Finanziato dall'Unione europea
NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA
ICSC
Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

Optimal compression and Simulation-Based Inference of the cosmic 21-cm signal

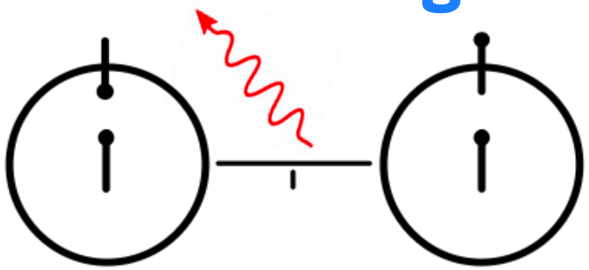
David Prelogović, Andrei Mesinger

Scuola Normale Superiore

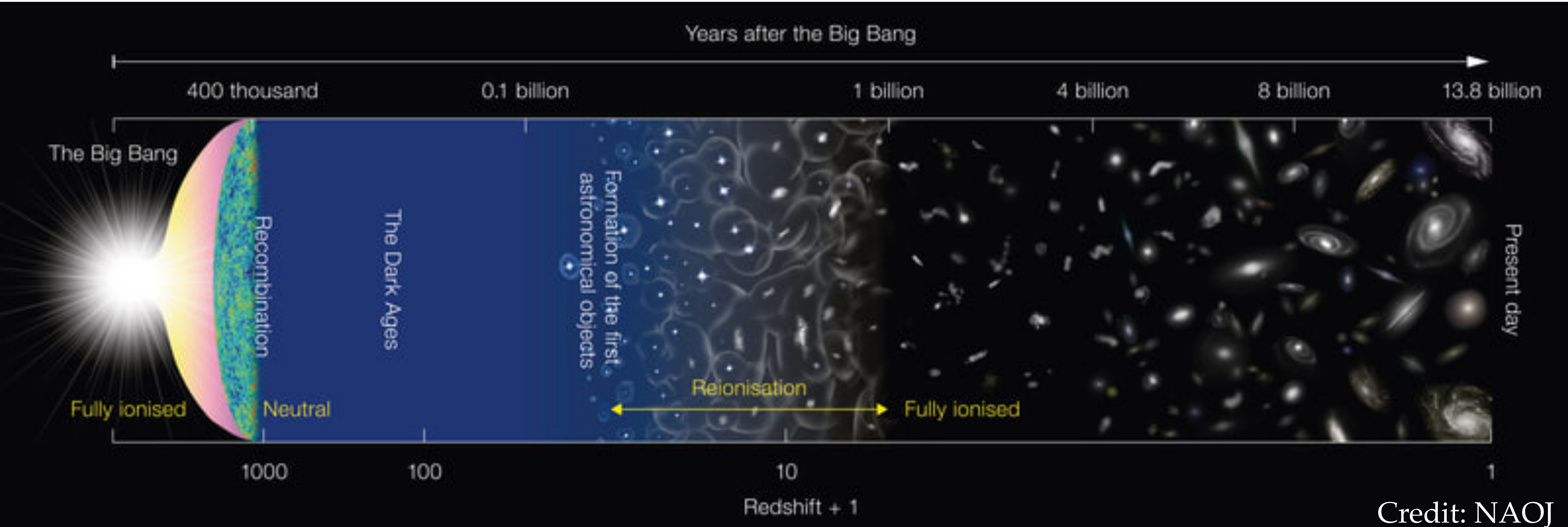
Spoke 3 General Meeting, Elba 5-9 / 05, 2024

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
ICSC Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

Cosmic 21-cm signal

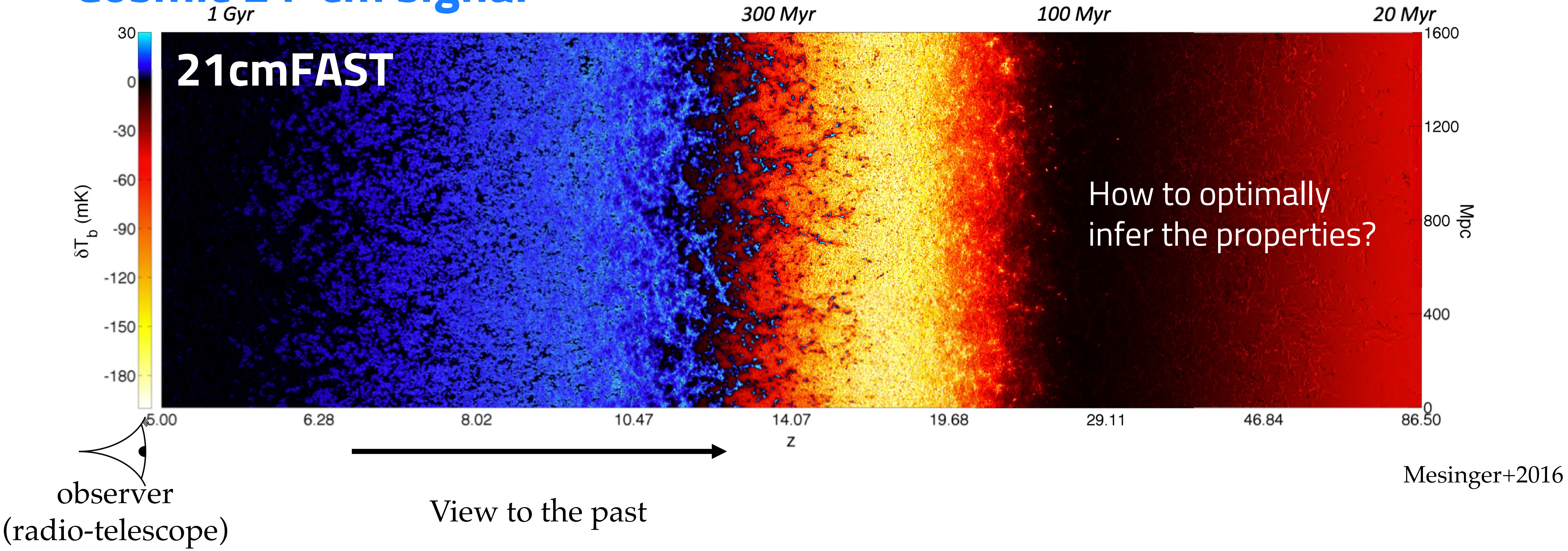


- Over 90% of the "normal" matter in the Universe is ***hydrogen***

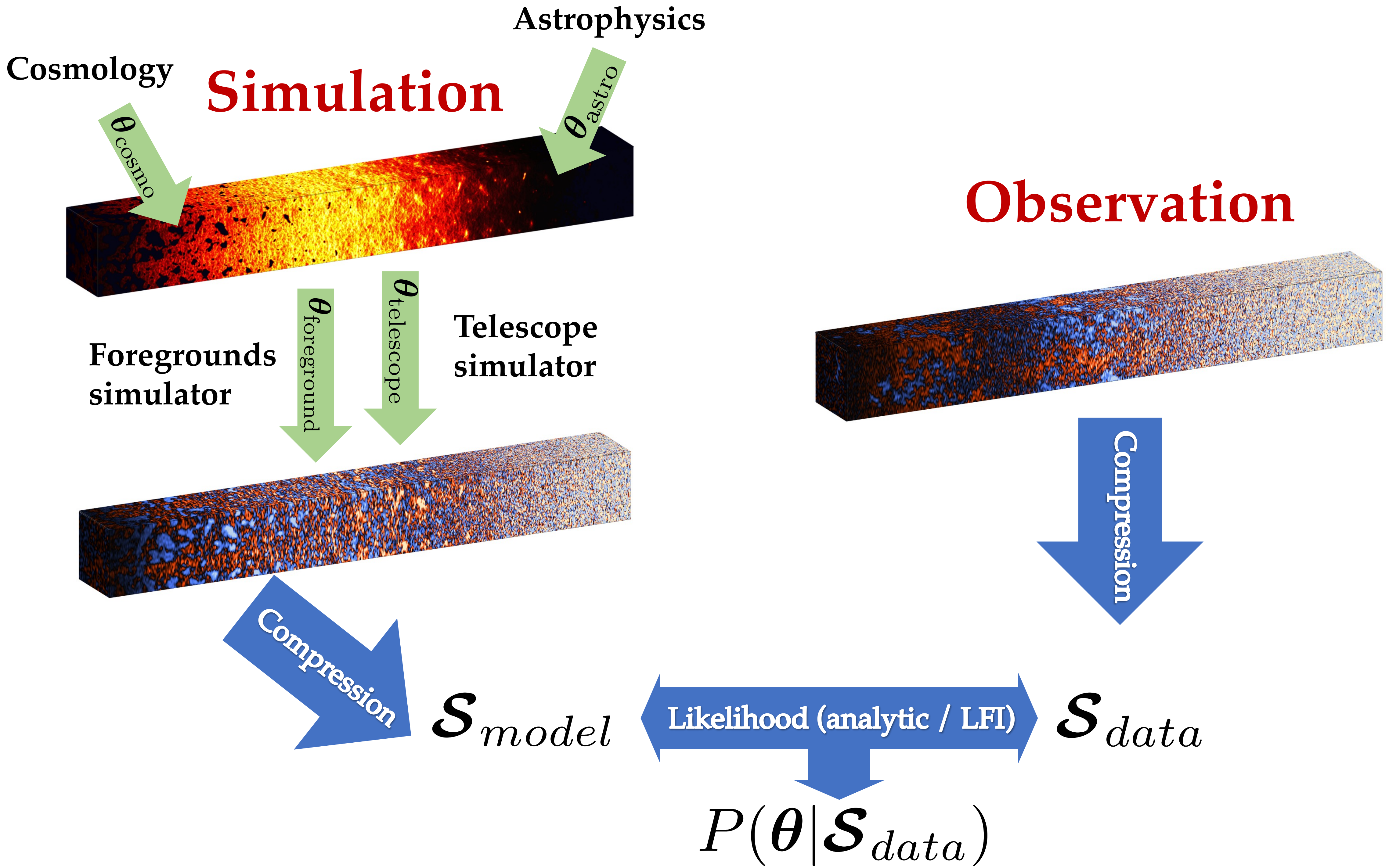


Credit: NAOJ

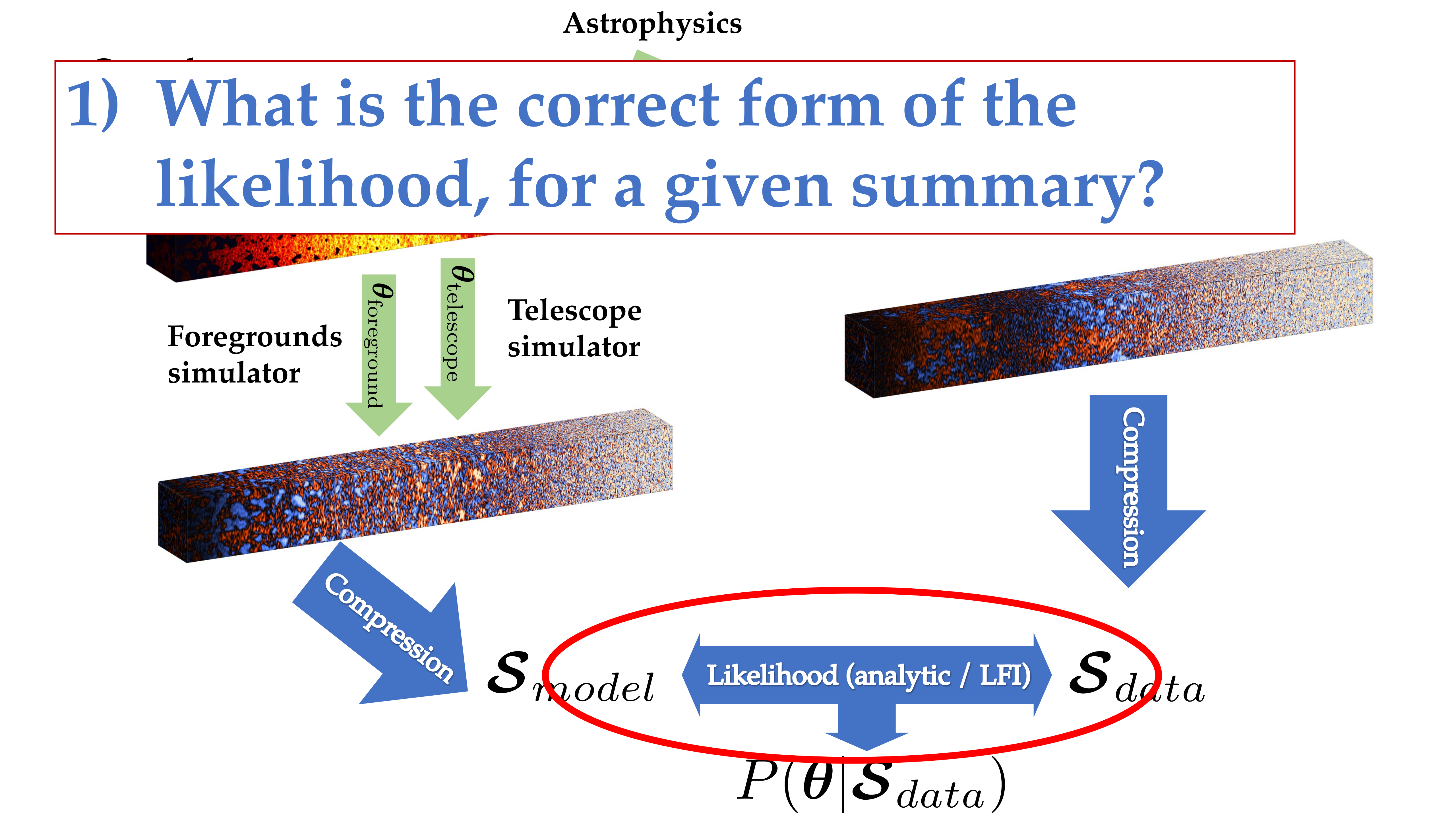
Cosmic 21-cm signal



- The properties of the ***unseen first galaxies*** are encoded in the timings and patterns of this signal!

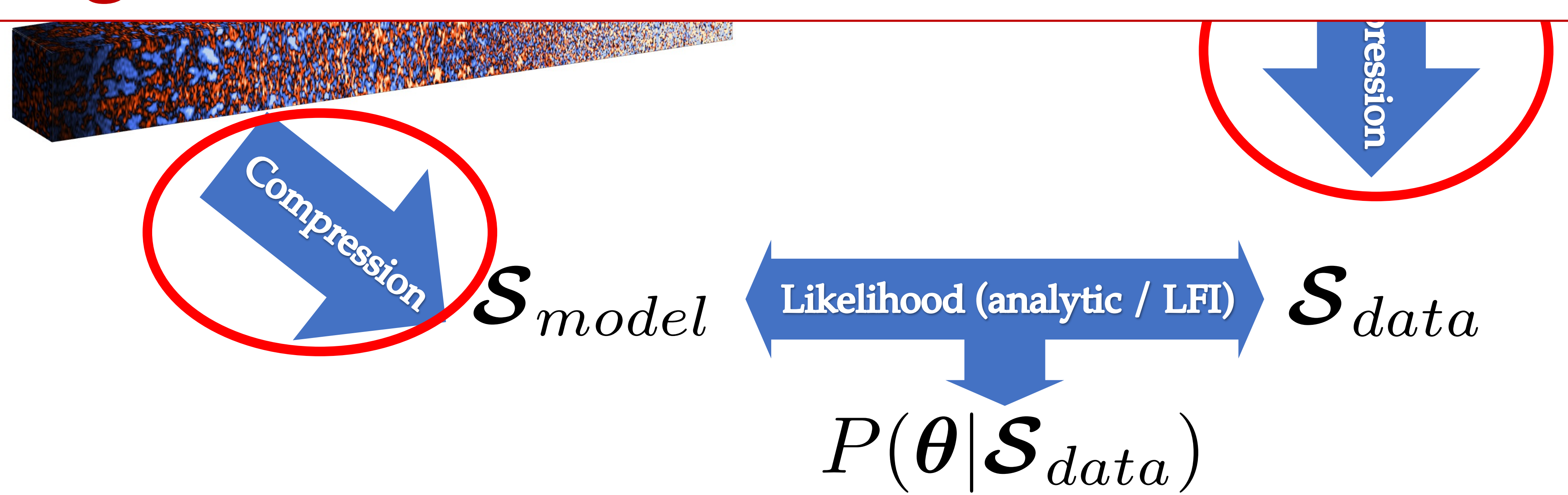


Astrophysics
Cosmology
Simulation
θ_{cosmo}
θ_{astro}
Observation
$\theta_{foreground}$
$\theta_{telescope}$
Telescope simulator
Foregrounds simulator
Compression
Compression
$\mathcal{S}_{model}$
Likelihood (analytic / LFI)
$\mathcal{S}_{data}$
$P(\theta|\mathcal{S}_{data})$


Astrophysics
1) What is the correct form of the likelihood, for a given summary?
Foregrounds simulator
$\theta_{foreground}$
$\theta_{telescope}$
Telescope simulator
Compression
Compression
s_{model}
Likelihood (analytic / LFI)
s_{data}
$P(\theta|s_{data})$

Astrophysics

1) **What is the correct form of the likelihood, for a given summary?**
2) **What summary extracts the most information from this non-Gaussian signal?**



Simulation-Based Inference for the 21-cm PS

Is the PS likelihood Gaussian?

$$\ln \mathcal{L}(\Delta^2_{21\,\mathrm{obs}}|\boldsymbol{\theta}) \propto -\frac{1}{2}\left[\Delta^2_{21\,\mathrm{obs}} - \mu(\boldsymbol{\theta})\right]^T \Sigma^{-1}\left[\Delta^2_{21\,\mathrm{obs}} - \mu(\boldsymbol{\theta})\right]$$

Prelogovic & AM (2023)
see also Zhao+2022

Is the PS likelihood Gaussian?

model mean $\mu(\theta, k, z)$
(averaging over stochasticity: cosmic variance, noise)

$$\ln \mathcal{L}(\Delta^2_{21\,\mathrm{obs}}|\boldsymbol{\theta}) \propto -\frac{1}{2}\left[\Delta^2_{21\,\mathrm{obs}} - \mu(\boldsymbol{\theta})\right]^T \Sigma^{-1} \left[\Delta^2_{21\,\mathrm{obs}} - \mu(\boldsymbol{\theta})\right]$$

Prelogovic & AM (2023)
see also Zhao+2022

Is the PS likelihood Gaussian?

model mean $\mu(\theta, k, z)$
(averaging over stochasticity: cosmic variance, noise)

$$\ln \mathcal{L}(\Delta^2_{21\,\mathrm{obs}}|\boldsymbol{\theta}) \propto -\frac{1}{2}\,[\Delta^2_{21\,\mathrm{obs}} - \mu(\boldsymbol{\theta})]^T\,\Sigma^{-1}\,[\Delta^2_{21\,\mathrm{obs}} - \mu(\boldsymbol{\theta})]$$

covariance $\Sigma(\theta, k_1, z_1, k_2, z_2)$

Prelogovic & AM (2023)
see also Zhao+2022

Is the PS likelihood Gaussian?

$$\ln \mathcal{L}(\Delta^2_{21\,\mathrm{obs}}|\boldsymbol{\theta}) \propto -\frac{1}{2}\left[\Delta^2_{21\,\mathrm{obs}} - \mu(\boldsymbol{\theta})\right]^T \Sigma^{-1} \left[\Delta^2_{21\,\mathrm{obs}} - \mu(\boldsymbol{\theta})\right]$$

$$\approx -\frac{1}{2}\sum_{k,z} \frac{\left[\Delta^2_{21\,\mathrm{obs}}(k,z) - \mu(\boldsymbol{\theta},k,z)\right]^2}{\sigma^2(\boldsymbol{\theta},k,z)},$$

variance (ignoring non-diagonal terms)

Prelogovic & AM (2023)
see also Zhao+2022

Is the PS likelihood Gaussian?

$$\ln \mathcal{L}(\Delta^2_{21\,\mathrm{obs}}|\boldsymbol{\theta}) \propto -\frac{1}{2}\left[\Delta^2_{21\,\mathrm{obs}} - \mu(\boldsymbol{\theta})\right]^T \Sigma^{-1} \left[\Delta^2_{21\,\mathrm{obs}} - \mu(\boldsymbol{\theta})\right]$$

$$\approx -\frac{1}{2}\sum_{k,z} \frac{\left[\Delta^2_{21\,\mathrm{obs}}(k,z) - \mu(\boldsymbol{\theta},k,z)\right]^2}{\sigma^2(\boldsymbol{\theta},k,z)},$$

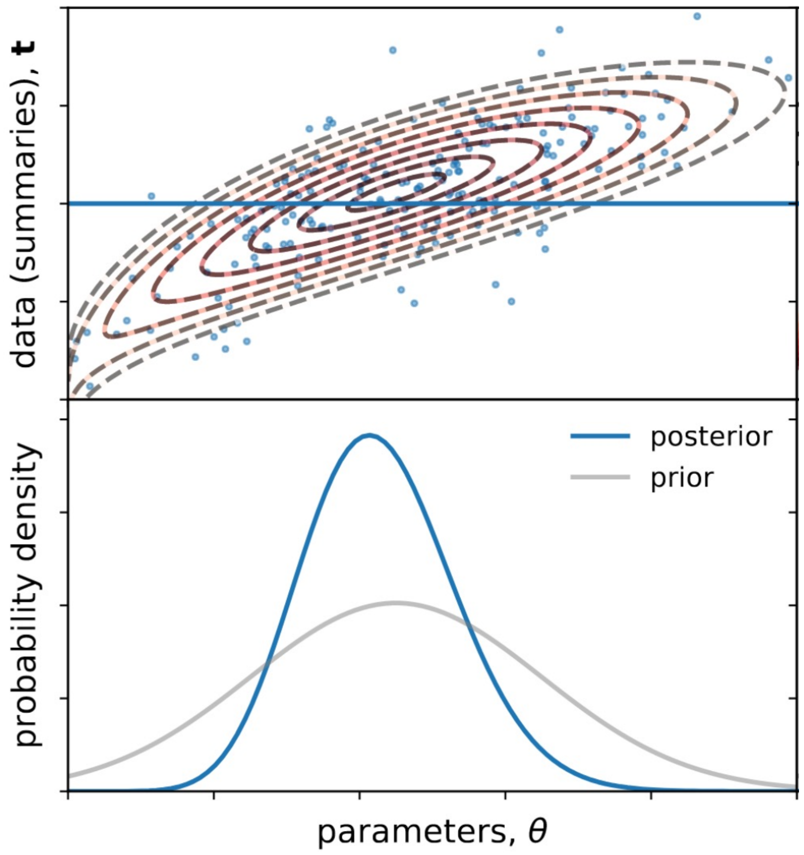
Common assumptions:

- *Is the likelihood Gaussian?*
- *Can we ignore non-diagonal terms in the covariance btw k and z?*
- *Can we estimate the mean from a single realization?*
- *Can we estimate (co)variance at a single, "fiducial" parameter value*

Prelogovic & AM (2023)
see also Zhao+2022

We can actually map the ***TRUE*** likelihood with Simulation Based Inference

Simulation-Based Inference (SBI)



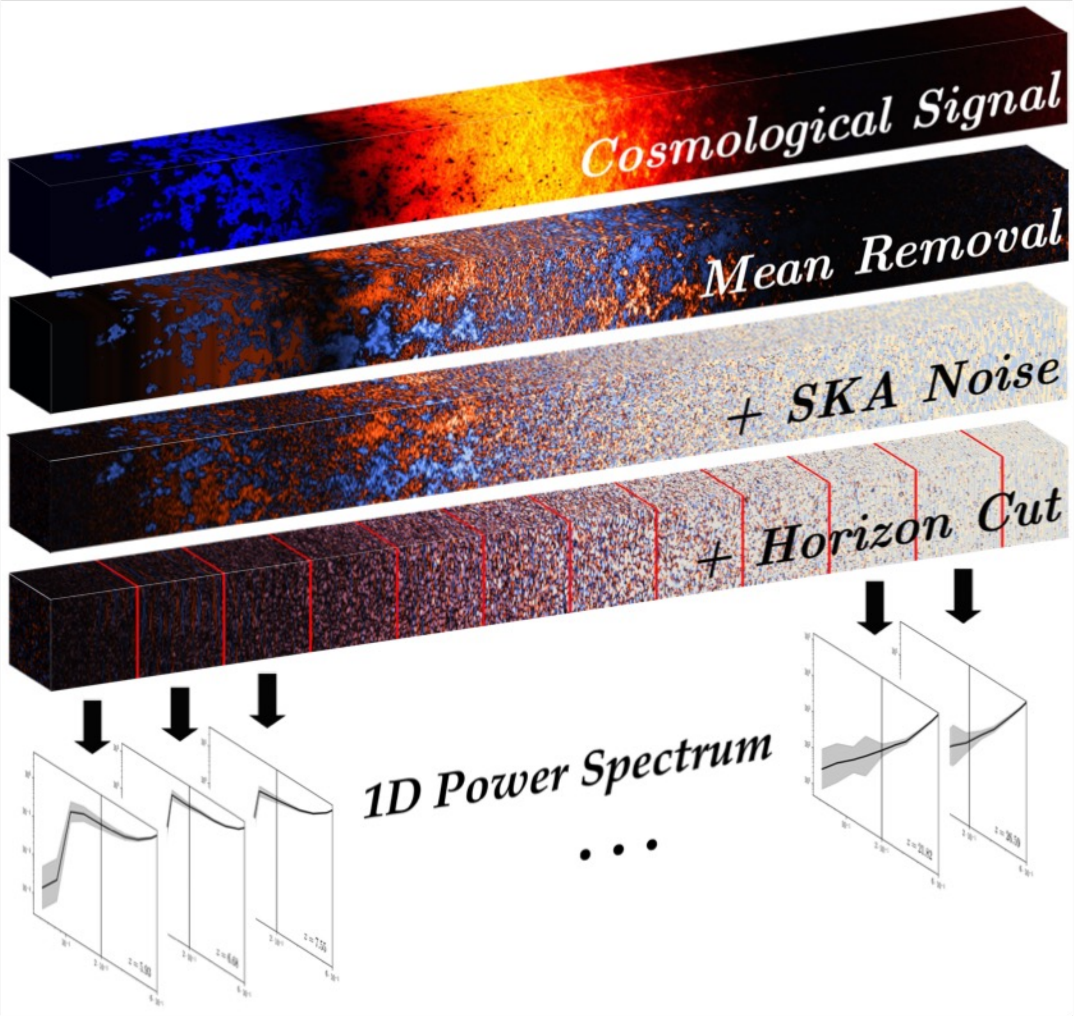
- Pull from the prior $\tilde{\theta} \sim p(\theta)$
- Pull from the likelihood $\tilde{t} \sim p(t|\tilde{\theta})$
  - Use simulator to simulate the data $$\tilde{d} = \text{simulator}(\tilde{\theta})$$
  - Optionally compress the data to summary $$\tilde{t} = \text{compressor}(\tilde{d})$$
- Repeat many times to construct a set $$\left\{(\tilde{t}_1, \tilde{\theta}_1), (\tilde{t}_2, \tilde{\theta}_2), \ldots, (\tilde{t}_N, \tilde{\theta}_N)\right\}$$
- It is assumed the simulator is perfect, i.e. it pulls from the actual likelihood

$$p(d|\tilde{\theta})$$

Alsing+2019

Simulation Based Inference (SBI)

***Precompute database* of forward models (varying cosmic ICs, noise, etc.) and *train density estimators* to fit the likelihood.**



Rapidly becoming popular in 21cm:
Zhao+2022ab, Saxena+2023

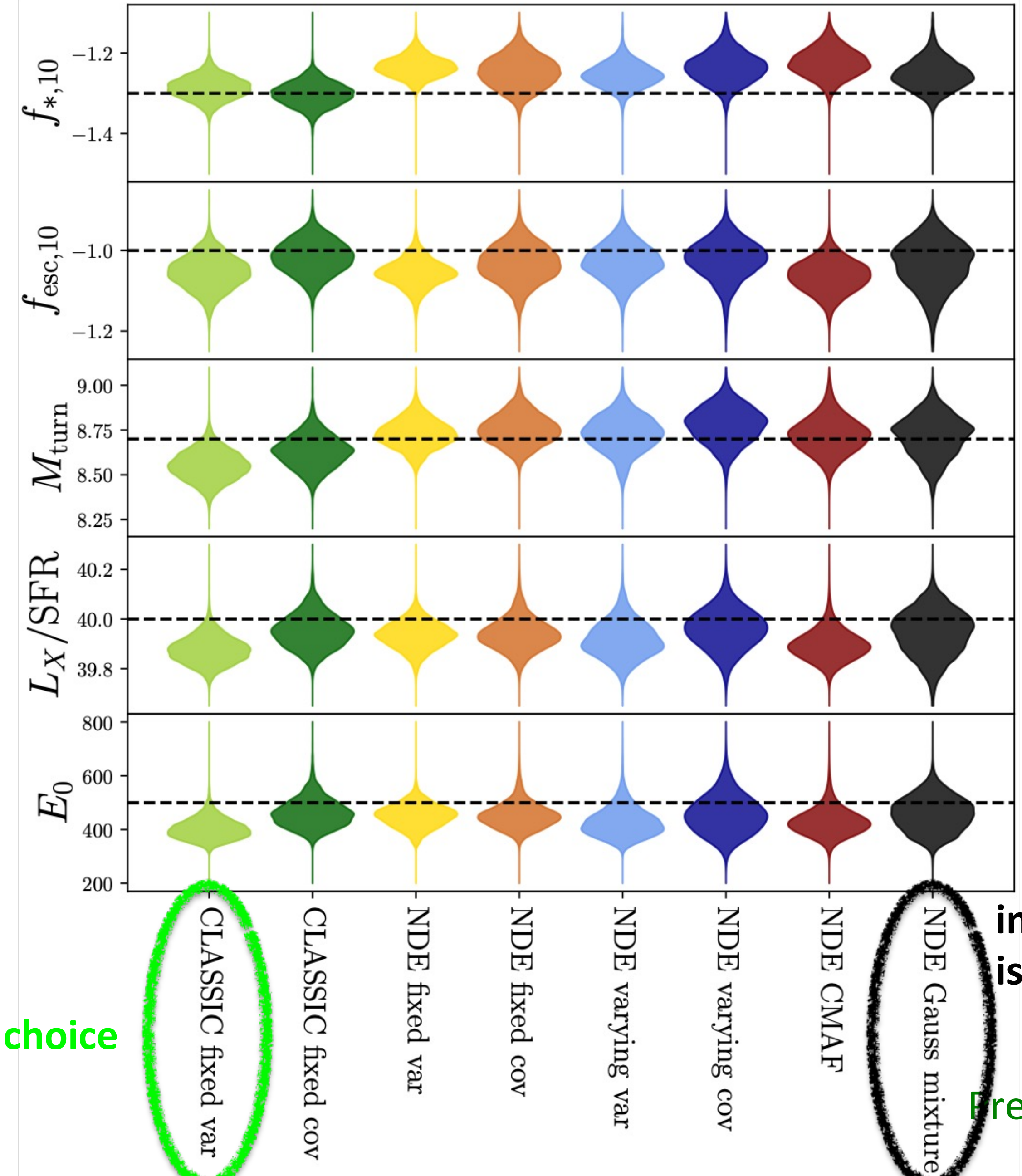
Prelogovic & AM (2023)

Use different functional forms to test the assumptions

	non-Gaussian	non-diagonal covariance	(co)variance is a function of $\boldsymbol{\theta}$	mean by averaging over realizations
CLASSIC fixed var	✗	✗	✗	✗ (single, well-chosen seed)
CLASSIC fixed cov	✗	✓	✗	✗ (single, well-chosen seed)
NDE fixed var	✗	✗	✗	✓
NDE fixed cov	✗	✓	✗	✓
NDE varying var	✗	✗	✓	✓
NDE varying cov	✗	✓	✓	✓
NDE CMAF	✓	N/A	N/A	N/A
NDE Gauss mixture	✓	N/A	N/A	N/A

Prelogovic & AM (2023)

Use different functional forms to test the assumptions



~ most common choice

implicit likelihood SBI is most accurate

Prelogovic & AM (2023)

Simulation Based Calibration (SBC)

- ***"prior" = "data averaged posterior"*** $P(\boldsymbol{\theta}) = \int P(\boldsymbol{\theta}|\tilde{\boldsymbol{y}})\, P(\tilde{\boldsymbol{y}}|\tilde{\boldsymbol{\theta}}) P(\tilde{\boldsymbol{\theta}})\, \mathrm{d}\tilde{\boldsymbol{y}}\, \mathrm{d}\tilde{\boldsymbol{\theta}}$

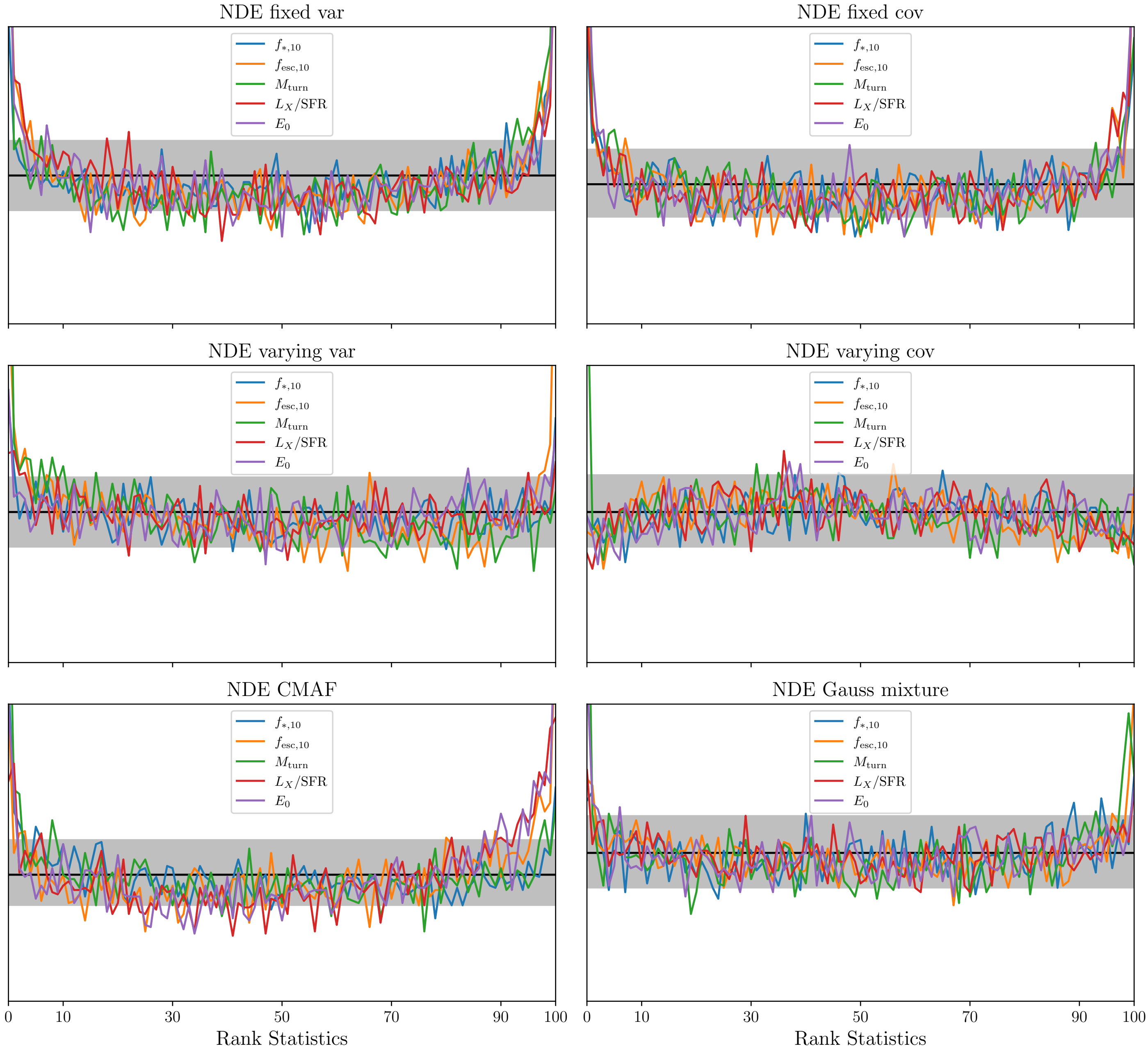
1. Pull from prior $\tilde{\boldsymbol{\theta}} \sim P(\boldsymbol{\theta})$
2. Pull the data from the likelihood $\tilde{\boldsymbol{y}} \sim P(\boldsymbol{y}|\tilde{\boldsymbol{\theta}}) \quad \Leftrightarrow \quad \tilde{\boldsymbol{y}} = \text{simulator}(\tilde{\boldsymbol{\theta}})$
3. Calculate the posterior the sample $P(\boldsymbol{\theta}|\tilde{\boldsymbol{y}})$
4. Repeat and average posteriors $P(\boldsymbol{\theta}) \approx \frac{1}{N} \sum_{i=1}^{N} P_i(\boldsymbol{\theta}|\tilde{\boldsymbol{y}}_i)$

SBC for 21-cm PS

- *Expensive to compute!*
  - *10 000 posteriors*

- However, once likelihood is trained, no new simulations are needed
  - "amortized inference"
- Procedure would be useful for classic inference too, but is impossible to compute

- NDE fixed var & cov – overconfident
- NDE varying var & cov – good
- NDE CMAF – overconfident
- NDE Gauss mixture – the best



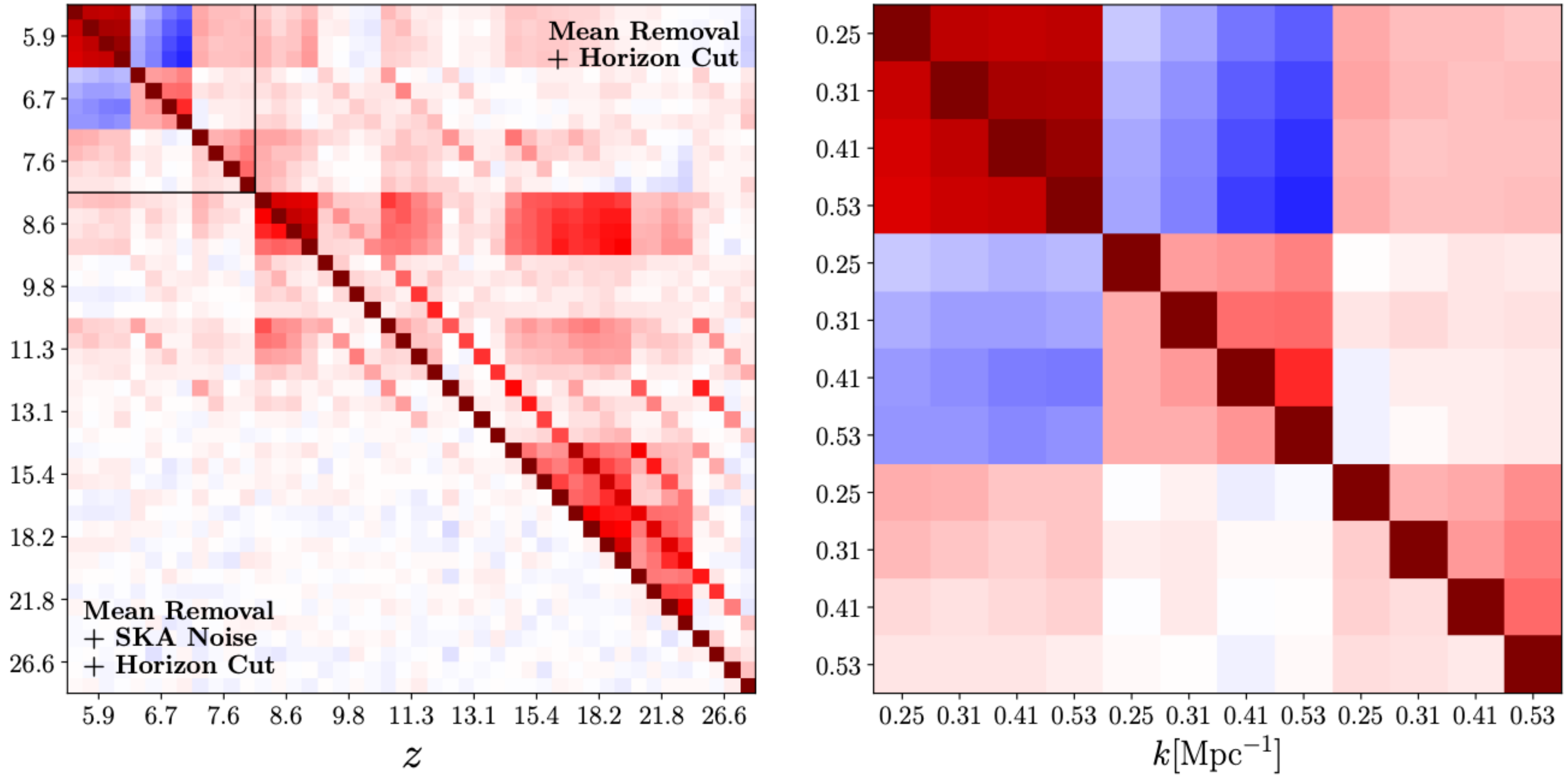
Prelogović & AM (2023)

So… Is the PS likelihood Gaussian?

- Is the likelihood close to Gaussian? —> ***close enough (assuming wedge excision)***

So… Is the PS likelihood Gaussian?

- Is the likelihood close to Gaussian? —> ***close enough***
- Can we ignore 2pt covariance btw k and z? —> ***No***



So… Is the PS likelihood Gaussian?

- Is the likelihood close to Gaussian? —> ***close enough***
- Can we ignore 2pt covariance btw k and z? —> ***No***
- Can we estimate the mean from a single realization? —> ***Only if you choose well***

So… Is the PS likelihood Gaussian?

- Is the likelihood close to Gaussian? —> ***close enough***
- Can we ignore 2pt covariance btw k and z? —> ***No***
- Can we estimate the mean from a single realization? —> ***Only if you choose well***
- Can we estimate (co)variance at a single, “fiducial” parameter value —> ***No***

So… Is the PS likelihood Gaussian?

- Is the likelihood close to Gaussian? —> ***close enough***
- Can we ignore 2pt covariance btw k and z? —> ***No***
- Can we estimate the mean from a single realization? —> ***Only if you choose well***
- Can we estimate (co)variance at a single, "fiducial" parameter value —> ***No***

But we don't have to worry about it because:

SBI gives *more accurate posteriors* using *an order of magnitude fewer simulations than MCMC/Nested sampling*

- no "emulation error" that is difficult to characterize over parameter space -> NDE error easier to mitigate
- once likelihood is learned, new inference from updated data can be done ~ instantaneously ("amortized" cost)

Prelogovic & AM (2023)

Optimal Compression of the 21-cm lightcones

Ad-hoc compressions

- Without good a-priori physical motivation, we cannot know what is THE optimal compression/summary, i.e. providing tightest recovery of astrophysics and cosmology

Ad-hoc compressions

- Without good a-priori physical motivation, we cannot know what is THE optimal compression/summary, i.e. providing tightest recovery of astrophysics and cosmology

Solution:

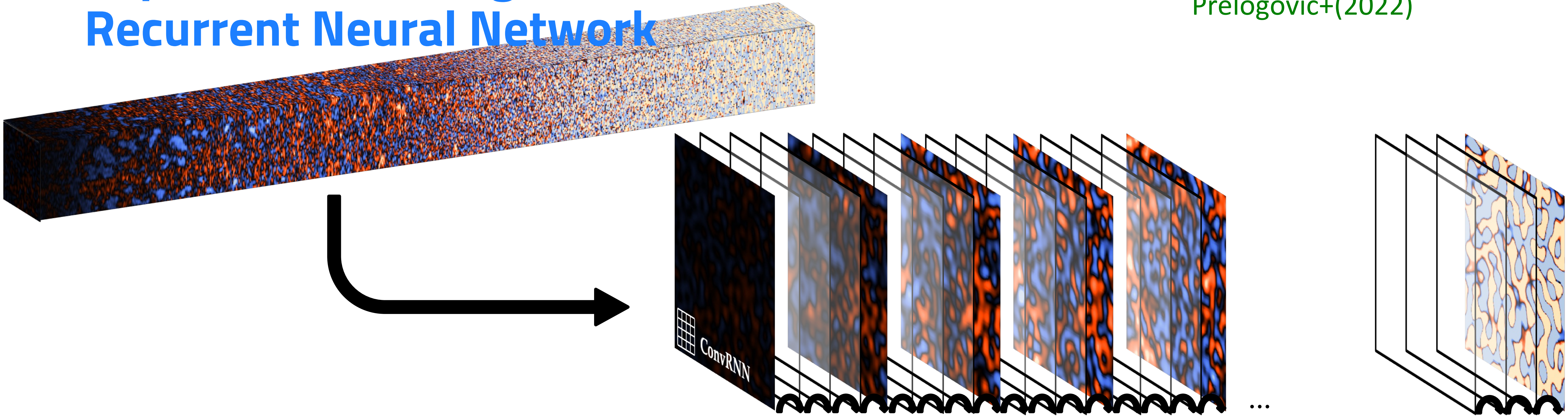
Let the machines figure it out for us!

(Neural Network)

- Gillet+2018
- La Plante & Ntampaka 2019
- Makinen+2020
- Mangena+2020
- Hortúa+2020
- Prelogović+2021
- +++

Supervised learning: Recurrent Neural Network

Prelogovic+(2022)



- RNN – encoding correlations across all frequency bins at once
- 2D Convolutional NN– local correlations in sky-plane

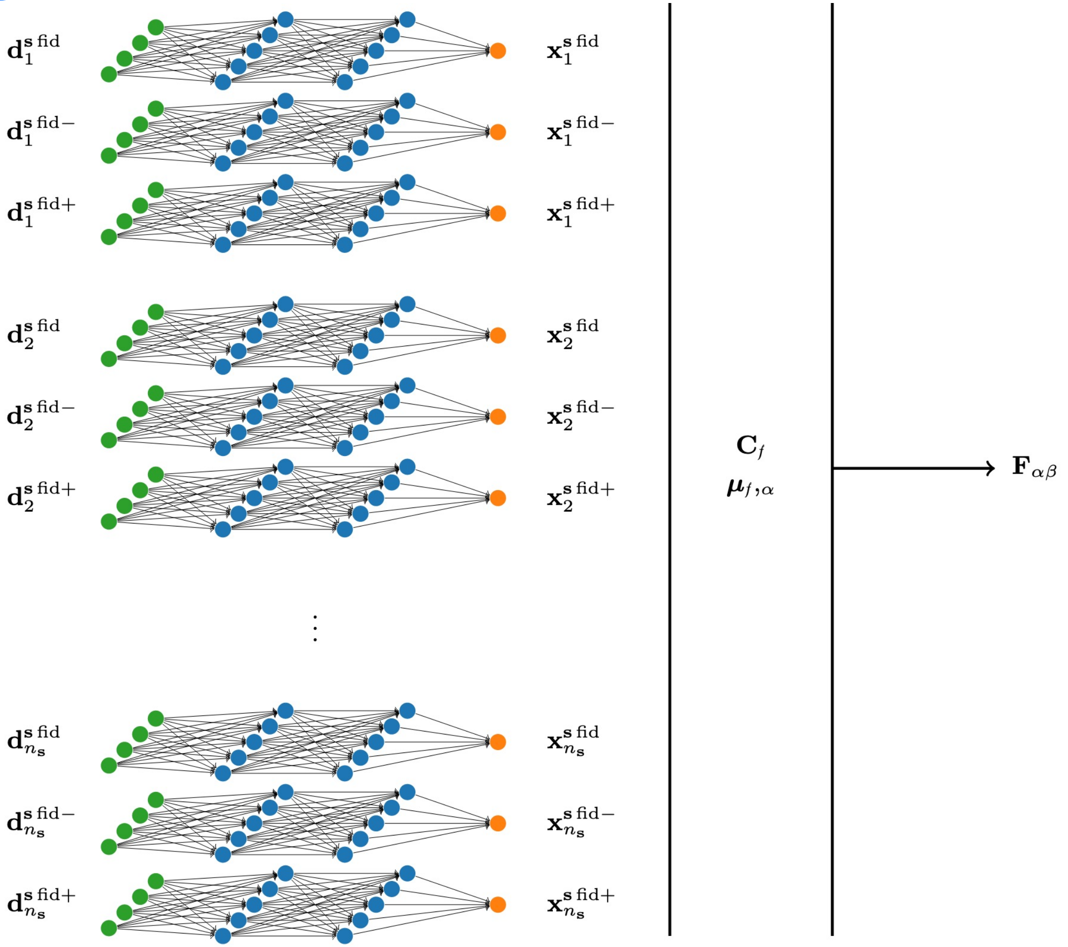
Unsupervised learning: Information Maximizing Neural Network

- Simulate the data at a fiducial parameter set: $\mathbf{d}(\boldsymbol{\theta}_{\rm fid})$
- Simulate around the fiducial parameters: $\mathbf{d}(\boldsymbol{\theta}^{+}_{\rm fid}), \mathbf{d}(\boldsymbol{\theta}^{-}_{\rm fid})$
- Calculate compressed summary:

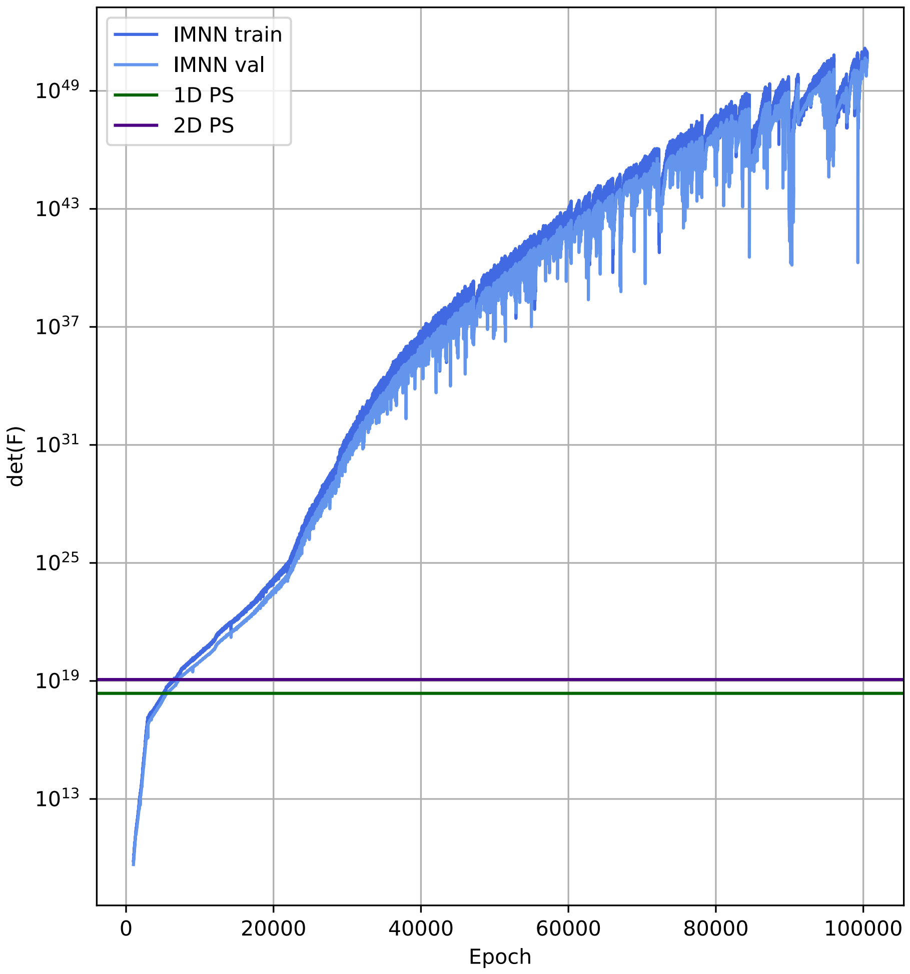
$$\mathbf{s}(\boldsymbol{\theta}) = \mathrm{NN}(\mathbf{d}(\boldsymbol{\theta}))$$

- Maximize Fisher information:

$$L = -\ln|F|$$



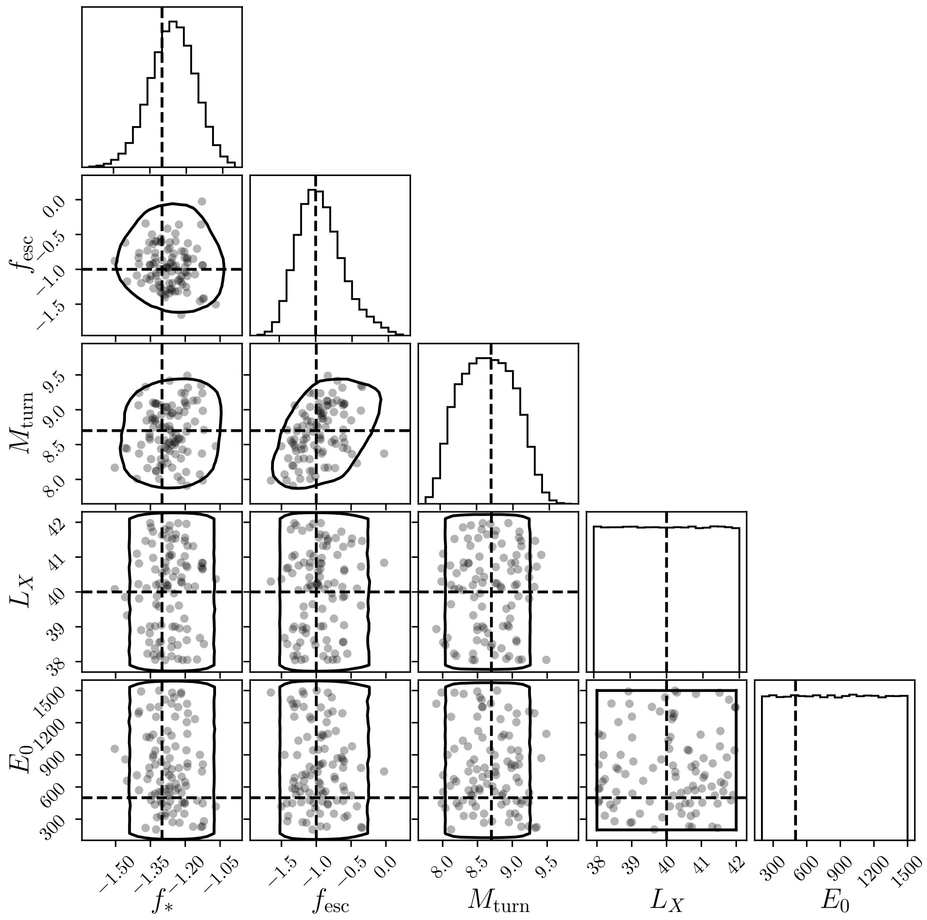
IMNN Fisher information at fiducial



Prelogović & Mesinger (2024)

over 40 orders of magnitude improvement in Fisher information wrt power spectrum! ~ 10^8 times lower variance in parameter recovery!

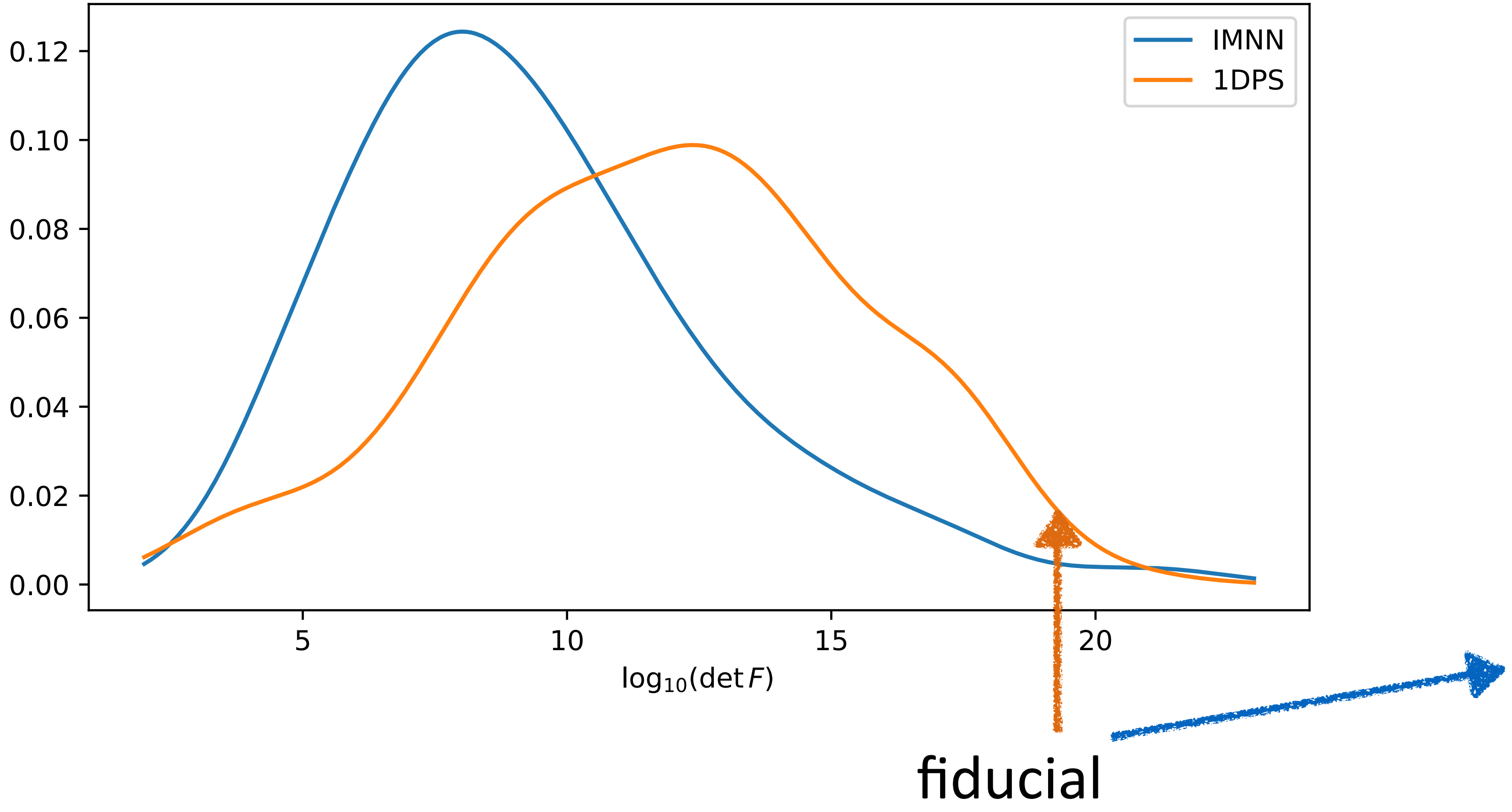
How optimal is the compression for different fiducial parameter choices?



sample your prior volume and evaluate Fisher information throughout -> *prior weighted Fisher information*

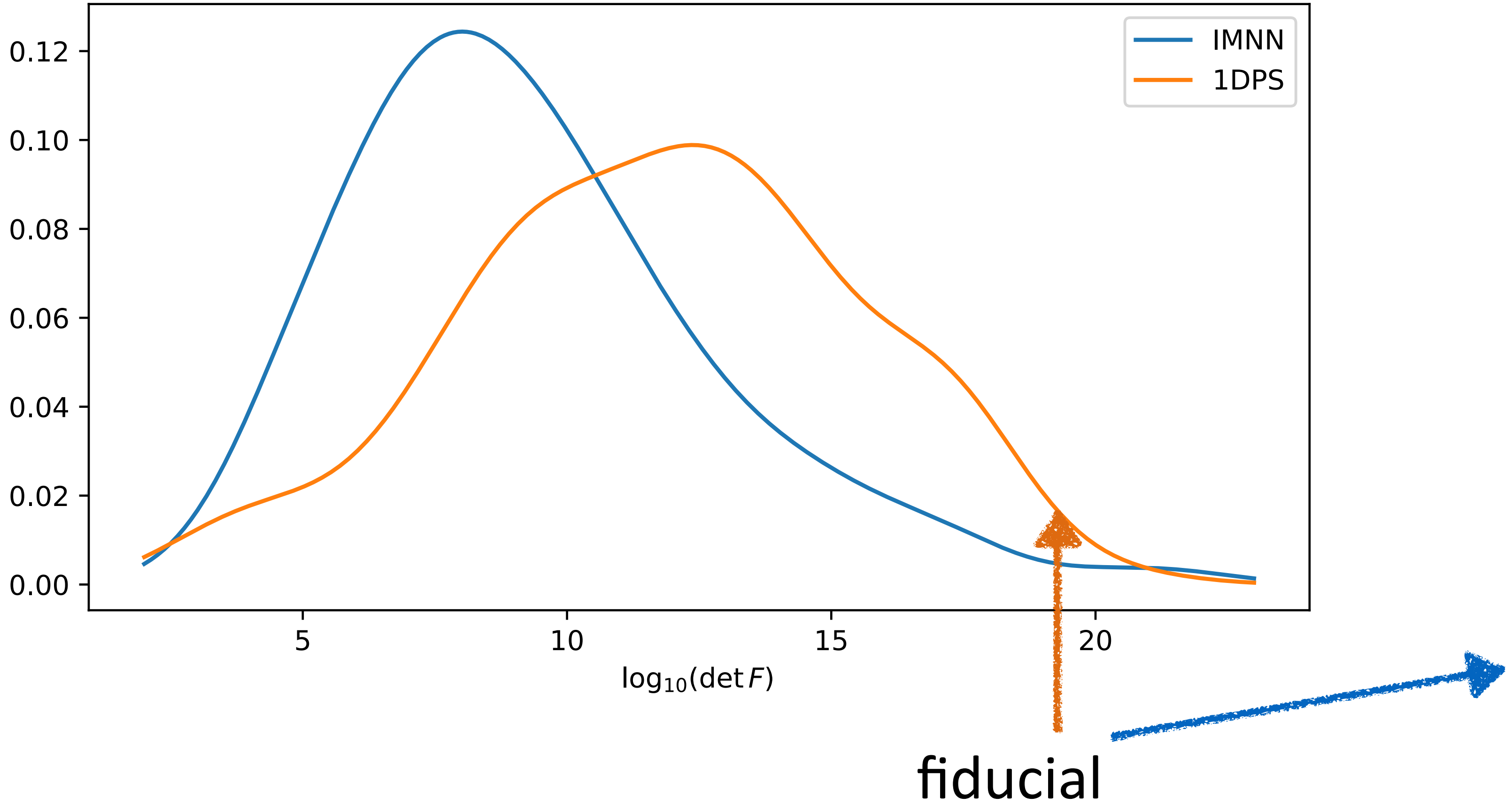
Prelogović & Mesinger (2024)

How optimal is the compression for different fiducial parameter choices?



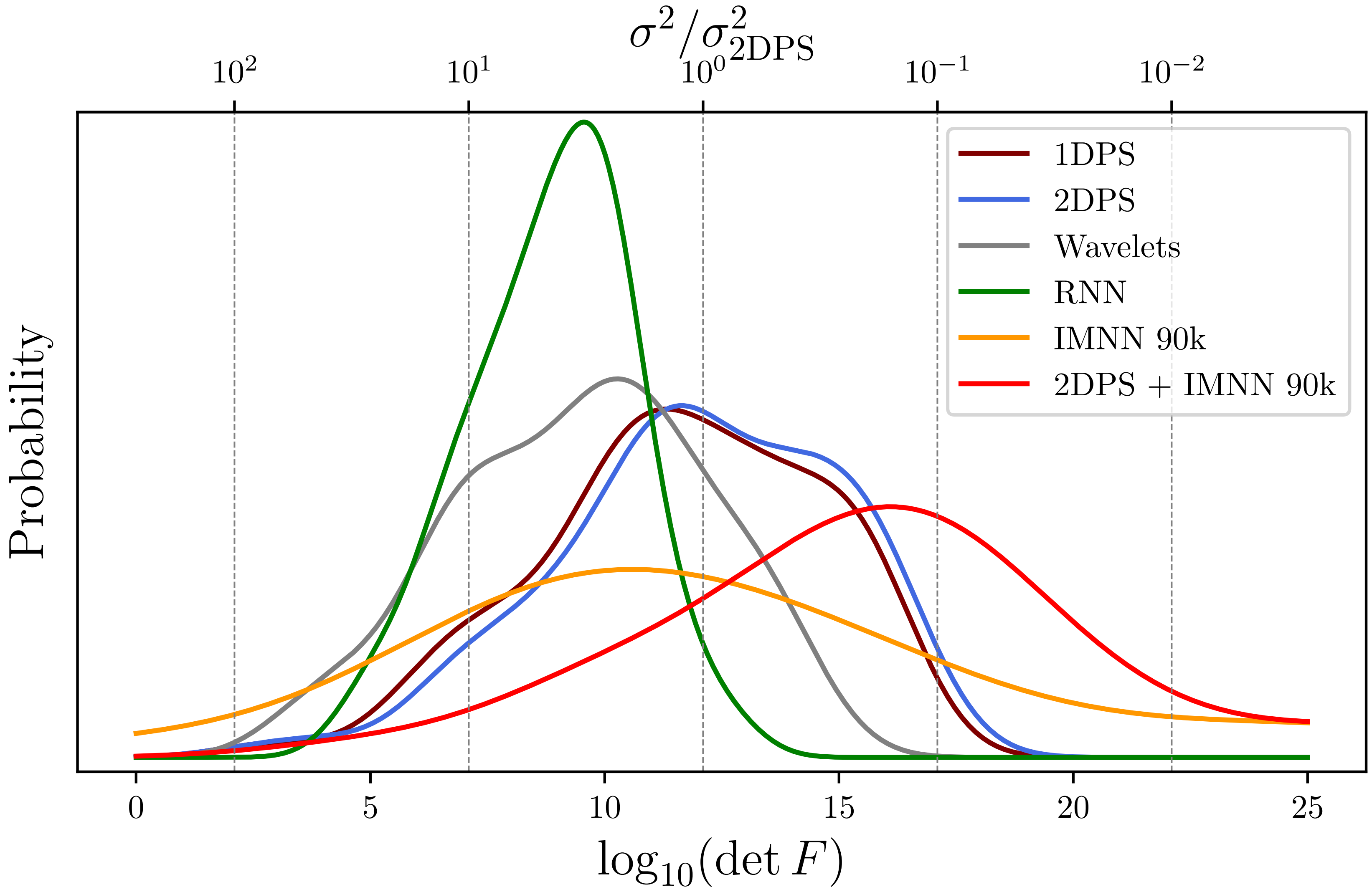
Prelogović & Mesinger (2024)

IMNN overspecializes to the Fisher of the *fiducial*… The mean Fisher over the prior is actually lower than the PS



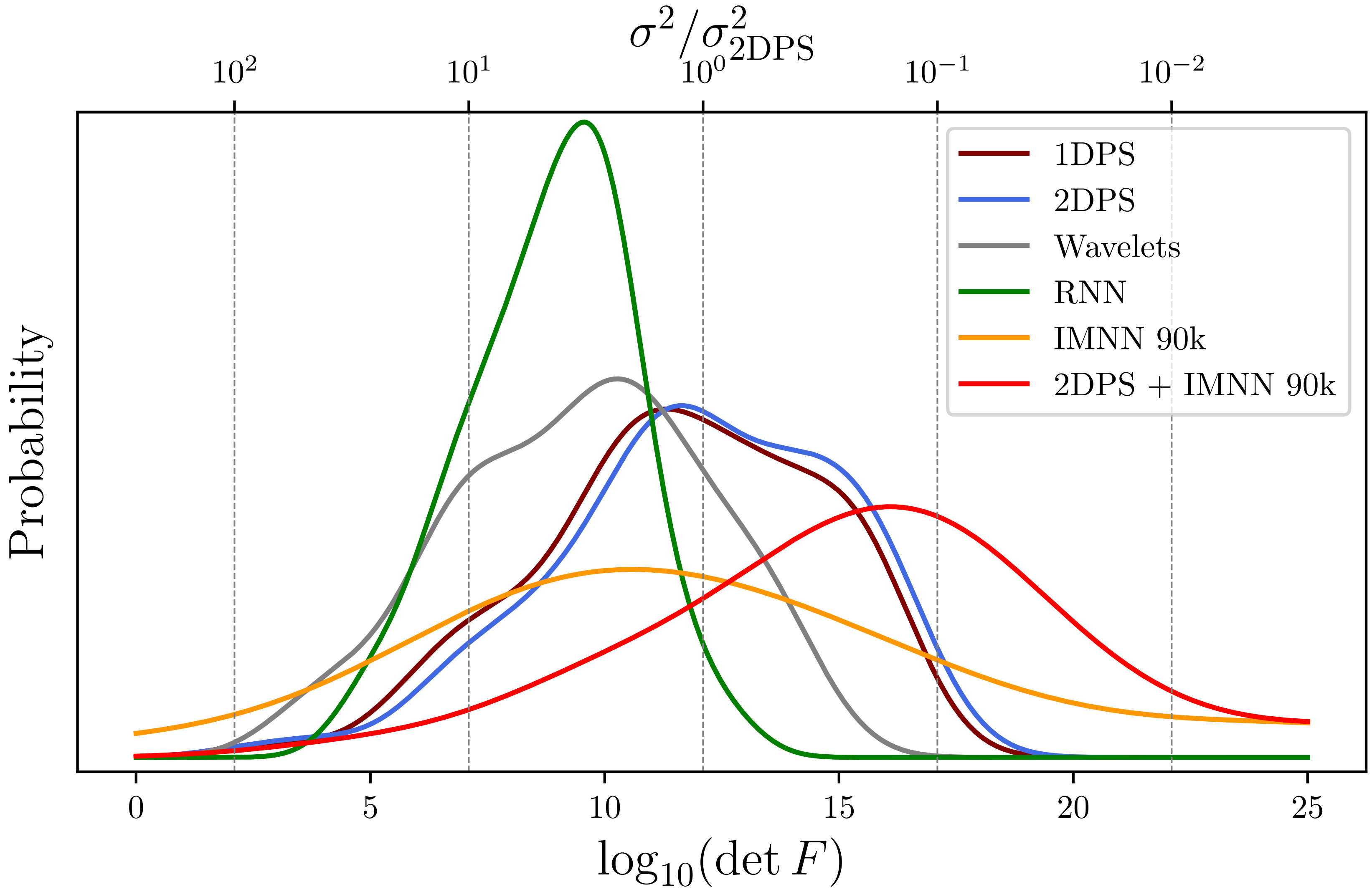
Prelogović & Mesinger (2024)

Comparing various 21cm summaries using the prior-weighted Fisher information



Prelogović & Mesinger 2024

Combining 2D PS & IMNN gives a constraining summary throughout the prior volume



Prelogović & Mesinger 2024

Summary

Milestones and KPIs

Bottlenecks:

Construction of databases:

- SBI database ~100 000 21cmFAST simulations
- IMNN database ~ 20 000
- Fisher database ~ 50 000

→ CPU resources (see Monday's talk)

NN trainings:

- IMNN – 5 000 GPU h per model
- RNN – 1 000 GPU h per model

→ GPU resources

- SBI NDEs – relatively cheap (~few GPUh per model)

Milestones and KPIs

Code release:
https://github.com/dprelogo/21cmLikelihoods
https://github.com/dprelogo/conditional_kde
https://github.com/21cmfast/21cmEMU
https://github.com/dprelogo/21cmIMNN

Publications:

Exploring the likelihood of the 21-cm power spectrum with simulation-based inference
arXiv:2305.03074 (Prelogovic & Mesinger; MNRAS 524, 2023)

21CMEMU: an emulator of 21CMFAST summary observables
arXiv:2309.05697 (Breitman, AM+; MNRAS 527, 2024)

How informative are summaries of the cosmic 21cm signal?
arXiv:2401.12277 (Prelogovic & Mesinger; A&A submitted)