Extending Neural Likelihood Estimators for Gravitation-Wave Analysis through Transfer Learning

Martijn Teunissen, Luca Negri, Anuradha Samajdar

NNV fall meeting 7 November 2025

Overview

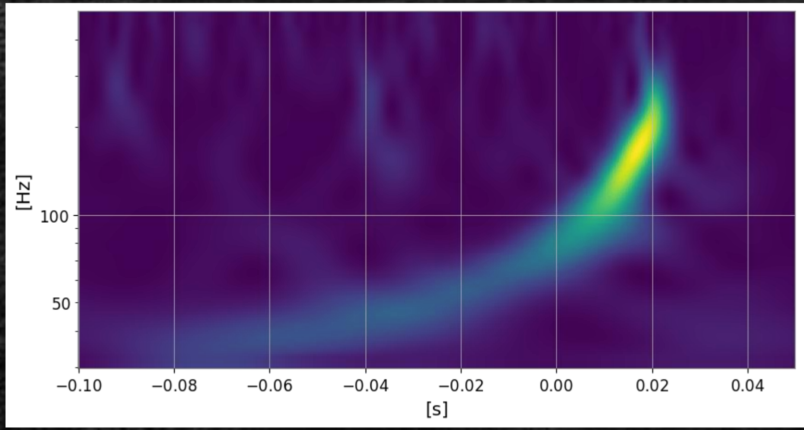
- Parameter Estimation
- FLEXible Neural Likelihood Estimator (FLEX)
- Transfer Learning with FLEX

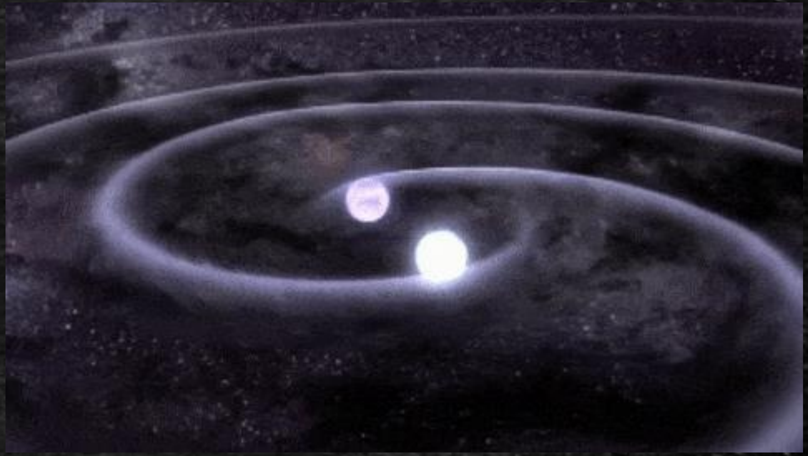
Paper: FLEX

https://arxiv.org/pdf/2509.17606

Parameter Estimation

- How do we link the data to the physics underneath?





Parameter Estimation with Bayes Theorem

- Posterior is the goal
- Evidence can be used for model comparison

- Between 9 and 17 parameters $\boldsymbol{\theta}$

- Posterior is found trough Stochastic sampling
    - $O(10^6)$, $O(10^7)$

Likelihood

Prior

$$p(\theta \mid D) = \frac{p(D \mid \theta)\, p(\theta)}{p(D)}$$

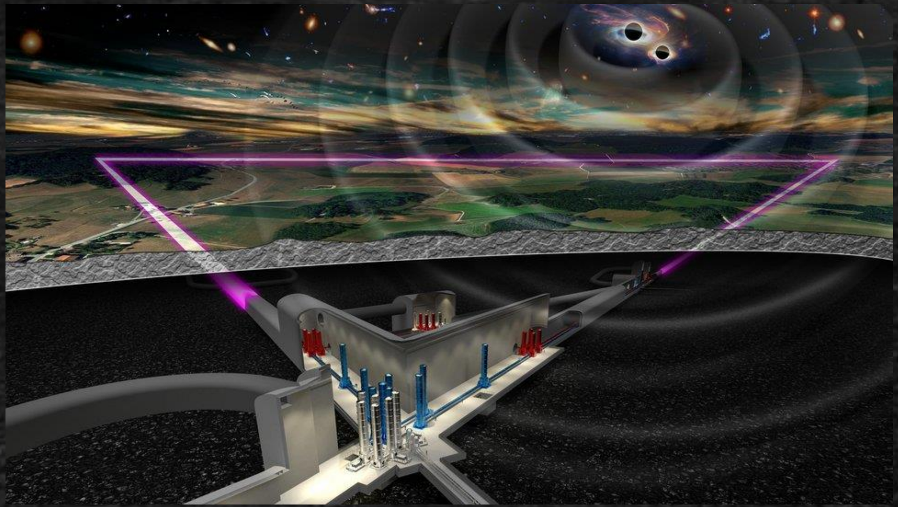
Posterior

Evidence

$$\boldsymbol{\theta} = (m_1, m_2, \mathbf{S}_1, \mathbf{S}_2, \alpha, \delta, \psi, \iota, r, t_c, \Phi_c)$$

Further challenges

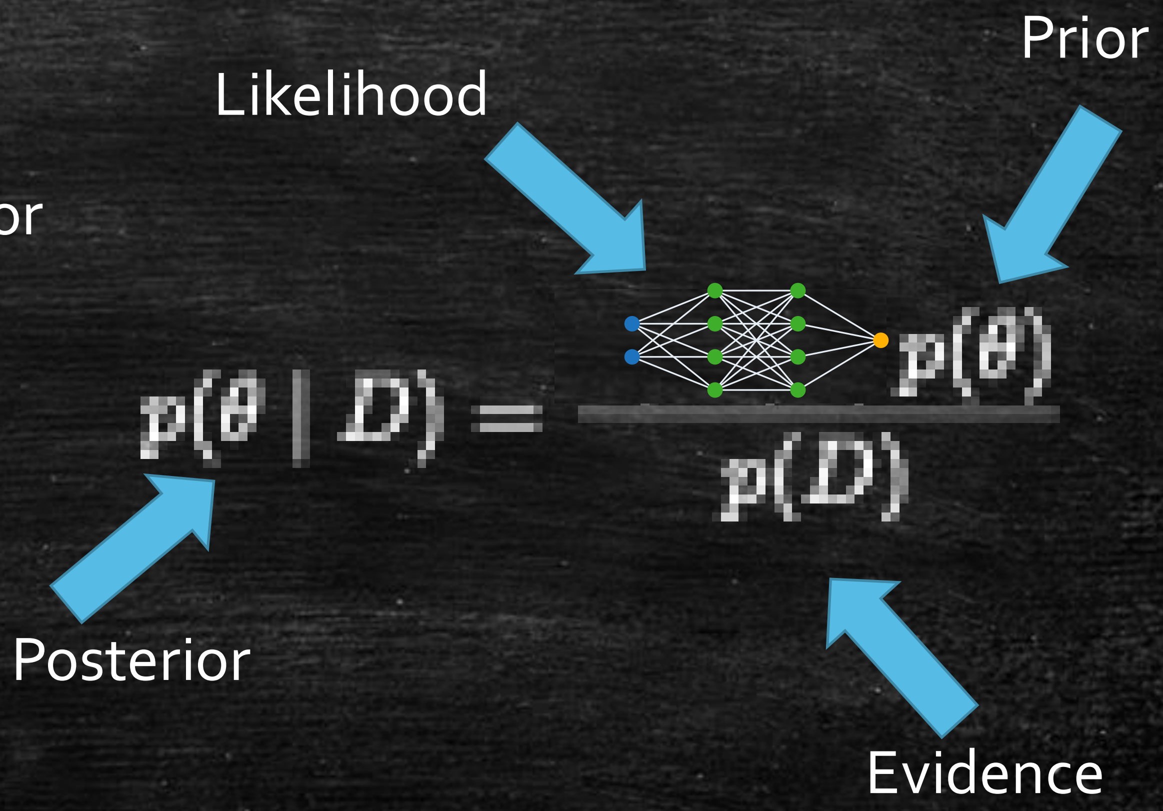
- More signals (x1000)
- Each signal will take longer to analyse
  - Louder
  - Longer
  - Overlapping
- Preparation means faster and accurate inference



A first step: FLEX

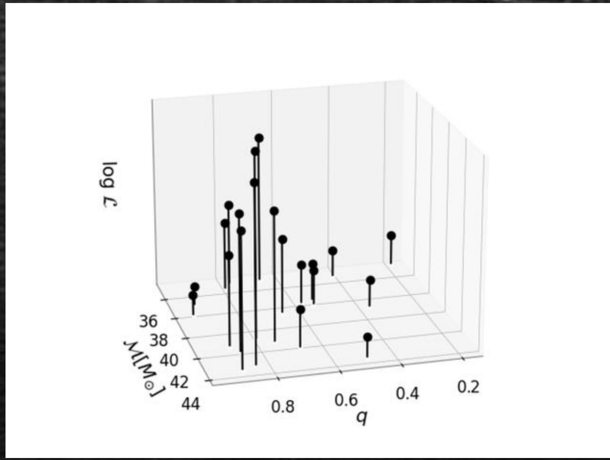
A FLEXible Neural Likelihood Estimator

- Train on the fly
- 1 ms vs 1μs

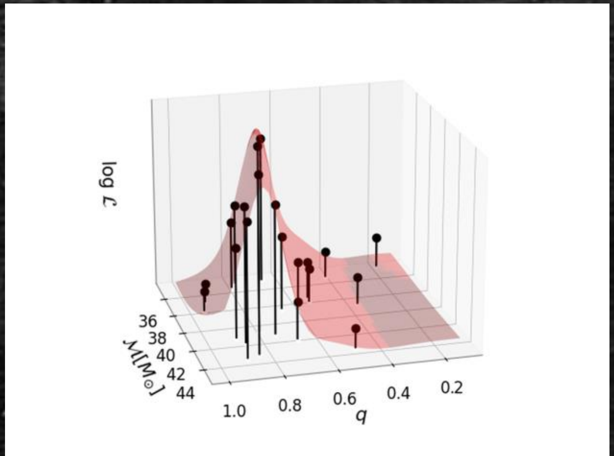


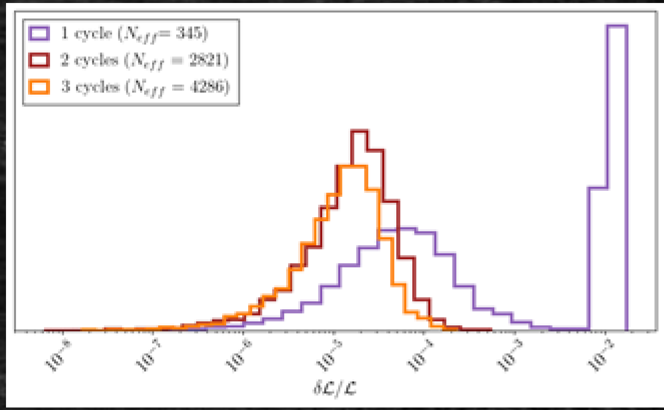
The FLEX Cycle

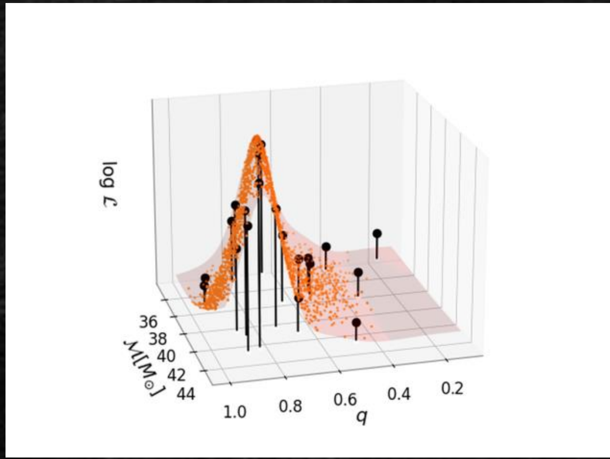
2: Train Neural Likelihood



1: Generate training samples



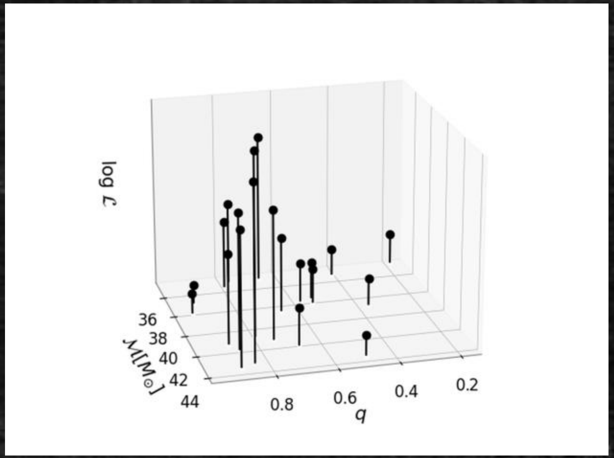




3: Run sampler
on the Neural likelihood

- 4: Check results and retrain

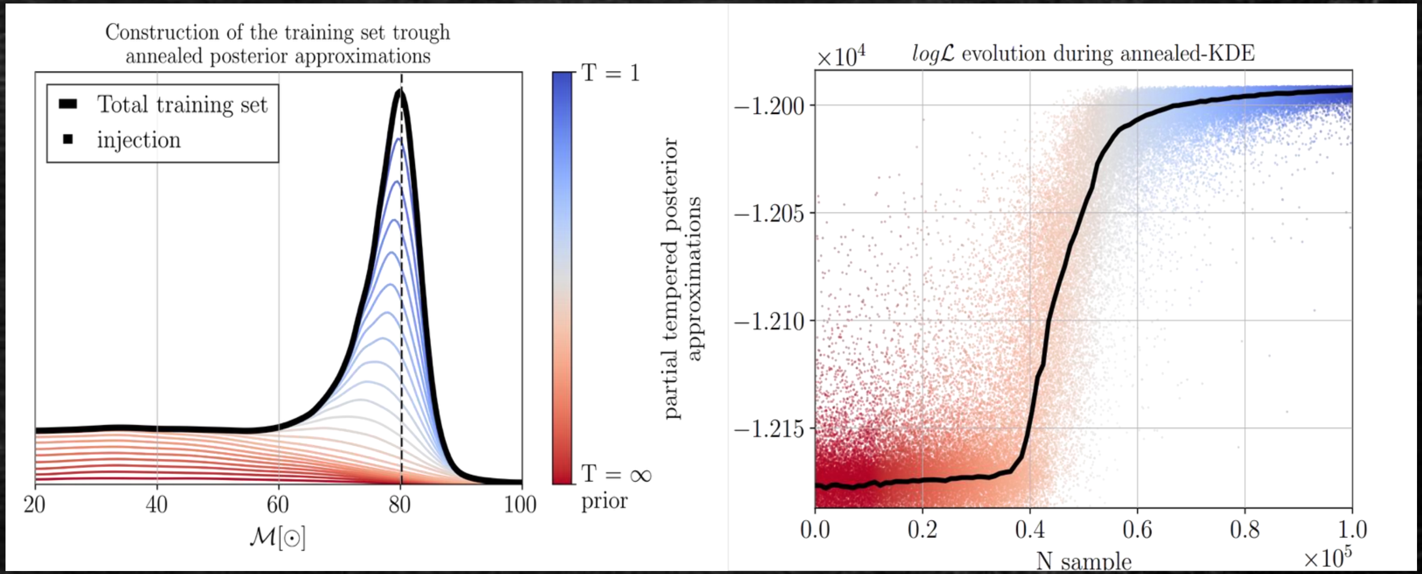
Phase 1: Generate training samples



Phase 1: Generate training samples

- Annealed series tempered Kernel Density Estimators (KDE)
- Interpolates between prior and posterior

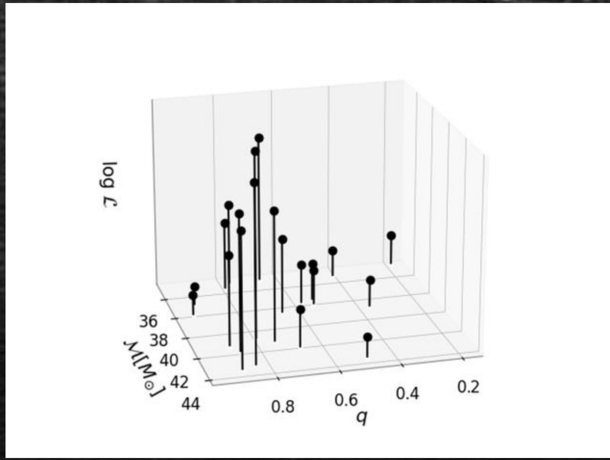
- $N_{train} << N_{MCMC} (10^{6-7})$
- $N_{initial} = 1e5$



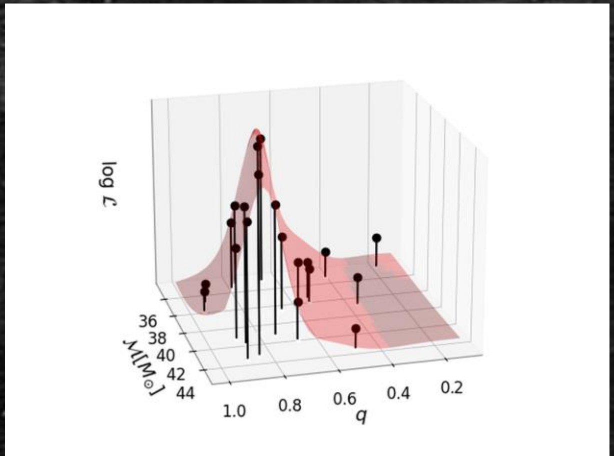
$$\theta_i \sim p(\theta|y, \mathcal{M})^{1/T}$$

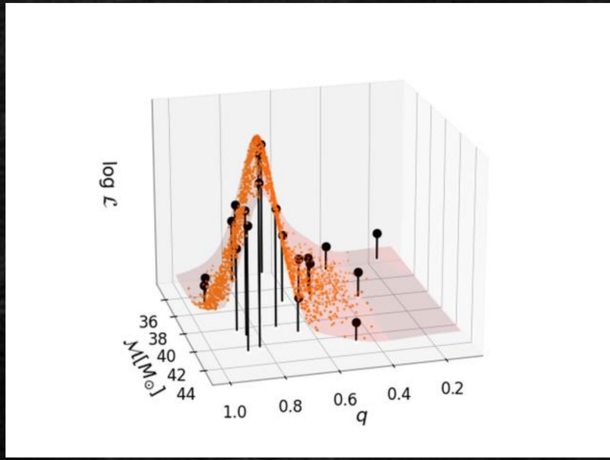
The FLEX Cycle

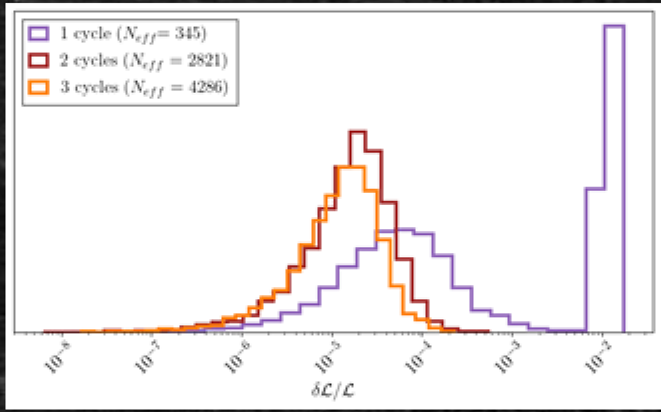
2: Train Neural Likelihood



1: Generate training samples





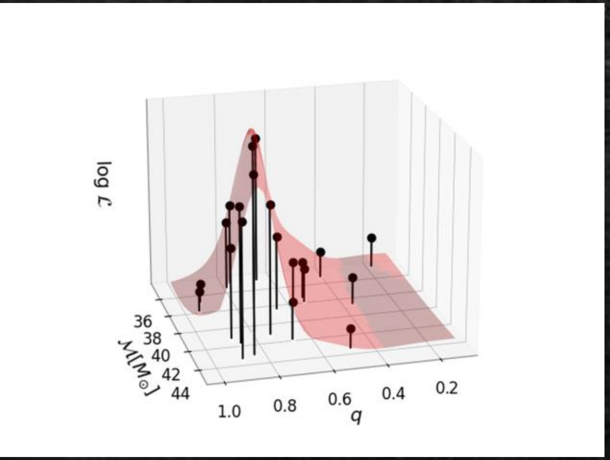


3: Run sampler
on the Neural likelihood

- 4: Check results and retrain

Phase 2: Train Neural Likelihood

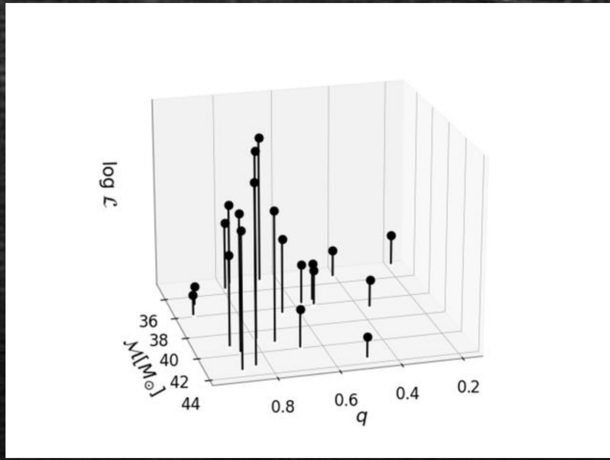
- Small ResNet, fully connected
- ~15k trainable parameters
- Fast to train, fast to analyze
- $N_{\mathrm{epochs}} = 700$



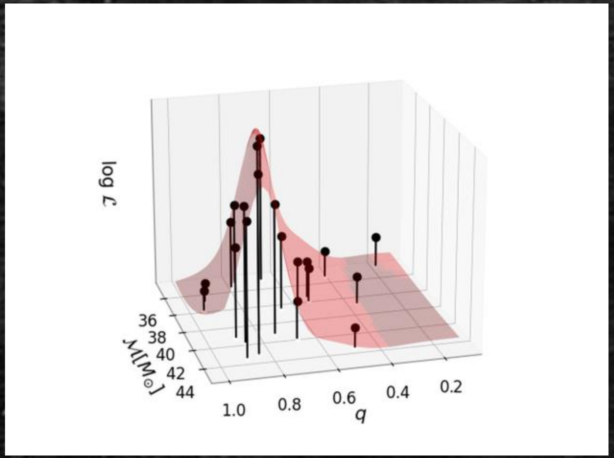
$$\mathrm{Loss} = \frac{1}{N_{\mathrm{t}}} \sum_{i=0}^{N_{\mathrm{t}}} \left(e^{\log \mathcal{L}(\theta_i)} - e^{\mathrm{NN}(\theta_i)}\right)^2 w_i$$

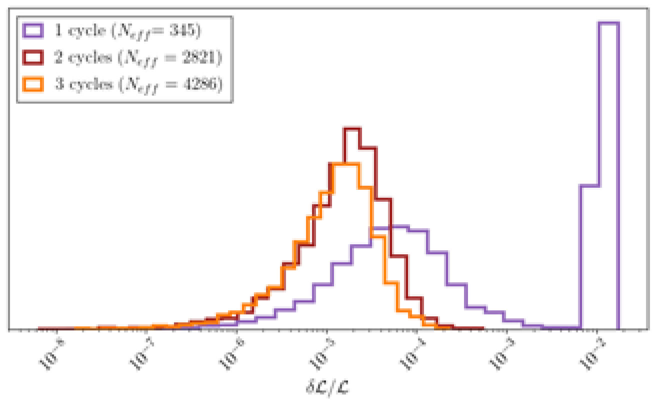
The FLEX Cycle

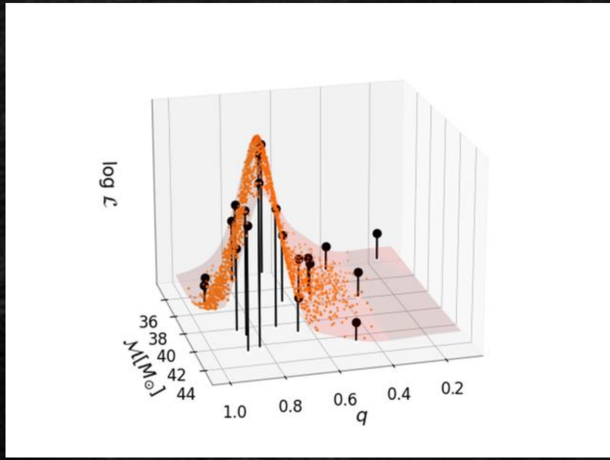
2: Train Neural Likelihood



1: Generate training samples





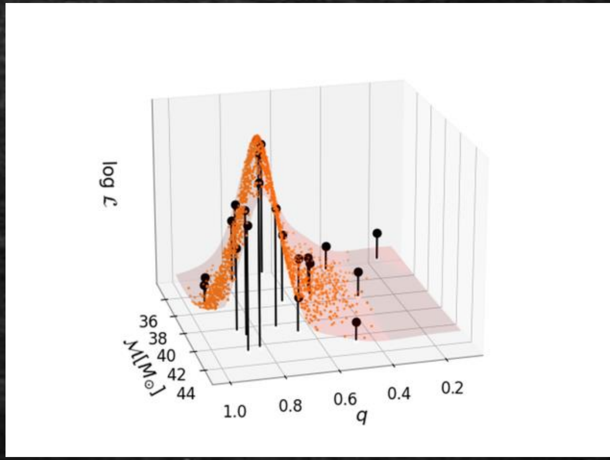


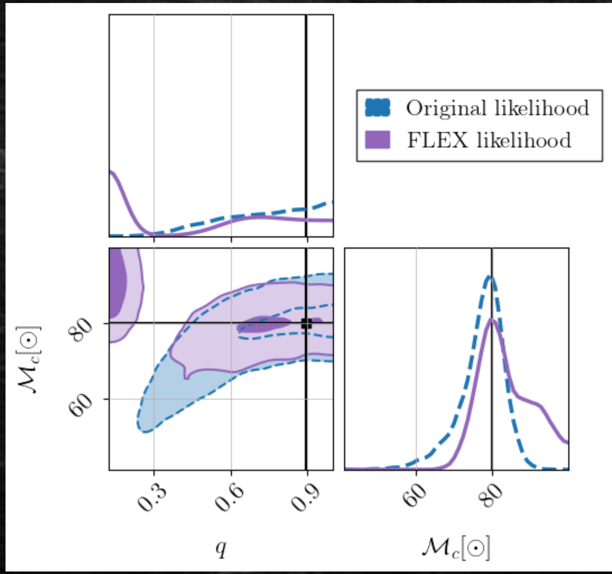
3: Run sampler
on the Neural likelihood

- 4: Check results and retrain

Phase 3: Run sampler on the Neural likelihood

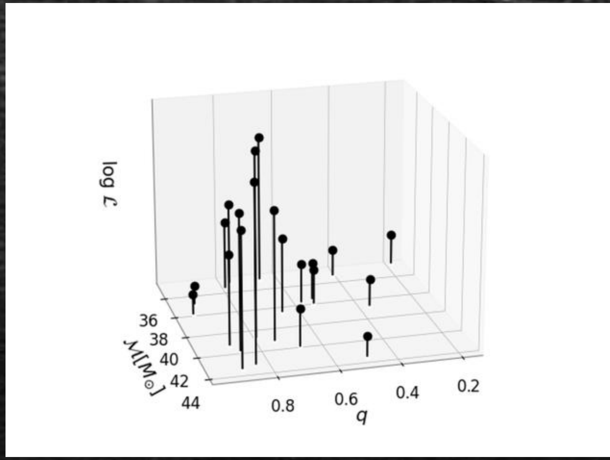
- MCMC with parallel tempering
- Modest hardware (CPU)
- ~ 1 minute



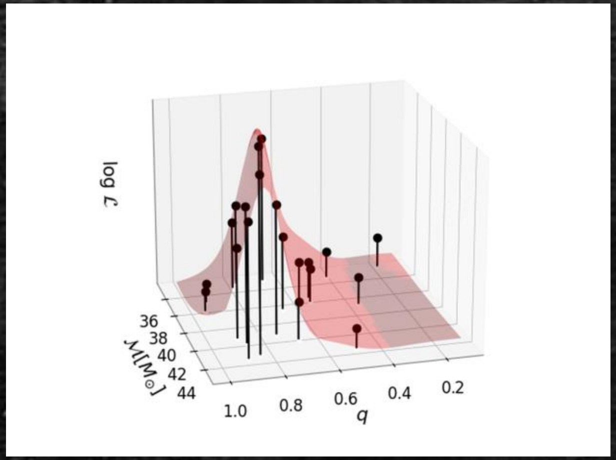


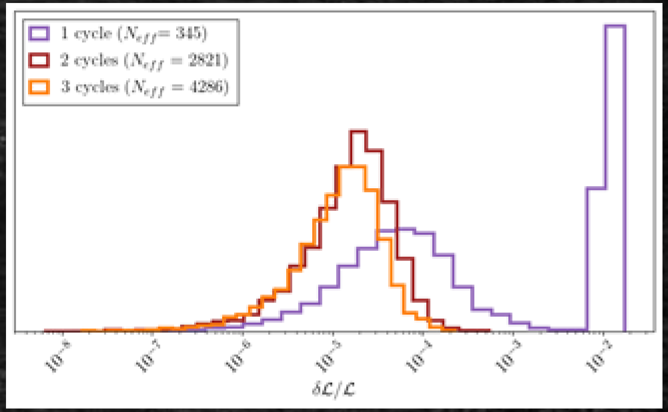
The FLEX Cycle

2: Train Neural Likelihood

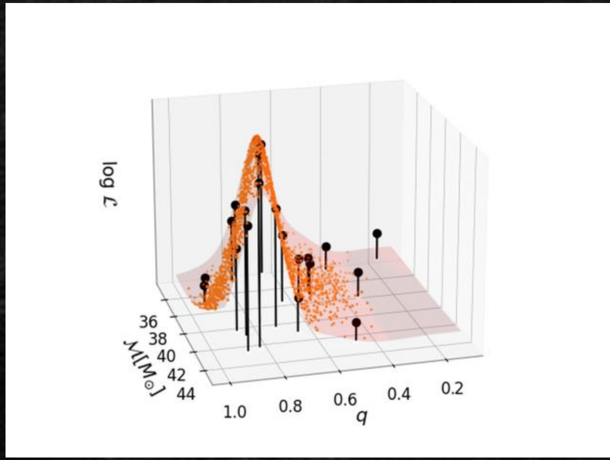


1: Generate training samples





3: Run sampler
on the Neural likelihood

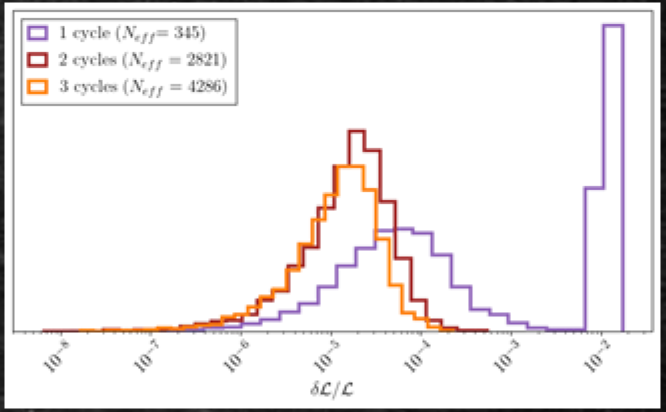


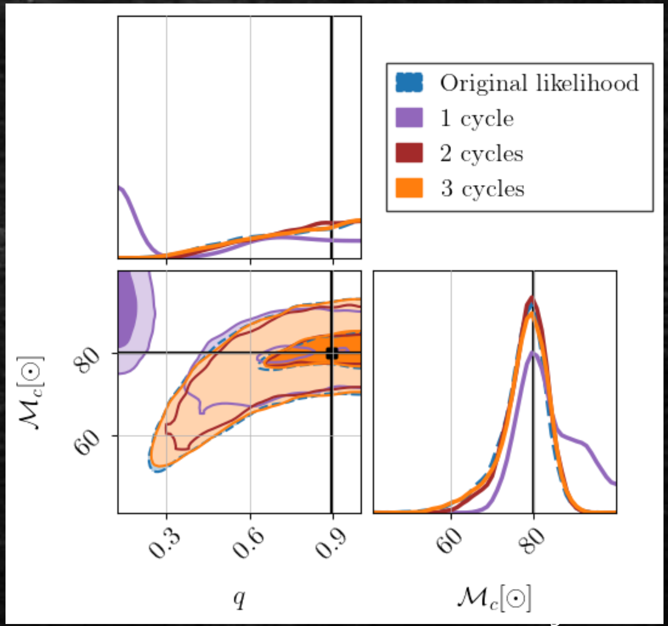
4: Check results and retrain

Phase 4: Check results and retrain

- Calculate the Effective Sample Size (ESS)
- If ESS/N_{post} < threshold
  - Append $N_{retrain\_samples}$ to samples
  - Restart cycle
  - Maximum of 6 cylces
  - Threshold = 50%
  - N_{post} = 5000
  - $N_{retrain\_samples}$ = 2000

$$w_i = \frac{\mathcal{L}(\theta_i)}{NN(\theta_i)} \qquad ESS = \frac{(\Sigma_i w_i)^2}{(\Sigma_i w_i^2)}$$

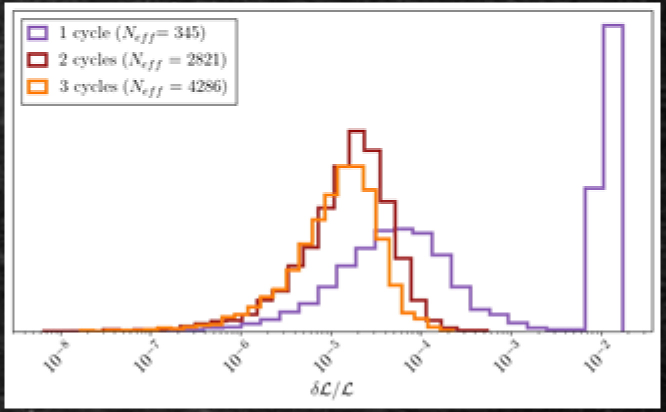


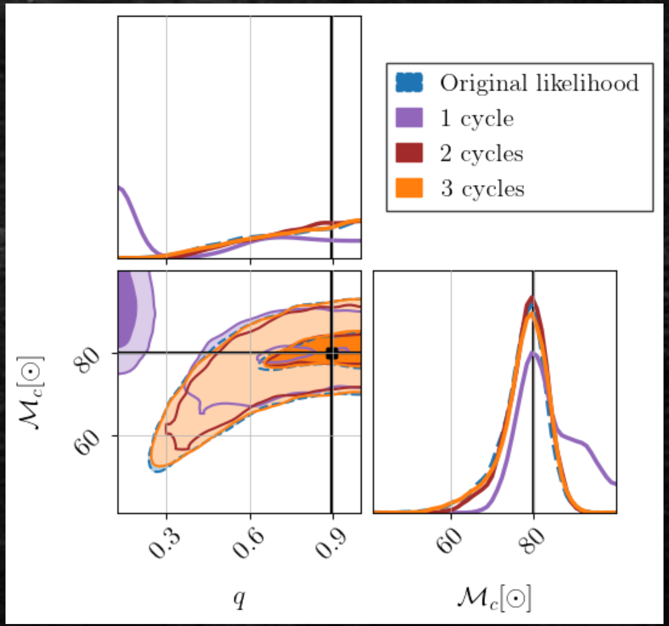


Phase 4: Check results and retrain

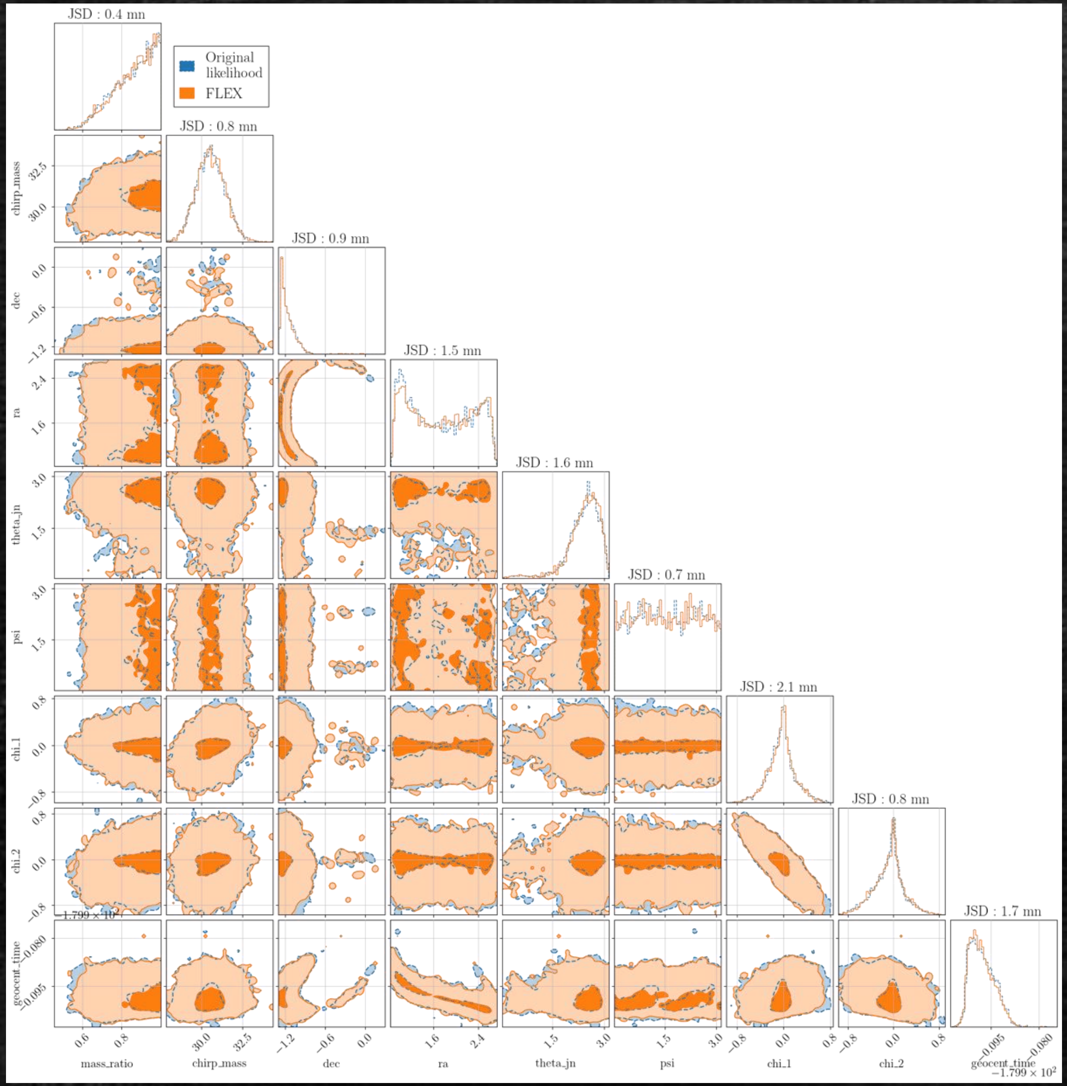
- Calculate the Effective Sample Size (ESS)
- If ESS/N_{post} > threshold
  - Convergence reached!
  - Terminate cylce

$$w_i = \frac{\mathcal{L}(\theta_i)}{\mathrm{NN}(\theta_i)} \qquad ESS = \frac{(\Sigma_i w_i)^2}{(\Sigma_i w_i^2)}$$





Standard test with real data: GW150914



The first GW signal detected

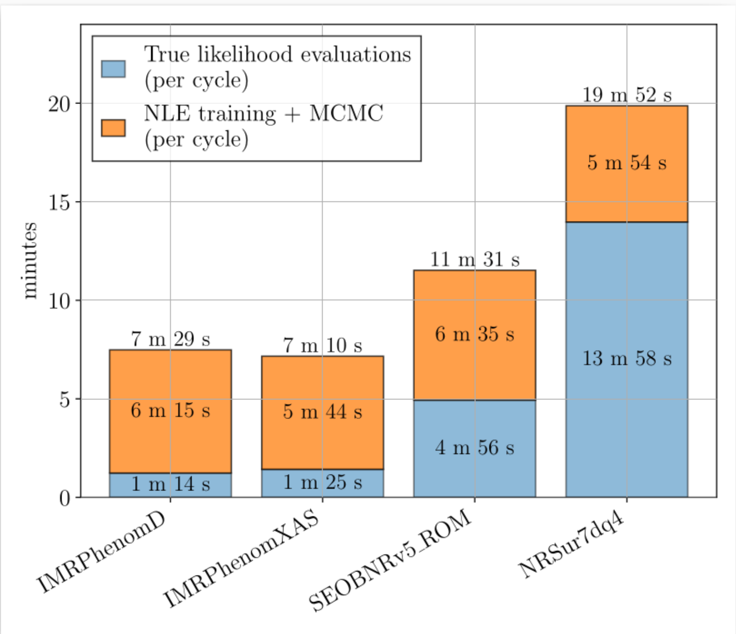
Model	Time	Number of likelihood evals
Eryn (blue)	~12 hours	~ 1.3e7
FLEX (orange)	~ 36 minutes	~ 1.8e5

60 times reduction

With much room for optimization

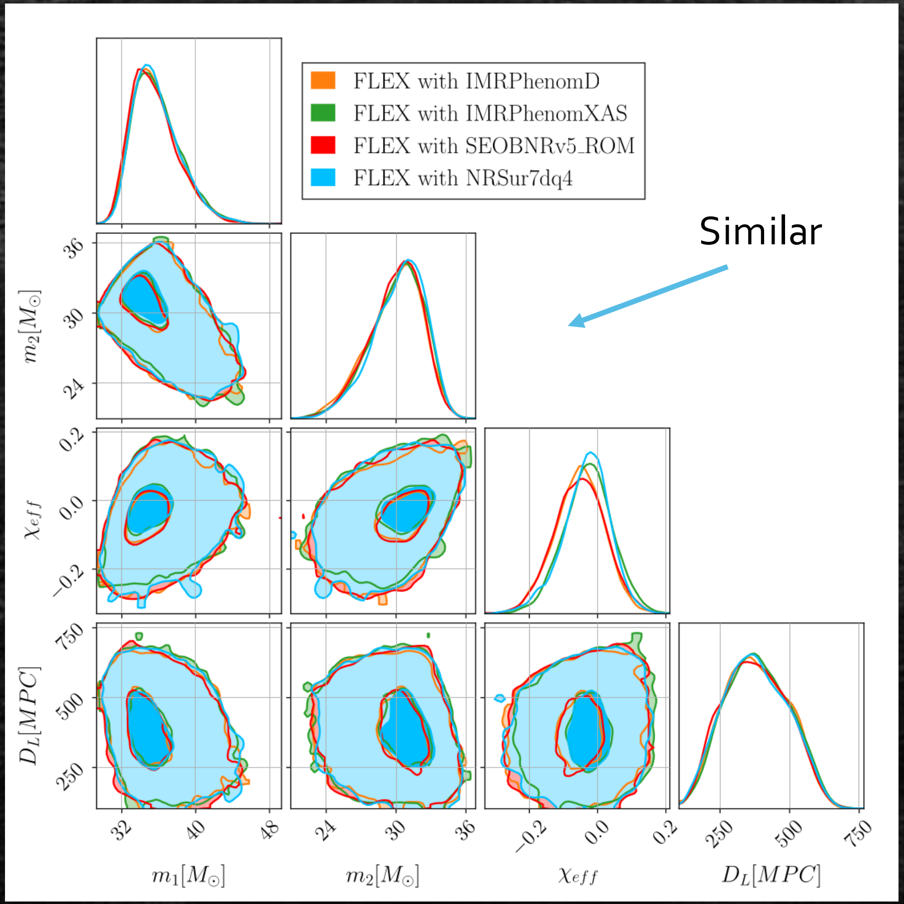
Different waveform models

- Different models used to describe the data
  - IMRPhenomD
  - IMRPhenomXAS
  - SEOBNRv5_ROM
  - NRSur7dq4



- Phenom waveforms are cheaper because they:
  - Are a hybrid of analytical approximations and fits to numerical relativity simulations
  - Don't need to solve expensive differential equations

Results: Robust on different waveform models



Model (FLEX)	Log evidence
IMRPhenomXAS	285.10
SEOBNRv5	285.05
NRSur7dq4	284.67
IMRPhenomD	285.2

Can we use this similarity?

- The idea: Use transfer learning to reduce the number of likelihood evaluations needed to train FLEX
  - Train FLEX on a computationally cheap waveform
    - IMRPhenomD
    - IMRPhenomXAS
  - Use what these NLE has learned to inform the training of FLEX on an expensive waveform
    - SEOBNRv5_ROM
    - NRSur7dq4

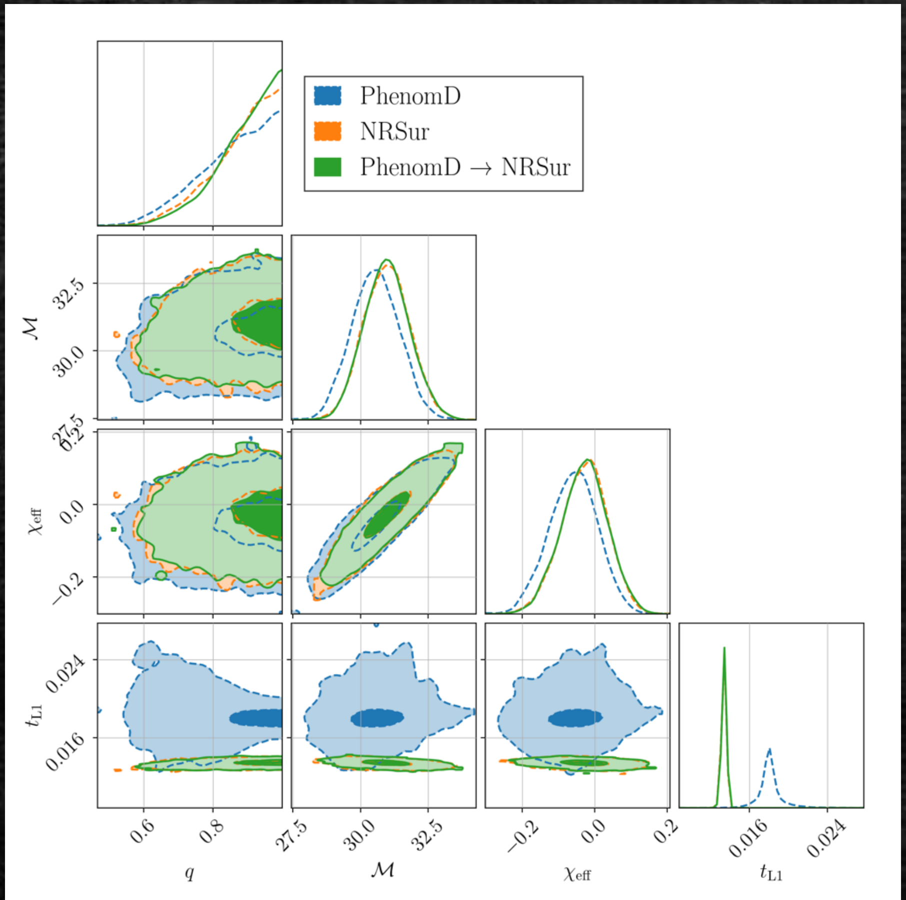
Method

- Transfer the NLE model weights
- Transfer a subset of the training samples an recalculate the loglikelihoods
  - Random selection
  - Vary the number of transferred samples

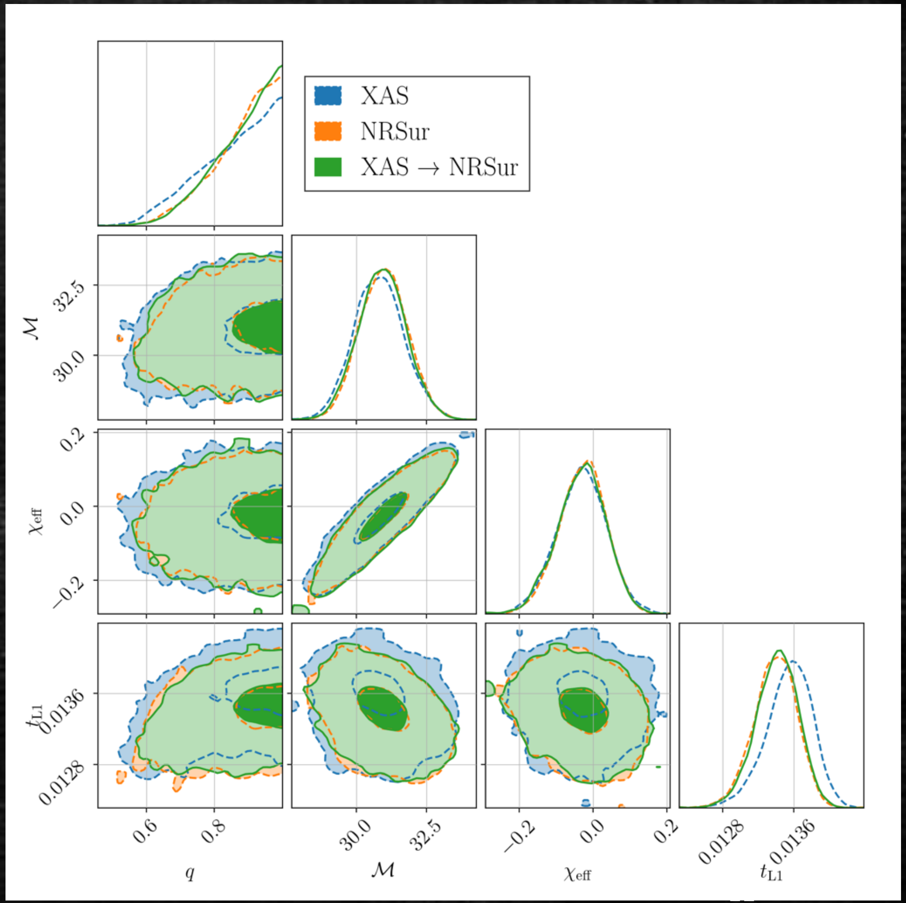
- Average number of True likelihood evaluations
- Jensen–Shannond divergence
  - Measures similarity between distributions
  - Below 3 mnats is essentially the same
- Do the FLEX runs converge

Results:NRSur

IMRPhenomD

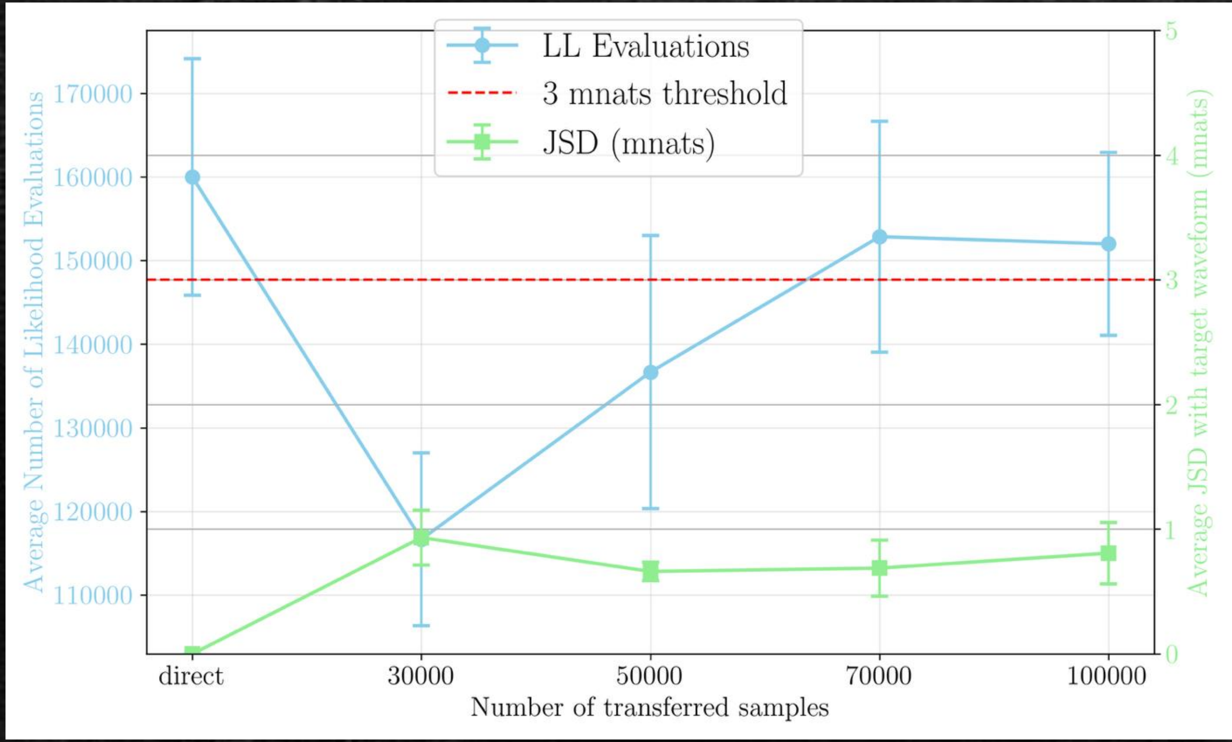


IMRPhenomXAS

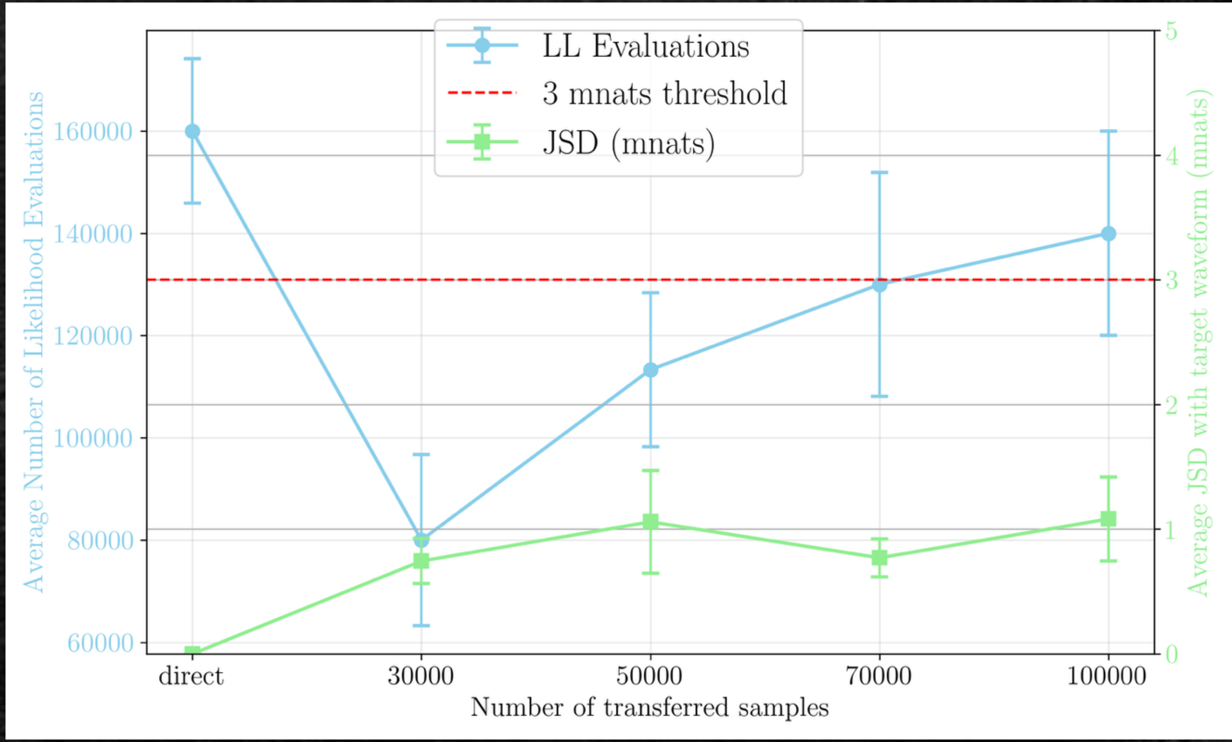


Results:NRSur

IMRPhenomD



IMRPhenomXAS



Conclusions and outlook

- Transfer Learning can reduce the number of likelihood evaluation without losing accuracy
- These results were a proof a concept
  - Get smarter about data transfer
    - Source likelihood weighted sample selection
    - Combine with FLEX KDE algorithm
  - Including a weighted training data set from the source model
  - Heterogeneous transfer learning
    - Train FLEX on a subset of parameters
    - Transfer to a run including more parameters/physics

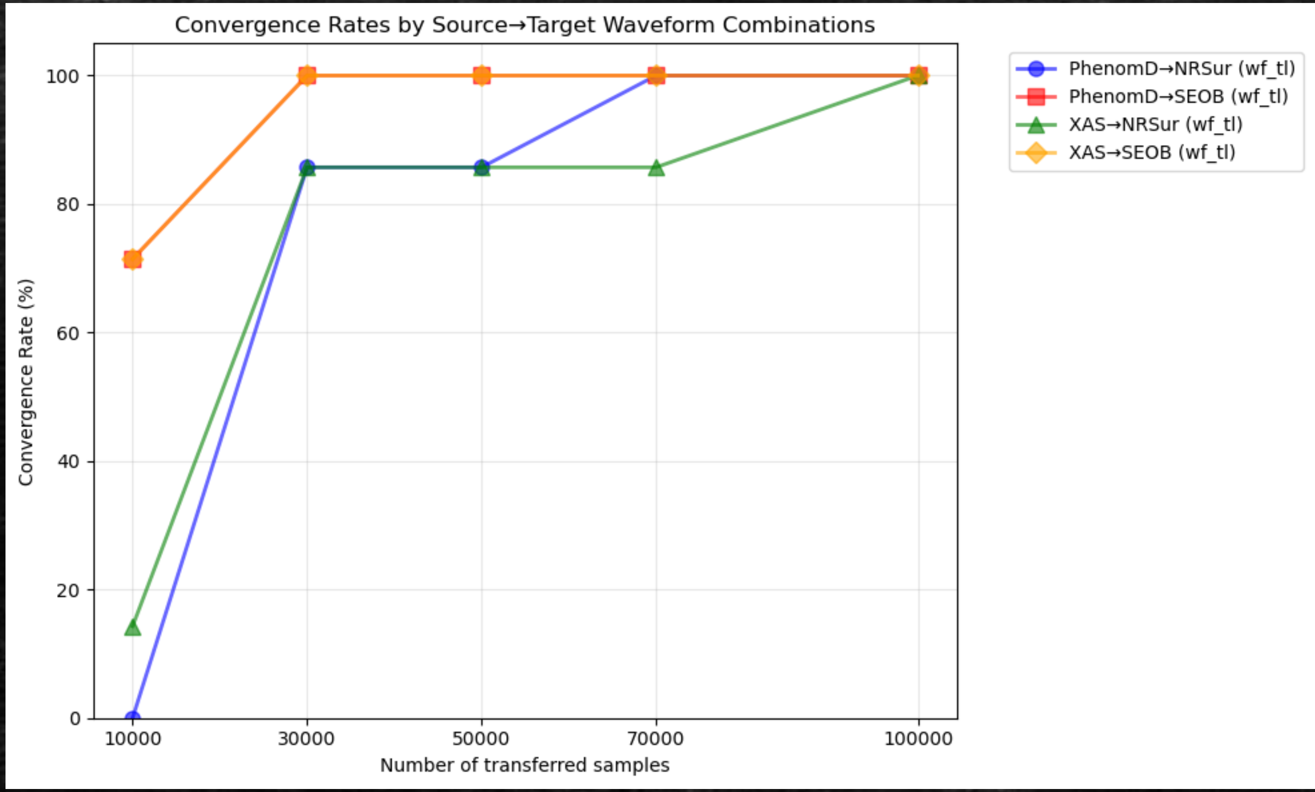
Paper: FLEX

https://arxiv.org/pdf/2509.17606

Questions?

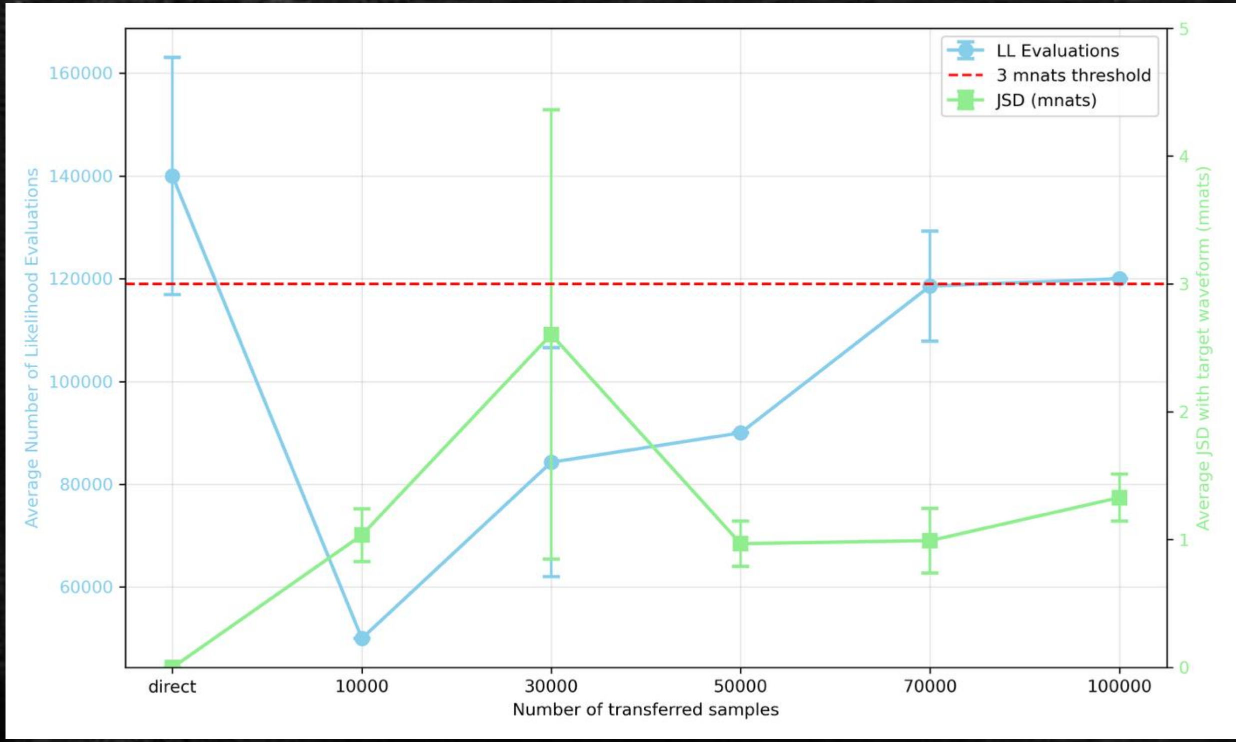
Results: Convergence

- Convergence direct runs 100%

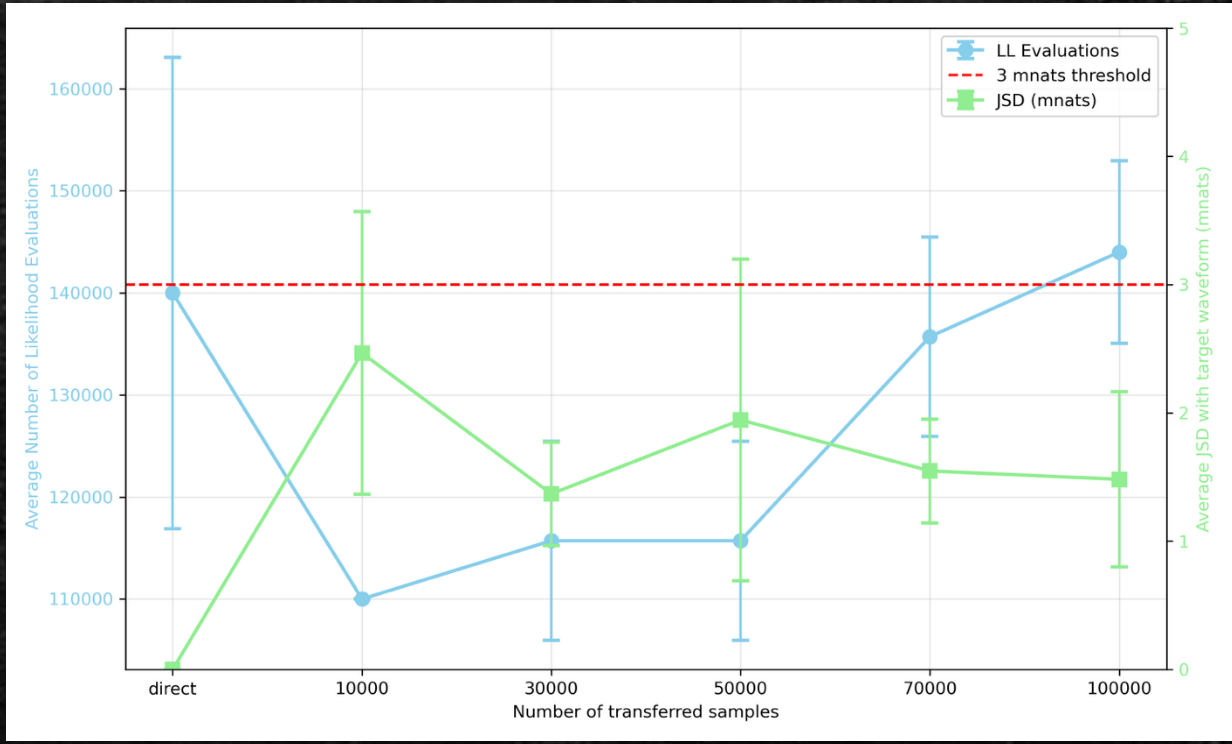


Results: Target SEOBNRv5_ROM

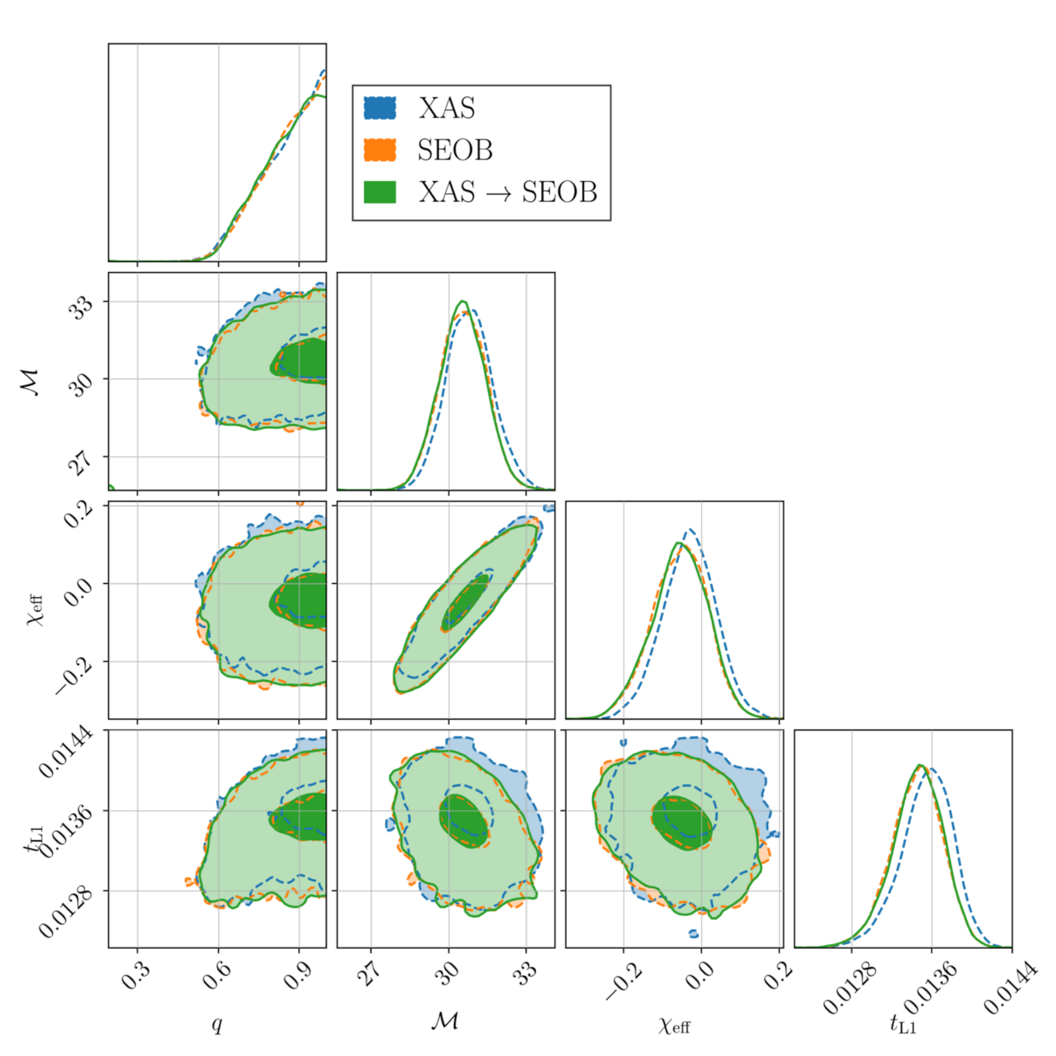
Source: IMRPhenomXAS

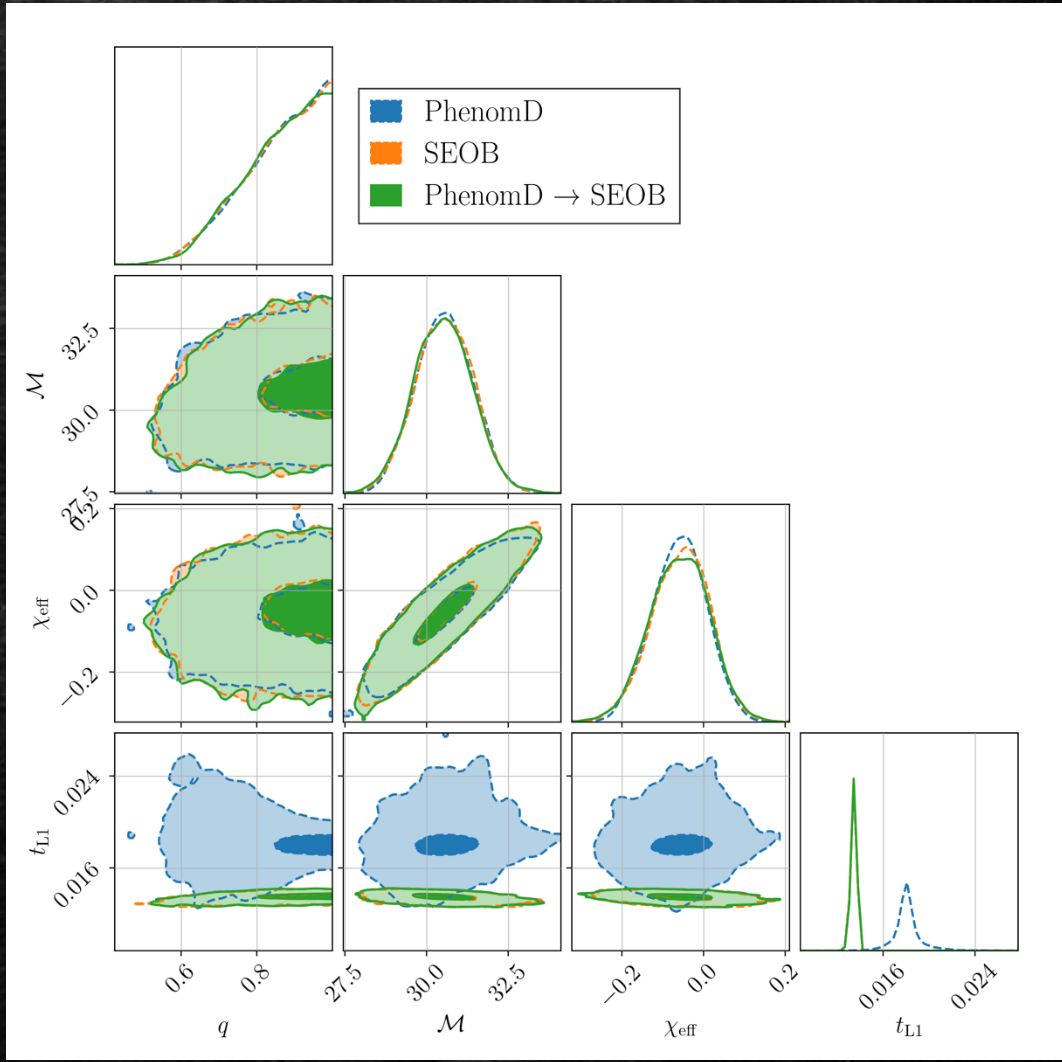


Source: IMRPhenomD



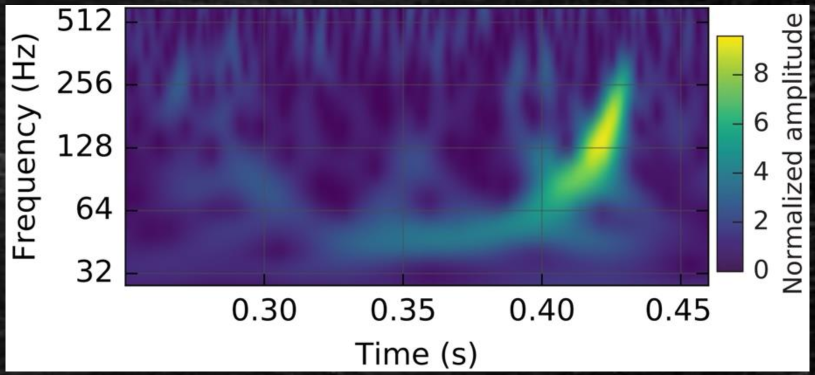
Results: SEOBNRv5_ROM



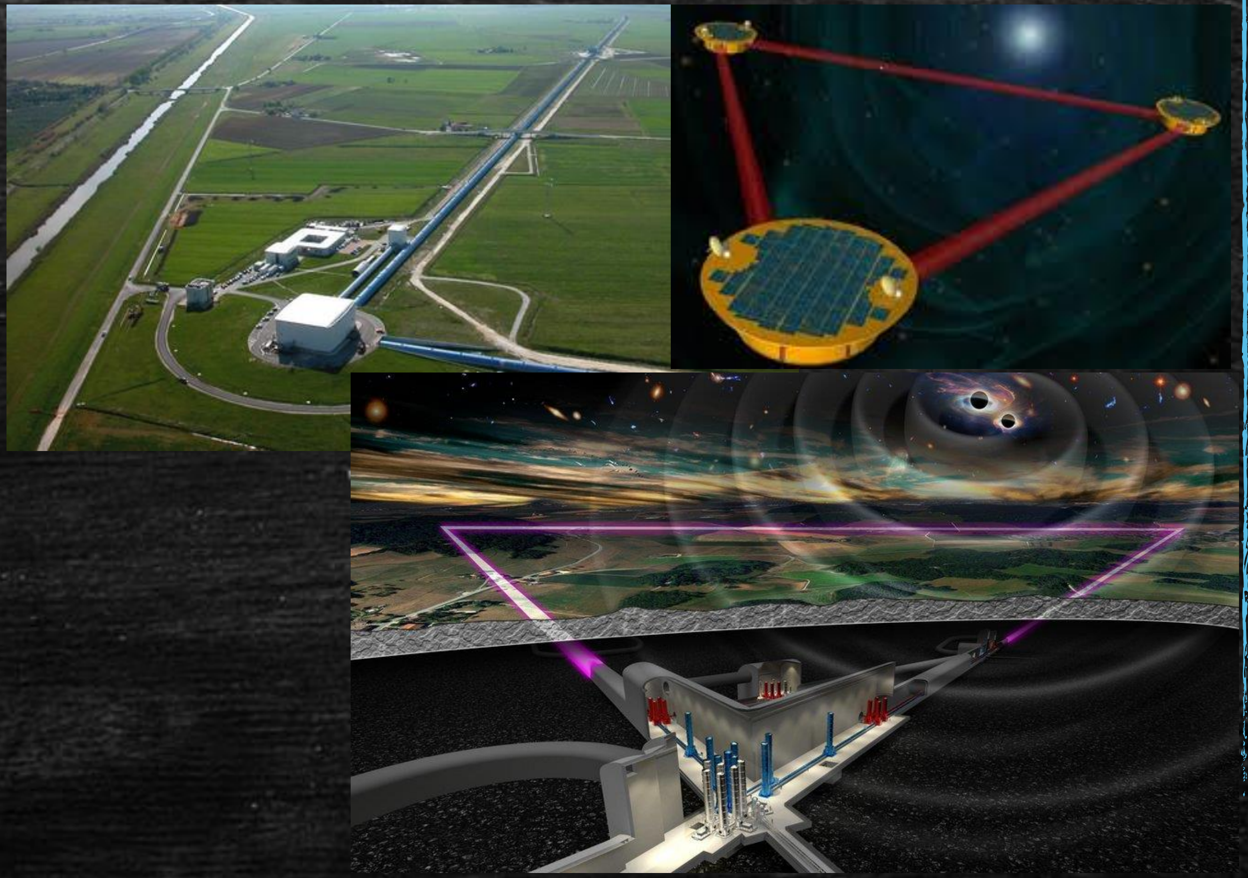


What signal type will we tackle?

- We start with the easiest problem in GWs
  - Heavy mass, short duration , BBHs [20-100 chirp mass]
  - SNR<40
  - 2-3 detector networks, current sensitivity
  - Aligned spin (no precession) IMRPhenomD
  - Distance & phase analytical marginalization
  - Sampling over 9 parameters
    - $[\mathcal{M}, q, \chi_1, \chi_2, \theta_{jn}, \psi, \mathrm{zenith}, \mathrm{azimuth}, t_{det}]$



Interferometers to detect GWs



Detected using interferometers