SEOBNRv5PHM_NNSur: a fast neural network waveform surrogate for generically precessing binaries

Chris Whittall & Geraint Pratten
BritGrav 2026
University of Cardiff

University of Southampton
UNIVERSITY OF BIRMINGHAM | GRAVITATIONAL WAVE ASTRONOMY

Reduced order surrogate modelling

- **Accurate** and **fast** waveform models needed for gravitational wave data analysis.

- Next-generation detectors need significant extension in waveform model coverage (**duration**, **precession** etc.)

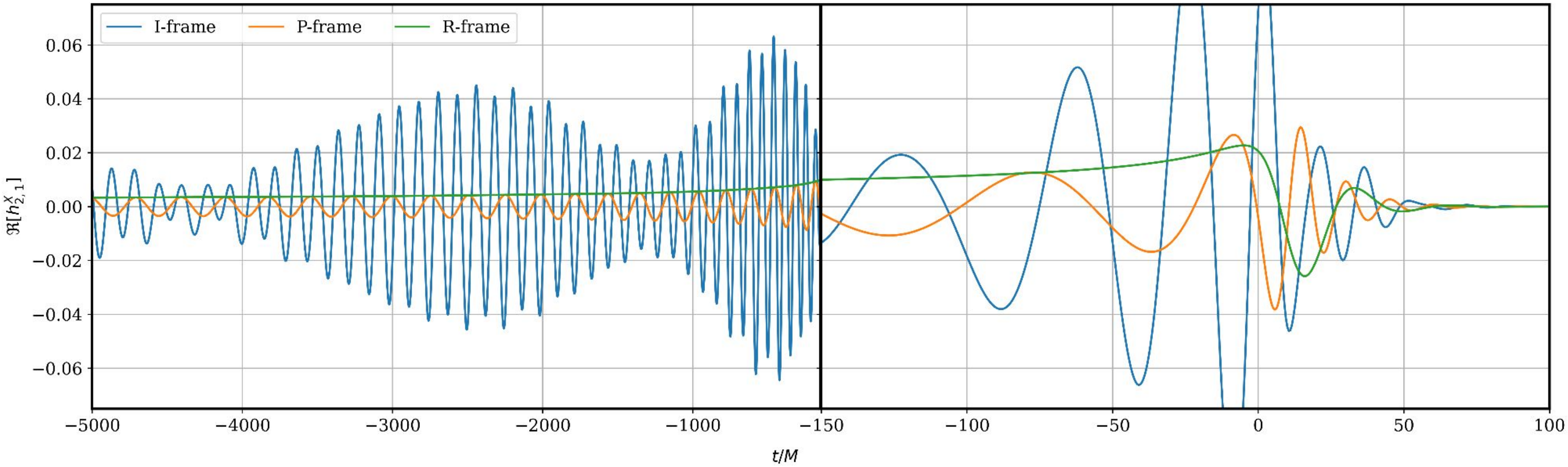
- Surrogate modelling: given a **slow base model**, train a model to **predict** its output for arbitrary parameters in some range.

- Did this for **SEOBNRv5PHM**, a fully precessing, quasicircular EOB model. [Ramos-Buades+ (2303.18046),...]

- Three basic stages: **decomposition**, **compression** and **prediction**.

Step 1: Waveform decomposition

Analyse precessing waveforms in appropriate frames of reference



I-frame: inertial source frame
P-frame: (non-inertial) co-precessing frame
R-frame: (non-inertial) co-rotating frame

Schmidt+ (1012.2879), Boyle+ (1110.2965), Schmidt+ (1207.3088), Boyle (1302.2919)

Step 1: Waveform decomposition

Break waveform up into pieces with simpler morphology

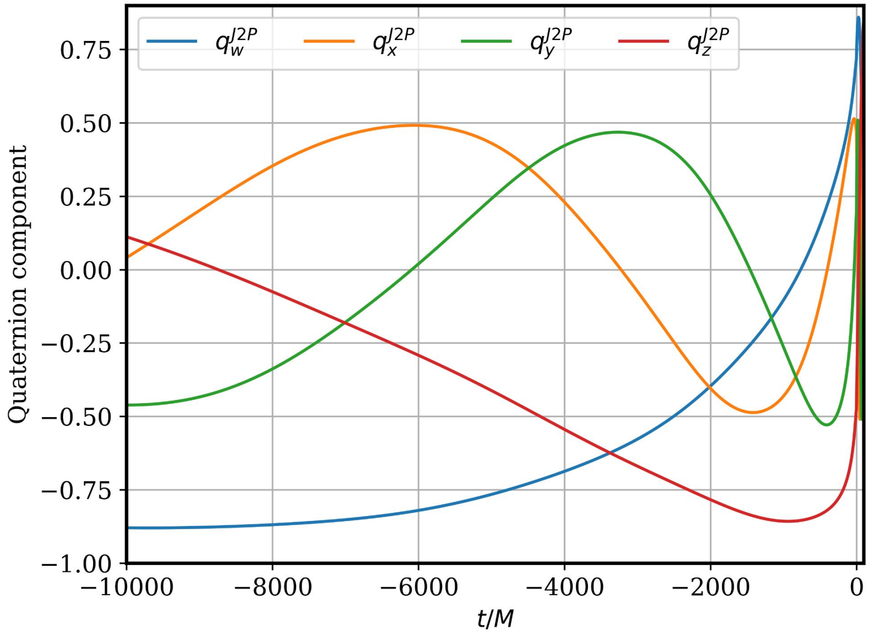
- **Quaternions describing rotation from inertial to co-precessing frame:**

$$\mathbf{q} = \cos\left(\frac{\theta}{2}\right) + \sin\left(\frac{\theta}{2}\right)\hat{\mathbf{n}}$$

- Orbital phase and co-orbital frame modes:

$$\phi_{\rm orb}(t) := \frac{1}{2}\arg\left[h^P_{2,2}(t)\right],$$

$$h^R_{\ell m}(t) = h^P_{\ell m}(t)e^{-im\phi_{\rm orb}(t)}.$$



Step 1: Waveform decomposition

Break waveform up into pieces with simpler morphology

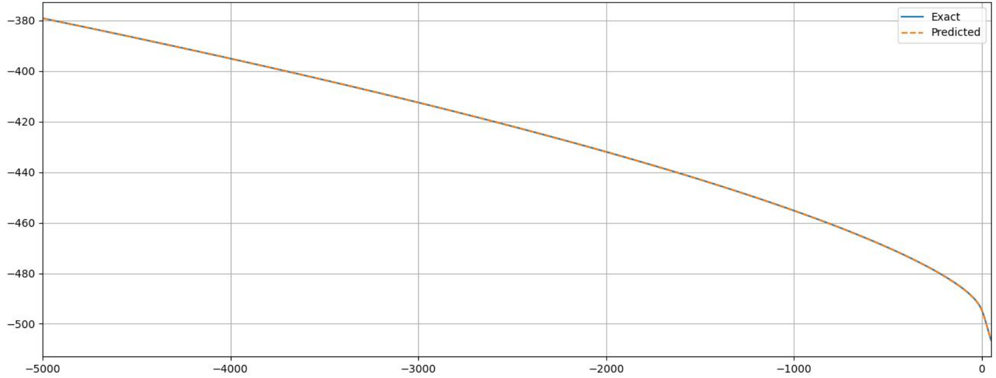
- Quaternions describing rotation from inertial to co-precessing frame:

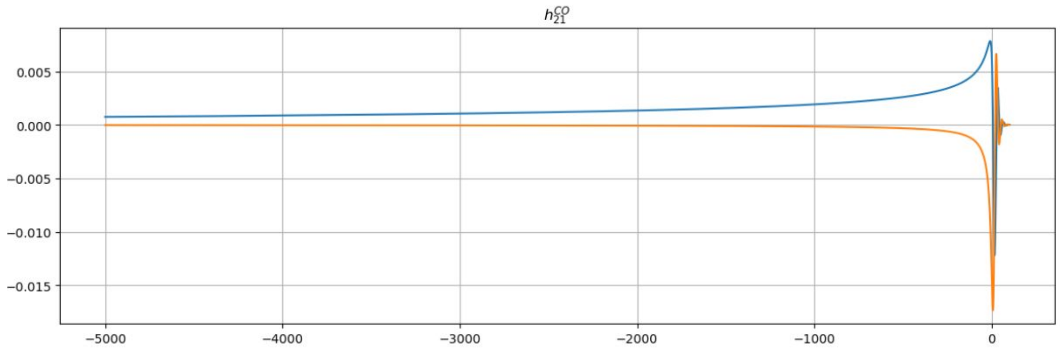
$$\mathbf{q} = \cos\left(\frac{\theta}{2}\right) + \sin\left(\frac{\theta}{2}\right)\hat{\mathbf{n}}$$

- **Orbital phase and co-orbital frame modes:**

$$\phi_{\rm orb}(t) := \frac{1}{2}\arg\left[h^P_{2,2}(t)\right],$$

$$h^R_{\ell m}(t) = h^P_{\ell m}(t)e^{-im\phi_{\rm orb}(t)}.$$





Step 2: data compression

[Field+ (1308.3565)
and many others….]

Construct a reduced basis for each waveform data piece

- Waveform pieces can be projected onto a finite dimensional basis $\{e_i(t)\}$:

$$f(t,\vec{\lambda}) \approx \mathcal{P}_n[f](t,\vec{\lambda}) := \sum_{i=1}^{n} c_i(\vec{\lambda}) e_i(t), \qquad (1)$$

Error in approximation (1) can be made arbitrarily small.

- Replace basis expansion (1) with an empirical interpolant:

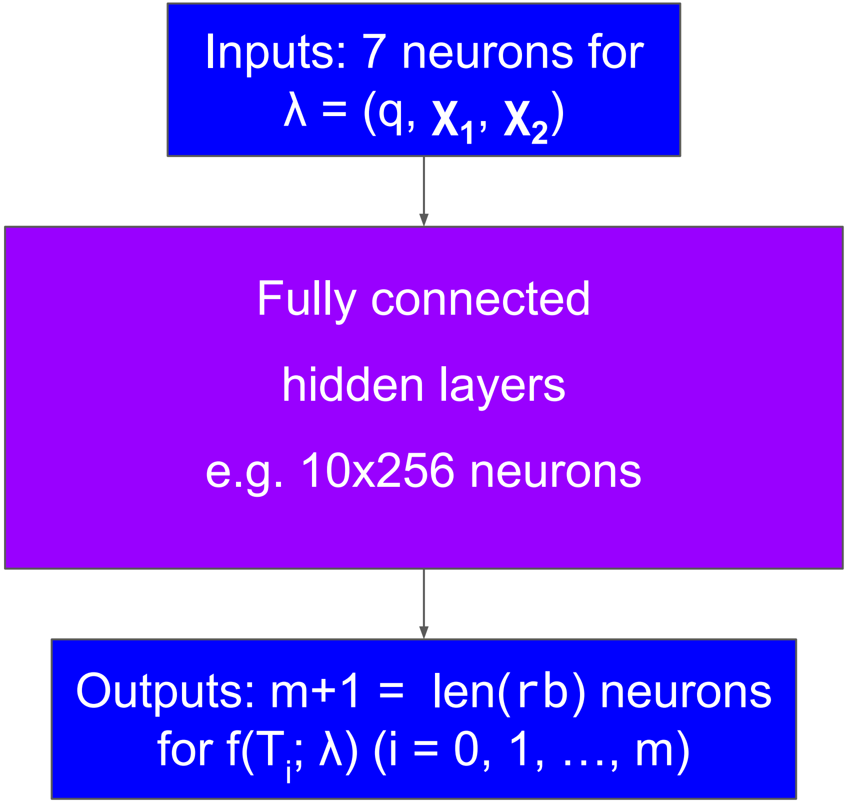
$$f(t,\vec{\lambda}) \approx \mathrm{EI}[f](t,\vec{\lambda}) := \sum_{j=0}^{n} B_j(t) f(T_j,\vec{\lambda}), \qquad (2)$$

Step 3: Prediction

Move from representative to predictive model

- Need to be able to **predict** $f(T_i; \lambda)$ for arbitrary parameters λ, for each T_i.
- Interpolate $f(T_i; \lambda)$ from the training set across parameter space using artificial neural networks
  - Interpolation across 7d parameter space $(q, \boldsymbol{\chi}_1, \boldsymbol{\chi}_2)$ challenging.
  - Make use of GPU accelerated frameworks.

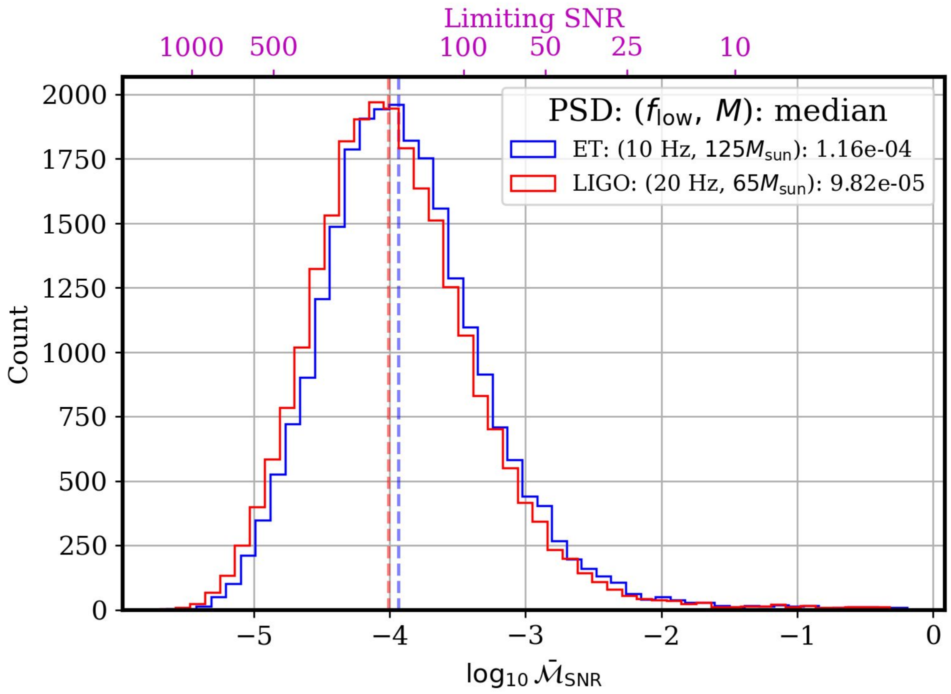
[Chua+ (1811.05491), Khan+ (2008.12932), …]



SEOBNRv5PHM_NNSur7dq10

- Reduced bases trained using 200k training waveforms.
- Neural networks trained on 1.2M waveforms (2.2M for the orbital phase)
- Covers mass ratios $q = m_1/m_2$ up to q=10, arbitrary spin magnitudes and directions.
- Returns time domain waveform, stretching from time 10,000M before merger through to 100M after.
- (2,2)-mode starting frequency < 20 Hz for all $M > 62.3M_{sun}$.

Results: faithfulness to SEOBNRv5PHM



- Mismatches with SNR-weighted average over inclination, phase and polarisation angle:

$$\bar{\mathcal{M}}_{\rm SNR} := 1 - \left(\frac{\sum_i \mathcal{M}_i^3 \rho_i^3}{\sum_i \rho_i^3} \right)^{1/3}$$

- Conservative estimate of “limiting SNR”,

$$\rho < \sqrt{\frac{N}{2\mathcal{M}_f}}.$$

above which we expect bias

Results: computational cost

Mean cost per waveform (wall time/batch size, averaged over parameter combinations)

	Wall time	**Speedup[3]**
Laptop CPU[1] **[single wf]**	12.52ms	5.2x
Laptop GPU[2] **[single wf]**	13.01ms	5.0x
Laptop GPU[2] **[250x batch]**	0.93ms	71x
Nvidia A100 **[1500 batch]**	0.16ms	410x
Nvidia A100 **[1500 batch, No copy to host][4]**	0.081ms	810x

[1] Single-thread on Intel Core Ultra 7 155H

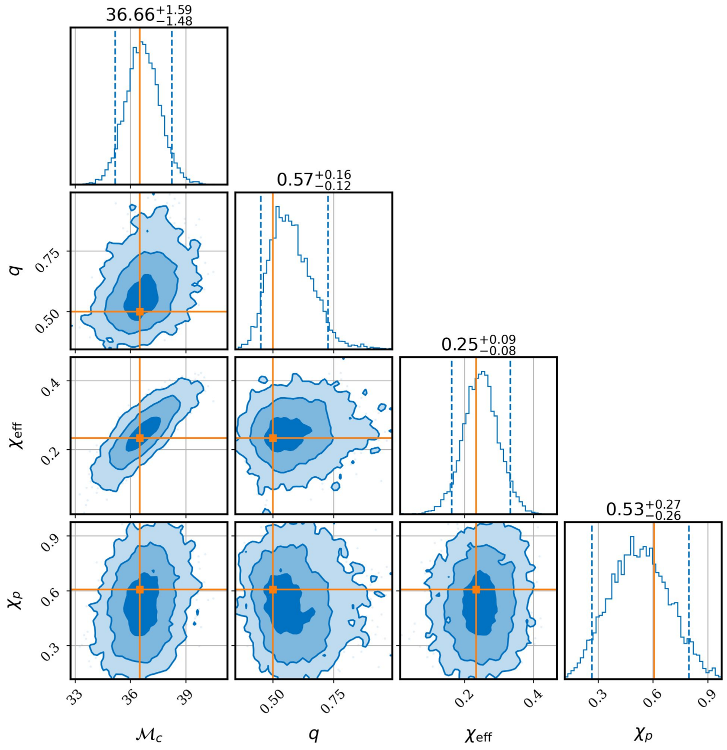
[2] Nvidia RTX 1000 Ada

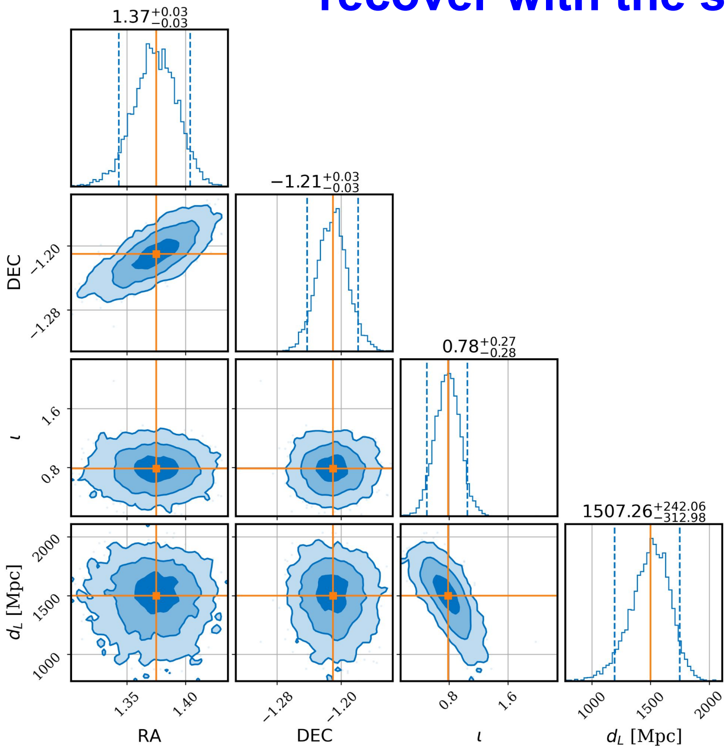
[3] Average SEOB cost = 65.6ms per waveform

[4] Keeping the final result in GPU memory rather than copying back to the CPU-accessible memory.

Results: injection-recovery

Inject SEOBNRv5PHM into zero noise, recover with the surrogate





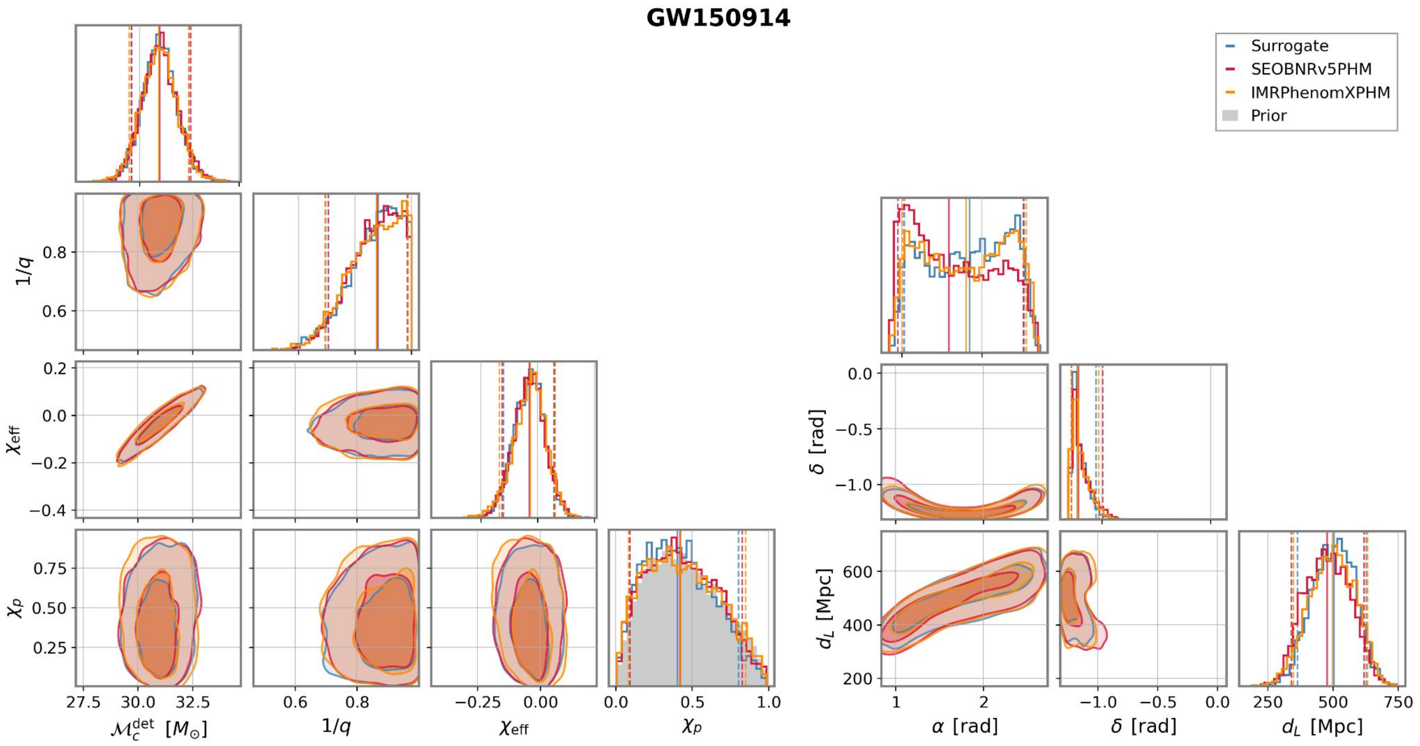
	Injected
m_1	60.0
m_2	30.0
a_1	0.7
θ_1	π/3
a_2	0.0

Network (HLV) optimal SNR ~ 25.7

$$\chi_{\rm eff} := \frac{q|\vec{\chi}_1|\cos\theta_1 + |\vec{\chi}_2|\cos\theta_2}{1+q},$$

$$\chi_p := \max\left\{|\vec{\chi}_1|\sin\theta_1, \frac{(4+3q)}{(4q+3)q}|\vec{\chi}_2|\sin\theta_2\right\}.$$

Results: GW150914



	Sur	SEOB	XPHM
Wall time	19hr 54m	63hr 49m	9hr 10m

~ 3.2x speedup compared to SEOBNRv5PHM

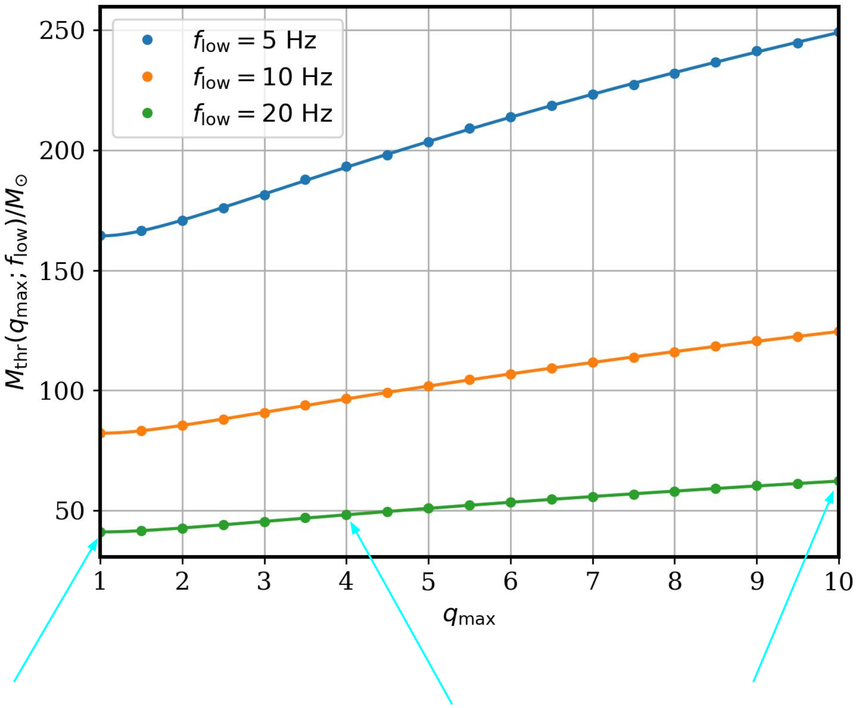
Conclusions

- Demonstrated neural network surrogate approach for 7d precessing problem.
- Median model mismatches ~ 1e-4 against base SEOBNRv5PHM.
- Speedup ~5x for single waveform on a laptop CPU or GPU
- Surrogate achieves best (average) performance when evaluating large waveform batches on the GPU.
    - Waveforms in as little as 81us amortized cost.
- Demonstrated practical application of surrogate to PE, with significant speedup over SEOBNRv5PHM.
    - Full potential requires samplers that can exploit batched waveform evaluation.

Coming to the arXiv next week

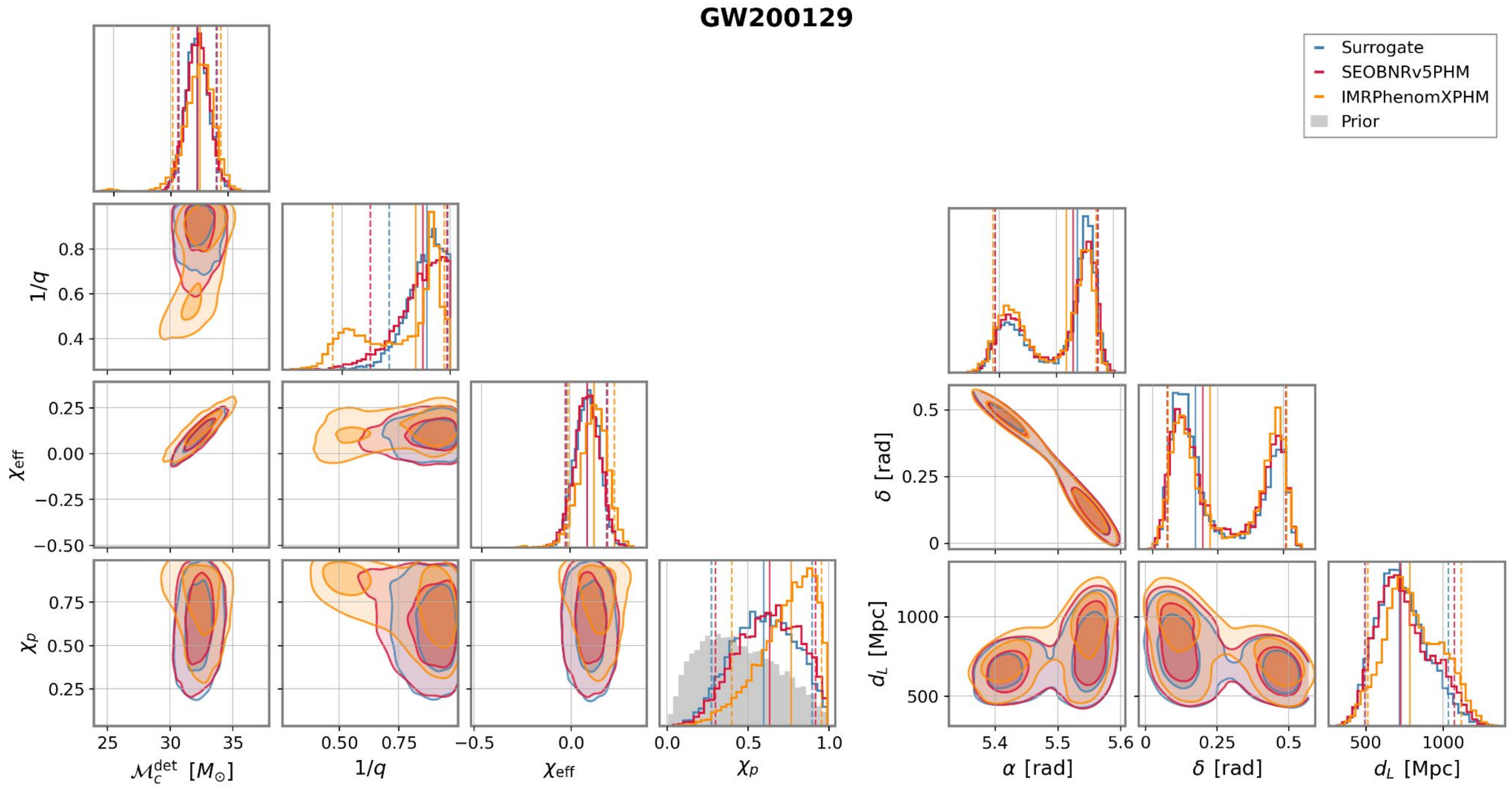
Bonus slide: threshold mass

- Waveform polarizations returned on a fixed geometric time grid [-10^4M, 100M] with spacing Δt = 0.5M.
- Starting frequency depends on the total mass.
- Threshold mass (right): minimum mass such that (2,2) mode starts below 5, 10, 20 Hz whenever M > M_{thr}.
- M_{thr} = 1.2 x 10^7 M_{sun} for f_{low} = 0.1 mHz, q_{max} = 10



M > 41.1M_{sun} **M > 48.3M_{sun}** **M > 62.3M_{sun}**

Bonus slide: GW200129



Bonus slide: GW250114