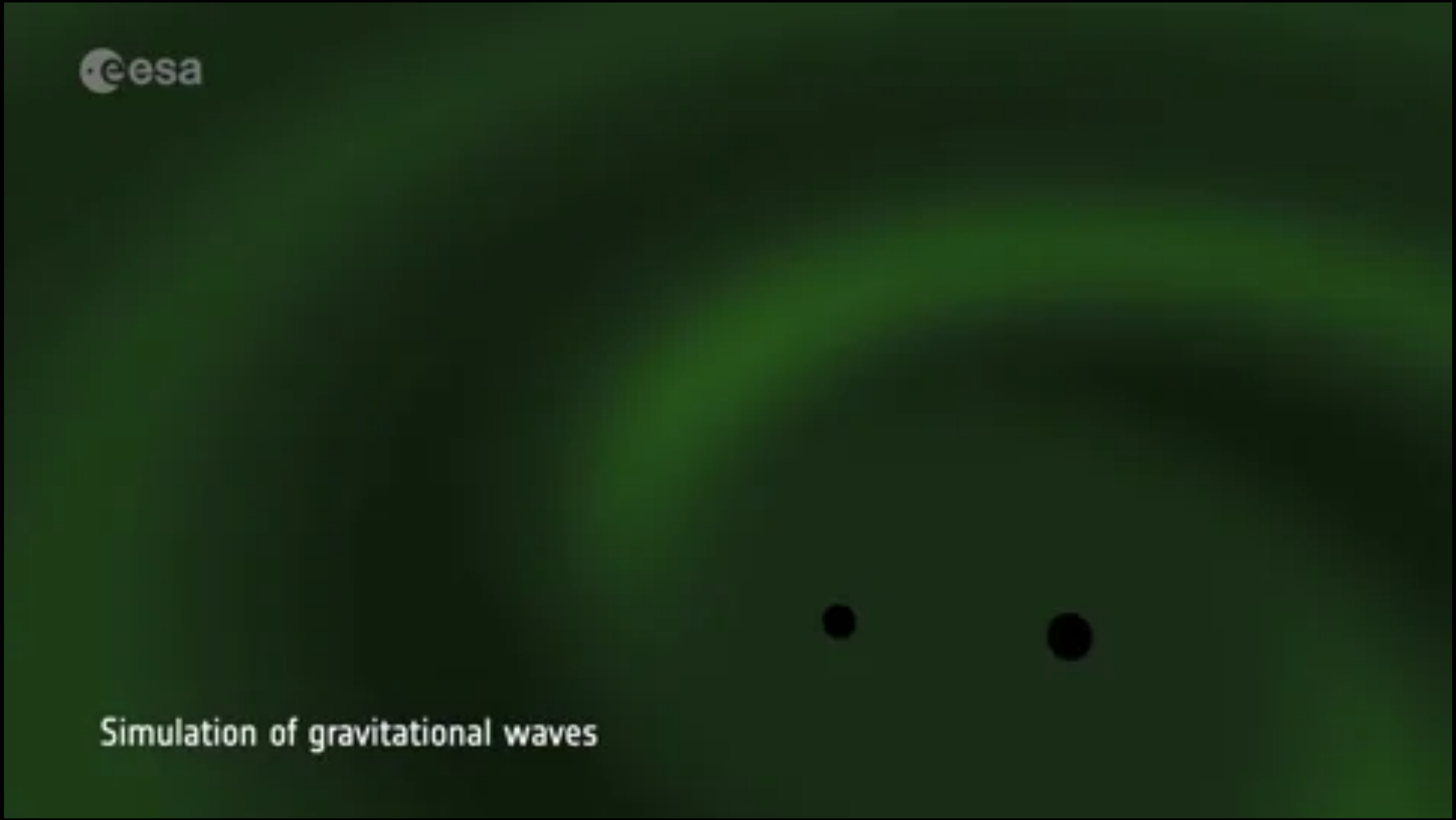
Talk by Annalena Kofler

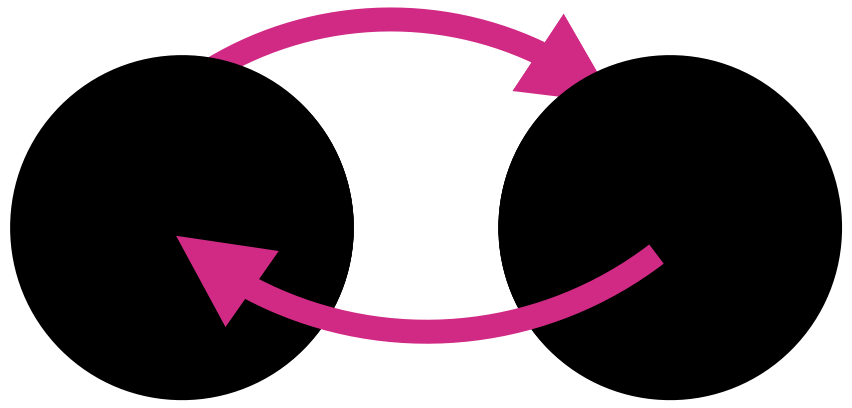
Image Credit: NASA

How can we observe black holes?


Simulation of gravitational waves

# What are gravitational waves?


Black holes merge → Emit gravitational wave → Measured in detectors



Universe

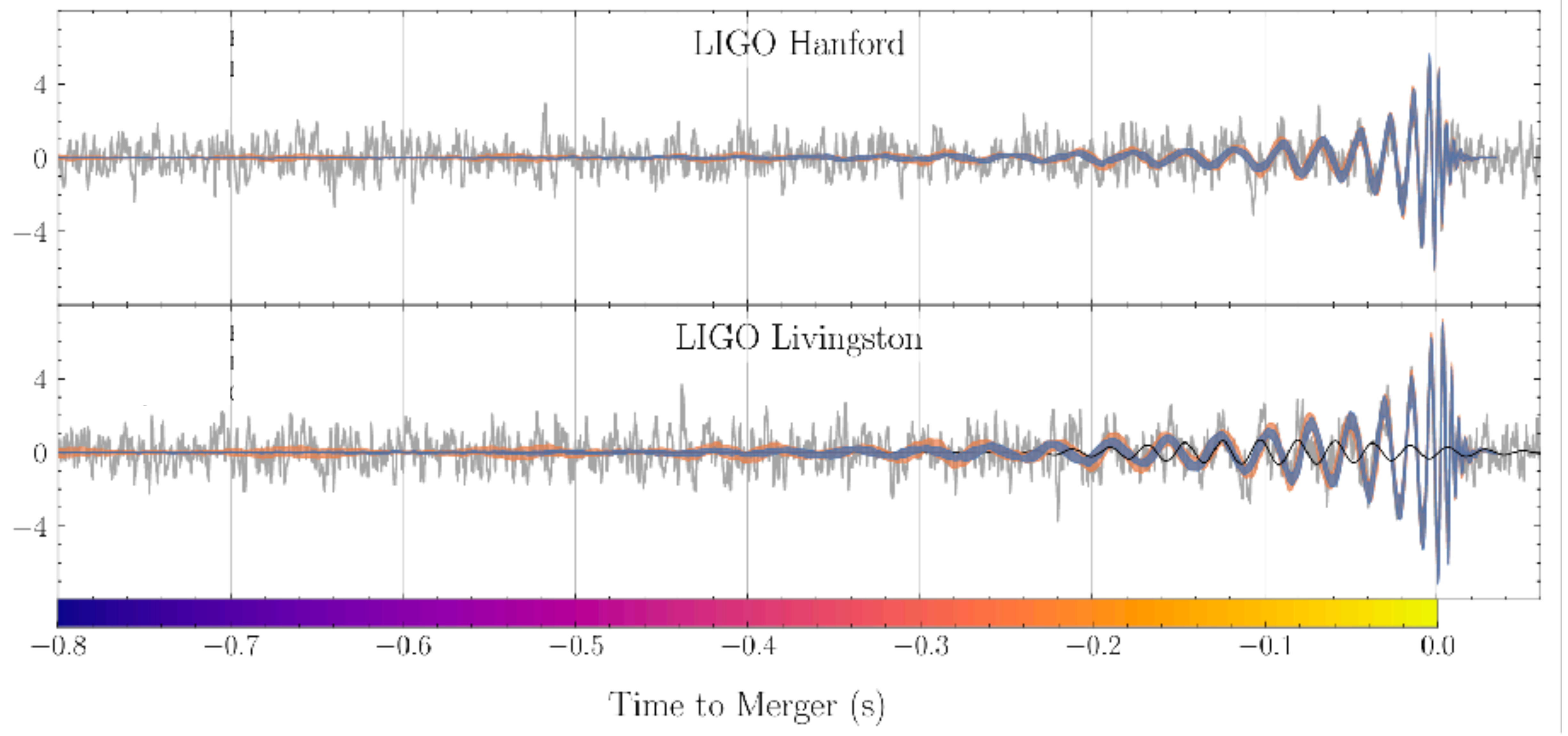
Described by physics parameters

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

Posterior Estimation

$p(\theta \mid d)$

Masses, spins, sky position, …



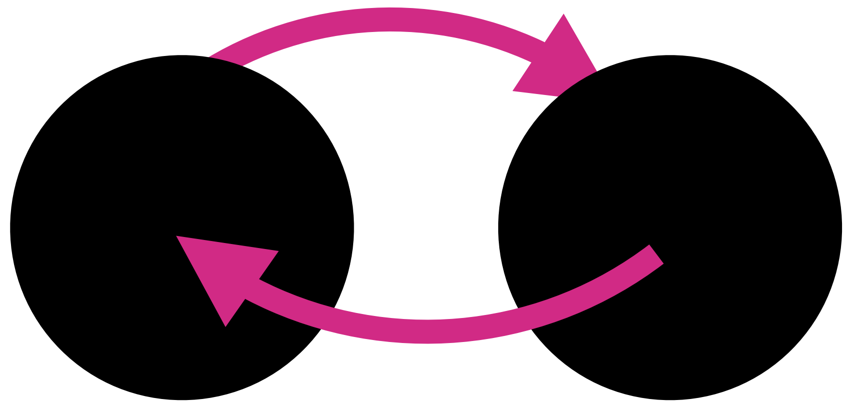

Measured data $d$



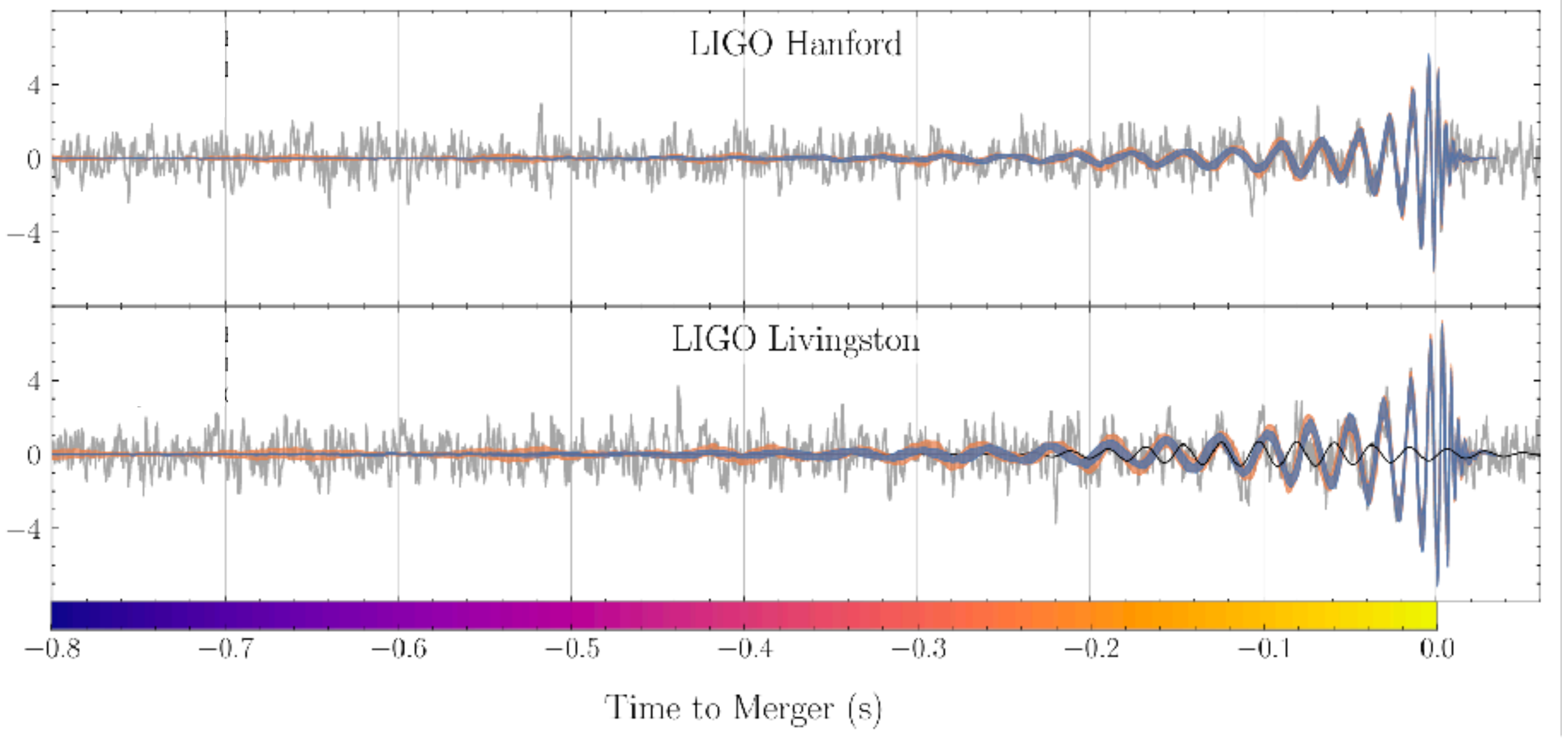
Gupte+, PRL 2025

# What are gravitational waves?


Black holes merge $\rightarrow$ Emit gravitational wave $\rightarrow$ Measured in detectors



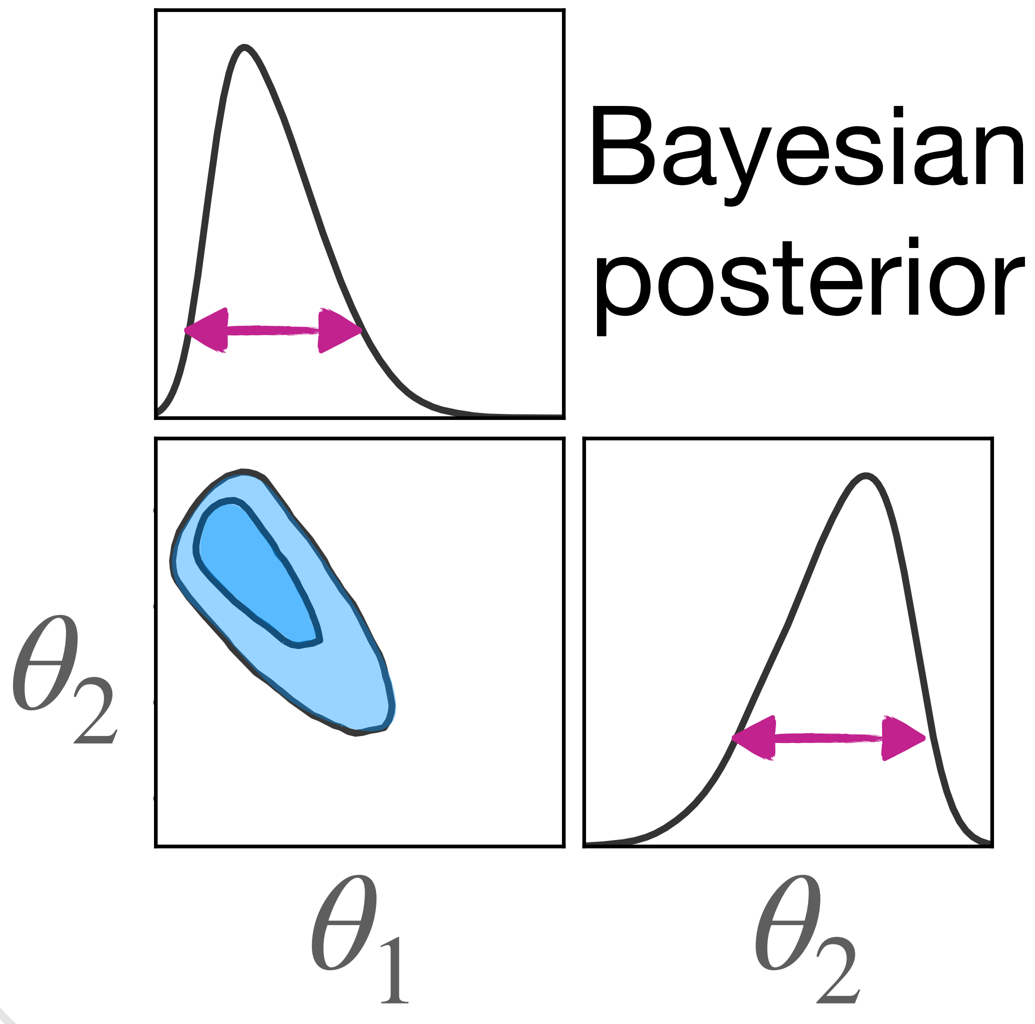
Universe




Measured data $d$

Posterior Estimation

$p(\theta \mid d)$




Bayesian posterior

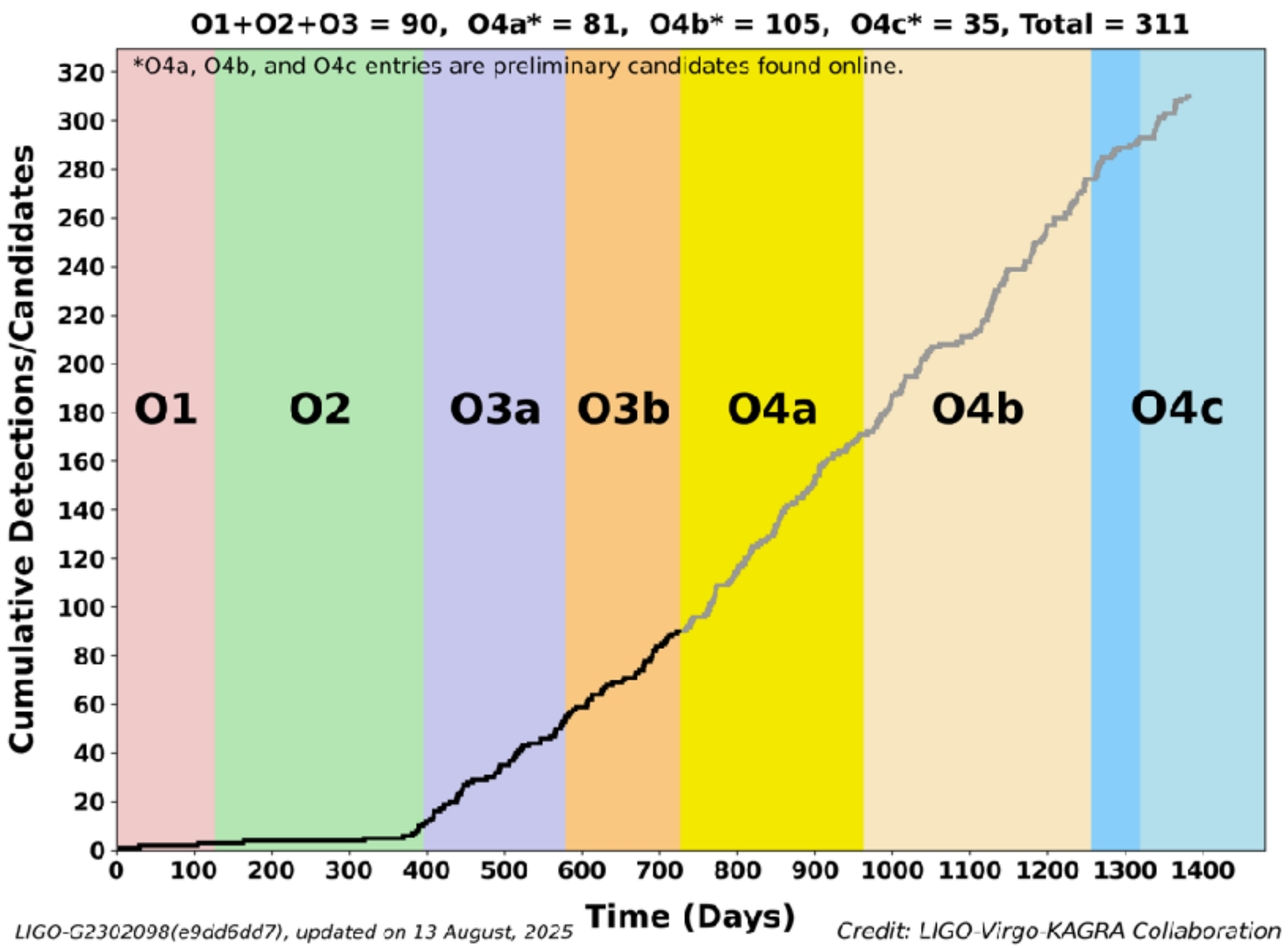




Gupte+, PRL 2025

# 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

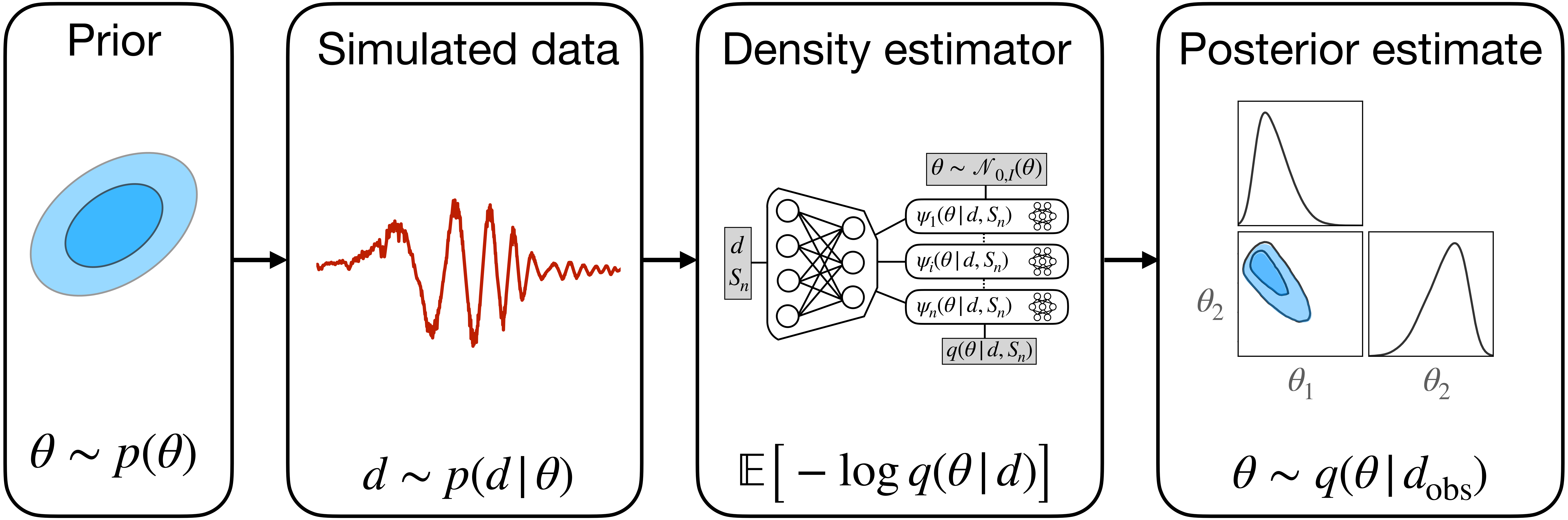



Figure inspired by Macke et al.
Cranmer+ 2020, arXiv:1911.01429

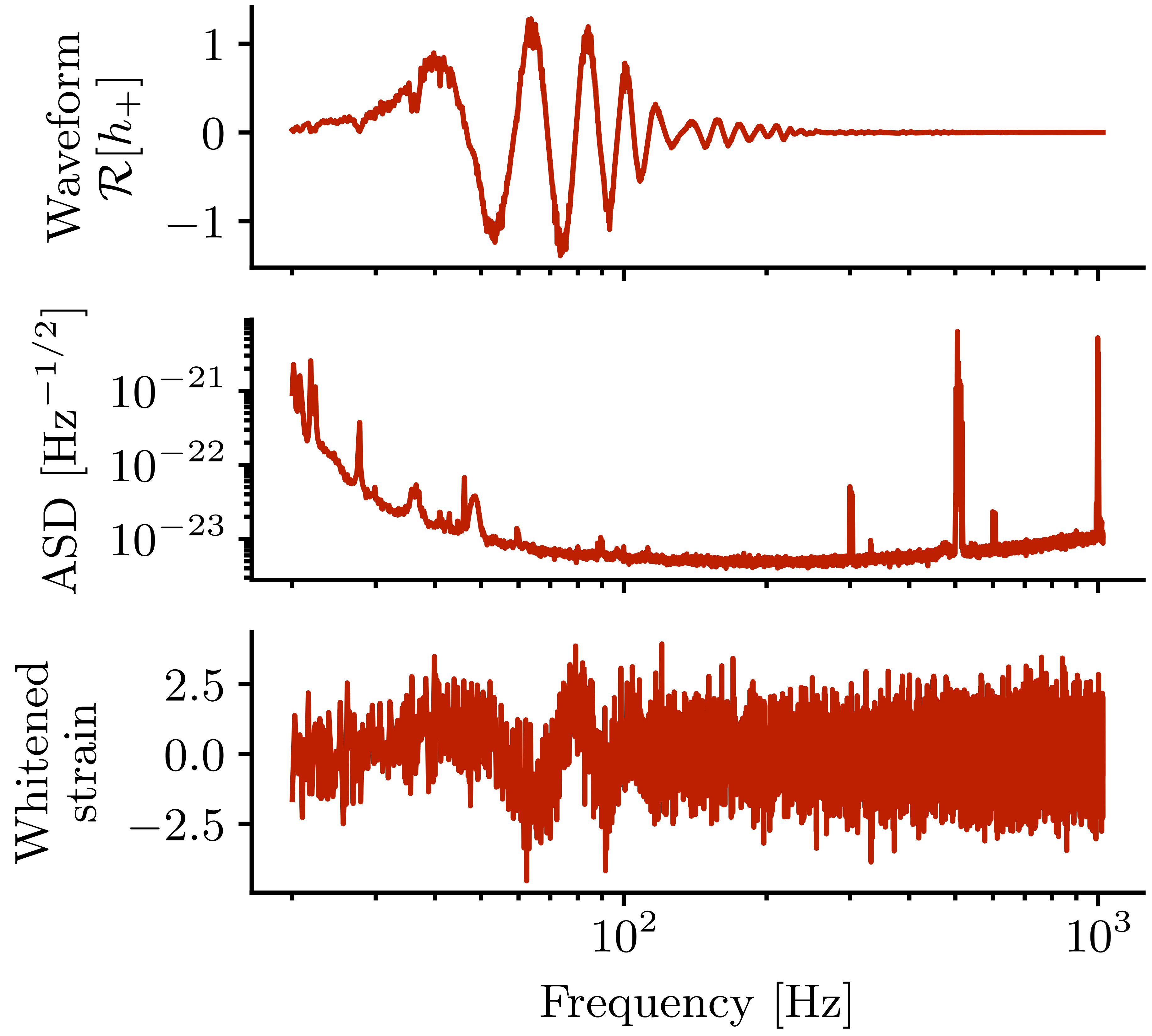
# NPE for gravitational waves

- 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



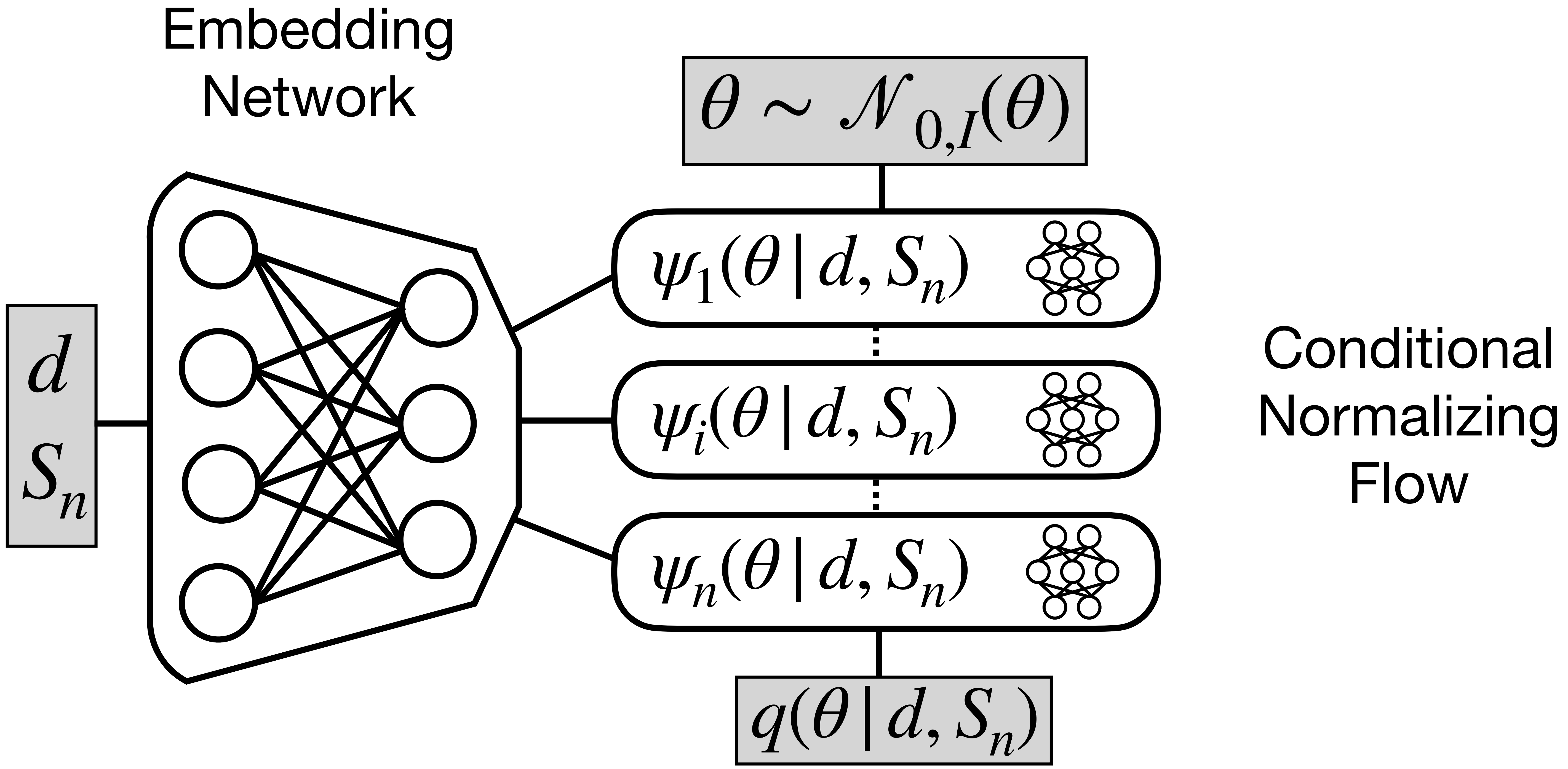





Dax+, PRL 2021, arXiv:2106.12594

# Training the model

- Provide data $d$ and noise curve $S_n$ to embedding network
- Train with negative log-likelihood loss $\mathcal{L} = -\mathbb{E}_{\theta \sim p(\theta), d \sim p(d|\theta)} \left[ \log q(\theta | d) \right]$



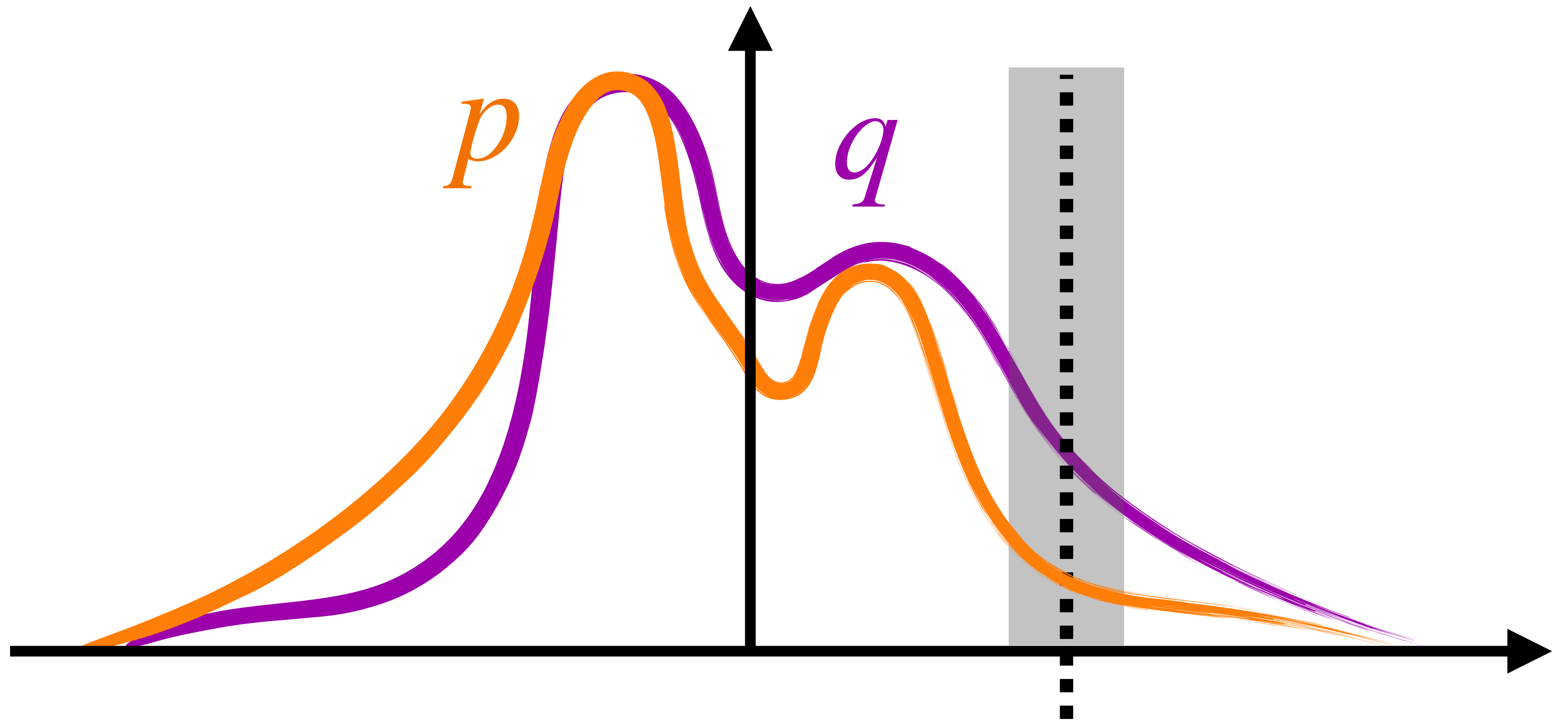

# But what if the model is wrong?

- 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

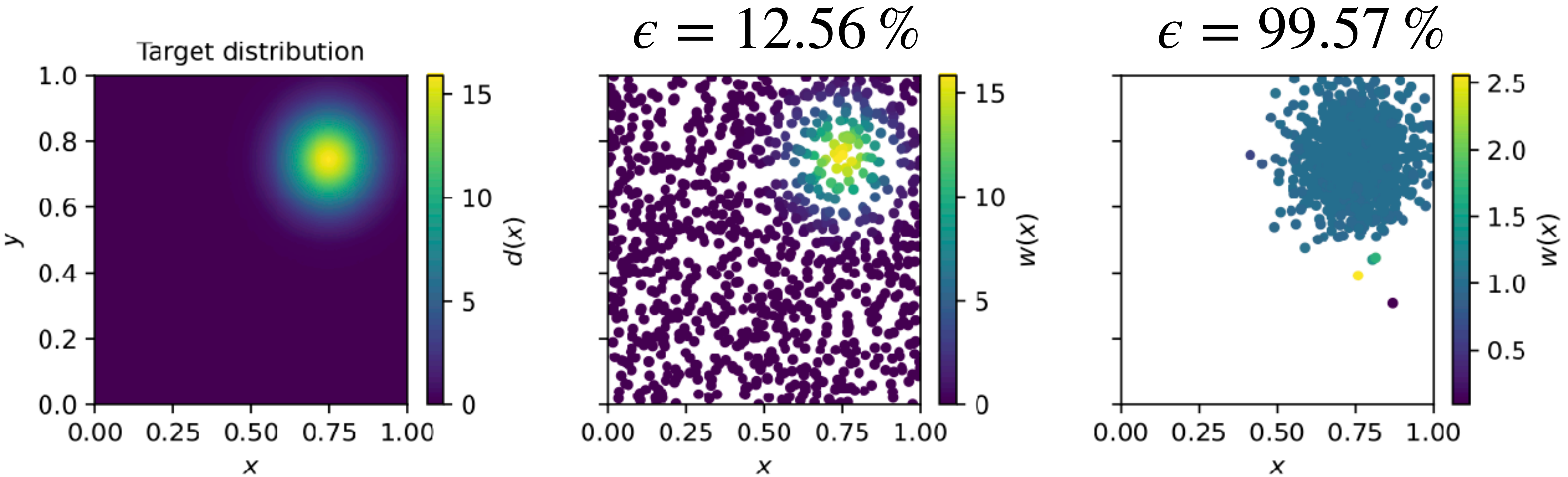



$q(\theta\,|\,d) > p(\theta\,|\,d)$

Low $w$

- Sample efficiency:

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

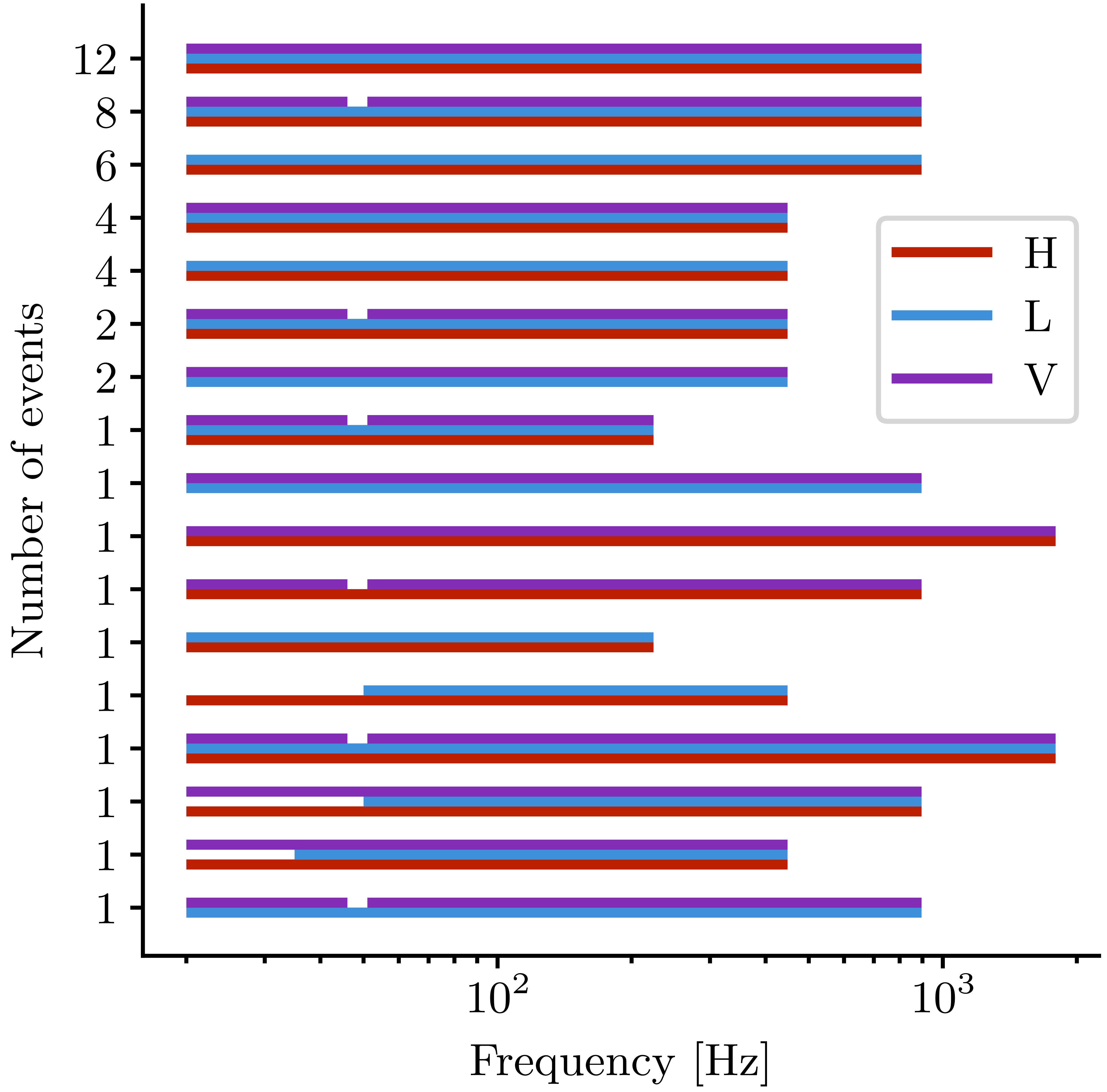





Dax+, PRL 2023, arXiv:2210.05686

# 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?

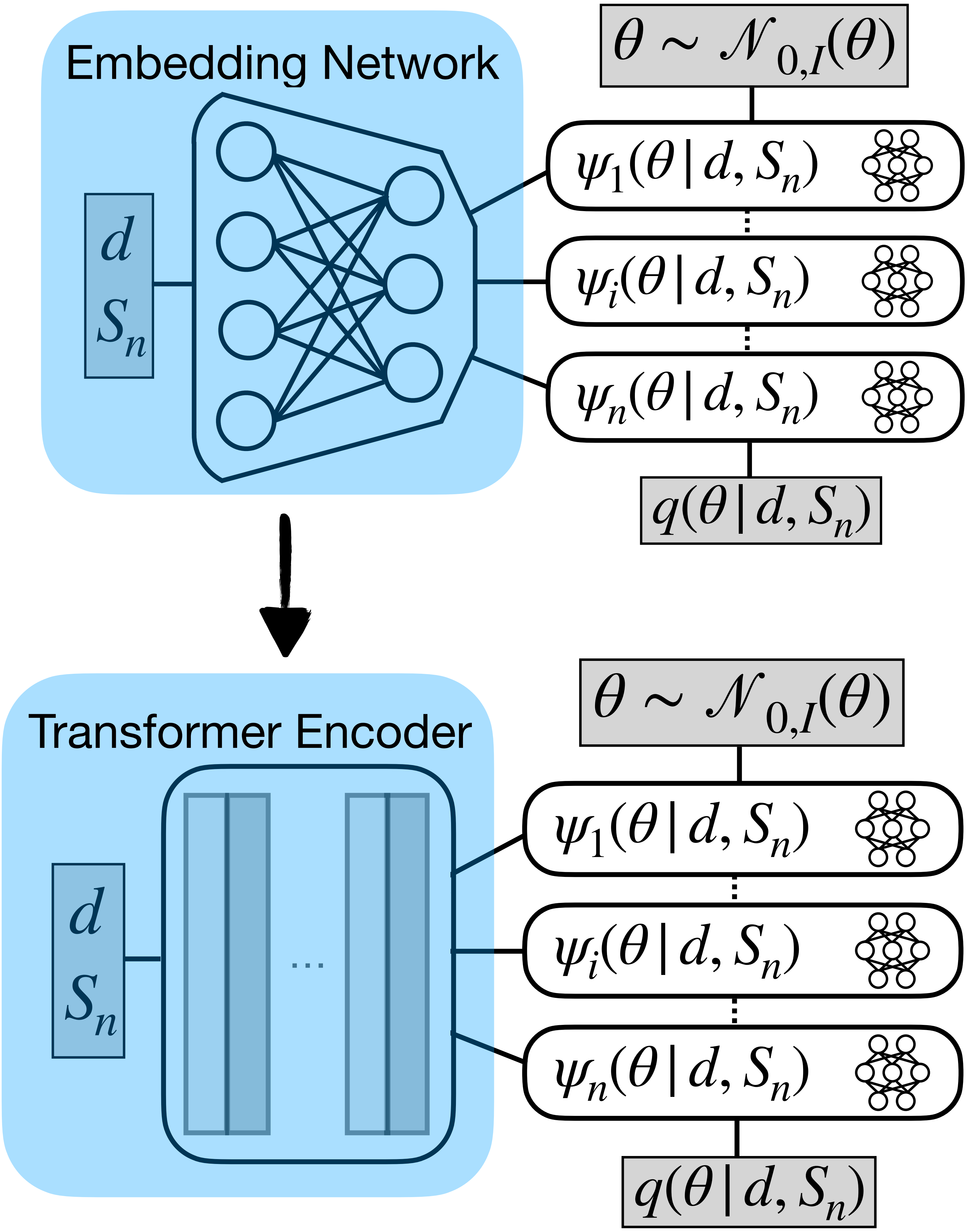
- Replace inflexible embedding network with transformer encoder[2]
- Train with signals of varying lengths

6 Tokens | I | love | gravit | ational | waves | ! |

9 Tokens | Gravit | ational | wave | data | analysis | is | the | best | . |

- Adjust data analysis settings at inference time

$\Rightarrow$ **DINGO-T1**

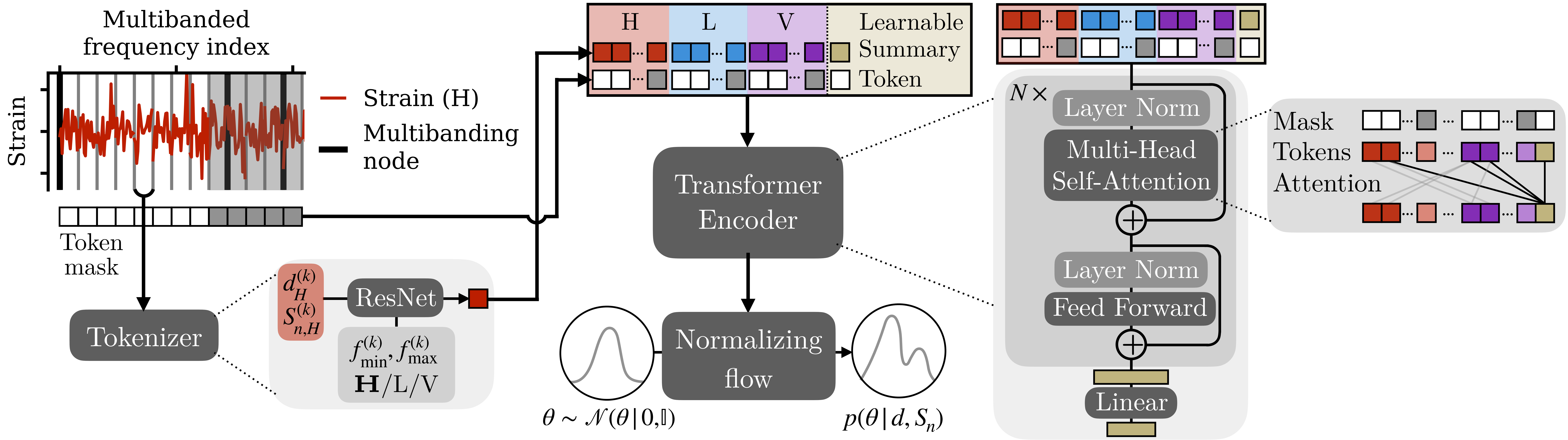





[2]Similar to Papalini+

# DINGO-T1 Architecture

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

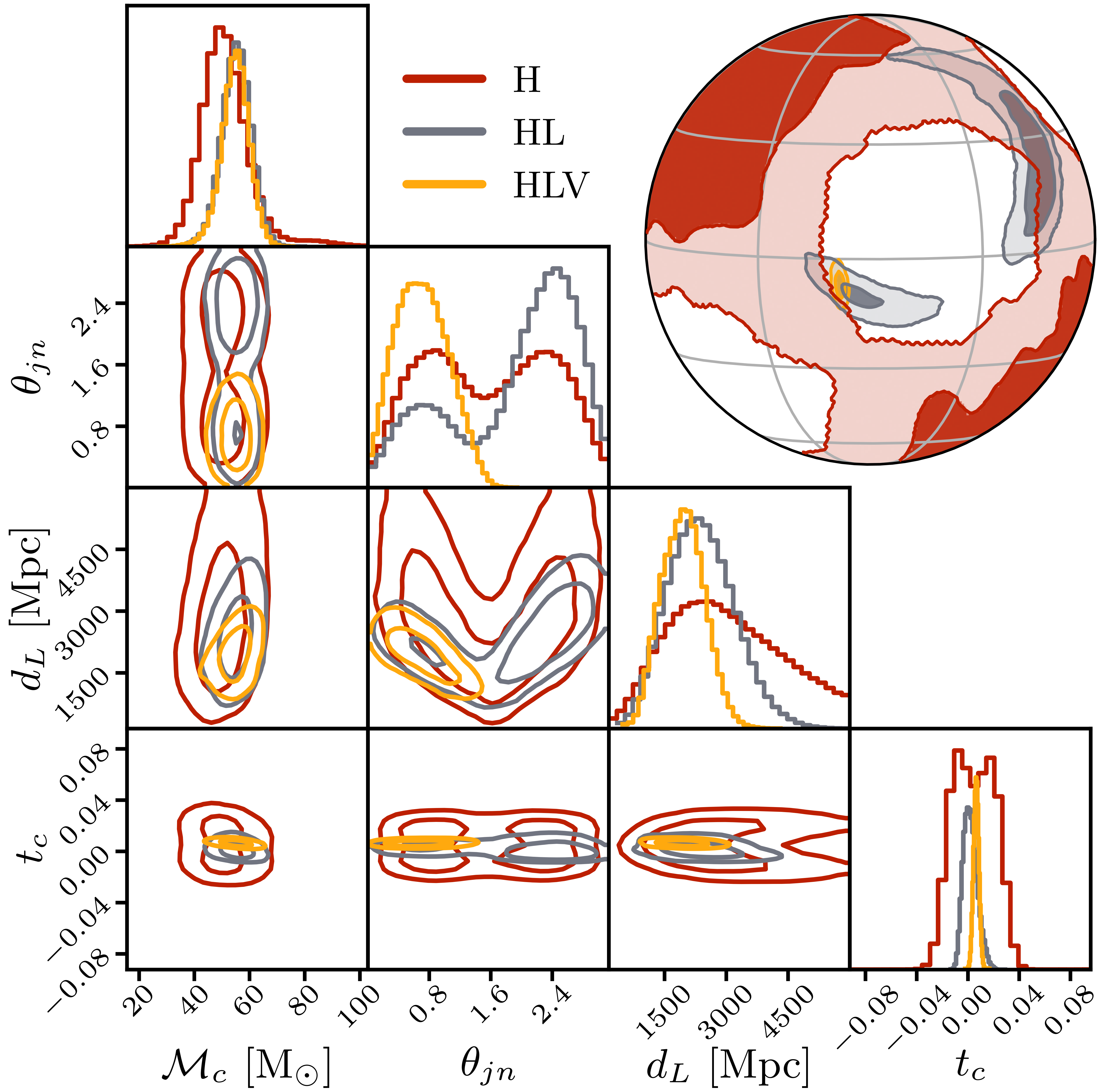





Kofler+, under review

# What can we do with DINGO-T1?

- Same event, different detectors



# Summary: DINGO-T1

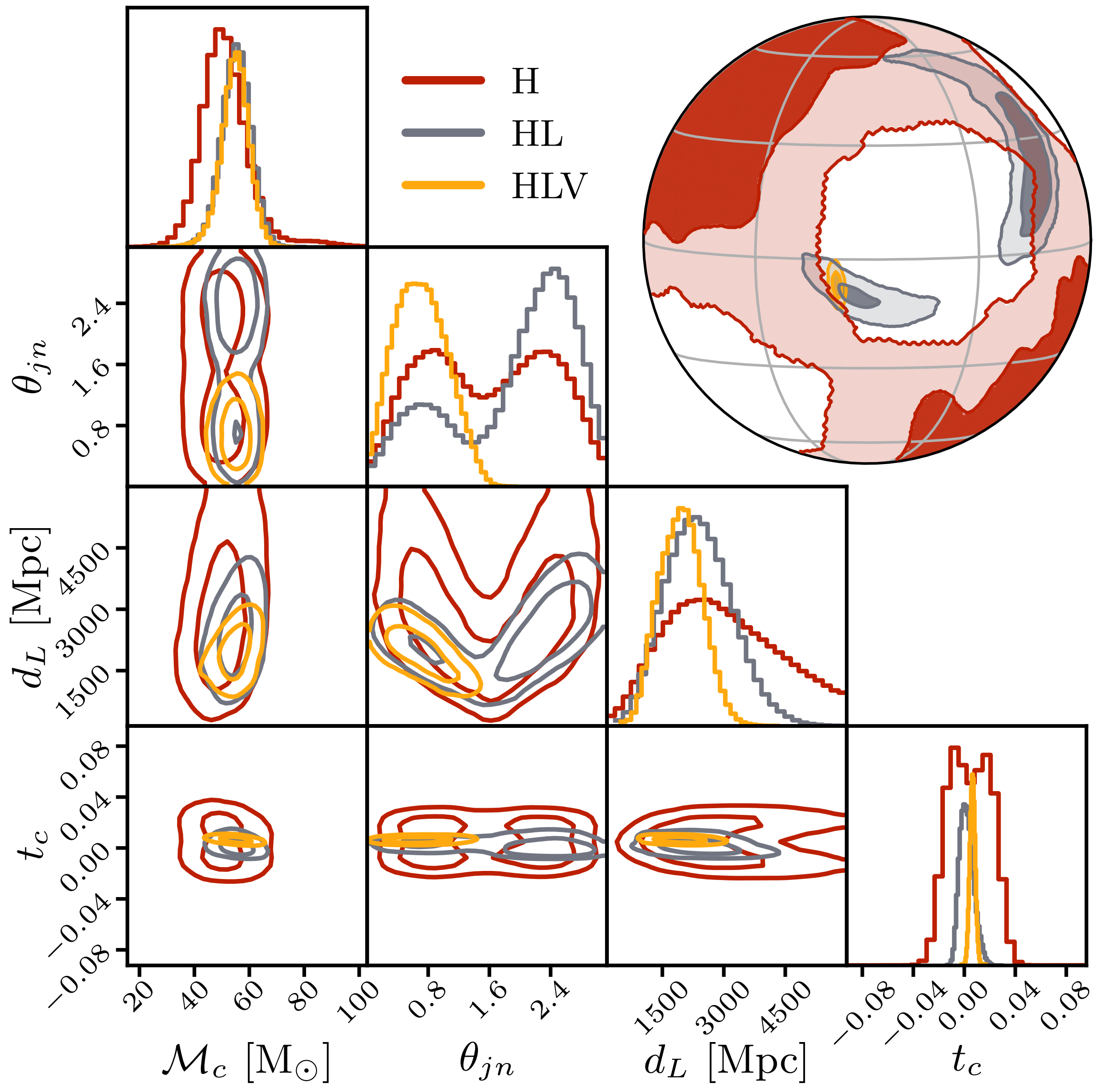
- Flexibility of DINGO-T1 allows us to do …

  … 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

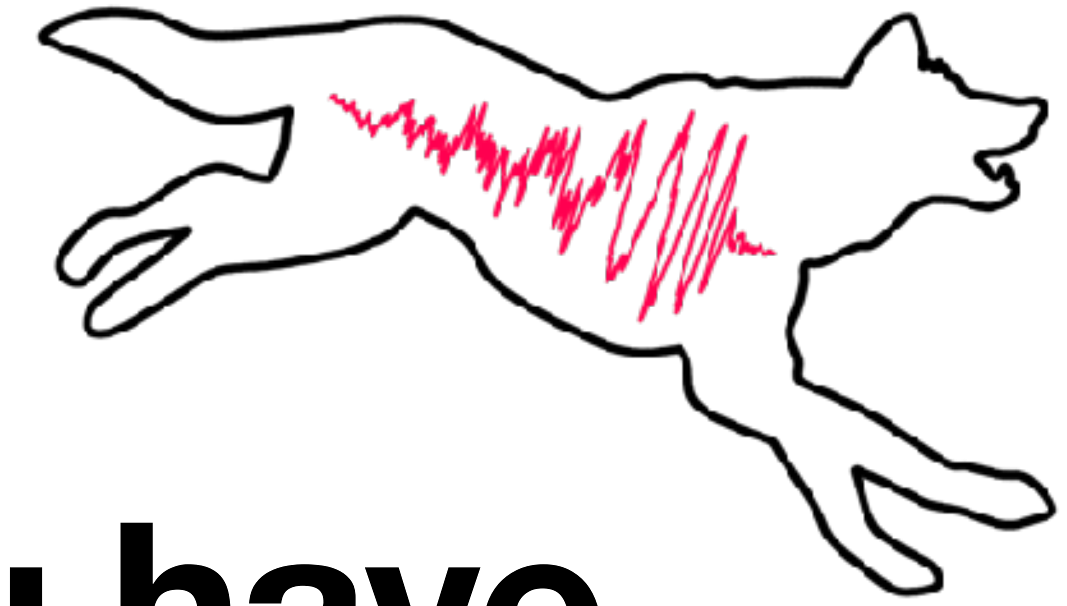

Vincent Berenz


Jonathan Gair


Jakob Macke


Bernhard Schölkopf


Alessandra Buonanno



# Do you have any questions?

MAX PLANCK INSTITUTE
FOR INTELLIGENT SYSTEMS