Title: Deep Probabilistic Models for Cosmological Analysis and Beyond

Speakers: Biwei Dai

Series: Colloquium

Date: March 05, 2024 - 11:00 AM

URL: https://pirsa.org/24030099

Abstract: Current and future weak lensing surveys contain significant information about our universe. However, their optimal cosmological analysis is challenging, with traditional analyses often resulting in information loss due to reliance on summary statistics like two-point correlation functions. While deep learning methods offer promise in capturing the complex non-linear features of these cosmological fields, they often suffer from issues such as inadequate uncertainty quantification, susceptibility to distribution shifts, and interpretability limitations, which hinder their scientific applicability. In this talk, I propose a novel approach leveraging generative probabilistic modeling with Normalizing Flows to learn the data likelihood function at the field level, facilitating more effective cosmological information extraction. This framework not only enables anomaly detection of distribution shifts to improve the robustness of the analysis, but also fostering interpretability via generated samples. I will also discuss incorporating physical prior knowledge, such as symmetries and multiscale structure, into the model architectures to improve their generalization capabilities. Finally, I will explore the broader implications of deep probabilistic models in physics, highlighting their potential applications in diverse areas ranging from astronomical observations to high-energy physics and lattice field theory.

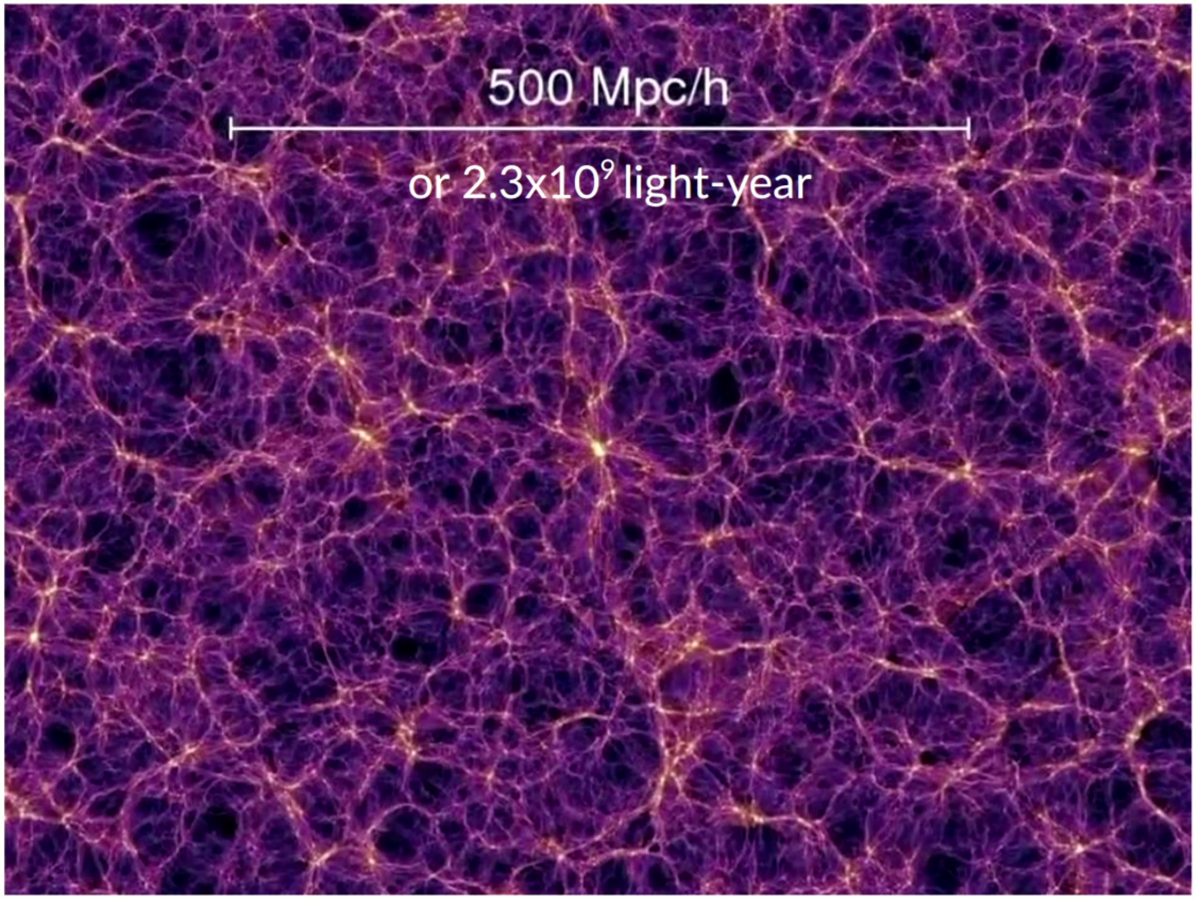
Zoom link

Deep Probabilistic Models for Cosmological Analysis and Beyond

Biwei Dai
UC Berkeley
March 5 @ Perimeter Institute

works with Uroš Seljak, Divij Sharma, Xiangchong Li, Rachel Mandelbaum, Francisco Villaescusa-Navarro, Richard Grumitt, George Stein

The Large-Scale Structure of the universe



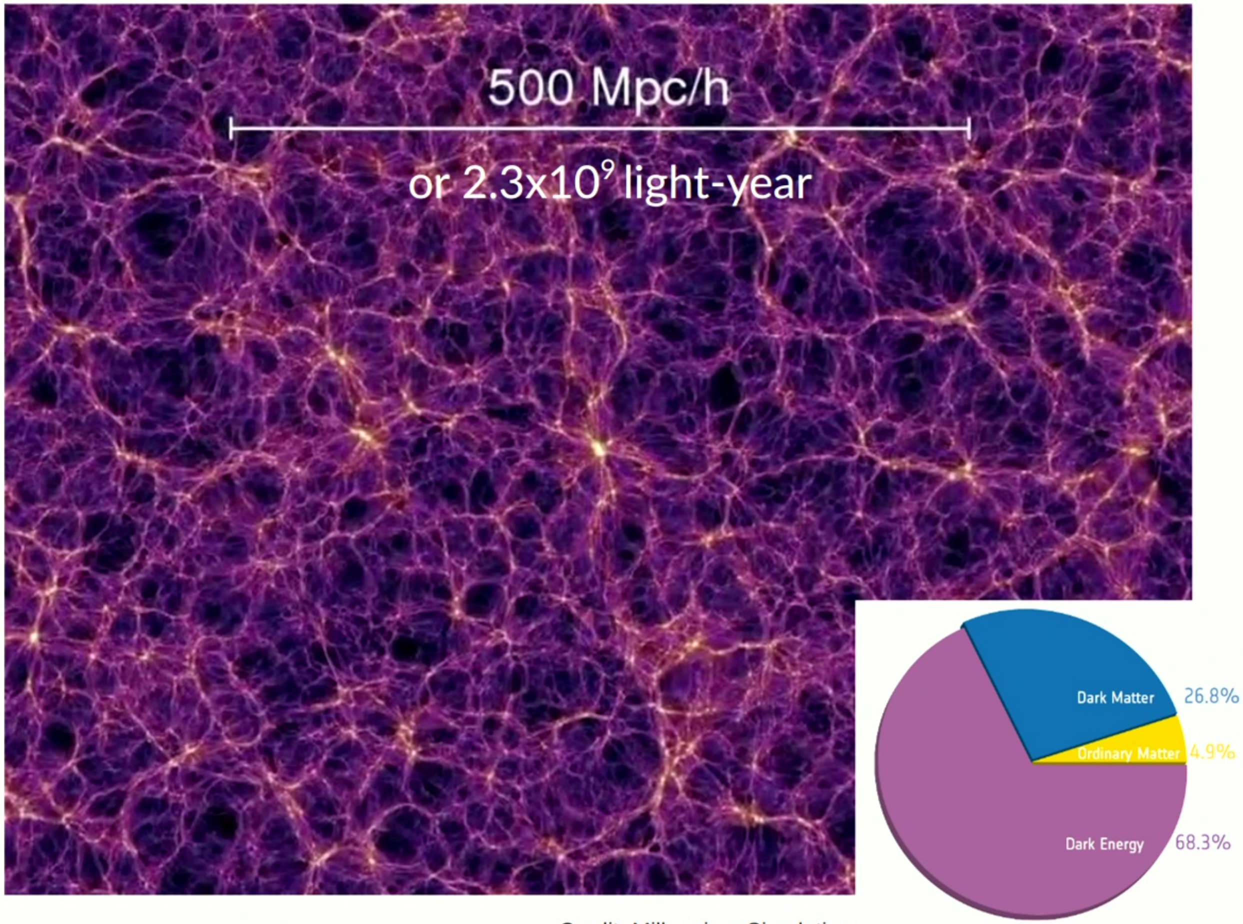
Credit: Millennium Simulation

The large-scale structure evolves under laws of gravity and expansion of the universe. It encodes significant information about our universe:

Values of cosmological parameters
Dark matter and dark energy
Gravity laws
Neutrino mass
Inflation, initial condition of the universe
New physics
...

2

The Large-Scale Structure of the universe

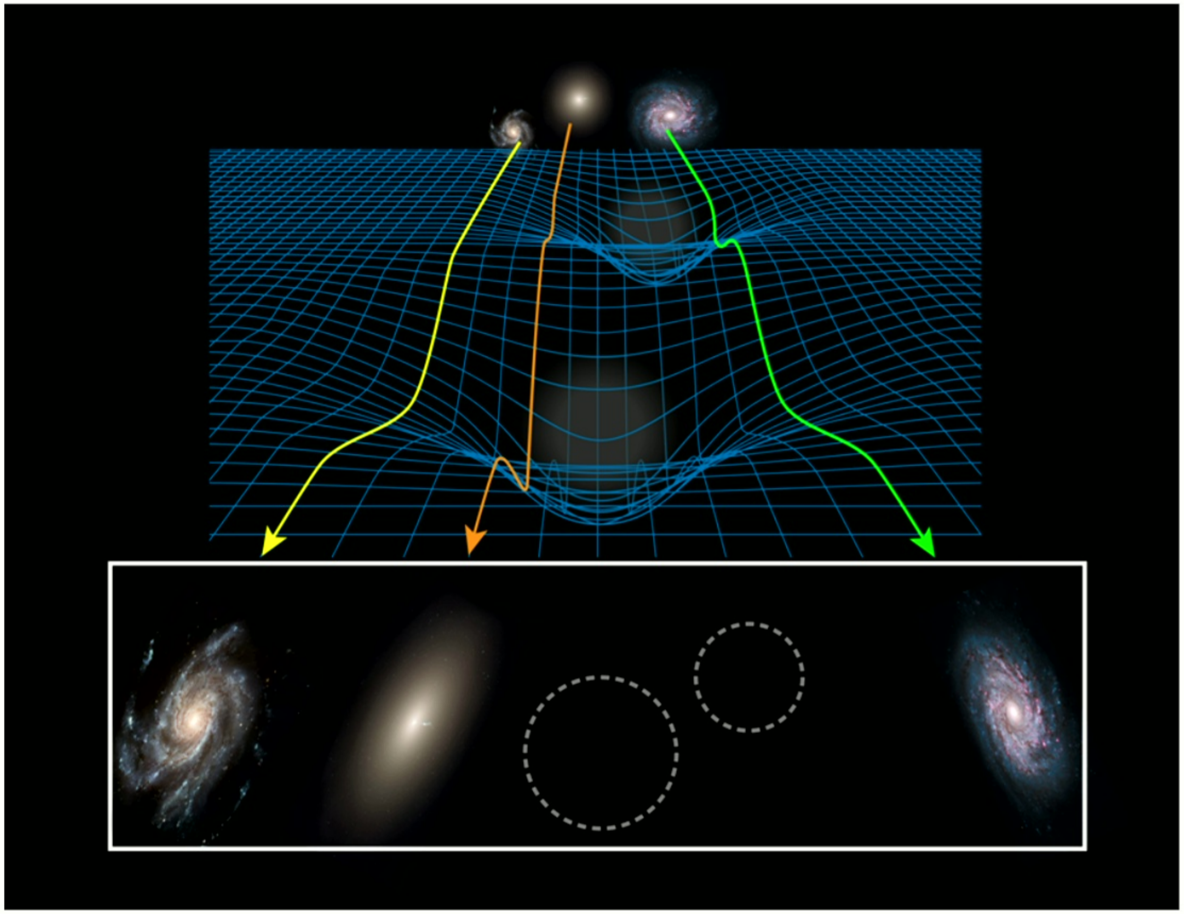


Credit: Millennium Simulation

The large-scale structure evolves under laws of gravity and expansion of the universe. It encodes significant information about our universe:

Values of cosmological parameters
Dark matter and dark energy
Gravity laws
Neutrino mass
Inflation, initial condition of the universe
New physics
...

Weak gravitational lensing



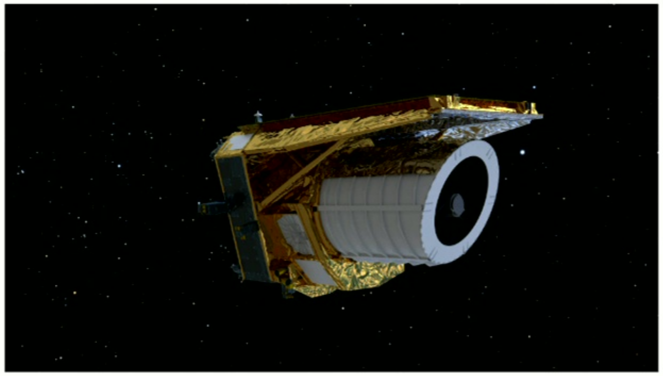
APS/Alan Stonebraker; galaxy images from STScI/AURA, NASA, ESA, and the Hubble Heritage Team

- The gravity of matter (baryonic or dark) warps the surrounding space-time and causes distortions in the perceived shapes of the background galaxies
- Reconstruct the matter distribution in our universe from coherent patterns of galaxy shapes
- Many methods in this talk can be applied to other LSS probes as well

Numerous weak lensing surveys are underway

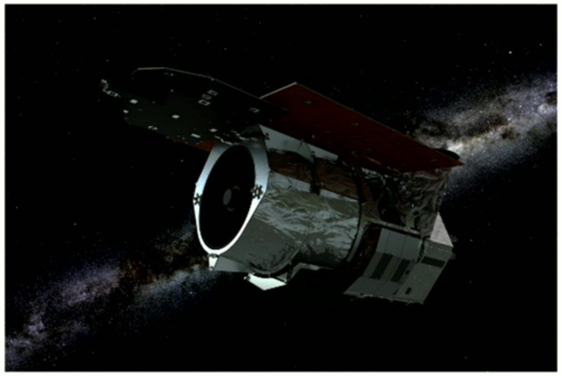
Dark Energy Survey (DES)

Hyper Suprime-Cam (HSC) Subaru Strategic Survey



Euclid telescope

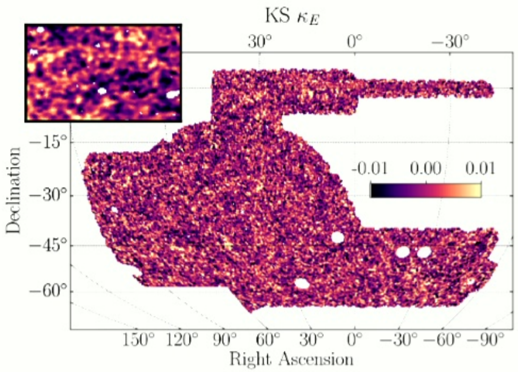
Rubin Observatory LSST



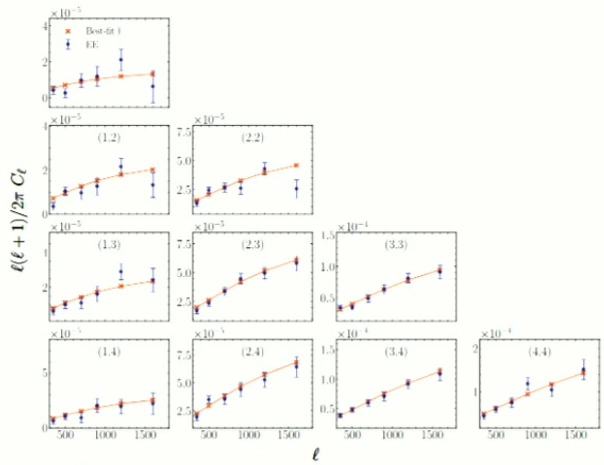
Roman space telescope

Cosmological analysis based on summary statistics

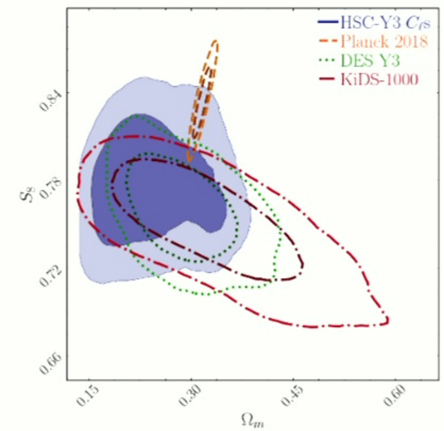
data x



summary statistics S



cosmological parameter y



Compressing data x to summary statistics S

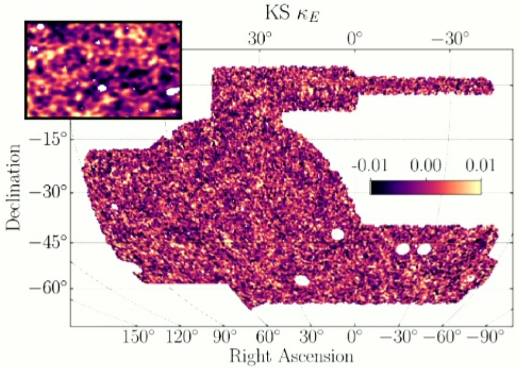
Gaussian likelihood on summary statistics S: p(S|y)

- ▷ Cosmological analysis based on two-point summary statistics: p(S|y) ⟶ p(y|S) = p(S|y)p(y)/p(S)
  - ○ For non-gaussian data, usually leads to **information loss**

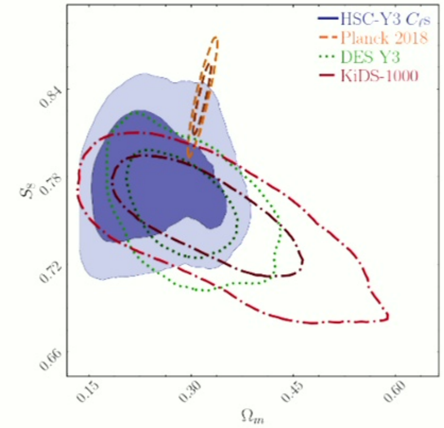
Credit: Jeffrey et al. 2021, Dalal et al. 2023

Field-level cosmological inference

data x



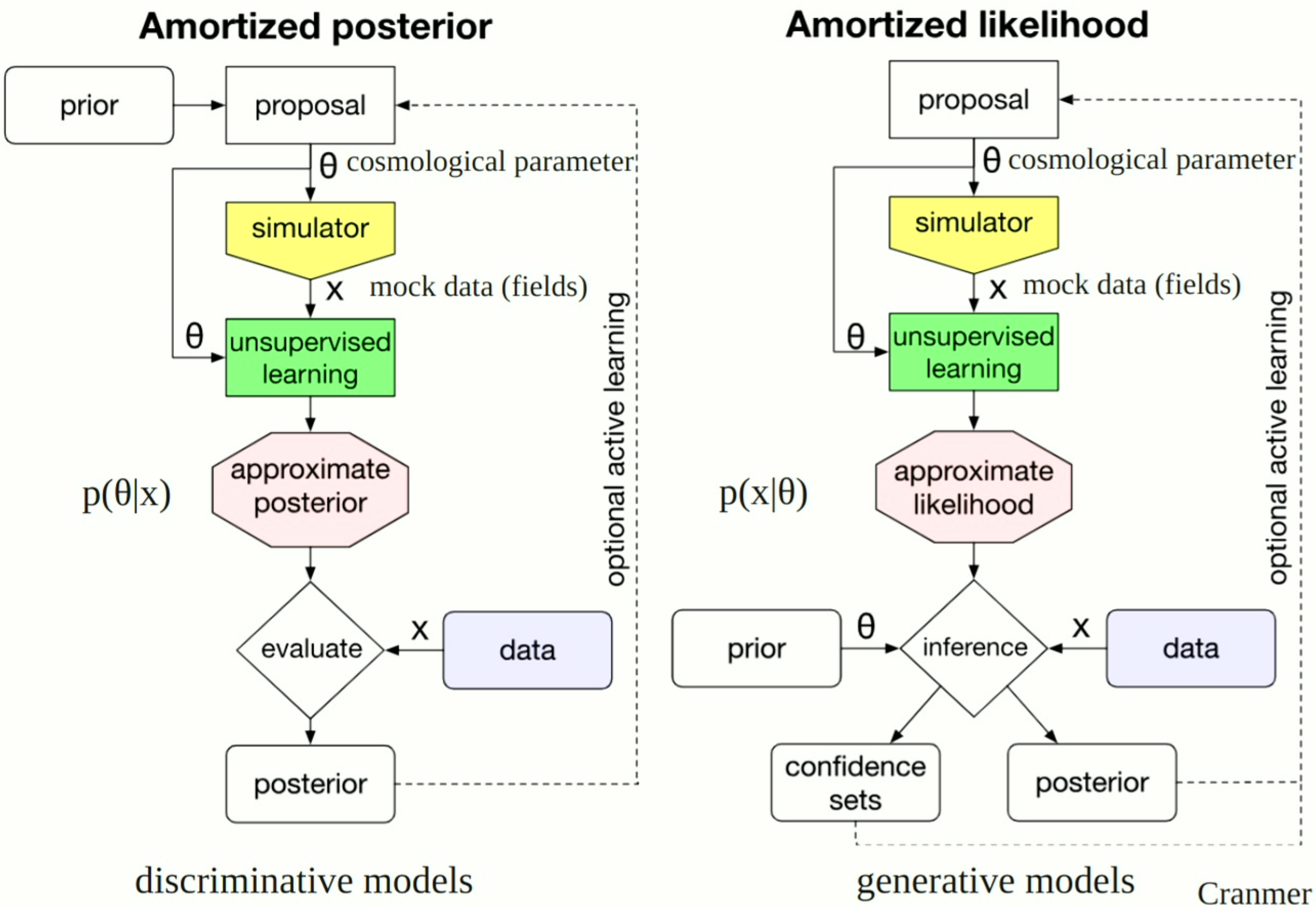
cosmological parameter y



- ▷ Field-level inference
  - ○ Pro: **No information loss** due to data compression.
  - ○ Deep learning allows us to directly extract information at the field level (simulation-based inference)
  - ○ (Reconstruction of initial conditions with forward modeling: $p(x|y) = \int p(x|z,y)p(z|y)dz$)

Credit: Jeffrey et al. 2021, Dalal et al. 2023

Simulation Based Inference (SBI)

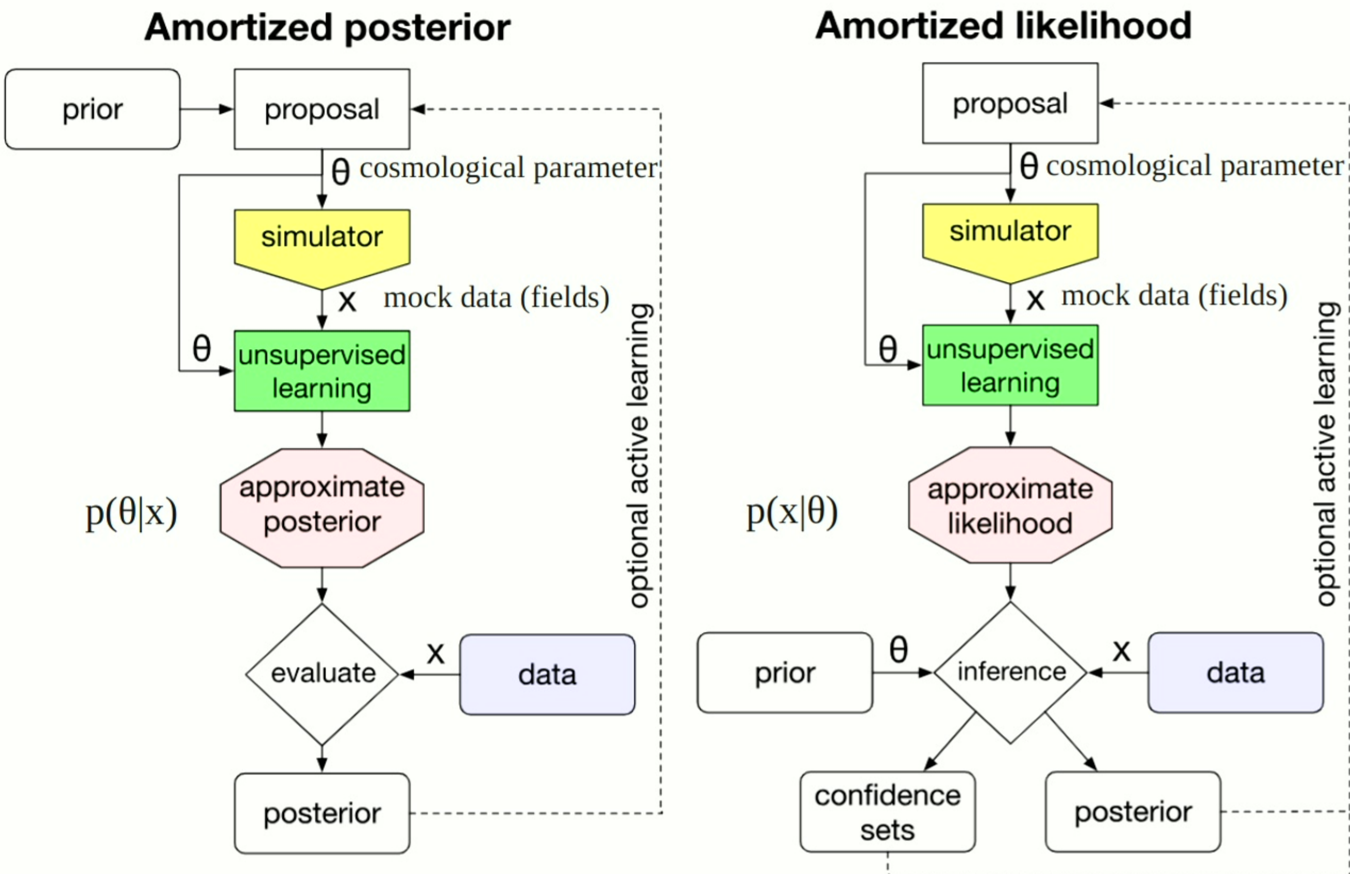


Yellow box: cosmological simulations, solving a system of PDEs coupling gravity, hydrodynamics, and various subgrid physics models such as star formation

Green box: machine learning models (normalizing flows) that take in $\{x,\theta\}_i$ and estimate $p(x|\theta)$ or $p(\theta|x)$.

Cranmer et al. 2020

Simulation Based Inference (SBI)



discriminative models

generative models

Cranmer et al. 2020

Yellow box: cosmological simulations, solving a system of PDEs coupling gravity, hydrodynamics, and various subgrid physics models such as star formation

Green box: machine learning models (normalizing flows) that take in $\{x,\theta\}_i$ and estimate $p(x|\theta)$ or $p(\theta|x)$.

Potential issues of SBI:

1. The simulations may not be accurate (distribution shift)
2. The ML model is a black box and lacks interpretability

9

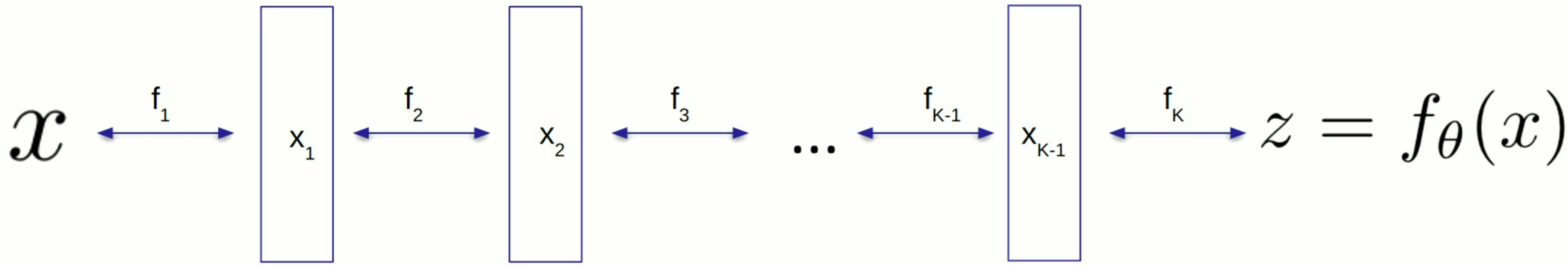
Outline

- ▷ Background: Cosmological data analysis and simulation-based inference
- ▷ Normalizing Flows (NFs) and what can they do for Physics
- ▷ Weak lensing analysis with generative NFs
  - **Optimal**: field-level information extraction
  - **Reliable**: anomaly detection of systematic effects
  - **Interpretable**: improve explainability with generated samples
  - Ongoing and future works
- ▷ Applications of NFs beyond cosmology
  - Anomaly detection of new physics in high energy physics
  - Lattice field variational inference
  - Speed up Bayesian sampling algorithm

Outline

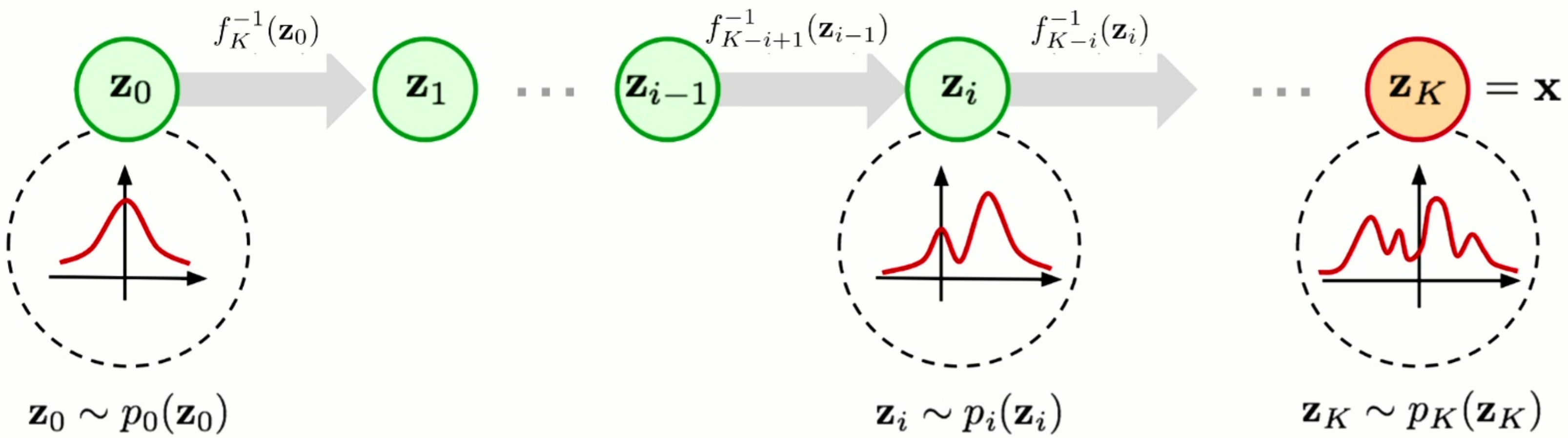
- ▷ Background: Cosmological data analysis and simulation-based inference
- ▷ Normalizing Flows (NFs) and what can they do for Physics
- ▷ Weak lensing analysis with generative NFs
  - **Optimal**: field-level information extraction
  - **Reliable**: anomaly detection of systematic effects
  - **Interpretable**: improve explainability with generated samples
  - Ongoing and future works
- ▷ Applications of NFs beyond cosmology
  - Anomaly detection of new physics in high energy physics
  - Lattice field variational inference
  - Speed up Bayesian sampling algorithm

Normalizing Flows



- ▷ Bijective mapping f between data x and latent variable z $(z = f(x),\ z \sim \pi(z))$
  - ○ **Evaluate density**: $p(x) = \pi(f(x))\ |\det(df/dx)|$
  - ○ **Sample**: $x = f^{-1}(z)\ \ (z \sim \pi(z))$

Normalizing Flows

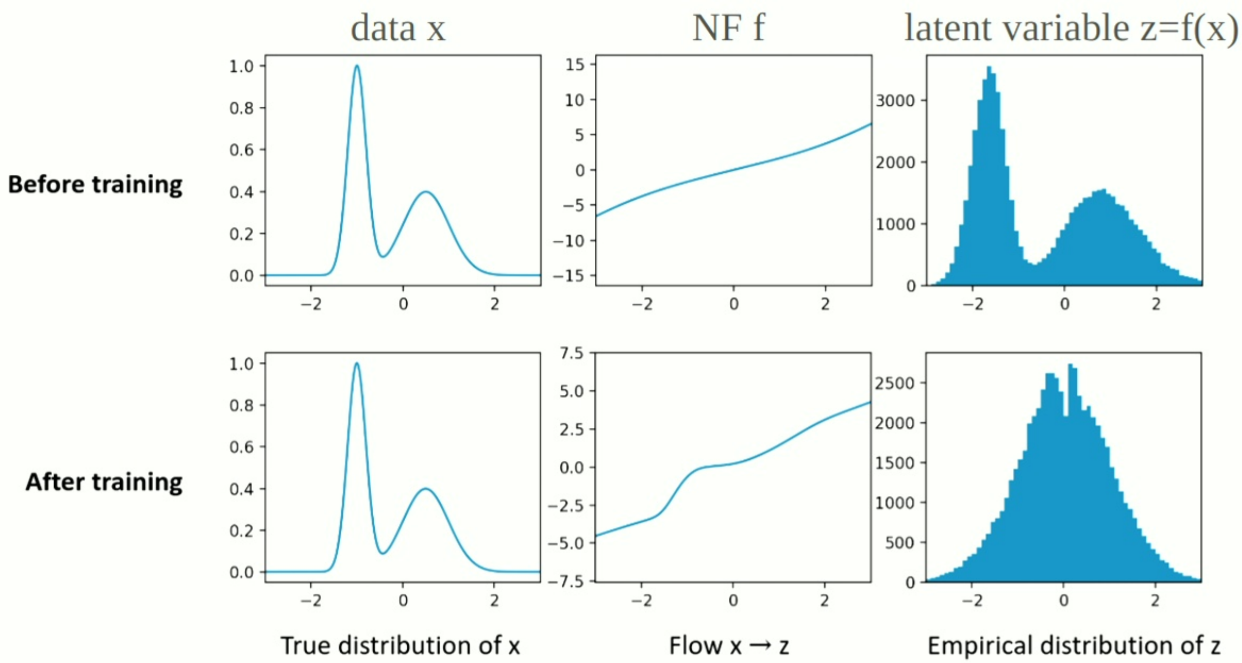


Credit: https://lilianweng.github.io/lil-log/2018/10/13/flow-based-deep-generative-models.html

- Bijective mapping f between data x and latent variable z (z = f(x), z ~ π(z))
  - **Evaluate density**: p(x) = π(f(x)) |det(df/dx)|
  - **Sample**: x = f^{-1}(z) (z ~ π(z))

Normalizing Flows

▷ 1D example



- Training objective: **<log p(x)>**

  p(x) = π(f(x)) |det(df/dx)|

  MLE estimation of parameters

  Equivalent to minimizing KL-divergence

- **Evaluate density**: p(x) = π(f(x)) |det(df/dx)|
- **Sample**: $x = f^{-1}(z)$ $(z \sim \pi(z))$

Credit: https://sites.google.com/view/berkeley-cs294-158-sp20/home

What can Normalizing Flows do for Physics?

Normalizing flows provide a powerful framework for high-dimensional density estimation (likelihood) and sampling

- Extract physical information (simulation-based inference)
- Fast sample generation
- Anomaly detection
  - Detect systematic effect (distribution shift)
  - Search for new physics
- Approximate intractable distributions
  - Variational Inference
  - Fast Bayesian sampling

Outline

- ▷ Background: Cosmological data analysis and simulation-based inference
- ▷ Normalizing Flows (NFs) and what can they do for Physics
- ▷ Weak lensing analysis with generative NFs
  - ○ **Optimal**: field-level information extraction
  - ○ **Reliable**: anomaly detection of systematic effects
  - ○ **Interpretable**: improve explainability with generated samples
  - ○ Ongoing and future works
- ▷ Applications of NFs beyond cosmology
  - ○ Anomaly detection of new physics in high energy physics
  - ○ Lattice field variational inference
  - ○ Speed up Bayesian sampling algorithm

Outline

- ▷ Background: Cosmological data analysis and simulation-based inference

- ▷ Normalizing Flows (NFs) and what can they do for Physics

- ▷ Weak lensing analysis with generative NFs
  - **Optimal**: field-level information extraction
  - **Reliable**: anomaly detection of systematic effects
  - **Interpretable**: improve explainability with generated samples
  - Ongoing and future works

- ▷ Applications of NFs beyond cosmology
  - Anomaly detection of new physics in high energy physics
  - Lattice field variational inference
  - Speed up Bayesian sampling algorithm

From Gaussian Random Fields, Lognormal Fields to TRENF

	NF transformation	Probability distribution function
Gaussian Random Fields	$\delta_{\mathrm{GRF}}(k) = \sqrt{P(k)}\delta_z(k)$	$p(\delta_{\mathrm{GRF}}) = \frac{1}{\sqrt{(2\pi)^n \prod_i P(k_i)}} \exp(-\sum_i \frac{\vert\delta_{\mathrm{GRF}}\vert^2}{2P(k_i)})$

18

From Gaussian Random Fields, Lognormal Fields to TRENF

	NF transformation	Probability distribution function
Gaussian Random Fields	$\delta_{\mathrm{GRF}}(k) = \sqrt{P(k)}\delta_z(k)$	$p(\delta_{\mathrm{GRF}}) = \frac{1}{\sqrt{(2\pi)^n \prod_i P(k_i)}} \exp(-\sum_i \frac{\lvert\delta_{\mathrm{GRF}}\rvert^2}{2P(k_i)})$
Lognormal Fields	$\delta_{\mathrm{LN}} = \exp(\delta_{\mathrm{GRF}}) - C$	$p(\delta_{\mathrm{LN}}) = p(\delta_{\mathrm{GRF}}) \lvert\frac{\partial \delta_{\mathrm{GRF}}}{\partial \delta_{\mathrm{LN}}}\rvert$

From Gaussian Random Fields, Lognormal Fields to TRENF

	NF transformation	Probability distribution function
Gaussian Random Fields	$\delta_{\mathrm{GRF}}(k) = \sqrt{P(k)}\delta_z(k)$	$p(\delta_{\mathrm{GRF}}) = \frac{1}{\sqrt{(2\pi)^n \prod_i P(k_i)}} \exp(-\sum_i \frac{\lvert\delta_{\mathrm{GRF}}\rvert^2}{2P(k_i)})$
Lognormal Fields	$\delta_{\mathrm{LN}} = \exp(\delta_{\mathrm{GRF}}) - C$	$p(\delta_{\mathrm{LN}}) = p(\delta_{\mathrm{GRF}}) \lvert \frac{\partial \delta_{\mathrm{GRF}}}{\partial \delta_{\mathrm{LN}}} \rvert$
Generalization	$f_i(x) = \Psi_i(\mathcal{F}^{-1} T_i(k) \mathcal{F} x)$	$p(f_i(x)) = p(x) \prod_r \lvert \Psi_i'^{-1}(x(r)) \rvert \prod_k T_i^{-1}(k)$

any monotonic, differentiable functions

1D function. Parametrize with spline functions

20

From Gaussian Random Fields, Lognormal Fields to TRENF

	NF transformation	Probability distribution function
Gaussian Random Fields	$\delta_{\mathrm{GRF}}(k) = \sqrt{P(k)}\delta_z(k)$	$p(\delta_{\mathrm{GRF}}) = \frac{1}{\sqrt{(2\pi)^n \prod_i P(k_i)}} \exp(-\sum_i \frac{\lvert\delta_{\mathrm{GRF}}\rvert^2}{2P(k_i)})$
Lognormal Fields	$\delta_{\mathrm{LN}} = \exp(\delta_{\mathrm{GRF}}) - C$	$p(\delta_{\mathrm{LN}}) = p(\delta_{\mathrm{GRF}}) \lvert \frac{\partial \delta_{\mathrm{GRF}}}{\partial \delta_{\mathrm{LN}}} \rvert$
Generalization	$f_i(x) = \Psi_i(\mathcal{F}^{-1} T_i(k) \mathcal{F} x)$	$p(f_i(x)) = p(x) \prod_r \lvert \Psi_i'^{-1}(x(r)) \rvert \prod_k T_i^{-1}(k)$

any monotonic, differentiable functions

1D function. Parametrize with spline functions

Translation and rotation symmetry
CNN-like architecture

From Gaussian Random Fields, Lognormal Fields to TRENF

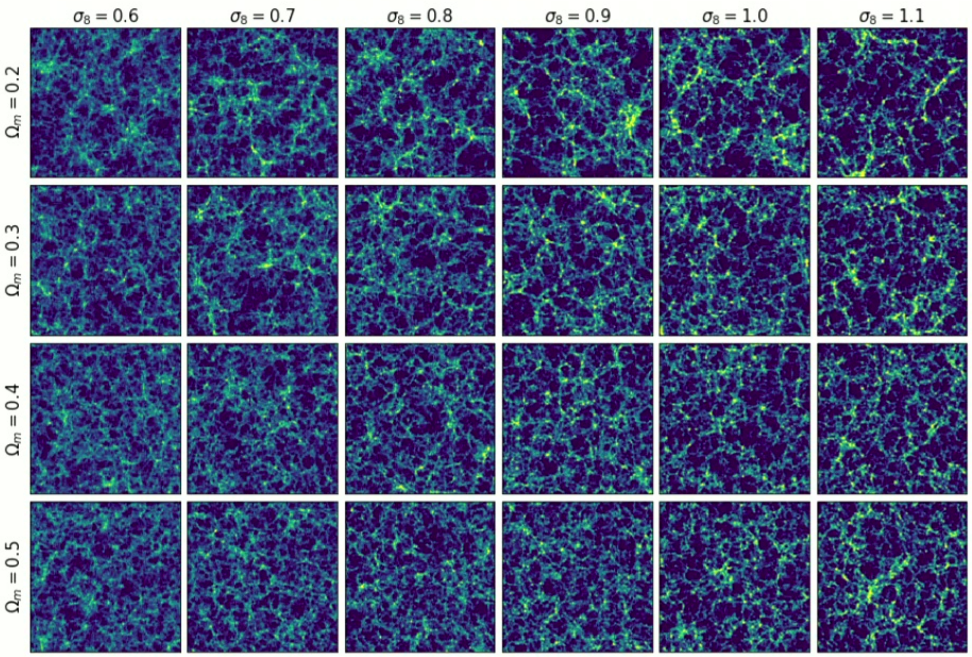
	NF transformation	Probability distribution function
Gaussian Random Fields	$\delta_{\mathrm{GRF}}(k) = \sqrt{P(k)}\delta_z(k)$	$p(\delta_{\mathrm{GRF}}) = \frac{1}{\sqrt{(2\pi)^n \prod_i P(k_i)}} \exp(-\sum_i \frac{\lvert\delta_{\mathrm{GRF}}\rvert^2}{2P(k_i)})$
Lognormal Fields	$\delta_{\mathrm{LN}} = \exp(\delta_{\mathrm{GRF}}) - C$	$p(\delta_{\mathrm{LN}}) = p(\delta_{\mathrm{GRF}})\lvert\frac{\partial \delta_{\mathrm{GRF}}}{\partial \delta_{\mathrm{LN}}}\rvert$
Generalization	$f_i(x) = \Psi_i(\mathcal{F}^{-1}T_i(k)\mathcal{F}x)$	$p(f_i(x)) = p(x)\prod_r \lvert\Psi_i'^{-1}(x(r))\rvert \prod_k T_i^{-1}(k)$
	any monotonic, differentiable functions; **1D** function. Parametrize with spline functions	Translation and rotation symmetry CNN-like architecture
Translation and Rotation Equivariant Normalizing Flow (TRENF)	$f = f_1 \circ f_2 \circ \cdots \circ f_n$	$p(f(x)) = p(x)\prod_i \lvert f_i'(x)\rvert^{-1}$

Dai, B. and Seljak, U., 2022. Translation and rotation equivariant normalizing flow (TRENF) for optimal cosmological analysis. *Monthly Notices of the Royal Astronomical Society*, *516*(2), pp.2363-2373.

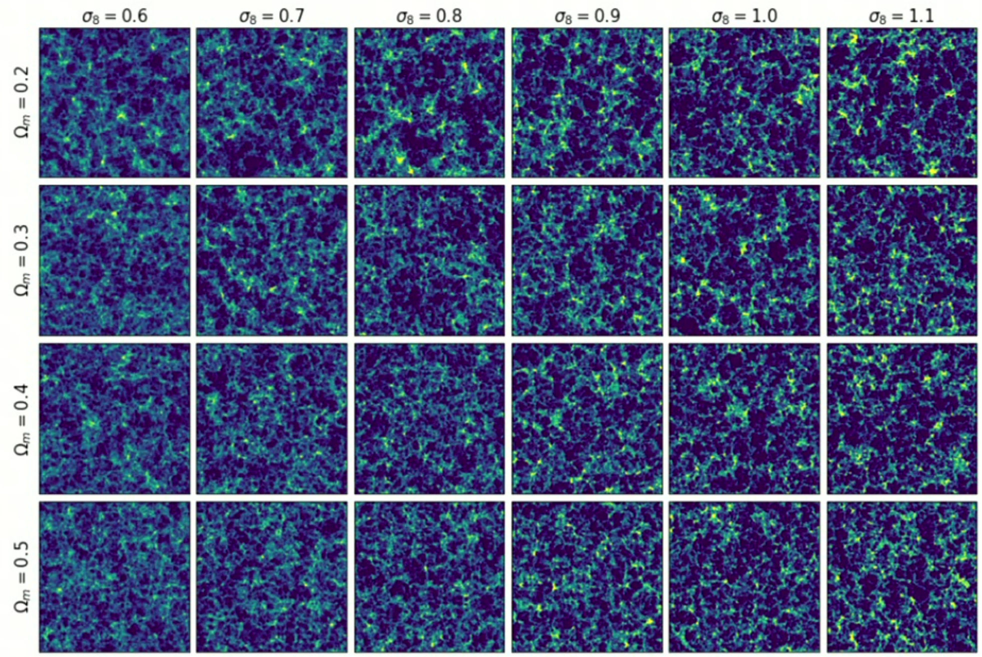
Results -- Samples

data x ⟷ NF f ⟷ Gaussian z

- data:



- TRENF samples:



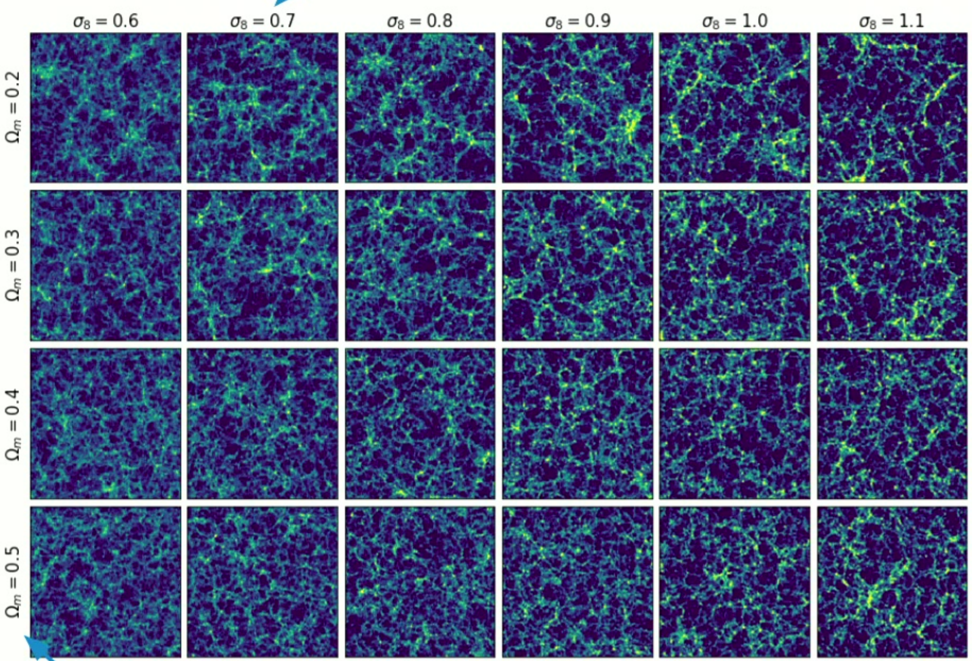
→ more clustering

Results -- Samples

data x ⟷ NF f ⟷ Gaussian z

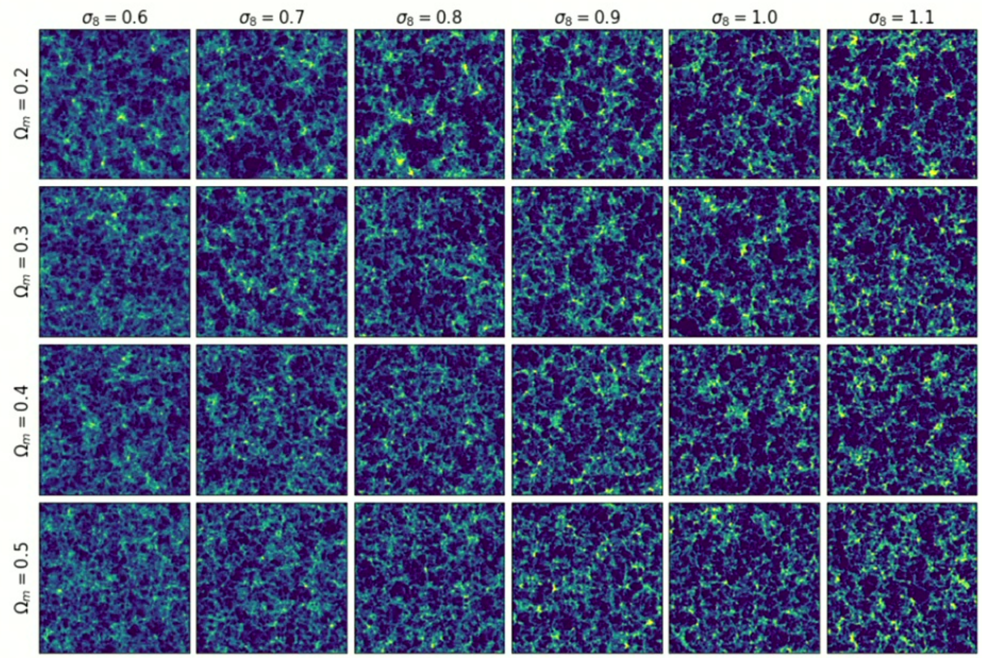
- data:

amplitude of matter fluctuations



mean matter density of the present-day universe

- TRENF samples:

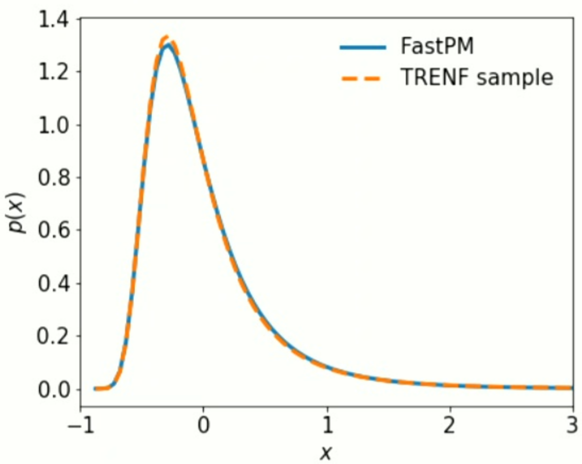


more clustering

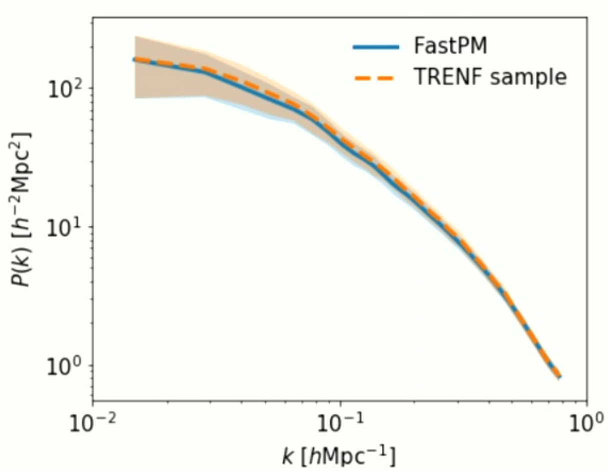
Results -- Summary Statistics of Samples

data x ⟷ (NF f) Gaussian z

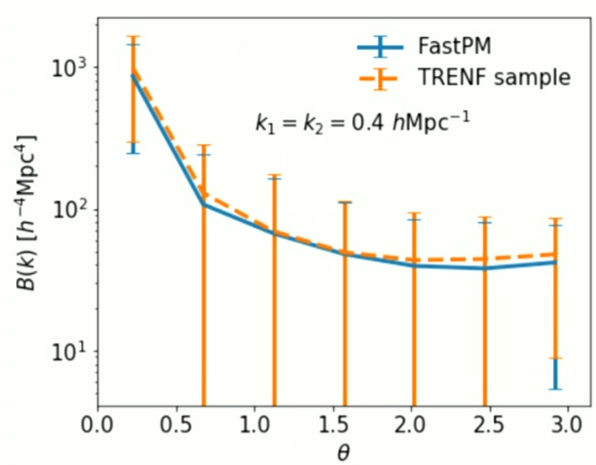
- 1-D probability distribution function



- power spectrum:



- bispectrum:

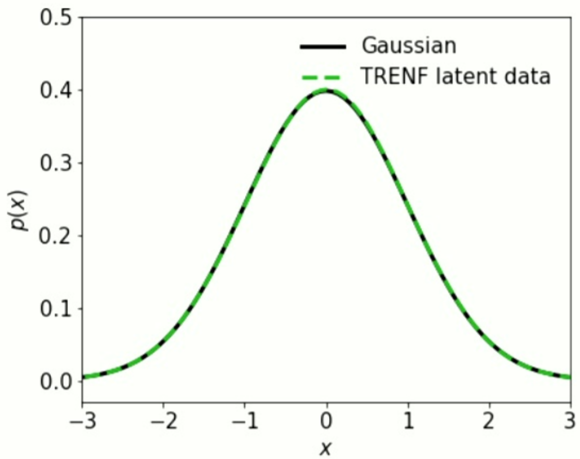


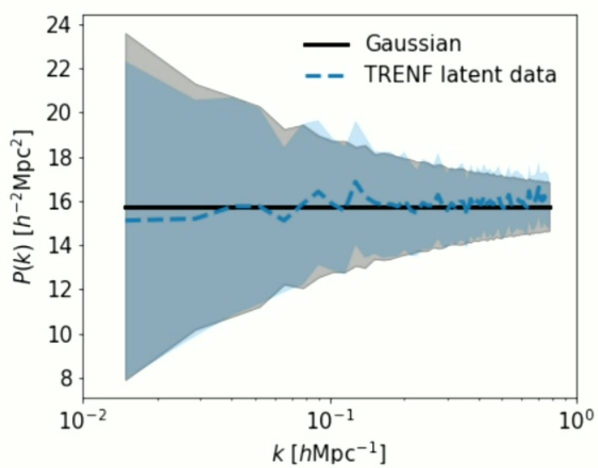
Measured over 10000 samples

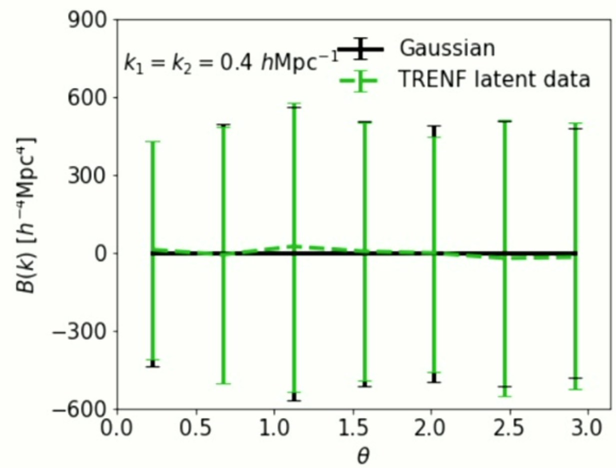
Results -- Summary Statistics of Latent data

data x ⟷ (NF f) Gaussian z

- 1-D probability distribution function
- power spectrum:
- bispectrum:





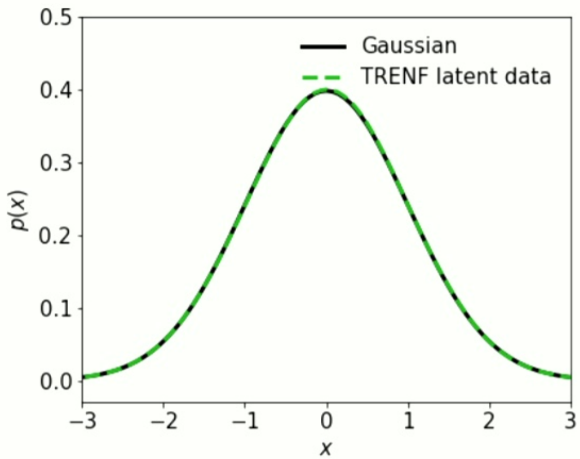


Measured over 10000 samples

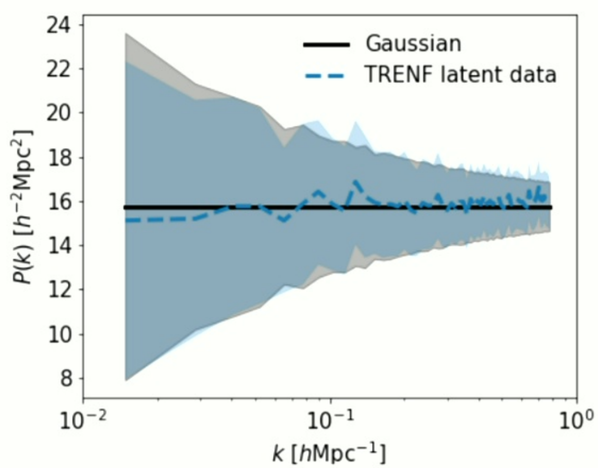
Results -- Summary Statistics of Latent data

data x ⟷ NF f ⟷ Gaussian z

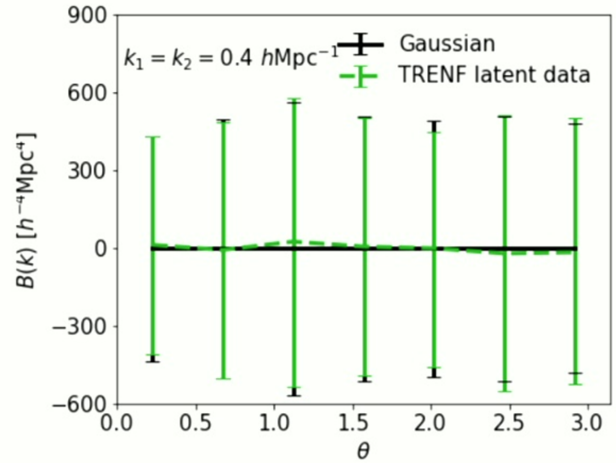
- 1-D probability distribution function



- power spectrum:



- bispectrum:

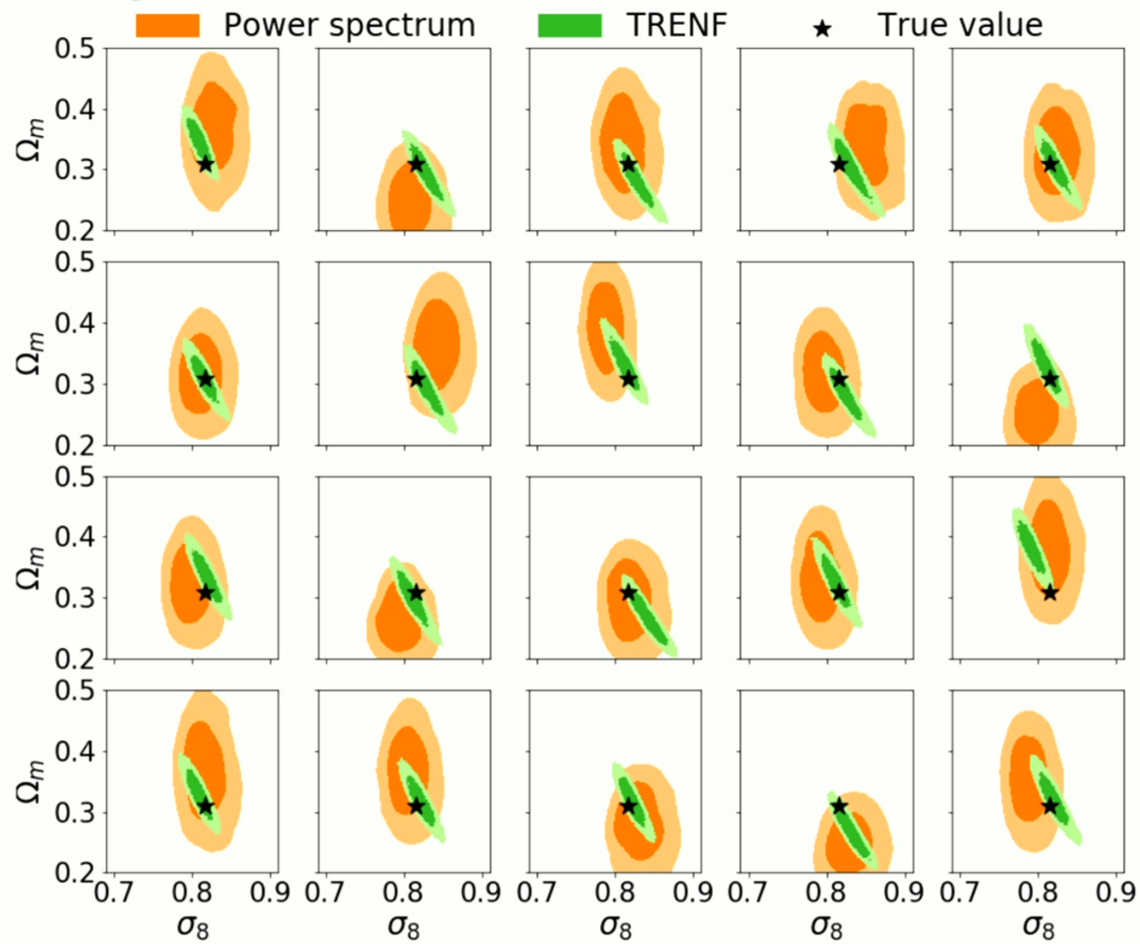


NF transformation removes the non-Gaussianity from gravitational effect.
Possible future direction: search for primordial non-Gaussianity in NF latent space

Measured over 10000 samples

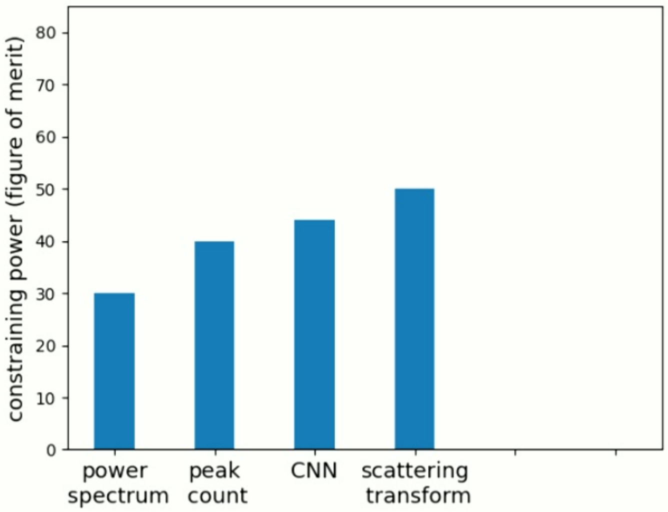
Results -- Posterior Analysis

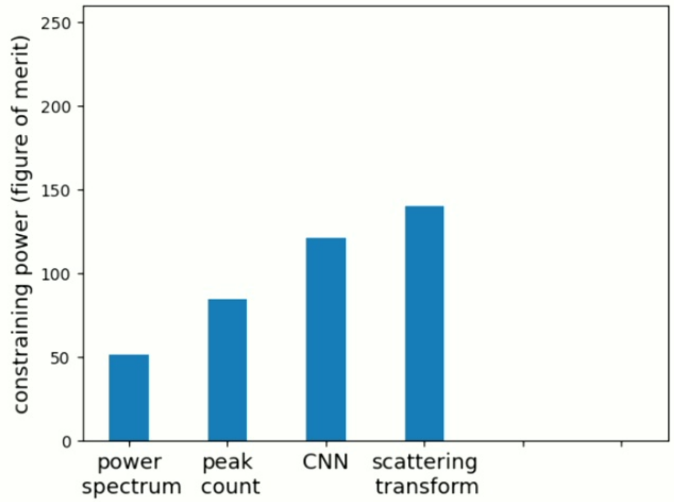
- Field-level likelihood function:
  $p(x|y) = \pi(f_y(x))\,|\det(df_y/dx)|$
- Full posterior distribution via $p(y|x)=p(x|y)p(y)/p(x)$
- Figure of merit (inverse of the area of the 68% confidence region):
  - **power spectrum: ~ 176**
  - **TRENF: ~ 995**
- The posterior is reliable: On 100 test data, 65 cases the true cosmology is within the 68% contour, and 95 cases the true cosmology is within the 95% region.

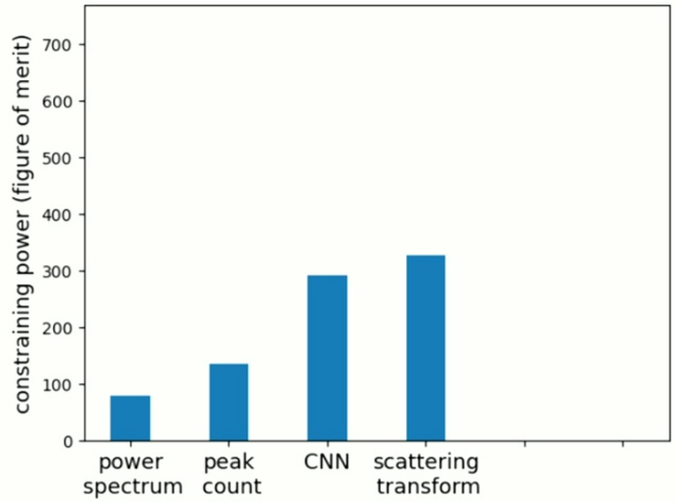


Performance on mock weak lensing maps

- Current surveys (n_g=10 arcmin^{-2})
- Upcoming surveys (n_g=30 arcmin^{-2})
- Optimistic scenario for a future-generation space-based survey (n_g=100 arcmin^{-2})







Ribli et al. 2019

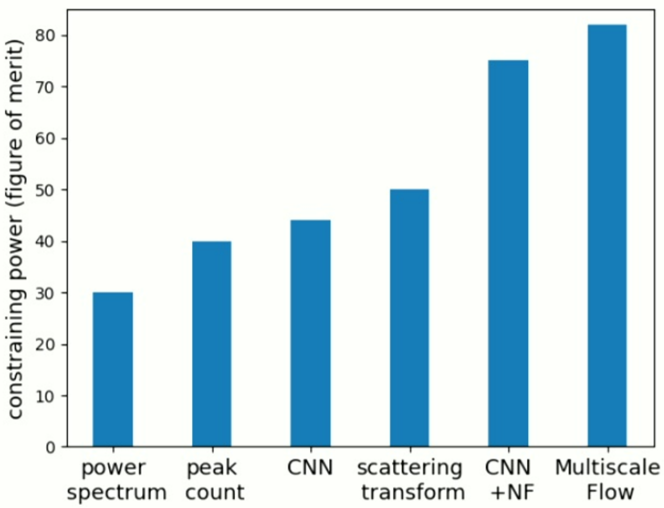
Cheng et al. 2021
Allys et al. 2021

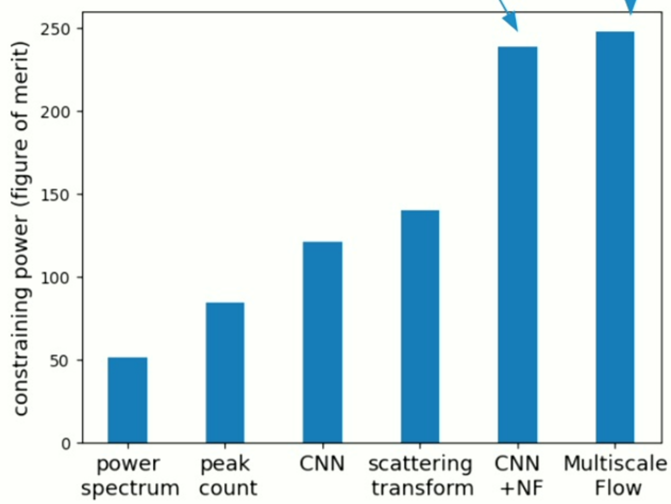
Performance on mock weak lensing maps

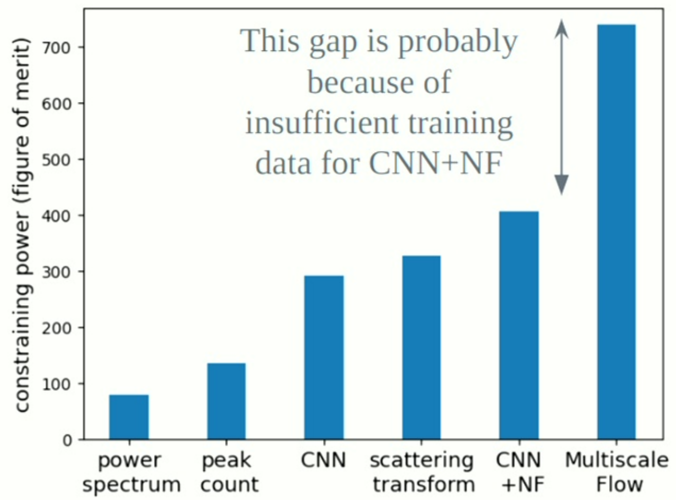
- Current surveys (n_g=10 arcmin^{-2})
- Upcoming surveys (n_g=30 arcmin^{-2})
- Optimistic scenario for a future-generation space-based survey (n_g=100 arcmin^{-2})

Discriminative model (learns posterior)

Generative model (learns likelihood)







Ribli et al. 2019

Cheng et al. 2021
Allys et al. 2021

Sharma, **Dai** & Seljak, in prep.

Dai & Seljak 2024

For current and upcoming surveys, generative and discriminative models lead to similar performance, potentially suggesting both may have extracted the full information content from the data

Outline

- ▷ Background: Cosmological data analysis and simulation-based inference

- ▷ Normalizing Flows (NFs) and what can they do for Physics

- ▷ Weak lensing analysis with generative NFs
  - **Optimal**: field-level information extraction
  - **Reliable**: anomaly detection of systematic effects
  - **Interpretable**: improve explainability with generated samples
  - Ongoing and future works

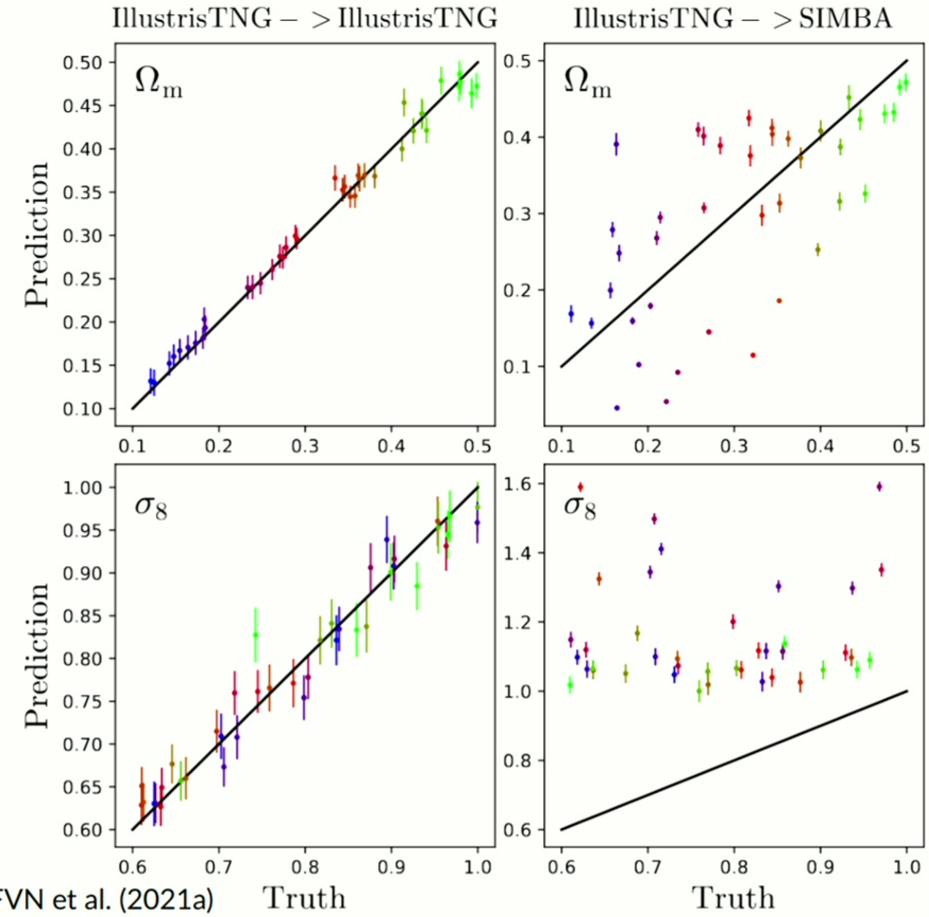
- ▷ Applications of NFs beyond cosmology
  - Anomaly detection of new physics in high energy physics
  - Lattice field variational inference
  - Speed up Bayesian sampling algorithm

Outline

- ▷ Background: Cosmological data analysis and simulation-based inference
- ▷ Normalizing Flows (NFs) and what can they do for Physics
- ▷ Weak lensing analysis with generative NFs
  - **Optimal**: field-level information extraction
  - **Reliable**: anomaly detection of systematic effects
  - **Interpretable**: improve explainability with generated samples
  - Ongoing and future works
- ▷ Applications of NFs beyond cosmology
  - Anomaly detection of new physics in high energy physics
  - Lattice field variational inference
  - Speed up Bayesian sampling algorithm

Robust cosmological data analysis

- ▷ How do we know we can trust our analysis?
- ▷ We need to assume that the simulations we trained on overlap with reality, but is this guaranteed?
- ▷ SBI models (CNN) trained on gas temperature maps of IllustrisTNG do not work on SIMBA.
- ▷ Marginalize over different baryon parameters, baryon models, N-body codes, etc.
- ▷ Tests to verify the reliability of the analysis



FVN et al. (2021a)

The analysis should pass various null tests / consistency tests

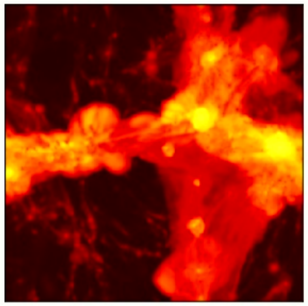
- Goodness-of-fit test with χ^2
  - The log-likelihood from generative models is a natural extension of the χ^2 statistic!

- Consistency test:
  - Robustness to different modeling choices
  - Internal consistency with different scales, redshift bins, different patches of the sky, etc.

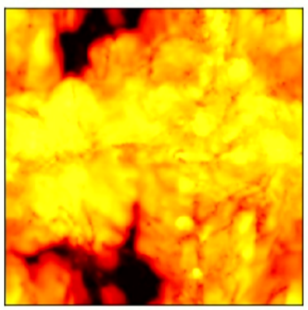
- Null tests with different systematic effects
  - e.g., B mode tests in weak lensing

Test 1: Goodness-of-fit test / Out-of-distribution detection

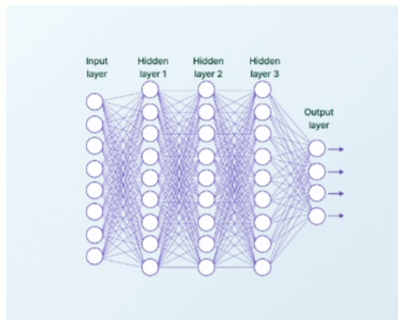
Training simulations

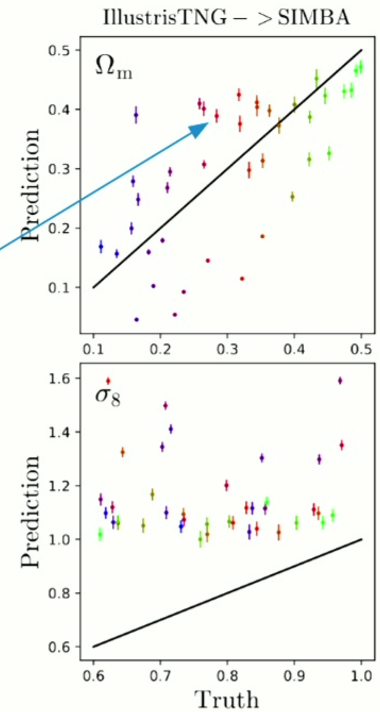


Test data / observation



Discriminative models

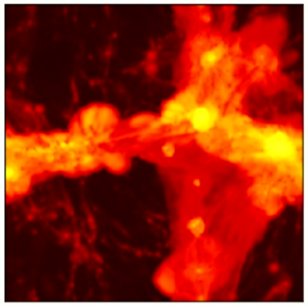




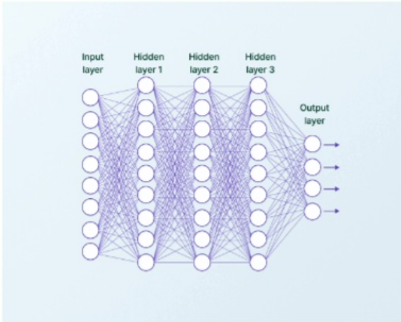
Biased parameter constraints due to distribution shifts, and we don't know it!

Test 1: Goodness-of-fit test / Out-of-distribution detection

Training simulations

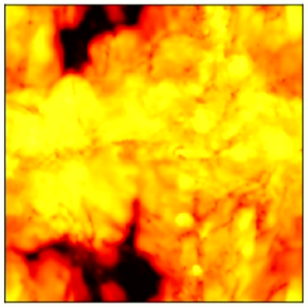


Generative models



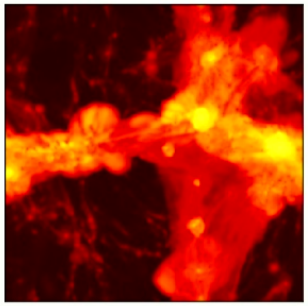
likelihood p(x|y)

Test data / observation

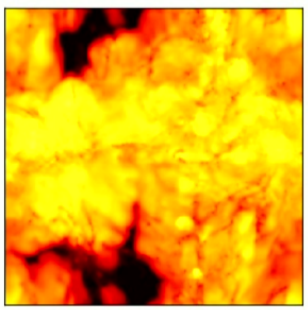


Test 1: Goodness-of-fit test / Out-of-distribution detection

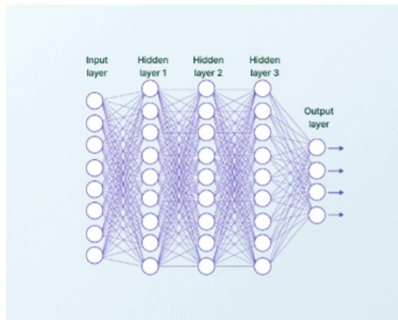
Training simulations



Test data / observation

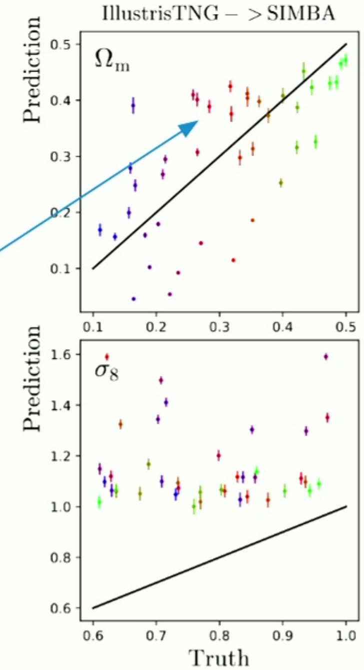


Generative models



likelihood p(x|y)

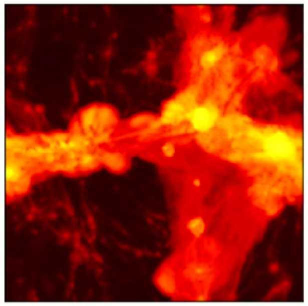
MCMC



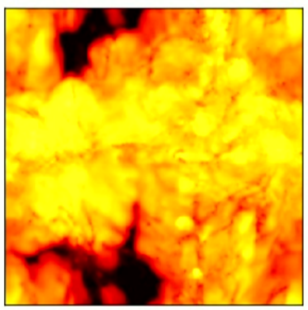
37

Test 1: Goodness-of-fit test / Out-of-distribution detection

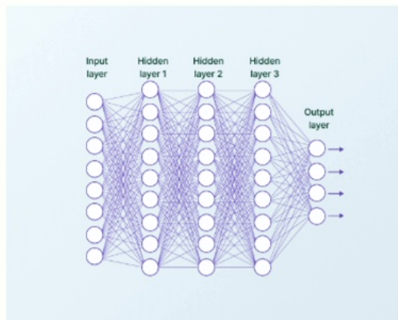
Training simulations



Test data / observation

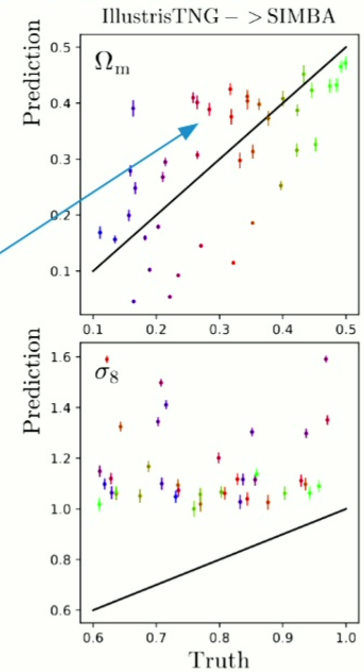


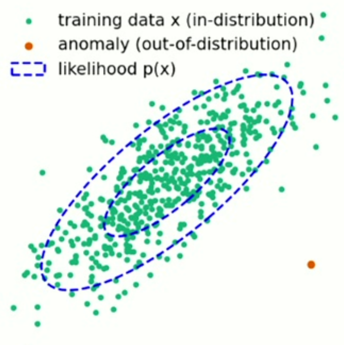
Generative models

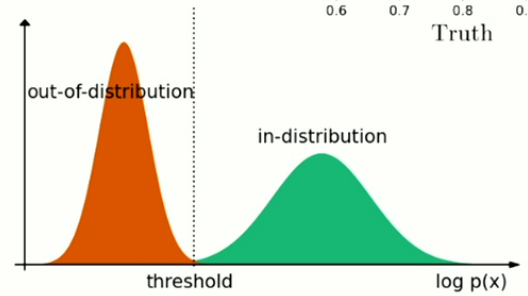


likelihood p(x|y)

MCMC

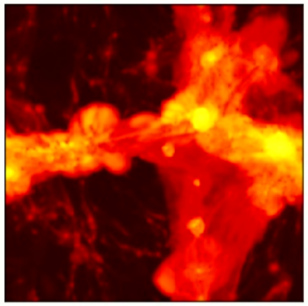




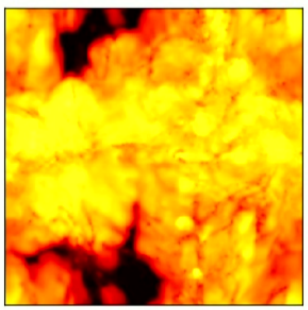


Test 1: Goodness-of-fit test / Out-of-distribution detection

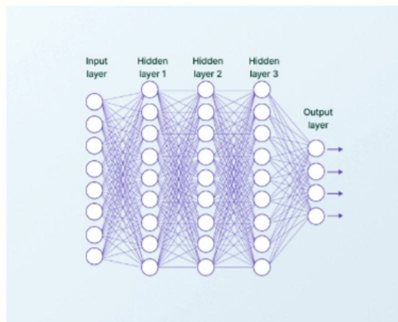
Training simulations



Test data / observation

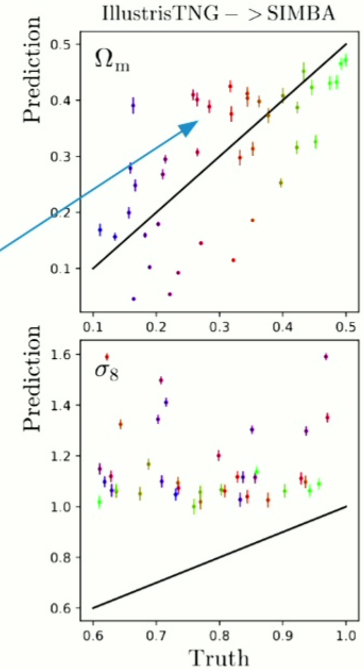


Generative models



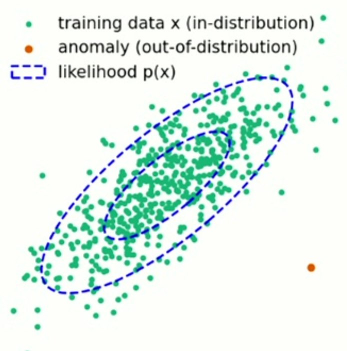
likelihood p(x|y)

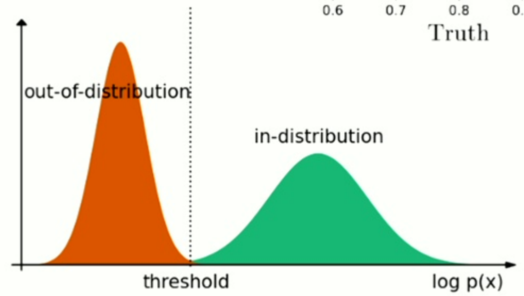
MCMC



The test data / observation doesn't look like training data, so we shouldn't trust our analysis!

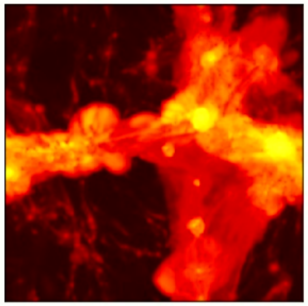
More conservative scale cuts, or improve the modeling of training simulations.



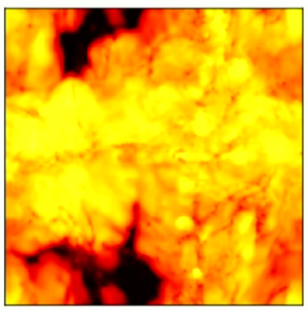


Test 1: Goodness-of-fit test / Out-of-distribution detection

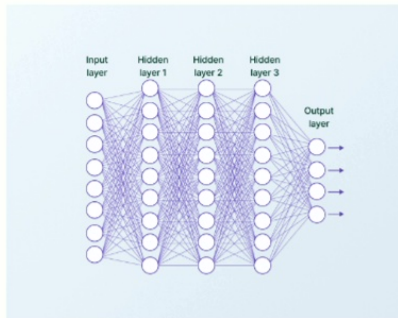
Training simulations



Test data / observation



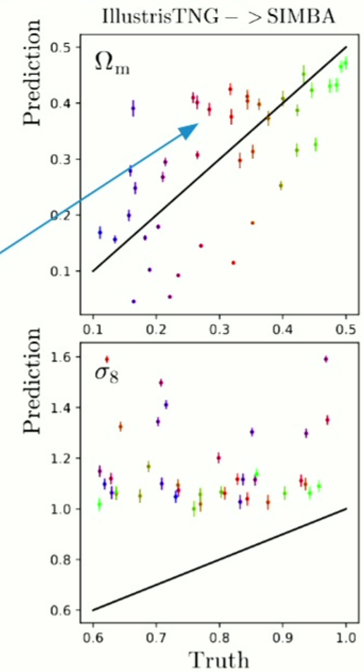
Generative models



likelihood p(x|y)

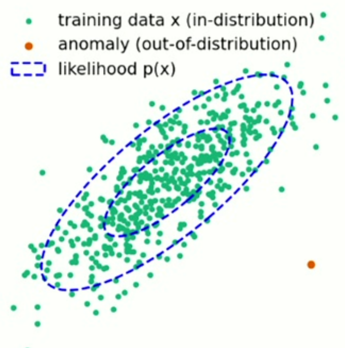
MCMC

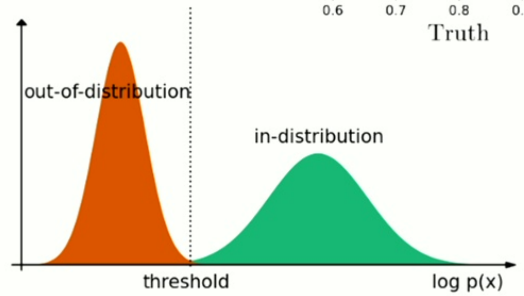
Generative NF models enable goodness-of-fit test to improve the robustness of analysis.



The test data / observation doesn't look like training data, so we shouldn't trust our analysis!

More conservative scale cuts, or improve the modeling of training simulations.





Test 2: Multiscale consistency test with Multiscale Flow

- Motivation: Multiscale analysis for robust constraints
  - Different scales are governed by different physics / systematics: the numerical / astrophysical effects normally happens on small scales, and PSF may influence very large scales
  - Separate and compare the information (likelihood) of different scales, and identify the part of the data that is contaminated by systematics

Dai, B. and Seljak, U., 2023. Multiscale Flow for Robust and Optimal Cosmological Analysis. *arXiv preprint arXiv:2306.04689*.

Test 2: Multiscale consistency test with Multiscale Flow

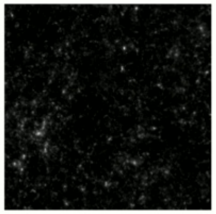
- Motivation: Multiscale analysis for robust constraints
  - Different scales are governed by different physics / systematics: the numerical / astrophysical effects normally happens on small scales, and PSF may influence very large scales
  - Separate and compare the information (likelihood) of different scales, and identify the part of the data that is contaminated by systematics

- Wavelet decomposition: recursively apply low-pass filters (scaling functions) and high-pass filters (wavelet functions) to the data. In each iteration, the data x_n with resolution 2^n is decomposed into a low-resolution approximation x_{n-1}, and detail coefficients of the remaining signal $x_{n-1,extra}$

Dai, B. and Seljak, U., 2023. Multiscale Flow for Robust and Optimal Cosmological Analysis. *arXiv preprint arXiv:2306.04689*.

Multiscale flow

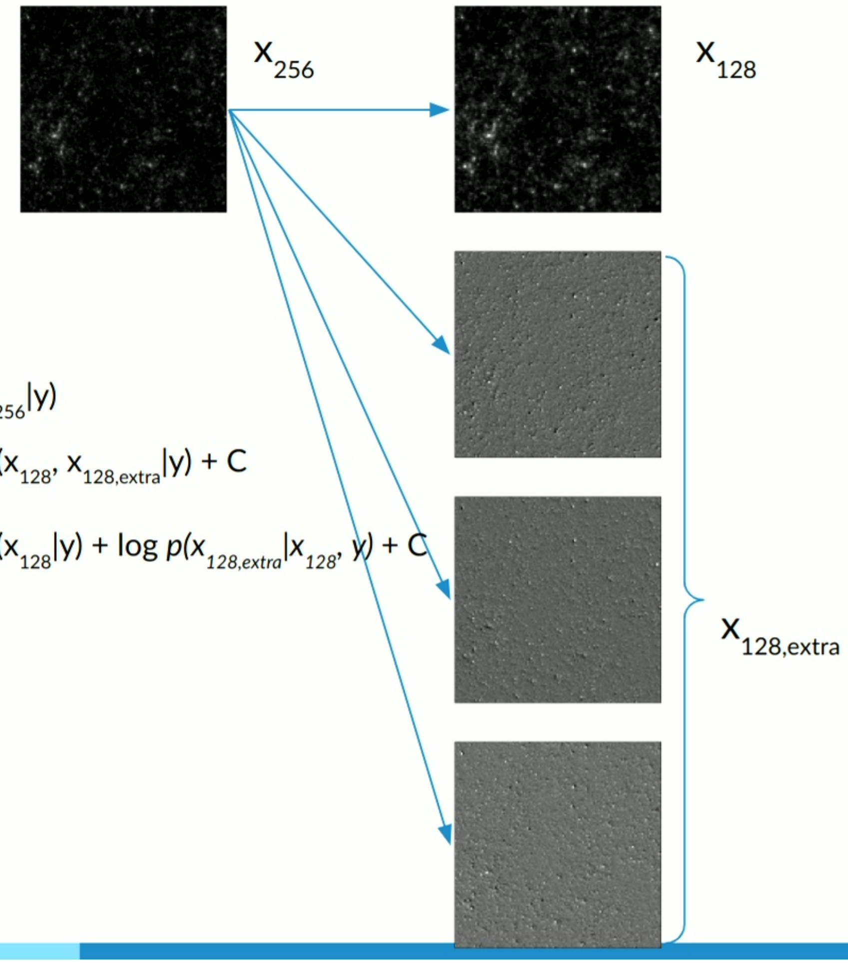
▷ Consider a cosmological field with 256^2 resolution:



$\log p(x_{256}|y)$

Multiscale flow

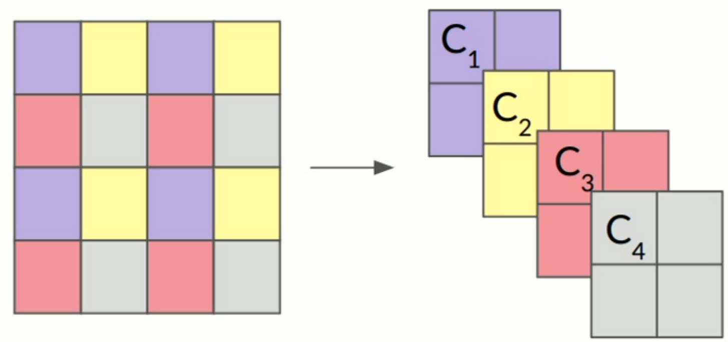
- Consider a cosmological field with 256^2 resolution:



$\log p(x_{256}|y)$

$= \log p(x_{128}, x_{128,extra}|y) + C$

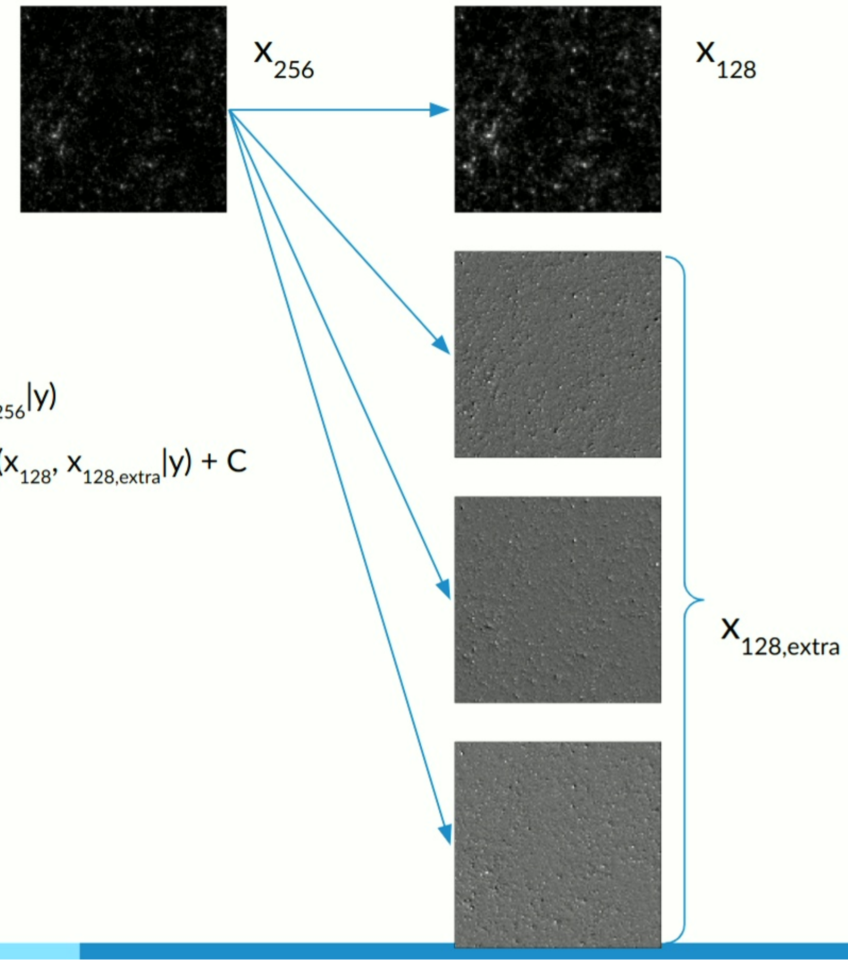
$= \log p(x_{128}|y) + \log p(x_{128,extra}|x_{128}, y) + C$



$$\begin{bmatrix} x_{128} \\ x^1_{128,\mathrm{extra}} \\ x^2_{128,\mathrm{extra}} \\ x^3_{128,\mathrm{extra}} \end{bmatrix} = \begin{bmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{bmatrix} \begin{bmatrix} C_1 \\ C_2 \\ C_3 \\ C_4 \end{bmatrix}$$

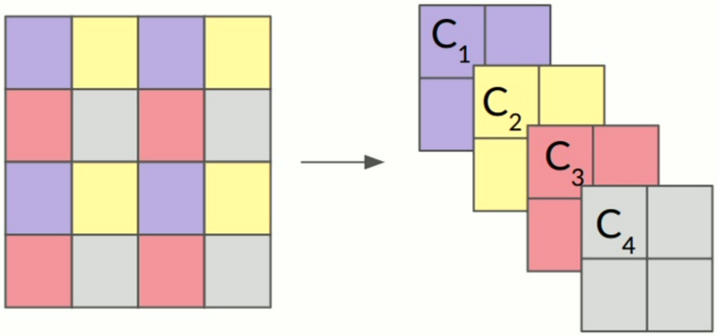
Multiscale flow

- Consider a cosmological field with 256^2 resolution:



$\log p(x_{256}|y)$

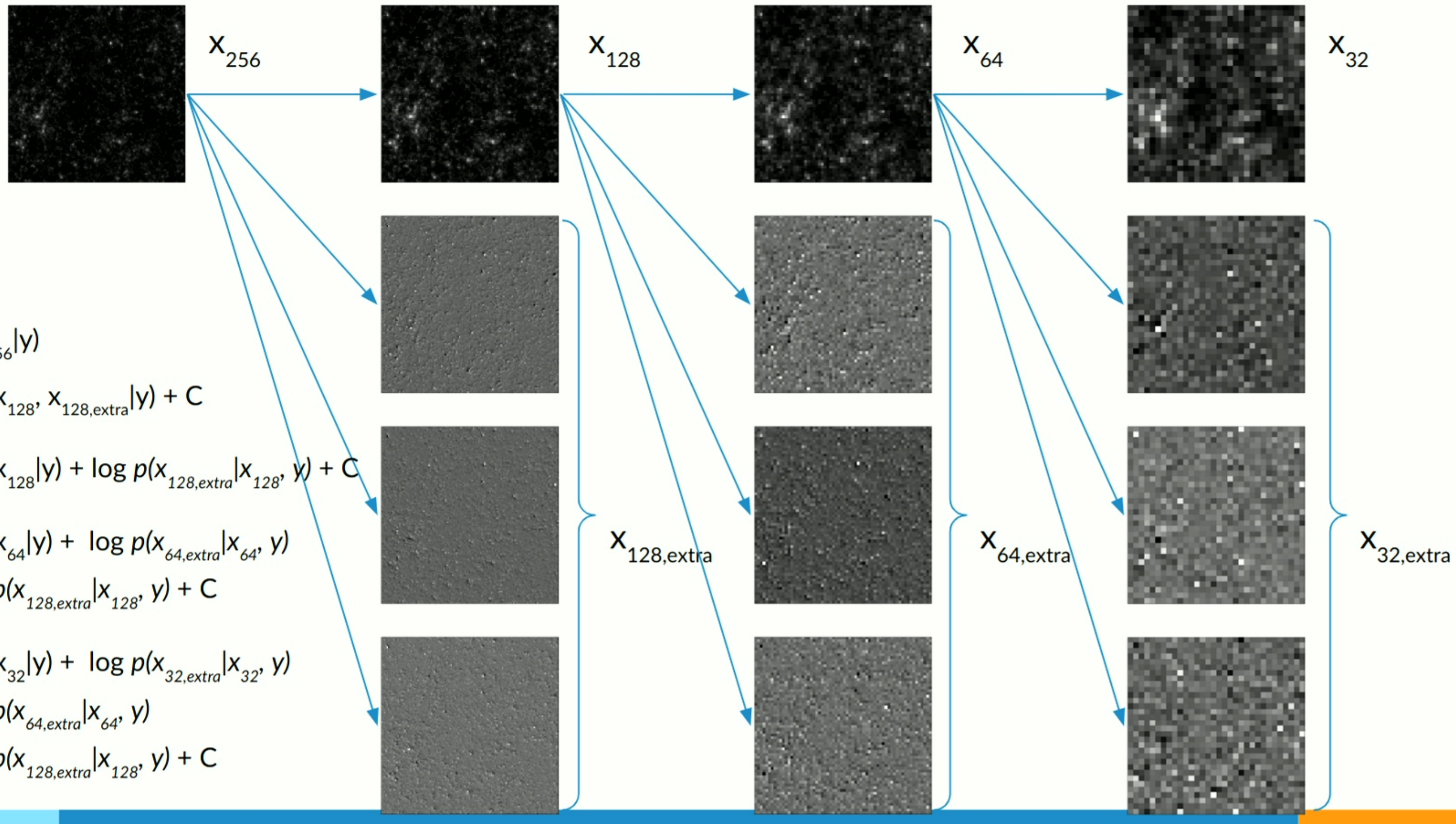
$= \log p(x_{128}, x_{128,extra}|y) + C$



$$\begin{bmatrix} x_{128} \\ x^1_{128,\text{extra}} \\ x^2_{128,\text{extra}} \\ x^3_{128,\text{extra}} \end{bmatrix} = \begin{bmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{bmatrix} \begin{bmatrix} C_1 \\ C_2 \\ C_3 \\ C_4 \end{bmatrix}$$

Multiscale flow

▷ Consider a cosmological field with 256^2 resolution:



$\log p(x_{256}|y)$

$= \log p(x_{128}, x_{128,extra}|y) + C$

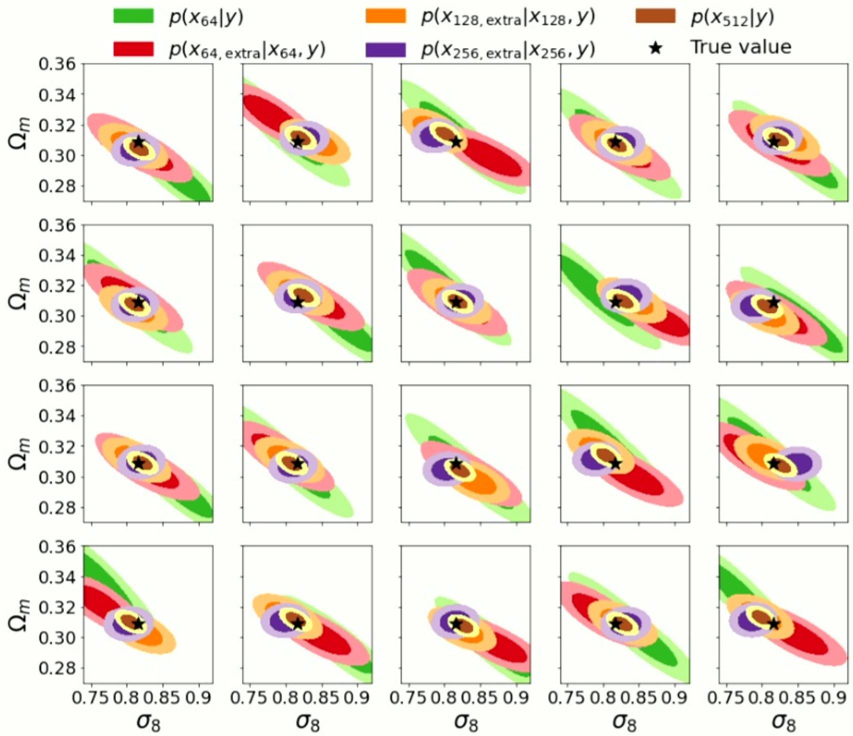
$= \log p(x_{128}|y) + \log p(x_{128,extra}|x_{128}, y) + C$

$= \log p(x_{64}|y) + \log p(x_{64,extra}|x_{64}, y)$
$\quad + \log p(x_{128,extra}|x_{128}, y) + C$

$= \log p(x_{32}|y) + \log p(x_{32,extra}|x_{32}, y)$
$\quad + \log p(x_{64,extra}|x_{64}, y)$
$\quad + \log p(x_{128,extra}|x_{128}, y) + C$

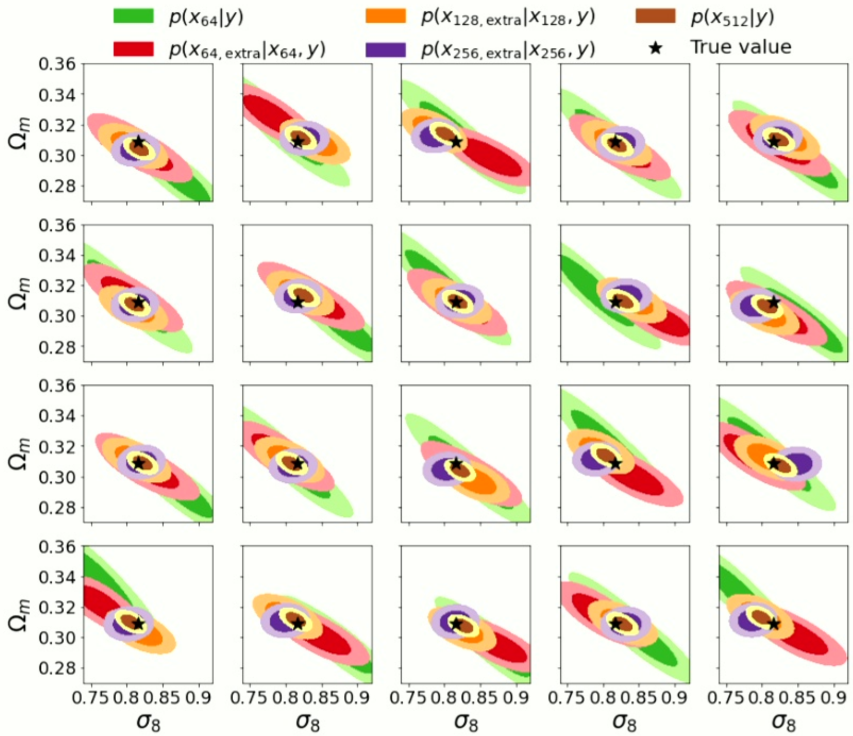
Distribution shift detection — noise miscalibration

- Consistent posteriors from different scales



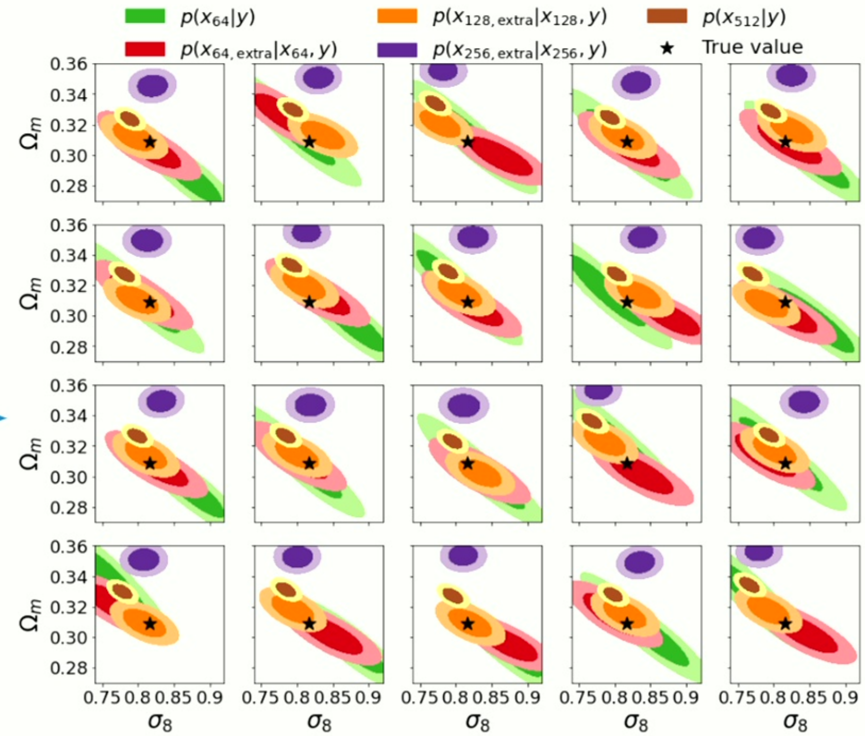
Distribution shift detection — noise miscalibration

- Consistent posteriors from different scales



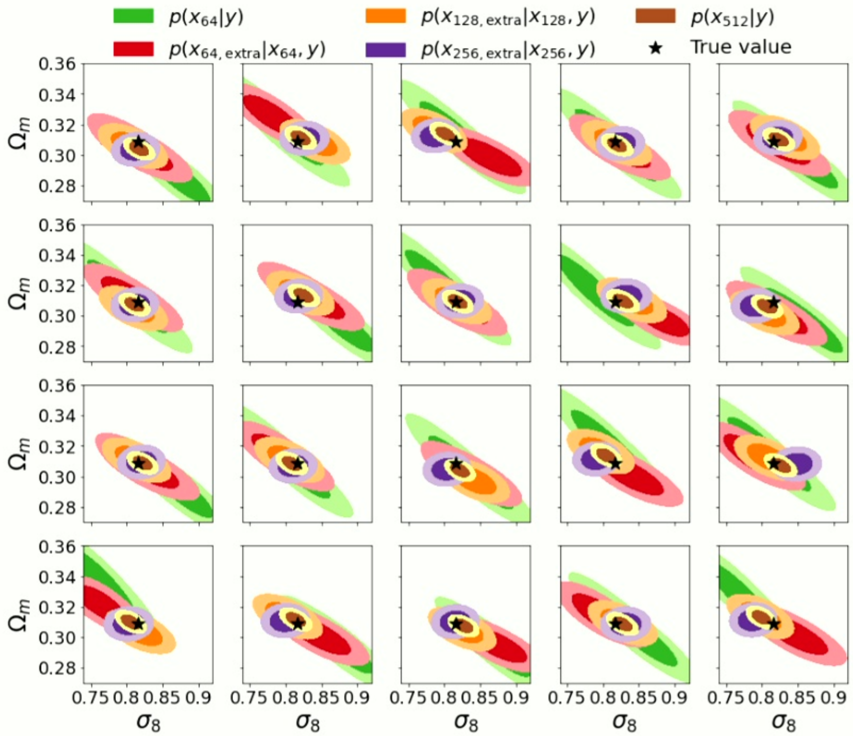
noise miscalibration

- Inconsistent small scale posterior



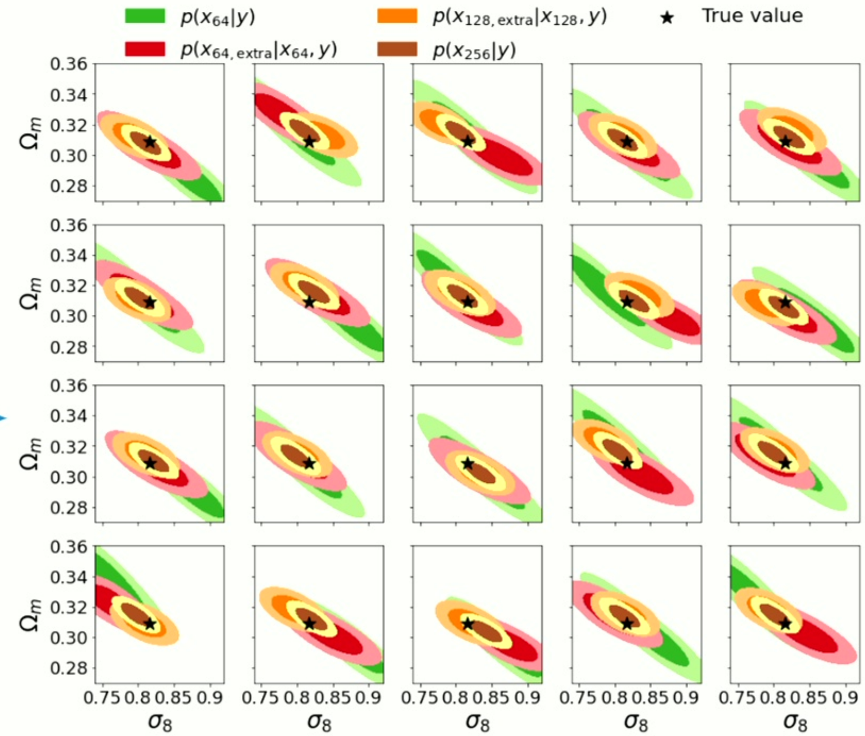
Distribution shift detection — noise miscalibration

- Consistent posteriors from different scales



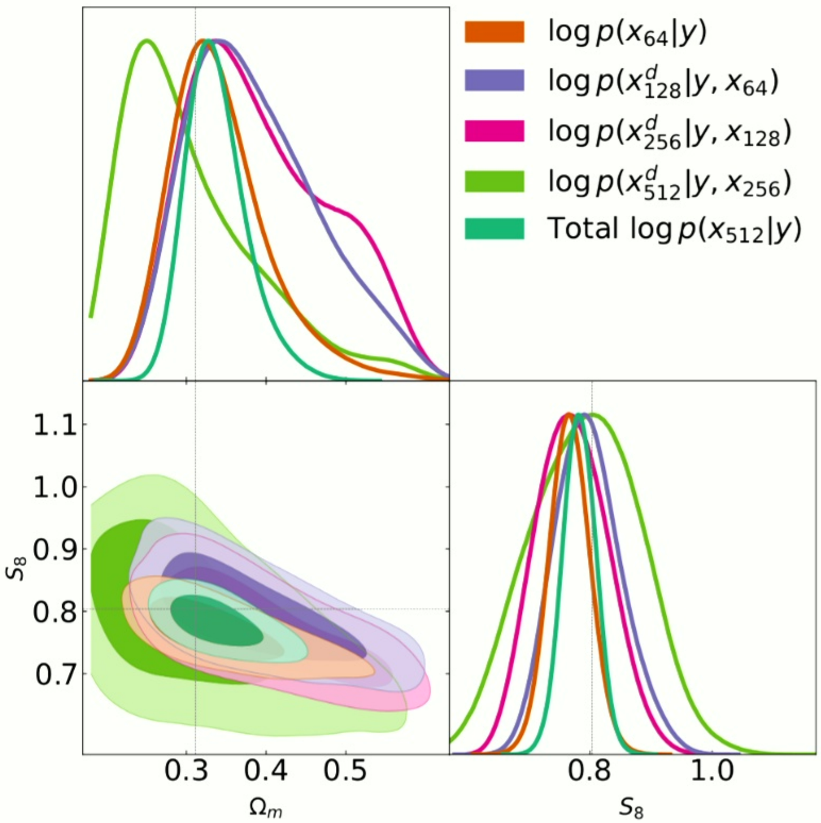
noise miscalibration

- Remove small scale information

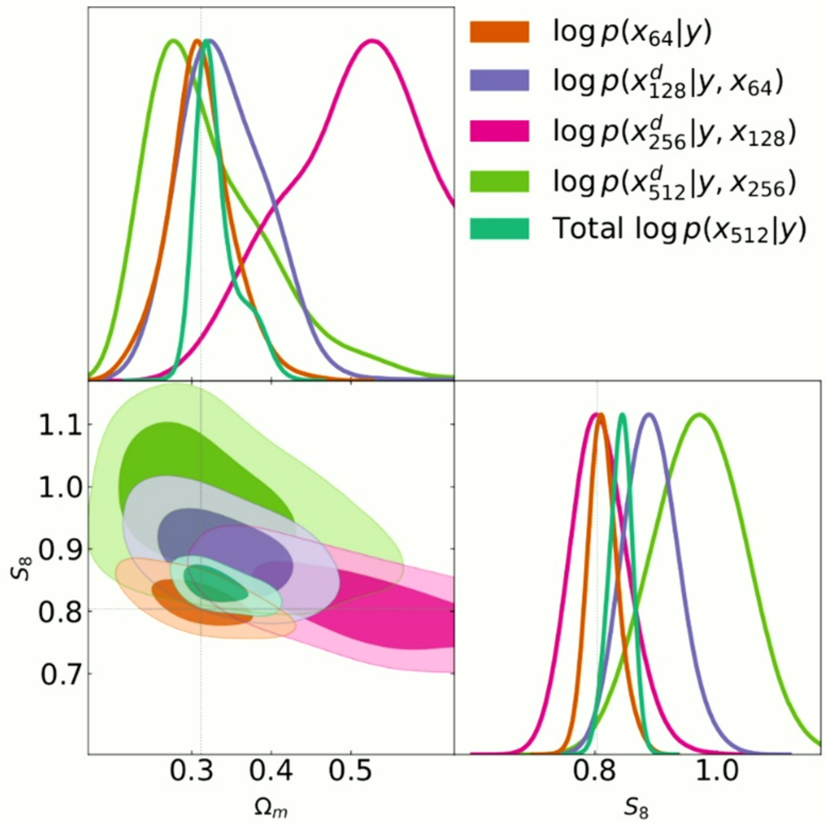


Distribution shift detection — baryon physics

- MSF trained on baryon models

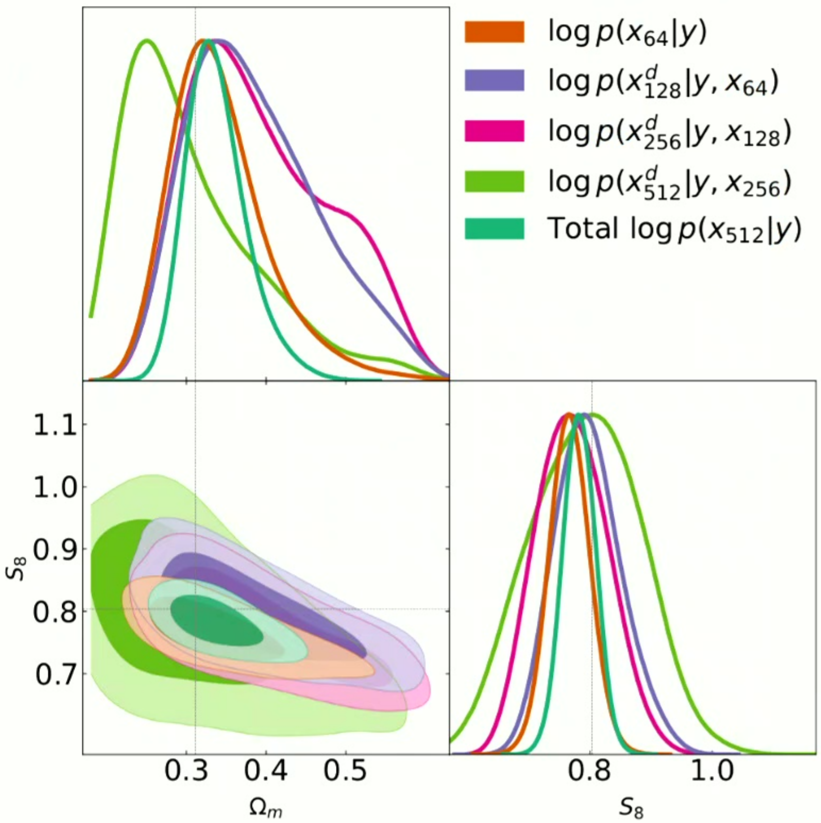


- MSF trained on DMO simulations

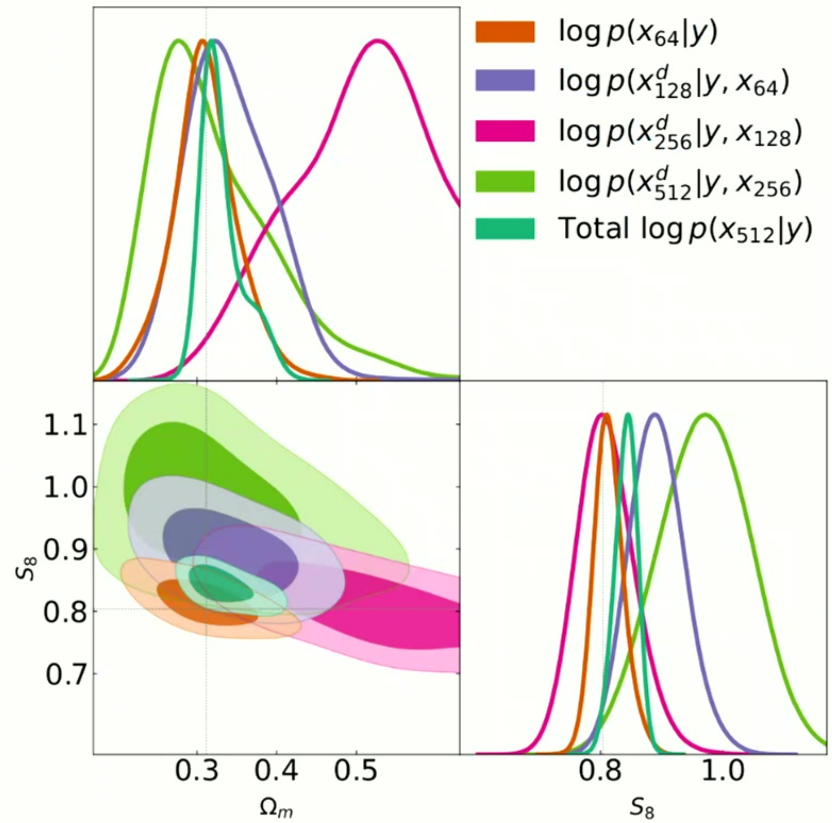


Distribution shift detection — baryon physics

- MSF trained on baryon models



- MSF trained on DMO simulations



51

Outline

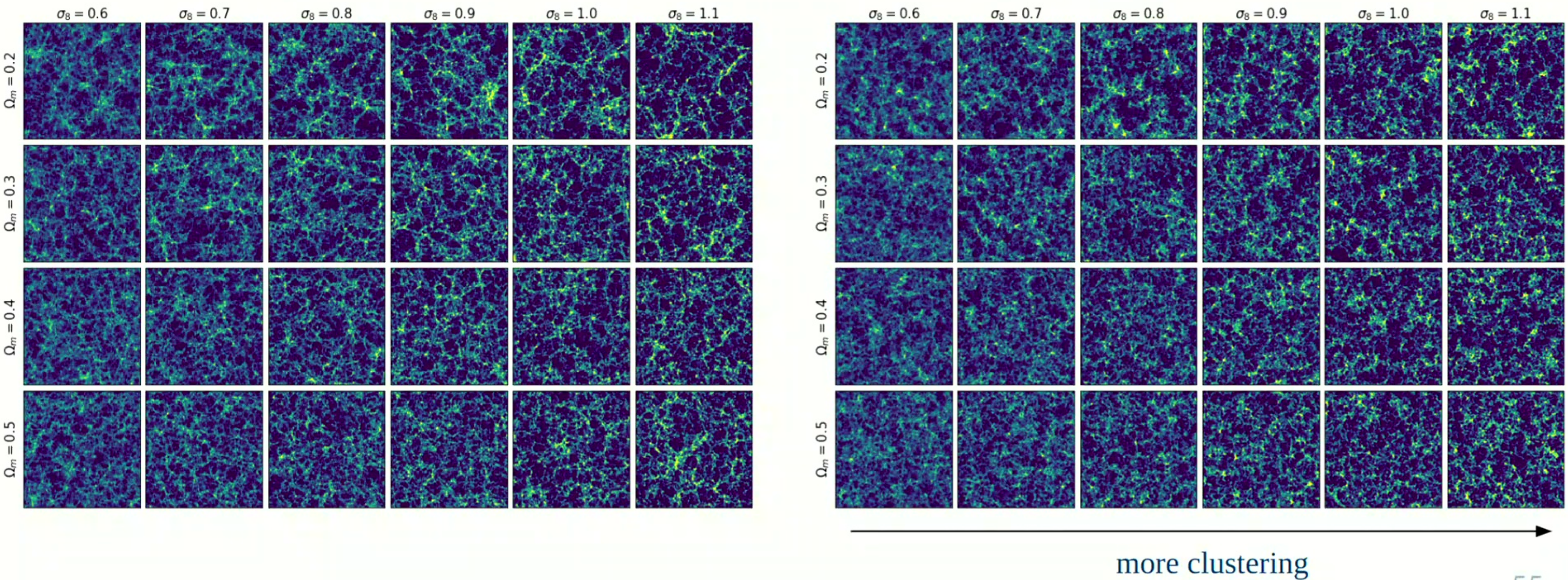
- ▷ Background: Cosmological data analysis and simulation-based inference
- ▷ Normalizing Flows (NFs) and what can they do for Physics
- ▷ Weak lensing analysis with generative NFs
  - **Optimal**: field-level information extraction
  - **Reliable**: anomaly detection of systematic effects
  - **Interpretable**: improve explainability with generated samples
  - Ongoing and future works
- ▷ Applications of NFs beyond cosmology
  - Anomaly detection of new physics in high energy physics
  - Lattice field variational inference
  - Speed up Bayesian sampling algorithm

Outline

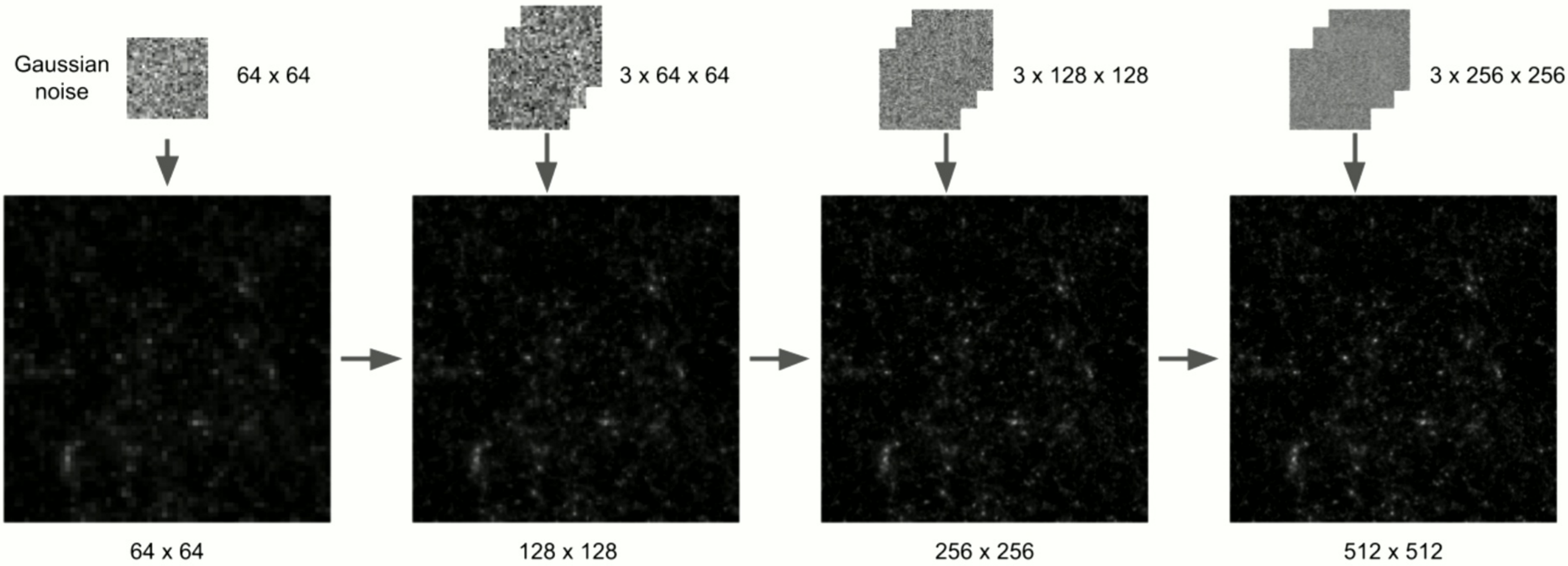
- ▷ Background: Cosmological data analysis and simulation-based inference
- ▷ Normalizing Flows (NFs) and what can they do for Physics
- ▷ Weak lensing analysis with generative NFs
  - ○ **Optimal**: field-level information extraction
  - ○ **Reliable**: anomaly detection of systematic effects
  - ○ **Interpretable**: improve explainability with generated samples
  - ○ Ongoing and future works
- ▷ Applications of NFs beyond cosmology
  - ○ Anomaly detection of new physics in high energy physics
  - ○ Lattice field variational inference
  - ○ Speed up Bayesian sampling algorithm

Fast LSS generation conditional on cosmology

- data:
- TRENF samples:

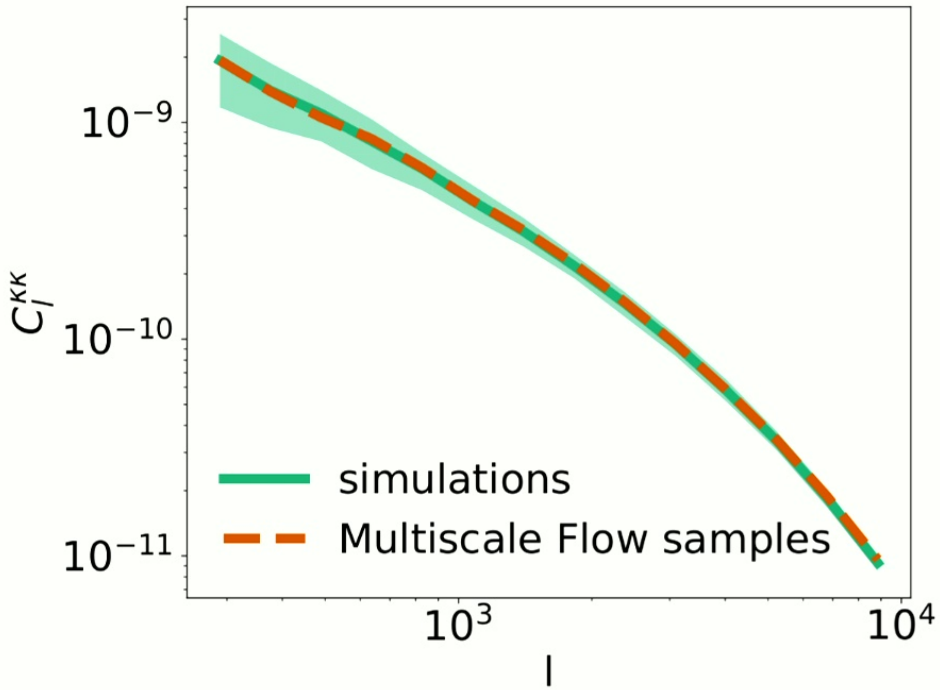


Sample generation & super-resolution

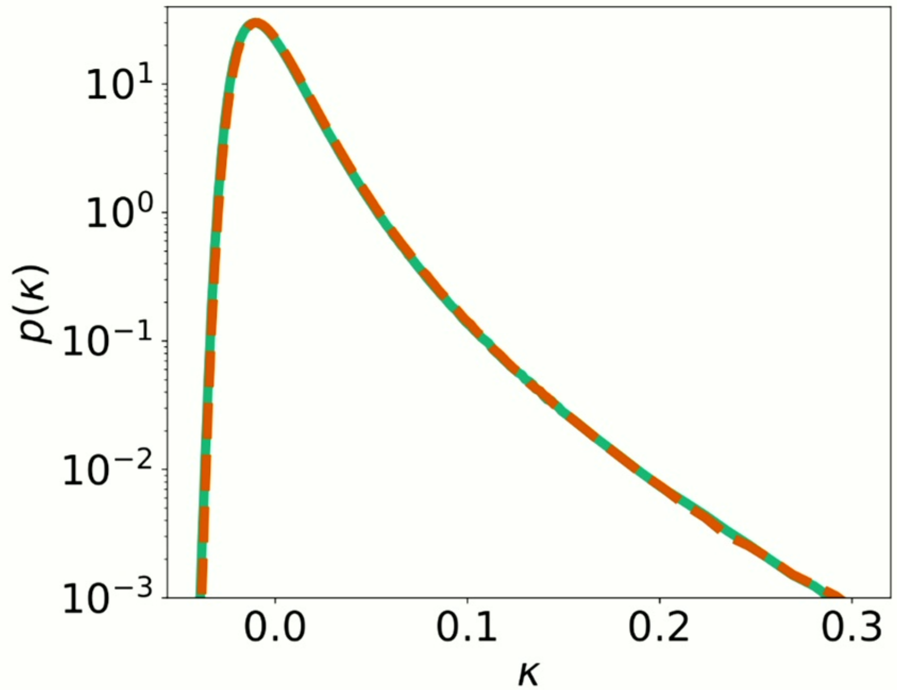


Sample generation & super-resolution

- power spectrum



- kappa probability distribution



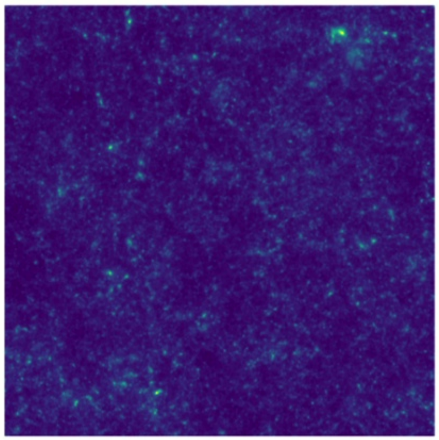
"Where is the extra information coming from?"

"You need to show why the other cosmological models are ruled out"

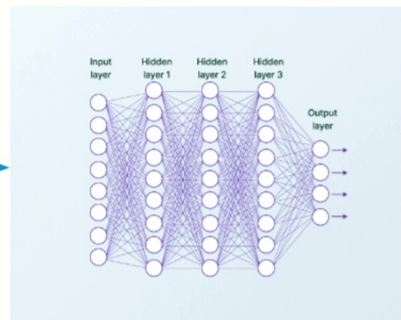
"Where is the extra information coming from?"

"You need to show why the other cosmological models are ruled out"

Input WL map



Generative models



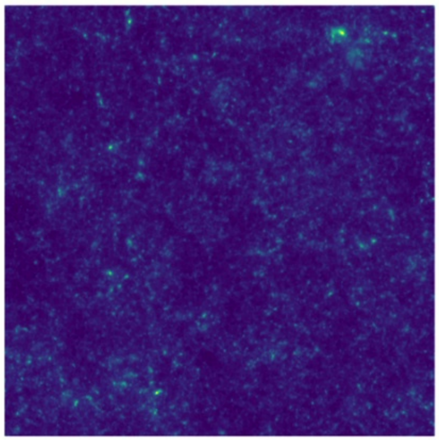
MCMC

$\sigma_8 = 0.76 \pm 0.02$

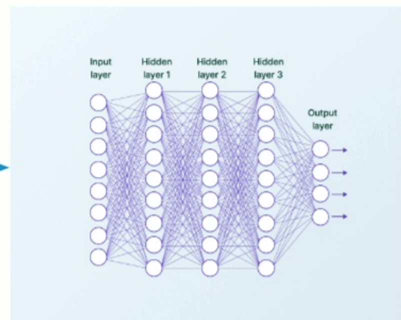
"Where is the extra information coming from?"

"You need to show why the other cosmological models are ruled out"

Input WL map



Generative models

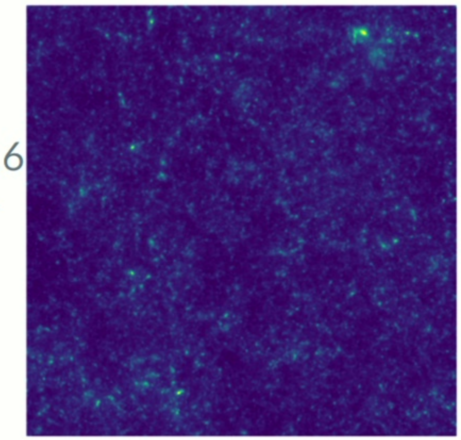


σ_8 = 0.816

MCMC

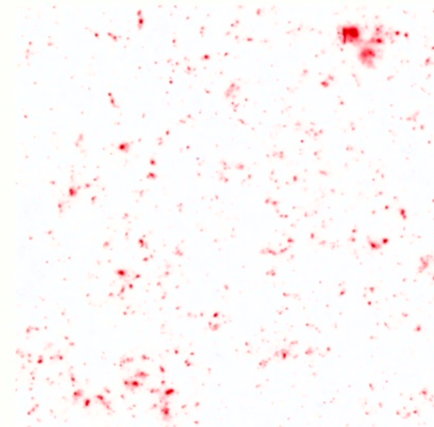
σ_8 = 0.76 ± 0.02

Generated sample



The same realization (latent code) as the input map, but assuming a different cosmology

Difference



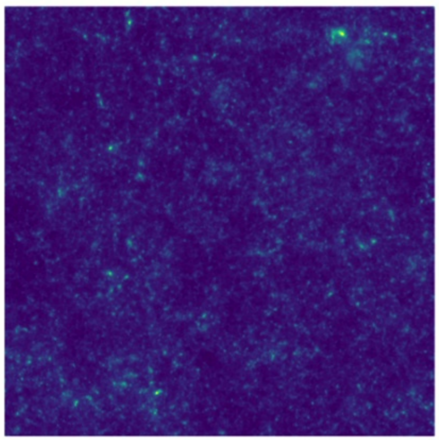
Generated sample - input map

“Where is the extra information coming from?”

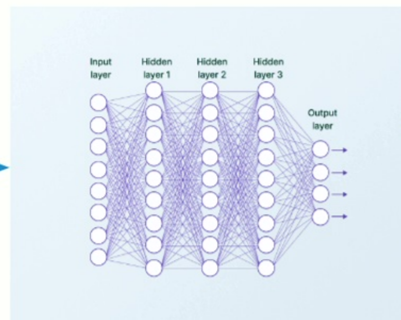
“You need to show why the other cosmological models are ruled out”

Generative models can visualize where the information is coming from, and how the constraints are made.

Input WL map



Generative models

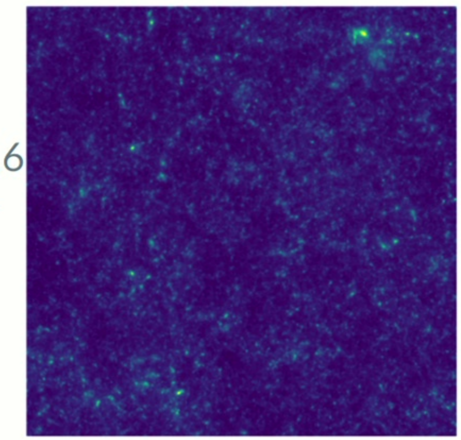


$\sigma_8 = 0.816$

MCMC

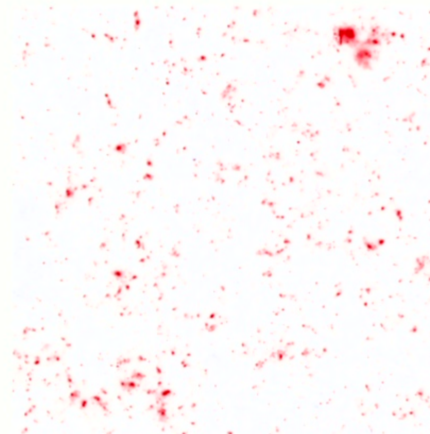
$\sigma_8 = 0.76 \pm 0.02$

Generated sample



The same realization (latent code) as the input map, but assuming a different cosmology

Difference



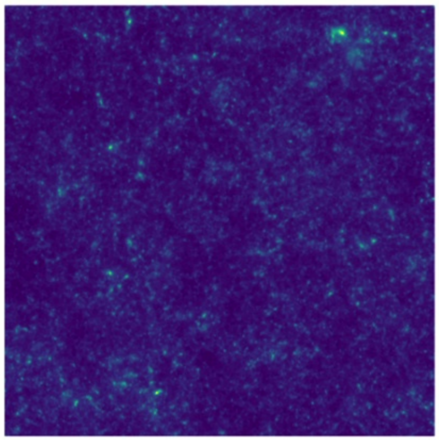
Generated sample - input map

"Where is the extra information coming from?"

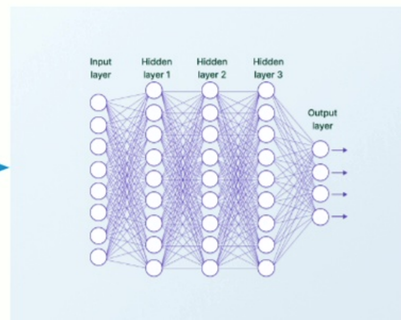
"You need to show why the other cosmological models are ruled out"

My model tells me that the halos from high σ_8 cosmology are too massive!

Input WL map



Generative models

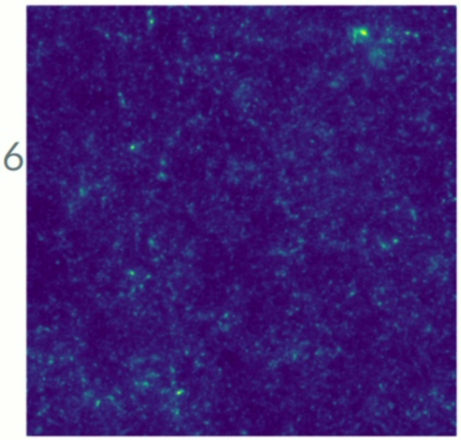


MCMC

$\sigma_8 = 0.76 \pm 0.02$

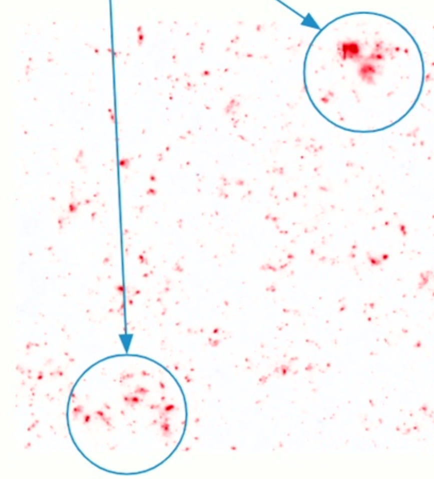
$\sigma_8 = 0.816$

Generated sample



The same realization (latent code) as the input map, but assuming a different cosmology

Difference



Generated sample - input map

Outline

- ▷ Background: Cosmological data analysis and simulation-based inference
- ▷ Normalizing Flows (NFs) and what can they do for Physics
- ▷ Weak lensing analysis with generative NFs
  - **Optimal**: field-level information extraction
  - **Reliable**: anomaly detection of systematic effects
  - **Interpretable**: improve explainability with generated samples
  - Ongoing and future works
- ▷ Applications of NFs beyond cosmology
  - Anomaly detection of new physics in high energy physics
  - Lattice field variational inference
  - Speed up Bayesian sampling algorithm

Outline

- ▷ Background: Cosmological data analysis and simulation-based inference
- ▷ Normalizing Flows (NFs) and what can they do for Physics
- ▷ Weak lensing analysis with generative NFs
  - **Optimal**: field-level information extraction
  - **Reliable**: anomaly detection of systematic effects
  - **Interpretable**: improve explainability with generated samples
  - Ongoing and future works
- ▷ Applications of NFs beyond cosmology
  - Anomaly detection of new physics in high energy physics
  - Lattice field variational inference
  - Speed up Bayesian sampling algorithm

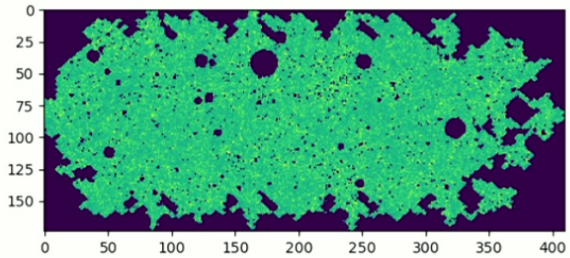
HSC weak lensing analysis with Multiscale Flow

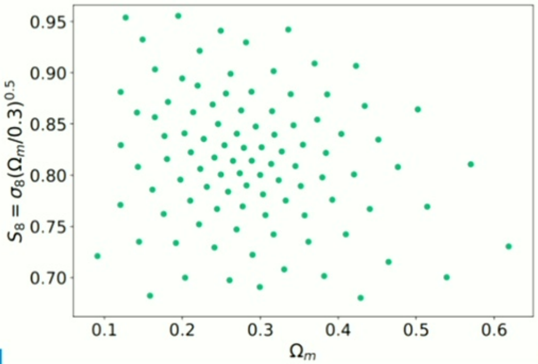
With *Xiangchong Li, Uroš Seljak and Rachel Mandelbaum*

- Subaru Hyper Suprime-Cam (HSC) WL survey
  - 330 nights of observation over 5-6 years
  - The deepest current WL survey, ideal for ML analysis

8.2m Subaru telescope

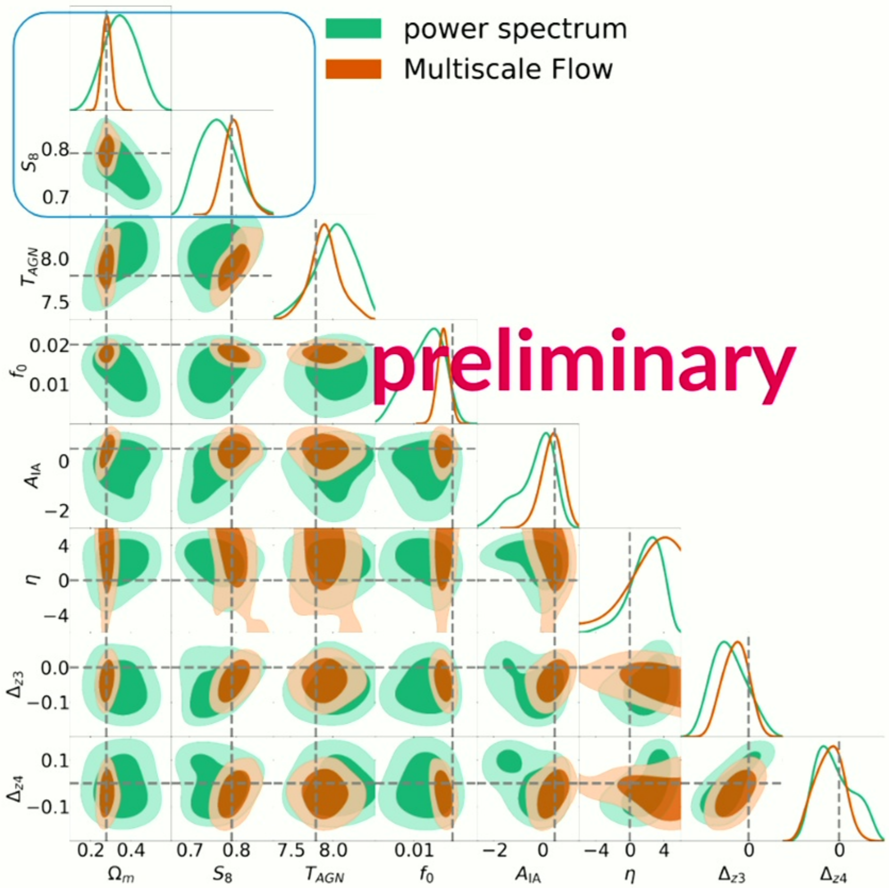
- Training maps:
  - O(100) high-resolution ray-tracing simulations with different cosmological models to simulate the large-scale structure
  - Build an ML emulator to add baryonic effect at the field level (Sharma, Dai, Villaescusa-Navarro and Seljak 2024)
  - Realistic systematic effects





HSC weak lensing analysis with Multiscale Flow

Cosmological constraints



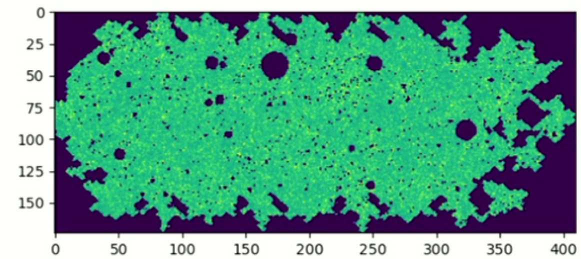
- Tests on mock data: significant improvement compared to traditional power spectrum analysis, after considering various systematic uncertainties

- From left to right:
  - the mean present-day matter density
  - a measure of the homogeneity of the Universe
  - 2 effective baryonic parameter
  - 2 intrinsic alignment parameter
  - 2 parameter of redshift estimation uncertainty

Future works — field-level weak lensing analysis

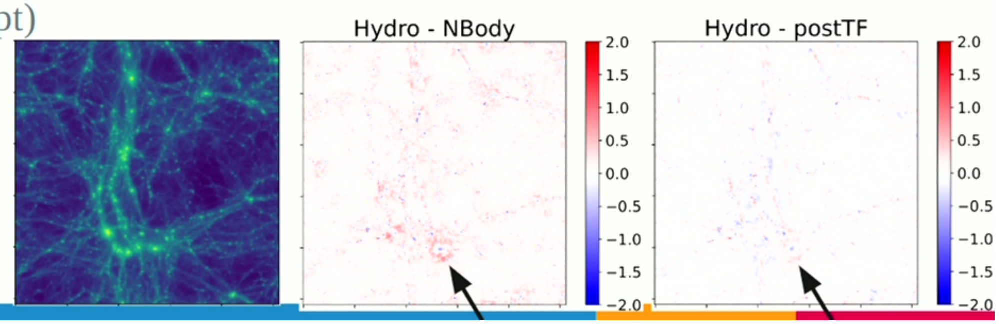
- Apply the field-level inference model I developed during my Ph.D. to weak lensing cosmic shear analysis

- Short term: HSC public galaxy shape catalog
  - Run 100 ray-tracing simulations with different cosmological parameters to generate weak lensing maps for training
  - Systematic effects: IA, baryon, photo-z, etc.
  - $O(10^5)$ training maps
  - Current stage: model validation on mock data



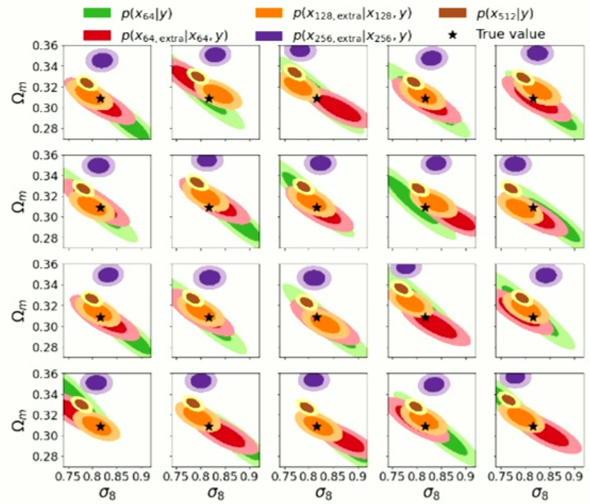
- Long term:
  - Upcoming surveys: Euclid, Rubin, Roman, etc.
  - Joint analysis with other probes (similar to 3x2pt)

- Improve the modeling of systematics
  - Baryonic physics
  - Intrinsic alignment
  - Source clustering effect



Future works — Reliable ML for cosmology and astrophysics

- Motivation: there is a significant risk of DL models not generalizing in the presence of the domain shift between training simulations and observations

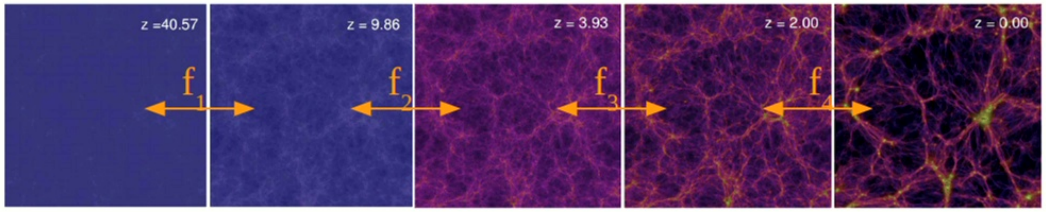


- Develop different tests to verify the reliability of the analysis
  - Goodness-of-fit test (anomaly detection / Out-of-distribution detection)
  - Consistency test: robustness to different modeling choices, scales, redshift bins, different patches of the sky, etc.
  - Null tests (e.g., B modes)
  - Blind analysis to reduce confirmation bias

- Weak lensing data challenge
  - Standardized dataset in Cosmology
  - Focus on discovering and minimizing the effects of systematic uncertainties

- Incorporate physical knowledge (e.g., EPT) into NFs and force latent space to be the same as IC (Eulerian simulation)

Measure PNG in latent space



Credit: Zhao et al. 2012

Outline

- ▷ Background: Cosmological data analysis and simulation-based inference

- ▷ Normalizing Flows (NFs) and what can they do for Physics

- ▷ Weak lensing analysis with generative NFs
  - **Optimal**: field-level information extraction
  - **Reliable**: anomaly detection of systematic effects
  - **Interpretable**: improve explainability with generated samples
  - Ongoing and future works

- ▷ Applications of NFs beyond cosmology
  - Anomaly detection of new physics in high energy physics
  - Lattice field variational inference
  - Speed up Bayesian sampling algorithm

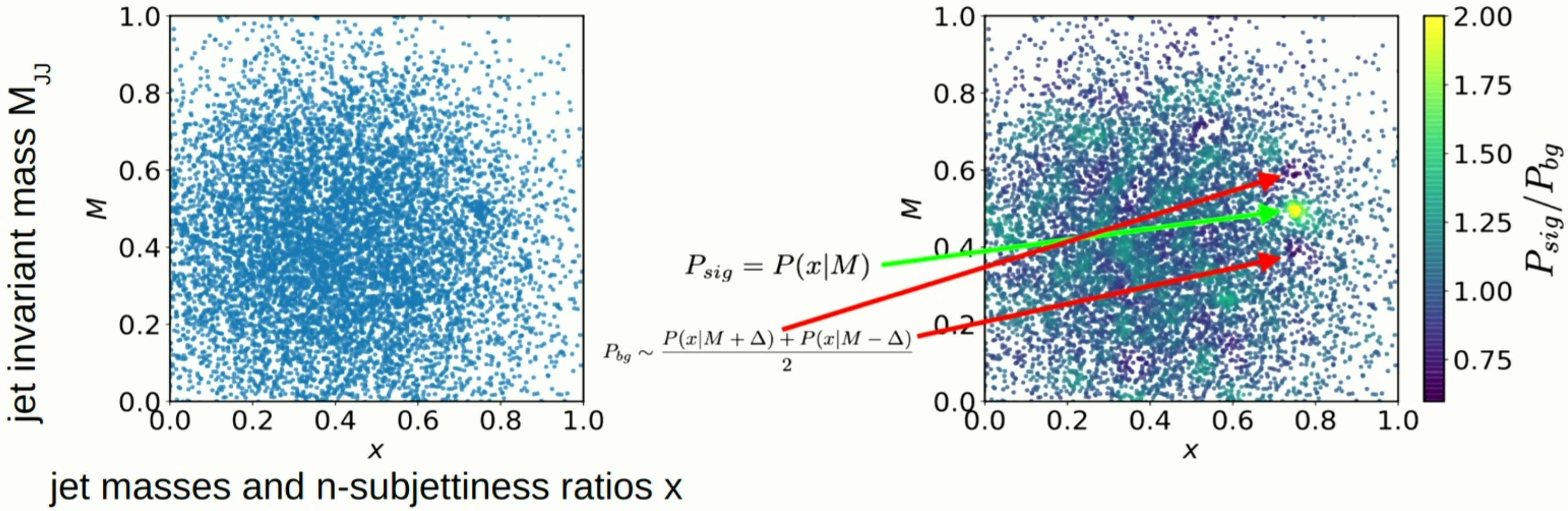
Anomaly detection of new physics

- Search for physics beyond the Standard Model (BSM): (small) unknown number of BSM resonant signal events buried in large number of SM background in LHC hadronic events

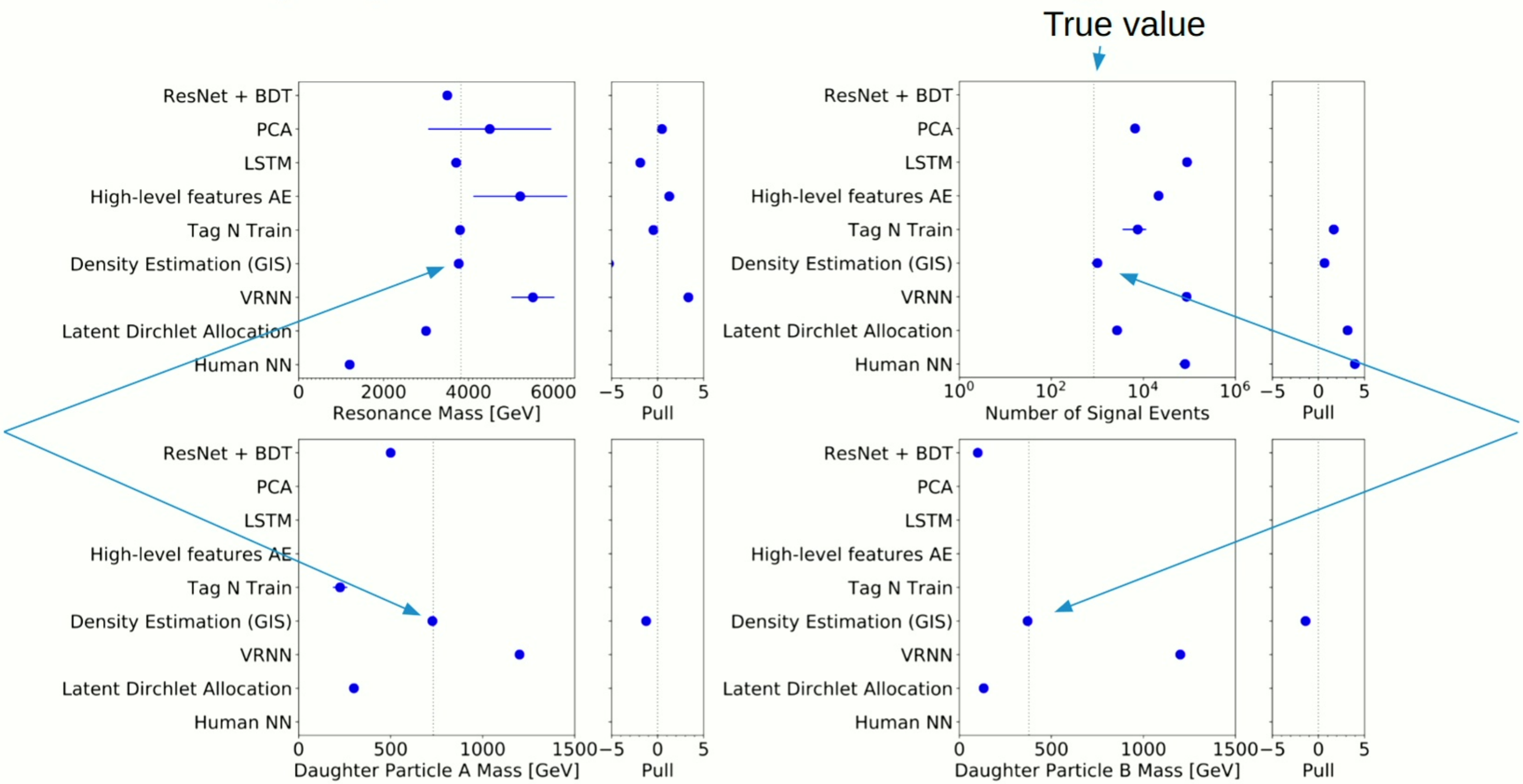
Anomaly detection of new physics

- Search for physics beyond the Standard Model (BSM): (small) unknown number of BSM resonant signal events buried in large number of SM background in LHC hadronic events

- Data-driven approach based on density estimation with normalizing flows (Dai & Seljak, ICML 2021):
  - Look for excess density in a narrow region of a parameter of interest, such as the invariant mass



LHC Olympics 2020 blind challenge results



Density estimation (Dai & Seljak, ICML 2021): the only entry with all correct answers

Lattice field variational inference

- Lattice field theory (LFT) is a robust and universal method to extract physical observables from a non-perturbative quantum field theory (QFT).
  - MCMC methods: hard due to critical slowing down, where the correlation of the MCMC samples increases dramatically close to critical temperature
  - Variational inference: approximate target $p(x)$ with a family of parameterized PDF $q_\theta(x)$

Lattice field variational inference

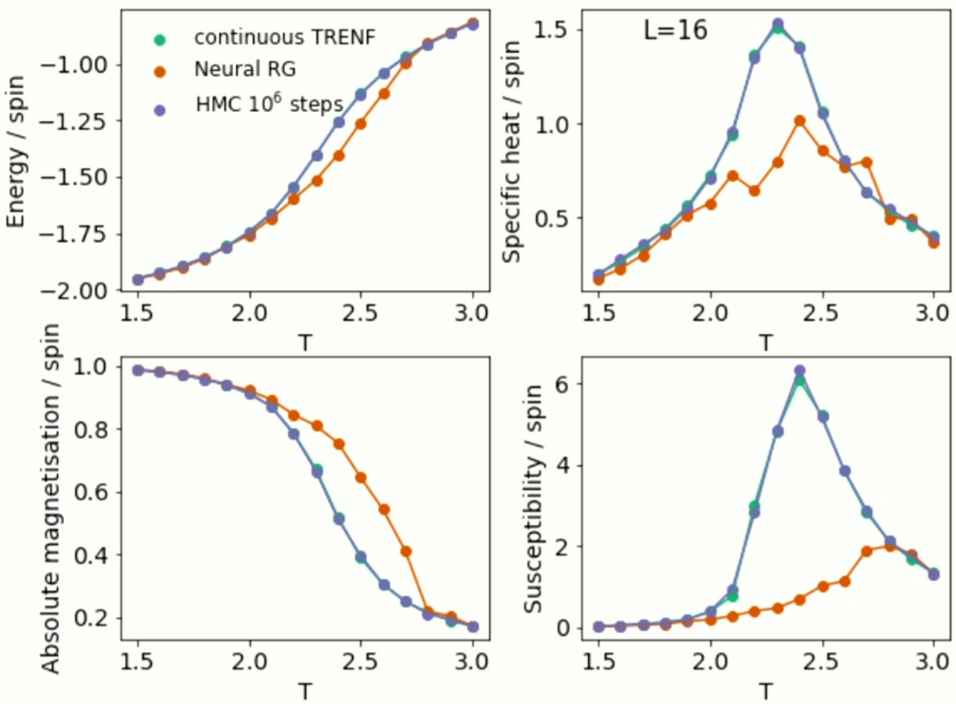
- Lattice field theory (LFT) is a robust and universal method to extract physical observables from a non-perturbative quantum field theory (QFT).
  - MCMC methods: hard due to critical slowing down, where the correlation of the MCMC samples increases dramatically close to critical temperature
  - Variational inference: approximate target p(x) with a family of parameterized PDF $q_\Theta(x)$

- 2D Ising model:
  - $\log p_\beta(s) = -\beta E(s) - \log Z_\beta$, where $E(s) = -0.5\, s^T J s - h^T s$
  - Hubbard-Stratonovich transformation rewrite the problem with continuous variables x

- Variational inference with TRENF
  - Minimize the reverse KL divergence between q(x) and p(x): $E_{q(x)}[\log q(x) - \log p(x)]$
  - Run independent Metropolis-Hasting MCMC algorithm with proposal distribution q(x)

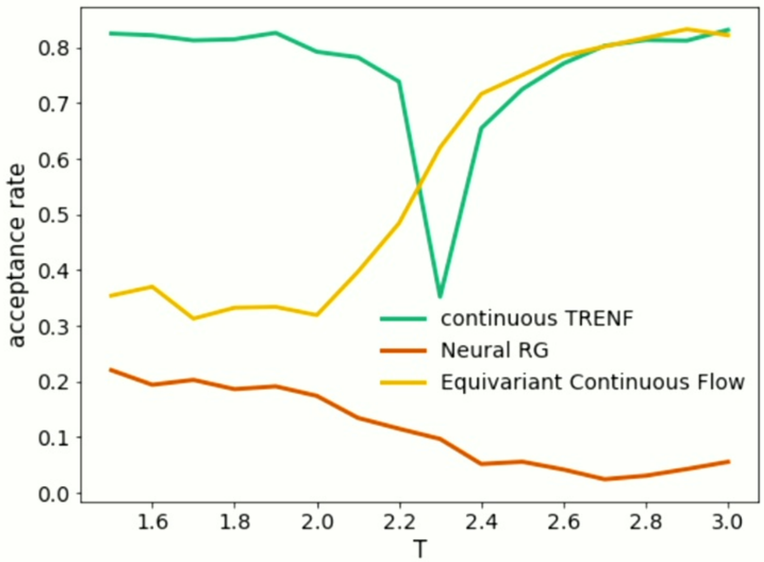
Ising model variational inference

- Thermodynamic quantities



- IMH acceptance rate

$$\alpha = \min\left\{1, \frac{\pi(\mathbf{x}')q(\mathbf{x})}{\pi(\mathbf{x})q(\mathbf{x}')}\right\}$$



Outline

- ▷ Background: Cosmological data analysis and simulation-based inference
- ▷ Normalizing Flows (NFs) and what can they do for Physics
- ▷ Weak lensing analysis with generative NFs
  - **Optimal**: field-level information extraction
  - **Reliable**: anomaly detection of systematic effects
  - **Interpretable**: improve explainability with generated samples
  - Ongoing and future works
- ▷ Applications of NFs beyond cosmology
  - Anomaly detection of new physics in high energy physics
  - Lattice field variational inference
  - Speed up Bayesian sampling algorithm

Outline

- ▷ Background: Cosmological data analysis and simulation-based inference
- ▷ Normalizing Flows (NFs) and what can they do for Physics
- ▷ Weak lensing analysis with generative NFs
  - **Optimal**: field-level information extraction
  - **Reliable**: anomaly detection of systematic effects
  - **Interpretable**: improve explainability with generated samples
  - Ongoing and future works
- ▷ Applications of NFs beyond cosmology
  - Anomaly detection of new physics in high energy physics
  - Lattice field variational inference
  - Speed up Bayesian sampling algorithm

Deterministic Langevin Monte Carlo (DLMC):

Fast Bayesian inference with Langevin dynamics + Normalizing Flows

- Goal: Sampling from an intractable probability distribution $p(x) = \exp(-U(x)) / Z$

Deterministic Langevin Monte Carlo (DLMC):

Fast Bayesian inference with Langevin dynamics + Normalizing Flows

- Goal: Sampling from an intractable probability distribution p(x) = exp(-U(x)) / Z

- Overdamped Langevin equation: $\dot{\boldsymbol{x}} = \boldsymbol{v} = -\boldsymbol{\nabla} U(\boldsymbol{x}) + \boldsymbol{\eta}$
  - The invariant distribution of particles x is the target distribution p(x)!
  - Main issue: highly correlated samples due to its stochastic nature

Deterministic Langevin Monte Carlo (DLMC):

Fast Bayesian inference with Langevin dynamics + Normalizing Flows

- Goal: Sampling from an intractable probability distribution p(x) = exp(-U(x)) / Z

- Overdamped Langevin equation: $\dot{\boldsymbol{x}} = \boldsymbol{v} = -\boldsymbol{\nabla} U(\boldsymbol{x}) + \boldsymbol{\eta}$
  - The invariant distribution of particles x is the target distribution p(x)!
  - Main issue: highly correlated samples due to its stochastic nature

- **DLMC: Speed up Langevin dynamics with Normalizing Flows (NFs)**
  - Convert the stochastic equation to deterministic to reduce correlation length:

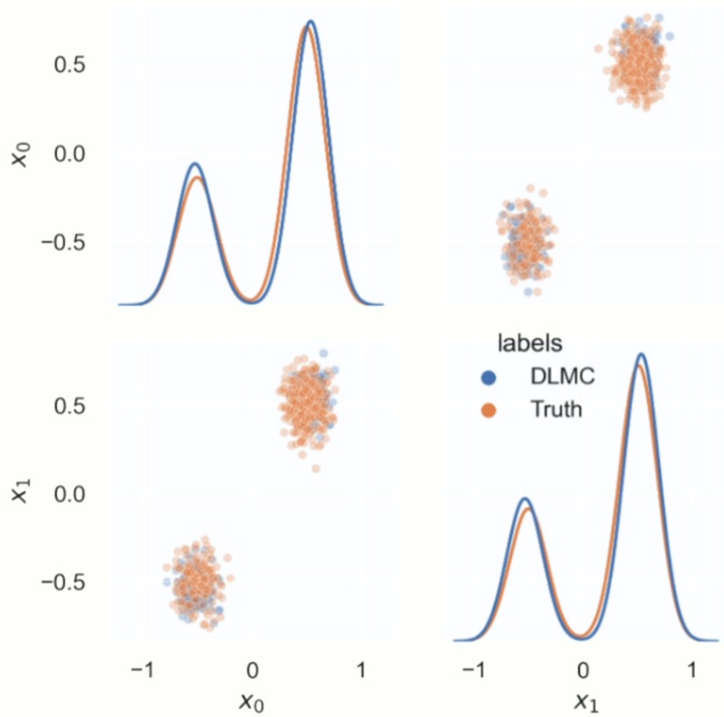
$$\dot{\boldsymbol{x}}(t) = \boldsymbol{v} = -\boldsymbol{\nabla}[U(\boldsymbol{x}) - V(\boldsymbol{x}(t))] \;, \quad \text{where} \quad V(\boldsymbol{x}(t)) = -\ln q(\boldsymbol{x}(t))$$

  - Unbiased: NF-based Metropolis-Hastings adjustment
  - NF-based Preconditioning

(Grumitt, Dai and Seljak, NeurIPS 2022)

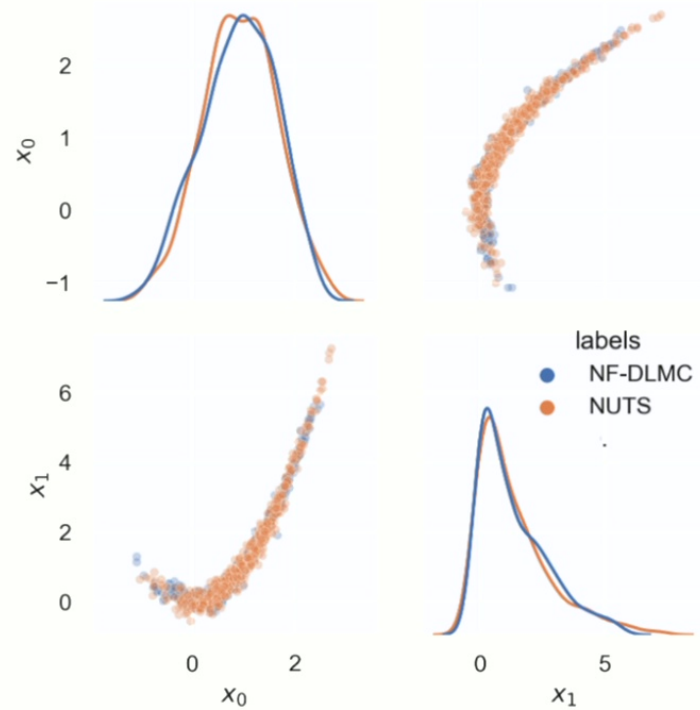
DLMC works for very hard problems

- 100D double gaussian



$\log V_{prior} / V_{posterior} = 415$

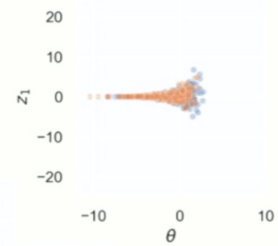
- 32D Rosenbrock function

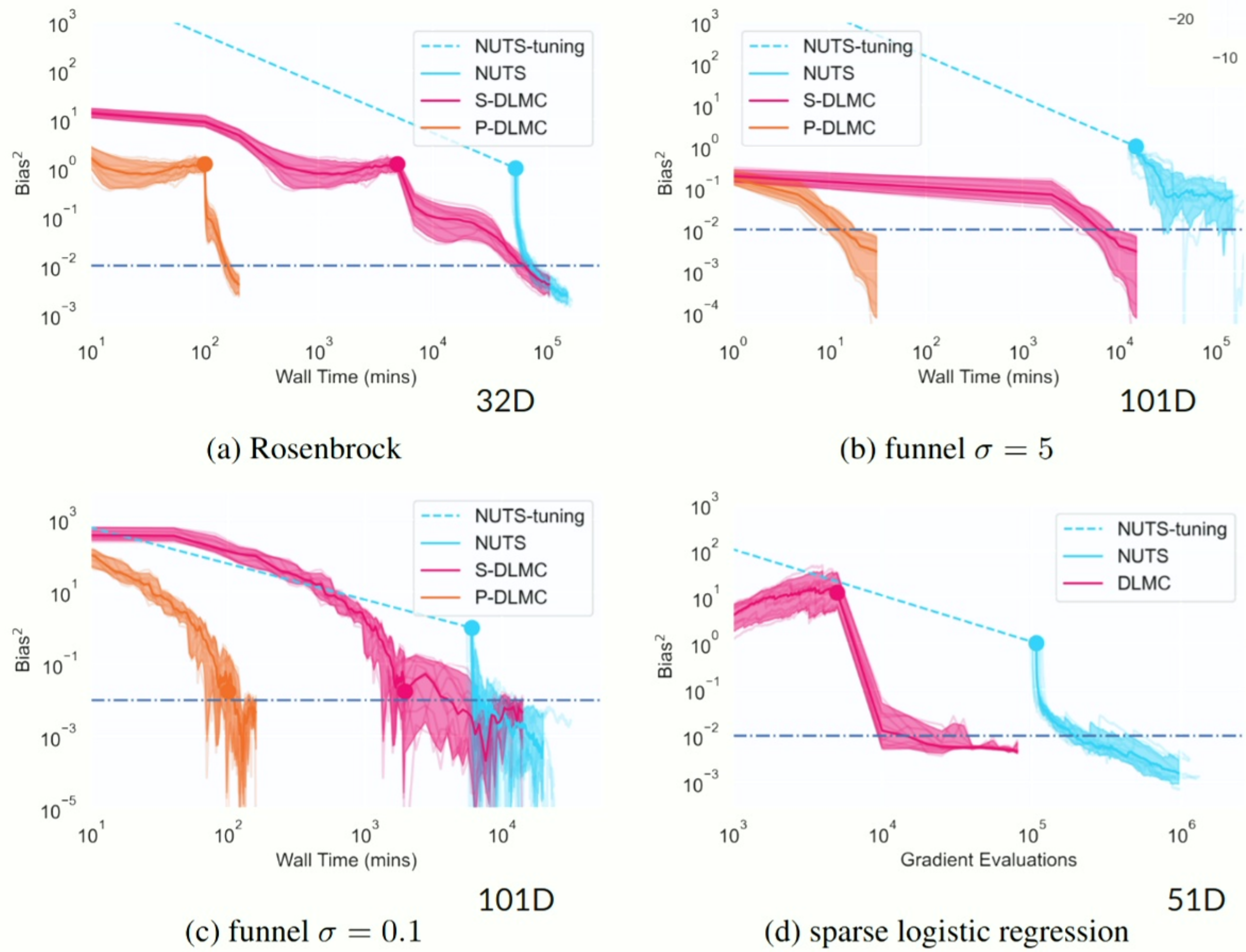


16 uncorrelated bananas

(Grumitt, Dai and Seljak, NeurIPS 2022)

Comparison with HMC





- ▷ DLMC is comparable or more efficient than HMC, and works for problems where HMC fails.
- ▷ DLMC is an ensemble algorithm and is embarrassingly parallelizable, which further reduces the wall-clock time by a factor of 10^2 - 10^3 (p-DLMC).
- ▷ DLMC's primary target application is to expensive likelihoods (wall clock time of seconds or more), where the NF cost is negligible.

(Grumitt, Dai and Seljak, NeurIPS 2022)

Conclusions (What can Normalizing Flows do for Physics?)

- Extract physical information (simulation-based inference)
- Fast sample generation
- Anomaly detection
  - Detect systematic effect (distribution shift)
  - Search for new physics
- Approximate intractable distributions
  - Variational Inference
  - Fast Bayesian sampling

Conclusions (What can Normalizing Flows do for Physics?)

- Extract physical information (simulation-based inference)
- Fast sample generation
- Anomaly detection
  - Detect systematic effect (distribution shift)
  - Search for new physics
- Approximate intractable distributions
  - Variational Inference
  - Fast Bayesian sampling

Cosmological analysis:

- We build NFs with physics prior knowledge (symmetry and multiscale structure) for (**optimal**) field-level cosmological information extraction
- **Interpretable**: NFs are able to generate realistic samples conditional on cosmology, which helps to visualize where the information is coming from
- **Reliable**: We develop goodness-of-fit test and multiscale analysis for anomaly detection of distribution shift
- We are applying the model to HSC weak lensing analysis. Stay tuned!

Conclusions (What can Normalizing Flows do for Physics?)

- Extract physical information (simulation-based inference)
- Fast sample generation
- Anomaly detection
  - Detect systematic effect (distribution shift)
  - Search for new physics
- Approximate intractable distributions
  - Variational Inference
  - Fast Bayesian sampling

Beyond Cosmology:

- **High Energy Physics**: NFs enable anomaly detection, and outperforms other approaches in a LHC Olympics blind challenge
- Lattice field theory: NFs generate realistic Ising model samples with accurate thermodynamic quantities
- Sampling: Deterministic Langevin Monte Carlo (DLMC) uses NFs to speed up Langevin dynamics, and outperforms HMC on various problems.

Conclusions (What can Normalizing Flows do for Physics?)

- Extract physical information (simulation-based inference)
- Fast sample generation
- Anomaly detection
  - Detect systematic effect (distribution shift)
  - Search for new physics
- Approximate intractable distributions
  - Variational Inference
  - Fast Bayesian sampling

Beyond Cosmology:

- High Energy Physics: NFs enable anomaly detection, and outperforms other approaches in a LHC Olympics blind challenge
- Lattice field theory: NFs generate realistic Ising model samples with accurate thermodynamic quantities
- **Sampling**: Deterministic Langevin Monte Carlo (DLMC) uses NFs to speed up Langevin dynamics, and outperforms HMC on various problems.

Conclusions (What can Normalizing Flows do for Physics?)

- Extract physical information (simulation-based inference)
- Fast sample generation
- Anomaly detection
  - Detect systematic effect (distribution shift)
  - Search for new physics
- Approximate intractable distributions
  - Variational Inference
  - Fast Bayesian sampling

Beyond Cosmology:

- High Energy Physics: NFs enable anomaly detection, and outperforms other approaches in a LHC Olympics blind challenge
- Lattice field theory: NFs generate realistic Ising model samples with accurate thermodynamic quantities
- **Sampling**: Deterministic Langevin Monte Carlo (DLMC) uses NFs to speed up Langevin dynamics, and outperforms HMC on various problems.