Accelerating cosmological inference from Euclid
with Marginal Neural Ratio Estimation

Guillermo Franco Abellán

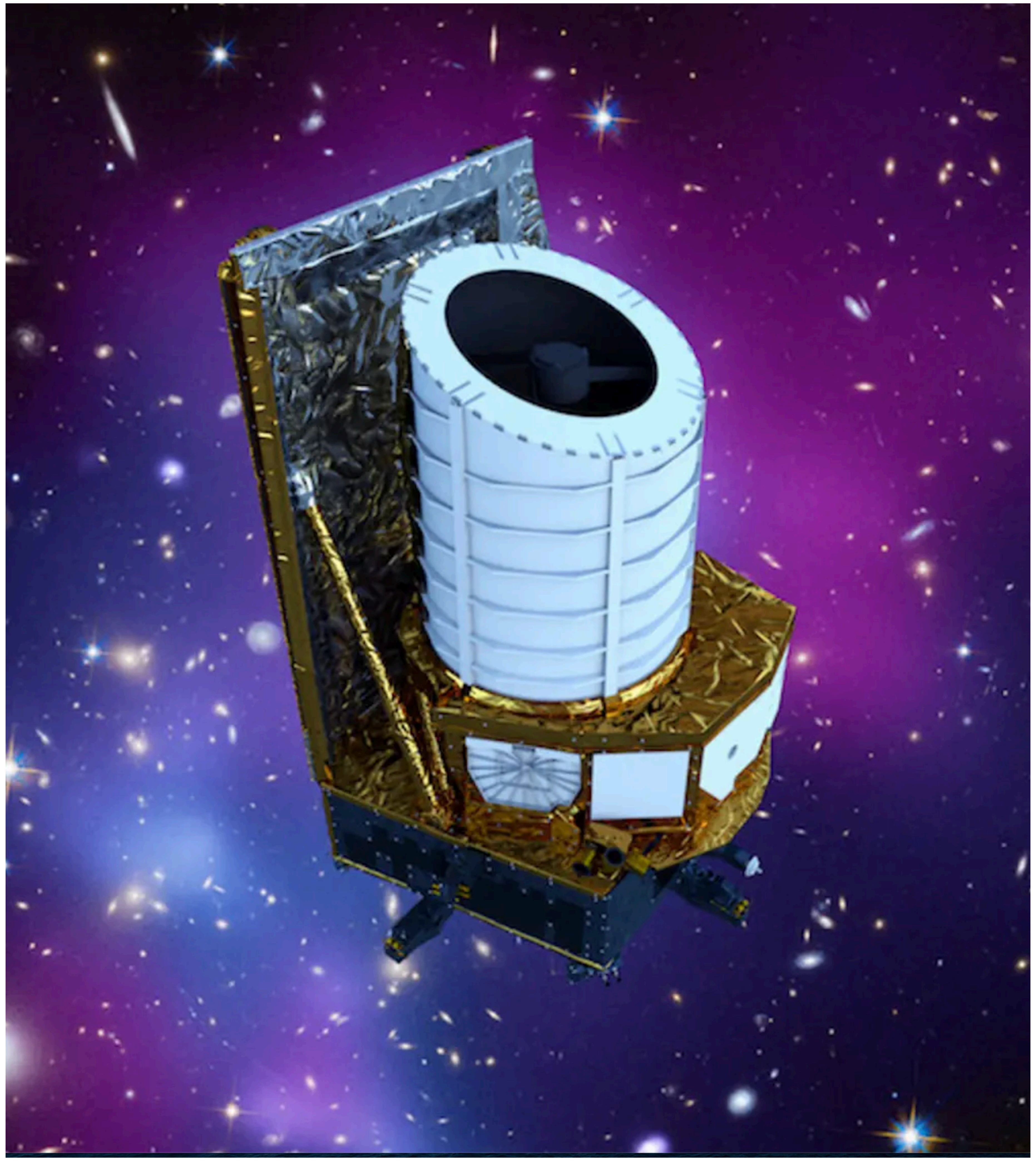
TUG - 11/10/2023

GRAPPA
GRavitation AstroParticle Physics Amsterdam

Euclid consortium

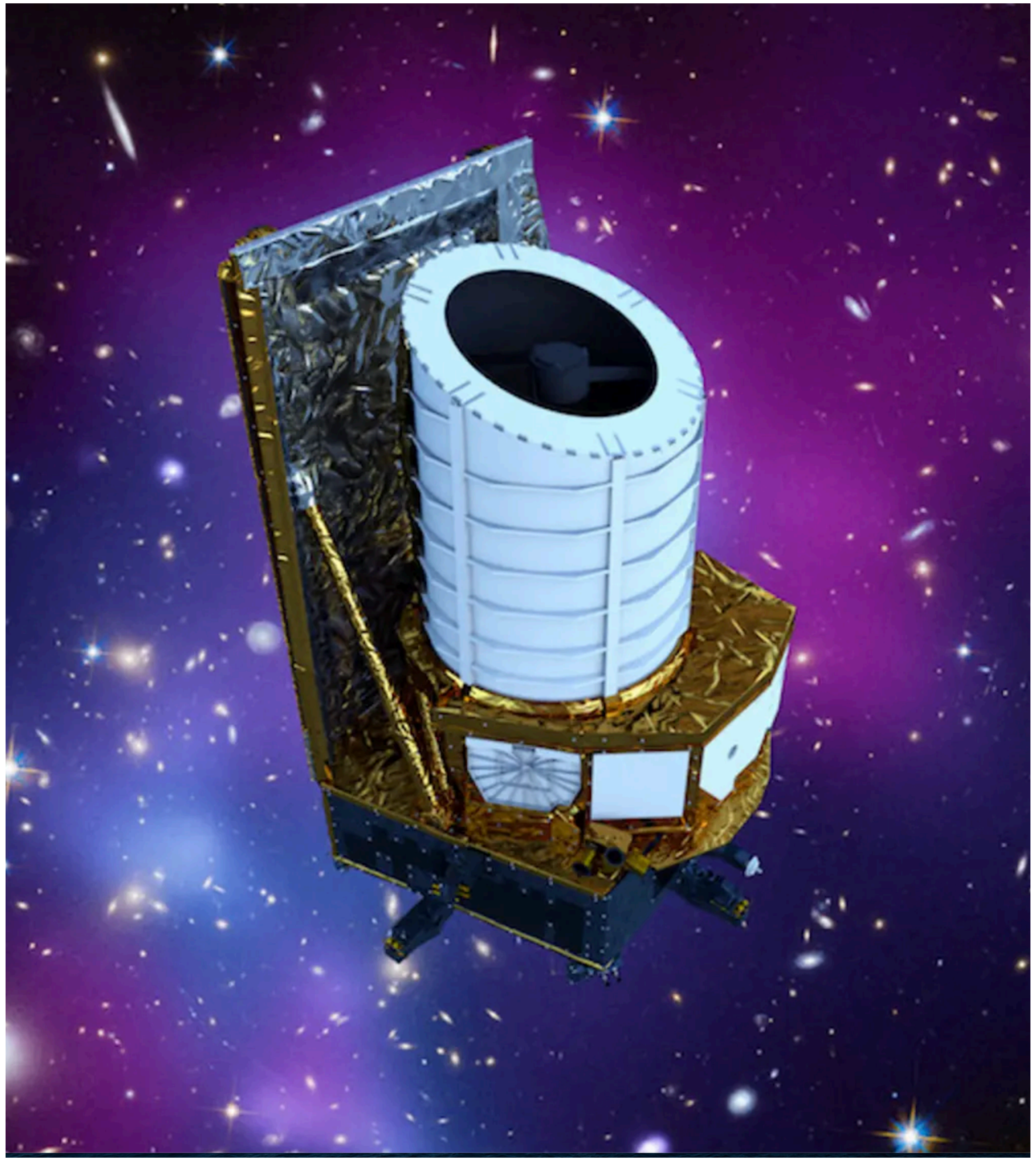
Ongoing work with
Guadalupe C. Herrera,
Matteo Martinelli,
Christoph Weniger,
& others

[ESA's Euclid space satellite]



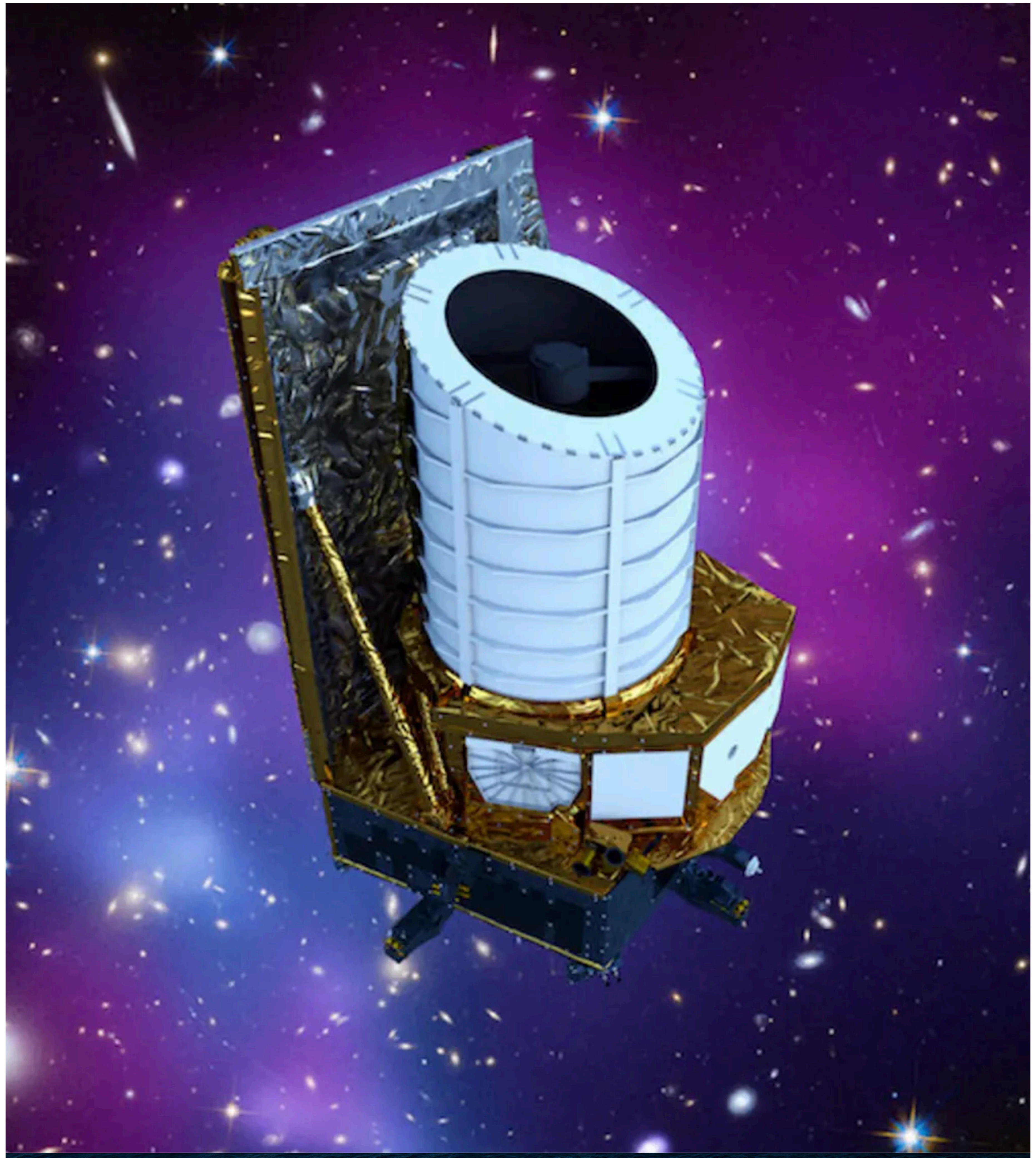
On July 1, Euclid was launched to L2

[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 release expected in 2025

Many more astronomical data to come...

esa gaia

WEBB SPACE TELESCOPE

SKA SQUARE KILOMETRE ARRAY

LiteBIRD

Einstein Telescope

DARK ENERGY SPECTROSCOPIC INSTRUMENT

VERA C. RUBIN OBSERVATORY

LSST Legacy Survey of Space and Time

CMB-S4 Next Generation CMB Experiment

lisa

Many more astronomical data to come…

Analysing these high-quality data will be challenging with standard methods

The curse of dimensionality

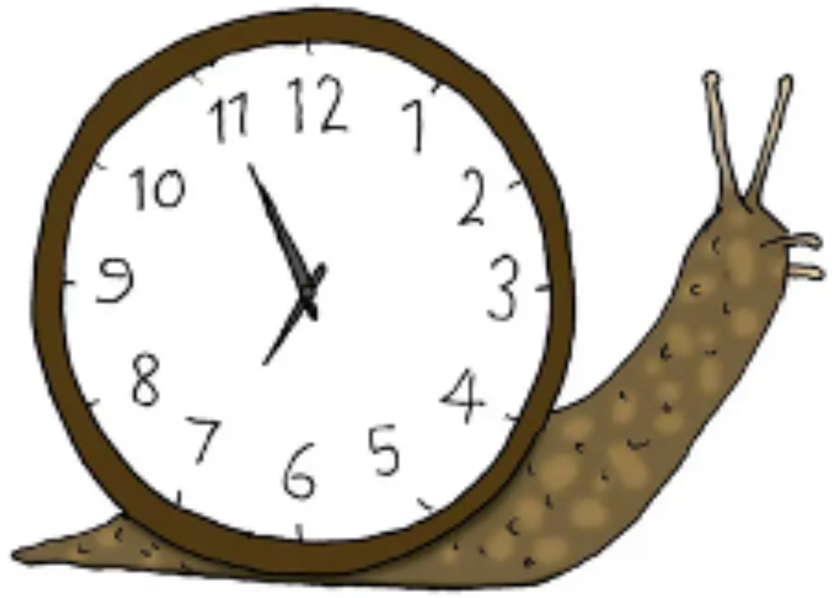
- Traditional likelihood-based methods (MCMC, Nested Sampling,...)
  → compute joint posterior and then marginalise

The curse of dimensionality

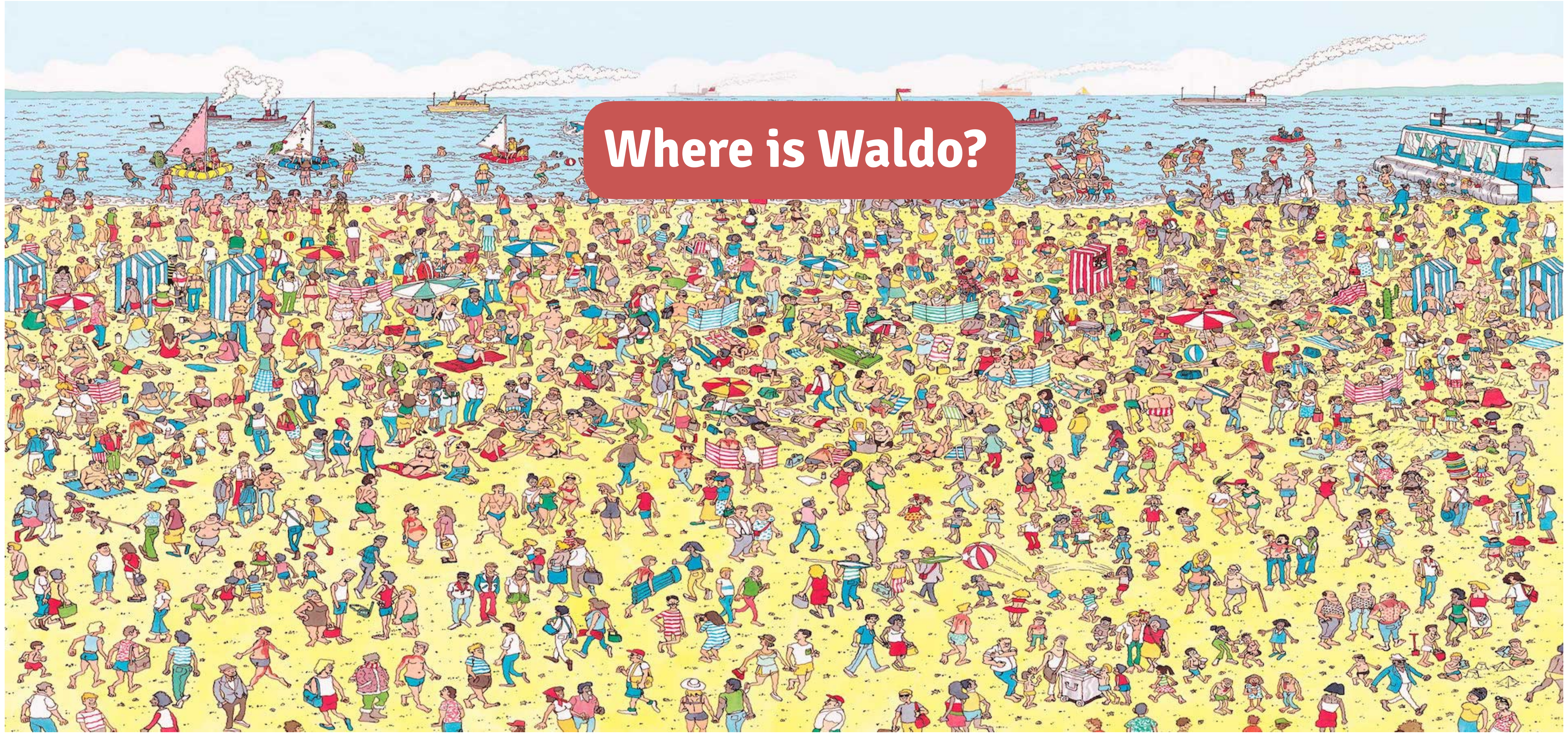
- Traditional likelihood-based methods (MCMC, Nested Sampling,...)
  ⟶ compute joint posterior and then marginalise

- Scale poorly with dimensionality of parameter space

Ex: For Euclid, we expect to have **+50 nuisance parameters**



The curse of dimensionality



Marginal posterior **Joint posterior**

$$P(z_{\text{waldo}} \,|\, x_0) = \int dz_{\text{Pierre}} dz_{\text{Thomas}} dz_{\text{Julien}} \dots dz_{\text{Killian}} \; P(z_{\text{Waldo}}, z_{\text{Pierre}}, z_{\text{Thomas}}, z_{\text{Julien}}, \dots, z_{\text{Killian}} \,|\, x_0)$$

Are there methods to overcome this problem?

Are there methods to overcome this problem?

Can machine learning be helpful?

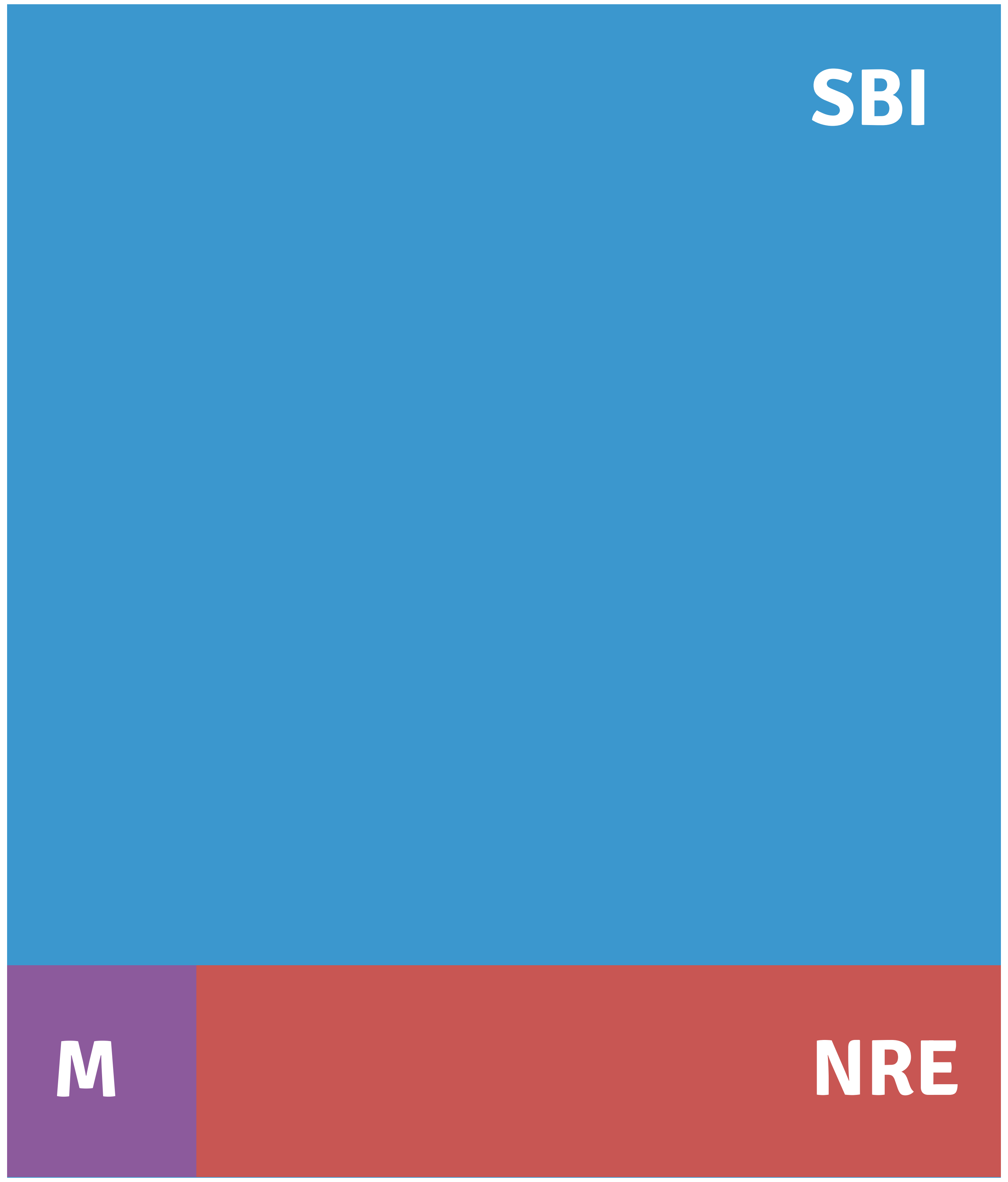
A NEW HOPE

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

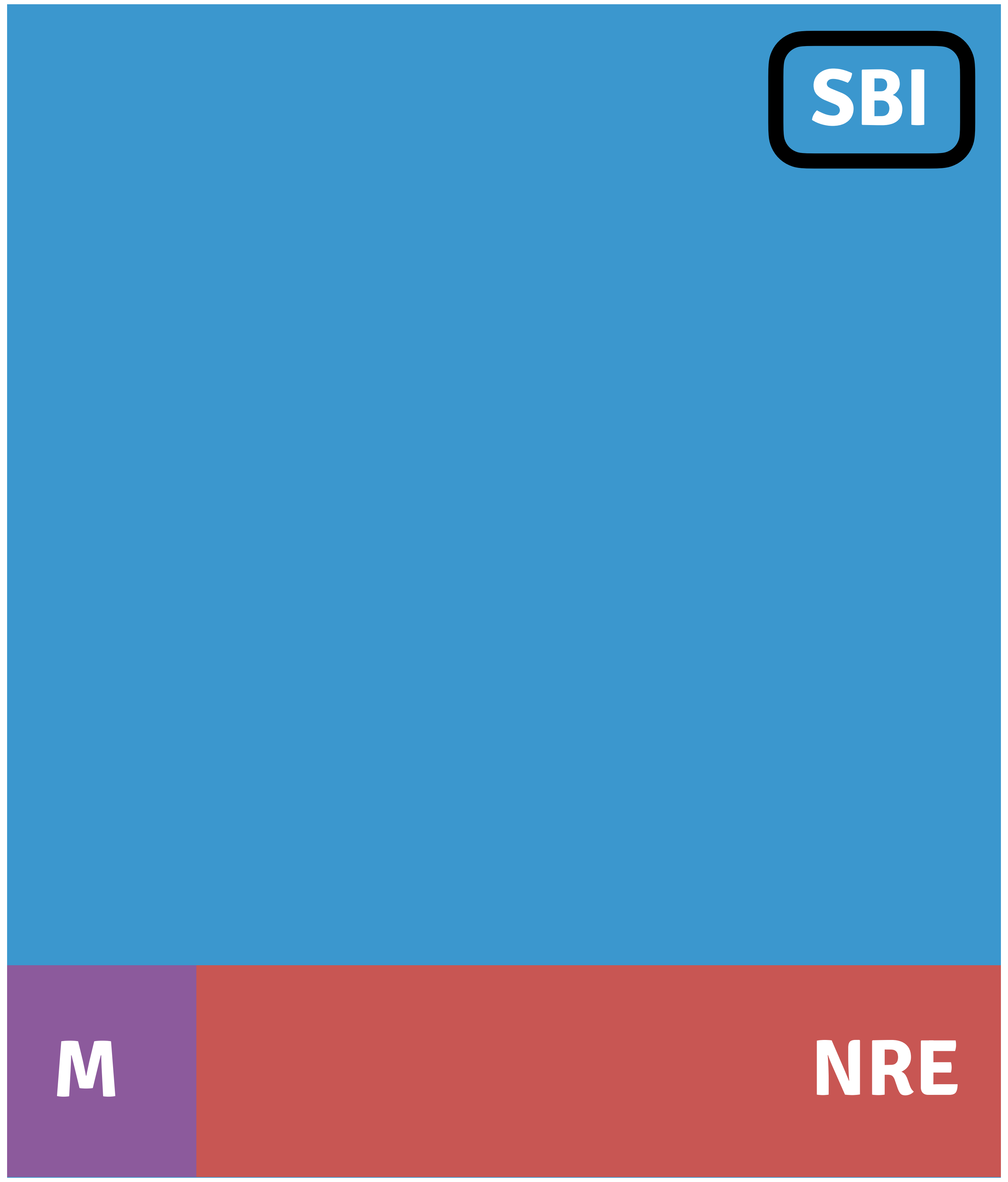
Implemented in Swyft* [Miller+ 20]

* Stop Wasting Your Precious Time

Marginal Neural Ratio Estimation

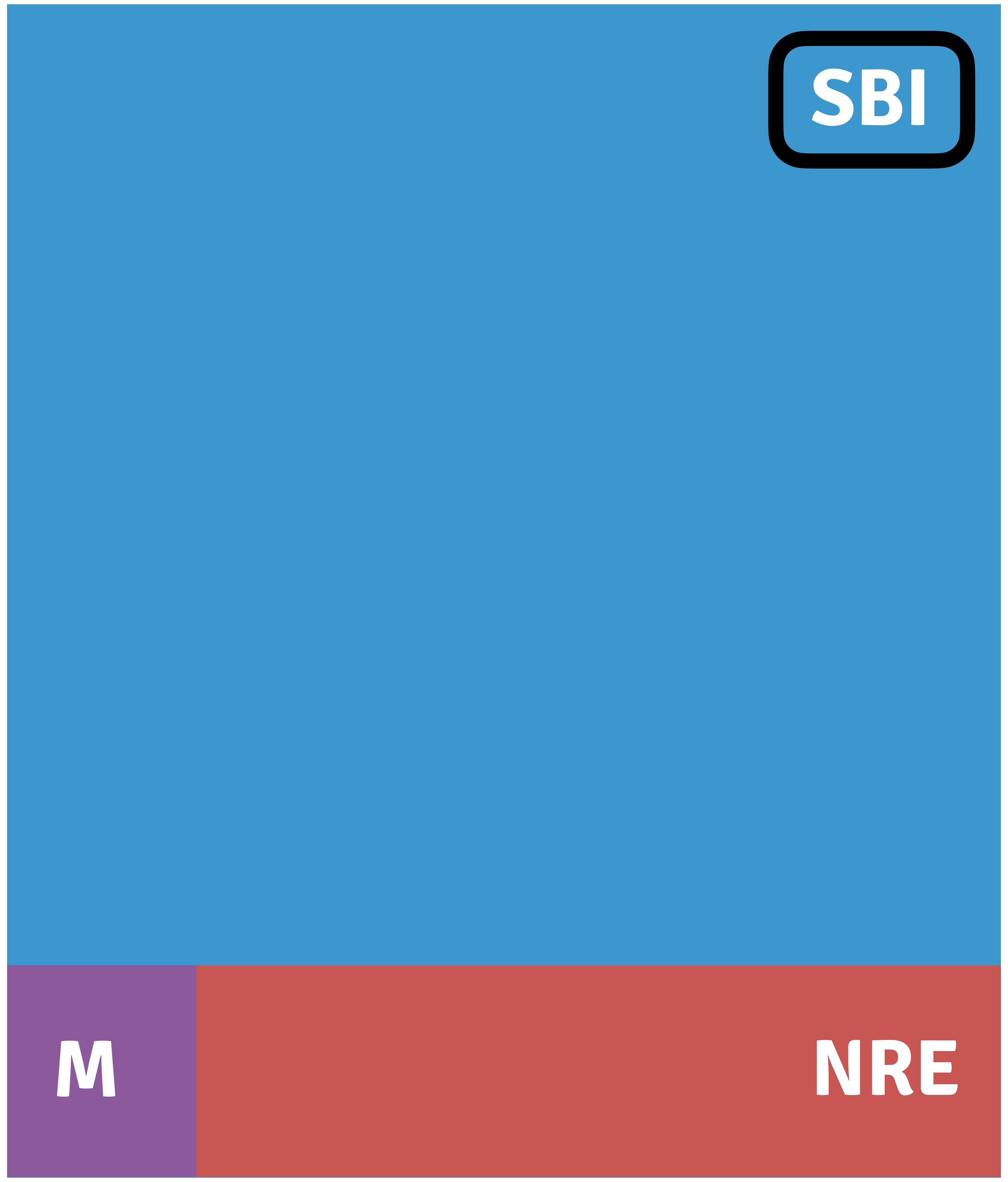


Marginal Neural Ratio Estimation



Simulation-based inference
(or likelihood-free inference)

Marginal Neural Ratio Estimation



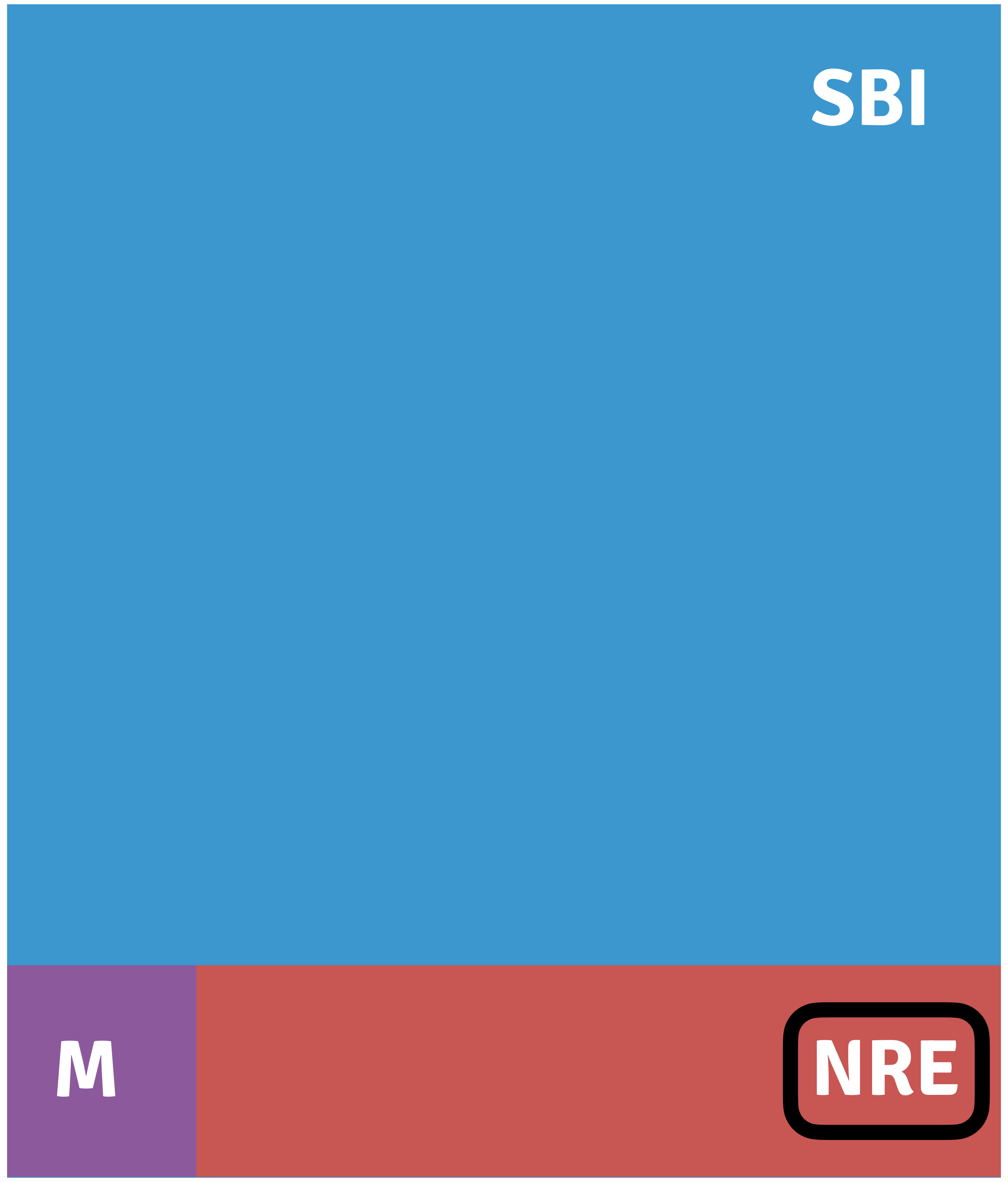
Simulation-based inference
(or likelihood-free inference)

↓

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

$$\mathbf{x} \sim p(\mathbf{x} \mid \mathbf{z})$$ (implicit likelihood)

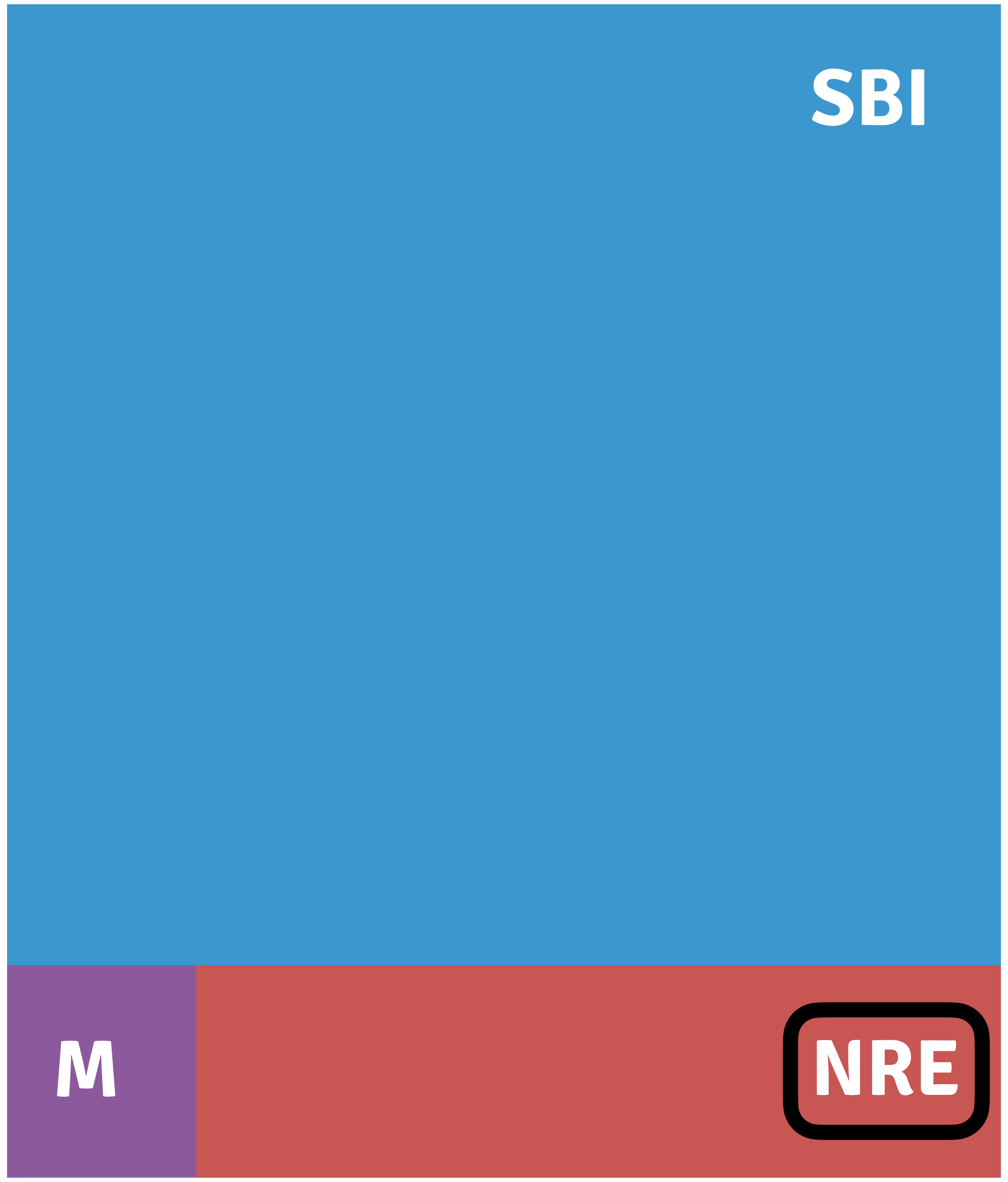
Marginal Neural Ratio Estimation



Neural Ratio Estimation

$$r(\mathbf{x};\mathbf{z}) = \frac{p(\mathbf{x}\,|\,\mathbf{z})}{p(\mathbf{x})} = \frac{p(\mathbf{z}\,|\,\mathbf{x})}{p(\mathbf{z})} = \frac{p(\mathbf{x},\mathbf{z})}{p(\mathbf{x})p(\mathbf{z})}$$

Marginal Neural Ratio Estimation



Neural Ratio Estimation

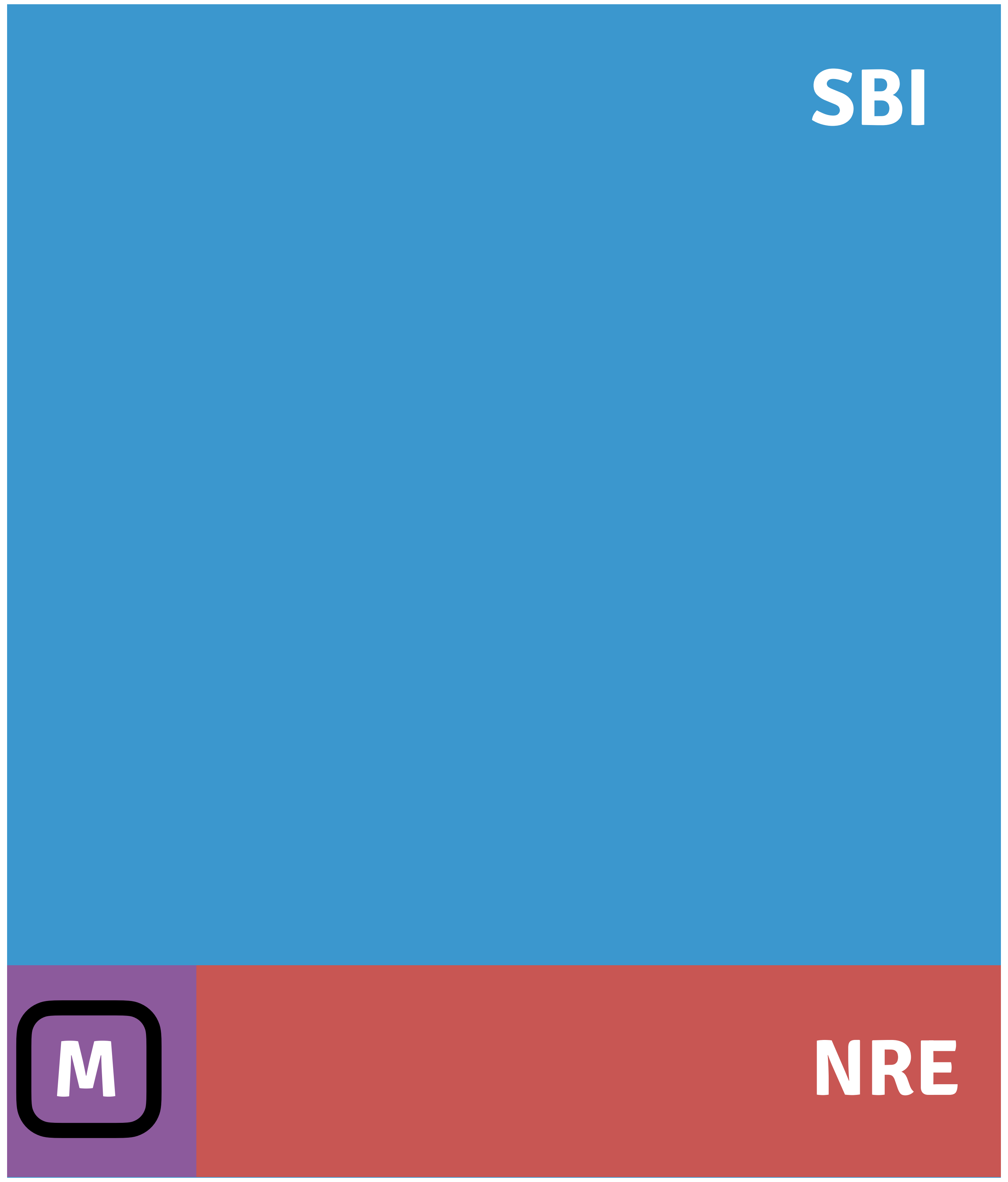
$$r(\mathbf{x}; \mathbf{z}) = \frac{p(\mathbf{x} \mid \mathbf{z})}{p(\mathbf{x})} = \frac{p(\mathbf{z} \mid \mathbf{x})}{p(\mathbf{z})} = \boxed{\frac{p(\mathbf{x}, \mathbf{z})}{p(\mathbf{x})p(\mathbf{z})}}$$

Rephrase inference as a binary classification problem, and solve it by training a NN on simulated data

$(\mathbf{x}, \mathbf{z}) \sim p(\mathbf{x}, \mathbf{z})$

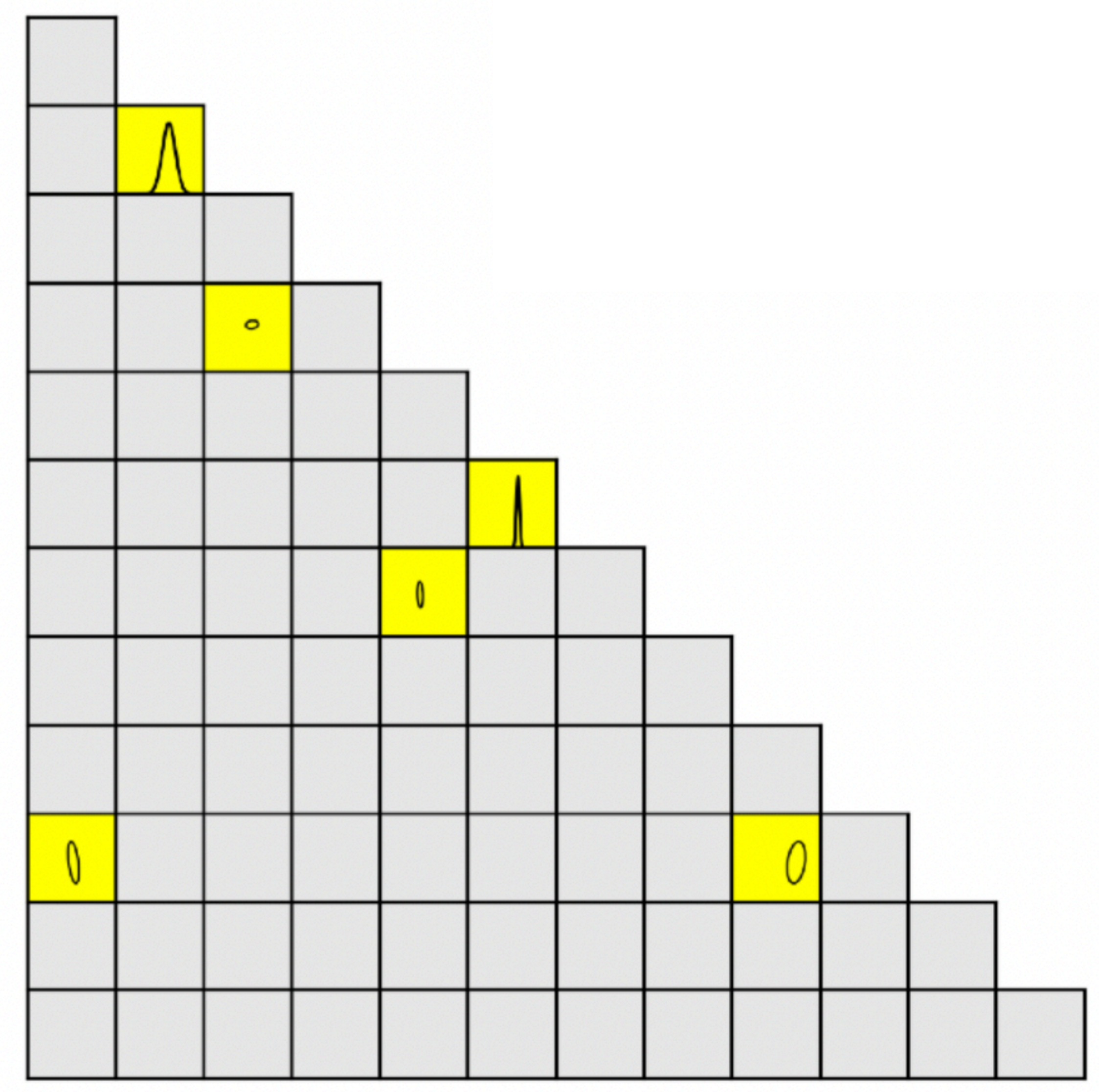
$(\mathbf{x}, \mathbf{z}) \sim p(\mathbf{x})p(\mathbf{z})$

Marginal Neural Ratio Estimation

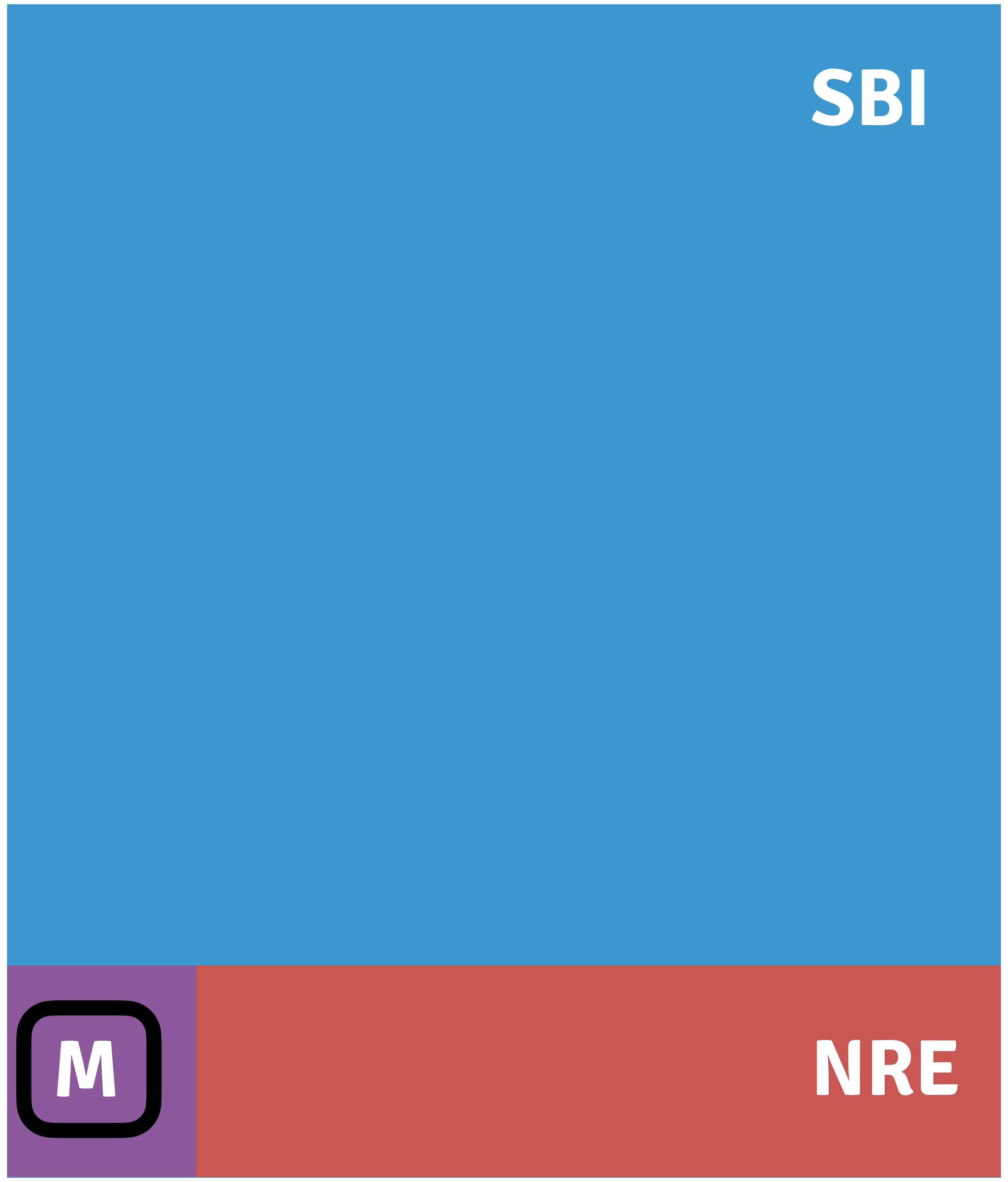


Marginal inference

Instead of estimating all parameters,
we can cherry-pick what we care about

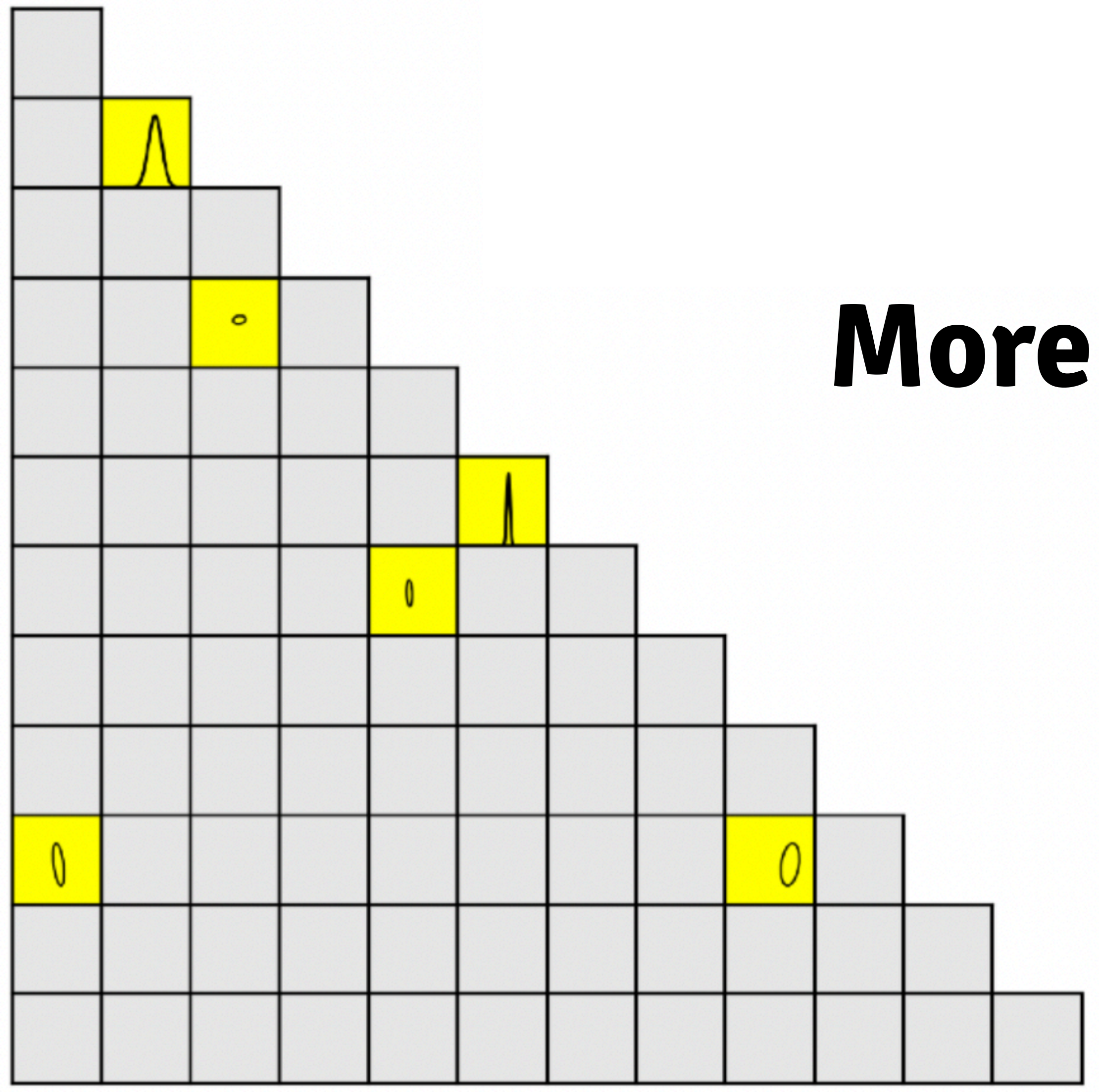


Marginal Neural Ratio Estimation



Marginal inference

Instead of estimating all parameters,
we can cherry-pick what we care about



More flexible & efficient

MNRE has been successfully applied in many contexts:

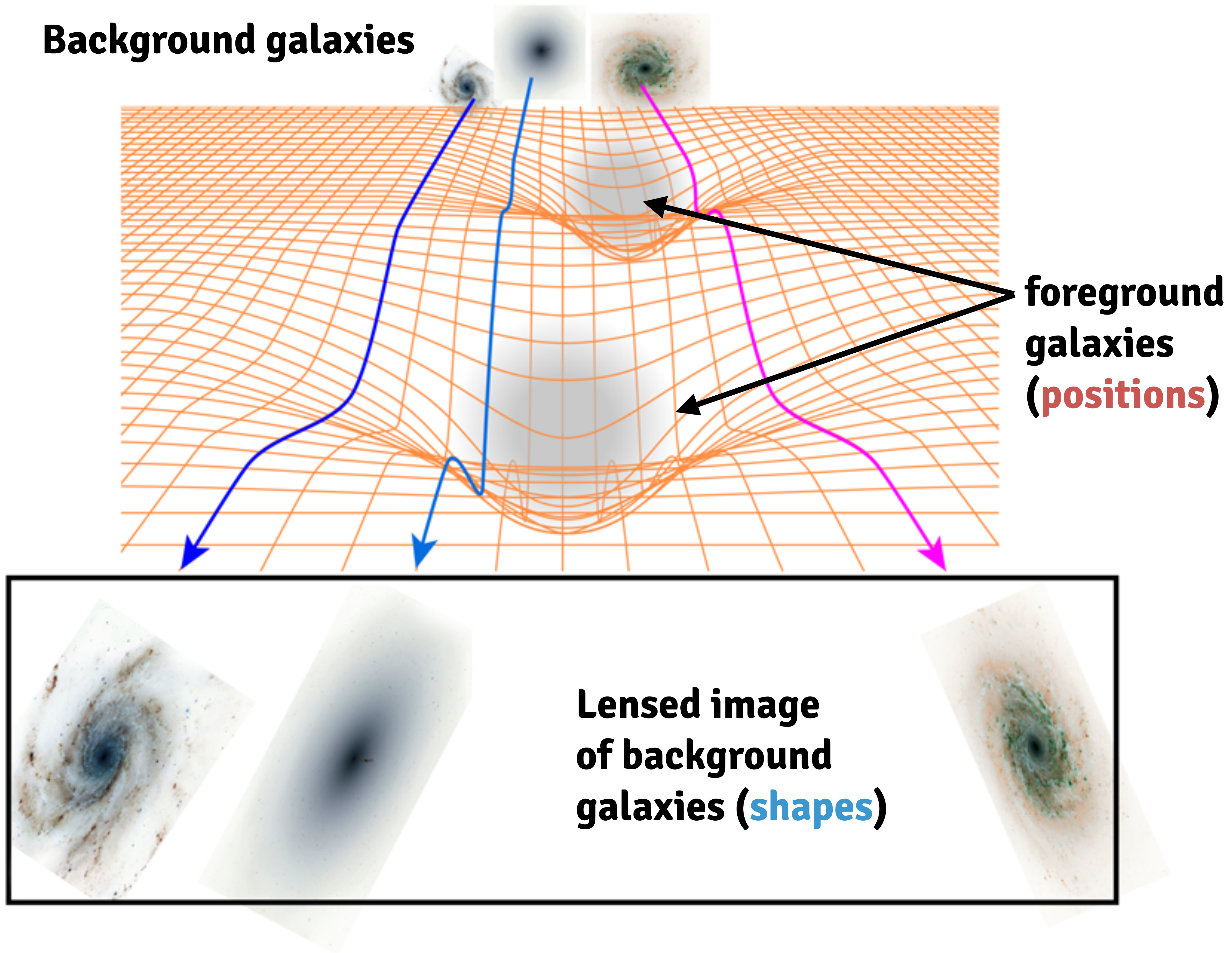
- **Strong lensing** [Montel+ 22]
- **Stellar Streams** [Alvey+ 23]
- **Gravitational Waves** [Bhardwaj+ 23] [Alvey+ 23]
- **CMB** [Cole+ 22]
- **21-cm** [Saxena+ 23]

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

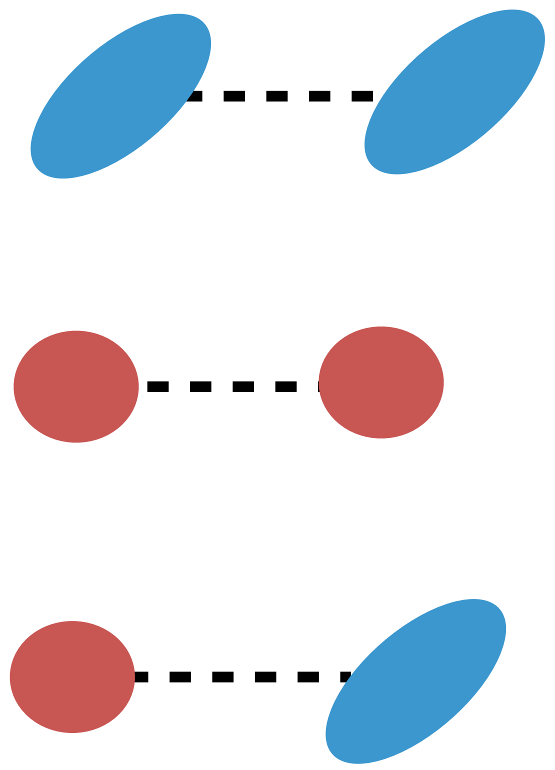
Which are the Euclid primary observables?



Which are the Euclid primary observables?

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

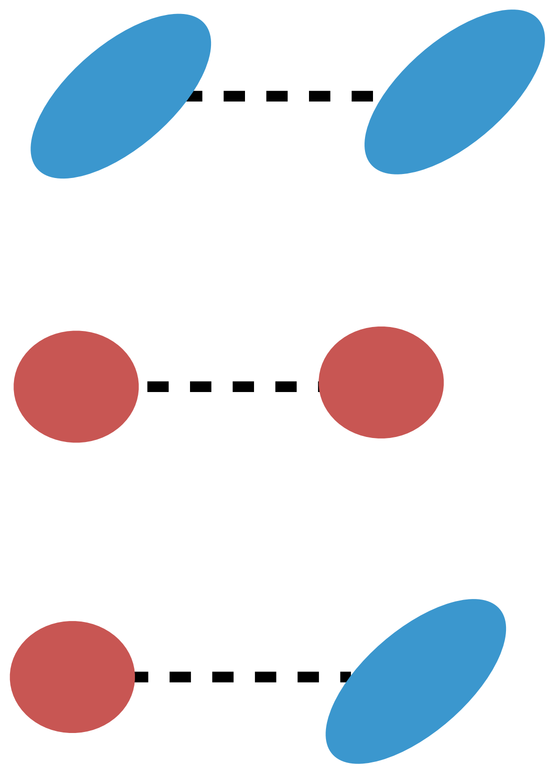
- **Cosmic Shear**
- **Galaxy clustering**
- **Galaxy-Galaxy lensing**



Which are the Euclid primary observables?

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

- **Cosmic Shear**
- **Galaxy clustering**
- **Galaxy-Galaxy lensing**

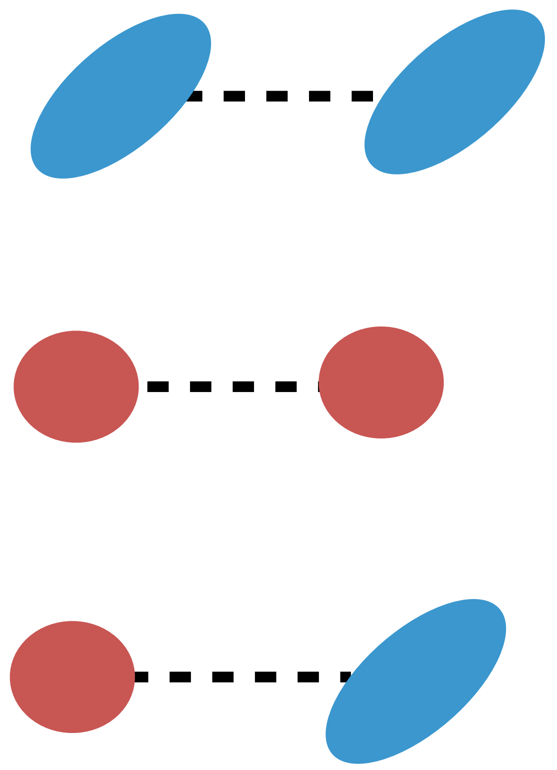


... measured for different tomographic redshift bins*

Which are the Euclid primary observables?

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

- **Cosmic Shear**
- **Galaxy clustering**
- **Galaxy-Galaxy lensing**



... measured for different
tomographic redshift bins*

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

Which are the Euclid primary observables?

- 3x2pt statistics described by a series of 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

Which are the Euclid primary observables?

- 3x2pt statistics described by a series of 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 power spectra!**

Which are the Euclid primary observables?

- 3x2pt statistics described by a series of power spectra:

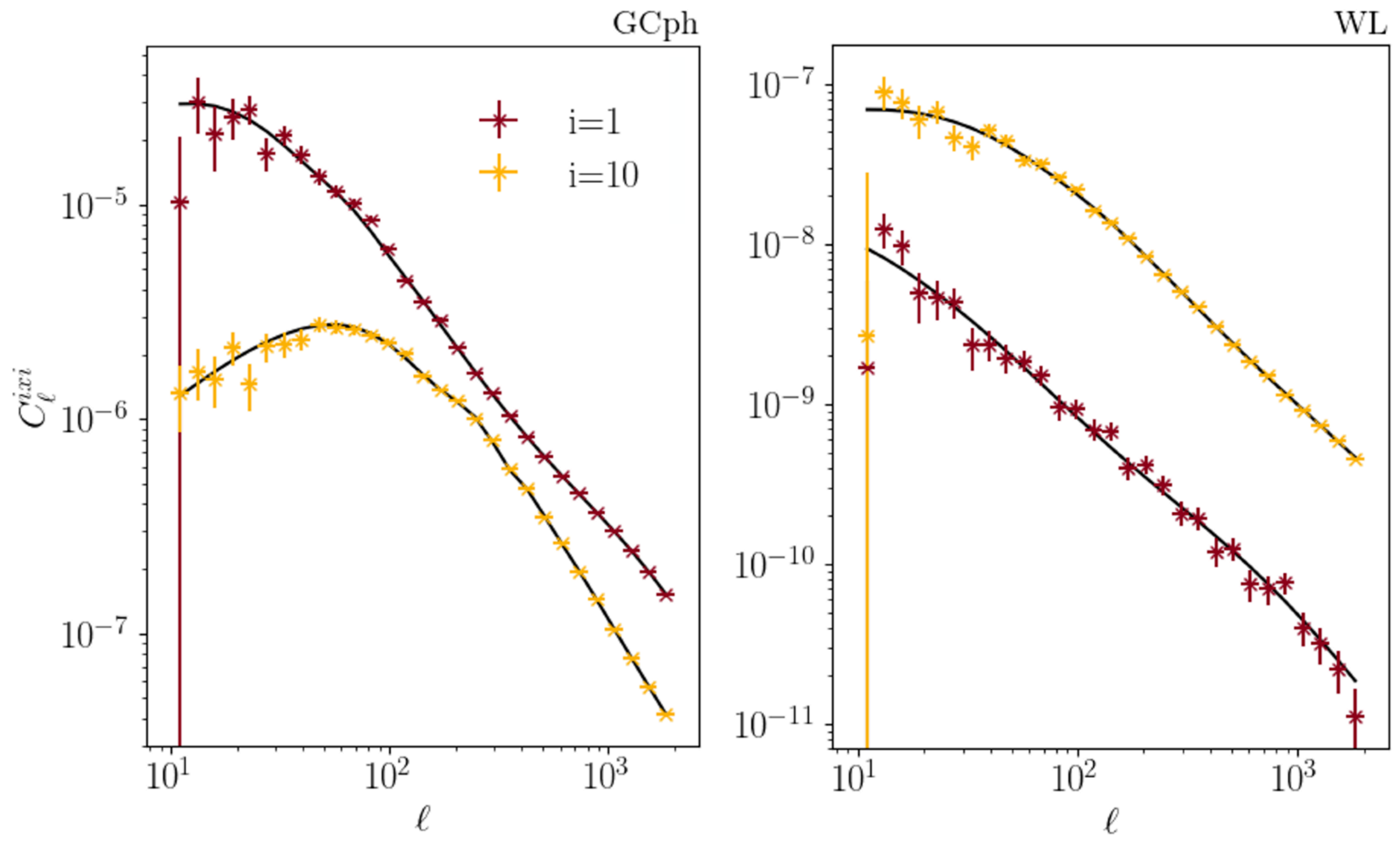
$$C_{ij}^{XY}(\ell) = \int dz \; W_i^X(z) W_j^Y(z) \; P_m(k_\ell, z)$$

Window functions — Matter power spectrum

- For 10 redshift bins up to z = 3

→ **+200 power spectra!**

Ex:



Preliminary results

- For **3x2pt**, we use a simulator based on a simplified Euclid likelihood (6 nuisance params)

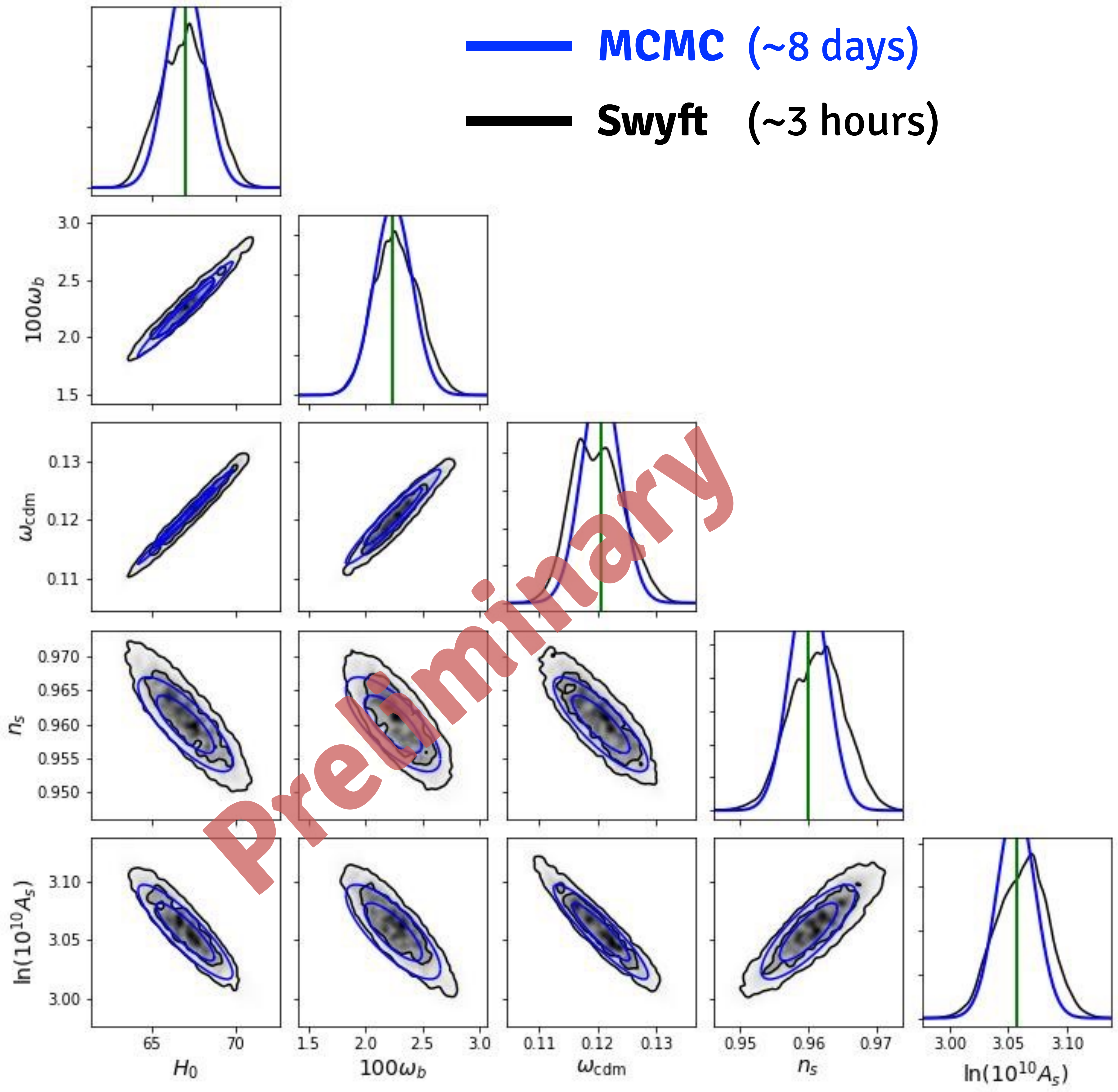
Preliminary results

- For **3x2pt**, we use a simulator based on a simplified Euclid likelihood (6 nuisance params)

- Perform mock data analysis on **ΛCDM** model (5 cosmo params)

Preliminary results

- For **3x2pt**, we use a simulator based on a simplified Euclid likelihood (6 nuisance params)
- Perform mock data analysis on **ΛCDM** model (5 cosmo params)



Conclusions

- Euclid will significantly deepen our understanding of the dark universe

Conclusions

- Euclid will significantly deepen our understanding of the dark universe

- To perform cosmological inference from Euclid data efficiently, we need to go beyond the MCMC

Conclusions

- Euclid will significantly deepen our understanding of the dark universe

- To perform cosmological inference from Euclid data efficiently, we need to go beyond the MCMC

- MNRE provides a highly efficient and flexible framework to analyse Euclid data

Conclusions

- Euclid will significantly deepen our understanding of the dark universe
- To perform cosmological inference from Euclid data efficiently, we need to go beyond the MCMC
- MNRE provides a highly efficient and flexible framework to analyse Euclid data

Early commissioning test image, Euclid VIS instrument

THANKS FOR YOUR ATTENTION

g.francoabellan@uva.nl

BACK-UP

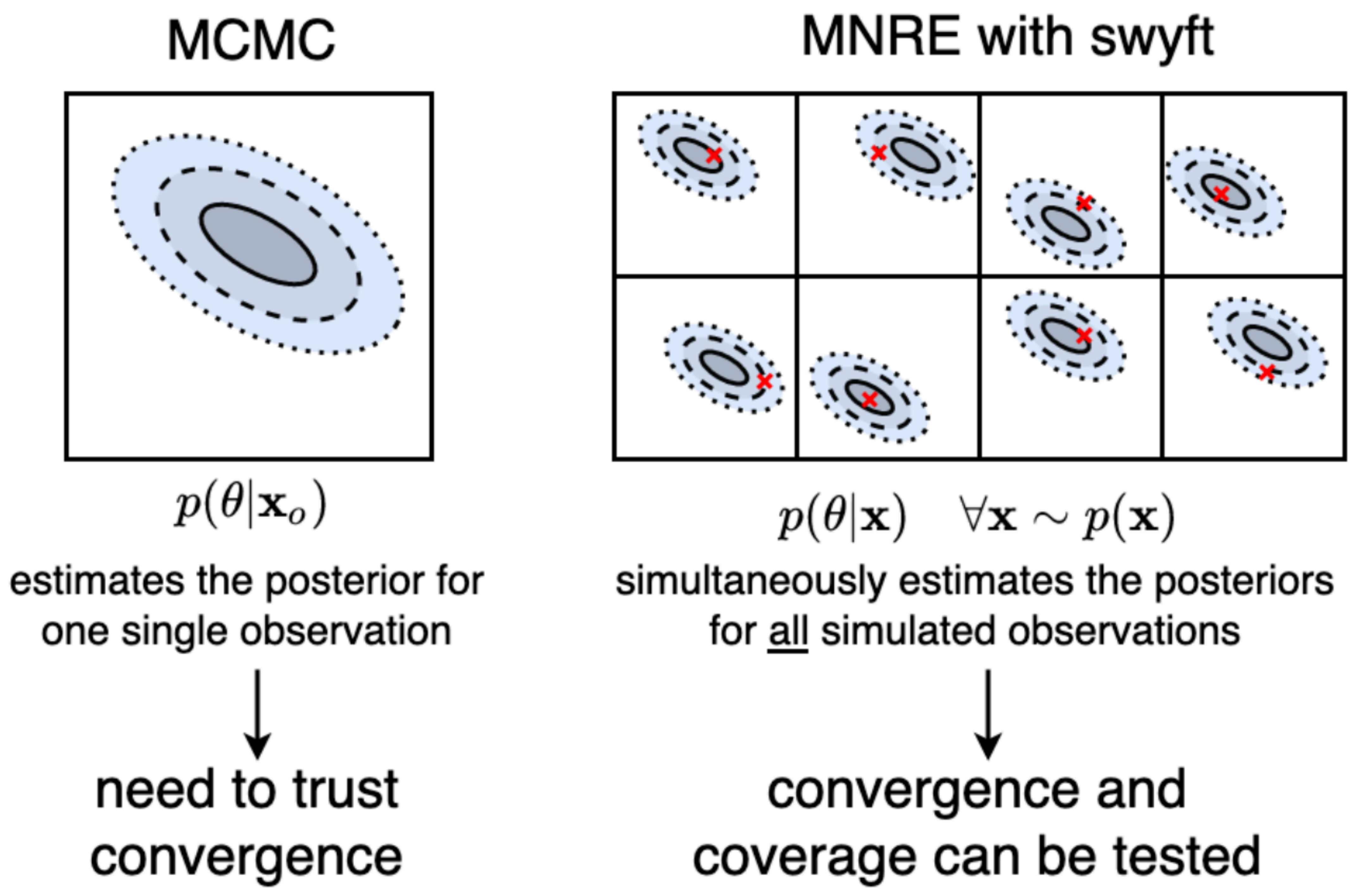
Can we trust our results?

- Even if NNs are often seen as "black boxes”, it is possible to perform statistical consistency tests which are imposible with MCMC

Can we trust our results?

- Even if NNs are often seen as "black boxes", it is possible to perform statistical consistency tests which are imposible with MCMC

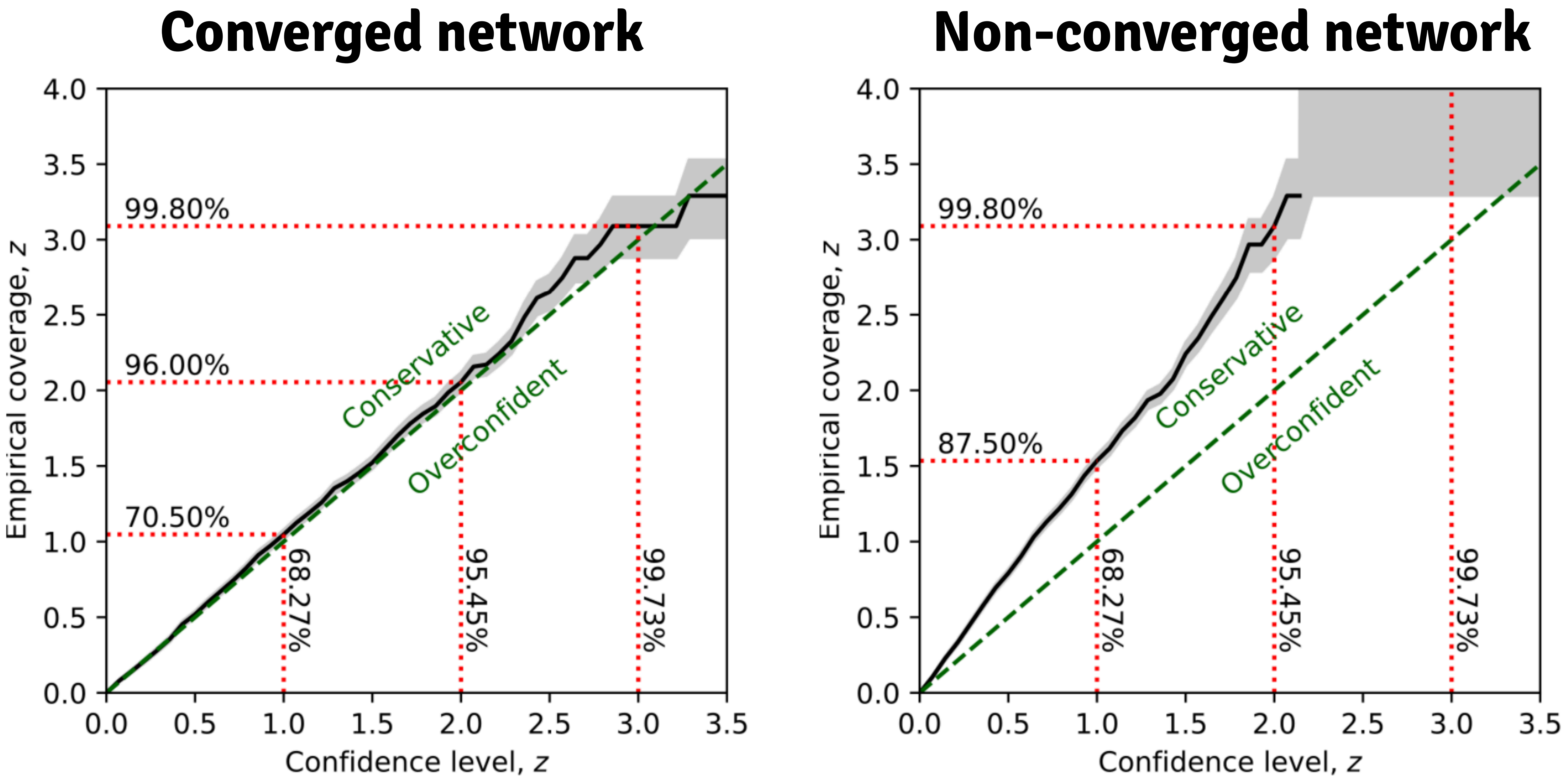
- Exploit MNRE's local amortization:



Cole+ 22

Can we trust our results?

- We can empirically estimate the Bayesian coverage



Cole+ 22