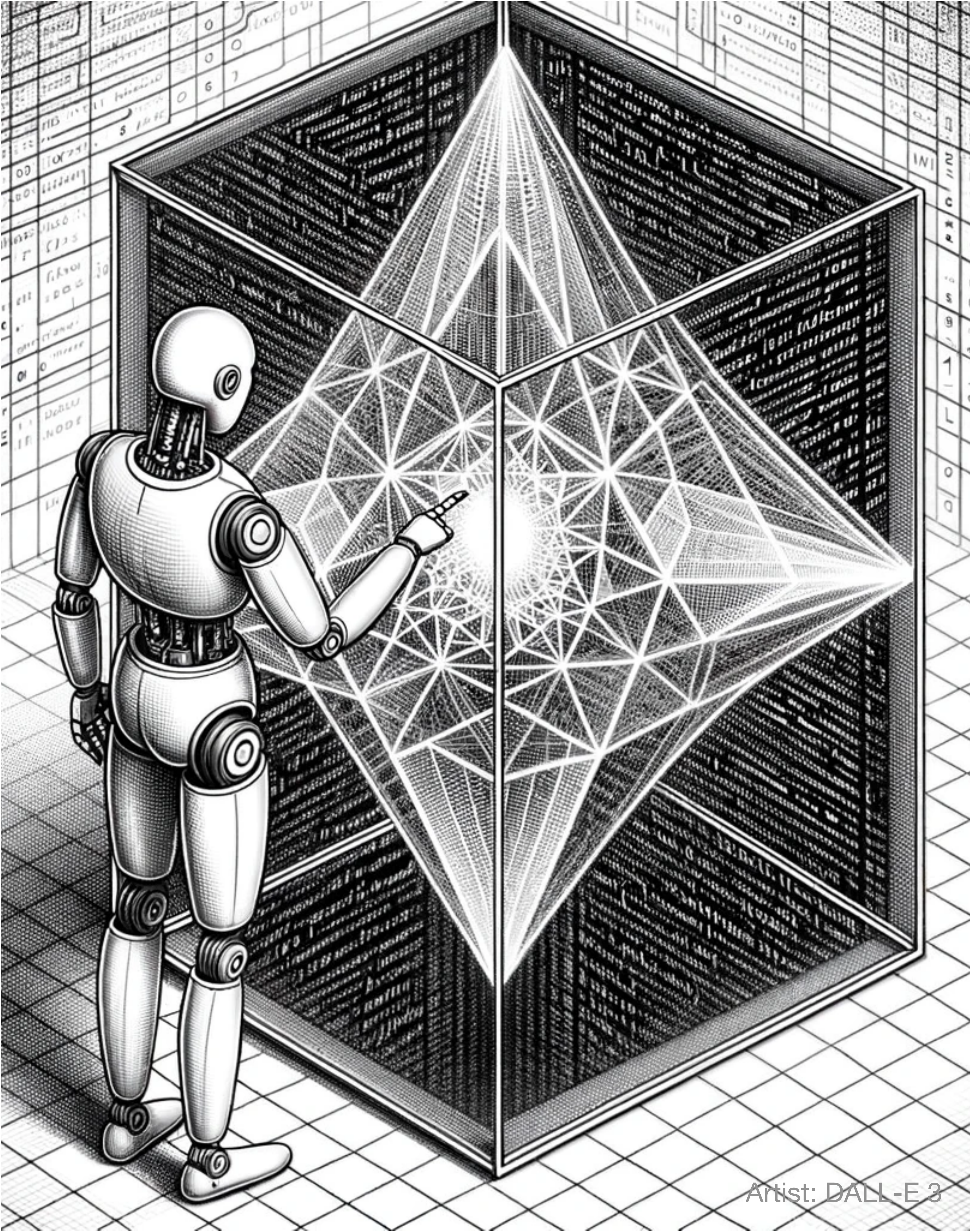


Artist: DALL-E 3

Large Physics Models through AI Prisms

Simulation-based inference for Stochastic Gravitational Wave Background Data Analysis

Christoph Weniger
GRAPPA, Institute of Physics, University of Amsterdam

netherlands eScience center

SURF

NWO

AI4Science Lab

erc

ERC CoG UnDark,
Proj. No. 864035

Belgian-Dutch Gravitational Wave Meeting 2023, Maastricht, 23 October 2023

GRAPPA
GRavitation AstroParticle Physics Amsterdam

UNIVERSITY OF AMSTERDAM

Likelihood-based inference via Sampling

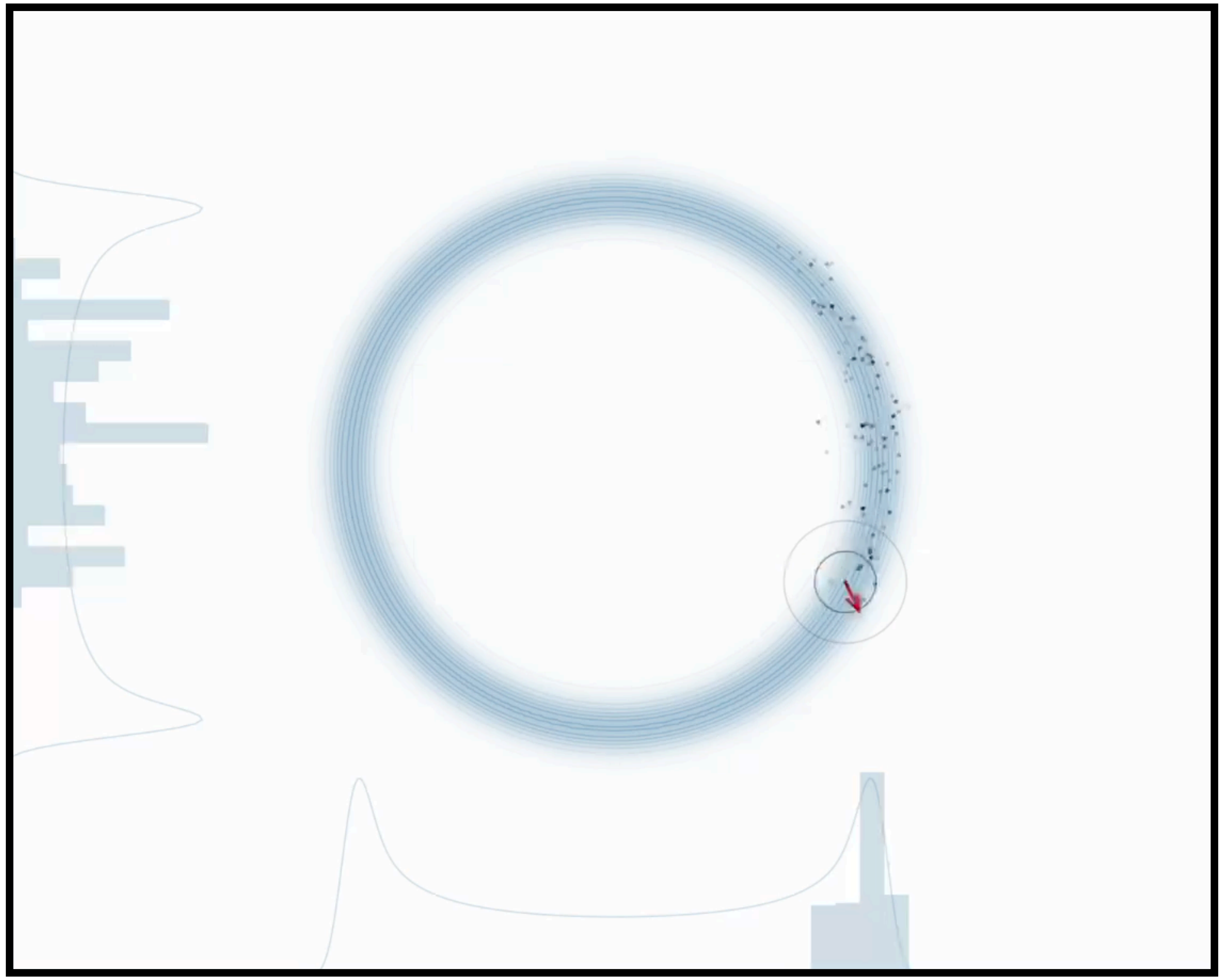
Metropolis Hastings, Nested sampling, …

Likelihood Prior

Parameters Data

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

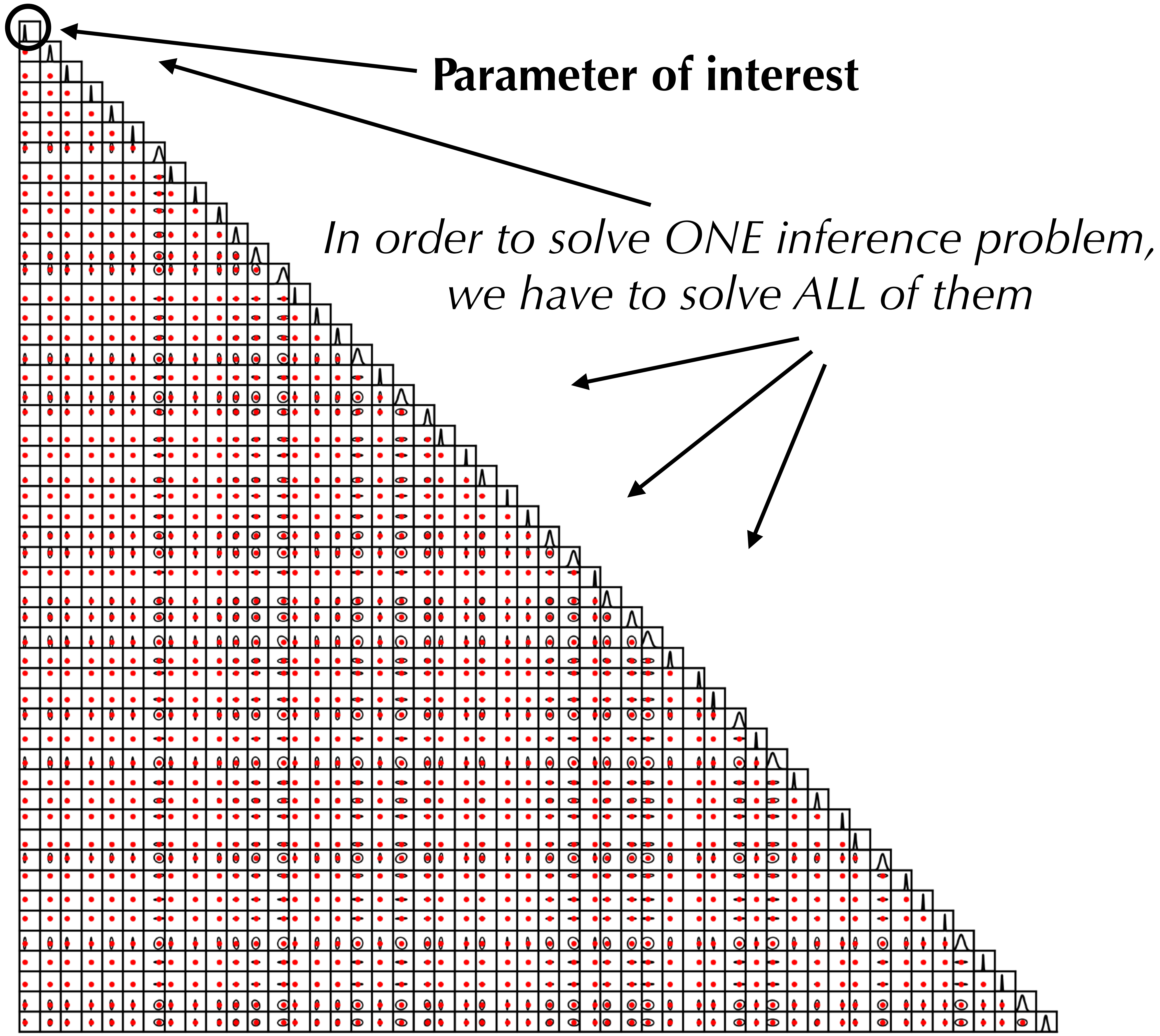
Posterior

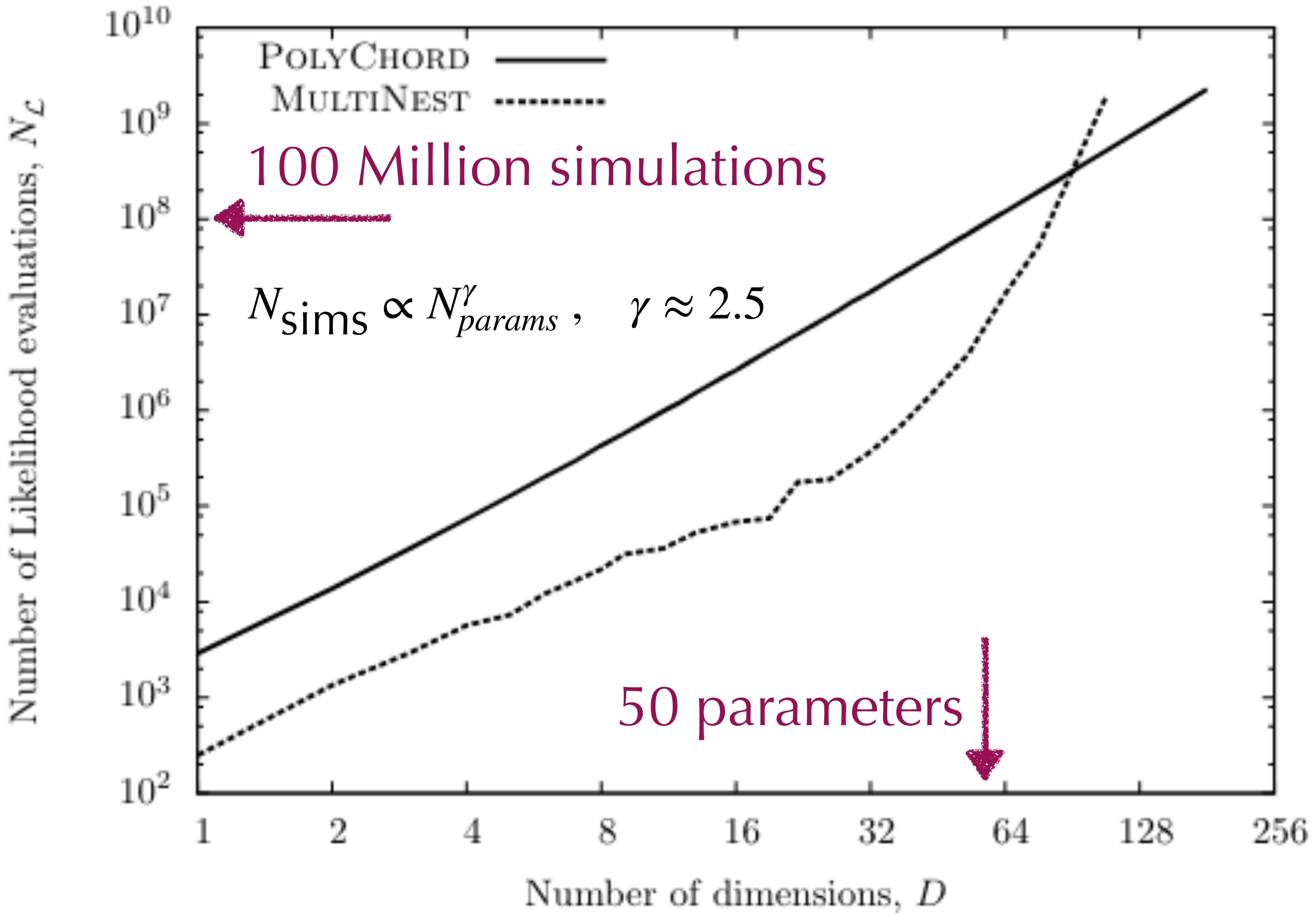


https://chi-feng.github.io/mcmc-demo/app.html?algorithm=RandomWalkMH

Speagle 1909.12313 (MCMC review)

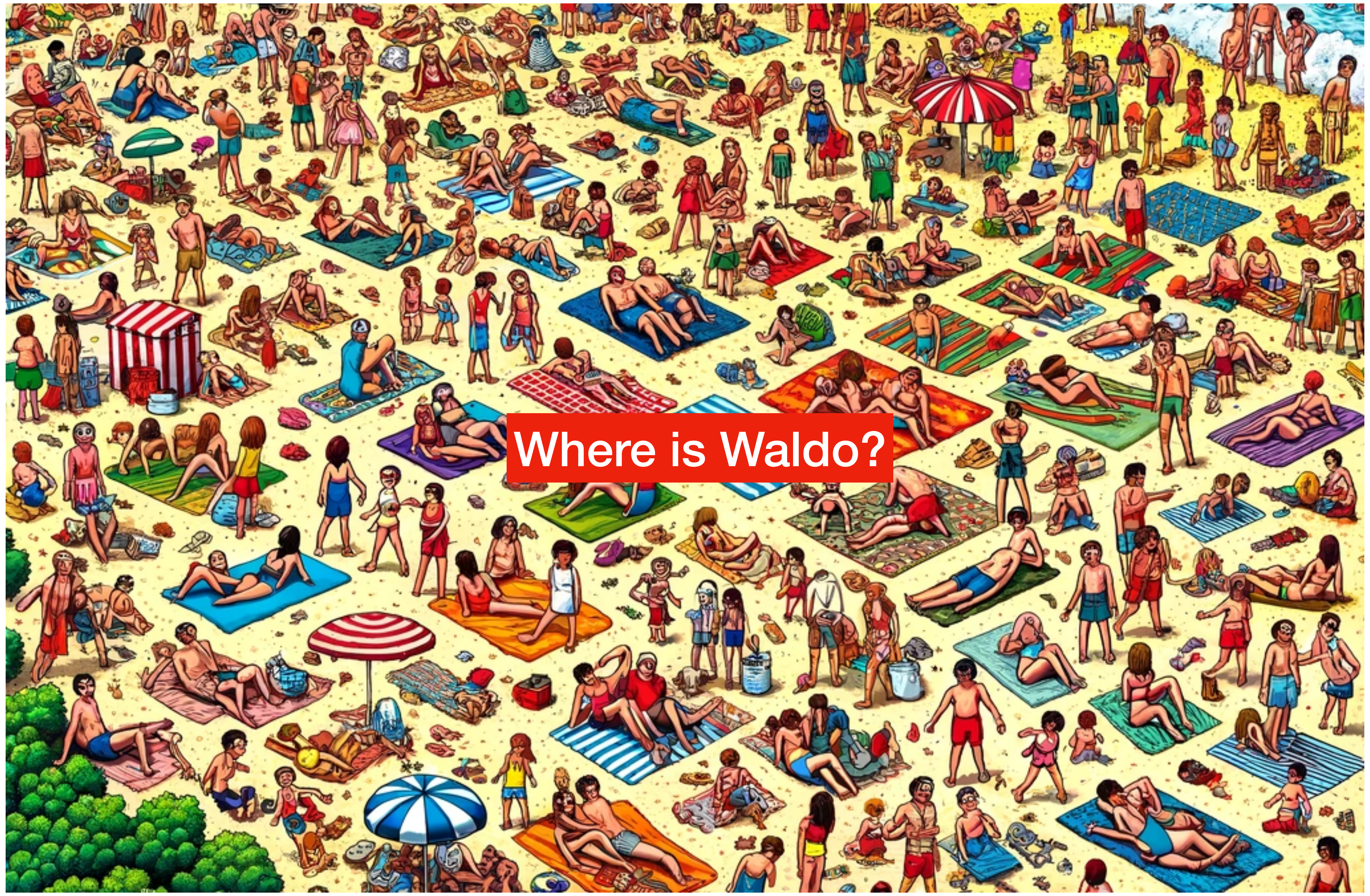
The curse of dimensionality





Feroz+ 0809.3437 (MultiNest)
Handley+ 1502.01856 (PolyChord)

The curse of dimensionality



$$q_\phi(\mathbf{z}_{\text{Waldo}} \mid \mathbf{x}_o) = \int d\mathbf{z}_{\text{Lucia}} d\mathbf{z}_{\text{Oleg}} d\mathbf{z}_{\text{Sibilla}} d\mathbf{z}_{\text{Dion}} \cdots d\mathbf{z}_{\text{Noemi}}\ q_\phi(\mathbf{z}_{\text{Waldo}}, \mathbf{z}_{\text{Lucia}}, \mathbf{z}_{\text{Oleg}}, \mathbf{z}_{\text{Sibilla}}, \mathbf{z}_{\text{Dion}}, \cdots, \mathbf{z}_{\text{Noemi}} \mid \mathbf{x}_o)$$

Same "physics" simulator, same result

Simulation-based inference
Train a neural network to find Waldo's marginal posterior.

Joined inference
Run MCMC to explore ultra-high-dimensional model for every single aspect in the image. Then marginalise.

Simulation-based inference (SBI)

Classic analog: Approximate Bayesian Computation*

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

Forward (Simulation)

Simulated data

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

Backward (Inference)

Amortised!

After generating simulation data, the posteriors for any observation can be generated instantaneously. Without re-simulation.

Data summary

$\mathbf{x} \rightarrow t(\mathbf{x})$

Simulate & compress

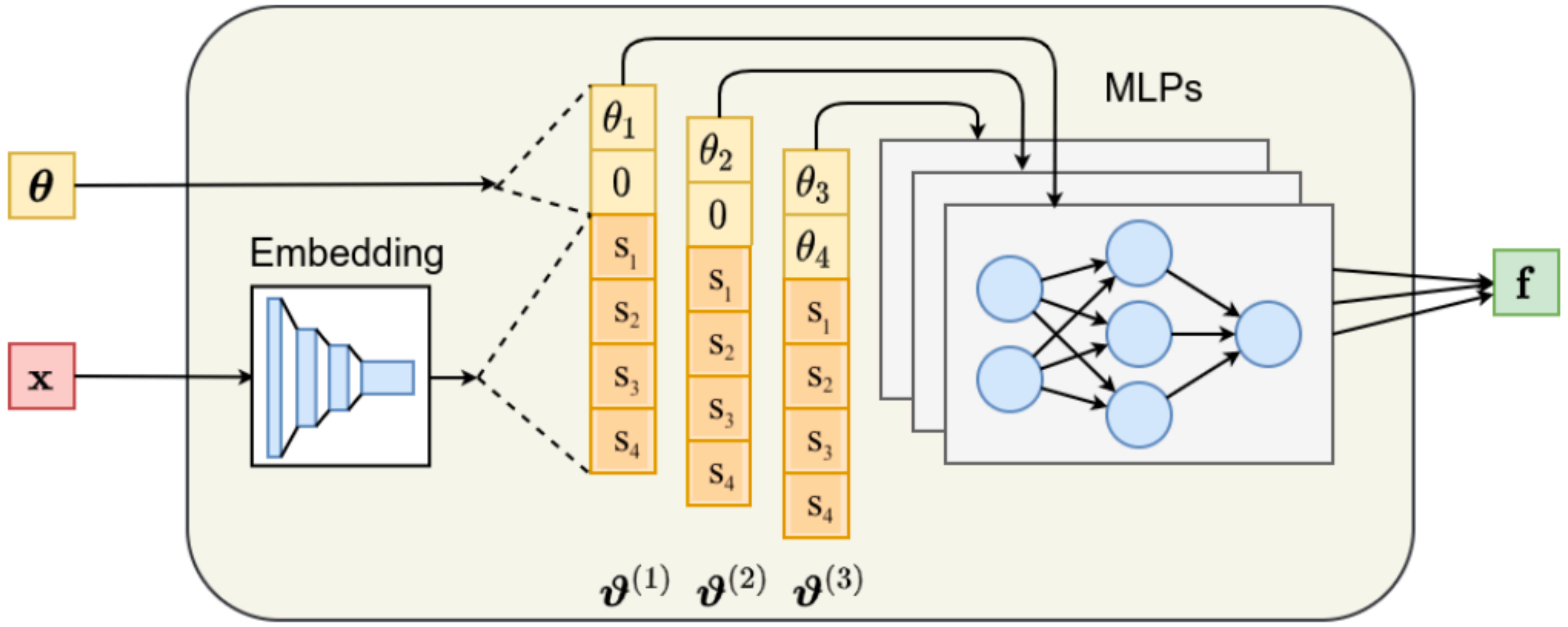
*this analogy might not exactly apply to all SBI algorithms but serves the purpose of providing some basic intuition for what is going on

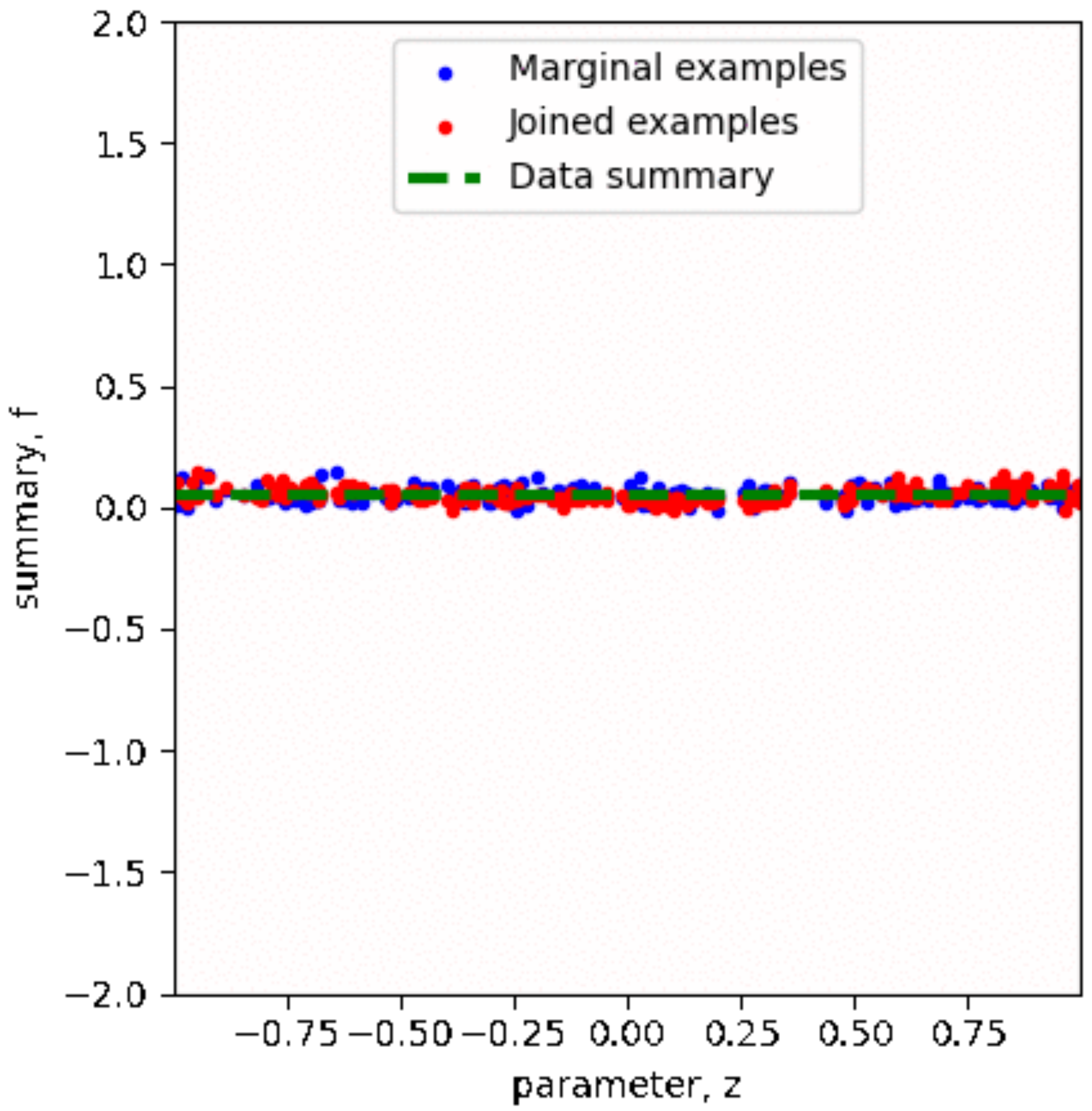
Reviews:
Marin+ 1101.0955
Sisson+ 1802.09720

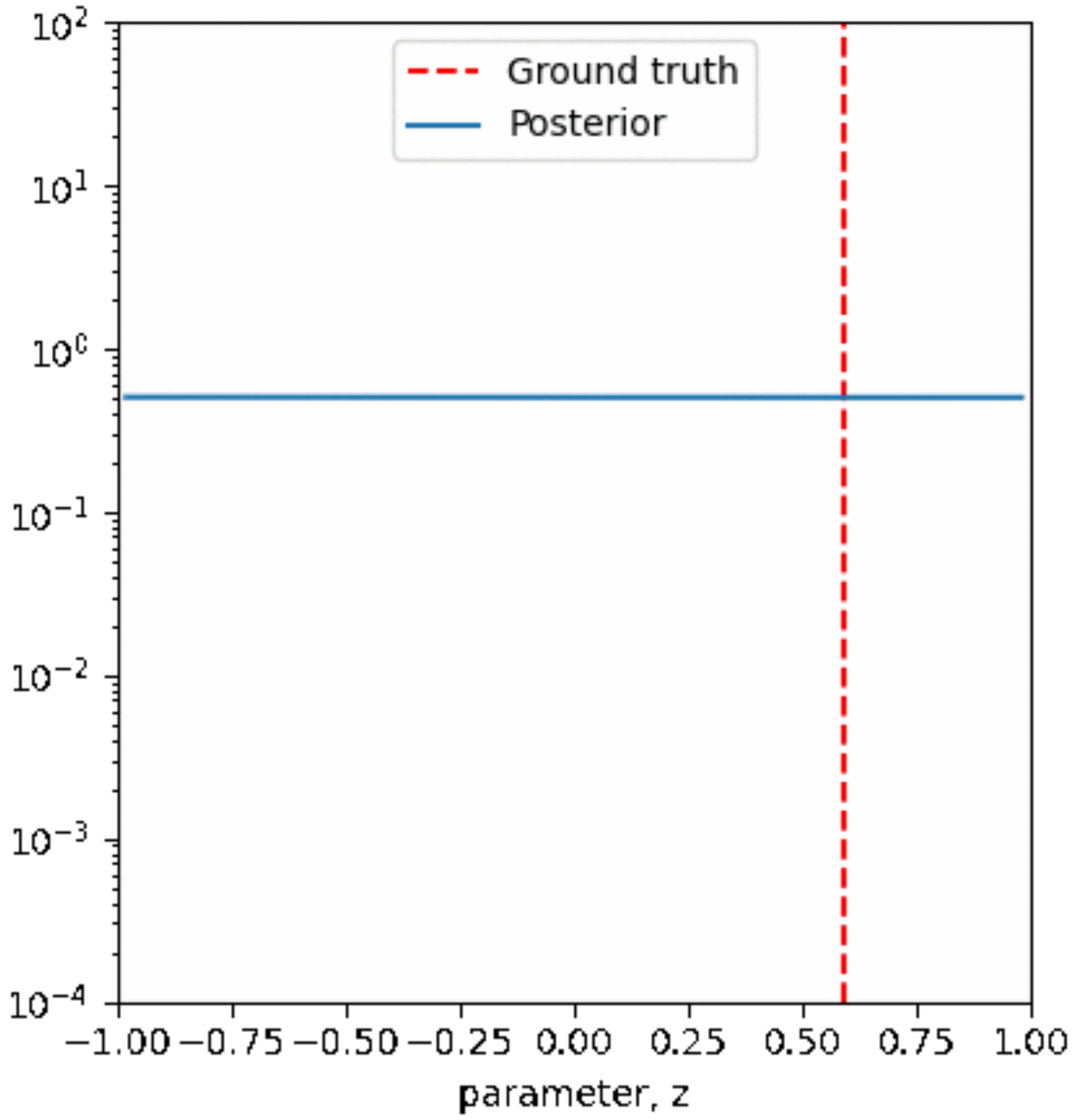
SBI with "neural ratio estimation"

Neural ratio estimation (NRE)
Train a neural network to discriminate

- Real sims: $z, \mathbf{x} \sim p(\mathbf{x} \mid z) p(z)$
- Scrambled sims: $z, \mathbf{x} \sim p(\mathbf{x}) p(z)$



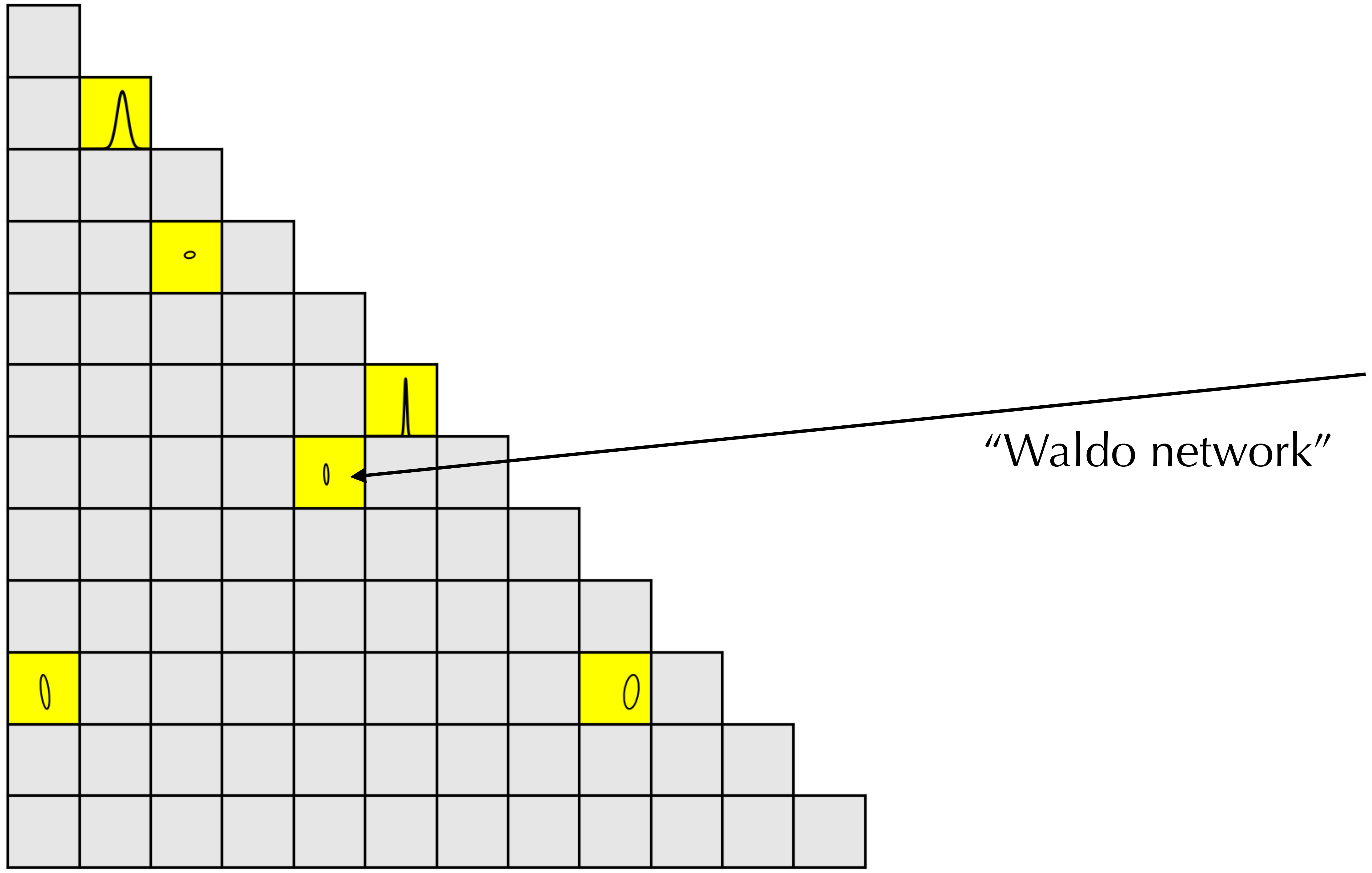




$$\ln r(\mathbf{x}, z) \equiv \mathrm{M}_\phi(\mathbf{s}, \mathbf{z}) \simeq \ln \frac{p(\mathbf{x}, \mathbf{z})}{p(\mathbf{x}) p(\mathbf{z})} = \ln \frac{p(\mathbf{z} \mid \mathbf{x})}{p(\mathbf{z})} = \ln \frac{p(\mathbf{x} \mid \mathbf{z})}{p(\mathbf{x})}$$

Miller+ 2011.13951, 2107.01214 - swyft & TMNRE

Take-away: SBI enables focused inference



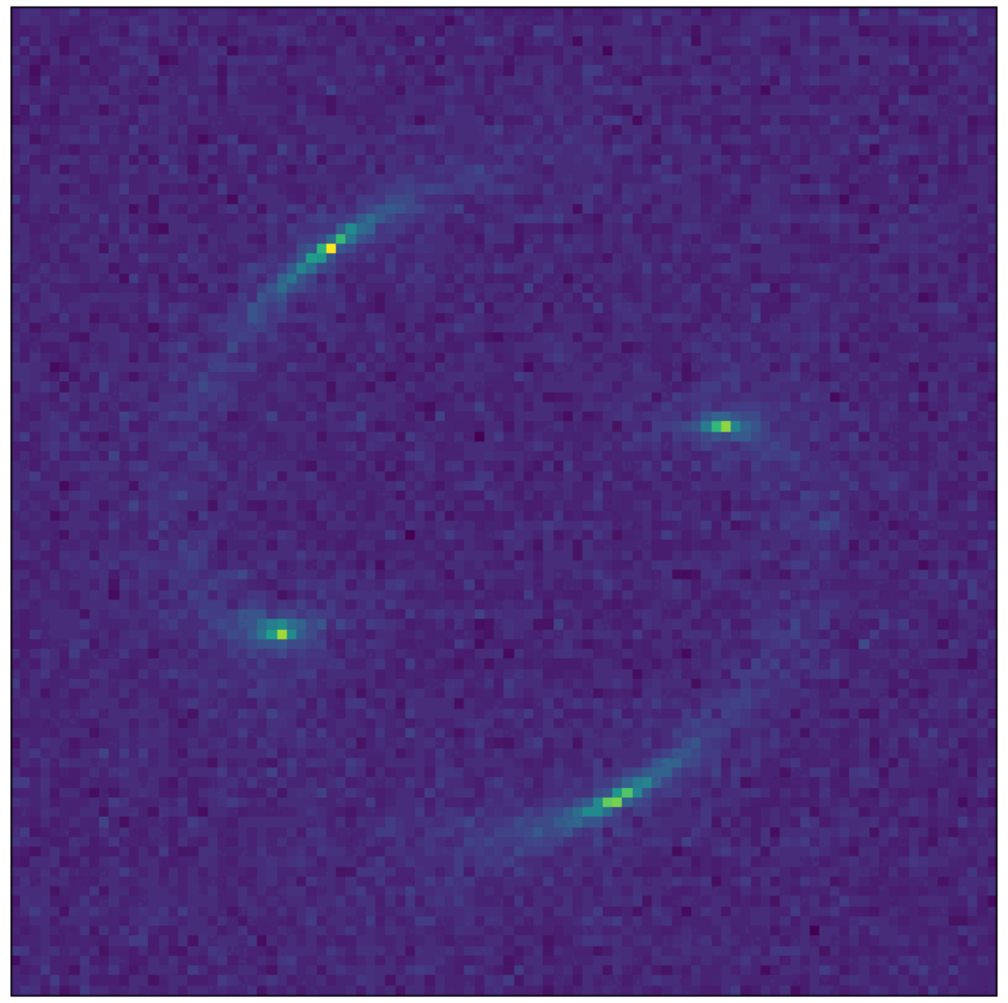
It turns out that solving 30 one-dim problems can be much easier than solving one 30-dim problem.

Sequential inference

Gaining precision through targeted simulations

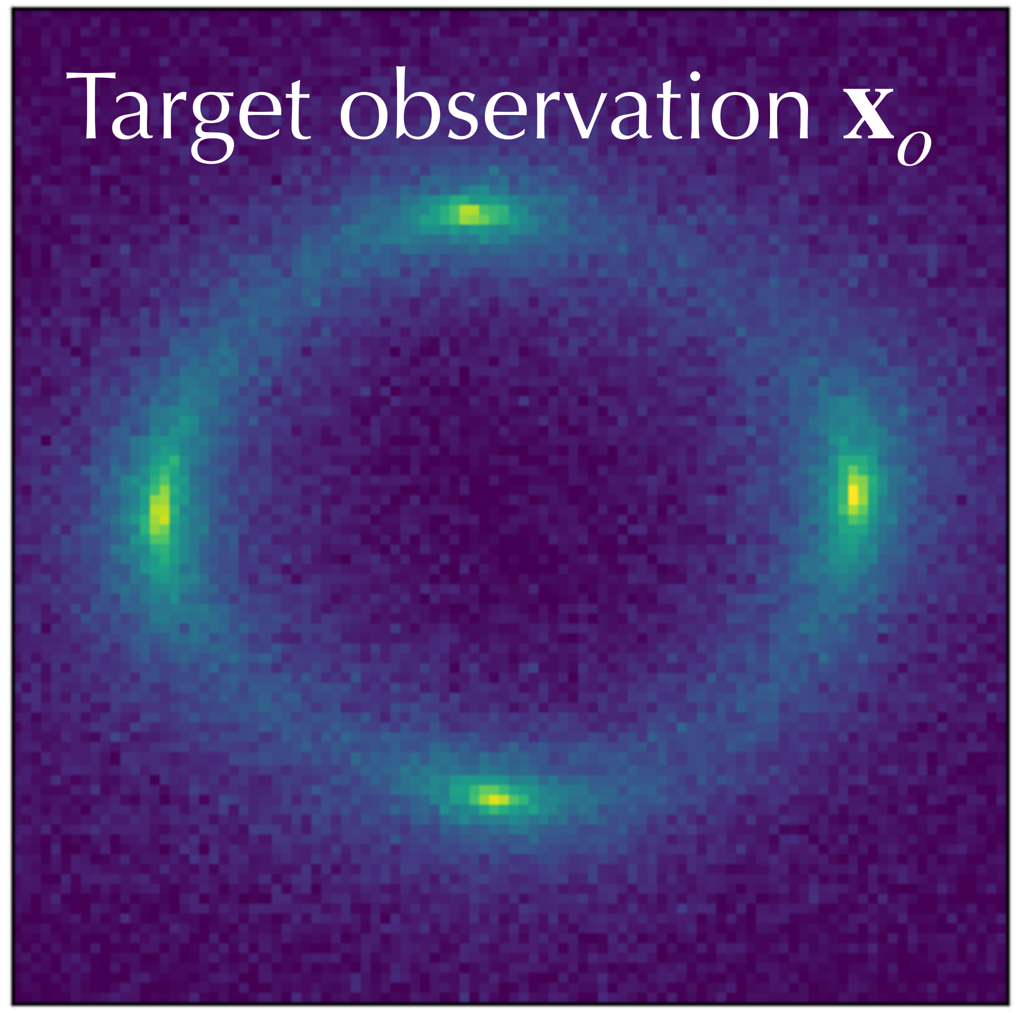
Samples from full prior:

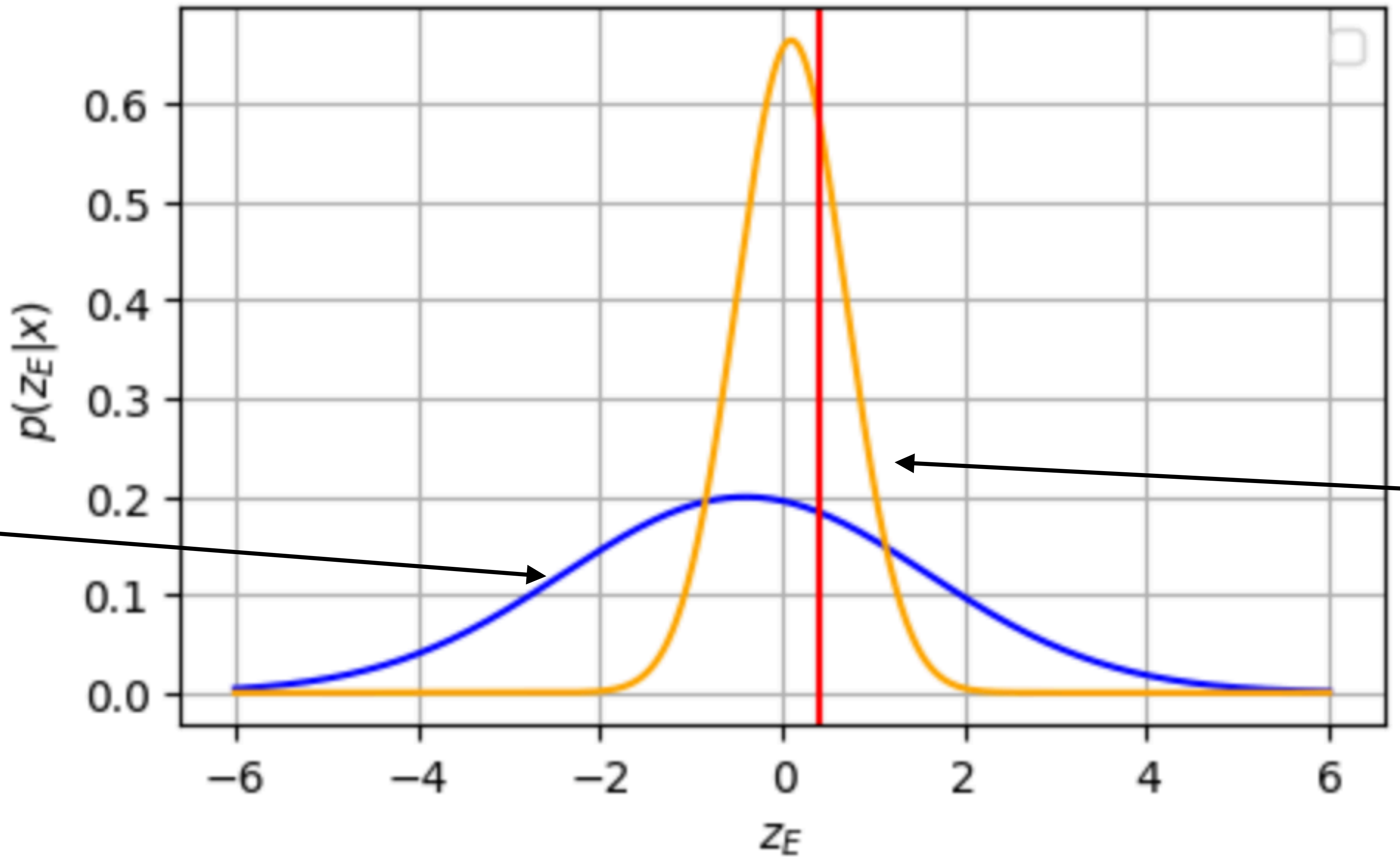
$\mathbf{x}, \mathbf{z} \sim p(\mathbf{x} \,|\, \mathbf{z}) p(\mathbf{z})$



NRE training

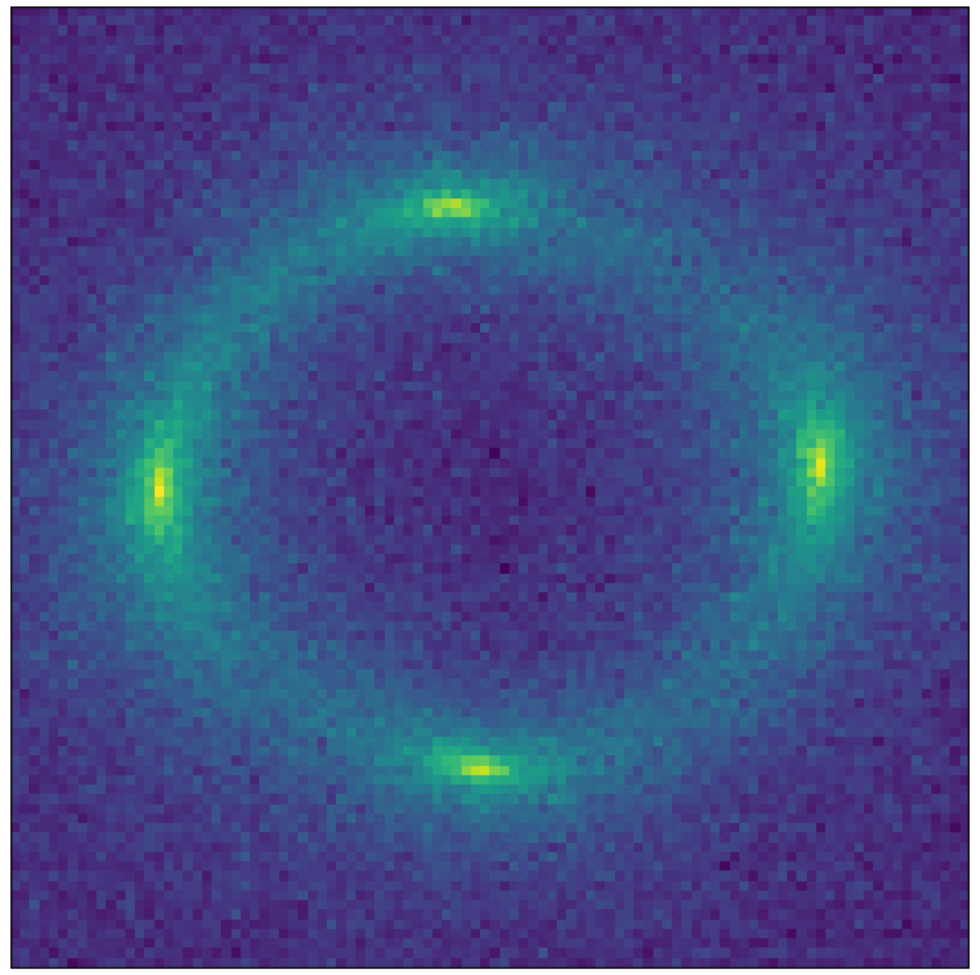
$r(\mathbf{x}; z_E)$





Samples from some constrained prior:

$\mathbf{x}, \mathbf{z} \sim p(\mathbf{x} \,|\, \mathbf{z}) \tilde{p}_{\mathbf{x}_o}(\mathbf{z})$



NRE training

$r(\mathbf{x}; z_E)$

Durkan+ 2002.03712 for a discussion

Image credit: Noemi Anau Montel

Marginal sequential inference

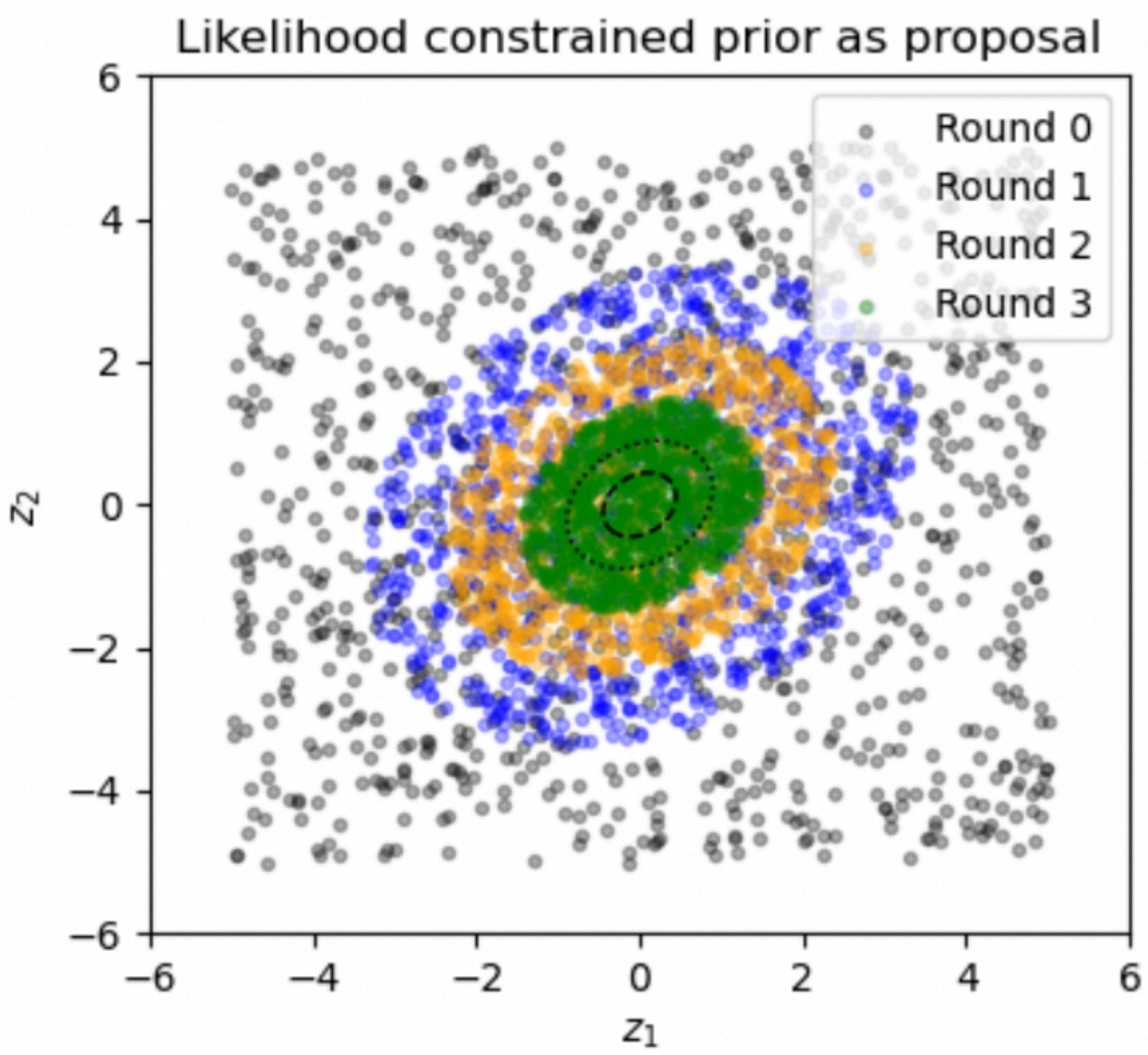
Key idea: Use a truncated version of the prior as proposal function.

$$\tilde{p}^{(R)}(\mathbf{z}) = \frac{1}{Z}\mathbb{I}(\mathbf{z} \in \Gamma^{(R-1)})p(\mathbf{z})$$

Miller+ 2011.13951, 2107.01214 - swyft & TMNRE

We use a hard likelihood constrained prior truncation scheme, excluding low likelihood regions estimated in previous rounds.

$$\Gamma^{(R)} = \{\mathbf{z} \in \mathbb{R}^N : \tilde{r}^{(R)}(\mathbf{x};\mathbf{z}) > \epsilon\} \qquad \tilde{r}^{(R)}(\mathbf{x};\mathbf{z}) \simeq \frac{p(\mathbf{x}\,|\,\mathbf{z})}{p(\mathbf{x})}$$



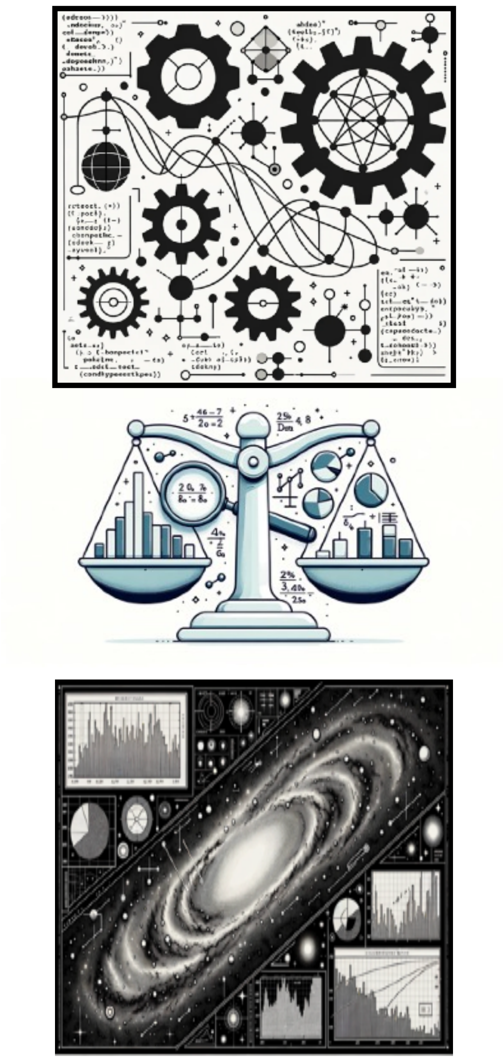
Doing this leaves the learned ratio unaffected, and marginal estimation becomes possible

$$q_\phi^{(R)}\left(z_1 \,|\, \mathbf{x}\right) \simeq \int dz_2 \ldots dz_N \; p(\mathbf{x} \,|\, \mathbf{z})\, \frac{1}{Z}\mathbb{I}(\mathbf{z} \in \Gamma^{(R-1)})p(z_2, \ldots, z_N) = \int dz_2 \ldots dz_N \; p(\mathbf{x} \,|\, \mathbf{z})p(z_2, \ldots, z_N) + \mathcal{O}(\epsilon)$$

See also Greenberg+ 1905.07488, Durkan+ 2002.03712 for sequential methods
Applied to NPE in Deistler+ 2210.04815

Large Physics Models through AI Prisms

Model & Data

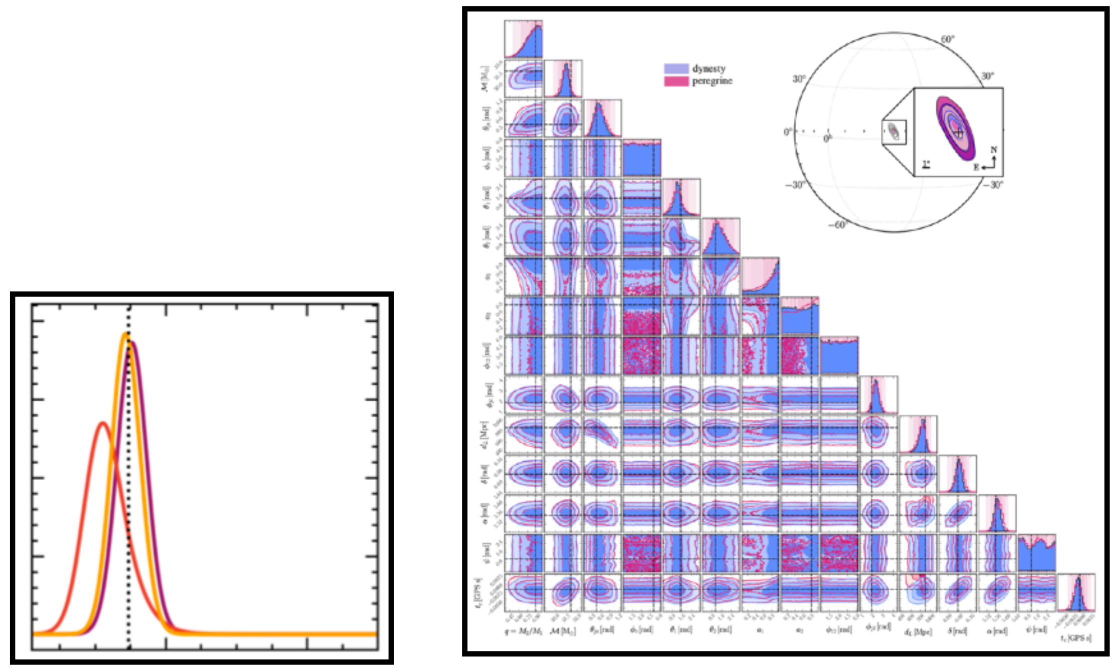


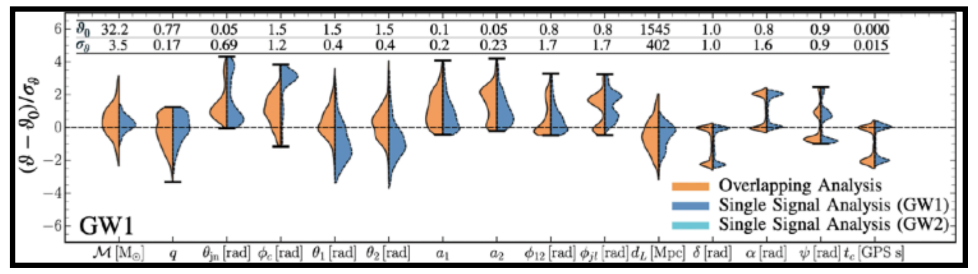
AI-based comparison

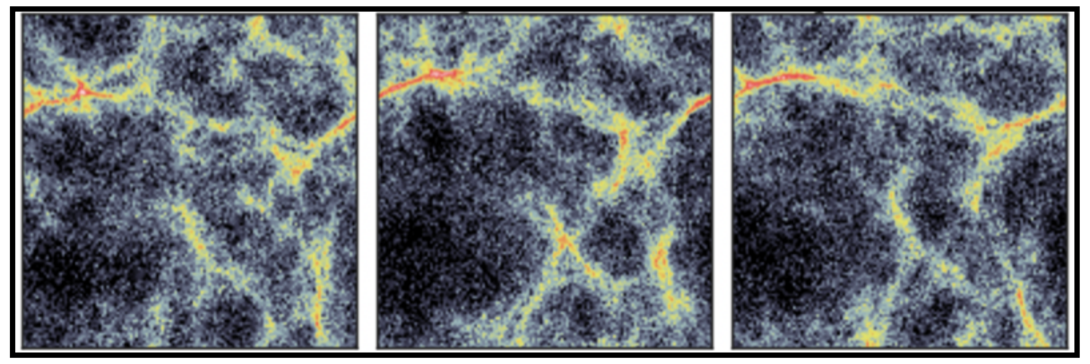
Inference

Refinement

Focused statistical insights







Ph.D. students & Postdocs

Noemi Anau Montel
(GRAPPA, UvA)

Kosio Karchev
(SISSA)

Florian List
(Vienna University)

Uddipta Bhardwaj
(GRAPPA, UvA)

Oleg Savchenko
(GRAPPA, UvA)

James Alvey
(GRAPPA, UvA)

Mathis Gerdes
(GRAPPA, UvA)

Anchal Saxena
(University of Groningen)

Guillermo Franco Abellan
(GRAPPA, UvA)

Ben Miller
(GRAPPA & AI4Science, UvA)

Applications: Strong lensing images, stellar streams, gravitational waves, 21cm cosmology, Planck cosmology, large scale structure data, point source populations, Ia SN cosmology, …

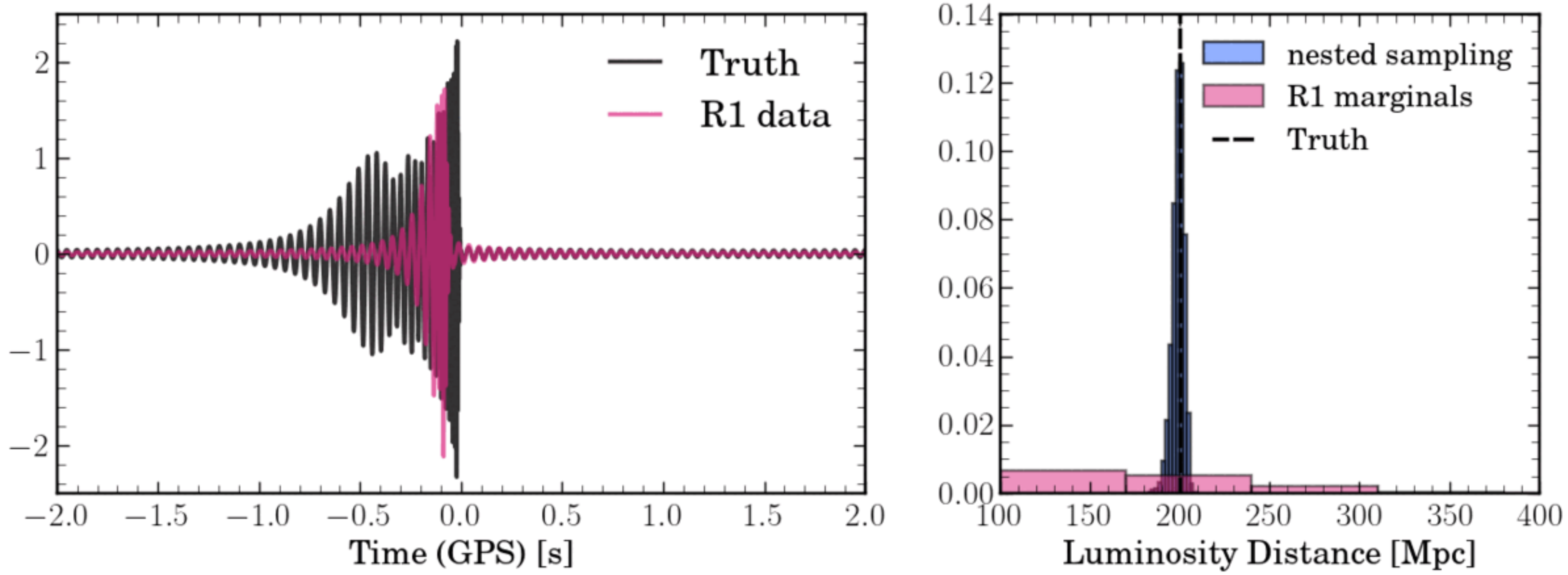
First applications to Gravitational waves

James Alvey
(GRAPPA, UvA)

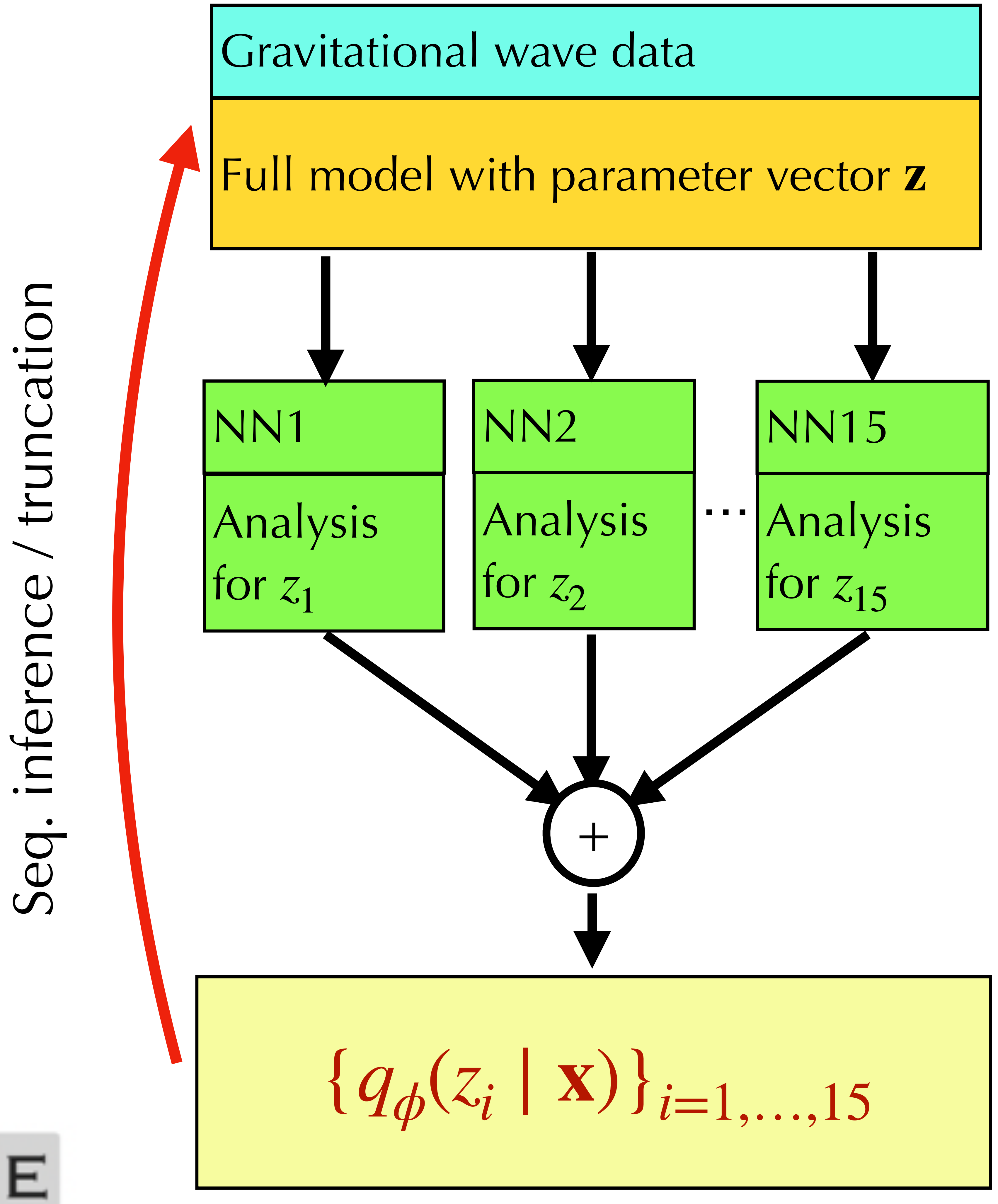
Uddipta Bhardwaj
(GRAPPA, UvA)

Gravitational wave parameter inference

Single signal



- Case study of binary-black hole mergers (200 Mpc, 900 Mpc distance)
- Waveform model: SEOBNRv4PHM (15 parameters)
- LIGO / Virgo detectors
- Inference of 15 1-dim posteriors in 8 "focus rounds"



PEREGRINE
Gravitational Wave Parameter Inference with Neural Ratio Estimation

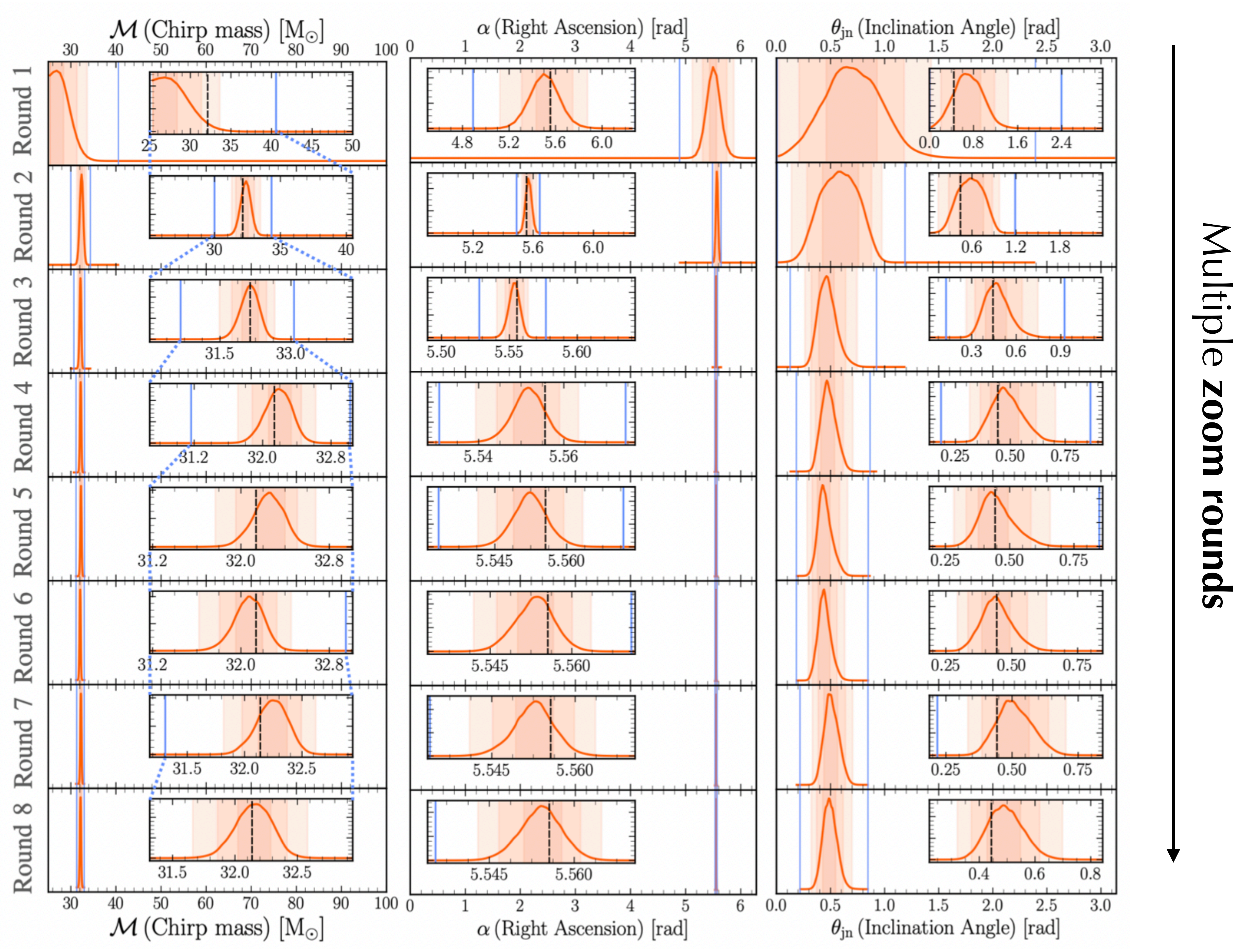
Bhardwaj+ 2304.02035

Related work: Dax+ 2106.12594

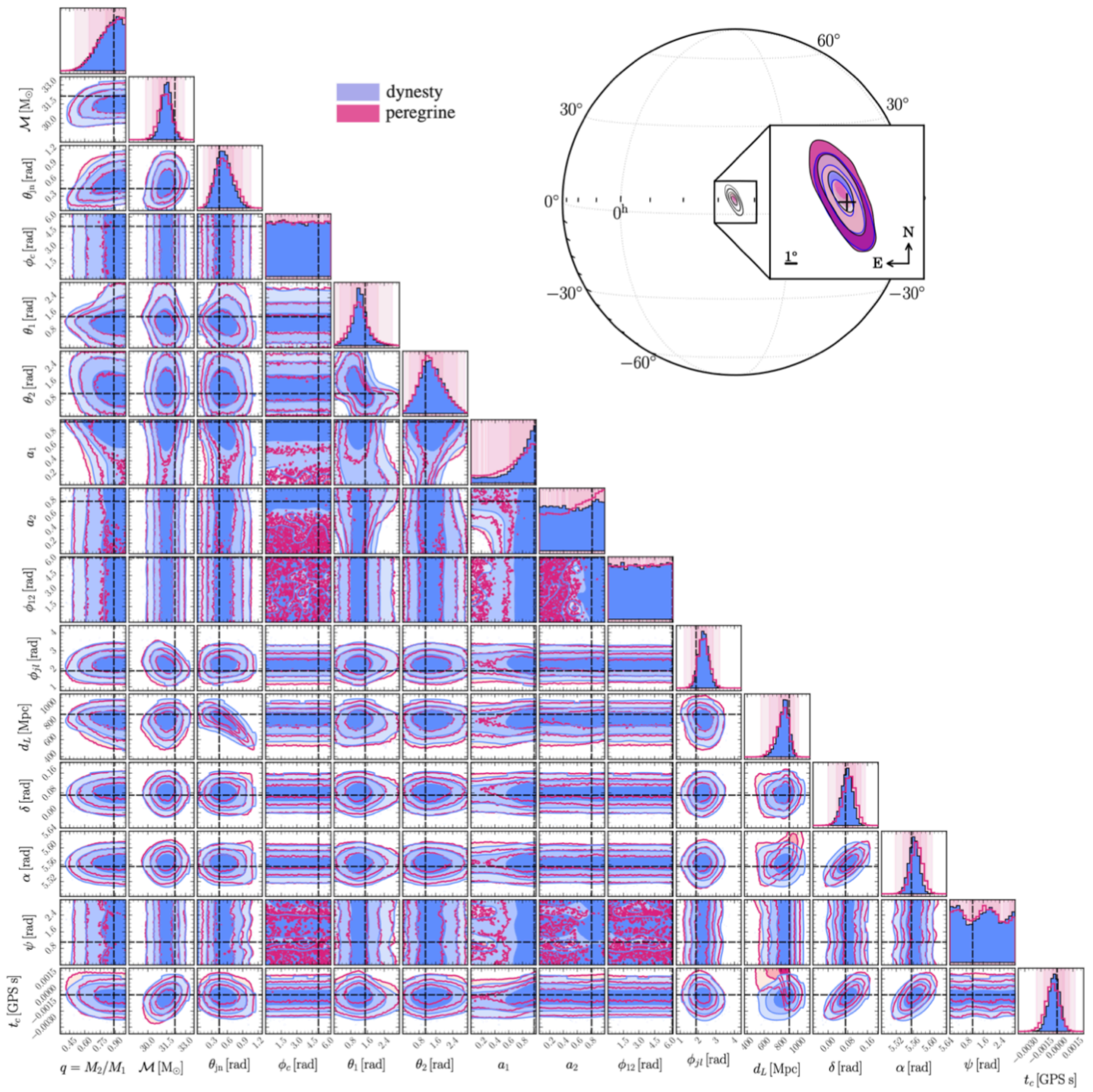
Gravitational wave parameter inference

Single signal

Step 1: Estimation of 1-dim posteriors (15 1-dim posteriors in 8 rounds)



Step 2: Estimation of 2-dim posteriors



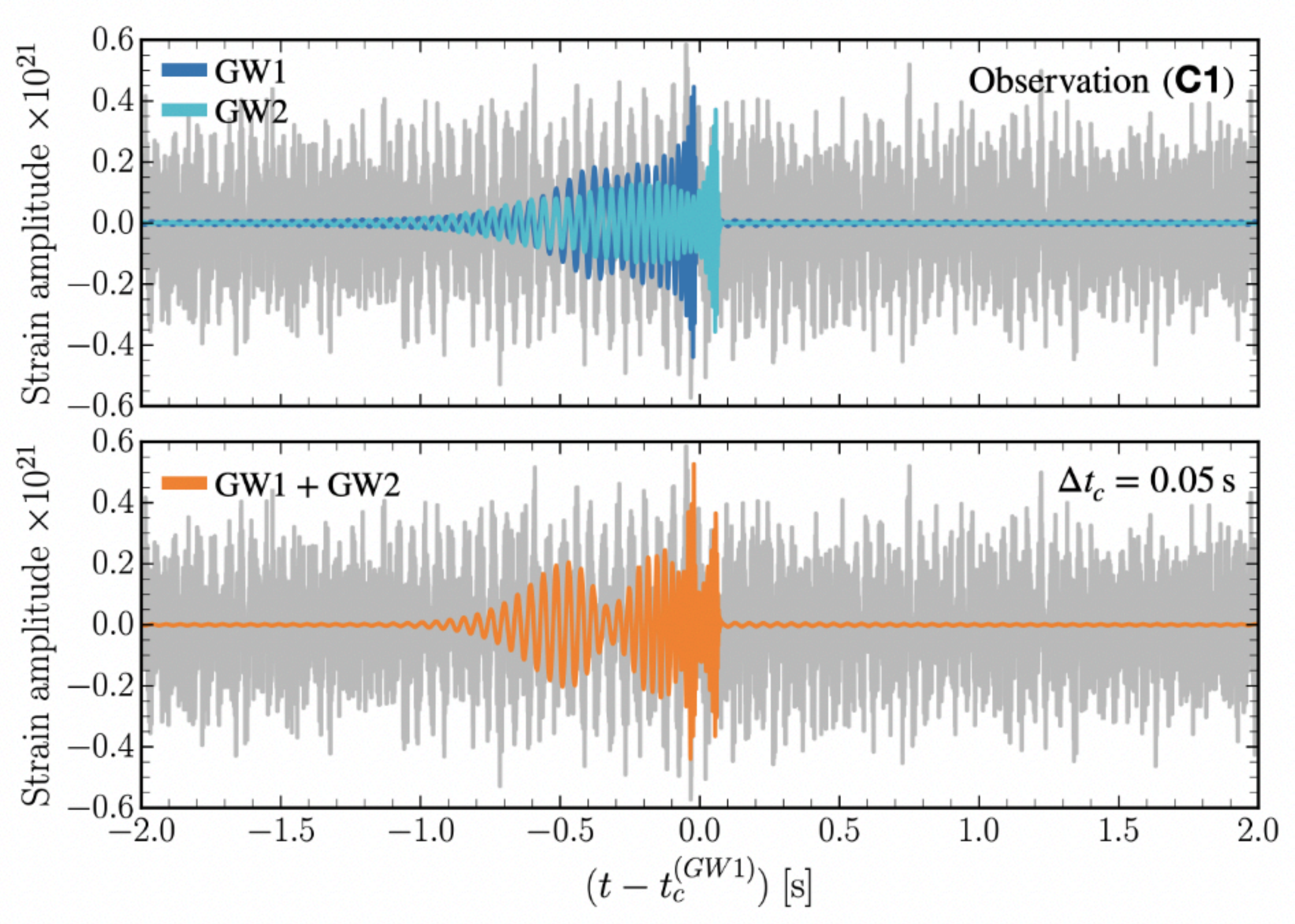
Estimation = $N(N-1)/2 = 105$ marginal 2-dim posteriors

- Essentially the same results as, e.g., dynesty
- 2% of simulation-runs compared to nested sampling

Bhardwaj+ 2304.02035

Gravitational wave parameter inference

Overlapping GW signals



- 2 x IMRPhenomXPHM
- 36 hours (instead of >20 days)
- Much faster than MCMC
- Higher precision than previous SBI attempts
- Precision only mildly degraded w.r.t. fits in absence of a second signal

Parameter	Prior Choice	Injection (GW1)	Injection (GW2)
Chirp mass $\mathcal{M}$ [$M_\odot$]	U(20, 100)	32.2	26.8
Mass ratio, q	U(0.125, 1)	0.77	0.81
Inclination angle θ_{jn} [rad]	sine(0, π)	0.05	1.8
Polarisation angle ψ [rad]	U(0, π)	0.9	0.4
Phase ϕ_c [rad]	U(0, 2π)	1.5	0.6
Tilt angles θ_1, θ_2 [rad]	sine(0, π)	1.5, 1.5	0.9, 1.2
Dimensionless spins a_1, a_2	U(0, 1)	0.1, 0.05	0.7, 0.3
Spin angles ϕ_{12}, ϕ_{jl} [rad]	U(0, 2π)	0.8, 0.8	0.6, 0.9
Right ascension α [rad]	U(0, 2π)	0.8	4.0
Declination δ [rad]	cosine($-\pi/2, \pi/2$)	1.0	-0.5
Merger time t_c [GPS s]	U(-0.1, 0.1)	0.0	
Merger time difference Δt_c [GPS s]	U(0, 1)		0.05(**C1**), 0.2(**C2**), 0.5(**C3**)
Luminosity Distance d_L [Mpc]	$U_{vol.}(100, 2500)^{\star}$	1545	636.5

(See also talk by Tomek Baka)

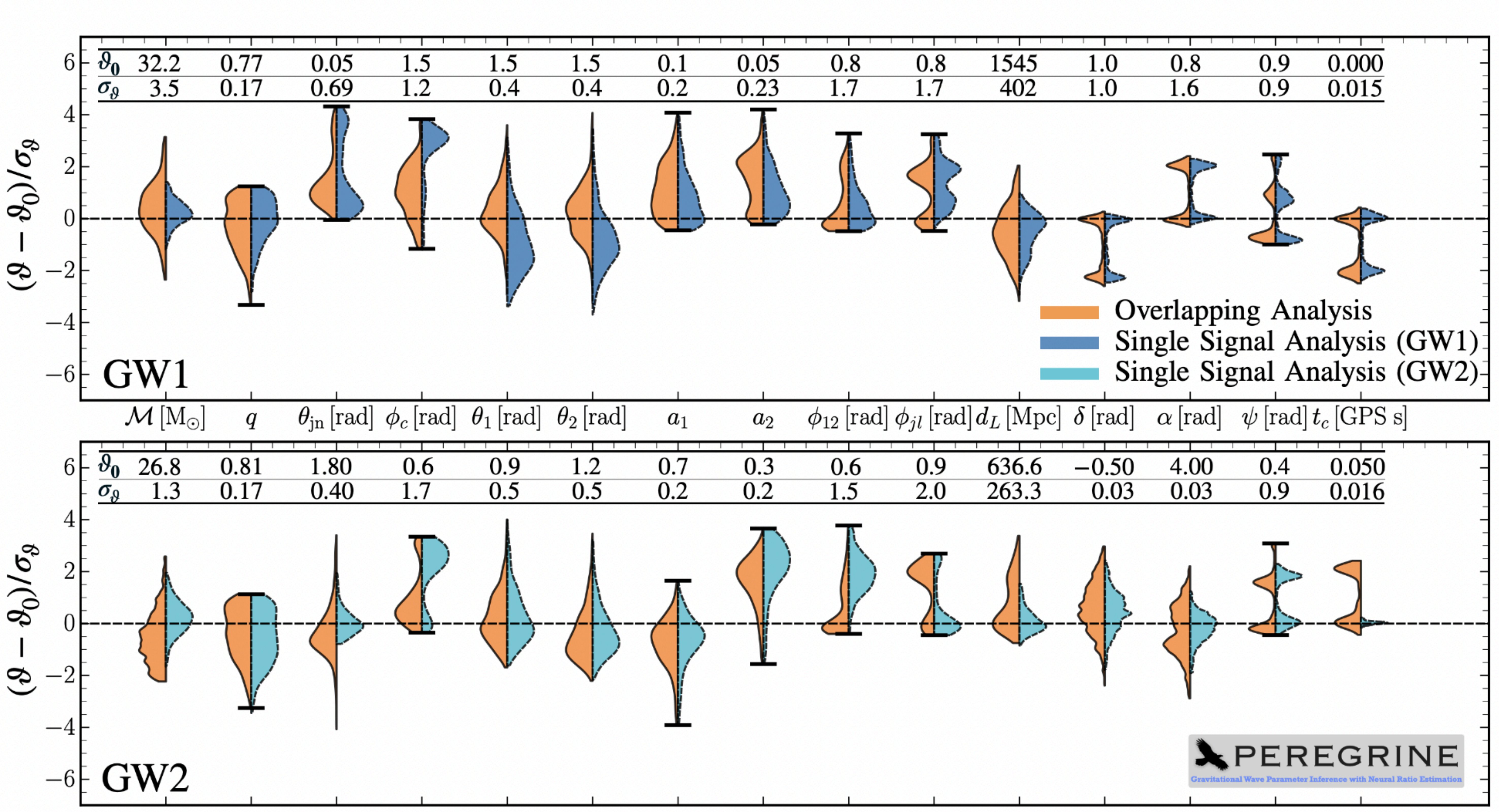
Alvey+ 2308.06318

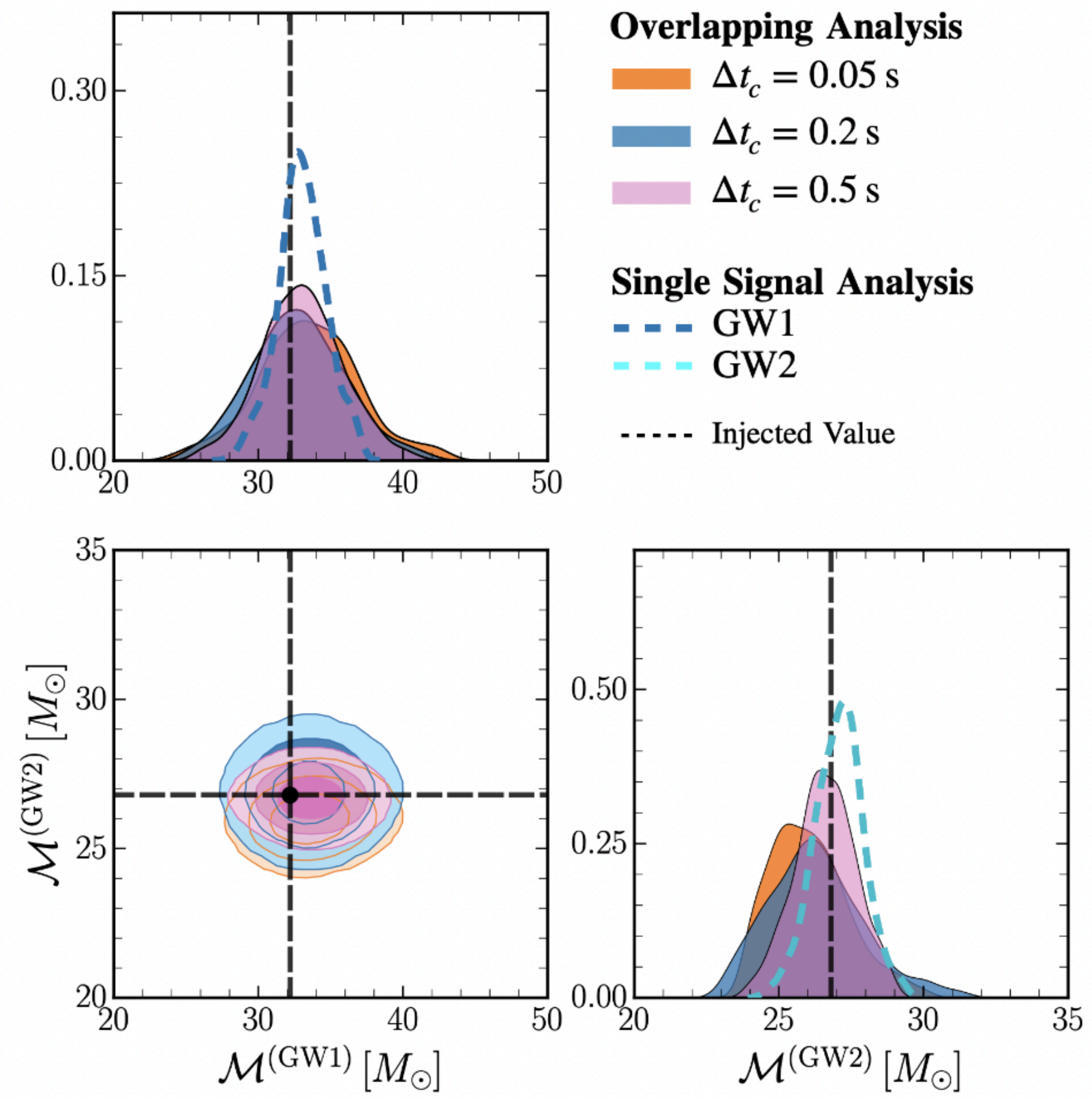
See also: Janquart+ 2211.01304 (sampling-based), Langendorff+ 2211.15097 (normalising flows)

Gravitational wave parameter inference

Solving overlapping GW signals

Check out Uddipta Bhardwaj's talk tomorrow!



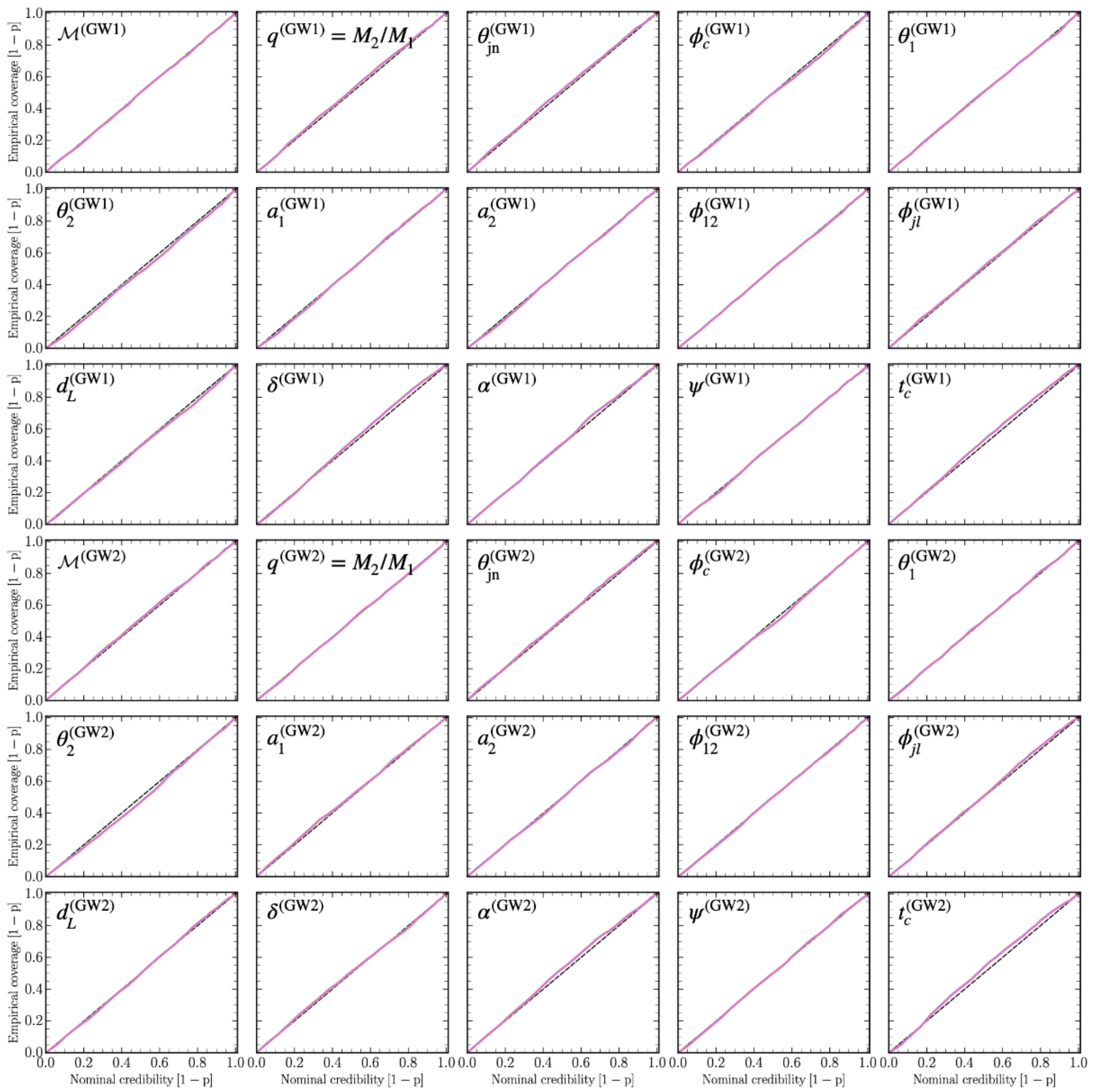


https://github.com/PEREGRINE-GW/peregrine

Alvey+ 2308.06318

Gravitational wave parameter inference

Bayesian credible interval coverage



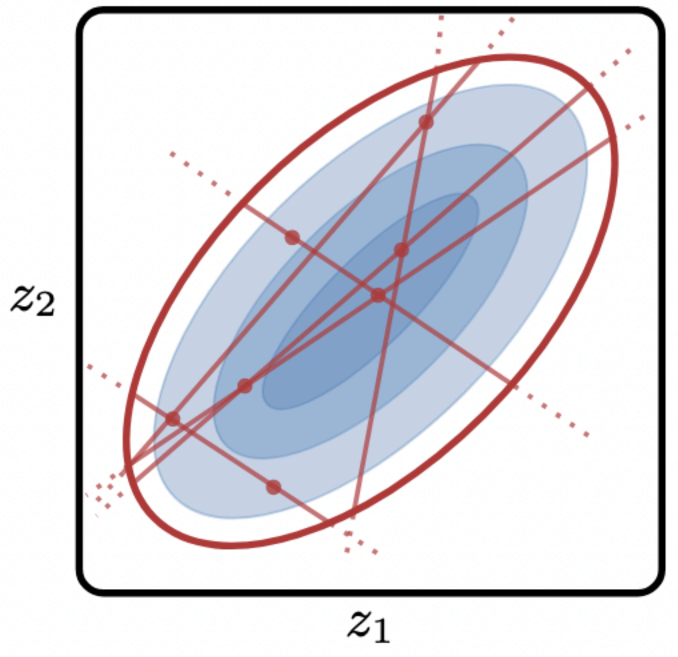
Alvey+ 2308.06318

(High-dim posteriors are possible as well)

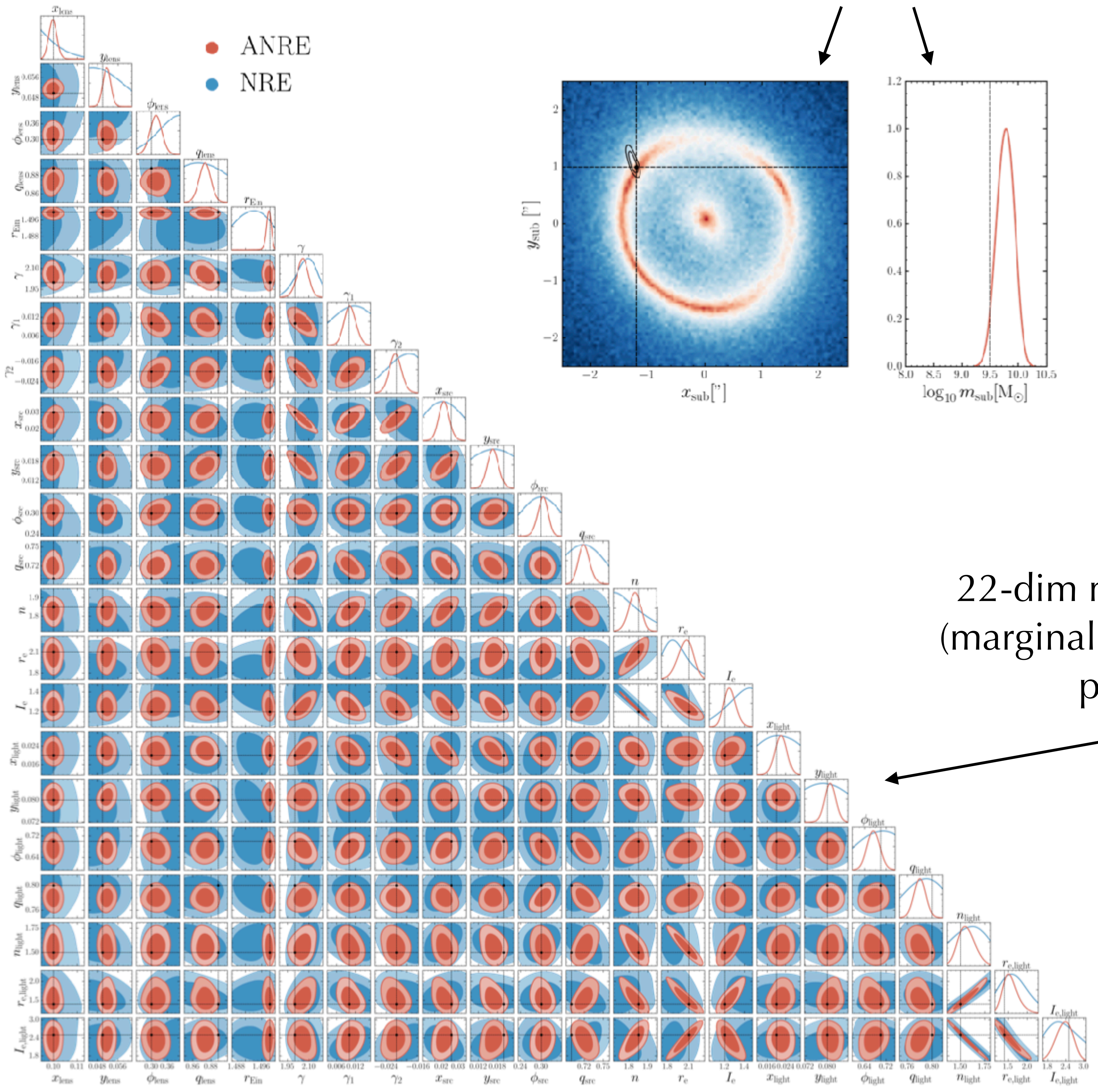
- Example: Strong gravitational galaxy-galaxy lensing
- Method: Autoregressive Neural Ratio estimation

$$p(\mathbf{z} \mid \mathbf{x}) = p(z_1 \mid \mathbf{x}) \prod_{i=2}^{N} p(z_i \mid \mathbf{x}, z_{1:i-1})$$

- Sampling: Nested sampling from joined likelihood



- 1-dim marginal perturber mass posterior
- 2-dim marginal perturber position posterior



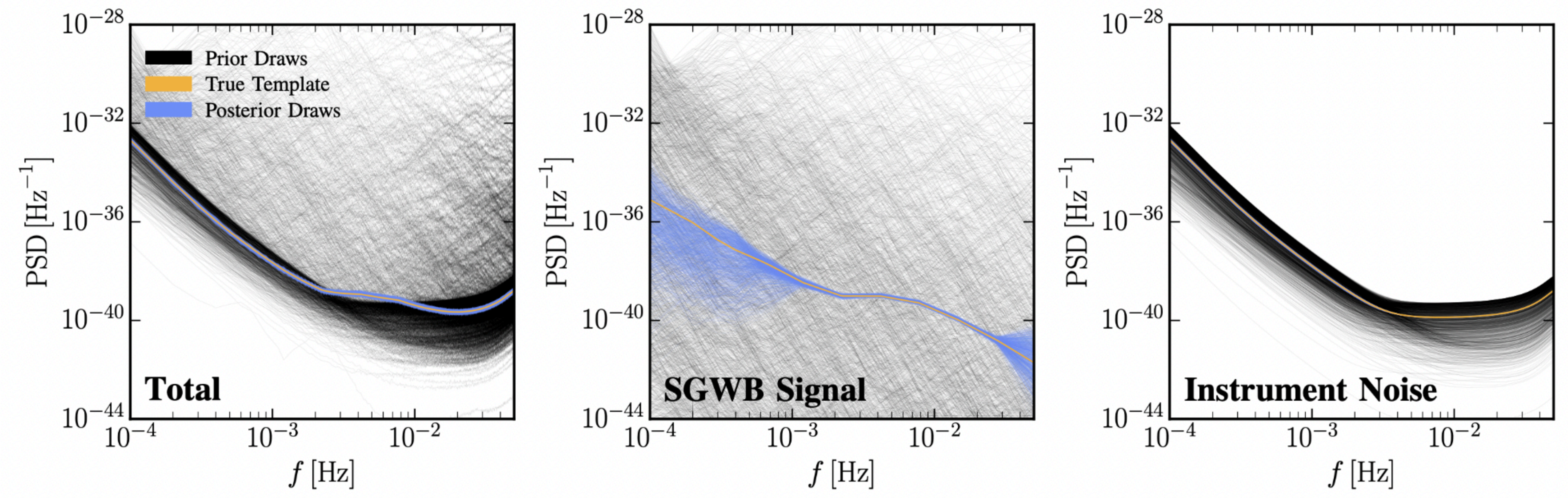
22-dim marginal posterior (marginalised over perturber population)

Anau Montel+ 2308.08597

Stochastic GW background

SAQQARA
Stochastic gravitational wave background analysis

Reconstructing an SGWB from mock LISA data



Simple case study

- 12 days of data, split in 10 chunks to estimate PSD, one TDI channel
- Varying signal model complexity
  - PL
  - agnostic 5-param or 10-param template
- Varying noise model complexity
  - Stationary LISA noise
  - stochastic transient signals (test mass noise, optical metrology system)

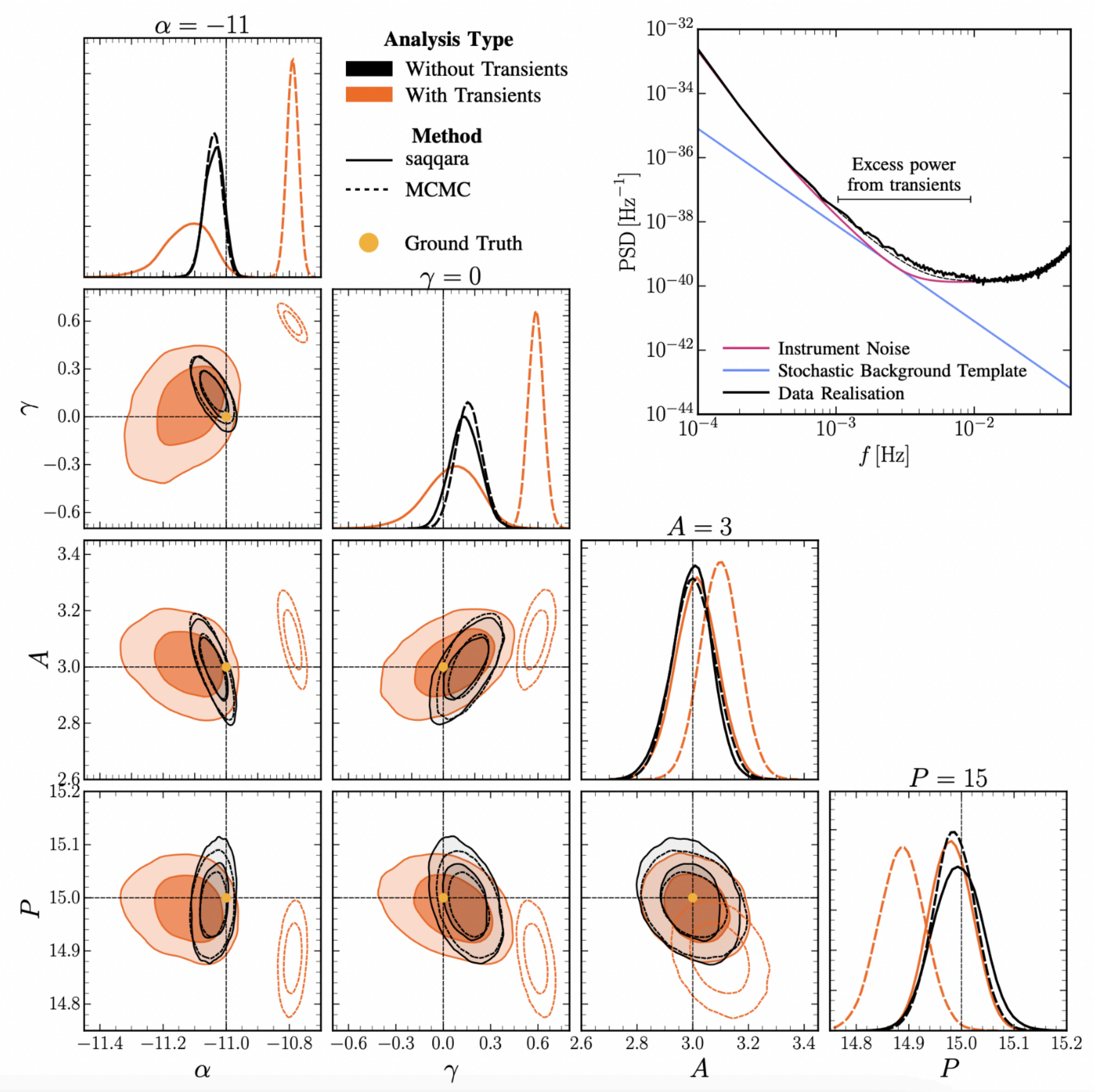
$$\textbf{power law:}\quad \Omega_{\rm GW}(f)h^2 = 10^{\alpha}\left(\frac{f}{\sqrt{f_{\rm min}f_{\rm max}}}\right)^{\gamma}$$

$$\textbf{agnostic:}\quad \Omega_{\rm GW}(f)h^2 = \sum_{i=1}^{N_{\rm bins}} 10^{\alpha_i}\left(\frac{f}{\sqrt{f_{{\rm min},i}f_{{\rm max},i}}}\right)^{\gamma_i} \times \Theta(f - f_{{\rm min},i})\Theta(f_{{\rm max},i} - f)$$

Alvey+ 2309.07954

Stochastic GW background

https://github.com/peregrine-gw/saqqara

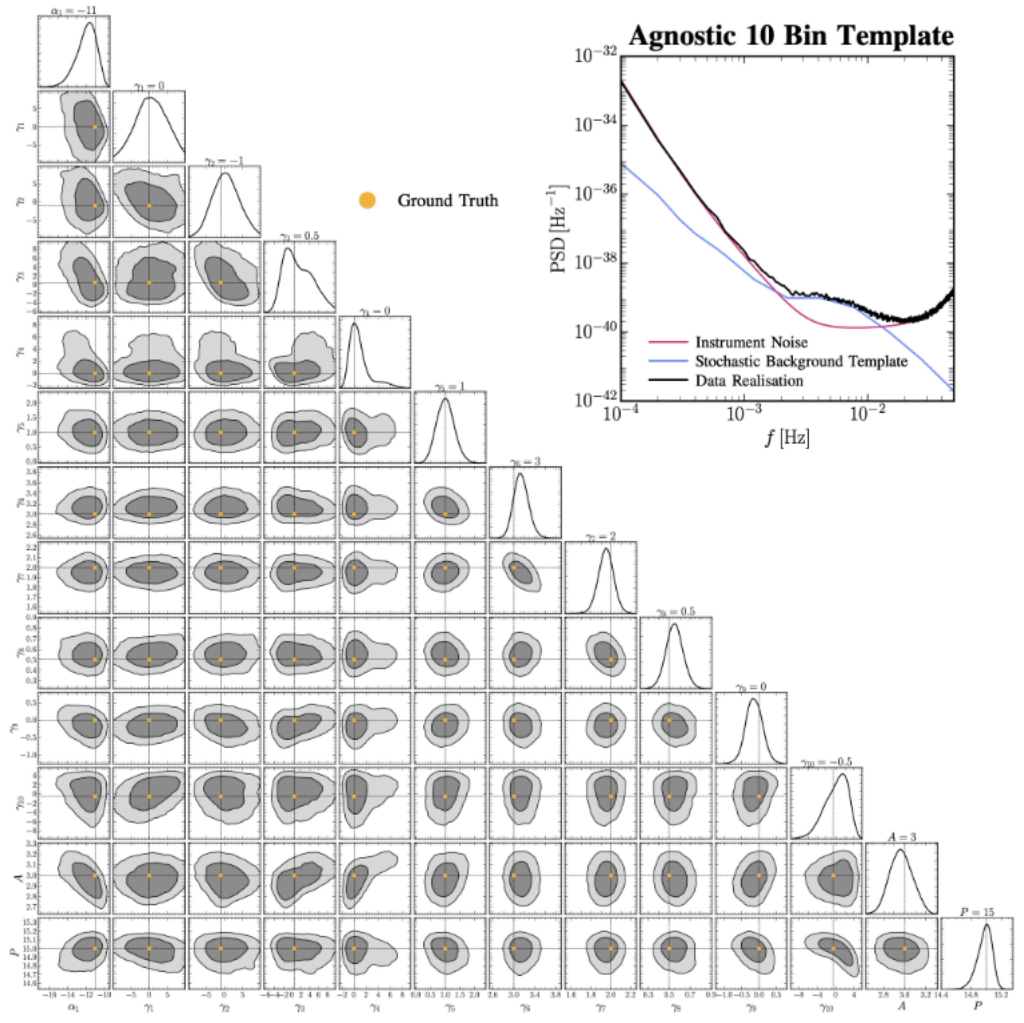


Power-law (with stochastic transient signal)

- In absence of transients, MCMC results are recovered with same precision
- In presence of transients, MCMC leads to biased results (when transients are neglected), while saqqara/Swyft correctly marginalises over transients (if present in simulation-data)

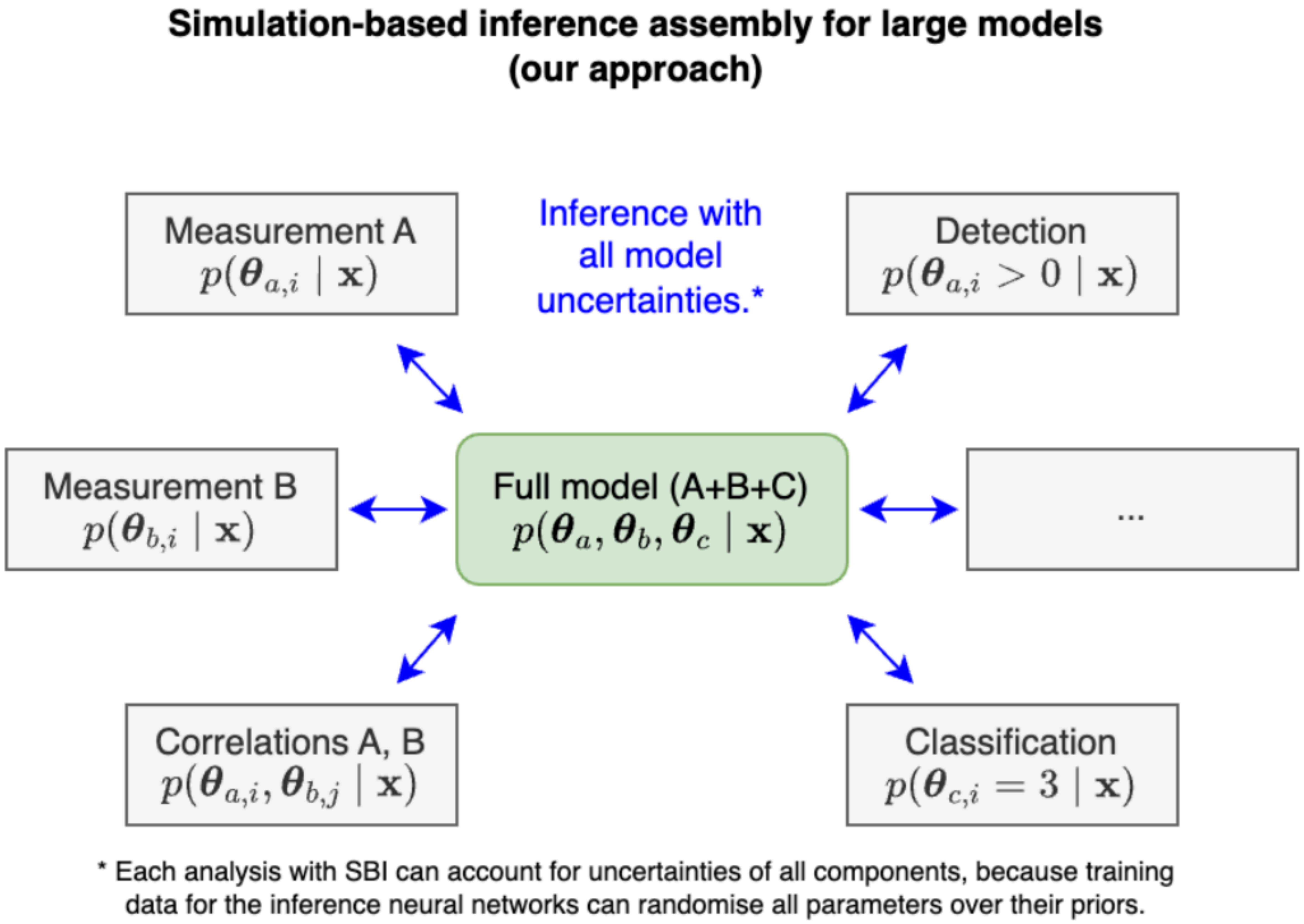
Agnostic templates

- Again, results are identical to MCMC, at significantly reduced computational costs
- Makes it possible to consider high-dimensional scenarios (here 10-parameter agnostic template)



Alvey+ 2309.07954

Goal: One large physics model from many perspectives



ML, GWs, …: Upcoming events in NL

ML4GW@NL: Machine learning for Gravitational Wave Research in the Netherlands

One-day mini-workshop, **Friday 8th December**, Utrecht or Amsterdam

https://indico.nikhef.nl/event/4878/

2nd Swyft workshop

4-day crash course in simulation-based inference with Swyft
5-9 February 2024, Amsterdam (TBC)

https://github.com/undark-lab/swyft

EuCAIFCon - European AI for Fundamental Physics Conference

Large "horizontal" conference on AI in particle physics, astroparticle, physics, gravitational waves, cosmology, nuclear physics … (eucaif.org)

30 April - 3 May 2024, Amsterdam

Conclusions

- The analysis of high-dimensional models is computationally expensive.
- Upcoming gravitational wave data provides a lot of high-dimensional analysis challenges (overlapping waveforms, lensed gravitational waves, all of LISA data)
- The unique capability of SBI is to perform **focused statistical inference**.
- Swyft/TMNRE is our attempt to provide efficient algorithms and tools for focused inference.
- We did first steps in applying SBI to GW analysis problems, with promising results.

Thanks!

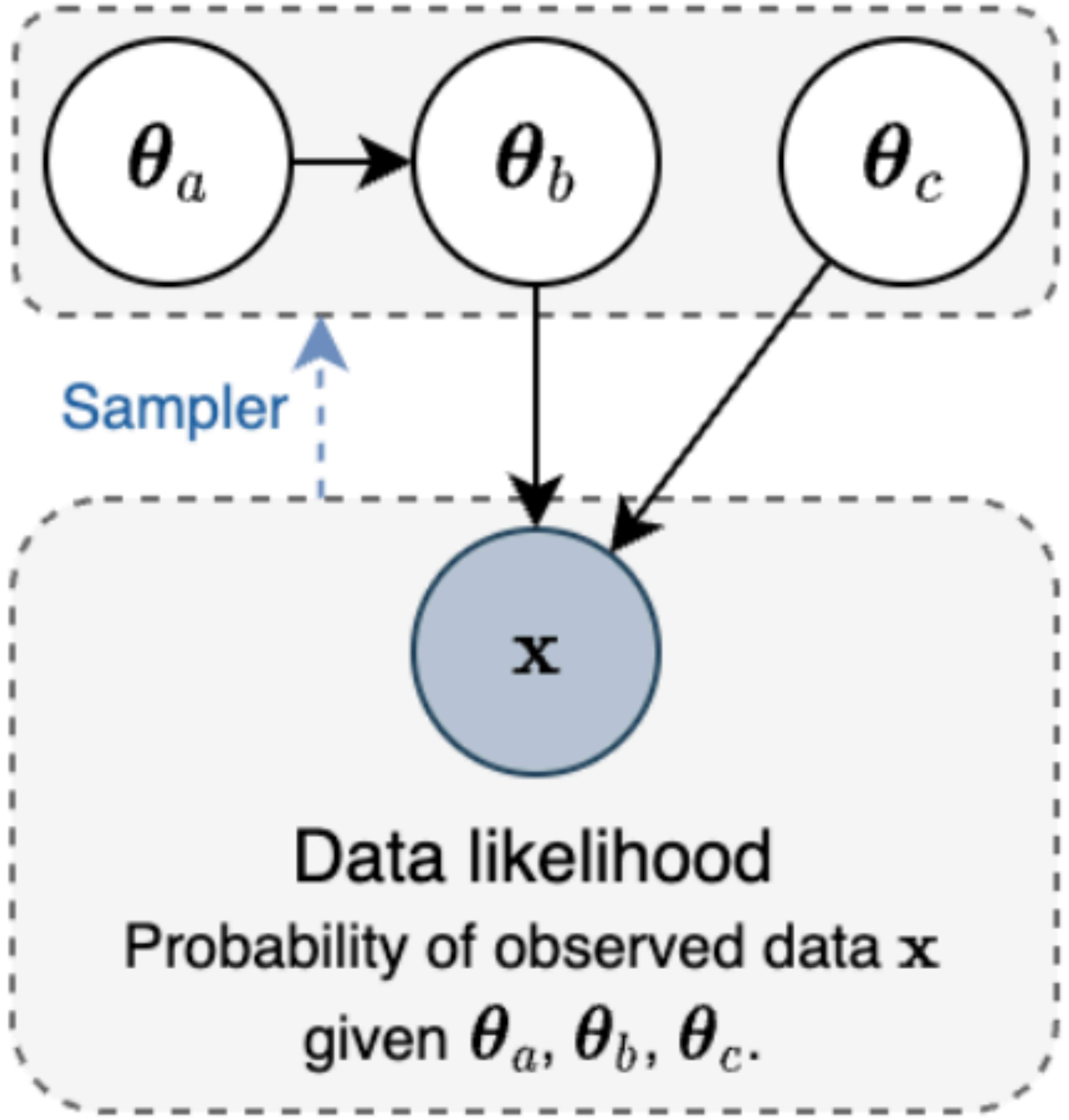
ERC CoG UnDark,
Proj. No. 864035

Backup

Simulation- vs likelihood-based techniques

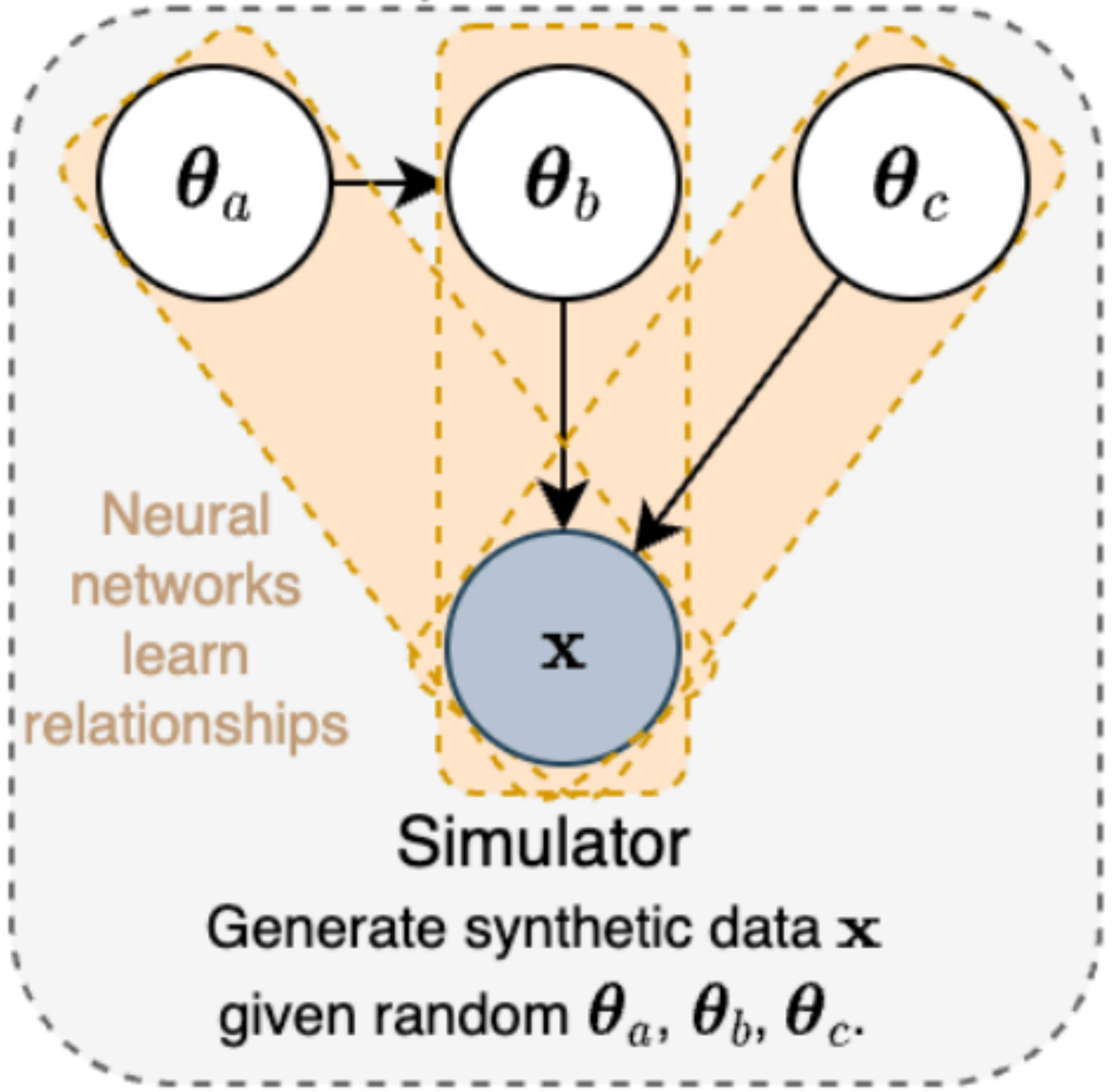
Likelihood-based inference (most common approach)

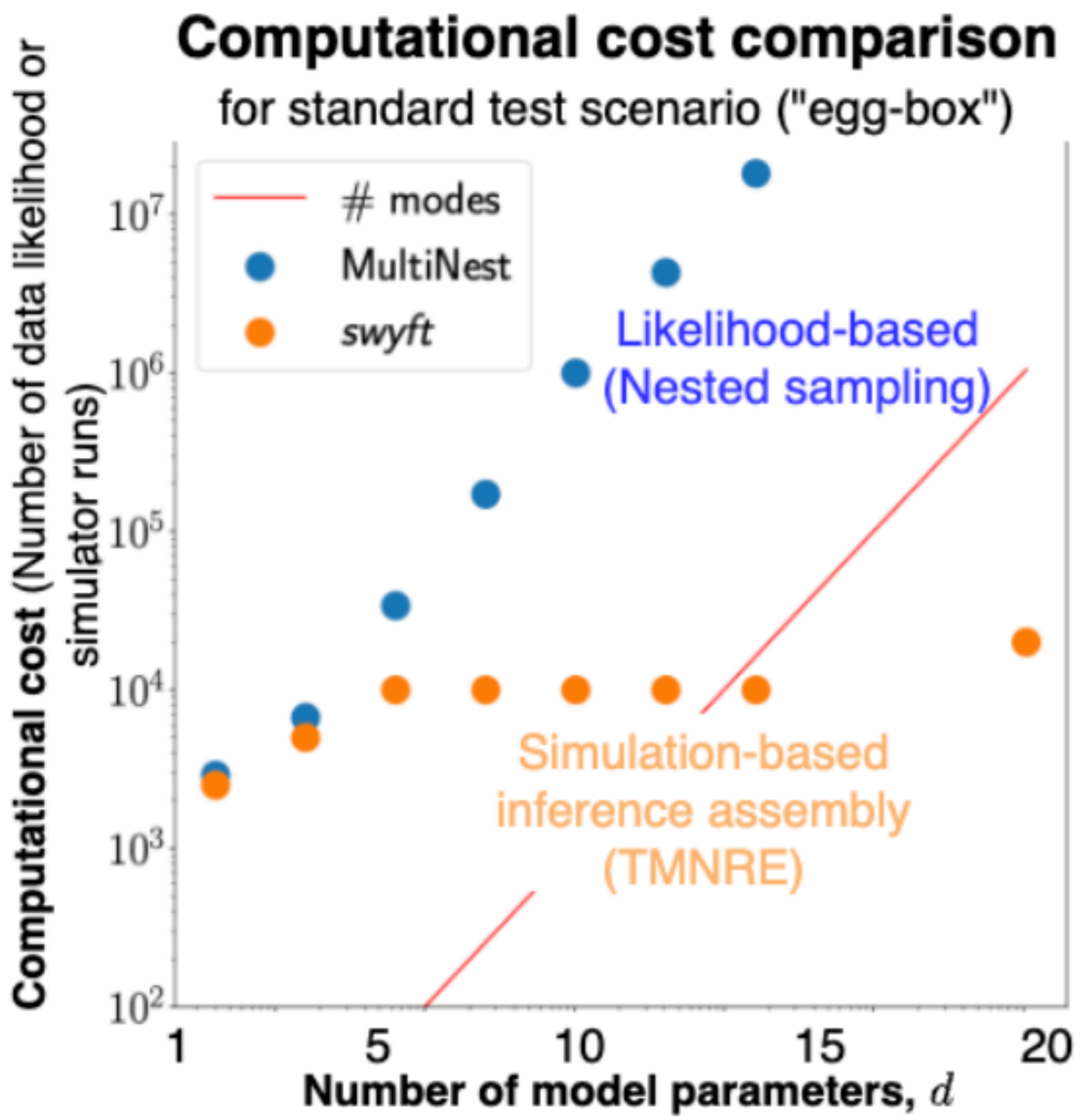
Monolithic joined inference of all model parameters at once



Simulation-based inference assembly (our approach)

Parallel marginal inference of distinct parameter blocks





$\mathbf{x}$: observable data

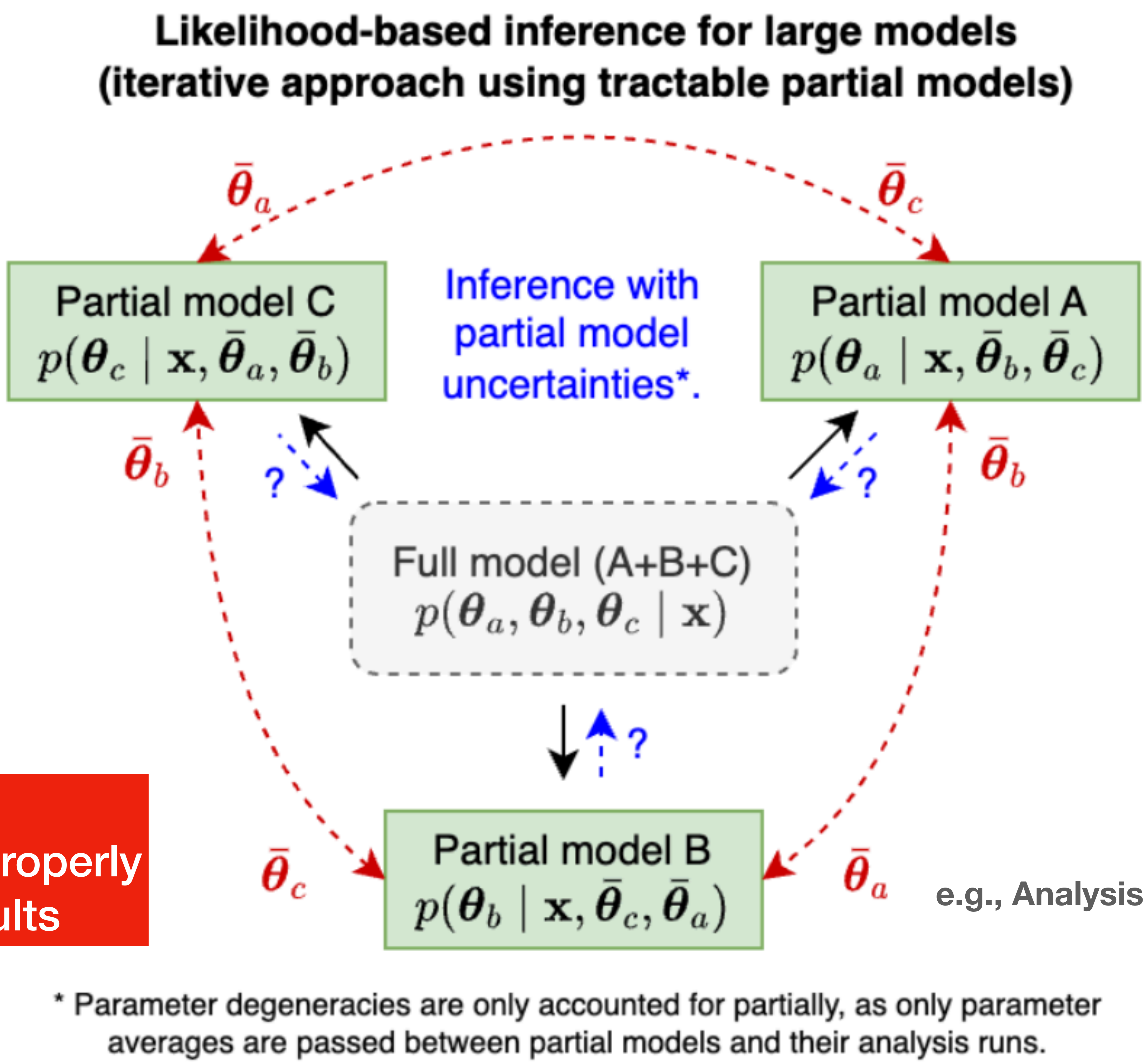
$\boldsymbol{\theta}_a$, $\boldsymbol{\theta}_b$, $\boldsymbol{\theta}_c$: different sets of model parameters, for an exemplary hierarchical model

(*e.g.*, cosmological parameters, $\boldsymbol{\theta}_a$, guide the position of galaxies today $\boldsymbol{\theta}_b$, and with instrumental systematic uncertainties $\boldsymbol{\theta}_c$, lead to data $\mathbf{x}$)

We got used to solve inference problems partially

e.g., Analysis of Fermi Bubbles

e.g., Analysis of point sources



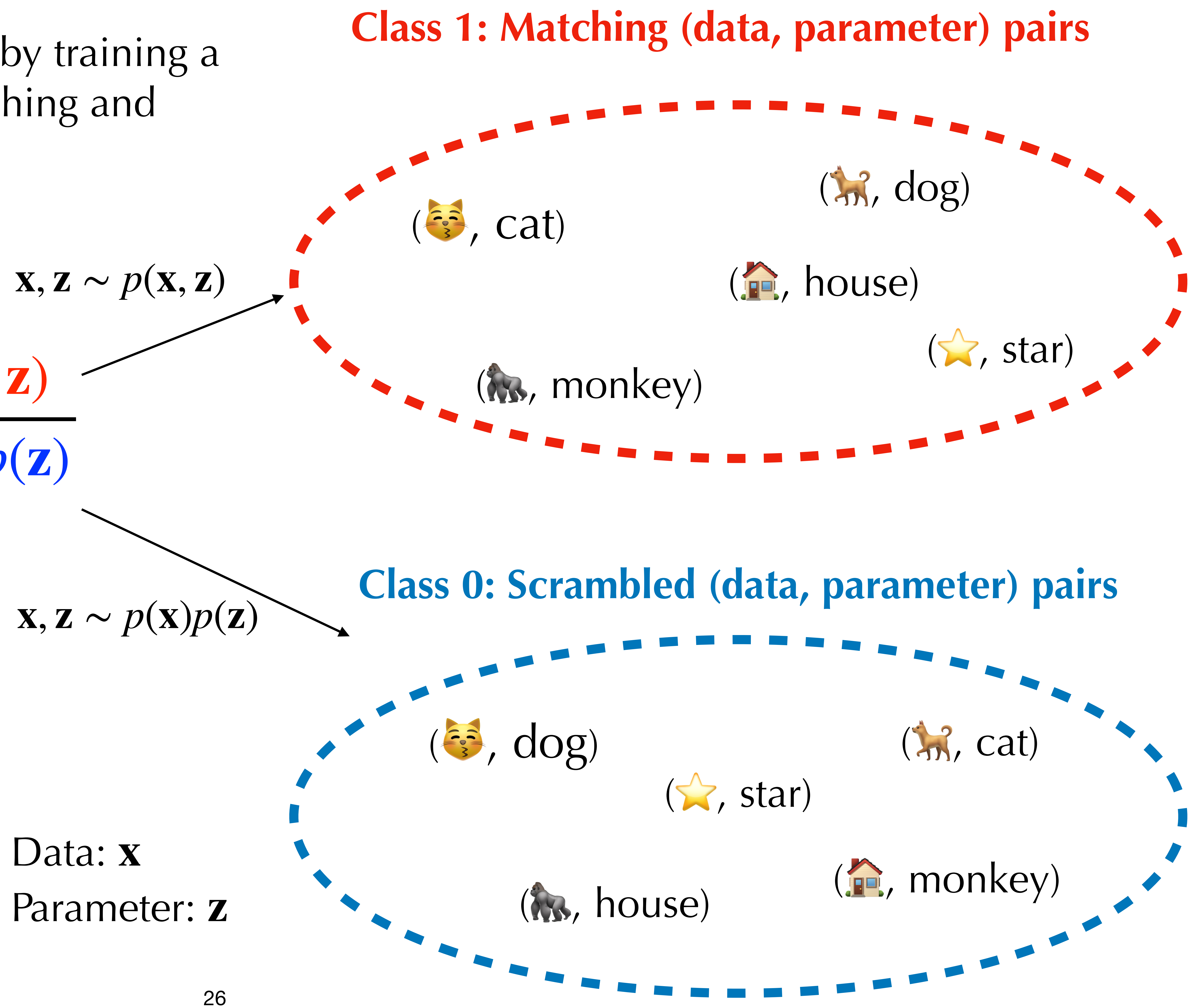
e.g., Analysis of dark gas

Problem:
Uncertainties are not propagated properly
=> Overconfident & biased results

NRE = binary classification

Strategy: We estimate posteriors-to-prior ratio by training a binary classifier to discriminate between matching and scrambled (data, parameter) pairs.

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



Data: $\mathbf{x}$

Parameter: $\mathbf{z}$

Hermans+ 1903.04057

Neural ratio estimation

TMNRE

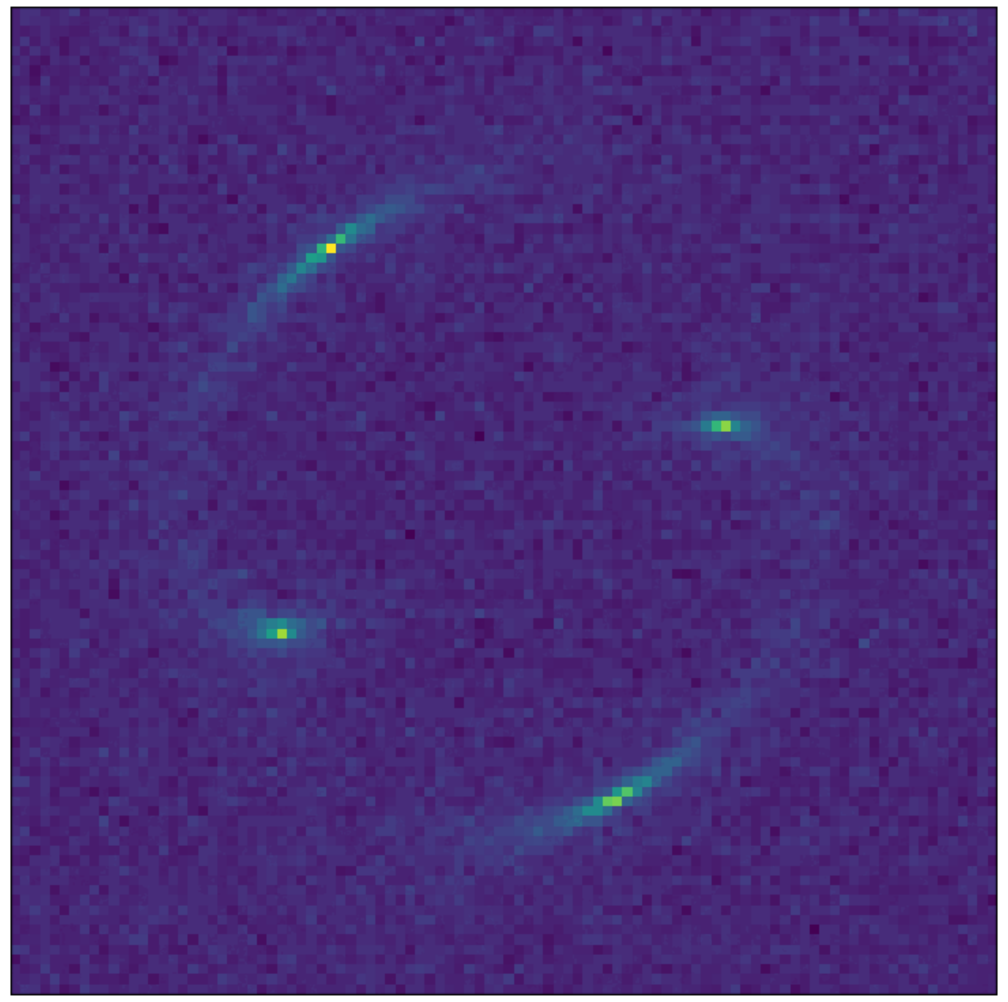
Algorithm

Generate training data

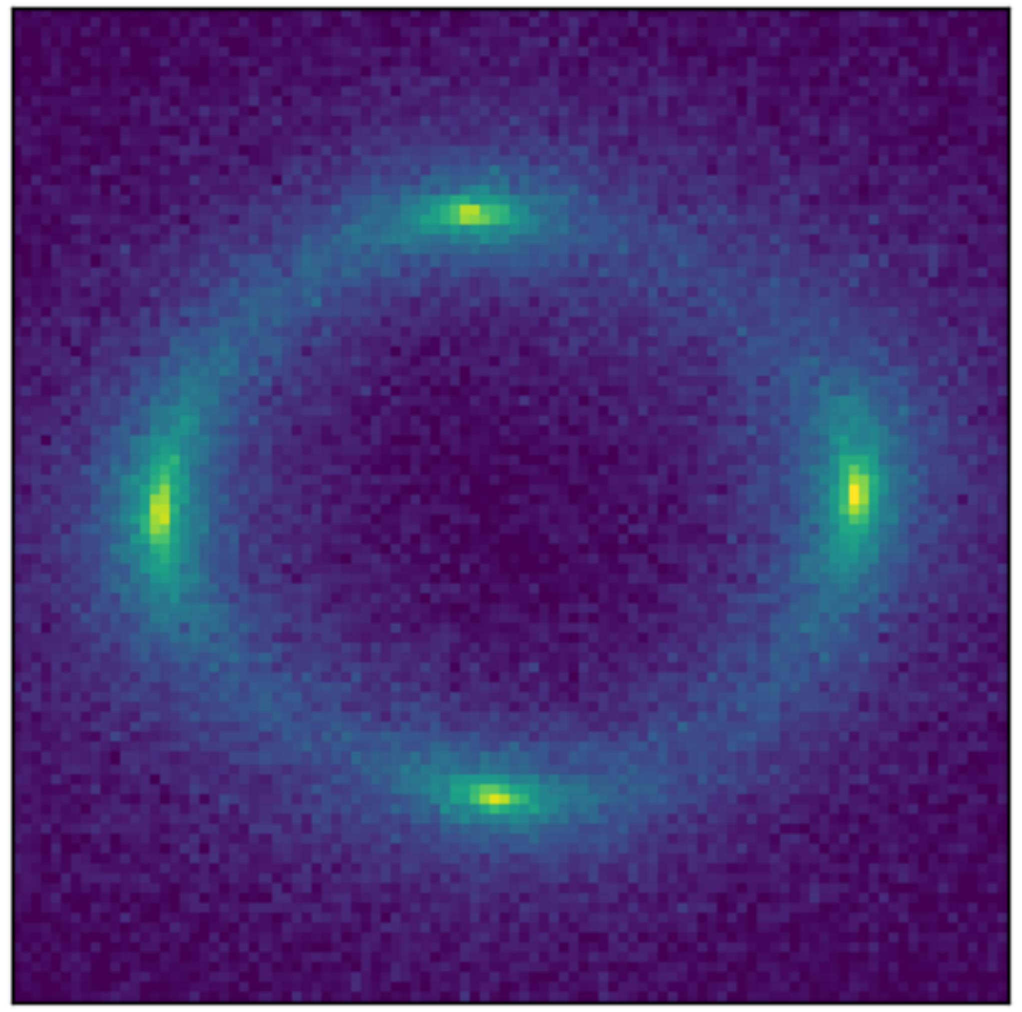
$$\mathscr{D} \equiv \{(\mathbf{x}_i, \mathbf{z}_i) \mid i = 1,2,\ldots,N\},\ \mathbf{x}, \mathbf{z} \sim p(\mathbf{x}, \mathbf{z}) = p(\mathbf{x} \mid \mathbf{z})p(\mathbf{z})$$

Initialise real-valued neural network, $f_\phi(\mathbf{x}, \mathbf{z})$

$\mathbf{x}_i$ for $i = 1,2,\ldots,N$



$\mathbf{x}_{\text{obs}}$



Train neural network using the mini-batch loss function

$$\mathscr{L}(\phi) = -\sum_{i \in B} \ln \sigma(f_\phi(\mathbf{x}_i, \mathbf{z}_i)) + \ln \sigma(-f_\phi(\mathbf{x}_i, \mathbf{z}_{P(i)}))$$

$Y = 1 : \mathbf{x}, \mathbf{z} \sim p(\mathbf{x}, \mathbf{z})$ $\qquad$ $Y = 0 : \mathbf{x}, \mathbf{z} \sim p(\mathbf{x})p(\mathbf{z})$

Here, B denotes a mini-batch, and P denotes random sample permutations.

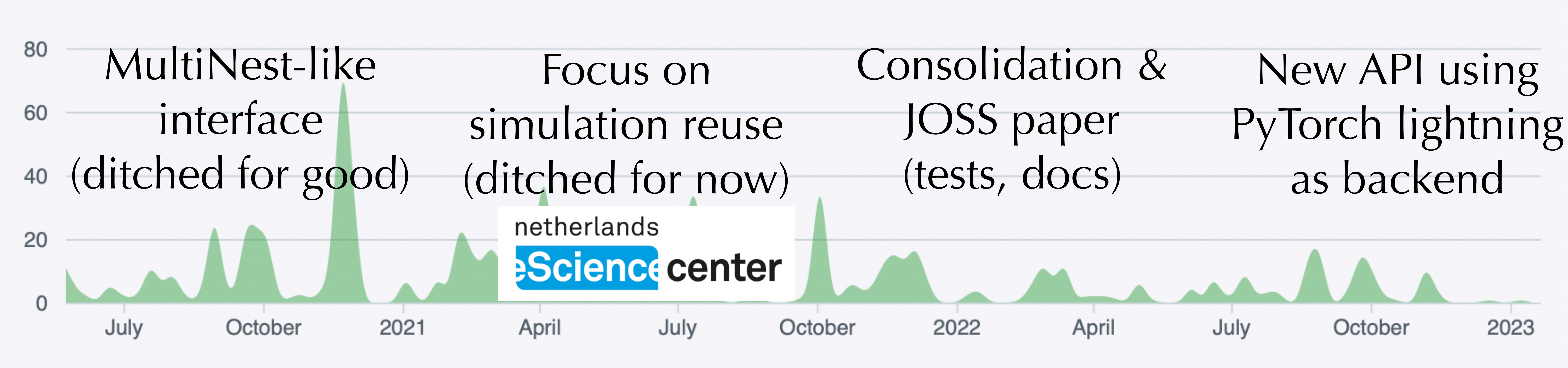
After training, $f_\phi(\mathbf{x}, \mathbf{z}) \approx \ln r(\mathbf{x}; \mathbf{z}) = \ln \dfrac{p(\mathbf{z} \mid \mathbf{x})}{p(\mathbf{z})}$

Swyft software package

Everybody being frustrated that code is changing all the time → ~Stability

"Are you using the latest Swyft version?"

Initial plan:
"Hey, writing a python module for TMNRE would be cool and impactful, ***should take 2-3 weeks****"*



3 years later…

Miller+ 2011.13951 Miller+ 2107.01214 Miller+ JOSS 2022

Methods that we worked with in our group

Variational inference

PyKeOps

Normalising flows

HMC

TMNRE

Gaussian processes

Probabilistic programming

Differentiable simulators

Scalable TMNRE

Density TMNRE

Hierarchical TMNRE

Image analysis TMNRE