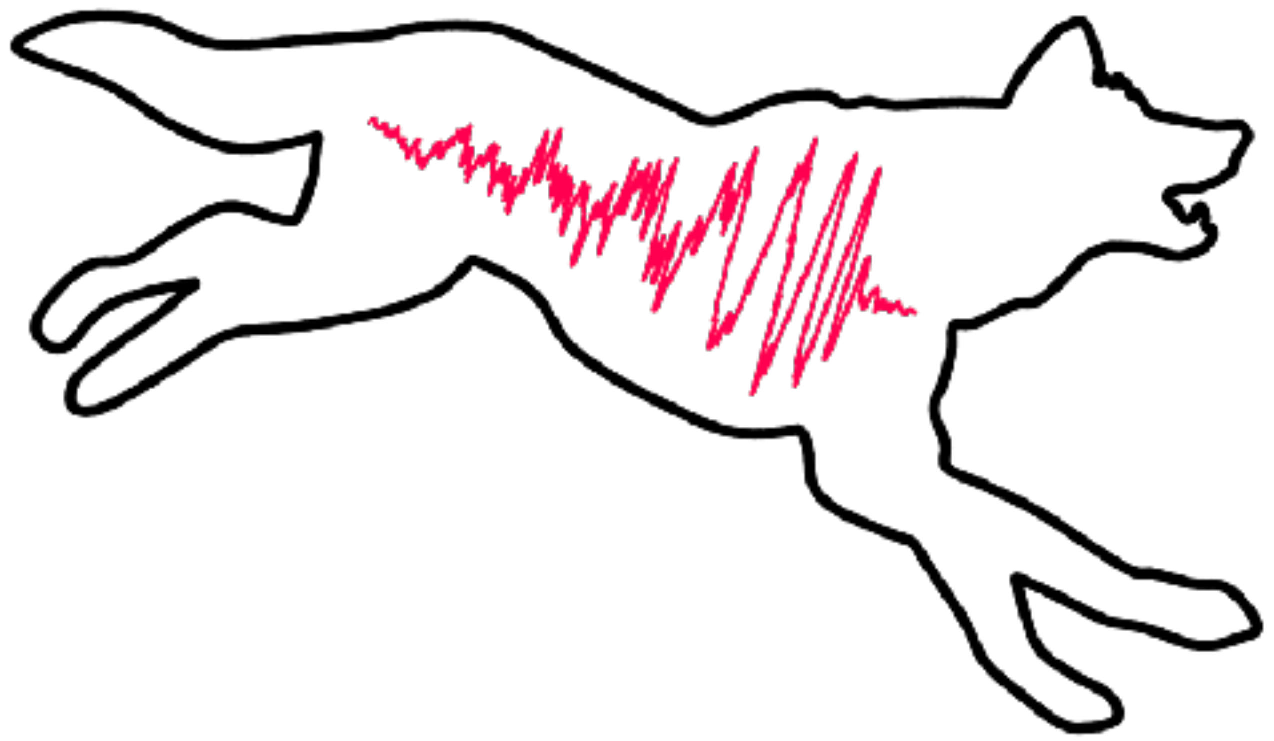
MAX PLANCK INSTITUTE
FOR INTELLIGENT SYSTEMS

Neural Posterior Estimation for Gravitational Waves

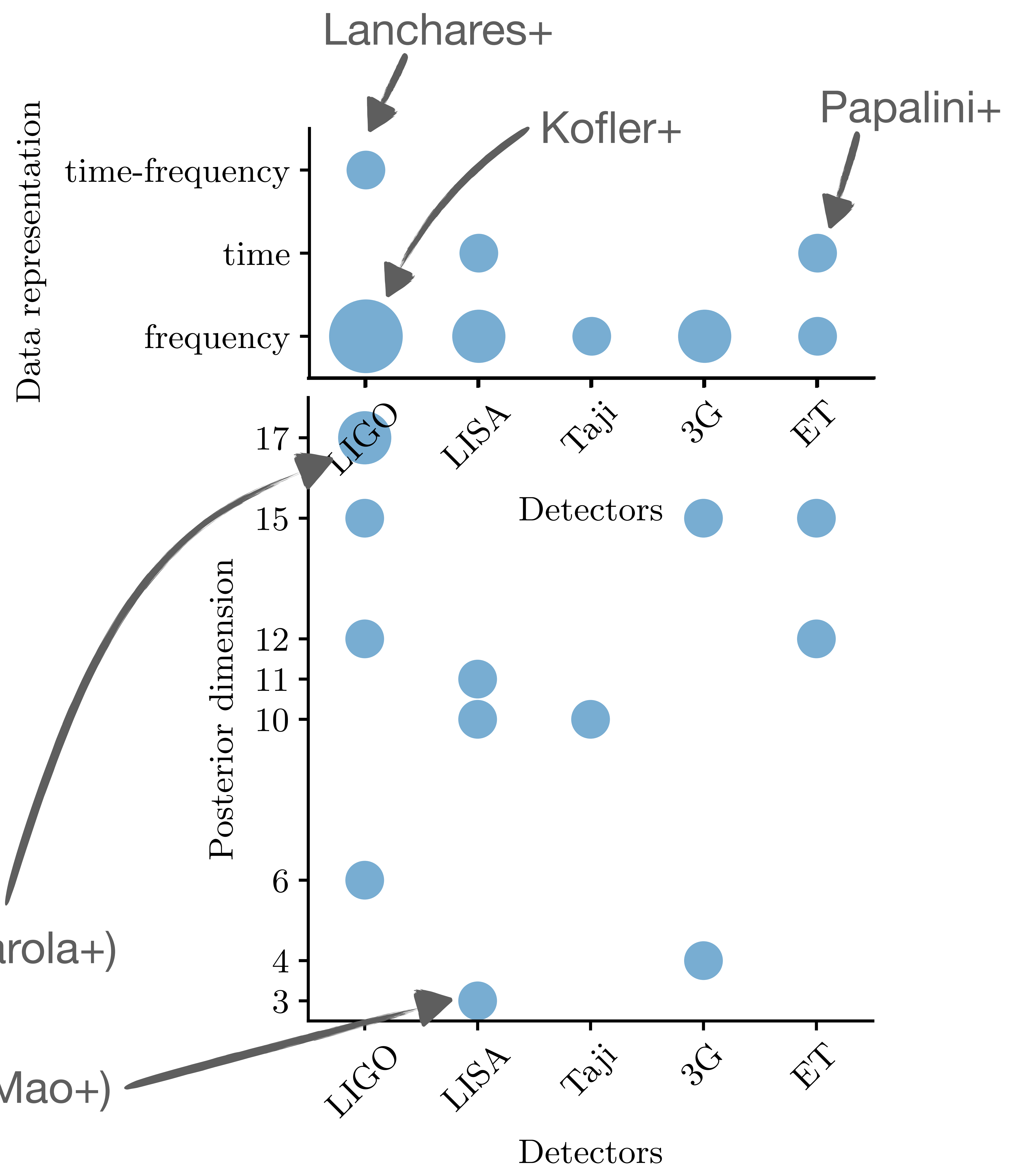


Annalena Kofler, 27.01.2026

GWfreeride workshop, Sexten

Trends in NPE for GWs

- Collected NPE papers from the last year
- Where are we going?
    - Toward future detectors
    - Beyond frequency domain data
    - Newly adopted architecture: transformers

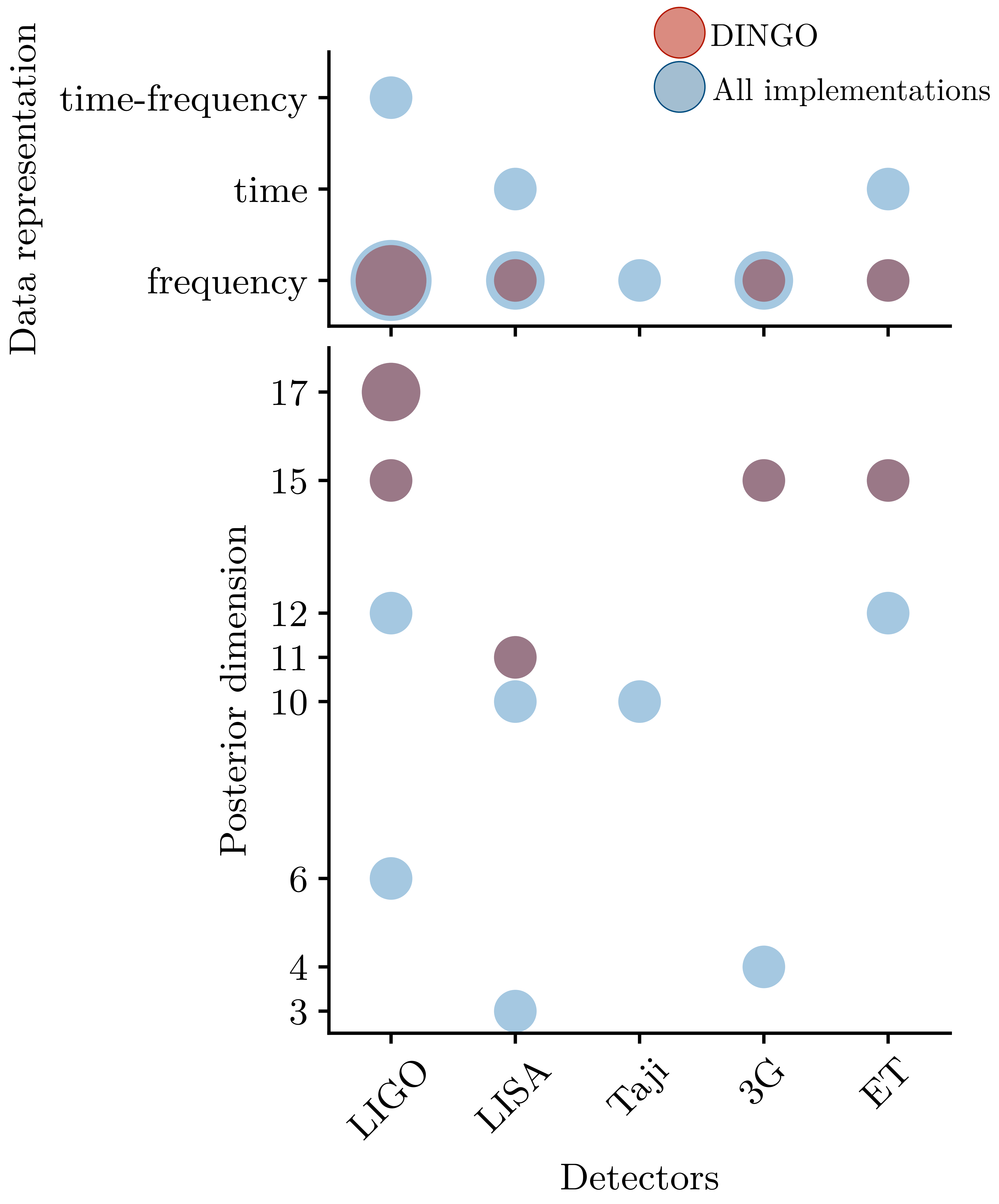


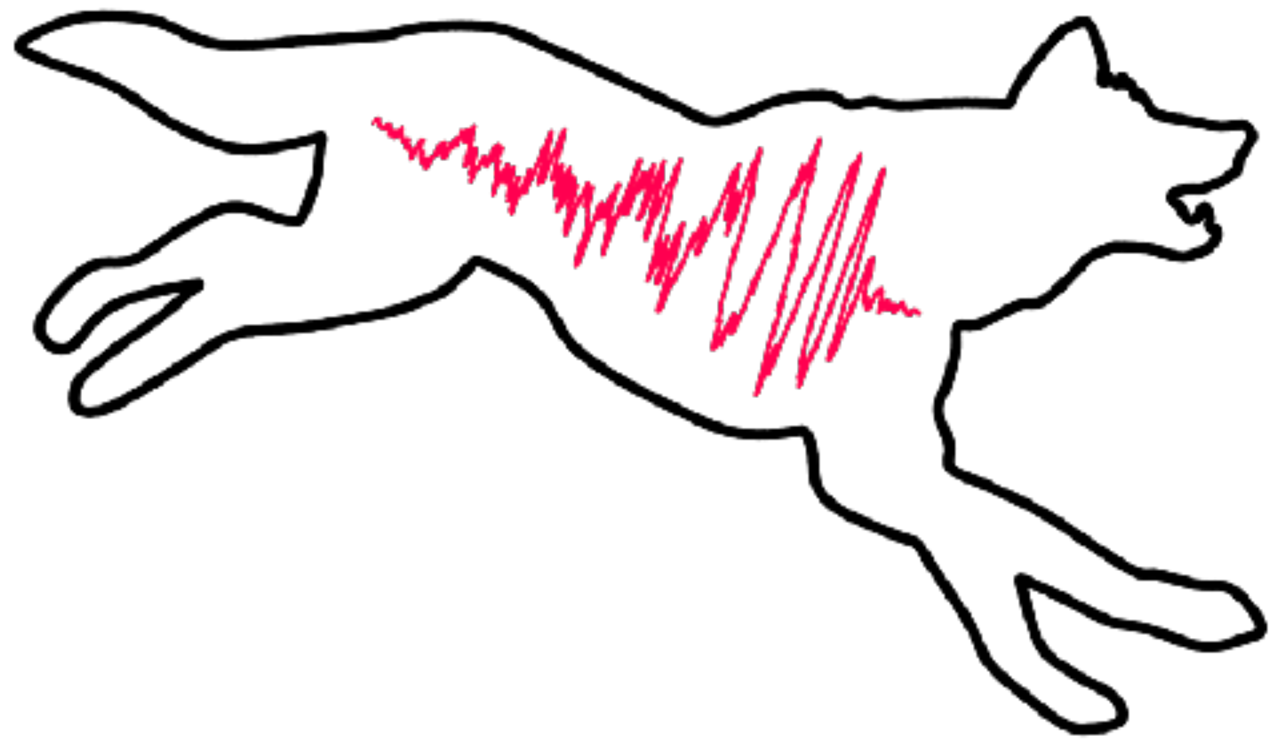
Gravitational lensing (Chan+, Caldarola+)

Analyzing 30-day signals (Mao+)

Trends in NPE for GWs

- Collected NPE papers from the last year
- Where are we going?
  - Toward future detectors
  - Beyond frequency domain data
  - Newly adopted architecture: transformers
- Several new papers building on DINGO





NPE with DINGO*

DINGO = Deep INference for Gravitational wave Observations

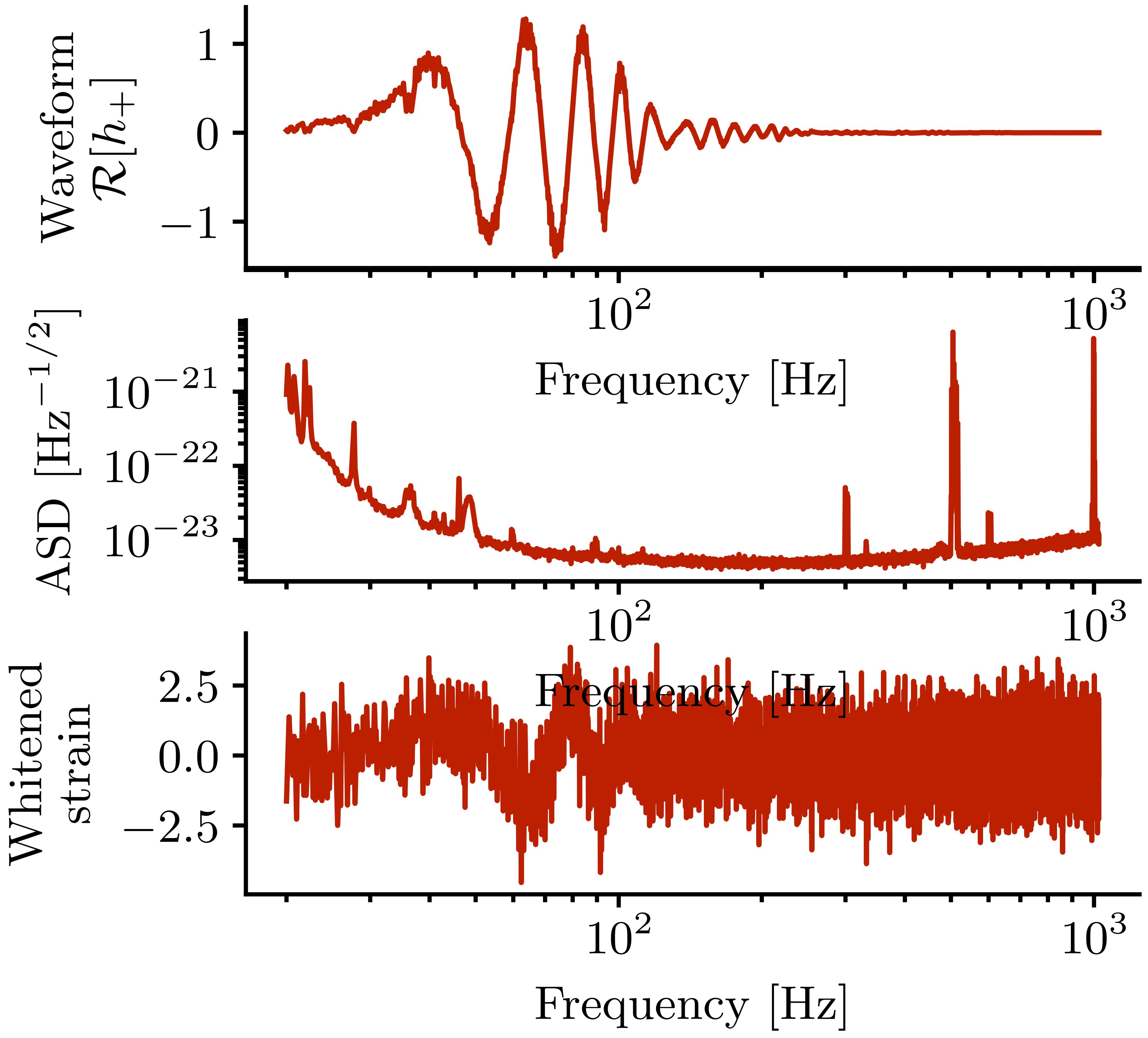
*and many other NPE-GW approaches

NPE for gravitational waves: Overview

- 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 $d^{(i)} = h(\theta) + n^{(i)}$



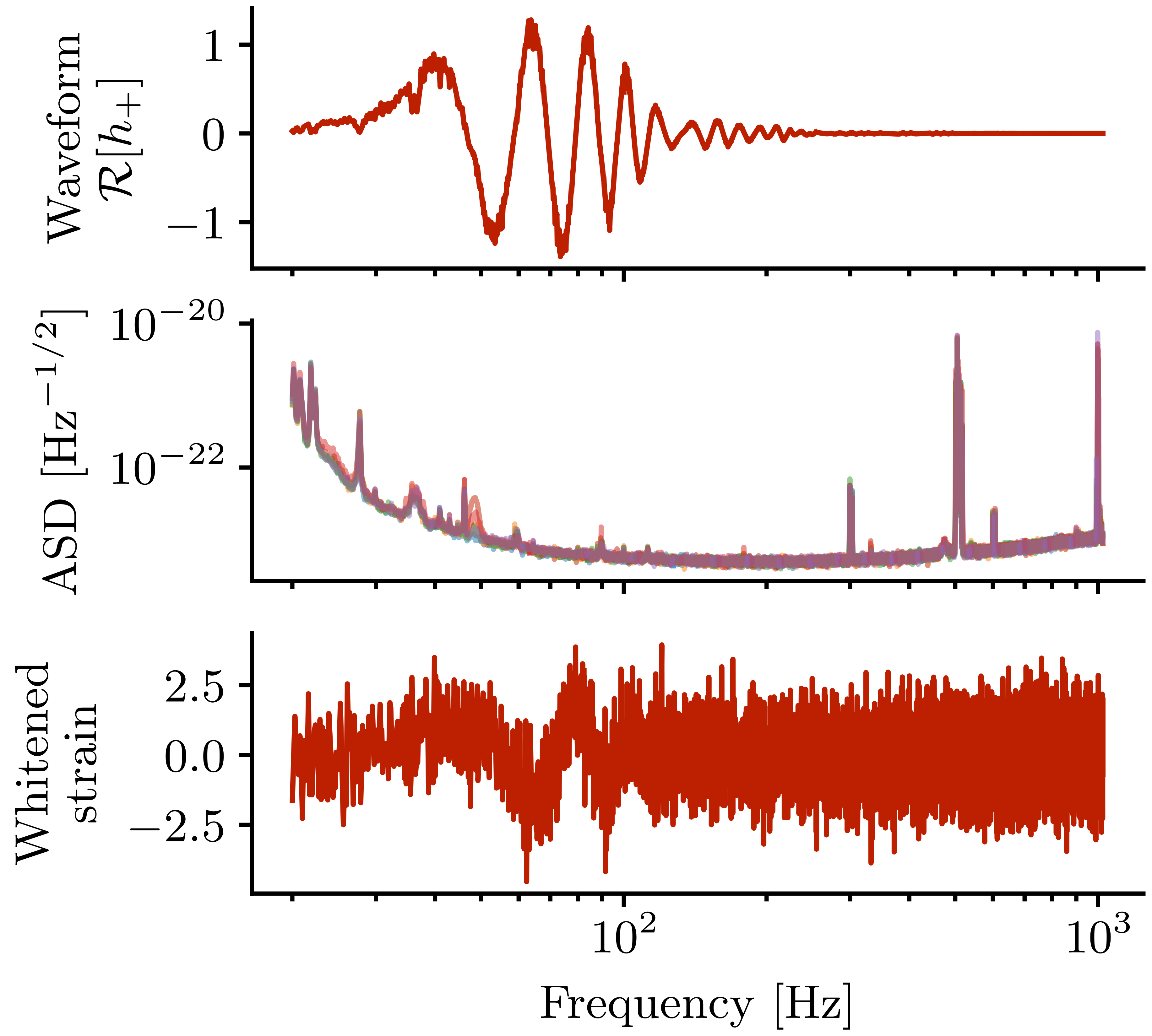
- Train density estimator

Dax+, PRL 2021, arXiv:2106.12594

Changing PSDs

- Detector noise $S_n(f)$ varies from event to event
  $\rightarrow$ Collection of PSDs $\{S_n^{(i)}(f)\}$

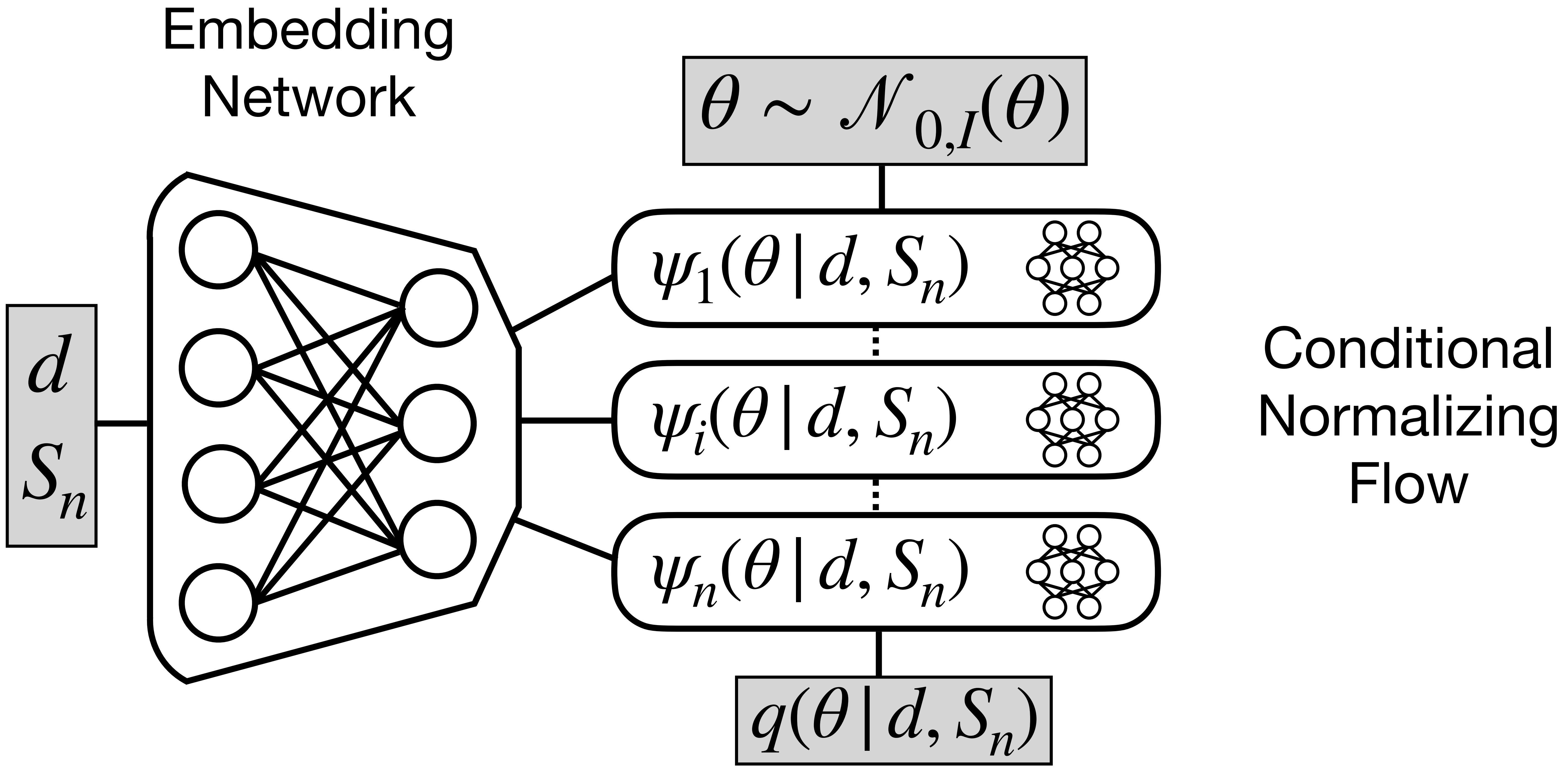
1. **Sample PSD** $S_n^{(i)} \sim p(S_n)$
2. Generate noise $n^{(i)} \sim \mathcal{N}(0,\ S_n^{(i)})$
3. Add signal $d^{(i)} = h(\theta^{(i)}) + n^{(i)}$



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 $\mathscr{L} = -\mathbb{E}_{\theta \sim p(\theta), d \sim p(d|\theta)} \left[\log q(\theta | d)\right]$

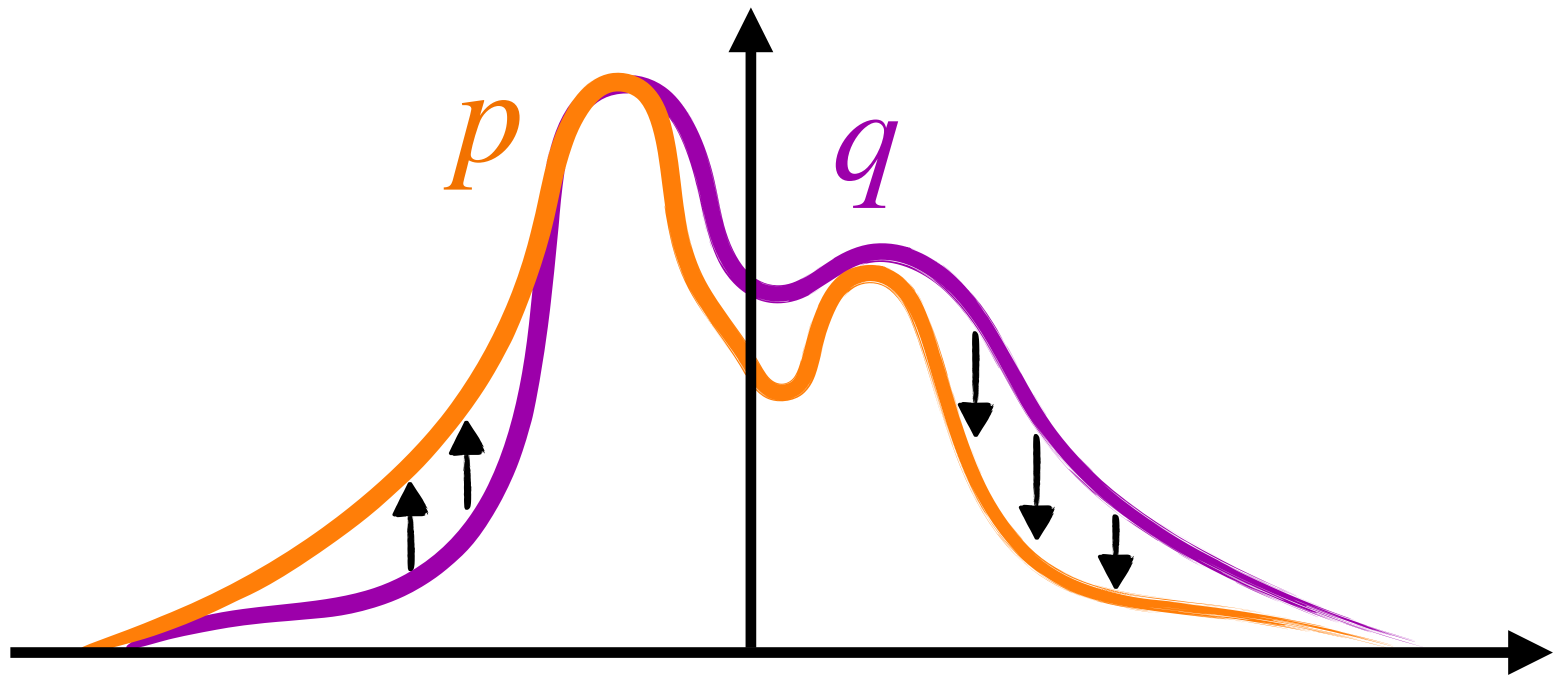


Dax+, PRL 2021, arXiv:2106.12594

But what if the NPE model is wrong?

- Importance sampling to validate and correct model

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

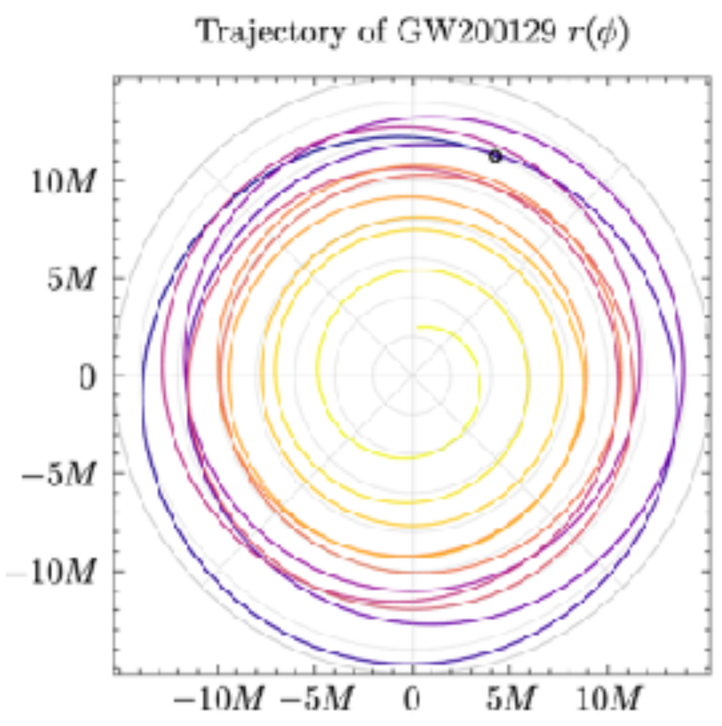


- Performance metric: sample efficiency

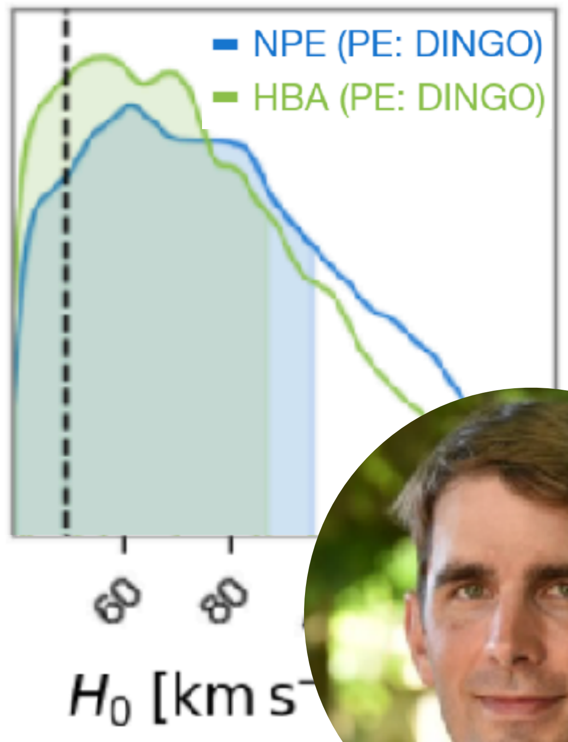
$$\epsilon = \frac{1}{N} \frac{\left(\sum_{i=1}^{N} w_i\right)^2}{\sum_{i=1}^{N} w_i^2}$$

Dax+, PRL 2023, arXiv:2210.05686

Where is DINGO used?



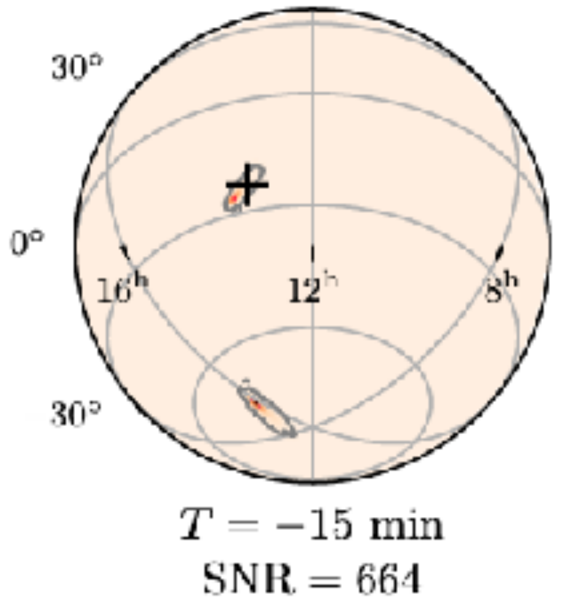
Eccentricity
Gupte+, PRD 2025

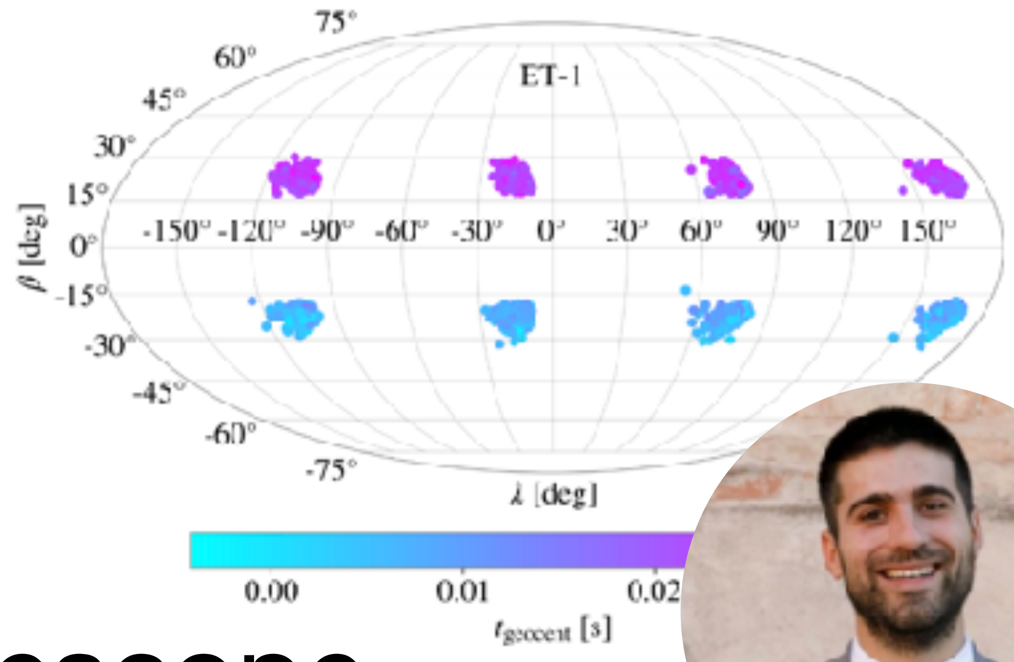


Populations
Leyde+, PRD 2024


Stephen

Binary Neutron Stars
Dax+, Nature 2025

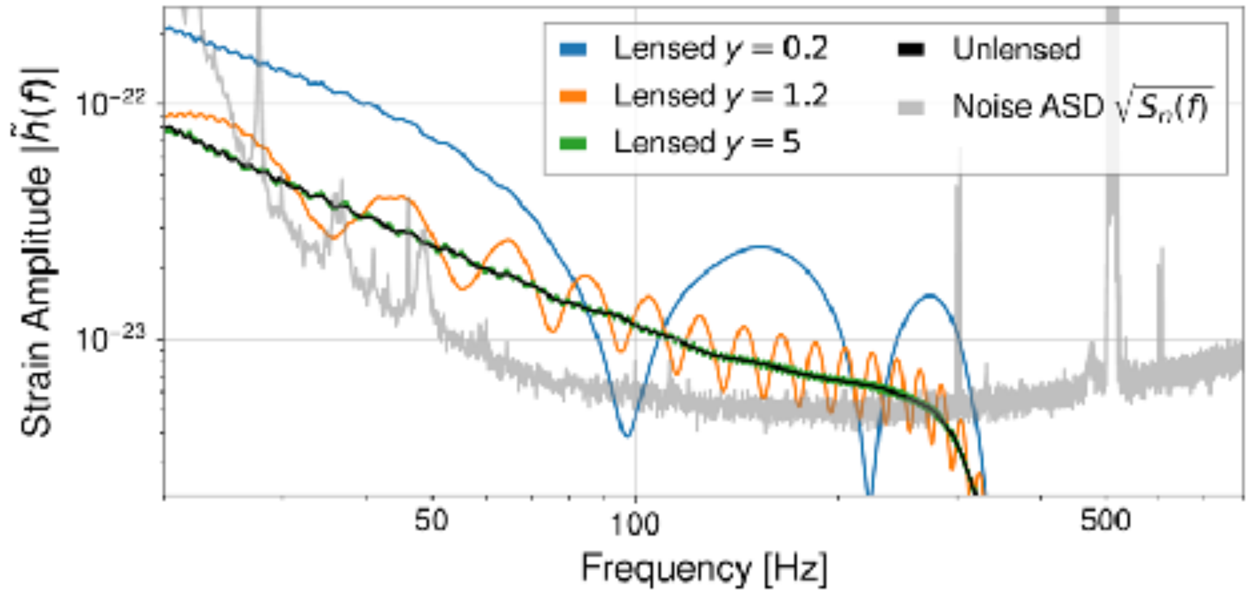




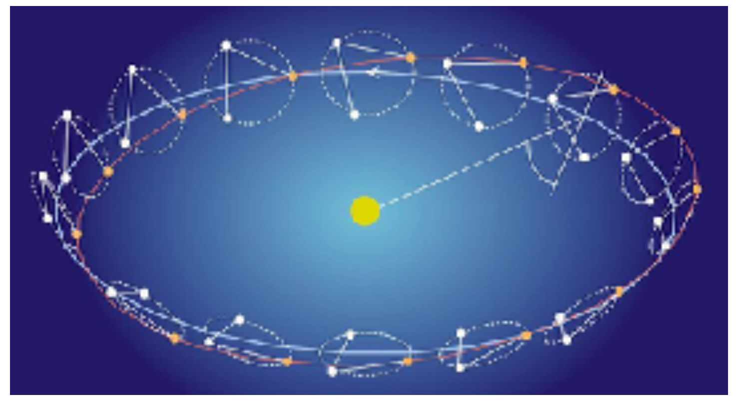
Einstein Telescope
Santoliquido+, PRD 2025


Filippo

GW Lensing
Caldarola+ arXiv 2025,
Chan+, arXiv 2025

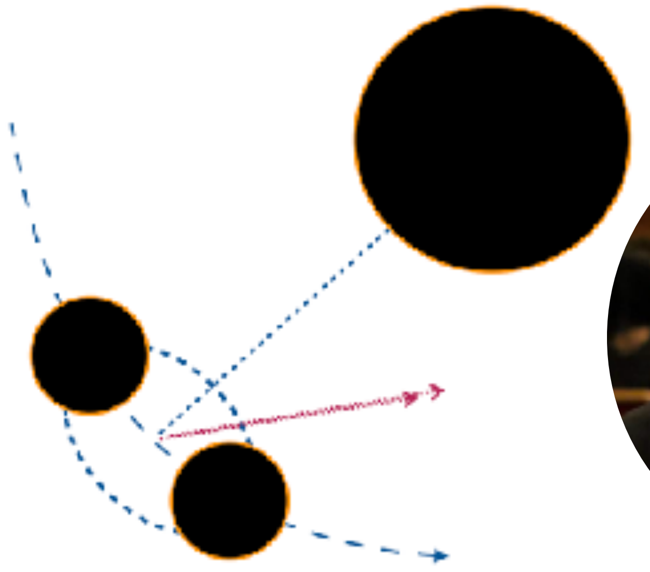


Transformers
Kofler+, arXiv 2025


jila.colorado.edu

LISA
Spadaro+, in prep.

Line-of-sight acceleration

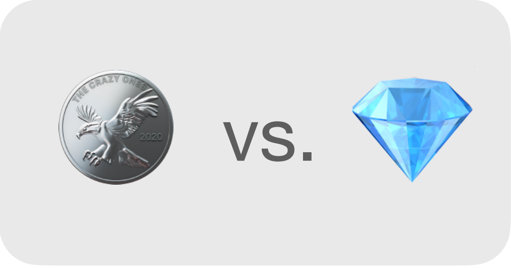



Alex

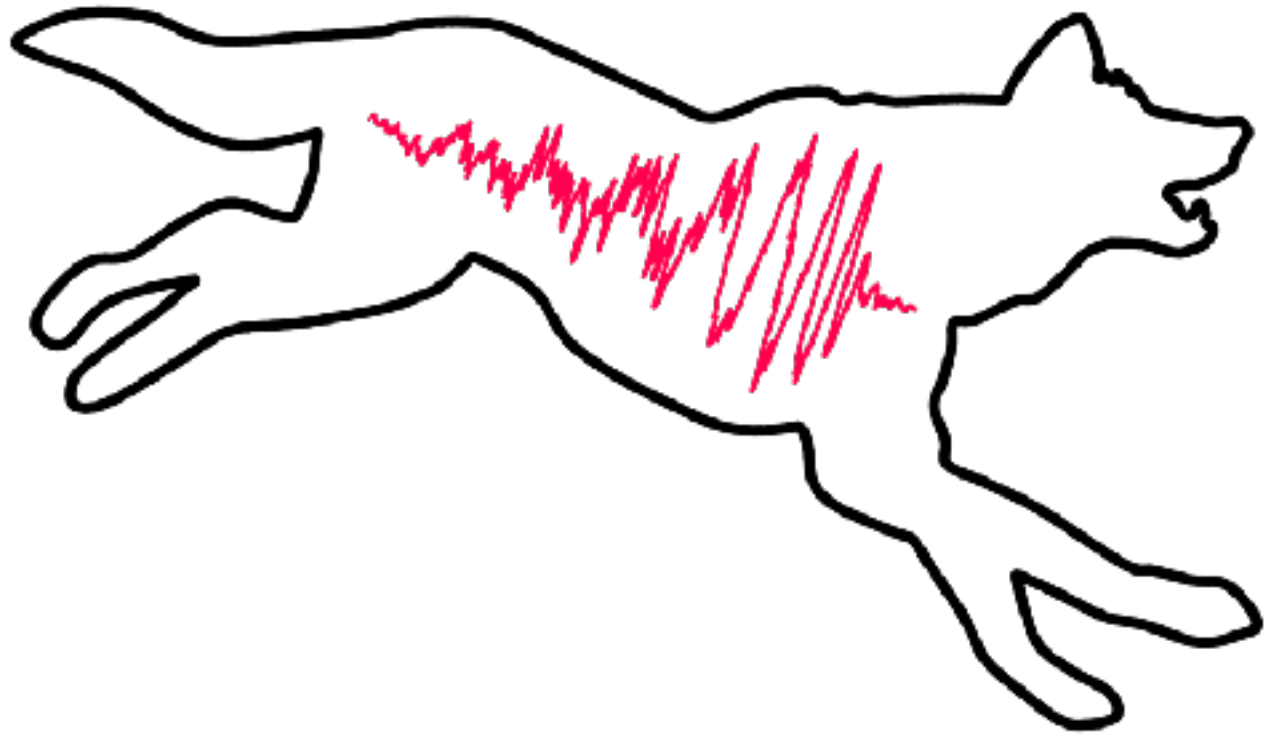
Reviewed and used
in the LVK collaboration


Cecilia

Multi-fidelity



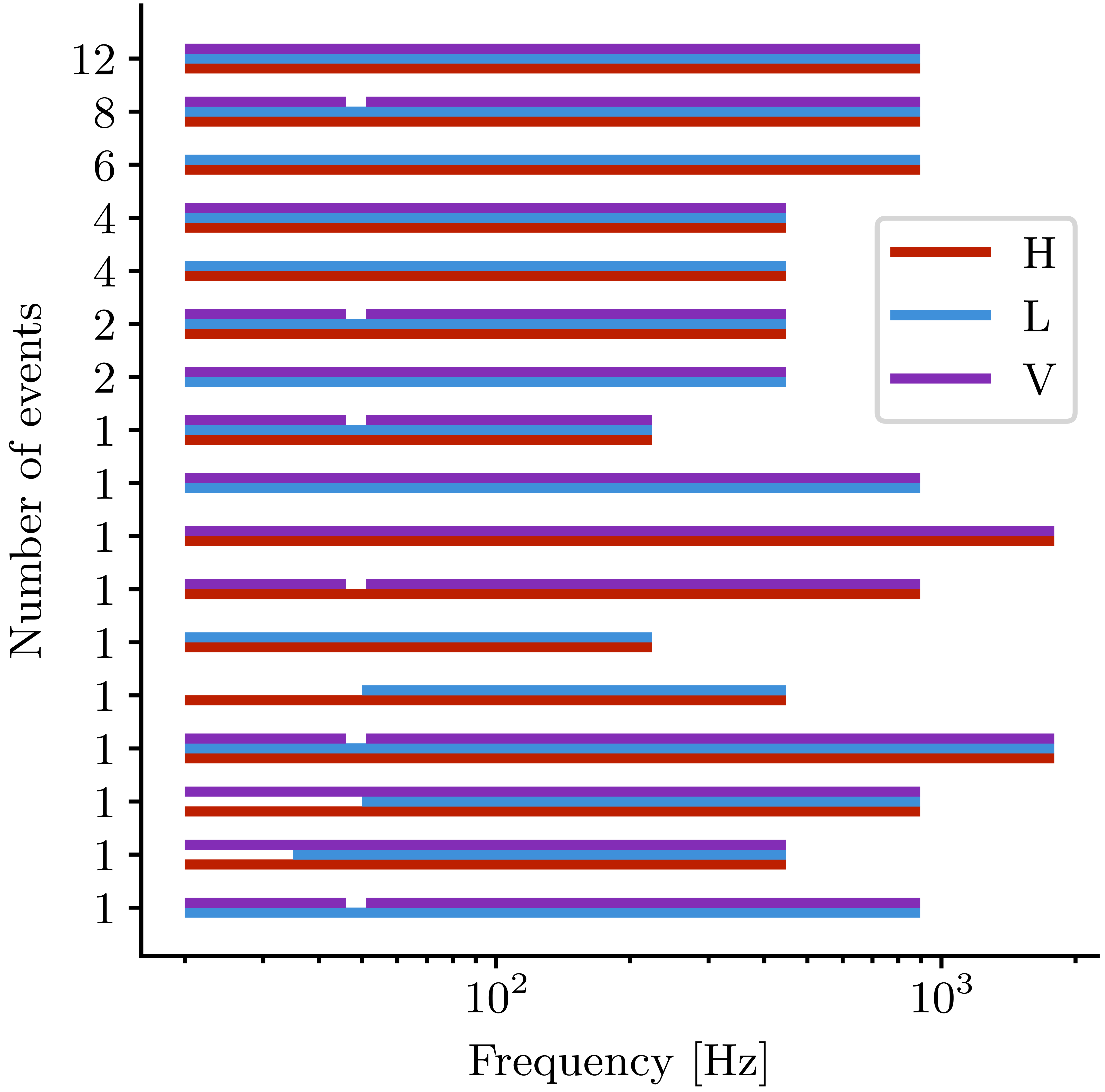
How to make NPE flexible?



Kofler+, 2025, arXiv:2512.02968

Why do we need flexible parameter estimation?

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



How do we make DINGO flexible?

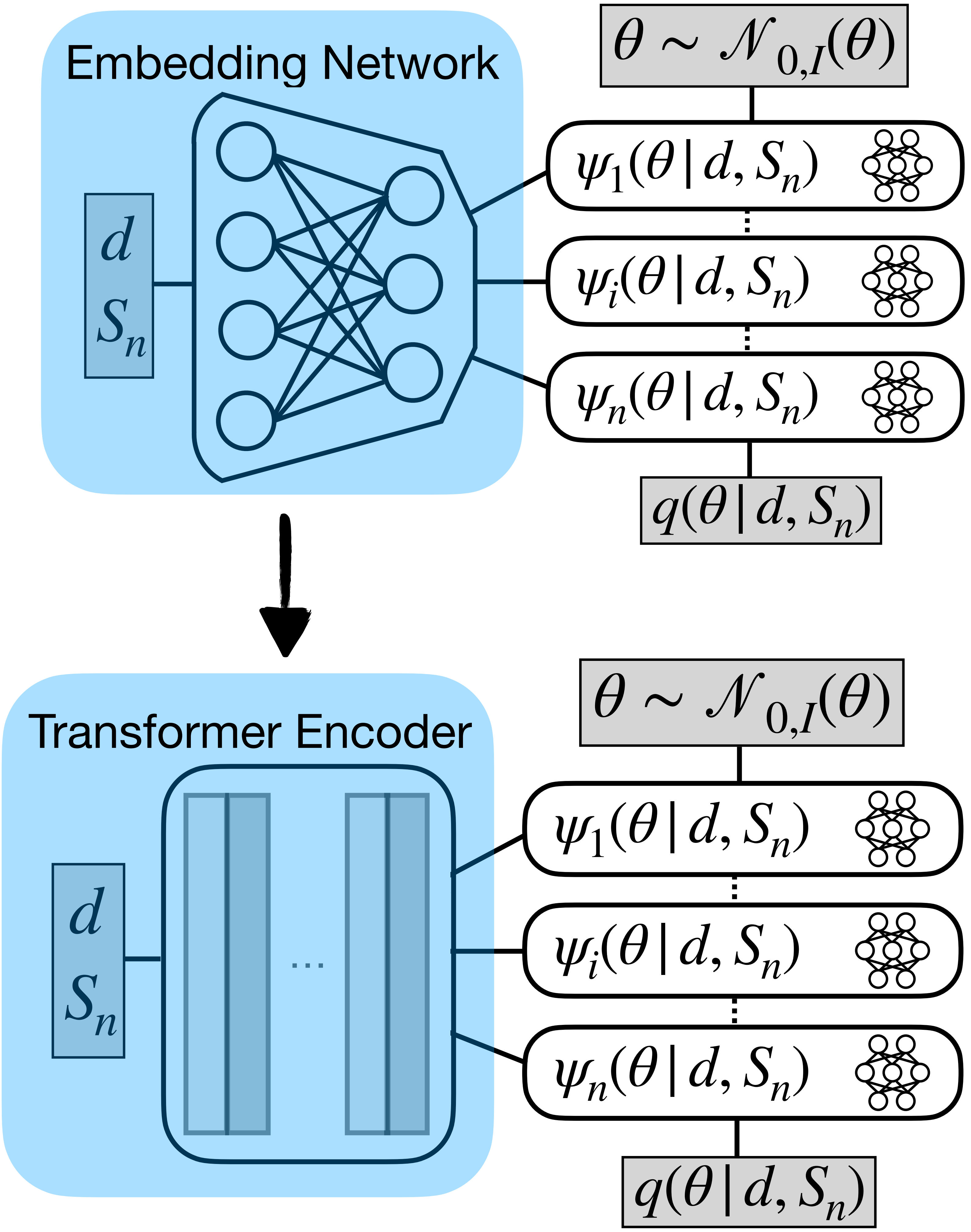
- Replace inflexible embedding network with transformer encoder[2]
- Train with signals of varying lengths

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

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

- Adjust data analysis settings at inference time

⇒ **DINGO-T1**

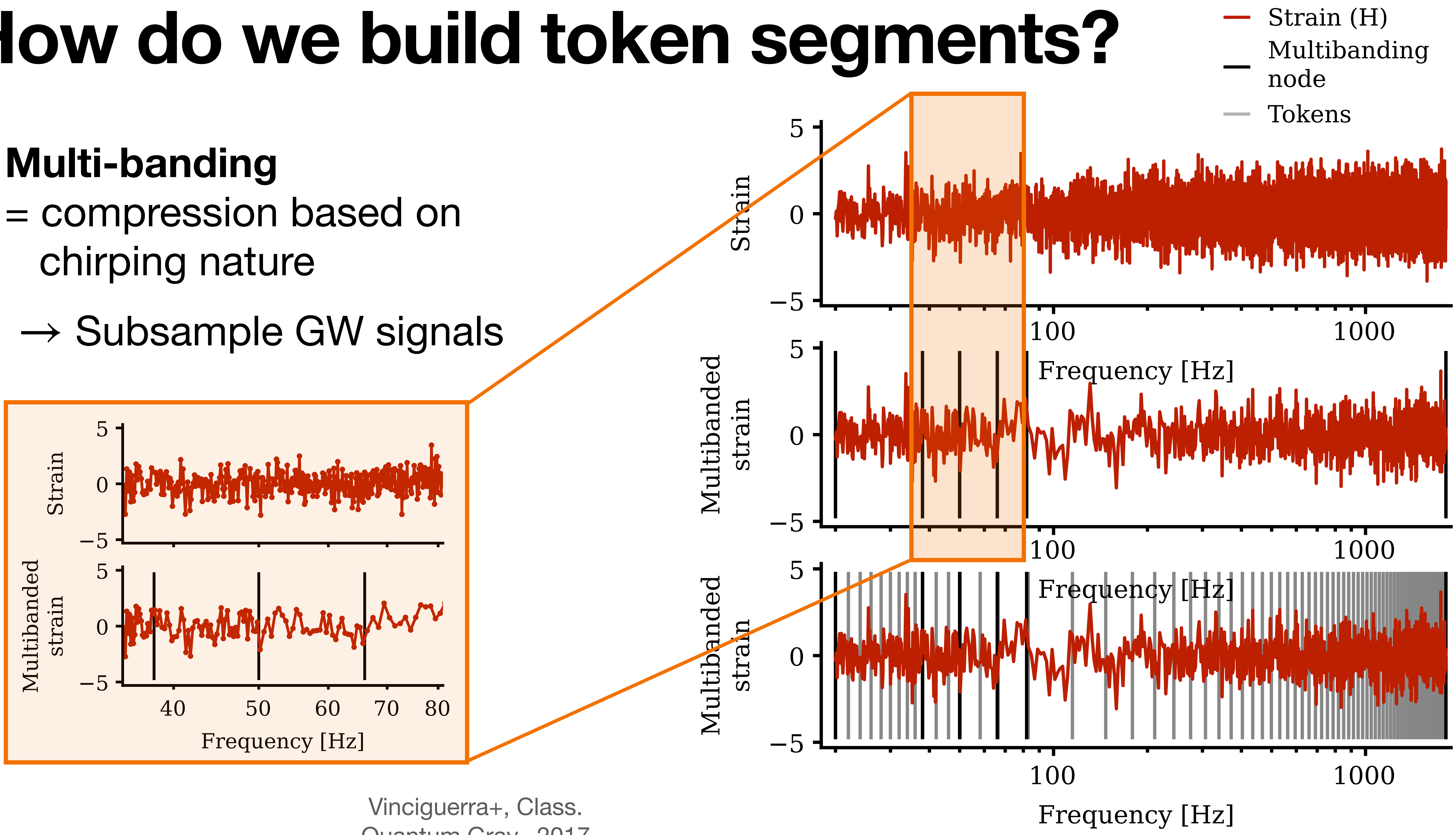


[2] Similar to Papalini+

How do we build token segments?

- **Multi-banding**
  = compression based on chirping nature

  → Subsample GW signals



Vinciguerra+, Class. Quantum Grav., 2017

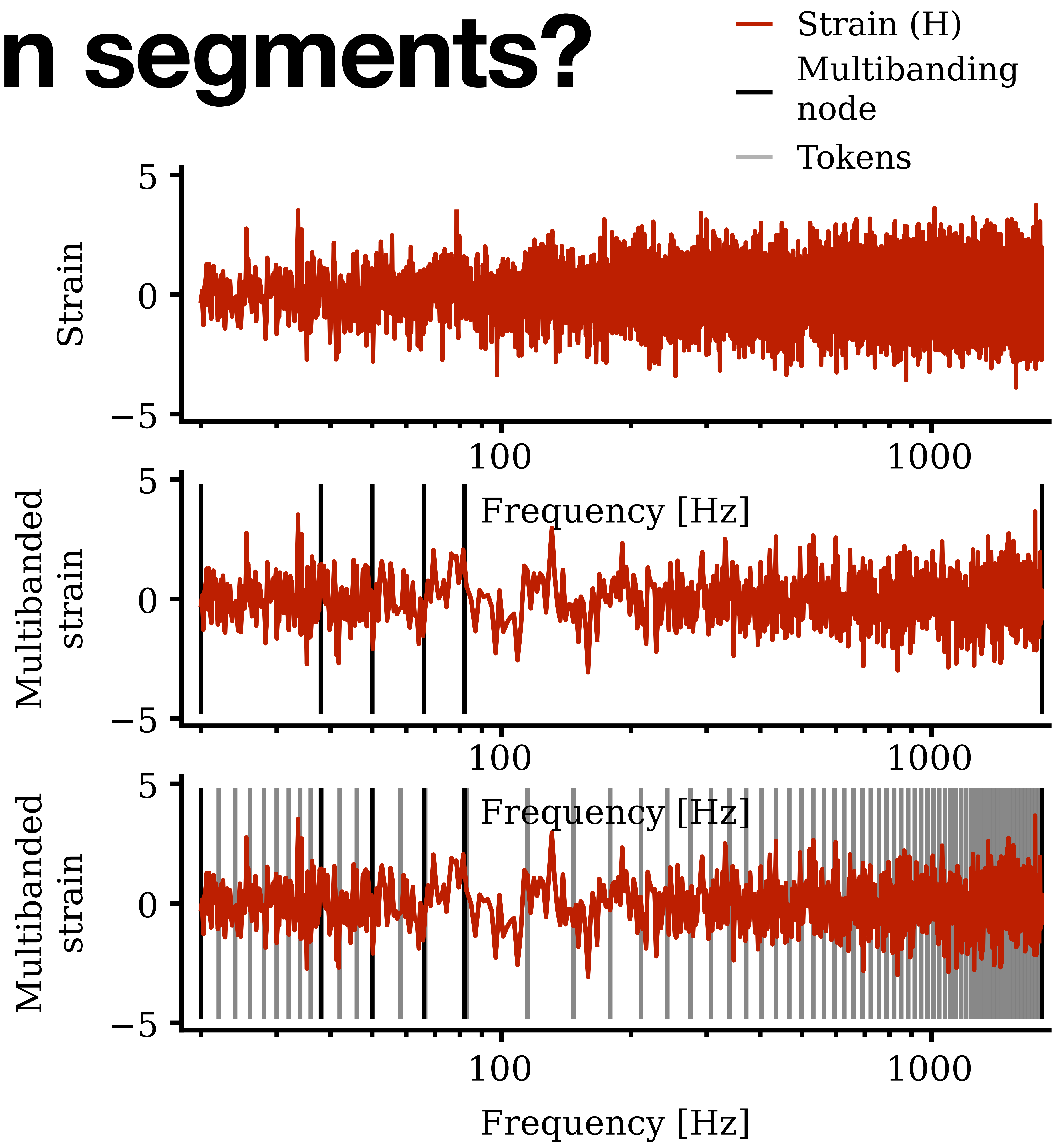
How do we build token segments?

- **Multi-banding**
  = compression based on
  chirping nature

  → Subsample GW signals

- Divide into equal segments
  with consistent resolution

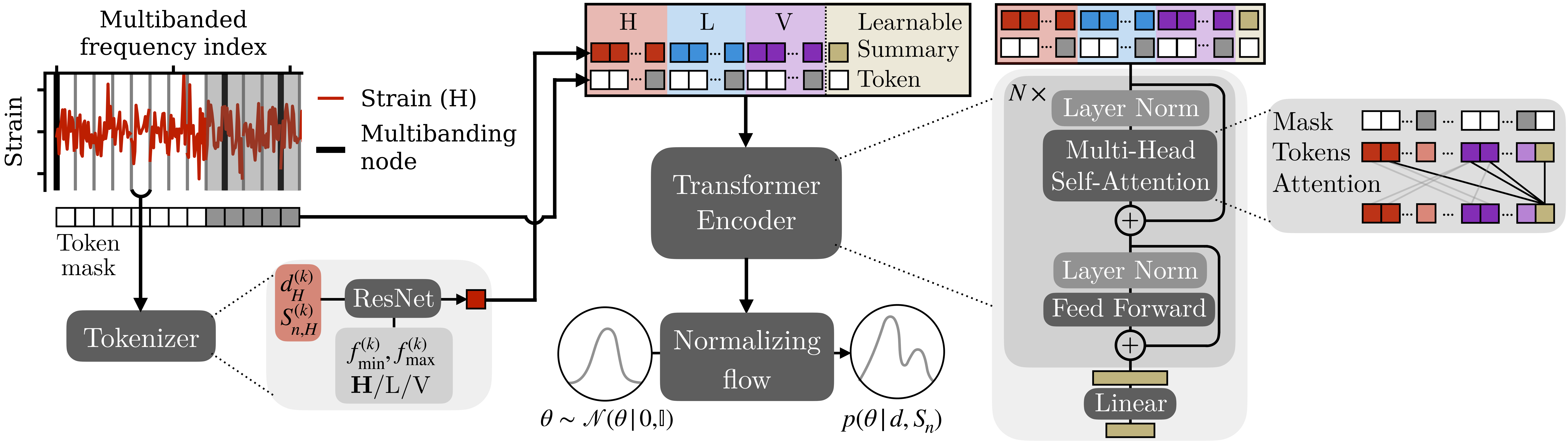
  → token segments



Vinciguerra+, Class. Quantum Grav., 2017

DINGO-T1 Architecture

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



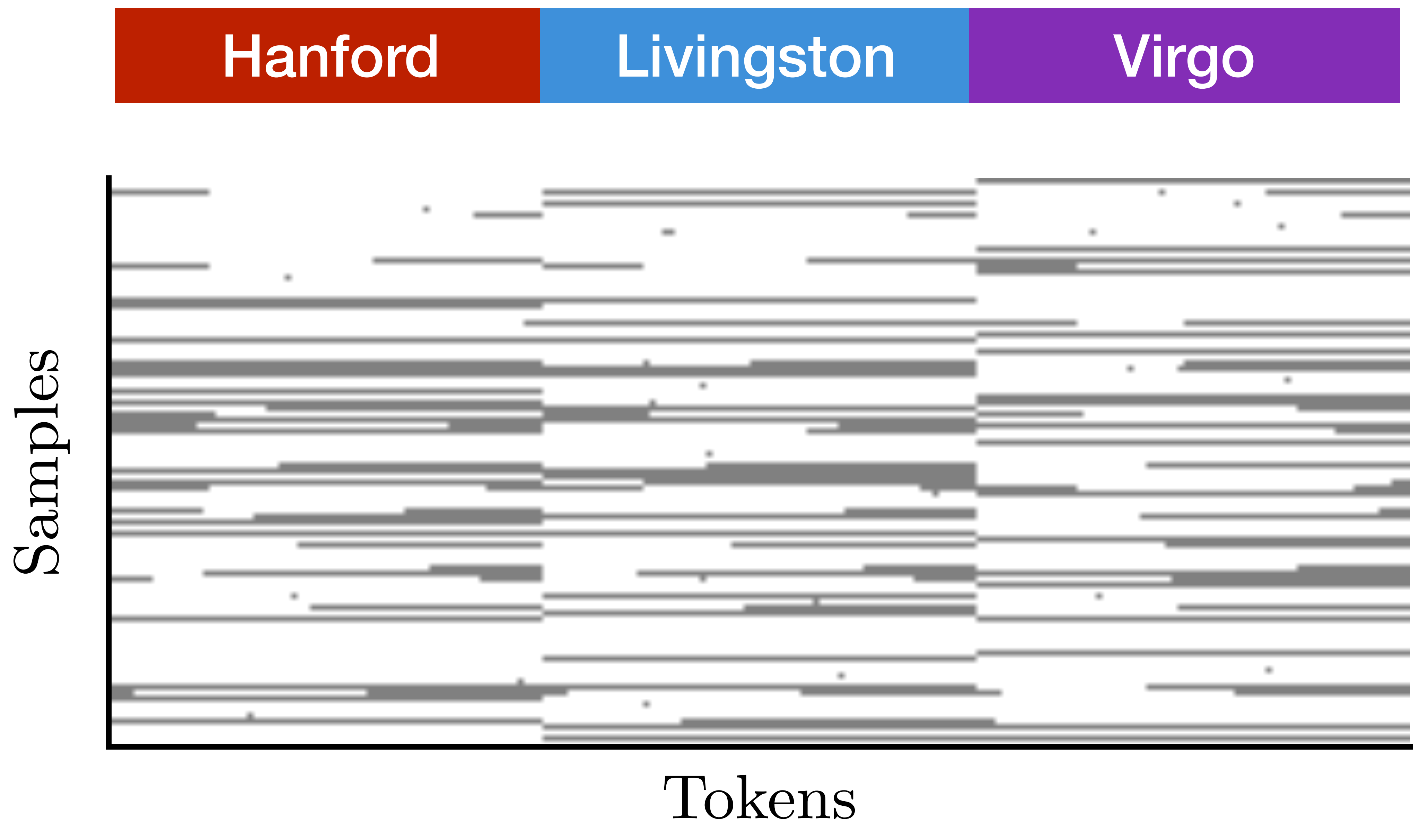
Training a flexible model

- Loss: $$\mathscr{L} = \mathbb{E}_{p(\theta)p(S_n)p(d|\theta,S_\mathrm{n})p(m)}\left[-\log q\left(\theta\,|\,m(d), m(S_n)\right)\right]$$

- Data-inspired masking strategy $m \sim p(m)$:

- On 8 A100 GPUs for $\sim$ 9 days

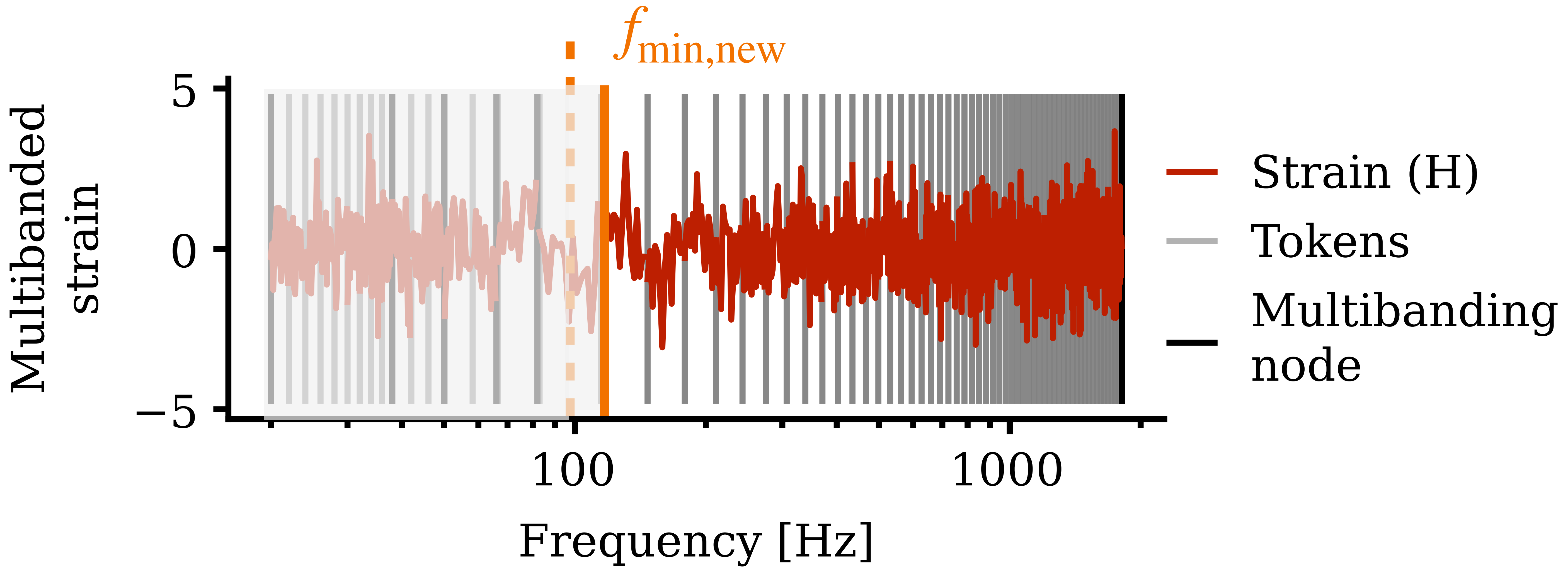
- Baseline: NPE with ResNet embedding network



Data analysis settings during inference

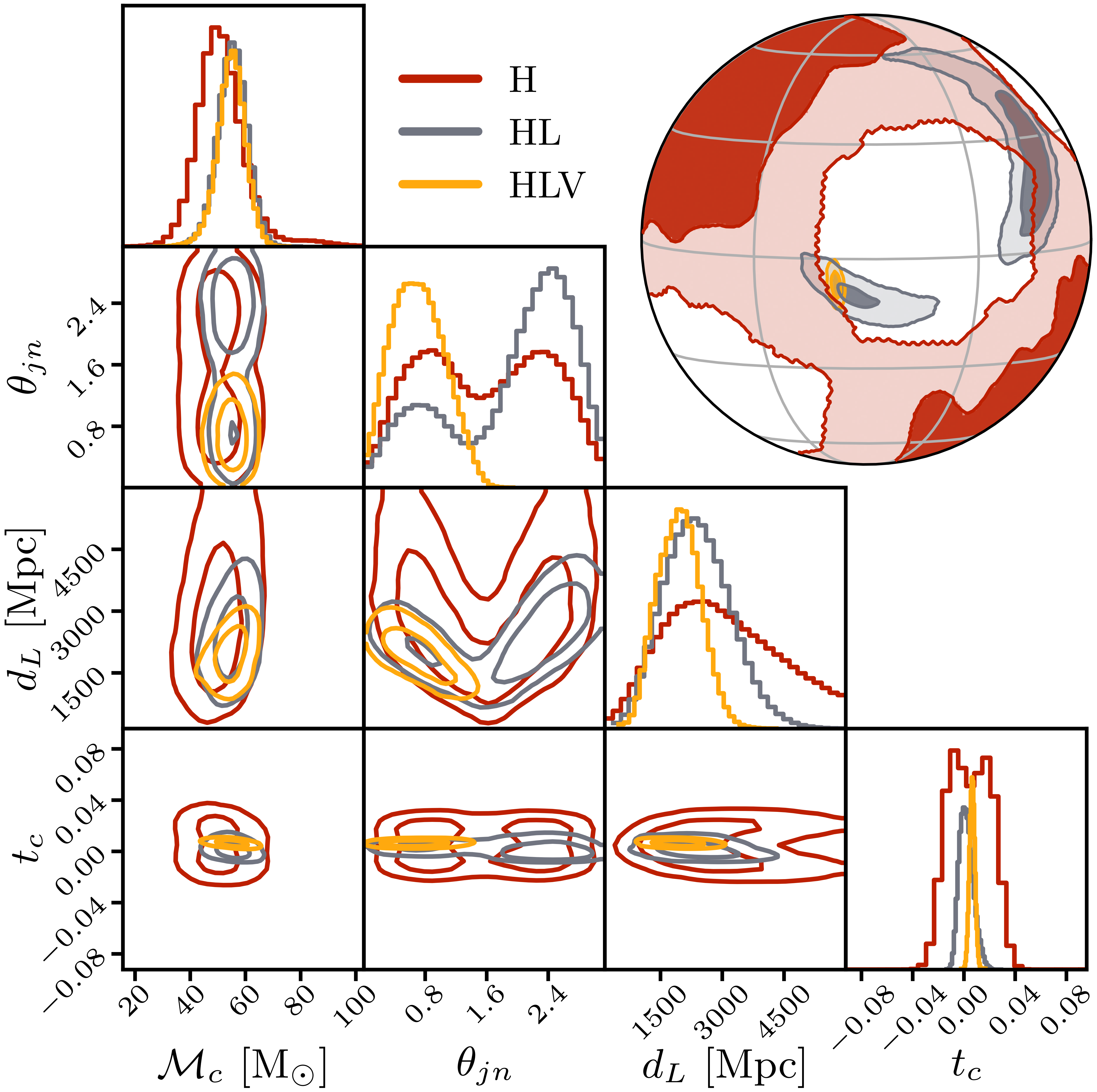
- Remove token inclusively
- Perform IS without multi-banding

→ Result independent of multi-banding and masking



What can we do with DINGO-T1?

- Same event, different detectors
- Example: GW190701_203306



What can we do with DINGO-T1?

- Fast inference
  - Initial sampling < 10 s
  - Importance sampling: 5 - 10 min
- Replace 264 individual PE runs
  → Computationally infeasible with standard samplers

Event	Detectors	HLV	HL	HV	LV	H	L	V
GW190408_181802	HLV	8.07 %	8.28 %	0.64 %	1.35 %	2.79 %	6.71 %	62.23 %
GW190413_052954	HLV	11.77 %	9.52 %	2.41 %	0.74 %	9.5 %	18.61 %	37.48 %
GW190413_134308$^{*\dagger L}$	HLV	4.52 %	7.84 %	8.29 %	1.36 %	21.16 %	11.78 %	32.2 %
GW190426_190642	HLV	0.79 %	2.27 %	7.41 %	0.0 %	29.53 %	0.04 %	49.52 %
GW190503_185404^{*L}	HLV	1.57 %	1.31 %	1.0 %	0.41 %	17.09 %	2.14 %	60.85 %
GW190513_205428^{*L}	HLV	0.16 %	0.52 %	4.19 %	2.04 %	11.74 %	4.17 %	22.9 %
GW190517_055101	HLV	2.17 %	4.28 %	1.79 %	0.27 %	1.36 %	0.36 %	13.06 %
GW190519_153544	HLV	4.26 %	0.85 %	3.71 %	4.88 %	3.38 %	9.09 %	59.66 %
GW190602_175927	HLV	13.09 %	13.18 %	28.09 %	5.4 %	27.24 %	15.95 %	26.58 %
GW190701_203306^{*L}	HLV	14.83 %	14.05 %	9.13 %	5.45 %	31.43 %	5.93 %	38.78 %
GW190706_222641	HLV	2.49 %	5.65 %	13.73 %	6.21 %	0.1 %	10.14 %	27.66 %
GW190727_060333$^{\cdot L}$	HLV	0.84 %	17.6 %	2.95 %	3.29 %	5.09 %	0.45 %	23.71 %
GW190803_022701	HLV	16.39 %	28.23 %	19.48 %	9.04 %	28.74 %	7.89 %	42.92 %
GW190828_063405	HLV	3.99 %	17.35 %	8.7 %	7.11 %	24.58 %	9.26 %	64.41 %
GW190915_235702	HLV	8.23 %	13.56 %	0.89 %	5.24 %	1.9 %	14.02 %	42.52 %
GW190916_200658	HLV	19.41 %	20.19 %	5.16 %	18.13 %	11.2 %	27.25 %	42.29 %
GW190926_050336	HLV	1.63 %	4.25 %	11.17 %	12.38 %	16.39 %	19.97 %	53.75 %
GW190929_012149	HLV	3.04 %	4.74 %	10.61 %	0.0 %	24.68 %	0.01 %	50.24 %
GW191127_050227$^{\diamond H}$	HLV	0.28 %	0.97 %	8.21 %	5.15 %	16.13 %	28.83 %	6.04 %
GW191215_223052	HLV	4.05 %	5.79 %	5.13 %	1.25 %	6.55 %	2.02 %	56.09 %
GW191230_180458	HLV	14.61 %	8.4 %	2.68 %	16.51 %	7.79 %	16.88 %	58.75 %
GW200129_065458$^{\ddagger L}$	HLV	0.29 %	0.29 %	2.16 %	0.01 %	14.15 %	0.07 %	2.66 %
GW200208_130117	HLV	11.58 %	12.15 %	5.17 %	5.99 %	32.27 %	8.9 %	17.75 %
GW200208_222617	HLV	2.86 %	2.47 %	4.73 %	0.05 %	5.55 %	18.82 %	18.48 %
GW200209_085452	HLV	2.52 %	11.06 %	3.65 %	16.62 %	9.66 %	34.08 %	41.28 %
GW200216_220804	HLV	2.2 %	12.56 %	12.41 %	7.44 %	21.27 %	10.96 %	52.1 %
GW200219_094415	HLV	5.12 %	7.36 %	3.27 %	2.77 %	11.76 %	11.09 %	1.9 %
GW200220_061928	HLV	10.17 %	18.1 %	6.07 %	7.68 %	8.67 %	11.95 %	61.93 %
GW200224_222234	HLV	5.49 %	7.18 %	1.65 %	6.85 %	10.2 %	20.89 %	4.01 %
GW200311_115853	HLV	9.2 %	5.56 %	11.57 %	4.81 %	7.71 %	4.23 %	3.64 %
GW190421_213856	HL	-	17.8 %	-	-	10.08 %	17.4 %	-
GW190514_065416$^{*\dagger L}$	HL	-	15.22 %	-	-	19.48 %	4.83 %	-
GW190521_074359	HL	-	1.28 %	-	-	7.24 %	1.26 %	-
GW190527_092055	HL	-	9.04 %	-	-	2.95 %	5.56 %	-
GW190719_215514	HL	-	14.56 %	-	-	17.56 %	0.05 %	-
GW190731_140936	HL	-	18.22 %	-	-	38.6 %	4.69 %	-
GW191109_010717$^{\diamond HL}$	HL	-	1.52 %	-	-	0.11 %	5.15 %	-
GW191204_110529	HL	-	0.07 %	-	-	19.49 %	13.28 %	-
GW191222_033537	HL	-	15.61 %	-	-	6.84 %	16.18 %	-
GW200128_022011	HL	-	8.28 %	-	-	38.67 %	18.79 %	-
GW200220_124850	HL	-	12.94 %	-	-	23.86 %	15.56 %	-
GW200306_093714	HL	-	0.09 %	-	-	6.0 %	0.08 %	-
GW190925_232845	HV	-	-	0.86 %	-	8.16 %	-	16.97 %
GW200302_015811	HV	-	-	1.4 %	-	2.25 %	-	31.8 %
GW190620_030421	LV	-	-	-	0.81 %	-	1.5 %	55.53 %
GW190630_185205	LV	-	-	-	1.15 %	-	7.98 %	16.08 %
GW190910_112807	LV	-	-	-	3.22 %	-	4.2 %	19.1 %
GW200112_155838	LV	-	-	-	5.79 %	-	16.7 %	20.93 %

Applications: Testing general relativity

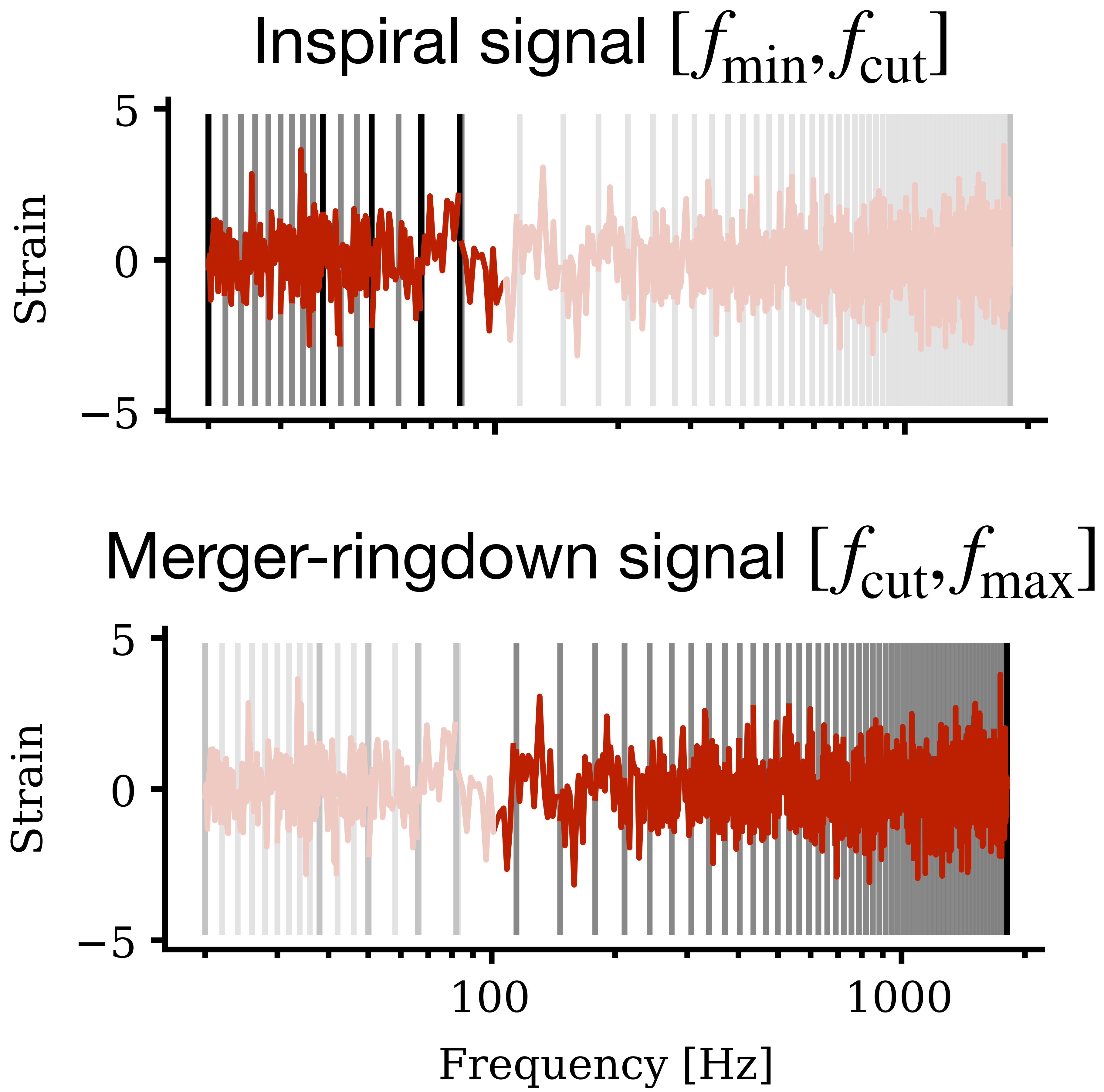
- Inspiral-merger-ringdown (IMR) consistency tests

- Same M_f & χ_f ?

- Quantify deviation:

$$\frac{\Delta M_f}{\overline{M}_f} = 2\frac{M_f^{\text{insp}} - M_f^{\text{postinsp}}}{M_f^{\text{insp}} + M_f^{\text{postinsp}}}$$

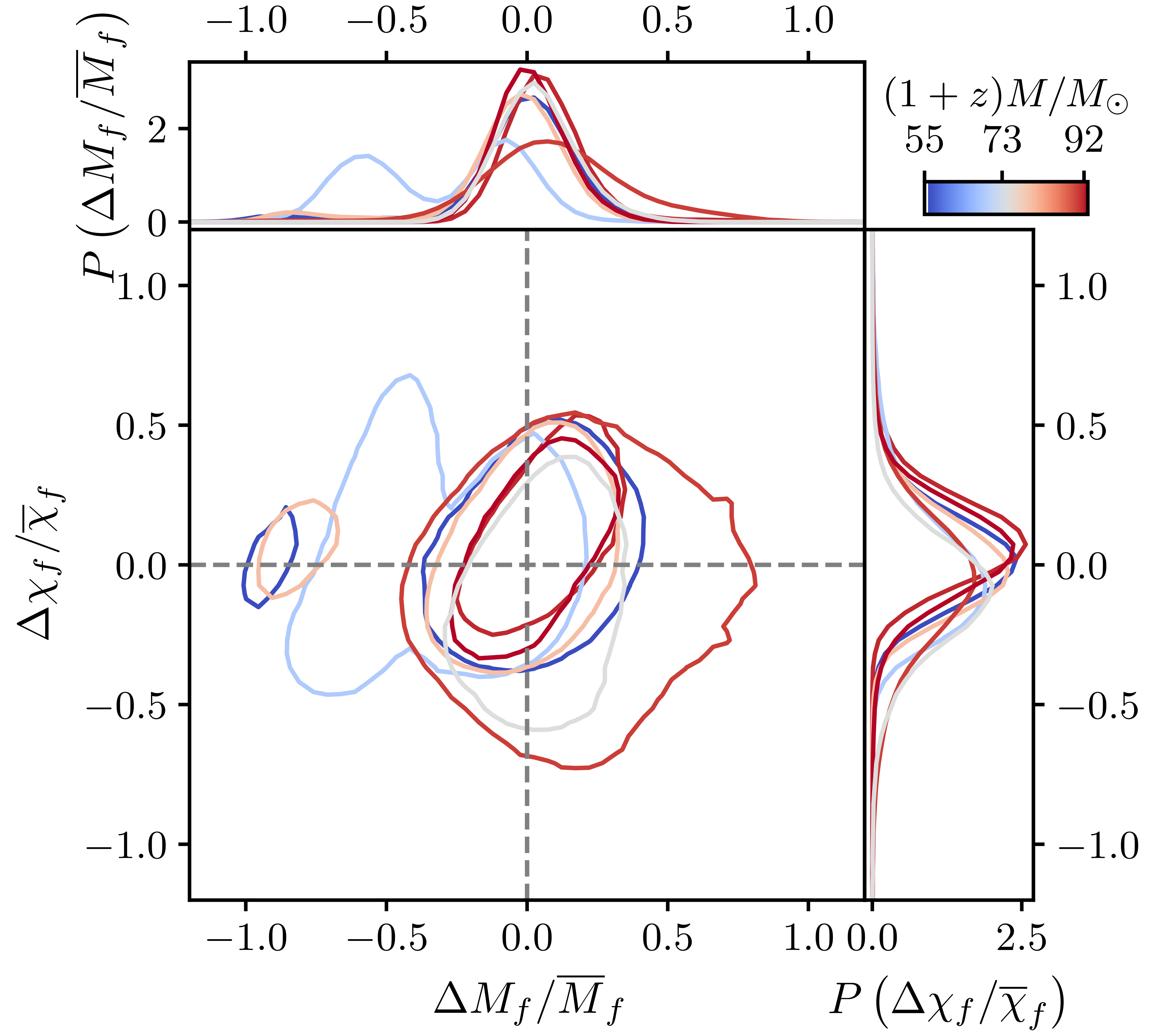
$$\frac{\Delta \chi_f}{\overline{\chi}_f} = 2\frac{\chi_f^{\text{insp}} - \chi_f^{\text{postinsp}}}{\chi_f^{\text{insp}} + \chi_f^{\text{postinsp}}}$$



Gosh+ Class. Quantum Grav. 2017, arXiv:1704.06784

Applications: Testing general relativity

- Inspiral-merger-ringdown (IMR) consistency tests
- Same M_f & χ_f ?
- 7 O3 events



Gosh+ Class. Quantum Grav. 2017, arXiv:1704.06784

Summary: DINGO-T1

- Flexibility of DINGO-T1 allows us to do …

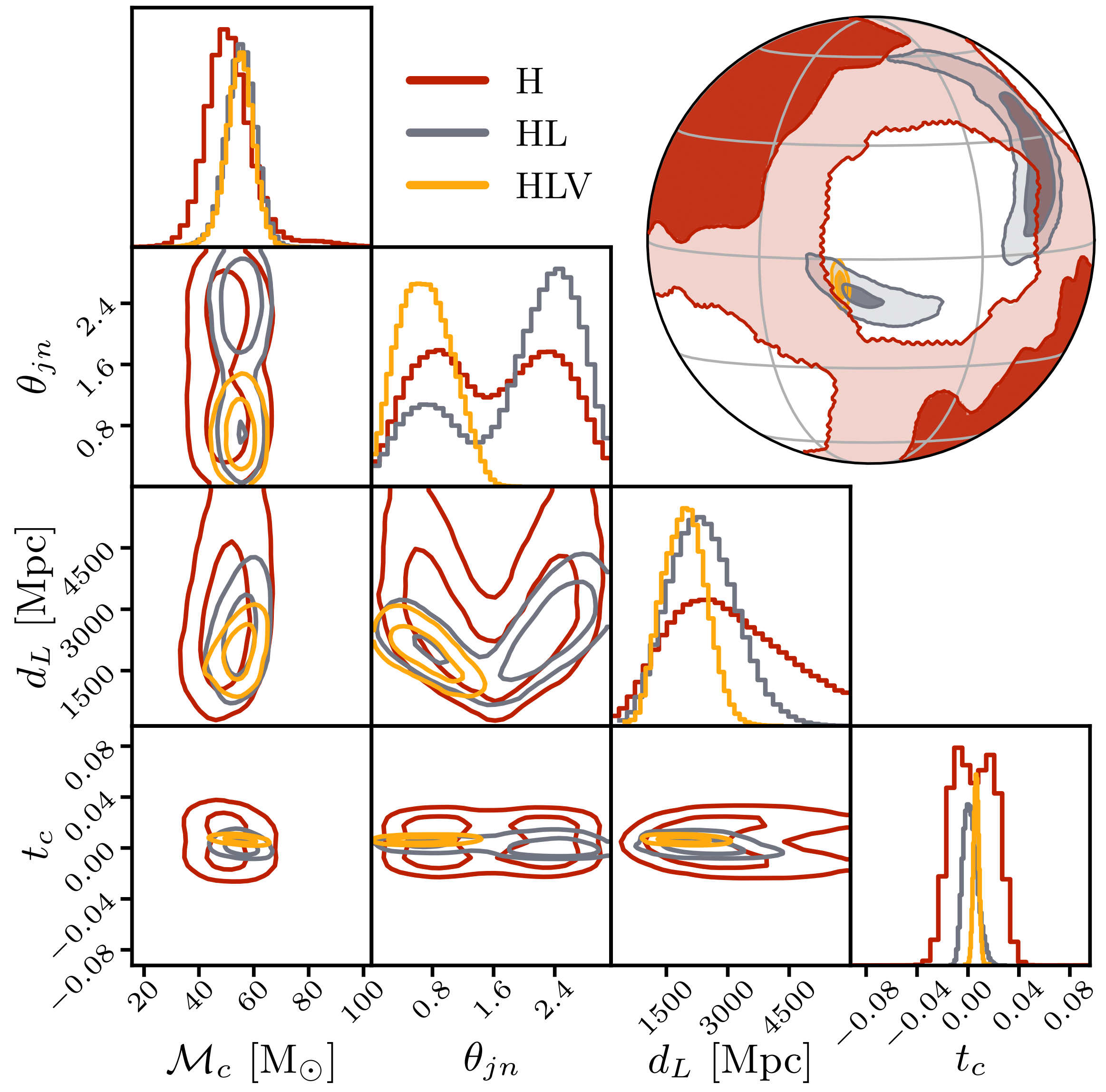
  … large scale PE analysis

  … fast re-analysis with different data analysis settings

  … IMR consistency tests

- All analyses in this paper would have required training **94 separate DINGO models!**

- Model & Tutorial online



Make this plot yourself!

NPE: Usefulness & Future directions

When to use NPE?

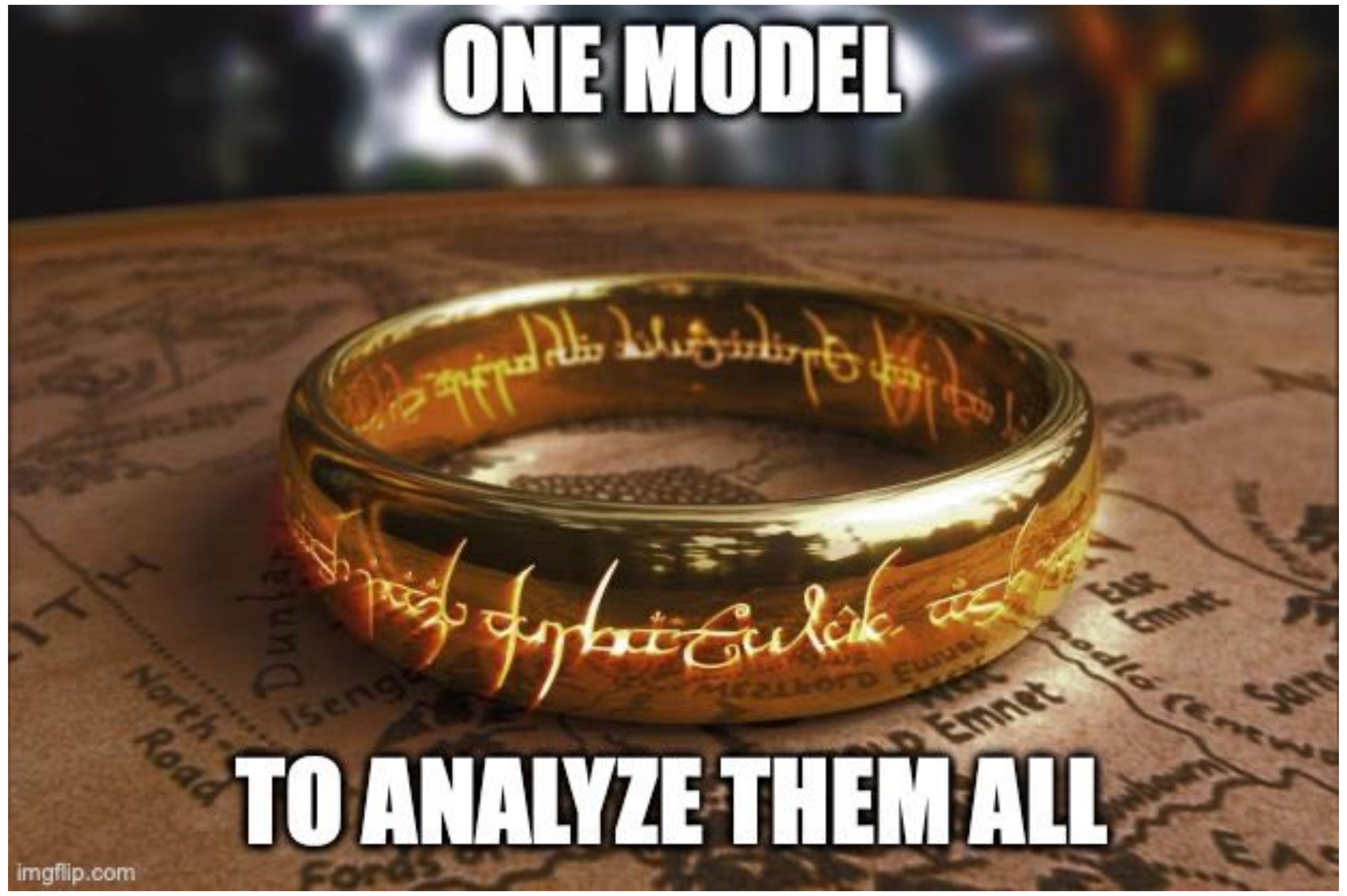
- Faset PE for a large number of injections or events
  DINGO training+inference: 1015 kWh, Bilby: 4000 kWh

Santoliquido+ PRD 2025, arXiv:2504.21087

- Expensive waveform models like EOB

Dreaming about the future:

- One model to analyze all events
- What data modality works the best?
- Real noise & trustworthy validation methods



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?