Power beyond the power spectrum: from cosmic origins to cosmic fate with computational models

Benjamin D. Wandelt

Early Computational Cosmology

- The Passemant astronomical clock (Versailles, 1749) keeps time, and computed the motion of the Copernican system, predicting eclipses and phases of the moon.

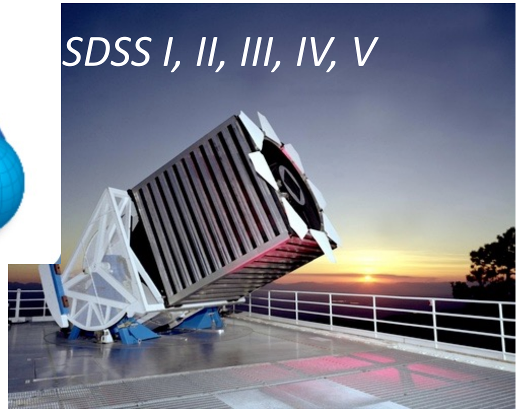
Exponential growth in observations

esa PLANCK

EUCLID
MAPPING THE GEOMETRY OF THE DARK UNIVERSE

SKA
SQUARE KILOMETRE ARRAY

SDSS I, II, III, IV, V



MegaPrime
CFH
MegaCam
Legacy Survey
CFHTLS

CMB-S4
Next Generation CMB Experiment

VERA C. RUBIN
OBSERVATORY

DARK ENERGY SPECTROSCOPIC INSTRUMENT

THE DARK ENERGY SURVEY

HYPER SUPRIME-CAM

ROMAN
WFIRST

(Your favorite survey here)

What we want to learn

Inflation non-Gaussianity

$A_s, n_s, r, f_{nl}, \ldots$ **Cosmic Origin**

Primordial gravitational waves

Neutrino mass Black holes

$\Omega_m, \Omega_b, m_\nu, \tau, \ldots$ **Cosmic Content**

Dark matter baryons Star formation

$H_0, \Omega_\Lambda, w_0, w_a, \ldots$ **Cosmic Fate**

Expansion speed Dark Energy

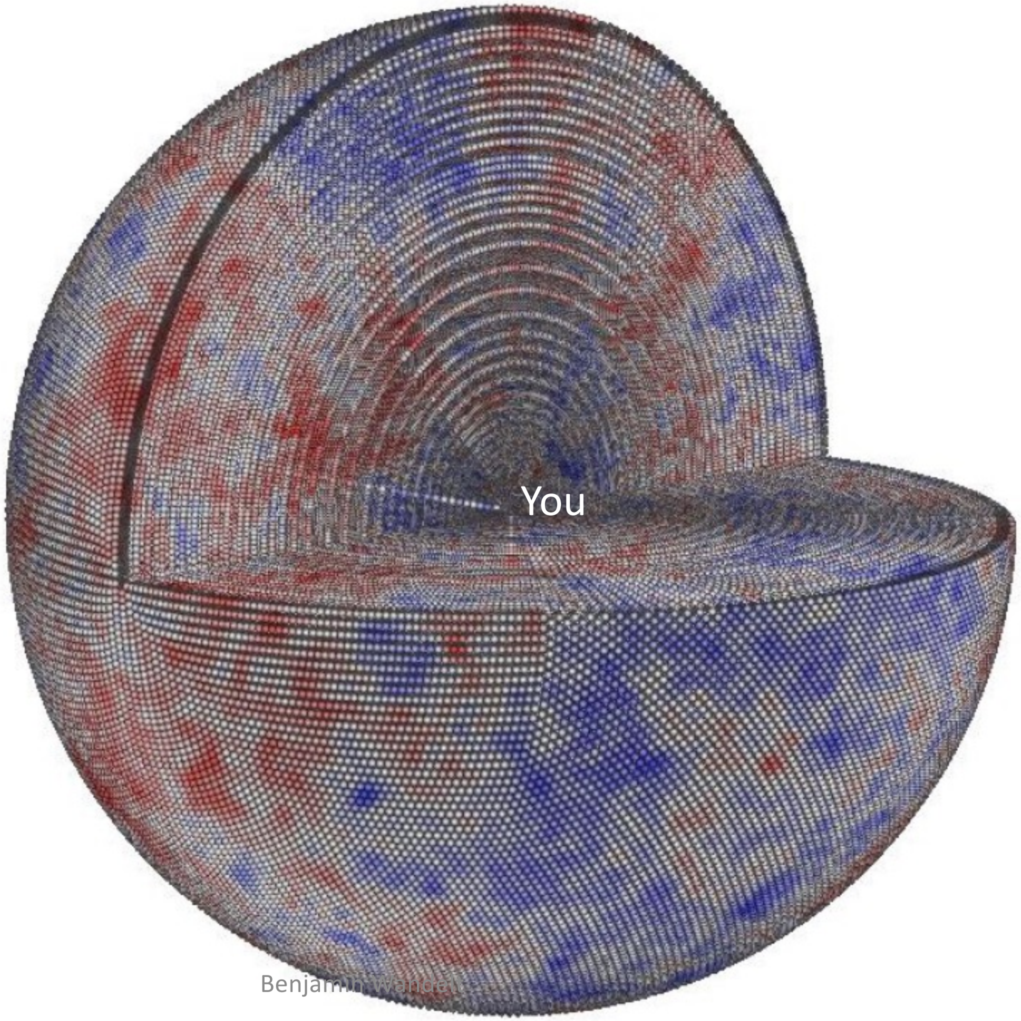
The "initial conditions"

Recipe for our universe

Dynamics

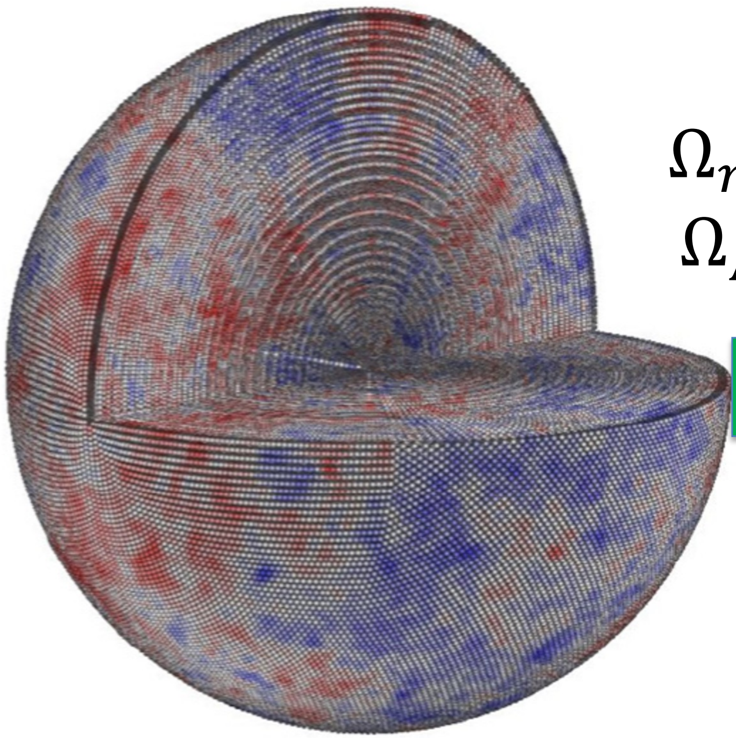
Expansion geometry

The Initial Conditions



Cosmological **Computation**

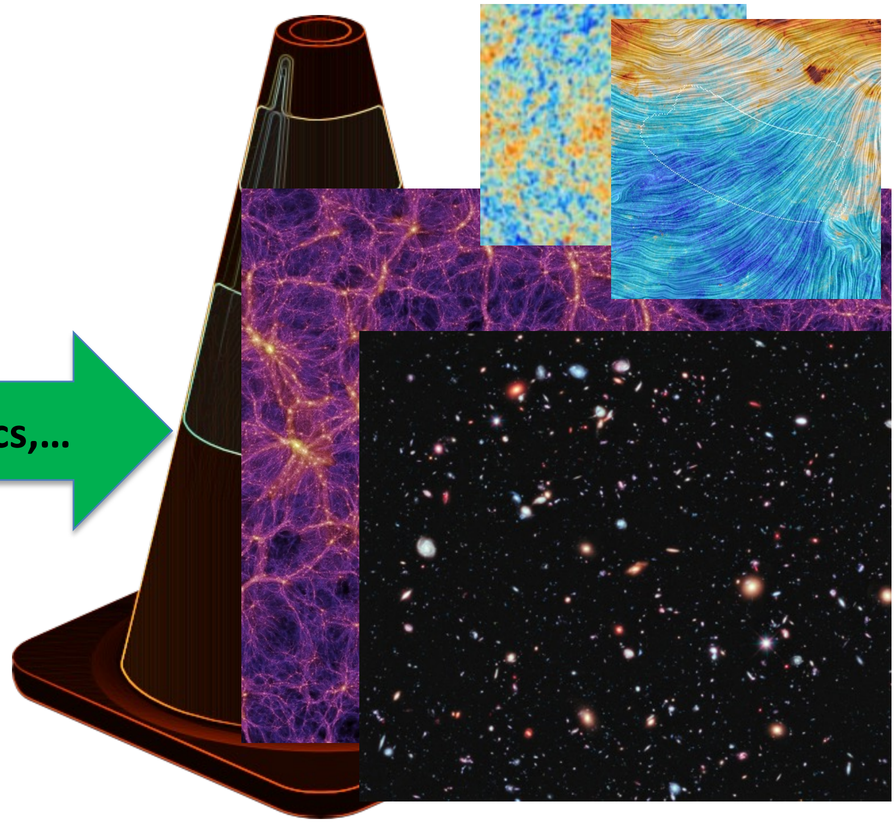
$A_s, n_s, r, f_{nl}, \ldots$



$\Omega_m, \Omega_b, m_\nu, \ldots$

$\Omega_\Lambda, w_0, w_a, \ldots$

Gravity, Hydrodynamics,...



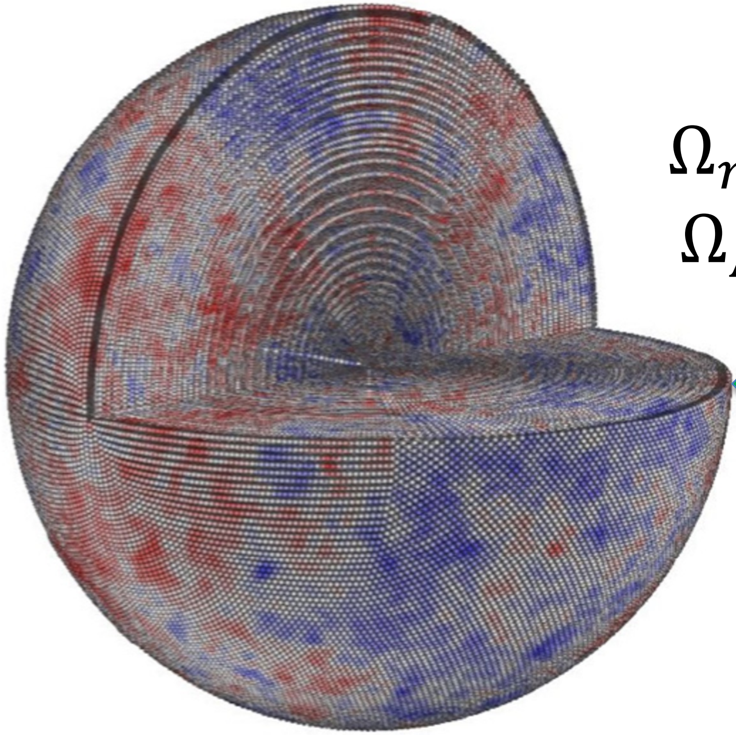
Initial conditions of the universe

The observed universe

Benjamin Wandelt

The Cosmological **Inference Problem**

$A_s, n_s, r, f_{nl}, \ldots$

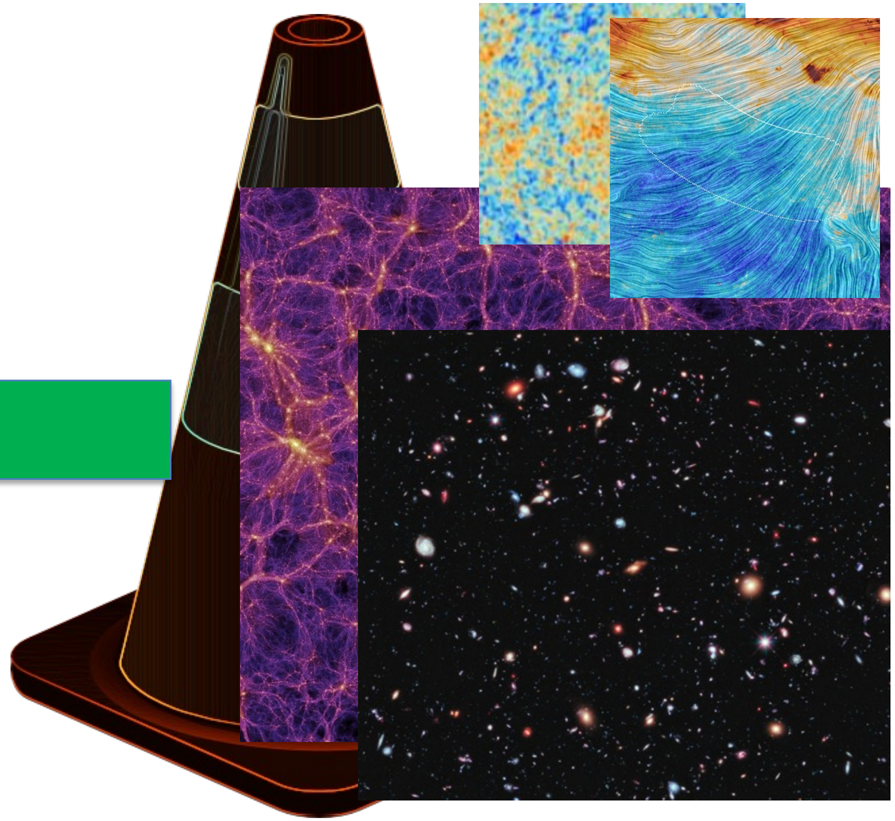


Initial conditions of the universe

$\Omega_m, \Omega_b, m_\nu, \ldots$

$\Omega_\Lambda, w_0, w_a, \ldots$

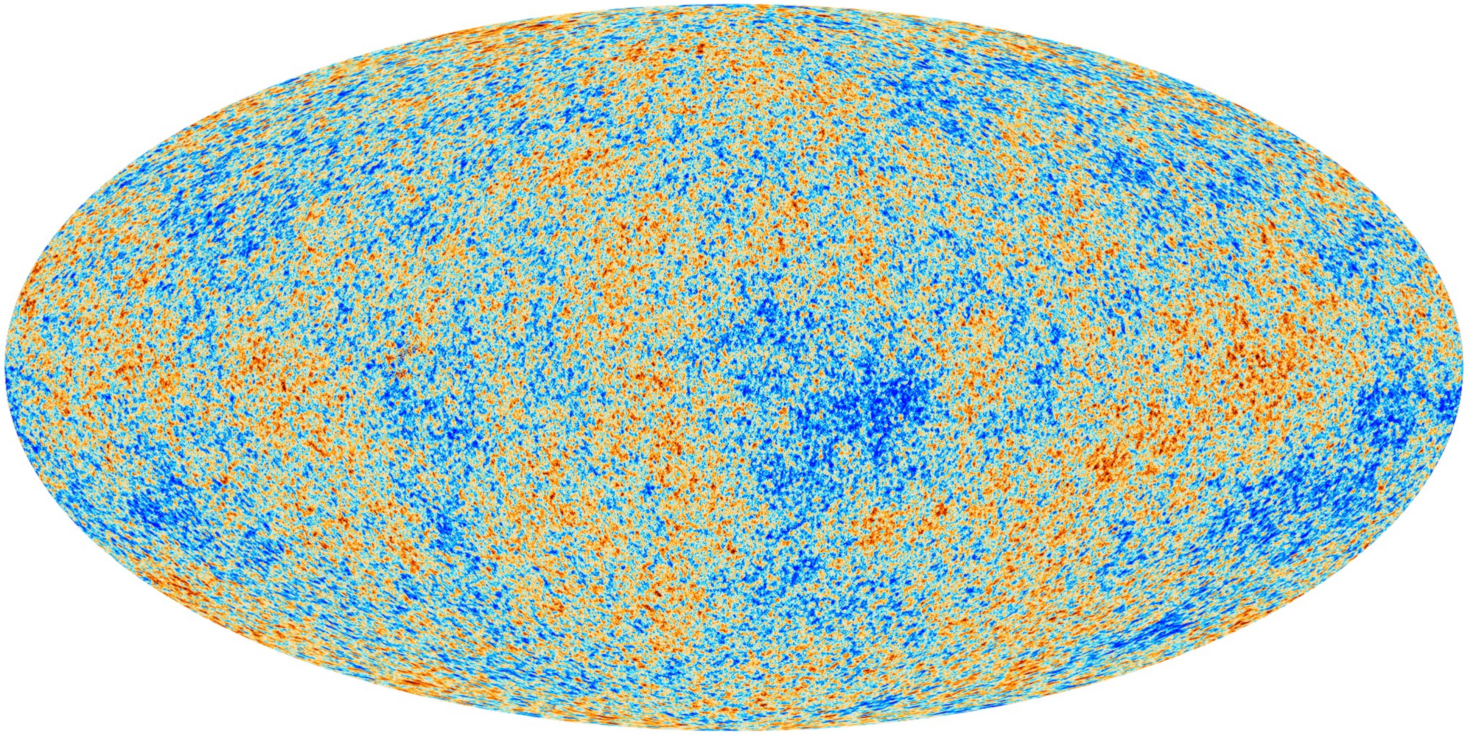
Inference



The observed universe

Benjamin Wandelt

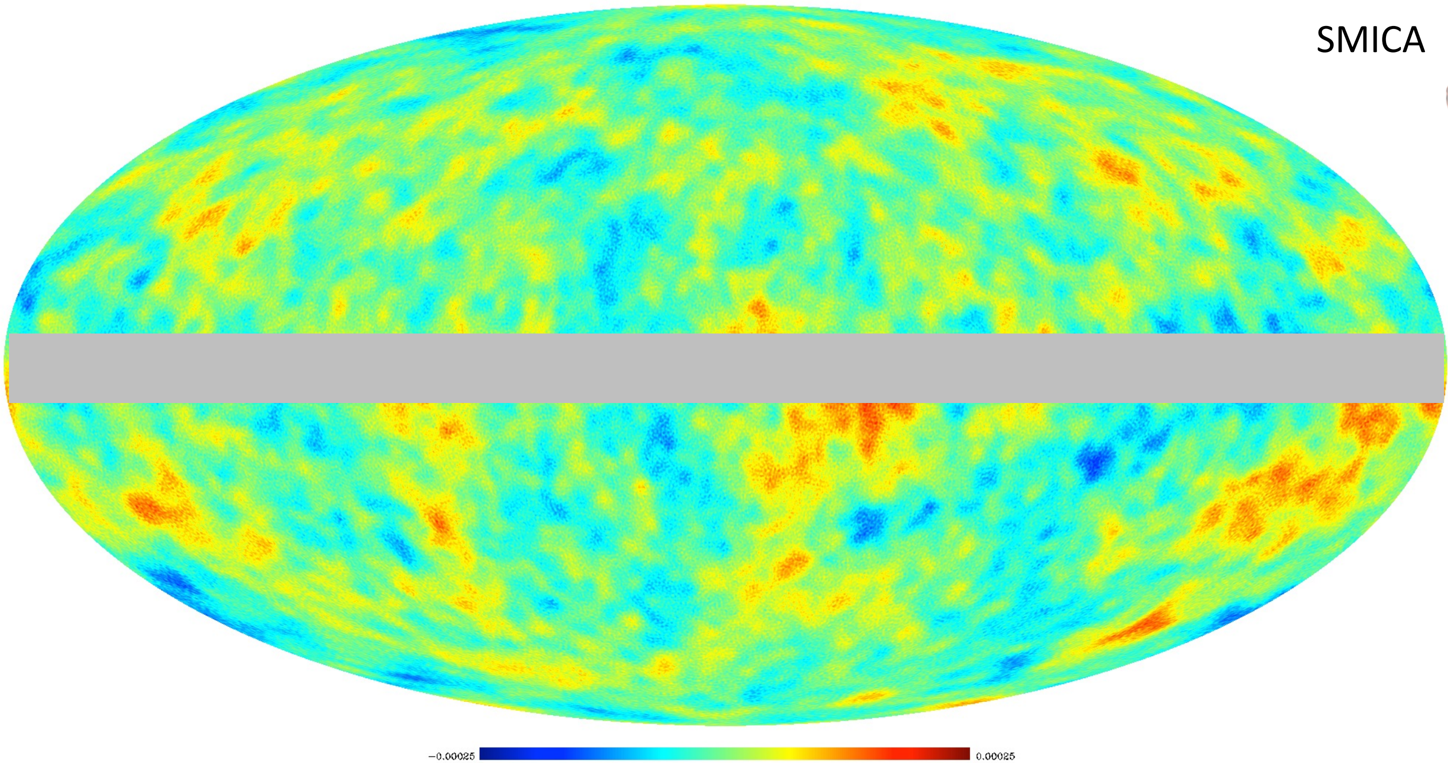
We have done it for the
Cosmic Microwave Background anisotropies;
a linear time machine

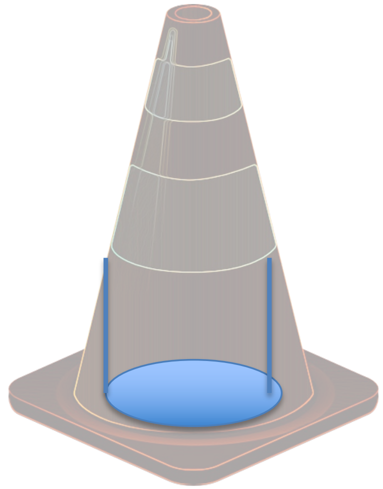


Benjamin Wandelt

Primordial curvature perturbations at $r = \eta_{CMB}$ from Planck T data

Method: Komatsu, Spergel, Wandelt 2005

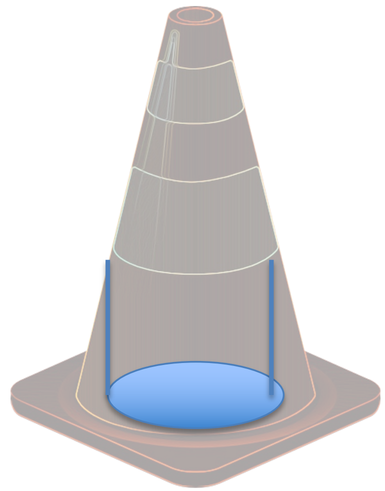




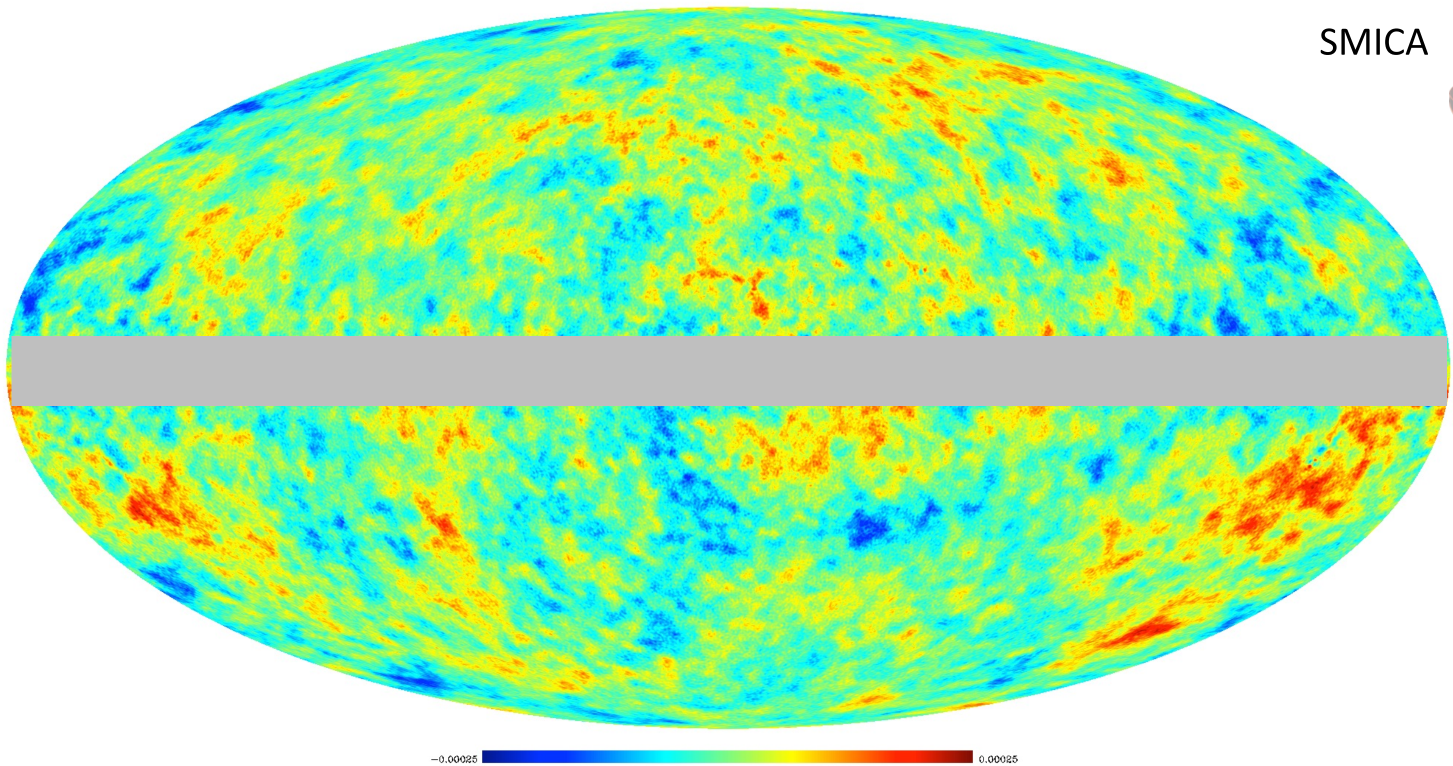
SMICA

Planck collaboration 2013 data

Primordial curvature perturbations at $r = \eta_{CMB}$ from both Planck T and polarization data



SMICA



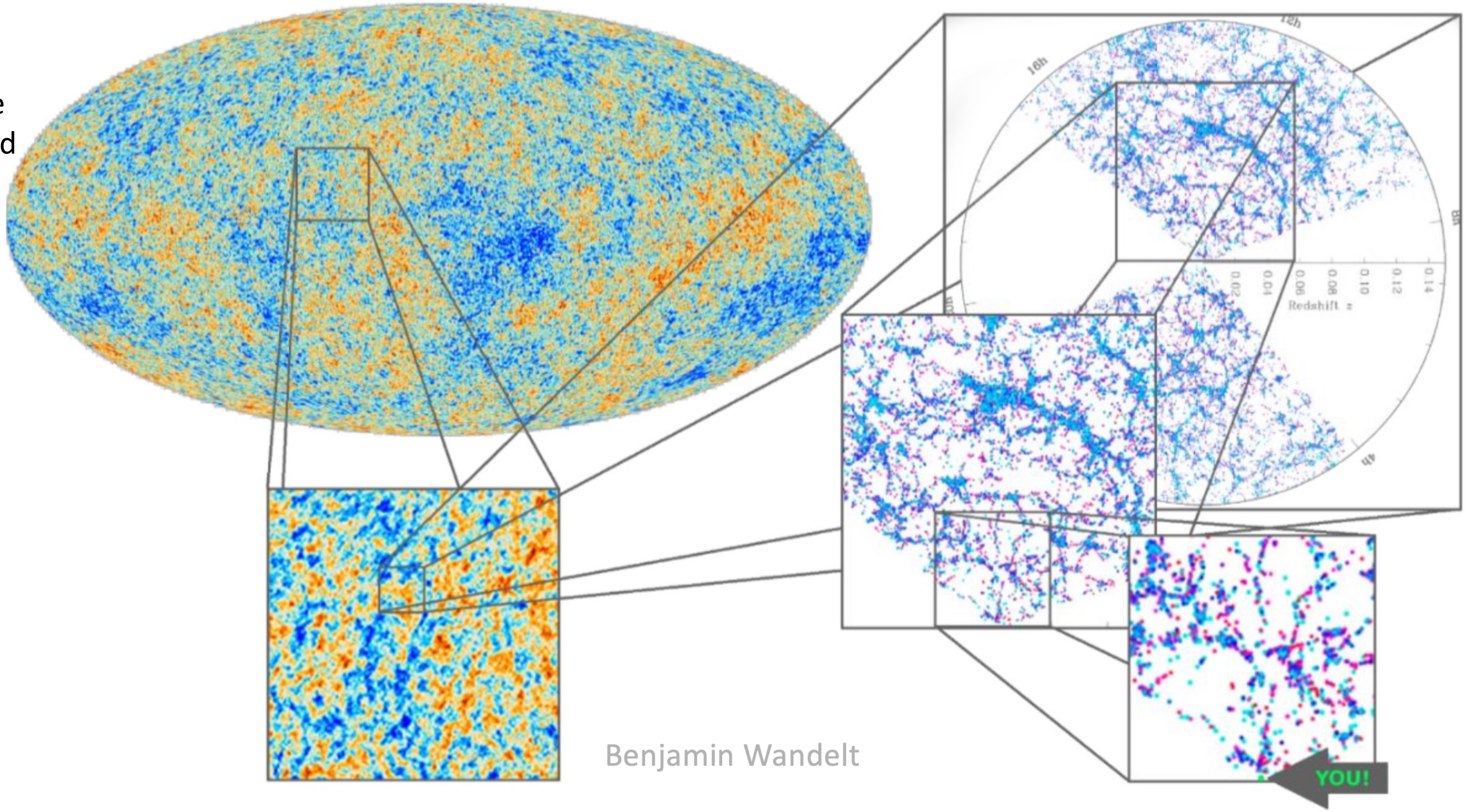
Planck collaboration
2013 data

Method: Yadav & Wandelt 2005

… leading to many results, e.g. Planck constraint
inflation (spectra, non-Gaussianity) and other early universe parad

Cosmological data covers a hierarchy of scales.
Smaller scales ➔ increasing complexity

Cosmic Microwave Background



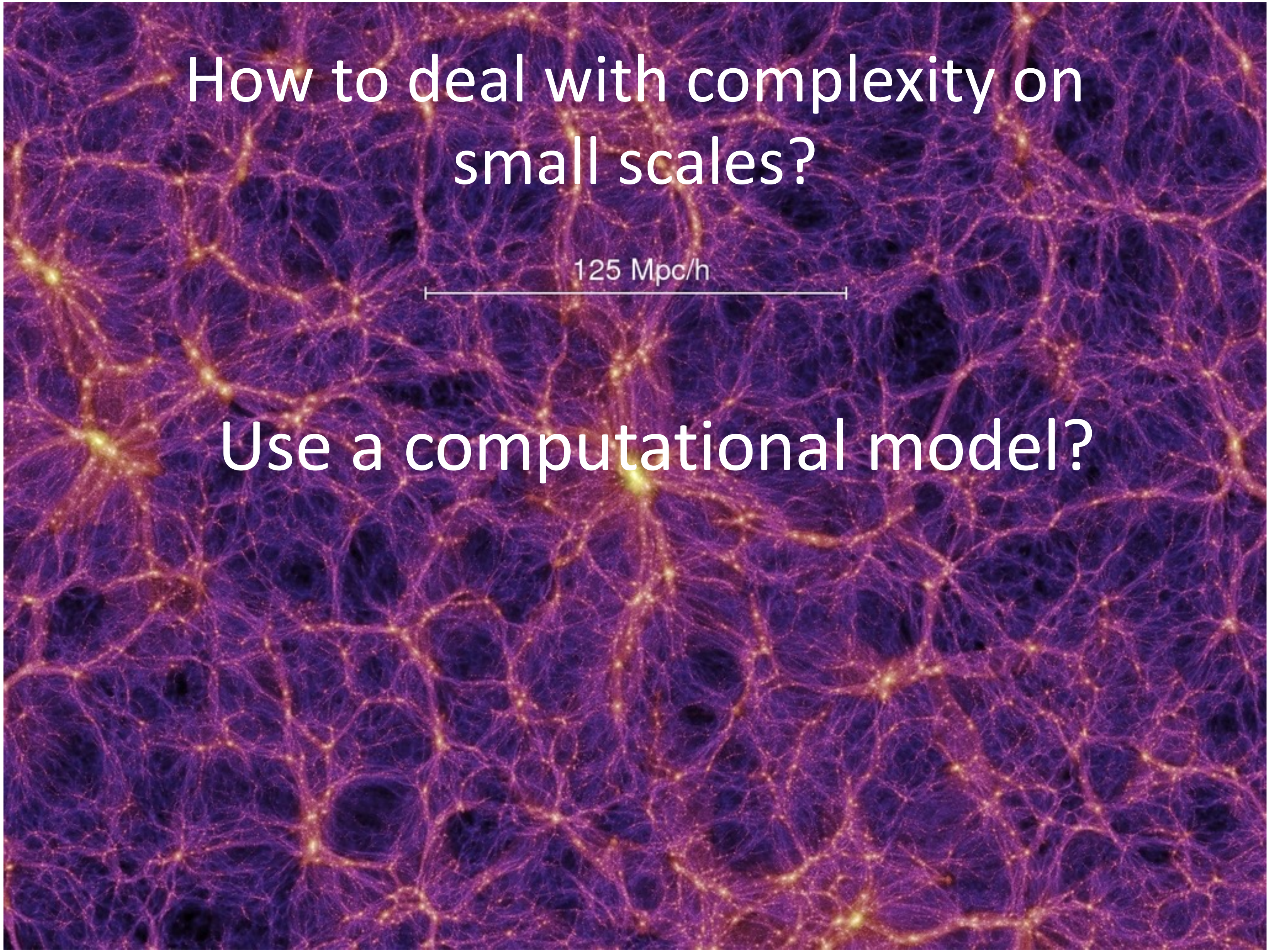
Benjamin Wandelt

Galaxy surveys


How to deal with complexity on small scales?
125 Mpc/h

How to deal with complexity on small scales?

Smooth away small scales?


How to deal with complexity on small scales?
125 Mpc/h
Use a computational model?

Can we analyze data if all we can do is simulate?

Yes!

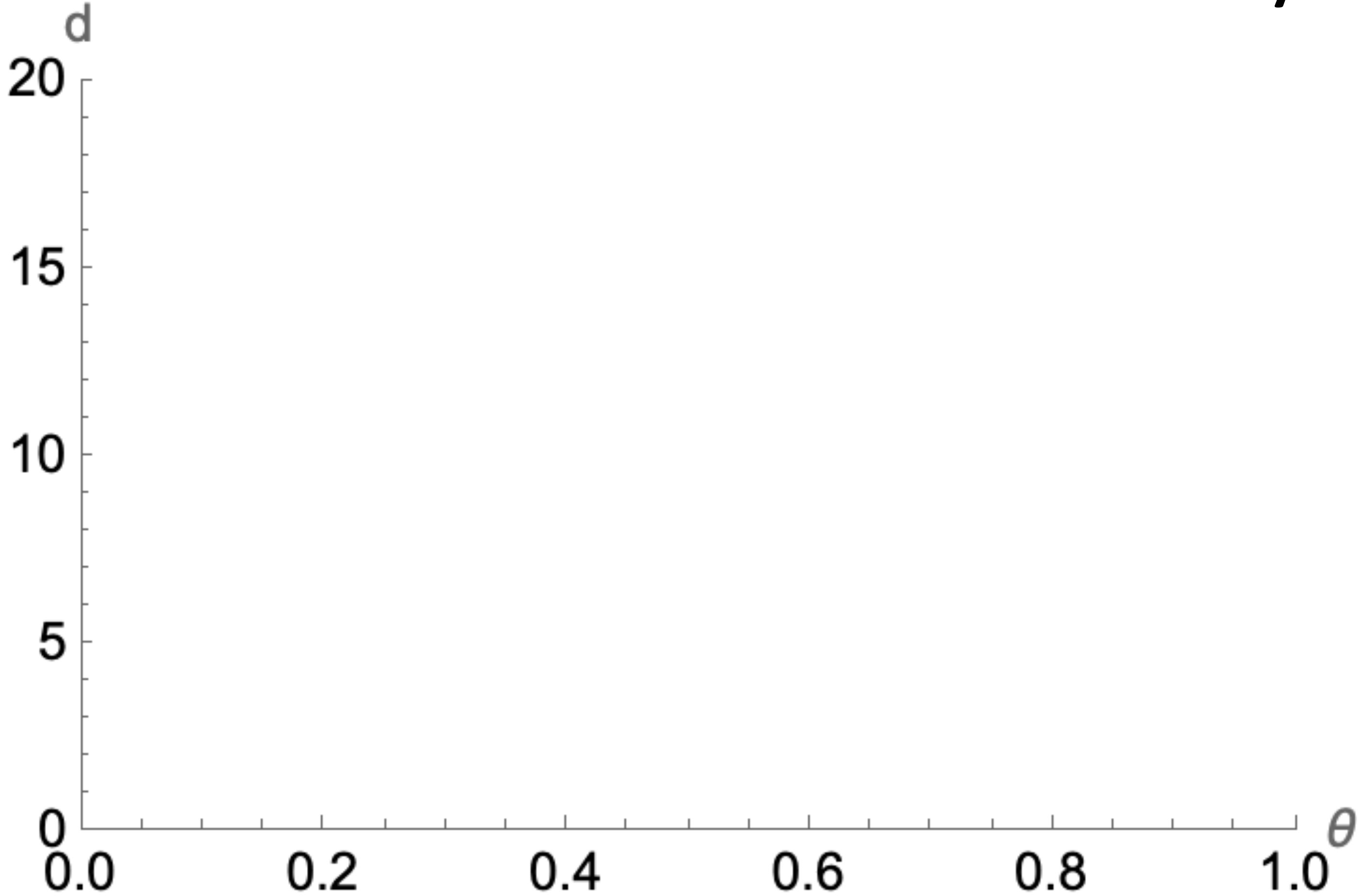
A major shift over the last 5 years.

Likelihood is represented implicitly through simulations

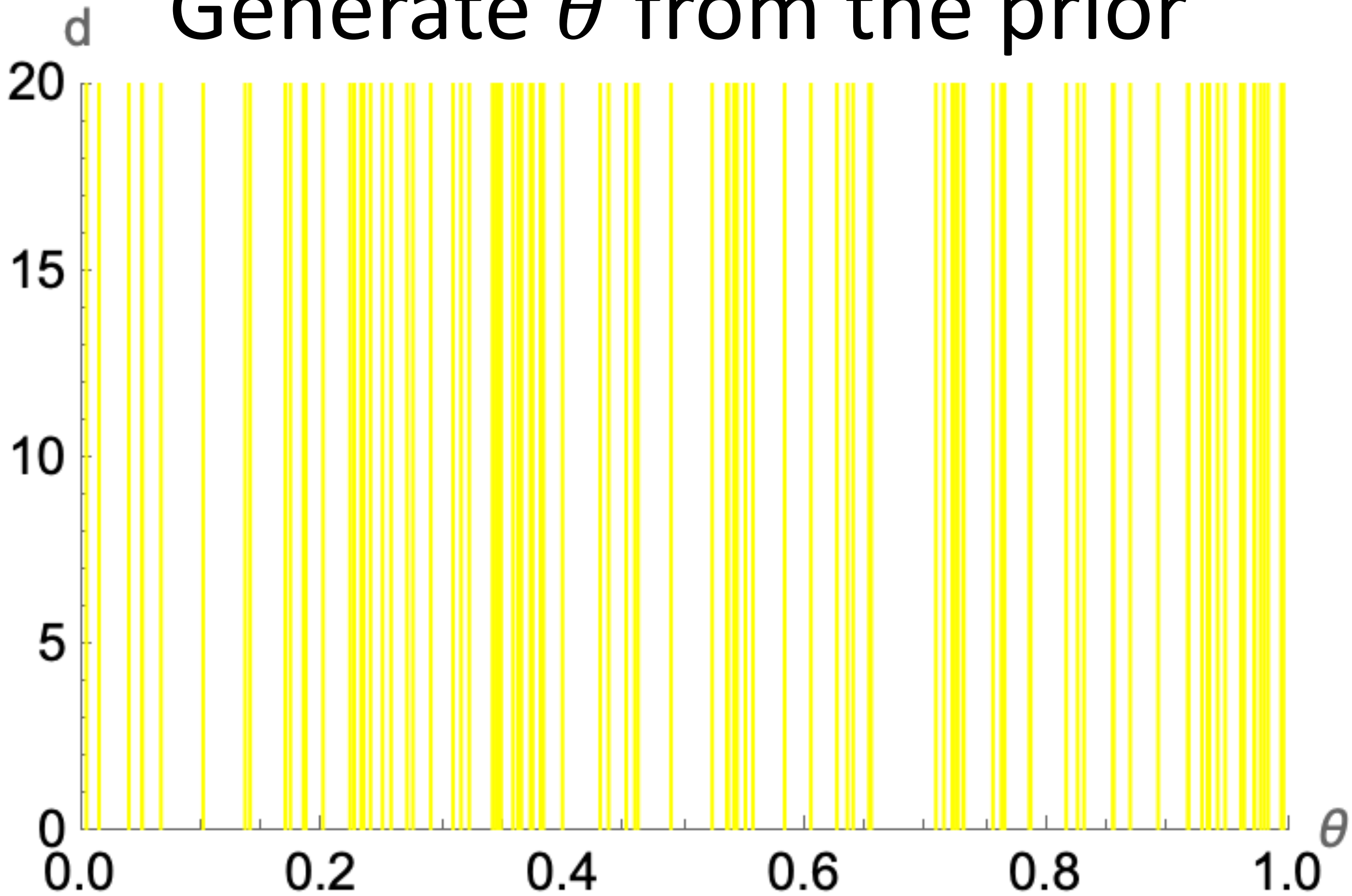
$$d \hookleftarrow p(d|\theta)$$

Let's do a simple example.

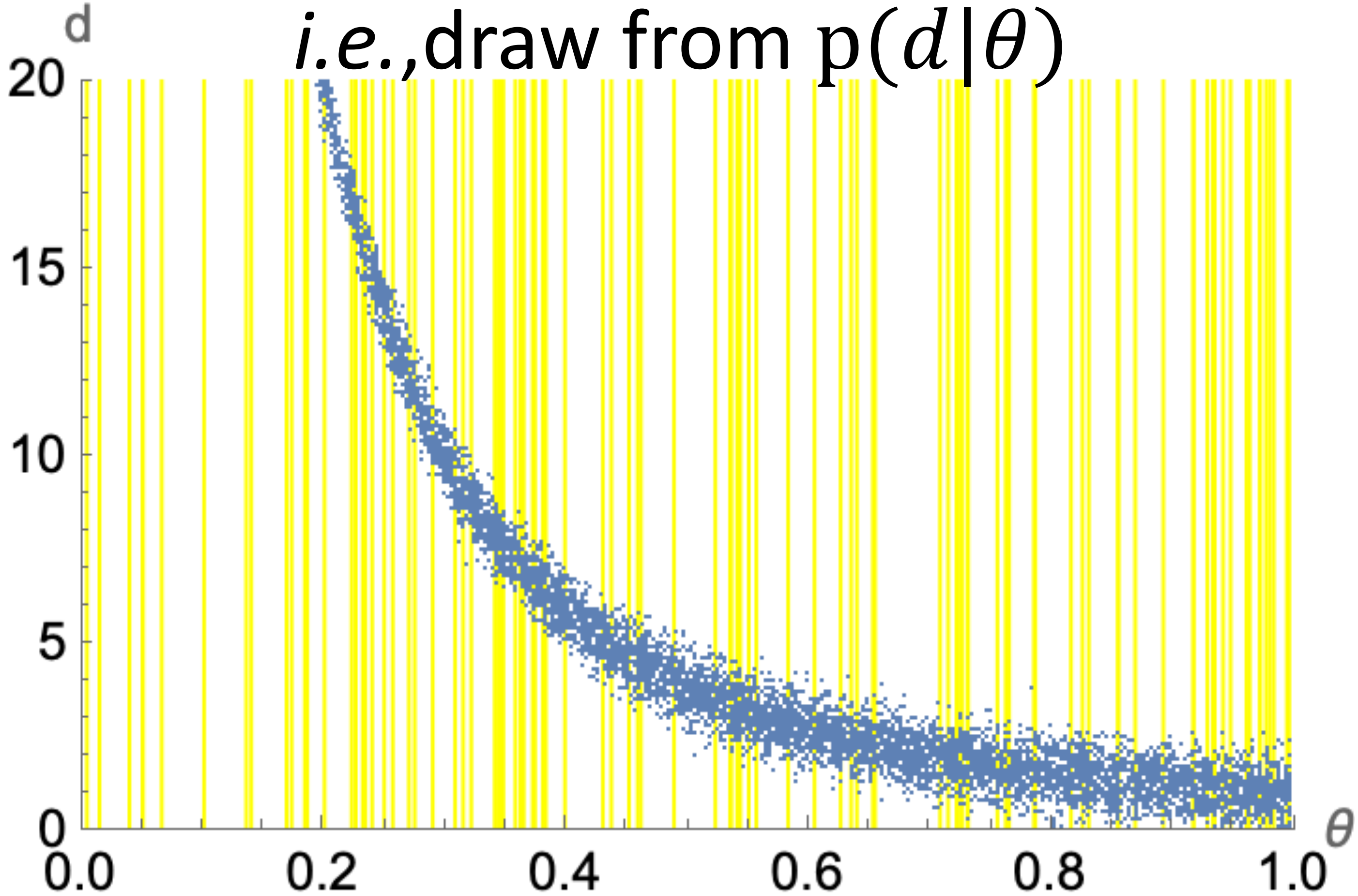
Inference from simulations only



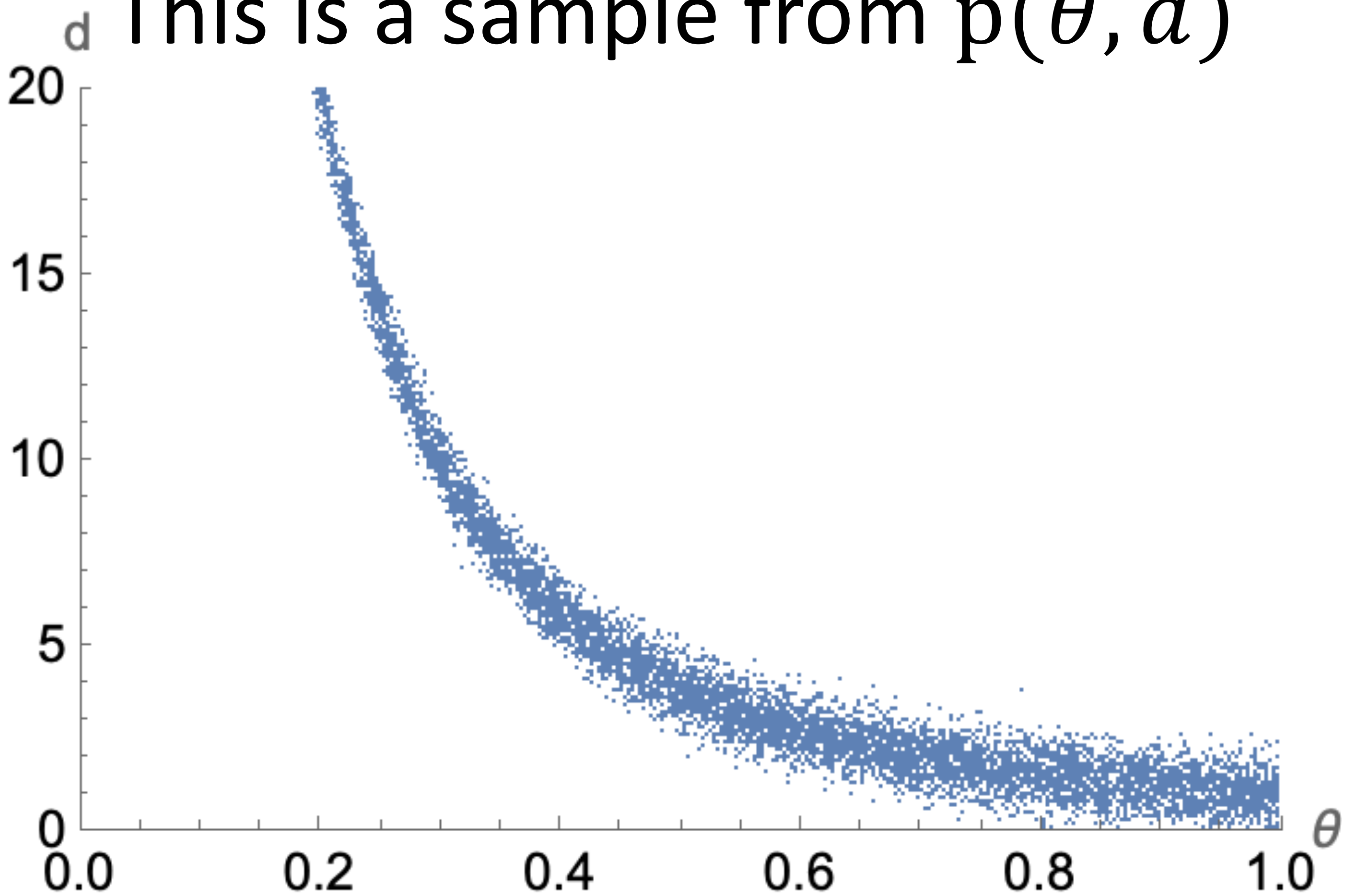
Generate θ from the prior



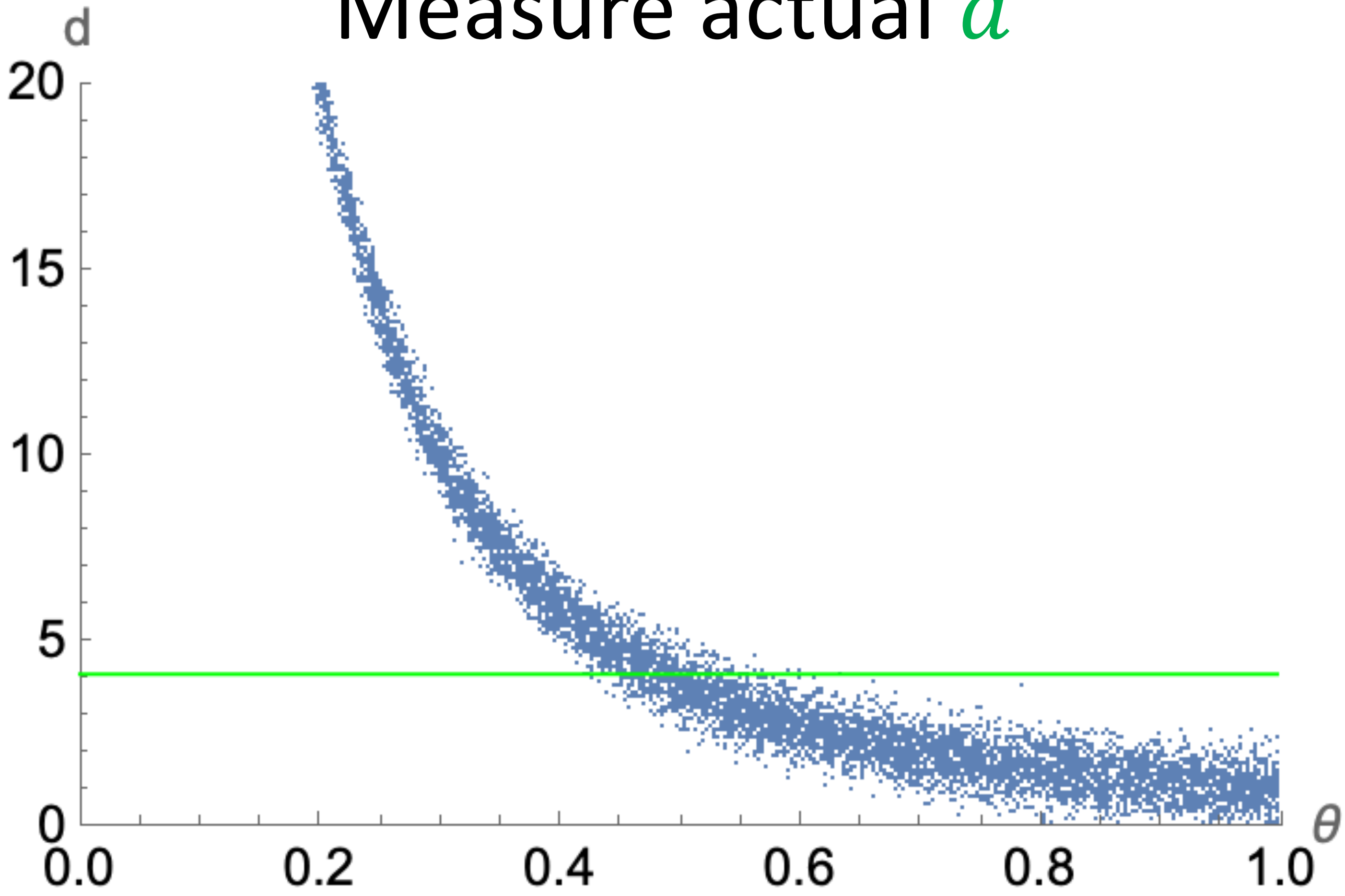
Simulate/generate d given θ
i.e.,draw from $\mathrm{p}(d|\theta)$



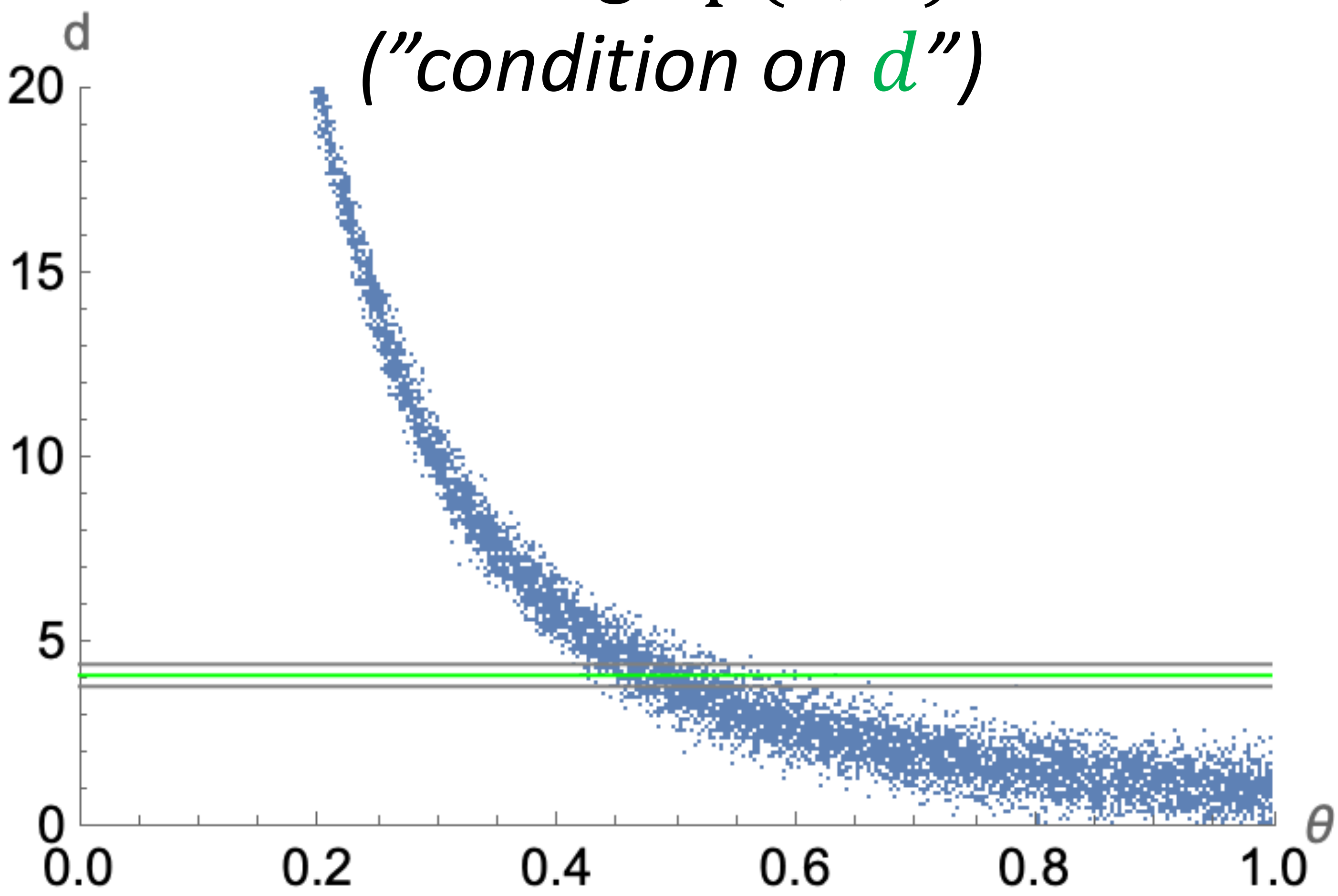
This is a sample from $p(\theta, d)$



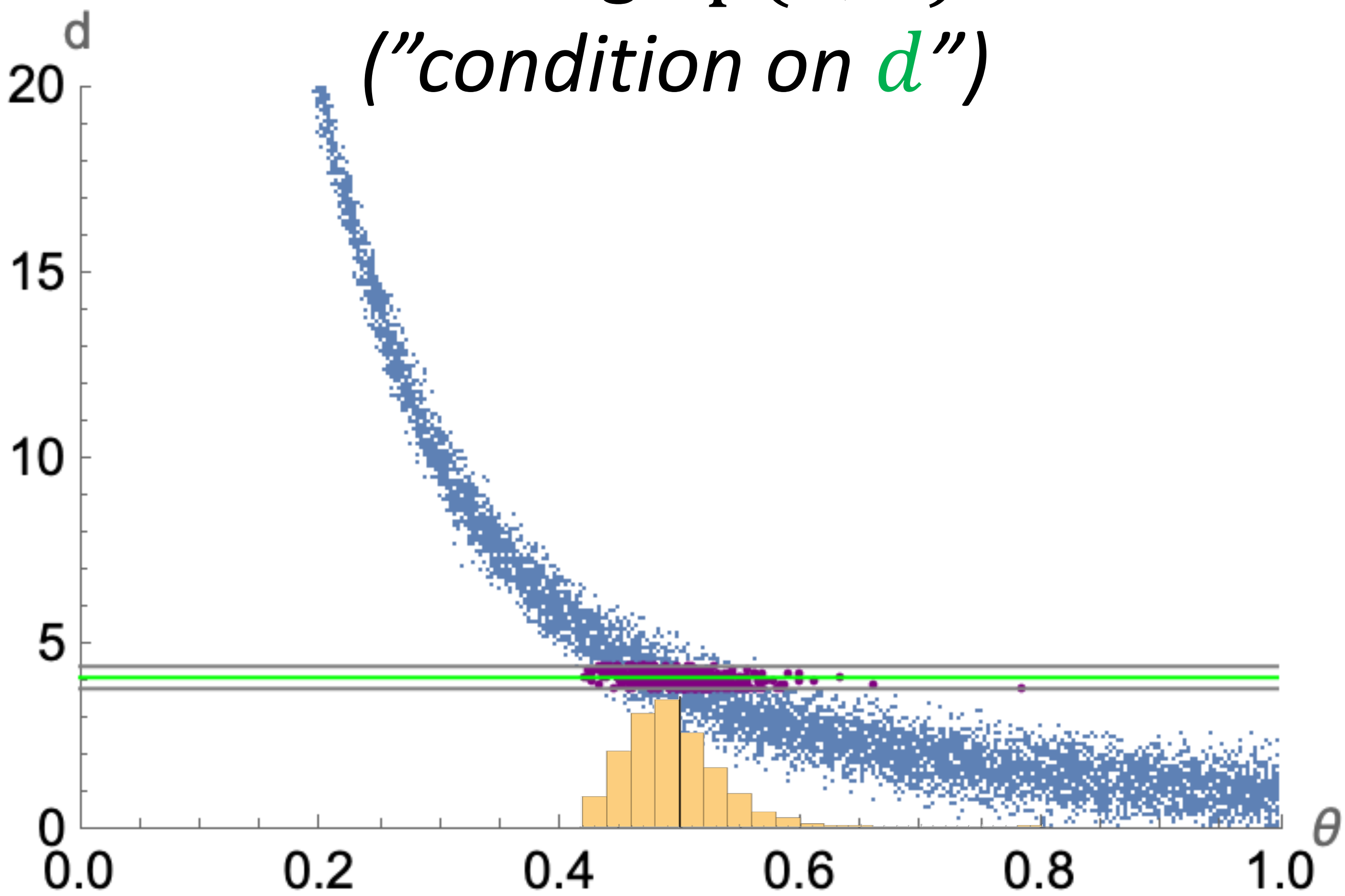
Measure actual d



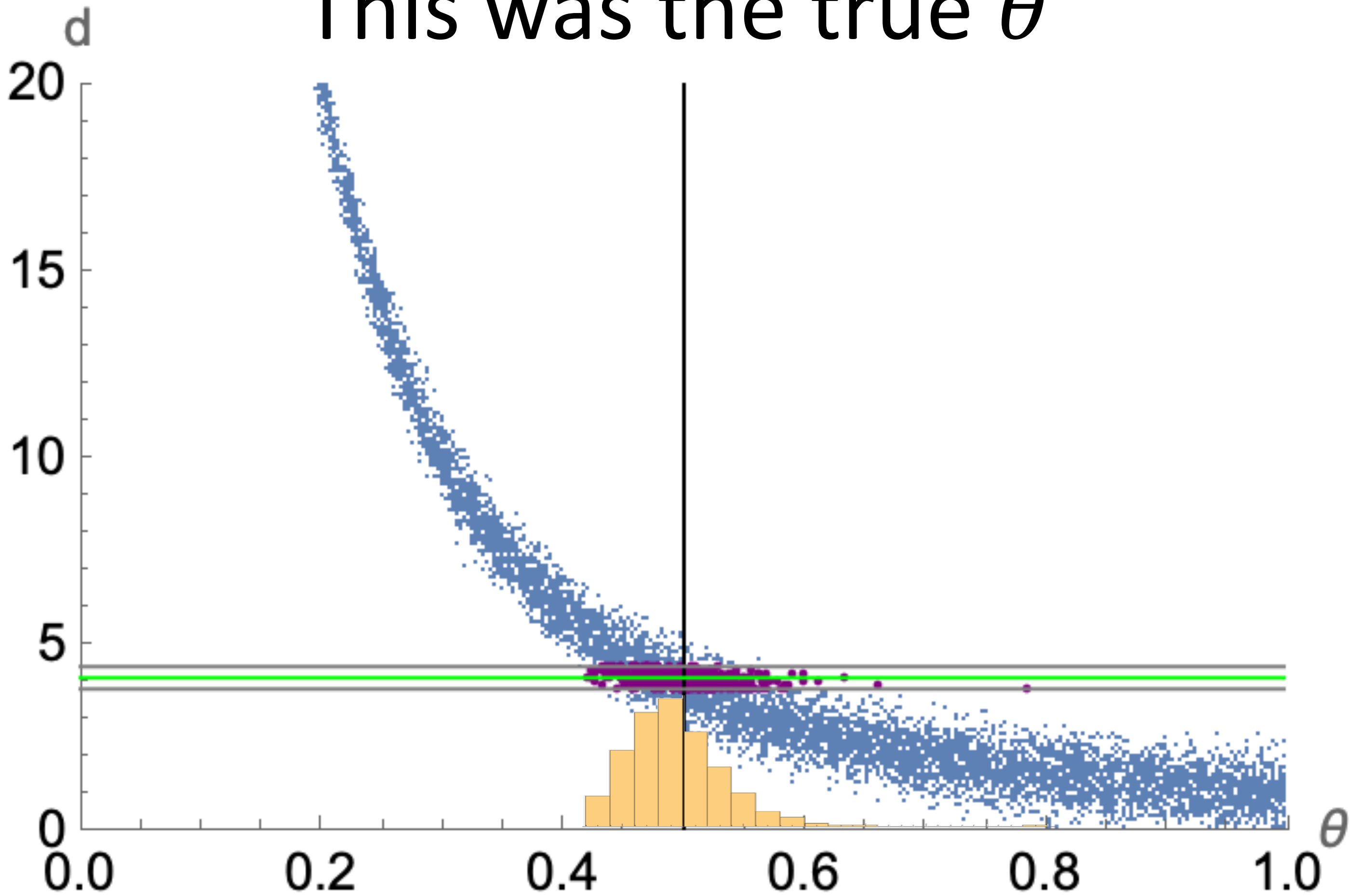
Slice through $\mathrm{p}(\theta, d)$ at d
("condition on d")



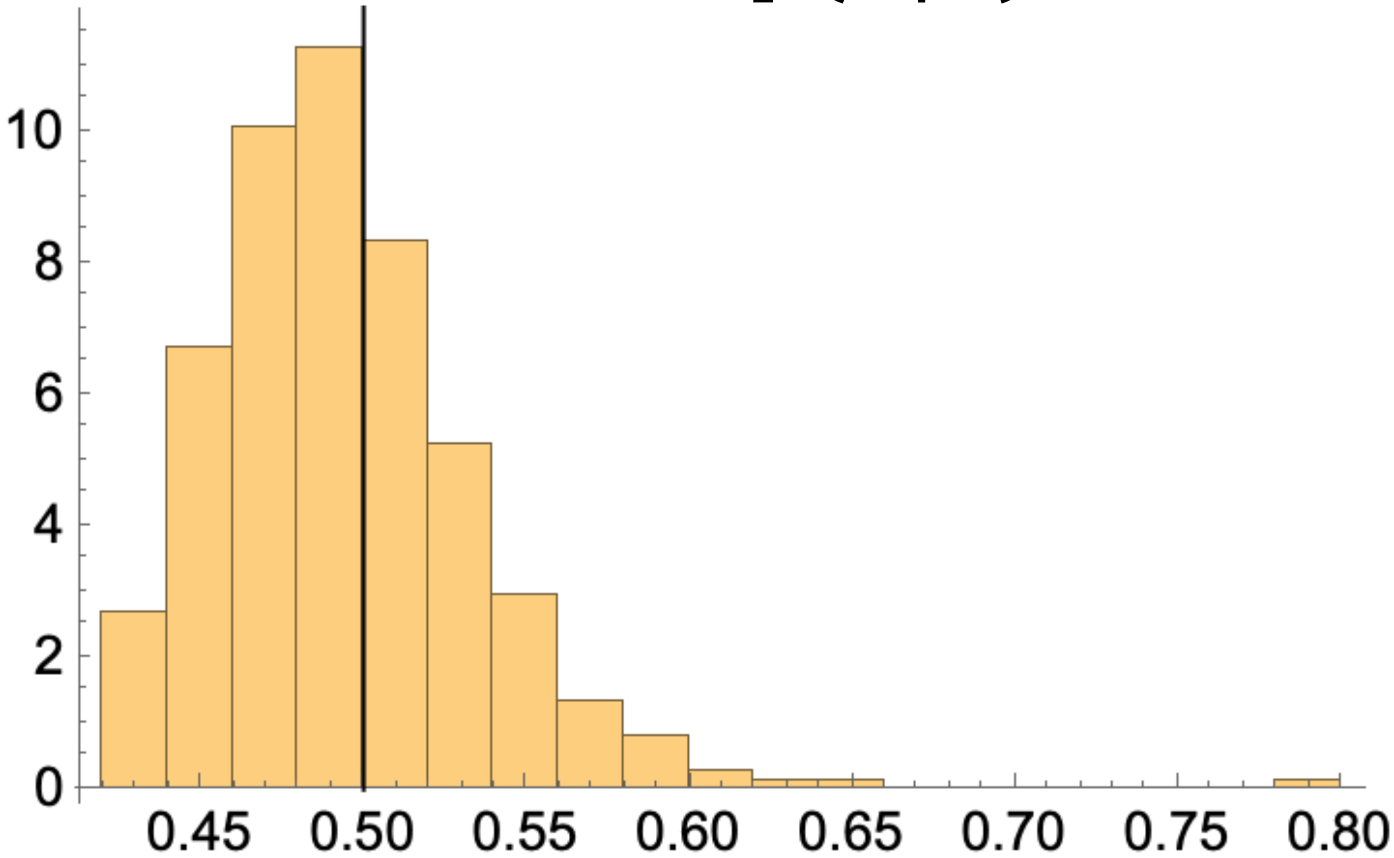
Slice through $\mathrm{p}(\theta, d)$ at d
("condition on d")



This was the true θ



Posterior $\mathrm{p}(\theta|d)$



This is ***Implicit*** **Inference**

- When likelihood and/or prior are not *explicitly* specified but *implicit* in…
  - simulations, generative models, labelled data.

- Various forms known as
  - Likelihood-free inference
  - Simulation-based inference
  - Approximate Bayesian Computation (ABC)

Machine learning takes us the rest of the way

- ***Recast inference problems as optimization problems***.

- Write down a loss that defines the problem
  - Parameterize the solution using a neural network
  - Minimize
  - Validate

First example: variational Bayes

- Define a parameterized family of distributions
- Minimize Kullback-Leibler loss between neural family and true likelihood

When using a neural density estimator this is DELFI, a (now) classic example of simulation-based inference.

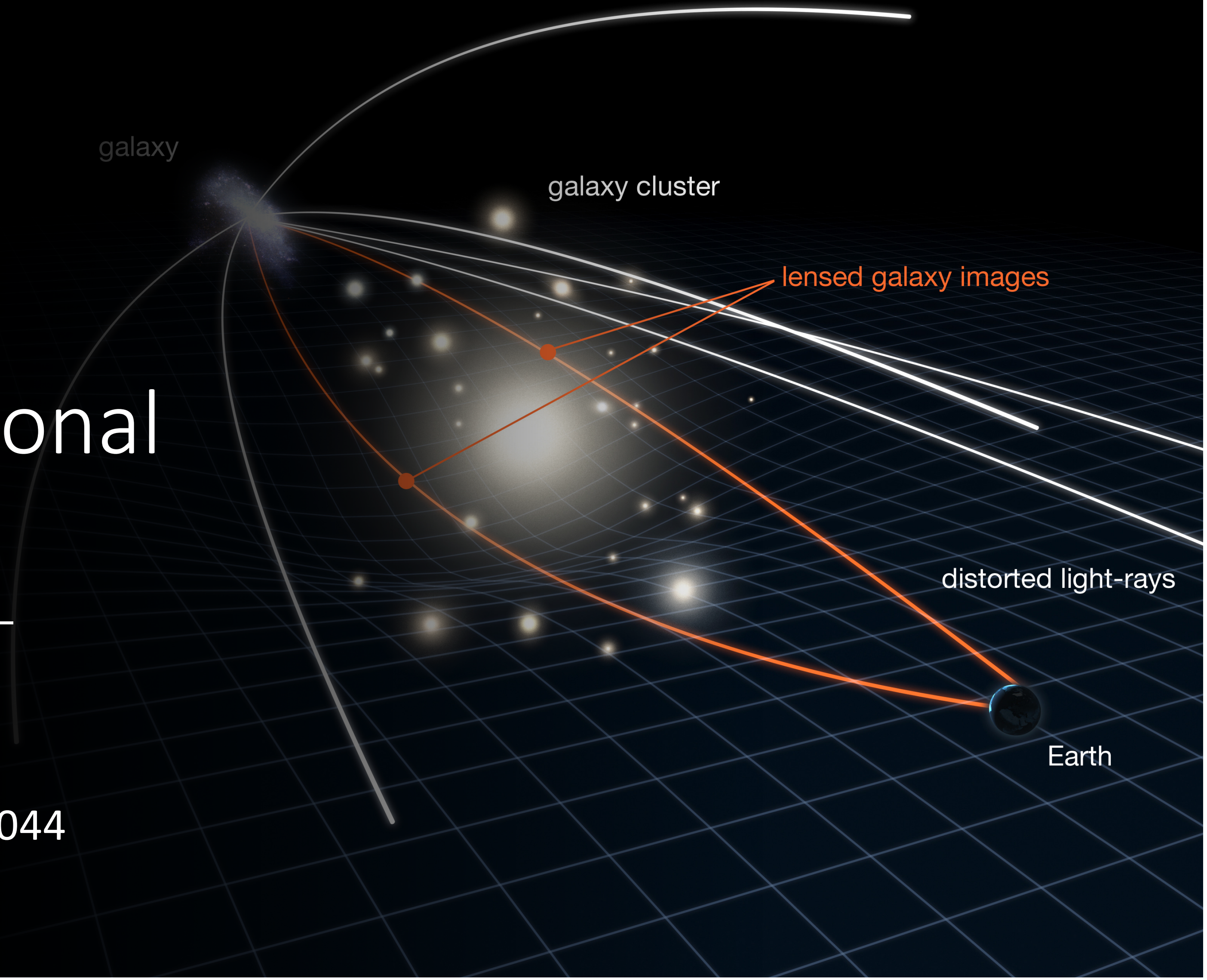
$$D_{\mathrm{KL}}(p^* \mid p) = \int p^*(\mathbf{t}|\boldsymbol{\theta}) \ln\left(\frac{p(\mathbf{t}|\boldsymbol{\theta};\mathbf{w})}{p^*(\mathbf{t}|\boldsymbol{\theta})}\right) d\mathbf{t}$$

$$-\ln U(\mathbf{w}|\{\boldsymbol{\theta},\mathbf{t}\}) = -\sum_{i=1}^{N_{\mathrm{samples}}} \ln p(\mathbf{t}_i|\boldsymbol{\theta}_i;\mathbf{w})$$

Papamakarios, Murray + coauthors, arXiv:1605.06376, 1705.07057, 1805.07226
Alsing, Feeney & Wandelt, arXiv: 1801.01497, 1903.01473

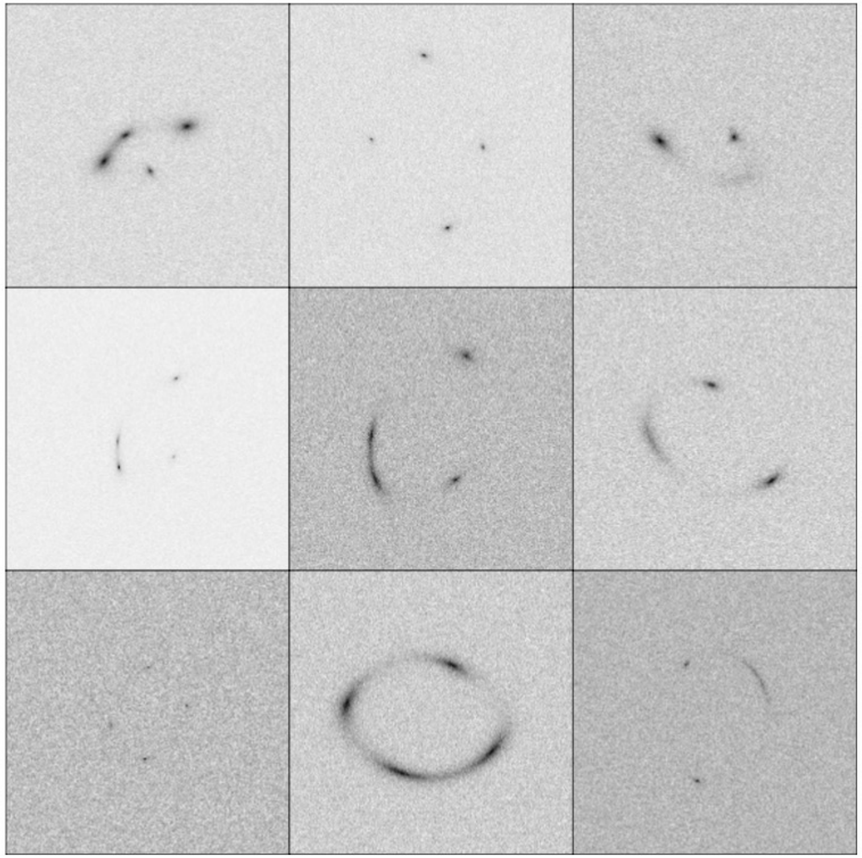
Example:

Strong Gravitational Lensing



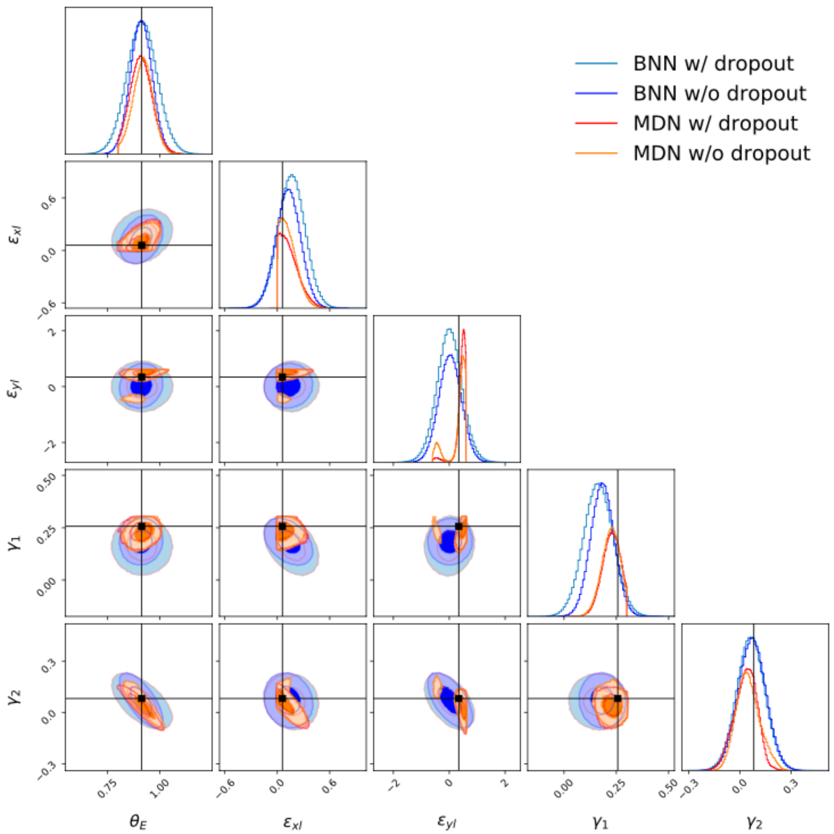
Legin et al. arXiv:2212.00044

Simulated images

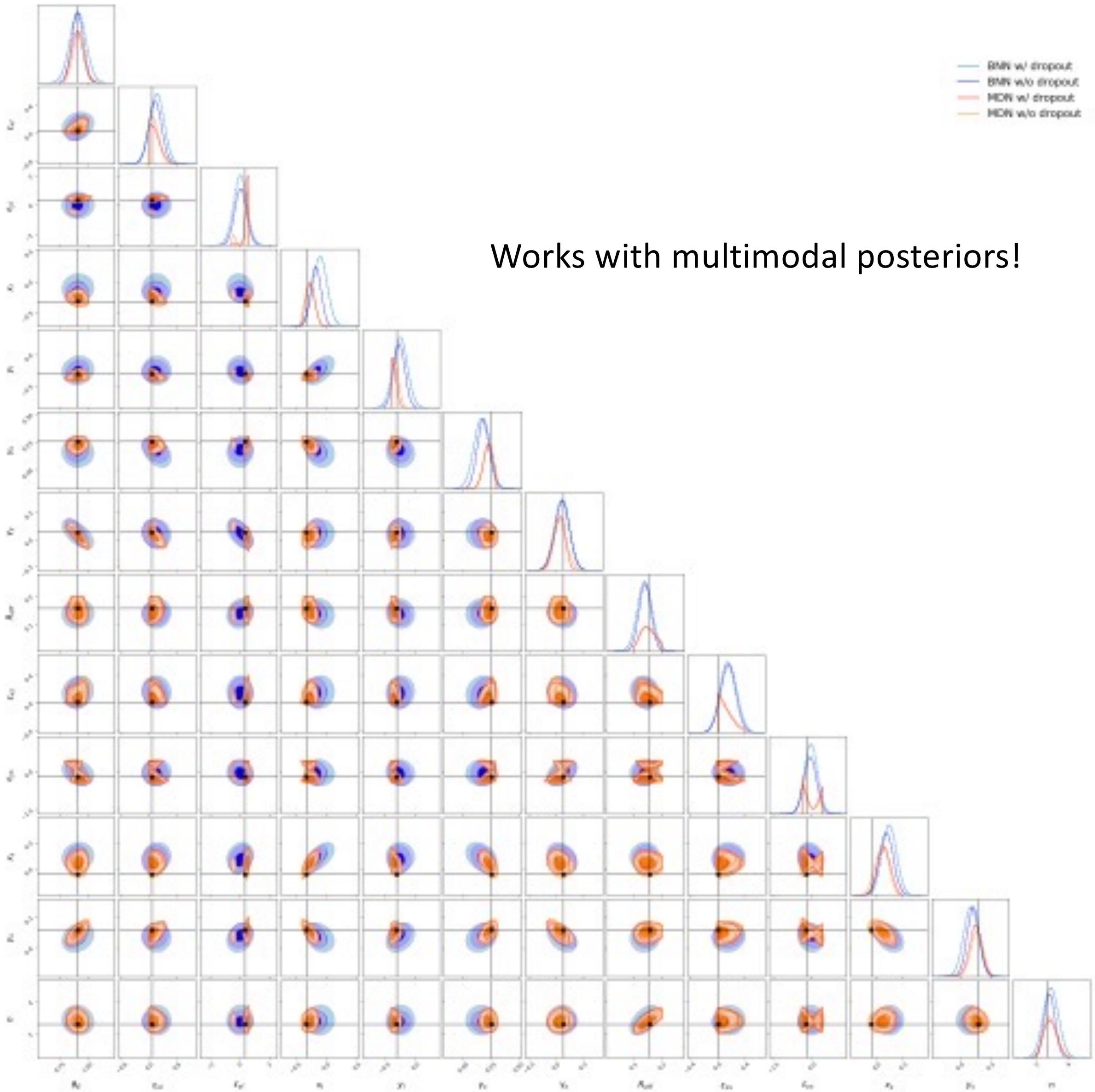


Legin et al arXiv 2212.00044

Inference

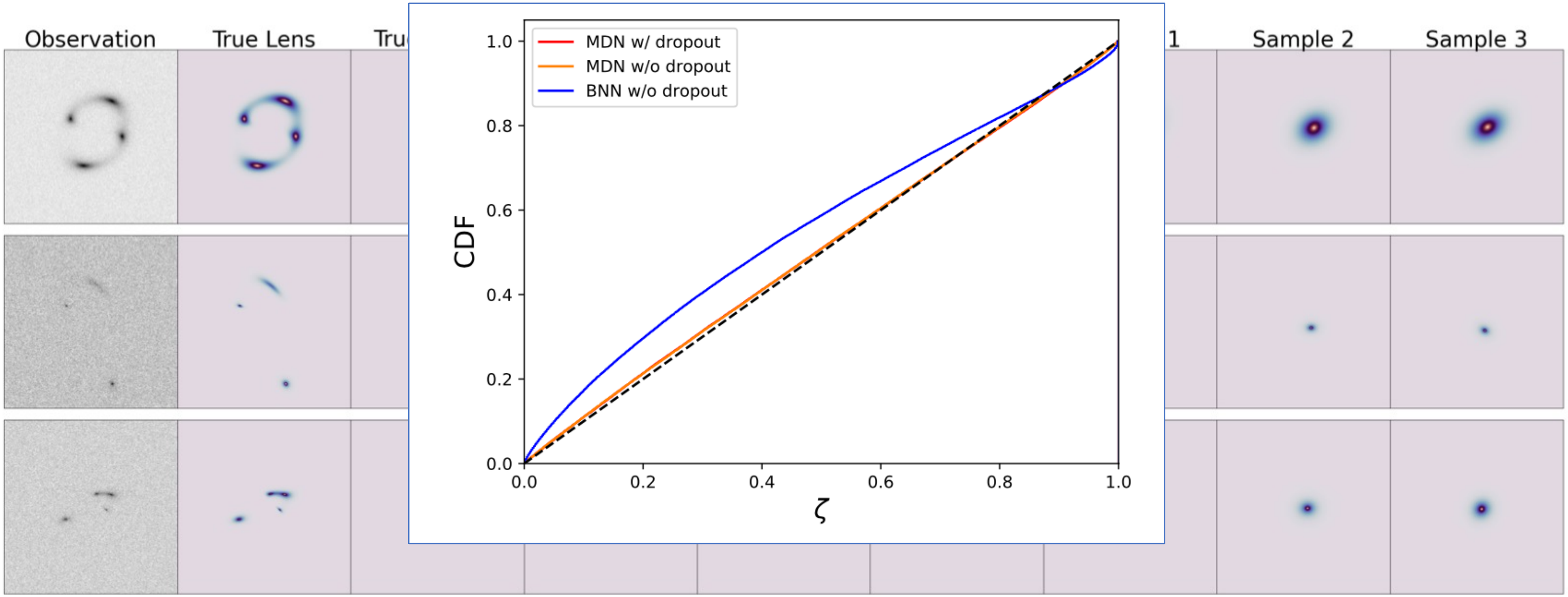


Legin et al arXiv 2212.00044



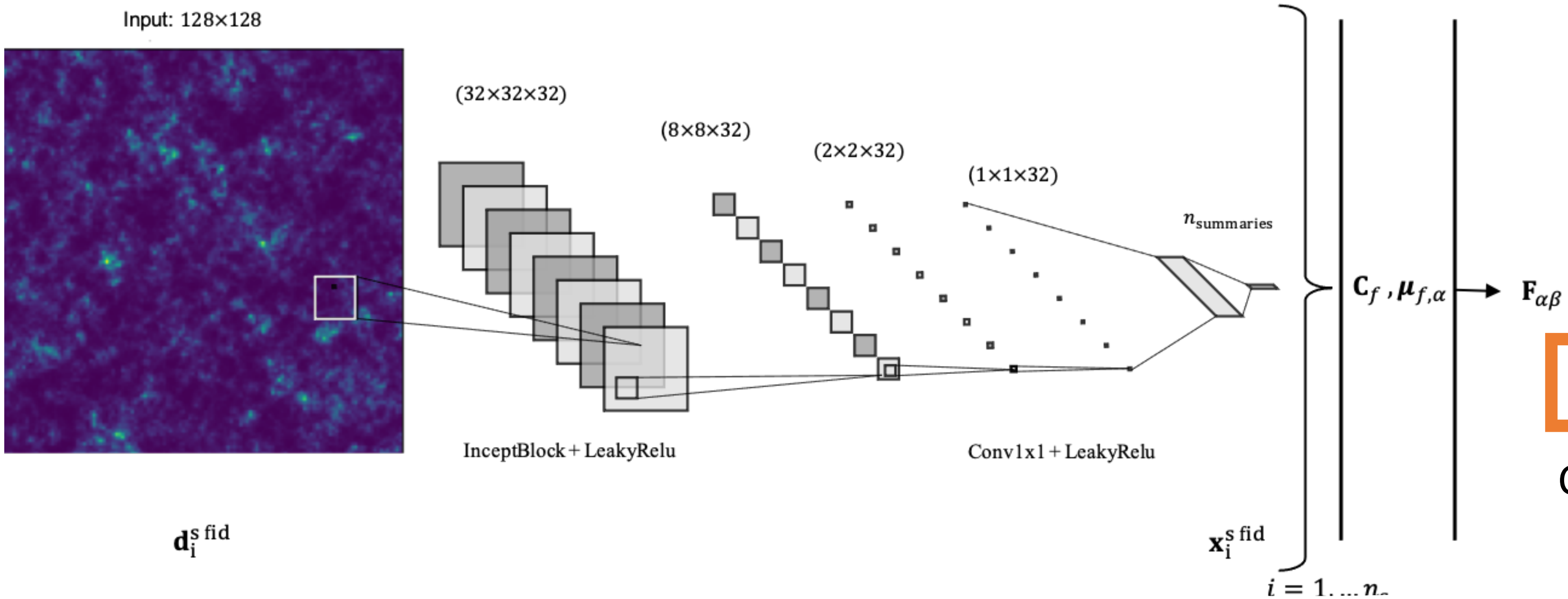
Works with multimodal posteriors!

Validation



Legin et al arXiv 2212.00044

Information maximizing neural networks: asymptotically optimal analysis, Fisher information, score computation *if you don't know the likelihood*

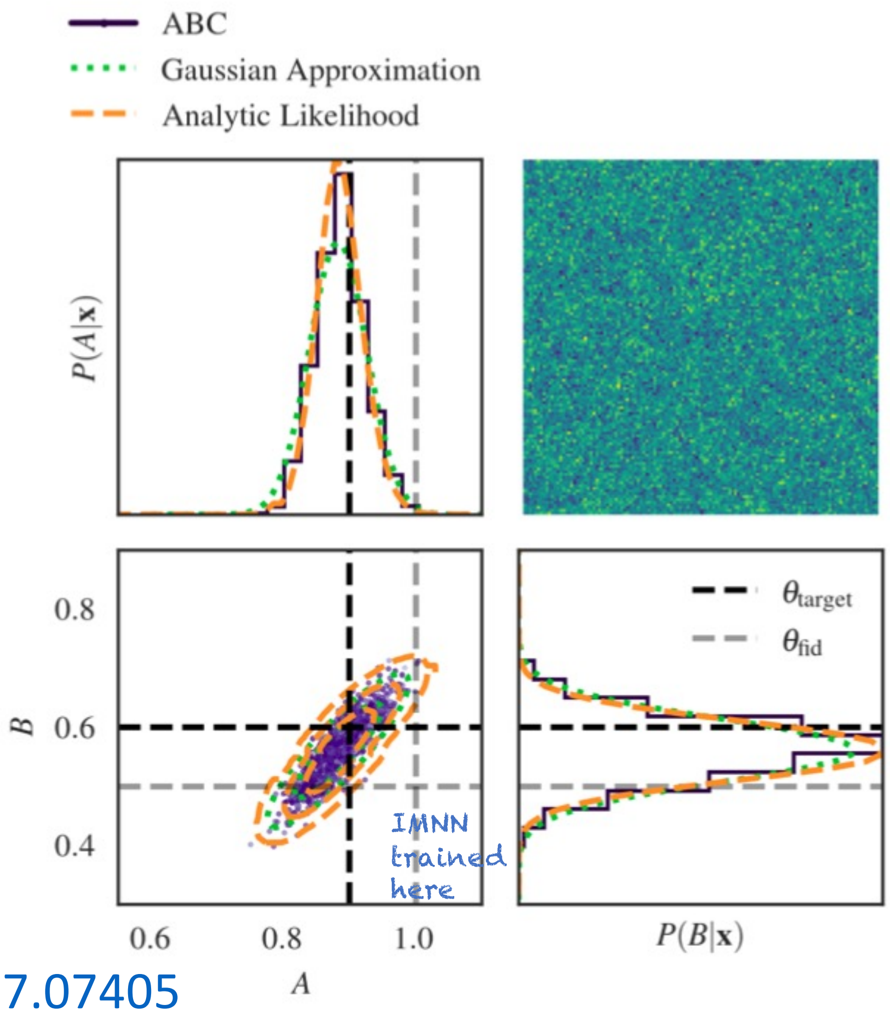


Optimal compression network minimizes information loss

$$\mathcal{L} = -\ln\det\langle F\rangle_{p(d|\theta_{\text{fid}})}$$

Charnock, Lavaux, Wandelt (arXiv:1802:03537)

IMNN recovers full info directly from the field



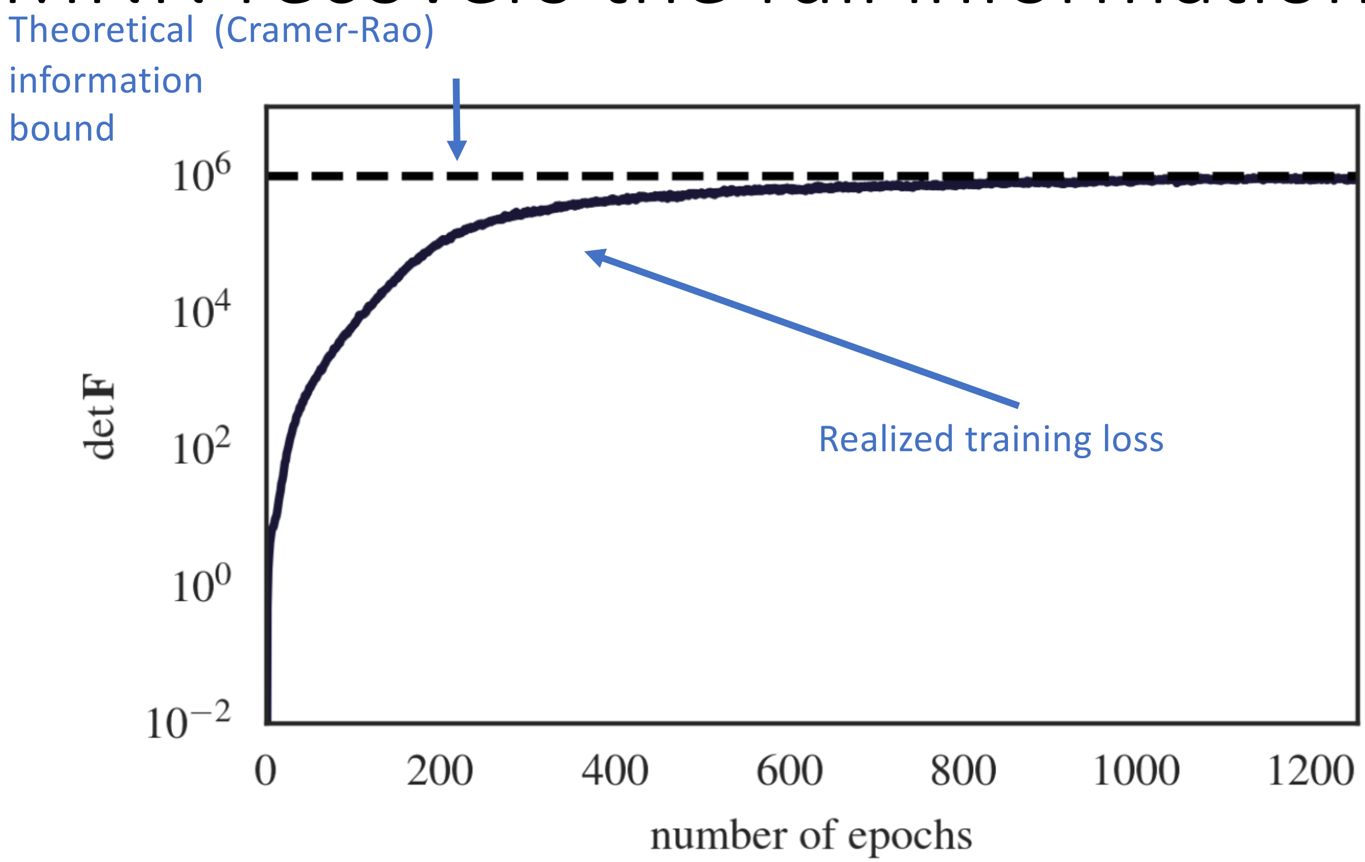
IMNN posterior

IMNN PML estimator

Exact likelihood

Makinen et al., arXiv:2107.07405

The IMNN recovers the full information



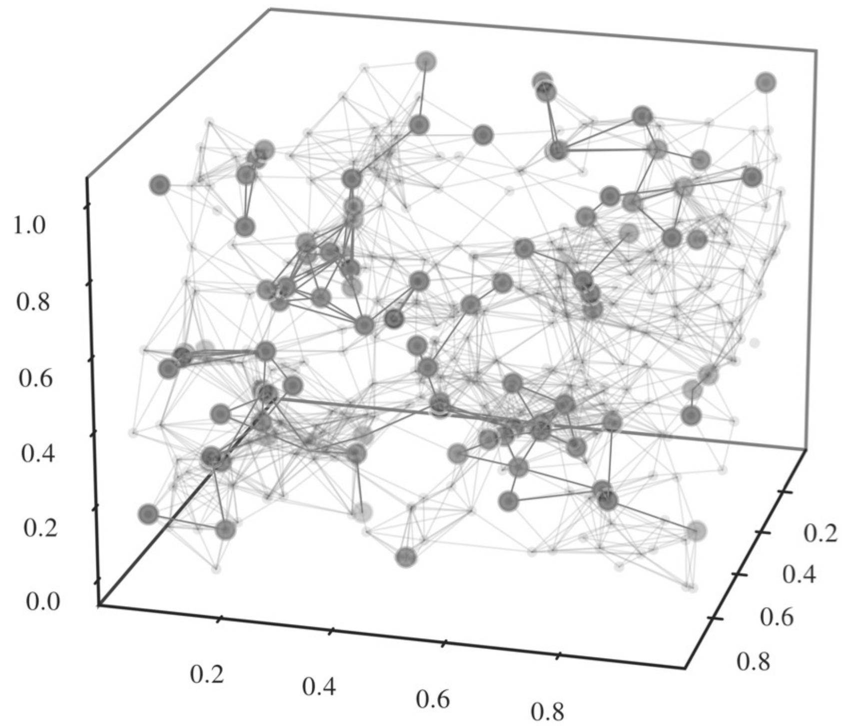
Makinen et al., arXiv:[2107.07405]

11 minutes on 1 GPU

Can define Fisher information and score on distributions of graphs

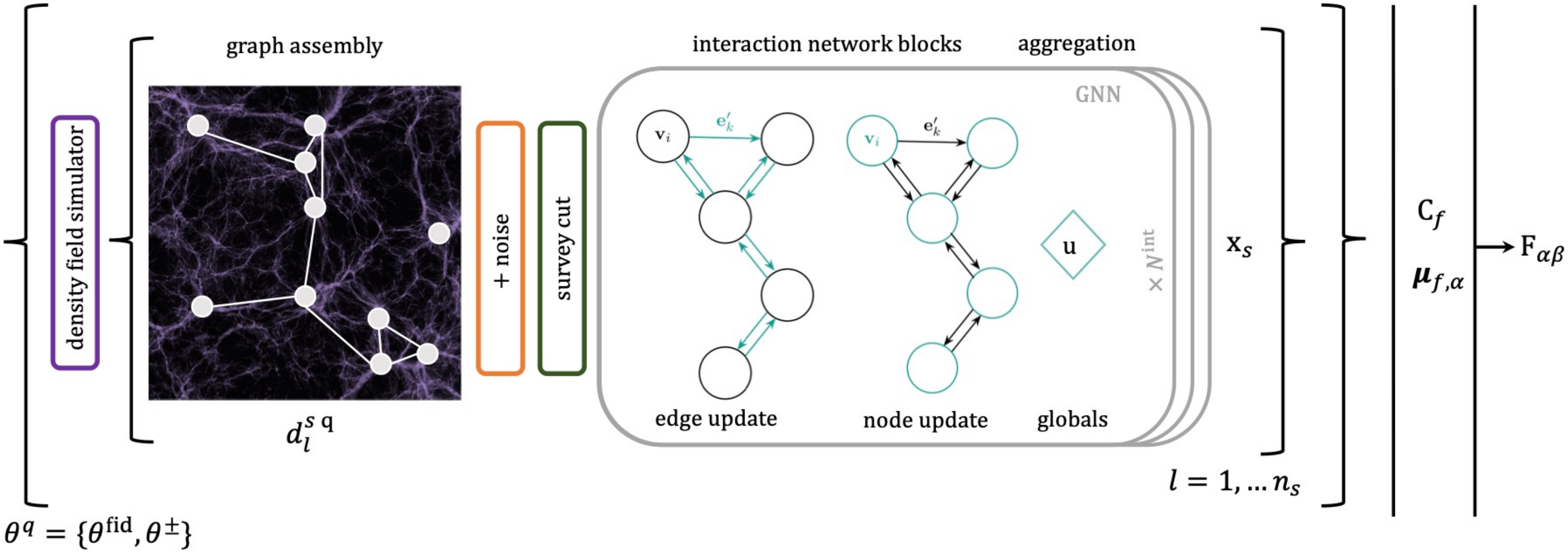
Example of using clusters of galaxies to infer cosmological parameters

Uses neurally derived Fisher score within pyDELFI.



Makinen et al. arXiv:2207.05202

Can define Fisher information and score on distributions of graphs



What if the number of parameters is large or simulations are scarce?

- General NDE becomes exponentially hard as number of dimensions increases.
- How do we handle high-dimensional problems?

- Simplify.

MOMENT AND POSTERIOR MARGINAL NETWORKS

Main idea: construct $\mathcal{F}(d)$,$\mathcal{G}(d)$ to go directly from data to posterior.

- **Moment networks**: obtain posterior moments directly from data by training NNs to solve

$$\langle\theta\rangle_{p(\theta|d)} = \underset{\mathcal{F}(d)}{\arg\min} \int ||\theta - \mathcal{F}(d)||_2^2 \, p(d,\theta) ddd\theta$$

$$\mathrm{Var}[\theta]_{p(\theta|d)} = \underset{\mathcal{G}(d)}{\arg\min} \int || \, ||\theta - \langle\theta\rangle_{p(\theta|d)}||_2^2 - \mathcal{G}(d)||_2^2 \, p(d,\theta) ddd\theta$$

(Jeffrey & Wandelt arXiv:2011.05991, presented at NeurIPS 2020)

Moment Network Example

Cosmology and astrophysics from full hydrodynamical simulations including black holes, star formation,…

CAMELS

Cosmology and Astrophysics with Machine Learning

Large suites of full, cosmological hydrosimulations as a function of cosmological parameters and astrophysics models with multiple codes (AREPO/Illustris, GIZMO/SIMBA, Astrid,…).

F. Villaescusa-Navarro, S. Genel, D. Angles-Alcazar et al. arXiv:2109.10915
F. Villaescusa-Navarro, D. Angles-Alcazar, S. Genel et al. arXiv:2010.00619

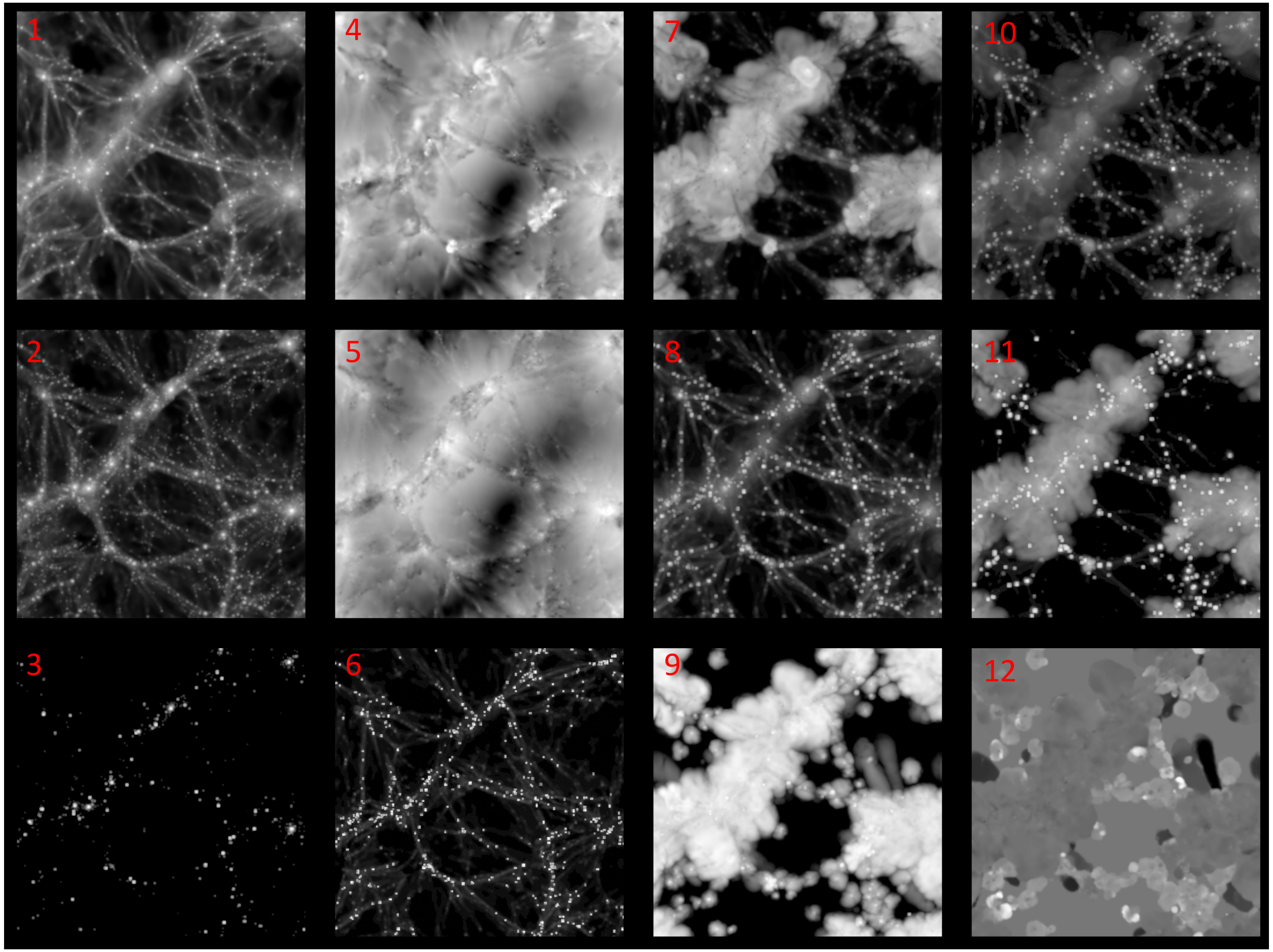
Cosmology on small scales with baryons

15 different 2-dimensional fields:

1. Gas mass
2. Dark matter mass
3. Stellar mass
4. Gas velocity
5. Dark matter velocity
6. Neutral hydrogen mass
7. Gas temperature
8. Electron density
9. Gas metallicity
10. Gas pressure
11. Magnetic fields
12. Mg/Fe
13. Total mass
14. N-body
15. All fields except dark matter

15,000 images per field from 1,000 CAMELS-IllustrisTNG simulations.
Each image:

- 250x250 pixels
- 25x25 $(Mpc/h)^2$
- 100 kpc/h resolution



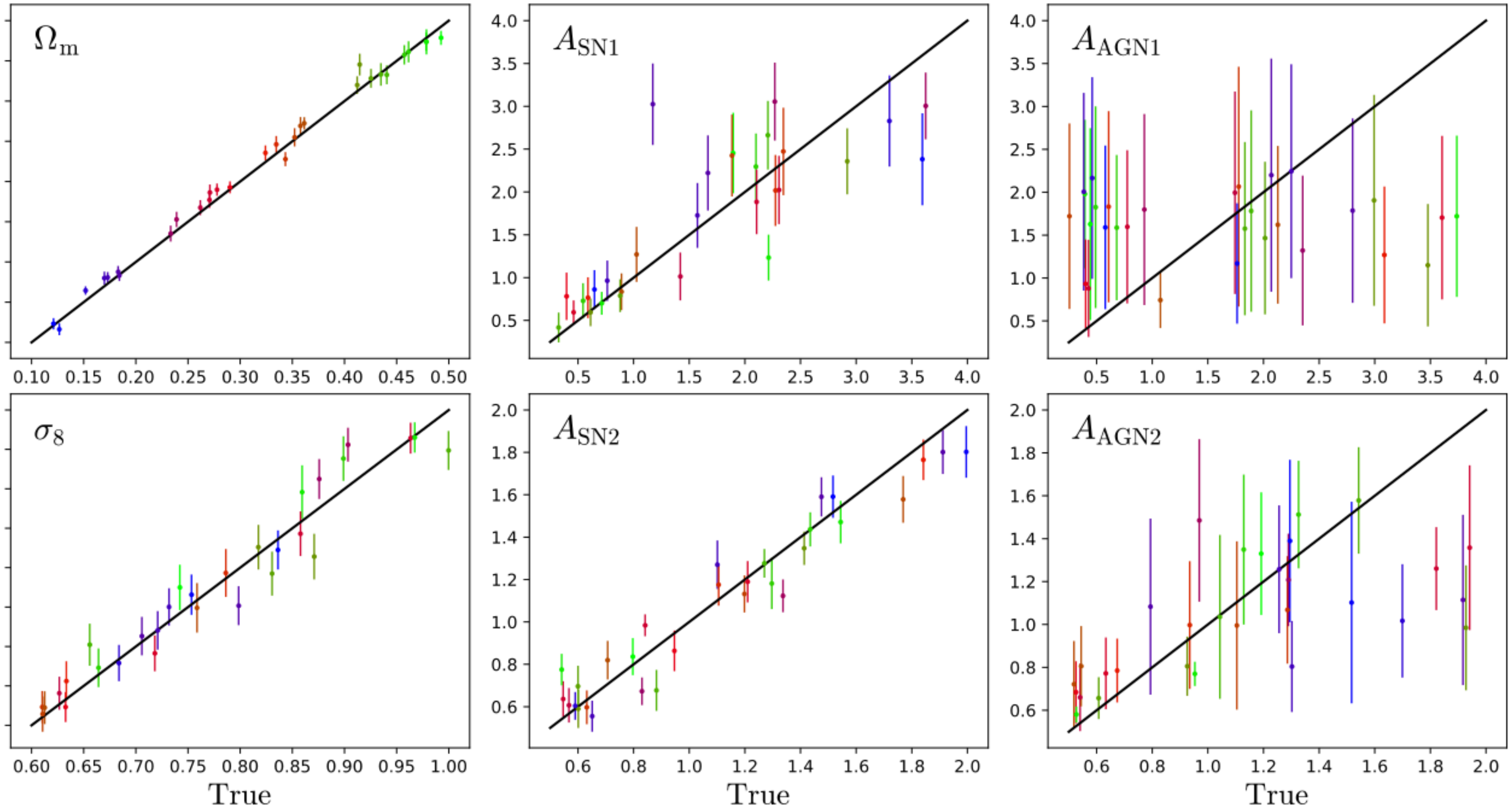
SBI: COSMOLOGY FROM SMALL-SCALE HYDRO

Villaescusa-Navarro, arXiv:2109.09747

Computing posterior means & variances from **gas temperature**

$$\mathcal{L} = \sum_{i=1}^{6} \log\left(\sum_{j\in\mathrm{batch}} (\theta_{i,j} - \mu_{i,j})^2\right) + \sum_{i=1}^{6} \log\left(\sum_{j\in\mathrm{batch}} \left((\theta_{i,j} - \mu_{i,j})^2 - \sigma_{i,j}^2\right)^2\right)$$

Posterior means & variances computed by **moment network** minimizing $\mathcal{L}$



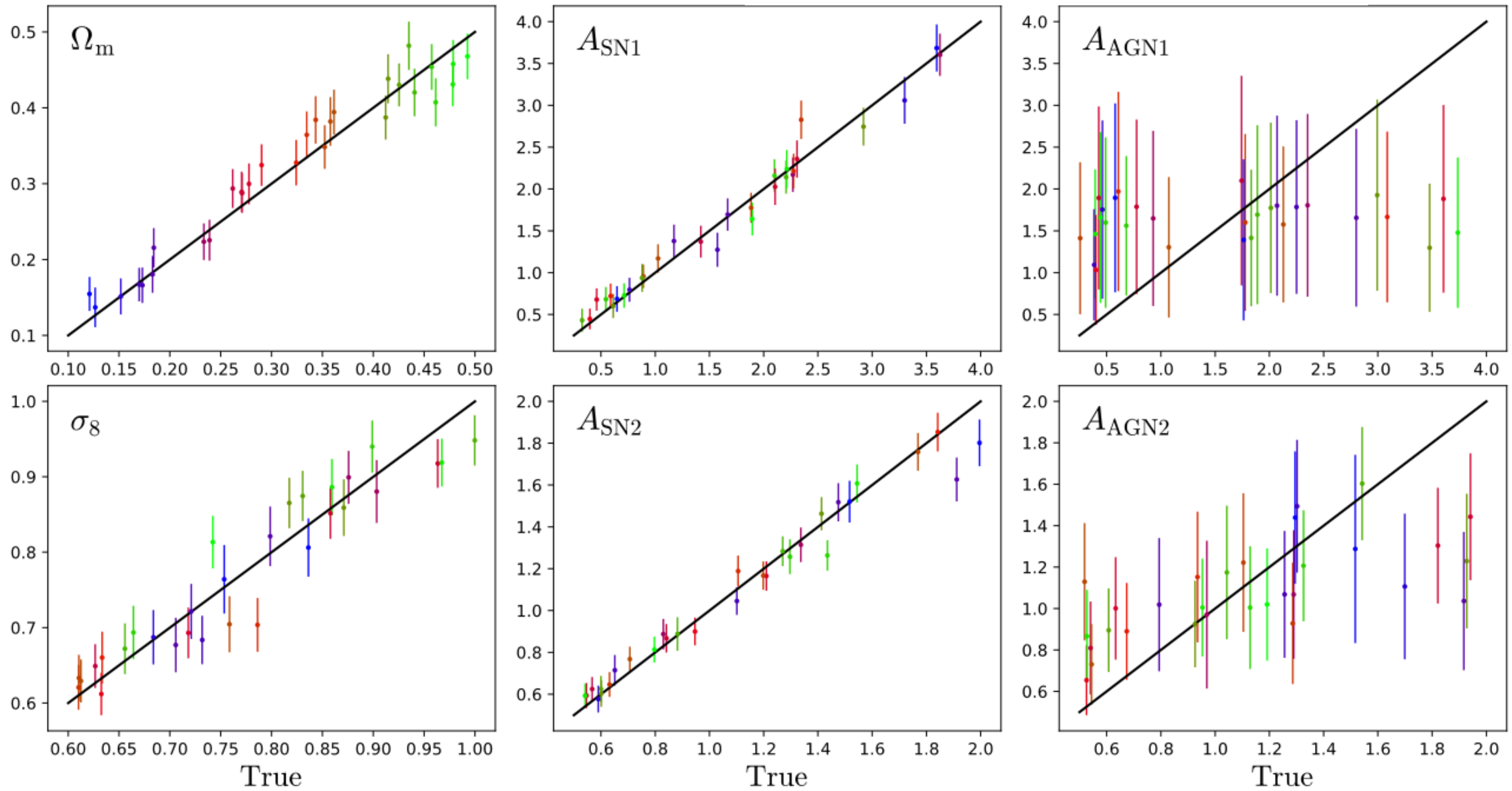
SBI: Cosmology from Small-Scale Hydro

Villaescusa-Navarro, arXiv:2109.09747

Computing posterior means & variances from **gas metallicity**

$$\mathcal{L} = \sum_{i=1}^{6} \log\left(\sum_{j\in\text{batch}} (\theta_{i,j} - \mu_{i,j})^2\right) + \sum_{i=1}^{6} \log\left(\sum_{j\in\text{batch}} \left((\theta_{i,j} - \mu_{i,j})^2 - \sigma_{i,j}^2\right)^2\right)$$

Posterior means & variances computed by **moment network** minimizing $\mathcal{L}$



What the cosmological AI tells us about the CAMELS Multifield Data set

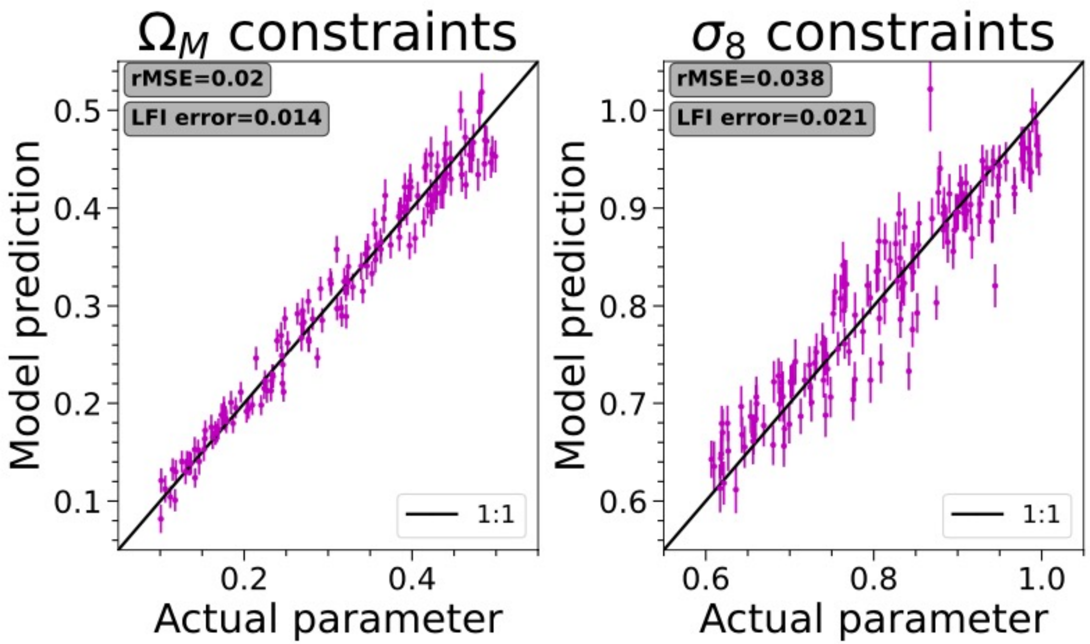
1. There is cosmological information on very small scales (100 kpc)

2. The hydro outputs contain *more* information than the dark matter density

3. For *total matter,* inferences are *robust* to baryonic physics (good news for weak lensing!)

Villaescusa-Navarro et al., arXiv:2109.09747, arXiv:2109.10360

Implicit Likelihood Inference with moment networks from surveys generated with **semi-analytic galaxy formation models**



Moment networks trained on SAMs run on 1000 DM sims (100 h^{-1} Mpc)^3 *stellar mass selected sample*

L. Perez et al. 2204.02408

What about inference about models?

Can we compute the evidence ratio if we don't know the likelihood?

$$\frac{p(M_i|d)}{p(M_j|d)} = \boxed{\frac{p(d|M_i)}{p(d|M_j)}} \frac{p(M_i)}{p(M_j)}$$

Bayes factor

Bayesian model comparison

Even if likelihood and posterior are explicitly given

- Likelihood can be costly to evaluate
- **Evidence can be hard to compute**

$$P(\boldsymbol{\theta}|\mathrm{d}) = \frac{P(\mathrm{d}|\boldsymbol{\theta})P(\boldsymbol{\theta})}{\boxed{P(\mathrm{d})}}$$

Jeffrey & Wandelt, in prep

Evidence Networks: example loss construction

Loss for model label *m*=0 or *m*=1:

$$\mathscr{V}(f(x), m) = m e^{-\frac{1}{2} f(x)} + (1 - m) e^{\frac{1}{2} f(x)} = e^{(\frac{1}{2} - m) f(x)}$$

Minimize over the prior drawn data:

$$I[f] = \int e^{-\frac{1}{2} f(x)} \, p(x, M_1) \; + e^{\frac{1}{2} f(x)} \, p(x, M_0) \, \mathrm{d}x$$

Optimised network *f* gives Bayes factor:

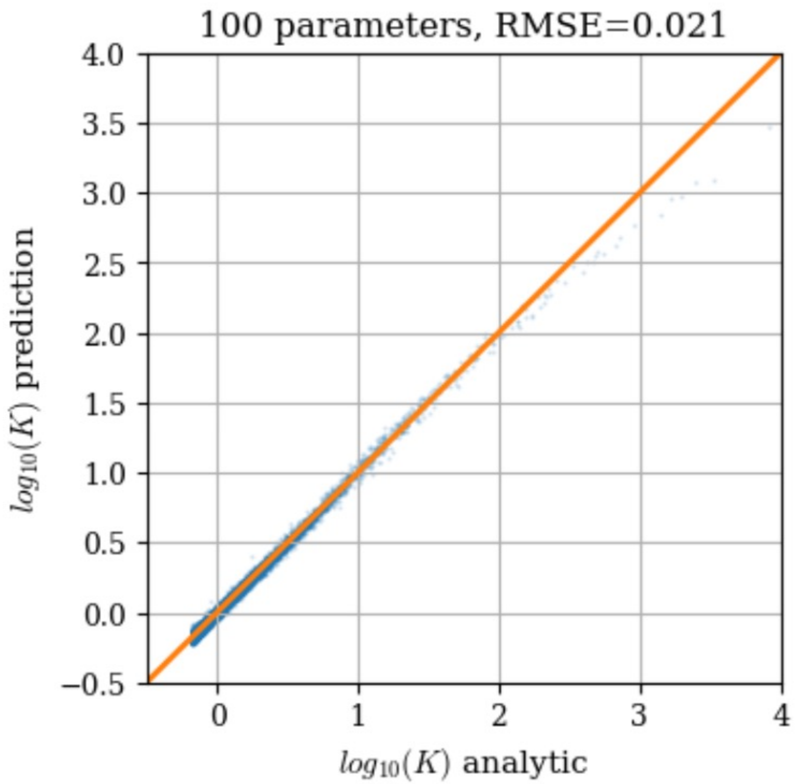
$$f_0(x) = \log\left(K \, \frac{p(M_1)}{p(M_0)}\right)$$

Jeffrey & Wandelt, in prep

Example: evidence ratio with 100 parameters

Evidence Networks

$$f_0(x) = \log\left(K \ \frac{p(M_1)}{p(M_0)}\right)$$



This evidence computation does not explicitly depend on number of parameters!

Jeffrey & Wandelt, in prep

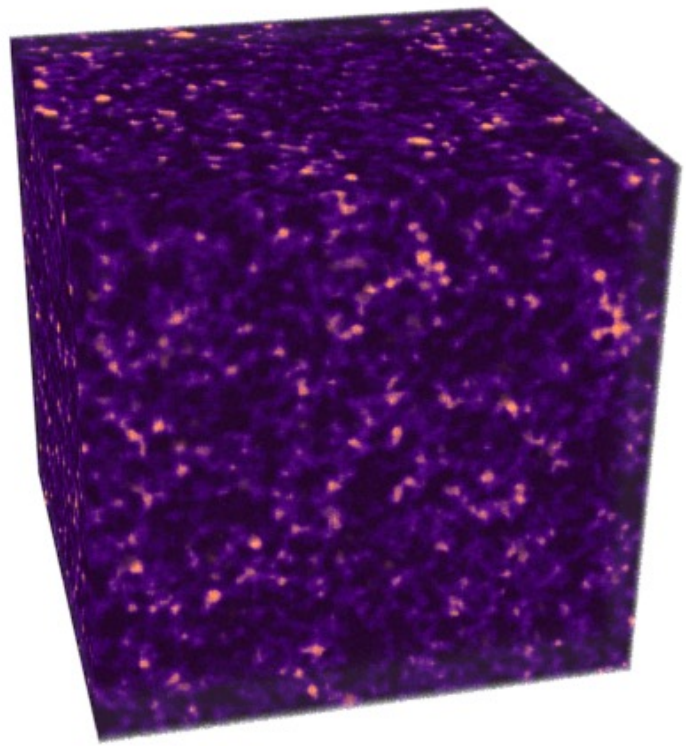
Example: cosmological initial conditions

We can build a *non-linear* time machine and
sample possible initial states
that could have given rise to our universe.

Ronan Legin et al, in prep

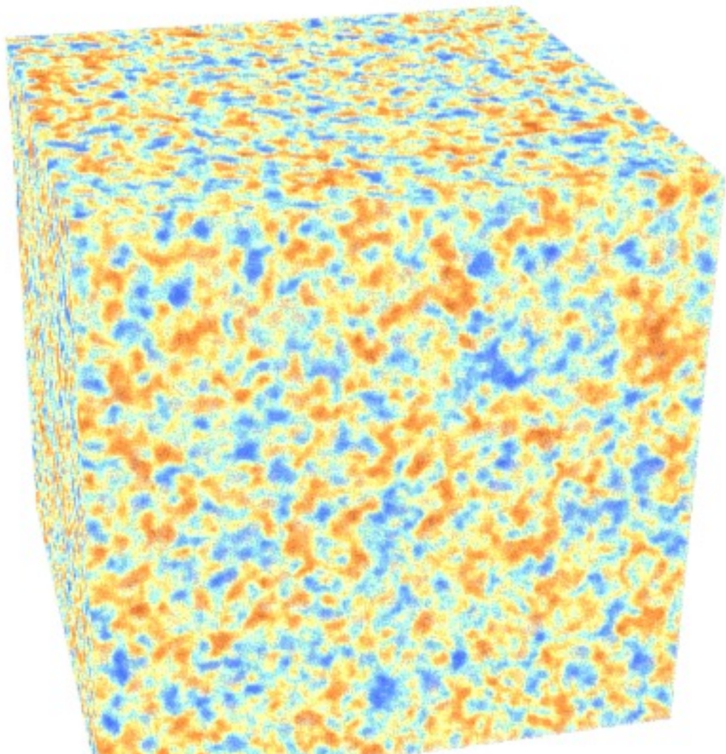
First full-field inference of initial conditions from fully non-linear density field

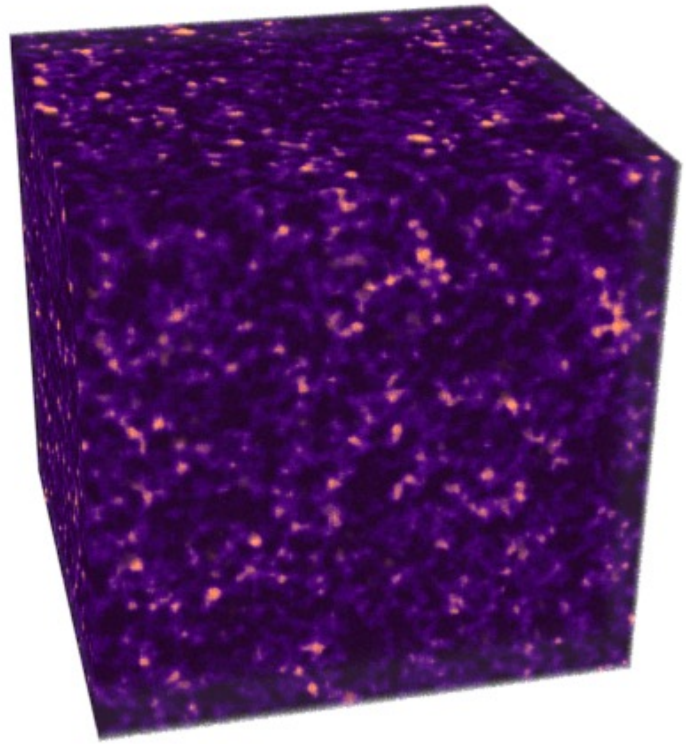
- 1 Gpc GADGET1024^3 simulation at z=0
- Binned on 128^3 grid



Ronan Legin et al, in prep

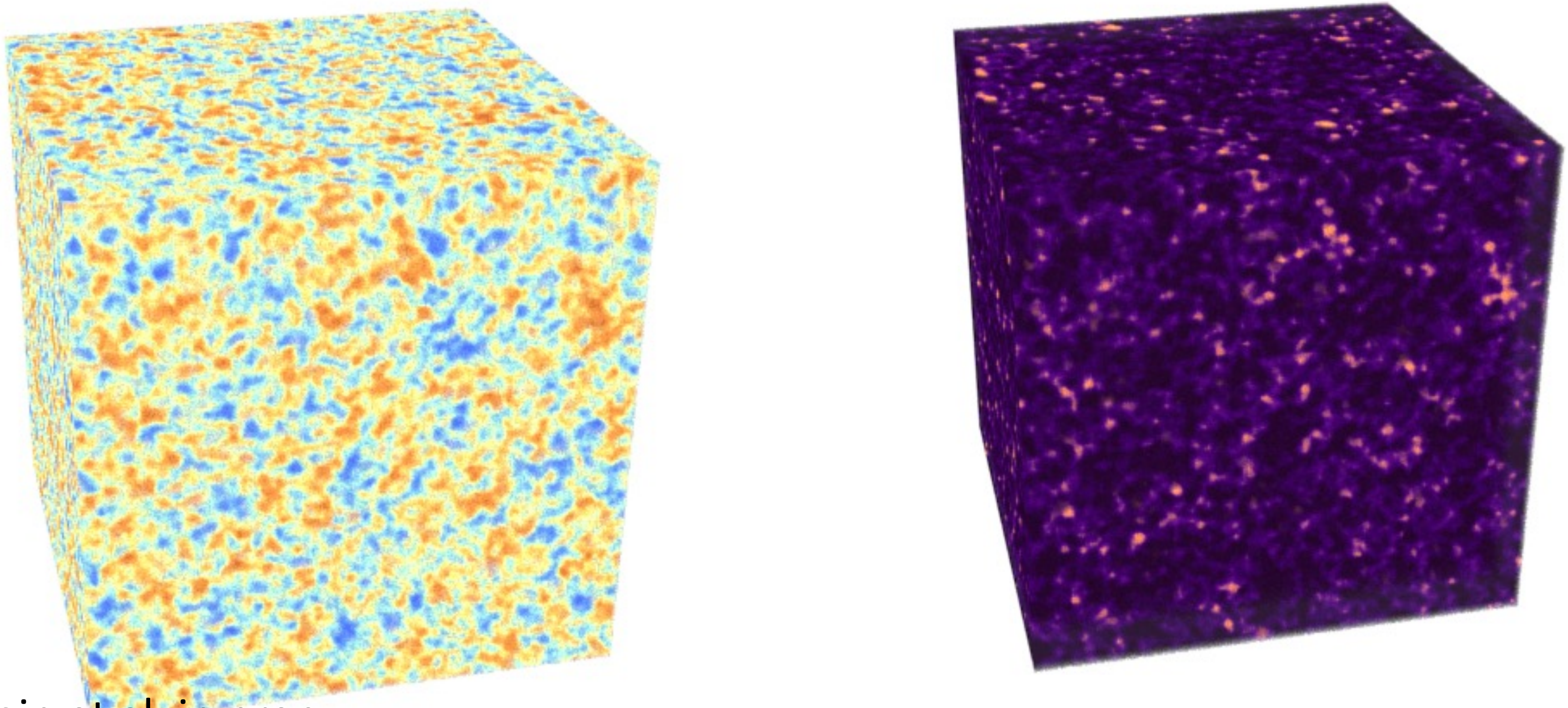
First full-field inference of initial conditions from fully non-linear density field





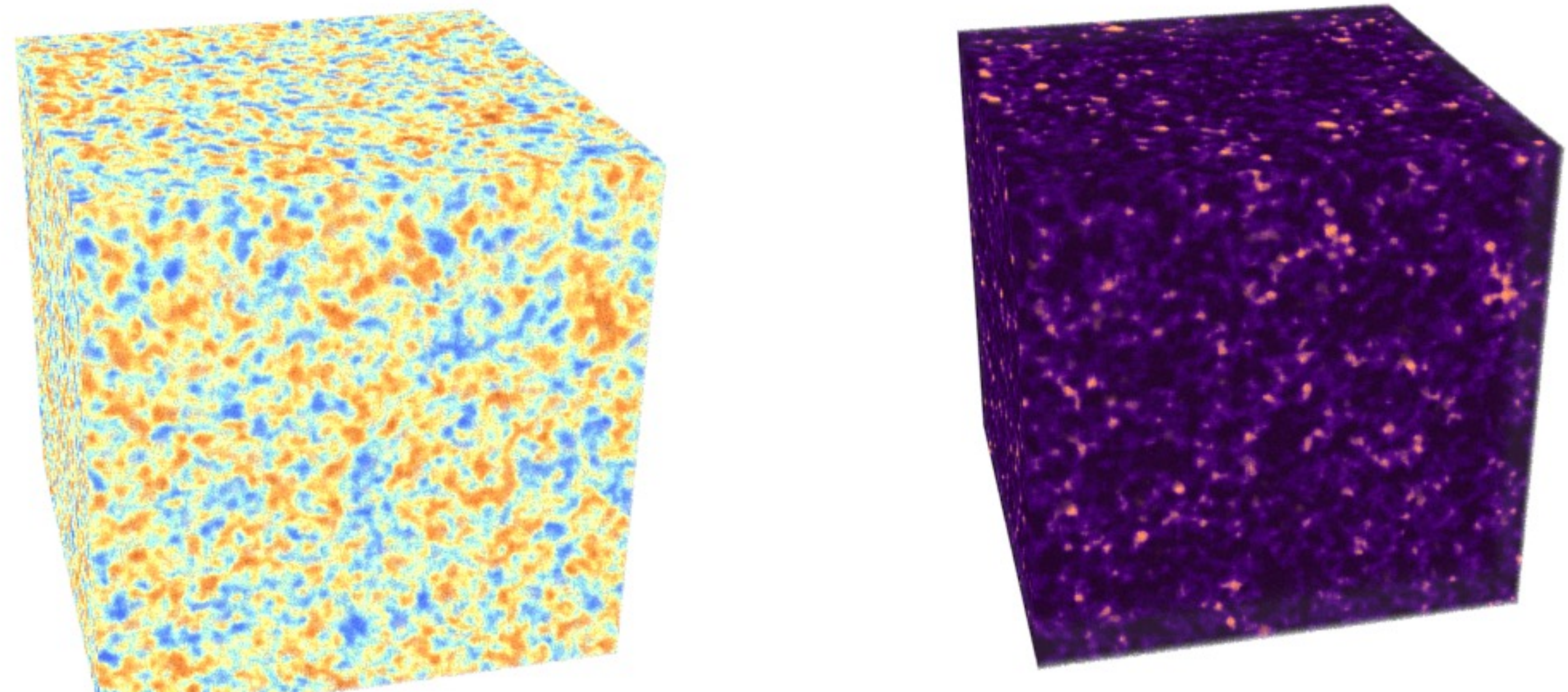
Ronan Legin et al, in prep

First full-field inference of initial conditions from fully non-linear density field



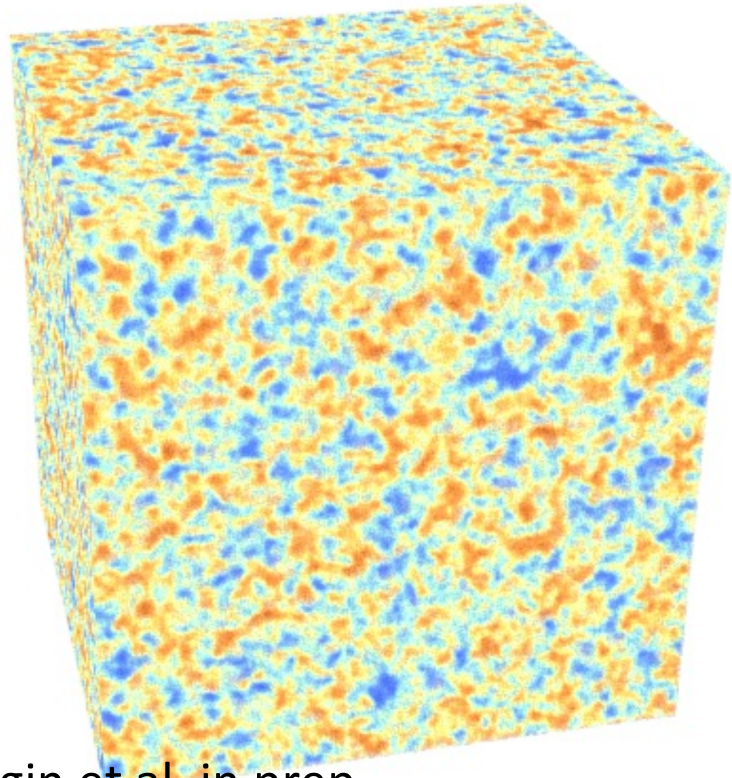
Ronan Legin et al, in prep

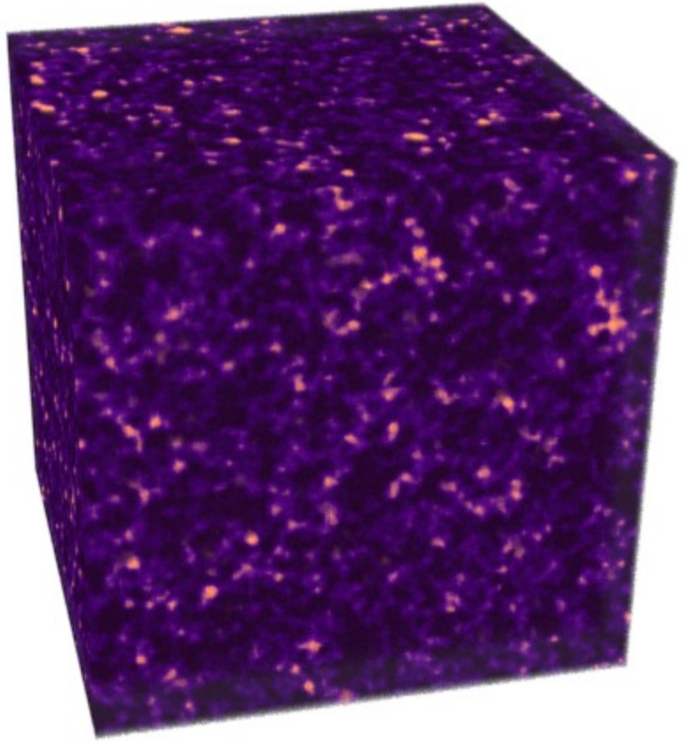
First full-field inference of initial conditions from fully non-linear density field



Ronan Legin et al, in prep

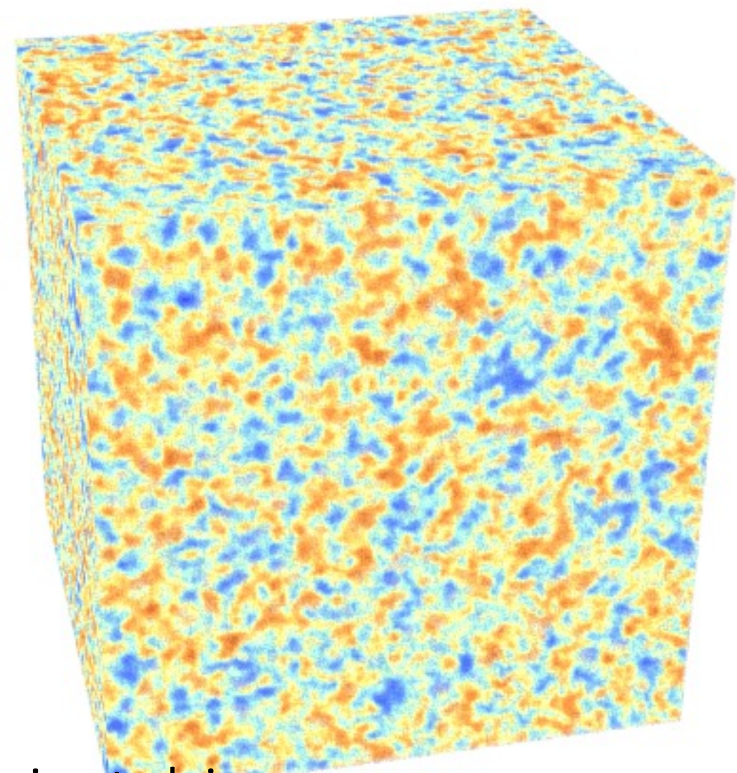
First full-field inference of initial conditions from fully non-linear density field

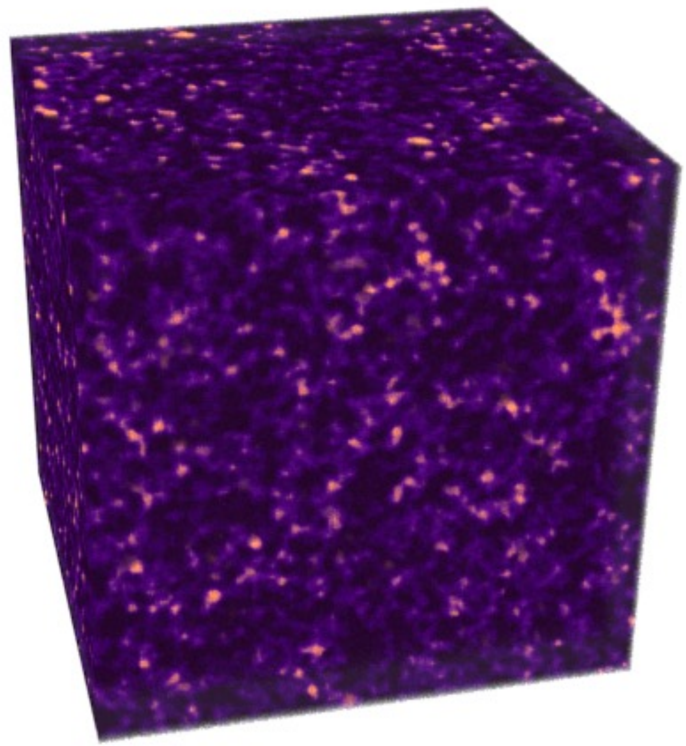




Ronan Legin et al, in prep

First full-field inference of initial conditions from fully non-linear density field





Ronan Legin et al, in prep

Conclusion

- Implicit inference gives full modeling freedom for parameter inference, Bayesian model comparison, and initial condition reconstruction
- We can measure the information content of quantities

- First 3D, fully non-linear initial condition reconstruction
- First marginalisation over cosmological parameters
- Ability to mask model shortcomings in a fine-grained way

Appendices

Proof that a trained IMNN computes the score of the unknown likelihood of the original data

$$\frac{\delta}{\delta t}\ln\det F = \frac{\delta}{\delta t}\ln \mu_{,\theta}\, C^{-1}\mu_{,\theta} = 0$$

$$\implies 2\mu_{,\theta}^{-1}\,\frac{\delta}{\delta t}\mu_{,\theta} - C^{-1}\frac{\delta}{\delta t}C =$$

$$2\mu_{,\theta}^{-1}\, p_{,\theta}\,(x|\theta) - 2C^{-1}(t-\mu)p(x|\theta) == 0$$

so

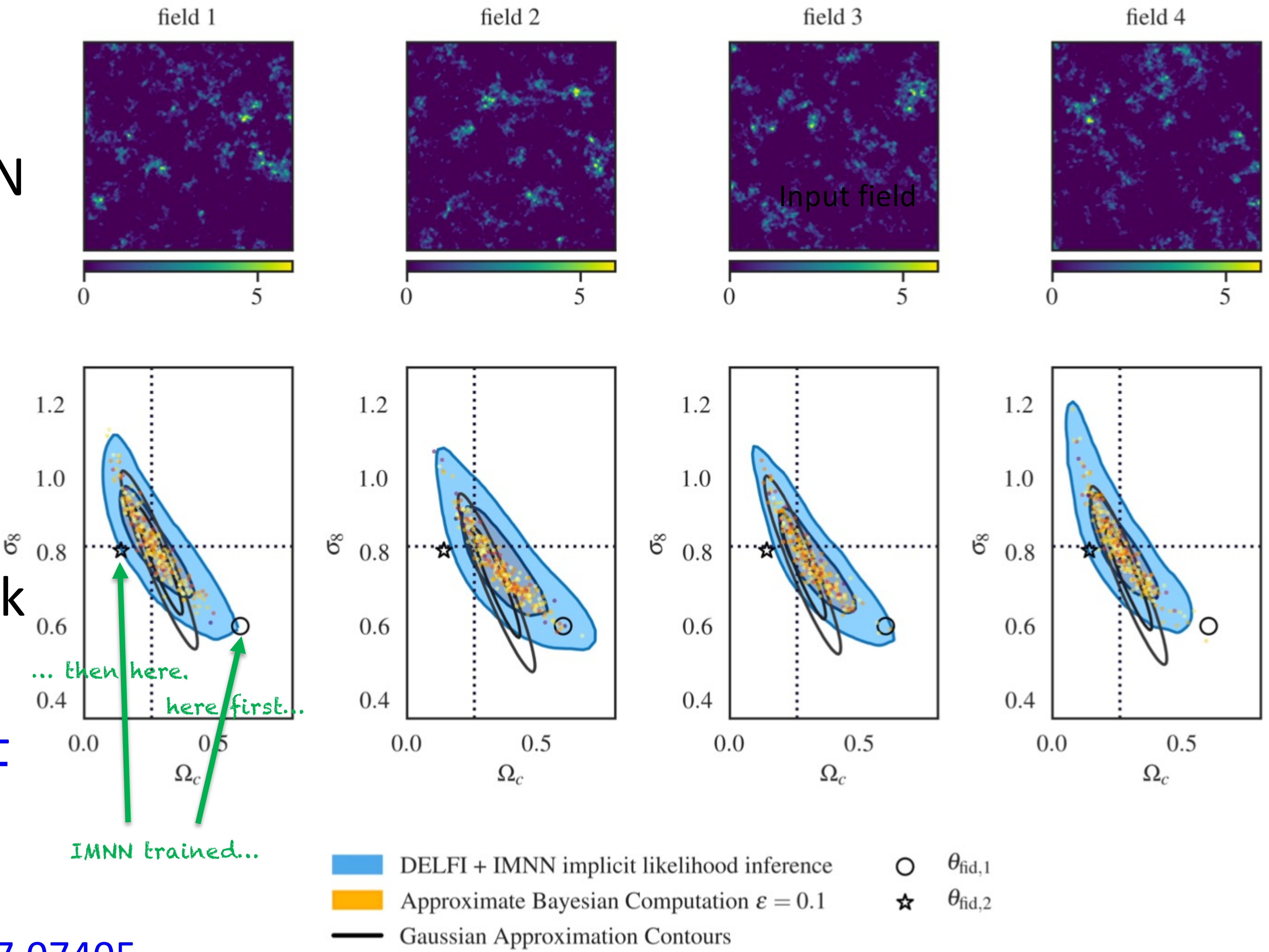
$$\mu_{,\theta}^{-1}\, p_{,\theta}\,(x|\theta) = C^{-1}(t-\mu)p(x|\theta)$$

Dividing by p we get

$$\partial_\theta \ln p(x|\theta) = \mu_{,\theta}\, C^{-1}(t-\mu).$$

This shows that the score function of the *unknown* original data likelihood is equal to the *Gaussian* score of the IMNN output.

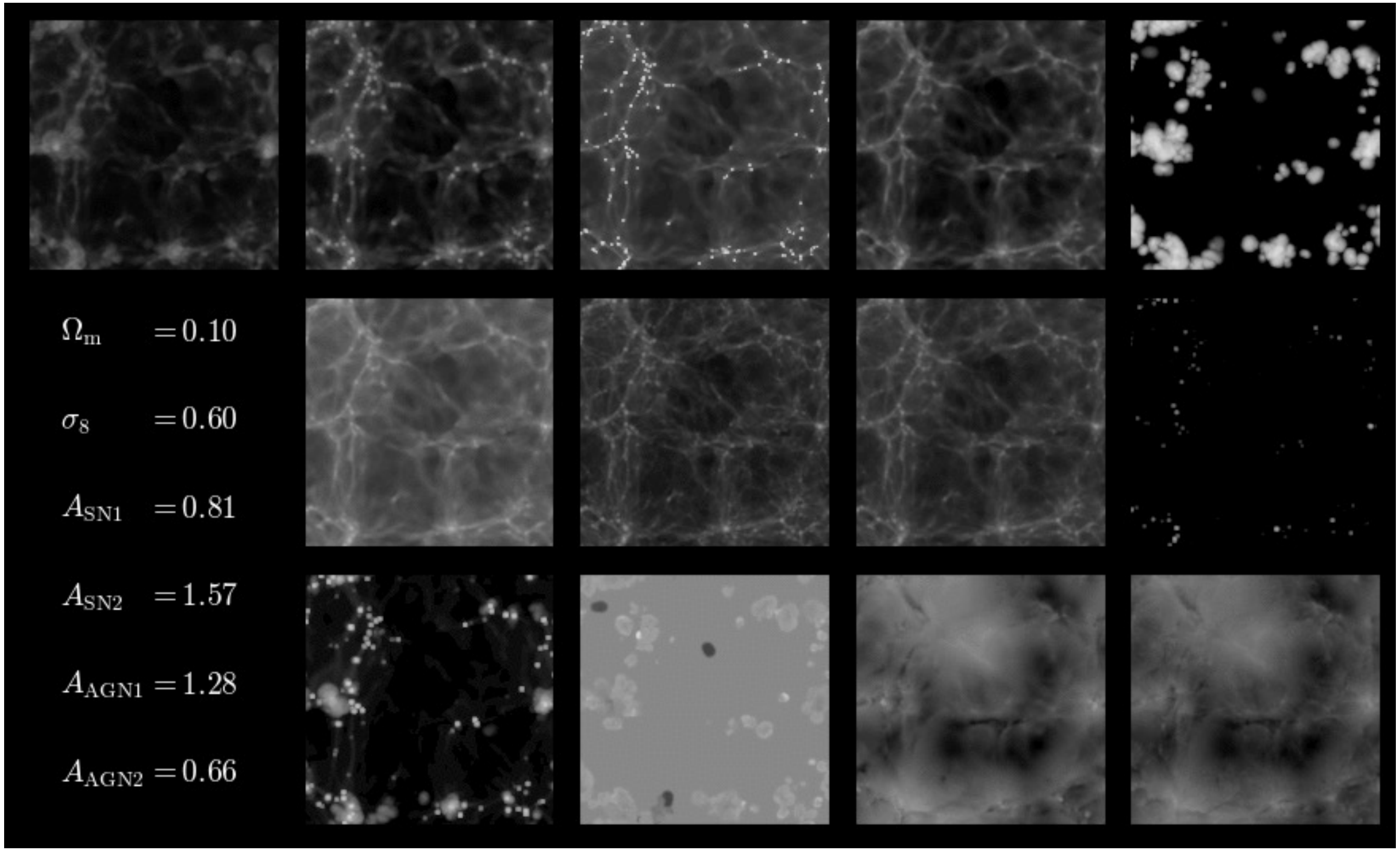
Non-Gaussian field inference with IMNN and DELFI



Available as interactive notebook tutorial at https://bit.ly/imnn-cosmo

Makinen et al., arXiv:[2107.07405](https://arxiv.org/abs/2107.07405)

CAMELS Multifield Dataset (CMD)



Paco Villaescusa-Navarro,
Shy Genel,
Daniel Angles-Alcazar, and
the CAMELS collaboration

13 fields from

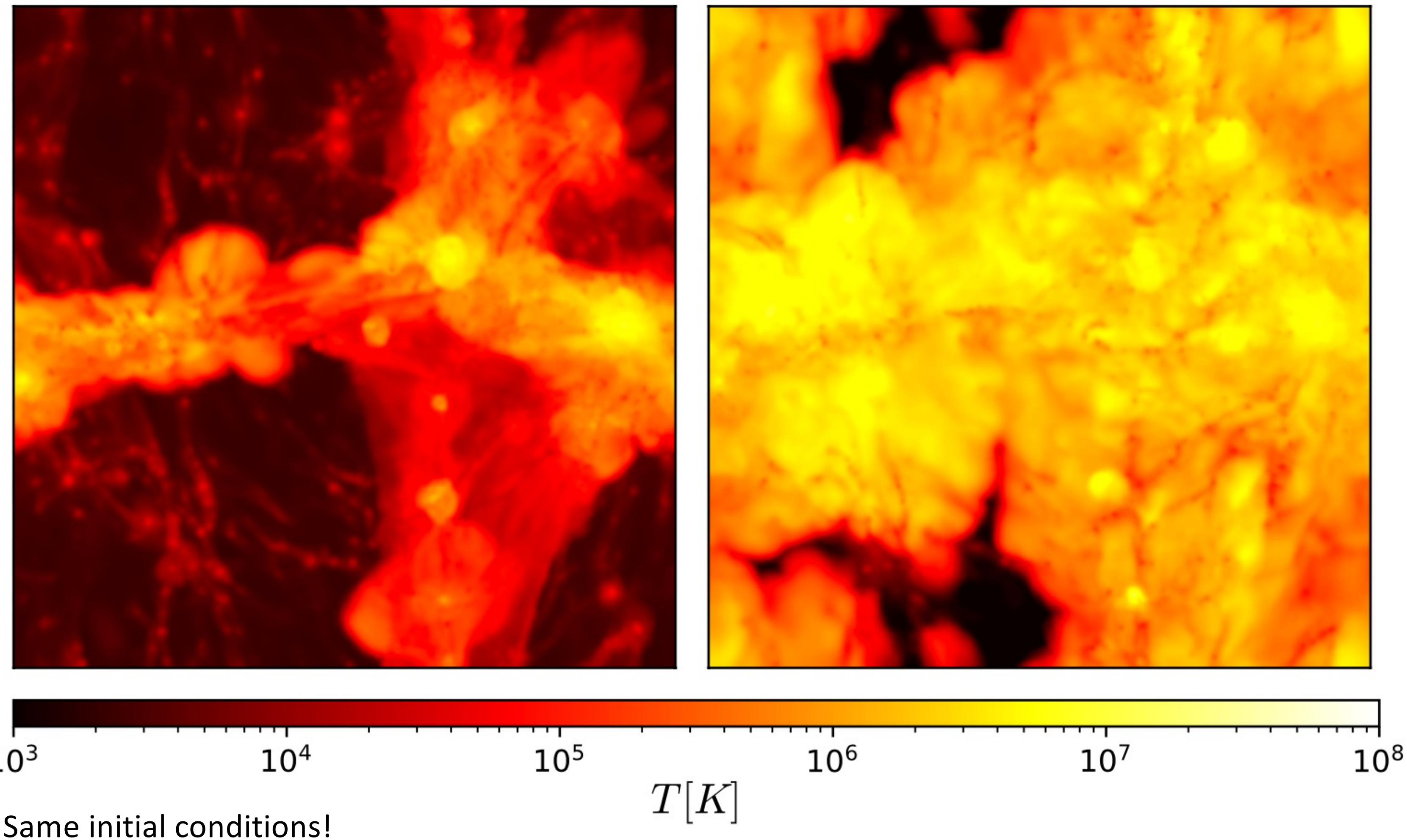
1000 IllustrisTNG sims
1000 SIMBA sims
and
2000 matched Nbody sims

arXiv:2109.10915

https://camels-multifield-dataset.readthedocs.io

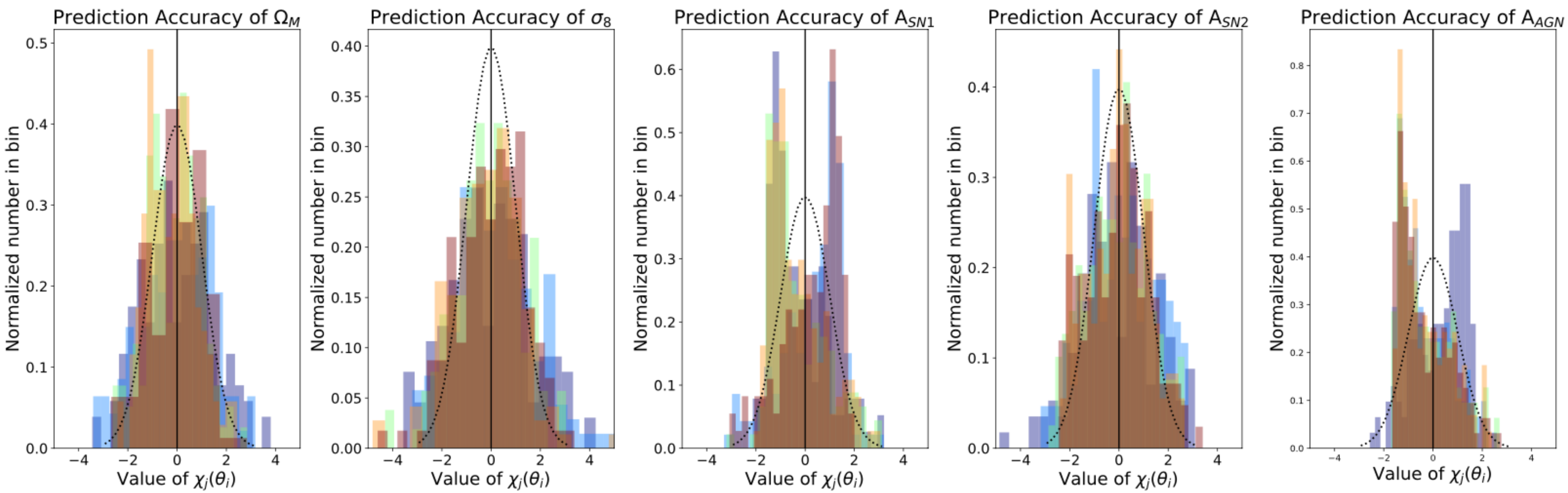
Villaescusa-Navarro et al., arXiv:2109.09747

Illustris TNG | SIMBA



Same initial conditions!

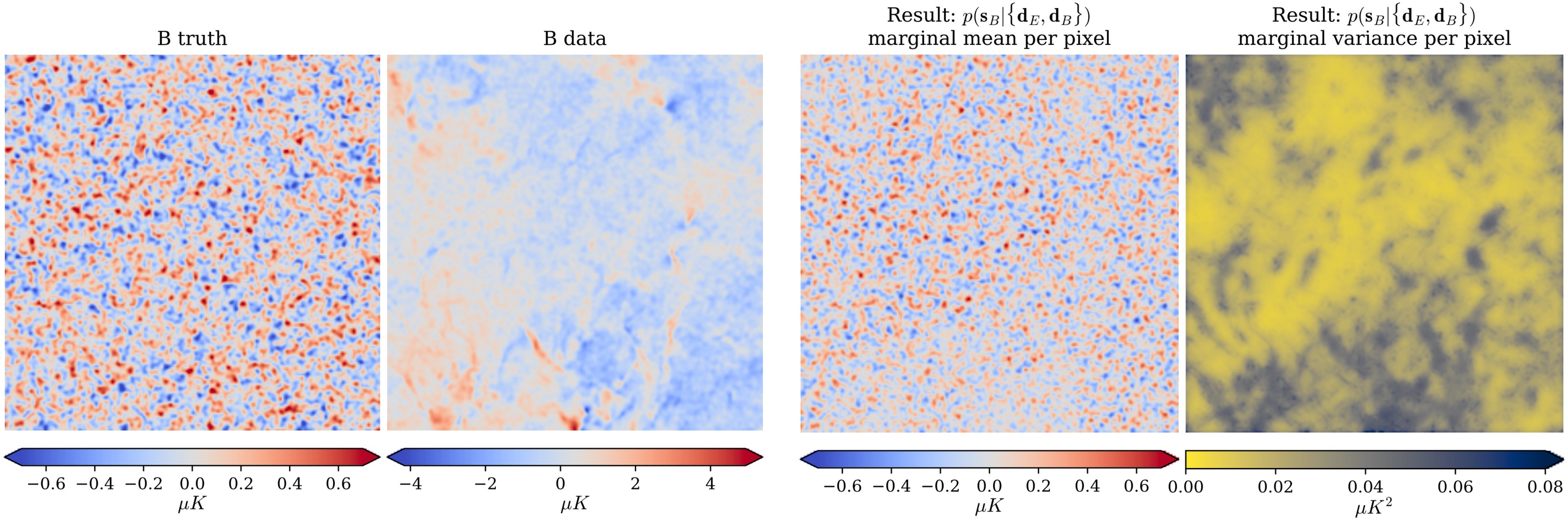
Implicit Likelihood Inference from surveys generated with **semi-analytic galaxy formation models**



Moment networks trained on SAMs run on 1000 DM sims (100 h^{-1} Mpc)^3 *stellar mass selected sample*

L. Perez et al. 2204.02408

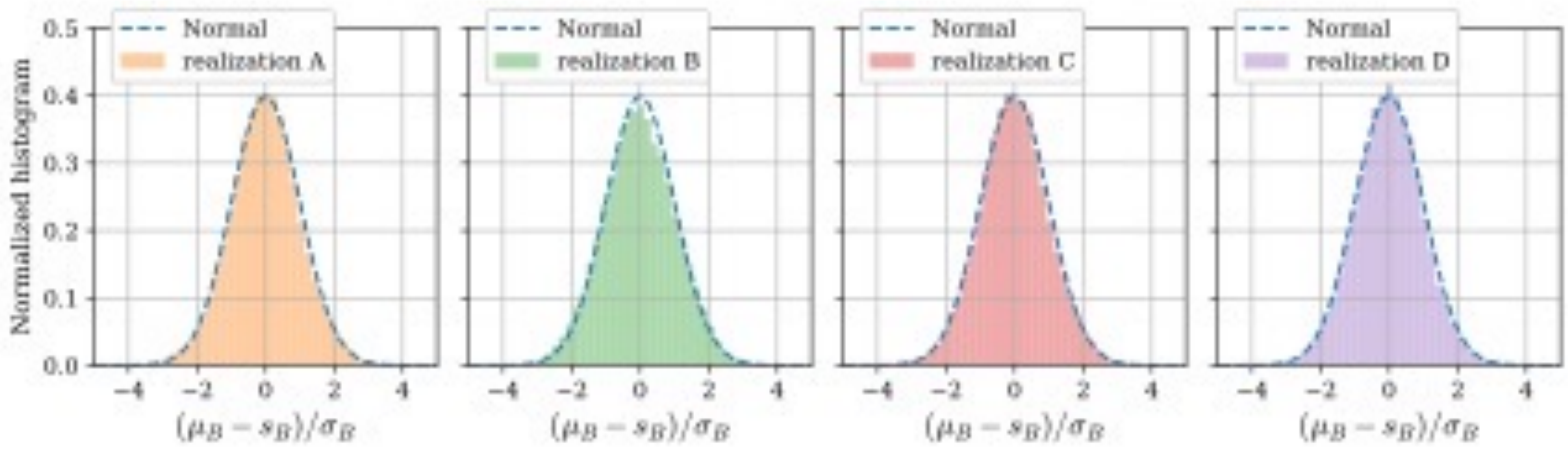
High dimensional application of Moment Networks: foreground removal



Uses a generative model based on Wavelet Phase Harmonics
based on a single training image!

Jeffrey, Boulanger, Wandelt, Regaldo-Saint Blancard, Allys, Levrier 2021, arXiv:2111.01138
(Allys et al. 2020; Regaldo-Saint Blancard et al. 2021; Jeffrey & Wandelt, arXiv:2011.05991)

Moment networks:
Posterior means and variances pass quantile test



Jeffrey, Boulanger, Wandelt, Regaldo-Saint Blancard, Allys, Levrier 2021, arXiv:2111.01138