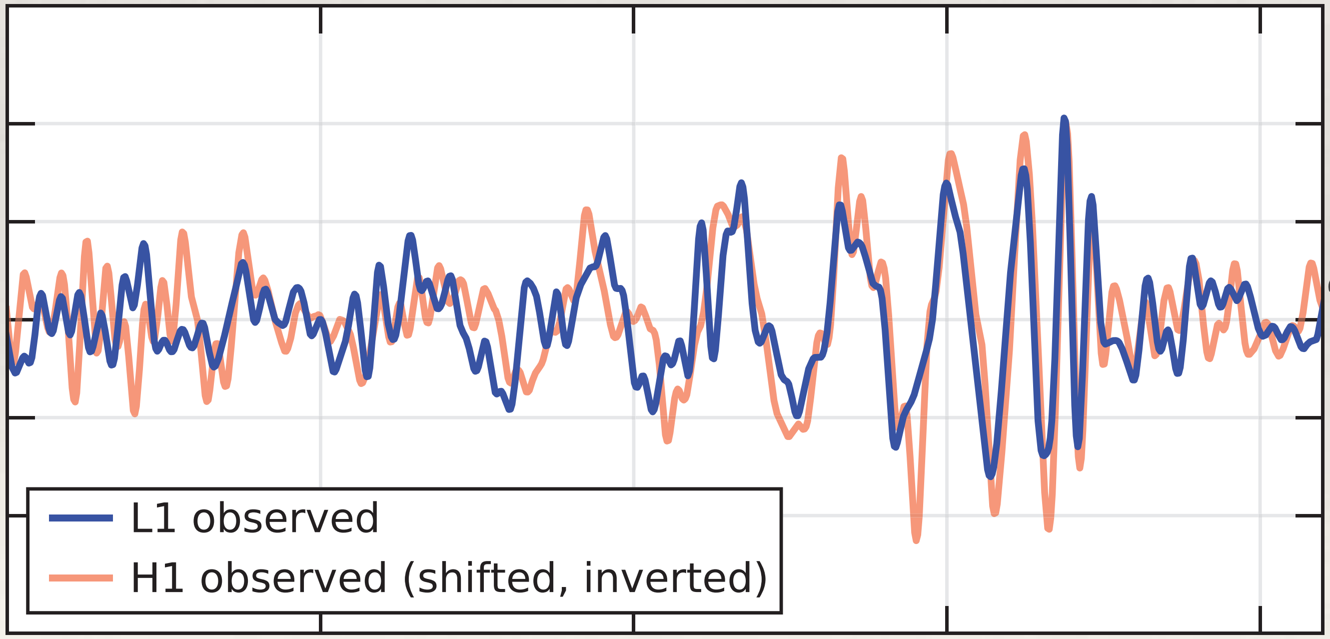
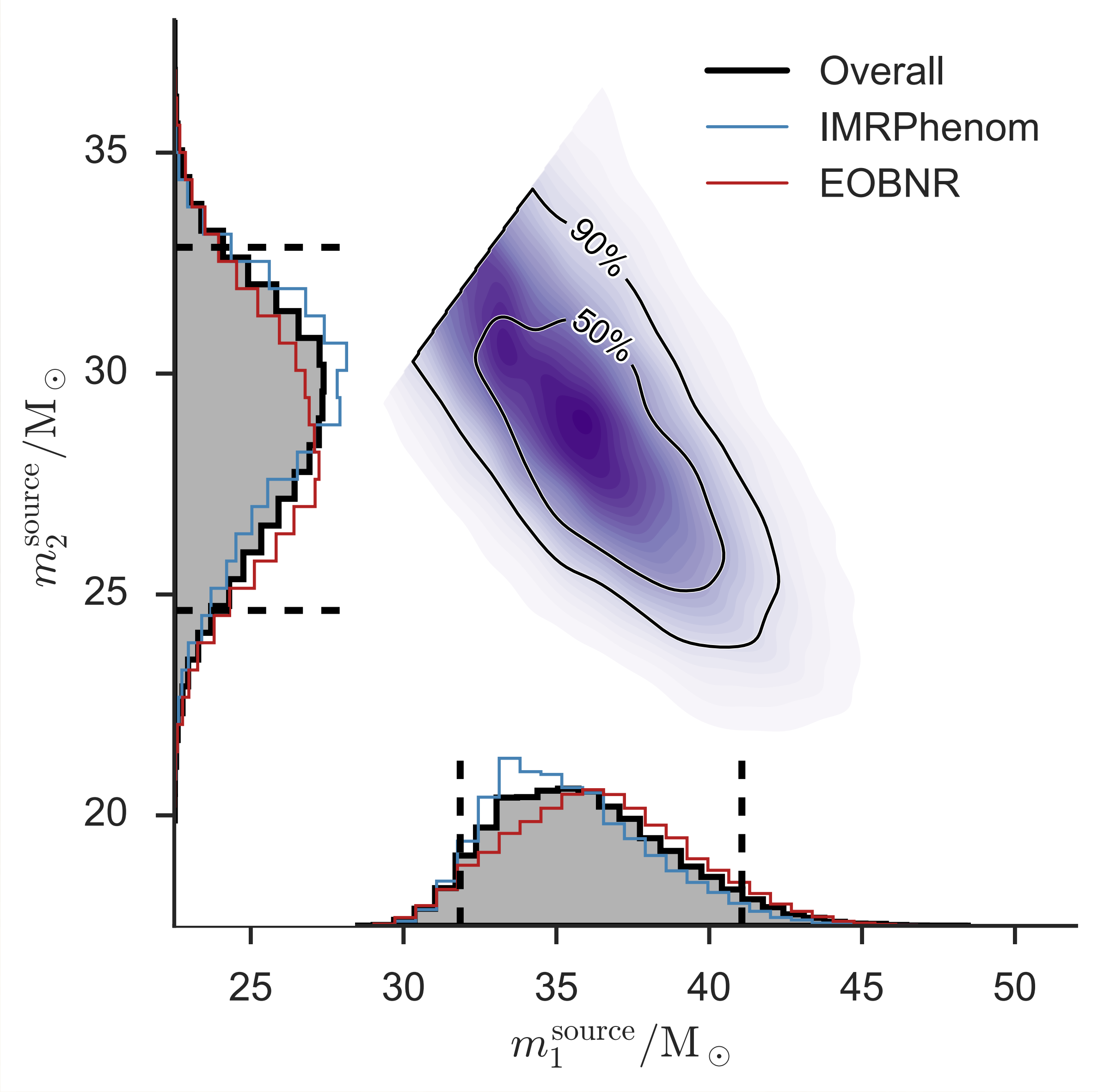
Rapid and robust parameter estimation for gravitational wave observations

IPAM workshop *Big Data in Multi-Messenger Astrophysics*, December 3rd 2021

Jonathan Gair, Albert Einstein Institute (Potsdam)

in collaboration with: **Stephen Green**, **Max Dax**, Jakob Macke, Michael Pürrer, Bernhard Schölkopf, Alessandra Buonanno



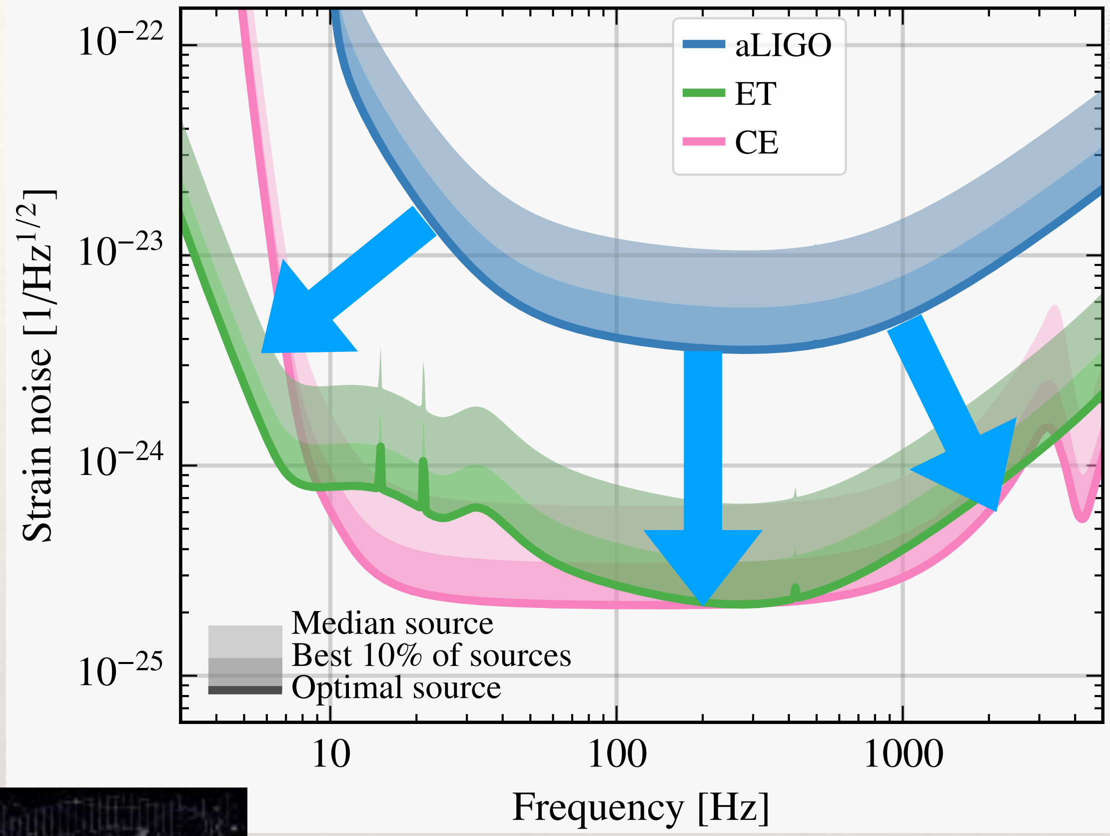


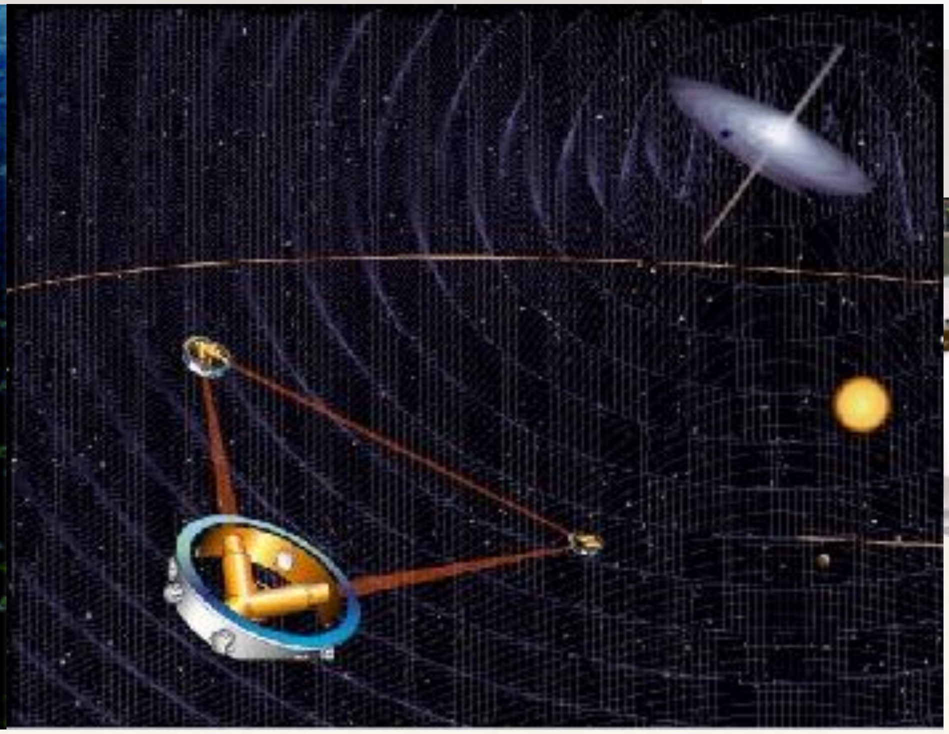
Talk outline

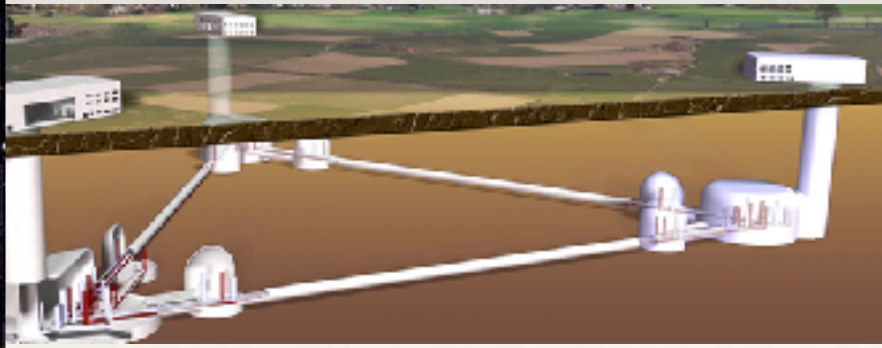
- Current and future gravitational wave detectors
- Gravitational wave parameter estimation
- Fast and accurate parameter estimation
    - Reduced order quadrature
    - Fast waveforms using machine learning methods
    - Neural posterior estimation
- Future challenges and opportunities

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*

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

- Typically we assume the detector output is a linear combination

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

- and that the noise is Gaussian, giving the likelihood

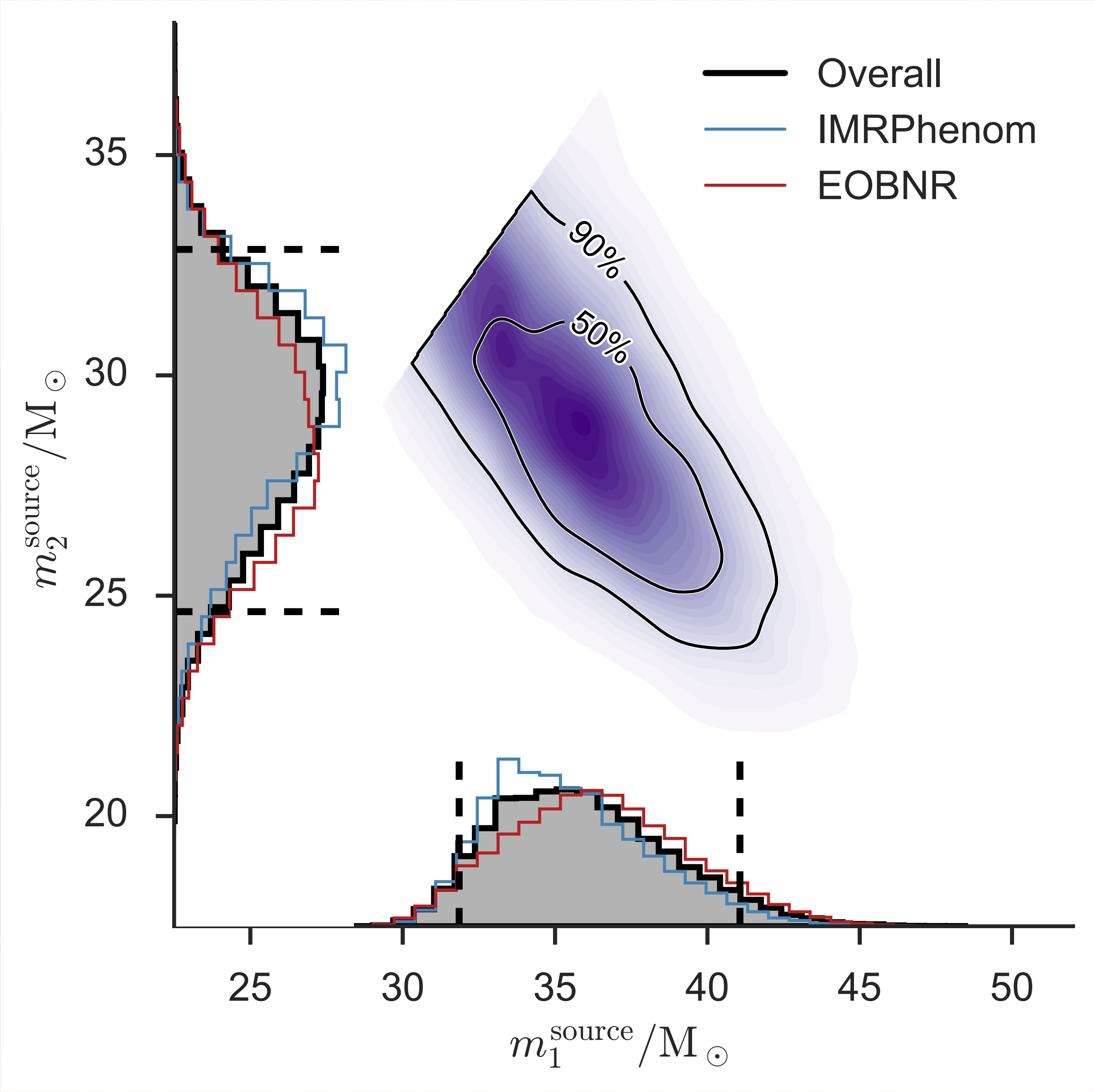
$$p(d|\vec{\theta}) \propto \exp\left[-\frac{1}{2}(d - h(\vec{\theta})|d - h(\vec{\theta}))\right] \qquad (a|b) = \int_{-\infty}^{\infty} \frac{\tilde{a}^*(f)\tilde{b}(f) + \tilde{a}(f)\tilde{b}^*(f)}{S_h(f)} \mathrm{d}f$$

- 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 fraction of that was PE.
- Lag between observation and publication of exceptional events mostly dominated by PE (re-)runs.

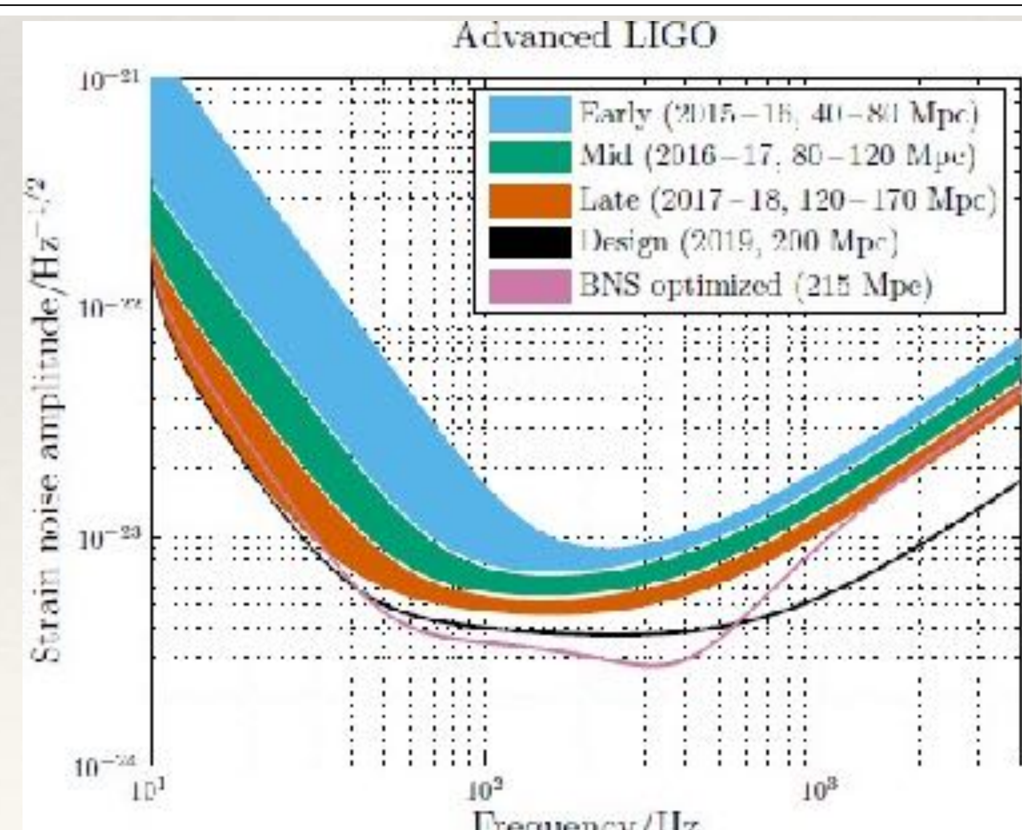
Primary black hole mass	$36^{+5}_{-4} M_\odot$
Secondary black hole mass	$29^{+4}_{-4} M_\odot$
Final black hole mass	$62^{+4}_{-4} M_\odot$
Final black hole spin	$0.67^{+0.05}_{-0.07}$
Luminosity distance	410^{+160}_{-180} Mpc
Source redshift z	$0.09^{+0.03}_{-0.04}$



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 applications: send triggers for EM follow-up.

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



Current solutions: Bayestar

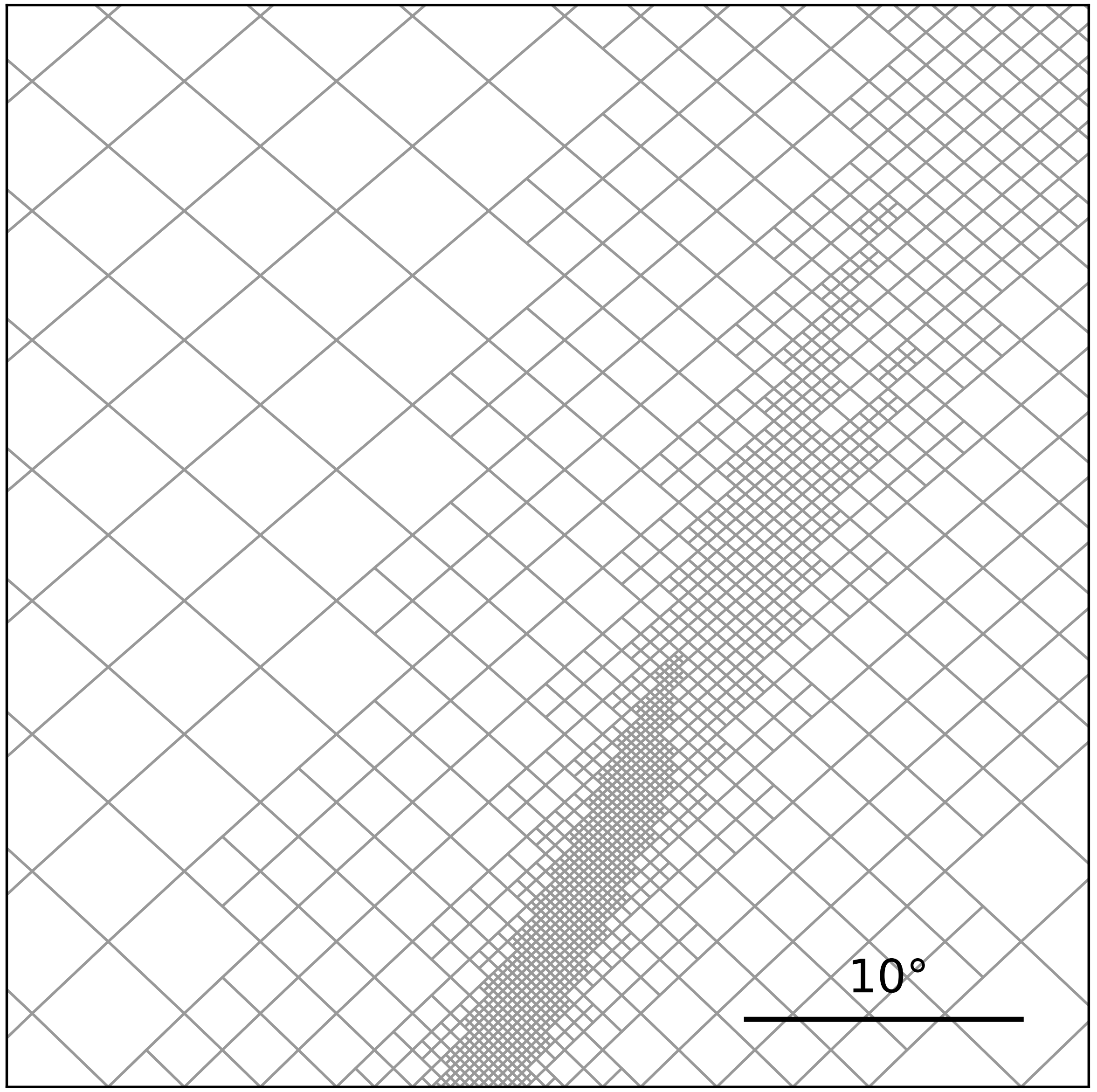
- Rapid PE is needed for multi-messenger applications, to send triggers out for EM follow-up. Current methods do this by approximating the likelihood.

- Bayestar employs the autocorrelation likelihood (likelihood evaluated at MLE parameter values)

$$\exp\left[-\frac{1}{2}\sum_i {\rho_i}^2 + \sum_i \rho_i \Re\left\{e^{-i\gamma_i} z_i^*(\tau_i)\right\}\right]$$

- Rapid marginalisation over parameters other than sky location achieved via integral approximation and look-up tables.

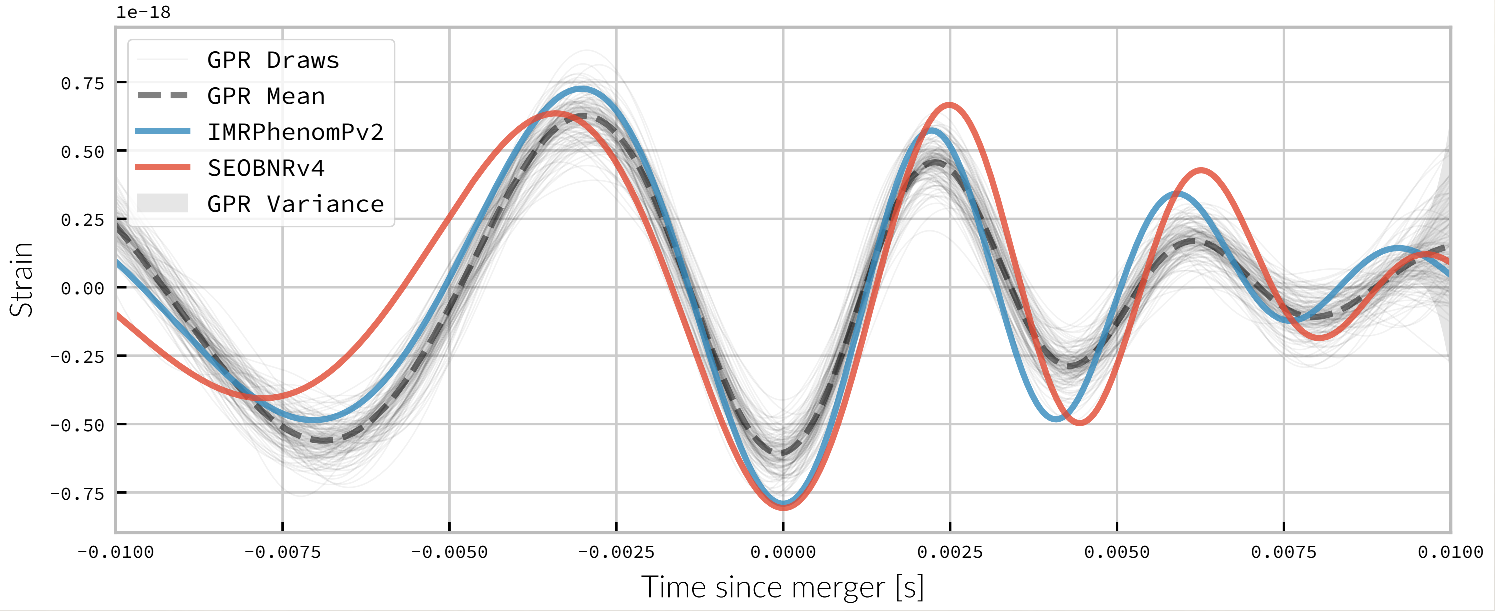
- Result is a sky map probability density.

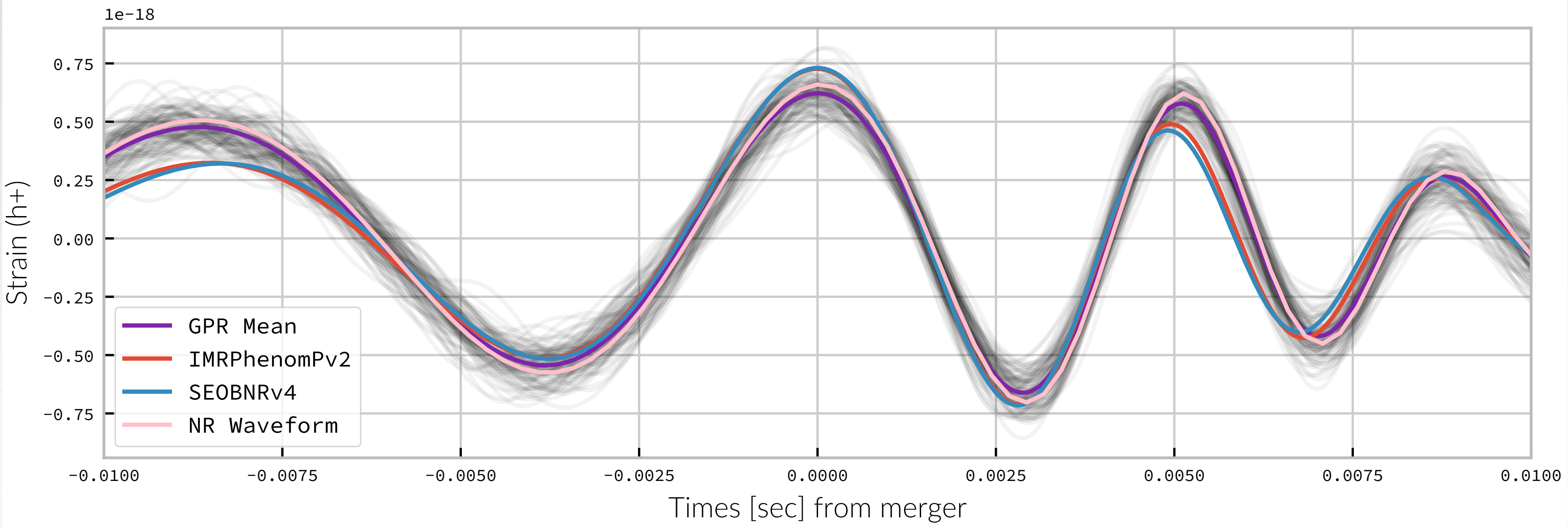


Singer & Price (2015)

Current solutions: Faster waveform models

- Faster waveform model evaluation facilitates faster PE with standard methods (see talk by Leïla Haegel). One example: Gaussian process regression (e.g., Williams et al. 2020) provides faster models that encode modelling uncertainties.





Current solutions: Reduced Order Modelling

- **Reduced order modelling** and **reduced order quadrature** can be used to produce computationally more efficient likelihood calculations (Field, Galley, Pürrer, Tiglio…).

    1. Construct a compact representation of the waveform space.

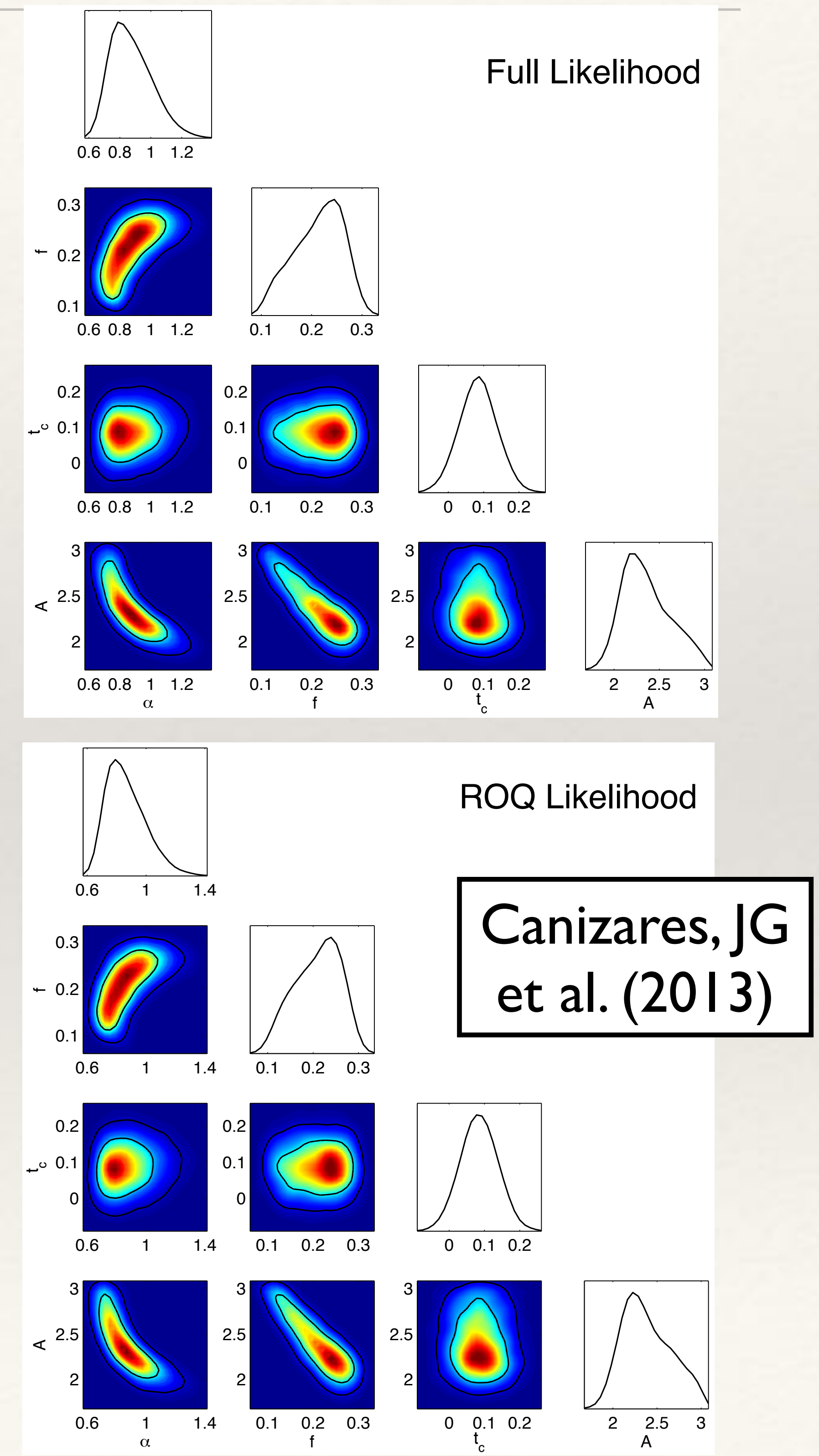
    2. Interpolate basis coefficients efficiently across parameter space.

    3. Construct ROQ approximation to likelihood by requiring waveform to match at specific frequencies

$$(h(\vec{\theta})|d) = 4Re \int_0^\infty \frac{\tilde{h}(\vec{\lambda})\, \tilde{d}^*(f)}{S_n(f)}\, \mathrm{d}f \approx 4Re \sum_{k=1}^{m} \omega_k\, h(F_k; \vec{\lambda})$$

    4. Evaluate waveform at specific F_k's using **surrogates**.

- Typically get order(s) of magnitude computational saving. Important tool for current LIGO analyses. Extensions to other waveform models and LISA non-trivial!



Canizares, JG et al. (2013)

New approaches: Neural posterior estimation

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

- Alternative: *simulation based inference*. Construct a neural network that generates samples from a distribution, $q(\theta|d)$, that approximates the target distribution, in this case the parameter posterior distribution, $p(\theta|d)$. Train the neural network by minimising the average *cross-entropy* with the true distribution

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

- Compute loss by simulation

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

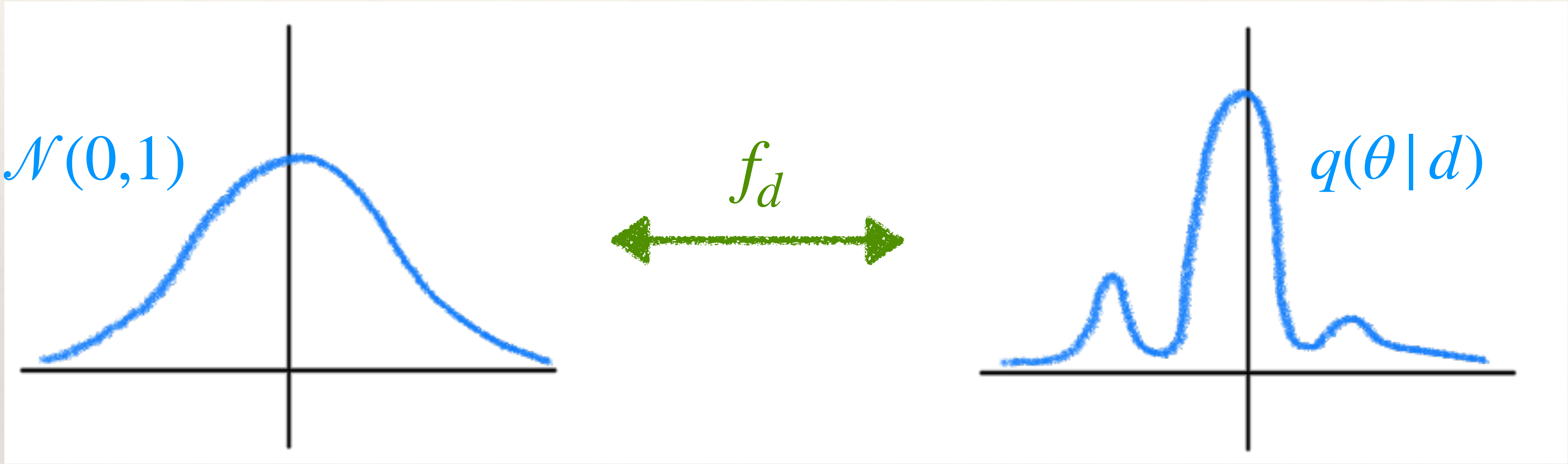
$$\textbf{Simulate } d^{(i)} \sim p(d|\theta^{(i)}); \quad d^{(i)} = h(\theta^{(i)}) + n^{(i)} \text{ with } n^{(i)} \sim p_{S_n}(n)$$

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

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

Normalizing flows

- A normalising flow maps 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.

Normalizing flows

- Build normalising flow from a sequence of *coupling transforms*

$$c_{d,i}(u) = \begin{cases} u_i & \text{if } i \leq D/2 \\ c(u_i; u_{i:\frac{D}{2}}, d) & \text{if } i > D/2 \end{cases}$$

- The coupling transforms must be differentiable and invertible.
- We use *spline flows* (Durkan et al. 2019), based on rational quadratic spline interpolation between a set of knots.
- A sequence of transforms can represent very complicated functions.

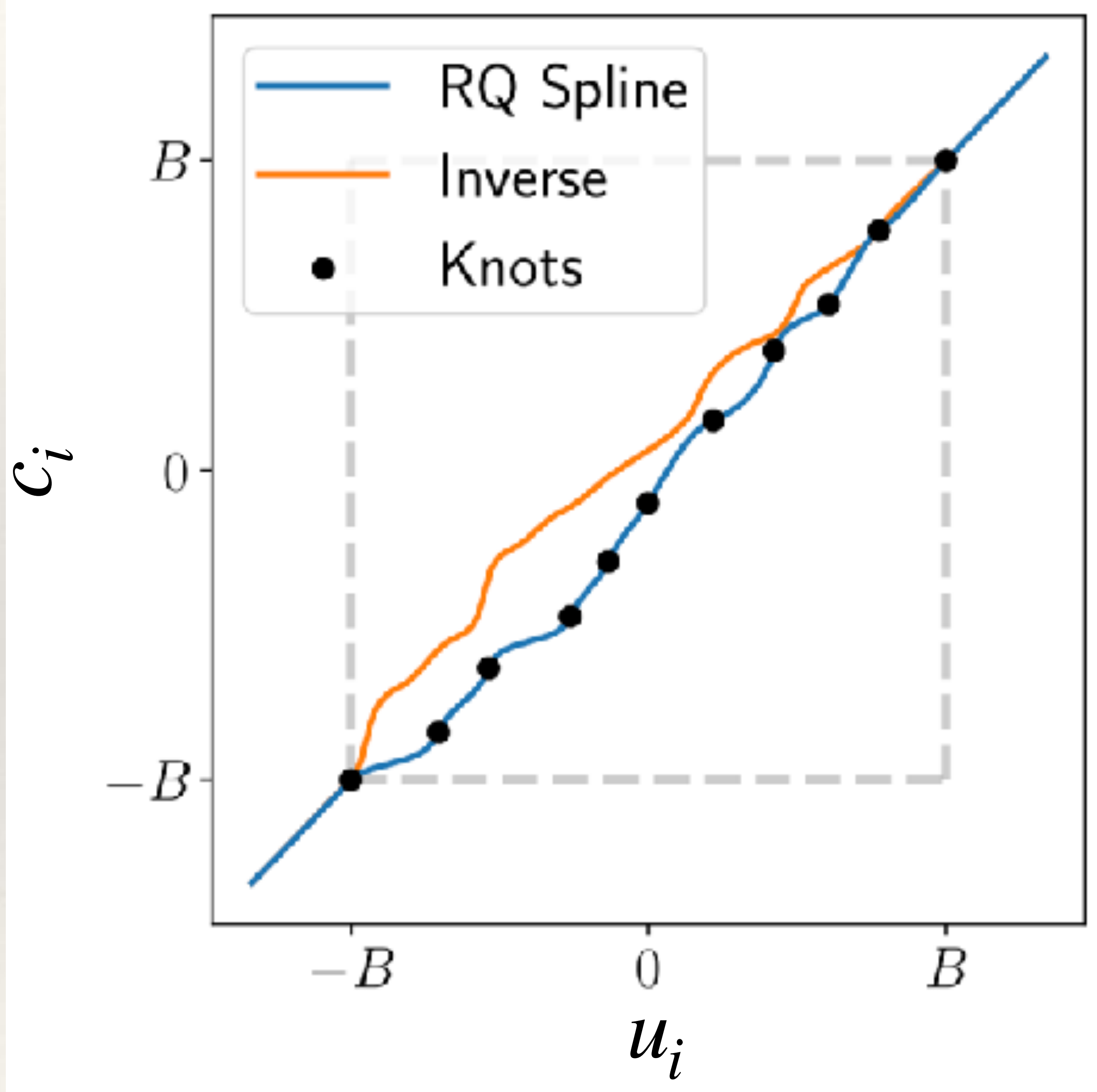


Figure: Durkan *et al* (2019)

NPE refinements: embedding network

- The existence of reduced bases shows that waveform bases can be compressed. Could impose this by hand, but more robust to learn this using an *embedding network*.

$N_{\mathrm{in}} = n_{\mathrm{det}}$ x 24099

(8033 frequency bins) x (Re + Im + PSD)
8s segment, sampled at 1024Hz,
f_{low} = 20Hz

Linear projection

Seed with principal components of *noise-free* waveforms

inductive bias to recognise waveforms

n_{det} x 400

Fully connected residual network

48 hidden layers

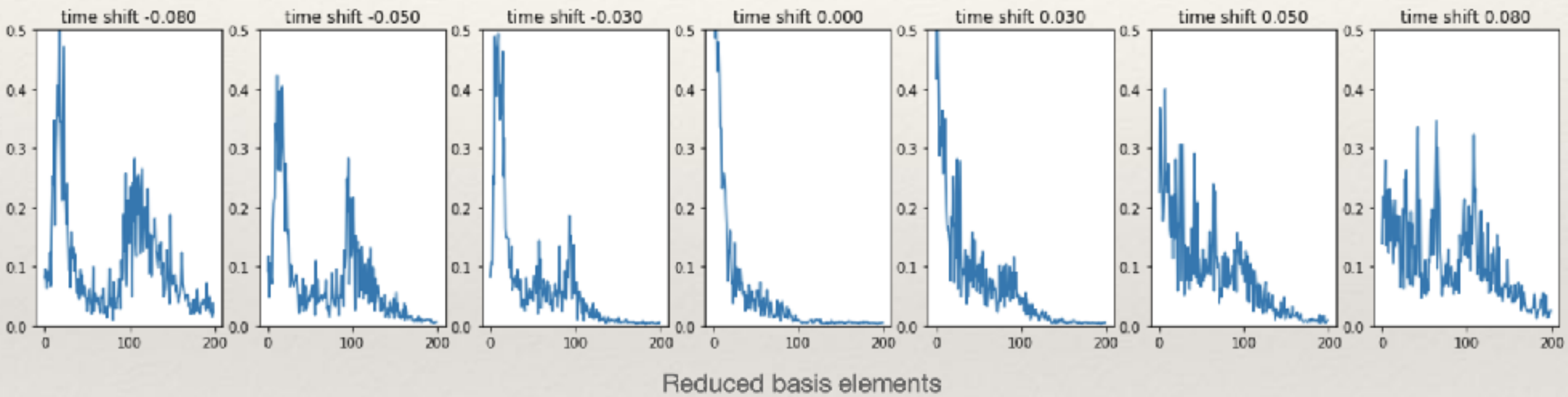
non-linear compression

N_{out} = 128

to flow

NPE refinements: group equivariant NPE

- Representing the time of coalescence, t_I, requires many reduced basis elements. Uses up a lot of training resources and freedom within the network.



- A change in time of coalescence in a single detector corresponds to a (trivial) transformation of the data and template. If the time shift is known, the waveform can be aligned and the learning process significantly simplified.

- Don't know this *a priori* and not an exact symmetry for a detector network.

NPE refinements: group equivariant NPE

- Introduce a blurred estimate of t_I, $\hat{t}_I$, into the parameter space.
- In training and inference, follow a Gibbs sampling procedure

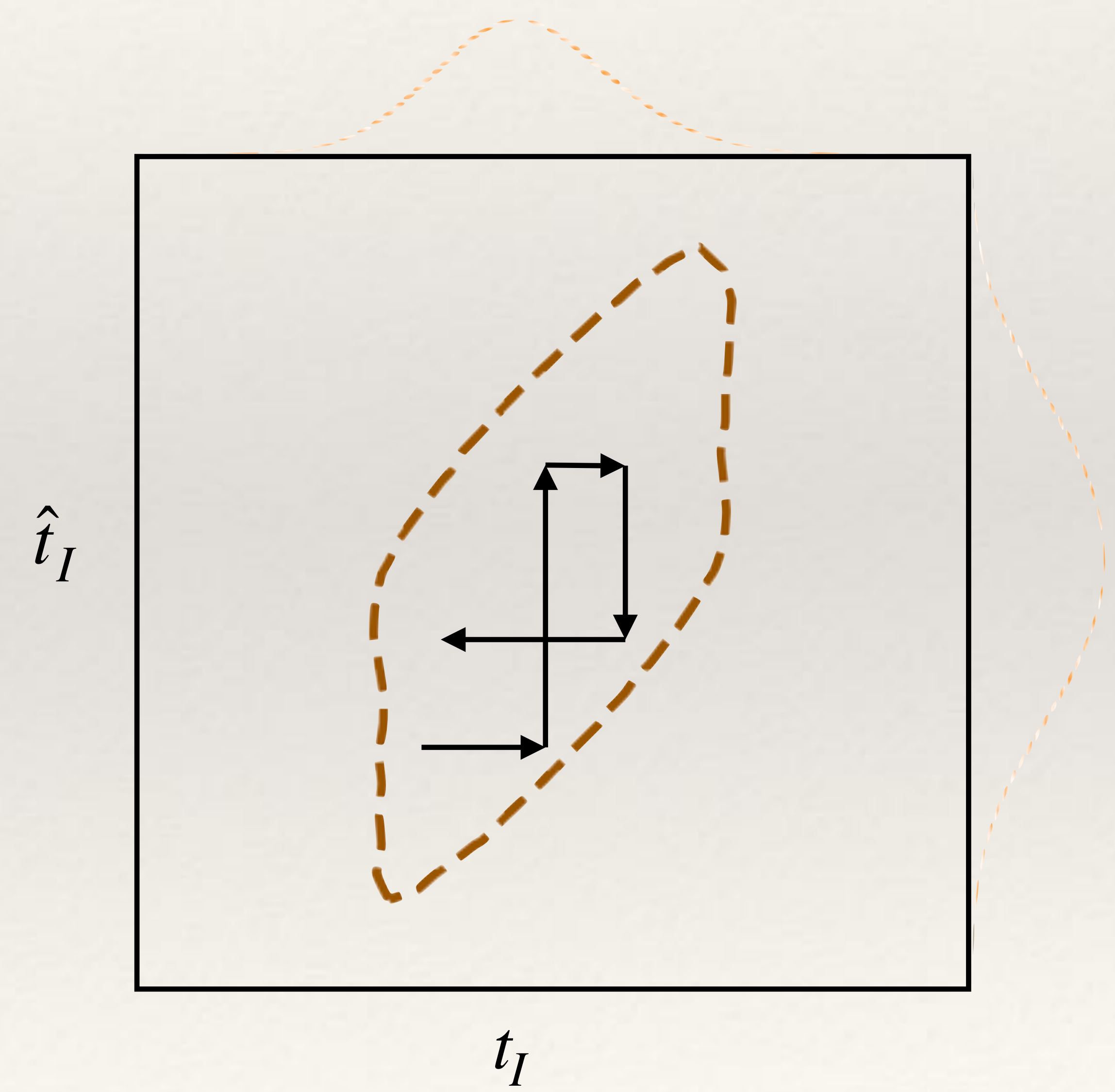
  1. Align data based on $\hat{t}_I$

  $$\theta \sim q(\theta | T_{-\hat{t}_I}(d), \hat{t}_I)$$

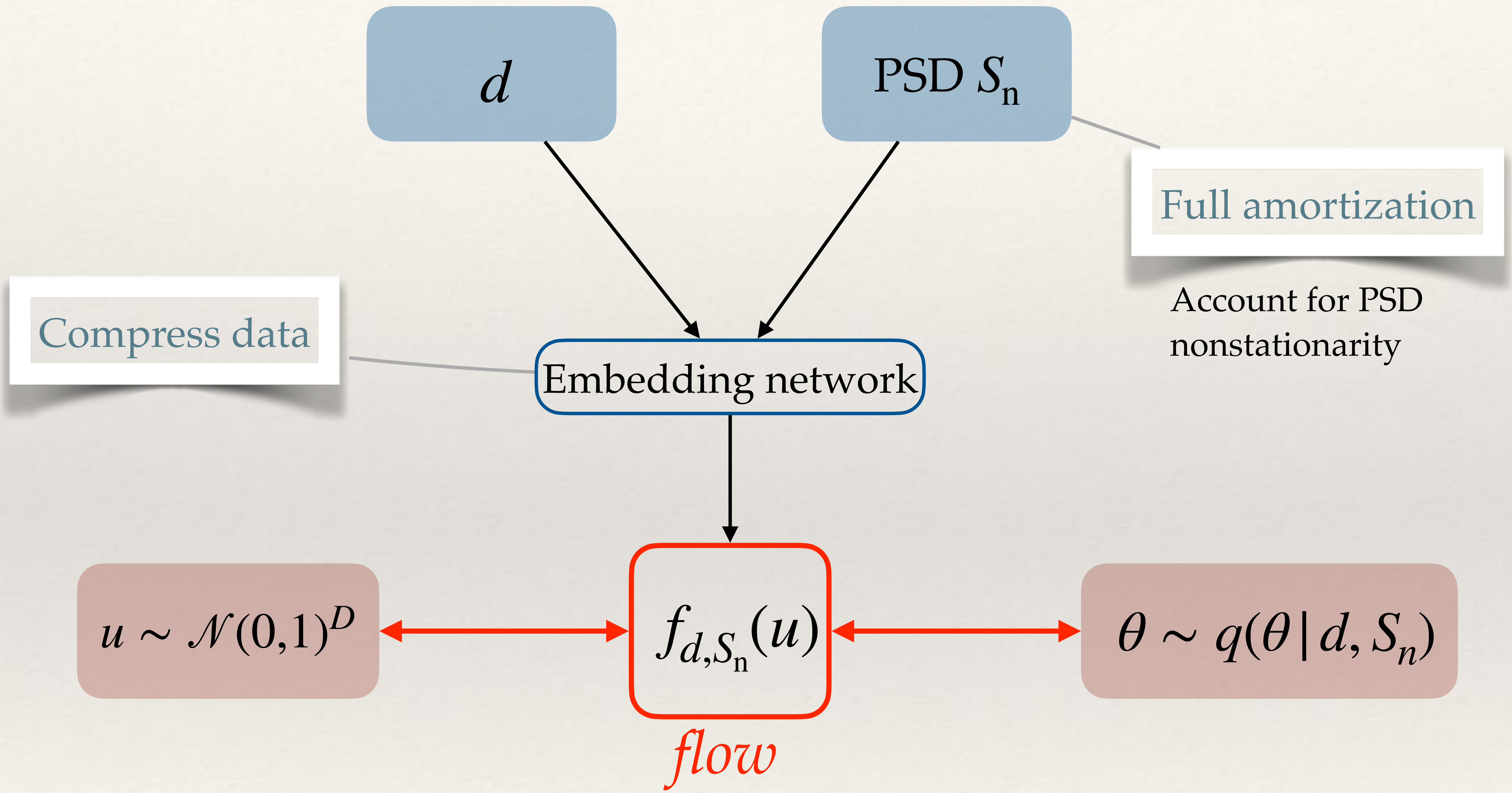
  2. Sample $\hat{t}_I$ from a fixed kernel

  $$\hat{t}_I \sim p(\hat{t}_I | t_I)$$

- Converges in O(10) iterations.
- GNPE exploits (near-) symmetries to simplify the learning task.



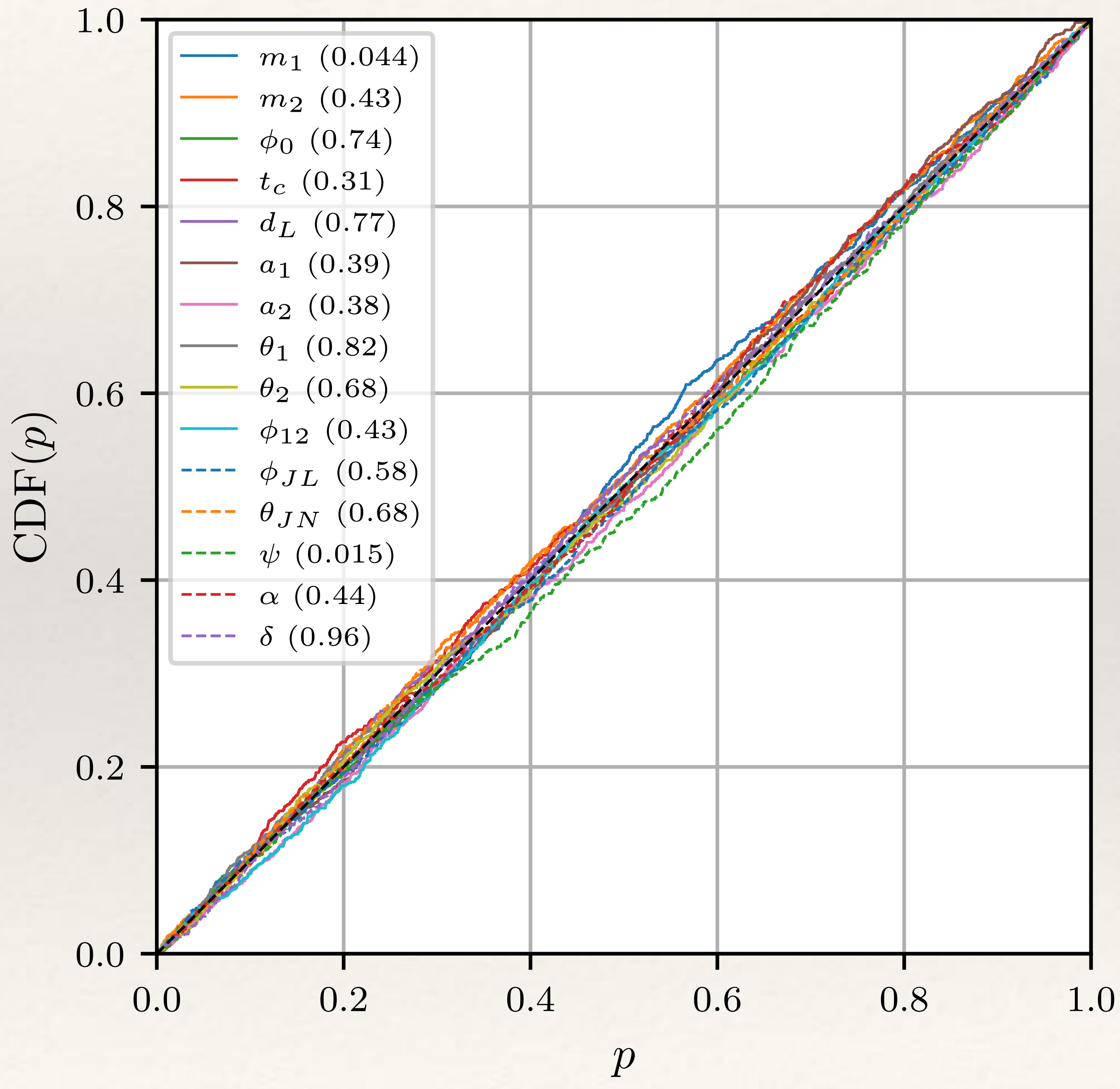
NPE network



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

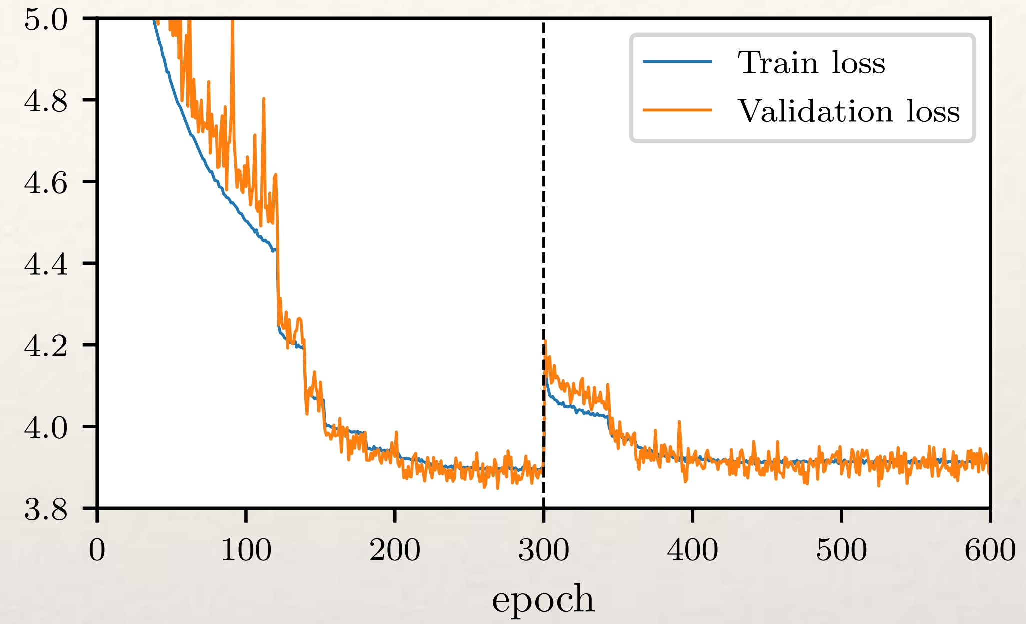
NPE validation

- Check **internal** or **within-distribution** consistency of network by generating a p-p plot.
- Check **external** or **out-of-distribution** consistency by comparing to posterior distributions computed for real observations using standard stochastic samplers.



NPE validation: GWTC-1 BBHs

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

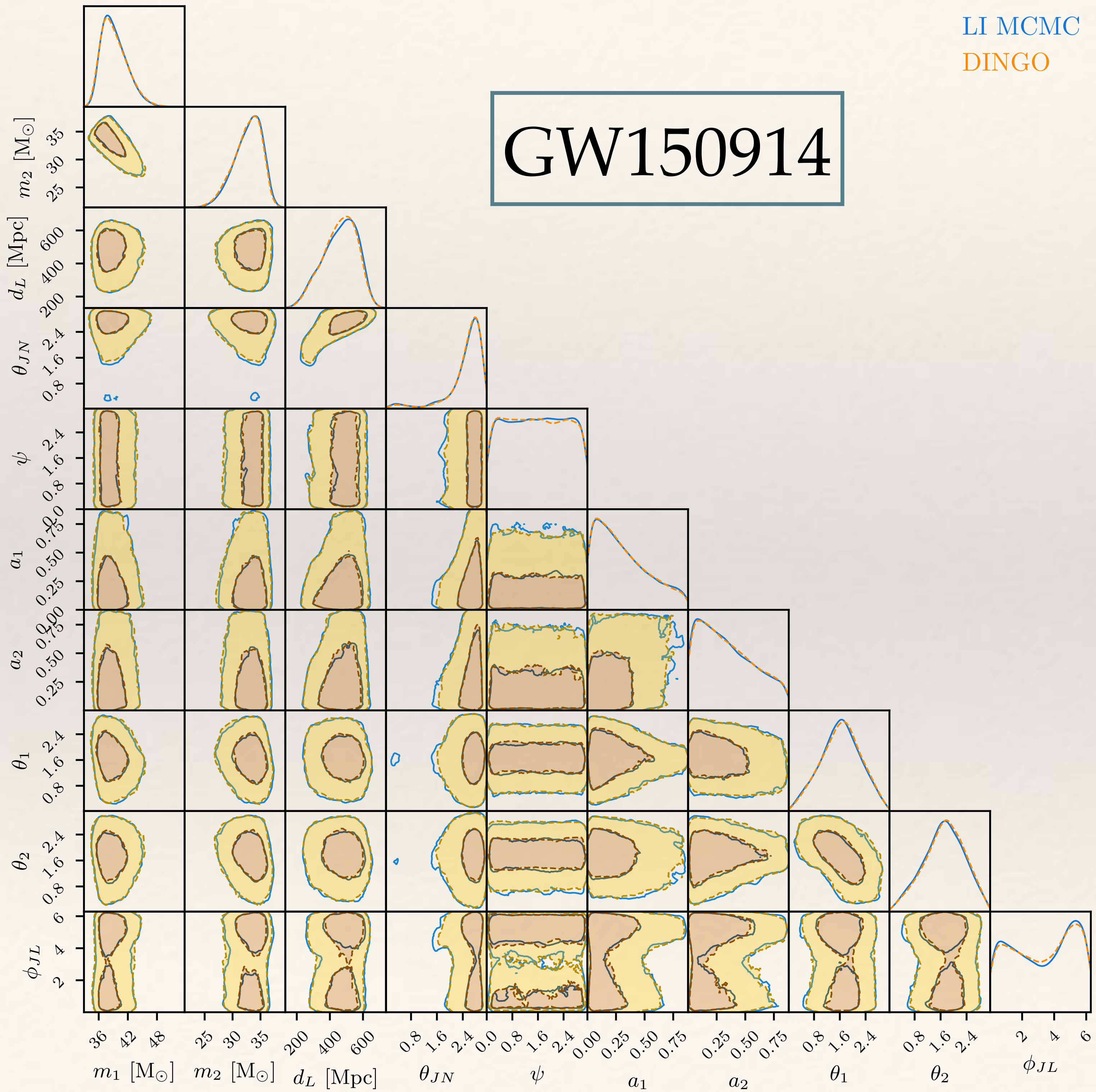


- + stationary Gaussian noise realisations consistent with given PSD

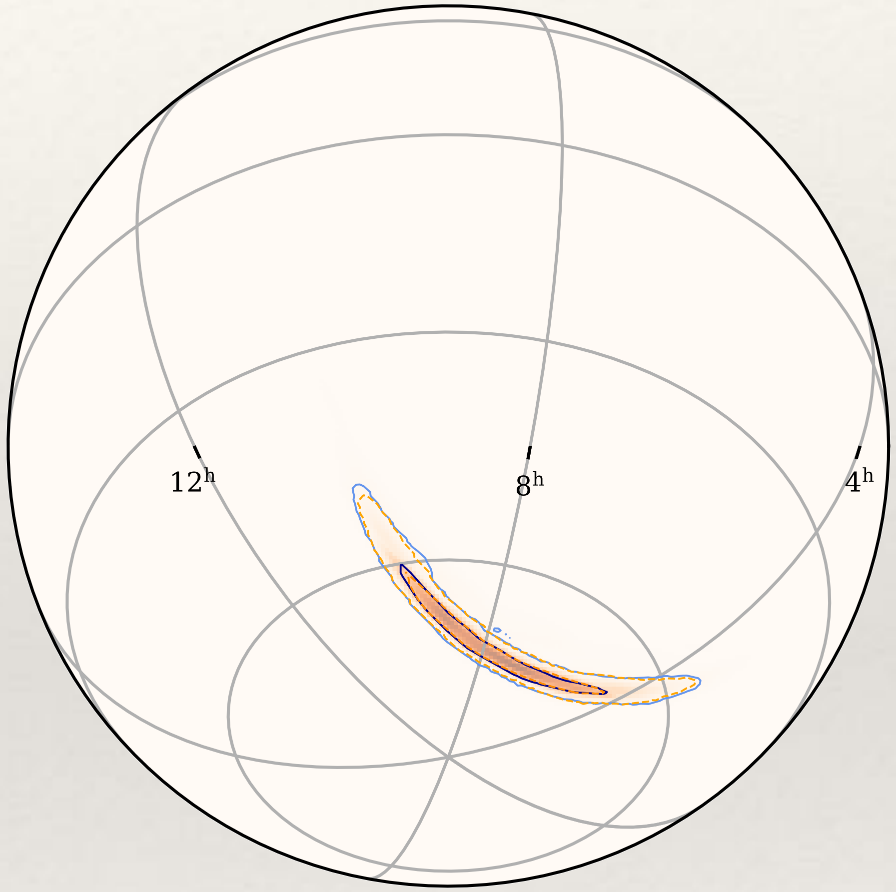
- Train several neural networks based on different noise level / number of detectors / distance range:

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

NPE validation: GWTC-1 BBHs



GW150914

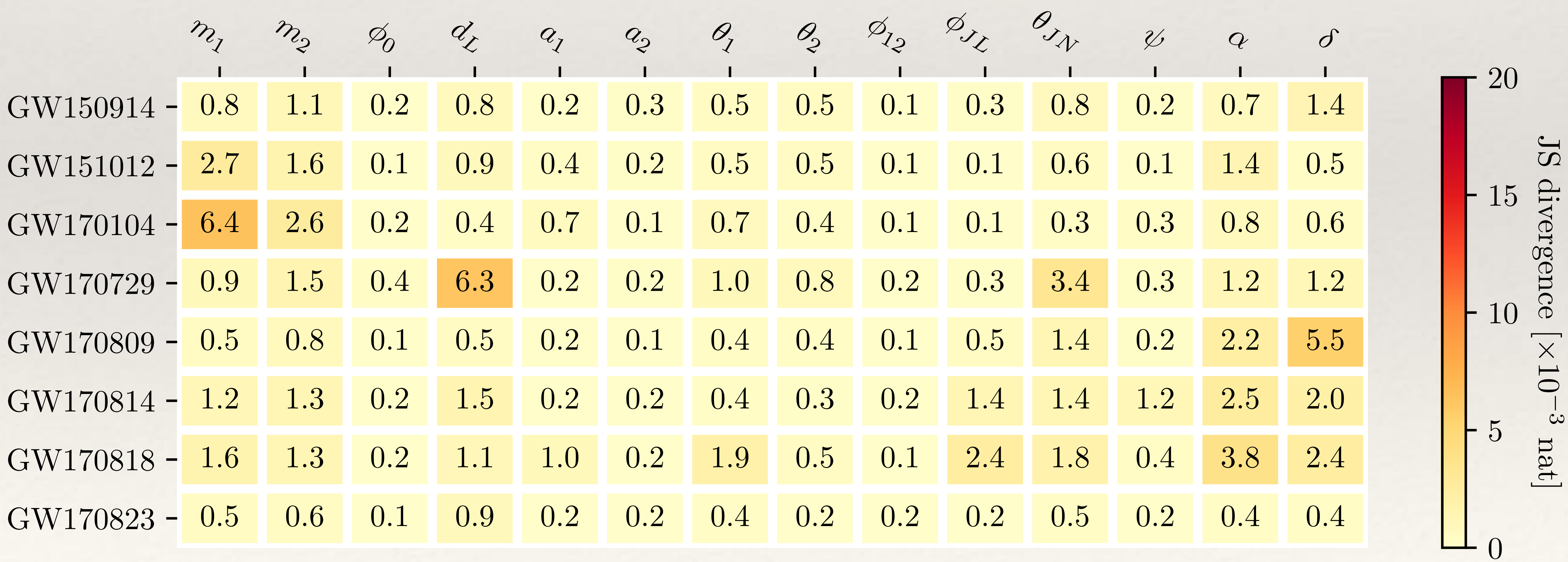


50,000 samples in ~ 20 s

NPE validation: 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*.



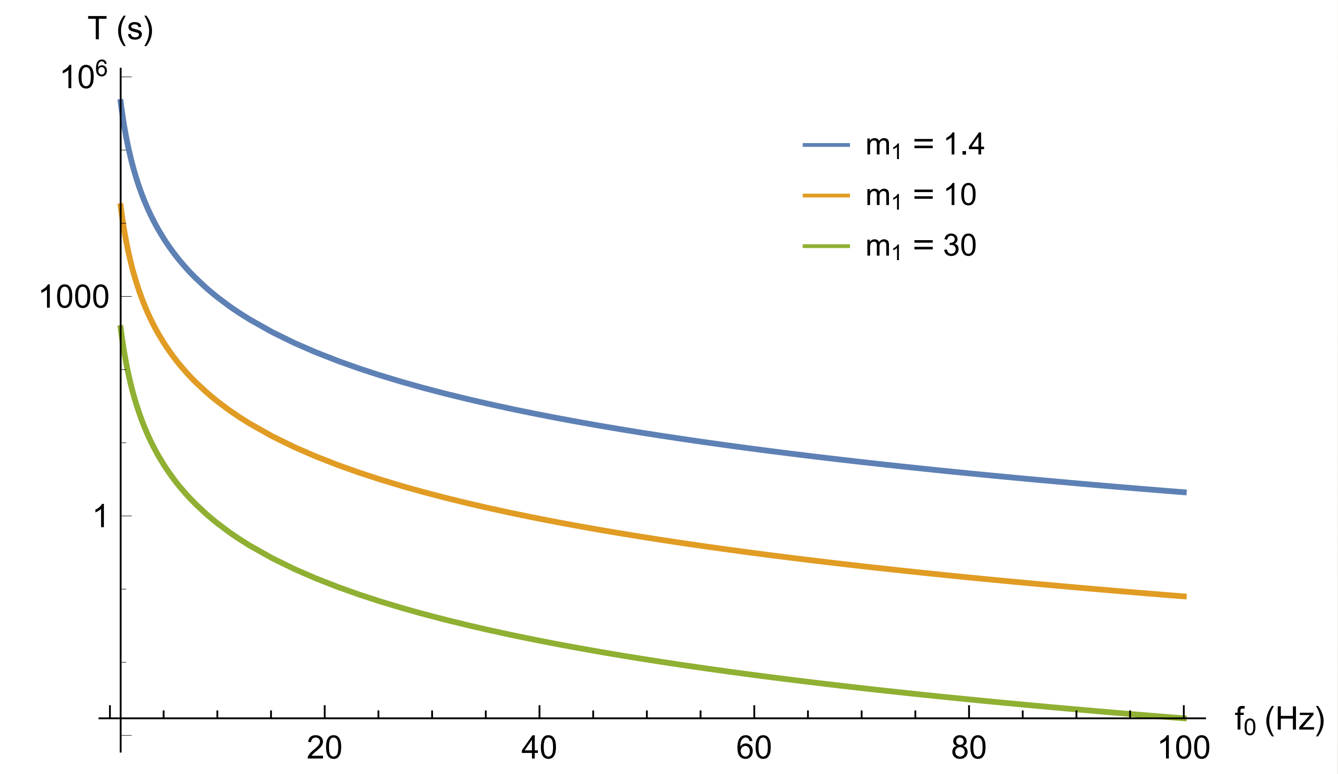
	m_1	m_2	ϕ_0	d_L	a_1	a_2	θ_1	θ_2	ϕ_{12}	ϕ_{JL}	θ_{JN}	ψ	α	δ
GW150914	0.8	1.1	0.2	0.8	0.2	0.3	0.5	0.5	0.1	0.3	0.8	0.2	0.7	1.4
GW151012	2.7	1.6	0.1	0.9	0.4	0.2	0.5	0.5	0.1	0.1	0.6	0.1	1.4	0.5
GW170104	6.4	2.6	0.2	0.4	0.7	0.1	0.7	0.4	0.1	0.1	0.3	0.3	0.8	0.6
GW170729	0.9	1.5	0.4	6.3	0.2	0.2	1.0	0.8	0.2	0.3	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

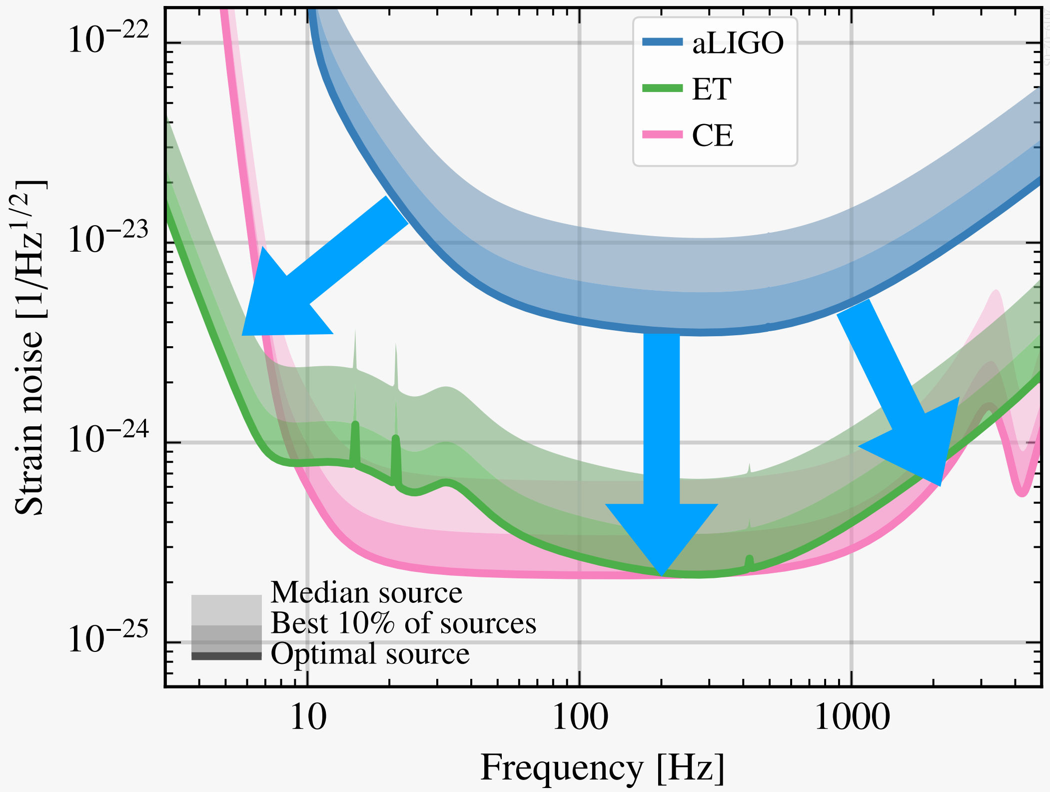
Future challenges: long waveforms

- Inspiral time from frequency f_0 for binary with $m_1=m_2$ is approximately

$$T \sim 980\,\mathrm{s} \left(\frac{m_1}{1.4\,M_\odot}\right)^{-\frac{5}{3}} \left(\frac{f_0}{10\,\mathrm{Hz}}\right)^{-\frac{8}{3}}$$

- As lower frequency cut off decreases from 30 Hz -> 6 Hz -> 2 Hz, BNS observation time increases from ~1 minute to ~1 hour to ~1 day.
- LISA sources observable for ~1 year.
- Need strategies to deal with long waveforms, e.g., reduced bases, relative binning, GNPE for intrinsic parameters.



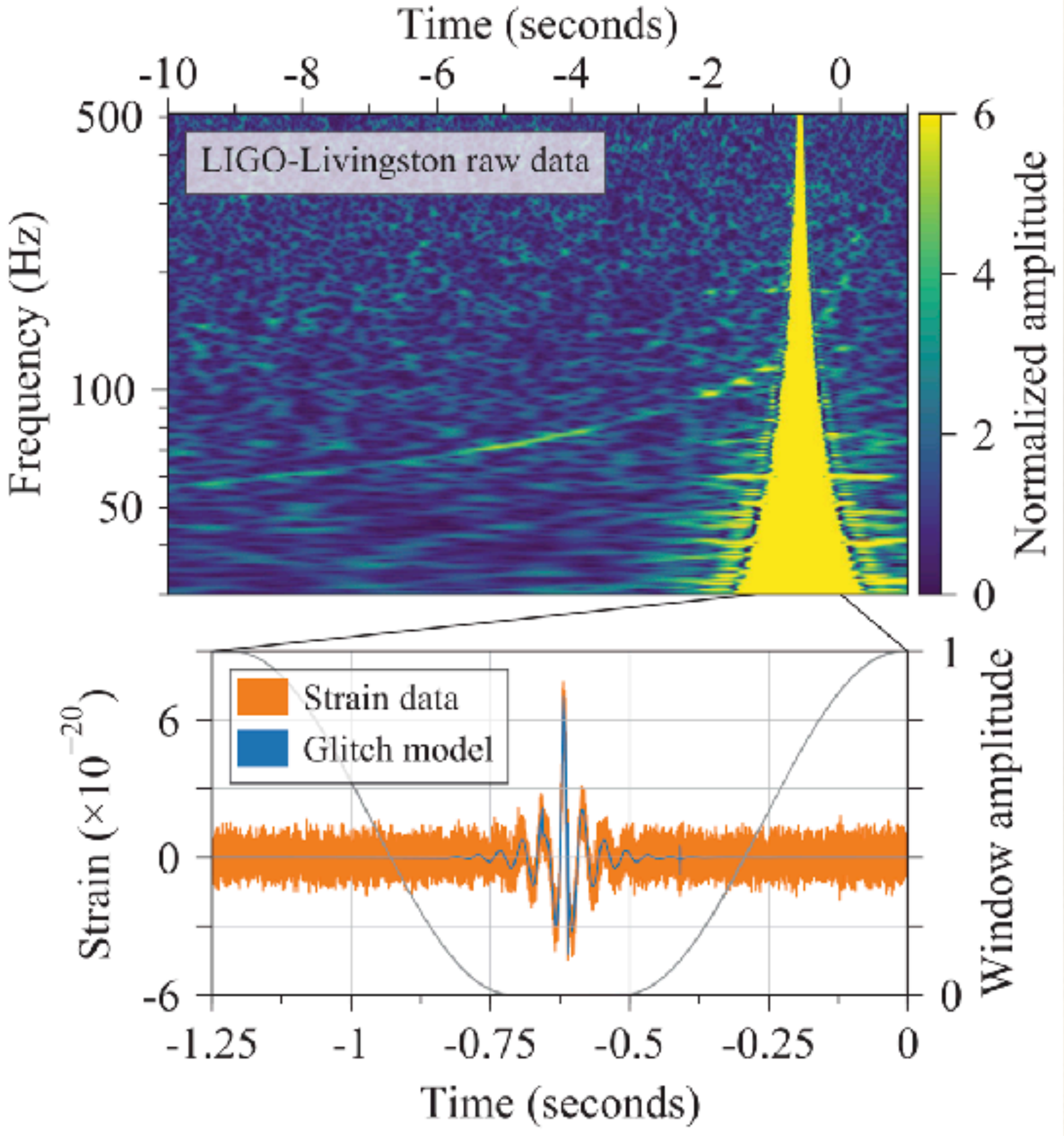


Future challenges: non-stationary noise

- BBH mergers are short enough that noise is approximately stationary.
- For longer signals this is less likely to be true: an issue for 3G detectors and LISA.
- NPE can readily accommodate alternative noise specifications, provided that we can simulate these (or use observed noise).

$$\textbf{Sample } \theta^{(i)} \sim p(\theta), \quad i = 1, \ldots, N$$
$$\textbf{Simulate } d^{(i)} \sim p(d|\theta^{(i)})$$
$$d^{(i)} = h(\theta^{(i)}) + n^{(i)}$$
$$\text{with } n^{(i)} \sim p_{S_n}(n)$$

Use other noise models



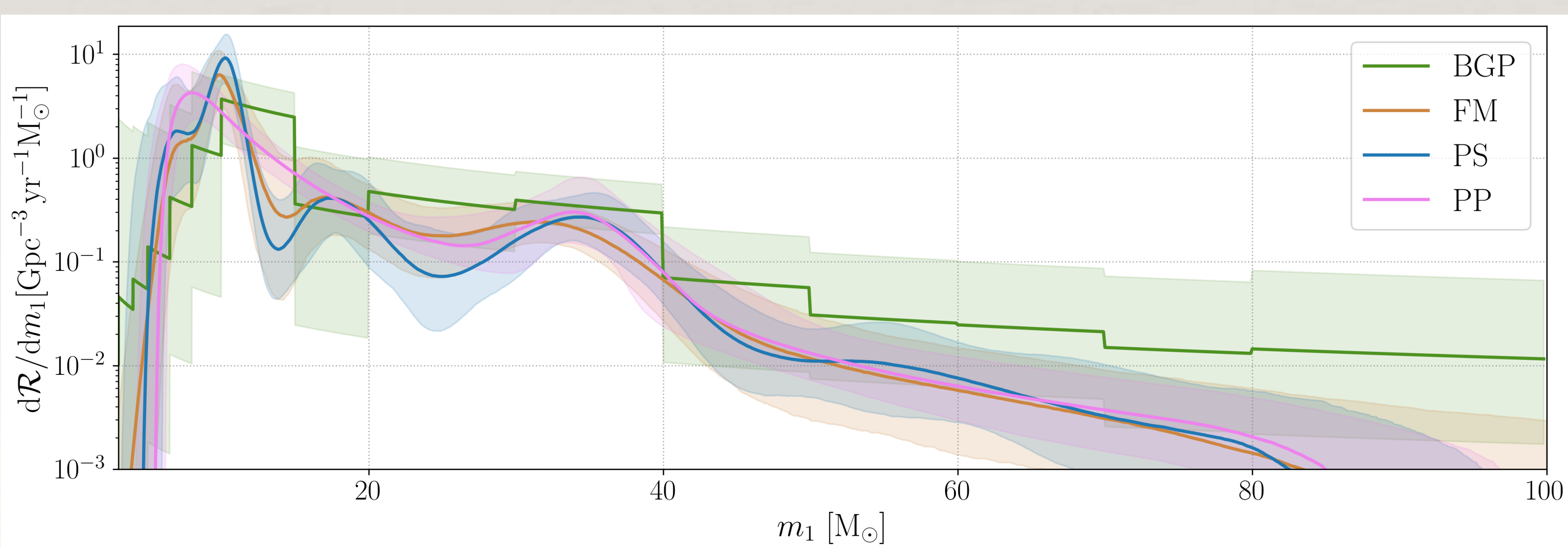
Future challenges: population inference

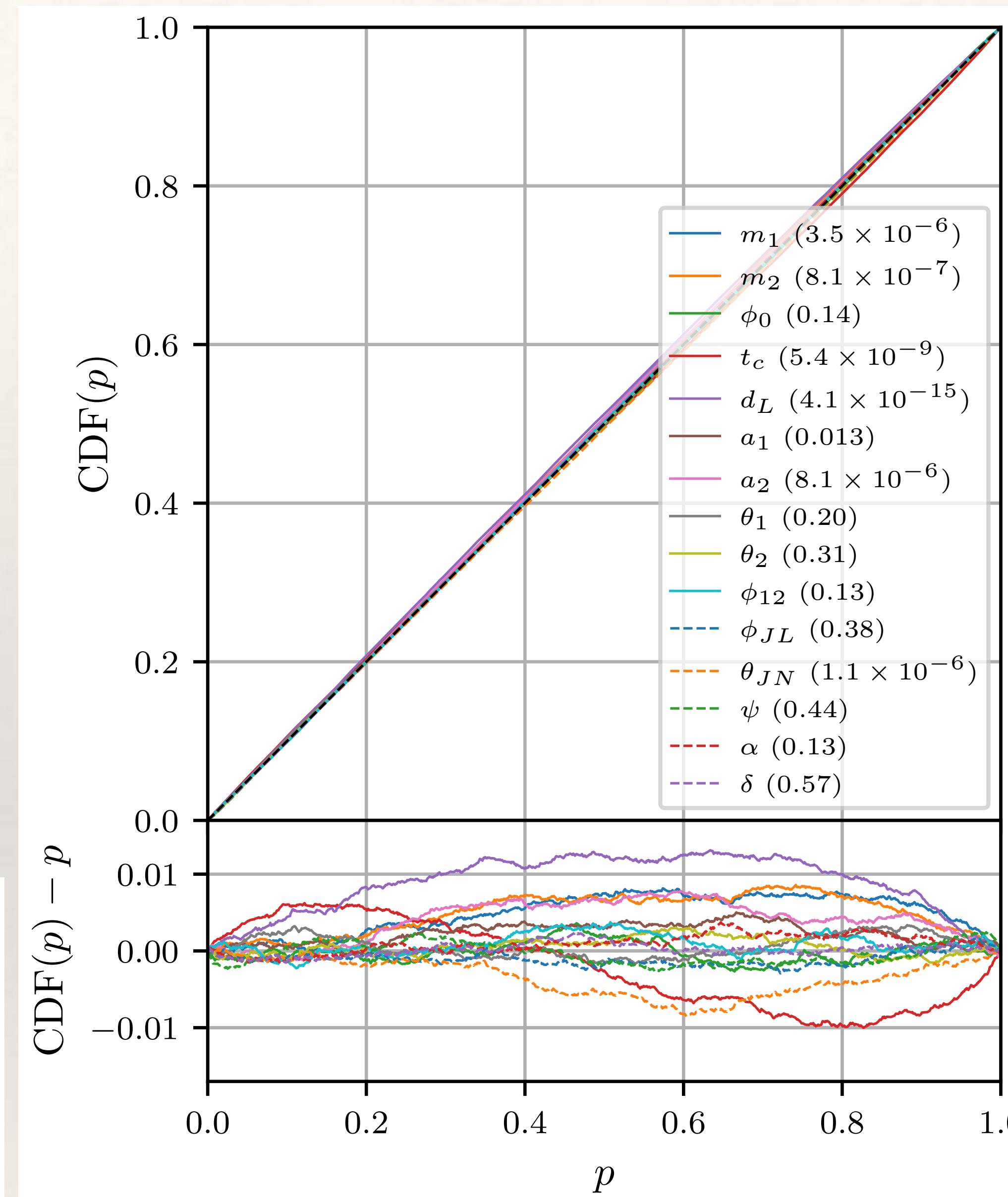
- Population inference uses Bayesian Hierarchical Modelling

$$p(\vec{\lambda}|\{\mathbf{d}_i\}) \propto p(\vec{\lambda}) \prod_{i=1}^{N} \int p(\mathbf{d}_i|\vec{\theta}_i)\, p(\vec{\theta}_i|\vec{\lambda})\, \mathrm{d}\vec{\theta}_i$$

- Inference relies on reweighting of posterior samples to alternative population parameter values.

- NPE's ability to generate many samples, hence resolving tails, could be crucial for accuracy.

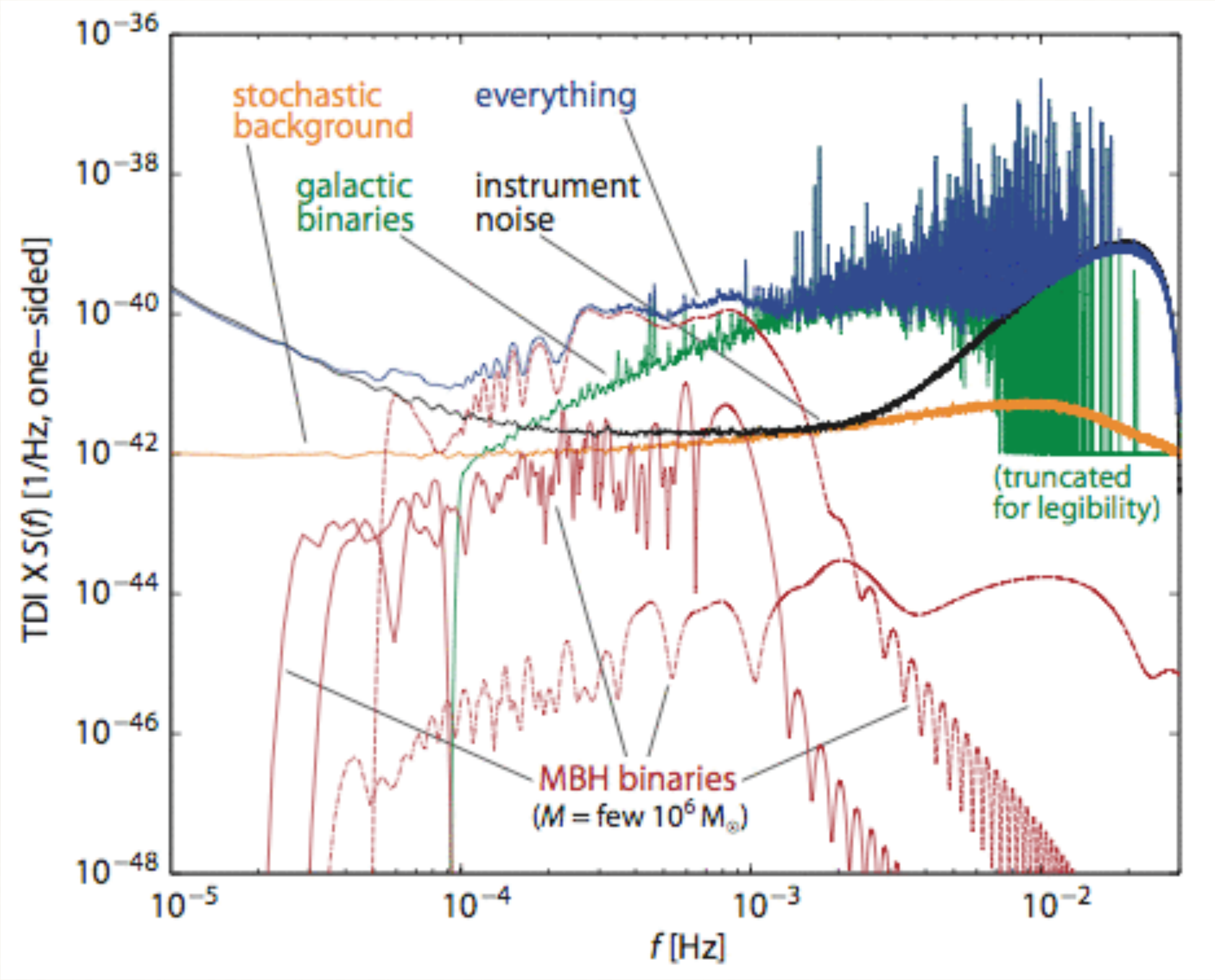




- **But** need to verify that NPE has trained sufficiently in the tails of the distribution.

Future challenges: overlapping sources

- Future detectors will observe sources that overlap in time and frequency.
- Need new NPE architectures to handle this, e.g.,
  - Fixed maximum number of sources - increase output dimensionality.
  - Iterative subtraction. Network operates on residuals.
- Training more complex, as network must learn multi-source and multi-residual data sets.



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**.
- Various approaches to accelerate conventional methods are being explored, including reduced order and surrogate modelling and Gaussian process regression.
- **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.
- **Group equivariant NPE** can be used to simplify training of a neural network by exploiting near symmetries.
- **Future detectors pose many developmental challenges**: long waveforms, non-stationary noise, new sources, overlapping sources, population inference….