MACHINE LEARNING FOR ASTROPHYSICS
2ND EDITION
CATANIA, 8-12 JULY, 2024

Extracting optimal information from cosmological surveys *with* **field-level inference** *and* **joint analyses**

Adrian Bayer

Princeton University / Simons Foundation

Catania
11 July 2024

PRINCETON UNIVERSITY

SIMONS FOUNDATION
FLATIRON INSTITUTE

MACHINE LEARNING FOR MEDICINE

2ND EDITION CATANIA, 8-12 JULY, 2024

Extracting optimal information from cosmological surveys *with* **field-level inference** *and* **joint analyses**

Adrian Bayer

Princeton University / Simons Foundation

Catania
11 July 2024

PRINCETON UNIVERSITY

SIMONS FOUNDATION

FLATIRON INSTITUTE

Google Translate with an Italian doctor

Google Translate

with an Italian doctor

The right Ok, go away Tiger again. Purulent tonsillitis Purlo in tonsillitis jokes because Forlan Translate Ok we need antibiotics, ok?

Google Translate with an Italian doctor

The right Ok, go away Tiger again. Purulent tonsillitis Purlo in tonsillitis jokes because Forlan Translate Ok we need antibiotics, ok?

I'm a singer 97 with 96 of frequencies 96 6ix9ine FC 90 oxygen 96 Epression snormann This is Twenty hundred 8.16pm, okay? He landed in Deutschland This World, do you understand?

Google Translate

with an Italian doctor

The right Ok, go away Tiger again. Purulent tonsillitis Purlo in tonsillitis jokes because Forlan Translate Ok we need antibiotics, ok?

I'm a singer 97 with 96 of frequencies 96 6ix9ine FC 90 oxygen 96 Epression snormann This is Twenty hundred 8.16pm, okay? He landed in Deutschland This World, do you understand?

ok, Enough, ok, you can go out. As long as you don't work in the refrigerator nowork in the refrigerator in the refrigerator, right?

Google Translate with an Italian doctor

The right Ok, go away Tiger again. Purulent tonsillitis Purlo in tonsillitis jokes because Forlan Translate Ok we need antibiotics, ok?

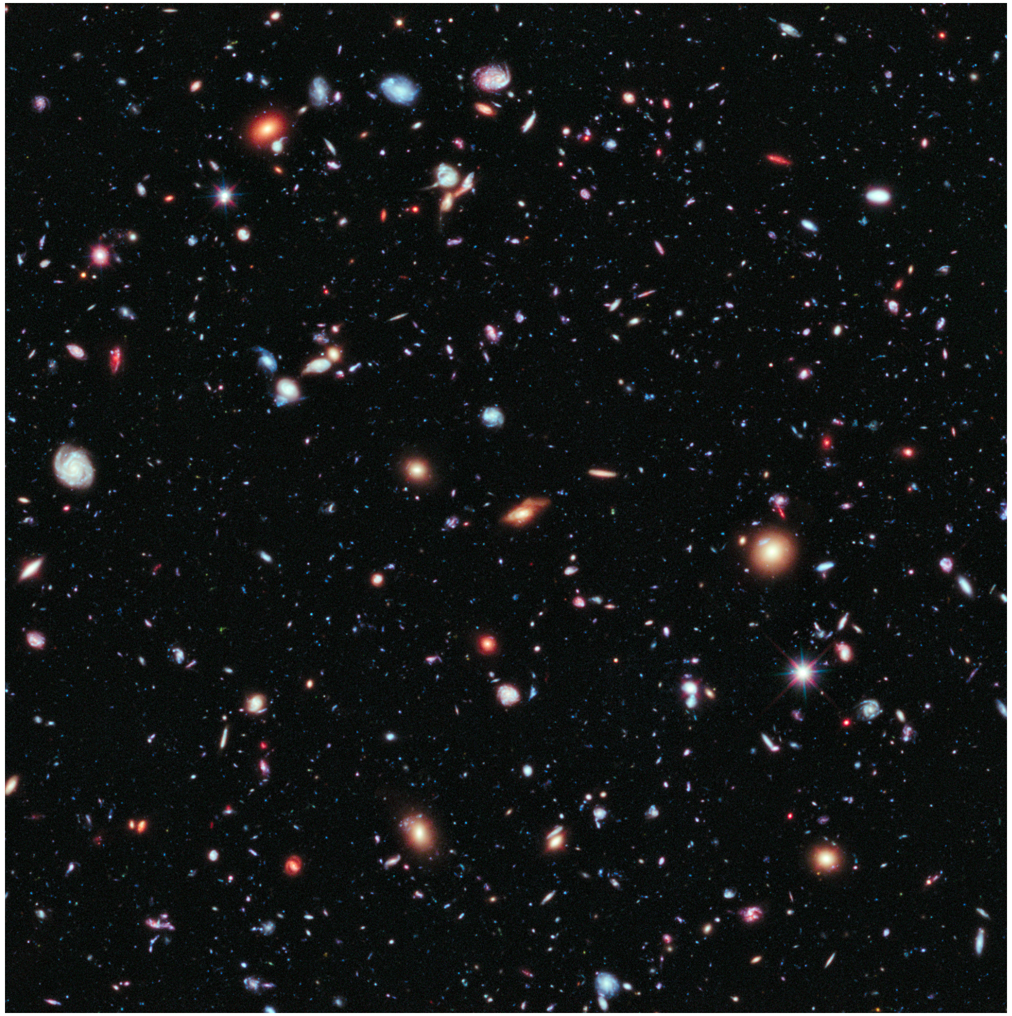
I'm a singer 97 with 96 of frequencies 96 6ix9ine FC 90 oxygen 96 Epression snormann This is Twenty hundred 8.16pm, okay? He landed in Deutschland This World, do you understand?

ok, Enough, ok, you can go out. As long as you don't work in the refrigerator nowork in the refrigerator in the refrigerator, right?

Should I point out Cold or Sud? Just don't talk and if not Speak to Filly. Don't talk so much.

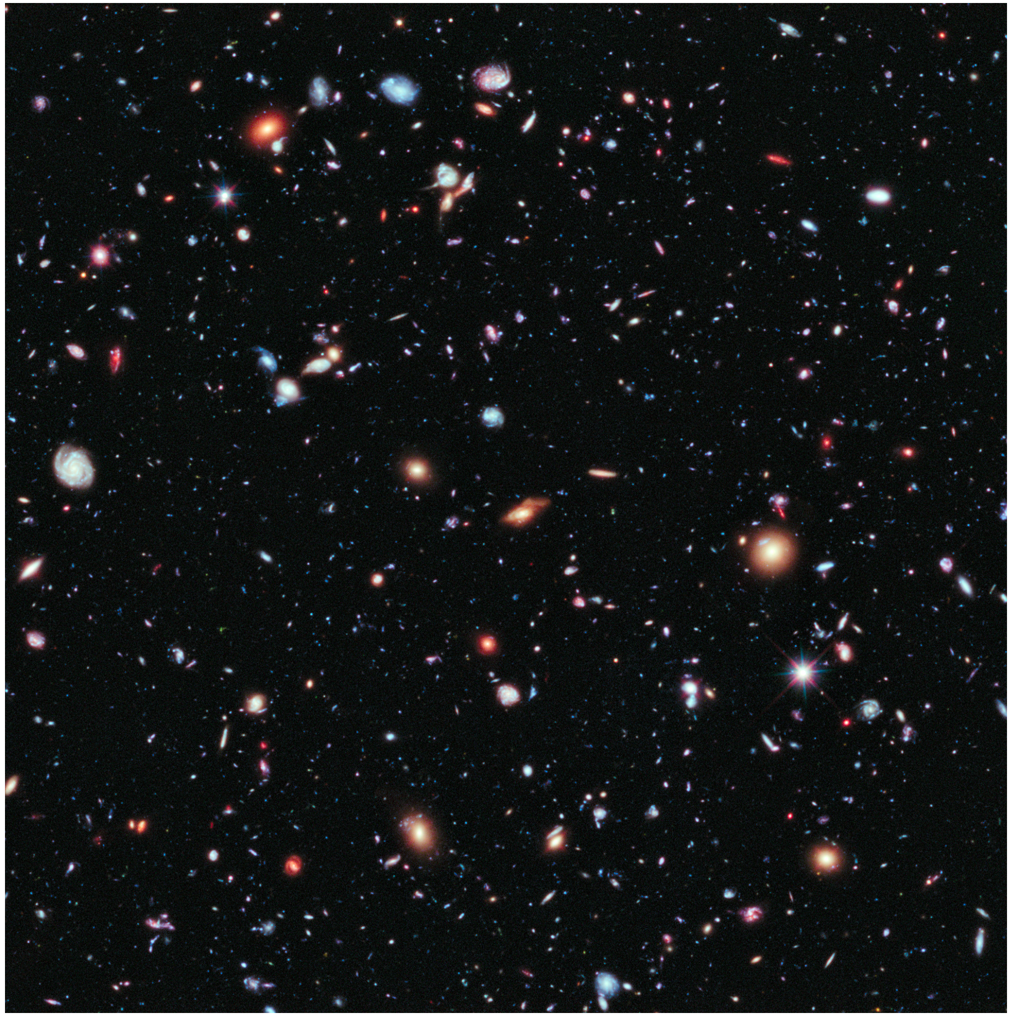
Cosmological Field-Level Inference *with* Microcanonical Langevin Monte Carlo

Cosmological
Field-Level Inference
with
Microcanonical
Langevin Monte Carlo



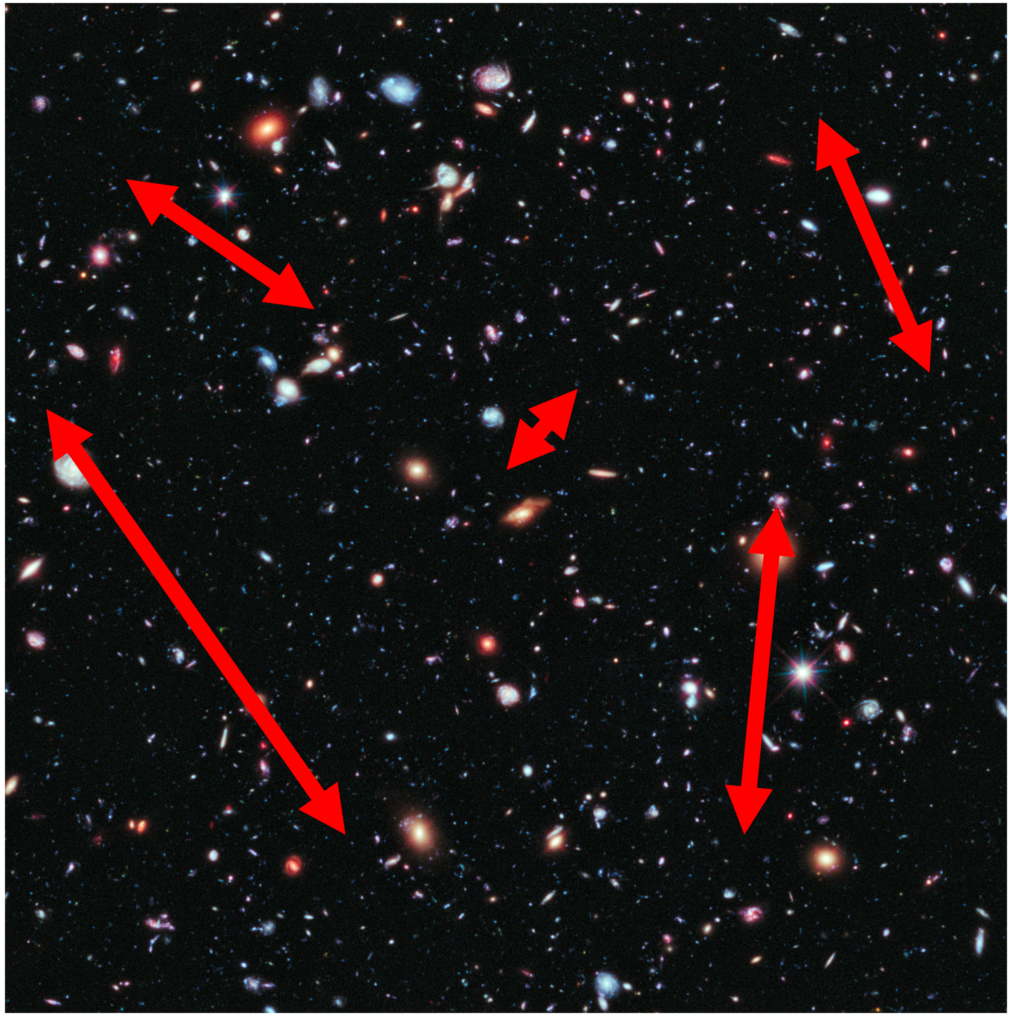
S_8 H_0 Ω_Λ f_{NL} M_ν

cosmological tensions? dark energy primordial NG neutrino mass



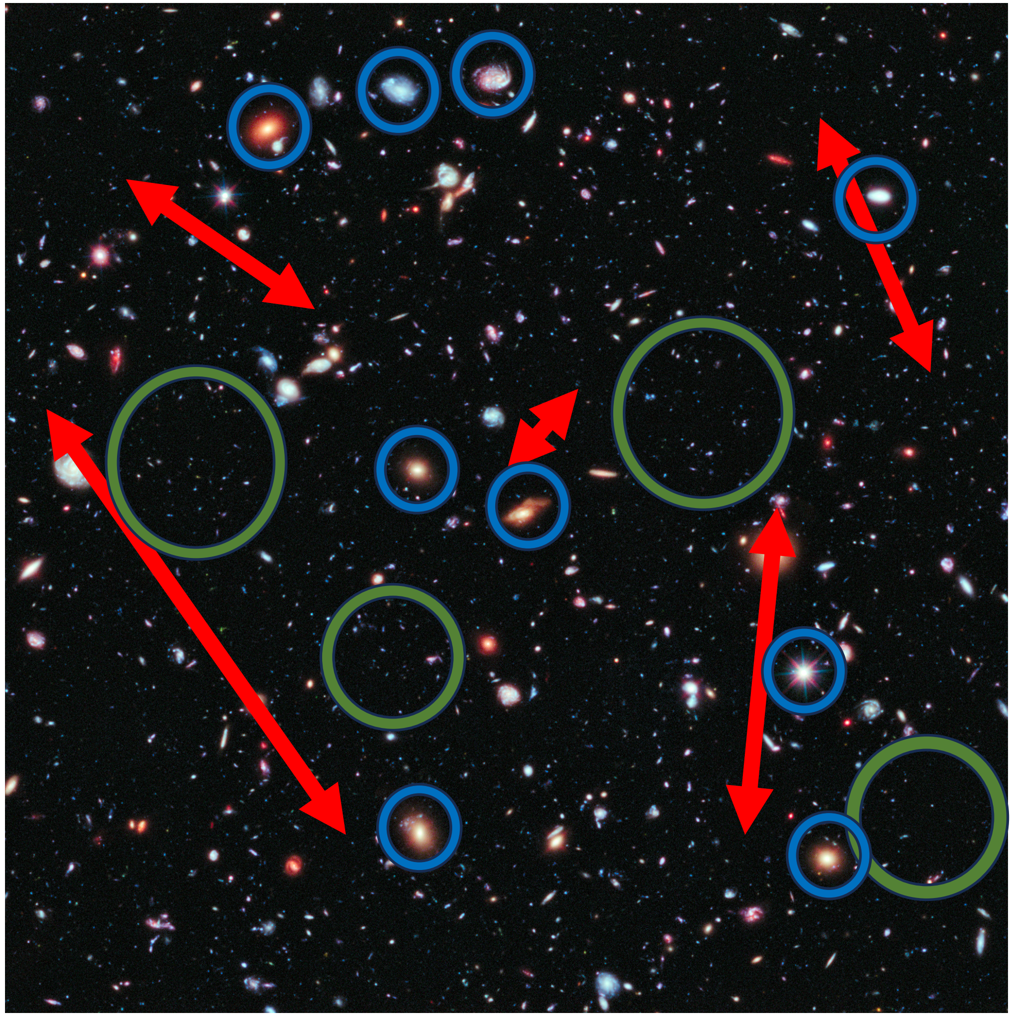
Traditional cosmology uses 2-pt information

but this is no longer optimal as we probe smaller scales



Higher-order statistics

can provide information beyond the 2-pt



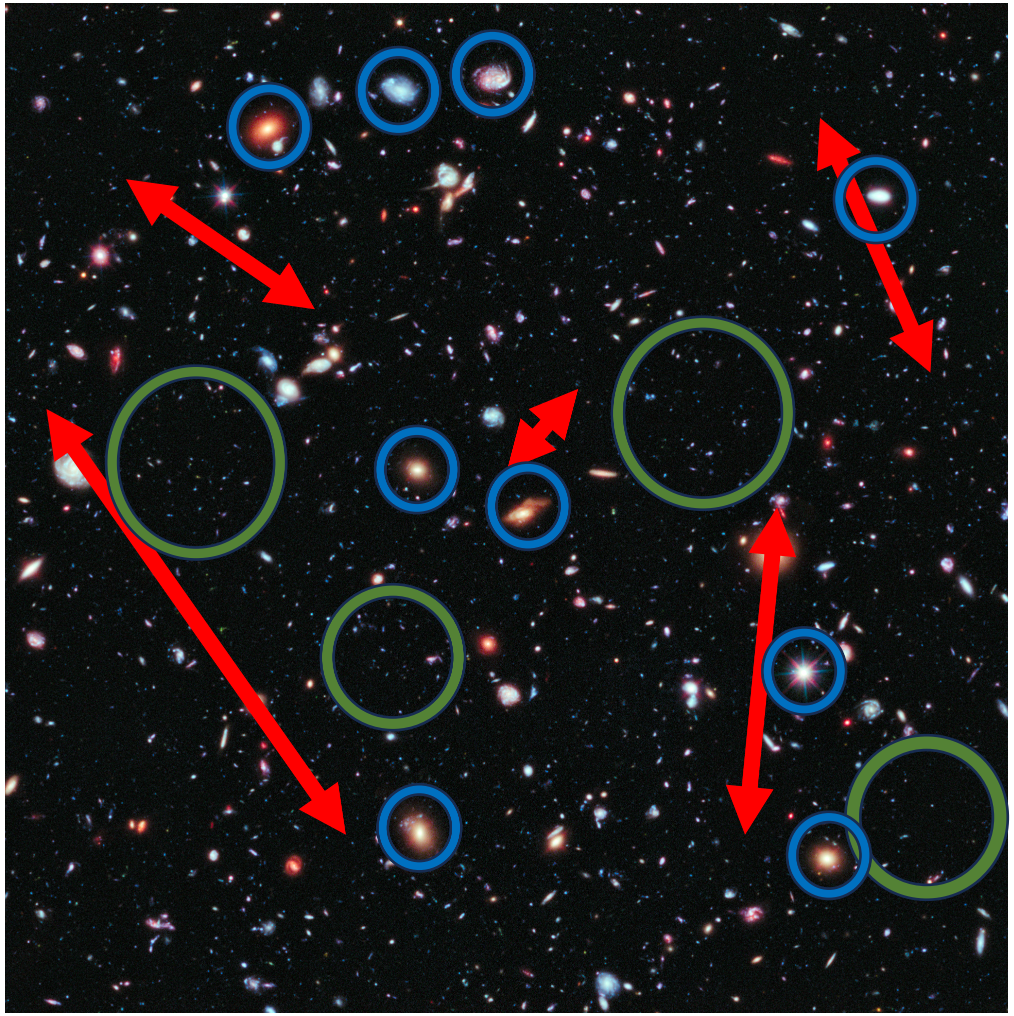
Halo Mass Function

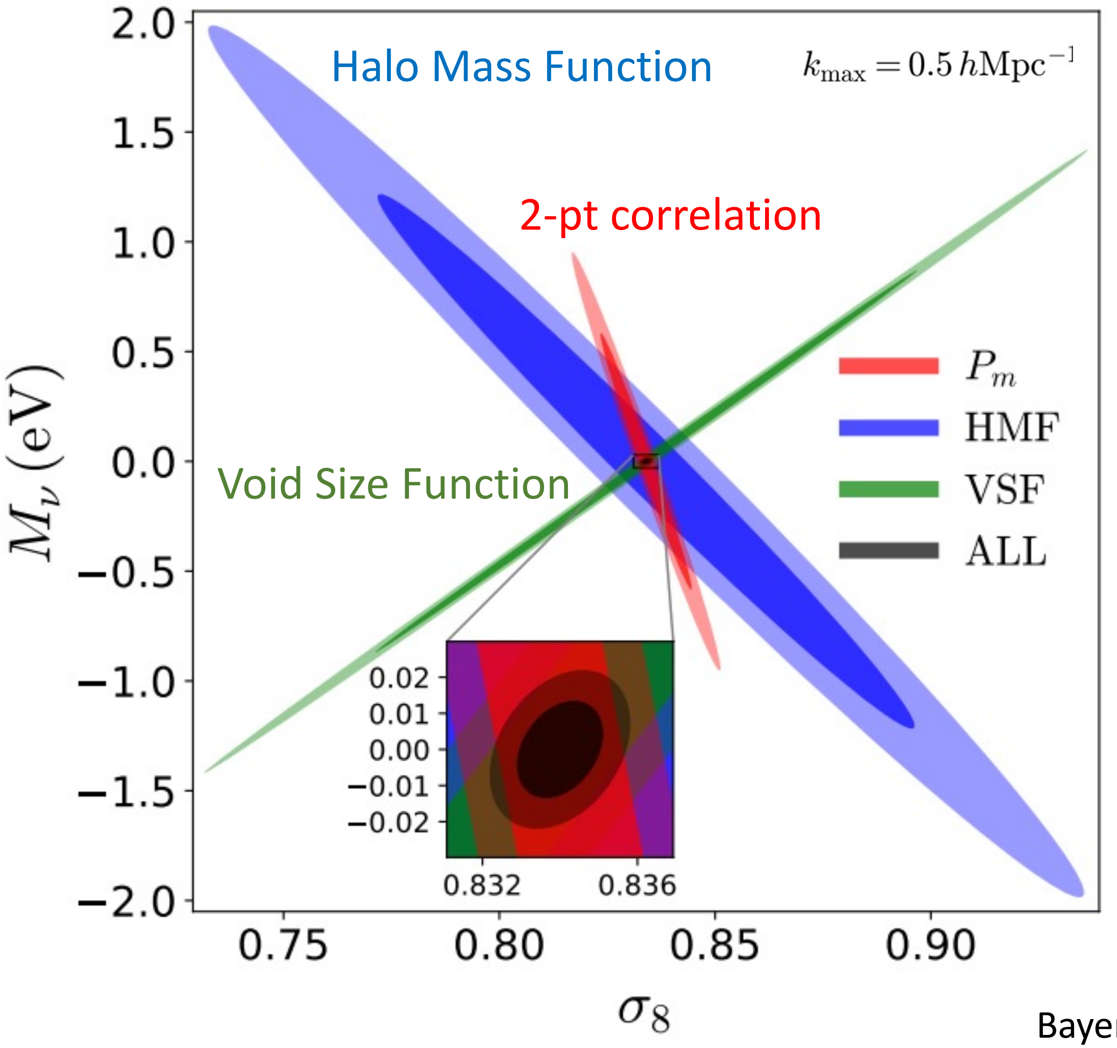
2-pt correlation

Void Size Function

Higher-order statistics

can provide information beyond the 2-pt



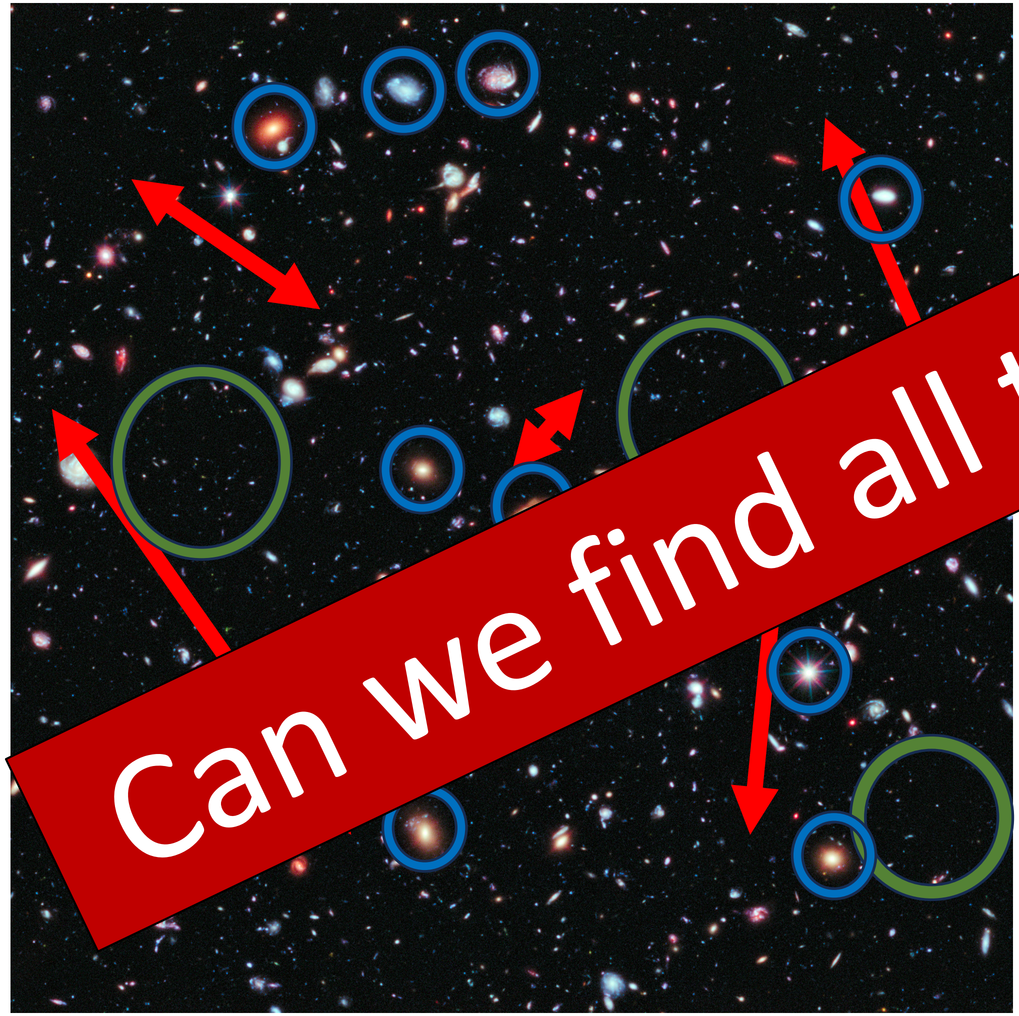


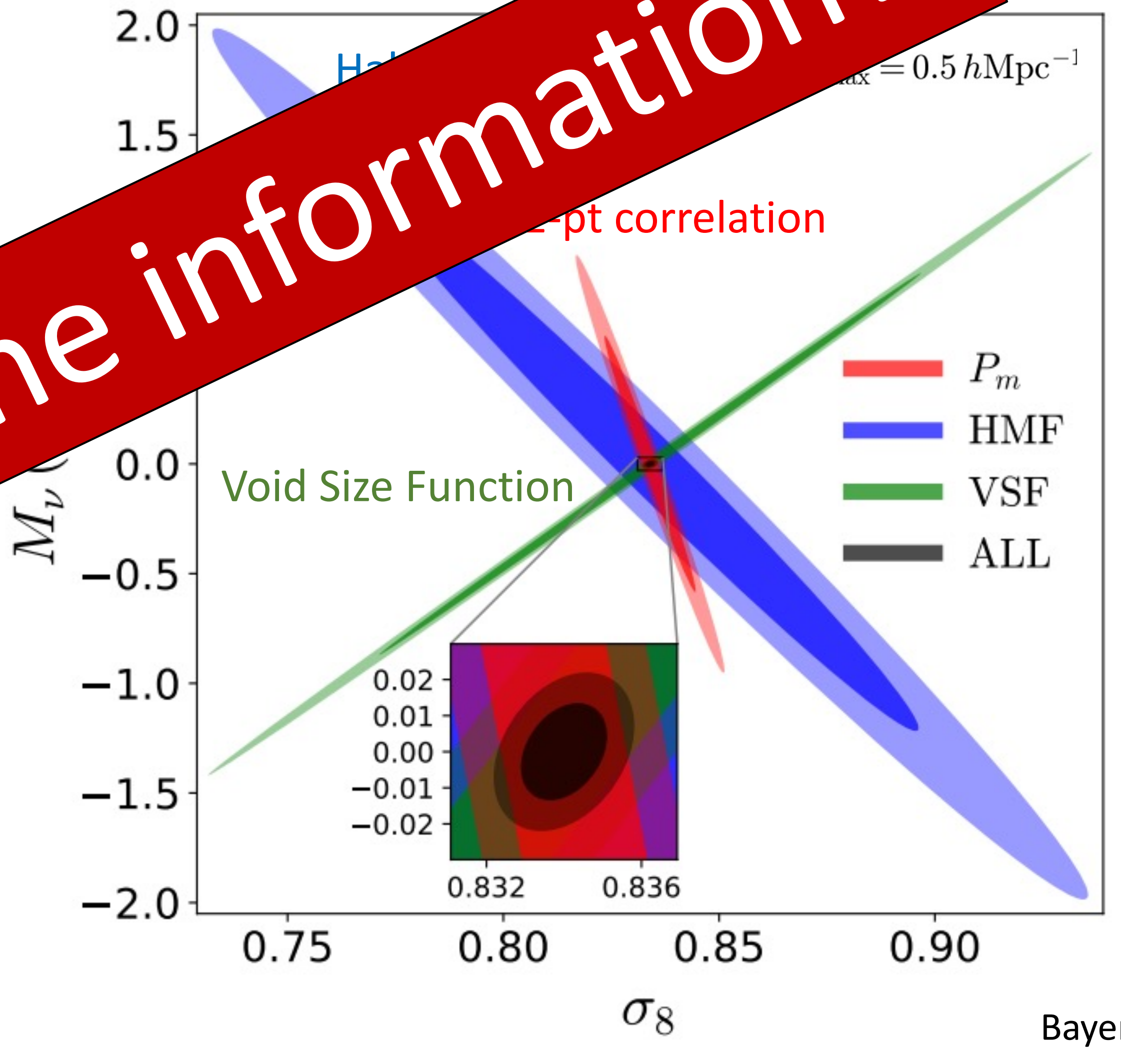
Bayer+ (2021)

Fake vs? Bayer, Banerjee, Seljak (2022)

Higher-order statistics

can provide information beyond the 2-pt





Can we find all the information?

Bayer+ (2021)

Fake vs? Bayer, Banerjee, Seljak (2022)

Cosmological

Field-Level Inference

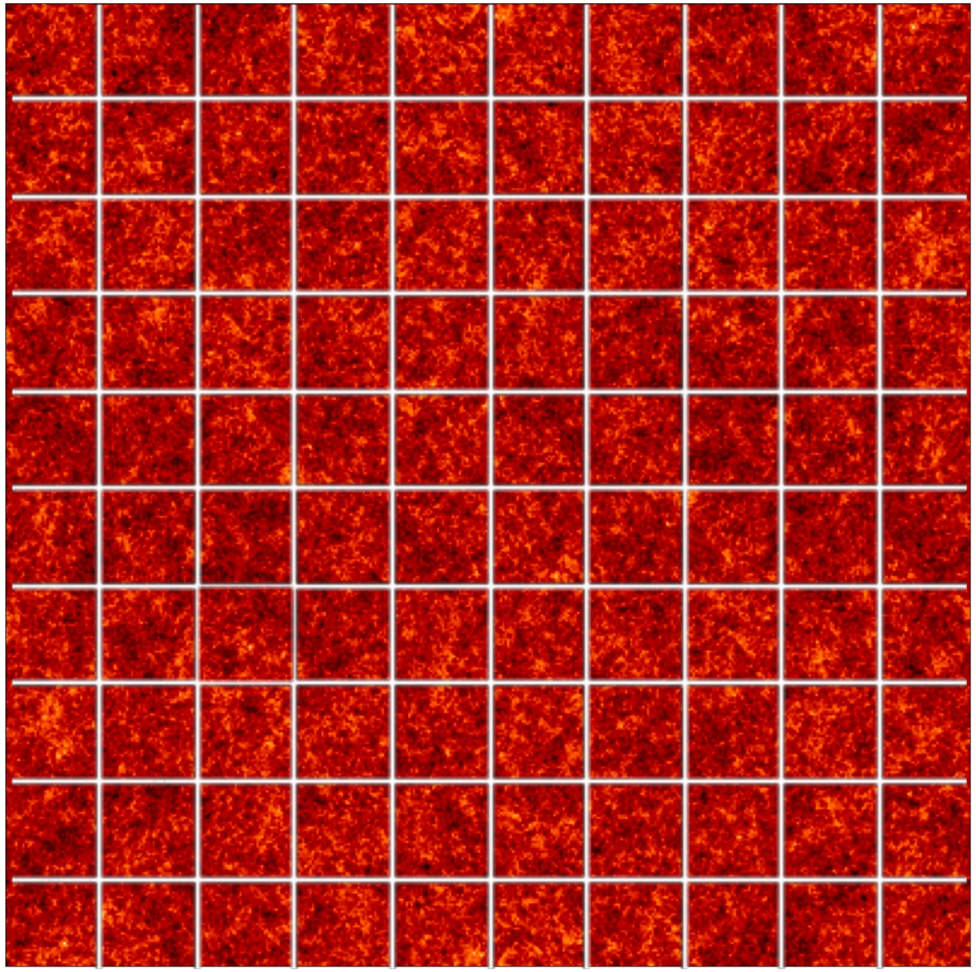
with

Microcanonical

Langevin Monte Carlo

Forward modeling

INITIAL s

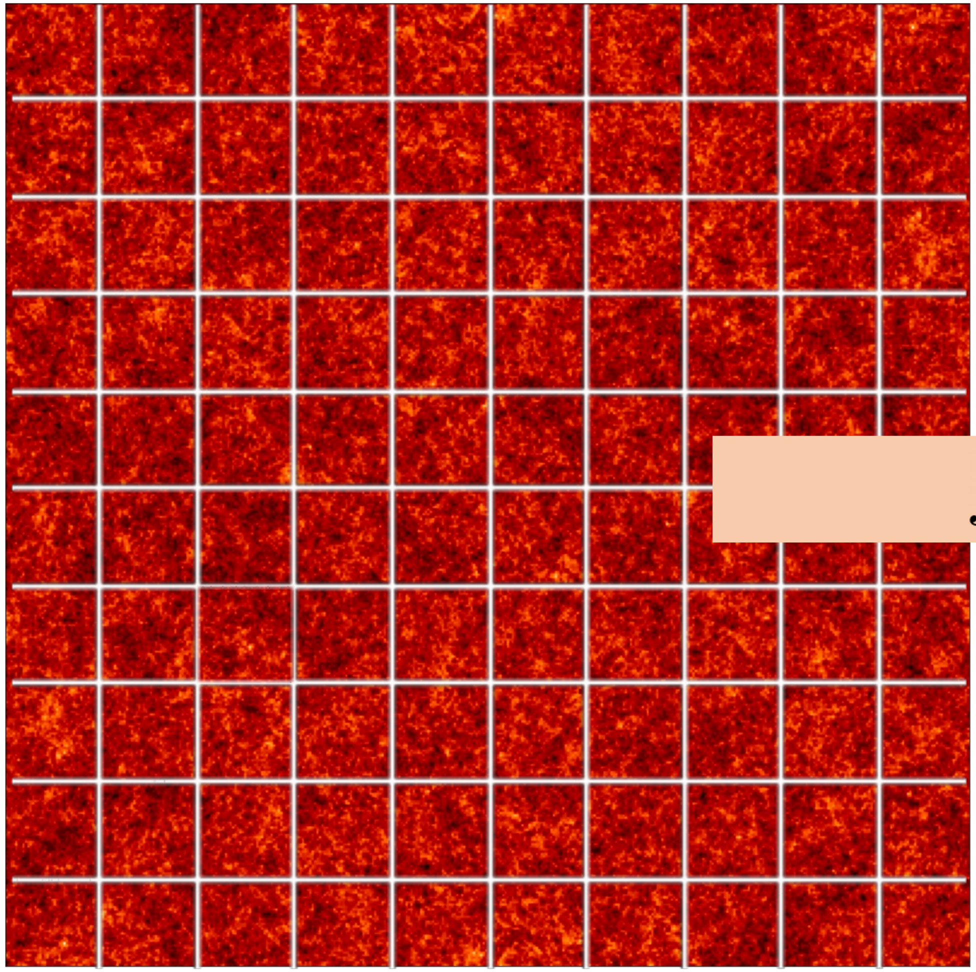


Kitaura & Ensslin (2007), Jasche & Wandelt (2012), Jasche & Lavaux (2017), Seljak et al. (2017), Schmidt et al. (2020), …

Forward modeling

s $\lambda = \{\sigma_8, \Omega_m, \ldots\}$

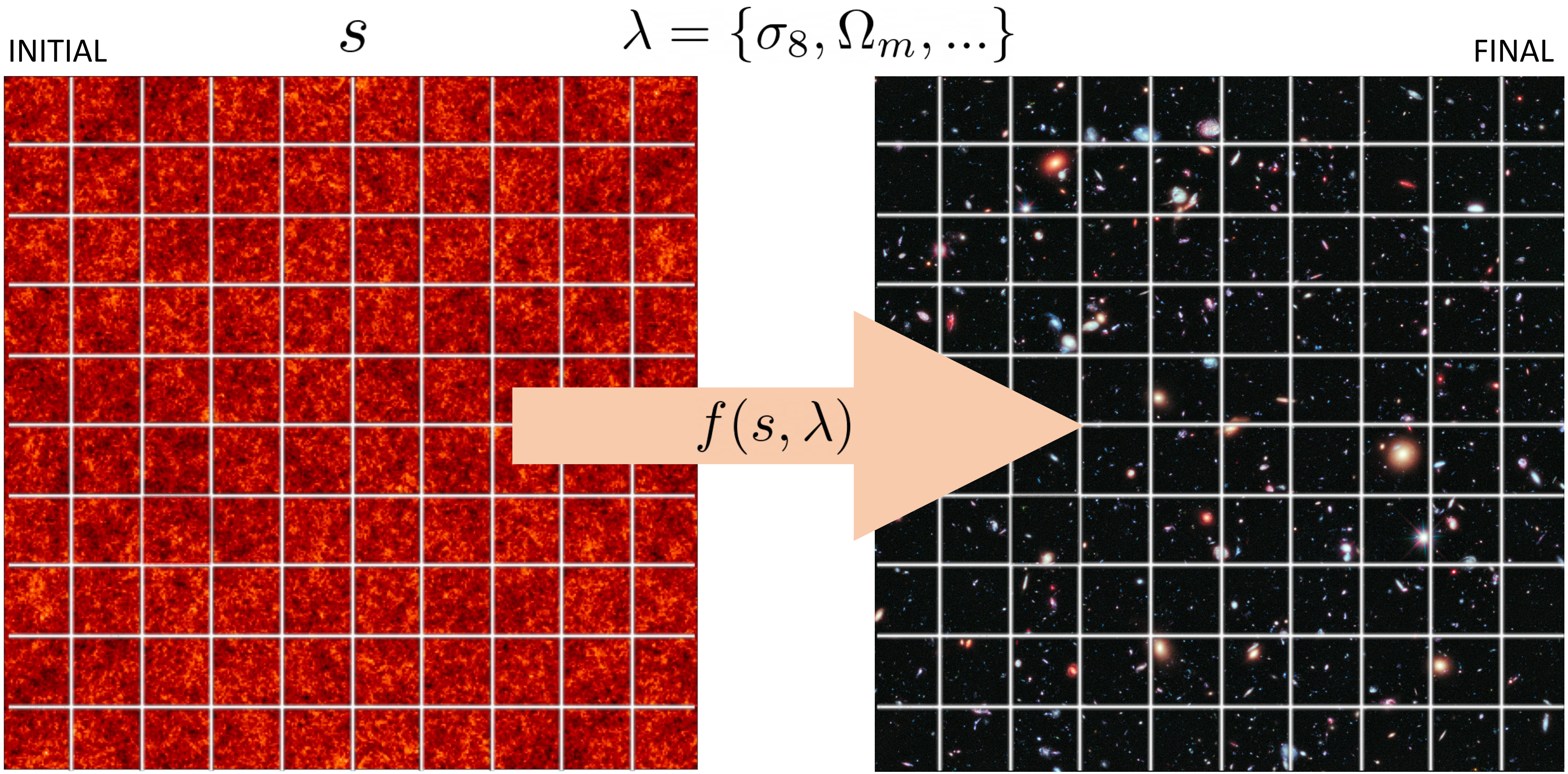
INITIAL



$f(s, \lambda)$

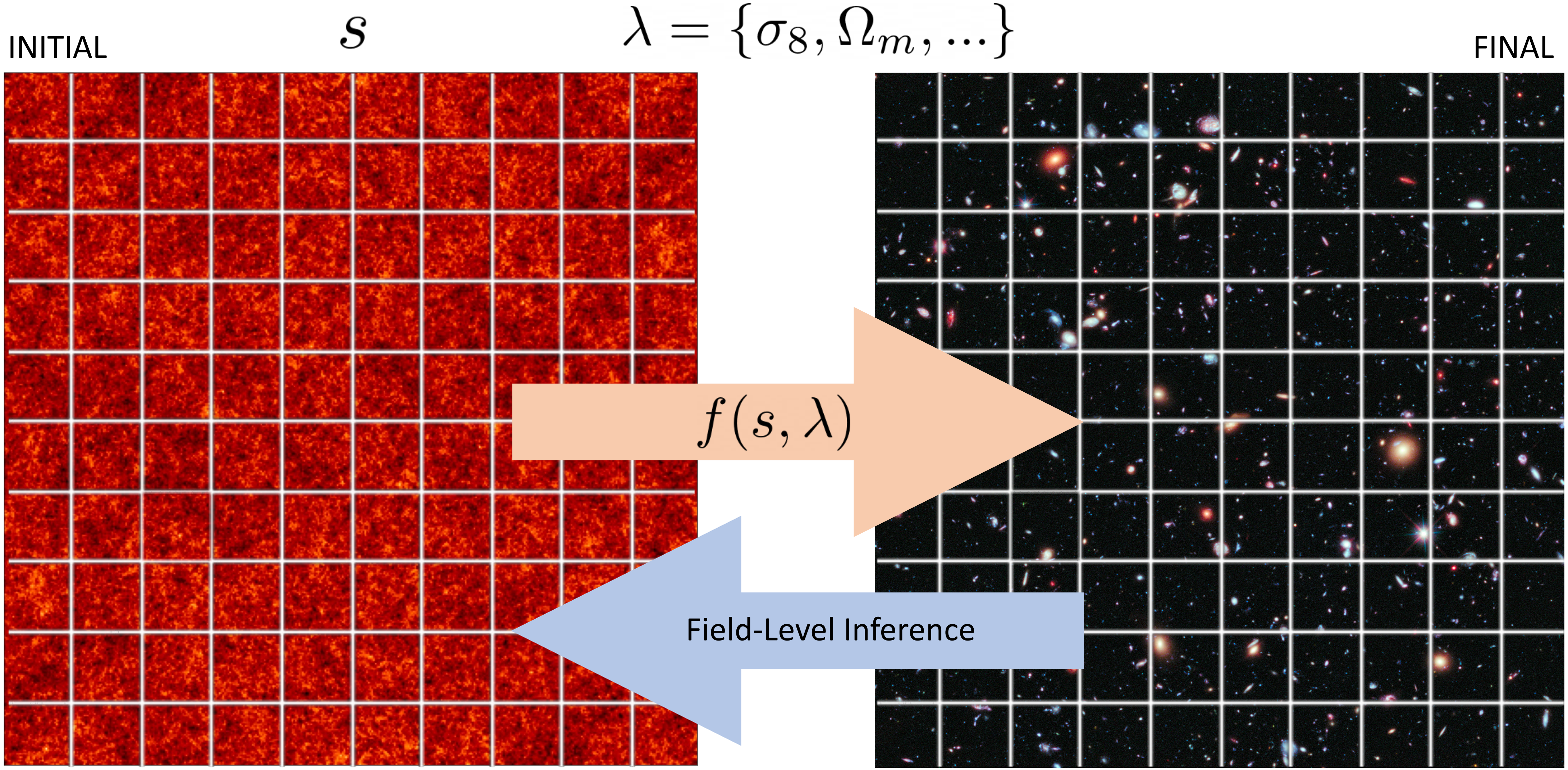
Kitaura & Ensslin (2007), Jasche & Wandelt (2012), Jasche & Lavaux (2017), Seljak et al. (2017), Schmidt et al. (2020), …

Forward modeling



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

Forward modeling



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

Field-Level Inference

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

Field-Level Inference

Given field data d and forward model f infer

initial modes s and cosmological parameters λ

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

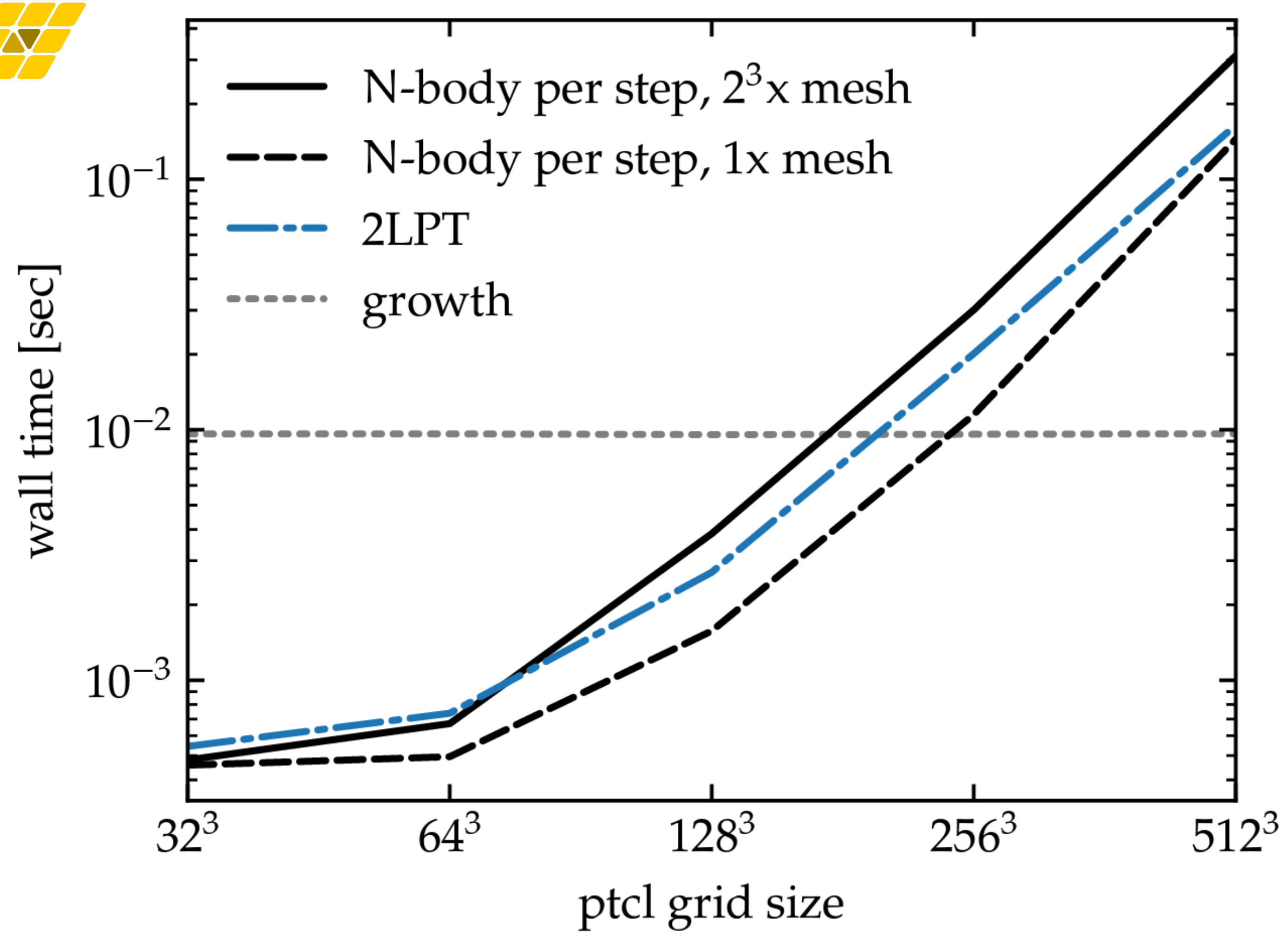
posterior likelihood prior

CHALLENGE: Multimillion dimensional parameter space!

1. Need fast forward model

2. Need differentiable forward model

3. Need fast sampler



1. Fast forward model ✅

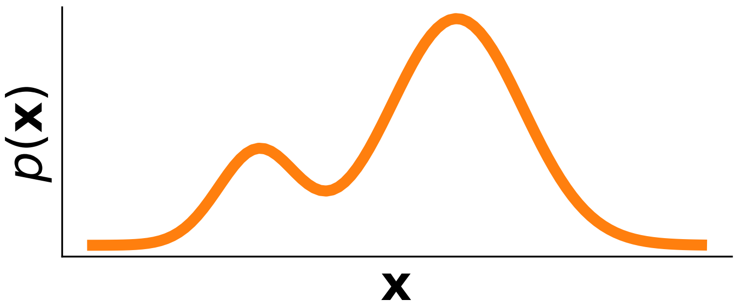
2. Differentiable forward model ✅

3. Need fast sampler

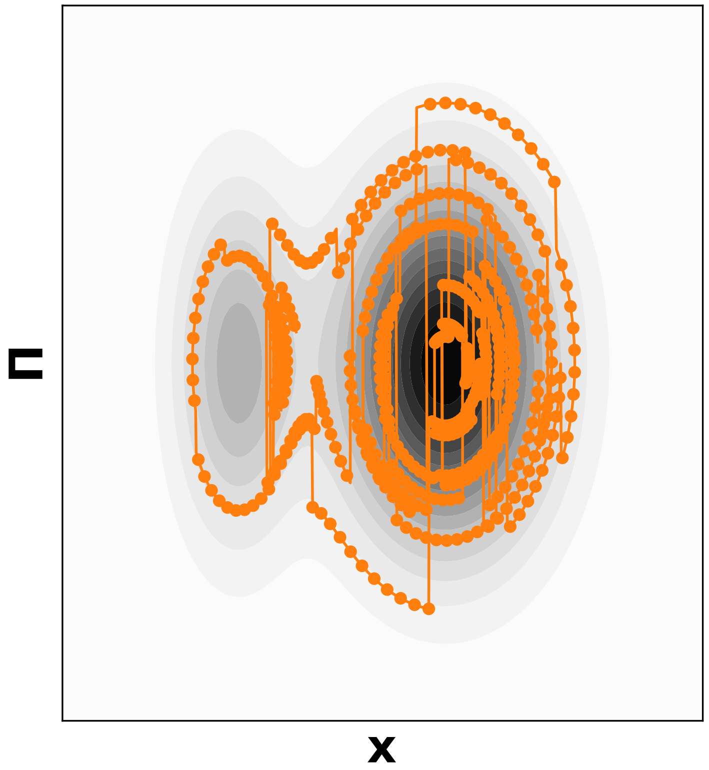
Li+2022

Cosmological Field-Level Inference *with* Microcanonical Langevin Monte Carlo

Canonical HMC

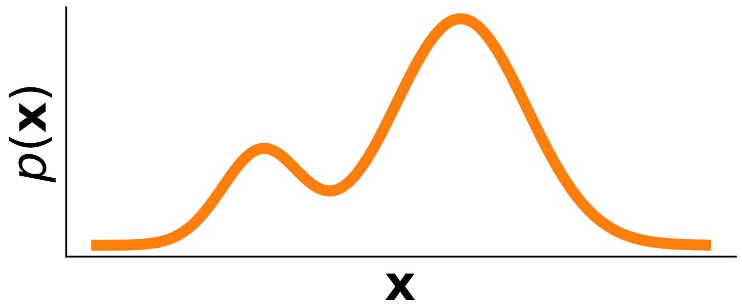


$p(\mathbf{x}, \mathbf{\Pi}) \propto e^{-H(\mathbf{x}, \mathbf{\Pi})}$

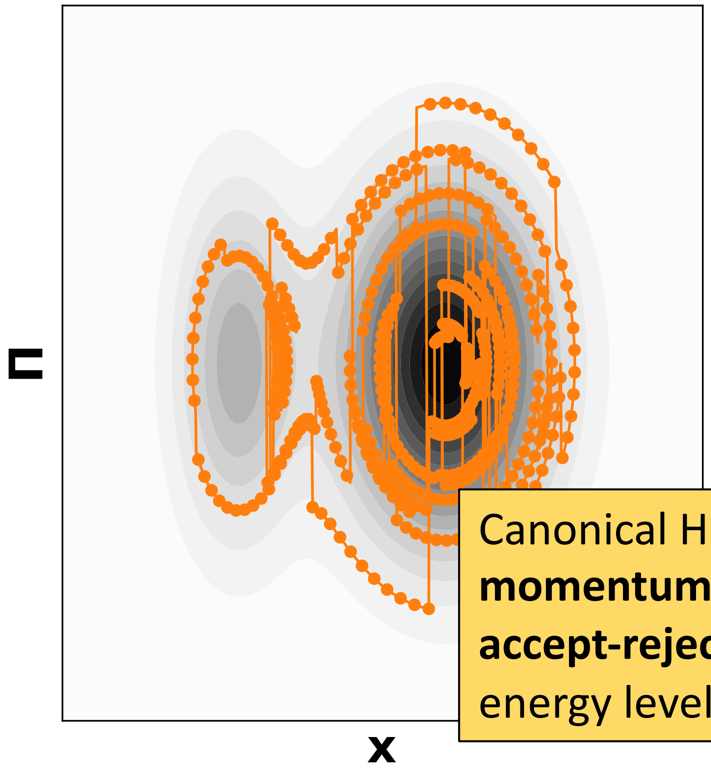


Robnik+2022

Canonical HMC



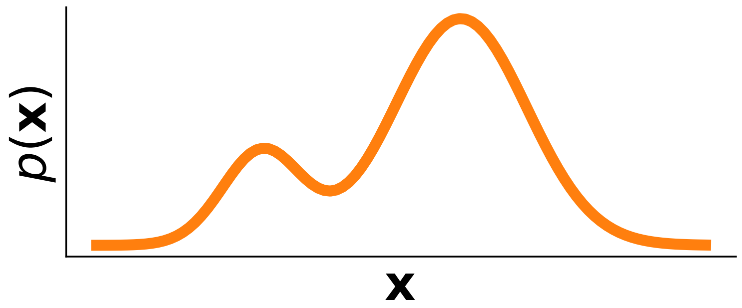
$$p(\mathbf{x}, \mathbf{\Pi}) \propto e^{-H(\mathbf{x}, \mathbf{\Pi})}$$



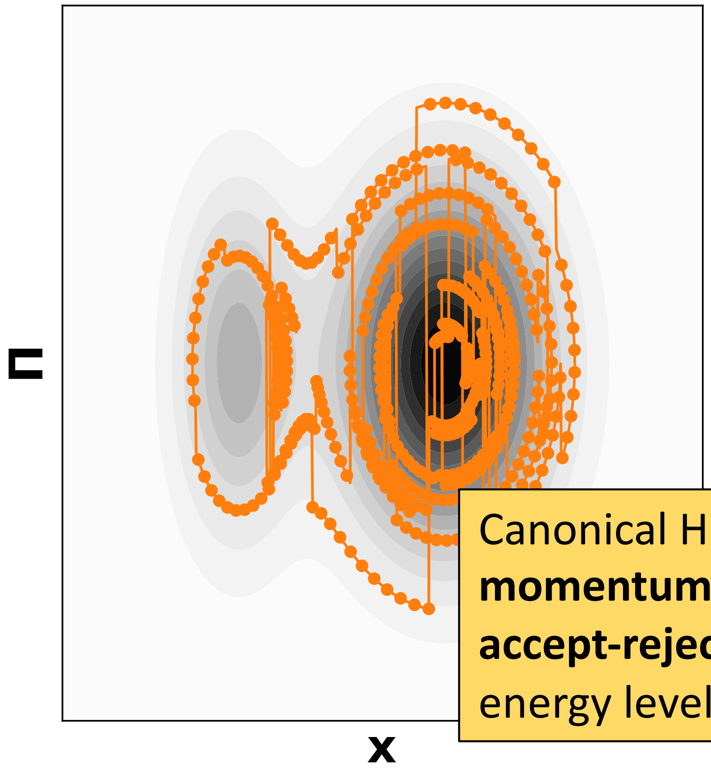
Canonical HMC requires **momentum resampling** and **accept-reject step** to change energy levels and converge

Robnik+2022

Canonical HMC

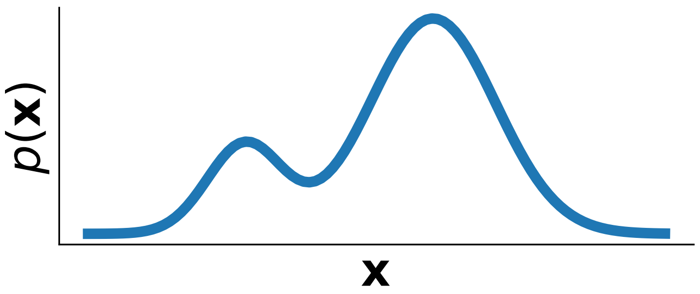


$$p(\mathbf{x}, \mathbf{\Pi}) \propto e^{-H(\mathbf{x}, \mathbf{\Pi})}$$

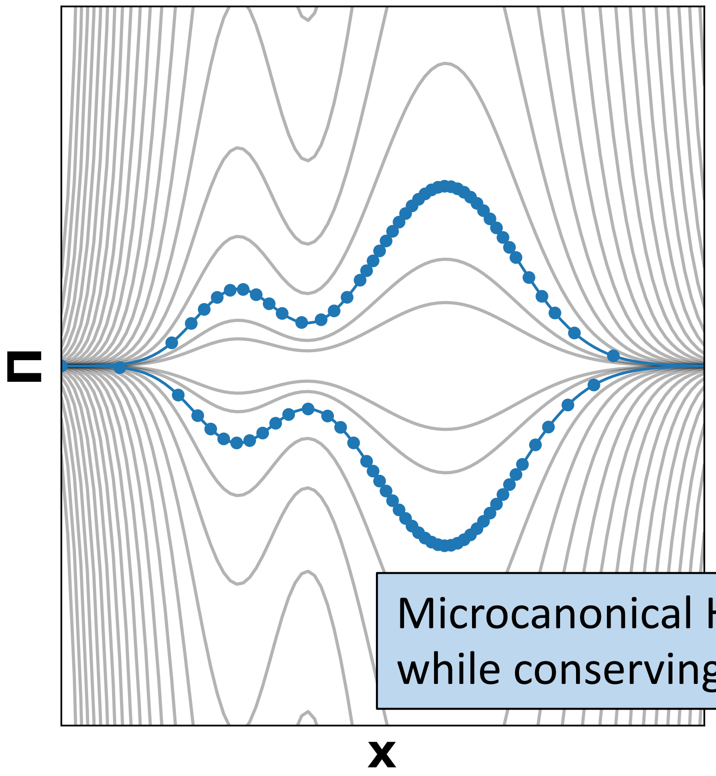


Canonical HMC requires **momentum resampling** and **accept-reject step** to change energy levels and converge

Microcanonical HMC



$$p(\mathbf{x}, \mathbf{\Pi}) \propto \delta(H(\mathbf{x}, \mathbf{\Pi}) - E)$$



Microcanonical HMC converges while conserving energy

Robnik+2022

Microcanonical Hamiltonian Monte Carlo

$$dz = u dt$$

$$du = -(d-1)^{-1}(1 - uu^T)\nabla \mathcal{L}(z)$$

MCHMC

Robnik+2022

Microcanonical Langevin Monte Carlo

$$dz = u\,dt$$

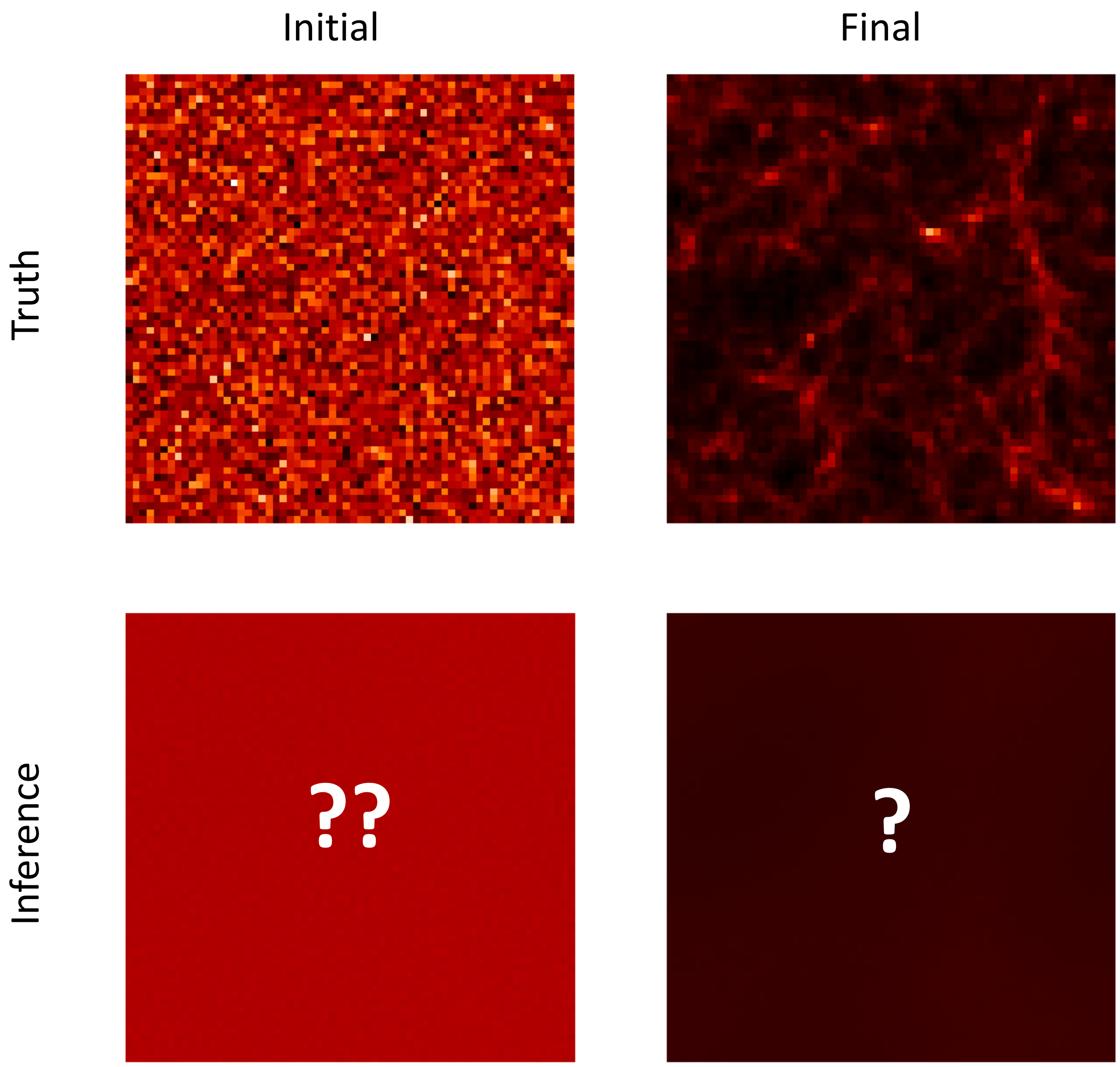
$$du = -(d-1)^{-1}(1 - uu^T)[\nabla \mathcal{L}(z) + \eta\, dW]$$

MCHMC **MCLMC**

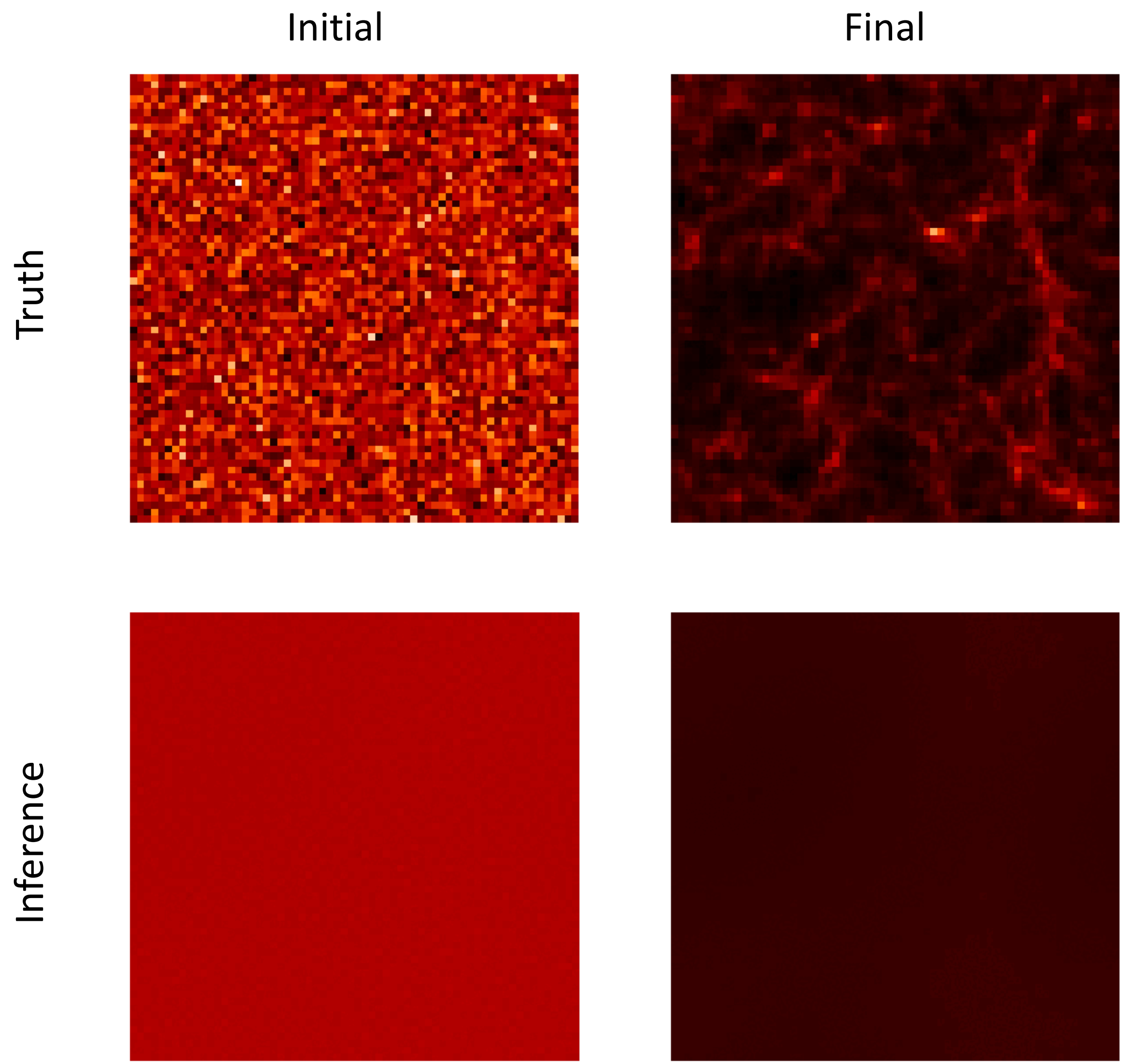
Improve ergodicity by including Langevin-like stochastic term

Robnik+2022

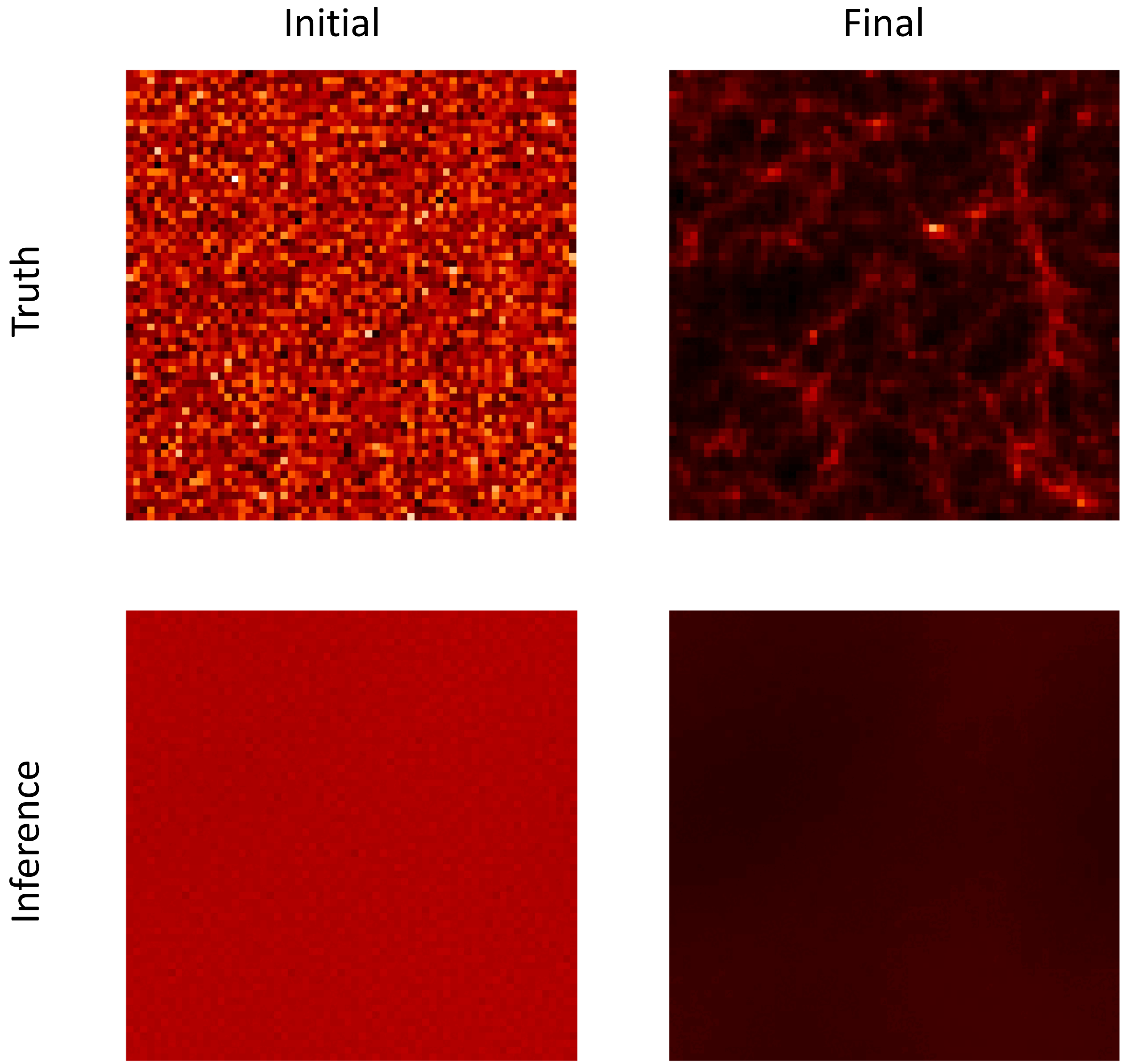
Cosmological Field-Level Inference *with* Microcanonical Langevin Monte Carlo



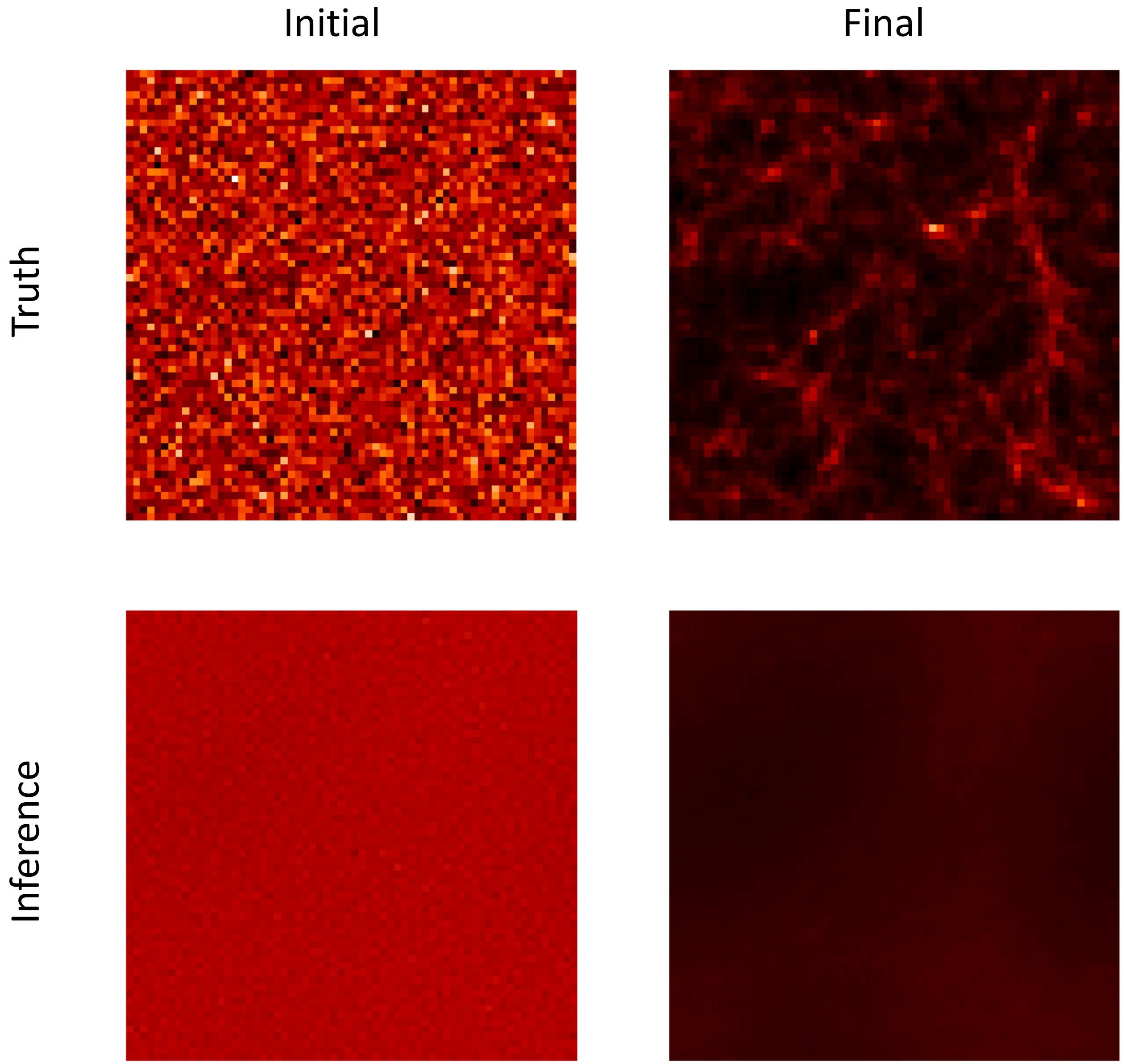
Bayer, Seljak, Modi (2023)



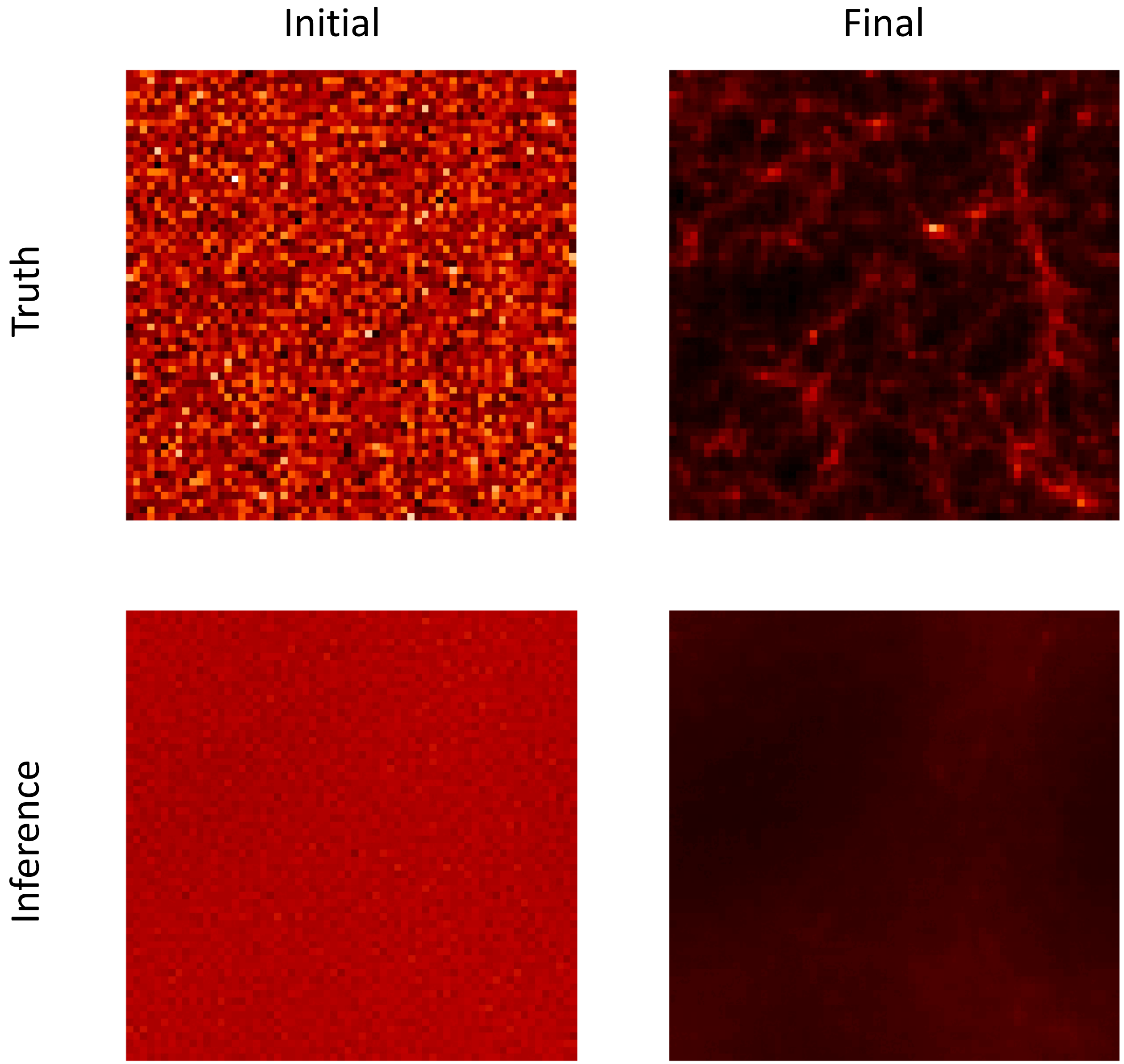
Bayer, Seljak, Modi (2023)



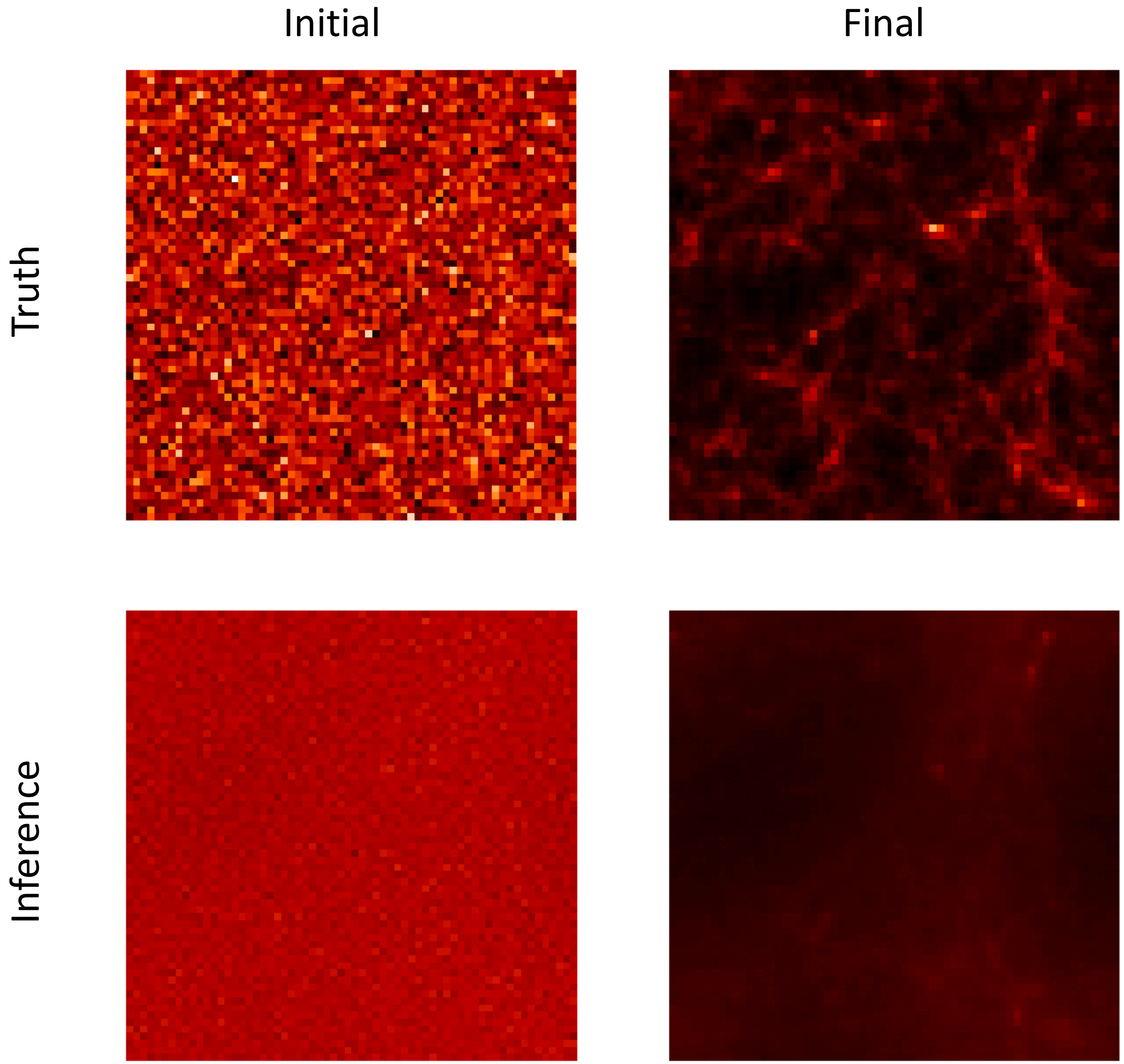
Bayer, Seljak, Modi (2023)



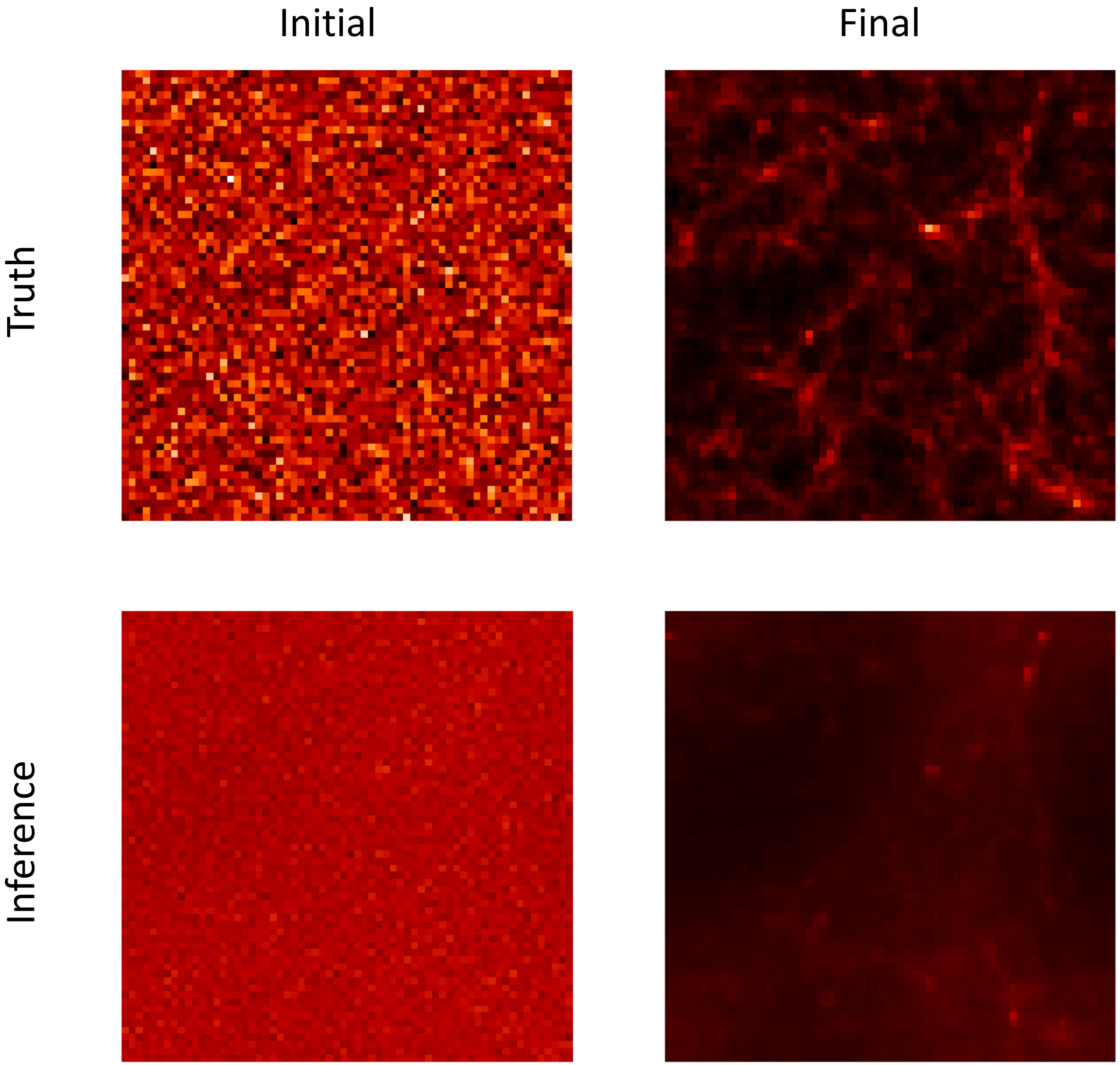
Bayer, Seljak, Modi (2023)



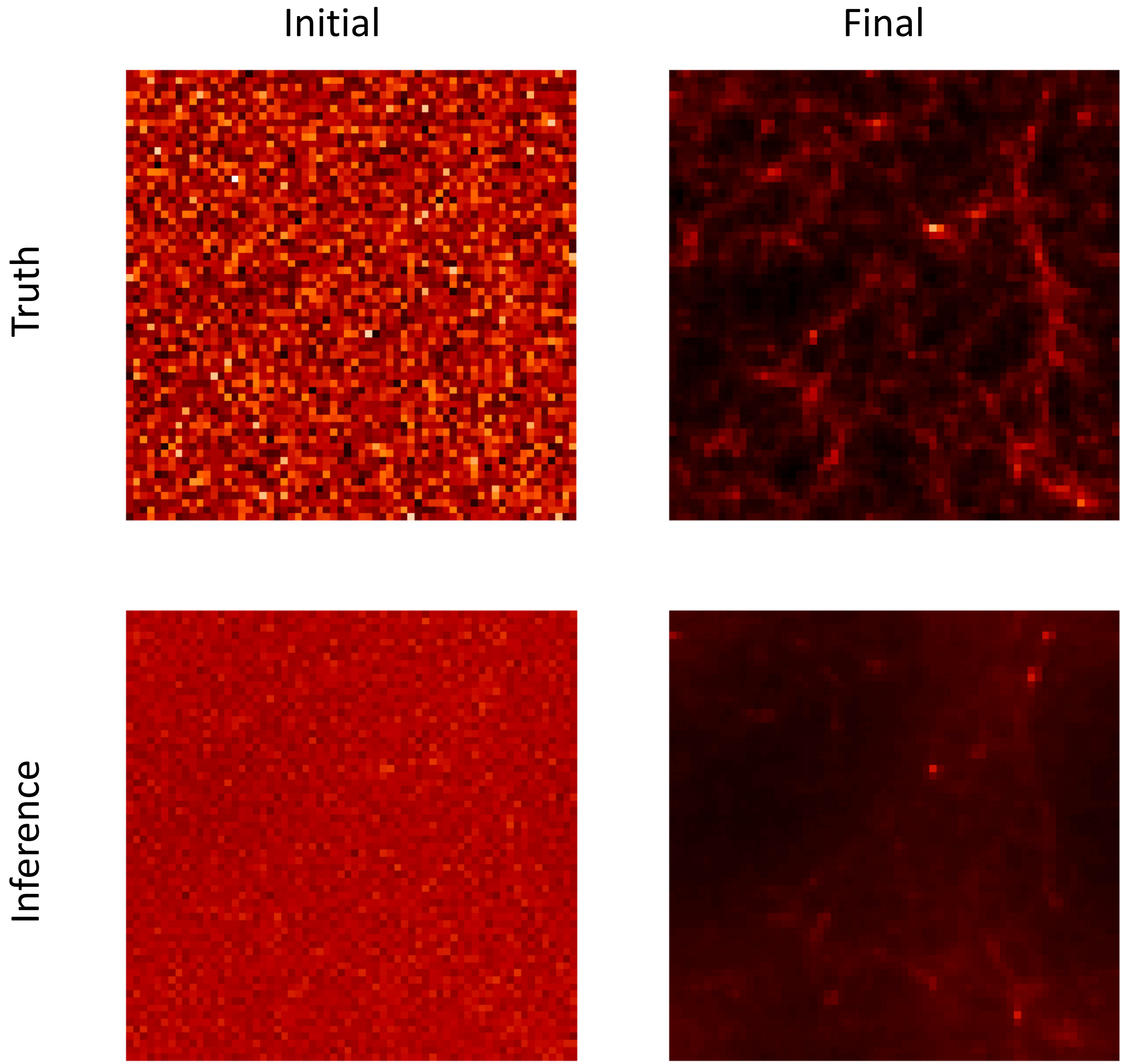
Bayer, Seljak, Modi (2023)



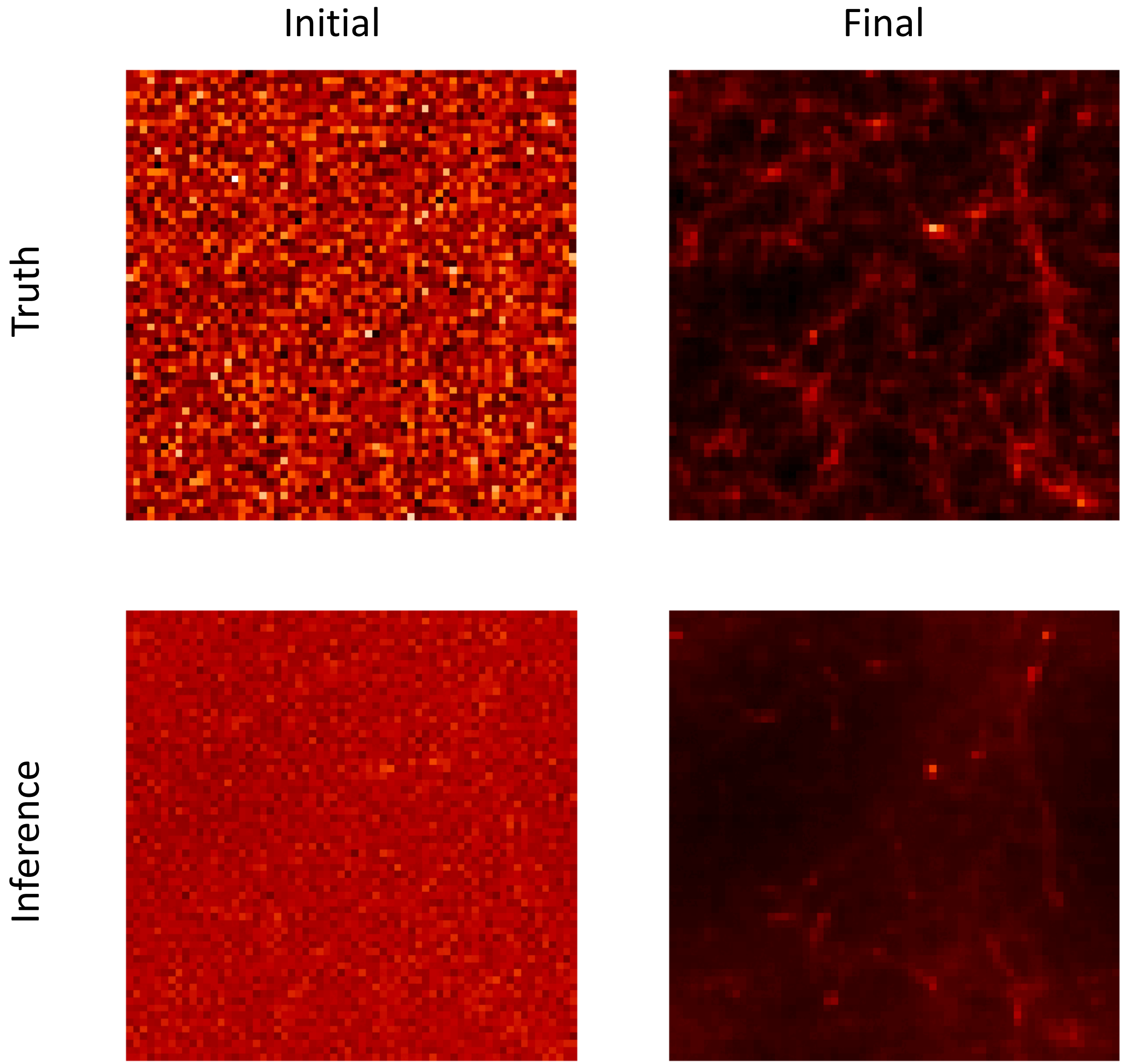
Bayer, Seljak, Modi (2023)



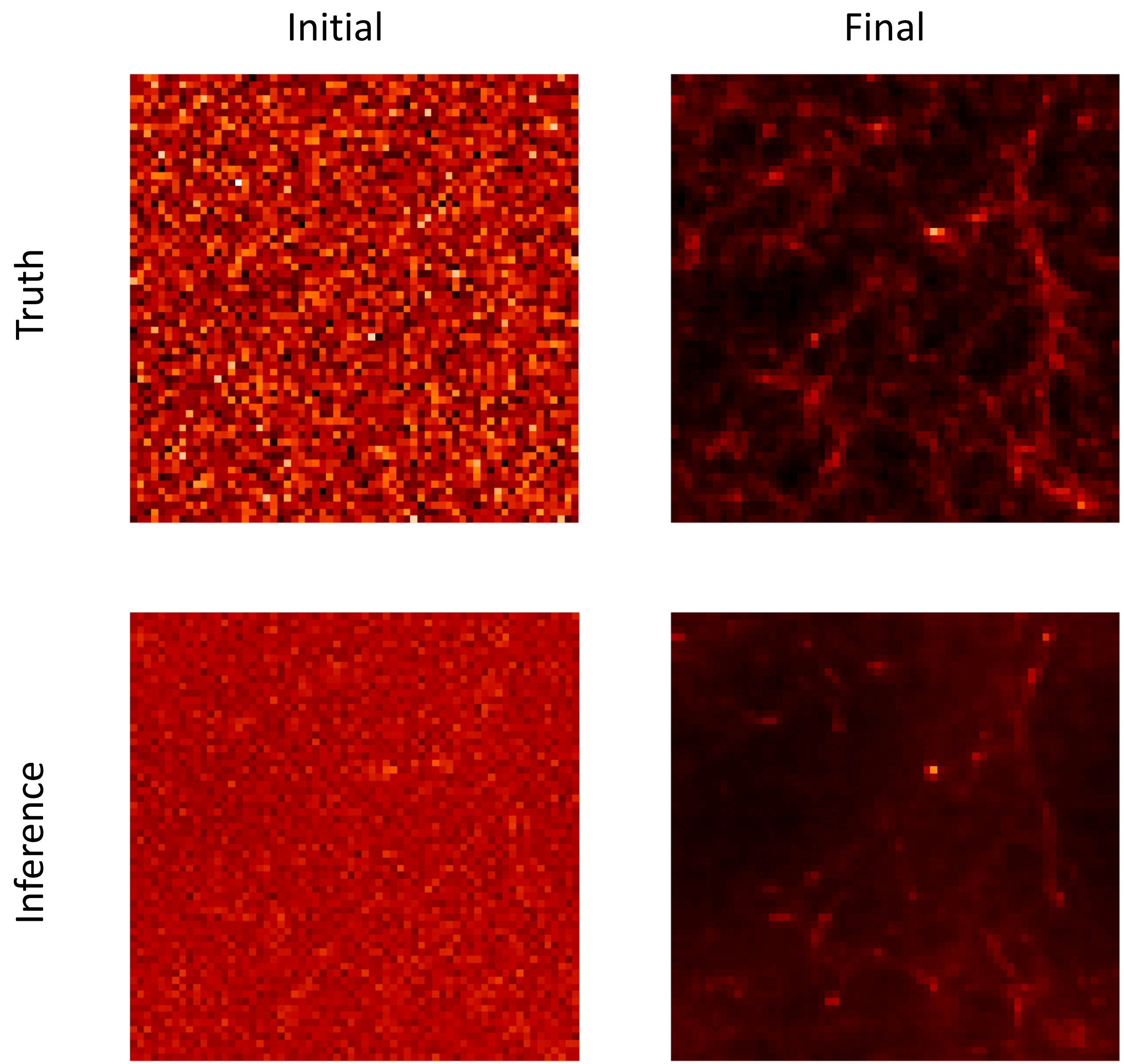
Bayer, Seljak, Modi (2023)



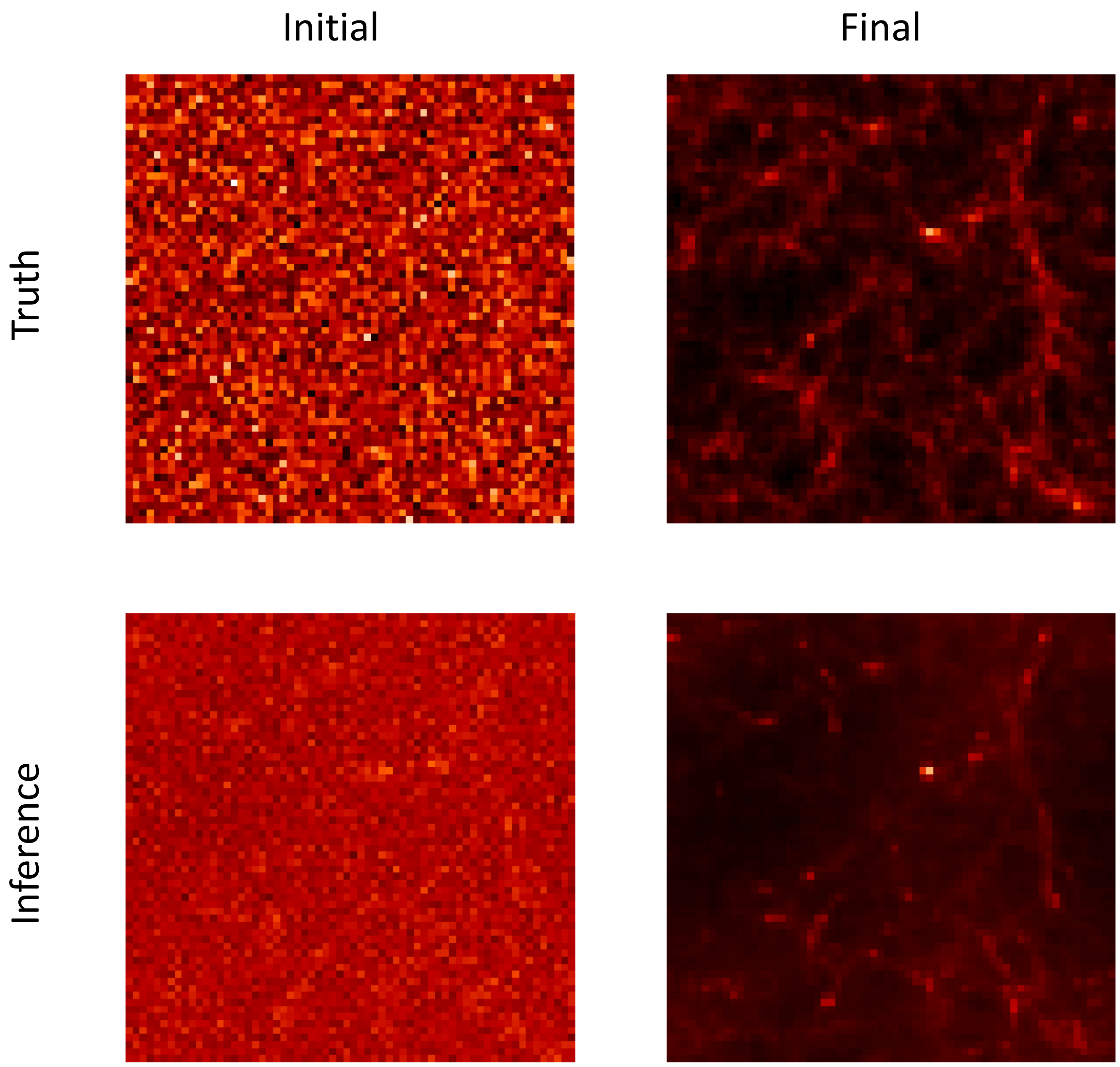
Bayer, Seljak, Modi (2023)



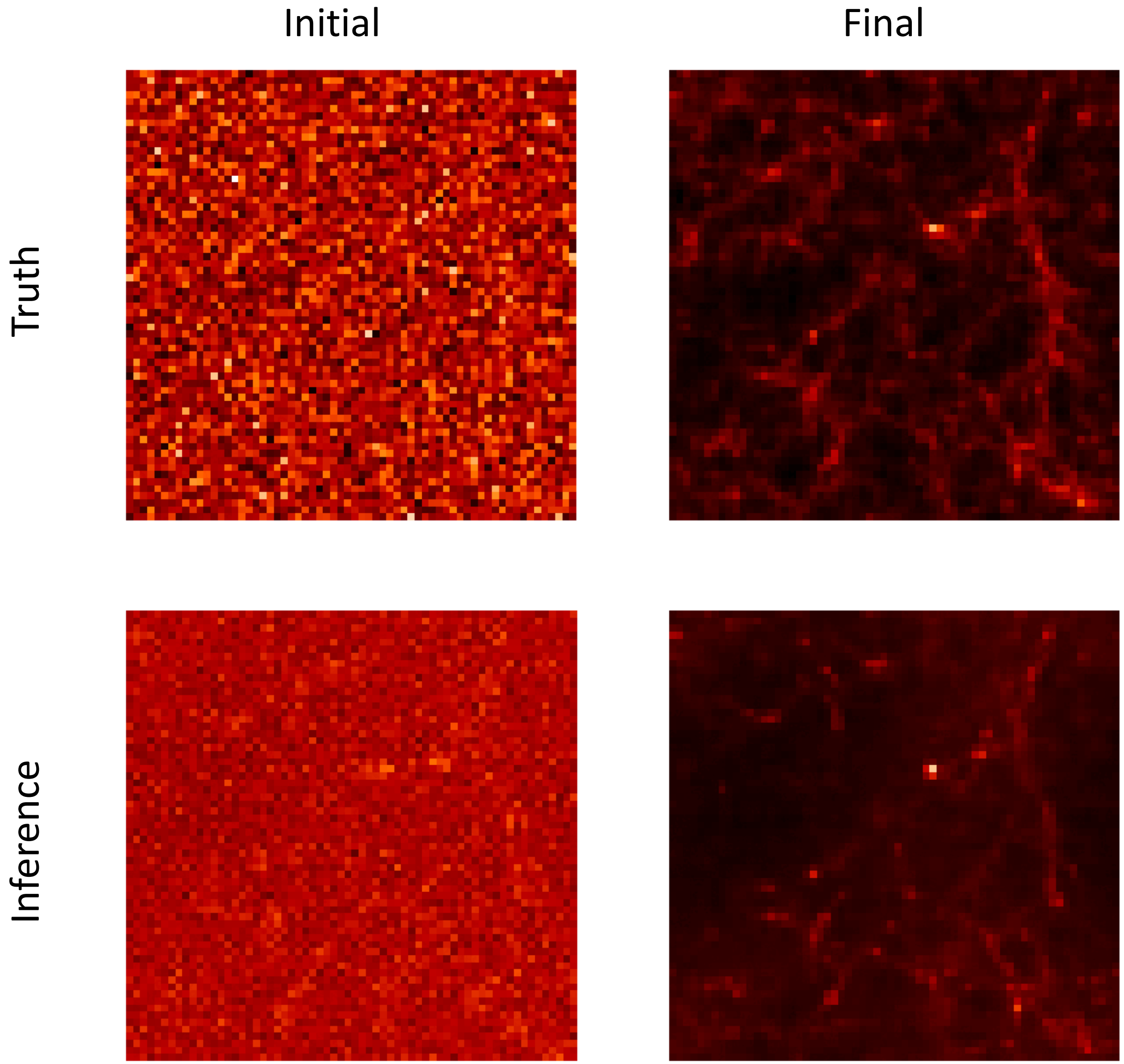
Bayer, Seljak, Modi (2023)



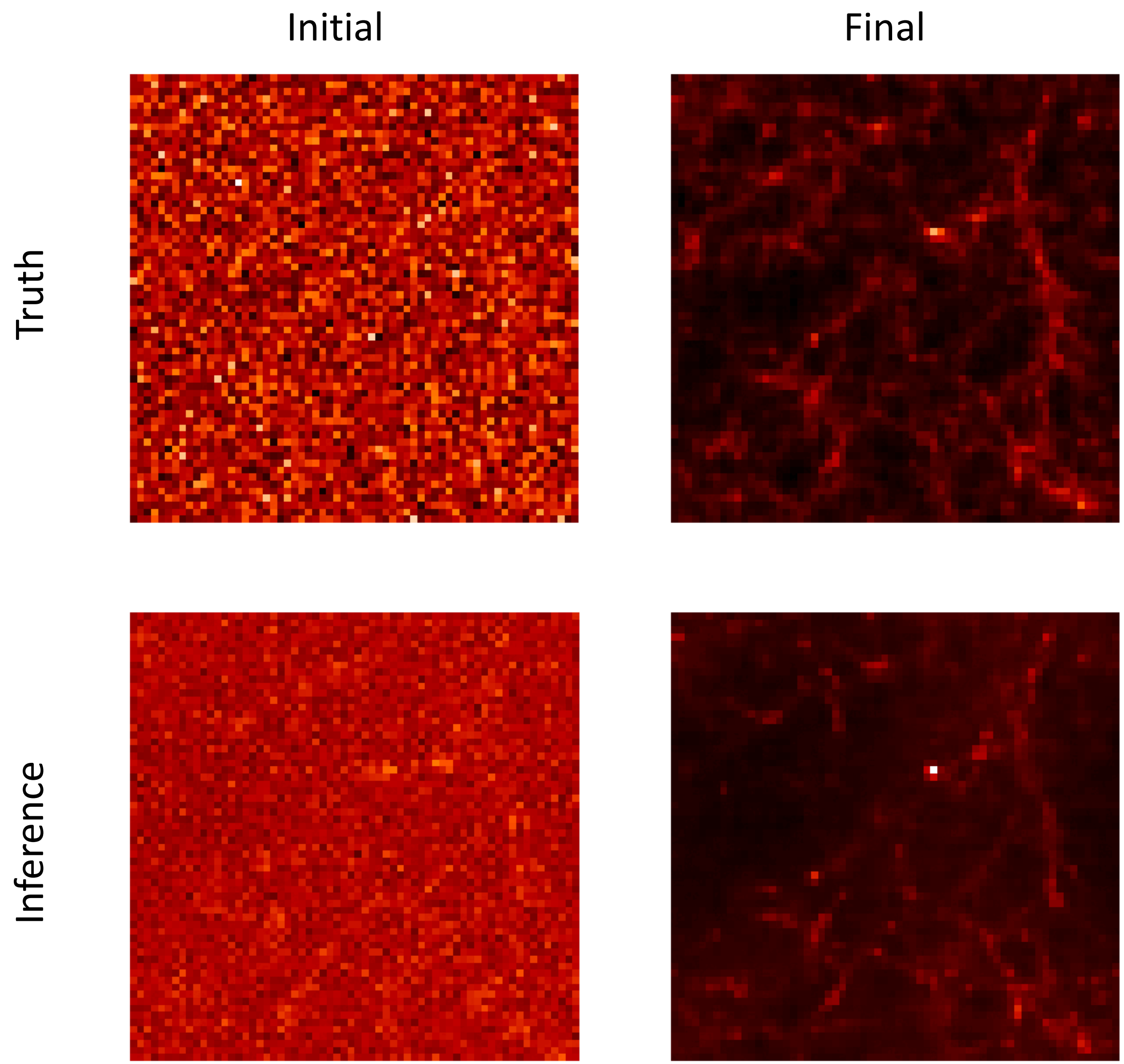
Bayer, Seljak, Modi (2023)



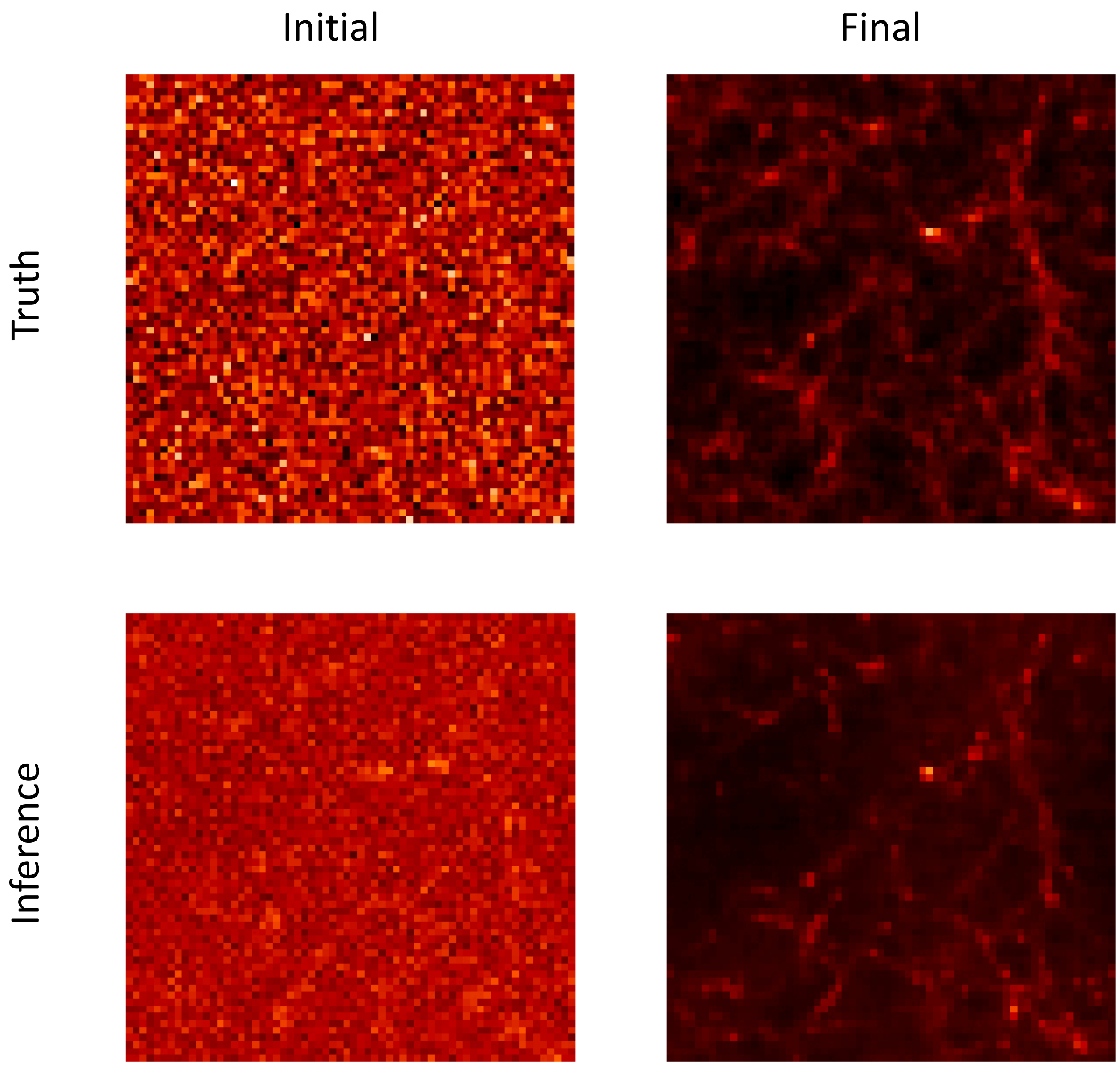
Bayer, Seljak, Modi (2023)



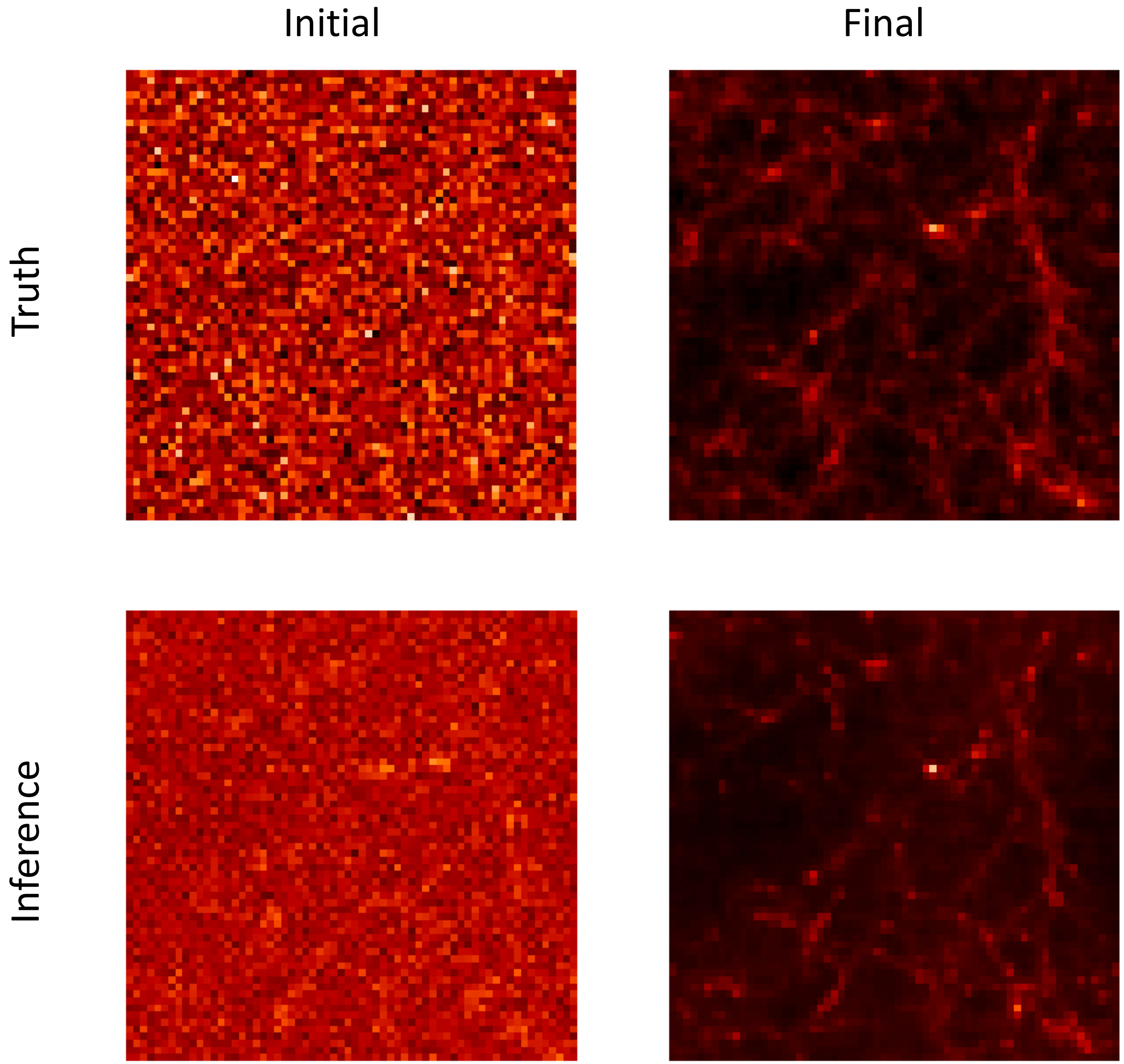
Bayer, Seljak, Modi (2023)



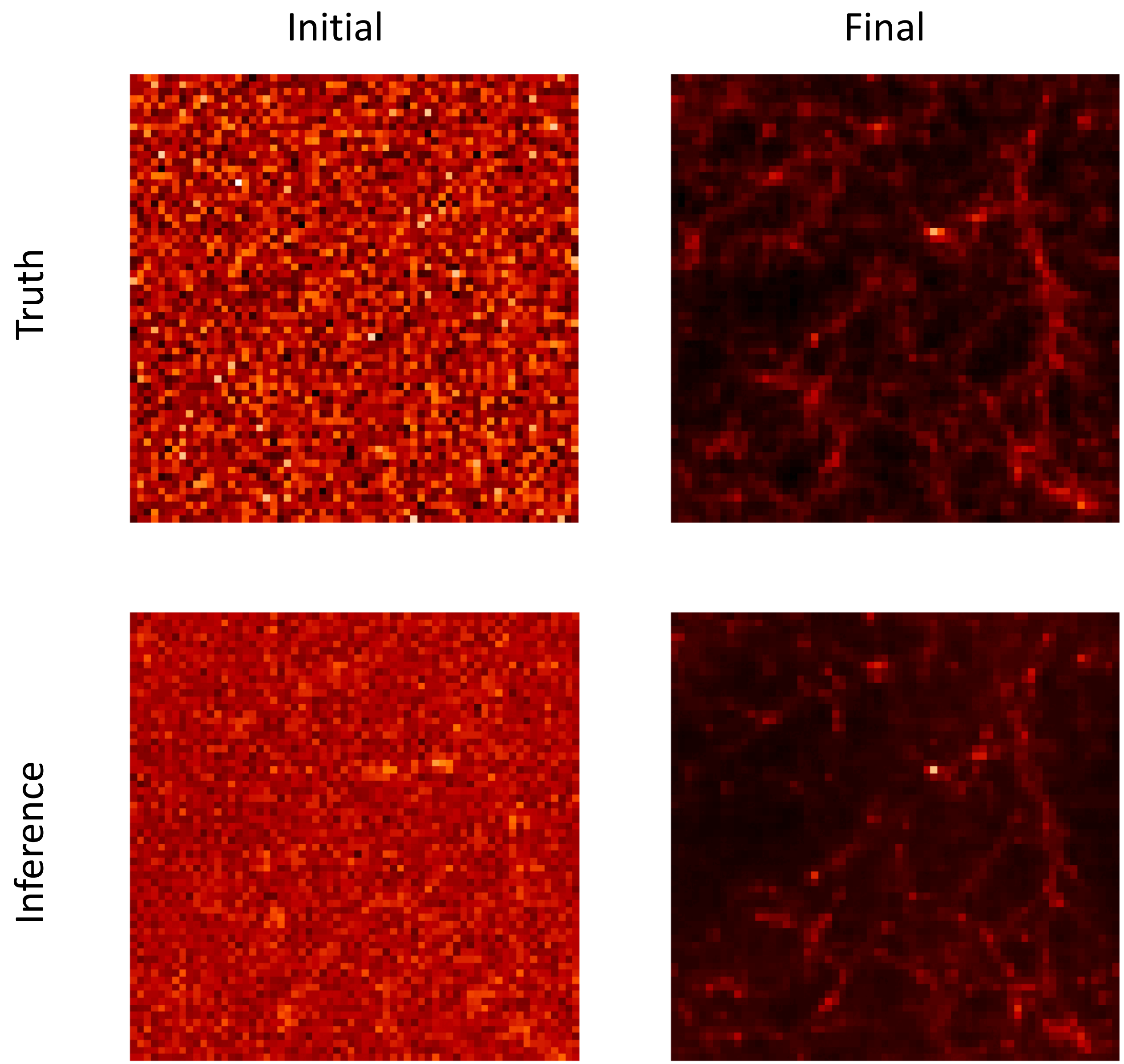
Bayer, Seljak, Modi (2023)



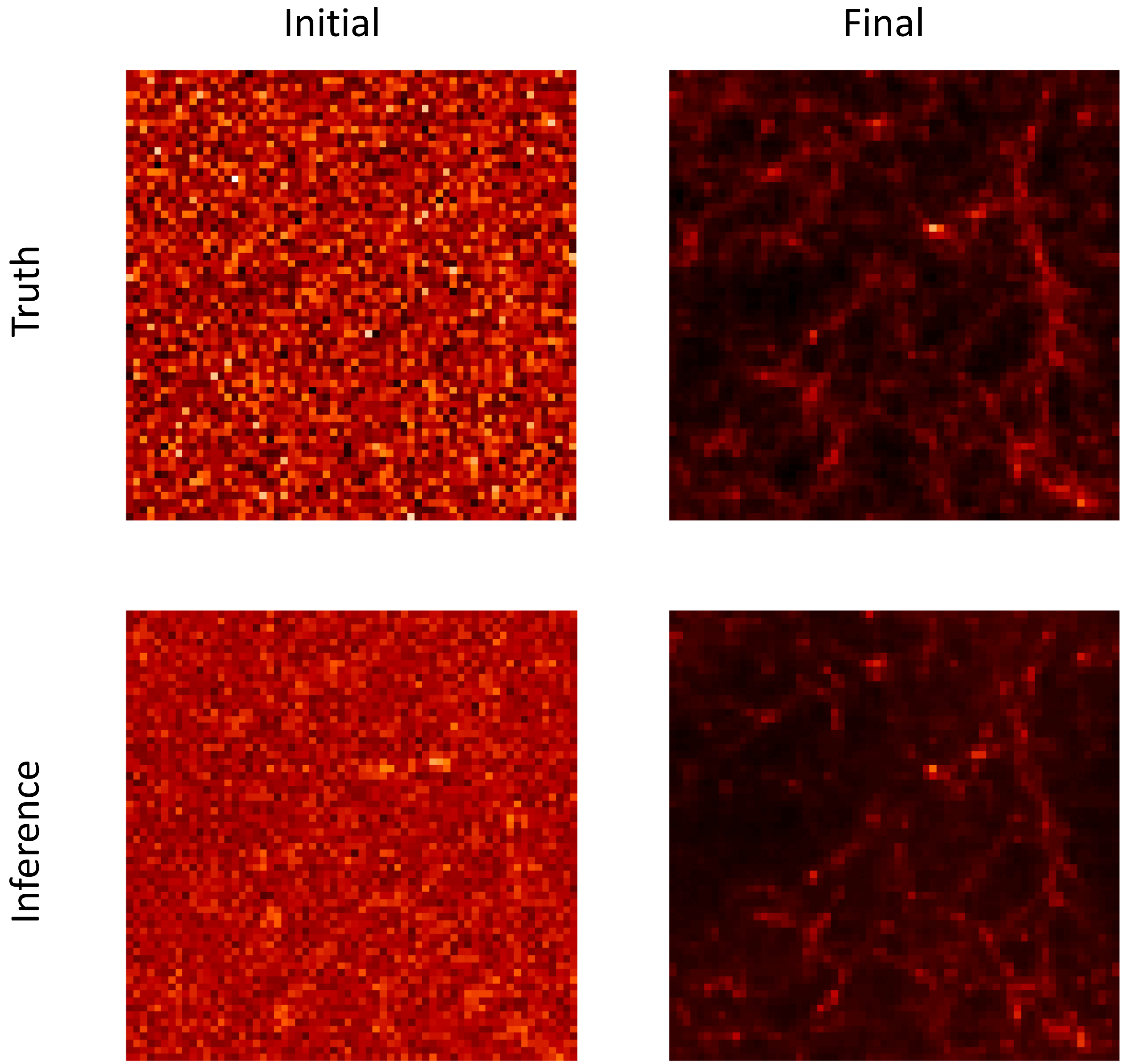
Bayer, Seljak, Modi (2023)



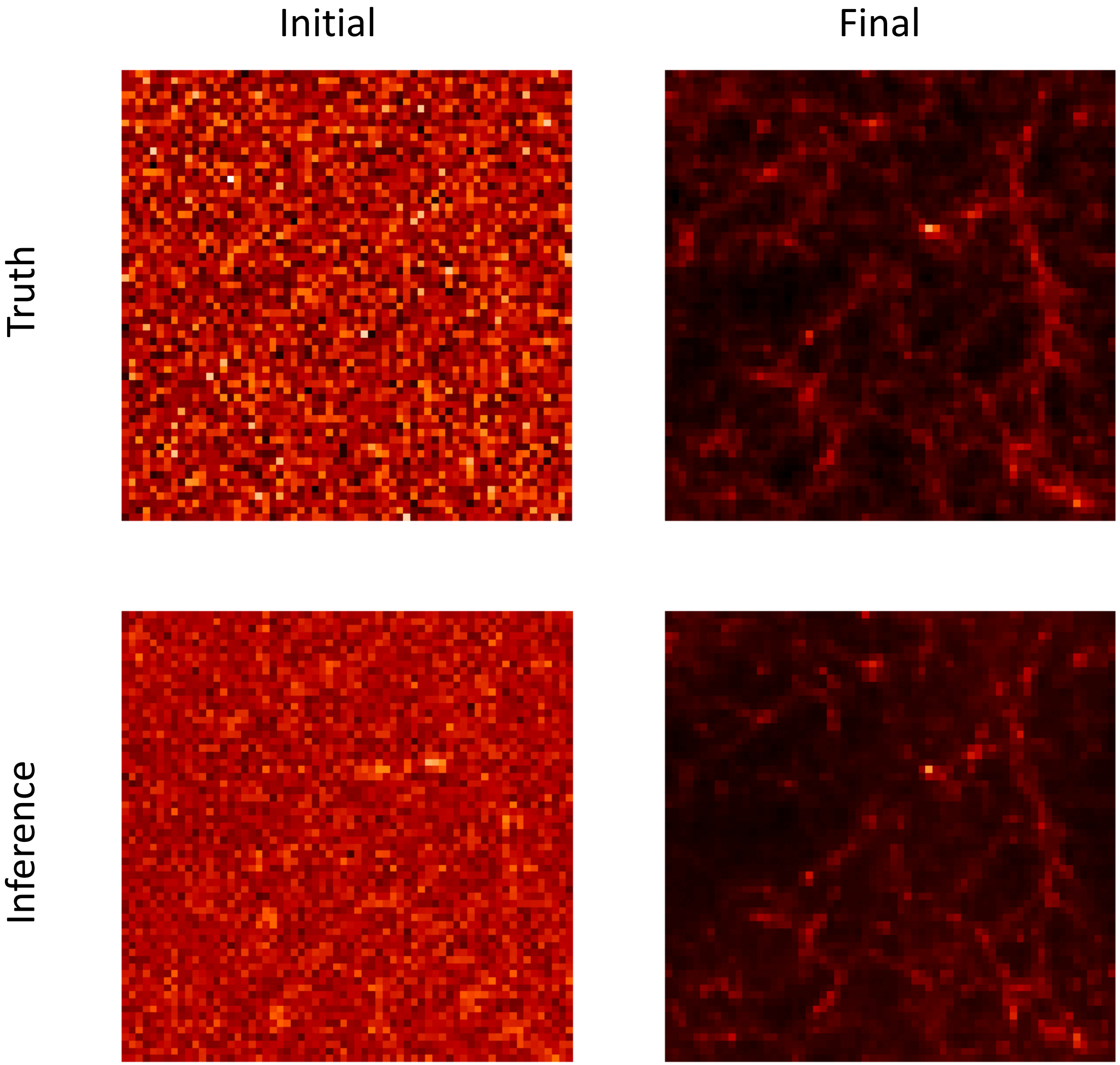
Bayer, Seljak, Modi (2023)



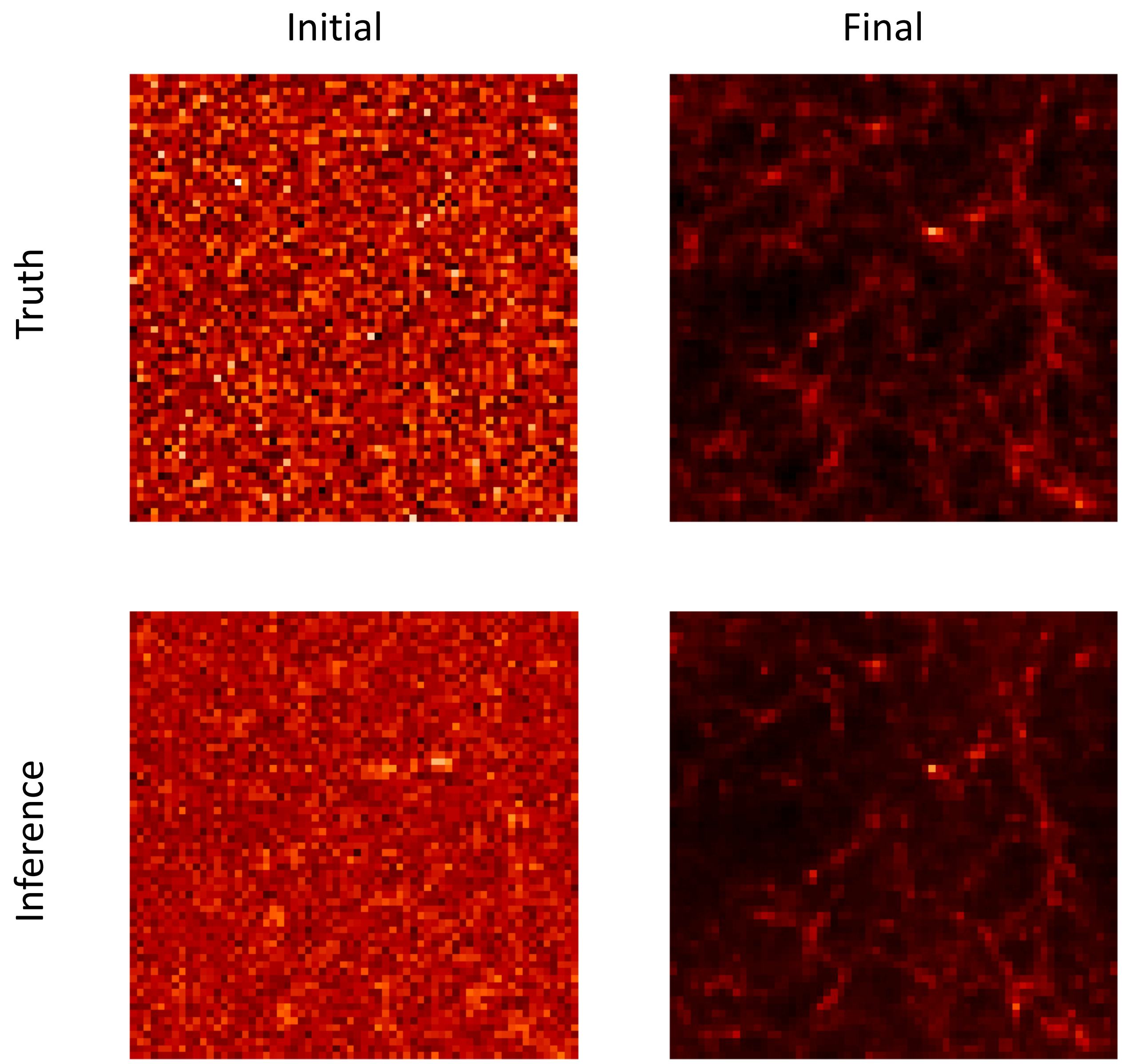
Bayer, Seljak, Modi (2023)



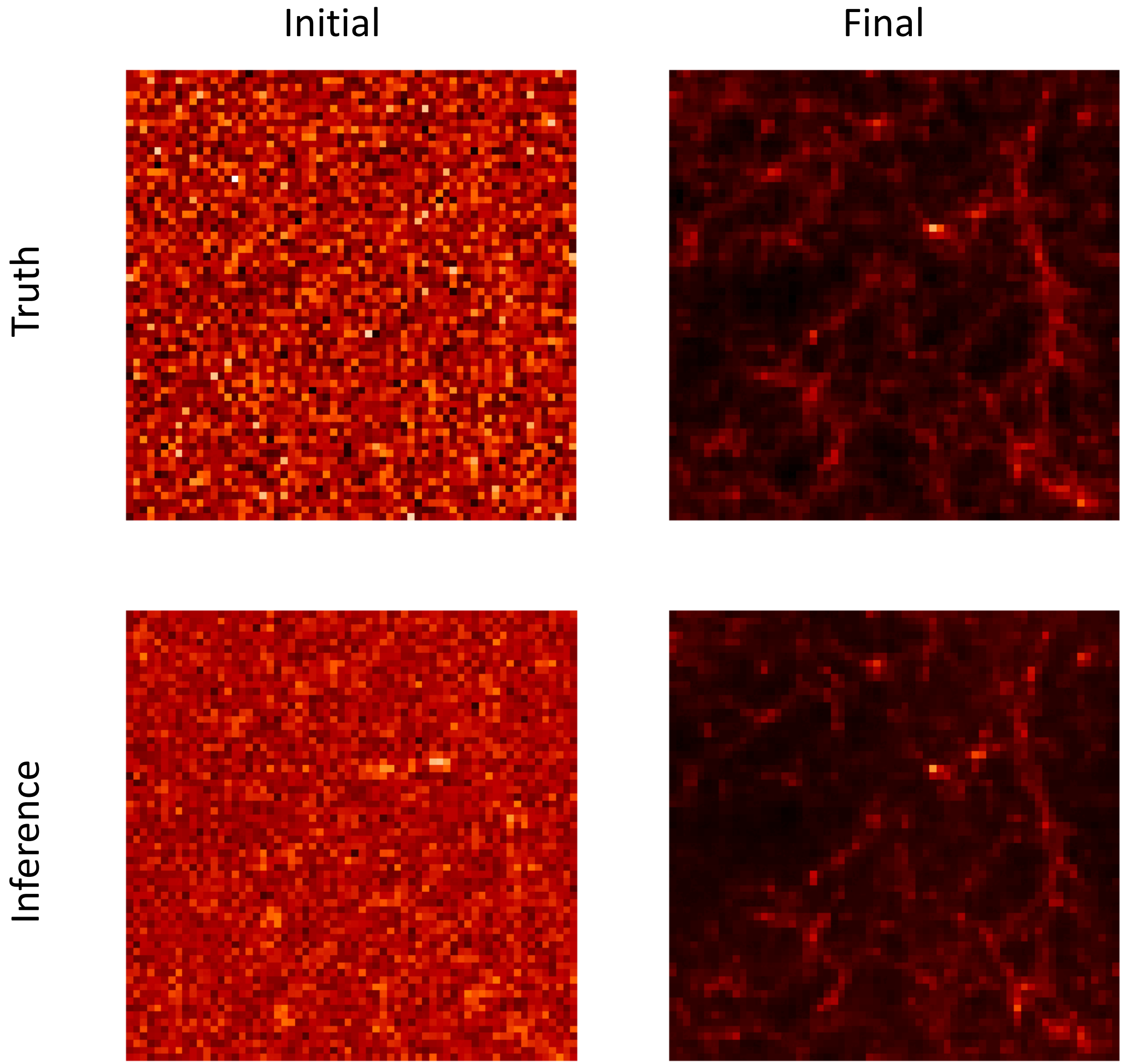
Bayer, Seljak, Modi (2023)



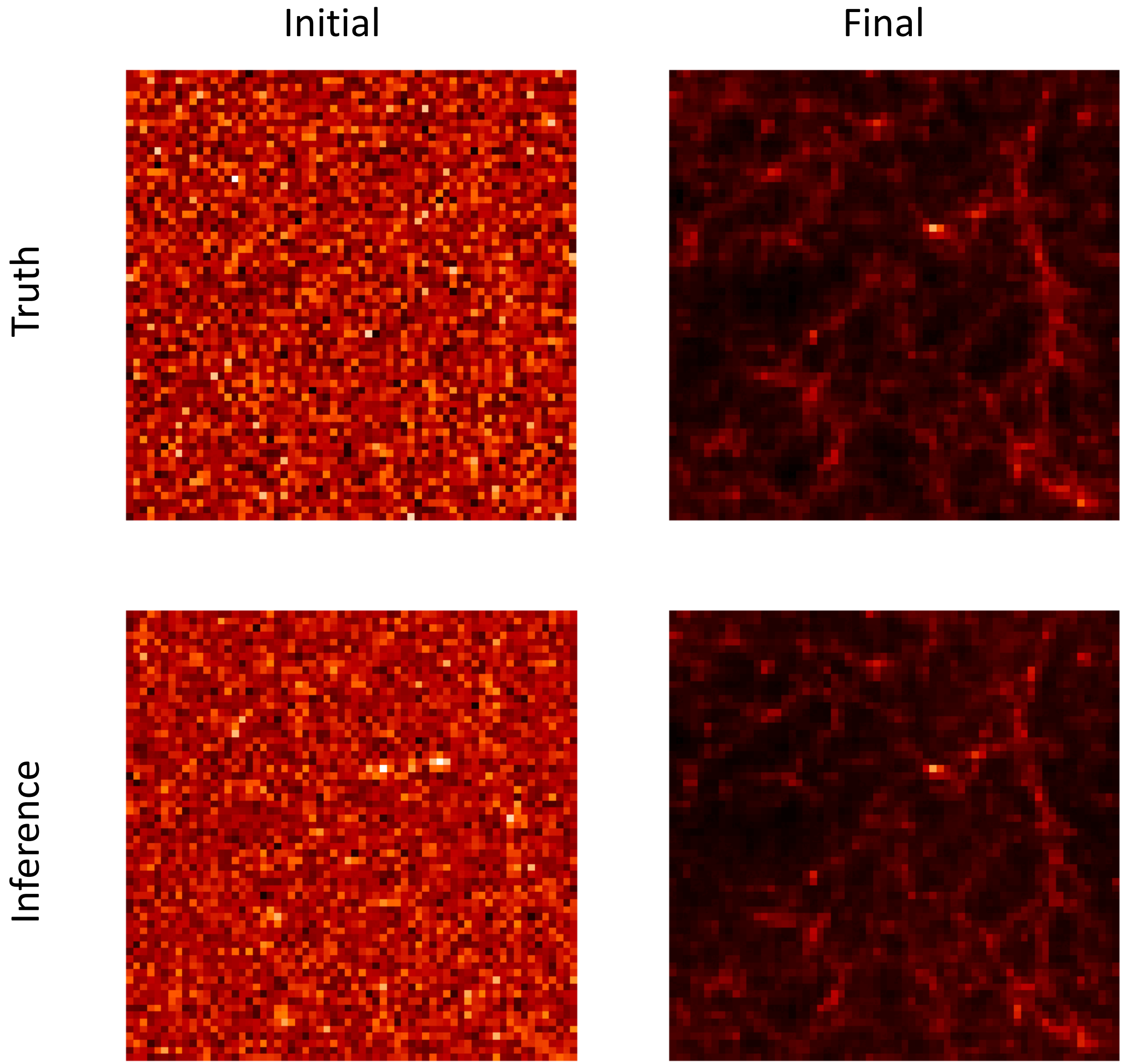
Bayer, Seljak, Modi (2023)



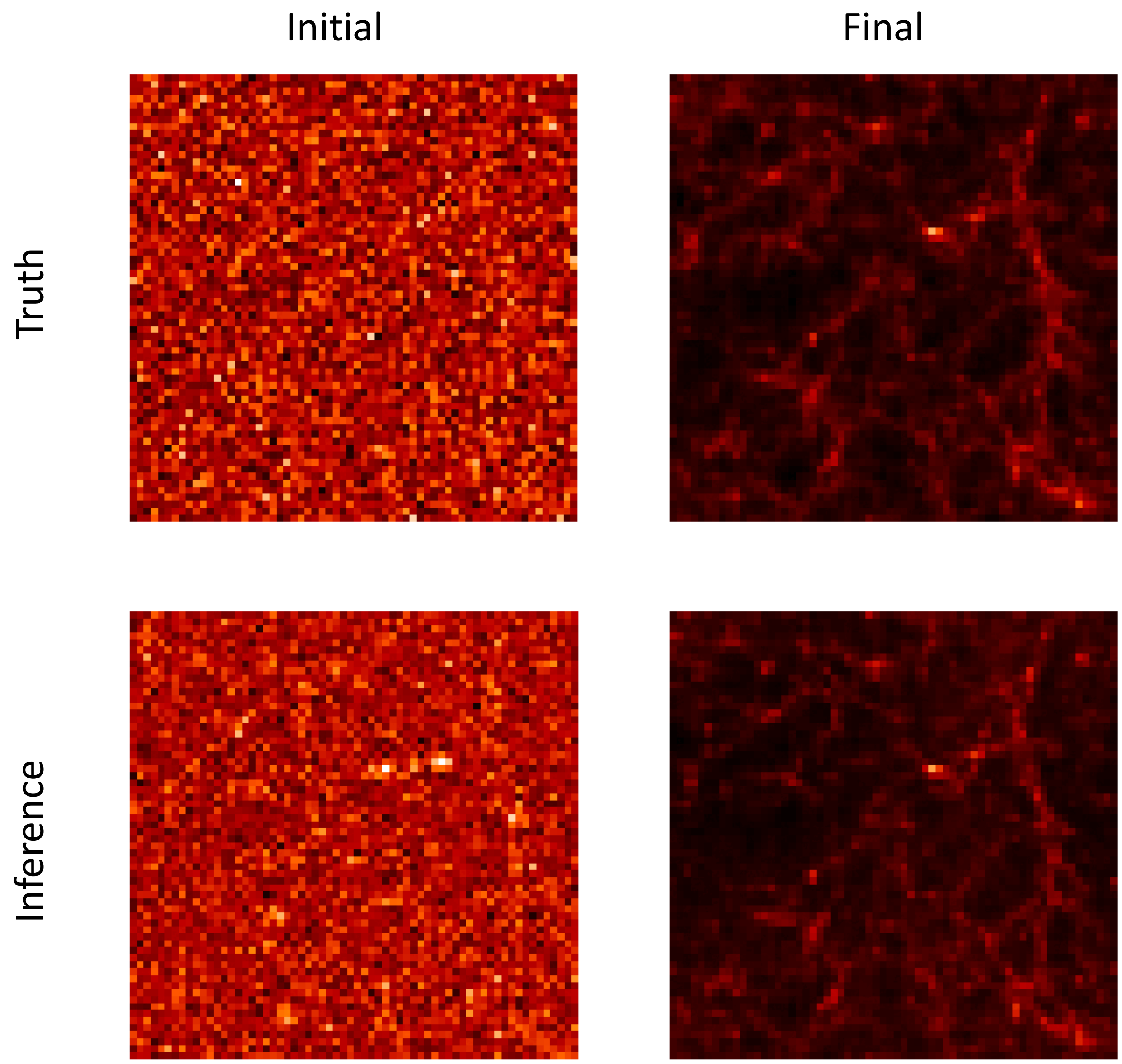
Bayer, Seljak, Modi (2023)



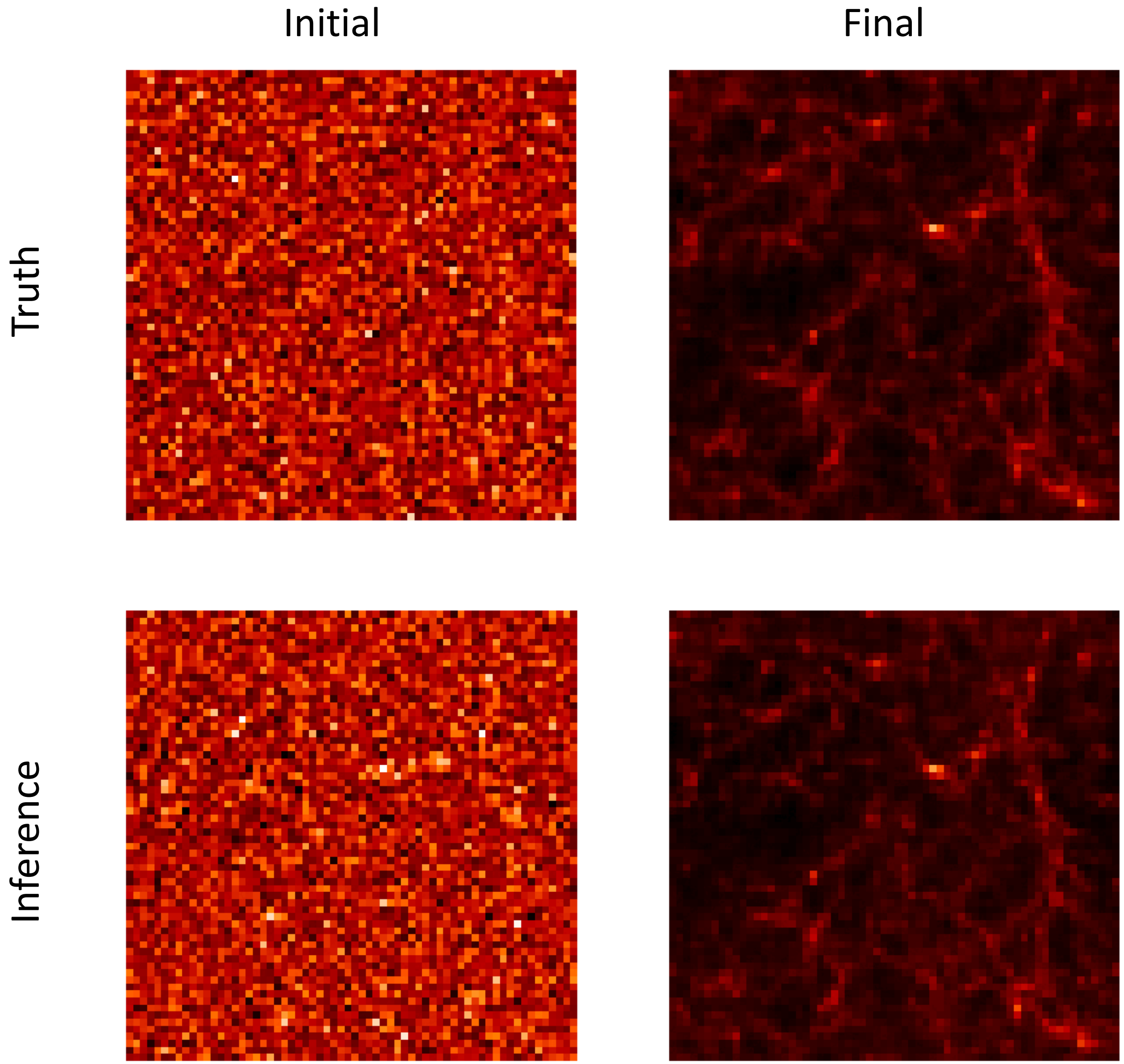
Bayer, Seljak, Modi (2023)



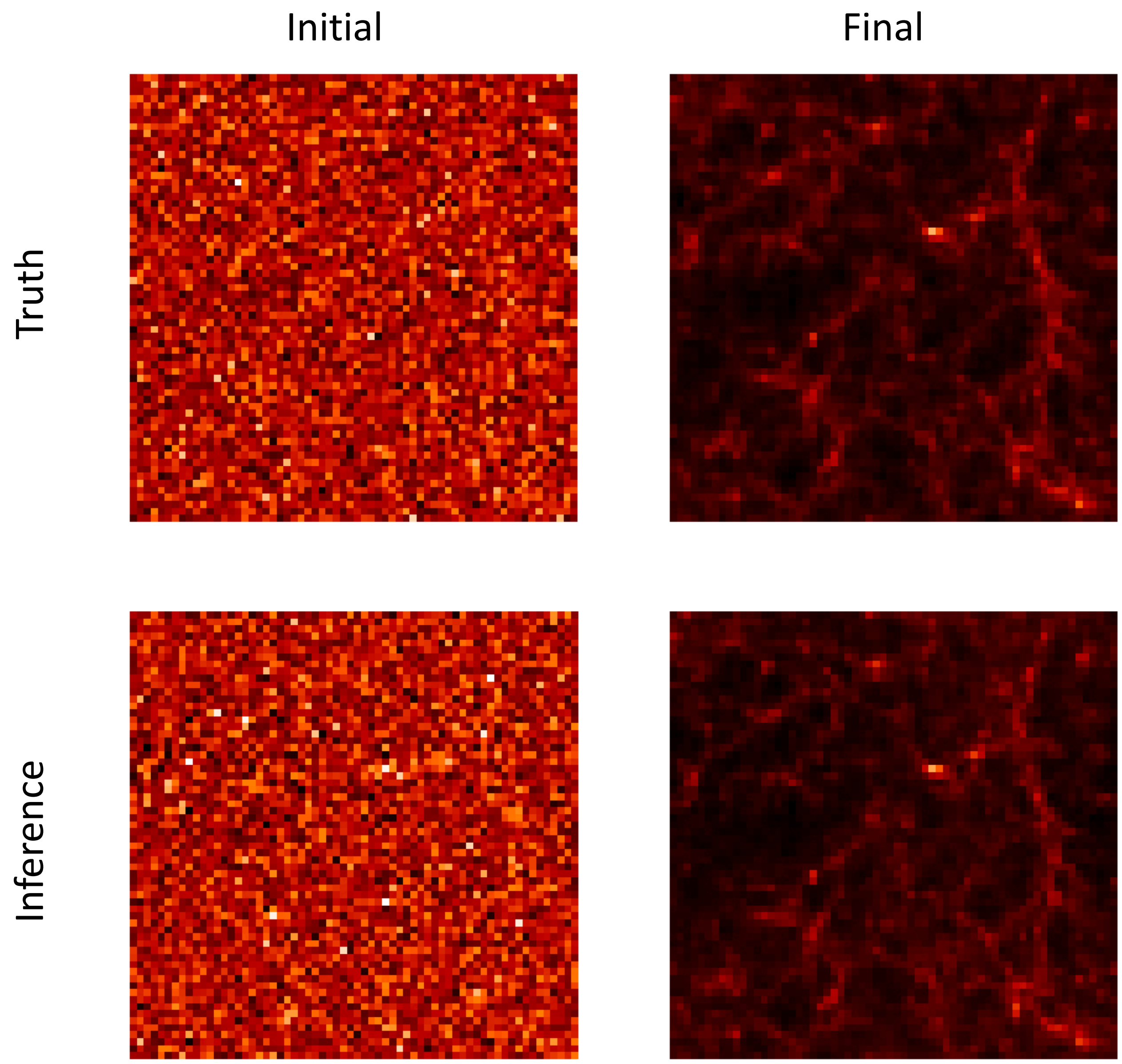
Bayer, Seljak, Modi (2023)



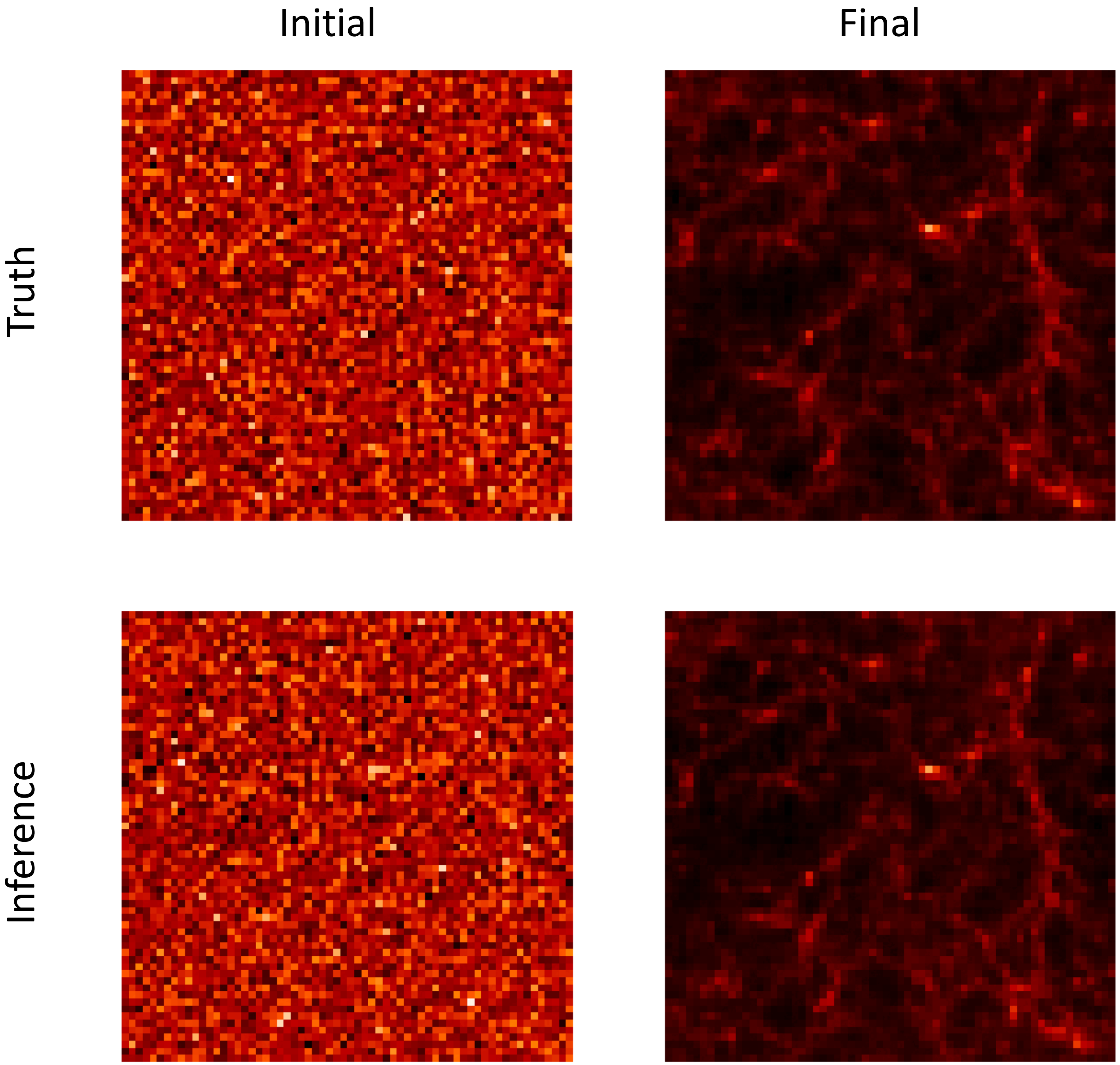
Bayer, Seljak, Modi (2023)



Bayer, Seljak, Modi (2023)



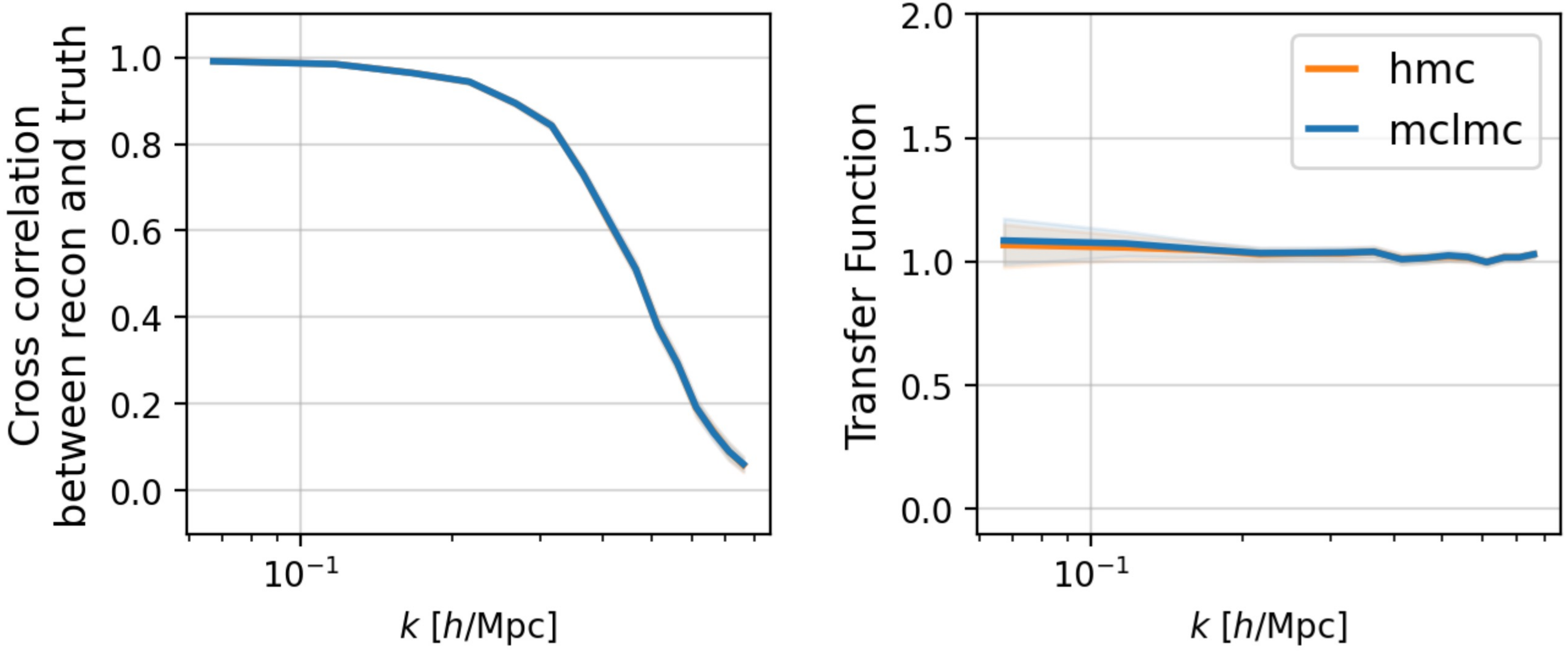
Bayer, Seljak, Modi (2023)



Bayer, Seljak, Modi (2023)

Samples of Initial Modes

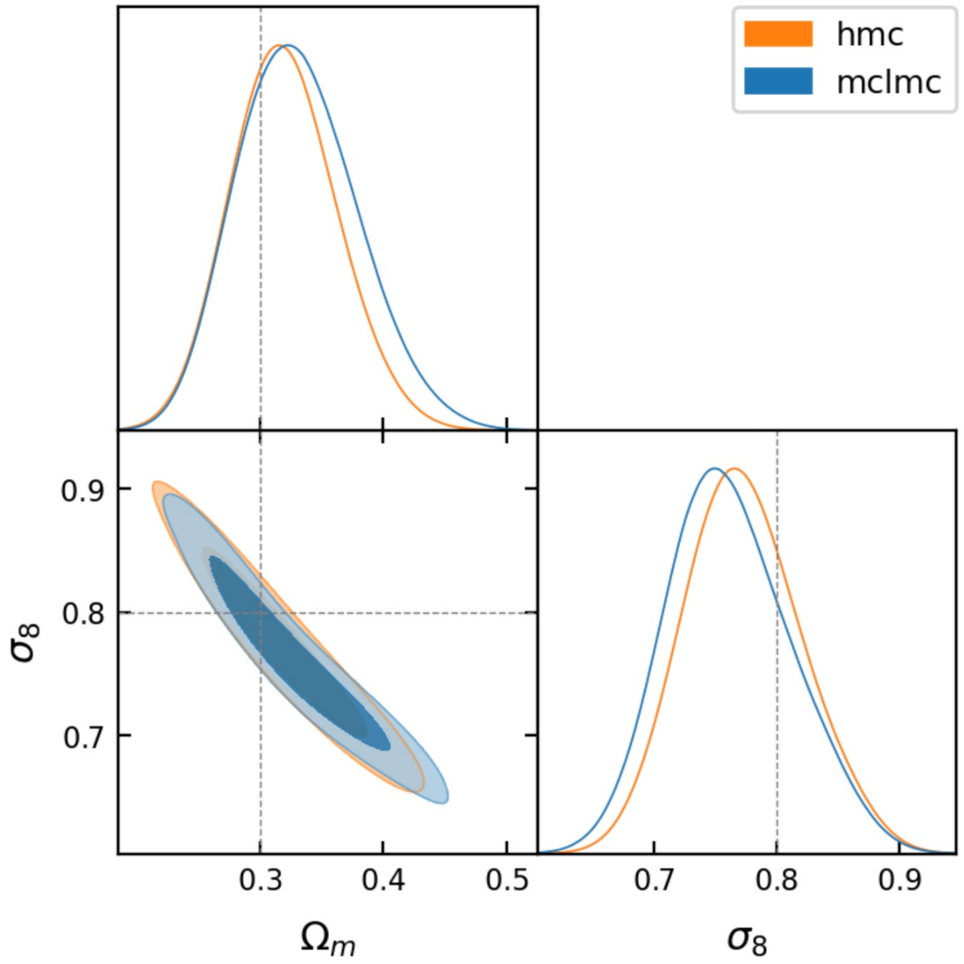
$d = 32^3 + 2 = 32{,}770$



Bayer, Seljak, Modi (2023)

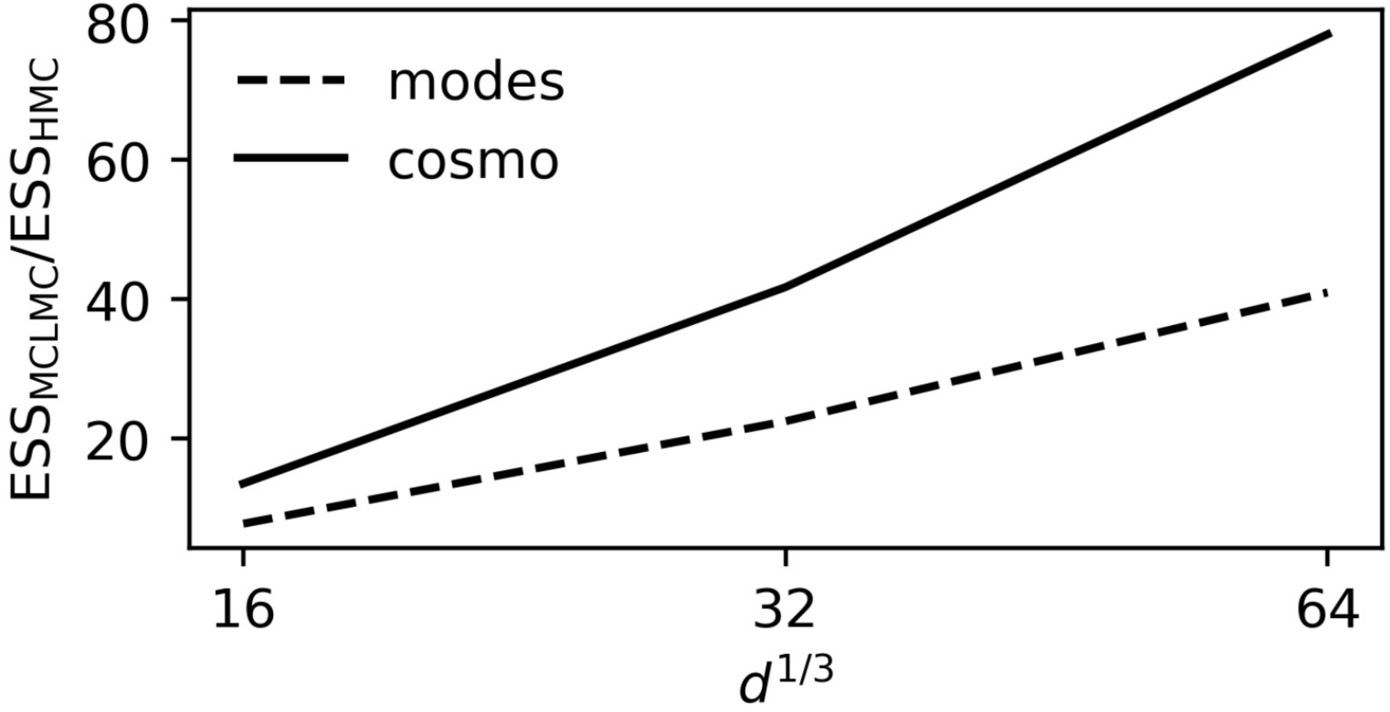
Samples of Cosmological Parameters

$d = 32^3 + 2 = 32{,}770$

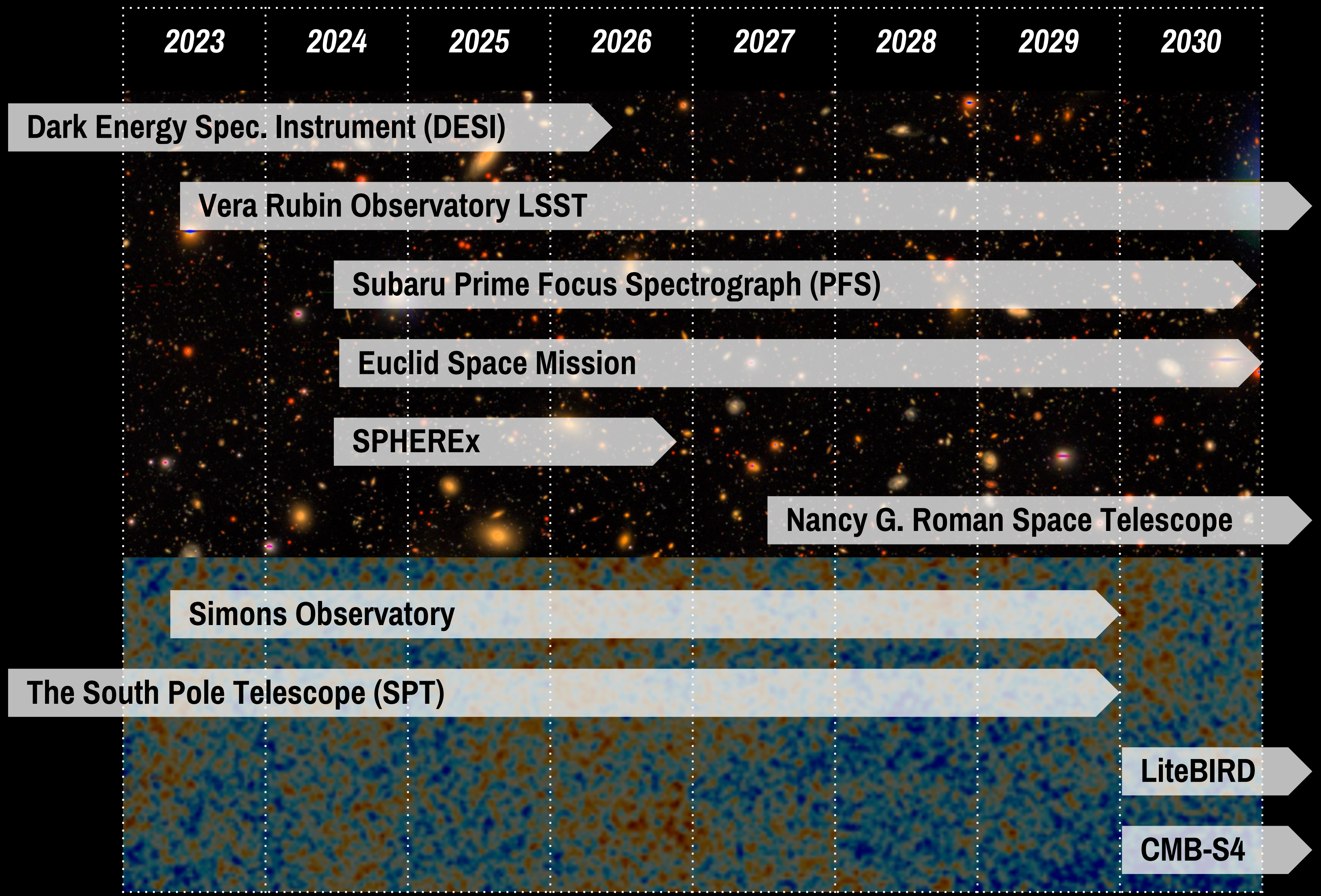


Bayer, Seljak, Modi (2023)

Efficiency improves with dimensionality!

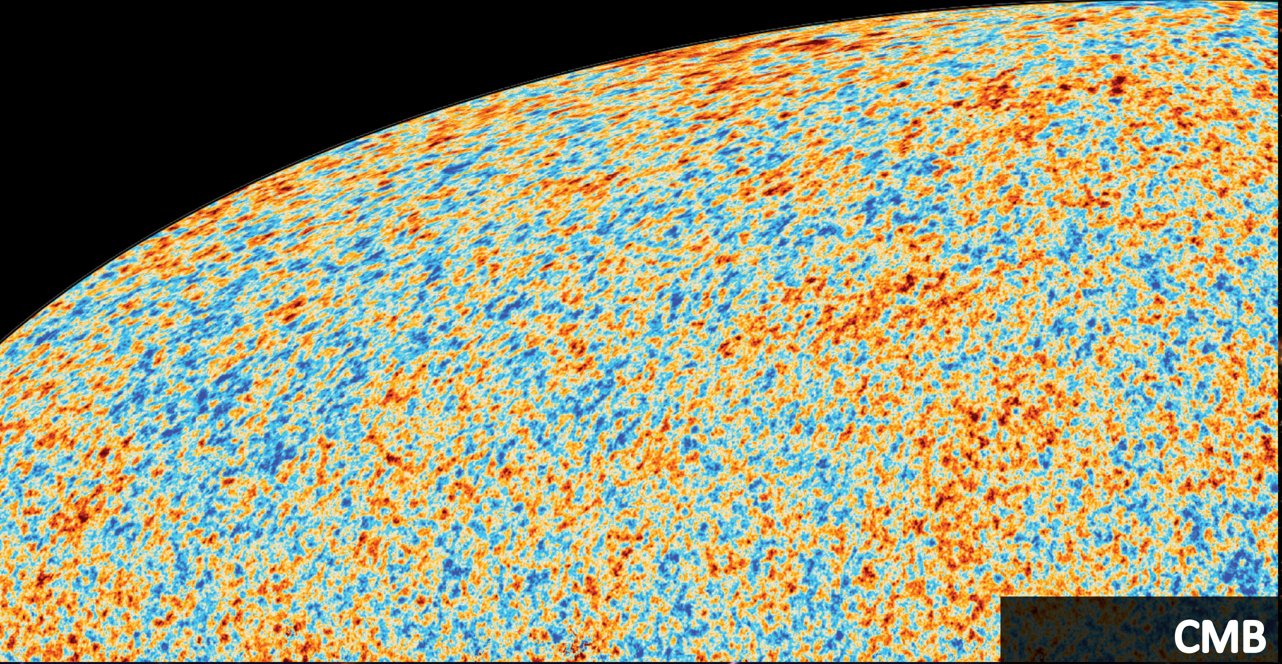


Bayer, Seljak, Modi (2023)


2023
2024
2025
2026
2027
2028
2029
2030
Dark Energy Spec. Instrument (DESI)
Vera Rubin Observatory LSST
Subaru Prime Focus Spectrograph (PFS)
Euclid Space Mission
SPHEREx
Nancy G. Roman Space Telescope
Simons Observatory
The South Pole Telescope (SPT)
LiteBIRD
CMB-S4

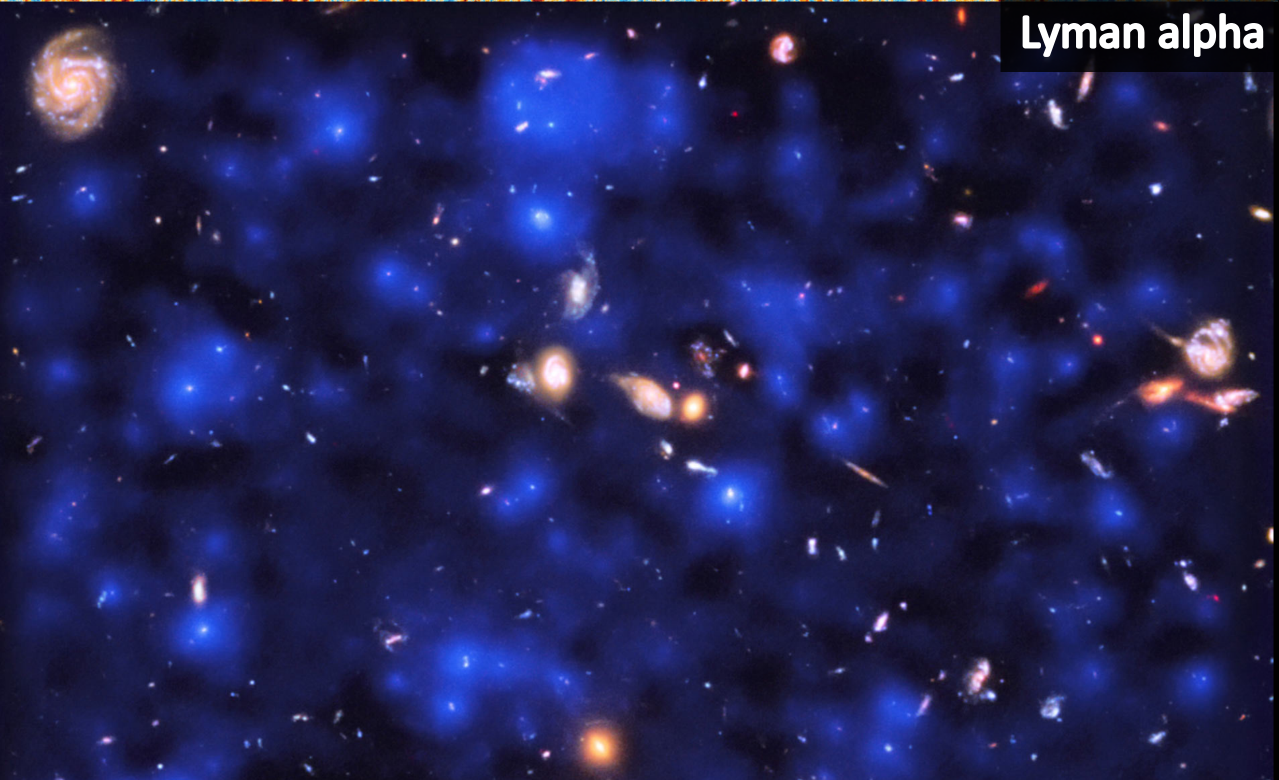
$$d \sim 512^3$$

CMB

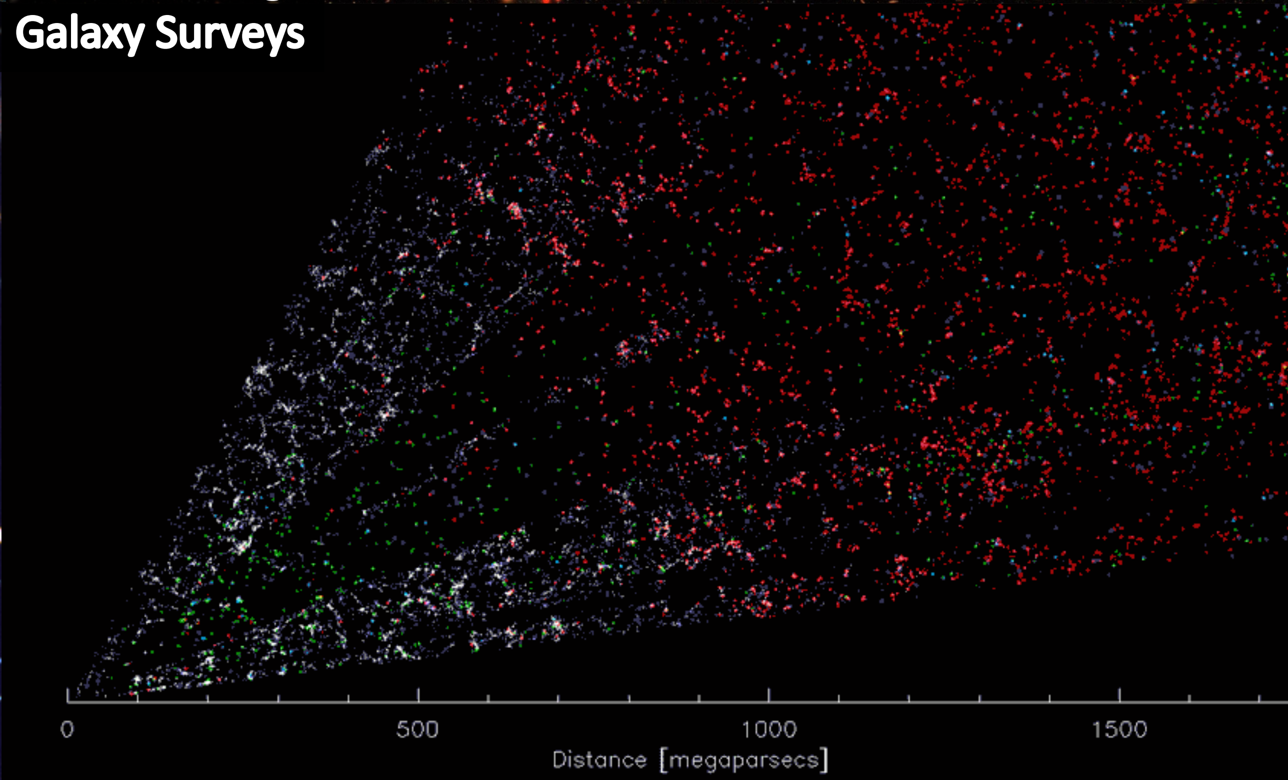


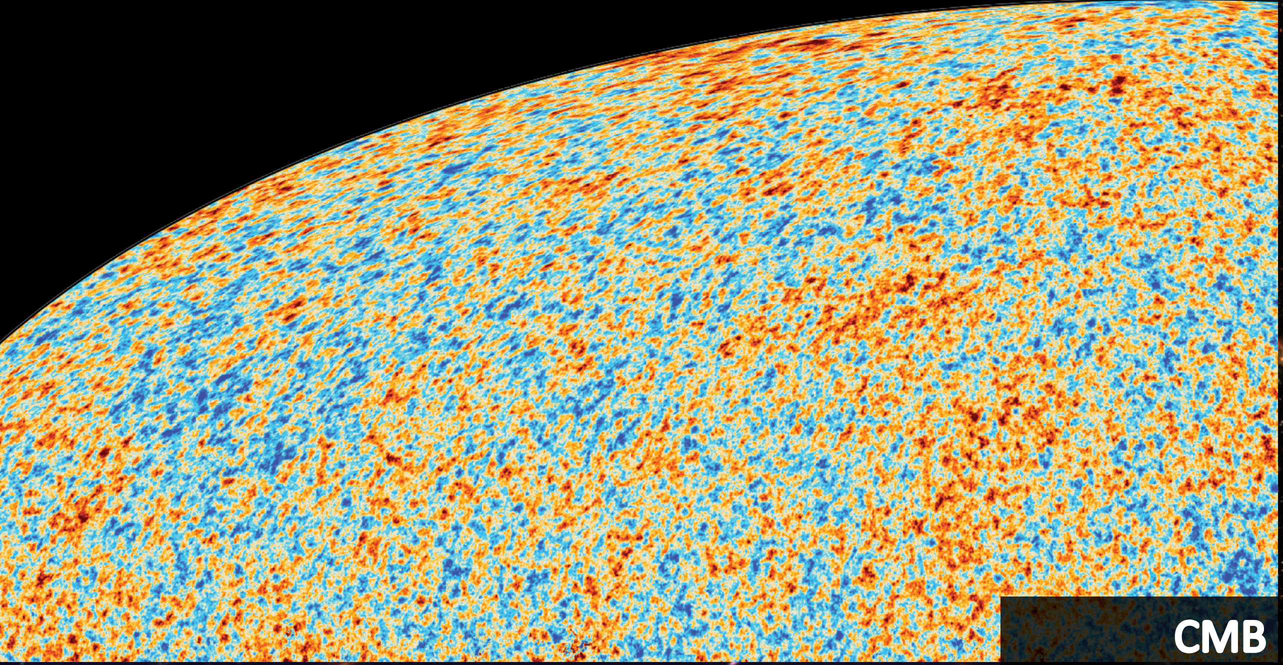
Weak Lensing

Lyman alpha



Galaxy Surveys





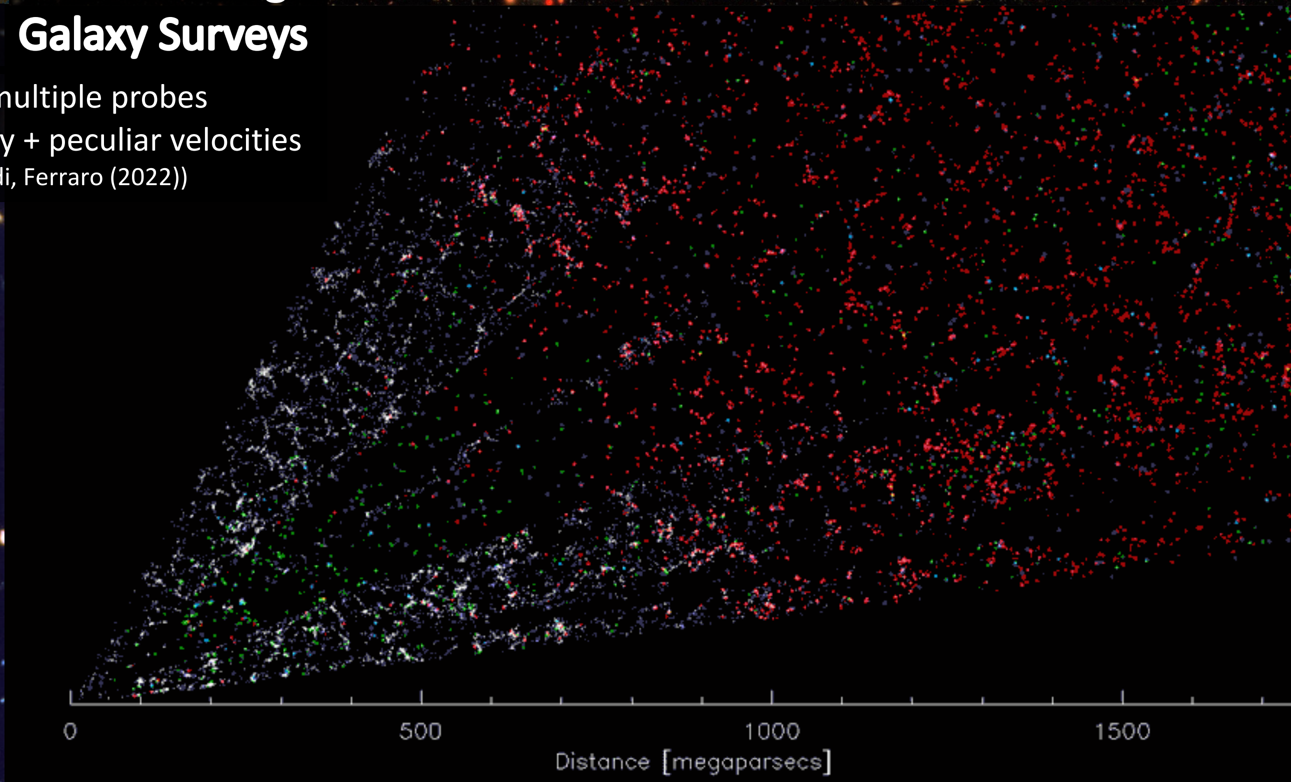
CMB **Weak Lensing**

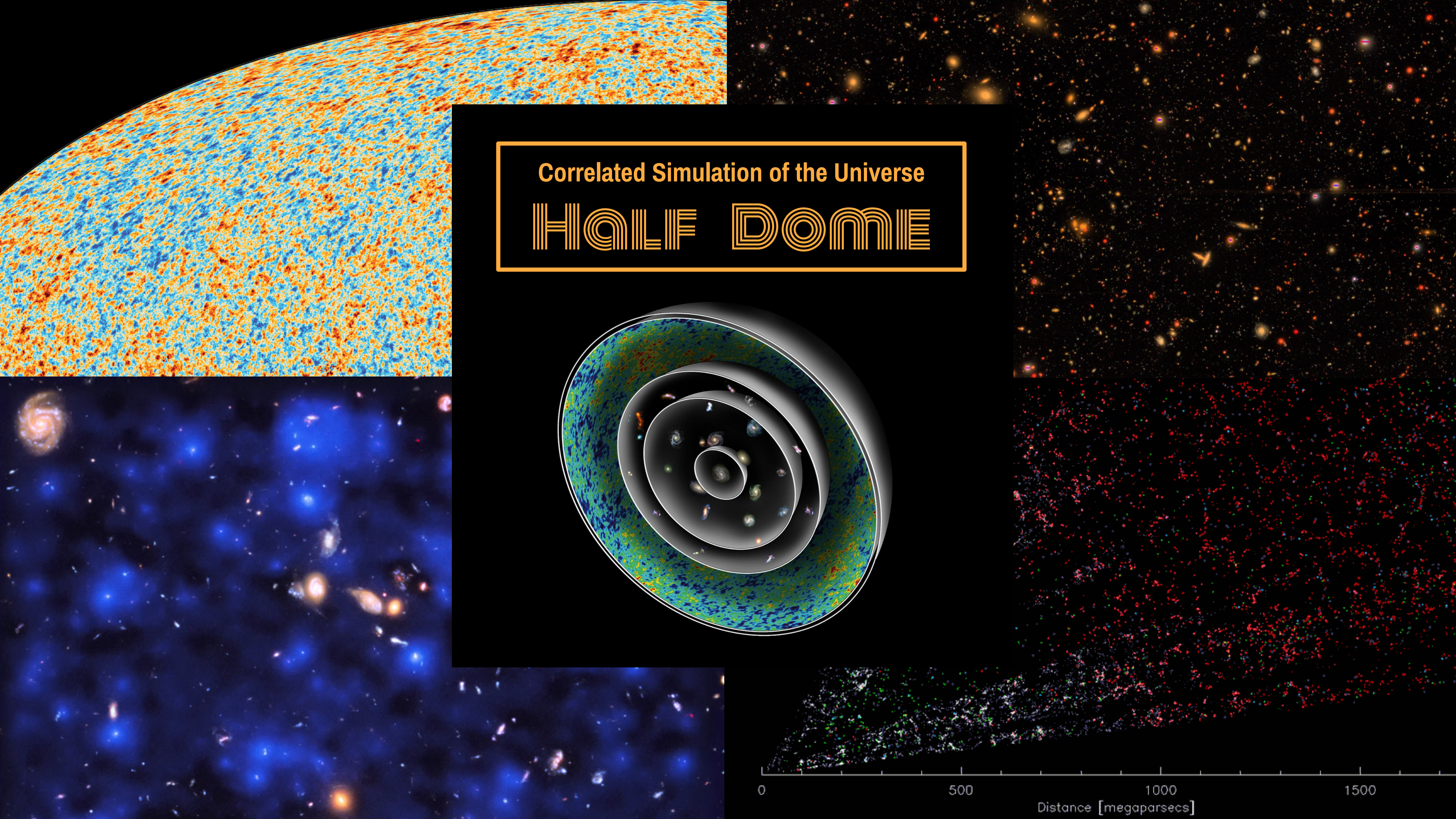
Lyman alpha **Galaxy Surveys**

Combine multiple probes

e.g. galaxy density + peculiar velocities

(Bayer, Modi, Ferraro (2022))




Correlated Simulation of the Universe
Half Dome
0
500
1000
1500
Distance [megaparsecs]

Simulation Comparison

	Sehgal+2010	**Websky** Stein+2020 Li+2022	**Agora** Omori 2022	**Stage IV requirements***
Box Size $N_{particles}$	1 Gpc/h 1024^3	7.7 Gpc 6144^3	1 Gpc/h 3840^3	a few Gpc/h
Min. M_{halo}	10^{13} $M_\odot$	1.2×10^{13} $M_\odot$	1.5×10^{9} $M_\odot$/h	10^{12} $M_\odot$/h
LSS observables	None	None	κ, clusters, LIM, +more to come	κ, galaxies,clusters
Number of realizations	1	1	1	10–100

* Inputs from SO, CMB-S4, LSST, DESI, PFS, SPHEREx, Roman collaborators

Simulation Comparison

	Sehgal+2010	**Websky** Stein+2020 Li+2022	**Agora** Omori 2022	**Stage IV requirements***
Box Size $N_{particles}$	1 Gpc/h 1024^3	7.7 Gpc 6144^3	1 Gpc/h 3840^3	a few Gpc/h
Min. M_{halo}	10^{13} $M_\odot$	1.2×10^{13} $M_\odot$	1.5×10^9 $M_\odot/h$	10^{12} $M_\odot/h$
LSS observables	None	None	κ, clusters, LIM, +more to come	κ, galaxies,clusters
Number of realizations	1	1	1	10–100

* Inputs from SO, CMB-S4, LSST, DESI, PFS, SPHEREx, Roman collaborators

Simulation Comparison

	Sehgal+2010	**Websky** Stein+2020 Li+2022	**Agora** Omori 2022	**Stage IV requirements***	**Half Dome** 2024-
Box Size $N_{particles}$	1 Gpc/h 1024^3	7.7 Gpc 6144^3	1 Gpc/h 3840^3	a few Gpc/h	3.5 Gpc/h, 6144^3
Min. M_{halo}	10^{13} $M_\odot$	1.2×10^{13} $M_\odot$	1.5×10^9 $M_\odot$/h	10^{12} $M_\odot$/h	10^{12} $M_\odot$/h
LSS observables	None	None	κ, clusters, LIM, +more to come	κ, galaxies,clusters	κ, galaxies, clusters, +more
Number of realizations	1	1	1	**10–100**	$11+1f_{NL}$ (more to come)

* Inputs from SO, CMB-S4, LSST, DESI, PFS, SPHEREx, Roman collaborators

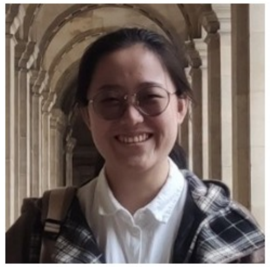
The Team


Adrian Bayer
Princeton/CCA


Jia Liu
Kavli IPMU

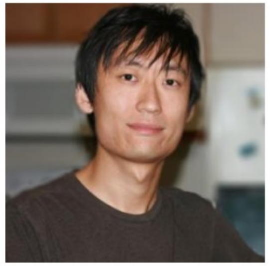

Zack Li
UC Berkeley/LBL


Joe DeRose
UC Berkeley/LBL


Yici Zhong
U Tokyo


Linda Blot
Kavli IPMU

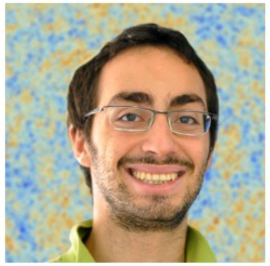

Marcelo Alvarez
LBL


Yu Feng
Google


Junjie Xia
Kavli IPMU


Alex Laguë
U Penn


Will Coulton
Cambridge

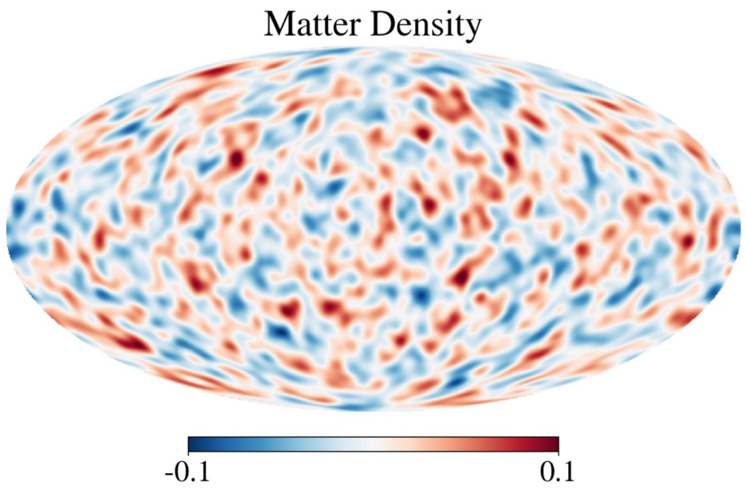

Giuseppe Puglisi
U Catania

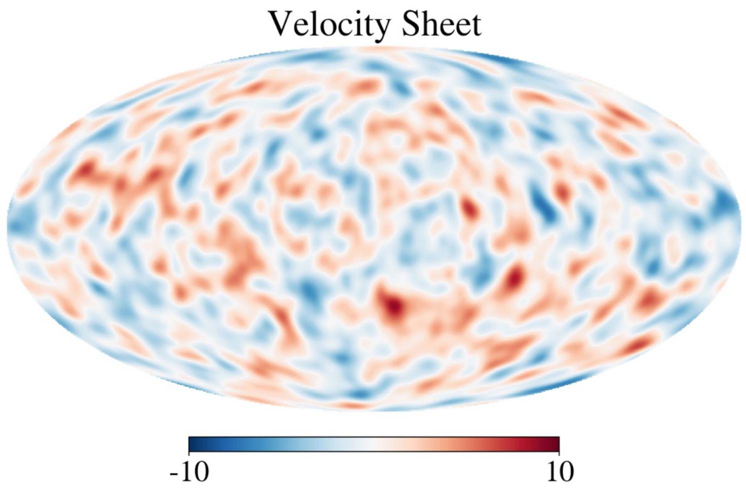

Hideki Tanimura
Kavli IPMU

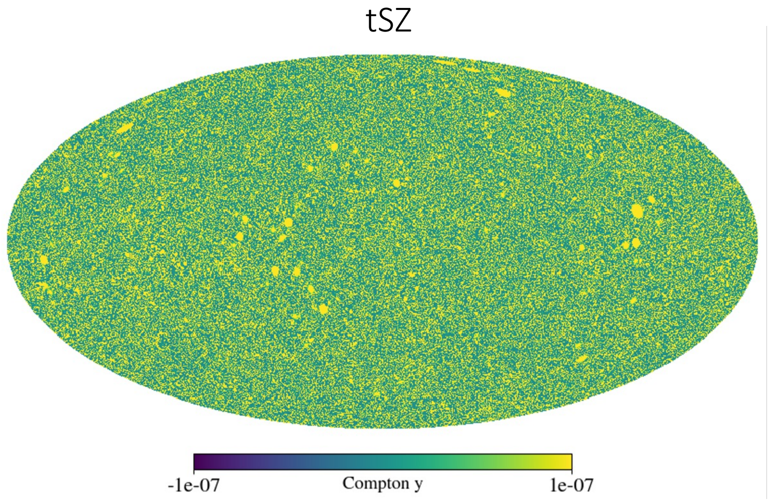

Mathew Madhavacheril
U Penn

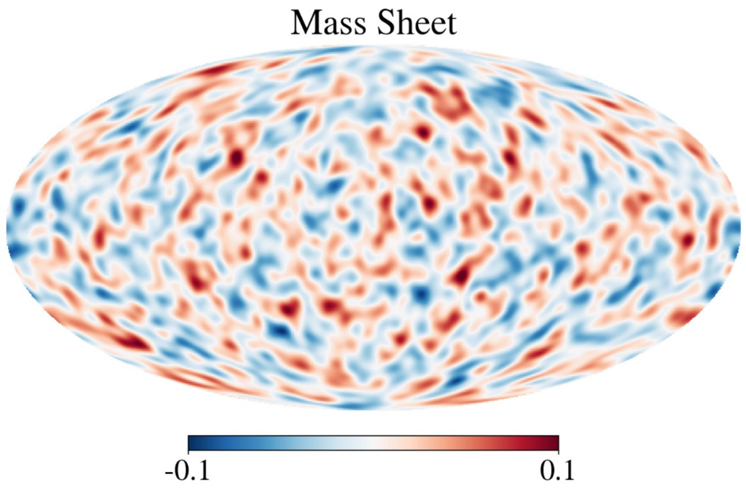
Preliminary Maps

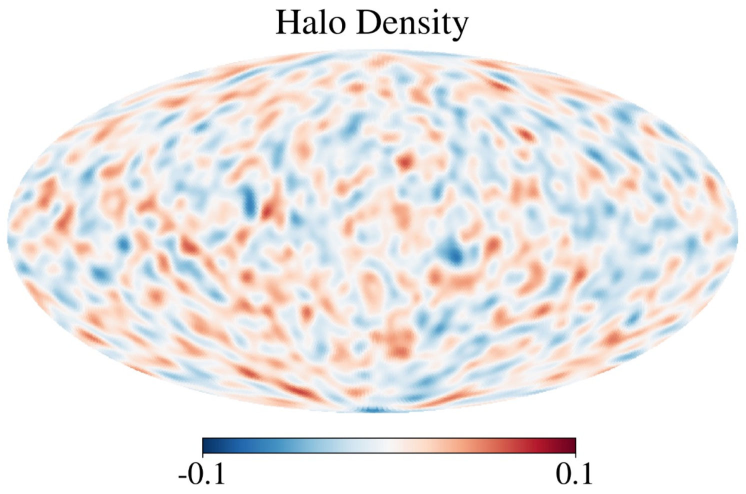
stay tuned for more!

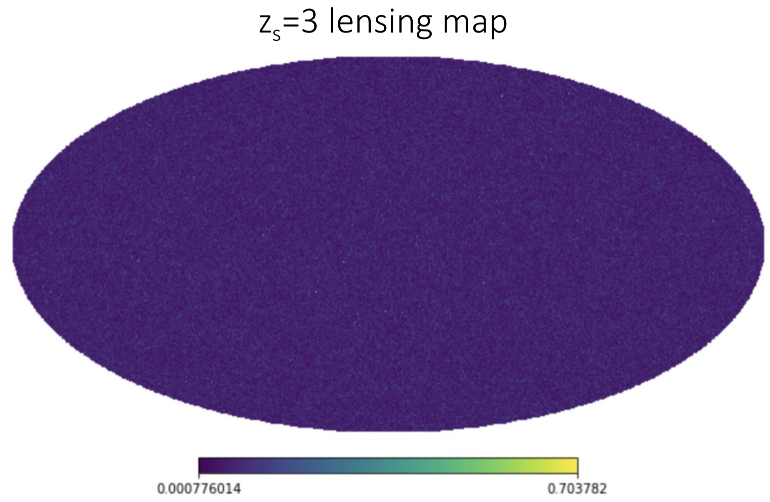


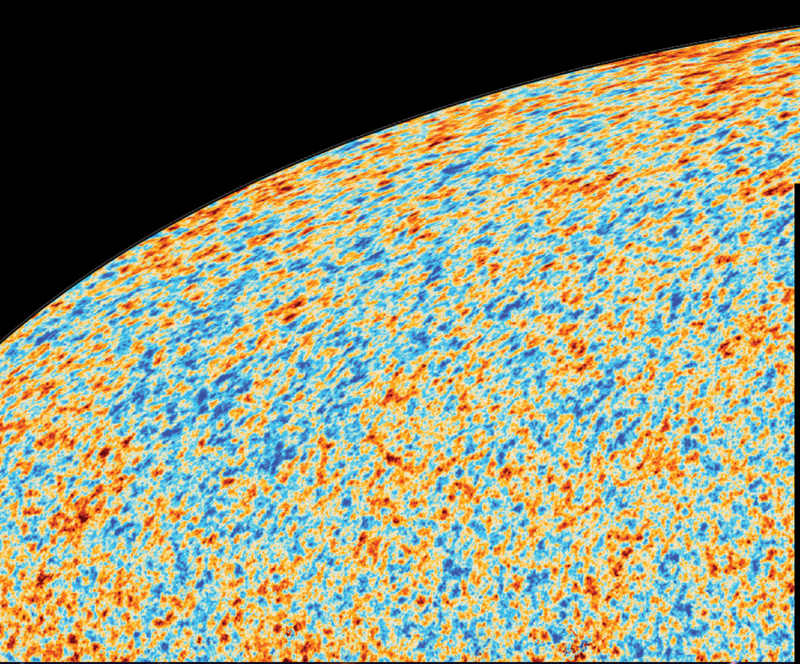


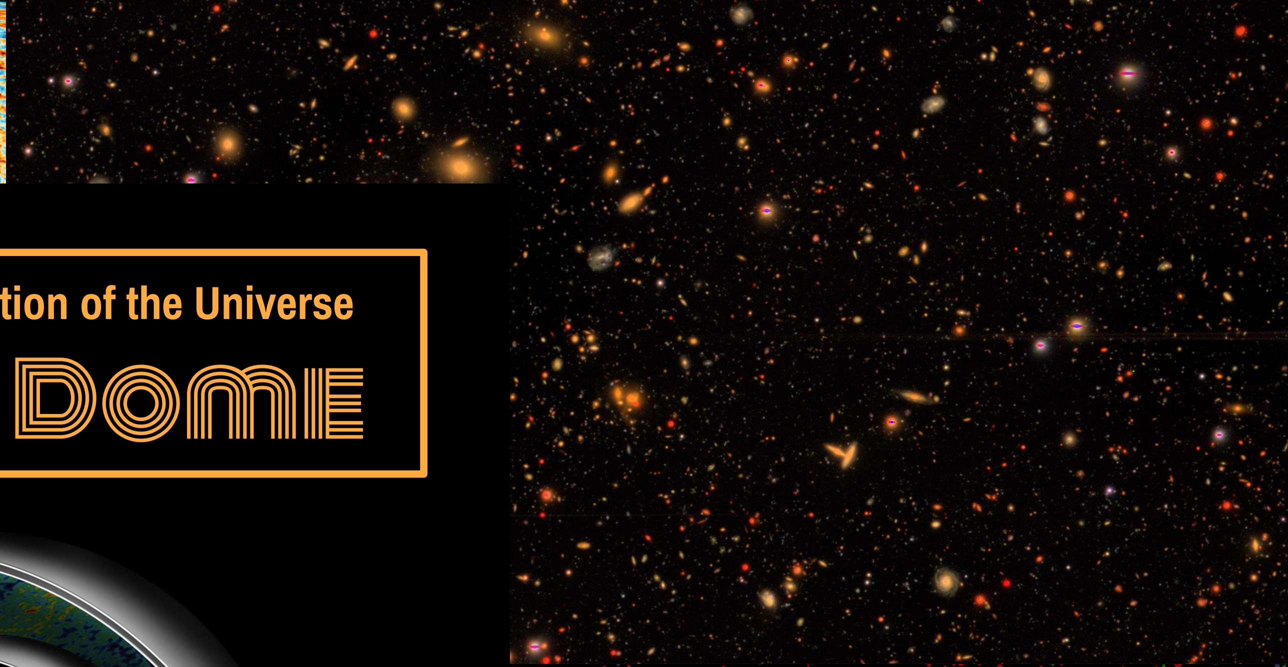






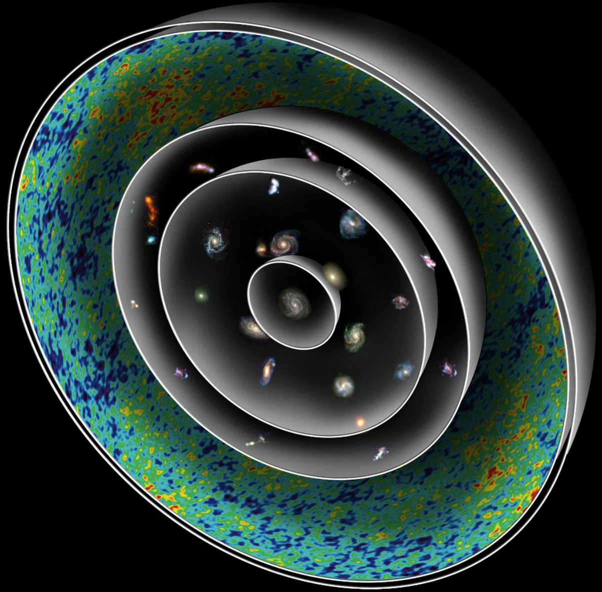


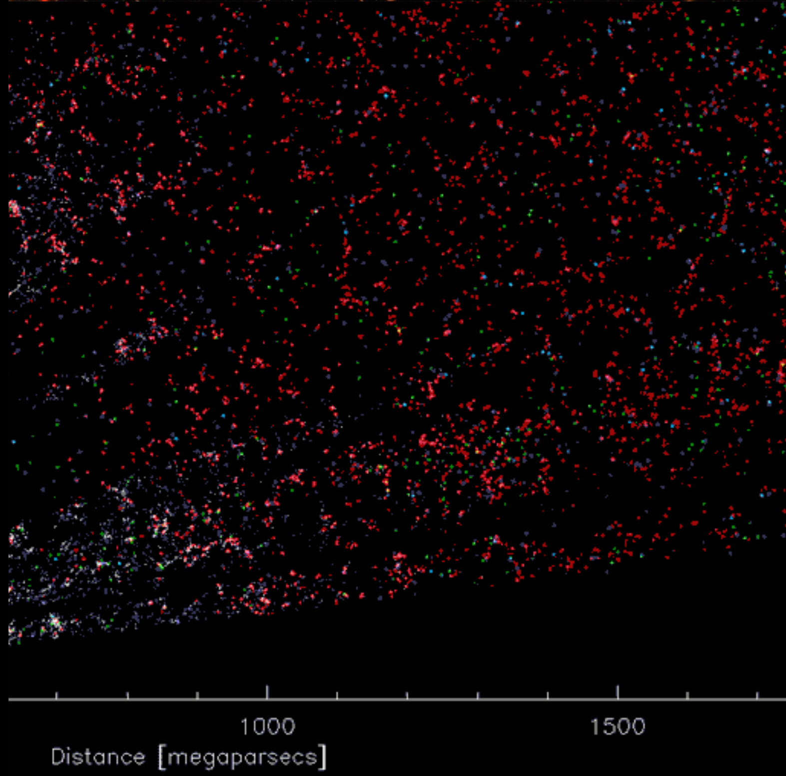




Correlated Simulation of the Universe

HALF DOME





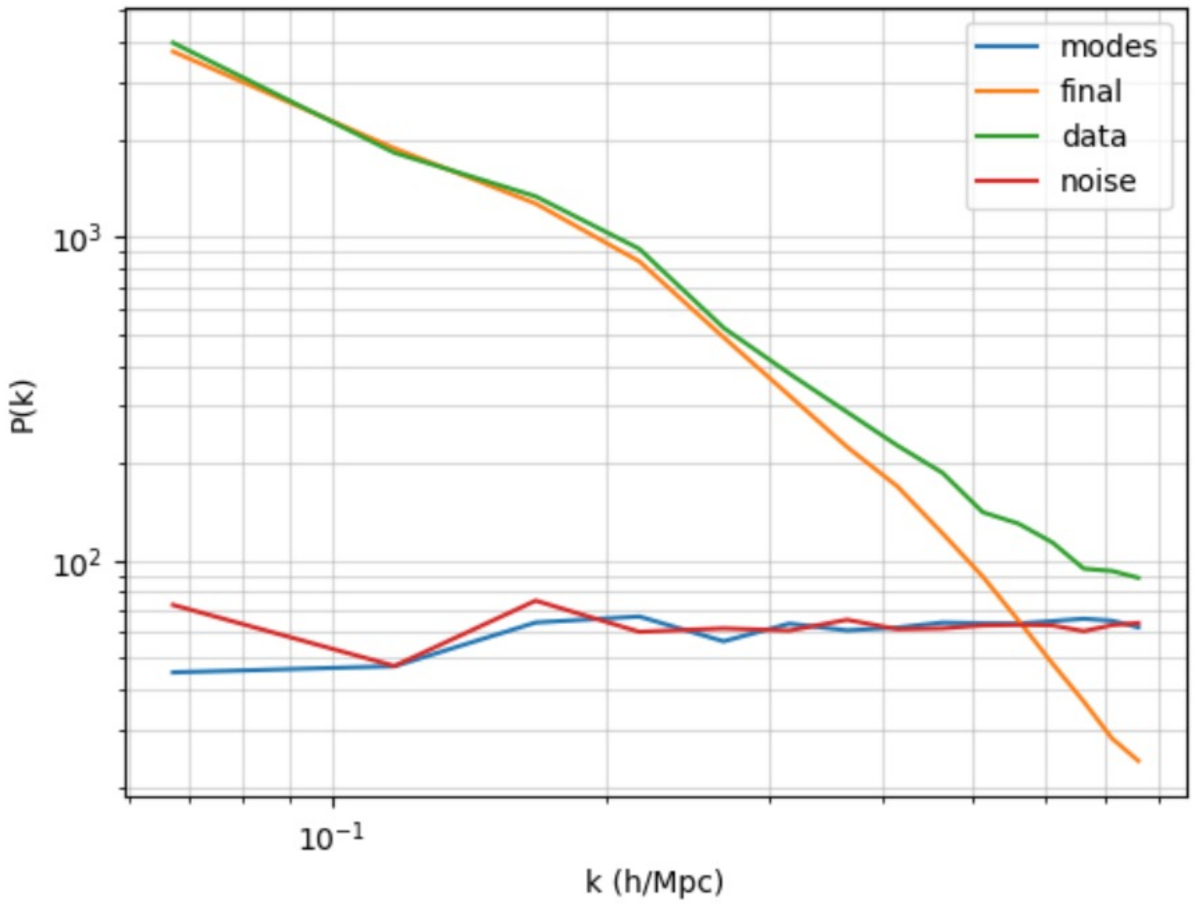
Thank you!

Adrian Bayer

http://adrianbayer.github.io

Experiments

- Nonlinear dark matter
  - 1LPT, 2LPT, 5-step PM
- Dimensionality
  - 16^3, 32^3, 64^3
- Voxel size 4 Mpc/h
- Inject noise



Bayer, Seljak, Modi (2023)

2.2. Hamiltonian Monte Carlo

The traditional approach for sampling in the context of field-level inference is HMC (Duane et al., 1987; Neal et al., 2011; Betancourt, 2017). Given a d-dimensional target distribution $p(z) \propto e^{-\mathcal{L}(z)}$, where $z \in \mathbb{R}^d$, HMC uses the gradient $\nabla\mathcal{L}(z)$ to improve the sampling efficiency compared no-gradient methods such as MH. It considers the Hamiltonian $H(z, \Pi)$, where Π is the canonical momentum, and samples the canonical ensemble in $2d$-dimension phase space, denoted by $p(z, \Pi) \propto e^{-H(z,\Pi)}$. The success of HMC relies on the tuning of the Hamiltonian such the the marginal of $p(z, \Pi)$ over Π converges to the target distribution,

$$p(z) \propto \int_{\mathbb{R}^d} d\Pi \, e^{-H(z,\Pi)}. \tag{2}$$

The most popular choice is the Hamiltonian of a free particle, $H(z, \Pi) = \frac{1}{2}\Pi^2(z) + \mathcal{L}(z)$, for which the solution is the set of ODEs,

$$\begin{aligned} dz &= udt, \\ du &= -\nabla\mathcal{L}(z)dt, \end{aligned} \tag{3}$$

where t is time and u is velocity. Following Hamiltonian dynamics ensures the trajectory conserves the Hamiltonian, or energy, allowing efficient exploration at a fixed energy level. Different energy levels must be explored to obtain an accurate set of samples in HMC, which is achieved by resampling the momentum Π according to its marginal distribution (a normal distribution) and results in inefficiencies (Betancourt, 2017). Moreover, to ensure the target distribution is converged to, HMC additionally requires an MH accept-reject step, which in turn requires a sufficiently small step size to ensure a frequent rate of acceptance.

2.3. Microcanonical Hamiltonian Monte Carlo

Unlike HMC which considers the marginal of the canonical distribution, the approach of MCHMC is to tune the Hamiltonian such that the microcanonical distribution marginalized over the momentum variables gives the target distribution, as follows

$$p(z) \propto \int_{\mathbb{R}^d} d\Pi \, \delta(H(z, \Pi) - E), \tag{4}$$

where $\delta(\cdot)$ denotes the delta function, and E is the energy. The motion of a particle under this Hamiltonian can be written as a set of ODEs as follows

$$\begin{aligned} dz &= udt, \\ du &= P(u)f(z)dt, \end{aligned} \tag{5}$$

where we have introduced the projection $P(u) \equiv (I - uu^T)$ and force $f(z) \equiv -\nabla\mathcal{L}(z)/(d-1)$ (Ver Steeg & Galstyan, 2021; De Luca & Silverstein, 2022; Robnik et al., 2022). The key difference to the HMC ODEs in Eqn. (3) is the projection operator. Unlike HMC, the MCHMC dynamics converges to the target distribution while maintaining a constant energy.

2.4. Microcanonical Langevin Monte Carlo

To further speed up reaching ergodicity, the ODEs can be modified by considering Langevin dynamics (Grenander & Miller, 1994; Girolami & Calderhead, 2011) such that,

$$dz = udt, \tag{6}$$

$$du = P(u)\left[f(z)dt + \eta dW\right], \tag{7}$$

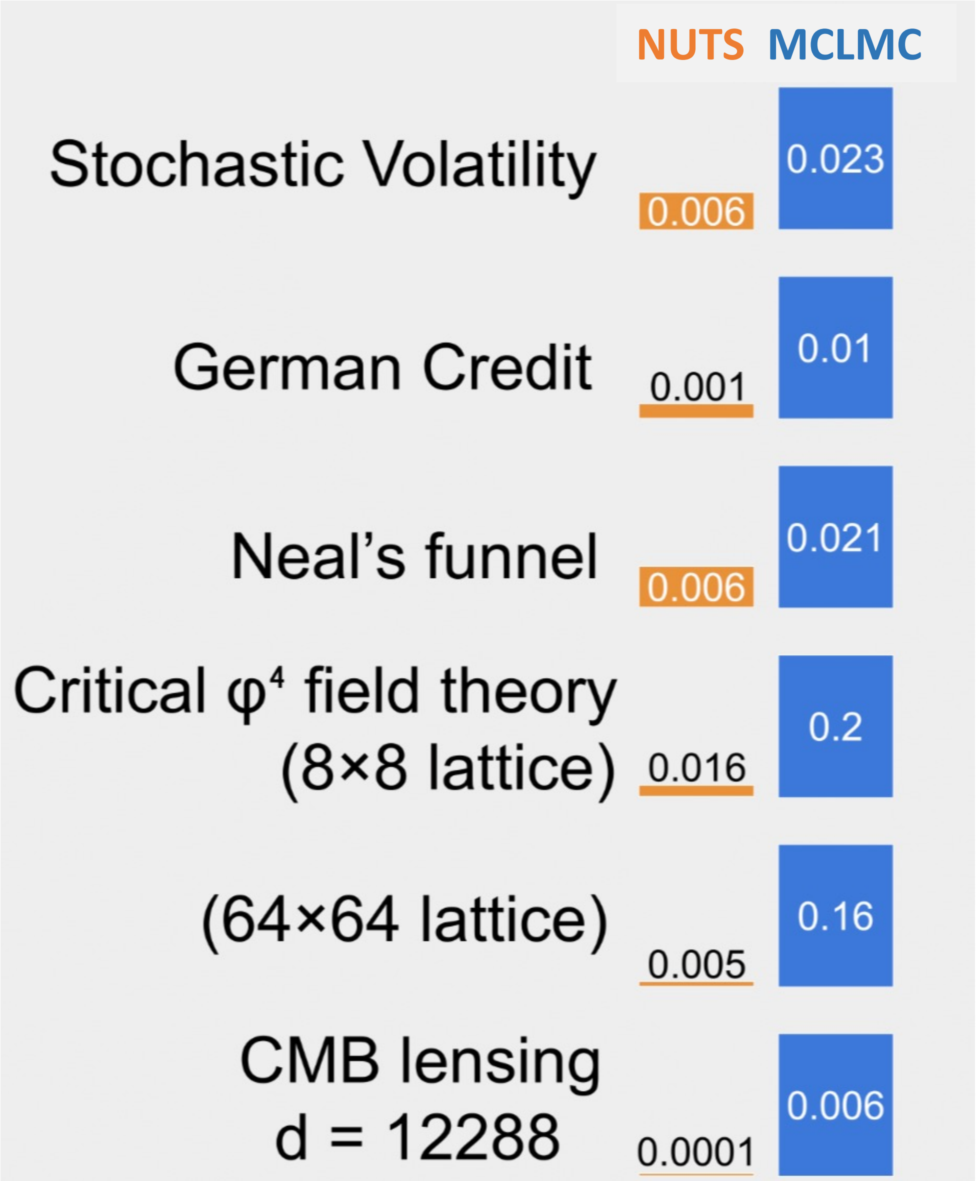
where η is a hyperparameter and W is a standard normal random vector. This additional term proportional to η can be understood physically as a diffusion term which enforces better exploration of the target, in turn boosting ergodicity.

MCLMC has two hyperparameters, the step size and the amount of noise η. Both of these parameters can be tuned during a burn-in stage by monitoring fluctuations in the energy and ensuring they are below a certain threshold.

Many other applications

- Lattice Field Theory (Robnik+2023)
- CMB (Bonici+2023)
- CMB lensing (Ruiz-Zapatero+, in prep)
- …

- Have an application? Let’s chat!



Talk structure: the audience is from diverse academic backgrounds, so we request that your talk addresses 3 aspects:

(1) the key scientific question you are trying to answer;
(2) why is AI/ML suitable for this question;
(3) performance comparison to previous (non-AI) methods.