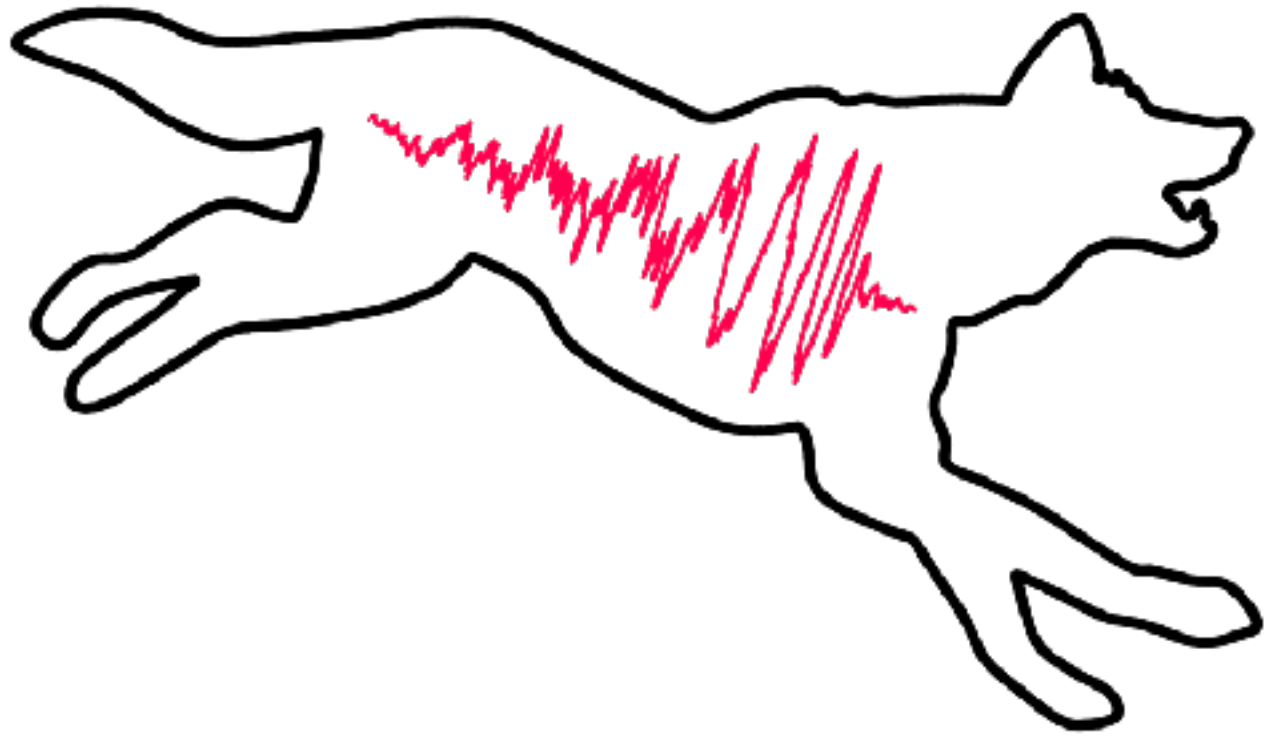
MAX PLANCK INSTITUTE
FOR INTELLIGENT SYSTEMS

Gravitational-wave parameter estimation with DINGO

From current to future ground-based detectors



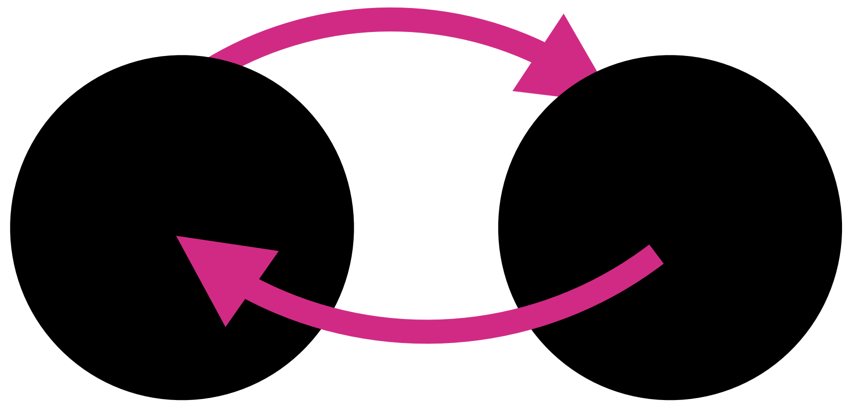
Annalena Kofler, 22.01.2026

RWTH Aachen

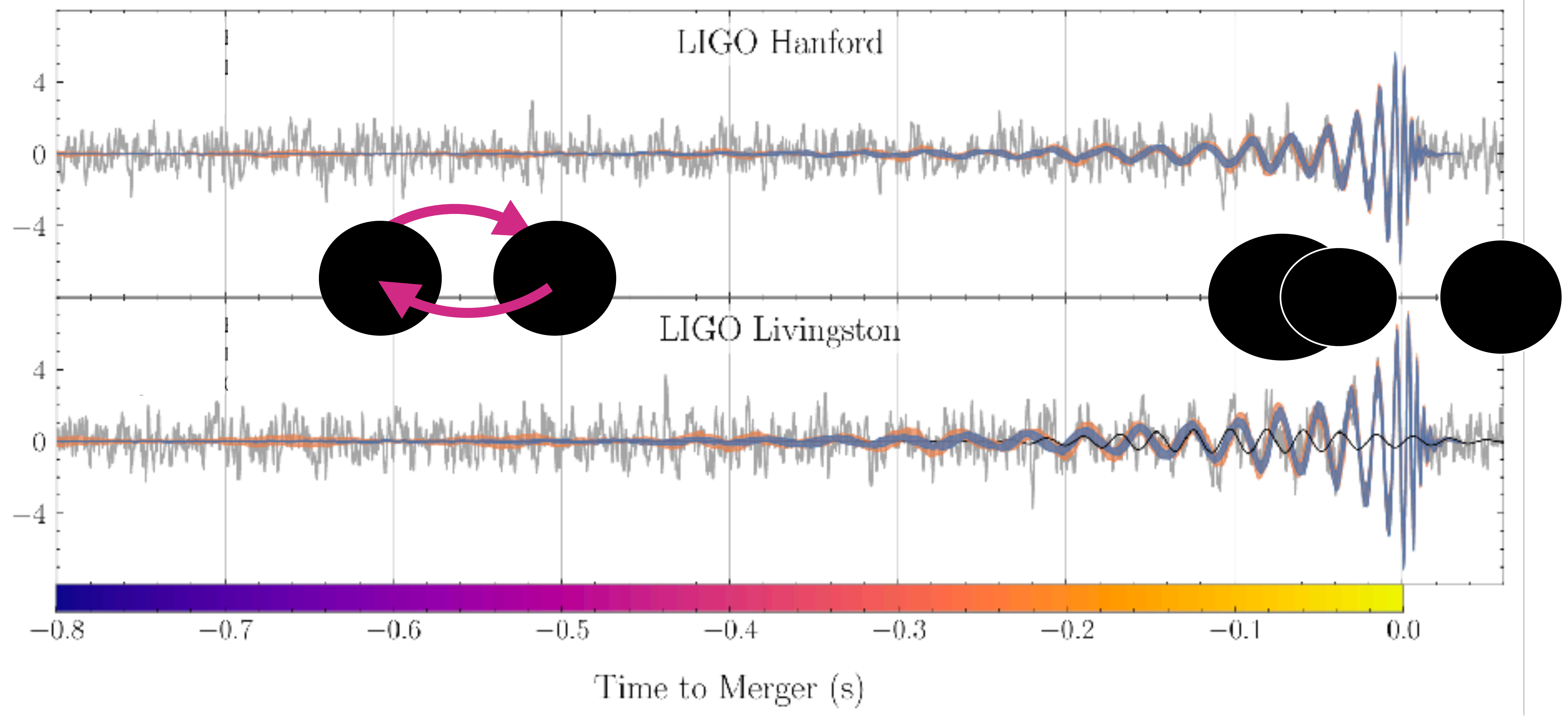
What are gravitational waves?

What are gravitational waves?

Black holes merge
(or other heavy objects)
→ Emit gravitational wave → Measured
in detectors



Universe



Described by
physics
parameters

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

Masses, spins, sky
position, orientation

Measured data d

Gupte+, PRL 2025

Why is GW science exciting?

New window into the universe!

Direct observation of two merging black holes

Physical Review Letters

FEATURED IN PHYSICS | EDITORS' SUGGESTION | OPEN ACCESS

Observation of Gravitational Waves from a Binary Black Hole Merger

B. P. Abbott[1], R. Abbott[1], T. D. Abbott[2], M. R. Abernathy[1], F. Acernese[3,4], K. Ackley[5], C. Adams[6], T. Adams[7], P. Addesso[3] *et al.* (LIGO Scientific Collaboration and Virgo Collaboration)

Phys. Rev. Lett. **116**, 061102 – **Published 11 February, 2016**

Estimating the expansion rate of the Universe

nature

Letter | Published: 16 October 2017

A gravitational-wave standard siren measurement of the Hubble constant

The LIGO Scientific Collaboration and The Virgo Collaboration, The 1M2H Collaboration, The Dark Energy Camera GW-EM Collaboration and the DES Collaboration, The DLT40 Collaboration, The Las Cumbres Observatory Collaboration, The VINROUGE Collaboration & The MASTER Collaboration

Nature 551, 85–88 (2017) | Cite this article

nature

Article | Published: 23 October 2019

Identification of strontium in the merger of two neutron stars

Science

Light curves of the neutron star merger GW170817/SSS17a: Implications for r-process nucleosynthesis

Heavy elements (gold, platinum, uranium, …) on Earth likely created by neutron star mergers

Testing matter under extreme conditions

Physical Review Letters

EDITORS' SUGGESTION

GW170817: Measurements of Neutron Star Radii and Equation of State

B. P. Abbott[1], R. Abbott[1], T. D. Abbott[2], F. Acernese[3,4], K. Ackley[5], C. Adams[6], T. Adams[7], P. Addesso[8], R. X. Adhikari[1] *et al.* (The LIGO Scientific Collaboration and the Virgo Collaboration)

Phys. Rev. Lett. **121**, 161101 – **Published 15 October, 2018**

THE ASTROPHYSICAL JOURNAL LETTERS

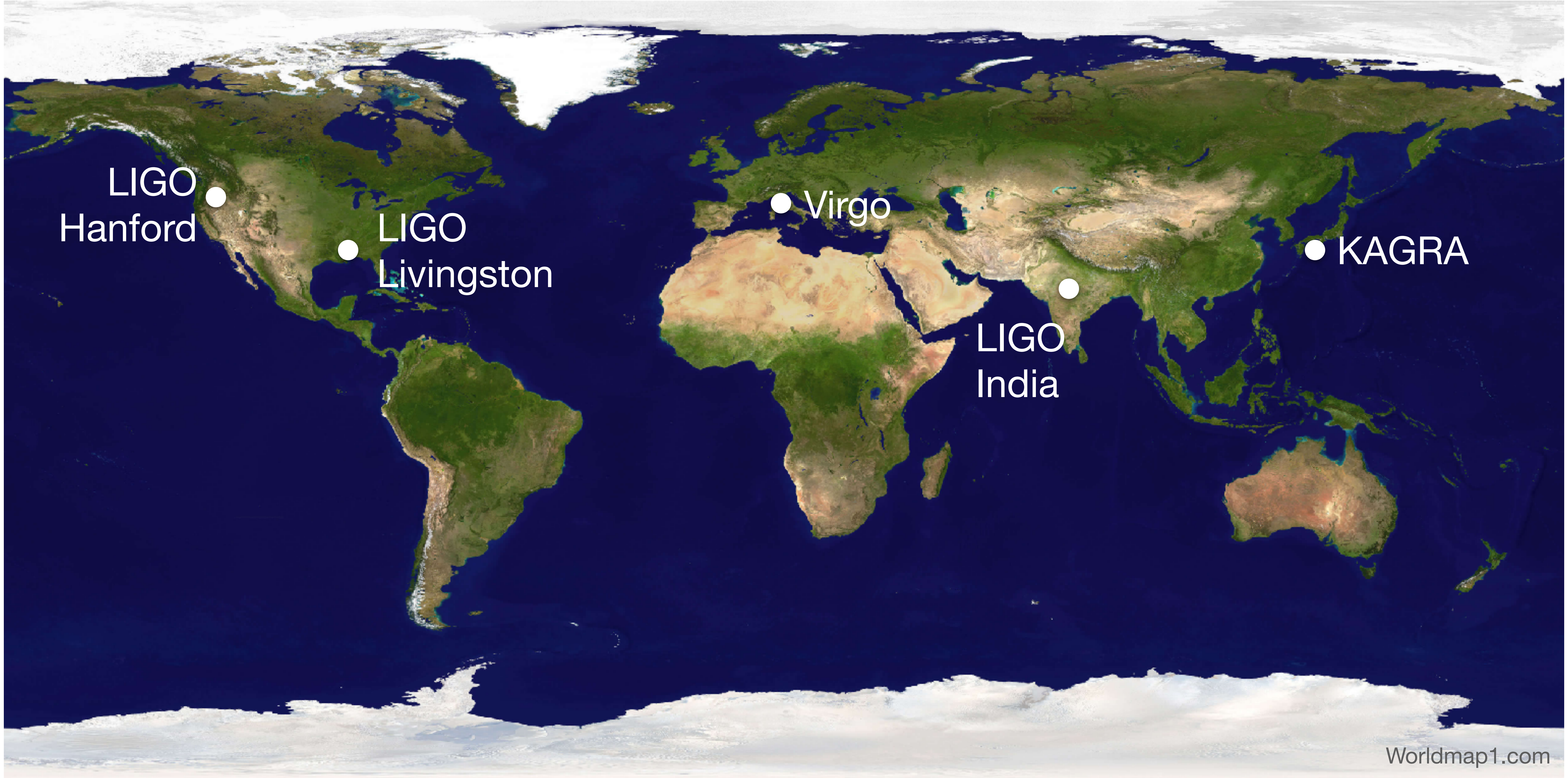
INTEGRAL Detection of the First Prompt Gamma-Ray Signal Coincident with the Gravitational-wave Event GW170817

V. Savchenko, C. Ferrigno, E. Kuulkers, A. Bazzano, E. Bozzo, S. Brandt, J. Chenevez,

Investigating fundamental concepts: Lorentz invariance, speed of gravity, equivalence principle

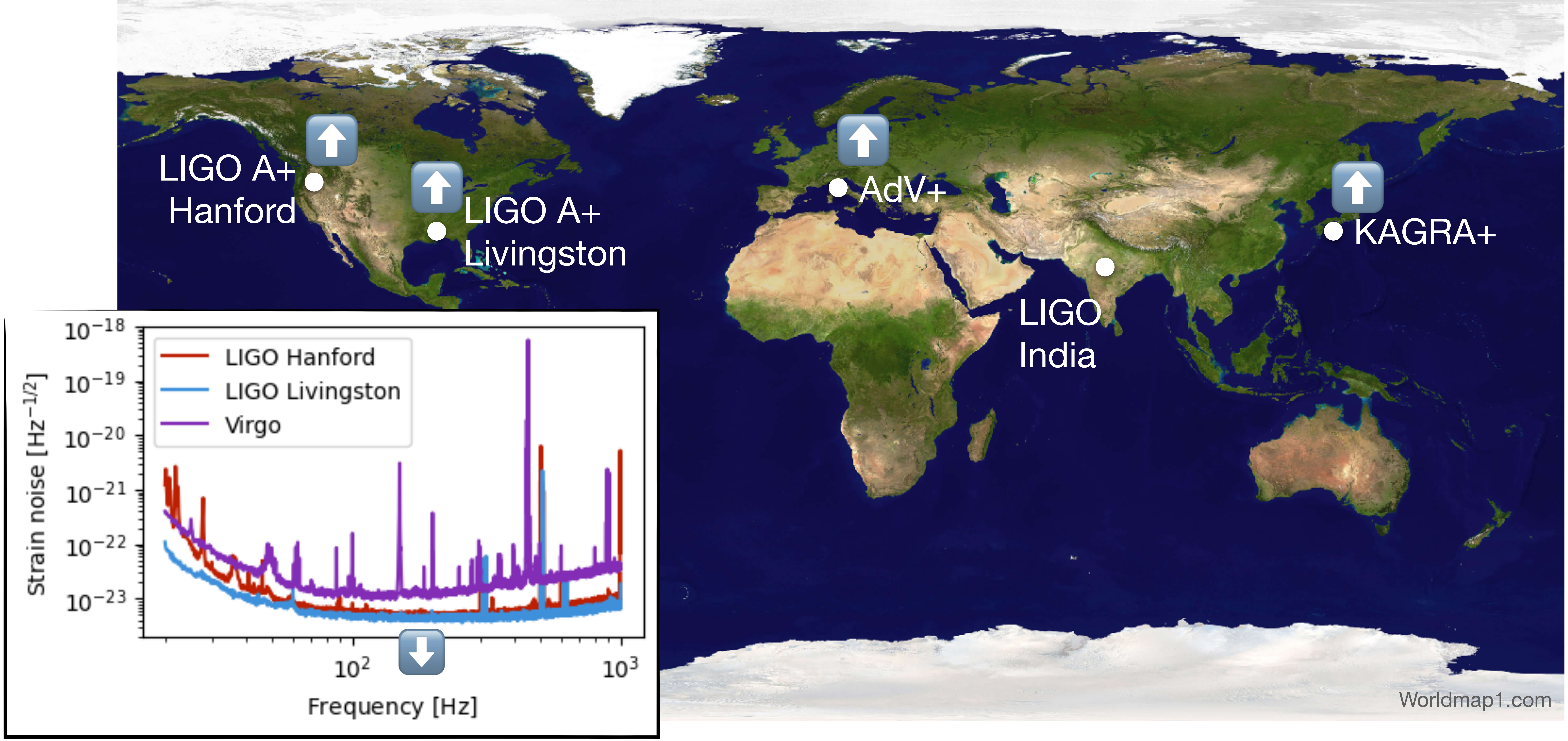
Promising future: Current & proposed detectors

Currently: 2G detectors



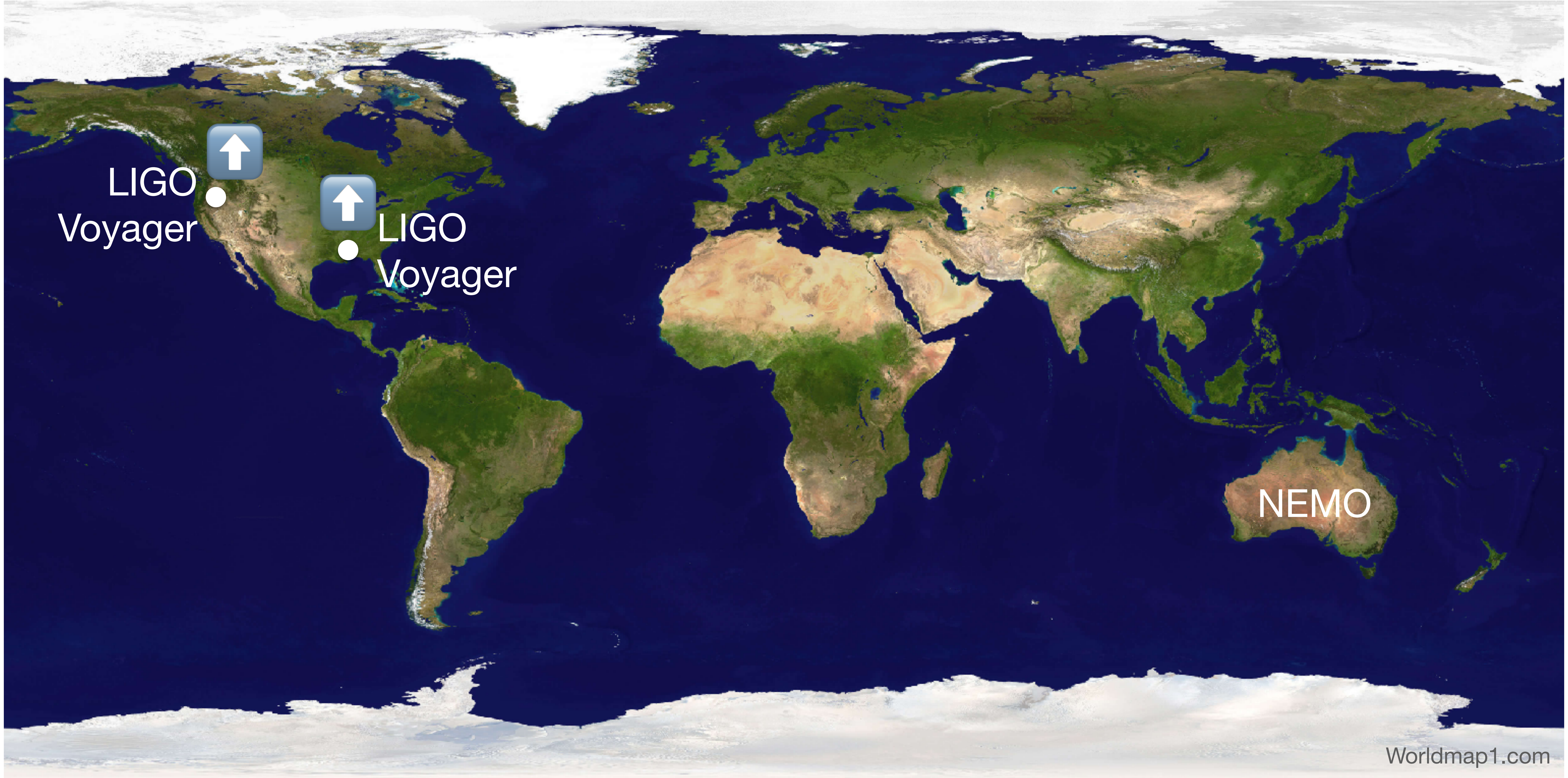
Promising future: Current & proposed detectors

Proposed: 2G+ detectors



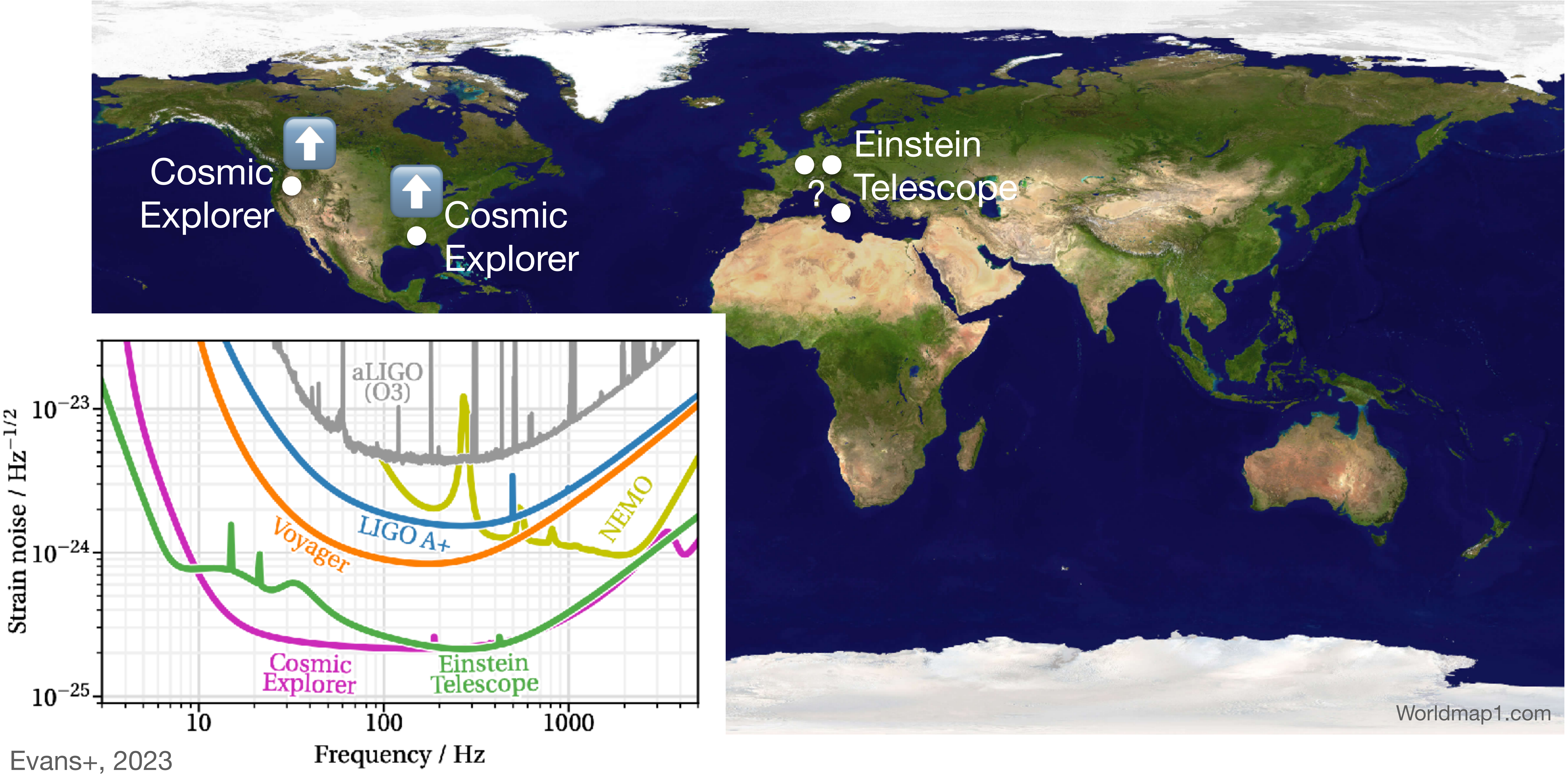
Promising future: Current & proposed detectors

Proposed: 2.5G detectors



Promising future: Current & proposed detectors

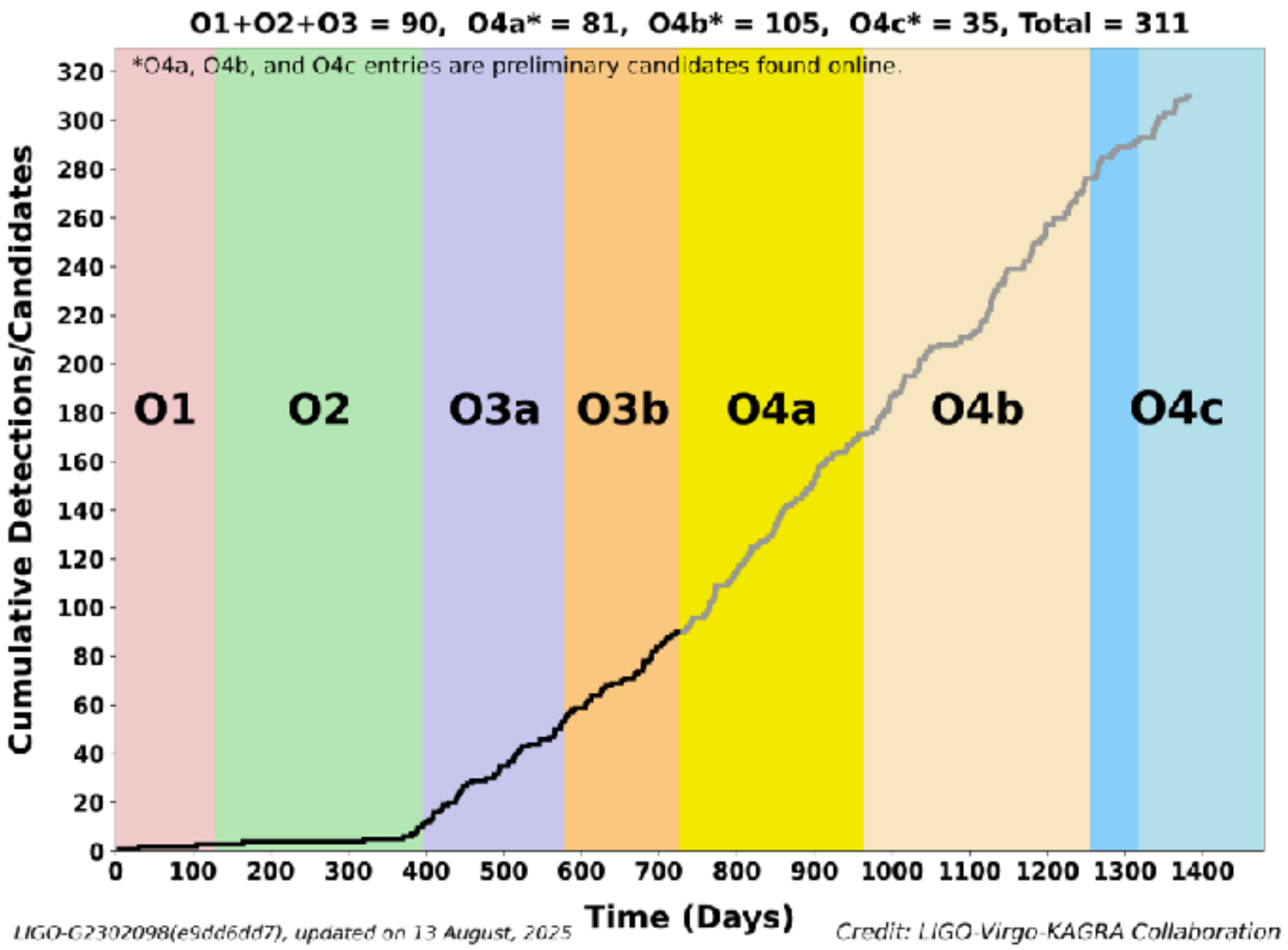
Proposed: 3G detectors



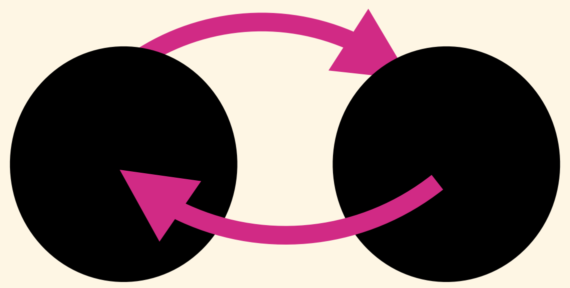
Evans+, 2023

What does increased sensitivity mean?

- Increasing number of detections:
  - Currently: ~ 5 per week
  - 3G detectors: ~ 200 per day
- Signals detectable longer before the merger → overlapping signals
- We need efficient analysis methods for…
  - Signal detection
  - Parameter estimation
    - Currently: **standard methods need minutes - hours for a single event**
    - **ML as a solution?**



What is parameter estimation?



Physics parameters

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

Masses, spins, sky position, orientation

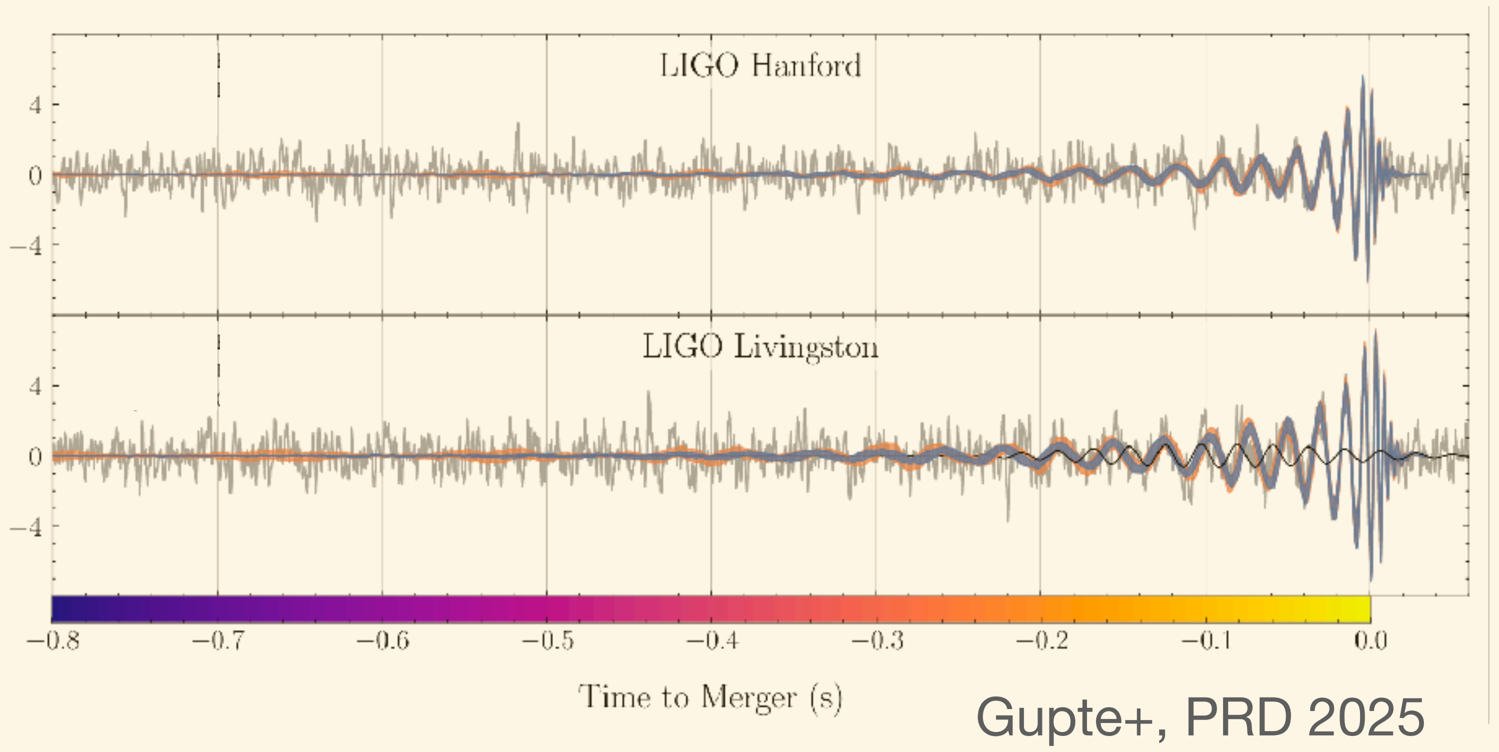
Universe

Simulator

Posterior Estimation

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

Measured data d

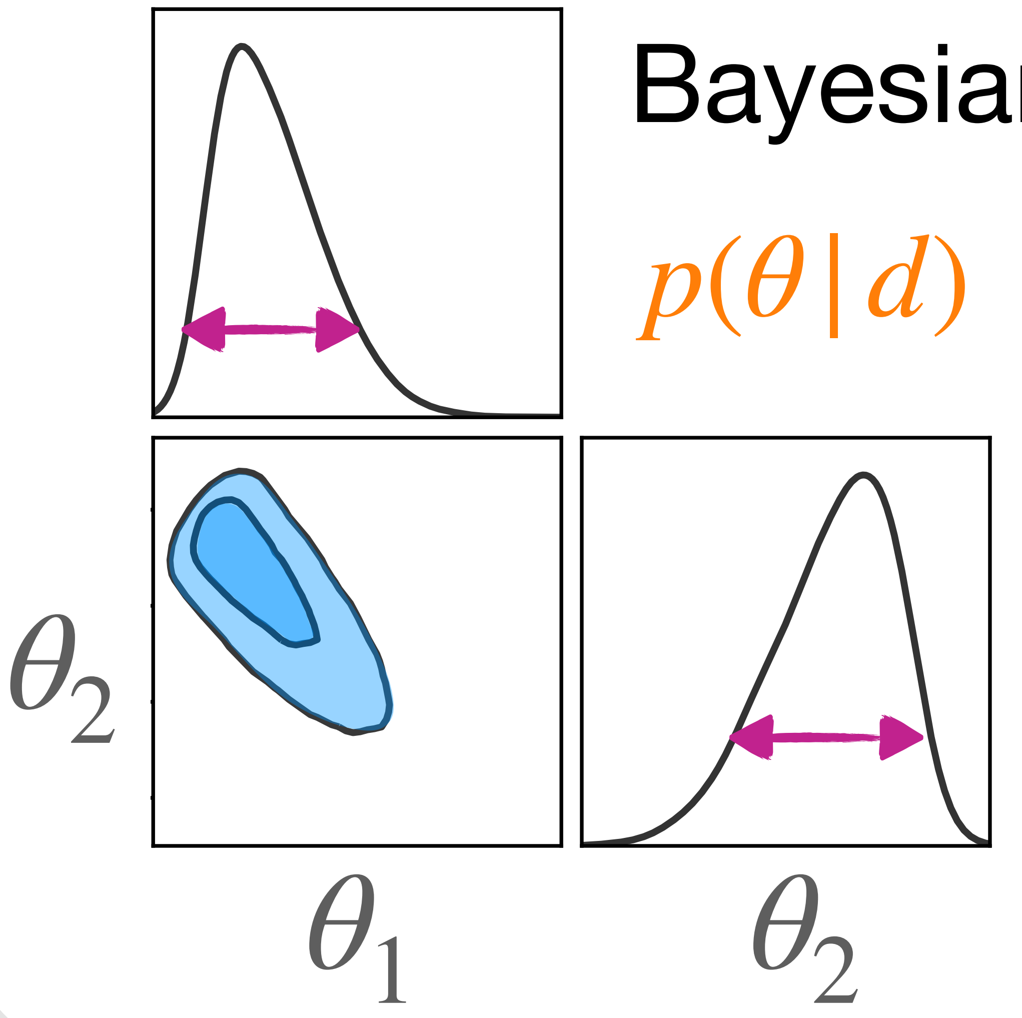


Gupte+, PRD 2025

Inverse direction with Bayesian inference:

$$\underbrace{p(\theta\,|\,d)}_{\text{Posterior}} = \frac{\overbrace{p(d\,|\,\theta)}^{\text{Likelihood}}}{p(d)}\, p(\theta)$$

Prior belief



Bayesian posterior

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

Uncertainties

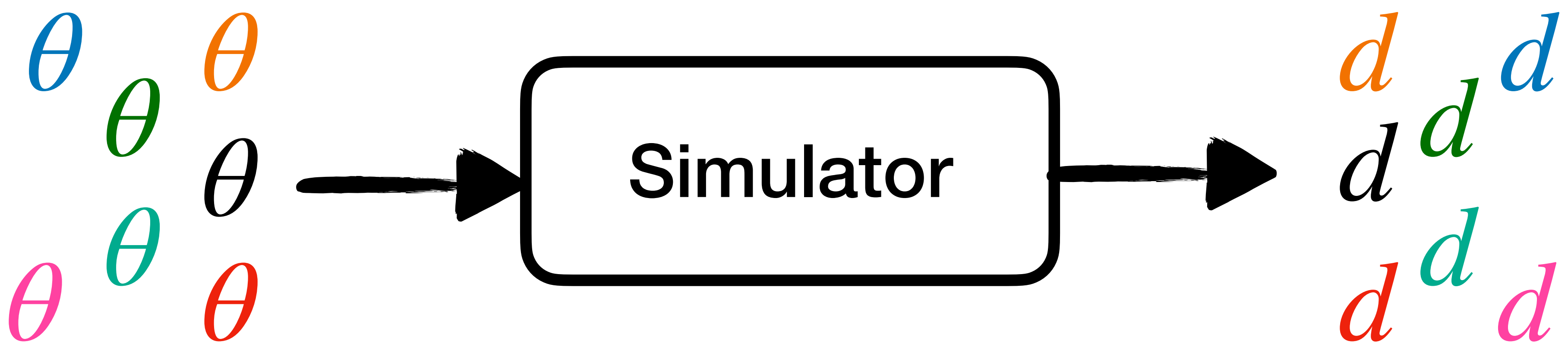
Correlations

Simulation-based inference*

***specifically: Neural Posterior Estimation (NPE)**

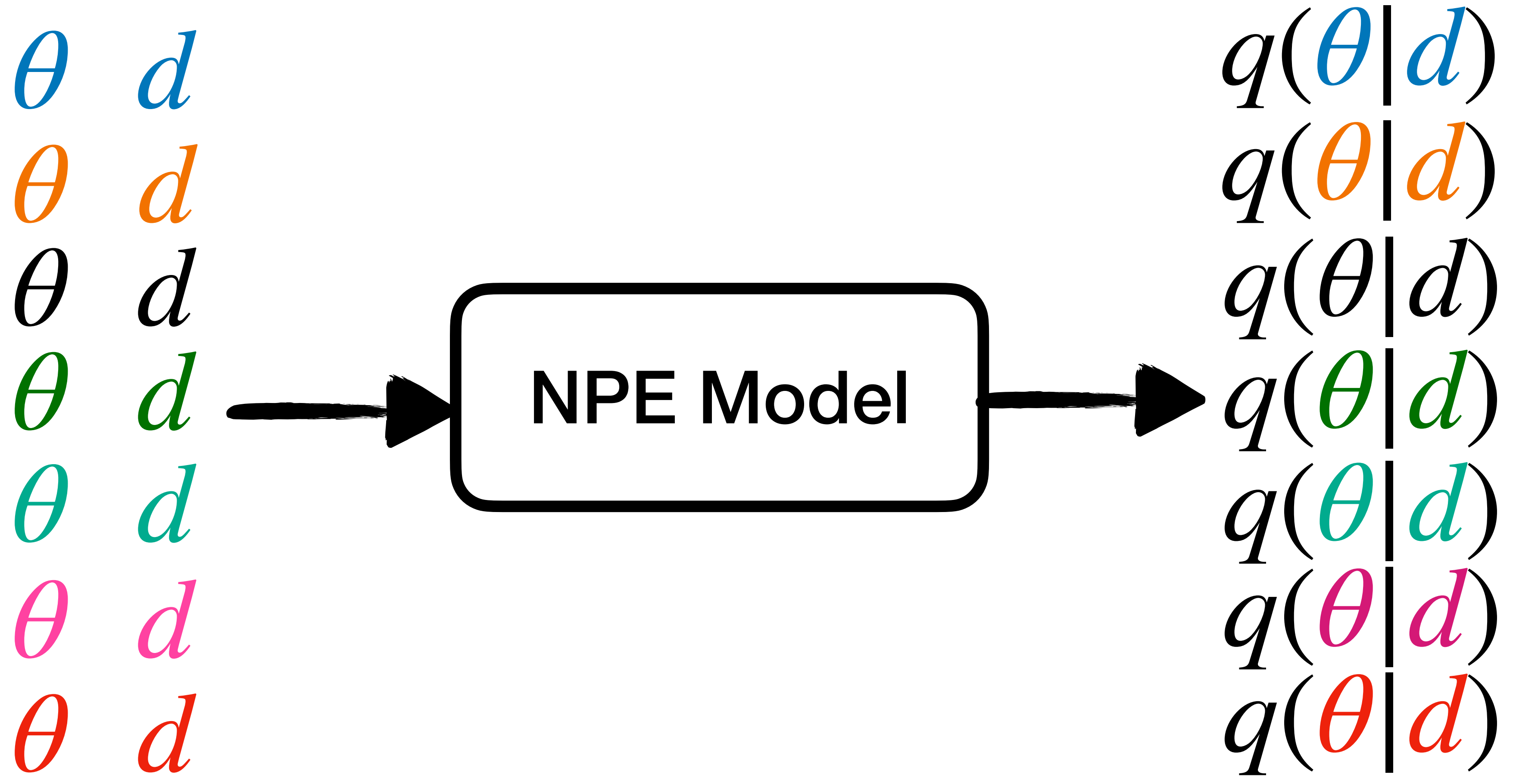
Idea: Train model to invert simulator

1. Generate large training data set with simulator

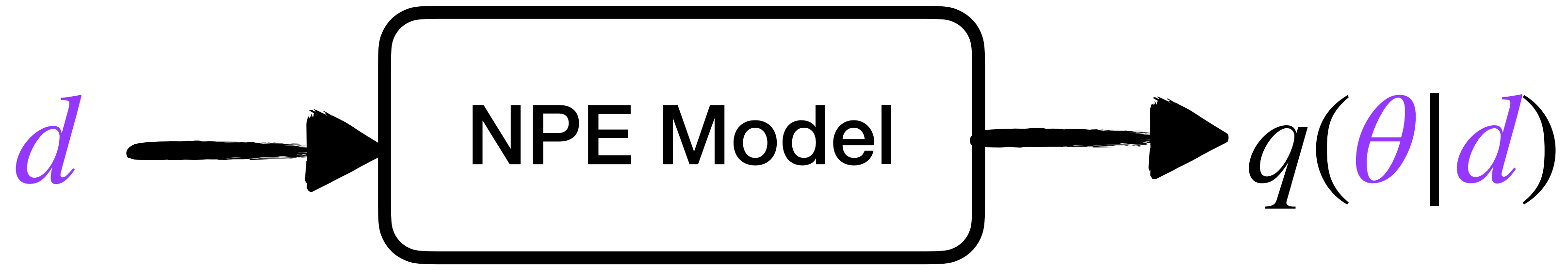


2. Train machine learning model q to approximate posterior

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



3. Evaluate model on measurement d to obtain $q(\theta|d)$



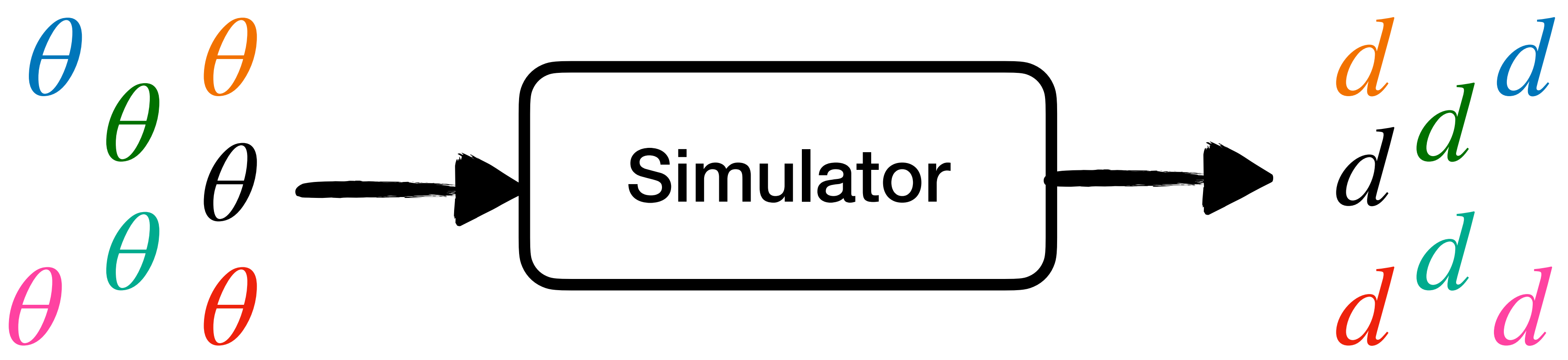
Cranmer+ 2020, arXiv:1911.01429

Simulation-based inference*

***specifically: Neural Posterior Estimation (NPE)**

Idea: Train model to invert simulator

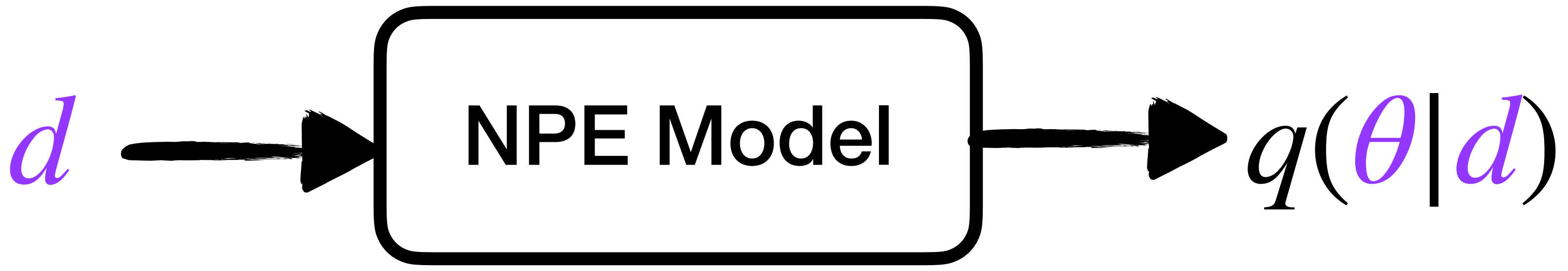
1. Generate large training data set with simulator



2. Train machine learning model q to approximate posterior

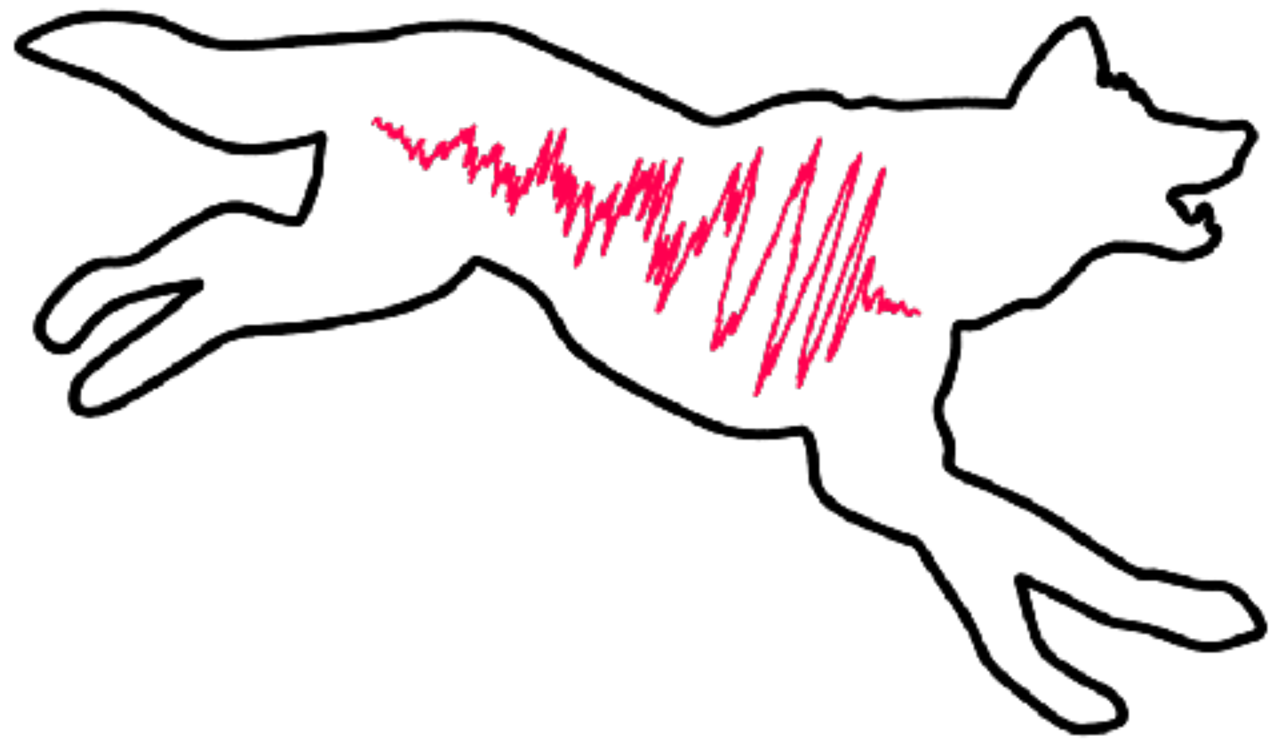
$$p(\theta \mid d) \approx q(\theta \mid d)$$

3. Evaluate model on measurement d to obtain $q(\theta|d)$



Amortization with NPE:
- Train once
- Evaluate on many events

Cranmer+ 2020, arXiv:1911.01429



NPE for gravitational waves with DINGO

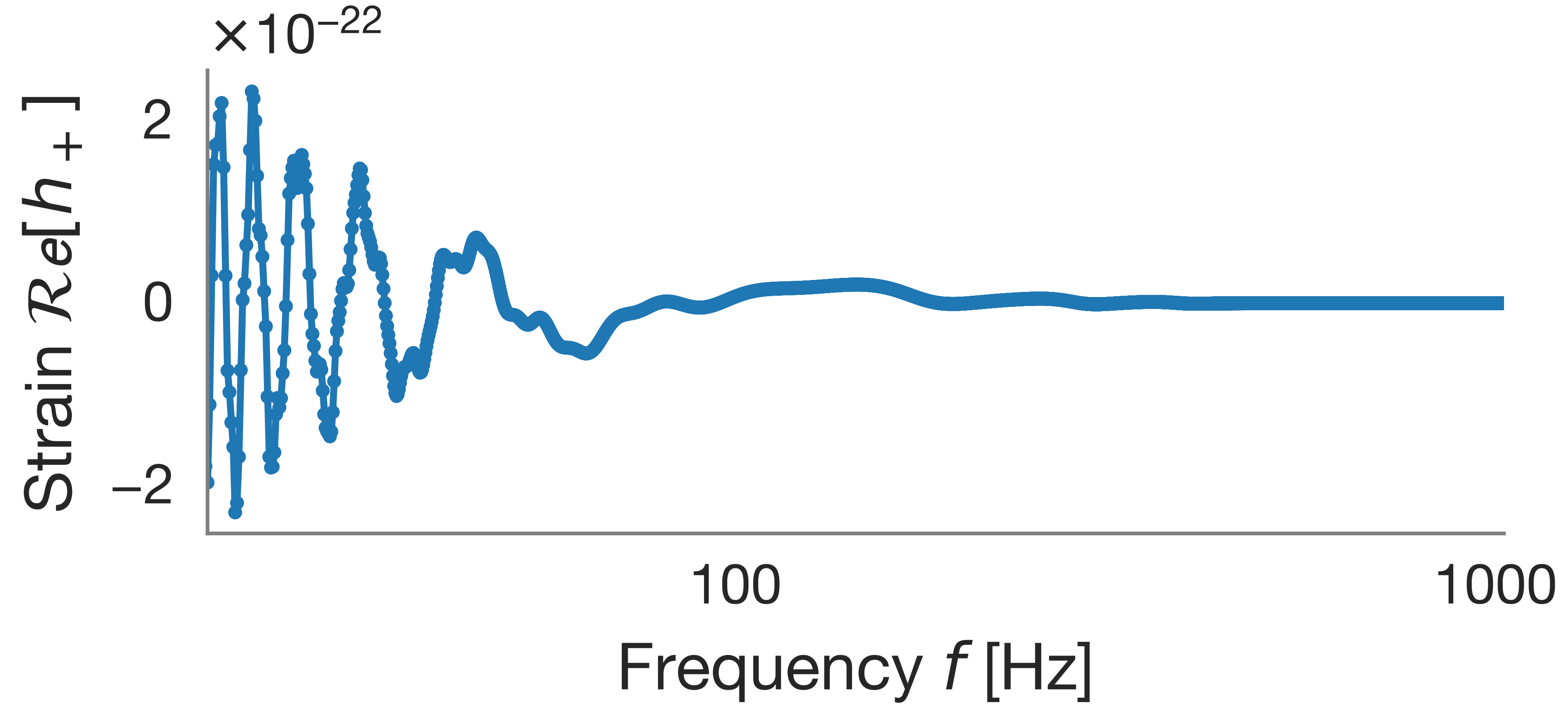
DINGO = <u>D</u>eep <u>IN</u>ference for <u>G</u>ravitational wave <u>O</u>bservations

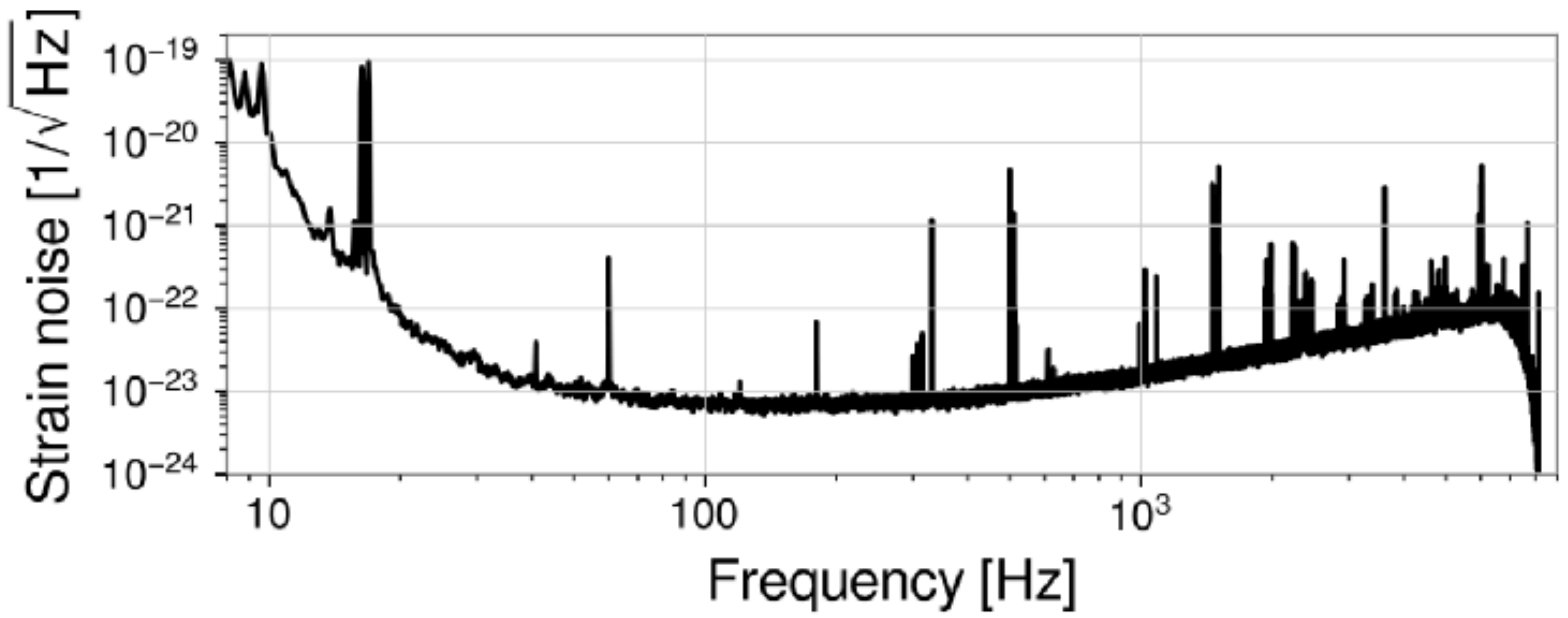
NPE for gravitational waves: Overview

- Generate simulated waveforms: $\theta \sim p(\theta), \;\; h = \mathrm{simulator}(\theta) \;\rightarrow \{\theta, h\}$

- Add realistic noise $S_n(f)$ to the waveform

  1. Sample noise $\quad n^{(i)} \sim \mathcal{N}(0,\, S_n^{(i)})$
  2. Add to waveform $\quad d^{(i)} = h(\theta^{(i)}) + n^{(i)}$





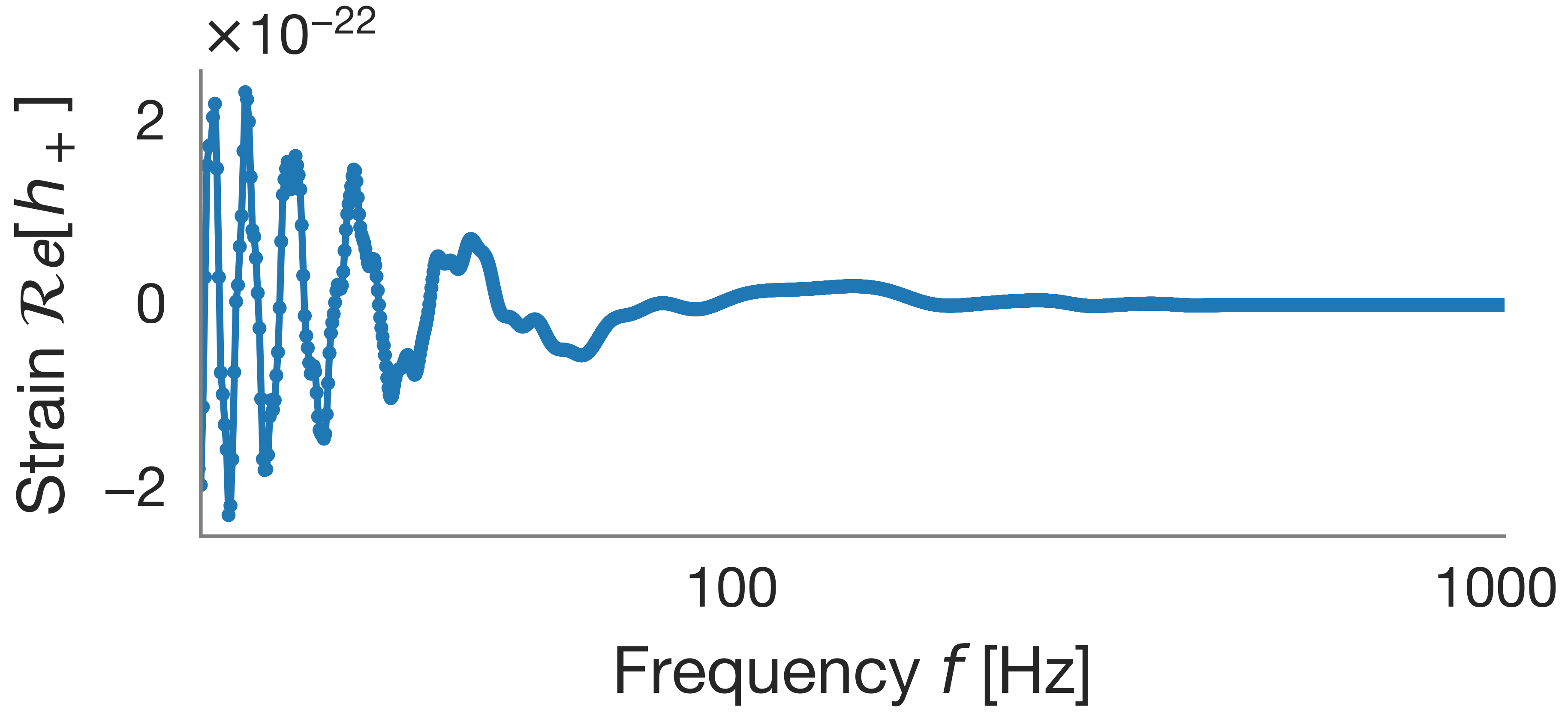
- Train density estimator

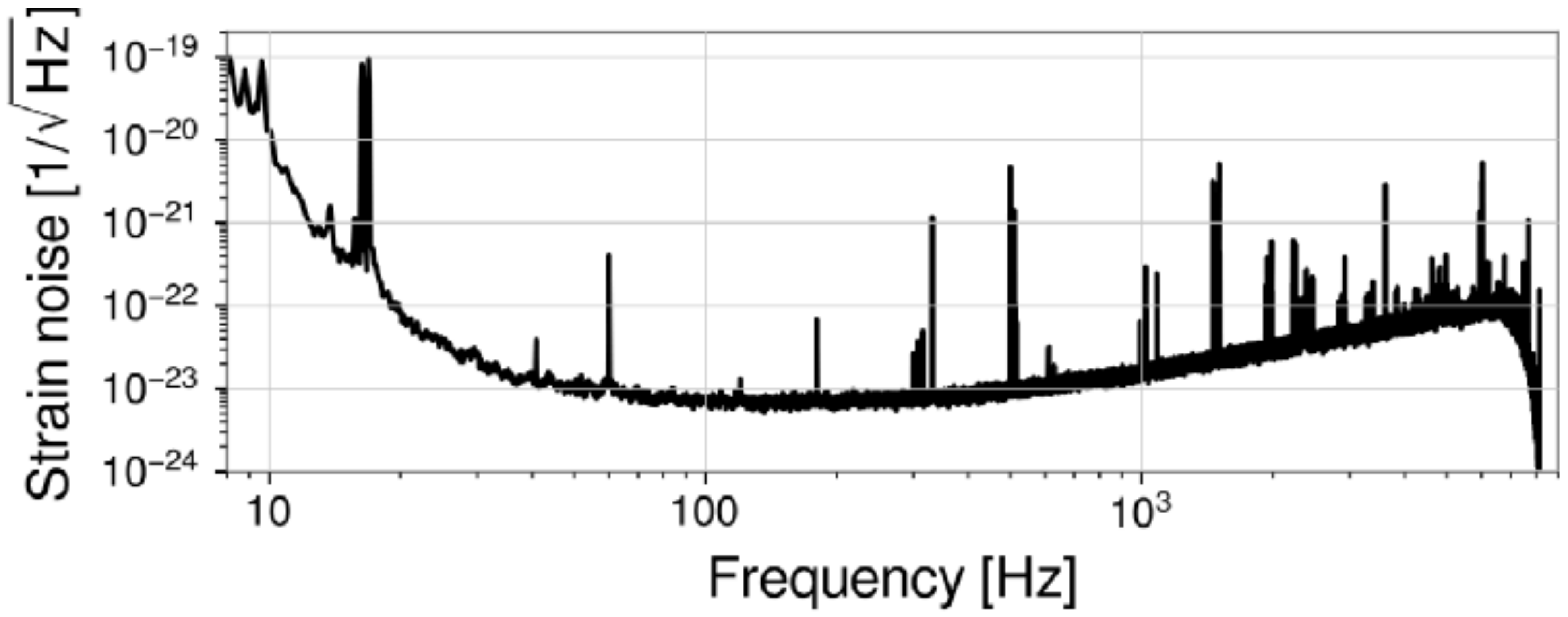
Dax+, PRL 2021, arXiv:2106.12594

Changing PSDs

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

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

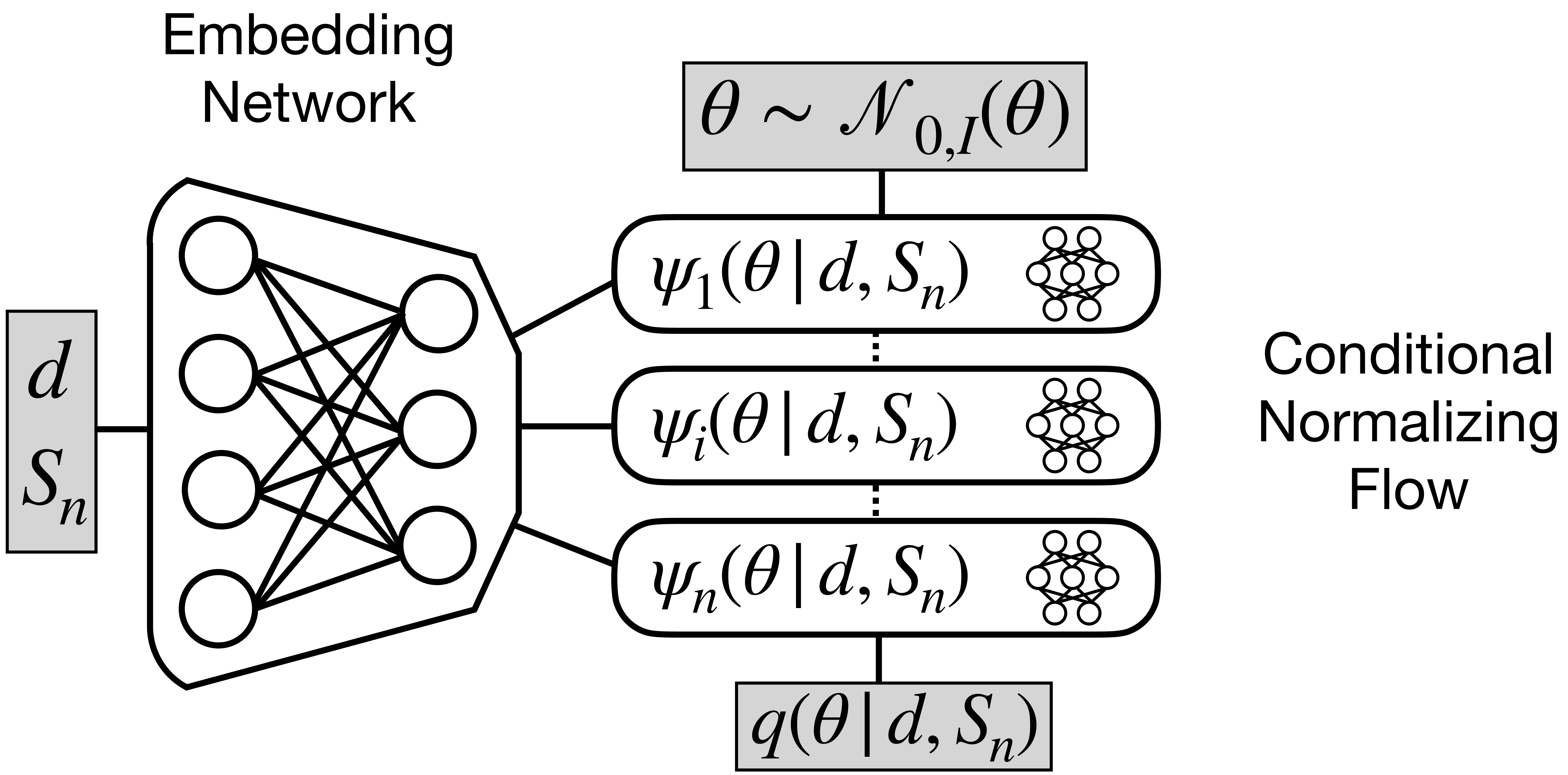




Dax+, PRL 2021, arXiv:2106.12594

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

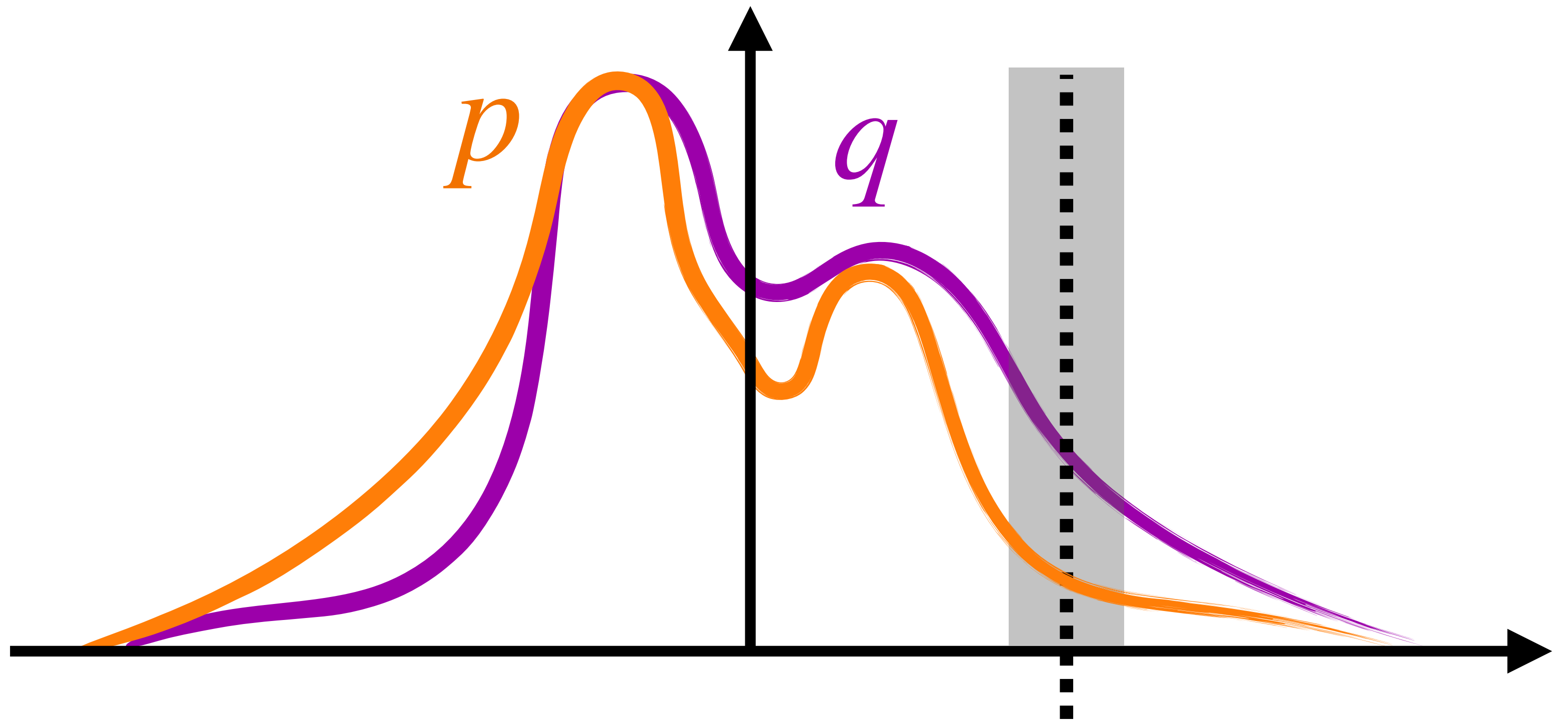
But what if the model is wrong?

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

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

$p(d\,|\,\theta)$: ∝ Gaussian noise

$p(\theta)$: Known

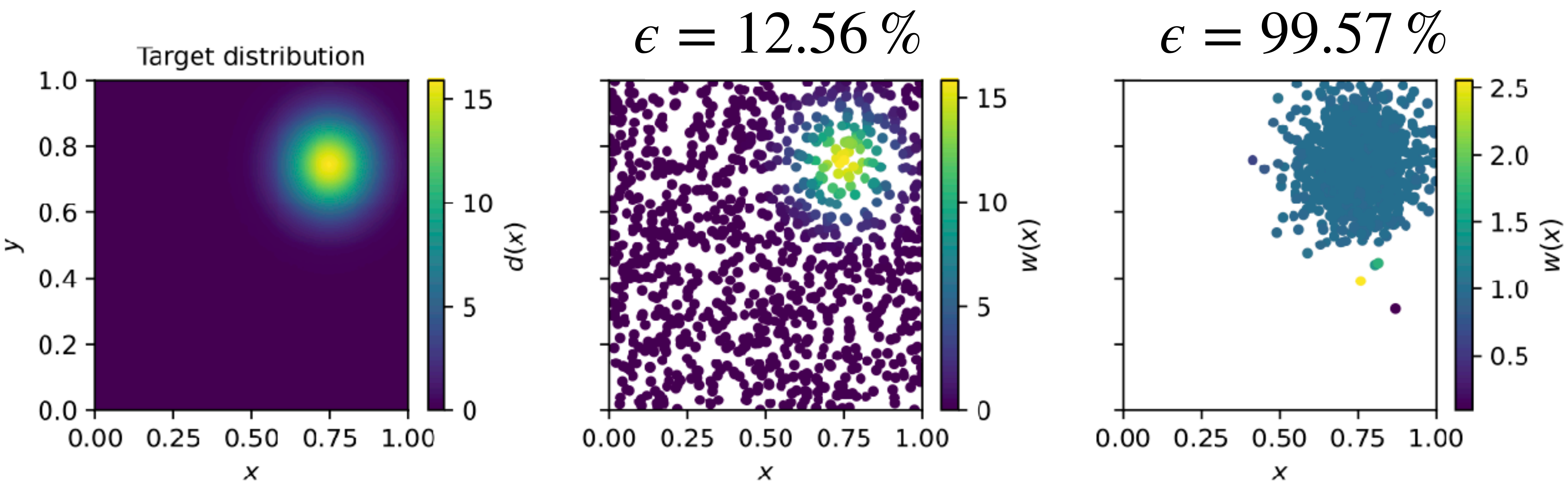


$q(\theta\,|\,d) > p(\theta\,|\,d)$

Low w

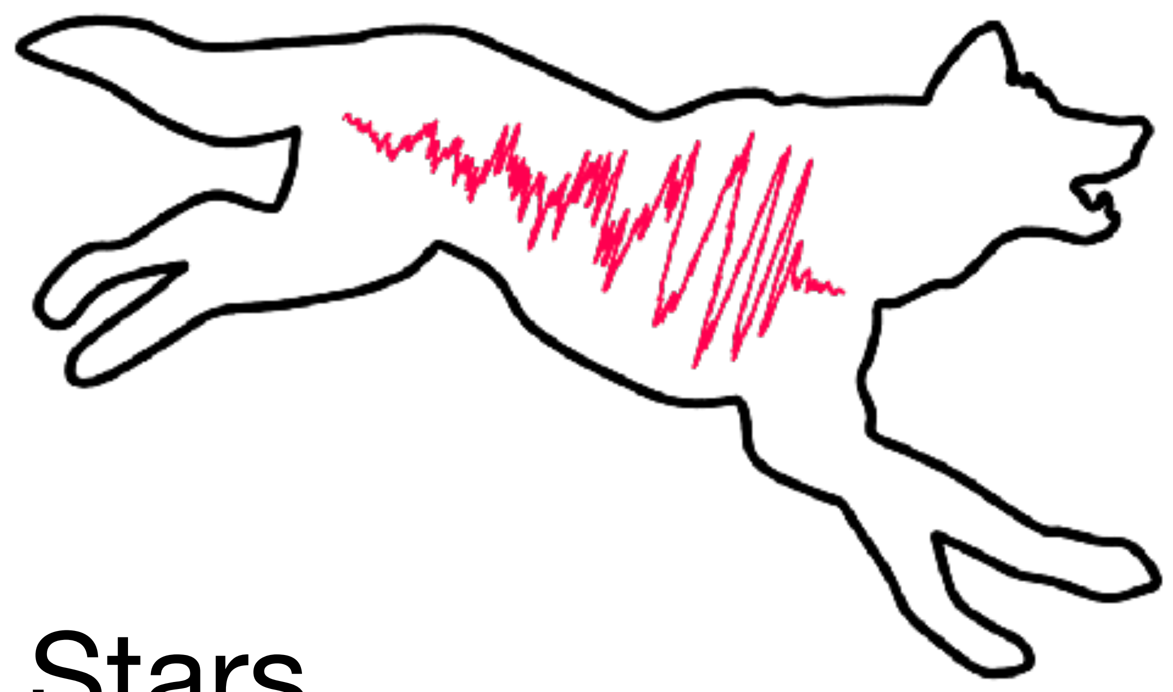
- Sample efficiency:

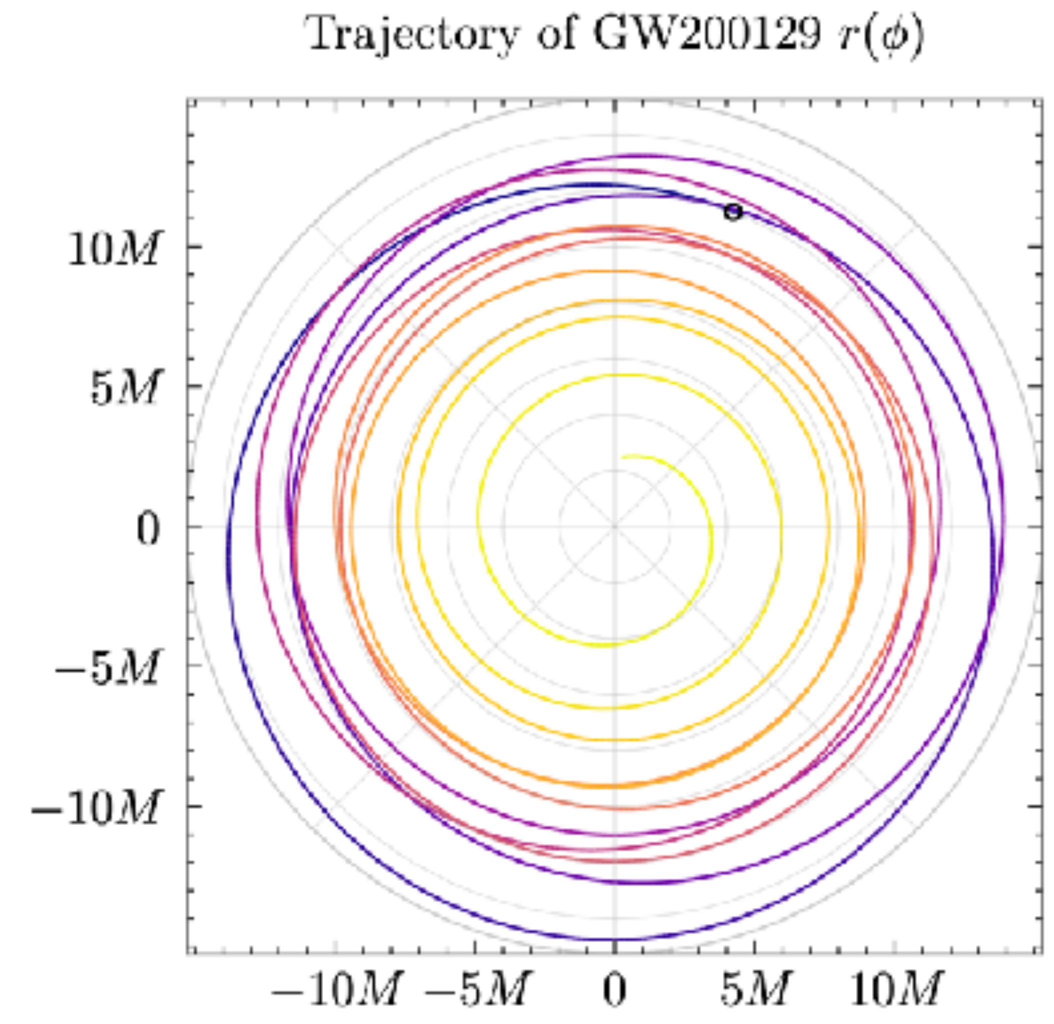
$$\epsilon = \frac{1}{N}\frac{\left(\sum_{i=1}^{N} w_i\right)^2}{\sum_{i=1}^{N} w_i^2}$$



Dax+, PRL 2023, arXiv:2210.05686

Where is DINGO used?



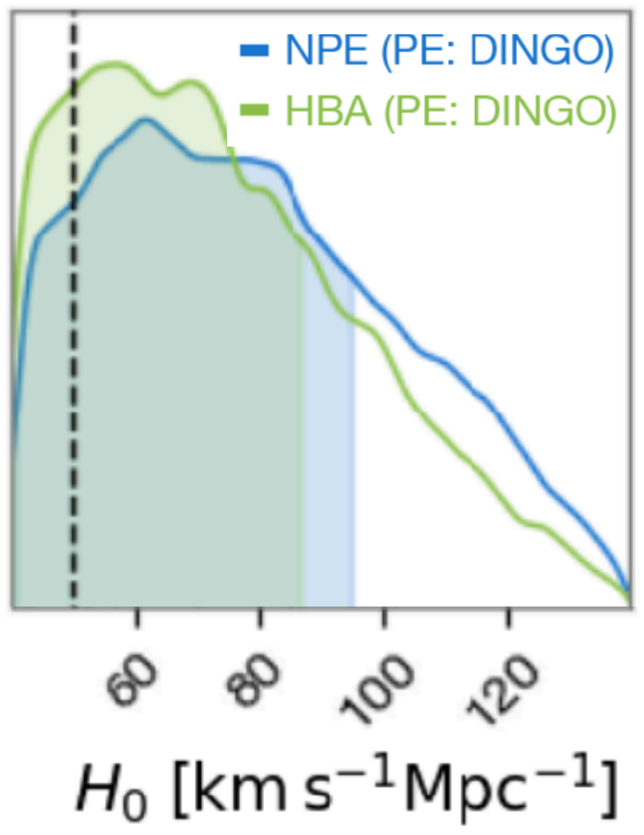


Eccentricity

Gupte+, PRD 2025

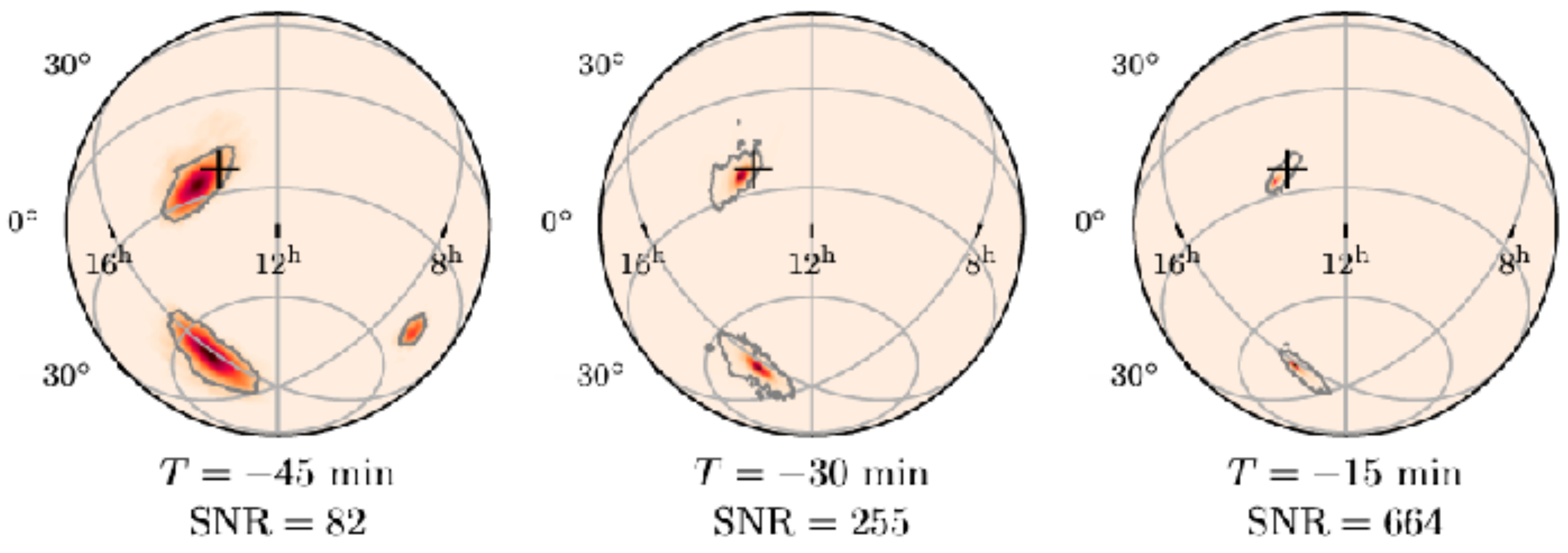
Populations

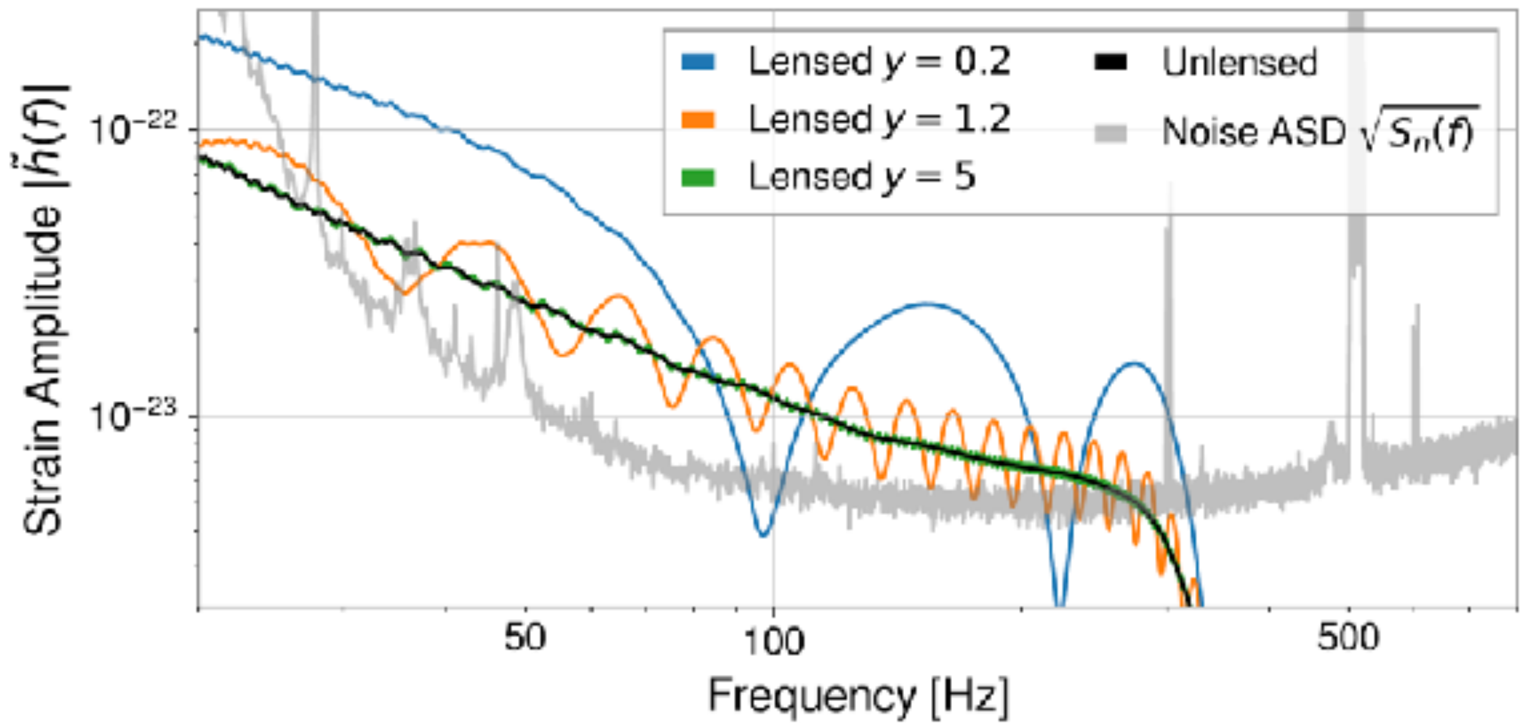
Leyde+, PRD 2024



Binary Neutron Stars

Dax+, Nature 2025





GW Lensing

Caldarola+ arXiv 2025,
Chan+, arXiv 2025

Reviewed and used within
the LVK collaboration

1) How can we make NPE flexible without retraining?

Kofler+, *Flexible Gravitational-Wave Parameter Estimation with Transformers*
arXiv:2512.02968v1, under review, 2025

2) What do we learn from using DINGO for the Einstein Telescope?

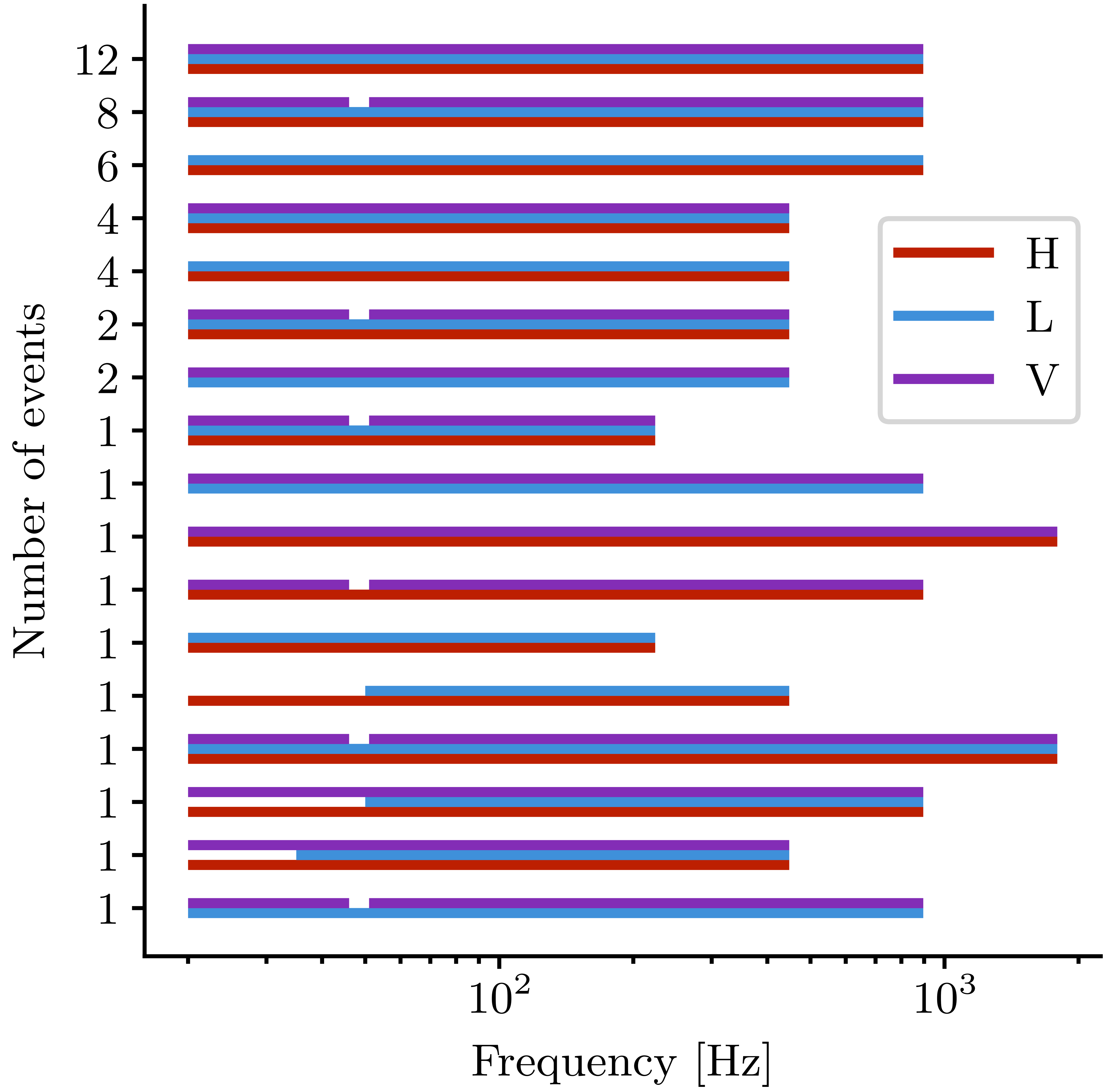
Santoliquido+, *Fast and accurate parameter estimation of high-redshift sources with the Einstein Telescope*
arXiv:2504.21087, PRD 2025

Filippo Santoliquido

Why do we need flexible NPE?

Example: Official O3 LVK analysis

- 48 events with 17 different data configurations
    - Different detectors
    - Different frequency ranges

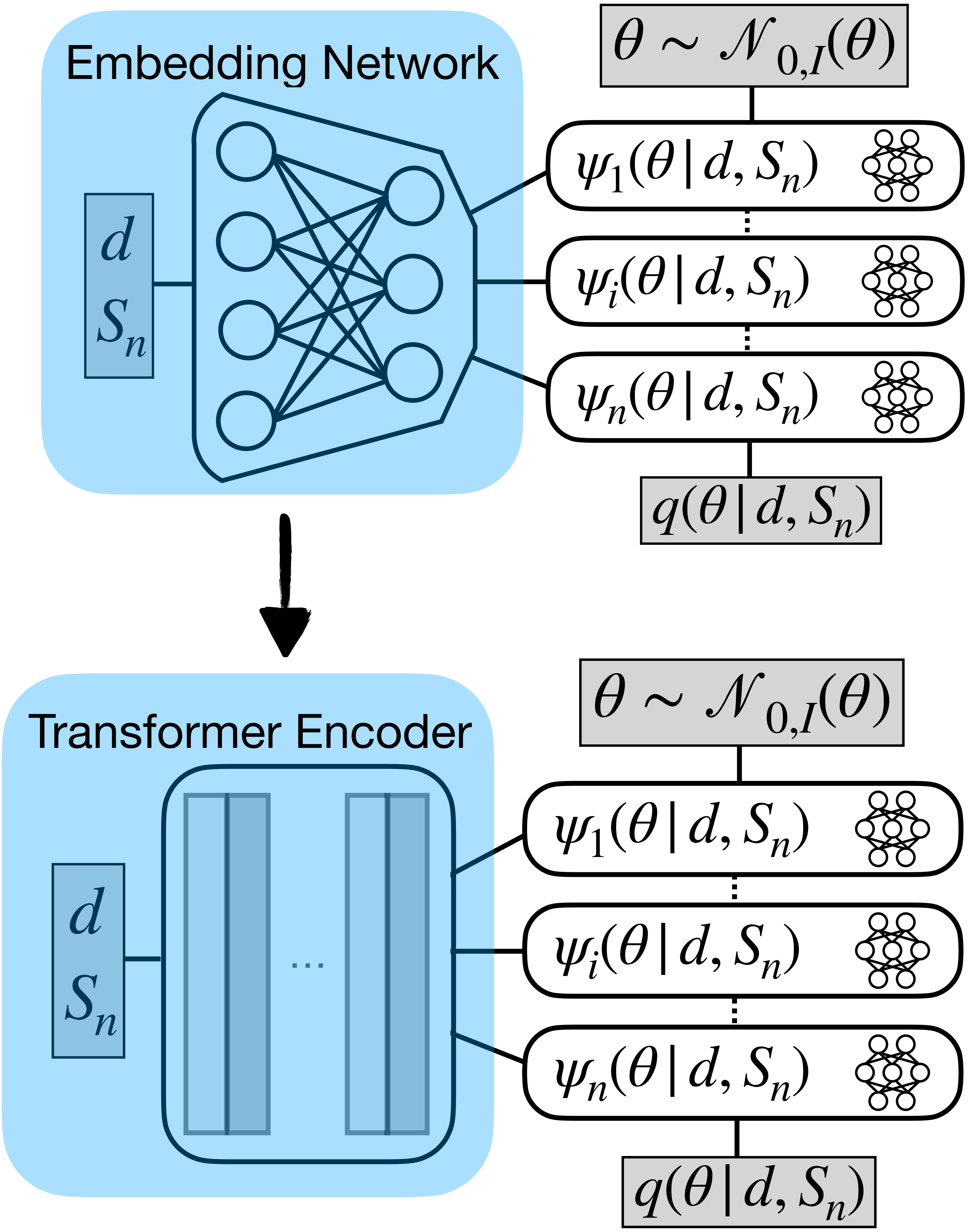


Problem:
Amortized ML approaches like DINGO cannot deal with changing inputs!

How do we make DINGO flexible?

- Replace fixed embedding network with transformer encoder[2]
- Train with signals of varying shapes
- Adjust data analysis settings at inference time

⇒ **DINGO-T1** (Transformer version 1)

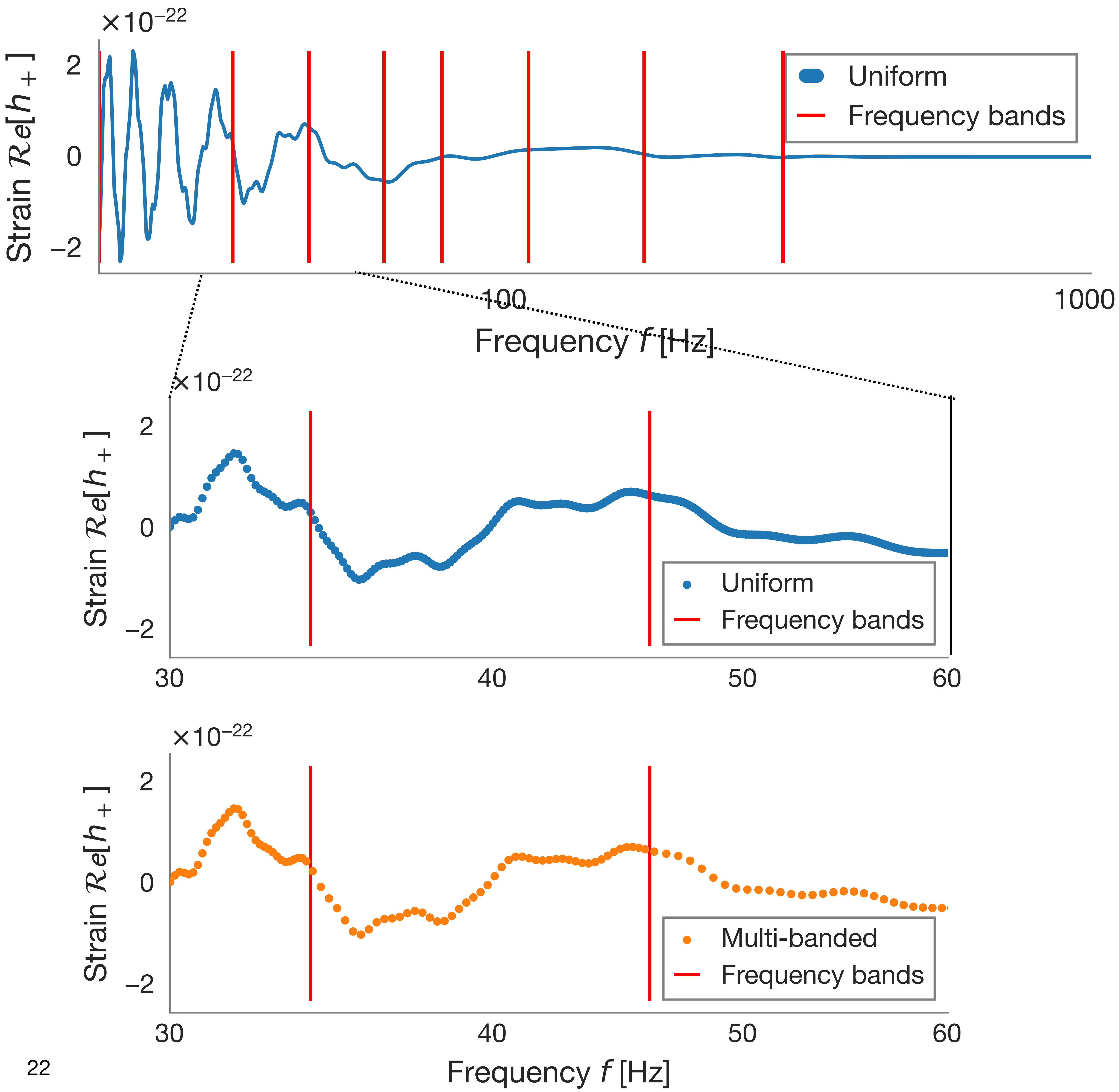


[2]Similar to Papalini+ MLST 2025

How do we build token segments?

- **Multi-banding**
  = compression based on physics domain knowledge

  → Subsample GW signals in frequency domain



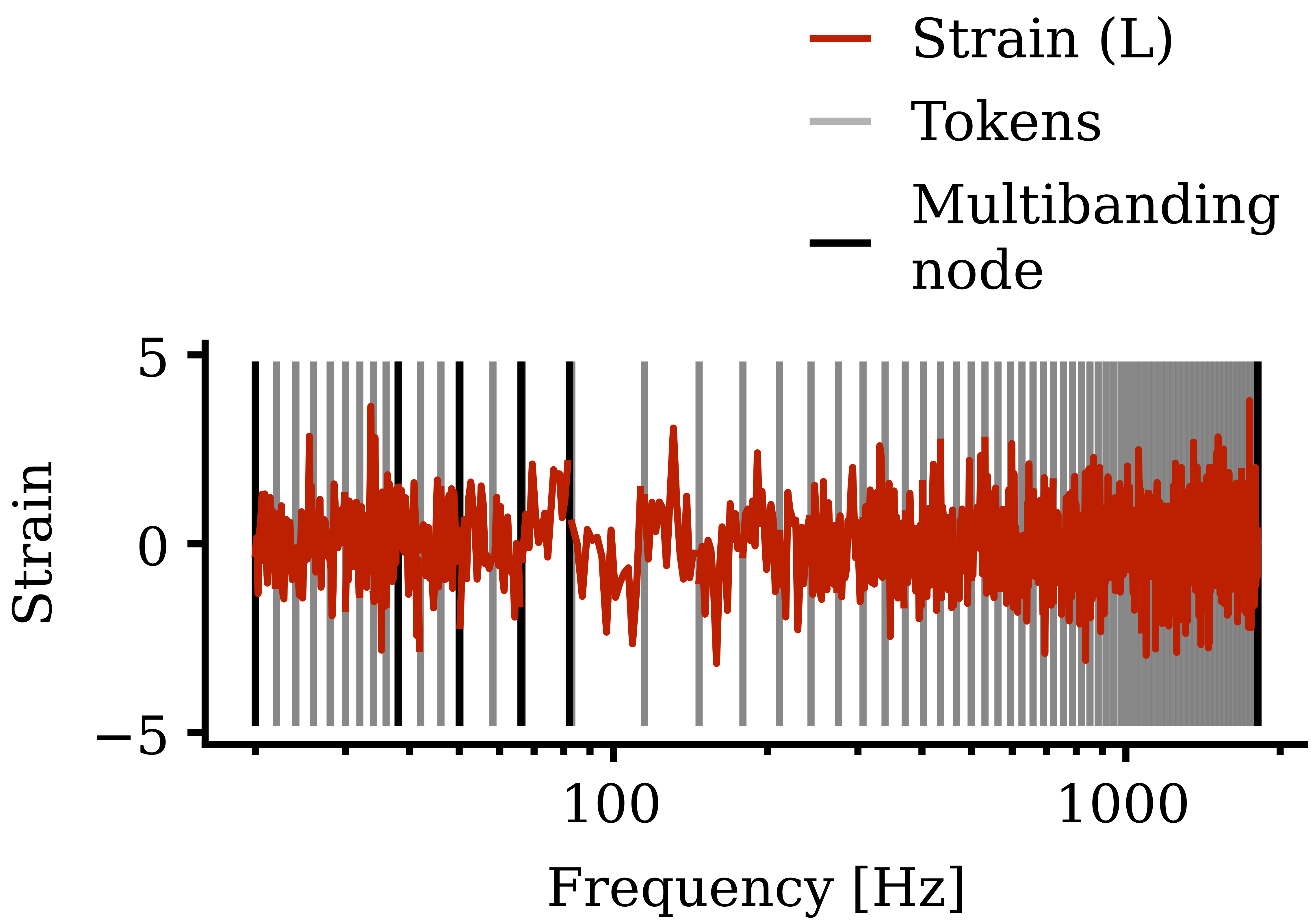
Vinciguerra+, Class. Quantum Grav., 2017

How do we build token segments?

- **Multi-banding**
  = compression based on physics domain knowledge

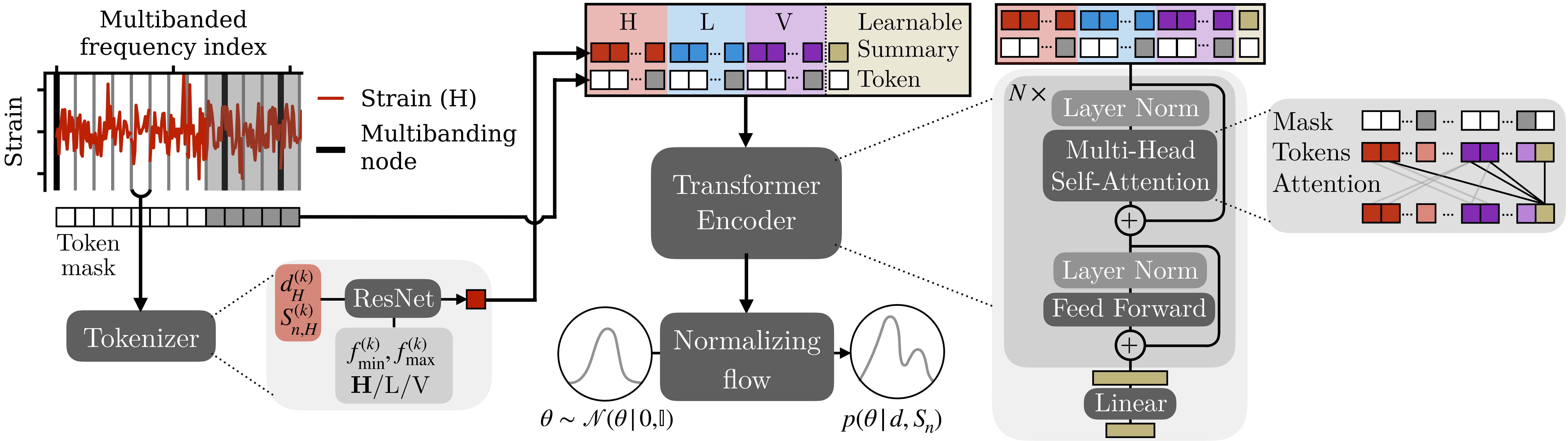
  → Subsample GW signals in frequency domain

- Divide into segments
  - Consistent resolution within a segment
  - Segments of consistent size



DINGO-T1 Architecture

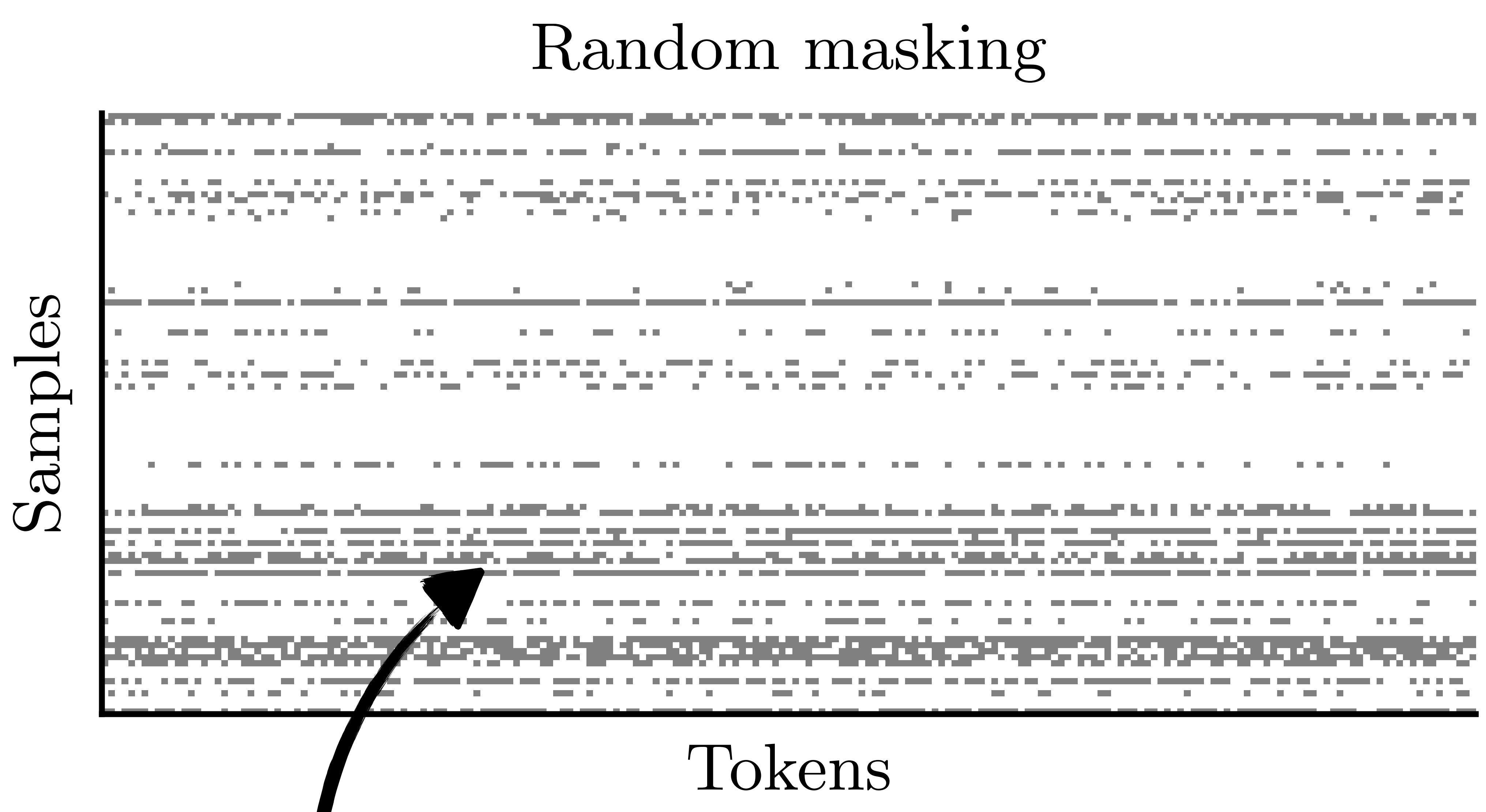
- Shared tokenizer across detector and frequencies
- Extract information via summary token
- End-to-end training



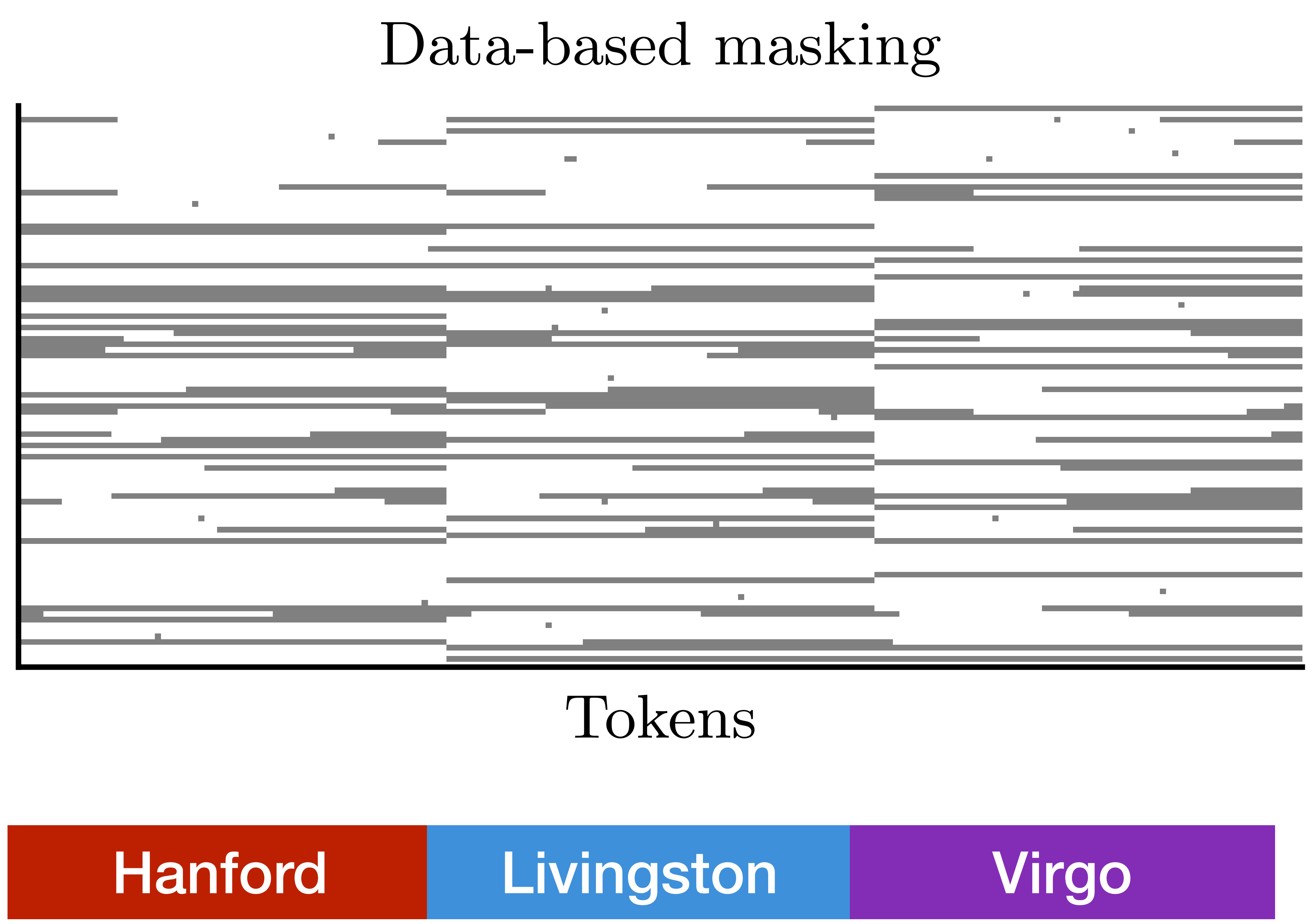
Training a flexible model

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

- Two masking strategies ($m \sim p(m)$):

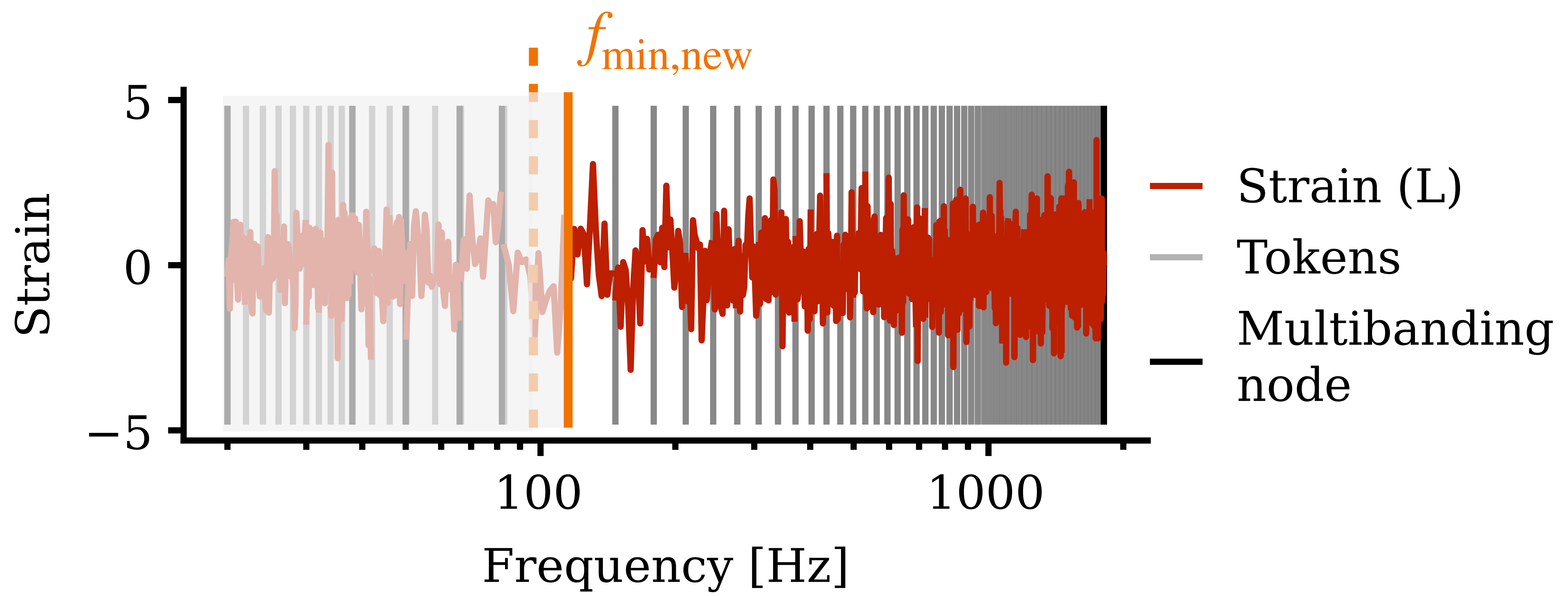


No physics domain knowledge included



Data analysis settings during inference

- Remove token inclusively by masking unwanted tokens
- Perform importance sampling on uniform frequency domain

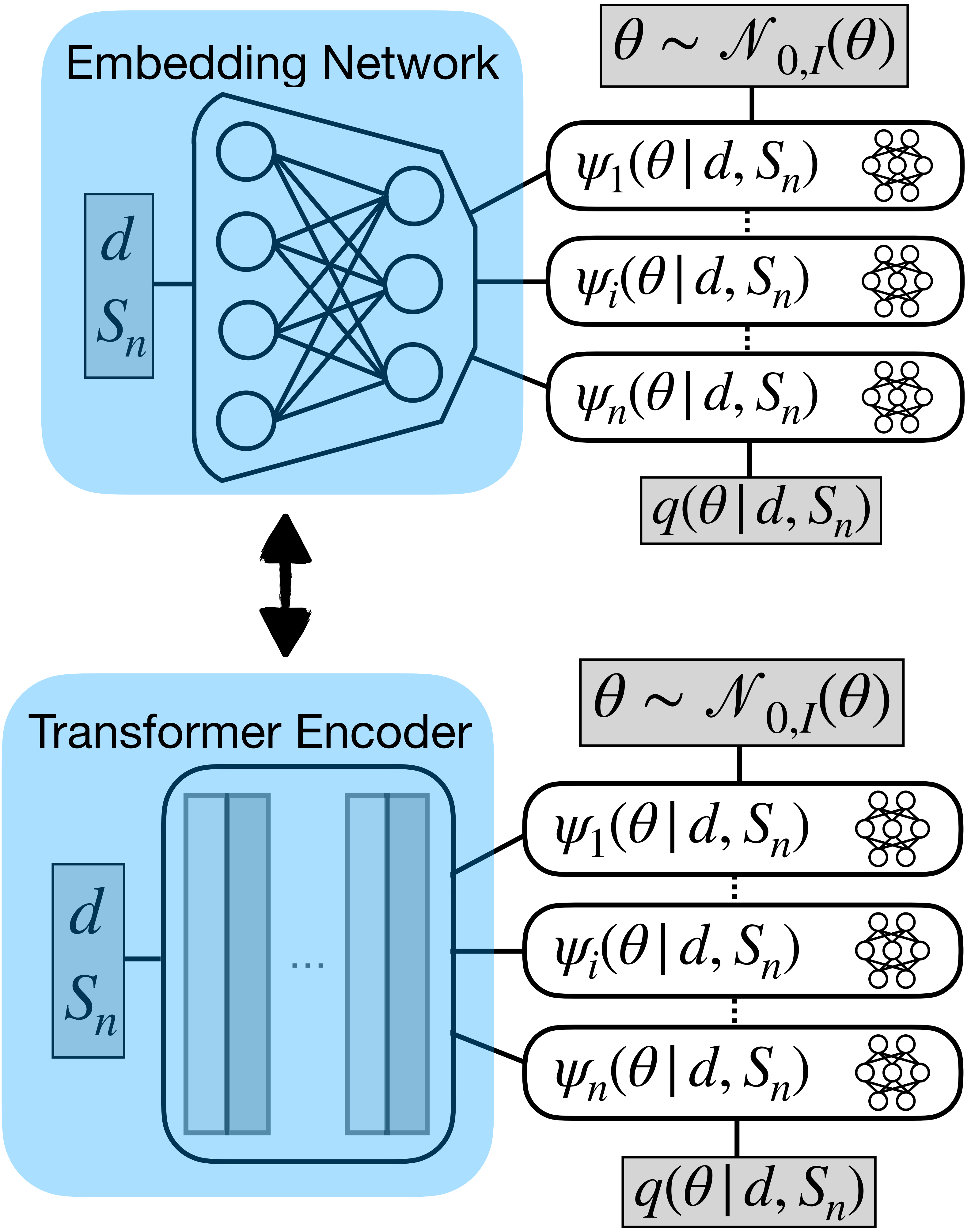


→ Result independent of multi-banding and masking

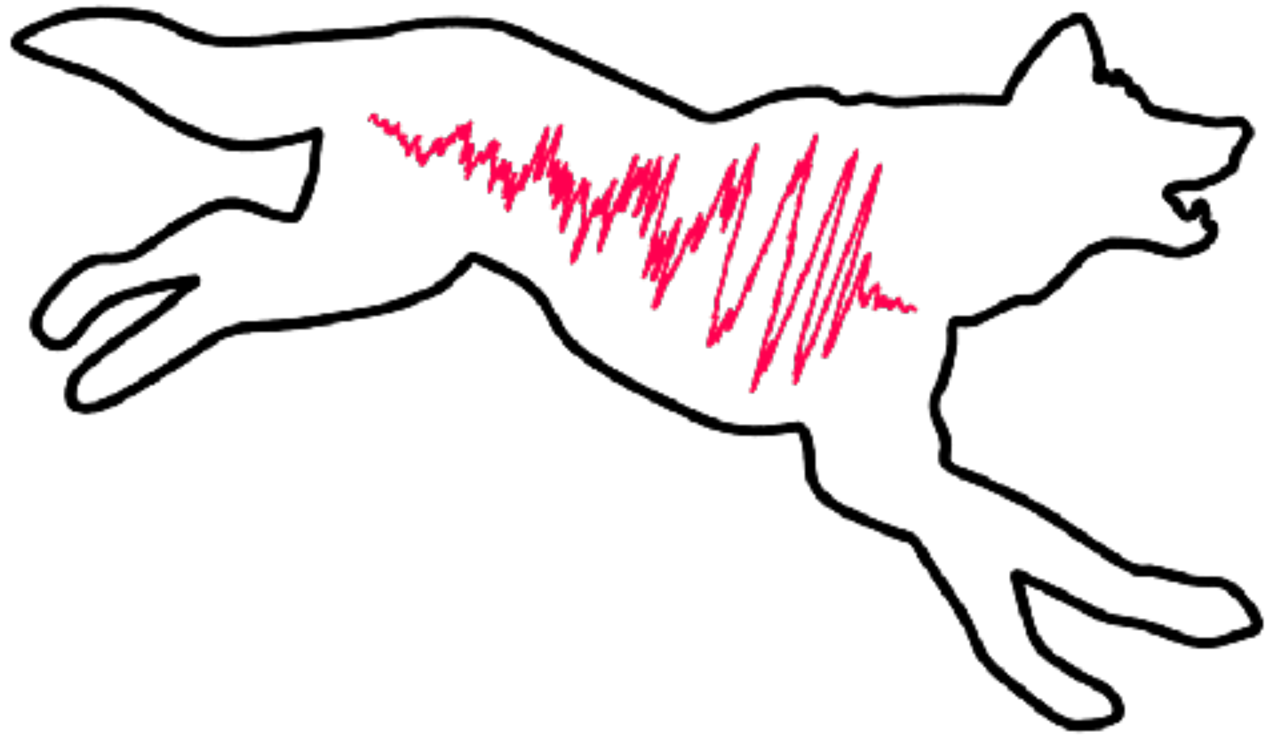
→ Correct for potential information loss

Training a flexible model

- Train with $2.5 \cdot 10^7$ IMRPhenomXPHM waveforms and $\sim 10^3$ noise curves
- Distributed data-parallel training on 8 A100 GPUs $\rightarrow \sim$ 8 days
- Baseline: NPE with ResNet embedding network



Results

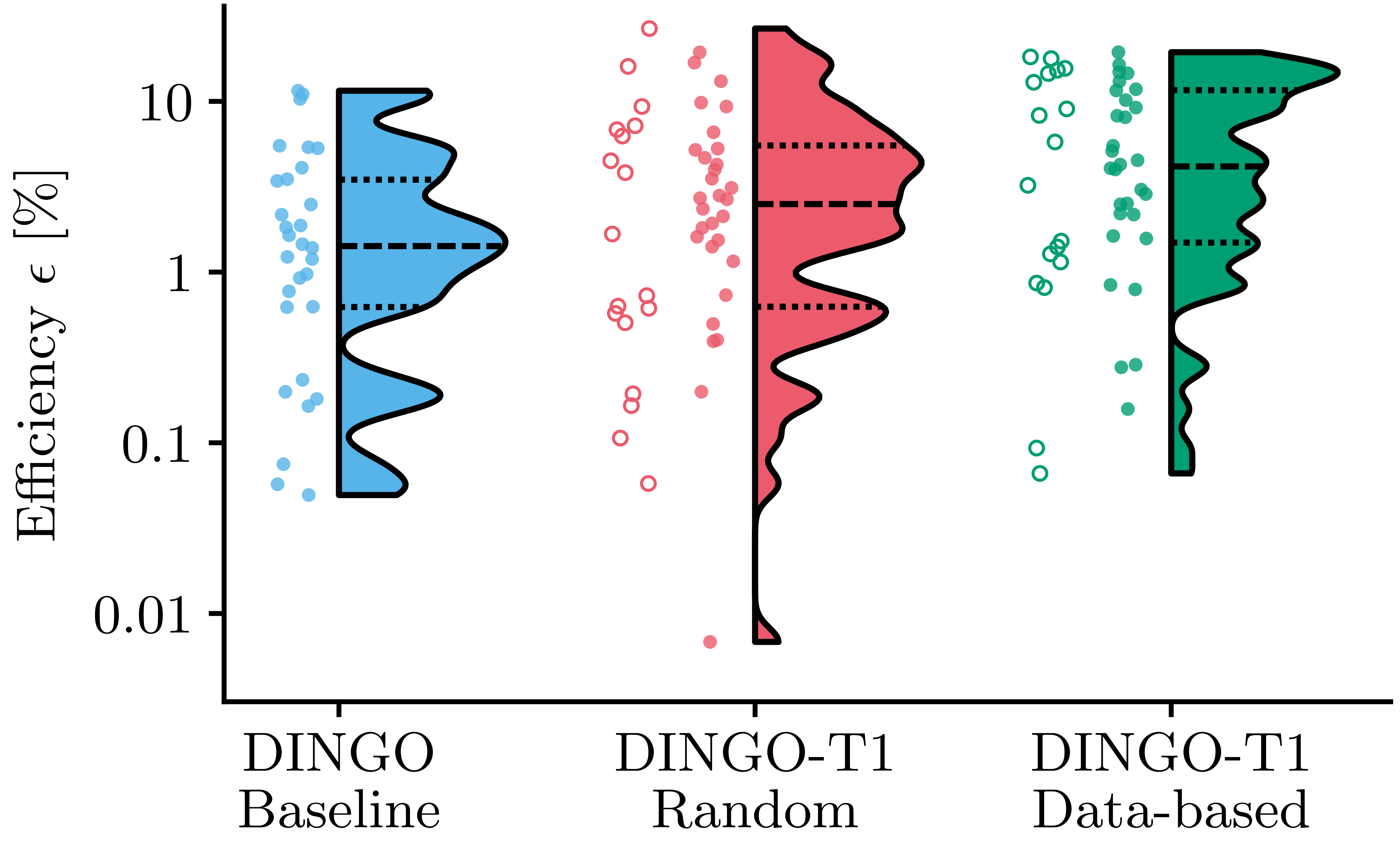


- What masking strategy works best?
- What can we do with a flexible model?

Comparing masking strategies

- Real data:
    - 48 events with 17 data analysis settings
- Distribution over sample efficiencies:

Median efficiency	2-detector	3-detector
DINGO-T1 Random	1.2%	2.7%
DINGO-T1 Data-based	**4.5%**	**4.2%**



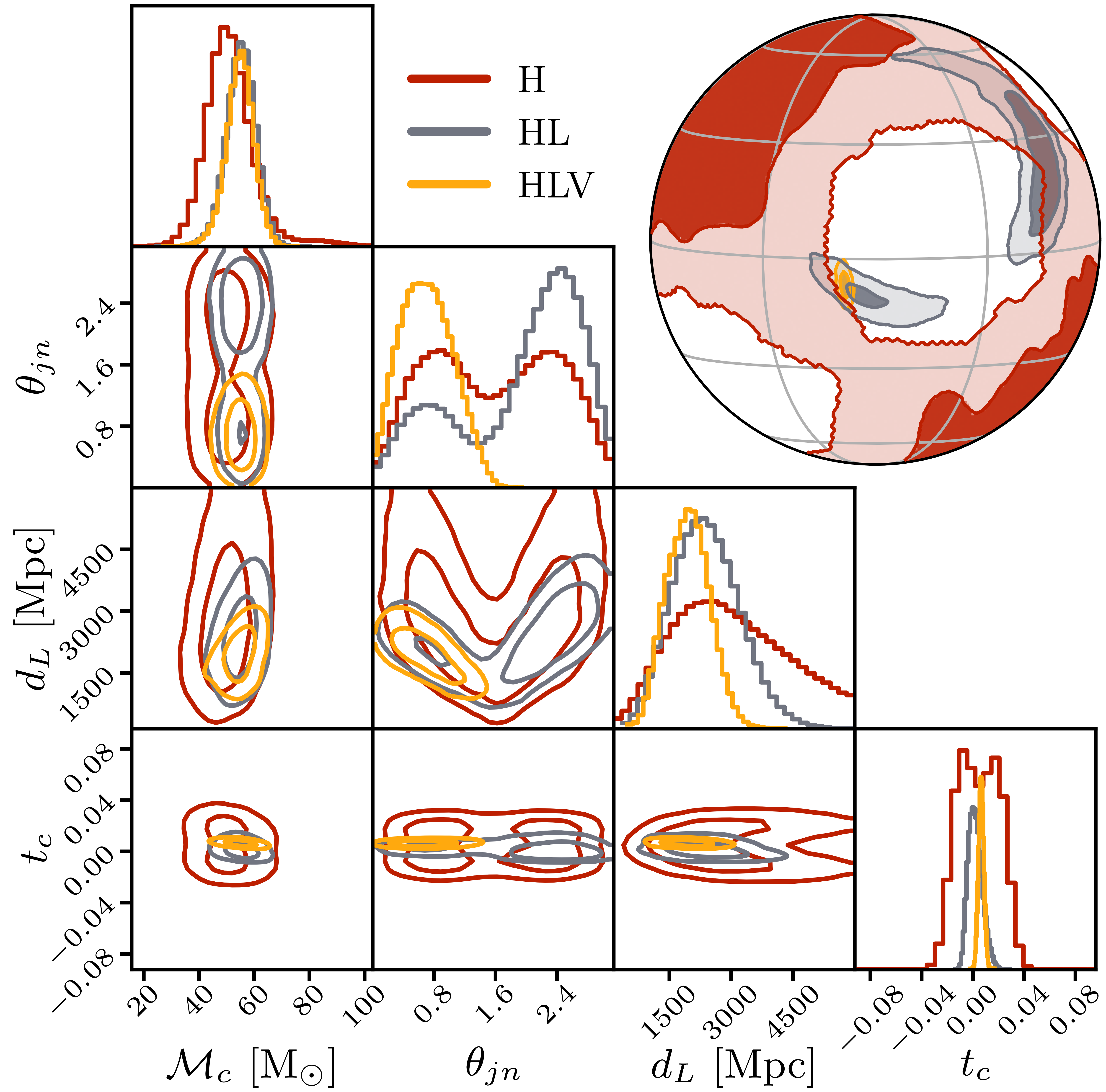
○ 2-detector events (HL, HV, LV)

● 3-detector events (HLV)

Conclusion: Data-based masking wins ⇒ DINGO-T1

Applications: Posteriors under different detectors

- How is the posterior affected when analyzing an event in different detectors?
- Sky position changes a lot (expected)



Applications: Posteriors under different detectors

- Fast inference with DINGO-T1
- Analyze all events in different detectors (264 individual PE runs)
- Computationally infeasible with standard samplers

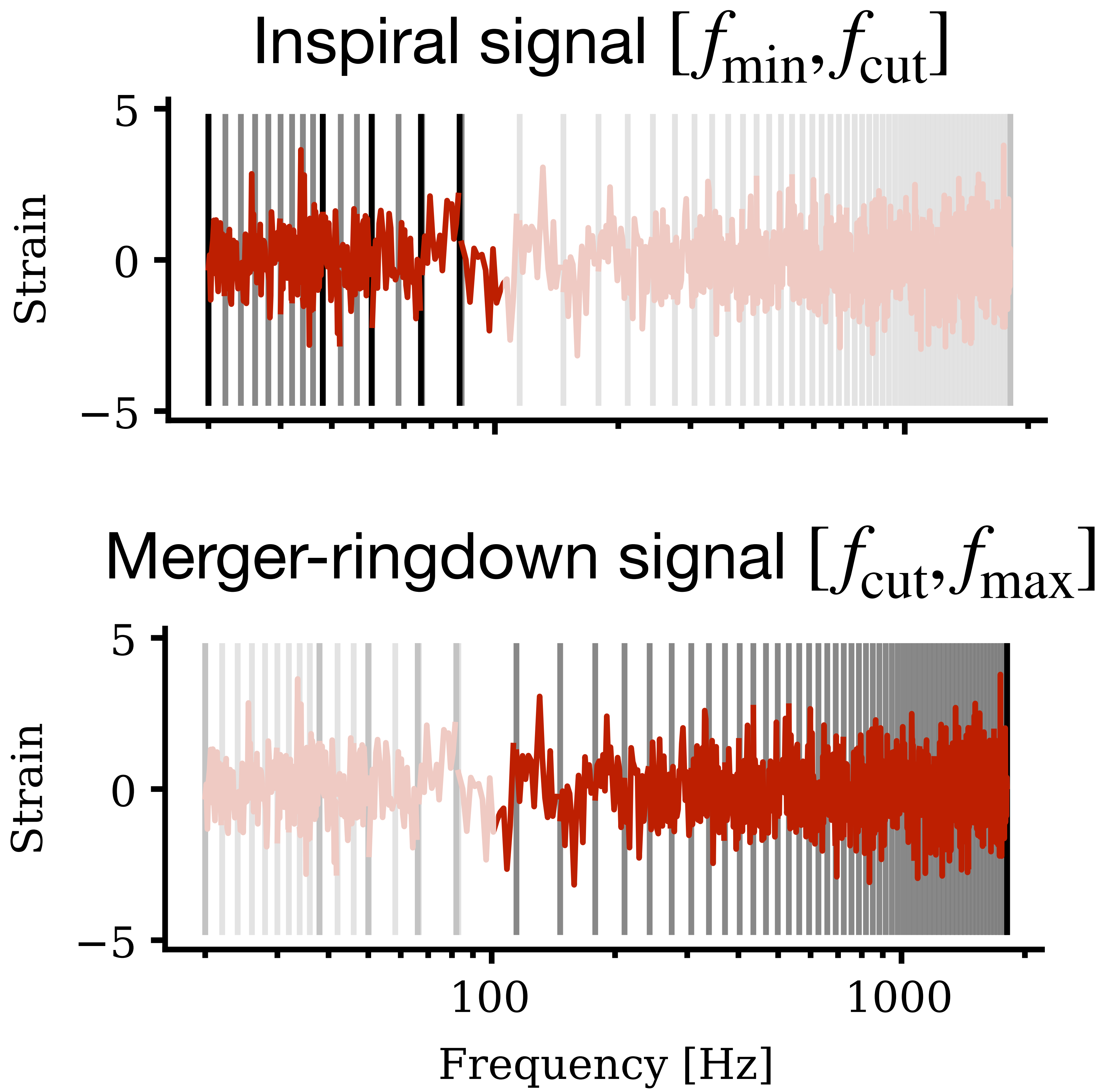
Event	Detectors	HLV	HL	HV	LV	H	L	V
GW190408_181802	HLV	8.07 %	8.28 %	0.64 %	1.35 %	2.79 %	6.71 %	62.23 %
GW190413_052954	HLV	11.77 %	9.52 %	2.41 %	0.74 %	9.5 %	18.61 %	37.48 %
GW190413_134308$^{*\dagger L}$	HLV	4.52 %	7.84 %	8.29 %	1.36 %	21.16 %	11.78 %	32.2 %
GW190426_190642	HLV	0.79 %	2.27 %	7.41 %	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^{*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^{*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 %	3.55 %	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
- Do we obtain the same M_f & χ_f from different parts of the signal?
- Quantify as deviation:

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

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



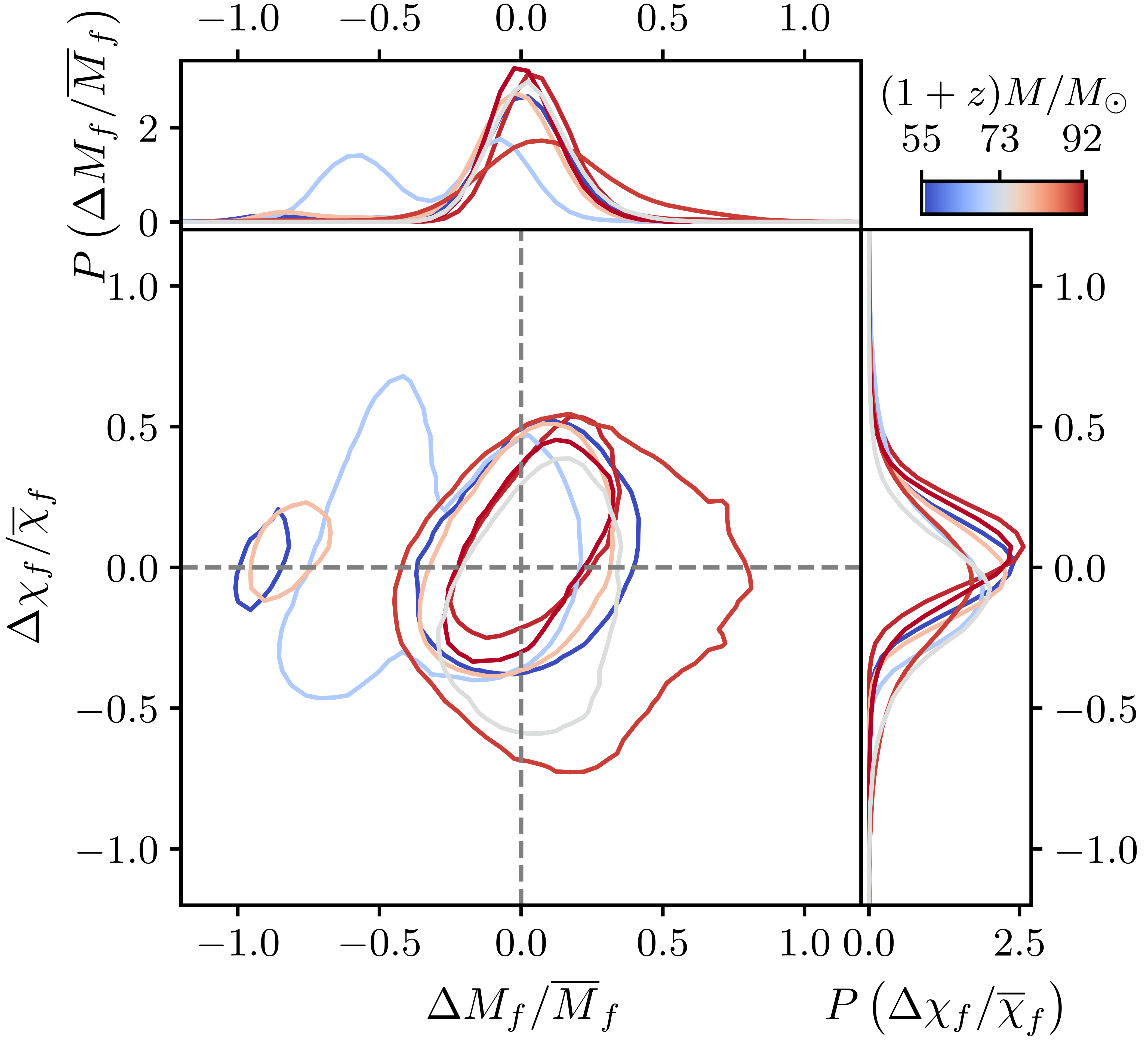
Gosh+ Class. Quantum Grav. 2017, arXiv:1704.06784

Applications: Testing general relativity

- 7 O3 events:

Event	Color	Frequency ranges [Hz] Inspiral	Frequency ranges [Hz] Postinspiral
GW190408_181802		[20, 164]	[164, 896]
GW190521_074359		[20, 105]	[105, 224]
GW190630_185205		[20, 135]	[135, 896]
GW190828_063205		[20, 132]	[132, 896]
GW200208_130117		[20, 98]	[98, 448]
GW200224_222234		[20, 107]	[107, 448]
GW200311_115853		[20, 122]	[102, 896]

- Theory consistent with data
  $\rightarrow$ peak at $(0,0)$



Summary: DINGO-T1

- Flexibility of DINGO-T1 allows us to do …

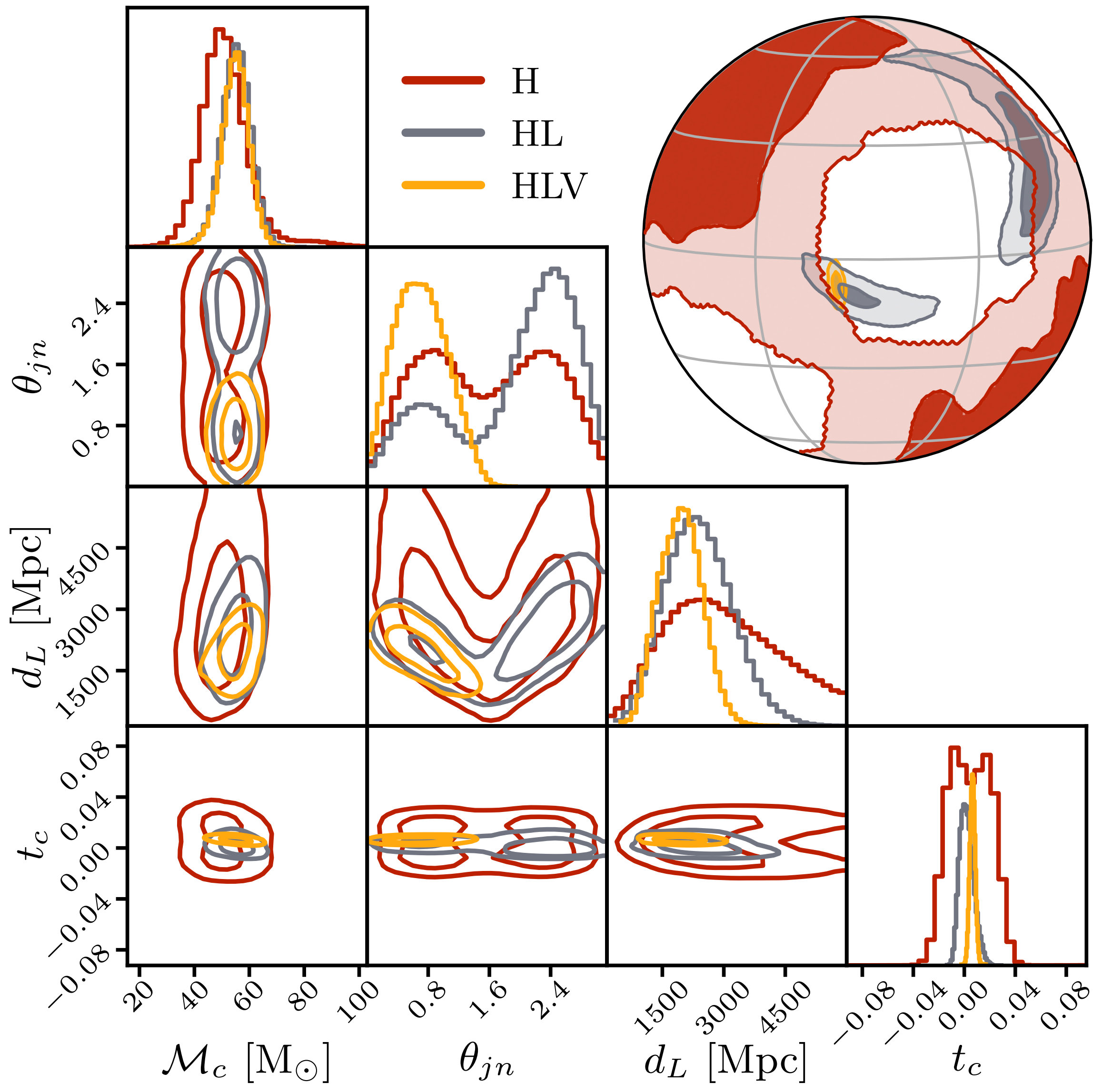
  … large scale PE analysis

  … 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: https://github.com/dingo-gw/dingo-T1



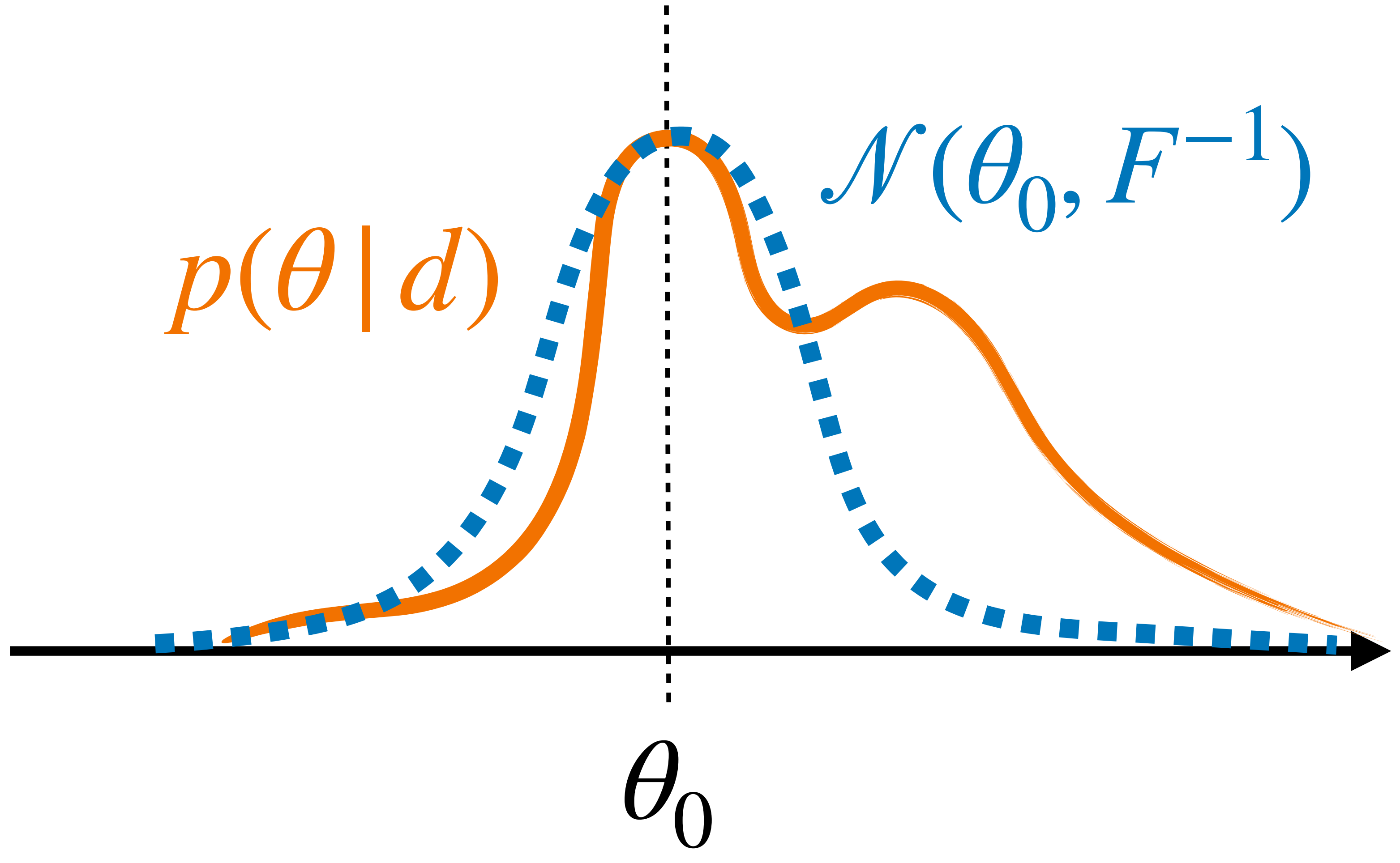
Make this plot yourself!

DINGO for the Einstein Telescope

Santoliquido+, *Fast and accurate parameter estimation of high-redshift sources with the Einstein Telescope*
arXiv:2504.21087, PRD 2025

Filippo Santoliquido

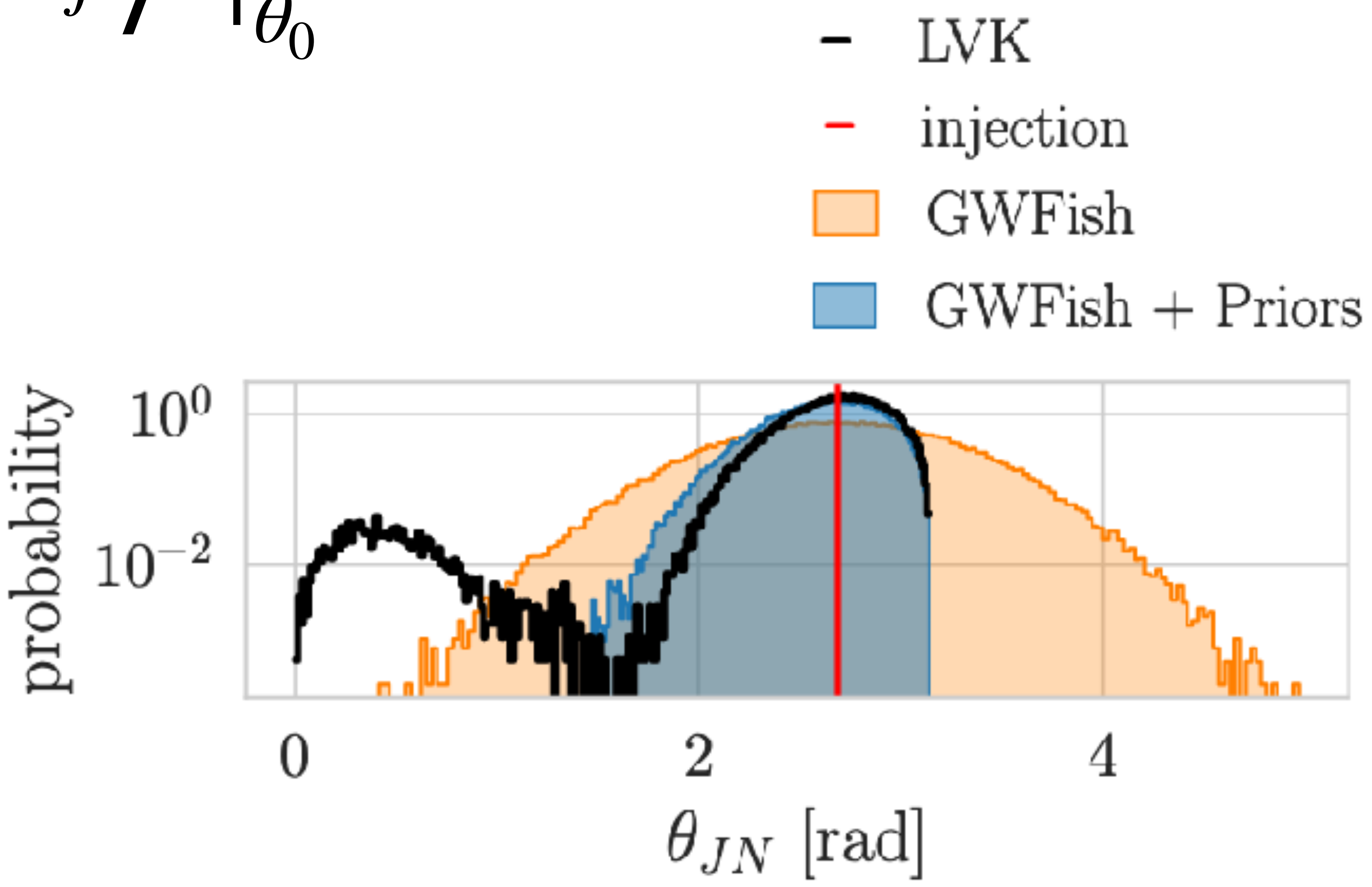
PE design studies for ET



- Based on **Fisher Information Matrix (FIM)** approximation
  - Multi-variate Gaussian posterior

$$p(\theta\,|\,d) \approx \mathcal{N}(\theta_0, F^{-1}) \quad \text{with} \quad F_{ij} = \left\langle \frac{\partial h}{\partial \theta_i}, \frac{\partial h}{\partial \theta_j} \right\rangle \Bigg|_{\theta_0}$$

- Ideally: Full posterior distribution
- Influence of FIM approximation on ET estimates?



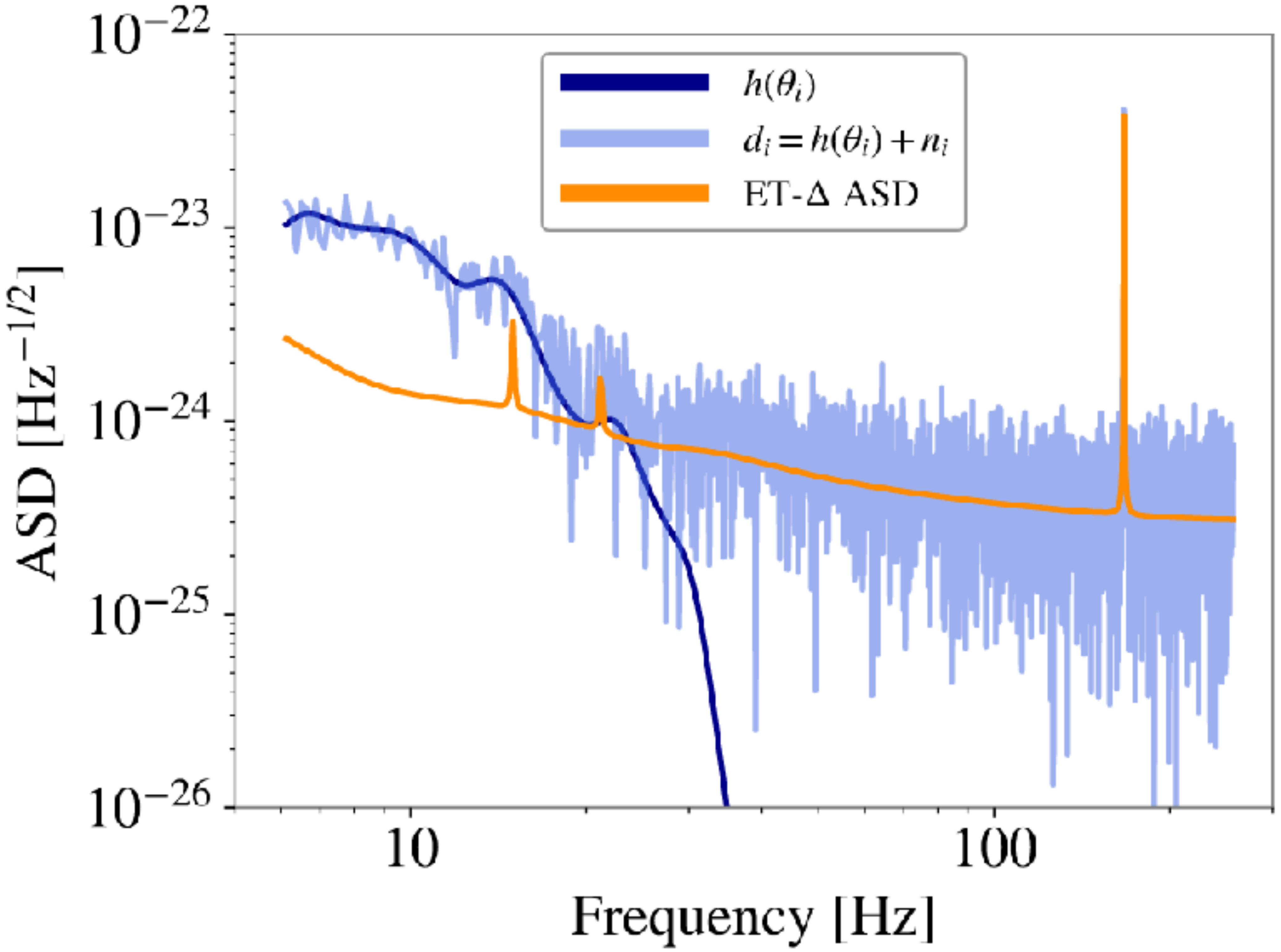
Dupletsa+, PRD 2025, arXiv:2404.16103

Which setting is most interesting?

- Moderate to high-redshift sources ($1 \leq z \leq 45$)
  $\rightarrow$ Broad posterior, FIM unreliable Vallisneri+, PRD 2008
  $\rightarrow$ Short signals

- Frequency range: $6 - 256$ Hz

- NPE DINGO architecture

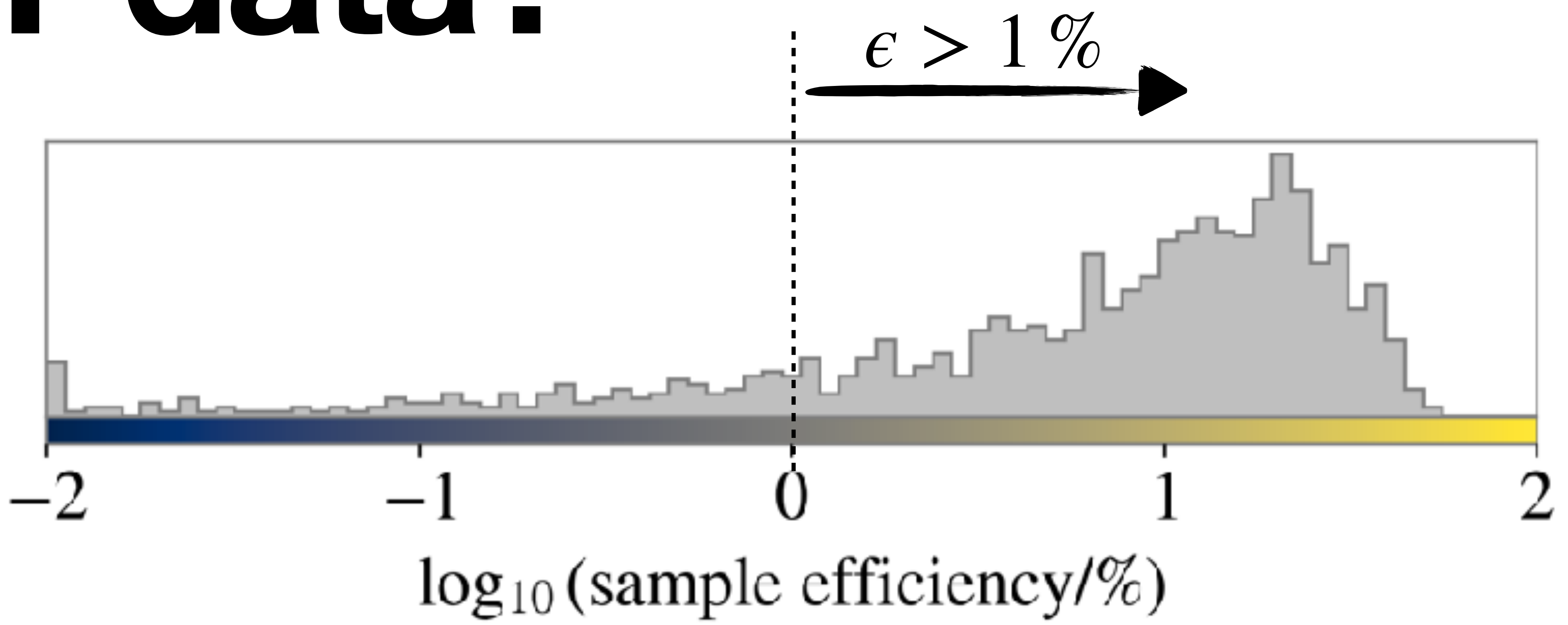


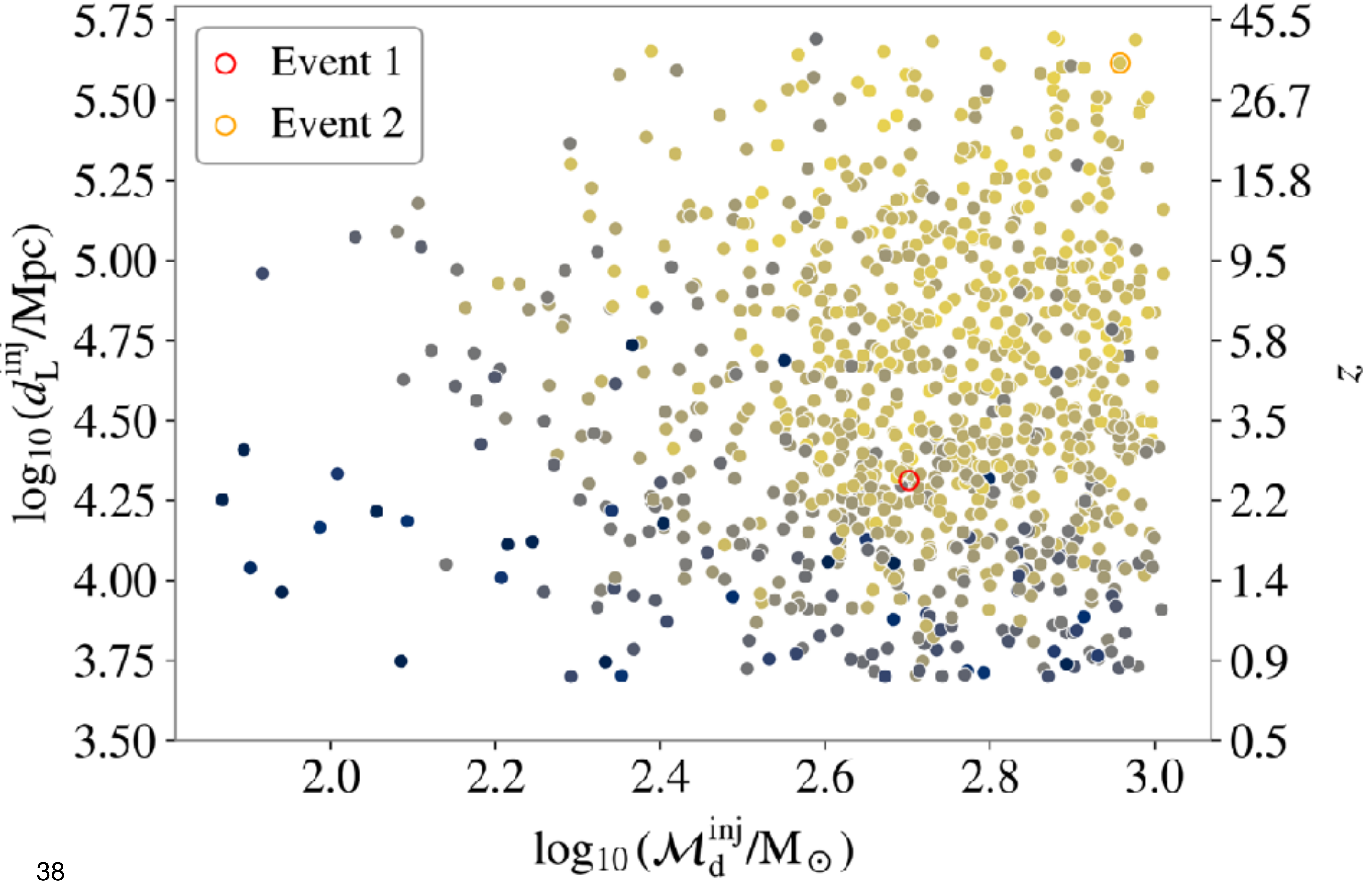
Does DINGO work on ET data?

- 1000 BBH simulations
- Lower distances / red-shifts
  → more difficult to learn
- Reduced energy cost:

  DINGO: ~1000kWh for training & inference
  Bilby: ~4000kWh

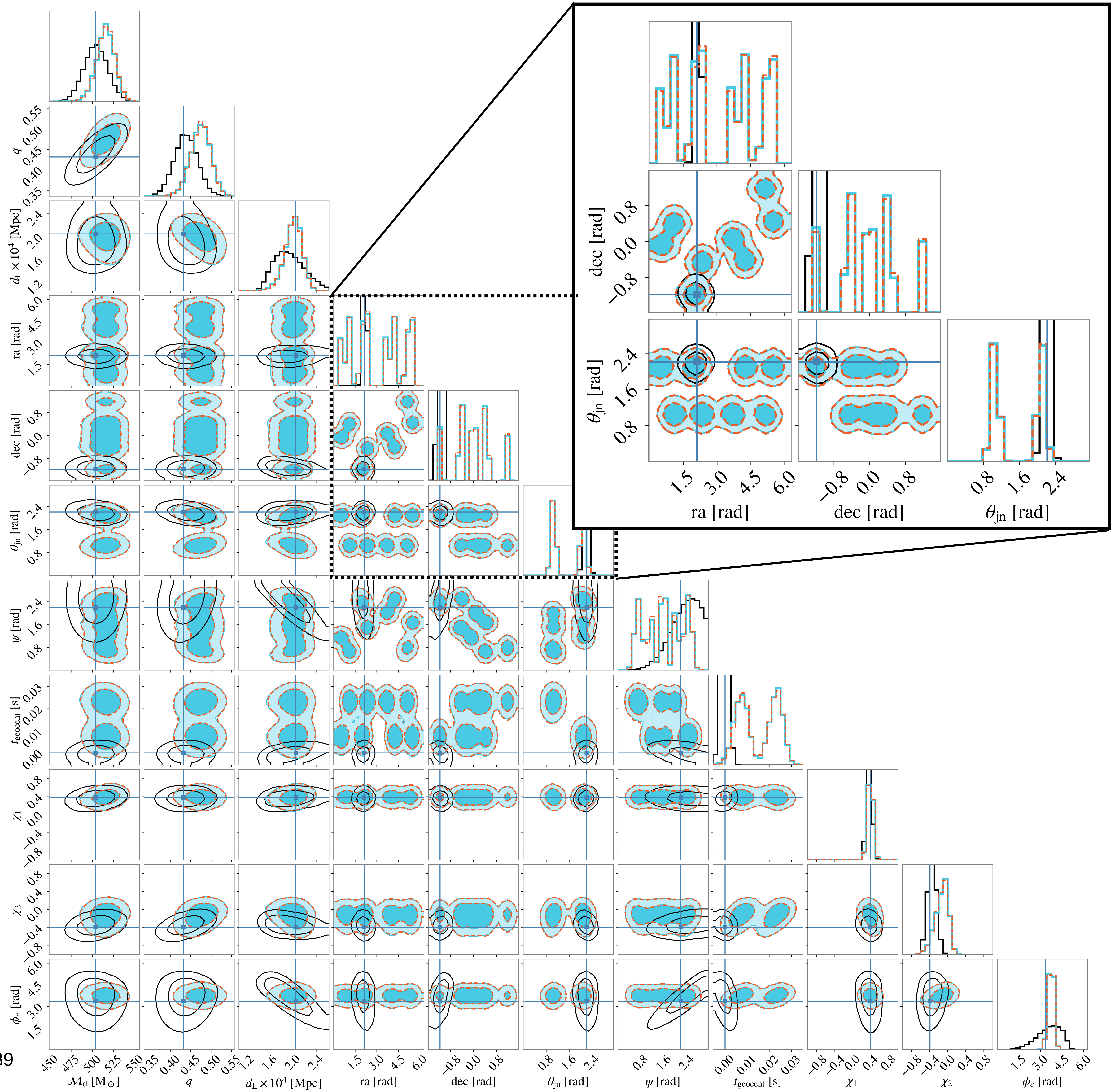
- It works!





FIM vs. full PE

- Compare:
  - Dingo-IS
  - Bilby
  - GWFish+Priors
- FIM misses multi-modality of sky position



Teaser: Comparing detector configurations with NPE

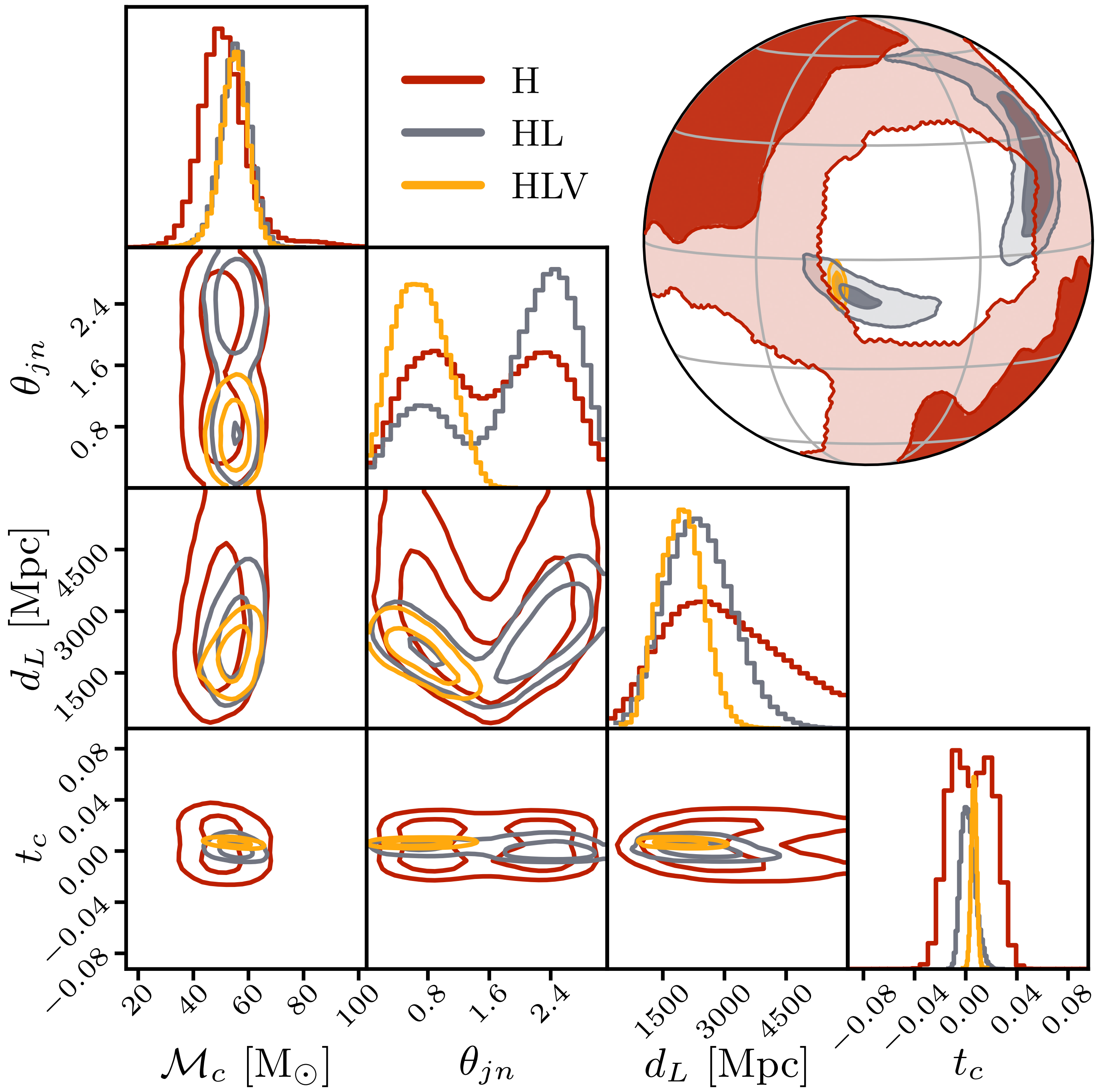
Santoliquido+, *Comparing next-generation detector configurations for high-redshift gravitational wave sources with neural posterior estimation*
arXiv:2512.20699

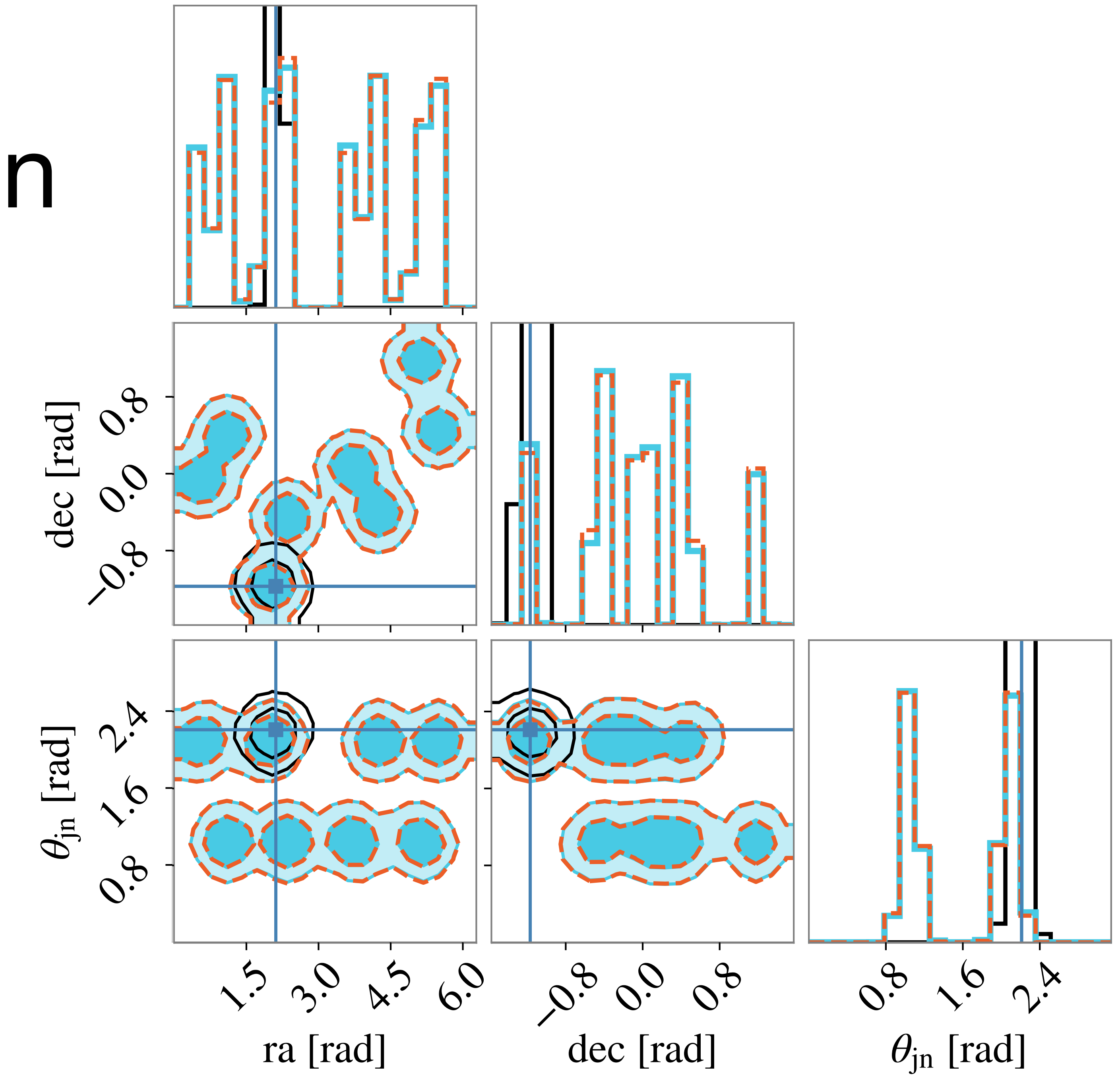
$\Delta\Omega_{90\%}$	$< 10\ \mathrm{deg}^2$	$< 100\ \mathrm{deg}^2$	$< 1000\ \mathrm{deg}^2$
Δ	0.7%	8.5%	26.2%
2L A	0.3%	4.4%	24.3%
2L MisA	3.6%	15.0%	34.8%
2L MisA + LHI	9.5%	29.6%	50.9%
1L + CE	17.3%	42.2%	69.7%
Δ + CE	31.0%	68.0%	93.2%
2L MisA + CE	36.7%	71.5%	94.2%

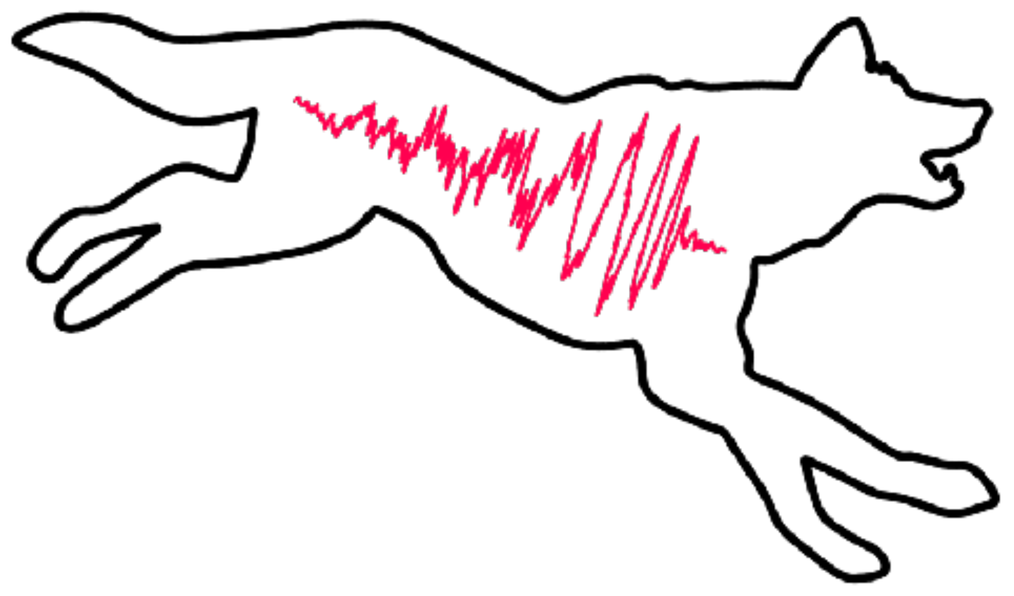
Filippo Santoliquido

Summary

- DINGO is a NPE framework for GW posterior estimation
- Train once, evaluate on many signals
- **Transformer**: Flexibility required for missing data
- **Application to ET:** improved estimates on sky localization







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

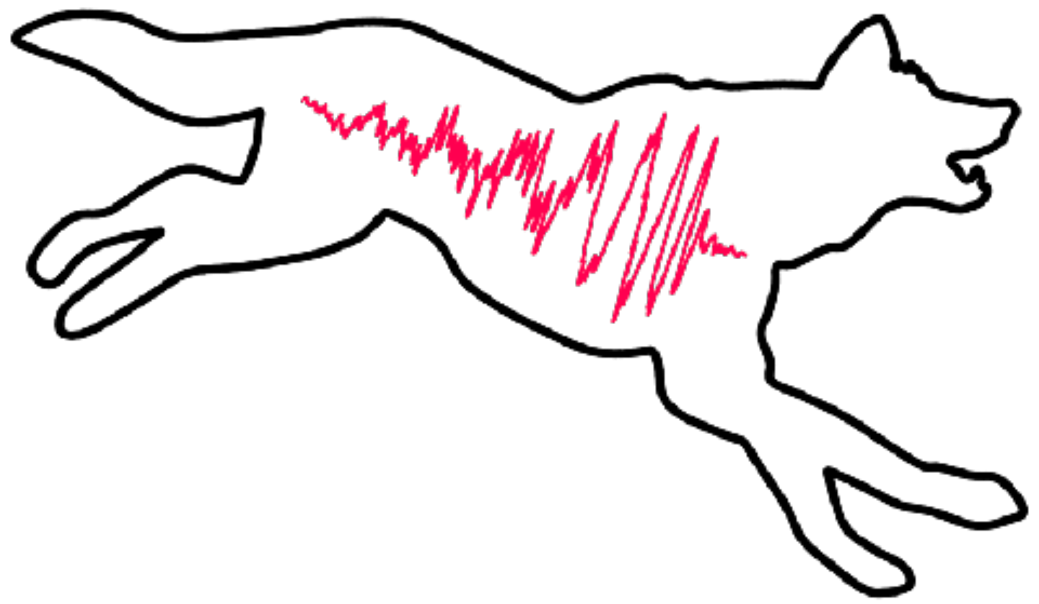

Jonathan Gair

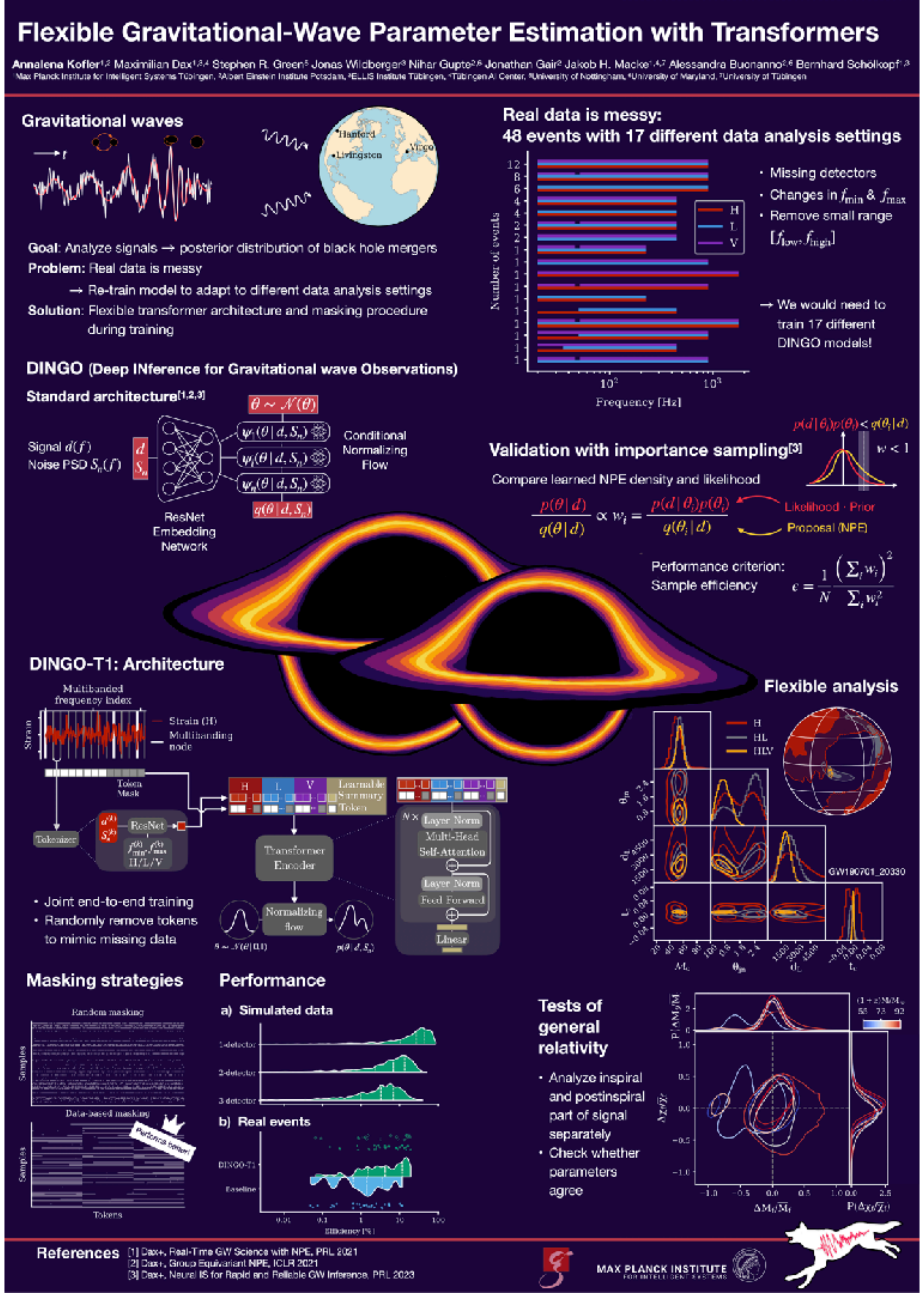

Jakob Macke


Bernhard Schölkopf


Alessandra Buonanno

Do you have any questions?





References

- Dax+, Real-Time Gravitational Wave Science with Neural Posterior Estimation. PRL 127, 2021
- Dax+, Neural Importance Sampling for Rapid and Reliable Gravitational Wave Inference, PRL 130, 2023
- Wildberger+, Flow Matching for Scalable Simulation-Based Inference, NeurIPS 2023
- Gupte+, Evidence for eccentricity in the population of binary black holes observed by LIGO-Virgo-KAGRA, arXiv:2404.14286v1, 2024
- Dax+, Real-time Gravitational-Wave Inference for Binary Neutron Stars using Machine Learning, Nature, 2025
- Kofler+, Flexible Graviational-Wave Parameter Estimation with Transformers
- Santoliquido+, Fast and accurate parameter estimation of high-redshift sources with the Einstein Telescope, arXiv:2504.21087, PRD 2025
- Santoliquido+, Comparing next-generation detector configurations for high-redshift gravitational wave sources with neural posterior estimation
  arXiv:2512.20699