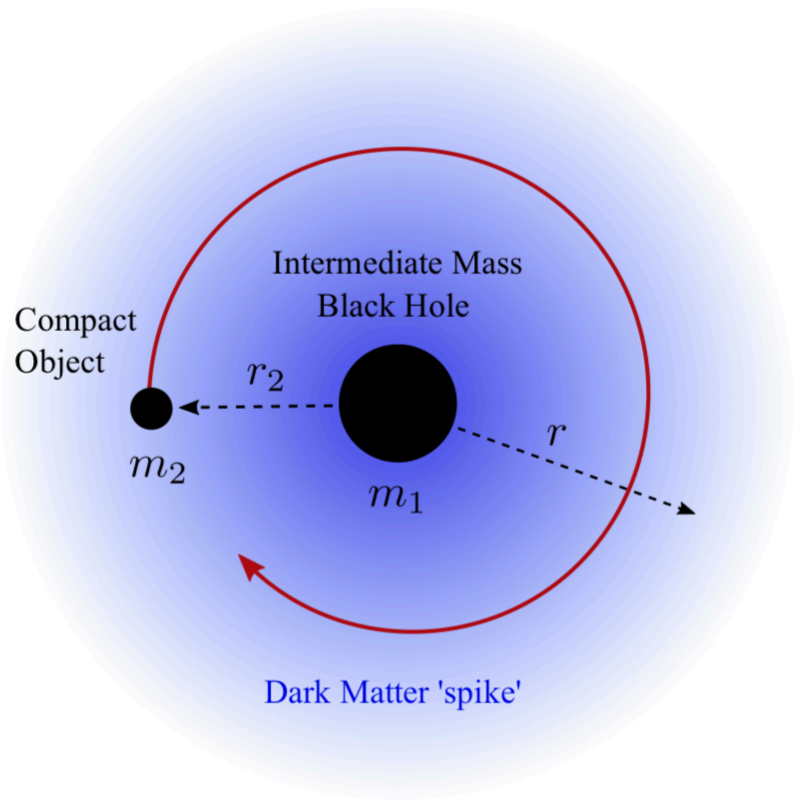
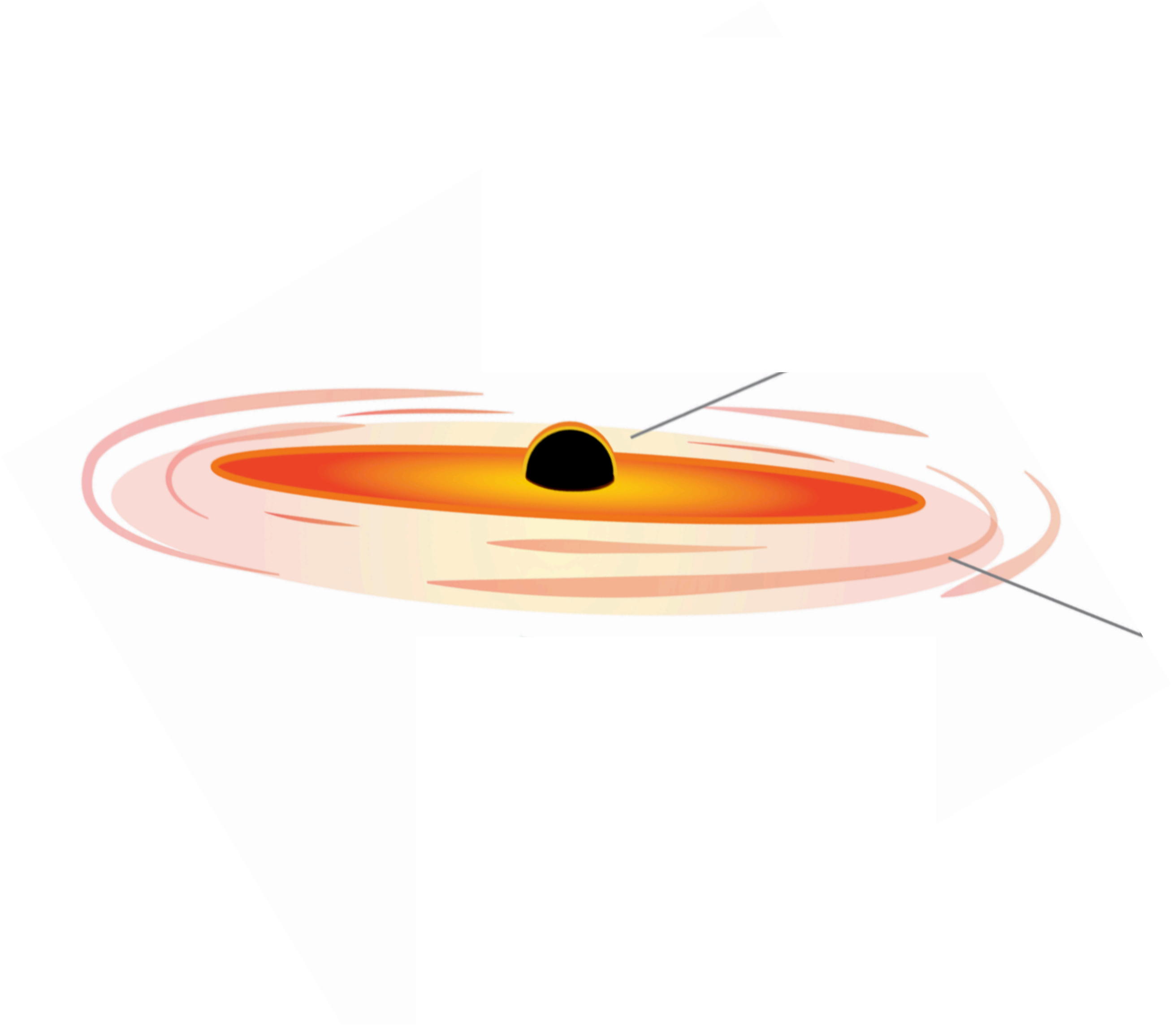
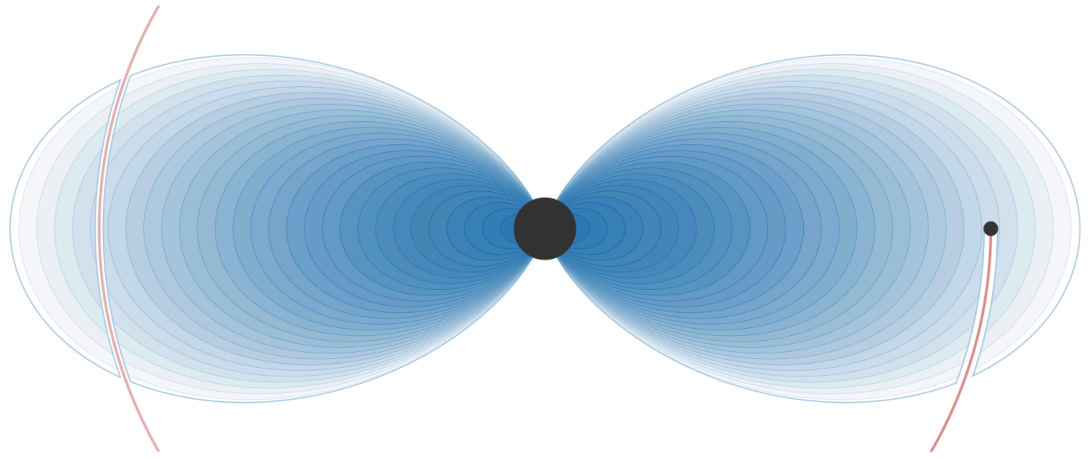
Why use machine-learning based tools for parameter estimation?

A dark matter case study







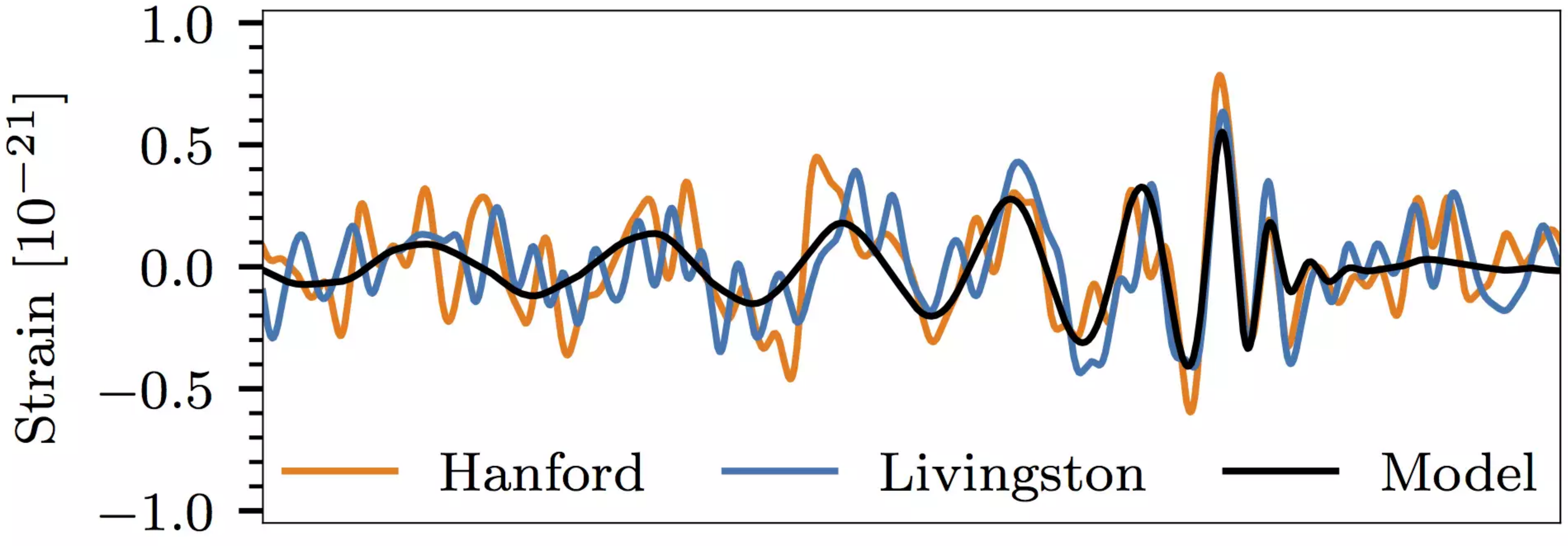
Philippa Cole, Postdoc Milano-Bicocca

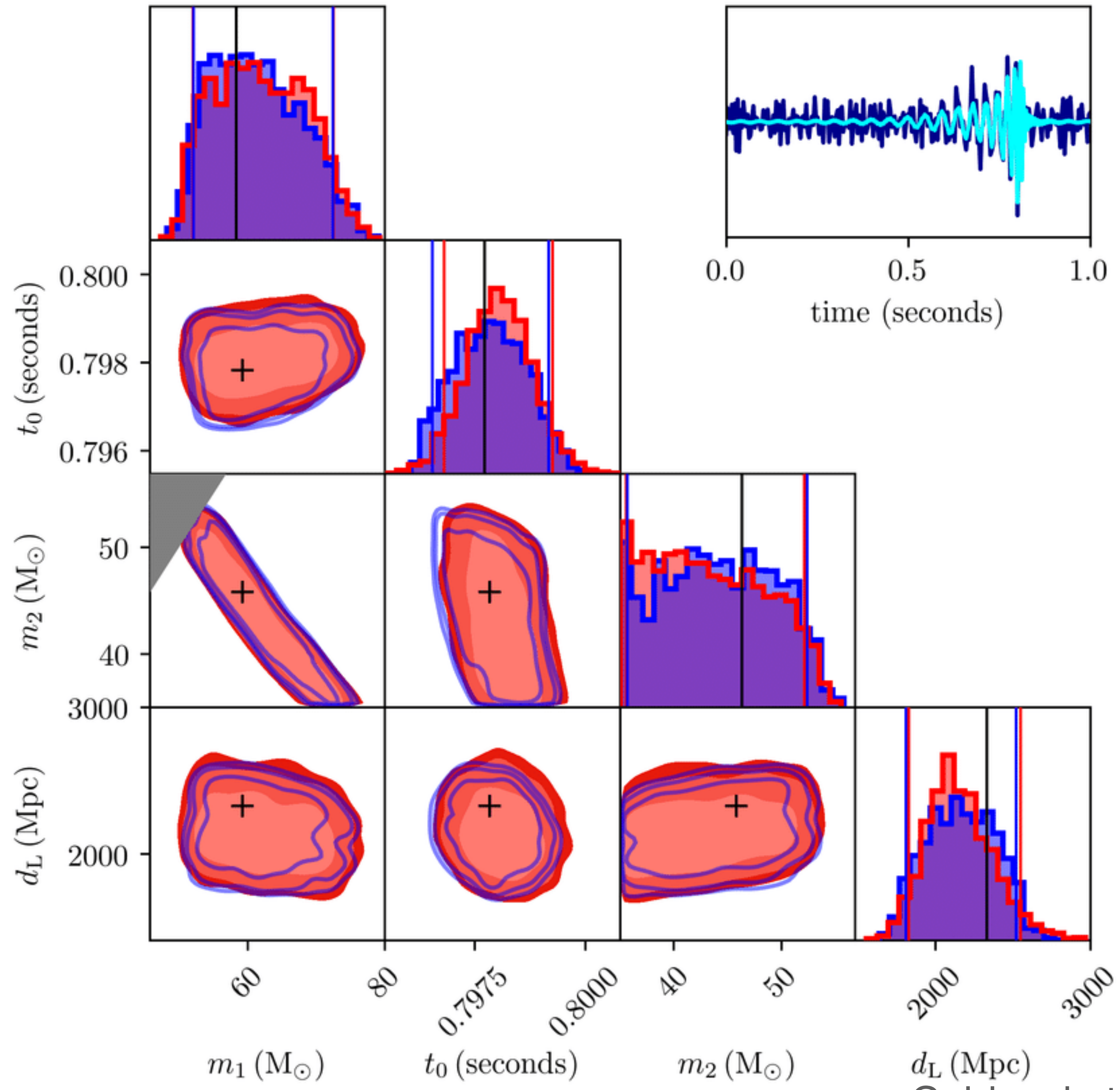
philippa.cole@unimib.it

Parameter estimation

- Given some data x
- And a model which is described by parameters θ
- Want to understand which values of the parameters in the model best fit the data.
- This is described by the posterior distributions $p(\theta|x)$ of the parameters θ
- Posteriors can be calculated according to Bayes theorem:

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





Gabbard et al 2019

MCMC and nested sampling tools can be used to calculate or estimate these posterior distributions

- They work really well when you have a likelihood, and when your data isn't too large

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

- But what if one or both of these two things aren't true for your problem?

- Simulation-based inference methods (also known as likelihood-free) can help

- Note that there are many other machine-learning based methods to tackle different problems with traditional sampling methods

SBI

Instead of likelihood - use a simulator or forward model

- Given a model $p(\theta, x)$ that takes some parameters θ and produces some simulated data x, we can see that this is equivalent to being able to sample from the likelihood (if we had known it) because:

- $p(\theta, x) = p(x | \theta)p(\theta)$

- In a nutshell: simulate some data, train a neural network to understand the mapping between the data and the parameters of the model, compute the posteriors

SBI

Instead of likelihood - use a simulator or forward model

- $p(\theta, x) = p(x \mid \theta)p(\theta)$

Different recipes

- Neural Posterior Estimation - directly estimate the posterior density $p(\theta \mid x)$ using e.g. normalising flows.

- Neural Likelihood Estimation - estimate the simulated data likelihood $p(x \mid \theta)$ and then use it to do MCMC or nested sampling.

- Neural Ratio Estimation - estimate the likelihood-to-evidence ratio $p(x \mid \theta)/p(x)$

SBI

Instead of likelihood - use a simulator or forward model

- $p(\theta, x) = p(x \mid \theta)p(\theta)$

Different recipes

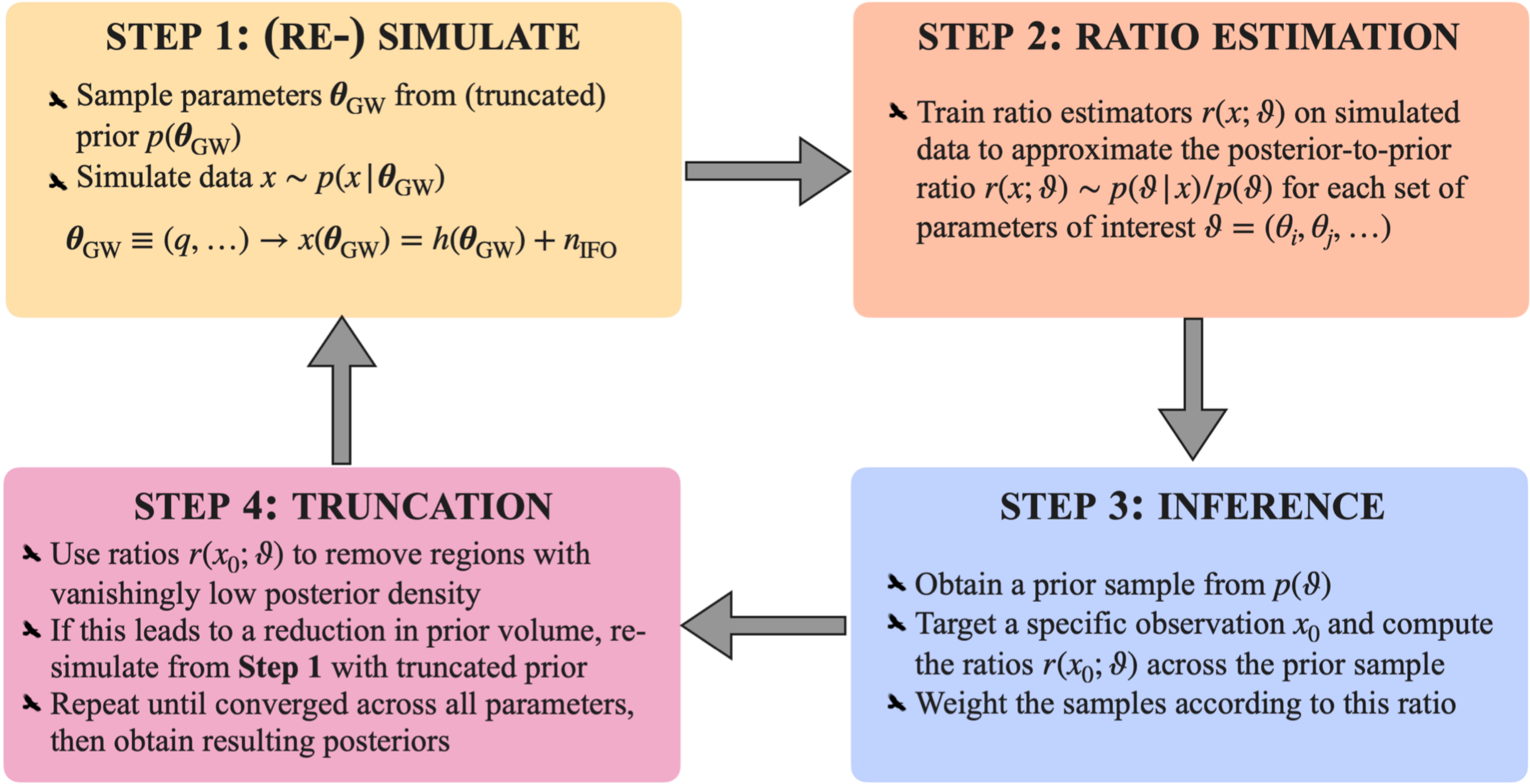
- Neural Posterior Estimation - directly estimate the posterior density $p(\theta \mid x)$ using e.g. normalising flows.

- Neural Likelihood Estimation - estimate the simulated data likelihood $p(x \mid \theta)$ and then use it to do MCMC or nested sampling.

- Neural Ratio Estimation - estimate the likelihood-to-evidence ratio $p(x \mid \theta)/p(x)$

SBI

Truncated Marginal Neural Ratio Estimation



Bhardwaj et al 2023

The loss function

- The training step requires you to minimise a loss function that teaches the network to recognise how your parameters relate to your data

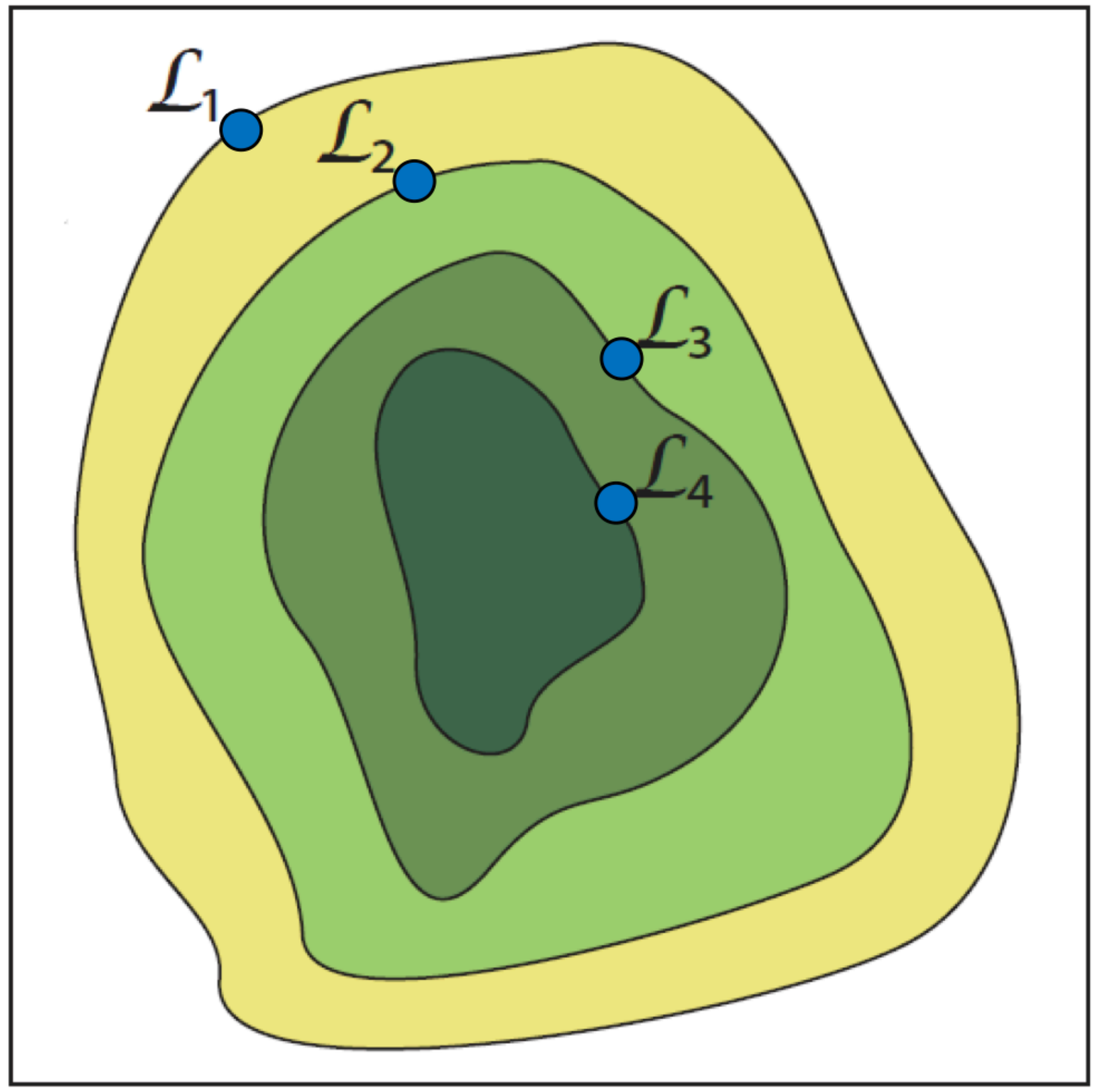
- In TMNRE, it turns this into a binary classification problem, so that the network learns to recognise jointly drawn pairs of parameters and data versus marginally (or independently) drawn after they've been shuffled

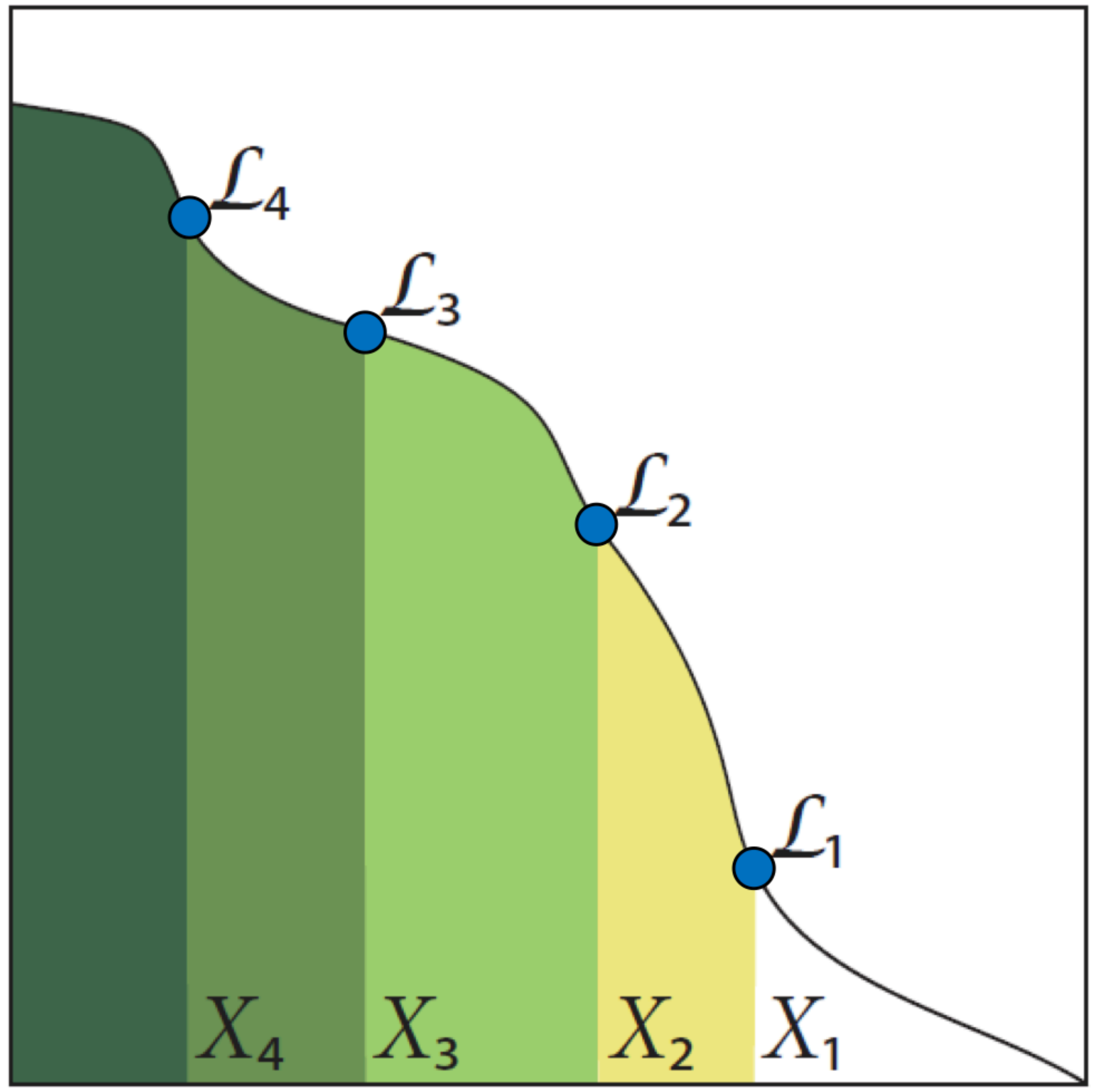
- The clever part is, that the classifier f_ϕ, which minimises the loss function is the log of the likelihood-to-evidence ratio $r(x, \theta)$

$$\mathcal{L}[f_\phi] = -\int \mathrm{d}x\mathrm{d}\vartheta \left[p(x,\vartheta)\ln\left(\sigma(f_\phi(x,\vartheta))\right) \right.$$
$$\left. + \, p(x)p(\vartheta)\ln\left(1-\sigma(f_\phi(x,\vartheta))\right)\right]$$

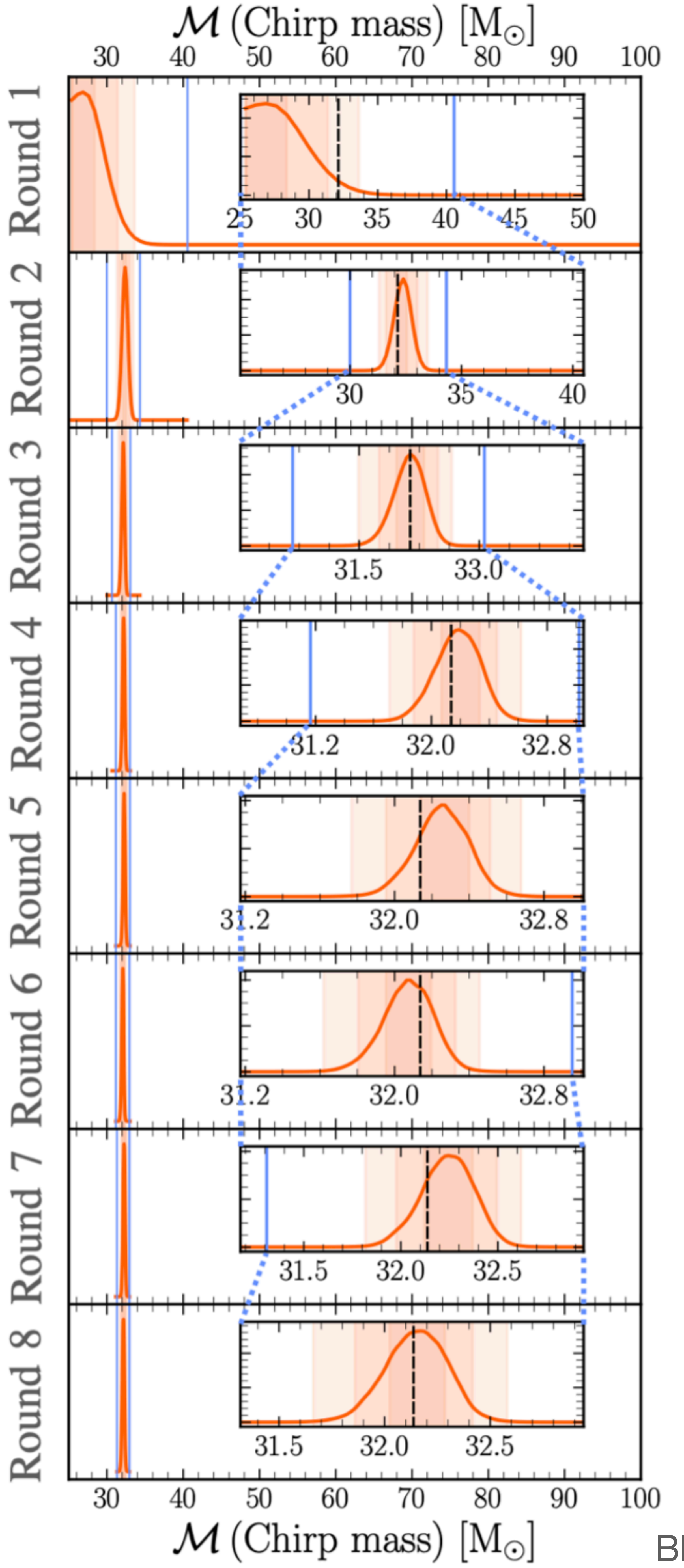
Truncated Marginal Neural Ratio Estimation

Truncation scheme is analogous to nested sampling contours of iso-likelihood





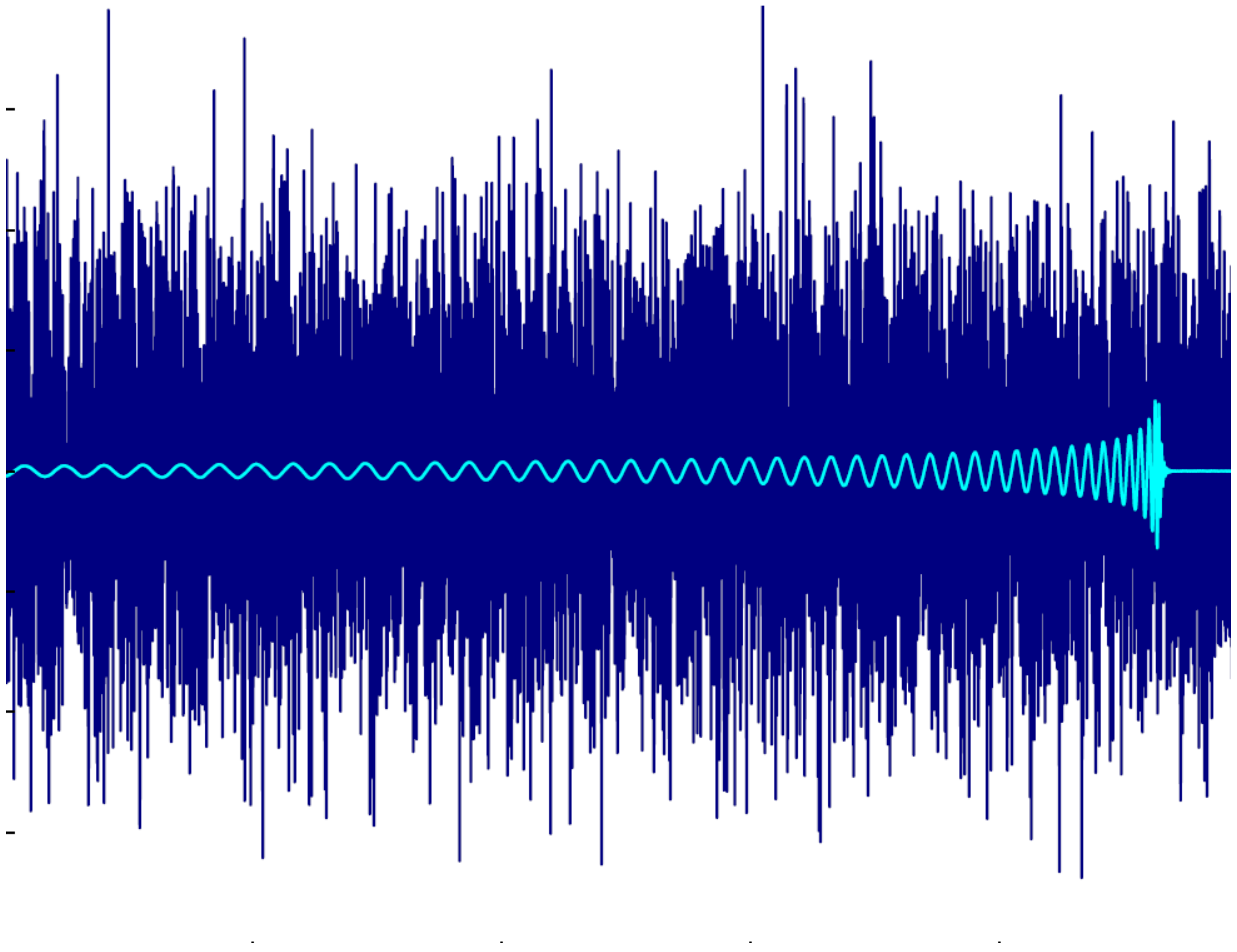
Feroz et al 2013



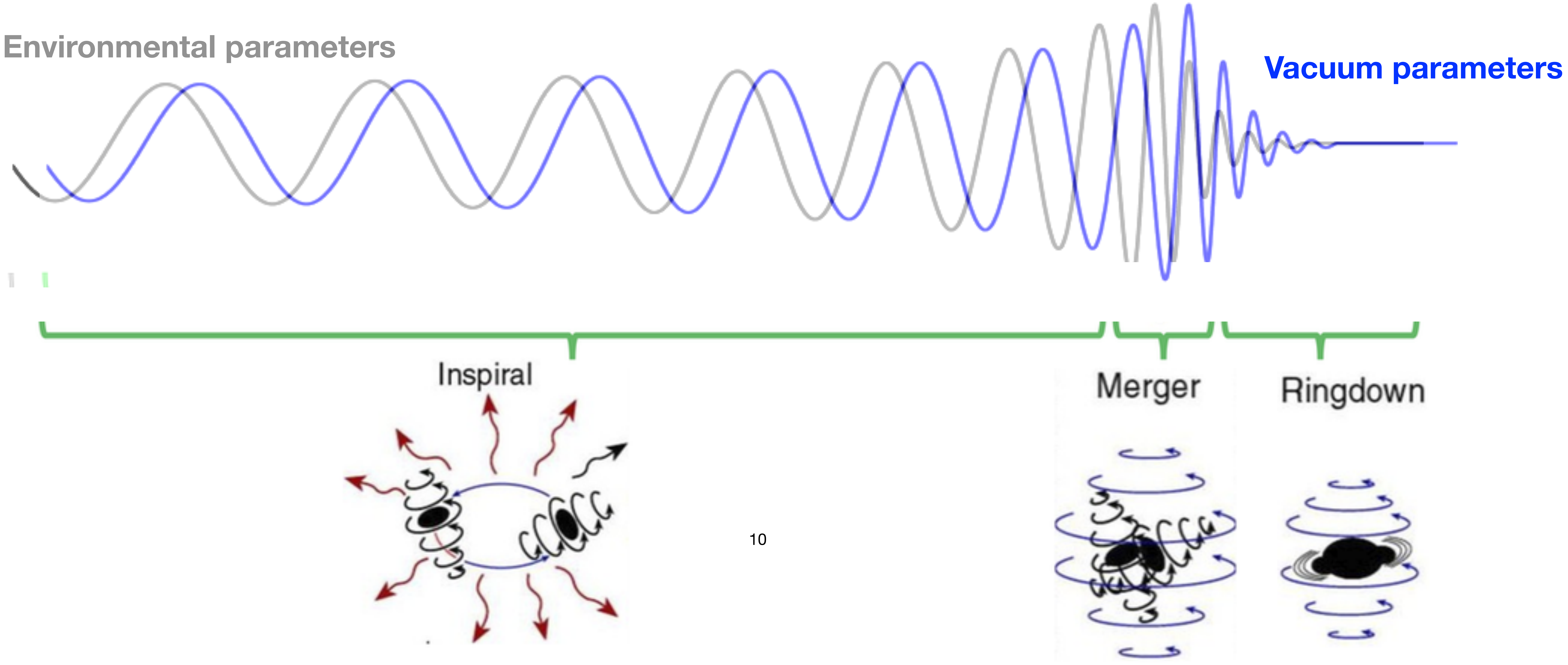
Bhardwaj et al. 2023

A case study, first with nested sampling: Vacuum or non-vacuum

- So far, all gravitational wave detections of LIGO/Virgo/KAGRA binary black hole mergers have been detected and measured assuming that they occurred in vacuum
- OK for short duration signals, but looking towards future interferometers, long duration signals may be affected by their environment



Over a long period of time, signals can look different depending on their environment, which manifests as a phase shift.



- Environmental effects can cause inspiral to either speed up or slow down with respect to vacuum case

- A dephasing to accumulate, which alters the gravitational waveform from the binary's inspiral

Vacuum parameters | Environment parameters

$$\dot{r} = \dot{r}_{\mathrm{GW}} + \dot{r}_{\mathrm{env}}$$

$$f(t) = \frac{1}{\pi}\sqrt{\frac{GM}{r(t)^3}}$$

$$\Phi(f) = \int_f^{f_{\mathrm{ISCO}}} \frac{\mathrm{d}t}{\mathrm{d}f'} f' \,\mathrm{d}f'$$

$$h_0(f) = \frac{1}{2}\frac{4\pi^{2/3}G_N^{5/3}\mathcal{M}^{5/3}f^{2/3}}{c^4}\sqrt{\frac{2\pi}{\ddot{\Phi}}}$$

Model

- Environmental effects can cause inspiral to either speed up or slow down with respect to vacuum case

- A dephasing to accumulate, which alters the gravitational waveform from the binary's inspiral

Vacuum parameters

Environment parameters

$$\dot{r} = \dot{r}_{\mathrm{GW}} + \dot{r}_{\mathrm{env}}$$

$$f(t) = \frac{1}{\pi}\sqrt{\frac{GM}{r(t)^3}}$$

$$\Phi(f) = \int_f^{f_{\mathrm{ISCO}}} \frac{\mathrm{d}t}{\mathrm{d}f'} f' \,\mathrm{d}f'$$

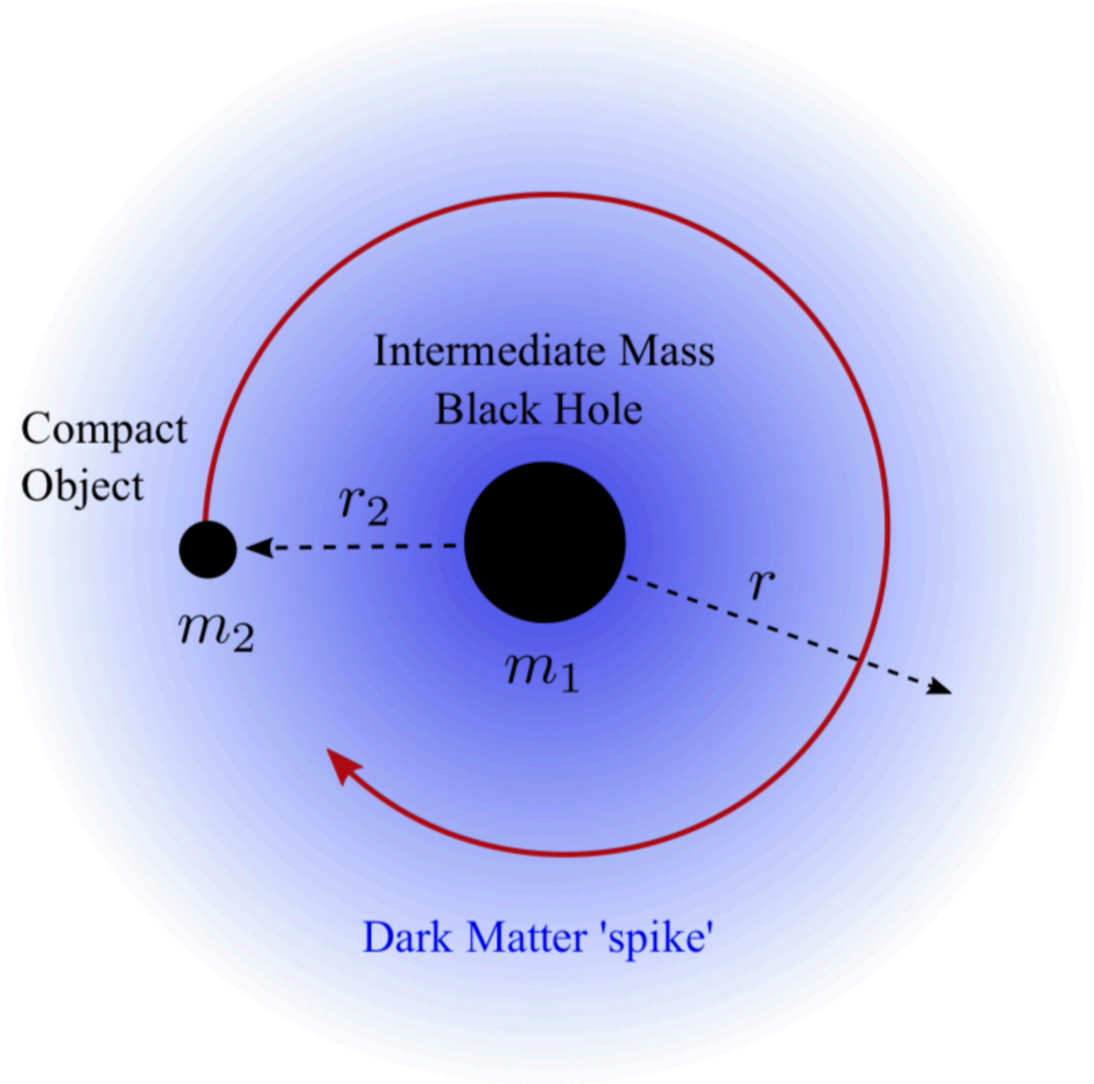
$$h_0(f) = \frac{1}{2}\frac{4\pi^{2/3}G_N^{5/3}\mathcal{M}^{5/3}f^{2/3}}{c^4}\sqrt{\frac{2\pi}{\ddot{\Phi}}}$$

Model

$$+n(f) = \text{Data}$$

Dark dress

Cold, collisionless dark matter

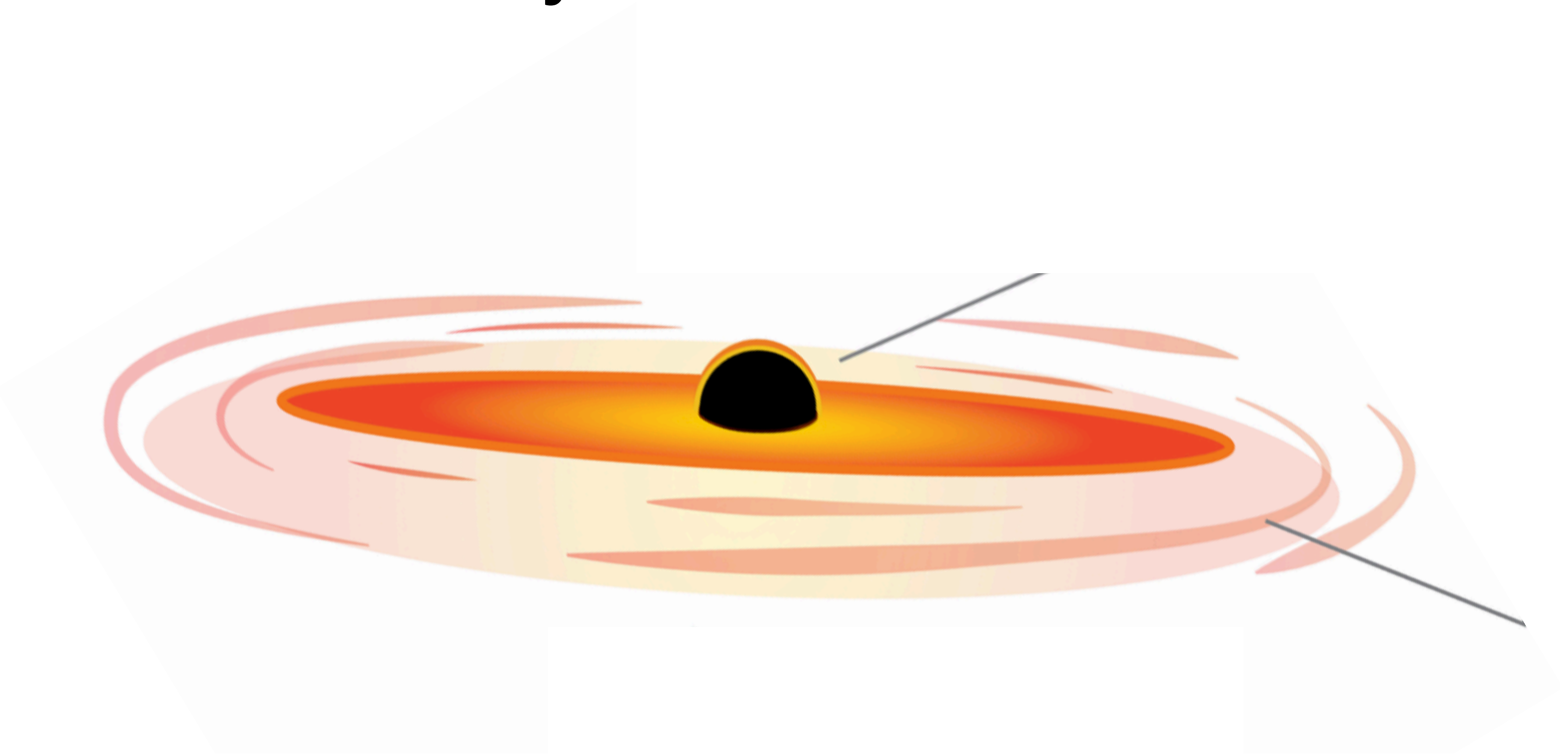


$$\rho(r) = \rho_6 \left(\frac{r_6}{r}\right)^{\gamma_s}$$

Eda et al. 2013, 2014
Gondolo, Silk 1999
Kavanagh et al. 2020
Coogan et al. 2021

Accretion disk

Baryonic matter



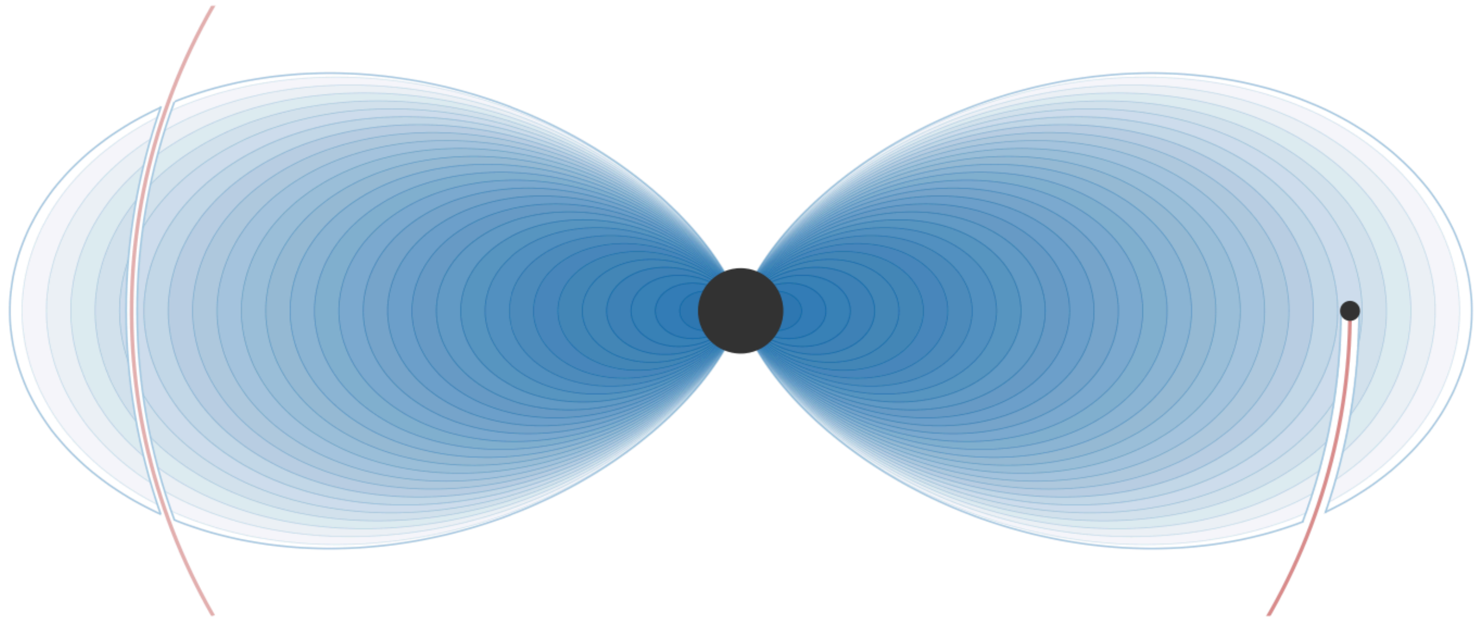
$$\Sigma(r) = \Sigma_0 \left(\frac{r}{r_0}\right)^{-1/2}$$

$$M = r/h$$

Goldreich & Tremaine 1980
Tanaka 2002
Derdzinski et al. 2020

Gravitational atom

Ultra-light bosons



$$\rho(\vec{r}) = M_c |\psi(\vec{r})|^2$$

$$\alpha \equiv G m_1 \mu \ll 1$$

Mass of light scalar field
$(10^{-10} - 10^{-20}\,\mathrm{eV})$

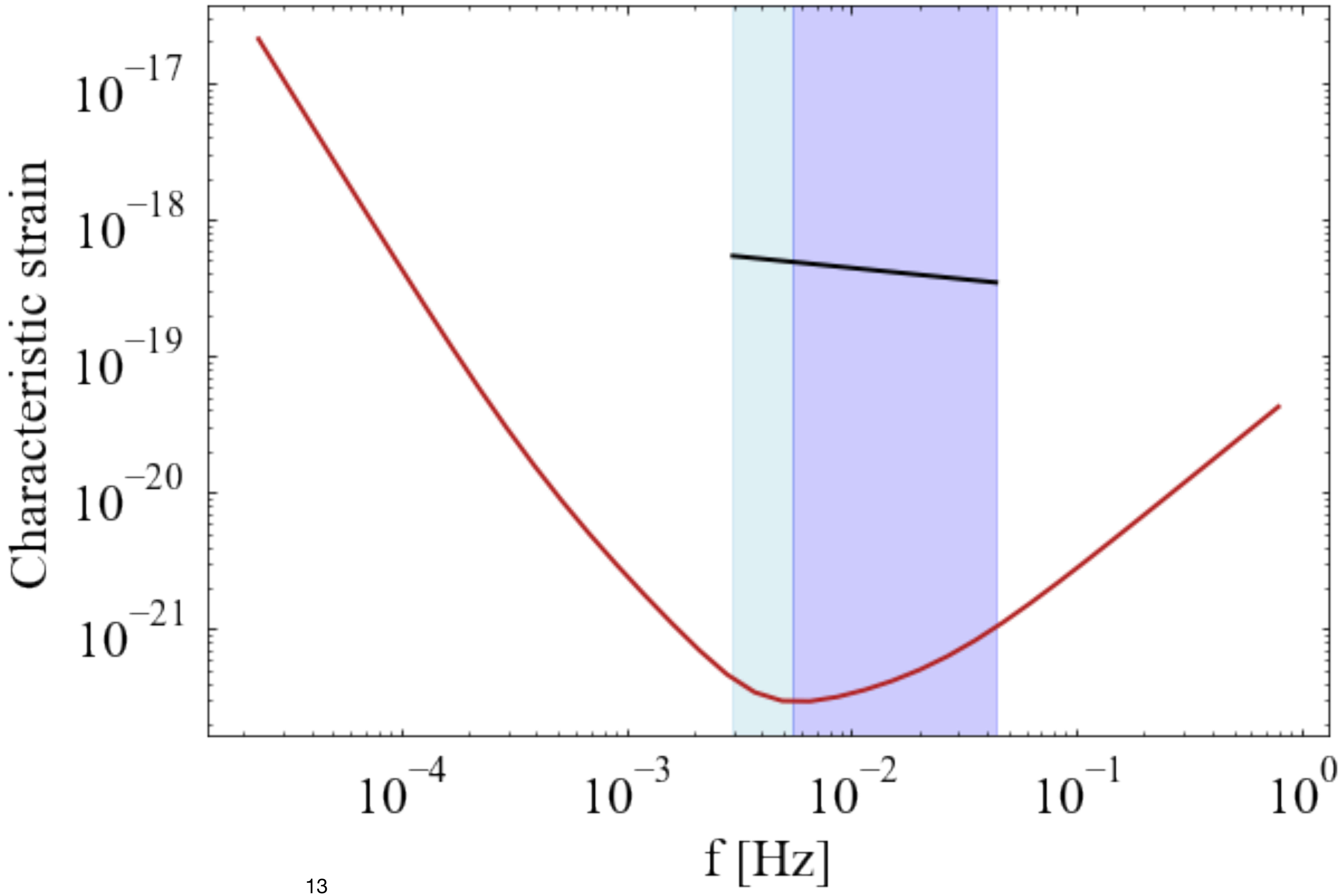
Baumann et al. 2019
Arvanitaki & Dubovsky 2010
Bauman et al. 2021, 2022

Credit: Sophia Dagnello, NRAO/AUI/NSF

Need to observe many cycles - i.e. long streams of data

- dephasing accumulates over thousands or millions of cycles
- small mass ratio $q = \frac{m_2}{m_1} < 10^{-2.5}$ so that environment survives
- systems possible sources for LISA and Einstein Telescope/Cosmic Explorer

$$m_1 = 10^5\,\mathrm{M}_\odot, \quad m_2 = 10\,\mathrm{M}_\odot$$



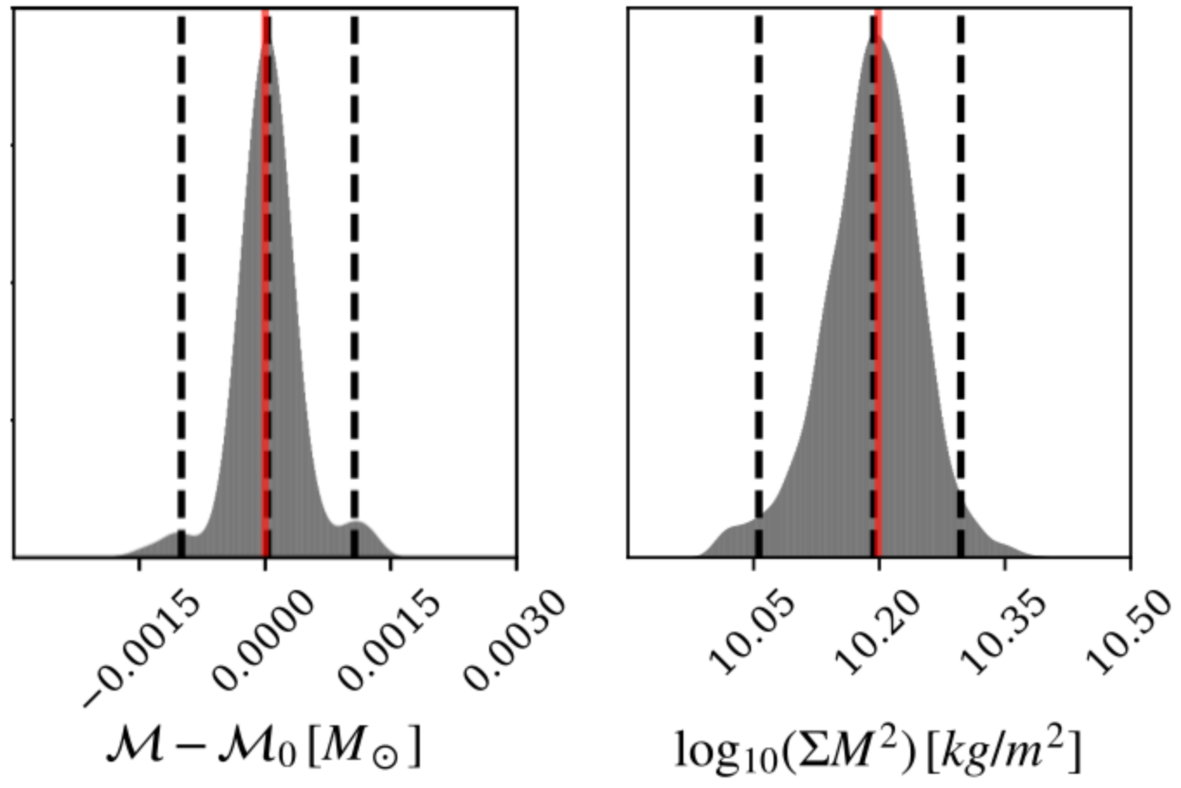
Why should we care about environmental effects?

- If we can measure the parameters of the environment via the dephasing in the waveform, chance to learn about the environment
- If we search the data with the wrong ‘template’ we might miss the signal
- If we do parameter estimation with the ‘wrong’ parameters, we might come up with biased results

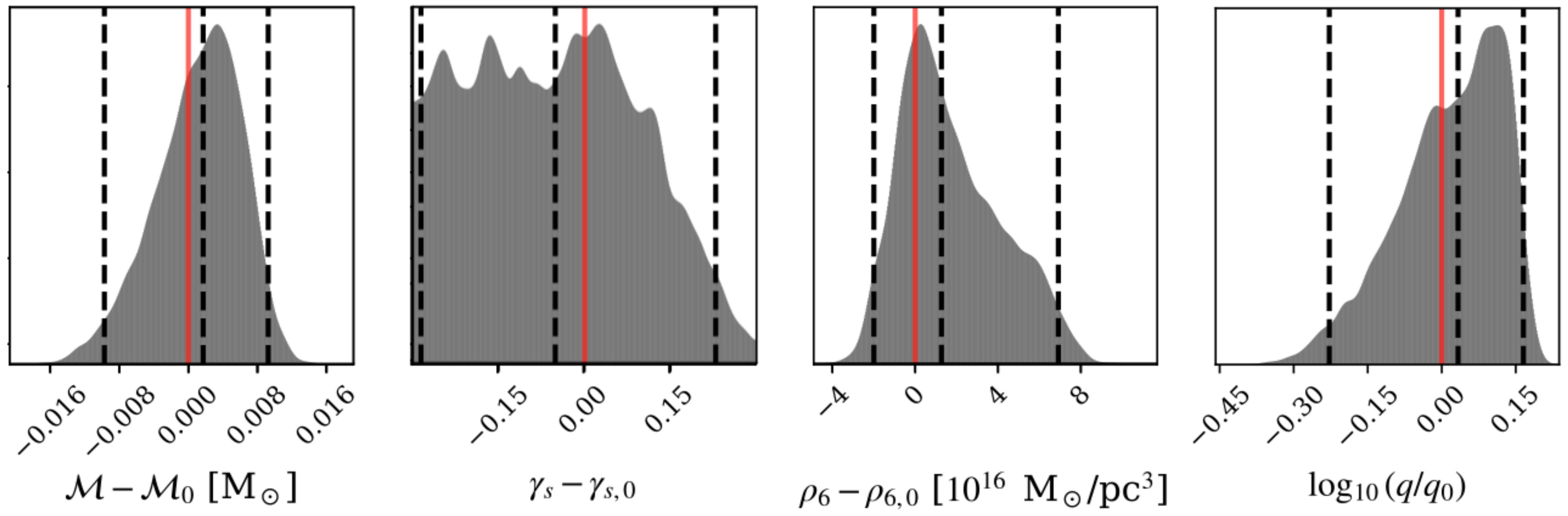
See also Barausse, Cardoso, Pani 2011

First attempt:
Parameter estimation using nested sampling

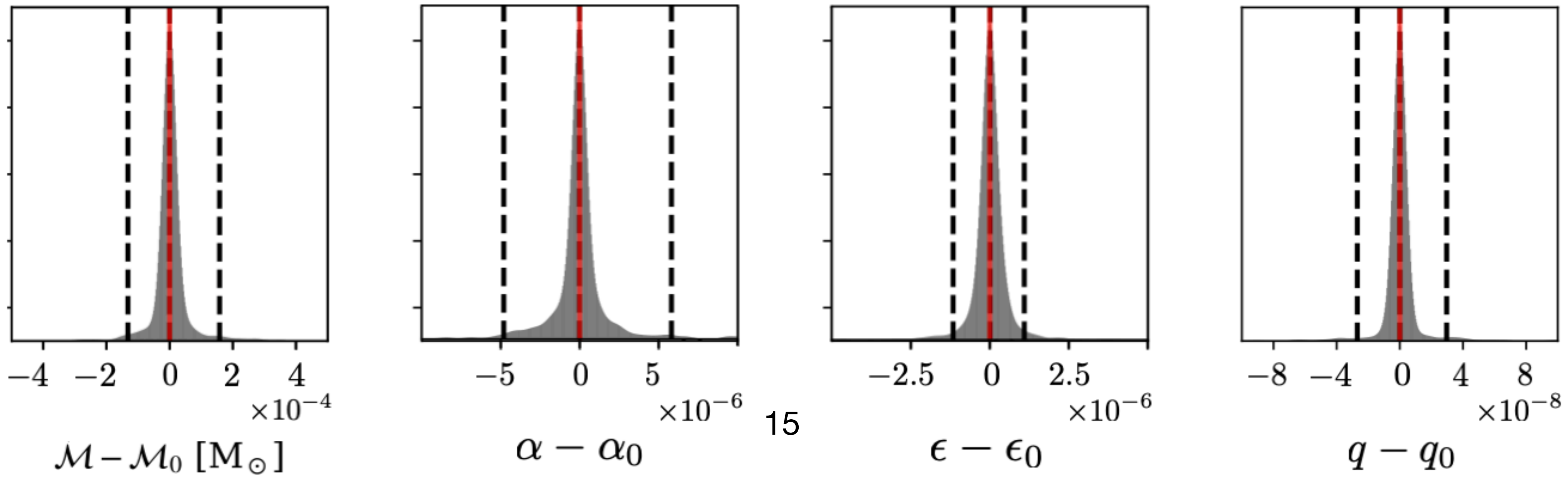
Accretion disk



Dark dress



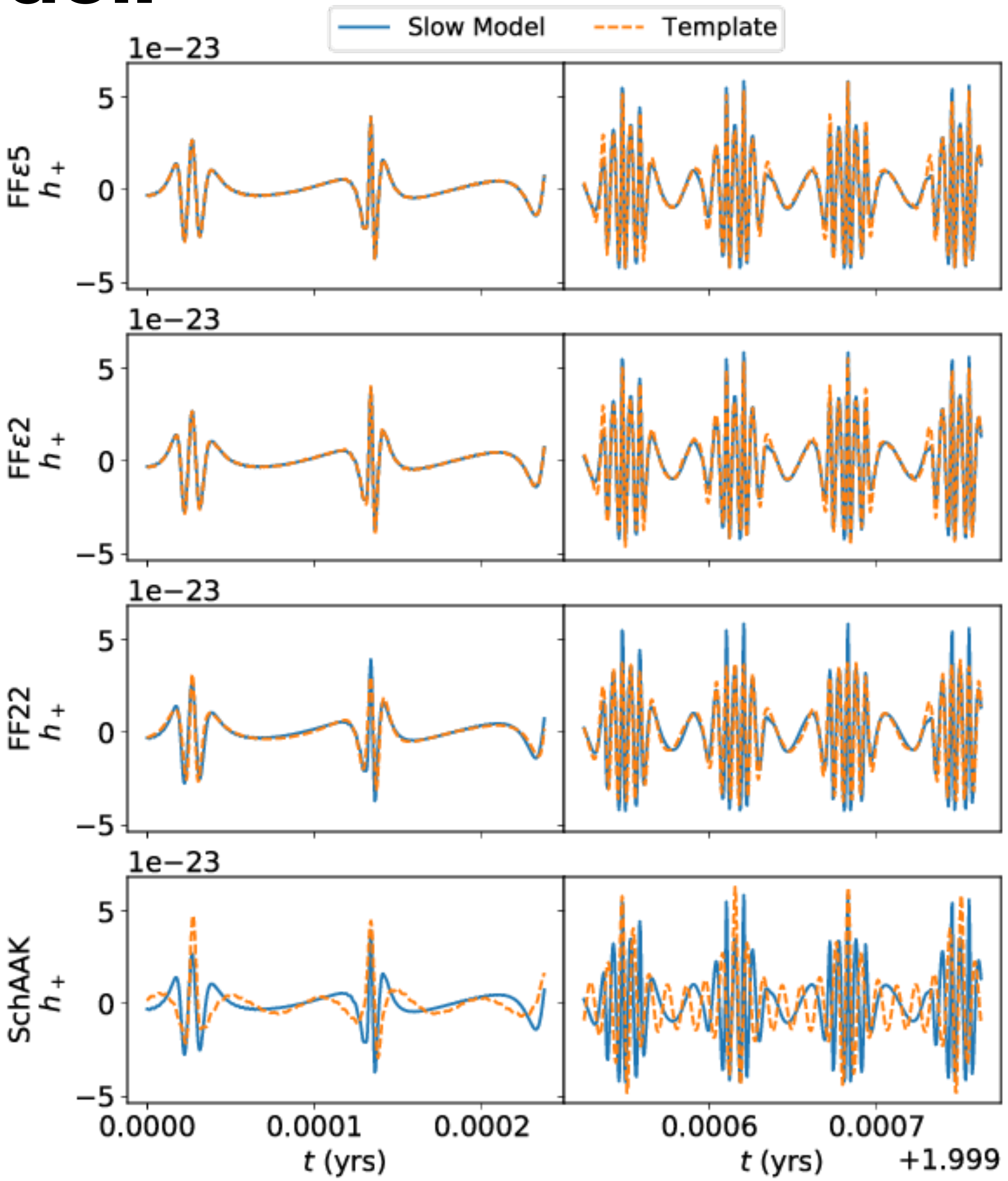
Gravitational atom



Cole et al. 2023

This was already computationally expensive in 4d… plus there's lots to add to the model:

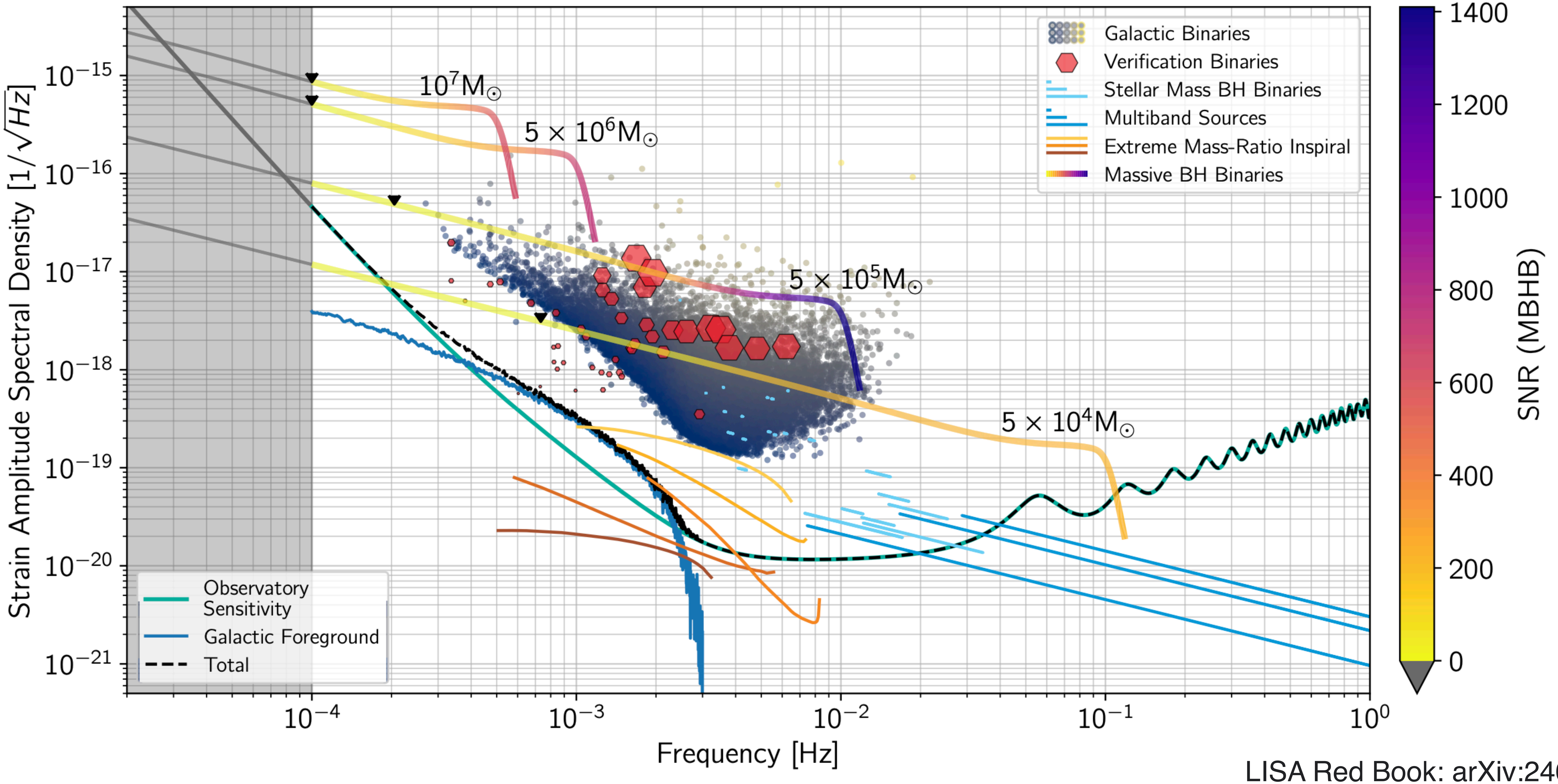
- Full parameter space to check for degeneracies with extrinsic parameters
- Include spins, eccentricity (EMRI waveforms (R))
- Include relativistic corrections
- Model the environments themselves more carefully



Katz et al. *Phys.Rev.D* 104 (2021) 6, 064047

As well as to the noise:

This part extremely tricky with likelihood-based analyses



LISA Red Book: arXiv:2402.07571

Our data is noisy and long… SBI is perfect!

With James Alvey and Uddipta Bhardwaj

- Want to flexibly add complexity to the signal and the noise models
- Likelihood-based methods expensive and maybe intractable for long duration signals including realistic noise
- Can simulation based inference help?

PEREGRINE

Gravitational Wave Parameter Inference with Neural Ratio Estimation

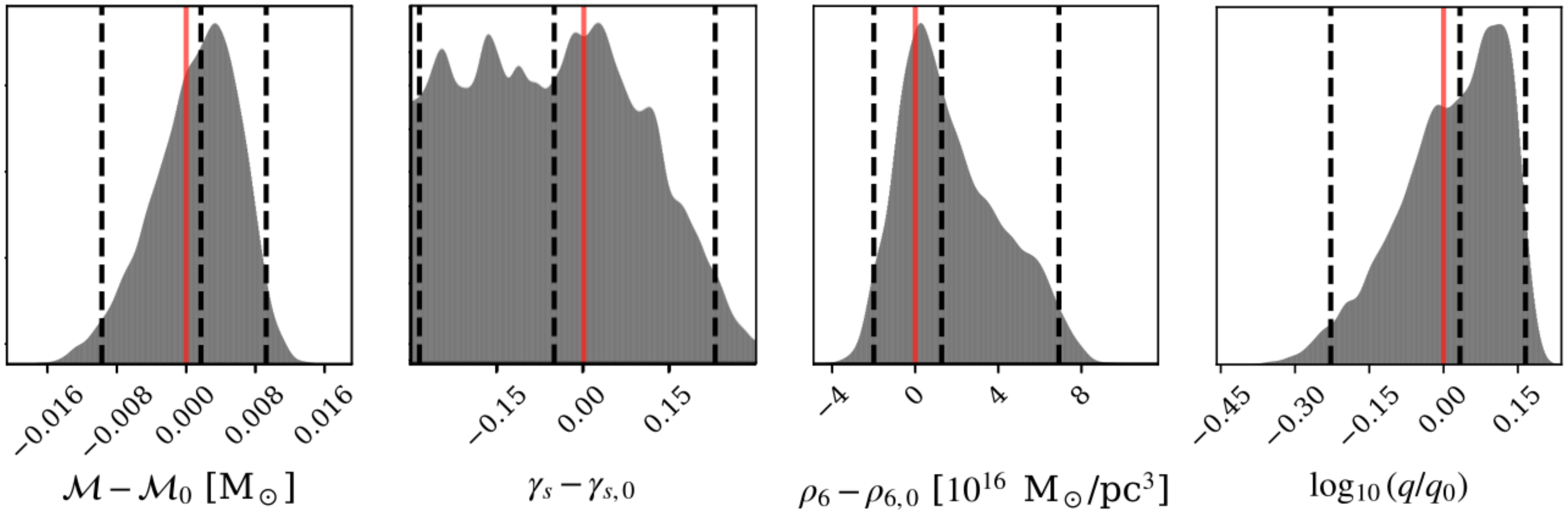
- Peregrine: Truncated Marginal Neural Ratio Estimation set up for gravitational wave signals

Bhardwaj et al, *Phys.Rev.D* 108 (2023) 4, 042004

Nested sampling expensive, and likelihood can't describe complicated noise

Current likelihood based method

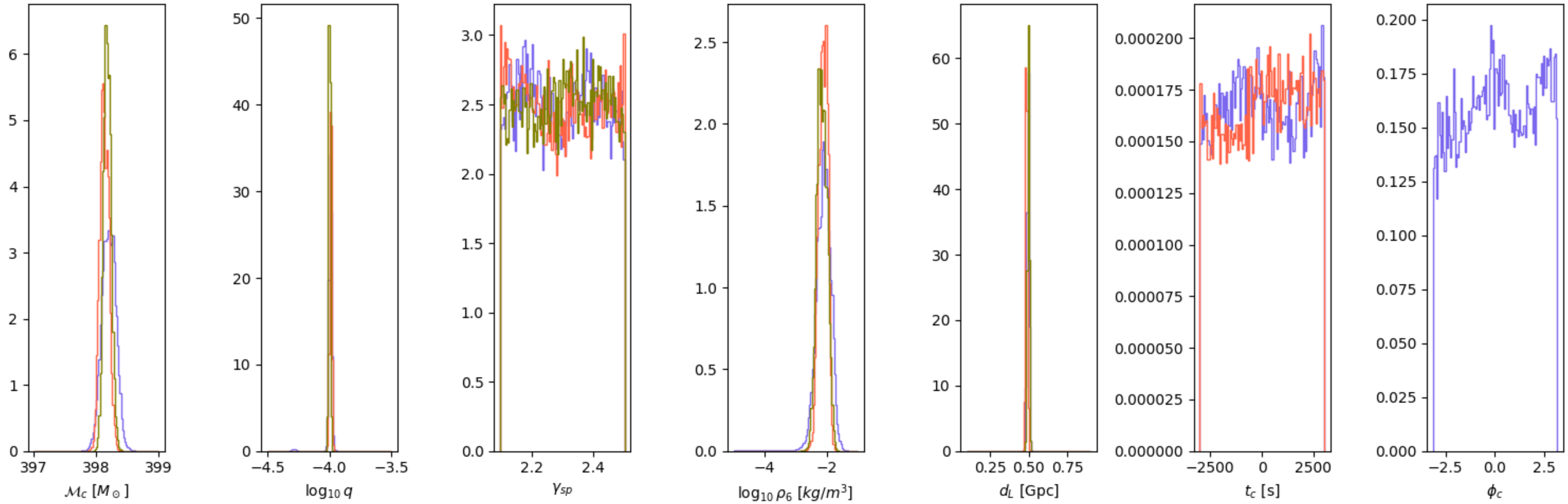
~2 million waveform evaluations for nested sampling, no noise



SBI less expensive, faster (GPUs) and noise included

30k waveform evaluations, plus (Gaussian) noise
Inference on new injection almost for free

With James Alvey and Uddipta Bhardwaj



(Moonshot) aim: full parameter space, realistic noise, EMRI waveforms

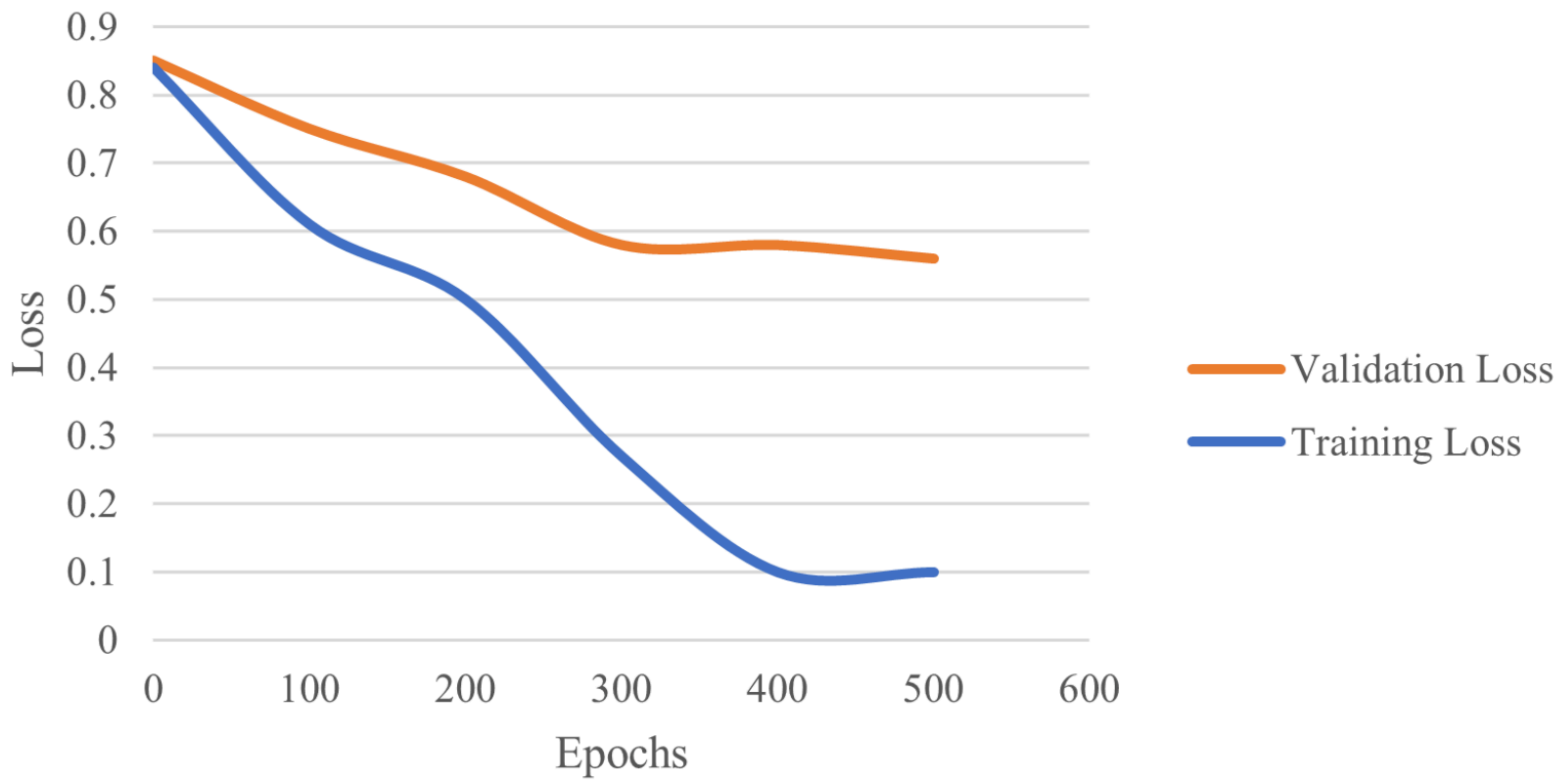
How do you know if it’s working?

- Training loss - the value of the loss function (which we’re trying to minimise) calculated according to the training data
- Validation loss - the value of the loss function (which we’re trying to minimise) calculated according to new data

Warning signs - underfitting

- Training loss - the value of the loss function (which we're trying to minimise) calculated according to the training data

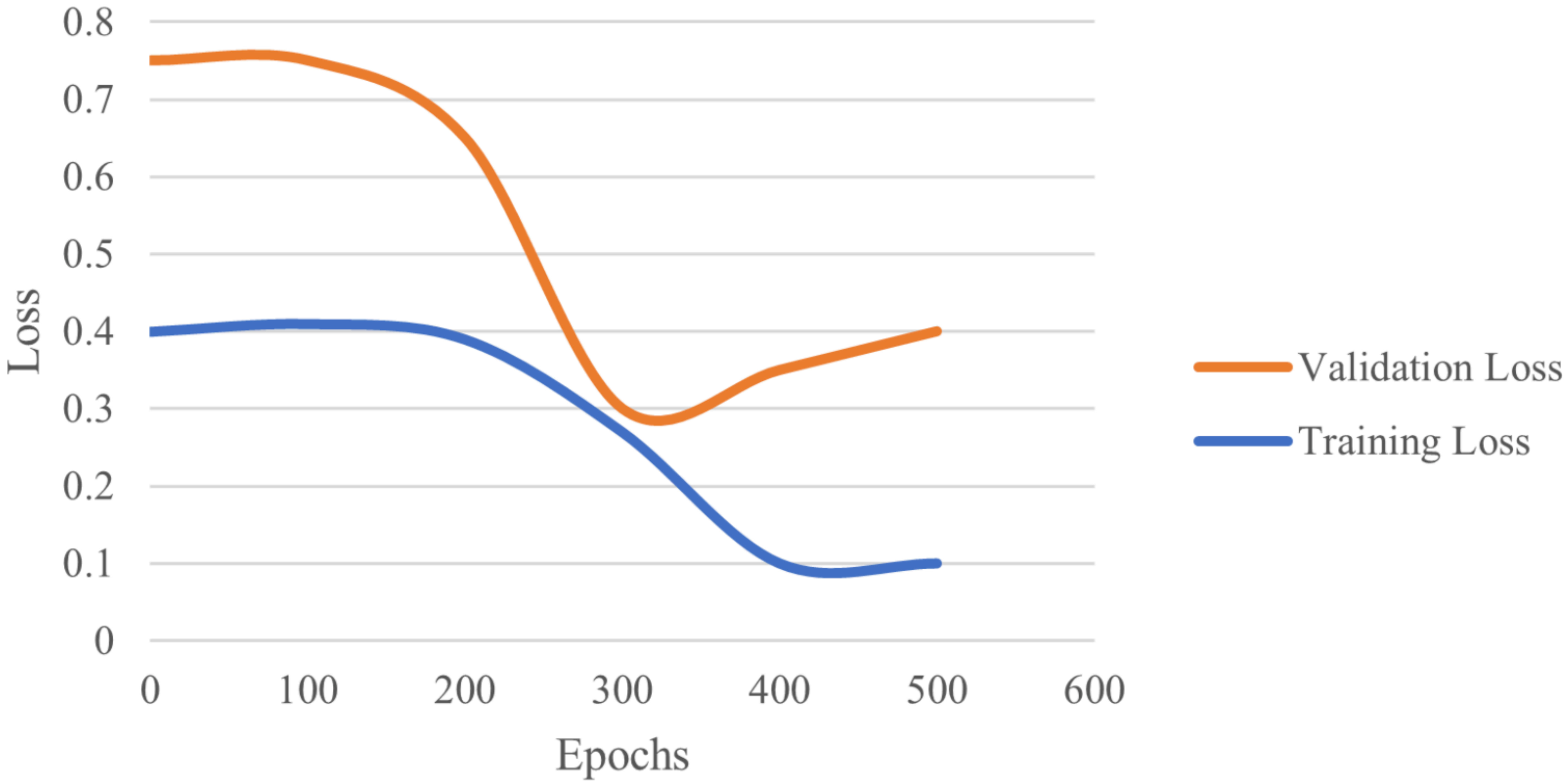
- Validation loss - the value of the loss function (which we're trying to minimise) calculated according to new data



Warning signs - overfitting

- Training loss - the value of the loss function (which we’re trying to minimise) calculated according to the training data

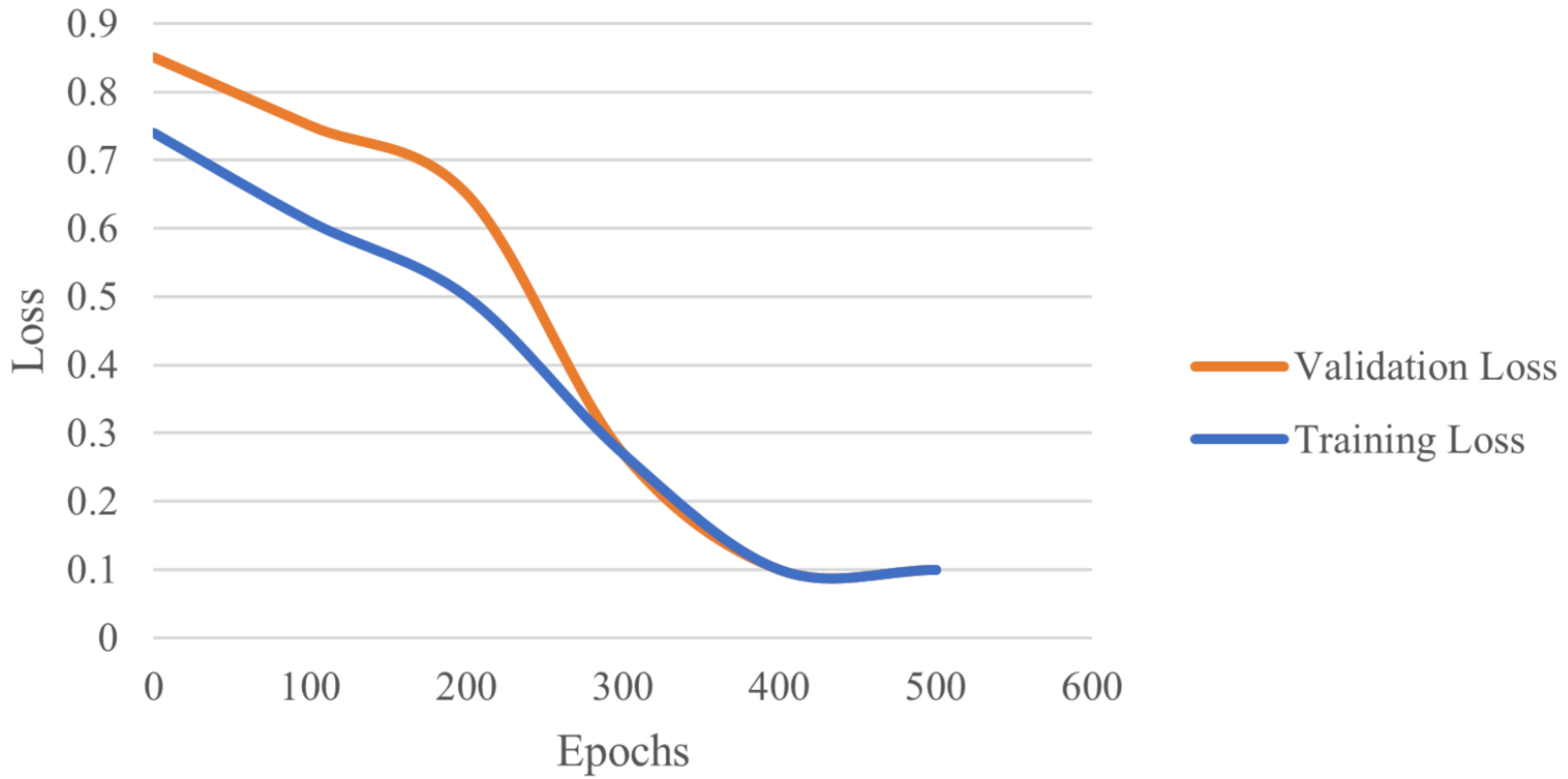
- Validation loss - the value of the loss function (which we’re trying to minimise) calculated according to new data



Just right…

- Training loss - the value of the loss function (which we're trying to minimise) calculated according to the training data

- Validation loss - the value of the loss function (which we're trying to minimise) calculated according to new data

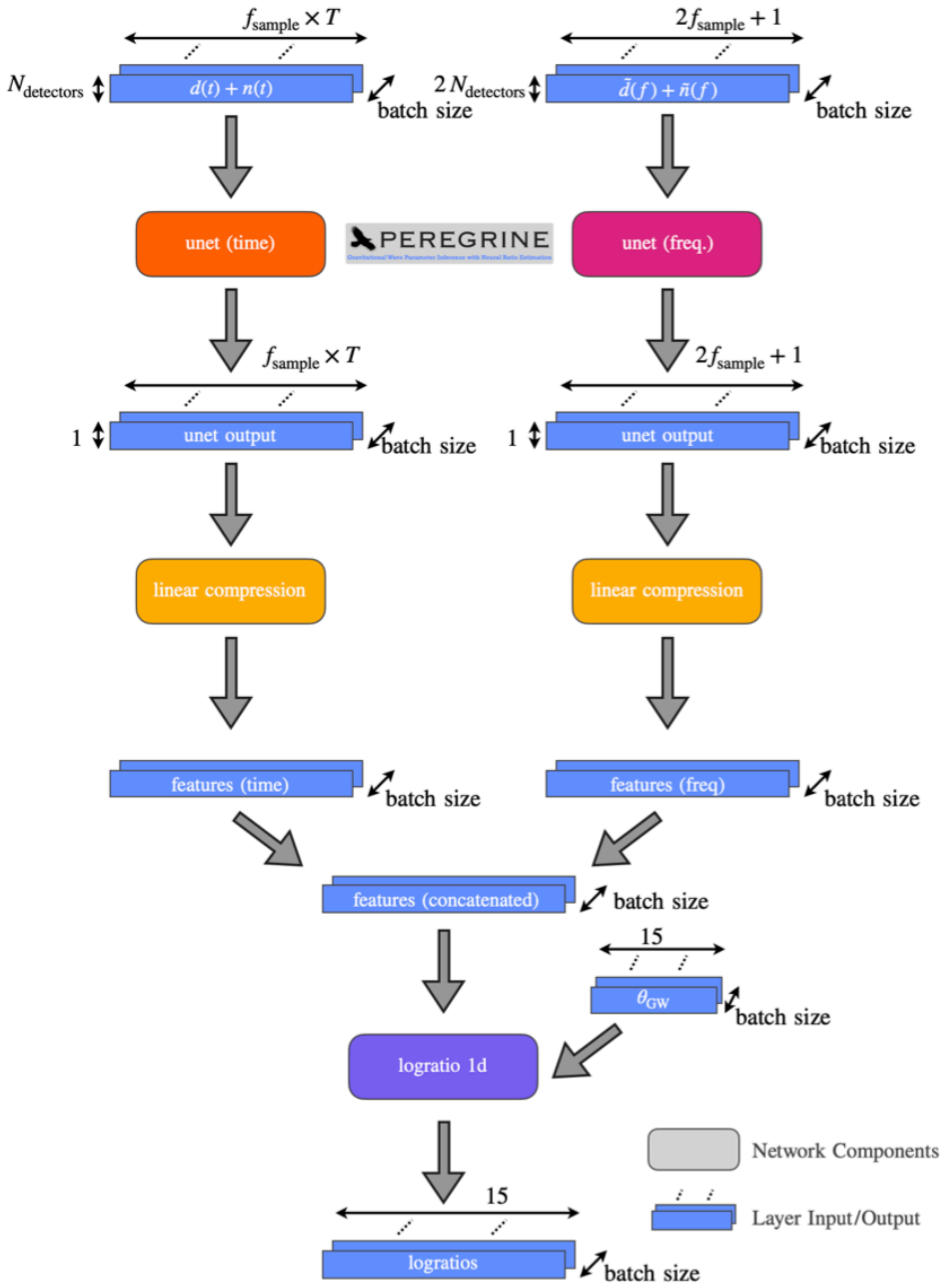


Summary

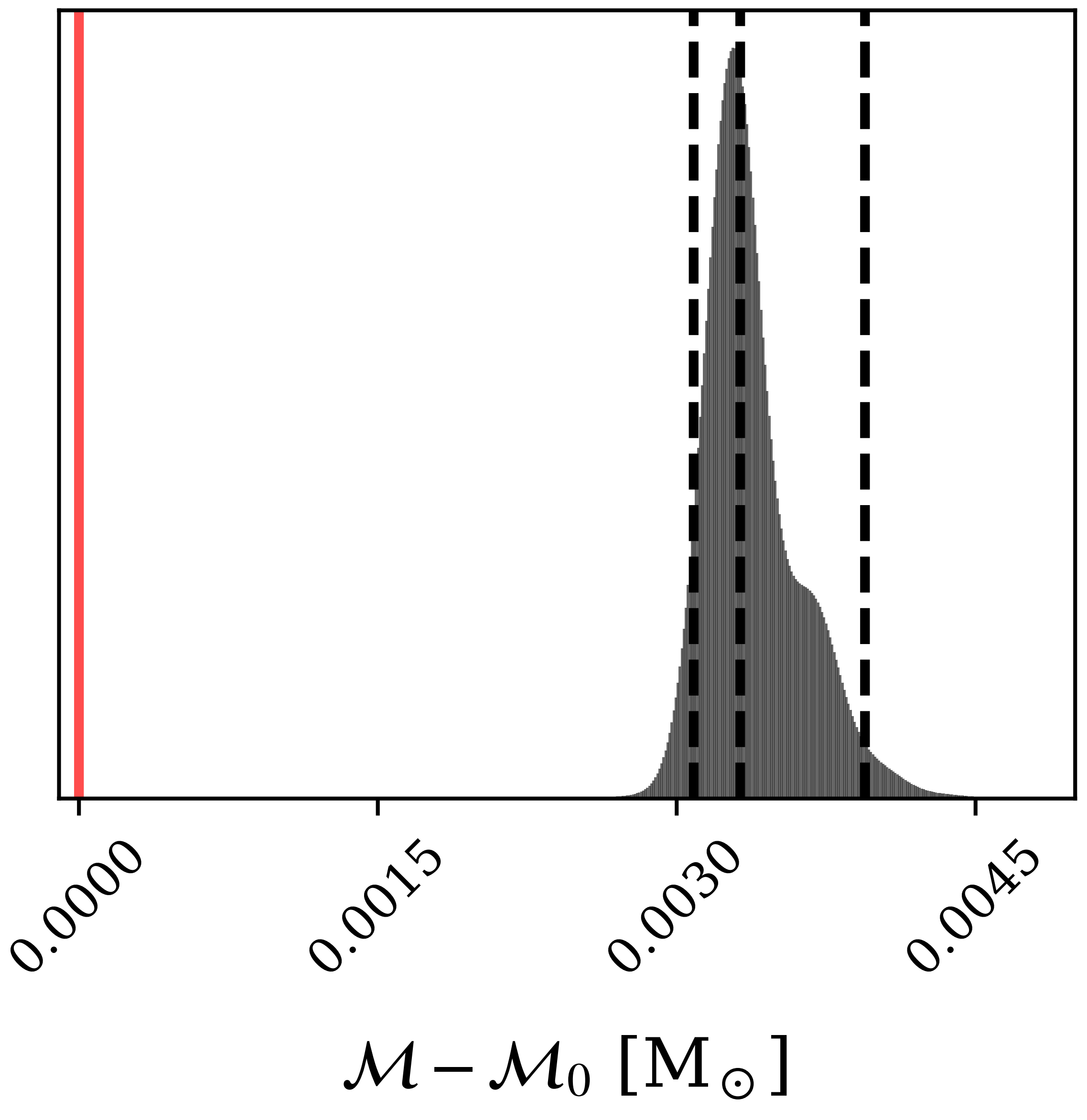
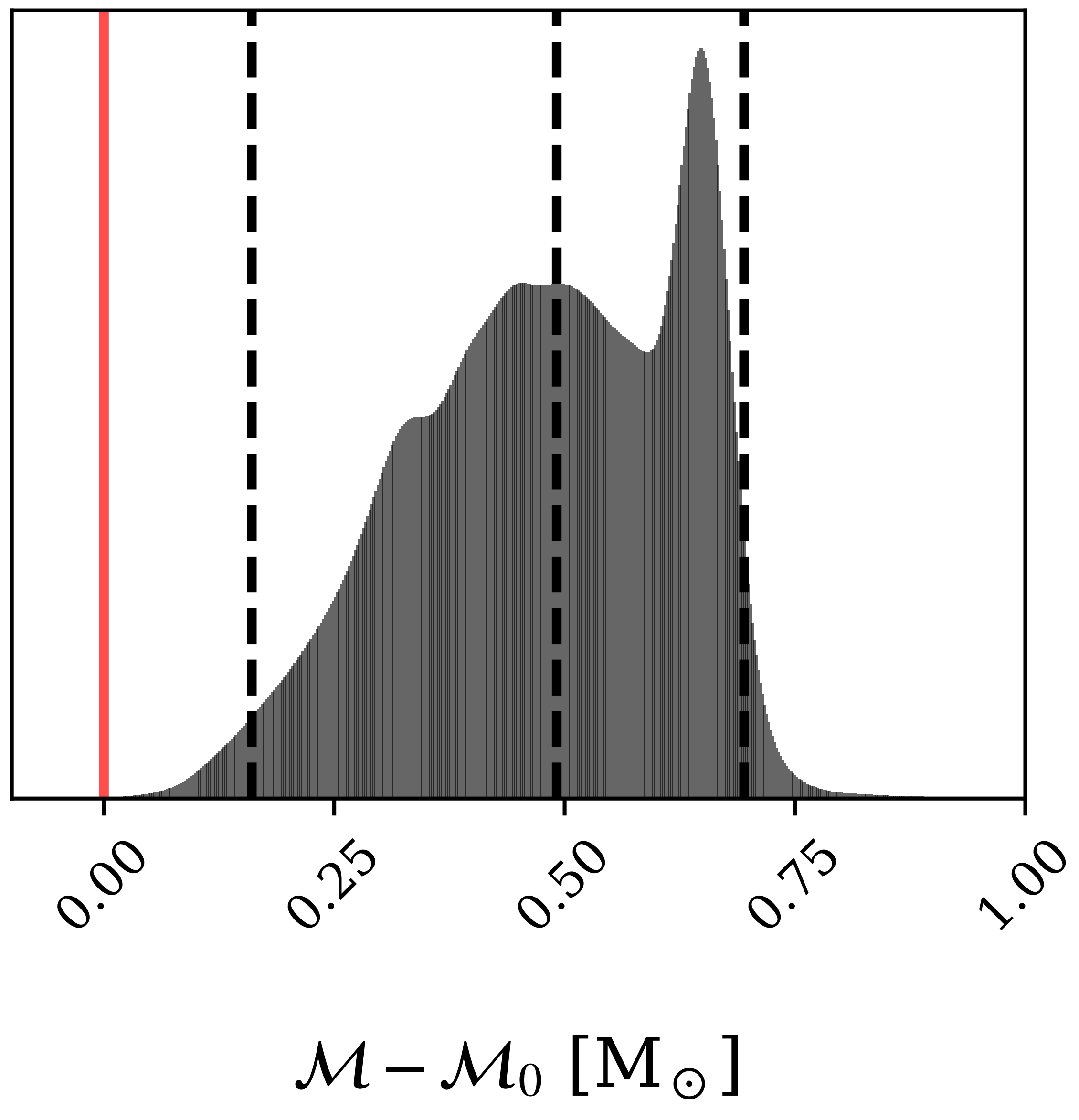
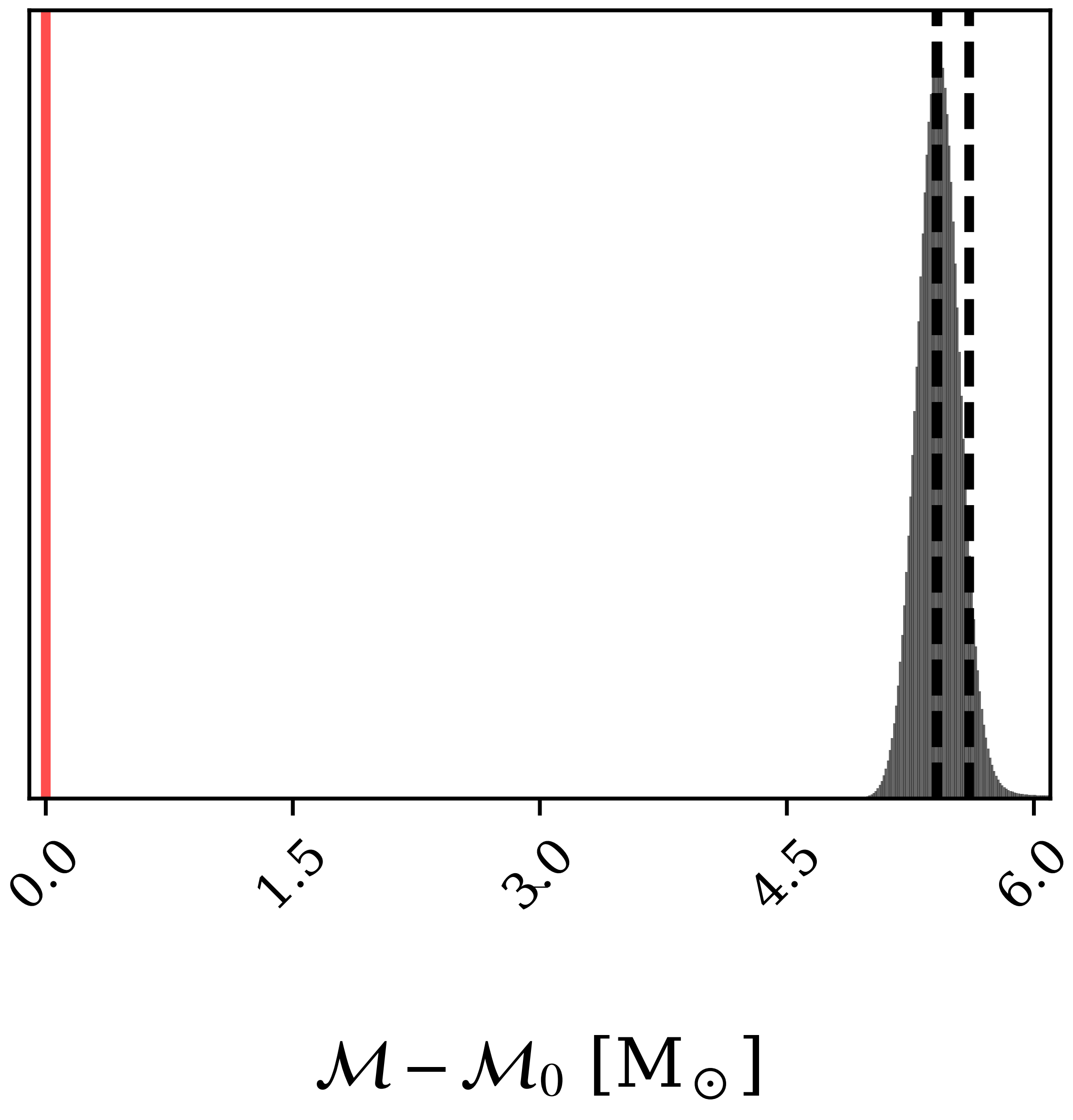
- We have an opportunity to learn about the nature of dark matter from long-duration gravitational waveforms
- These waveforms are complicated and the noise is too
- Likelihood-based methods for parameter estimation struggle in this regime
- Simulation-based inference methods may be able to help because they require no knowledge of a likelihood (and can be less computationally expensive)
- Current, ongoing research with lots of developments in the community everyday!

Network architecture

```
self.unet_f = Unet(
    n_in_channels=2,
    n_out_channels=1,
    sizes=(16, 32, 64, 128, 256),
    down_sampling=(4,2,2,2),
)
```



Another issue is calculating Bayes factors: Parameter estimation with using nested sampling - using the wrong model

	Accretion disk signal	Dark dress signal	Grav atom signal
Vacuum template			

Cole et al. 2023

See also Hannuksela et al. arXiv:1804.09659
Maselli et al. arXiv:2106.11325

Bayesian model comparison shows confident preference for correct model over any other environment - for this you need the evidence - which nested sampling estimates for free… but what about with SBI?

$\log_{10} \mathcal{B}$	**Dark dress signal**	**Accretion disk signal**	**Gravitational atom signal**
Vacuum template	34	6	39
Dark dress template	-	3	39
Accretion disk template	17	-	33
Gravitational atom template	24	6	-