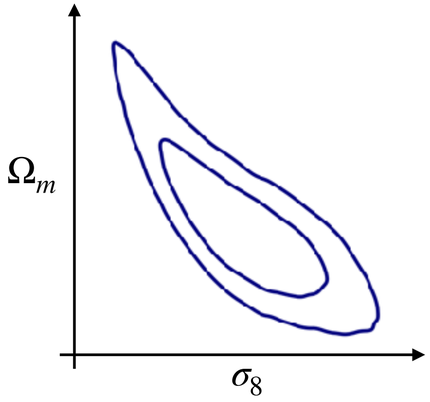
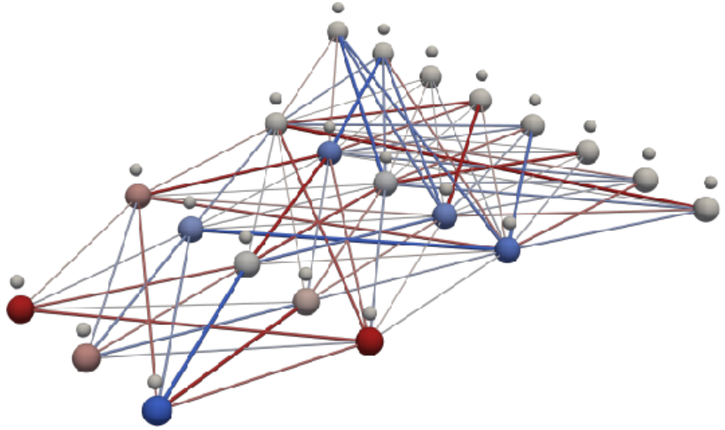
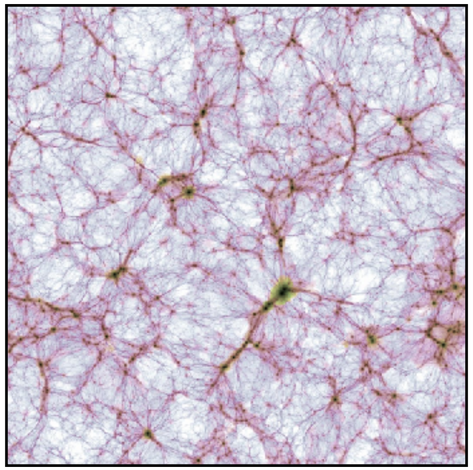
SBI has its own Dodelson-Schneider effect (but it knows!)







Jed Homer

Workshop on "Cosmological Inference" | MPA | 07/07/26

2412.02311

LMU PHYSIK

ORIGINS Excellence Cluster

mcml Munich Center for Machine Learning

Why simulation-based inference (SBI)?

$$p(\theta \,|\, \hat{x}) \propto p(\hat{x} \,|\, x[\theta], \theta)\, p(\theta)$$

- Likelihood $p(\hat{x} \,|\, \theta)$ may be non-Gaussian
- Expectation $x[\theta]$ is a complex non-linear function of θ
- Covariance Σ may depend on θ
- n_x is increasing with new surveys, inversion of $\hat{\Sigma}$

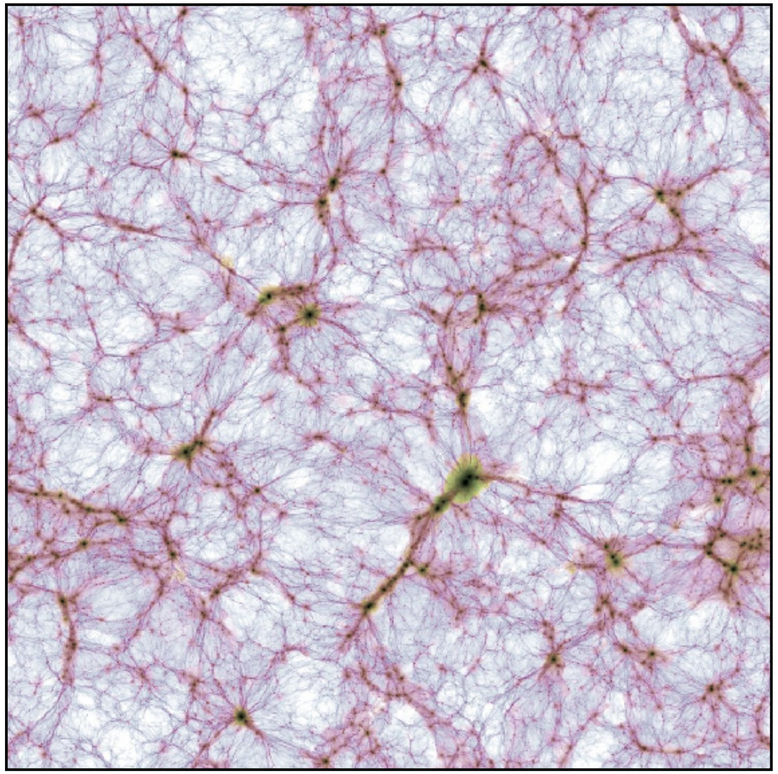
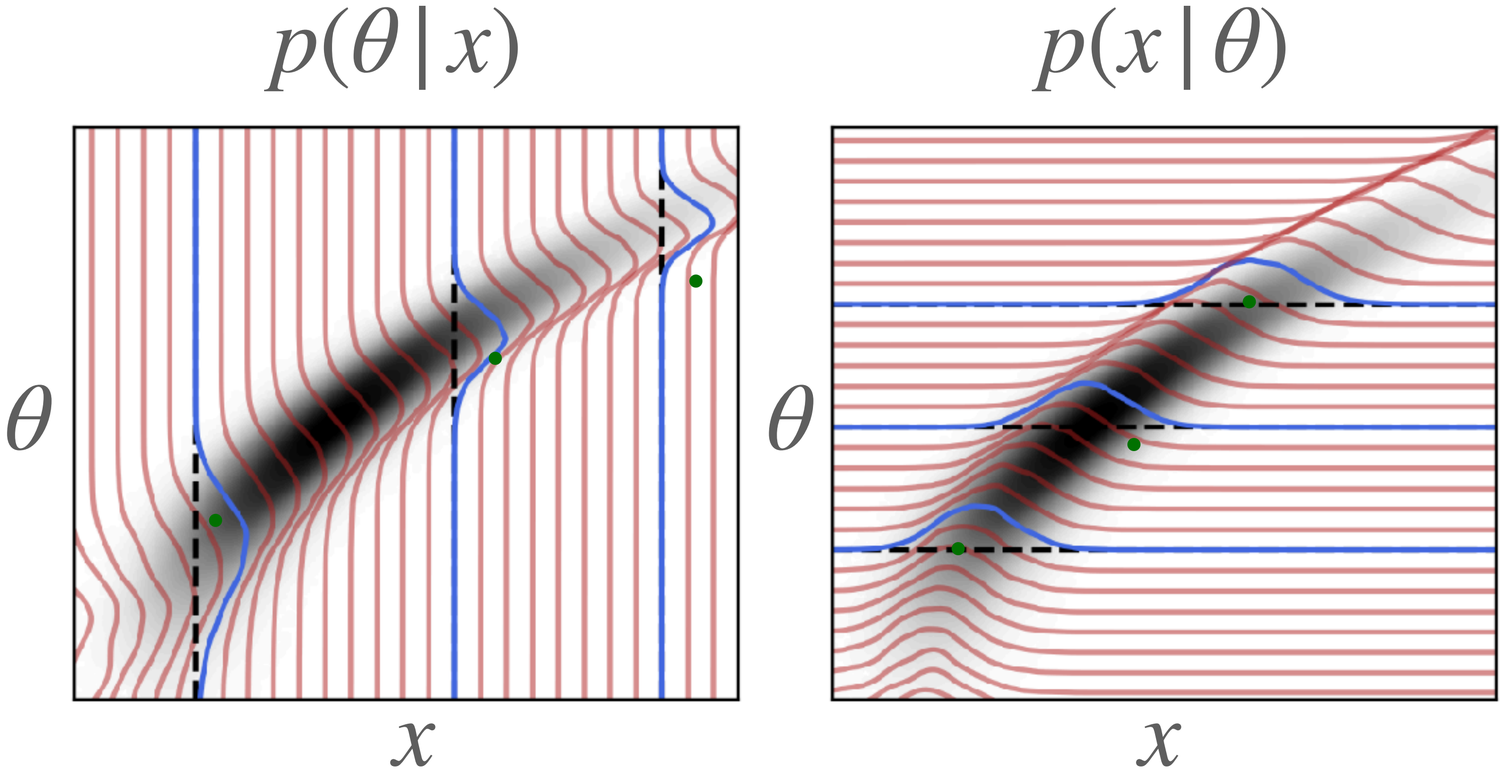


Image: IllustrisTNG

Simulation-based inference (SBI)

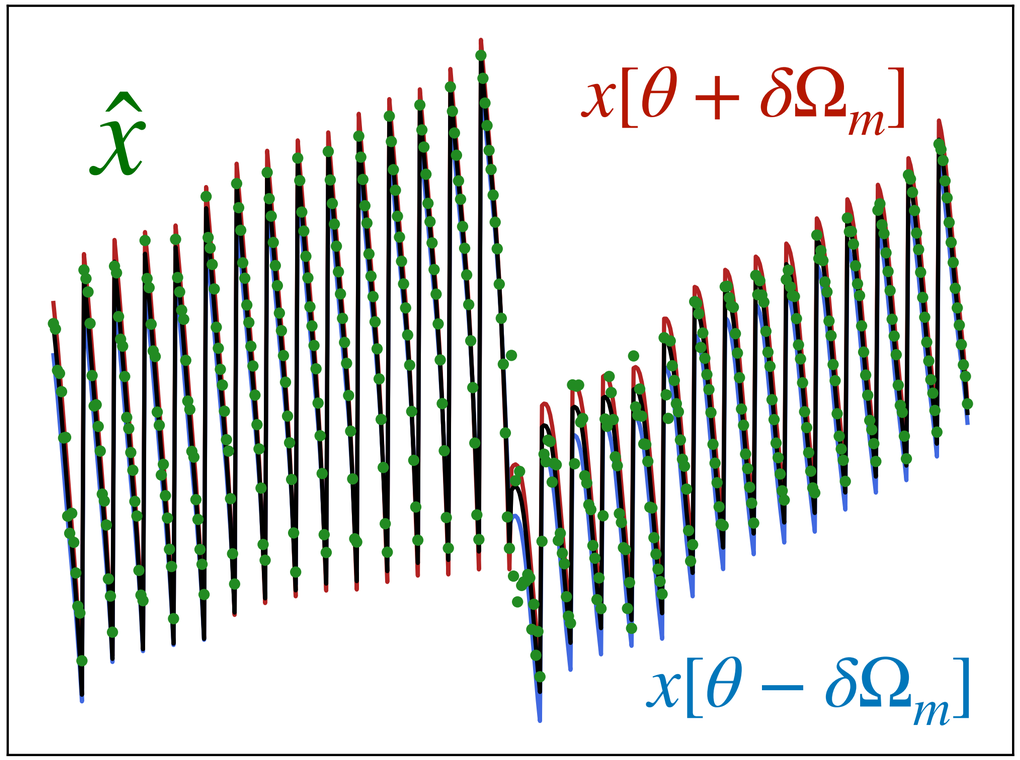


- **Analytic methods**: model PDF across space (x, θ)

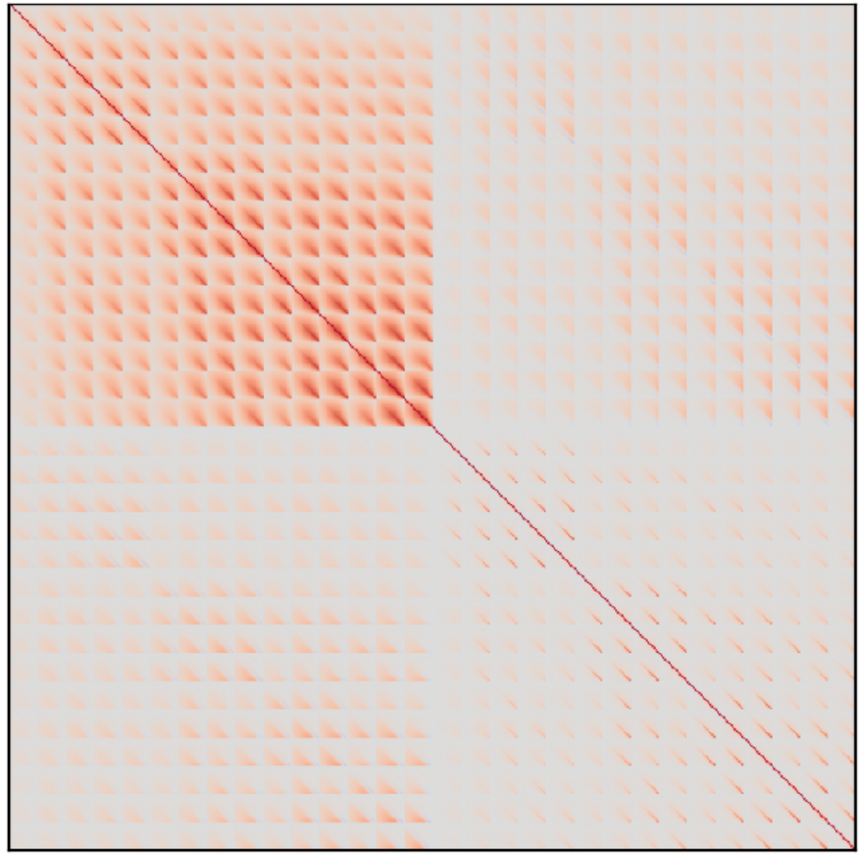
- **SBI**: fit the expectation $x[\theta]$, covariance Σ and PDF $p(\hat{x} \mid \theta)$ from simulations (x, θ)

$$\hat{x} \sim \mathcal{G}\left[\hat{x} \,|\, x[\theta], \Sigma\right]$$

$x[\theta]$

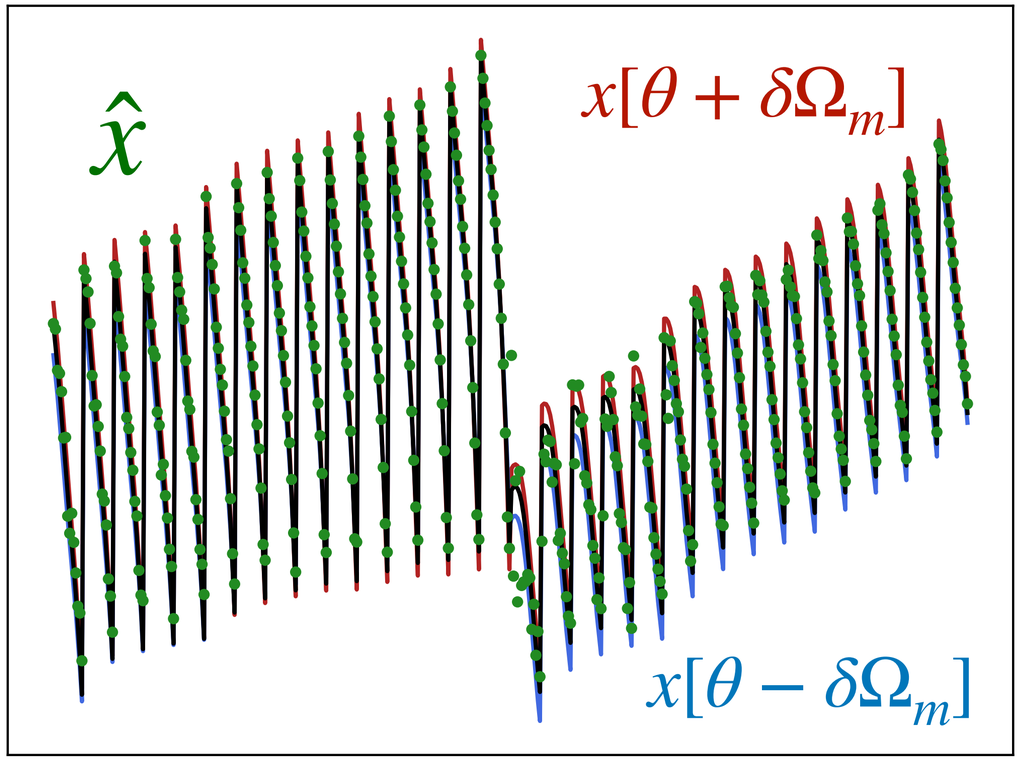


Σ

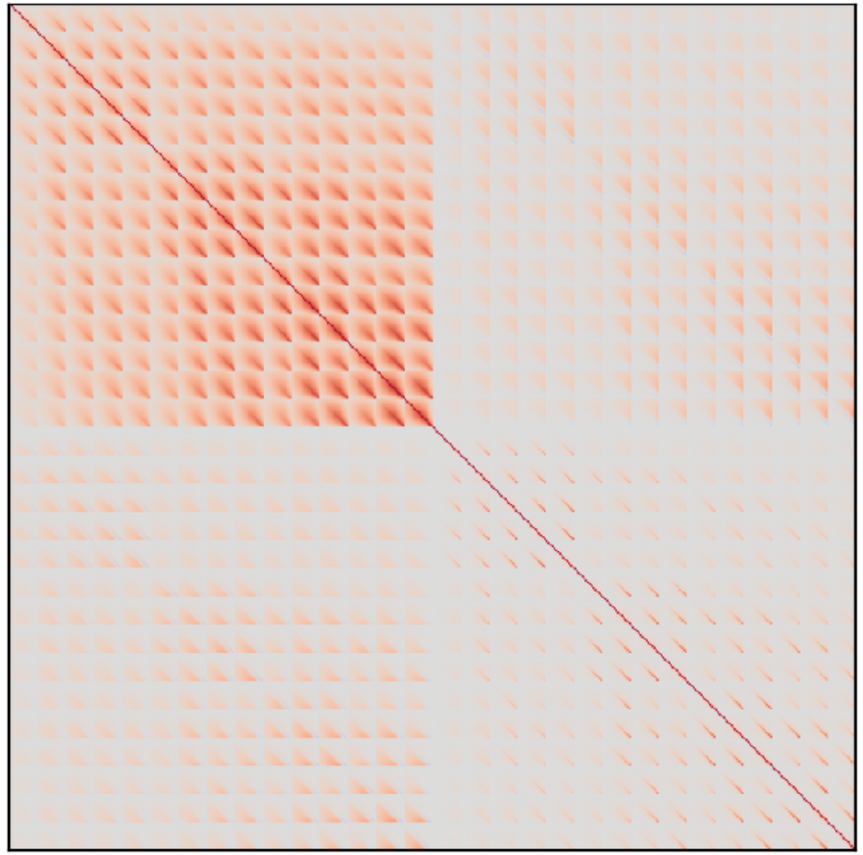


$$\hat{x} \sim \mathcal{G}\left[\hat{x} \,|\, x[\theta], \Sigma\right]$$

$x[\theta]$



Σ



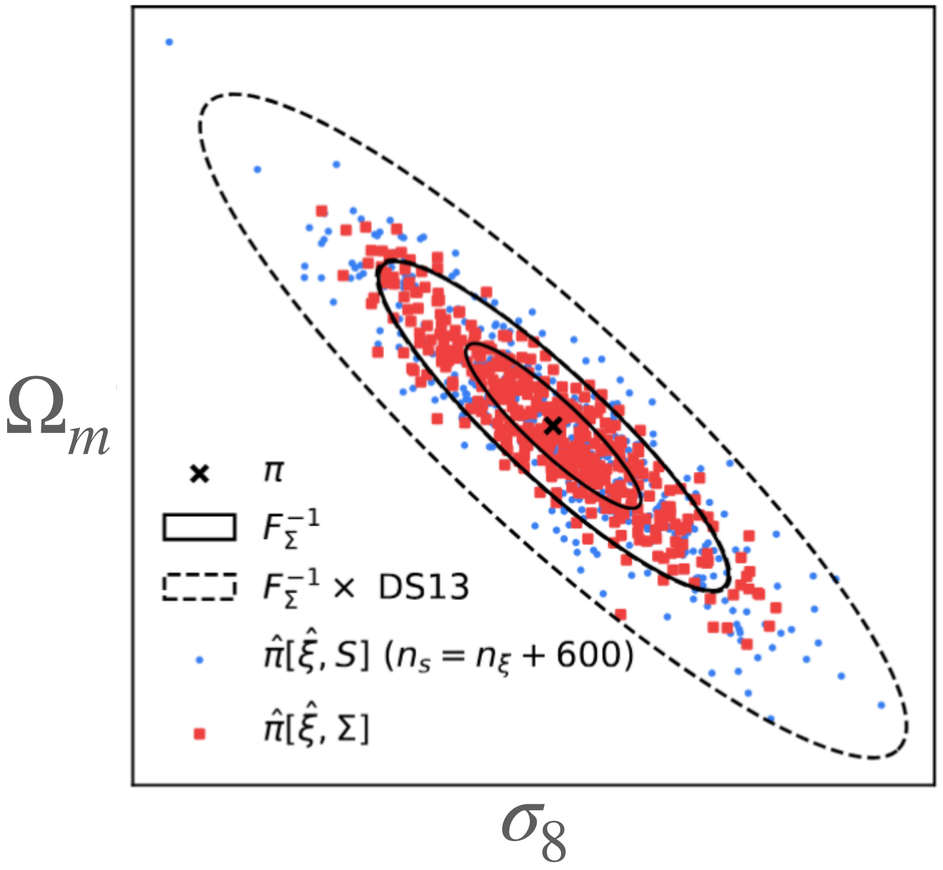
Best-fit $\hat{\theta}$ with estimated covariance $\hat{\Sigma} \neq \Sigma$

- Estimating **best-fit** $\hat{\theta}$ from $\hat{x}$ with an **estimated covariance** $\hat{\Sigma}$

$$\hat{x} \sim \mathcal{G}\left[\hat{x} \,|\, x[\theta], \Sigma\right]$$

$$\hat{\theta} = \theta + M\left[\hat{\Sigma}^{-1}\right]\left(\hat{x} - x[\theta]\right)$$

$\hat{\Sigma}^{-1}$ Σ^{-1}

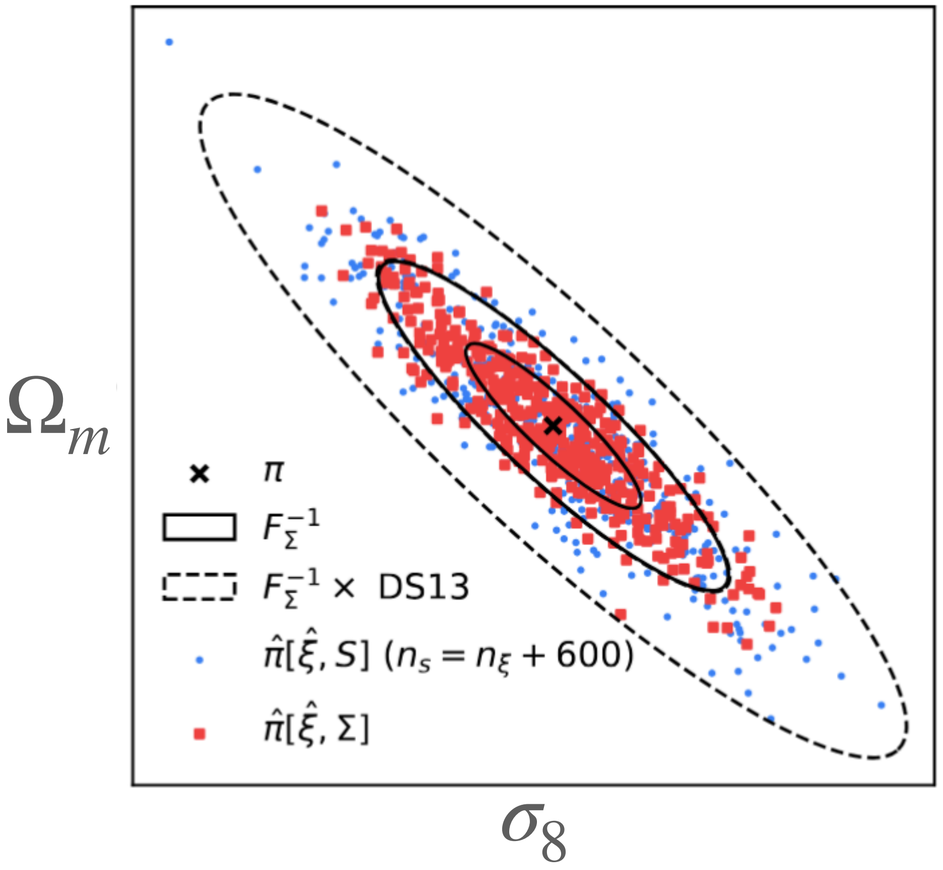


The Dodelson-Schneider Effect

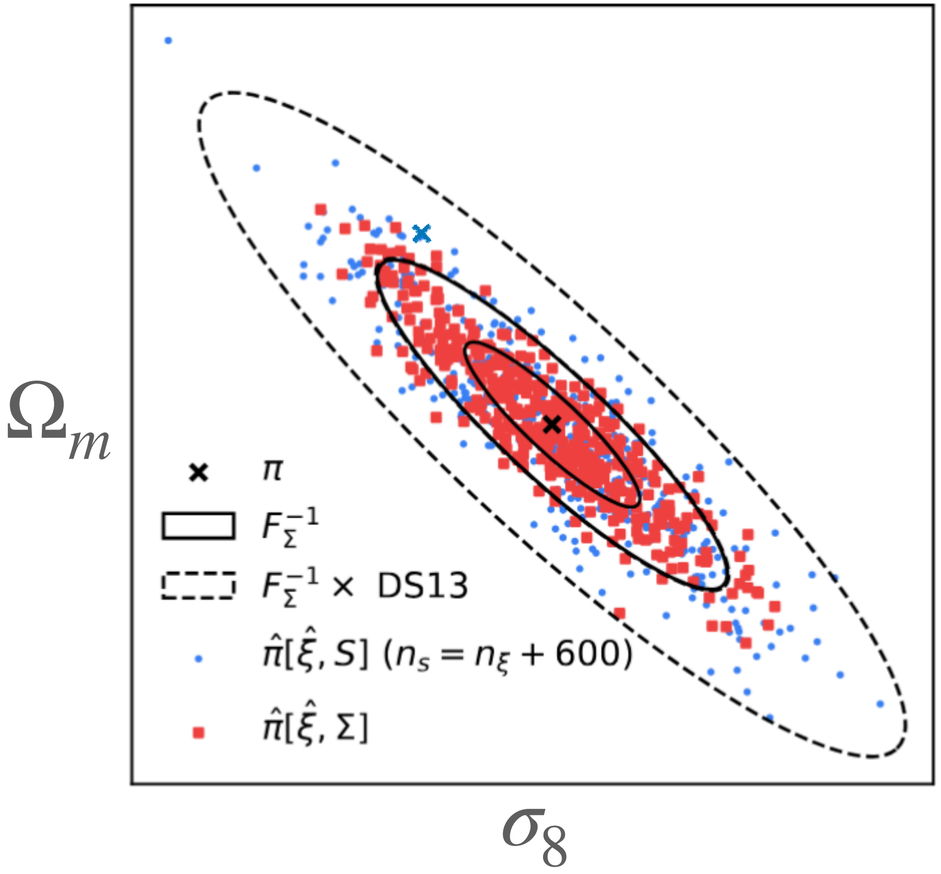
- Estimate the covariance $\hat{\Sigma}$ with n_s simulations

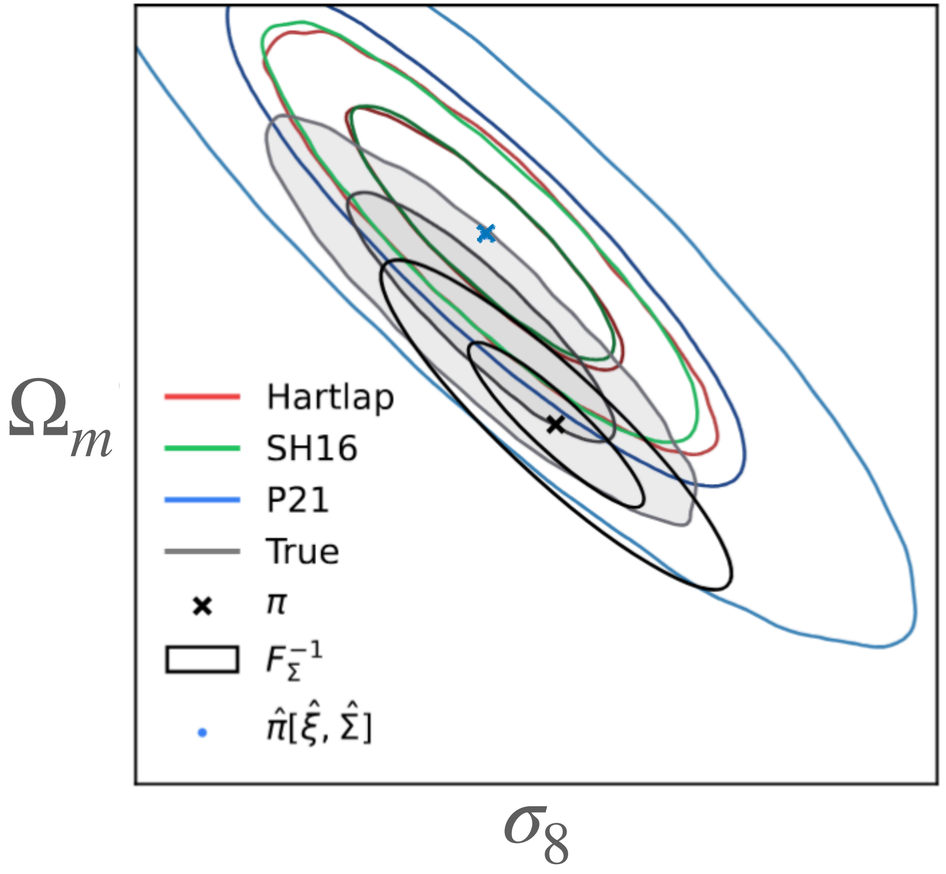
$$\hat{\theta} = \theta + M\left[\hat{\Sigma}^{-1}\right]\left(\hat{x} - x[\theta]\right)$$

$$\hat{\Sigma}_\theta = \Sigma_\theta\left[1 + \frac{n_x - n_\theta}{n_s - n_x}\right]$$



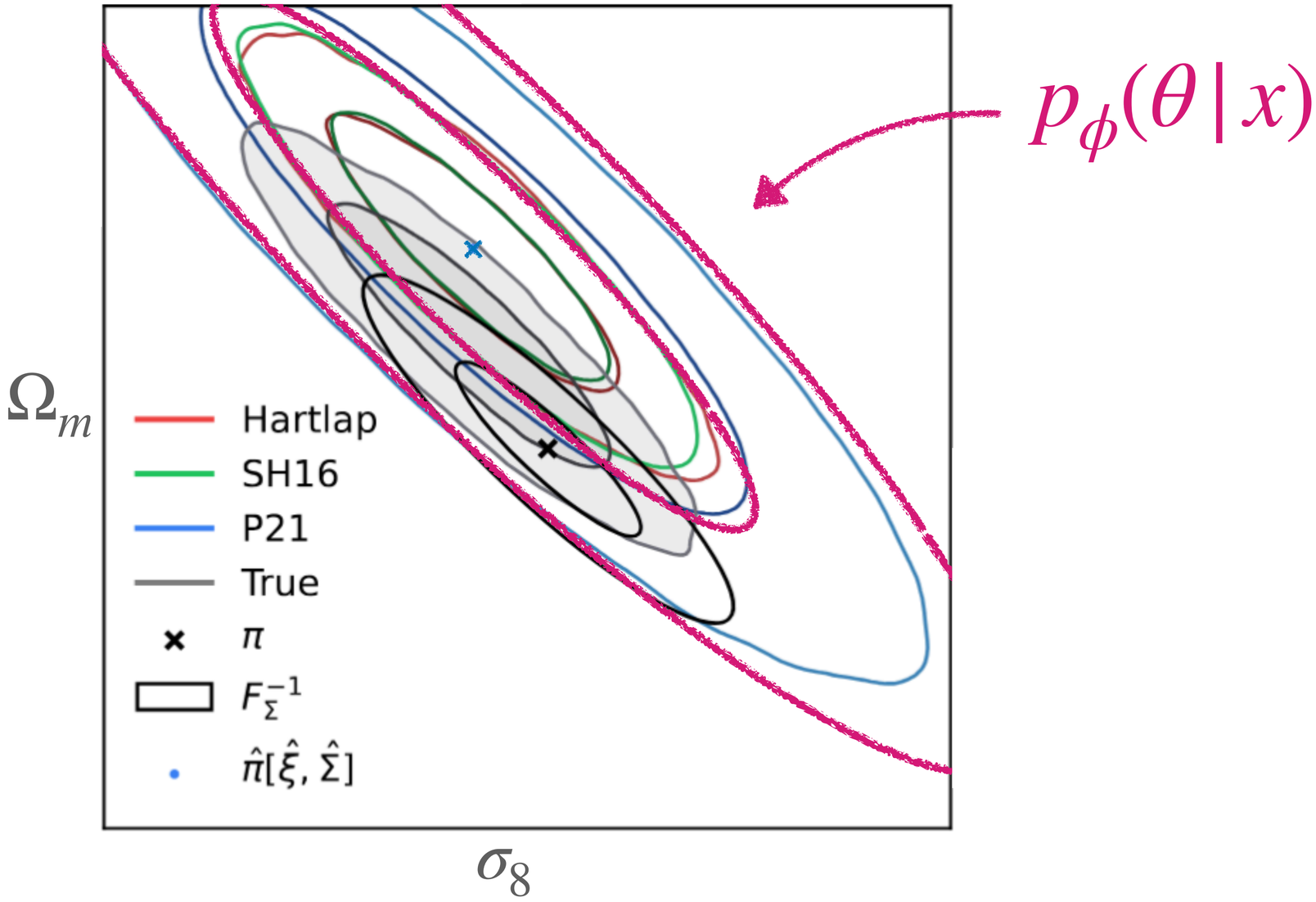
Problem: estimating θ from $\hat{x}$ when $\hat{\Sigma} \neq \Sigma$



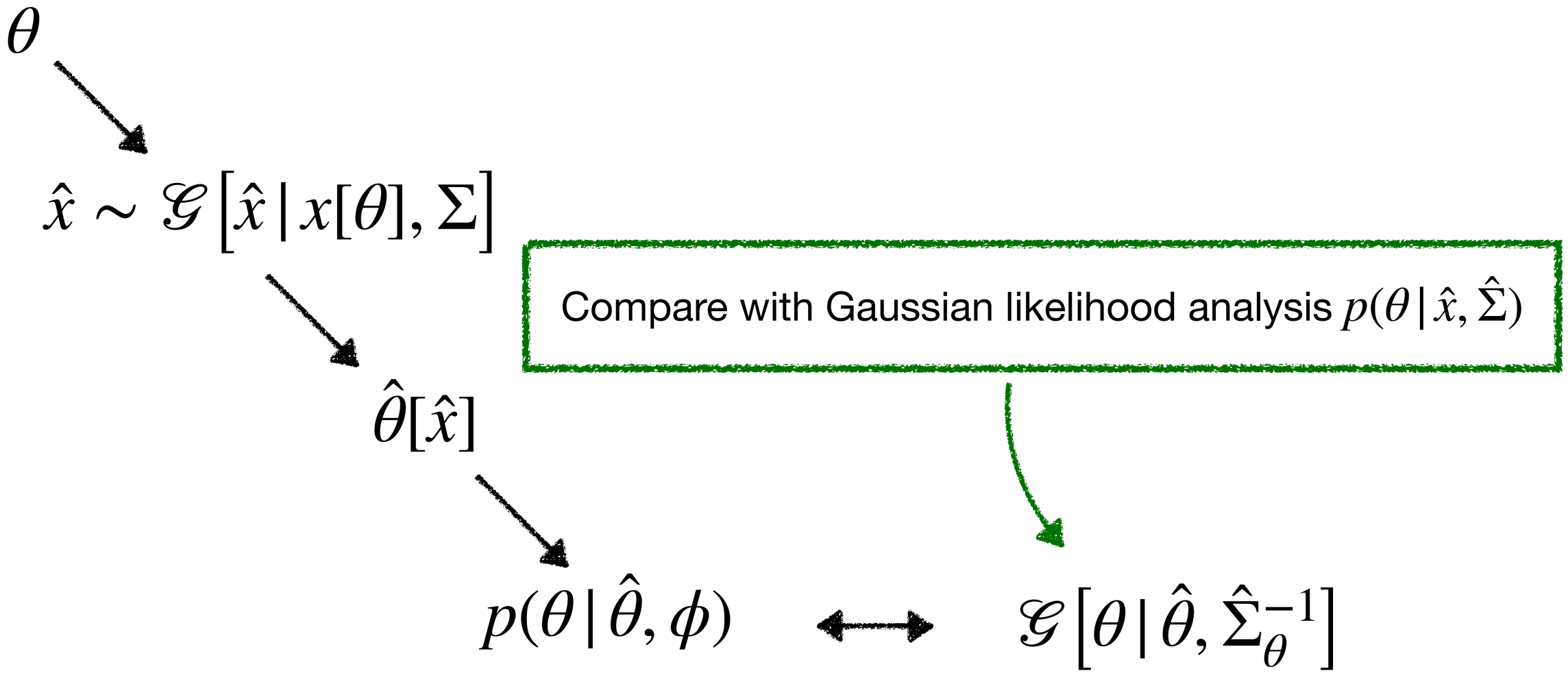


Noise in $\hat{\Sigma} \neq \Sigma$ adds scatter to best-fit $\hat{\pi}$. **Not accounted for in Gaussian ansatz.**

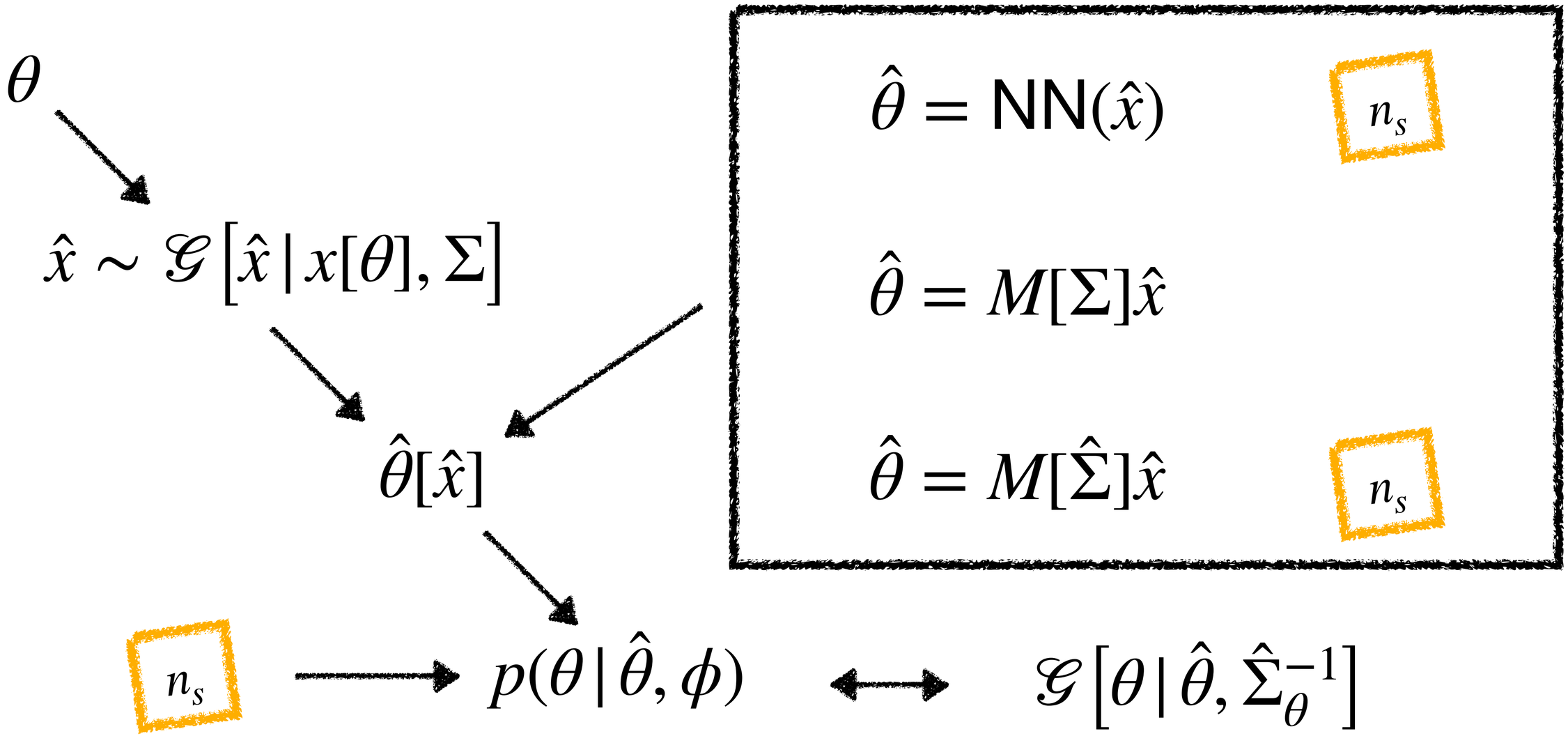
What contour does SBI draw?



Analytic testing of SBI



Analytic testing of SBI



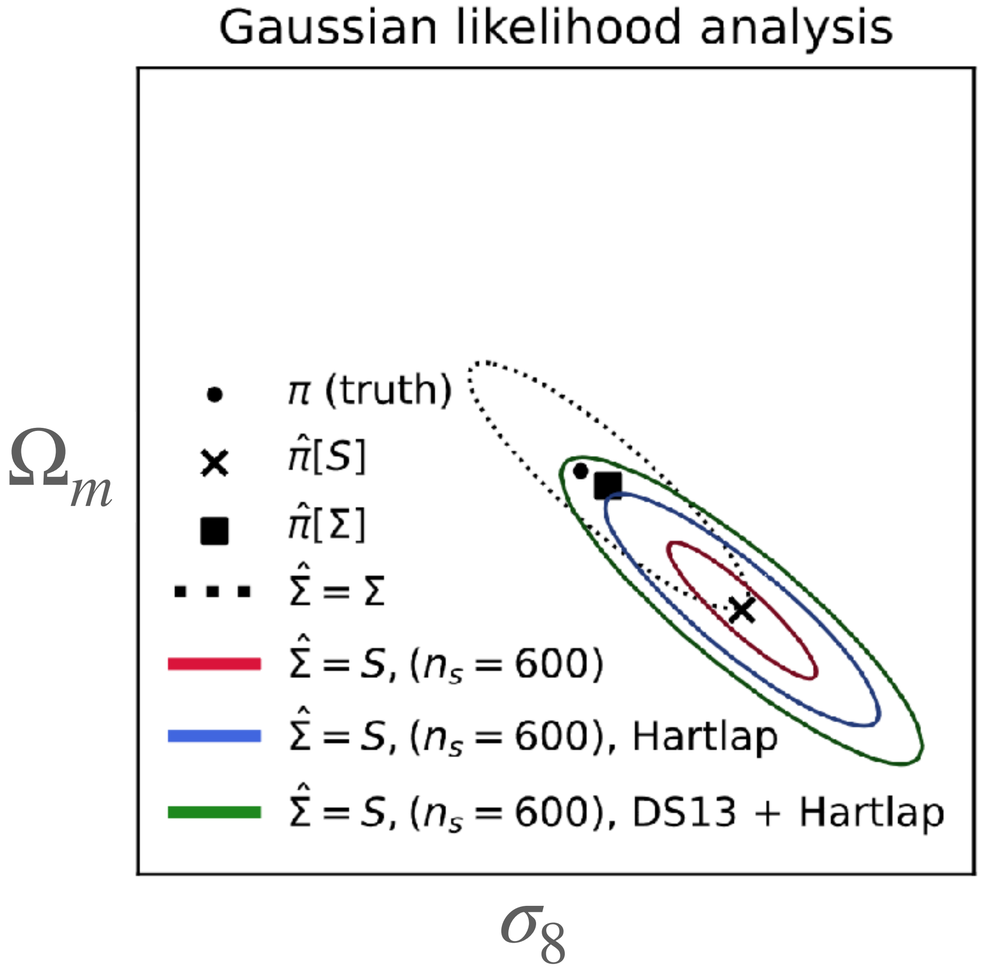
Analytic testing of SBI

$$\theta \rightarrow \hat{x} \sim \mathcal{G}\left[\hat{x} \,|\, x[\theta], \Sigma\right] \rightarrow \hat{\theta}[\hat{x}] \rightarrow p(\theta \,|\, \hat{\theta}, \phi) \longleftrightarrow \mathcal{G}\left[\theta \,|\, \hat{\theta}, \hat{\Sigma}_\theta^{-1}\right]$$

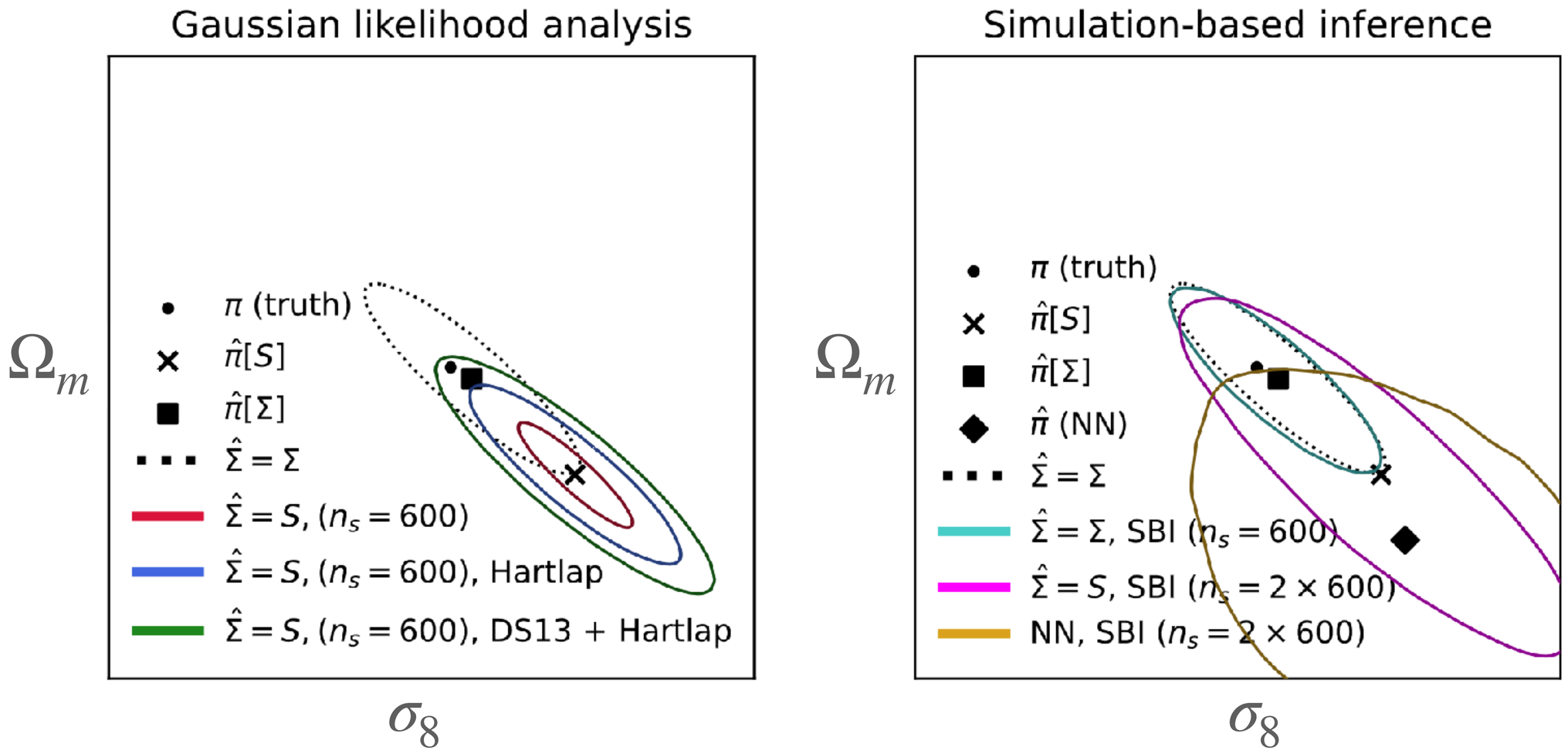
- **Here**: Σ or $\hat{\Sigma}$ parameterises the optimal compression
- **Compression does not remove the noise in $\hat{\Sigma}$!**

$$p(x \,|\, x[\theta], \hat{\Sigma}) \;\leftrightarrow\; p(\hat{\theta} \,|\, \theta, \hat{\Sigma}_{\hat{\theta}}) \;\leftrightarrow\; p(\hat{\theta} \,|\, \mu_{\text{NN}}, \hat{\Sigma}_{\text{NN}})$$

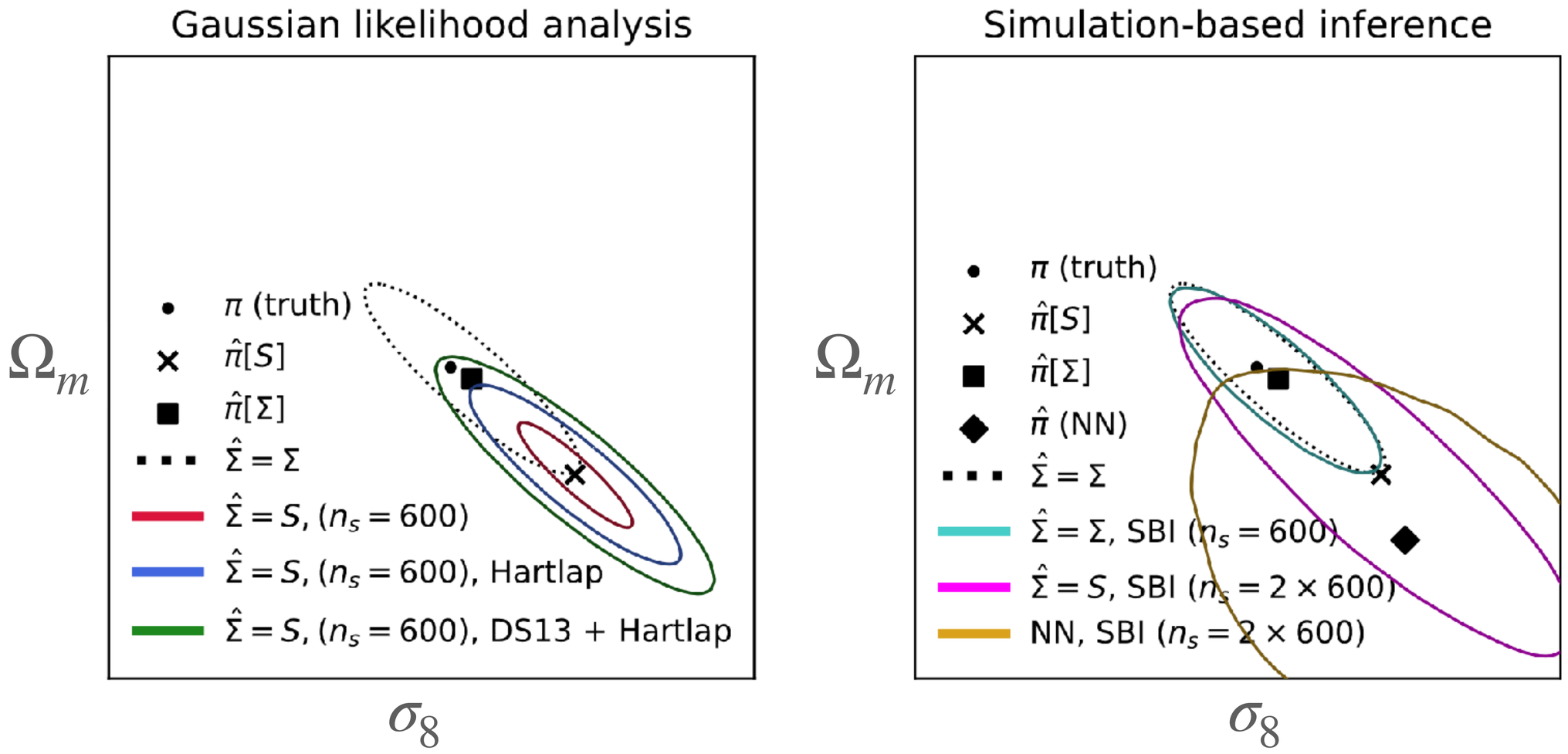
The good and bad news about SBI



The good and bad news about SBI

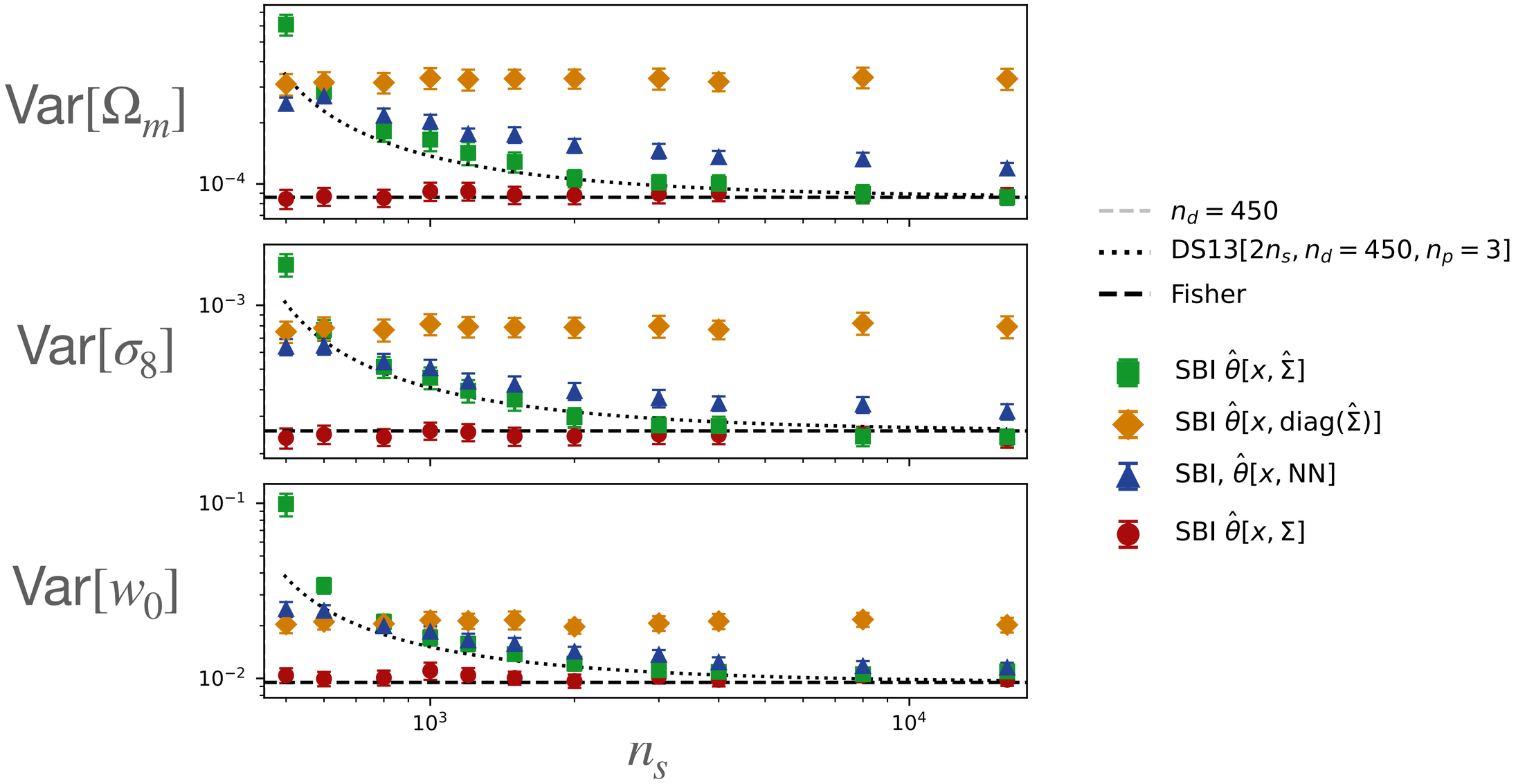


The good and bad news about SBI

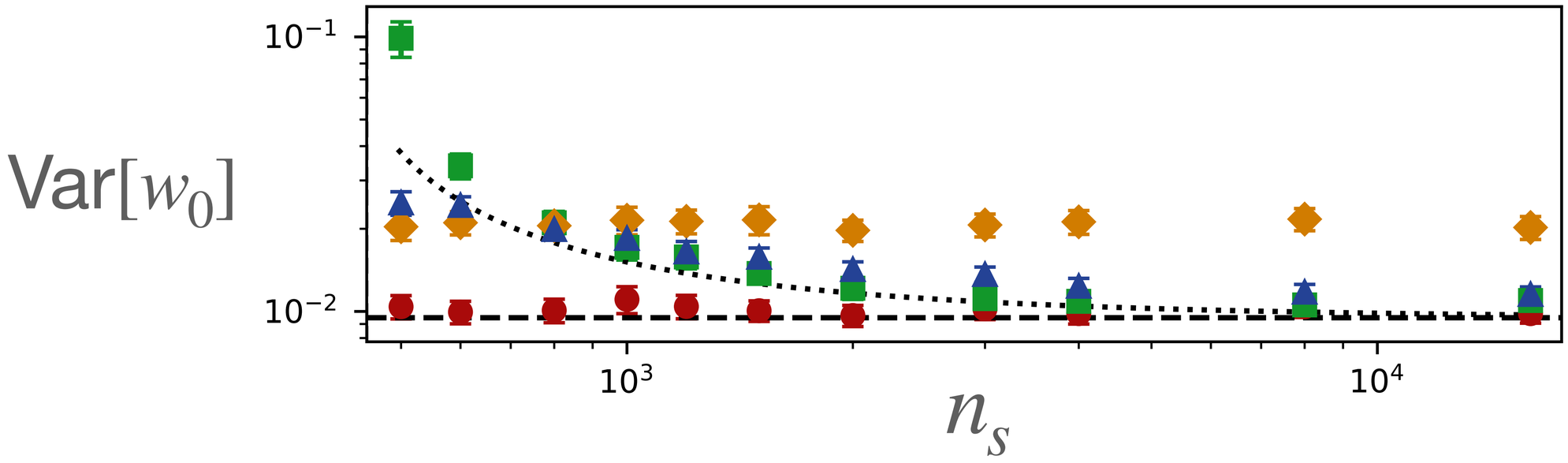


SBI is ***aware*** of the Dodelson-Schneider effect but it is ***inefficient*** in its response!

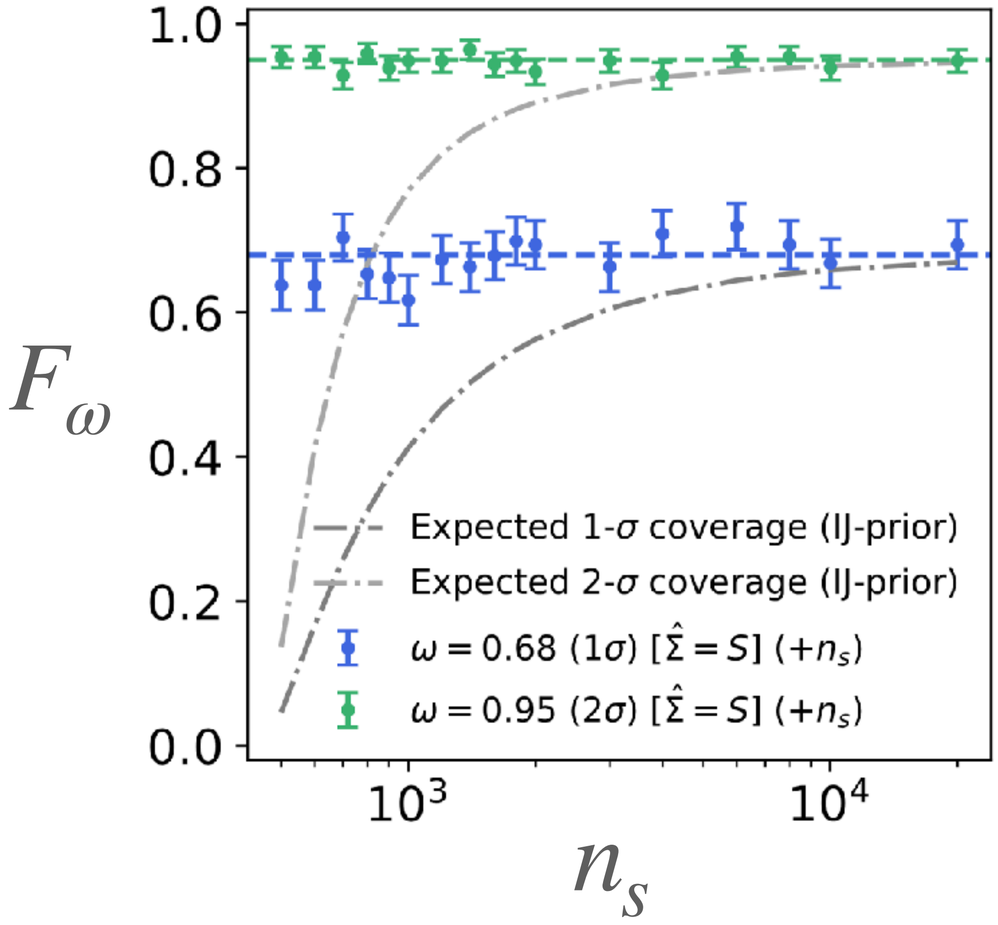
Posterior widths: SBI vs. Gaussian likelihood analysis



Posterior widths: SBI vs. Gaussian likelihood analysis



Posterior coverages: SBI



Good news

- Coverages are nominal, widths are inflated $\Longrightarrow$ **SBI *knows* about the Dodelson-Schneider effect and corrects for it!**

- **Low n_s with high n_x $\Longrightarrow$ NN + NDE returns tighter constraints** (than $\mathscr{G}$ + $\hat{\Sigma}$), where Σ is insurmountable (analytically or through simulations)

- **No assumptions on $p(x \mid \theta)$** and **optimal compression** $\Longrightarrow$ SBI fits true $p(x \mid \theta)$ with low n_s

Bad news

- **SBI returns diluted parameter constraints w.r.t. a Gaussian likelihood analysis for the same** n_s for Stage-III surveys

- **Compression is the crux**: we don't know how to compress data optimally, and multi-probe analyses increase n_x
  - **Moving noise around the analysis**: choose either $p(x \,|\, x[\theta], \hat{\Sigma}),\ p(\hat{\theta} \,|\, \theta, \hat{\Sigma}_{\hat{\theta}}),\ p(\hat{\theta} \,|\, \mu_{\text{NN}}, \hat{\Sigma}_{\text{NN}})$
    - **the neural density estimator likelihood fit cannot extract information that is lost here!**

- **Nuisances** and **complexity in** $x[\theta]$, **parameter dependence of** Σ and **non-Gaussian** $p(x \,|\, \theta)$ will dilute information further, it is unclear how SBI responds to this...

- **Hyperparameter optimisation:** when n_x is large, training set is not as large as we'd like, meaning we overfit more to this during hyperparameter optimisation!

Two questions about SBI

- Does SBI produce "correct" posteriors in a set of repeated experiments, with no assumptions on the likelihood PDF?
  - How does this vary with n_s ?

Today

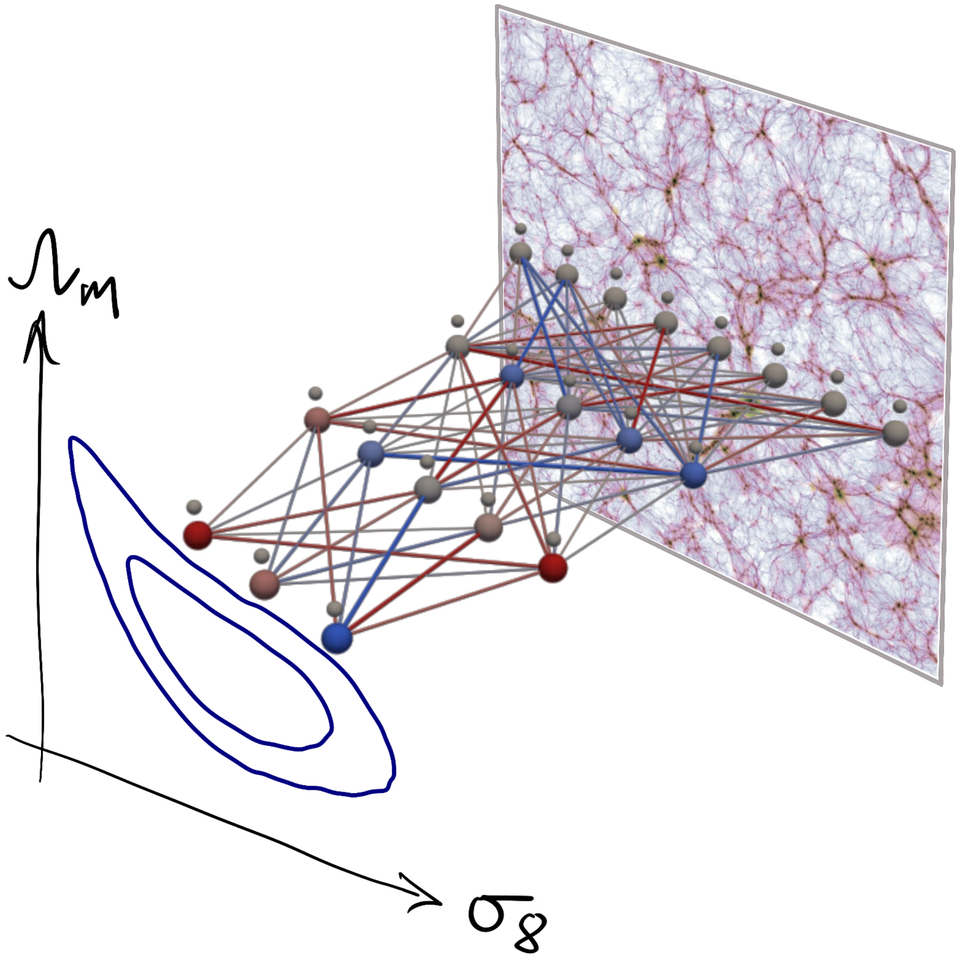
- Does SBI obtain more information than a Gaussian likelihood analysis for non-Gaussian statistics (e.g. the matter PDF) with a non-linear expectation $x[\theta]$?

Homer++26 (in prep.)

homerjed/sbiax

homerjed.github.io

jed.homer@physik.lmu.de



2412.02311

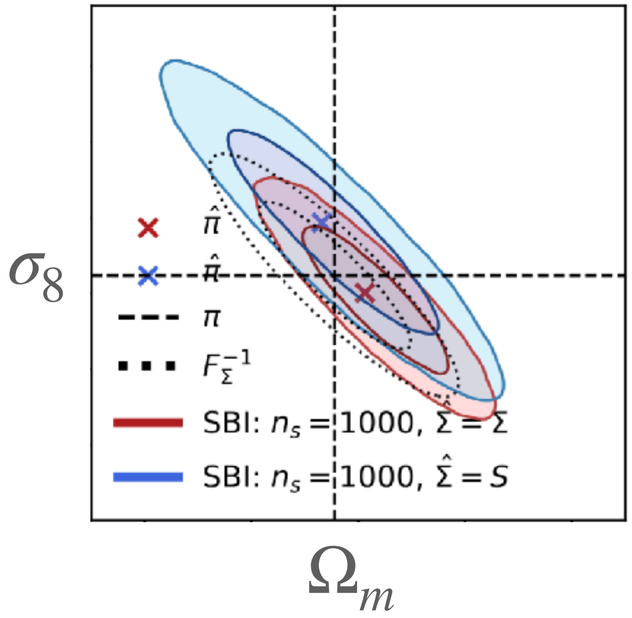
LMU PHYSIK

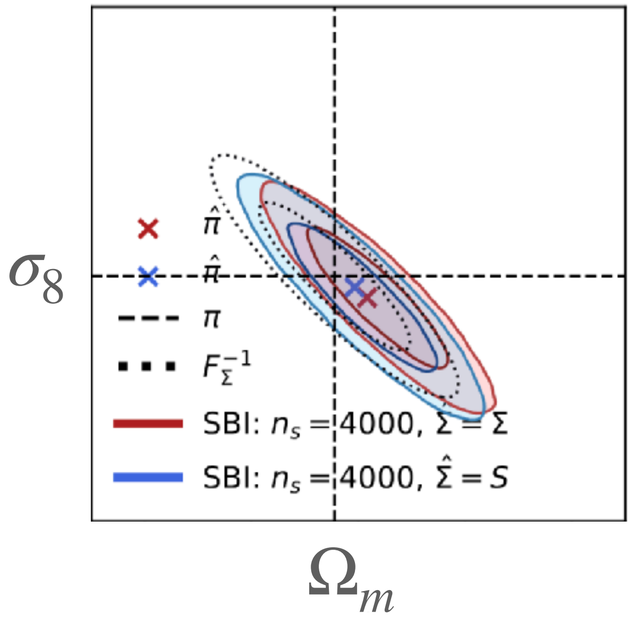
ORIGINS Excellence Cluster

mcml Munich Center for Machine Learning

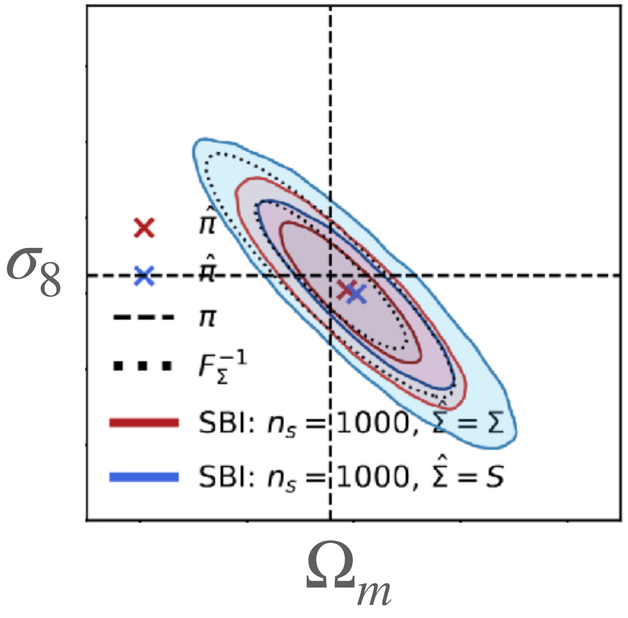
Reconstructed posteriors

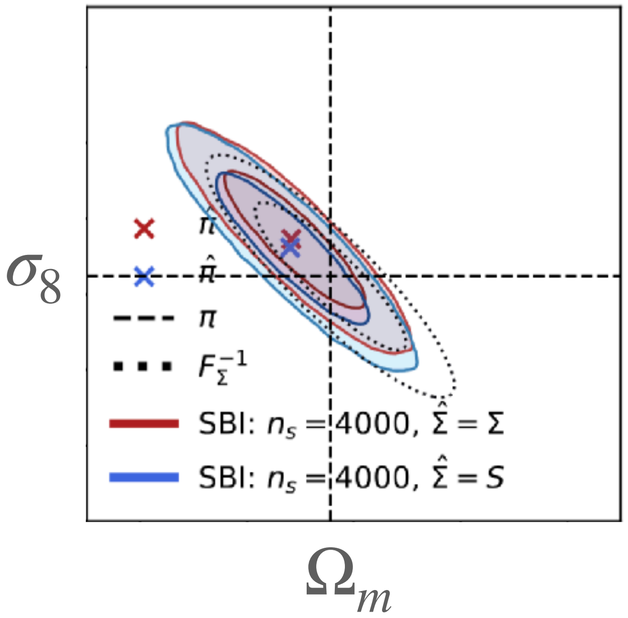
MAFs





CNFs





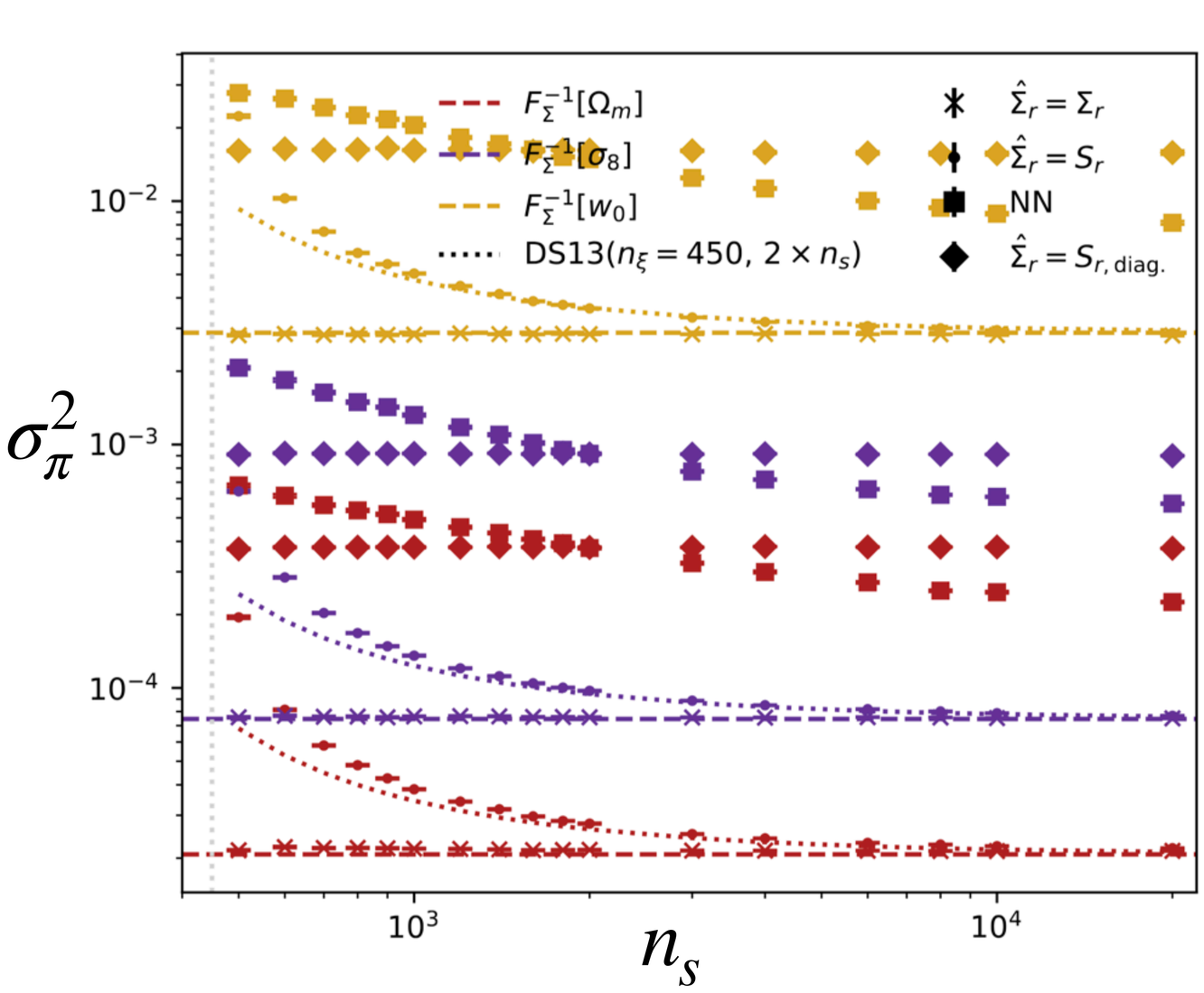
$n_s = 1000$

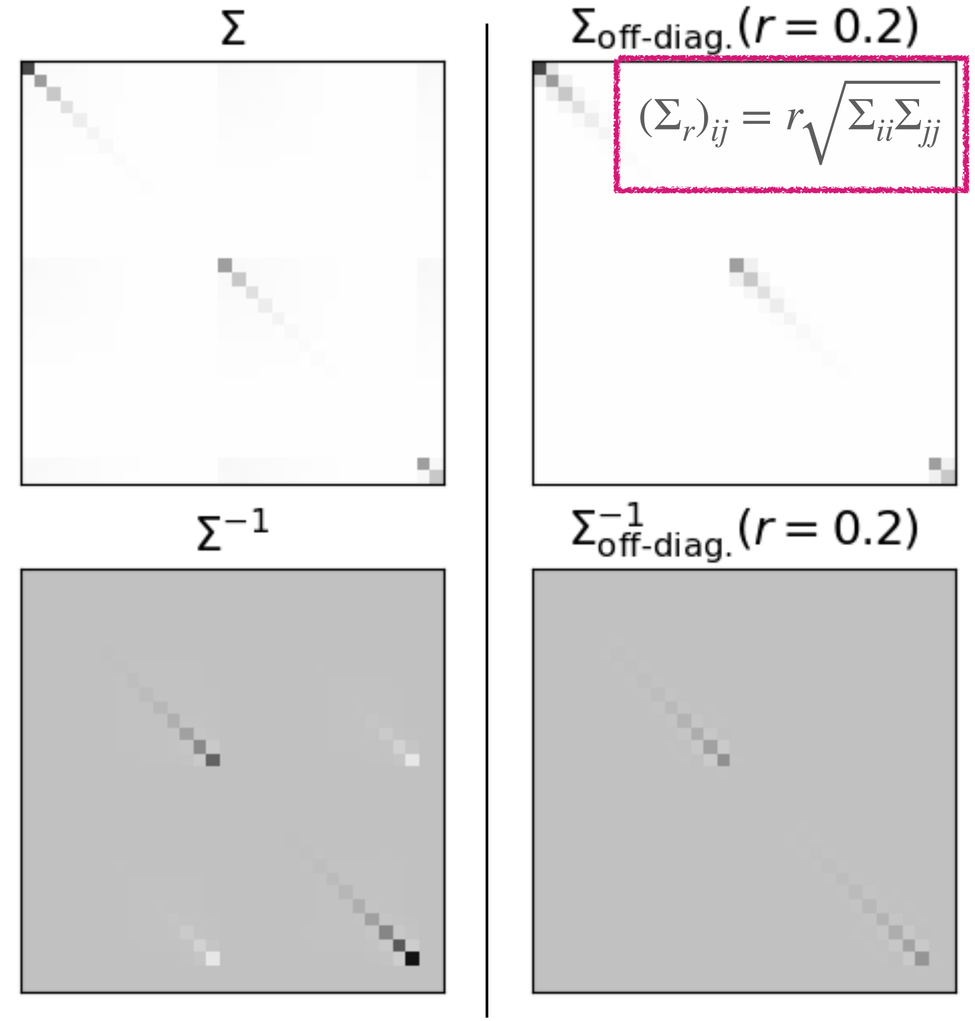
$n_s = 1000 \times 2$

$n_s = 4000$

$n_s = 4000 \times 2$

Where does NN compression fail?

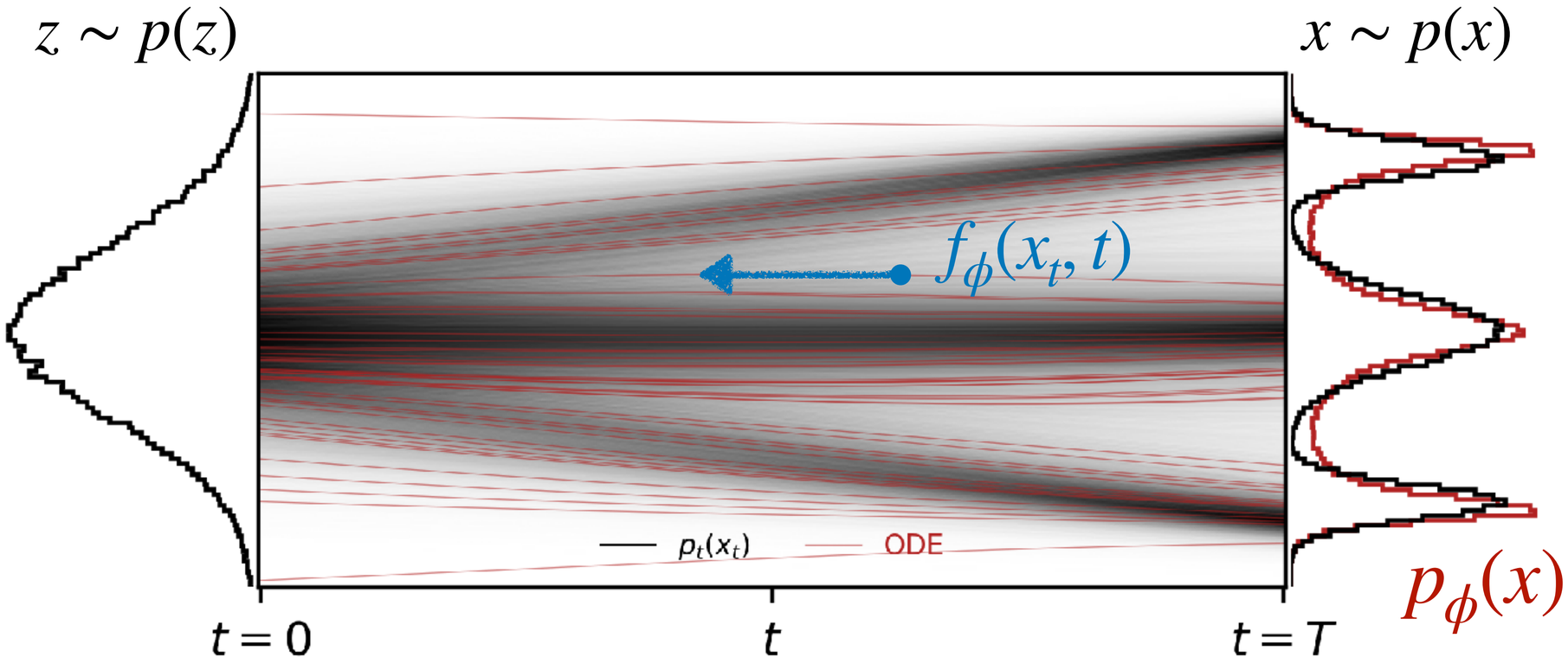




- NN fails to summarise when data has a (strongly) non-diagonal covariance!

Hyperparameter optimisation

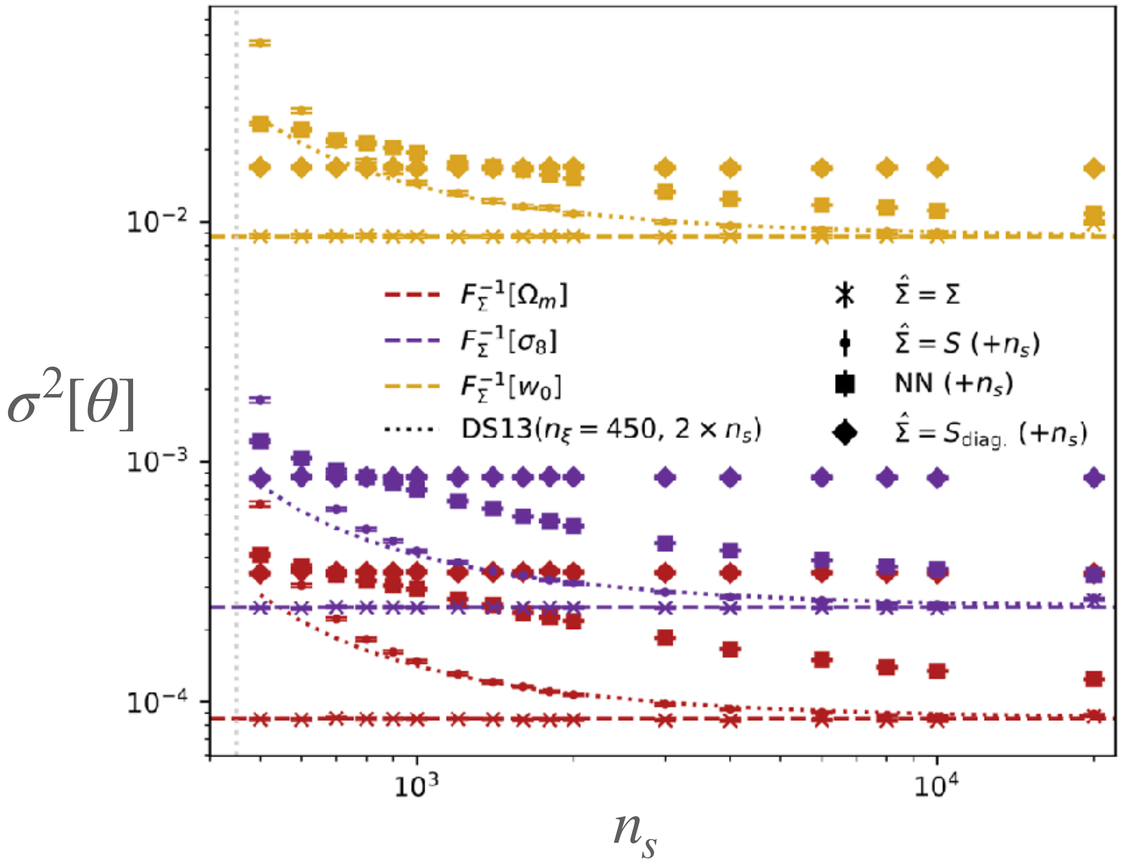
Continuous normalising flows



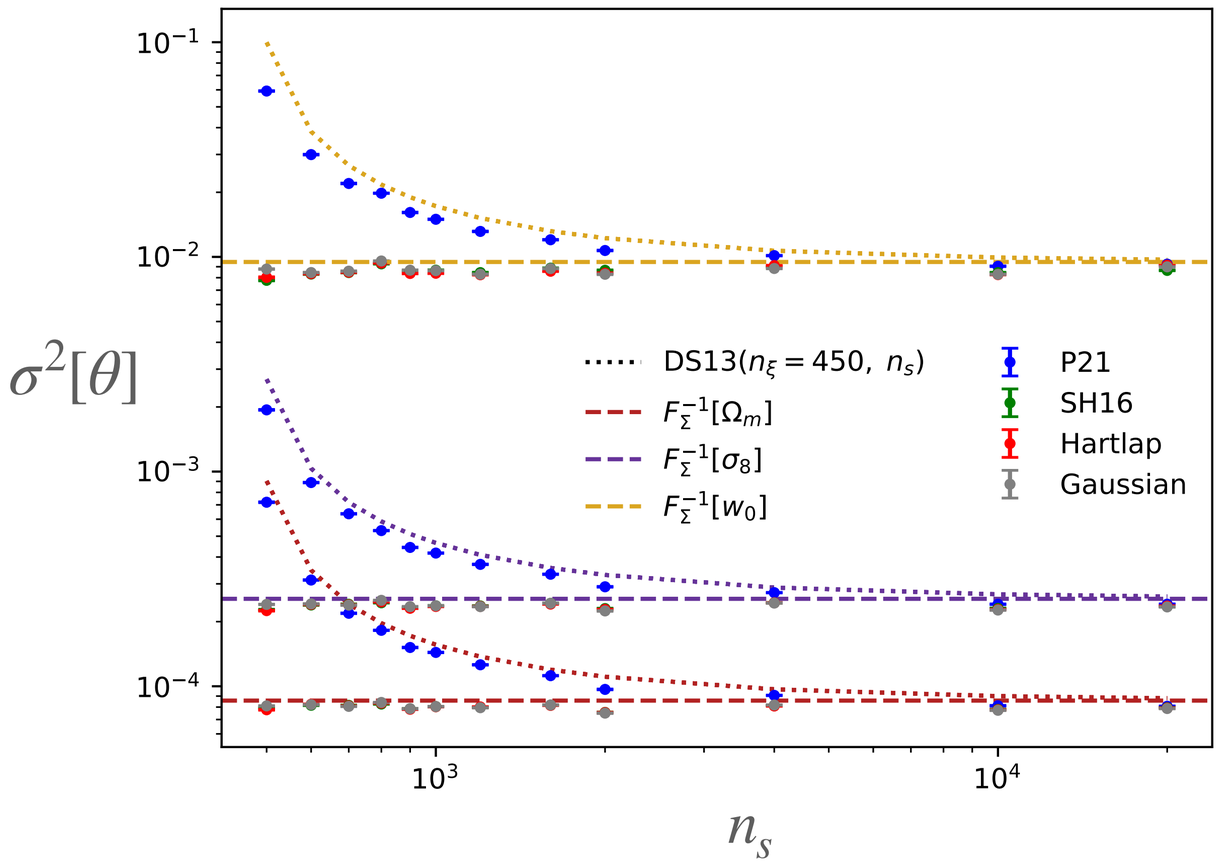
$$\log p_\phi(x) = \log p(z) + \int_{t=0}^{t=T} \mathrm{d}t \, \nabla_{x_t} \cdot f_\phi(x_t, t)$$

SBI vs. Gaussian likelihood analysis

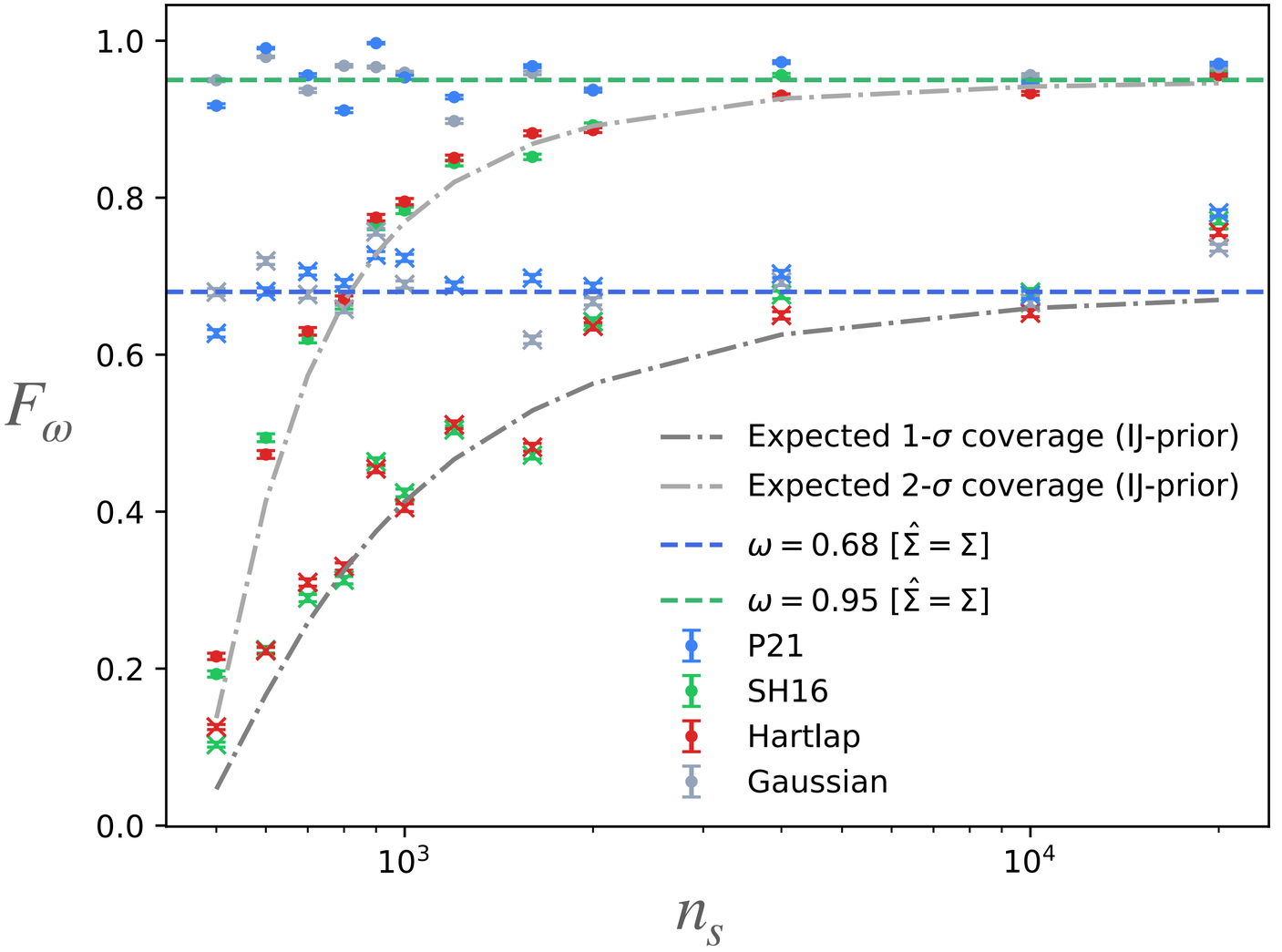
Posterior widths - SBI



Posterior widths - Gaussian likelihood analysis

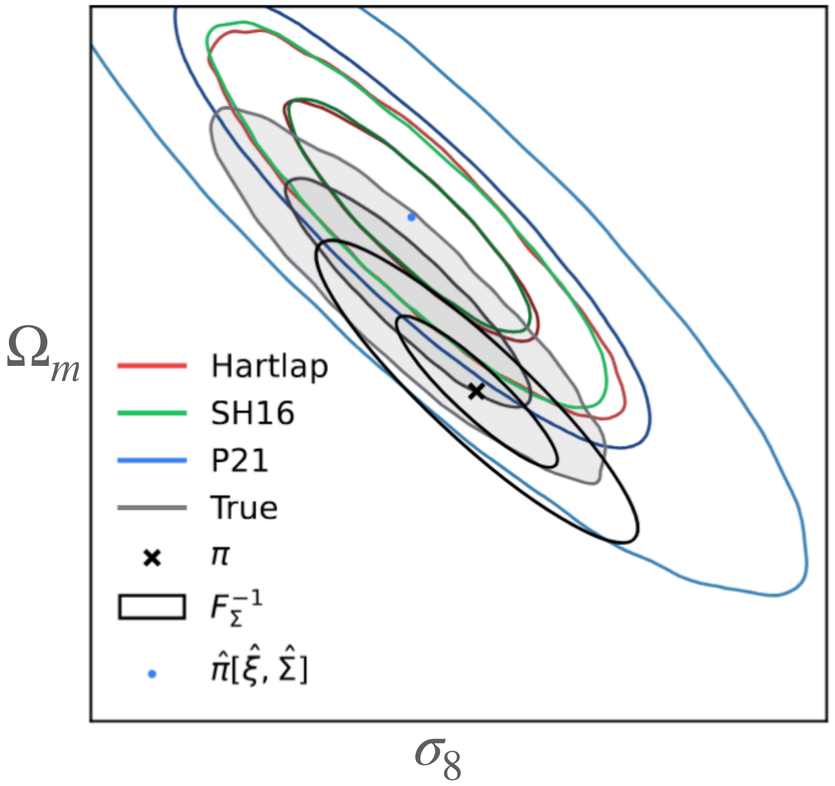


Posterior coverages - Gaussian likelihood analysis



The Percival++ posterior

Percival, Friedrich, Sellentin & Heavens 2021



$$p(\theta \,|\, \hat{x}, \hat{\Sigma}) = \int \mathrm{d}\Sigma \; p(\theta, \Sigma \,|\, \hat{x}, \theta)$$

$$= \int \mathrm{d}\Sigma \; p(\hat{x} \,|\, \theta, \Sigma) p(\hat{\Sigma} \,|\, \Sigma) p(\theta, \Sigma)$$

$$p(\theta, \Sigma) \propto |\Sigma|^{-\frac{m - n_S + n_X - 1}{2}}$$

$$\left\langle (\theta - \hat{\theta})(\theta - \hat{\theta})^T \right\rangle_{\theta|\hat{x},\hat{\Sigma},\hat{x}} = \left\langle (\hat{\theta} - \theta)(\hat{\theta} - \theta)^T \right\rangle_{\hat{\Sigma},\hat{x}} \rightarrow m$$