Park *et al* 2020 https://arxiv.org/abs/2012.00042

Large-Scale Gravitational Lens Modeling with Bayesian Neural Networks for Accurate and Precise Inference of the Hubble Constant

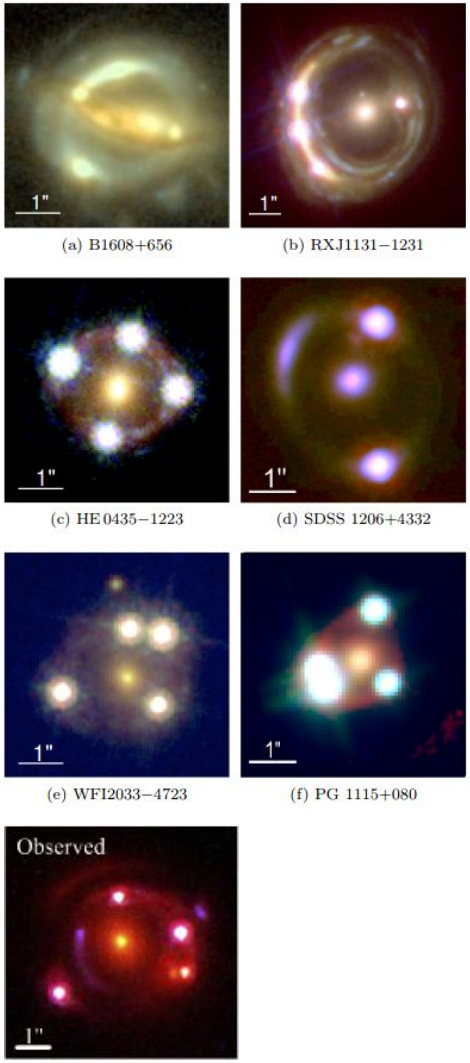
Ji Won Park[1, 2], Sebastian Wagner-Carena[1], Simon Birrer[1], Phil Marshall[1, 2], Joshua Yao-Yu Lin[3], Aaron Roodman[1, 2] (LSST DESC)

Time-Domain Cosmology with Strong Gravitational Lensing, 1/26/2020

[1] Kavli Institute for Particle Astrophysics and Cosmology, Stanford University
[2] SLAC National Accelerator Laboratory
[3] University of Illinois at Urbana-Champaign

LSST DESC Dark Energy Science Collaboration

Motivation

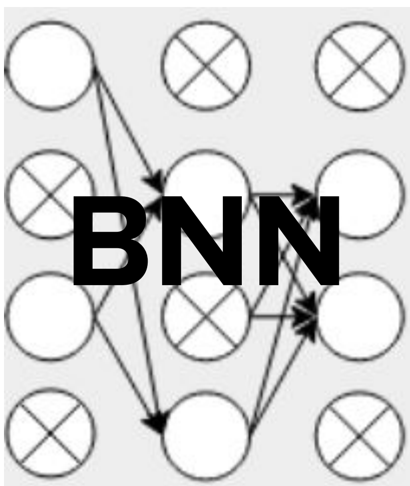


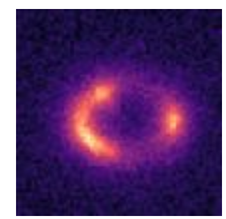
- We need 40 time-delay lenses and 200 non-time-delay lenses to reach ~1% precision (Birrer and Treu 2020, TDCOSMO V).
- LSST is expected to discover 100s of time delay lenses and 10,000s of non-time delay lenses (Collett 2015, Oguri and Marshall 2010).
- Current lens modeling method relies on **forward modeling** the images.
  - Time-intensive
  - Lens-by-lens basis
- Joint hierarchical inference over hundreds of lenses requires an ***efficient***, ***unbiased*** method with a ***uniform*** approach to modeling. → **Bayesian neural networks!**

Wong et al 2019, Shajib et al 2020

Lens modeling with Bayesian neural networks (BNNs) *applied to H0 inference*

- BNNs can approximate the **full posterior** over the target quantities, rather than a point estimate (Denker & Lecun 1991).
- They show promising accuracy on galaxy-galaxy lenses (Hezaveh *et al* 2017, Levasseur *et al* 2017, Wagner-Carena *et al* 2020). So do non-Bayesian variants (Pearson et al 2019, Schuldt et al 2020).

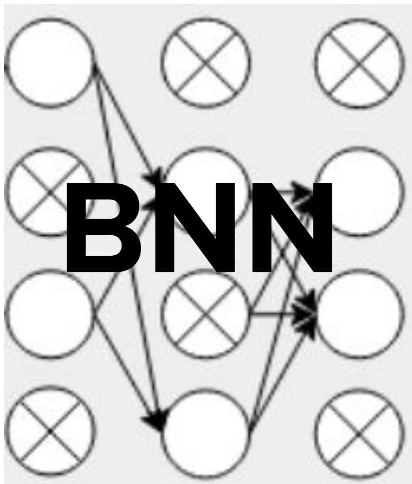


$p(\text{lens model parameters}|$  $)$

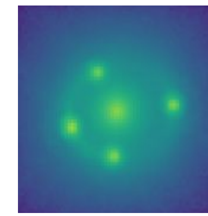
Are they good enough for cosmography?

Pipeline overview

1. BNN lens modeling



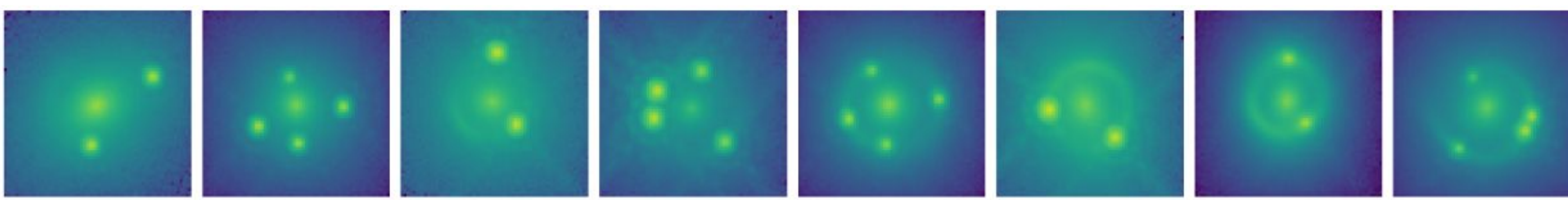
PEMD + external shear

$p(\text{lens model parameters, source position}\,|$  $)$

11 “BNN target parameters”

See the paper and code for implementation details!

- Simulated HST-like images (assuming follow-up)
- Training, test sets drawn *i.i.d.*



Pipeline overview

2. Propagating BNN-inferred posterior into H0 inference (simplified)

Setup: time delays were simulated with 0.25-day errors (very accurate) and redshifts assumed to be known.

H0 posterior for a single lens was

$$p(H_0|\mathrm{image}, \Delta t) \propto p(H_0) p\left(\mathrm{image}, \Delta t | H_0\right)$$ Bayes' rule

$$\propto p(H_0) \int p\left(\Delta t | \mathrm{BNN\ target\ params}, H_0\right)$$

Time delay likelihood

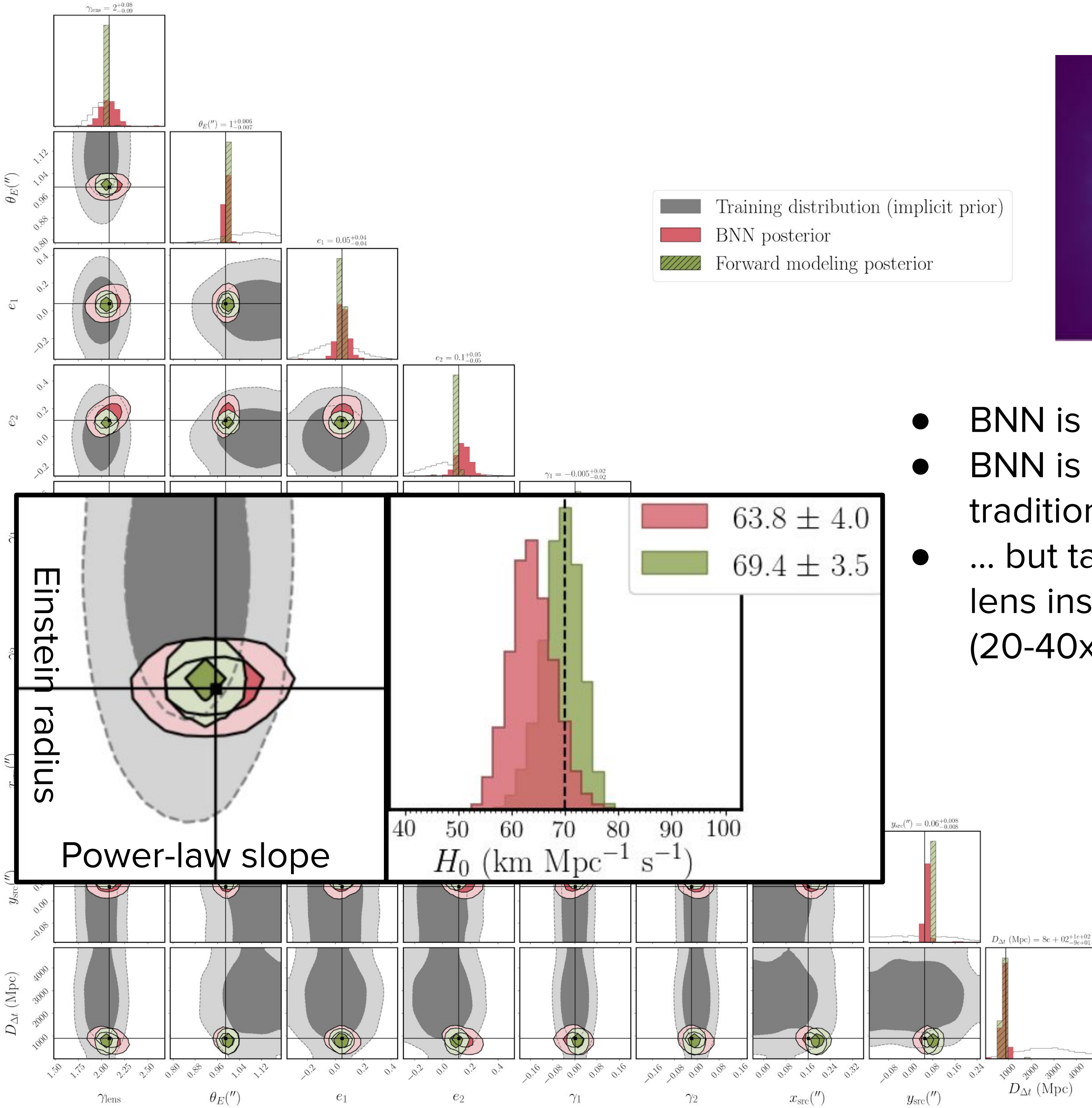
$$\times p\left(\mathrm{BNN\ target\ params} | \mathrm{image}\right)$$

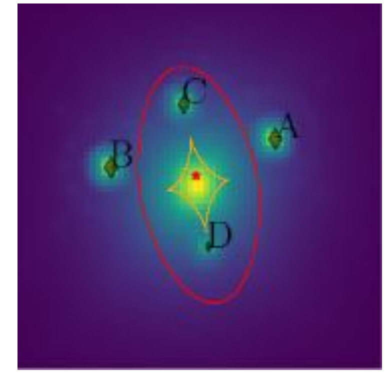
BNN-inferred posterior

$$\times \cdots$$

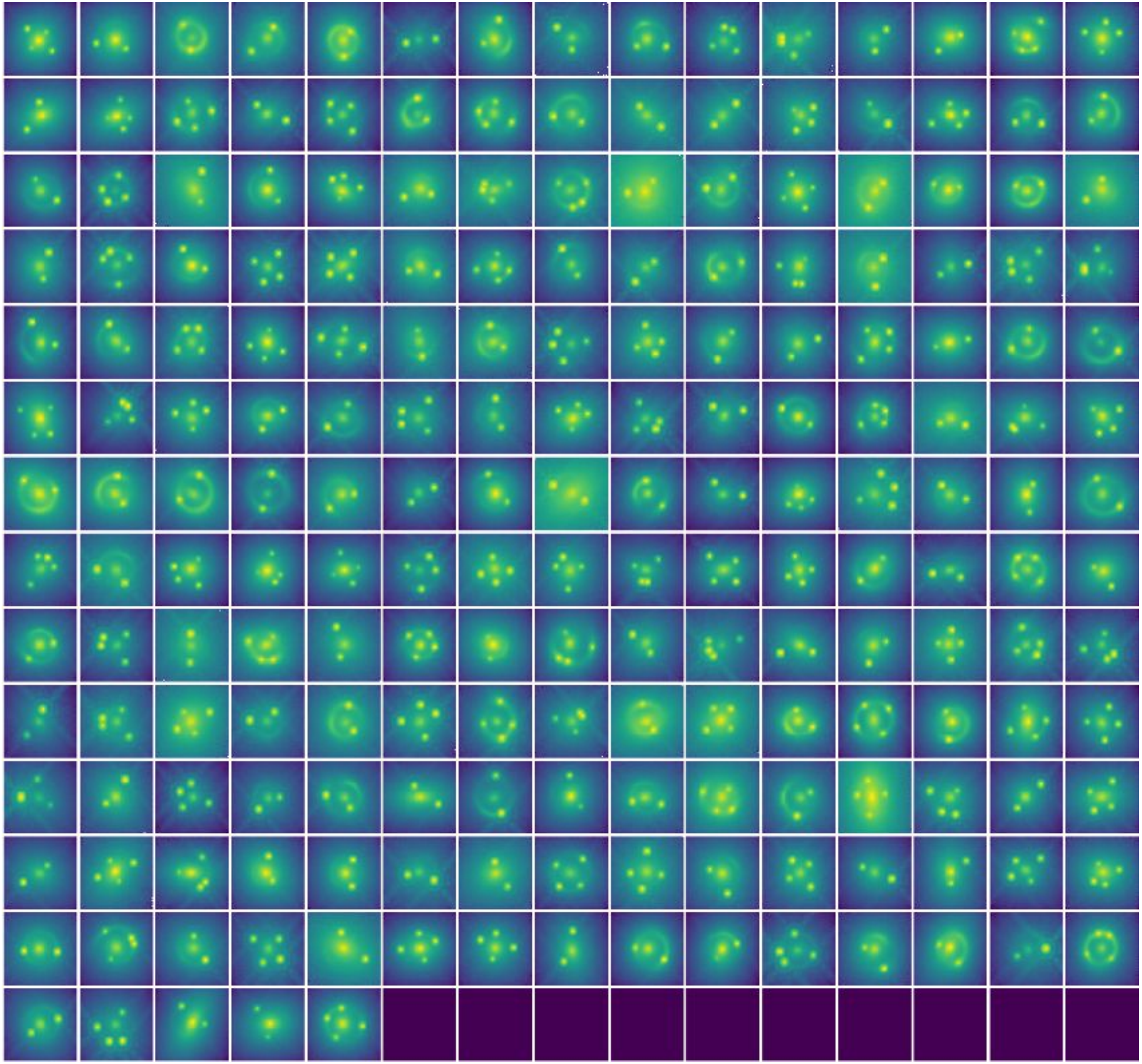
$$d(\mathrm{BNN\ target\ params})$$

- MCMC sample over $D_{\Delta t}$ (equivalently, H0) jointly with the BNN target parameters
- Combine across 200 lenses

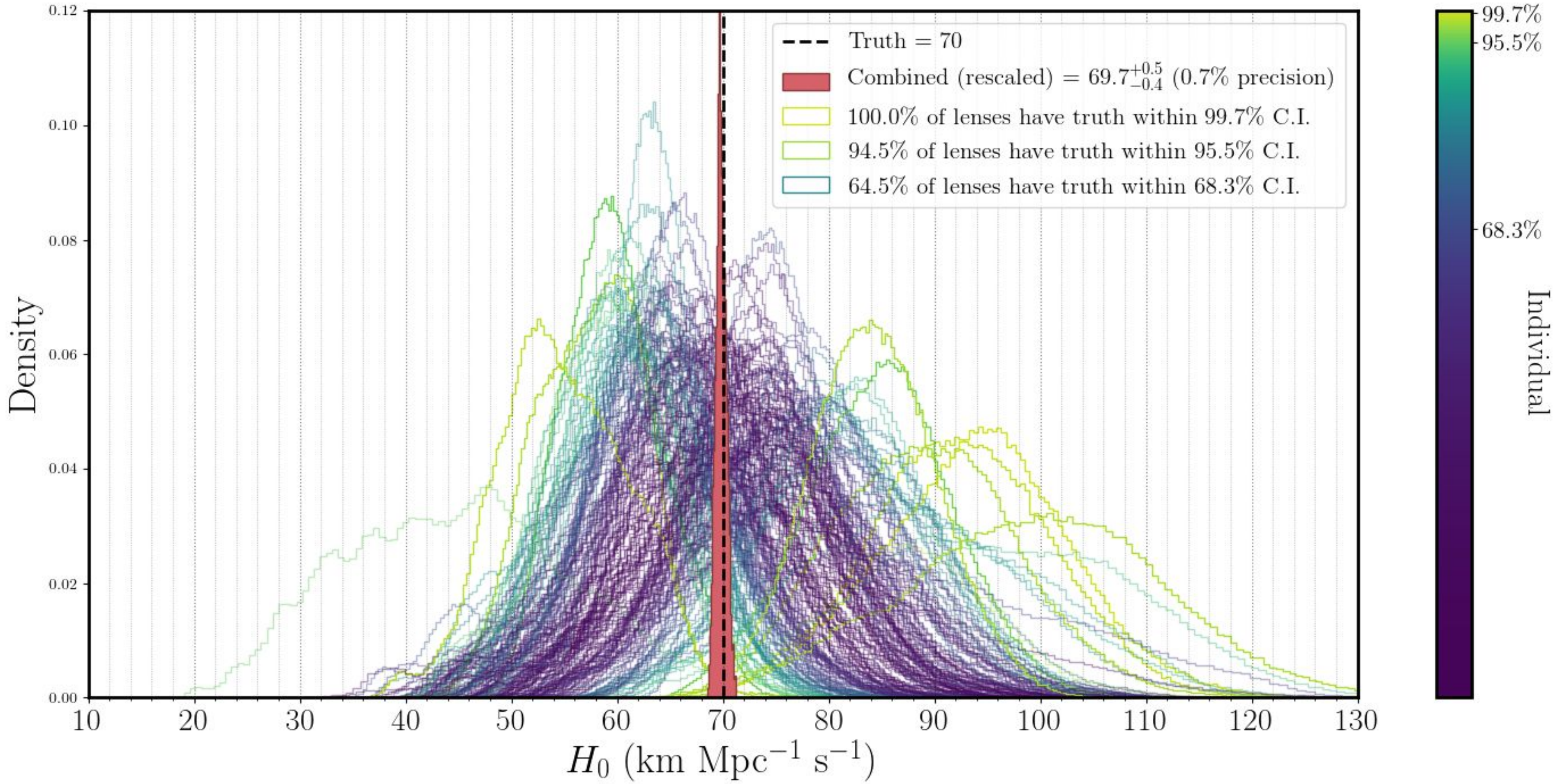




- BNN is accurate
- BNN is less precise than traditional forward modeling
- … but takes 9 minutes per lens instead of 3-6 hours (20-40x speedup)



Power of rapid ensemble inference



- Confidence in statistical agreement with truth (70 km/Mpc/s)
- 9% average precision achieved per lens
- 0.7% precision achieved from 200 lenses, with no evidence of bias

Code and data releases

- Our datasets were generated using Baobab, a wrapper around Lenstronomy that samples the parameters and renders the images.

https://github.com/jiwoncpark/baobab

Baobab

build passing | docs passing | coverage 27%

- The H0 inference pipeline is implemented in H0rton.

https://github.com/jiwoncpark/h0rton

h0rton - Deep Modeling of Strong Gravitational Time Delay Lenses for Bayesian Inference of the Hubble Constant

build passing | docs passing | coverage 74% | pypi package 1.0 | license MIT | astro-ph.IM arXiv:2012.00042

DOI 10.5281/zenodo.4300382 | powered by AstroPy

Key takeaways

Our paper is the first to use BNNs for time delay cosmography.

- ✓ BNN lens modeling is sufficiently **accurate for cosmography** for individual time delay lenses, given our model assumptions.
- ✓ When propagated into a joint H0 inference over 200 lenses, the BNN lens models enable **precise (0.7%) and unbiased H0 recovery**.
- ✓ Our pipeline is **efficient**, enabling large-scale systematics tests on a 1-day development cycle -- unfeasible for traditional forward modeling.

The method and software we present in this paper promise to become core infrastructure in large-scale hierarchical inference for H_0, as the cosmology community prepares to beat down systematics for a large sample of lenses soon to become available.

Bonus slides

Computational efficiency

3. Cosmological sampling **~19hr**

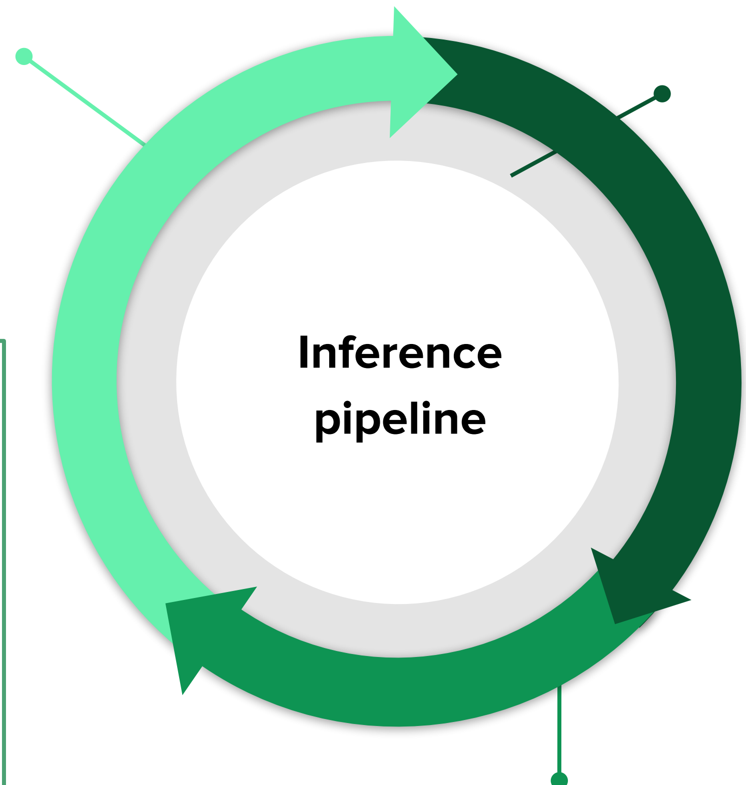
- Evaluating the BNN lens posterior takes <4μs for 200 lenses (quick)
- MCMC sampling for H0 takes 6 minutes per lens (bottleneck)

9min/lens or **1.2 day total** for test set of 200 lenses

A complete experiment can be performed on a 1-day development cycle.

Compare to: 3-6hr/lens for traditional forward modeling (20 -40x speedup)*

**Lenstronomy* (Birrer *et al* 2018) *benchmark*



1. Data generation **~6hr**

512K training set,
5K validation set,
200 test set

2. Training **~ 5hr**
on 16GB NVIDIA Tesla P100 GPU

To-do list

Our pipeline is inherently versatile and allow extensions in many directions.

- Hierarchical inference framework (Wagner-Carena *et al* 2020) to recover the population prior over the BNN target parameters
  - Robust to out-of-sample data
  - Can incorporate information from non-time delay lenses
- Complex lens models
- Complex source models
- Stress tests with image artifacts, e.g. cosmic rays
- LSST-like image features, e.g. multiband data, PSF, noise
- Simultaneously inferring redshifts
- Likelihood-free inference

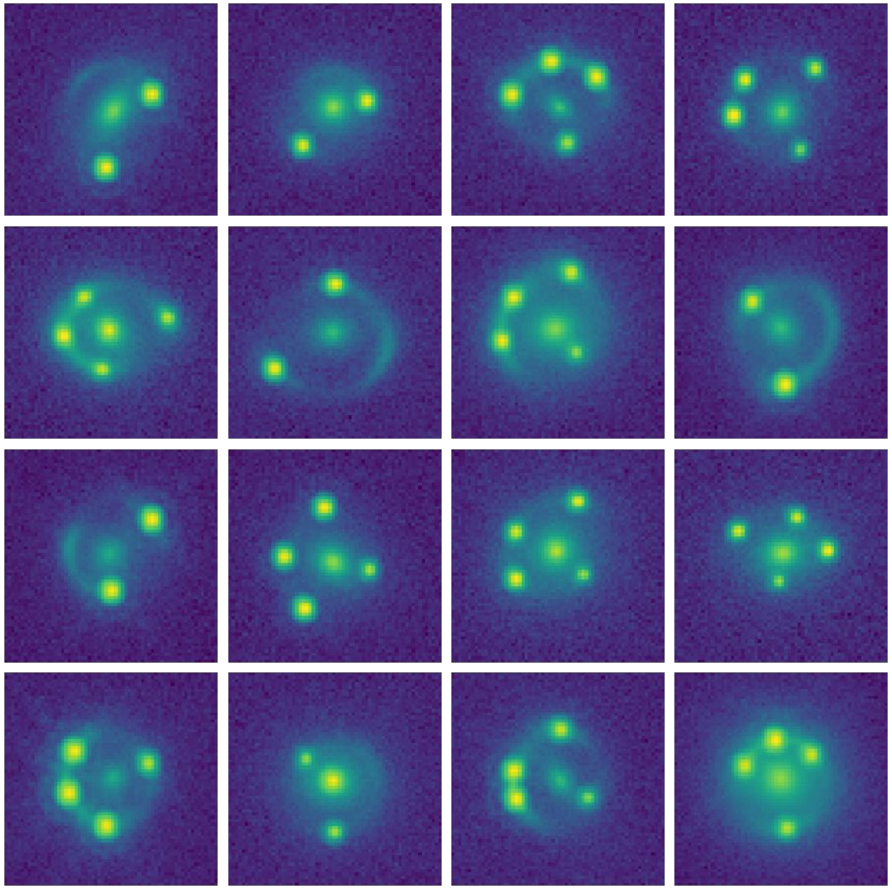
The goal of cosmography places extra demands on lens modeling.

- We require high precision on the power-law mass slope (γ'), as its uncertainty propagates directly into D_dt.
- BNN needs to predict additional parameters.
  - Source position
    - ~mas precision desired (Birrer *et al* 2019)
  - Lens galaxy size
    - Required for velocity dispersion estimation
    - This means the image should include the lens light, which is an extra challenge for the convolutional engine (Pearson *et al* 2019).
  - Host galaxy size
    - To capture known degeneracy with γ'

Data generation: model assumptions

Motivated by the Time Delay Lens Modeling Challenge (TDLMC; Ding *et al* 2019), we simulate HST-like images with the assumed profiles:

- Elliptical power-law lens (PEMD)
- External shear
- Elliptical Sersic lens light
- Elliptical Sersic host galaxy light
- Point-source quasar



Detector and observation conditions:

- Drizzled HST PSFs → 0”.08/pixel
- WFC3/IR F160W band
- 64 x 64 pixels
- Sky, readout, CCD noise

The H0 inference stage

Assumptions: test prior = implicit prior, flat ΛCDM with fixed $\mathbf{\Omega}_m$=0.3. Our H0 prior = [50, 90] (Mpc/km/s)

For each individual lens, we MCMC sample from the joint posterior. Sampling was c ke:

$$
\begin{aligned}
p(H_0|\Delta t, \sigma_V, d^{(k)}) \\
\propto p(H_0) \int & p(\{\Delta t\}^{(k)}|D_{\Delta t}(H_0), \xi_{lens}^{(k)}, \xi_{light}^{(k)}, \kappa_{ext}^{(k)}) \\
& \times p(\sigma_V^{(k)}|D_{\Delta t}(H_0), \xi_{lens}^{(k)}, \xi_{light}^{(k)}, \kappa_{ext}^{(k)}, r_{ani}^{(k)}) \\
& \times p(\xi_{lens}^{(k)}, \xi_{light}^{(k)}|d^{(k)}) \\
& \times p(\kappa_{ext}^{(k)})p(r_{ani}^{(k)}) \quad d\left(\xi_{lens}^{(k)}, \xi_{light}^{(k)}\right) d\kappa_{ext}^{(k)} dr_{ani}^{(k)}. \qquad (3)
\end{aligned}
$$

Time delay likelihood

Velocity dispersion likelihood

BNN lens posterior

When combining the lenses in the test set of 200 lenses, we

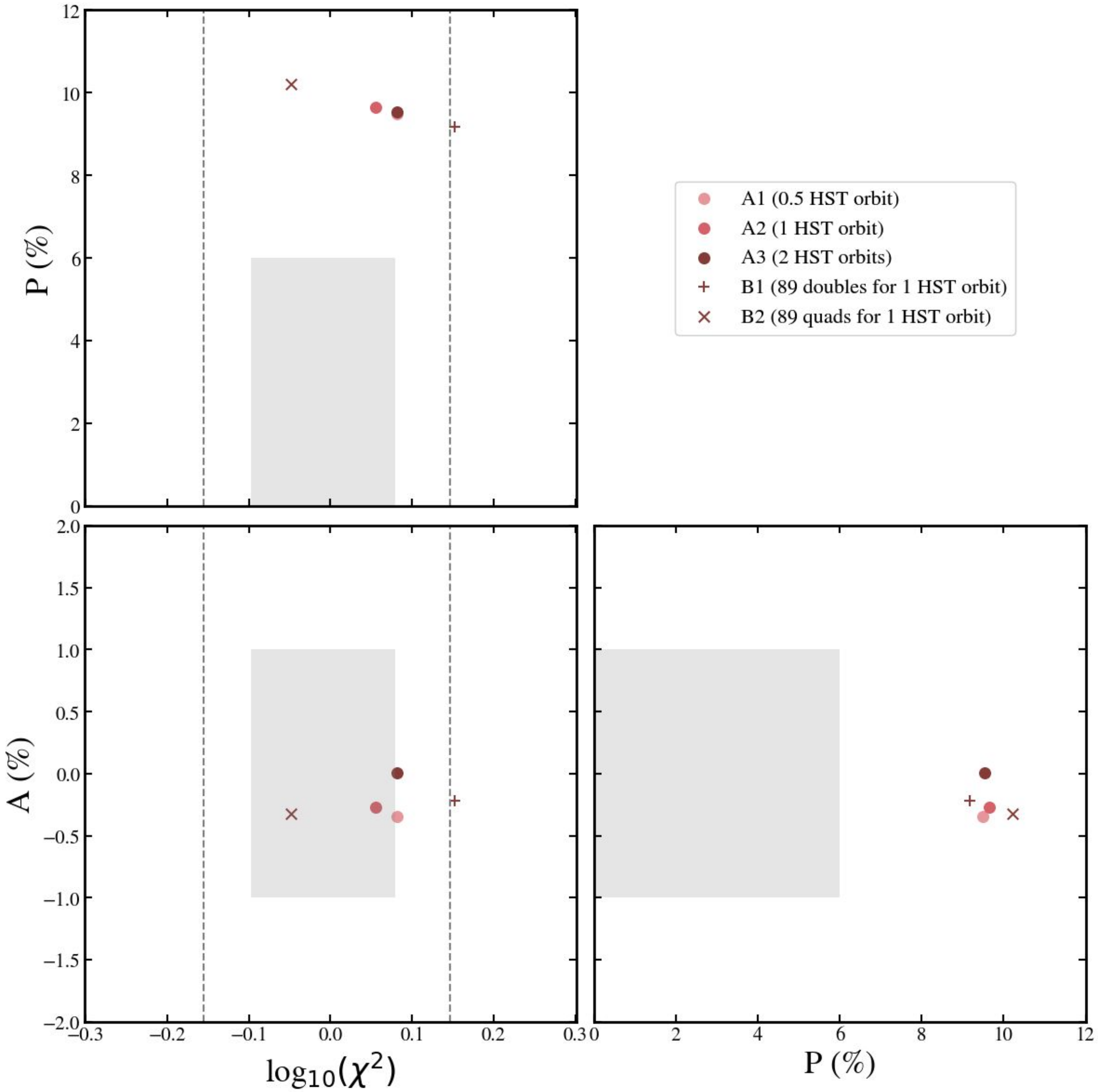
- Interpret individual D_dt posteriors as KDE
- MCMC sample from the combined D_dt posterior

Time Delay Lens Modeling Challenge (TDLMC) metrics

- **Precision:** average fractional estimated H0 uncertainty $$P \equiv \frac{1}{N_{success}} \sum_k \frac{\hat{\sigma}_k}{H_0}$$
- **Accuracy:** average fractional bias $$A \equiv \frac{1}{N_{success}} \sum_k \frac{\hat{\mu}_k - H_0}{H_0}$$
- **Goodness:** standard reduced χ^2 $$\chi^2 \equiv \frac{1}{N_{success}} \sum_i \left(\frac{\hat{\mu}_k - H_0}{\sigma_k}\right)^2$$

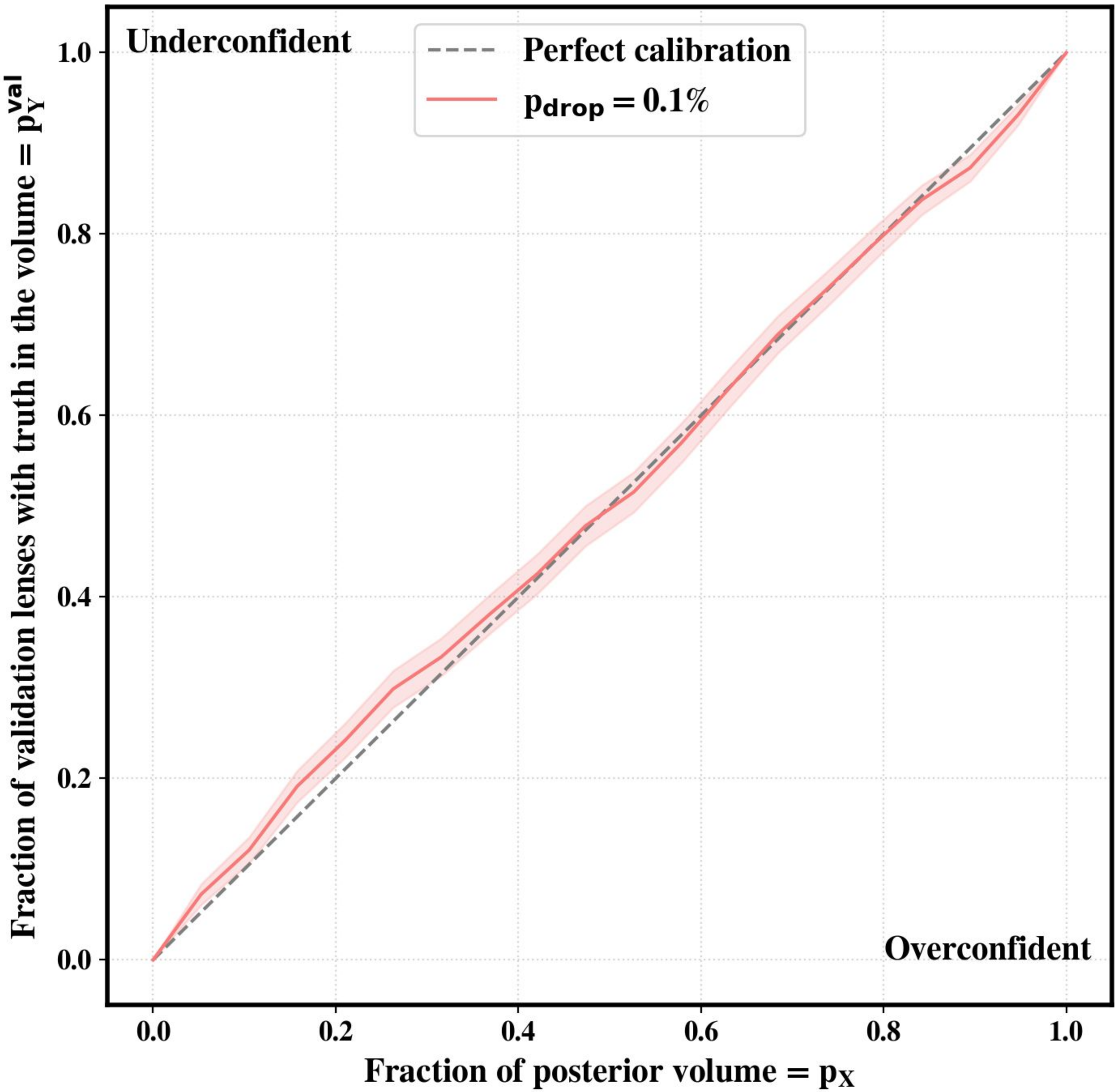
Note that TDLMC metrics weight each lens equally, regardless of the assigned uncertainty.

TDLMC metrics



Our experiments are accurate but fall behind in the precision.

Note: TDLMC metrics weight all lenses equally.


Underconfident
Perfect calibration
$p_{drop} = 0.1\%$
Overconfident
Fraction of validation lenses with truth in the volume = p_Y^{val}
Fraction of posterior volume = p_X
0.0
0.2
0.4
0.6
0.8
1.0

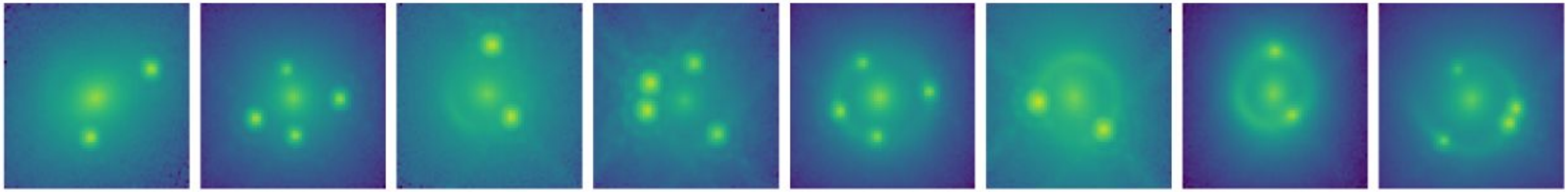
Simulated dataset and model assumptions

Model components

- Elliptical power-law lens (PEMD)
- External shear
- Elliptical Sersic lens light
- Elliptical Sersic host galaxy light
- Point-source AGN

Detector and observation conditions

- 512K simulated 64x64 HST-like images in the IR band
- Sky, readout, CCD noise



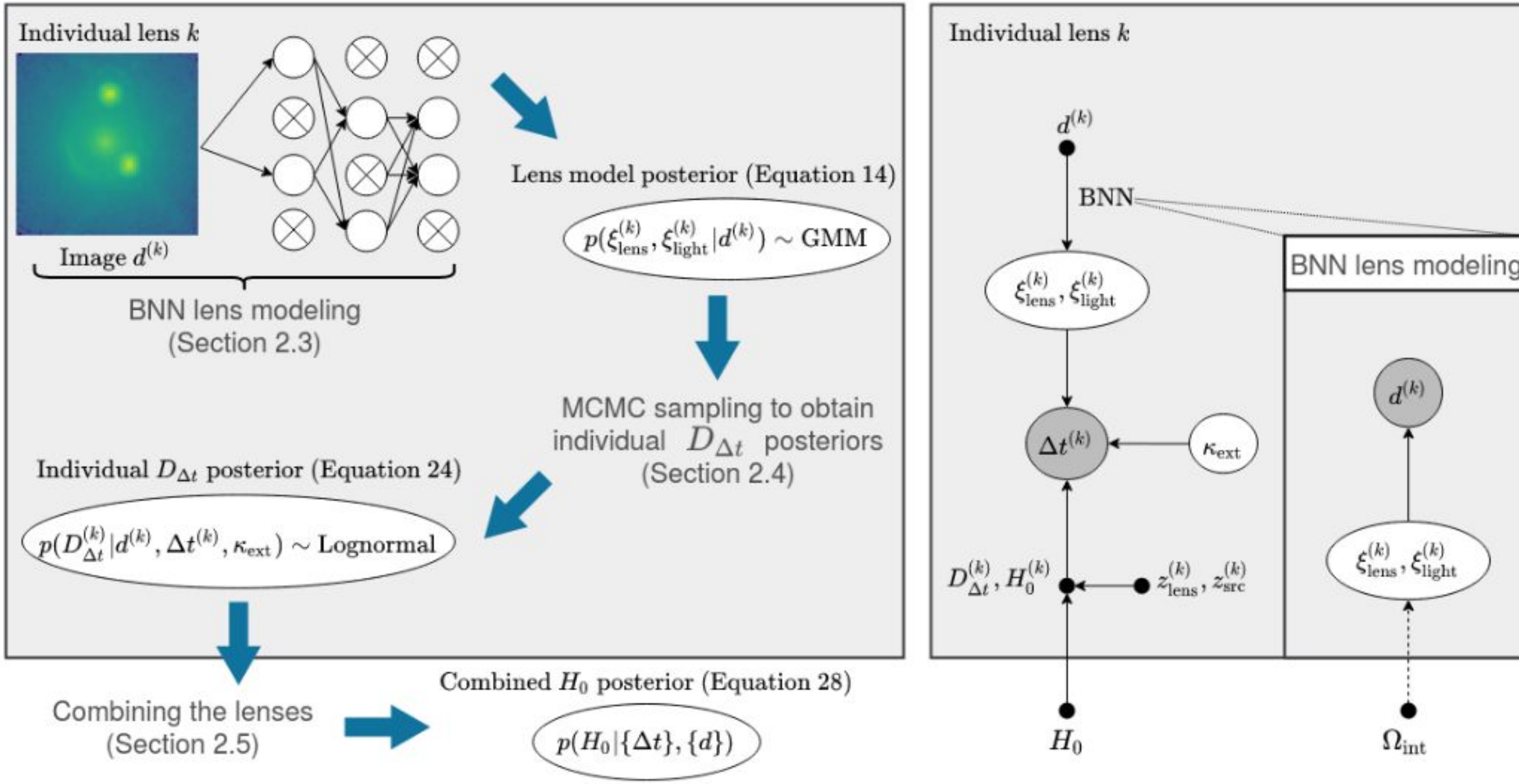
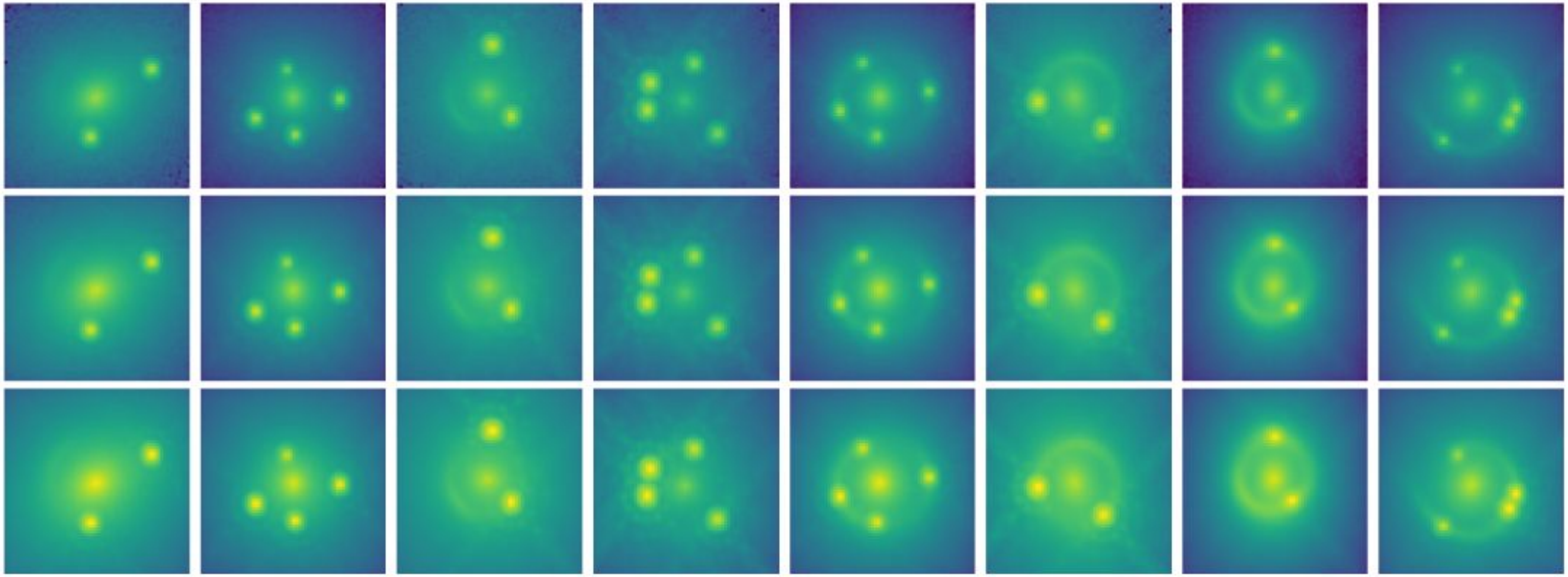


FIG. 1.— Left: illustration of the H_0 inference pipeline in the form of a flowchart. Right: the dependence relation shown as a probabilistic graphical model (PGM). Dots refer to delta functions, or fixed values; shaded ovals refer to observed values, or data; and unshaded ovals refer to random variables.



Parameter	Prior	Description
Flat ΛCDM cosmology		
H_0 (km Mpc^{-1} s^{-1})	$U(50, 90)$	Hubble constant
$\Omega_{\rm m}$	$\delta(0.3)$	Mass density
Mass profile		
$\xi^{(k)}_{\rm lens}, \xi^{(k)}_{\rm light}$	BNN-inferred lens model posterior (see Equation 19)	PEMD, external shear, source position/size
Line of sight		
$\kappa_{\rm ext}$	$\frac{1}{1-\kappa_{\rm ext}} \sim N(1, 0.025)$	External convergence

TABLE 2
SUMMARY OF MODEL PARAMETERS THAT ENTER INTO H_0 INFERENCE.

Parameter	Distribution
Lens redshift	$z_{\rm lens} \sim N(0.5, 0.2)$
Source redshift	$z_{\rm src} \sim N(2, 0.4)$
Lens galaxy	
Elliptical power-law mass	
Lens center ($''$)	$x_{\rm lens}, y_{\rm lens} \sim N(0, 0.07)$
Einstein radius ($''$)	$\theta_E \sim N(1.1, 0.1)$
Power-law slope	$\gamma_{\rm lens} \sim N(2.0, 0.1)$
Axis ratio	$q_{\rm lens} \sim N(0.7, 0.15)$
Orientation angle (rad)	$\phi_{\rm lens} \sim U(-\pi/2, \pi/2)$
Elliptical Sersic light	
Magnitude	$m_{\rm lens*} \sim U(19, 17)$
Half-light radius ($''$)	$R_{\rm lens*} \sim N(0.8, 0.15)$
Sersic index	$n_{\rm lens*} \sim N(3, 0.55)$
Axis ratio	$q_{\rm lens*} \sim N(0.85, 0.15)$
Orientation angle (rad)	$\phi_{\rm lens*} \sim U(-\pi/2, \pi/2)$
Environment	
External shear modulus	$\gamma_{\rm ext} \sim U(0, 0.05)$
Orientation angle (rad)	$\phi_{\rm ext} \sim U(-\pi/2, \pi/2)$
External convergence	$\kappa_{\rm ext} \sim N(0, 0.025)$
Host galaxy	
Elliptical Sersic light	
Host center ($''$)	$x_{\rm src}, y_{\rm src} \sim U(-0.2, 0.2)$
Host magnitude	$m_{\rm src} \sim U(25, 20)$
Half-light radius ($''$)	$R_{\rm src} \sim N(0.35, 0.05)$
Sersic index	$n_{\rm src} \sim N(3, 0.5)$
Axis ratio	$q_{\rm src} \sim N(0.6, 0.1)$
Orientation angle (rad)	$\phi_{\rm src} \sim U(-\pi/2, \pi/2)$
AGN	
Point source	
AGN magnitude	$m_{\rm AGN} \sim U(22.5, 20)$

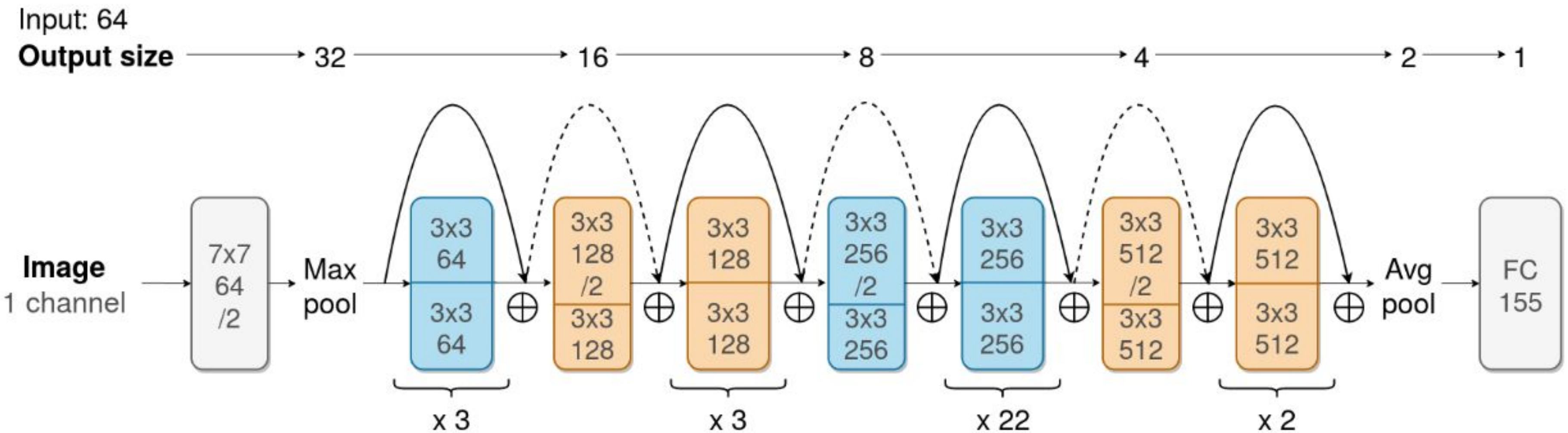


FIG. 3.— The `ResNet101` network architecture used for the convolutional engine of the BNN. The size of the square feature maps evolves through the layers as indicated on the top. Rectangular boxes contain convolutions of the indicated kernel size and channel number (width). Strides of 2 are denoted as /2. Note that blue and orange boxes are two stacked convolutions. Curved arrows indicate shortcut connections; the solid ones preserve the input feature dimension and dotted ones double the number of channels and halve the feature map resolution. Not shown are the 1D dropout layers, which were inserted before every convolutional layer and before the final fully-connected layer. Batch normalization and ReLU layers followed each convolution as well.

Individual H0 inference

$$
\begin{aligned}
& p\left(H_0|\Delta t^{(k)}, d^{(k)}\right) \\
& \propto p(H_0) p\left(\Delta t^{(k)}, d^{(k)}|H_0\right) \\
& \propto p(H_0) \int p\left(\Delta t^{(k)}|D_{\Delta t}^{(k)}(H_0), \xi_{\text{lens}}^{(k)}, \xi_{\text{light}}^{(k)}, \kappa_{\text{ext}}^{(k)}\right) \\
& \quad \times p\left(\xi_{\text{lens}}^{(k)}, \xi_{\text{light}}^{(k)}|d^{(k)}\right) \\
& \quad \times p\left(\kappa_{\text{ext}}^{(k)}\right) \quad d\left(\xi_{\text{lens}}^{(k)}, \xi_{\text{light}}^{(k)}\right) d\kappa_{\text{ext}}^{(k)}.
\end{aligned}
\tag{24}
$$

$$
\begin{aligned}
& p\left(\Delta t^{(k)}|D_{\Delta t}^{(k)}(H_0), \xi_{\text{lens}}^{(k)}, \xi_{\text{light}}^{(k)}, \kappa_{\text{ext}} = 0\right) \\
& \times p\left(\xi_{\text{lens}}^{(k)}, \xi_{\text{light}}^{(k)}|d^{(k)}\right).
\end{aligned}
\tag{25}
$$

MCMC objective

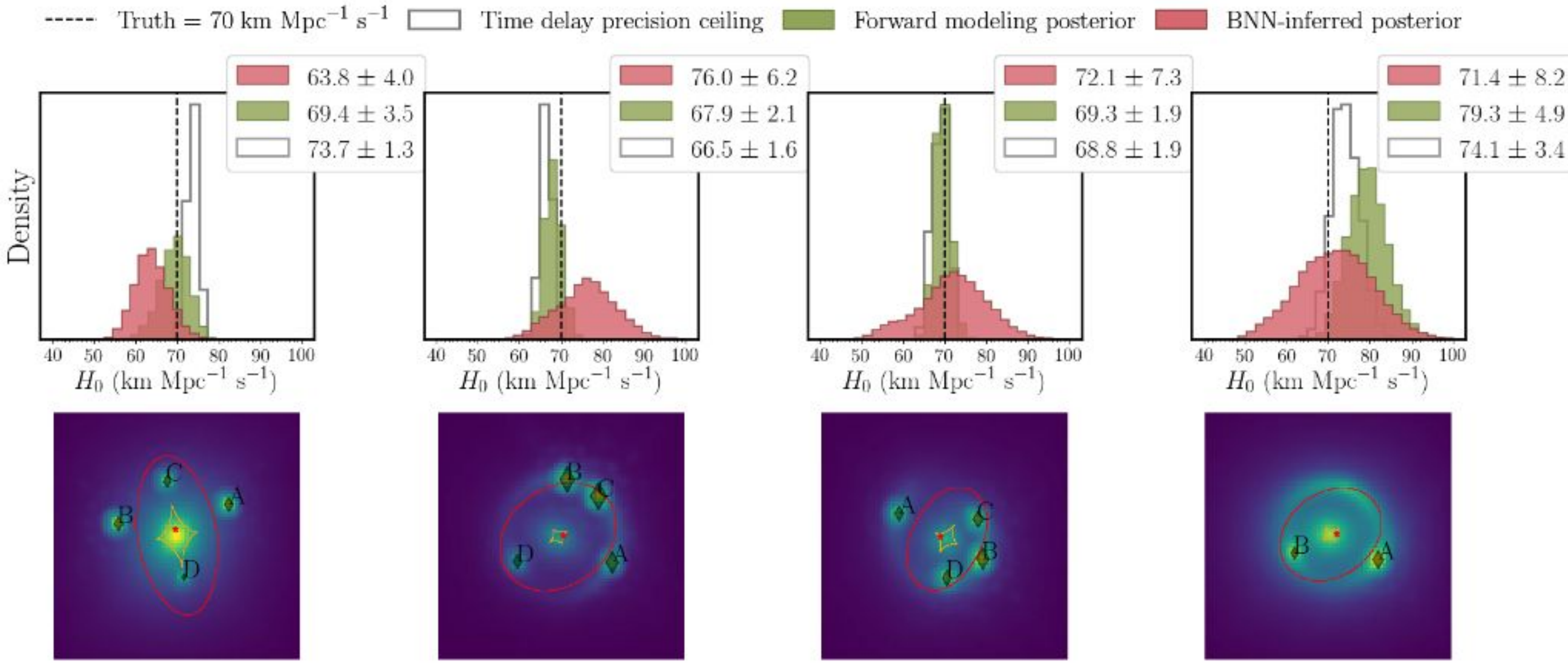
Joint-sample H0 inference

$$
\begin{aligned}
&p\left(H_0|\{\Delta t\},\{d\}\right) \\
&\propto p(H_0)p\left(\{\Delta t\},\{d\}|H_0\right) \\
&\propto p(H_0)\prod_k \int p\left(\Delta t^{(k)}|D_{\Delta t}^{(k)}(H_0),\xi_{\rm lens}^{(k)},\xi_{\rm light}^{(k)},\kappa_{\rm ext}^{(k)}\right) \\
&\quad \times p\left(\xi_{\rm lens}^{(k)},\xi_{\rm light}^{(k)}|d^{(k)}\right) \\
&\quad \times p\left(\kappa_{\rm ext}^{(k)}\right) \quad d\left(\xi_{\rm lens}^{(k)},\xi_{\rm light}^{(k)}\right) d\kappa_{\rm ext}^{(k)}.
\end{aligned}
\tag{27}
$$

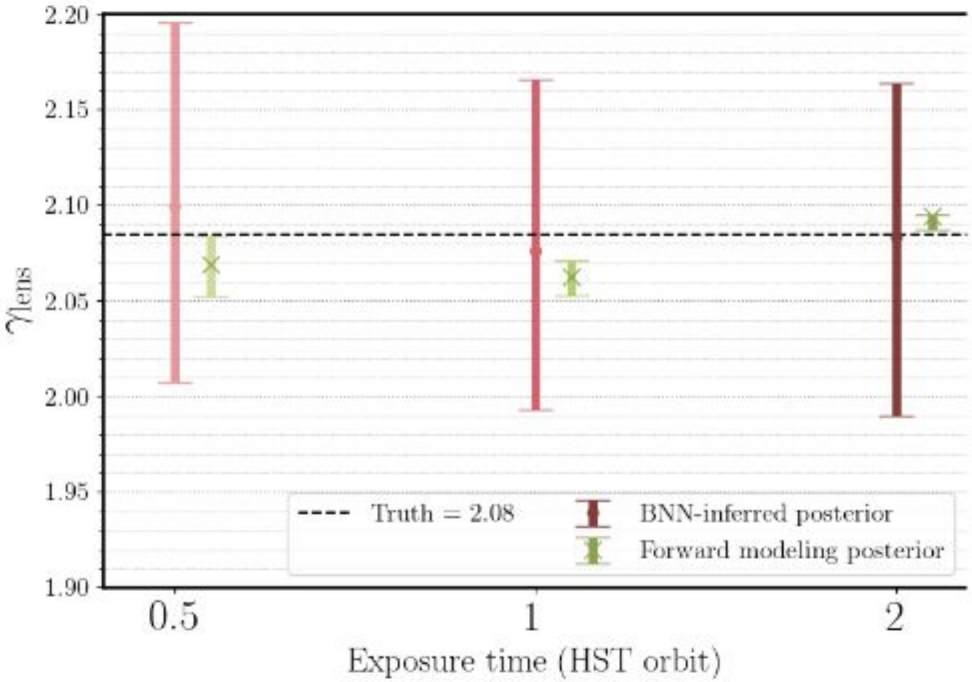
$$
p\left(H_0|\{\Delta t\},\{d\}\right) \propto p(H_0)\prod_k p\left(\Delta t^{(k)},d^{(k)}|D_{\Delta t}^{(k)}(H_0)\right). \tag{28}
$$

$$
\prod_k p\left(\Delta t^{(k)},d^{(k)}|D_{\Delta t}^{(k)}(H_0)\right). \tag{29}
$$

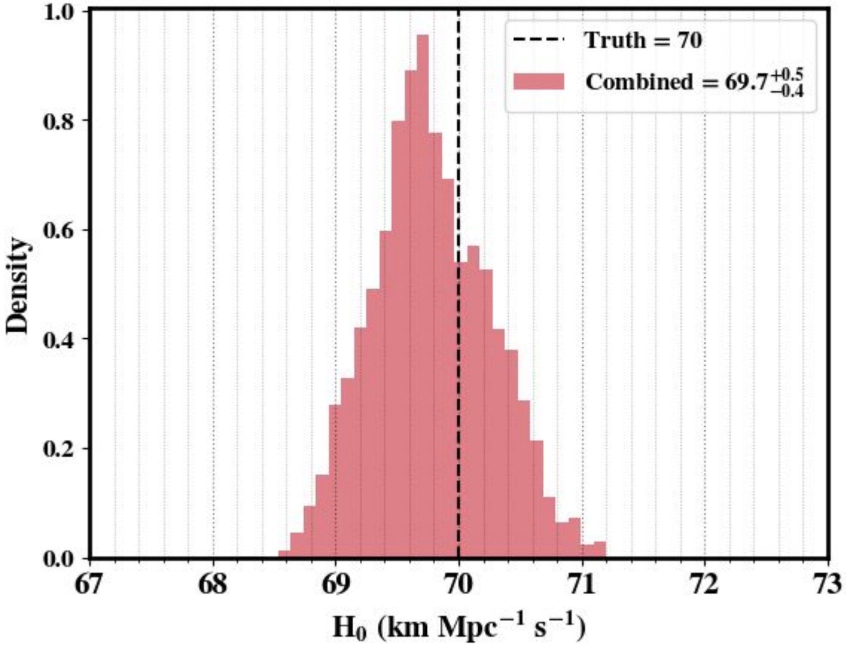
MCMC objective


Truth = 70 km Mpc^{-1} s^{-1}
Time delay precision ceiling
Forward modeling posterior
BNN-inferred posterior
63.8 ± 4.0
69.4 ± 3.5
73.7 ± 1.3
76.0 ± 6.2
67.9 ± 2.1
66.5 ± 1.6
72.1 ± 7.3
69.3 ± 1.9
68.8 ± 1.9
71.4 ± 8.2
79.3 ± 4.9
74.1 ± 3.4
Density
H_0 (km Mpc^{-1} s^{-1})
A
B
C
D

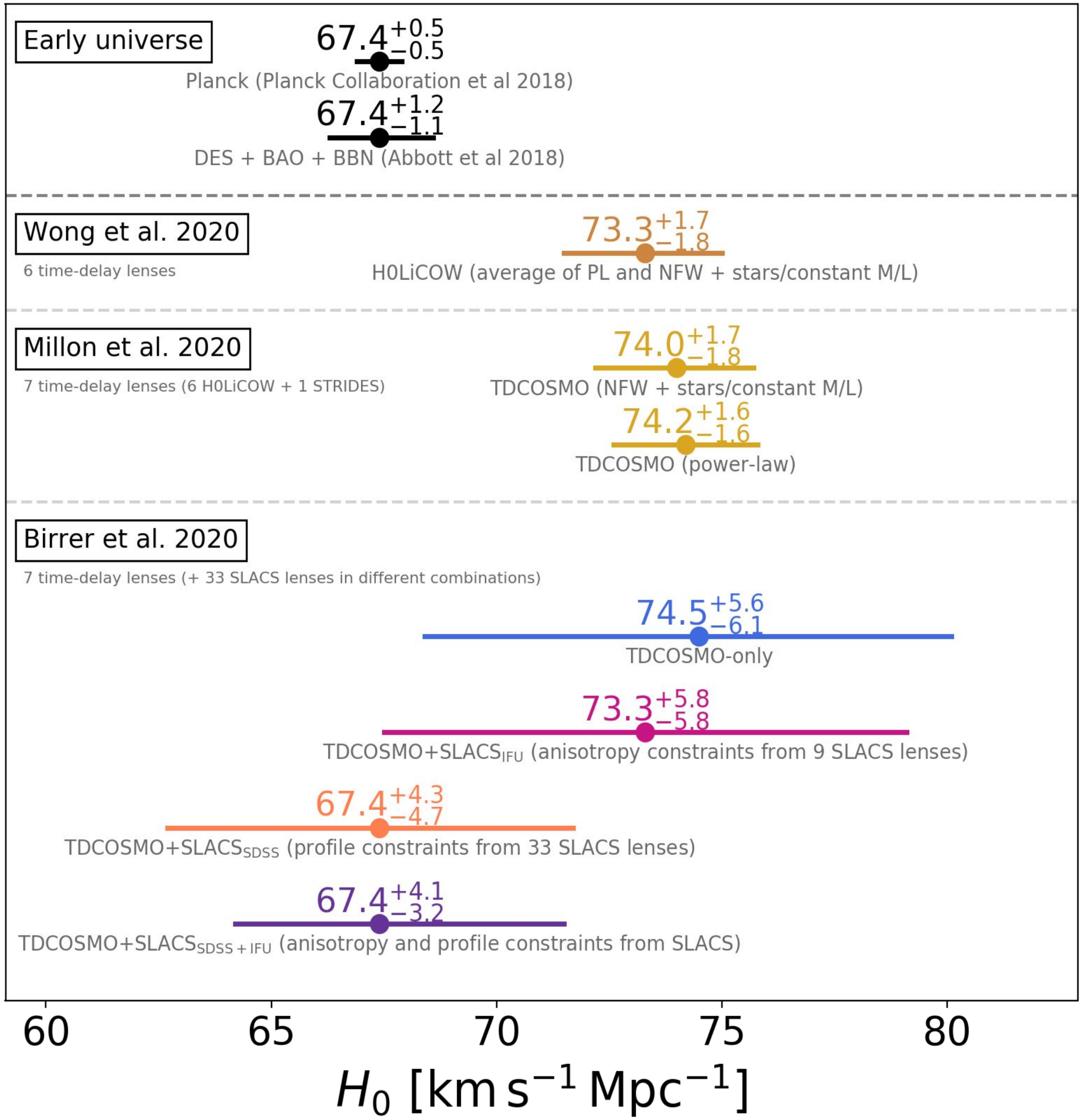
Comparison with forward modeling for 1 lens



Our combined H0 posterior



- Our posterior is consistent with the truth.
- The center of our posterior only 0.28 km/Mpc/s away from 70.0, which is 0.4% in H0 and ~50% of the posterior width.
- Statements beyond this level of precision will require knowing the shape of the D_dt posterior, out to the tails.
- Even if there exists some bias within that 0.4%, in applications to real data, errors due to other systematics would far outweigh this level of bias.


Early universe
67.4 +0.5 −0.5
Planck (Planck Collaboration et al 2018)
67.4 +1.2 −1.1
DES + BAO + BBN (Abbott et al 2018)
Wong et al. 2020
6 time-delay lenses
73.3 +1.7 −1.8
H0LiCOW (average of PL and NFW + stars/constant M/L)
Millon et al. 2020
7 time-delay lenses (6 H0LiCOW + 1 STRIDES)
74.0 +1.7 −1.8
TDCOSMO (NFW + stars/constant M/L)
74.2 +1.6 −1.6
TDCOSMO (power-law)
Birrer et al. 2020
7 time-delay lenses (+ 33 SLACS lenses in different combinations)
74.5 +5.6 −6.1
TDCOSMO-only
73.3 +5.8 −5.8
TDCOSMO+SLACS IFU (anisotropy constraints from 9 SLACS lenses)
67.4 +4.3 −4.7
TDCOSMO+SLACS SDSS (profile constraints from 33 SLACS lenses)
67.4 +4.1 −3.2
TDCOSMO+SLACS SDSS + IFU (anisotropy and profile constraints from SLACS)
60
65
70
75
80
H_0 [km s^{-1} Mpc^{-1}]