Simulation-based inference for gravitational-wave science

Konstantin Leyde, 21/5/2025 @Bayesian Deep Learning for Cosmology and Astrophysics

Plan: simulation-based inference (SBI) in GW science

Introduction

- What are **gravitational waves** (GWs)?
- **GW measurement**
- **Sources and their physics**

SBI for GWs

- **Fast inference for binary black holes and neutron stars**
  - **Key ideas for speed-up of** $\times 1000$
- **Hierarchical Bayesian inference problem**

Gravitational-wave astronomy: Introduction

- **Framework: General Relativity**
- **Definition**
  - **Perturbations of space-time**, i.e. gravitational field
- Propagate at **speed of light**
- **Very weak,** $\mathcal{O}(10^{-21})$

➡Need **high velocities, large masses**

➡**Sources: binaries of black holes or neutron stars**

Metric perturbations, aka GWs

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}$$

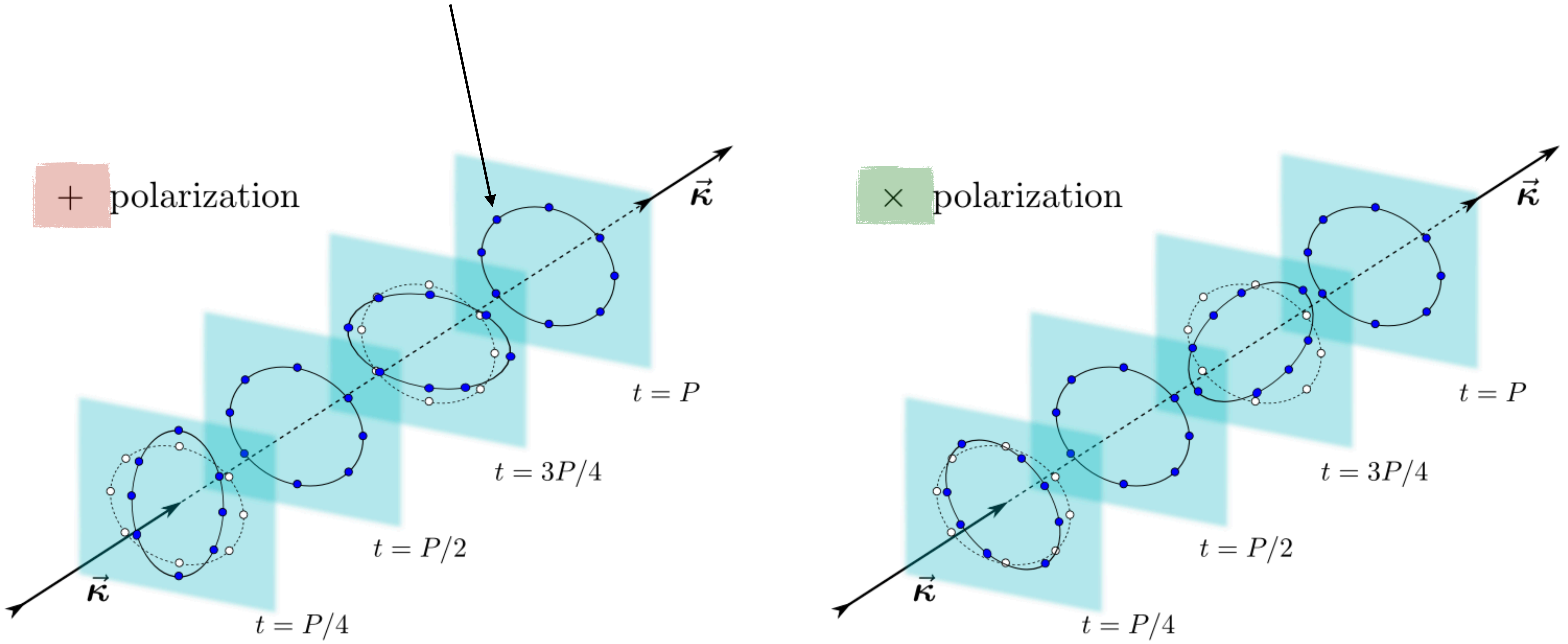
Metric

Flat spacetime

Observable effect (far from source)

Bishop and Rezzolla, 2016

- Effect on ring of **free-falling test particles**



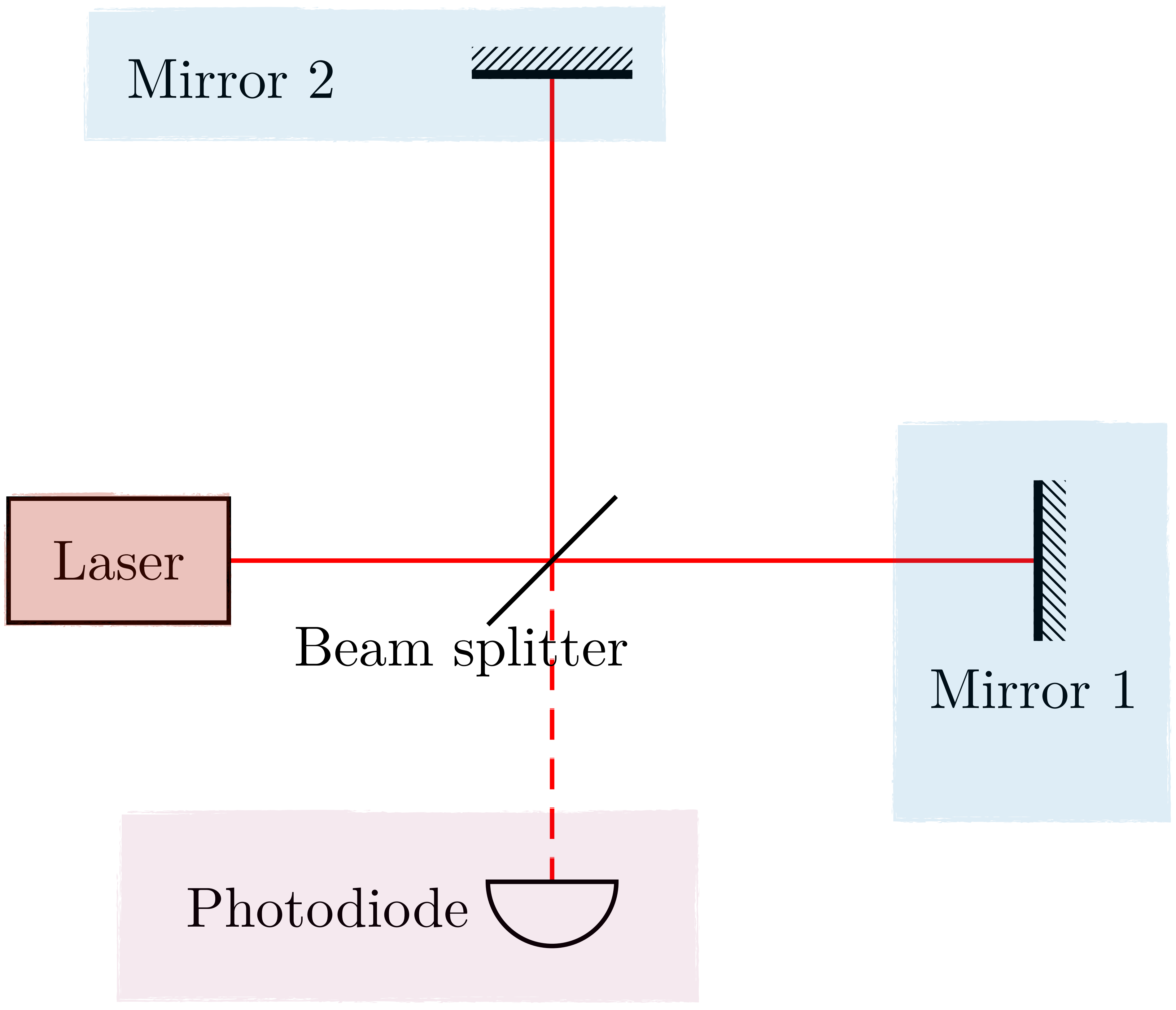
- Experiments to **measure distance to high precision**!

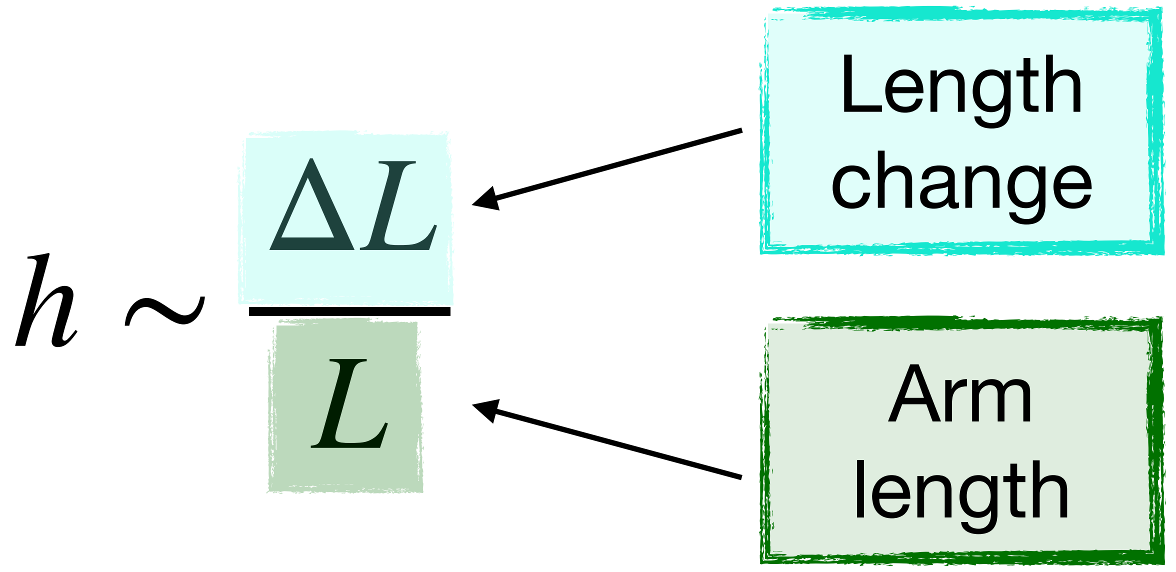
$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}$$

$$h_{\mu\nu} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & h_+ & h_\times & 0 \\ 0 & h_\times & -h_+ & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \cos(2\pi f_{\mathrm{GW}}(t - z/c))$$

GW detection

- **Km-sized Michelson interferometer**
  - **Laser cavity** in each arm
  - **Free-falling mirrors** in two-dimensions, isolated from ground motion
  - **Mirror displacement** $\rightarrow$ **Interference pattern shift**
- **Detector output: $h(t)$ relative arm length time series**
- True instrument ∞-more complex

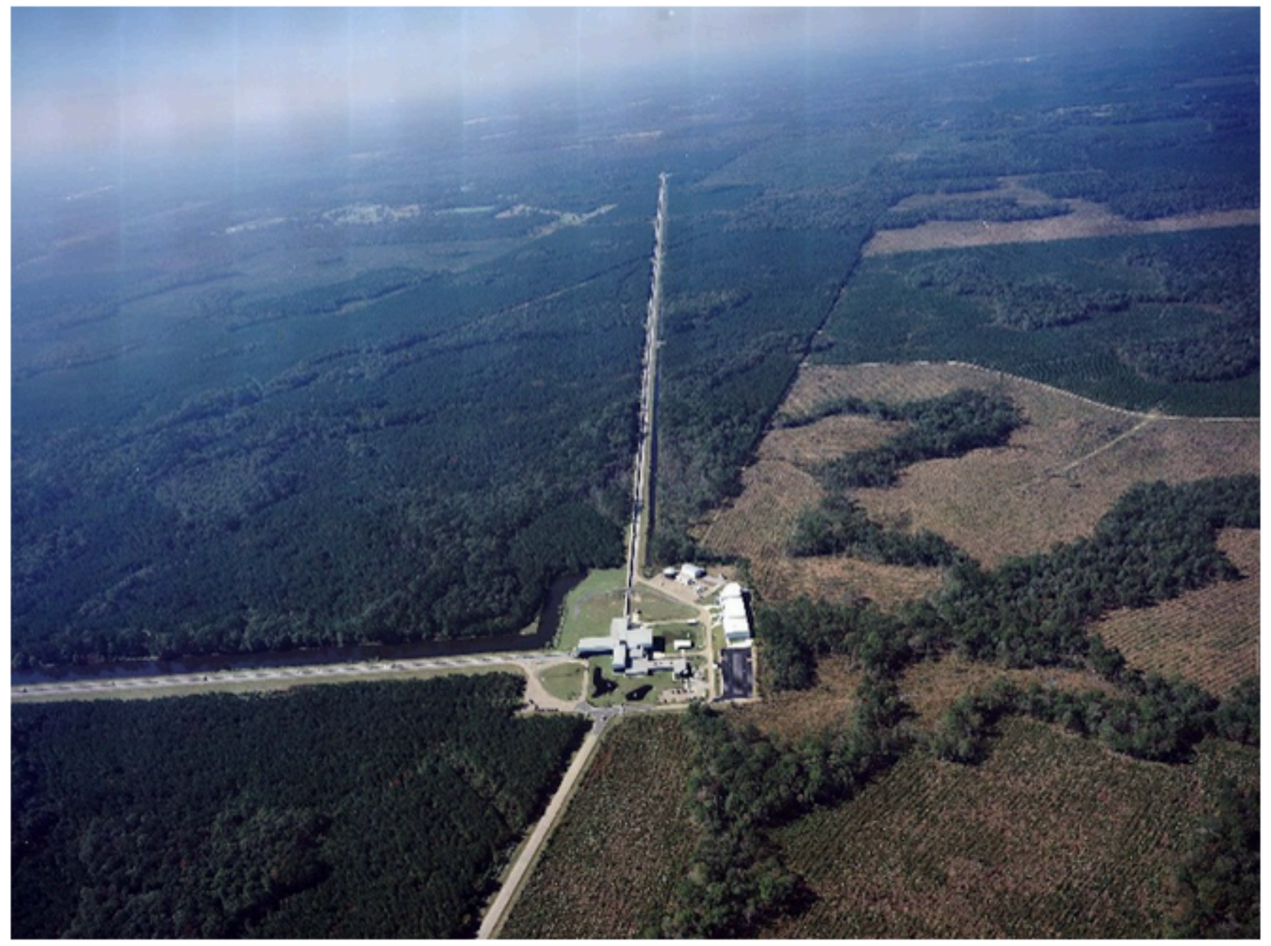




Current ground-based detectors

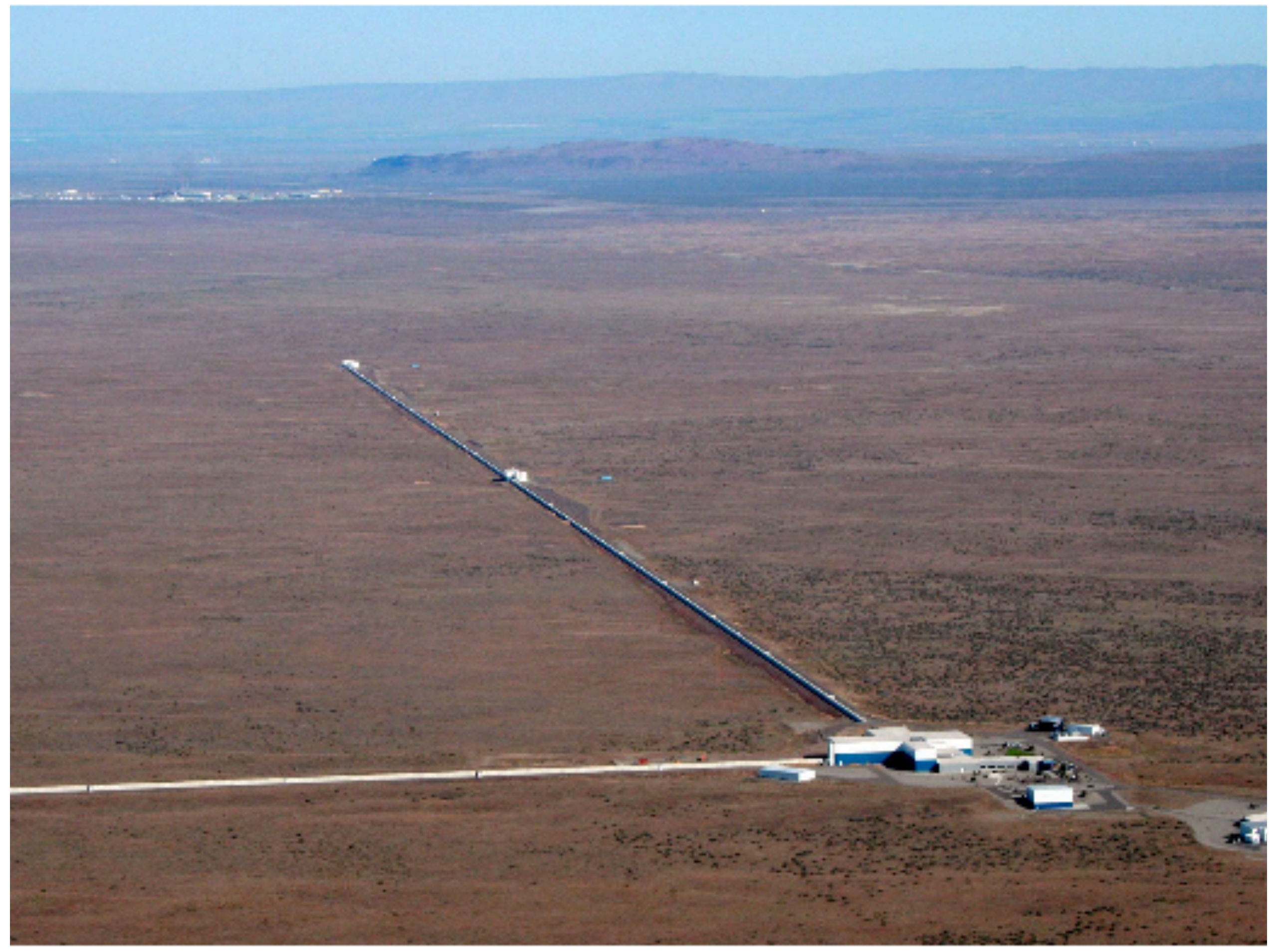
- Four detectors
  - **LIGO** (2 USA)
  - **Virgo** (1 Italy)
  - **KAGRA** (1 Japan)
- On-going: **“O4c”-run**
- **Upgrades planned**

LIGO

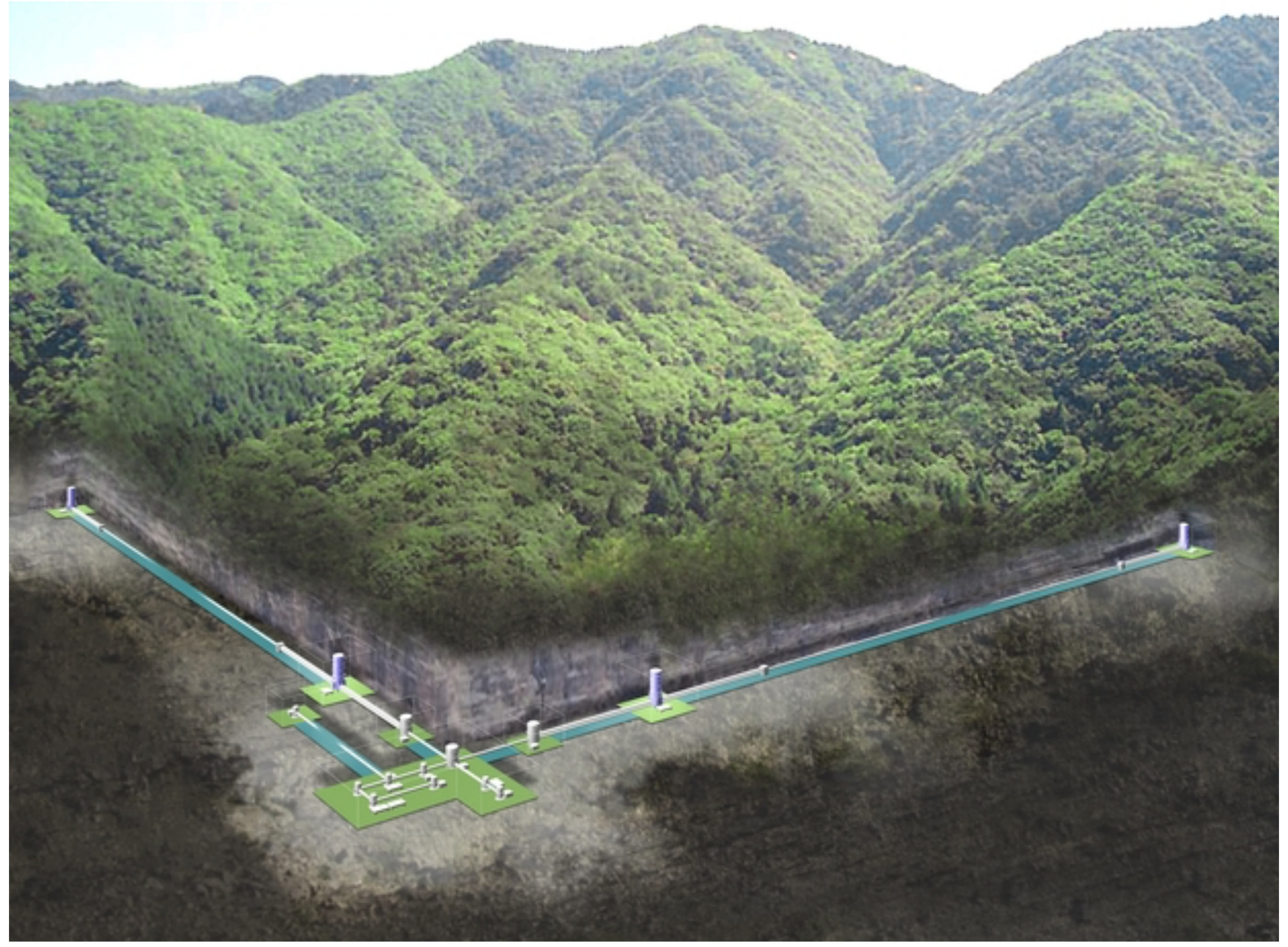


Virgo

LIGO



KAGRA

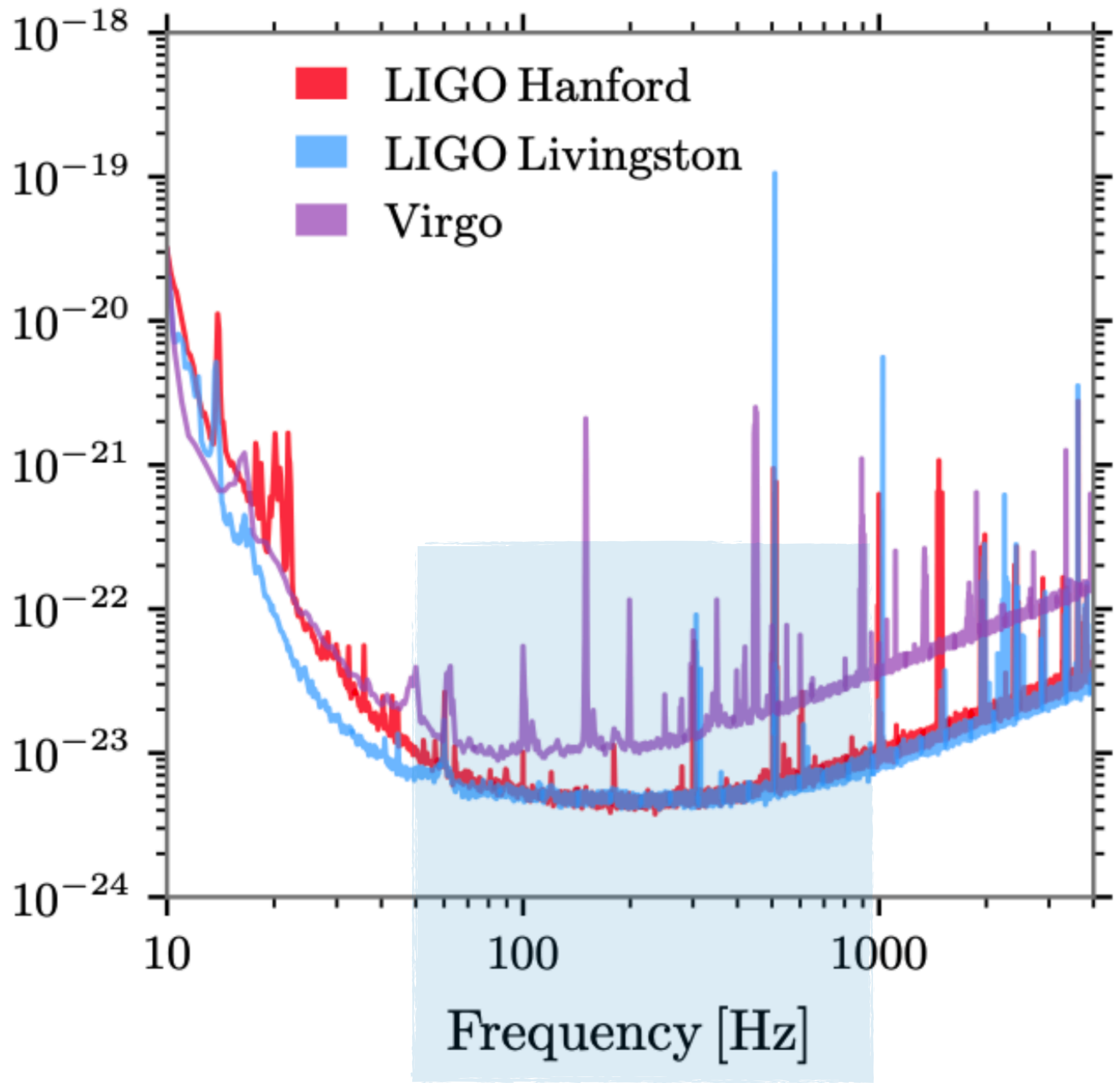


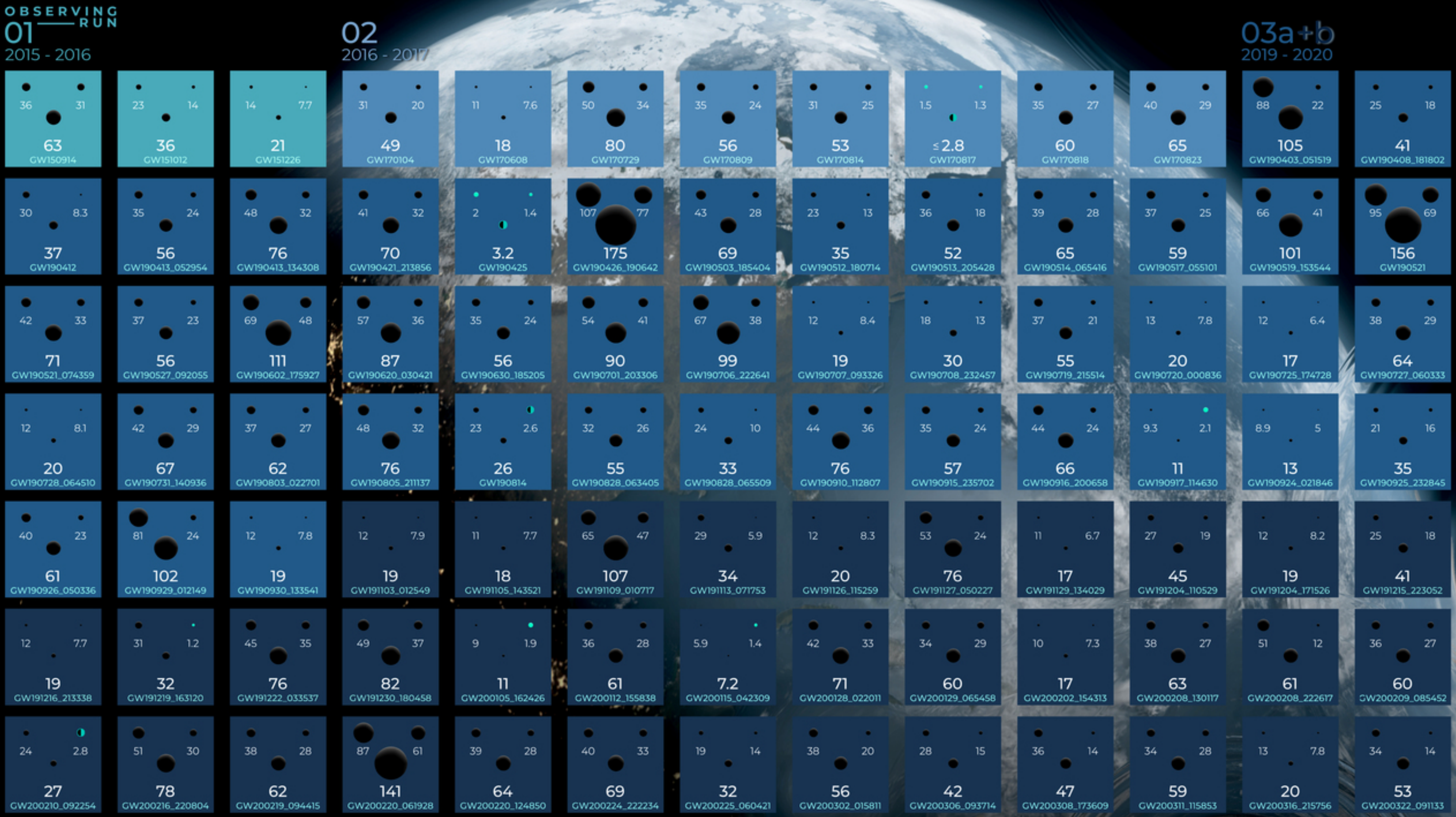
Detector noise properties

- **Noise properties**: average over realizations
  - Work in **Fourier space**
  - $\langle n(f) n(f') \rangle = S_{\mathrm{n}}(f) \delta(f - f')$
- Estimate **noise PDF** $\mathbb{P}(n)$
  - **Model: Gaussian with short duration excess ("glitches")**
  - Important later for **binary parameter inference**

Detectors

$\sqrt{S_{\mathrm{n}}(f)}$

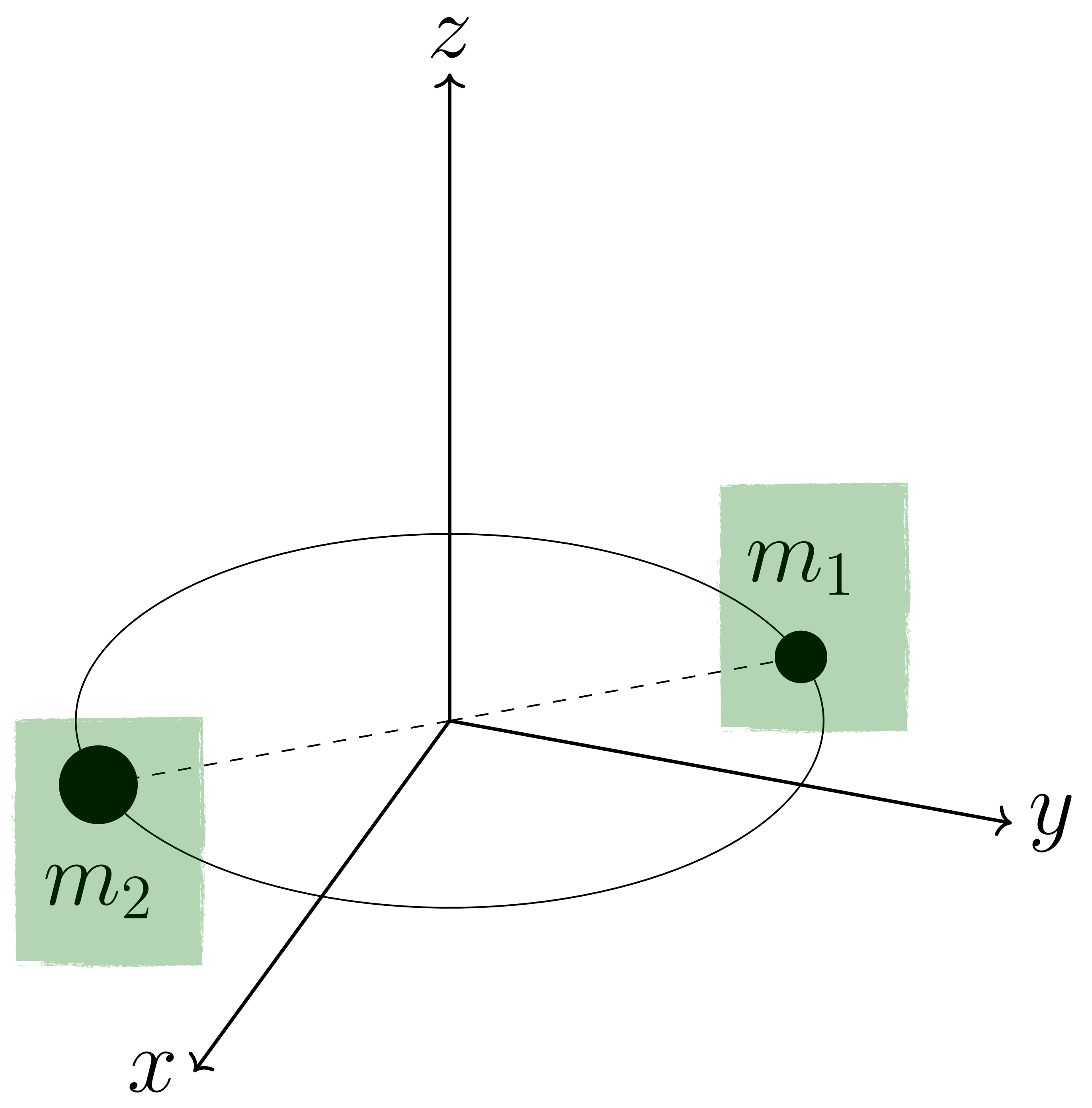




Credit: Carl Knox (OzGrav, Swinburne University of Technology).

Gravitational wave sources

- **Sources with large** $h(f)$ at detector's most sensitive frequencies
- **Compact binary mergers**
  - Black holes
  - Neutron stars
- **Potential: not yet observed**
  - Supernovae
  - Continuous waves
  - Tidal disruption events, etc.



Compact binary mergers

- **Intrinsic properties: masses, spins**
- **Loss of energy and angular momentum through GW emission**

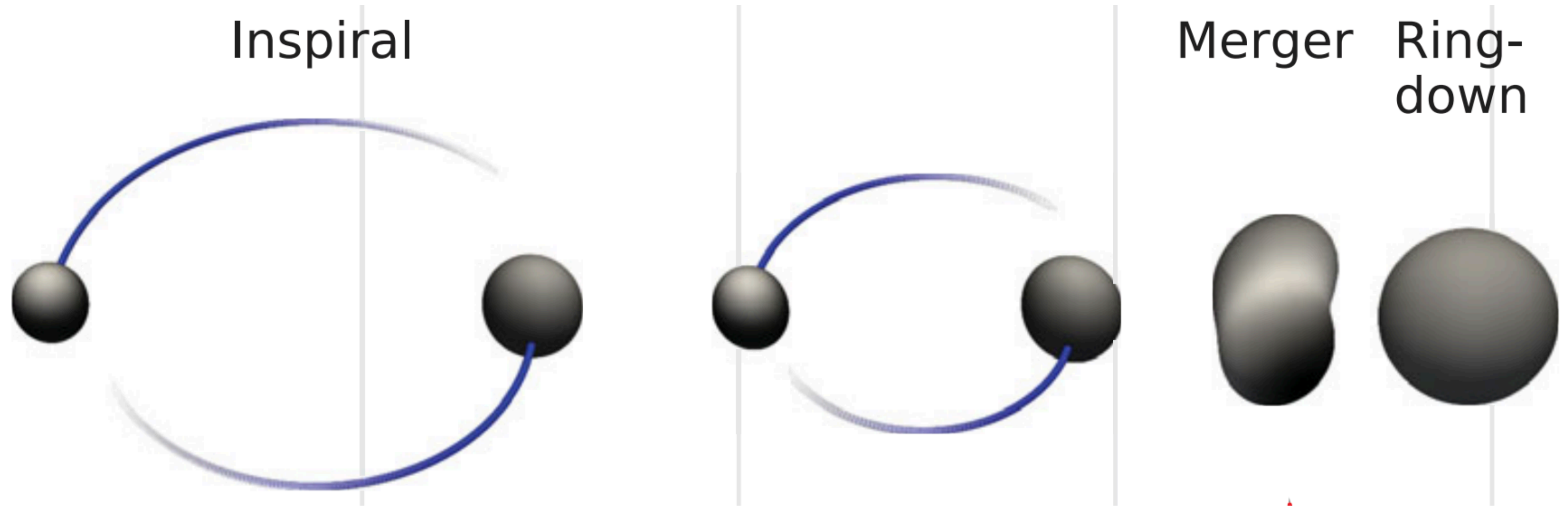
➡ **Binary separation shrinks**

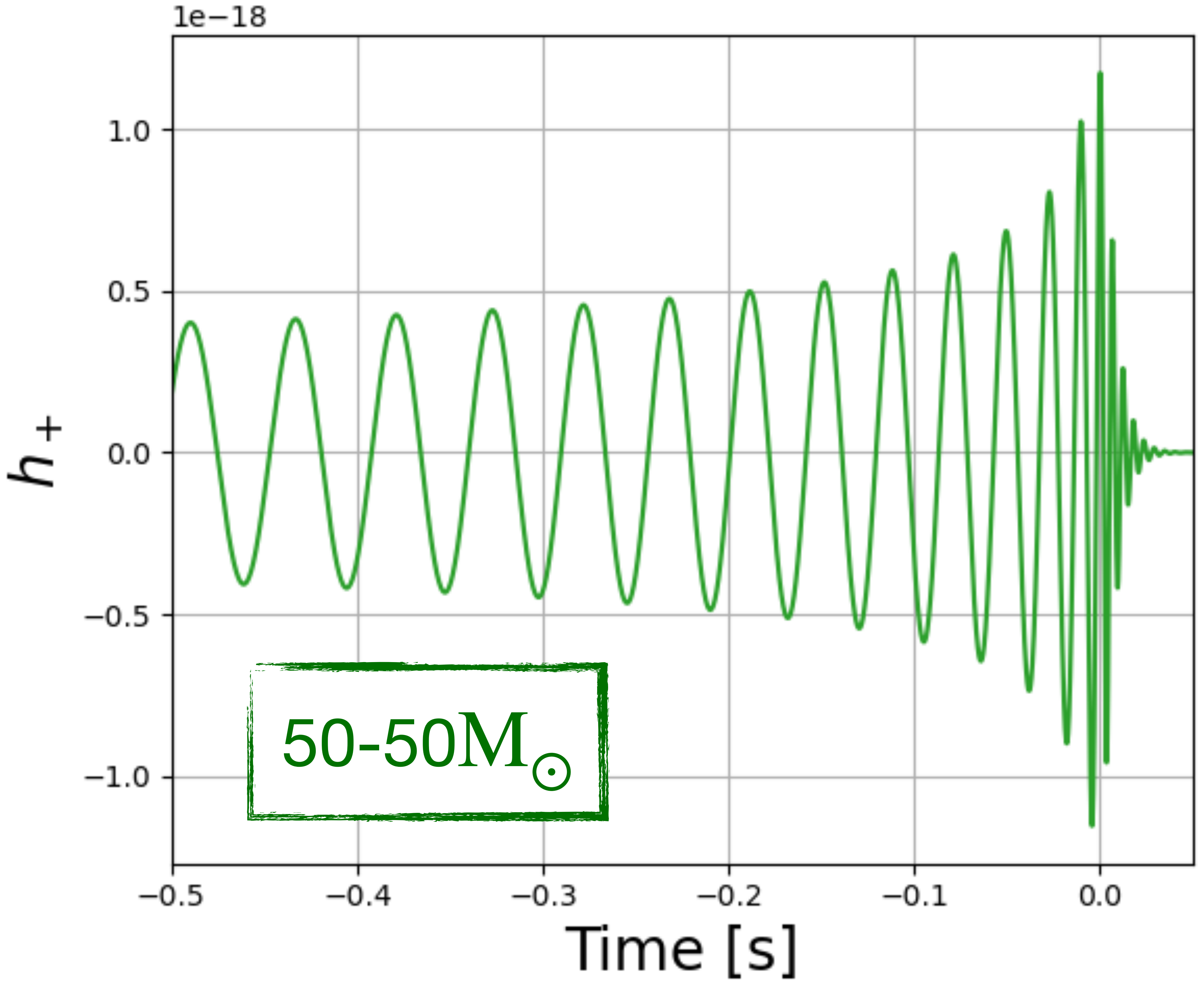
➡ **Kepler: frequency increase**

➡ **Amplitude increase: *"Chirp"***

Chirp mass $\mathcal{M} = \frac{(m_1 m_2)^{3/5}}{(m_1 + m_2)^{1/5}}$

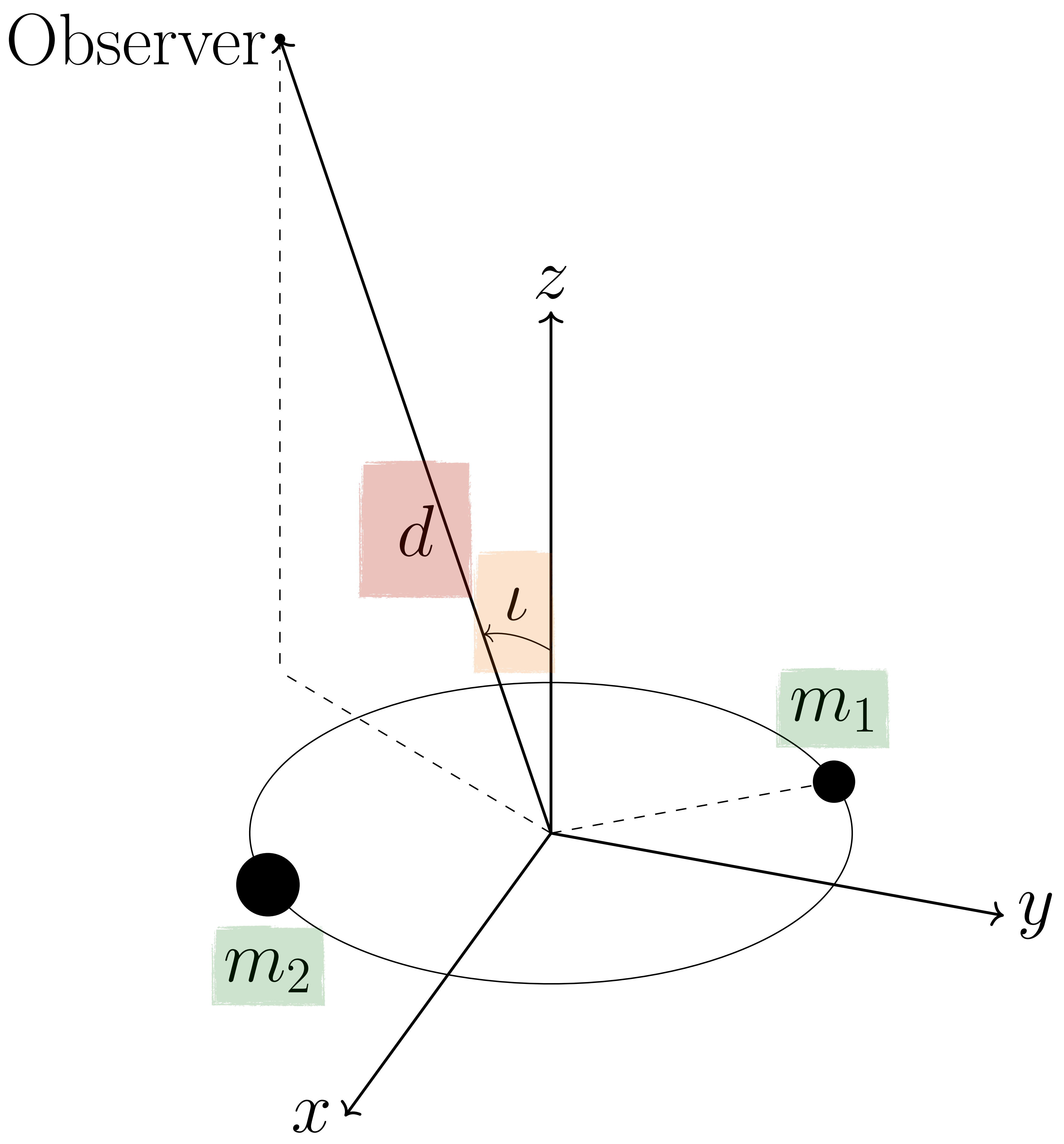
$\dot{f} \propto \mathcal{M}^{5/3} f^{11/3}$





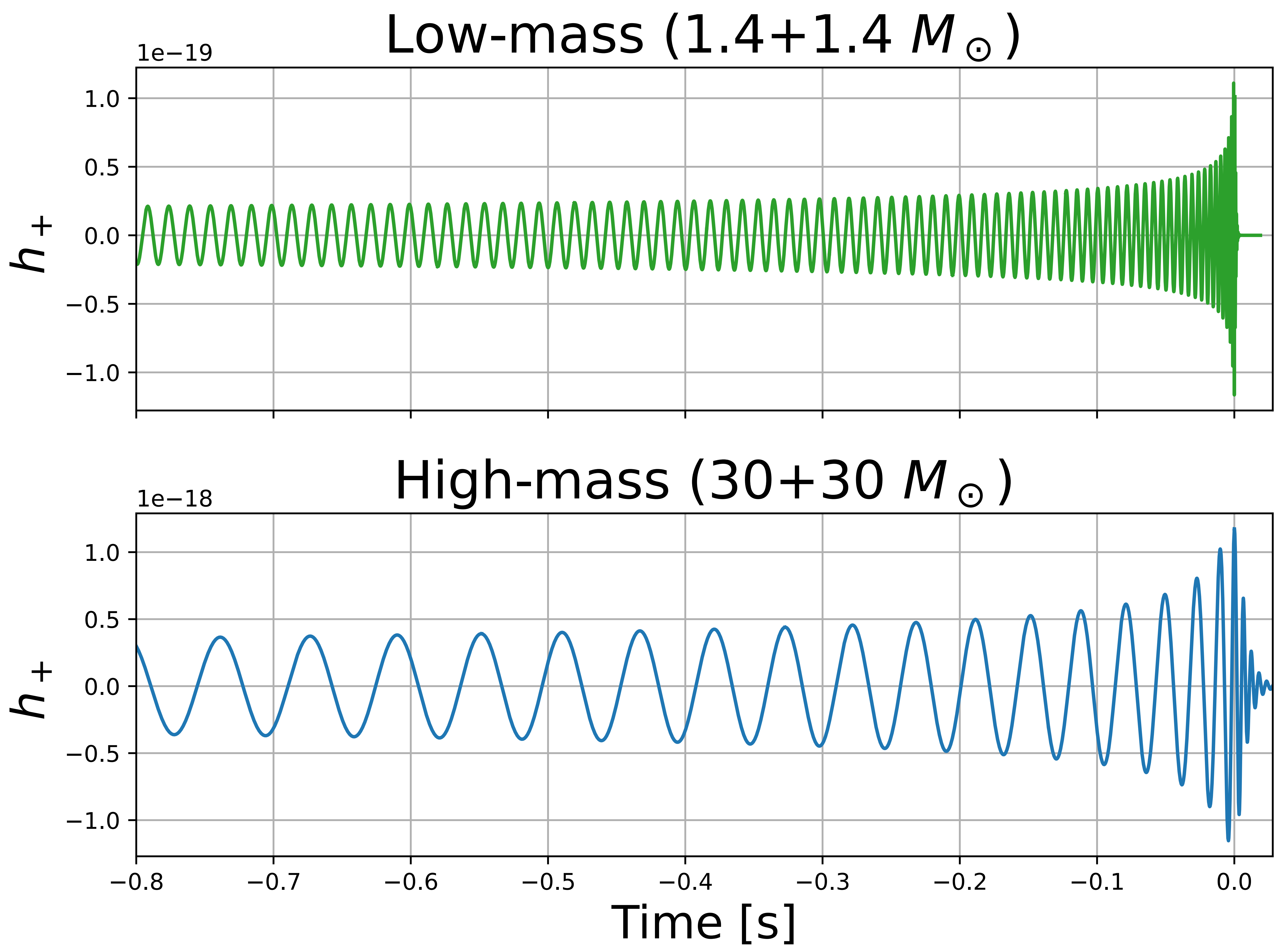
Compact binary mergers

- **Intrinsic properties: masses, spins**
- **Extrinsic properties: sky position, distance, inclination, reference time, reference phase, etc.**
- Computational cost division:
  - **Intrinsic: costly**, **extrinsic: cheap**
- **Signal model ~understood: different waveform models**
  - **Signal generation:** $\mathcal{O}(10^{-4} - 1)$ **sec**



Signal dependence on parameters

- **Lighter signals**
  - **Merger at higher frequencies**
  - **Long duration (BNS up to minutes, BBH sub-second)**
  - **Low amplitude**
- **Spins induce beating pattern**



1Mpc signal

Dedicated data analysis pipelines

A. **Detect short-duration binary signal** in noisy data

B. Estimate **astrophysical binary parameters**

C. Extract science from **populations of all binaries**

Dedicated data analysis pipelines

A. **Detect short-duration binary signal** in noisy data

B. Estimate **astrophysical binary parameters**

C. Extract science from **populations of all binaries**

GW parameters

Noise

$$\underbrace{\mathscr{D}}_{\text{GW data}} = \underbrace{s(\overbrace{\theta}^{\text{GW parameters}})}_{\text{Signal}} + \overbrace{n}^{\text{Noise}}$$

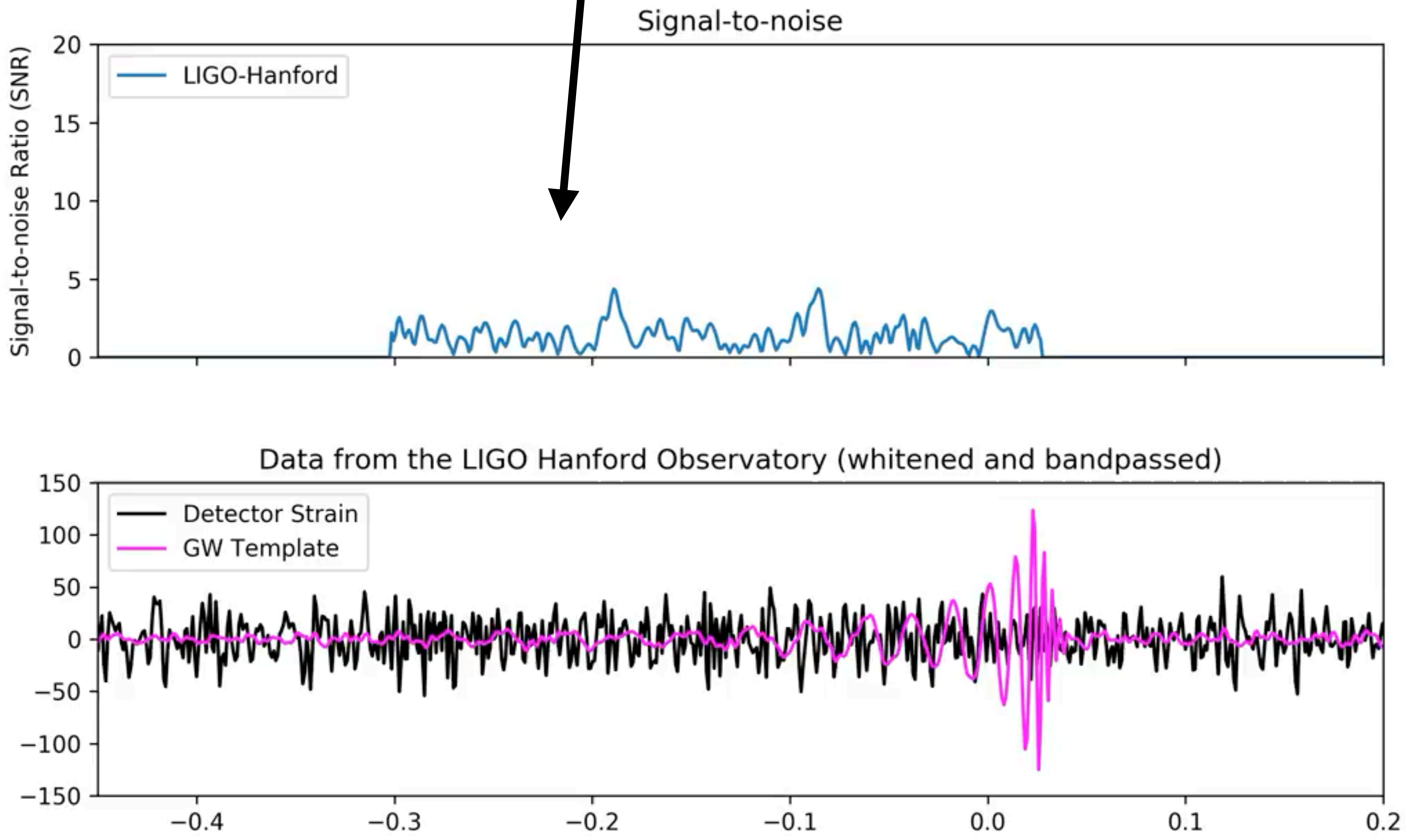
GW data

Signal

(1) Modeled searches: *"Look for known signals"*

- **Find short-duration signals** in noisy data
    - **Signal + Noise model**
    - **Compute overlap**
    - **Binary parameters unknown: precompute dataset of signals, i.e. template bank**
- **Assess probability of data of astrophysical origin**
    - **Coincidence** in several detectors **→ higher confidence**

$$\frac{\mathbb{P}(\text{Data}\,|\,\text{Signal} + \text{Noise})}{\mathbb{P}(\text{Data}\,|\,\text{Noise})}$$



Alex Nitz

(2) GW parameter estimation

- **Characterize binary properties and uncertainties**
  - ~15D parameter space: masses, spins, sky position, etc.

➡**Bayesian inference** approach

- **Likelihood:**
  - $\mathbb{P}(\mathscr{D}\,|\,\theta) = \mathbb{P}(\text{Residual noise}\,|\,\theta)$
  - Depends on **detector noise**
  - **Integral in Fourier-space**

Signal

$$\mathscr{D} = s(\theta) + n$$

Data Noise

Likelihood

$$\mathbb{P}(\theta\,|\,\mathscr{D}) \propto \mathbb{P}(\mathscr{D}\,|\,\theta)\mathbb{P}(\theta)$$

Posterior Prior

$\subset$

$$\int \mathrm{d}f \frac{\mathscr{D}(f)s(f;\theta)^*}{S_\mathrm{n}(f)}$$

Amplitude spectral density^2

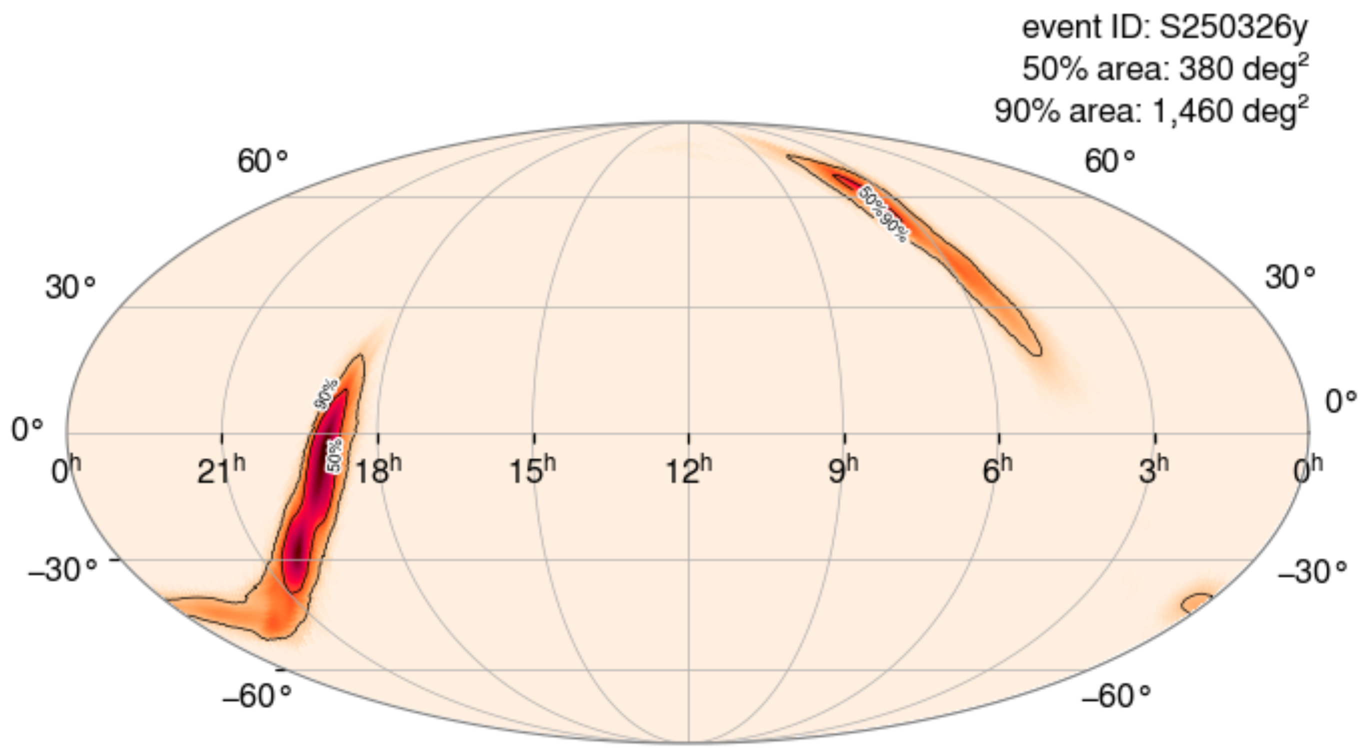
(2) Likelihood-based GW parameter inference

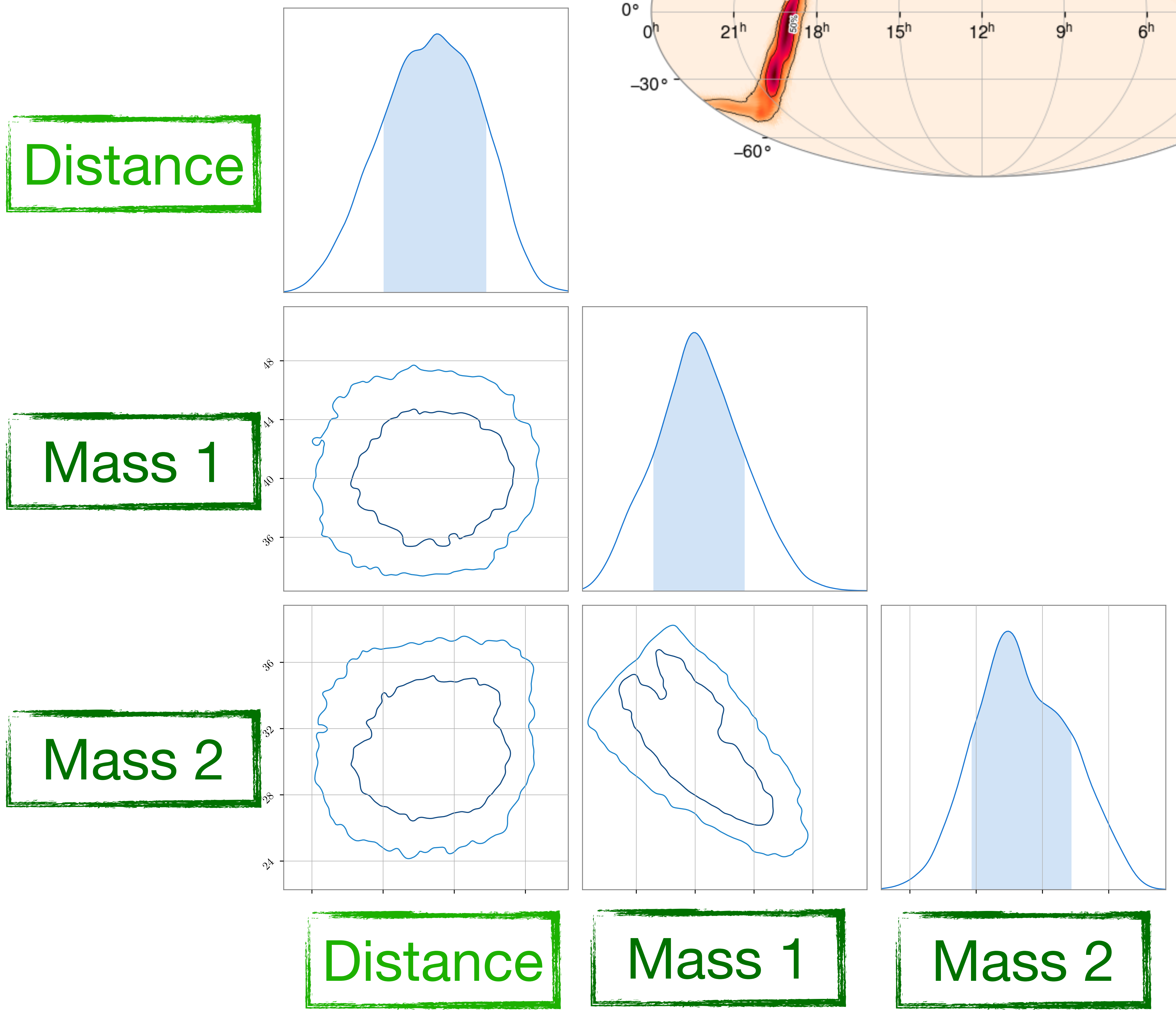
- **15 parameters; multi-modal posterior**

  ➡MCMC, nested sampling

  ➡**Hours for BBH, days for BNS**

- Alternatives
  - **Hamiltonian MC**, but most waveform models **not auto-differentiable (cf. E. Porter's talk)**
  - **Simulation-based inference applications** (see later)
- Optimization
  - **Multi-banding, heterodyning**, etc.





Heterodyning (or relative binning)

Cornish 1007.4820,
Zackay et al. 1806.08792

- **Likelihood: frequency integral**
  - Naive: **fine grid** since **integrand rapidly oscillating**
- **Reference waveform removes rapidly oscillating signal**
  - E.g. maximum likelihood parameters from searches
  - **Pre-inference: reference integral on fine frequency grid**
  - **Inference**: difference in likelihood integral on **coarse frequency grid**

➡ 10^{2-3} data reduction

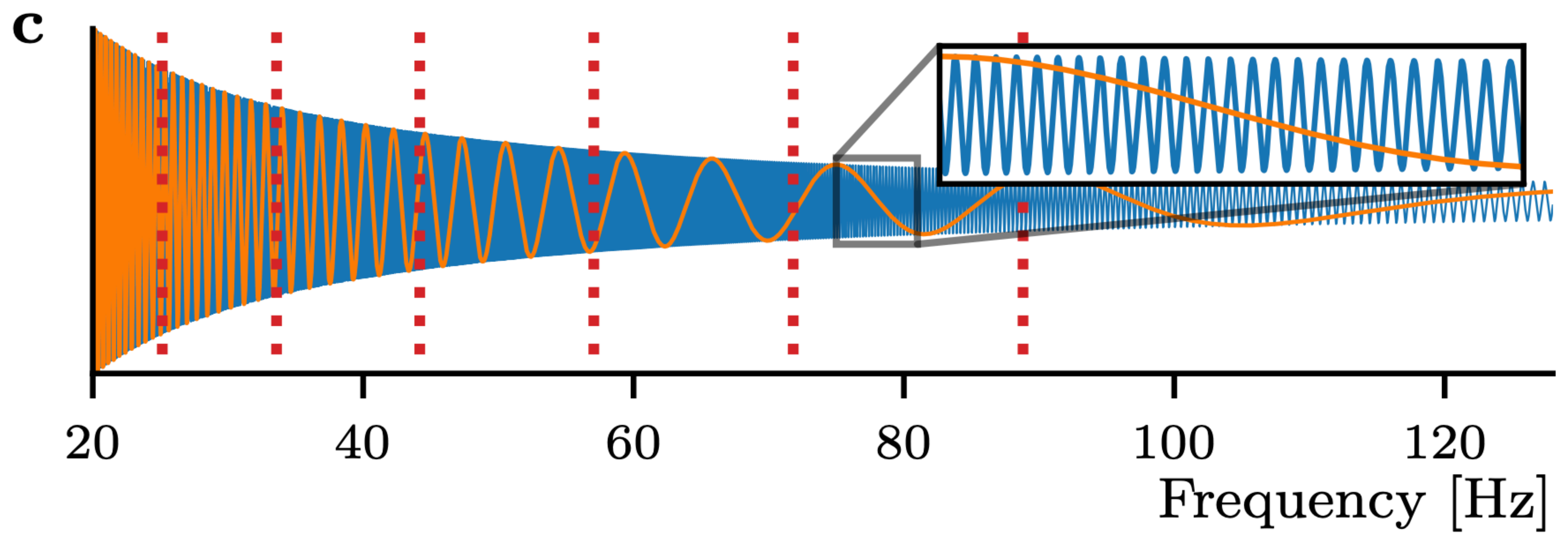
$$s(f) = A(f)\exp\left[i\Psi(f)\right]$$

Amplitude (slowly-oscillating)

Phase (rapidly-oscillating)

Ratio slowly-oscillating near reference $s(f;\theta_{\rm ref})$

$$r(f;\theta) = \frac{s(f;\theta)}{s(f;\theta_{\rm ref})}$$



2407.09602

Multi-banding

Vinciguerra et al. 1703.02062
García-Quirós et al. 2001.10897

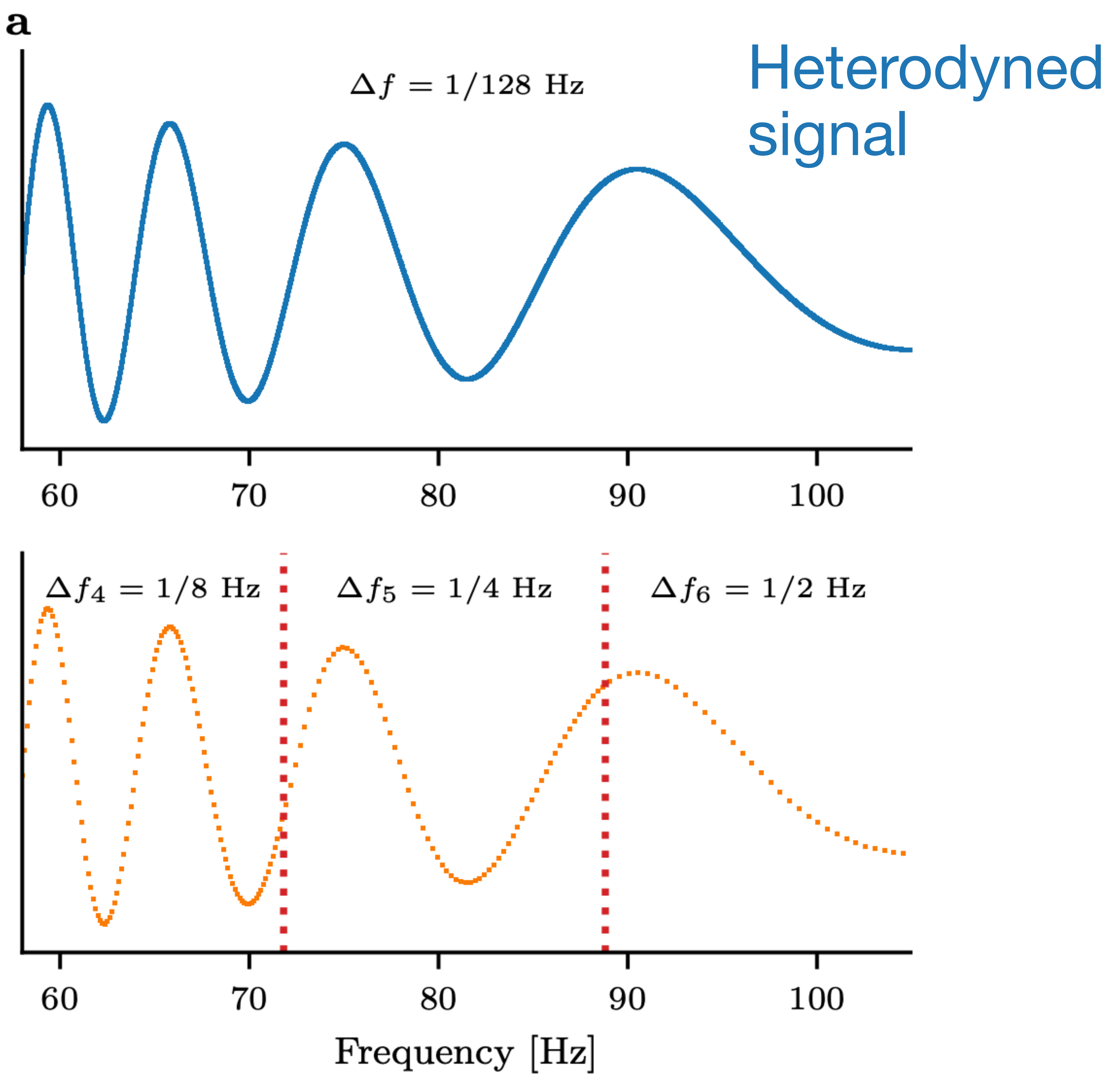
- **Slowly-varying** signal

  ➡**Reduced frequency resolution for same information**

  - $\Delta f_i = 2^i \Delta f_0$

- **Data compression** of $\times 60$

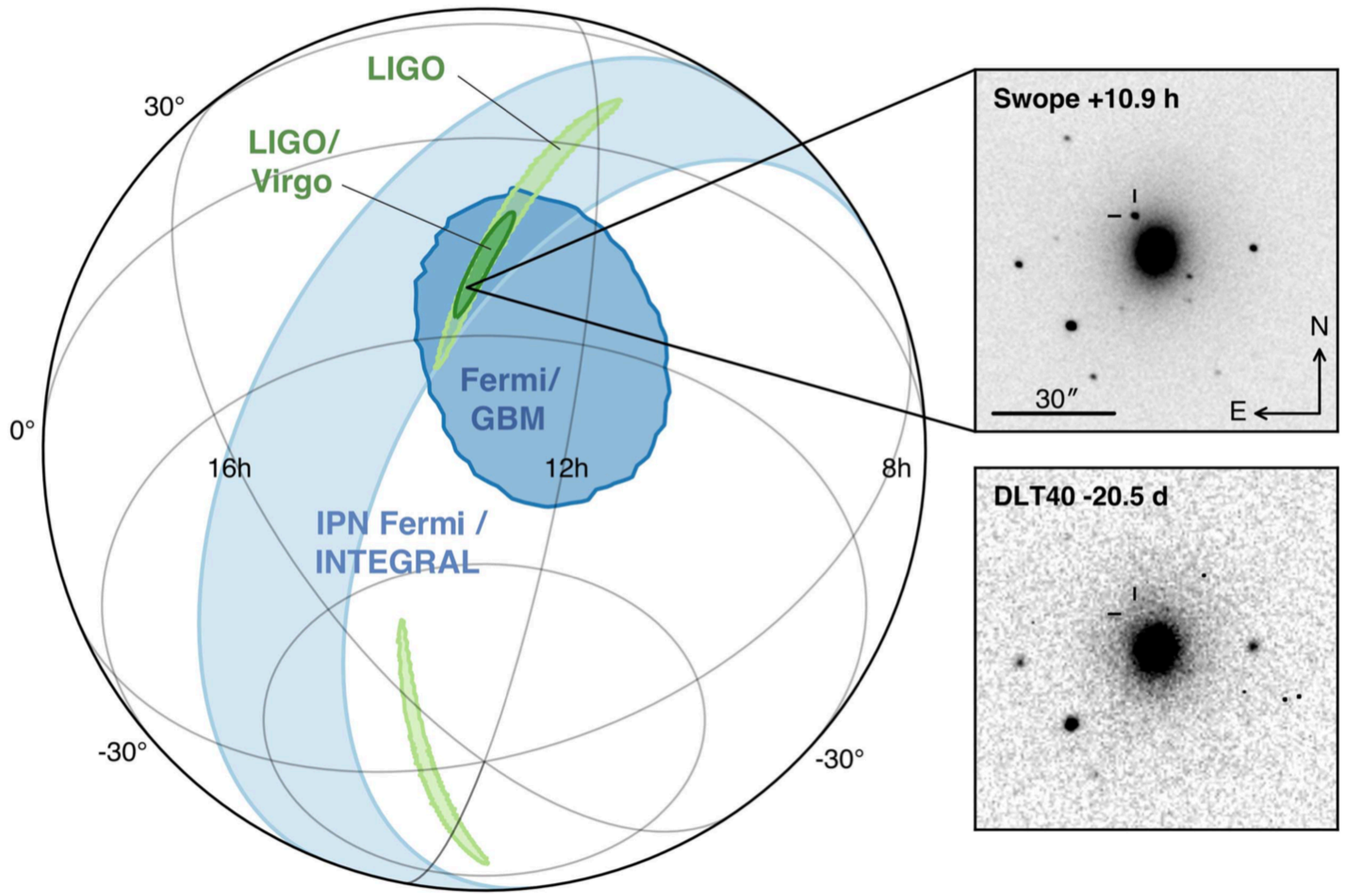
  ➡Important for **later SBI application**



Dax et al. 2407.09602

(2) Example of GW parameter estimation

- **Merger of binary neutron star** (GW170817)
- **Triangulation** → sky position
- **70 telescopes** follow-up: neutron-rich, fast-moving, rapidly-cooling cloud
  - **Host galaxy**: NGC4993
- Implications for cosmology and modified gravity (see later)
- **Need: fast and robust GW parameter estimation**



(3) Population analyses

Data

- Observations: ~200 binary mergers
- **Formation channels, cosmology, etc. impact binary signals**

Characterize **overall binary distribution**

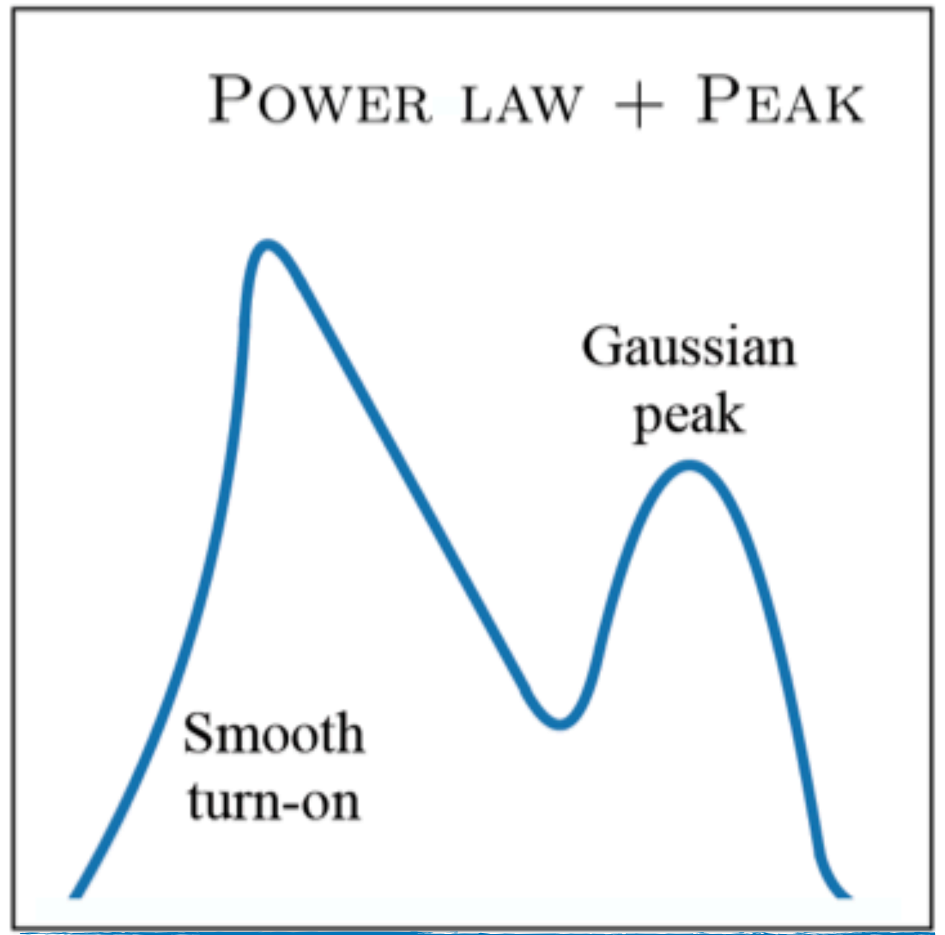
Questions

- **Probe formation of black holes and neutron stars**
- **Test cosmological model**
- **Verify theory of gravity**

(3) Hierarchical Bayesian inference

- **Hyperparameters Λ**
  - **Mass/spin/redshift** distribution parameters
  - **Cosmological** parameters
  - **Modified gravity** parameters

$$\mathbb{P}(\theta\,|\,\Lambda)$$



Mass of the heavier component

- **Infer hyperparameters Λ** from GW catalog $\{\mathscr{D}\}$ from
  - **Astrophysical distribution**
  - **Single-event posterior samples**
  - **Selection bias**

$$\mathbb{P}(\Lambda\,|\,\{\mathscr{D}\}) \propto \mathbb{P}(\{\mathscr{D}\}\,|\,\Lambda)\mathbb{P}(\Lambda)$$

Simulation-based inference

Applications of ML in GW science (non-exhaustive)

Motivation SBI

- Speed
- Non-tractable noise (non-Gaussian)
- Data reduction

SBI examples

- **GW parameter estimation (PE)**
  - **Hunter et al. 1909.06296 (Conditional Variational Autoencoders)**
  - **Delaunoy et al. 2010.12931 (CNN)**
  - **Dax et al. 2106.12594, 2111.13139, 2305.17161 (NF, flow-matching)**
  - **Bhardwaj et al. 2304.02035 (NF)**
  - **Xiong et al. 2405.09475 PE in non-Gaussian noise (NF)**
  - **Raymond et al. 2406.03935 GW PE for intermediate-mass**
- **Stochastic gravitational wave background data**
  - **Alvey et al. 2309.07954 (TMNRE)**
- **GW Cosmology**
  - **Stachurski et al. 2310.13405 H0 Dark Siren Measurement (NF)**
  - **Leyde et al. 2311.12093 H0 mass spectrum method (NF)**

Other examples for non-SBI

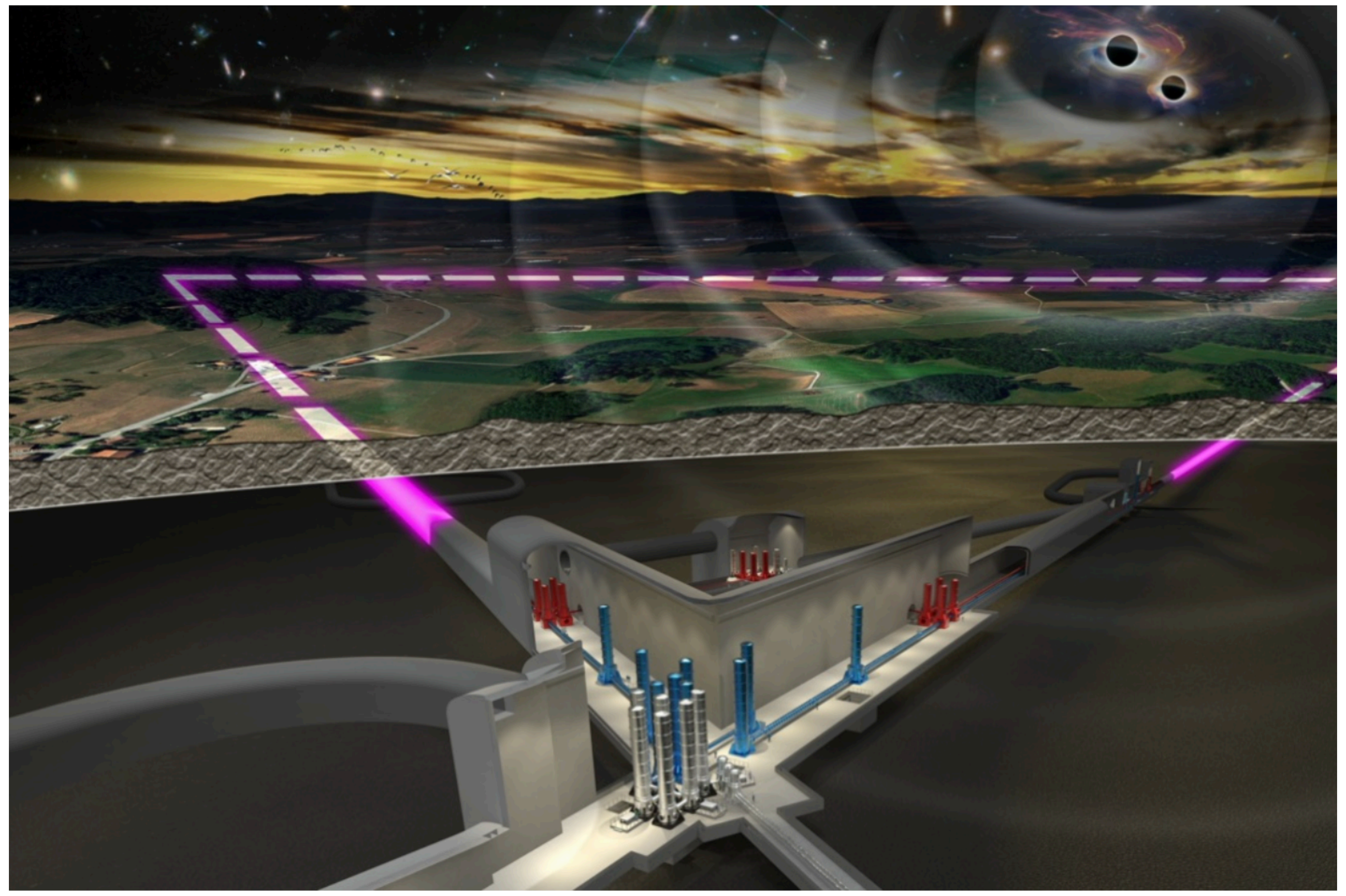
- **GW parameter estimation**
  - **Williams et al. 2102.11056 (Nested sampling with NFs)**
  - **Legin et al. 2410.19956 (Score-Based Likelihood Characterization)**
- **Signal vs noise classification (Detection)**
  - **Cavaglia et al. 2002.04591 (Genetic Programming)**
  - **Jadhav et al. 2010.08584 (Deep Learning), Jadhav 2306.11797 (Deep Learning)**
  - **Baltus et al. 2104.00594 (Deep Learning)**
  - **Lopez et al. 2112.06608 (Gaussian Process Regression)**
  - **Andres-Carcasona et al. 2212.02829 (Deep Learning)**
  - **Chatterjee et al. 2207.14522 (Deep Learning)**
  - **Boudart 2210.04588 (CNN), Boudart and Fays 2201.08727 (CNN)**
  - **Fernandes et al. 2303.13917 Glitch classification (ResNet)**
  - **Shah et al. 2306.13787 (Random Forest)**
  - **Trovato et al. 2307.09268 (Deep Learning)**
- **Interferometer Control**
  - **Coughlin et al. 1611.09812 (Deep Learning)**
  - **Biswas et al. 1910.12143**
  - **Ma and Vajente 2302.07921 (Probabilistic NN, Gated Recurrent Unit)**
- **Denoising**
  - **Vajente et al. 1911.09083 (Deep NN)**
  - **Ormiston et al. 2005.06534 (CNN)**
  - **Yu and Adhikari 2111.03295 (CNN)**
  - **Torres-Forné et al 2002.11668 (Deep Learning)**
- **Population studies**
  - **Wong et al. 2206.04062 Inverse population synthesis**
  - **Wong and Gerosa 1909.06373 Population model emulator (Gaussian Process Regression)**
  - **Wong et al. 2002.09491 Population model emulator (NF)**
  - **Ruhe et al,. 2211.09008 Population model emulator (NF)**
- **Etc.**

See Cuoco et al. 2412.15046

GW astronomy in the future

Virgo

- Operational until ~2035
- Detect $\mathcal{O}(1000)$ events / year



Einstein Telescope
(Image: NIkhef)

- Operational from ~2035
- Detect $\mathcal{O}(10^5)$ events / year

Simulation-based inference

- **<u>Idea</u>: posterior approximation**
  - Model parameters θ, data $\mathscr{D}$
  - **True posterior:** $\mathbb{P}(\theta \,|\, \mathscr{D})$
  - **Approximate model:** $\mathbb{Q}(\theta \,|\, \mathscr{D})$
- **Generative model with model parameters** w
  - $\mathbb{Q}(\theta \,|\, \mathscr{D}) \rightarrow \mathbb{Q}(\theta \,|\, \mathscr{D}, w)$
- **Optimize loss(w) s.t.** $\mathbb{Q} \approx \mathbb{P}$

Papamakarios and Murray 1605.06376, Lueckmann et al. 2017, Greenberg et al. 2019, Cranmer et al. 1911.01429

The loss

➡**Minimize distance($\mathbb{P}$,$\mathbb{Q}$)**

➡Example: **Kullback-Leibler**

$$\mathrm{KL}(\mathbb{P},\mathbb{Q}) = \int \mathrm{d}\theta\, \mathbb{P} \log\left(\frac{\mathbb{P}}{\mathbb{Q}}\right)$$

$$\mathrm{KL}(\mathbb{P},\mathbb{Q}) = 0 \quad \text{for } \mathbb{P} = \mathbb{Q}$$

$$\mathrm{KL}(\mathbb{P},\mathbb{Q}) > 0 \quad \text{for } \mathbb{P} \neq \mathbb{Q}$$

- **Evaluation?** Avoid costly
  - **Integral over** θ
  - **Computation of** $\mathbb{P}(\theta\,|\,\mathscr{D})$

$$
\begin{aligned}
\text{Loss} &= \mathbb{E}_{\mathbb{P}(\mathscr{D})}\left[\mathsf{KL}(\mathbb{P},\mathbb{Q})\right] \\
&= \dots \\
&= -\,\mathbb{E}_{\mathbb{P}(\theta,\mathscr{D})} \log(\mathbb{Q}) + \# \\
&\approx -\sum_{\theta,\mathscr{D}\sim\mathbb{P}(\theta,\mathscr{D})} \log(\mathbb{Q}) + \#
\end{aligned}
$$

Def. conditional probability

Dependent on w

Independent of w

- Only forward model

  ➡Generation of (parameter-data) pairs

  - No appearance of $\mathbb{P}(\theta|\mathscr{D})$

    ➡No explicit likelihood, *"likelihood-free"*

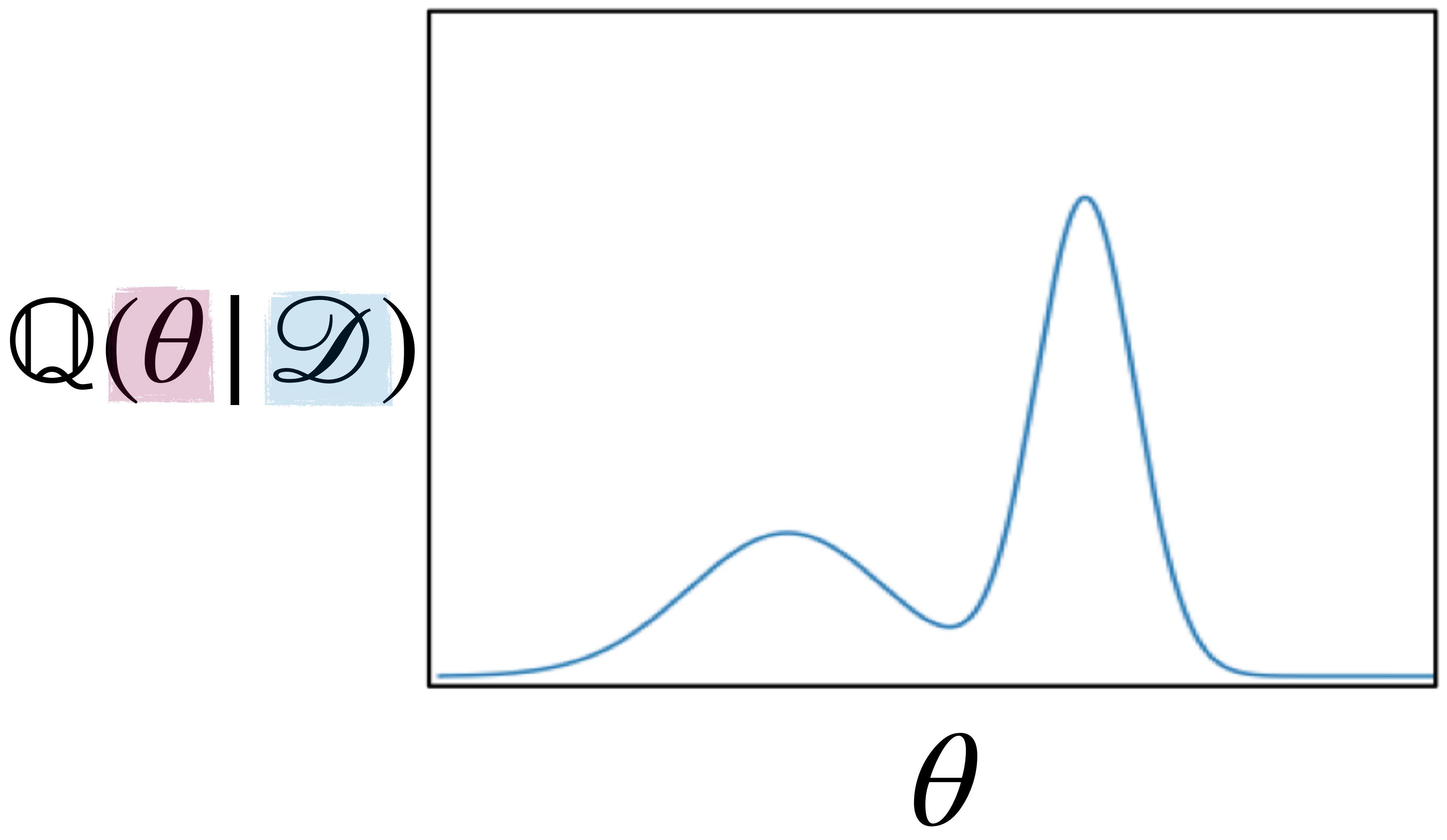
- Require: evaluation of $\mathbb{Q}(\theta|\mathscr{D})$

$$\text{Loss} \approx -\sum_{\theta,\mathscr{D}\sim\mathbb{P}(\theta,\mathscr{D})} \log(\mathbb{Q})$$

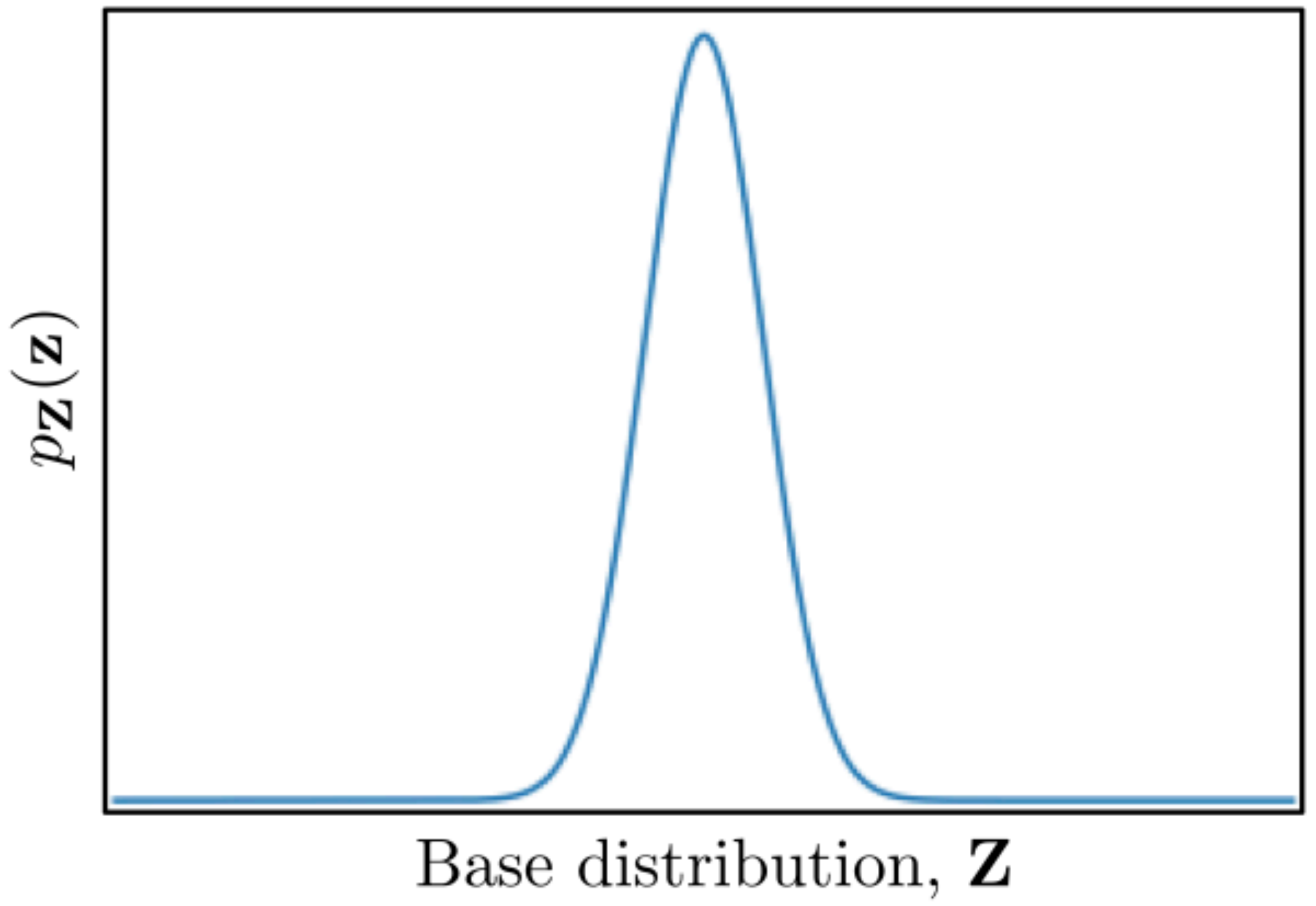
Papamakarios and Murray 1605.06376, Lueckmann et al. 2017, Greenberg et al. 2019, Cranmer et al. 1911.01429

Normalizing flows

- **Generative model** from machine learning
- **Simple PDF via non-linear transformation to** $\mathbb{Q}(\theta \,|\, \mathscr{D})$
- **Transformation g data-dependent**



$$\boxed{\theta := g(Z, \text{data}; w)} \longleftarrow$$



Tabak and Vanden-Eijnden 2010, Tabak and Turner 2013, Durkan et al. 2019, Kobyzev et al. 2019

Dingo: fast parameter estimation

Dax et al. 2106.12594, 2111.13139

- **Build** $\mathcal{O}(10^7)$ **waveforms in intrinsic parameters** (masses, spins)
- **Trick (1): Data augmentation**
  - **Draw extrinsic parameters and noise during training**
- **Trick (2): Normalize data**
  - **Whiten 2-point statistic**
- **Trick (3): Reduce data dimension**
  - **Standardize** arrival time, ***"group-equivariant NPE"***

Group equivarient neural posterior estimation

Dax et al. 2111.13139

- **Reduce data dimension**/simplify learning problem

➡**Standardize** arrival time/sky position

- *"How should I shift my data s.t. the signal occurs at all detectors at the same time?"*
- True reference time **unknown: blur arrival time with kernel**

➡Estimate pose and GW parameters θ jointy via **Gibbs sampling**

➡**Iterative problem (for real data: 30 it.)**

$$\theta \to T_g\theta \qquad \mathscr{D} \to \mathscr{D}_g$$

Pose

Gibbs

$$\mathbb{Q}(\theta \,|\, \mathscr{D}_g, g)$$

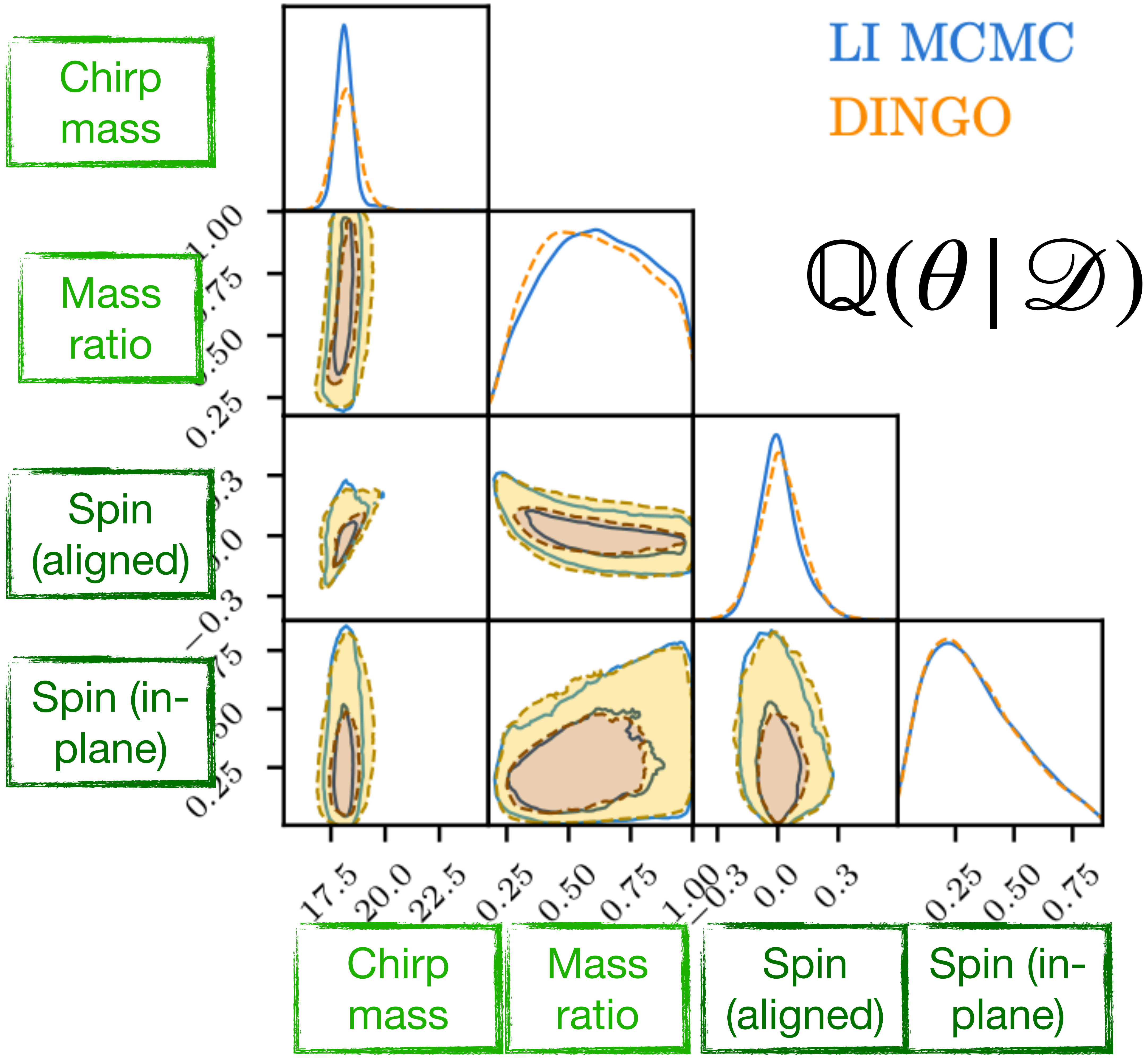
$$\mathbb{Q}(g \,|\, \mathscr{D}_g, \theta) = \mathrm{kernel}(\theta)$$

Dingo - binary black holes

Dax et al. 2106.12594, 2111.13139

- **Fully SBI:** $\mathcal{O}(10^4)$ samples in **<1min**
- **Importance sampling**
  - Assign new weight to **samples** if likelihood available
  - $\mathcal{O}(10^4)$ **samples in <1h**
  - Sometimes: **low sample size (for high SNR events)**

$$p_{\text{new}} \propto \frac{\mathbb{P}(\theta \,|\, \mathcal{D})}{\mathbb{Q}(\theta \,|\, \mathcal{D})}$$



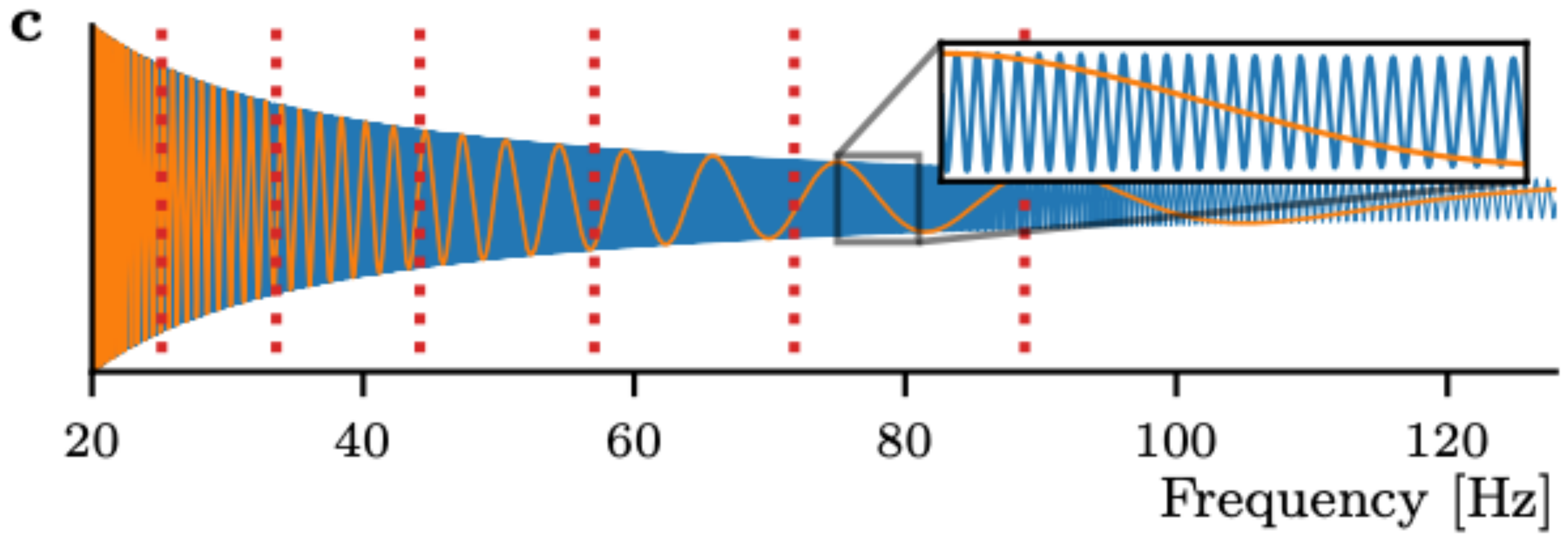
Dingo - binary neutron stars

Dax et al. 2407.09602

- Challenge:
  - Low mass → **long duration** signals
  - Potential EM → fast inference
- Trick (1): **Data reduction**
  - **Divide out analytic signal evolution (heterodyning)**
  - **Adapt frequency bins to speed of frequency evolution,** $\delta f \propto 1/\delta t$ **(multi-banding)**

  ➡**Less frequency bins for same data content**
- Trick (2): **Importance-sampling**
  - Correct for eventual biases
- Trick (3): **Prior conditioning**
  - **See next slide**

Full signal Simplified signal



$$\mathscr{D} \rightarrow \mathscr{D}_{\widetilde{\mathscr{M}}}$$

Cornish 1007.4820, Zackay et al. 1806.08792
Vinciguerra et al. 1703.02062, García-Quirós et al. 2001.10897

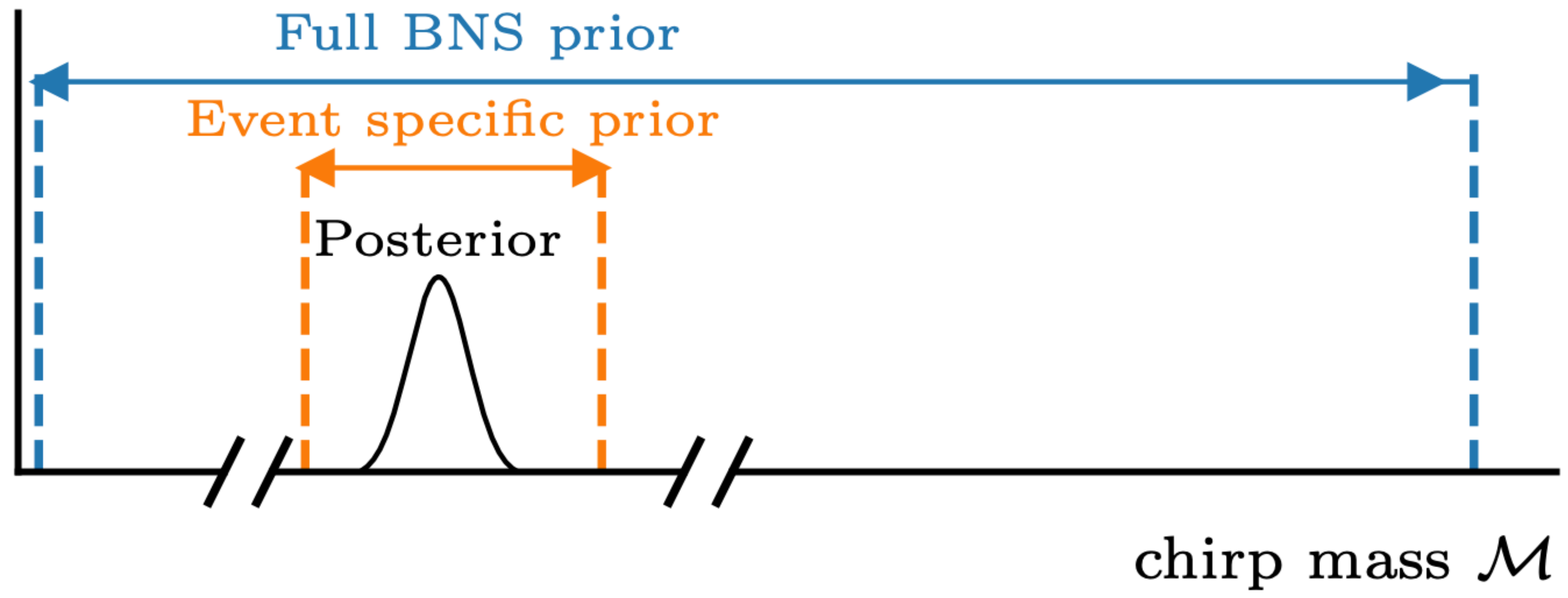
Prior conditioning

Dax et al. 2407.09602

- BNS signals very long

➡Tight constraints on the chirp mass $\mathscr{M}$

- Introduce **hyperprior** $\widetilde{\mathscr{M}}$



$2\Delta\mathscr{M} = 0.01\mathrm{M}_{\odot}$

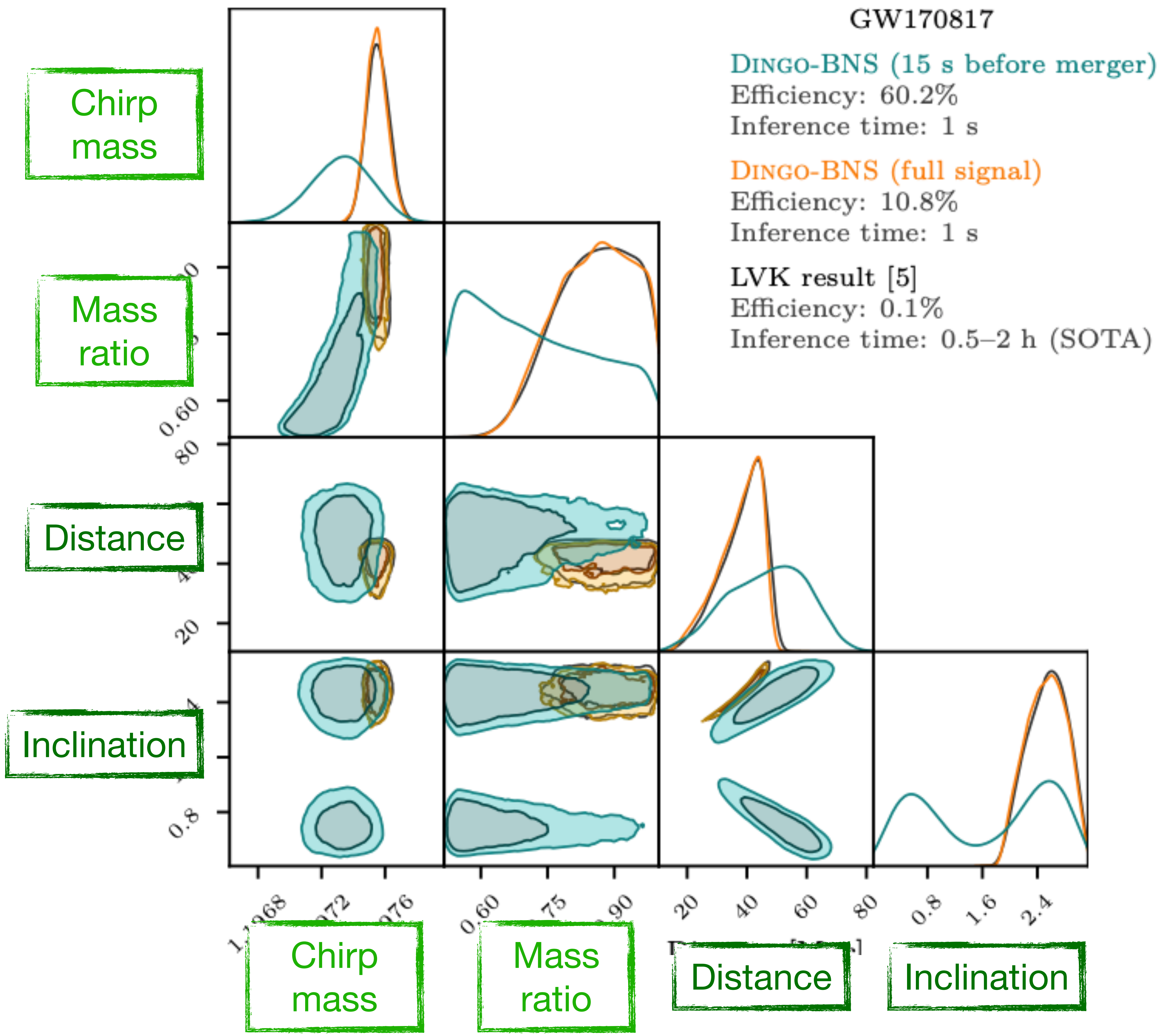
- **Condition** model on data and $\widetilde{\mathscr{M}}$
  - $\mathscr{Q}(\mathscr{M}\,|\,\mathscr{D}) \to \mathscr{Q}(\delta\mathscr{M} = \mathscr{M} - \widetilde{\mathscr{M}}\,|\,\mathscr{D}_{\widetilde{\mathscr{M}}}, \widetilde{\mathscr{M}})$

➡Prior volume reduction $\mathscr{O}(100)$

- **Training**
  - Draw prior mean $\widetilde{\mathscr{M}}$
  - Draw $\mathscr{M} \sim \mathbb{P}(\widetilde{\mathscr{M}})$
  - Produce data $\mathscr{D} \sim \mathbb{P}(\mathscr{D}\,|\,\mathscr{M})$
  - Optimize $\mathscr{Q}(\delta\mathscr{M}\,|\,\mathscr{D}_{\widetilde{\mathscr{M}}}, \widetilde{\mathscr{M}})$

Dingo - binary neutron stars

- **Results in ~1sec**
  - **Fast EM follow-up**
- Up to **1h duration**
  - **Future GW data**
- Analyze incoming data
  - **Pre-merger detection**
- Challenges
  - **Higher-order modes: no analytic reference signal**



SBI for hierarchical GW studies

KL, Green, Toubiana and Gair. PRD 109 (2024) 6, 064056

Motivation

- Current analyses **cannot analyze large datasets**

Approach

- **Normalizing flow approximates hyperparameter posterior**
- **Machine learning pipeline for event summary**

Λ hyperparameters

- Hubble constant
- Marginal parameters

$$\mathbb{P}(\Lambda \,|\, \mathrm{GW\ catalog})$$

Application of Hierarchical Bayesian inference: H_0 measurement

- H_0 **impacts luminosity distance-redshift relation**
- Observed **mass scales redshifted**

➡ $m_\mathrm{d} = (1+z)m_\mathrm{s}$ Source-frame mass

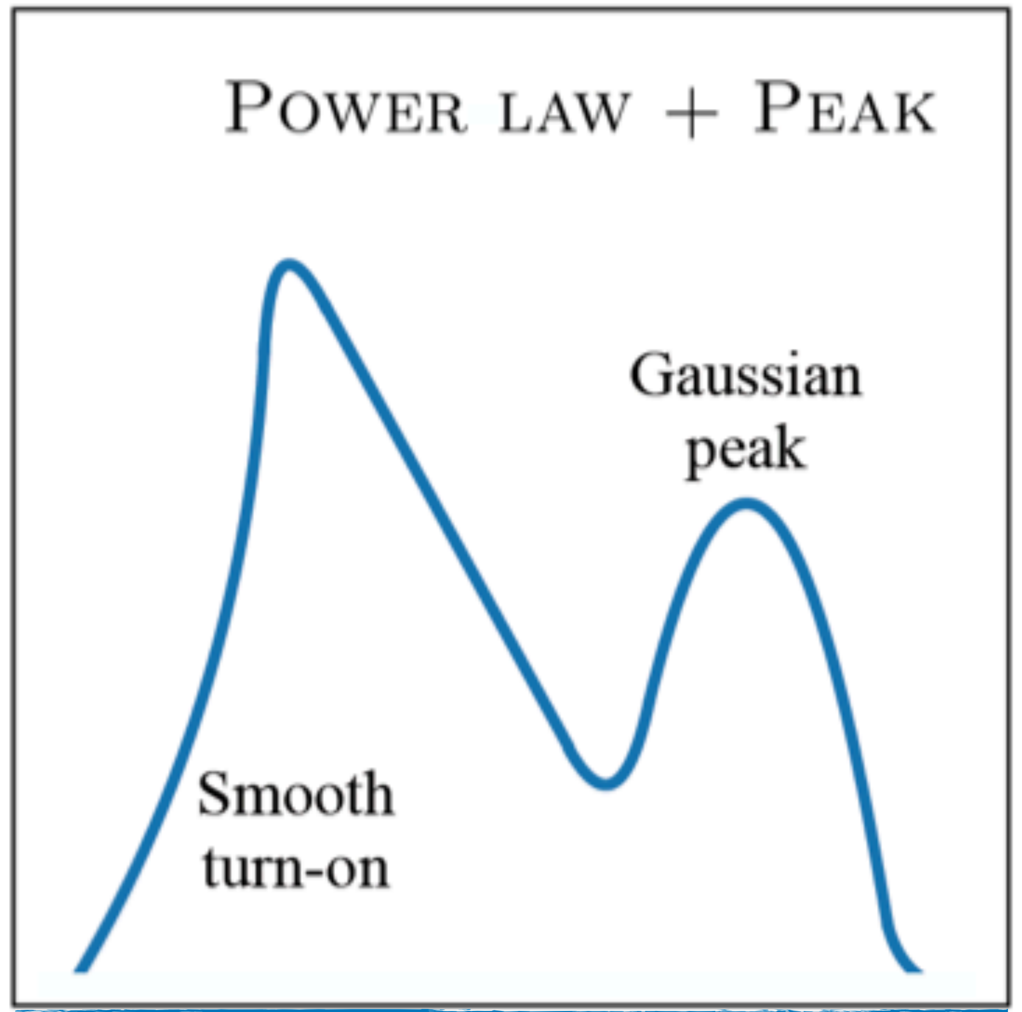
Detector-frame mass

- **Jointly constrain source-frame mass distribution and H_0**

➡ ***"Mass spectrum method"***

➡ **Only GW, combination with galaxy catalogs possible**

m_s



Mass of the heavier component

SBI for hierarchical GW studies

$$\text{Loss} \approx - \sum_{\Lambda,\{\mathscr{D}\}\sim\mathbb{P}(\Lambda,\{\mathscr{D}\})} \log\left(\mathbb{Q}\right)$$

Data now: selected biased post-processed GW signals

Draw Universe Λ → **Draw binary mergers/ produce signals** → **Detector data/ detection** → **Post-process data**

SBI for hierarchical GW studies

$$\text{Loss} \approx - \sum_{\Lambda,\{\mathscr{D}\}\sim\mathbb{P}(\Lambda,\{\mathscr{D}\})} \log\left(\mathbb{Q}\right)$$

Data now: selected biased post-processed GW signals

Draw Universe Λ

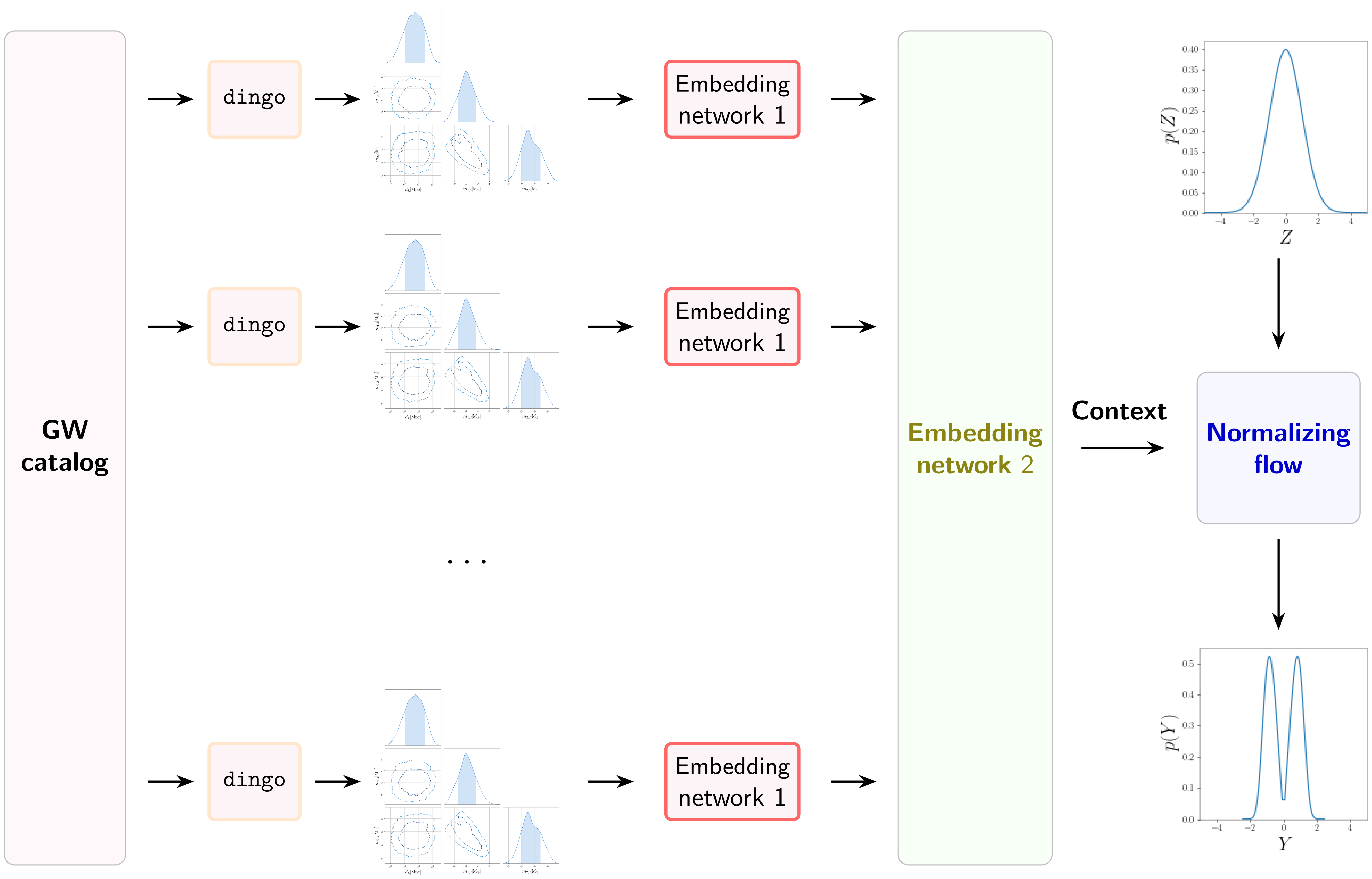
Draw binary mergers/ produce signals

Detector data/ detection

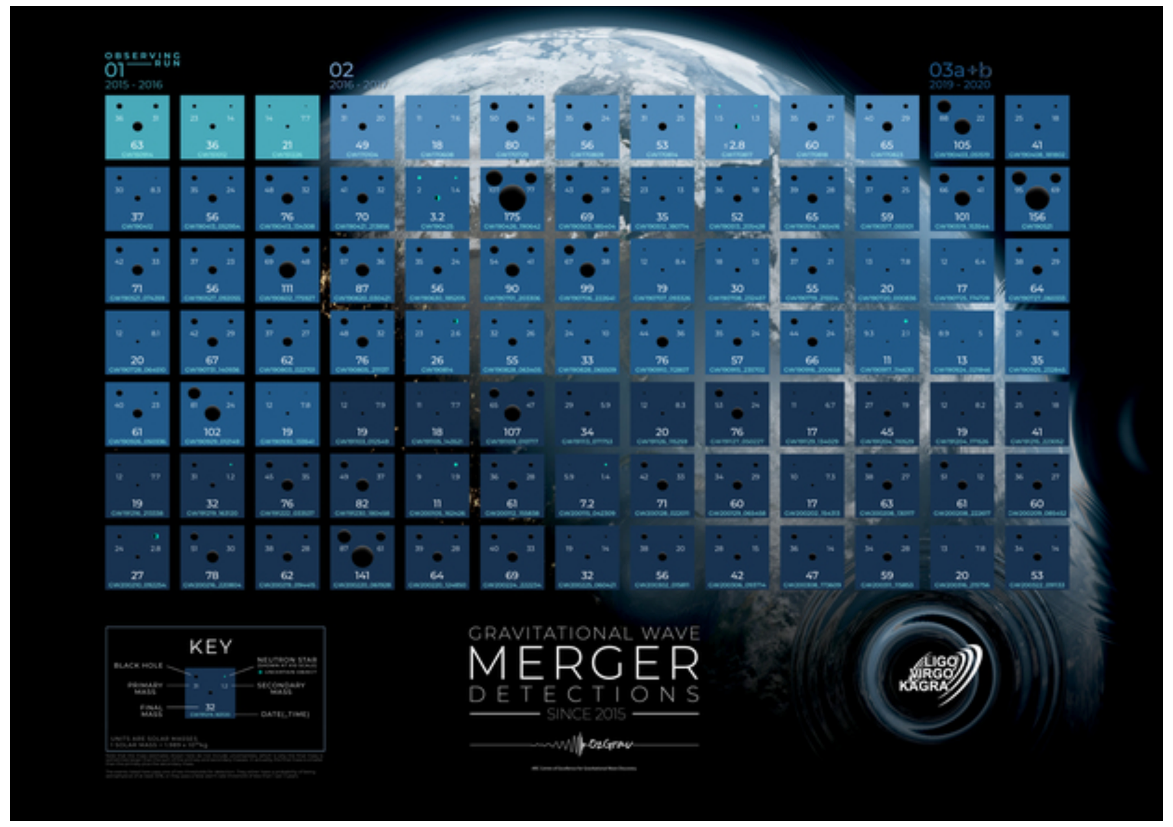
Post-process data

Dingo

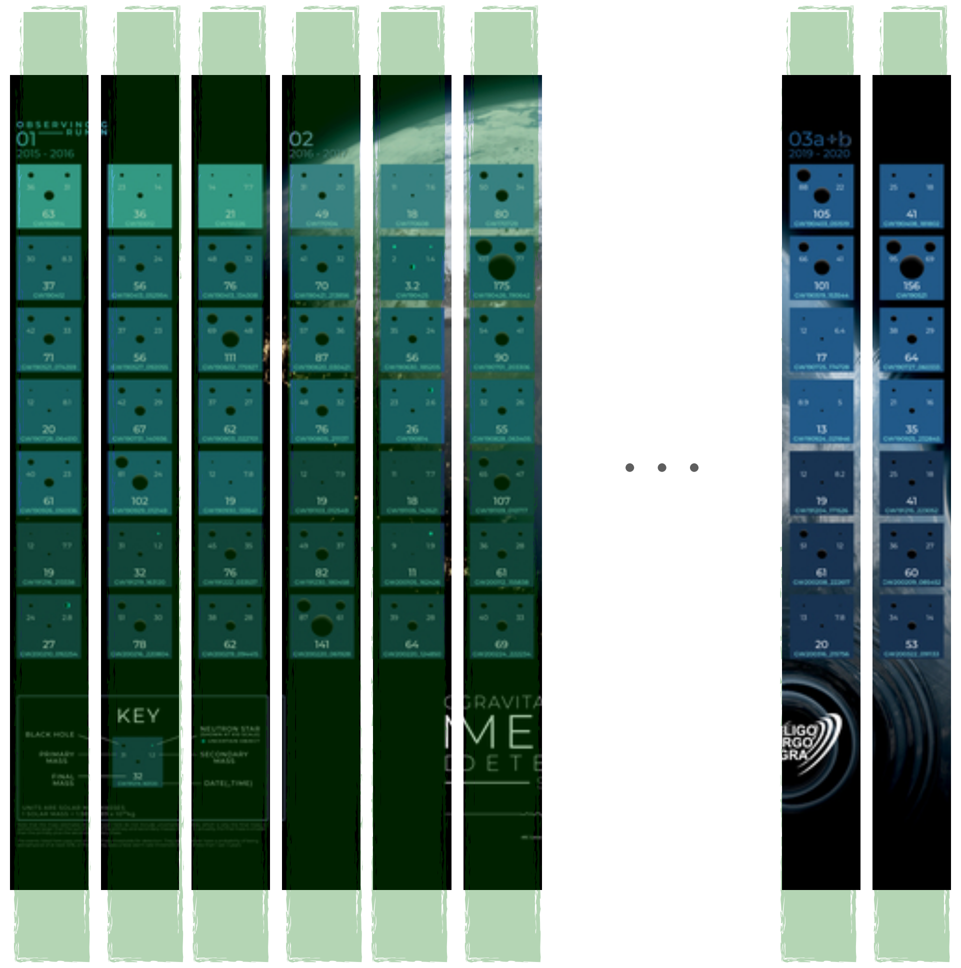
Model architecture



Split problem in multiple sub-problems



Batch catalog



- Analyze each batch individually
- Combine via importance sampling

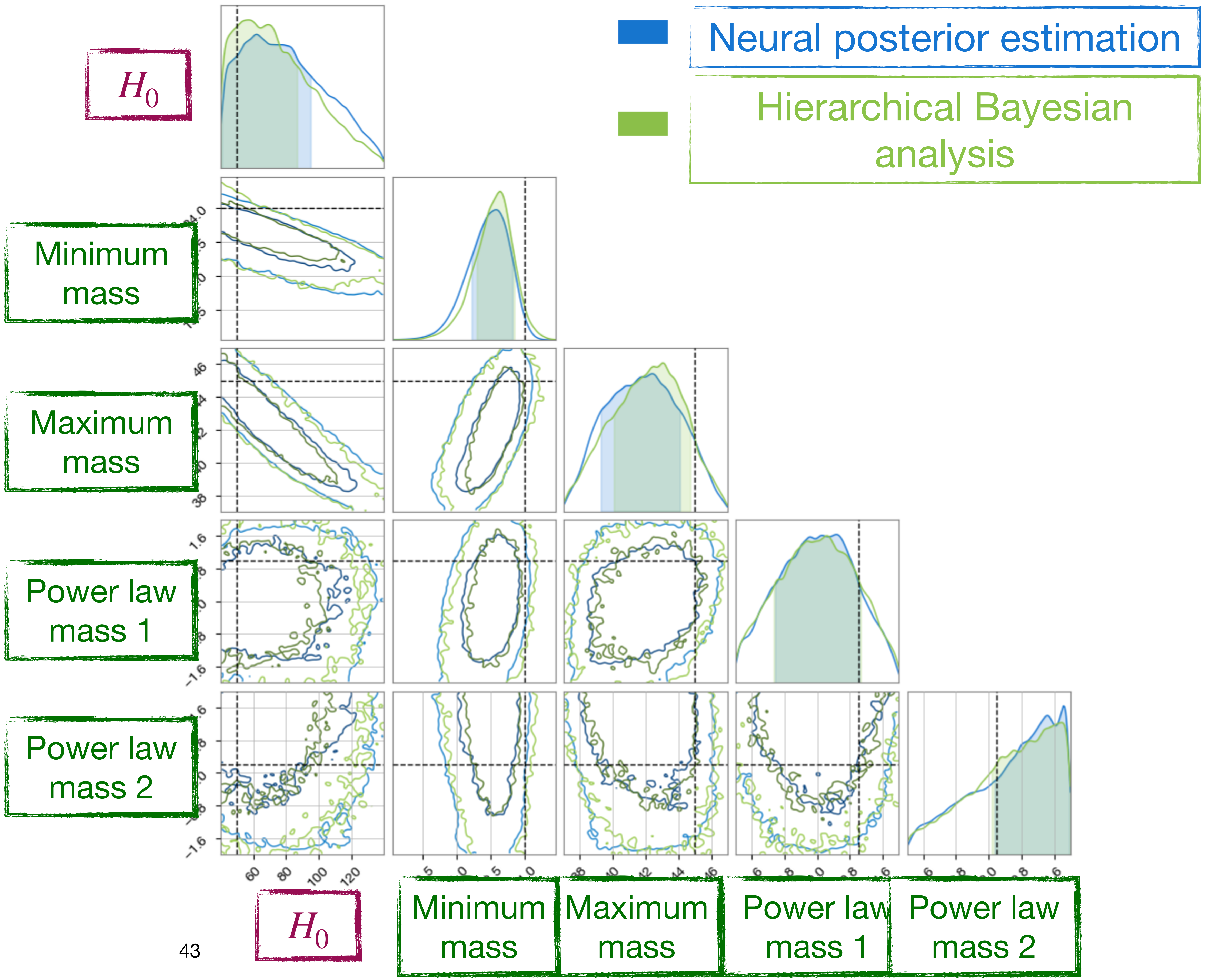
Approximate posterior from batch of events I

$$\mathbb{Q}(\Lambda \,|\, \mathscr{D}_C) \propto \mathbb{Q}(\Lambda)\, \mathbb{Q}(\mathscr{D}_C \,|\, \Lambda) \approx \frac{1}{\mathbb{Q}(\Lambda)^{N_{\rm obs}-1}} \prod_{I \subset C} \mathbb{Q}(\Lambda \,|\, \mathscr{D}_{i \in I})$$

C the GW catalog

Results (60 simulated events)

- **Simulated data**: truncated power law mass model
- **LIGO detectors at O1 sensitivity**
- **3 min vs. 8h computation time**
- **Discrepancy** on a fraction of populations

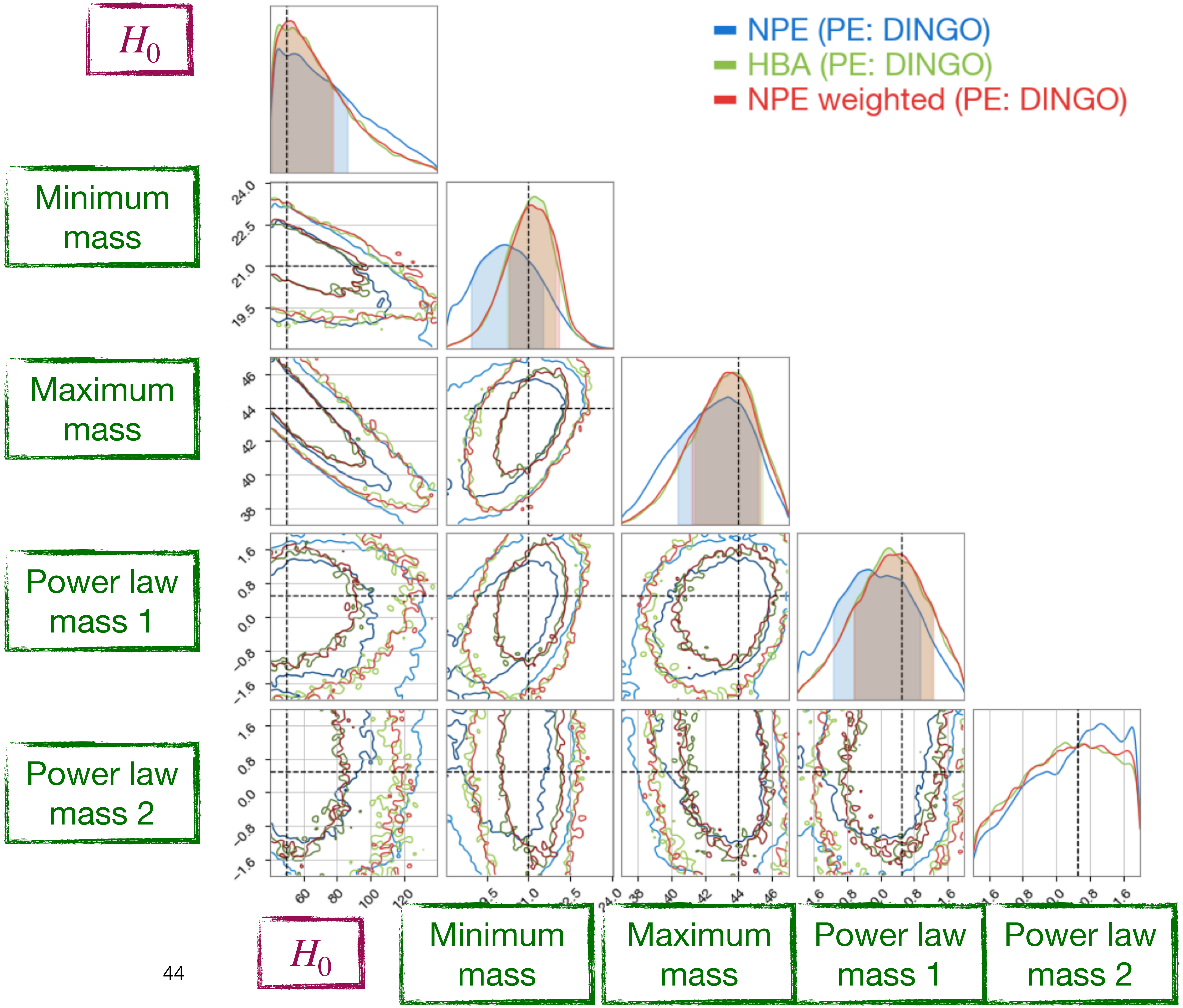


Importance sampling for SBI

- **Conventional** and **SBI** can **differ**
- SBI always **more conservative** ⟺ larger support than conventional
- If **likelihood available**

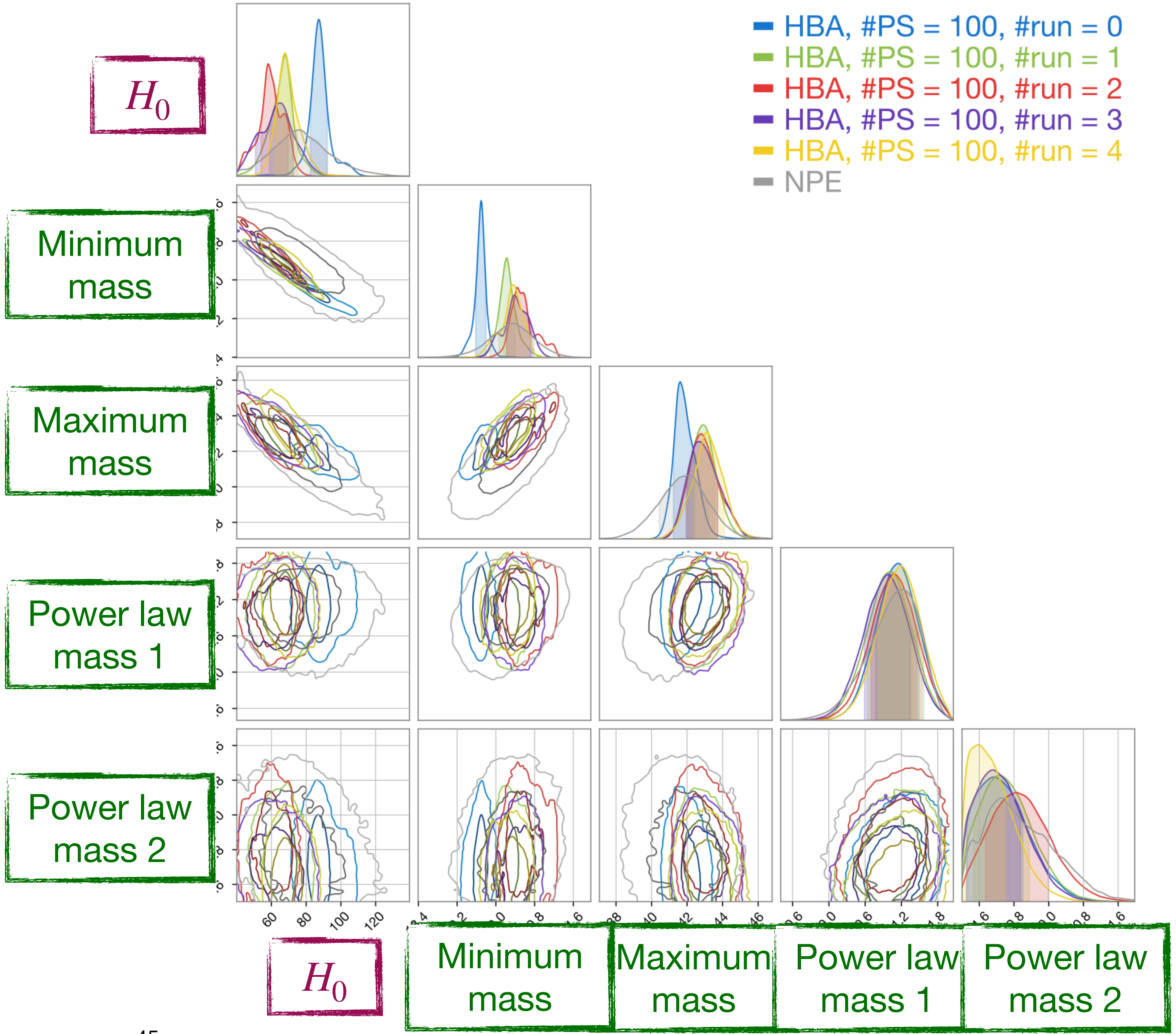
  ➡**Importance sampling**

  ➡**Fewer evaluations**



SBI for hierarchical GW studies

- **Complete freedom in representation of data**
- Fully SBI
  - **Likelihood-free**
  - Marginalize over **insufficient data representation**
- **Fast inference**
  - Validation via **coverage tests, etc.**



Conclusions: simulation-based inference in GW science

- **Promising avenue** for current and future GW data
  - **Successful demonstration** of **binary black hole and binary neutron star data**
    - **Fast, pre-merger parameter estimation**
    - **Even with importance sampling:** $1\,\%$ **of likelihood calls**
  - **Proof-of-concept for hierarchical studies**
- **Open questions**
  - **Full SBI results without importance sampling?**
  - **Promise of amortized inference?**
  - **Data reduction for higher-order modes?**
  - **Scaling of population inference for large datasets?**
  - **Re-training models for different population models?**

Back-up slides

Loss of group equivarient posterior

2111.13139

$$\mathcal{L}_{\text{GNPE}} = \mathbb{E}_{p(\theta)}\mathbb{E}_{p(x|\theta)}\mathbb{E}_{p(\hat{g}|\theta)}\left[-\log q\left(\theta|T_{\hat{g}^{-1}}x, \hat{g}\right)\right]$$