Learning the Universe

Principled AI for Cosmology

Benjamin D. Wandelt

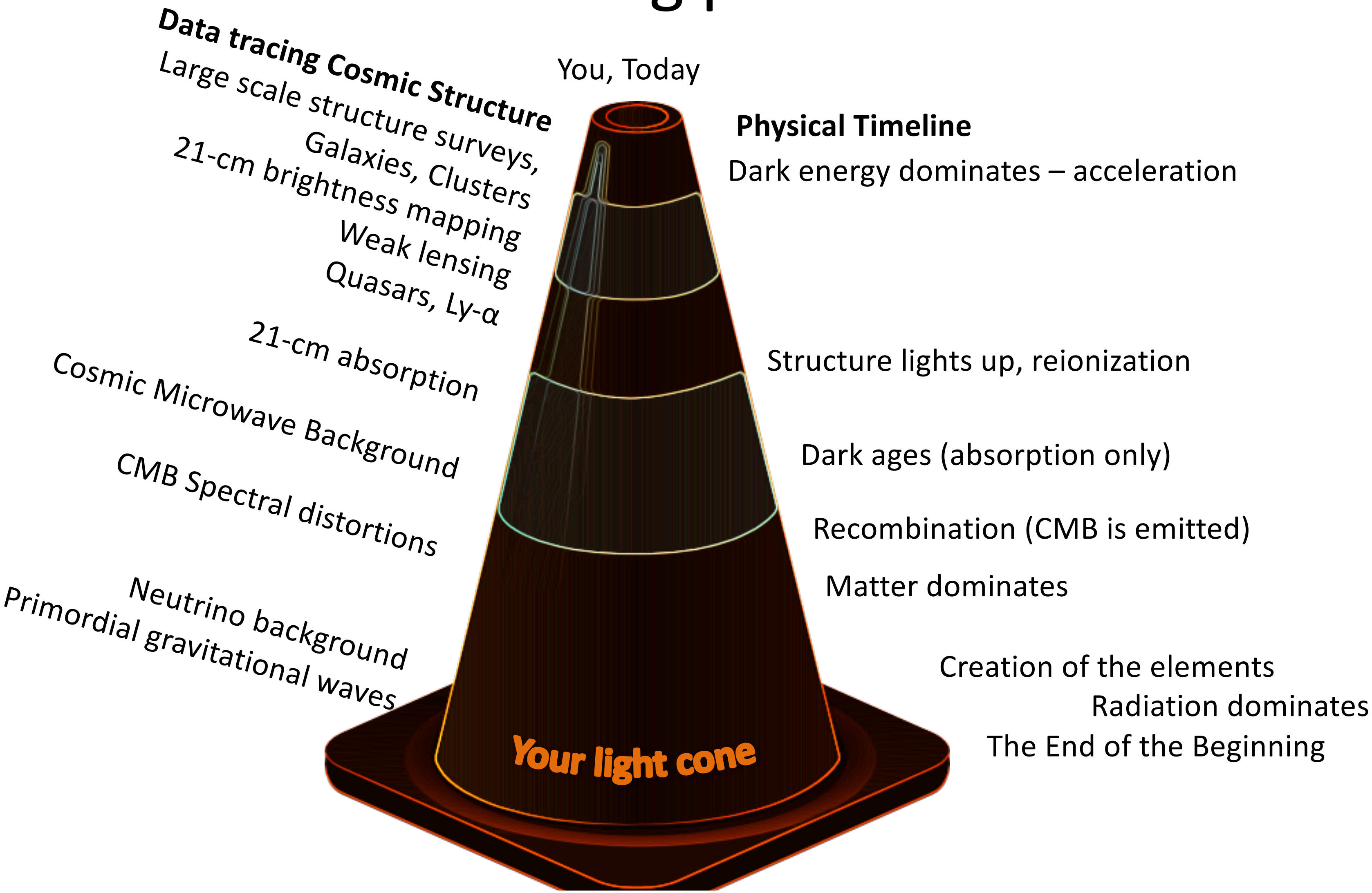
SCIENCES SORBONNE UNIVERSITÉ

FLATIRON INSTITUTE

IAP

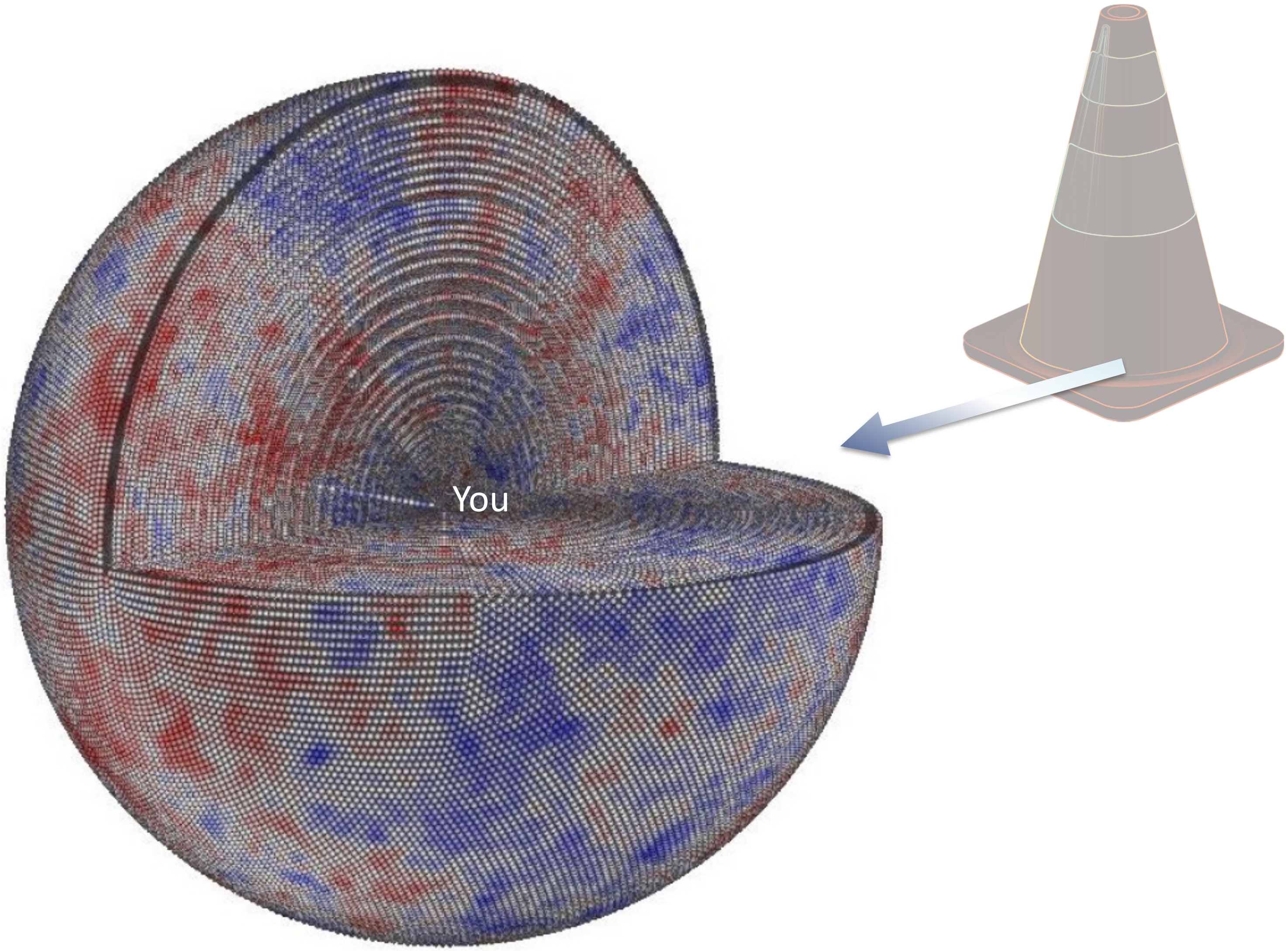
Institut Lagrange de Paris

The big picture



The initial conditions of the universe live on the base of the light cone

(curvature perturbations)

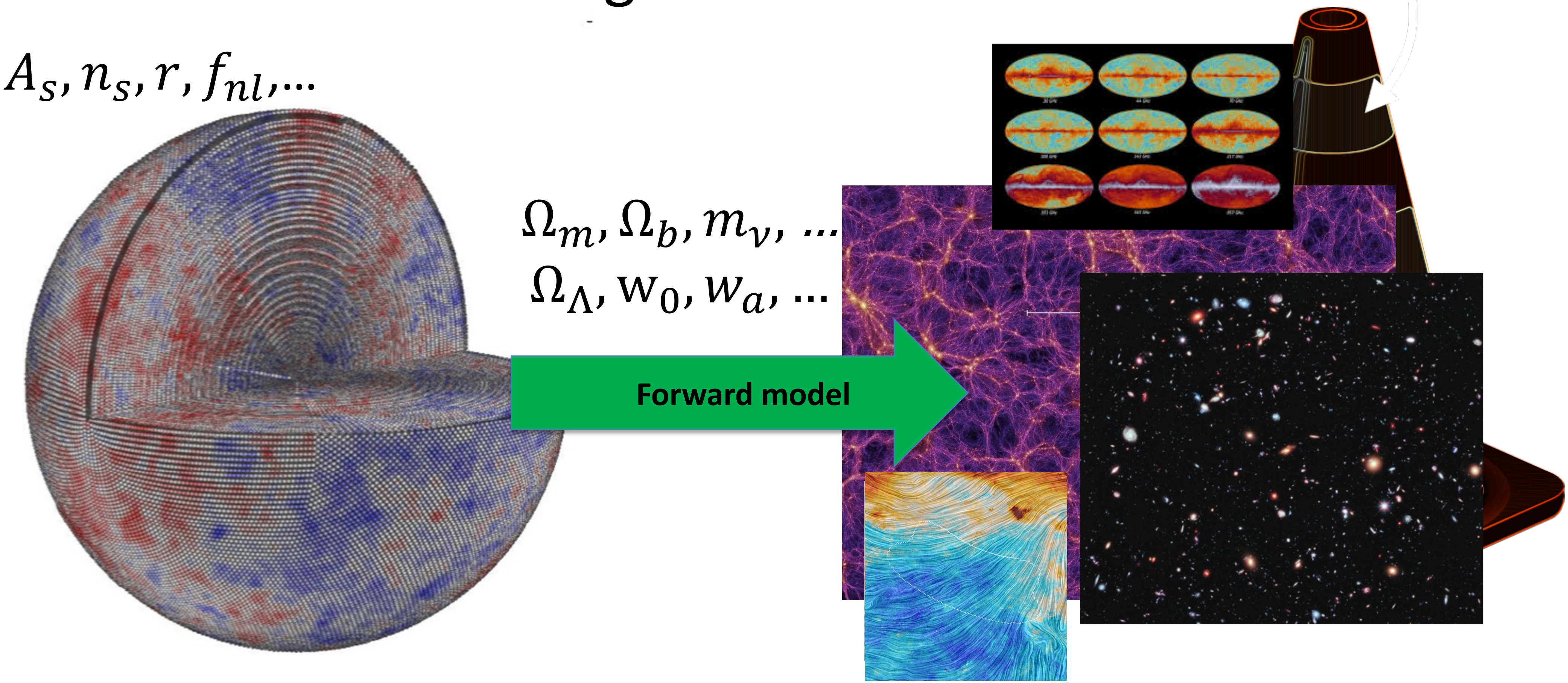


Benjamin Wandelt

We are sampling our past light cone exponentially fast

(Your favorite survey here)

The Cosmological Inference Problem



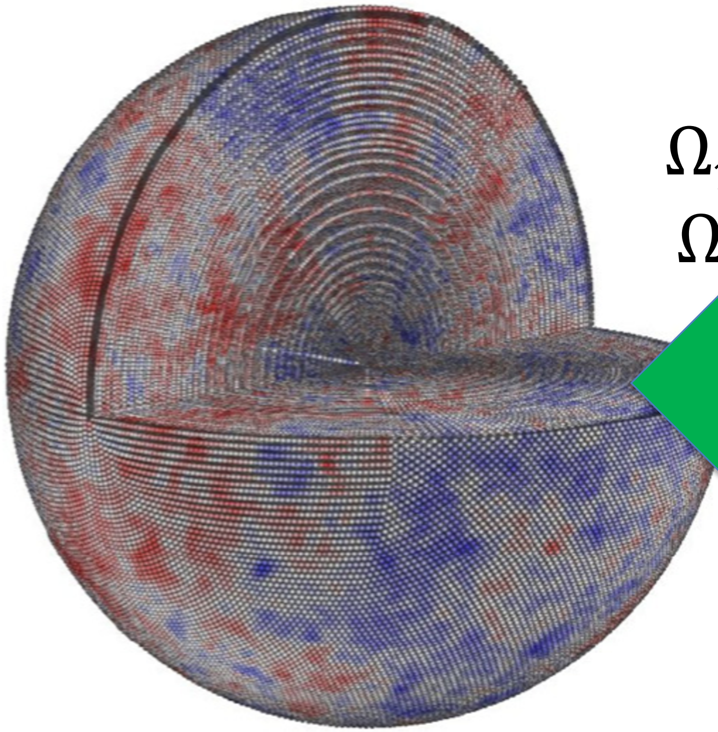
Initial conditions of the universe

The observed universe

Benjamin Wandelt

The Cosmological Inference Problem

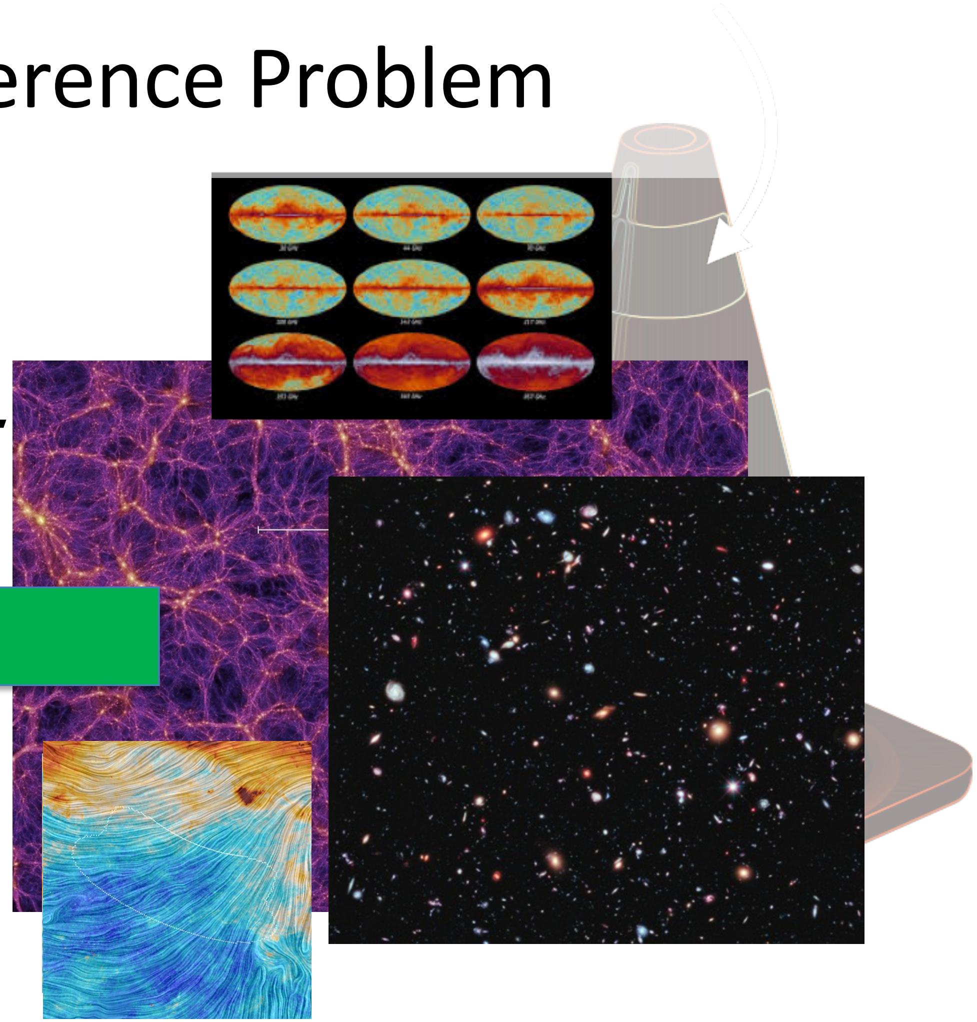
$A_s, n_s, r, f_{nl},...$



$\Omega_m, \Omega_b, m_\nu, \ ...$

$\Omega_\Lambda, w_0, w_a, \ ...$

Inference



Initial conditions

The observed universe

What we want to learn from cosmological inference

Initial conditions

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

Growth of perturbations

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

Dynamics

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

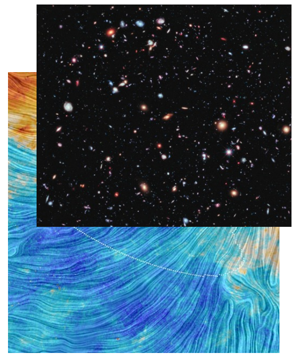
Expansion geometry

How to science, Bayesianly

1. Write down full physical and stochastic model of data given parameter.
2. Get data.

   ➔ **Likelihood**

3. Specify prior
4. Write down posterior

What if d =  ?

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

5. Explore posterior for fixed data as a function of parameters

Standard solution

Pick summaries: Power spectrum, bispectrum, counts, …

Compute predictions of these summaries: theory, simulations, emulation…

Approximate likelihood: often Gaussian

Risks of standard solution

Inadequate physics approximations (non-linear regime, (g)astrophysics, systematics, instruments…)

Inadequate statistical approximations (likelihood form) can lead to tensions

How do we know the chosen summaries exhaust the information content?

Let’s start with the initial conditions

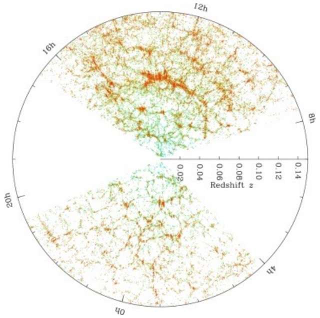
Initial condition reconstruction using Explicit Likelihood Inference: a fully generative *probabilistic* forward model of galaxy surveys

BORG: *Bayesian Origin Reconstruction from Galaxies*

- Gaussian prior + **Gravity** + likelihood for galaxies

  (includes Particle-Mesh or LPT gravity solver, survey model, bias model, automatic noise level calibration, selection function, mask, …)

- Hamiltonian Markov Chain **with >10^7 parameters**



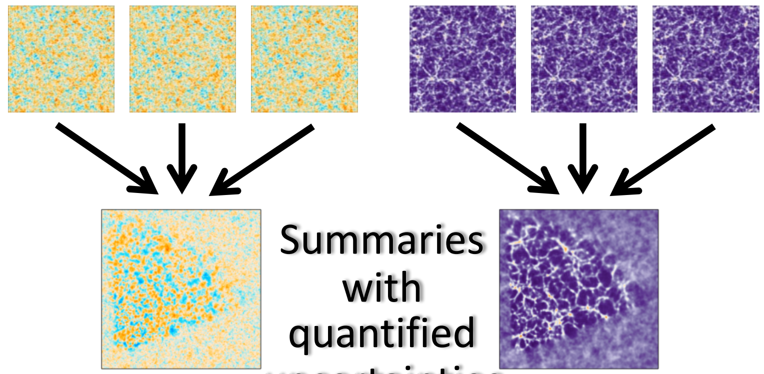
Observations

(galaxy catalog + meta-data: selection functions, completeness…)

→ BORG →

Initial conditions, and inferred dark matter densities

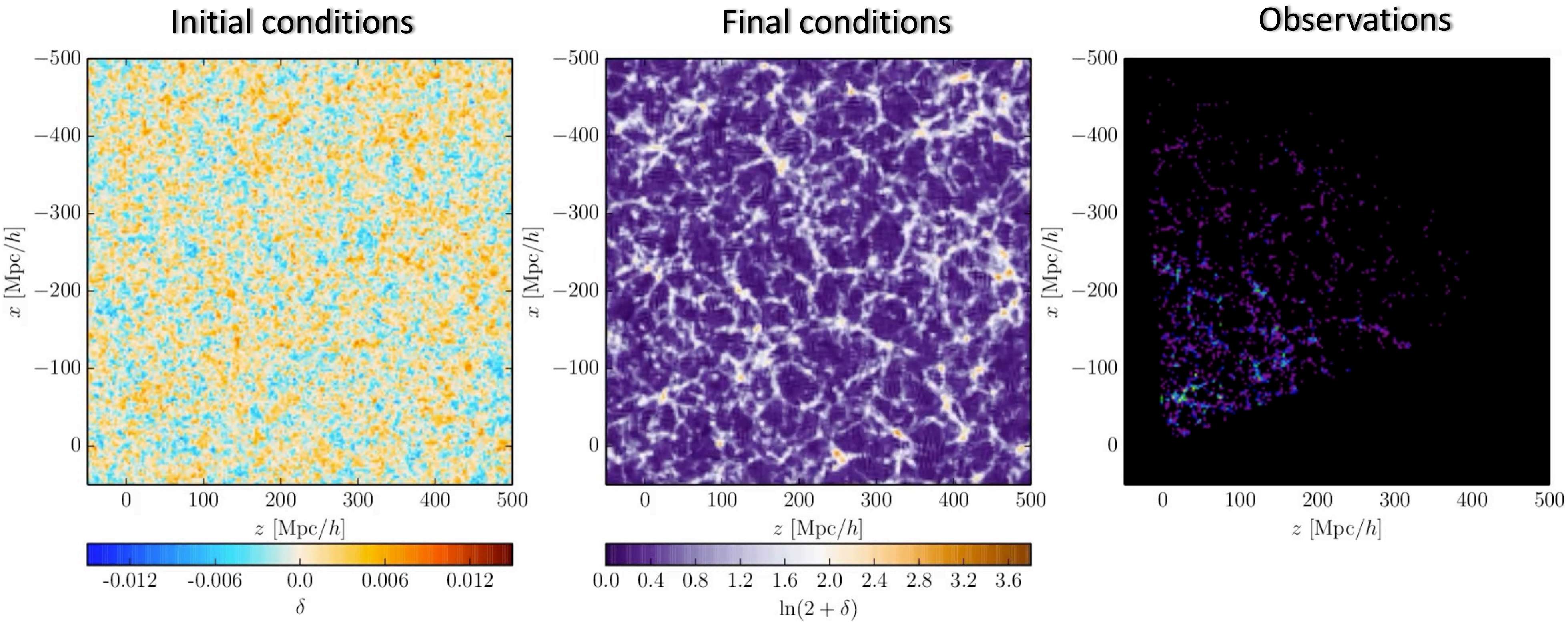
z=100 z=0



Summaries with quantified uncertainties

Jasche & Wandelt 2013, arXiv:1203.3639
Jasche, Leclercq & Wandelt 2015, arXiv:1409.6308
Lavaux & Jasche 2017…

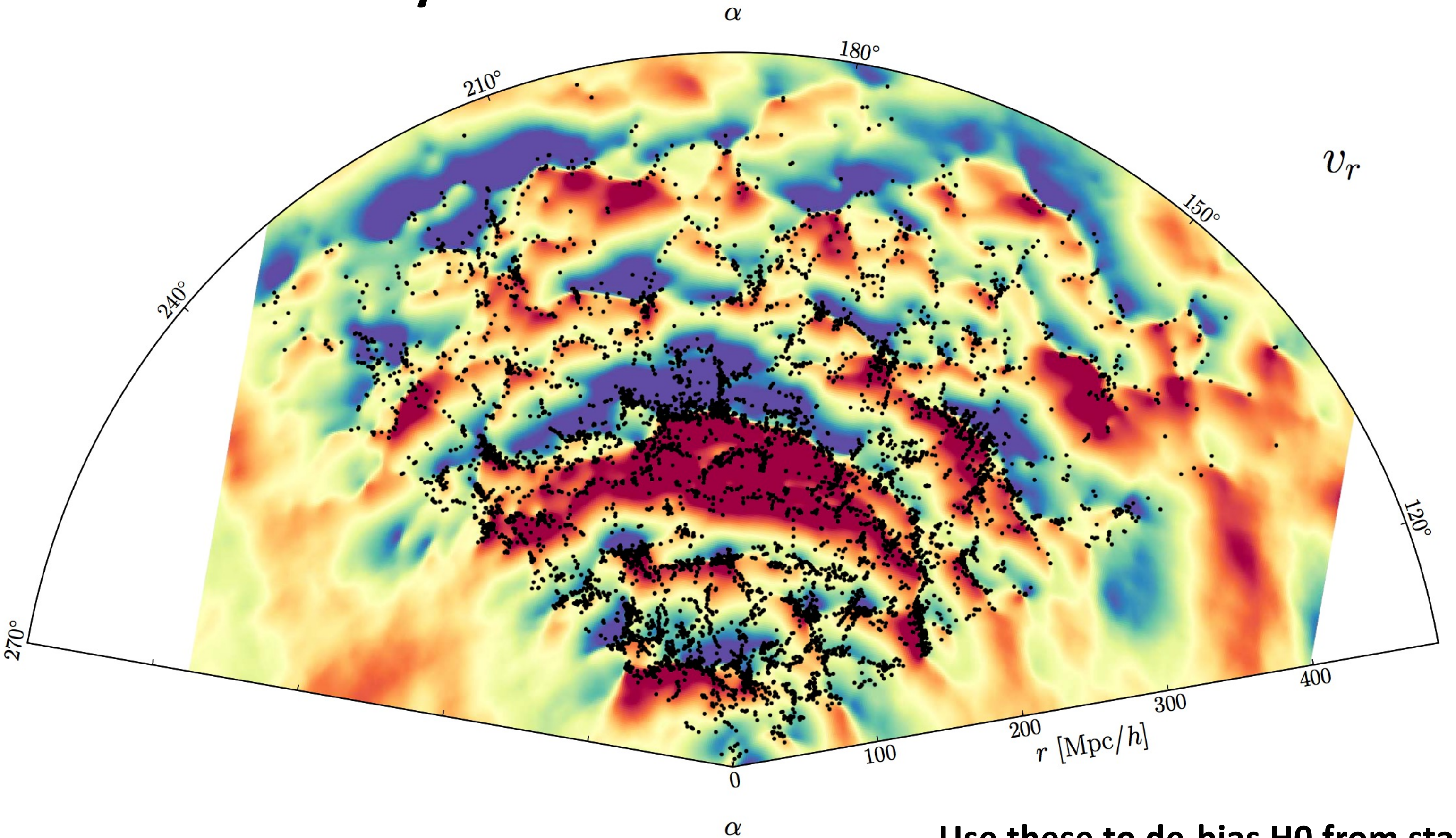
Bayesian cosmological initial conditions from real data since 2013



e.g. Jasche & Wandelt 2013, arXiv:1203.3639; Jasche, Leclercq & Wandelt 2014, arXiv:1409.6308

See full bibliography and current status at **http://aquila-consortium.org**

Example Bayesian LCDM results: dynamical velocities



Leclercq et al. 2017

Use these to de-bias H0 from standard sirens: Mukherjee et al arXiv:1909.08627

So is that it – are we done? Problem solved?

- The full statistical power even of current data is enormous
- Need:
  - more reality in the data model;
  - better ability to project/cut/mask the data for cosmological inference to become insensitive to remaining model error
  - Build in robustness to model mis-specification or residual model error using *physical principles.*

The role of Machine Learning

In principle, fully *ab initio,* physics-based models like BORG allow the tightest possible confrontation of models and data and eliminate the risk of "forgetting" an informative summary.

But is it really practical to write down a likelihood that includes ***everything?***

Principled used of physics-based Machine Learning (ML) can help in connecting physical models to data.

Neural Physical Engines: Modeling **bias** with ML

- We designed a new type of neural network to learn (cosmological) physics
- The network encodes relevant approximate physical symmetries/constraints
  - Translation invariance
  - Local rotational invariance
  - Locality
- A neural network with only **17 parameters**!

- Use it in the BORG framework **as a bias layer to map DM density to halos**
  Charnock, Lavaux, Wandelt, Boruah, Jasche, Hudson (arXiv:1909.06379)

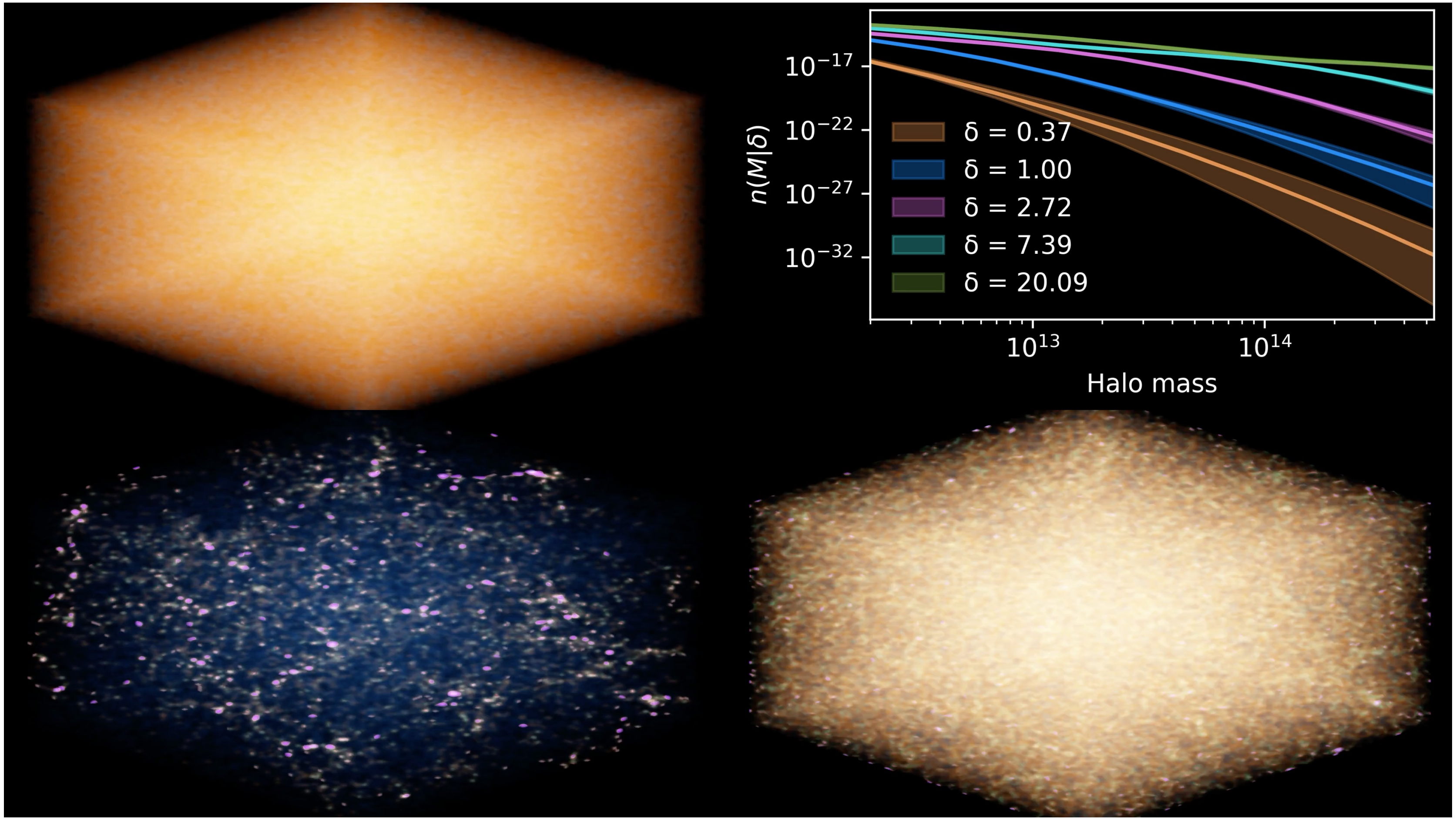
- *This allows **zero shot learning**:* needs *no* training data!
- A fully Bayesian neural network with data-driven MCMC inference of network parameters and cosmological initial conditions

Neural physical engines for inferring the halo mass distribution function

Charnock, Lavaux, Wandelt, Boruah, Jasche, Hudson (arXiv:1909.06379)

DM reconstruction (shown at z=0) and Initial conditions (not shown)

Simulated data: halo distribution



Neural forward model of halo distribution within BORG

Too much, too fast?
Let's relax and focus on geometrical tests

Cosmographic/geometric tests probe aspects of the data that are robust to model misspecification

This avoids having to model the full complexity of the data.

Can we use this geometrical approach to do cosmological inference with BORG?

- Going to a geometric approach decouples the “bias” model from cosmological parameters
- By *only keeping the cosmological parameter dependence in coordinate mapping w*e can use BORG to do a generalized, non-Gaussian, field-based “Alcock-Paczynski” on the light cone

A field-based Alcock-Paczynski test
(not just 2-point stats!)

Coordinate Transformation

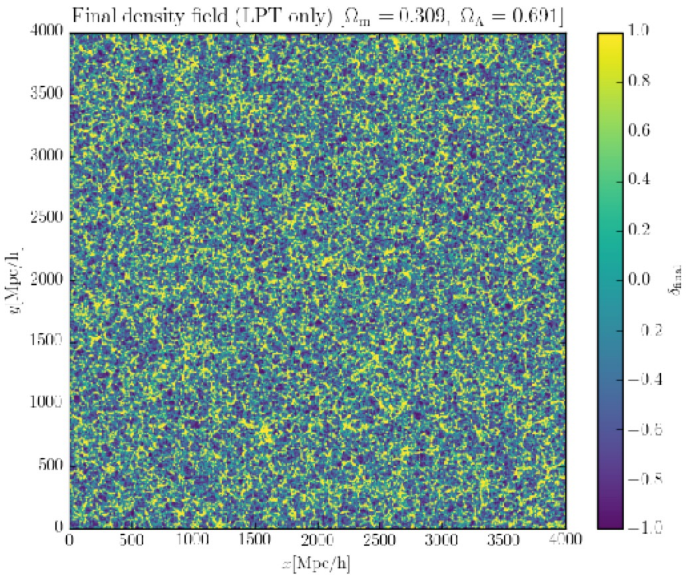
(Alcock & Paczyński 1979)

$(z, \vec{u})$ $\xleftrightarrow{\{\Omega_m, \Omega_\Lambda, \Omega_k, \omega, H_0\}}$ $(d, \vec{u})$

Redshift space ⟷ **Comoving space**

- Distortions due to assumption of incorrect cosmological parameters
- Structure: **Spherical → Ellipsoidal**
- Statistical distribution: **Isotropic → Anisotropic**

comoving space



$$d = \int_{z_1}^{z_2} \frac{1}{cH(z)}$$

$$H(z) = H_0(\Omega_m(1+z)^3 + \Omega_k(1+z)^2 + \Omega_\Lambda)^{\frac{1}{2}}$$

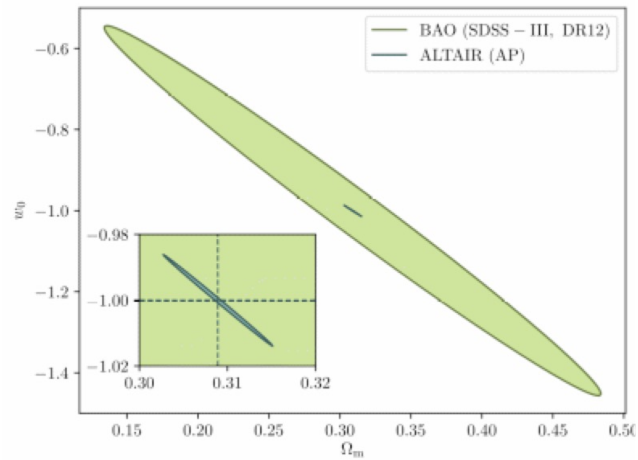
Kodi Ramanah et al., arXiv 1808.07496

High precision inferences

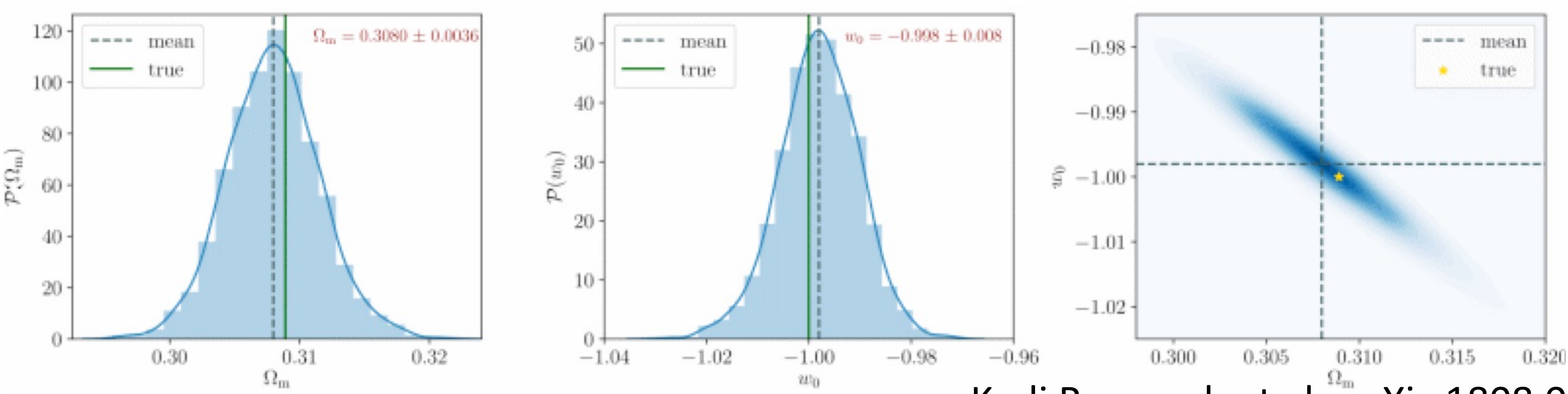
- Probing deep redshift range; geometric distortion due to cosmic expansion is highly informative

$$\{\Omega_{\rm m} = 0.3080 \pm 0.0036, \;\; w_0 = -0.998 \pm 0.008\}$$

Comparison to standard BAO constraints

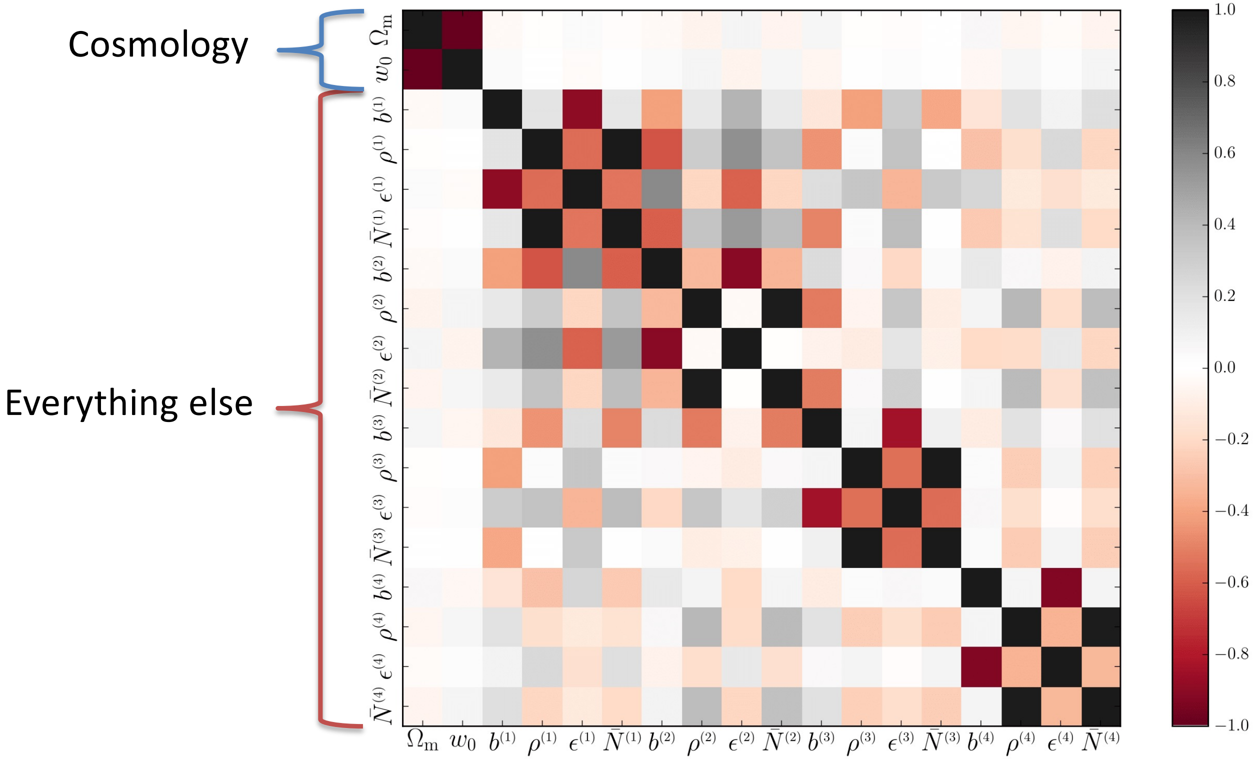


Marginal & joint posteriors



Kodi Ramanah et al., arXiv 1808.07496

Focusing on geometry works: Cosmology and bias parameters decouple!



Kodi Ramanah et al., arXiv 1808.07496

Relaxed?

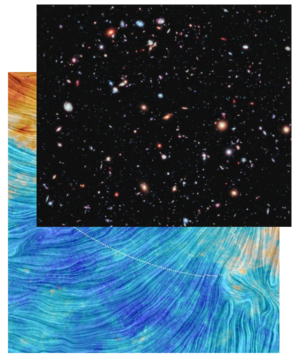
Good! Let’s get back to solving the full problem!

How to science, Bayesianly

1. Write down full physical and stochastic model of data given parameter.
2. Get data.

   ➔ **Likelihood**

3. Specify prior
4. Write down posterior

What if d =  ?

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

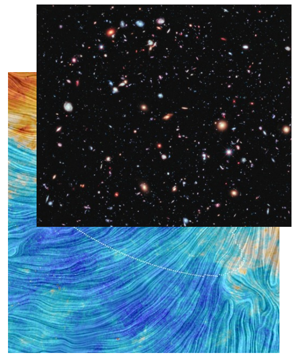
5. Explore posterior for fixed data as a function of parameters

How to science, Bayesianly

1. Write down full physical and stochastic model of data given parameter.
2. Get data.

   ➔ **Likelihood**

3. Specify prior
4. Write down posterior

What if d =  ?

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

5. Explore posterior for fixed data as a function of parameters

The full problem

To succeed we need more freedom than a traditional likelihood approach can provide:

- FREEDOM to make our physical model anything we want
- FREEDOM to project/summarize/cut/mask our data any way we want

Simulating data is **much easier** than deriving an accurate likelihood.

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

Simulations are draws from the likelihood

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

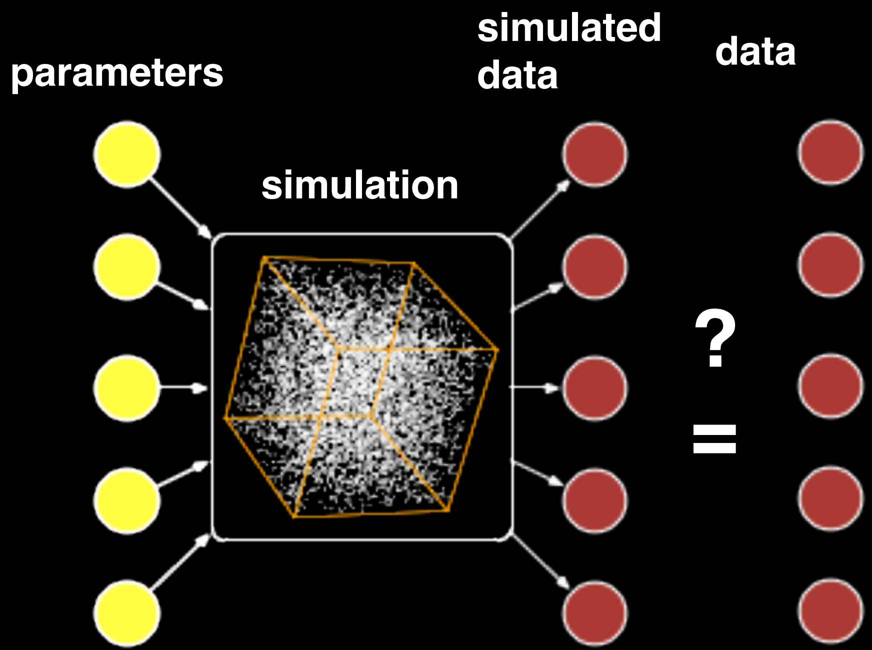
$$\mathbf{d}^* \leftarrow \text{simulation}(\mathbf{d}^*|\boldsymbol{\theta})$$

Team: Implicit likelihood methods (a.k.a. simulation-based or likelihood-free inference)

Justin Alsing, Tom Charnock, Stephen Feeney, Francisco V-N, Niall Jeffrey, Lucas Makinen, Nicolas Chartier

Guilhem Lavaux, …

Simulation based inference



Draw from prior:

$$\boldsymbol{\theta} \leftarrow P(\boldsymbol{\theta})$$

Simulate data:

$$\mathbf{d}^* \leftarrow P(\mathbf{d}^*|\boldsymbol{\theta})$$

If $\rho(\mathbf{d}^*, \mathbf{d}) < \epsilon$

accept;

else:

reject;

In the limit $\epsilon \to 0, \; \{\boldsymbol{\theta}\} \leftarrow P(\boldsymbol{\theta}|\mathbf{d})$

Benjamin Wandelt

Simulation-based inference

How to reduce data-space?

Draw from prior:

$\boldsymbol{\theta} \leftarrow P(\boldsymbol{\theta})$

Simulate data:

$\mathbf{D}^* \leftarrow P(\mathbf{D}^*|\boldsymbol{\theta})$

If $\rho(\mathbf{D}^*, \mathbf{D}) < \epsilon$:

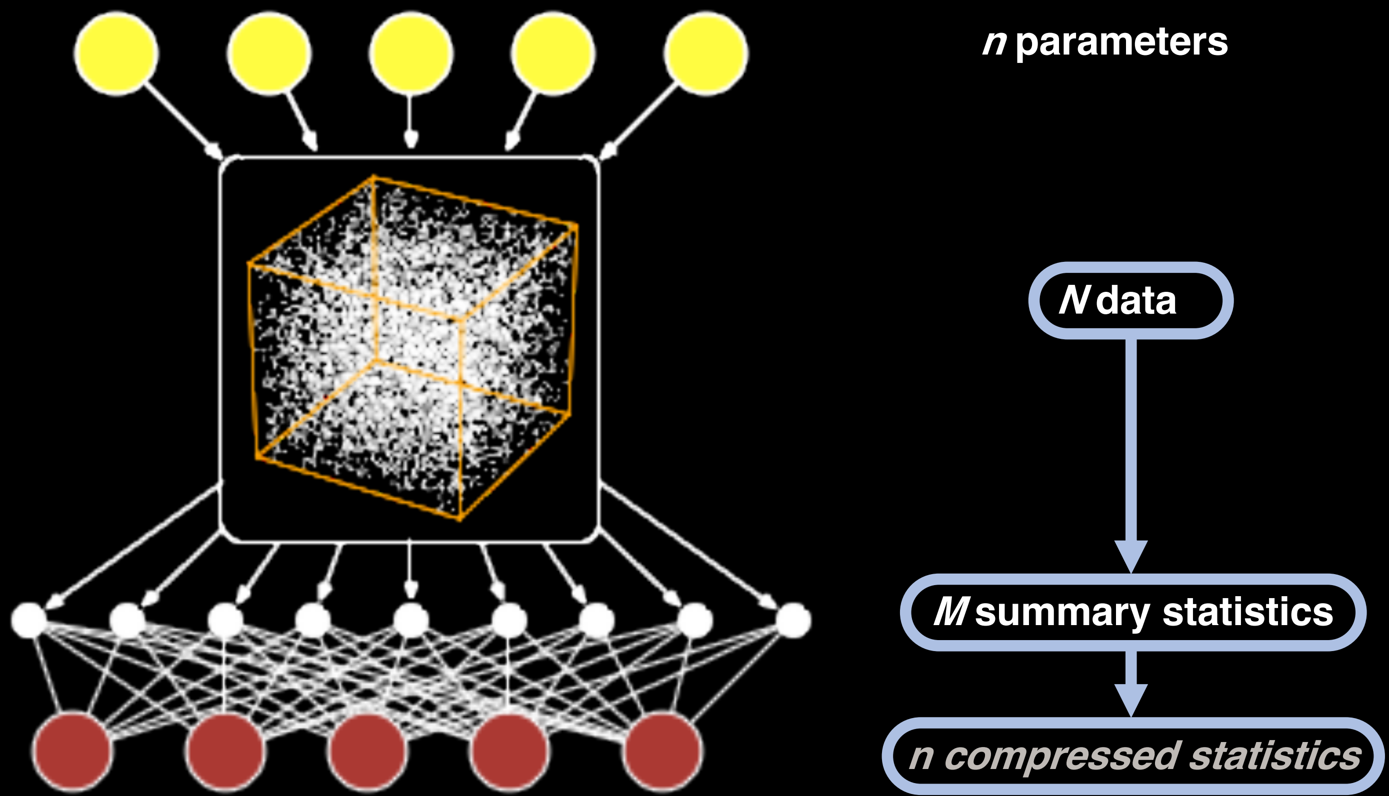
accept;

else:

reject;

In the limit $\epsilon \rightarrow 0, \quad \{\boldsymbol{\theta}\} \leftarrow P(\boldsymbol{\theta}|\mathbf{D})$

Reducing data space: massive data compression



Score compression: Alsing & Wandelt arXiv:1712.00012; Heavens, Jimenez & Lahav 2000

Simulation-based inference

How to reduce data-space? ✅

How to explore parameter-space?

In the limit $\epsilon \rightarrow 0, \quad \{\boldsymbol{\theta}\} \leftarrow P(\boldsymbol{\theta}|\mathbf{D})$

Machine Learning to the rescue!

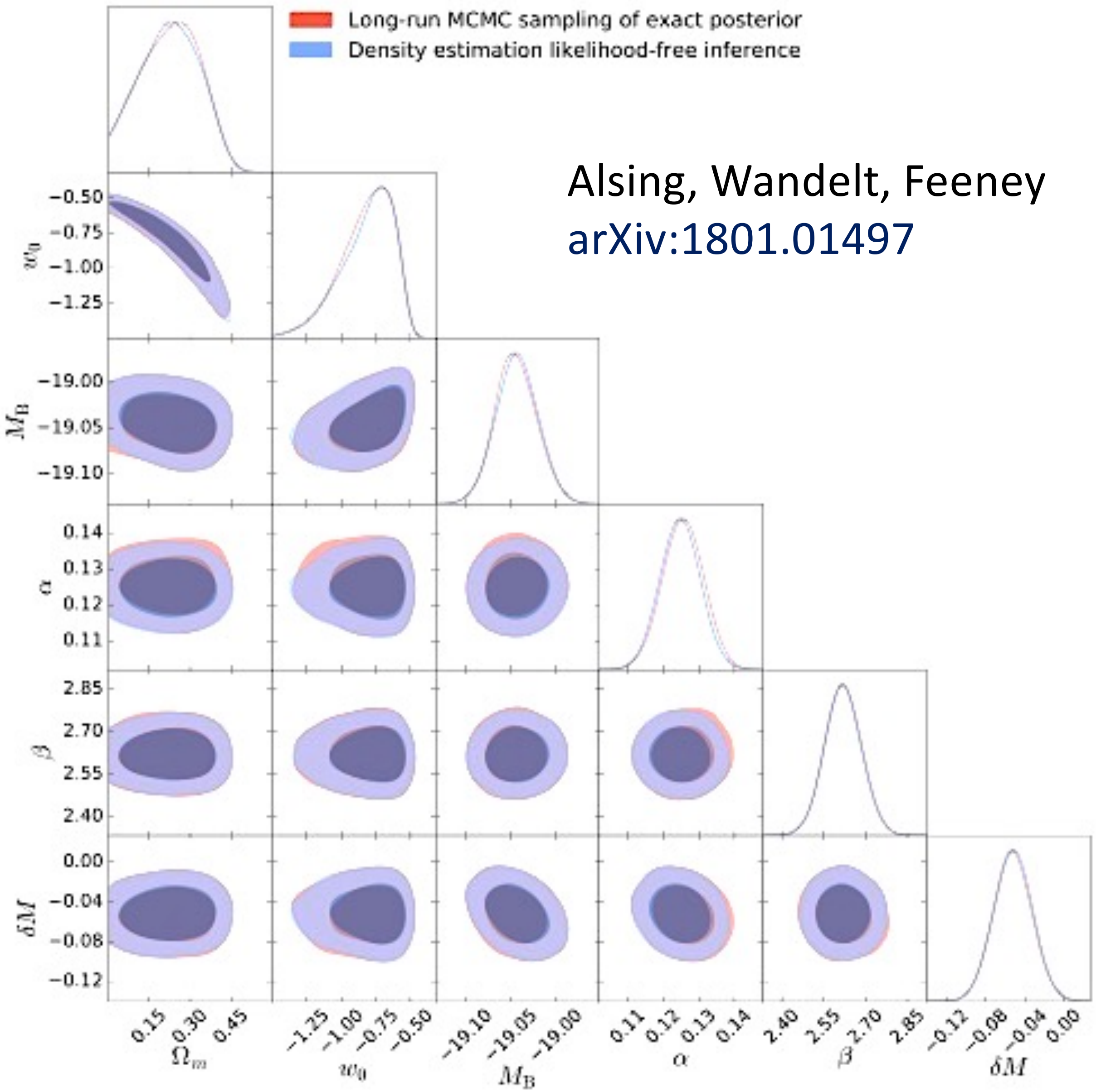
Density estimation Likelihood free inference (DELFI):

Directly learn probability density of parameters and compressed data

Alsing, Feeney & Wandelt arXiv: 1801.01497

DELFI
Posterior inference works…

and is much faster than Explicit Likelihood Inference with MCMC!



Alsing, Wandelt, Feeney
arXiv:1801.01497

(O(1000) simulations)

Density Estimation Likelihood-Free Inference

- New *nuisance-hardened* compression greatly reduces required number of simulations and allows many more parameters (Alsing & Wandelt arXiv:1903.01473).

- New version of DELFI now released including neural density estimators to fit the likelihood (Alsing, Charnock, Feeney, Wandelt arXiv:1903.00007)
  - Includes active learning for deciding where to run simulations

But what if you don't know how to compute informative summaries of your data?

Machine Learning to the rescue!

Automatic Physical Inference with Information Maximizing Neural Networks (IMNN)

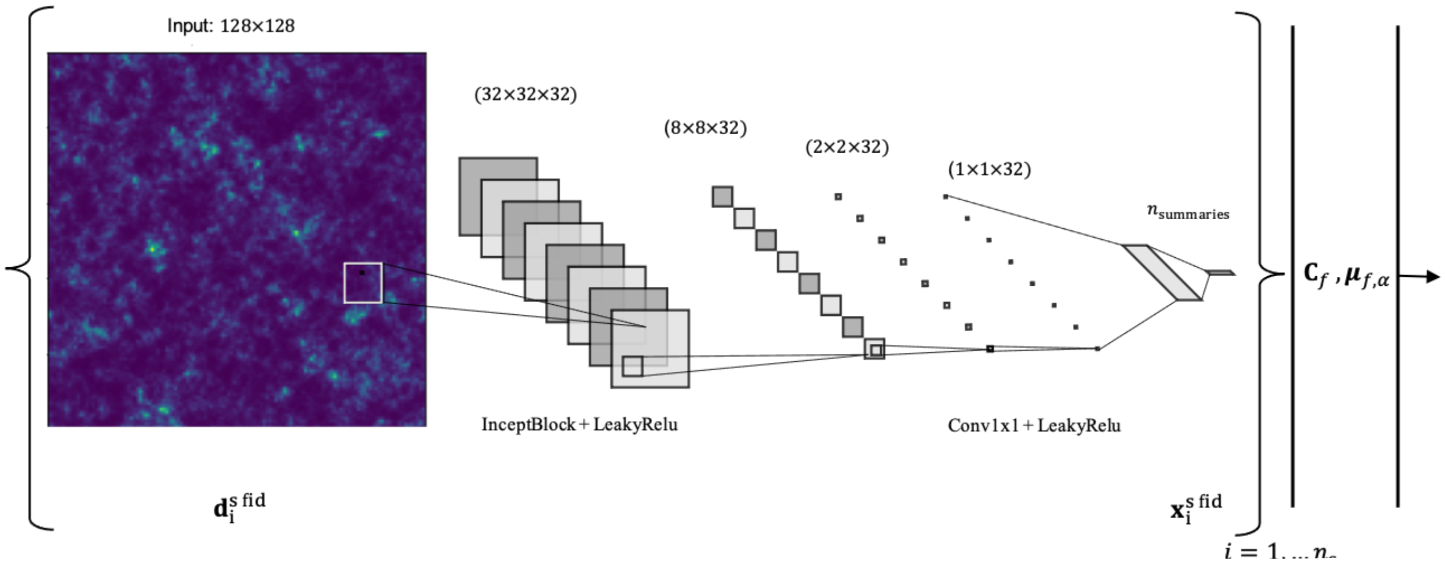
Charnock, Lavaux, Wandelt (arXiv:1802:03537)

- Goal: remove the need to "guess" heuristic, informative summaries of the data
- Setup: a neural network that maps the data into a small set of maximally informative *summaries*

- Training uses physical simulations of the model at a fiducial point in parameter space
- The achieved loss on a test set is meaningful – it's the information content of the data.
- Can prove that the IMNN computes the optimal (score) compression without knowing the likelihood (Wandelt et al., in prep)

SBI Parameter Inference Using Optimal Compression

- Idea: Likelihood is implicitly defined through forward simulations.
- Fit likelihood with neural density model (or accept/reject parameters based on similarity of simulations to data)
- Compress data for dimensional reduction.
- ML safety through identical data and simulation pipelines.
- *Optimal information summaries* of the data found by neural networks trained on physical simulations: Information Maximizing Neural Networks (IMNN) or Regression Networks
- IMNN optimal loss on a test set is the recovered information content of the data.
- The IMNN training loss provably defines the optimal (LH score) compression at the fiducial model *purely based on simulations.*



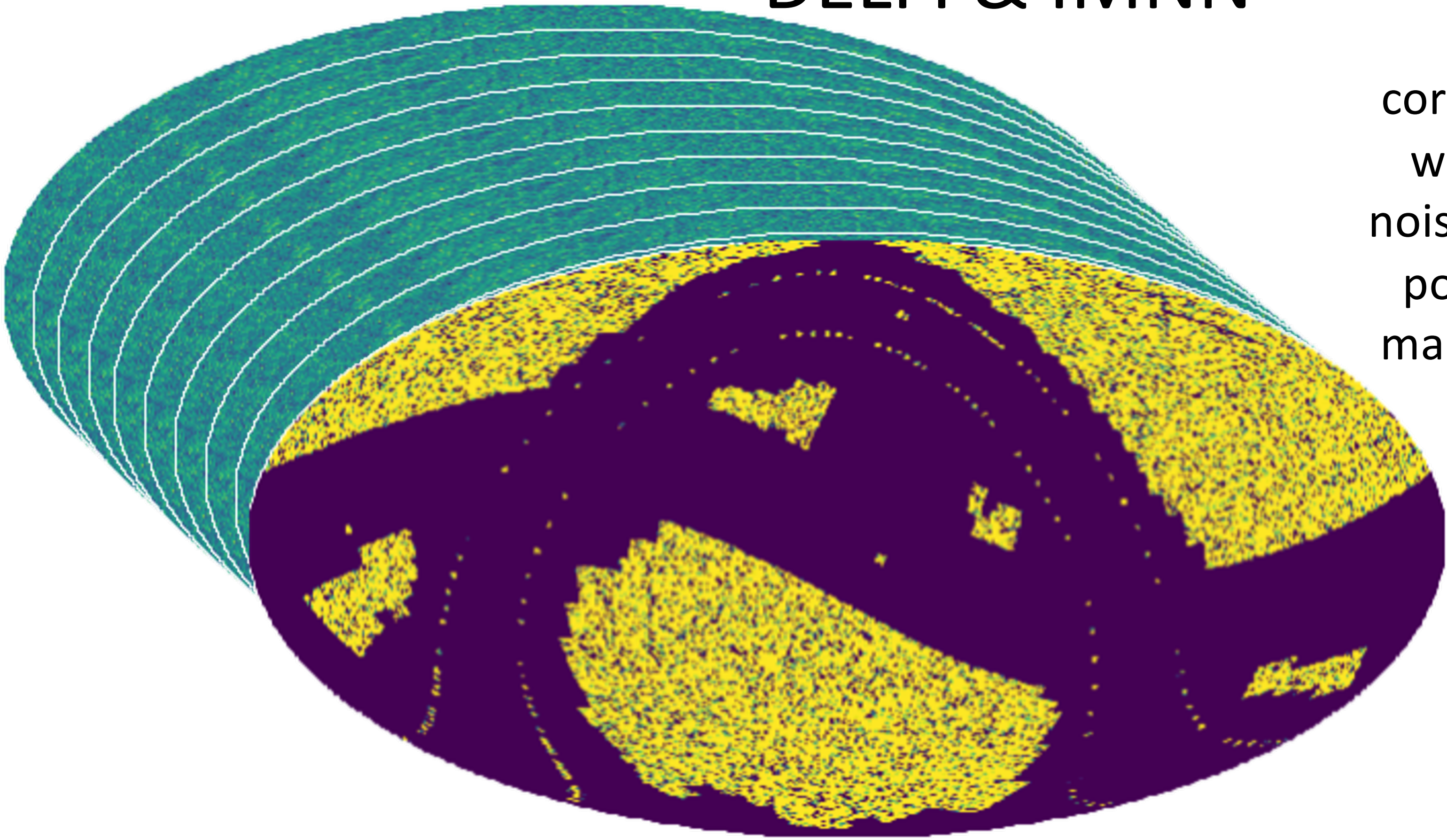
Optimal compression network minimizes information loss

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

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

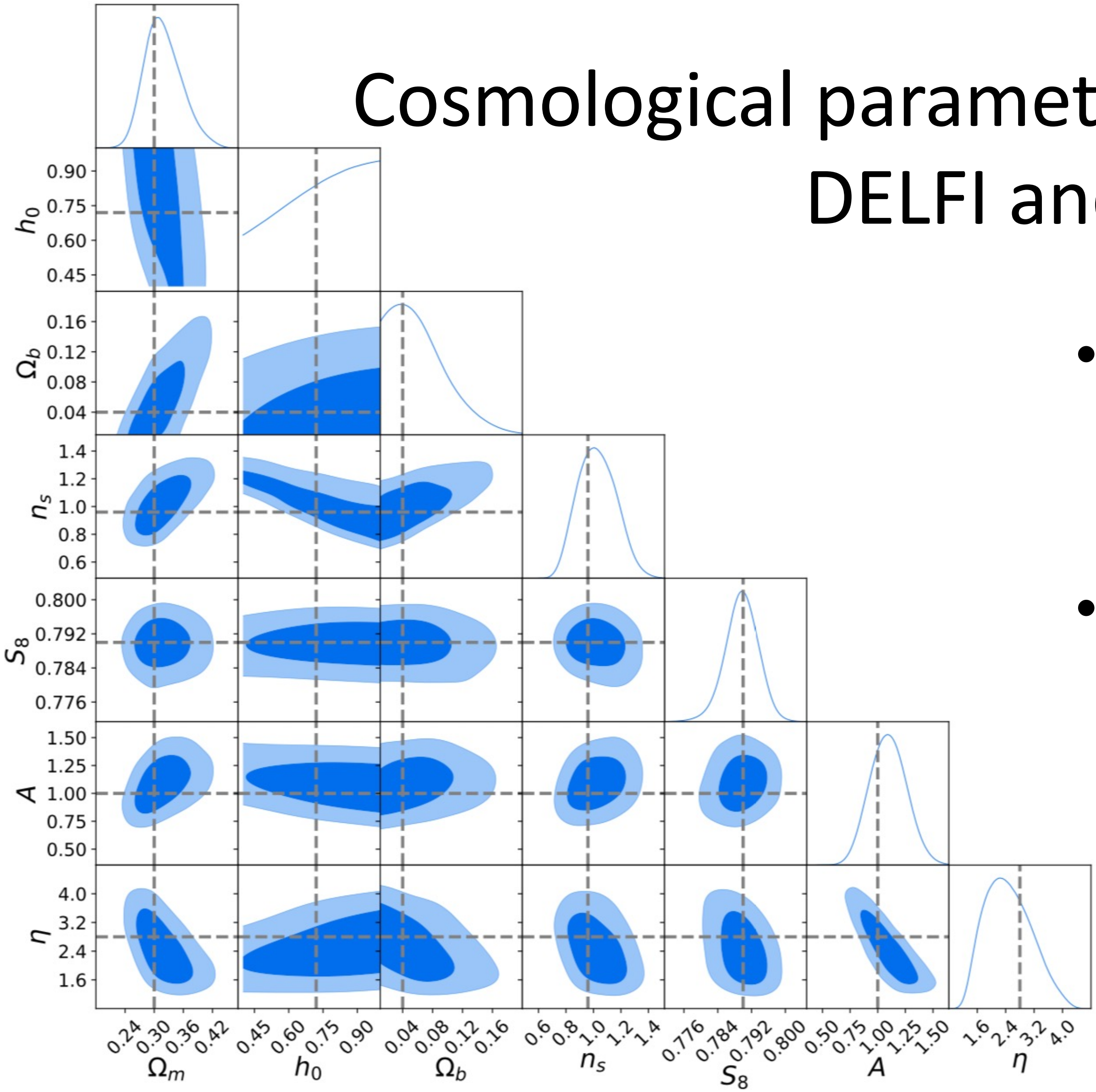
bitbucket.org/tomcharnock/imnn/

Example application: weak lensing tomography with DELFI & IMNN



10 spherical shells of correlated shear simulation, with Euclid-like mask and noise, pre-compressed to the power spectrum and then massively compressed using IMNN

Taylor, et al., arXiv: 1904.05364



Cosmological parameter inferences using DELFI and IMNN

- First Bayesian **weak lensing analysis with Non-Gaussian lensing potential**
- Enabled by DELFI and IMNN

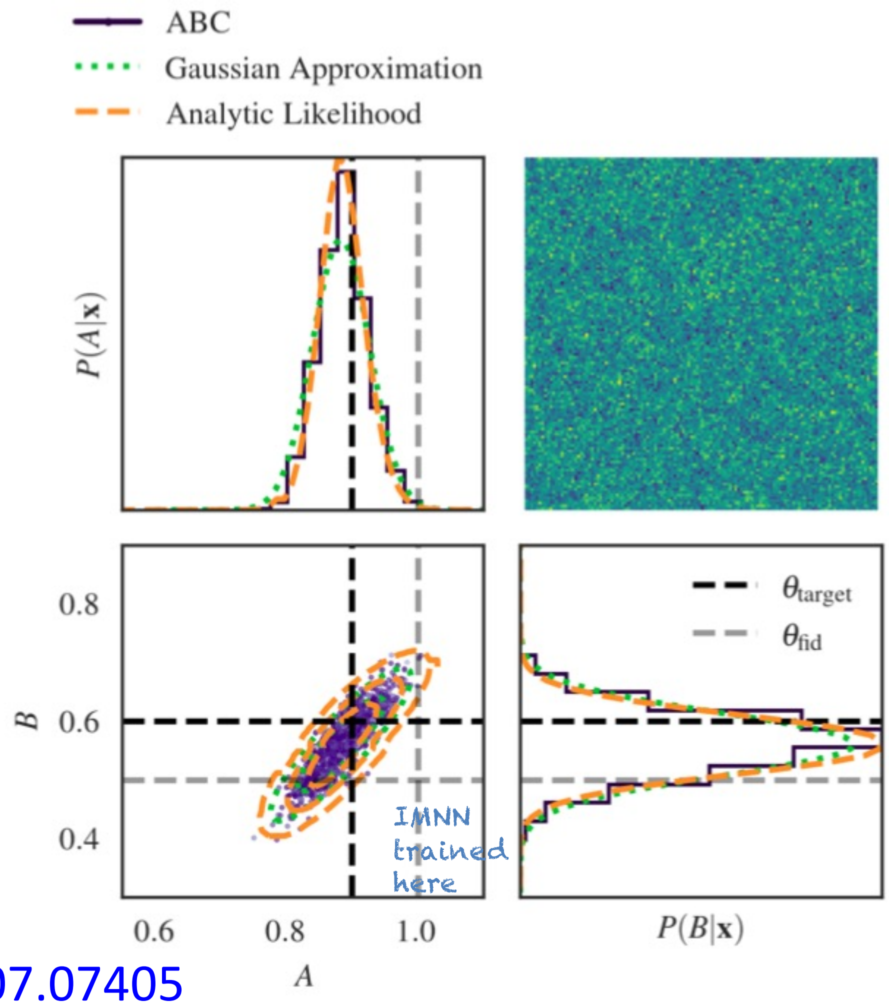
Taylor, et al., arXiv: 1904.05364

(see also Diaz Rivero & Dvorkin arxiv:2007.05535)

Field-Based Implicit Likelihood Inference with Information Maximizing Neural Networks

Benjamin Wandelt

IMNN recovers full info directly from Gaussian field



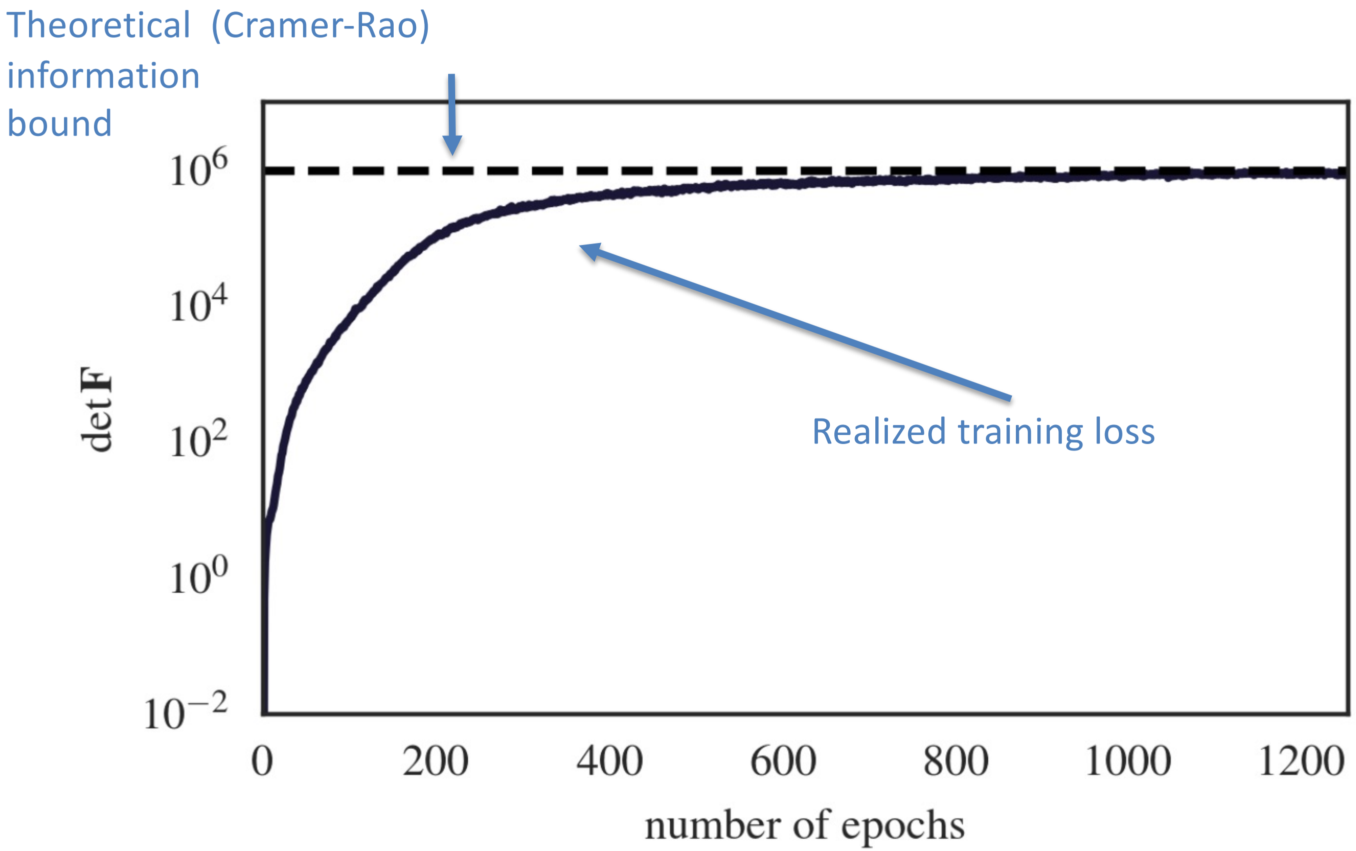
IMNN posterior

IMNN PML estimator

Exact likelihood

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

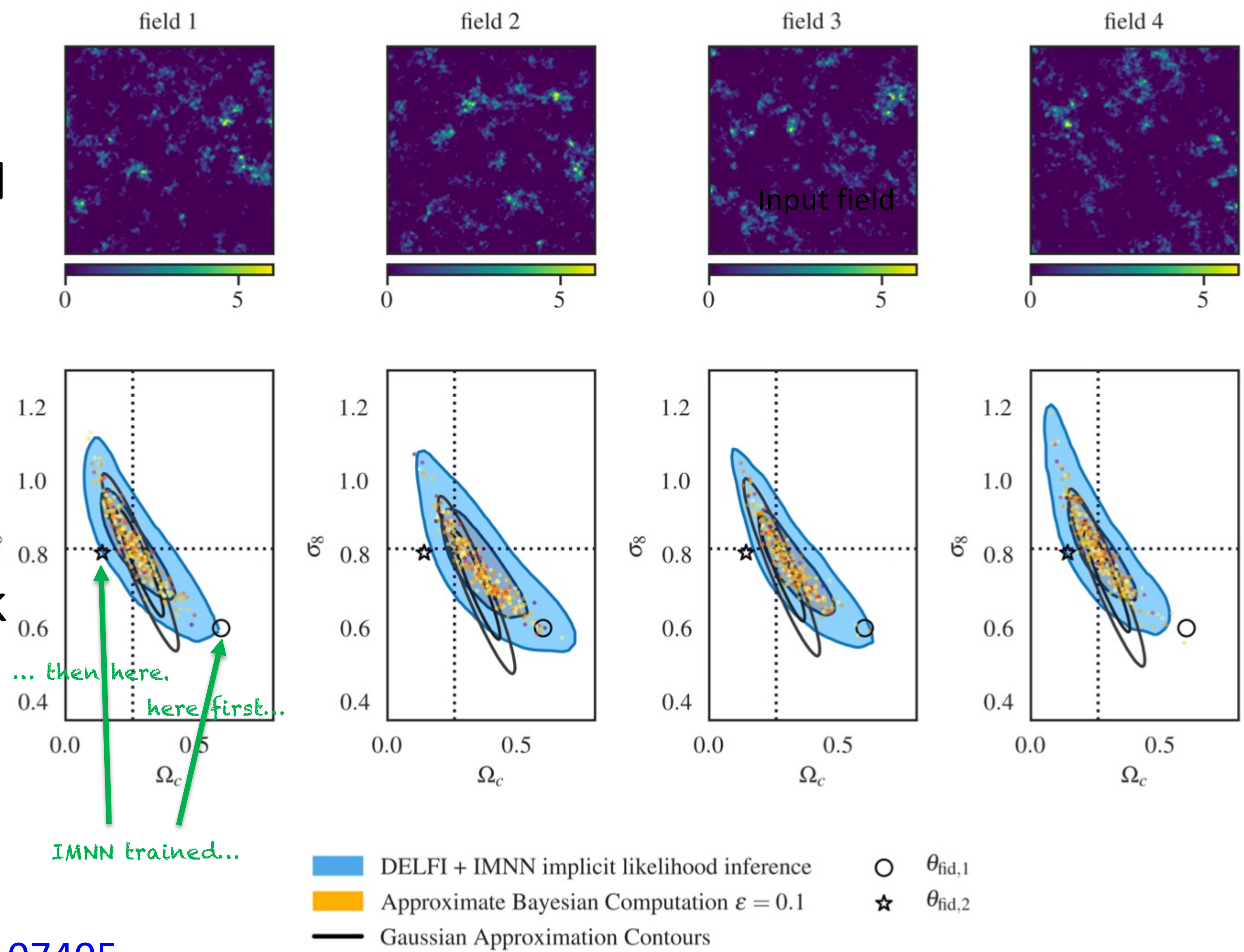
The IMNN recovers the full information



11 minutes on 1 GPU

Makinen et al., arXiv:2107.07405

Non-Gaussian field inference with IMNN and DELFI



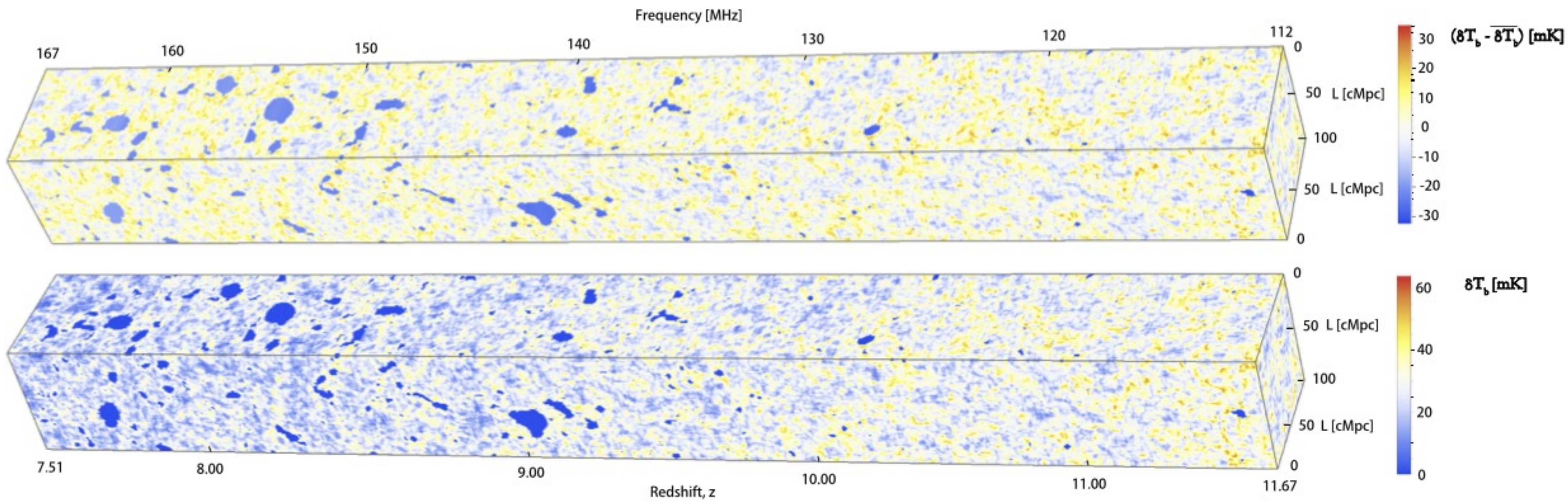
Available as interactive notebook tutorial at https://bit.ly/imnn-cosmo

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

Field-Based, Implicit Likelihood Inference with squared-loss trained Regression Network (DELFI 3D-CNN)

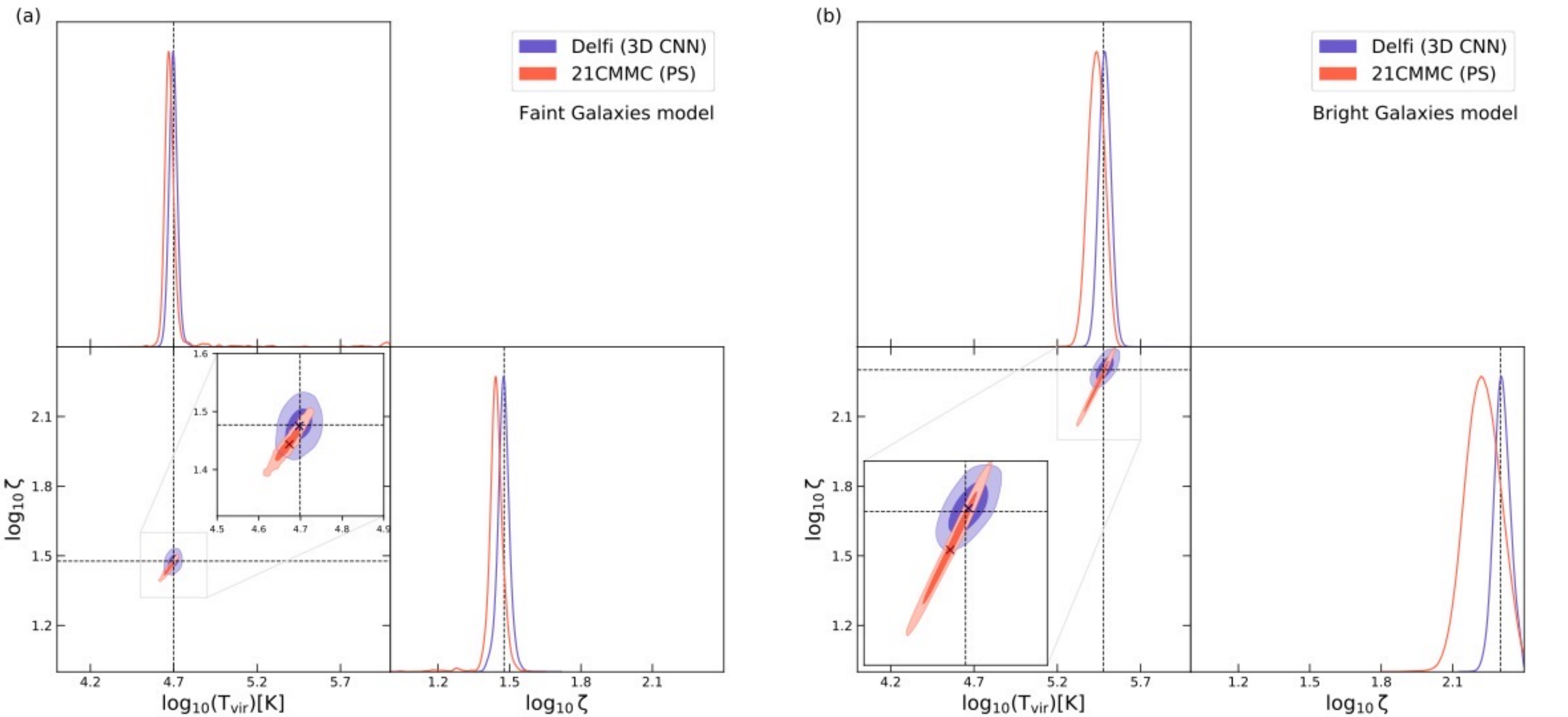
Benjamin Wandelt

Implicit Likelihood Inference to Infer Reionization Parameters from 21cm Light Cones



Zhao, Mao, Cheng, Wandelt arXiv:2105.03344

Implicit Likelihood Inference to Infer Reionization Parameters from 21cm Light Cones



Zhao, Mao, Cheng, Wandelt arXiv:2105.03344

Sounds complicated…
can we go straight to the answer?

SBI WITH MOMENT AND POSTERIOR MARGINAL NETWORKS

Main idea: skip compression step – go directly from data to posterior.

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

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

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

- **Marginal posterior networks**: obtain low-dimensional posterior marginals directly from data by minimizing Kullback-Leibler divergence

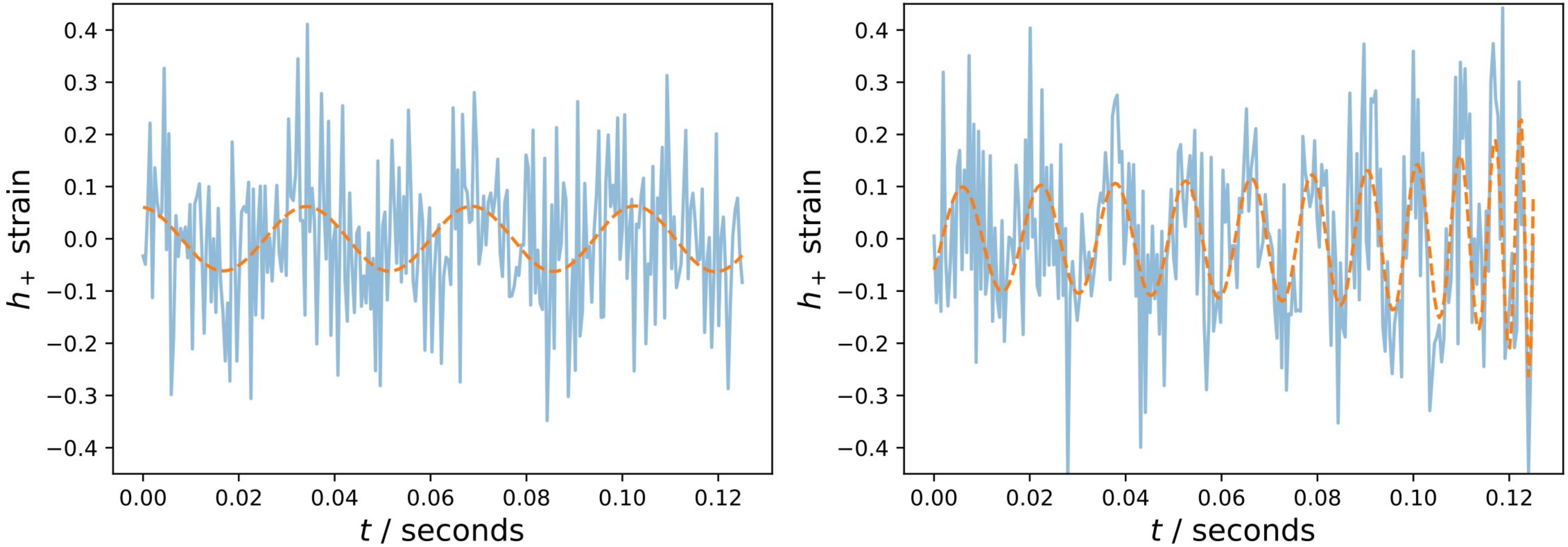
$$\int \ln q(\theta_i, \theta_j | d, w) p(d,\theta) d\theta \, dd$$

over network weights of a conditional neural density estimator *q.*

Solves curse of dimensionality **through a combination of direct neural estimates of posterior moments and low-dimensional posterior marginals.**

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

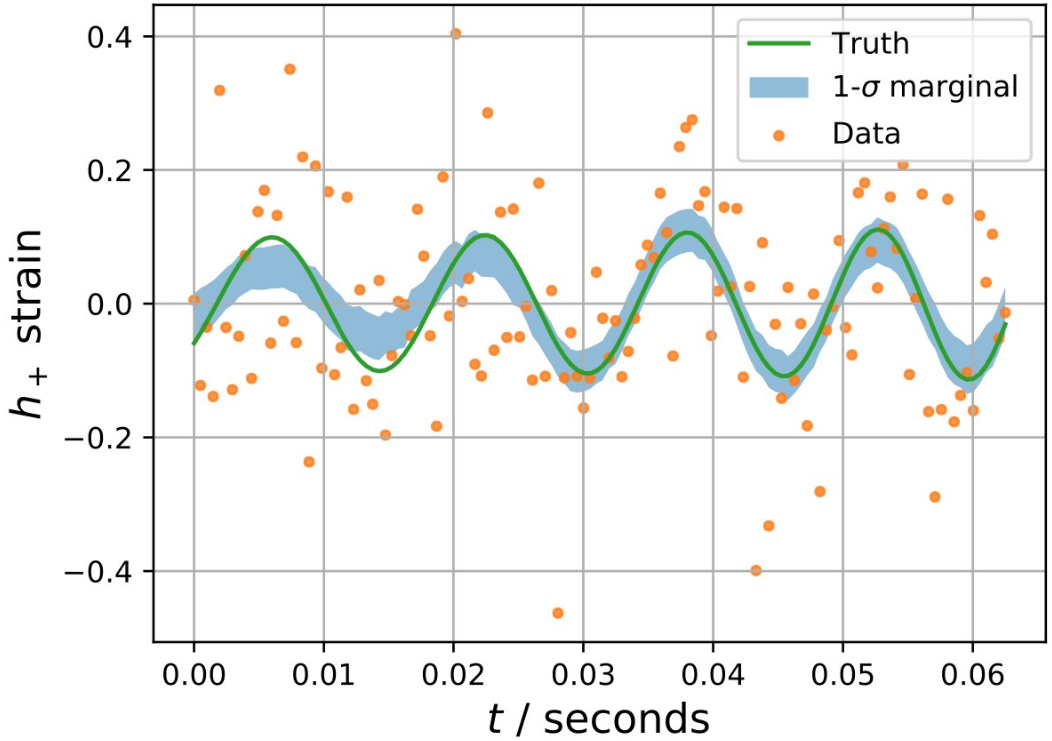
Example: Inference from BBH Mergers



BBH merger simulations, LIGO noise

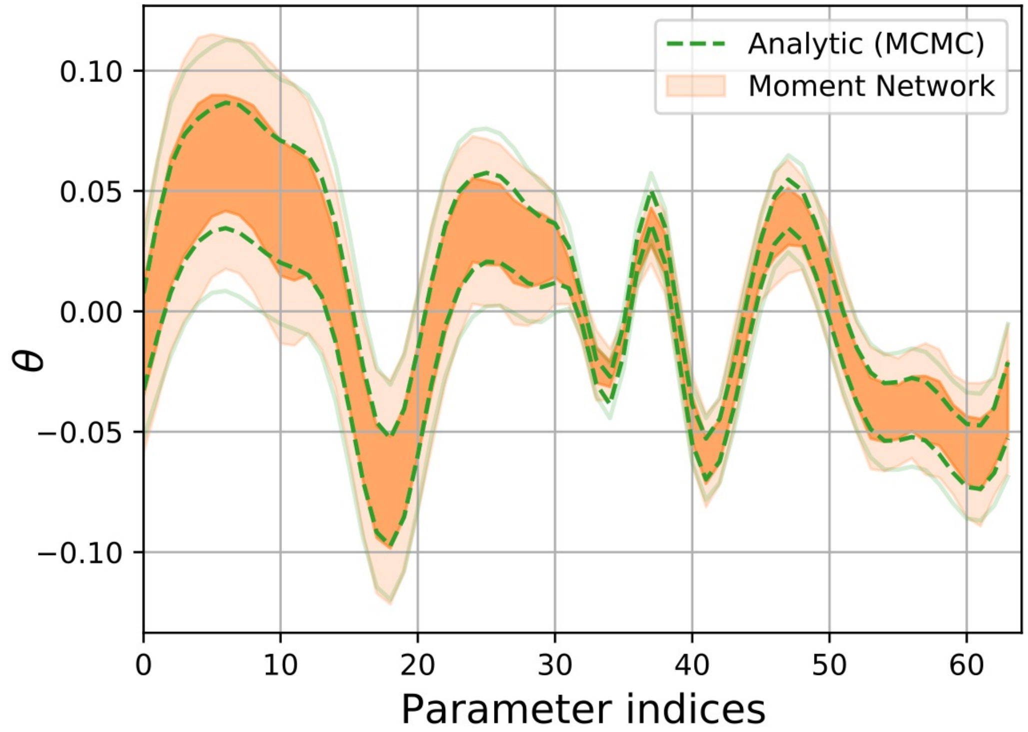
Jeffrey & Wandelt, arXiv:2011.05991

Signal Reconstruction from BBH Mergers with Moment Networks



Simulation-Based Inference (SBI) reconstruction of BBH merger simulations, using simulated LIGO noise

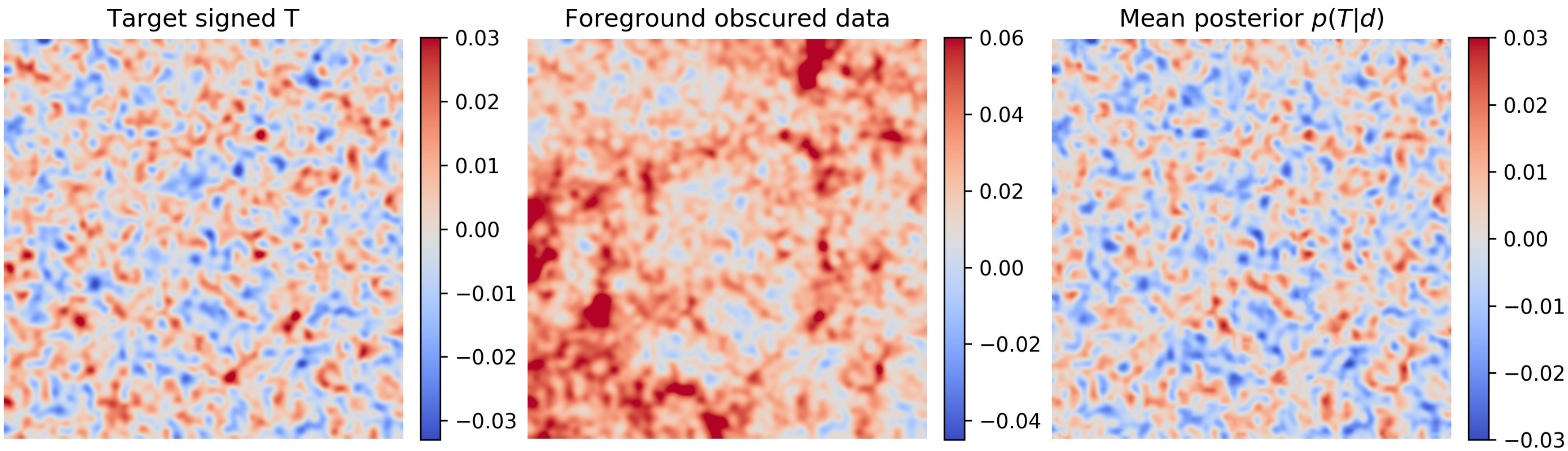
O(100) parameters!



Validation

Jeffrey & Wandelt, arXiv:2011.05991

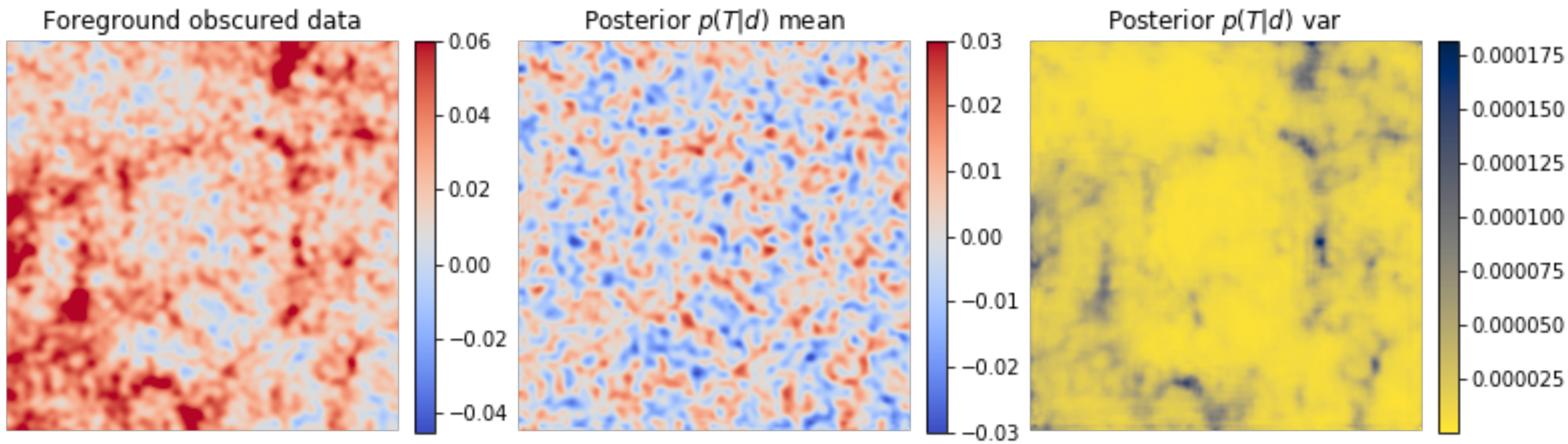
CMB Foreground Cleaning with Moment Networks



~10^5 parameters!

Jeffrey & Wandelt, arXiv:2011.05991

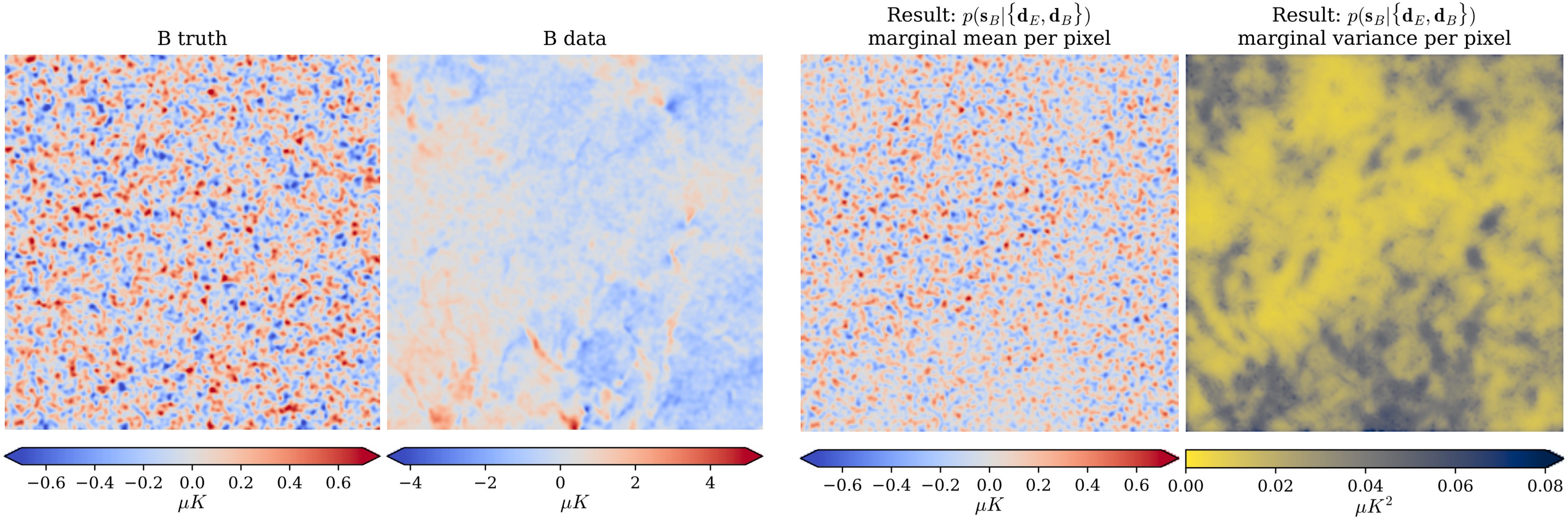
CMB Foreground Cleaning with Moment Networks



Moment Network computes Bayesian posterior means and variances for ~10^5 parameters

Jeffrey & Wandelt, arXiv:2011.05991

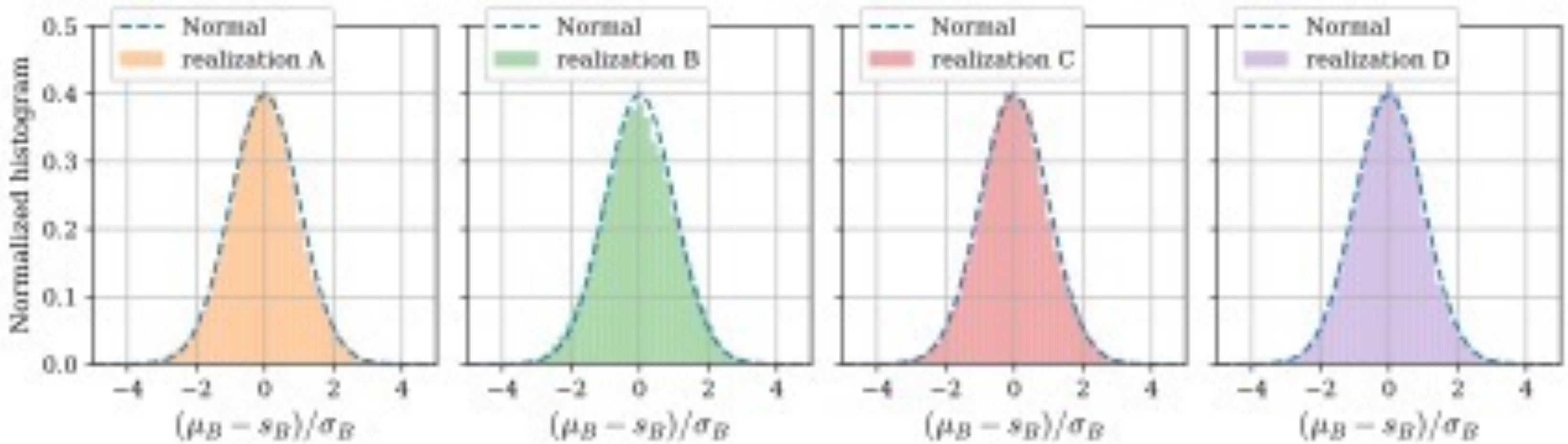
Moment networks: non-Gaussian B mode inference



Trained using a single training image at a single frequency and generative model based on Wavelet Phase Harmonics

Jeffrey, Boulanger, Wandelt, Regaldo-Saint Blancard, Allys, Levrier 2021, submitted.
(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, submitted

Approaching the full cosmological inference problem

CAMELS

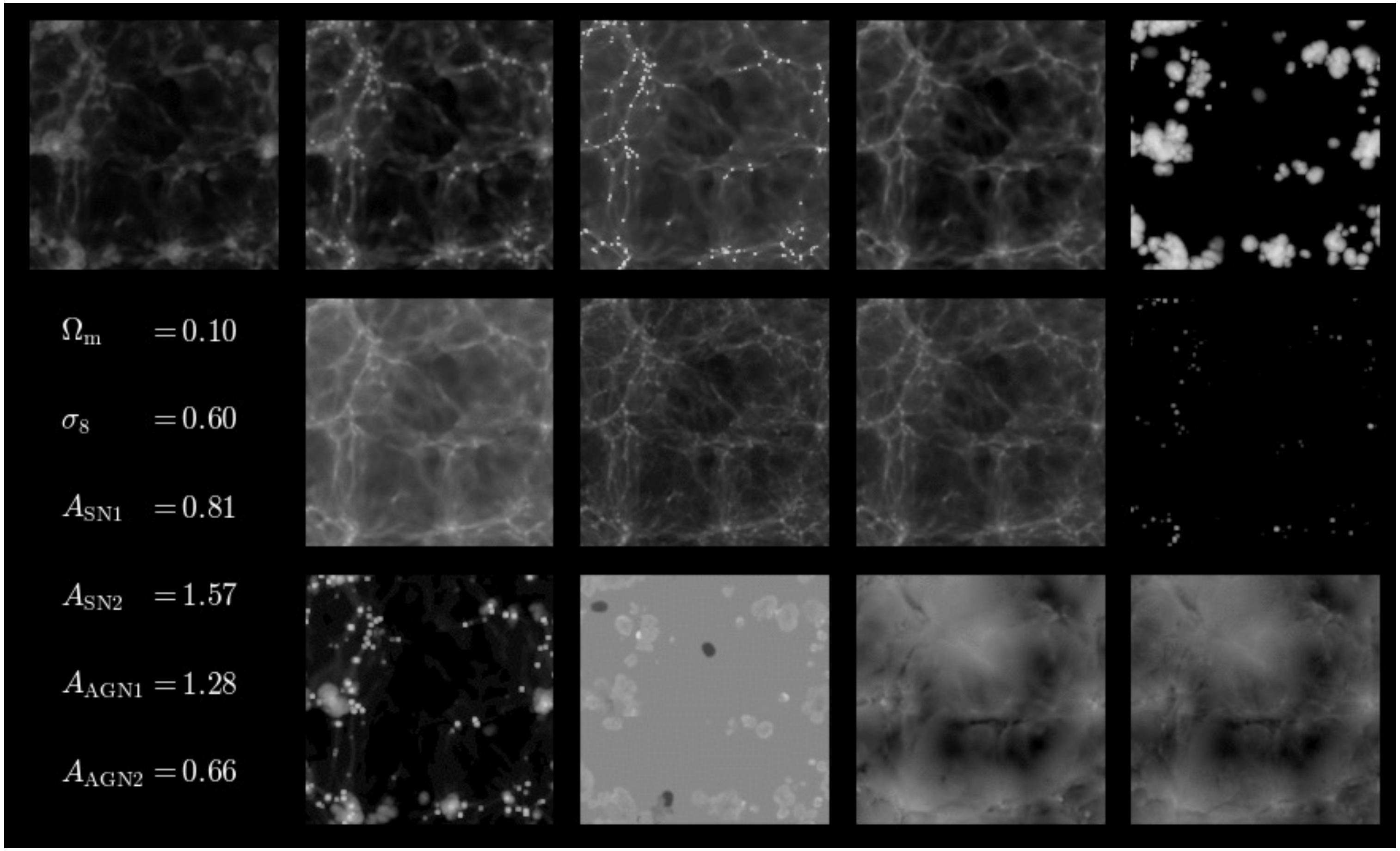
Cosmology and Astrophysics with Machine Learning

Collaborative project to generate large suites of full, cosmological hydrosimulations as a function of cosmological parameters and astrophysics models with two different codes (AREPO/Illustris & GIZMO/SIMBA).

Use to train and validate machine learning surrogates, and likelihood-free, simulation-based inference.

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

CAMELS Multifield Dataset (CMD)
The MNIST for cosmology?



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

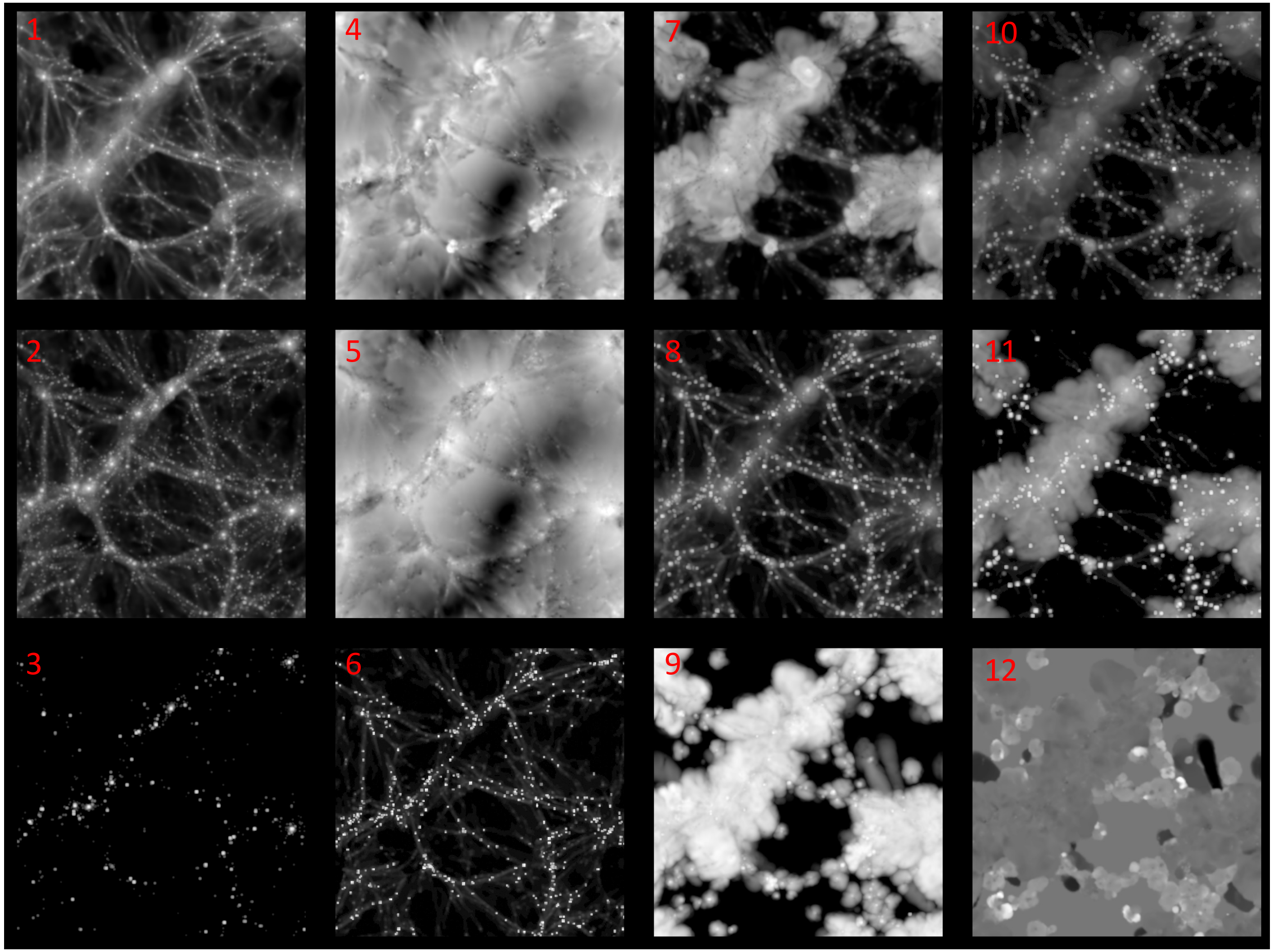
Cosmology on small scales with baryons

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



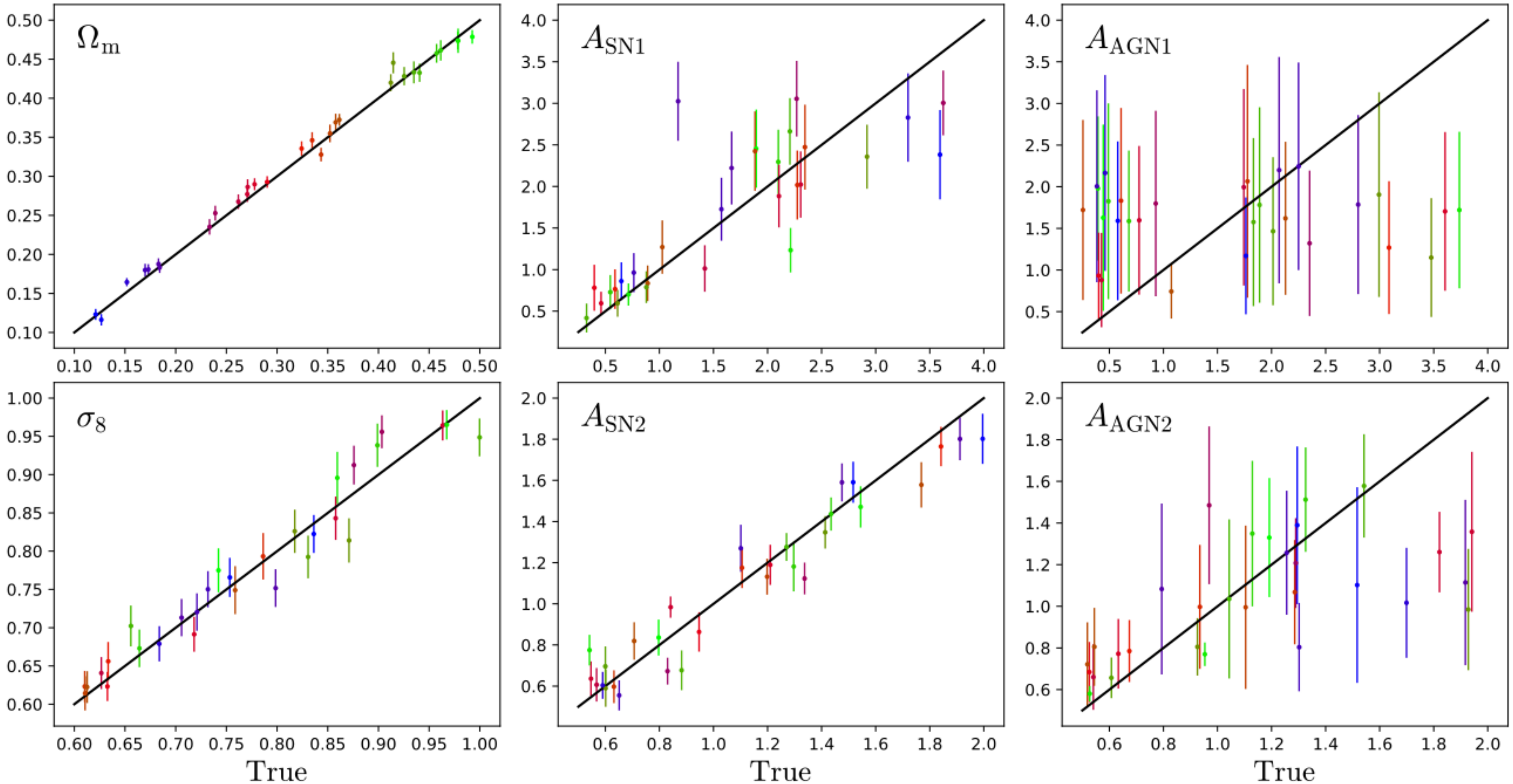
SBI: Cosmology from Small-Scale Hydro

Villaescusa-Navarro, arXiv:2109.09747

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

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

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



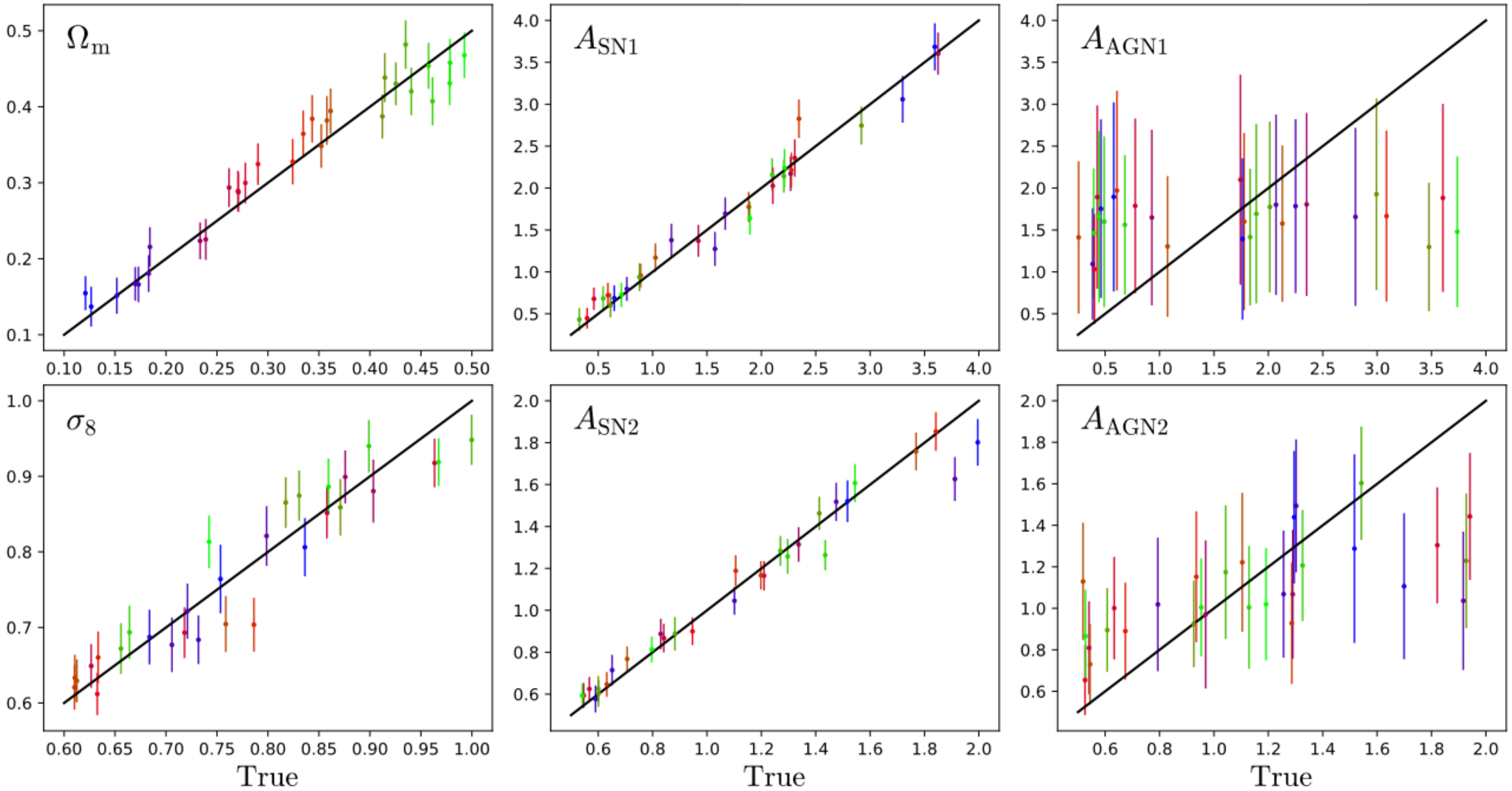
SBI: Cosmology from Small-Scale Hydro

Villaescusa-Navarro, arXiv:2109.09747

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

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

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

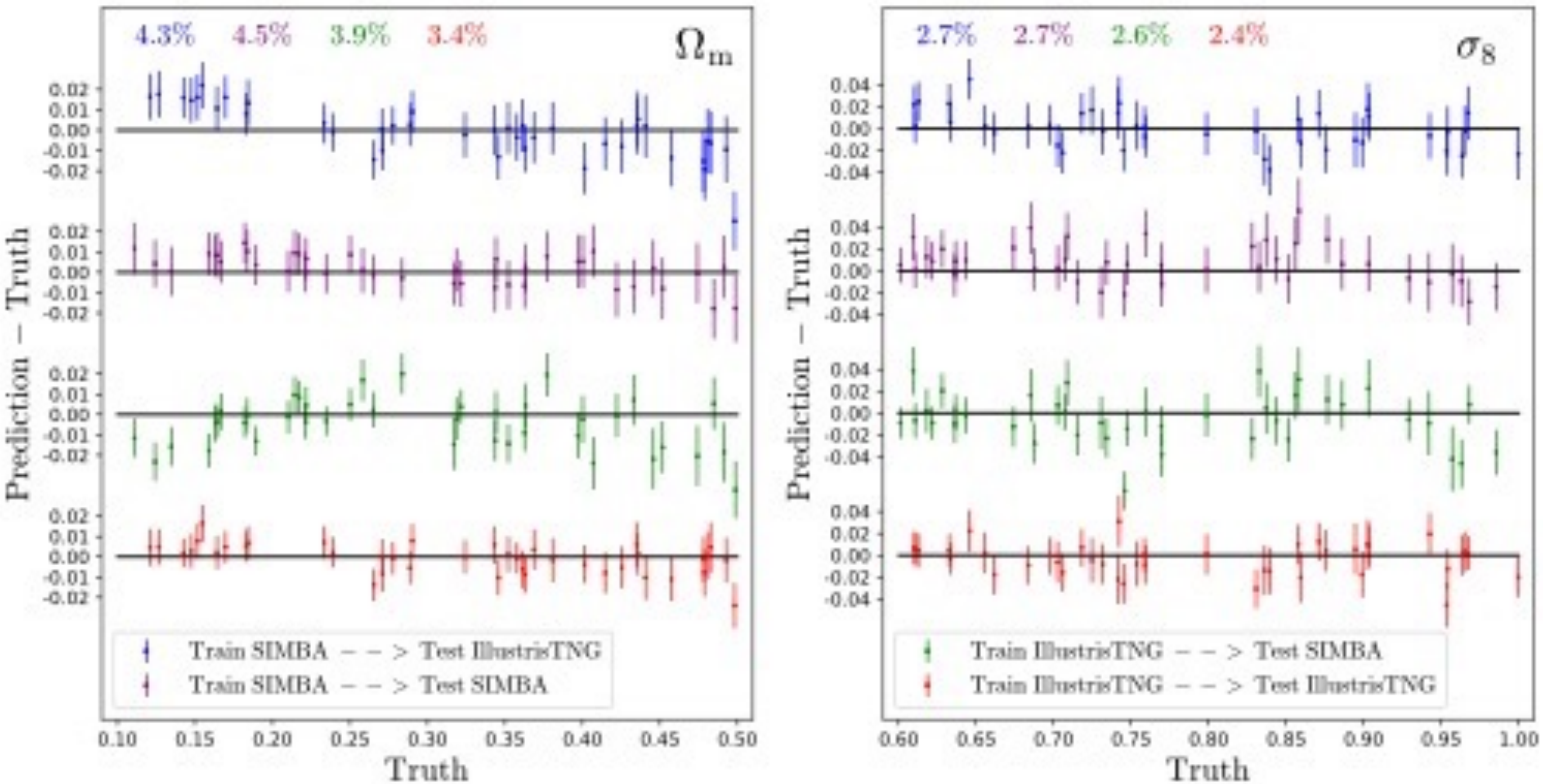


First results from cosmological AI on CAMELS Multifield Data set

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!)

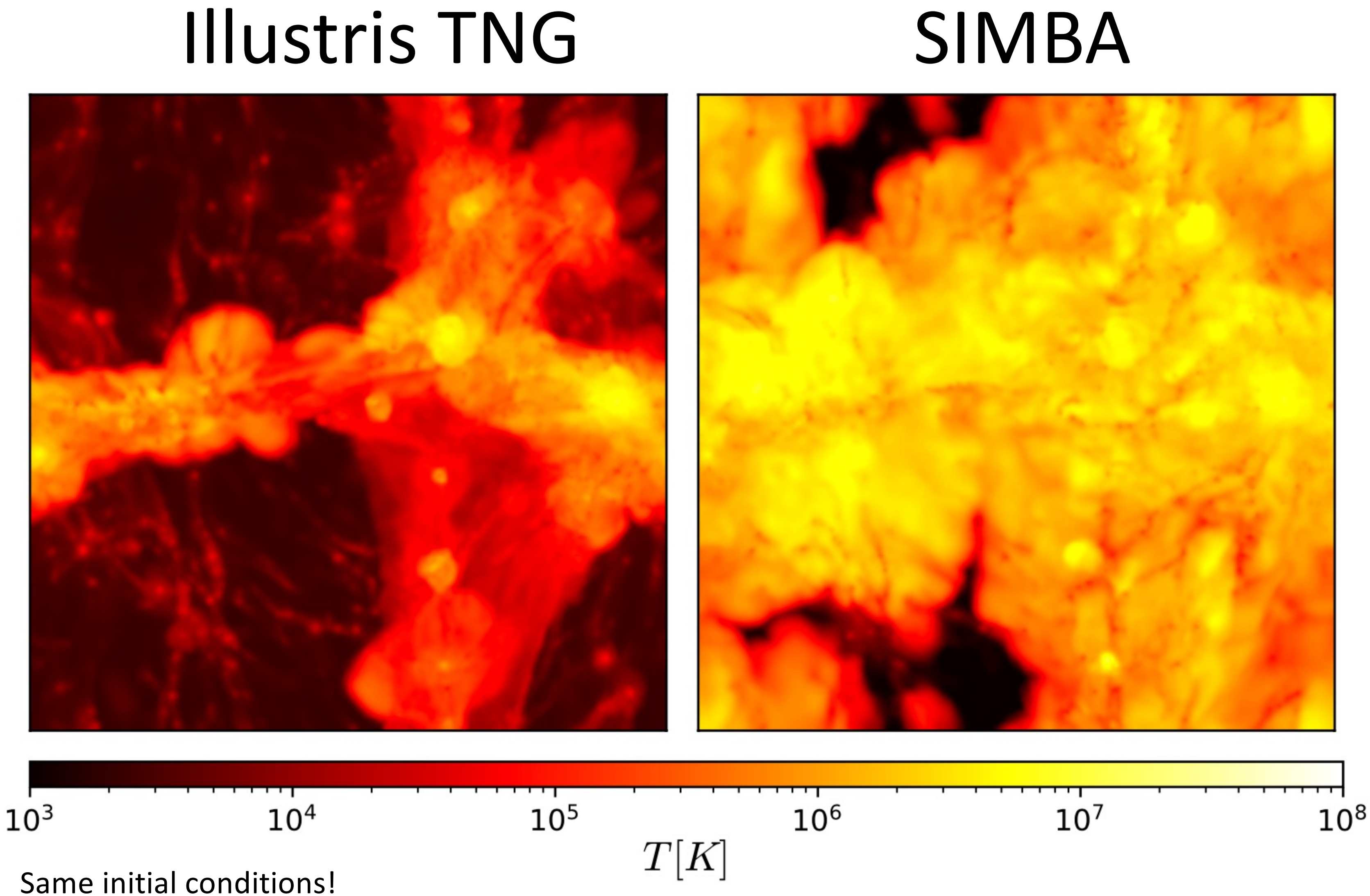
Villaescusa-Navarro et al., arXiv:2109.09747, arXiv:2109.10360

Cosmology robust to baryonic physics



Villaescusa-Navarro et al., arXiv:2109.10360

Villaescusa-Navarro et al., arXiv:2109.09747



Same initial conditions!

Meeting the Theory and Simulation Challenges in the Age of Implicit Likelihood Inference

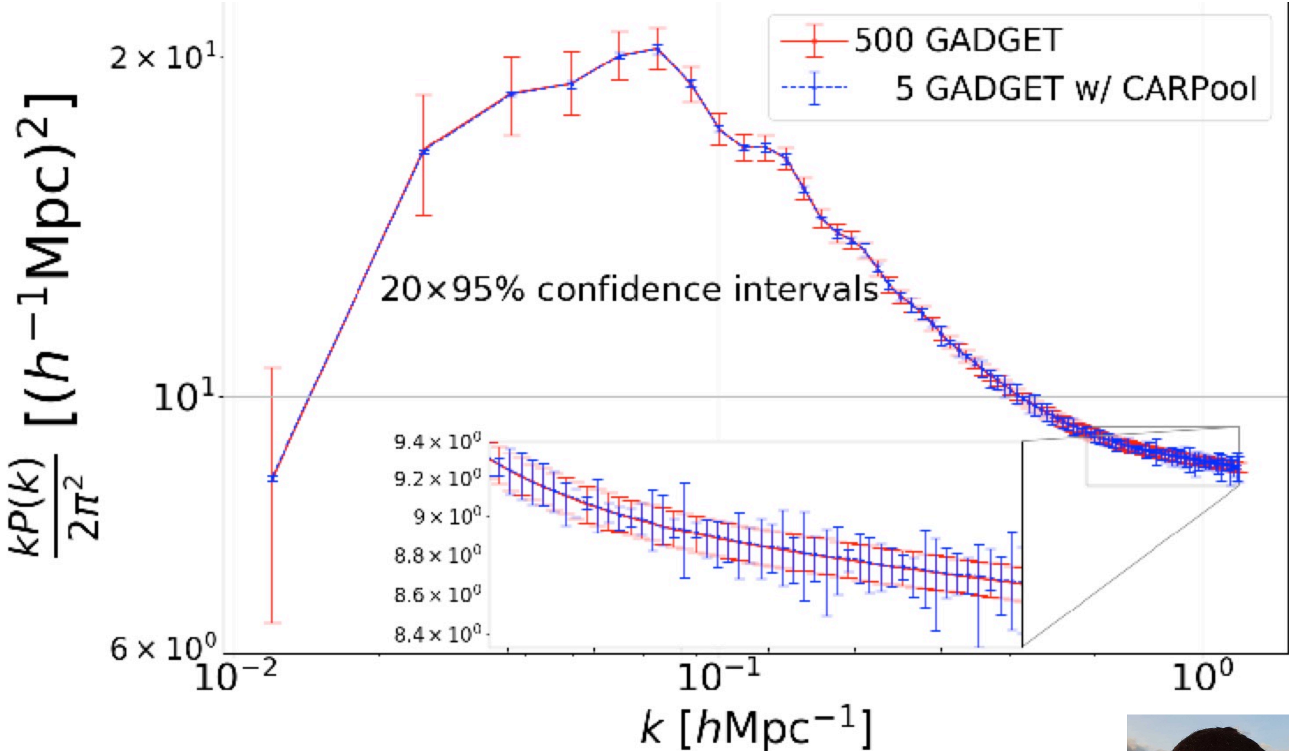
Using recent advances in computational physics, stats, and machine learning

Benjamin Wandelt

N. Chartier et al: CARPool reduces the number of needed simulations by orders of magnitude

Convergence
Acceleration by
Regression and
Pooling

uses fast, approximate surrogates to give **unbiased, low-variance** estimates of full simulation results.

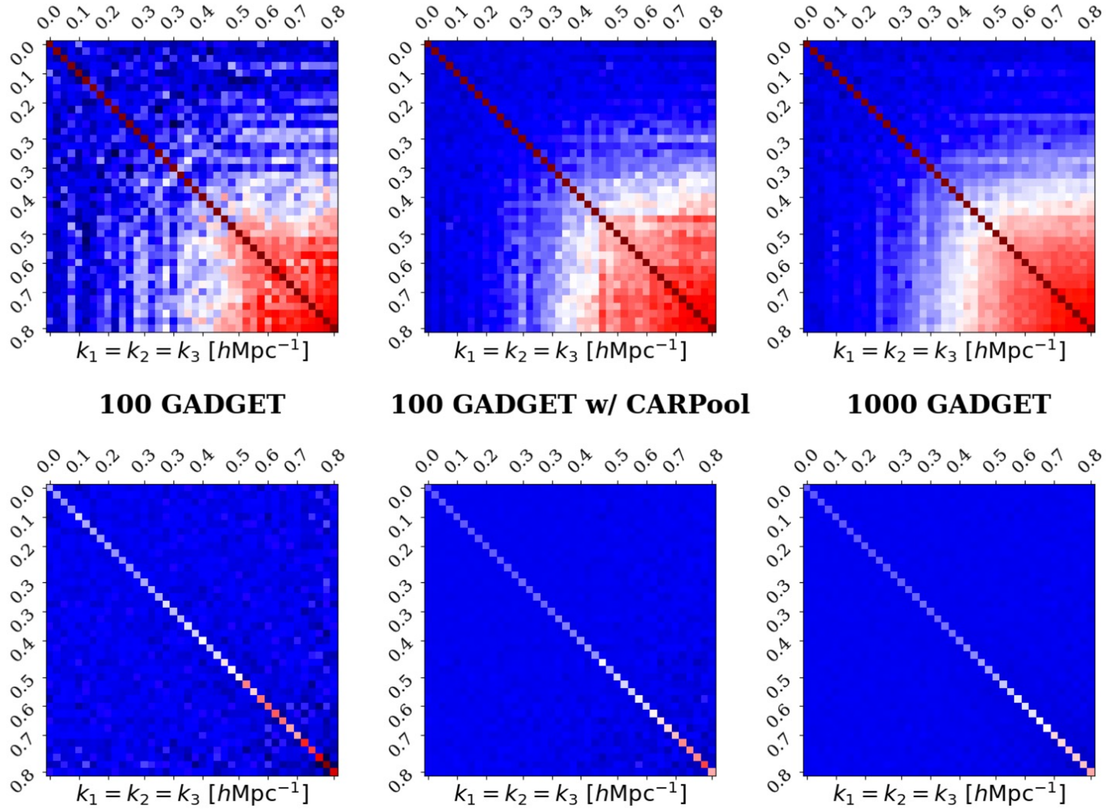


N. Chartier et al, **arXiv:2009.08970**

N. Chartier et al: CARPool Covariance reduces the number of simulations by orders of magnitude

Covariance matrices and inverses

10 fold reduction in number of simulations for comparable accuracy



N. Chartier et al, arXiv:2106.11718

CARPool as a Non-perturbative, Statistical Approach to "Perturbation Theory"

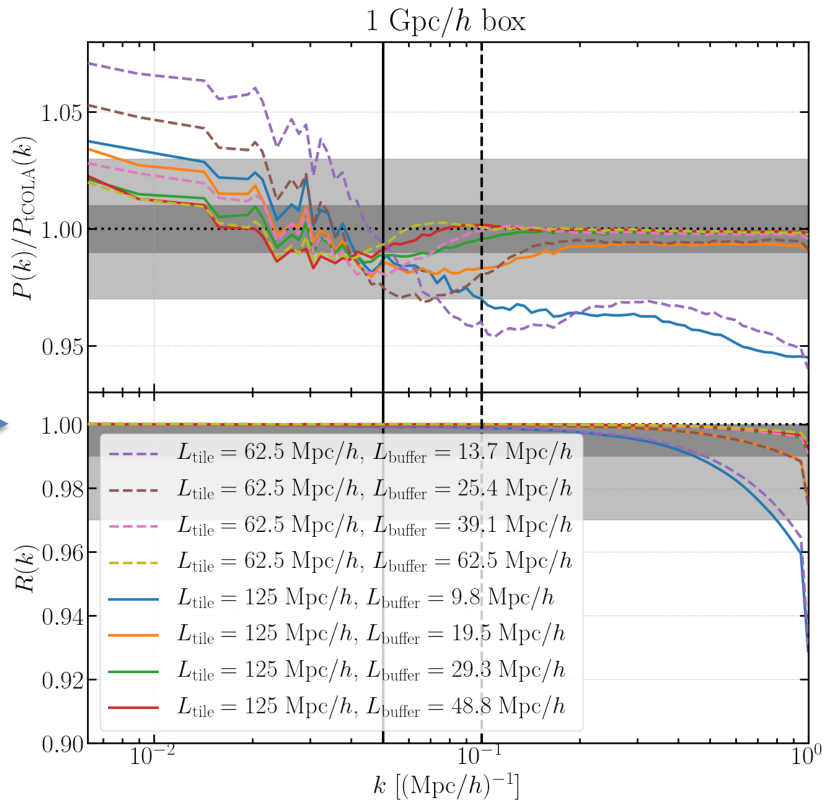
- Take existing set of numerical simulations for Model A (e.g. Quijote for LCDM)
- For Model B, include new physics (e.g. Modified Gravity,...)
- Using Model A solutions as "surrogates" and apply **CARPool:**
  - Run a *few* simulations for Model B that are *correlated* with existing set (e.g. same initial conditions)
  - Use Model A solutions to subtract statistical fluctuations.

- Result: precision expectation values and covariances for the new model with only a handful of simulations

N. Chartier, et al, **arXiv:2009.08970**

Leclercq et al: Simbelmynë: Perfectly Parallel n-body sims.

Opens up new ways to do larger and deeper n-body sims on a broad range of computational architectures



<1% error on P(k) small scales

100% correlation with serial simulation

Leclercq et al: **arXiv:2003.04925**

F. Villaescusa-Navarro et al.: The QUIJOTE simulations
to train machine learning surrogates

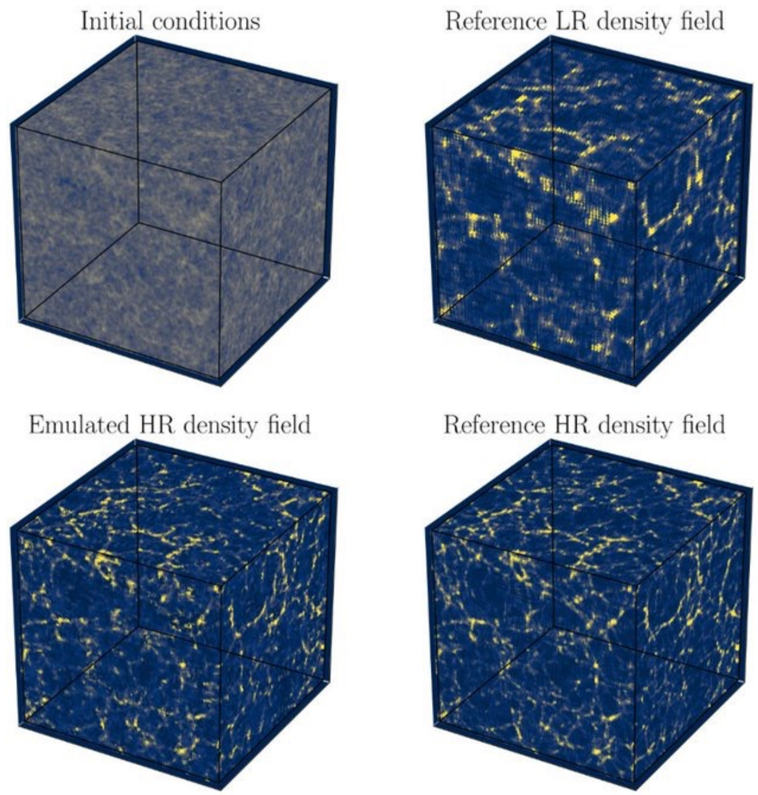
- Largest release of N-body simulation data to date
  - 43,100 full GADGET 3 simulations (1 Gpc)^3, 512^3 or 1024^3 particles
  - ~1 PB of data
- Goal: quantify statistics information content of non-Gaussian non-linear density field about cosmological parameters
- Includes full dark matter snapshots, halo and void catalogues, and many pre-computed statistics.

Excellent tool for training machine learning surrogates.

Villaescusa-Navarro et al, **arXiv:1909.05273**

Kodi Ramanah et al: **neural super-resolution** of n-body simulations

Uses a Wasserstein-GAN to generate high-resolution n-body output from low-res result and high-res initial conditions with ~ 1% accuracy.



Kodi Ramanah et al, **arXiv:2001.05519**

Conclusions

- Will be awash in data. Many advances in cosmology hinge on solving the cosmological inference problem. Let's solve it!
- We now have a tool set to attack this problem based on advances in physics, stats, and machine/deep learning
  - Full physical forward model inference such as BORG
  - Neural physical engine layer
  - Massive data compression (IMNN)
  - Likelihood-free, simulation-based inference (DELFI, Moment Networks)

- New approaches to solve the simulation problem, e.g
  - Perfectly parallel sims with Simbelmynë
  - Reduction of number of sims with CARPool variance reduction
  - High-performance neural surrogates

Codes and Data

BORG and related projects: aquila-consortium.org

IMNN: bitbucket.org/tomcharnock/imnn/

DELFI: github.com/justinalsing/pydelfi

The Quijote Simulations: github.com/franciscovillaescusa/Quijote-simulations

The Camels Simulations: camel-simulations.org

Simbelmynë perfectly parallel n-body code: simbelmyne.florent-leclercq.eu