Machine Learning the Universe

Bhuvnesh Jain
Center for Particle Cosmology
University of Pennsylvania

Collaborators: Minsu Park, Kunhao Zhong, Marco Gatti, Supranta Bouruah;
& Dark Energy Survey
Papers: arXiv:2310.17557, 2405.10881, 2403.01368, 2411.04759,
arXiv:2502.04158, 2502.06687

See also: Subaru-HSC lensing survey!

Outline

- How do we map dark matter in the universe?
- How do we compare observed maps to theory?
- How can we do cosmology with deep learning?

The expanding universe

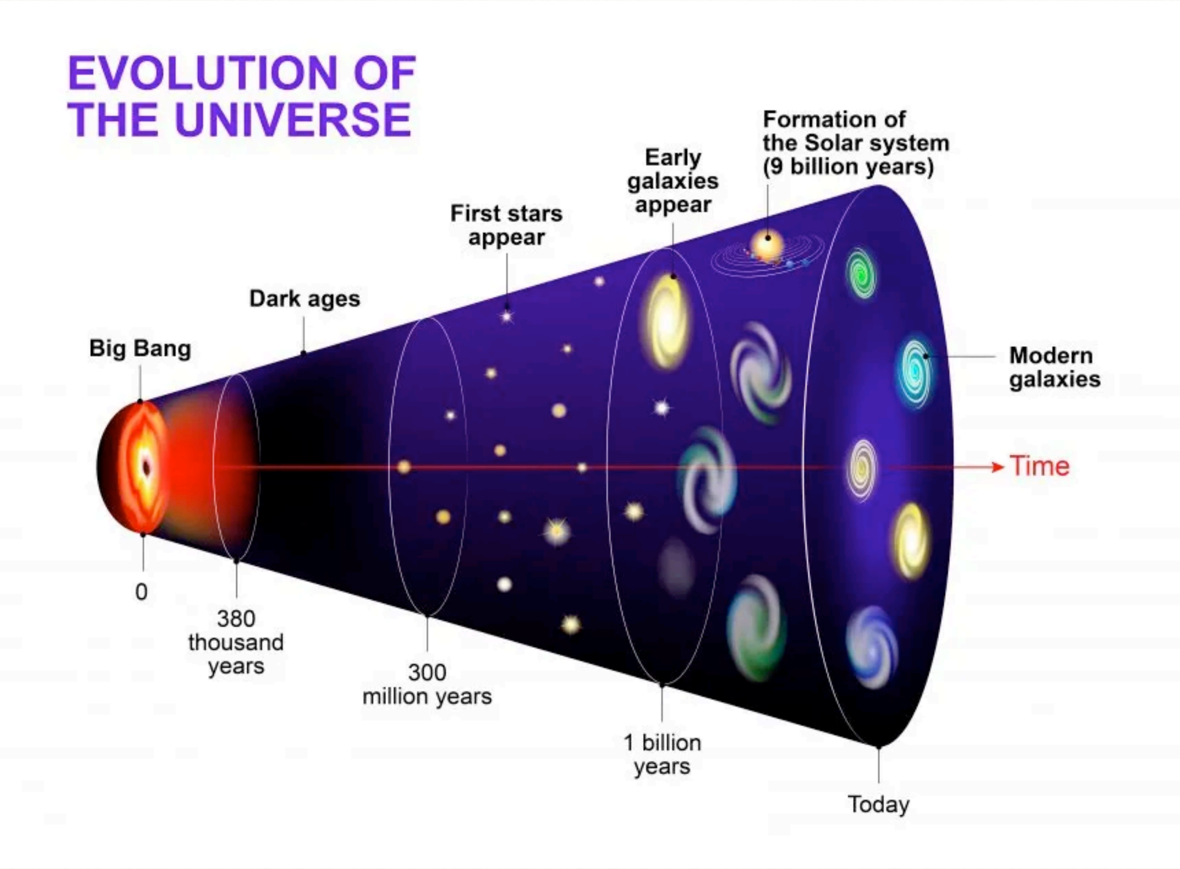


Figure from [Cherifa Bochra Soltani]

Gravitational lensing

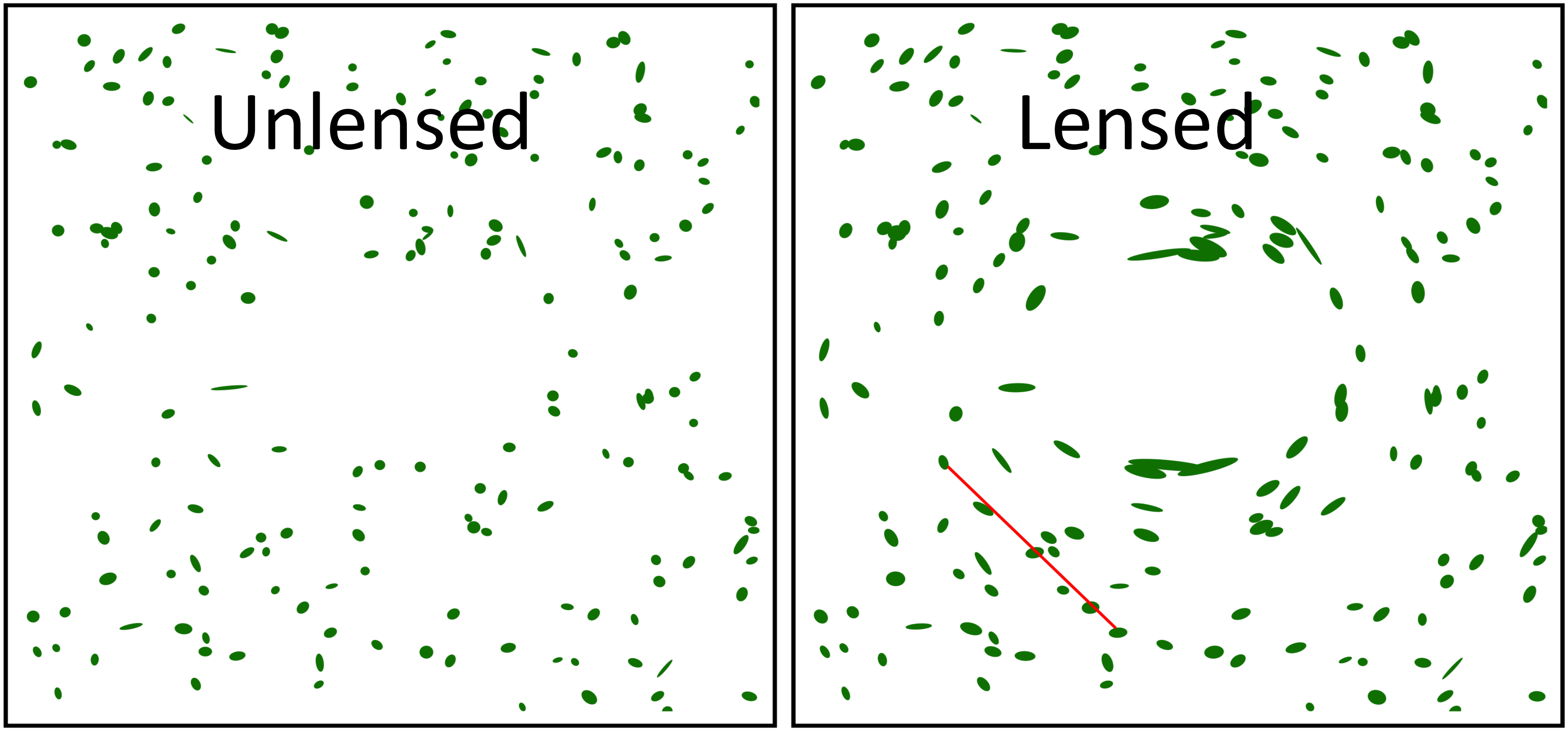
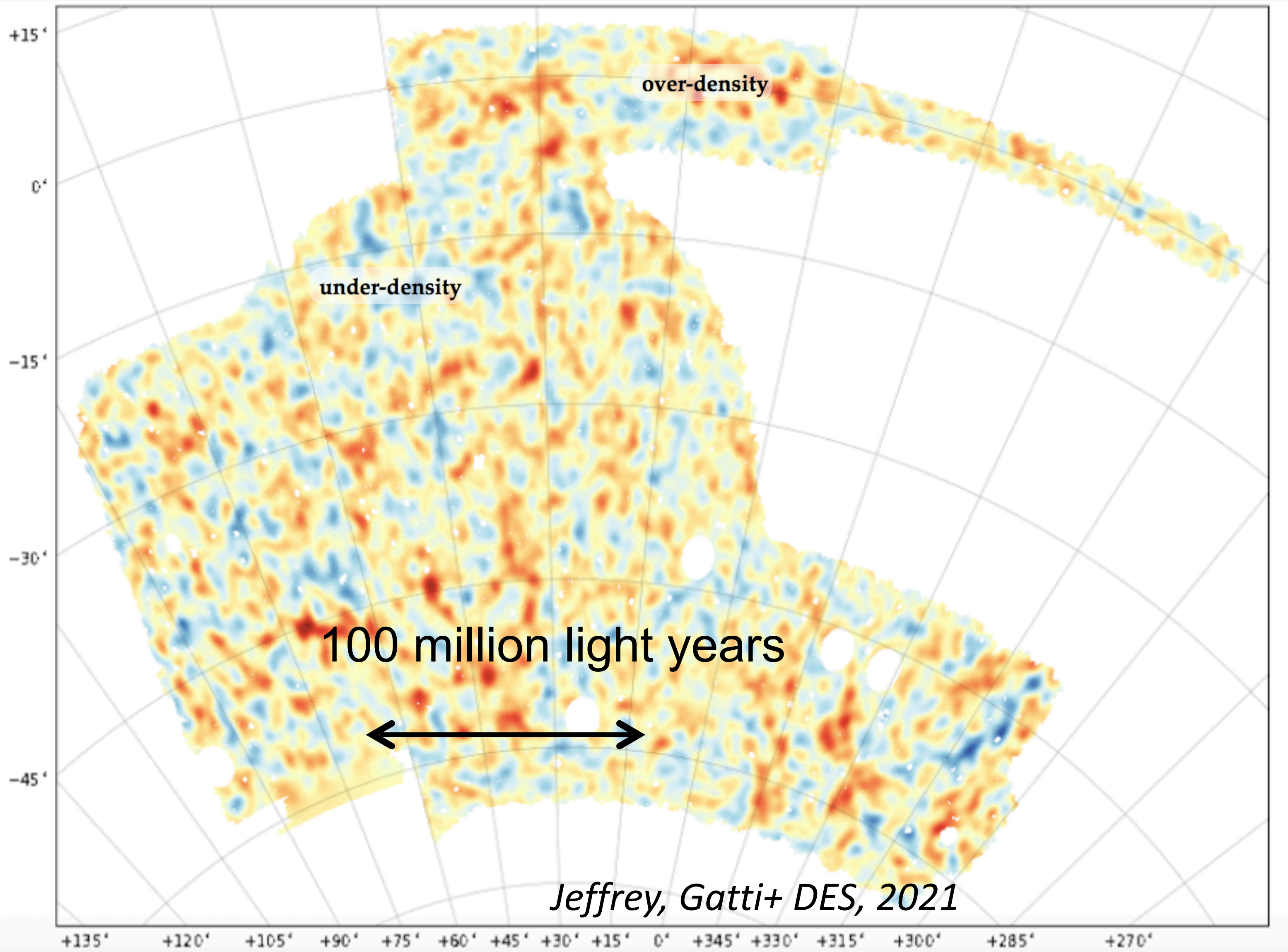


Figure from Jim Bosch

Measured shapes of galaxies help us map the full mass distribution.
In reality the signal is a 0.1 percent coherent pattern!

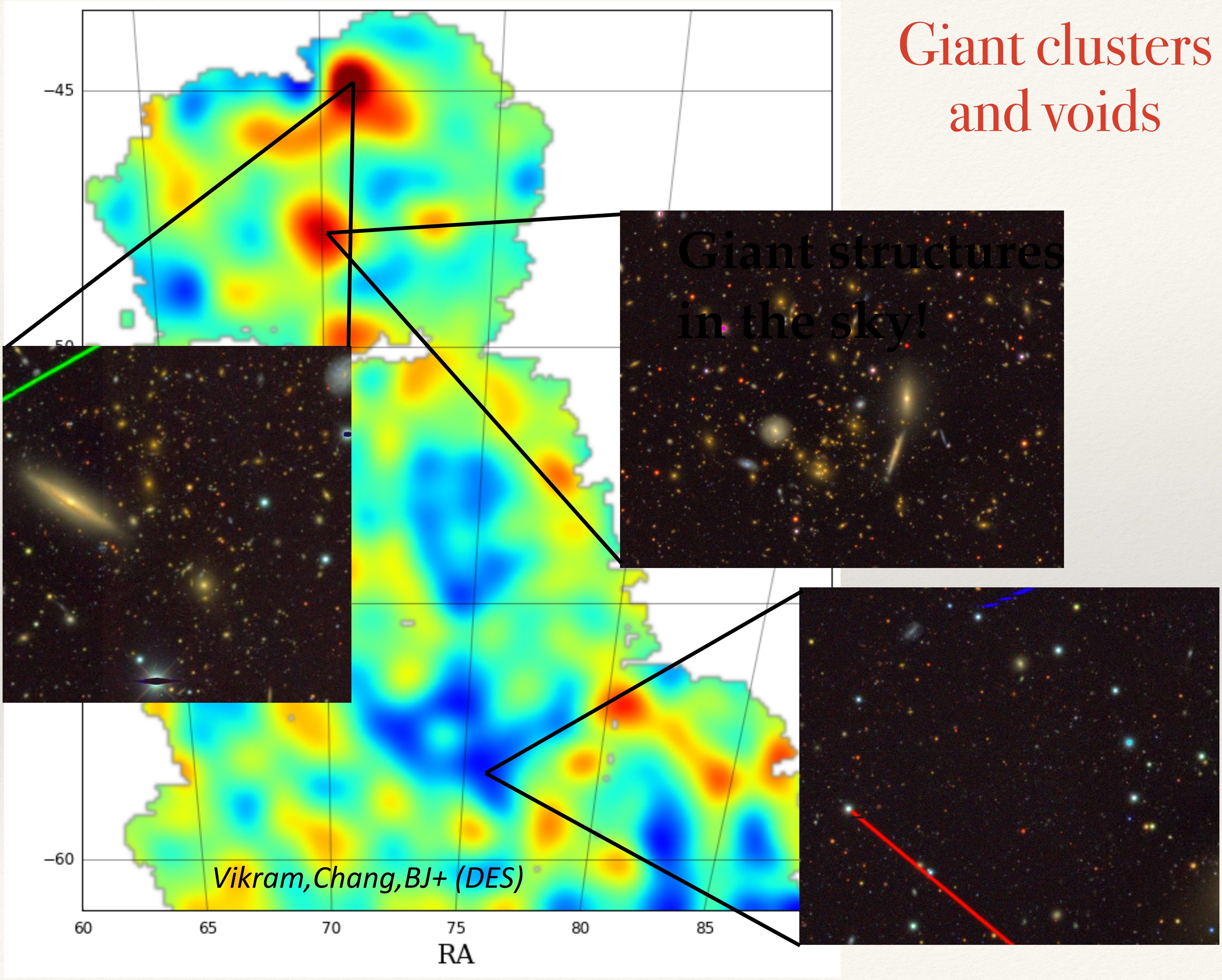
A large scale map of dark matter



Jeffrey, Gatti+ DES, 2021

Lensing mass map: using the lensing distortions of ~100 million galaxies

Giant clusters and voids



Giant structures in the sky!

Vikram,Chang,BJ+ (DES)

But first, what do we do with all this data?

Two cosmic puzzles: the H_0 and S_8 tensions

- **The H_0 tension: The universe is expanding faster than ‘predicted’, i.e. using the baseline model of cosmology, ΛCDM, to extrapolate early time measurements to the present.**

- **The S_8 tension: Growth of large-scale structure in the universe is slower than predicted.**
    - The S_8 tension shows up in multiple measurements of the late time universe but only at ~2-sigma.

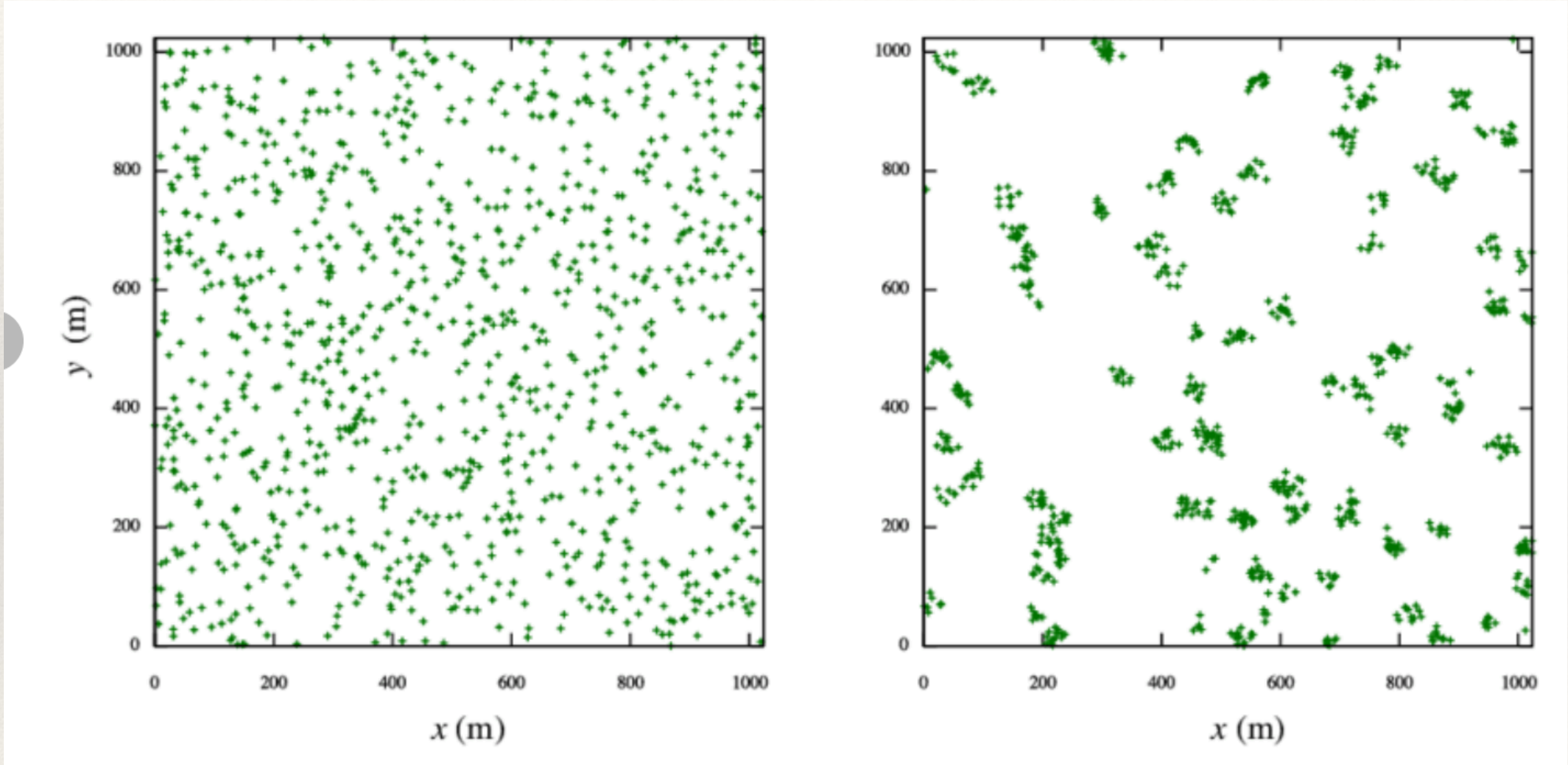
- The nature of dark matter and dark energy + the two cosmic puzzles + the origin of the Big Bang universe drive much of the research in cosmology

How does data confront theory?

- The standard model of cosmology has 6+ parameters — numbers that define
  - The energy budget: densities of dark matter, baryons, dark energy, neutrinos
  - Initial fluctuations: amplitude and slope of the mass power spectrum
  - H, expansion rate / Age of the universe
- Measurements are compared to predictions by varying these parameters using Bayesian statistics + MCMC sampling of parameter space
- Key questions answered by this exercise:
  - Is the model a good fit?
  - What are the values of its parameters? Do they agree with other surveys?
  - Is an alternative model a better fit?

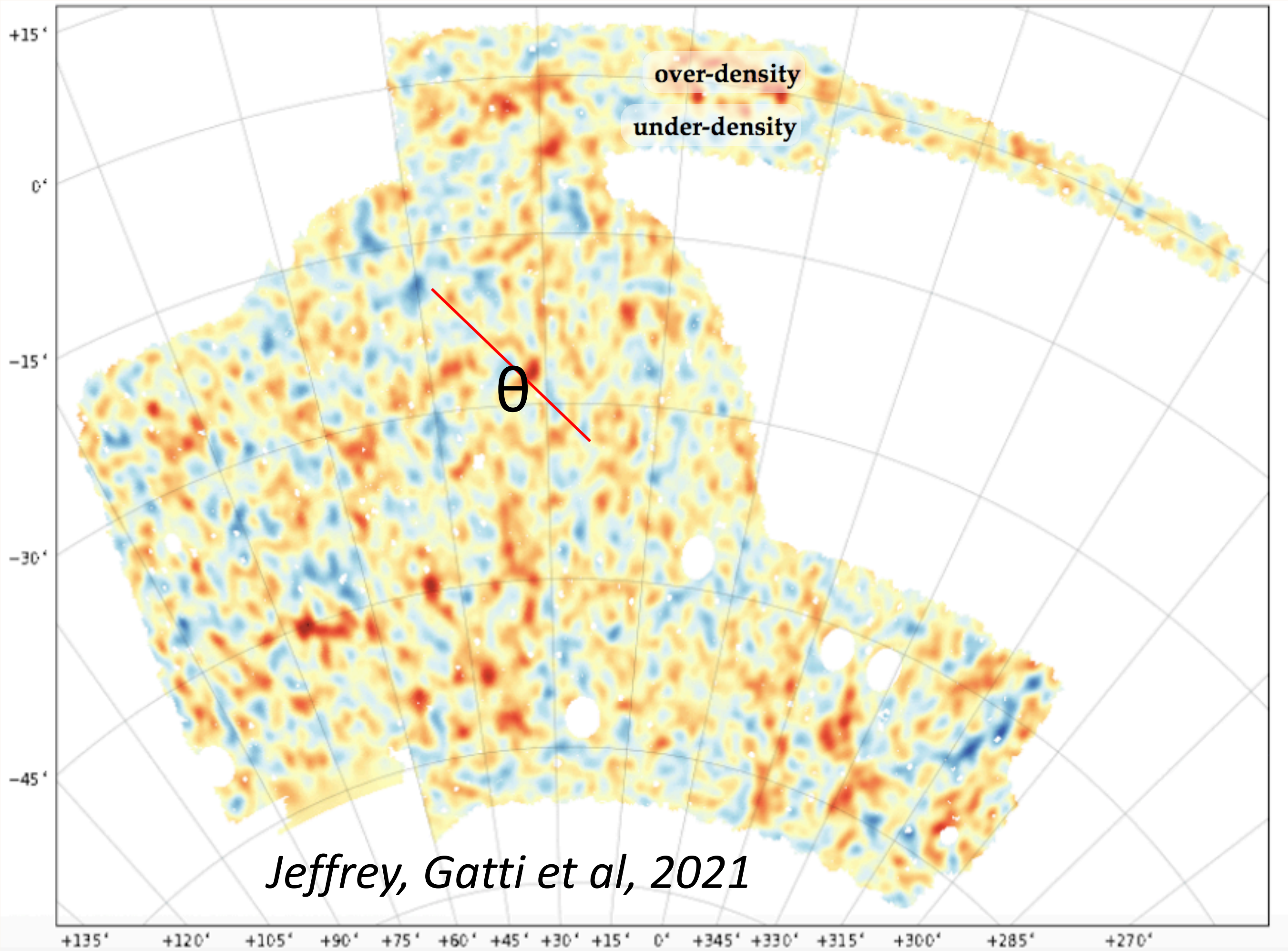
  - We will start with a conventional approach, then dive into Deep Learning

Random vs clustered points



How can we measure clustering?

Dark Energy Survey mass map



Jeffrey, Gatti et al, 2021

Lensing mass map: using the lensing distortion for ~100 million galaxies

Clustering statistics

We measure the 2-point correlation function between two variables as a function of angular separation, θ:

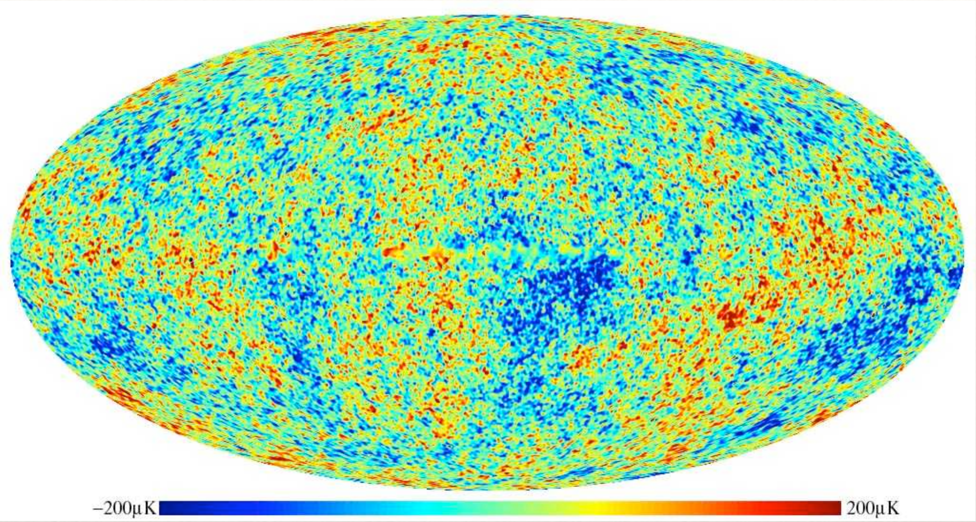
$$\left\langle Density(\theta_1)Density(\theta_2)\right\rangle \text{where}\theta = |\theta_1 - \theta_2|$$

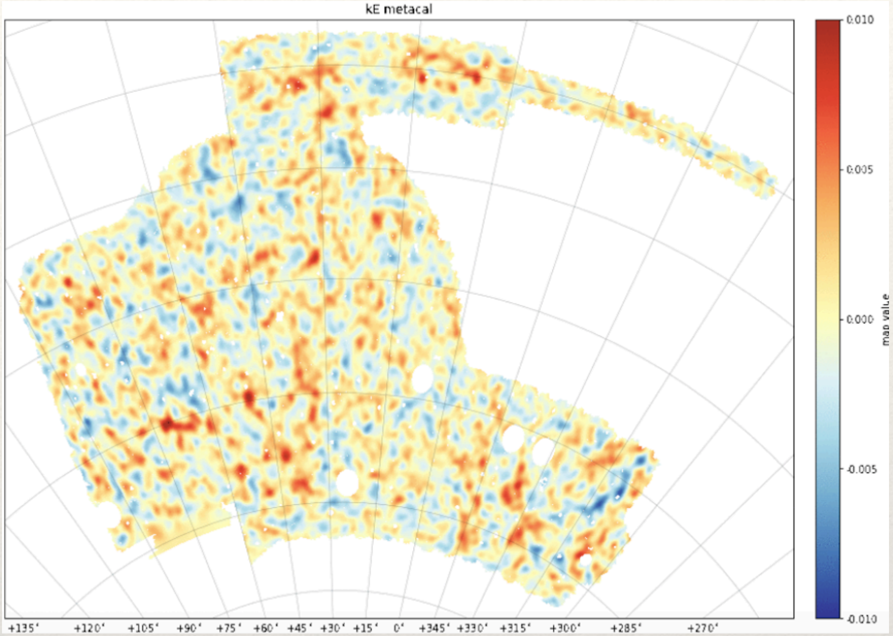
The idea is to measure the strength of fluctuations on different length scales.

The 2-point function compresses the data very effectively.

Its Fourier transform is the power spectrum.

Spatial clustering in the early vs. late universe





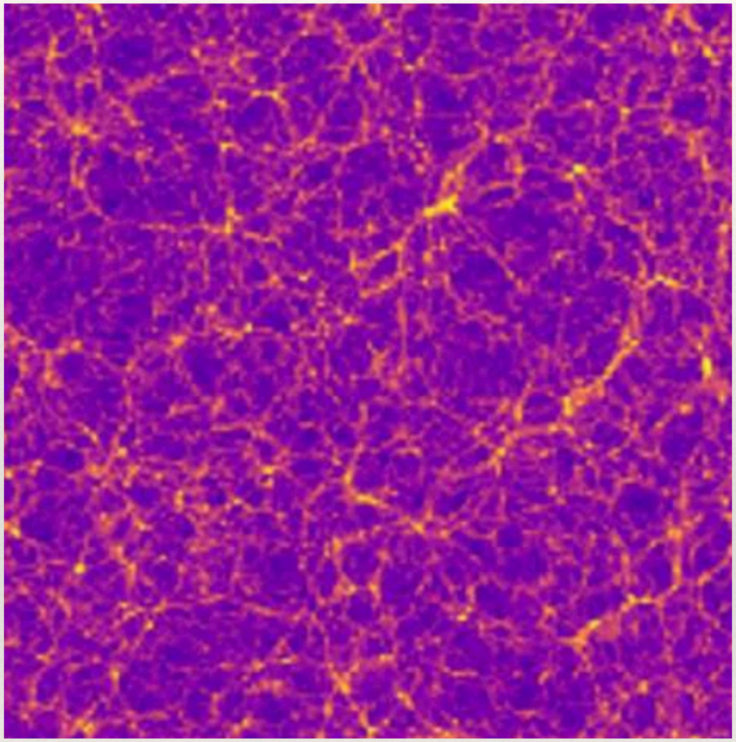
Age = 400,000 years

Age = 10 billion years

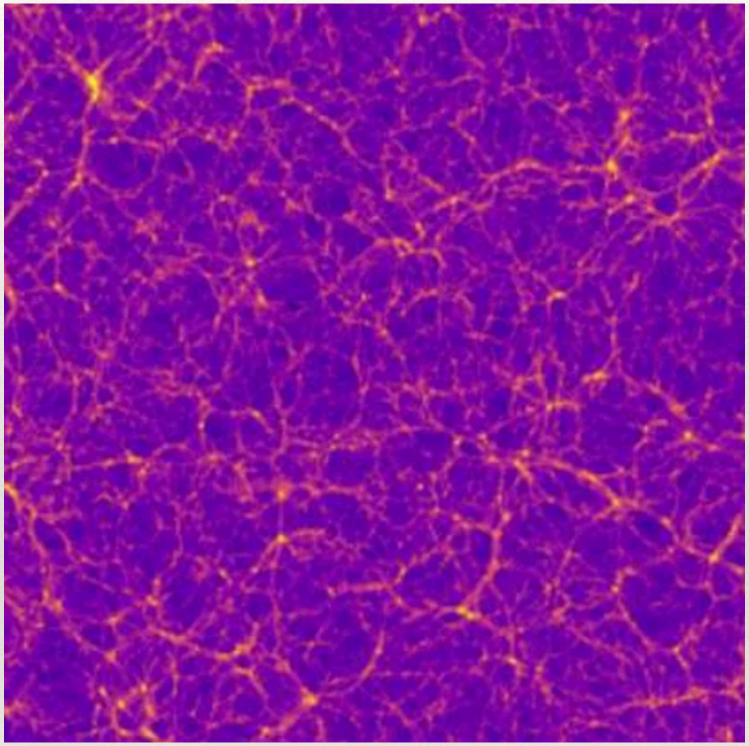
The universe ~today is less clustered than ‘predicted’ by the theory.

This is the ‘second cosmic puzzle’, called the S8 tension.

Structure mismatch – the S_8 tension



Predicted by theory

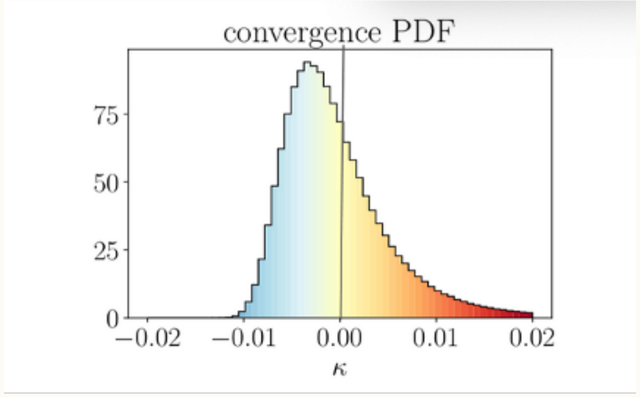


Inferred from lensing data

How can we compare these maps, using all the information they contain?

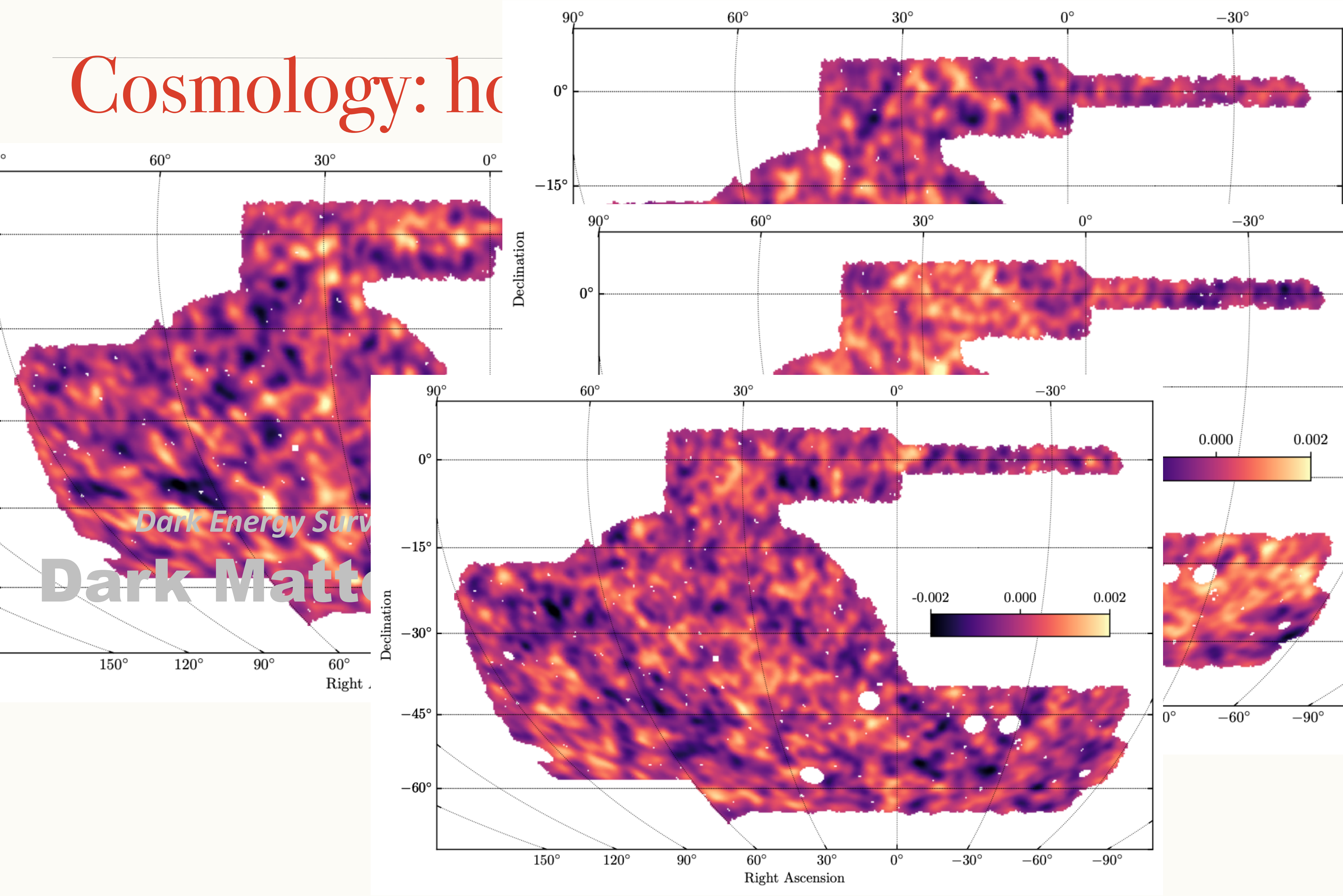
Can we extract more information?

- 2-point correlations capture all the information in a Gaussian field but not from the observed maps of the evolved universe.
- Mining 'non-Gaussian' regime can ~double the information content. How?
  - N-point correlations
  - Clusters and voids
  - Topological statistics
  - Field level inference with **deep learning** or BHM



Leaping into the unknown with deep learning

Cosmology: ho



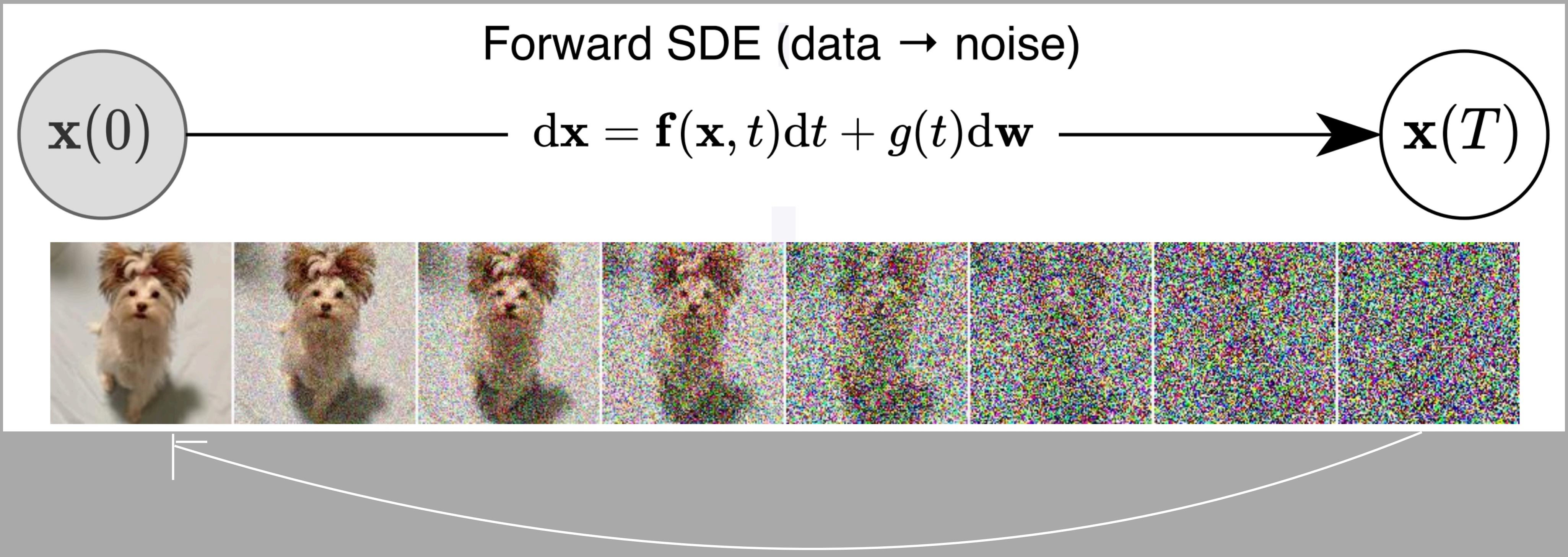
Where astronomy meets AI

Deep Learning for Gravitational Lensing

- Weak lensing is well suited for Deep Learning: the uncertain physics (baryonic feedback and Intrinsic Alignments) is well bounded*
- We can generate simulated maps to train a deep learning model and then apply it to extract information from survey maps
- First, let's build a model to **generate** large-scale structure.

Diffusion — the main idea

Progressively add noise to the image until we just have white noise
(We completely control this step via a Stochastic Differential Equation)



Forward SDE (data → noise)

$\mathbf{x}(0)$

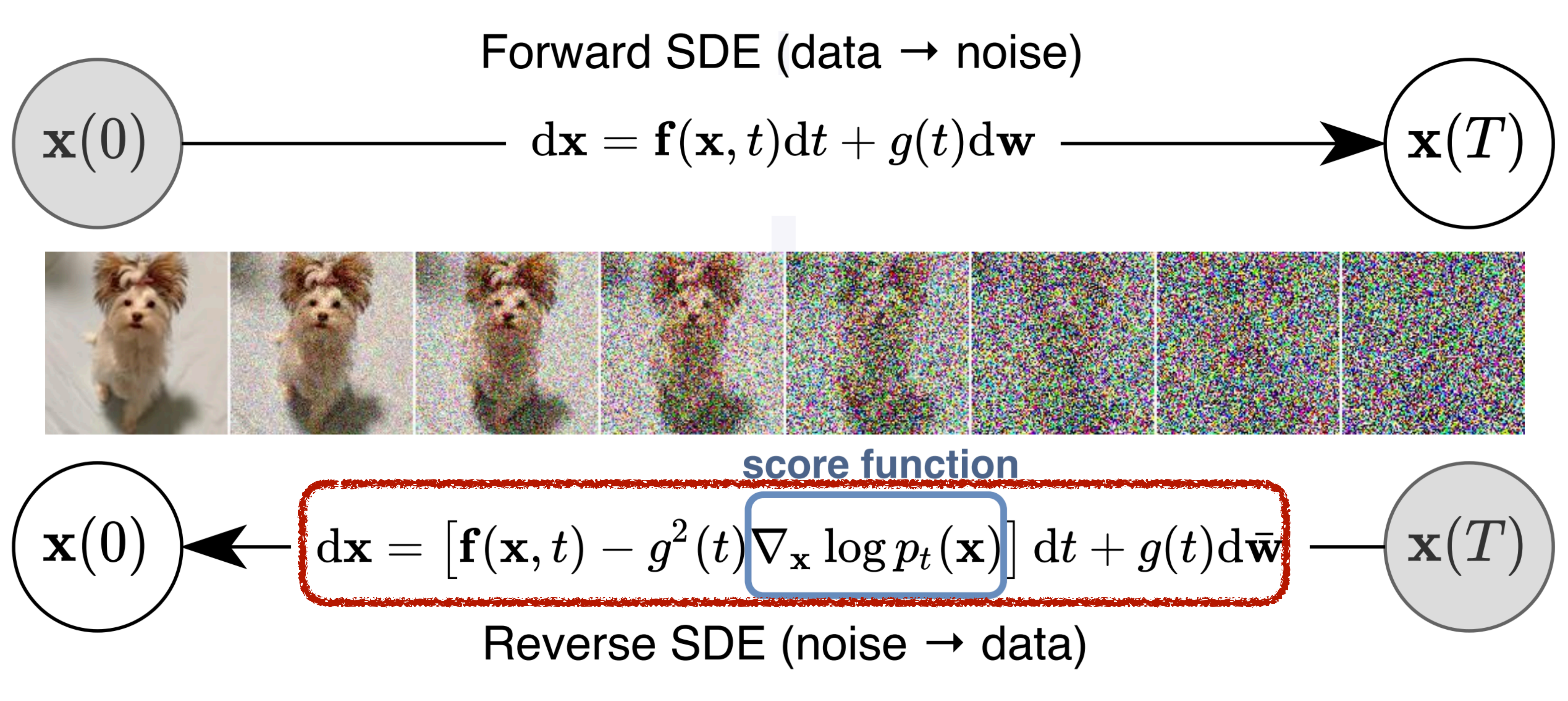
$$\mathrm{d}\mathbf{x} = \mathbf{f}(\mathbf{x}, t)\mathrm{d}t + g(t)\mathrm{d}\mathbf{w}$$

$\mathbf{x}(T)$

Slide from Supranta Bouruah

Why diffusion is useful for generative modeling?

Important result: Any SDE of this form can be reversed!



Forward SDE (data → noise)

$$\mathbf{x}(0) \longrightarrow \mathrm{d}\mathbf{x} = \mathbf{f}(\mathbf{x},t)\mathrm{d}t + g(t)\mathrm{d}\mathbf{w} \longrightarrow \mathbf{x}(T)$$

$$\mathbf{x}(0) \longleftarrow \boxed{\mathrm{d}\mathbf{x} = \left[\mathbf{f}(\mathbf{x},t) - g^2(t)\overset{\text{score function}}{\boxed{\nabla_{\mathbf{x}}\log p_t(\mathbf{x})}}\right]\mathrm{d}t + g(t)\mathrm{d}\bar{\mathbf{w}}} \longleftarrow \mathbf{x}(T)$$

Reverse SDE (noise → data)

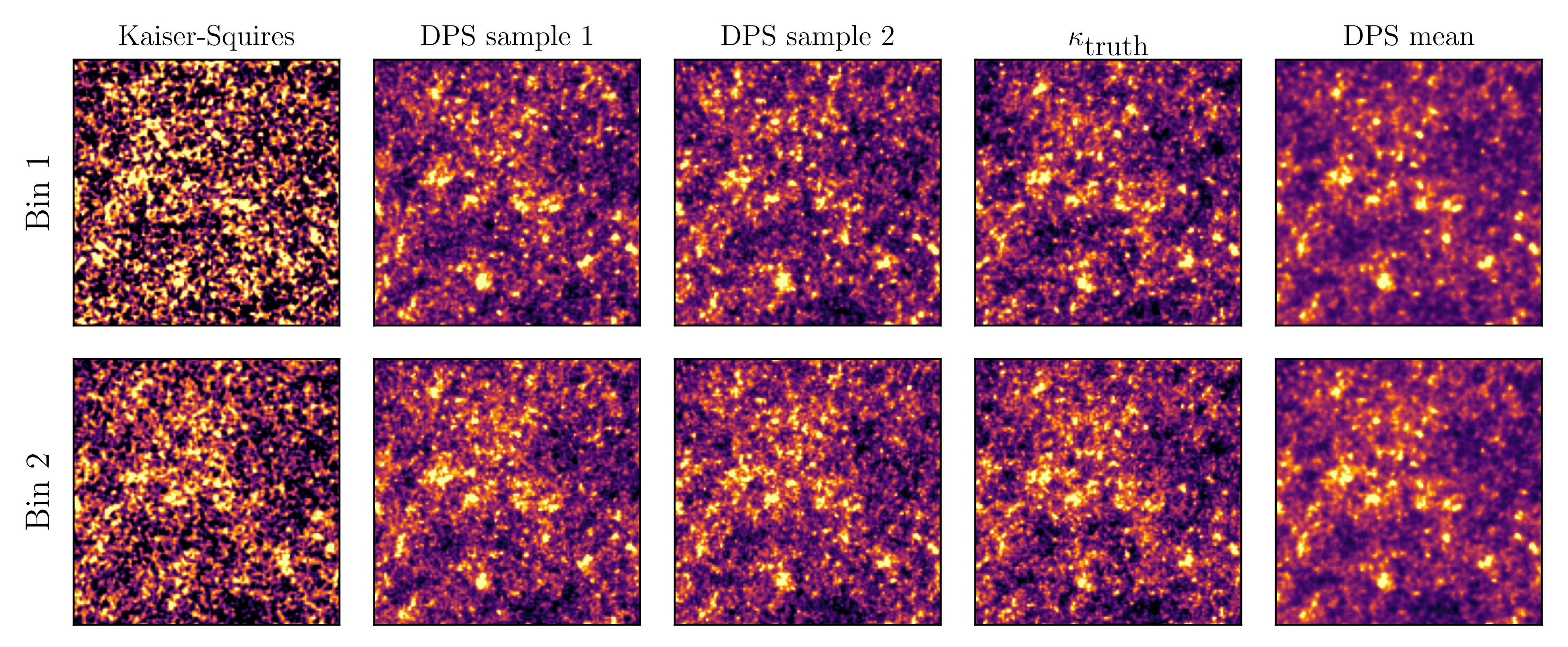
Reverse stochastic differential equation requires the gradient of the log probability, a.k.a ***score***

Train neural networks to predict the score, at different noise level

Once trained, neural networks can be used to transport latent space noise to samples from the data distribution

Slide from Supranta Bouruah

Diffusion models: noisy data ➡ underlying field



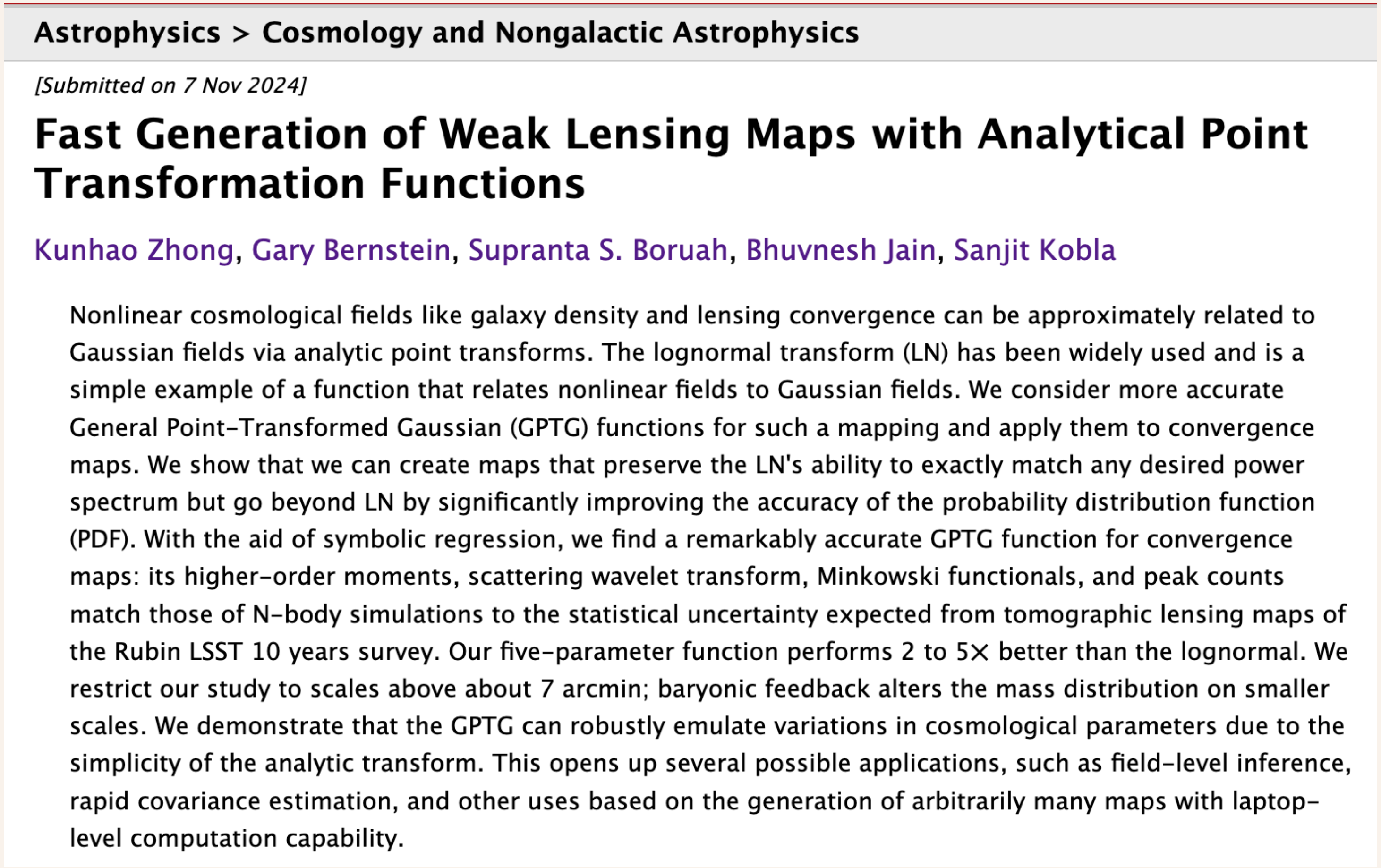
- Use generative AI to learn the ‘prior’ distribution of mass maps
- Reconstruct mass map from noisy survey data with no additional training!
- High resolution maps with **uncertainty** estimates — fast **and** robust

Bouruah, Jacob, BJ 2025, arXiv:2502.04158

And now for an ‘anti-ML’ result

- We found an analytical method that also generates simulation quality maps!
- It is accurate at lower resolution — good for real surveys, and still highly nonlinear.

Analytical Lensing Maps



Astrophysics > Cosmology and Nongalactic Astrophysics

[Submitted on 7 Nov 2024]

Fast Generation of Weak Lensing Maps with Analytical Point Transformation Functions

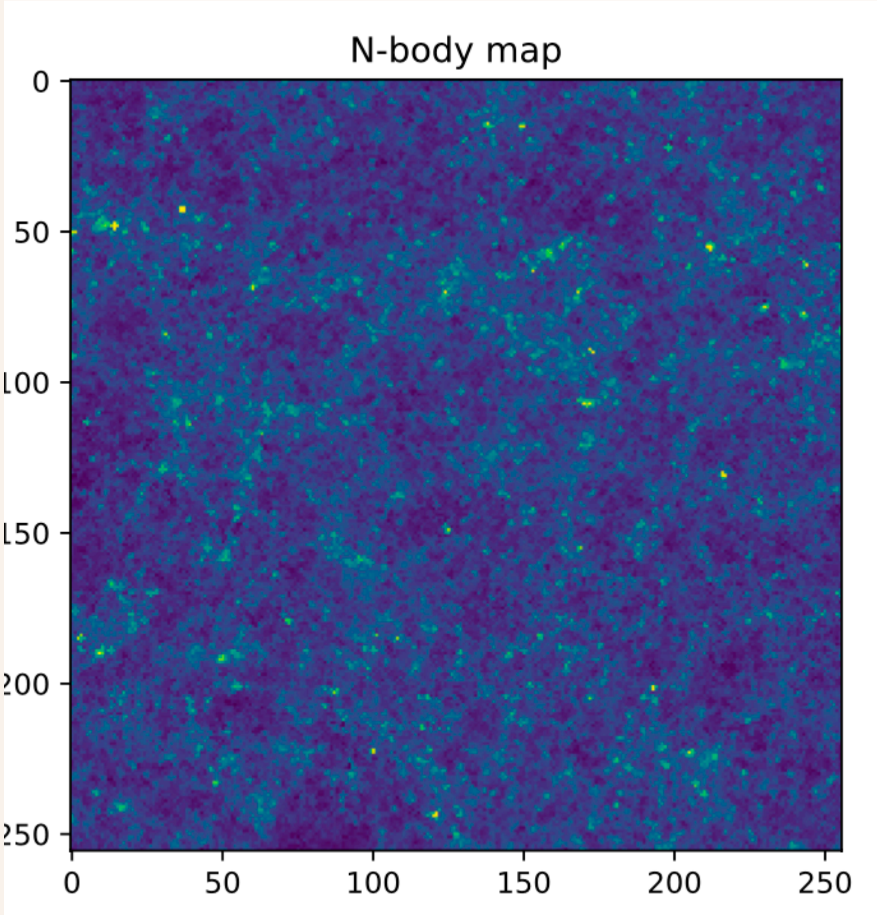
Kunhao Zhong, Gary Bernstein, Supranta S. Boruah, Bhuvnesh Jain, Sanjit Kobla

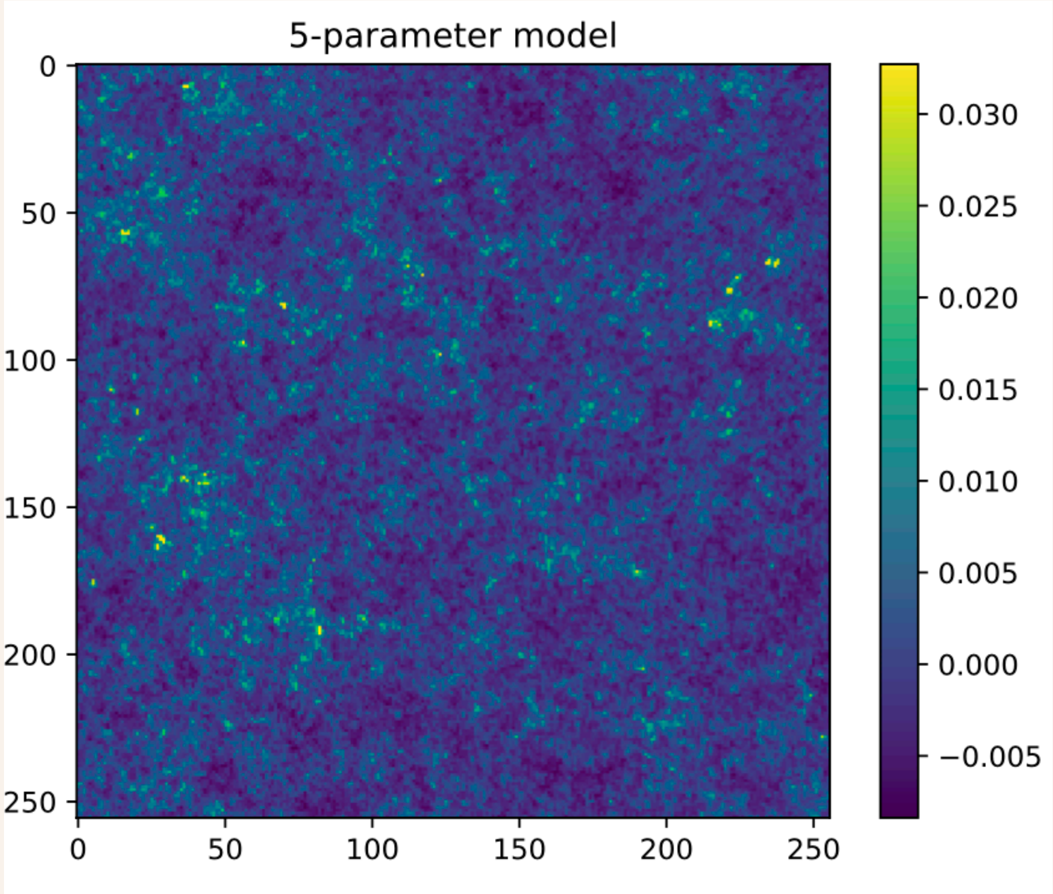
Nonlinear cosmological fields like galaxy density and lensing convergence can be approximately related to Gaussian fields via analytic point transforms. The lognormal transform (LN) has been widely used and is a simple example of a function that relates nonlinear fields to Gaussian fields. We consider more accurate General Point-Transformed Gaussian (GPTG) functions for such a mapping and apply them to convergence maps. We show that we can create maps that preserve the LN's ability to exactly match any desired power spectrum but go beyond LN by significantly improving the accuracy of the probability distribution function (PDF). With the aid of symbolic regression, we find a remarkably accurate GPTG function for convergence maps: its higher-order moments, scattering wavelet transform, Minkowski functionals, and peak counts match those of N-body simulations to the statistical uncertainty expected from tomographic lensing maps of the Rubin LSST 10 years survey. Our five-parameter function performs 2 to 5× better than the lognormal. We restrict our study to scales above about 7 arcmin; baryonic feedback alters the mass distribution on smaller scales. We demonstrate that the GPTG can robustly emulate variations in cosmological parameters due to the simplicity of the analytic transform. This opens up several possible applications, such as field-level inference, rapid covariance estimation, and other uses based on the generation of arbitrarily many maps with laptop-level computation capability.

Three steps:

1. Gaussianize nonlinear field using its pdf.
2. Match power spectrum exactly.
3. Generate new Gaussian field and invert to get new maps.

Simulated vs. Analytical mass maps





We can pick up some subtle differences by eyes, but all statistical tests pass at the accuracy of real surveys. We can generate millions of ‘analytic’ kappa maps on a laptop! Current simulations take days to run on a supercomputer.

Another analytical method applied to a variety of physical phenomena

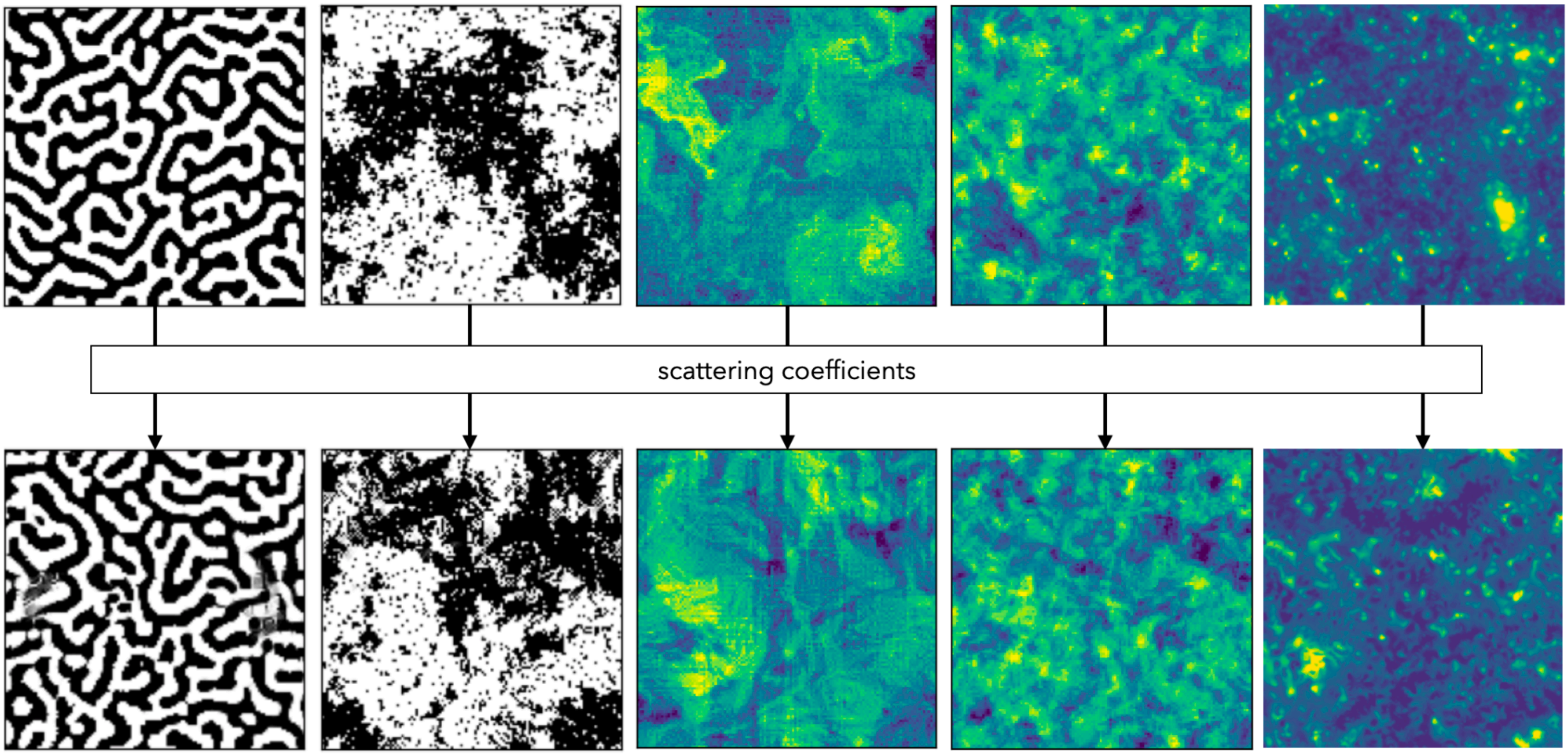
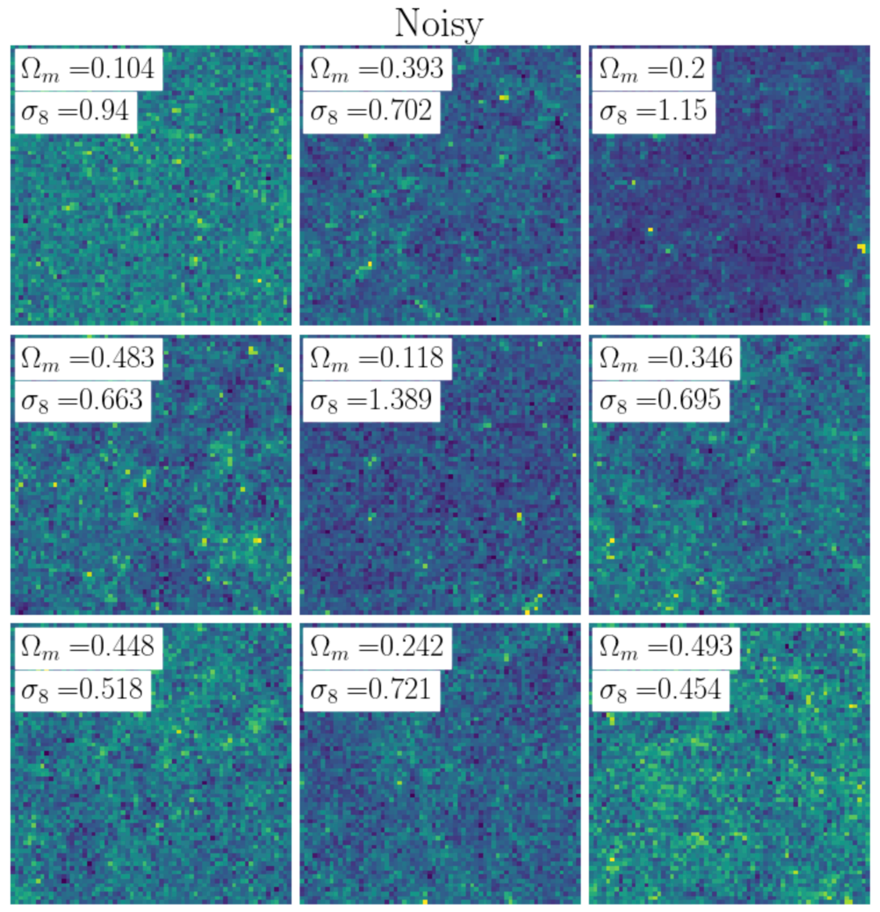


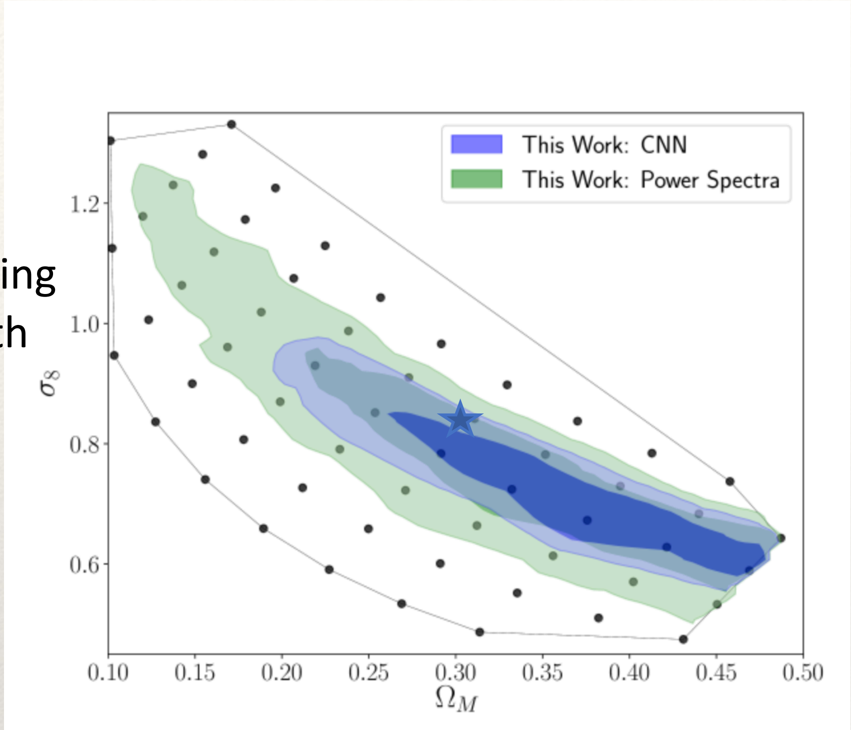
Figure 2. Texture synthesis using the scattering transform for a variety of physical fields: Turing pattern, Ising model, ocean turbulence, solar surface, cosmic matter density. The upper panels show input 2-D fields from simulations or observations. The lower panels show randomly generated fields with scattering coefficients matching their upper counterparts.

Wavelet like transform can reconstruct non-Gaussian stochastic fields. *Cheng & Menard 2021*

Back to Deep Learning: Application to Data



Clustering Strength



Density

Lensing map-> cosmology via a Convolutional Neural Net.

End-to-end analysis: factor of 2 gain in information! *Fluri+ 2019, 2022 and others*

Deep learning summary

- We have seen how to generate and reconstruct dark matter maps.
- Analytical methods keep showing up and helping or replacing ML!
- Deep learning framework for cosmology (including inference) is powerful.
- Validating and community acceptance are at early stages.

Open questions

- What does deep learning learn? Hard to tell, but we must try!
    - Fruitful area of collaboration with computer scientists, radiologists…

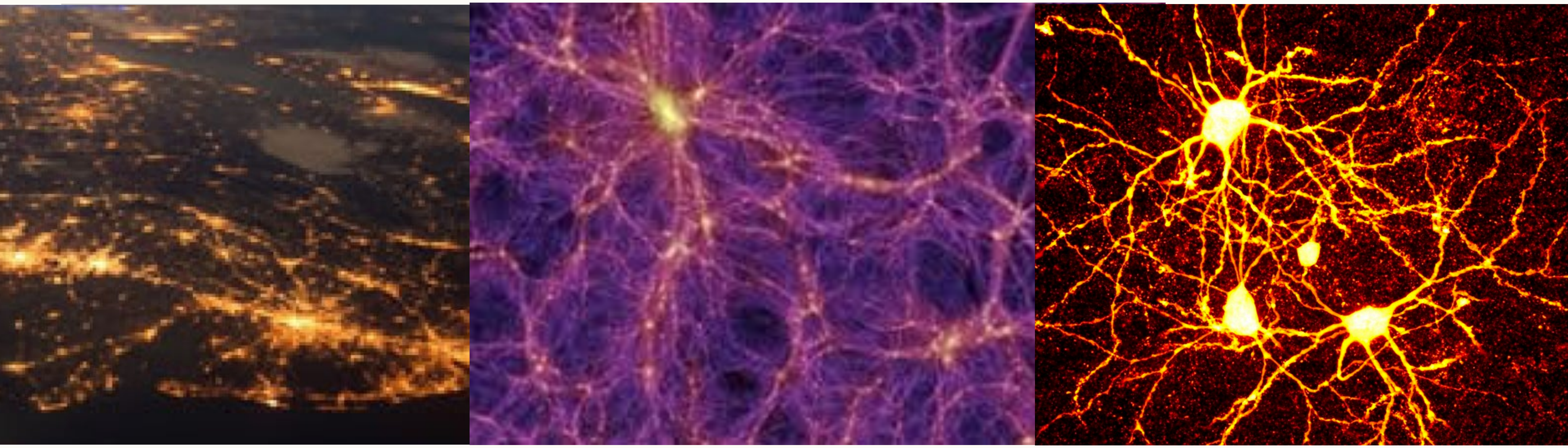
- How much information do these maps have?
    - Deep learning extracts all the information…under certain assumptions. How close are we to that regime in different problems in the natural sciences?

- What are the common features of natural phenomena?
    - Hierarchical structures, long range interactions, energy flow…??
      Can these shared features enable a multi-modal ‘foundation model’, a chatGPT for science?

How to approach ML in science: practical tips

- There are many tools of ML and many architectures for Deep Learning
    - Courses are good, but don't hesitate to jump in when an opportunity comes up (chatGPT and co are great tutors!).
- Learn about CNNs and Transformers — the oldest and youngest Deep Learning models.
    - CNNs for images & Transformers for language, time series data and more.
    - Transformers for science, an informal review: Tanoglidis, BJ, Qu. arXiv:2310.12069
- Use AI assisted coding and Huggingface to download the best performing software.
    - And follow developments in the ML literature — use arXiv + blogs + AI summaries

Patterns, Patterns, Patterns



Patterns of ‘light’ in cities, cosmology and neurons.

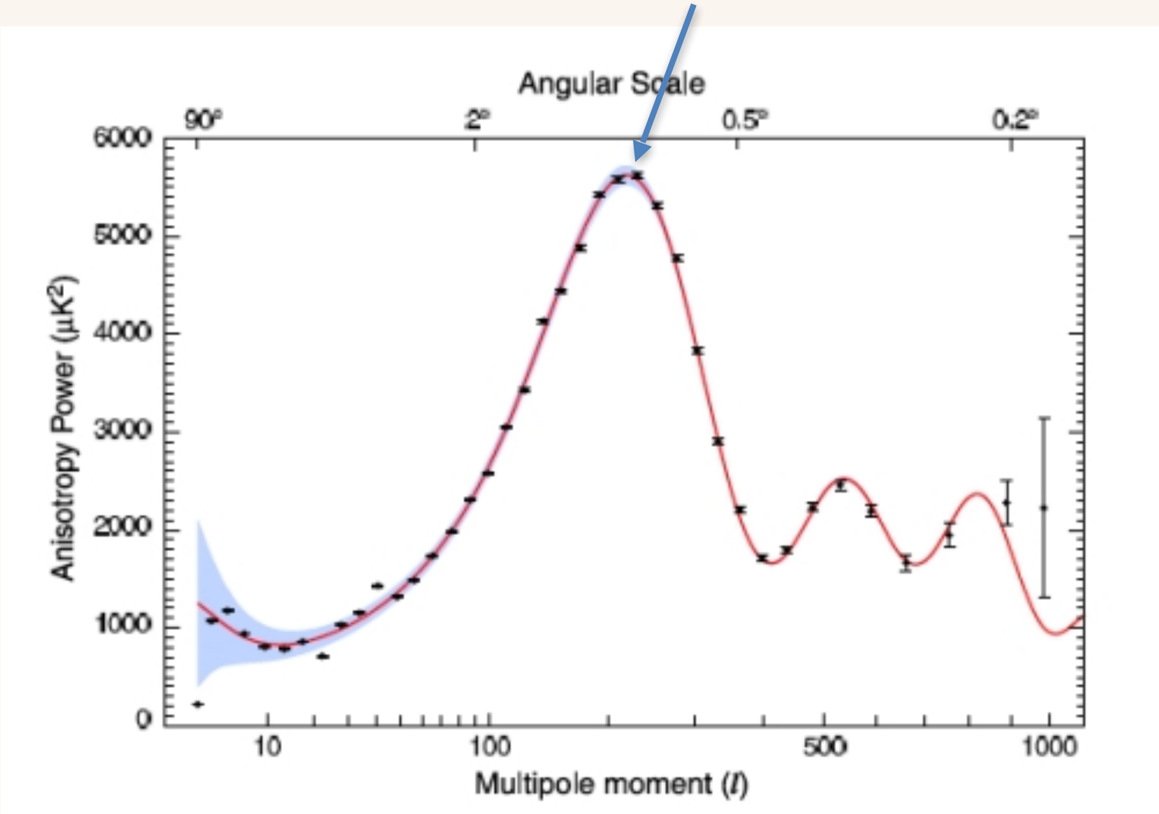
A great time for scientists and AI experts to work together!

Thank you Hitoshi, Masahiro and all the organizers!

Spare slides

The CMB power spectrum

1 degrees on sky

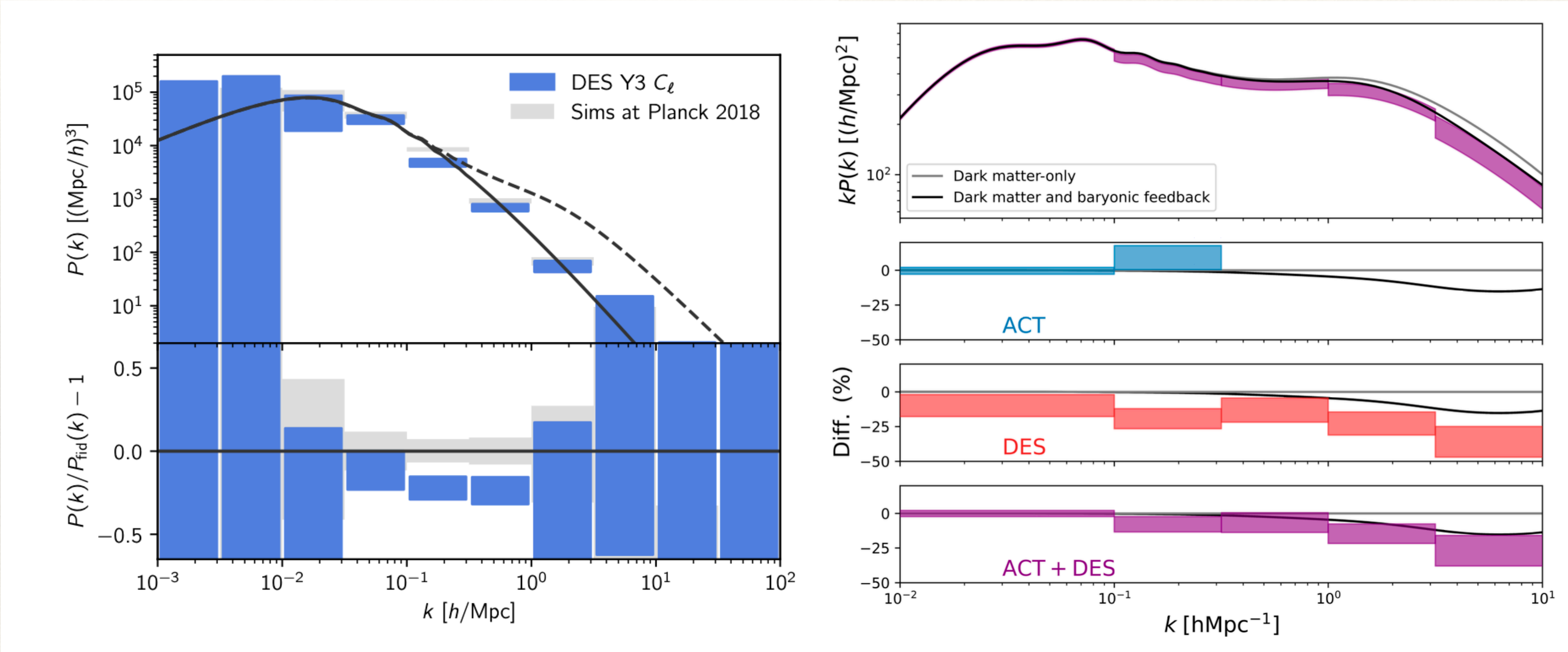


Power spectra of CMB temperature fluctuations.

Blue curves: best fit ΛCDM model with 6+ free parameters.

Planck Collaboration 2018

Structure mismatch – the S_8 tension



Reconstructing P(k,z): a first pass

Doux, BJ+ (DES) 2022; Sarmiento+ 2025, arXiv: 2502.06687