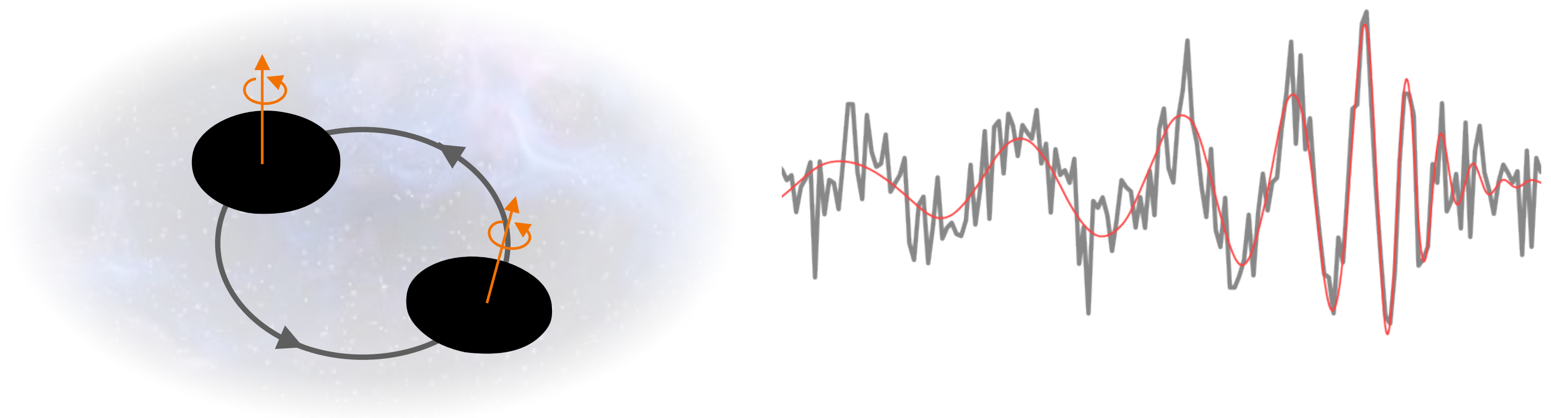
Gravitational Wave Analysis with Machine Learning

Note: the pdf version breaks some overlays and animations (e.g. the gifs on slides 4, 19)

Maximilian Dax, MPI for Intelligent Systems (Tübingen), mdax@tue.mpg.de



[1] Dax et al., Real-Time Gravitational Wave Science with Neural Posterior Estimation, Phys. Rev. Lett. **127**, 241103 (2021)
[2] Dax et al., Group equivariant neural posterior estimation, ICLR 2022
[3] Dax, et al., Neural Importance Sampling for Rapid and Reliable Gravitational-Wave Inference, Phys.Rev.Lett. **130**, 171403 (2023)

MAX PLANCK INSTITUTE
FOR INTELLIGENT SYSTEMS

Collaborators: Green, Gair, Wildberger, Buchholz, Gupte, Pürrer, Deistler, Macke, Buonanno, Schölkopf

PG PARENTAL GUIDANCE RECOMMENDED. THE FOLLOWING CONTAINS PHYSICS-UNRELATED CONTENT.

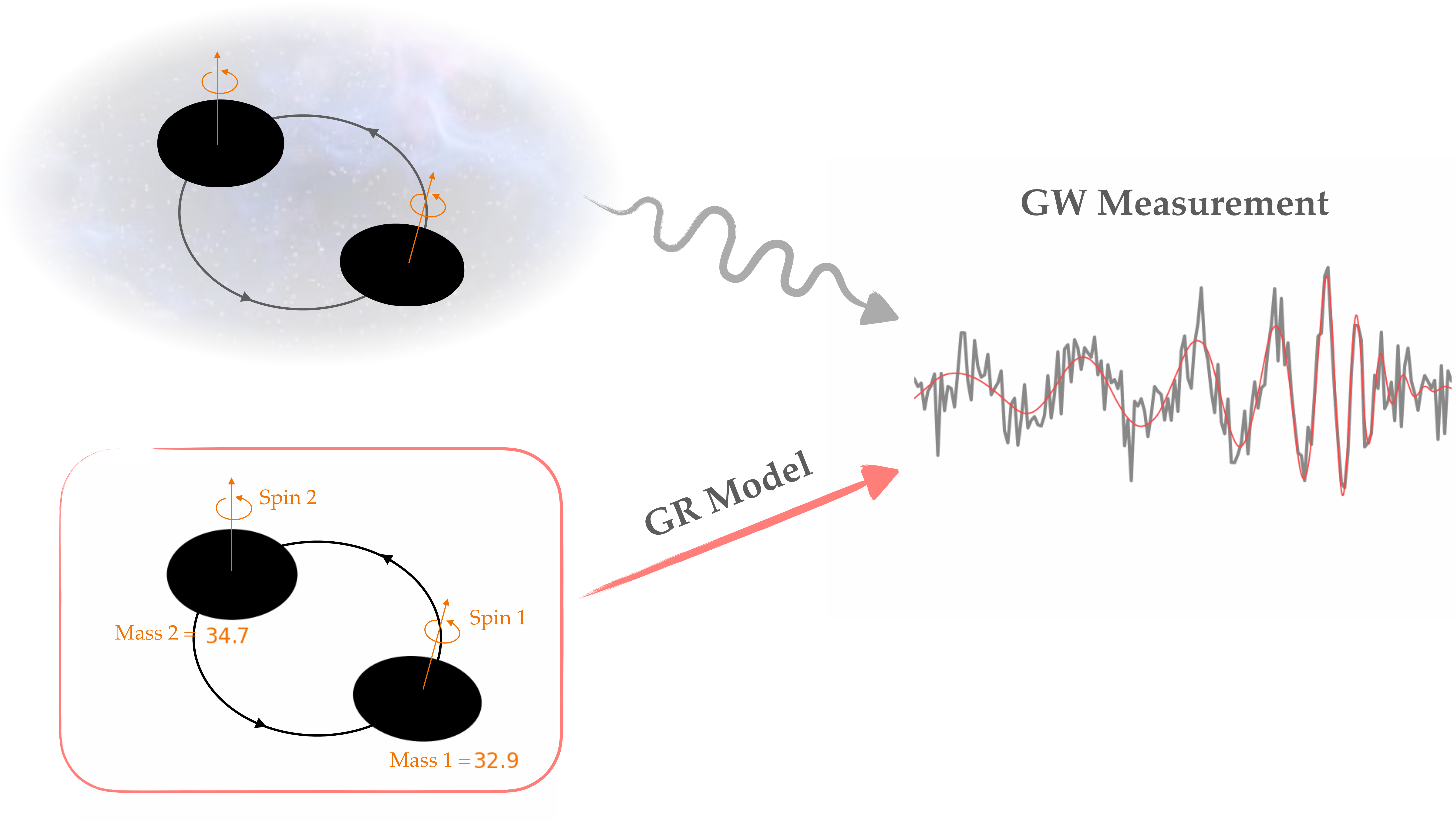
Turning **known-unknowns**
into **known-knowns**,

and thereby help to investigate the
unknown-unknowns.

I. Real-Time GW Inference

[1] Dax et al., Real-Time Gravitational Wave Science with Neural Posterior Estimation, Phys. Rev. Lett. **127**, 241103 (2021)

Gravitational wave analysis: comparing data to models



GW Measurement

GR Model

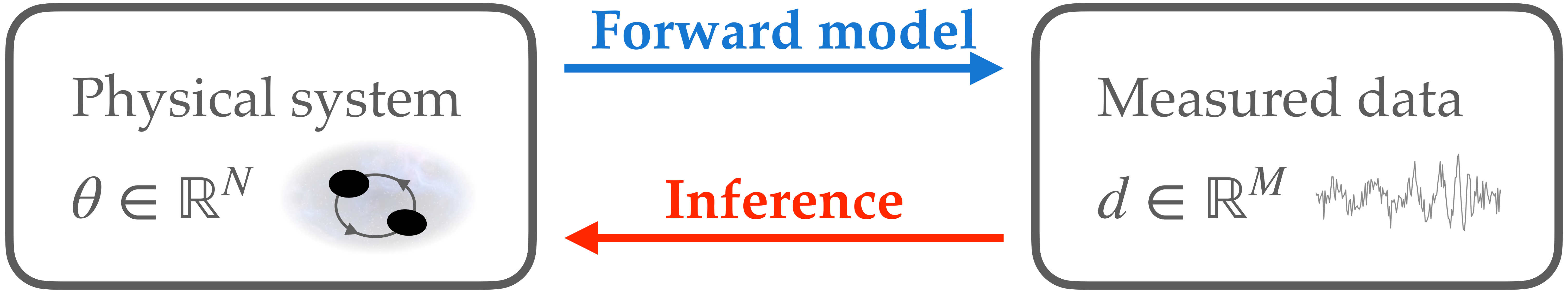
General relativity (GR)

- Black hole mergers emit gravitational waves (GWs)
- GW shape depends on the black hole properties
  15 parameters: masses, spins, …

GW analysis

Decode GW information to characterize the black holes

Inverse problems in science



Physical system

$\theta \in \mathbb{R}^N$

Forward model

Inference

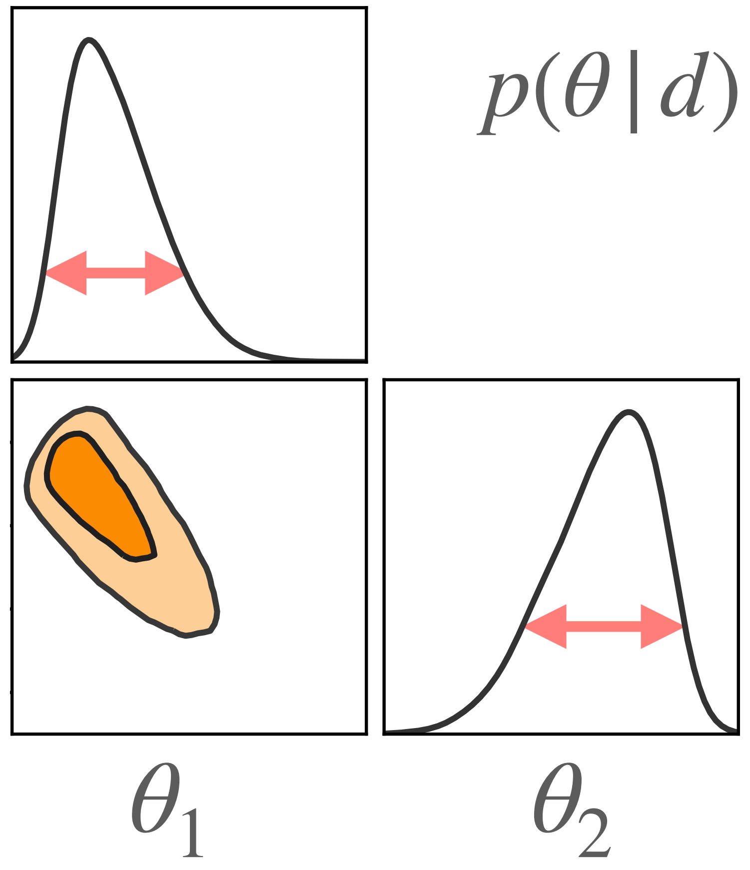
Measured data

$d \in \mathbb{R}^M$

- **Forward direction $\theta \to d$** is defined by a simulator, $d \sim p(d\,|\,\theta)$

- **Inverse direction** with Bayesian inference

$$p(\theta\,|\,d) = \frac{\overbrace{p(d\,|\,\theta)}^{\text{forward model}}\ \overbrace{p(\theta)}^{\text{prior belief}}}{p(d)}$$



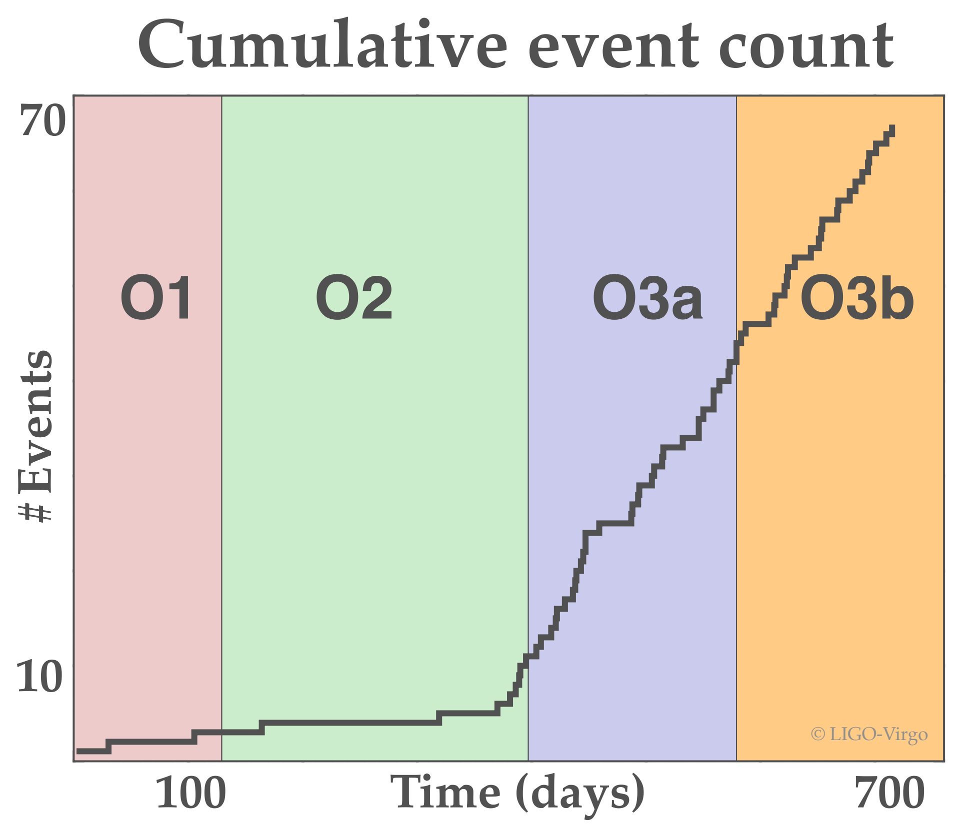
$p(\theta\,|\,d)$ **Bayesian posterior**

→ uncertainties

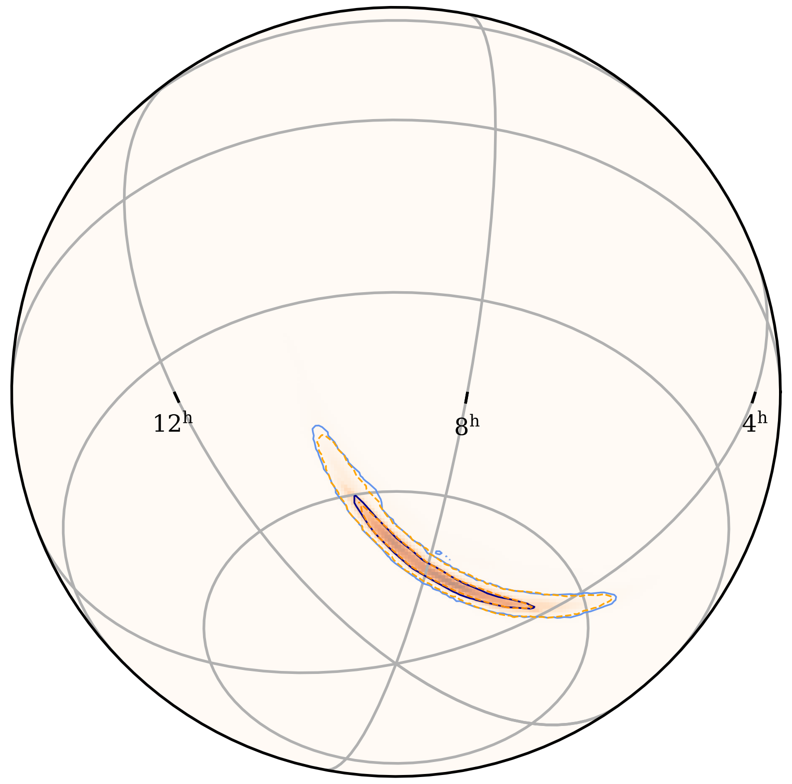
→ correlations

Limitations of conventional GW inference [e.g., MCMC]

- **Speed** (~days)

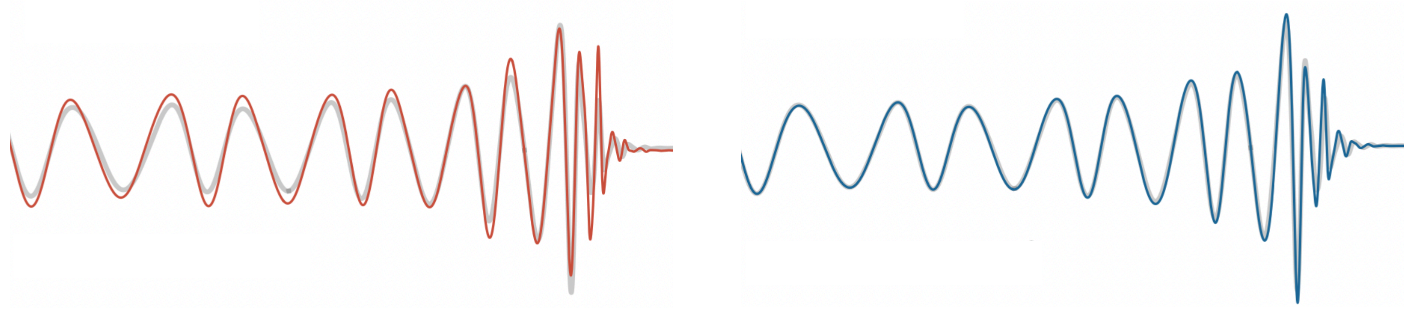


computationally costly
(increasing event rate!)

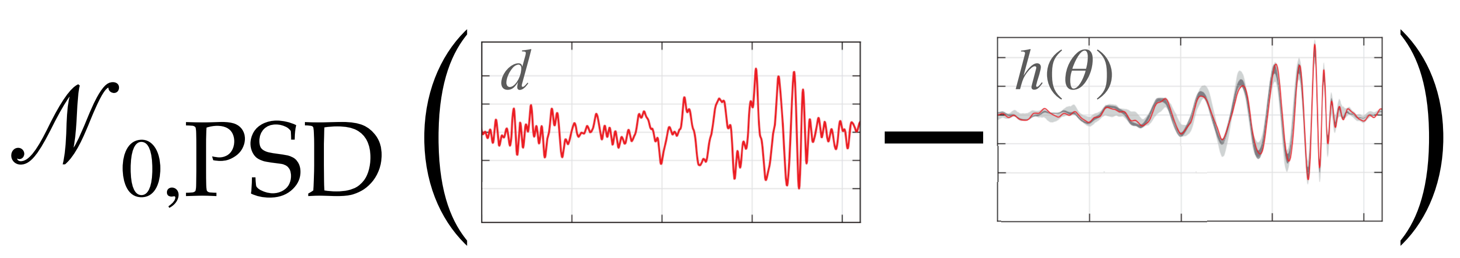


no fast localization
for e.m. follow-up

- **Accuracy**



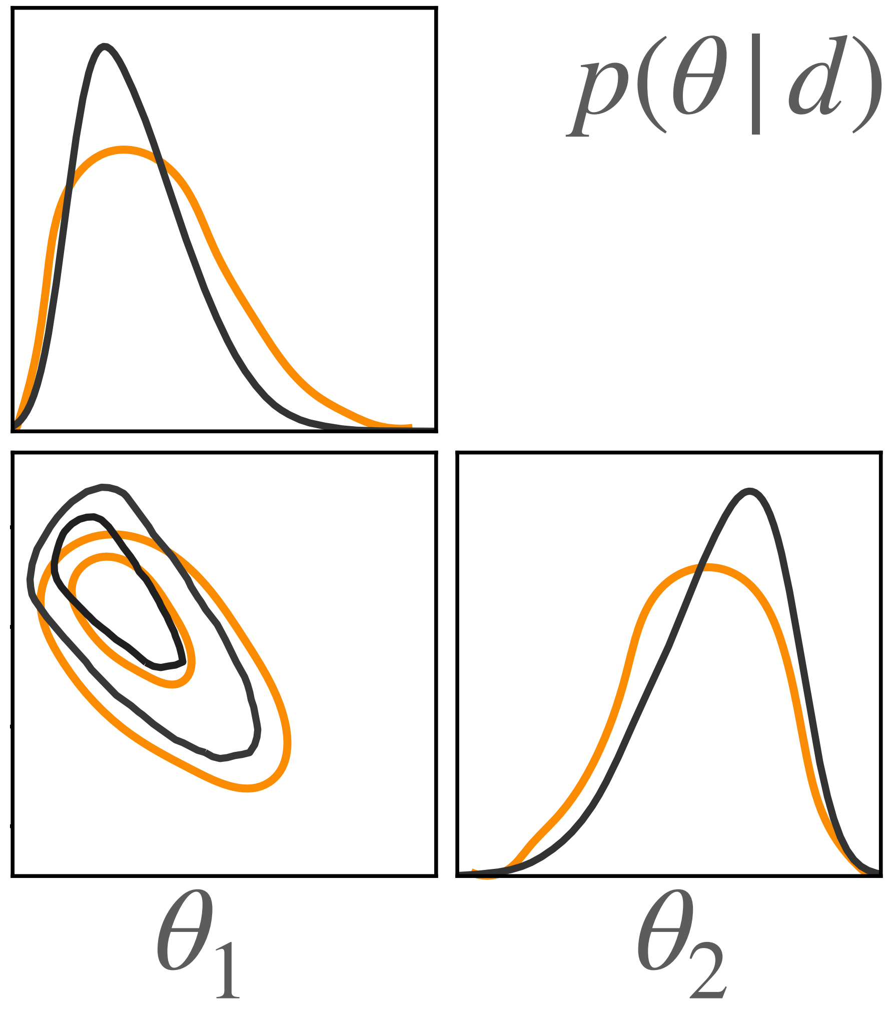
don’t scale to
high-quality GW models



$$\mathcal{N}_{0,\mathrm{PSD}}\left(d - h(\theta)\right)$$

require tractable likelihood
⇒ need noise model

Amortised simulation-based inference (SBI)



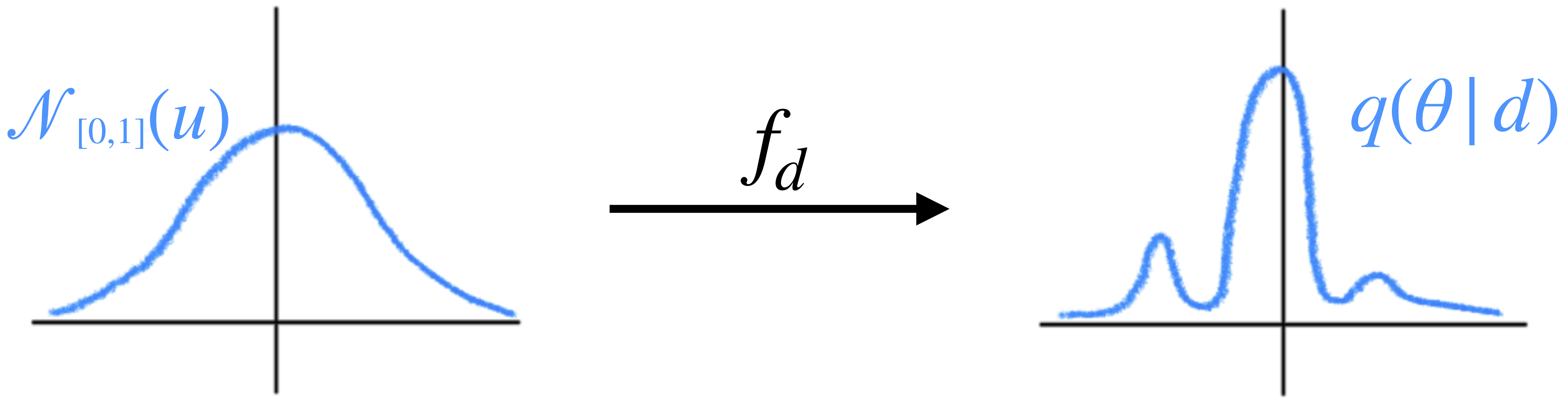
- Traditionally: inference with **stochastic samplers** (e.g., MCMC)
  - need tractable likelihood $p(d|\theta)$ → **simplifying assumption**
  - need many likelihood evaluations → **expensive**

- Amortised SBI: **fit neural network** $q(\theta|d)$ **as surrogate for** $p(\theta|d)$
  - Train with samples $d \sim p(d|\theta)$ → **no simplifying assumption**
  - Perform inference with trained network → **cheap**

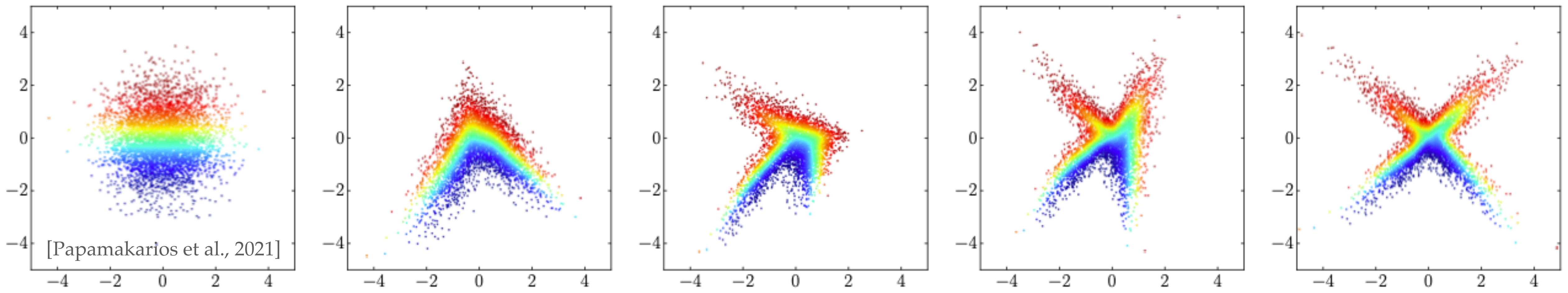
- Requirements
  1. Expressive **density estimator** $q(\theta|d)$ (N-dim density, conditional on d)
  2. **Training strategy** s.t. $q(\theta|d) = p(\theta|d)\ \forall d$

Normalising flows Rezende & Mohamed, ICML 2015

- Idea: transform base distribution $\mathcal{N}_{[0,1]}$ to $q(\theta | d)$ via f_d



- Flexible f_d achieved by composition of simple transforms



[Papamakarios et al., 2021]

- Normalizing flows can be made **arbitrarily expressive**

$$\theta = f_d(u), \quad u \sim \mathcal{N}_{[0,1]}(u)$$

$$q(\theta | d) = \mathcal{N}_{[0,1]}\left(f_d^{-1}(\theta)\right) \left| \det J_{f_d}^{-1} \right|$$

f_d parameterised using neural network
with learnable parameters ϕ

Neural posterior estimation (NPE)

Papamakarios & Murray, NeurIPS 2016

- Minimize

average over d

D_{KL} for fixed d

$$D_{\mathrm{KL}}\left(p\,|\,q\right) = \int \mathrm{d}d\, p(d) \int \mathrm{d}\theta\, p(\theta\,|\,d)\, \log\left(\frac{p(\theta\,|\,d)}{q(\theta\,|\,d)}\right)$$

$$= \int \mathrm{d}d\, p(d) \int \mathrm{d}\theta\, \frac{p(\theta)\, p(d\,|\,\theta)}{p(d)} \left[-\log q(\theta\,|\,d) + \cancel{\log p(\theta\,|\,d)}\right] \quad = \text{const.}$$

$$\sim \int \mathrm{d}\theta\, p(\theta) \int \mathrm{d}d\, p(d\,|\,\theta) \left[-\log q(\theta\,|\,d)\right] = -\,\mathbb{E}_{\theta\sim p(\theta)}\mathbb{E}_{d\sim p(d|\theta)}\left[\log q(\theta\,|\,d)\right]$$

Neural posterior estimation (NPE) Papamakarios & Murray, NeurIPS 2016

- Minimize

$$D_{\mathrm{KL}}\left(p\,|\,q\right) = -\,\mathbb{E}_{\theta\sim p(\theta)}\mathbb{E}_{d\sim p(d|\theta)}\left[\log q(\theta\,|\,d)\right] \quad + \mathrm{const.}$$

- Monte Carlo approximation: train flow by minimizing loss L across dataset $\mathcal{D}$

$$\boxed{L = -\,\log q_{\phi}(\theta\,|\,d), \qquad \mathcal{D} = \left\{\theta^{(i)}, d^{(i)}\right\}_{i=1}^{N}, \quad \theta^{(i)} \sim p(\theta),\ d^{(i)} \sim p(d\,|\,\theta^{(i)})}$$

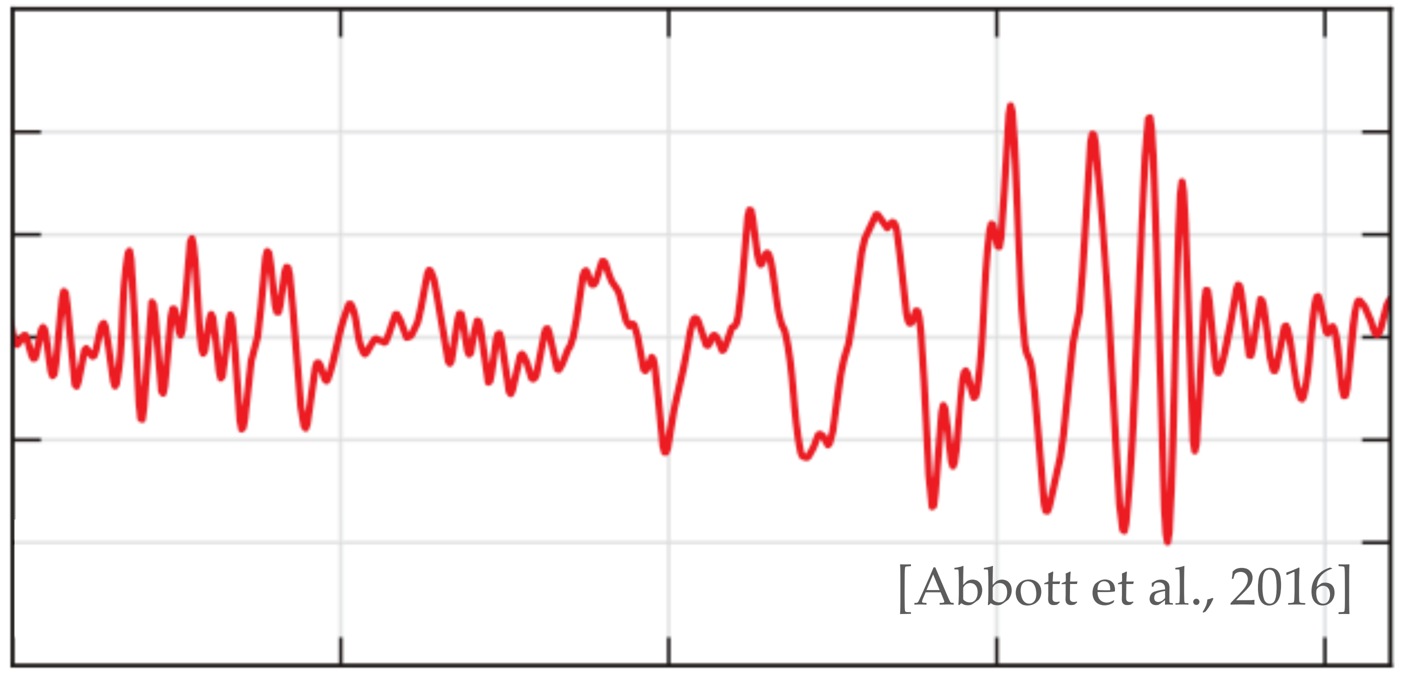
- Minimization of D_{KL} + arbitrarily expressive ⇒ **perfect recovery of posterior**

- NPE uses same ingredients as MCMC (prior + likelihood), but **only requires samples**

GW simulator

- GW measurement: **signal** + **noise**

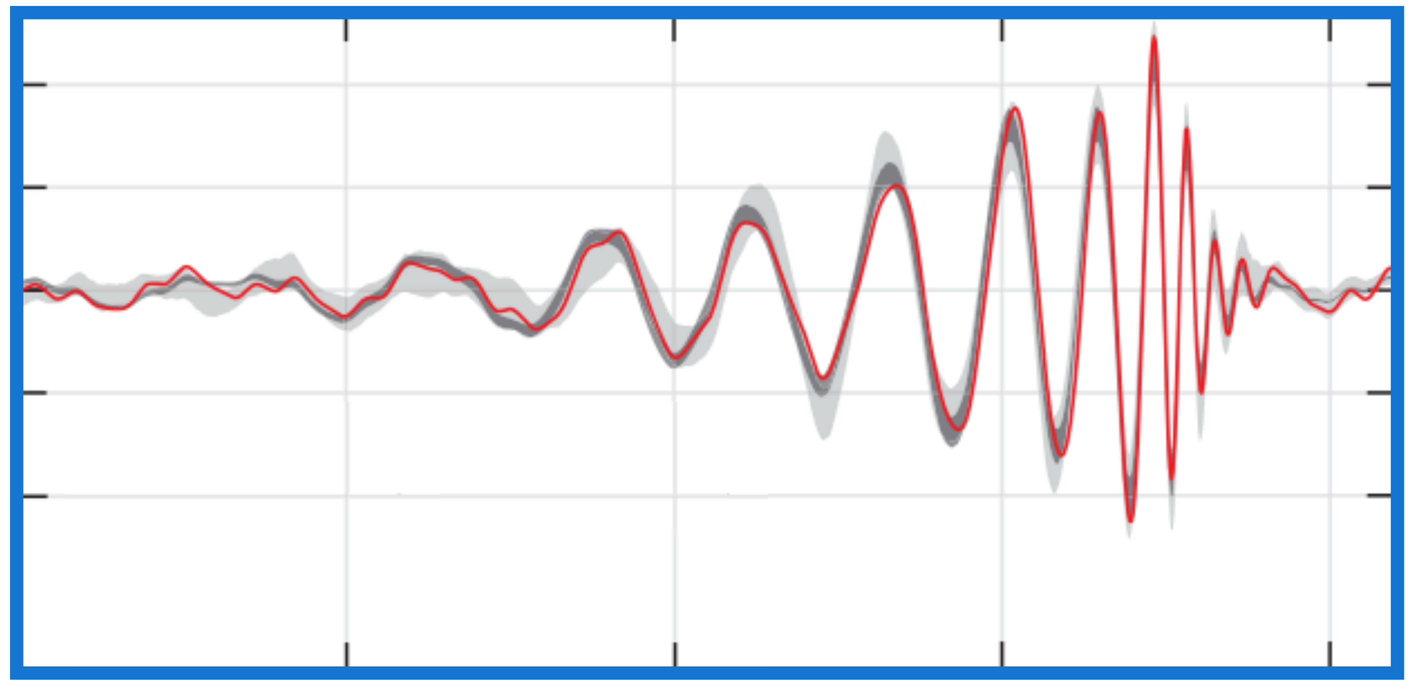
$d \sim p(d|\theta)$



[Abbott et al., 2016]

$=$

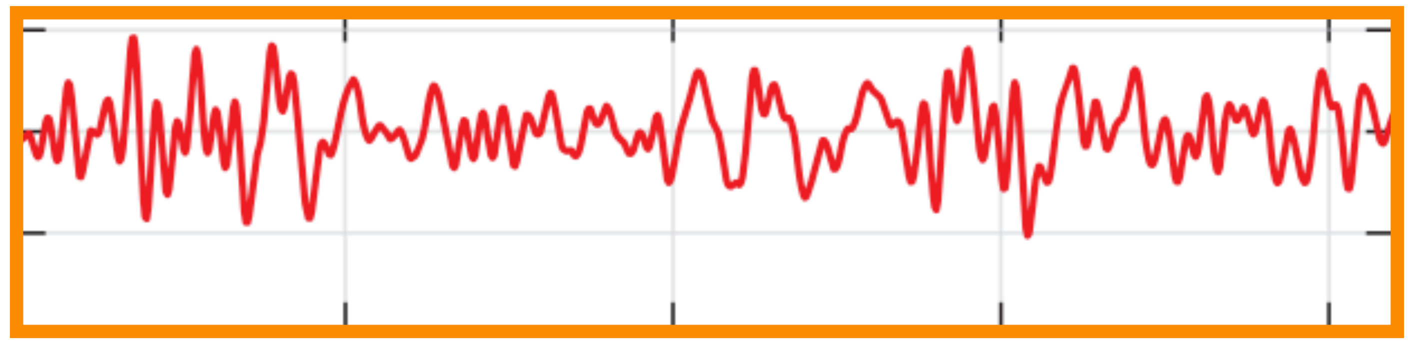
$h(\theta)$



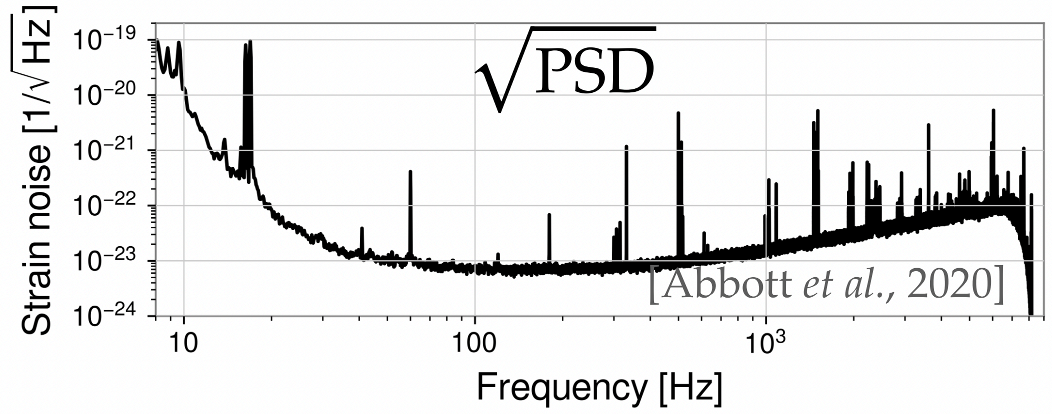
GW signal, from general relativity model

$+$

$n \sim \mathcal{N}(0,S_n)$



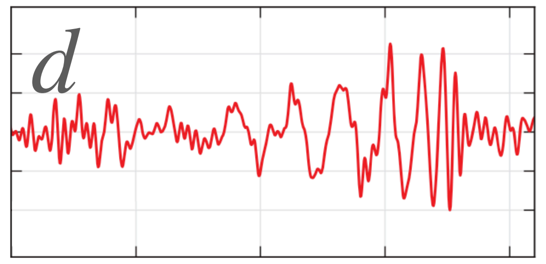
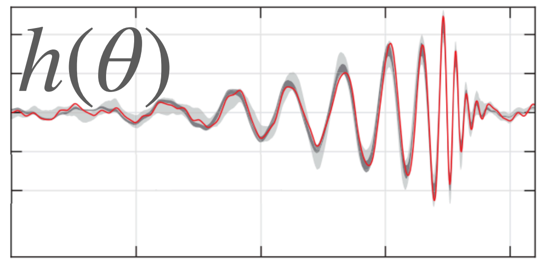
stationary Gaussian detector noise



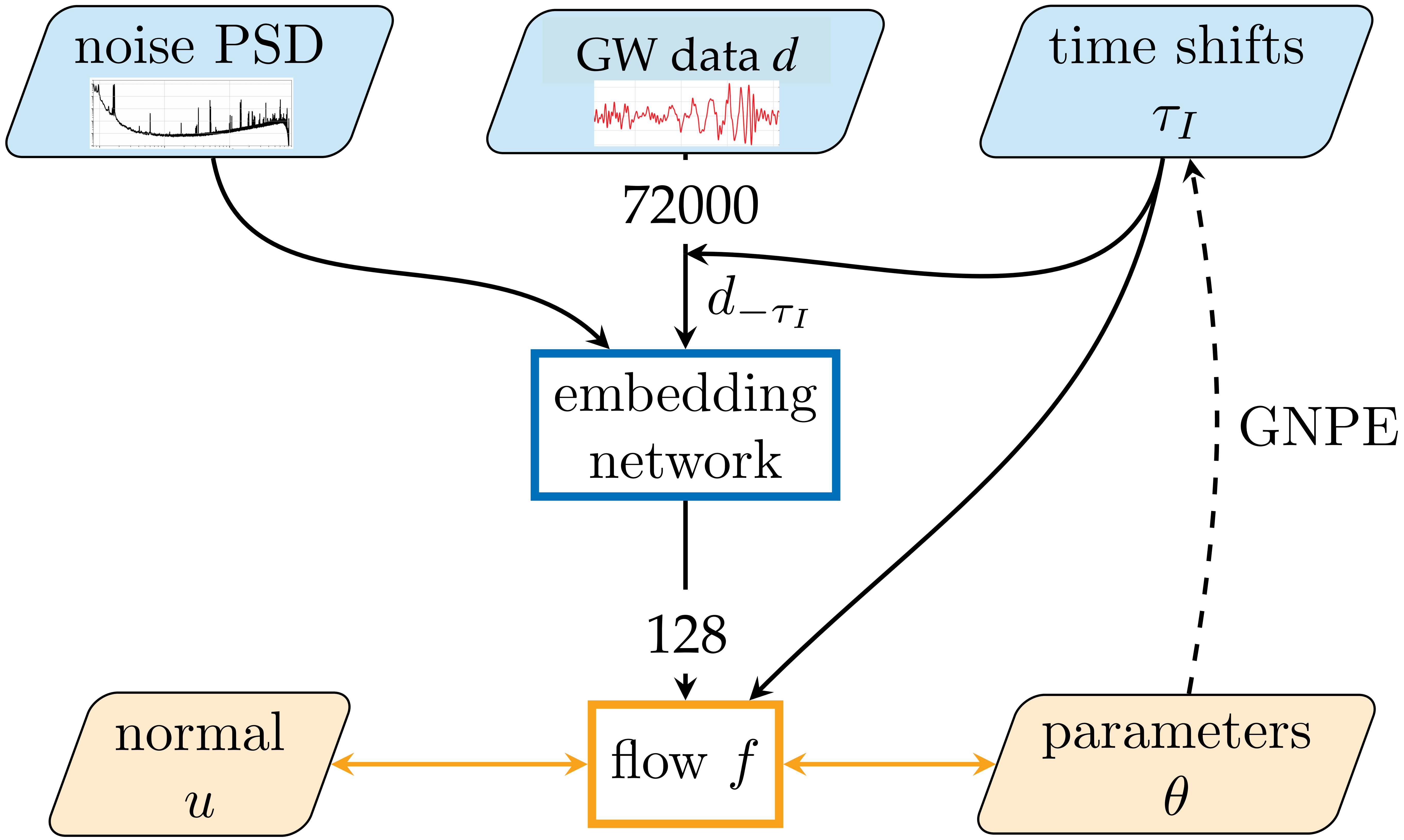
[Abbott *et al.*, 2020]

- Assumption: stationary Gaussian noise

- Tractable likelihood

$p(d|\theta) = \mathcal{N}_{[\mu=0,\sigma^2=\text{PSD}]}\Big($  $-$  $\Big)$

Inference network


noise PSD

GW data d

time shifts τ_I

72000

$d_{-\tau_I}$

embedding network

GNPE

128

normal u

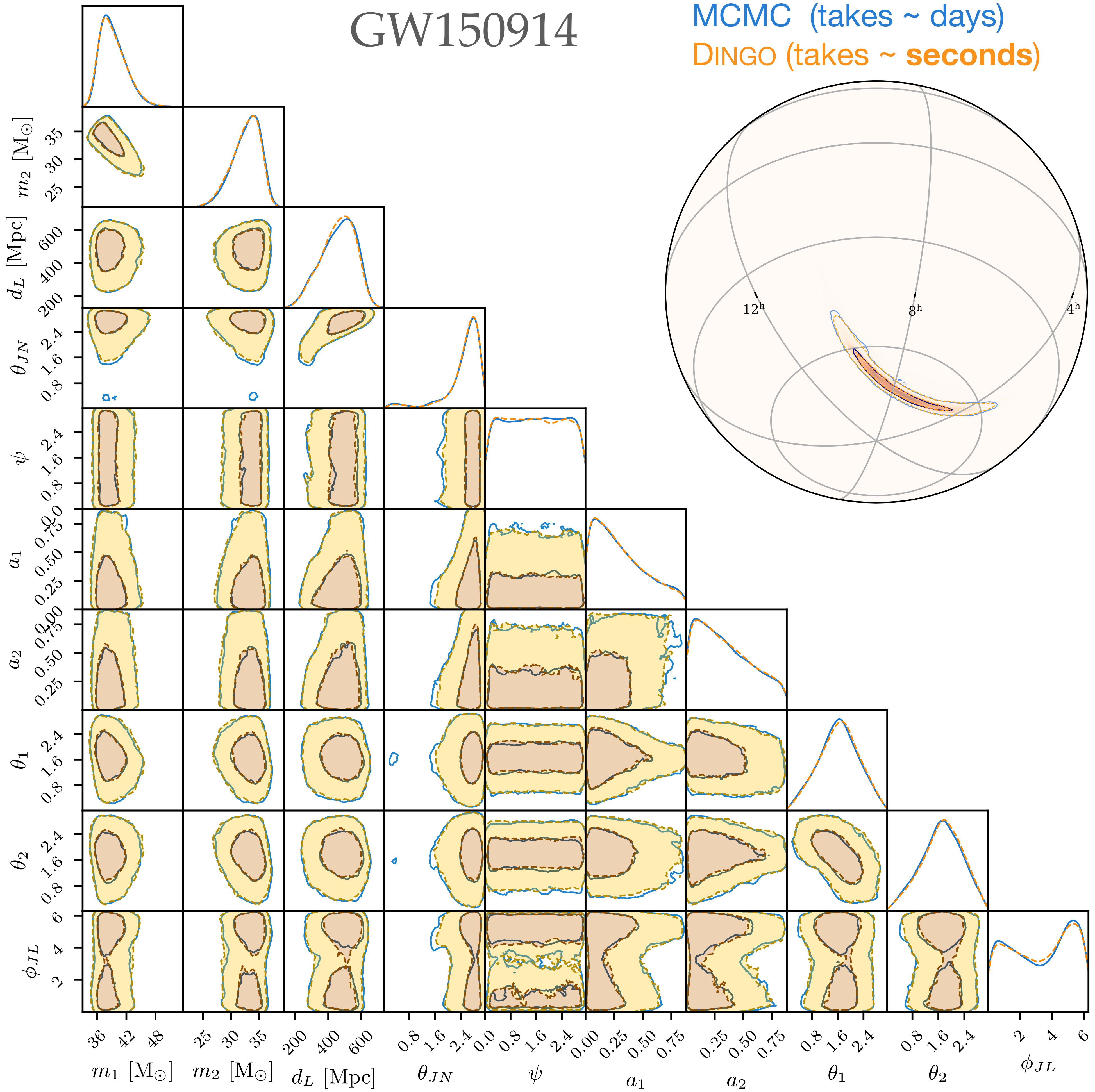
flow f

parameters θ

DINGO: Deep INference for Gravitational-wave Observations

- **Embedding network**
  - 46M learnable parameters
  - compresses data 72K → 128
  - first layer seeded with SVD
- **Neural spline flow** Durkan *et al.*, NeurIPS 2019
  - 94M learnable parameters
- **PSD Conditioning**
  → instant tuning to noise level
- **Training**
  - End-to-end
  - 2B simulations
  - 3 days on A100 with batch size 4096
- **Inference**
  20 seconds per event

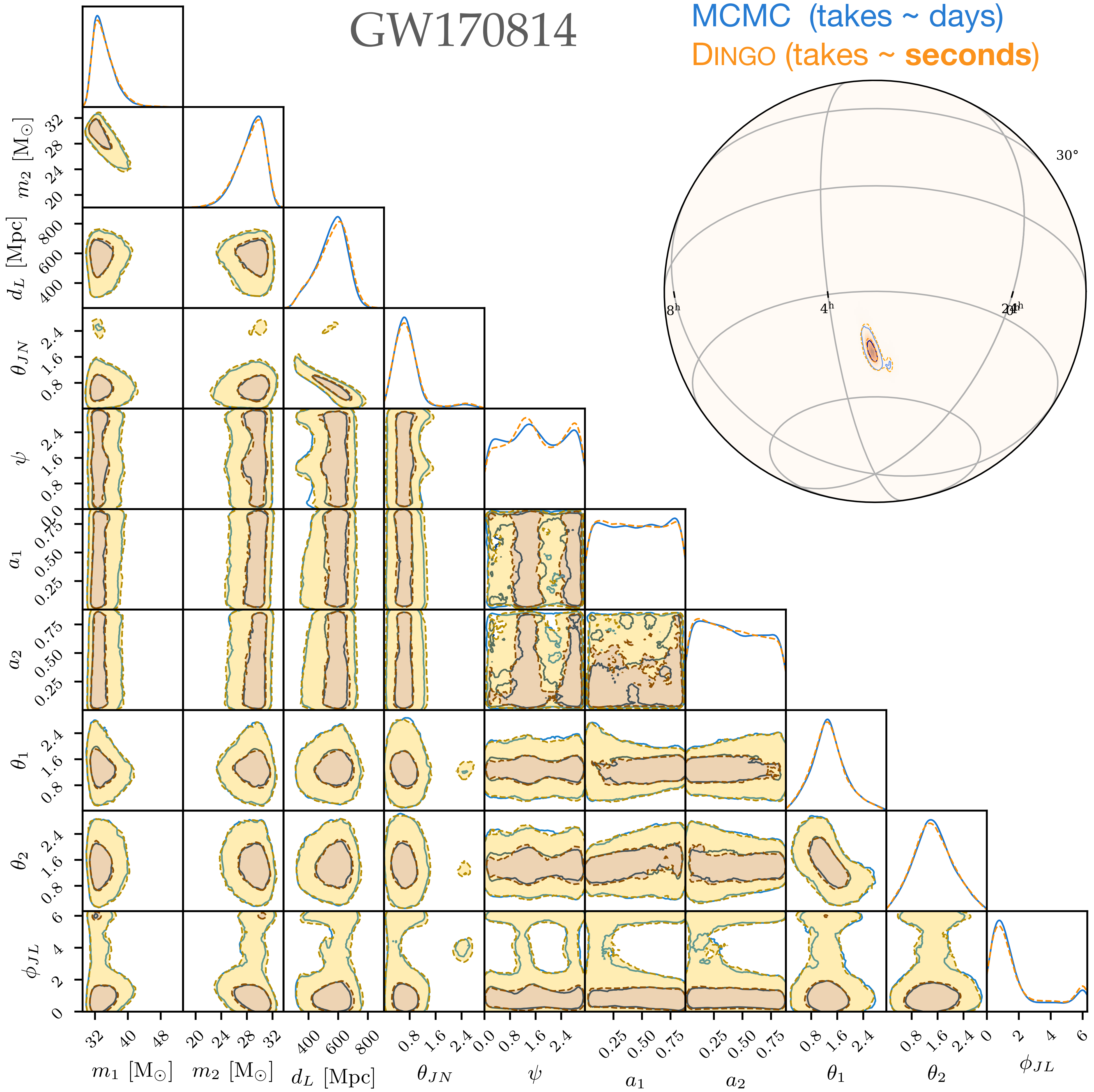
Evaluation on real events



GW150914

MCMC (takes ~ days)

DINGO (takes ~ **seconds**)

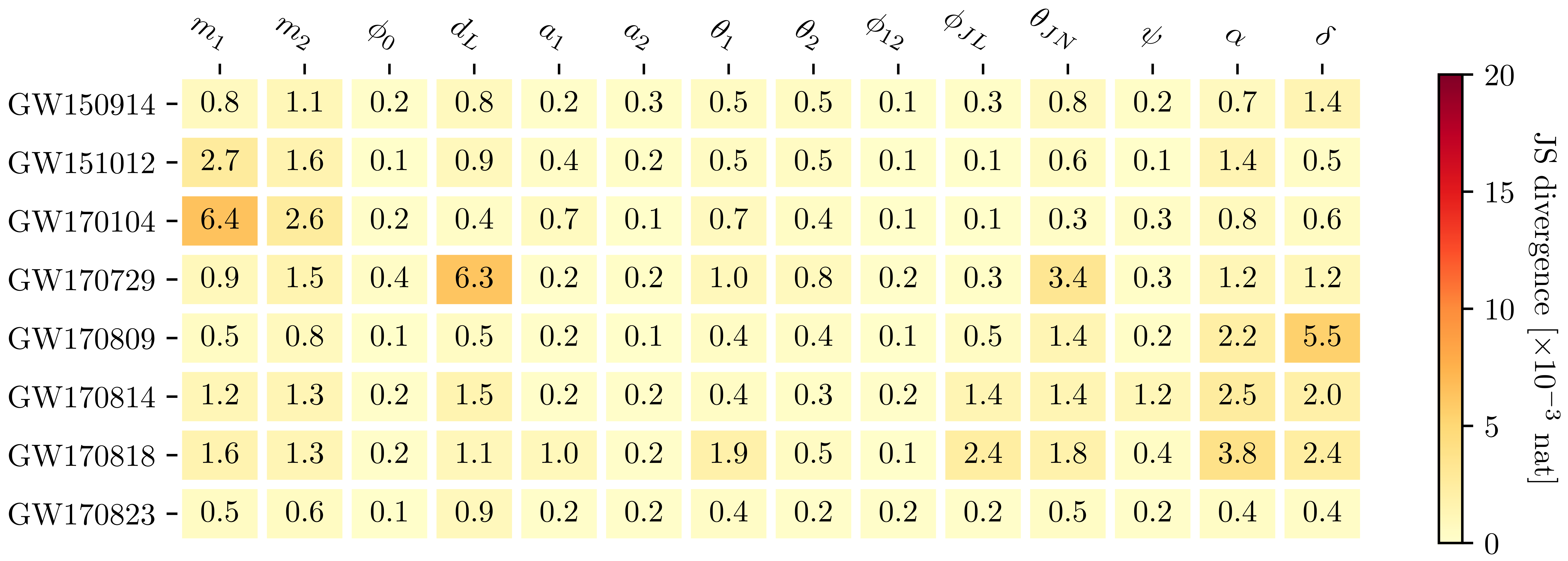


GW170814

MCMC (takes ~ days)

DINGO (takes ~ **seconds**)

Quantitative evaluation on real events



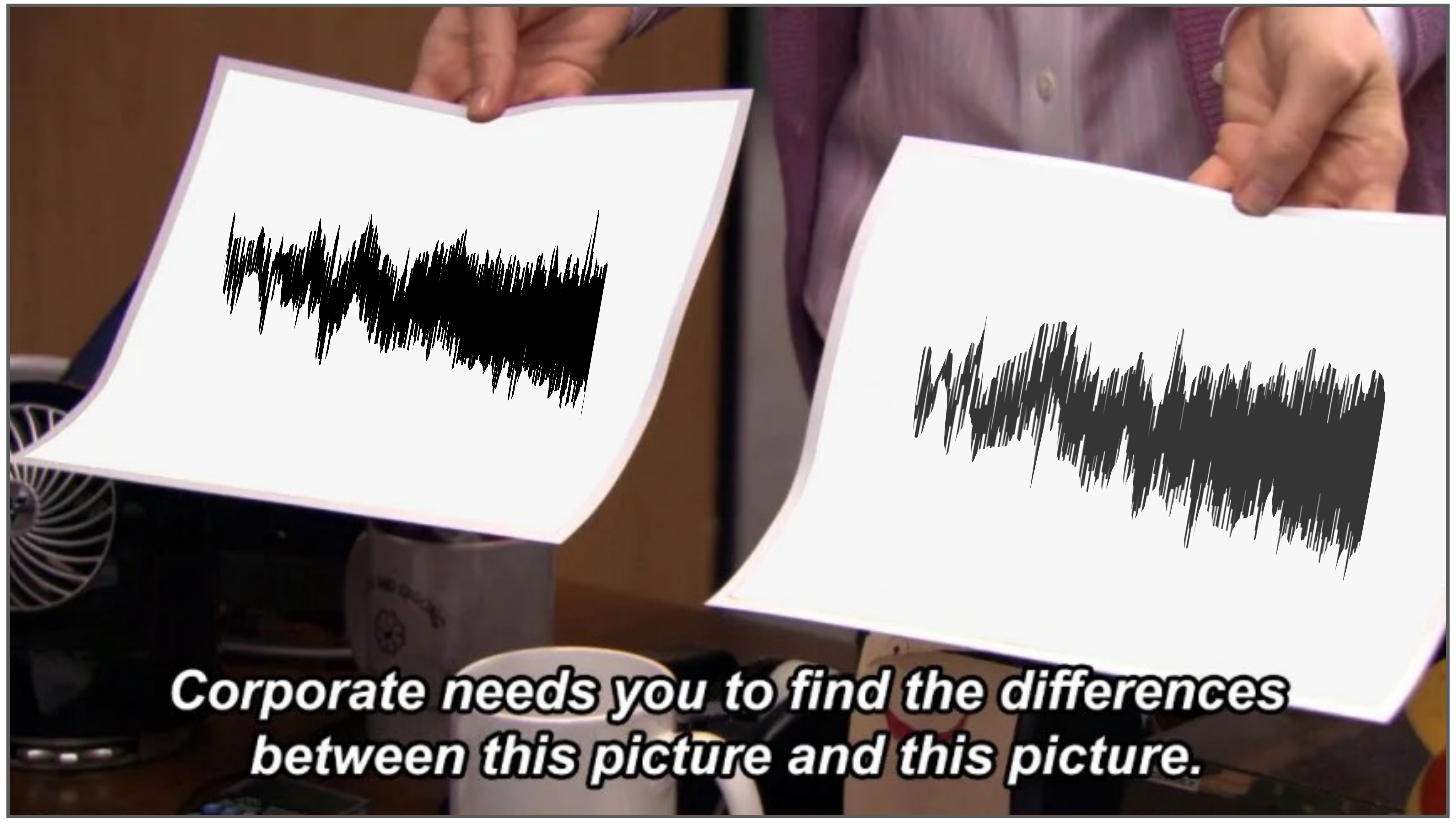
	m_1	m_2	ϕ_0	d_L	a_1	a_2	θ_1	θ_2	ϕ_{12}	ϕ_{JL}	θ_{JN}	ψ	α	δ
GW150914	0.8	1.1	0.2	0.8	0.2	0.3	0.5	0.5	0.1	0.3	0.8	0.2	0.7	1.4
GW151012	2.7	1.6	0.1	0.9	0.4	0.2	0.5	0.5	0.1	0.1	0.6	0.1	1.4	0.5
GW170104	6.4	2.6	0.2	0.4	0.7	0.1	0.7	0.4	0.1	0.1	0.3	0.3	0.8	0.6
GW170729	0.9	1.5	0.4	6.3	0.2	0.2	1.0	0.8	0.2	0.3	3.4	0.3	1.2	1.2
GW170809	0.5	0.8	0.1	0.5	0.2	0.1	0.4	0.4	0.1	0.5	1.4	0.2	2.2	5.5
GW170814	1.2	1.3	0.2	1.5	0.2	0.2	0.4	0.3	0.2	1.4	1.4	1.2	2.5	2.0
GW170818	1.6	1.3	0.2	1.1	1.0	0.2	1.9	0.5	0.1	2.4	1.8	0.4	3.8	2.4
GW170823	0.5	0.6	0.1	0.9	0.2	0.2	0.4	0.2	0.2	0.2	0.5	0.2	0.4	0.4

JS Divergence between MCMC and Dingo

- **Mean JSD = 0.0009 nat** (MCMC vs. MCMC: 0.0007)
- Posteriors regarded as **indistinguishable if JSD $\leq$ 0.0020 nat** → fulfilled for 90% of marginals
- Deviation on similar level as deviation between different stochastic samplers (LALInference vs. bilby)

II. Symmetries

[1] Dax et al., Real-Time Gravitational Wave Science with Neural Posterior Estimation, Phys. Rev. Lett. **127**, 241103 (2021)
[2] Dax et al., Group equivariant neural posterior estimation, ICLR 2022



*up to time shifts

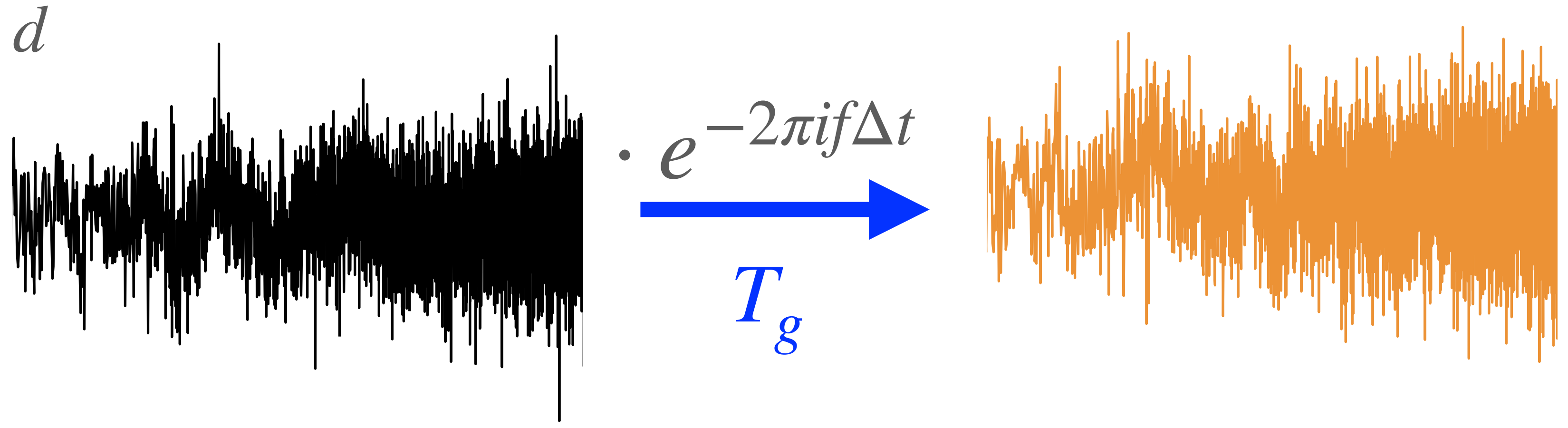
Symmetries in GW inference

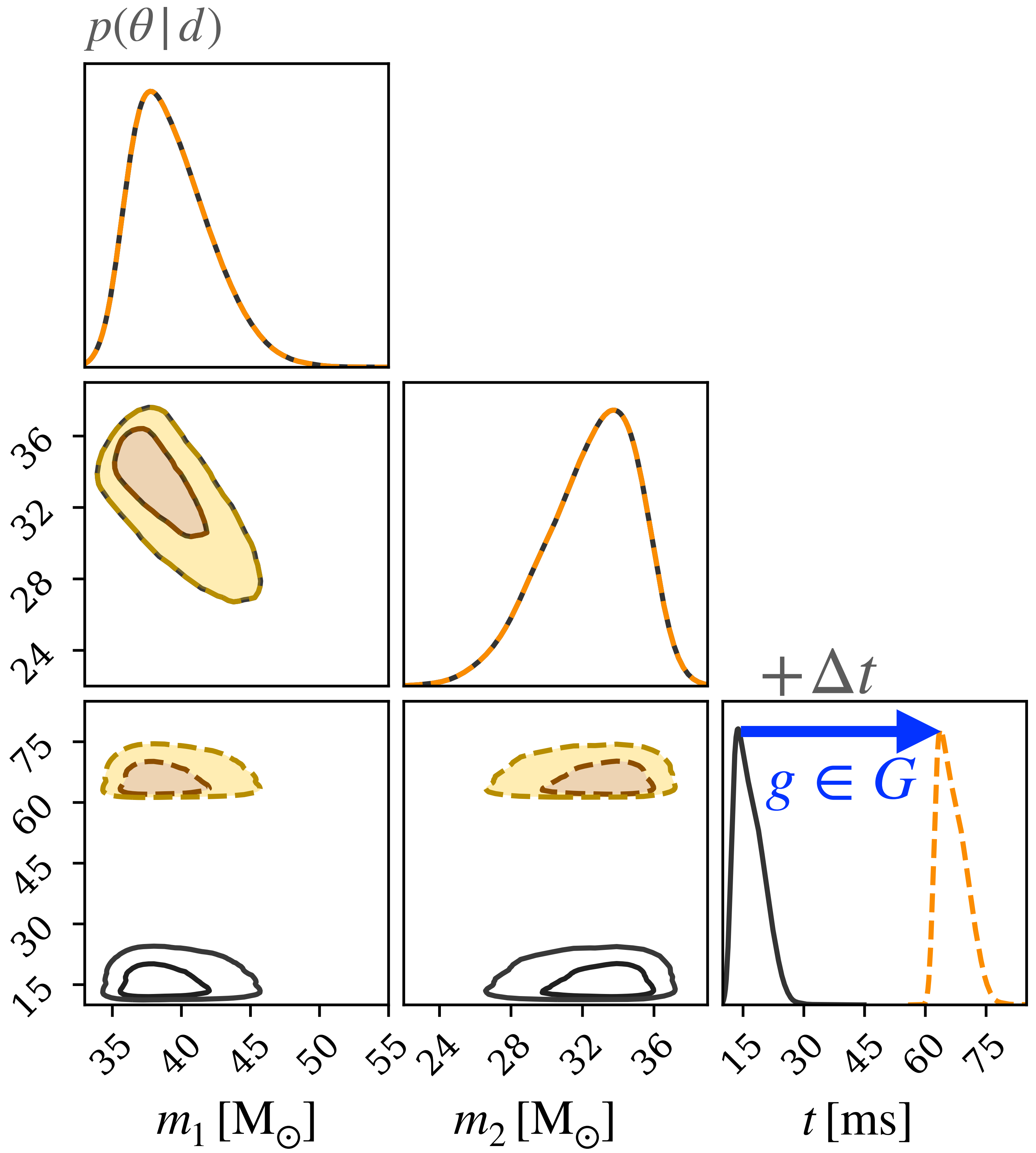
- Equivariance (covariance) under time shift

$$p(\theta\,|\,d) = p(g\theta\,|\,T_g d)\,|\det J_g| \qquad \forall g \in G$$

- NPE learns such symmetries from simulation data
  ⇒ requires network and training capacity

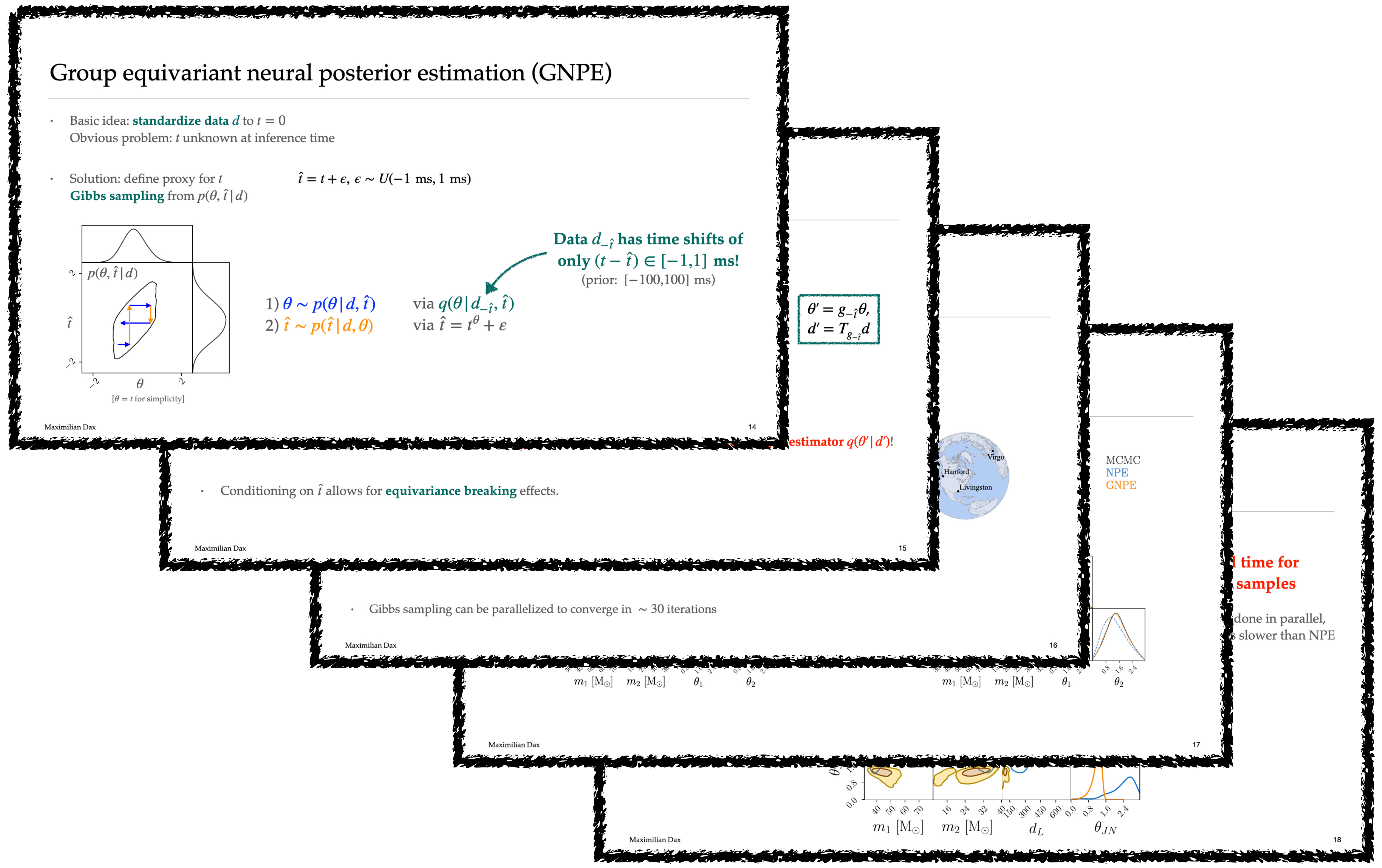
- How can we **enforce such symmetries**?





Group equivariant neural posterior estimation (GNPE)

GNPE simplifies inverse problems with symmetries



Group equivariant neural posterior estimation (GNPE)

- Basic idea: **standardize data** d to $t = 0$
  Obvious problem: t unknown at inference time
- Solution: define proxy for t $\quad \hat{t} = t + \epsilon,\ \epsilon \sim U(-1\ \text{ms}, 1\ \text{ms})$
  Gibbs sampling from $p(\theta, \hat{t}\,|\,d)$

[$\theta = t$ for simplicity]

1) $\theta \sim p(\theta\,|\,d, \hat{t})$ $\quad$ via $q(\theta\,|\,d_{-\hat{t}}, \hat{t})$
2) $\hat{t} \sim p(\hat{t}\,|\,d, \theta)$ $\quad$ via $\hat{t} = t^{\theta} + \varepsilon$

Data $d_{-\hat{t}}$ has time shifts of only $(t - \hat{t}) \in [-1,1]$ ms!
(prior: $[-100,100]$ ms)

- Conditioning on $\hat{t}$ allows for **equivariance breaking** effects.

- Gibbs sampling can be parallelized to converge in ~ 30 iterations

If you are interested talk to me during the breaks or find the paper at

arXiv:2111.13139

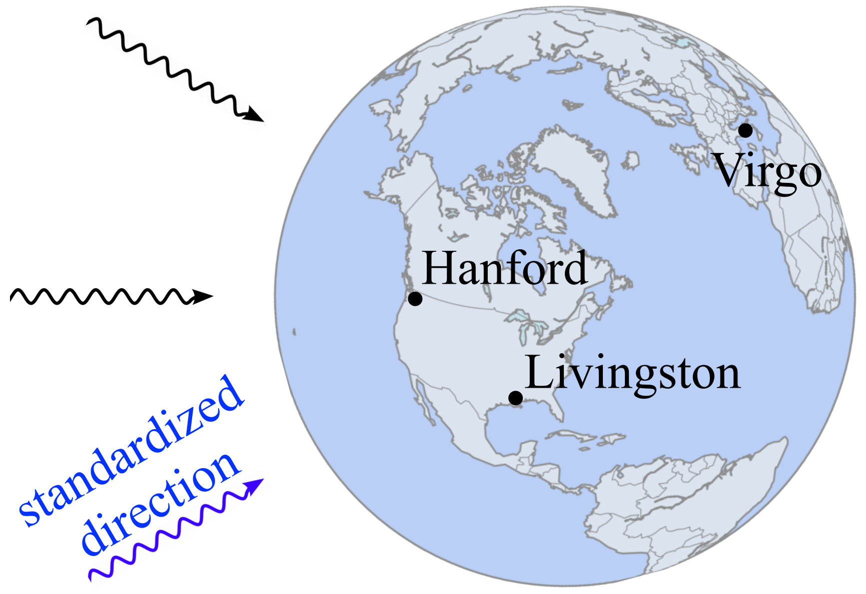
GNPE (cont'd)

- GNPE is a generic way to integrate (approximate) **symmetries** into ***conditional* density estimation**.

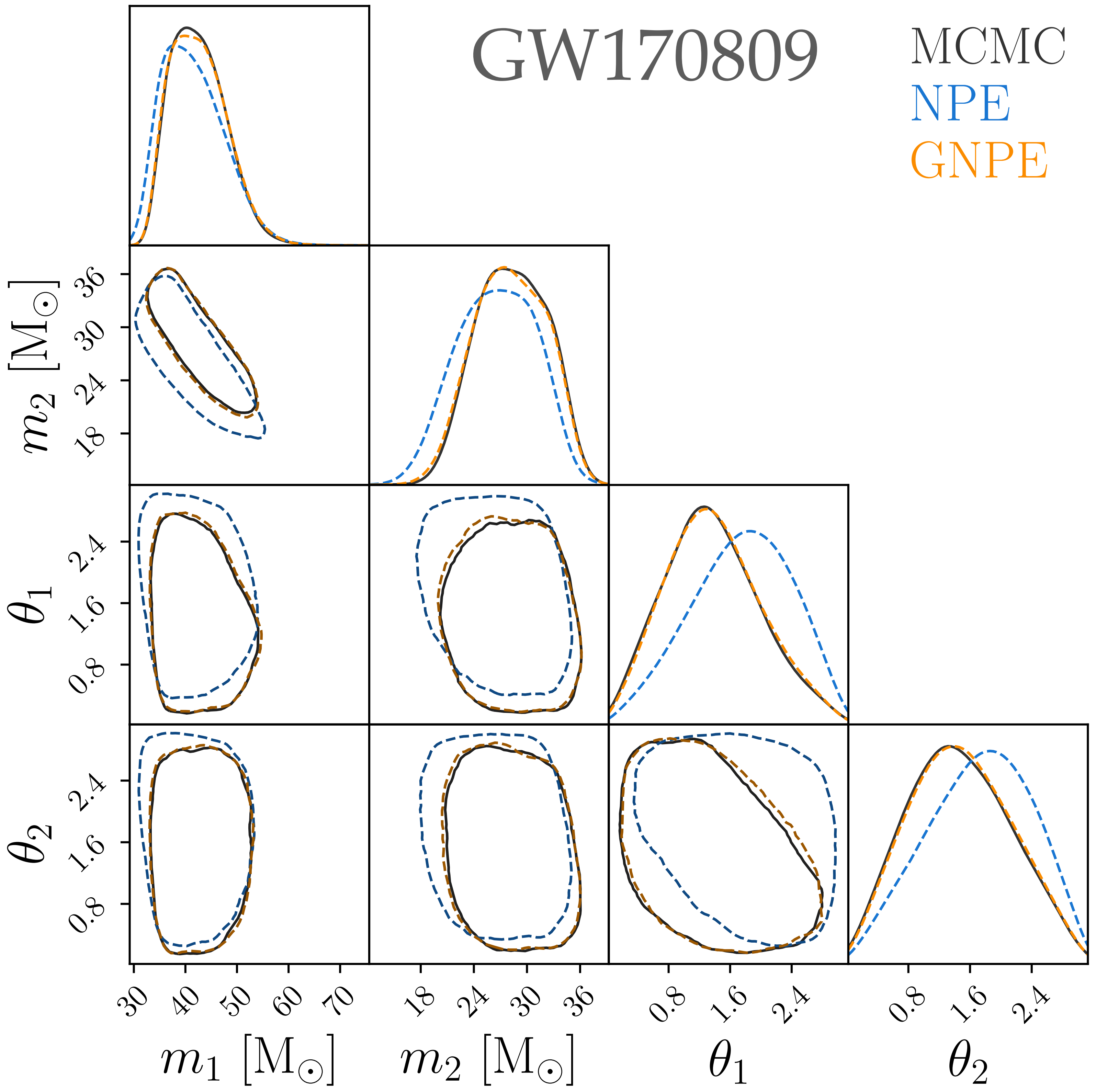
- **Exact equivariances:** enforced by construction
- **Approximate equivariances:** used to constrain data distribution

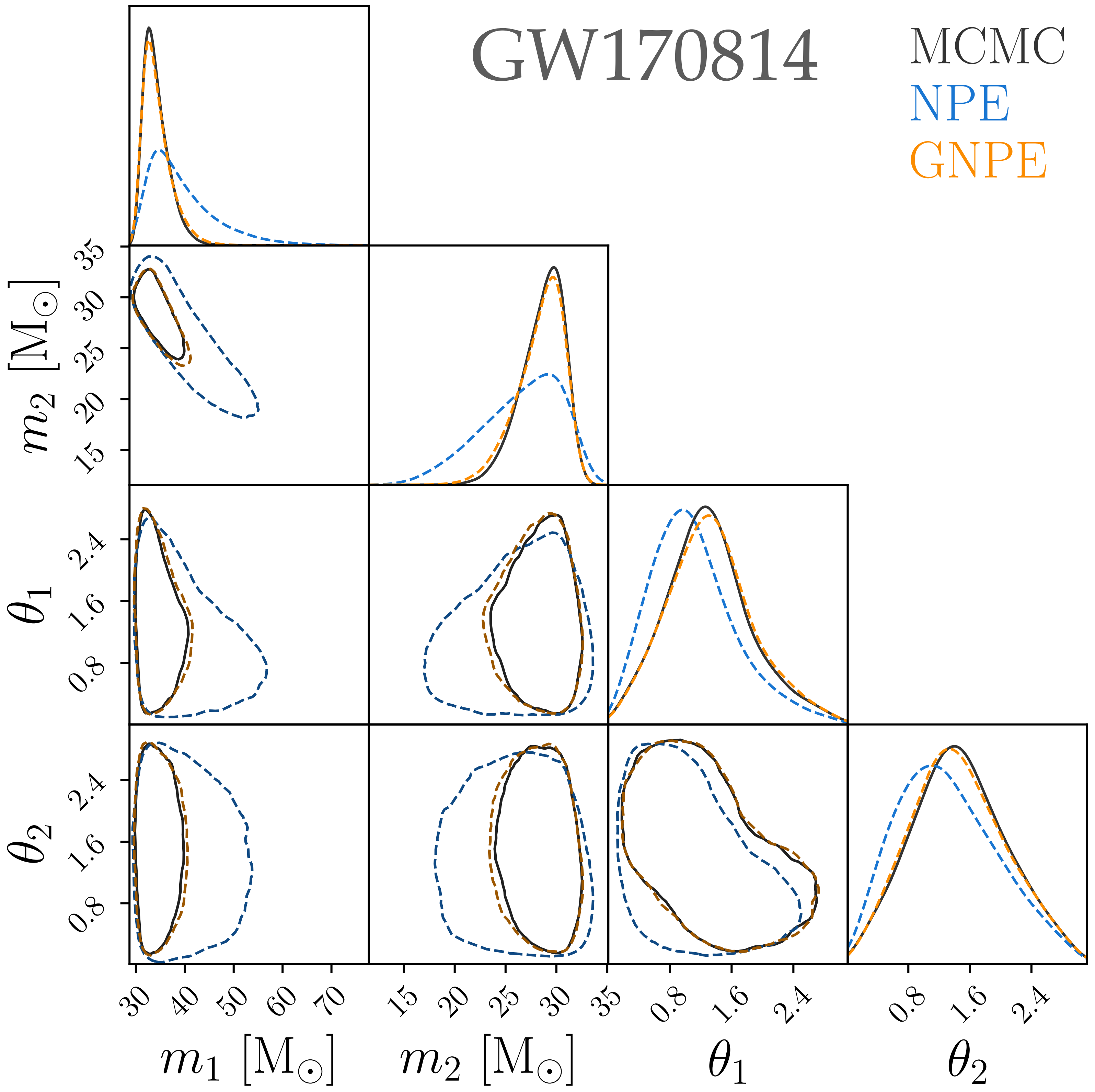
- Architecture independent
  ⇒ compatible with domain-specific architectures

- GW inference
  - Exact **symmetry under time shift**
  - Approximate **symmetry under rotation** of incident direction

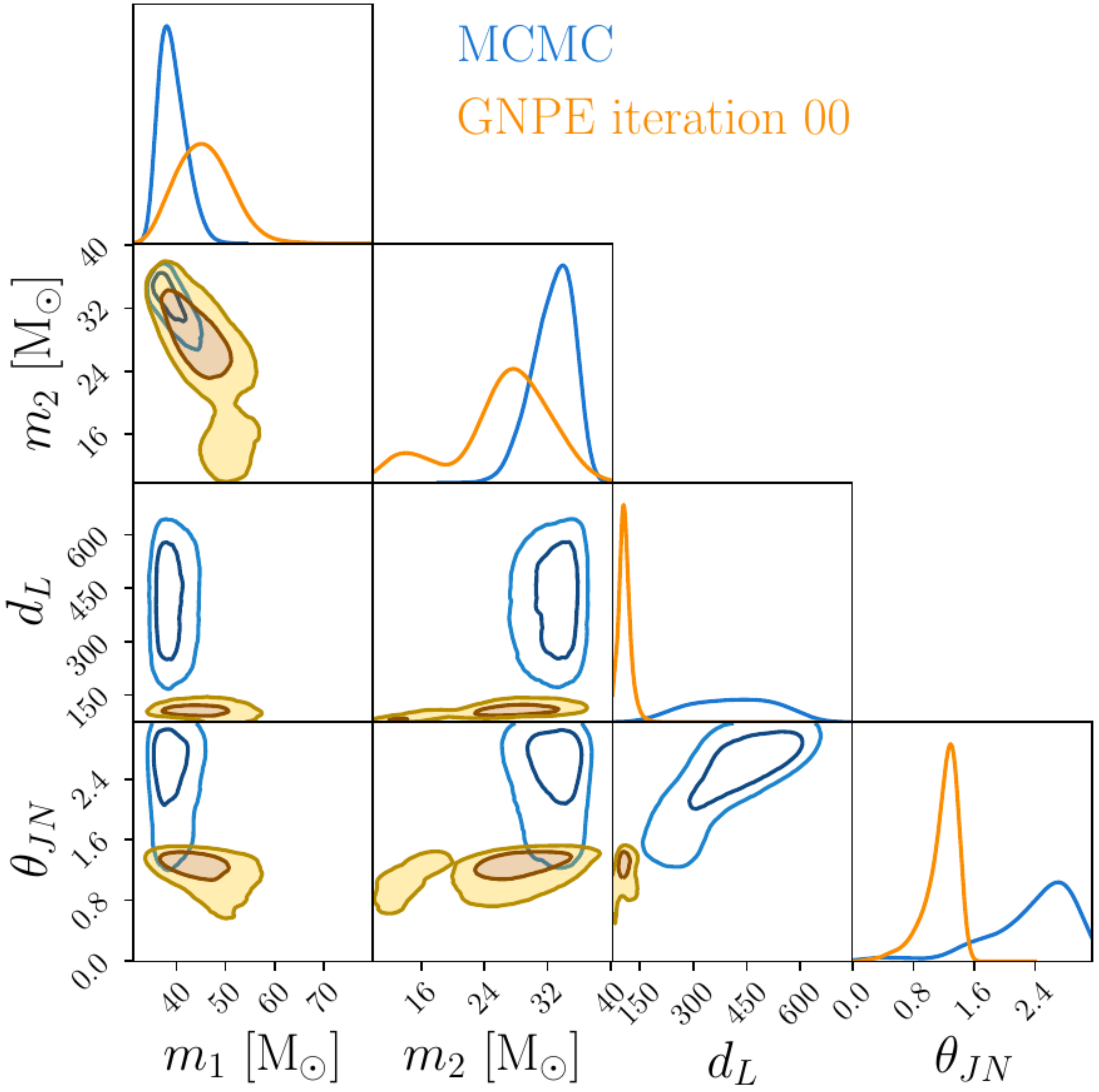


NPE vs. GNPE


GW170809
MCMC
NPE
GNPE
m_2 [M$_\odot$]
θ_1
θ_2
m_1 [M$_\odot$]
m_2 [M$_\odot$]
θ_1
θ_2
30 40 50 60 70
18 24 30 36
0.8 1.6 2.4
0.8 1.6 2.4


GW170814
MCMC
NPE
GNPE
m_2 [M$_\odot$]
θ_1
θ_2
m_1 [M$_\odot$]
m_2 [M$_\odot$]
θ_1
θ_2
30 40 50 60 70
15 20 25 30 35
0.8 1.6 2.4
0.8 1.6 2.4

GNPE iterations



Real time

50k samples

Sampling done in parallel,
$\sim N_{it.}$ times slower than NPE

III. Verification through Importance Sampling

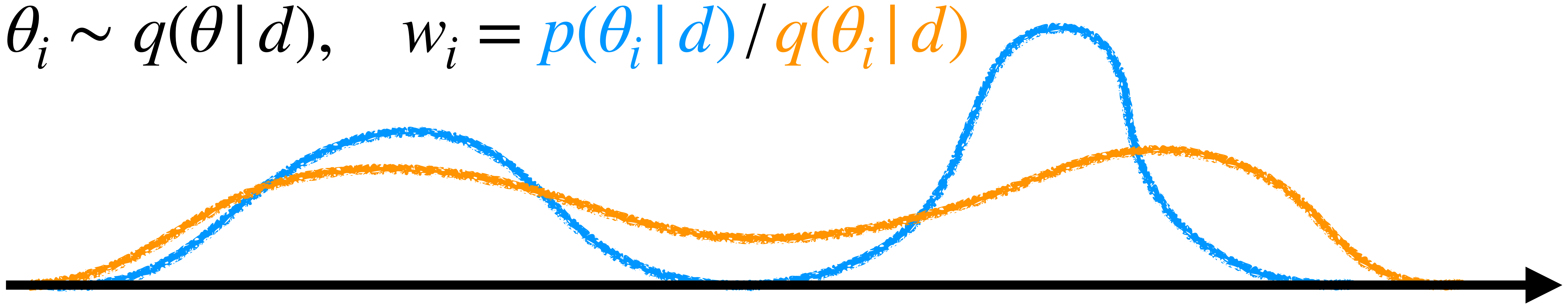
[3] Dax*, Green* et al., Neural Importance Sampling for Rapid and Reliable Gravitational-Wave Inference, Phys.Rev.Lett. **130**, 171403 (2023)

[from reddit]

Importance sampling

- Express target distribution $p(\theta \,|\, d)$ via proposal distribution $q(\theta \,|\, d)$, generate weighted samples from $p(\theta)$:

$$\theta_i \sim q(\theta \,|\, d), \quad w_i = p(\theta_i \,|\, d) \,/\, q(\theta_i \,|\, d)$$



$\Rightarrow$ **Produce exact results by reweighting DINGO samples with likelihood**

- Importance sampling requires **supp**$(p) \subseteq$ **supp**(q) $[p \equiv p(\theta \,|\, d), \quad q \equiv q(\theta \,|\, d)]$

- DINGO minimizes forward KL-divergence $\mathrm{KL}(p \,|\, q)$ which is probability-mass covering!
  supp$(p) \nsubseteq$ **supp**$(q) \Rightarrow$ **diverging DINGO loss**

verified empirically

Importance sampling diagnostics and evidence

- Effective sample size n_{eff} related to variance of the weights.
  Sample **efficiency ϵ quantifies the quality** of the DINGO proposal distribution.

$$n_{\text{eff}} = \frac{\left(\Sigma_i w_i\right)^2}{\Sigma_i \left(w_i^2\right)} \qquad \epsilon = \frac{n_{\text{eff}}}{n} \in (0,1]$$

⇒ don’t need ground truth posterior for verification

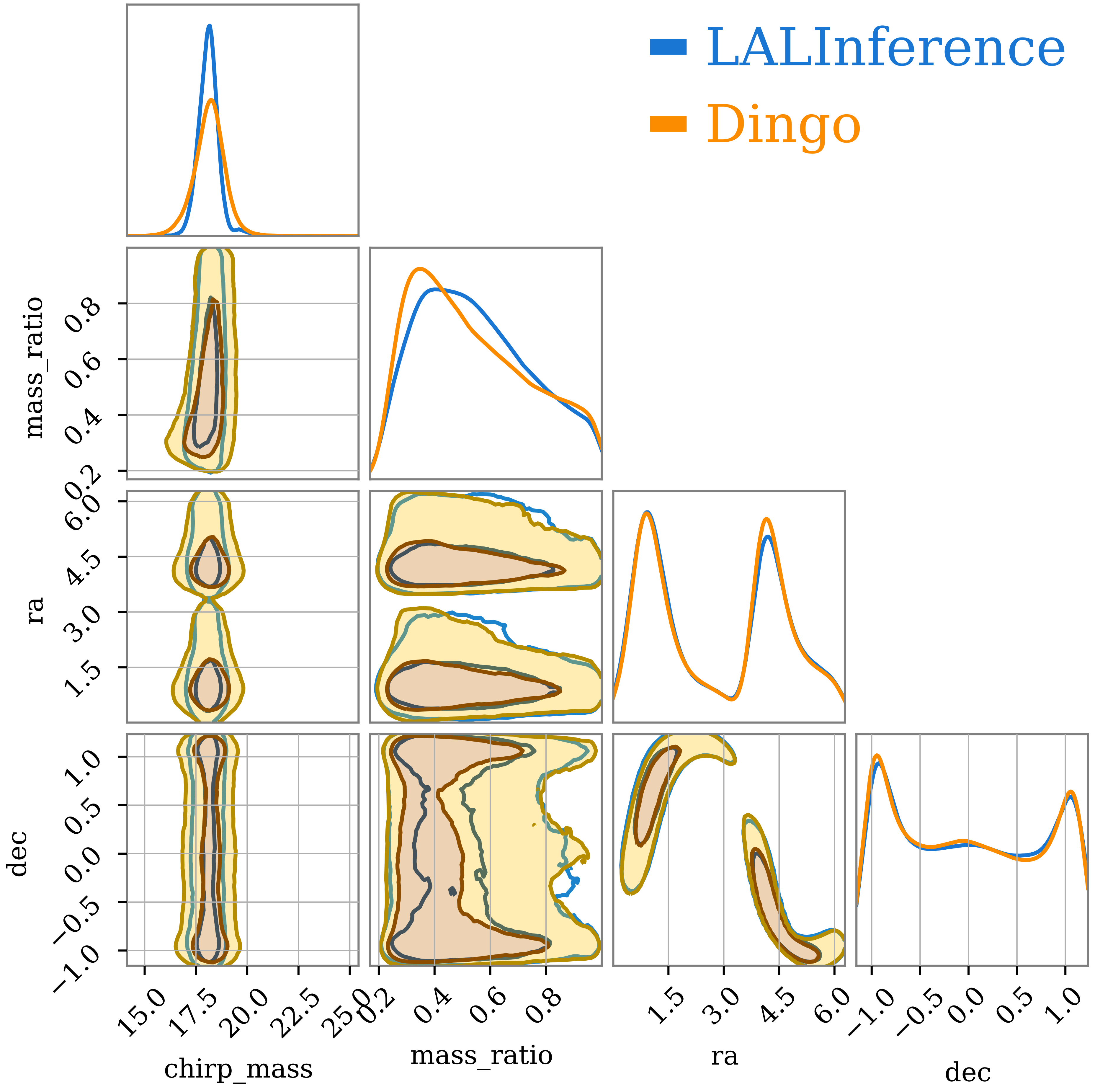
- **Bayesian evidence** related to normalisation of weights.

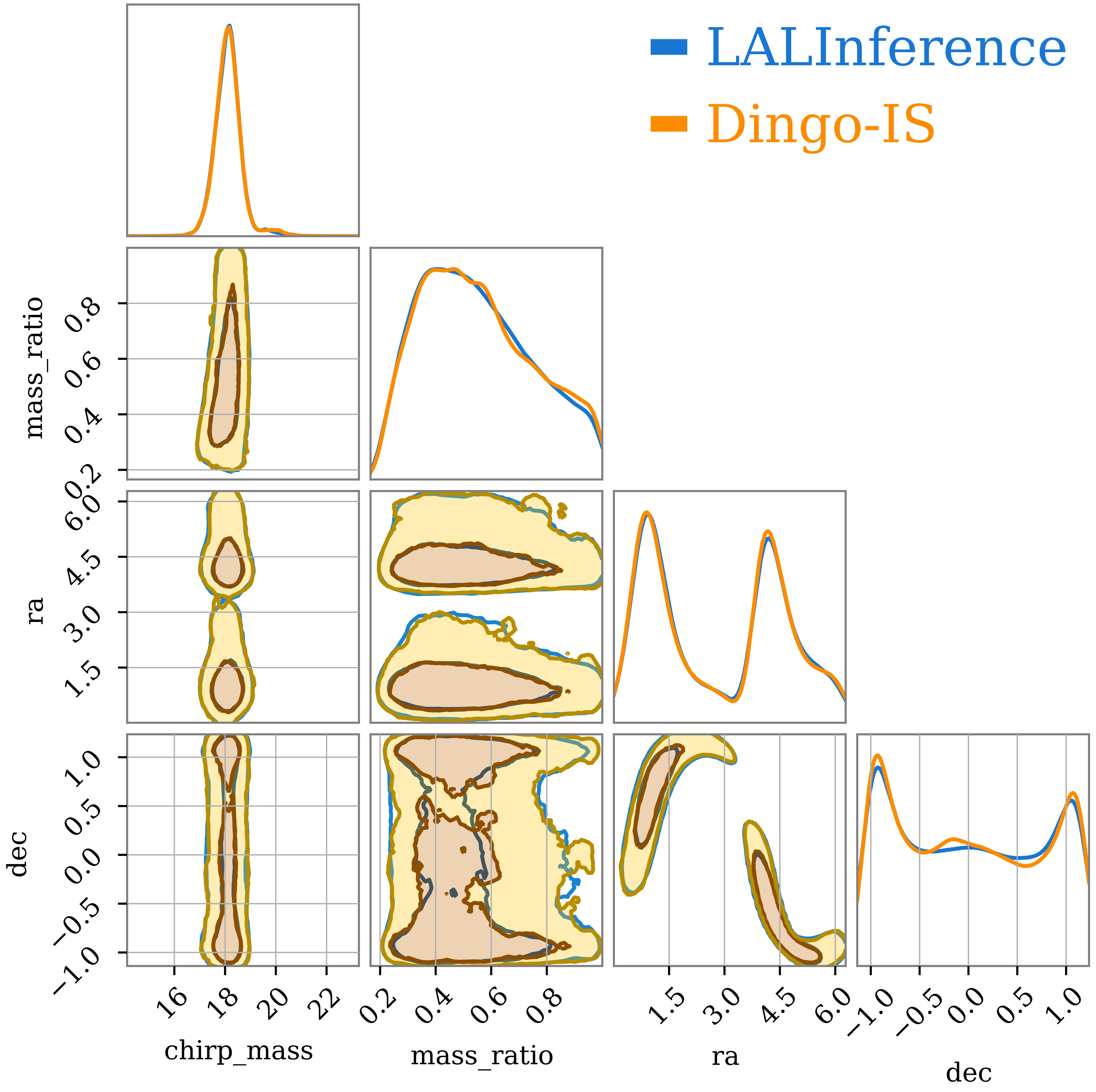
$$p(d) = \frac{1}{n} \sum_i w_i \qquad \sigma_{\log p(d)} = \sqrt{\frac{(1-\epsilon)}{(n \cdot \epsilon)}}$$

⇒ Unbiased estimate, variance scales with $1/n$

DINGO-IS: qualitative results

GW151012,
IMRPhenomXPHM





⇒ Even when DINGO results are off, IS results match LALInference well

DINGO-IS: quantitative results

- Evaluation on 42 real GW events (third LVK observing run)

- DINGO-IS IS **~100x more sample efficient** than MCMC
  - Median $\epsilon = 10.9\%$ for GW model IMRPhenomXPHM
  - Median $\epsilon = 4.4\%$ for GW model SEOBNRv4PHM

[$\epsilon = 36.8\%$ for IMRPhenomPv2]

- Estimate Bayesian evidence **10x more precise** than nested sampling

- **High-accuracy inference for SEOBNRv4PHM** (MCMC takes ~ months)
  ⇒ IS only way of verifying results

- Low sample efficiency flags with low ϵ: **adversarial attacks, glitches, failures of GW model, …**

Summary

- DINGO is a **fast and accurate inference method for GW** data analysis
  - Inference in seconds to minutes
  - Independently verified with importance sampling
  - 10x precision improvement for Bayesian evidence
  - Reviewed for PE at LIGO

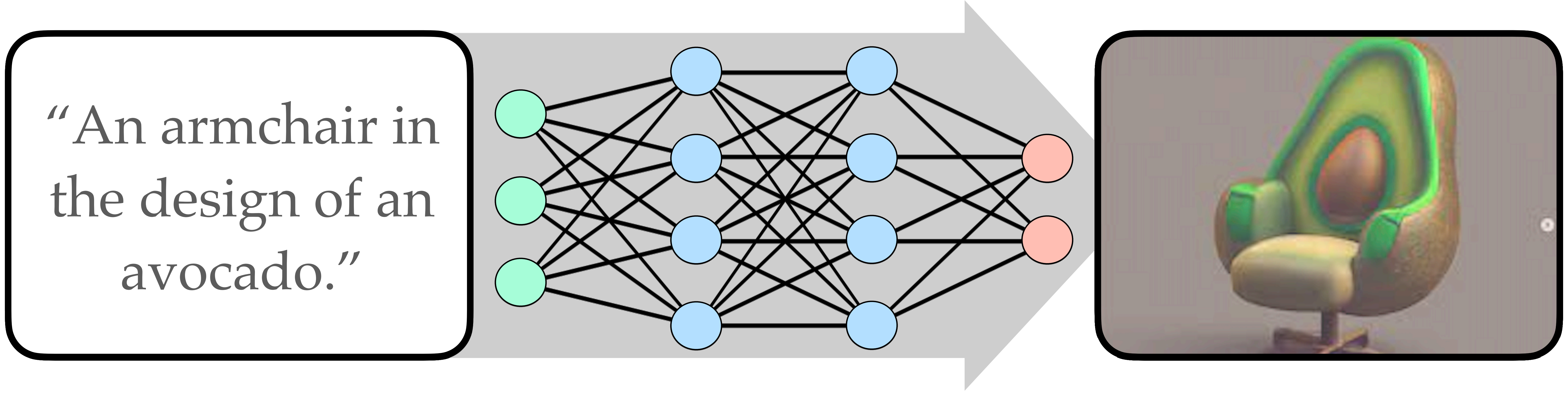
github.com/dingo-gw/dingo

- Tested for various BBH models (IMRPhenomPv2, IMRPhenomXPHM, SEOBNRv4PHM); **Working on BNS** and NSBH

- Working on **various science cases** (eccentricity, population studies, testing GR)

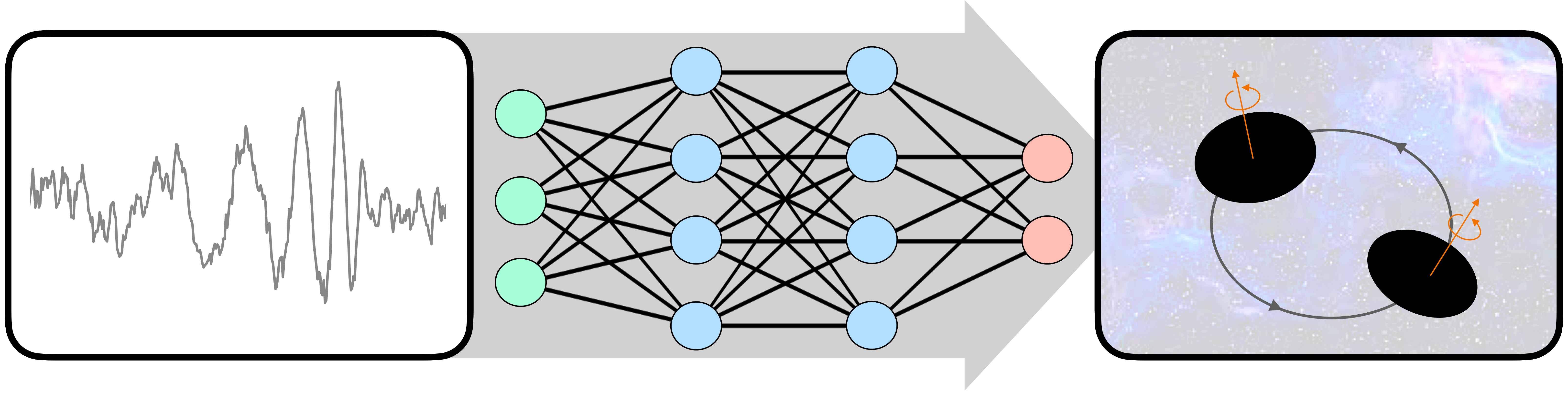
- If you need **fast GW inference** or encounter a seemingly intractable/expensive **inverse problems** in your research, feel free to reach out :)

The bigger picture



"An armchair in the design of an avocado."

⇒ Synthesize individual samples from distribution



⇒ Extract summary statistics to describe distribution