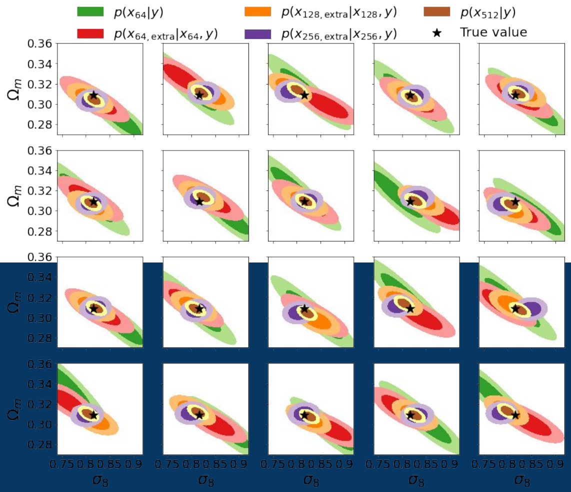
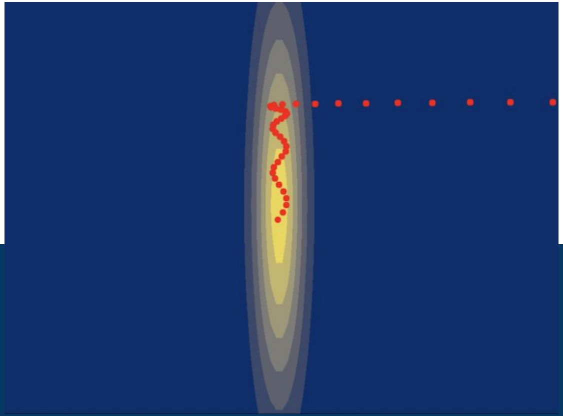
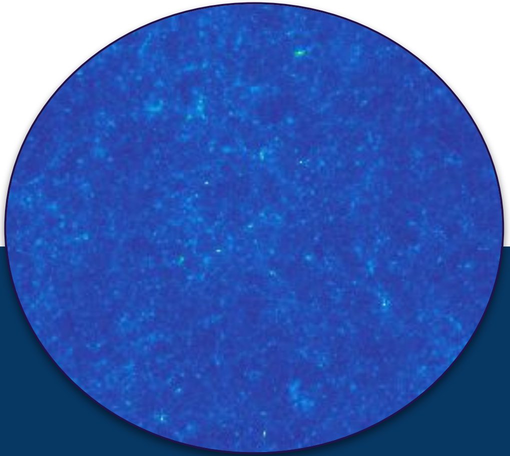
Cosmology with AI







Uroš Seljak

UC Berkeley

Lawrence Berkeley National Lab

BERKELEY CENTER *for* COSMOLOGICAL PHYSICS

Why do observational cosmology?

- Big discovery potential: new phenomenon, new fundamental physics…

To achieve it we need:

- Ability to observe new things: new instruments, telescopes...
- Develop (computational) models to explain the observable phenomena, while also modeling all the nuisances
- Develop computational analysis methods that are **optimal**, **robust against systematics, interpretable**, and **computationally feasible**…
- There is of course no guarantee that we will be able to discover something big, but we try to do our best
- **AI/ML** can help us achieve these goals, often with much less effort than traditional methods

How to use AI in cosmology

1) The goal of AI is to extract scientific information using a data driven approach. Typical application is **data inference**, e.g. learn about probability distribution of some **parameters y** from **data x p(y|x),** other applications are **anomaly detection, generative modeling** etc.

2) In AI this is done by using **training data** (X,Y) to learn p(y|x) of parameters y we wish to extract from some real data x which is not in the training data set

3) Often we do not have real data to train on. We may however have simulations. In this case simulations become training data. The corresponding concept is called **Simulation Based Inference** (also known as Likelihood Free Inference, Implicit Likelihood etc.)

4) Astronomy data are complex: images, spectra, photometry etc. there are a number of scientific tasks one can do with it. It makes sense to pre-process the data to facilitate and accelerate the tasks. This is called **Foundation Model**.

5) In the (near?) future almost all of technical work will be done by **agents**: you prompt what you want to get done and agents will perform the task for you

Simulation Based Inference

Use simulation pairs (X,Y) to train an inference model

1) **Discriminative**: were we wish to determine parameters y from simulations x: a common **discriminative training objective is L2 loss** $\mathbf{min_w \sum (y_w(x)\text{-}y_{true})^2}$ over all the training data (X,Y_{true}).
2) **Posterior Estimation**: more generally, we can predict posterior **p(y|x),** which may be a Gaussian defined by the mean $y_{mean}(x)$ and the variance $\sigma^2(x)$, or some other distribution **p(y|x)**.

3) We do this by replacing on the last layer the L2 loss with a Normalizing Flow that maximizes on training data log **p(y|x)**

4) **Generative**: learn conditional probability density of the data p(x|y)

5) Use Bayes Theorem to do inference: p(y|x)=p(x|y)p(y)/p(x)

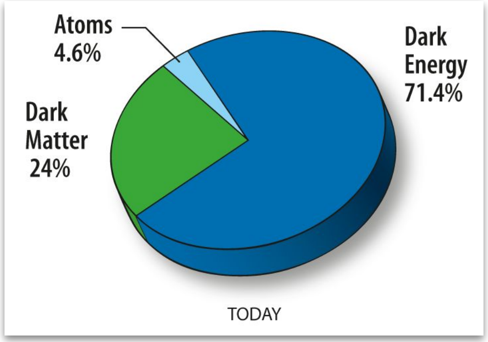
This approach gives richer set of applications: **anomaly detection, generative modeling** etc.

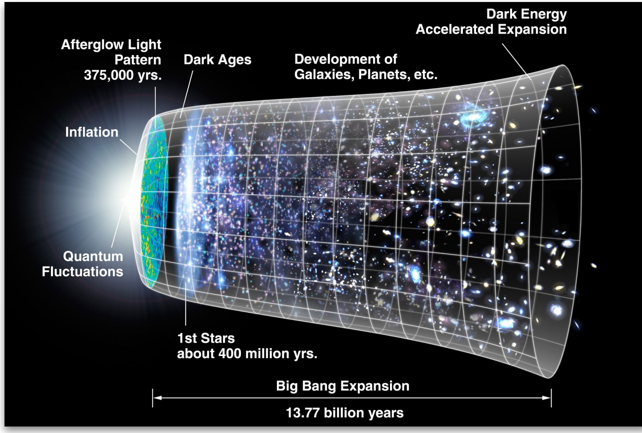
Prelude

Cosmology: status report

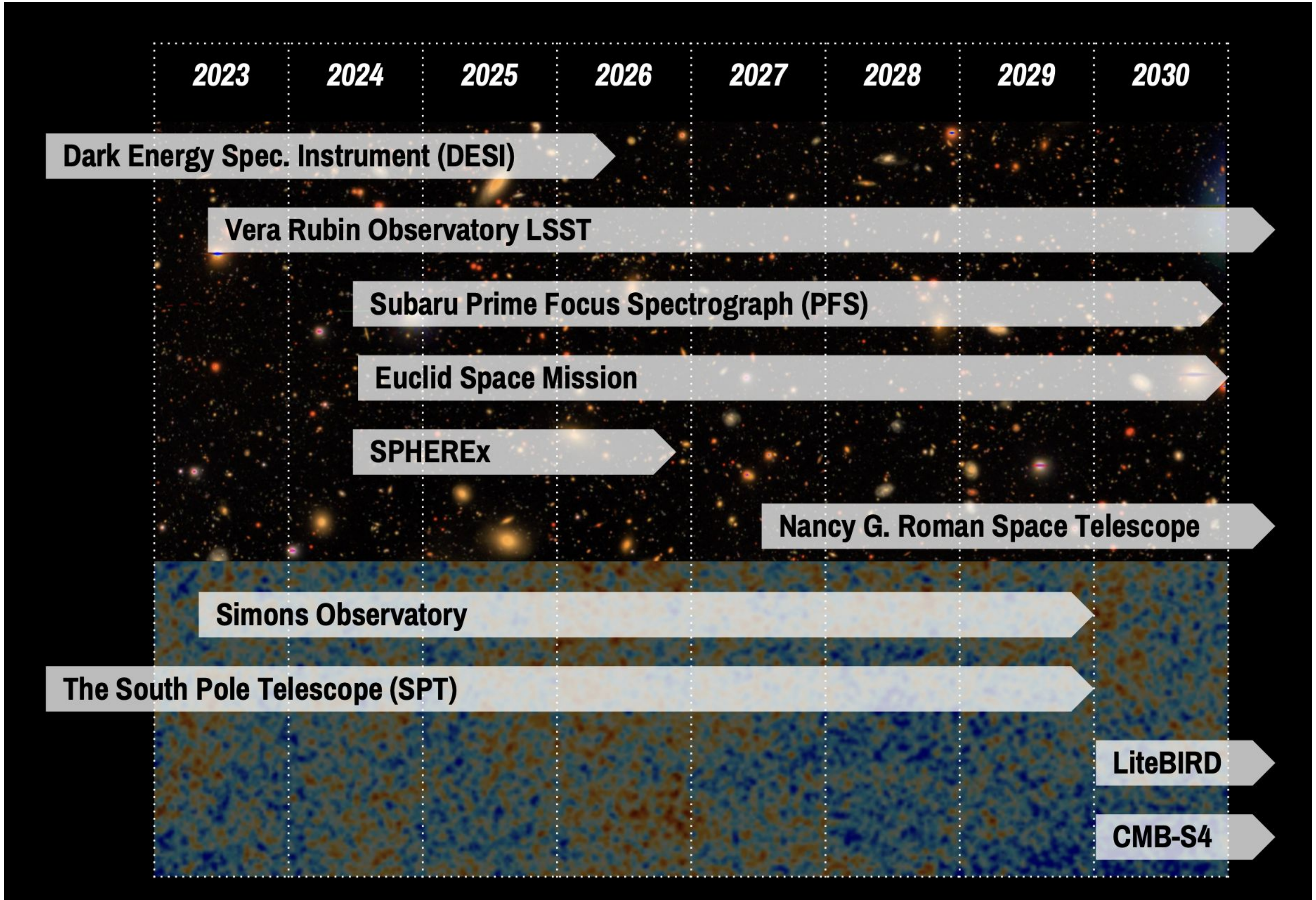
What is our current knowledge of cosmology?

- **We have a standard 5 parameter model!** It contains dark matter, baryons and cosmological constant, and the initial conditions from inflation defined by 2 parameters
- There are known unknowns that people are searching for. Example: B modes. It will be a tremendous achievement of experimentalists if these are found
- There may be surprises (unknown unknowns) with a big discovery potential
- There are many claimed anomalies, which if real would be pointing to a big discovery, but some are almost certainly **noise or systematics**
- Example: Hubble tension, sigma8 tension, dark energy tension





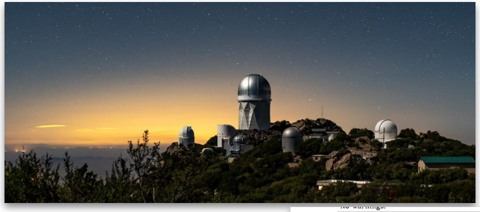
We are in the golden era of cosmology: huge investments in new surveys



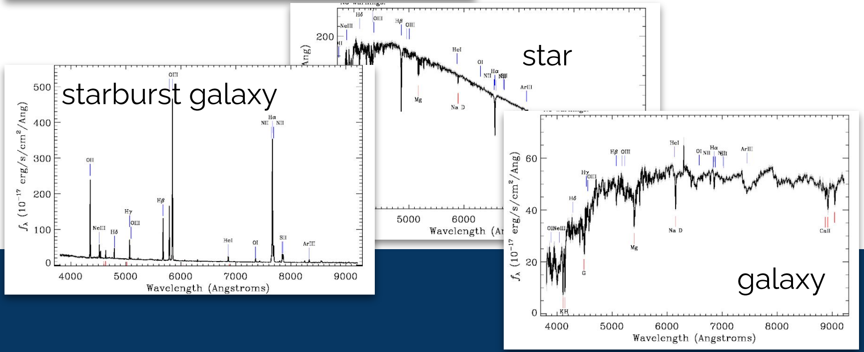
Possibly out of date figure

New surveys

Dark Energy Spectroscopic Instrument (DESI)



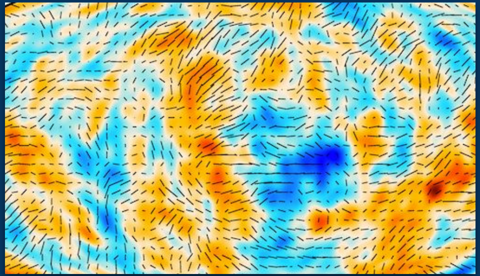
40 million galaxies
10 million stars



Vera Rubin Observatory (LSST)

20 billion galaxies
17 billion stars
20 terabyte data/day

CMB Simons Observatory

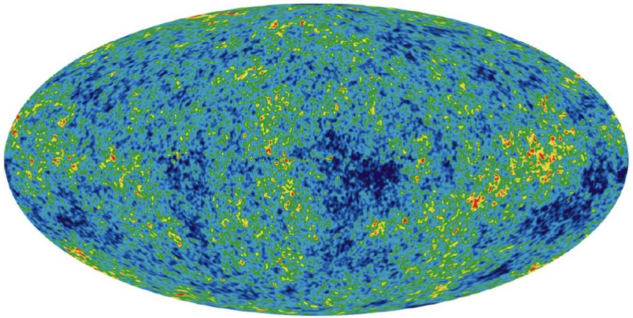


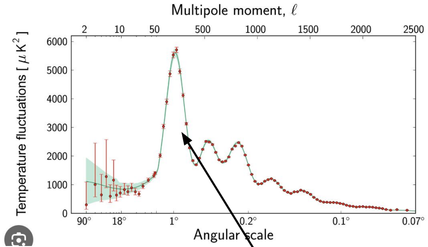
Cosmic Microwave Background

Current and optimal cosmological data analysis

- **Current analysis: power spectrum or correlation function**

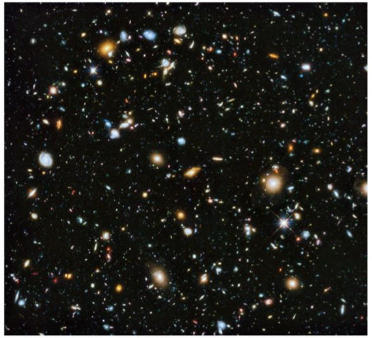
Gaussian density field
Fully described by the
power spectrum

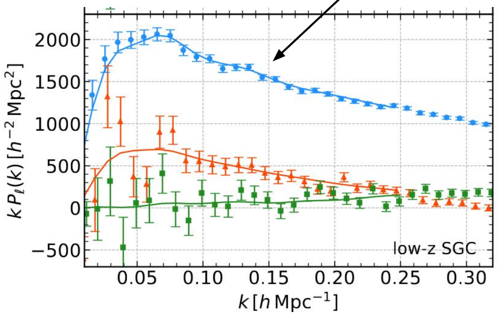




Baryonic acoustic oscillations (BAO)

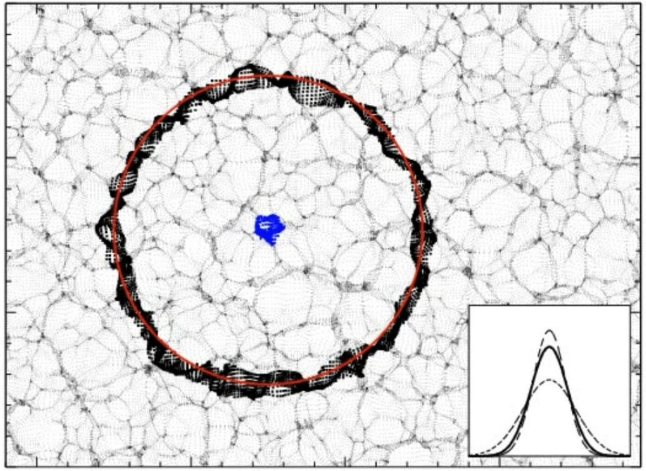
Non-Gaussian density field
P(k), B(k), peaks, voids, …





Current and optimal cosmological data analysis

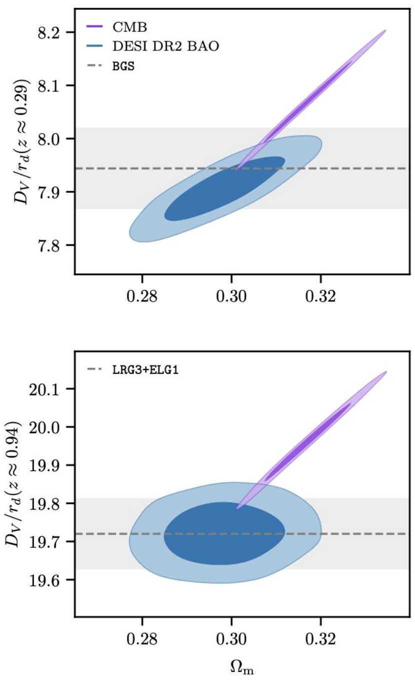
- **Current analysis: power spectrum or correlation function**
- **BAO reconstruction goes beyond P(k): partially removes BAO damping**
- **Can we do better?**
- **Optimal:** Extract **maximum** amount of information from the data

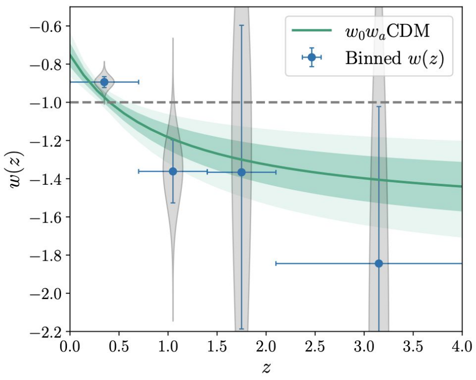


Padmanabhan etal 2012

State of DESI BAO

- Recent DESI BAO results suggest 2.5-3 sigma evidence for new physics, possibly time dependent dark energy
- This evidence comes from discrepancy between DESI and Planck
- This is an extraordinary claim, if true
- This is based on 3 years of data: full survey is 5 years, so BAO errors will decrease by 25%
- Can we do better?



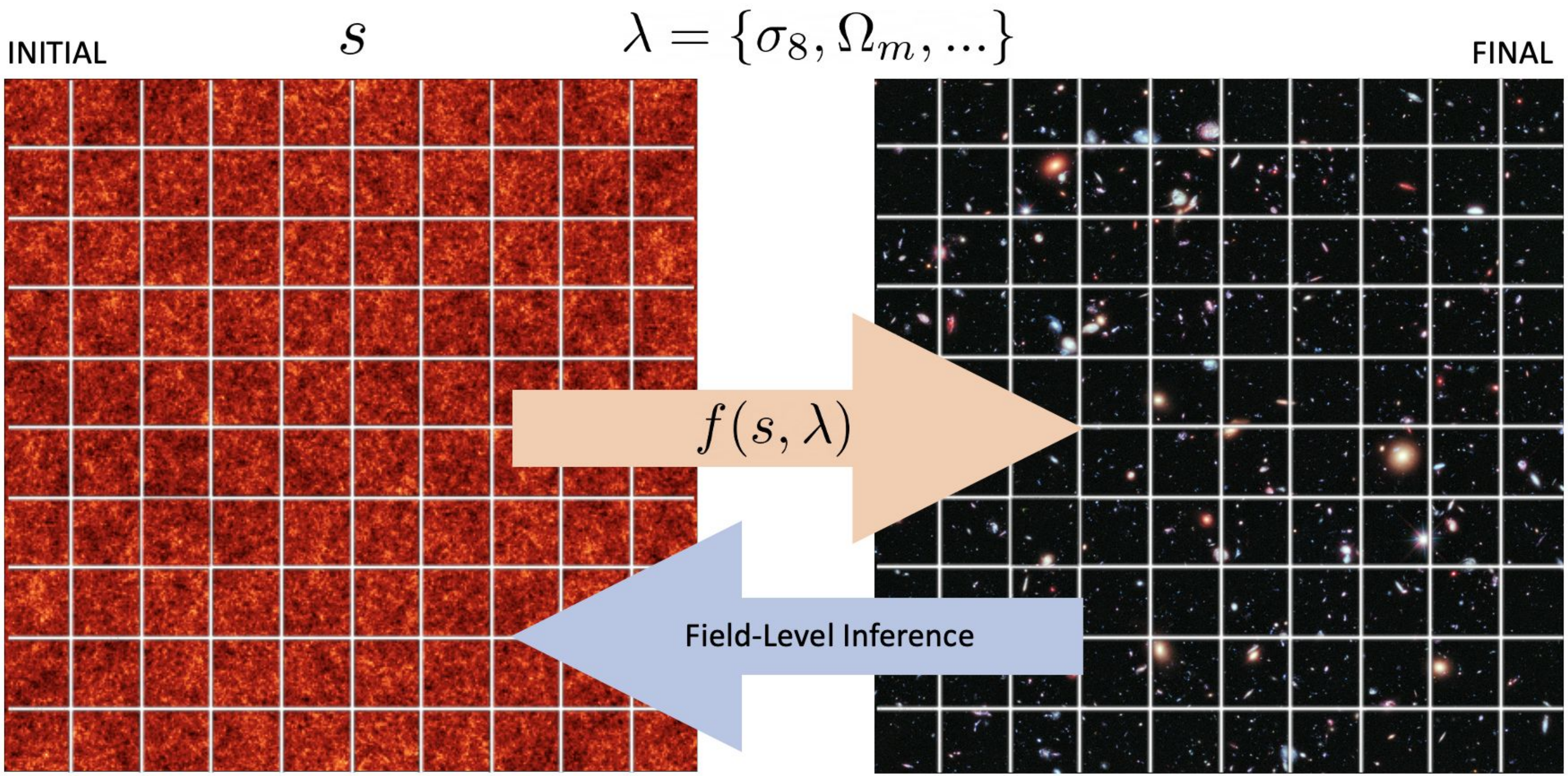


Part 1:

Baryonic Acoustic Oscillations with Field level Inference

Work led by Adrian Bayer and Liam Parker (based on previous work by Chirag Modi)

Field level inference with a forward model: hierarchical Bayesian analysis



Kitaura & Ensslin (2007), Jasche & Wandelt (2012), Jasche & Lavaux (2017), Seljak et al. (2017), Schmidt et al. (2020), ...

Field level inference with a forward model: hierarchical Bayesian analysis

Given field data d and forward model f infer

initial modes s and cosmological parameters λ

$$-2\log P(s,\lambda|d) = \sum_{\vec{k}} \left[\frac{|d - f(s,\lambda)|^2}{N} + \frac{|s|^2}{\mathcal{P}(\lambda)}\right]_{\vec{k}}$$

posterior likelihood prior

CHALLENGE: Multimillion dimensional parameter space!

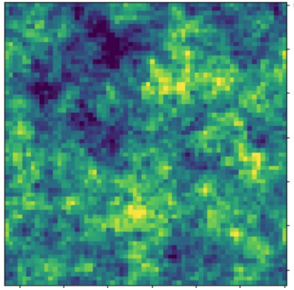
1. Need fast forward model

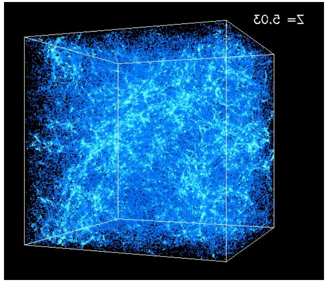
2. Need differentiable forward model

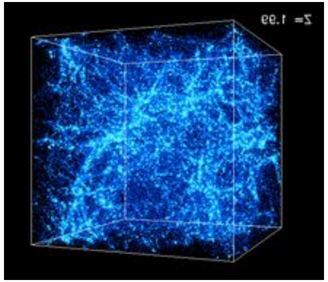
3. Need fast sampler

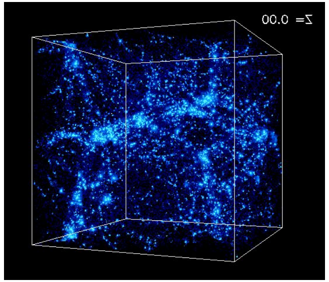
Generating fast forward model with N-body simulations

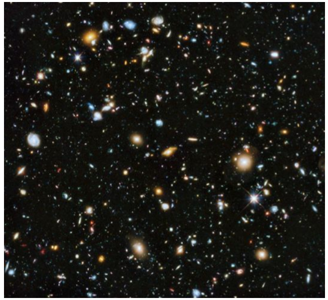
cosmological structure formation: an N-body simulation and a galaxy formation model
A lot of progress in recent years with **differentiable fast** simulations (**FastPM, FlowPM**, **PMWD**…)











source: http://cosmicweb.uchicago.edu/filaments.html

random initial fluctuations **structure formation** **galaxies**

random realization is independent from cosmological parameters

depend on cosmological parameters

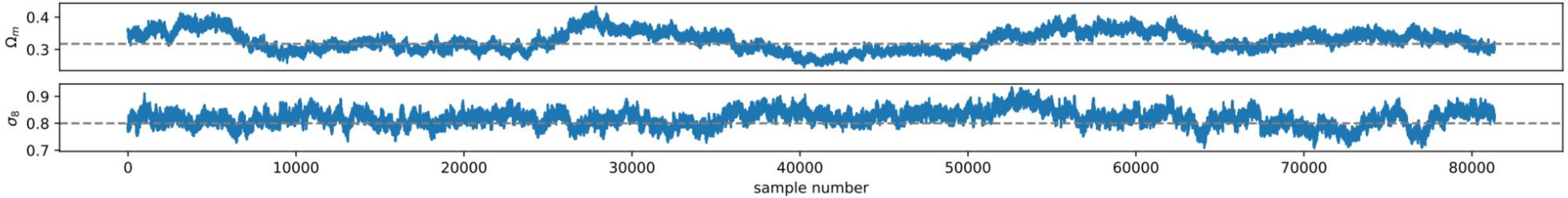
We can now run hundred million particles in a matter of seconds

Major problem in high dimensional MCMC sampling: correlation between samples

MCMC sampling for cosmology is very high dimensional (we would like to scale to 10^9 initial modes or more)

In cosmology it has been pursued using Hamiltonian Monte Carlo for initial conditions (Jasche & Wandelt 2013, Lavaux etal 2019, Porqueres etal 2022)

Example for d=500,000 (Porqueres etal 2023): correlation lengths of order 5-10k, expected to grow to 50-100k for d=5x10^9



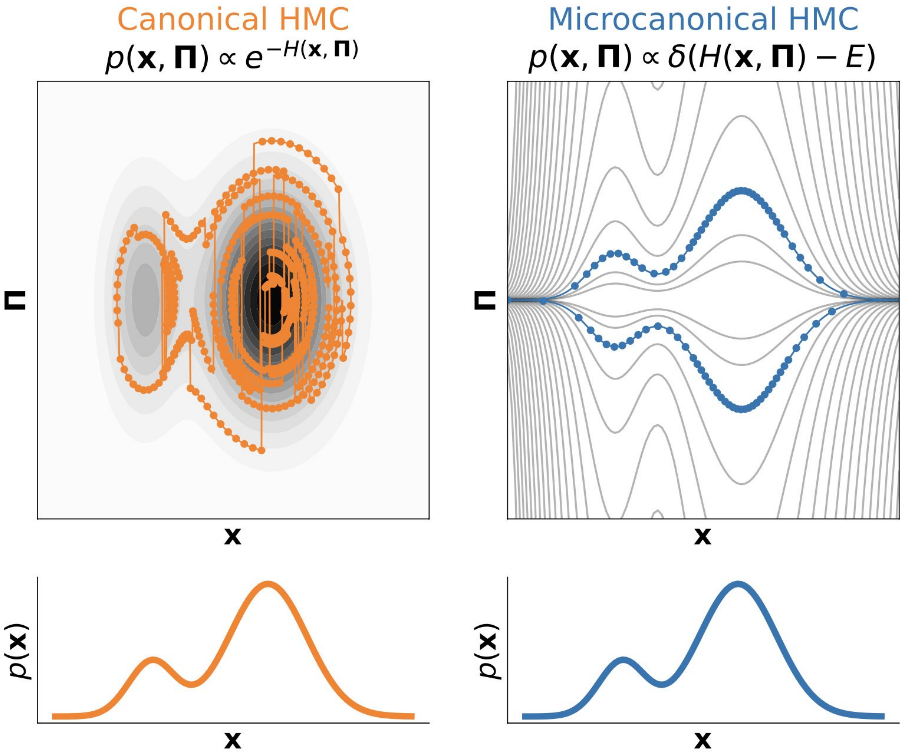
In HMC each sample requires 10-50 likelihood calls, so we expect of order 10^6 forward simulations for each independent sample. We need at least 100 samples. This is not currently feasible

This problem took us on a long and very fruitful sidetrack: creation of a new gradient based sampler

MicroCanonical Hamiltonian Monte Carlo

J. Robnik, B. de Luca, E. Silverstein, US (2022)

$$\frac{\|\mathbf{\Pi}\|^2}{2} - \log p_{\text{target}}(\boldsymbol{x})$$



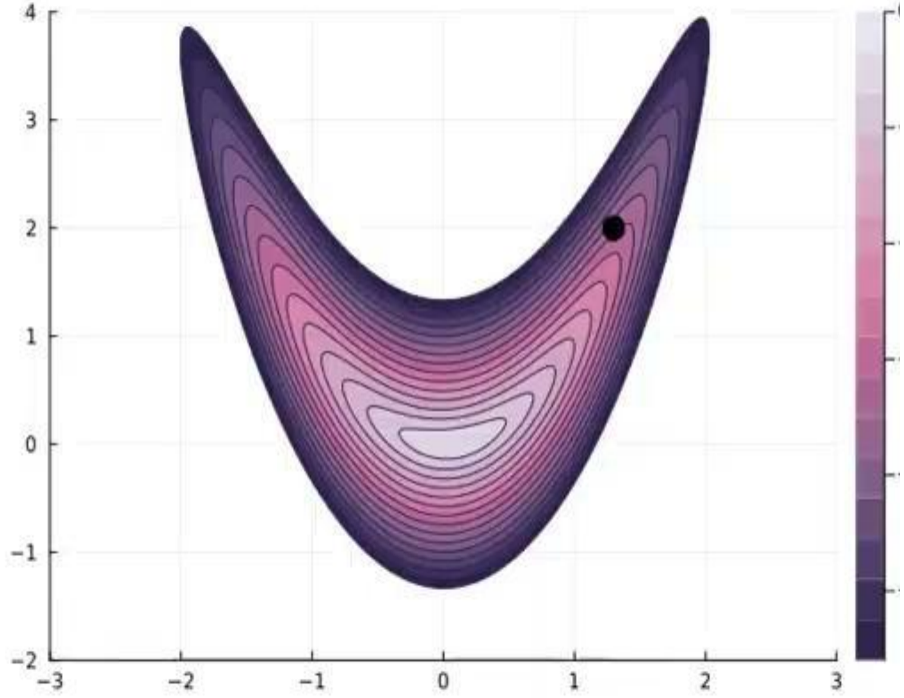
$$\longrightarrow \quad H = \frac{d}{2} \log \frac{\|\mathbf{\Pi}\|^2}{d} - \log p(\boldsymbol{x})$$

$$\Updownarrow$$

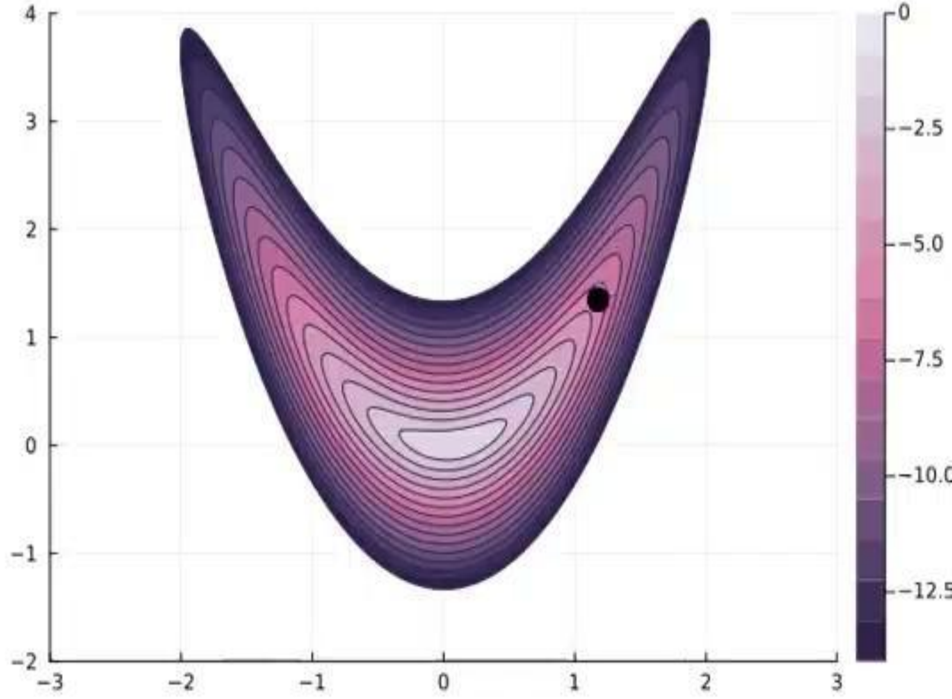
$$H = \frac{\|\mathbf{\Pi}\|^2}{2m(\mathbf{x})}$$

particle moves slowly where the target density is high

HMC MCHMC



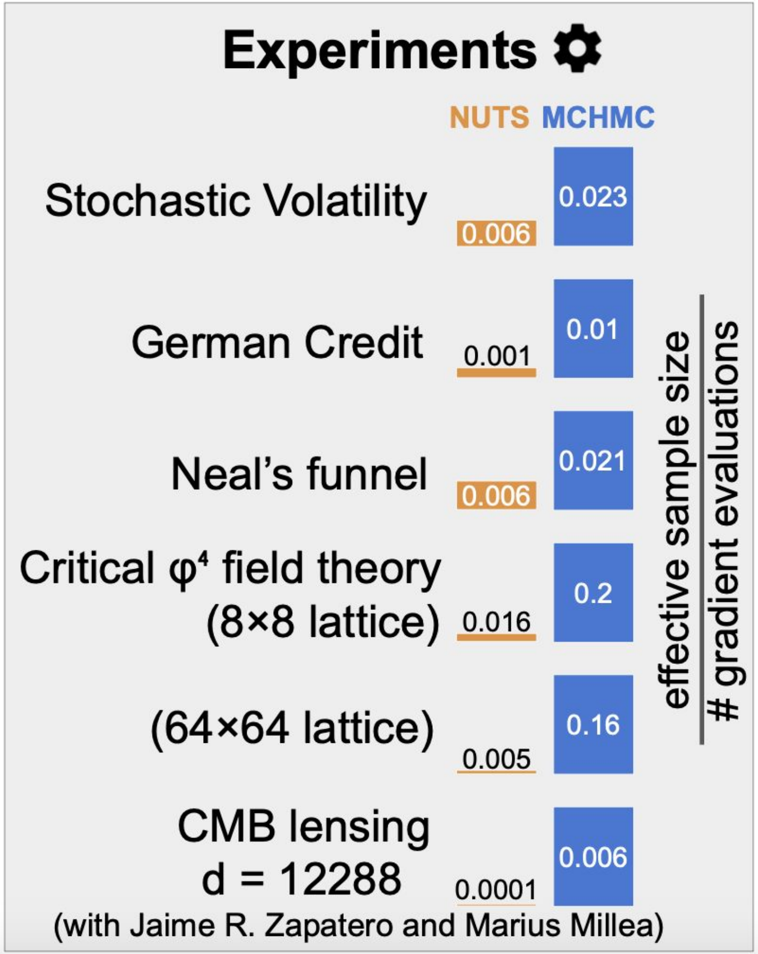
HMC: standard Hamiltonian dynamics, fast through the high density, slow in low density



MCHMC: fast through low density

- Severalfold faster sampling in low dimensions
- **Orders of magnitude** faster sampling in **high dimensions**
- Automatic **hyperparameter tuning**
- Extremely **fast burn-in**
- Provable convergence
- Python (JAX), Julia and C++ implementations

https://github.com/JakobRobnik/MicroCanonicalHMC

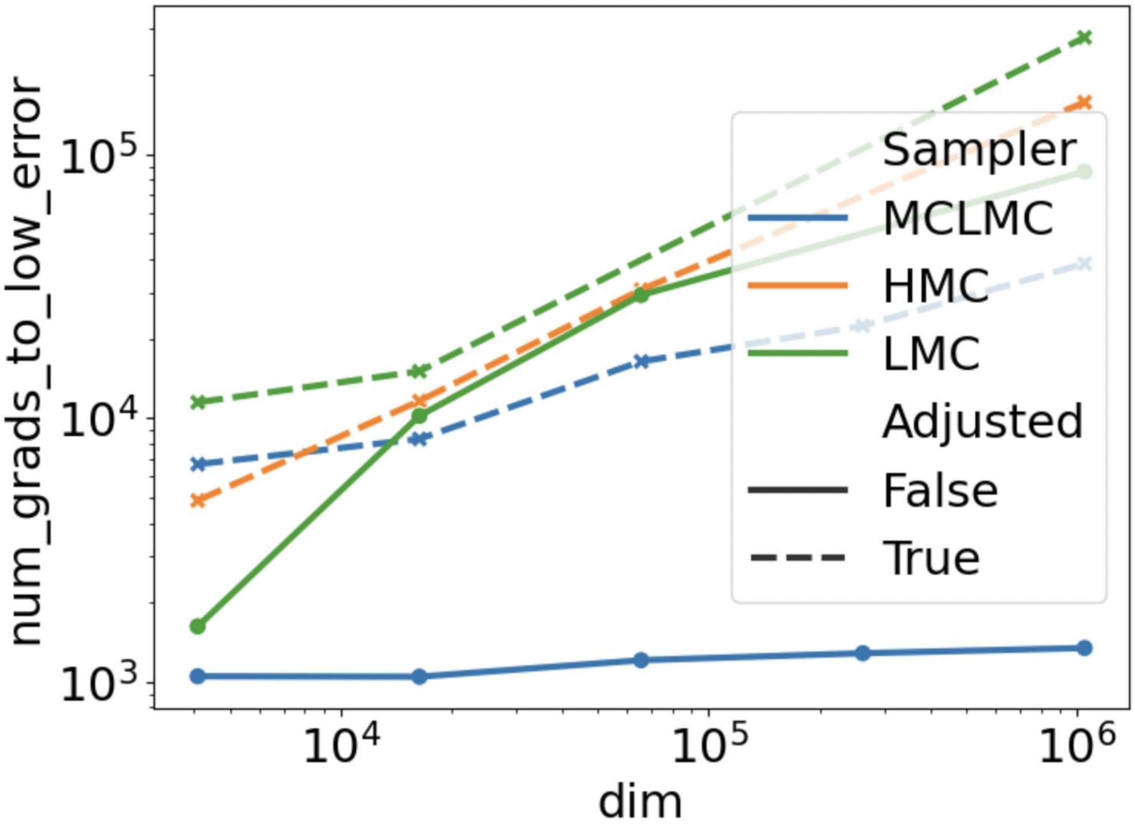


Recent developments in MicroCanonical Dynamics

- Microcanonical Langevin Monte Carlo: continuous noise (Robnik & US 2024)
- Metropolis Adjusted MCLMC/MCHMC: asymptotically unbiased (Robnik, Cohn-Gordon & US 2025)
- Ensemble MCLMC: EMAUS (Robnik & US 2025) is 2-3 orders of magnitude faster than NUTS
- All of the MCLMC/MCHMC papers are published in AI/ML conferences/journals

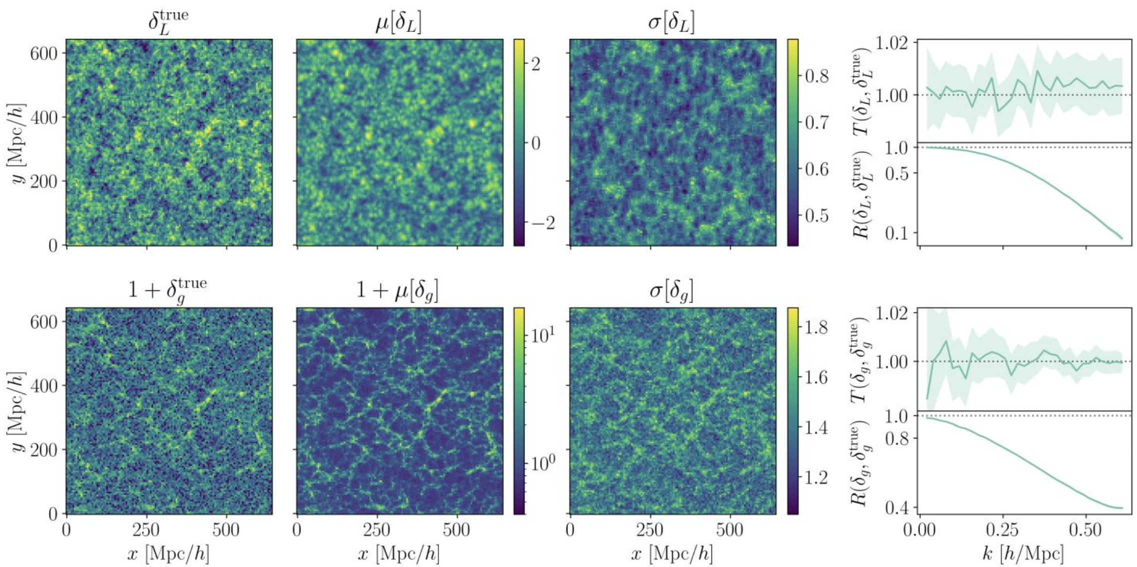
	NUTS (SEQUENTIAL)	MCLMC (SEQUENTIAL)	EMAUS (TOTAL)	EMAUS (PER CHAIN)
BANANA	64200	5000	4400	17
ILL CONDITIONED GAUSSIAN	113900	33500	68600	270
SPARSE LOGISTIC REGRESSION	10400	4200	28700	110
BROWNIAN MOTION	5400	1700	17900	70
ITEM RESPONSE THEORY	7300	1600	21000	80
STOCHASTIC VOLATILITY	29900	10000	281600	1100

High dimensional applications: MCLMC vs HMC/LMC



Cohn-Gordon, Robnik, US 2025

Back to cosmology: Initial condition reconstruction



Simon-Onfroy, Lanusse, de Mattia 2025, IC reconstruction on 64^3 mesh

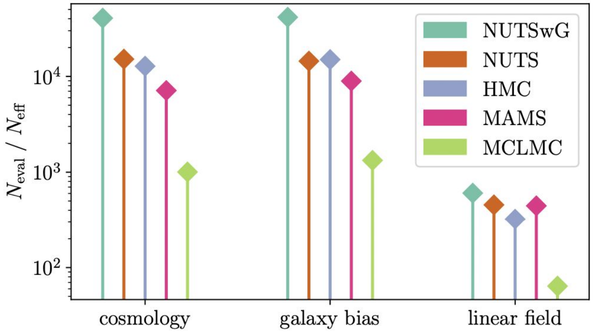
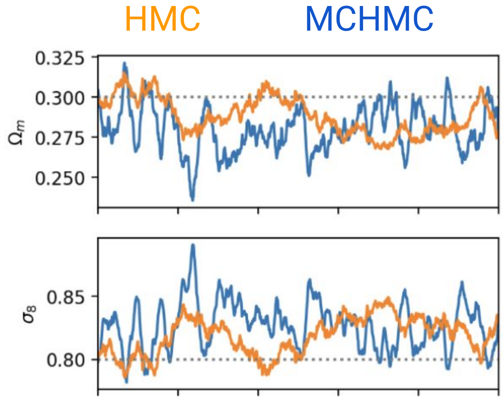


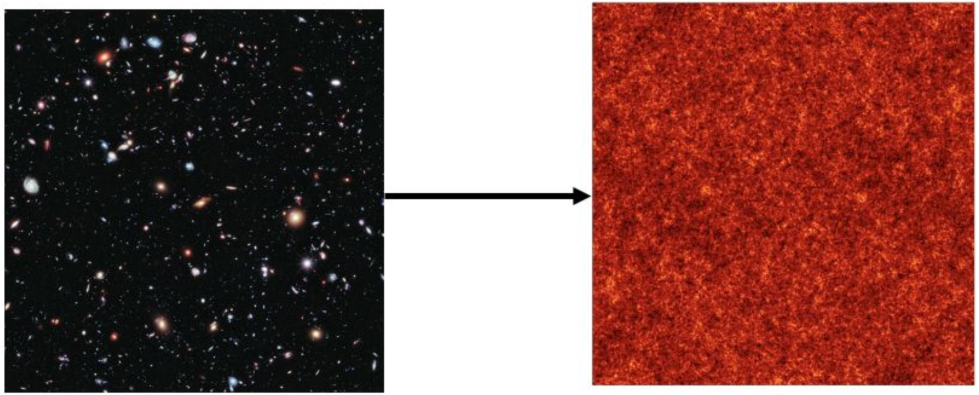
Figure 5: Number of model evaluations per effective sample for each sampler and each set of parameters (cosmology, galaxy bias and linear field).

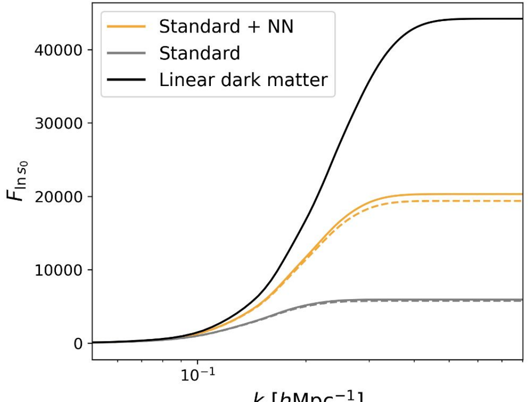
Back to cosmology: optimal BAO reconstruction

- We have developed two methods:
- Explicit Bayesian sampling of initial conditions (US et al 2017, Modi etal 2018, Bayer et al 2024): MCLMC has significantly shorter correlations
- CNN based training from standard reconstruction to initial conditions (Parker, Bayer & US 2025)

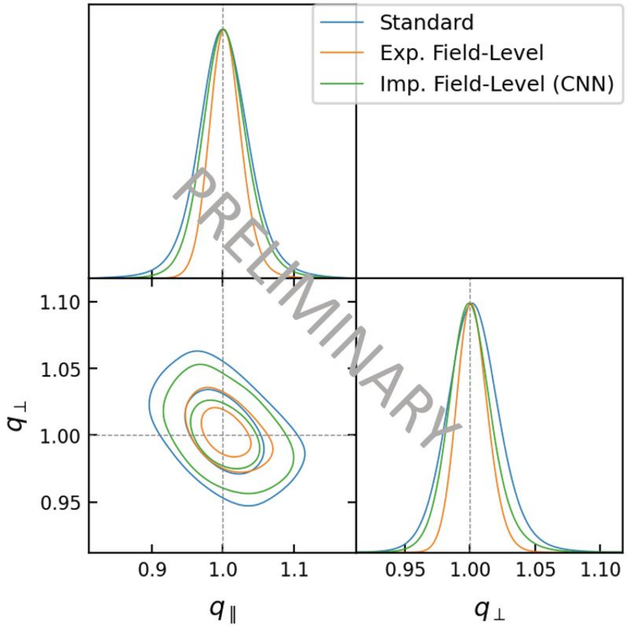


CNN





End to end BAO test on simulations (Bayer et al, in prep)



Test on simulations

Potentially we can improve the Figure of Merit by up to a factor of 2 (equivalent to 2 times the volume)!

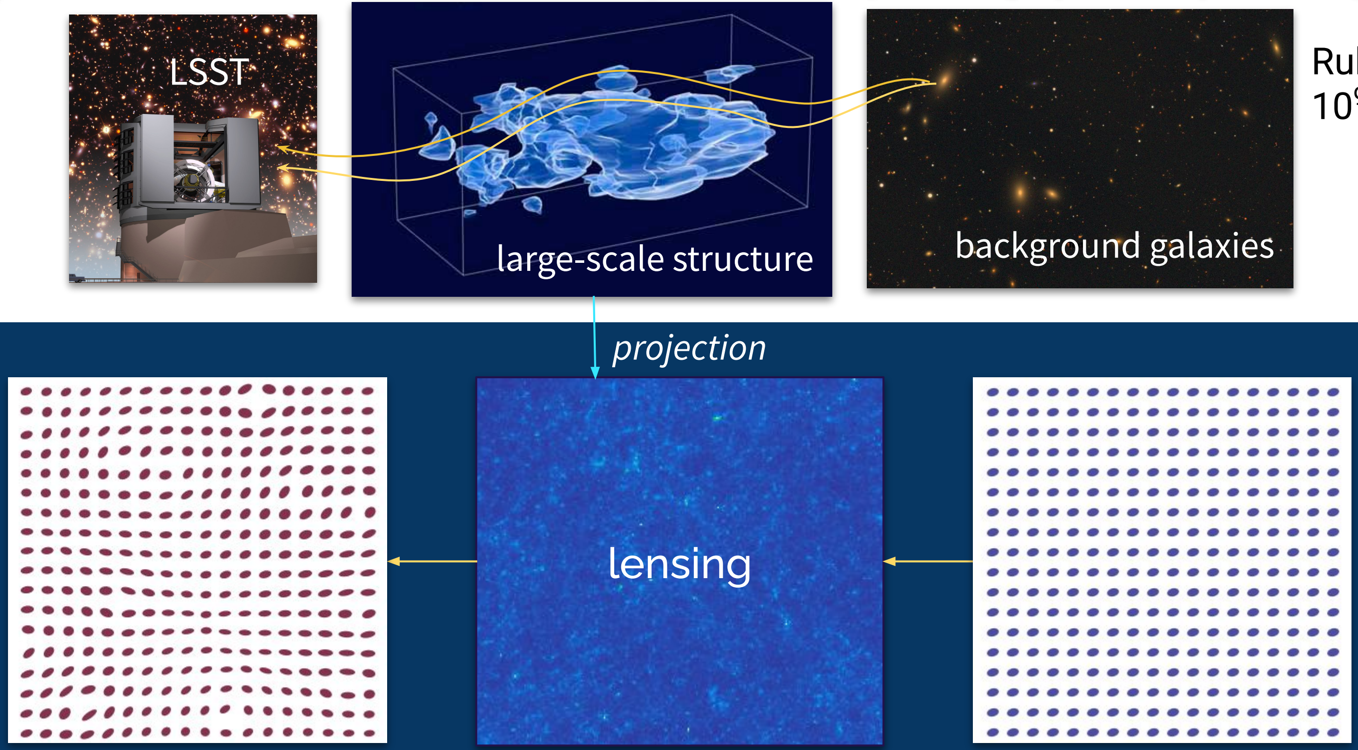
We will be able to test DESI results: 3 sigma can become 4 sigma (or 2 sigma if the central value changes)

Robust and Optimal Analysis of Weak Gravitational Lensing with Normalizing Flows

Work led by Biwei Dai

Weak Lensing of Galaxies

Current WL P(k) constraints: low sigma8, recently moving up due to photoz (eg KIDS)



Rubin data size:
10^9 galaxies

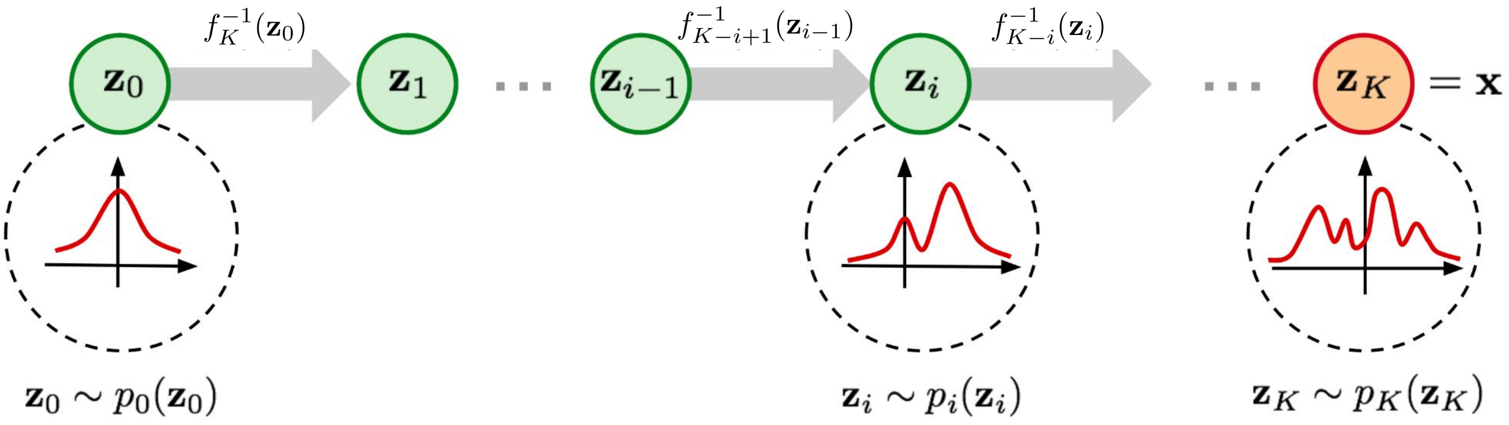
Optimal cosmological data analysis with AI/ML

- **Optimal:** Extract **maximum** amount of information from the data

- **Supervised learning with AI/ML** to predict posterior $p(y|x)$: discriminative
- **Learning the likelihood $p(x|y)$ directly with AI/ML: generative**
  - Con: High dimensionality of x.
  - Model: Normalizing Flows

Use symmetries and multiscale architecture to reduce the curse of dimensionality

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

What can Normalizing Flows do for Astronomy?

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

- Extract physical information (simulation-based inference)
- Fast sample generation
- Anomaly detection
  - Detect systematic effect (distribution shift)
  - Search for new physics/astrophysics

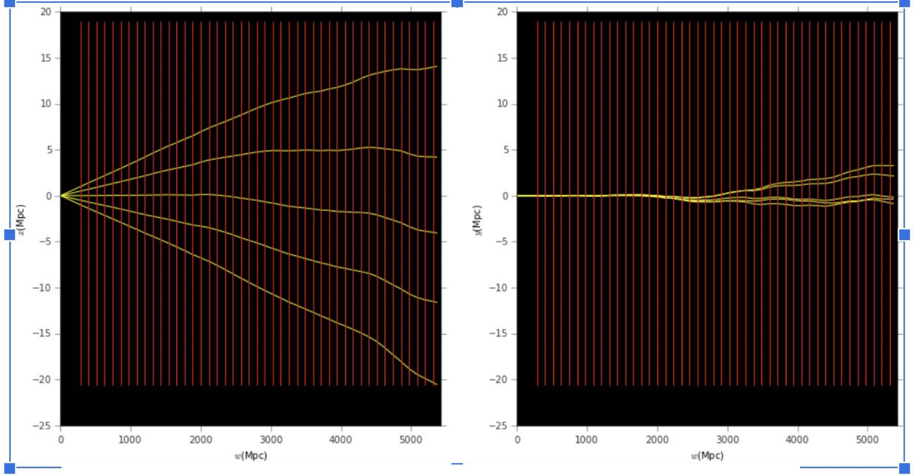
Cosmology specific considerations: translation/rotation symmetries, correlations on all scales

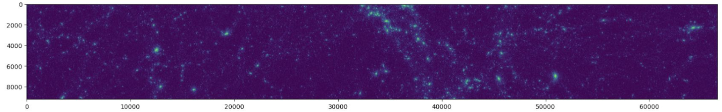
1) MultiScale Flow (Dai & Seljak 2023)
2) Translational and Rotational Normalizing Flow (Dai & Seljak 2021)

Creating training data from fast simulations (FastPM)

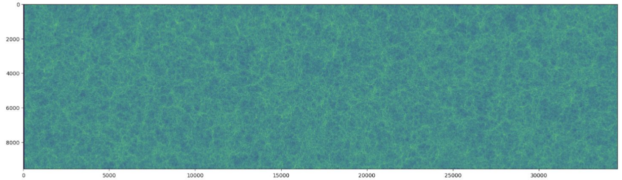
Ray-tracing

- multi–lens–plane algorithm from LensTools package



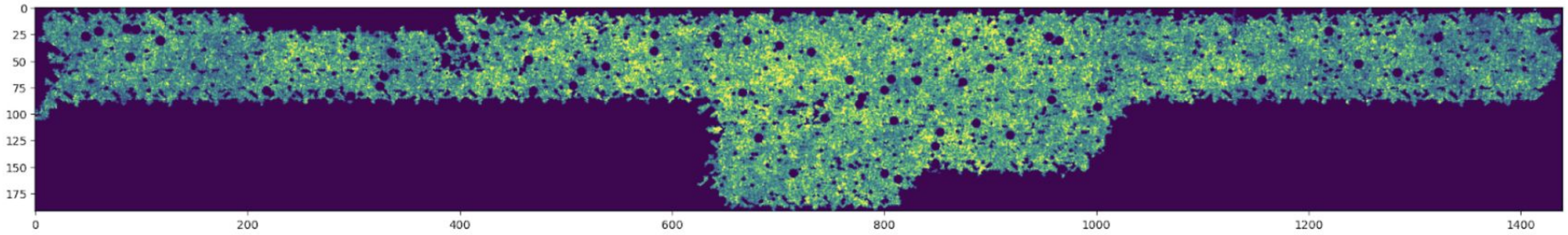


Another example lensing plane at higher redshift:



Stack all the lens planes, add survey mask

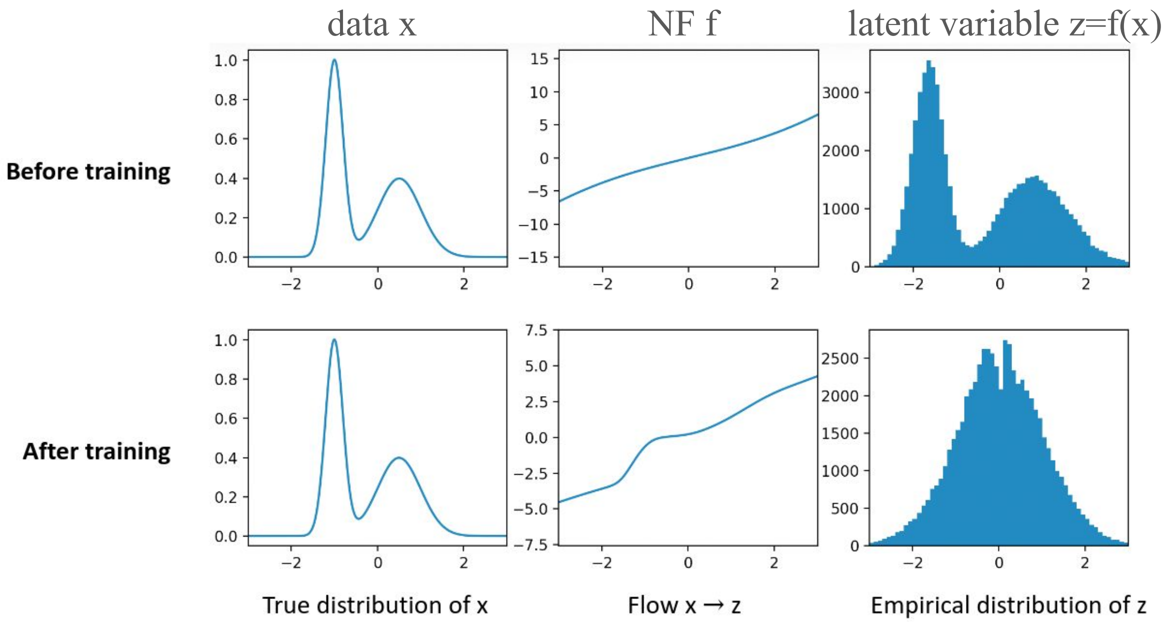
HypersupremeCam (HSC) data on Subaru
Later Rubin (LSST) etc.
Real world complications easy to add: survey mask, noise etc.



We now have realizations of data x
For each simulation we can make many maps by making different projections
Repeat the process for different values of cosmological parameters y
y: we can vary initial density amplitude, matter density, dark energy, neutrino mass etc.

Normalizing Flow training

- 1D example



- **Training objective**: maximize **log p(x) over training data**

  $p(x) = \pi(f(x))\,|\det(df/dx)|$

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

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

How to design Normalizing Flows for physics?

NF by itself requires:

- Easy evaluation of the inverse
- Easy evaluation of the Jacobian determinant

Physics applications require:

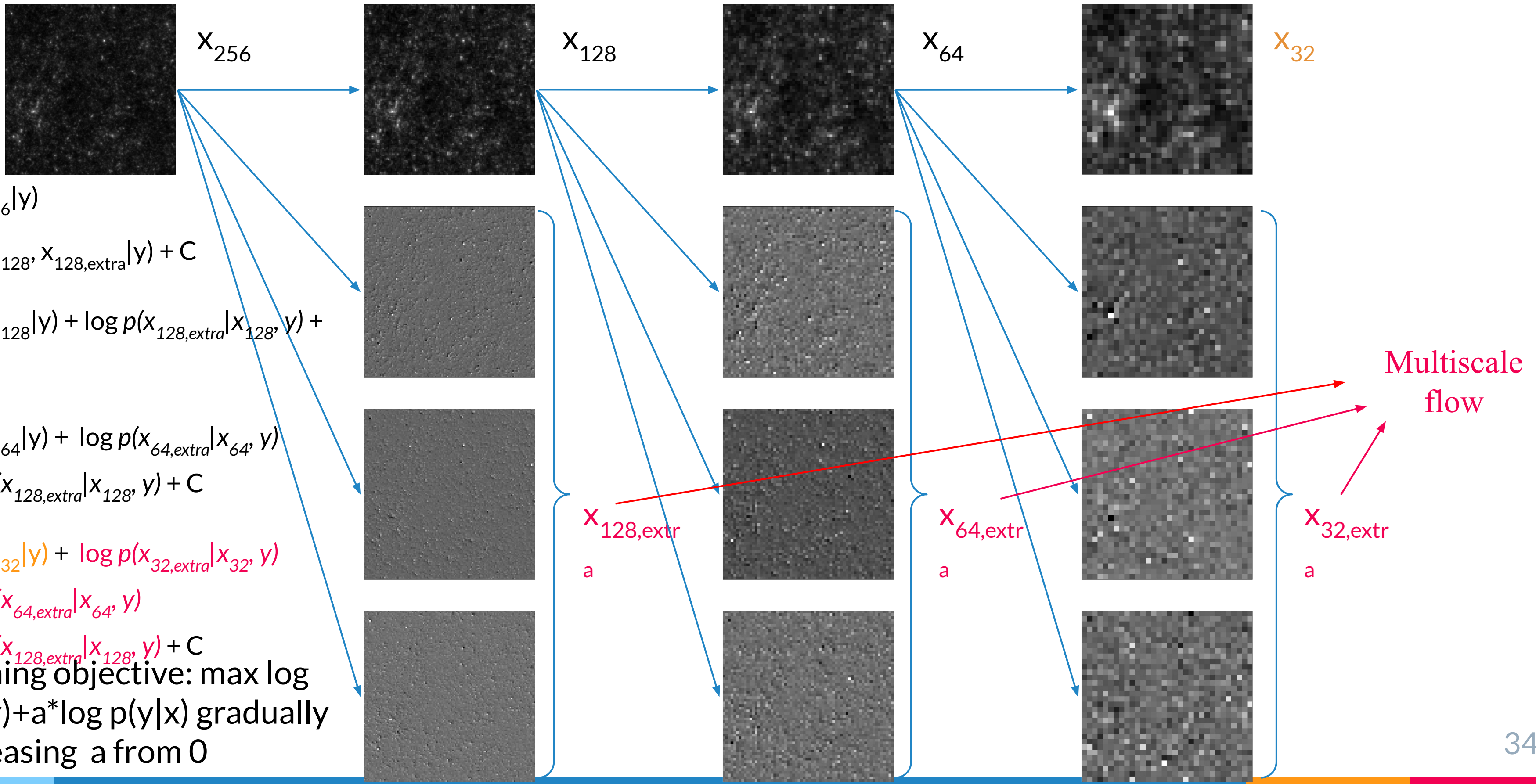
- Spatial Structure modeling, high dimensionality (pixels, voxels): coarse graining, multi-scale correlations
- Symmetries (translation, rotation etc.)
- Ability to learn from very stochastic data

How do we parametrize our NF, given these considerations?

1) MultiScale Flow (Dai & Seljak 2023)
2) Translational and Rotational Normalizing Flow (Dai & Seljak 2021)

Multiscale flow (MSF): a wavelet based Normalizing flow (Dai & US, 2023)

- Consider a cosmological field with 256^2 resolution:



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

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

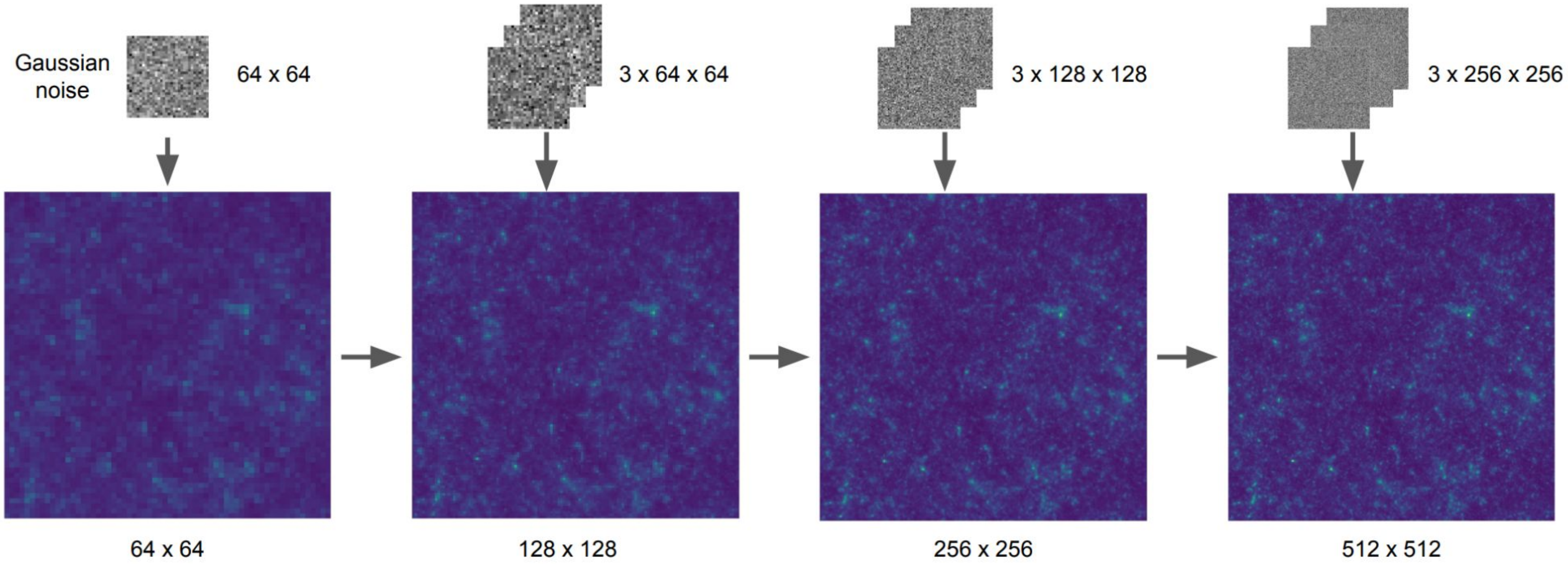
$= \log p(x_{128}|y) + \log p(x_{128,extra}|x_{128}, y) + C$

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

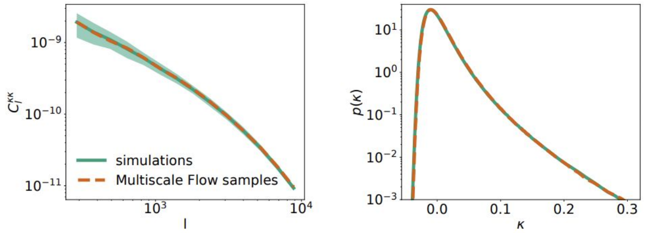
$= \log p(x_{32}|y) + \log p(x_{32,extra}|x_{32}, y) + \log p(x_{64,extra}|x_{64}, y) + \log p(x_{128,extra}|x_{128}, y) + C$

Training objective: max log p(x|y)+a*log p(y|x) gradually increasing a from 0

MSF: A fast and accurate simulator of WL maps



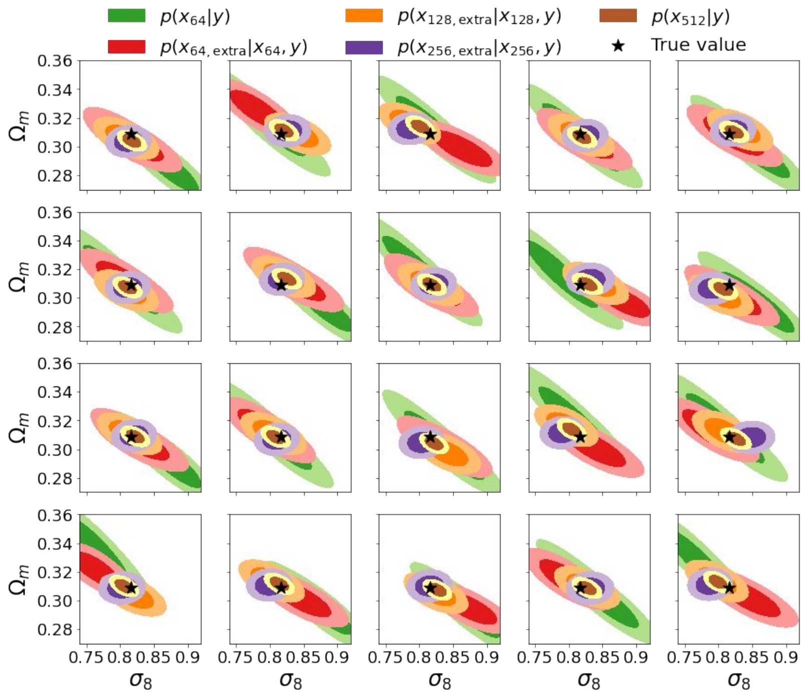
Perfect agreement with real simulations on the power spectrum and one point distribution



Notice how stochastic the data is

MSF: a Reliable Uncertainty Quantification with NF likelihood

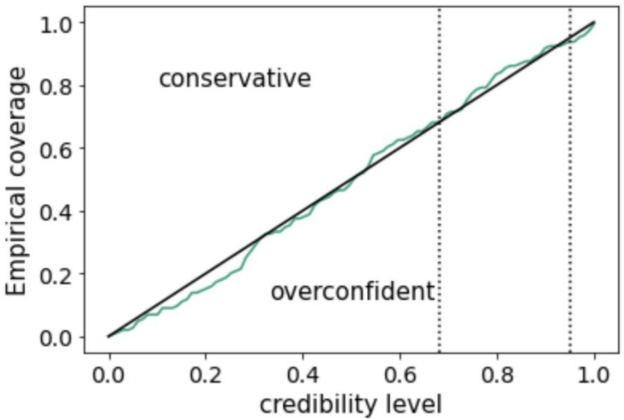
- Consistent posteriors from different scales



Posterior p(y|x)=p(x|y)p(y)/p(x)
prior p(y) is assumed to be flat here

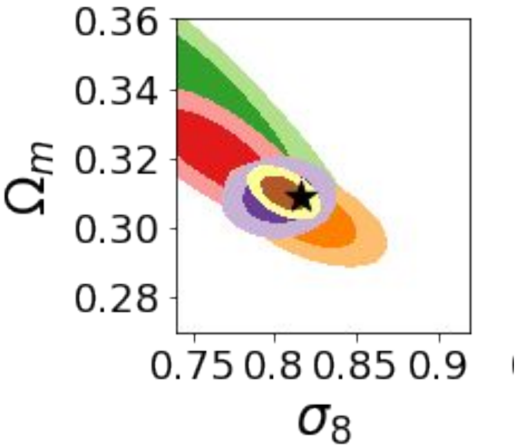
On simulated data the errors are properly calibrated and in agreement with frequentist notions of error quantification (68/95% of true simulation values are within 68/95% posterior contours)

Empirical coverage of Multiscale Flow posterior:

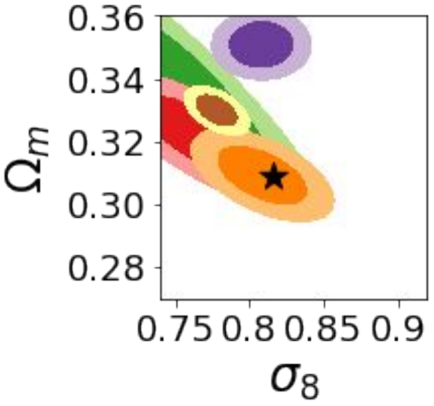


Robustness — how to identify unknown unknowns (anomalies)?
1) scale dependence of unknown systematic effects on WL maps

- Consistent posteriors from different scales
- Inconsistent small scale posterior



unknown
systematic
effect
(gaussian
noise)



One of the few ways to identify unknown unknowns!

2) anomaly detection with density estimation (Diao, Dai, US 2025)

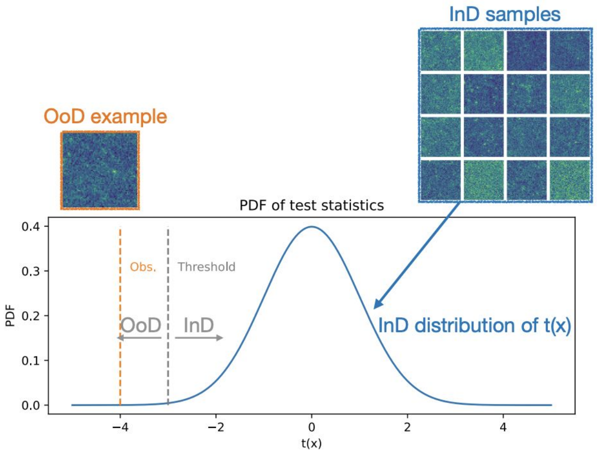
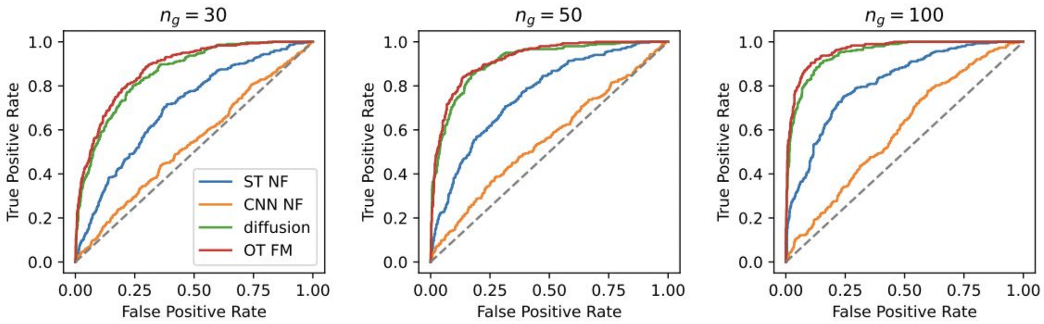


Figure 1. An illustration of OoD detection pipeline. A probability distribution function (PDF) of InD sample test statistics is calculated and shown in blue solid line, and an empirical threshold in grey dashed line is chosen to identify an OoD sample with $t(\boldsymbol{x})$ smaller than the threshold, shown in orange dashed line.



Method	$n_g = 30$	$n_g = 50$	$n_g = 100$
Diffusion	0.85	0.90	0.94
OT FlowMatching	0.87	0.92	0.95
MultiscaleFlow[38] [4]	$\gtrsim 0.65$	–	–
CNN	0.54	0.55	0.60
ST coefficient	0.73	0.77	0.81

Two independent ways to identify unknown unknowns!

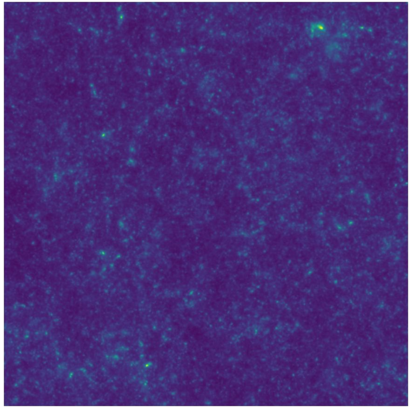
Interpretability

“Where is the extra information coming from?”

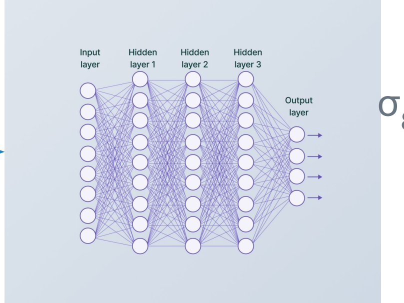
“Why are the other cosmological models ruled out?”

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

Input WL map



Generative models

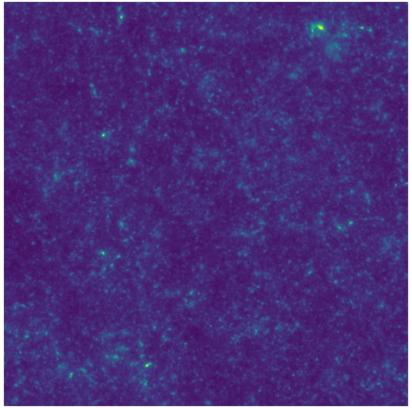


MCMC

$\sigma_8 = 0.76 \pm 0.02$

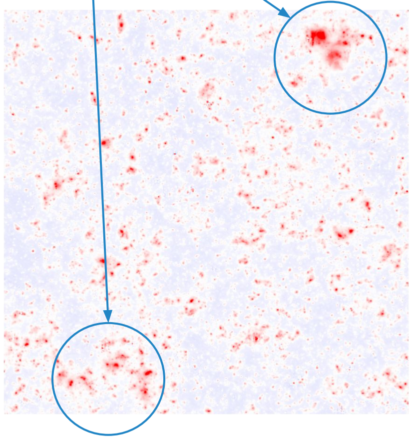
$\sigma_8 = 0.816$

Generated sample



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

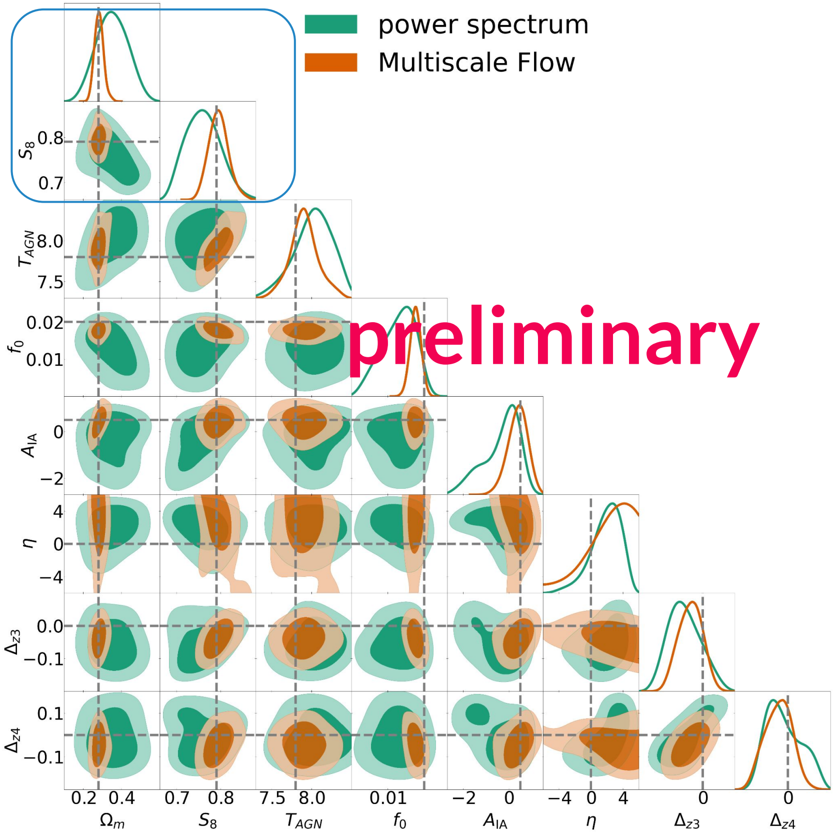
Difference



Generated sample - input map

Multiscale Flow versus power spectrum for HSC

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
- **Stay tuned for upcoming HSC analysis**

We have WL Neurips 2025 data challenge!

Led B. Dai and P-W Chang

Weak Lensing ML Uncertainty Competition

FAIR Universe

Overview

The competition tasks will be structured into **2 phases**:

- **Phase 1: Cosmological Parameter Estimation**

  Participants will develop models that:

  - Accurately infer cosmological parameters (Ω_m, S_8) from the weak lensing image data.
  - Quantify uncertainties via the 68% confidence intervals of the parameters of interest. $(\hat{\sigma}_{\Omega_m}, \hat{\sigma}_{S_8})$.

Scoring metrics:

KL divergence between the true Gaussian-like posterior distribution and the Gaussian with the predicted mean and standard deviation:

$$\text{score}_{\text{ inference}} = -\frac{1}{N_{\text{test}}}\sum_{i=0}^{N_{\text{test}}}\left\{\frac{\left(\hat{\Omega}_m - \Omega_m^{\text{truth}}\right)^2}{\hat{\sigma}^2_{\Omega_m}} + \frac{\left(\hat{S}_8 - S_8^{\text{truth}}\right)^2}{\hat{\sigma}^2_{S_8}} + \log\left(\hat{\sigma}^2_{\Omega_m}\right) + \log\left(\hat{\sigma}^2_{S_8}\right) + \lambda\left[\left(\hat{\Omega}_m - \Omega_m^{\text{truth}}\right)^2 + \left(\hat{S}_8 - S_8^{\text{truth}}\right)^2\right]\right\}$$

$\lambda \equiv 10^3$: penalty factor for bad point estimates

- **Phase 2: Out-of-Distribution Detection**

  Some test data will be generated with different physical models (OoD), leading to some distribution shifts with respect to the test data in Phase 1. Participants will develop models that:

  - Identify test data samples inconsistent with the training distribution (OoD detection).
  - Provide probability estimates indicating data conformity to training distributions.

Binary cross-entropy:

$$\text{score}_{\text{OoD}} = \frac{1}{N_{\text{test}}}\sum_{i=0}^{N_{\text{test}}}\left(y_i \log\left(p_i\right) + \left(1 - y_i\right)\log\left(1 - p_i\right)\right)$$

where $y_i = 1$ if the dataset is InD; $y_i = 0$ if the dataset is OoD, and p_i is the probability estimates predicted by the model.

Weak Lensing ML Uncertainty Competition

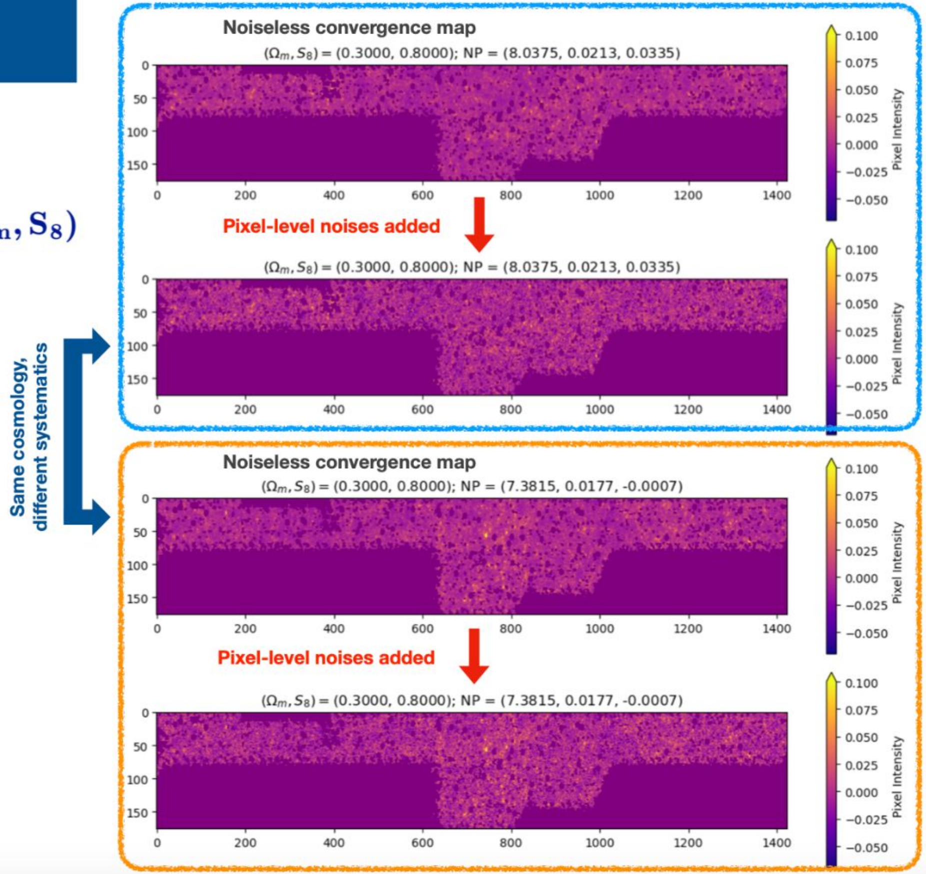
FAIR Universe

Overview: Dataset

- Mock galaxy catalogs predicted with N-body simulations and lens algorithms at 101 cosmological parameters (Ω_m, S_8)

- Pixelized 2D weak lensing images: convergence maps

- The model must take into account the systematic uncertainties from **4 realistic systematic effects**:
  (1) 2 baryonic effect uncertainties
  (2) 1 photometric redshift uncertainty
  (3) Pixel-level noises



Po-Wen Chang (LBNL)

Weak Lensing ML Uncertainty Competition

Overview: Phase 1 Dataset

The participants will be provide with:

- **Public training set:**
  - Image data; shape = (101, 256, 1424, 176)
  - Label shape = (101, 256, 5)

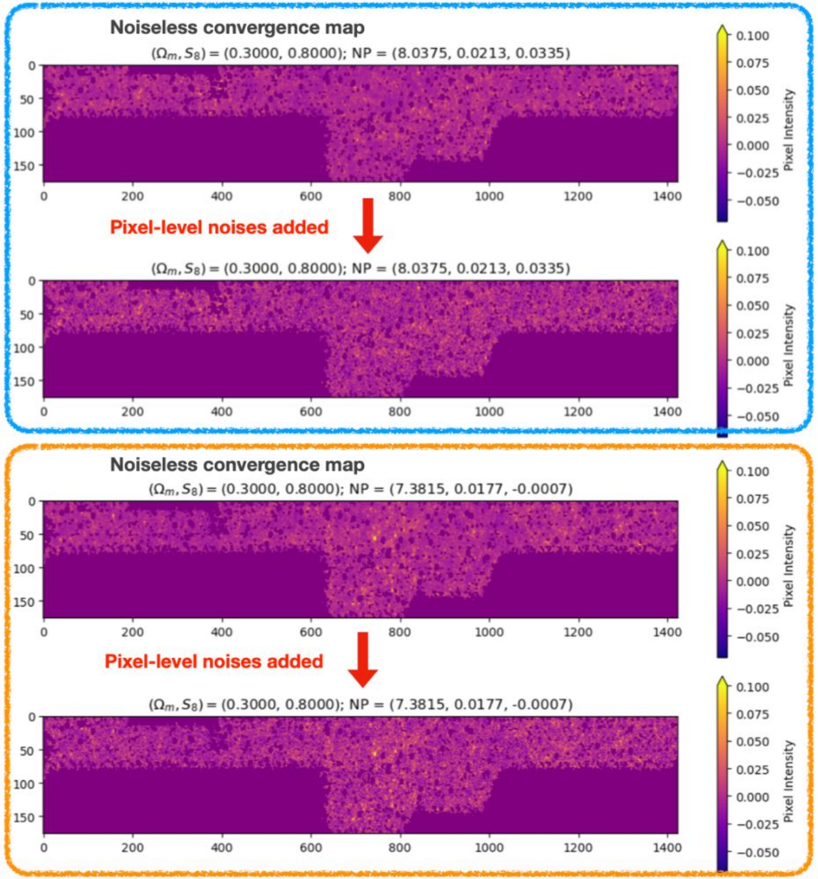
  101 = Realizations of cosmological models; each characterized with 2 parameters of interest $(\mathbf{\Omega_m, S_8})$

  256 = Realizations of 3 nuisance parameters for systematics (1) and (2)

  (1424, 176) = Image dimension

  5 = 2 parameters of interest $(\mathbf{\Omega_m, S_8})$ + 3 nuisance parameters for systematics (1) and (2)

  - **The provided training set is noiseless.** Participants can generate pixel-level noise (3) to augment their training data using a simple function we provide



Po-Wen Chang (LBNL)

Weak Lensing ML Uncertainty Competition

FAIR Universe

Overview: Phase 1 Dataset

The participants will be provide with:

- **Public test set:**
  - Image data; shape = (4000, 1424, 176)

    4000 = Number of test images

    (1424, 176) = Image dimension
  - The test images are generated with random cosmological parameters, random nuisance parameters, and random pixel-level noises.

Phase 1 Evaluation

The true parameters $(\mathbf{\Omega_m^{truth}}, \mathbf{S_8^{truth}})$ of the public test set are unknown to the participants.

Participants will submit predictions of

- **Cosmological parameters** $(\hat{\mathbf{\Omega}}_\mathbf{m}, \hat{\mathbf{S}}_\mathbf{8})$
- **Their uncertainties** $(\hat{\boldsymbol{\sigma}}_{\mathbf{\Omega_m}}, \hat{\boldsymbol{\sigma}}_{\mathbf{S_8}})$

to Codabench, our competition platform.

The model performance will then be evaluated with the hidden ground truth based on our scoring metrics.

Weak Lensing ML Uncertainty Competition

FAIR Universe

Status and timeline

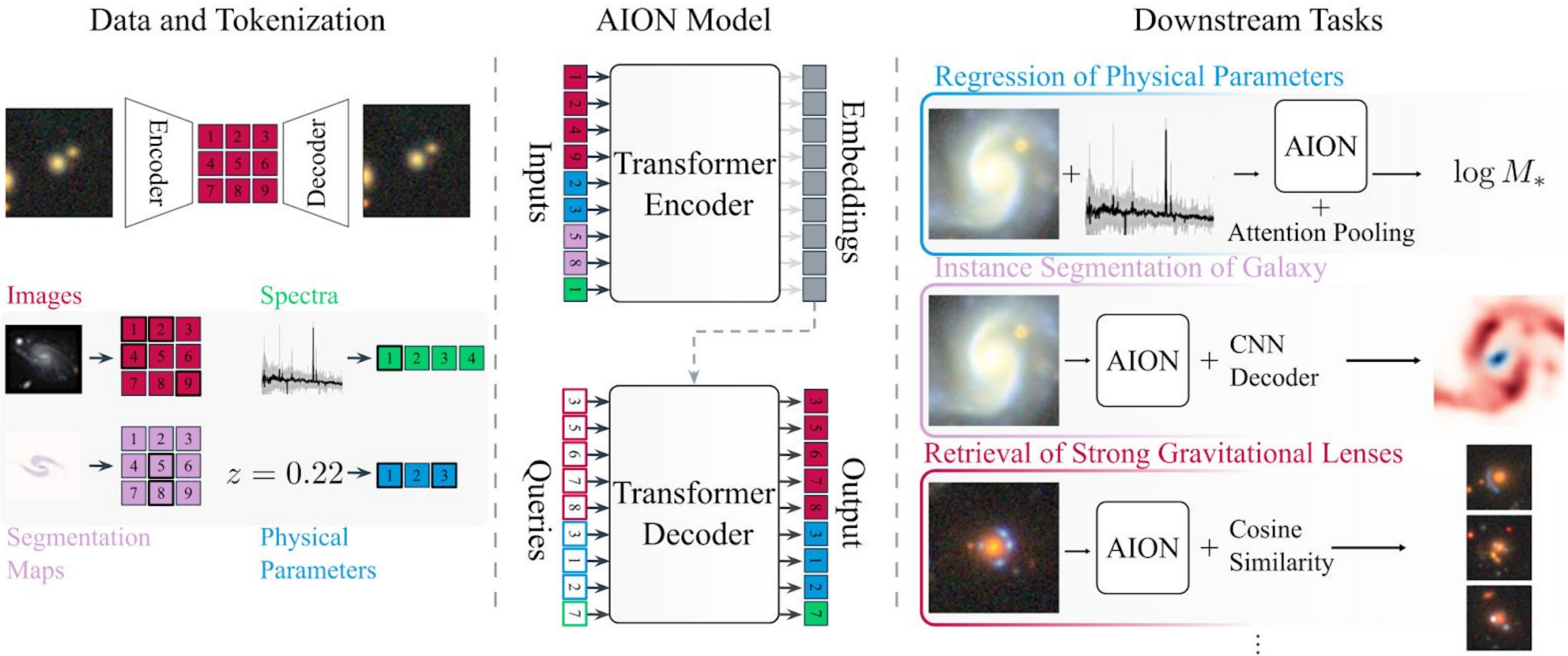
- This challenge has been accepted as one of the ***NeurIPS 2025 competitions***.

Envisioned competition schedule		
Competition Phase	**Date**	**Description**
Phase 1	Aug – Early Nov 2025	Open submissions
	Early Nov – Mid Nov 2025	Evaluating top submissions on hidden dataset
	Mid Nov 2025	Announcement of winners
Phase 2	Sep – Mid Dec 2025	Open submissions
	Mid Dec 2025 – Jan 2026	Evaluating top submissions on hidden dataset
	Jan 2026	Announcement of winners
NeurIPS 2025	Early Dec 2025	Winners will be invited to our workshop

Coming soon!

U.S. DEPARTMENT of ENERGY · BERKELEY LAB · CHA LEARN · université PARIS-SACLAY · UNIVERSITY of WASHINGTON · UCI University of California, Irvine · IAS INSTITUTE FOR ADVANCED STUDY

Foundation models (Parker etal 2025)



Agents

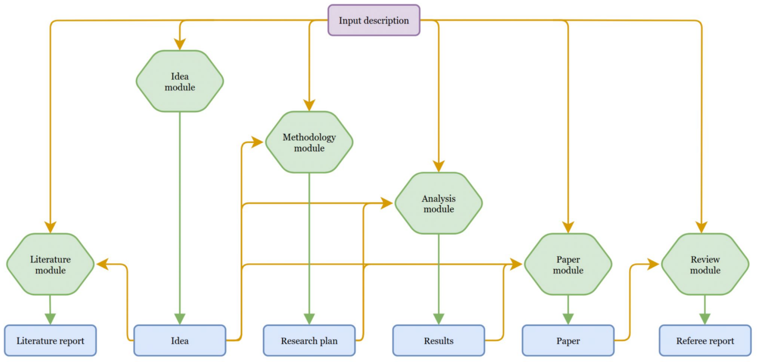
Large language models may enable a new way to do science

Working prototypes exist (e.g. L. Parker's LEO): agent under the hood collects the data, curates runs MCMC etc., outputs results, notebooks and github repository

DENARIO project: end to end paper writing

This could be how we do science in near future (next couple of years)

It will give us the freedom to spend less time on coding/debugging and more time on **thinking**!



Denario architecture.

Summary

- State of cosmology: at present standard Lambda CDM is holding up well, despite the (time dependent) noise (sigma8 is too low → sigma8 is too high)
- But there are many opportunities to improve current data analysis techniques in cosmology and test the standard model, sometimes with a large improvement in performance (e.g. **Weak Lensing, BAO**)
- Any improvement in knowledge extraction comes with a discovery potential. Maybe we will get lucky and make a big discovery (increasing BAO Figure of Merit to check DESI time dependent dark energy claim would be a good place to start!).
- **AI** can help us do the data analysis better, faster, cheaper…