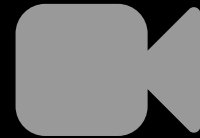


Field Level Inference

A biased Review

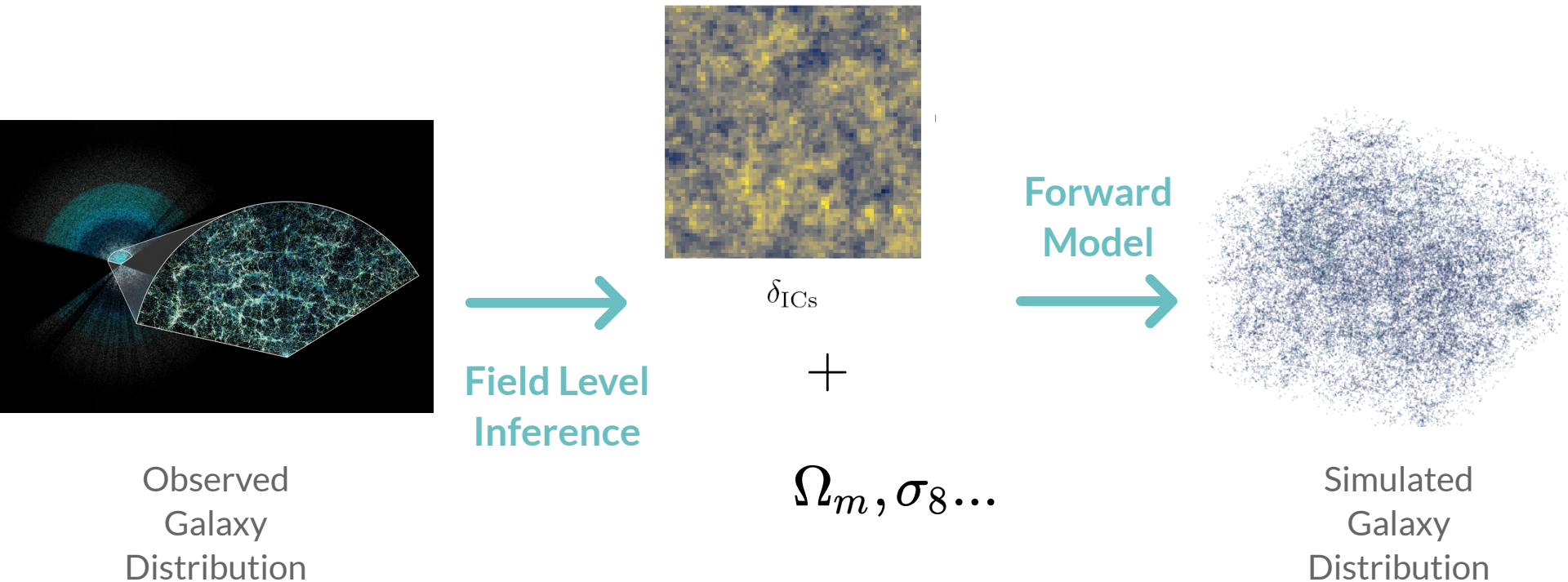


Carolina Cuesta-Lazaro

 IAIFI Fellow

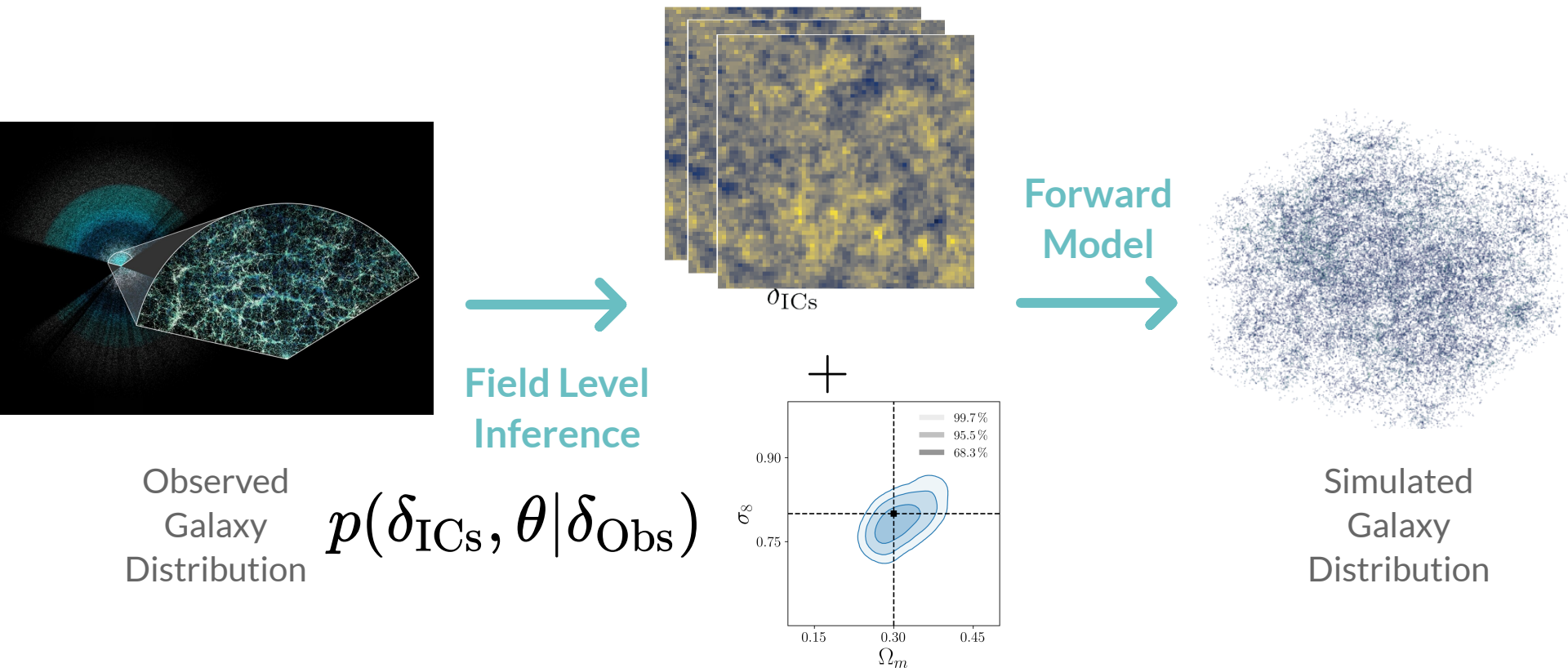
[Video Credit: N-body simulation Francisco Villaescusa-Navarro]

What is field-level inference?



***A digital twin of our Universe
(= bye bye Cosmic Variance)***

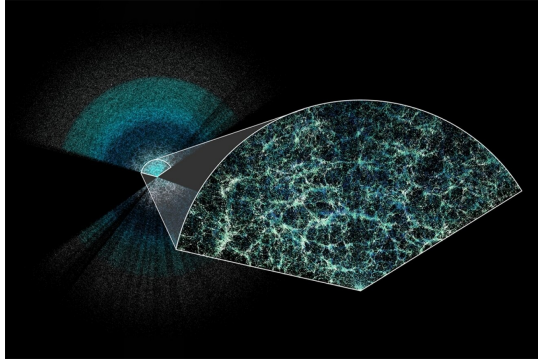
What is field-level inference?



***A digital twin of our Universe
(= bye bye Cosmic Variance)***

Why field-level inference?

Optimal constraints

$$p(\text{Cosmology} \mid \text{Image})$$


Why field-level inference?

Optimal constraints

$$p\left(\text{Cosmology} \mid \begin{array}{l} \text{N-point functions} \\ \text{Counts-in-cell} \\ \text{Marked tpcf's} \\ \text{Wavelets} \\ \text{Voids} \end{array}\right)$$

Why field-level inference?

Optimal constraints

$$p(\text{Cosmology} \mid \begin{array}{l} \text{N-point functions} \\ \text{Counts-in-cell} \\ \text{Marked tpcfs} \\ \text{Wavelets} \\ \text{Voids} \end{array})$$

M-th order. forward model: info on $N \leq M+1$

["On the Connection between Field-Level
Inference and N-point Correlation Functions"
Schmidt]

Why field-level inference?

Optimal constraints

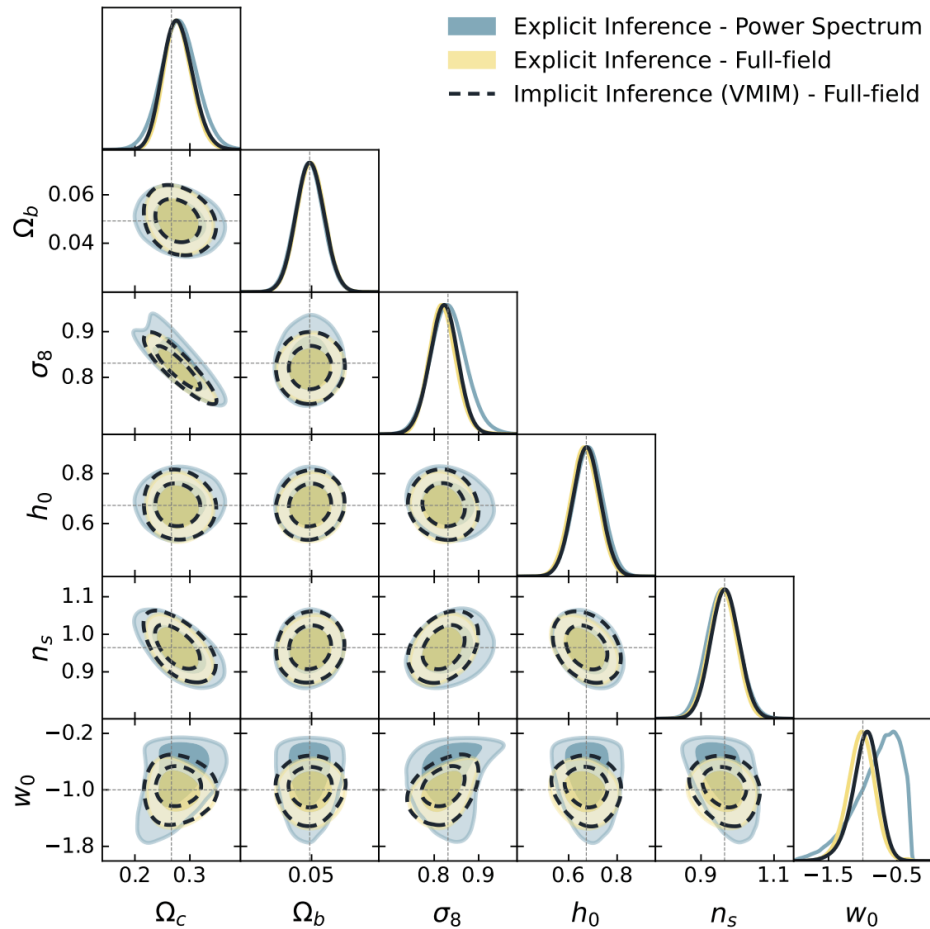
$$p\left(\begin{array}{c} \text{Cosmology} \end{array} \middle| \begin{array}{c} N\text{-point functions} \\ \text{Counts-in-cell} \\ \text{Marked tpcfs} \\ \text{Wavelets} \\ \text{Voids} \end{array}\right)$$

M-th order. forward model: info on $N \leq M+1$

["On the Connection between Field-Level
Inference and N-point Correlation Functions"
Schmidt]

Do we really need to infer the ICs?

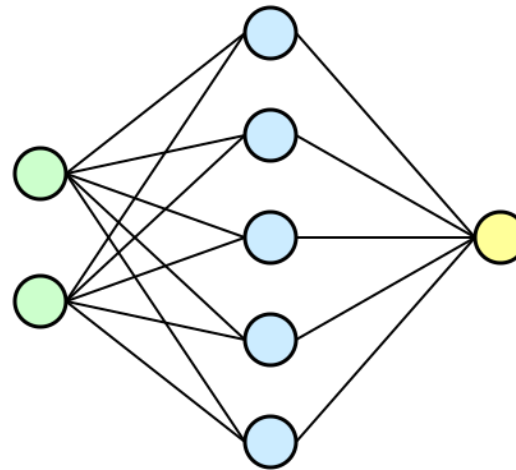
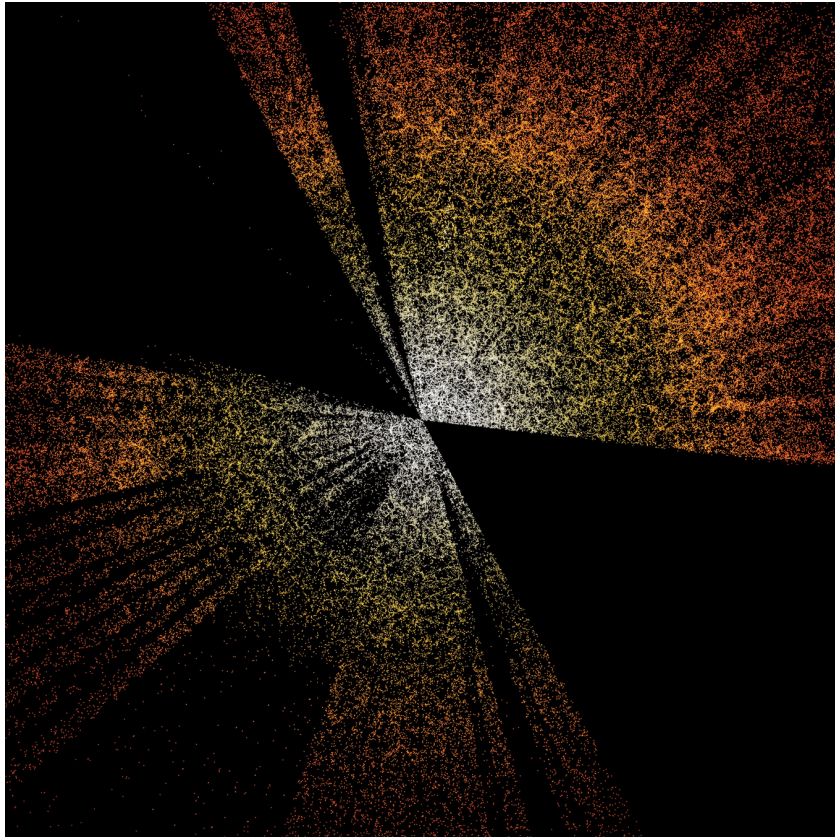
Neural Posterior Estimation -> Optimal Summaries



["Optimal Neural Summarisation for Full-Field Weak Lensing
Cosmological Implicit Inference" Lanzieri et al]

Neural Compression

$\mathcal{X}$ High-Dimensional

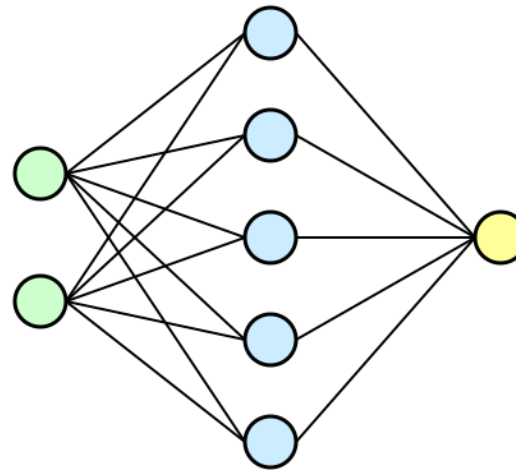
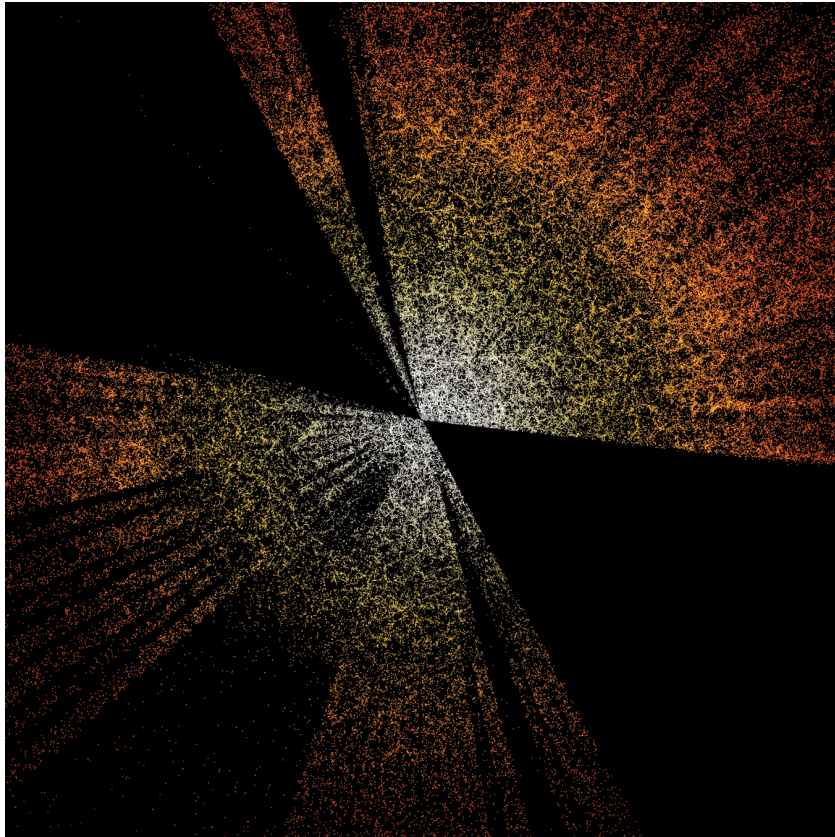


Low-Dimensional

$$s = F_{\eta}(x)$$

Neural Compression

$\mathcal{X}$ High-Dimensional



Low-Dimensional

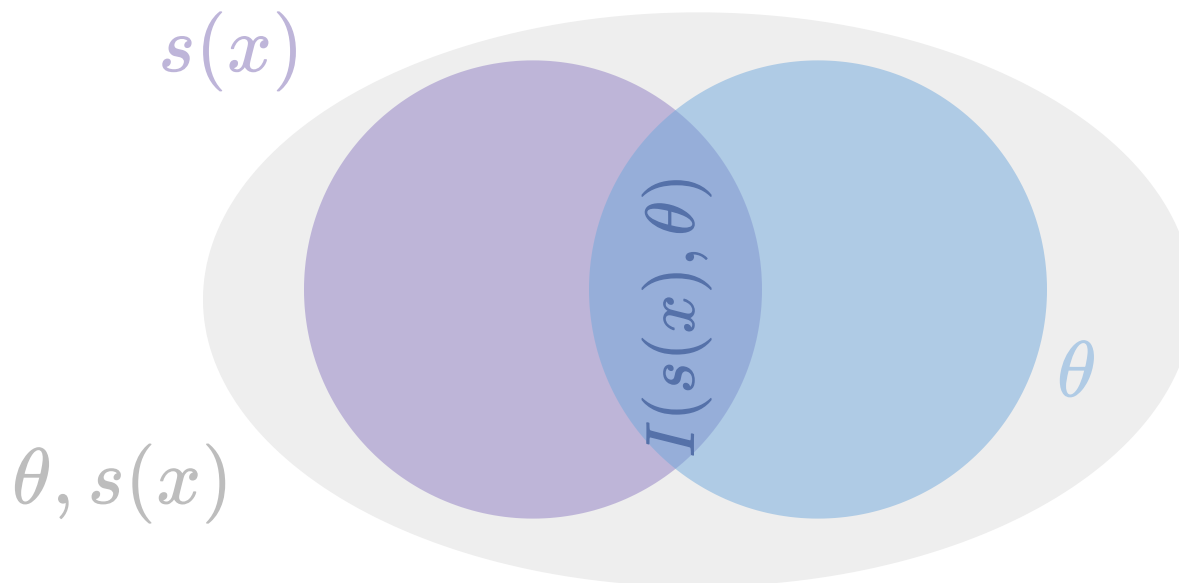
$$s = F_{\eta}(x)$$

s is sufficient iif

$$p(\theta|x) = p(\theta|s)$$

Neural Compression: MI

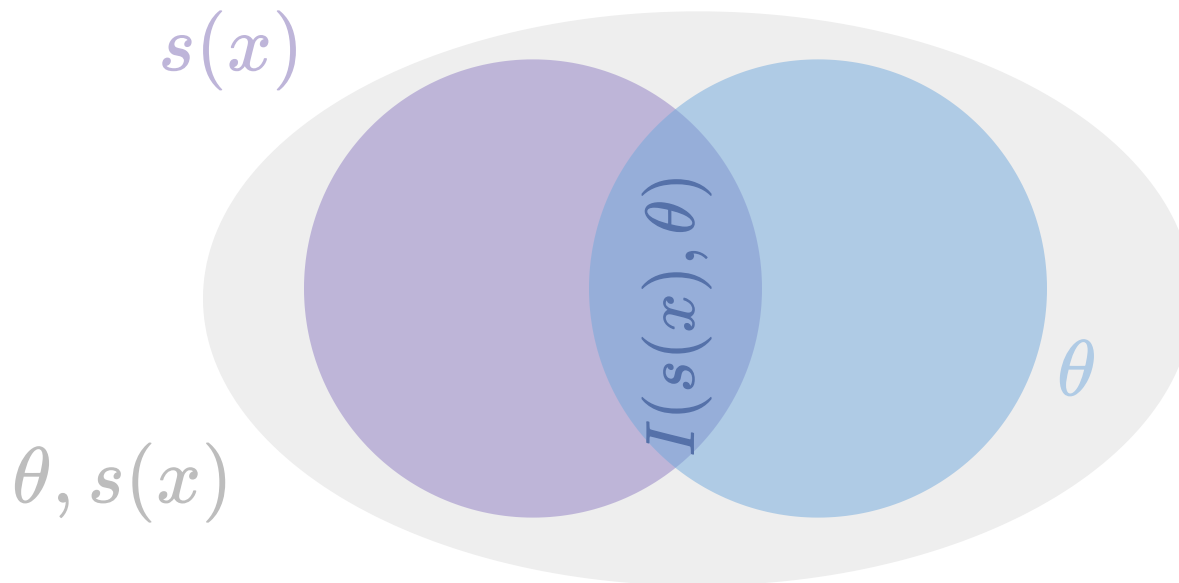
Maximise
Mutual Information $I(s(x), \theta)$



Neural Compression: MI

Maximise
Mutual Information $I(s(x), \theta)$

$$I(\theta, s(x)) = D_{\text{KL}}(p(\theta, s(x)) \parallel p(\theta)p(s(x)))$$

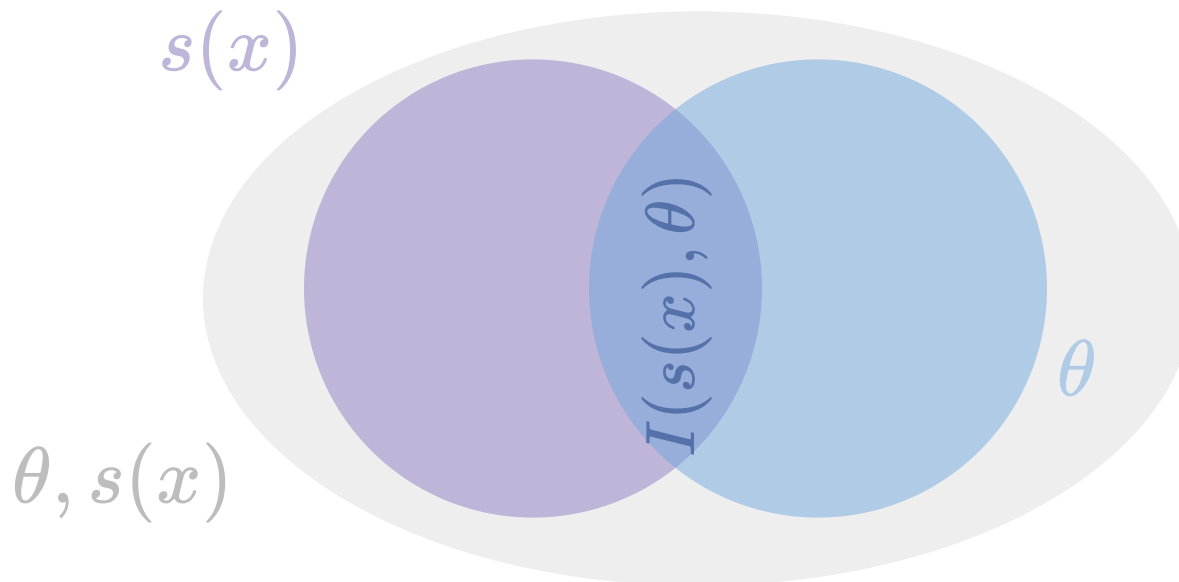


Neural Compression: MI

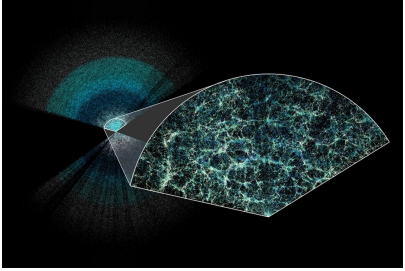
Maximise
Mutual Information $I(s(x), \theta)$

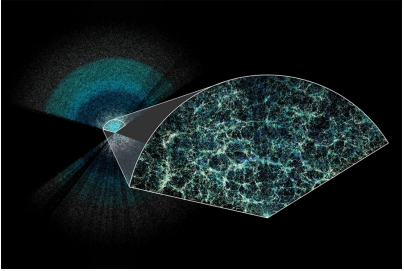
$$I(\theta, s(x)) = D_{\text{KL}}(p(\theta, s(x)) \parallel p(\theta)p(s(x)))$$

$\theta, s(x)$ independent $\rightarrow p(\theta, s(x)) = p(\theta)p(s(x))$



What field level inference isn't

$$p(\text{Cosmology} \mid S(\text{))$$

$$p(\text{Cosmology} \mid \text{)$$

What field level inference isn't

$$p(\text{Cosmology} \mid S(\text{img}))$$

$$p(\text{Cosmology} \mid \text{img})$$

["A point cloud approach to generative modeling for galaxy surveys at the field level"]

Cuesta-Lazaro and Mishra-Sharma
arXiv:2311.17141]

Is Field-Level Inference worth it?

Optimal Summaries

$$p(\theta | S(\delta_{\text{Obs}}))$$

FLI

$$p(\delta_{\text{ICs}}, \theta | \delta_{\text{Obs}})$$

Is Field-Level Inference worth it?

Optimal Summaries

$$p(\underline{\theta} | S(\delta_{\text{Obs}}))$$

$\mathcal{O}(10)$

FLI

$$p(\delta_{\text{ICs}}, \theta | \delta_{\text{Obs}})$$

Is Field-Level Inference worth it?

Optimal Summaries

$$\begin{array}{cc} p(\theta | \underline{S}(\delta_{\text{Obs}})) & \\ \mathcal{O}(10) & \mathcal{O}(10 - 100) \end{array}$$

FLI

$$p(\delta_{\text{ICs}}, \theta | \delta_{\text{Obs}})$$

Is Field-Level Inference worth it?

Optimal Summaries

$$p(\theta | \underline{S}(\delta_{\text{Obs}}))$$

$\mathcal{O}(10)$ $\mathcal{O}(10 - 100)$

FLI

$$p(\underline{\delta_{\text{ICs}}}, \theta | \underline{\delta_{\text{Obs}}})$$

$\mathcal{O}(10^9)$ $\mathcal{O}(10^9)$

Is Field-Level Inference worth it?

Optimal Summaries

$$p(\underline{\theta} | \underline{S(\delta_{\text{Obs}})})$$

$\mathcal{O}(10)$ $\mathcal{O}(10 - 100)$

Robustness?

Training simulations are IID

Low dimensional inference

FLI

$$p(\underline{\delta_{\text{ICs}}}, \theta | \underline{\delta_{\text{Obs}}})$$

$\mathcal{O}(10^9)$ $\mathcal{O}(10^9)$

Very high dimensional inference!

Is Field-Level Inference worth it?

Optimal Summaries

$$p(\theta | \underline{S}(\delta_{\text{Obs}}))$$

$\mathcal{O}(10)$ $\mathcal{O}(10 - 100)$

Robustness?

Training simulations are IID

Low dimensional inference

FLI

$$p(\underline{\delta_{\text{ICs}}}, \theta | \underline{\delta_{\text{Obs}}})$$

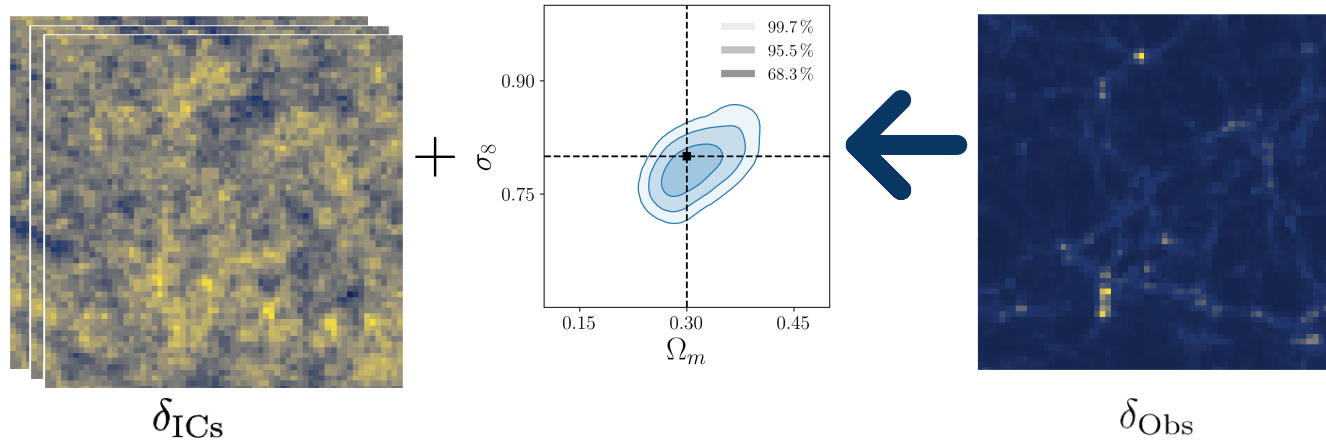
$\mathcal{O}(10^9)$ $\mathcal{O}(10^9)$

Very high dimensional inference!

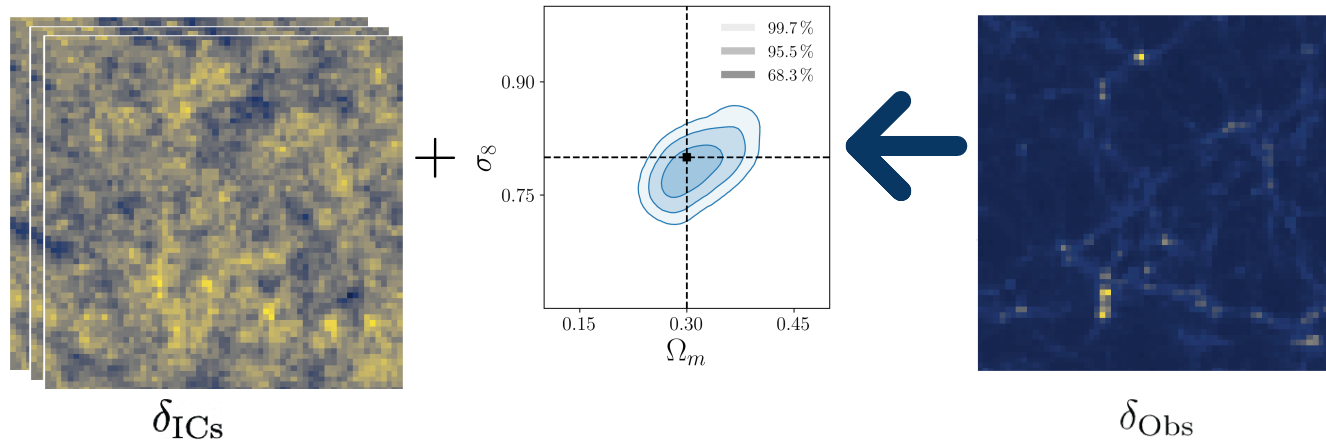
Same pixel-level fidelity required

Number of simulations needed?

Why field-level inference?



Why field-level inference?



Reconstructing ALL latent variables:

Dark Matter distribution

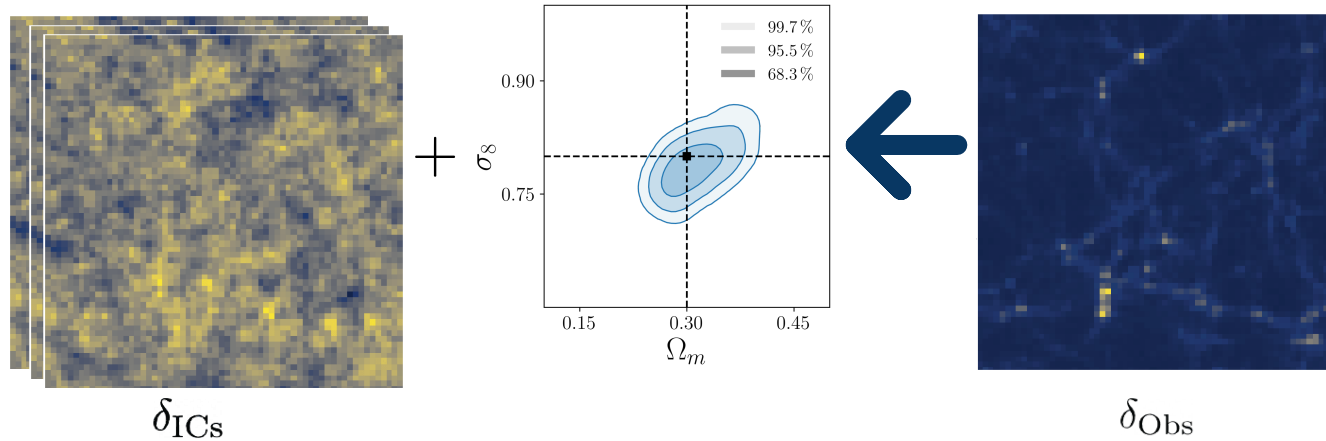
Entire formation history

Peculiar velocities

Constraining Inflation:

Inferring primordial non-gaussianity

Why field-level inference?



Reconstructing ALL latent variables:

Dark Matter distribution

Entire formation history

Peculiar velocities

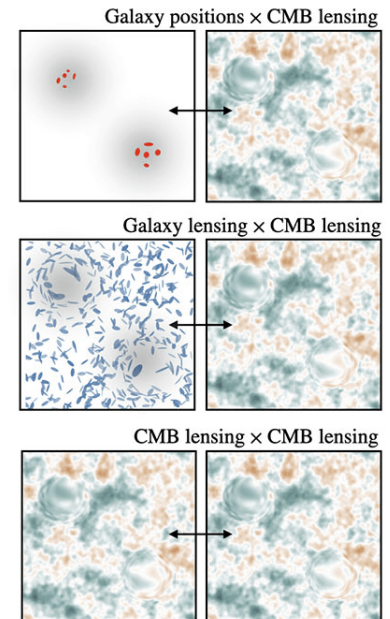
Constraining Inflation:

Inferring primordial non-gaussianity

Predictive:

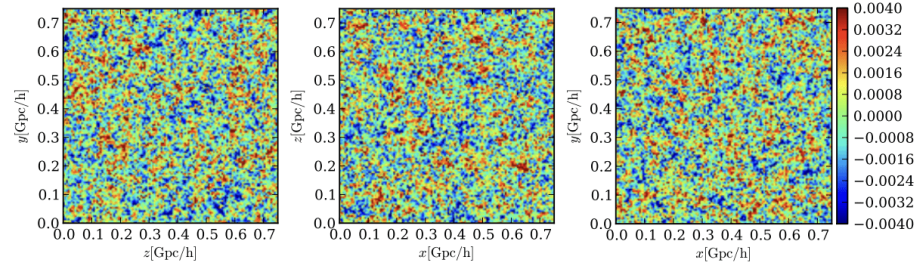
Cross-Correlation with
other probes without
Cosmic Variance

[Image Credit: Yuuki Omori]

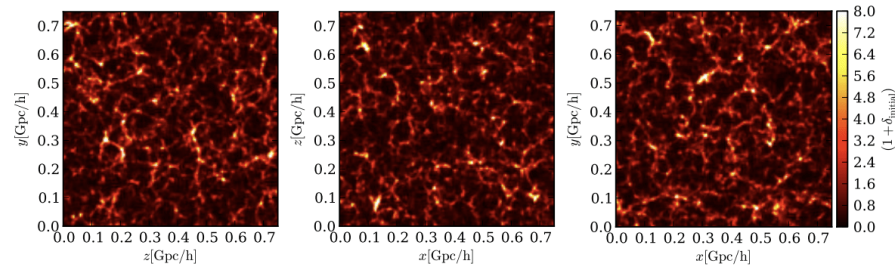


Bayesian Origin Reconstruction from Galaxies (BORG)

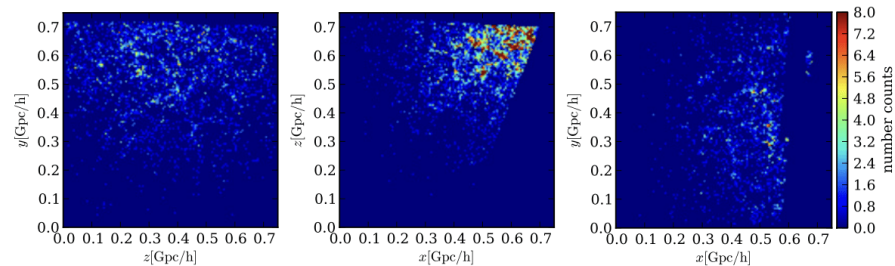
Initials



Finals



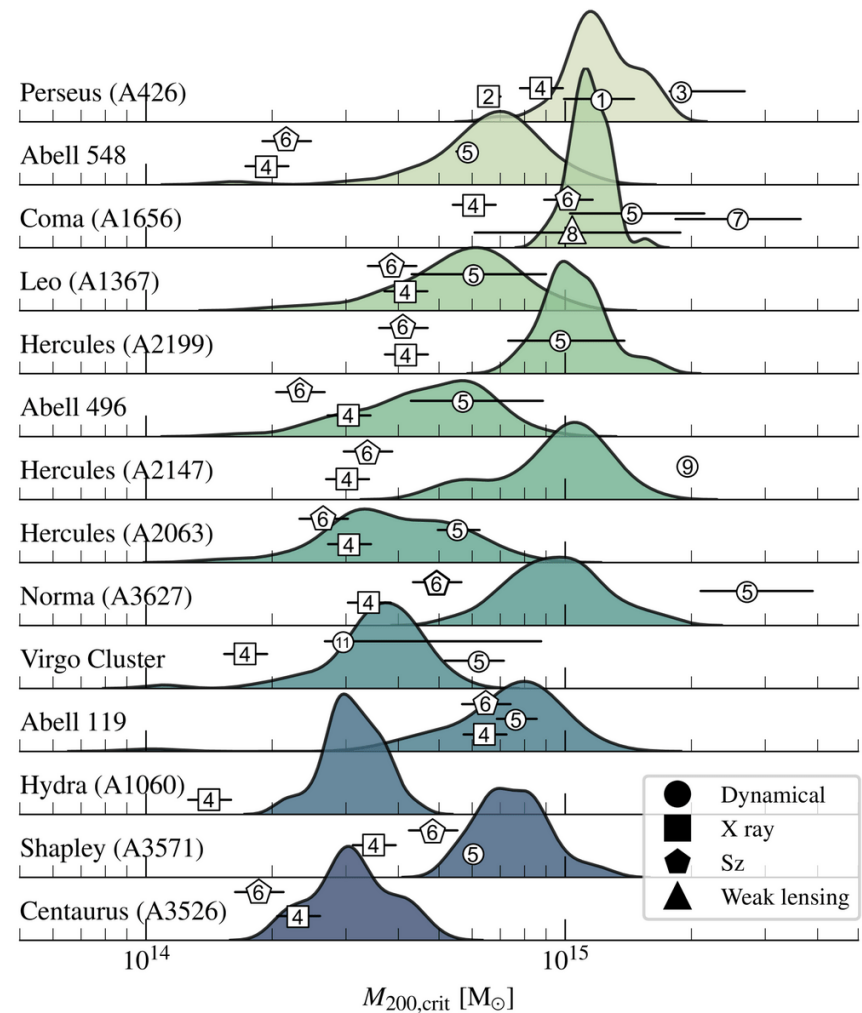
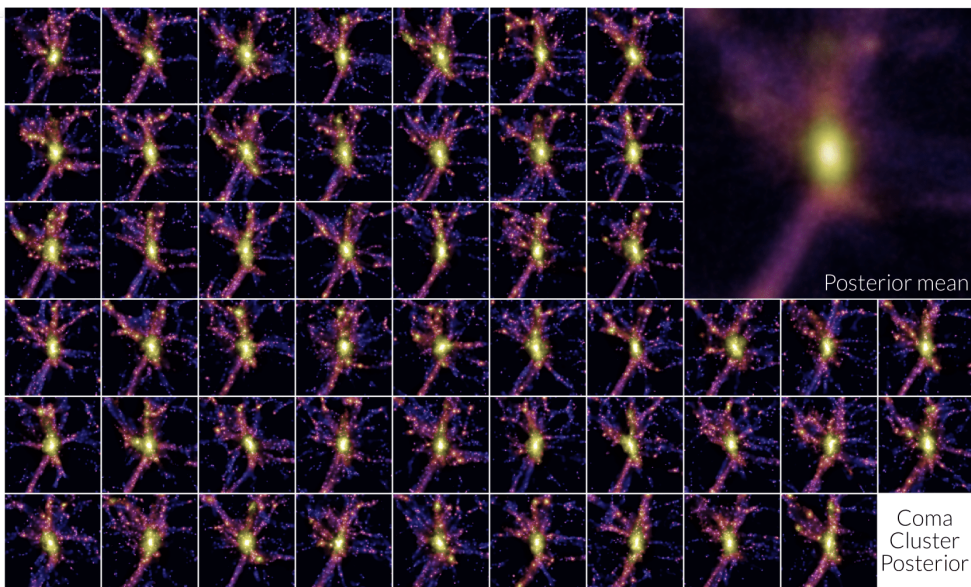
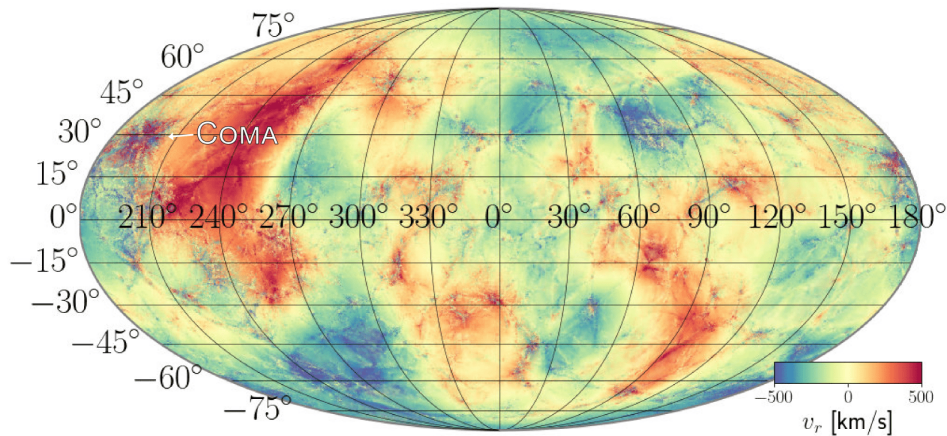
Galaxies



"Bayesian physical reconstruction of initial conditions from large scale structure surveys" Jasche, Wandelt (2012)

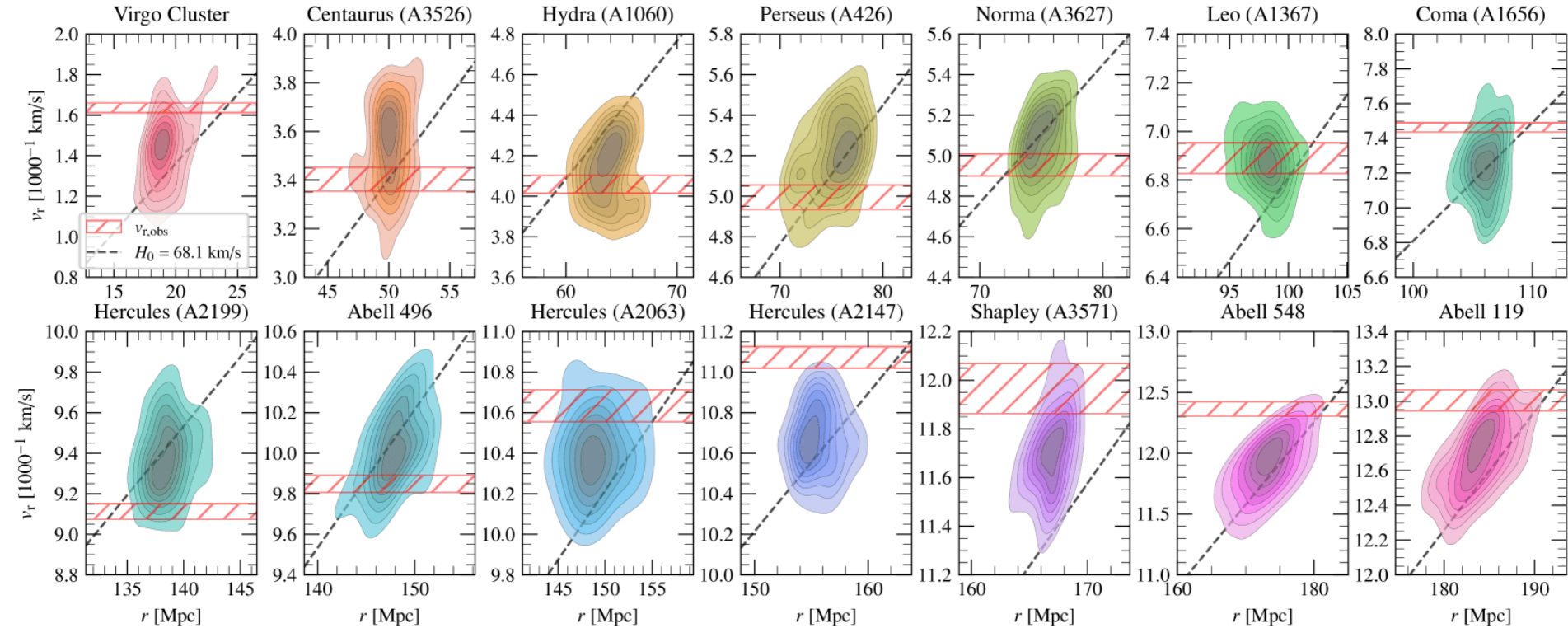
"Systematic-free inference of the cosmic matter density field from SDSS3-BOSS data" Lavaux, Jasche, Leclercq

The Local Universe without CV



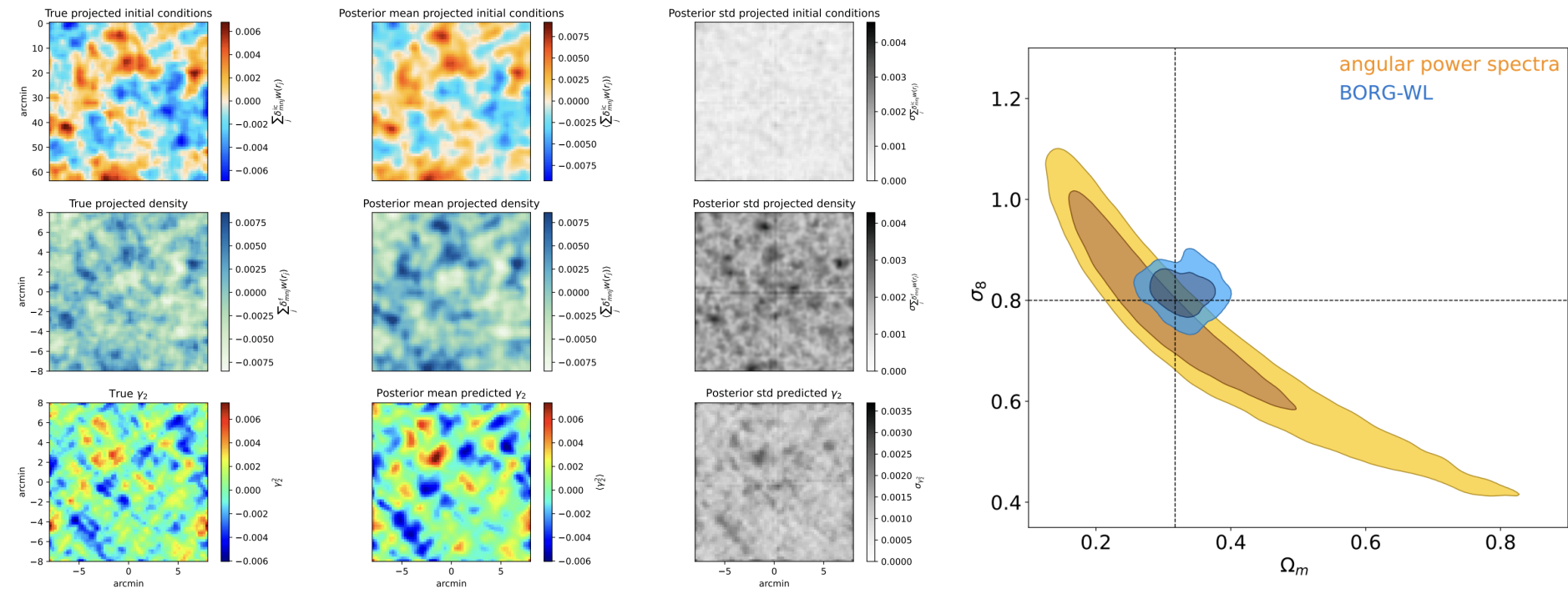
"The Manticore Project I: a digital twin of our cosmic neighbourhood from Bayesian field-level analysis" MacAlpine et al

The Hubble diagram



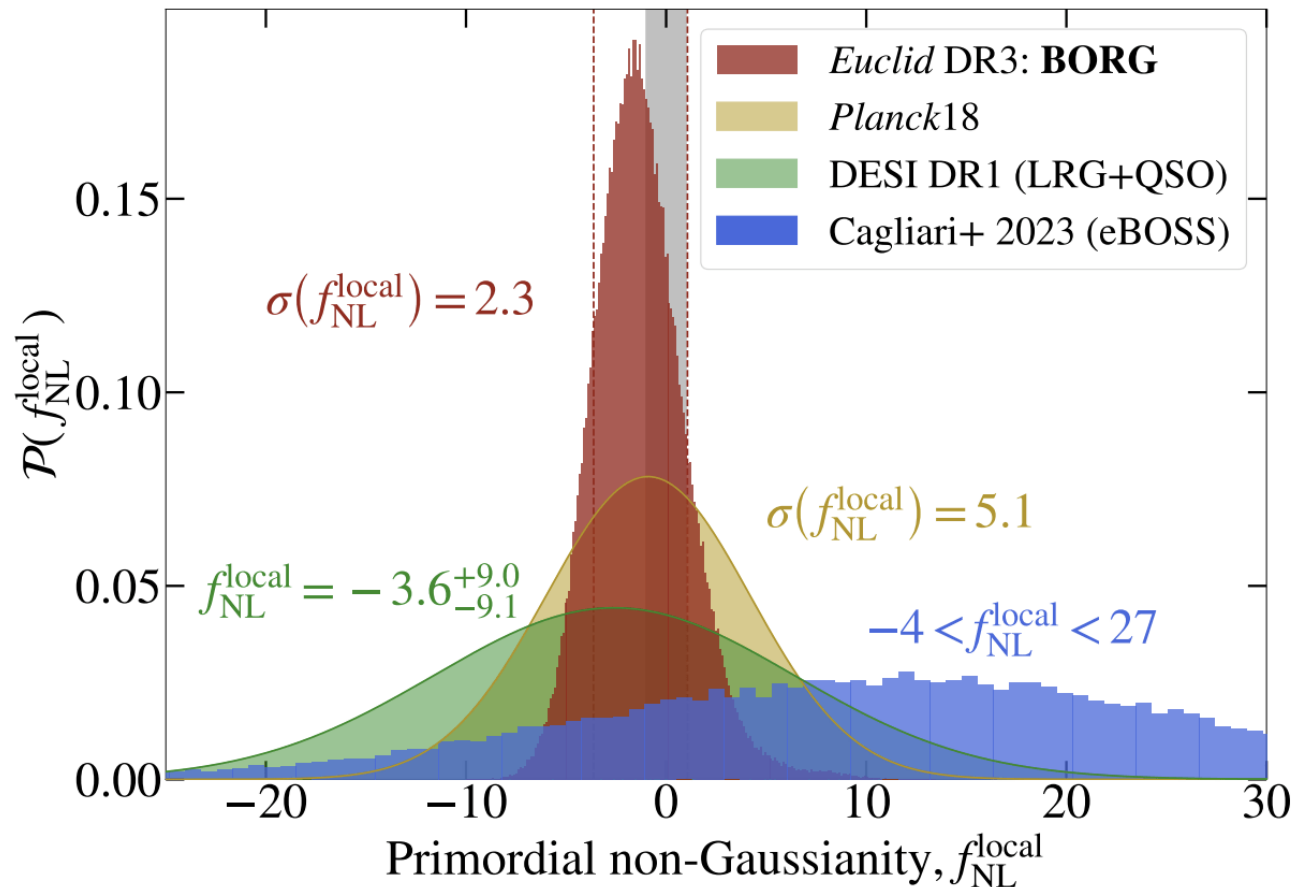
"The Manticore Project I: a digital twin of our cosmic neighbourhood from Bayesian field-level analysis" MacAlpine et al

FLI for Cosmic Shear



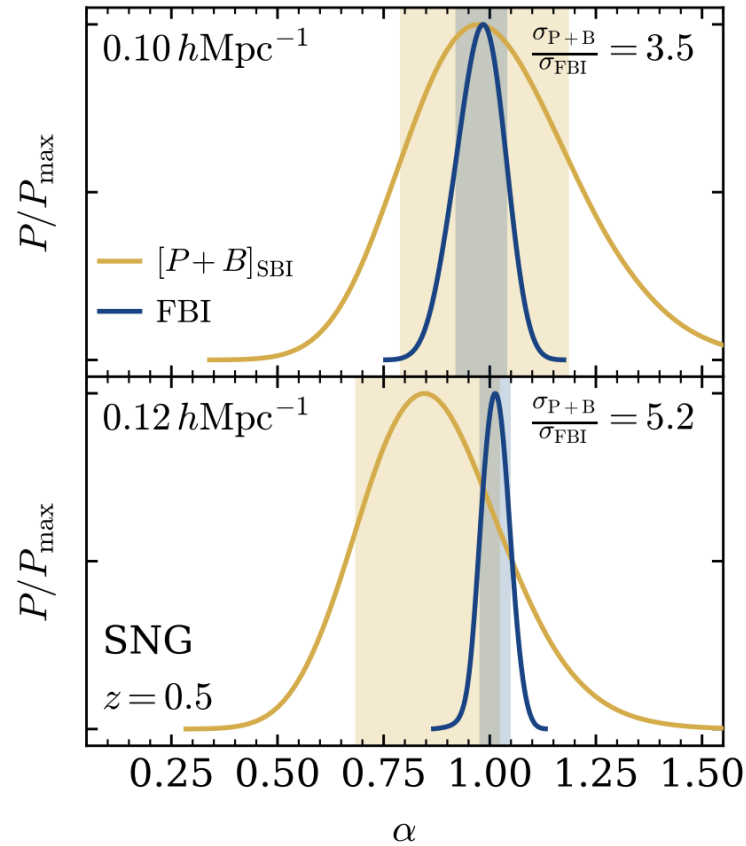
"Field-level inference of cosmic shear with intrinsic alignments and baryons" Porqueres et al

FLI for Primordial Non-Gaussianity



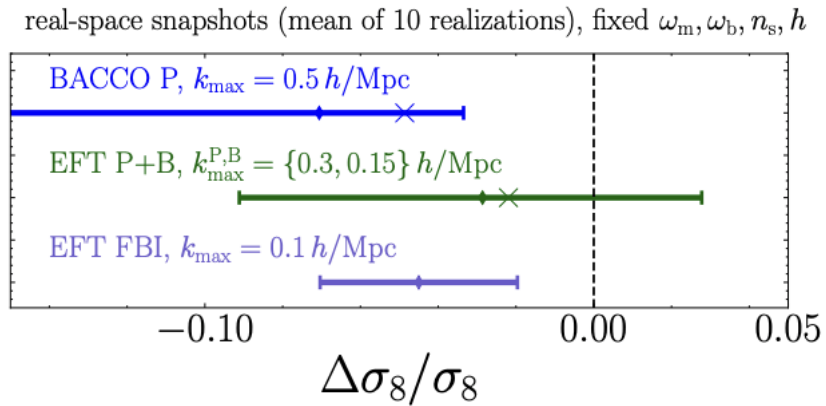
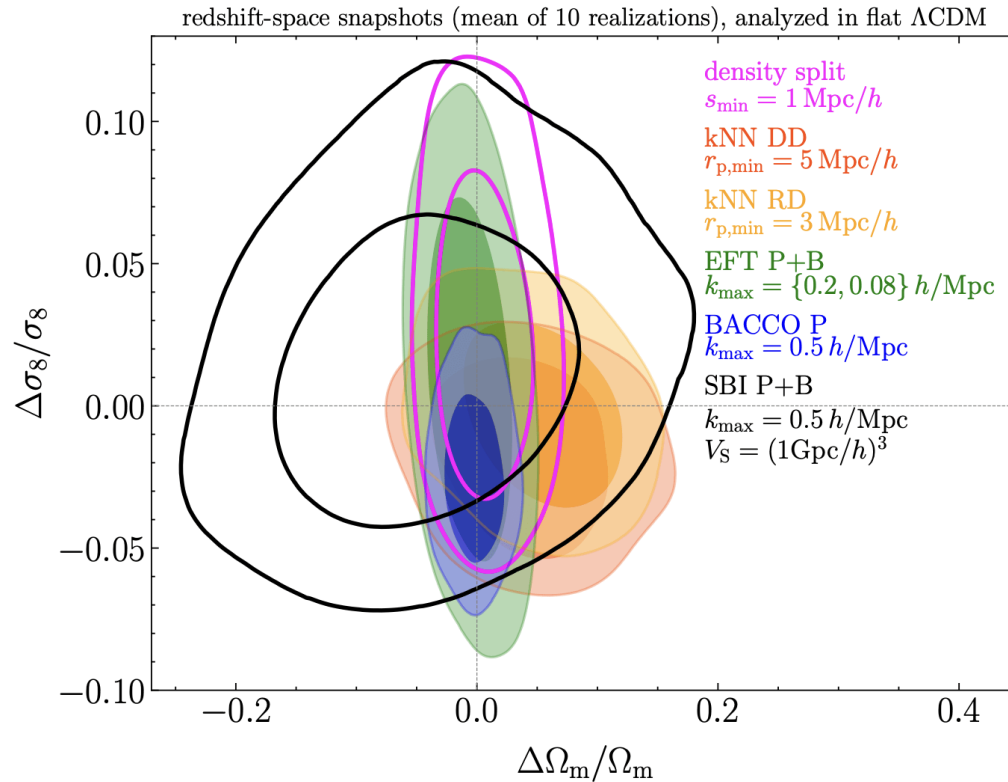
"Euclid: Field-level inference of primordial non-Gaussianity and cosmic initial conditions" Andrews et al

EFT at the Field Level



"How much information can be extracted from galaxy clustering at the field level?" Nguyen, Schmidt, Tucci, Reinecke, Kostić

The Beyond2pt Challenge



["A Parameter-Masked Mock Data Challenge for Beyond-Two-Point Galaxy Clustering statistics" Krause et al arXiv:arXiv:2405.02252]

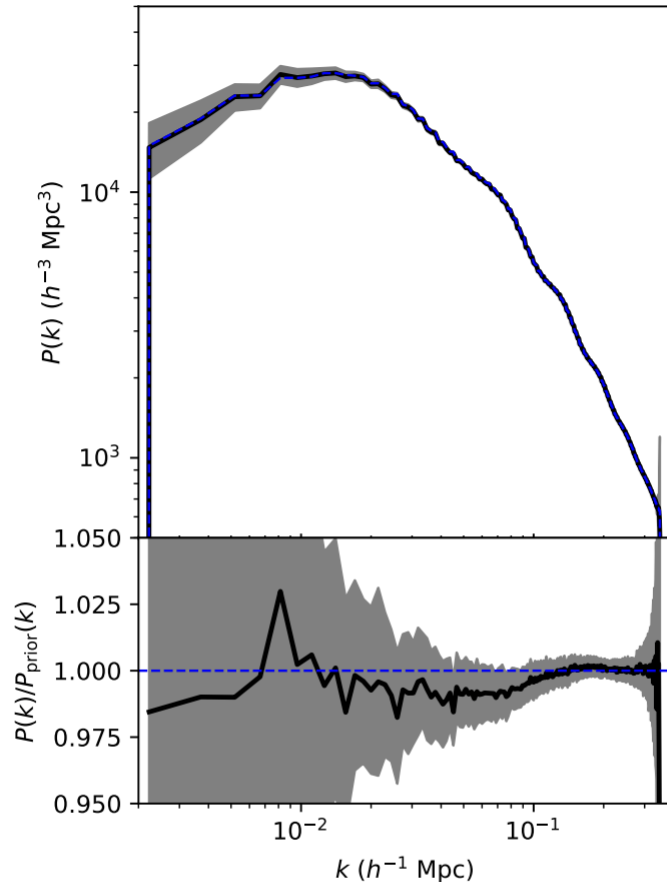
How well does the forward model fit the data?

"With four parameters I can fit an elephant, and with ~~five~~ I can make ~~him wiggle his trunk.~~"

$\mathcal{O}(10^9)$

LCDM fit DESI

Validating FLI: Testing the mean

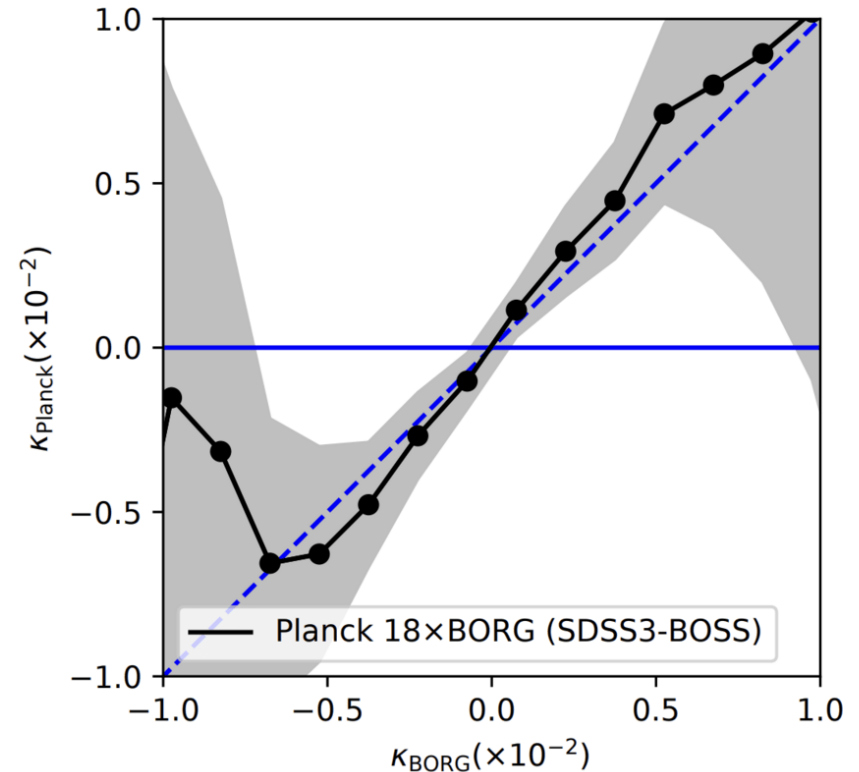


Internal Consistency

Initial $P(k)$ consistency with LCDM

"Systematic-free inference of the cosmic matter density field from SDSS3-BOSS data"

Lavaux, Jasche, Leclercq



Cross-Validation

Reconstructing lensing convergence from Planck

Validating FLI: Testing the error bars

$$\text{Credible region (CR)} \quad \int_{\Theta} \hat{p}(\theta \mid x = x_0) d\theta = 1 - \alpha$$

Validating FLI: Testing the error bars

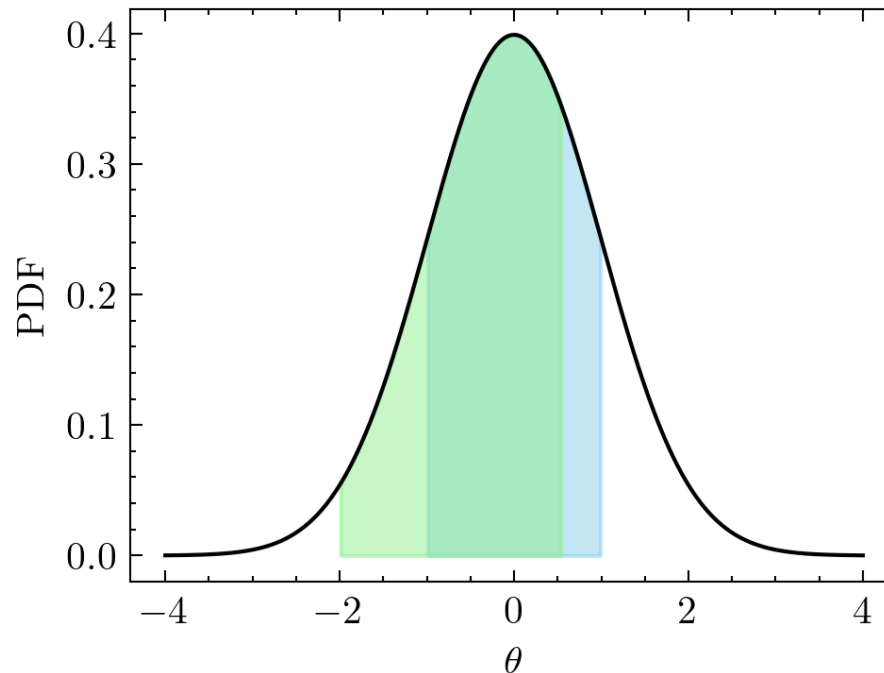
$$\text{Credible region (CR)} \quad \int_{\Theta} \hat{p}(\theta \mid x = x_0) d\theta = 1 - \alpha$$

Not unique

Validating FLI: Testing the error bars

Credible region (CR) $\int_{\Theta} \hat{p}(\theta \mid x = x_0) d\theta = 1 - \alpha$

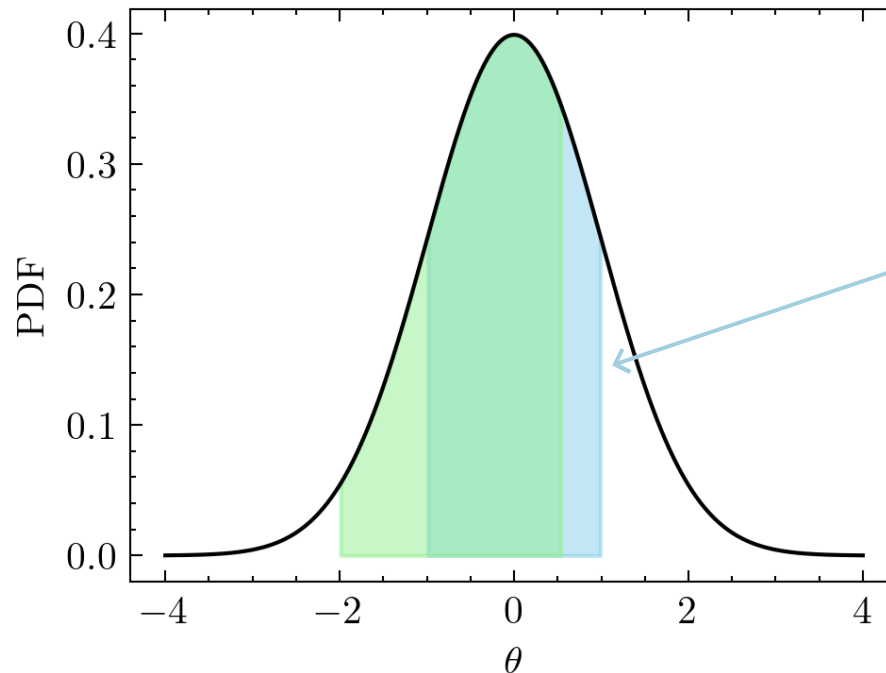
Not unique



Validating FLI: Testing the error bars

Credible region (CR) $\int_{\Theta} \hat{p}(\theta \mid x = x_0) d\theta = 1 - \alpha$

Not unique



$\mathcal{H}$

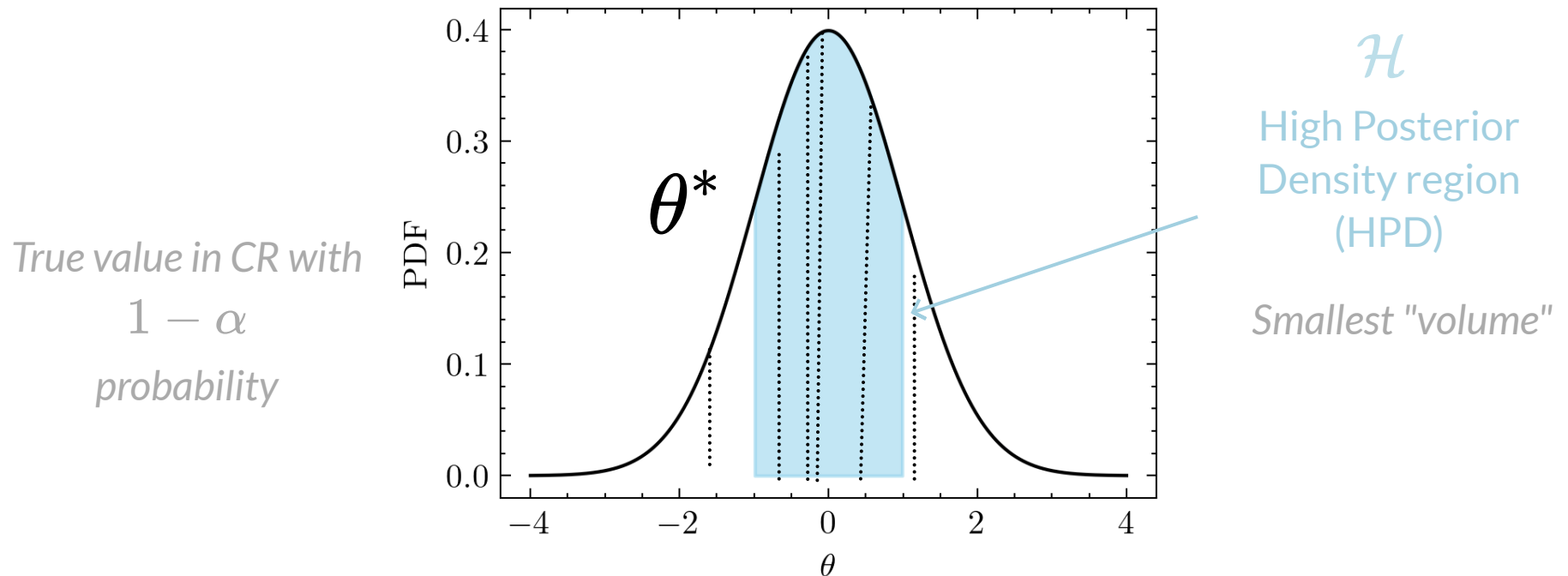
High Posterior
Density region
(HPD)

Smallest "volume"

Validating FLI: Testing the error bars

Credible region (CR) $\int_{\Theta} \hat{p}(\theta \mid x = x_0) d\theta = 1 - \alpha$

Not unique

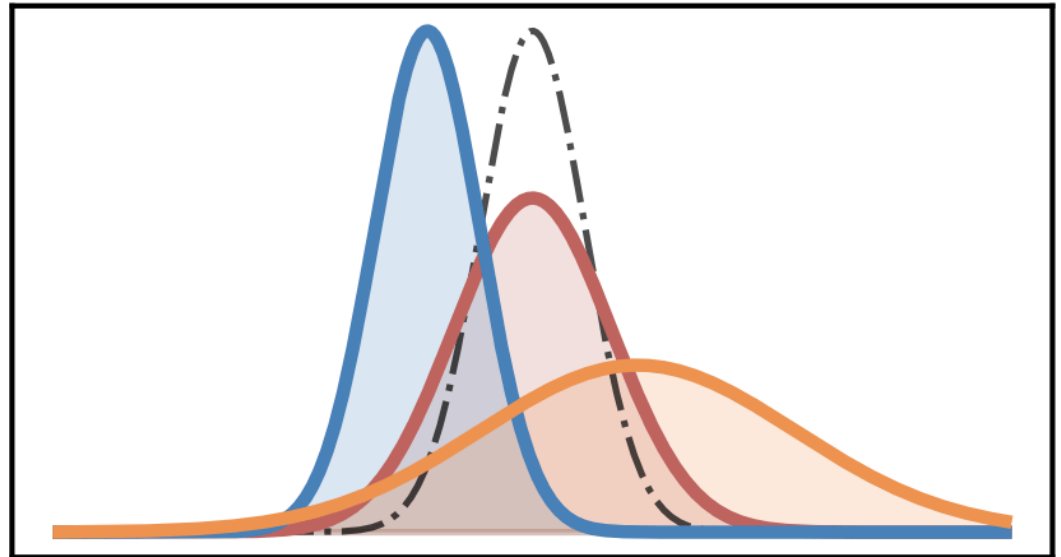
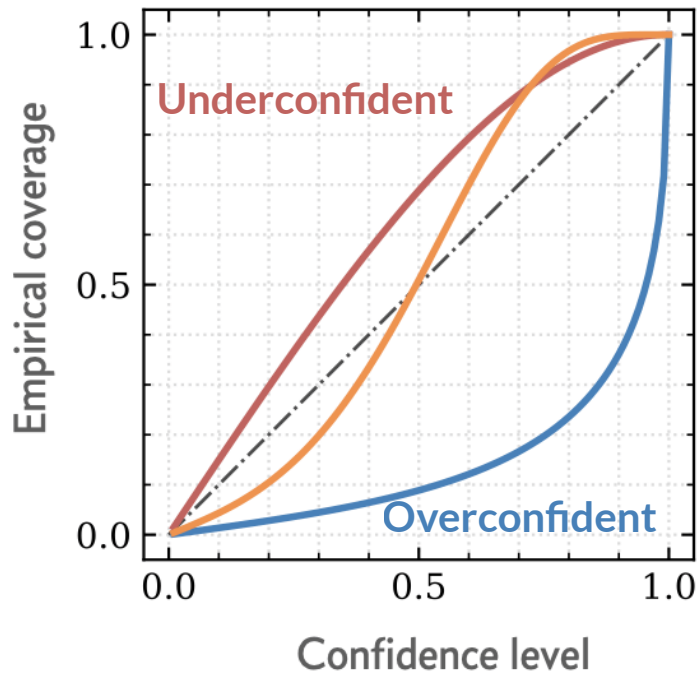


**Empirical
Coverage
Probability
(ECP)**

$$\text{ECP} = \mathbb{E}_{p(x, \theta)} \left[1 \left[\theta \in \mathcal{H}_{\hat{p}(\theta|x)}(1 - \alpha) \right] \right]$$

**Empirical
Coverage
Probability
(ECP)**

$$\text{ECP} = \mathbb{E}_{p(x, \theta)} \left[1 \left[\theta \in \mathcal{H}_{\hat{p}(\theta|x)}(1 - \alpha) \right] \right]$$



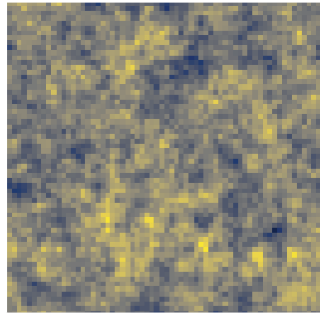
["Investigating the Impact of Model Misspecification
in Neural Simulation-based Inference"
Cannon et al arXiv:2209.01845]

The FLI ingredients:

A) Gravity

B) Galaxy biasing

C) Survey Systematics

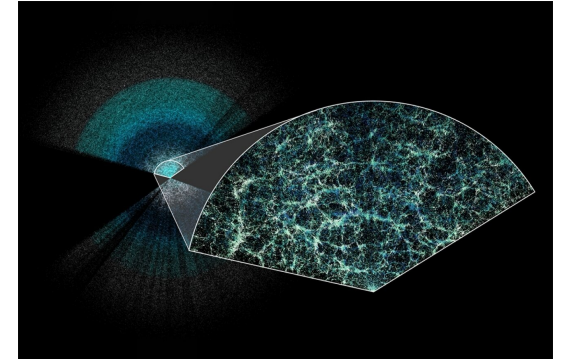


δ_{ICs}

1) Forward Model



2) Sampling Method



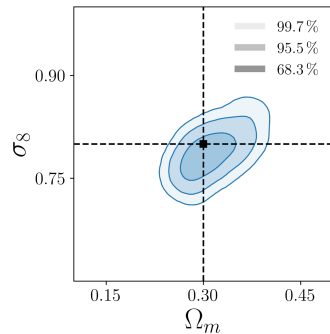
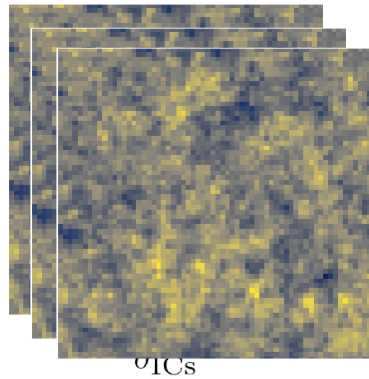
$\Omega_m, \sigma_8 \dots$

The FLI ingredients:

A) Gravity

B) Galaxy biasing

C) Survey Systematics

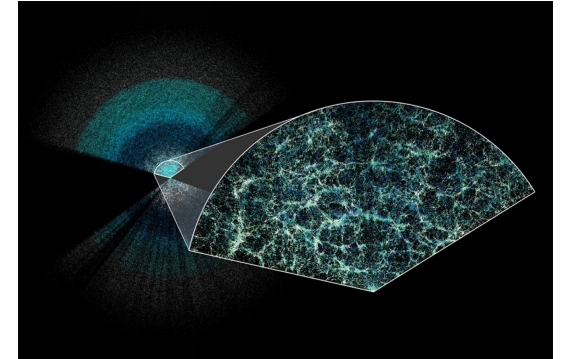


1) Forward Model



2) Sampling Method

$$p(\delta_{\text{ICs}}, \theta | \delta_{\text{Obs}})$$



Gravity

Effective Field Theories

Particle Mesh

N-body

Scale
Range

$$k_{\max} \approx 0.5h^{-1}\text{Mpc}$$

Compute
- Galaxy
formation

Compute
- Galaxy
formation

Efficient
Sampling



Gravity

Effective Field Theories

Particle Mesh

N-body

Scale
Range

$$k_{\max} \approx 0.5h^{-1}\text{Mpc}$$

Compute
- Galaxy
formation

Compute
- Galaxy
formation

Efficient
Sampling



*Neural Network
Corrections*



$$\mathbf{F}_{\theta}(\mathbf{x}, a) = \frac{3\Omega_m}{2} \nabla \left[\phi^{\text{PM}}(\mathbf{x}) + \phi_{\theta}^{\text{corr}}(\mathbf{x}, a, \phi^{\text{PM}}, \delta^{\text{PM}}) \right]$$

["Hybrid Physical Neural ODEs for Fast N-body
simulations" Lanzieri, Lanusse, Starck]



Gravity

Effective Field Theories

Particle Mesh

N-body

Scale
Range

$$k_{\max} \approx 0.5h^{-1}\text{Mpc}$$

Compute
- Galaxy
formation

Compute
- Galaxy
formation

Efficient
Sampling



*Neural Network
Corrections*



*Neural Network
Emulators*

$$\mathbf{F}_{\theta}(\mathbf{x}, a) = \frac{3\Omega_m}{2} \nabla \left[\phi^{\text{PM}}(\mathbf{x}) + \phi_{\theta}^{\text{corr}}(\mathbf{x}, a, \phi^{\text{PM}}, \delta^{\text{PM}}) \right]$$

["Hybrid Physical Neural ODEs for Fast N-body
simulations" Lanzieri, Lanusse, Starck]

["Field Level Neural
Network Emulator for
Cosmological N-body
Simulations"
Jamieson et al]



Galaxy Bias

Effective Field Theories

Robust

Clear assumptions

Large Volumes

Large Scales

Galaxy formation?

Empirical HOD/SHAM

Fast

Clear assumptions

Large Volumes

Galaxy formation?

Hydrodynamics

Self consistent
predictions

MTNG ~ 500 Mpc/h

Directly? linked to
physical processes



Galaxy Bias

Effective Field Theories

Robust

Clear assumptions

Large Volumes

Large Scales

Galaxy formation?

["Full-shape analysis with
simulation-based priors:
Constraints on single field
inflation from BOSS" Ivanov,
Cuesta-Lazaro, Mishra-Sharma,
Oblujen, Toomey arXiv:2402.13310]

Empirical HOD/SHAM

Fast

Clear assumptions

Large Volumes

Galaxy formation?

Hydrodynamics

Self consistent
predictions

MTNG ~ 500 Mpc/h

Directly? linked to
physical processes



Galaxy Bias

Effective Field Theories

Robust

Clear assumptions

Large Volumes

Large Scales

Galaxy formation?

["Full-shape analysis with simulation-based priors: Constraints on single field inflation from BOSS" Ivanov, Cuesta-Lazaro, Mishra-Sharma, Oblujen, Toomey arXiv:2402.13310]

Empirical HOD/SHAM

Fast

Clear assumptions

Large Volumes

Galaxy formation?

Hydrodynamics

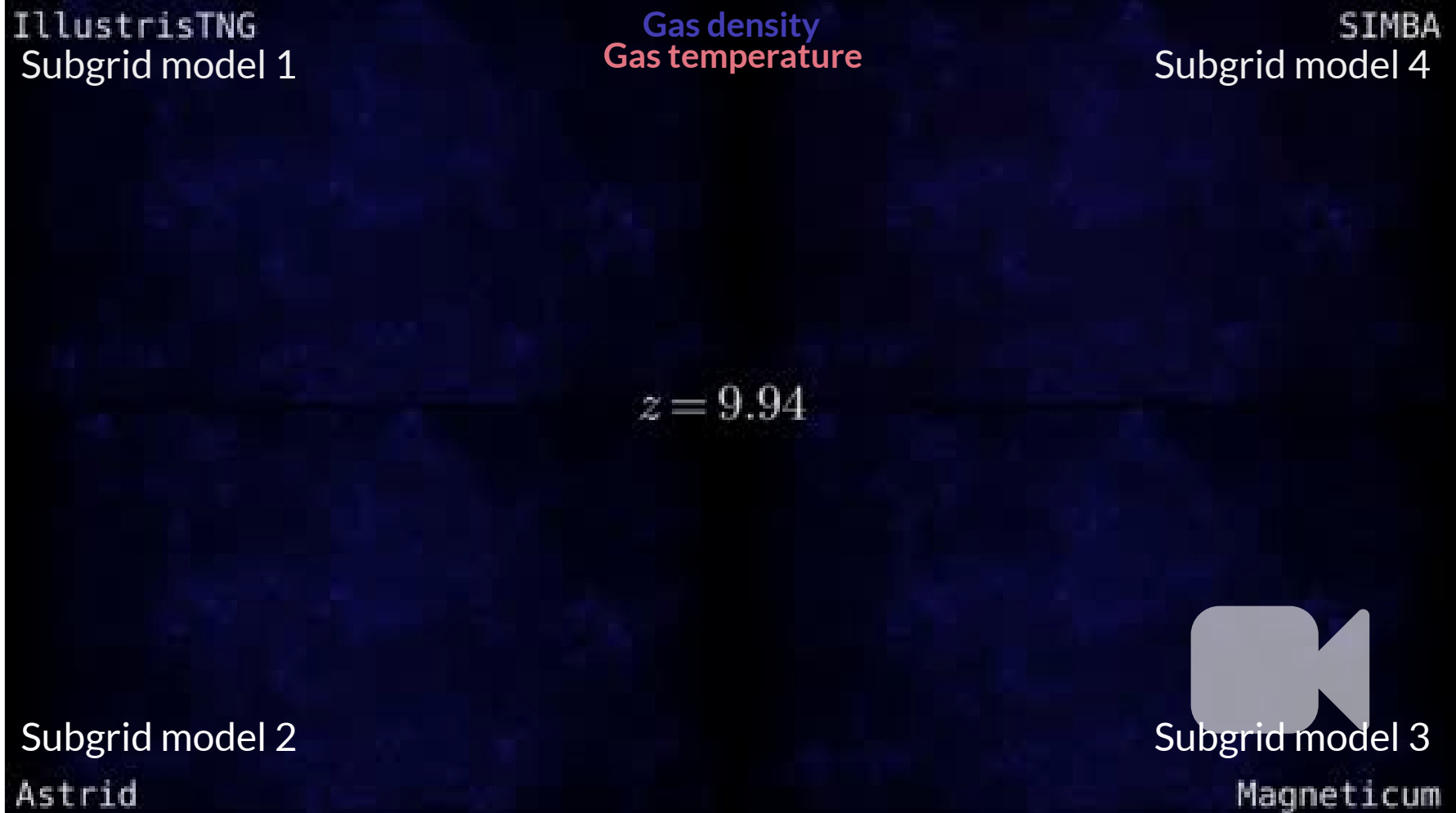
Self consistent predictions

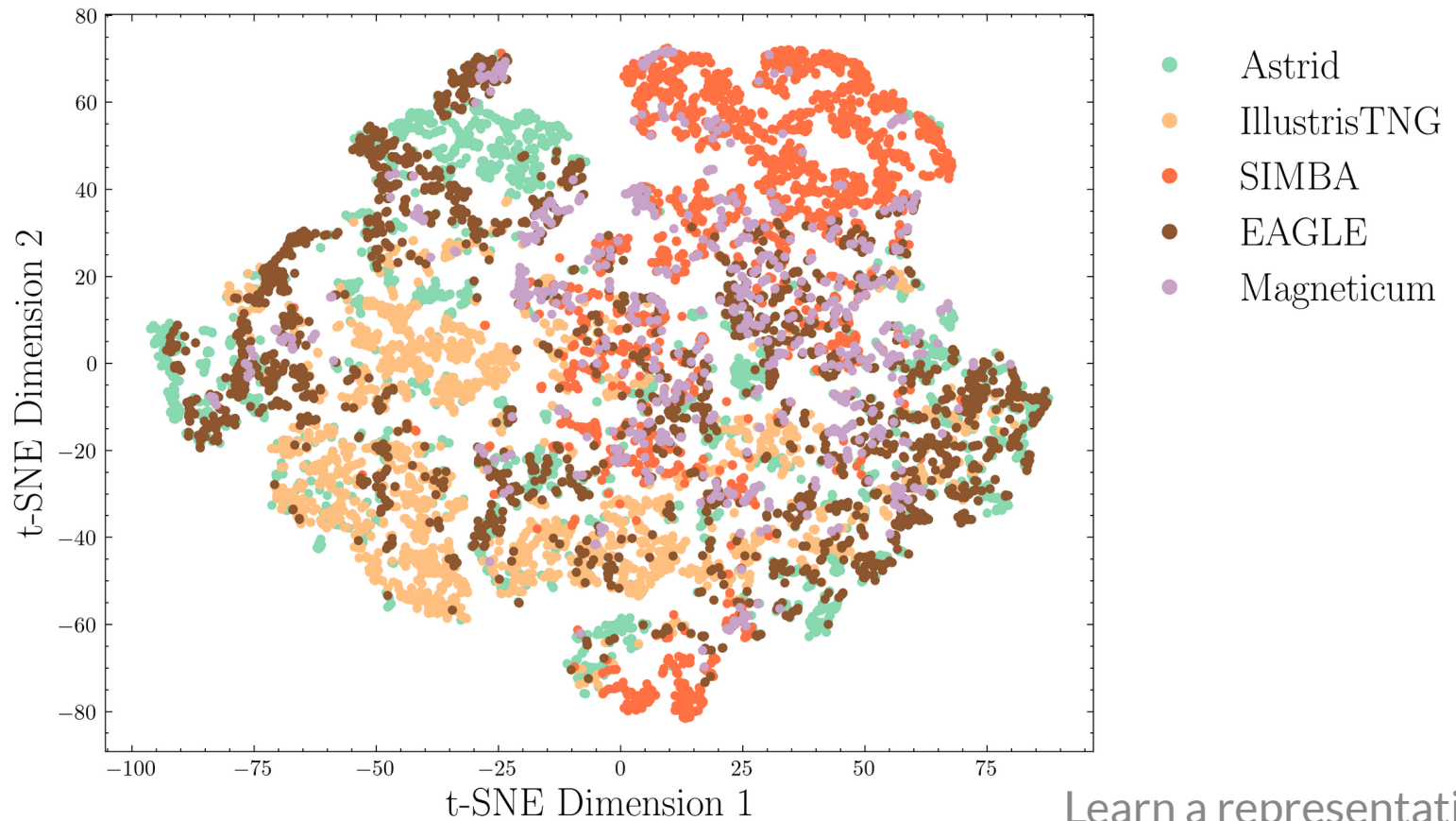
MTNG ~ 500 Mpc/h

Directly? linked to physical processes

["Differentiable Cosmological Hydrodynamics for Field-Level Inference and High Dimensional Parameter Constraints" Horowitz, Lukic arXiv:2502.02294]





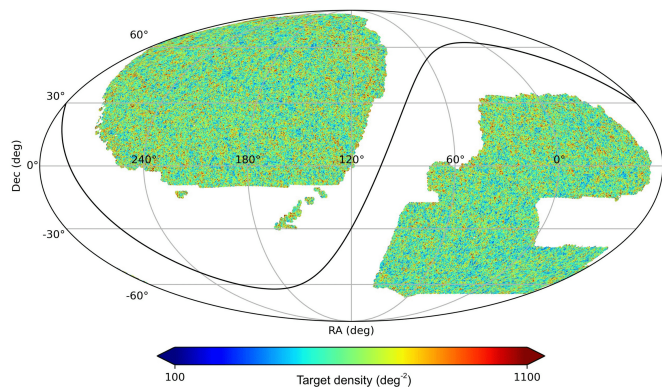


Learn a representation
for feedback

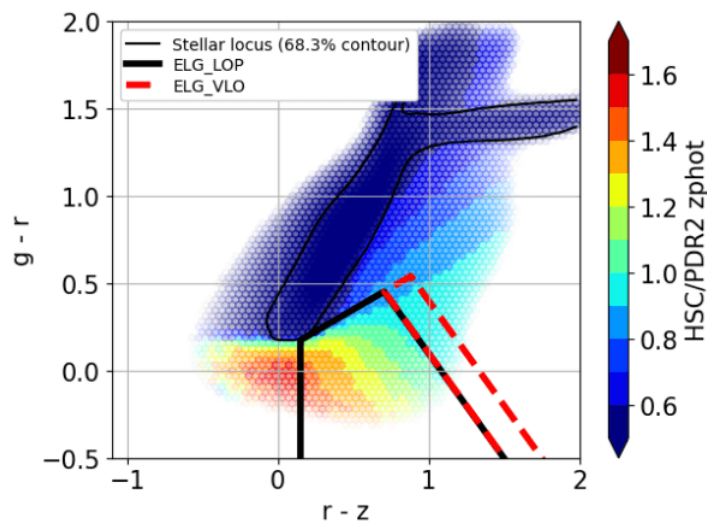
$$p(\text{Baryonic fields} \mid \text{Dark Matter}, z)$$

Survey Forward Model: The Usual Suspects

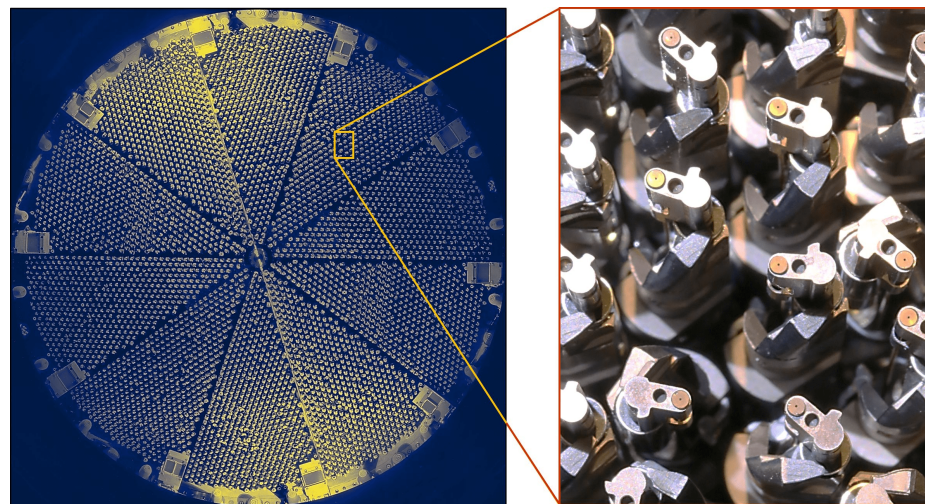
Survey Mask



Target Selection

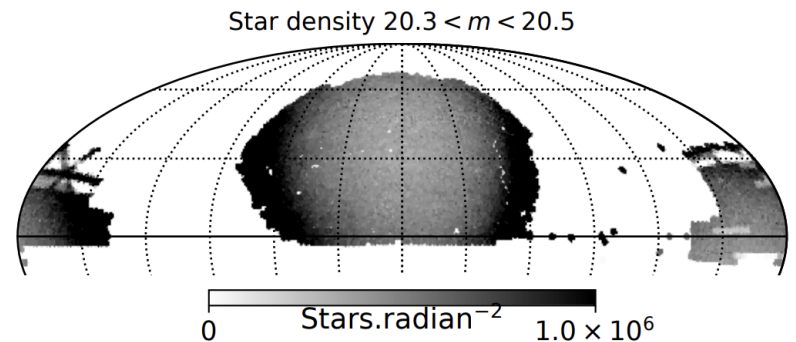
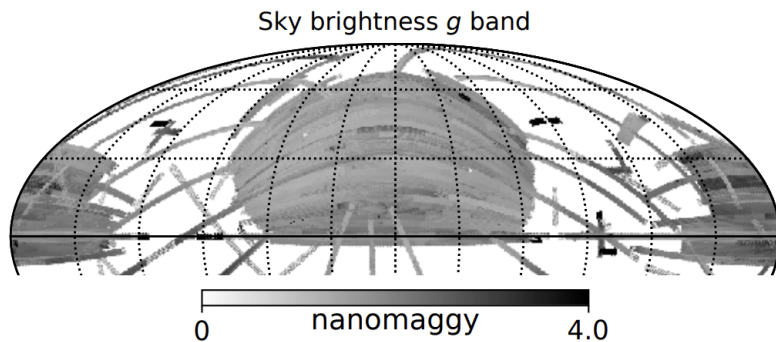
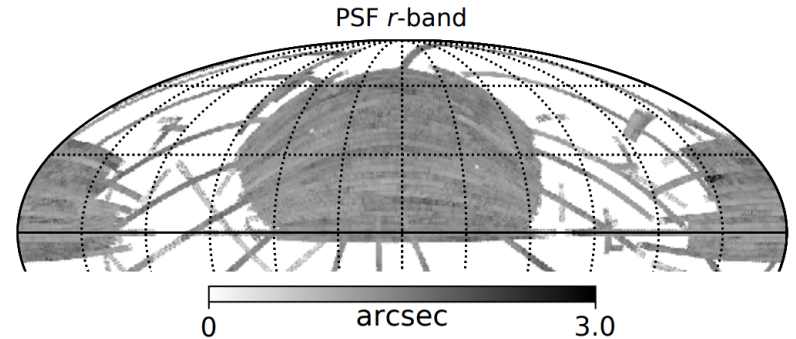
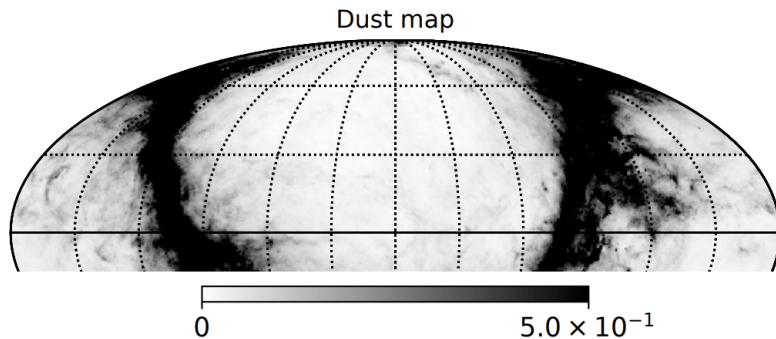


Fiber Collisions



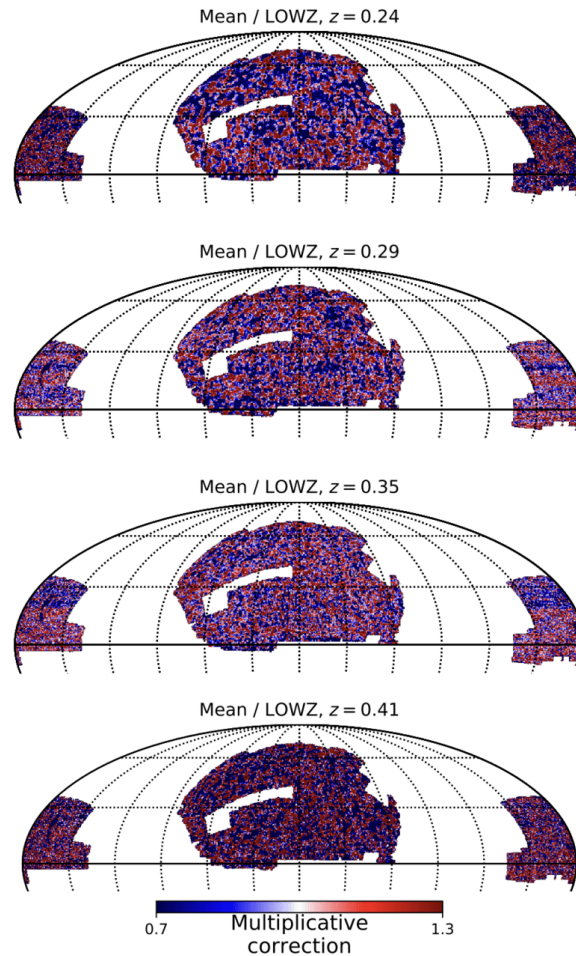
(tracer dependent priorities)

Survey Forward Model: Know Unknowns



"Systematic-free inference of the cosmic matter density field from SDSS3-BOSS data"
Lavaux, Jasche, Leclercq

Survey Forward Model: Know Unknowns



"Systematic-free inference of the cosmic matter density field from SDSS3-BOSS data"
Lavaux, Jasche, Leclercq

The data likelihood

Poisson

$$P(n_i|\lambda_i) = \frac{\lambda_i^{n_i} e^{-\lambda_i}}{n_i!}$$

Gaussian

$$P(n_i|\mu_i) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(n_i - \mu_i)^2}{2\sigma^2}\right)$$

The data likelihood

Poisson

$$P(n_i|\lambda_i) = \frac{\lambda_i^{n_i} e^{-\lambda_i}}{n_i!}$$

Gaussian

$$P(n_i|\mu_i) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(n_i - \mu_i)^2}{2\sigma^2}\right)$$

More testing in realistic scenarios

"Impacts of the physical data model on the forward inference of
initial conditions from biased tracers"

Nguyen, Schmidt, Lavaux, Jasche

Sampling Methods

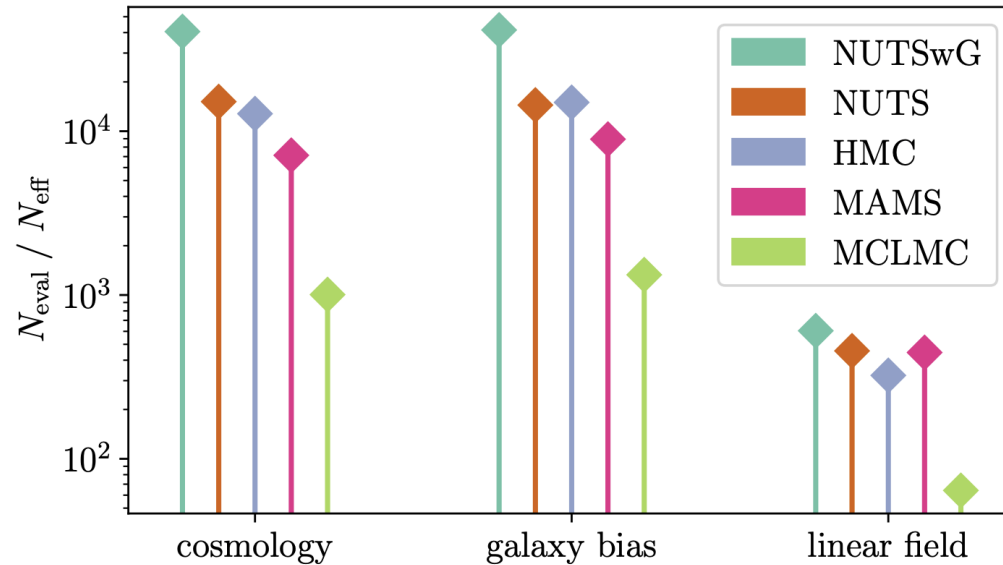


Figure 5: Number of model evaluations per effective sample for each sampler and each set of parameters (cosmology, galaxy bias and linear field).

"Benchmarking field-level cosmological inference from galaxy redshift surveys" Simon-Onfroy, Lanusse, De Mattia

"Microcanonical Hamiltonian Monte Carlo" Robnik, De Luca, Silverstein, Seljak

"Field-Level Inference with Microcanonical Langevin Monte Carlo"
Bayer, Seljak, Modi

FLI Forward Models

BORG: 2LPT Gravity, Power-law galaxy bias

FLI Forward Models

BORG: 2LPT Gravity, Power-law galaxy bias

1) *Current analysis rely on **simple forward models** pushed to small scales (+ differentiable)*

FLI Forward Models

BORG: 2LPT Gravity, Power-law galaxy bias

1) Current analysis rely on ***simple forward models*** pushed to small scales (+ differentiable)

2) Likelihood is ***complex*** for realistic scenarios, but can get samples from simulator

FLI Forward Models

BORG: 2LPT Gravity, Power-law galaxy bias

1) Current analysis rely on **simple forward models** pushed to small scales (+ differentiable)

2) Likelihood is **complex** for realistic scenarios, but can get samples from simulator

3) Either not sampling cosmology at all, or struggling to

FLI Forward Models

BORG: 2LPT Gravity, Power-law galaxy bias

- 1) Current analysis rely on **simple forward models** pushed to small scales (+ differentiable)
- 2) Likelihood is **complex** for realistic scenarios, but can get samples from simulator
- 3) Either not sampling cosmology at all, or struggling to
- 4) Not amortized -> Rigorous testing (coverage) becomes extremely hard

Meanwhile, on Earth...

Contrastive Learning

A folk music band of anthropomorphic autumn leaves playing bluegrass instrument

BigGAN

VAEs

GANS



2006

2014

2017

2019

2022

2023

Deep Belief Networks

0 0
1 1
2 2
3 3
... ..

Normalising Flows



Diffusion Models



Probabilistic ML has made *high dimensional inference tractable*

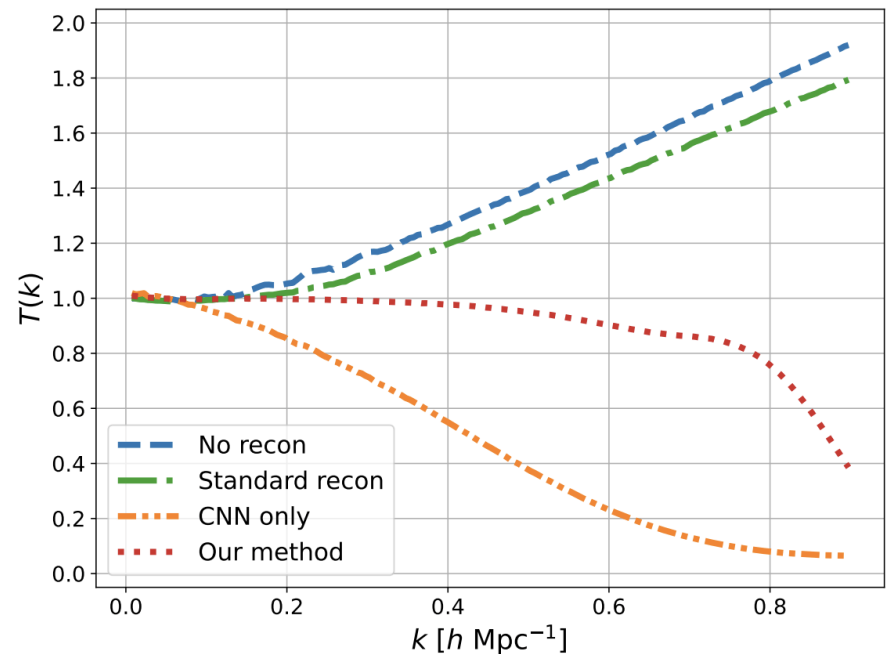
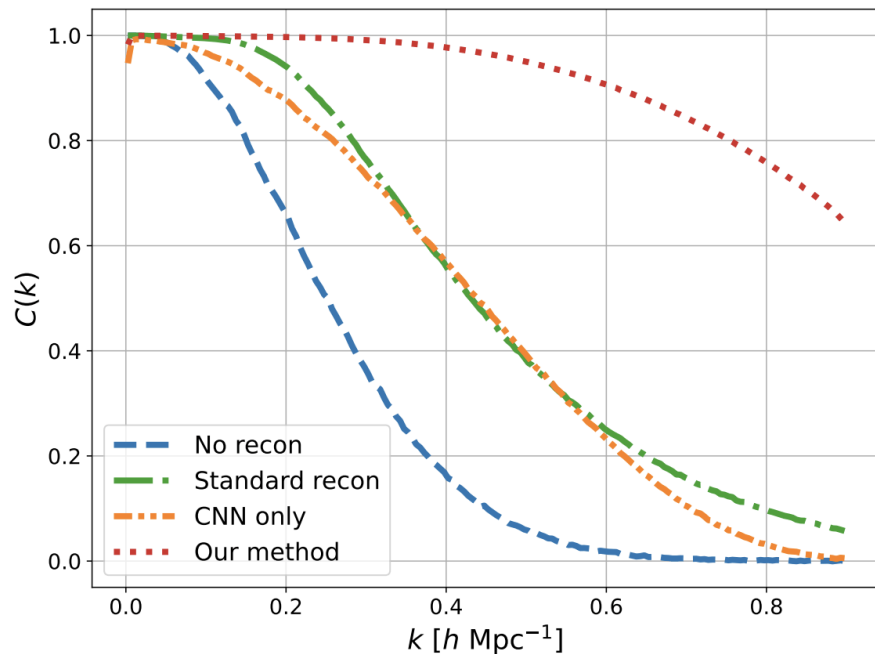
$p(\text{World}|\text{Prompt})$

["Genie 2: A large-scale foundation model"
Parker-Holder et al]

$$p(\text{Drug}|\text{Properties})$$

Learning to sample complex forward models

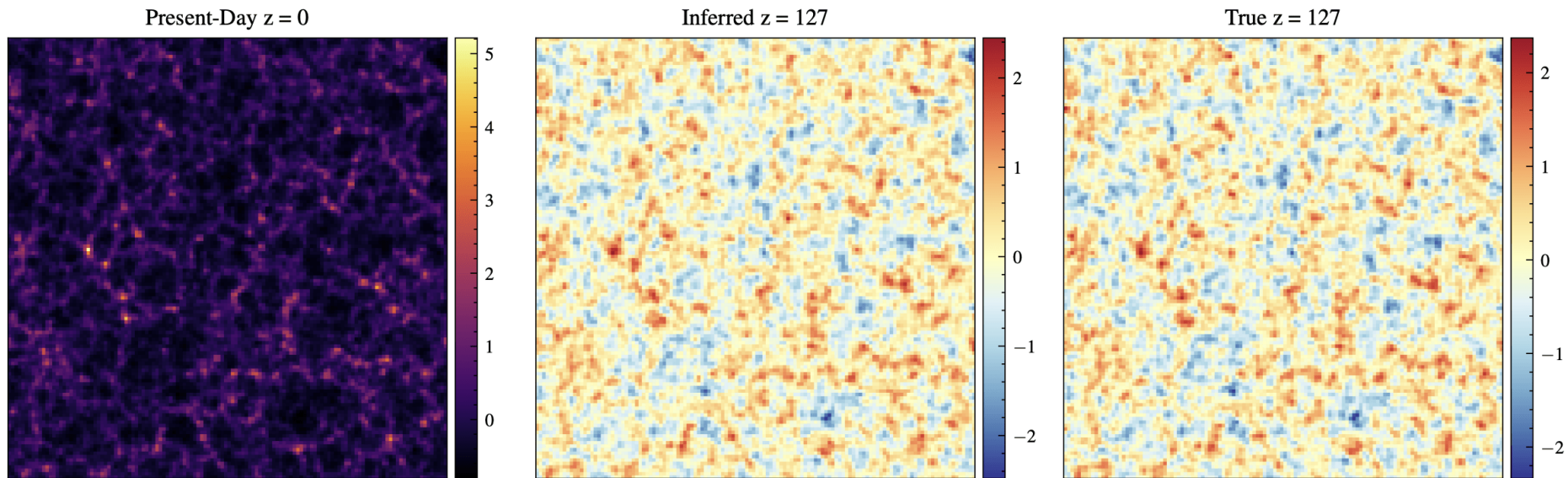
1) Learning the posterior mean with deterministic models



"Reconstructing Cosmological Initial Conditions from Late-Time Structure with Convolutional Neural Networks"
Schallue, Eisenstein

Learning to sample complex forward models

2) Learning to sample with generative models

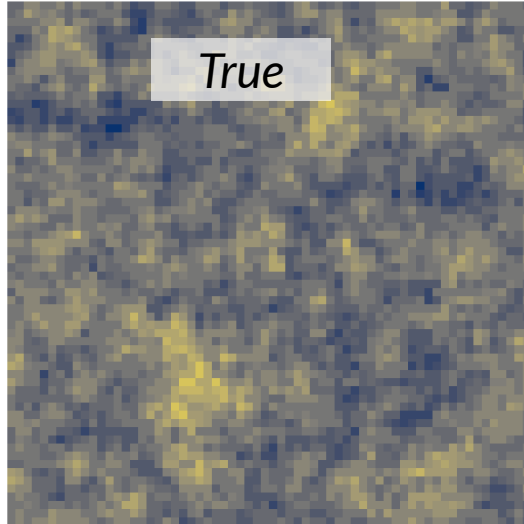


"Posterior Sampling of the Initial Conditions of the Universe from Non-linear Large Scale Structures using Score-Based Generative Models" Legin et al

Stochastic Interpolants

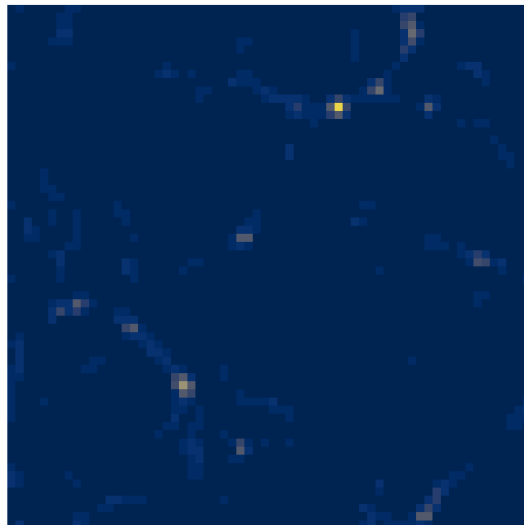
$$p(\delta_{\text{ICs}}, \theta | \delta_{\text{Obs}})$$

True



δ_{ICs}

δ_{Obs}

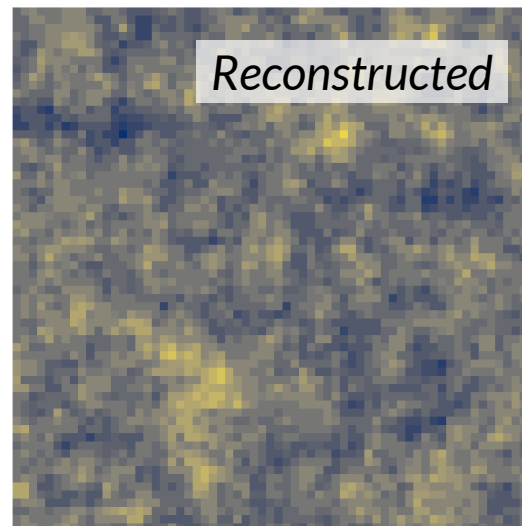
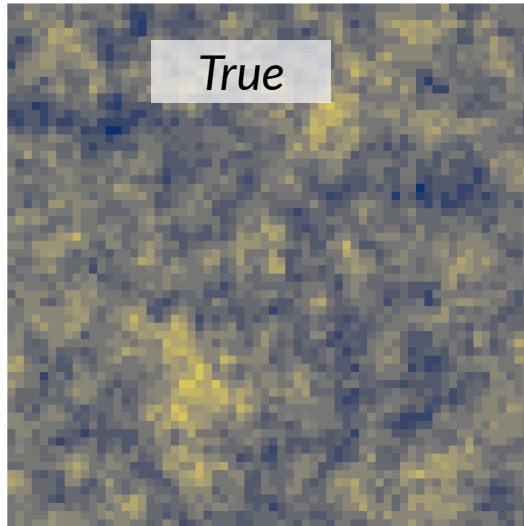


"Joint cosmological
parameter inference and
initial condition
reconstruction with
Stochastic Interpolants"
Cuesta-Lazaro, Bayer,
Albergo et al
NeurIPs ML4PS 2024 Spotlight
talk

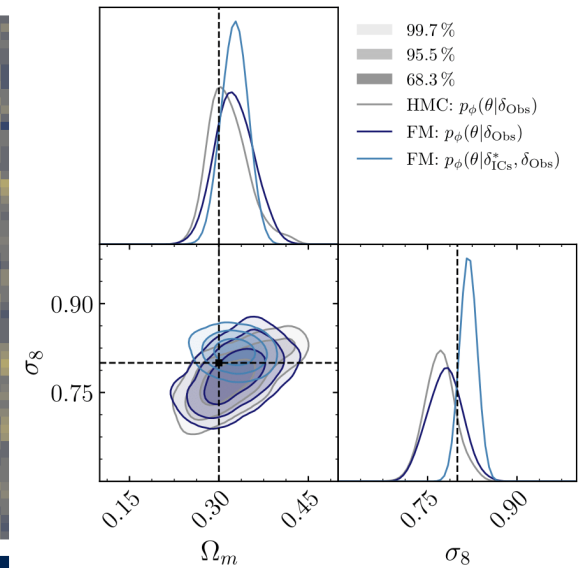
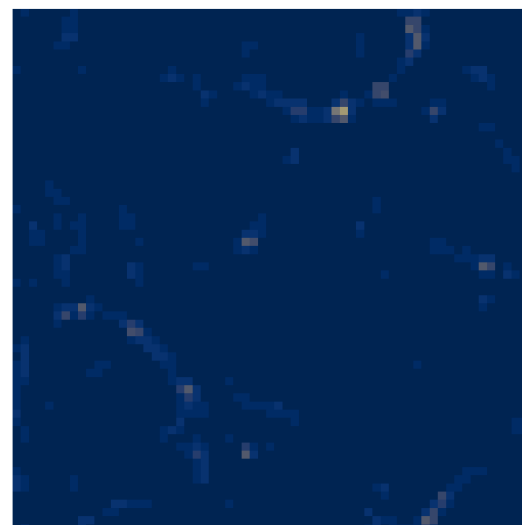
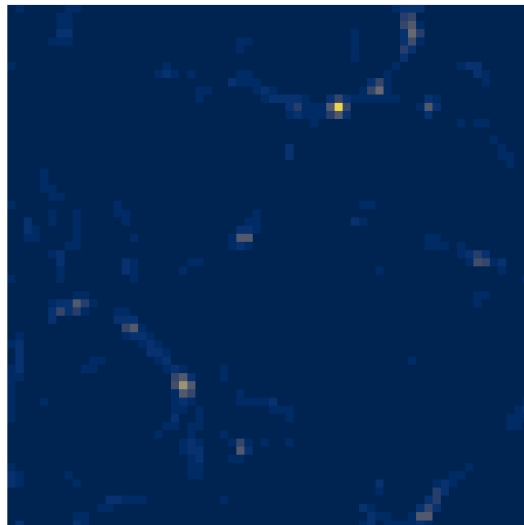
Stochastic Interpolants

$$p(\delta_{\text{ICs}}, \theta | \delta_{\text{Obs}})$$

δ_{ICs}

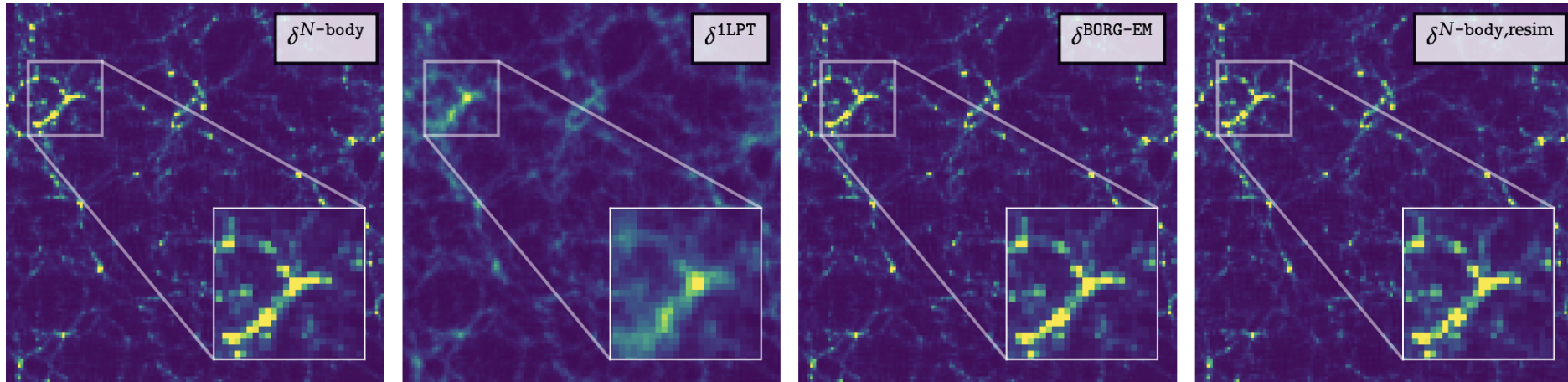


δ_{Obs}



"Joint cosmological
parameter inference and
initial condition
reconstruction with
Stochastic Interpolants"
Cuesta-Lazaro, Bayer,
Albergo et al
NeurIPs ML4PS 2024 Spotlight
talk

Simulators at scale



"Bayesian Inference of Initial Conditions from Non-Linear Cosmic Structures using Field-Level Emulators"

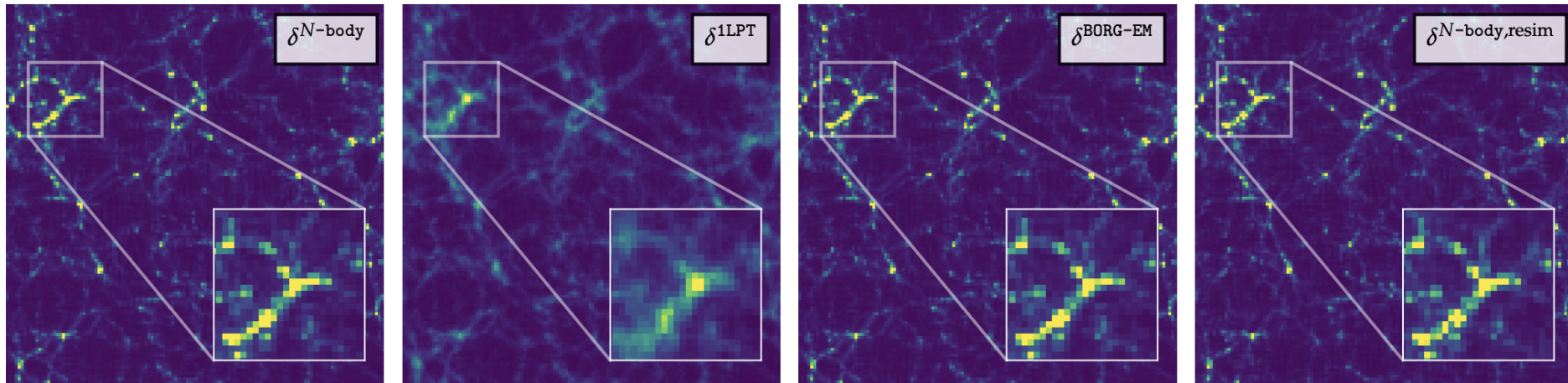
Doerer et al

"Field Level Neural Network Emulator for Cosmological N-body Simulations"

Jamieson et al

Carolina Cuesta-Lazaro IAIFI/MIT @ Sexten 2025

Simulators at scale



(Tested on matter, differences likely worse for galaxies)

"Bayesian Inference of Initial Conditions from Non-Linear Cosmic Structures using Field-Level Emulators"

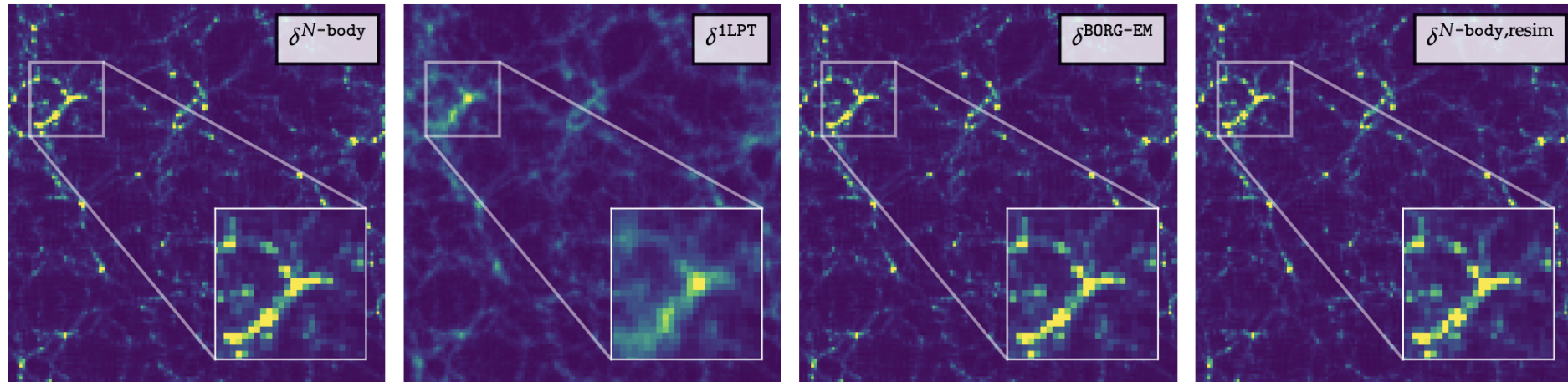
Doerer et al

"Field Level Neural Network Emulator for Cosmological N-body Simulations"

Jamieson et al

Carolina Cuesta-Lazaro IAIFI/MIT @ Sexten 2025

Simulators at scale



(Tested on matter, differences likely worse for galaxies)

Structure Formation Model	t_{forward} [s]	t_{adjoint} [s]
1LPT	$< 10^{-1}$	$< 10^{-1}$
BORG-EM (1LPT + emulator) [★]	1.6	2.6
1LPT + mini-emulator [★]	0.4	0.6
BORG-PM (COLA, $n_{\text{steps}} = 20$, forcesampling= 4) [†]	~ 8	~ 8
N -body (P-Gadget-III, $n_{\text{steps}} = 1664$) [†]	$\sim 6 \times 10^2$	–

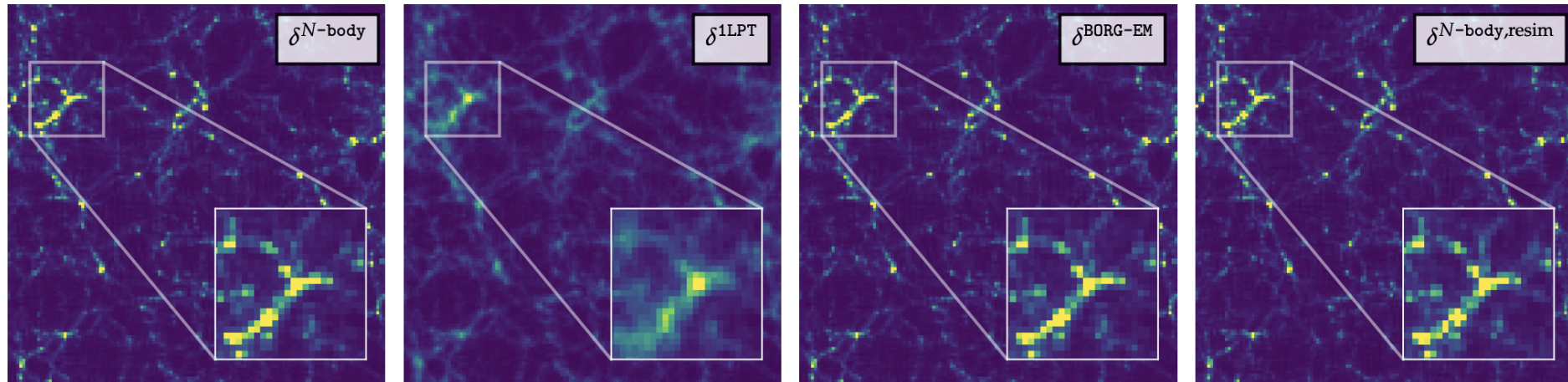
"Bayesian Inference of Initial Conditions from Non-Linear Cosmic Structures using Field-Level Emulators"

Doeser et al

"Field Level Neural Network Emulator for Cosmological N-body Simulations"

Jamieson et al

Simulators at scale



(Tested on matter, differences likely worse for galaxies)

Structure Formation Model	t_{forward} [s]	t_{adjoint} [s]
1LPT	$< 10^{-1}$	$< 10^{-1}$
BORG-EM (1LPT + emulator) [★]	1.6	2.6
1LPT + mini-emulator [★]	0.4	0.6
BORG-PM (COLA, $n_{\text{steps}} = 20$, forcesampling= 4) [†]	~ 8	~ 8
N -body (P-Gadget-III, $n_{\text{steps}} = 1664$) [†]	$\sim 6 \times 10^2$	–

Speed up perhaps not so impressive, but scaling with N -body resolution may be

"Bayesian Inference of Initial Conditions from Non-Linear Cosmic Structures using Field-Level Emulators"

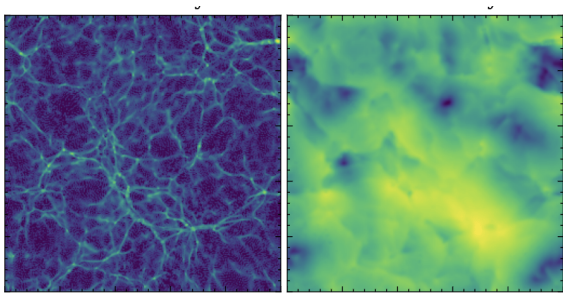
Doeser et al

"Field Level Neural Network Emulator for Cosmological N-body Simulations"

Jamieson et al

Hydro Simulators at scale

Particle Mesh for Gravity



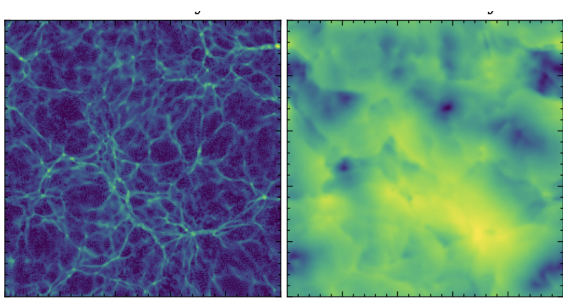
CAMELS Volumes

$25h^{-1}\text{Mpc}$

"BaryonBridge: Interpolants
models for fast
hydrodynamical simulations"
Horowitz, Cuesta-Lazaro,
Yehia (in prep)

Hydro Simulators at scale

Particle Mesh for Gravity



CAMELS Volumes

$25h^{-1}\text{Mpc}$

$+C, A$

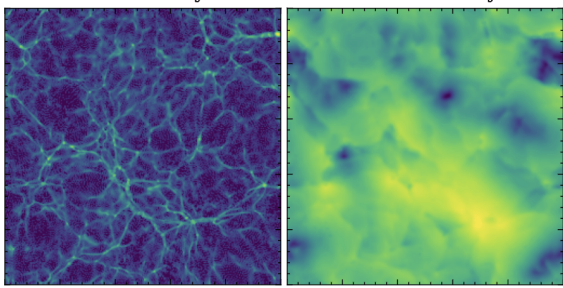
1000 boxes with varying
cosmology and *feedback* models

"BaryonBridge: Interpolants
models for fast
hydrodynamical simulations"

Horowitz, Cuesta-Lazaro,
Yehia (in prep)

Hydro Simulators at scale

Particle Mesh for Gravity



$$p(\text{Baryons}|\text{DM}, \mathcal{C}, \mathcal{A})$$



CAMELS Volumes

$25h^{-1}\text{Mpc}$

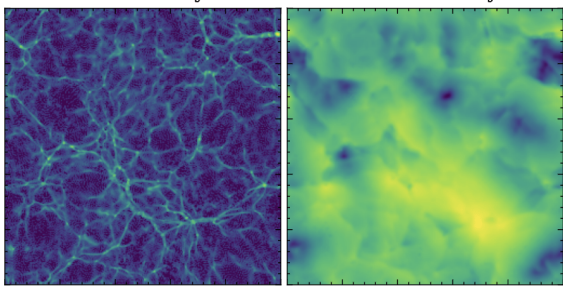
$+\mathcal{C}, \mathcal{A}$

1000 boxes with varying
cosmology and *feedback* models

"BaryonBridge: Interpolants
models for fast
hydrodynamical simulations"
Horowitz, Cuesta-Lazaro,
Yehia (in prep)

Hydro Simulators at scale

Particle Mesh for Gravity



CAMELS Volumes

$25h^{-1}\text{Mpc}$

$p(\text{Baryons}|\text{DM}, \mathcal{C}, \mathcal{A})$



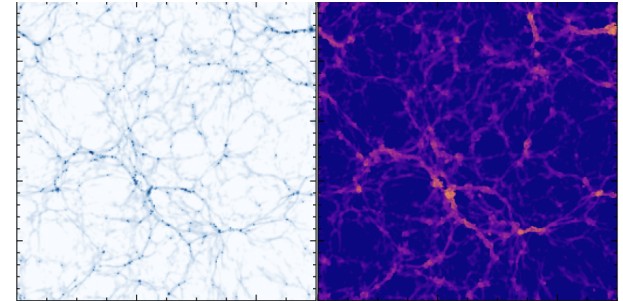
$+\mathcal{C}, \mathcal{A}$

1000 boxes with varying
cosmology and *feedback* models

Gas Properties

Density

Temperature

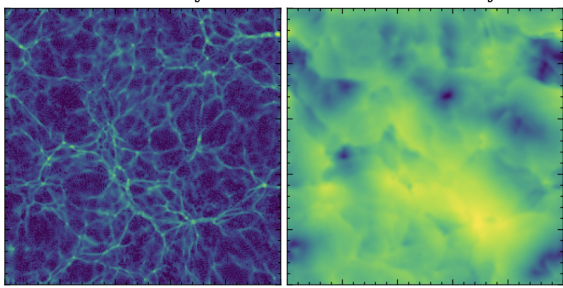


"BaryonBridge: Interpolants
models for fast
hydrodynamical simulations"

Horowitz, Cuesta-Lazaro,
Yehia (in prep)

Hydro Simulators at scale

Particle Mesh for Gravity



CAMELS Volumes

$25h^{-1}\text{Mpc}$

$$p(\text{Baryons}|\text{DM}, \mathcal{C}, \mathcal{A})$$



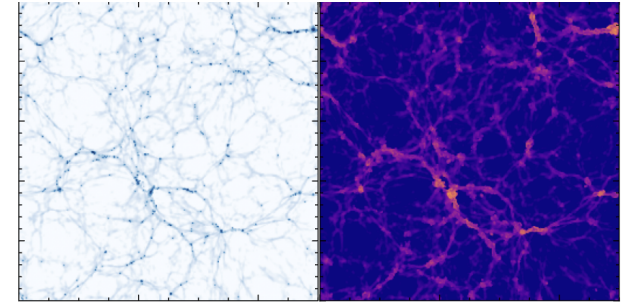
$+\mathcal{C}, \mathcal{A}$

1000 boxes with varying
cosmology and *feedback* models

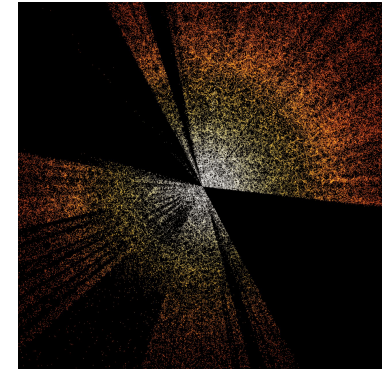
Gas Properties

Density

Temperature



Galaxy Distribution

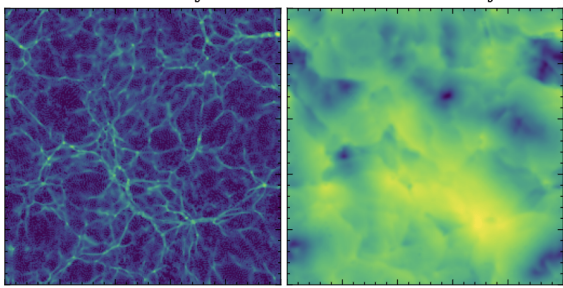


"BaryonBridge: Interpolants
models for fast
hydrodynamical simulations"

Horowitz, Cuesta-Lazaro,
Yehia (in prep)

Hydro Simulators at scale

Particle Mesh for Gravity



CAMELS Volumes

$25h^{-1}\text{Mpc}$

1000 boxes with varying
cosmology and *feedback* models

$$p(\text{Baryons}|\text{DM}, \mathcal{C}, \mathcal{A})$$

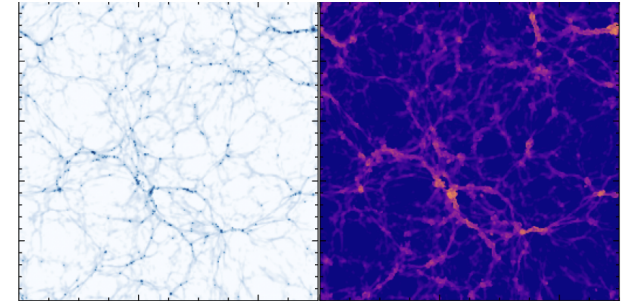


$+\mathcal{C}, \mathcal{A}$

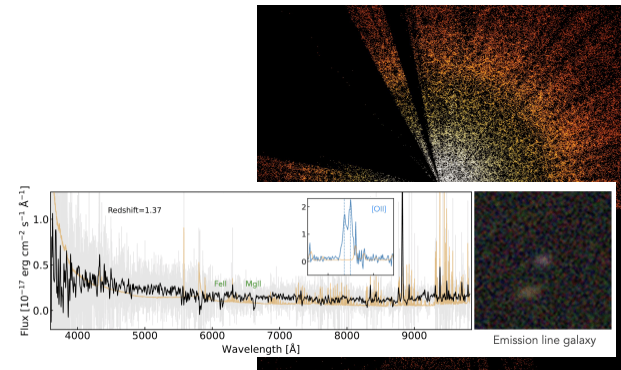
Gas Properties

Density

Temperature



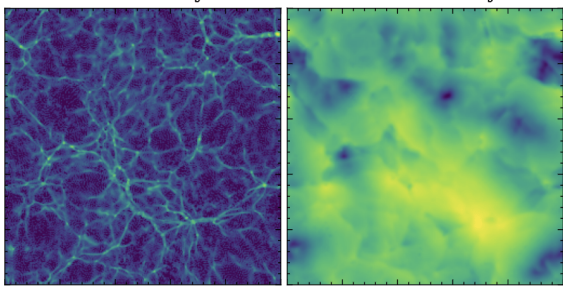
Galaxy Distribution



"BaryonBridge: Interpolants
models for fast
hydrodynamical simulations"
Horowitz, Cuesta-Lazaro,
Yehia (in prep)

Hydro Simulators at scale

Particle Mesh for Gravity



CAMELS Volumes

$25h^{-1}\text{Mpc}$

$$p(\text{Baryons}|\text{DM}, \mathcal{C}, \mathcal{A})$$



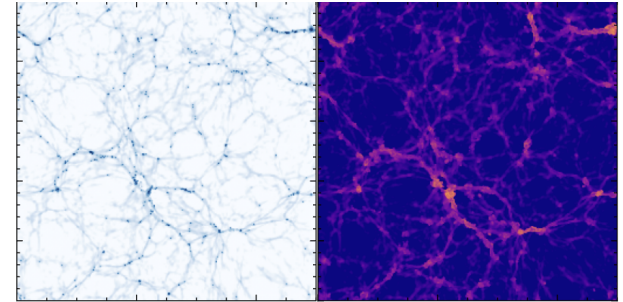
$+\mathcal{C}, \mathcal{A}$

1000 boxes with varying
cosmology and *feedback* models

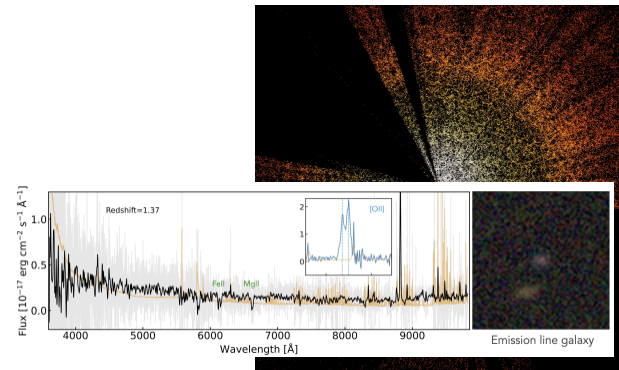
Gas Properties

Density

Temperature



Galaxy Distribution



Current model optimised for Lyman Alpha forest

7 GPU minutes for a 50 Mpc simulation

130 million CPU core hours for TNG50

Carolina Cuesta-Lazaro IAFI/MIT @ Sexten 2025

"BaryonBridge: Interpolants
models for fast
hydrodynamical simulations"
Horowitz, Cuesta-Lazaro,
Yehia (in prep)

The Roadmap

The Roadmap

- 1) Need to develop *better validation metrics* (requires better validation suites)

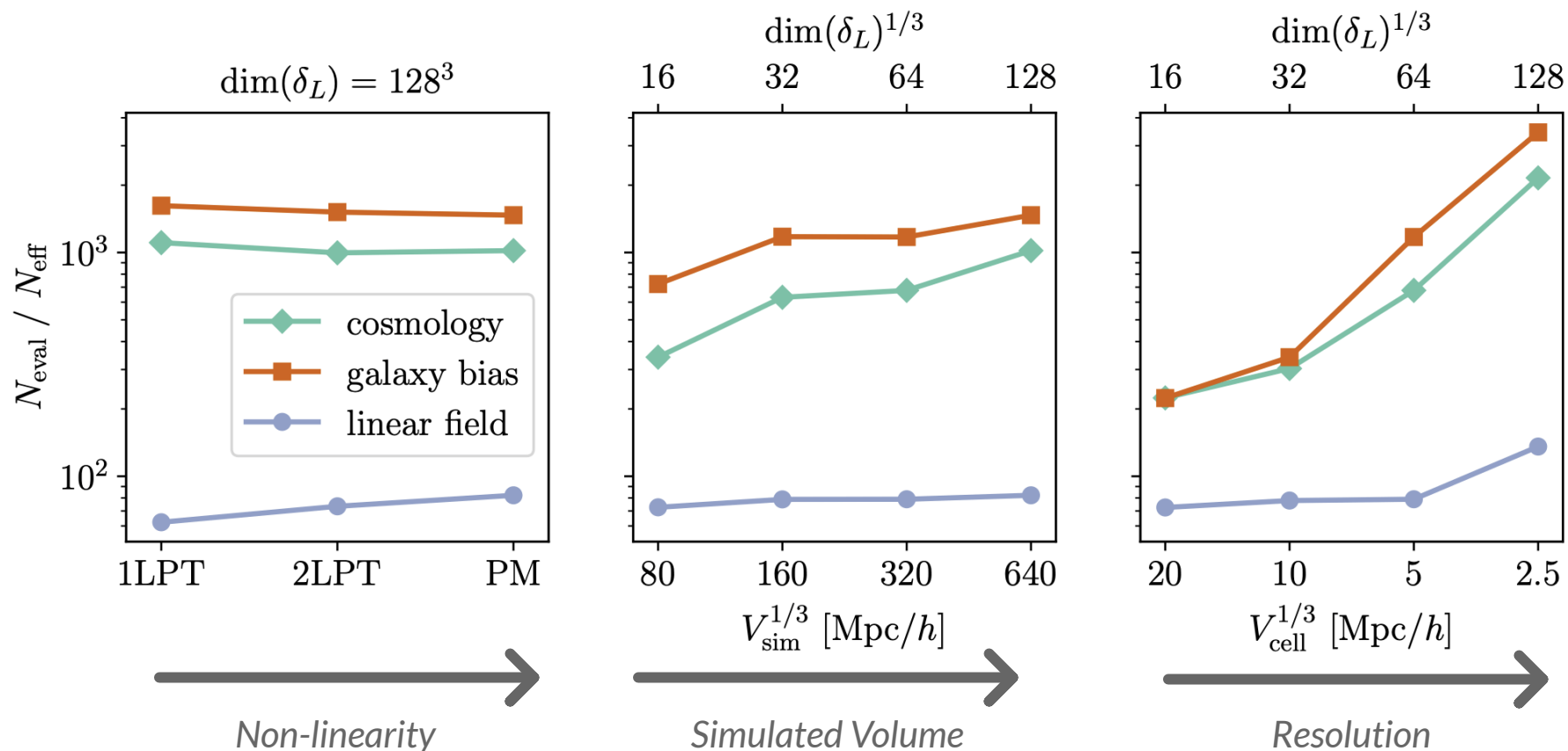
The Roadmap

- 1) Need to develop *better validation metrics* (requires better validation suites)
- 2) Assess the robustness of **field**-level inference via *parameter-masked* mock challenges in realistic scenarios (example Beyond2pt)

The Roadmap

- 1) Need to develop *better validation metrics* (requires better validation suites)
 - 2) Assess the robustness of **field**-level inference via *parameter-masked* mock challenges in realistic scenarios (example Beyond2pt)
 - 3) Development of open source ecosystems for more plug and play models
- Field level analysis too complex for one group to develop a robust framework!*

FLI Benchmarks

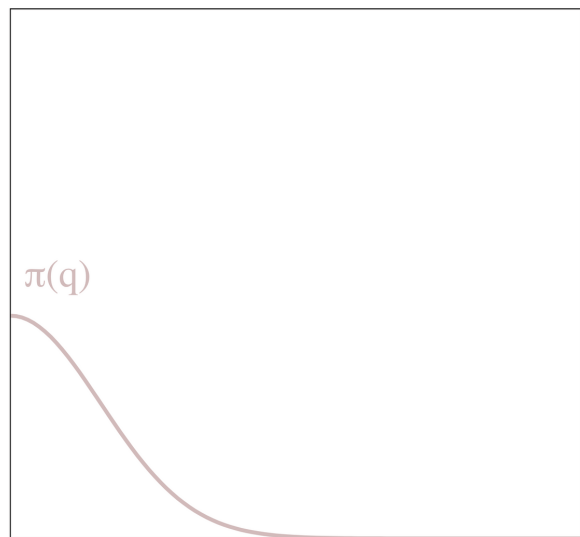


"Benchmarking field-level cosmological inference from galaxy redshift surveys" Simon-Onfroy, Lanusse, De Mattia

Sampling Methods

Why do we need gradients to sample in high dimensions?

$$\mathbb{E}_{\pi}[f] = \int_Q dq \, \pi(q) f(q)$$



$|q - q_{\text{Mode}}|$

Distance from the mode

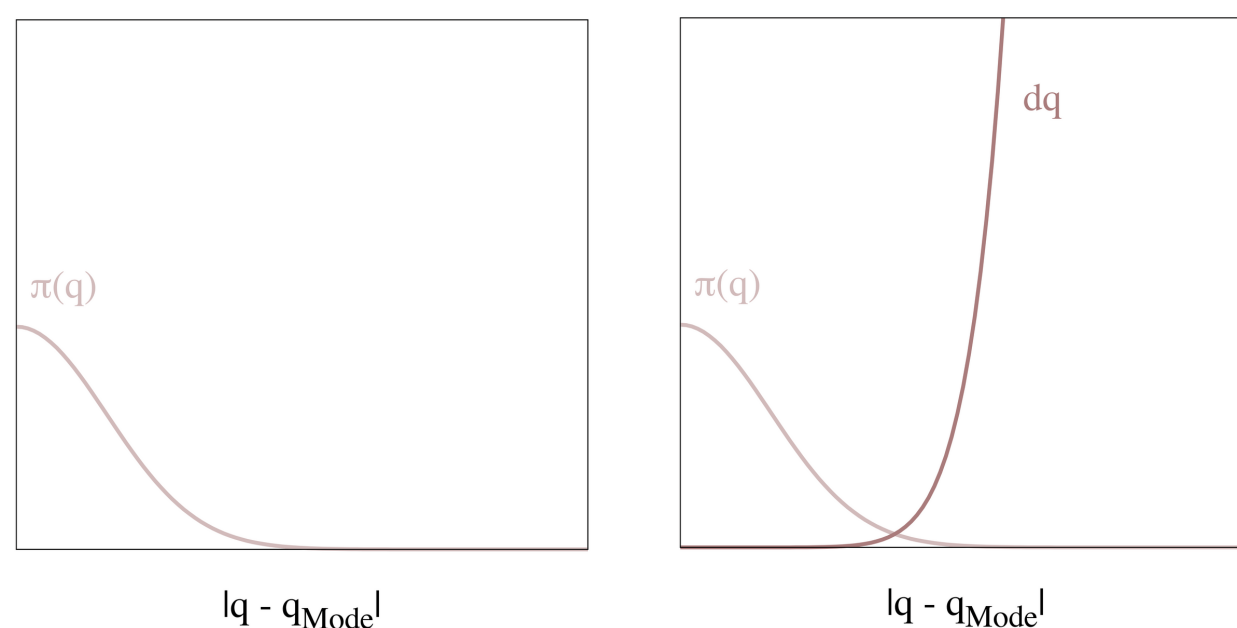
Image Credit: "Probabilistic Computation" Michael Betancourt

https://betanalpha.github.io/assets/case_studies/probabilistic_computation.html

Sampling Methods

Why do we need gradients to sample in high dimensions?

$$\mathbb{E}_{\pi}[f] = \int_Q dq \, \pi(q) f(q)$$



Distance from the mode

Image Credit: "Probabilistic Computation" Michael Betancourt

https://betanalpha.github.io/assets/case_studies/probabilistic_computation.html

Sampling Methods

Why do we need gradients to sample in high dimensions?

$$\mathbb{E}_{\pi}[f] = \int_Q dq \, \pi(q) f(q)$$

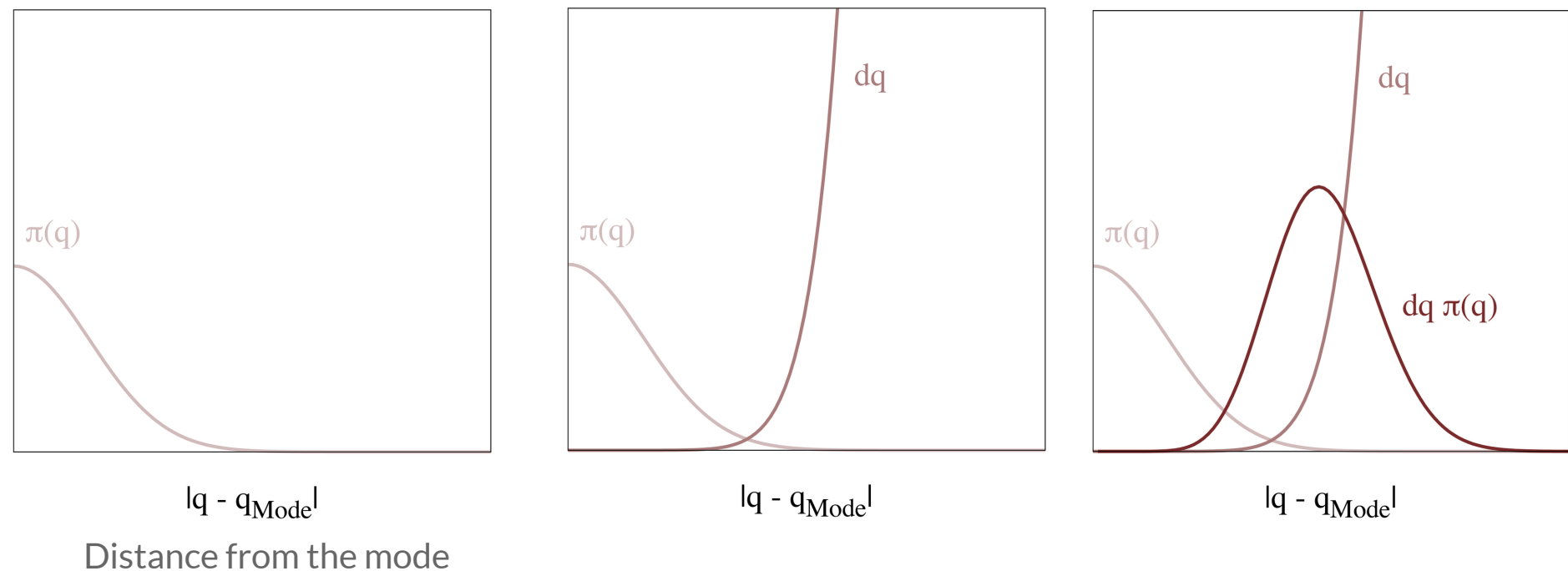


Image Credit: "Probabilistic Computation" Michael Betancourt

https://betanalpha.github.io/assets/case_studies/probabilistic_computation.html

Sampling Methods

Why do we need gradients to sample in high dimensions?

$$\mathbb{E}_{\pi}[f] = \int_Q \underbrace{dq \pi(q)}_{\text{Typical Set}} f(q)$$

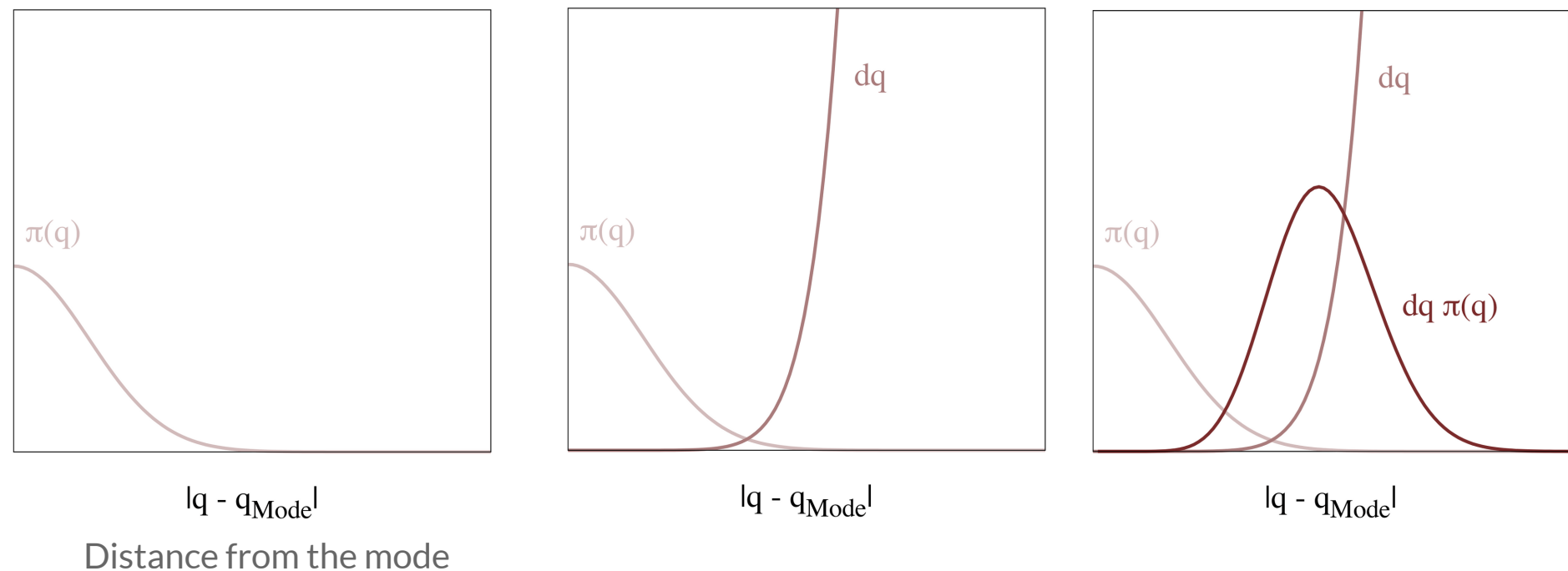


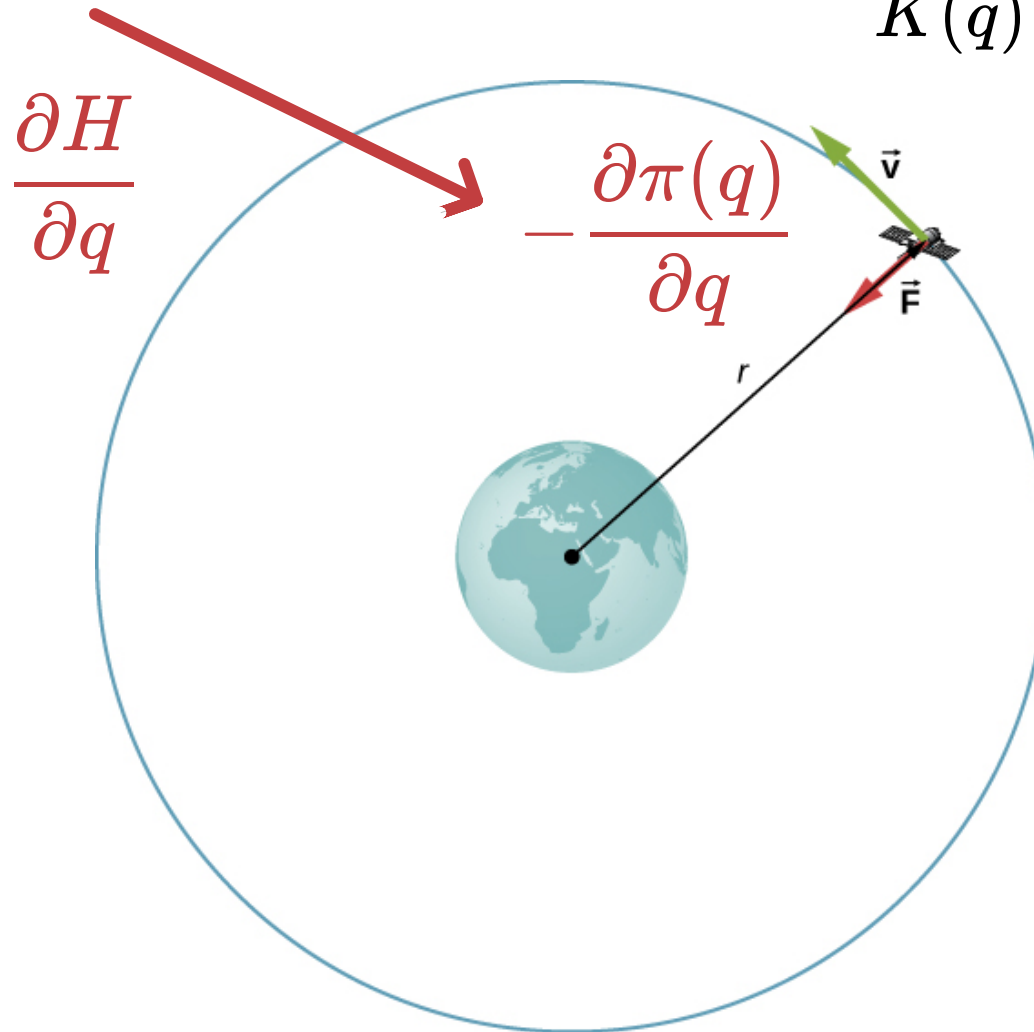
Image Credit: "Probabilistic Computation" Michael Betancourt

https://betanalpha.github.io/assets/case_studies/probabilistic_computation.html

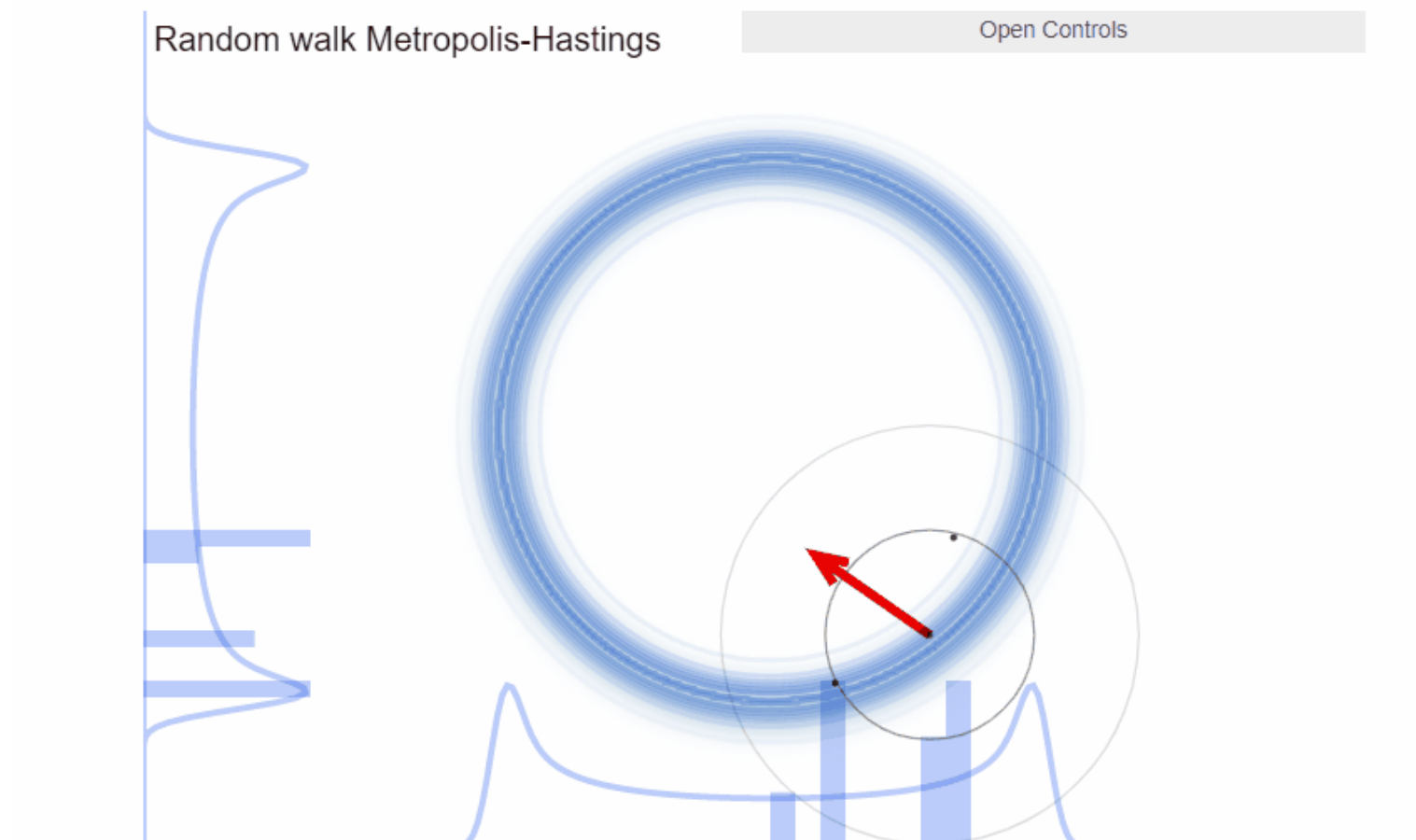
$$H(q, p) = U(q) + K(p)$$

$$U(p) = -\log \pi(p)$$

$$K(q) = \frac{1}{2}p^T p$$



Random Walks are isotropic, will bring you outside the typical set



Can we find a sampling algorithm that keeps samples in the typical set whilst moving far?

Image Credit: "The Markov-chain Monte Carlo Interactive Gallery" Chi Feng
<https://chi-feng.github.io/mcmc-demo/>

Hamiltonian Monte Carlo

Find trajectories of similar Energy (p) that are far away

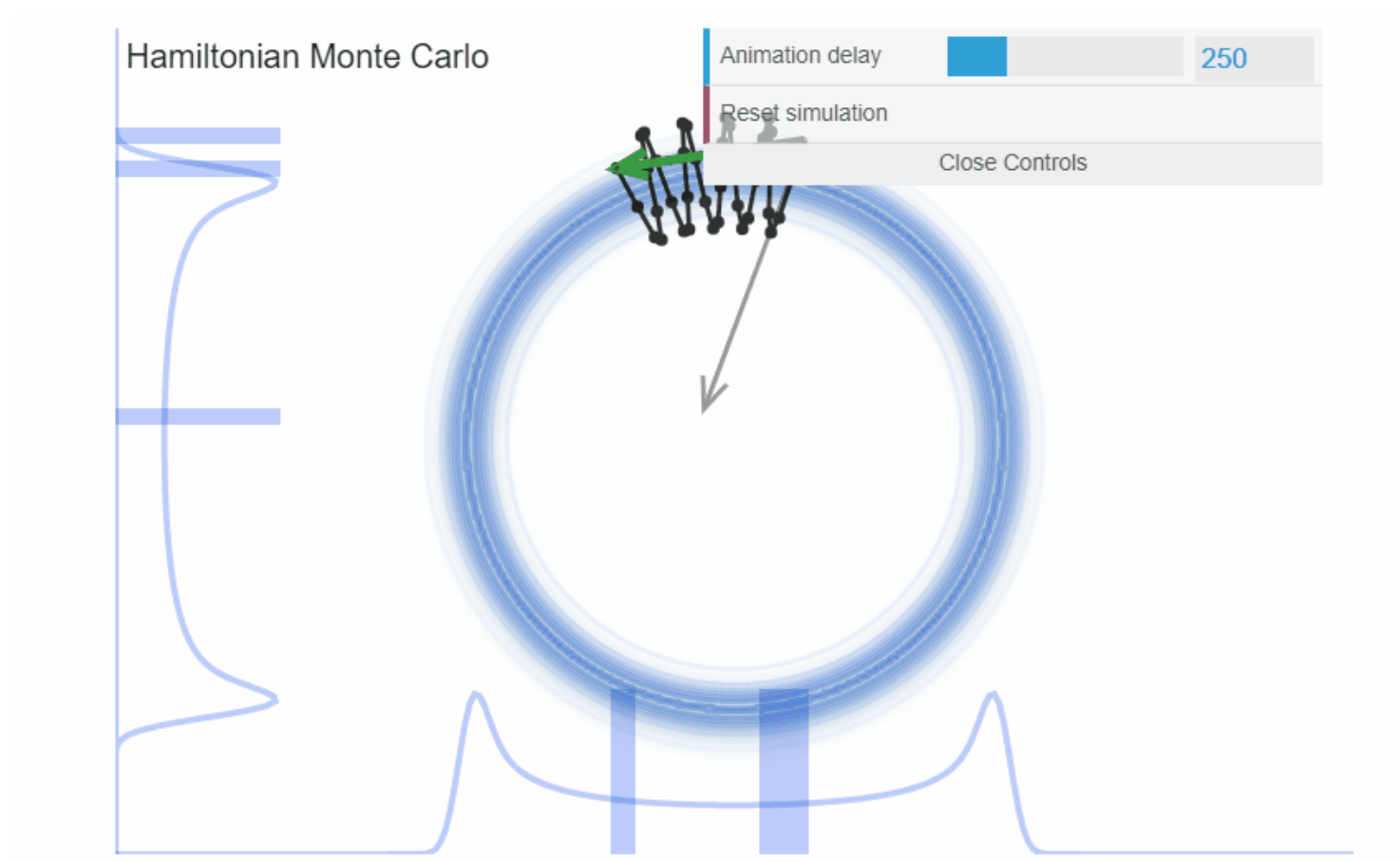
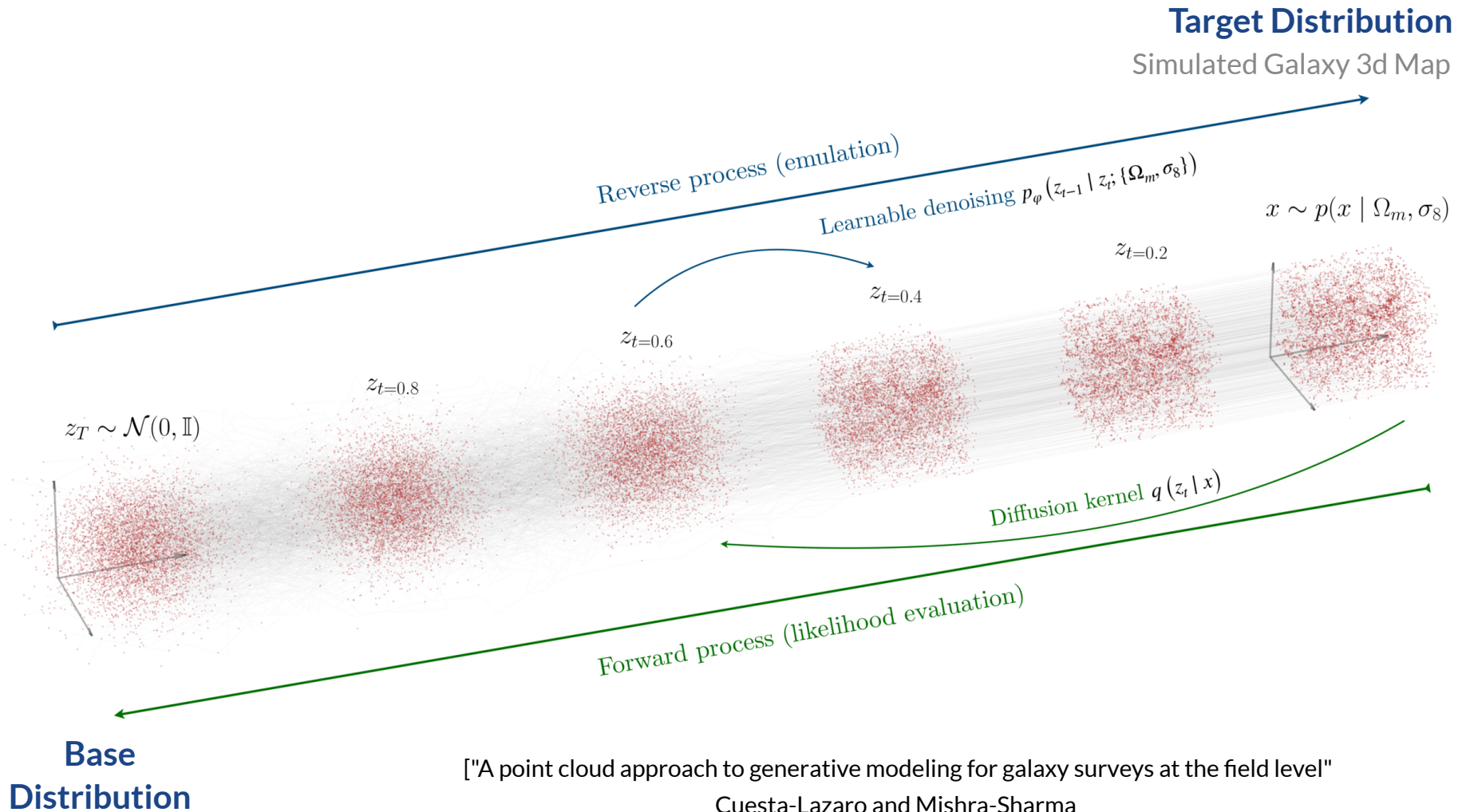


Image Credit: "The Markov-chain Monte Carlo Interactive Gallery" Chi Feng
<https://chi-feng.github.io/mcmc-demo/>

Learning likelihoods at the field-level

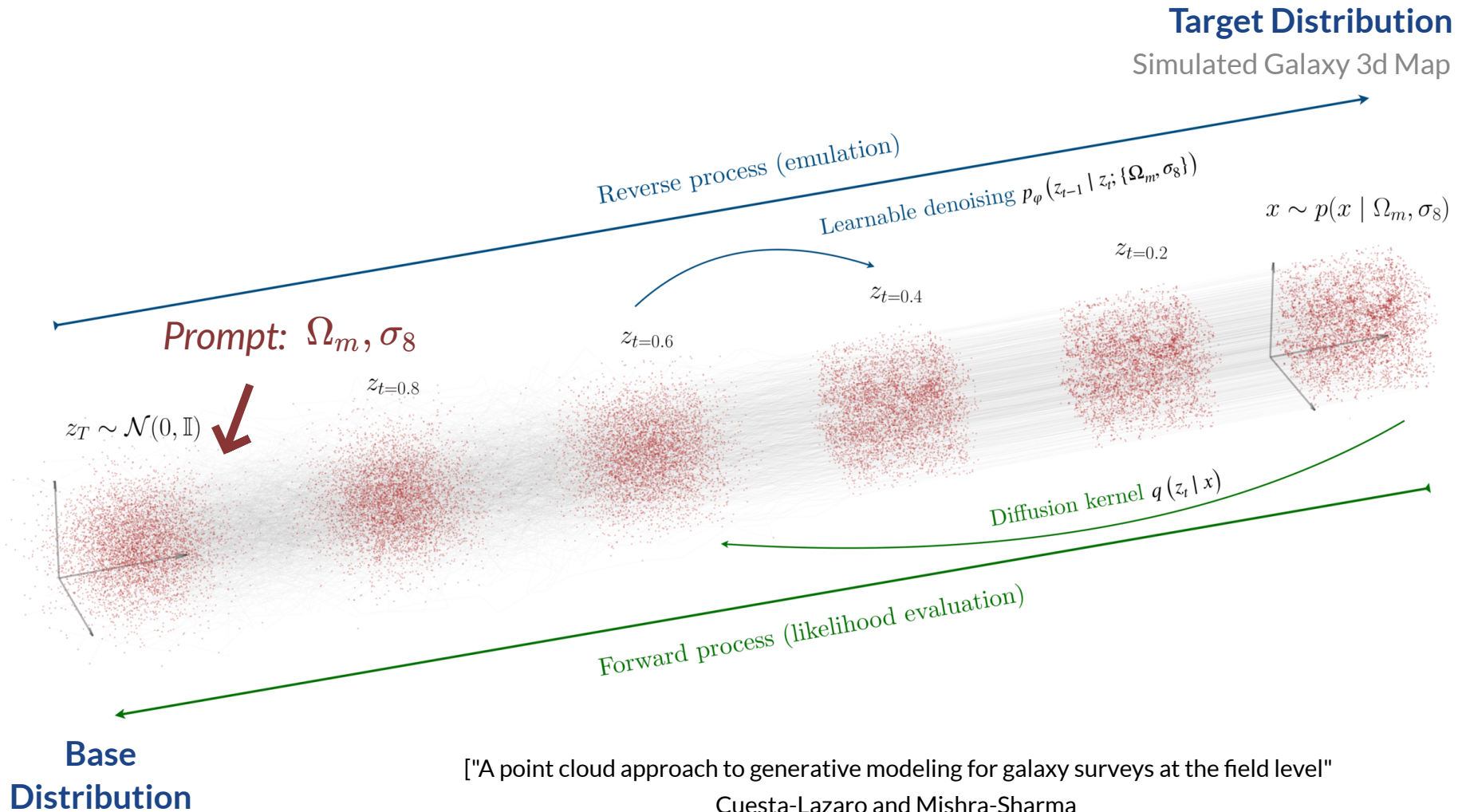


["A point cloud approach to generative modeling for galaxy surveys at the field level"]

Cuesta-Lazaro and Mishra-Sharma

International Conference on Machine Learning ICML AI4Astro 2023, Spotlight talk, arXiv:2311.17141]

Learning likelihoods at the field-level



["A point cloud approach to generative modeling for galaxy surveys at the field level"]

Cuesta-Lazaro and Mishra-Sharma

International Conference on Machine Learning ICML AI4Astro 2023, Spotlight talk, arXiv:2311.17141]