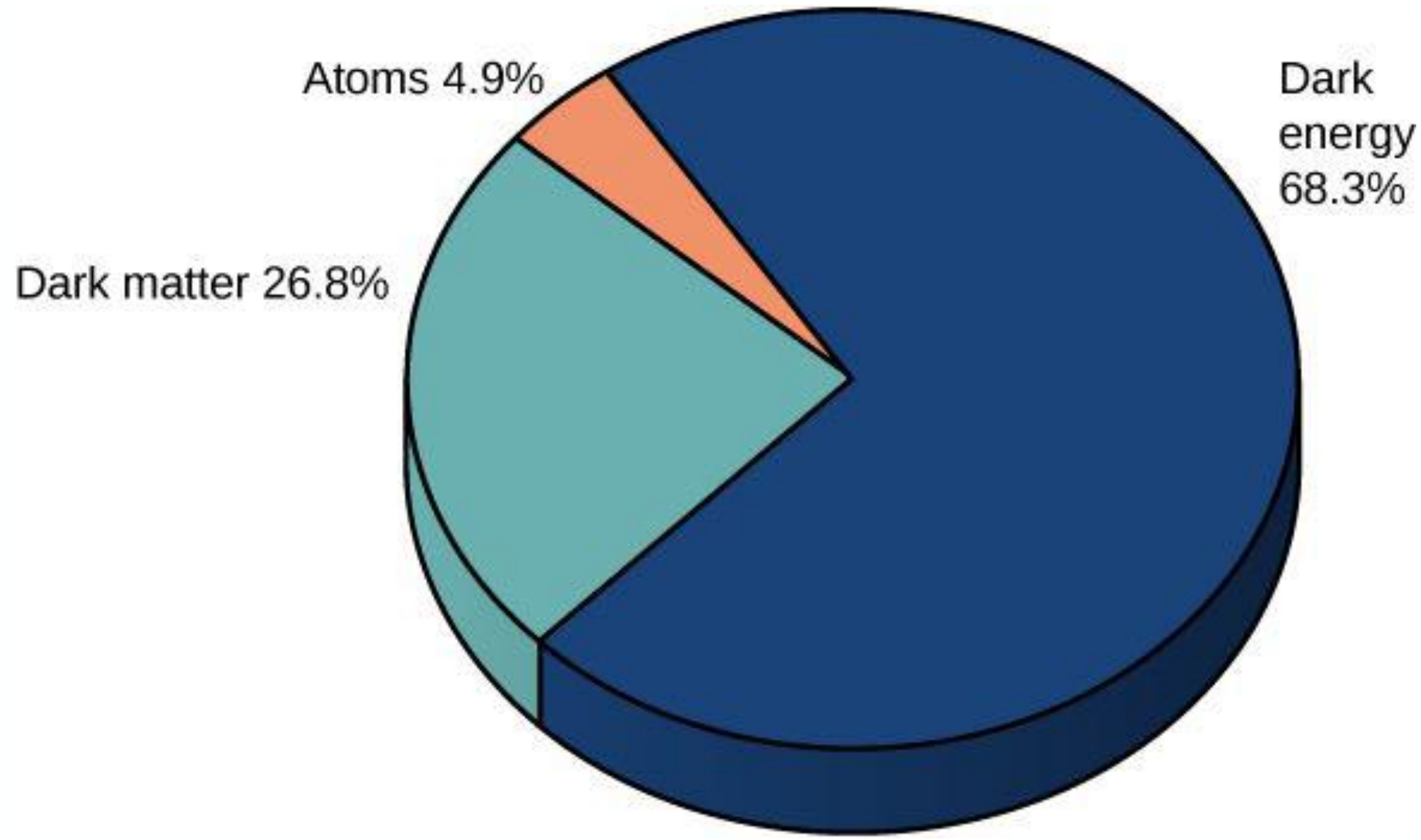
Laurence Perreault-Levasseur

Ciela Institute
Université de Montréal
Mila
Flatiron Institute
Center for Computational Astrophysics


Cosmic
Microwave
Background
First stars
Stars,
galaxies
develop
Expansion
accelerates
Today
Inflation
Quantum
fluctuations

INFLATIONARY-ΛCDM

Inflation +

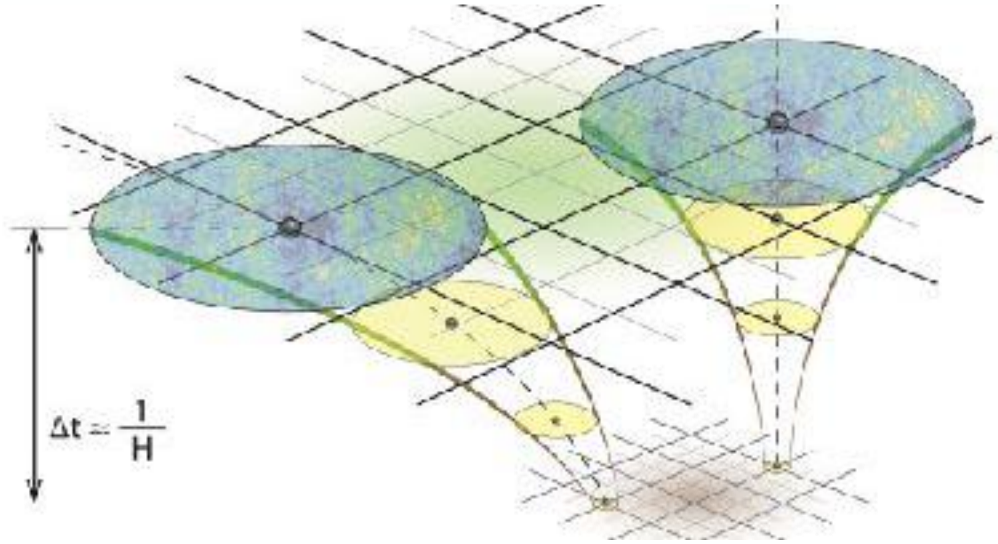


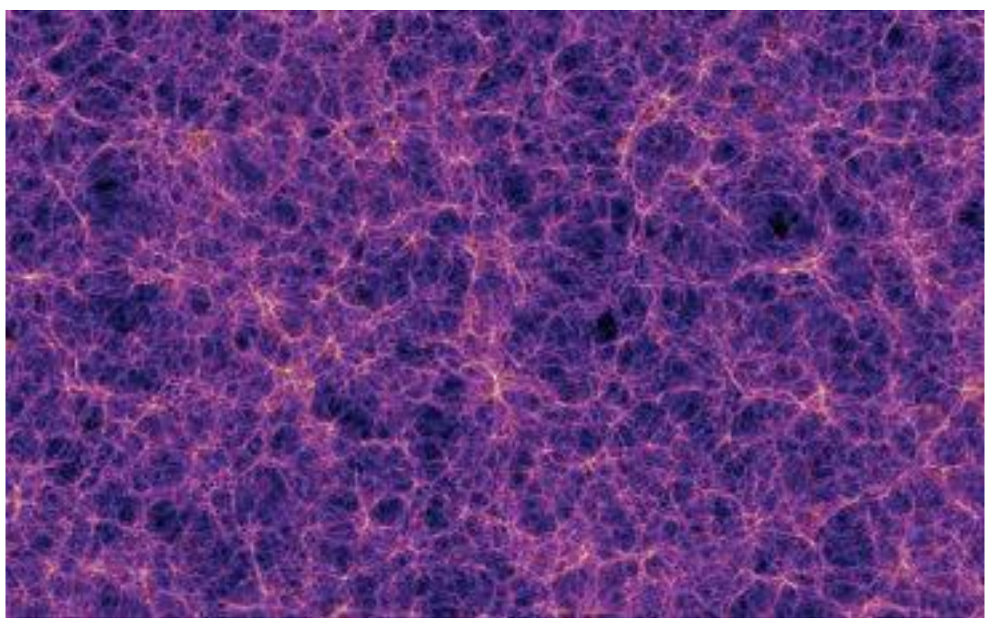
At this point,
are we done with cosmology?

Inflationary-ΛCDM

Inflation:

The nature, properties, or origin of the field causing inflation are completely unknown





Dark Matter:

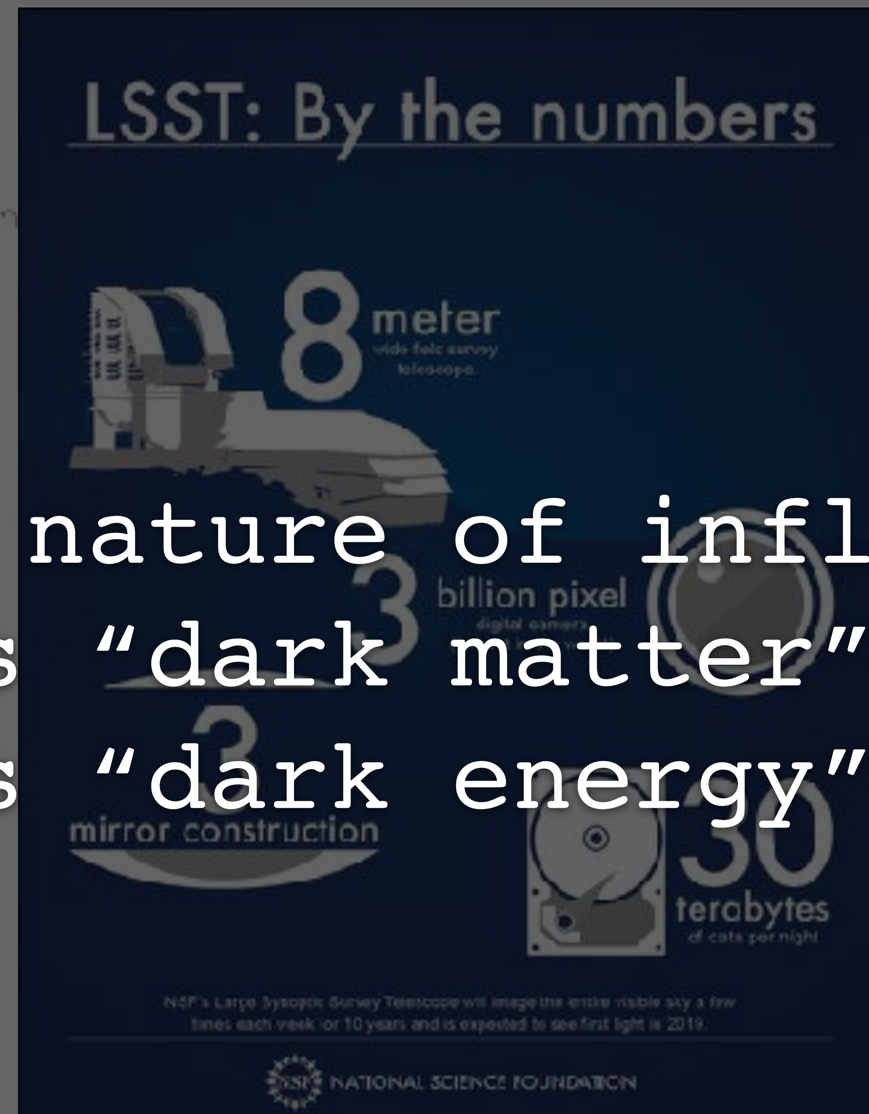
What is known:
only that it exists and gravitates

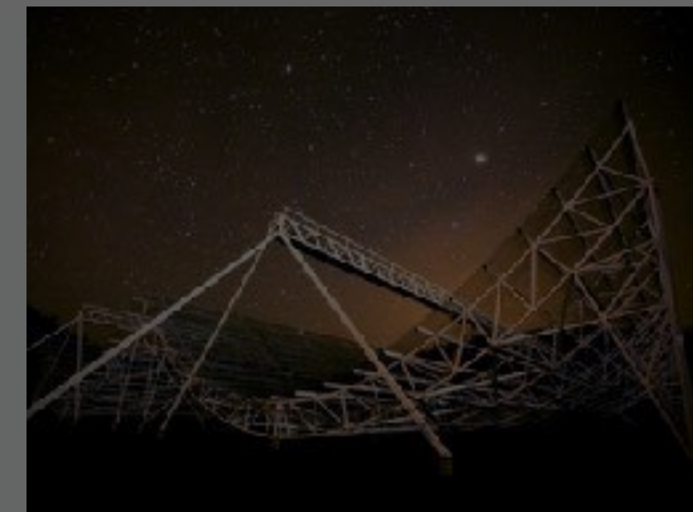
Dark Energy:

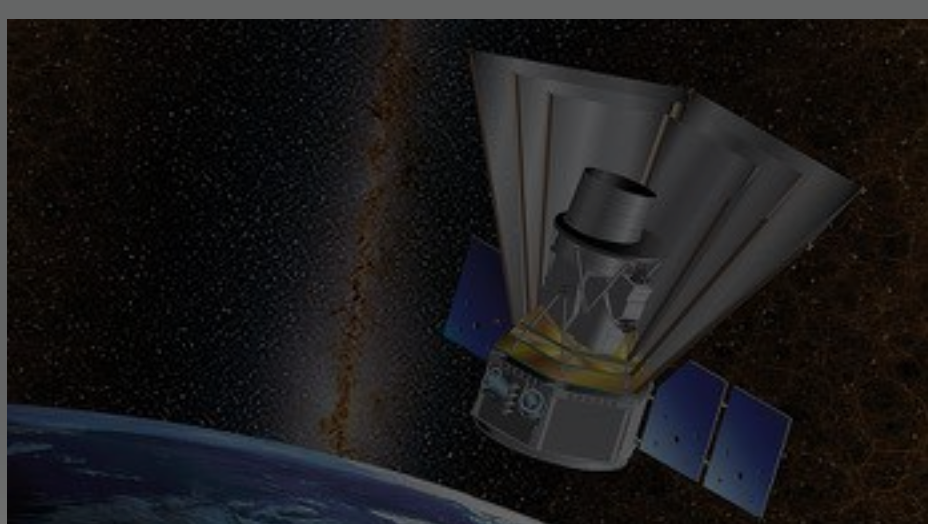
Is not even remotely understood.

Modern Cosmology



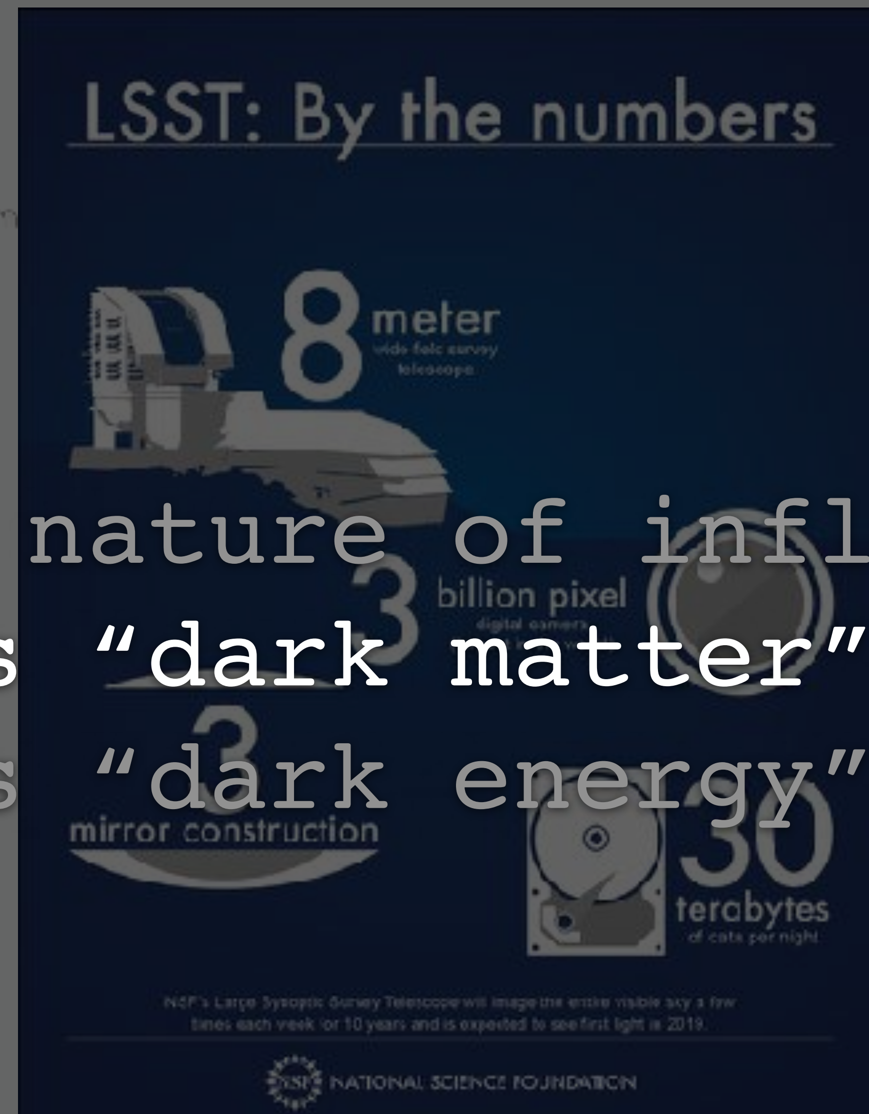


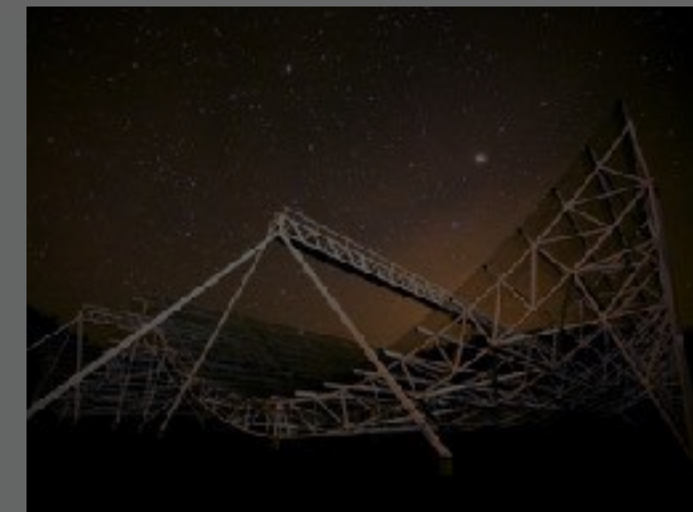


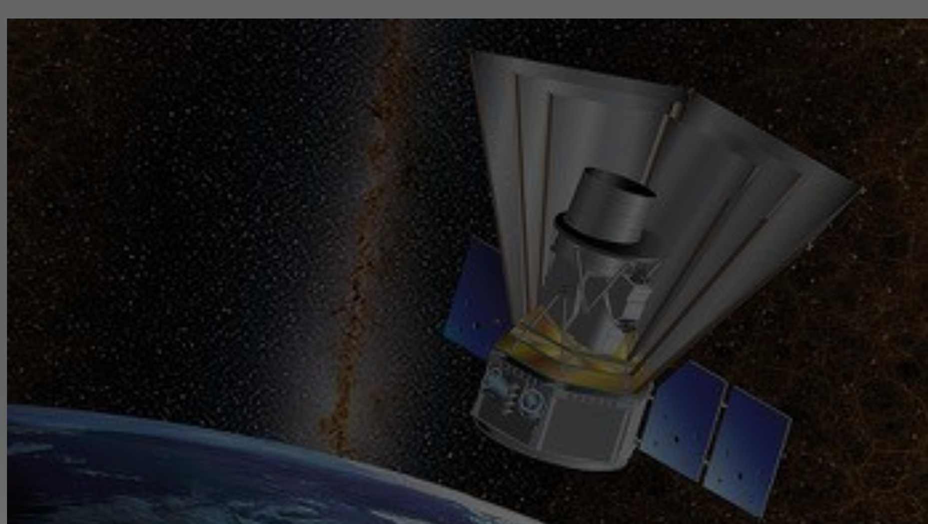




What is the nature of inflation?
What is “dark matter”?
What is “dark energy”?

Modern Cosmology





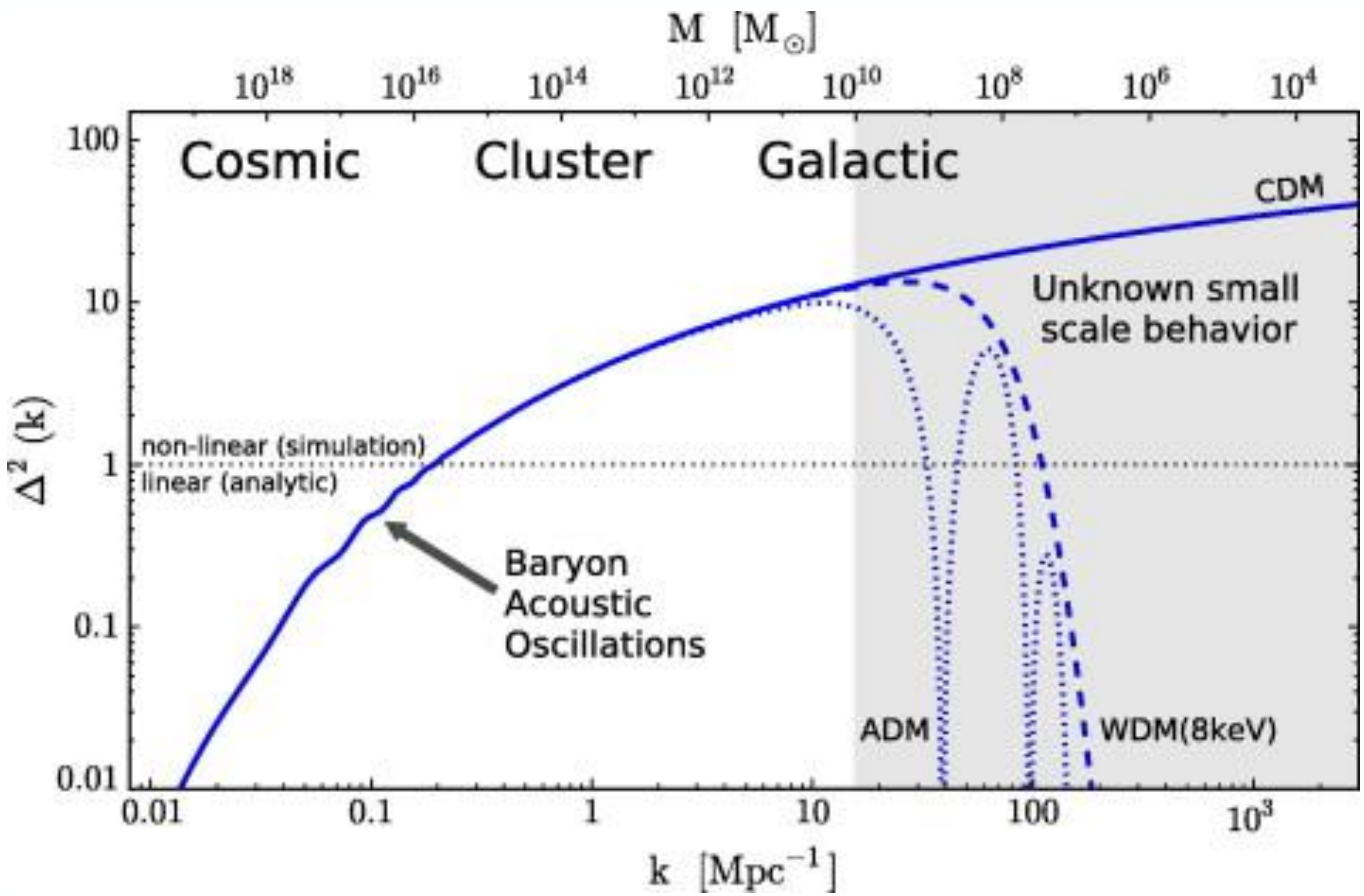


What is the nature of inflation?

What is "dark matter"?

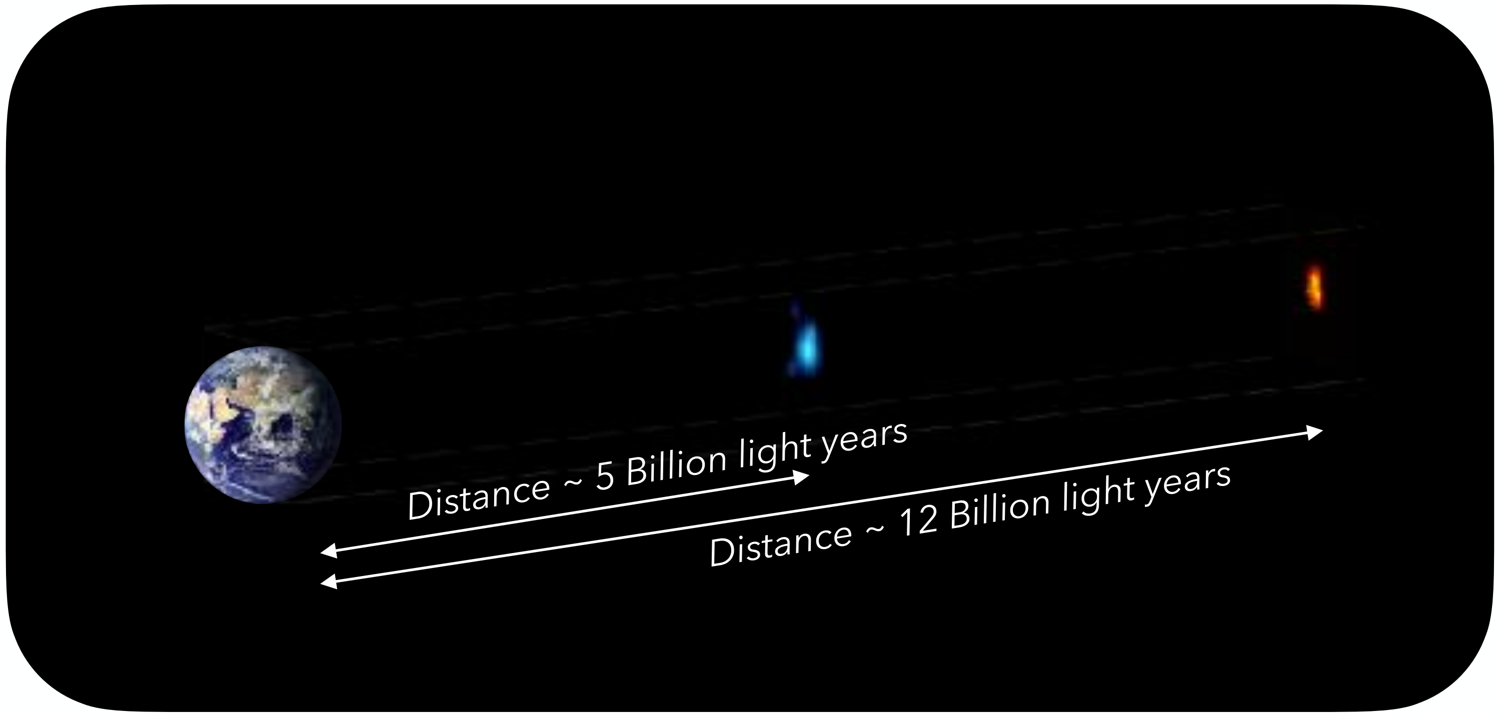
What is "dark energy"?

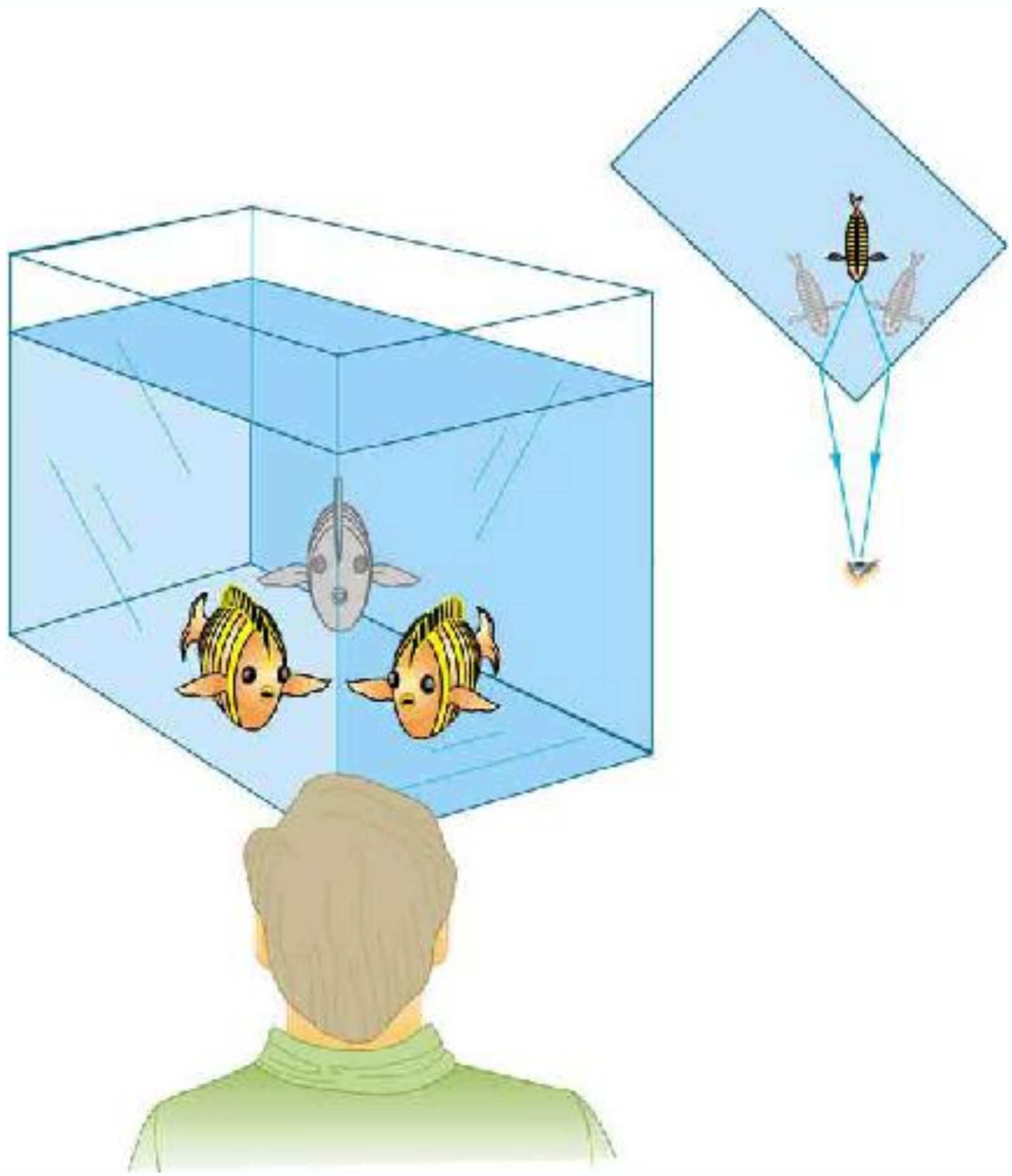
Matter power spectrum

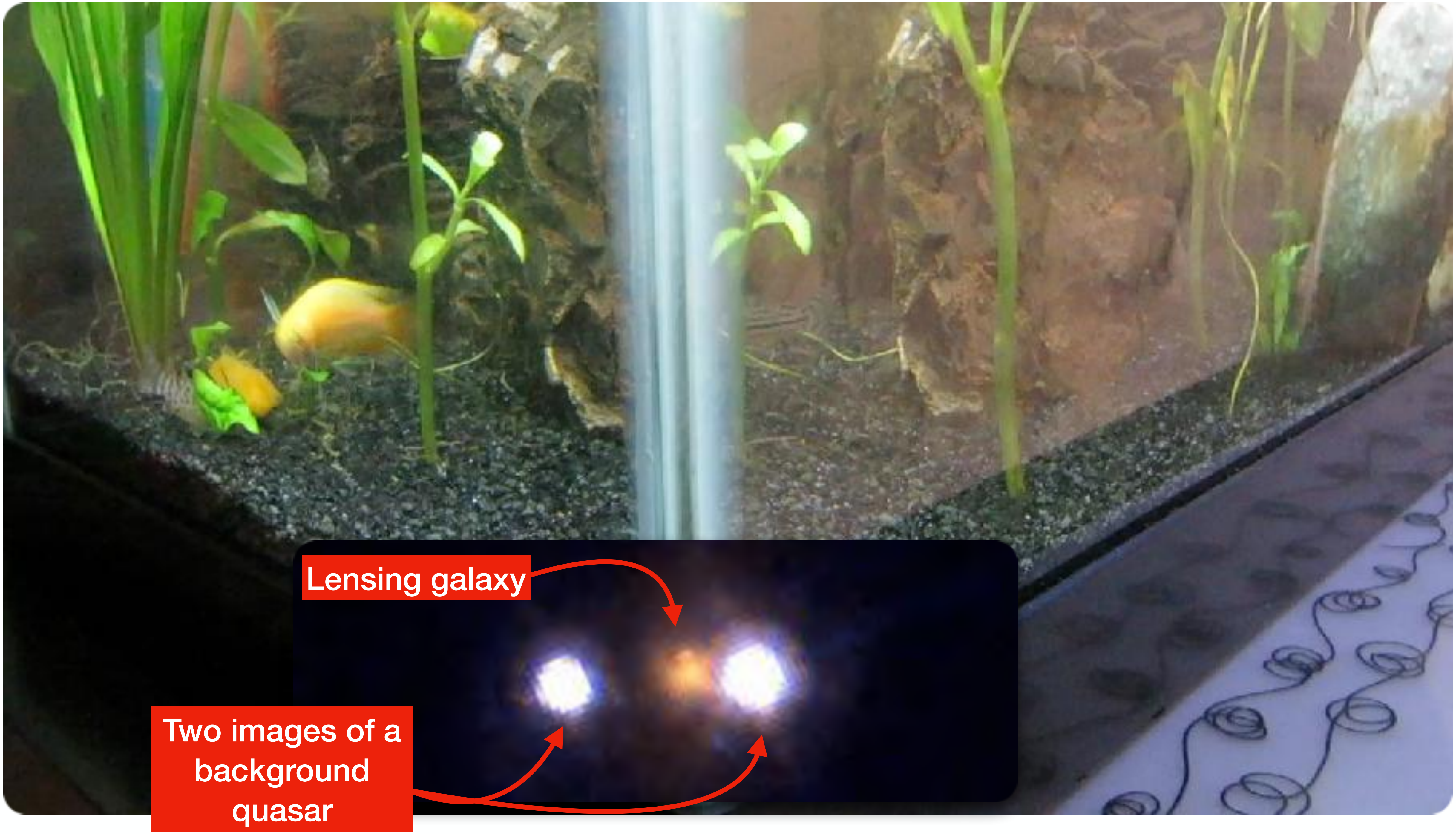


Strong Gravitational Lensing

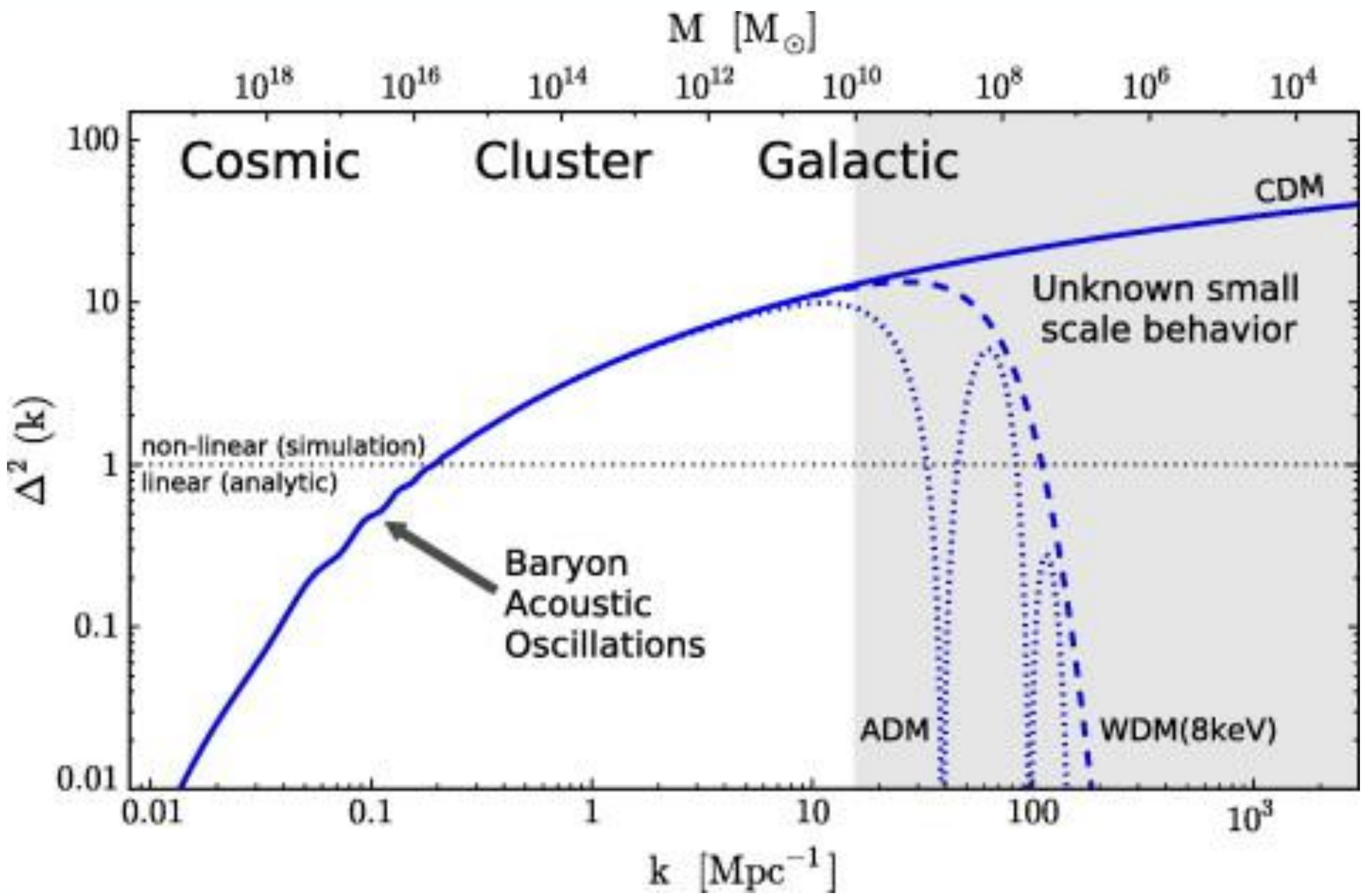
Formation of **multiple images** of a single distant object due to the **deflection of its light** by the **gravity** of intervening structures.

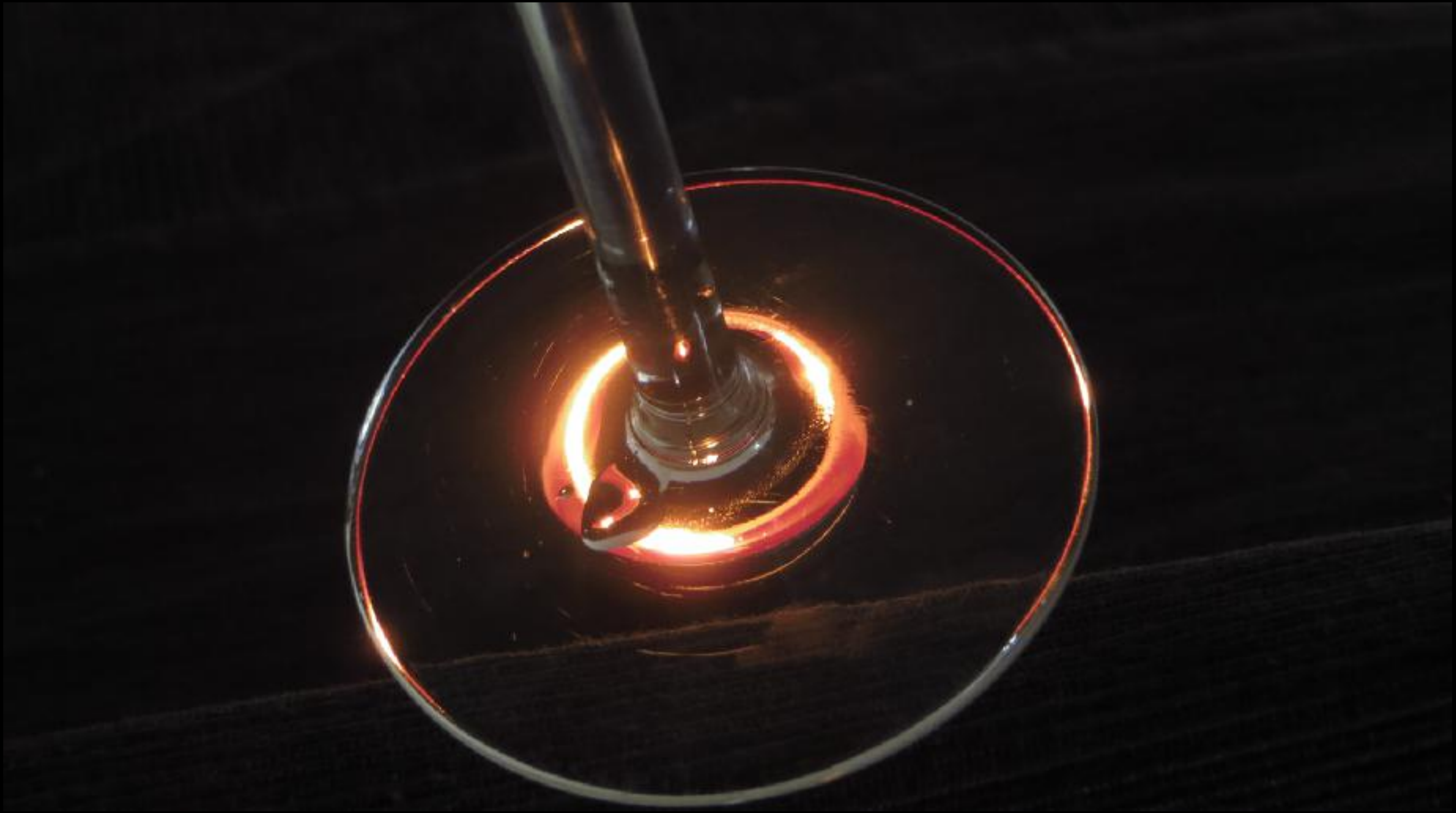





Lensing galaxy
Two images of a background quasar

Matter power spectrum





SCIENCE MOTIVATIONS FOR STRONG LENSING

- **Foreground structure**

Use lensing to probe the **distribution of matter** in the lensing galaxies.

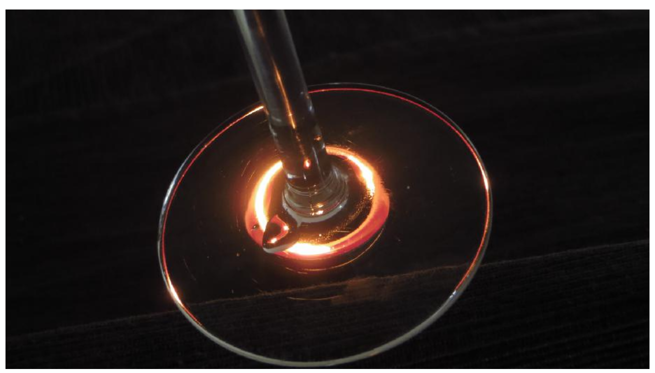
- **Background source**

Use strong lensing as a **cosmic telescope**.

- **Cosmology**

Use lensing to probe the **cosmological parameters (H_0)**

Lensing analysis



Data

1: Morphology of the background source (the true, undistorted image of the candle)

2: Matter distribution in the lens (the shape of the wineglass)

Source Parameters (linear) → x

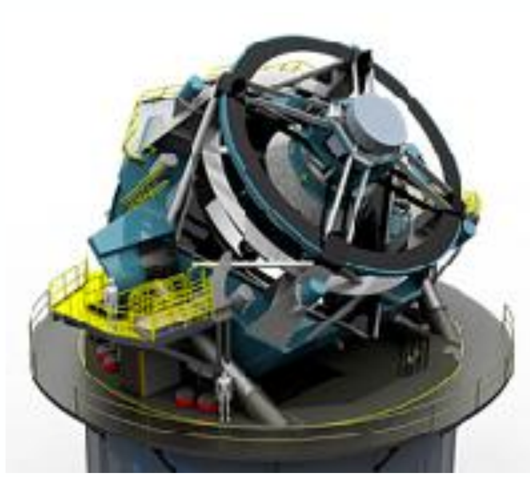
Data → $$y = L(p)x + n$$

Lens Parameters (non-linear) → p

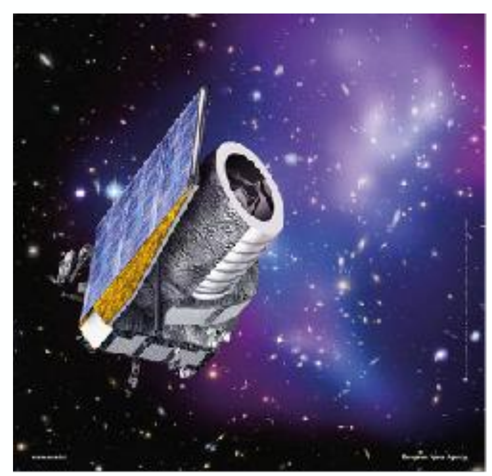
Noise → n

LOOKING INTO THE FUTURE

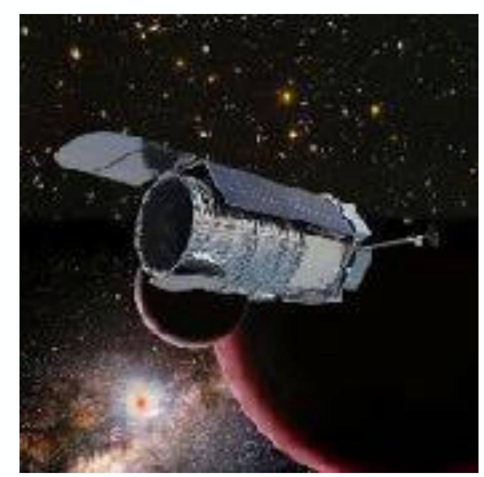
In the next few years, we're expecting to discover more than 170,000 new lenses.



LSST



Euclid consortium



NANCY GRACE ROMAN SPACE TELESCOPE

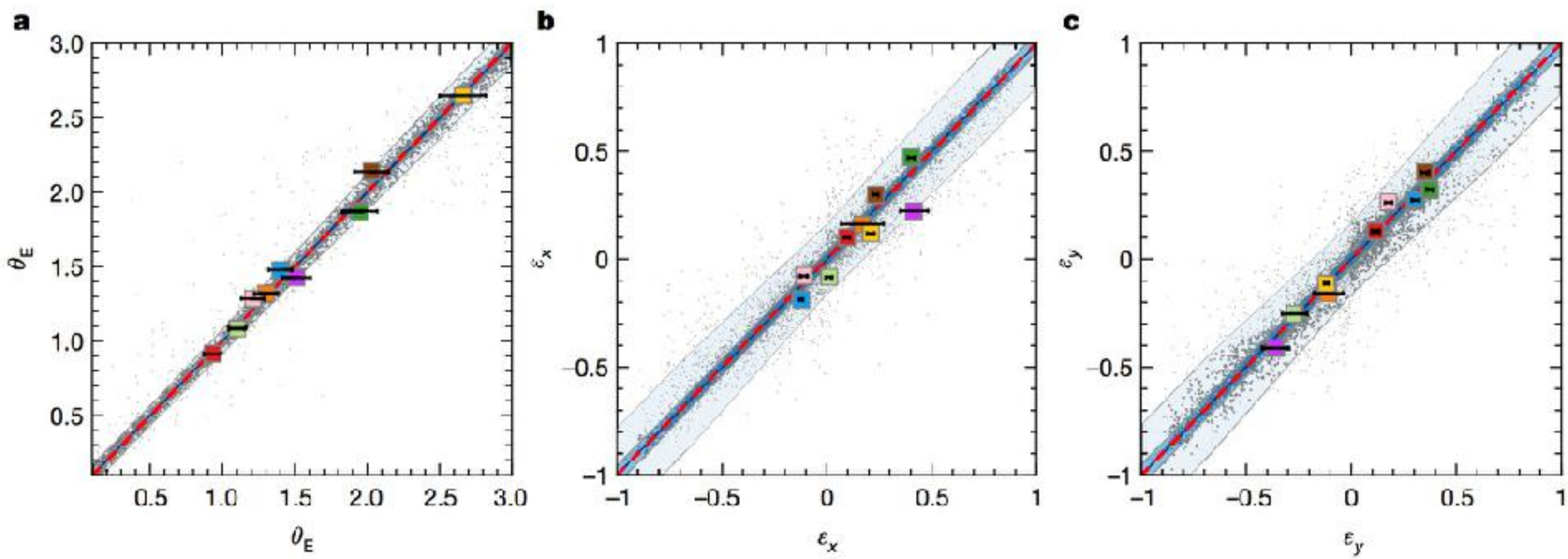
Methods for the future:

How are we going to analyze 170,000 lenses?

- Lens modeling is **very slow**.
- Simple lens model takes ~3 days

=> **1,400 years !**

ESTIMATING THE MATTER DISTRIBUTION PARAMETERS WITH CNNS

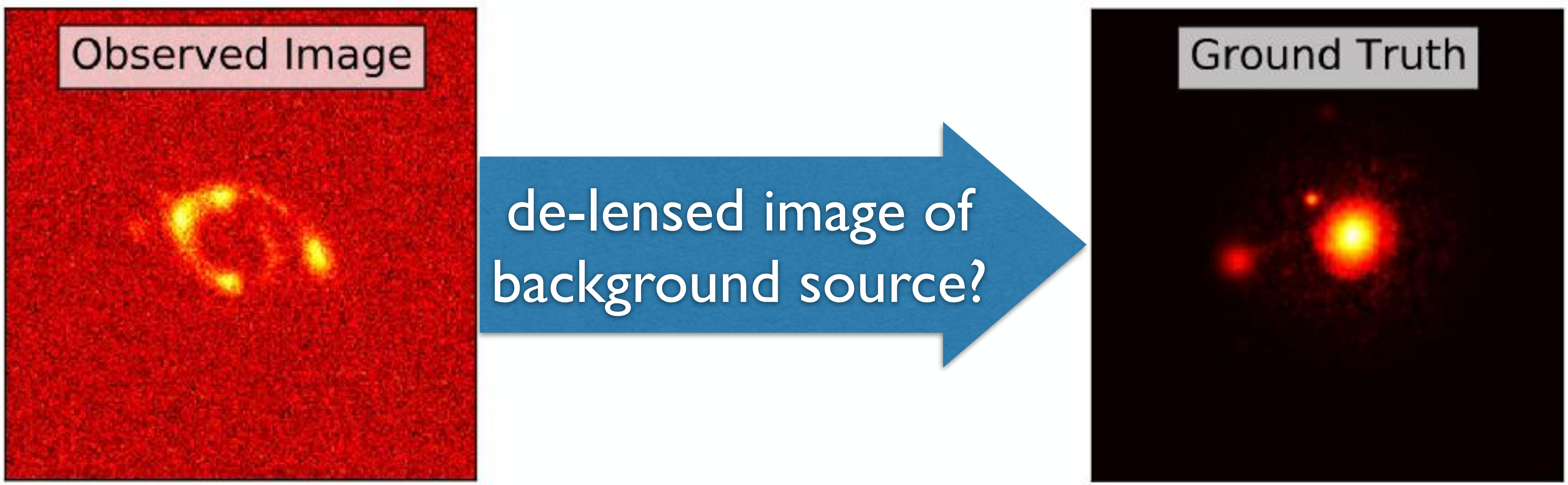


10 million times faster than traditional lens modeling.

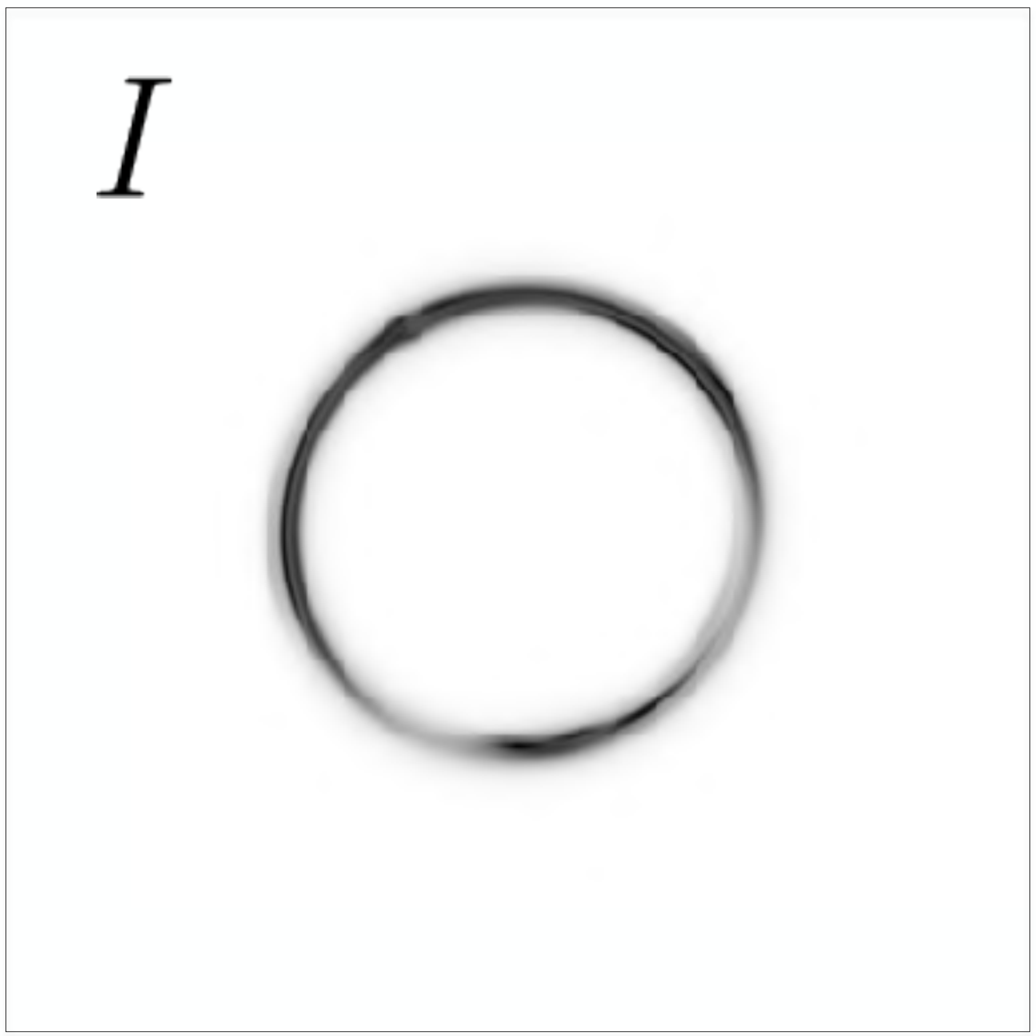
0.01 seconds on a **single GPU**

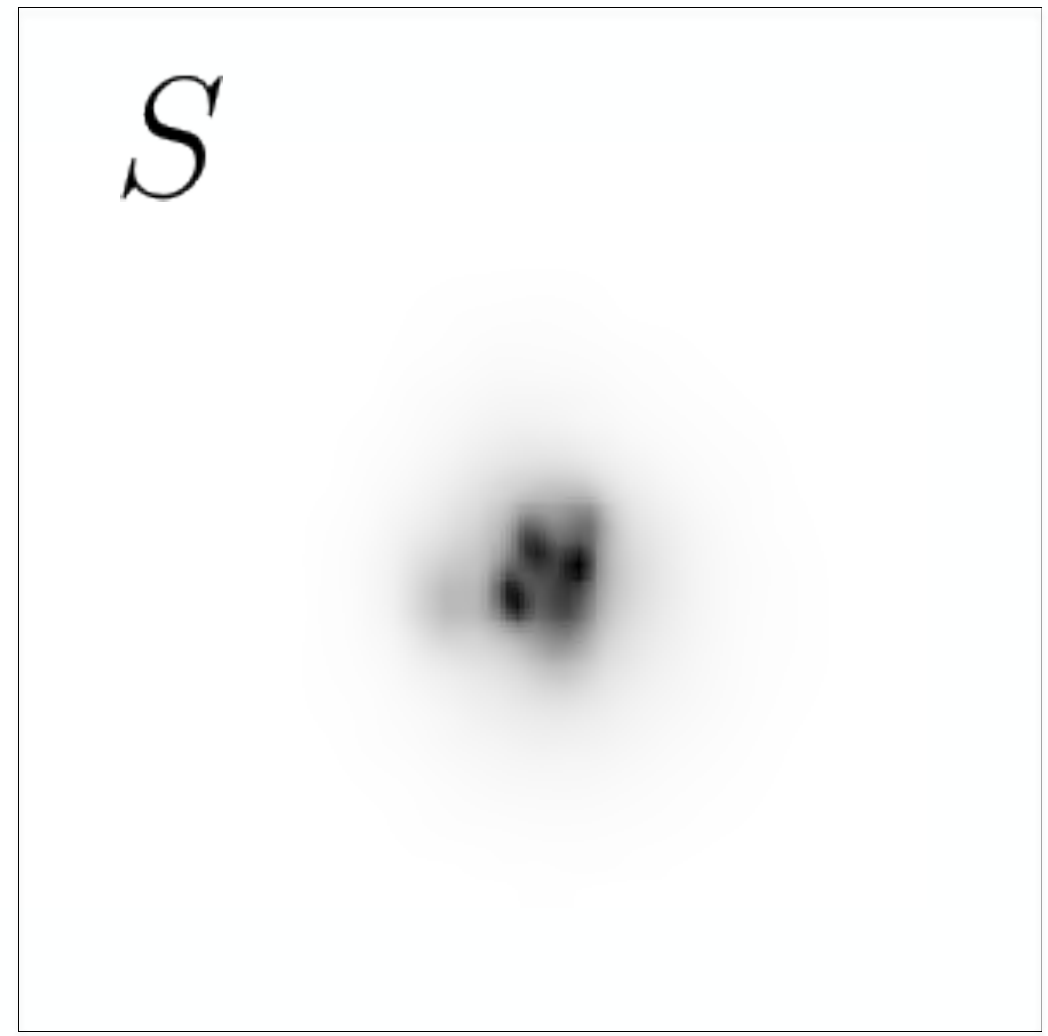
Hezaveh, Perreault Levasseur, Marshall, Nature, 2017

Undistorted image of the background source



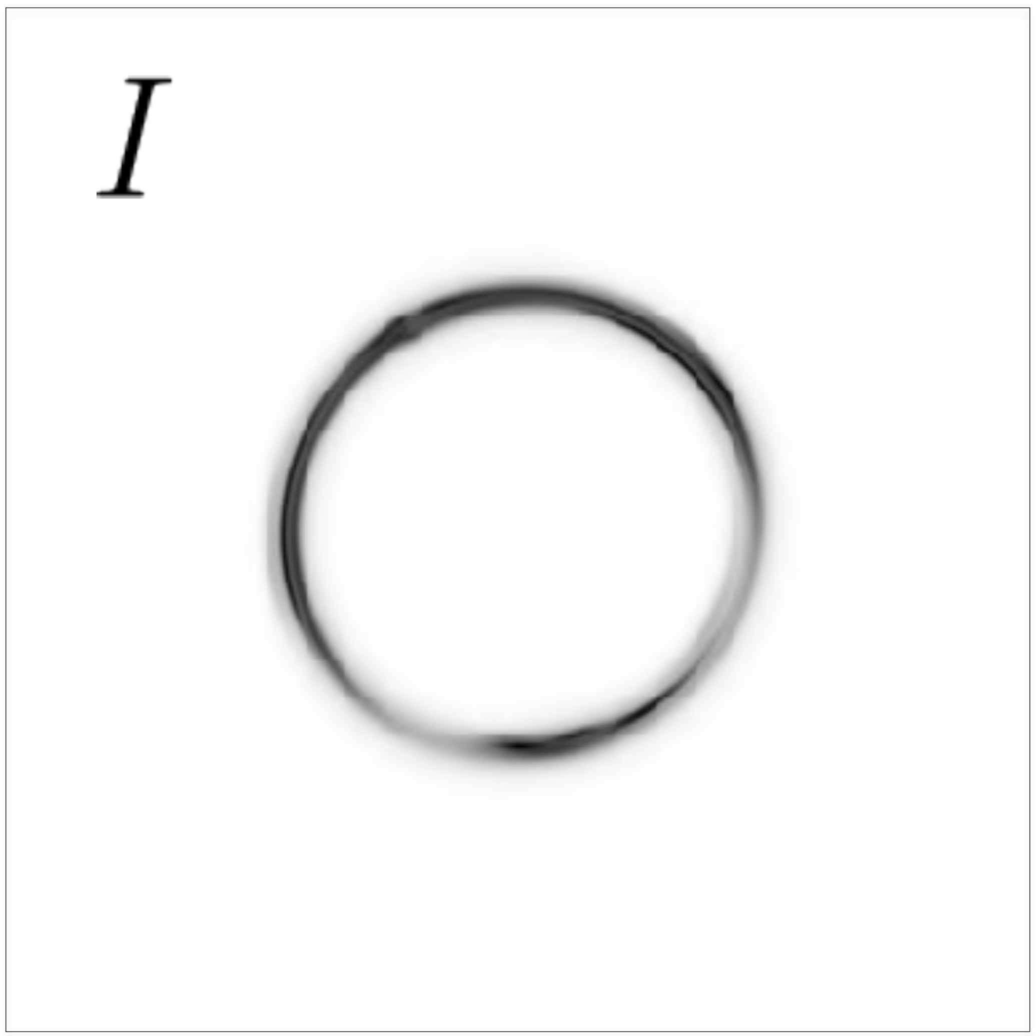
Pixel values of the background source are linear parameters

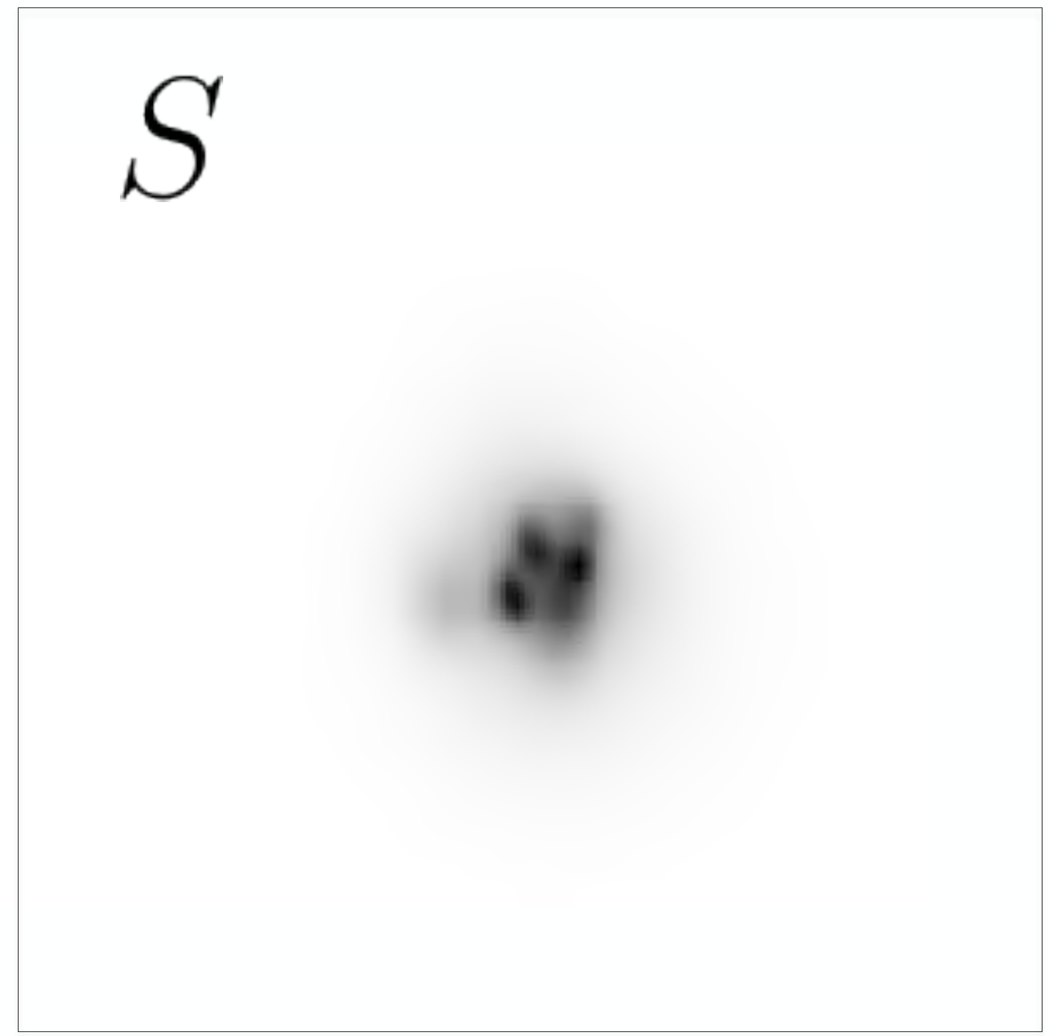




$$I = L(p)S$$

Pixel values of the background source are linear parameters

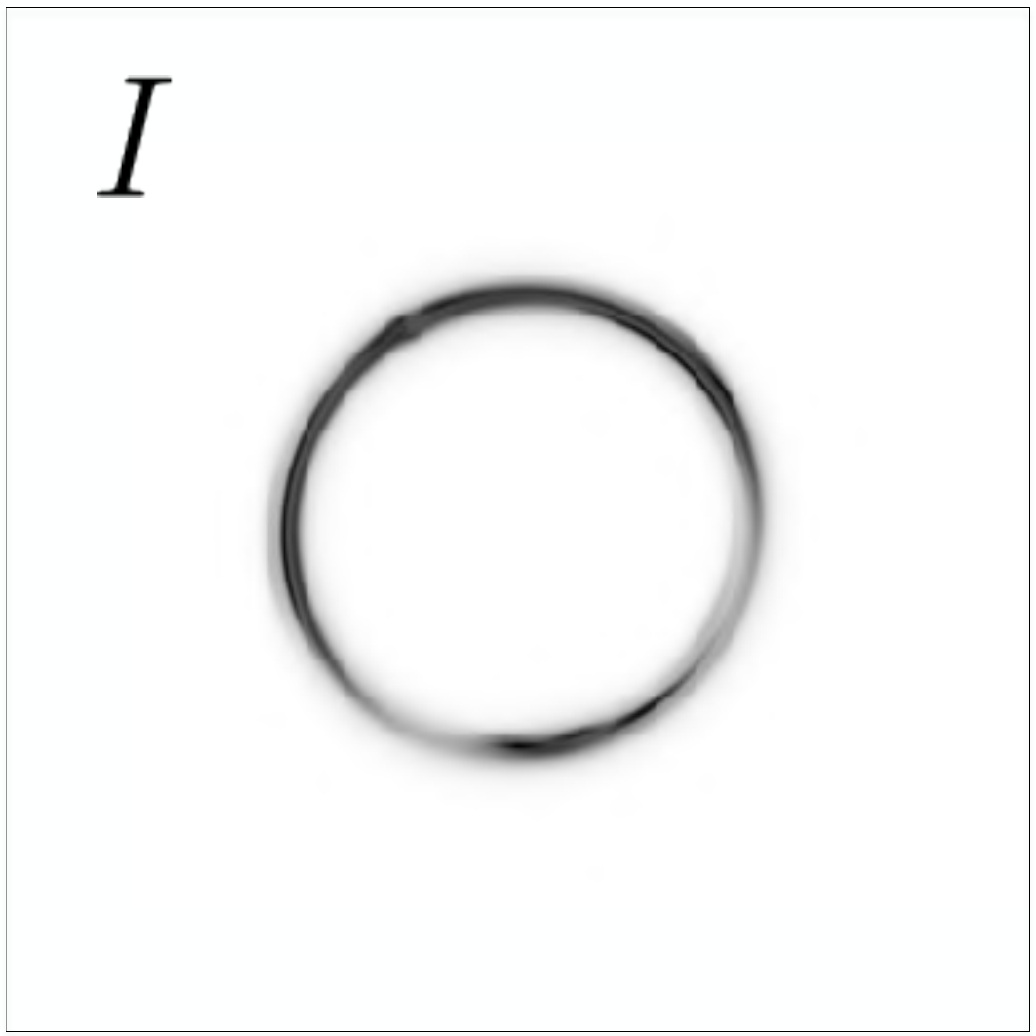


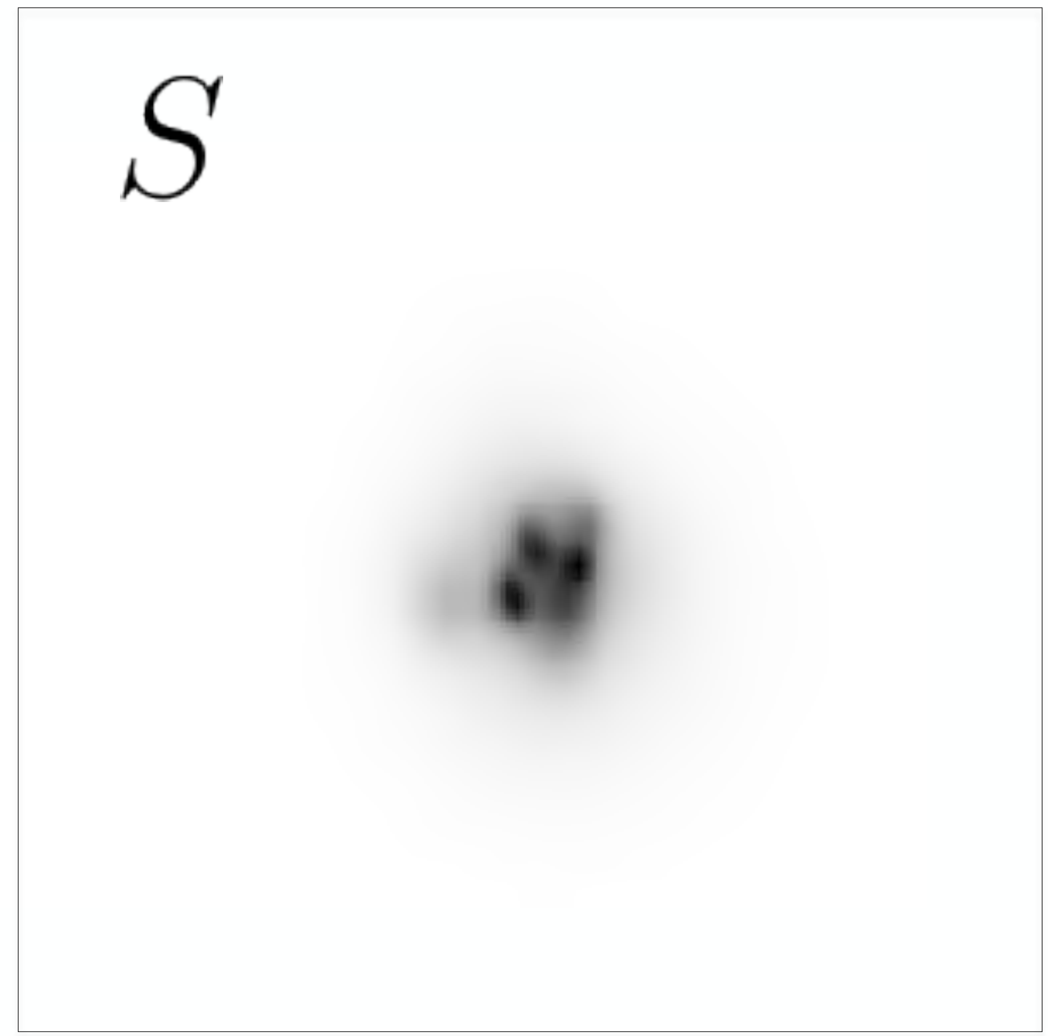


$$I = L(p)S$$

$$S = (L^T C_N^{-1} L)^{-1} L^T C_N^{-1} D$$

Pixel values of the background source are linear parameters

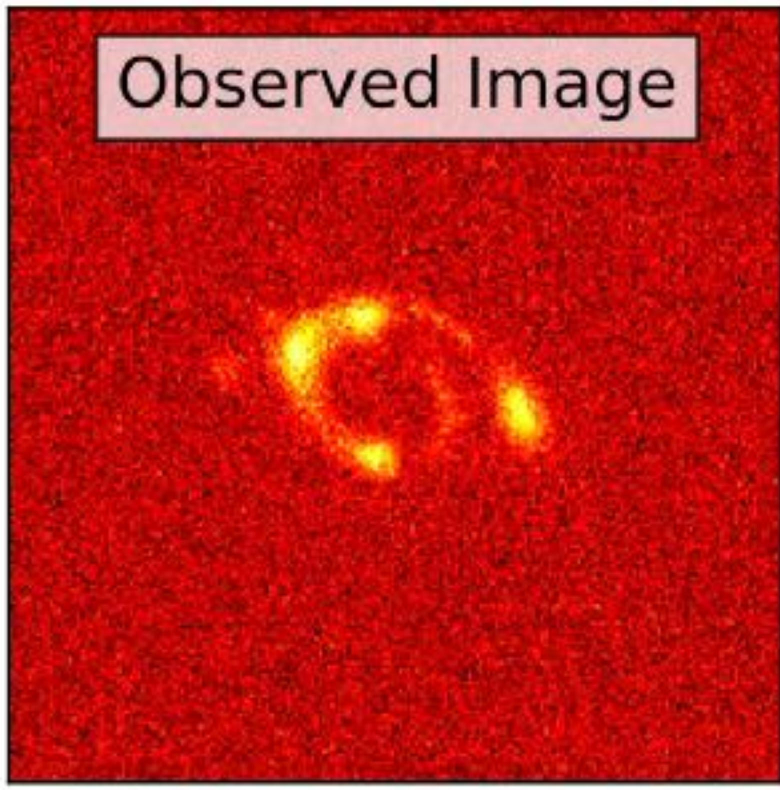




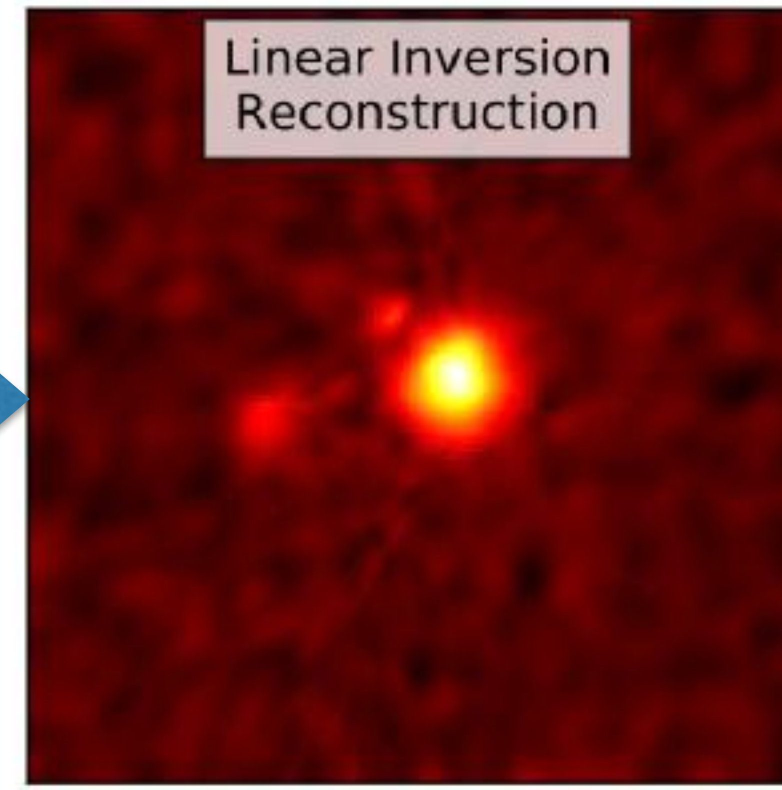
$$I = L(p)S$$

$$S = (L^T C_N^{-1} L + C_p^{-1})^{-1} L^T C_N^{-1} D$$

UNDISTORTED IMAGE OF THE BACKGROUND SOURCE



de-lensed image of background source



$$S = (L^T C_N^{-1} L + C_p^{-1})^{-1} L^T C_N^{-1} D$$

The Recurrent Inference Machine

The RIM is designed to solve problems of the form:

$$\mathbf{y} = \mathbf{f}(\mathbf{x}) + \mathcal{N}$$

In this model, assuming the noise is Gaussian, the likelihood of $\mathbf{y}$ given $\mathbf{x}$ is:

$$\mathcal{L}(\mathbf{y}|\mathbf{x}) \propto \exp\left[-\frac{\mathbf{1}}{\mathbf{2}}(\mathbf{y} - \mathbf{f}(\mathbf{x}))^{\mathbf{T}}\mathbf{C}_{\mathcal{N}}^{-\mathbf{1}}(\mathbf{y} - \mathbf{f}(\mathbf{x}))\right]$$

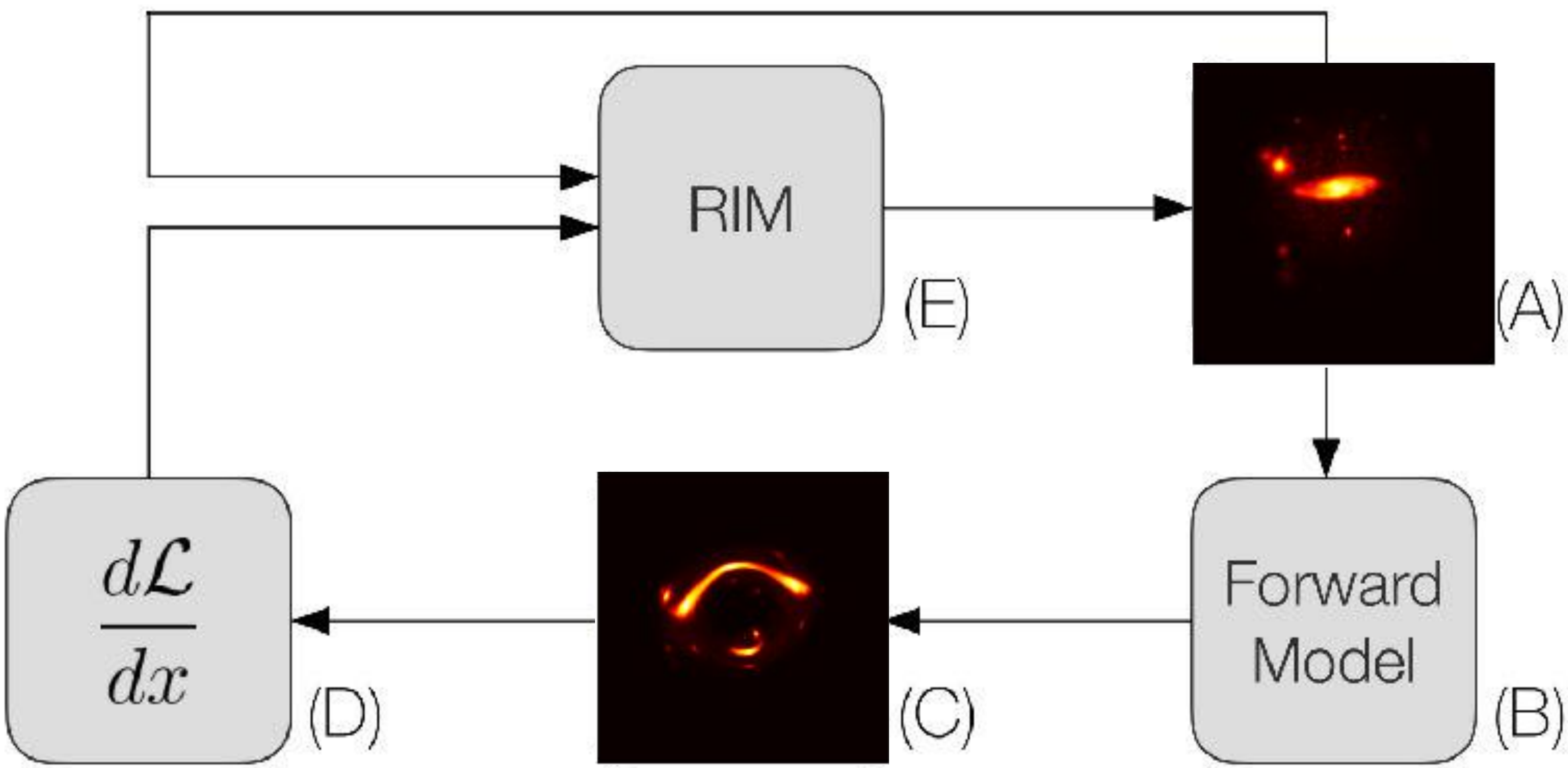
The RIM solves the above equation for $\mathbf{x}$ recursively. At every time step, it takes its current estimate of $\mathbf{x}$, $\mathbf{x}_t$, and the gradient $\nabla_{\mathbf{x}}\mathcal{L}|_{\mathbf{x_t}}$ to output an update to the estimate $\Delta\mathbf{x_t}$ so that:

$$\mathbf{x_{t+1}} = \mathbf{x_t} + \mathbf{\Delta x_t}$$

=> very similar idea to what a downhill optimizer does

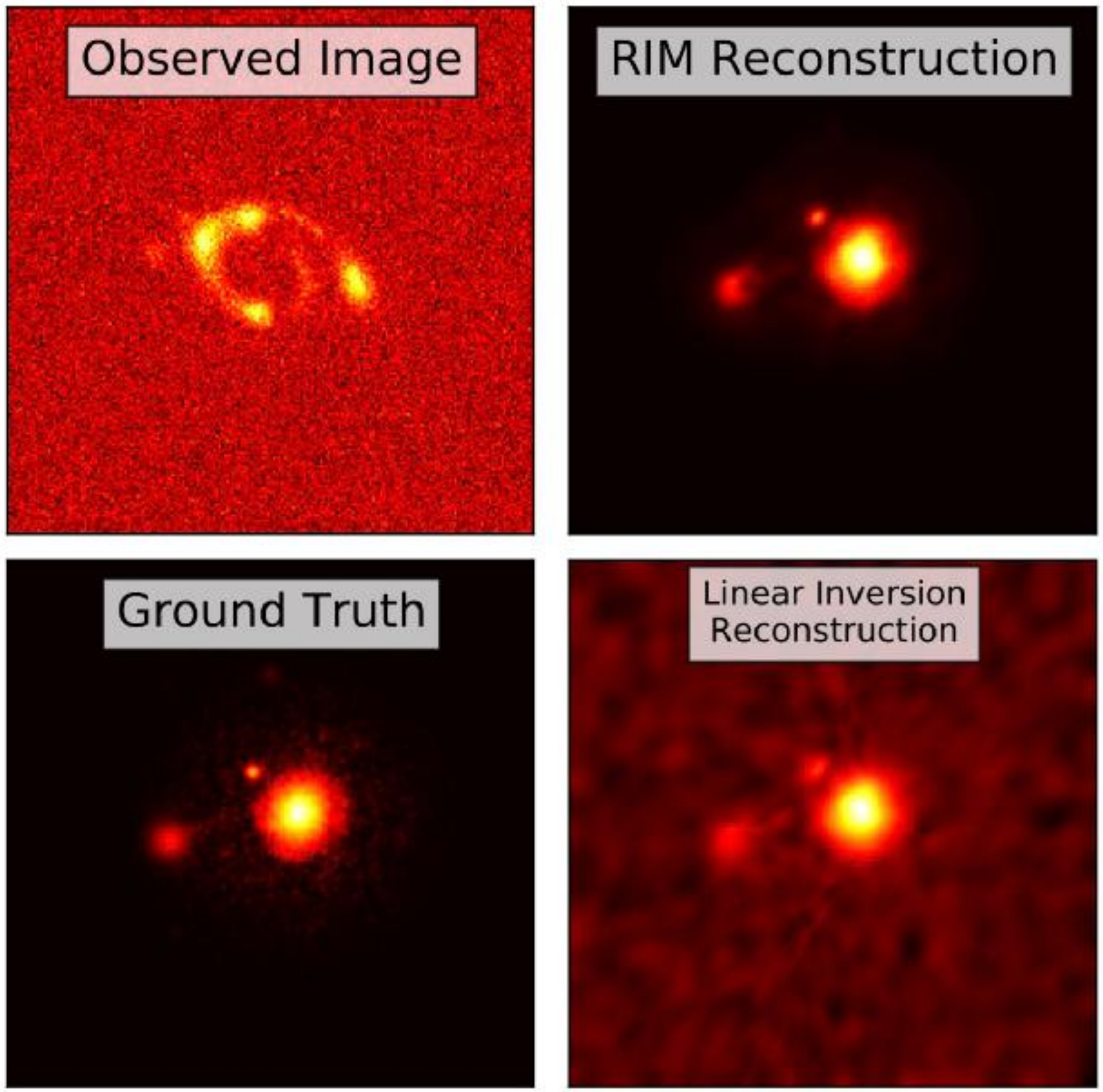
Putzky & Welling, 2017

Estimating the background source images with the Recurrent Inference Machine



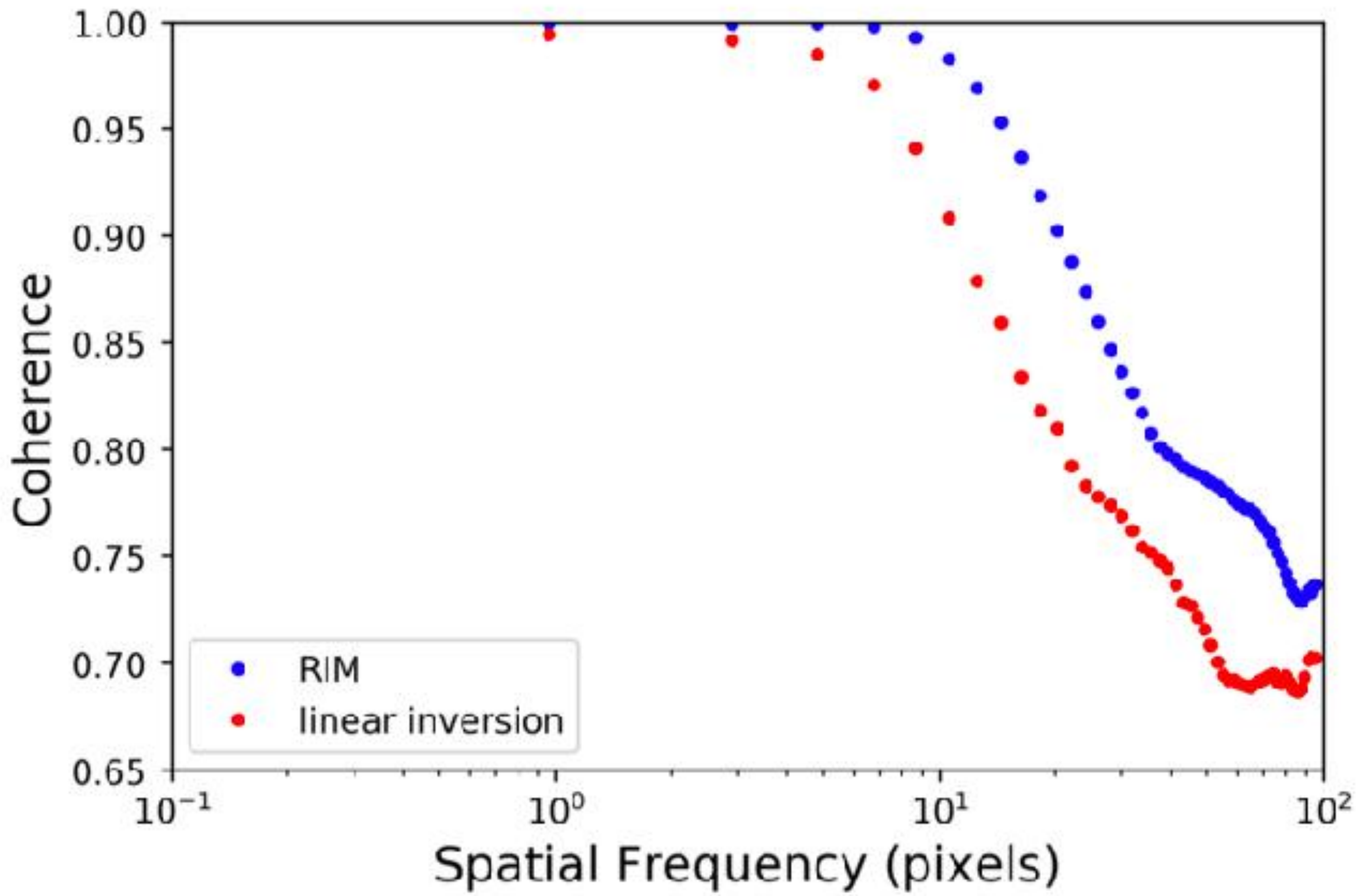
Morningstar, Perreault Levasseur et al., 2019

Undistorted image of the background source With the Recurrent Inference Machine (RIM)



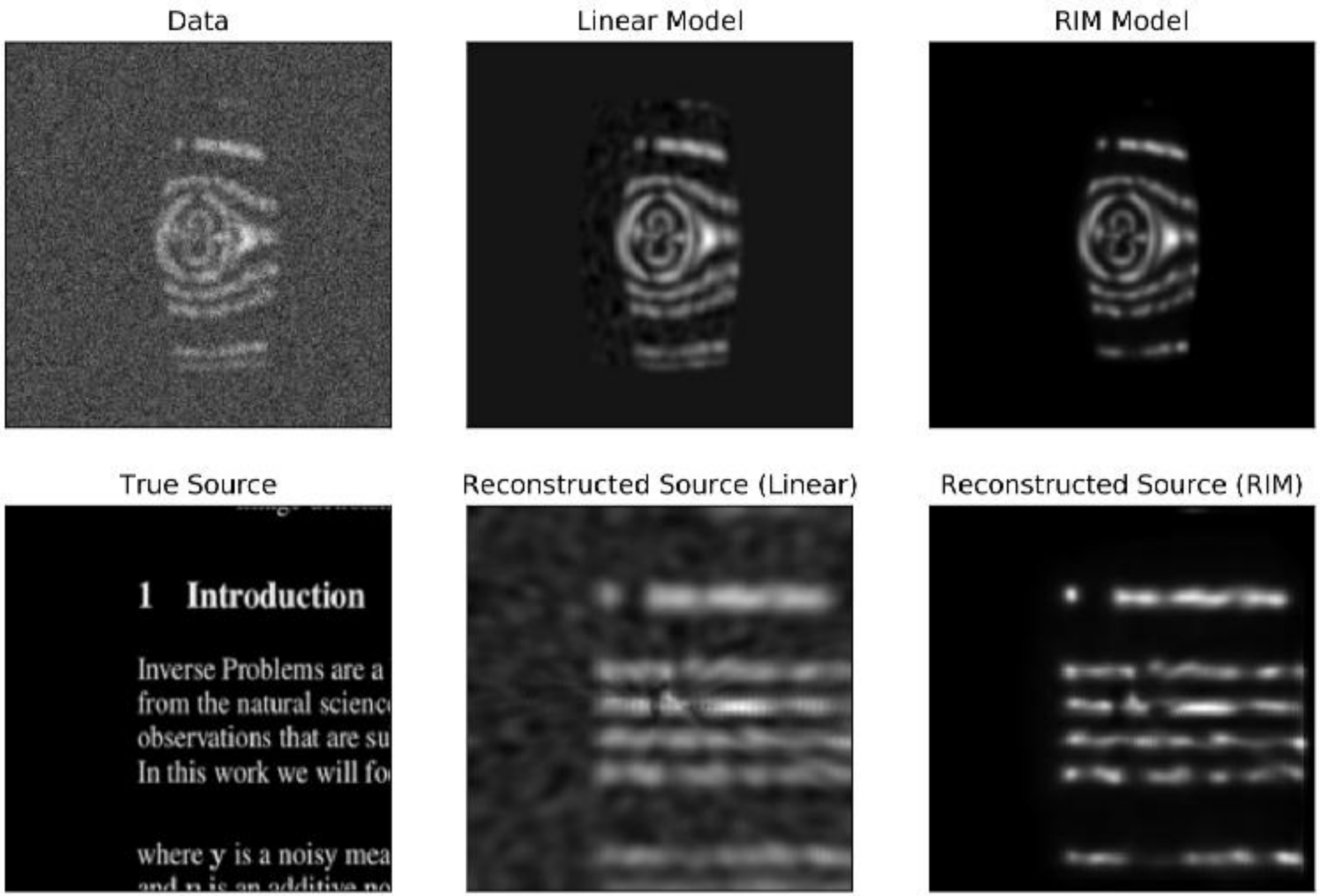
Morningstar, Perreault Levasseur et al., 2019

Background source reconstruction: Comparison to maximum likelihood methods

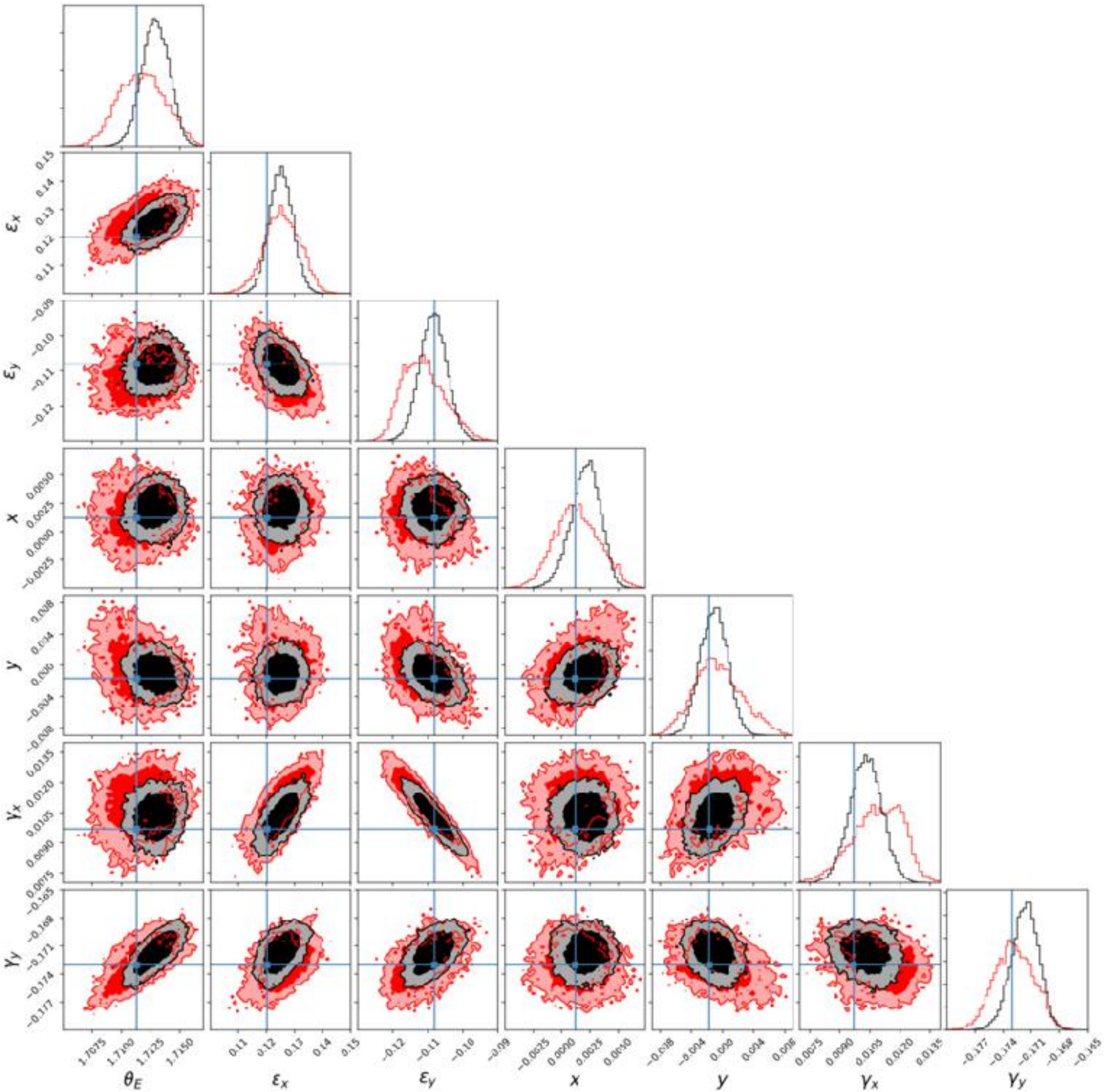


Morningstar, Perreault Levasseur et al., 2019

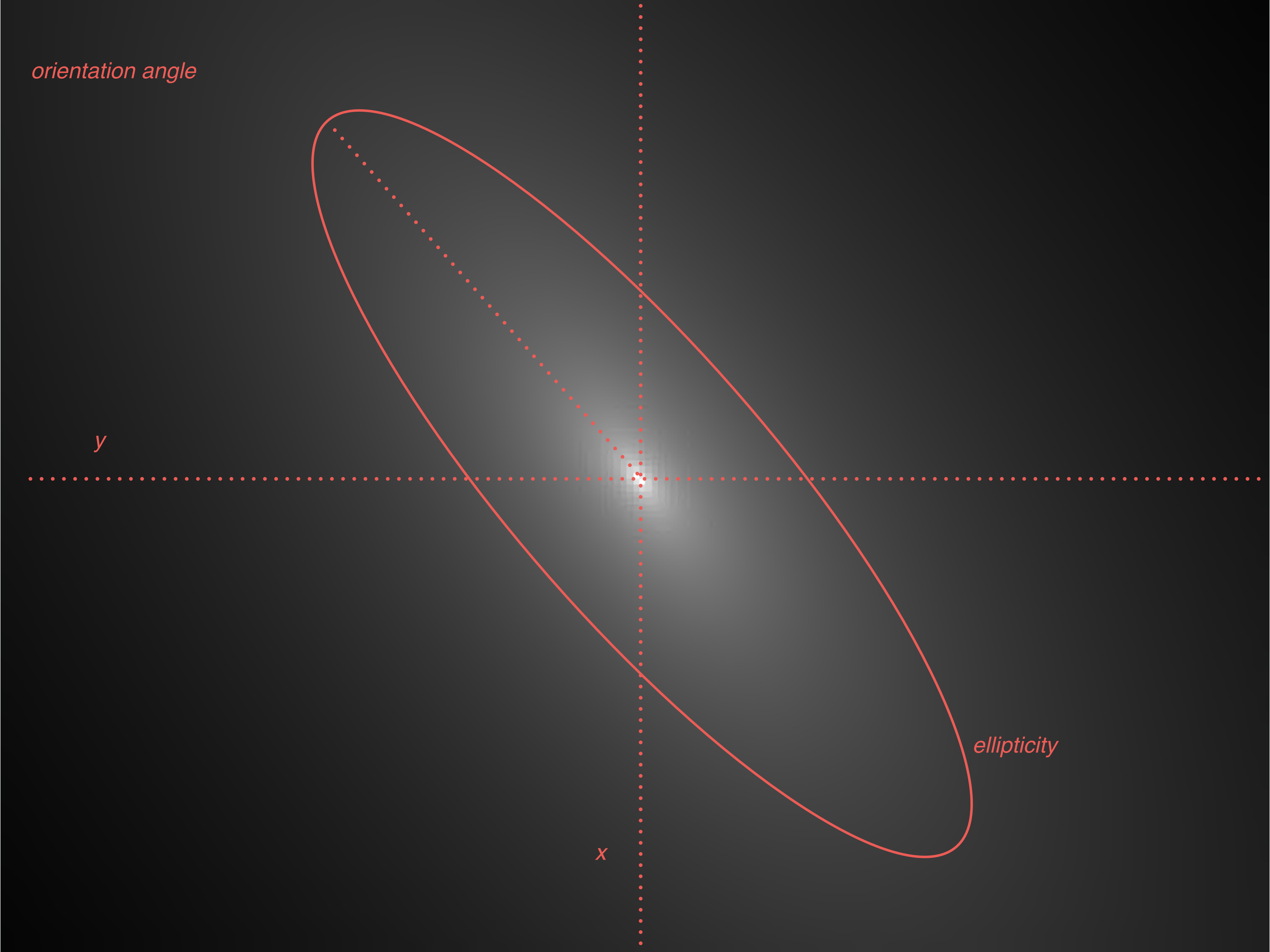
EXAMPLES OUTSIDE THE TRAINING DATA



Morningstar, Perreault Levasseur et al., 2019

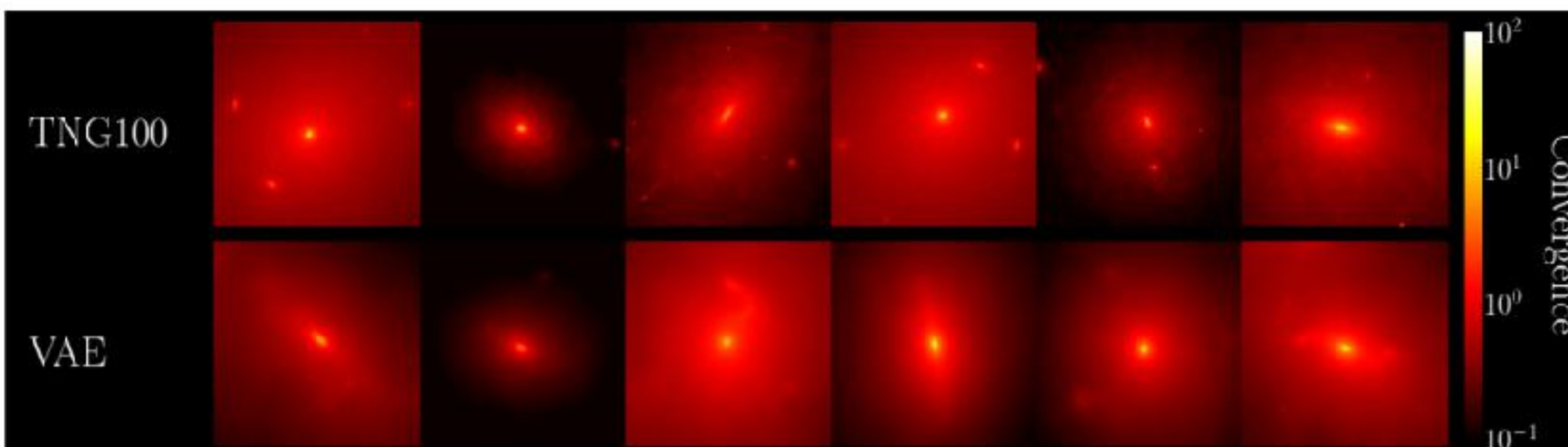


θ_E
ϵ_x
ϵ_y
x
y
γ_x
γ_y

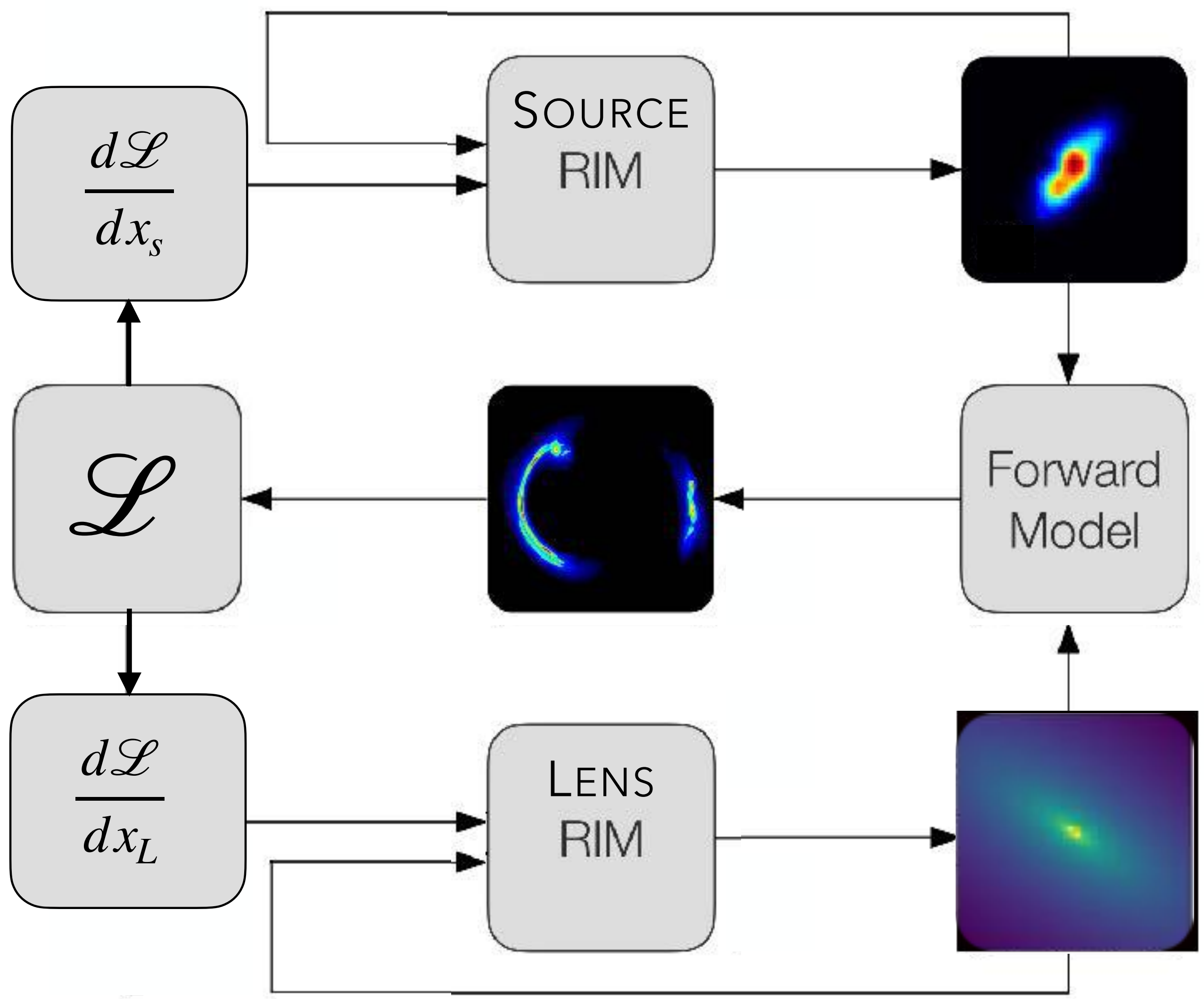

orientation angle
y
x
ellipticity

Simulated galaxies generated with a Variational Autoencoder

Alexandre Adam

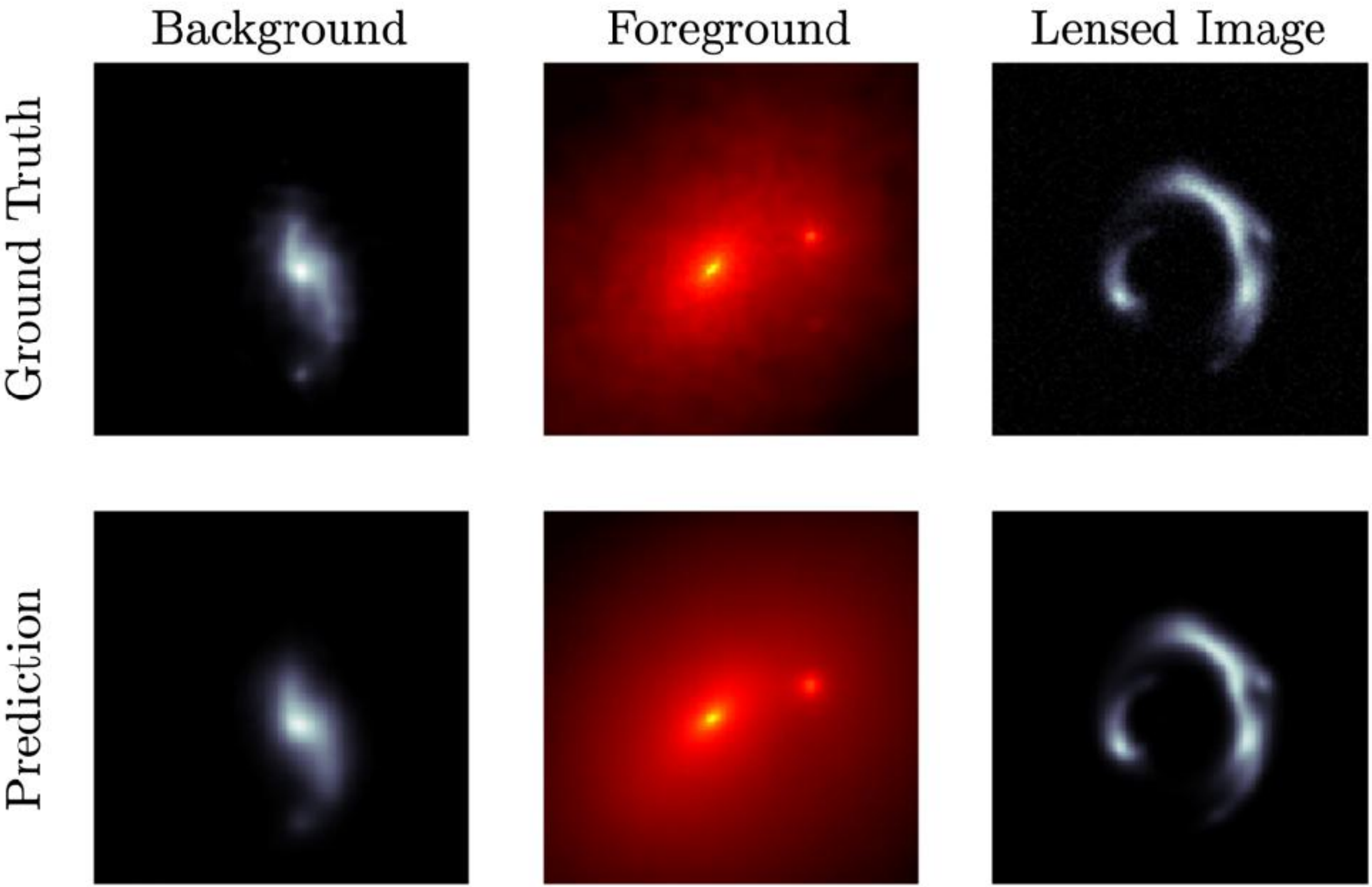


SOLVING ENTIRE LENSING SYSTEMS WITH THE RECURRENT INFERENCE MACHINE

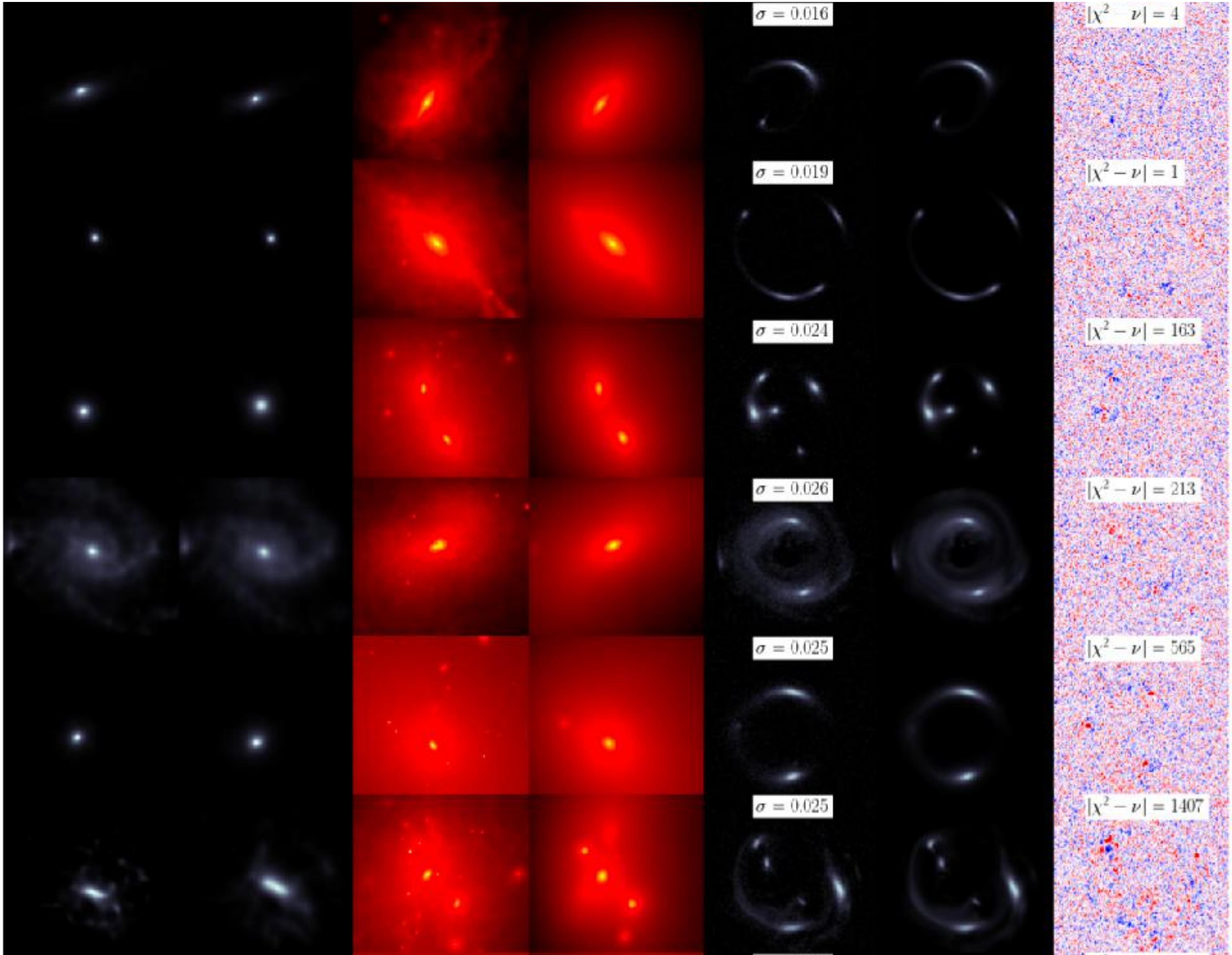


Adam, Perreault Levasseur, Hezaveh, 2023

TRAINING ON HYDRODYNAMICAL SIMULATIONS

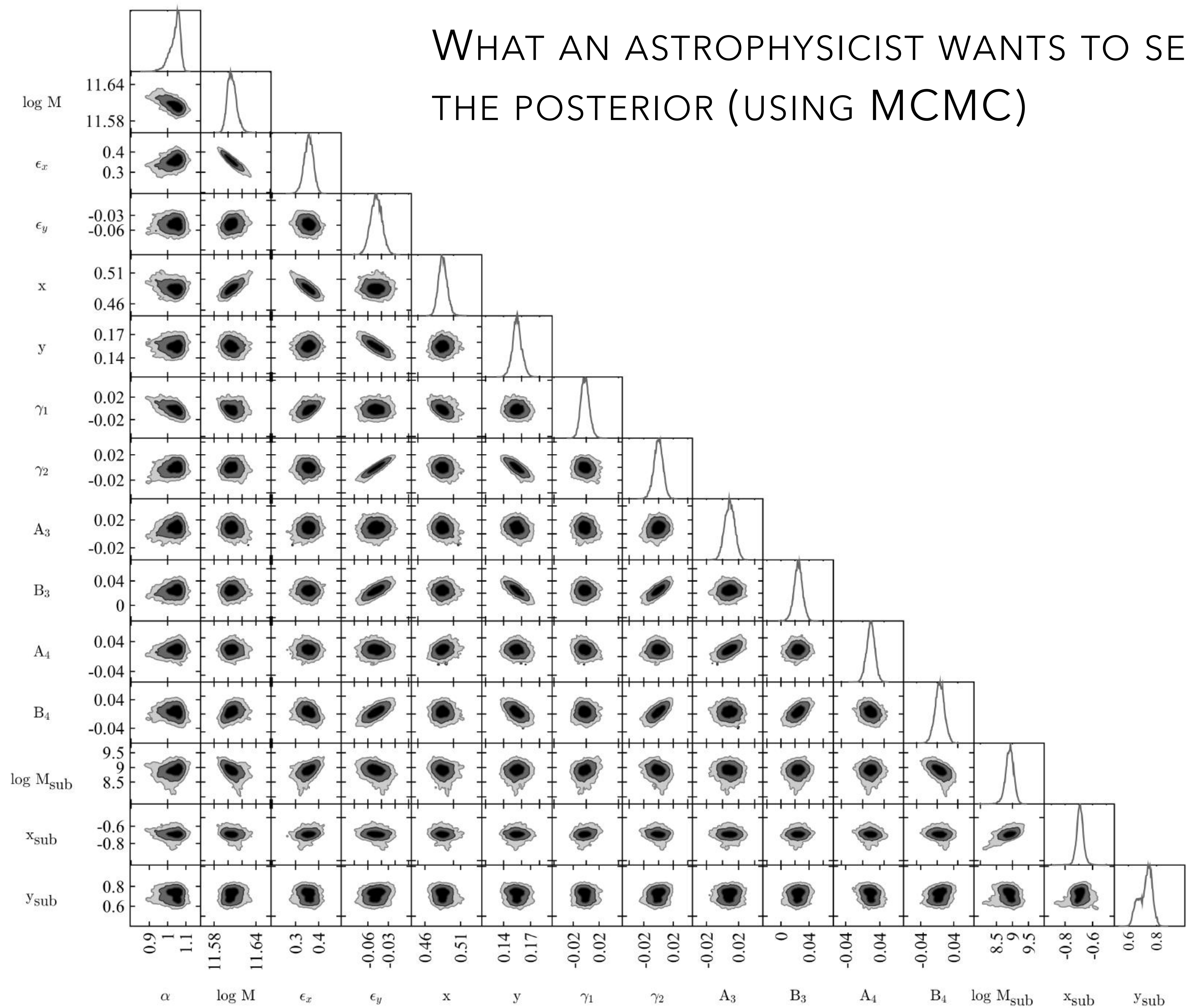


Adam, Perreault-Levasseur, Hezaveh., 2023



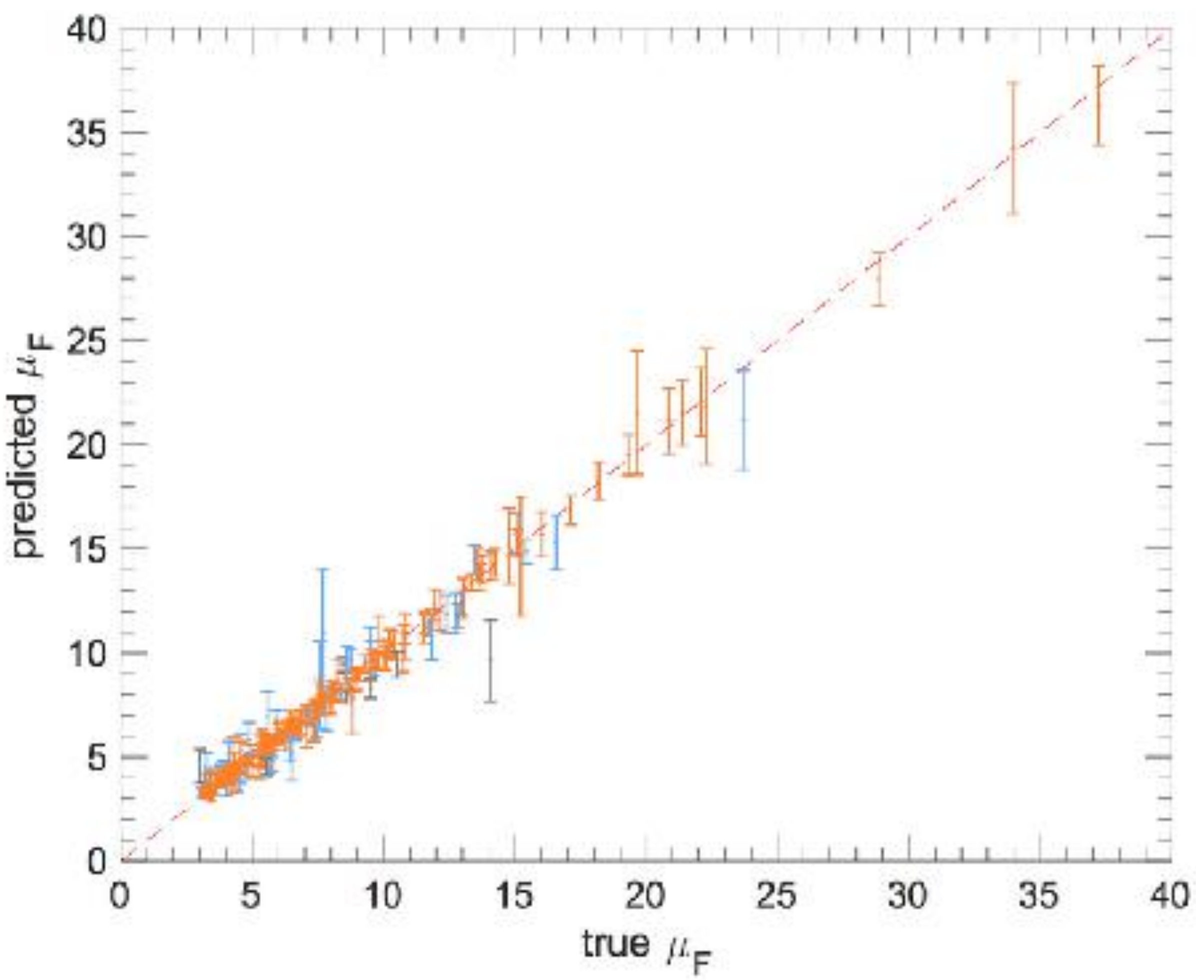
Adam, Perreault Levasseur, Hezaveh., 2023

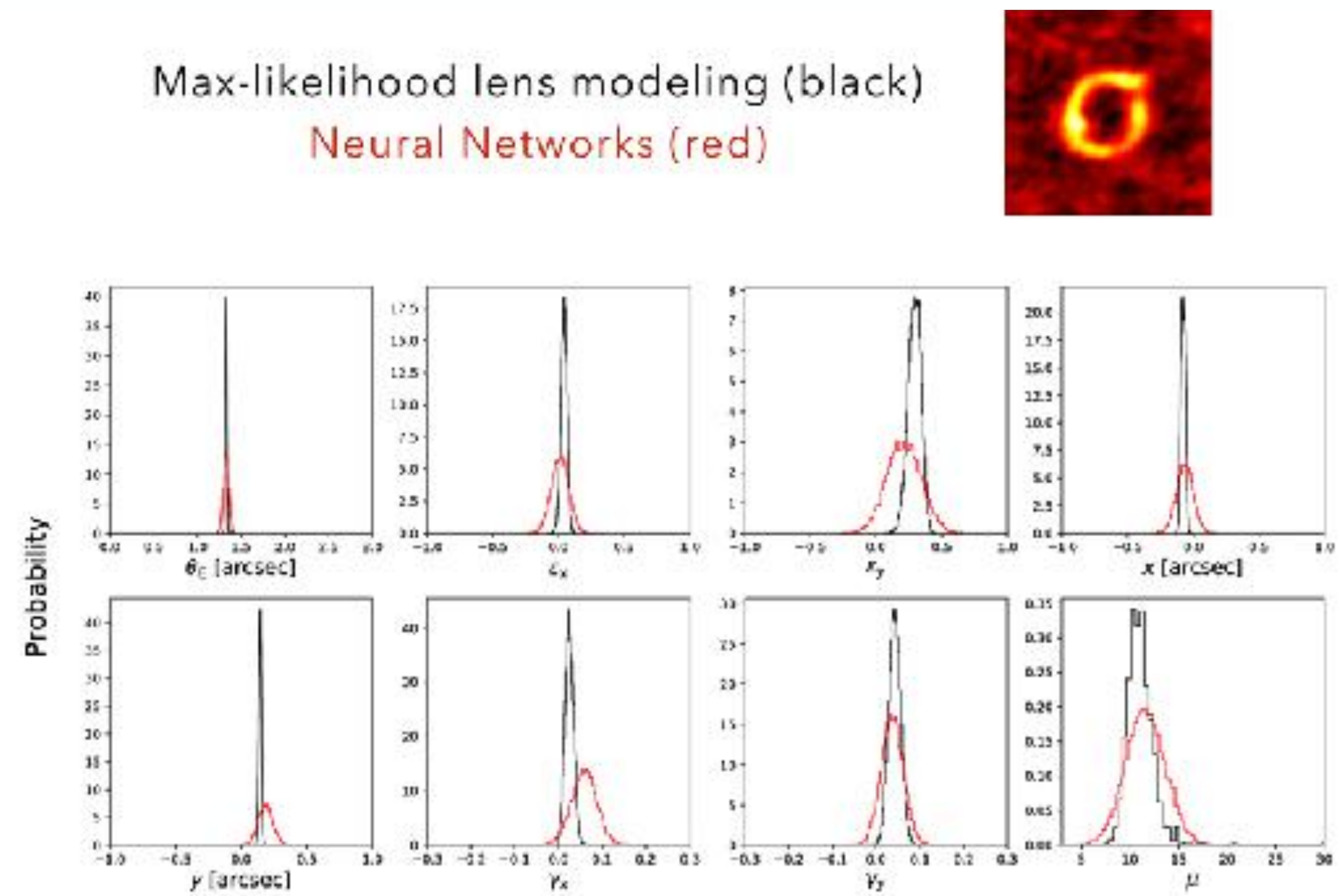
WHAT AN ASTROPHYSICIST WANTS TO SEE: THE POSTERIOR (USING MCMC)



Hezaveh, …,LPL, et al. ApJ 2016

UNCERTAINTY ESTIMATION WITH APPROXIMATE BAYESIAN NEURAL NETWORKS





Perreault-Levasseur et al., 2017

Morningstar et al., 2018

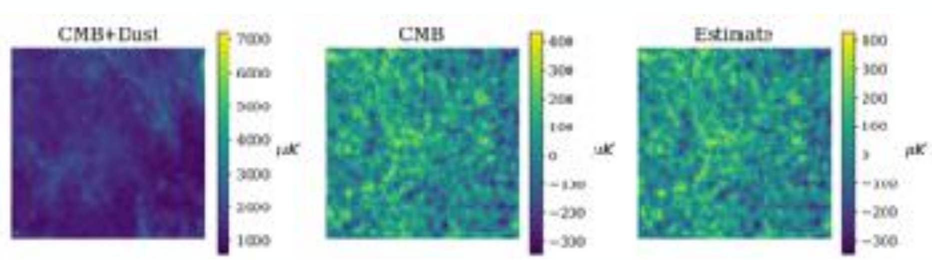
Variational Inference to Approximate Bayesian Neural Networks

Pros:

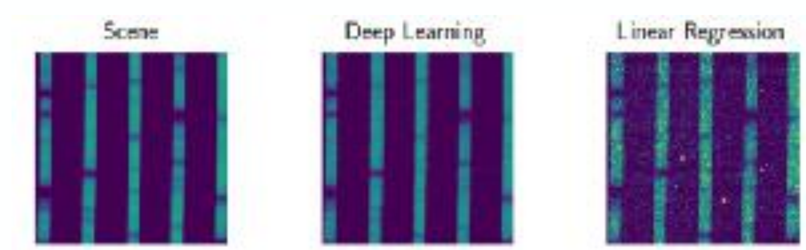
- Amortized.
- Requires few hundred forward passes at evaluation time (to collect samples). Still very fast.
- Marginalizes implicitly over parameters we do not wish to explicitly model.
- With good coverage probabilities, one can use importance sampling of the output distribution to get an unbiased posterior. (Provided one can actually write this posterior)

Caveats:

- The variational distributions (Bernouilli) are extremely simplistic, therefore even if we attempt to use them to approximate the true weight distributions, that approximation could be bad and yield inaccurate uncertainties.



CMB Cleaning



IR Spectrometer De-noising

Same problem remains regardless of the variational distribution used: there is no way of quantifying how well we approximate the true weight distributions

Simulation-Based inference

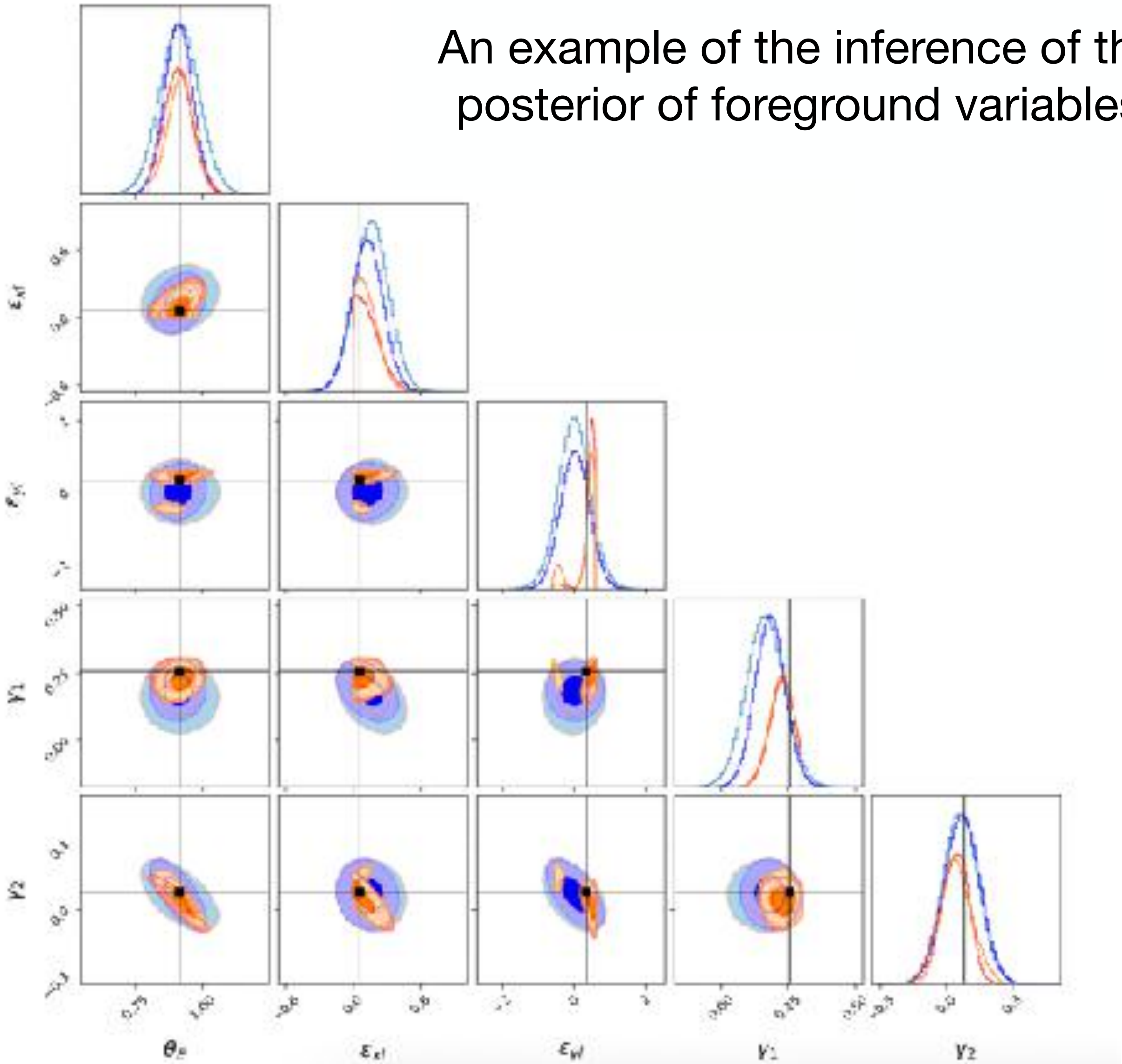
$$P(\theta \,|\, \hat{\theta}_x)$$

Legin, Hezaveh, Perreault Levasseur, Wandelt
NeurIPS 2021 - Physical Sciences Workshop

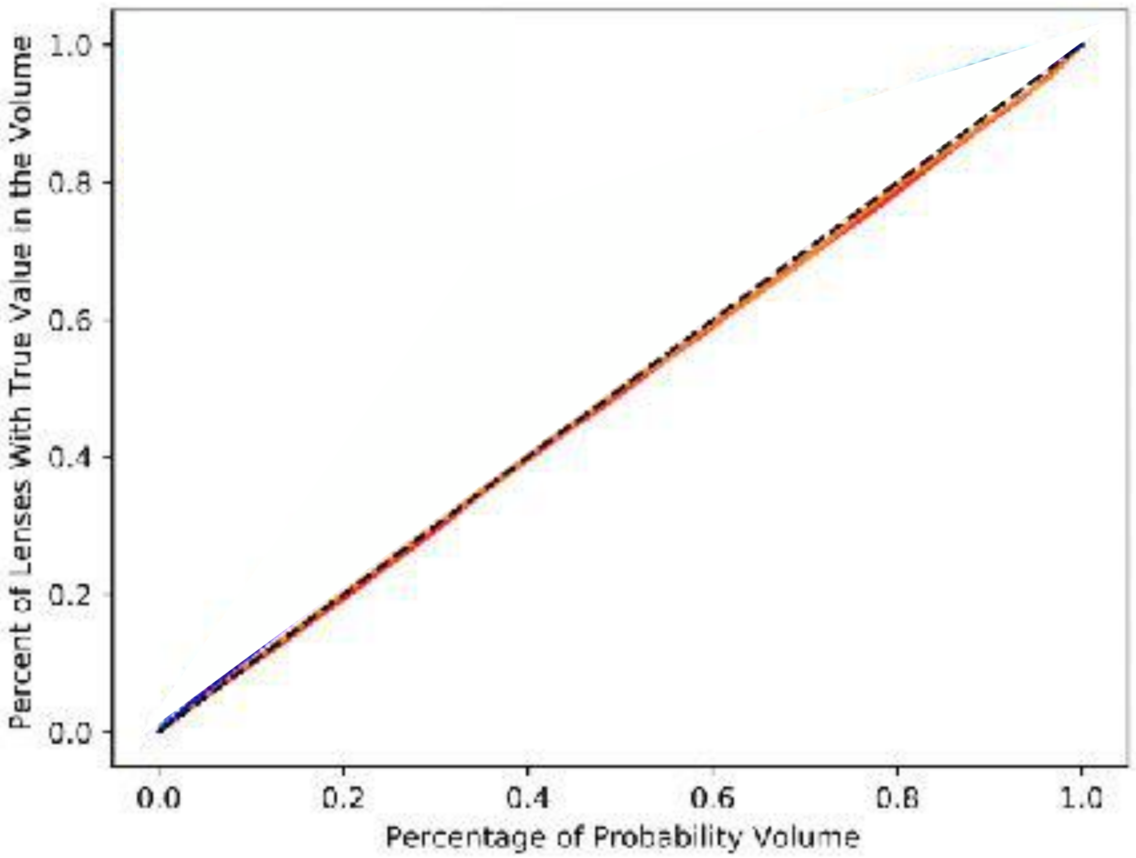
UNCERTAINTY ESTIMATION WITH SIMULATION-BASED INFERENCE METHODS

Ronan Legin

An example of the inference of the posterior of foreground variables



Coverage probabilities



R. Legin, Y. Hezaveh, L. Perreault Levasseur, B. Wandelt, ApJ 2022

Uncertainty estimation with simulation-based inference methods

Pros:

- Use the **power of ML to find a compress statics**, and even if it is biased, we can get unbiased error estimate, the only drawback would be sub-optimal precision. (Provided the simulation pipeline is accurate!)
- A well-defined statistical framework that can: be relatively **fast**, deal with **complex distributions**, model **joint posteriors**.
- Use a neural density estimator to get the joint distribution p(data, parameter), no need for the epsilon parameter in ABC.
- Can **change the prior** from data point to data point without retraining the ML compressor.
- Once we have the posterior, can **generate samples** that are consistent with data (this is really important for 'interrogating the black box')

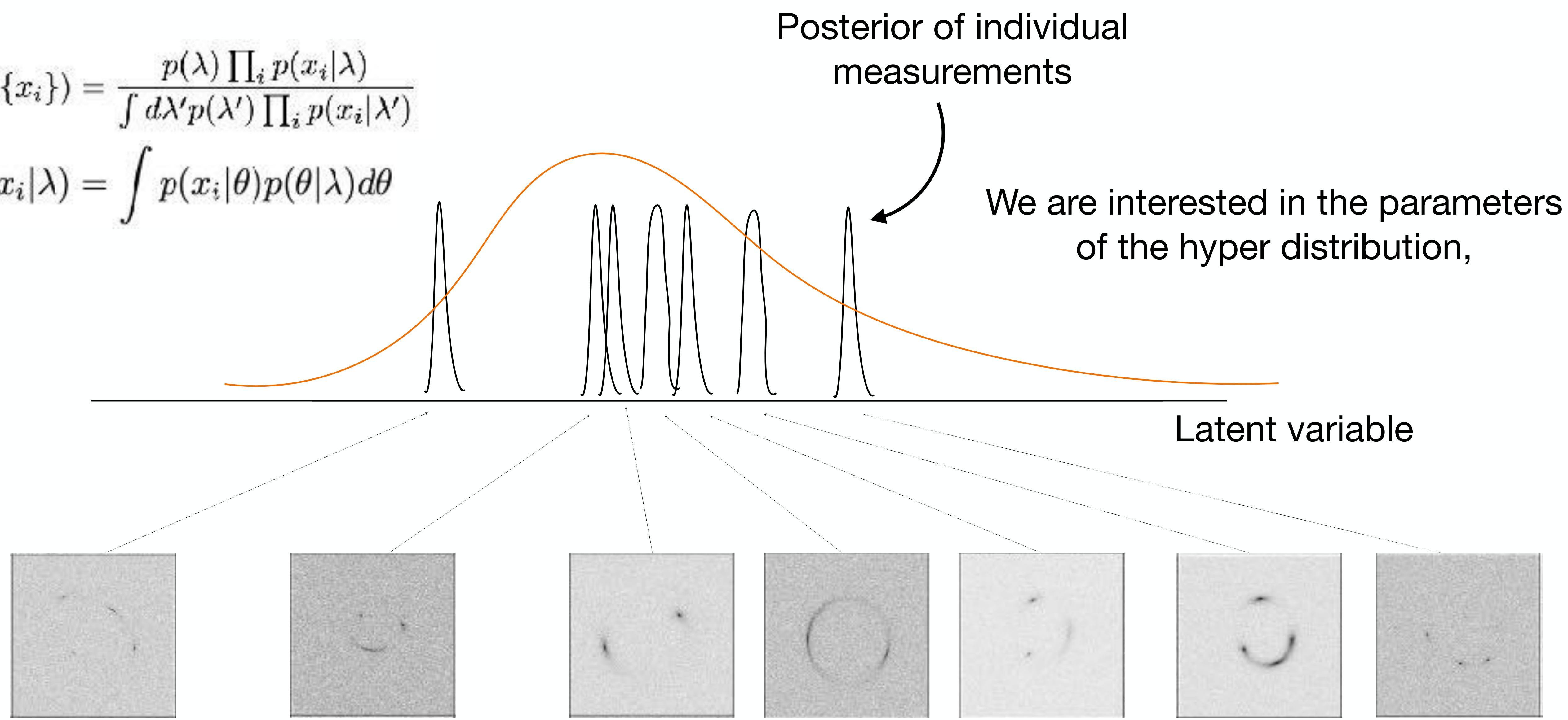
Caveats:

- Hard to marginalize implicitly over parameters, we need to **explicitly** model them.
- We don't model the uncertainty of the density estimator itself. (But it's a fairly simple ML model, and except for very pathological problems it's reasonable to expect that we are in interpolation mode).
- Limited to **low-dimensional posteriors** (10s maximum).
- Requires an **accurate simulation pipeline**.

Hierarchical Bayesian inference

$$p(\lambda|\{x_i\}) = \frac{p(\lambda)\prod_i p(x_i|\lambda)}{\int d\lambda' p(\lambda')\prod_i p(x_i|\lambda')}$$

$$p(x_i|\lambda) = \int p(x_i|\theta)p(\theta|\lambda)d\theta$$



Hierarchical Bayesian inference

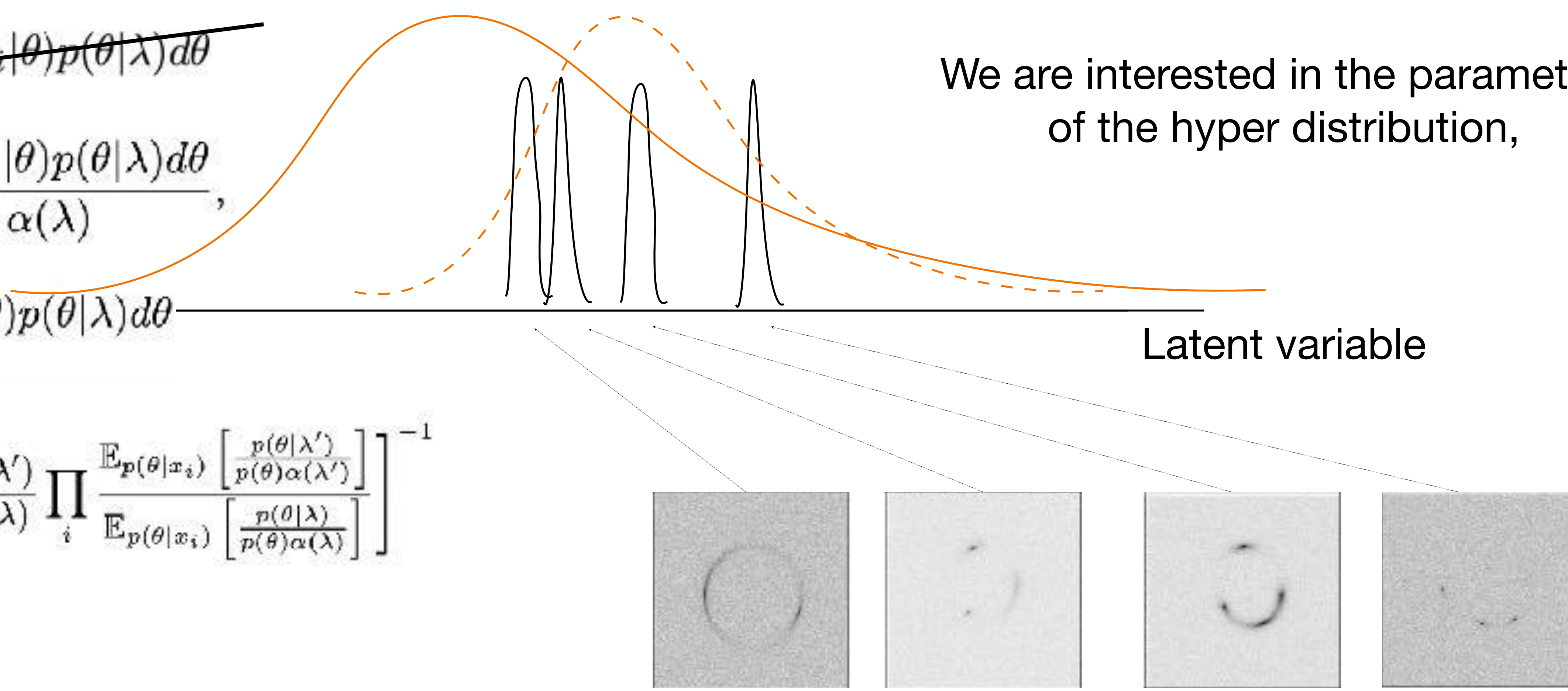
$$p(\lambda|\{x_i\}) = \frac{p(\lambda)\prod_i p(x_i|\lambda)}{\int d\lambda' p(\lambda')\prod_i p(x_i|\lambda')}$$

~~$$p(x_i|\lambda) = \int p(x_i|\theta)p(\theta|\lambda)d\theta$$~~

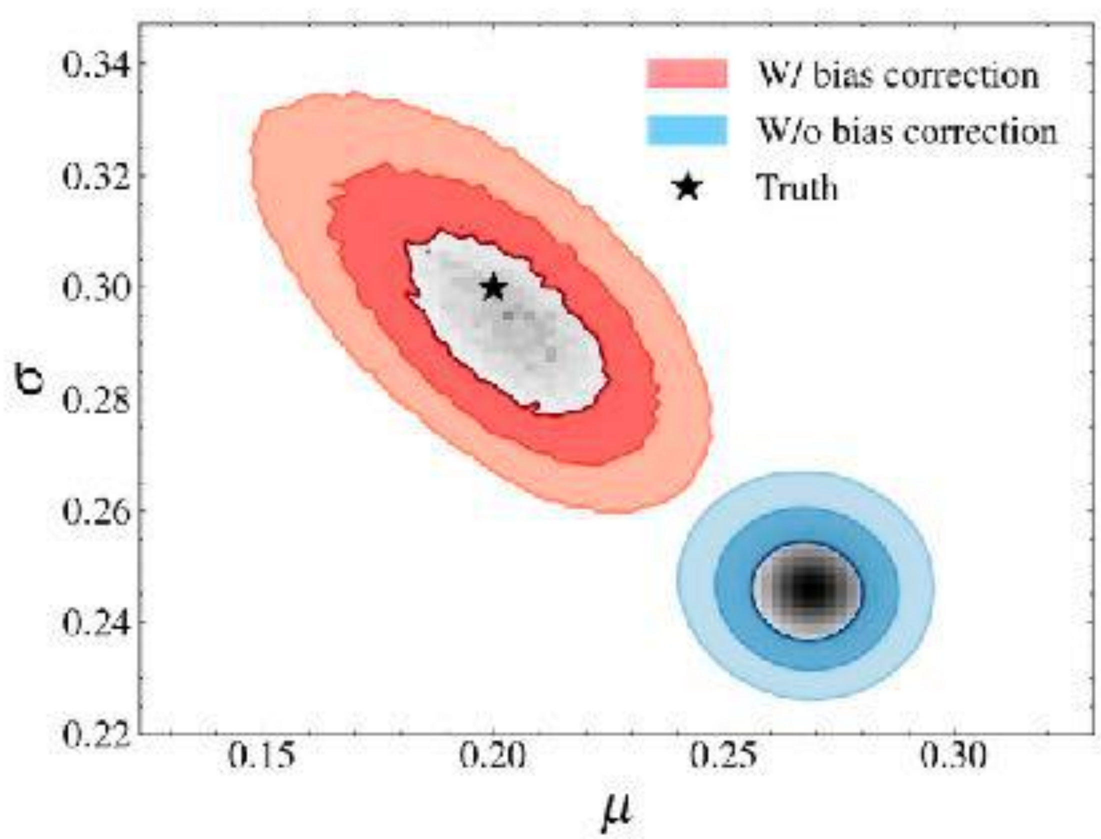
$$p(x_i|\lambda) = \frac{\int p(x_i|\theta)p(\theta|\lambda)d\theta}{\alpha(\lambda)},$$

$$\alpha(\lambda) = \int p_{\rm det}(\theta)p(\theta|\lambda)d\theta$$

$$p(\lambda|\{x_i\}) = \left[\int d\lambda' \frac{p(\lambda')}{p(\lambda)}\prod_i \frac{\mathbb{E}_{p(\theta|x_i)}\left[\frac{p(\theta|\lambda')}{p(\theta)\alpha(\lambda')}\right]}{\mathbb{E}_{p(\theta|x_i)}\left[\frac{p(\theta|\lambda)}{p(\theta)\alpha(\lambda)}\right]}\right]^{-1}$$



Hierarchical Bayesian inference

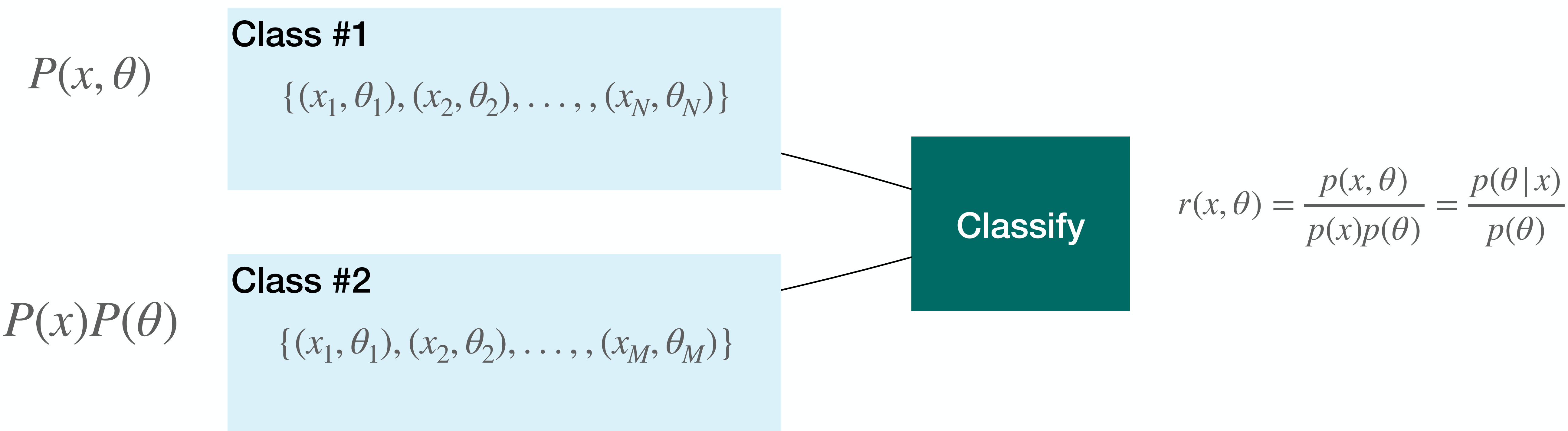


Ronan Legin

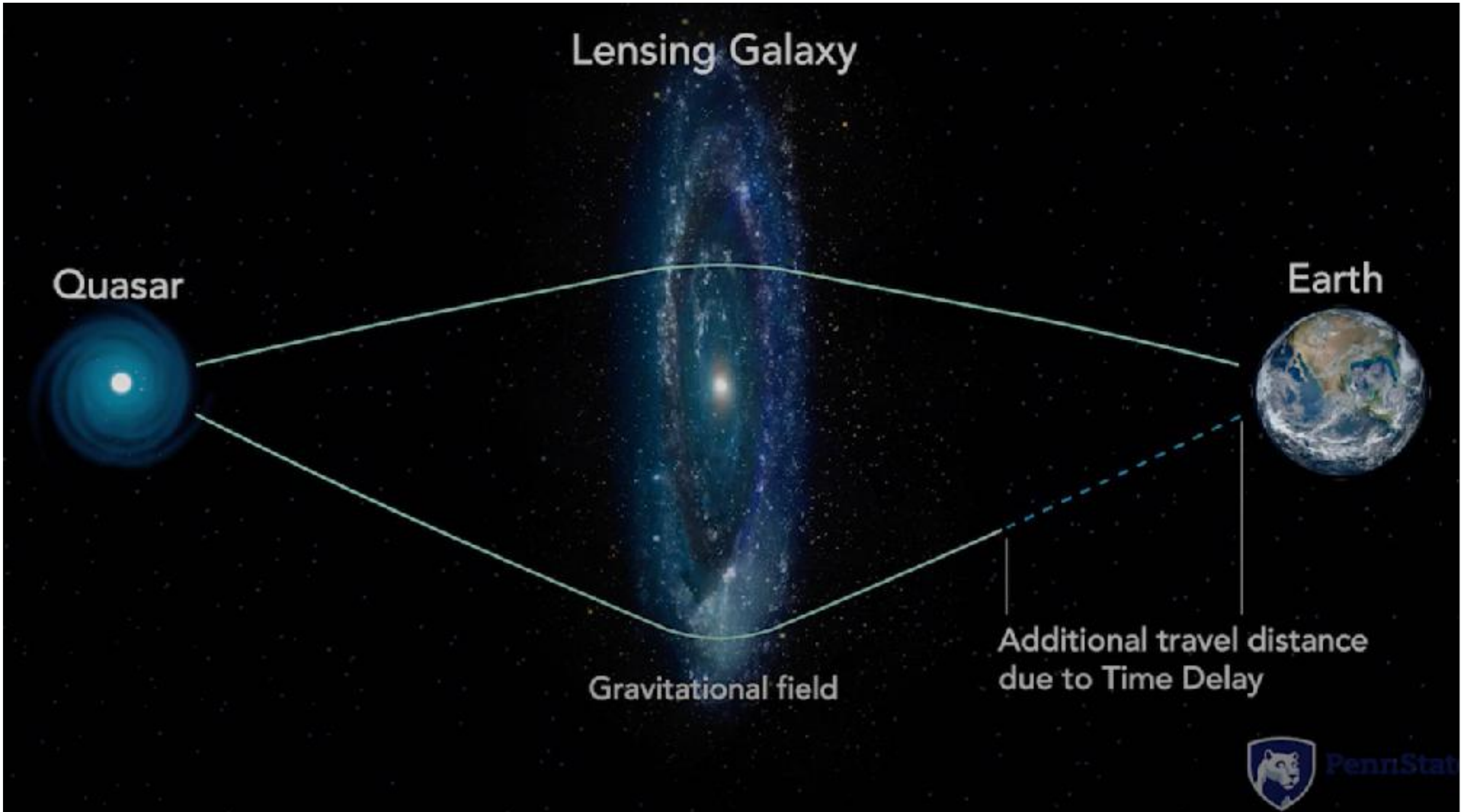
Connor Stone

Legin, Stone, Hezaveh, Perreault Levasseur, ICML 2022 - Machine Learning for Astrophysics Workshop

Neural Ratio Estimators

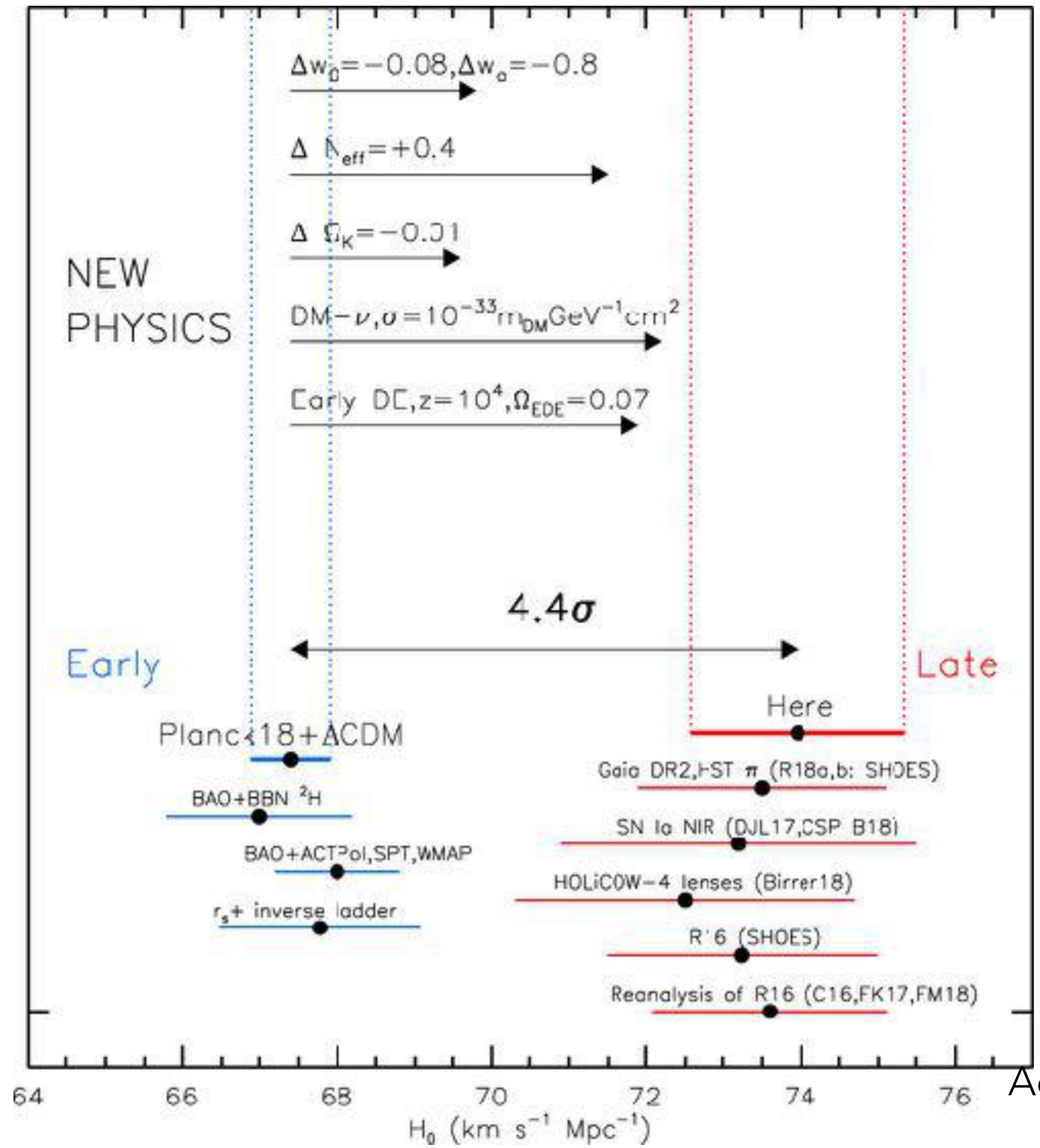


H0 Inference with Time Delay Cosmography



THE HUBBLE CONSTANT

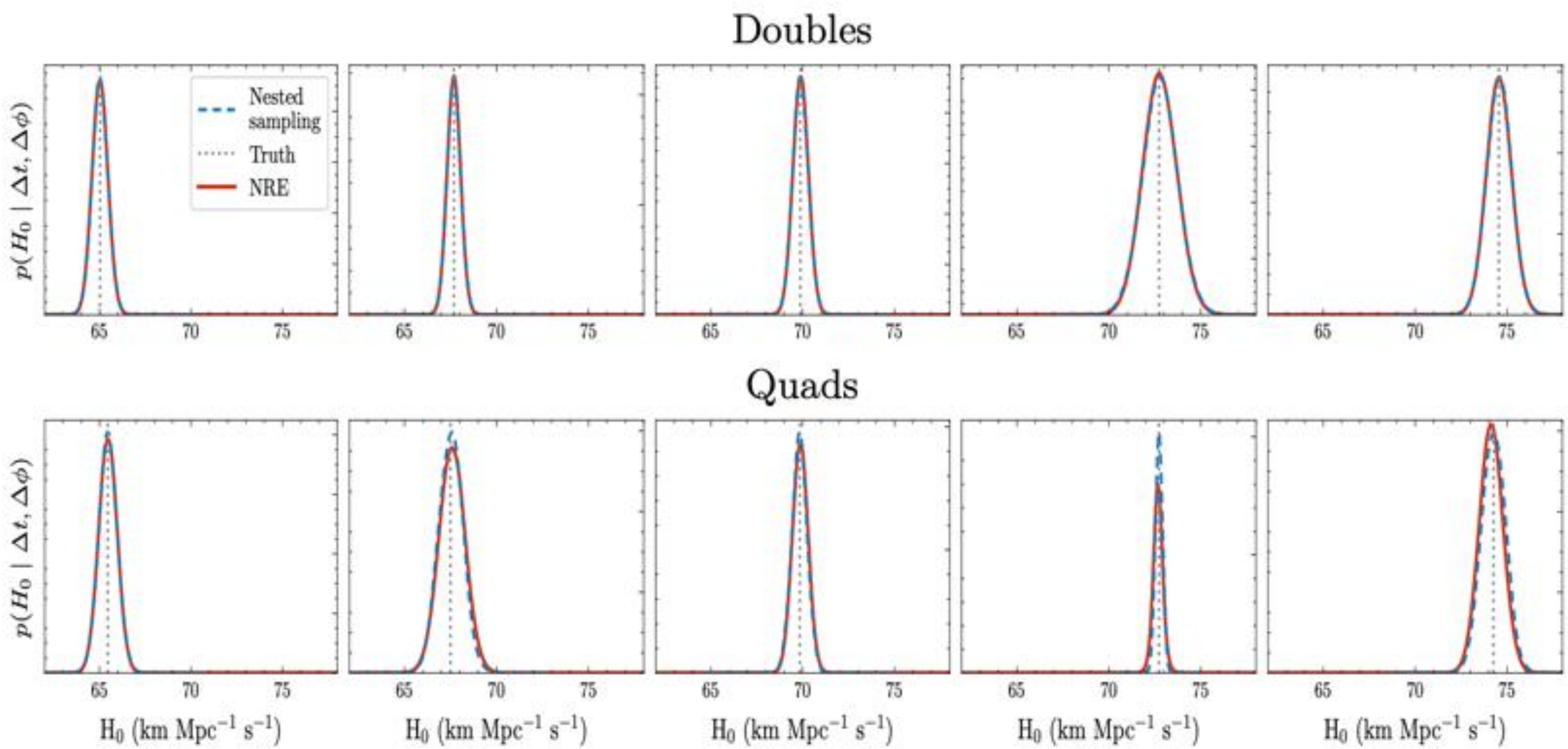
DISCREPANCY BETWEEN MEASUREMENTS



Adam G. Riess *et al* 2019
ApJ **876** 85

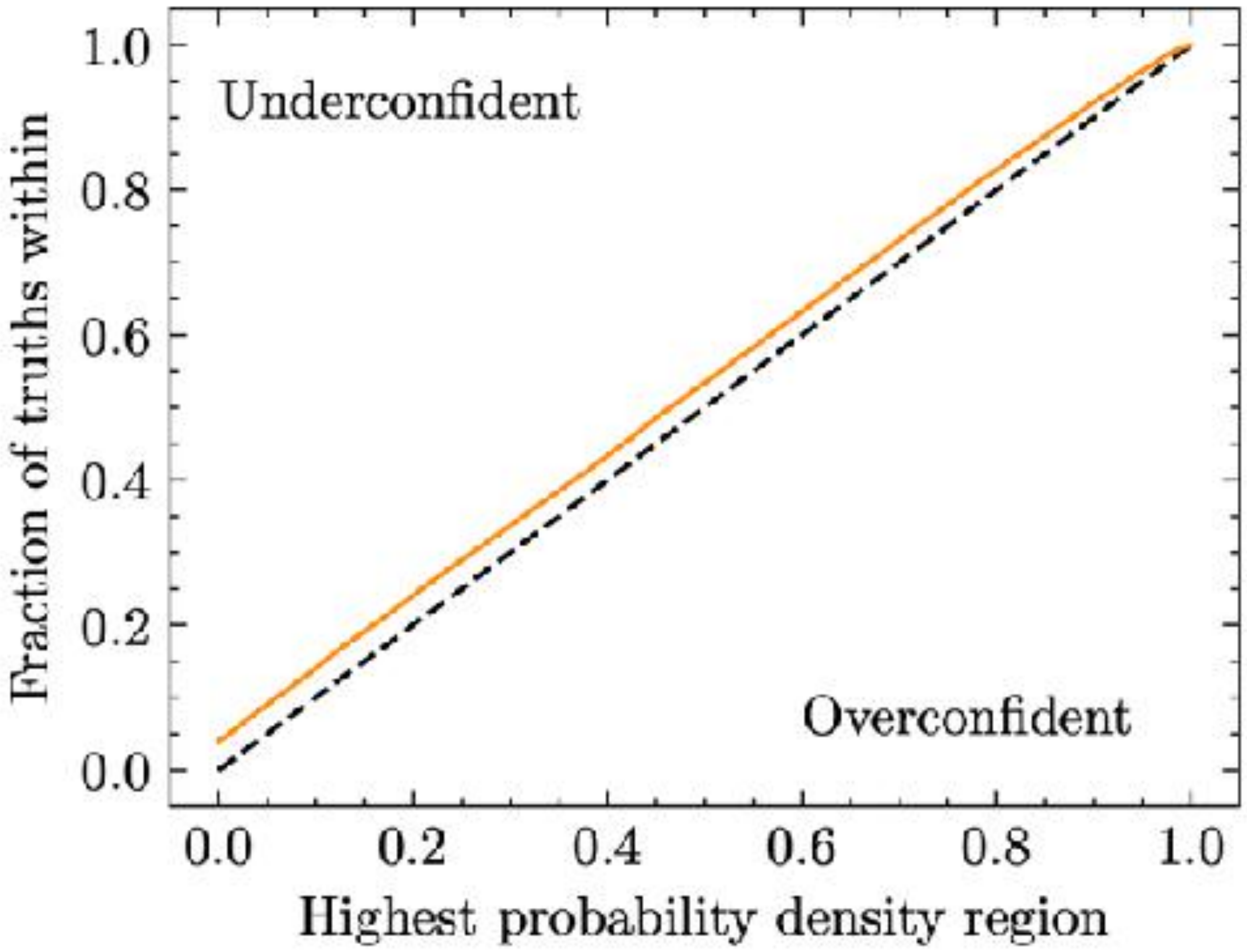
H0 Inference with Neural Ratio Estimators

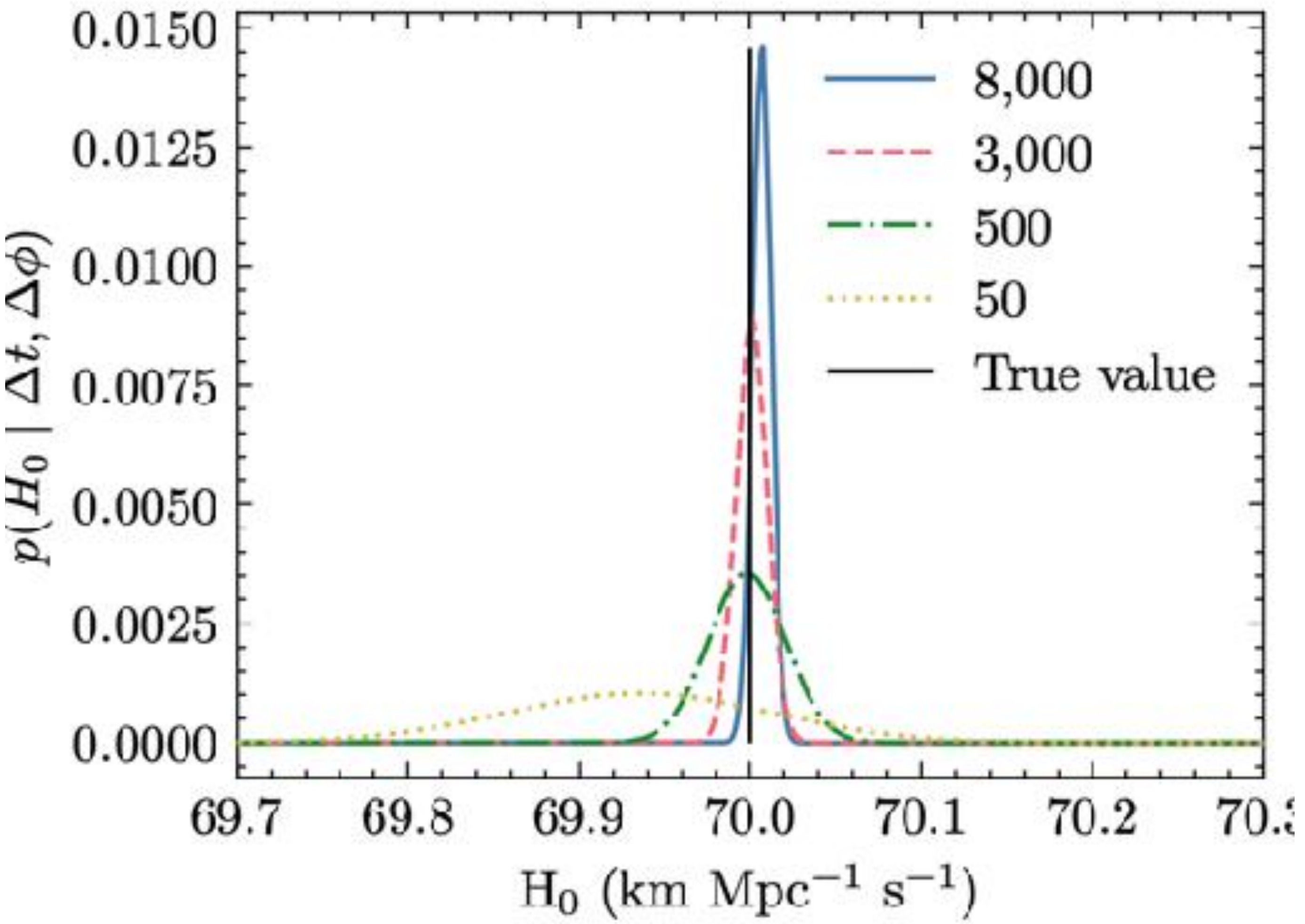
Ève Campeau-Poirier



H0 Inference with Neural Ratio Estimators

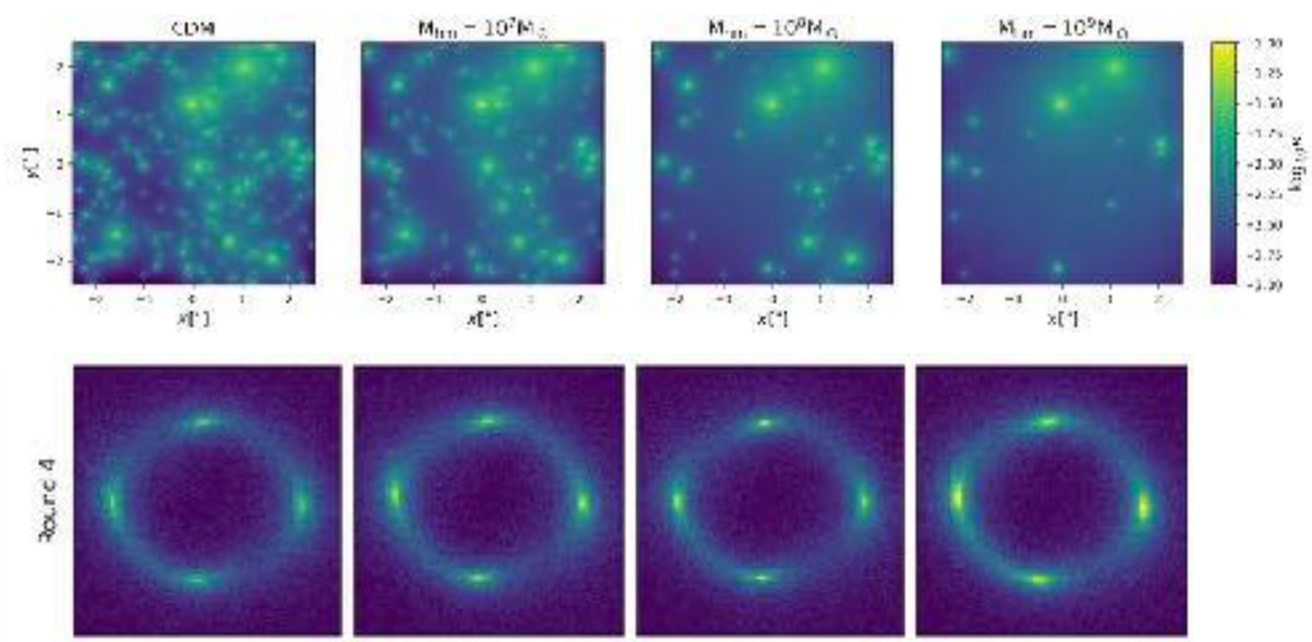
Ève Campeau-Poirier

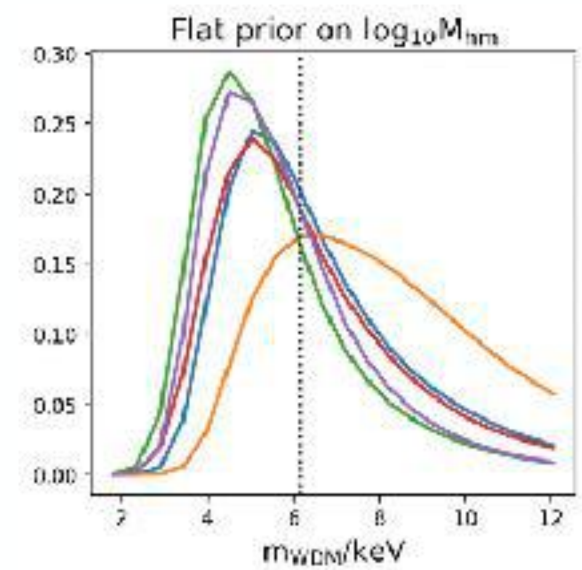




Campeau-Poirier et al. ICML 2023 ML4Astro Workshop

Estimating the dark matter particle temperature with Neural Ratio Estimators





Adam Coogan

Anau Montel, Coogan et al. 2022
Coogan et al. , NeurIPS 2020 ML4PS Workshop

Ratio estimation methods

Pros:

- Can marginalize implicitly over large number of nuisance parameters

Caveats:

- Because we have marginalized, we've lost the capability to generate samples consistent with the observations.

- So far: no real way of quantifying the uncertainty of the ratio estimator itself. All the guarantees are in terms of convergence to a specific ratio in the limit of perfect training. Is this always realistic?

TACKLING AN UNSOLVED PROBLEM:
HIGH DIMENSIONAL INFERENCE

A previously unsolved problem in all of astrophysics (and other sciences):

How do we infer the posteriors of high-dimensional parameters (e.g., an image or spectra)?

Obstacles:

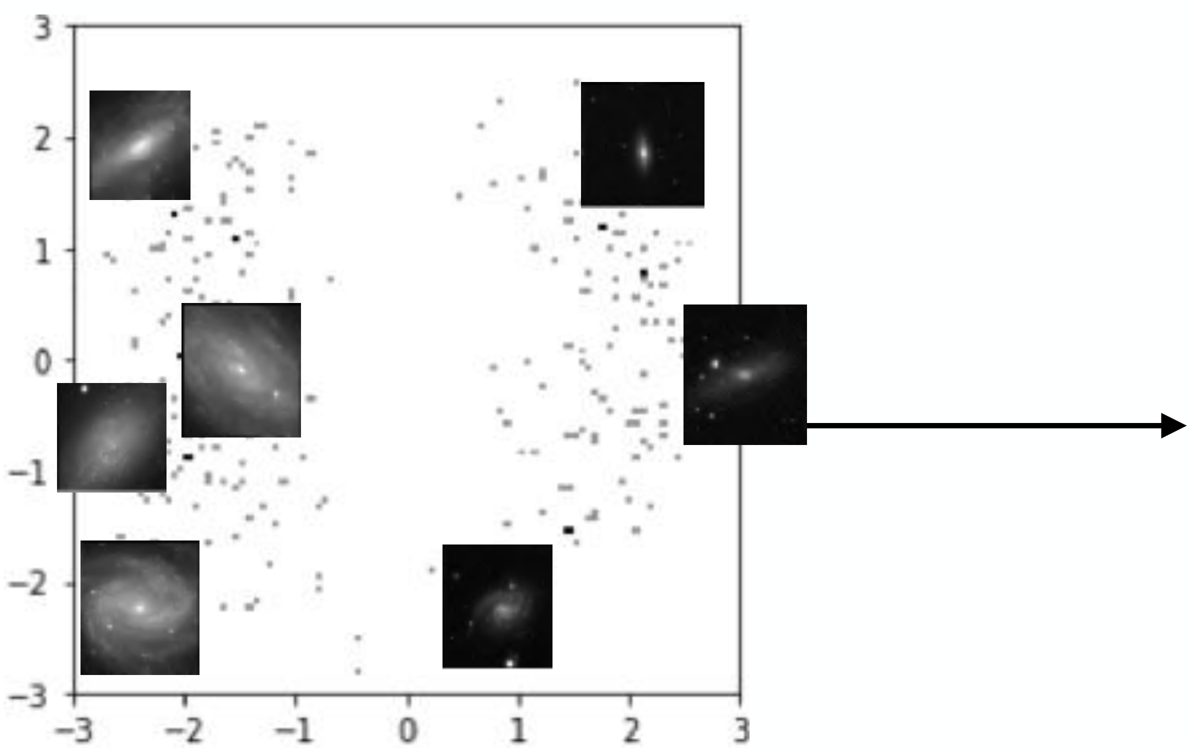
1) How do we encode complex priors

2) How we sample such high-dimensional posteriors (even if we could compute them)

LEARNING THE PRIOR EXPLICITLY

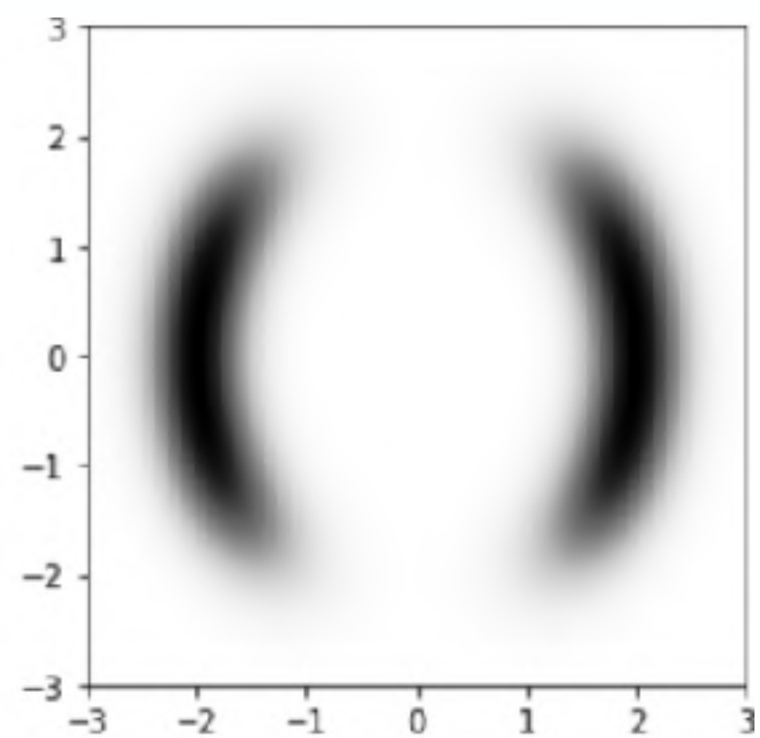
Can we learn our high-dimensional prior explicitly from data?
i.e. can we learn a generative model that will produce samples from that distribution?

How can we do this from samples (e.g. data)? Modeling the density?



Training data

$\{\mathbf{x}_1, \ldots, \mathbf{x}_n,\} \overset{iid}{\sim} \pi_{data}(\mathbf{x})$



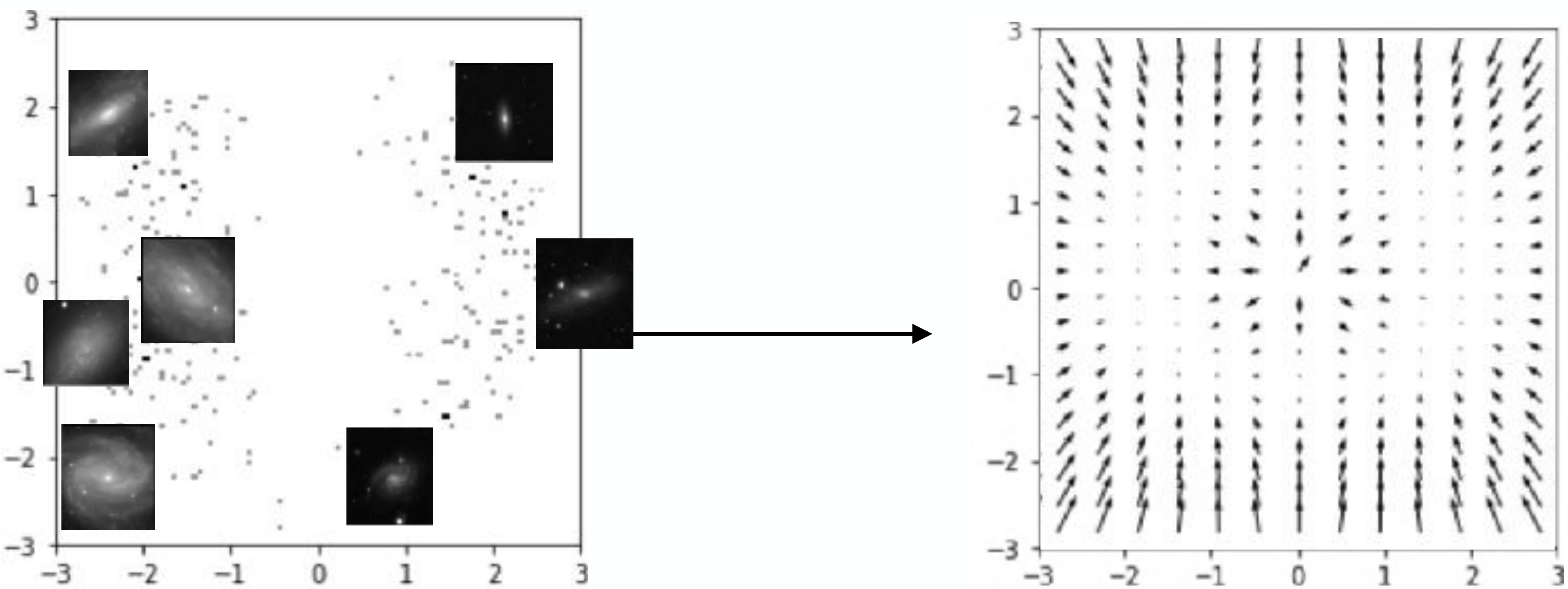
Density function

$p_\theta(\mathbf{x}) \approx \pi_{data}(\mathbf{x})$

Score Modeling

Turns out that if I want to sample a distribution, the only thing I need to learn is its ***score***, which does not include the normalization constant and only uses local information

$$\mathbf{s}(\mathbf{x}) = \nabla_{\mathbf{x}} \log(\pi(\mathbf{x}))$$



Training data

$\{\mathbf{x}_1, \ldots, \mathbf{x}_n,\} \overset{iid}{\sim} \pi_{data}(\mathbf{x})$

Score function

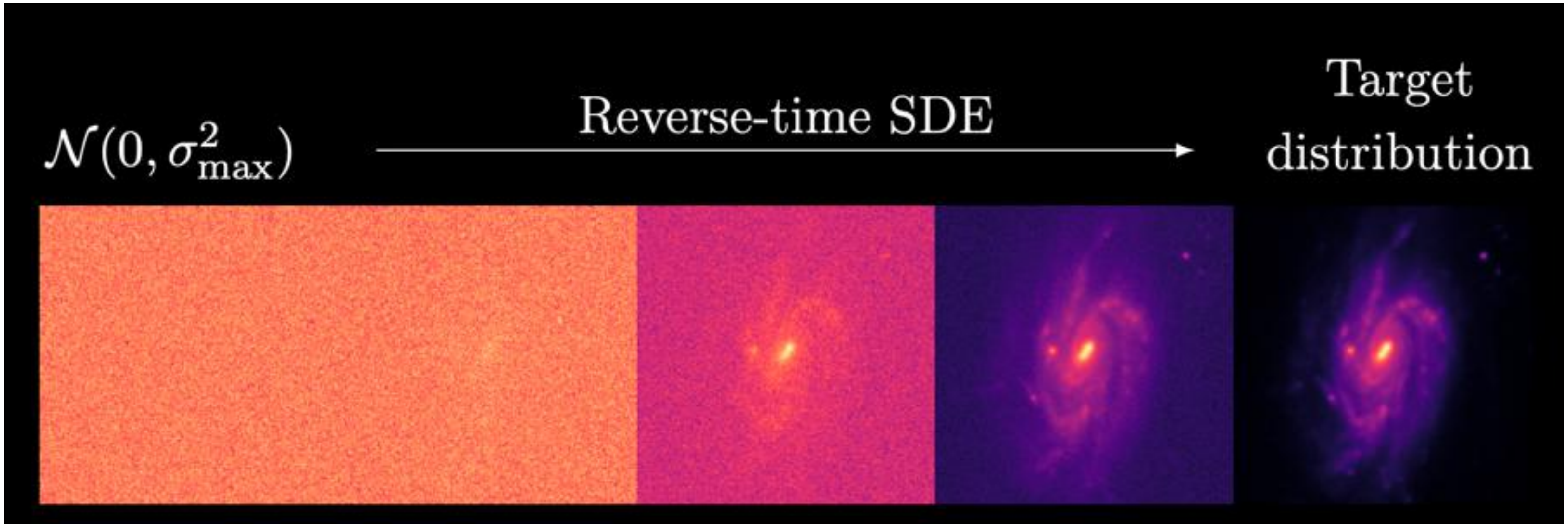
$\mathbf{s}_\theta(\mathbf{x}) \approx \nabla \log(\pi(\mathbf{x}))$

Score-Based Modeling

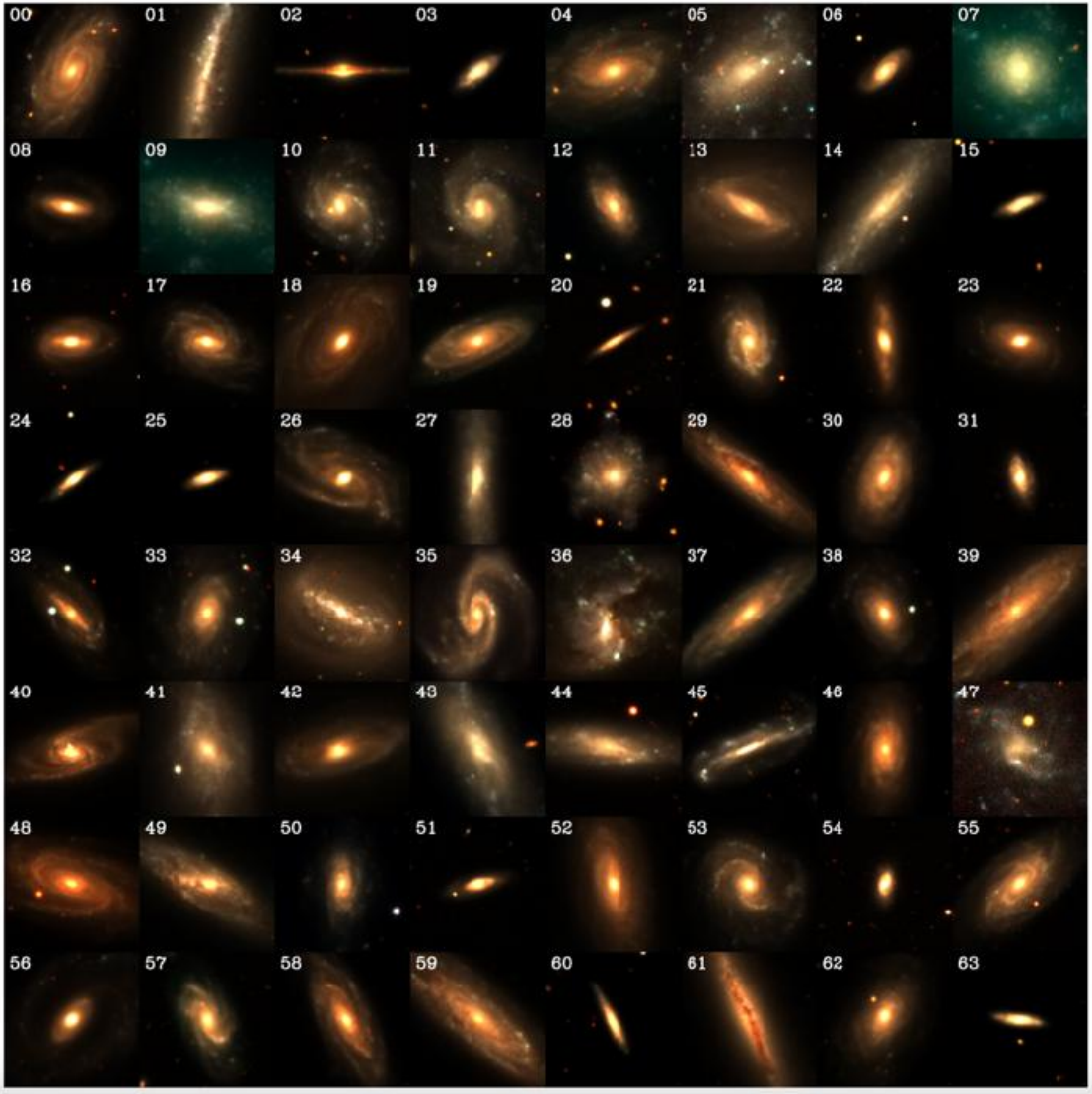
Alexandre Adam

We model the score of the prior

$$s_\theta(x) \equiv \nabla_x \log p_\theta(x)$$



Adam et al. NeurIPS 2022 ML4PS workshop

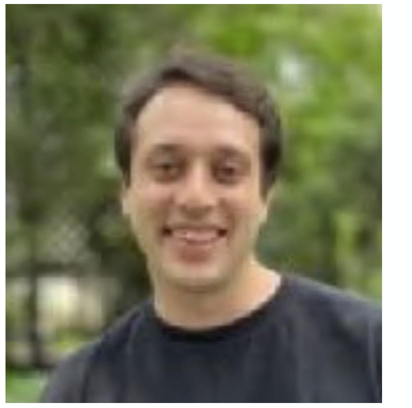


Connor Stone

http://www.mjjsmith.com/thisisnotagalaxy/

Smith et al. arXiv:2111.01713

Score-based Modeling


Alexandre Adam

To sample from the posterior, the score of the likelihood is all that we need:

$$\nabla_x \log p(x \mid y) = \nabla_x \log p(y \mid x) + \nabla_x \log p_\theta(x)$$

We can calculate the likelihood score analytically if we assume it's Gaussian.

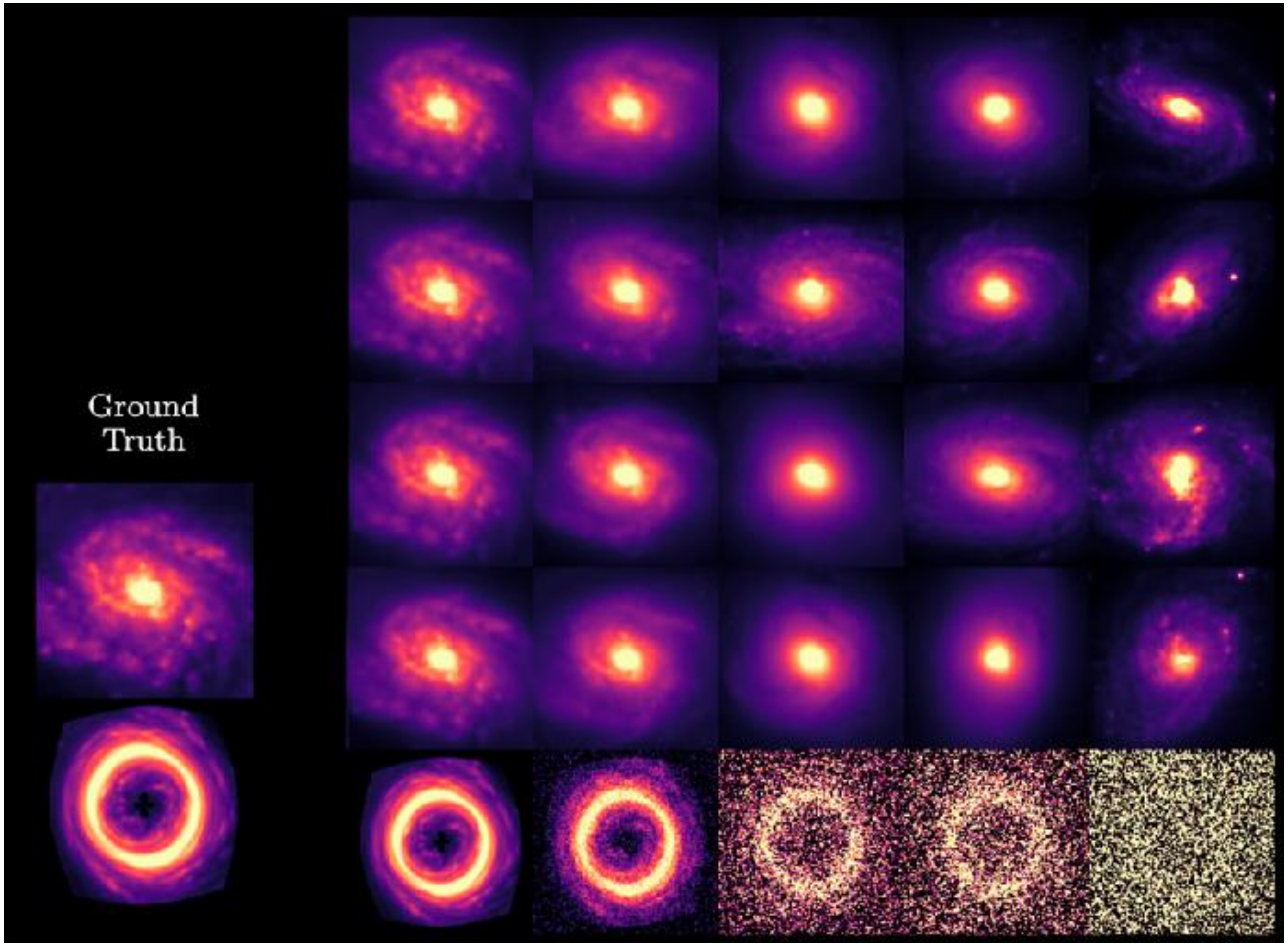
This is the prior score we learnt from the training data

SCORE-BASED MODELING

Alexandre Adam

To sample from the posterior, the score of the likelihood is all that we need:

$$\nabla_x \log p(x \mid y) = \nabla_x \log p(y \mid x) + \nabla_x \log p_\theta(x)$$

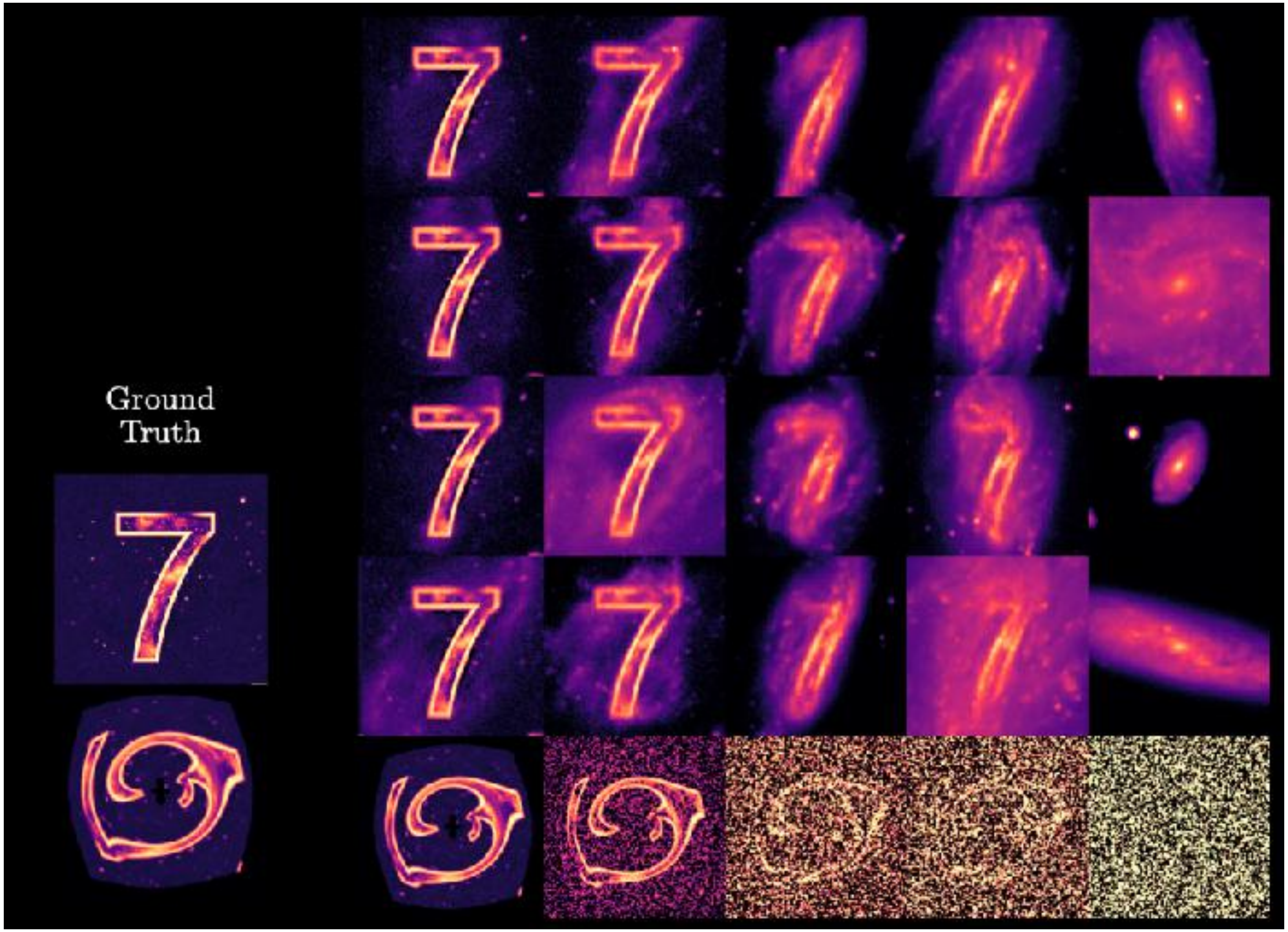


Out of Distribution Tests

Alexandre Adam

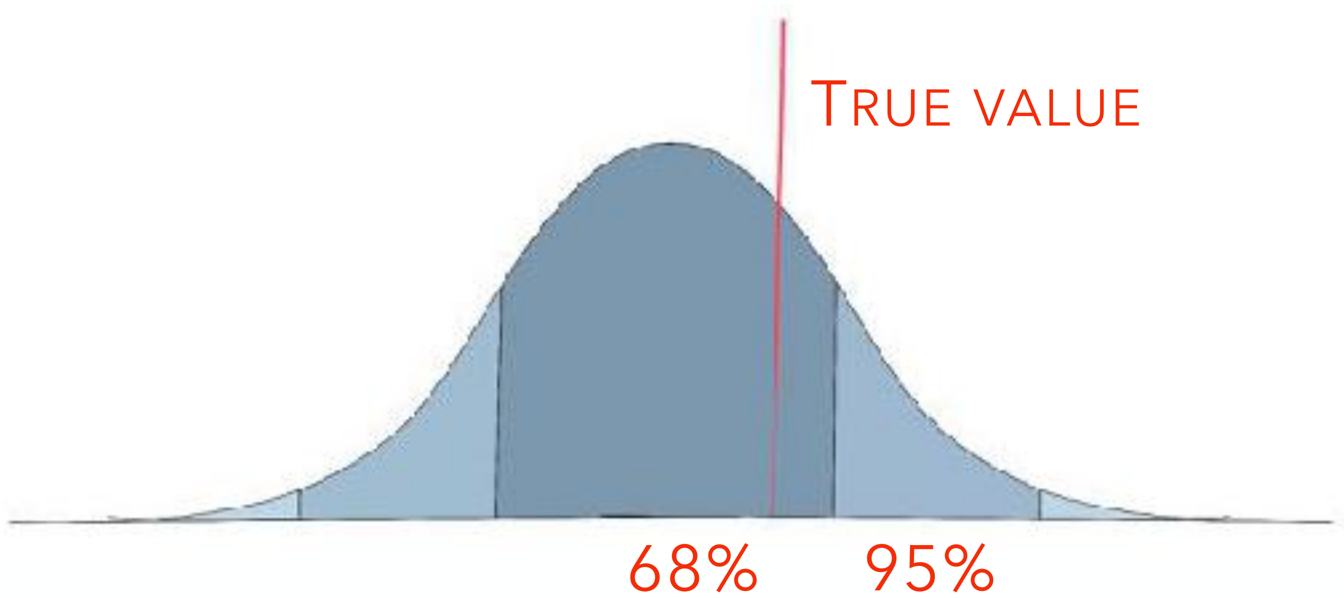
To sample from the posterior, the score of the likelihood is all that we need:

$$\nabla_x \log p(x \mid y) = \nabla_x \log p(y \mid x) + \nabla_x \log p_\theta(x)$$



Are these uncertainties accurate?

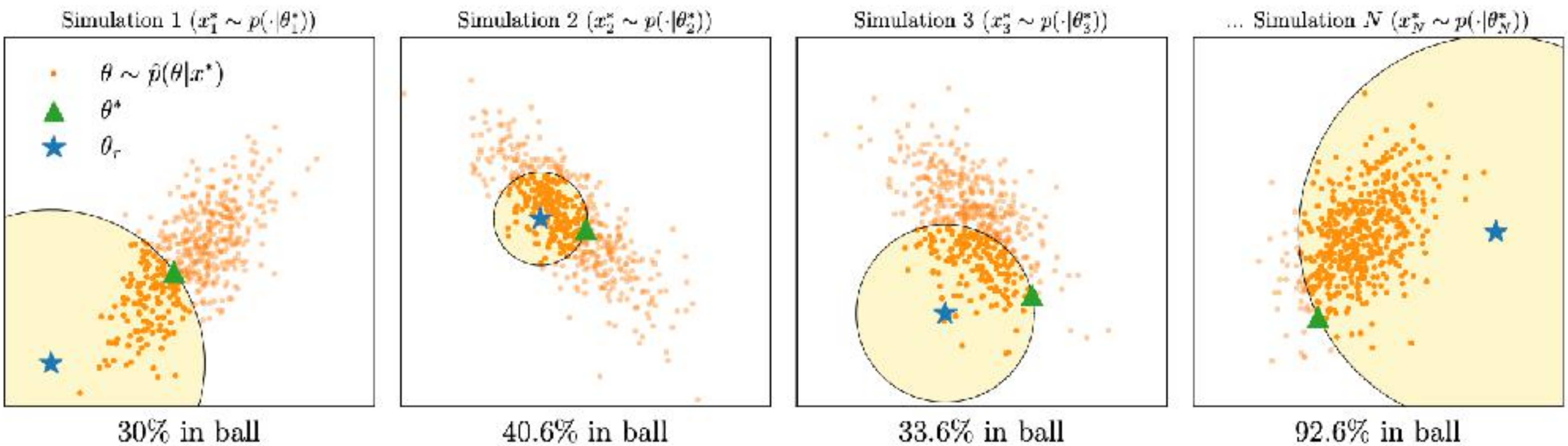
Coverage probability of a confidence interval is the proportion of the time that the interval contains the true value of interest.



For an accurate interval estimator, the coverage probability is equal to its confidence level

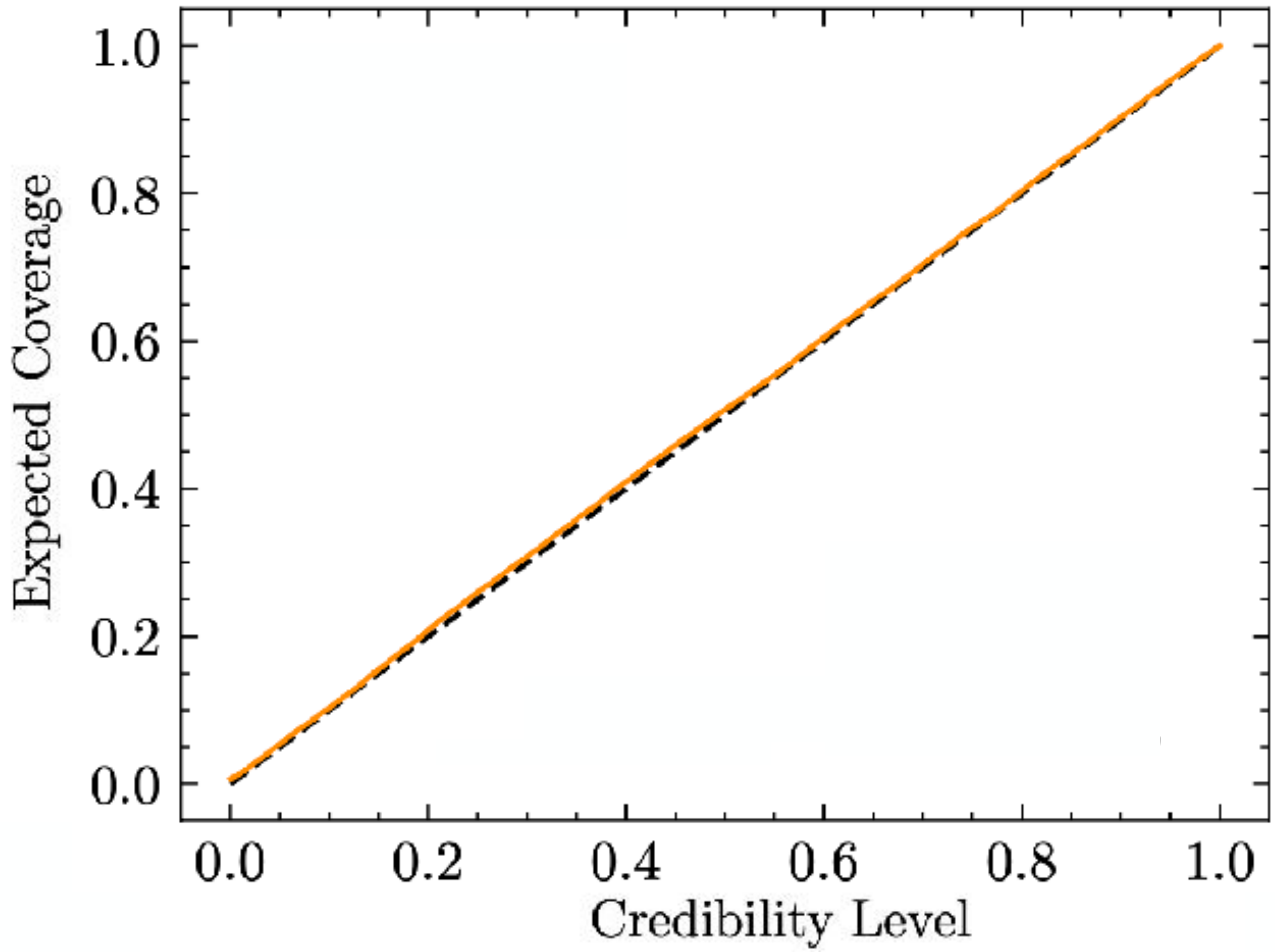
Coverage Test for Accuracy

Pablo Lemos



Lemos, et al. ICML 2023, 2302.03026

Coverage Test for Accuracy

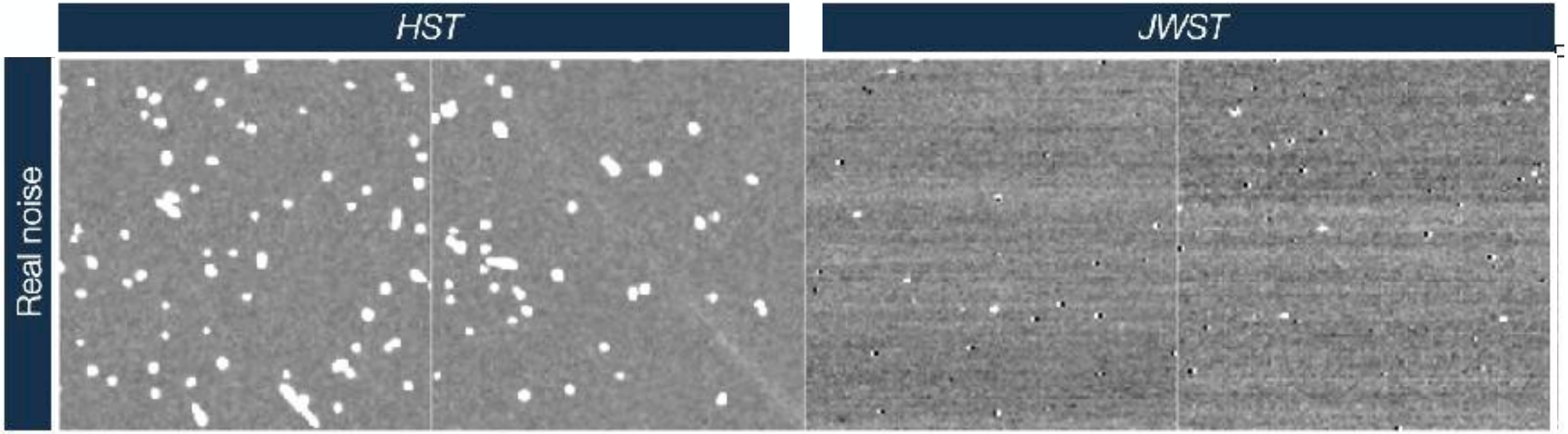


Lemos, et al. ICML 2023, 2302.03026

DEALING WITH REALISTIC NOISE: BEYOND GAUSSIANITY

Alexandre Adam

Ronan Legin

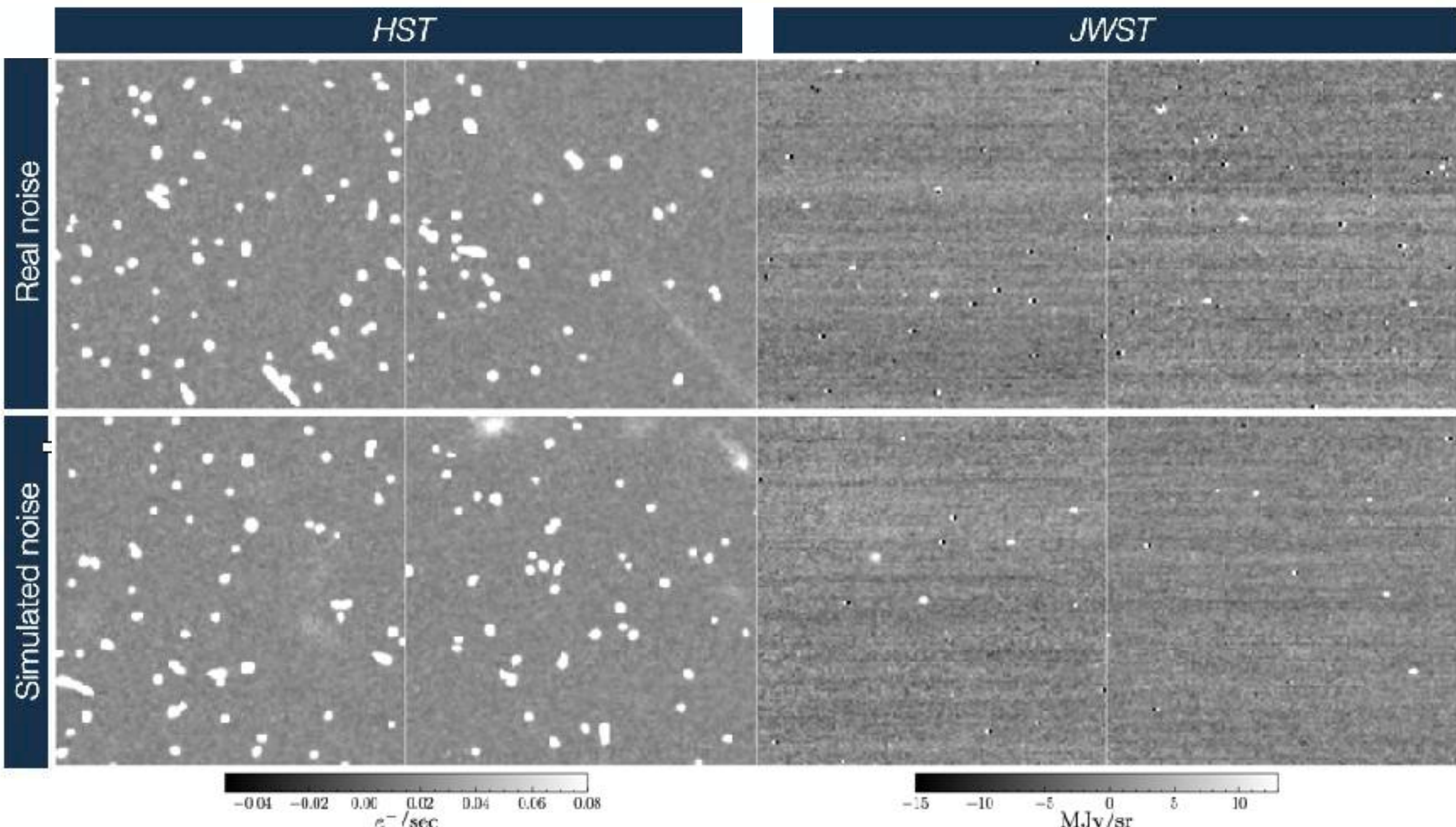


Legin, Adam, et al. 2302.03046

Dealing With Realistic Noise: Beyond Gaussianity
SLIC: Score-based LikelIhood Characterization

Since we have learnt a generative model of the additive noise, it can now be used in a simulation pipeline to get new, independent realizations of noise:

$$P(\mathbf{x}_{\mathrm{O}}|\eta) = Q(\mathbf{x}_{\mathrm{O}} - \mathbf{M}(\eta))$$



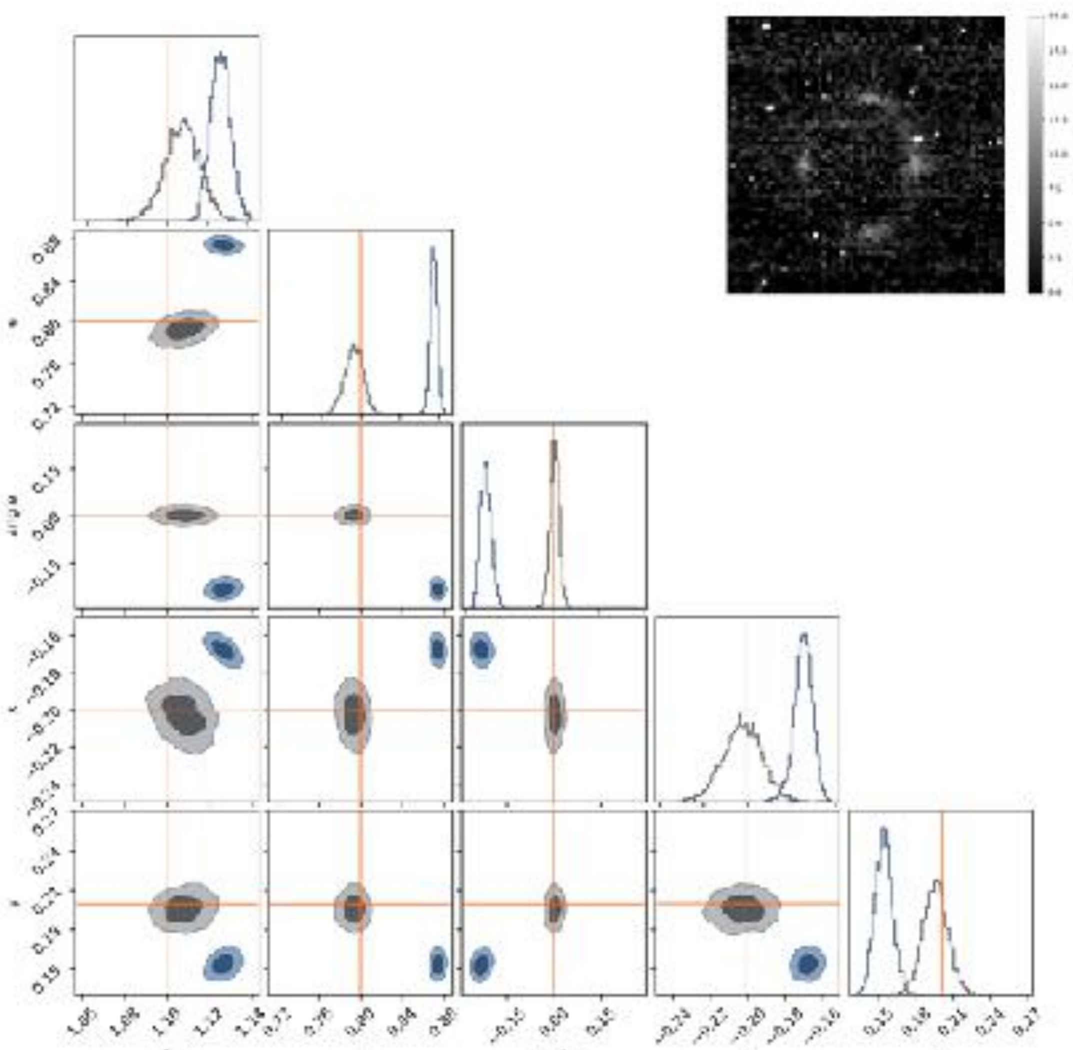
Legin, Adam, et al. 2302.03046

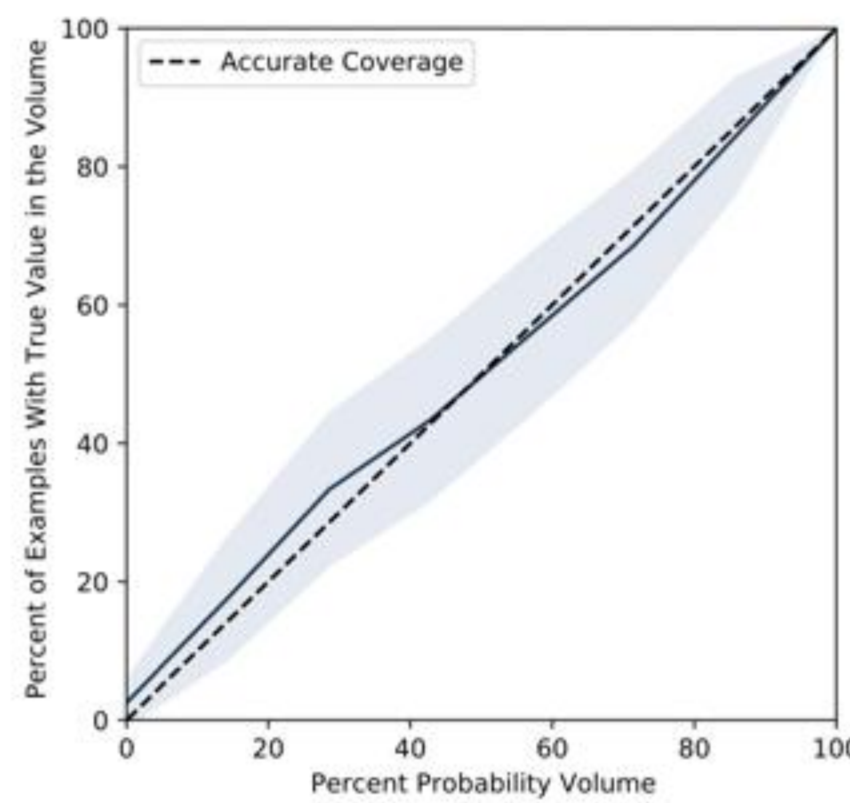
Dealing With Realistic Noise: Beyond Gaussianity
SLIC: Score-based Likelihood Characterization

$$\mathbf{s}(\mathbf{x}_0) = \partial \log Q(\mathbf{x}_0)/\partial \mathbf{x}$$

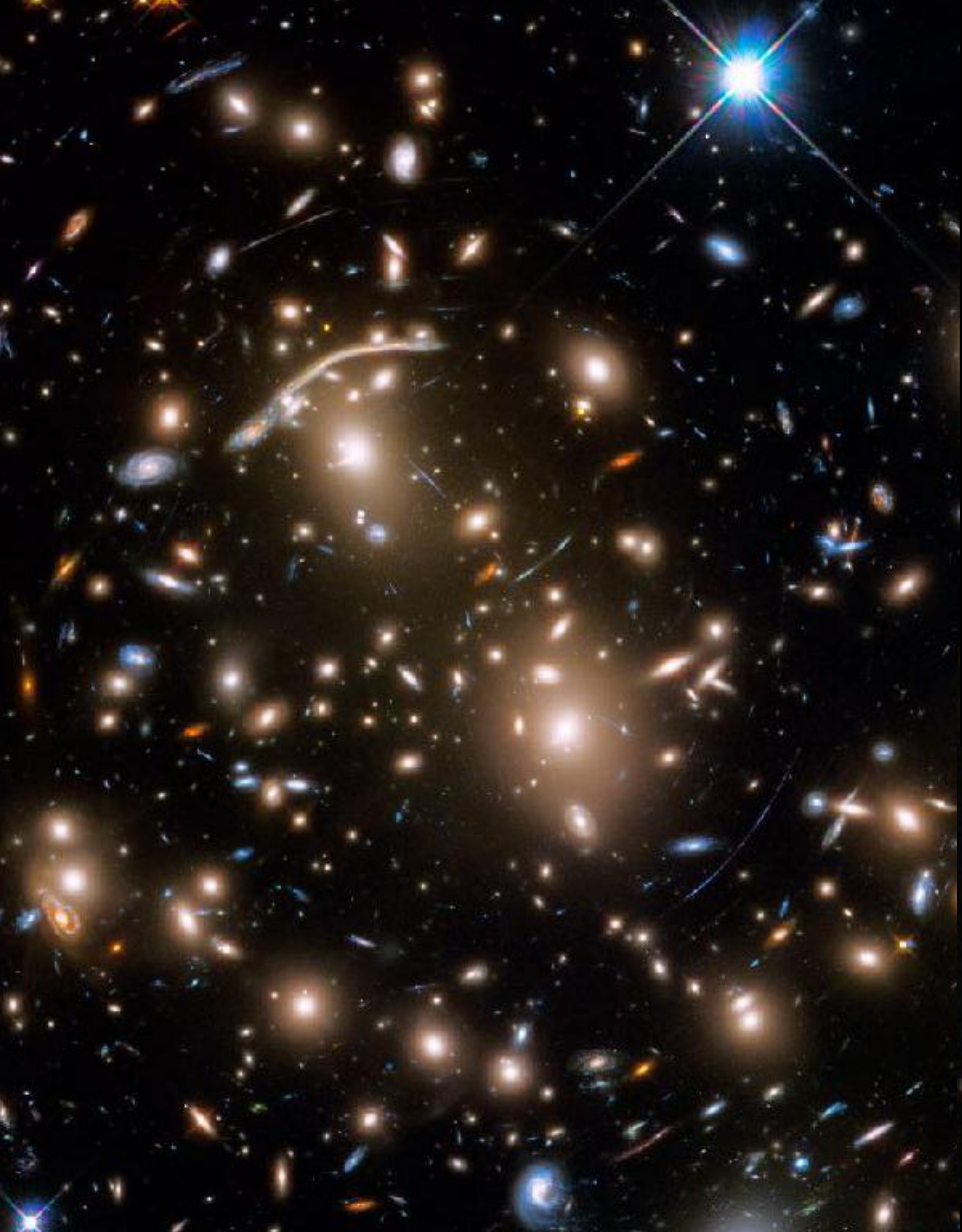
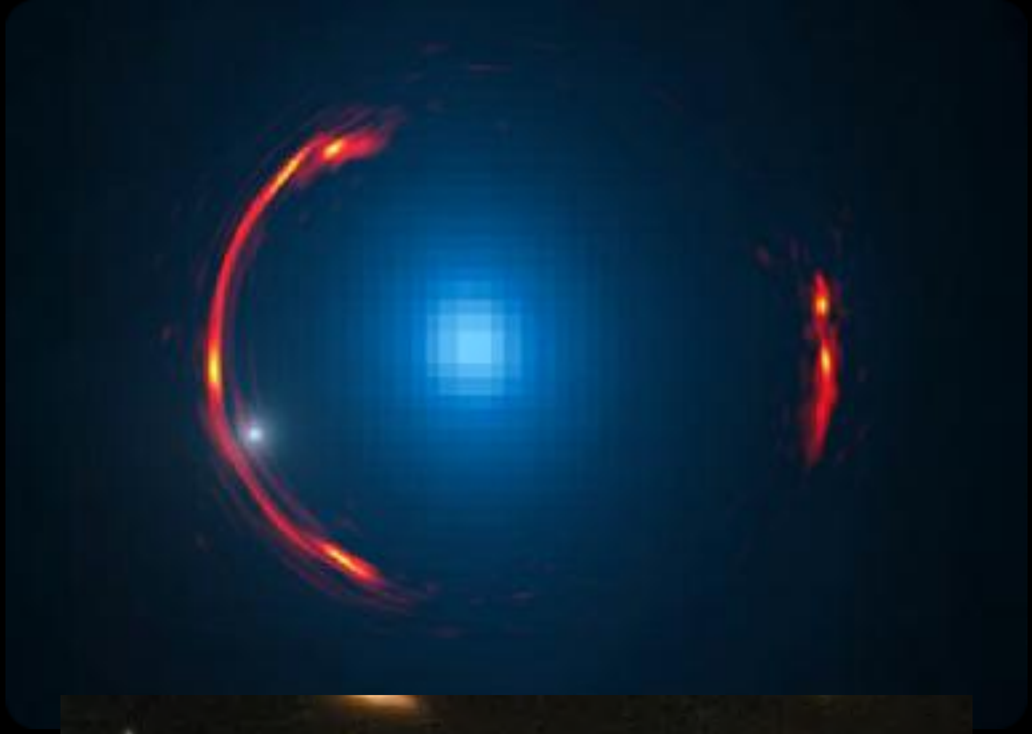
$$P(\mathbf{x}_\mathrm{O}|\eta) = Q(\mathbf{x}_\mathrm{O} - \mathbf{M}(\eta))$$

$$\eta_{i+1} = \eta_i + \tau\, \boldsymbol{\nabla}_\mathbf{x} \log Q(\mathbf{x}_\mathrm{o} - \mathbf{M}(\eta))\boldsymbol{\nabla}_\eta M(\eta_i) + \sqrt{2\tau}\xi$$





Legin, Adam, et al. 2302.03046

Ciela Institute

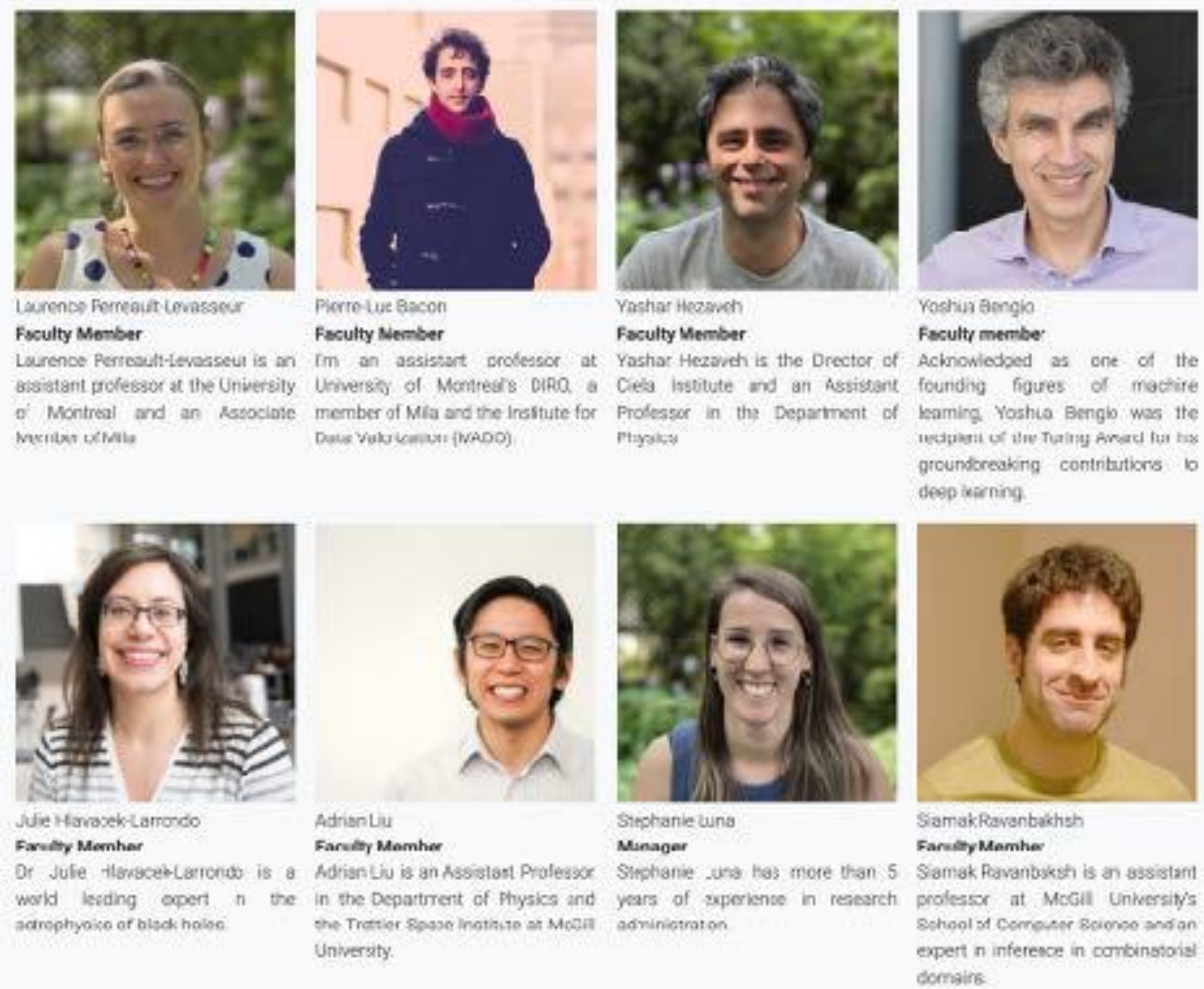
Explore Algorithmically

Montreal Institute for Astrophysics and Machine Learning

Ciela is a research institute associated with the Université de Montréal, with the mission to contribute to breakthrough discoveries in astrophysics by developing new computational data analysis methods. It supports a unique interdisciplinary community of world-leading researchers in astrophysics and machine learning. Read about our research.

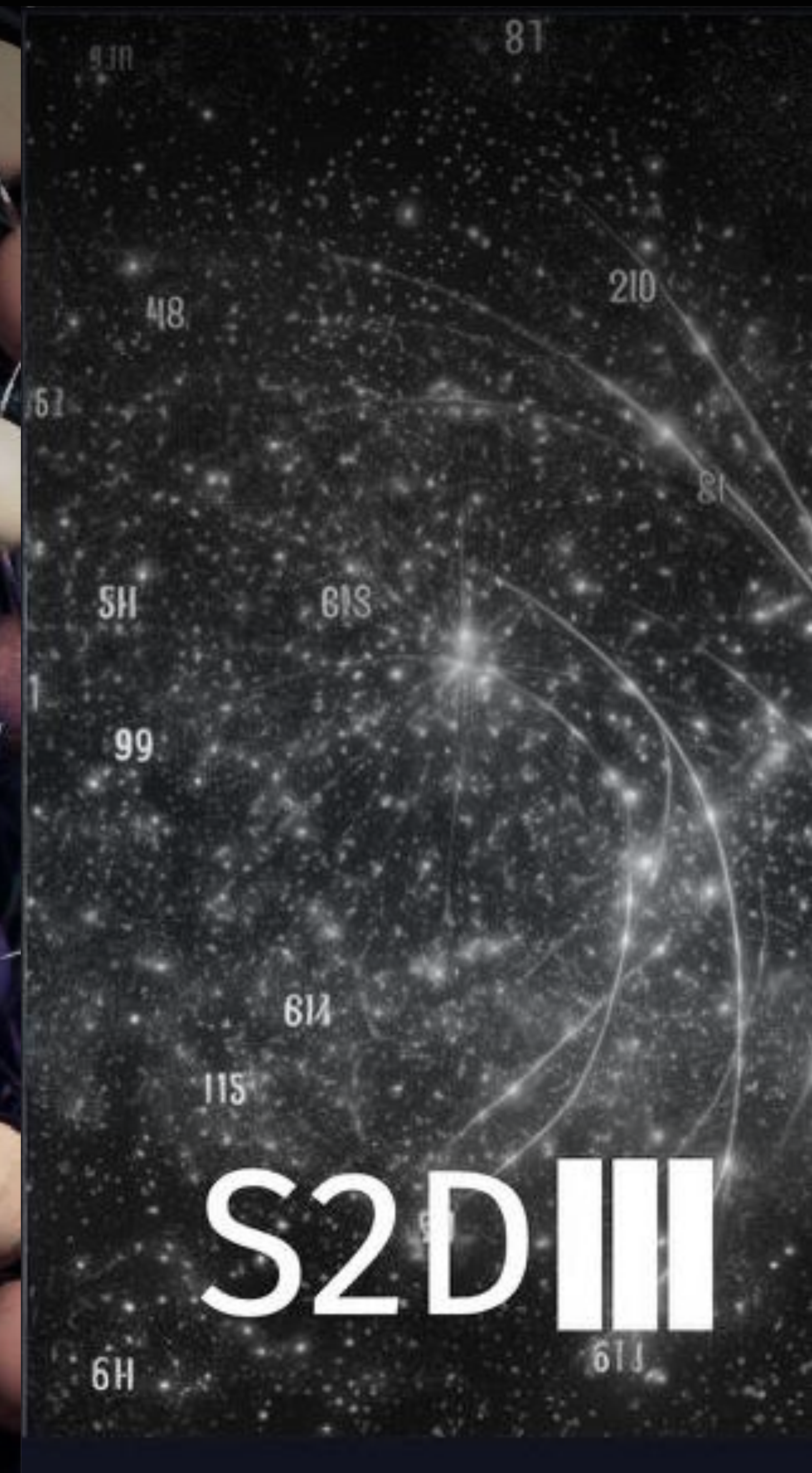
Ciela Institute

The institute's mission is to contribute to breakthrough discoveries in astrophysics and cosmology by developing innovative data analysis and machine learning methods. It also aims to use unsolved and challenging data analysis problems in astrophysics to push the boundaries of machine learning and to make significant advances that can contribute to the successful and widespread use of machine learning in other fields of science.




ASTROMATIC
2022
JULY 31 AUGUST 6
CITY OF MONTREAL


Ciela
Mila
Physical Sciences seminars


S2D

Strong lensing simulation pipeline: Caustic

A fast, AI-empowered, differentiable, extremely modular simulation pipeline for all your strong lensing needs.

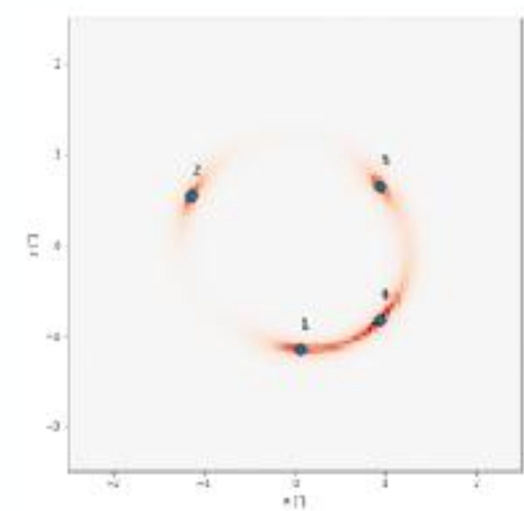
1) Lens and source from analytic profiles or pixelated images/densities

2) Multiplane lensing

3) Line of sight mass distributions

4) Fast microlensing simulations

5) Time-delays



Connor Stone

Adam Coogan

Andi Filipp

Alex Adam

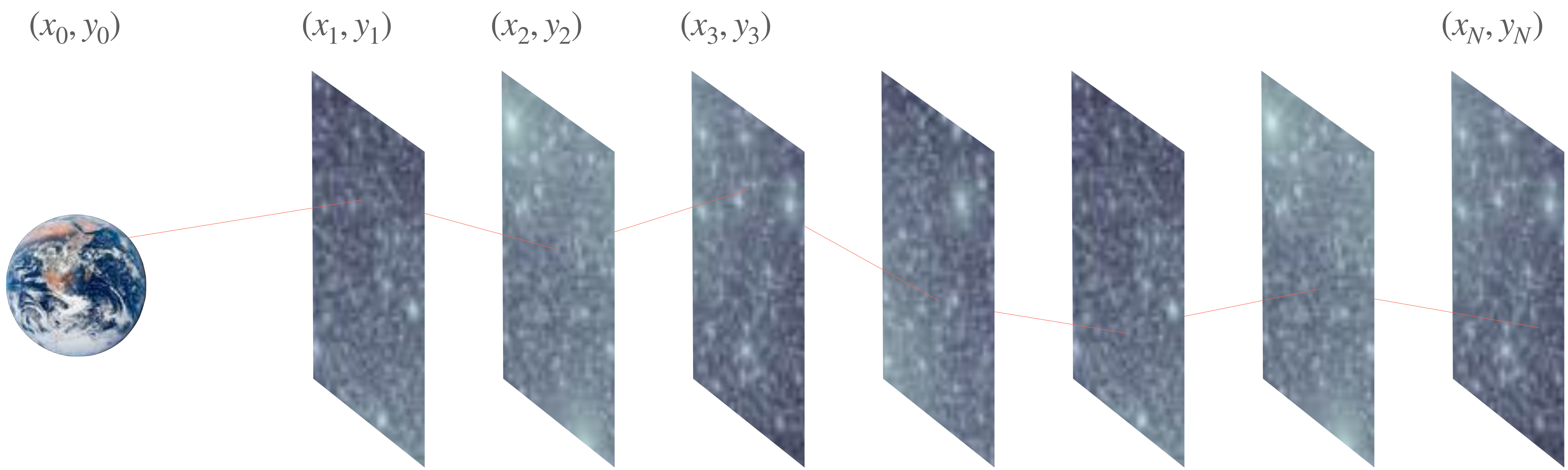
Misha Barth

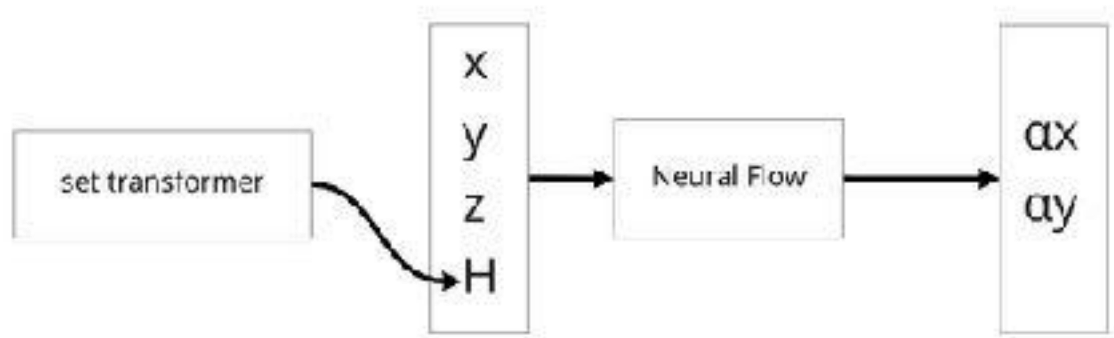
Charles Wilson

https://github.com/Ciela-Institute/caustic

https://github.com/Ciela-Institute/caustic-analyses

SPEEDING UP THE SIMULATIONS



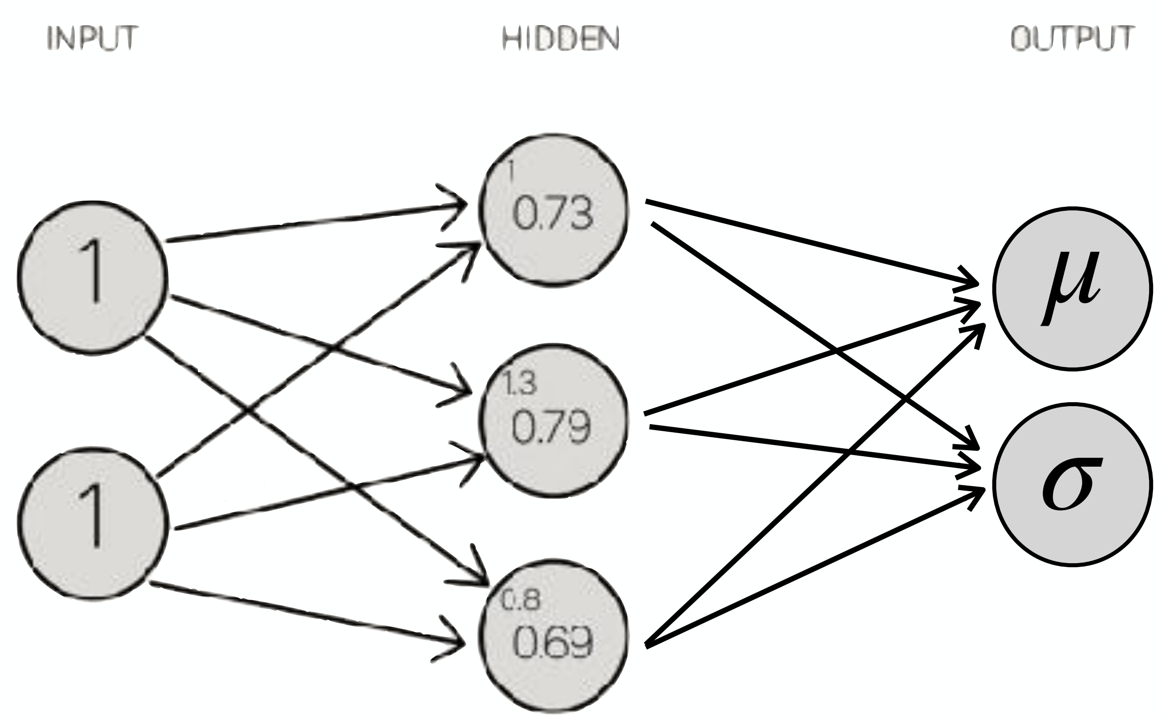


Charles Wilson

Bayesian Neural Network

(Aleatoric uncertainties)

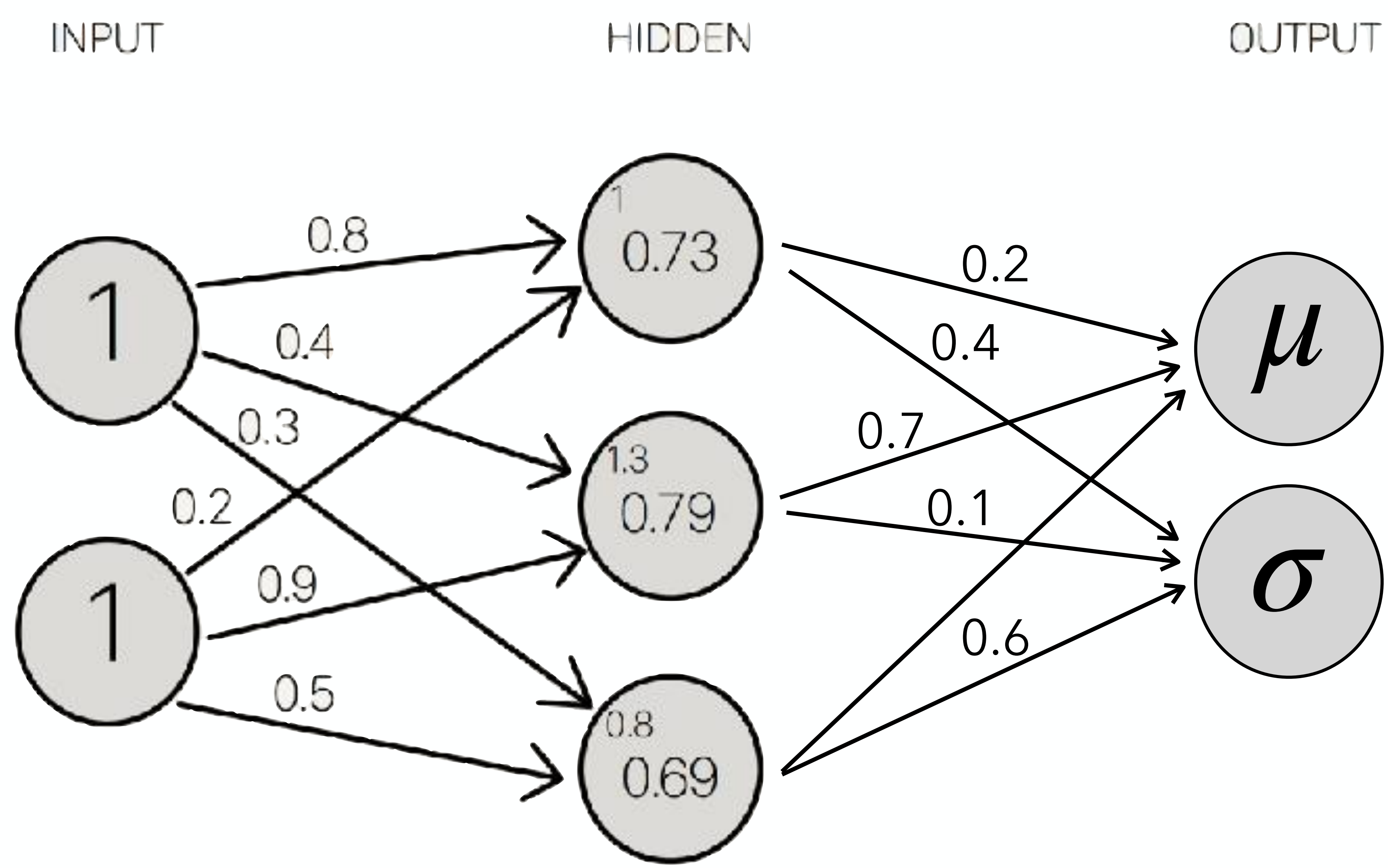
The neural network predicts its own uncertainties



$$\mathcal{L}(\mathbf{y}_n, \hat{\mathbf{y}}_n(\mathbf{x}_n, \omega)) \propto \sum_k \frac{-1}{2\sigma_k^2} ||y_{n,k} - \hat{y}_{n,k}(\mathbf{x}_n, \omega)||^2 - \frac{1}{2} \log \sigma_k^2$$

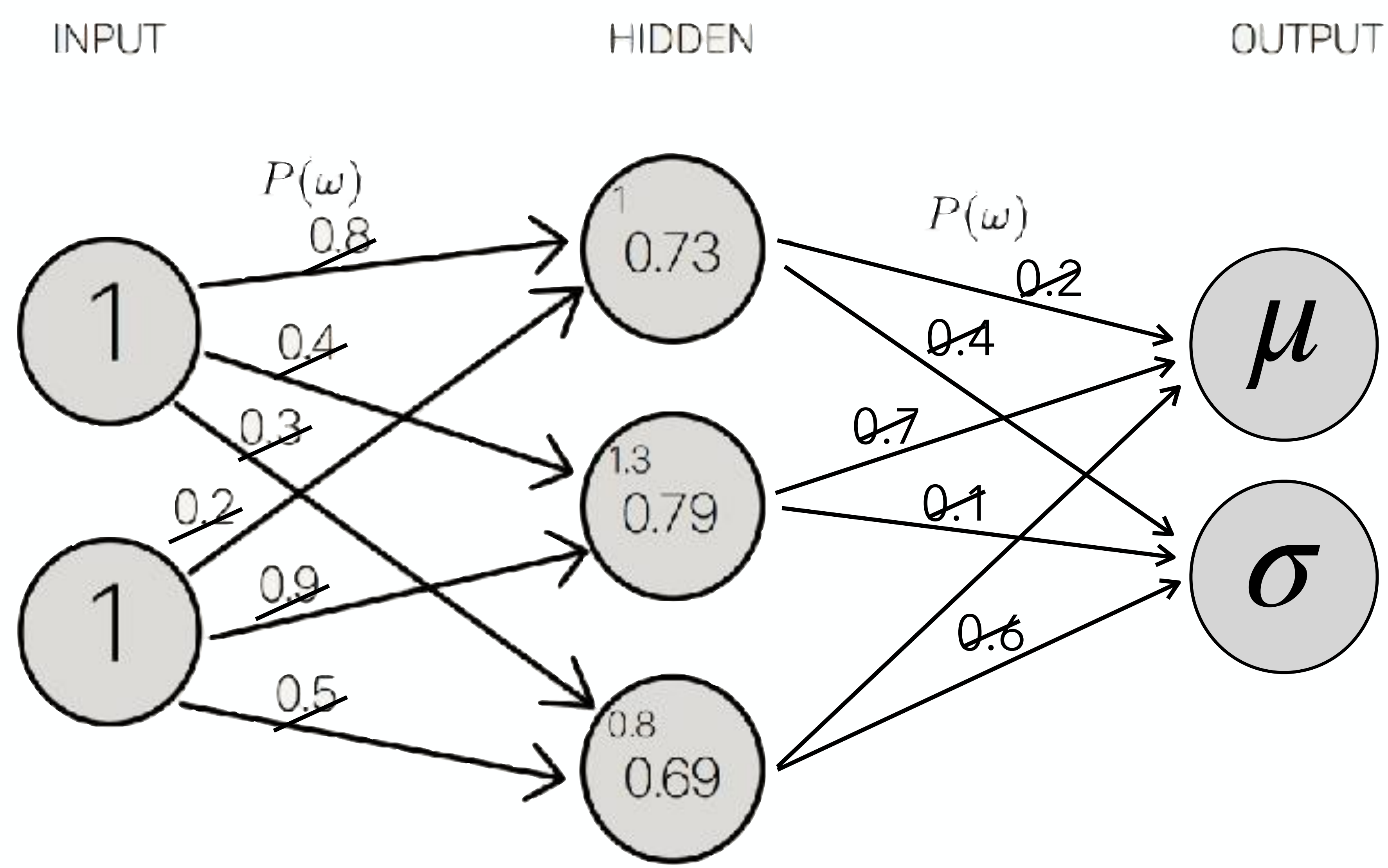
Standard Neural Networks:

Weight have fixed, deterministic values



Bayesian Neural Networks:

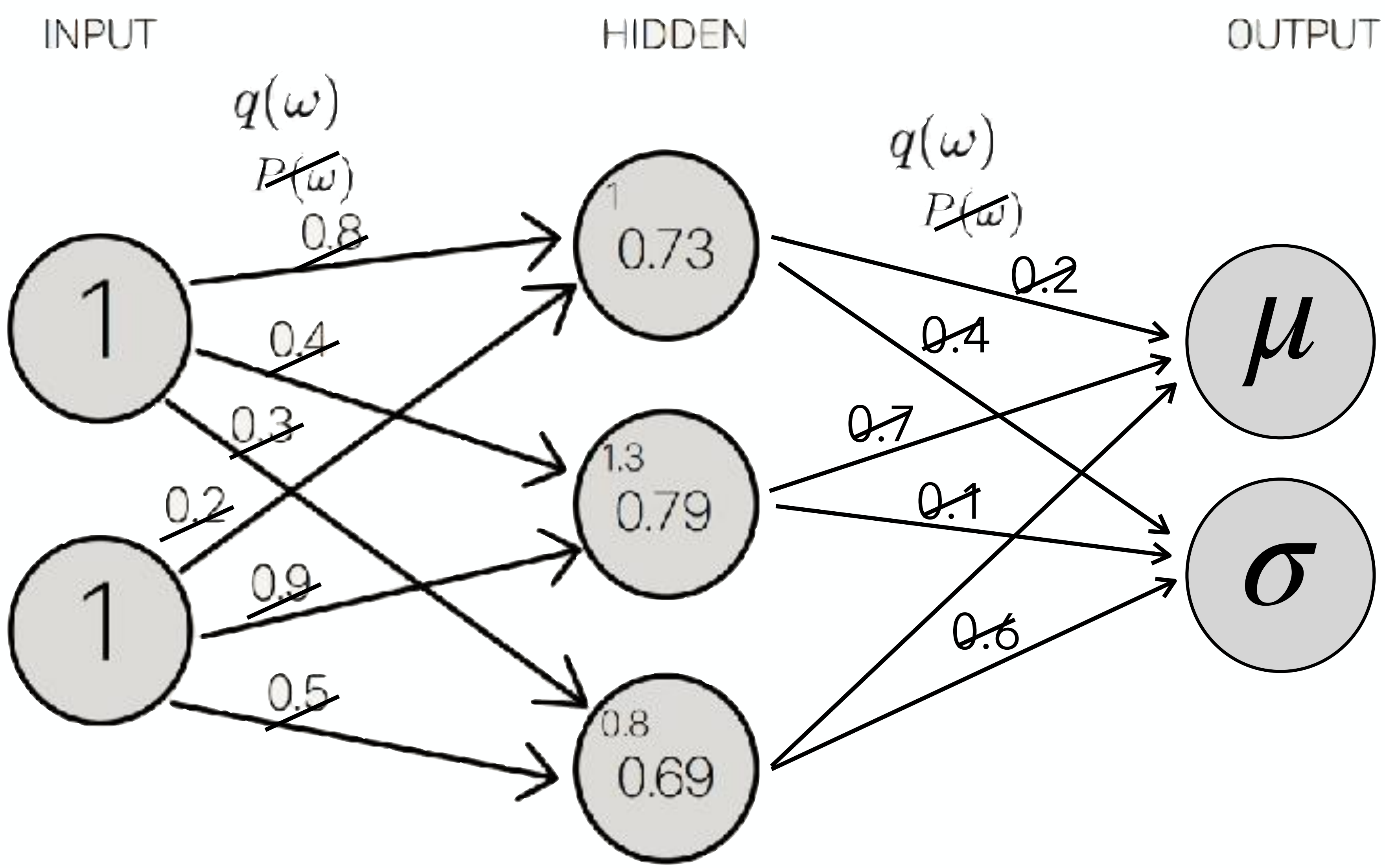
Instead of fix values, weights are defined by probability distributions



[using variational inference]

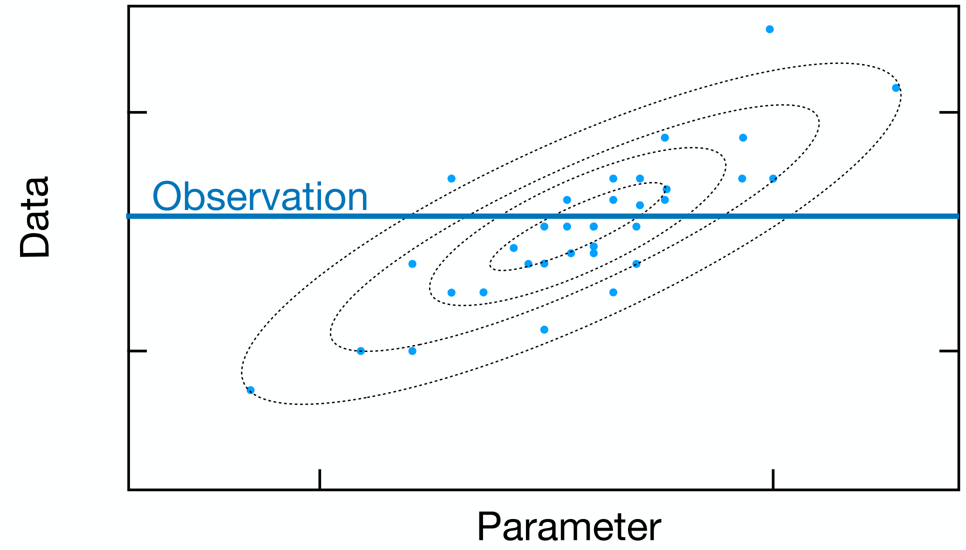
Variational Inference

Replace $P(\omega)$ by a distribution with a simple analytic form, $q(\omega)$, (e.g., a Gaussian).



Likelihood-free inference

Simulation-based inference: produce lots of simulations to populate a parameters-data graph, and at test time cut through that graph to get an accurate posterior



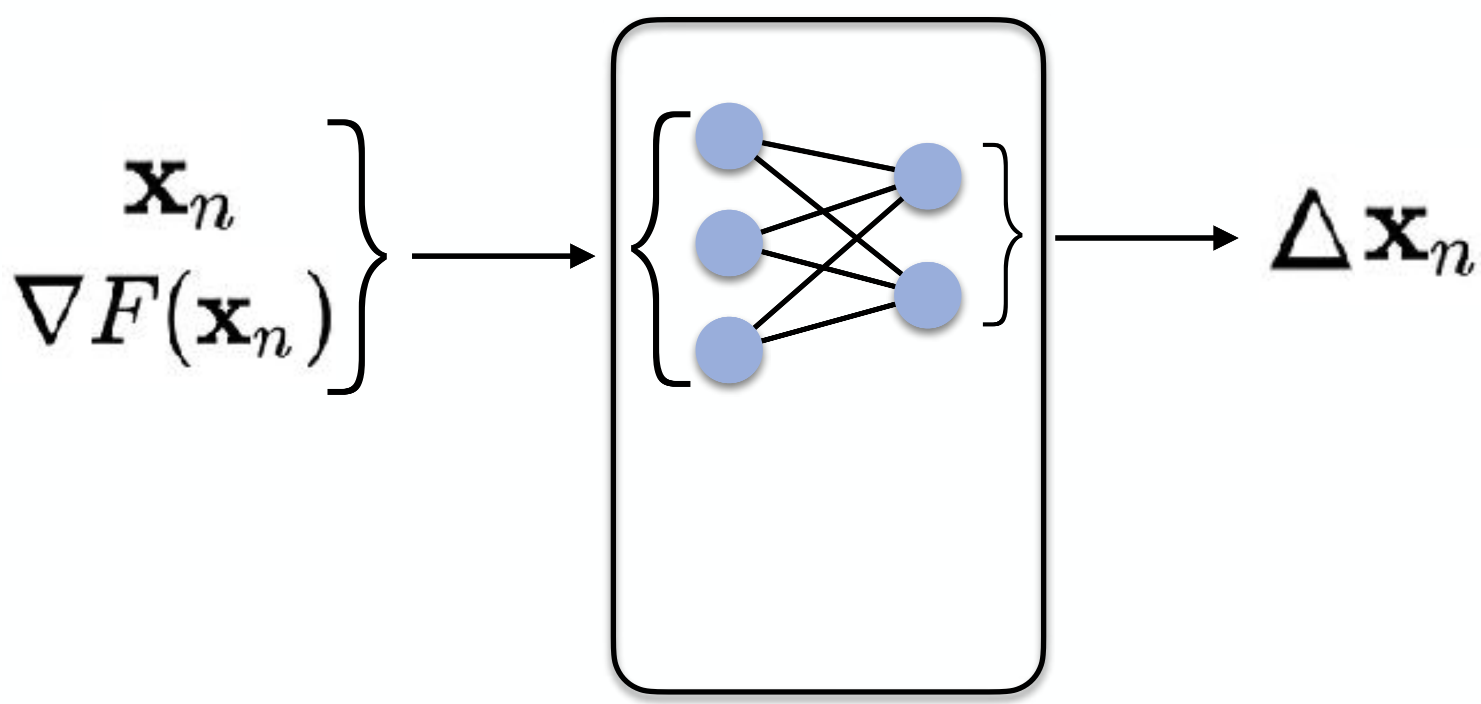
It's possible to use deep learning for automated feature extraction and data compression, and then run your preferred LFI framework.
=> Allows to both harness the power of NNs and be fully Bayesian, it's the best of both world!!

Train a neural network to be a gradient descent optimizer

$$\mathbf{x}_{n+1} = \mathbf{x}_n - \underbrace{\gamma_n \nabla F(\mathbf{x}_n)}_{\Delta \mathbf{x}_n}$$

Train a neural network to be a gradient descent optimizer

$$\mathbf{x}_{n+1} = \mathbf{x}_n - \underbrace{\gamma_n \nabla F(\mathbf{x}_n)}_{\Delta \mathbf{x}_n}$$



Train a neural network to be a gradient descent optimizer

$$\mathbf{x}_{n+1} = \mathbf{x}_n - \underbrace{\gamma_n \nabla F(\mathbf{x}_n)}_{\Delta \mathbf{x}_n}$$