Higher Order Statistics for Neutral Hydrogen Intensity Mapping

Pauline Gorbatchev

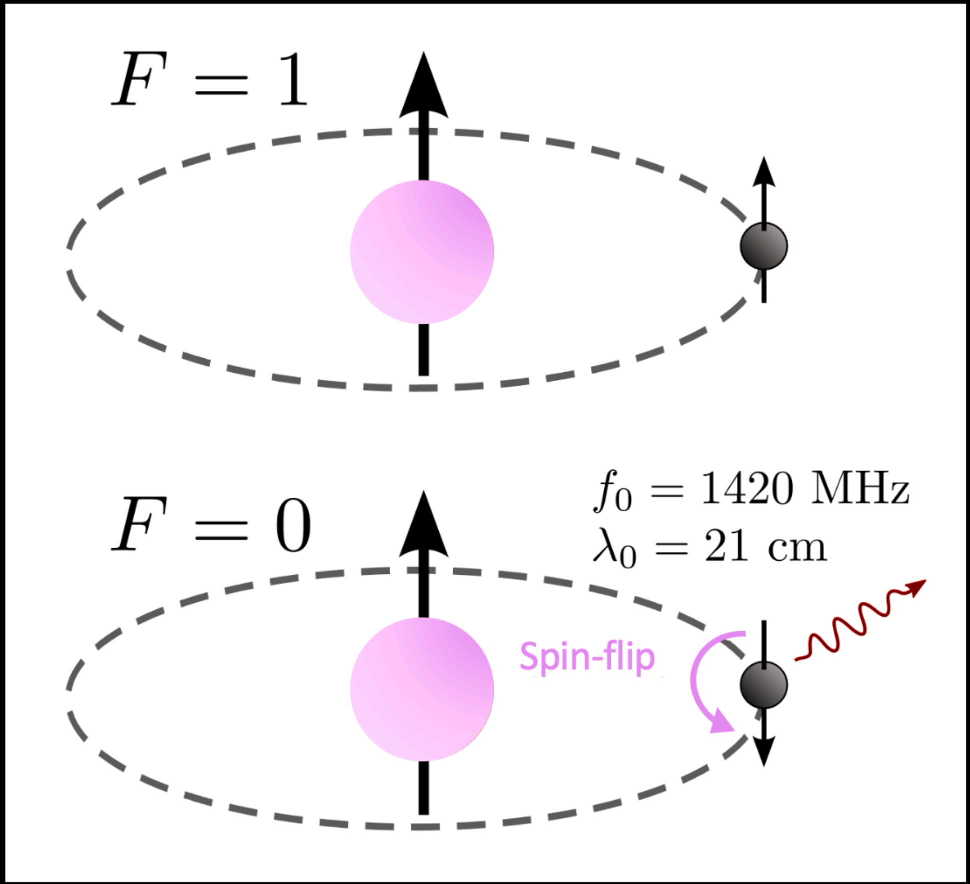
2025 Joint ARGOS-TITAN-TOSCA workshop

TITAN ARTIFICIAL INTELLIGENCE IN ASTROPHYSICS
FORTH INSTITUTE OF COMPUTER SCIENCE
Signal Processing Laboratory
CosmoStat
cea
Funded by the European Union

Physical background

- Hydrogen: most abundant element in the Universe.
- After reionization, HI is located inside galaxies.

$\Longrightarrow$ **biased tracer** of the underlying **matter distribution** of the Universe.



HI hyperfine transition

Why using 21cm line ?

Benefits:

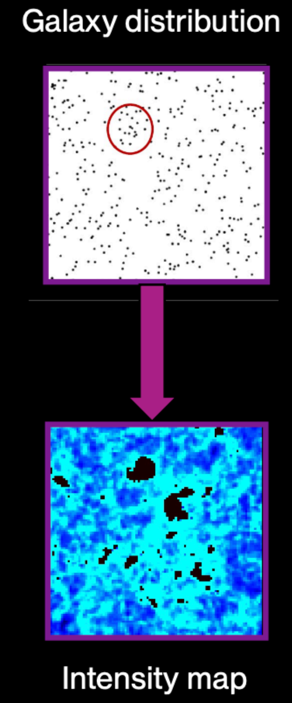
- Can be **measured from earth** (penetrates the atmosphere).
- thermal noise in HI surveys is less important than shot noise in galaxy surveys $\Longrightarrow$ HI analysis is **more constraining than galaxies.**

Uses:

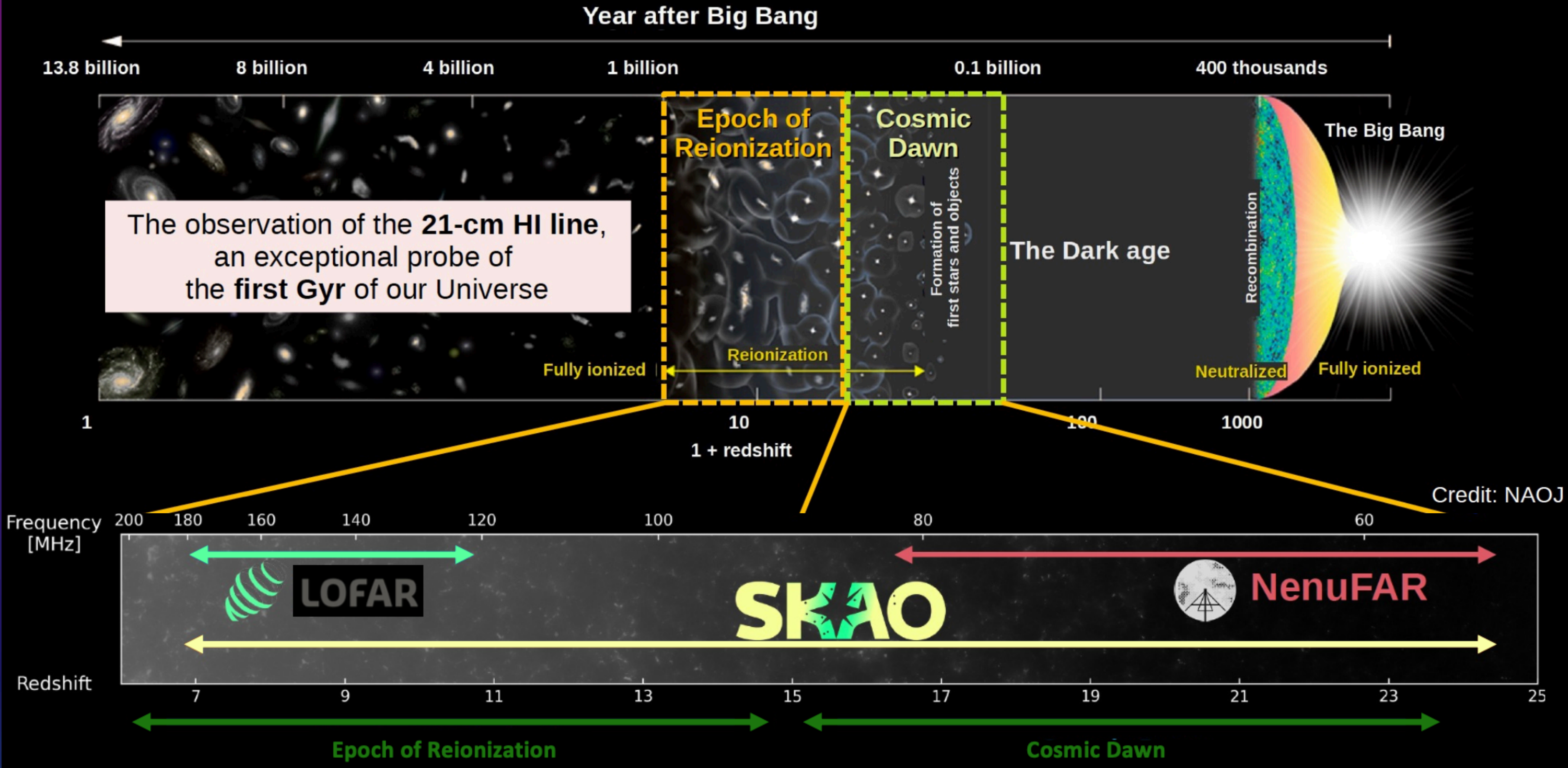
- Reconstruct **DM density fields.**
- **Map 3D Large Scale Structures** of the Universe.
- Complementary measurement to optical surveys to **constrain cosmological parameters.**

What is Intensity Mapping?

- Measurement of redshift and intensity of HI over the **whole sky.**
- Treats HI signal as a **diffuse background.**
- Large cosmological volume.
- Less costly, less time consuming.
- High spectral resolution $\Longrightarrow$ **high redshift resolution.**
- Individual galaxy detection not needed for LSS study.

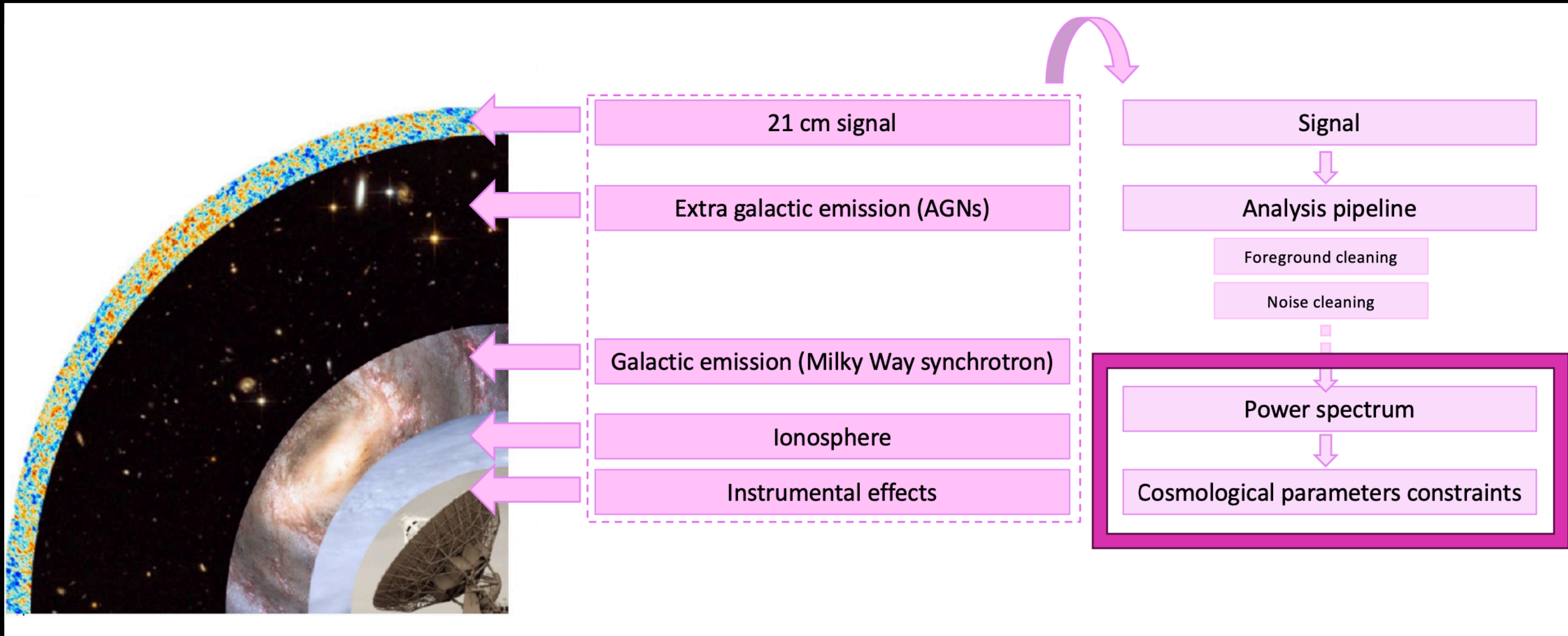


State of the Art for HI IM



Measurements scheme

Credits: :Marta Spinelli



Limitations of the power spectrum

- **Gaussian Assumption**:
  - The power spectrum is most effective for Gaussian random fields.
- **Non-Gaussian Features**:
  - The universe exhibits non-Gaussian features due to non-linear growth of structures and primordial non-Gaussianities.
- **Loss of Information**:
  - Higher order interactions and complex structures are not captured by the power spectrum.
  - Important information about the morphology and connectivity of cosmic structures is lost.

Make use of Higher order statistics which are sensitive to the non-Gausianities.

Lognormal simualtions

CAMB + GLASS

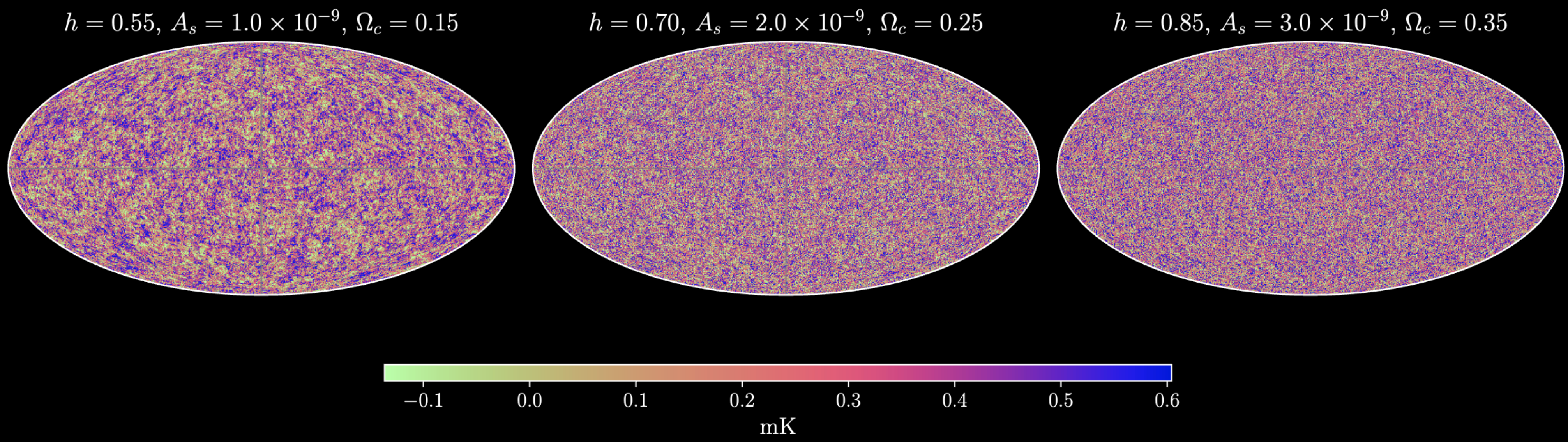
Input parameters: Ω_c, A_s and h (Ω_b fixed)

Different realizations (seed)

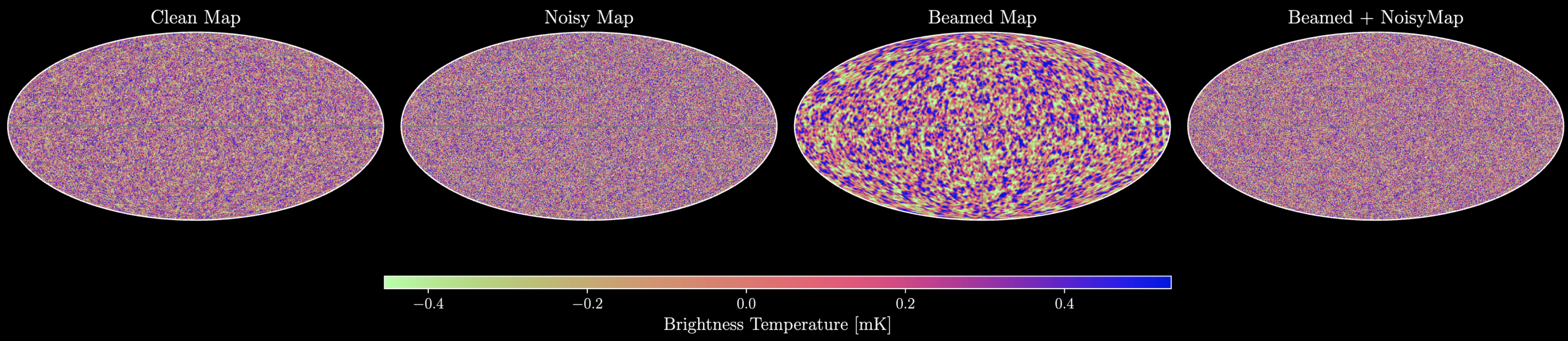
Option of adding beam and/or noise (MeerKLASS-like)

Output: Full sky HI brightness Temperature map of resolution 1024

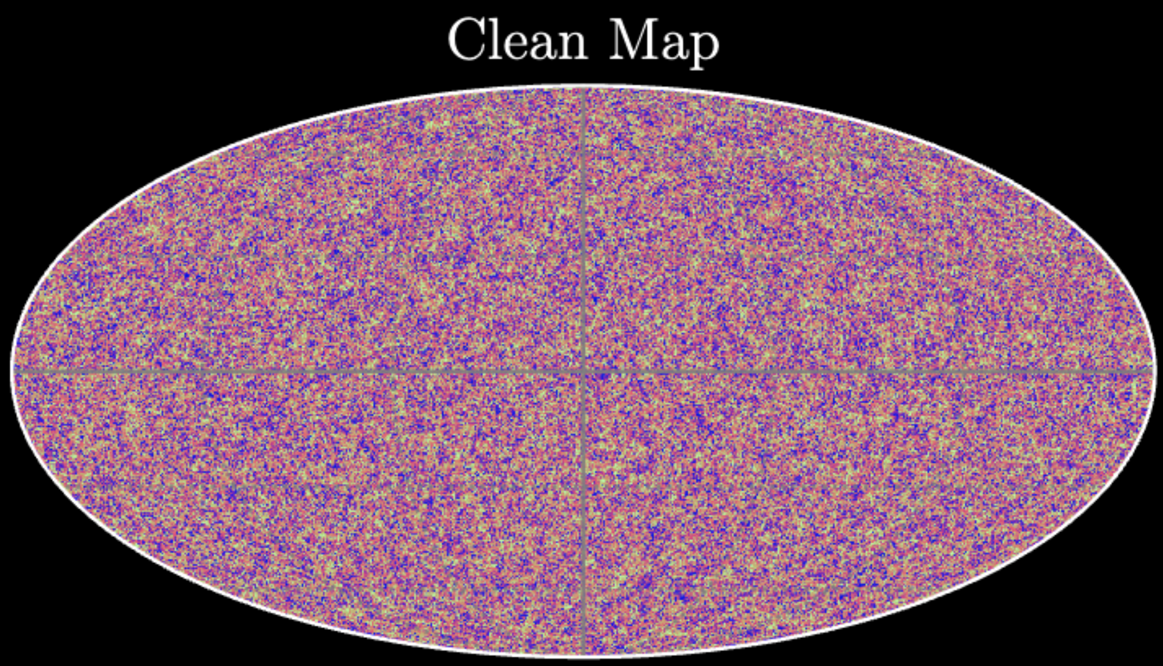
Maps across cosmologies

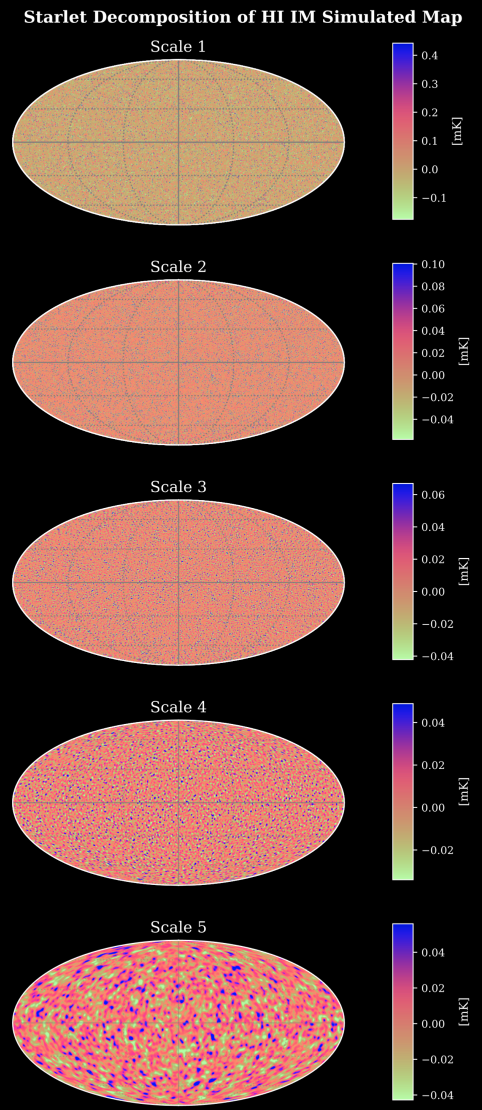


Instrumental effects



Starlet transform

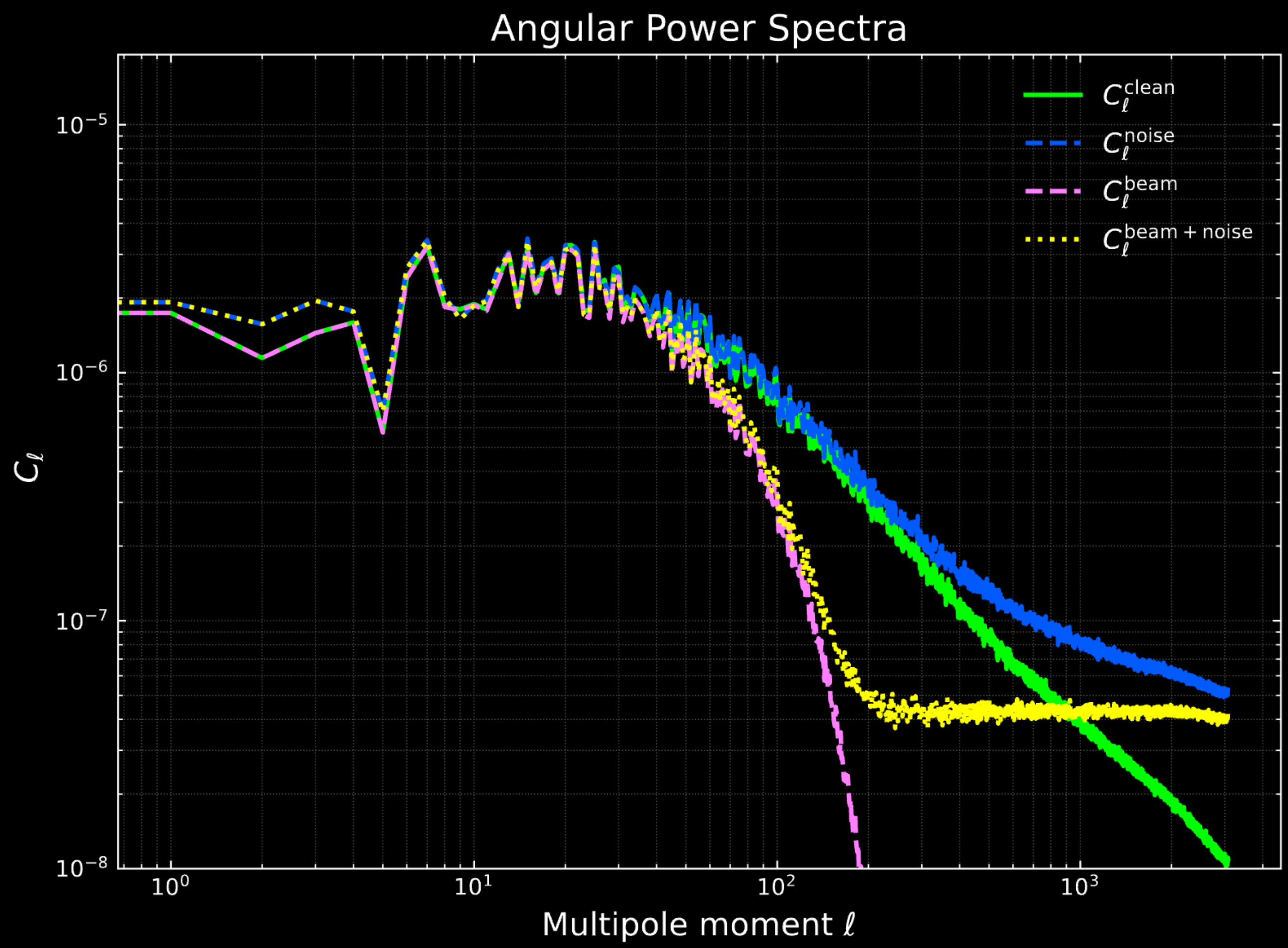




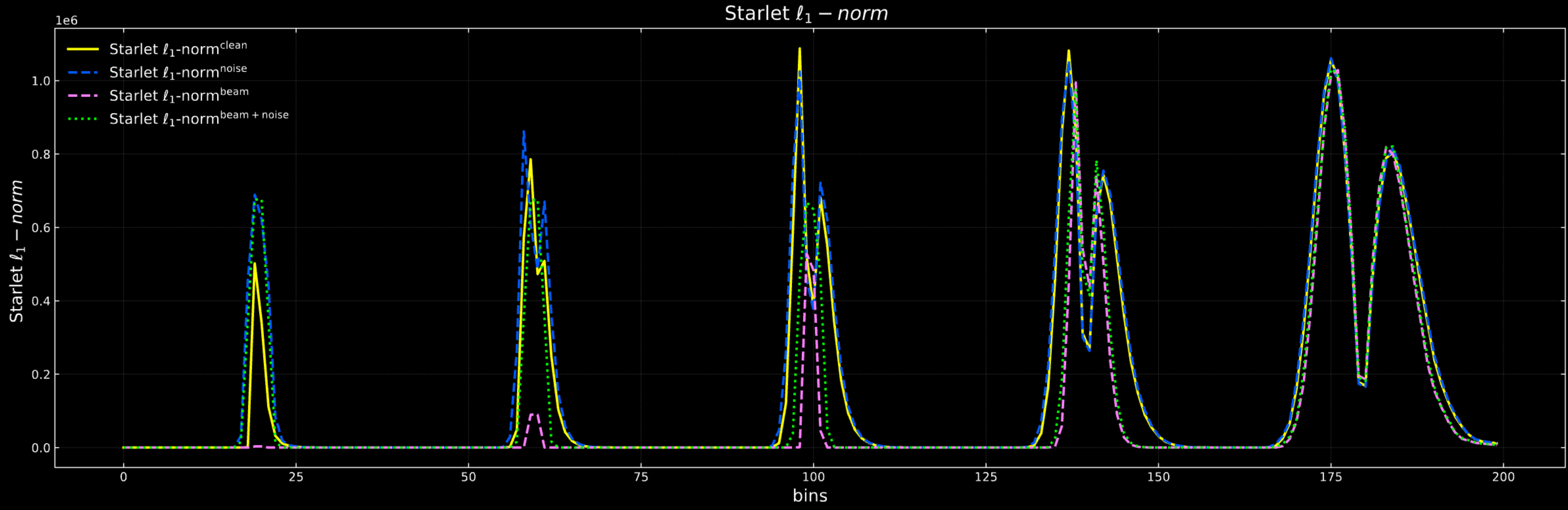
Decomposes the map into a set of wavelet coefficients + coarse residual map

$$M(\hat{n}) = \sum_{j=1}^{J} w_j(\hat{n}) + c_J(\hat{n}).$$

Summary statistics



Summary statistics

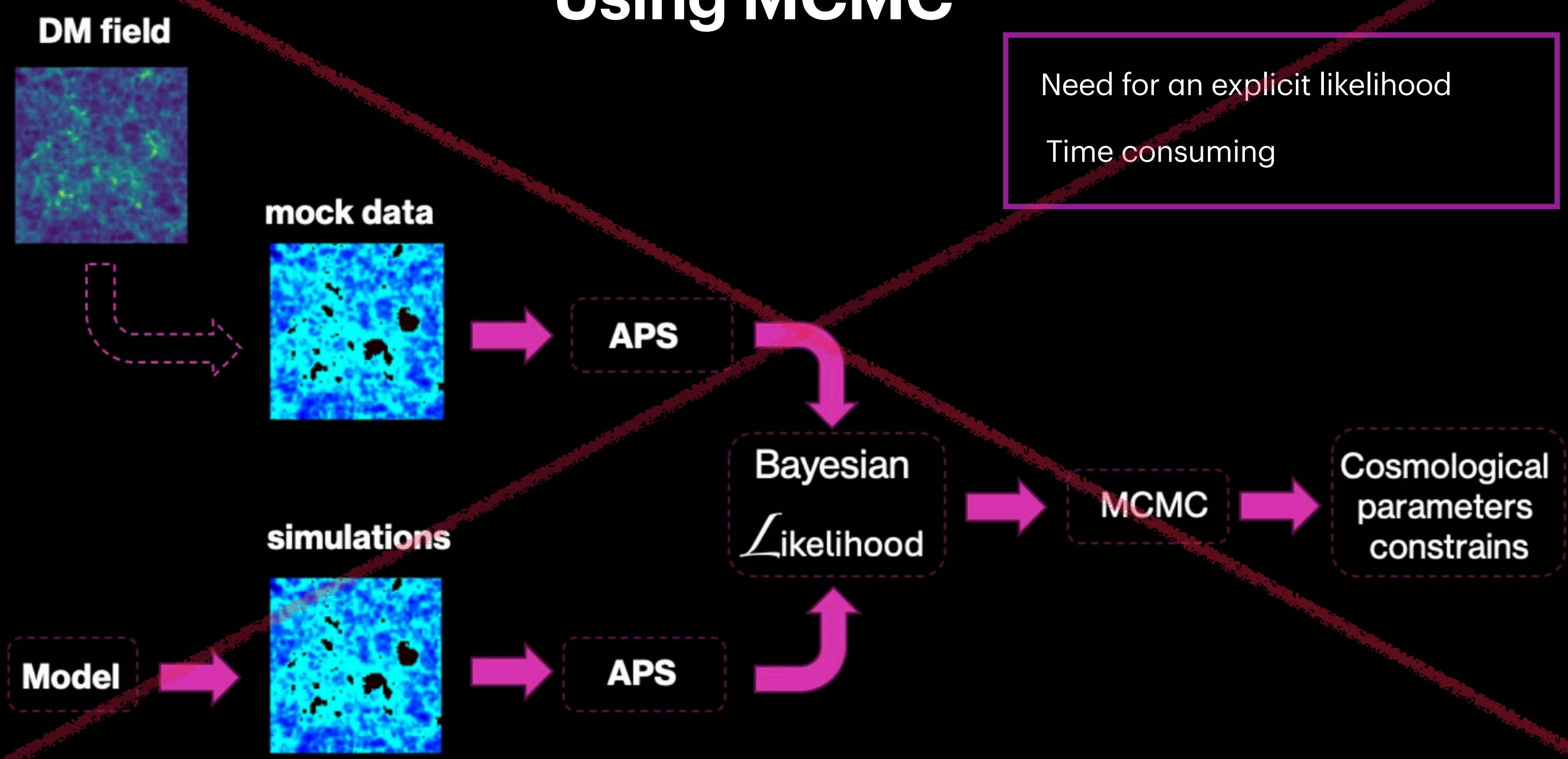


Bayesian cosmological parameter inference

Using MCMC

Need for an explicit likelihood

Time consuming



Simulation Based Inference (SBI)

Splitting of data set 10 000

Training set: 7 000

Validation set: 2 000

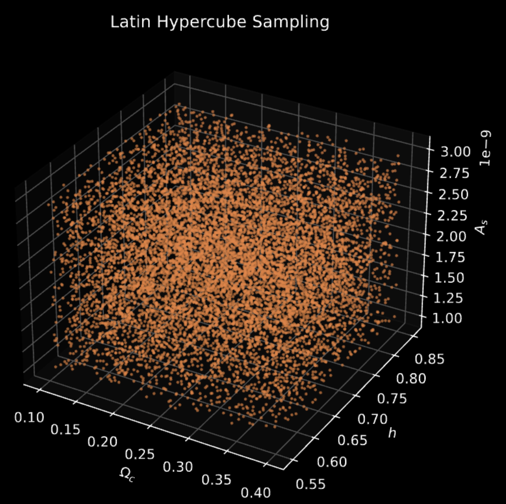
Test set: 1 000

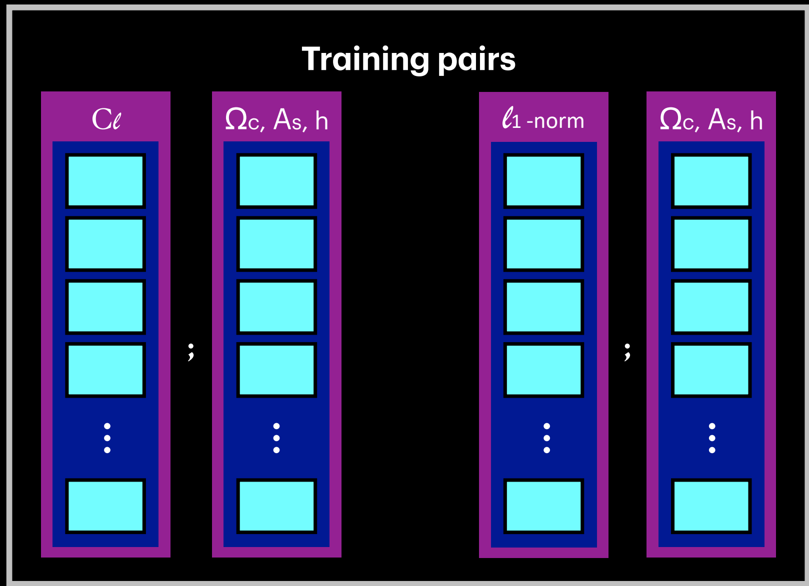
JAXILI

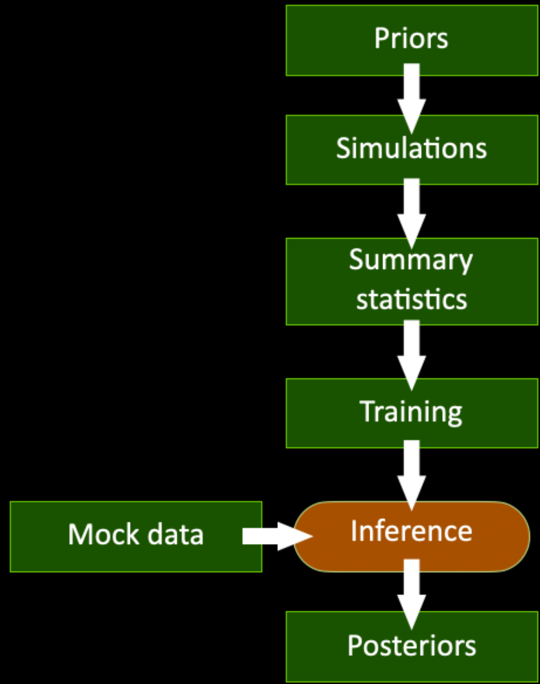
No Need for an explicit likelihood

Training set computed **once**

Training time - few minutes (with JAXILI)

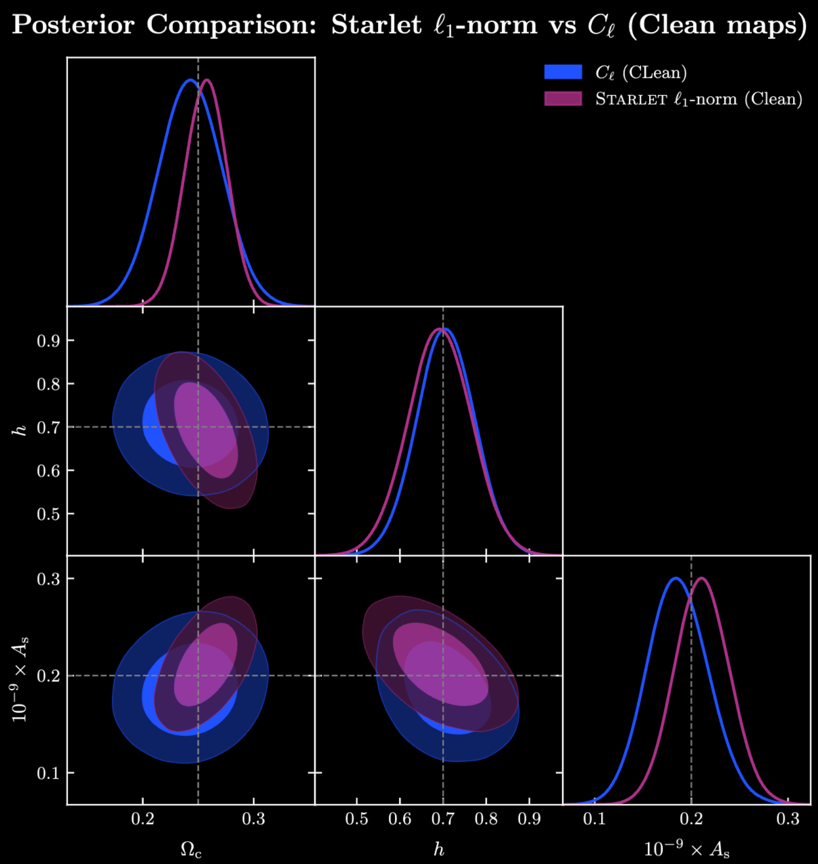


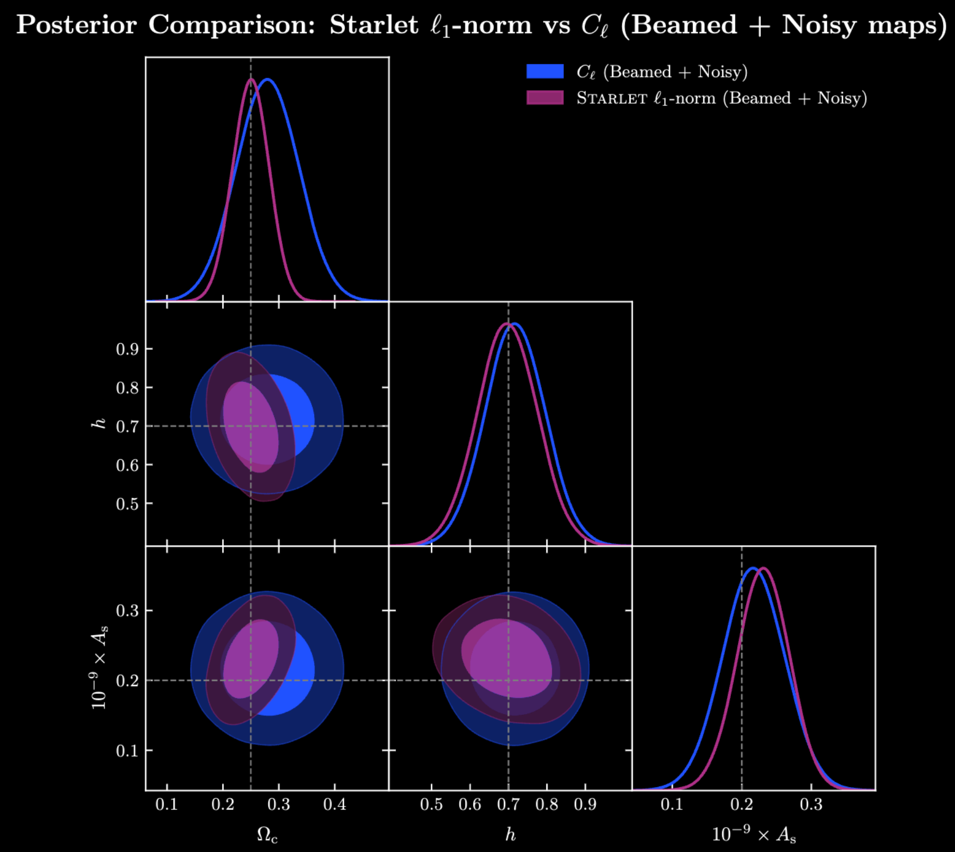




Results

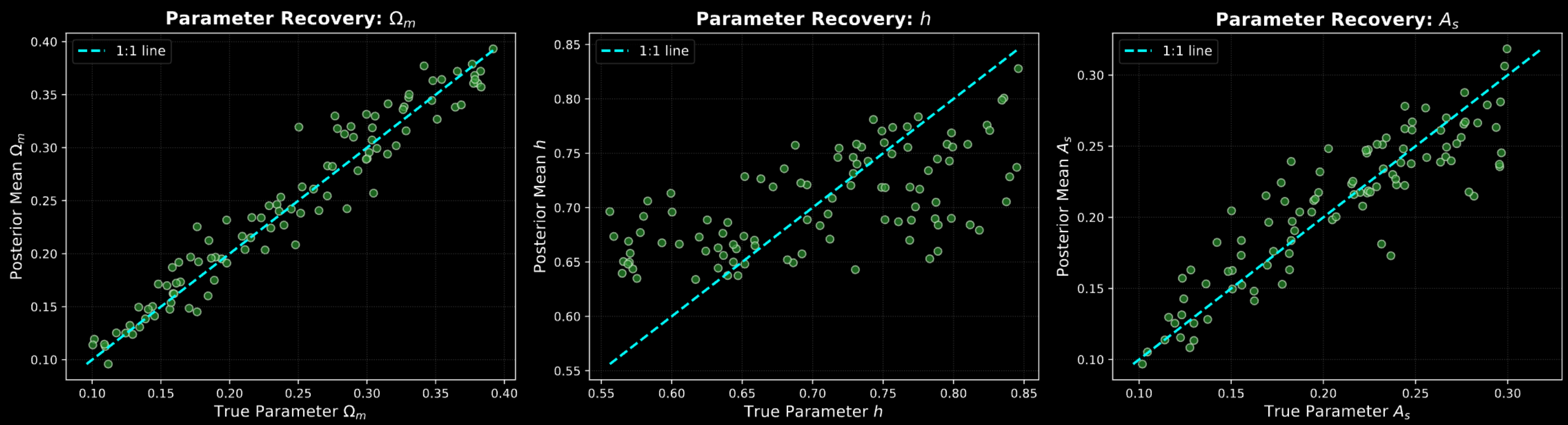
Cosmological parameter constrains (Ωm, h, As).



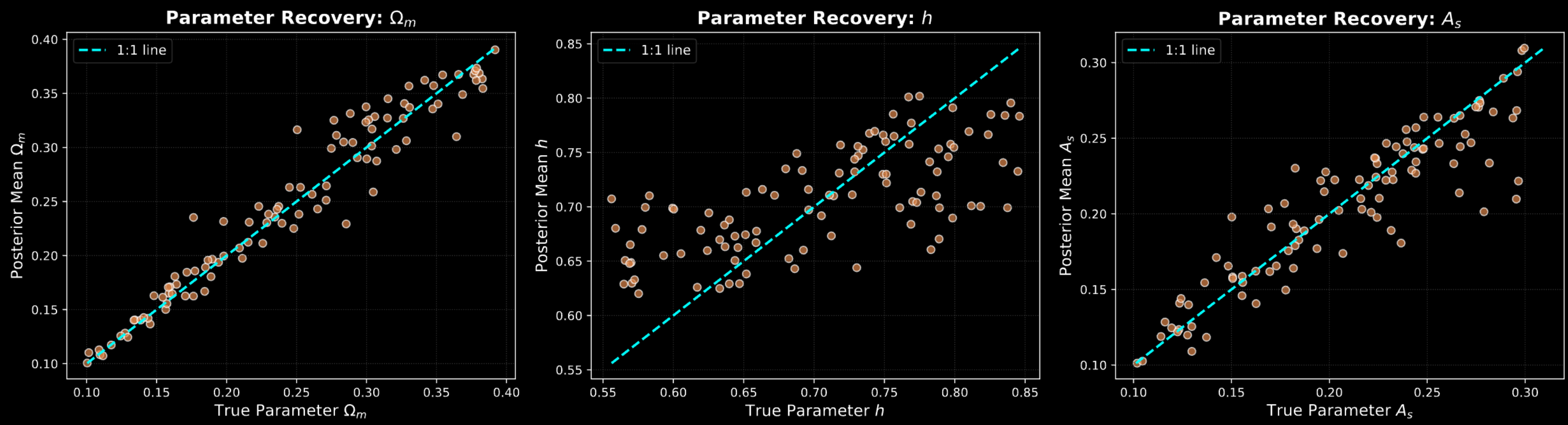


Clean maps

Starlet l1 norm

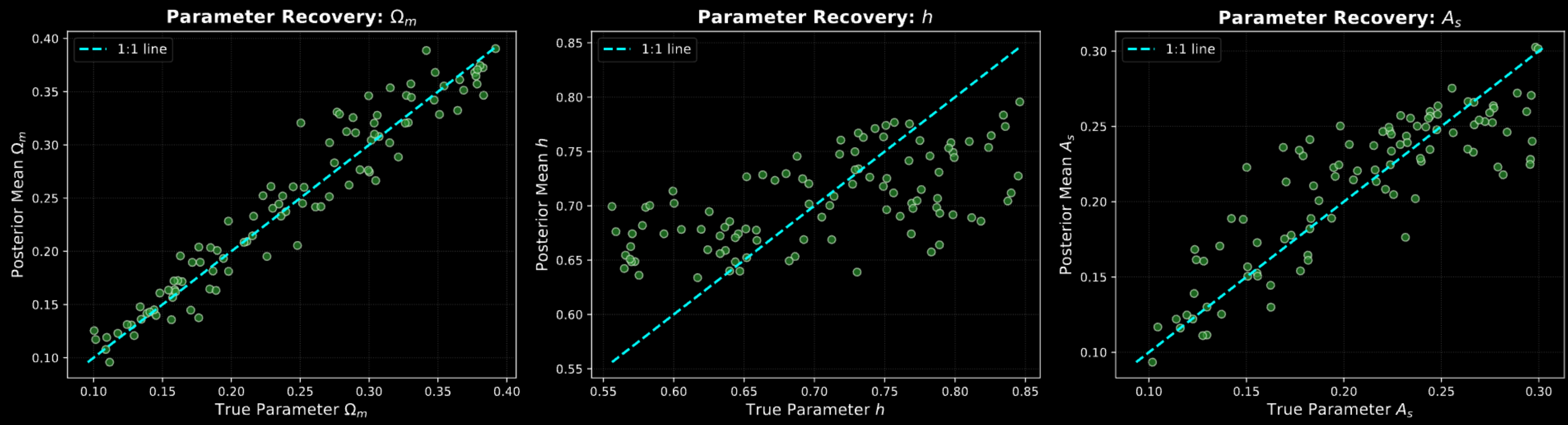


Cl

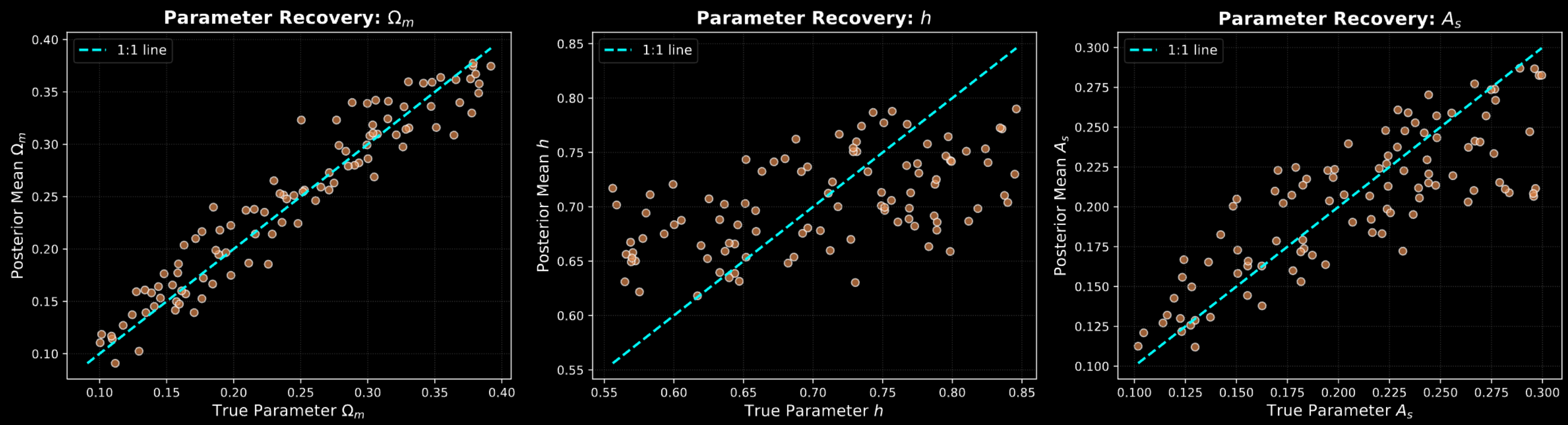


Beamed + noisy maps

Starlet l1 norm



Cl



Next steps

N-body simulations instead of lognormal maps

Addition of **other systematics** (e.g. satellites trails) for more **realistic case**