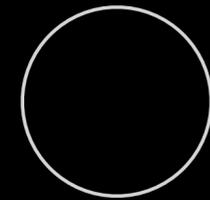
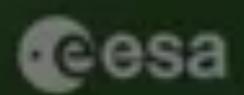


How can we observe black holes?

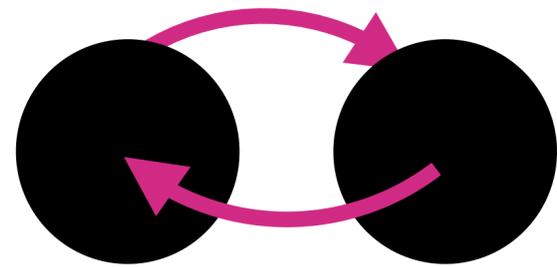




Simulation of gravitational waves

What are gravitational waves?

Black holes merge → Emit gravitational wave → Measured in detectors



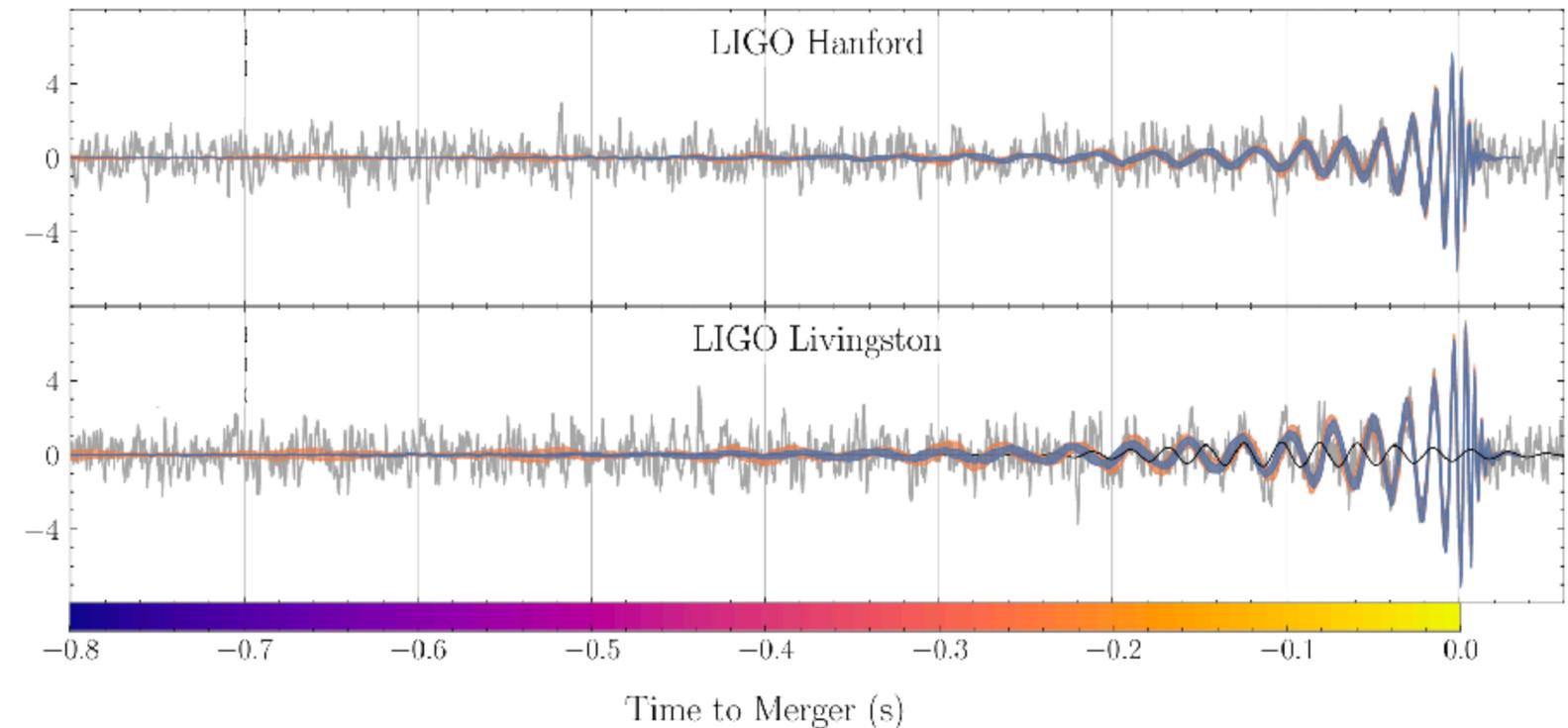
Described by physics parameters

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

Masses, spins, sky position, ...

Universe

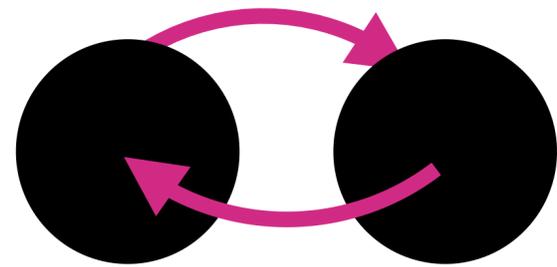
Posterior Estimation
 $p(\theta | d)$



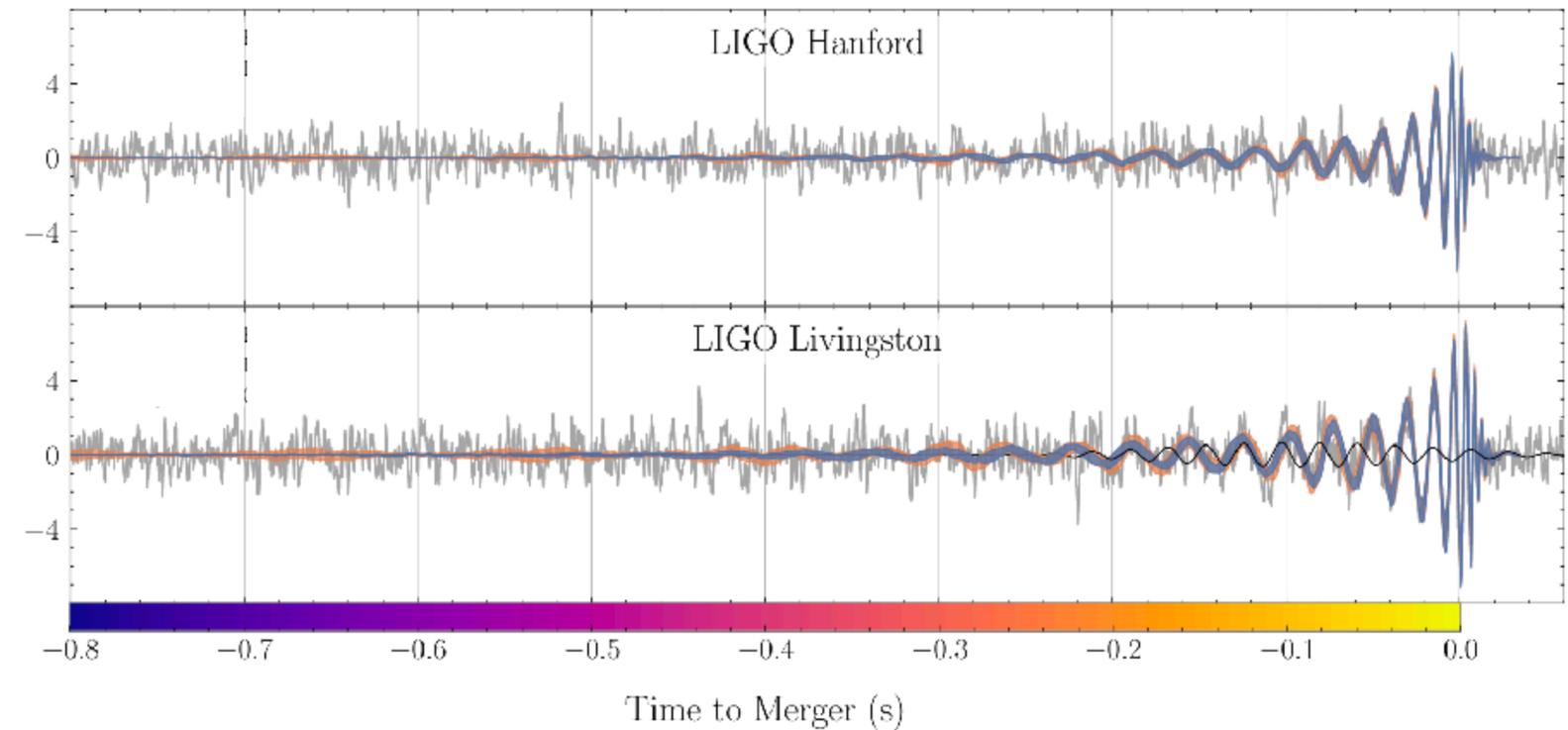
Measured data d

What are gravitational waves?

Black holes merge → Emit gravitational wave → Measured in detectors



Universe

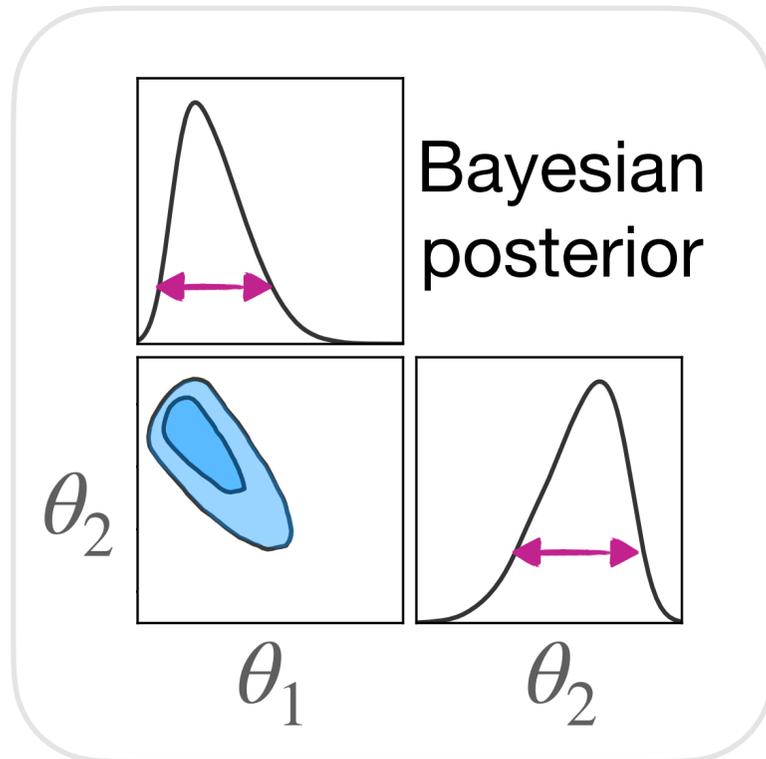


Measured data d

Posterior Estimation



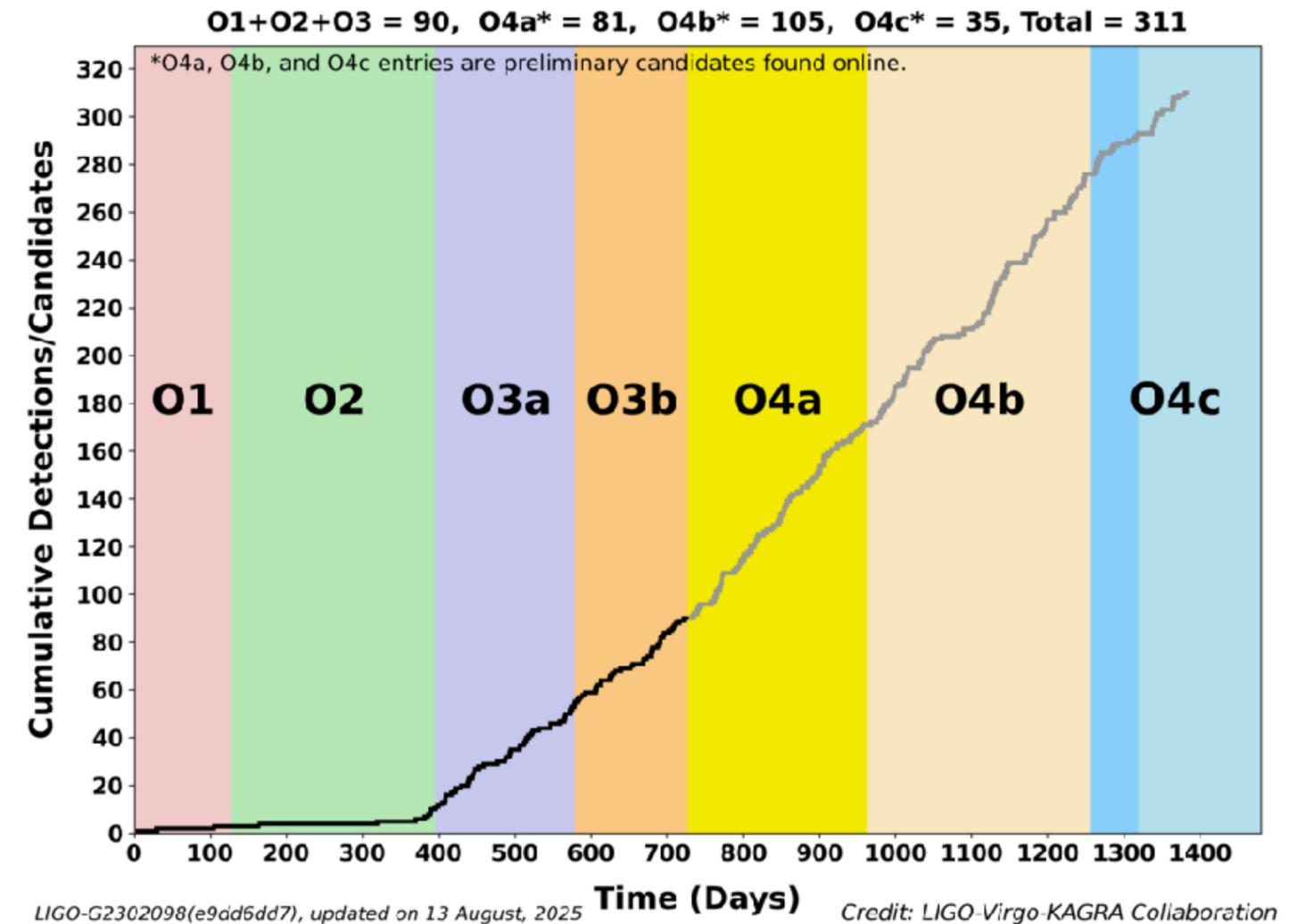
$$p(\theta | d)$$



Bayesian posterior

Why do we need ML?

- Increasing number of events:
 - Currently: ~ 5 per week
 - Future: ~ 200 per day
- Standard methods need **minutes - hours** for a **single event**
- **ML to speed it up!**



Neural Posterior Estimation (NPE)

A simulation-based inference technique

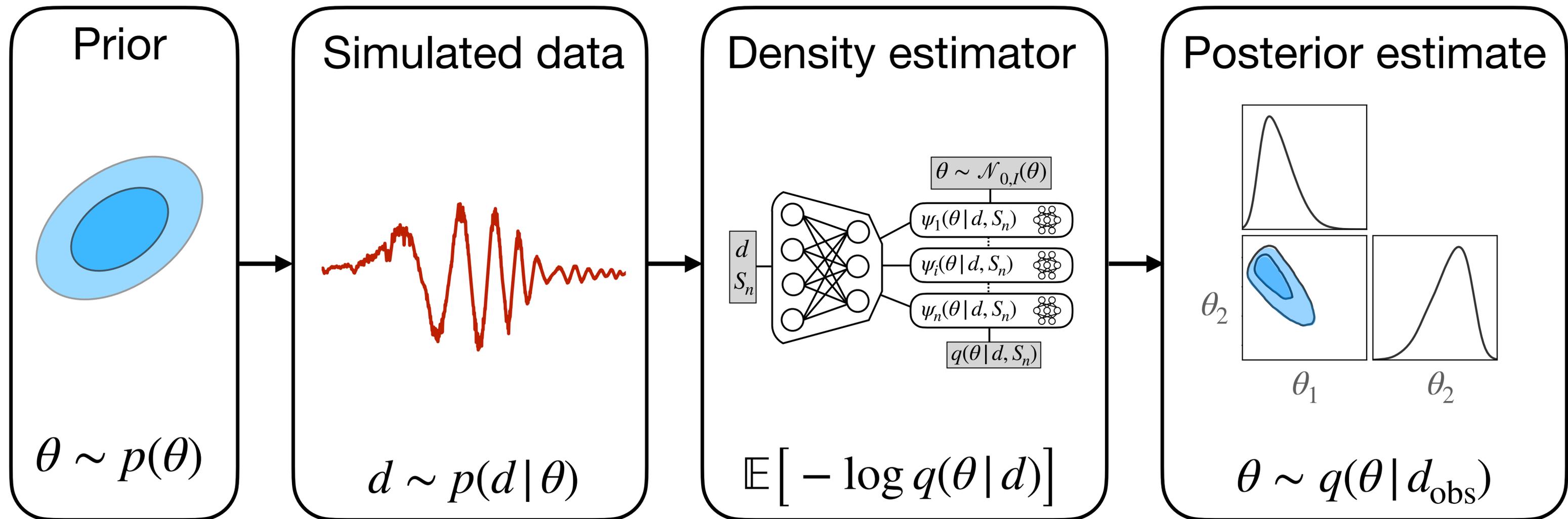
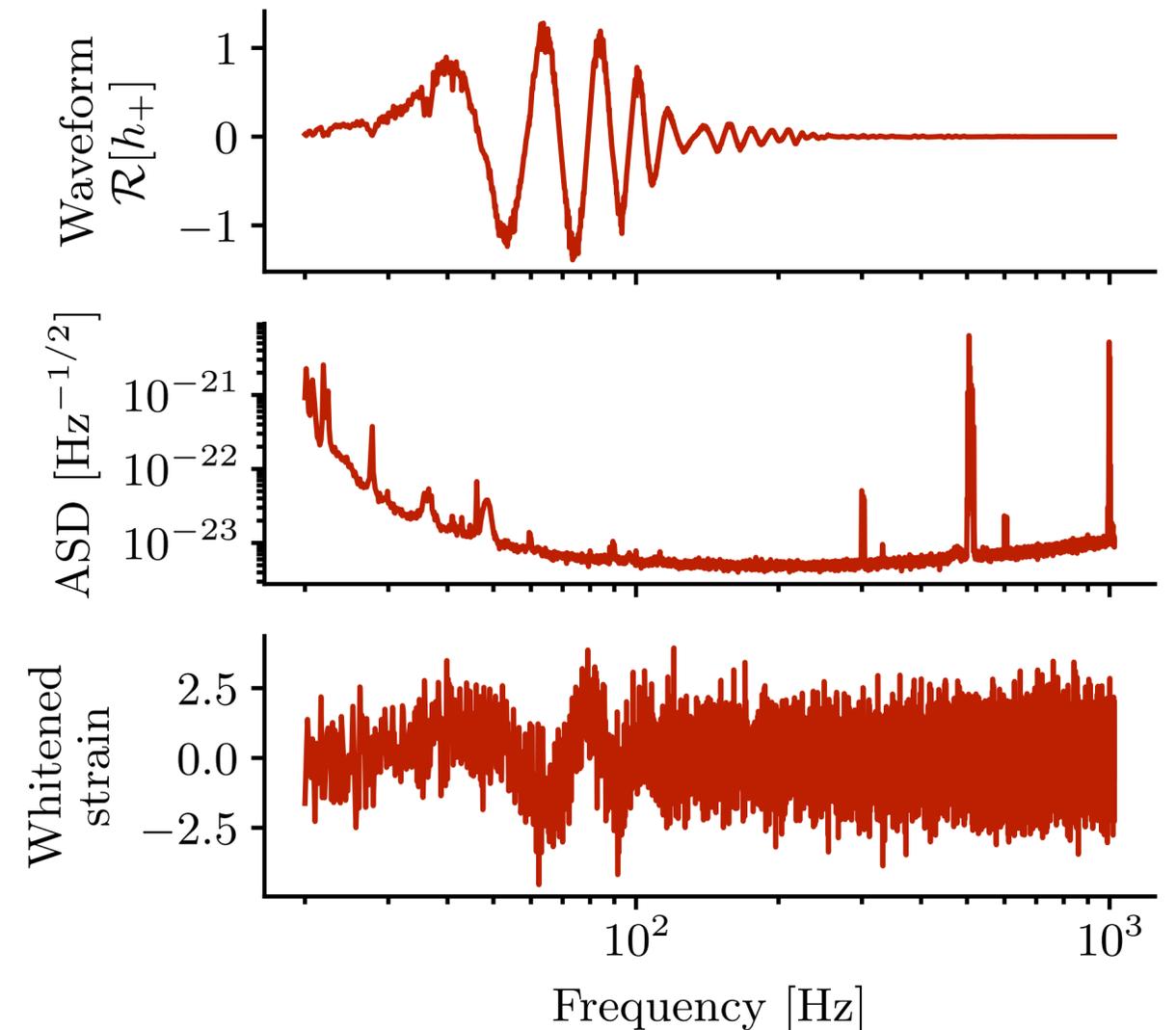


Figure inspired by Macke et al.
Cranmer+ 2020, arXiv:1911.01429

NPE for gravitational waves

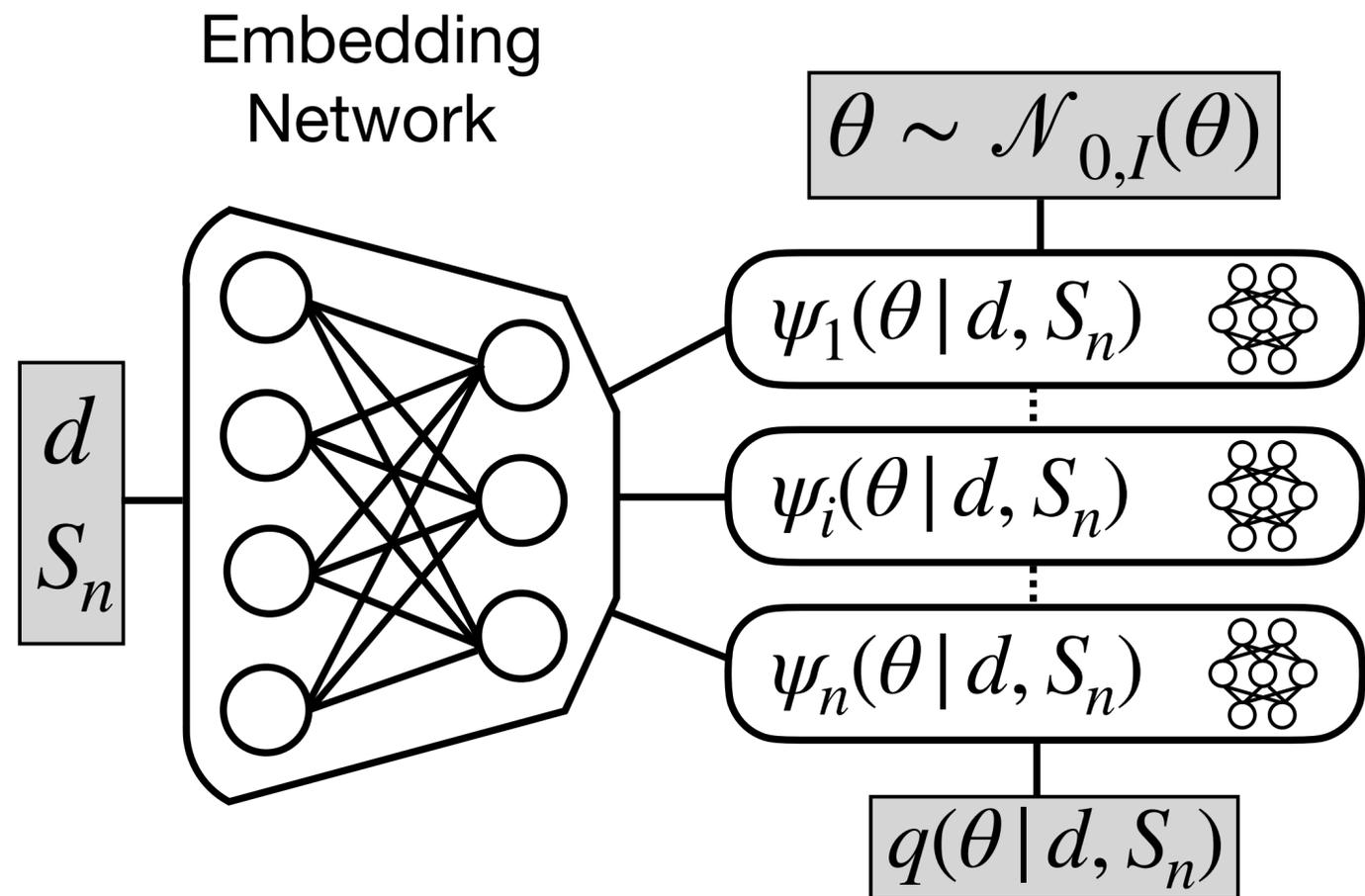
- Generate simulated waveforms: $\theta \sim p(\theta)$, $h = \text{simulator}(\theta) \rightarrow \{\theta, h\}$
- Add realistic noise $S_n(f)$ to the waveform
 1. Sample noise $n^{(i)} \sim \mathcal{N}(0, S_n^{(i)})$
 2. Add to waveform $d^{(i)} = h(\theta) + n^{(i)}$
- Train density estimator



Training the model

- Provide data d and noise curve S_n to embedding network

- Train with negative log-likelihood loss $\mathcal{L} = - \mathbb{E}_{\theta \sim p(\theta), d \sim p(d|\theta)} [\log q(\theta | d)]$



Conditional
Normalizing
Flow

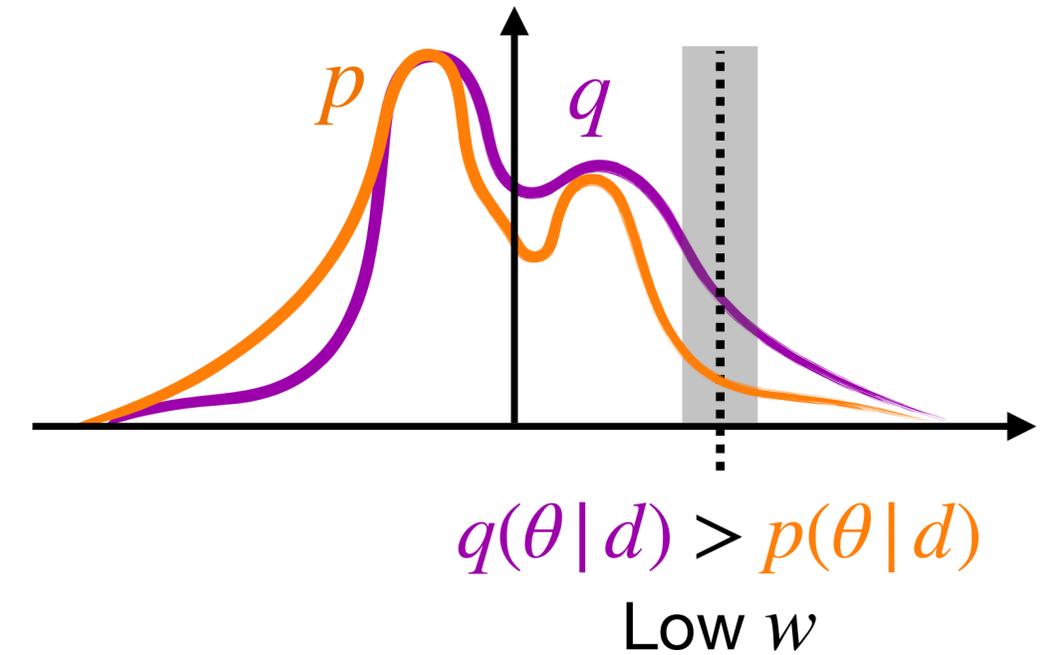


But what if the model is wrong?

- Importance sampling to validate model & reweigh samples towards true posterior

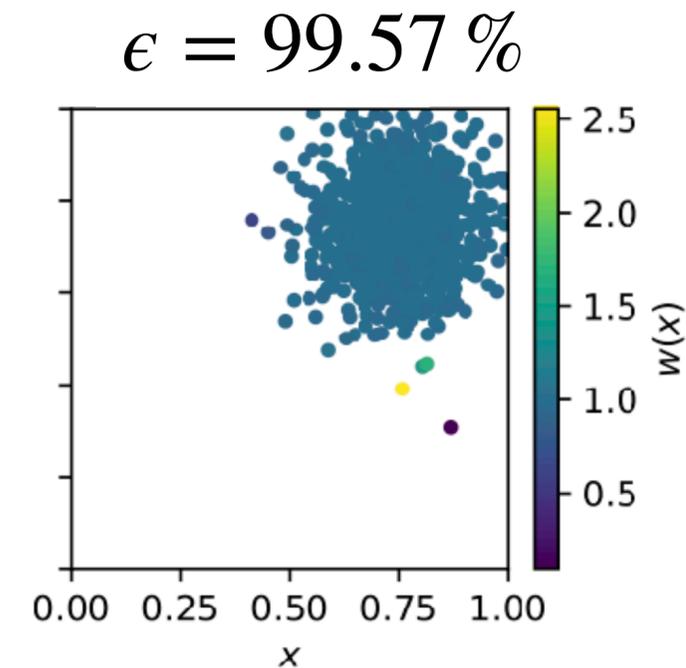
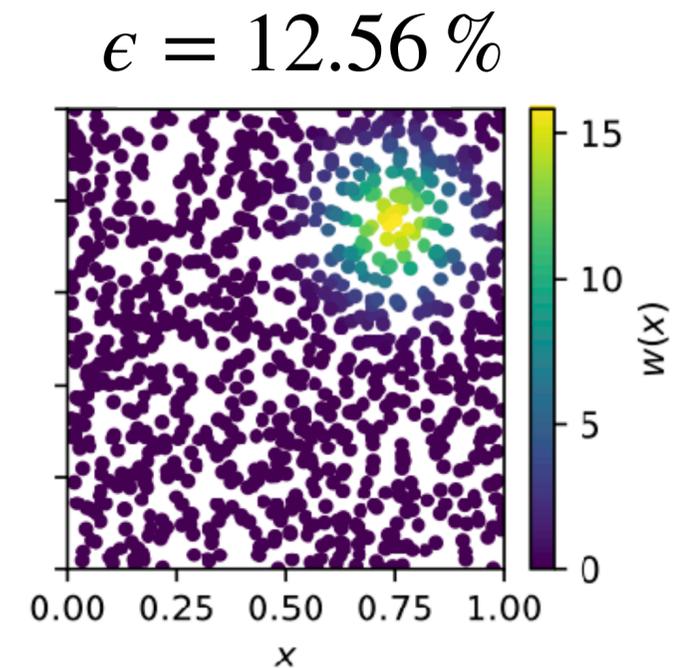
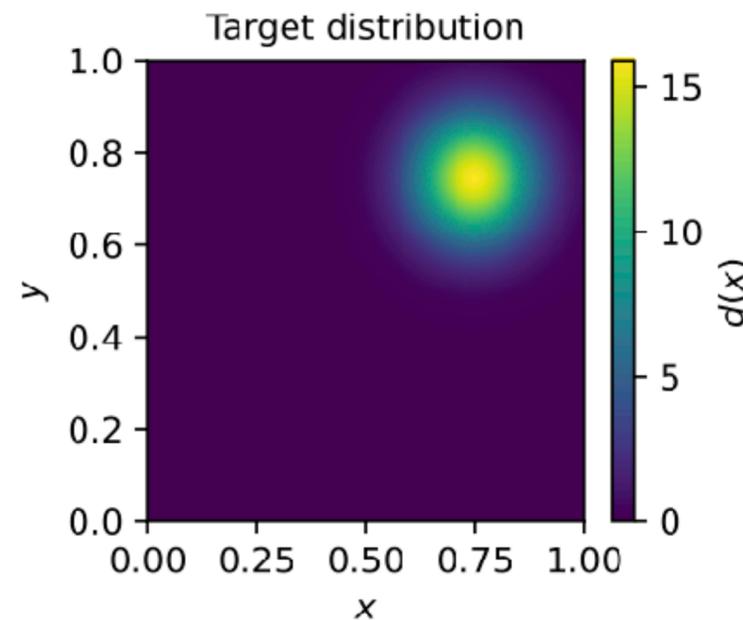
$$\frac{p(\theta | d)}{q(\theta | d)} \propto w = \frac{p(d | \theta) p(\theta)}{q(\theta | d)}$$

\propto Gaussian noise
 Known



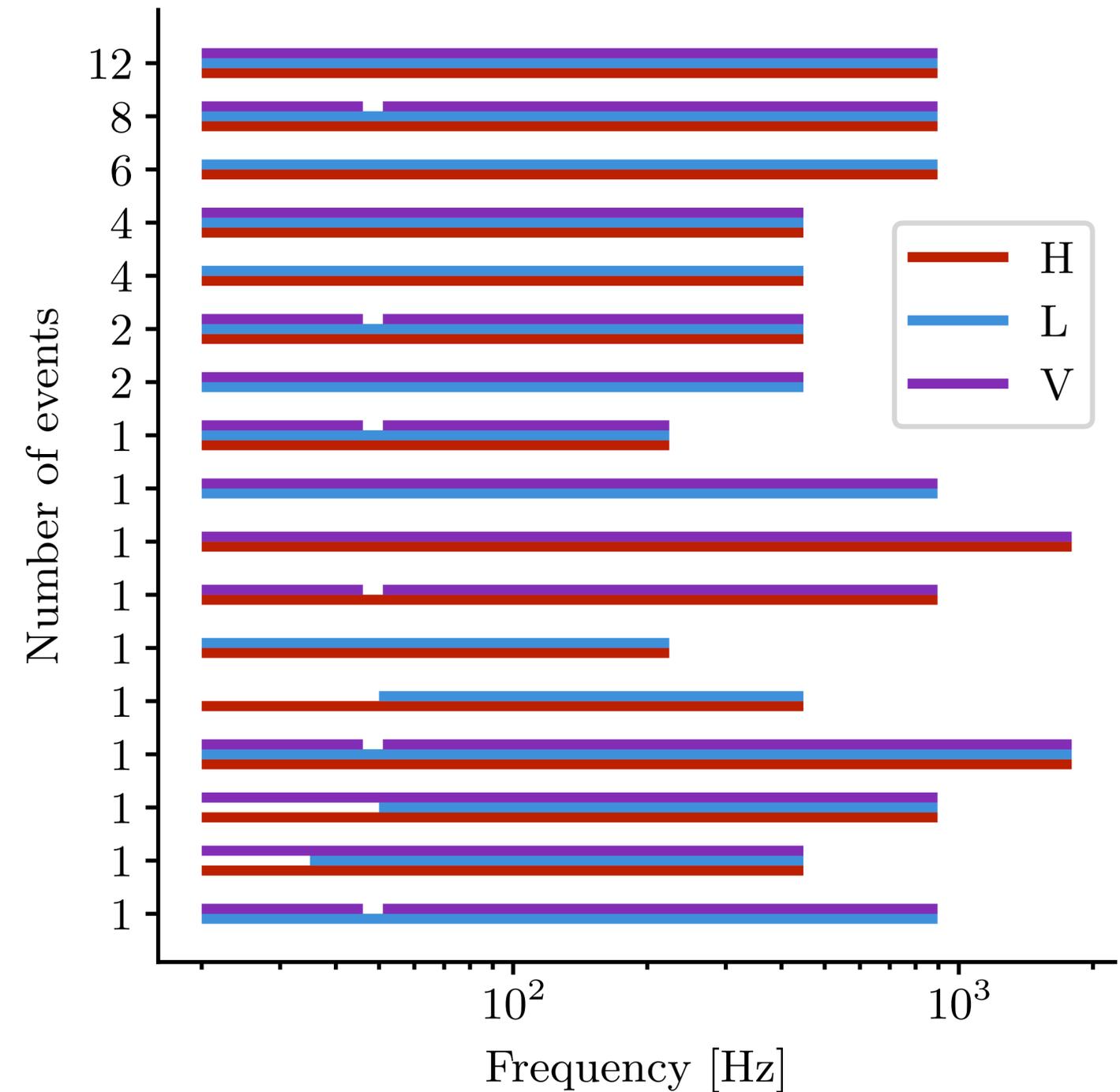
- Sample efficiency:

$$\epsilon = \frac{1}{N} \frac{\left(\sum_{i=1}^N w_i \right)^2}{\sum_{i=1}^N w_i^2}$$



Real data is messy

- Data analysis settings vary
 - Detectors
 - Frequency ranges
 - ...
- **NPE cannot deal with changing inputs**
→ **Retraining required**



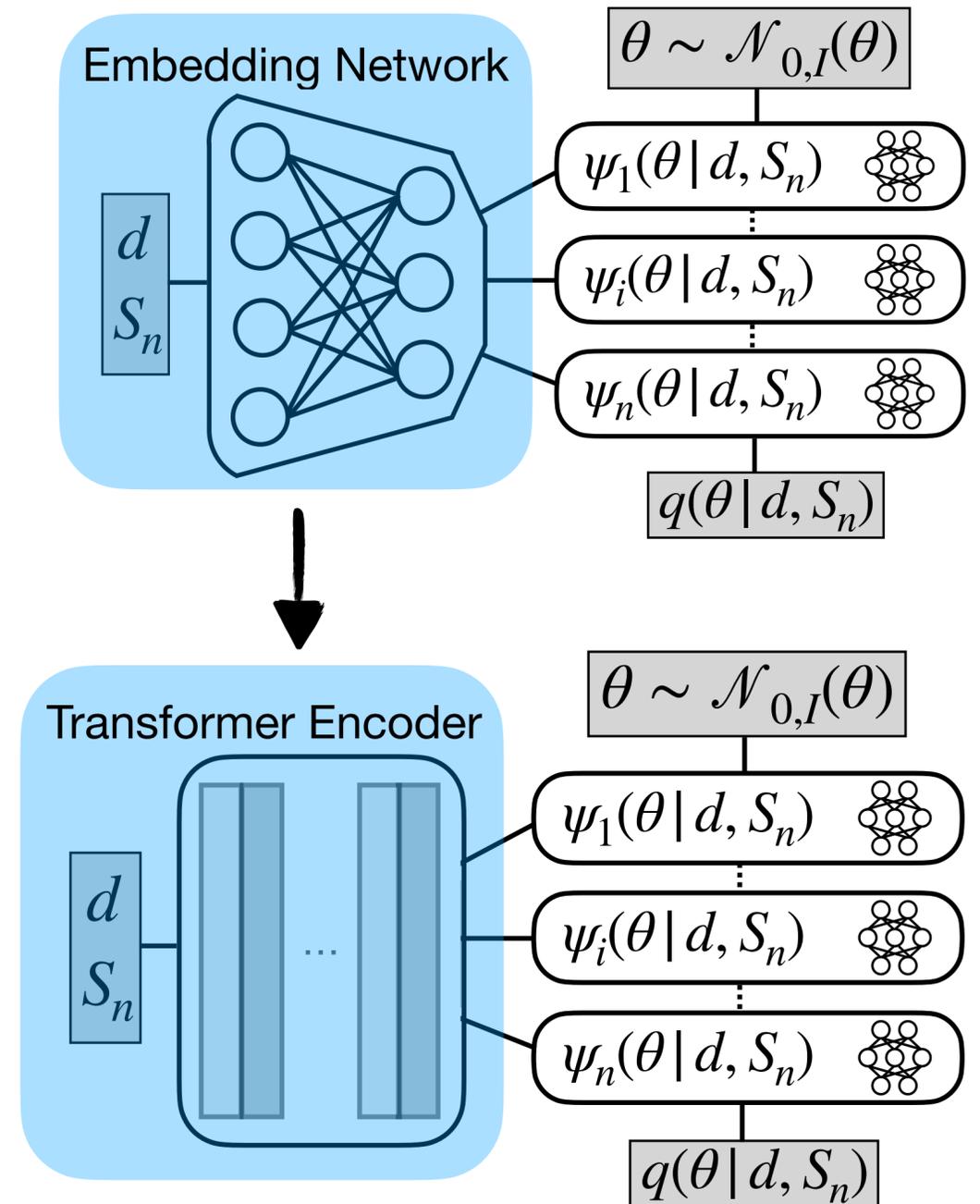
How do we make DINGO flexible?

- Replace inflexible embedding network with transformer encoder²
- Train with signals of varying lengths

6 Tokens | love gravitational waves!
9 Tokens | Gravitational wave data analysis is the best.

- Adjust data analysis settings at inference time

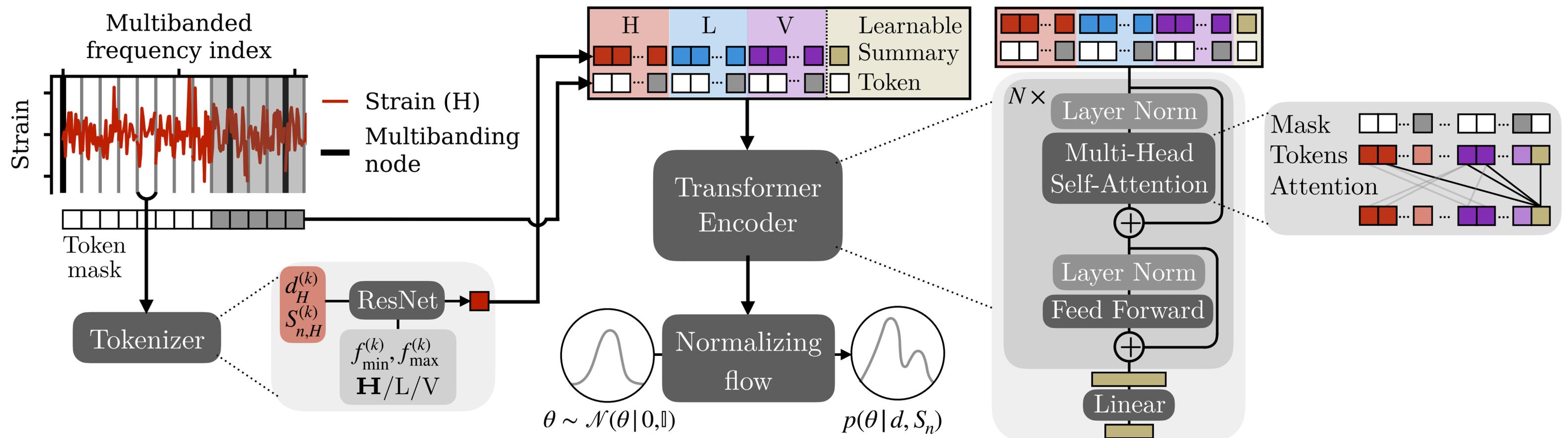
⇒ **DINGO-T1**



²Similar to Papalini+

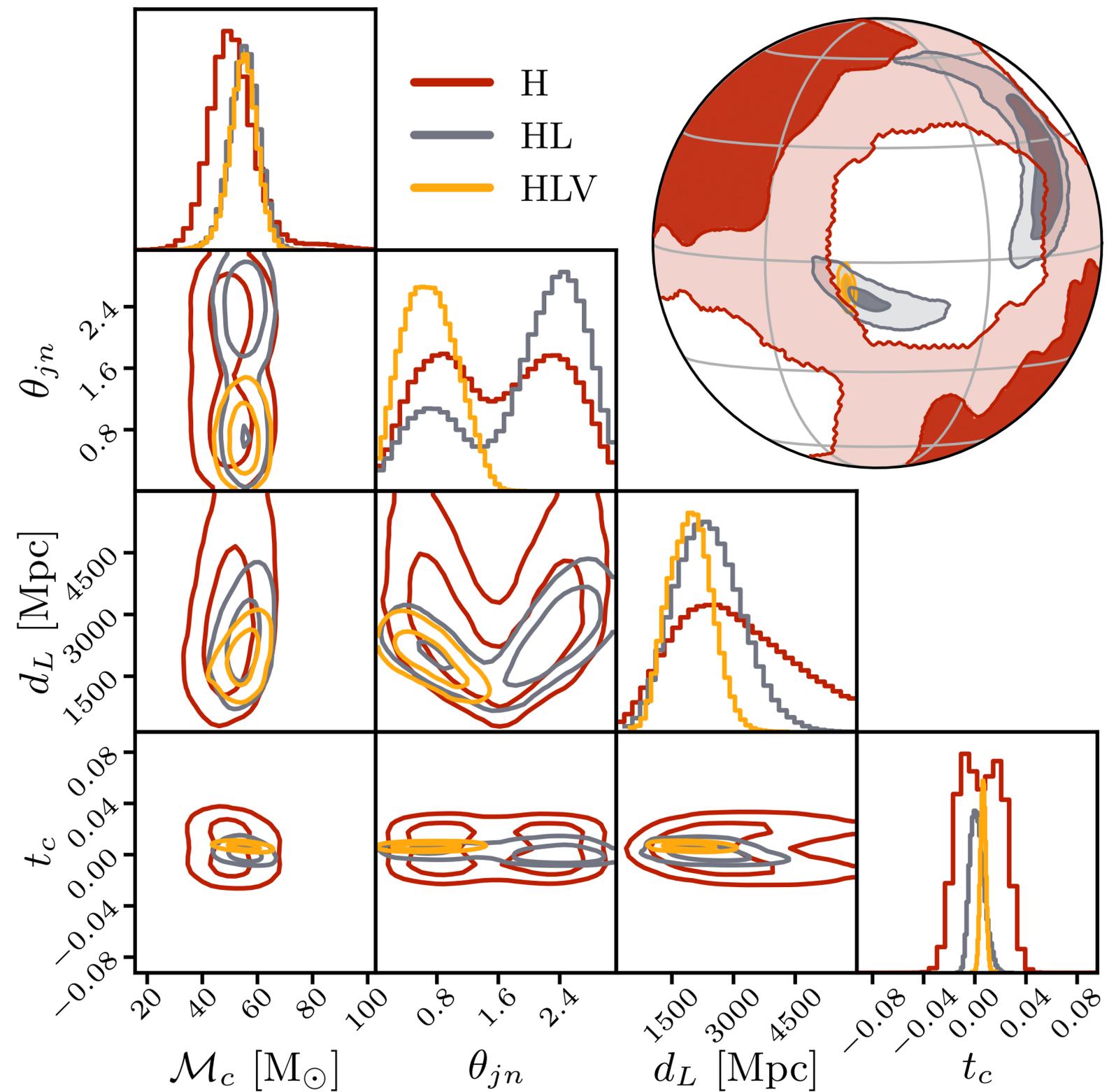
DINGO-T1 Architecture

- Shared tokenizer across detector and frequencies
- Extract information via summary token
- End-to-end training



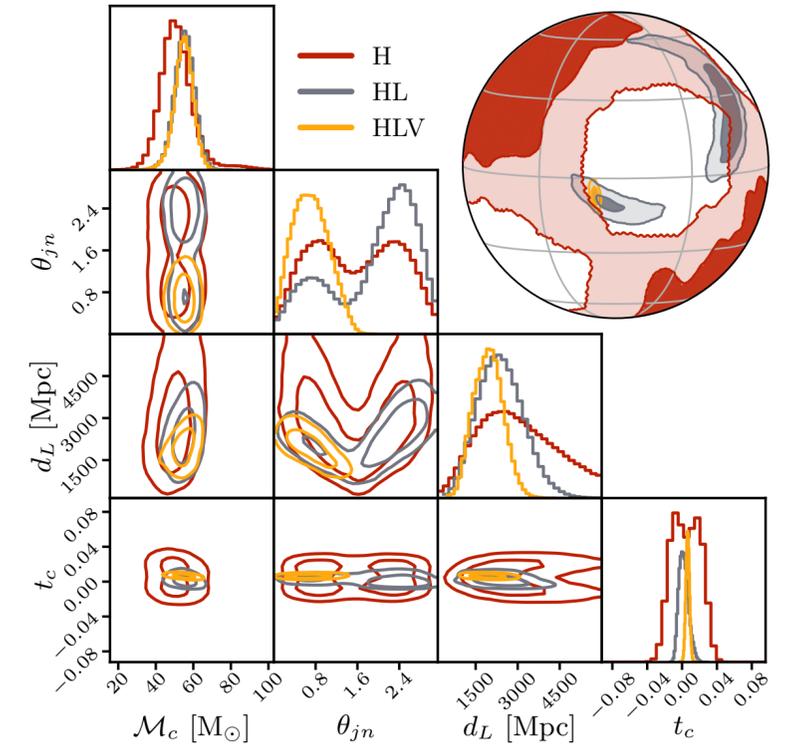
What can we do with DINGO-T1?

- Same event, different detectors



Summary: DINGO-T1

- Flexibility of DINGO-T1 allows us to do ...
 - ... quickly analyze a lot of events
 - ... change the settings at inference time
- All analyses in this paper would have required training **94 separate DINGO models!**
- Model & Tutorial online



Make this plot yourself!



The DINGO Pack



Maximilian Dax



Stephen Green



Annalena Kofler



Nihar Gupte



Alex Roussopoulos



Samuel Clyne



Ashwin Girish



Cecilia Fabbri



Lorenzo Pompili



Alexandre Göttel



Michael Pürrer



Vincent Berenz



Jonathan Gair



Jakob Macke

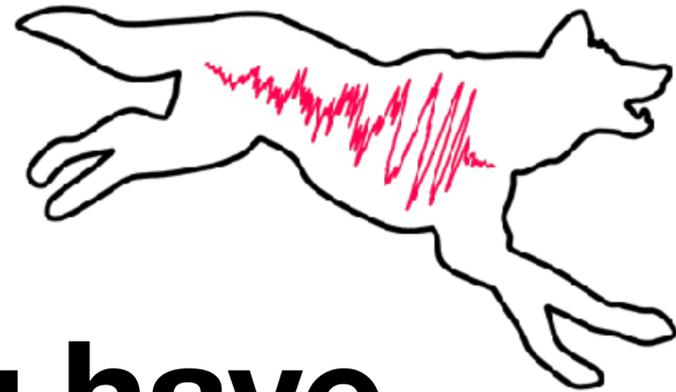


Bernhard Schölkopf



Alessandra Buonanno

Do you have any questions?



MAX PLANCK INSTITUTE
FOR INTELLIGENT SYSTEMS



Flexible Gravitational-Wave Parameter Estimation with Transformers

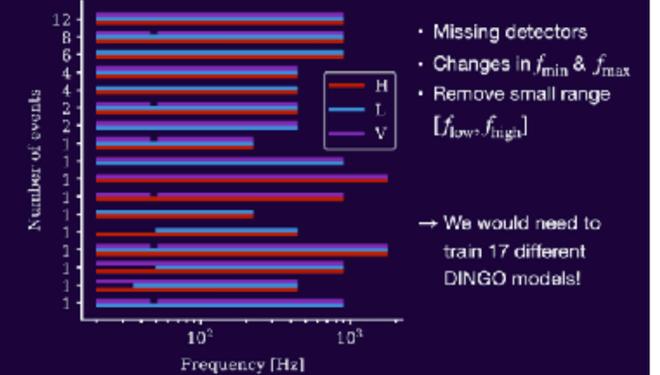
Annalena Kofler^{1,2}, Maximilian Dax^{1,2,4}, Stephen R. Green⁵, Jonas Wildberger³, Nihar Gupta^{2,6}, Jonathan Gair², Jakob H. Macke^{1,4,7}, Alessandra Buonanno^{2,6}, Bernhard Schölkopf^{1,2}
¹Max Planck Institute for Intelligent Systems Tübingen, ²Albert Einstein Institute Potsdam, ³ELLIS Institute Tübingen, ⁴Tübingen AI Center, ⁵University of Nottingham, ⁶University of Maryland, ⁷University of Tübingen

Gravitational waves



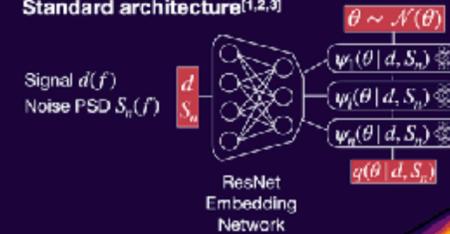
Goal: Analyze signals \rightarrow posterior distribution of black hole mergers
 Problem: Real data is messy
 \rightarrow Re-train model to adapt to different data analysis settings
 Solution: Flexible transformer architecture and masking procedure during training

Real data is messy:
48 events with 17 different data analysis settings



DINGO (Deep INference for Gravitational wave Observations)

Standard architecture^[1,2,3]

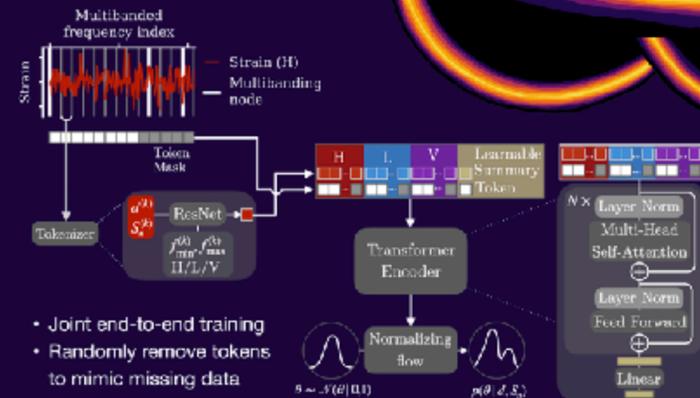


Validation with importance sampling^[3]

Compare learned NPE density and likelihood
 $\frac{p(\theta | d)}{q(\theta | d)} \propto w_i = \frac{p(d | \theta_i)p(\theta_i)}{q(\theta_i | d)}$
 Likelihood \cdot Prior
 Proposal (NPE)
 $w < 1$
 Performance criterion: Sample efficiency
 $\epsilon = \frac{1}{N} \left(\frac{\sum_i w_i}{\sum_i w_i^2} \right)^2$

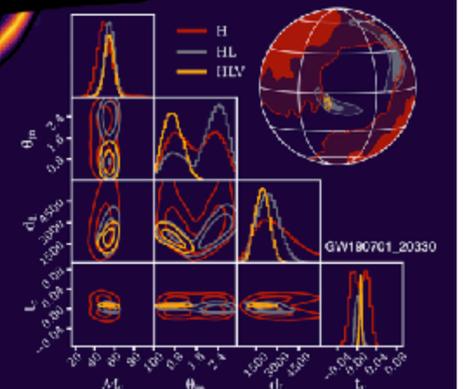


DINGO-T1: Architecture



- Joint end-to-end training
- Randomly remove tokens to mimic missing data

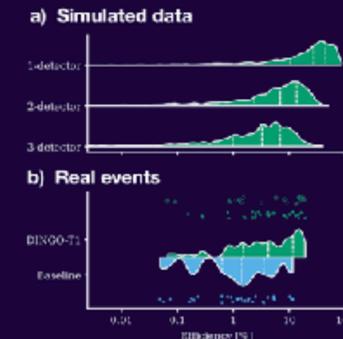
Flexible analysis



Masking strategies

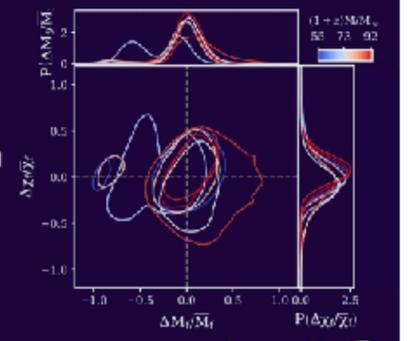


Performance



Tests of general relativity

- Analyze inspiral and postinspiral part of signal separately
- Check whether parameters agree



References [1] Dax+, Real-Time GW Science with NPE, PRL 2021
 [2] Dax+, Group Equivariant NPE, ICLR 2021
 [3] Dax+, Neural IS for Rapid and Reliable GW Inference, PRL 2023