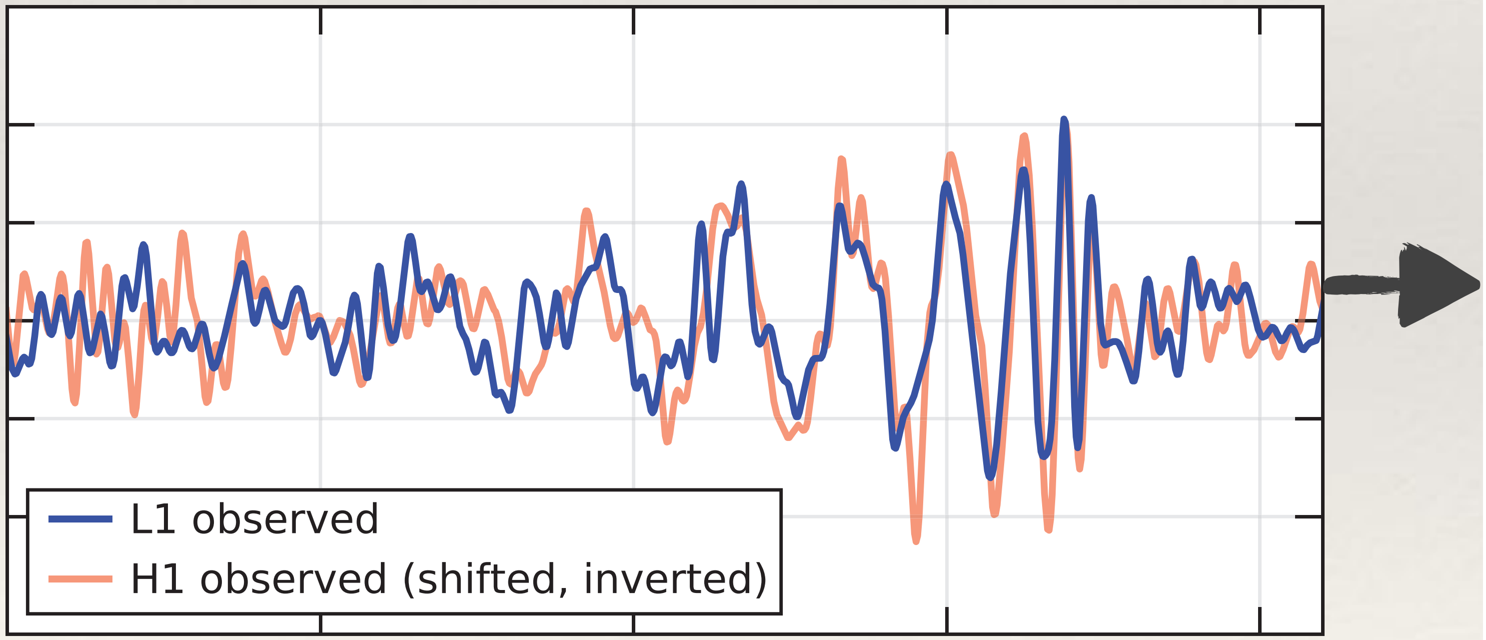
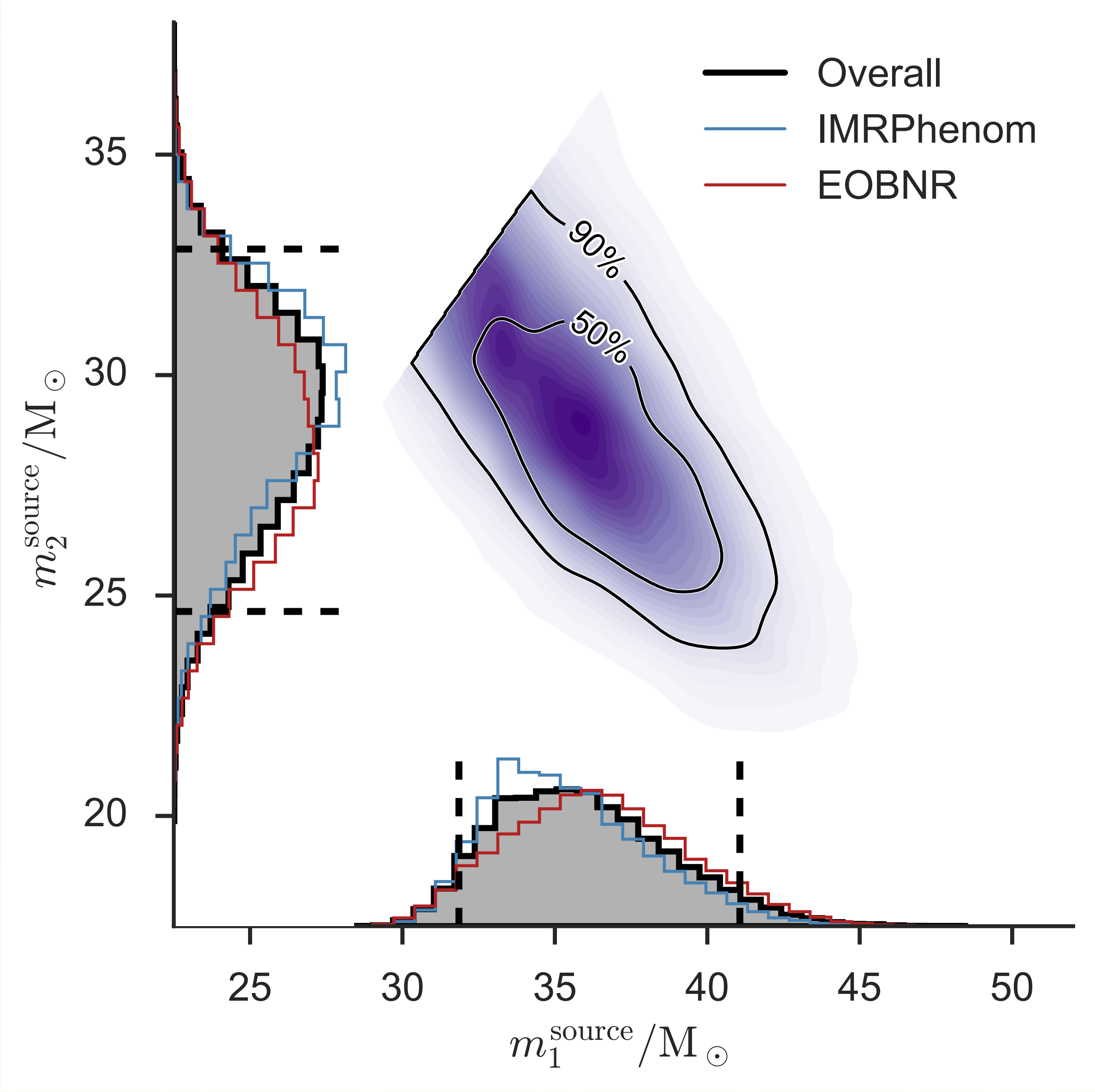
Machine learning for gravitational wave inference

International conference on the physics of the two infinities, Kyoto, March 28th 2023

Jonathan Gair, Max Planck Institute for Gravitational Physics (Albert Einstein Institute), Potsdam

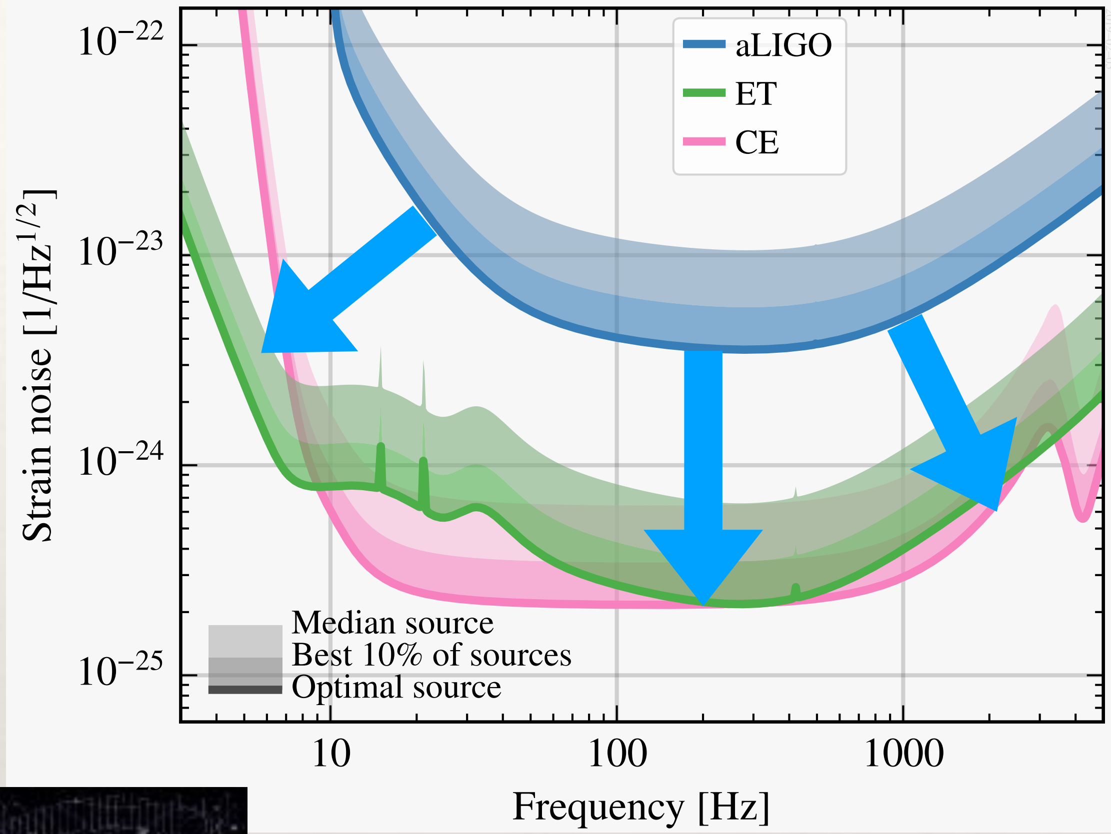
collaborators: **Stephen Green**, **Max Dax**, **Jonas Wildberger**, Jakob Macke, Michael Pürrer, Bernhard Schölkopf, Alessandra Buonanno

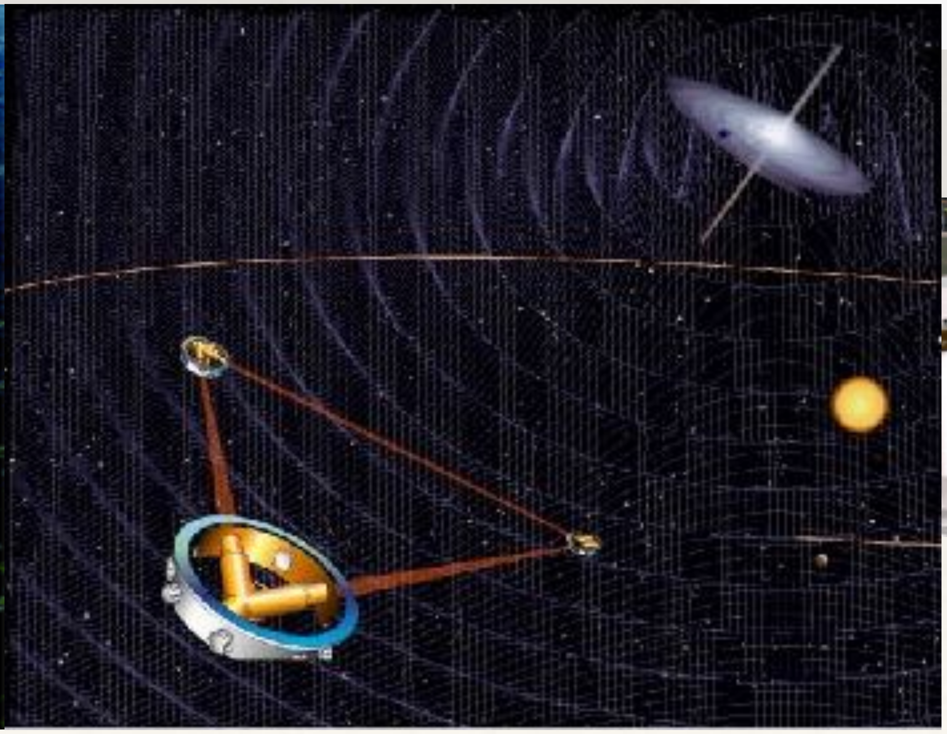


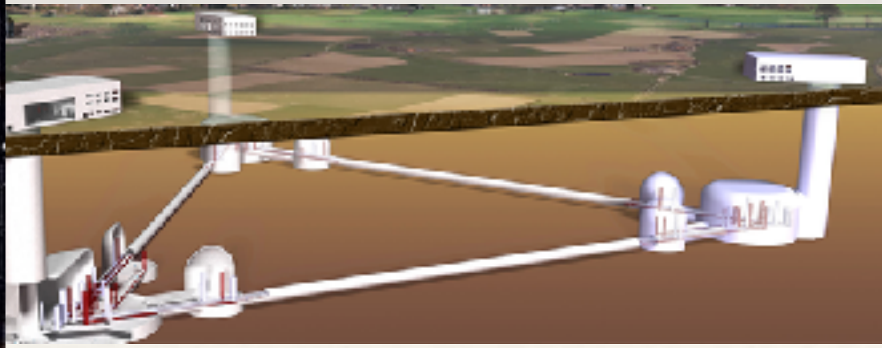


Current and future detectors

- **LIGO/Virgo/KAGRA**: Ground-based interferometers currently operating. 90+ (likely) astrophysical sources observed to date, over three observing runs.
- **LISA**: space-based interferometer to launch in ~2035, operating in mHz band. ESA-led; NASA contributions,
- **3G**: next generation ground-based detector concepts under development. **Einstein Telescope** (Europe) and **Cosmic Explorer** (US). To start operation in ~2030s.







Overview of GW parameter estimation

- GW parameter estimation typically uses Bayesian inference, in which we obtain samples from the *posterior distribution* after specifying a *prior distribution* and the *likelihood*

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

- To specify the likelihood, we typically assume the detector output is a linear combination

$$s(t) = n(t) + h(t;\vec{\theta})$$

- and that the noise is Gaussian and stationary, giving the likelihood

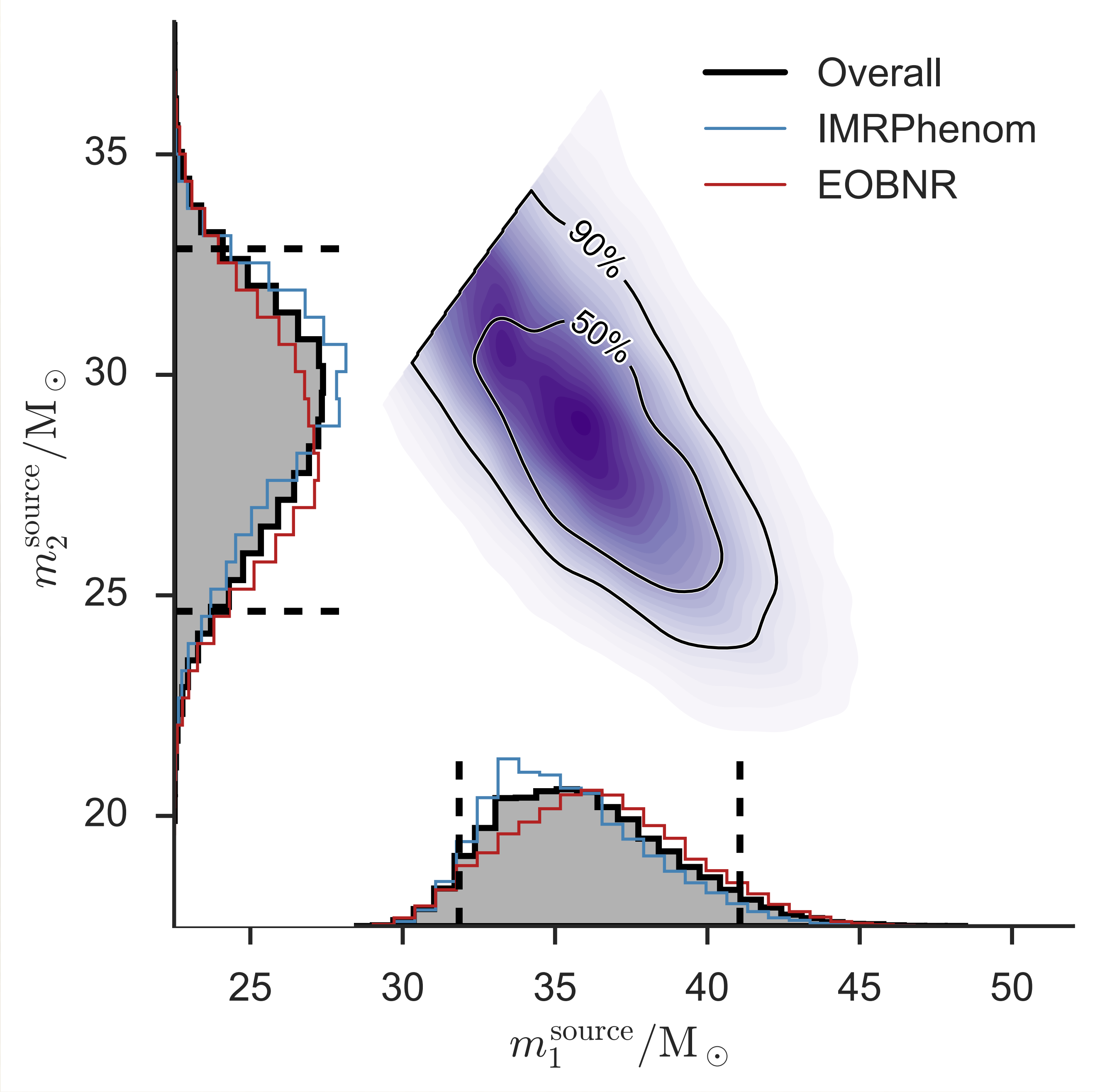
$$p(d|\vec{\theta}) \propto \exp\left[-\frac{1}{2}(d - h(\vec{\theta})|d - h(\vec{\theta}))\right] \qquad (a|b) = \int_{-\infty}^{\infty} \frac{\tilde{a}^*(f)\tilde{b}(f) + \tilde{a}(f)\tilde{b}^*(f)}{S_h(f)}\mathrm{d}f$$

- Inference typically uses *Markov Chain Monte Carlo* or other stochastic sampling methods to draw samples from the posterior distribution - needs **millions** of likelihood evaluations, which rely on constructing **expensive** waveform models.

Computational cost: GW150914

- The analysis of GW150914 used 50 million CPU hours (20,000 PCs running for 100 days). A significant part of that was for PE.
- Lag between observation and publication of exceptional events mostly dominated by PE (re-)runs.

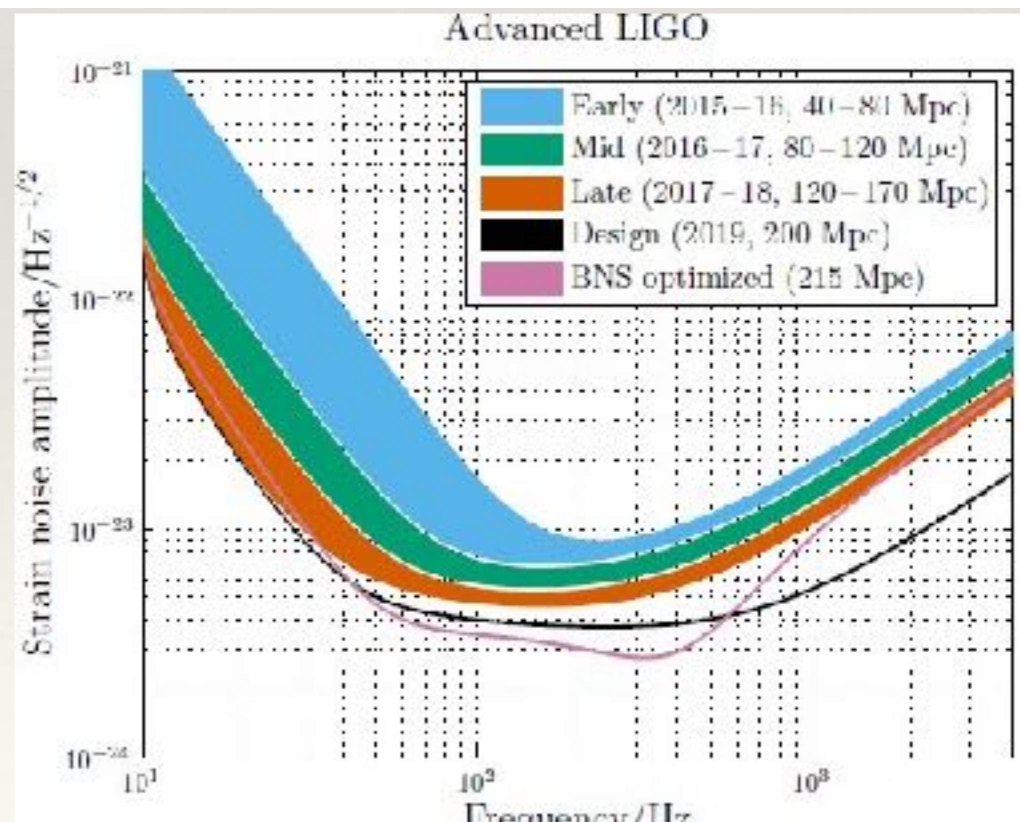
Primary black hole mass	$36^{+5}_{-4} M_\odot$
Secondary black hole mass	$29^{+4}_{-4} M_\odot$
Final black hole mass	$62^{+4}_{-4} M_\odot$
Final black hole spin	$0.67^{+0.05}_{-0.07}$
Luminosity distance	410^{+160}_{-180} Mpc
Source redshift z	$0.09^{+0.03}_{-0.04}$



Challenges in GW parameter estimation

- Future detectors will have more events: expect to move from ~1 event/week to several/day.
- Future detectors will have wider bandwidths: new types of source, longer waveforms and hence more expensive PE.
- Sources for LISA (and to a lesser extent 3G detectors) will overlap in time and frequency.
- Fast PE needed for multi-messenger: send triggers for follow-up.
- Need faster inference: accelerated waveform models, faster likelihoods, novel sampling techniques etc.

Epoch			2015 – 2016	2016 – 2017	2018 – 2019	2020+	2024+
Planned run duration			4 months	9 months	12 months	(per year)	(per year)
Expected burst range/Mpc		LIGO	40 – 60	60 – 75	75 – 90	105	105
		Virgo	—	20 – 40	40 – 50	40 – 70	80
		KAGRA	—	—	—	—	100
Expected BNS range/Mpc		LIGO	40 – 80	80 – 120	120 – 170	190	190
		Virgo	—	20 – 65	65 – 85	65 – 115	125
		KAGRA	—	—	—	—	140
Achieved BNS range/Mpc		LIGO	60 – 80	60 – 100	—	—	—
		Virgo	—	25 – 30	—	—	—
		KAGRA					
Estimated BNS detections			0.05 – 1	0.2 – 4.5	1 – 50	4 – 80	11 – 180
Actual BNS detections			0	1			
90% CR	% within	5 deg^2	< 1	1 – 5	1 – 4	3 – 7	23 – 30
		20 deg^2	< 1	7 – 14	12 – 21	14 – 22	65 – 73
		median/deg^2	460 – 530	230 – 320	120 – 180	110 – 180	9 – 12
Searched area	% within	5 deg^2	4 – 6	15 – 21	20 – 26	23 – 29	62 – 67
		20 deg^2	14 – 17	33 – 41	42 – 50	44 – 52	87 – 90



Neural posterior estimation

- Stochastic sampling relies on being able to evaluate the likelihood, $p(d|\theta)$, which requires a new waveform evaluation at each sampling step.

- Alternative: construct a neural network that generates samples from $q(\theta|d)$, a distribution that approximates the parameter posterior distribution, $p(\theta|d)$. Train by minimising the average *cross-entropy* with the true distribution

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

- This is *simulation based inference*. We compute the loss using simulated data

$$\textbf{Sample } \theta^{(i)} \sim p(\theta), \quad i = 1, \ldots, N$$

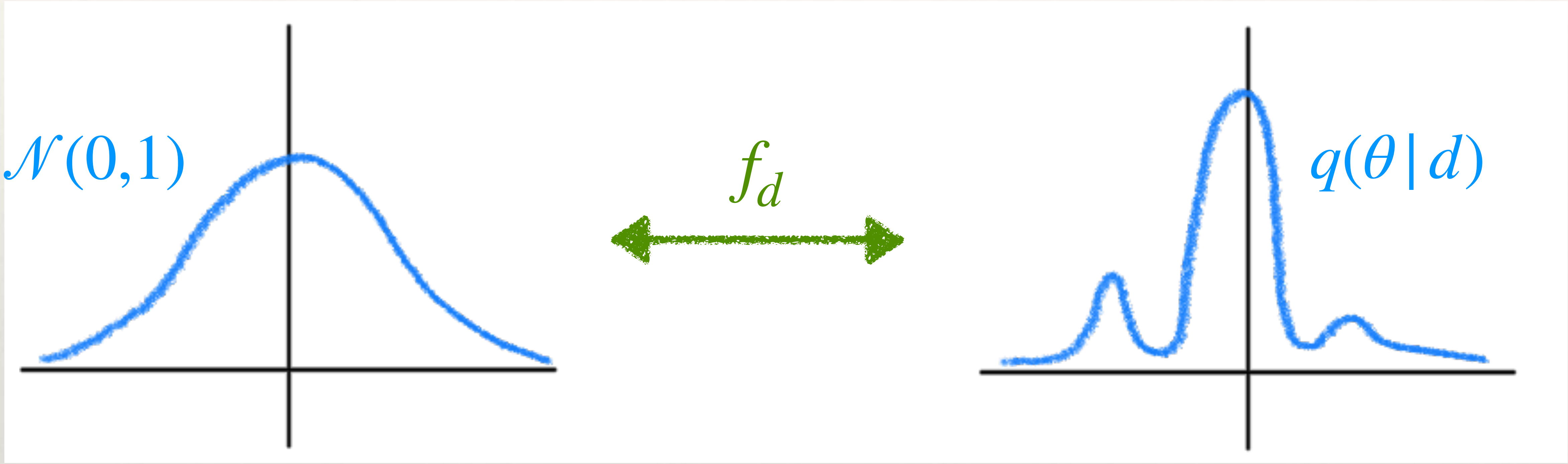
$$\textbf{Simulate } d^{(i)} \sim p(d|\theta^{(i)}); \quad d^{(i)} = h(\theta^{(i)}) + n^{(i)} \text{ with } n^{(i)} \sim p_{S_n}(n)$$

$$\textbf{Compute } L \approx \frac{1}{N}\sum_{i=1}^{N}\left[-\log q(\theta^{(i)}|d^{(i)})\right]$$

- **Advantages**: *likelihood-free*, *amortised* cost of waveform generation, *flexible*.

Normalizing flows

- A normalising flow represents a complex distribution as a mapping of a simple one.



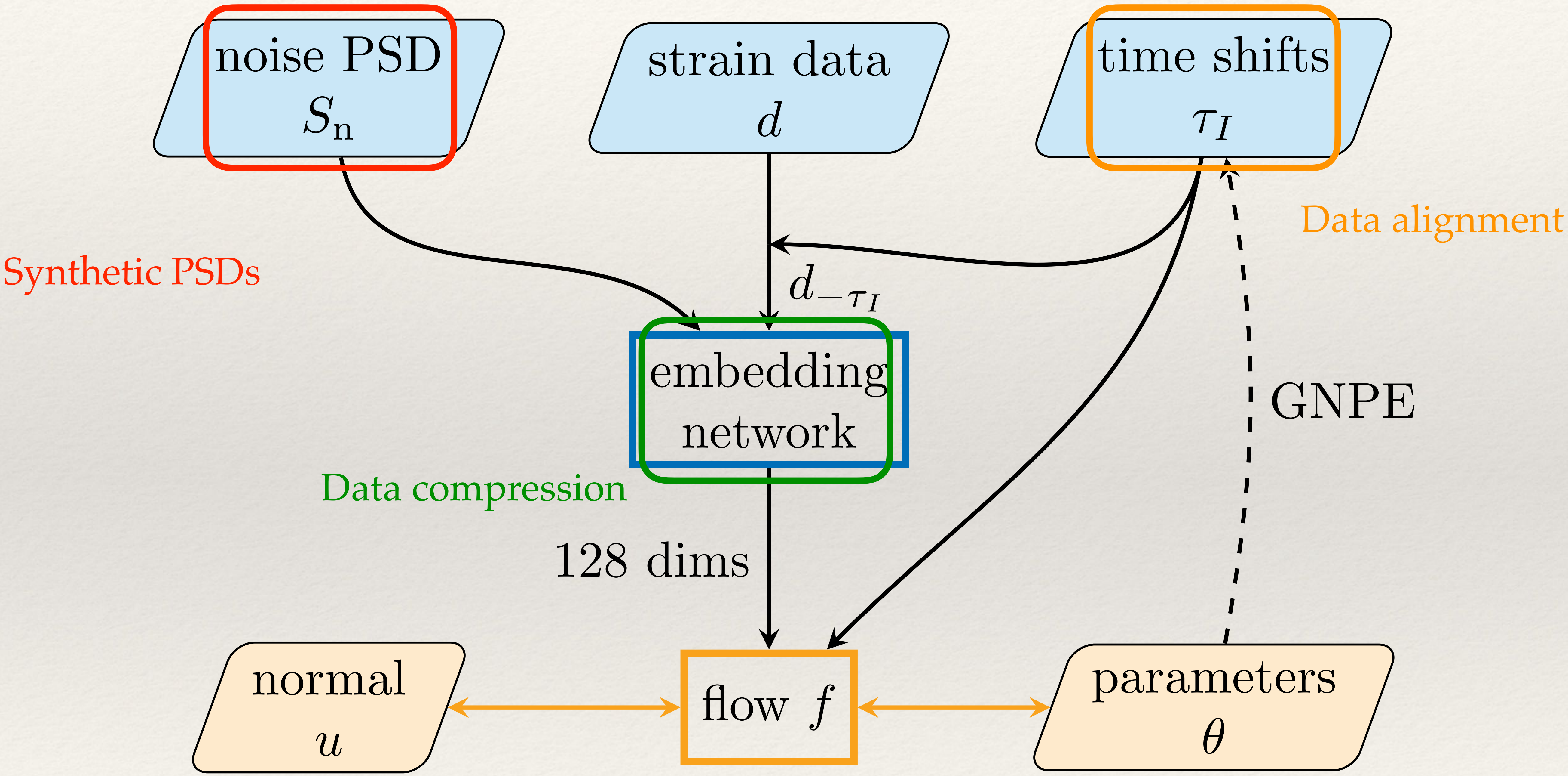
- Construct target distribution using

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

- Want mapping to be invertible and have a simple Jacobian determinant. Can represent a normalising flow with these properties using a neural network.

- We use normalising flows built from a sequence of coupling transforms described by quadratic splines (*spline flows* Durkan et al. 2019).

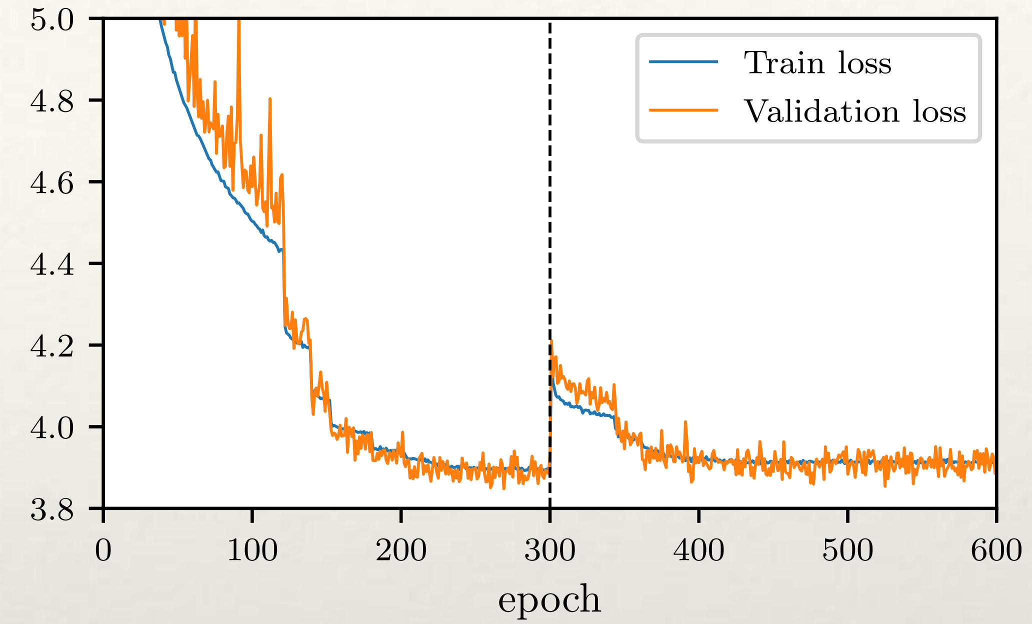
Refinements



Big neural networks: $\approx$ 350 layers and 150 million parameters

Results: GWTC-1 BBHs

- Used 5×10^6 waveforms for training
    - IMRPhenomPv2
    - $T = 8$ s, $f_{\mathrm{min}} = 20$ Hz, $f_{\mathrm{max}} = 1024$ Hz
    - 15D parameter space
    - $m_1, m_2 \in [10,80]$ $\mathrm{M}_\odot$

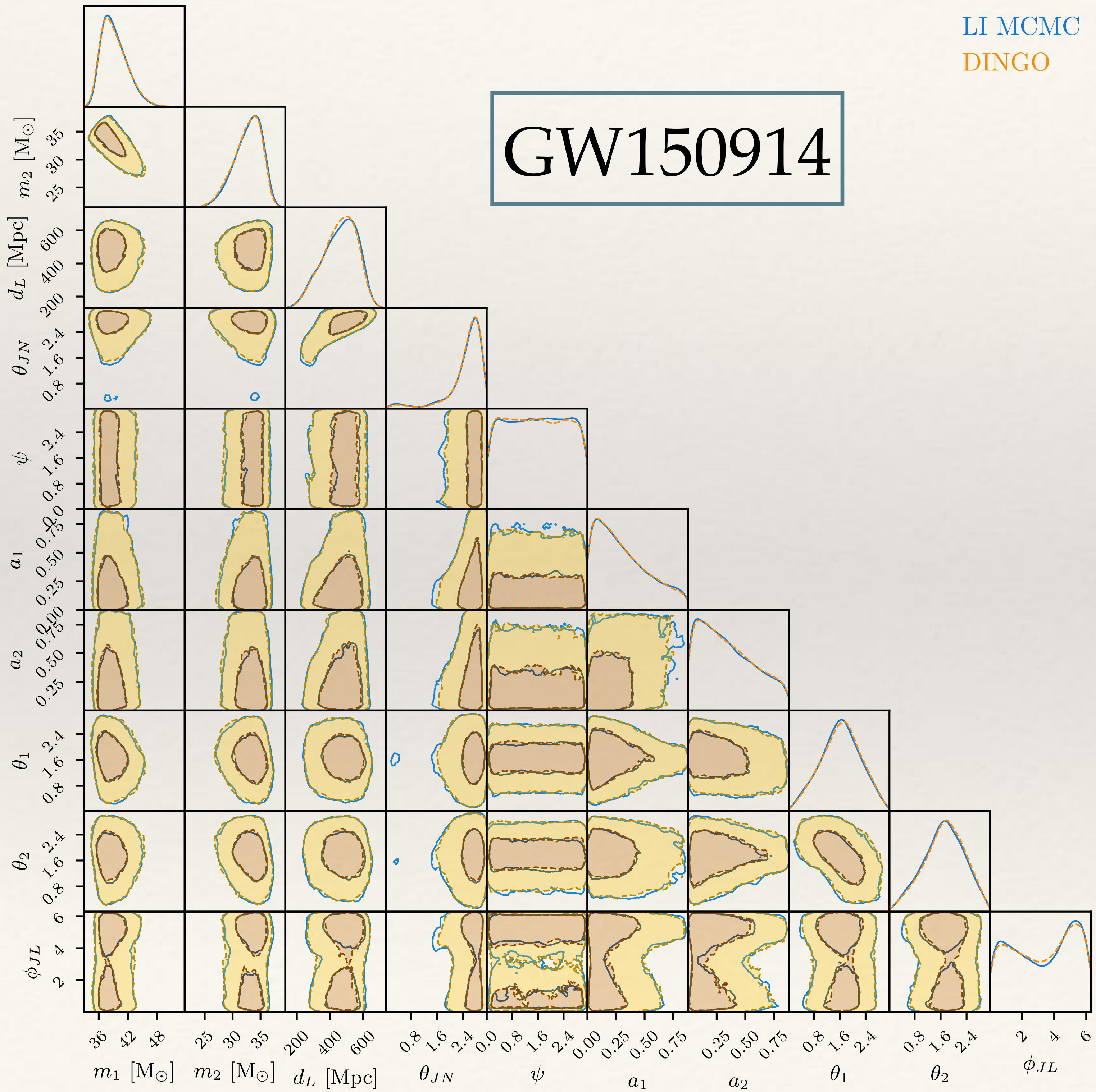


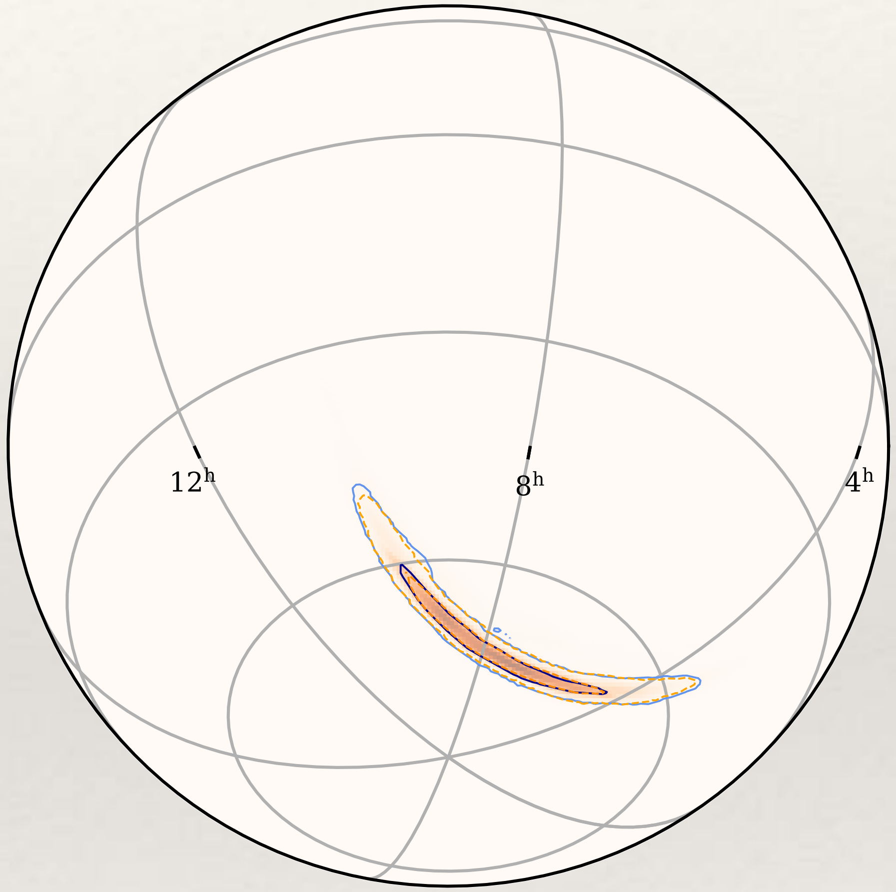
- + stationary Gaussian noise realisations consistent with measured PSDs

- Trained several neural networks based on different noise level / number of detectors / distance range:

Observing run	Detectors	Distance range [Mpc]
O1	HL	[100, 2000]
O2	HL	[100, 2000]
		[100, 6000]
	HLV	[100, 1000]

Results: GWTC-1 BBHs





50,000 samples in ~ 20 s

Results: GWTC-1 BBHs

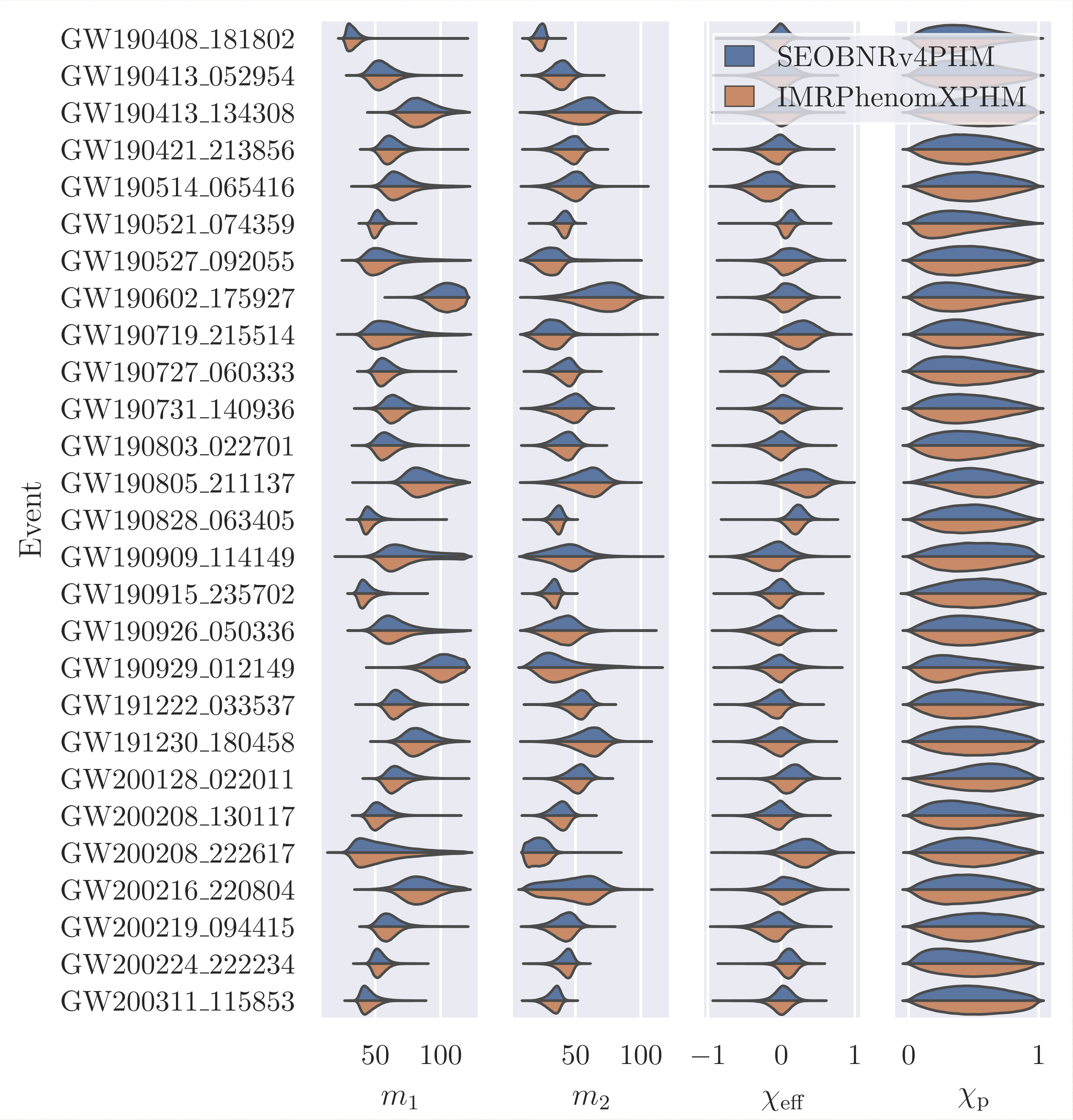
- Compare NPE posteriors to "standard" posteriors generated by *LALInference* and *Bilby*. Use JS divergence as a metric for comparison.

- JS divergence less than 2 nats considered *indistinguishable*.

	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 [$\times 10^{-3}$ nat] (colour scale: 0, 5, 10, 15, 20)

Results: O3 events

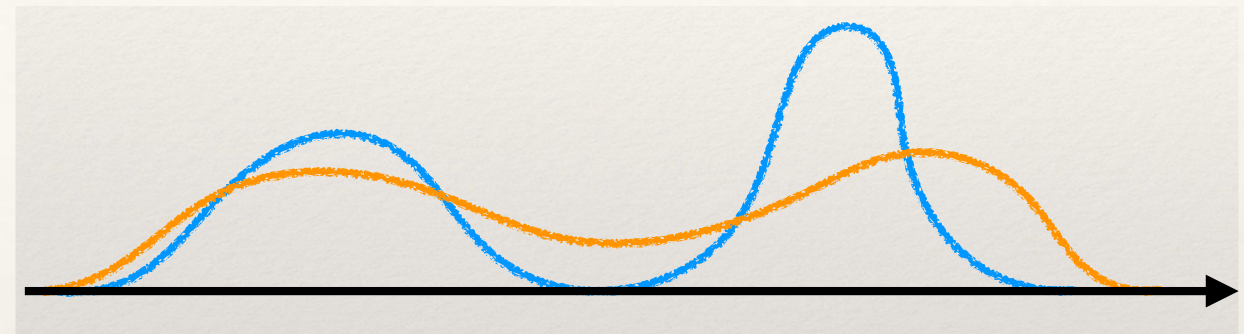


Used to analyse 30 BBHs from GWTC-3 using precessing, higher mode waveform models.

First robust PE to be obtained using *SEOBNRv4PHM* for many of these events.

Result validation: importance sampling

- Can reweight samples to target density using *importance sampling*.
- NPE samples suitable for reweighting because *probability mass covering*.
- Several advantages
  - correct posterior inaccuracies.
  - effective sample size, n_{eff}, provides a metric of quality of samples.
  - evidence can be directly computed.
  - neural networks extrapolate unpredictably to out-of-distribution (OOD): low n_{eff} indicates OOD data.
- **But**: no longer likelihood-free! :-(



target (prior x likelihood)

$$w_i \propto \frac{p(\theta_i)p(d\,|\,\theta_i)}{q(\theta_i\,|\,d)}$$

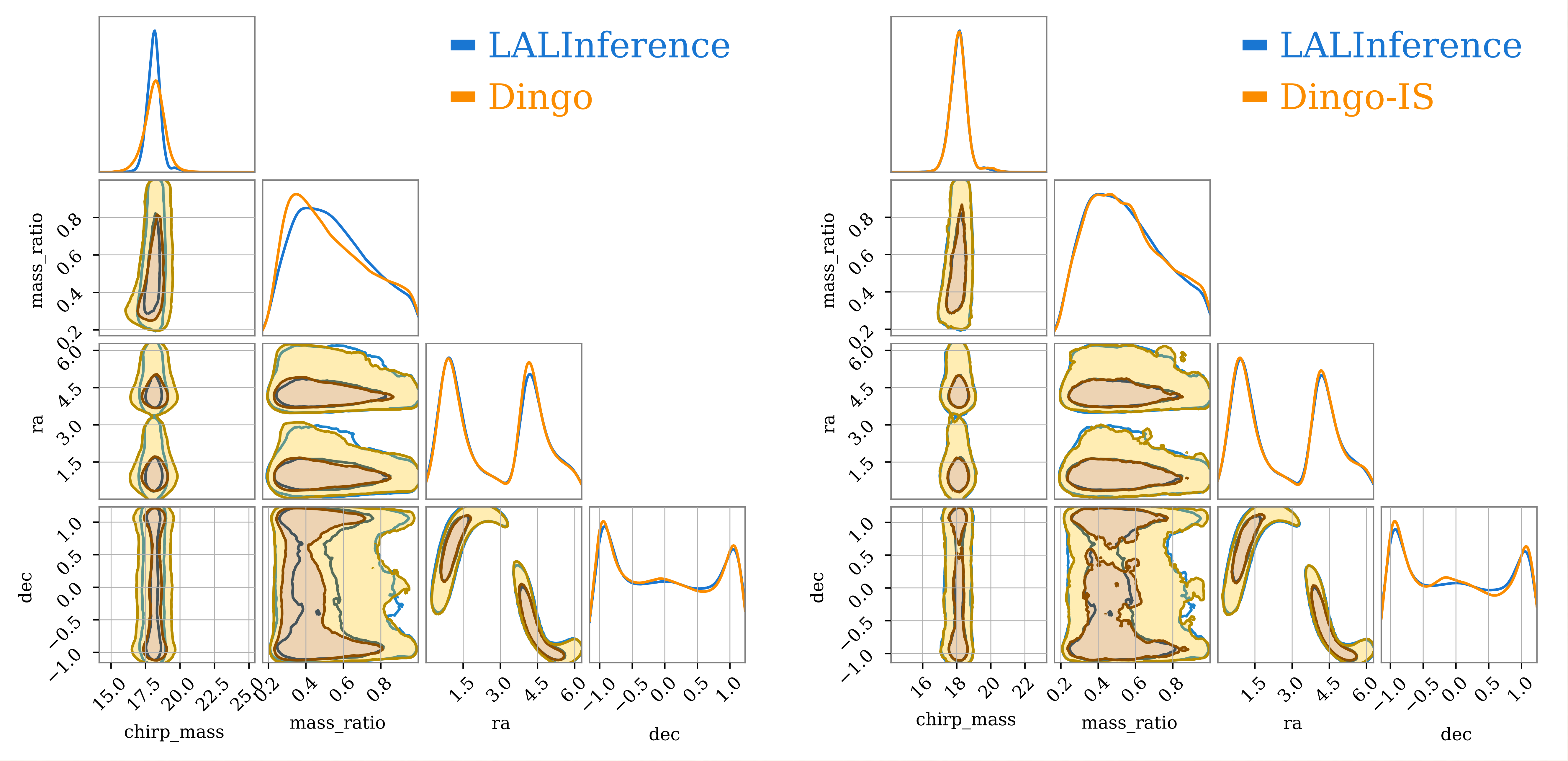
proposal (NPE)

$$n_{\text{eff}} = \frac{\left(\sum_i w_i\right)^2}{\sum_i w_i^2}$$

$$p(d) \approx \frac{1}{n}\sum_{i=1}^{n} w_i$$

Importance sampling: posteriors

- Importance-reweighted posteriors show closer agreement with standard sampling



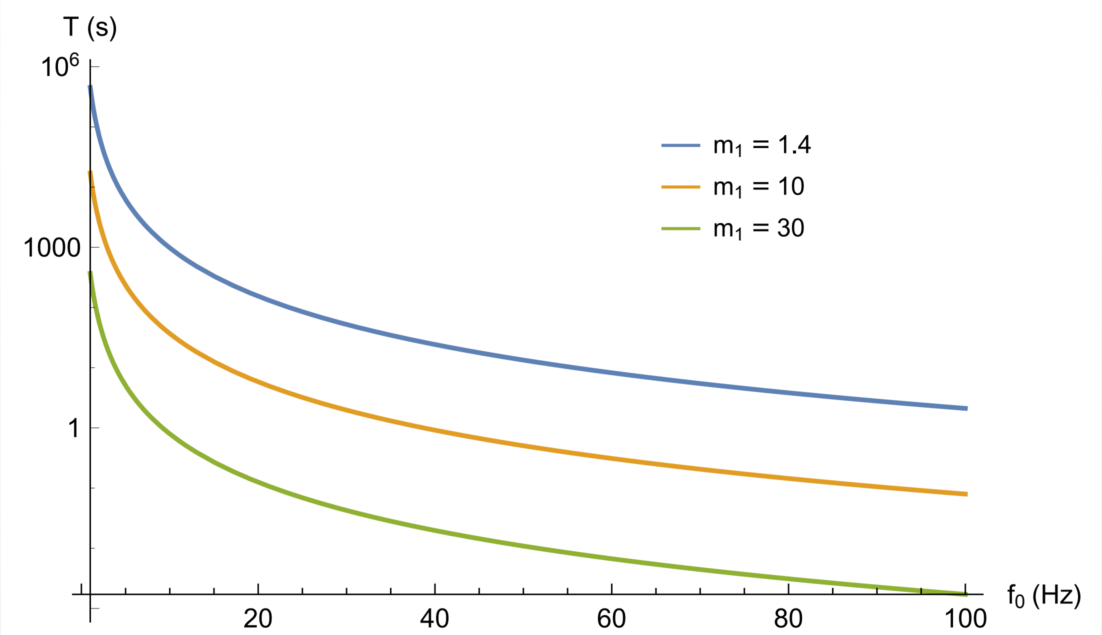
DINGO-IS

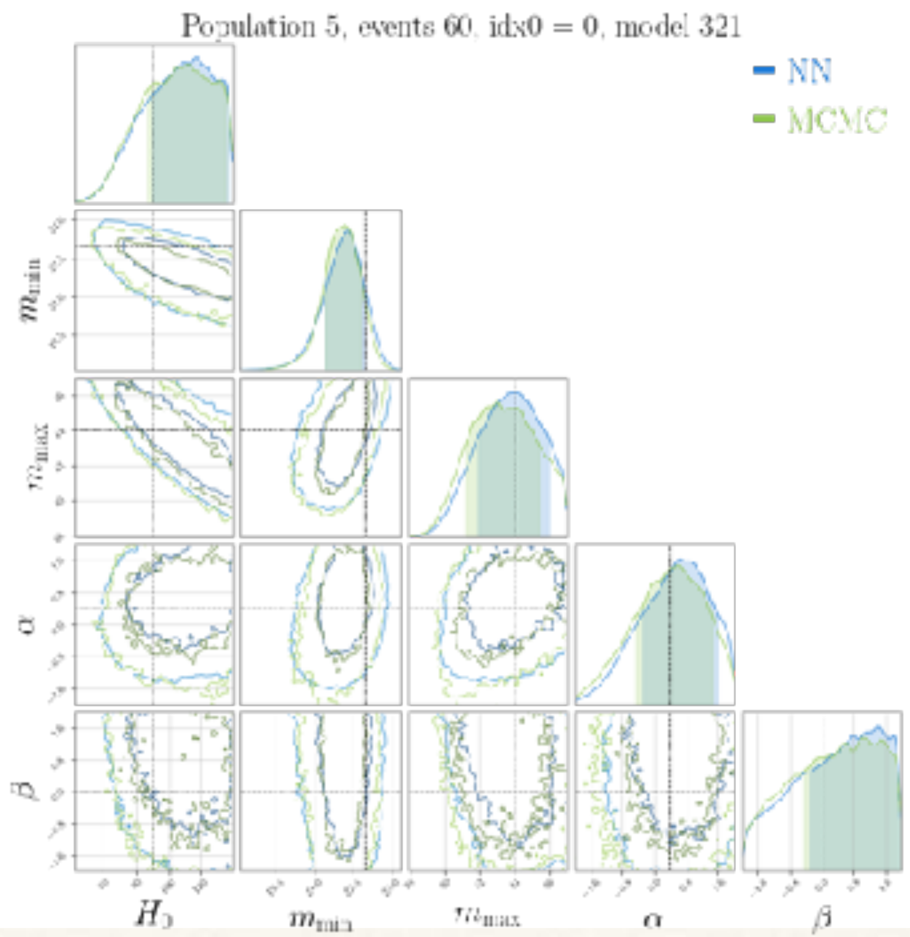
Event	$\log p(d)$	ϵ	Event	$\log p(d)$	ϵ	Event	$\log p(d)$	ϵ
GW190408	−16178.332 ± 0.012	6.9%	GW190727	−15992.017 ± 0.009	10.3%	GW191230	−15913.798 ± 0.009	12.2%
_181802	−16178.172 ± 0.010	9.3%	_060333	−15992.428 ± 0.005	30.8%	_180458	−15913.918 ± 0.010	8.8%
GW190413	−15571.413 ± 0.006	22.5%	GW190731	−16376.777 ± 0.005	32.6%	GW200128	−16305.128 ± 0.013	6.1%
_052954	−15571.391 ± 0.005	26.3%	_140936	−16376.763 ± 0.005	31.0%	_022011	−16304.510 ± 0.007	18.3%
GW190413	−16399.331 ± 0.009	12.4%	GW190803	−16132.409 ± 0.006	21.4%	‡GW200129	−16226.851 ± 0.109	0.1%
_134308	−16399.139 ± 0.014	4.7%	_022701	−16132.408 ± 0.005	27.8%	_065458	−16231.203 ± 0.051	0.4%
GW190421	−15983.248 ± 0.008	15.3%	GW190805	−16073.261 ± 0.006	20.0%	GW200208	−16136.381 ± 0.007	16.6%
_213856	−15983.131 ± 0.010	9.4%	_211137	−16073.656 ± 0.007	16.6%	_130117	−16136.531 ± 0.009	11.2%
GW190503	−16582.865 ± 0.022	2.0%	GW190828	−16137.220 ± 0.009	12.2%	GW200208	−16775.200 ± 0.011	7.4%
_185404	−16583.352 ± 0.027	1.4%	_063405	−16136.799 ± 0.010	9.1%	_222617	−16774.582 ± 0.021	2.2%
GW190513	−15946.462 ± 0.043	0.6%	GW190909	−16061.634 ± 0.011	7.4%	GW200209	−16383.847 ± 0.009	12.5%
_205428	−15946.581 ± 0.017	3.4%	_114149	−16061.275 ± 0.016	3.8%	_085452	−16384.157 ± 0.025	1.6%
GW190514	−16556.466 ± 0.009	11.6%	GW190915	−16083.960 ± 0.015	20.8%	GW200216	−16215.703 ± 0.017	3.4%
_065416	−16556.314 ± 0.017	3.5%	_235702	−16083.937 ± 0.027	4.8%	_220804	−16215.540 ± 0.018	3.1%
GW190517	−16271.048 ± 0.027	1.3%	GW190926	−16015.813 ± 0.019	2.8%	GW200219	−16133.457 ± 0.011	9.6%
_055101	−16272.428 ± 0.034	0.9%	_050336	−16015.861 ± 0.009	12.1%	_094415	−16133.157 ± 0.017	4.0%
GW190519	−15991.171 ± 0.008	15.2%	GW190929	−16146.666 ± 0.018	3.2%	GW200220	−16303.782 ± 0.007	17.3%
_153544	−15991.287 ± 0.068	0.2%	_012149	−16146.591 ± 0.021	2.4%	_061928	−16303.087 ± 0.026	1.5%
GW190521	−16008.876 ± 0.008	13.4%	GW191109	−17925.064 ± 0.025	1.7%	GW200220	−16136.600 ± 0.008	13.2%
_074359	−16008.037 ± 0.015	4.2%	_010717	−17922.762 ± 0.041	0.6%	_124850	−16136.519 ± 0.037	0.7%
GW190527	−16119.012 ± 0.008	13.8%	GW191127	−16759.328 ± 0.019	2.7%	GW200224	−16138.613 ± 0.006	22.5%
_092055	−16118.781 ± 0.013	6.1%	_050227	−16758.102 ± 0.029	1.2%	_222234	−16139.101 ± 0.006	21.4%
GW190602	−16036.993 ± 0.006	25.0%	‡GW191204	−15984.455 ± 0.015	4.2%	‡GW200308	−16173.938 ± 0.013	6.0%
_175927	−16037.529 ± 0.006	23.5%	_110529	−15983.618 ± 0.063	0.3%	_173609	−16173.692 ± 0.025	1.7%
GW190701	−16521.381 ± 0.040	0.6%	GW191215	−16001.286 ± 0.013	5.8%	GW200311	−16117.505 ± 0.011	7.4%
_203306	−16521.609 ± 0.010	10.1%	_223052	−16000.846 ± 0.052	0.4%	_115853	−16117.583 ± 0.009	11.9%
GW190719	−15850.492 ± 0.008	13.4%	GW191222	−15871.521 ± 0.007	16.5%	‡GW200322	−16313.568 ± 0.307	0.0%
_215514	−15850.339 ± 0.011	8.0%	_033537	−15871.450 ± 0.005	25.8%	_091133	−16313.110 ± 0.105	0.1%

Table II. 42 BBH events from GWTC-3 analyzed with DINGO-IS. We report the log evidence $\log p(d)$ and the sample efficiency ϵ for the two waveform models IMRPhenomXPHM (upper rows) and SEOBNRv4PHM (lower rows). Highlighting colors indicate the sample efficiency (green: high; yellow: medium; orange/red: low); DINGO-IS results can be trusted for medium and high ϵ (see Supplemental Material). Events in gray suffer from data quality issues [1, 21]. ‡See remarks on these events in text.

Future challenges

- *DINGO* (Deep INference for Gravitational wave Observations) under review to be used by LVK for PE.
- Various extensions to pursue:
    - ***Additional physics***: eccentricity, GR deviations, lensing etc. (*in progress*).
    - ***Signal duration***: extension to BBH with component masses below 10, NSBH and BNS is a priority.
    - ***Population inference***: use DINGO to provide samples to standard methods or output population posterior directly.
    - ***Instrumental realism***: exploit simulation based inference to move away from Gaussian approximation.
    - ***Overlapping signals***: necessary for 3G detectors and LISA.





Summary

- Gravitational wave science relies on obtaining **parameter posterior distributions** for all observed sources. Multi-messenger applications require **rapid estimation** of sky position, and perhaps other parameters.
- Current PE codes are computationally intensive—need new methods that are **fast**, **robust** and **accurate**.
- **Neural posterior estimation** is a new, machine learning approach that now has comparable performance to standard methods in a fraction of the time. Training cost is amortised, allowing **near real-time analysis** of new observations.
- Code under internal review to be used within the LVK for analysis of future events.
- **Many developmental challenges remain**: long waveforms, non-stationary noise, new sources, overlapping sources, population inference….