Statistical cosmology

Benjamin Wandelt

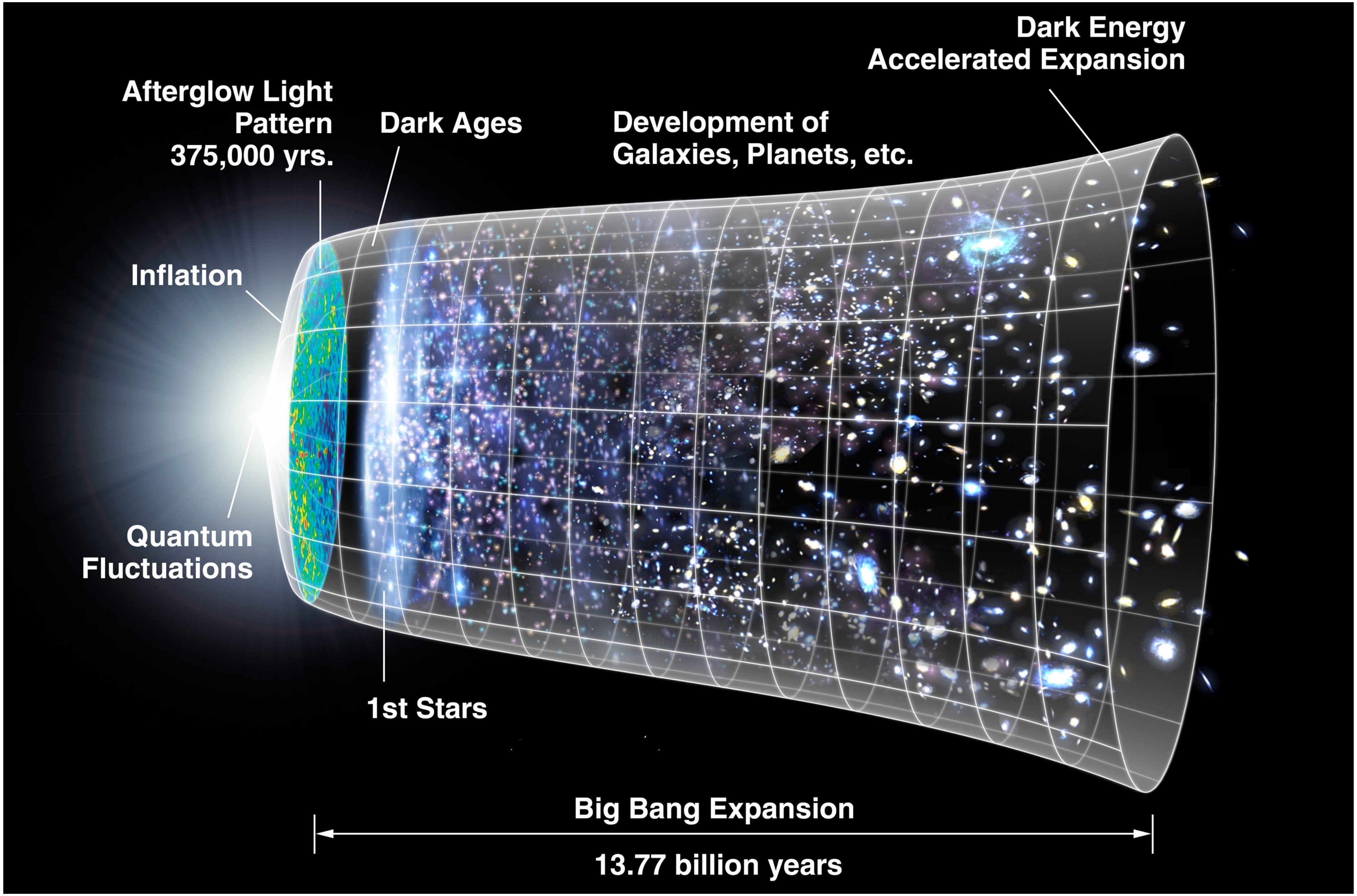
Lectures at the 2023 Summer School in Cargese

Reading

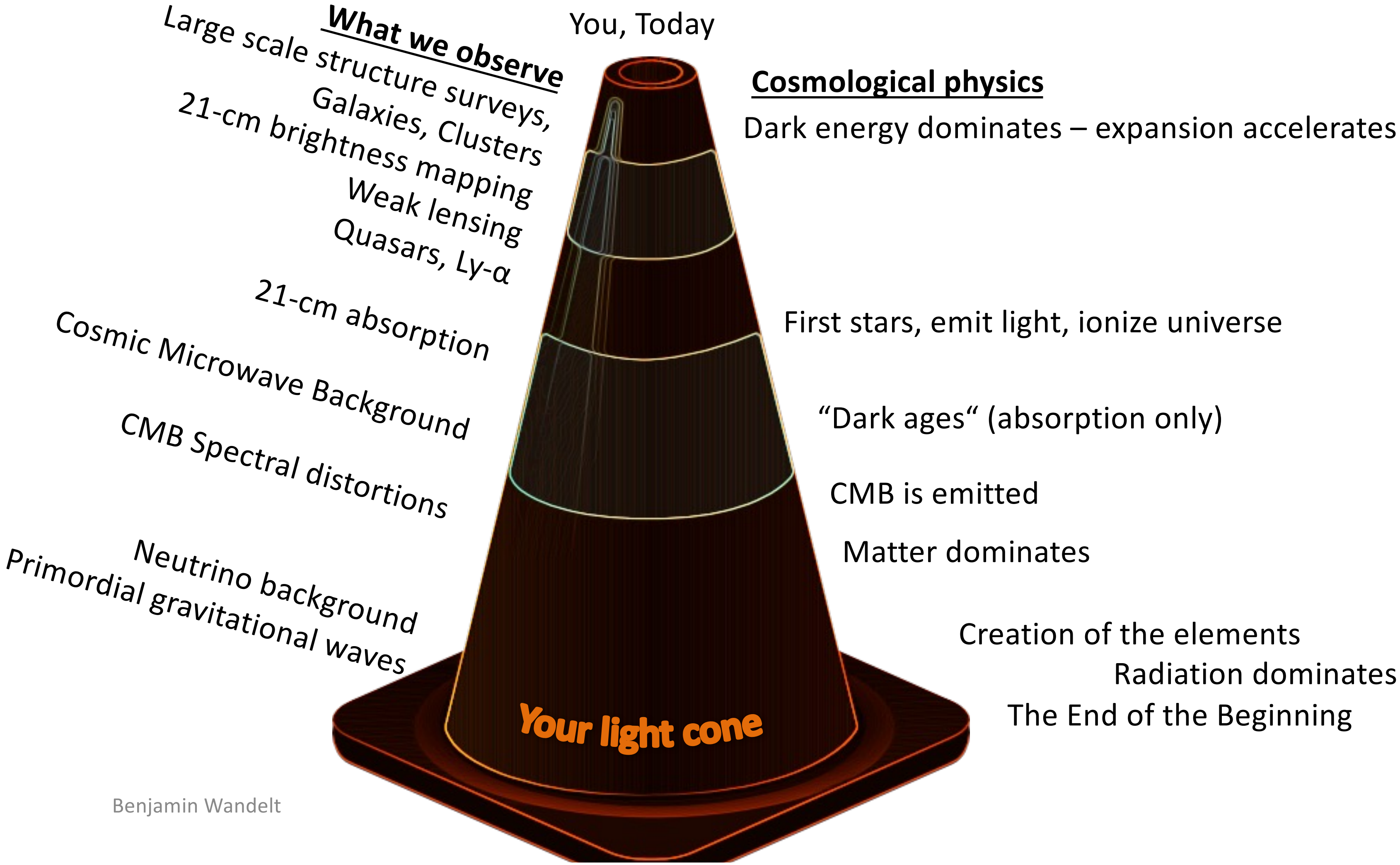
- Leclercq, Pisani & Wandelt 2014, *Cosmology: from theory to data, from data to theory*, arXiv:1403.1260
- Alsing & Wandelt 2017, *Generalized massive optimal data compression*, arXiv:1712.00012

What we want to know

- How did the Universe begin?
- How did structure appear in the Universe?
- How did it evolve until today?
- What is the Universe made of?
- What are the properties of dark matter?
- What are the properties of dark energy?
- What is the geometry and the symmetry of the Universe?

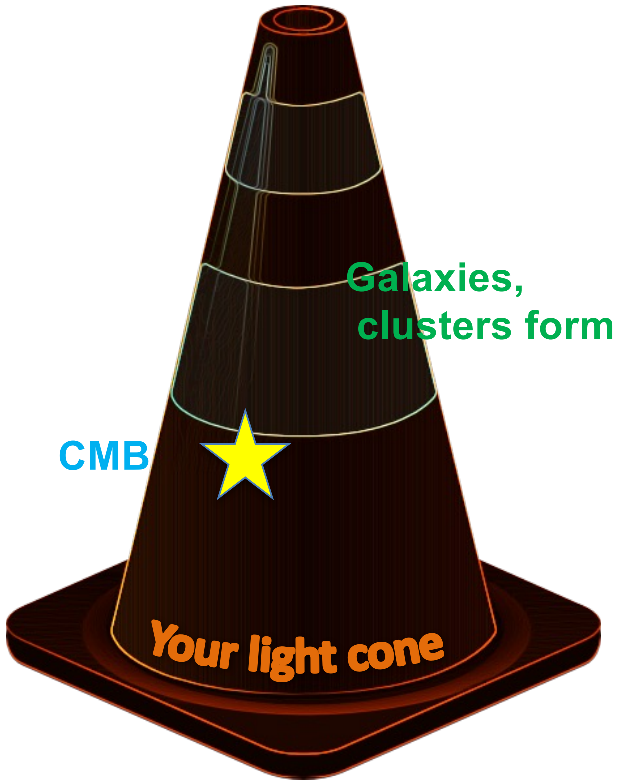

Dark Energy
Accelerated Expansion
Afterglow Light
Pattern
375,000 yrs.
Dark Ages
Development of
Galaxies, Planets, etc.
Inflation
Quantum
Fluctuations
1st Stars
Big Bang Expansion
13.77 billion years

The playground of cosmological physics



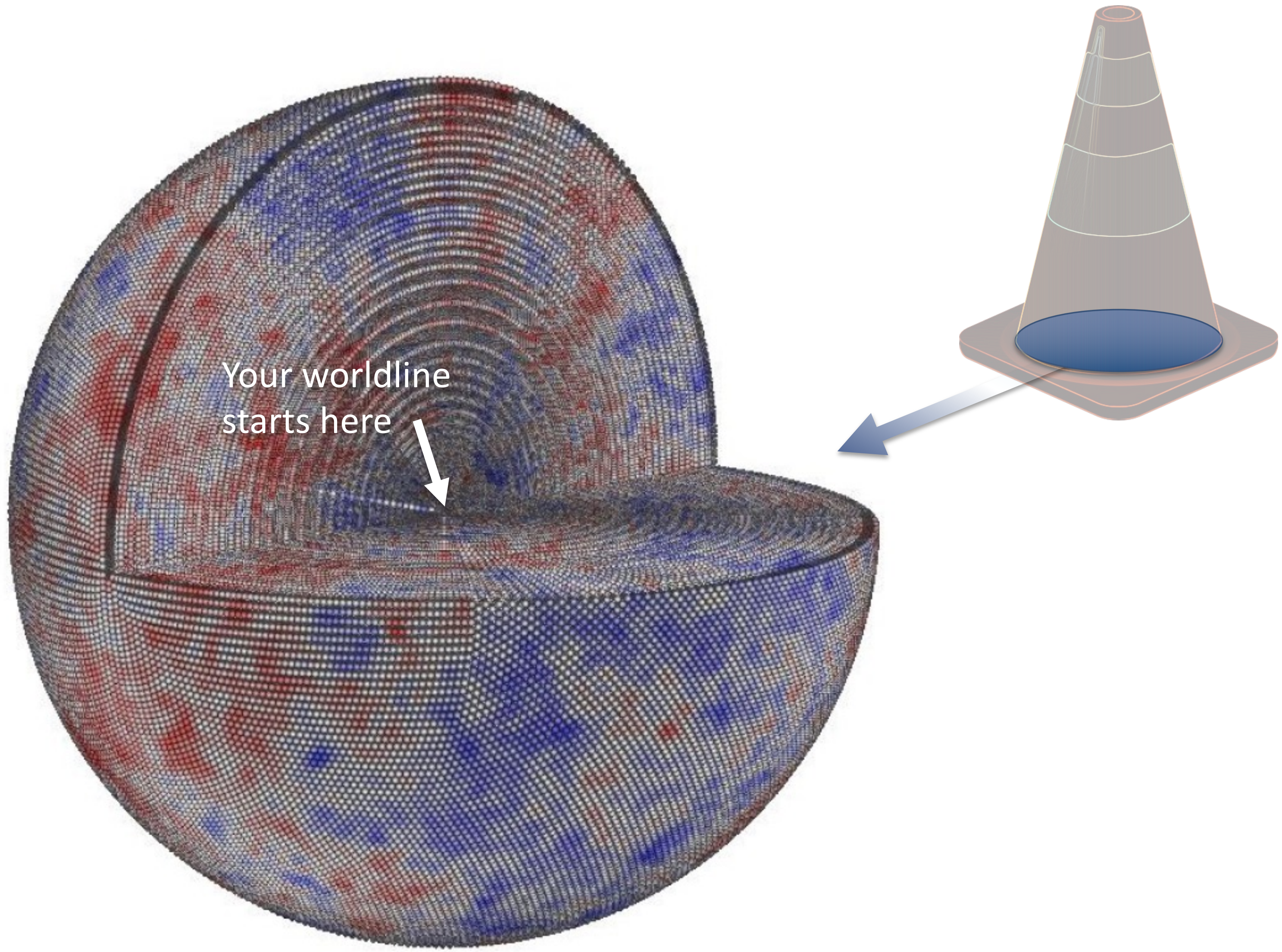
Benjamin Wandelt


A Journey of Light
through Space and Time


You, Today
Galaxies,
clusters form
CMB
Your light cone

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

(curvature perturbations)



Benjamin Wandelt

What cosmologists want to learn

Inflation non-Gaussianity

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

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

Statistics of the “initial conditions”

Recipe for our universe

Dynamics

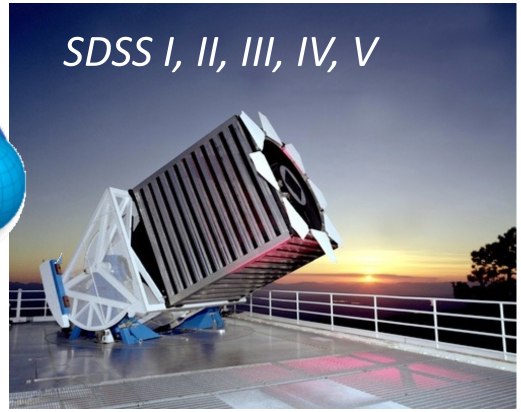
Expansion geometry

Surveys are sampling our past light cone exponentially fast!

PLANCK

EUCLID
Mapping the Geometry of the Dark Universe

SKA
Square Kilometre Array



MegaPrime
CFH
MegaCam
Legacy Survey
CFHTLS

THE DARK ENERGY SURVEY

HYPER SUPRIME-CAM

CMB-S4
Next Generation CMB Experiment

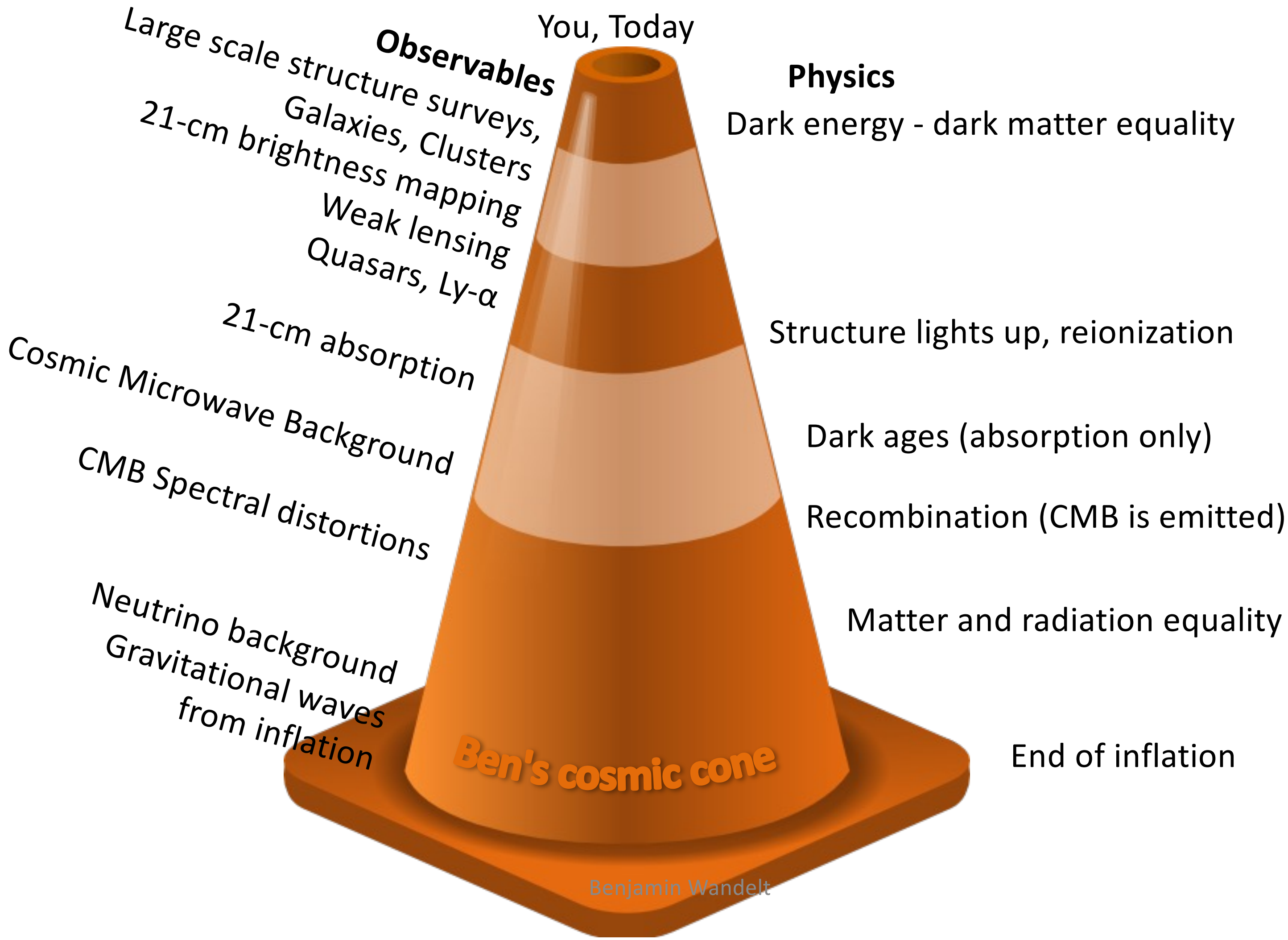
VERA C. RUBIN
OBSERVATORY

(Your favorite survey here)

Dark Energy Spectroscopic Instrument

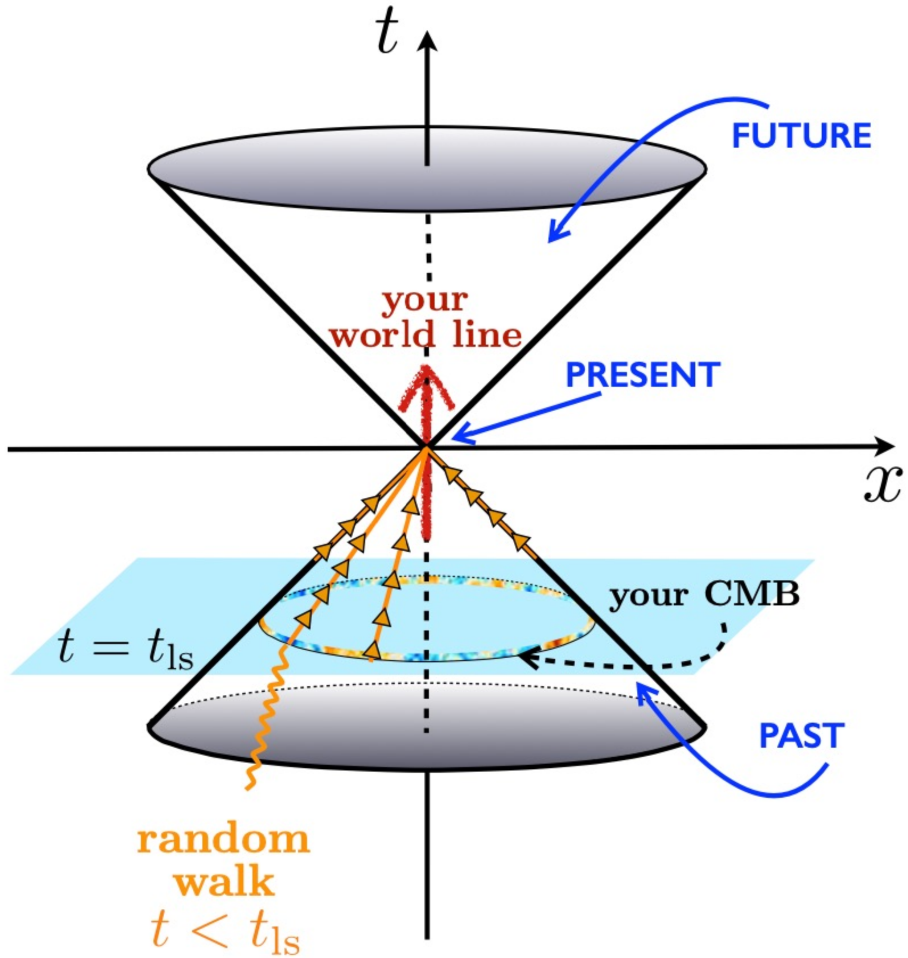
What is there to observe?
What about the universe is "knowable?"

Causal structure of your Universe



Some exercises with the causal structure of the Universe

- Draw a causal diagram.
- Where is your past? Mark it on the diagram.
- Define “the present.” Then show where that is on the diagram.
- Mark the future on the diagram.
- Draw in your entire existence (your world line from birth to death).



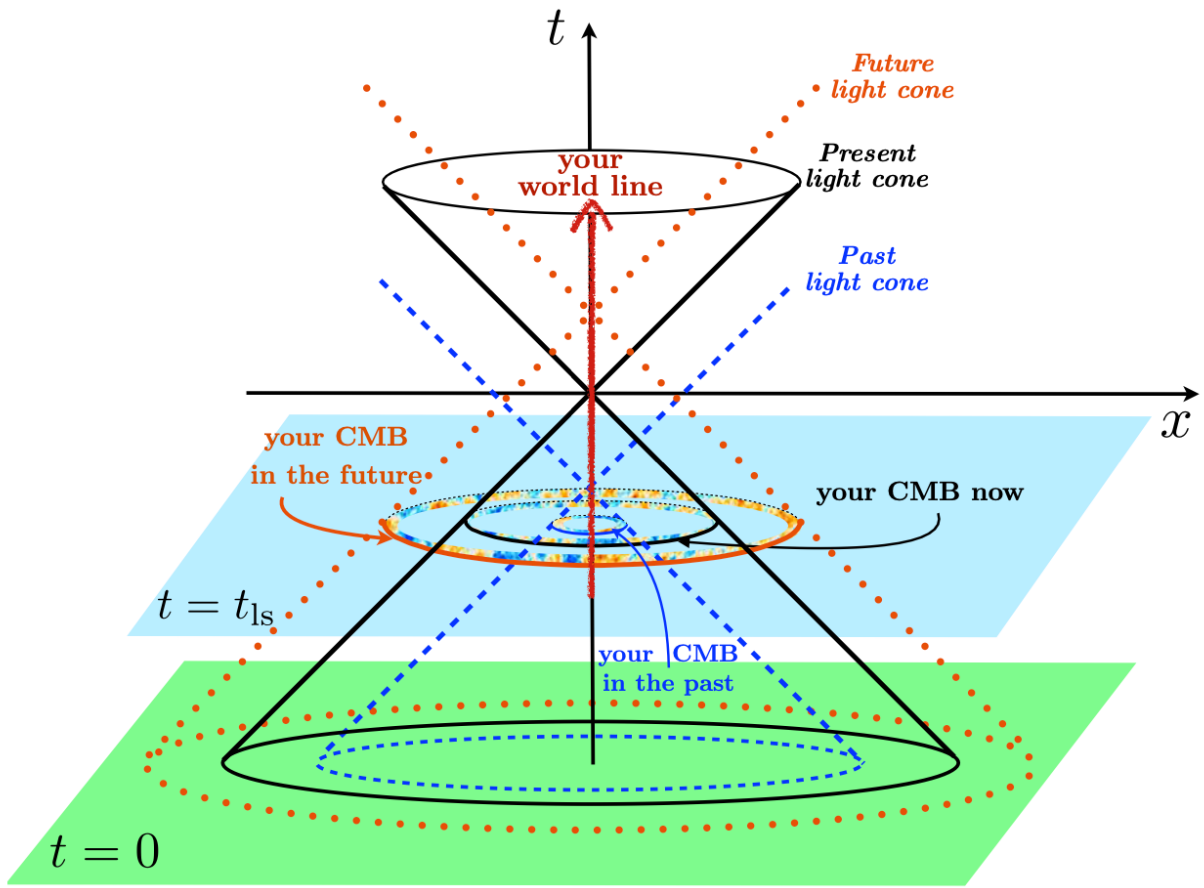
Benjamin Wandelt

Some more exercises

- Draw another causal diagram
- Define “seeing” as receiving all the photons that reach you NOW, in this moment, from everywhere in the Universe, from all directions and without obstructions. Draw the 3-volume of what you see right now on the diagram.
- Draw the 4-volume corresponding to everything you have ever seen and you will ever see.

More exercises

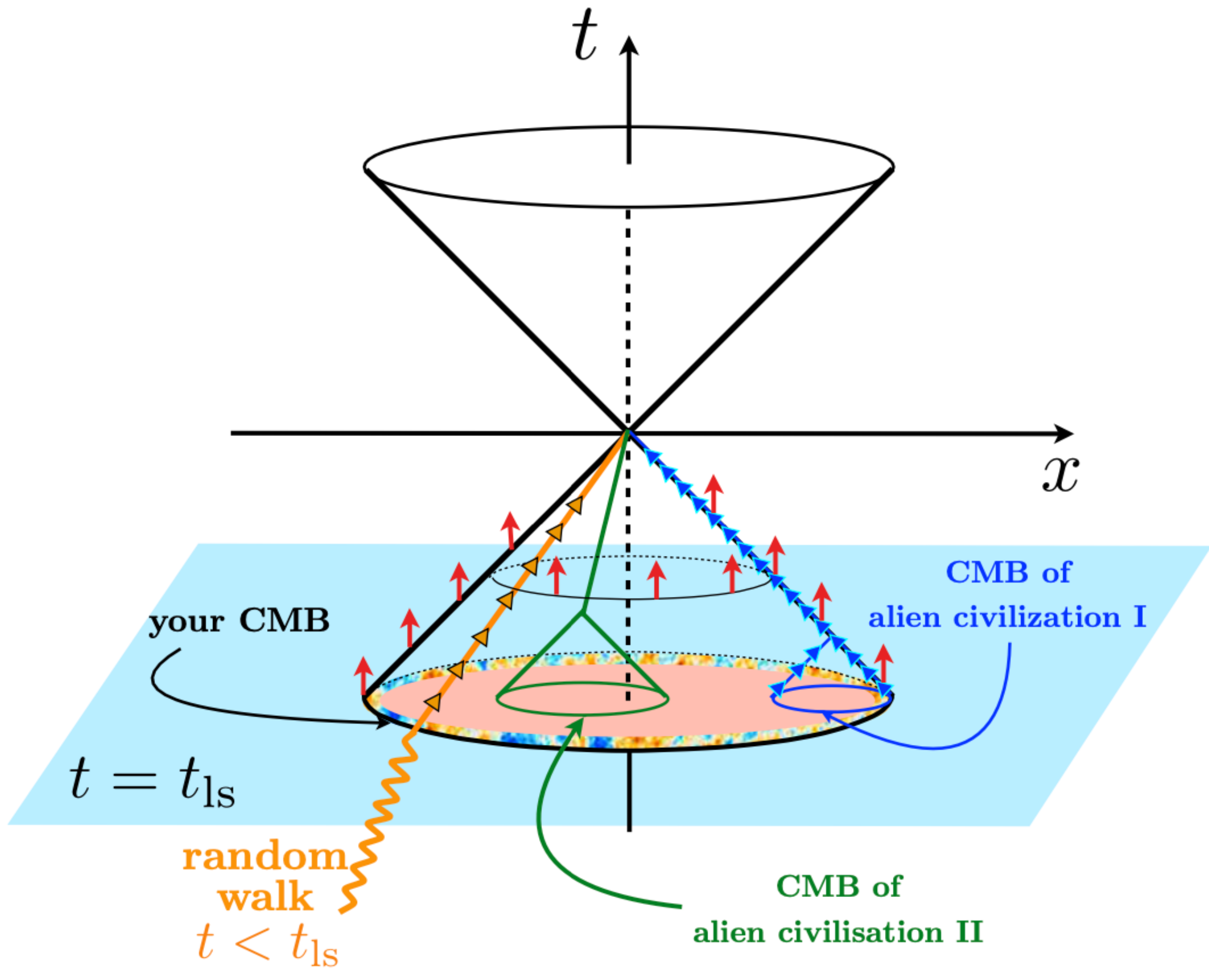
- Draw another causal diagram
- Draw the 4-volume of everything you can learn about indirectly from what you see right now.



Benjamin Wandelt

Indirect observation – friendly aliens

- Draw another causal diagram
- Draw the 2-sphere corresponding to the cosmic microwave background that you see (the surface of last scattering, SLS)
- On your diagram, illustrate how we could learn more about the cosmic microwave background if friendly aliens had observed it for us in the past and sent us radio transmissions through cosmic time.



Benjamin Wandelt

Indirect observation – time capsule

- A *time capsule* is a durable container kept in a safe place, which protects a message or objects intended for future generations.
- Can you think of cosmological observables that are like *time capsules*?

Causal diagrams give quick and correct answers to interesting questions

- Is the CMB really the ultimate way to observe the largest possible scales?
  - No. In principle, causality allows accessing the entire volume
- How is it possible to see “super-Horizon” scales in the CMB?
  - These are scales that *were* super-Horizon at the time the CMB was emitted
  - They are now sub-Horizon

How do these observable signals relate to what we want to know?

"Do not believe any observational result until it is confirmed by theory." (Eddington)

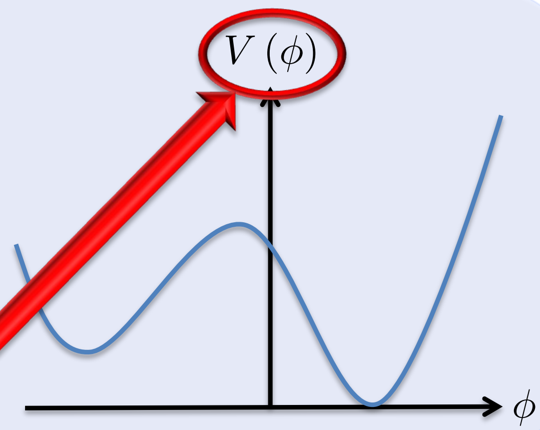
Cosmological Inflation

- Is a high energy phase of accelerated expansion in the early Universe $\ddot{a} > 0$
- Solves the Hot Big Bang horizon and flatness problem $\mathrm{d}s^2 = -\mathrm{d}t^2 + a^2(t)\,\mathrm{d}\vec{x}^2$
- Can be implemented with a single scalar field

$$S = -\int \mathrm{d}^4 x \sqrt{-g}\left[\frac{1}{2} g^{\mu\nu}\partial_\mu\phi\partial_\nu\phi + V(\phi)\right]$$

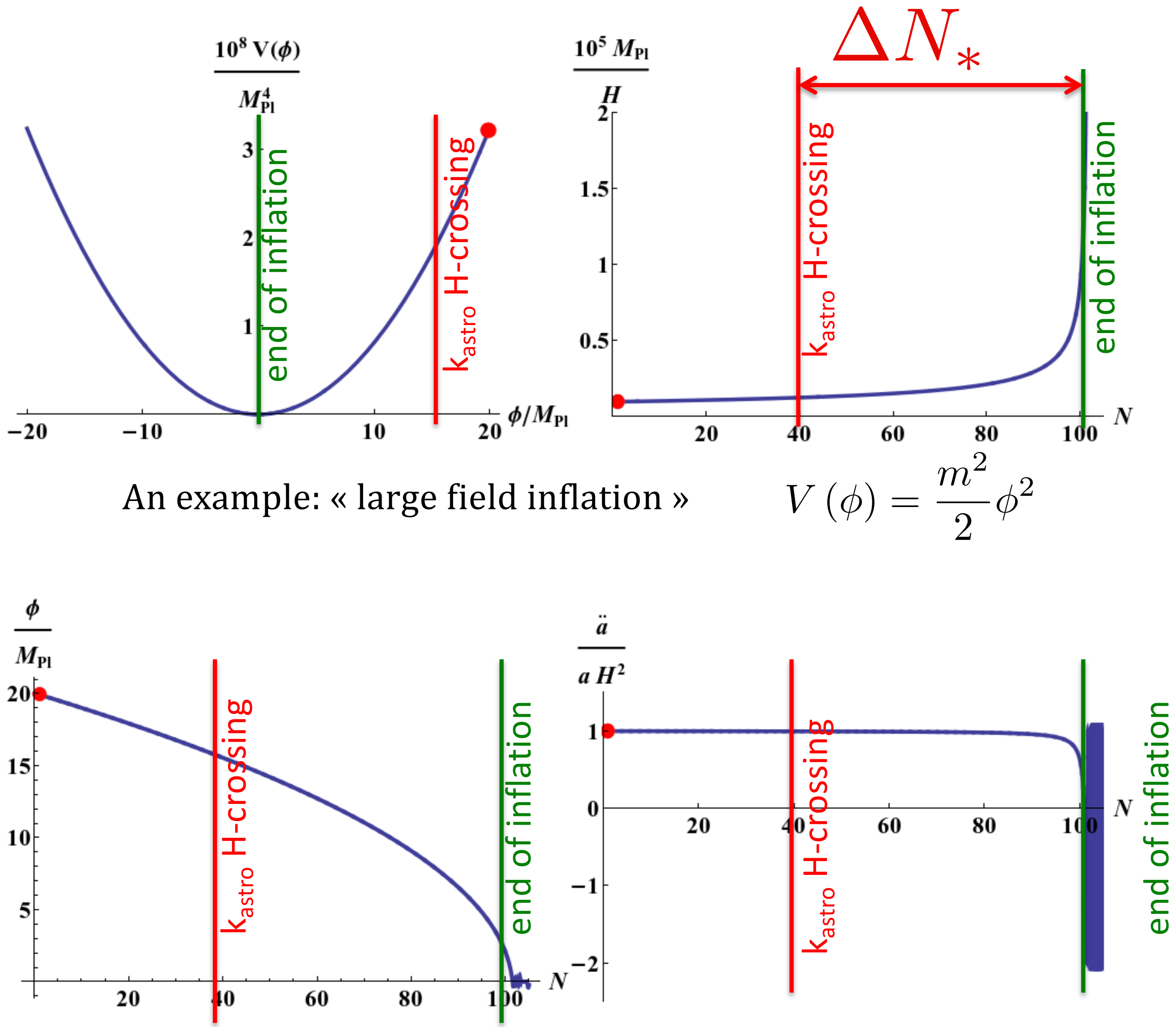
$$\Rightarrow \begin{cases} \rho = \frac{1}{2}\left(\dot{\phi}\right)^2 + V(\phi) \\ p = \frac{1}{2}\left(\dot{\phi}\right)^2 - V(\phi) \end{cases}$$

$$\ddot{a}/a = -\frac{1}{6M_\mathrm{P}^2}(\rho + 3p) \quad \Longrightarrow \quad V(\phi) \gg \dot{\phi}^2$$



- Combined with QM, accounts for an almost scale invariant power spectrum

credit: Vincent Vennin

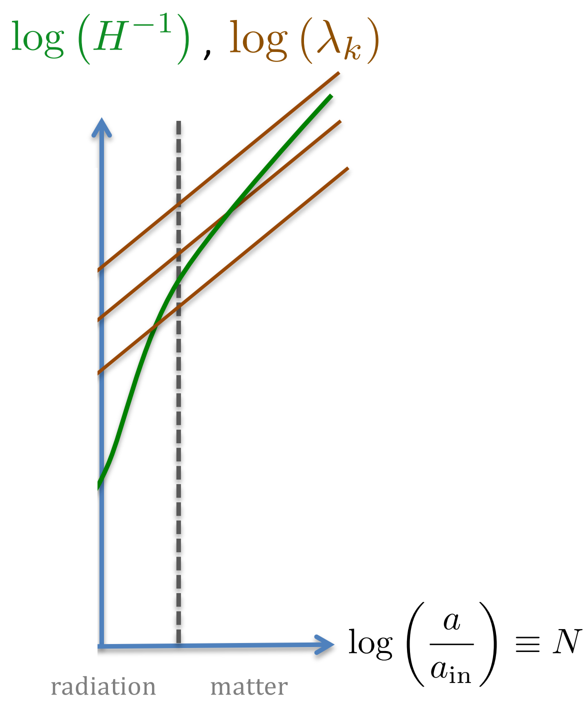


An example: « large field inflation » $V(\phi) = \frac{m^2}{2}\phi^2$

credit: **Vincent Vennin**

Cosmological Inflation

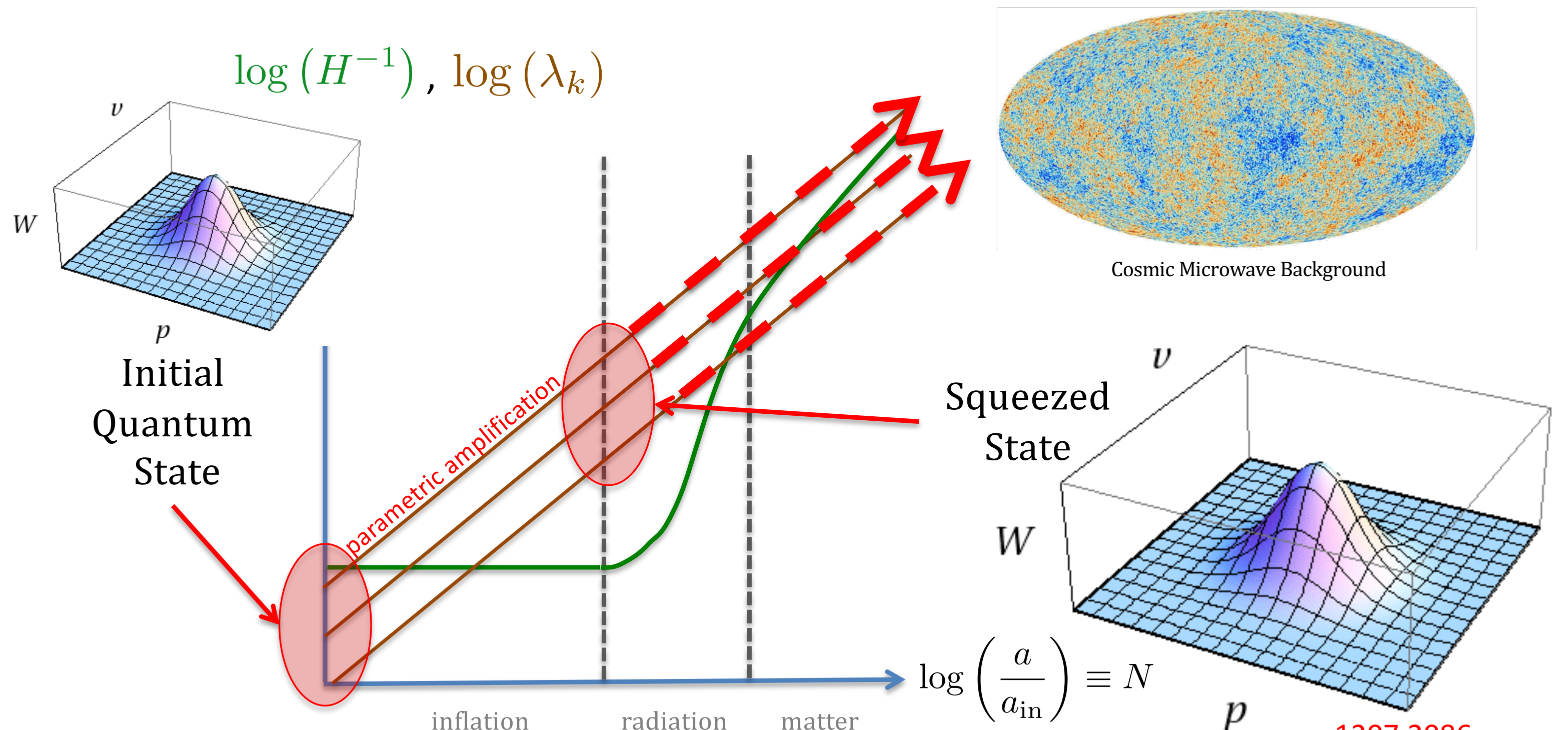
Solves the Horizon Problem



credit: Vincent Vennin

Cosmological Inflation

Quantized fluctuations evolved over an
expanding background



1207.2086
J.Martin, V. Vennin &
P.Peter

credit: **Vincent Vennin**

Scalar Power Spectrum

Cosmological Fluctuations:

- are combined gauge invariant perturbations of the metric and of the inflaton field v
- are the seeds of temperature anisotropies in the CMB and the structure and dynamics of matter $\quad v \propto \frac{\delta T}{T}$
- Follow a parametric amplifying equation of motion

$$v_{\boldsymbol{k}}'' + \left[k^2 - \frac{\left(a\sqrt{\epsilon_1}\right)''}{a\sqrt{\epsilon_1}}\right] v_{\boldsymbol{k}} = 0$$

Power Spectrum:

$$P_v(k) = \frac{k^3}{2\pi^2}\left\langle \hat{v}_k^2 \right\rangle$$
$$= \frac{a^2 H^2}{8\pi^2 M_{\mathrm{Pl}}^2 \epsilon_{1\diamond}}\left[1 - (2\epsilon_{1\diamond} + \epsilon_{2\diamond} + \cdots)\ln\frac{k}{k_\diamond} + \left(2\epsilon_{1\diamond}^2 + \epsilon_{1\diamond}\epsilon_{2\diamond} + \frac{\epsilon_{2\diamond}^2}{2} - \frac{\epsilon_{2\diamond}\epsilon_{3\diamond}}{2} + \cdots\right)\ln^2\frac{k}{k_\diamond} + \cdots\right]$$

Spectral index $\quad n_{\mathrm{S}} = \left.\frac{\mathrm{d}\ln P}{\mathrm{d}\ln k}\right|_{k_*}$

$$n_{\mathrm{S}}^{\mathrm{Planck}} \sim 0.96$$

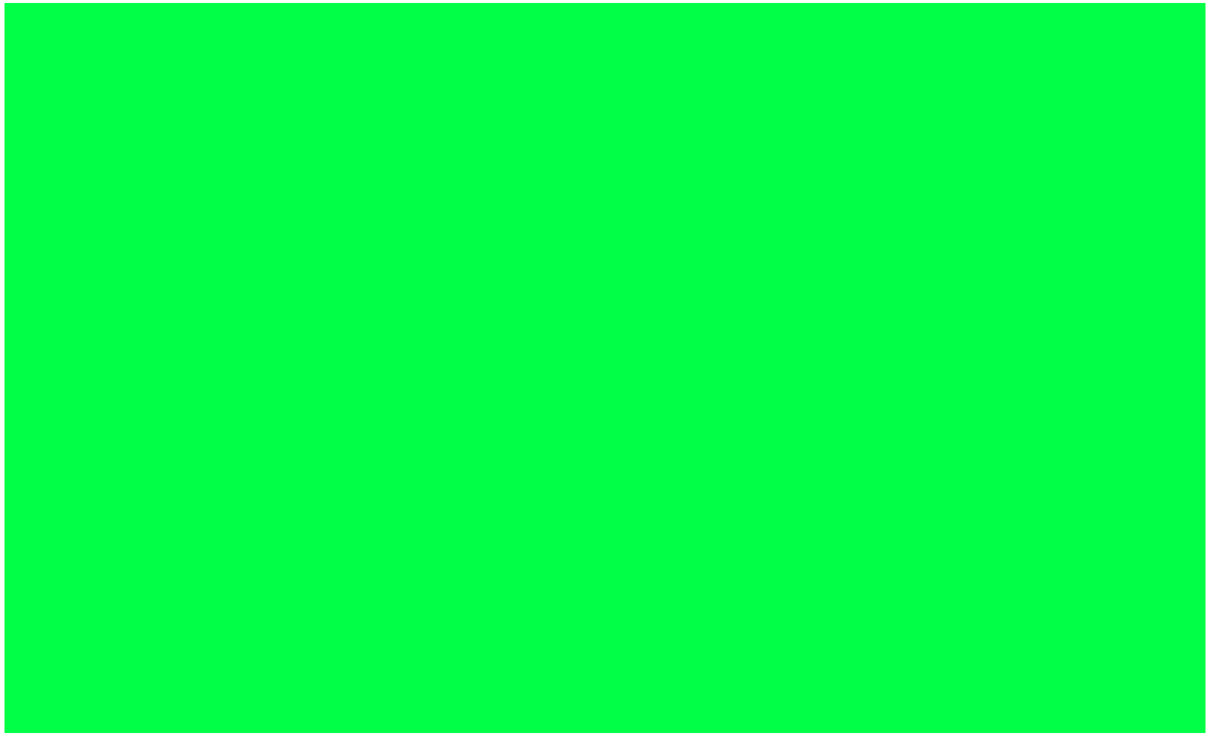
Gravity waves:

$$r = \frac{P_h(k_*)}{P_v(k_*)} = 16\epsilon_{1*} + \cdots$$

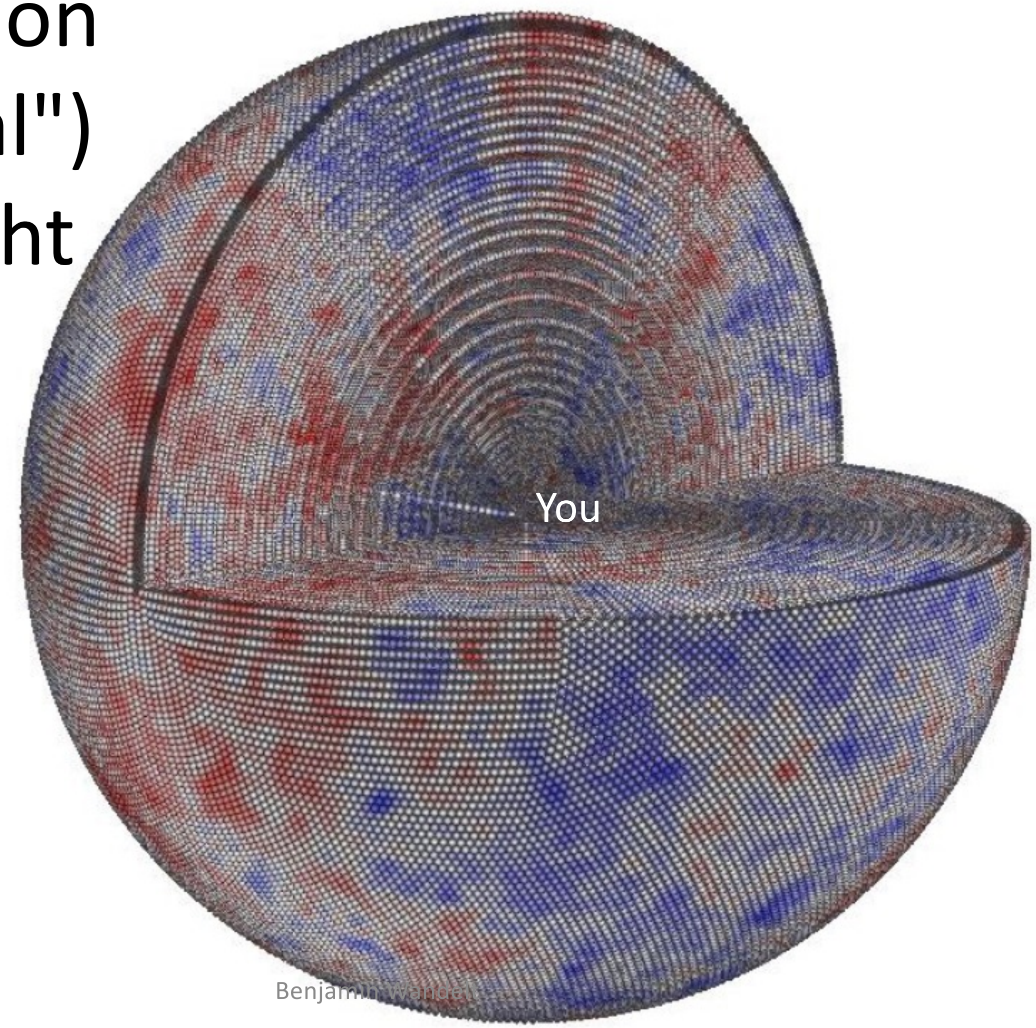
Benjamin Wandelt

credit: Vincent Vennin

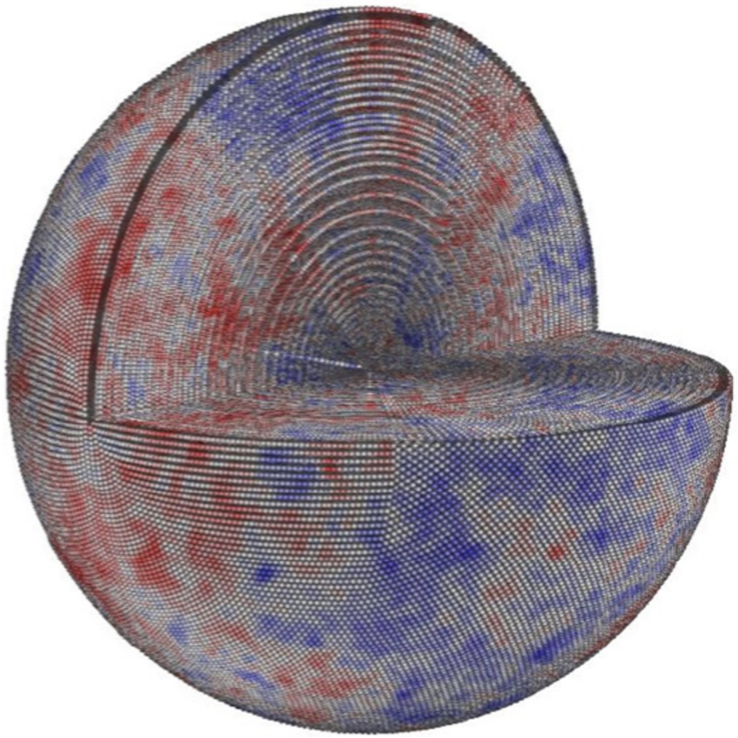
Cartoon of perturbations arising from quantum fluctuations



Curvature perturbation ("potential") on the light cone



Primordial perturbations give rise to all observations of cosmic structure



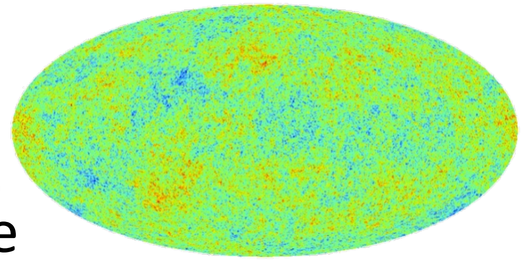
Radiative Transfer

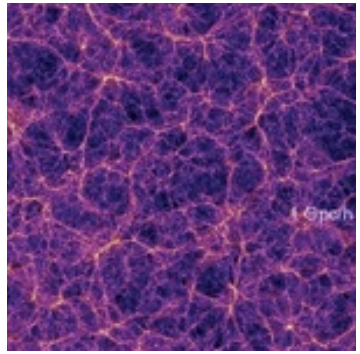
\+

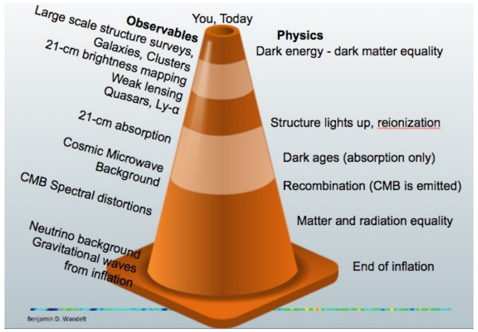
Gravity

\+

Astro-physics

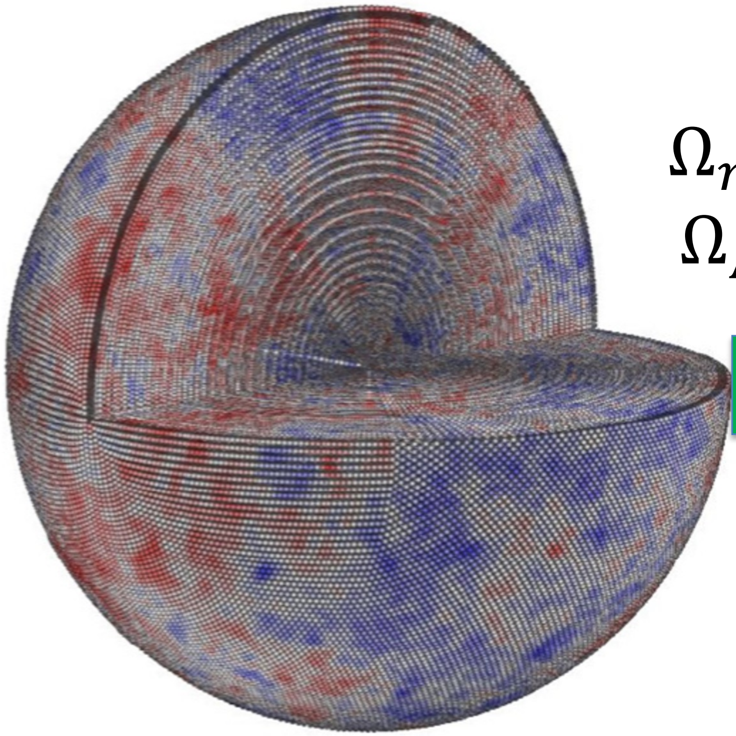






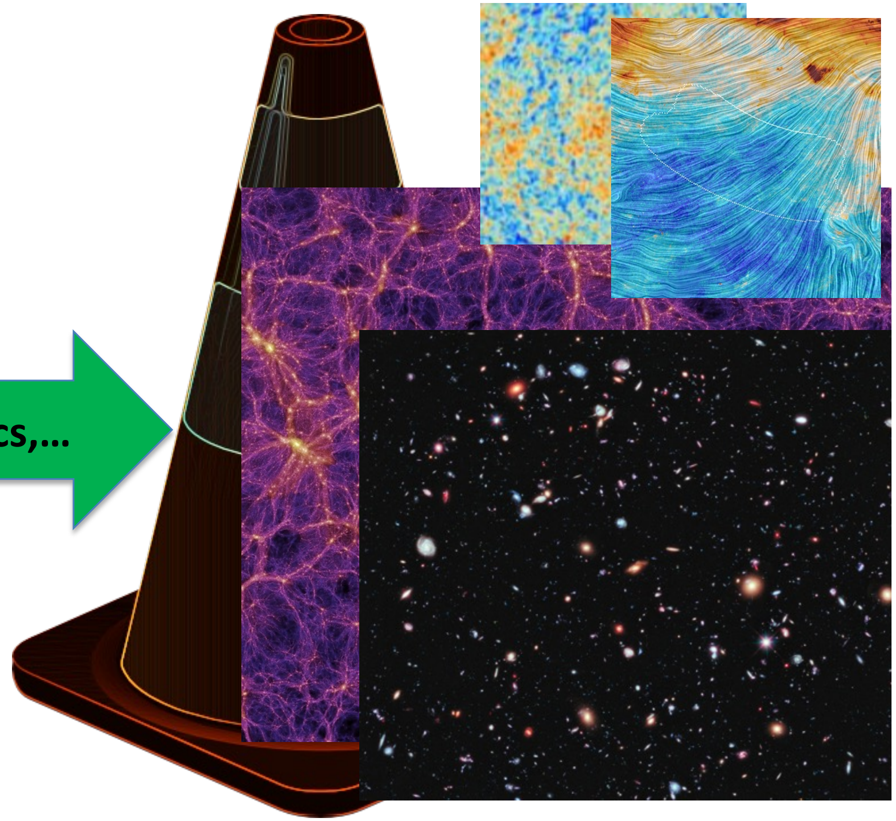
Cosmological **Computation**

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



$\Omega_m, \Omega_b, m_\nu, \ldots$
$\Omega_\Lambda, w_0, w_a, \ldots$

Gravity, Hydrodynamics,...



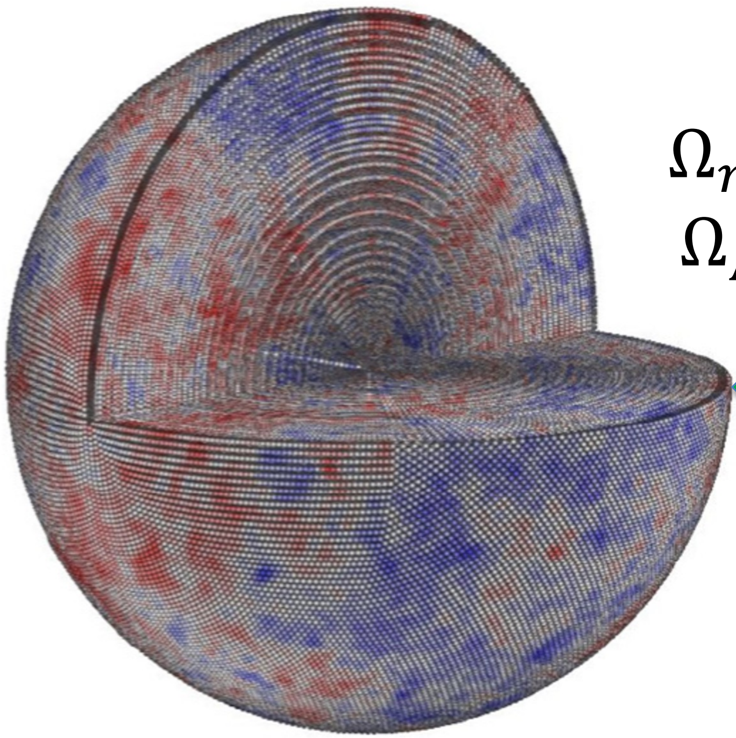
Initial conditions of the universe

The observed universe

Benjamin Wandelt

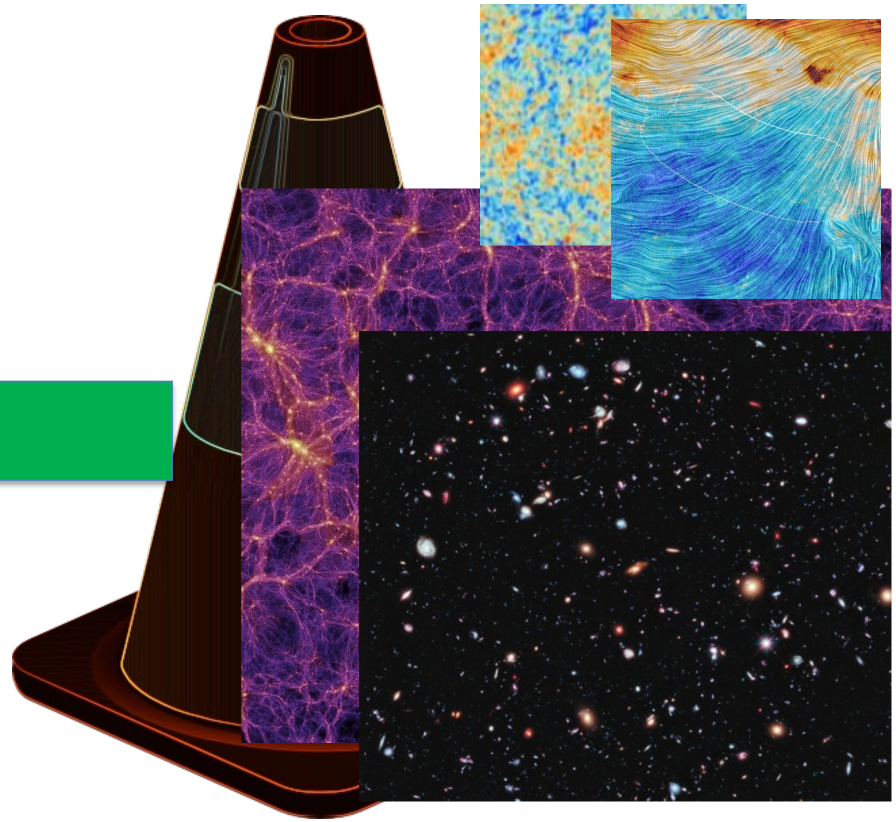
The Cosmological **Inference Problem**

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



$\Omega_m, \Omega_b, m_\nu, \ldots$
$\Omega_\Lambda, w_0, w_a, \ldots$

Inference

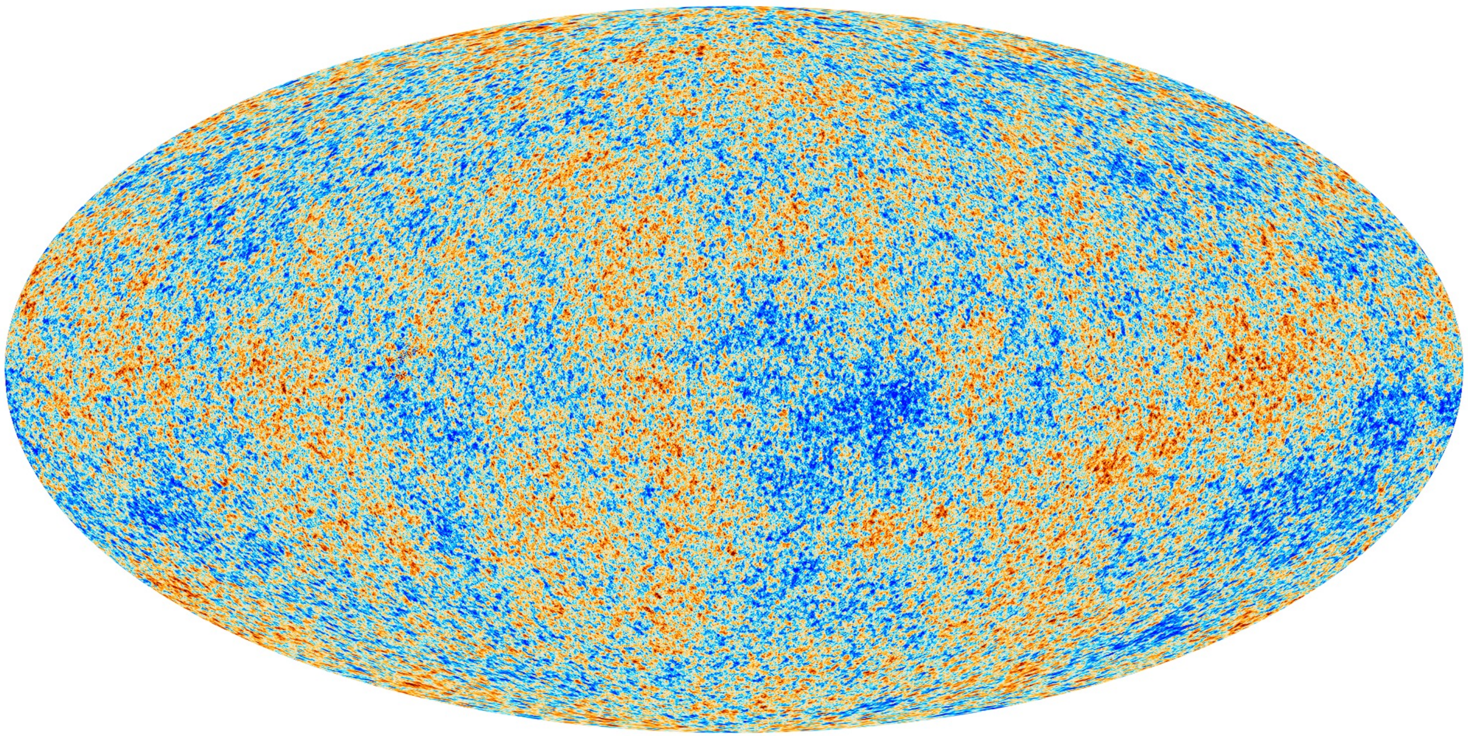


Initial conditions of the universe

The observed universe

Benjamin Wandelt

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



Simplicity is power: exploiting symmetries and physics

- Most general way the observable CMB anisotropy can be related linearly related to initial perturbations: represent map in terms of a “Fourier” expansion. The coefficients relate back to the initial conditions like so

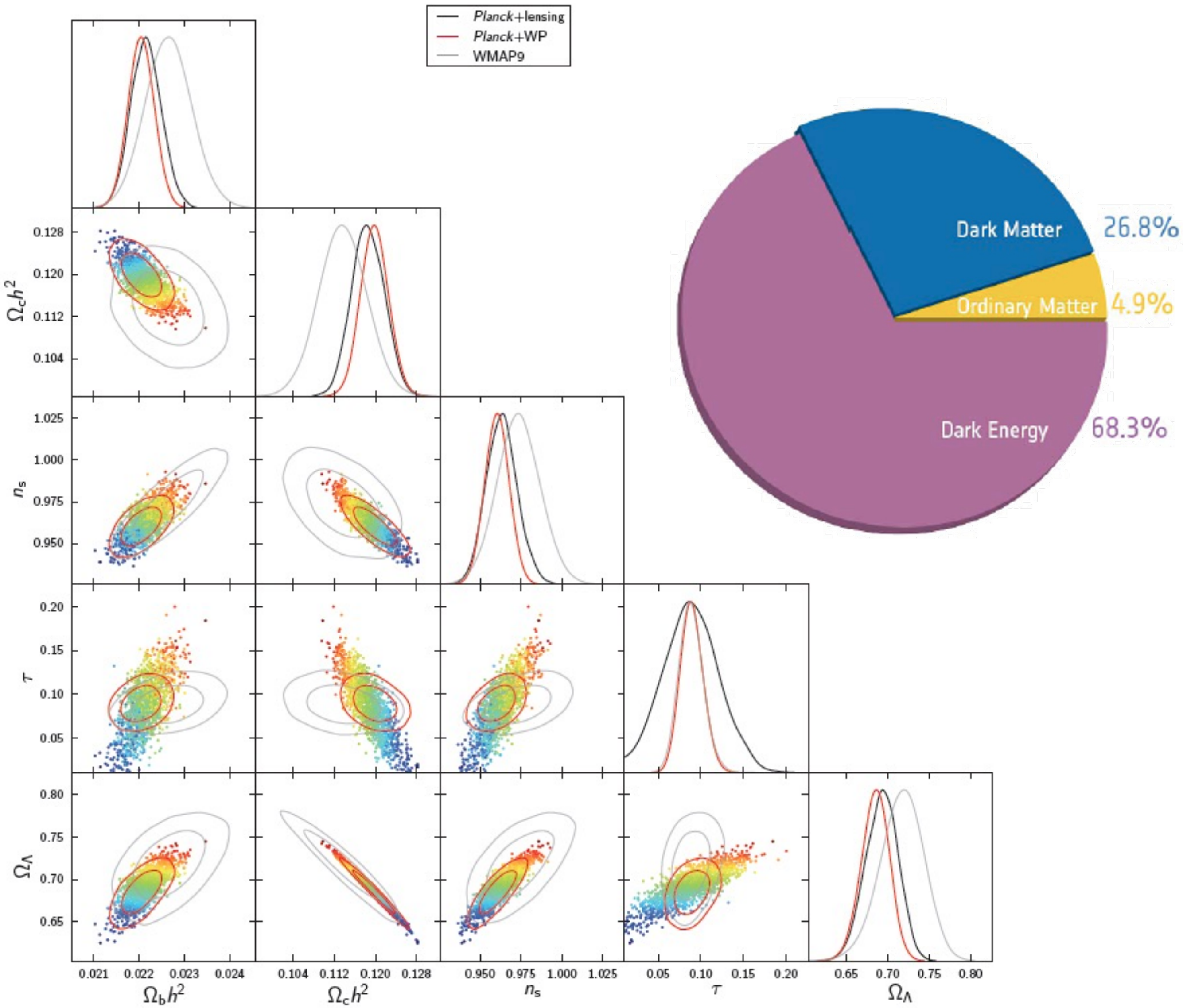
$$a_{lm} = \int d^3k \phi_k g_{lm\,k}$$

- In a homogeneous and isotropic universe with small fluctuations all the physics can be encoded in a linear transfer function ***g*** that *must* have the form

$$g_{lm\,k} = g_l(|k|) i^l Y^*_{lm}(\hat{k})$$

- ***g*** *is the solution to a set of diff eqs (fluid equations coupled to GR)*

Cosmological parameters from Planck's CMB maps



So far we have dealt with (nearly) Gaussian random fields

- For standard inflation the initial perturbations are very nearly Gaussian
- The primary anisotropies in CMB temperature and polarization are *linear functions* of the initial perturbations (to a very good approximation)

What is a Gaussian Random Field?

- This is something even some working cosmologists are confused about.
- I will take some time to explain this and answer your questions

What is a Gaussian Random Field?

$$p(x \mid C, \mu) = \frac{e^{-\frac{1}{2}(x-\mu)^T C^{-1}(x-\mu)}}{\sqrt{|2\pi C|}}$$

That's it!

The power of Gaussian fields

- Completely specified by C and μ
  - Can calculate all moments etc.
- Marginals (integrals) and conditionals (sections) can be calculated using linear algebra
- Nice mathematical properties
  - C^∞ and fast decay at ∞ so all polynomial moments exist.
- Nice physical properties
  - Sums of independent random variates, even non-Gaussian ones, and sometimes even dependent non-Gaussian ones, tend to be Gaussian (central limit theorem)

(Mis-)conceptions about Gaussian Random Fields

- True or false: Gaussian random fields have independent Fourier (or momentum) modes.
- False in general: only true for homogeneous random fields

(Mis-)conceptions about Gaussian Random Fields

- True or false: Gaussian Random Fields are defined as fields with random phases

- Not true in general:
  - First of all: poorly defined ("random?")
  - Argument *works* as follows: *if* a field is homogenous *and* Gaussian, *then* the phases are independent and drawn uniformly from [0,2π]
  - But does *not* work in reverse: can of course construct non-Gaussian field with independent phases drawn uniformly from [0,2π]

(Mis-)conceptions about Gaussian Random Fields

- True or false: Histograms of Gaussian Random Fields are Gaussian
- Not true in g… …lection of n independent… …For large n histogram will converge to a *mixture* of Gaussians!

Need to be careful about **tests** of Gaussianity – they can easily be confused by anisotropy/inhomogeneity.

$$hist \leftarrow \frac{1}{n}\sum_{i=1}^{n} g(x_i|\mu_i, C_{ii})$$

Benjamin Wandelt

Can we use GRFs to build a linear physics time machine?

- Let's take our CMB sky to be *y* and the slice of primordial potential we'd like to reconstruct as *x*.
- Then can use the formulae for the conditional density of x given y to build the optimal inference of x given y.

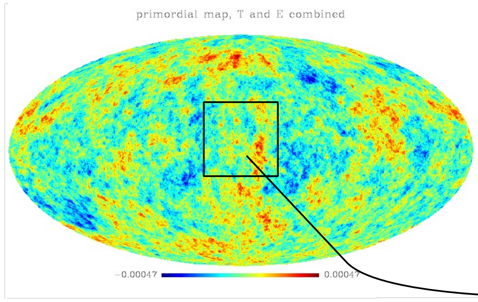
$$\mu = \begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix}$$

$$C = \begin{pmatrix} C_{xx} & C_{xy} \\ C_{yx} & C_{yy} \end{pmatrix}$$

$$\mu_{x|y} = \mu_x + C_{xy} C_{yy}^{-1} (y - \mu_y)$$
$$C_{x|y} = C_{xx} - C_{xy} C_{yy}^{-1} C_{yx}$$

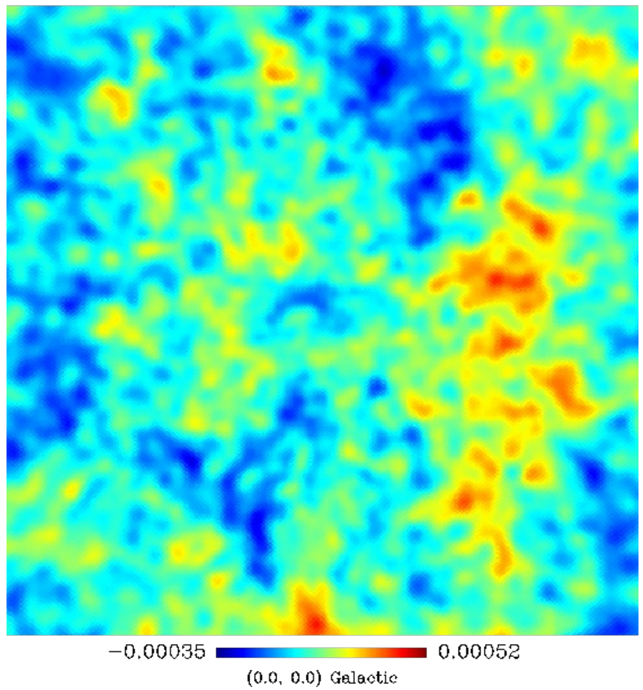
Our linear physics CMB time machine

simulation



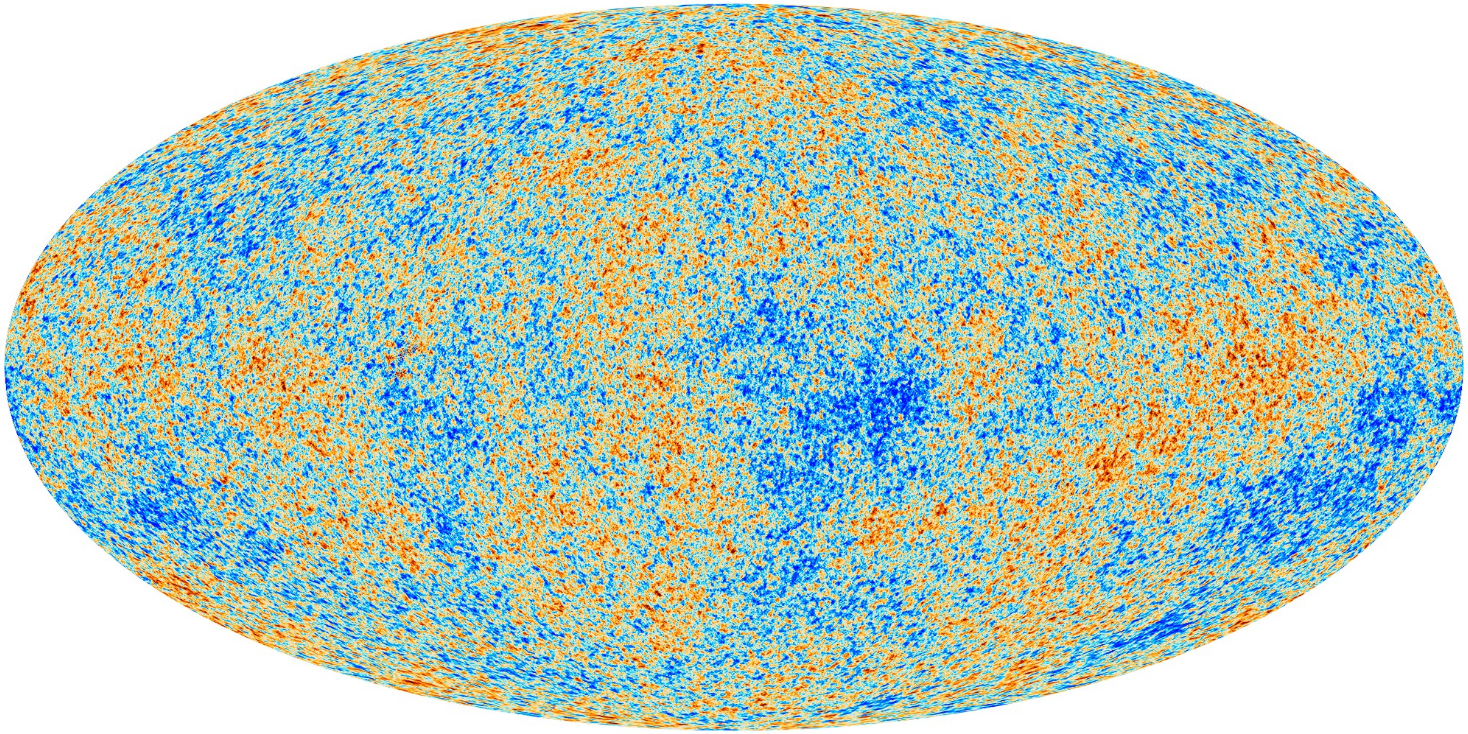
- The CMB T&E anisotropies map the Universe at t=380,000
- By "inverting" linear physics we can infer primordial curvature perturbation and test model predictions for the power spectrum and beyond.

Primordial curvature fluctuations

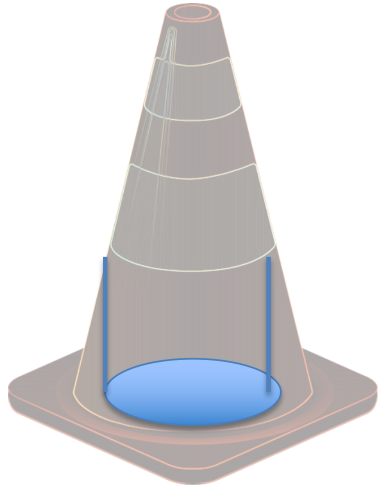


Komatsu, Spergel, Wandelt (2005)
Yadav and Wandelt (2005)

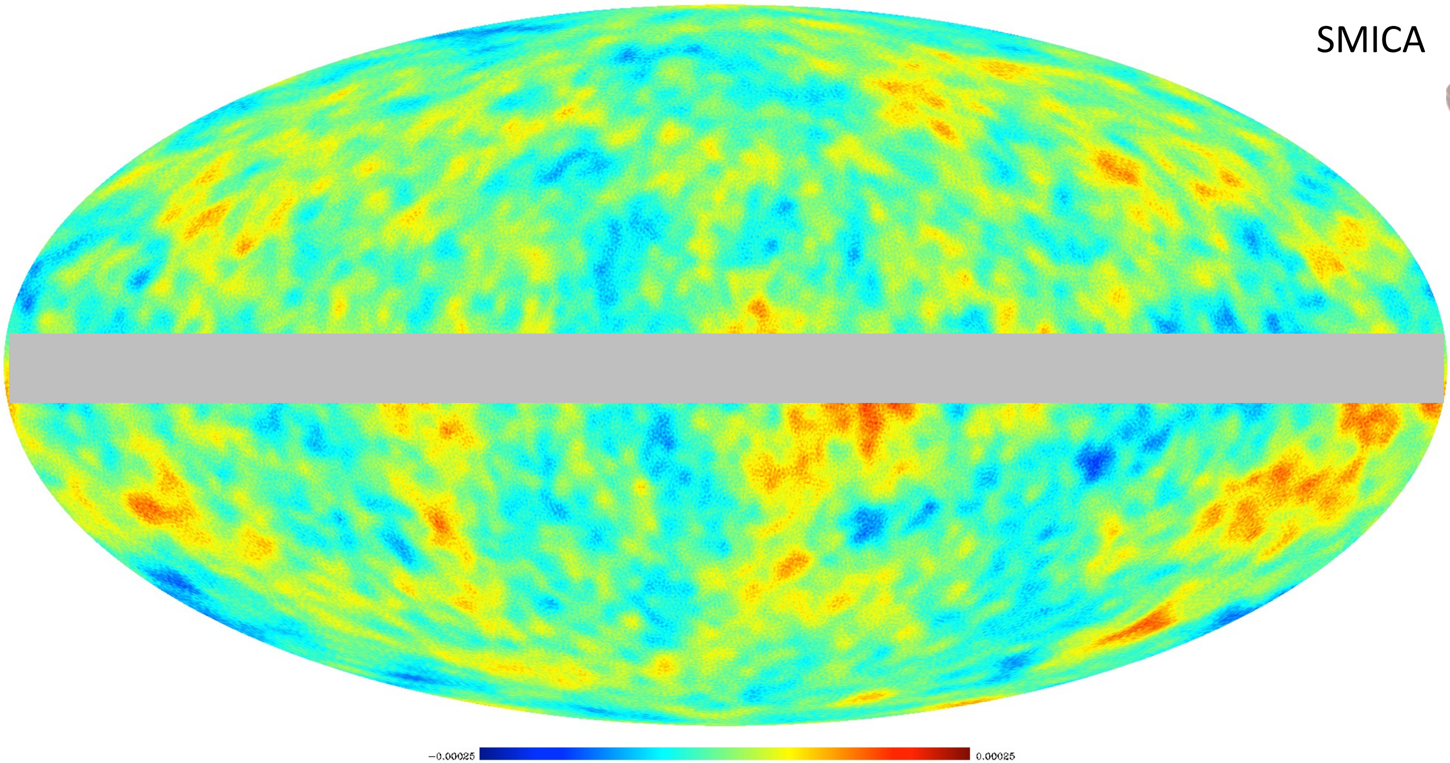
Let’s do it for our actual Cosmic Microwave Background anisotropies



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



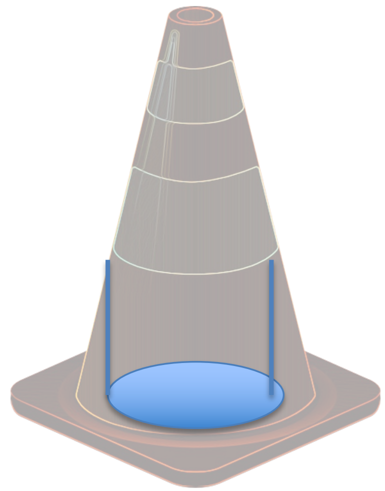
SMICA

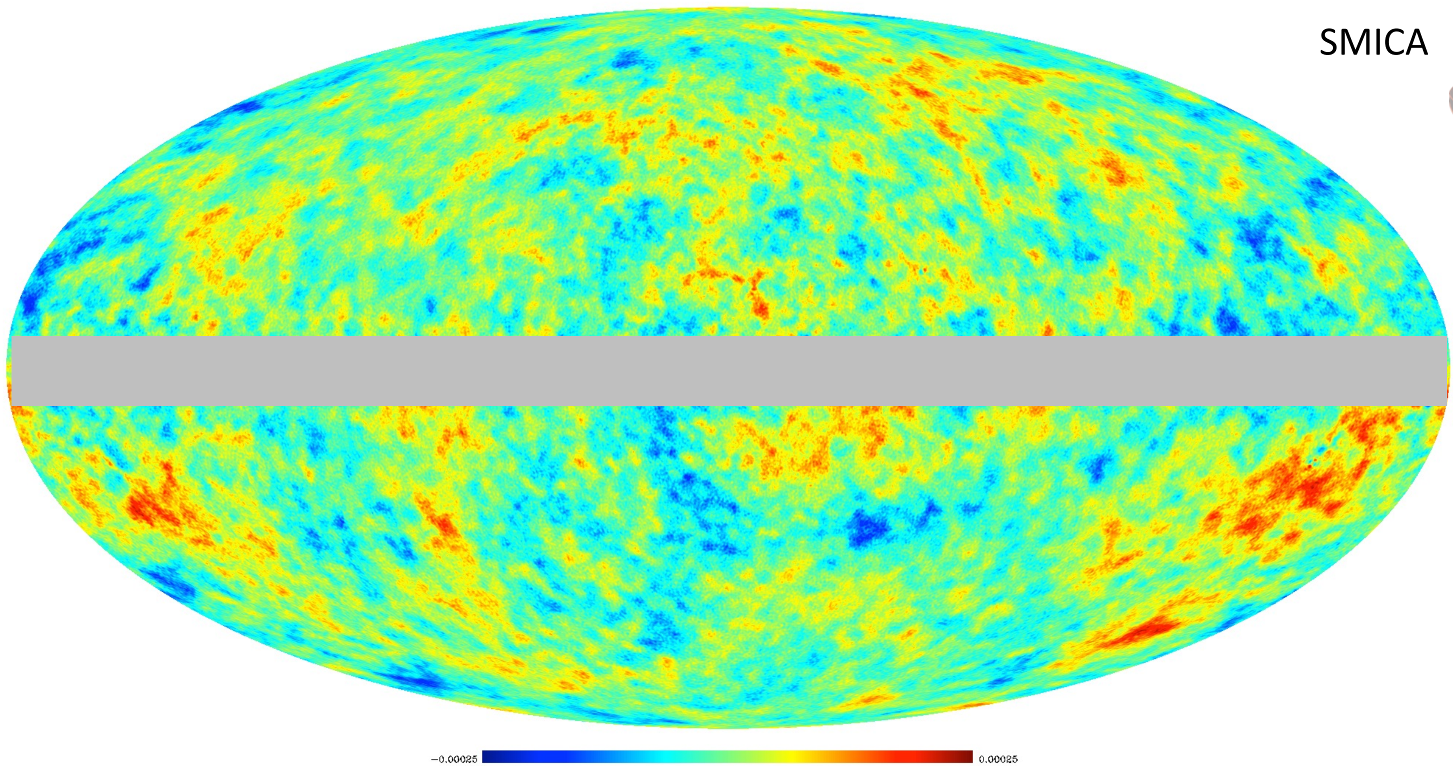


Planck collaboration 2013 data

Method: Komatsu, Spergel, Wandelt 2005

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





Planck collaboration
2013 data

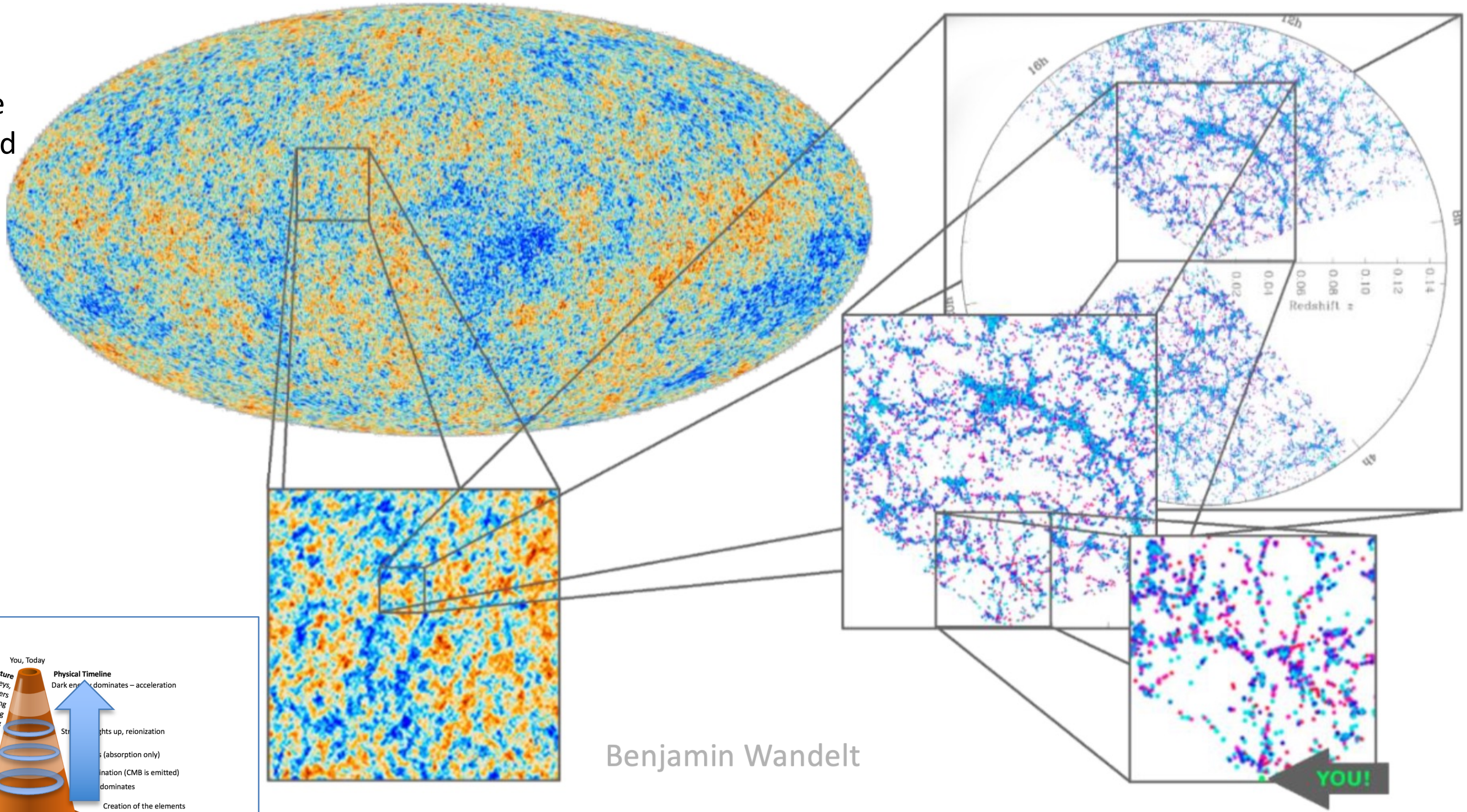
Method: Yadav & Wandelt 2005

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

Beyond Gaussian fields

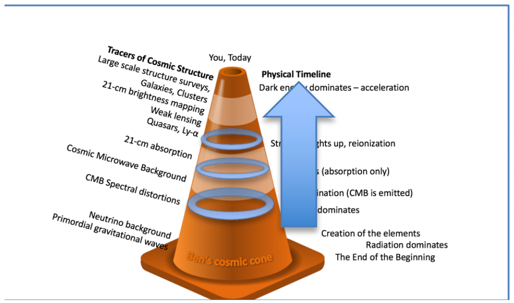
Cosmological data covers a hierarchy of scales on the past light cone. Smaller scales ➔ increasing complexity

Cosmic Microwave Background



Galaxy surveys

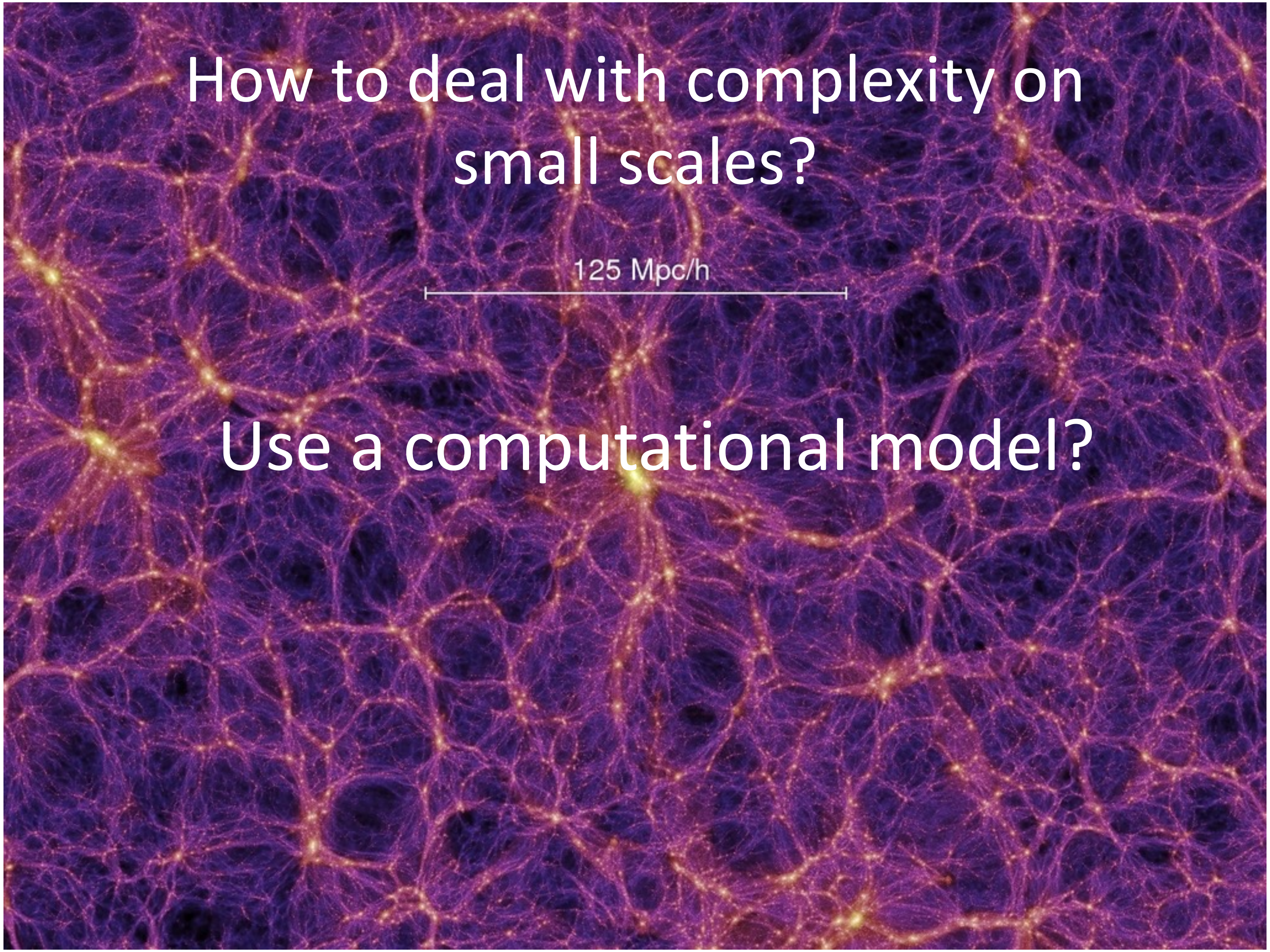
Benjamin Wandelt



How to deal with complexity on small scales?

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?

Beyond the linear and the Gaussian

- Even for Gaussian fields, non-Gaussian statistics arise for covariance estimation (power spectrum inference, parameter inference)
- 21st century cosmology deals with non-linear, non-Gaussian problems:
  - Gravitational non-linearity! Galaxy formation! Gastrophysics!
  - Primordial non-Gaussianity?
  - Data imperfections and systematics!

Current practice

1. Pick summary statistics (e.g., galaxy number, pair counts, triplets,…)

2. Compute predictions for these summaries using theory or simulations

3. Approximate likelihood (often Gaussian)

Risks of current practice

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

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

Inadequate statistical approximations (leading to tensions?)

We can do this

- Dropping the Gaussianity assumption boils down the problem to this:

"If I have a way to predict x from y then I can use x to constrain y"

$$\mathbf{P(y|x)P(x)=P(x|y)P(y)}$$

I think I've seen that before:

$$p(\theta|d) = \frac{p(d|\theta)p(\theta)}{p(d)}$$

$$p(theory|data) = \frac{p(data|theory)p(theory)}{p(data)}$$

➔ Lighting quick intro to Bayesian inference

Bayes' theorem

$$p(\theta|d) = \frac{p(d|\theta)p(\theta)}{p(d)}$$

$p(d|\theta)$: likelihood

$p(\theta)$: prior

$p(\theta|d)$: posterior

$p(d)$: evidence

What is Bayesian analysis

One sentence summary of Bayesian stats:

Whatever is uncertain gets a pdf.

$$p(\theta|d) = \frac{p(d|\theta)p(\theta)}{(d)}$$

$p(d|\theta)$: data *before it is known* given parameters

$p(\theta)$: parameters in the absence of data

$p(\theta|d)$: parameters after the data are known

$p(d)$: data *before it is known (*unknown parameters)

Inputs to Bayesian analysis

- Priors
  - Modeling assumptions, both theoretical and experimental
  - We often have excellent physical motivation for choosing priors in cosmology
  - Specifying priors means getting the assumptions out in the open
    - Whenever someone (Bayesian or frequentist) tells you "we did not have to assume anything!" DO NOT TRUST THEM.

- Data
  - e.g. CMB data, Galaxy catalog, etc,…
  - e.g. Detailed survey specifications

Output of Bayesian analysis

- Key point: the output is the **posterior density**

$$p(\theta|d) = \frac{p(d|\theta)p(\theta)}{p(d)}$$

Output of Bayesian analysis

- Ok, you plug in your data into Bayes' theorem. Now what?
- **<u>Explore</u>** the posterior

$$p(\theta|d) = \frac{p(d|\theta)p(\theta)}{p(d)}$$

  - Visualize
  - Compute summaries e.g.
    - mean and variance of each parameter, marginalizing over all others
    - Means and variances of groups of parameters
  - Validate
  - perform model/prior checking
    - Simulate data using parameters drawn from the posterior and see if it agrees with your data

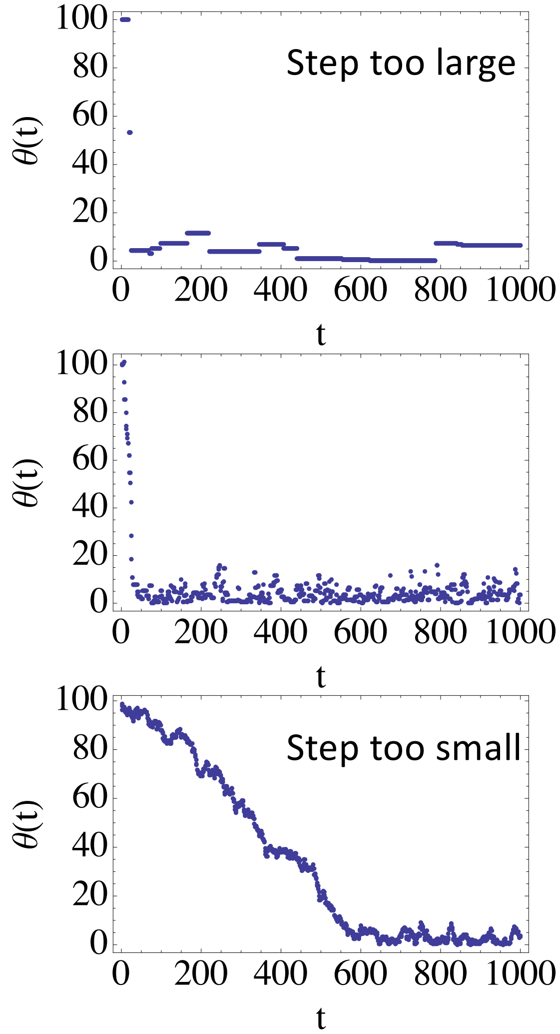
M-H sampling from the posterior

- Practical approaches generate a correlated "Markov Chain" with *n* steps, that converges to the posterior p(θ) for $n \to \infty$
- E.g Metropolis-Hastings
  - Specify proposal pdf q(θ'|θ)
    - Draw θ' from q(θ'|θ)
    - If $$a = \frac{p(\theta')q(\theta)}{p(\theta)q(\theta')} > 1$$
      then accept θ'; otherwise accept with probability *a*.

M-H problem

- M-H can be inefficient if proposal pdf q(θ') is suboptimal
  - Hard to find good q(θ') if number of parameters is high (>10)
  - Typically the chain moves very slowly
    - Either due to tiny step size
    - Or because a tiny fraction of proposals are accepted
  - Can be diagnosed using lagged auto-correlation of the chain

$$\xi(\Delta) = \int \theta(t)\theta(t+\Delta)dt$$



Hamiltonian Monte Carlo

- Duane, Kennedy, Pendleton & Roweth (1987) proposed the following trick:
- Think of log-posterior as a potential $\psi(\theta) = -\ln p(\theta)$
- Introduce conjugate momenta p_i, one for each parameter
- Choose Gaussian pdf for $\{p_i\}$ to get Hamiltonian
- $$H = \sum_{ij} \frac{1}{2} p_i M^{-1}_{ij} p_j + \psi(\theta)$$
- Then do M-H step with the proposal defined by drawing from $p(\{p_i\})$and then moving θ using Hamiltonian integrator respecting symplectic symmetry
- This conserves the Hamiltonian and hence gives acceptance probability 1 !!!
- In the end just ignore $\{p_i\}$ to get $p(\theta) = \int p(\theta, p) Dp$

Bayesian model comparison

- Second level inference
  – If we have more than one model and prior I can compare their relative probability after having seen the data:

Evidence for first level inference of model M_1

$$\frac{p(M_1|d)}{p(M_2|d)} = \frac{p(d|M_1)p(M_1)}{p(d|M_2)p(M_2)}$$

Evidence for first level inference of model M_1

  – Gives quantitative answers to everyday cosmology and astrophysics questions!

Example questions answered by model comparison

- Is the universe flat?
- Is gravity GR or a modification?
- Initial conditions: Gaussian or non-Gaussian?
- Isocurvature: yes or no?
- Is the equation of state of the dark energy -1?

- Did I detect…
  - a source?
  - a spectral line?
  - gravitational waves?
  - a reionization bubble?
  - non-zero neutrino mass?

Why Bayesian model comparison?

- It answers the question most people want to ask
- Which of these do you find more intuitive:
  - Frequentist null-hypothesis testing: “Given an ensemble of infinitely many fictional Gaussian Universes, what is the probability of the non-Gaussianity estimate from the ensemble being larger than the estimate from the data?”
  - Bayesian model comparison: “What is the relative probability of cold or warm dark matter given the data?”

Example application of Bayesian inference

Using this we can now build *non-linear* time machines using Bayesian posterior exploration

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

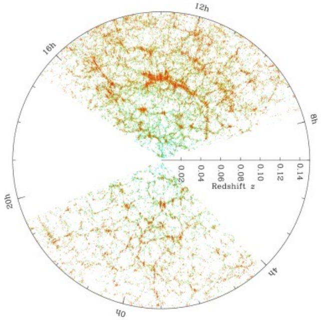
BORG: *Bayesian Origin Reconstruction from Galaxies*

- 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**

N-body
512^3
ICs-> mock in
10-15 secs
on 800 cores



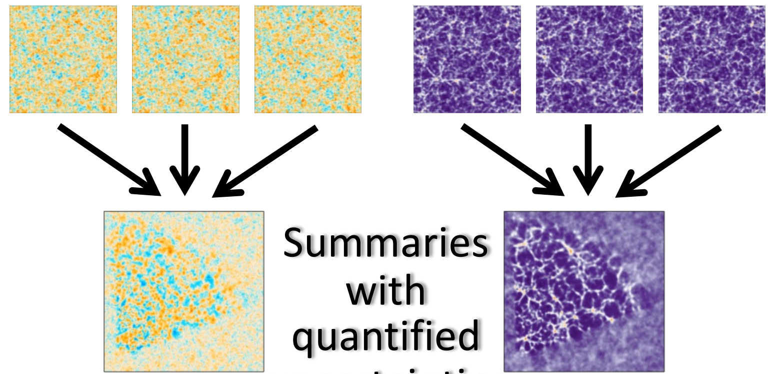
Observations

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

→ BORG →

Initial conditions, and inferred dark matter densities

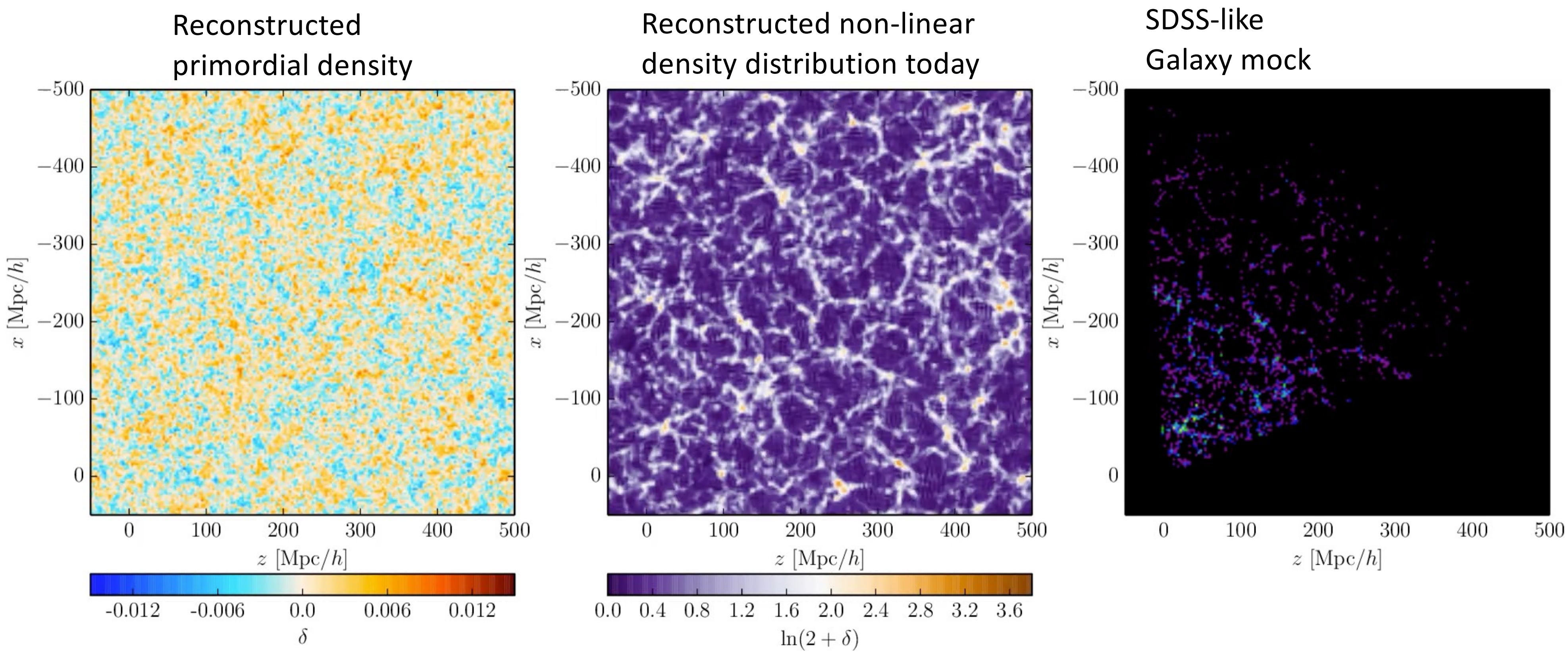
z=100 z=0



Summaries with quantified uncertainties

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

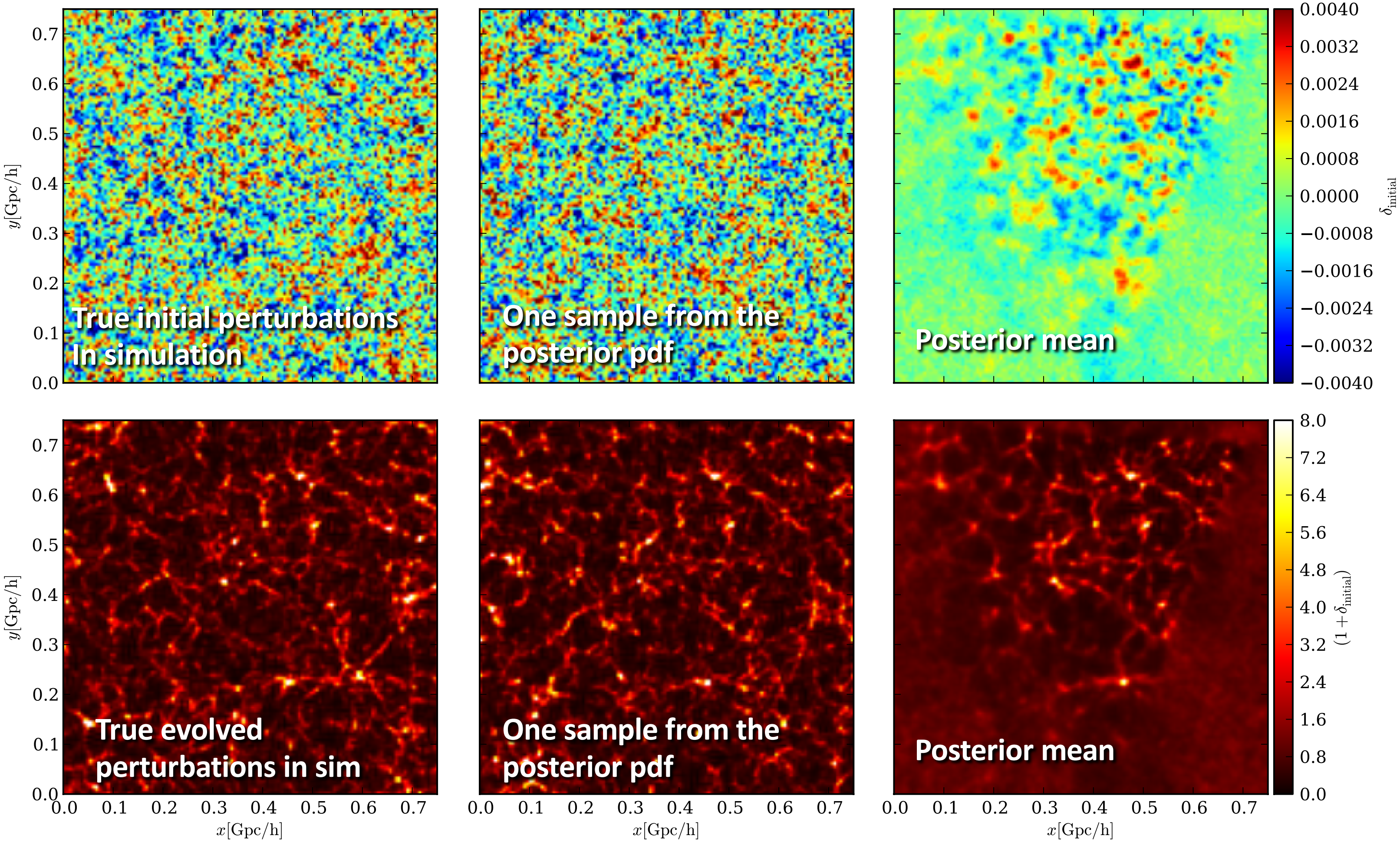
Markov Chain through Initial Conditions (density slices in 3D field)



The movie shows Bayesian physical reconstruction of initial conditions from Large Scale Structure

Benjamin Wandelt

Bayesian physical reconstruction of initial conditions from Large Scale Structure



How to think about samples from the posterior

- “Samples from the posterior density” – what does this mean?
- I find it helpful to think of each sample as a possible version of the truth.
- The variation between samples quantifies the uncertainty that results from having, e.g.
  - only one Universe (this is a more precise version of “cosmic variance”)
  - Imperfect data (mask, finite volume, finite number of galaxies, photo-z)

BORG is currently the most advanced explicitly likelihood analysis framework for galaxy surveys.

So… are we done? Problem solved?

No.

Challenges as a Bayesian Scientist

Even if likelihood and posterior are assumed to be known

- Posterior computation can be challenging
- Good MCMC can be hard
- Likelihood can be costly to evaluate
- Evidence can be hard to compute

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

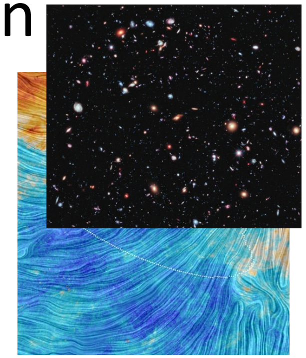
And most importantly:

- The full statistical power even of current data is enormous
- The key issue is ***model misspecification***

How to do science as a Bayesian

1. Work out the full physical and stochastic model of data given parameters θ. This is the **Likelihood**.
2. Get data d.
3. Specify prior
4. Write down posterior

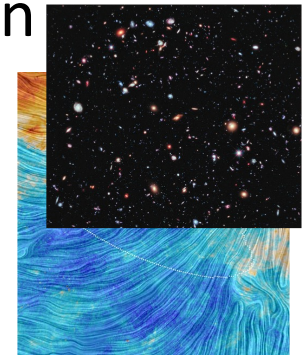
$$P(\theta|\mathrm{d}) = \frac{\cancel{P(\mathrm{d}|\theta)}\,P(\theta)}{P(\mathrm{d})}$$

What if d =  ?

5. Explore posterior for fixed data as a function of parameters, e.g. on a grid or using Markov Chain Monte Carlo.
6. Done!

How to do science as a Bayesian

1. Work out the full physical and stochastic model of data given parameters θ. This is the **Likelihood**.
2. Get data d.
3. Specify prior
4. Write down posterior

What if d =  ?

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

5. Explore posterior for fixed data as a function of parameters, e.g. on a grid or using Markov Chain Monte Carlo.

What if we can't write down the prior?

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

Yes!

A major shift over the last 5 years.

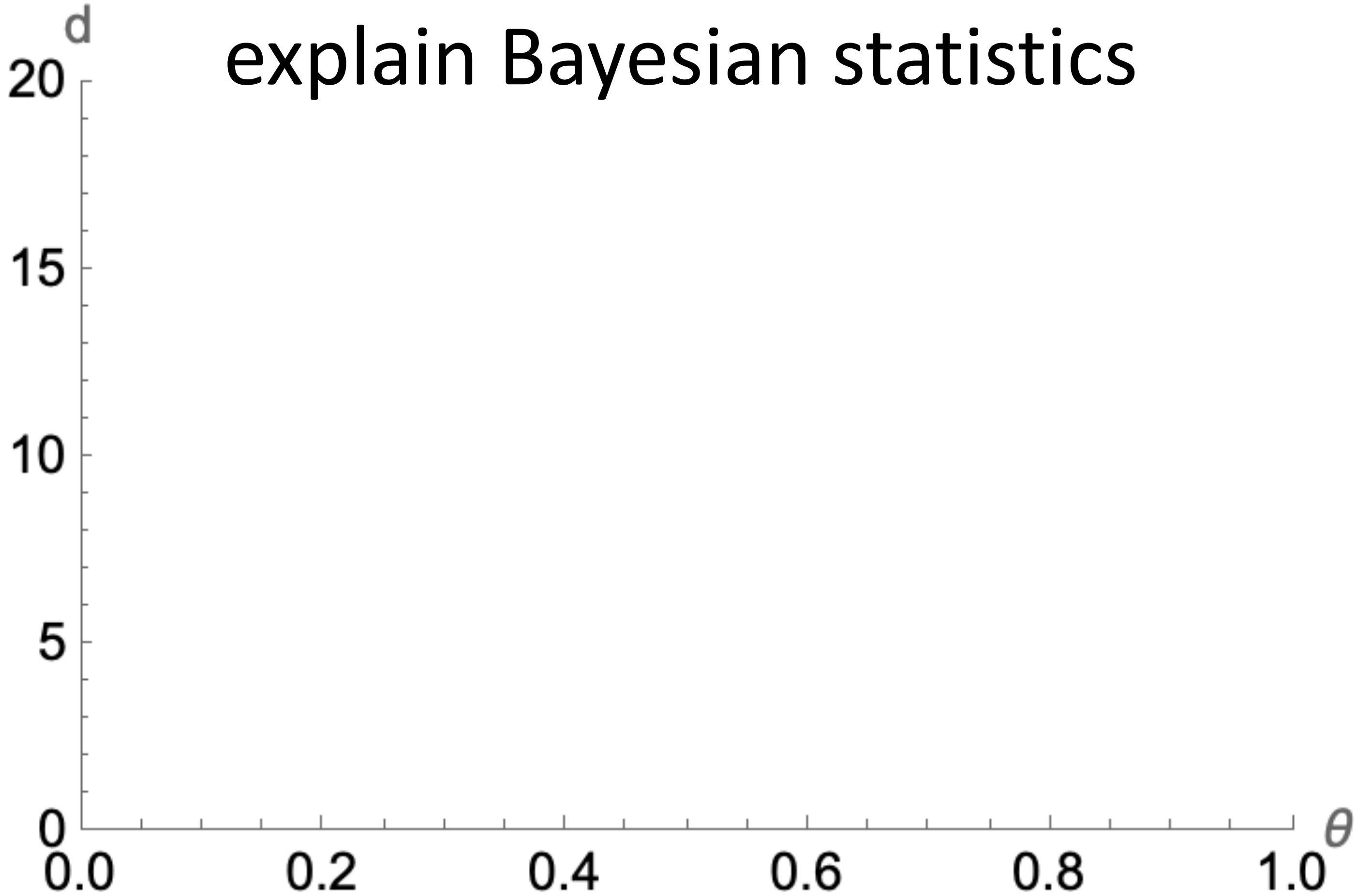
Likelihood is represented implicitly through simulations

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

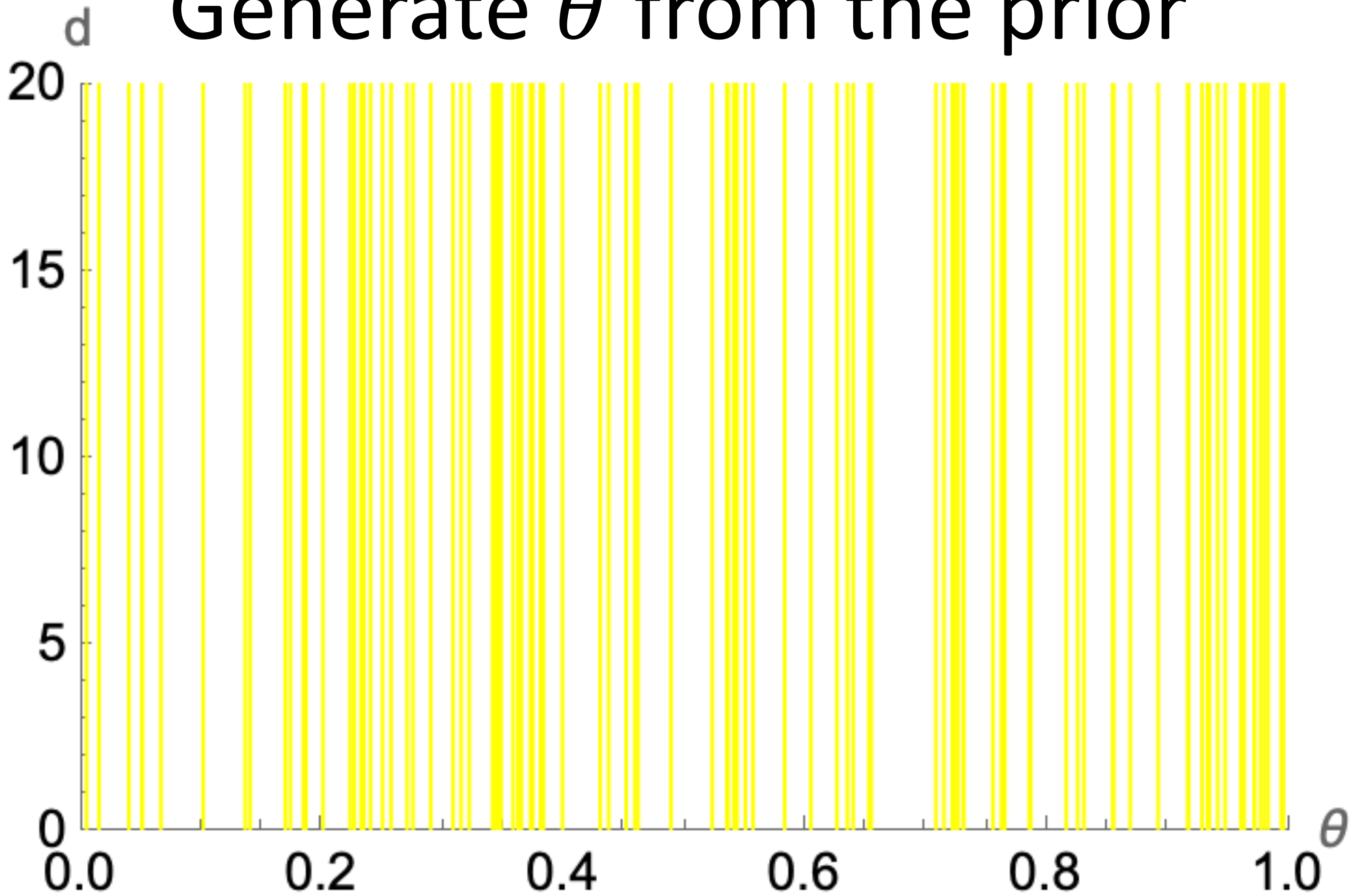
Let's do a simple example.

Challenge: Keep a running count of the number of likelihood and prior evaluations!

The best diagram to explain Bayesian statistics

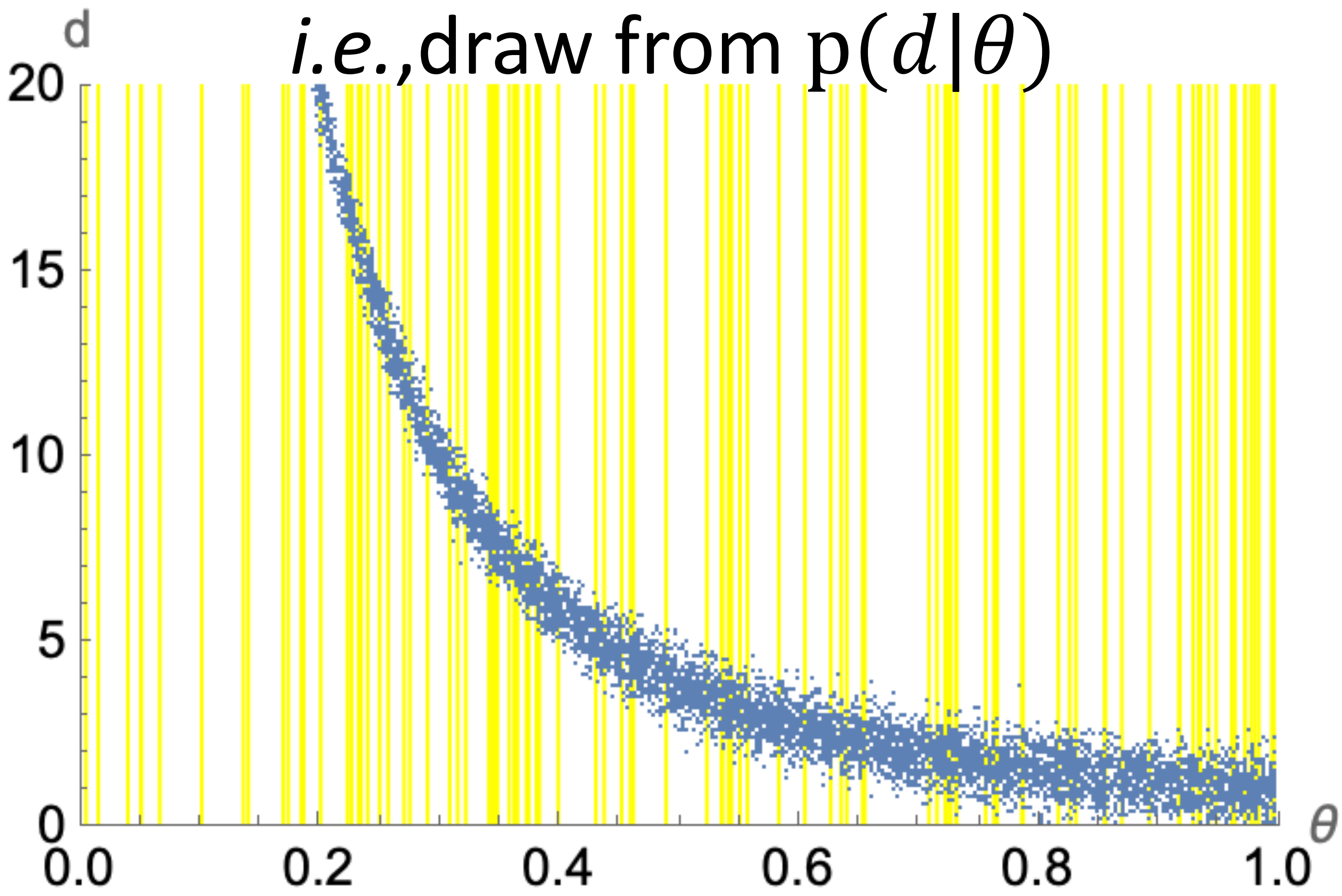


Generate θ from the prior

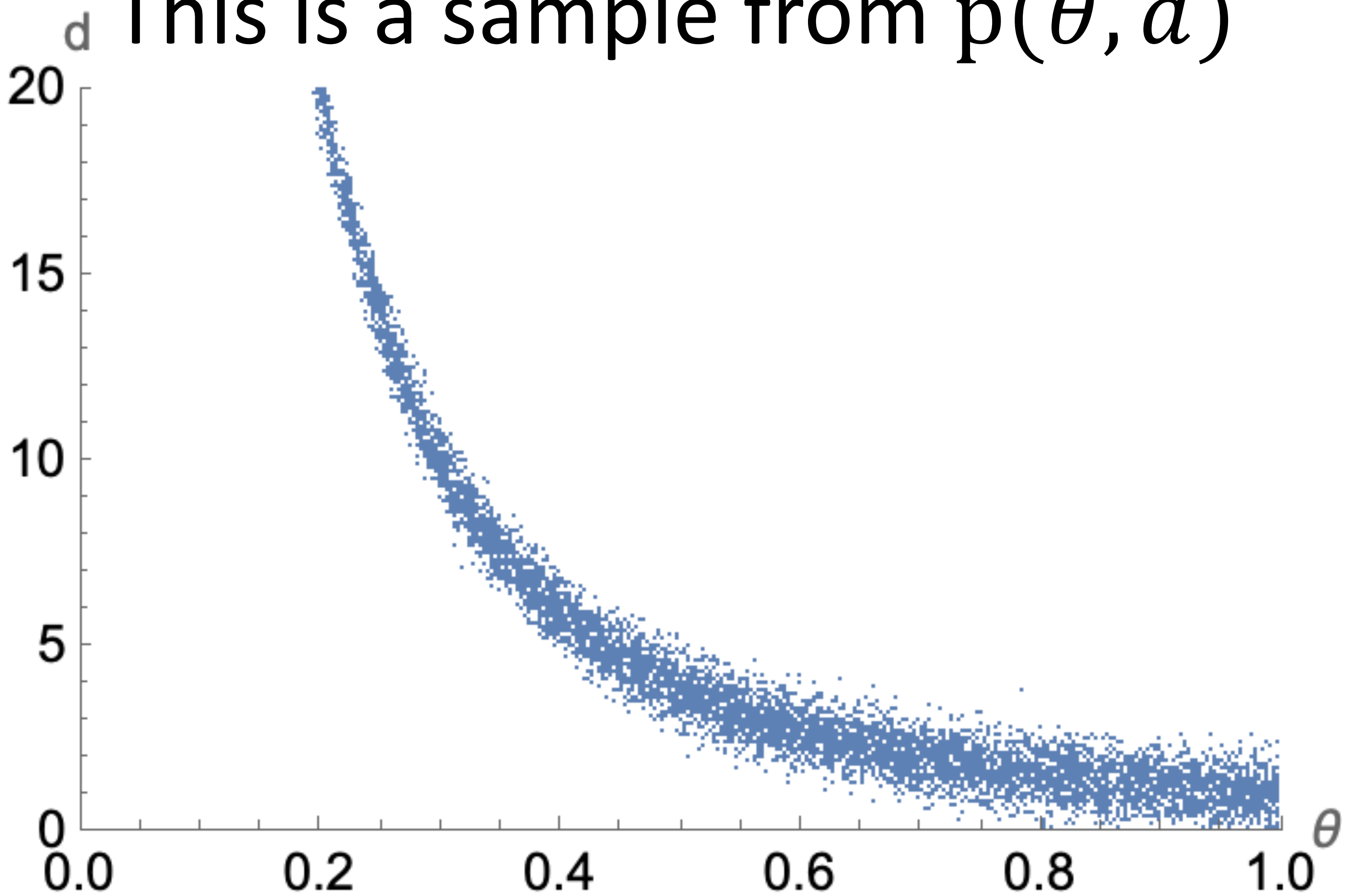


Simulate/generate d given θ

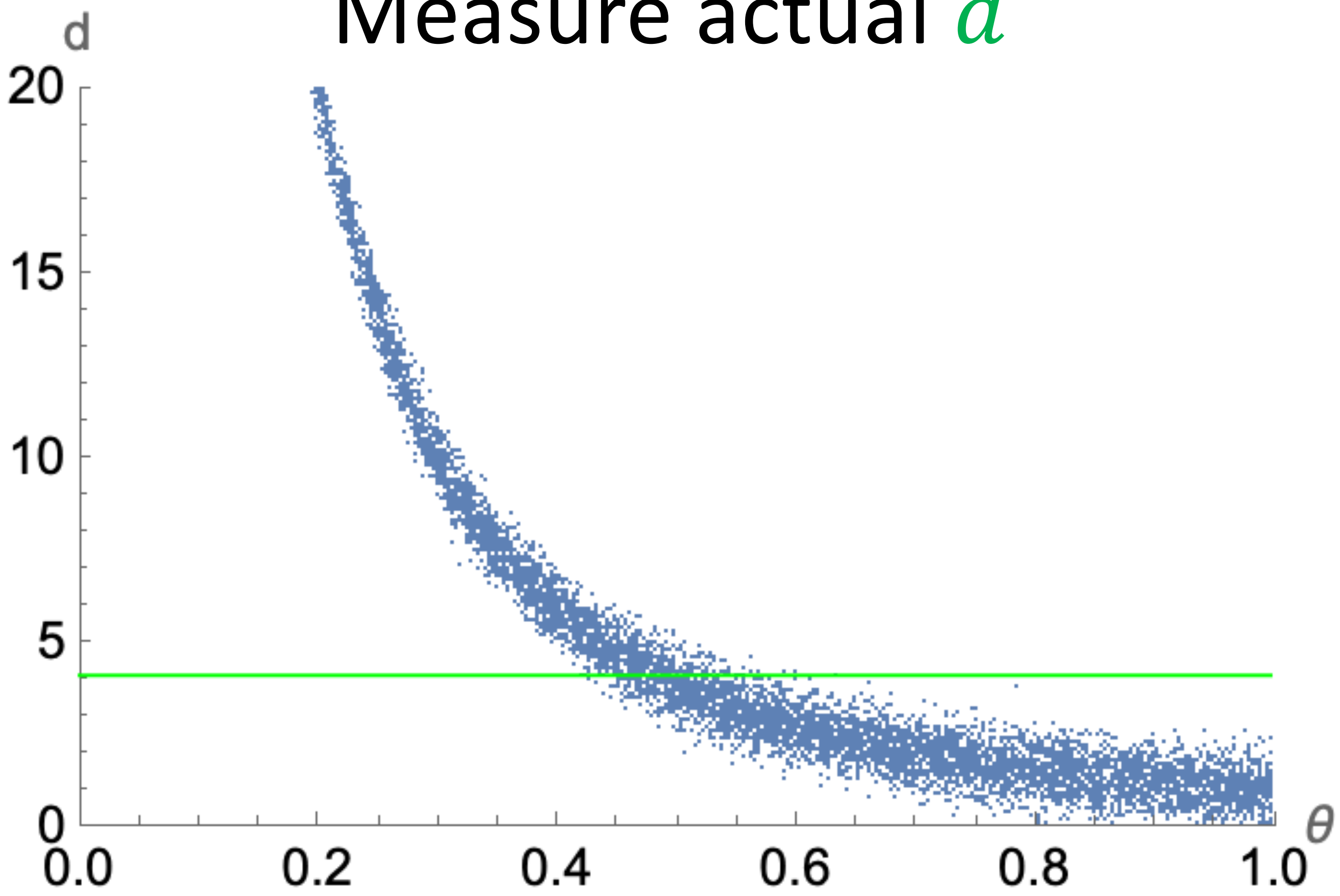
i.e.,draw from $\mathrm{p}(d|\theta)$



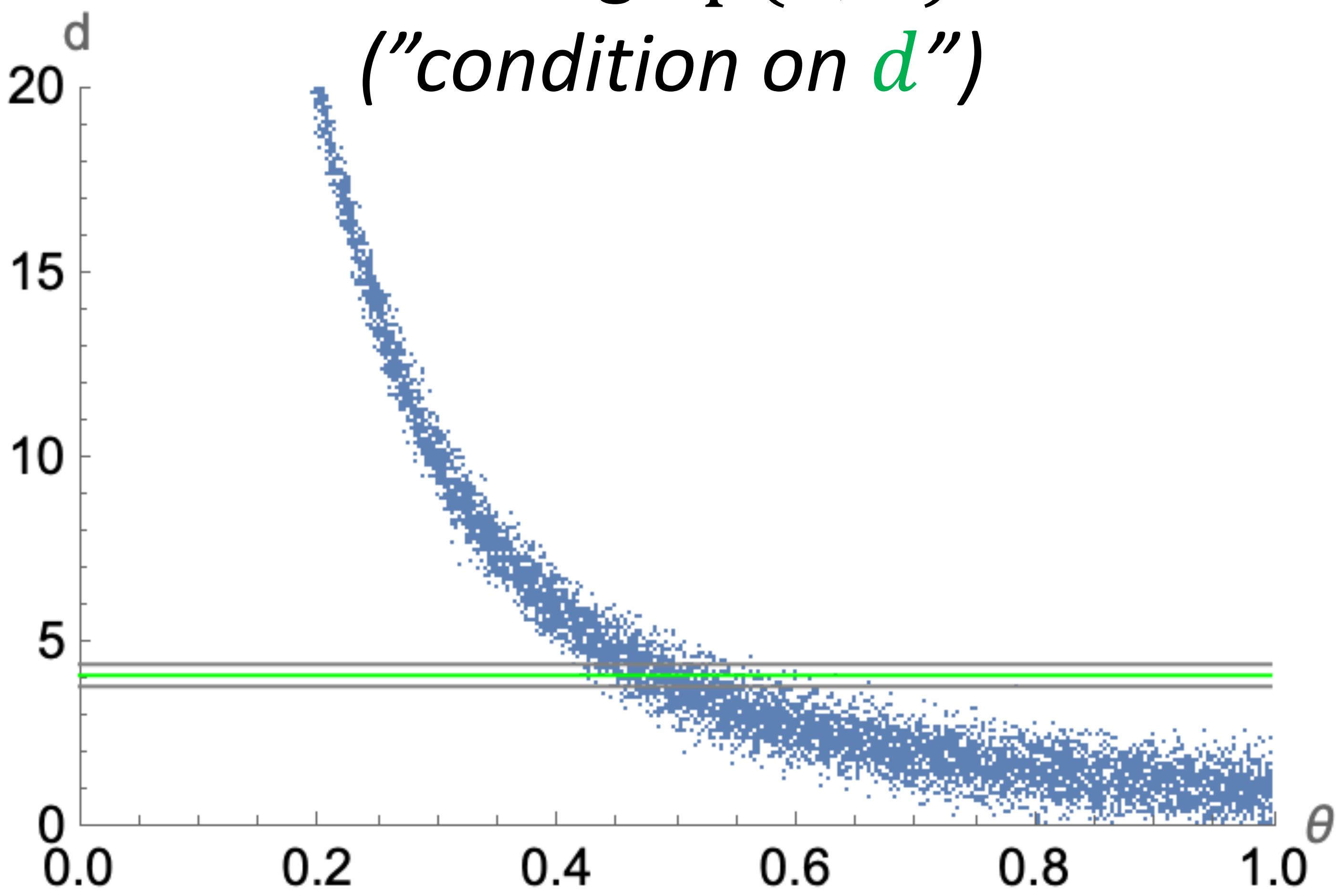
This is a sample from $\mathrm{p}(\theta, d)$



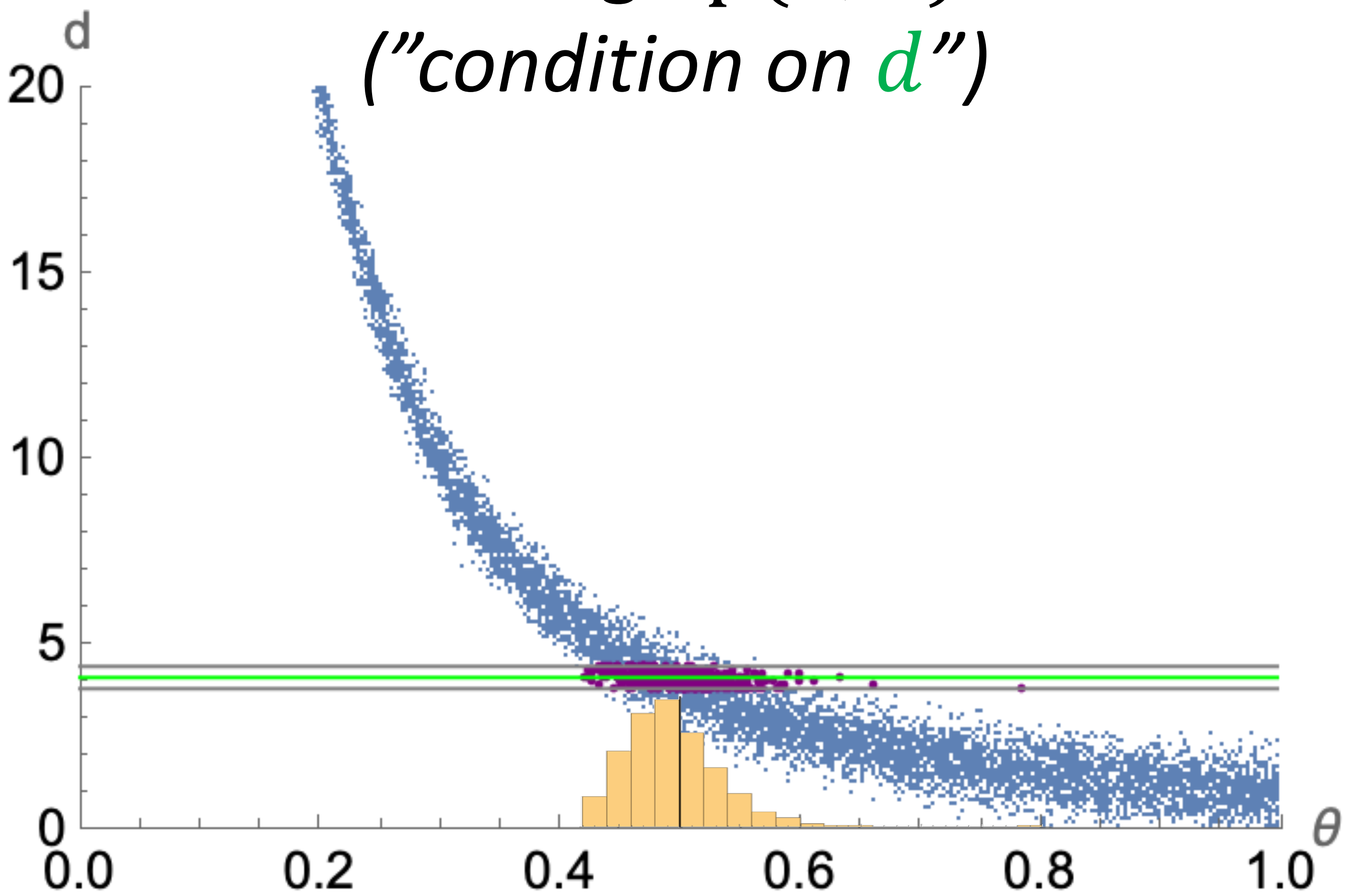
Measure actual d



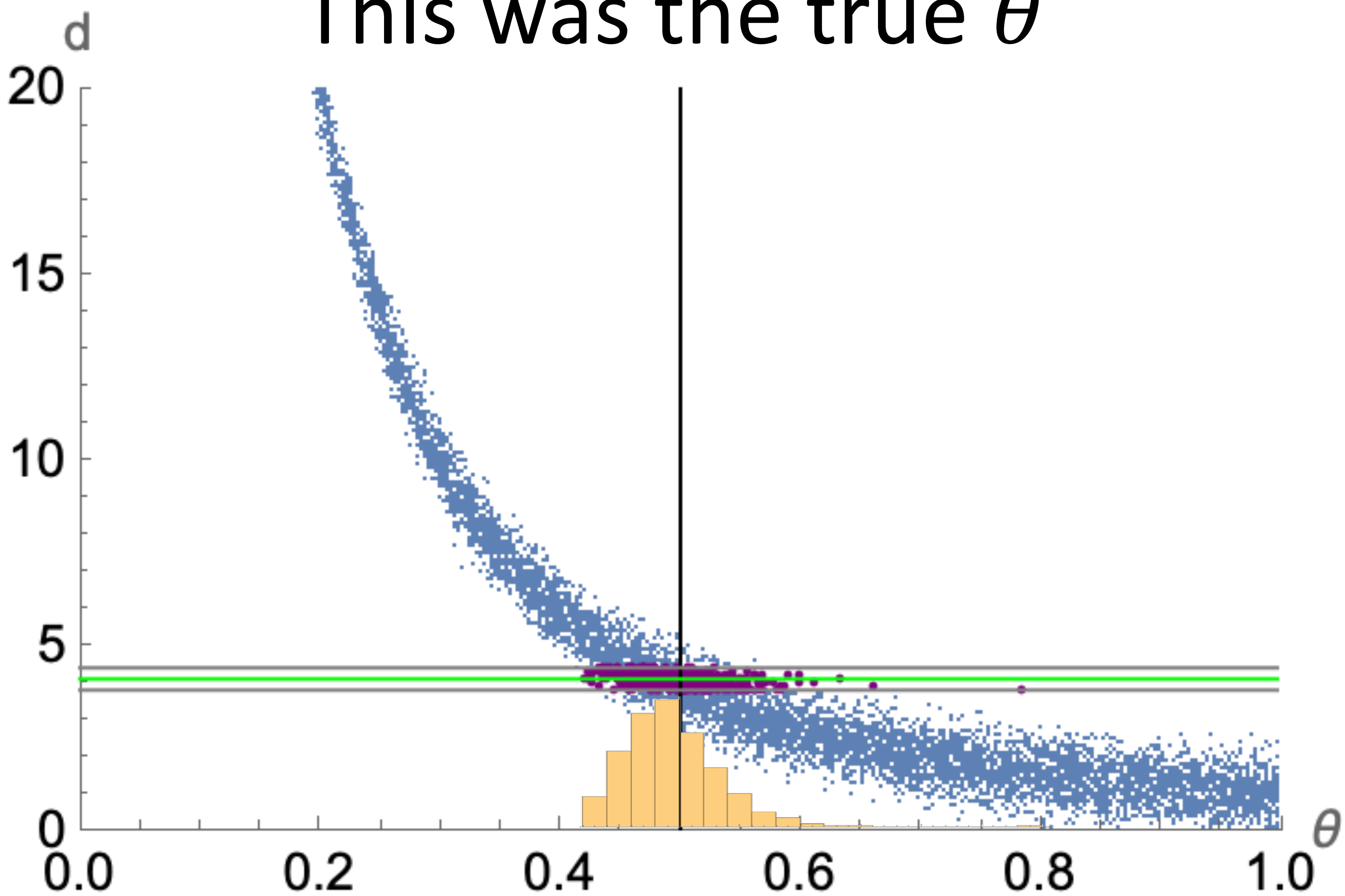
Slice through $\mathrm{p}(\theta, d)$ at d
("condition on d")



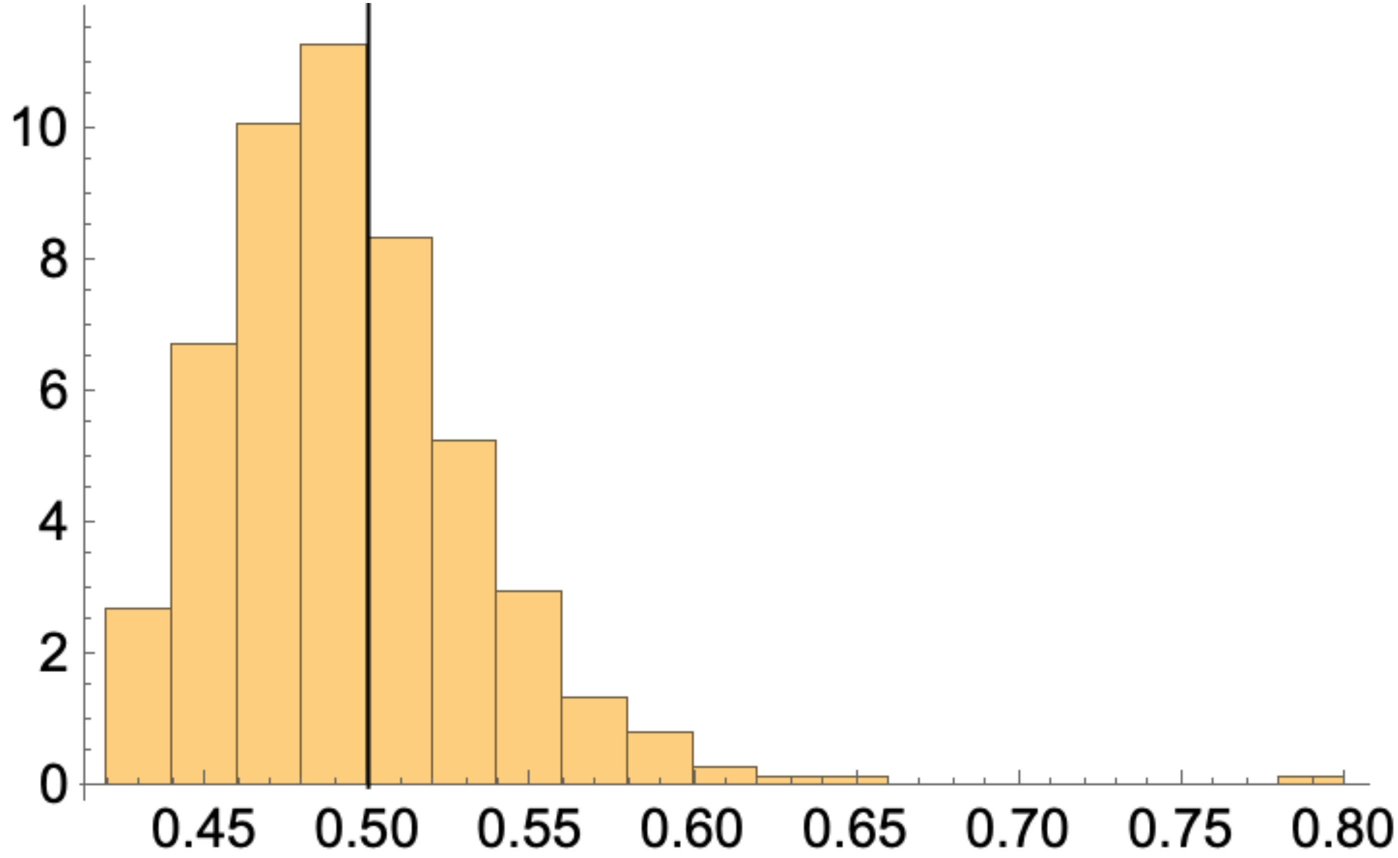
Slice through $\mathrm{p}(\theta, d)$ at d
("condition on d")



This was the true θ



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



Simulated data are nothing other than draws from the likelihood!

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

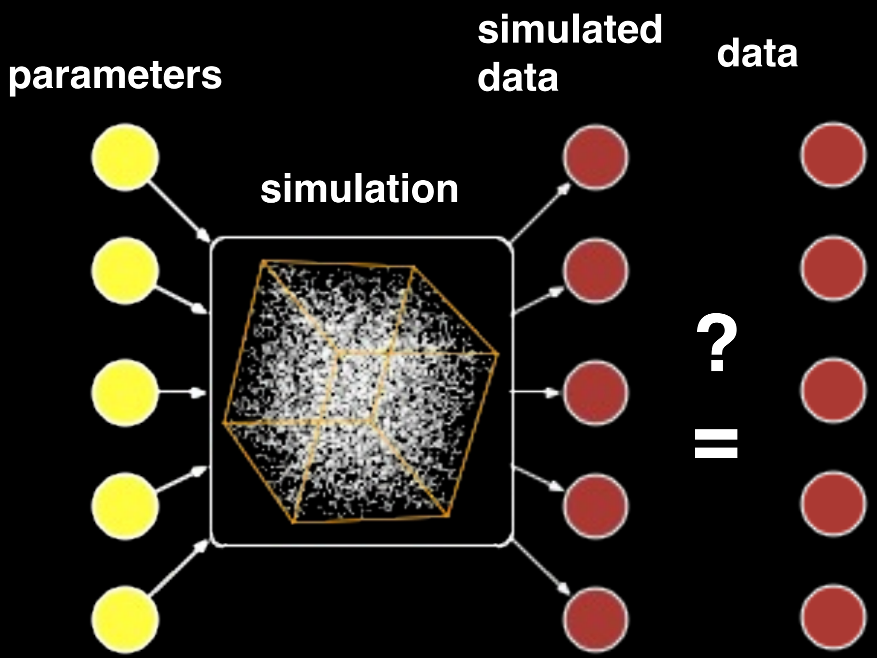
$$\mathbf{d}^* \leftarrow \mathrm{simulation}(\mathbf{d}^*|\boldsymbol{\theta})$$

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)

The ABC algorithm for implicit 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;

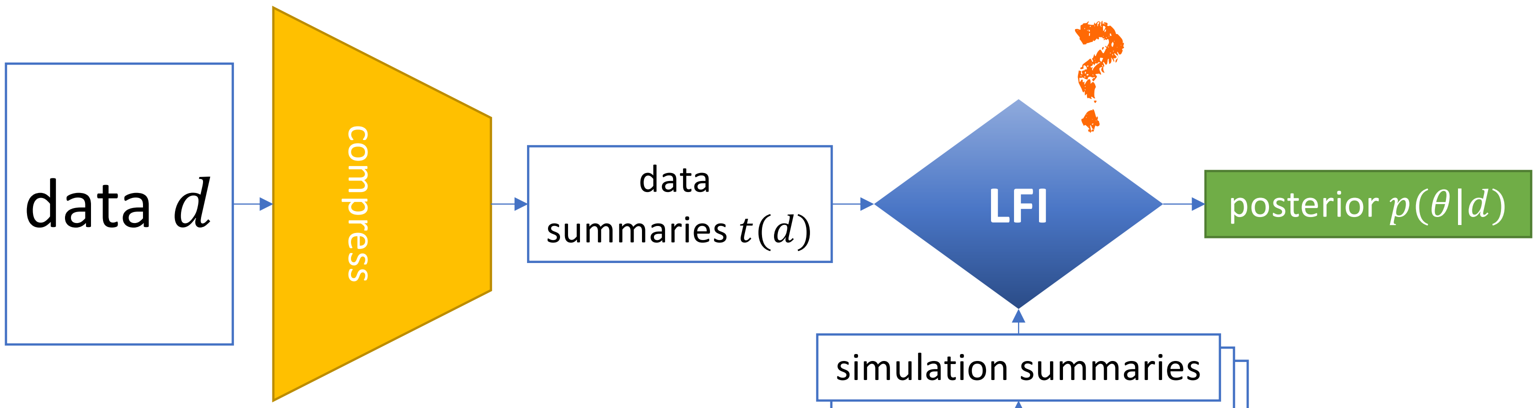
In the limit $\epsilon \rightarrow 0, \ \{\boldsymbol{\theta}\} \leftarrow P(\boldsymbol{\theta}|\mathbf{d})$

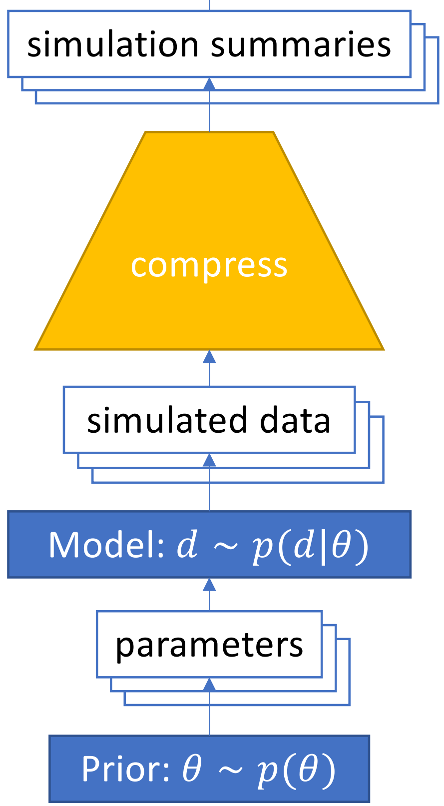
Benjamin Wandelt

Challenges for implicit inference

- The simplest version, ABC, becomes exponentially difficult when the data dimension is high.
    - Curse of dimensionality
- Need to find informative summaries of the data.
- Or find smarter approaches to solving the problem.

Implicit Inference with summaries





The statistical model is learned from simulations.

Alsing, Wandelt & Feeney
arXiv:1801.01497

Alsing, Charnock, Wandelt & Feeney
arXiv:1903.00007

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

Density Estimation Likelihood-Free Inference

Directly learn conditional probability density of compressed data given parameters, *e.g.*, in terms of a *mixture density* or a neural density estimator, *e.g.*, a *conditional normalizing flow.*

Alternative: likelihood-ratio estimation to do maximum likelihood estimation

LRE: Cranmer, Pavez & Louppe, arXiv1506.02169
NDE: Papamakarios, Murray + coauthors, arXiv:1605.06376, arXiv:1705.07057, arXiv:1805.07226
NDE: pyDELFI: Alsing, Feeney & Wandelt, arXiv: 1801.01497, arXiv:1903.01473
Truncated LRE: Miller et al arXiv:2107.01214, Cole et al arXiv:2111.08030

Density Estimation Likelihood-Free Inference

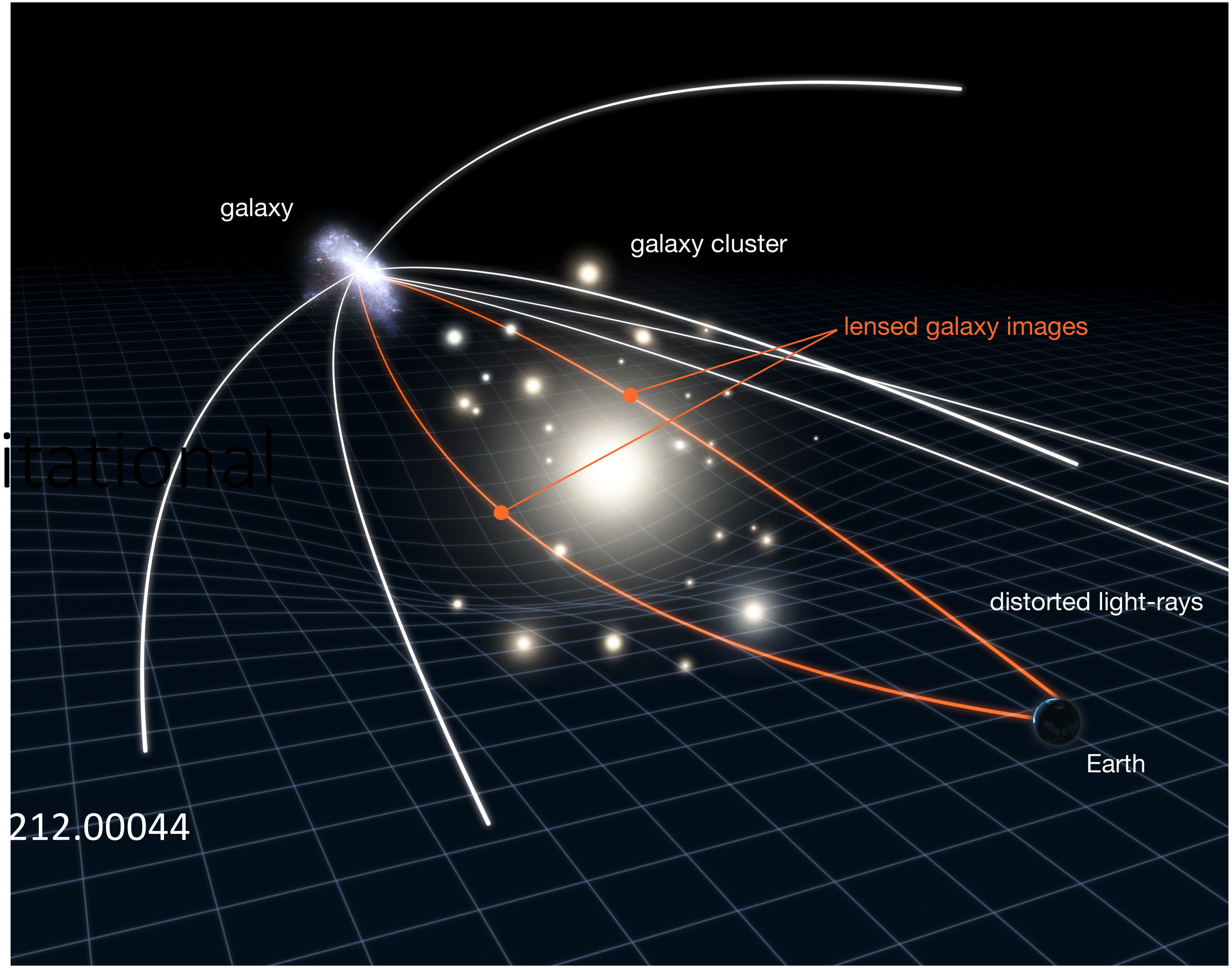
- *Nuisance-hardened* compression greatly reduces required number of simulations and allows many more parameters (Alsing & Wandelt arXiv:1903.01473).

- pyDELFI includes ensembles of conditional neural density estimators to fit the likelihood (Alsing, Charnock, Feeney, Wandelt arXiv:1903.00007)
  - Includes active learning for deciding where to run simulations
  - Also includes very fast, GPU-accelerated MCMC chains to explore posterior.

- SBI is a widely used package (but does not build ensembles out of the box)

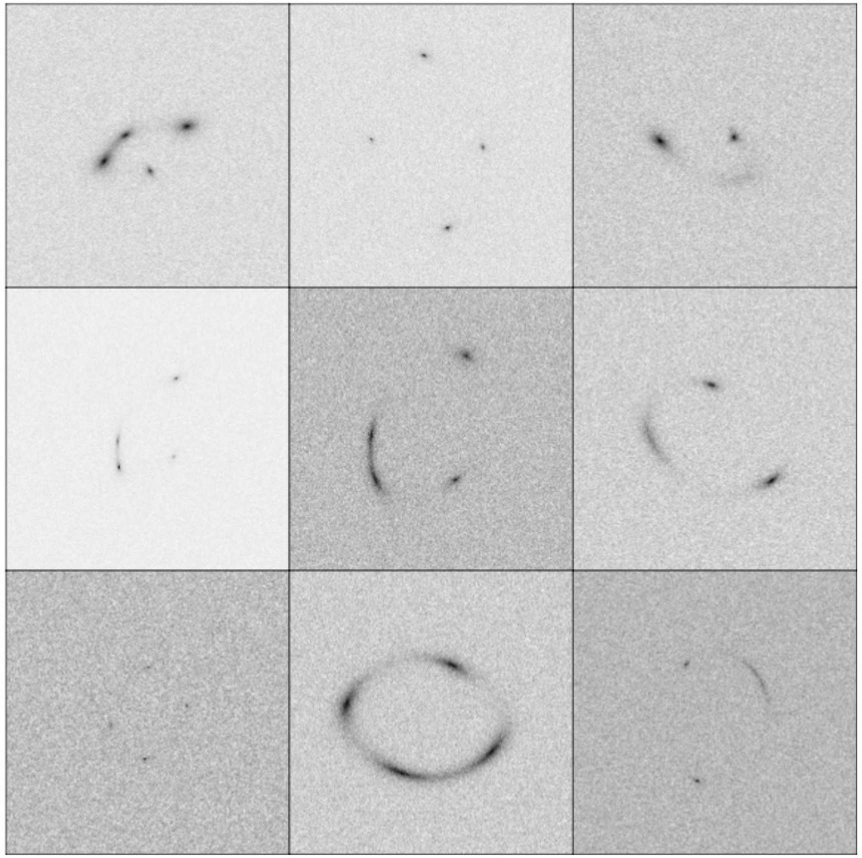
Example:

Strong Gravitational Lensing



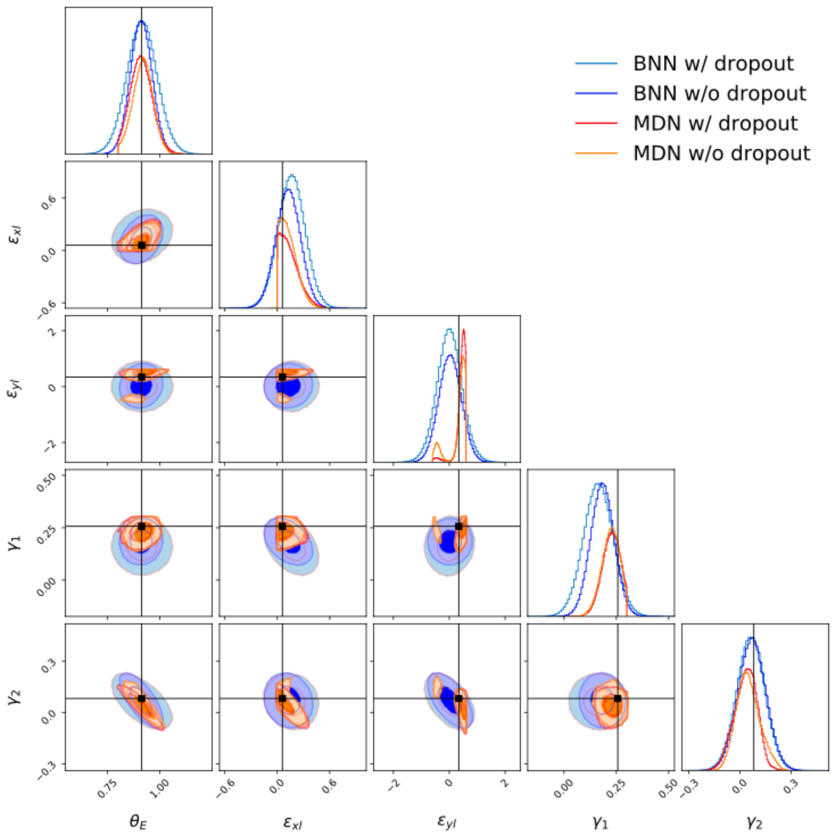
212.00044

Simulated images

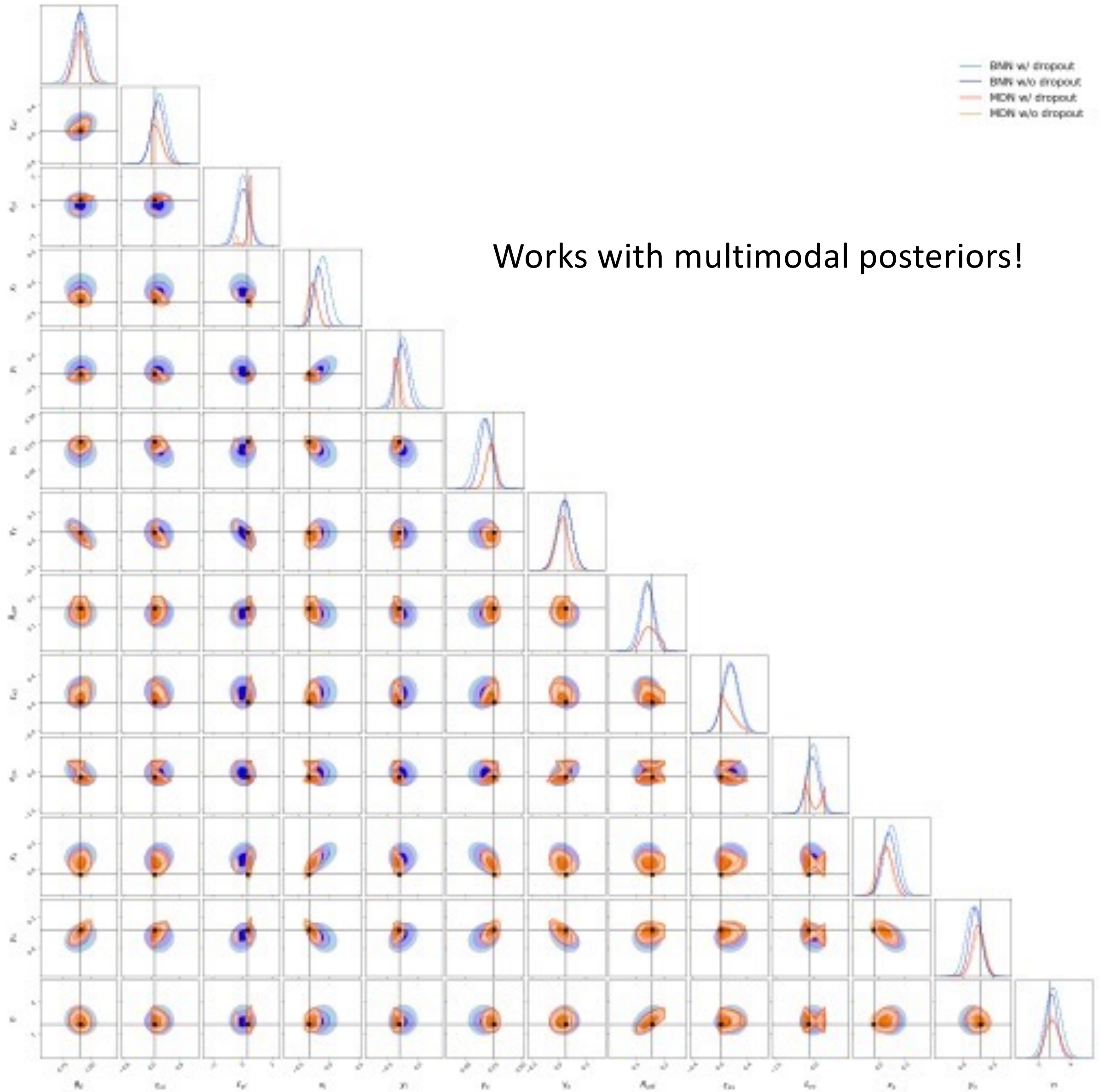


Legin et al arXiv 2212.00044

Inference

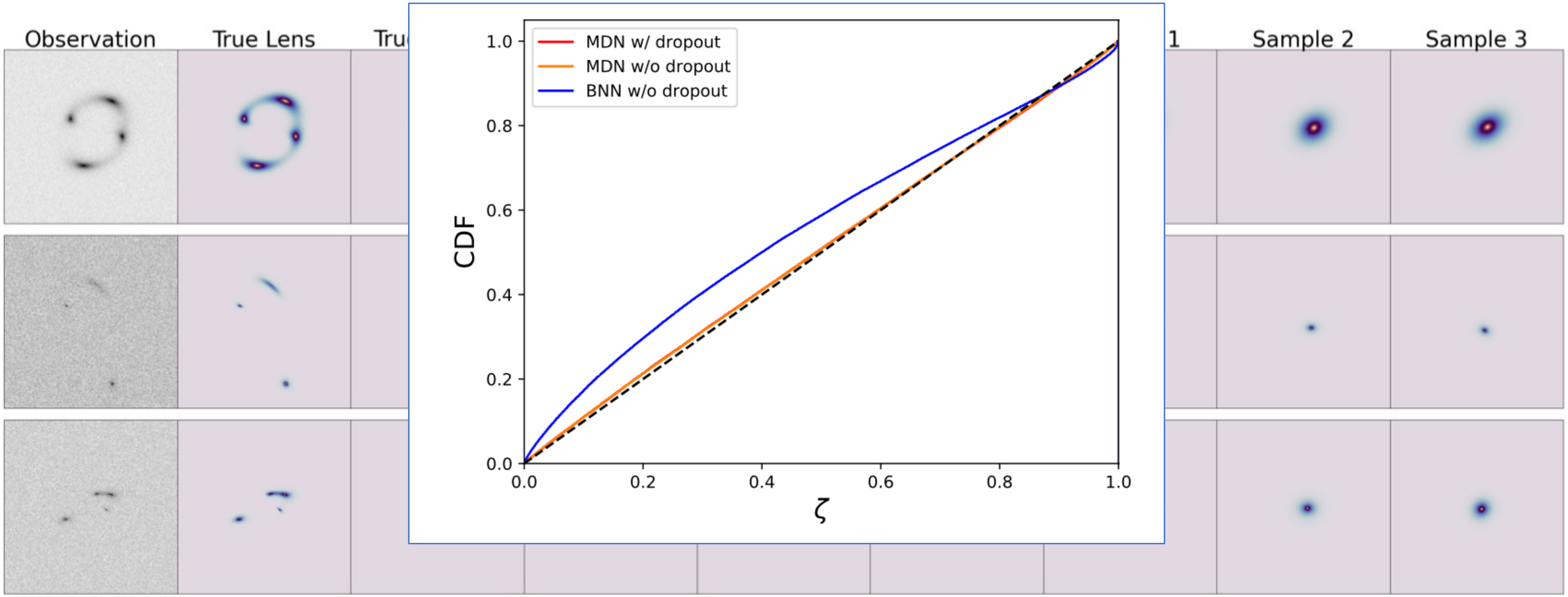


Legin et al arXiv 2212.00044



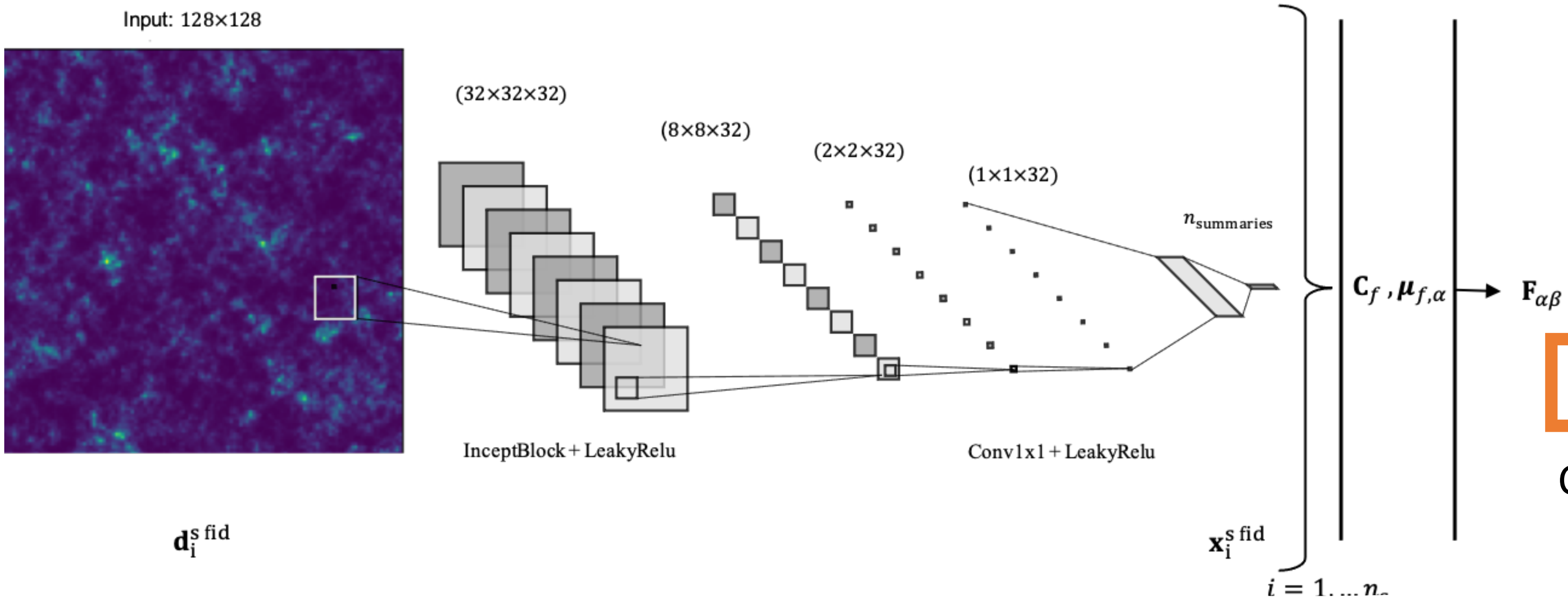
Works with multimodal posteriors!

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*

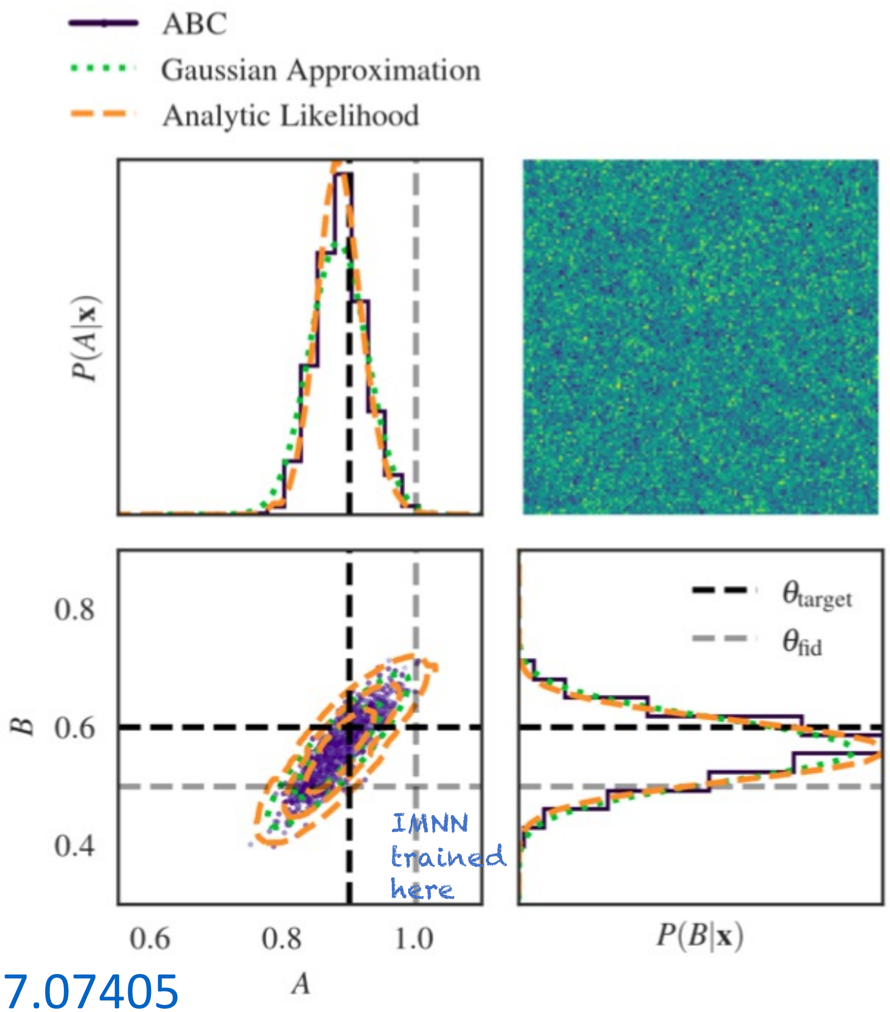


Optimal compression network minimizes information loss

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

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

IMNN recovers full info directly from the field



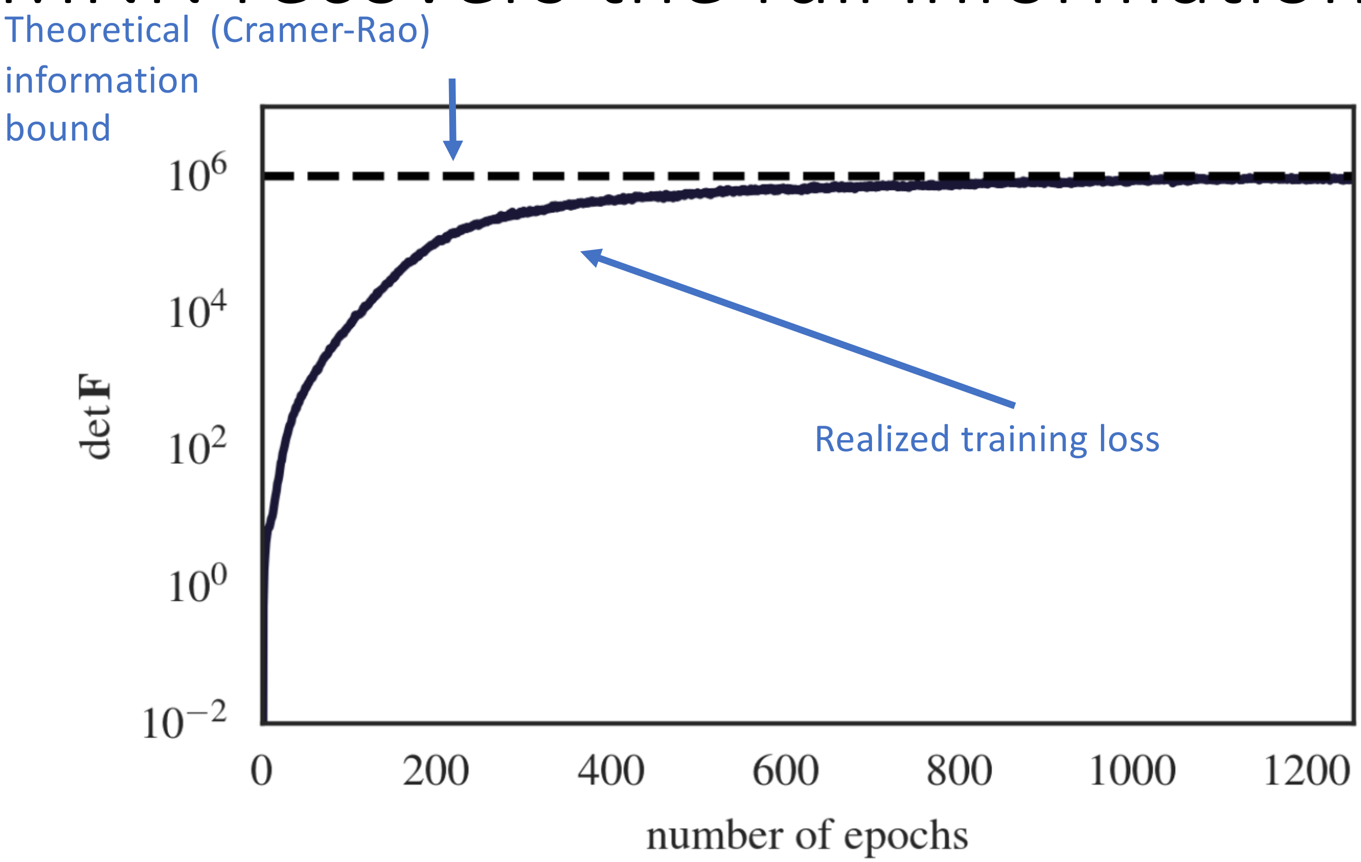
IMNN posterior

IMNN PML estimator

Exact likelihood

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

The IMNN recovers the full information

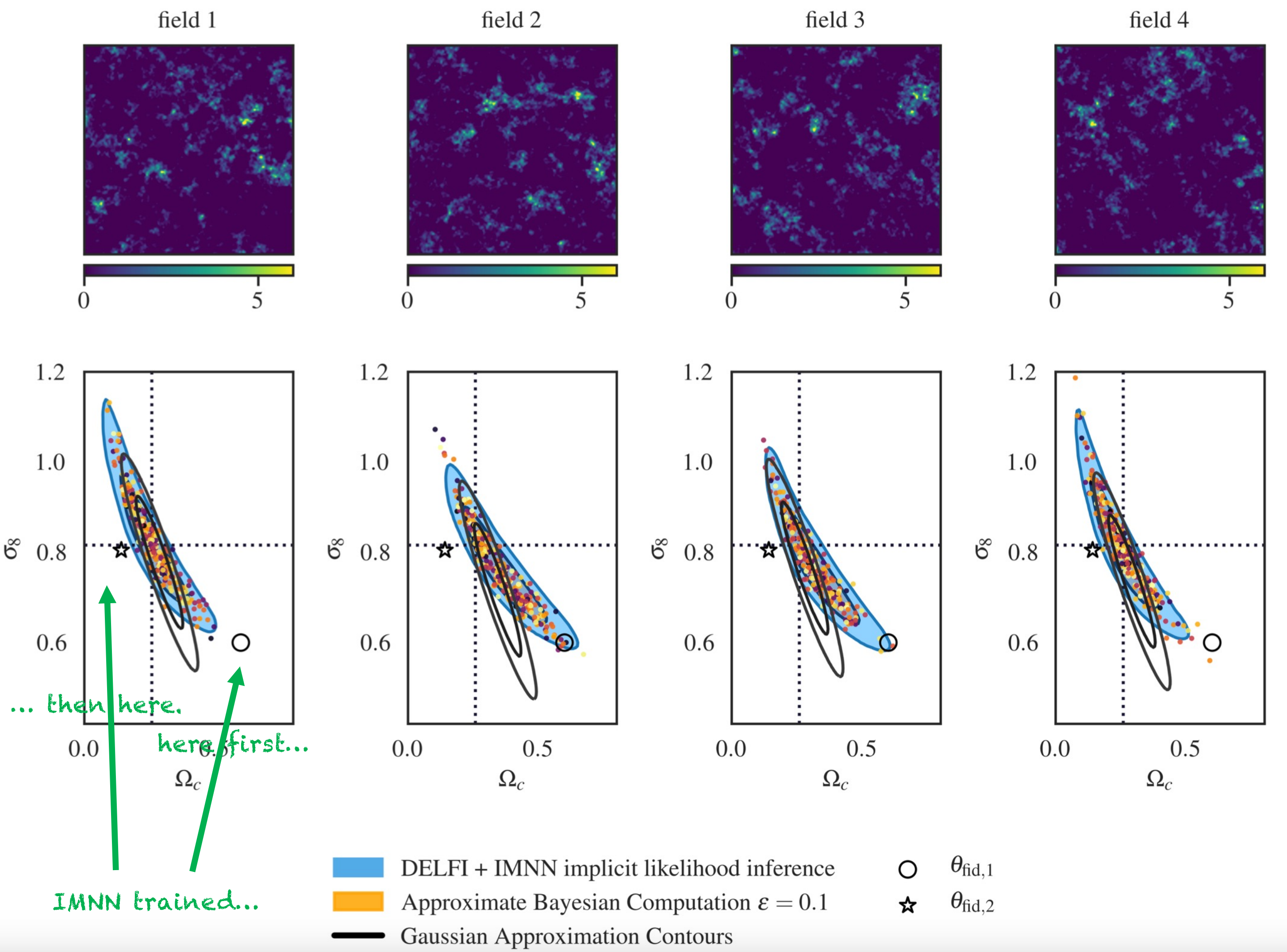


11 minutes on 1 GPU

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

Non-Gaussian field inference with IMNN and DELFI

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

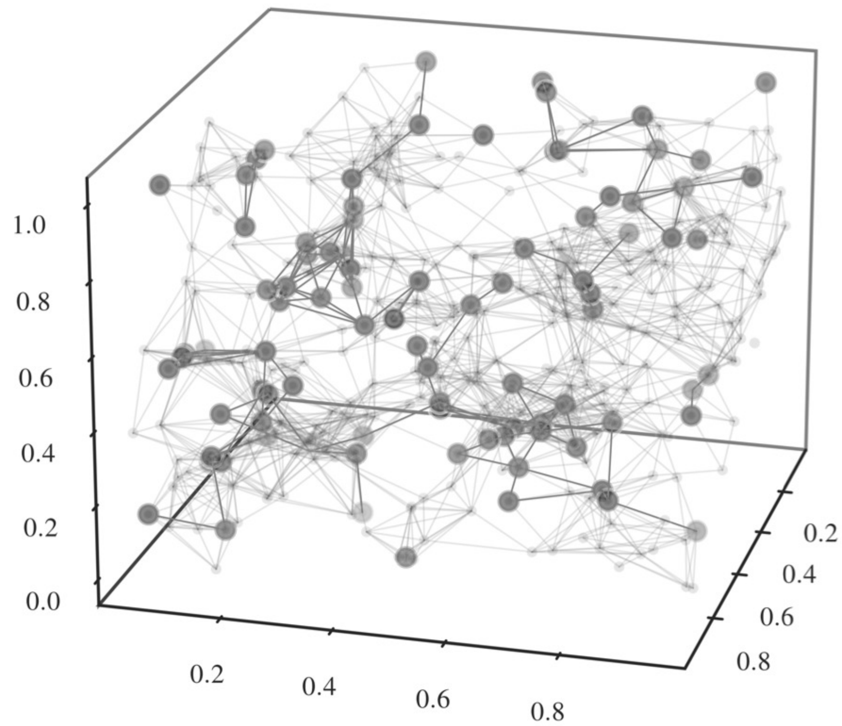


Makinen et al., arXiv:[2107.07405](2107.07405)

Can define Fisher information and score on distributions of graphs

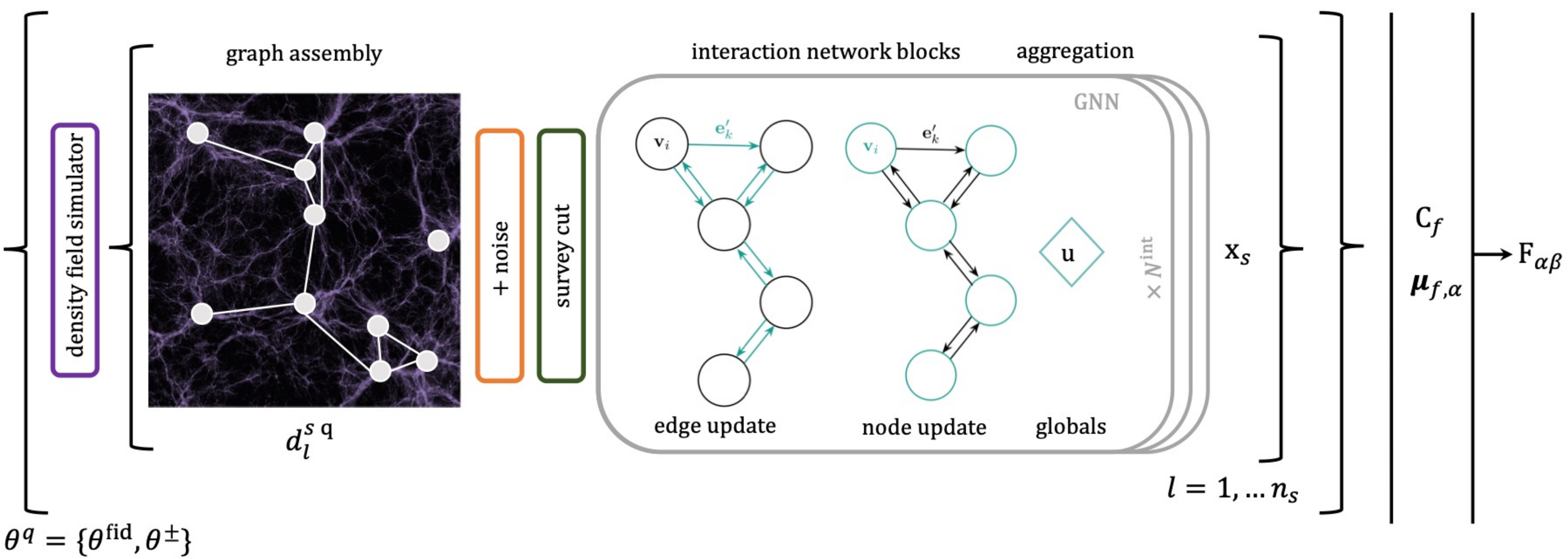
Example of using clusters of galaxies to infer cosmological parameters

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

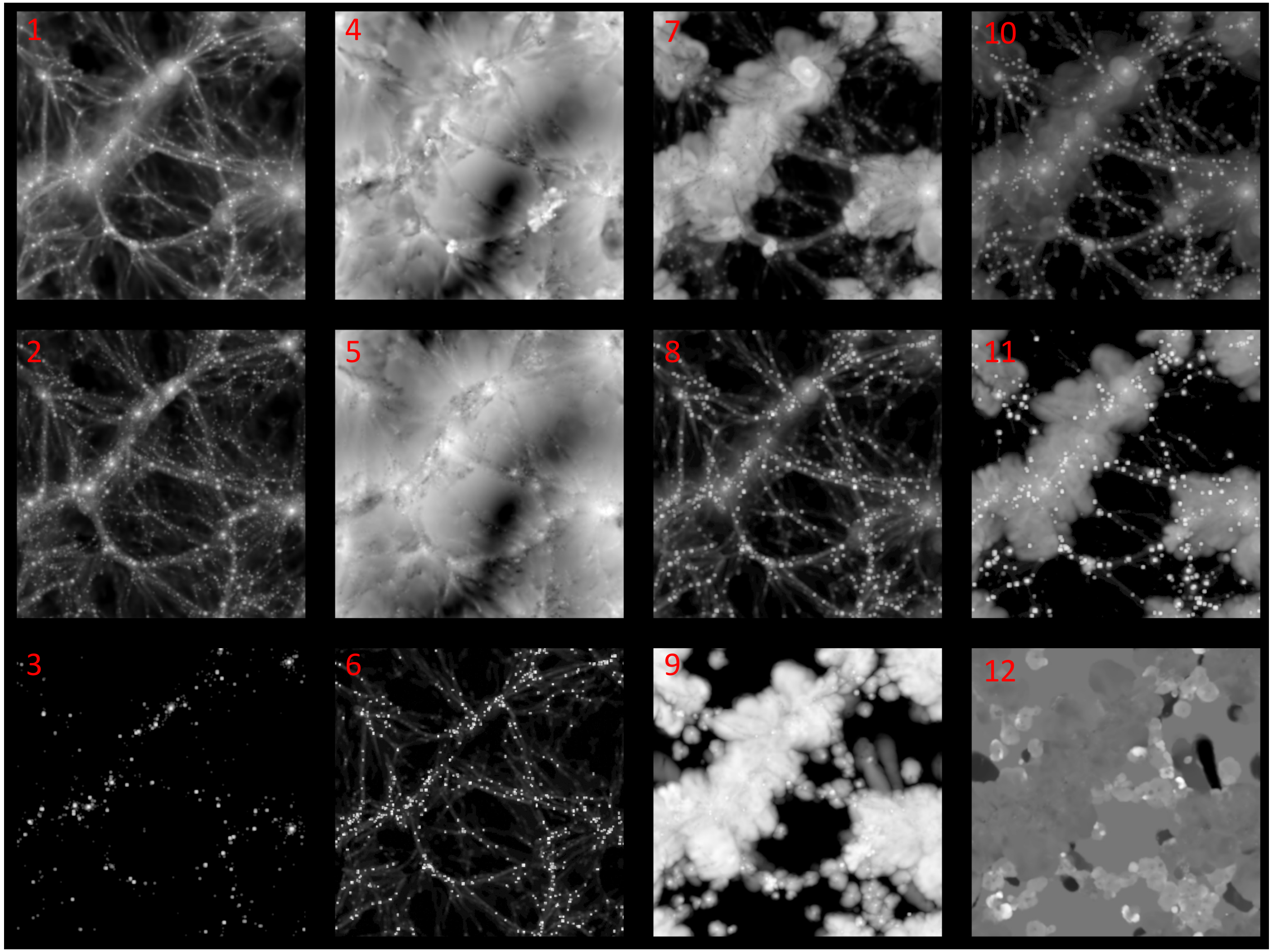
Cosmology on small scales with baryons

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 $(\mathrm{Mpc}/h)^2$
- 100 kpc/h resolution



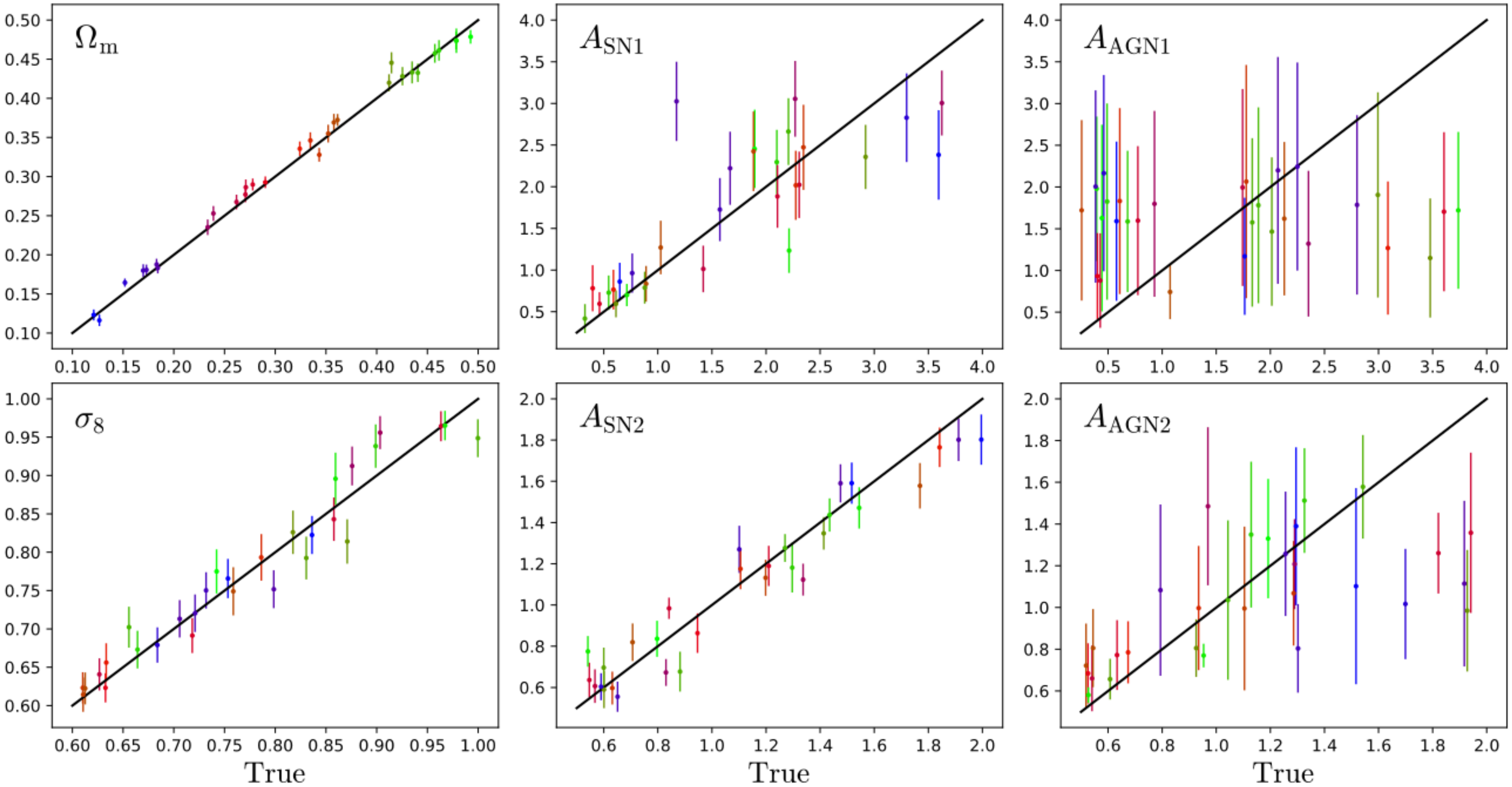
SBI: Cosmology from Small-Scale Hydro

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 $\mathcal{L}$



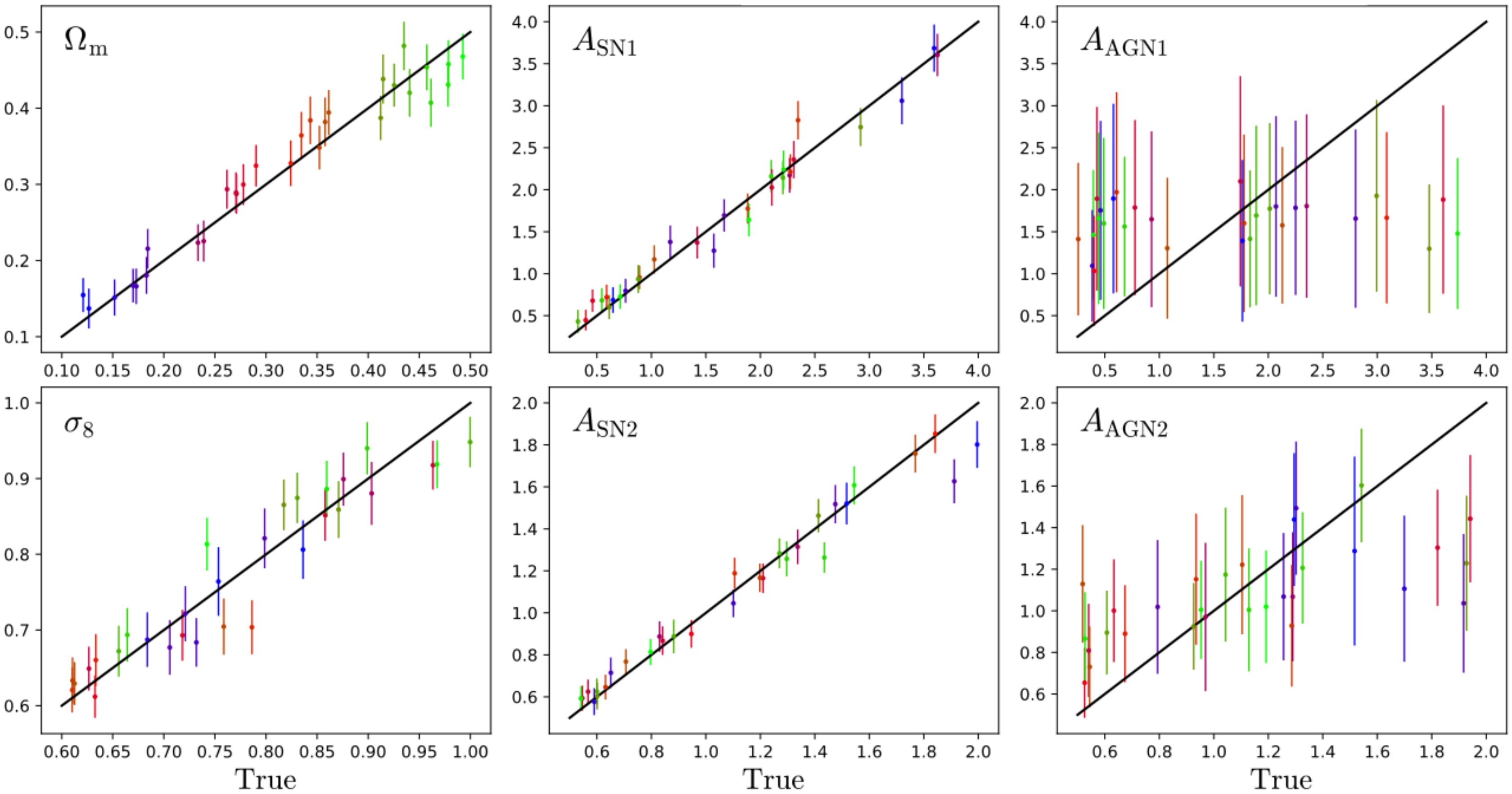
SBI: Cosmology from Small-Scale Hydro

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 $\mathcal{L}$



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

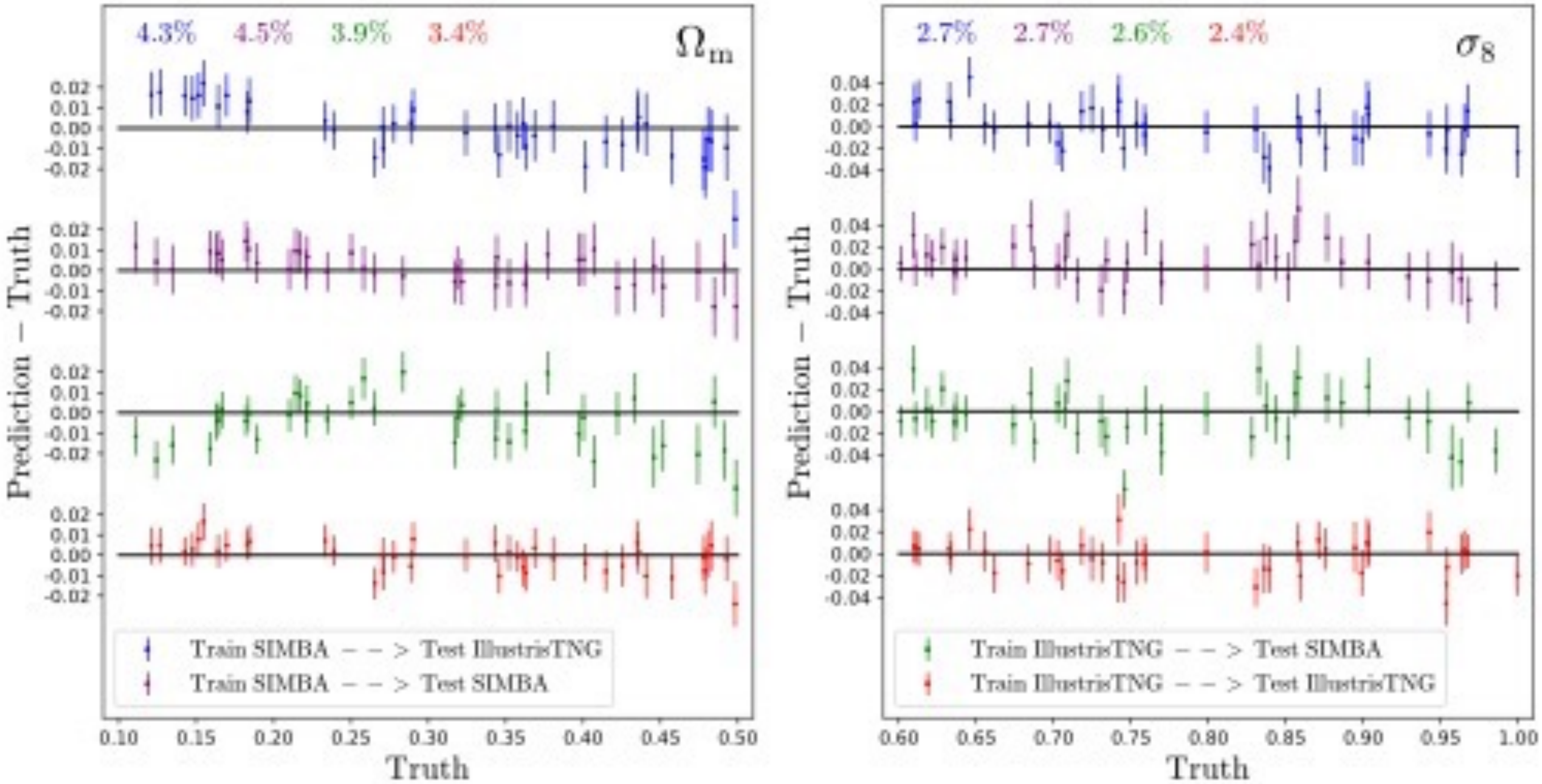
1. There is cosmological information on very small scales (100 kpc)

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

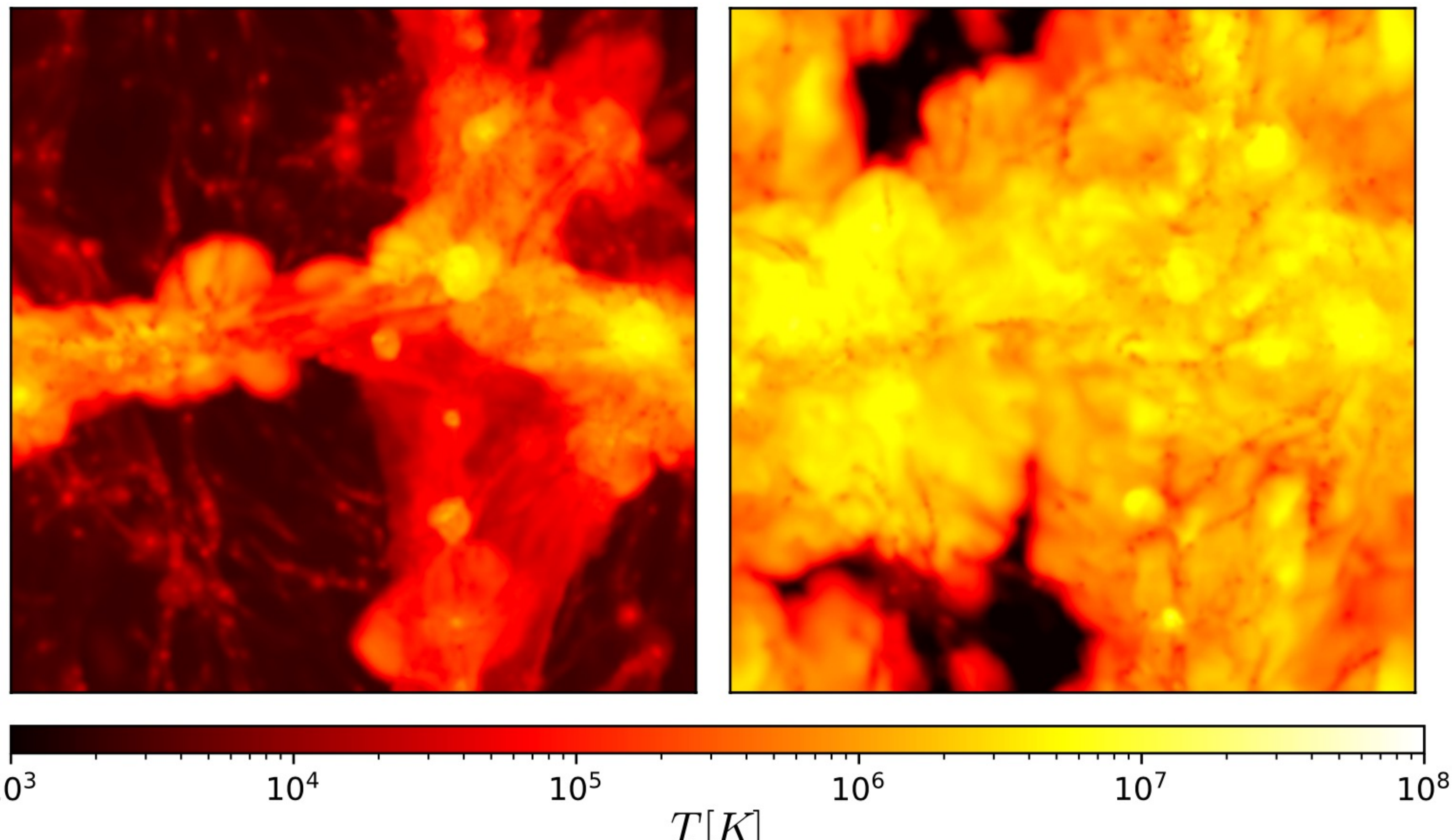
Cosmology robust to baryonic physics



Villaescusa-Navarro et al., arXiv:2109.10360

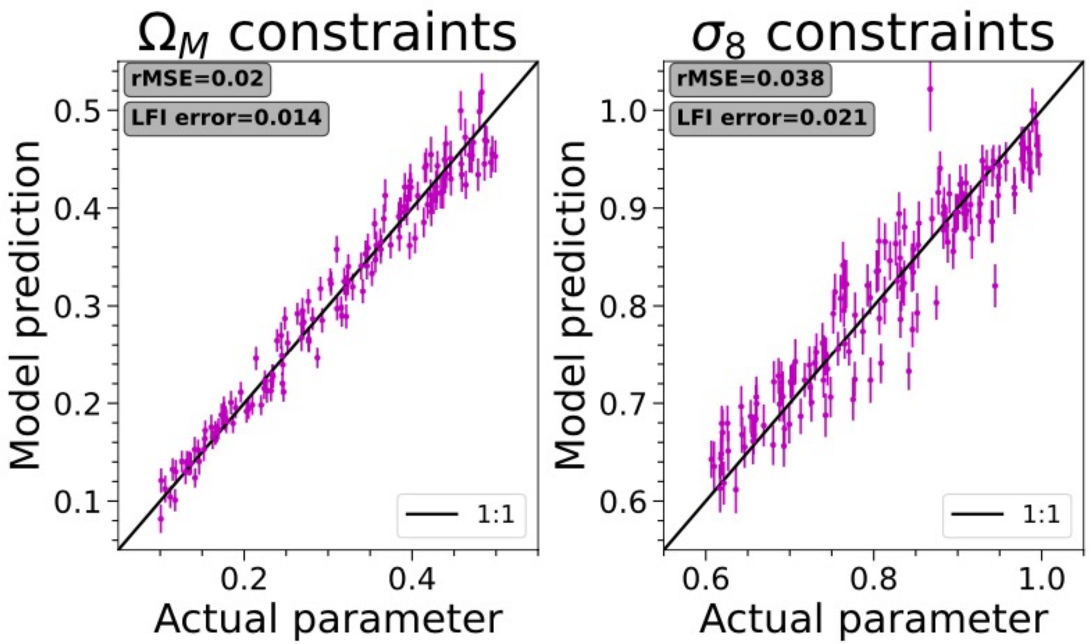
Villaescusa-Navarro et al., arXiv:2109.09747

Illustris TNG SIMBA



Same initial conditions!

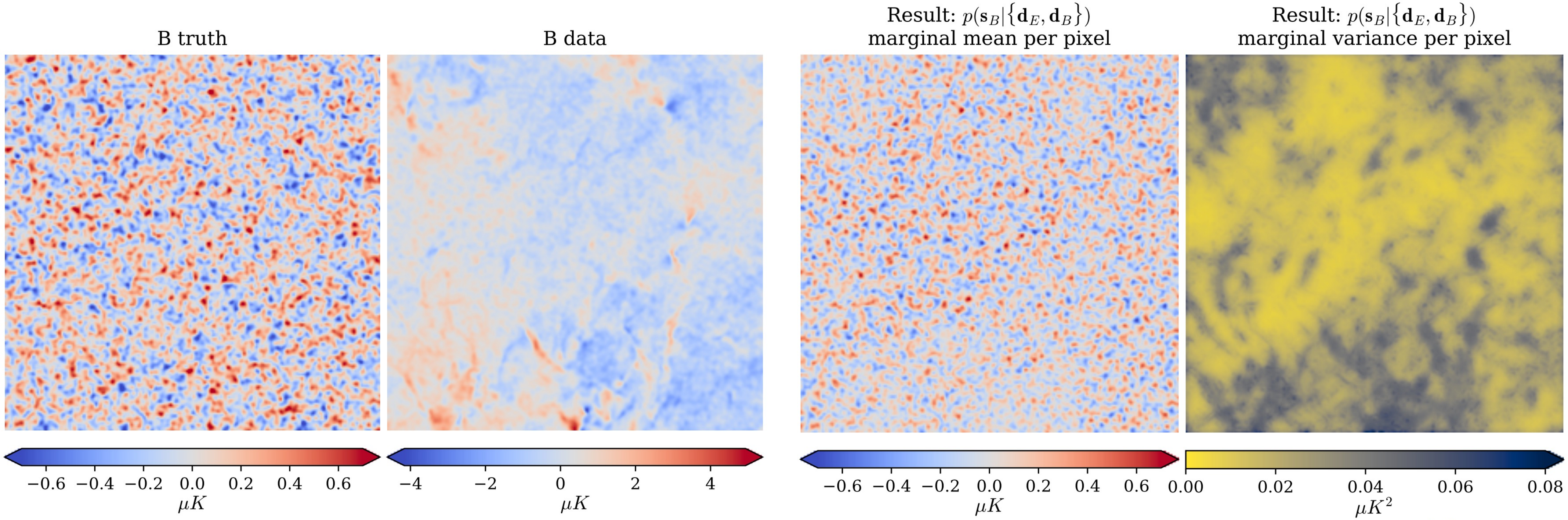
Implicit Likelihood Inference with moment networks from surveys generated with **semi-analytic galaxy formation models**



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

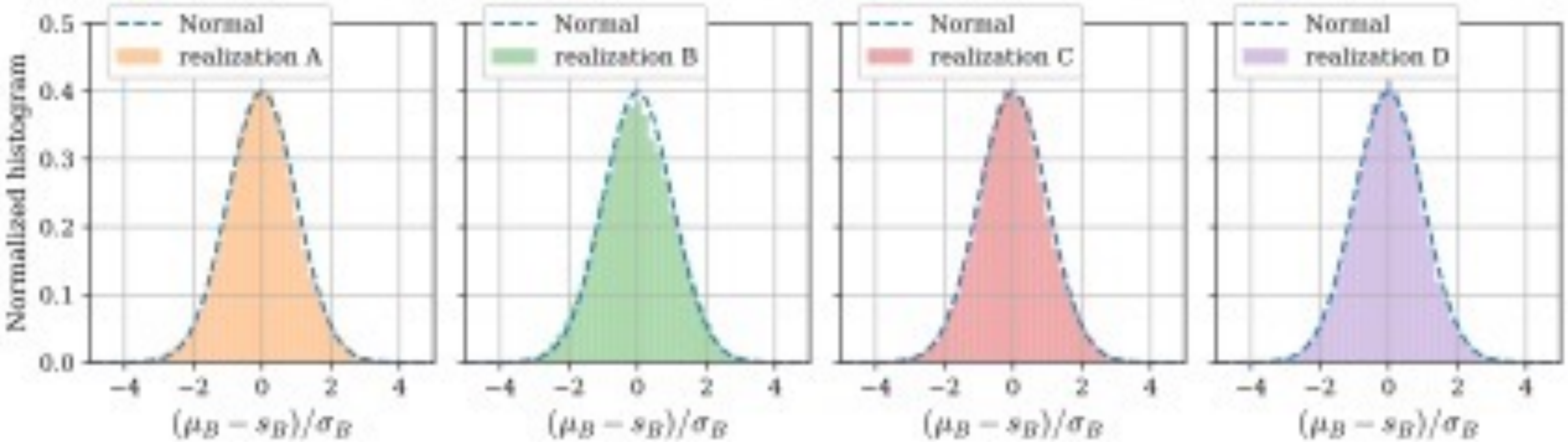
High dimensional application of Moment Networks (with **a single training image**)



Uses a generative model based on Wavelet Phase Harmonics

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

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 ***K***

Bayesian model comparison

Even if likelihood and posterior are explicitly given

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

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

$$\Longrightarrow P(d|M) = \int P(d|\theta,M)P(\theta|M)d\theta$$

Jeffrey & Wandelt, in prep

Evidence Networks: example loss construction

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

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

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:

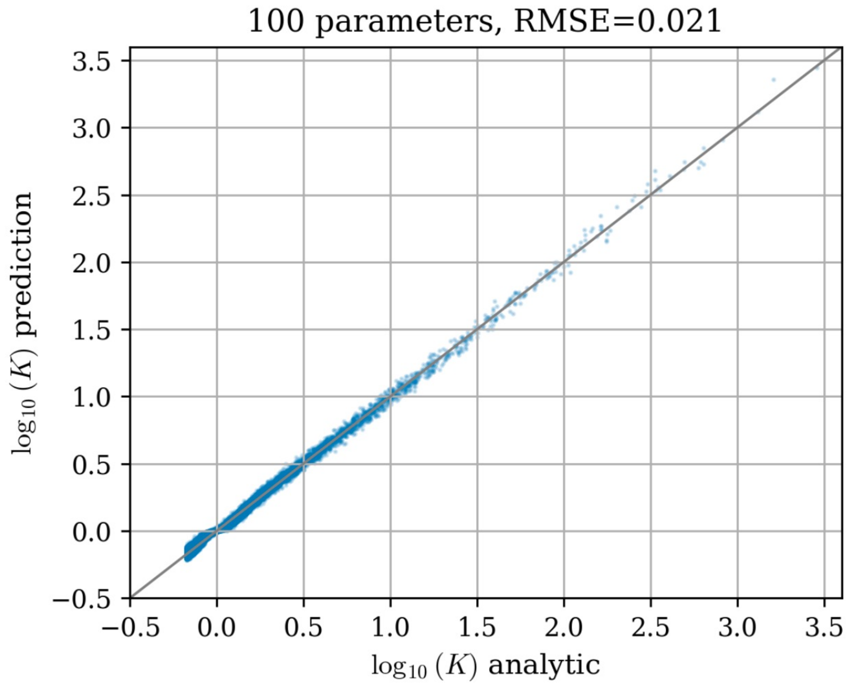
$$f_0(x) = \log\left(K\ \frac{p(M_1)}{p(M_0)}\right)$$

Jeffrey & Wandelt, in prep

Example: evidence ratio with 100 parameters

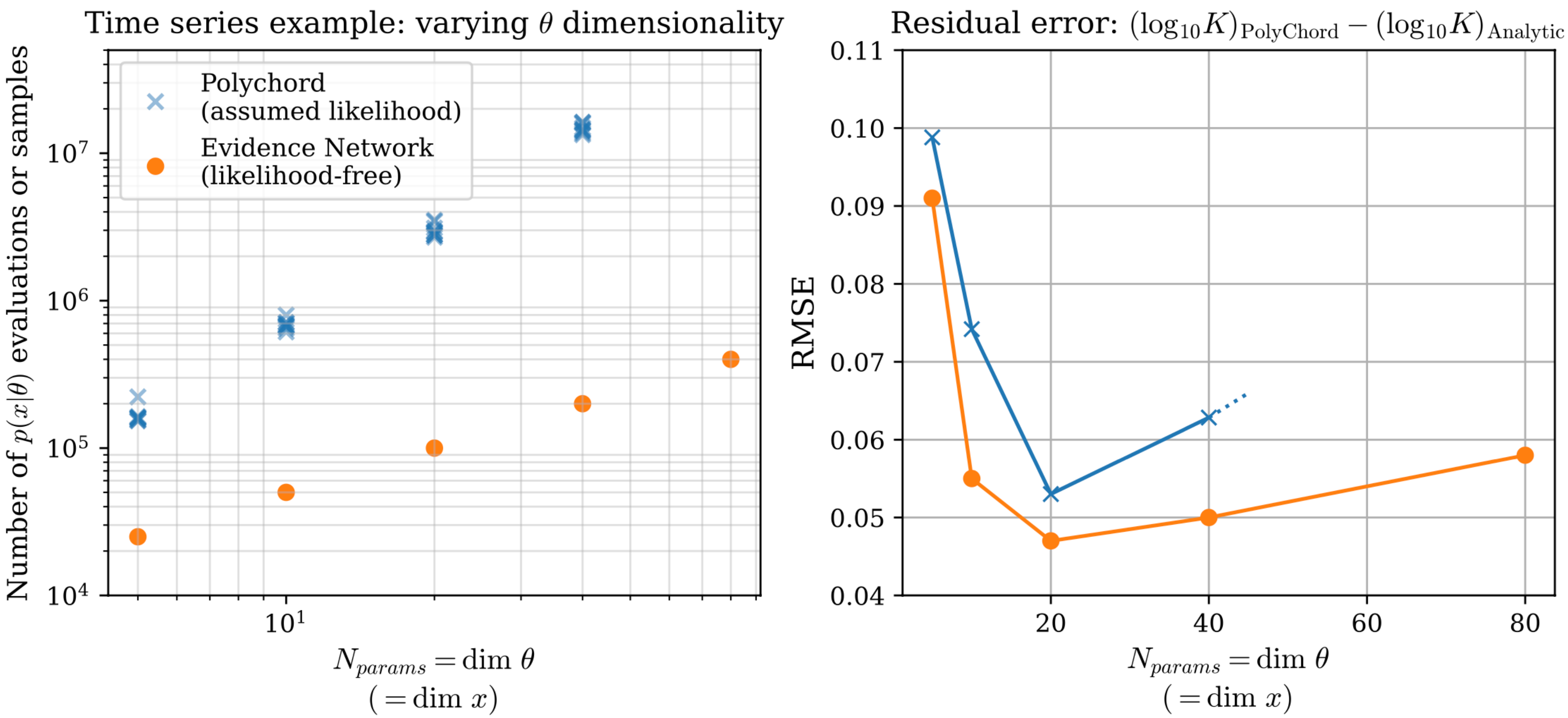
Evidence Networks with a variant of the exponential loss.

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



Jeffrey & Wandelt, in prep

Evidence nets: more accurate and faster than nested sampling



Computational cost of evidence network includes time to generate sims and train. Application to a given data set is nearly instantaneous.

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.

R. Legin *et al.*, arxiv:2304.03788

Score-based Diffusion model

- Consider a random walk of images
- Initialise with initial conditions
- Add Gaussian noise at every step
- This has an *attractor:* a Gaussian noise distribution

- Then sample by solving a series of inference problems to go from Gaussian noise back to a sample of the initial conditions
- If the number of steps is large enough, each step is a Gaussian inference problem!
- Train a neural network on simulations to learn the posterior mean for each of these steps

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

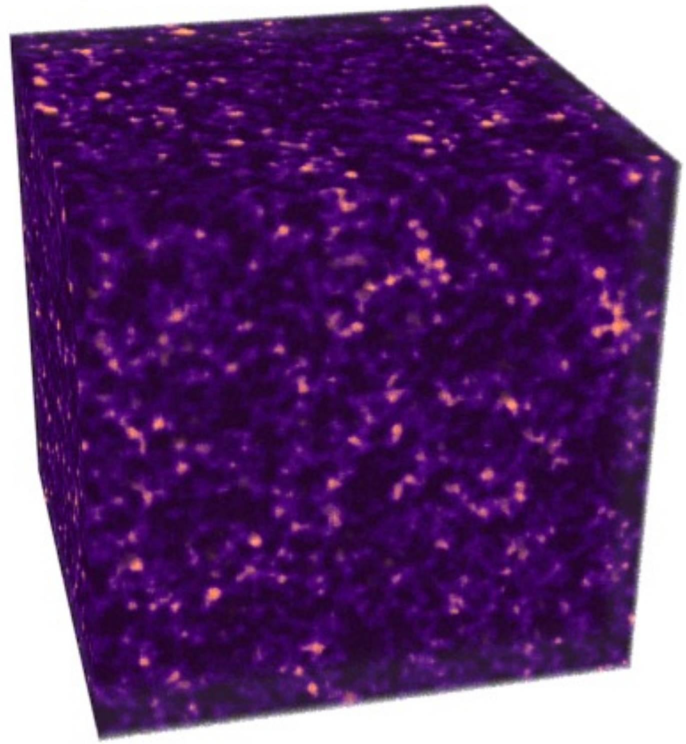
- Largest release of N-body simulation data to date
  - 43,100 full GADGET 3 simulations (1 Gpc)^3, 512^3 or 1024^3 particles
  - ~1 PB of data
- Goal: quantify statistics information content of non-Gaussian non-linear density field about cosmological parameters
- Includes full dark matter snapshots,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

Excellent tool for training machine learning surrogates.

Villaescusa-Navarro et al, arXiv:1909.05273

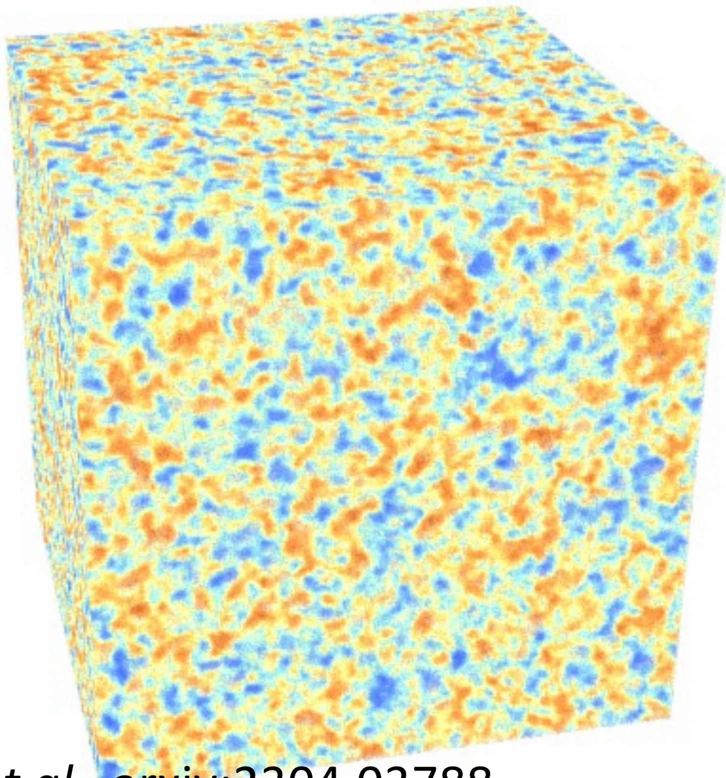
First full-field inference of initial conditions from fully non-linear density field

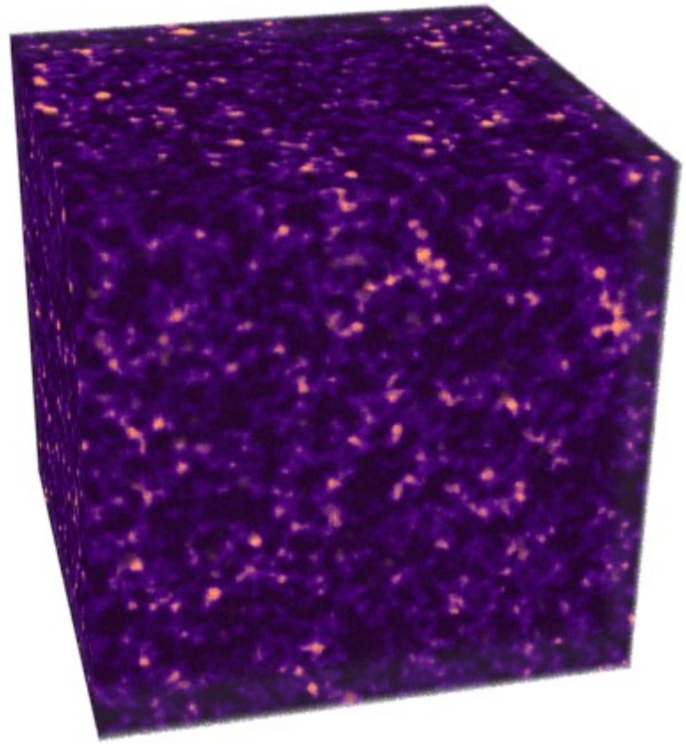
- 1 Gpc GADGET1024^3 simulation at z=0
- Binned on 128^3 grid



R. Legin *et al.*, arxiv:2304.03788

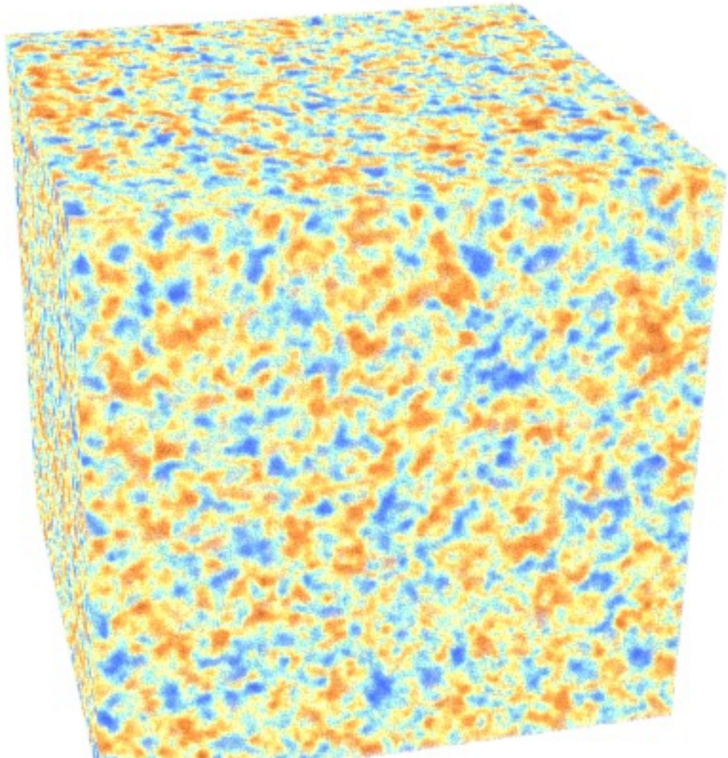
First full-field inference of initial conditions from fully non-linear density field

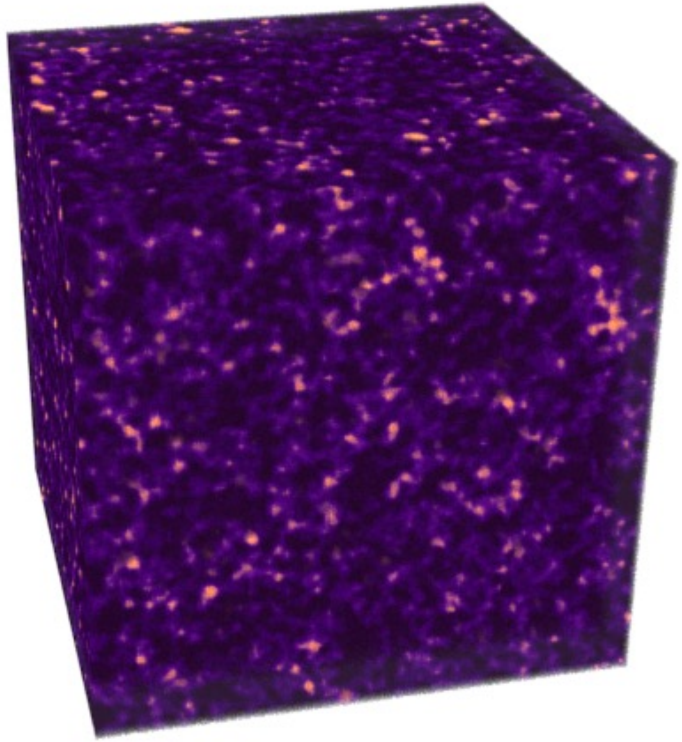




R. Legin *et al.*, arxiv:2304.03788

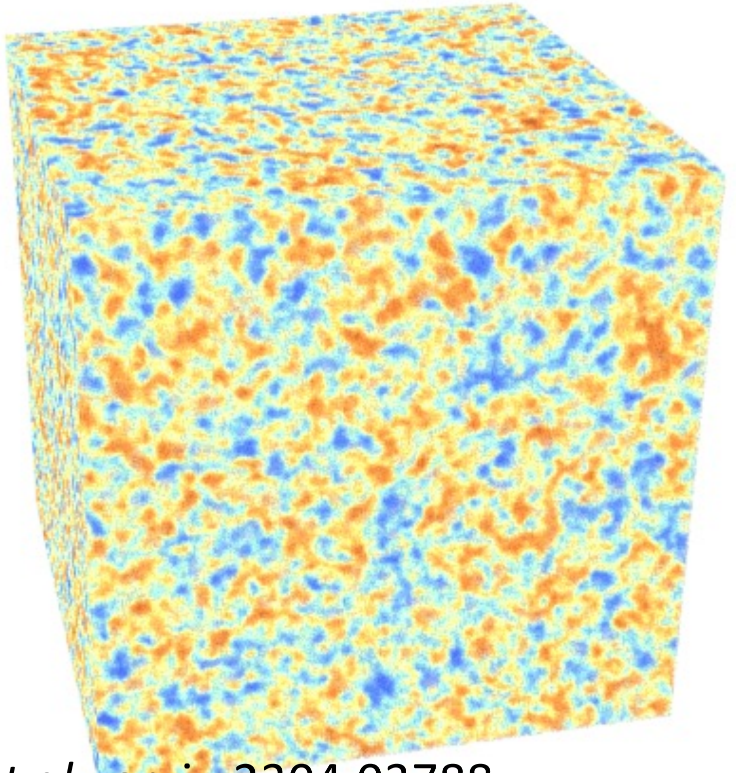
First full-field inference of initial conditions from fully non-linear density field

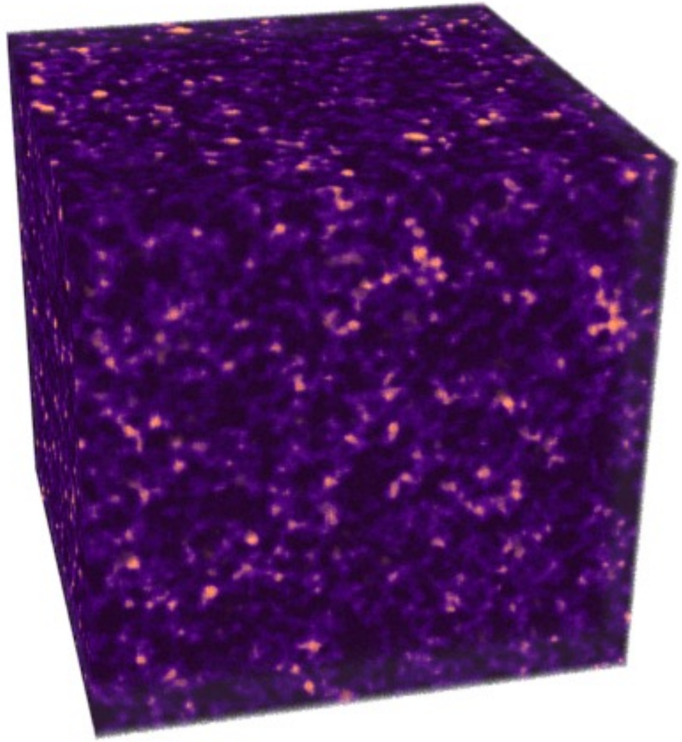




R. Legin *et al.*, arxiv:2304.03788

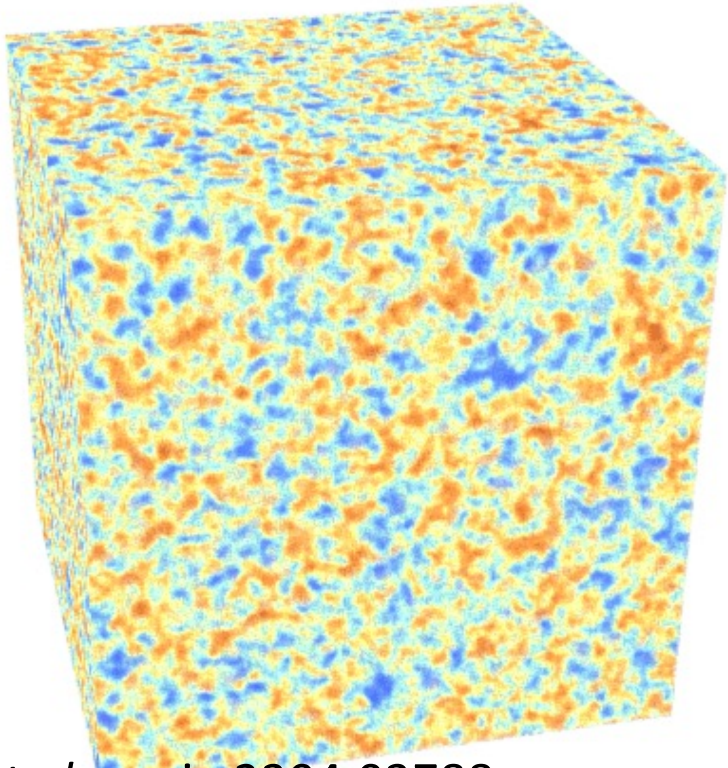
First full-field inference of initial conditions from fully non-linear density field

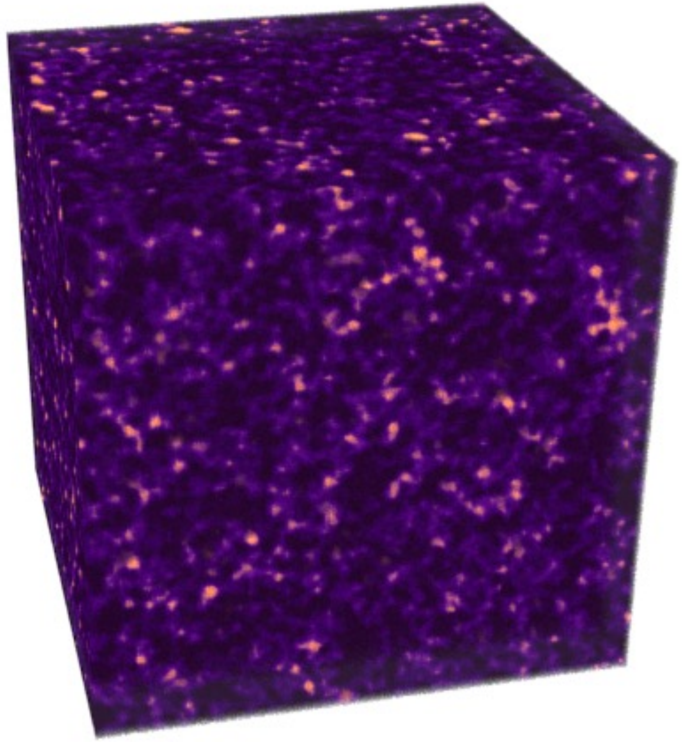




R. Legin *et al.*, arxiv:2304.03788

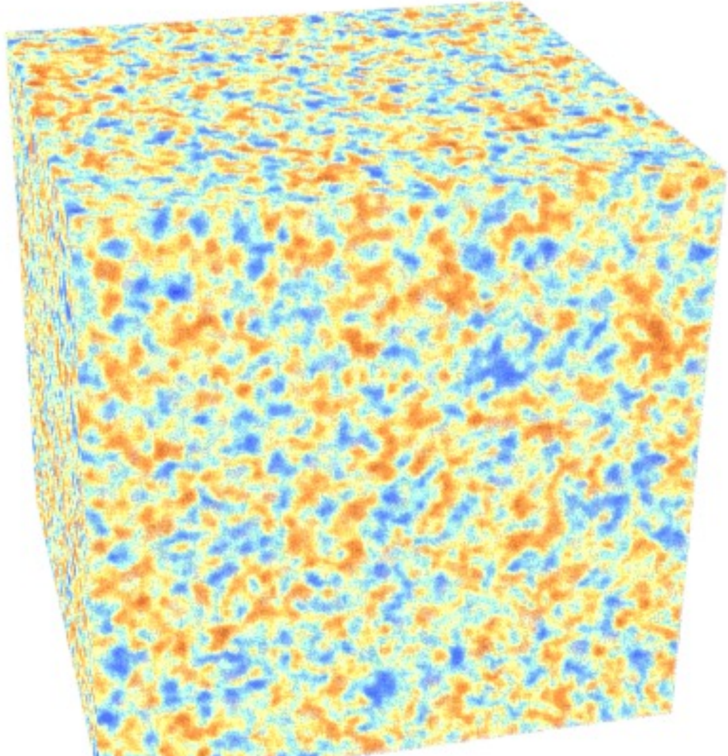
First full-field inference of initial conditions from fully non-linear density field

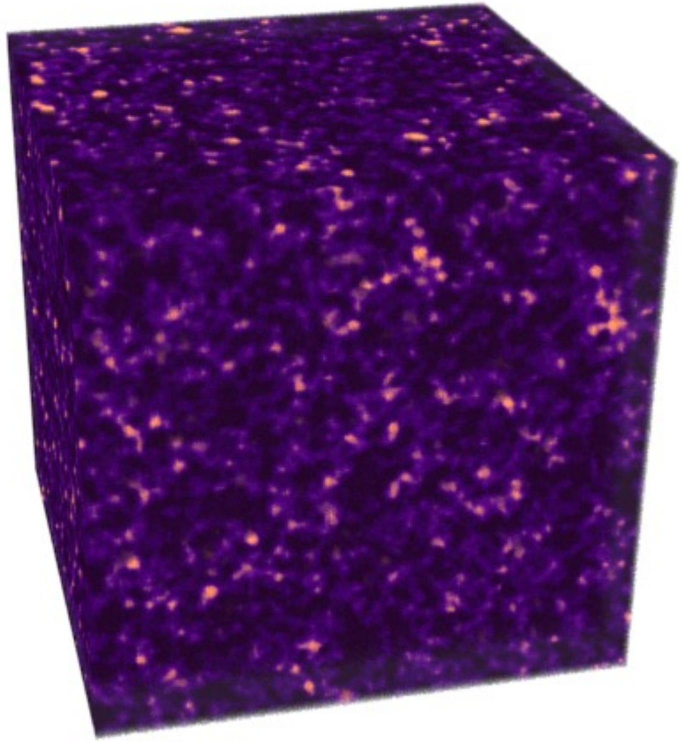




R. Legin *et al.*, arxiv:2304.03788

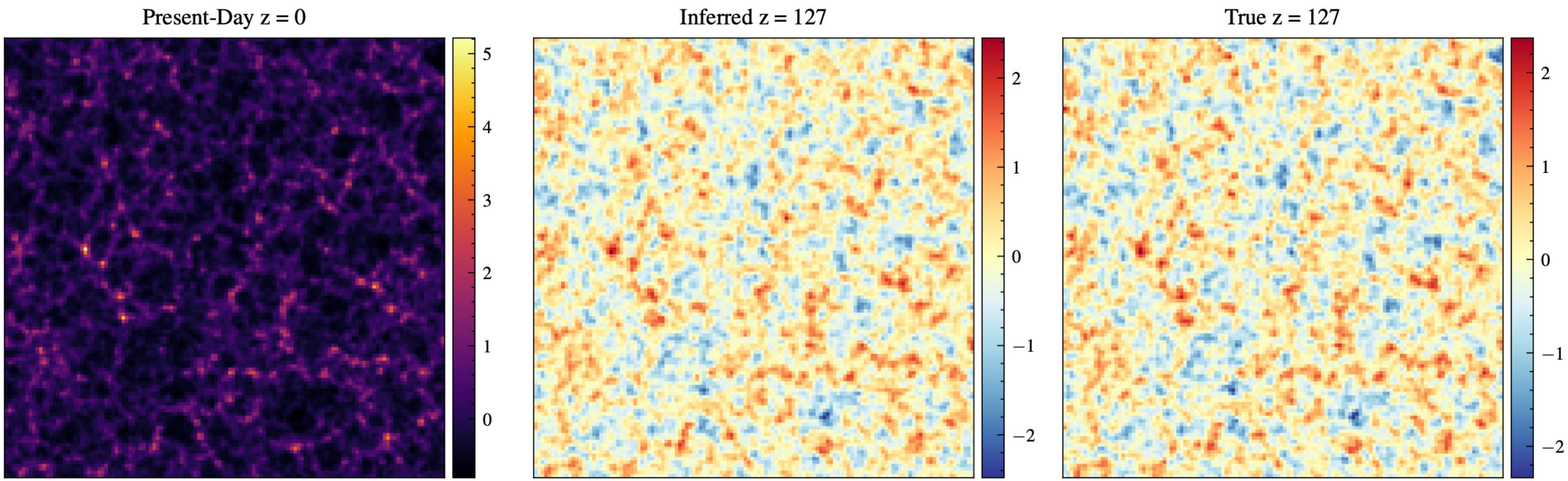
First full-field inference of initial conditions from fully non-linear density field



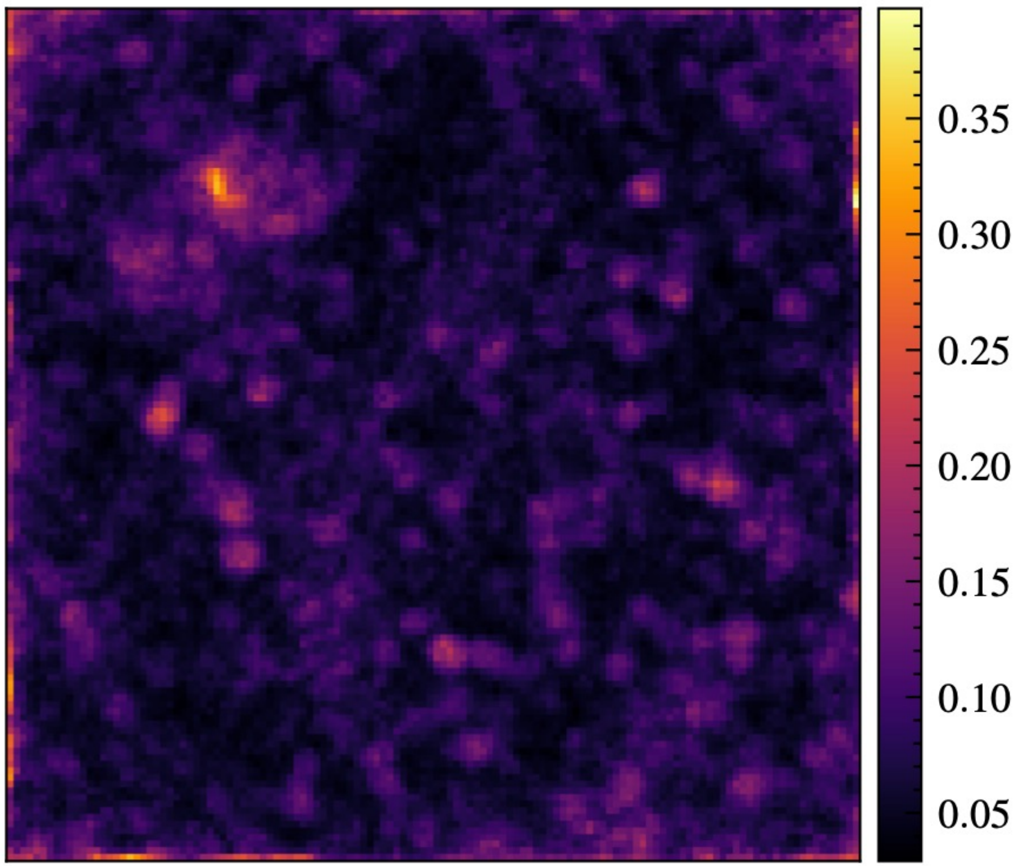


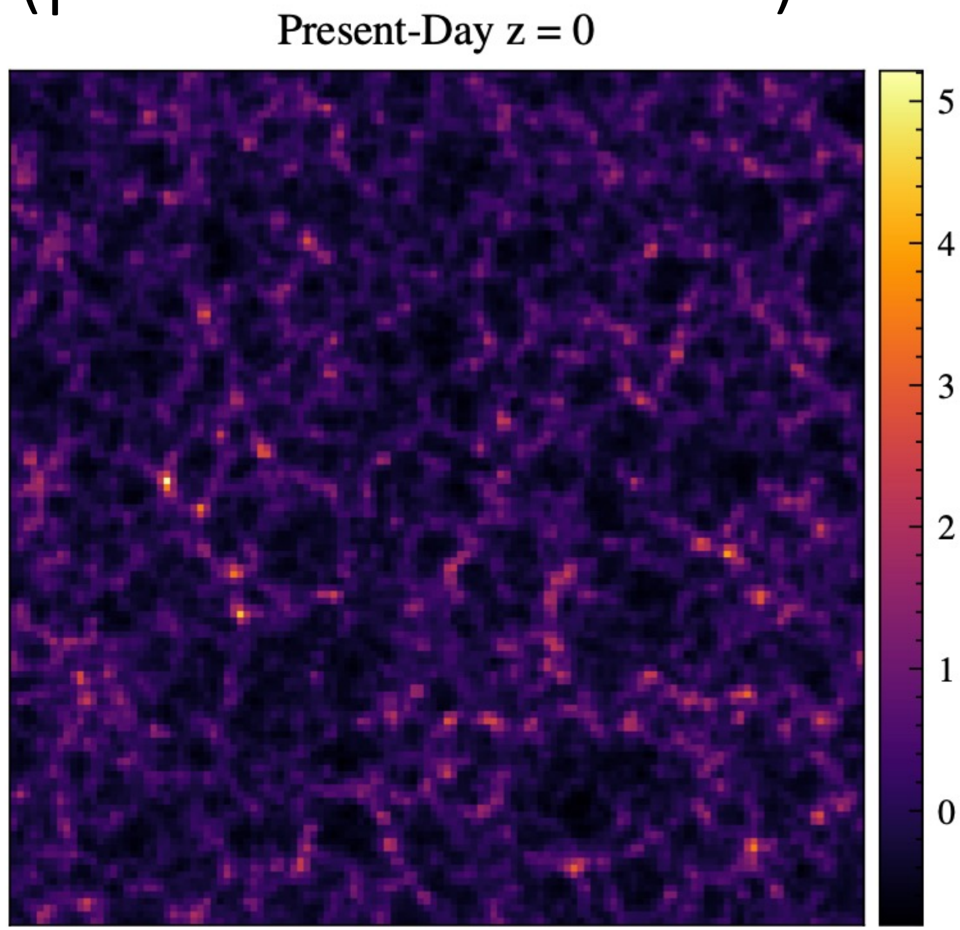
R. Legin *et al.*, arxiv:2304.03788

Faithful reconstruction…



… including uncertainties (posterior variance)

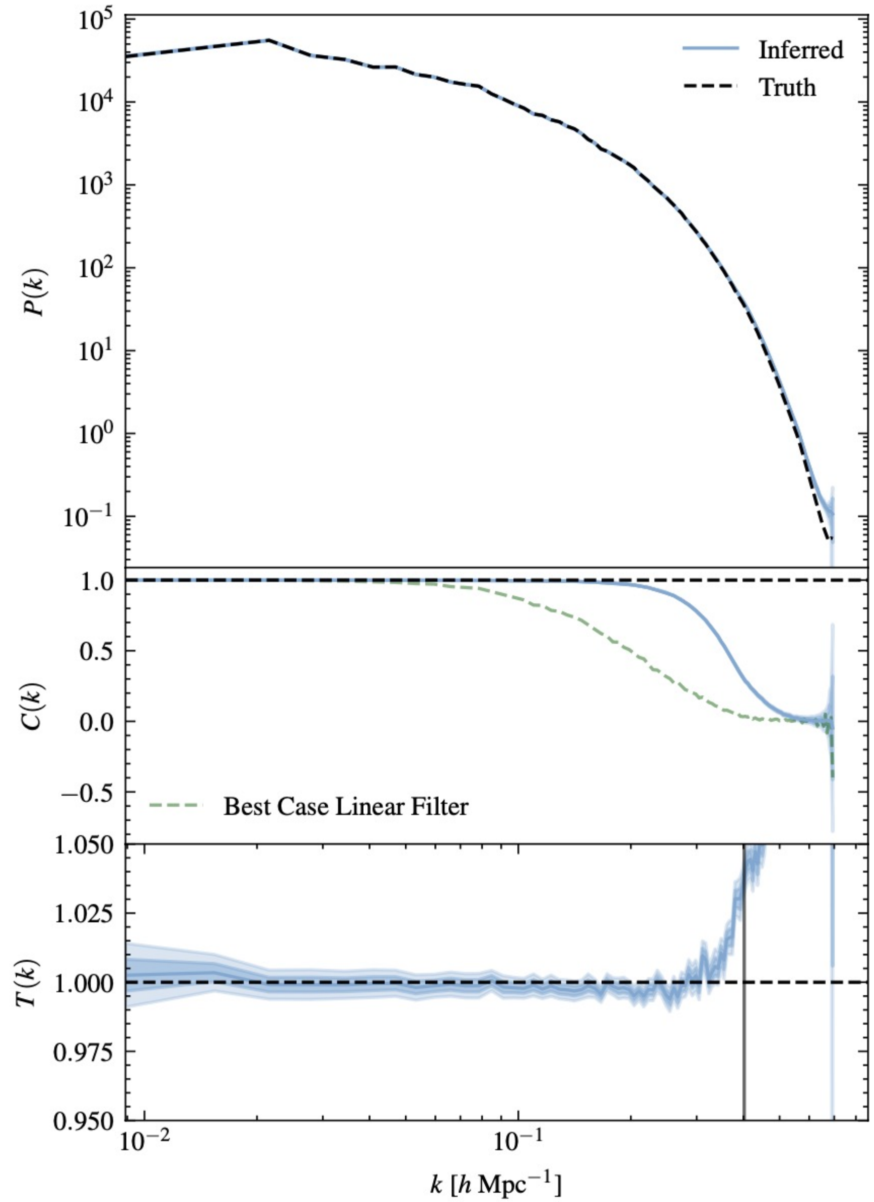




Accurate reconstructions

Points to note:

- full non-linear gravity
- No need for differentiability



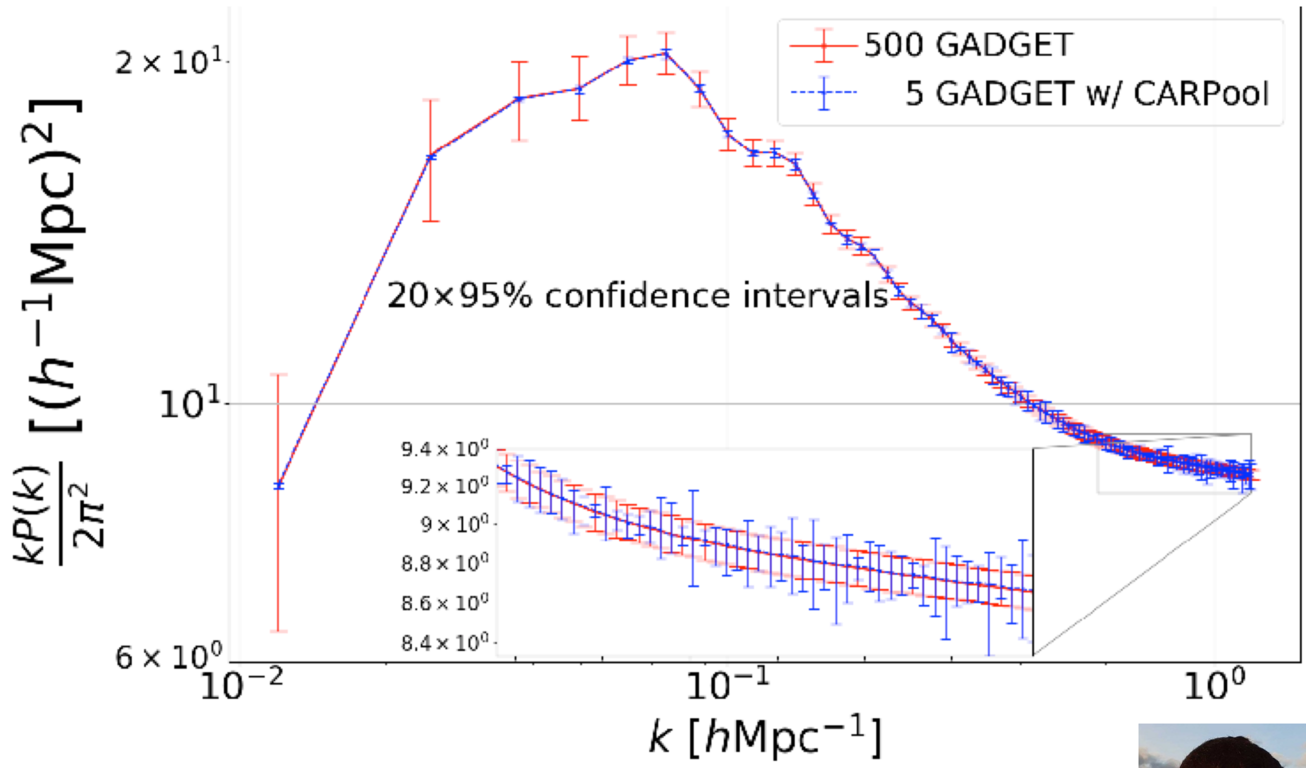
Bonus material

- How do we get all these simulations?

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.

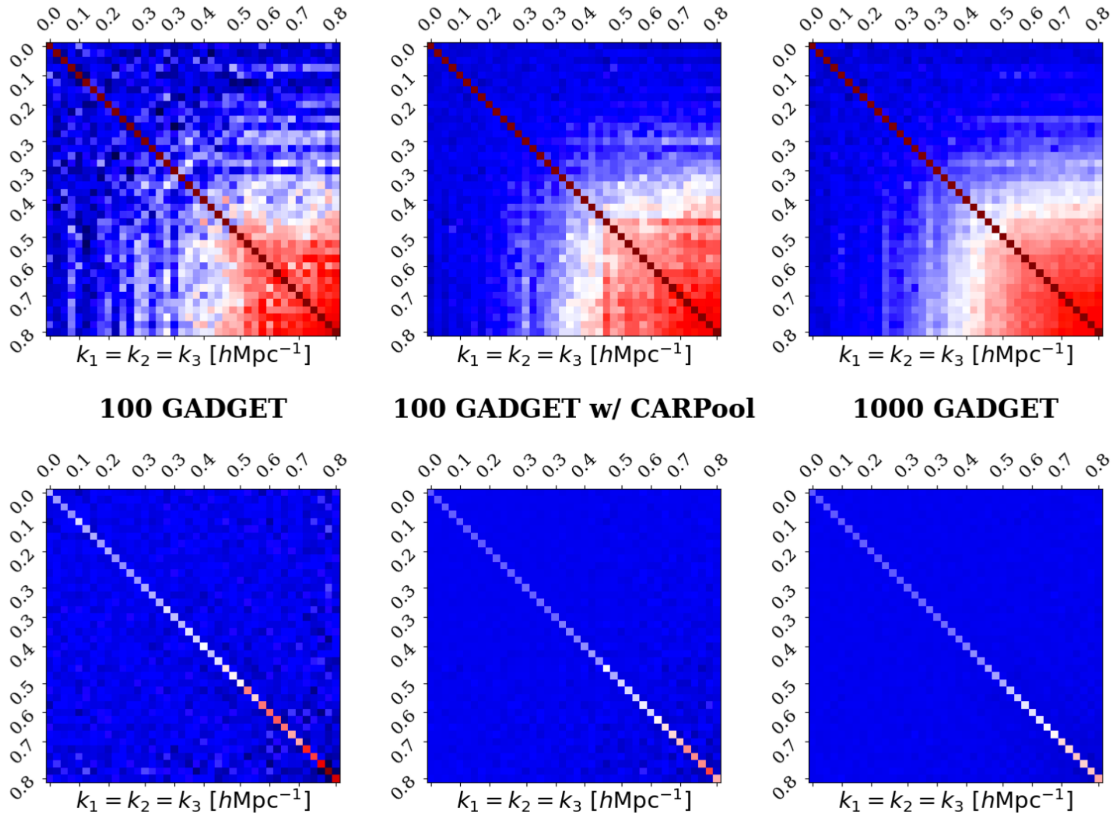


N. Chartier et al, arXiv:2009.08970

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

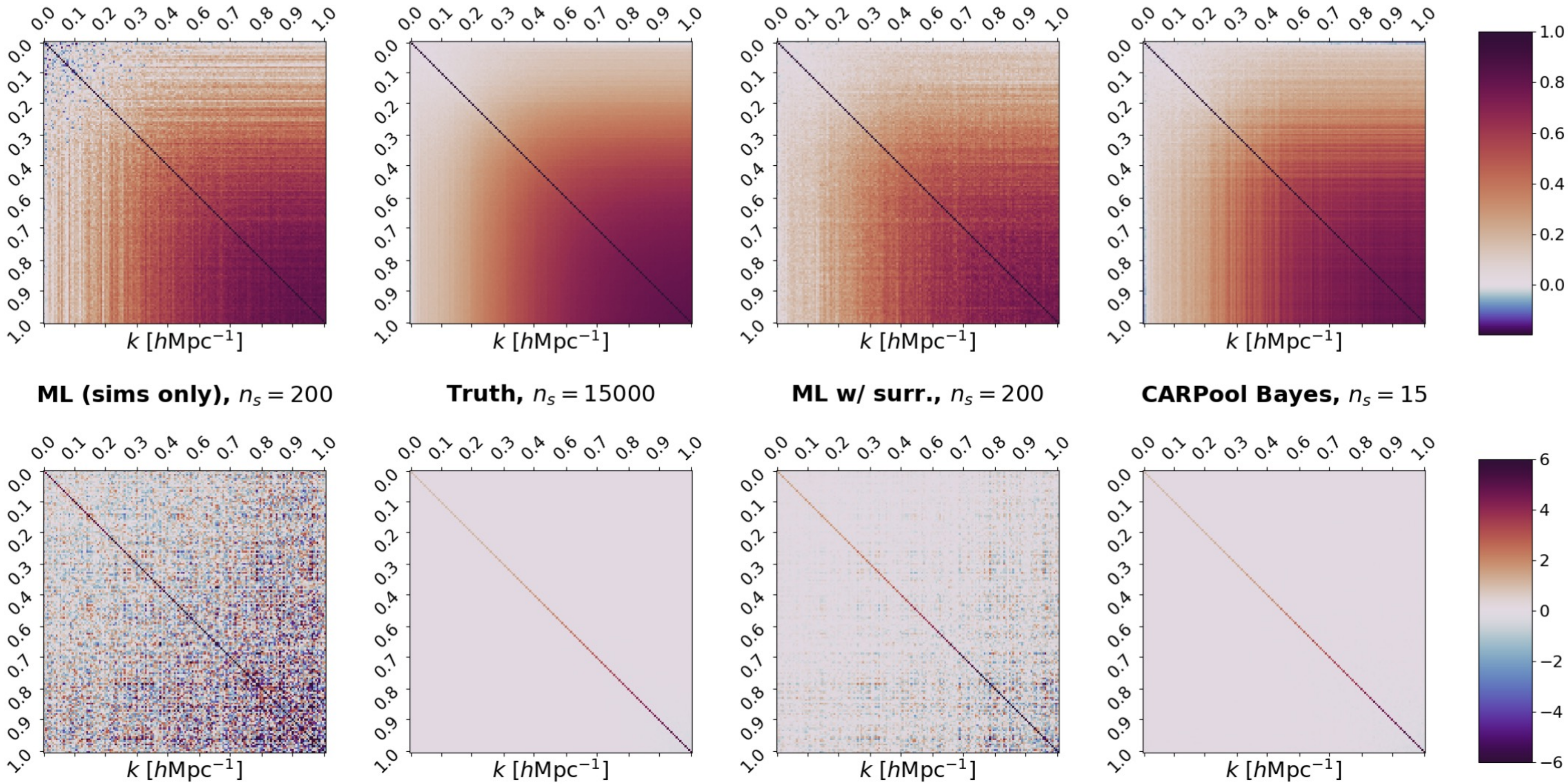
Covariance matrices and inverses

10 fold reduction in number of simulations for comparable accuracy



N. Chartier et al, arXiv:2106.11718

New: Bayesian CARPool estimators for the covariance



Chartier & Wandelt, arXiv:2204.03070

The CARPool Bootstrap: a Non-perturbative, Statistical Approach to "Perturbation Theory"

- Take existing set of numerical simulations for Model A
- For Model B, change parameters, include new effects
- Use 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

Chartier & Wandelt, arXiv:2204.03070

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

Benjamin Wandelt

Easy evidence calculation

- If you have a nested model, and if the prior for the additional parameter is independent of the other parameters, evidence calculation becomes simple
  - the **Savage-Dickey** ratio

$$\frac{p(d|M_{\theta,\zeta})}{p(d|M_{\theta,\zeta=0})} = \frac{p(\zeta=0|M_{\theta,\zeta})}{p(\zeta=0|d,M_{\theta,\zeta})}$$