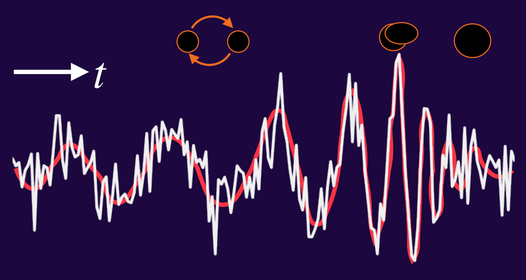
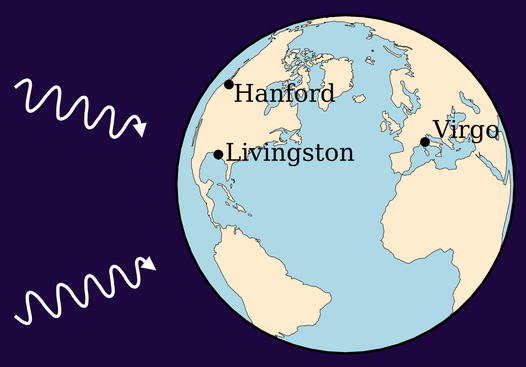
# Flexible Gravitational-Wave Parameter Estimation with Transformers

**Annalena Kofler**[1,2] Maximilian Dax[1,3,4] Stephen R. Green[5] Jonas Wildberger[3] Nihar Gupte[2,6] Jonathan Gair[2] Jakob H. Macke[1,4,7] Alessandra Buonanno[2,6] Bernhard Schölkopf[1,3]
[1]Max Planck Institute for Intelligent Systems Tübingen, [2]Albert Einstein Institute Potsdam, [3]ELLIS Institute Tübingen, [4]Tübingen AI Center, [5]University of Nottingham, [6]University of Maryland, [7]University of Tübingen

## Gravitational waves



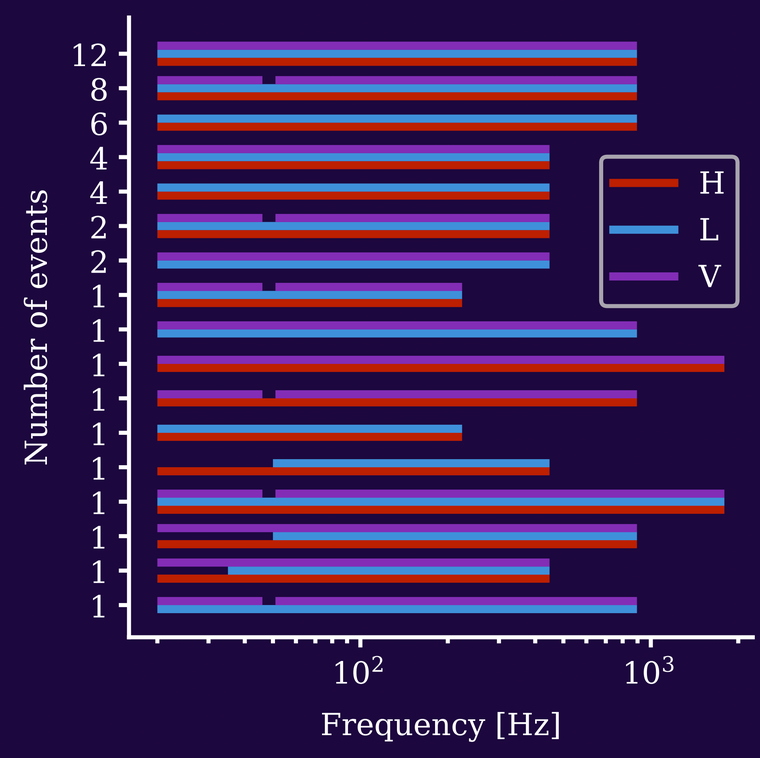

**Goal**: Analyze signals → posterior distribution of black hole mergers

**Problem**: Real data is messy

→ 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

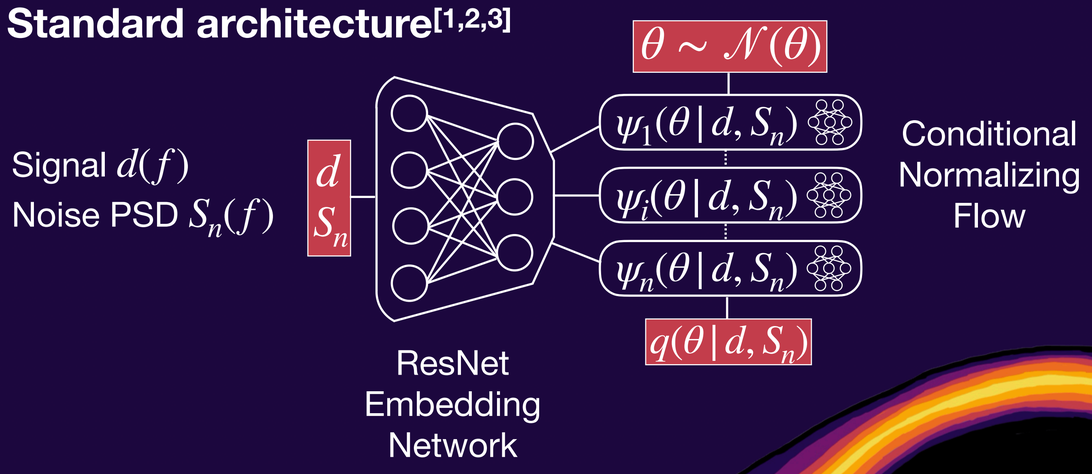



- Missing detectors
- Changes in $f_{\min}$ & $f_{\max}$
- Remove small range $[f_{\text{low}}, f_{\text{high}}]$

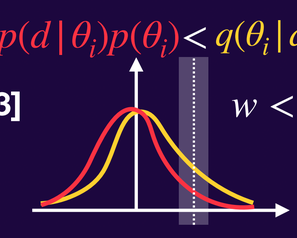
→ We would need to train 17 different DINGO models!

## DINGO (Deep INference for Gravitational wave Observations)

### Standard architecture[1,2,3]




## Validation with importance sampling[3]




Compare learned NPE density and likelihood

$$\frac{p(\theta \mid d)}{q(\theta \mid d)} \propto w_i = \frac{p(d \mid \theta_i)p(\theta_i)}{q(\theta_i \mid d)}$$

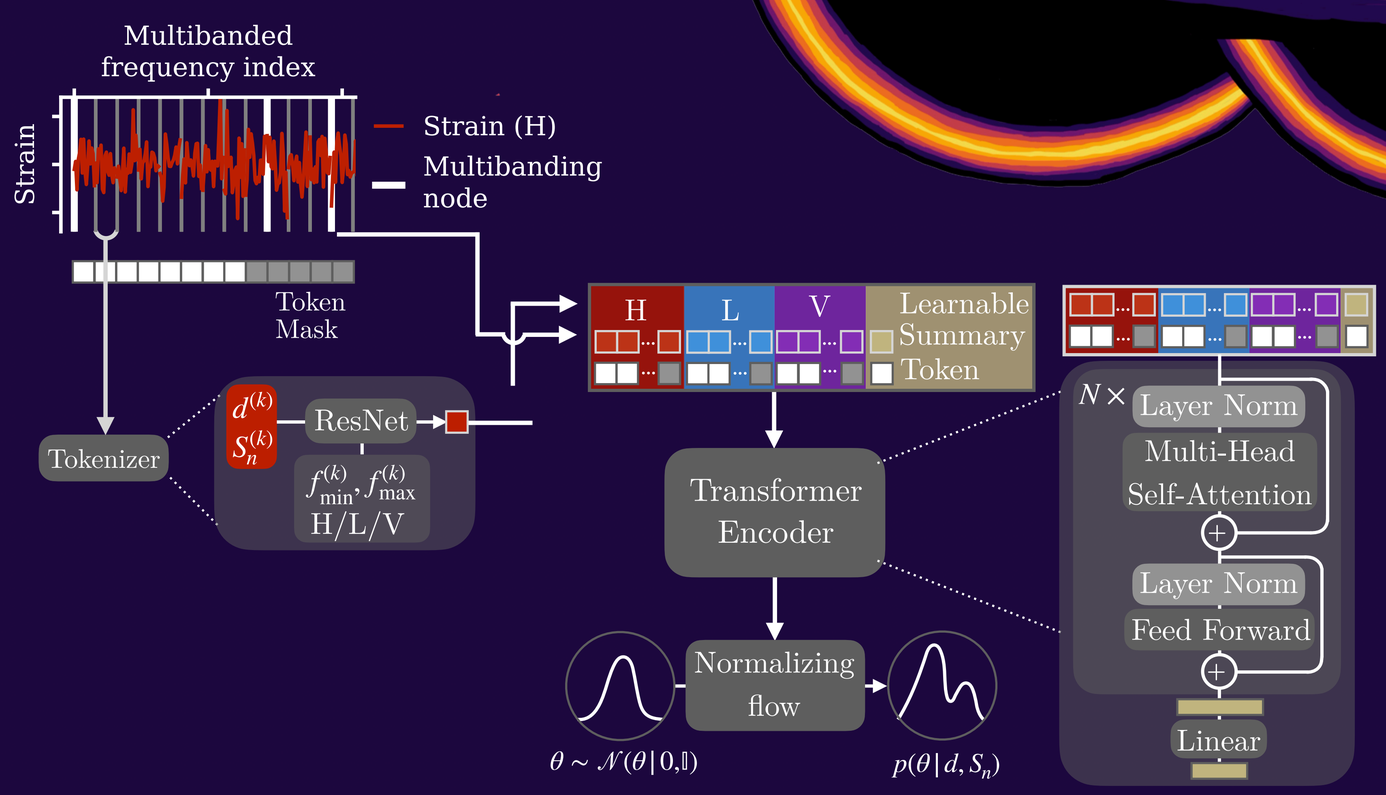
Likelihood · Prior

Proposal (NPE)

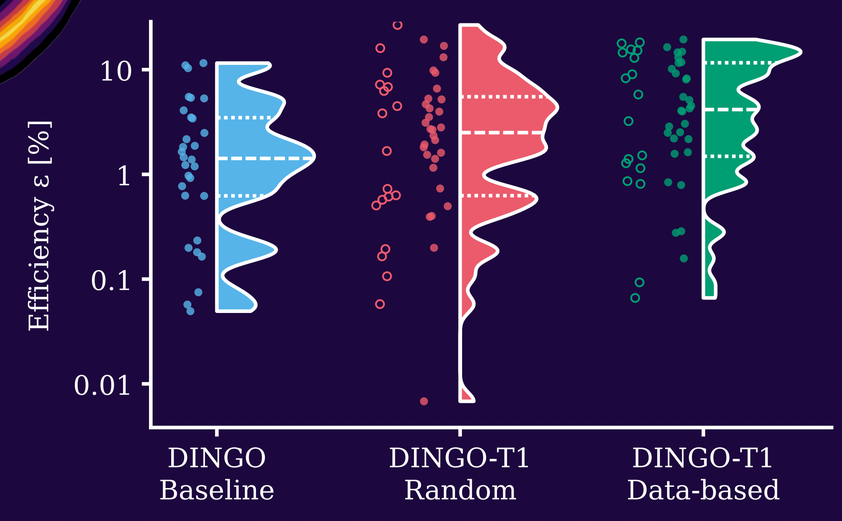
Performance criterion: Sample efficiency

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

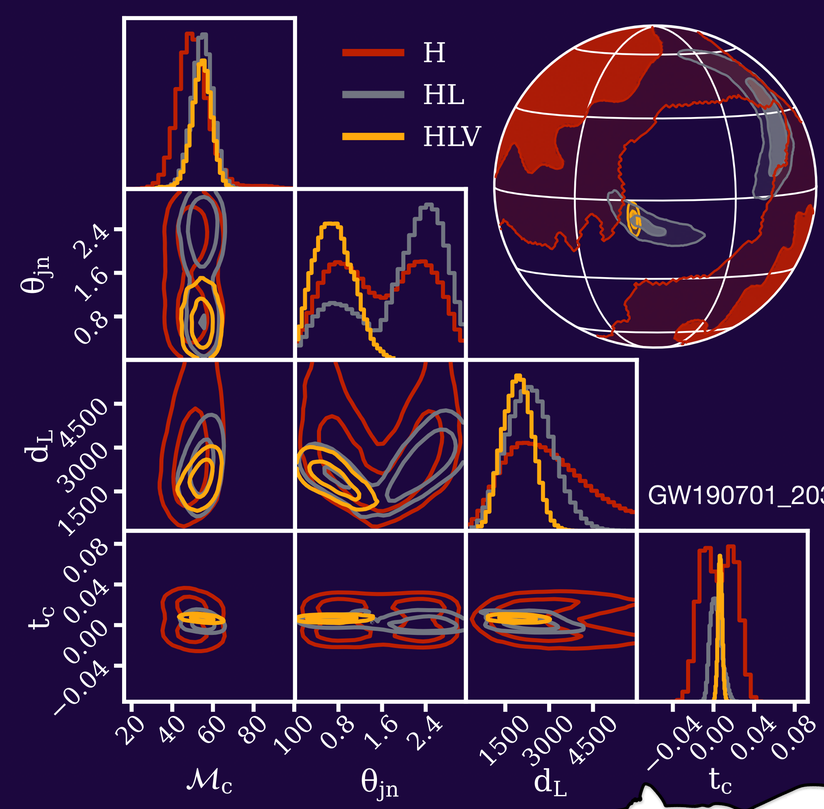
## DINGO-T1: Architecture



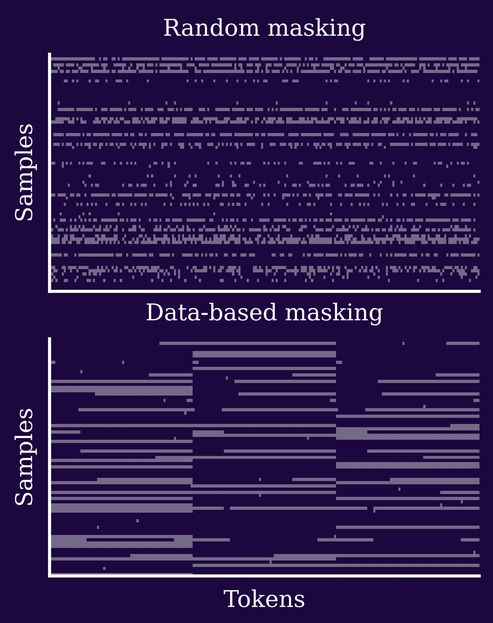

## Performance for 48 events

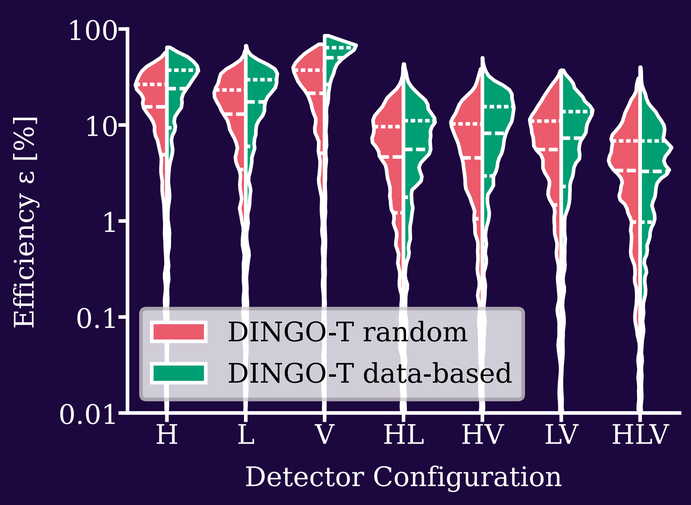



## Flexible analysis




## Masking strategies during training




## 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

MAX PLANCK INSTITUTE FOR INTELLIGENT SYSTEMS