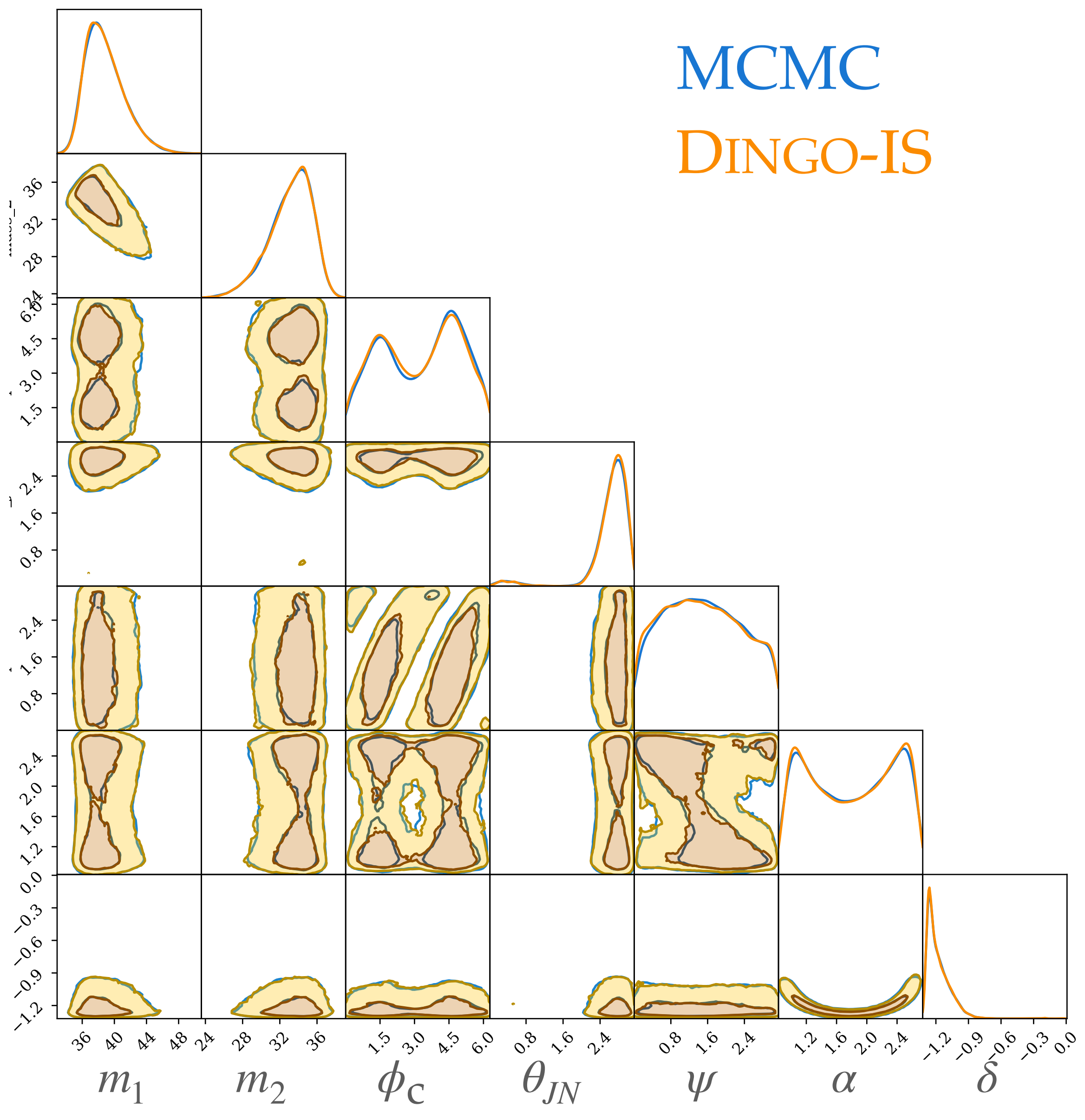


# Real-Time Gravitational Wave Science with DINGO

---

Maximilian Dax mdax@tue.mpg.de
Max Planck Institute for Intelligent Systems, Tübingen

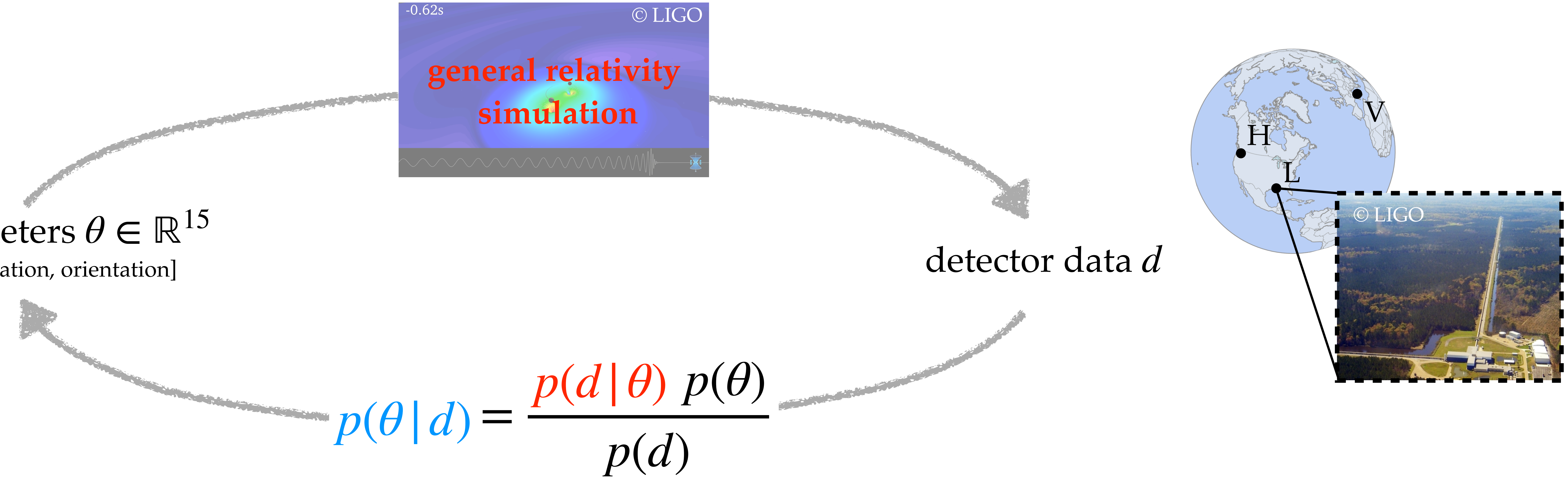
MAX PLANCK INSTITUTE
FOR INTELLIGENT SYSTEMS

[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*, Green* et al., Neural Importance Sampling for Rapid and Reliable Gravitational-Wave Inference, Preprint (2022)

With **Green**, Gair, Wildberger, Pürrer, Deistler, Macke, Buonanno, Schölkopf

# GW inference

Black hole mergers emit gravitational waves (GWs), which encode information about the source



source parameters $\theta \in \mathbb{R}^{15}$

[masses, spins, location, orientation]

detector data $d$

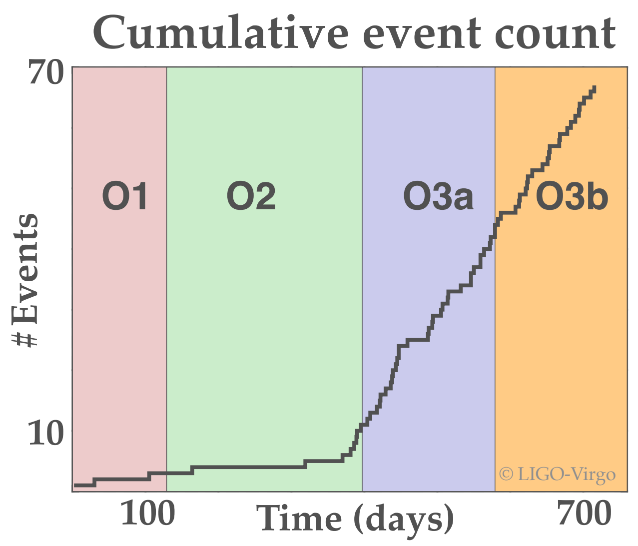
$$p(\theta|d) = \frac{p(d|\theta)\, p(\theta)}{p(d)}$$

GW inference: sample $\theta \sim p(\theta|d)$ to **characterize astrophysical source**.

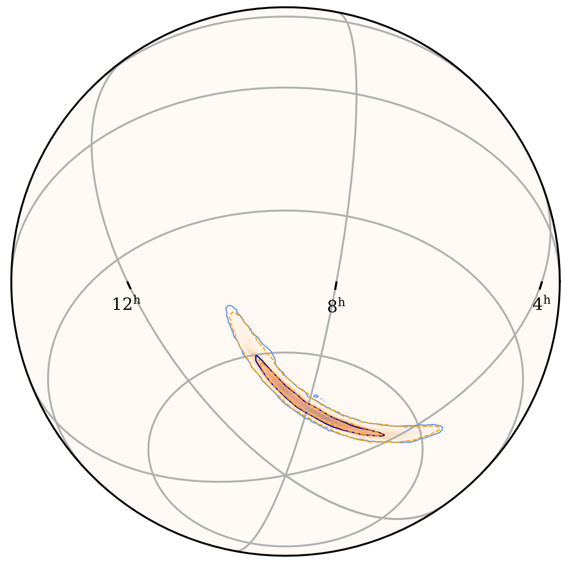
MCMC requires $\sim 10^6$ likelihood evaluations

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

- **Speed** (~days)

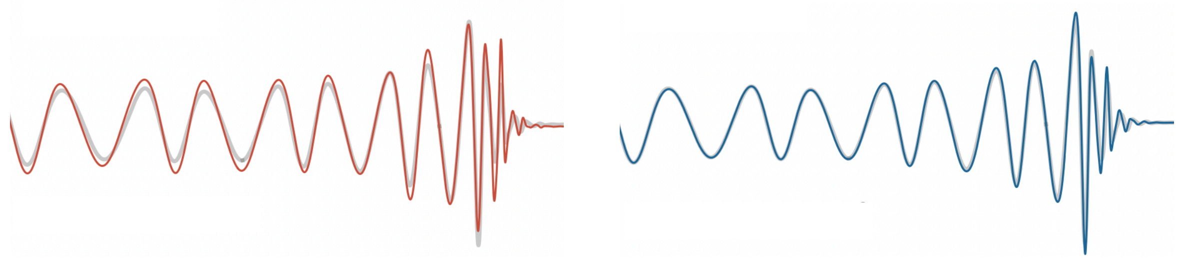


computationally costly
(increasing event rate!)

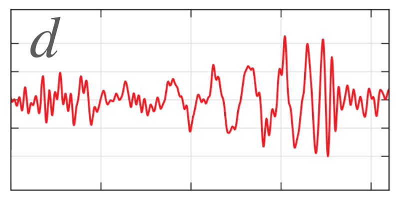


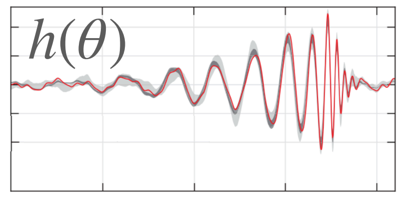
no fast localization
for e.m. follow-up

- **Accuracy**



don't scale to
high-quality GW models

$\mathcal{N}_{0,\mathrm{PSD}}\Bigg($ 

$-$ 

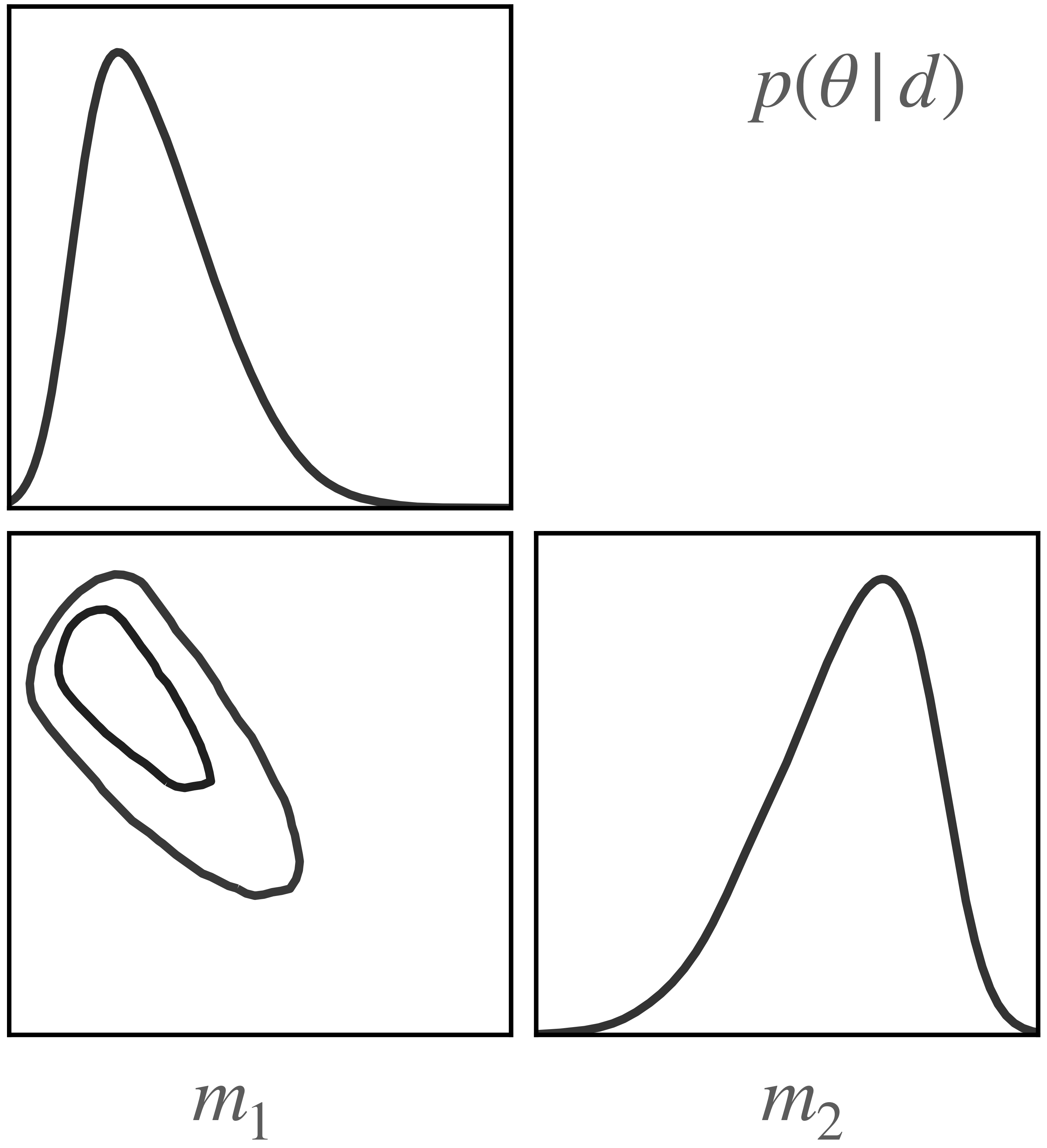
$\Bigg)$

require tractable likelihood
⇒ need noise model

# I. Real-Time GW Inference

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

# Amortized inference

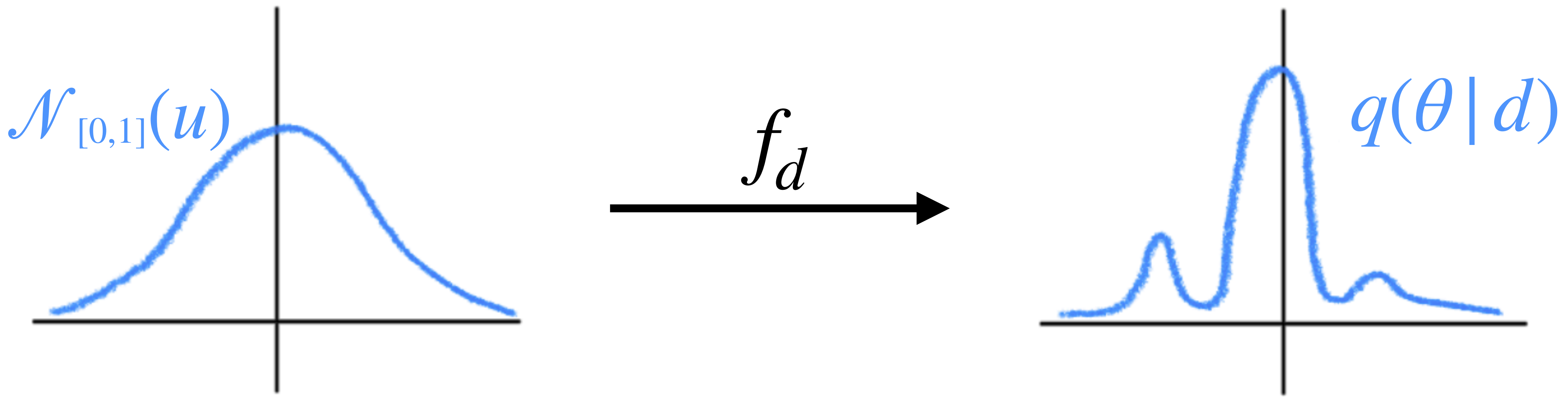


Idea: Learn parameterization of $p(\theta|d)$ using the GW model (**expensive**), use if for inference of all future events (**cheap**).

1. Expressive **parameterization** $q(\theta|d)$ (15D density, conditional on $d$)
2. **Training strategy** s.t. $q(\theta|d) = p(\theta|d)\ \forall d$

# Normalizing flows [Rezende & Mohamed, 2015]

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

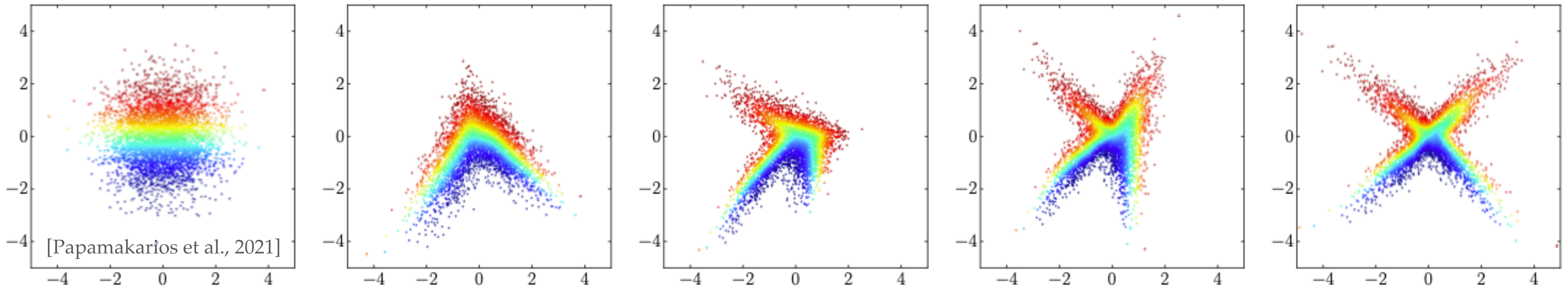


$$\theta = f_d(u), \quad u \sim \mathcal{N}_{[0,1]}(u)$$

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

$f_d$ parameterized using neural network
with learnable parameters $\phi$

- Flexible $f_d$ achieved by composition of simple transforms



[Papamakarios et al., 2021]

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

# Neural posterior estimation (NPE) [Papamakarios & Murray, 2016]

- Minimize

average over $d$

$D_{\mathrm{KL}}$ 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, 2016]

- Minimize

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

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

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

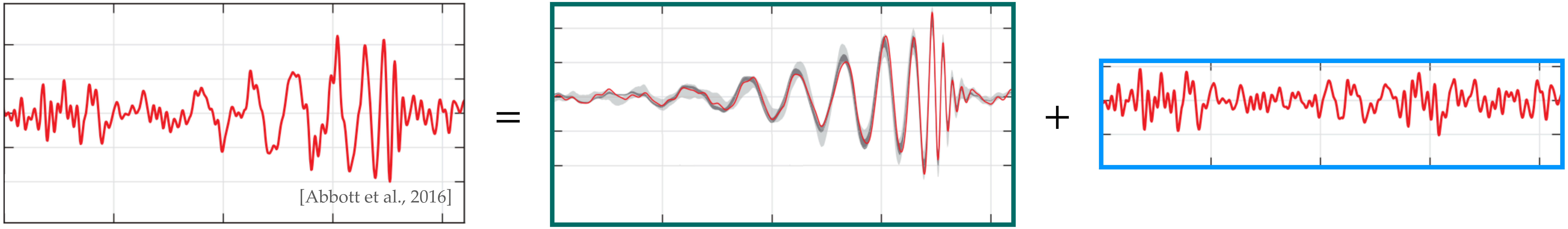
- Minimization of $D_{\mathrm{KL}}$ + arbitrarily expressive ⇒ **perfect recovery of posterior**

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

# GW likelihood

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

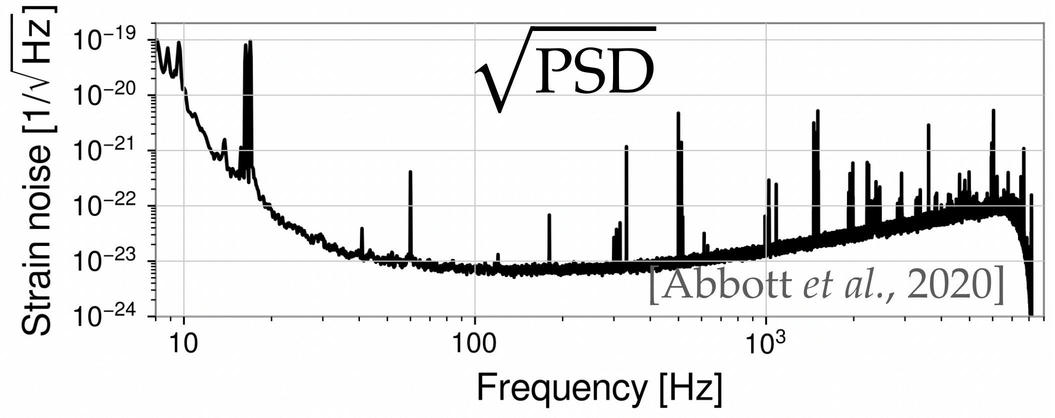
$d \sim p(d|\theta)$ $\qquad$ $h(\theta)$ $\qquad$ $n \sim \mathcal{N}(0,S_n)$



[Abbott et al., 2016]

GW signal, from general relativity model

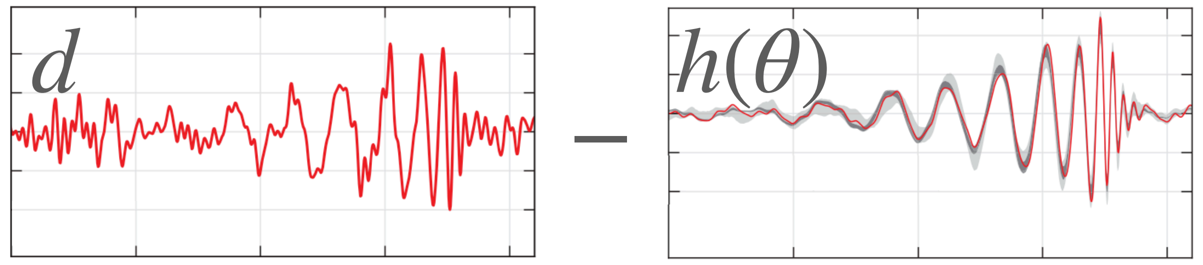
stationary Gaussian detector noise



[Abbott *et al.*, 2020]

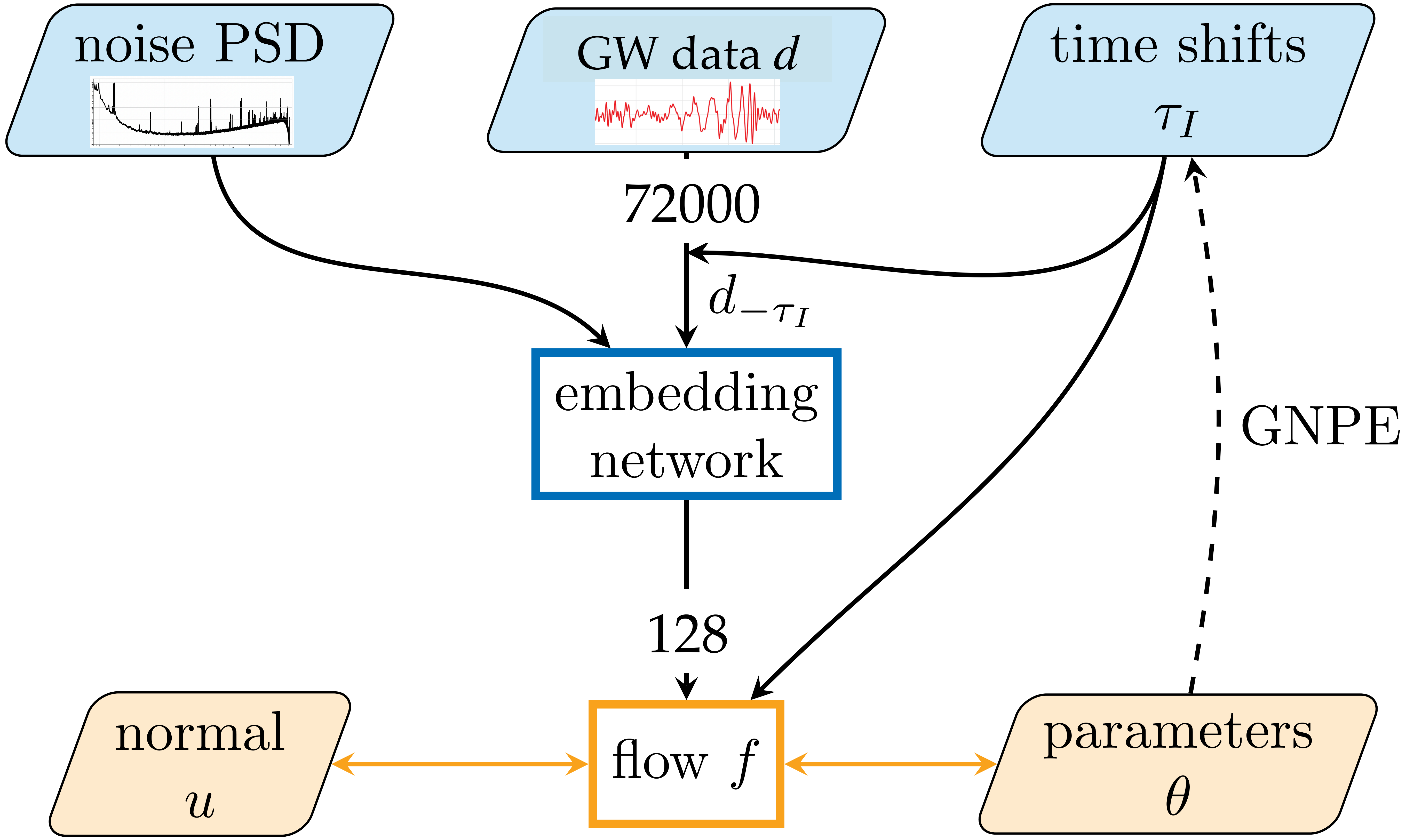
- Assumption: stationary Gaussian noise

- Tractable likelihood



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

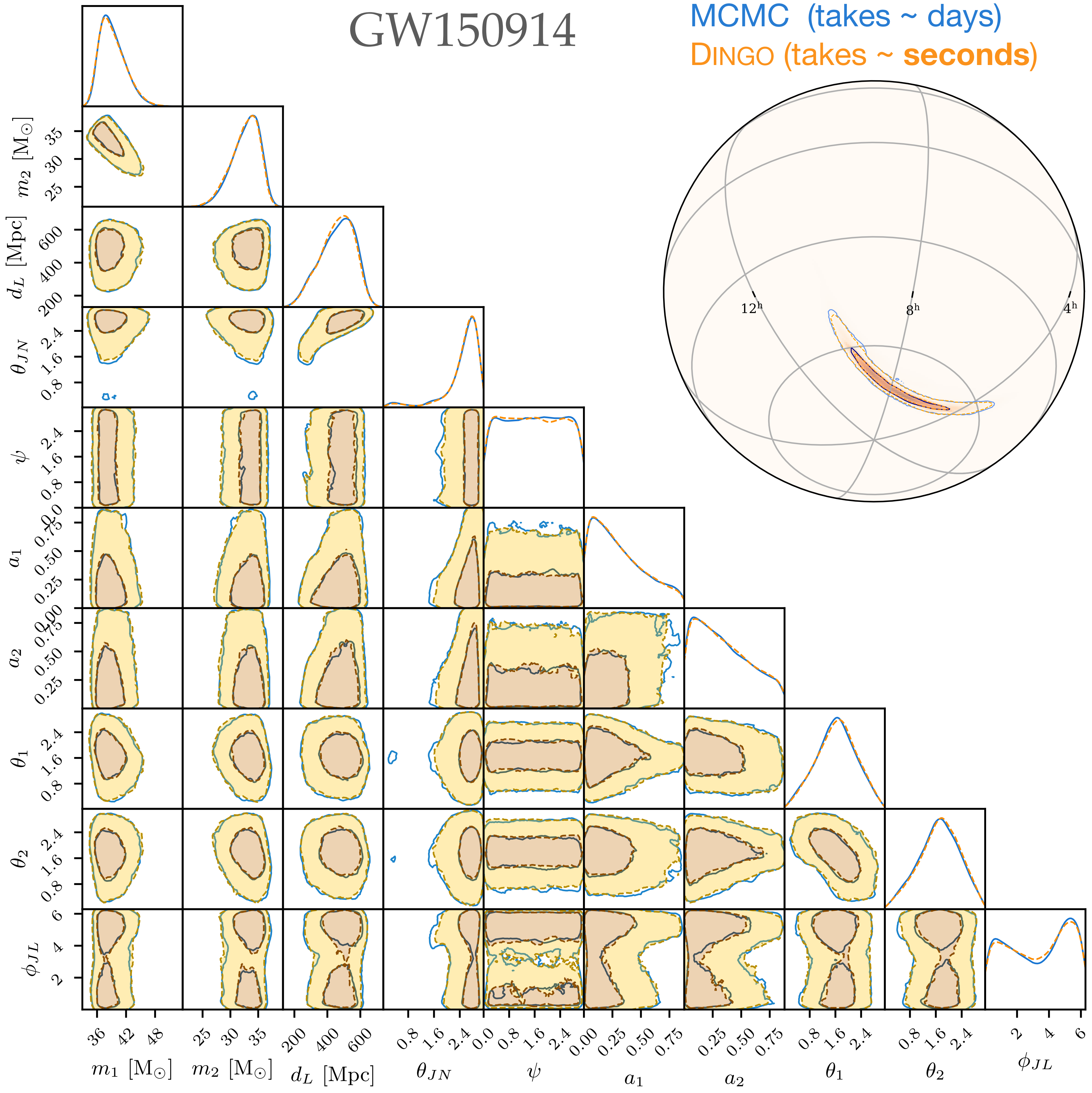
# Inference network

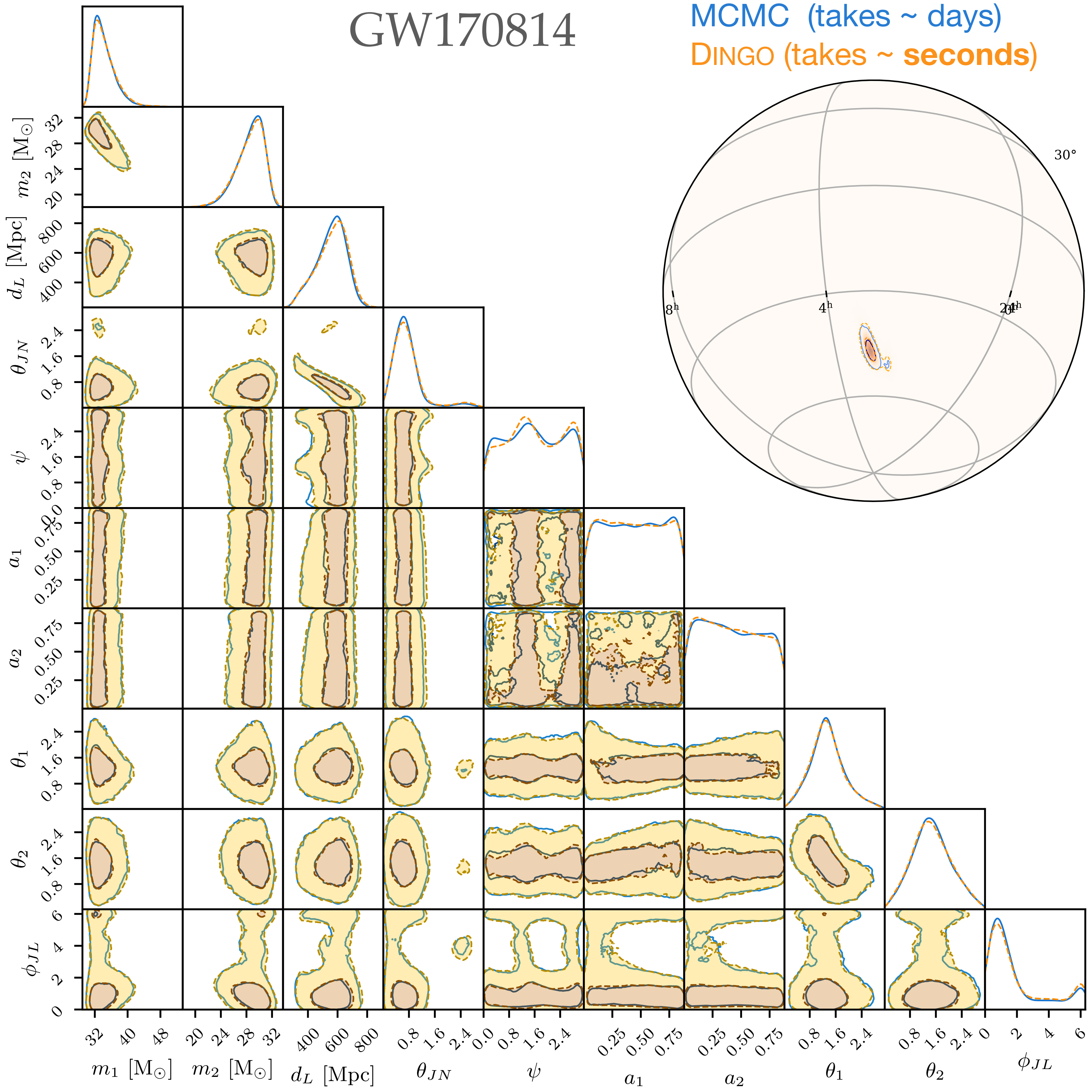


**DINGO: Deep INference for Gravitational-wave Observations**

- **Embedding network**
  - 46M learnable parameters
  - compresses data 72K → 128
  - first layer seeded with SVD
- **Neural spline flow** [Durkan *et al.*, 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**

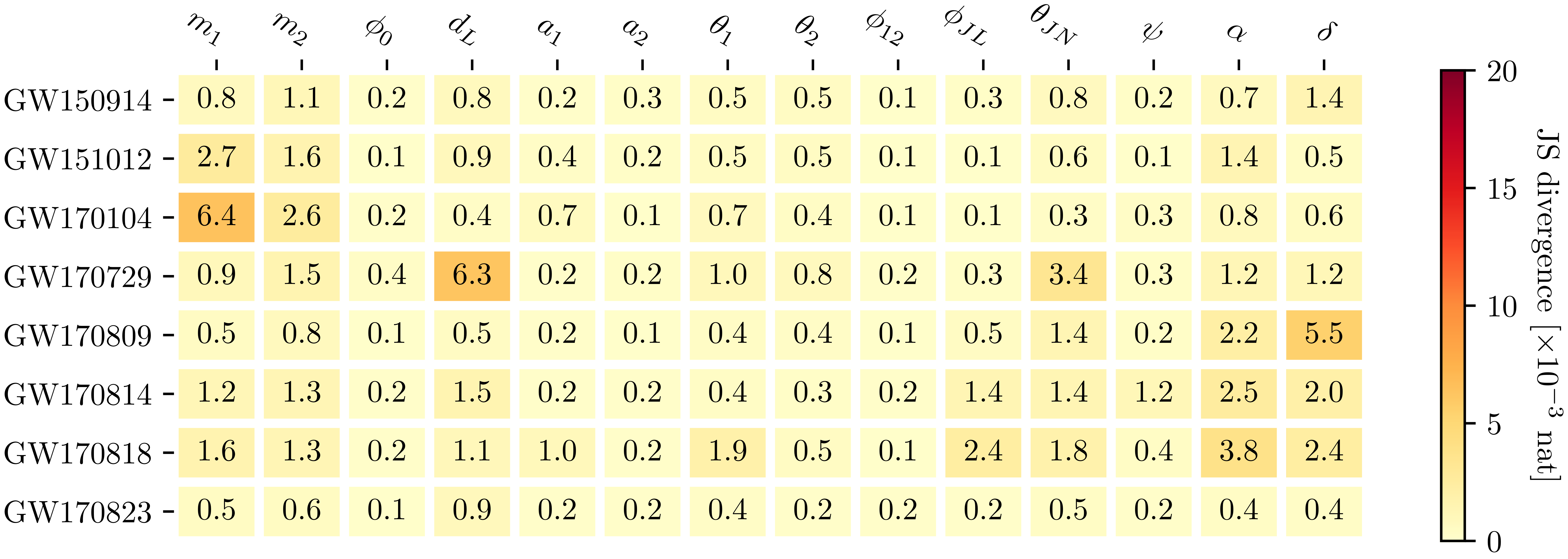
# Evaluation on real events





Maximilian Dax

# Quantitative evaluation on real events



| | $m_1$ | $m_2$ | $\phi_0$ | $d_L$ | $a_1$ | $a_2$ | $\theta_1$ | $\theta_2$ | $\phi_{12}$ | $\phi_{JL}$ | $\theta_{JN}$ | $\psi$ | $\alpha$ | $\delta$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 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 ≤ 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

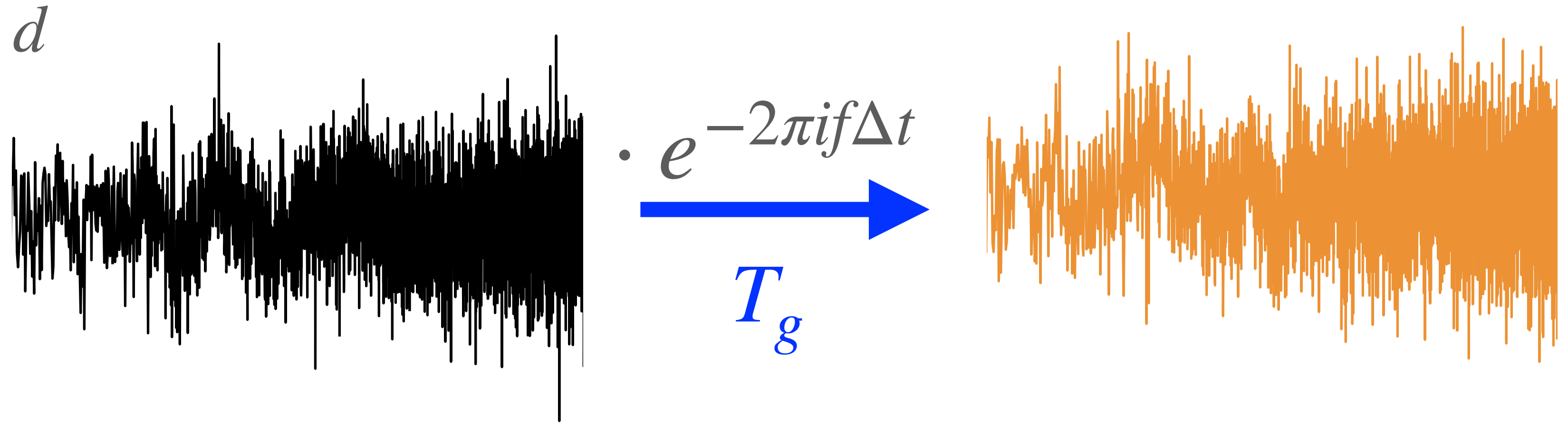
# Symmetries in GW inference

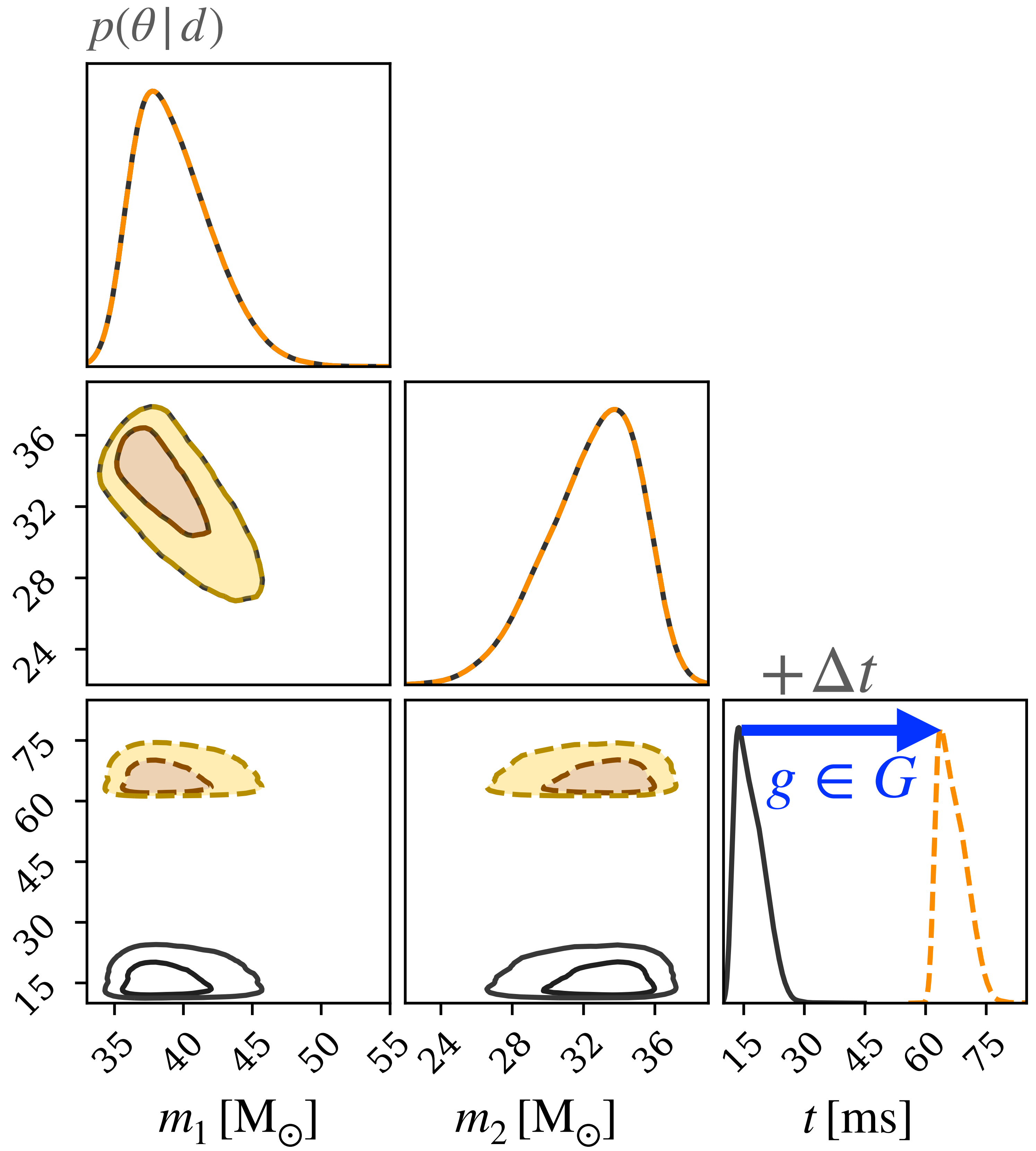
- Equivariance (covariance) under time shift

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

- NPE learns such symmetries from simulation data
  $\Rightarrow$ requires network and training capacity

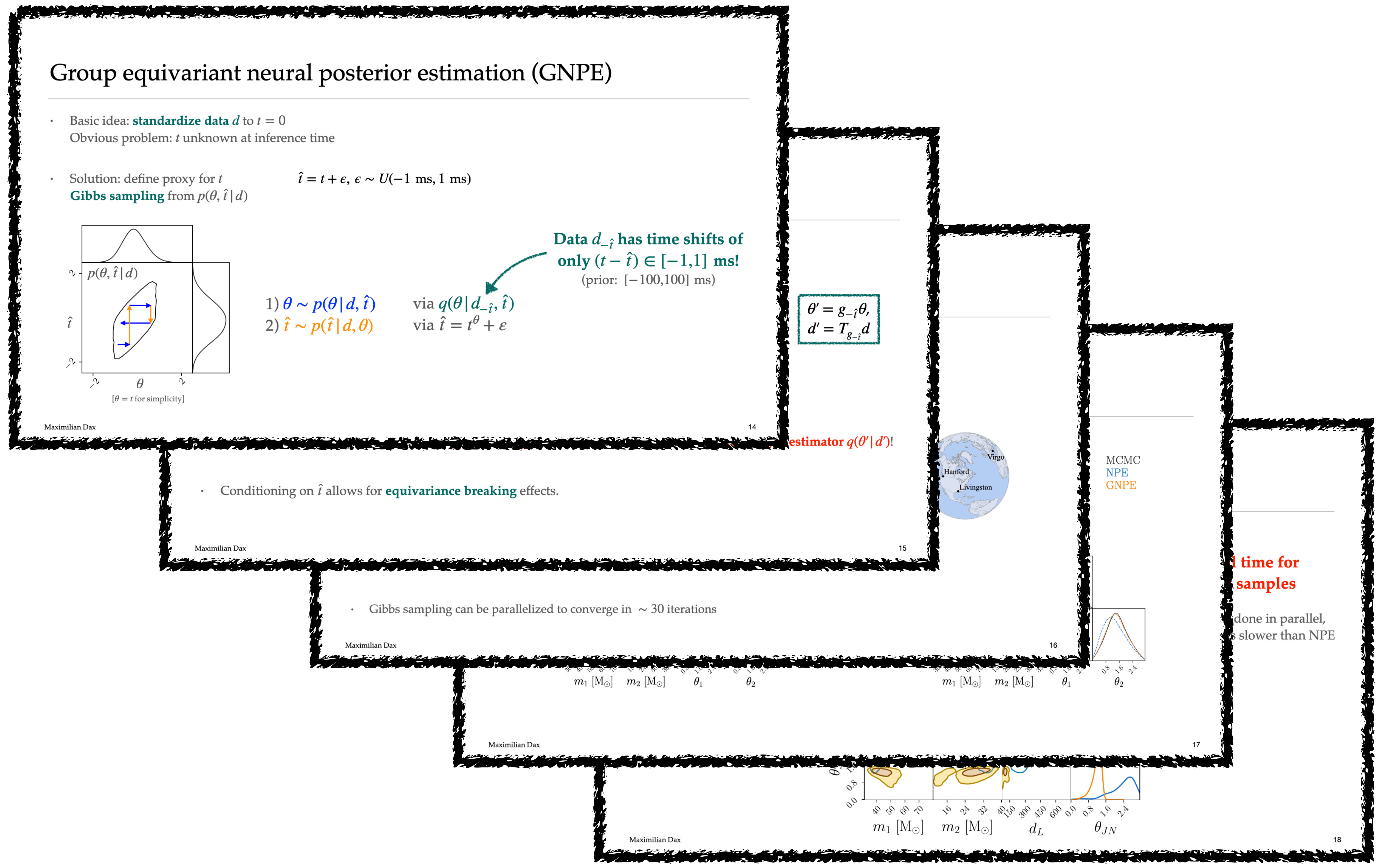
- How can we **enforce such symmetries**?





# 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 $t$ $\hat{t} = t + \epsilon,\ \epsilon \sim U(-1\ \text{ms}, 1\ \text{ms})$
  **Gibbs sampling** from $p(\theta, \hat{t}\,|\,d)$

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

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

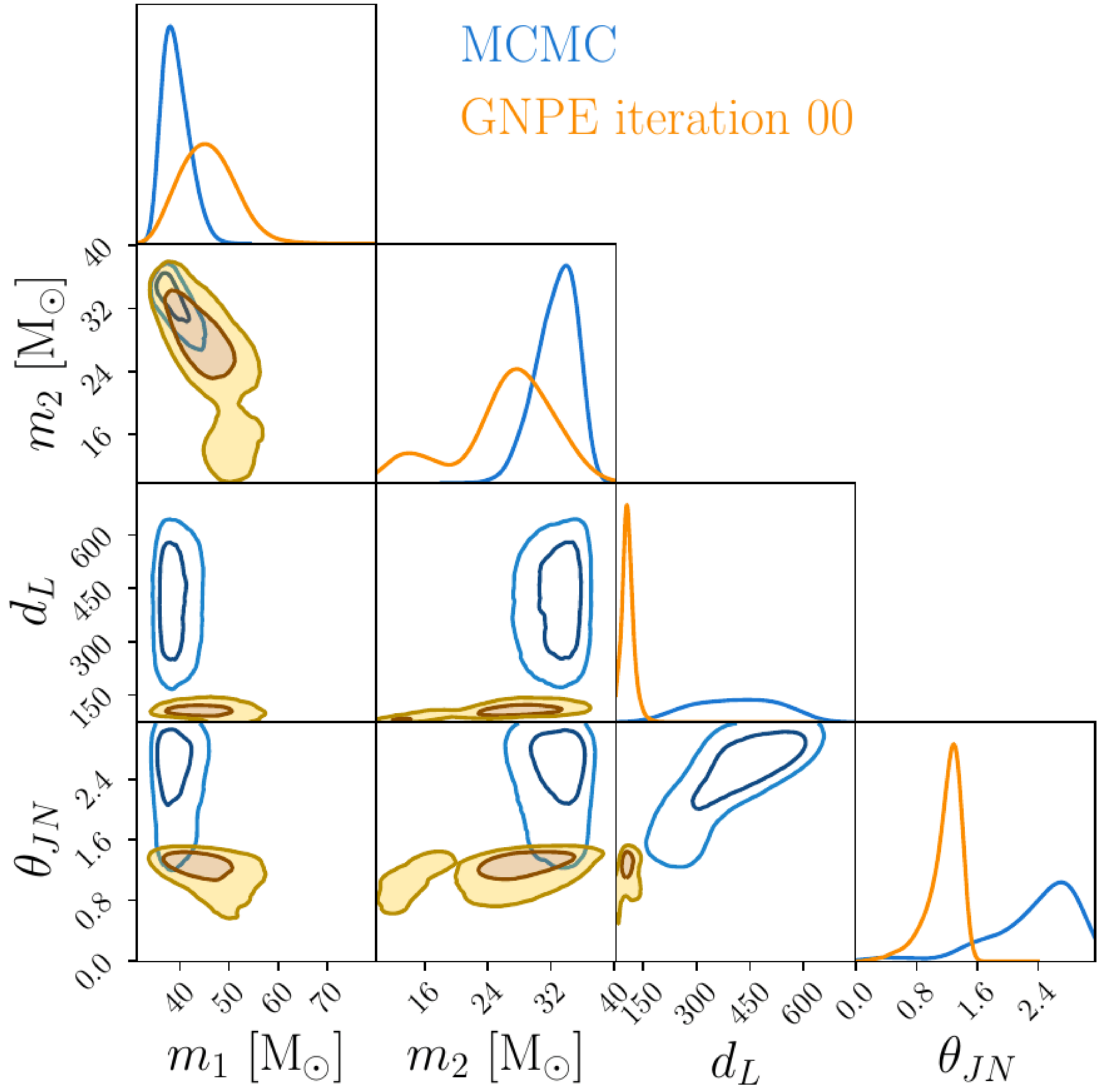
Maximilian Dax 14

- Conditioning on $\hat{t}$ allows for **equivariance breaking** effects.
- Gibbs sampling can be parallelized to converge in $\sim 30$ iterations

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

arXiv:2111.13139

# GNPE iterations



**Real time**

(50k samples)

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


THIS IS YOUR MACHINE LEARNING SYSTEM?
YUP! YOU POUR THE DATA INTO THIS BIG PILE OF LINEAR ALGEBRA, THEN COLLECT THE ANSWERS ON THE OTHER SIDE.
WHAT IF THE ANSWERS ARE WRONG?
JUST STIR THE PILE UNTIL THEY START LOOKING RIGHT.
DATA
ANSWERS

[from reddit]

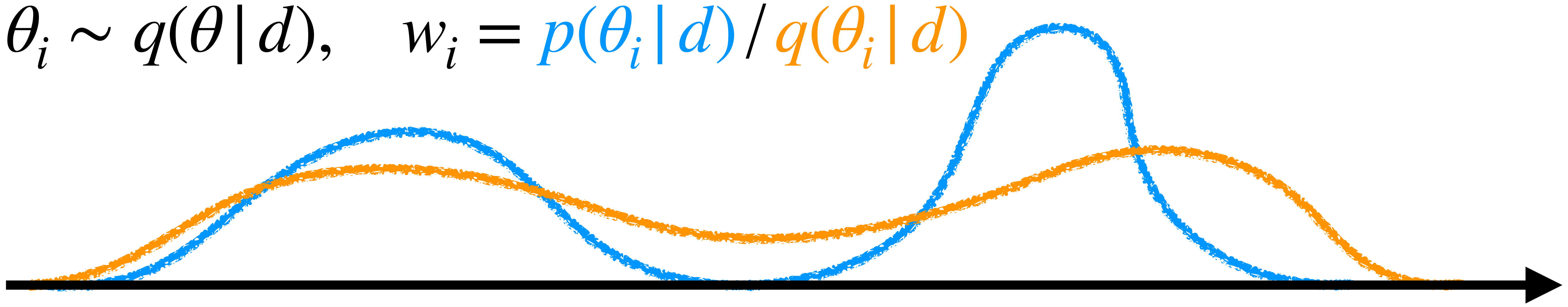
# III. Verification through Importance Sampling

[3] Dax*, Green* et al., Neural Importance Sampling for Rapid and Reliable Gravitational-Wave Inference, Preprint (2022)

# Importance sampling

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

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



⇒ **Produce exact results by reweighting DINGO samples with likelihood**

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

- DINGO minimizes forward KL-divergence $\mathrm{KL}(p \,|\, q)$ which is probability-mass covering!
  $\mathbf{supp}(p) \not\subseteq \mathbf{supp}(q) \Rightarrow$ **diverging DINGO loss**

**verified empirically**

# Importance sampling diagnostics and evidence

- Effective sample size $n_{\text{eff}}$ related to variance of the weights.
  Sample **efficiency $\epsilon$ 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**

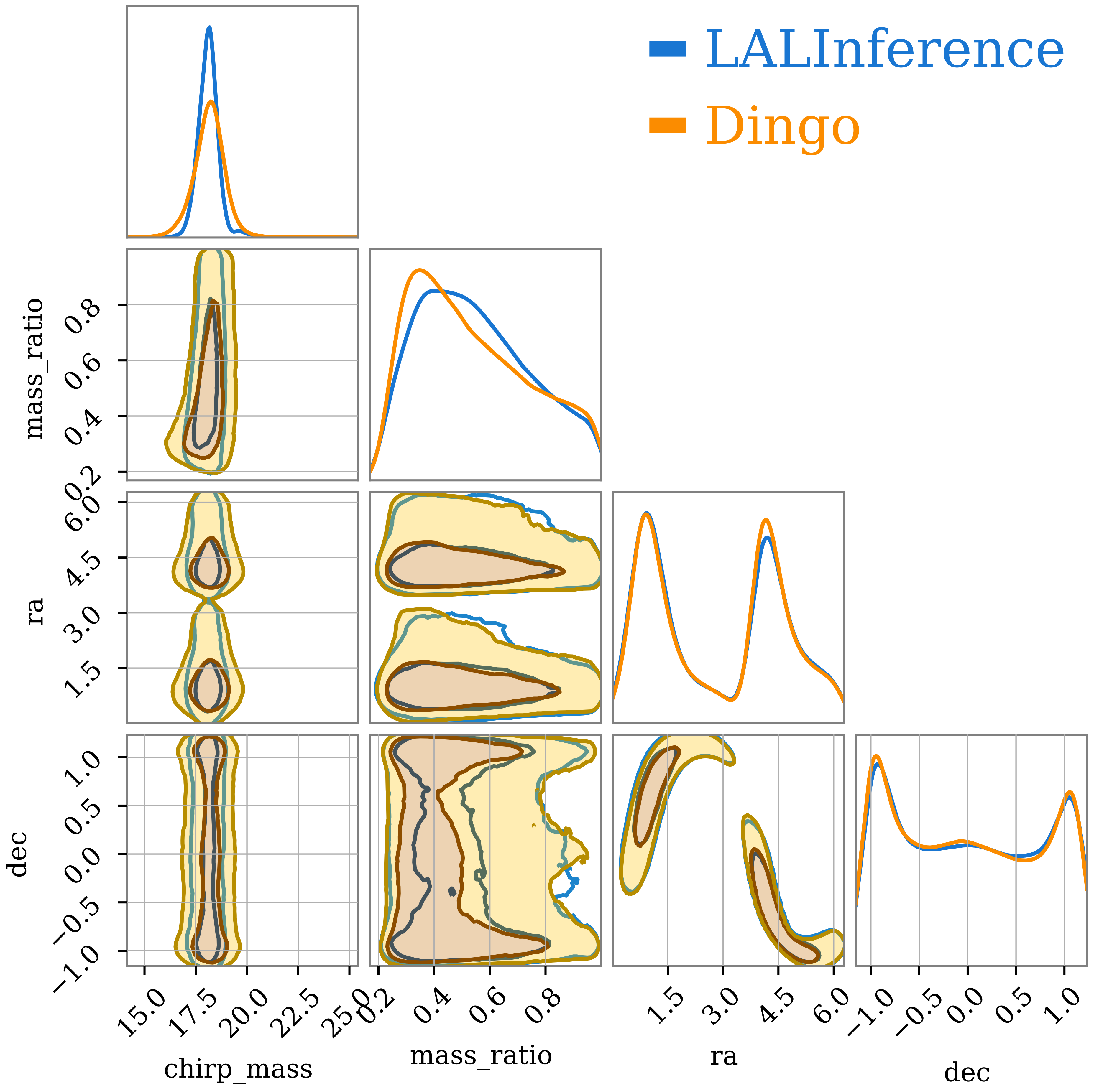
- **Bayesian evidence** related to normalisation of weights.

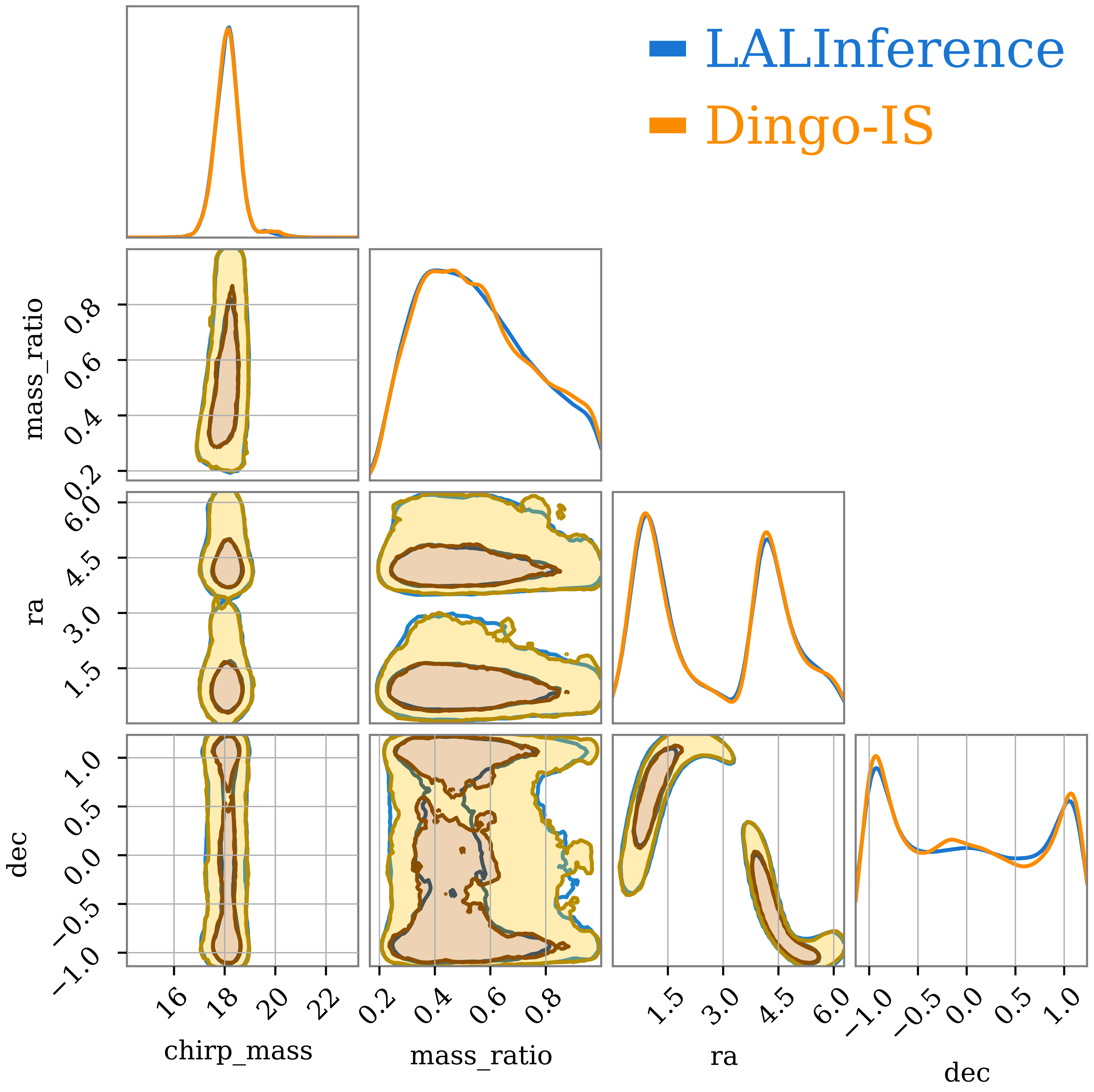
$$p(d) = \frac{1}{n} \sum_i w_i \qquad \sigma_{\log p(d)} = \sqrt{\frac{(1-\epsilon)}{(n \cdot \epsilon)}}$$

**⇒ Unbiased estimate, 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)
  $\Rightarrow$ IS only way of verifying results

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

# Summary

- DINGO: **Rapid GW inference** with **production-level accuracy**
  **⇒ 1000 times faster than MCMC**

- Symmetry-aware **GNPE framework crucial** for performance
  → GNPE is a generic extension, applicable to general inference problems

- DINGO-IS: Importance sampling for verification and correction
  → **exact results**, robust to OOD data
  → **100x efficiency** gain, **10x precision** gain (evidence) compared to MCMC

- Future work: binary neutron stars, noise model free inference, new density estimators

- DINGO code under review for production use @LIGO-Virgo-KAGRA