Optimising bayesian inference in cosmology
with Marginal Neural Ratio Estimation

Guillermo Franco Abellán

GRAPPA GRavitation AstroParticle Physics Amsterdam

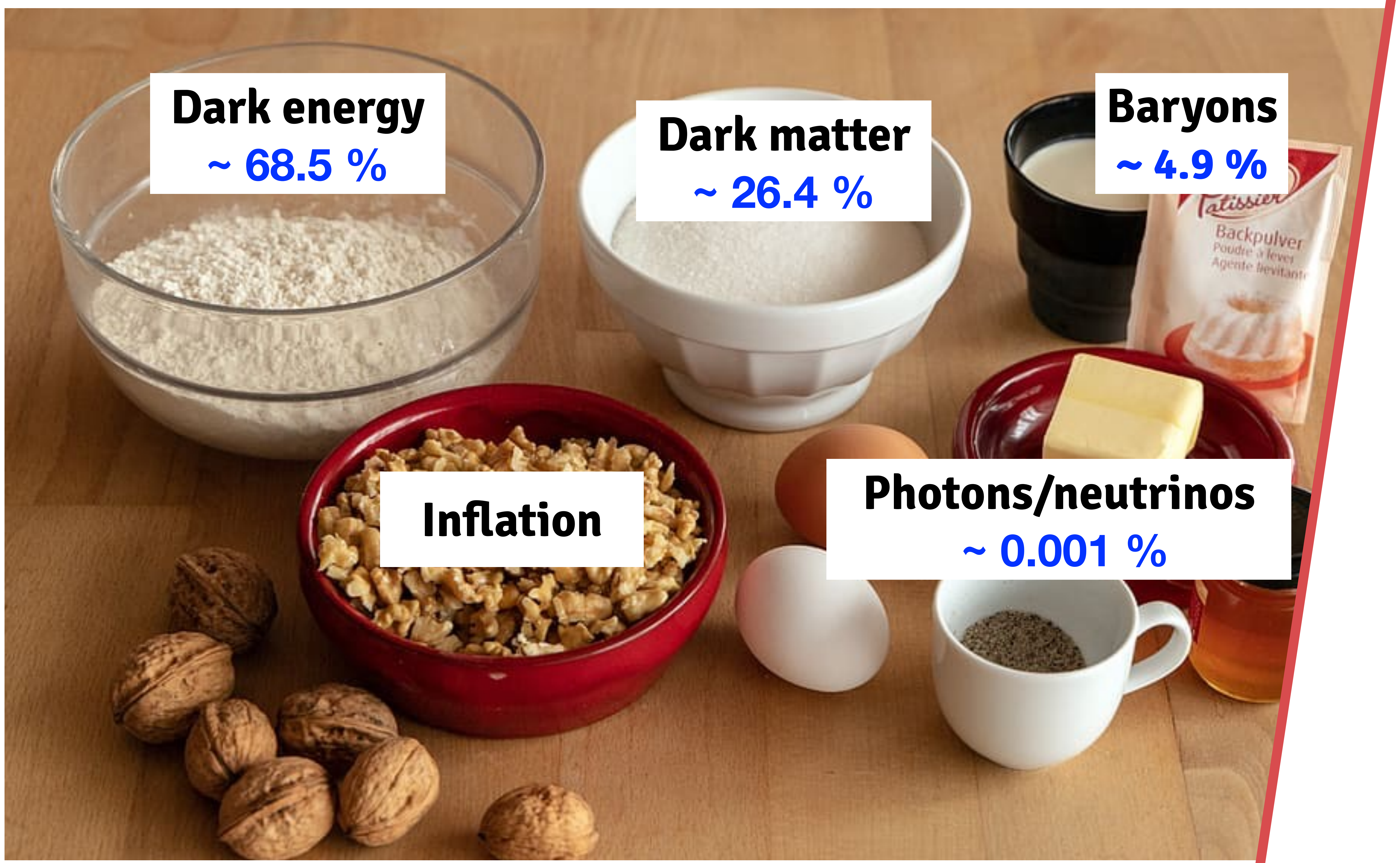
Euclid consortium

Based on arXiv:2401.XXXX

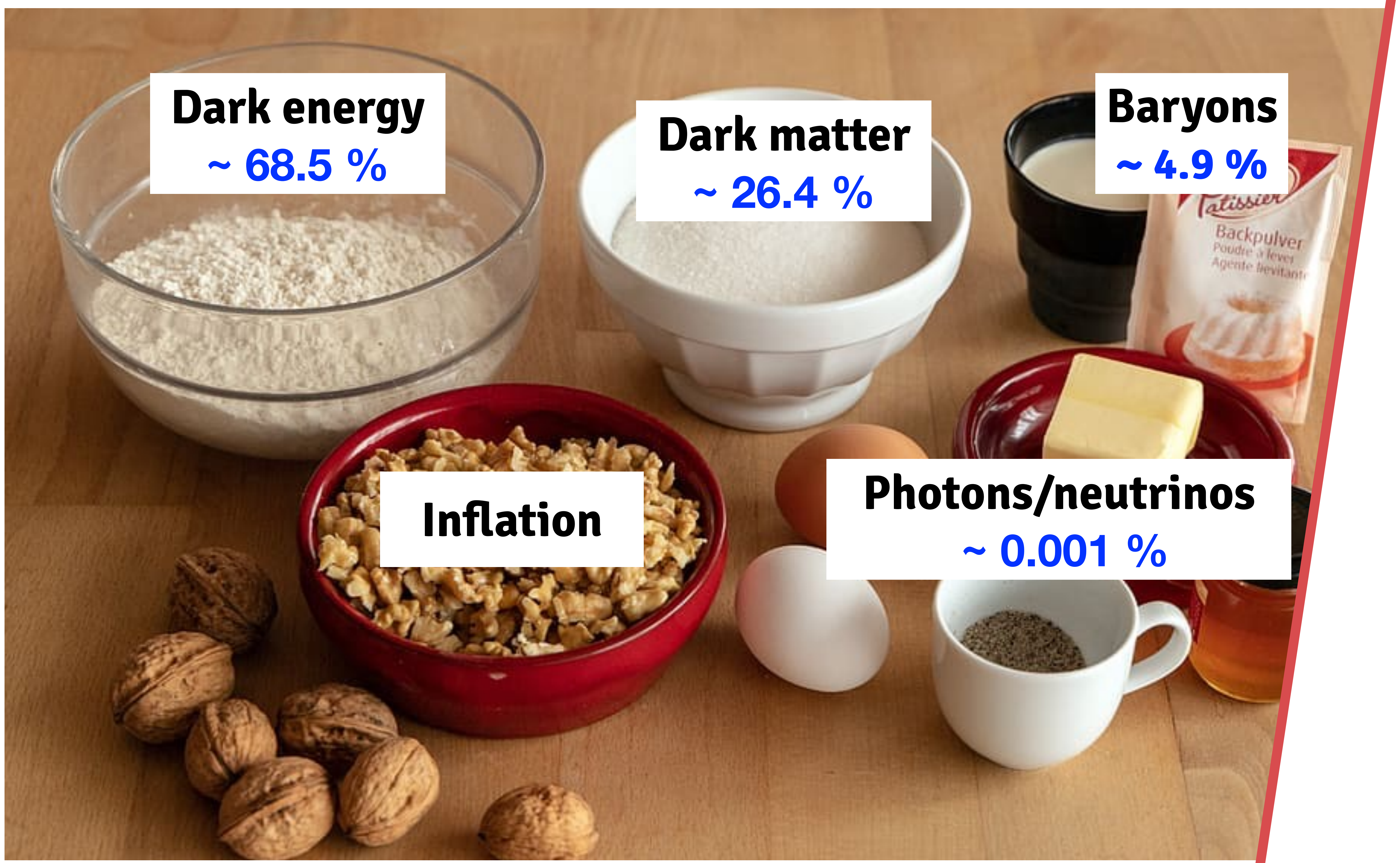
with Guadalupe C. Herrera,
Matteo Martinelli,
Oleg Savchenko,
& Christoph Weniger

Centre Physique Théorique Marseille - 12/01/2024

Concordance ΛCDM model of cosmology:



Concordance ΛCDM model of cosmology:

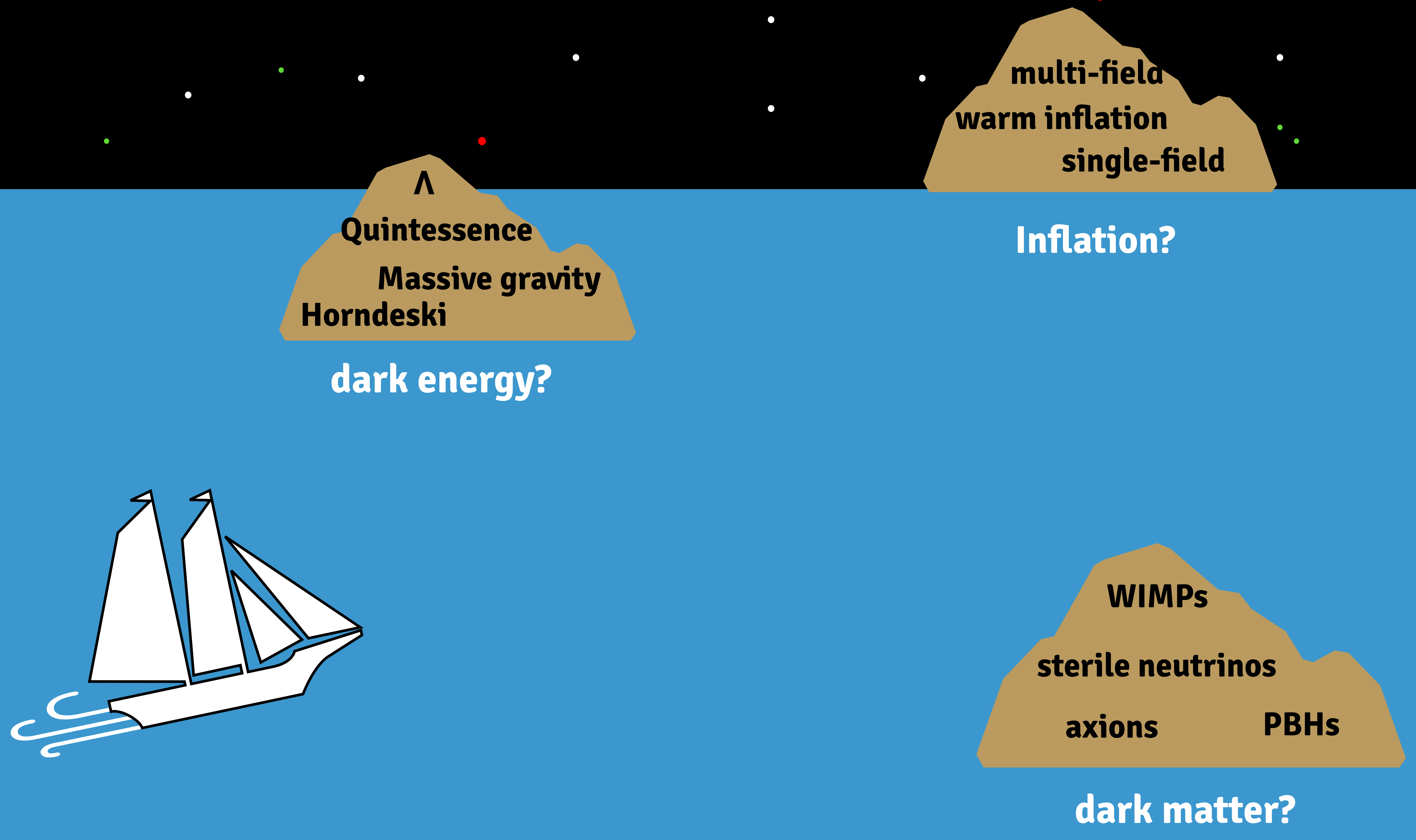


Only 6 free parameters:

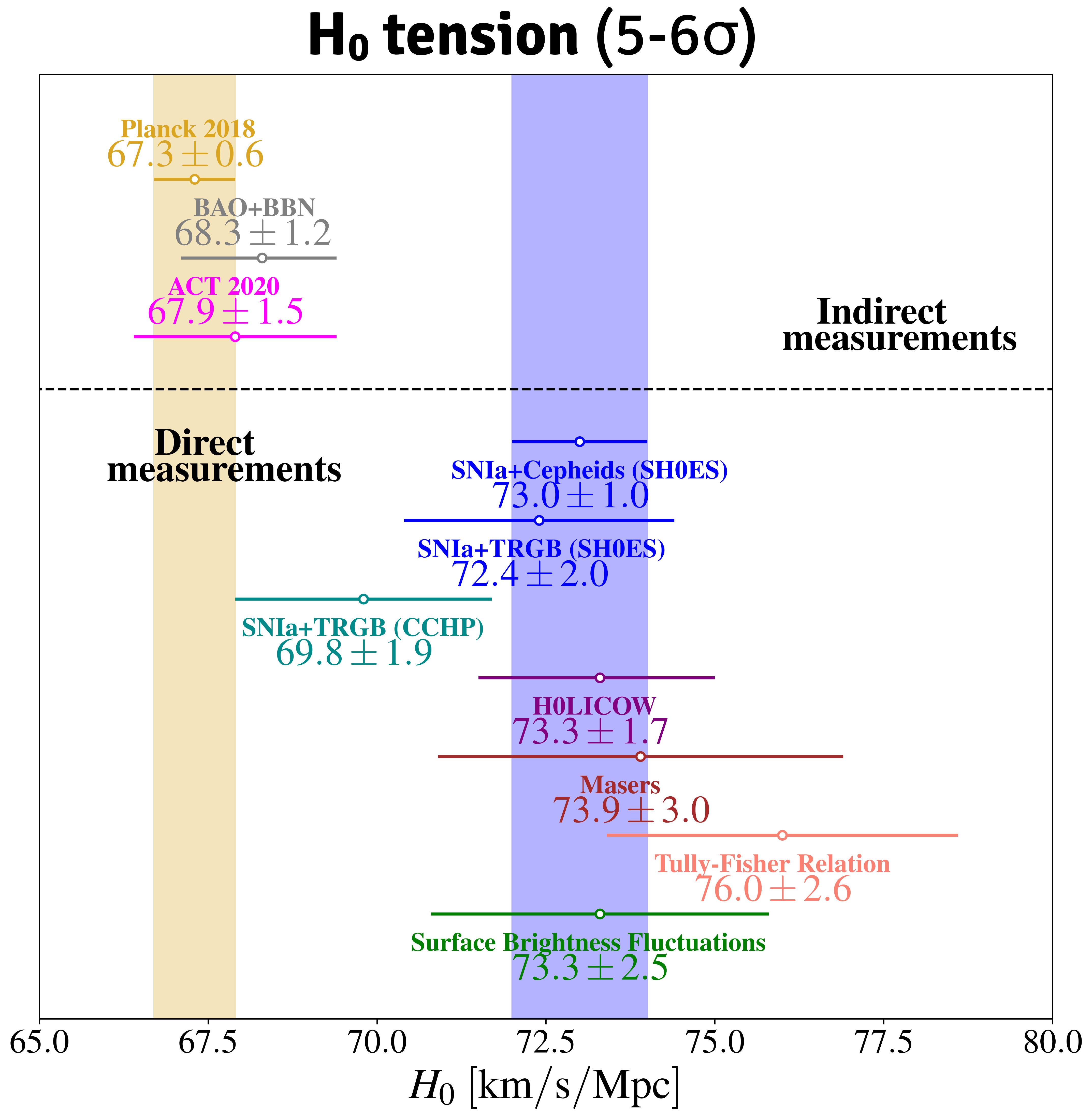
ω_c ω_b H_0

A_s n_s τ_{reio}

However, the nature of the **dark sector** remains **unknown**

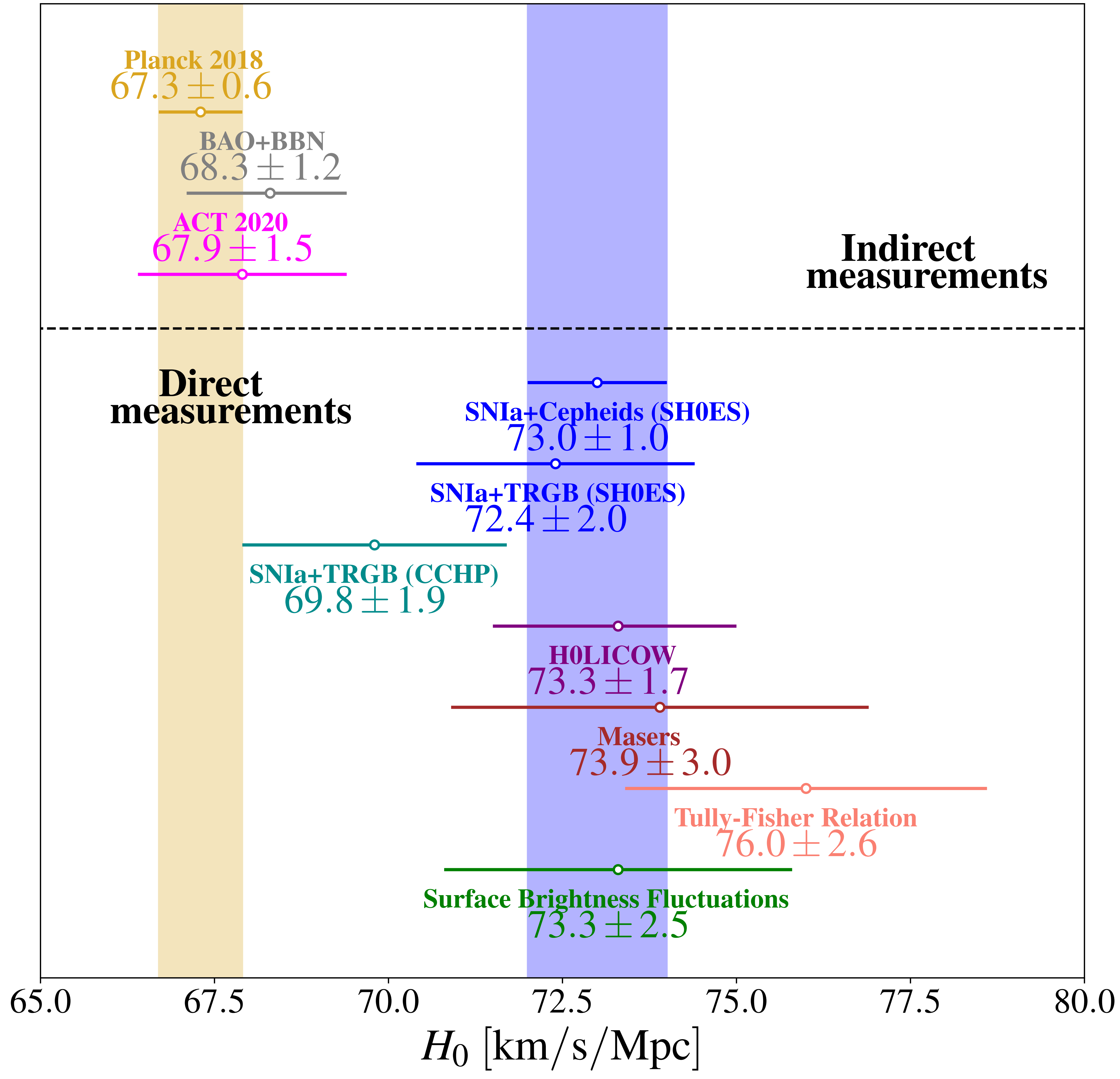


In addition, discrepancies have emerged

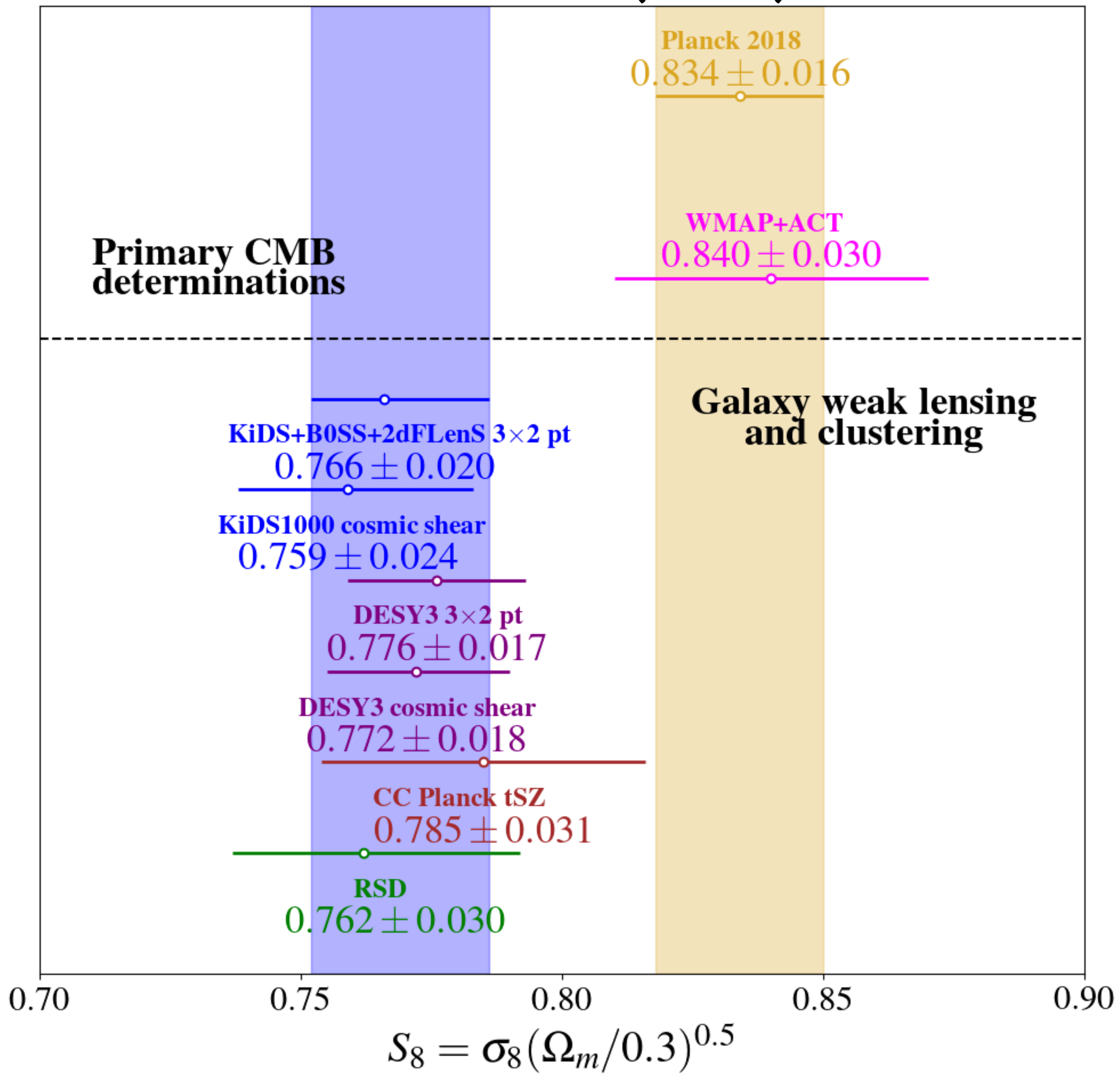


In addition, discrepancies have emerged

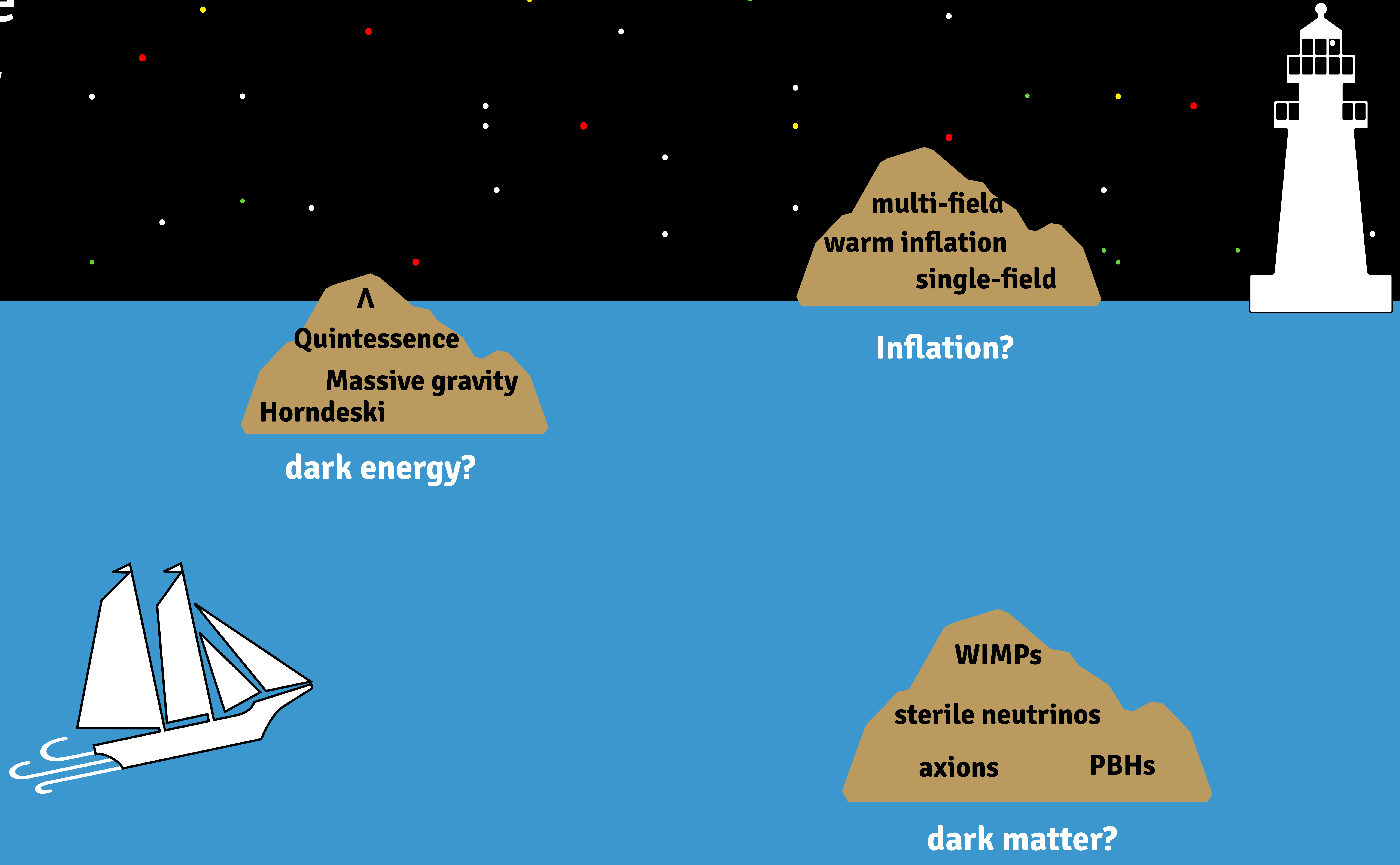
H_0 tension (5-6σ)



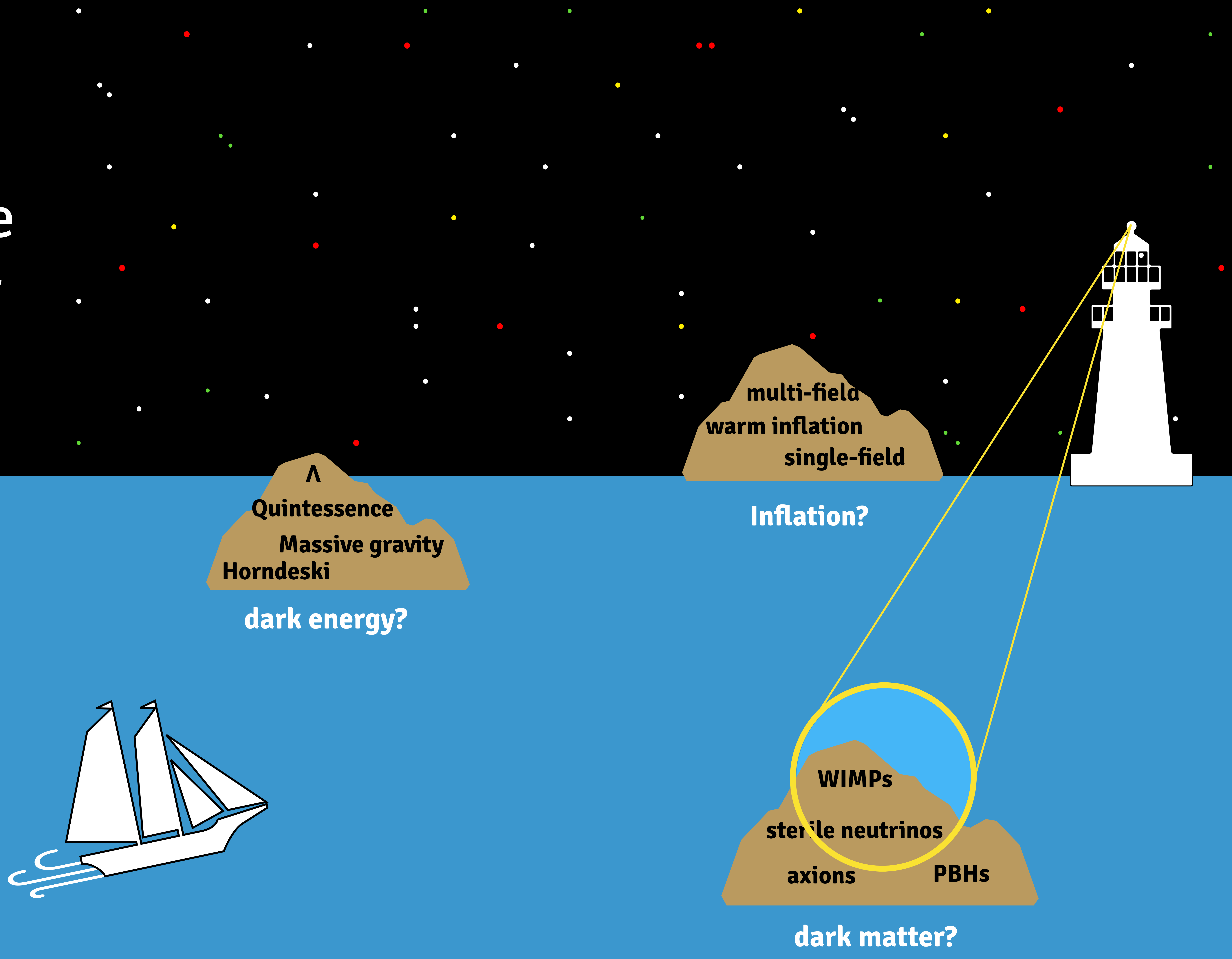
S_8 tension (2-3σ)



Cosmic tensions can **shed some light** on the mysterious dark sector



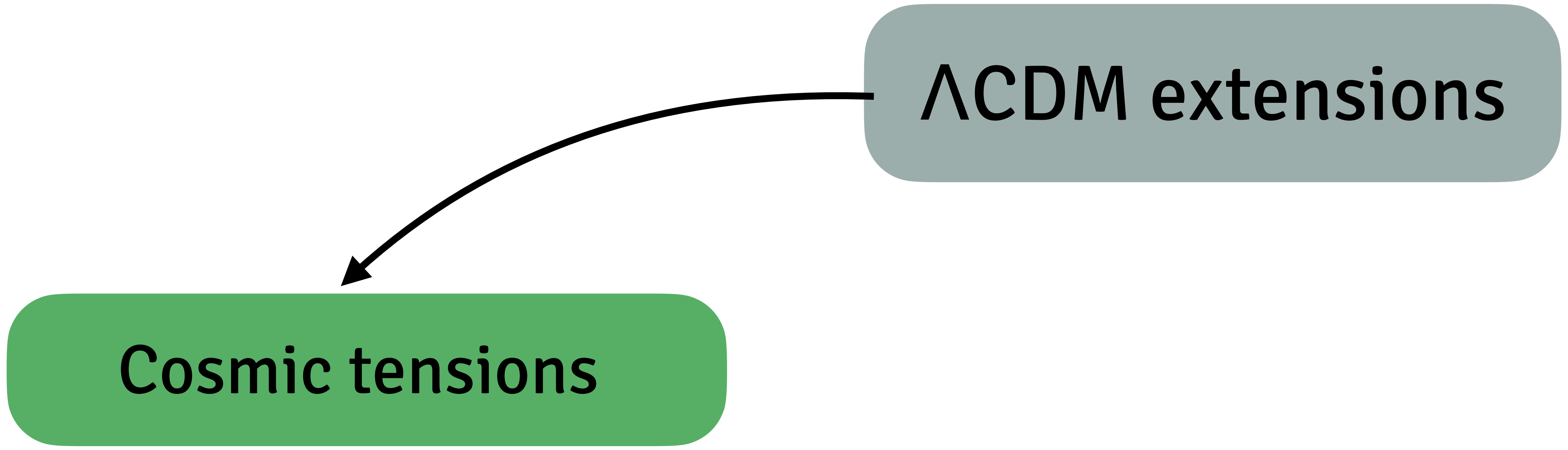
Cosmic tensions can **shed some light** on the mysterious dark sector



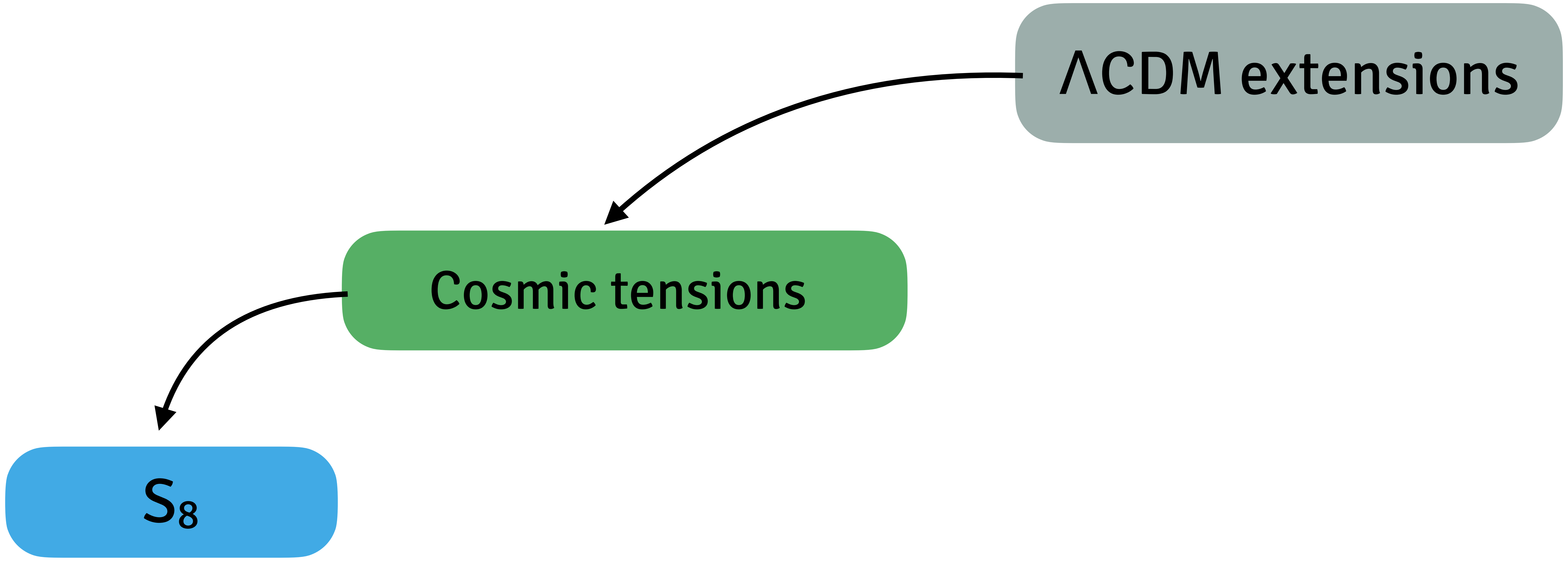
My work so far...

ΛCDM extensions

My work so far...



My work so far...



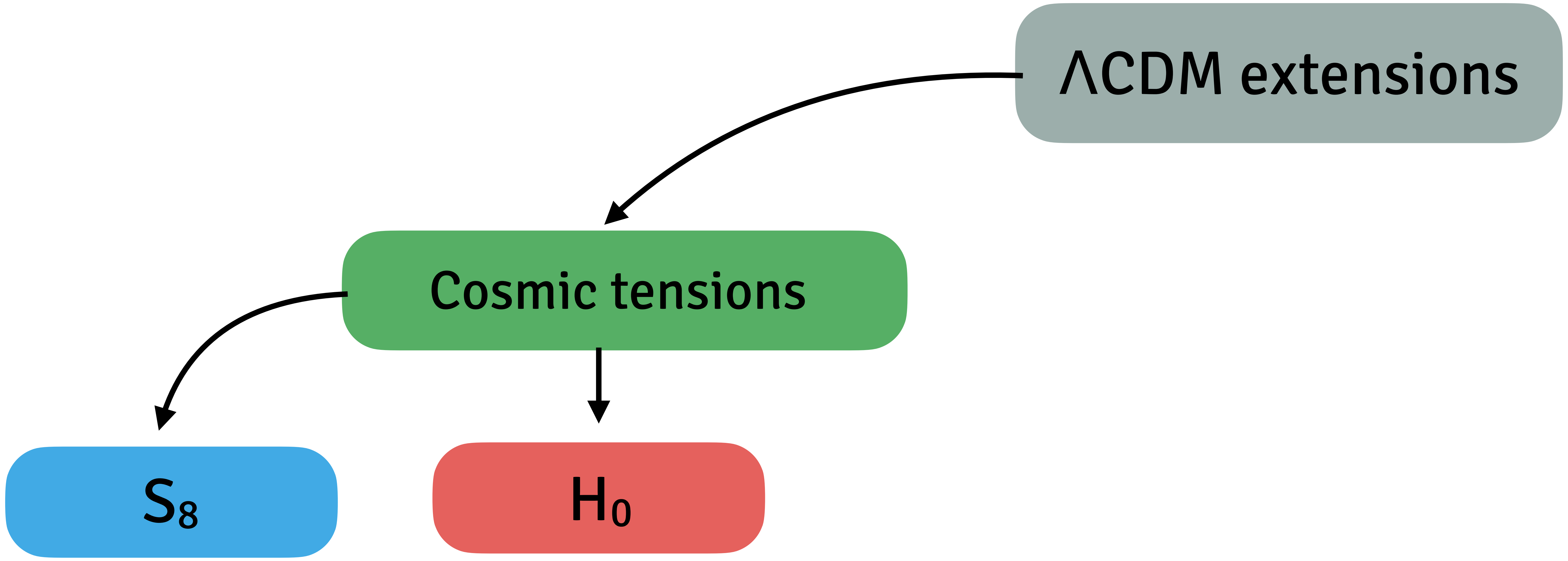
Decaying DM

[GFA, Murgia+ 20]

[GFA, Murgia+ 21]

[Simon, GFA+ 22]

My work so far...



Decaying DM

[GFA, Murgia+ 20]
[GFA, Murgia+ 21]
[Simon, GFA+ 22]

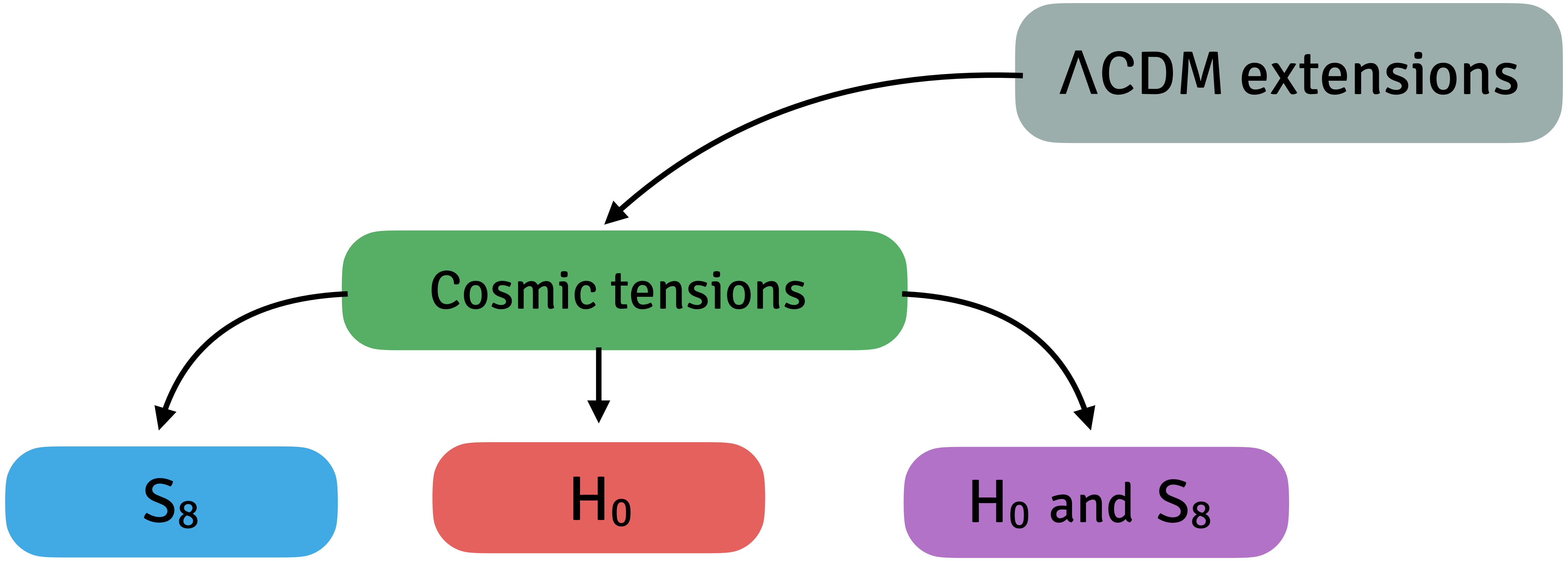
Early Modified Gravity

[GFA, Braglia+ 23]

Early Dark Energy

[Murgia, GFA+ 20]

My work so far...



Decaying DM

[GFA, Murgia+ 20]

[GFA, Murgia+ 21]

[Simon, GFA+ 22]

Early Modified Gravity

[GFA, Braglia+ 23]

Early Dark Energy

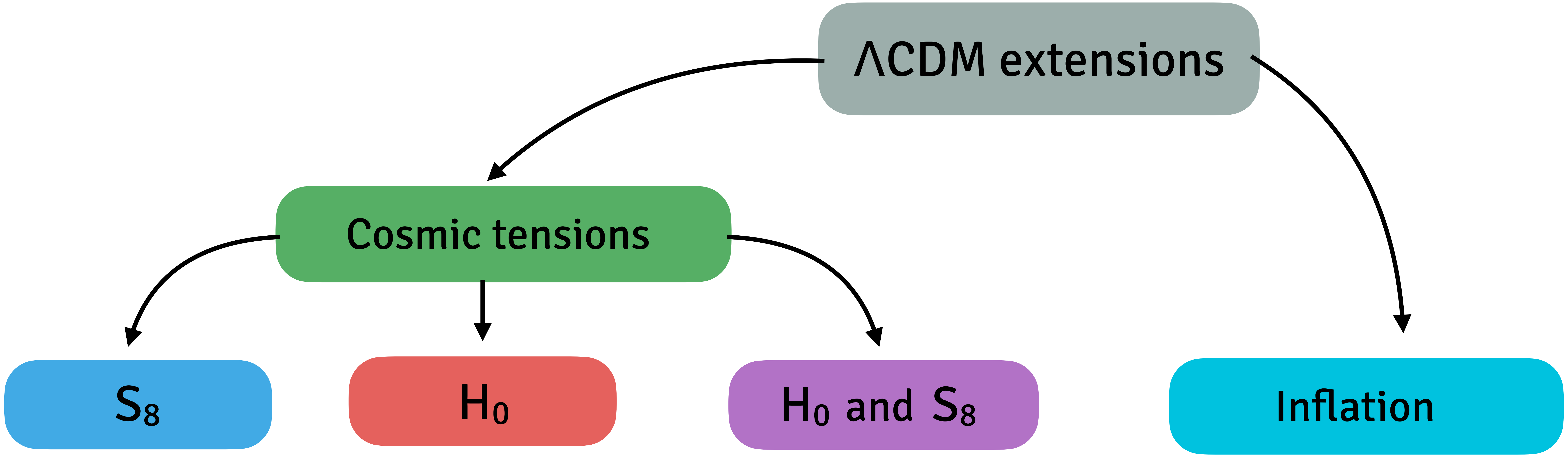
[Murgia, GFA+ 20]

Interacting Stepped DR

[Schöneberg & GFA 22]

[Schöneberg, GFA+ 23]

My work so far...



Decaying DM

[GFA, Murgia+ 20]

[GFA, Murgia+ 21]

[Simon, GFA+ 22]

Early Modified Gravity

[GFA, Braglia+ 23]

Early Dark Energy

[Murgia, GFA+ 20]

Interacting Stepped DR

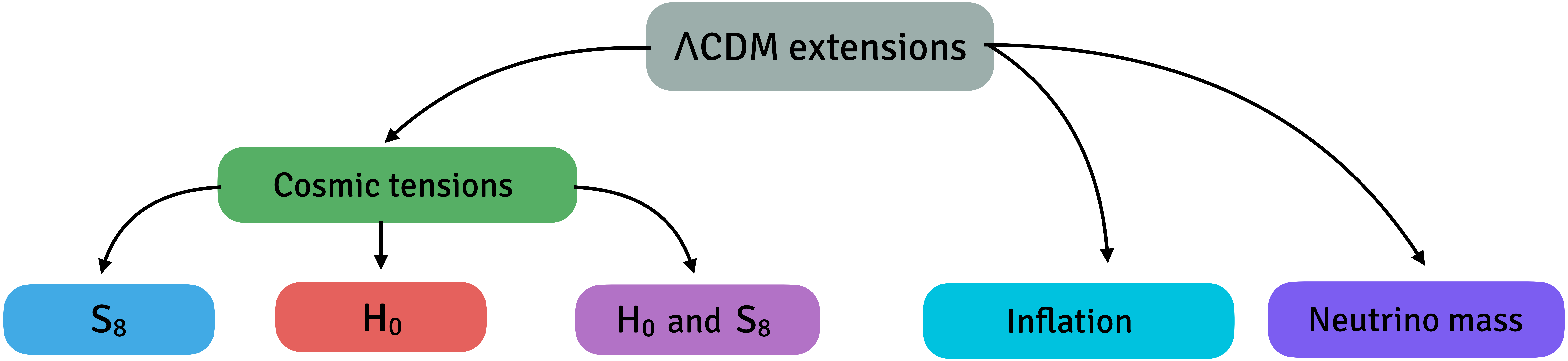
[Schöneberg & GFA 22]

[Schöneberg, GFA+ 23]

Ultracompact minihalos

[GFA & Facchinetti 23]

My work so far...



Decaying DM
[GFA, Murgia+ 20]
[GFA, Murgia+ 21]
[Simon, GFA+ 22]

Early Modified Gravity
[GFA, Braglia+ 23]

Early Dark Energy
[Murgia, GFA+ 20]

Interacting Stepped DR
[Schöneberg & GFA 22]
[Schöneberg, GFA+ 23]

Ultracompact minihalos
[GFA & Facchinetti 23]

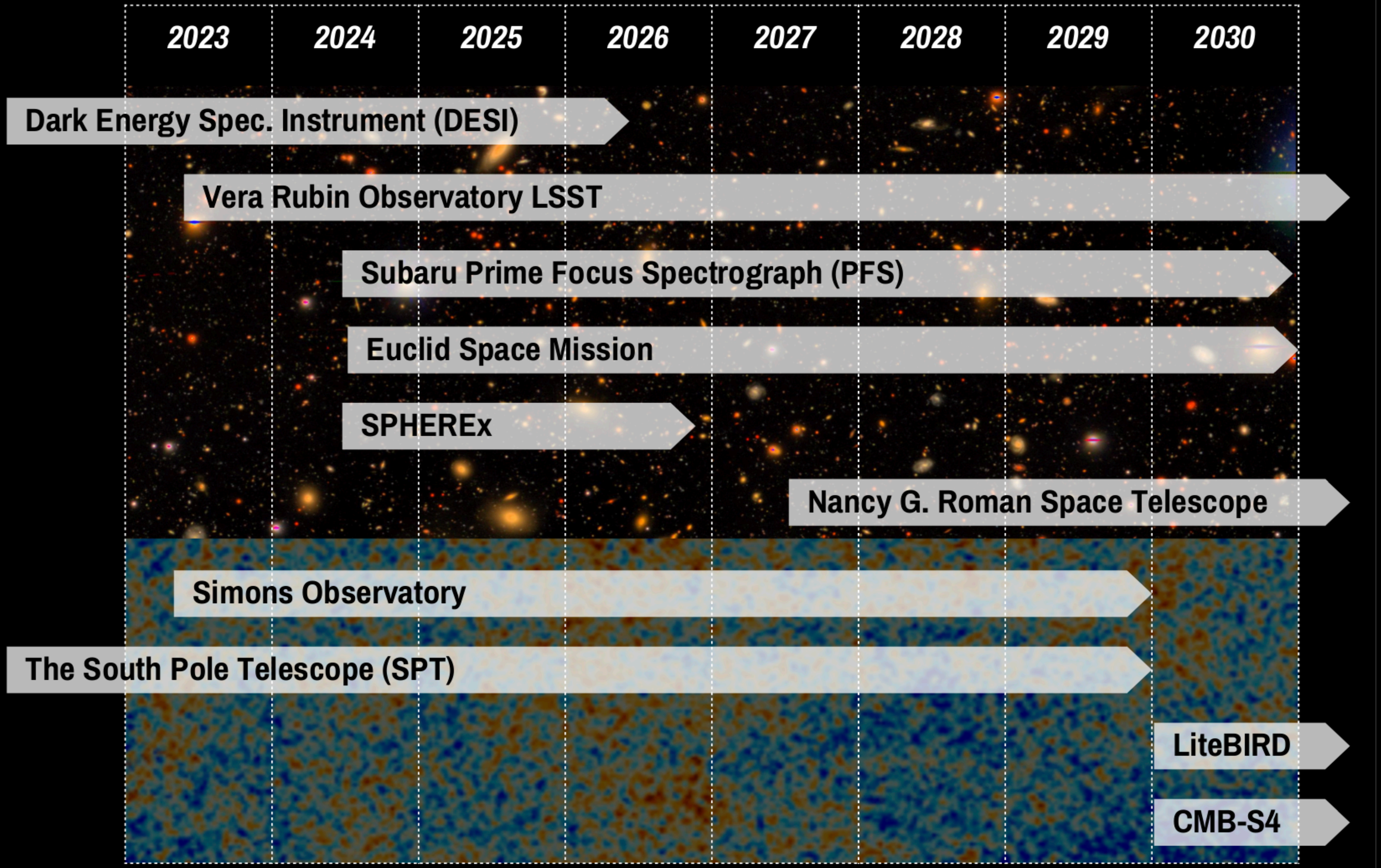
Decaying neutrinos
[GFA, Chacko+ 21]

Still no smoking-gun signature of new physics...

Still no smoking-gun signature of new physics...

...is there hope to establish a new concordance model?

Next-generation cosmological data is becoming available



Analysing these high-quality data will be extremely challenging with traditional methods

Outline

I. Why we need to go **beyond MCMC**

II. Our new approach: **Marginal Neural Ratio Estimation**

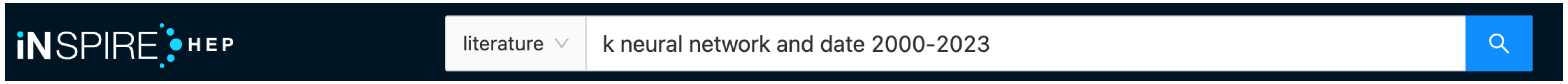
III. Applying MNRE to **Euclid** observables

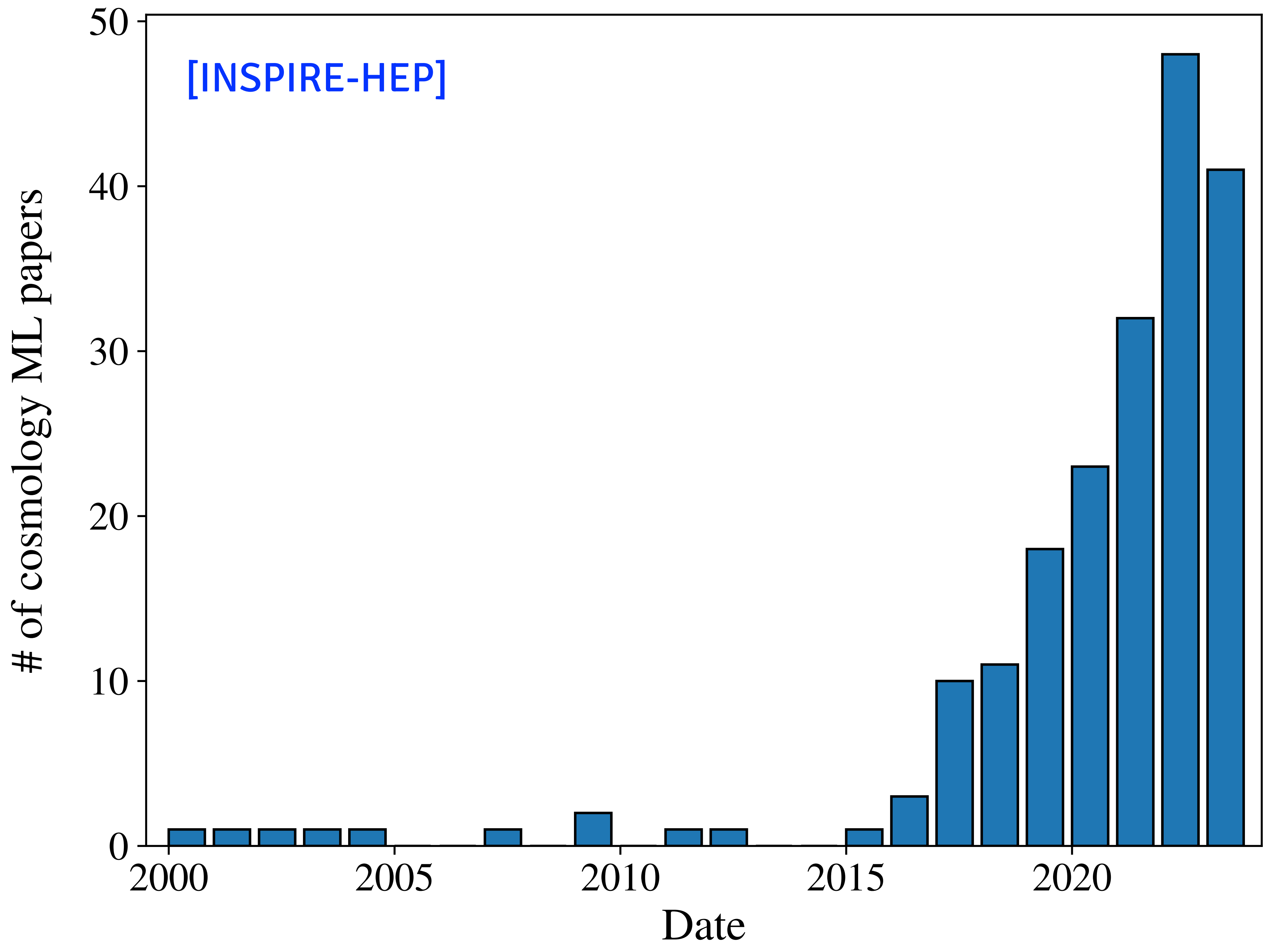
Outline

I. Why we need to go **beyond MCMC** ✓

II. Our new approach: **Marginal Neural Ratio Estimation**

III. Applying MNRE to **Euclid** observables





Machine learning is having a **strong impact** in cosmology

Two main approaches in ML

Two main approaches in ML

Emulators

to achieve **ultra-fast** evaluations of cosmological observables

Two main approaches in ML

Emulators

to achieve **ultra-fast** evaluations of cosmological observables

New statistical methods

to improve the sampling in **high-dimensional** parameter spaces

Two main approaches in ML

Emulators

to achieve **ultra-fast** evaluations of cosmological observables

New statistical methods

to improve the sampling in **high-dimensional** parameter spaces

This talk

Bayesian inference

Likelihood

Posterior $\quad$ Prior

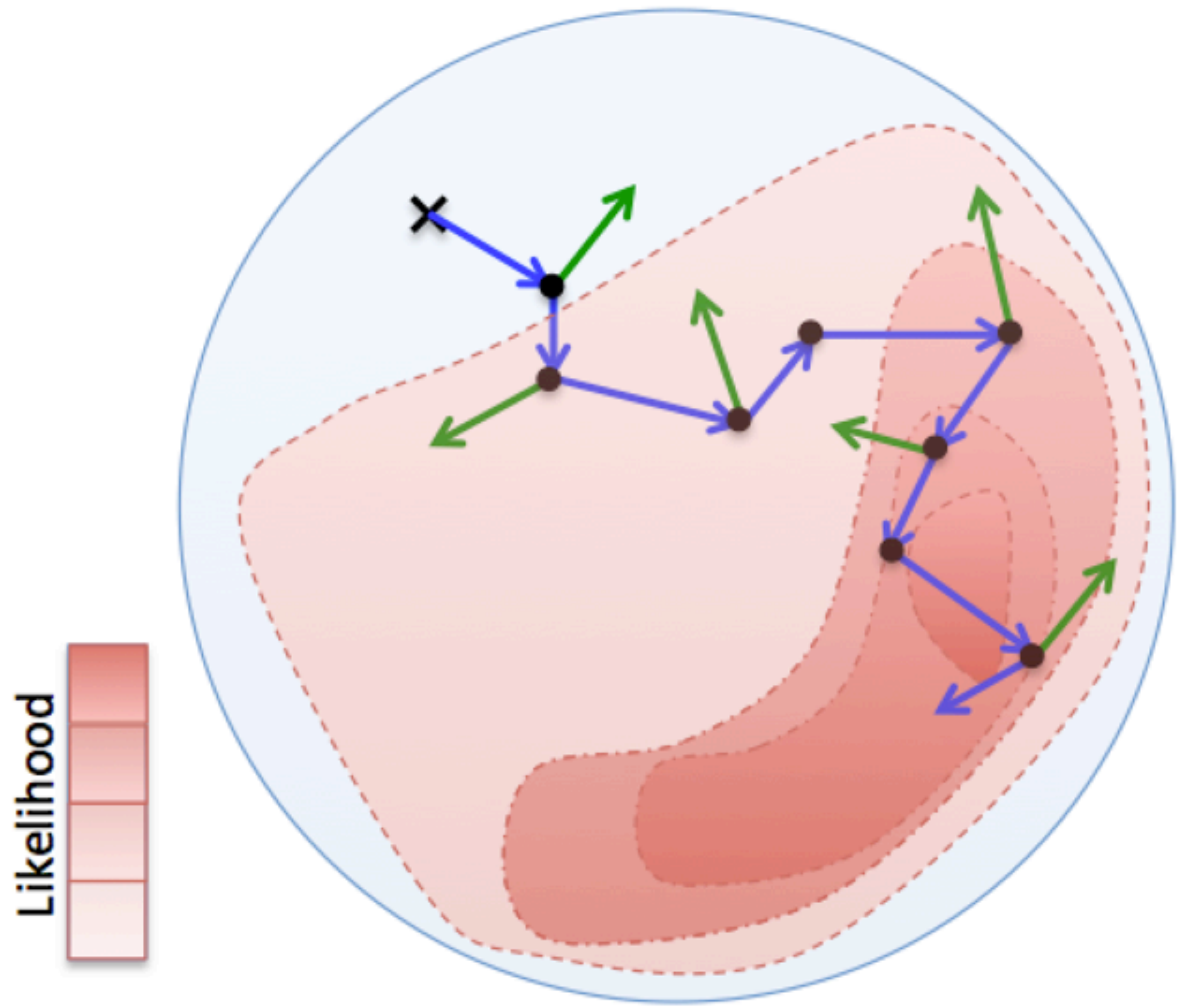
$$p(\boldsymbol{\theta} \mid \mathbf{x}) = \frac{p(\mathbf{x} \mid \boldsymbol{\theta})\, p(\boldsymbol{\theta})}{p(\mathbf{x})}$$

Evidence

$\mathbf{x}$: Data

$\boldsymbol{\theta}$: Parameters

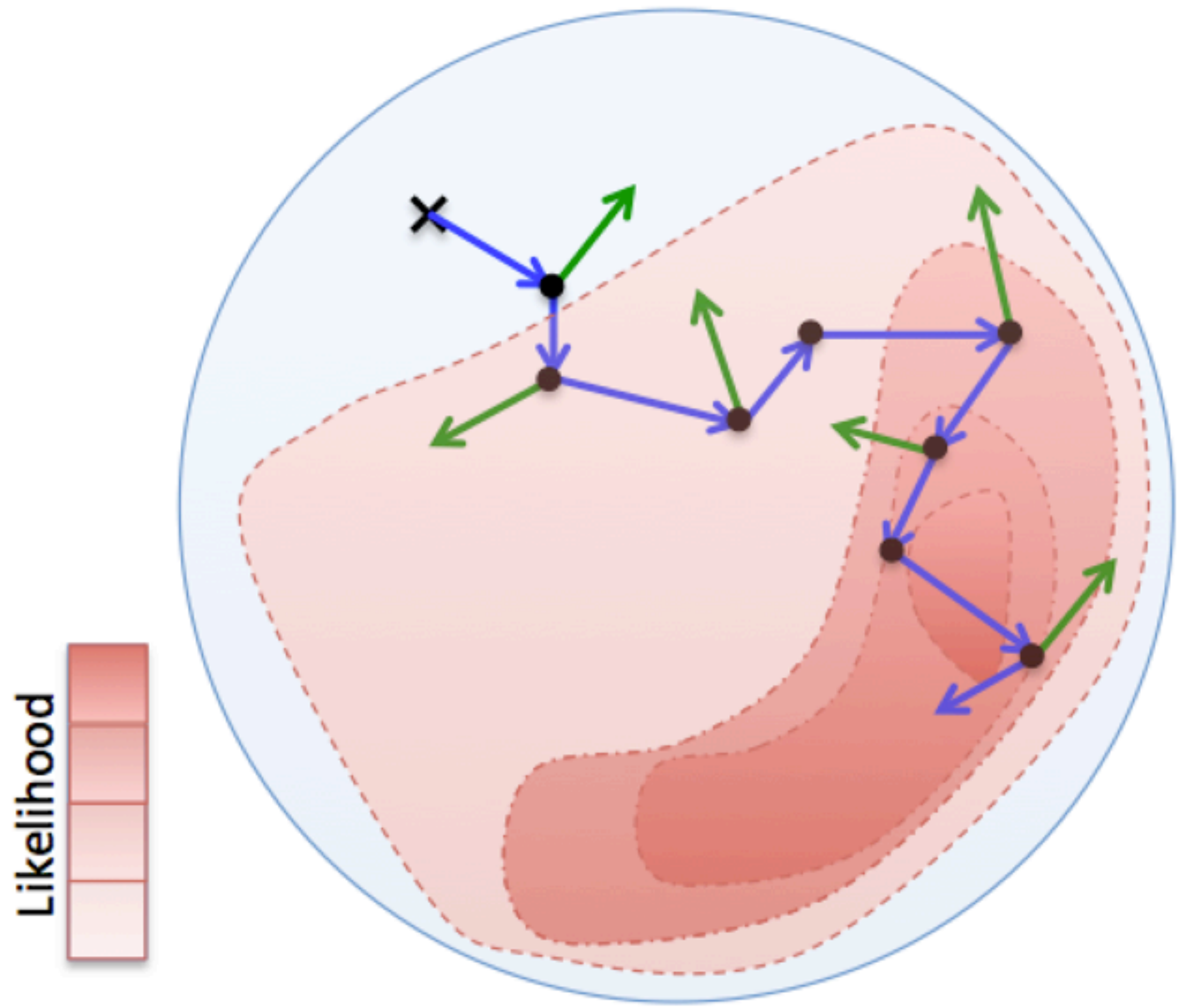
Metropolis-Hastings algorithm



Traditional likelihood-based methods (MCMC, Nested Sampling,...) allow to get samples from the full joint posterior

$$\boldsymbol{\theta} \sim p(\boldsymbol{\theta} \,|\, \mathbf{x}), \qquad \boldsymbol{\theta} \in \mathbb{R}^D$$

Metropolis-Hastings algorithm



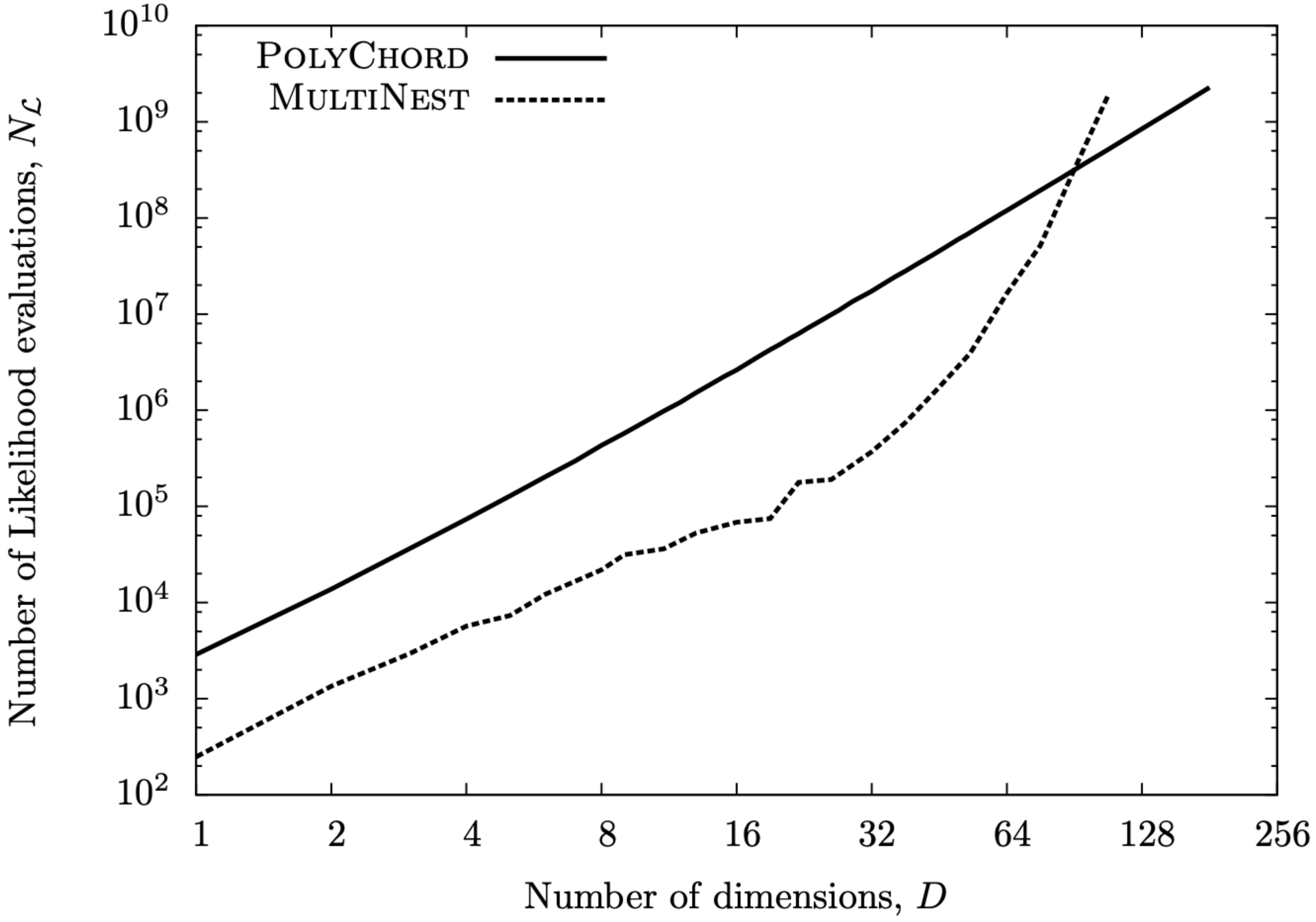
- Traditional likelihood-based methods (MCMC, Nested Sampling,...) allow to get samples from the full joint posterior

$$\boldsymbol{\theta} \sim p(\boldsymbol{\theta} \,|\, \mathbf{x}), \qquad \boldsymbol{\theta} \in \mathbb{R}^D$$

- Then we marginalise to get posteriors of interest

The curse of dimensionality

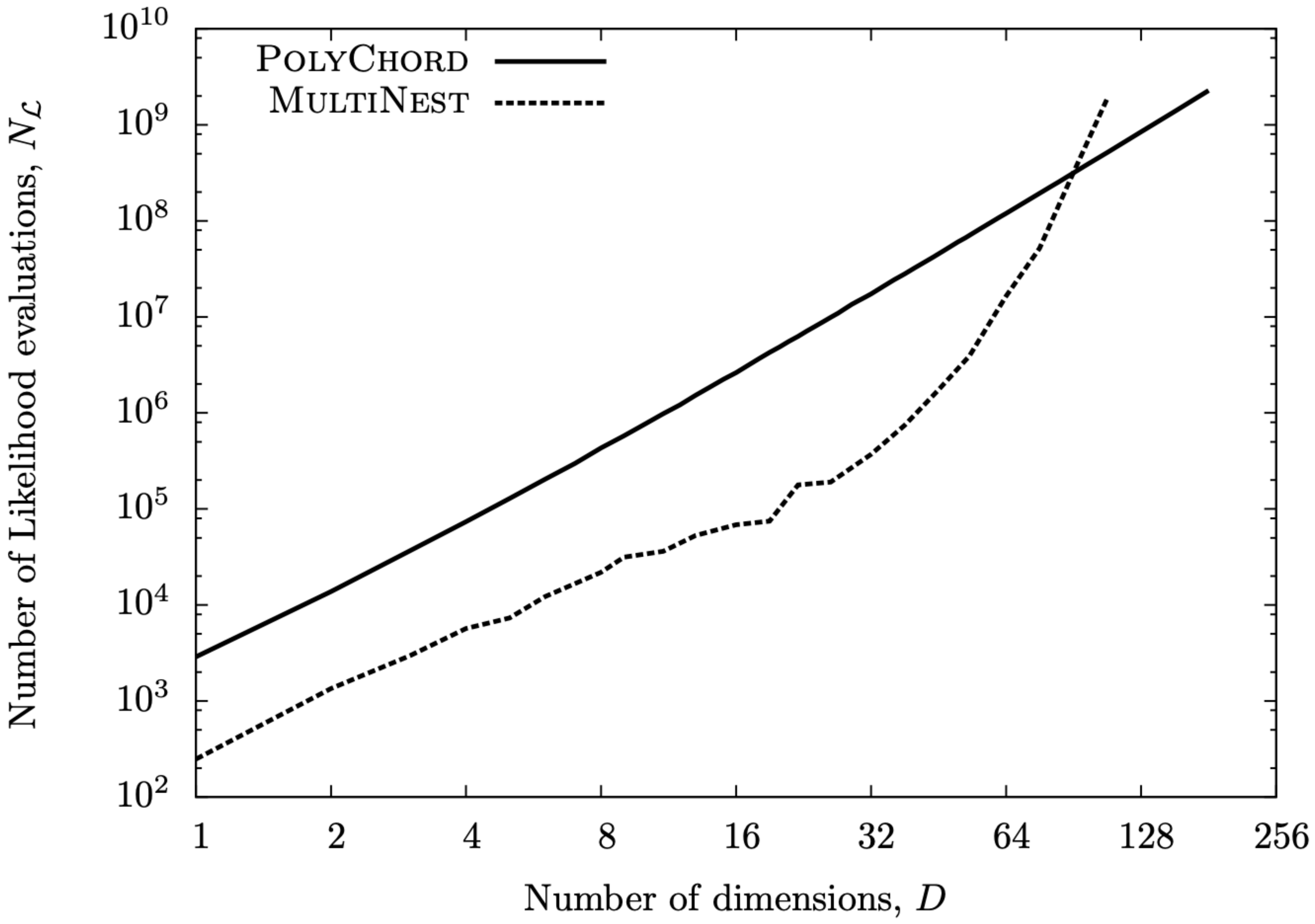
These methods scale poorly with the dimensionality of the parameter space



[Handley+ 15]

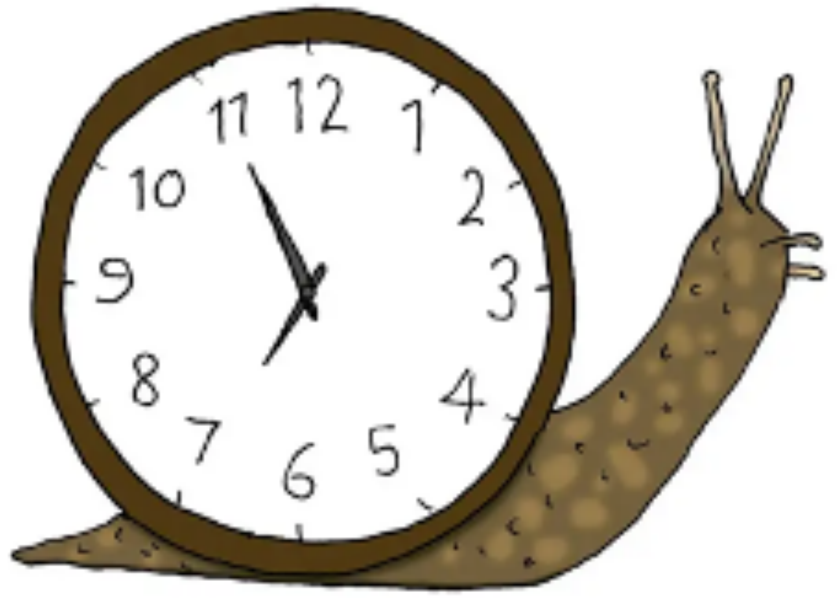
The curse of dimensionality

These methods scale poorly with the dimensionality of the parameter space



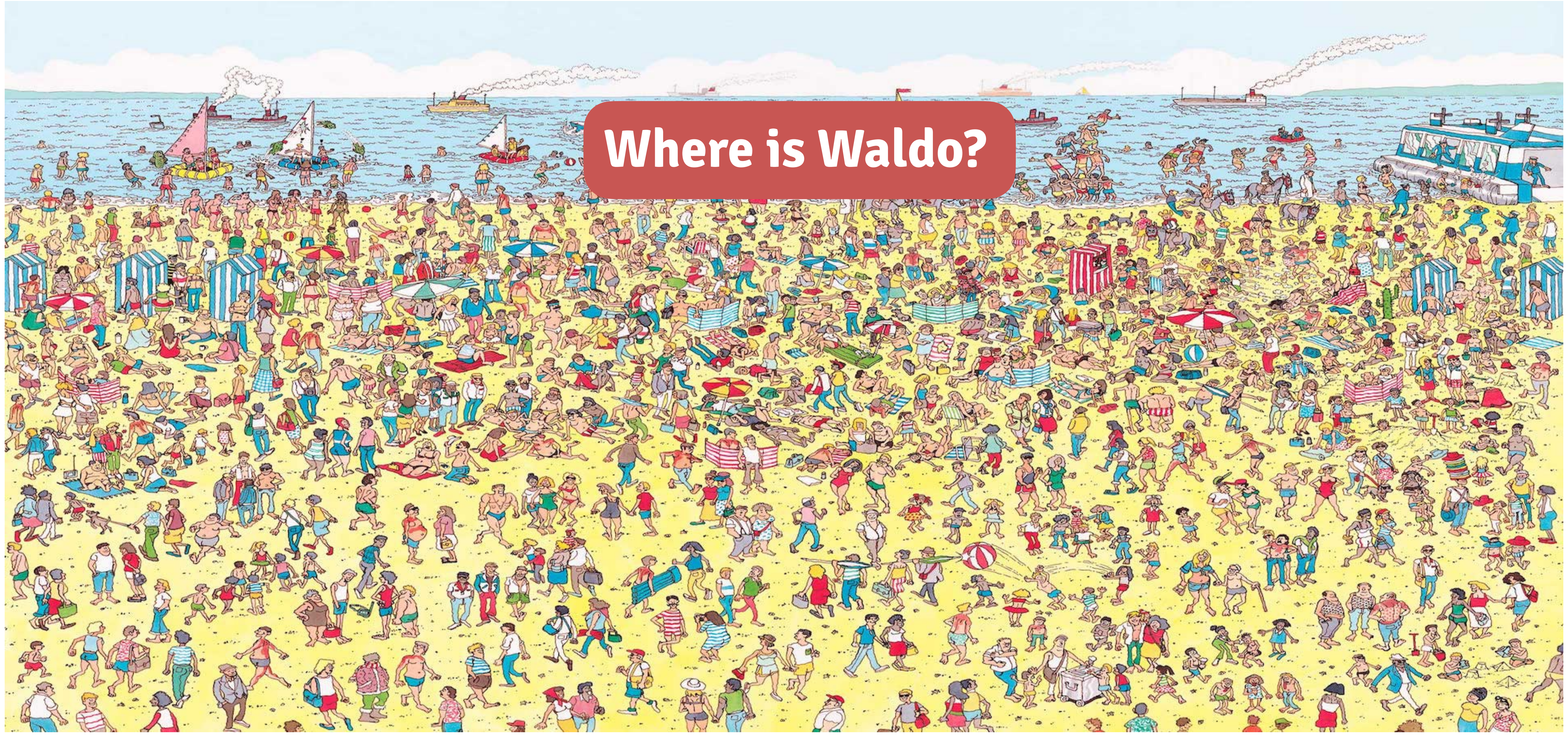
[Handley+ 15]

Ex: For Euclid, we expect to have **+100 nuisance parameters**



Weeks...

The curse of dimensionality



Marginal posterior

Joint posterior

$$P(\theta_{\text{waldo}} \,|\, x_0) = \int d\theta_{\text{Pierre}} d\theta_{\text{Theo}} d\theta_{\text{Julien}} \ldots d\theta_{\text{Hugo}} \; P(\theta_{\text{Waldo}}, \theta_{\text{Pierre}}, \theta_{\text{Theo}}, \theta_{\text{Julien}}, \ldots, \theta_{\text{Hugo}} \,|\, x_0)$$

Are there methods to overcome this problem?

Are there methods to overcome this problem?

Can machine learning be helpful?

A NEW HOPE

MNRE = **M**arginal **N**eural **R**atio **E**stimation

Implemented in Swyft* [Miller+ 20]

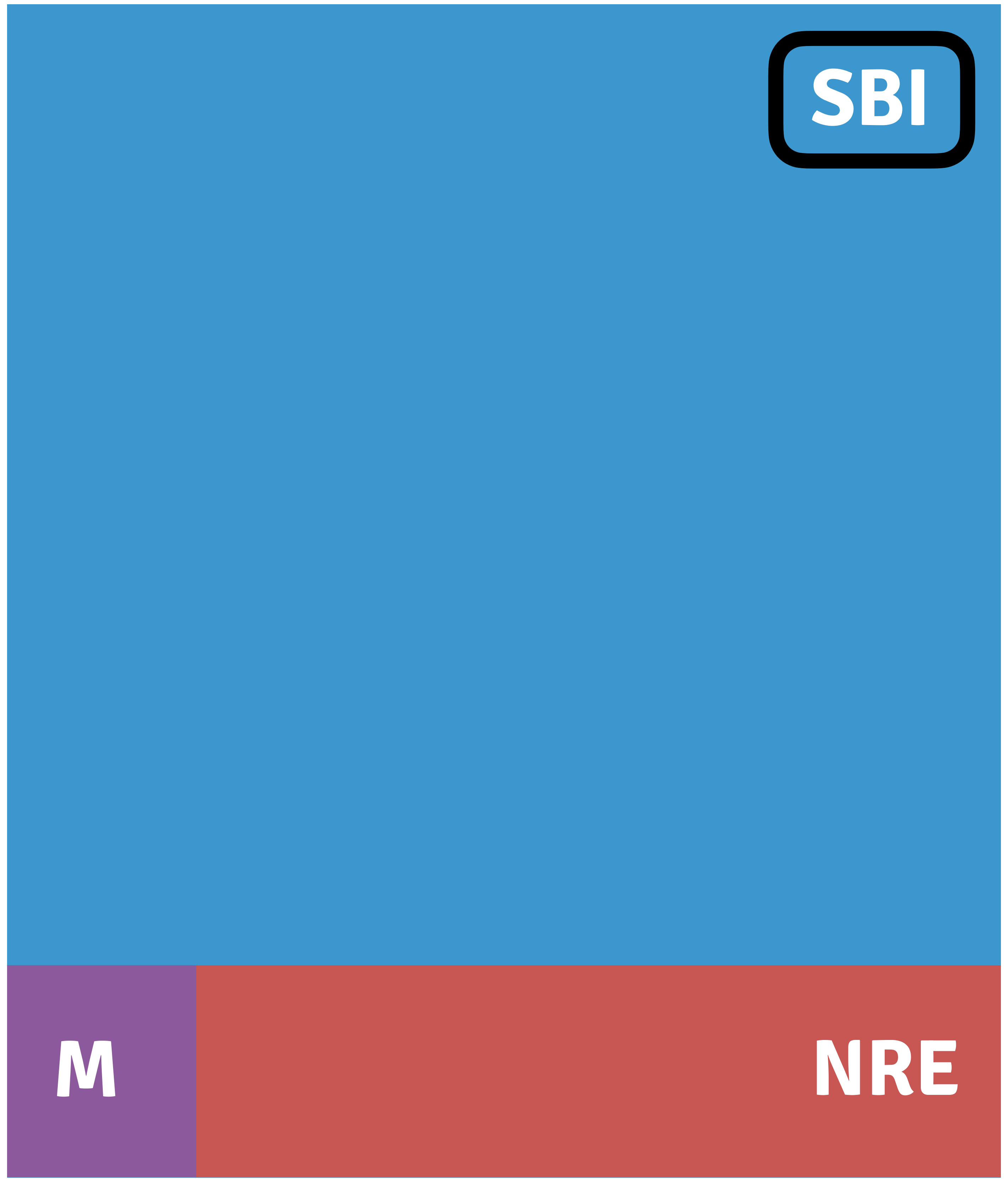
* Stop Wasting Your Precious Time

I. Why we need to go **beyond MCMC**

II. Our new approach: **Marginal Neural Ratio Estimation** ✓

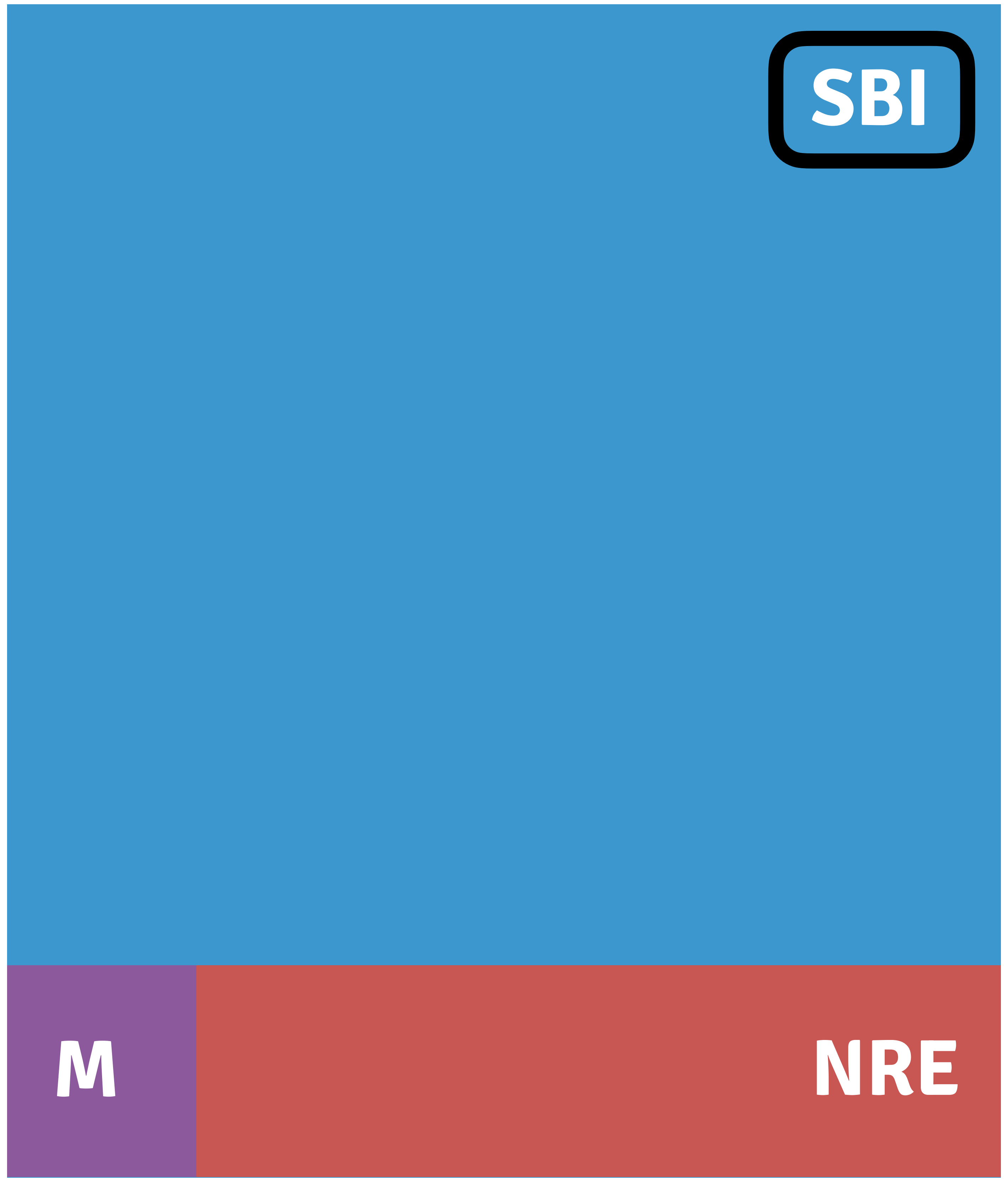
III. Applying MNRE to **Euclid** observables

Marginal Neural Ratio Estimation



Simulation-based inference
(or likelihood-free inference)

Marginal Neural Ratio Estimation



Simulation-based inference
(or likelihood-free inference)

↓

Stochastic simulator that maps from model parameters **θ** to data **x**

$\mathbf{x} \sim p(\mathbf{x} \,|\, \boldsymbol{\theta})$ (implicit likelihood)

We can simulate N samples that can be used as training data for a neural network

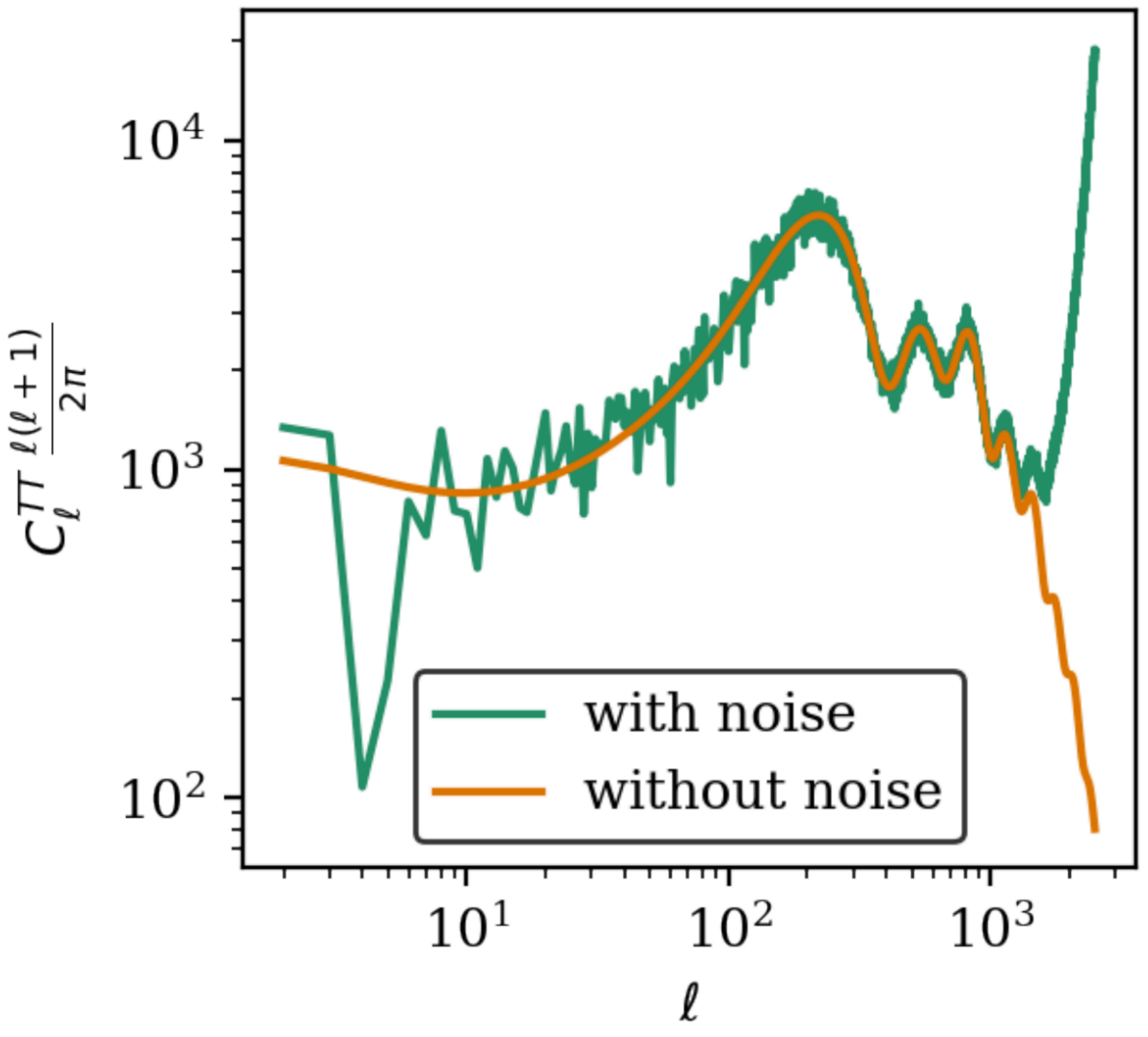
$$\{(\mathbf{x}^{(1)}, \boldsymbol{\theta}^{(1)}), (\mathbf{x}^{(2)}, \boldsymbol{\theta}^{(2)}), \ldots, (\mathbf{x}^{(N)}, \boldsymbol{\theta}^{(N)})\}$$

We can simulate N samples that can be used as training data for a neural network

$$\{(\mathbf{x}^{(1)}, \boldsymbol{\theta}^{(1)}), (\mathbf{x}^{(2)}, \boldsymbol{\theta}^{(2)}), \ldots, (\mathbf{x}^{(N)}, \boldsymbol{\theta}^{(N)})\}$$

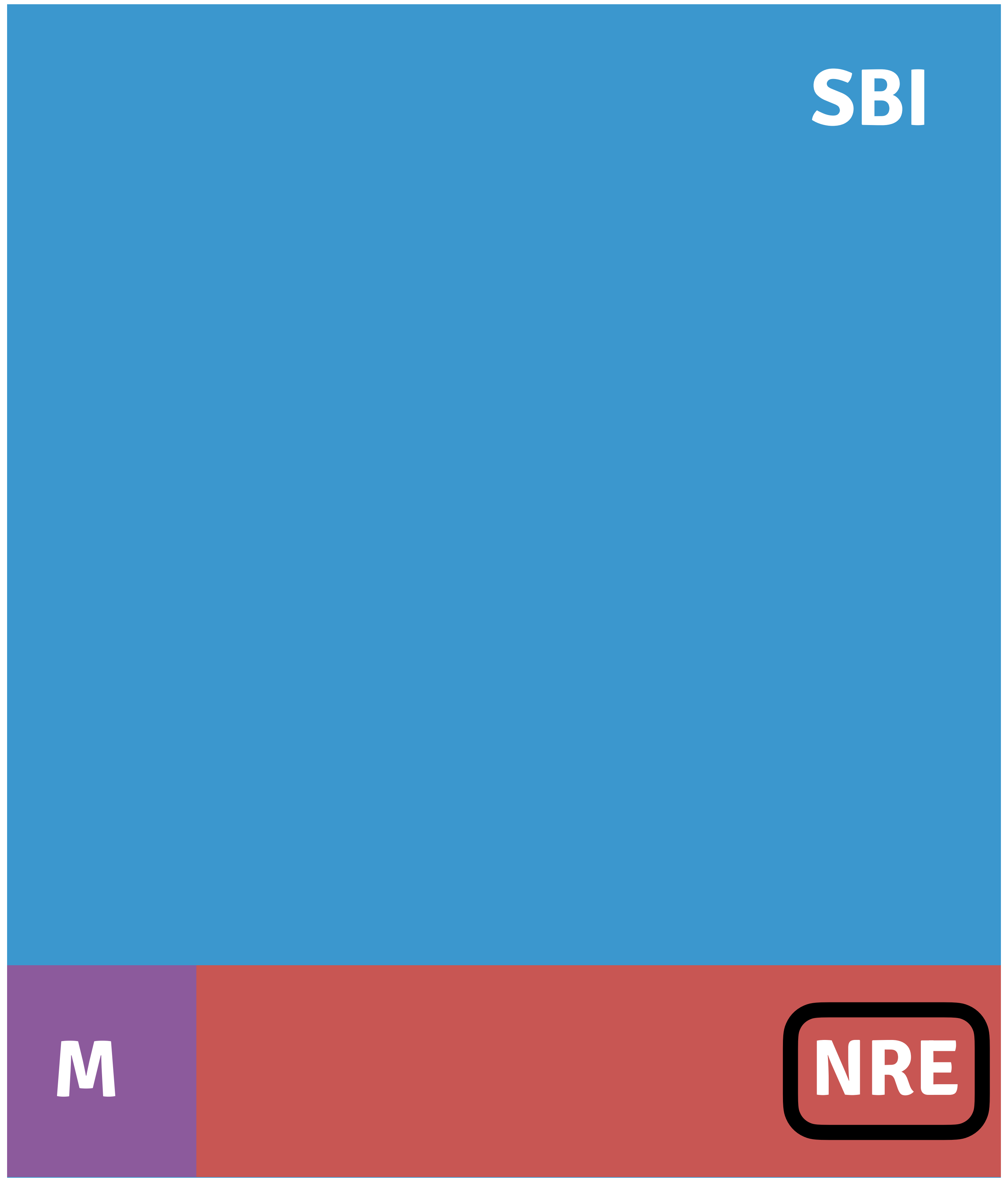
Ex: CMB simulator

$$\boldsymbol{\theta} \rightarrow C_\ell(\boldsymbol{\theta}) \rightarrow C_\ell(\boldsymbol{\theta}) + N_\ell$$



Cole+ 22

Marginal Neural Ratio Estimation

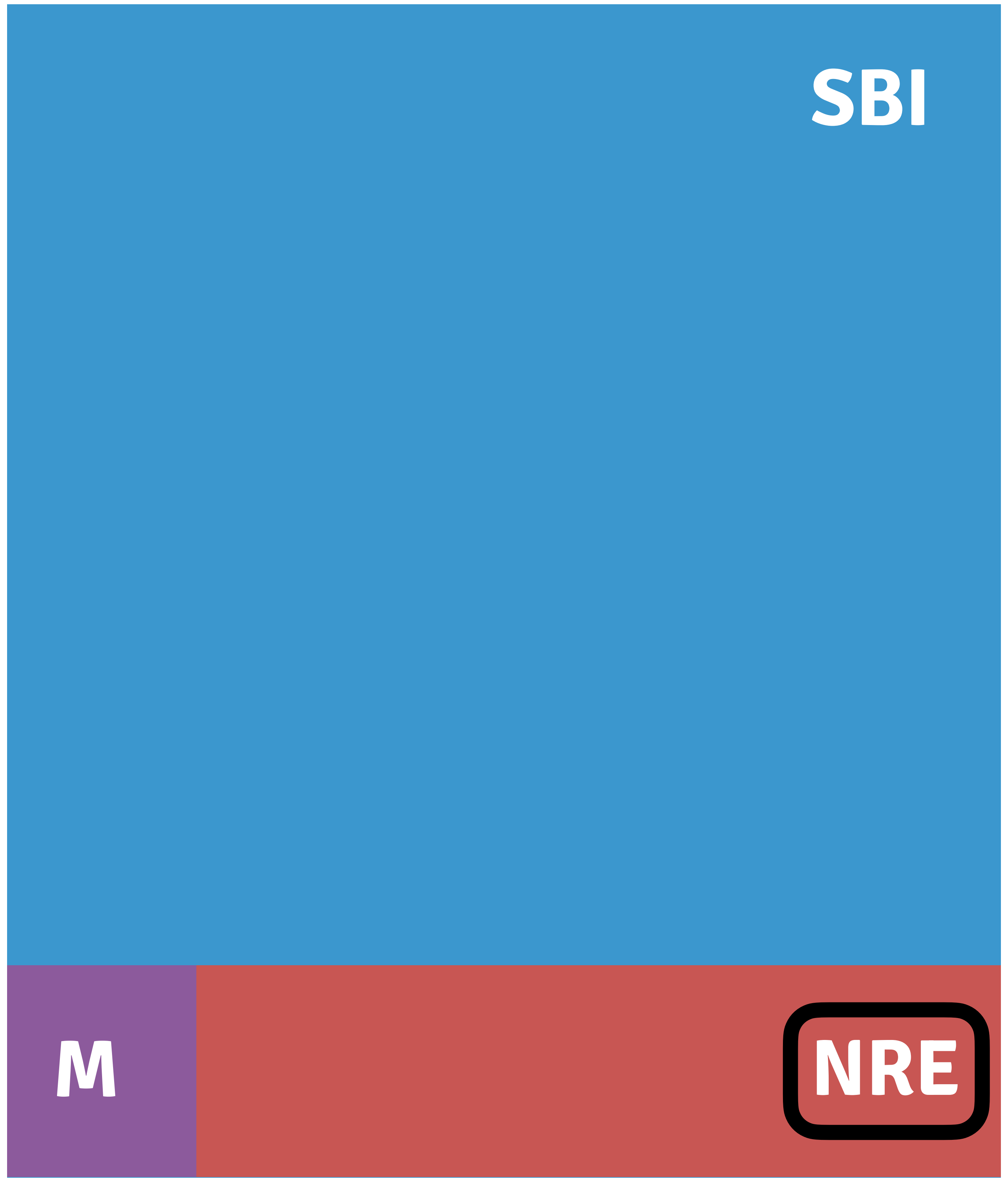


Neural Ratio Estimation

Instead of directly estimating the posterior, estimate:

$$r(\mathbf{x};\boldsymbol{\theta}) = \frac{p(\mathbf{x},\boldsymbol{\theta})}{p(\mathbf{x})p(\boldsymbol{\theta})}$$

Marginal Neural Ratio Estimation



Neural Ratio Estimation

Instead of directly estimating the posterior, estimate:

$$r(\mathbf{x};\boldsymbol{\theta}) = \frac{p(\mathbf{x},\boldsymbol{\theta})}{p(\mathbf{x})p(\boldsymbol{\theta})}$$

It's easy to show that:

$$\frac{p(\mathbf{x},\boldsymbol{\theta})}{p(\mathbf{x})p(\boldsymbol{\theta})} = \frac{p(\boldsymbol{\theta}\,|\,\mathbf{x})}{p(\boldsymbol{\theta})}$$

(posterior-to-prior ratio)

$\boldsymbol{\theta} \sim p(\boldsymbol{\theta})$ Draw labels **(cat, dog)**

$\boldsymbol{x} \sim p(\boldsymbol{x})$ Draw images

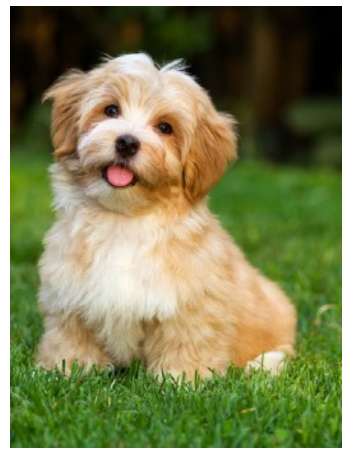
$\boldsymbol{\theta} \sim p(\boldsymbol{\theta})$ Draw labels **(cat, dog)**

$\boldsymbol{x} \sim p(\boldsymbol{x})$ Draw images

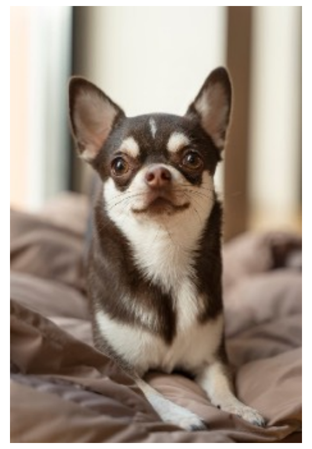
$(\mathbf{x}, \boldsymbol{\theta}) \sim p(\mathbf{x}, \boldsymbol{\theta})$ (jointly drawn)

Cat

Dog



Dog



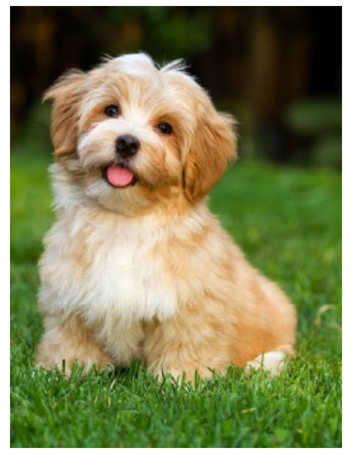
$\boldsymbol{\theta} \sim p(\boldsymbol{\theta})$ Draw labels **(cat, dog)**

$\boldsymbol{x} \sim p(\boldsymbol{x})$ Draw images

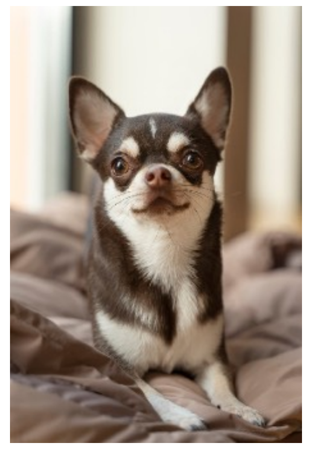
$(\mathbf{x}, \boldsymbol{\theta}) \sim p(\mathbf{x}, \boldsymbol{\theta})$ (jointly drawn)

Cat

Dog



Dog



$(\mathbf{x}, \boldsymbol{\theta}) \sim p(\mathbf{x})p(\boldsymbol{\theta})$ (marginally drawn)

Cat

Dog

Cat

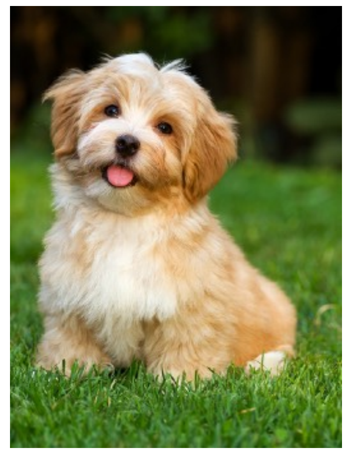
$\boldsymbol{\theta} \sim p(\boldsymbol{\theta})$ Draw labels **(cat, dog)**

$\boldsymbol{x} \sim p(\boldsymbol{x})$ Draw images

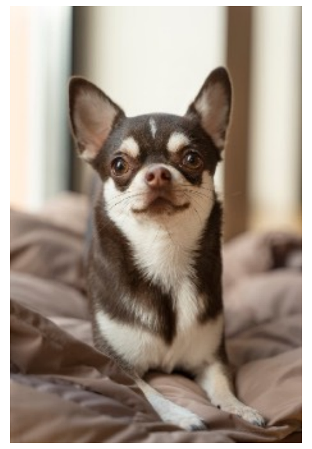
$(\mathbf{x}, \boldsymbol{\theta}) \sim p(\mathbf{x}, \boldsymbol{\theta})$ (jointly drawn)

Cat

Dog



Dog



$(\mathbf{x}, \boldsymbol{\theta}) \sim p(\mathbf{x})p(\boldsymbol{\theta})$ (marginally drawn)

Cat

Dog

Cat

Given some (x, θ) pair, are they drawn jointly or marginally?

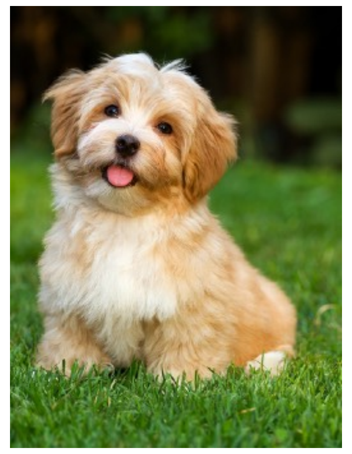
$\boldsymbol{\theta} \sim p(\boldsymbol{\theta})$ Draw labels **(cat, dog)**

$\boldsymbol{x} \sim p(\boldsymbol{x})$ Draw images

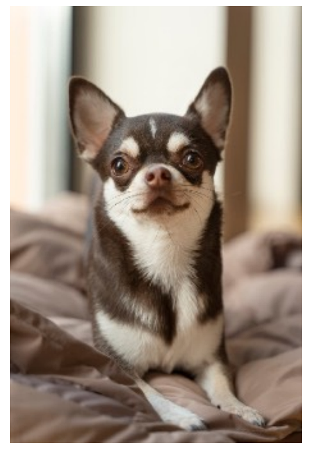
$(\mathbf{x}, \boldsymbol{\theta}) \sim p(\mathbf{x}, \boldsymbol{\theta})$ (jointly drawn)

Cat

Dog



Dog



$(\mathbf{x}, \boldsymbol{\theta}) \sim p(\mathbf{x})p(\boldsymbol{\theta})$ (marginally drawn)

Cat

Dog

Cat

Given some (x, θ) pair, are they drawn jointly or marginally?

→ Rephrase inference as a **binary classification problem**

Strategy: train a neural network $d_\phi(\mathbf{x}, \boldsymbol{\theta}) \in [0,1]$ as a binary classifier, so that

$$d_\phi(\mathbf{x}, \boldsymbol{\theta}) \simeq 1 \quad \text{if} \quad (\mathbf{x}, \boldsymbol{\theta}) \sim p(\mathbf{x}, \boldsymbol{\theta}) = p(\mathbf{x} \mid \boldsymbol{\theta}) p(\boldsymbol{\theta})$$

$$d_\phi(\mathbf{x}, \boldsymbol{\theta}) \simeq 0 \quad \text{if} \quad (\mathbf{x}, \boldsymbol{\theta}) \sim p(\mathbf{x}) p(\boldsymbol{\theta})$$

Note: Φ denotes all the network parameters

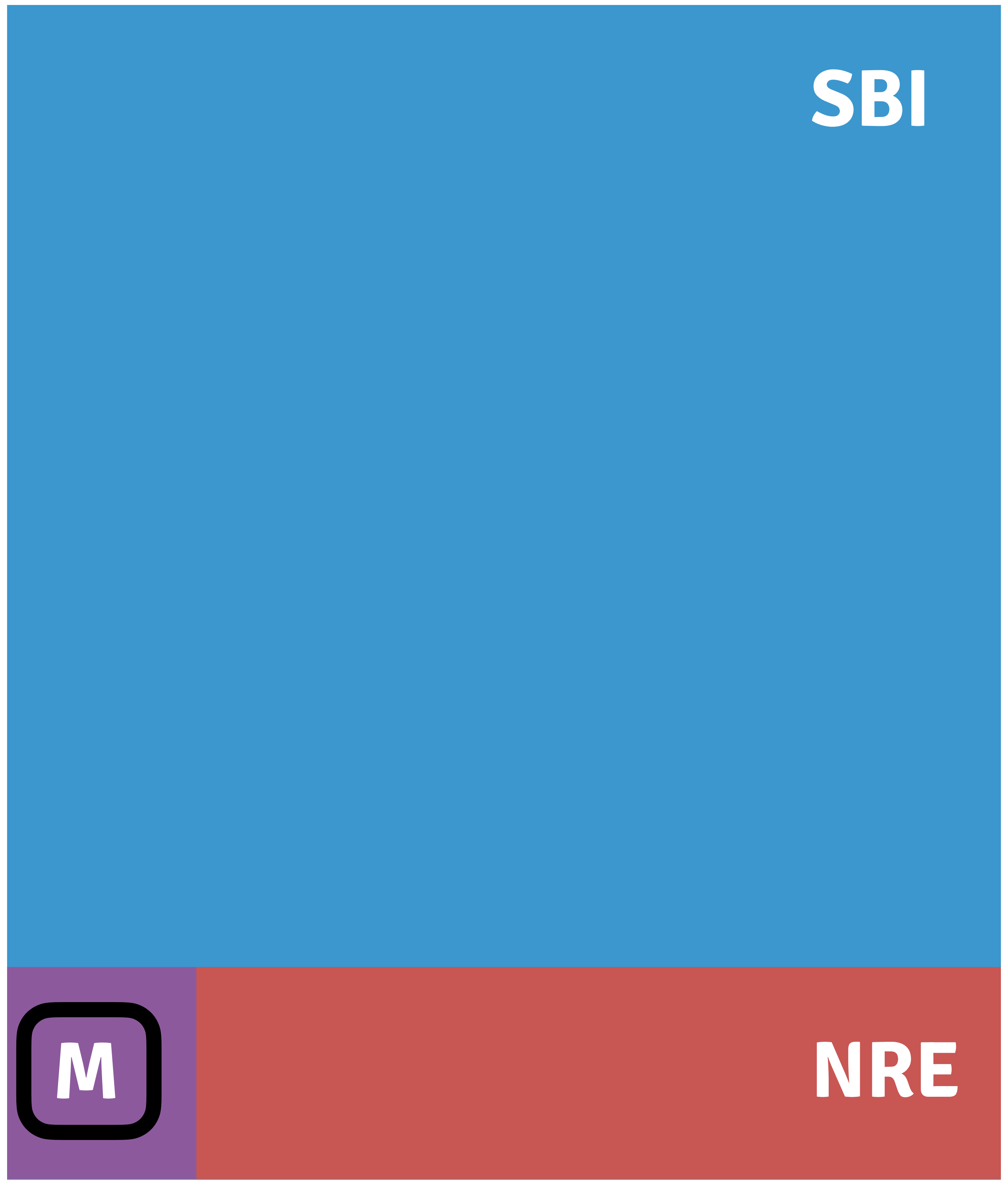
We have to minimise a **loss function** w.r.t. the network params. Φ

$$L[d_\phi(\mathbf{x},\boldsymbol{\theta})] = -\int d\mathbf{x}d\boldsymbol{\theta}\left[p(\mathbf{x},\boldsymbol{\theta})\ln(d_\phi(\mathbf{x},\boldsymbol{\theta})) + p(\mathbf{x})p(\boldsymbol{\theta})\ln(1-d_\phi(\mathbf{x},\boldsymbol{\theta}))\right]$$

which yields

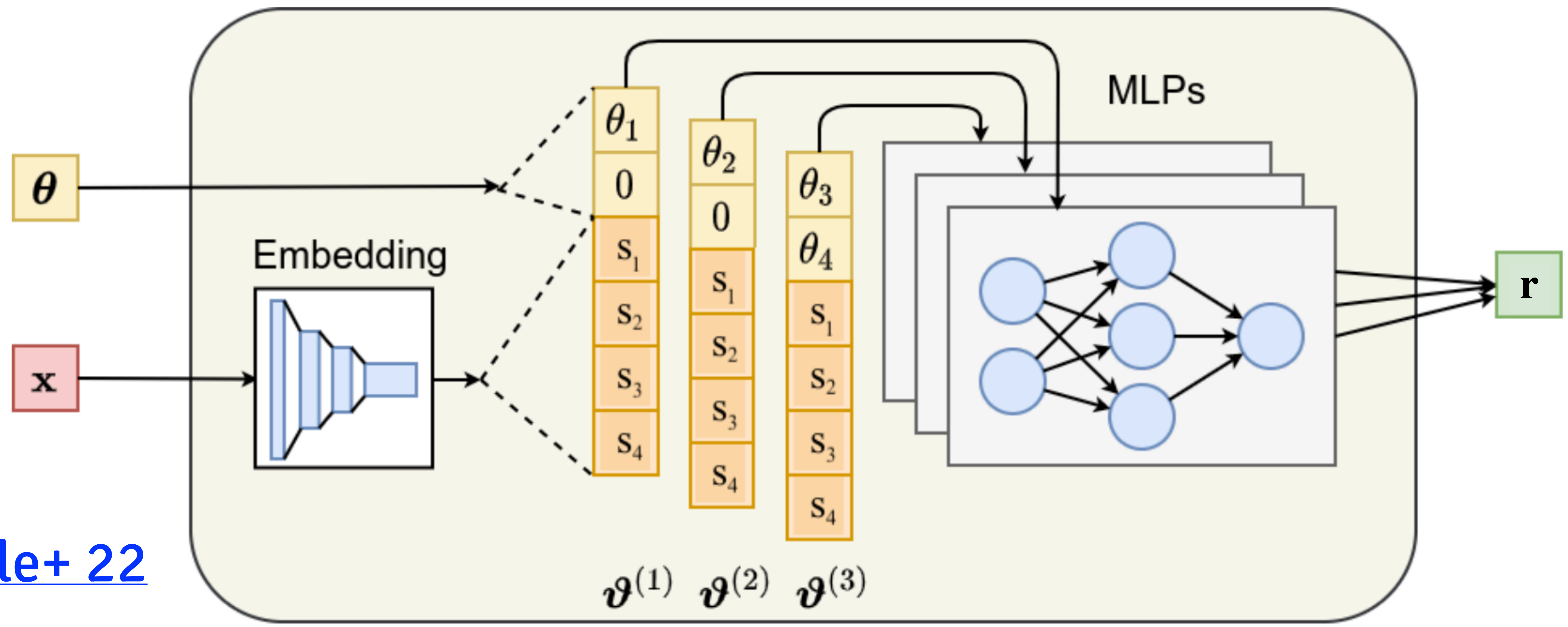
$$d_\phi(\mathbf{x},\boldsymbol{\theta}) \simeq \frac{p(\mathbf{x},\boldsymbol{\theta})}{p(\mathbf{x},\boldsymbol{\theta}) + p(\mathbf{x})p(\boldsymbol{\theta})} = \frac{r(\mathbf{x};\boldsymbol{\theta})}{r(\mathbf{x};\boldsymbol{\theta})+1}$$

Marginal Neural Ratio Estimation



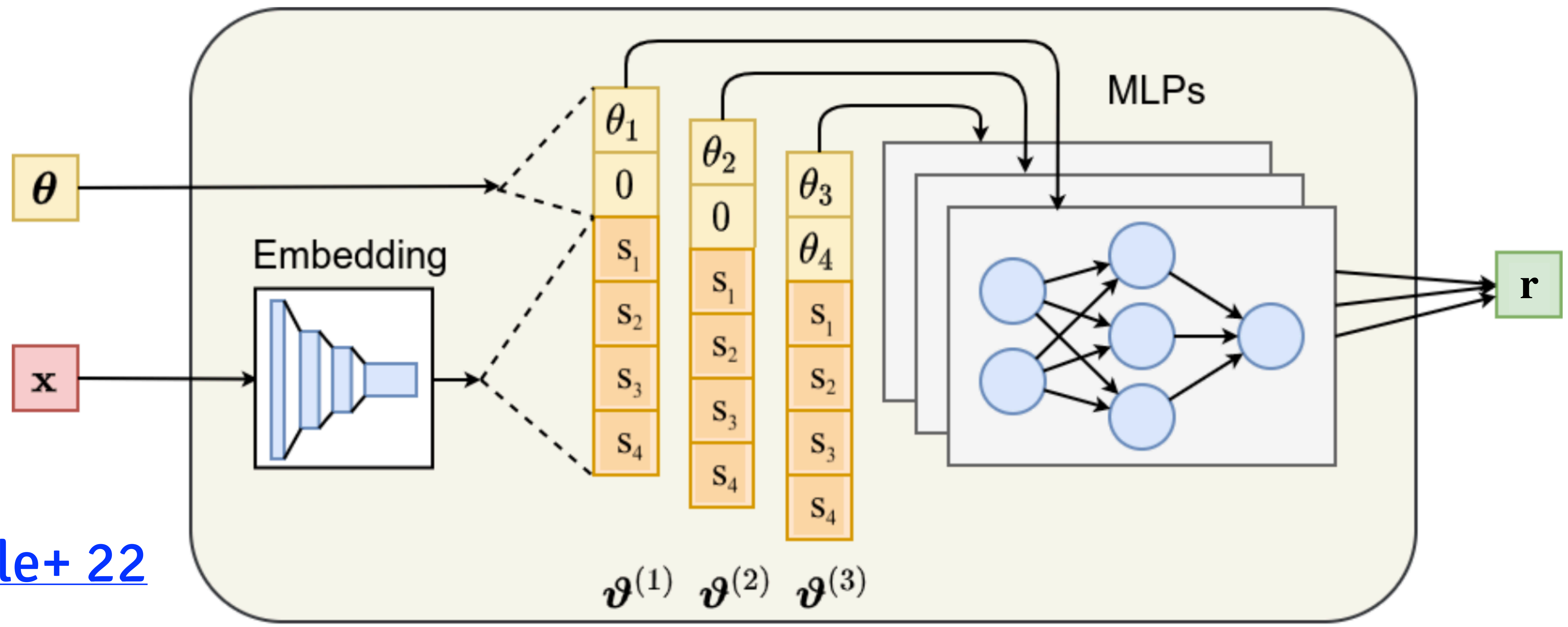
Marginal inference

We can directly target marginal posteriors of interest, and forget about the rest



$\boldsymbol{\theta}$
x
Embedding
θ_1
0
s_1
s_2
s_3
s_4
θ_2
0
s_1
s_2
s_3
s_4
θ_3
θ_4
s_1
s_2
s_3
s_4
MLPs
r
$\boldsymbol{\vartheta}^{(1)}$ $\boldsymbol{\vartheta}^{(2)}$ $\boldsymbol{\vartheta}^{(3)}$

Cole+ 22



Cole+ 22

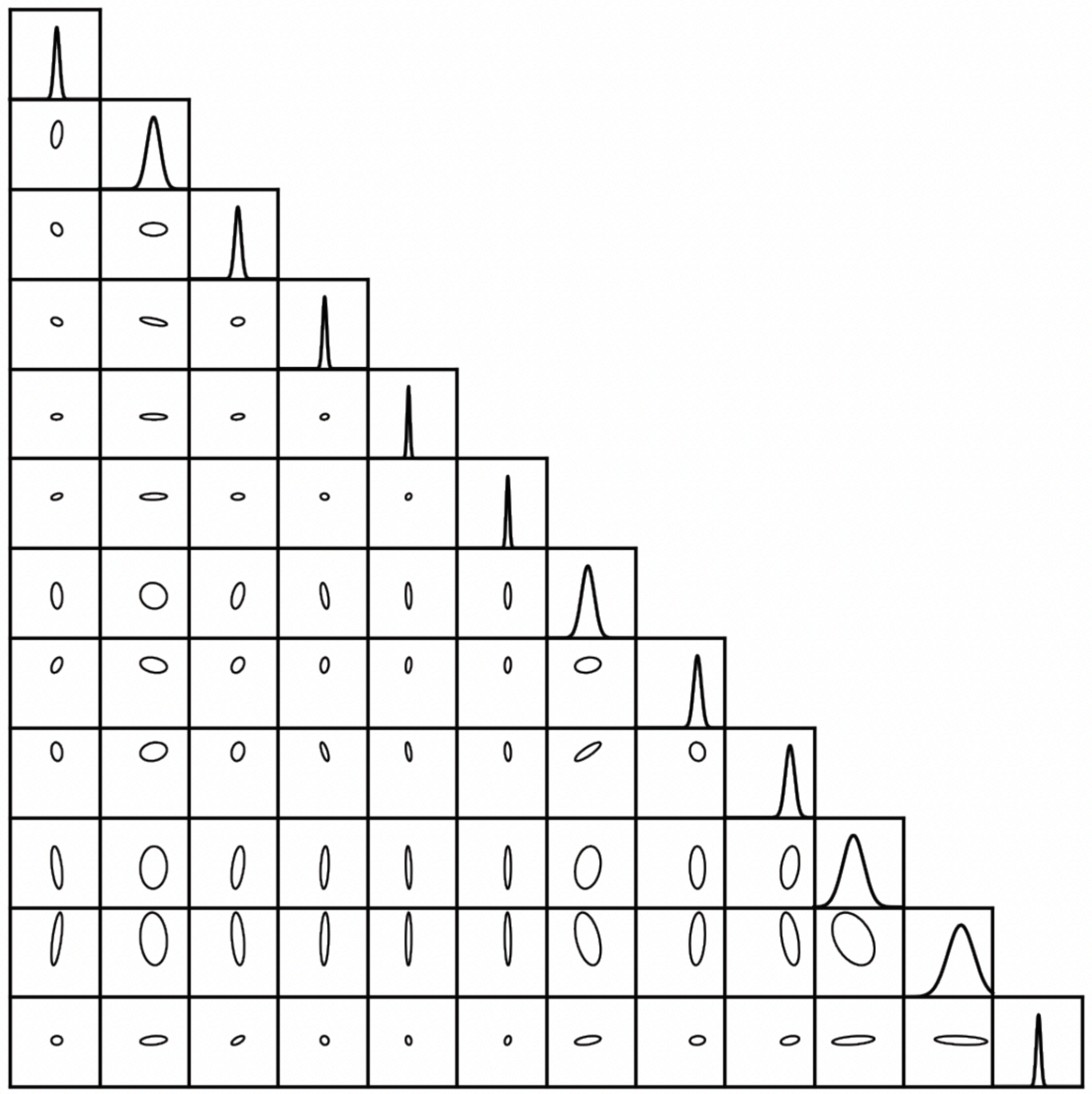
Estimates

$$p(\theta_1 \mid \mathbf{x}), \quad p(\theta_2 \mid \mathbf{x}), \quad p(\theta_3, \theta_4 \mid \mathbf{x})$$

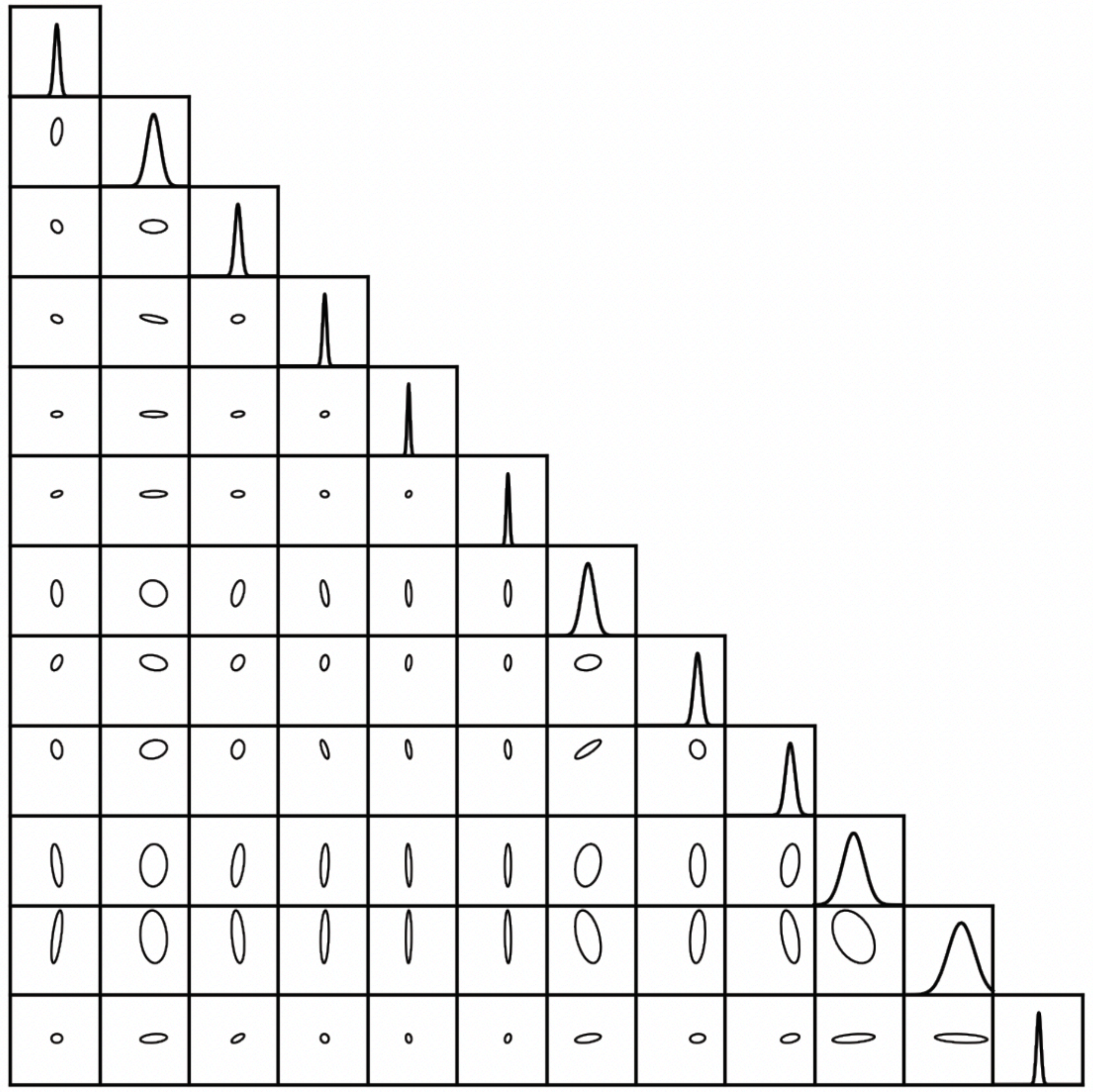
Does **not** estimate

$$p(\theta_1, \theta_2 \mid \mathbf{x}), \quad p(\theta_1, \theta_2, \theta_3 \mid \mathbf{x}), \ldots$$

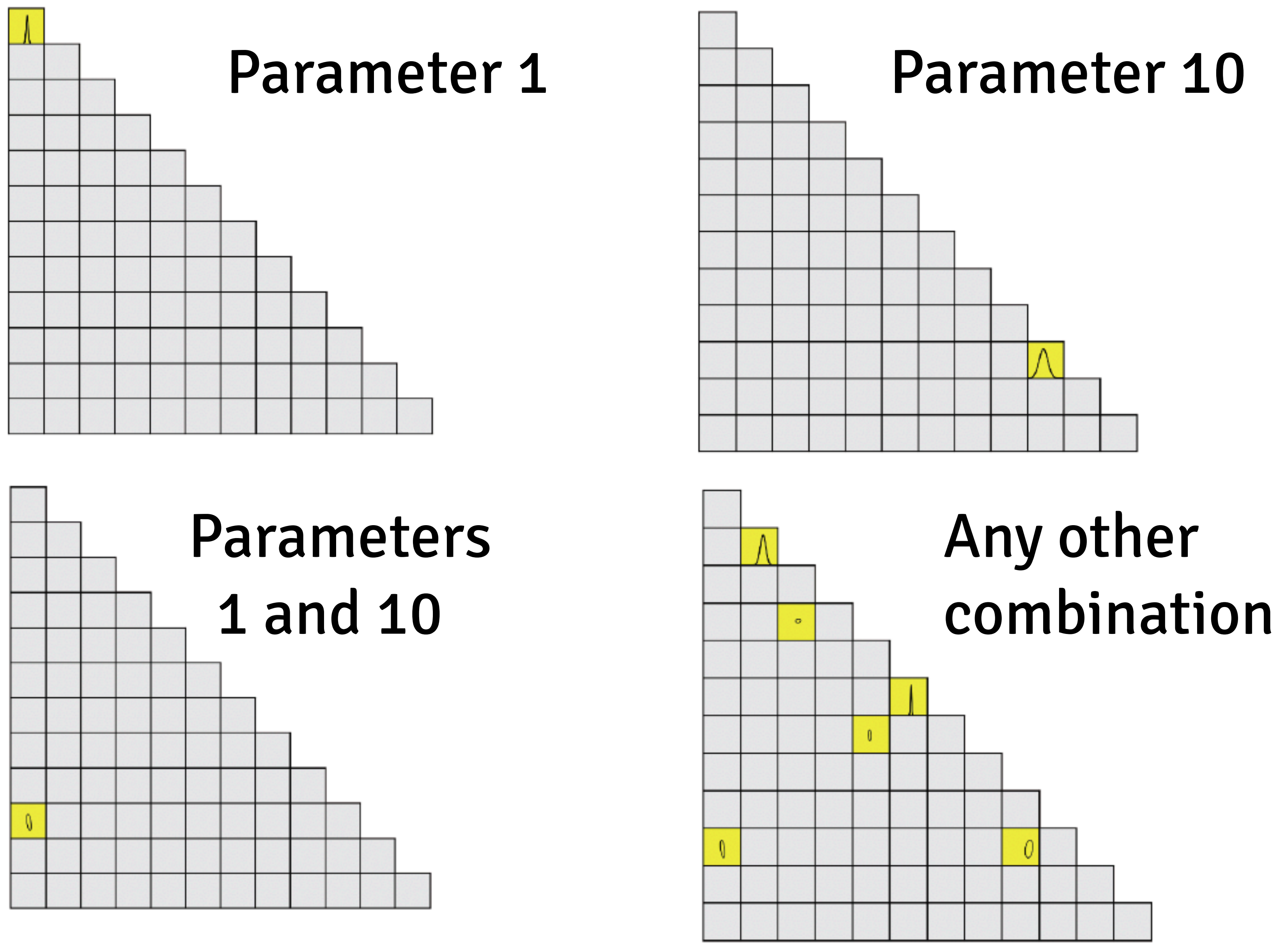
Instead of estimating all parameters...



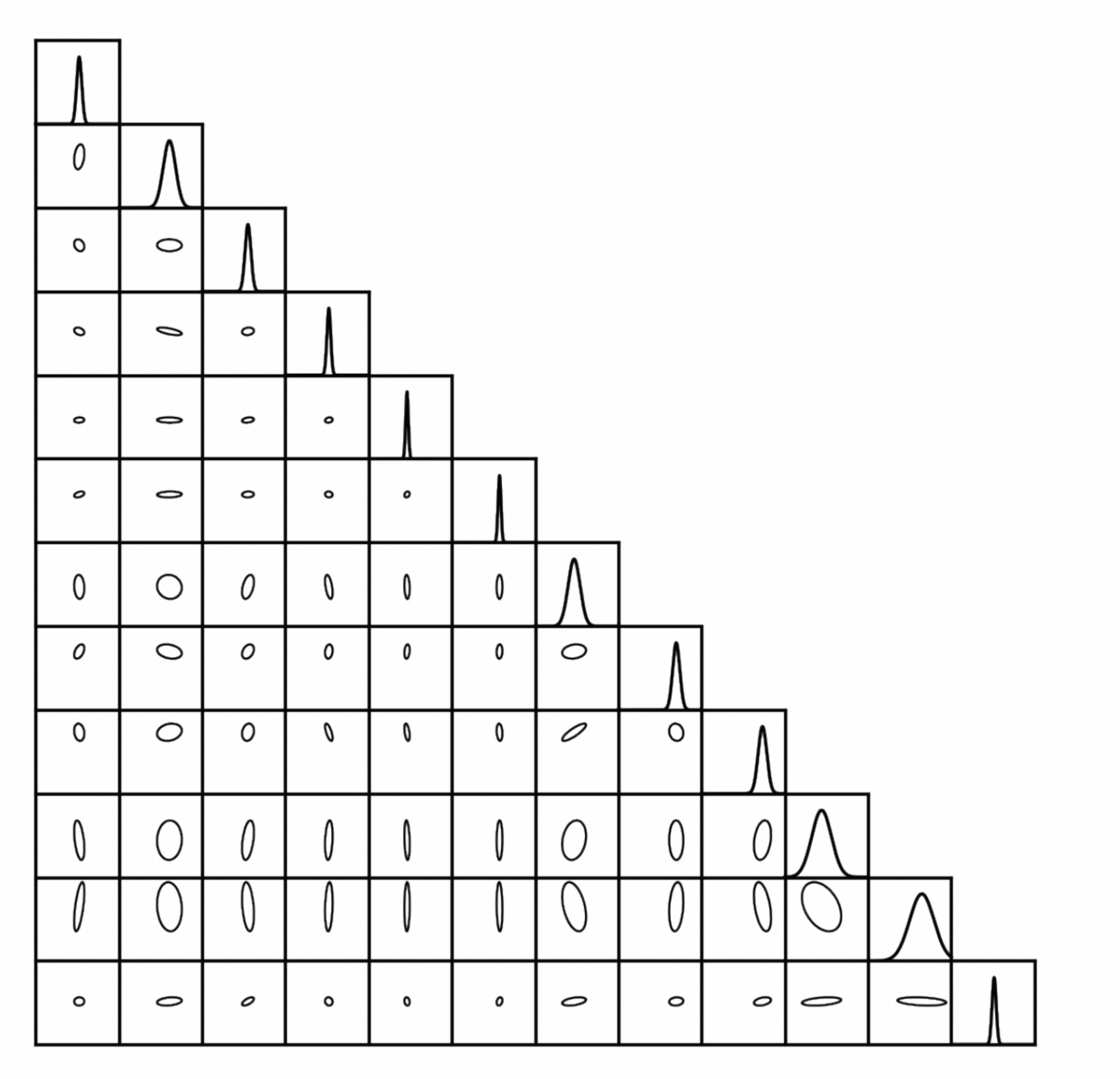
Instead of estimating all parameters...



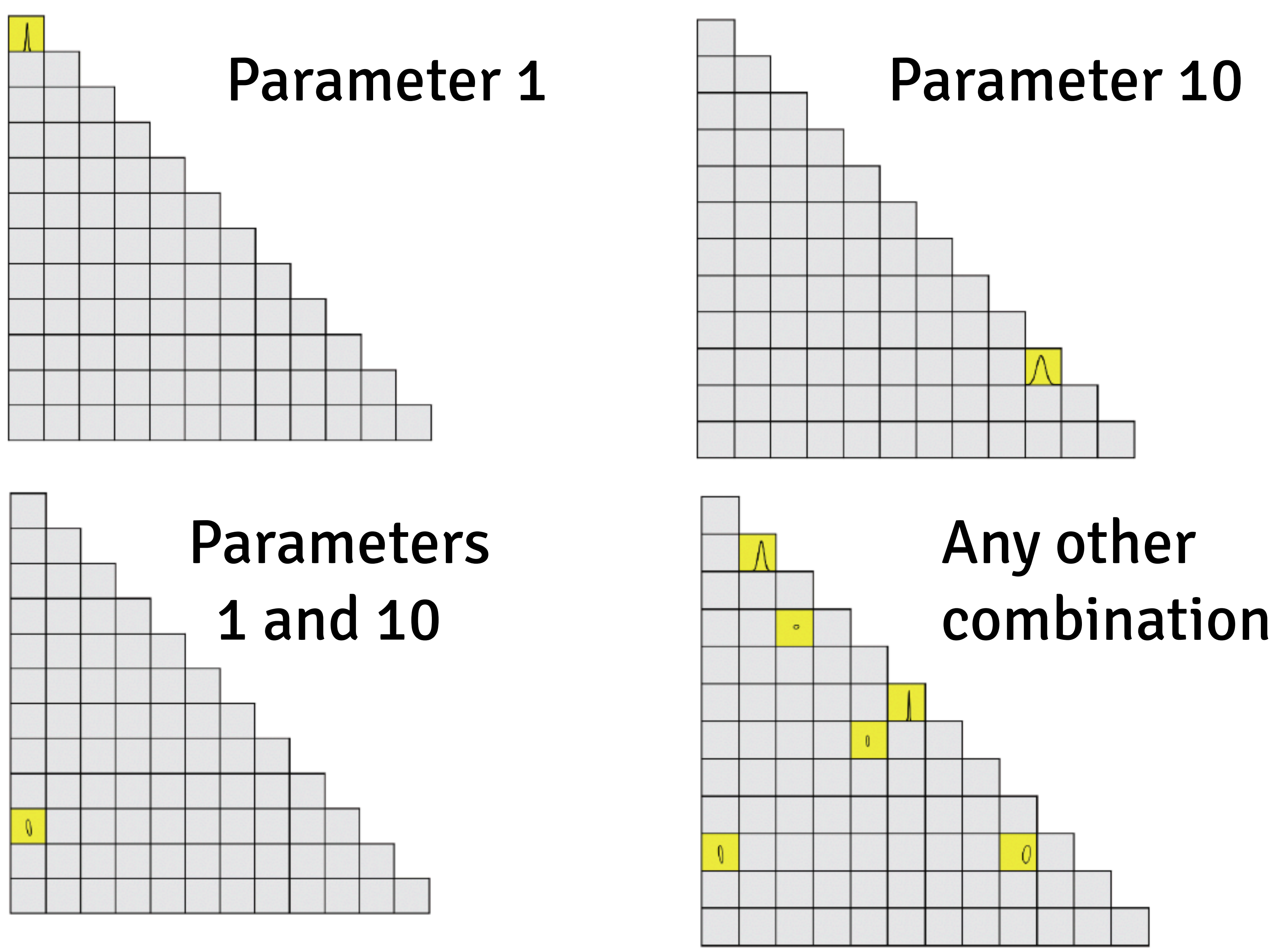
... we can cherry-pick what we care about



Instead of estimating all parameters...



... we can cherry-pick what we care about



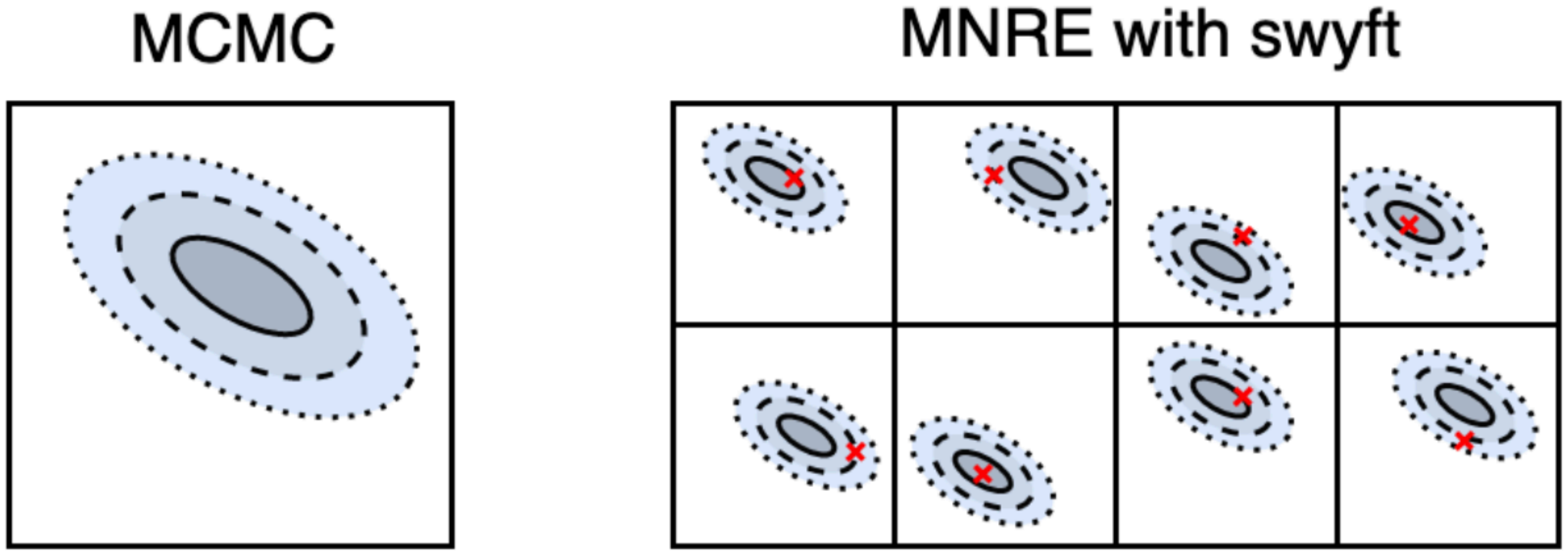
Much more flexible much more efficient!

But can we trust our results?

But can we trust our results?

...even if NNs are often seen as "black boxes”, it is possible to perform statistical consistency tests which are impossible with MCMC

Exploit MNRE's local amortization:



$p(\theta|\mathbf{x}_o)$

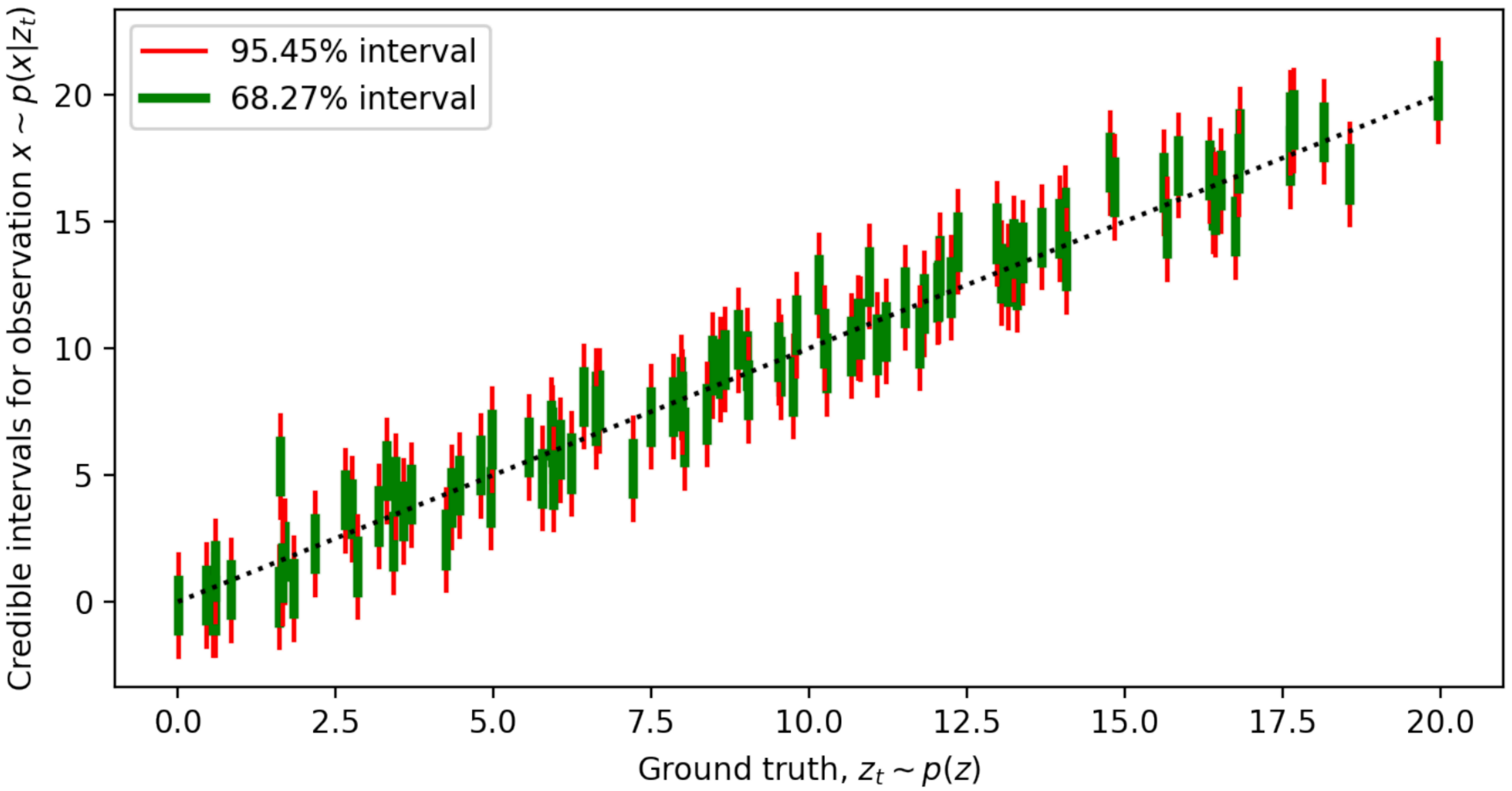
estimates the posterior for one single observation

$p(\theta|\mathbf{x}) \quad \forall \mathbf{x} \sim p(\mathbf{x})$

simultaneously estimates the posteriors for all simulated observations

Cole+ 22

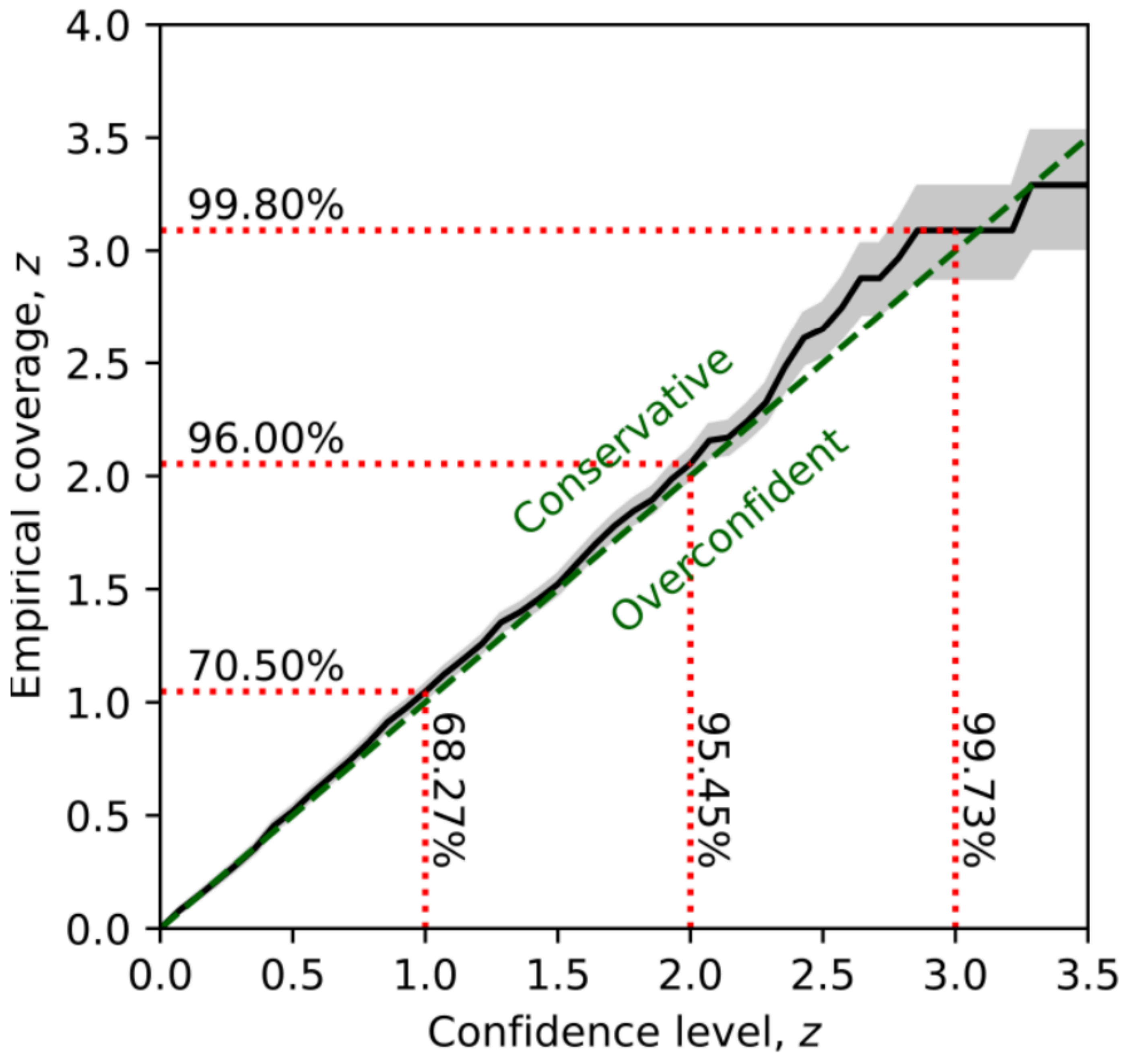
We can empirically estimate the Bayesian coverage



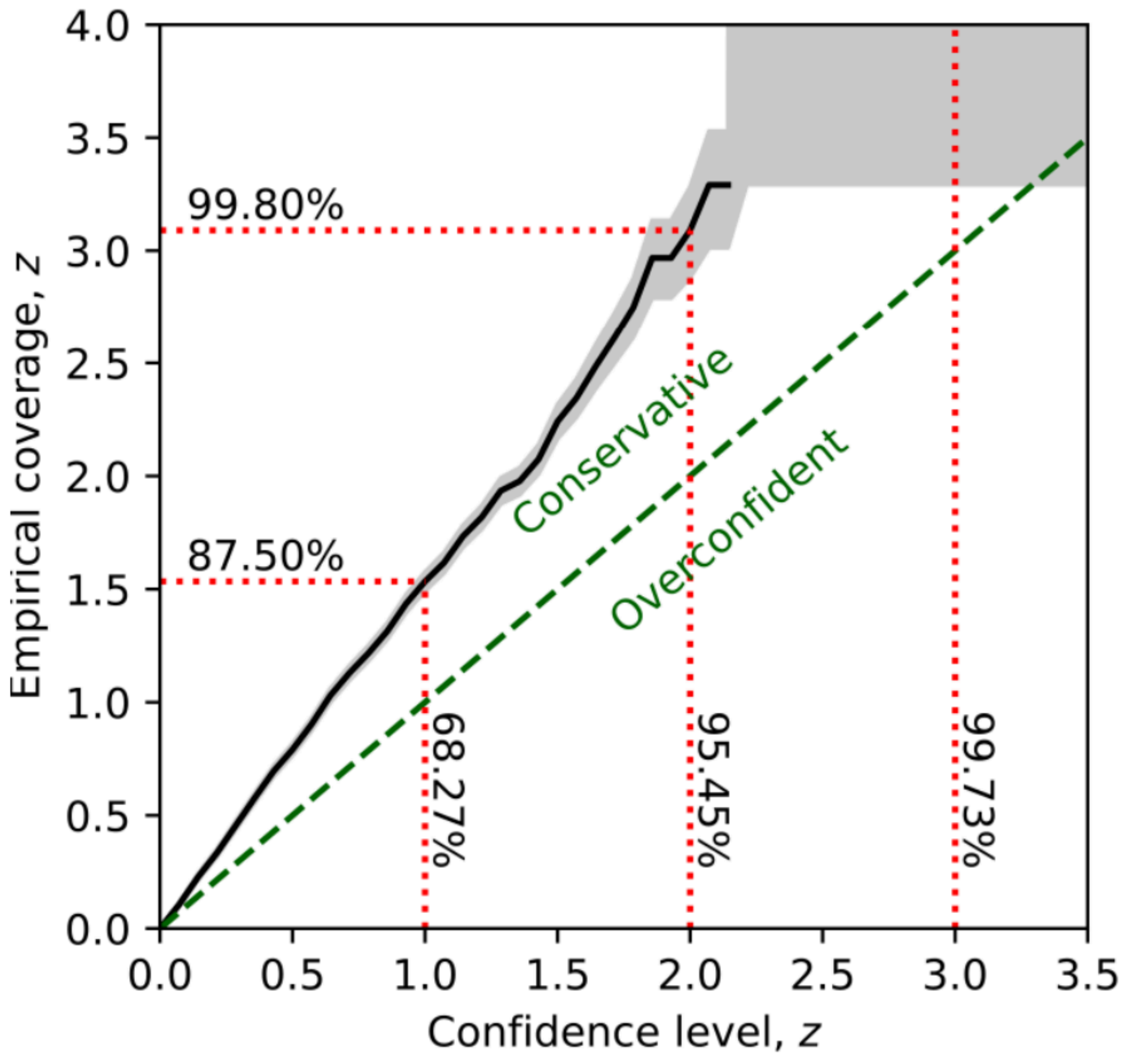
Cole+ 22

We can empirically estimate the Bayesian coverage

Converged network



Non-converged network



[Cole+ 22]

MNRE has been successfully applied in many contexts:

- **Strong lensing** [Montel+ 22]
- **Stellar Streams** [Alvey+ 23]
- **Gravitational Waves** [Bhardwaj+ 23] [Alvey+ 23]
- **CMB** [Cole+ 22]
- **21-cm** [Saxena+ 23]

MNRE has been successfully applied in many contexts:

- **Strong lensing** [Montel+ 22]
- **Stellar Streams** [Alvey+ 23]
- **Gravitational Waves** [Bhardwaj+ 23] [Alvey+ 23]
- **CMB** [Cole+ 22]
- **21-cm** [Saxena+ 23]

Our goal: apply MNRE to **Euclid** primary observables

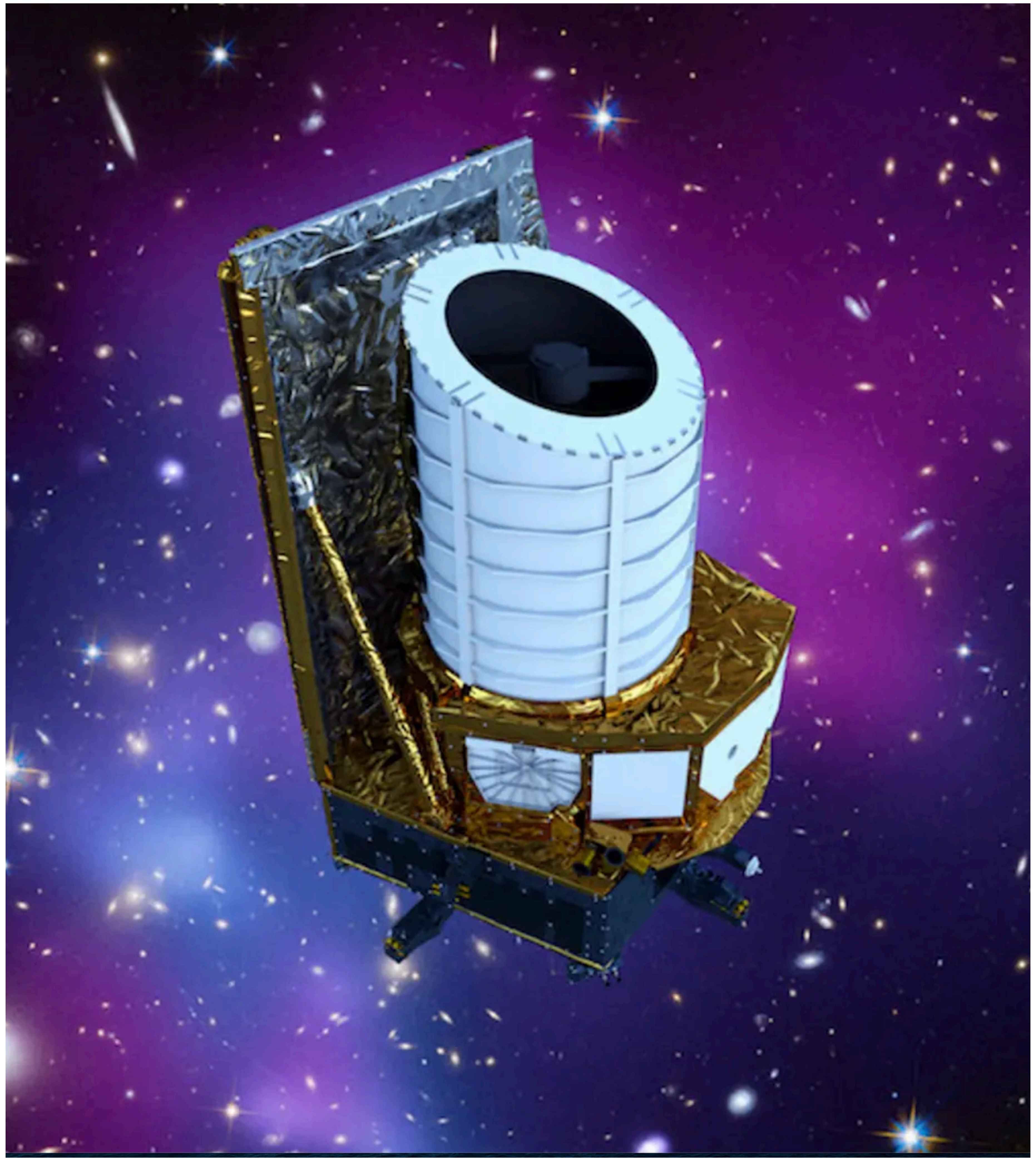
I. Why we need to go **beyond MCMC**

II. Our new approach: **Marginal Neural Ratio Estimation**

III. Applying MNRE to **Euclid** observables

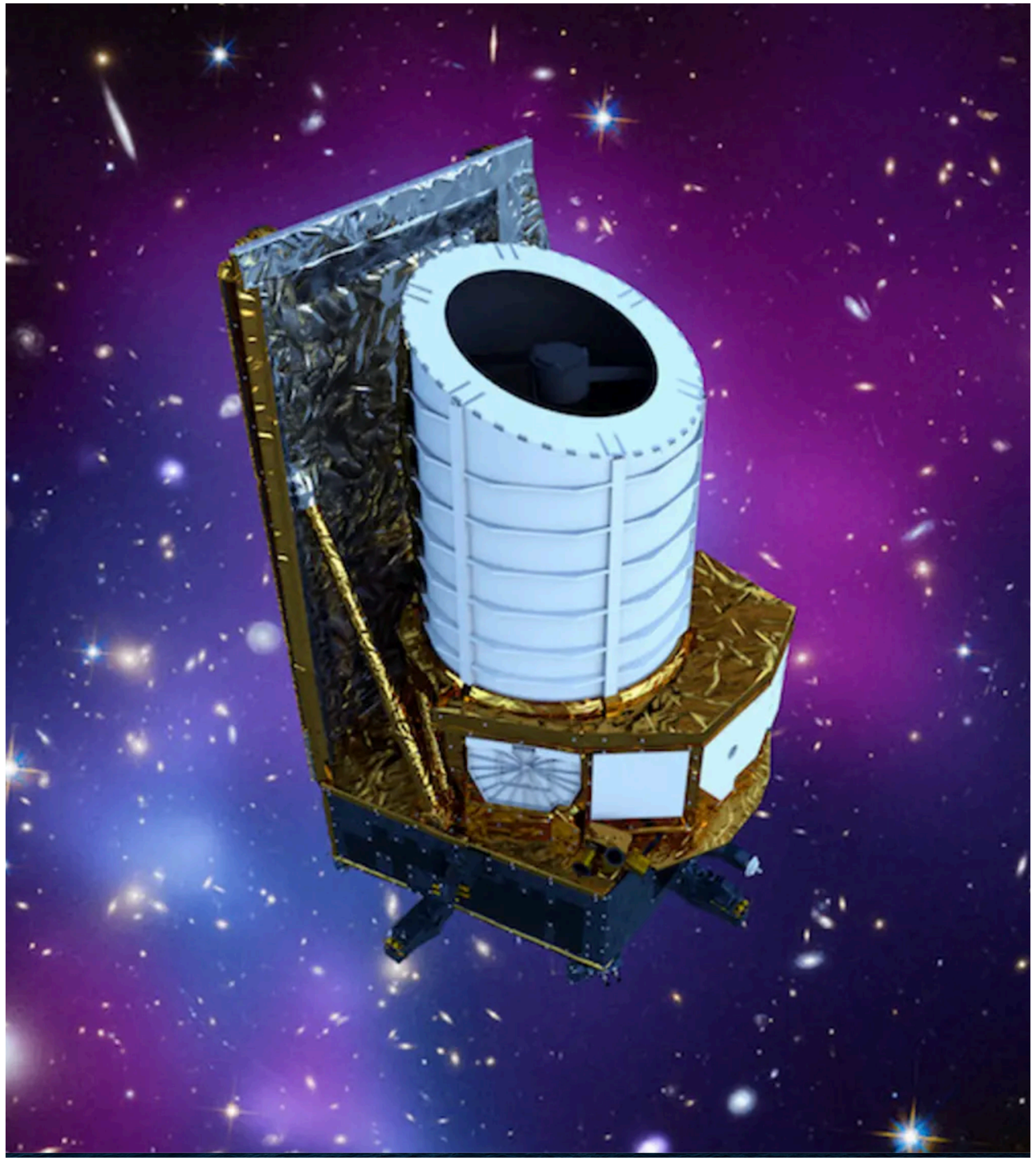
- On July 1, Euclid was launched to L2

[ESA’s Euclid space satellite]



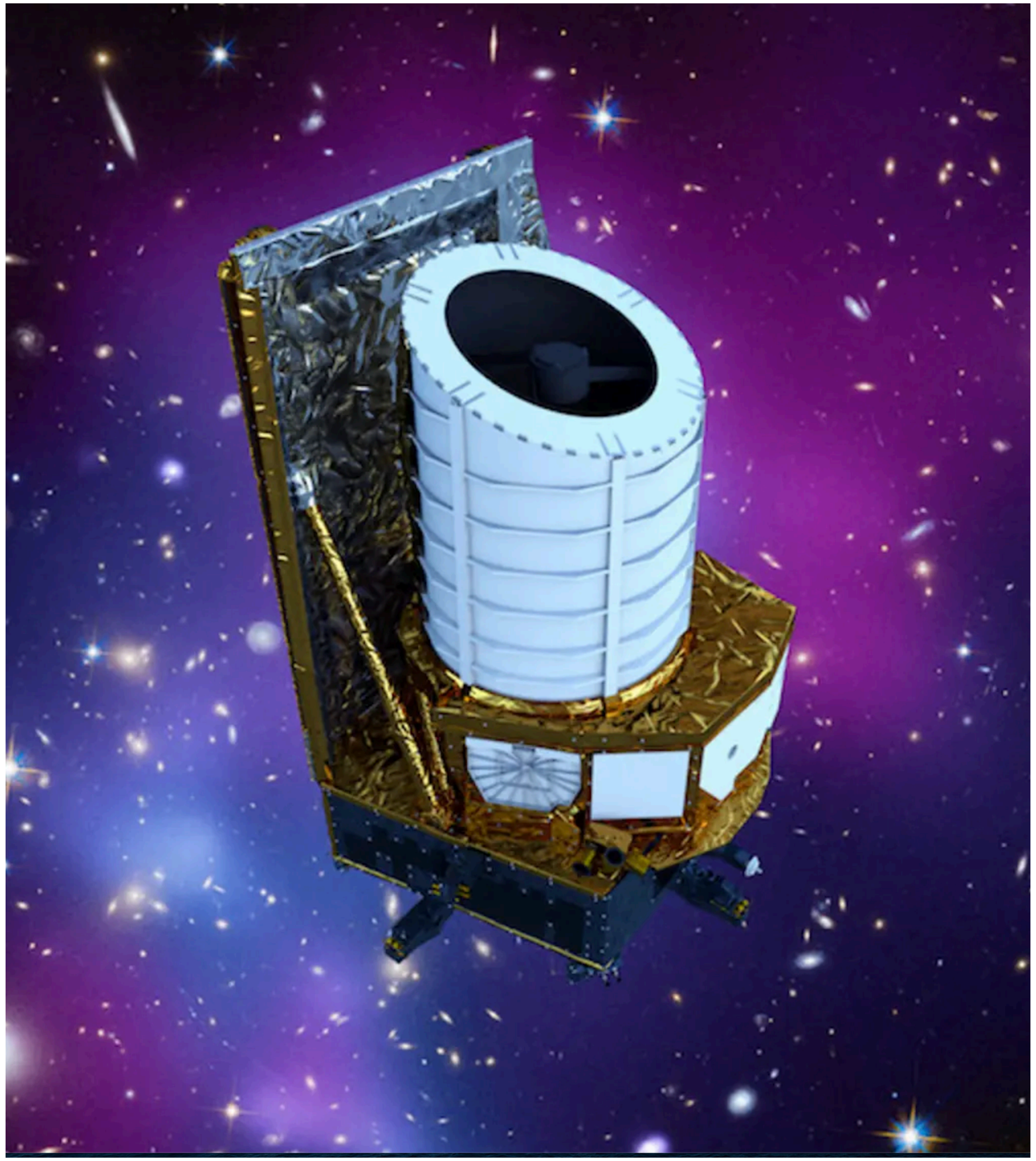
- On July 1, Euclid was launched to L2
- Over the next 6 years, Euclid will measure the shapes and redshifts of billions of galaxies, across ~1/3 of the sky

[ESA's Euclid space satellite]

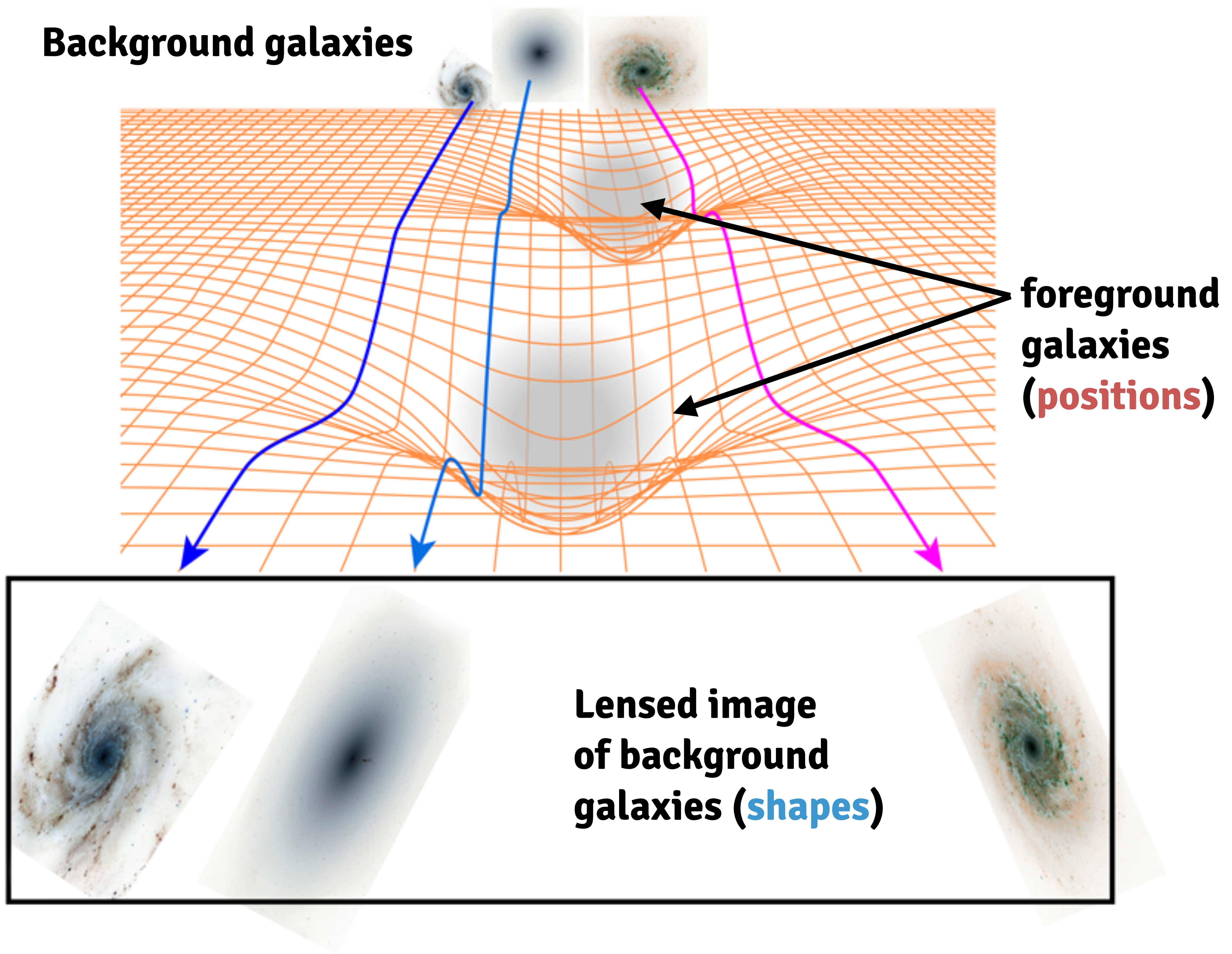


- On July 1, Euclid was launched to L2
- Over the next 6 years, Euclid will measure the shapes and redshifts of billions of galaxies, across ~1/3 of the sky
- First public data expected in 2025

[ESA's Euclid space satellite]



Which are the Euclid primary observables?

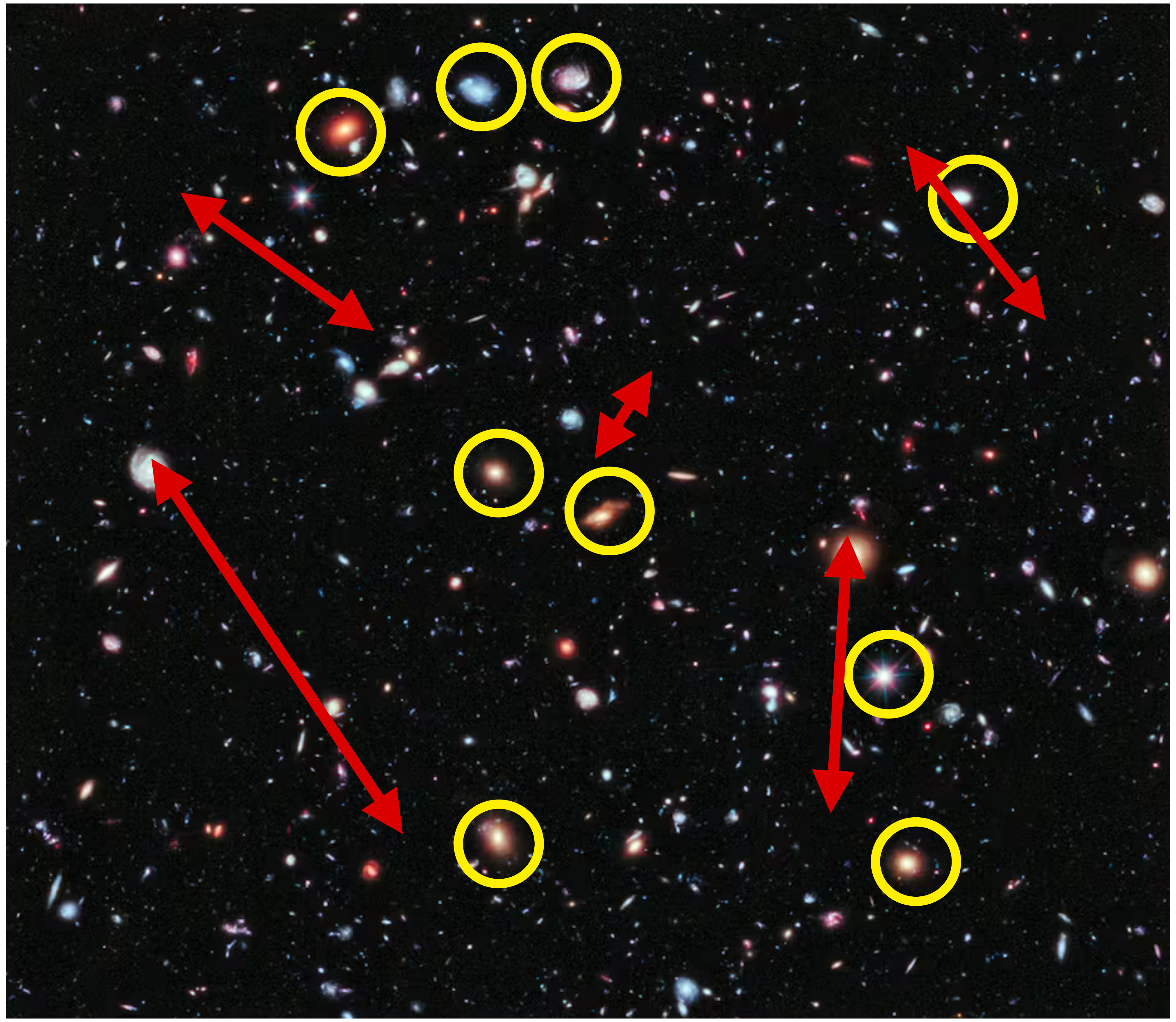


Which are the Euclid primary observables?

Which are the Euclid primary observables?

2-point function

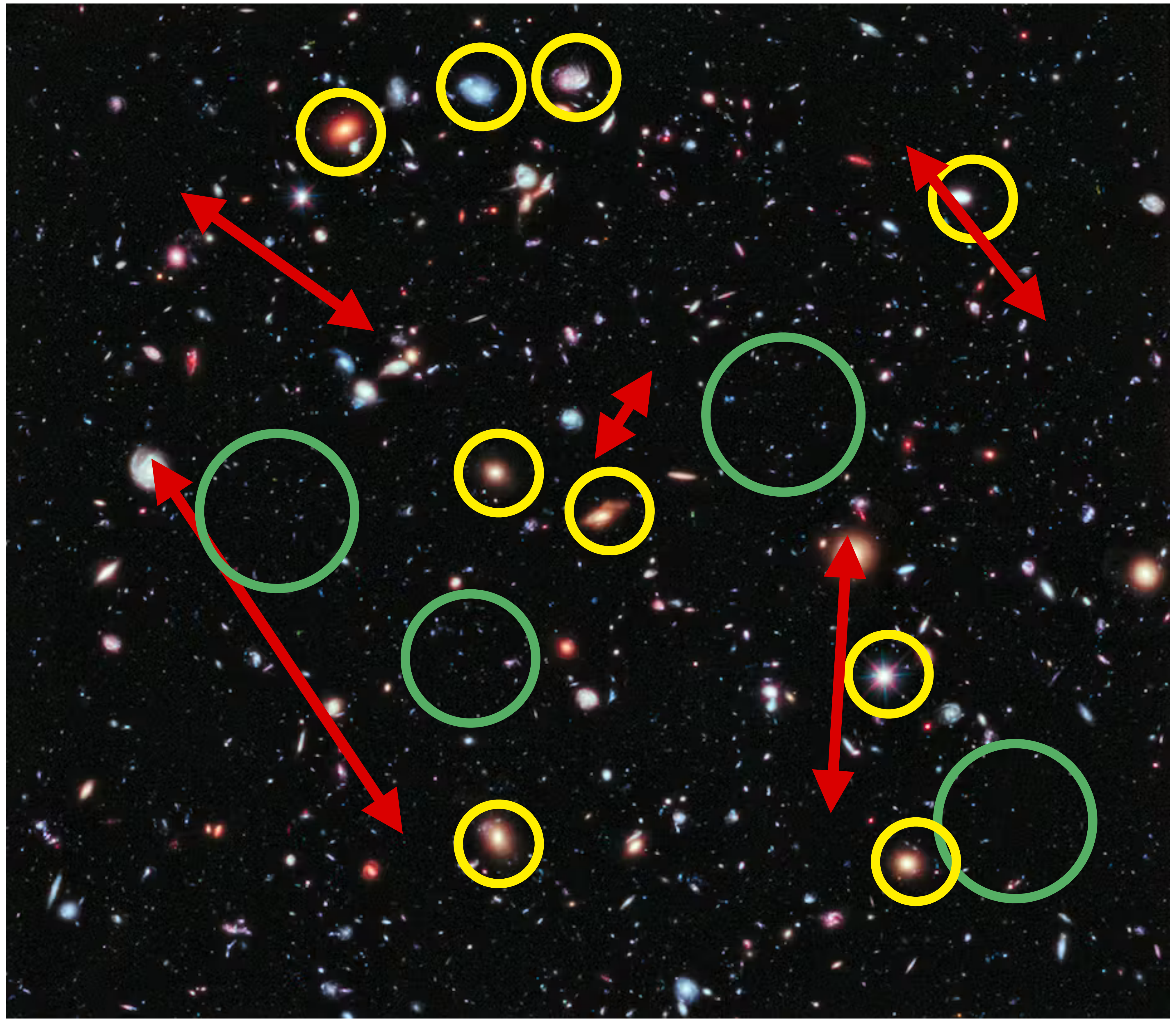
Which are the Euclid primary observables?



2-point function

Halo mass function

Which are the Euclid primary observables?

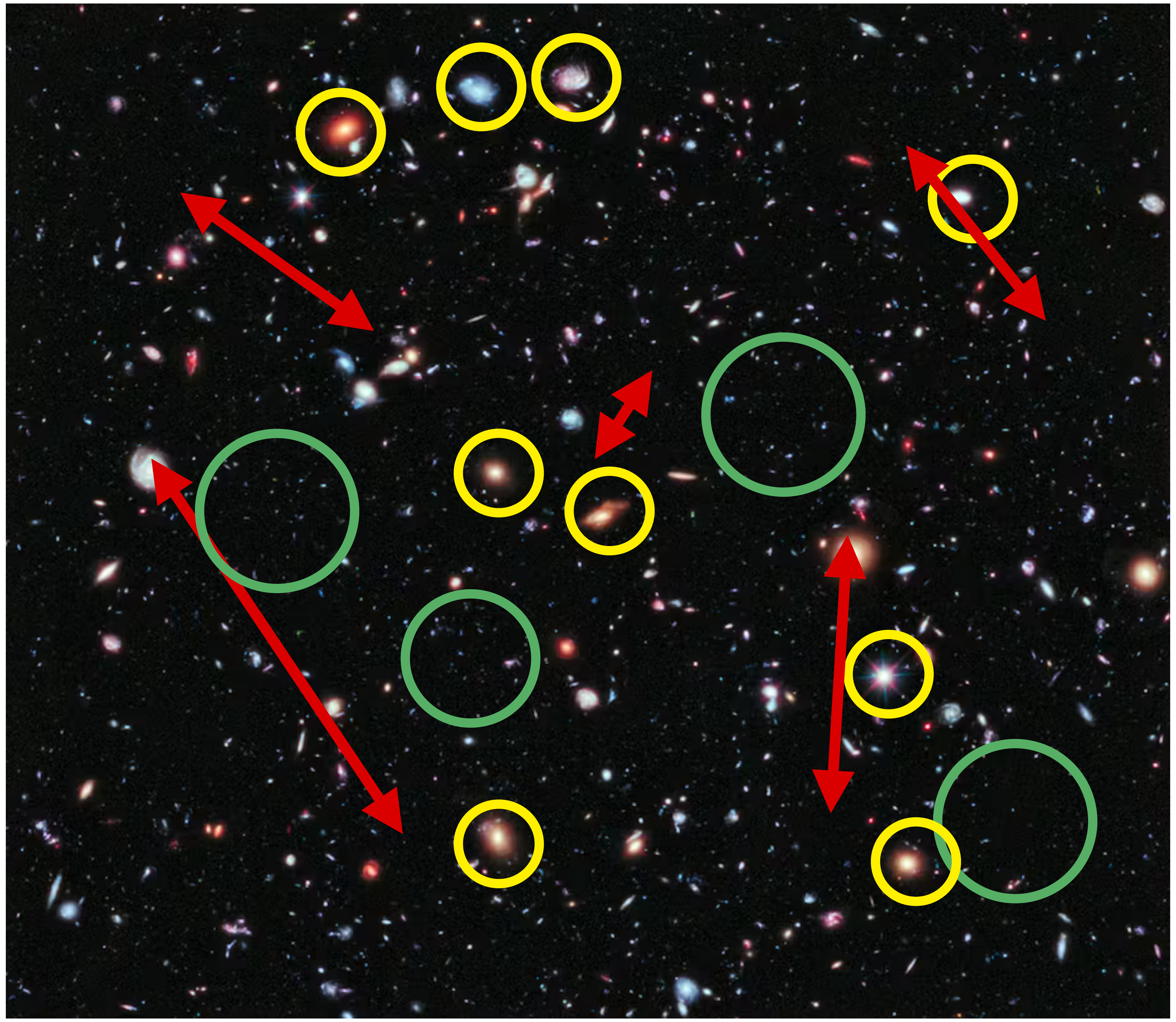


2-point function

Halo mass function

Void Size function

Which are the Euclid primary observables?



2-point function ✓

Halo mass function

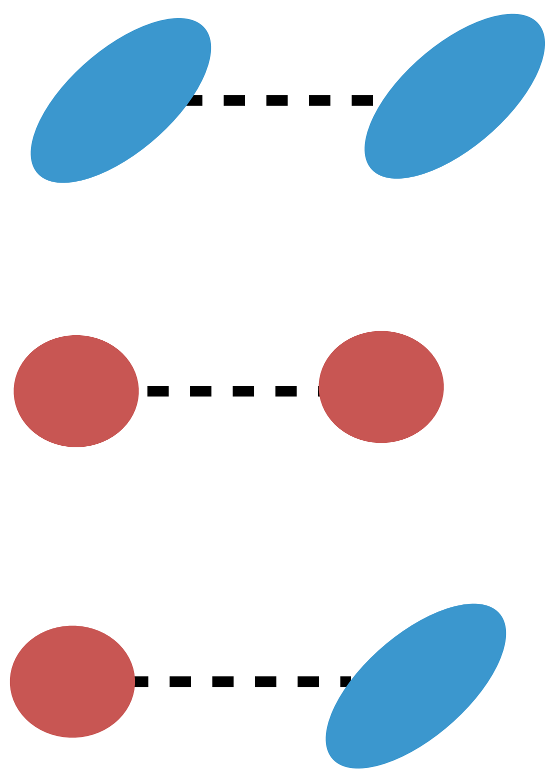
Void Size function

Summarise maps of positions/shapes
using three 2-point statistics (**3x2pt**):

Cosmic Shear

Galaxy clustering

Galaxy-Galaxy lensing

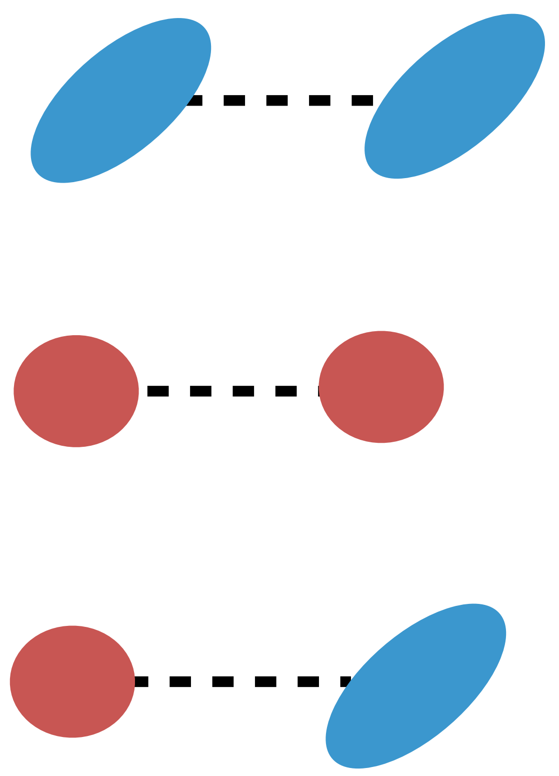


Summarise maps of positions/shapes
using three 2-point statistics **(3x2pt)**:

Cosmic Shear

Galaxy clustering

Galaxy-Galaxy lensing



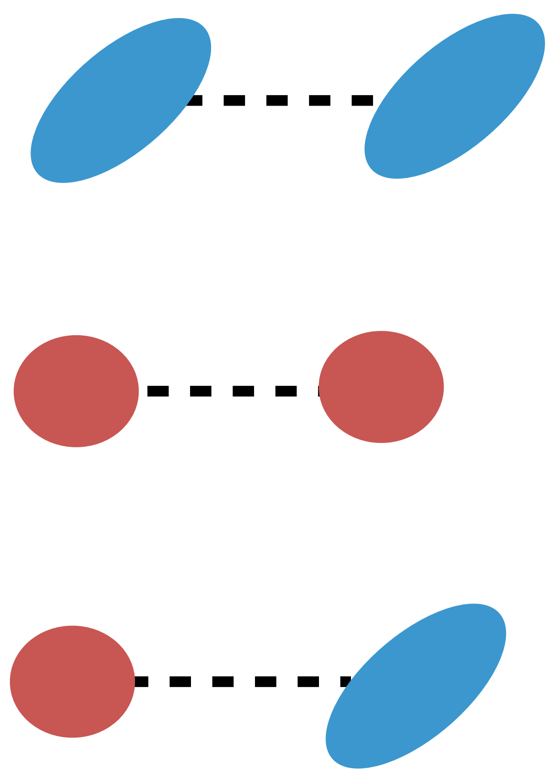
... measured for different
tomographic redshift bins

Summarise maps of positions/shapes
using three 2-point statistics **(3x2pt)**:

Cosmic Shear

Galaxy clustering

Galaxy-Galaxy lensing



... measured for different
tomographic redshift bins

Note: We consider only **photometric redshifts,** but Euclid will also
create a spectroscopic survey

3x2pt statistics described by power spectra

$$C_{ij}^{XY}(\ell) = \int dz\ W_i^X(z) W_j^Y(z)\ P_m(k_\ell, z)$$

Window functions

Matter power spectrum

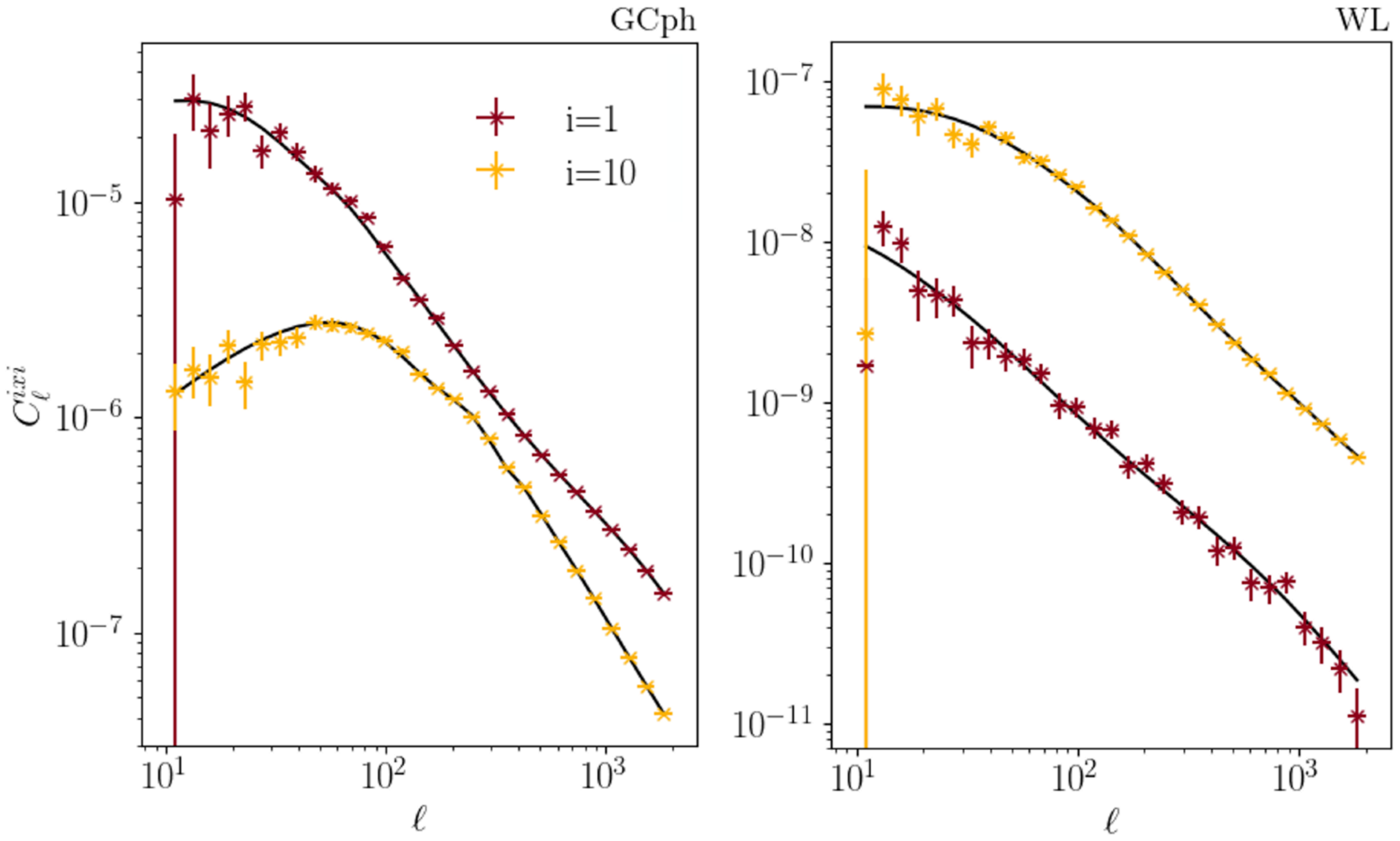
3x2pt statistics described by power spectra

$$C_{ij}^{XY}(\ell) = \int dz \; W_i^X(z) W_j^Y(z) \; P_m(k_\ell, z)$$

Window functions

Matter power spectrum

Ex:



3x2pt statistics described by power spectra

$$C_{ij}^{XY}(\ell) = \int dz \; W_i^X(z) W_j^Y(z) \; P_m(k_\ell, z)$$

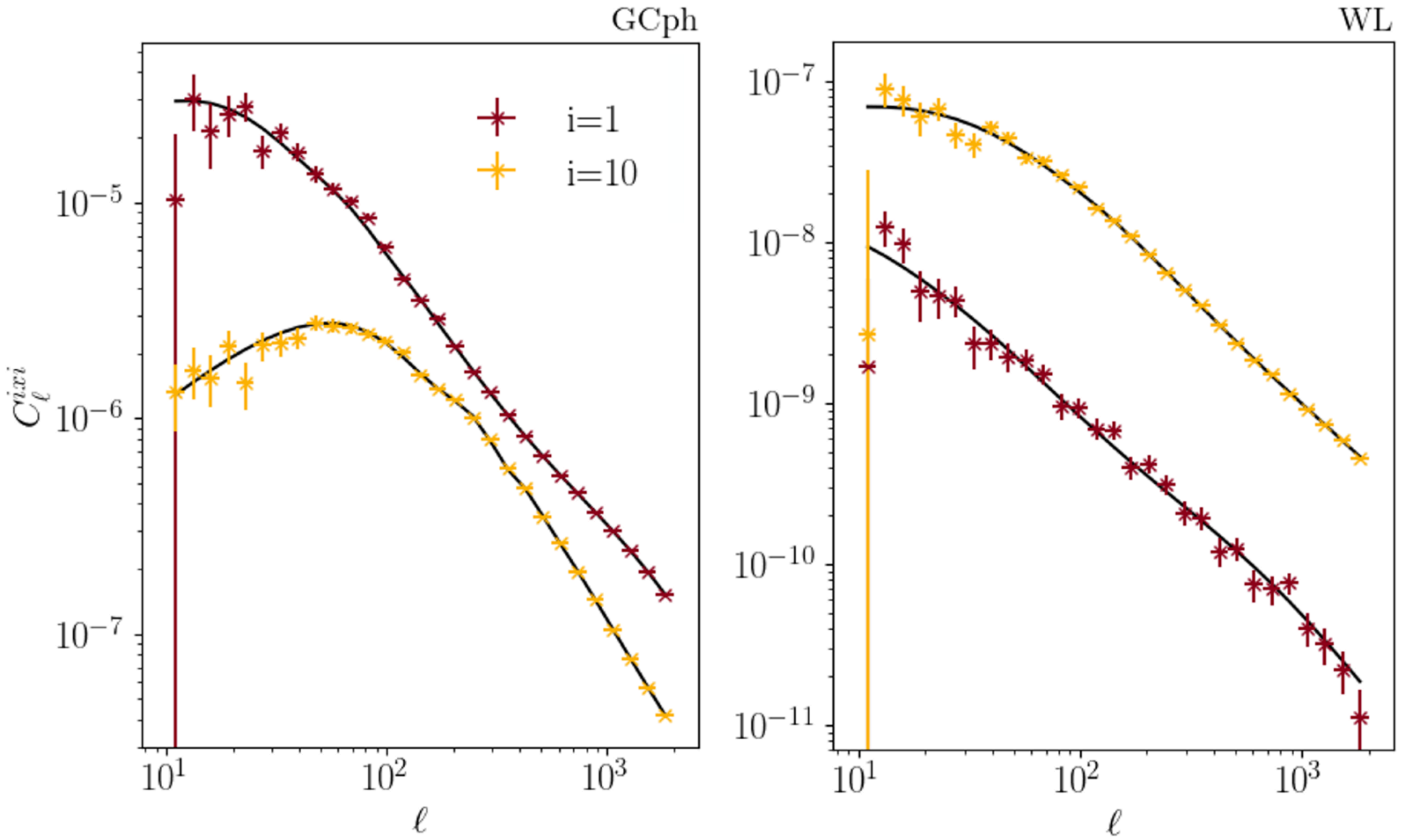
Window functions

Matter power spectrum

For 10 redshift bins up to z = 3

→ **+200 independent spectra!**

Ex:



Two main ingredients:

Two main ingredients:

- **Simulator of 3x2pt statistics**, based on a simplified Euclid likelihood (gaussian, **7 nuisance params**)

Two main ingredients:

- **Simulator of 3x2pt statistics**, based on a simplified Euclid likelihood (gaussian, **7 nuisance params**)

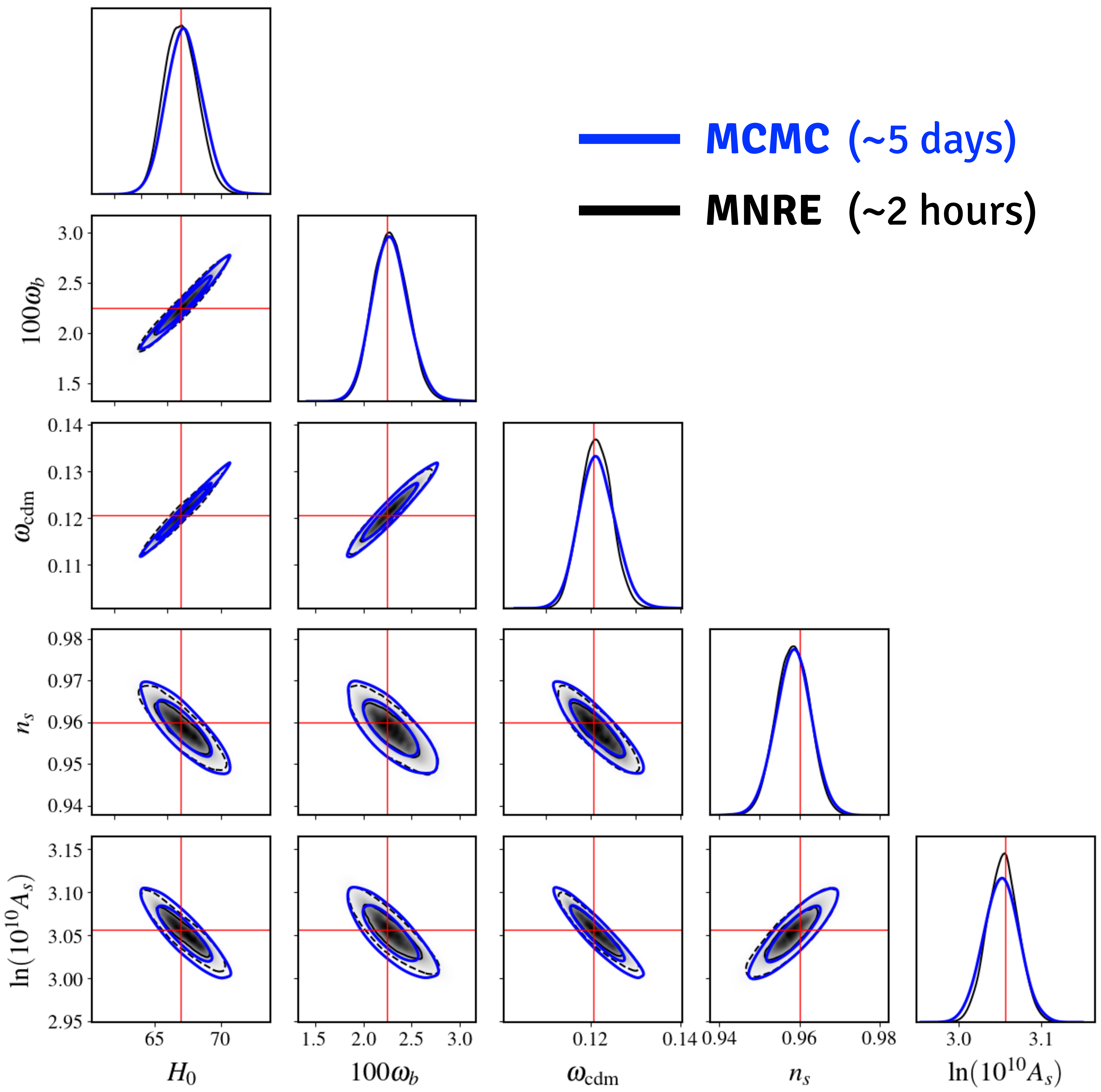
- **Network: ** Linear map that **compresses** all spectra into a few features

Results

Mock data analysis
on **ΛCDM** model
(5 cosmo params)

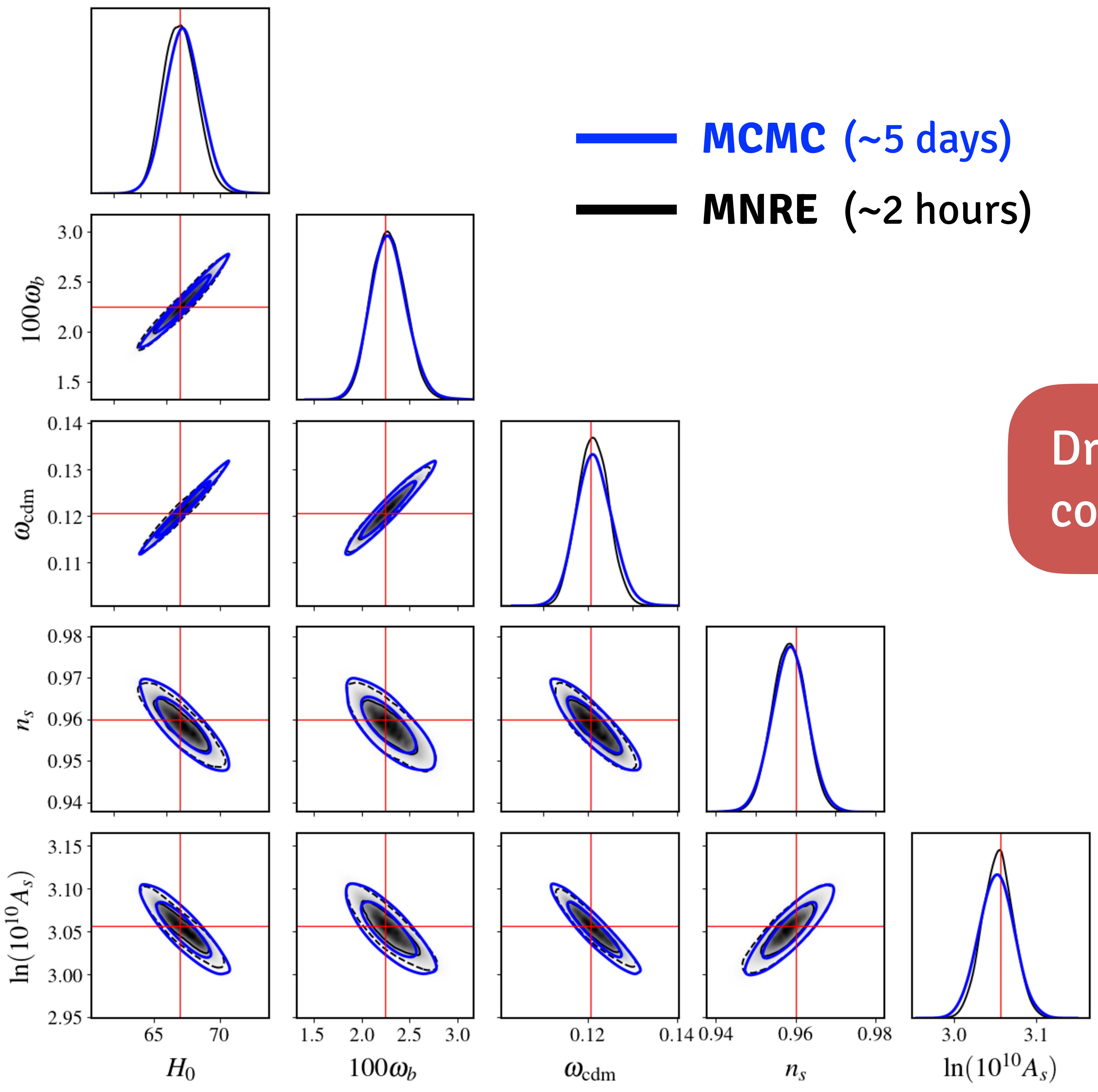
Results

Mock data analysis on **ΛCDM** model (5 cosmo params)



Results

Mock data analysis on **ΛCDM** model (5 cosmo params)



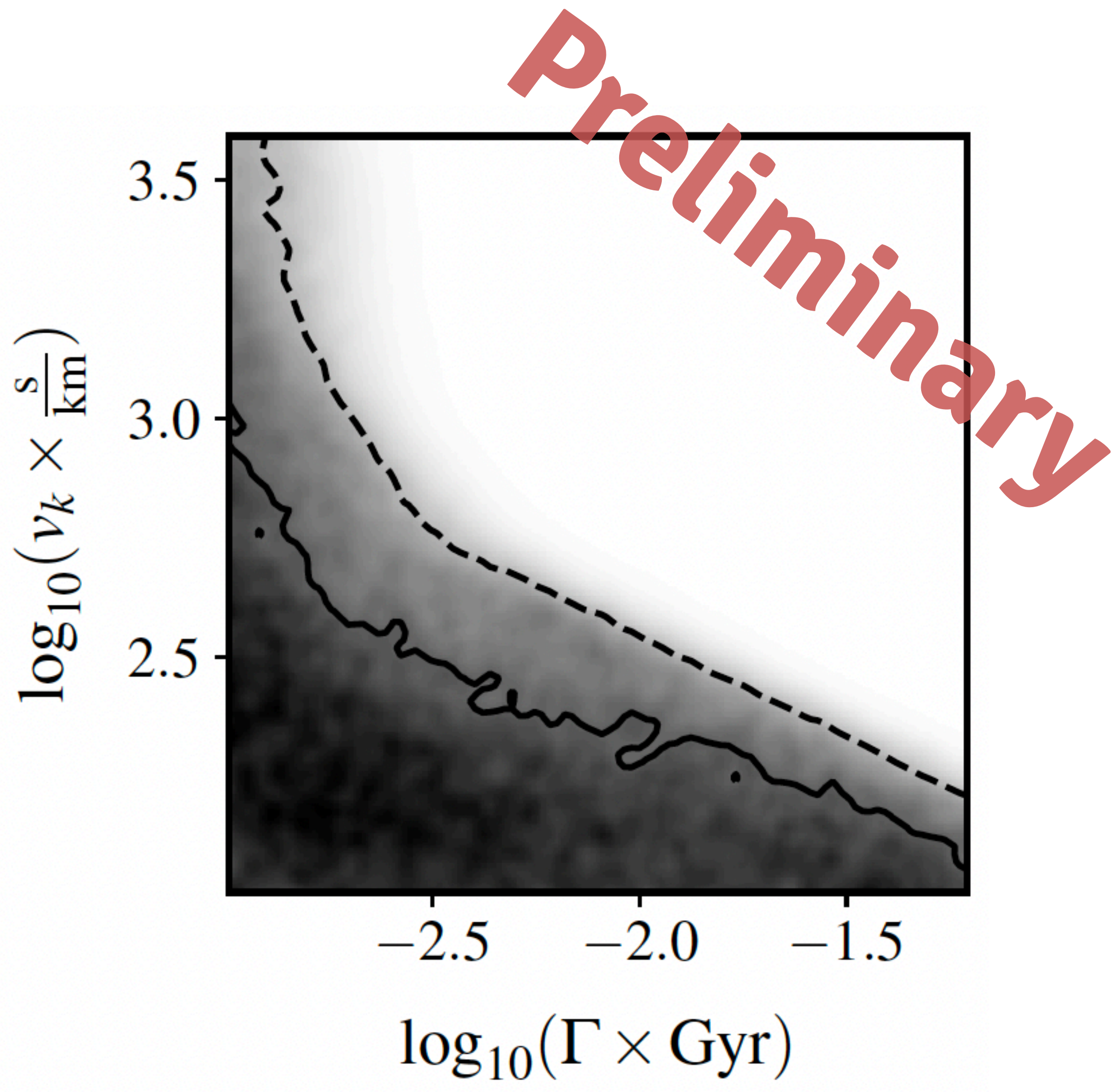
Dramatic reduction in computational time!

Next steps

- Use more realistic likelihood/simulator (many more nuisance pars.)

Next steps

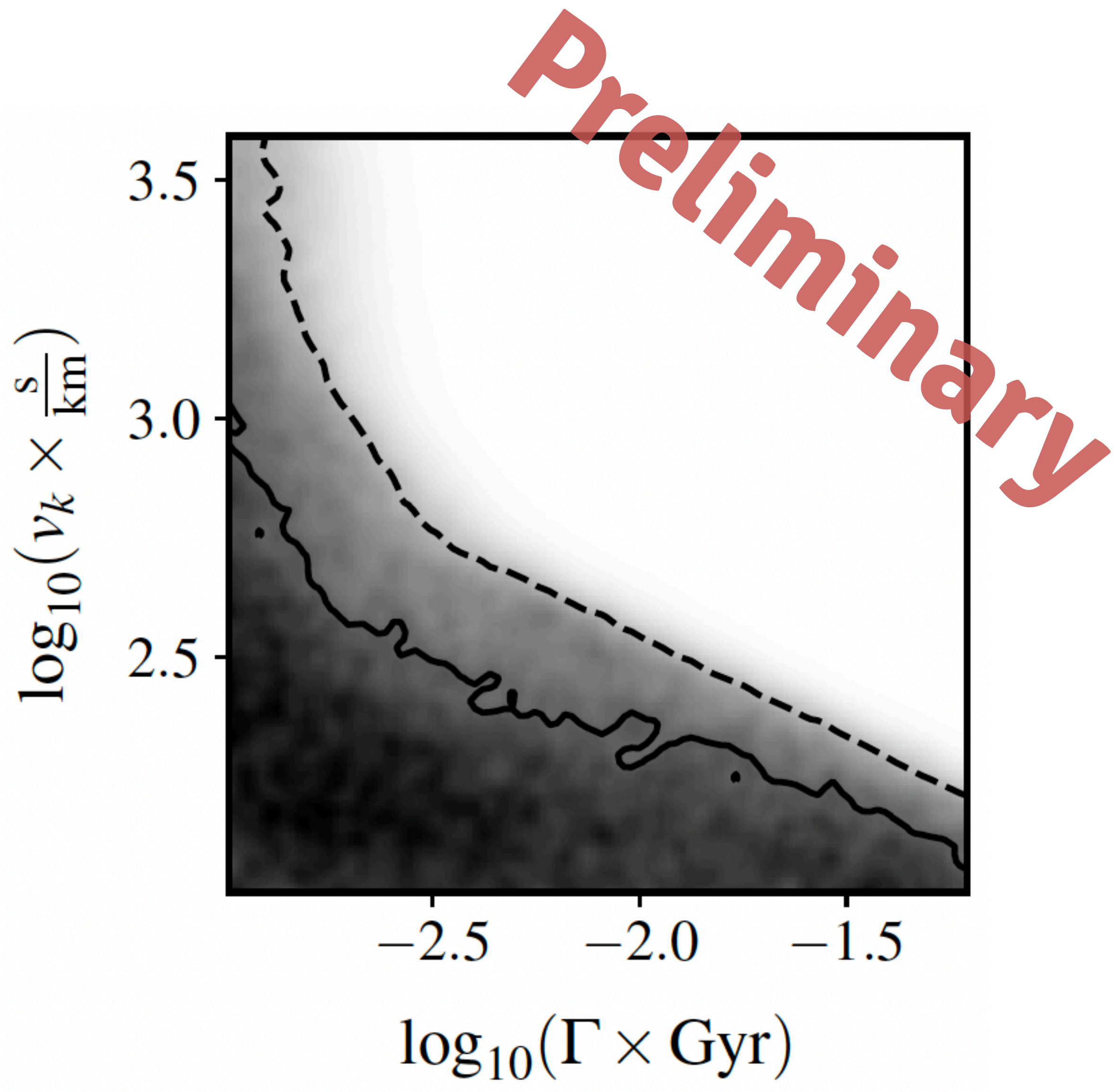
- Use more realistic likelihood/simulator (many more nuisance pars.)
- Consider various ΛCDM extensions (i.e. decaying DM)



Next steps

- Use more realistic likelihood/simulator (many more nuisance pars.)

- Consider various ΛCDM extensions (i.e. decaying DM)



- Apply MNRE to a very wide variety of cosmic data

Conclusions

- To learn as much as we can about the dark sector from future data, we need to go beyond traditional methods such as MCMC

- Using MNRE, we can analyse Euclid data (and potentially any other cosmic data) in a much more efficient and flexible way than MCMC

Euclid's view of the Perseus cluster of galaxies

Conclusions

- To learn as much as we can about the dark sector from future data, we need to go beyond traditional methods such as MCMC

- Using MNRE, we can analyse Euclid data (and potentially any other cosmic data) in a much more efficient and flexible way than MCMC

THANKS FOR YOUR ATTENTION

g.francoabellan@uva.nl

BACK-UP

Another big advantage of MNRE: simulation re-use

It is interesting to see how constraints change with different data combinations

It is interesting to see how constraints change with different data combinations

BUT

with MCMC, one has to **restart chains** for each experimental configuration

It is interesting to see how constraints change with different data combinations

BUT

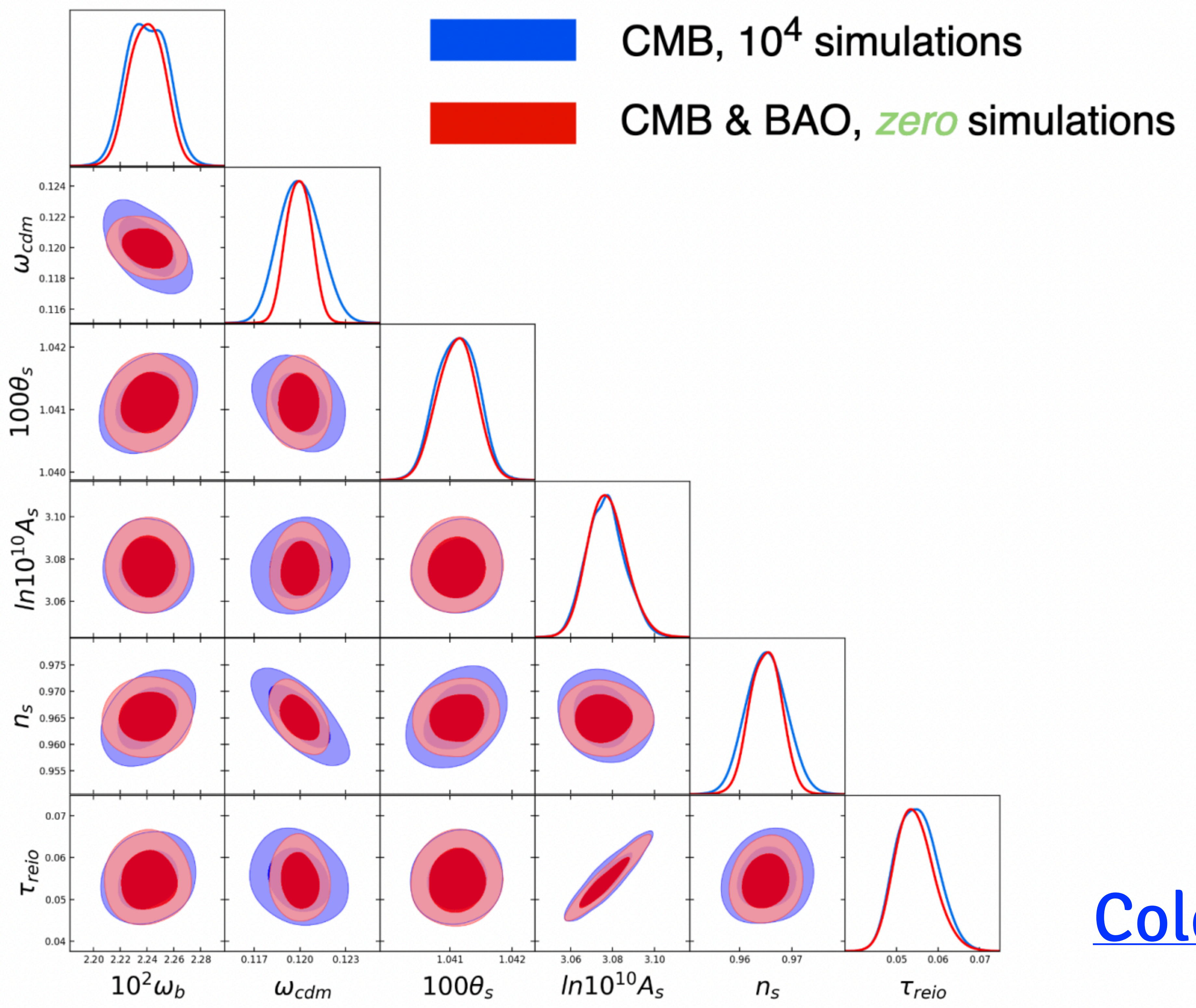
with MCMC, one has to **restart chains** for each experimental configuration

In MNRE it is possible to re-use simulations for different data combinations

The idea is to simulate all the data at once, and then train different inference networks for different data combinations

The idea is to simulate all the data at once, and then train different inference networks for different data combinations

Ex: Planck+BAO



Cole+ 22