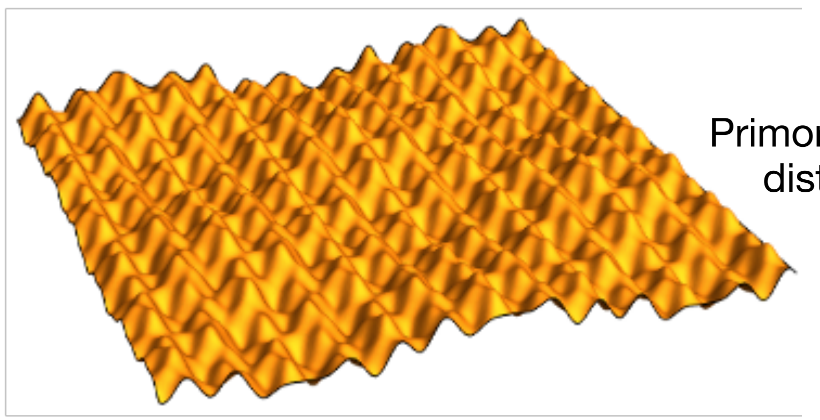
Two machine learning based methods for field-level inference of cosmological parameters

Machine Learning at GGI - Conference
09/05/2022

Moritz Münchmeyer, UW Madison

Primordial physics vs structure formation

(Near)-Gaussian initial conditions from inflation

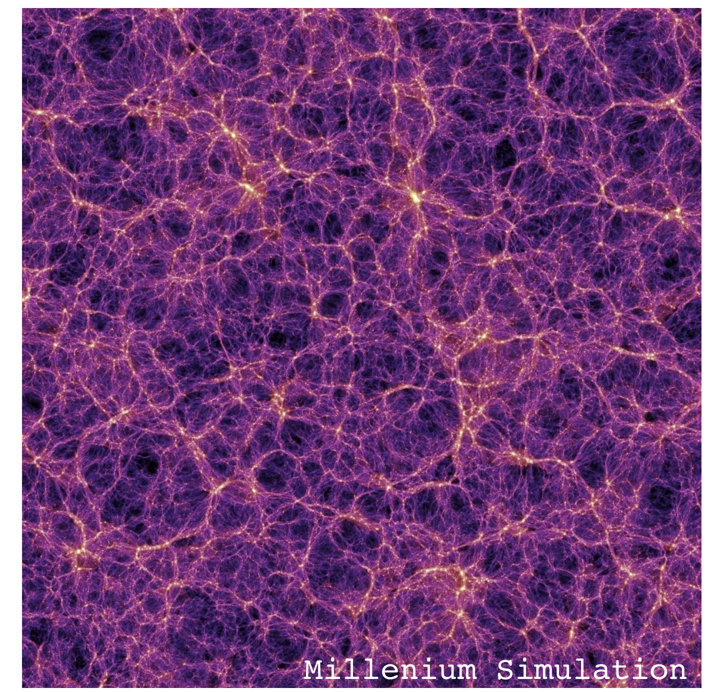


Primordial matter distribution

13.7 billion years

Cosmological time evolution

Matter distribution today

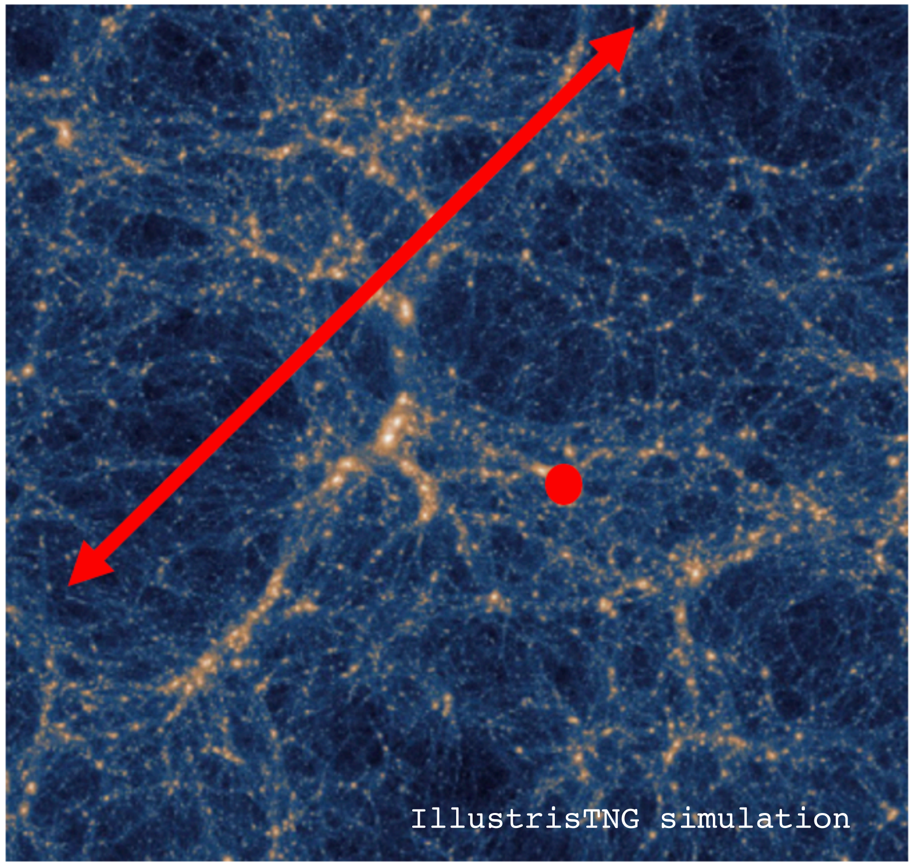


Observed galaxies

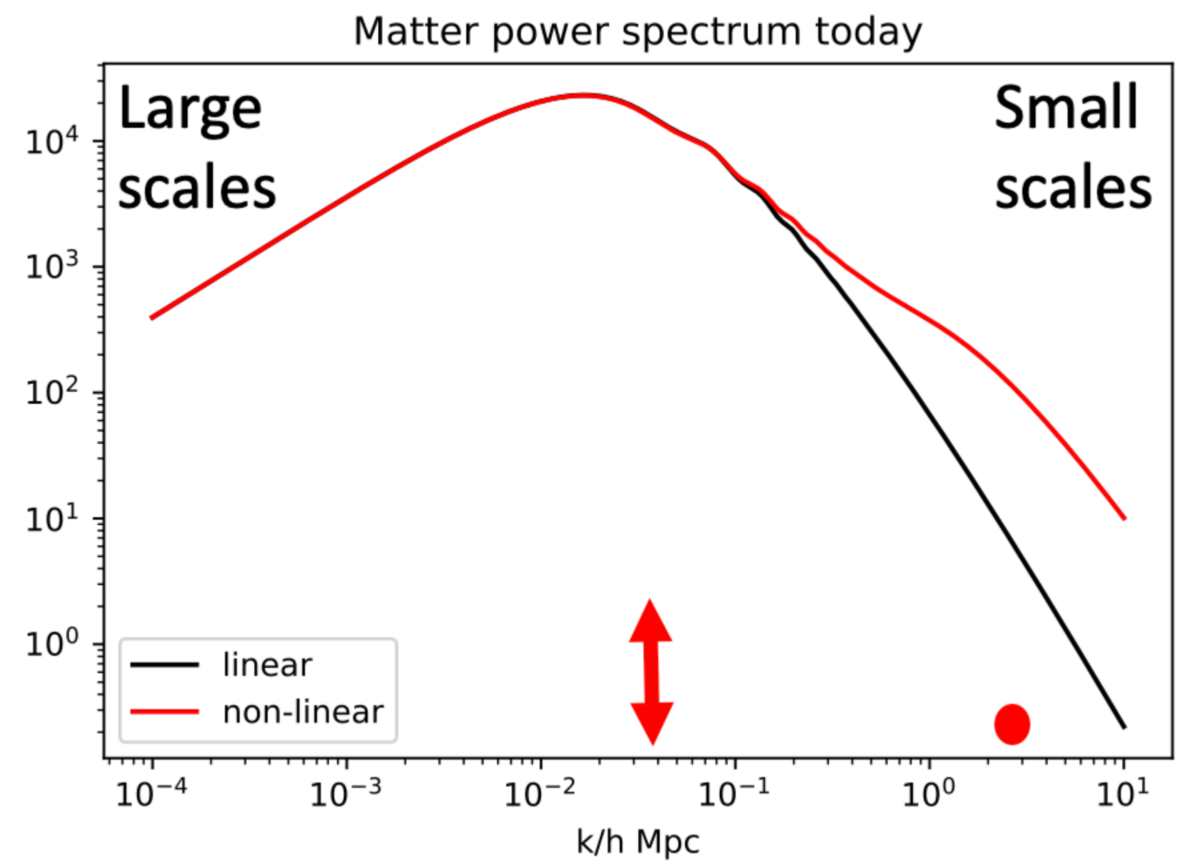
inference

Goal: Measure parameters of initial conditions and evolution from the matter (galaxy) distribution observed today

Different scales in cosmology



Power spectrum



Large scales evolve linearly → **Easy to use for cosmology**

Small scales are very nonlinear and depend on complicated astrophysics → **Hard to use for cosmology but MUCH more information in principle**

Chances and challenges for Machine Learning

- For many upcoming experiments **classic data analysis methods** (e.g. power spectrum) **become very suboptimal**.

- Instead, many groups develop methods based on **forward modelling, simulation based inference, machine learning** and combinations thereof.

- General challenges:

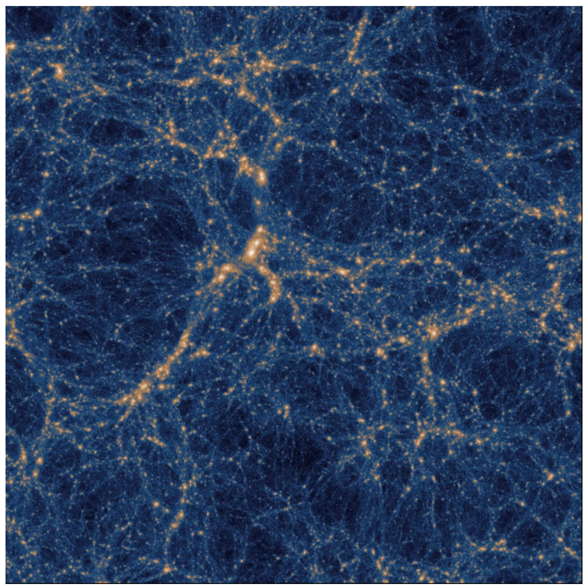
  - If these methods are applied at small enough scales so that gains in statistical sensitivity are large, they become **sensitive to uncertain small-scale baryonic physics**.

  - **Simulations are very costly**, so that naive sampling over them is intractable. Need approximations; hard to quantify error bars.

- **This talk**: 2 new methods that help with these problems in some cases.

Flavours of simulation-based inference

Neural Network estimates



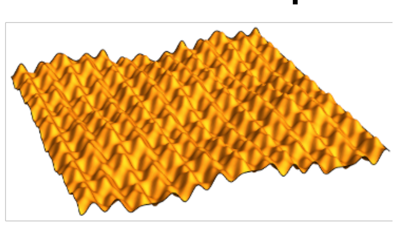
Neural Network

Cosmological parameters: f_{NL}, A_s etc.

Train NN on large set of simulations which are hopefully reliable in this domain (this is being studied intensely of course).

Forward modelling

Primordial perturbations

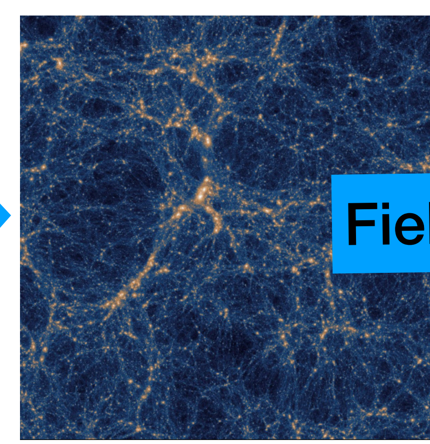


“Known” PDF

Differentiable forward model

Can be based on:
N-body simulations
Analytic PT
Machine Learning

Matter/Galaxy distribution

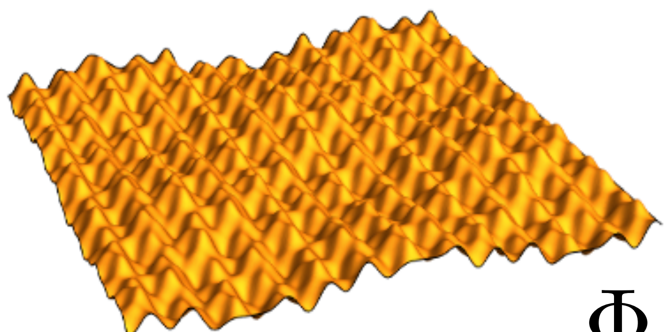


Field-level Likelihood

Cosmological parameters: f_{NL}, A_s etc.

Primordial density perturbations

- Models of the early universe (inflation) make different predictions for the statistical properties of the matter and radiation distribution at the beginning of the universe.



Φ

Primordial perturbations =
initial conditions of the universe

- To good approximation perturbations are Gaussian, but small "**non-Gaussianities**" are predicted by different models.

- A gaussian field is fully defined by its **power spectrum** $P(k) \propto \langle \Phi(k)\Phi(k) \rangle$

- Non-Gaussian fields have **higher order correlations** $\langle \Phi(k_1)...\Phi(k_N) \rangle$
- **Goal: Detect or constrain non-Gaussianity.**

Primordial non-Gaussianity

- **Planck CMB constrained the 3-point correlation function (bispectrum):**

$$\langle \Phi_{\mathbf{k}_1} \Phi_{\mathbf{k}_2} \Phi_{\mathbf{k}_3} \rangle \propto f_{NL}$$

- Roughly: Non-Gaussianity is constrained to be ~10^{-4} smaller than Gaussian part. The minimum possible value is ~10^{-7}.

- There are many different shapes of 3-point correlation functions.

- **Near term goal**: **Probe multi-field inflation = "local non-Gaussianity"**

  - Current constraint f_{NL} = -0.9 ± 5.1 (from Planck) must get 10 times tighter.

- **Long term goal: Detect masses, couplings and spins of primordial fields.**

Part 1: A robust neural network enhanced method for local non-Gaussianity inference

Based on: arxiv: 2205.12964

Collaborators:

Utkarsh Giri,
Postdoc at UW Madison

Kendrick Smith,
Perimeter Institute

Warmup: Measuring σ_8 with a CNN

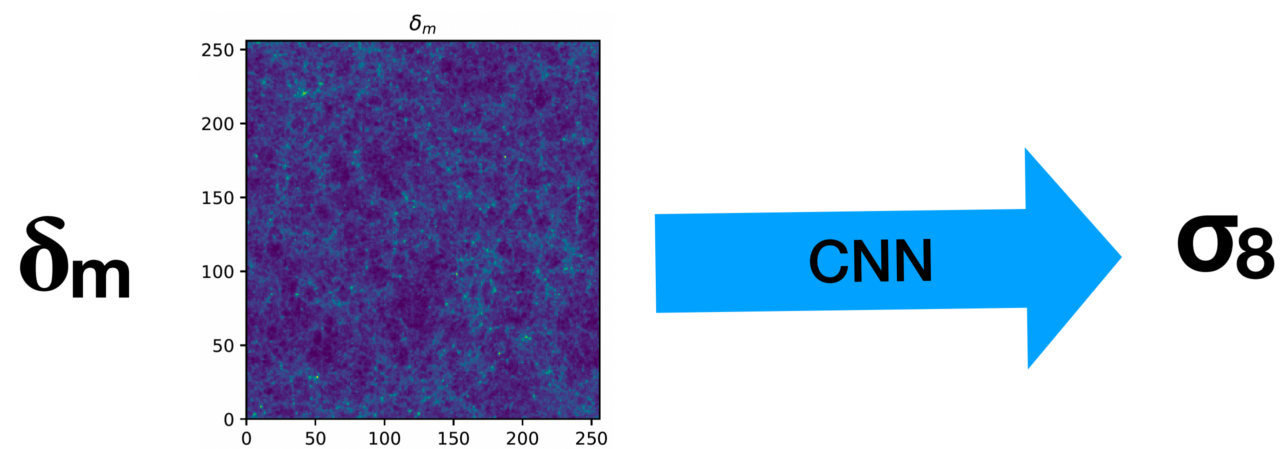
CNNs give **very strong constraints on cosmological parameters**, IF they are allowed to use "very" non-linear scales. Given enough capacity and training they should give the optimal constraint.

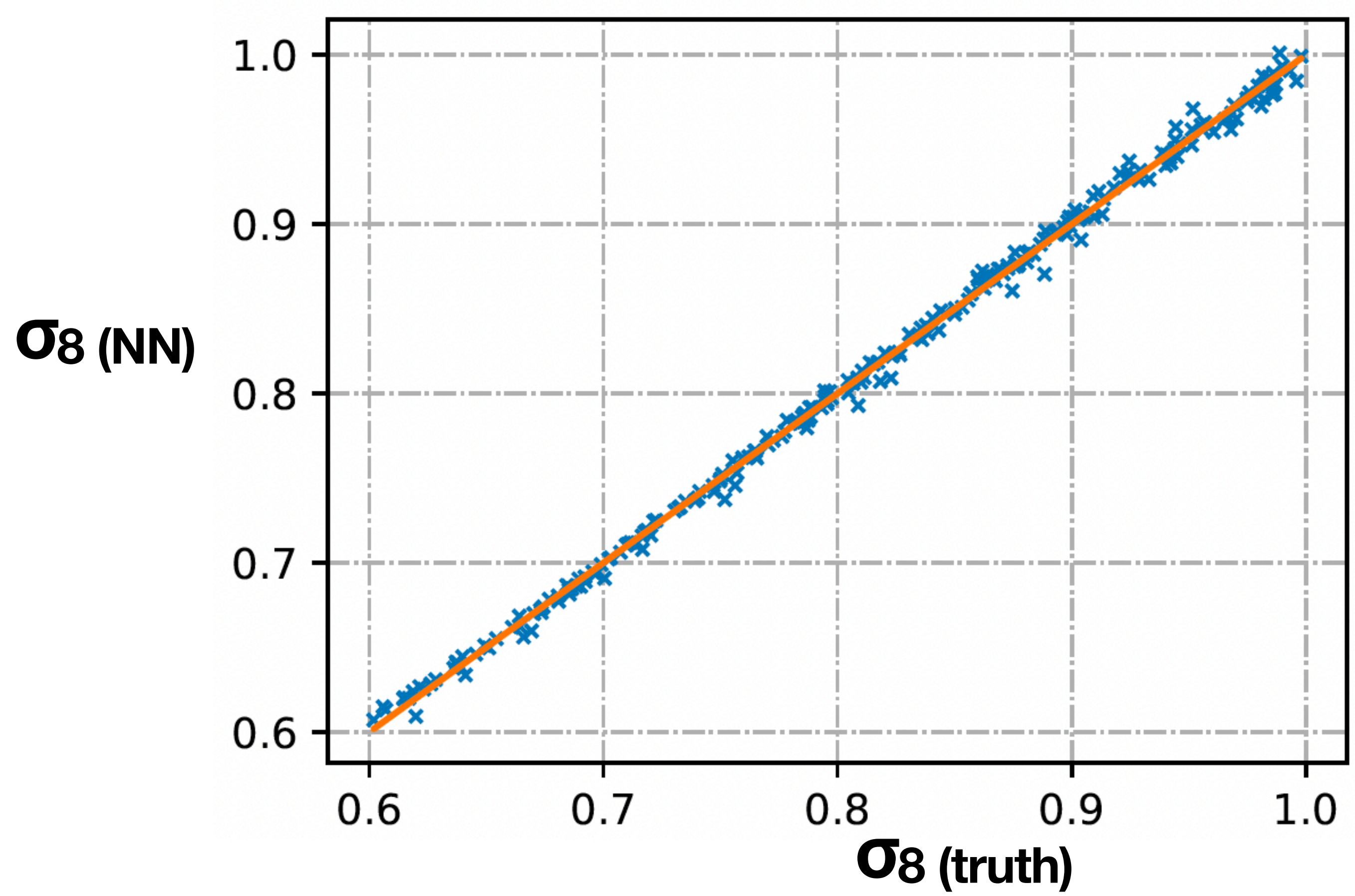
Example:

Train **CNN to estimate the linear fluctuation amplitude** (called σ_8) from the matter field.

For a cosmological volume V=1 Gpc^3 and with 256^3 pixels, the precision is ~ percent level.

Problem: not trustworthy.

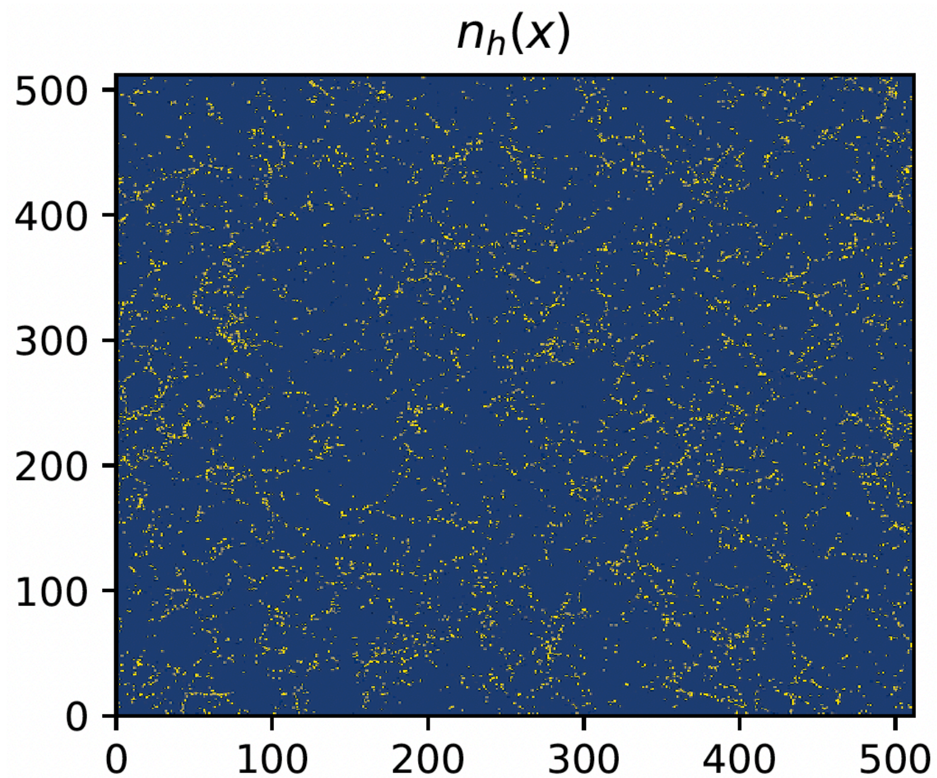




Warmup: Measuring σ_8 with halo counting

- A **simpler way to measure σ_8 would be simply count how many halos/galaxies are in a given cosmological volume**.
- In a given simulation, n_{halo} is very sensitive to σ_8.
- However nobody would believe this measurement because it is **highly dependent on uncertain small-scale physics** that governs the formation of halos/galaxies.

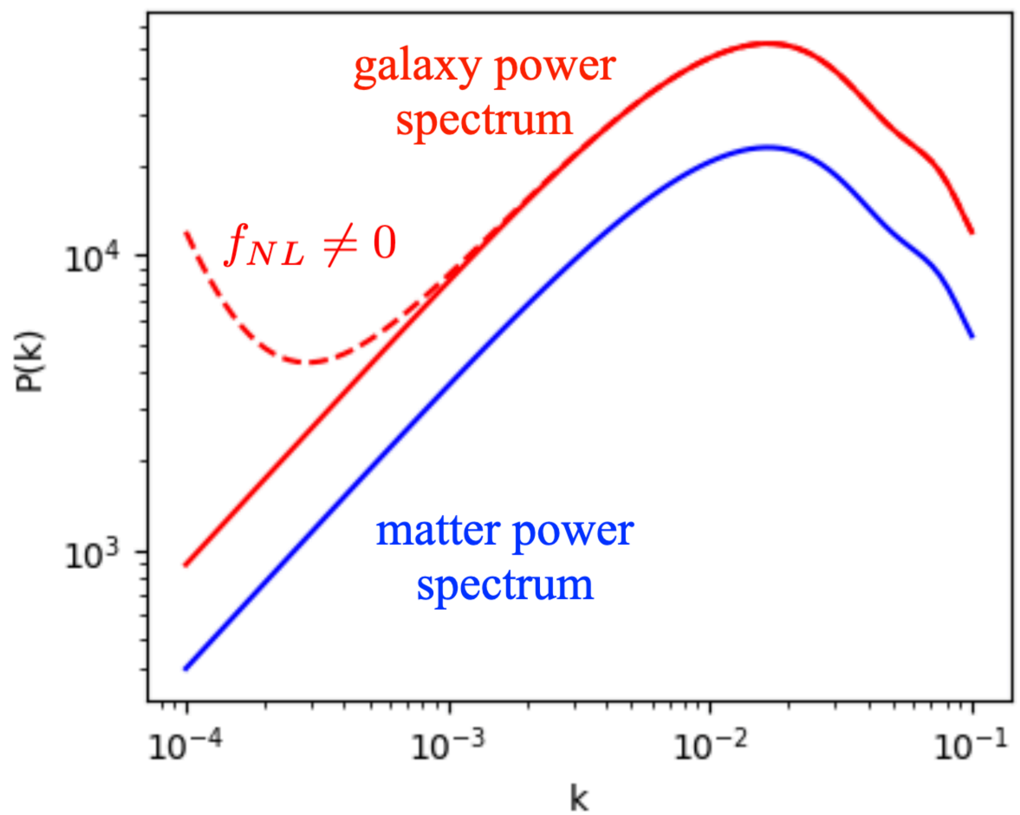
n_h



σ_8

Scale-dependent bias of the halo field

Local non-Gaussianity f_{NL} generates an excess clustering on large scales. This effect is called "**scale dependent bias**" of the halo field.



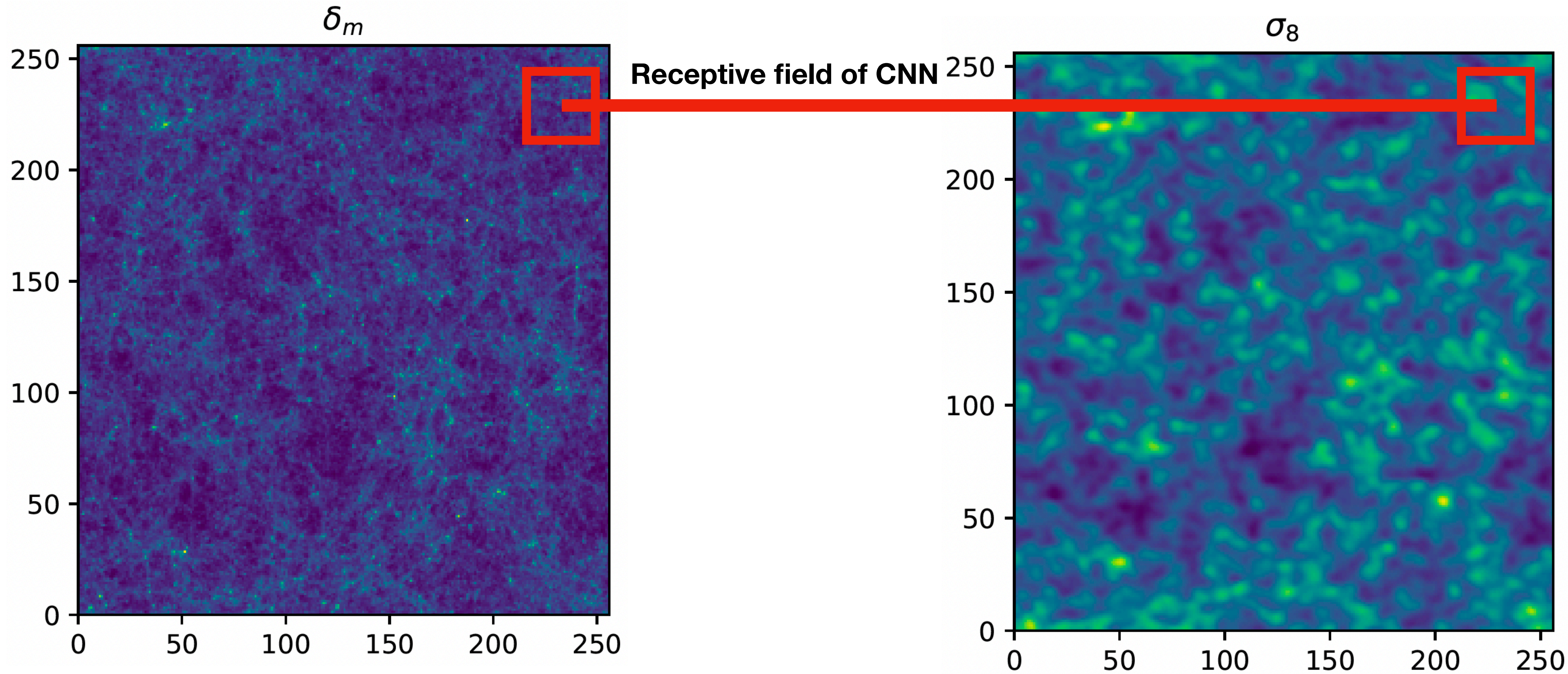
$$\delta_h(\mathbf{k}_L) = \left(b_h + \beta_h \frac{f_{NL}}{k_L^2} \right) \delta_m(\mathbf{k}_L)$$

A famous result in cosmology is that the "kink" in the power-spectrum cannot be introduced by non-linear astrophysics. This **robustness is ultimately a consequence of Einstein's Equivalence Principle**.

We have a symmetry protected observable. We now want to enhance its SNR with a NN without spoiling the robustness.

CNN for local σ_8 measurements

Idea: Local non-Gaussianity f_{NL} is **a large-scale modulation of local power**. Therefore if we have a **neural network that optimally probes local power (i.e. local σ_8)**, it would be the ideal field to base an f_{NL} estimate on.

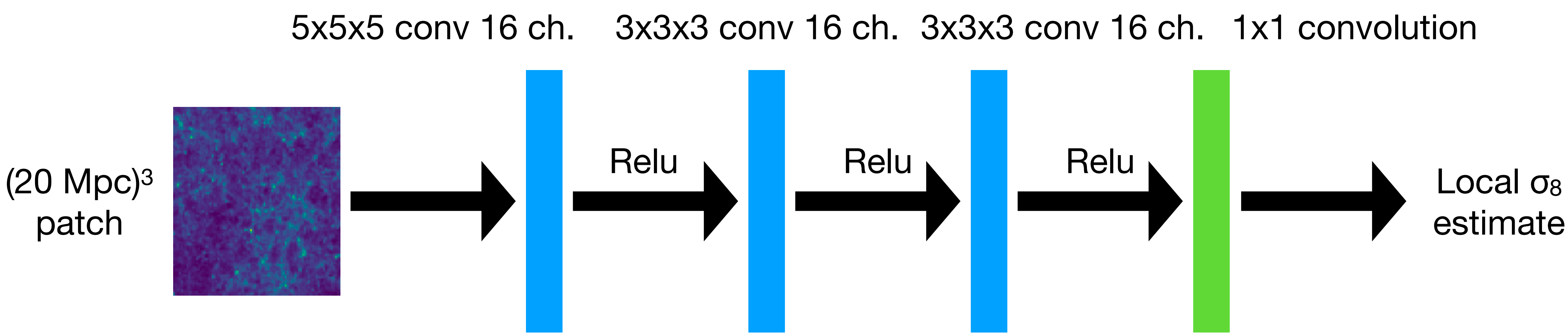


This local **NN output field will also have a scale-dependent f_{NL} bias.**

Details of the implementation

- We use a simple fully convolutional CNN, which **can run on any size of input data** (no GPU memory problem for realistic data sets!).

- **Simple CNN architecture** (not very optimized):

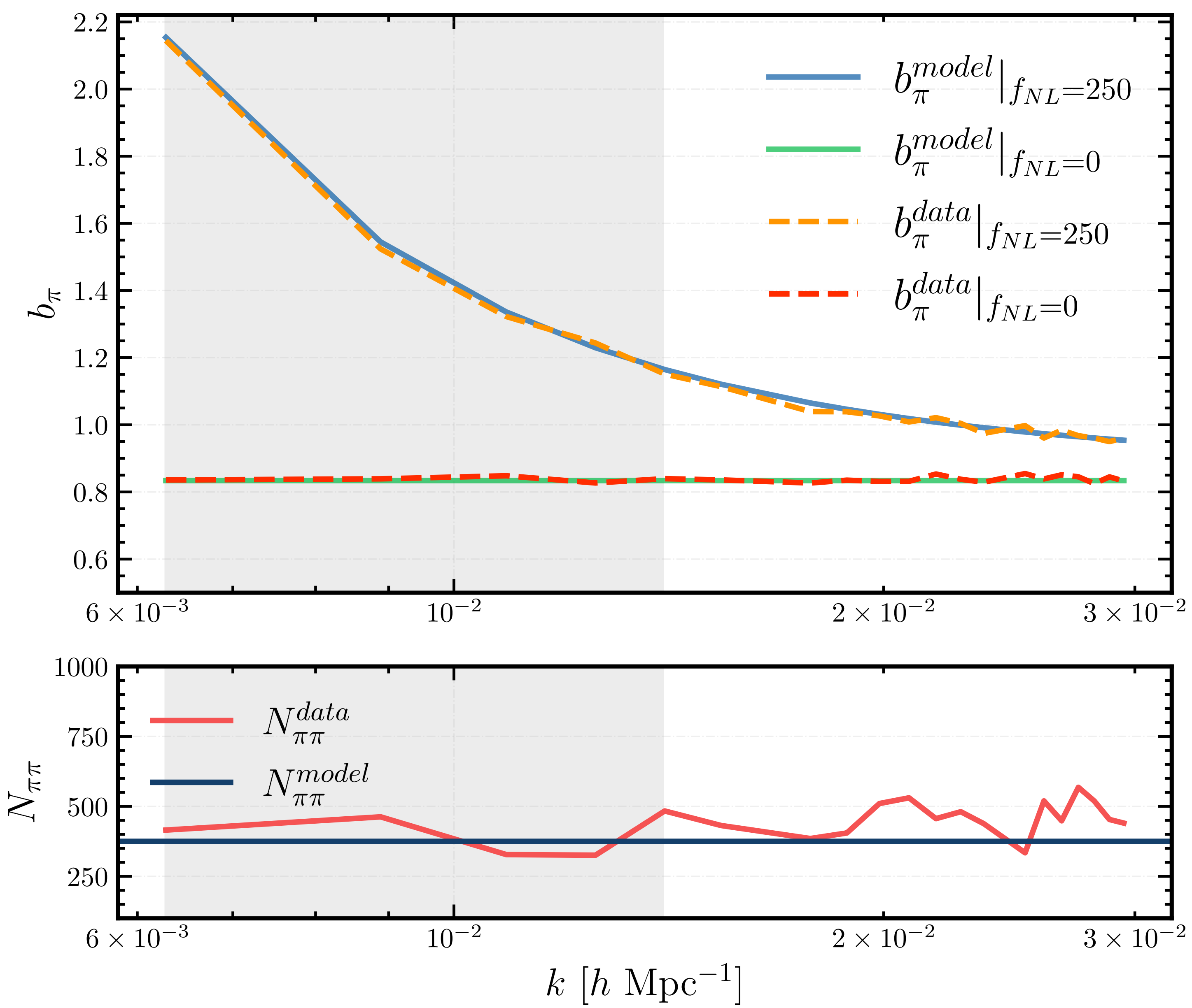


- Training: **800 Quijote Gaussian (!) simulations with different σ_8 values.**

- After training, we analyze independent simulations and run an **MCMC power spectrum chain on the CNN generated field to determine f_{NL} and bias.**

The analytic model matches the NN output

“Halo bias” of the neural network output field π



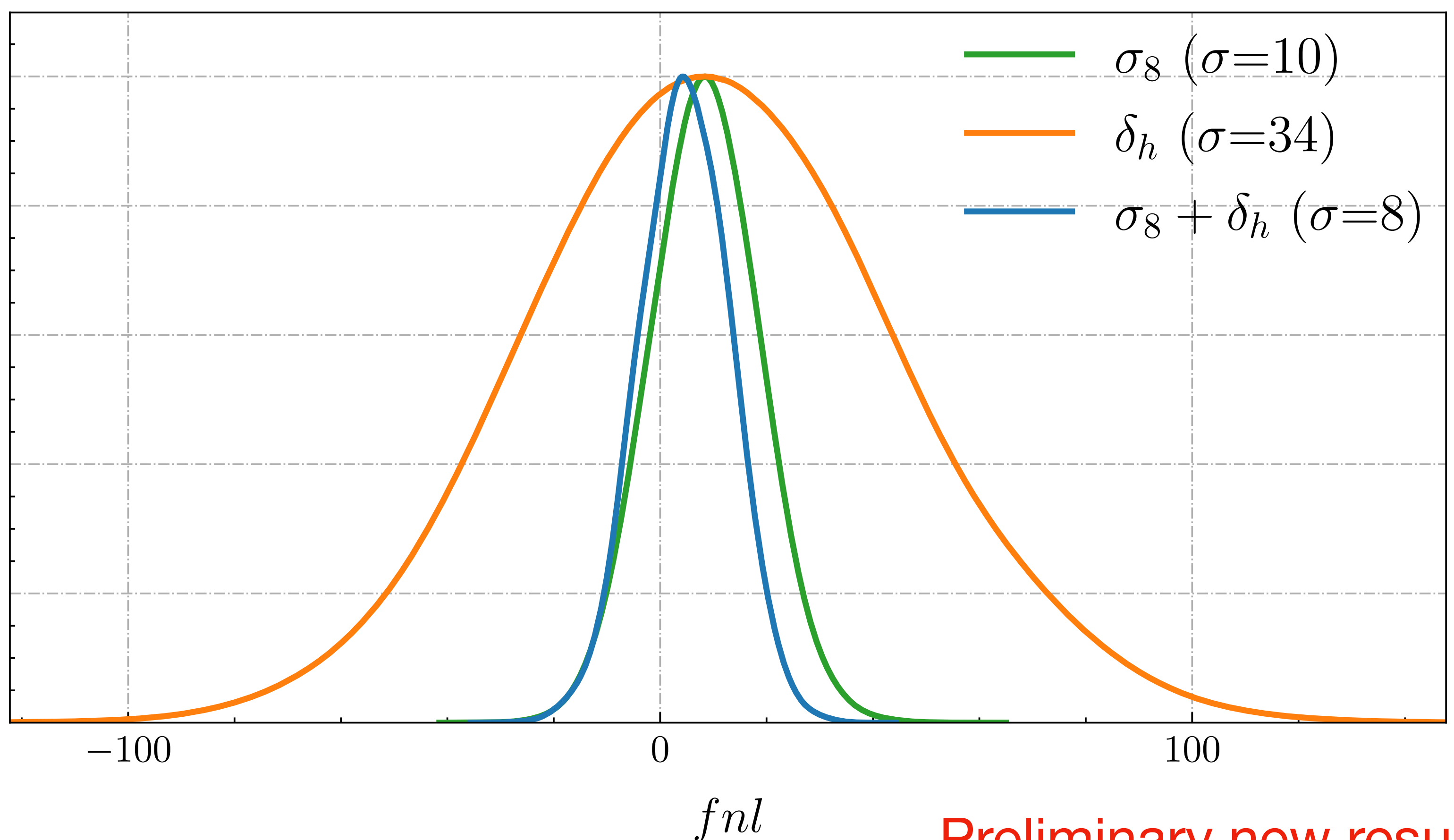
Flat noise power spectrum

The signal power spectrum and noise power spectrum of the NN output behave exactly as we predict analytically.

We also have some a proof that our method is optimal under certain circumstances.

Result: 3x improvement on f_{NL}

Comparison of fNL constraint between ordinary scale-dependent halo bias and our new result where the mass-weighted halo field is first processed with a neural network.



For the Quijote simulation suite, which has a realistic halo density, we find that our NN outperforms halo bias by a significant factor (here about a fact 3).

The method should work on real data with a significant improvement factor.

Preliminary new result, not in the paper

Side note: Recently halo catalogues have been analyzed using **graph neural networks** (see talk by Farida Farsian). A local graph neural network would also work with our approach and perhaps be a bit more sensitive.

Summary

- We **combined an analytic method to measure f_{NL} with a neural network to get the best of both worlds**: robustness to baryonic effects and optimal statistical sensitivity.
- **Next step: Determine the f_{NL} constraint our method can reach on upcoming surveys such as Rubin Observatory**
- **Application to similar problems**
    - A **large-scale modulation of small scale power** is a very common setup in cosmology. E.g. quadratic estimators for various fields.
    - If the small-scale fields are substantially non-Gaussian, a local neural network operation can improve the SNR. Again, **small-scale physics will show up as biases that can be marginalized,** but won't affect robustness.

Part 2: Applications of normalizing flows to cosmology at field level

Based on: arXiv:2105.12024, Neurips 2021 Workshop Machine Learning and the Physical Sciences + new work

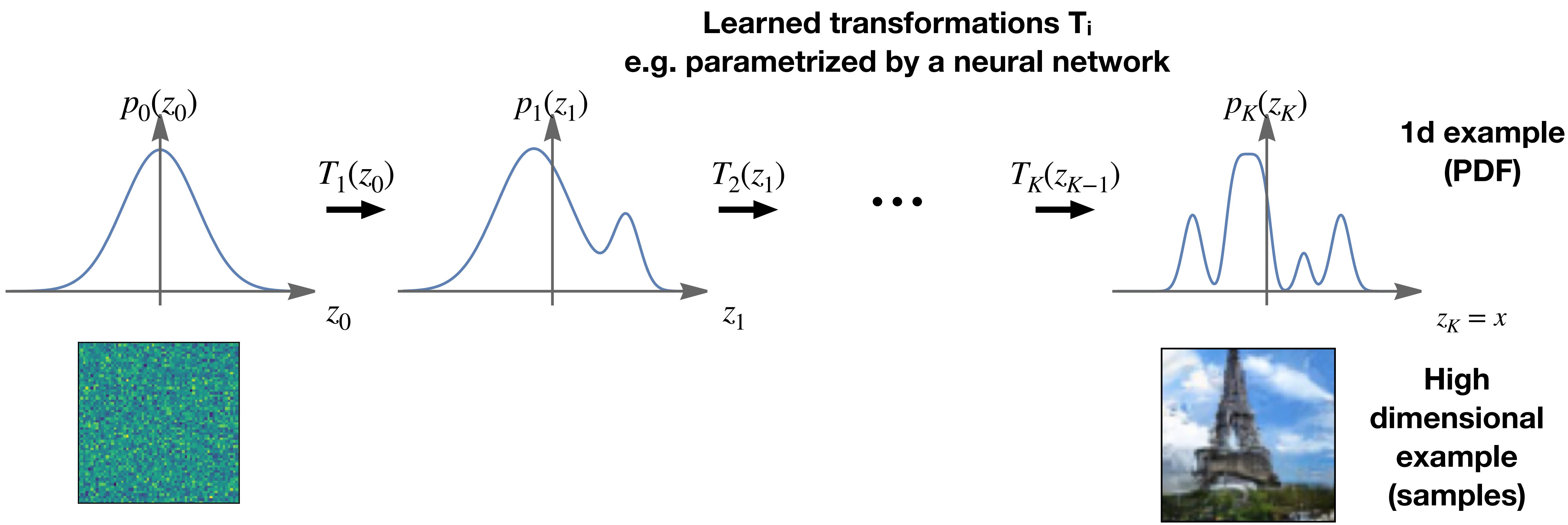
Collaborators:

Adam Rouhiainen,
Grad student at UW Madison

Utkarsh Giri,
Postdoc at UW Madison

Quick introduction to normalizing flows

- Normalizing flow: Series of learned transformations that **deform a simple base distribution into a complicated target distribution**.



- Difference with most other ML methods: We learn a probability distribution, rather than an arbitrary input->output mapping.

- Recently used in physics in particular in lattice QCD (e.g. review 2101.08176) and likelihood free inference (e.g. 2105.12024)

Mathematical formulation

- Transformation T (the "flow") → Change of variables of PDF

$$\mathbf{x} = T(\mathbf{u})$$

$$p_x(\mathbf{x}) = p_{\mathrm{u}}(\mathbf{u}) \, |\det J_T(\mathbf{u})|^{-1}$$

- Architecture choice: We **want T(u), it's inverse T^{-1}(x) and the Jacobian J_T(u) to be computationally efficient** to apply the method to large dimensions.

- After training **two basic operations can be performed**:
  - **Sampling from the distribution** (forward mode)

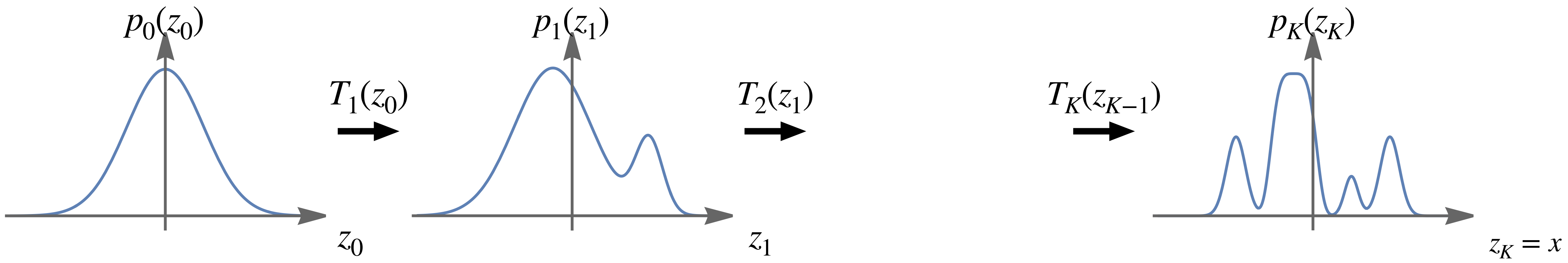
    Base distribution sample u → **Target sample x**

  - **Exact density evaluation** (backward mode)

    Sample x → **p(x)**

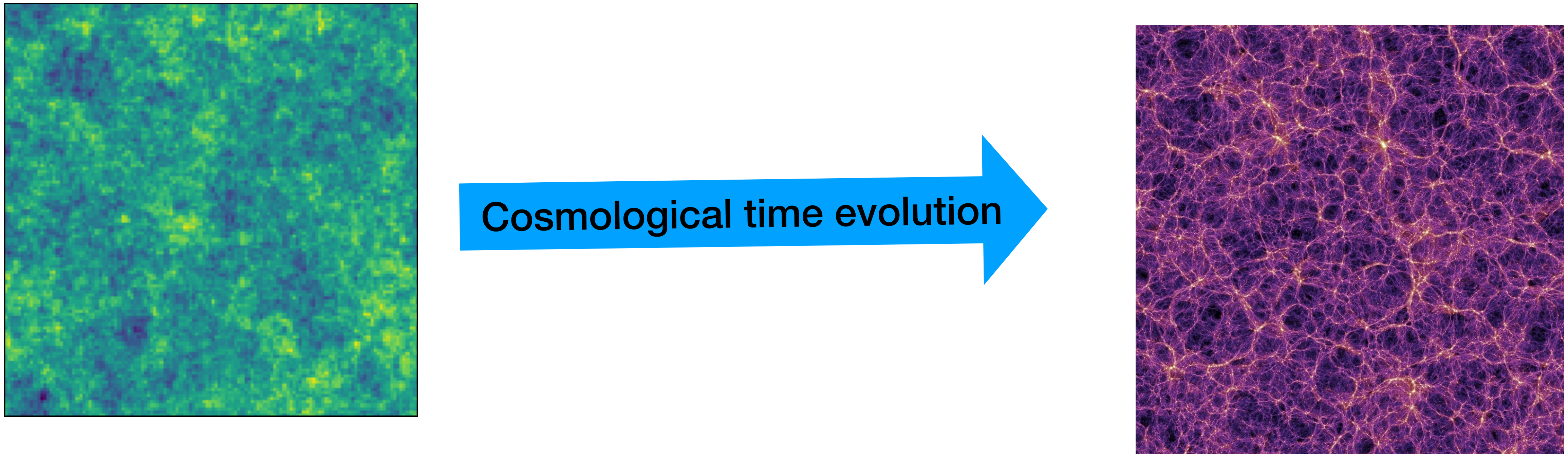
Analogy with cosmic structure formation

(Near-) Gaussian initial conditions PDF morphs into complicated late-time matter distribution.



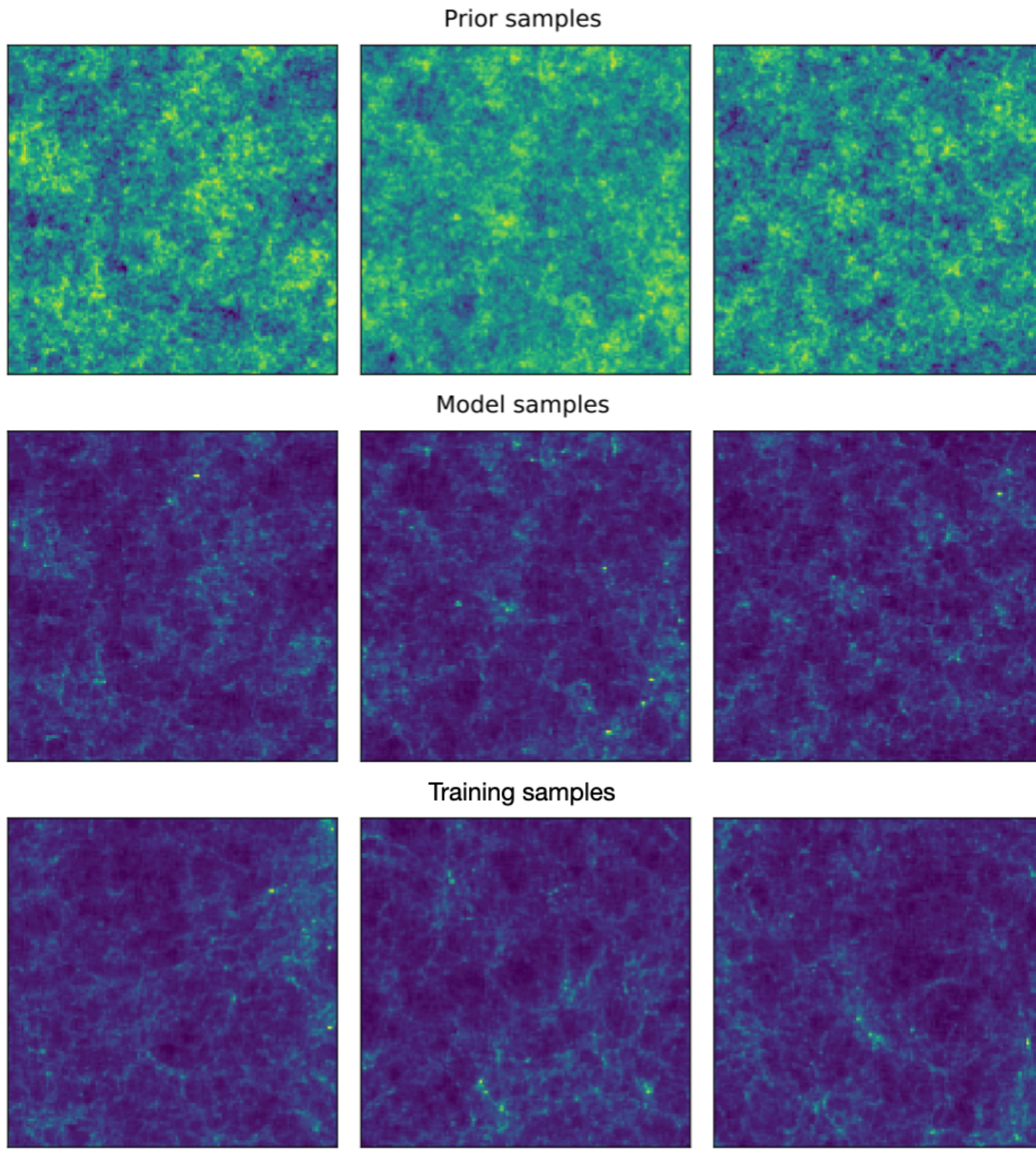
Gaussian primordial matter perturbations

Non-gaussian matter/galaxy distribution today



Flowing from a correlated Gaussian to the matter distribution

In cosmology: Flow from a physically motivated prior PDF: The gaussian field with the right power spectrum.



Flow: RealNVP

The flow learned to deform a Gaussian PDF into a highly non-Gaussian PDF

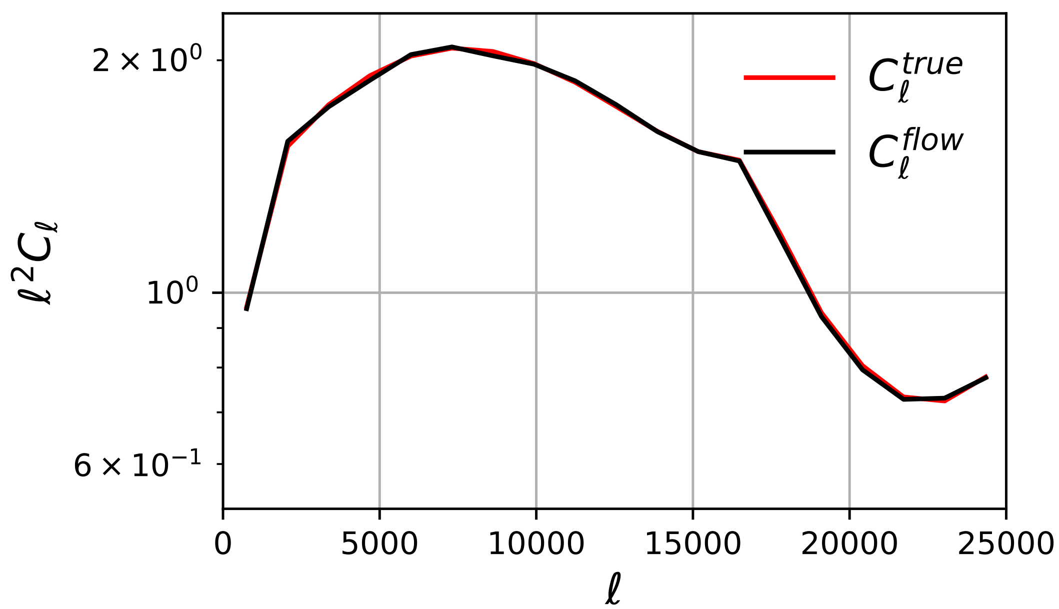
Density peaks match, as in physical structure formation (even though this is not explicitly trained for or needed).

$125\ h^{-1}$ Mpc

About 128^3 particles per field

Quality control

- **Power spectra and non-Gaussianity agree very well (here "Glow" flow).**



these fNL values are normalized so that 1 is O(1) non-gaussianity

$\tilde{f}_{NL}^{\text{local}}$ training	$\tilde{f}_{NL}^{\text{local}}$ flow	$\tilde{f}_{NL}^{\text{equi}}$ training	$\tilde{f}_{NL}^{\text{equi}}$ flow
0.85 ± 0.40	0.85 ± 0.37	0.61 ± 0.24	0.63 ± 0.30

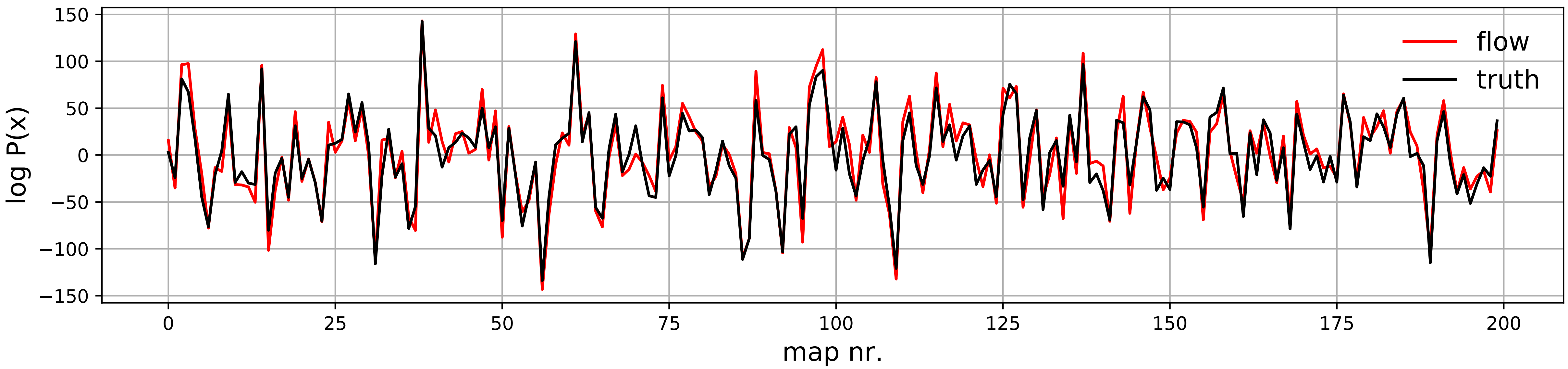
Sigmas are per sample

- **Normalizing Flows can learn to sample from the non-Gaussian PDF of the cosmological matter field. Does density evaluation also work?**

Is density evaluation working on IID samples?

We tried flowing from uncorrelated Gaussian noise to a physical correlated Gaussian, to compare p_{flow} and analytic p_{true} to build confidence.

p_{flow} and p_{true} for 200 fields (zero-centered):



The cross-correlation coefficient between $\ln p_{\mathrm{flow}}$ and $\ln p_{\mathrm{true}}$ is $r \approx 0.993$.
This shows how flow can be used for density evaluation with high accuracy.

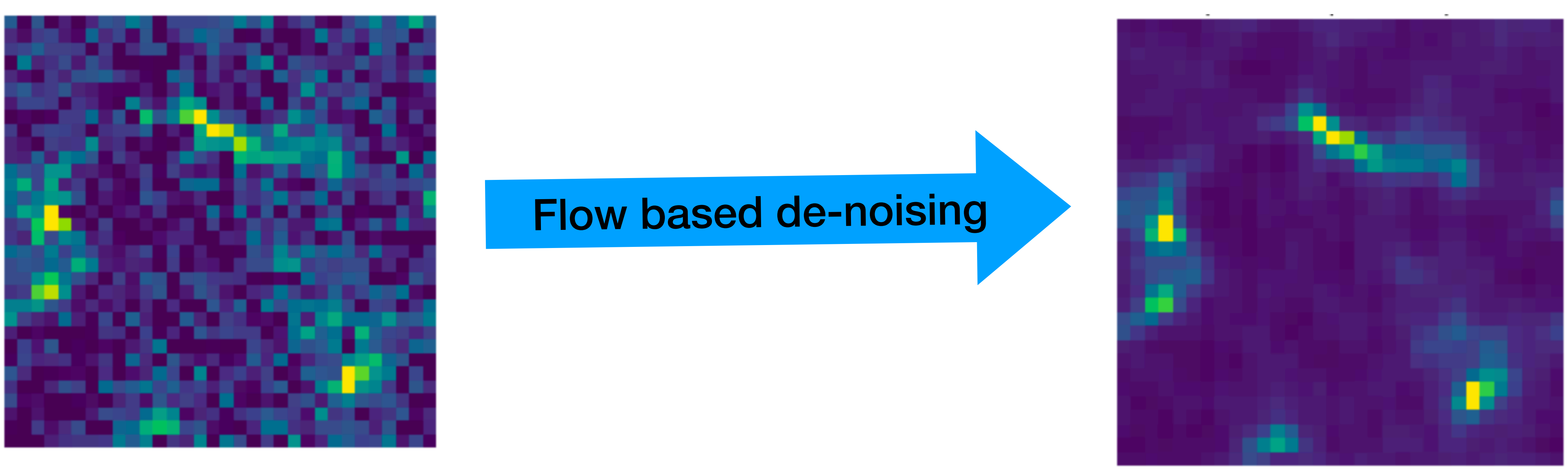
However: **Quality of density estimation depends on training set size and number of parameters of the flow and dimension of the map.**

Application: non-Gaussian priors

- In data analysis in cosmology we often make use of Gaussian priors. This is no longer justified for very high resolution observations. Using the trained normalizing flow we can now **include non-Gaussian priors** (mildly OOD)**:**

$$\ln p(y \mid d) = -\frac{1}{2}(y-d)^{\mathrm{T}} N^{-1}(y-d) - \ln p_{\mathrm{flow}}(y)$$

- We use a flow trained on simulations of the matter distribution. Then we use this knowledge of the matter PDF to **de-noise an observation of a matter field by maximizing the posterior**.



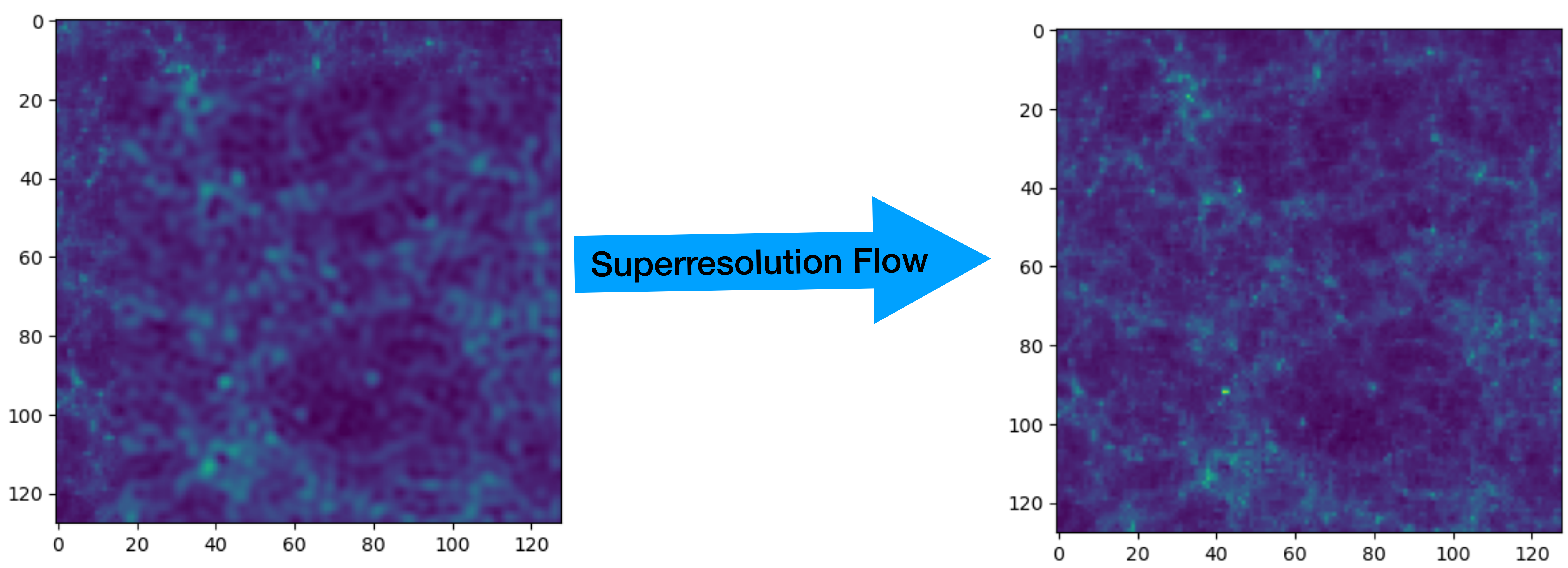
Similar to denoting with score matching.

Applications: Superresolution emulators

- In a second application we use **the flow as a generative model**, with a well-defined probabilistic interpretation.

- A **conditional flow** can learn how non-Gaussian small-scale structure reacts to large-scale structure, probabilistically, at field level. $P(\text{small-scale structure} \,|\, \text{large-scale structure})$

Example (using a conditional RealNVP flow):



Summary

- Flows at field level are an interesting tool for cosmology. They are **dual purpose for inference and data generation**.

- We are currently exploring in particular:

  - Flow performance for **non-Gaussian priors** (smaller networks, more symmetries, better OOD performance). Application to **very non-Gaussian small-scale field observations (e.g. kSZ)**.

  - Conditional flows to model **large-scale to small-scale coupling**.

  - Include **cosmological parameter dependence** in the flow (see Uros Seljak talk)

  - **Variational inference** of the field posterior with flows.