SBI has its own Dodelson-Schneider effect (but it knows that it does)

Jed Homer, Oliver Friedrich & Daniel Gruen

2412.02311

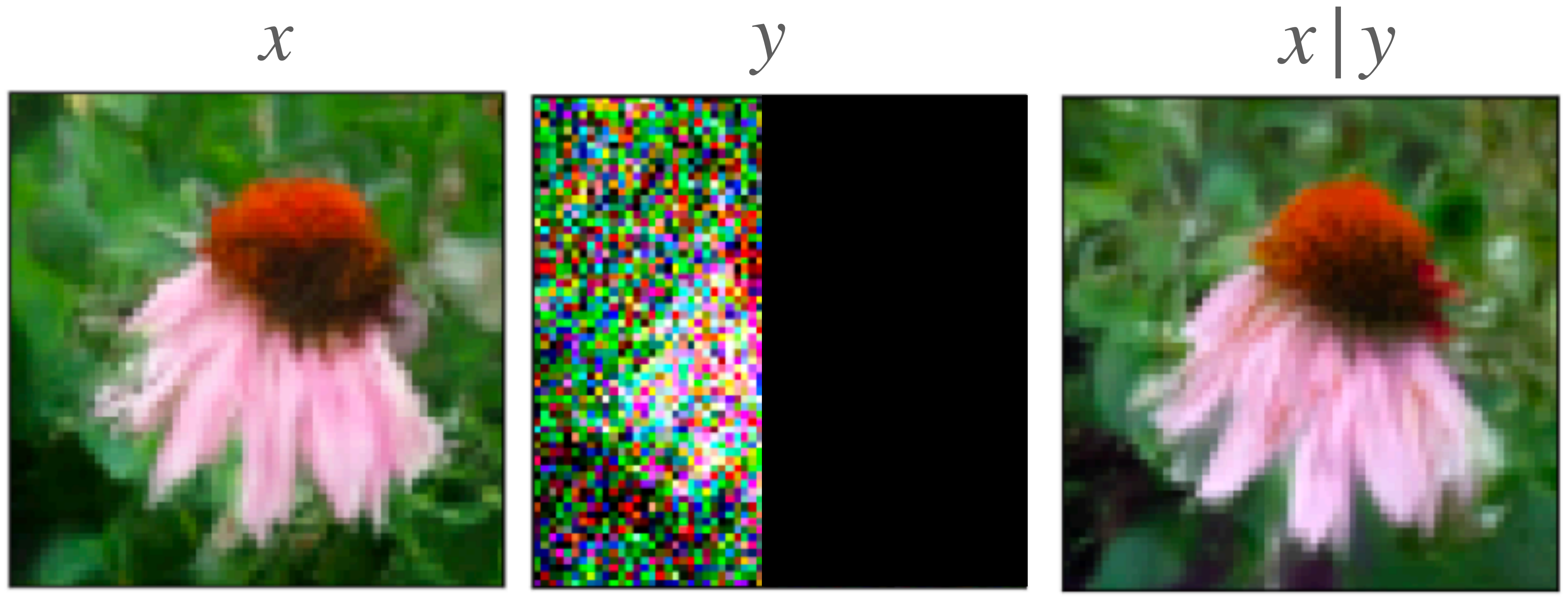
ODSL AI4Science - 14/02/25

LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

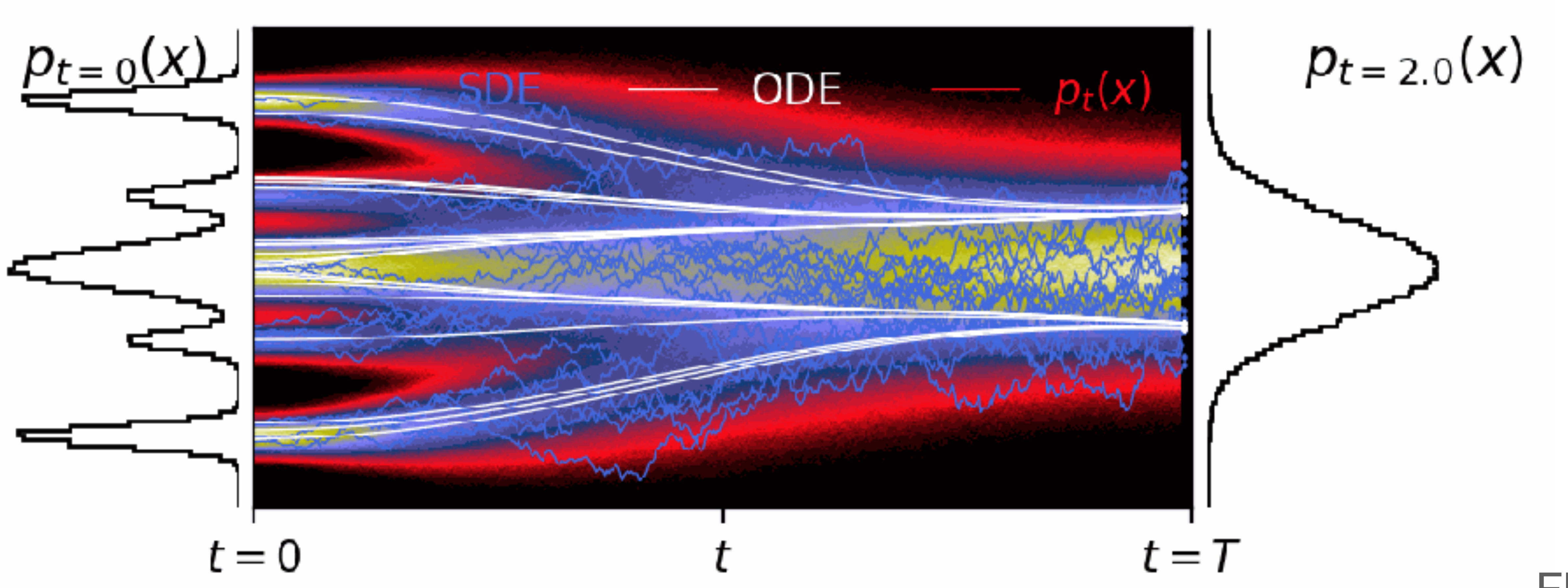
Munich Center for Machine Learning

Hello!

Bayesian inference in cosmology with generative models

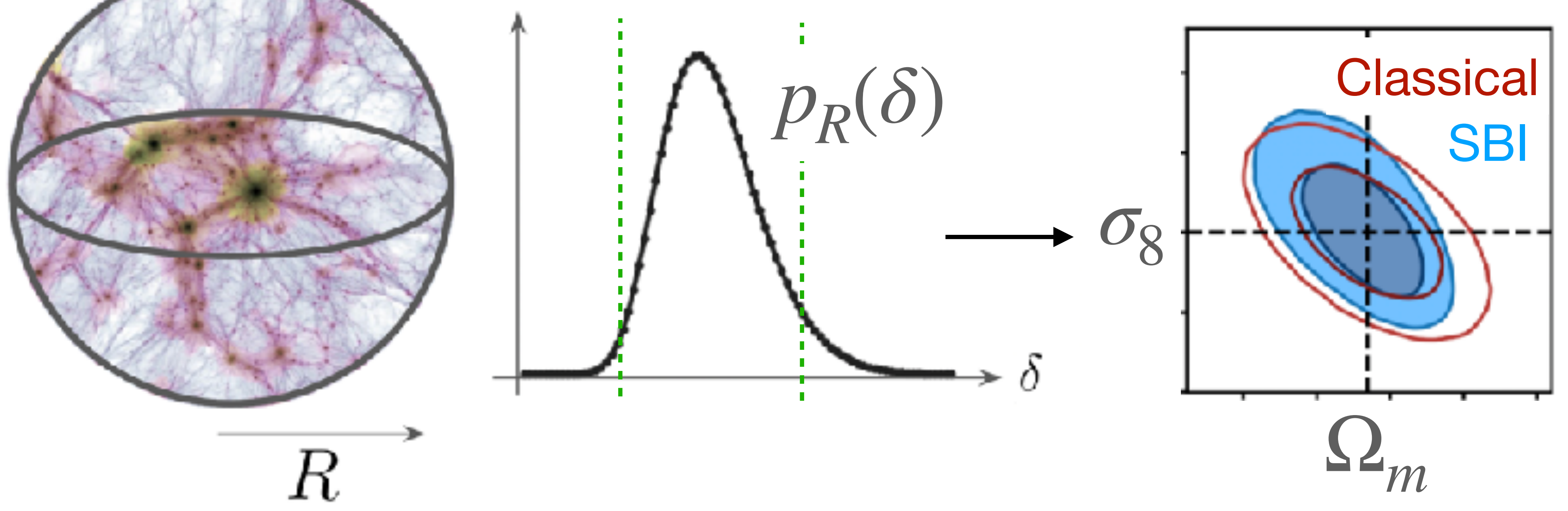


Deep generative signal priors with JAX-NIFTy

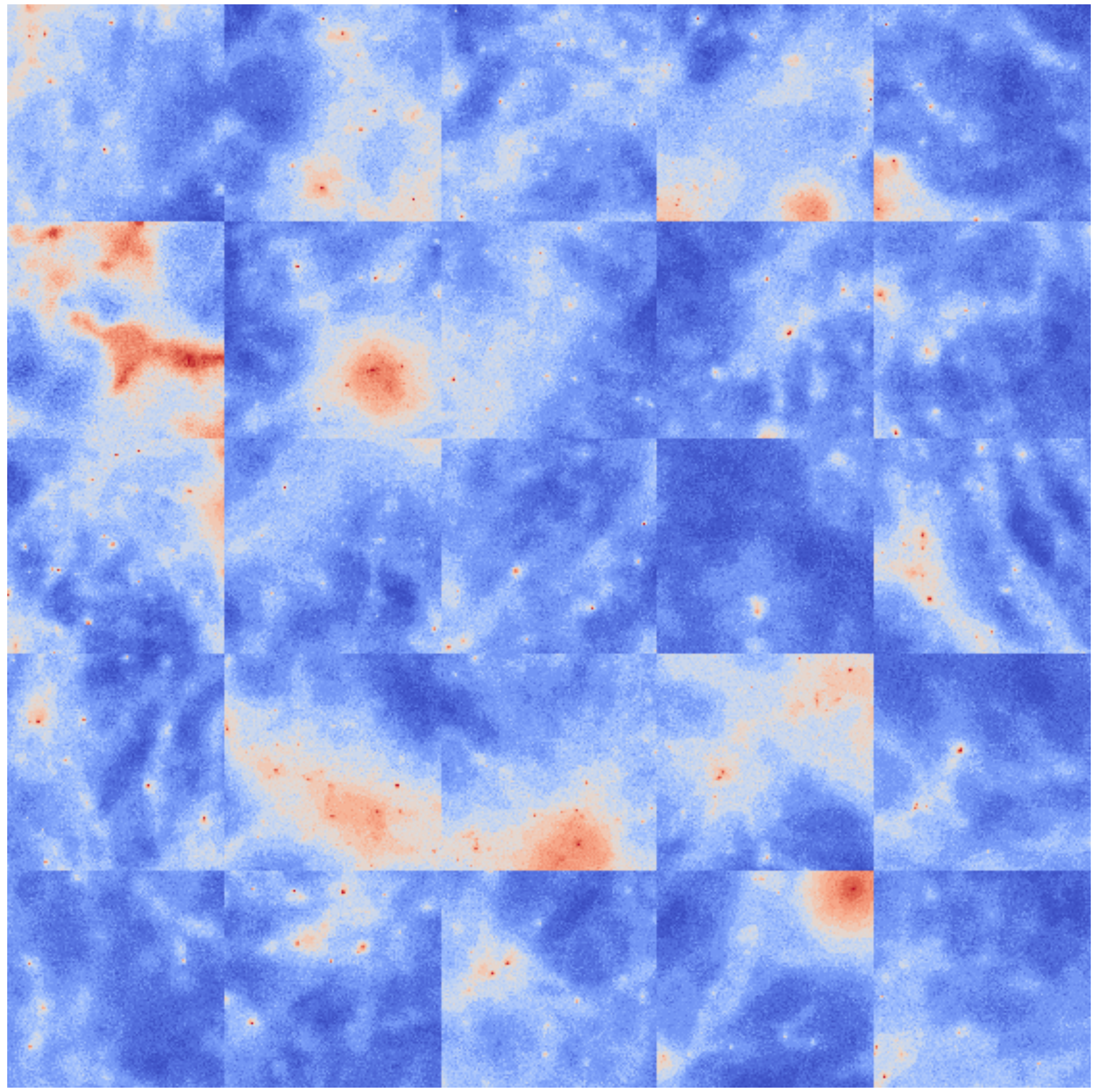


Generative diffusion processes: homerjed/sbgm

Homer, Friedrich, Uhlemann & Gruen (in prep.)



Unlocking information content of the 1pt PDF with SBI homerjed/sbiax

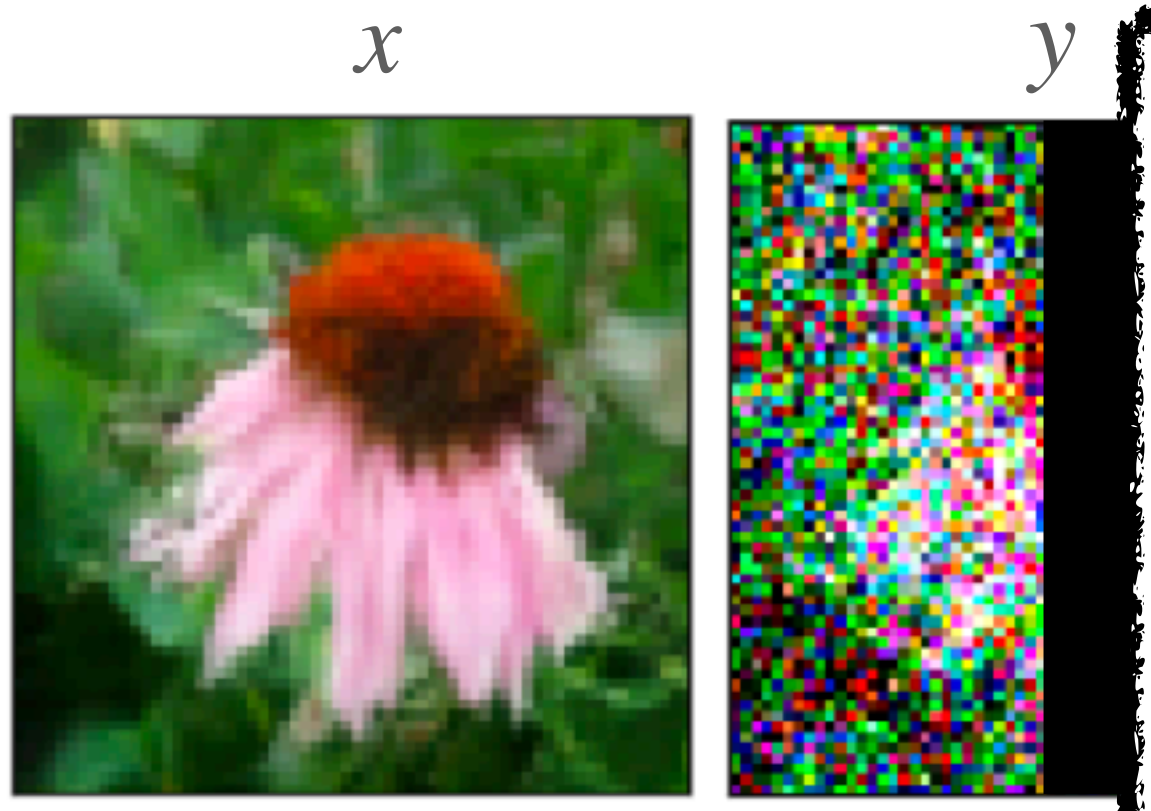


Flow-matching baryonification (Homer, Bucko & Kacprzak, ETH Zürich)

homerjed/rectified_flows

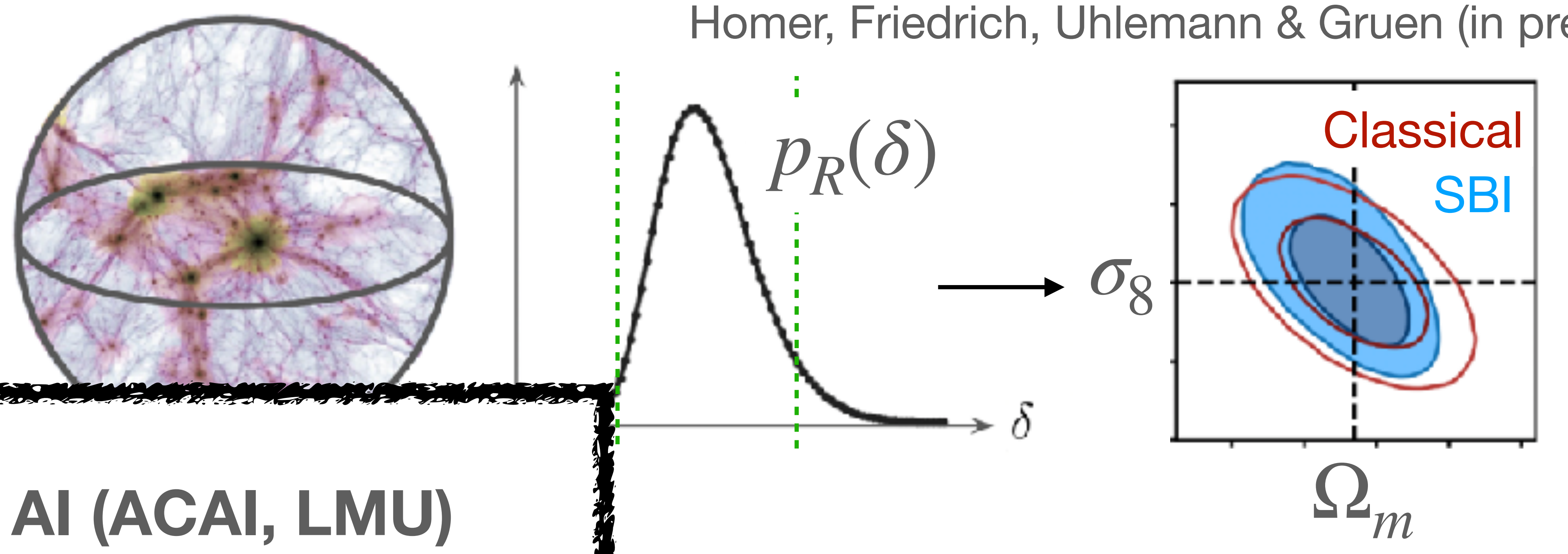
Hello!

Bayesian inference in cosmology with generative models



Deep generative signal priors

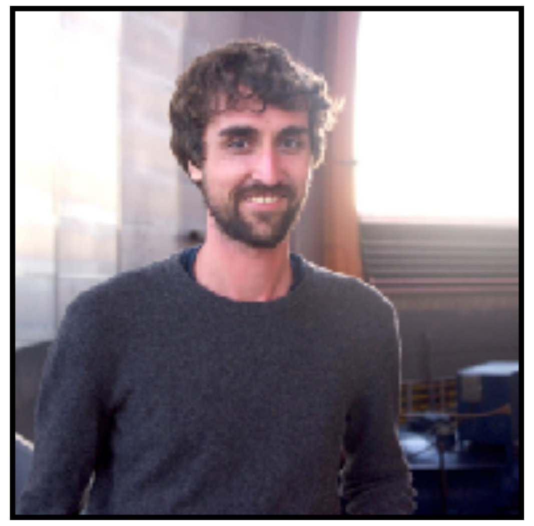
Homer, Friedrich, Uhlemann & Gruen (in prep.)



…of the 1pt PDF with SBI homerjed/sbiax

Astrophysics, Cosmology and AI (ACAI, LMU)

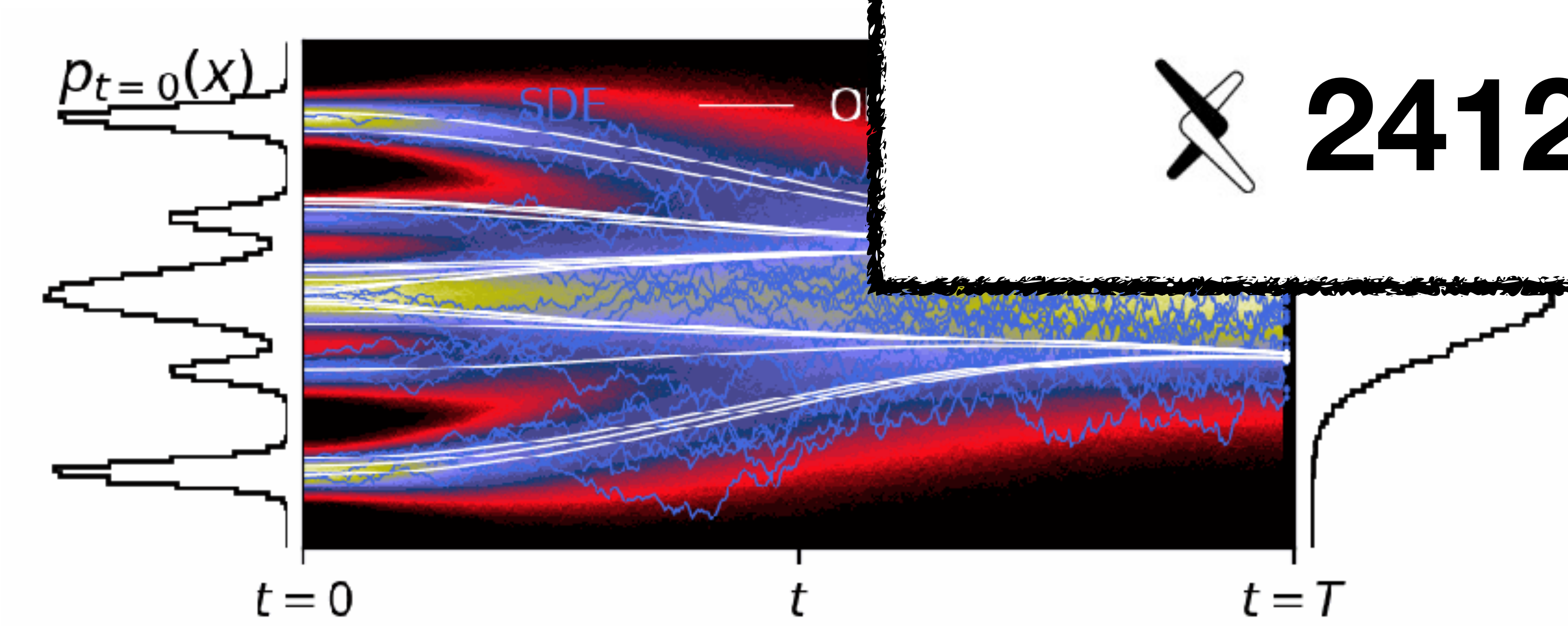
Jed Homer (USM, LMU)



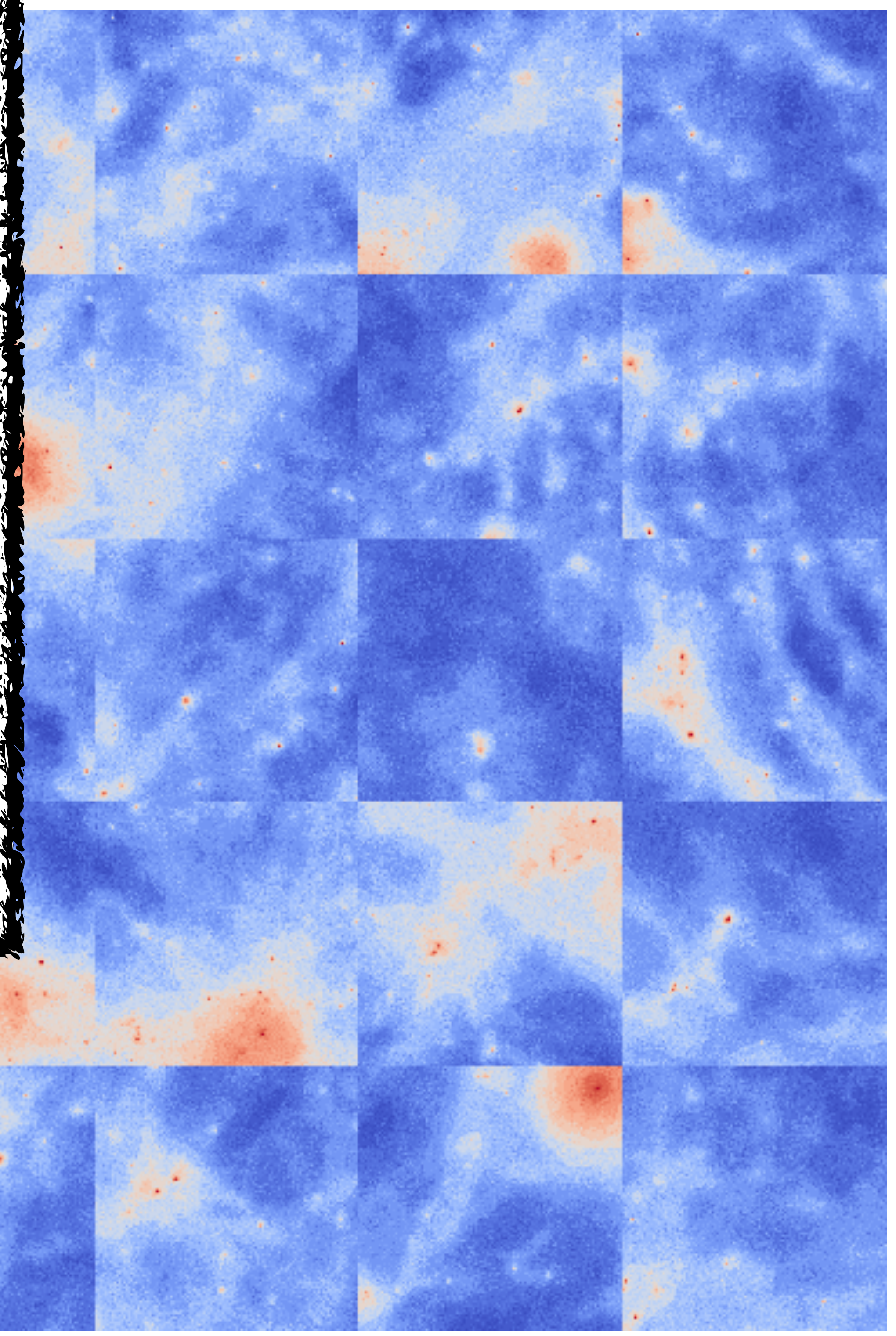
Daniel Gruen (USM, LMU)

Oliver Friedrich (USM, LMU)

2412.02311



Generative diffusion processes: homerjed/sbgm



Flow-matching baryonification (Homer, Bucko & Kacprzak, ETH Zürich)

homerjed/rectified_flows

Posterior $\propto$ Likelihood $\times$ Prior

$$p(\xi[\pi] \,|\, \hat{\xi}) \propto p(\hat{\xi} \,|\, \xi[\pi]) \times p(\pi)$$

$$p(\pi \,|\, \hat{\xi}) \propto \exp\left(-\frac{1}{2}\chi^2[\pi, \hat{\xi}, \hat{\Sigma}] \right) p(\pi)$$

Gaussian likelihood

$$\hat{\Sigma}, \hat{\Sigma}^{-1}, \xi[\pi]$$

Covariance, precision and model

Posterior $\propto$ Likelihood $\times$ Prior

Reasons for using SBI:

1. **Likelihood** $p(\hat{\xi} \,|\, \xi[\pi])$ is **non-Gaussian,**
2. **Model** $\xi[\pi]$ is a complex non-linear function of π,
3. Statistic $\hat{\xi}$ is **inaccurately predicted in simulations,**
4. Modelling **covariance** $\Sigma[\pi]$ dependence on π.

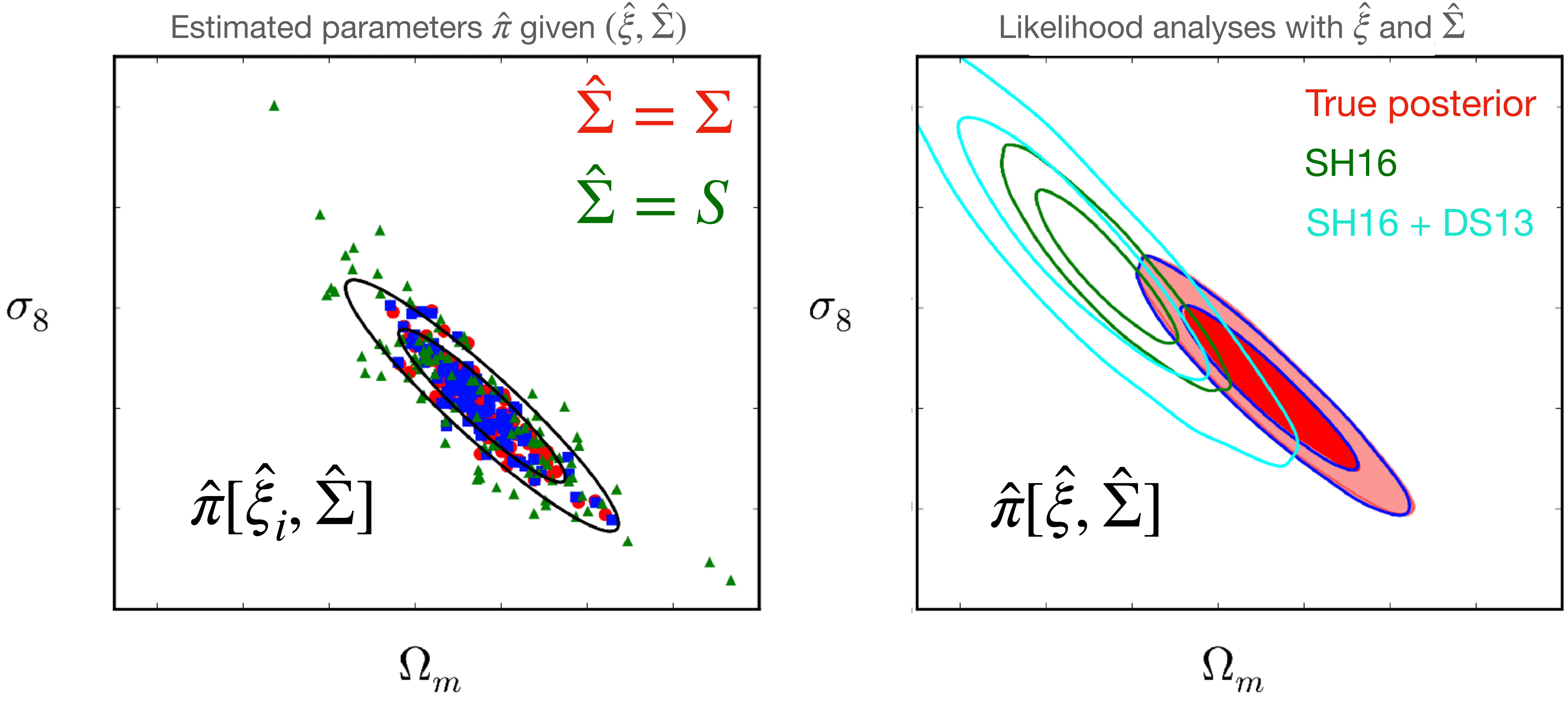
...but does it do what it says it does?

Covariance, precision and model

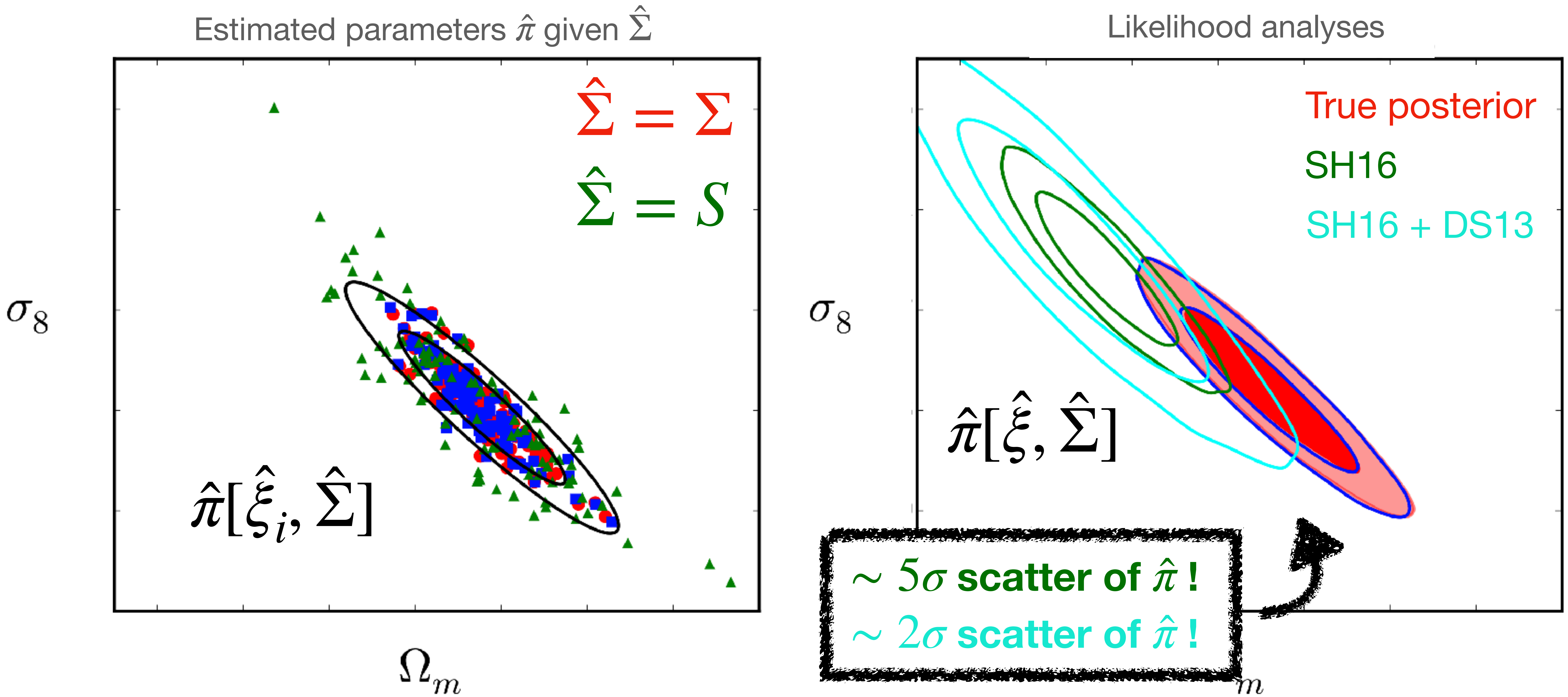
Covariance matrix estimation

Building an accurate Gaussian likelihood

How not knowing Σ affects best-fit $\hat{\pi}$



How not knowing Σ affects best-fit $\hat{\pi}$



$\sim 5\sigma$ **scatter of** $\hat{\pi}$ **!**

$\sim 2\sigma$ **scatter of** $\hat{\pi}$ **!**

Problem: Noise in $\hat{\Sigma} = S \neq \Sigma$ scatters best-fit $\hat{\pi}$. **Not accounted for in Gaussian ansatz.**

How not knowing Σ affects best-fit $\hat{\pi}$

The Dodelson-Schneider Correction

$\rightarrow$ Correct **posterior** for noise in ***location*** $\hat{\pi}$ of contours from $\hat{\Sigma} = S$

$$F^{-1}_{\hat{\Sigma}^{-1}} \approx \left[1 + \frac{n_\xi - n_\pi}{n_s - n_\xi}\right] F^{-1}_{\Sigma} \qquad (n_s \gg n_\xi \gg n_\pi)$$

Dodelson & Schneider 2013

Problem: Noise in $\hat{\Sigma} = S \neq \Sigma$ scatters best-fit $\hat{\pi}$. **Not accounted for in Gaussian ansatz.**

An analytic solution* to fixing posterior coverage (when you don't know Σ)

$$p(\pi \,|\, \hat{\xi}, S) \propto p(\hat{\xi} \,|\, \pi, \Sigma) p(\Sigma \,|\, S) p(\pi)$$

Posterior given S

posterior on Σ

- Choose a **prior** such that *over repeated experiments* $(\hat{\xi}, S)$:

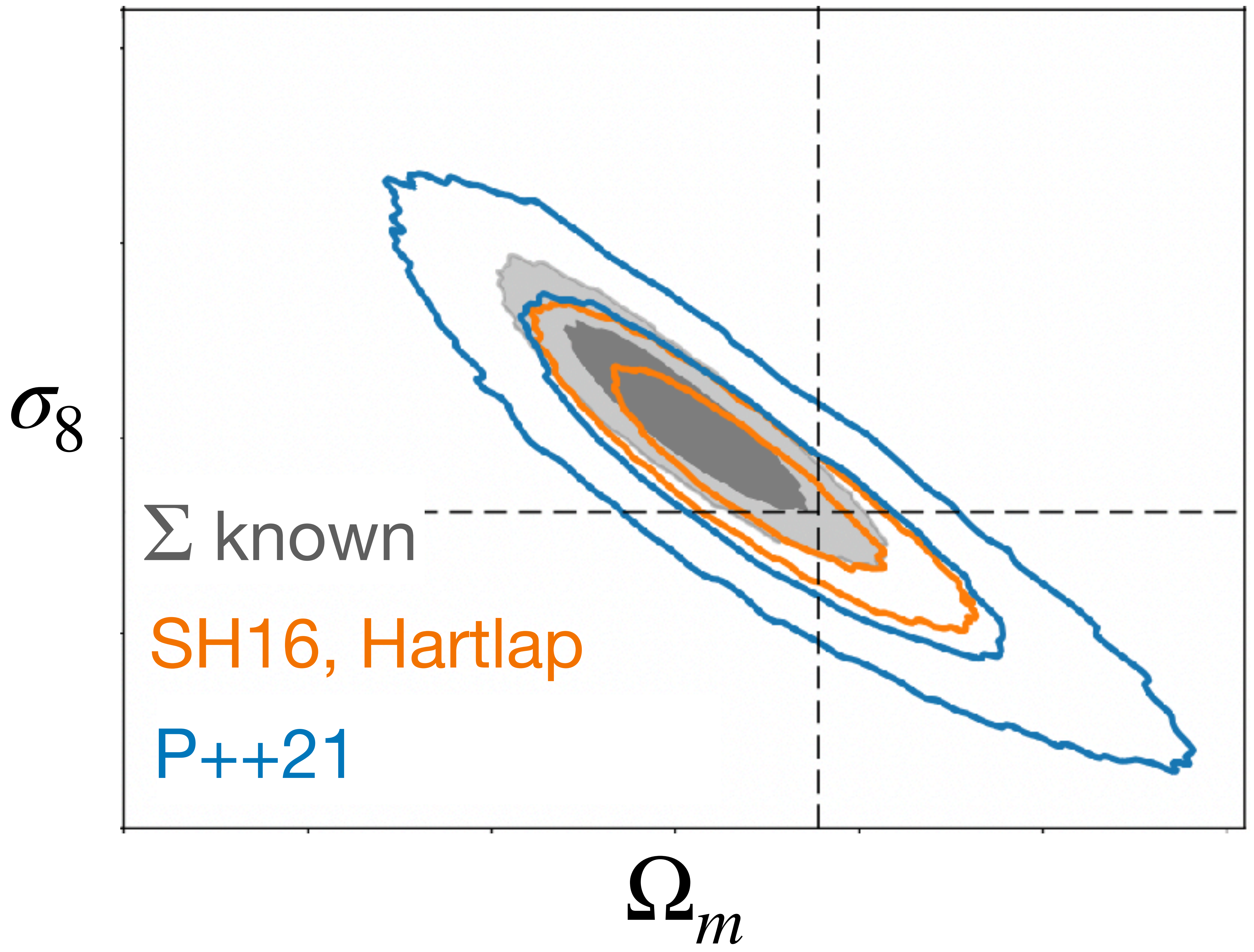
$$\langle (\hat{\pi} - \pi)(\hat{\pi} - \pi)^T \rangle_{\hat{\xi}, S} \approx \langle (\pi - \hat{\pi})(\pi - \hat{\pi})^T \rangle_{\hat{\xi}, S}$$

Frequentist Bayesian

Percival++14, Sellentin+16, **Percival++21**

The Bayesian approach...

Example posterior



- Corrected coverage + accounting for unknown Σ

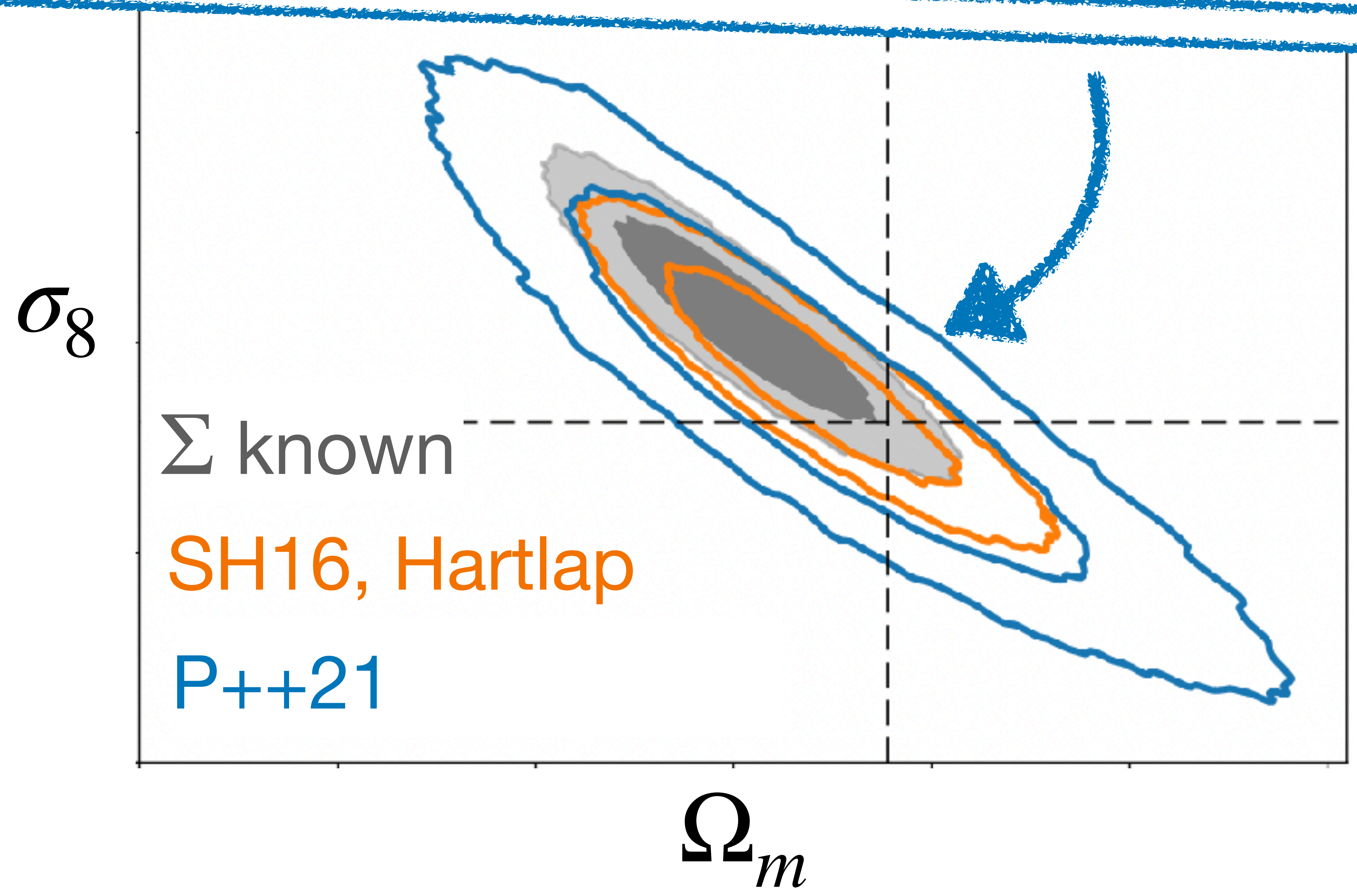
$\implies$ **larger posterior widths**

Percival++21

Examples: homerjed/frequentist_matching_priors

The Bayesian approach...

This is the solution a machine is aiming for!



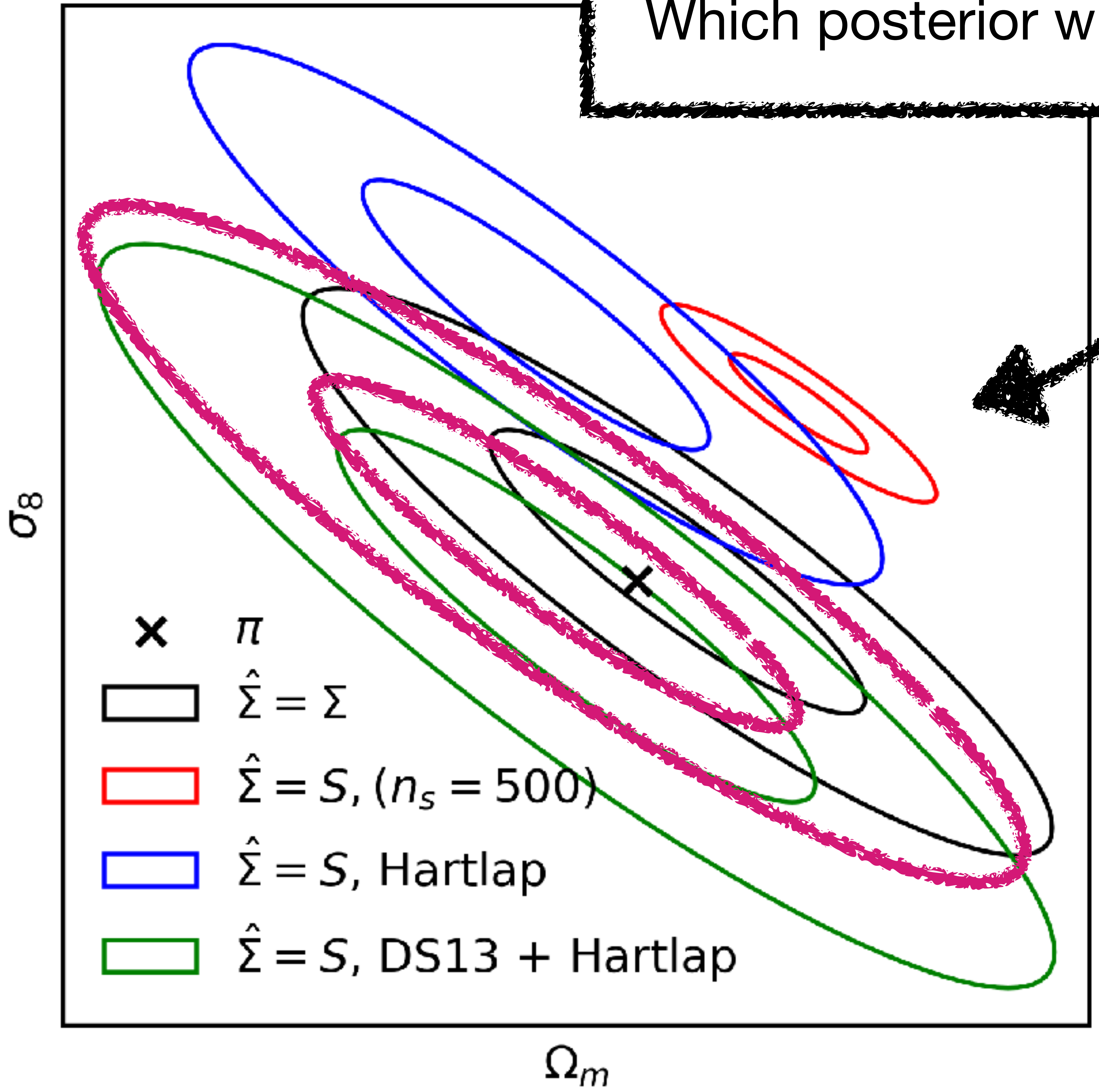
- Corrected coverage + accounting for unknown Σ

$\Longrightarrow$ **larger posterior widths**

Percival++21

Examples: homerjed/frequentist_matching_priors

Questions about SBI

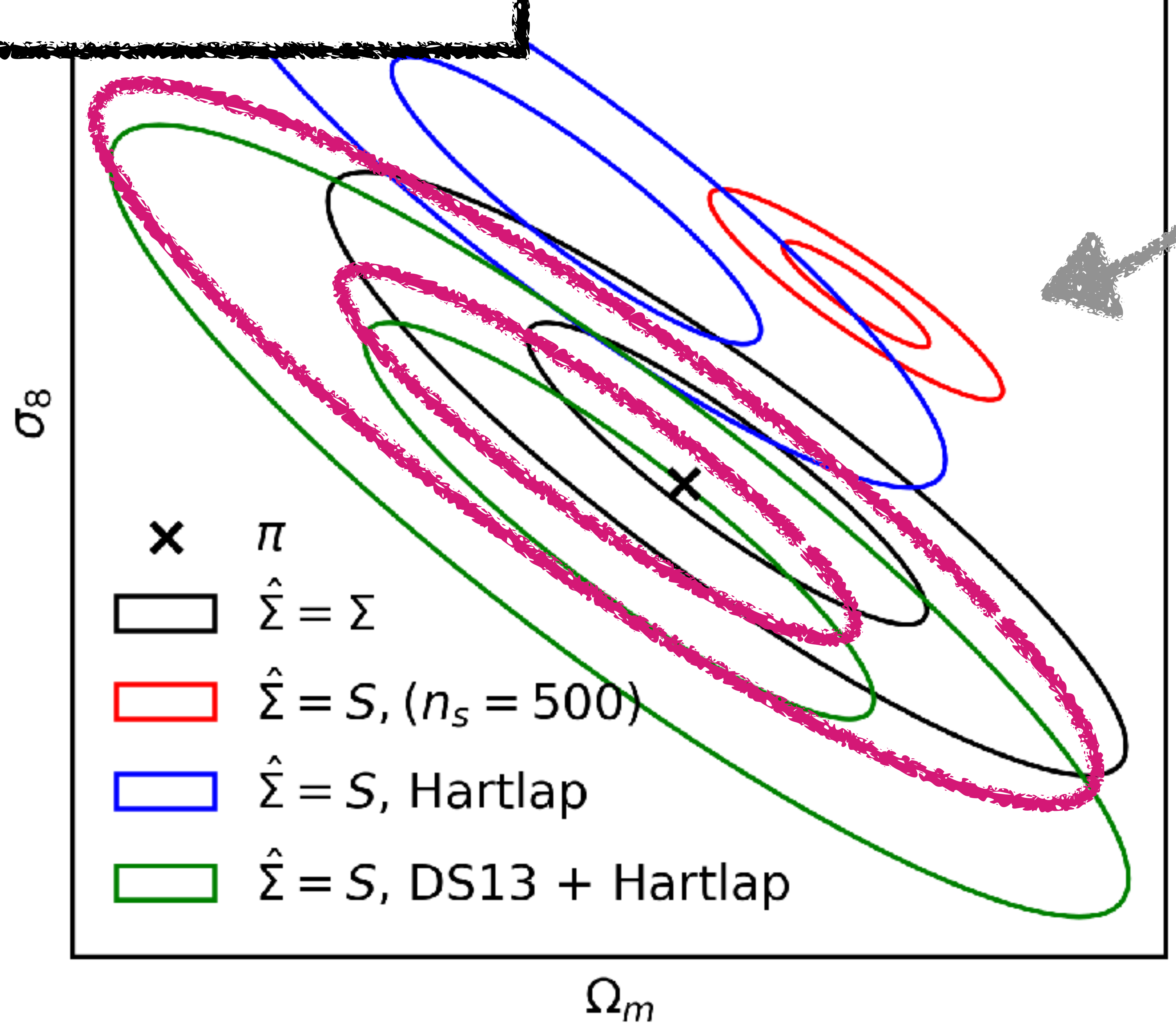


Which posterior will SBI obtain on <u>average</u>?

Questions about SBI

Are the <u>posteriors inflated</u> w.r.t. a Gaussian likelihood analysis for the same n_s?

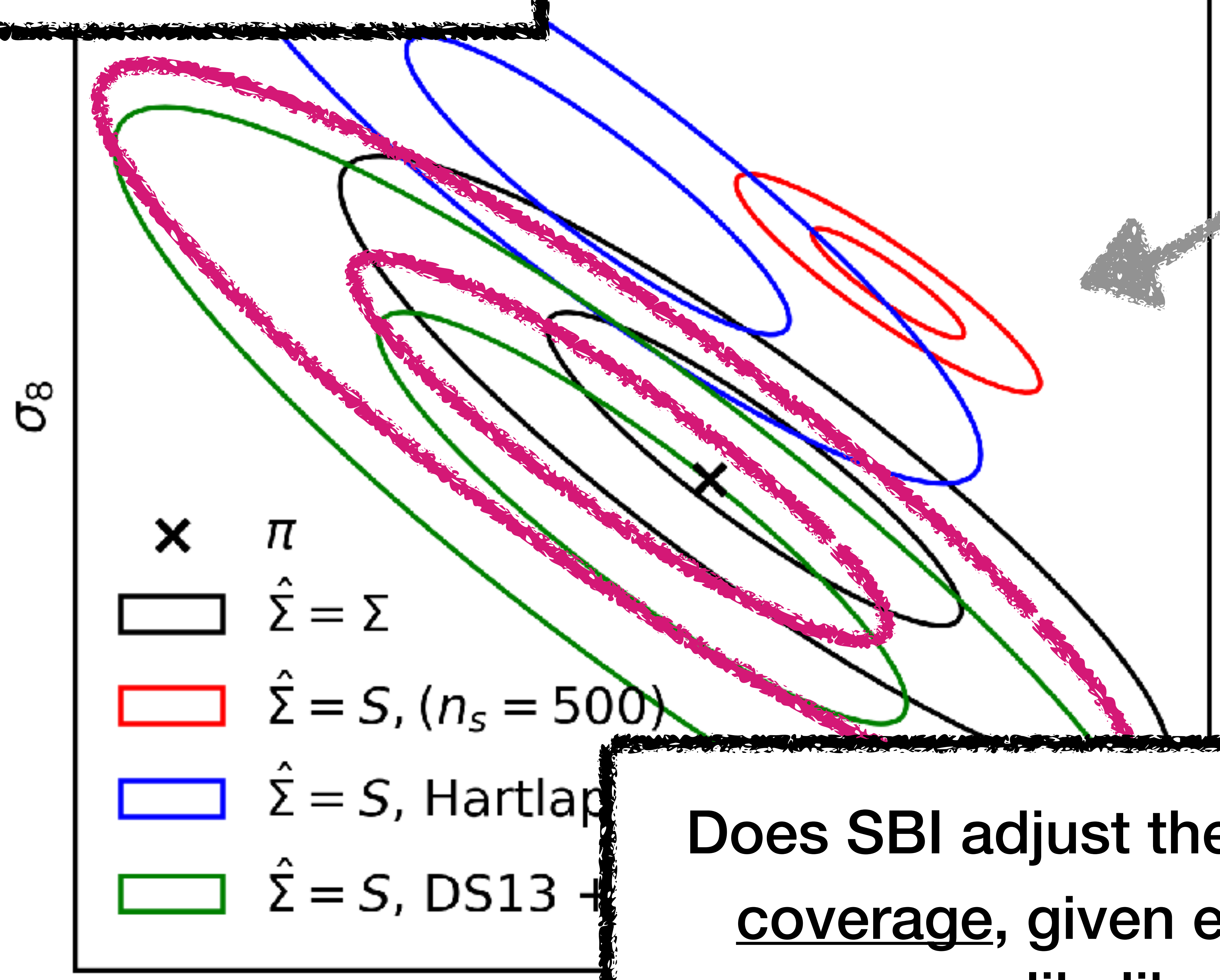
Which posterior will SBI obtain on <u>average</u>?



Questions about SBI

Are the <u>posteriors inflated</u> w.r.t. a Gaussian likelihood analysis for the same n_s?

Which posterior will SBI obtain on <u>average</u>?

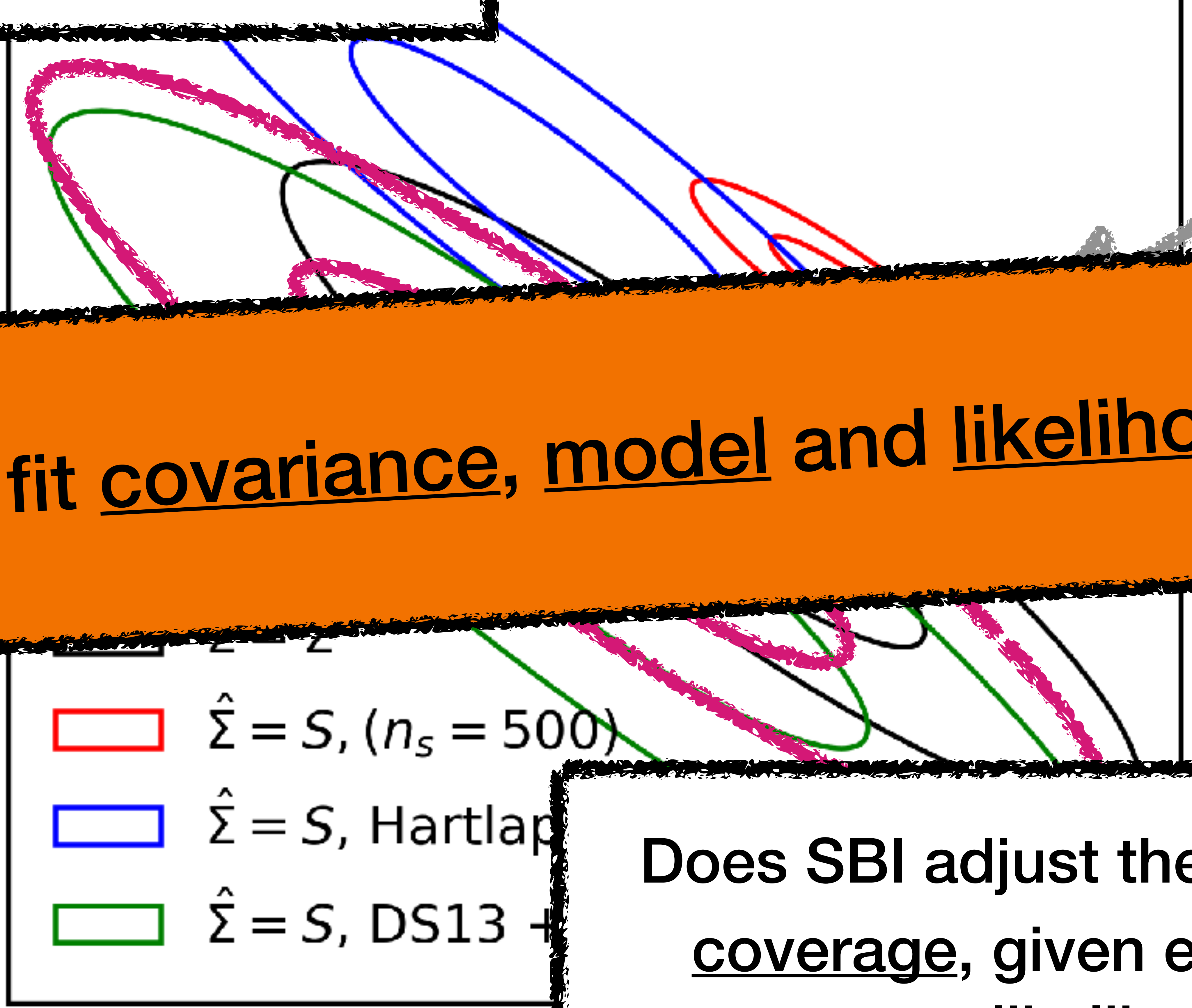


Does SBI adjust the posterior width to fix its <u>coverage</u>, given effects of $\hat{\Sigma} \neq \Sigma$ on the likelihood function?

Questions about SBI

Are the posteriors inflated w.r.t. a Gaussian likelihood analysis for the same n_s?

Which posterior will SBI obtain on average?



...in SBI we fit covariance, model and likelihood shape!

Does SBI adjust the posterior width to fix its coverage, given effects of $\hat{\Sigma} \neq \Sigma$ on the likelihood function?

Challenging SBI

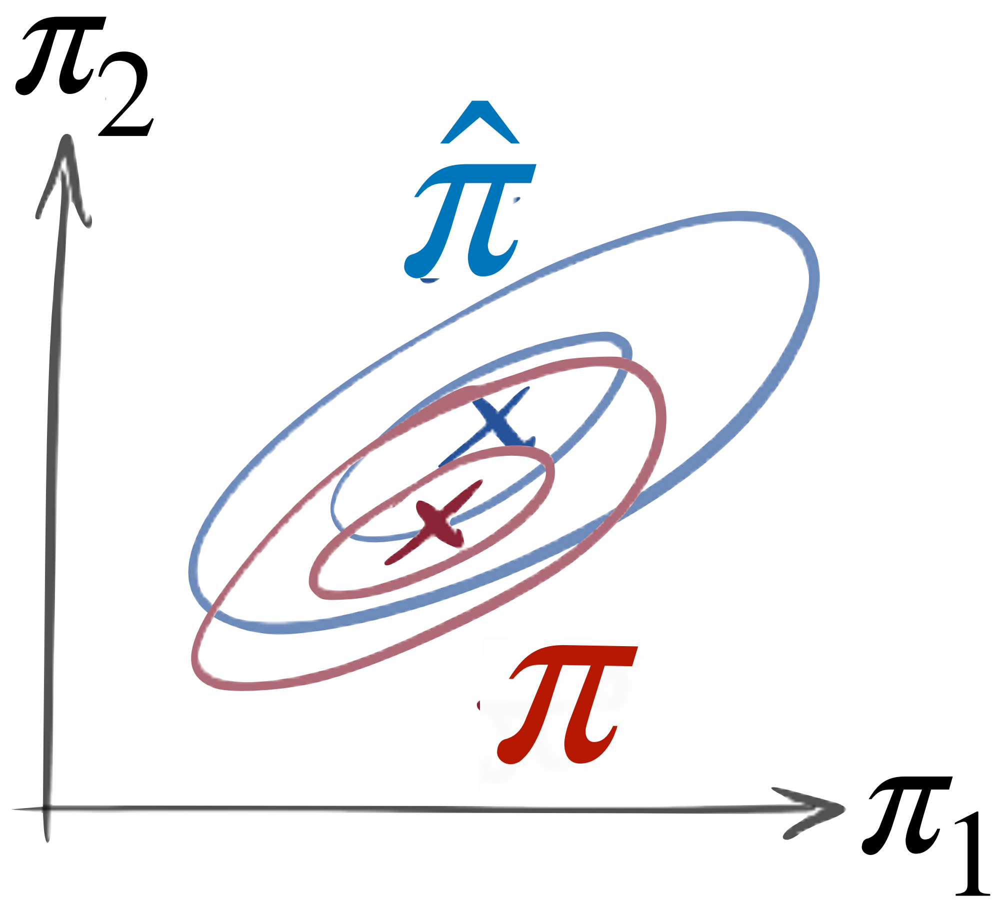
Testing likelihoods built by machines

An experiment with SBI

- $\hat{\xi} \leftarrow$ linearised model for DES-Y3 shear 2pt functions.

$$\pi \rightarrow \hat{\xi} \sim \mathcal{G}[\hat{\xi} \,|\, \xi[\pi], \Sigma] \rightarrow \hat{\pi} = f(\hat{\xi}) \rightarrow p_\phi(\xi \,|\, \pi) \rightarrow$$

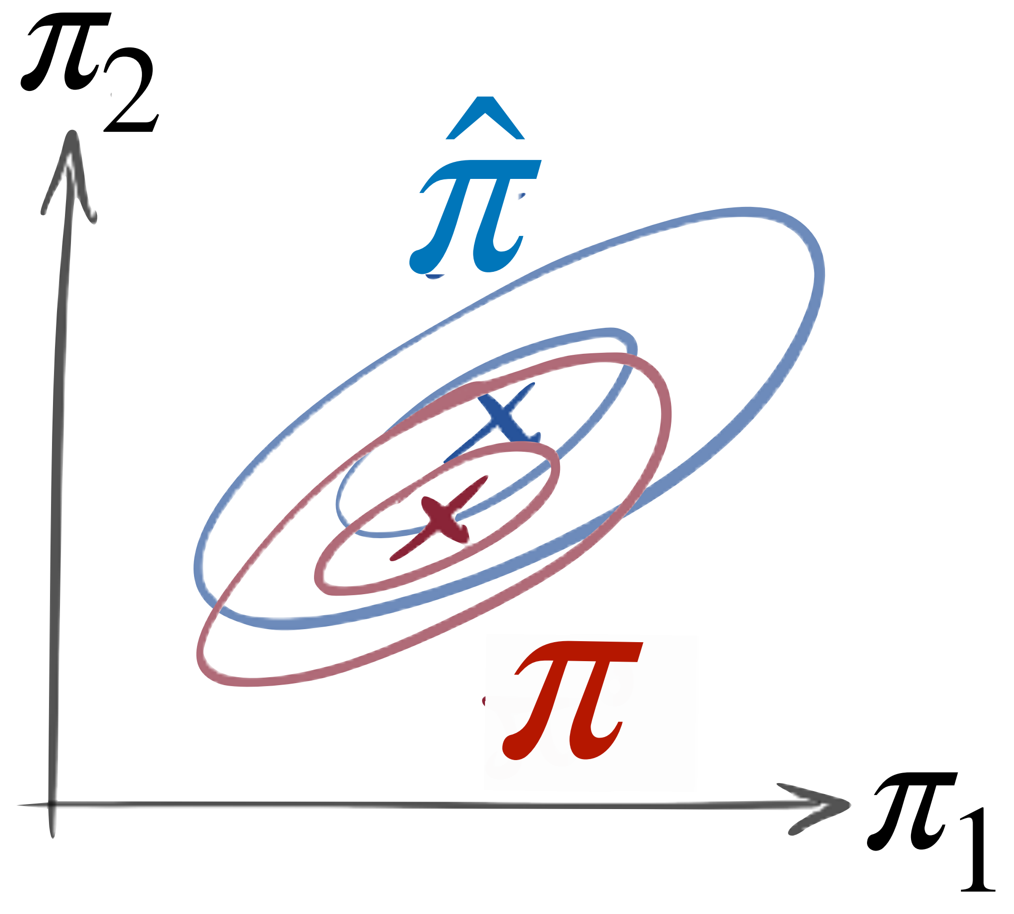
Measure Compress Infer



An experiment with SBI

$\pi \rightarrow \hat{\xi} \sim \mathcal{G}[\hat{\xi} \,|\, \xi[\pi], \Sigma] \rightarrow \hat{\pi} = f(\hat{\xi}) \rightarrow p_\phi(\xi \,|\, \pi) \rightarrow$

Measure — Compress — Infer



Linear compression

$$\hat{\pi} = M\hat{\xi}$$

$\hat{\Sigma}$: $2 \times n_s$

Σ: n_s

Neural network compression

$$\hat{\pi} = f_\psi(\xi)$$

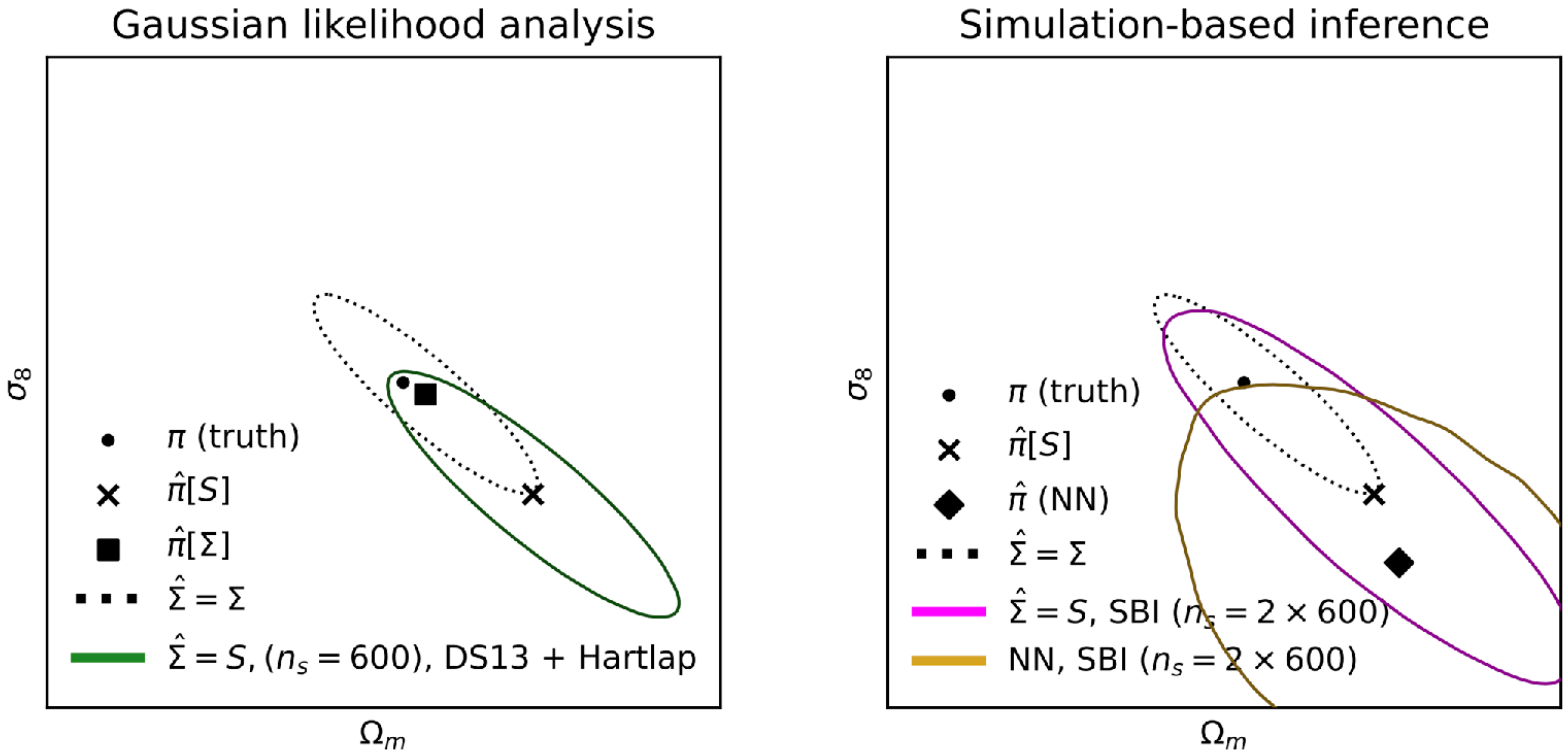
$2 \times n_s$

An … with SBI

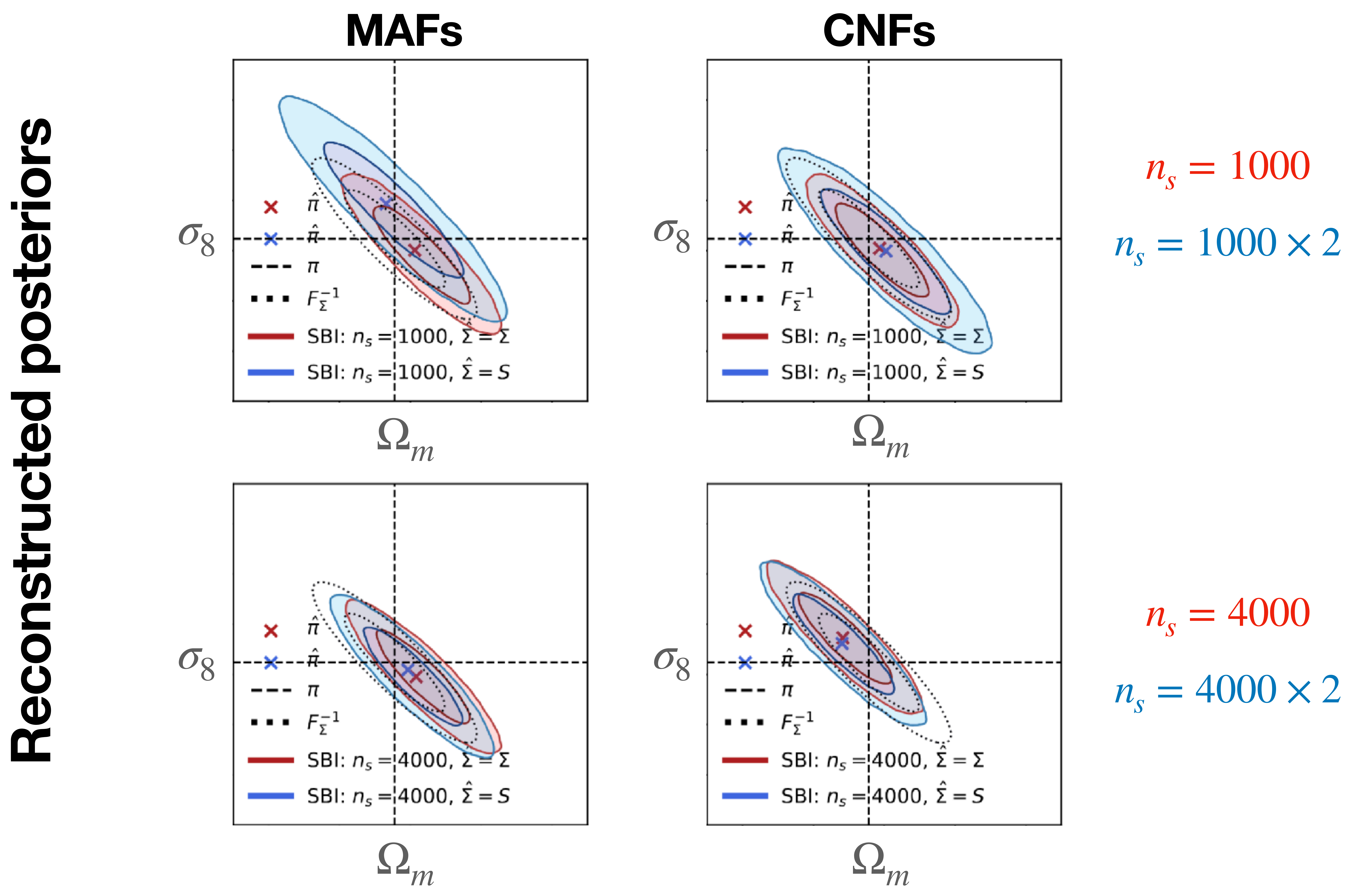
Possible questions

- Need Σ or $\hat{\Sigma}$ to parameterise our <u>optimal compression</u>
- Why not just use a neural network? It doesn't invert $\hat{\Sigma}$?
- Why not field-level, which doesn't use compression?
- <u>Compression doesn't simply "remove the noise"!</u>

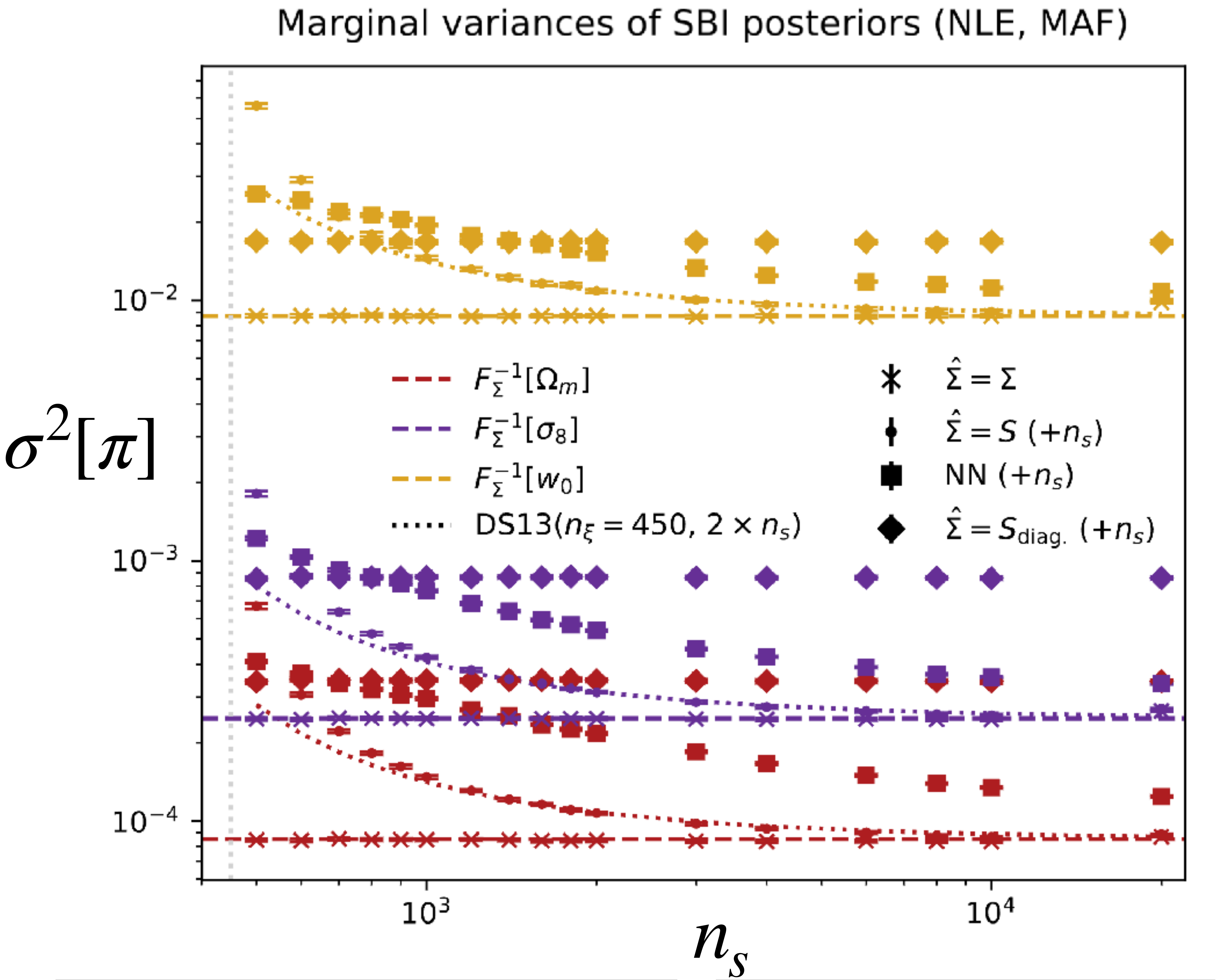
Spoiler (in one universe):



SBI is *aware* of the Dodelson-Schneider effect but it is *inefficient* in its response!

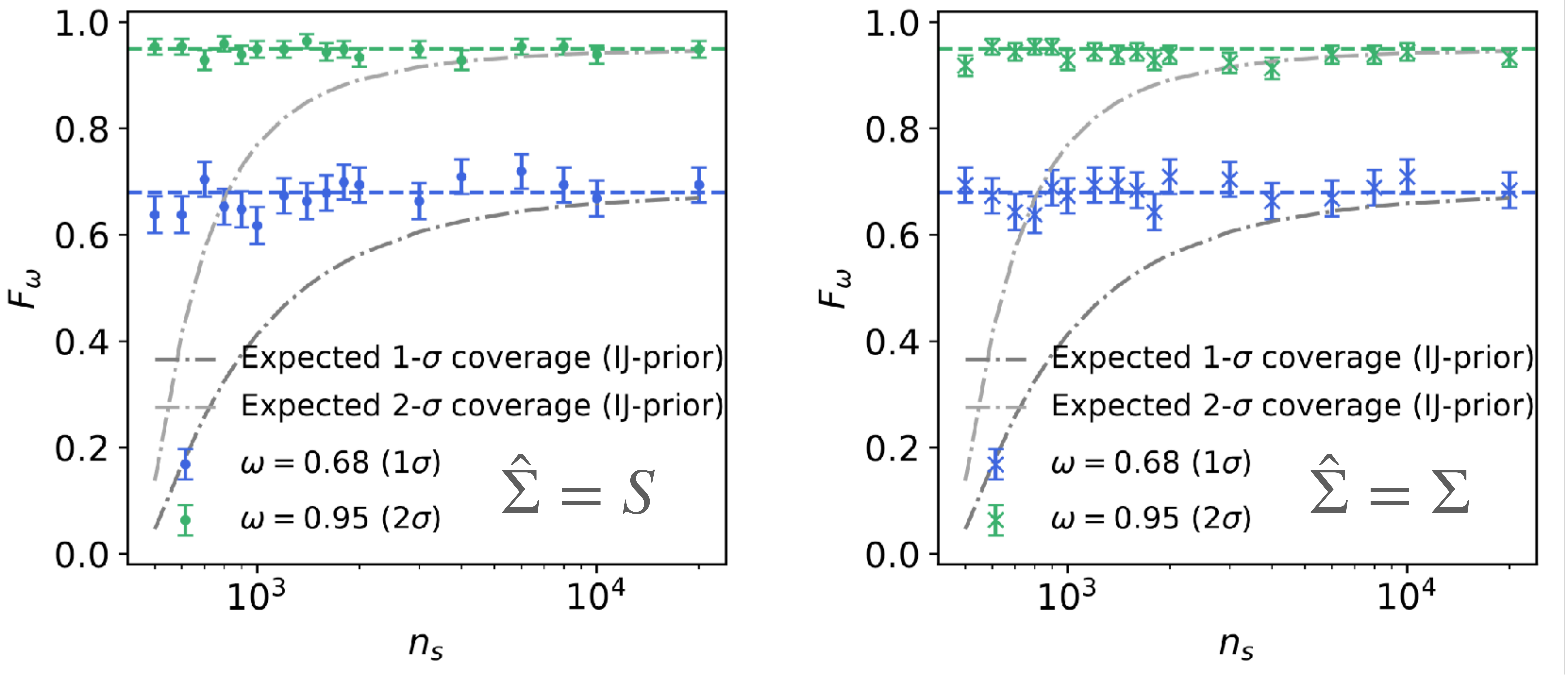

Reconstructed posteriors
MAFs
CNFs
$\hat{\pi}$
$\hat{\hat{\pi}}$
π
F_Σ^{-1}
SBI: $n_s = 1000$, $\hat{\Sigma} = \Sigma$
SBI: $n_s = 1000$, $\hat{\Sigma} = S$
SBI: $n_s = 4000$, $\hat{\Sigma} = \Sigma$
SBI: $n_s = 4000$, $\hat{\Sigma} = S$
σ_8
Ω_m
$n_s = 1000$
$n_s = 1000 \times 2$
$n_s = 4000$
$n_s = 4000 \times 2$

SBI posterior widths



- **Does SBI recover the errors it should?**

SBI posterior coverage



- **Does SBI assign correct probability to posterior credible intervals?**

SBI posterior coverage

So what is SBI doing?

- Coverages are corrected ✓
- Widths are inflated ✓
  - for low n_s there is additional posterior inflation!

→ SBI '*knows*' about the Dodelson-Schneider effect and corrects for it

Conclusions

1. Using **SBI** with...

- **optimal** $\hat{\pi}$, **true** $\xi[\pi]$, **Gaussian** $p(\hat{\xi} \,|\, \xi[\pi])$,
- cutting-edge density estimation techniques,

...obtains diluted parameter constraints compared to

- **a Gaussian likelihood analysis,**
- **and the same number of simulations** n_s **, a modest** n_ξ **and** n_π **,**
- **but SBI does what it says it does!**

FM-priors (Percival++21), PME (Friedrich+17)

Conclusions

2. Given what is **<u>required for analyses of LSS statistics</u>...**

- worse when you don't know how to **summarise your data optimally**,
- your **covariance** Σ has strong **non-diagonal** structure (+ an NN for compression),
- there are many nuisance parameters,
- and your model $\xi[\pi]$ is **complex and non-linear**, for a **non-Gaussian** statistic.

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Munich Center for Machine Learning

Thank you

2412.02311

🌐 **homerjed.github.io** ✉️ **jed.homer@physik.lmu.de**

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

2412.02311

Munich Center for Machine Learning

sbiax

Fast, lightweight and parallel simulation-based inference.

JOSS 10.21105/joss.07606 | repo status Active | arXiv 2412.02311

homerjed/sbiax

```
> pip install sbiax
> cd examples/
```

homerjed.github.io

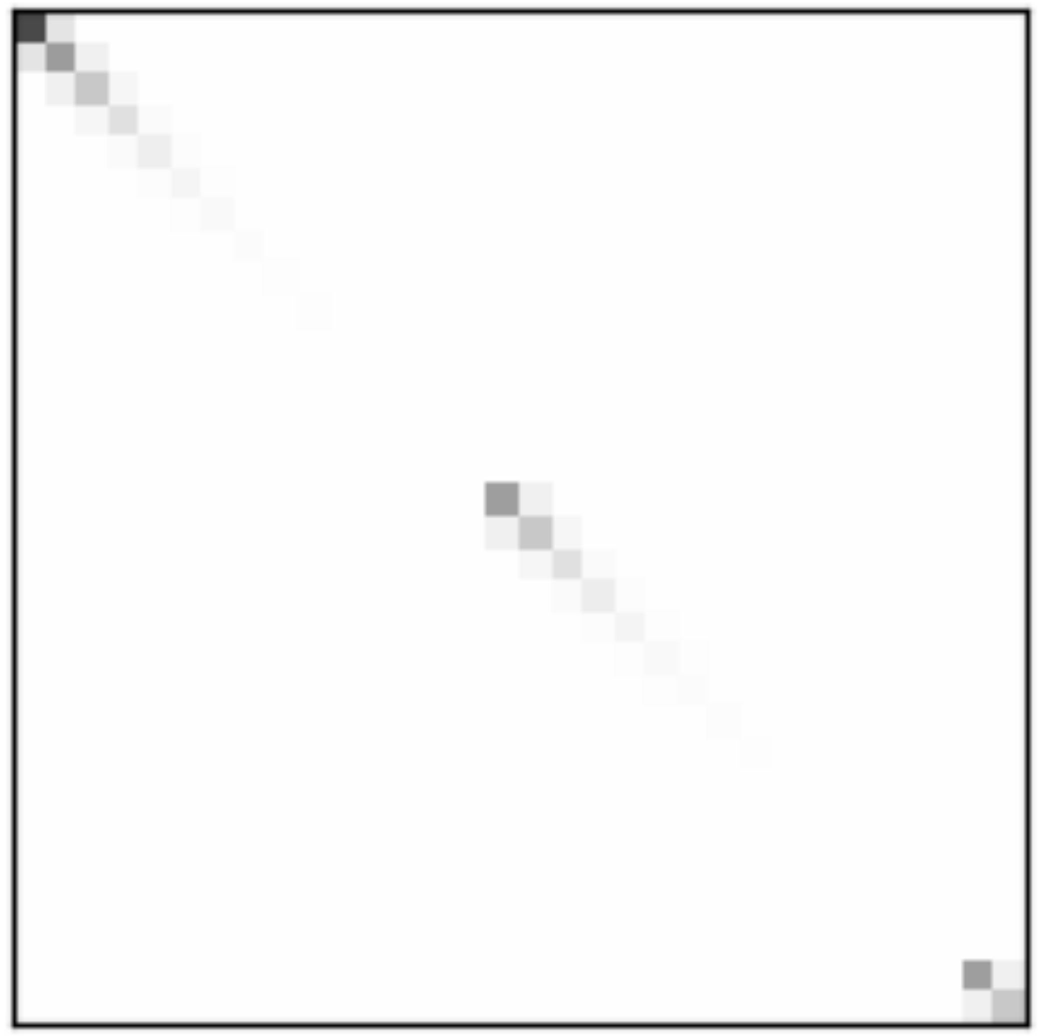
jed.homer@physik.lmu.de

Where does NN compression fail?

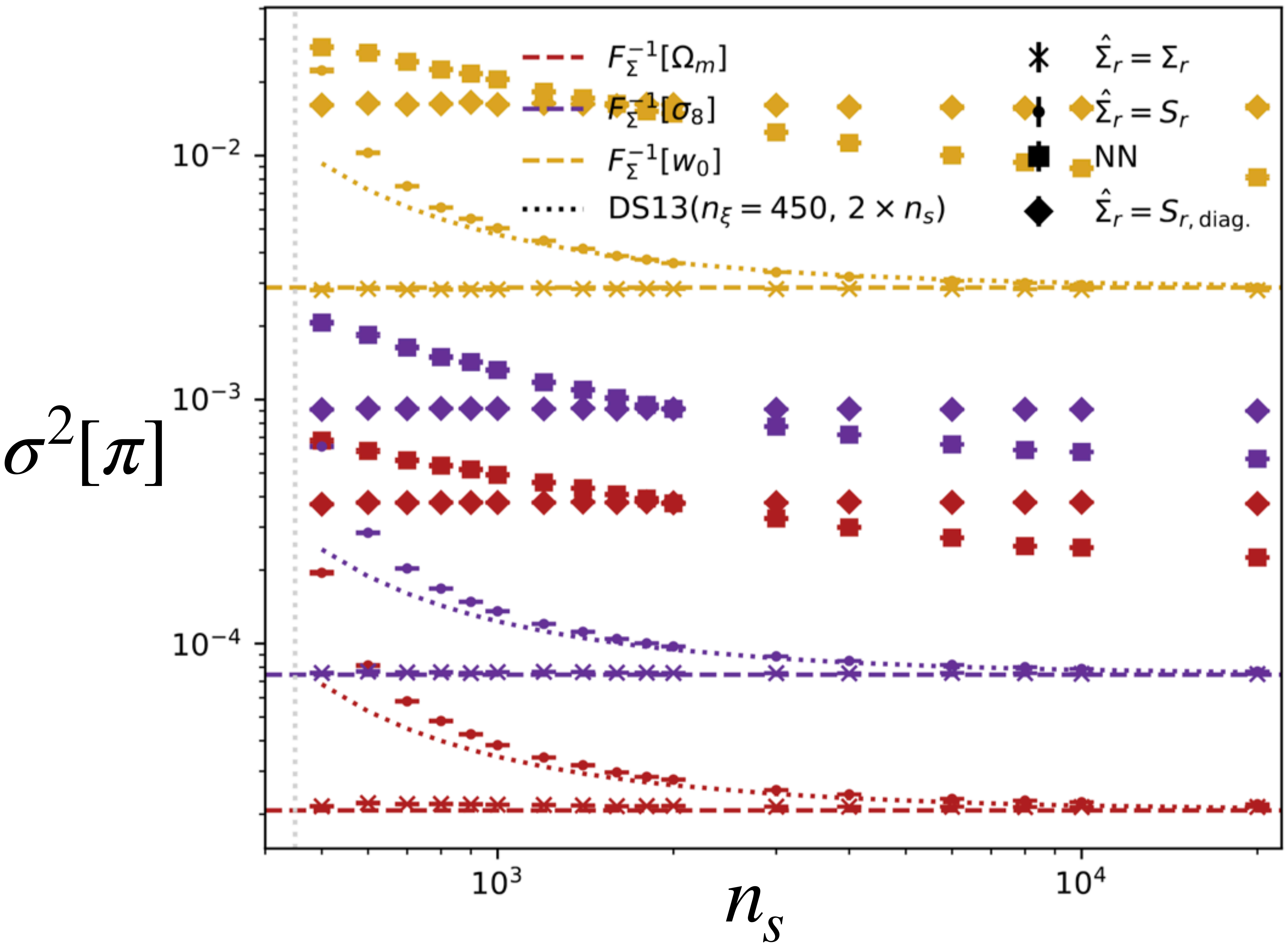
→ **Test a change to the covariance structure** $\Sigma \rightarrow \Sigma_r$

$$(\Sigma_r)_{ij} = \begin{cases} \Sigma_{ij}, & \text{if } i = j \\ r\sqrt{\Sigma_{ii}\Sigma_{jj}}, & \text{if off-diagonal} \\ 0, & \text{else} \end{cases}$$

$\Sigma_r \,, (r = 0.2)$

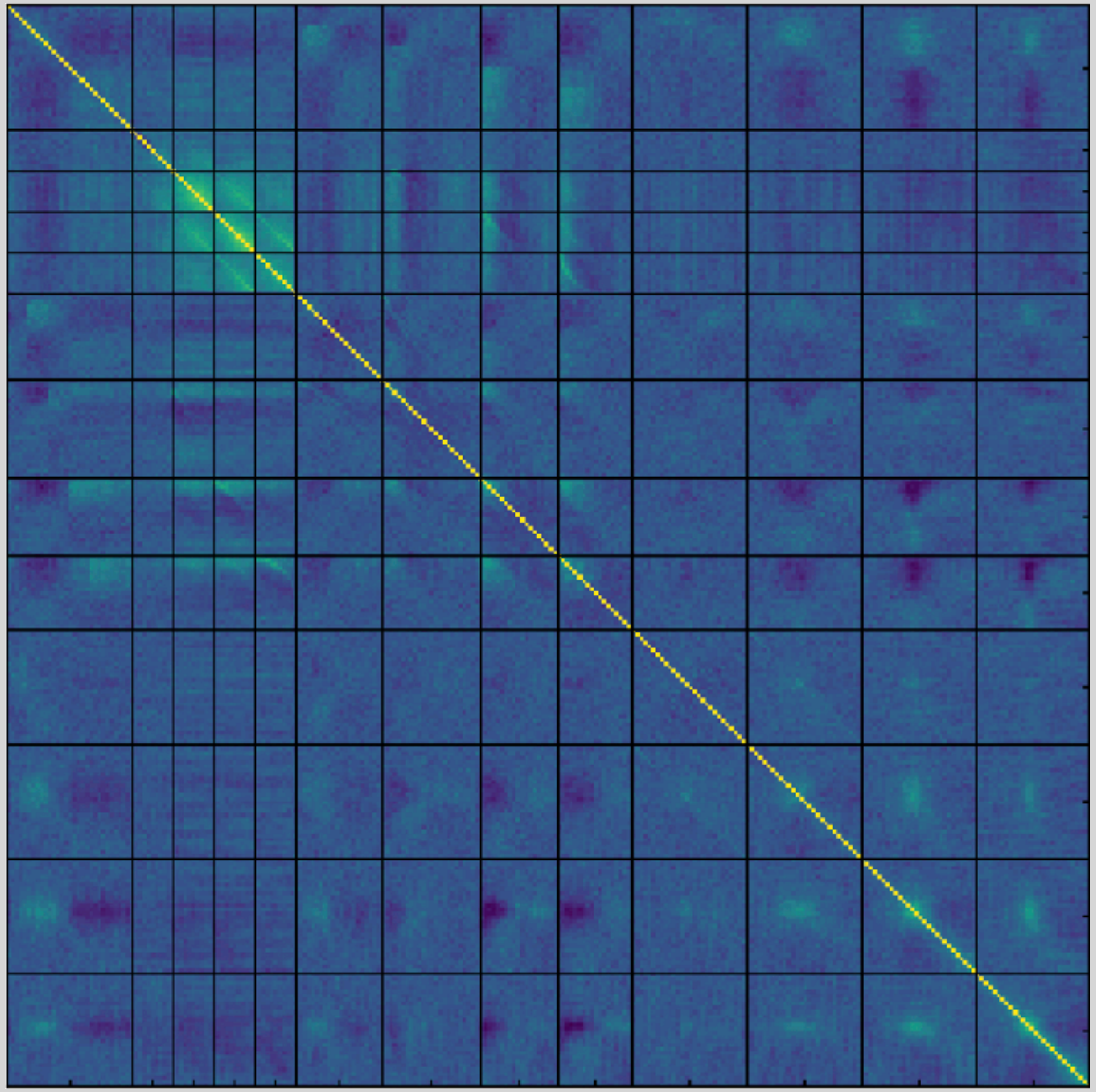


Where does NN compression fail?

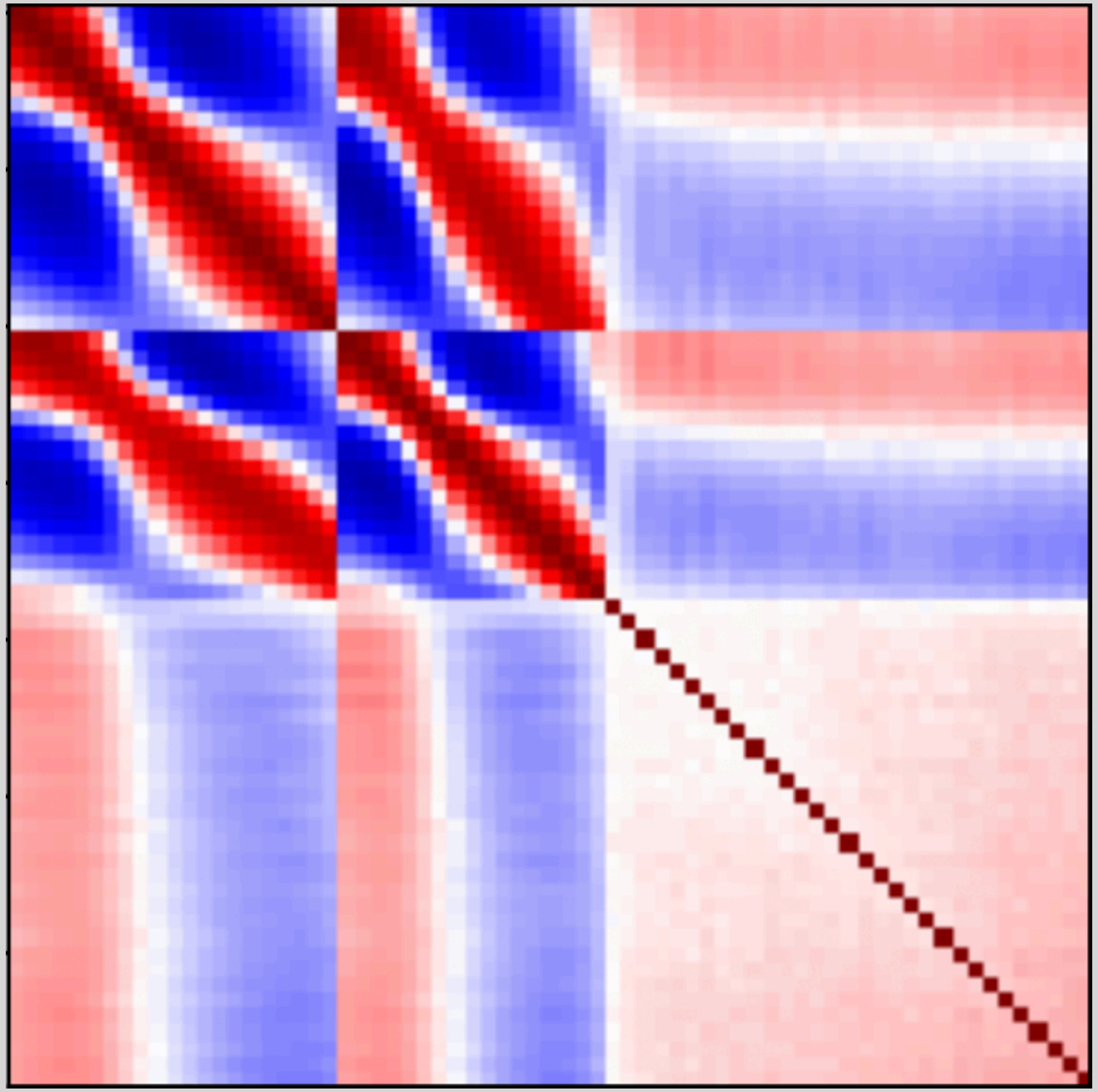


- **NN fails to summarise when data covariance has large off-diagonal elements!**

Unfavourable Σ's are easy to find in cosmology!

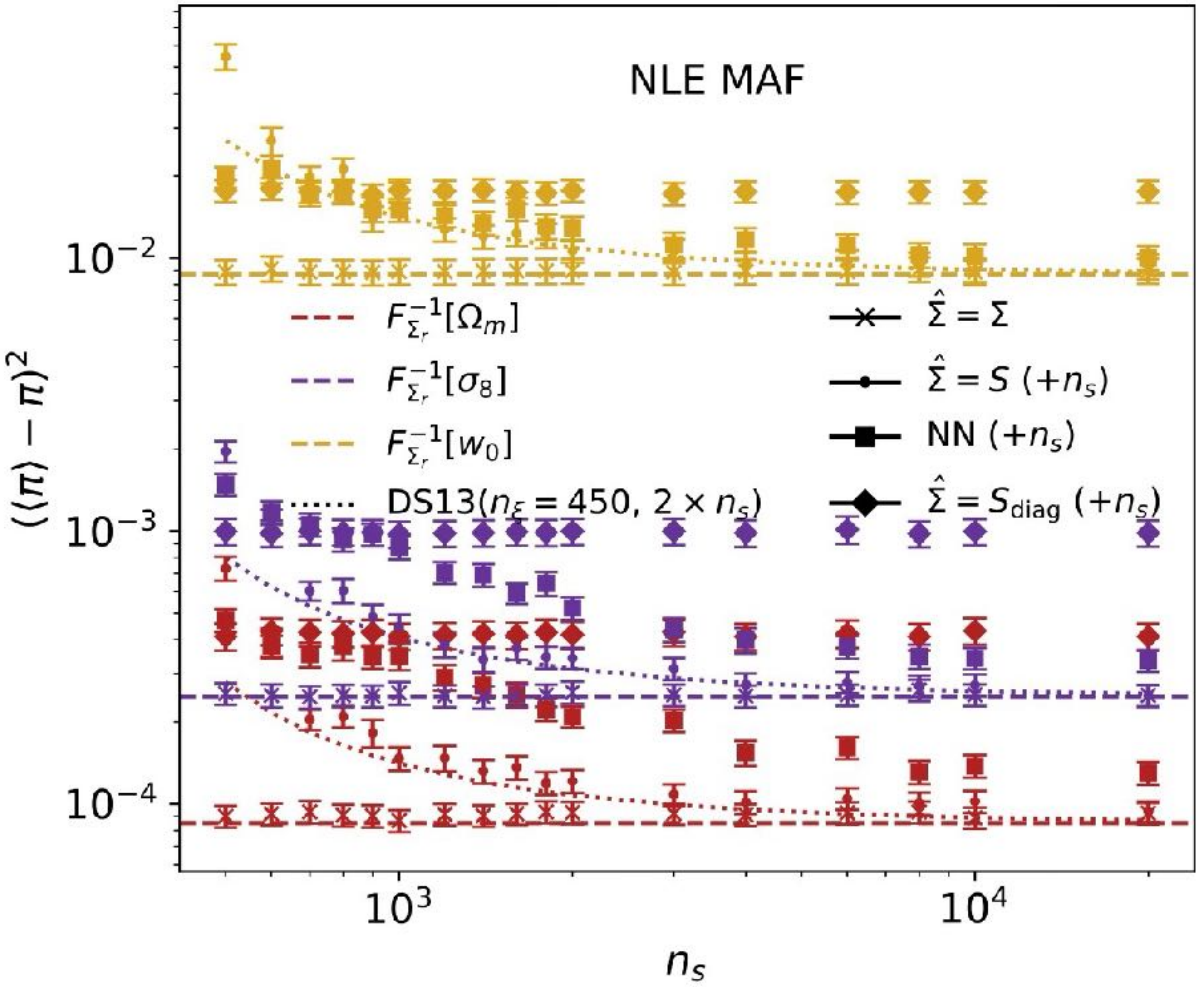


Σ for weak lensing peaks (Davies++2021).



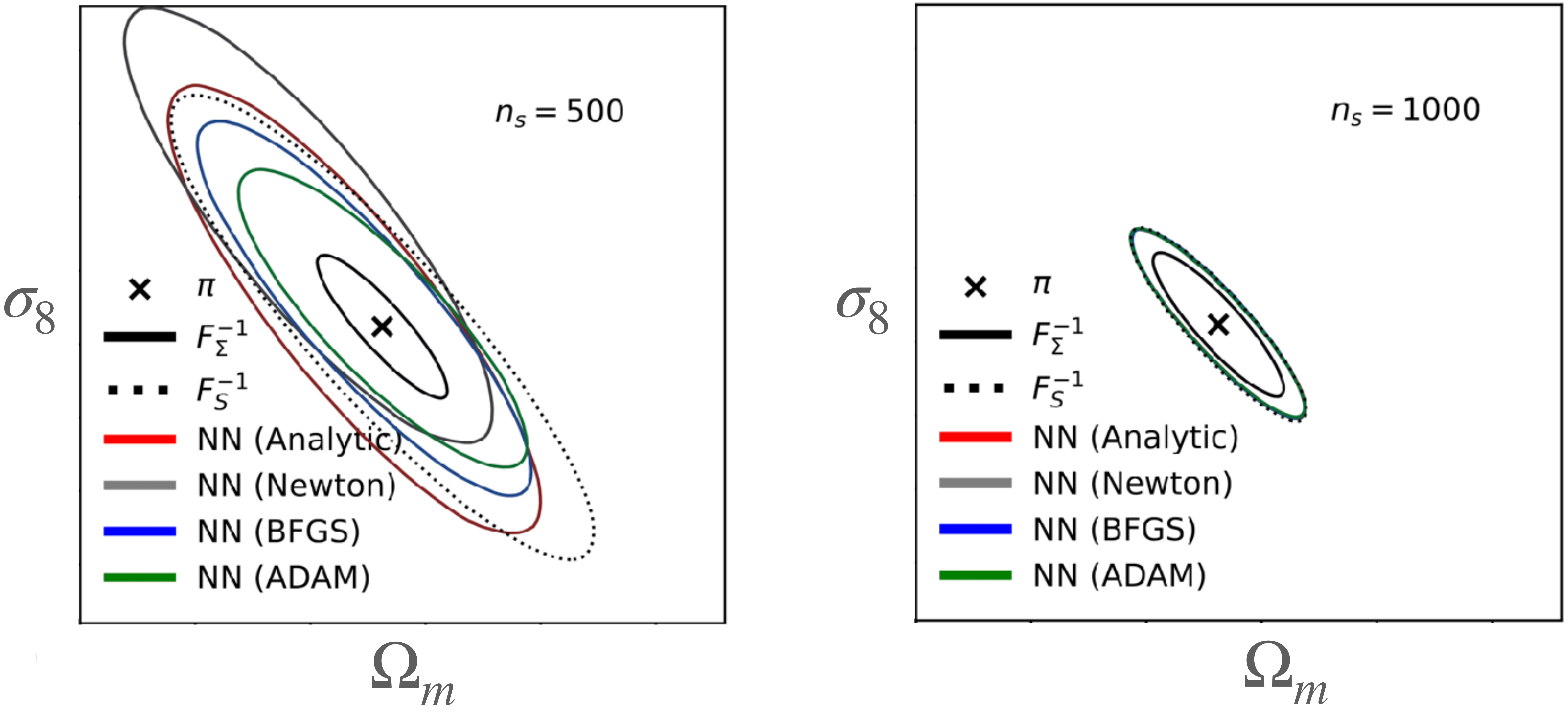
Σ for matter PDF + $P(k)$, 3 scales, redshift zero (Uhlemann++2019).

SBI creates non-Gaussian posteriors, so what about $\langle\pi\rangle_{\pi|\hat{\xi}} \neq \pi$?



- Is there a Dodelson-like effect in the density estimation?
- In addition to the scatter from compression with $\hat{\Sigma} = S$?

Future: much ado about neural networks



- **<u>How does</u>** $\hat{\pi}[\hat{\xi}]$ **<u>from a neural network scatter on average</u>** for low n_s?
    - calculation of summary scatter for non-linear model,
    - optimisation has a regularising effect.

Future: interpretable likelihoods from machines

- **Current density estimation methods are not interpretable**
  - What is the difference between $p_\phi(\xi\,|\,\pi)$ and a Gaussian linear model?
  - Can we fit a model for $\xi[\pi]$ and $p(\hat{\xi}\,|\,\xi[\pi])$?
- **Solution may *not* lay in the machine learning literature... yet**
  - 10 years of flows,
  - diffusion, FM, ... poorer density estimation.