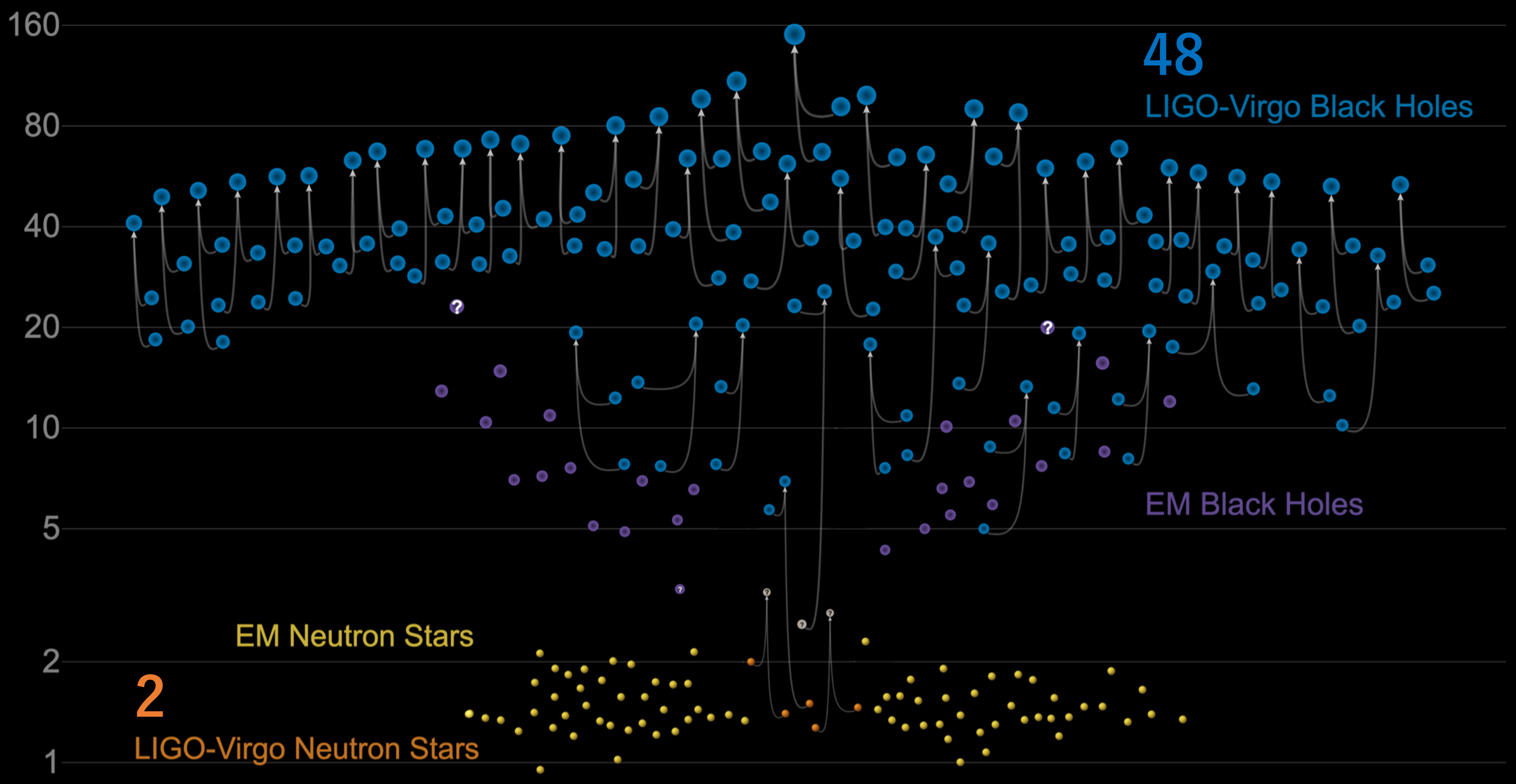
Rapid Parameter Estimation of Gravitational Waves from Compact Binary Coalescence

Soichiro Morisaki

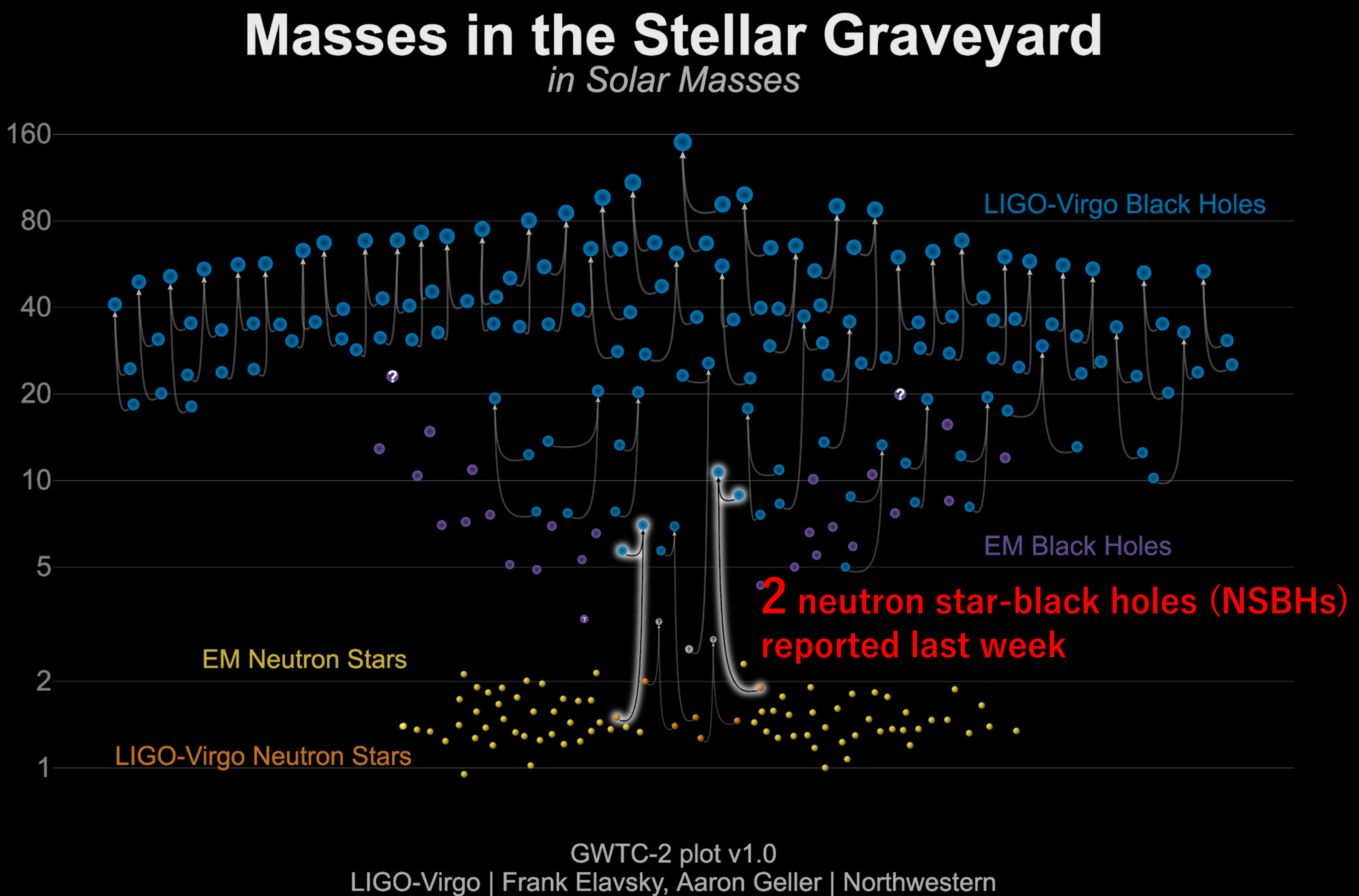
University of Wisconsin-Milwaukee

Masses in the Stellar Graveyard

in Solar Masses



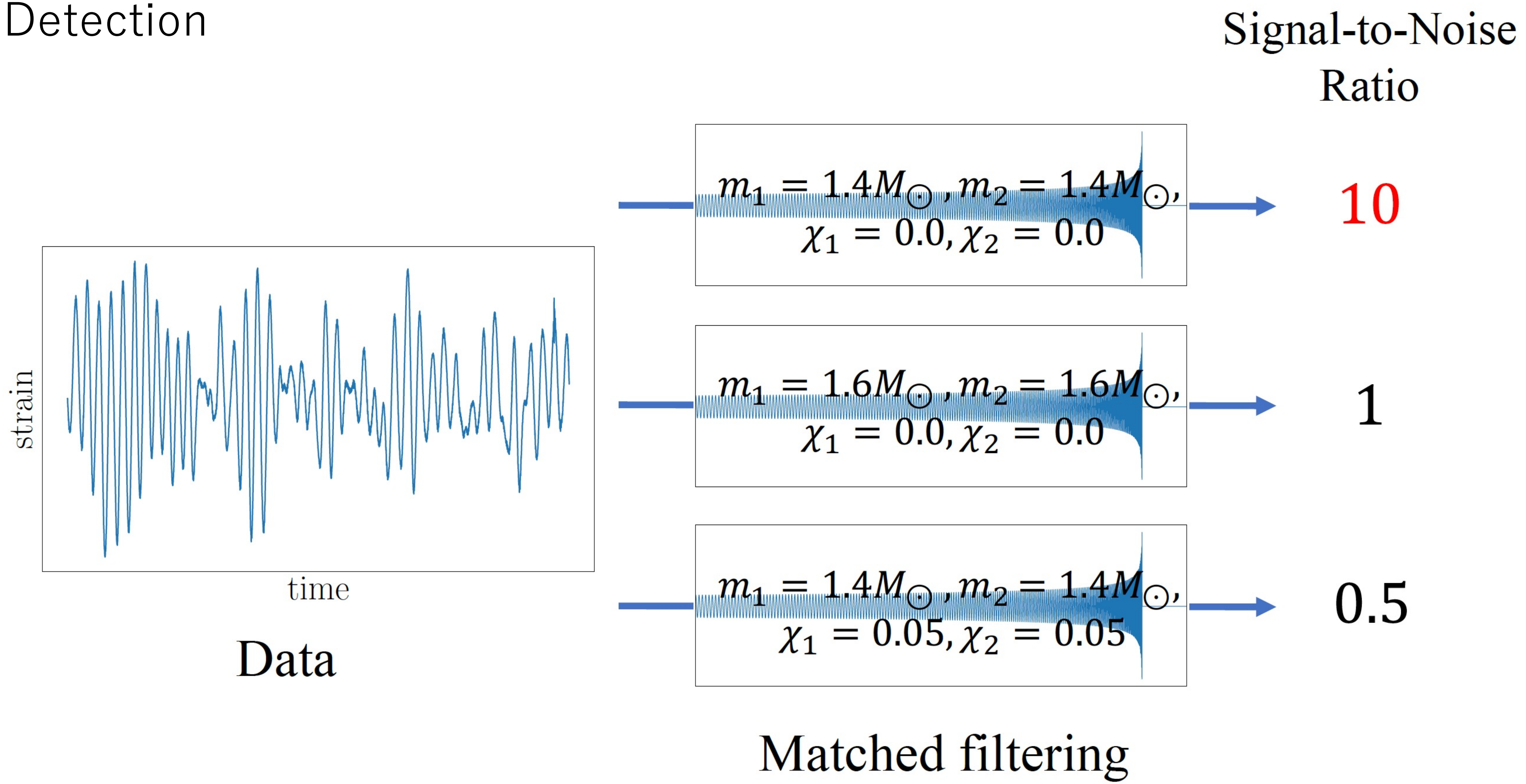
GWTC-2 plot v1.0
LIGO-Virgo | Frank Elavsky, Aaron Geller | Northwestern



Masses in the Stellar Graveyard
in Solar Masses
160
80
40
20
10
5
2
1
LIGO-Virgo Black Holes
EM Black Holes
EM Neutron Stars
LIGO-Virgo Neutron Stars
2 neutron star-black holes (NSBHs)
reported last week
GWTC-2 plot v1.0
LIGO-Virgo | Frank Elavsky, Aaron Geller | Northwestern

Analysis workflow of compact binary coalescence

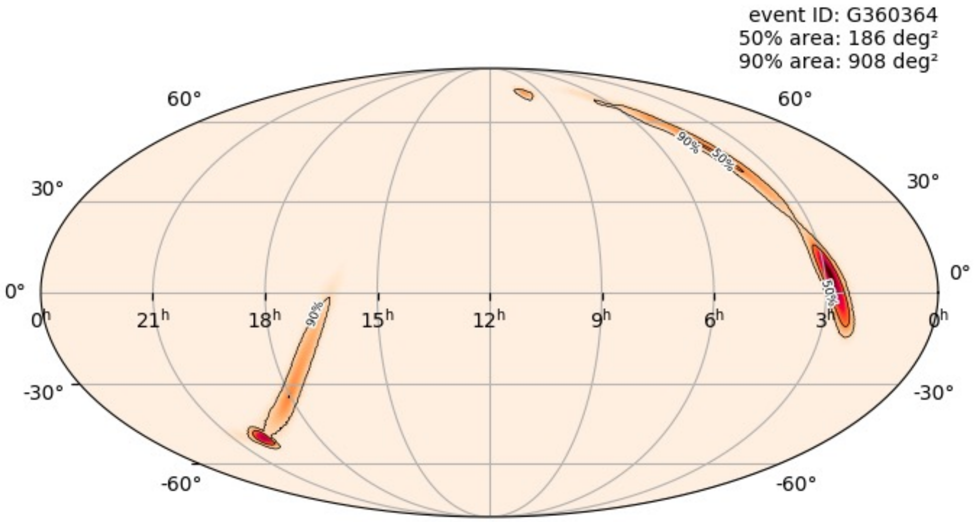
(1) Detection



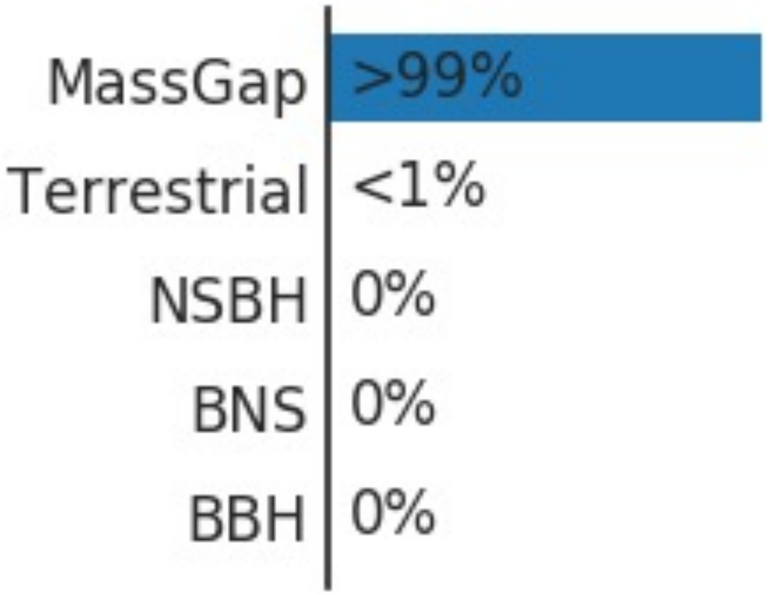
Analysis workflow of compact binary coalescence

(2) Low-latency inference

Produced in ~seconds, but subject to search bias.



Bayestar skymap of GW200115



Low-latency classification of GW200115

Analysis workflow of compact binary coalescence

(3) Bayesian parameter estimation (**PE**)

- Stochastic sampling of posterior,

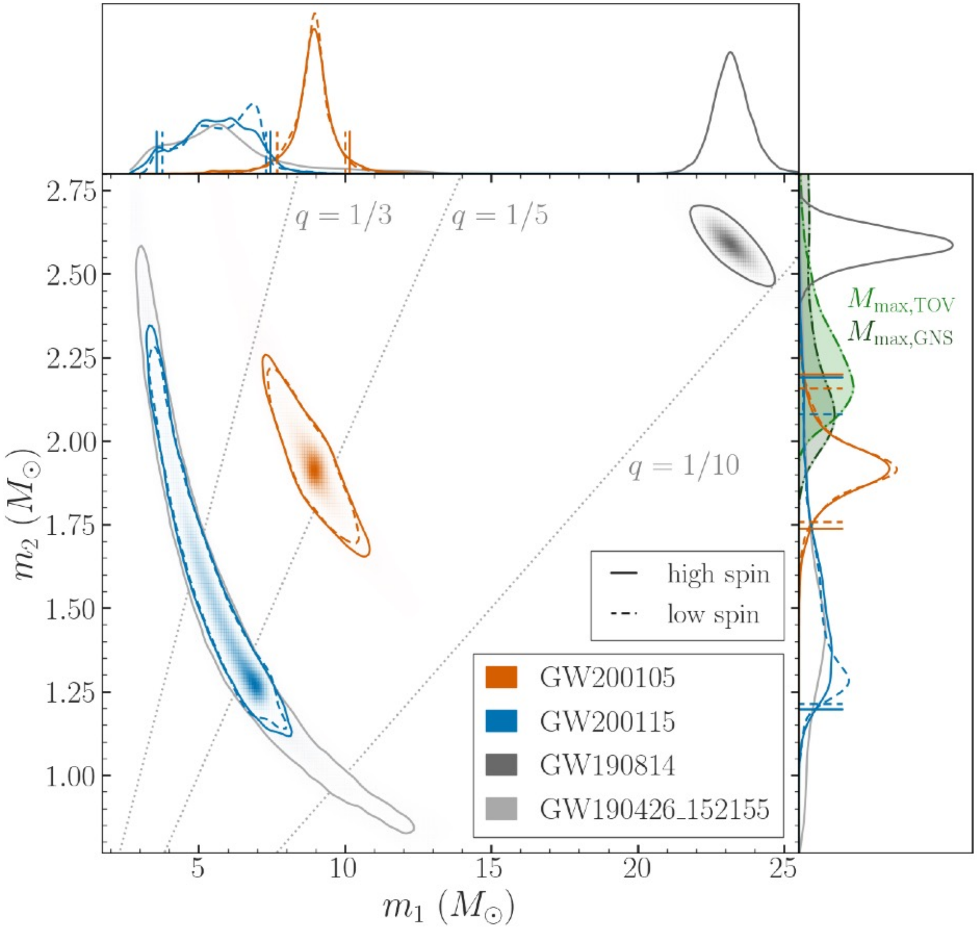
$$p(\theta|d) \propto \underbrace{p(\theta)}_{\text{prior}}\underbrace{p(d|\theta)}_{\text{likelihood}}.$$

θ: source parameters

(masses, spins, location etc.)

d: data

- More accurate than the initial estimates.
  - ➢Take into account measurement errors.
  - ➢Finer resolution of masses and spins
  - ➢Include more physics (precession, higher-order moments, tides etc.)



Estimated masses of GW200105 and GW200115

R. Abbott *et al.*, Astrophys. J. Lett. **915** (2021).

PE is computationally costly

The median latency in O3 is 59 hours.

PE of binary neutron star (BNS) can take ~years without any approximate methods.

Too slow for follow-up observations of BNS or NSBH.

More detections make it tough.

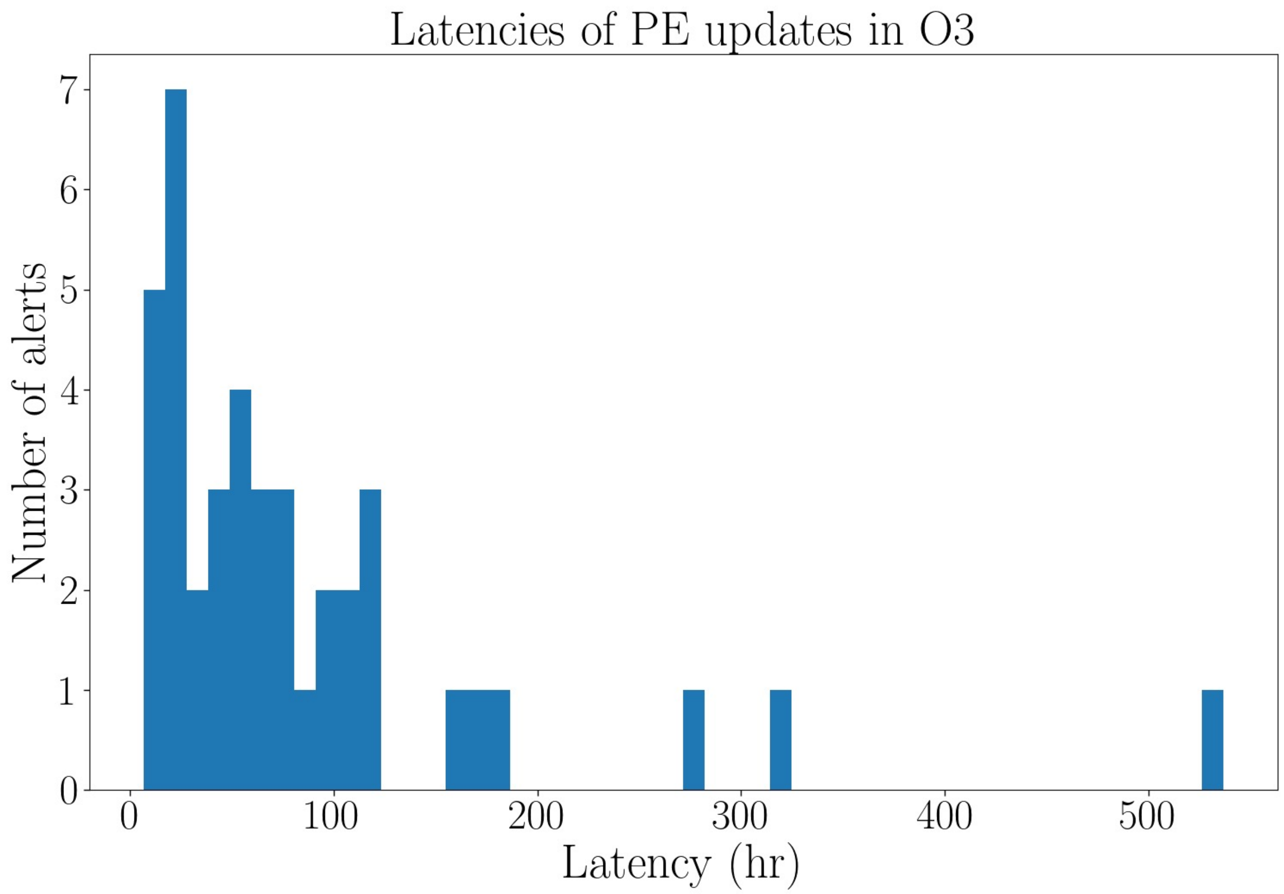


Figure: Latencies of PE updates in O3

Why is PE costly?

PE computes likelihood more than millions of times.

$$\ln p(d|\theta) = -\frac{1}{2}\big(d - h(\theta), d - h(\theta)\big) = \big(d, h(\theta)\big) - \frac{1}{2}\big(h(\theta), h(\theta)\big) + \text{const.}$$

$h(\theta)$: template waveform

The inner products involve the frequency-domain waveform $\tilde{h}(f_l;\theta)$.

$$\big(d, h(\theta)\big) = \frac{4}{T}\Re\left[\sum_{l=1}^{\lfloor (N-1)/2 \rfloor} \frac{\tilde{d}_l^* \tilde{h}(f_l;\theta)}{S_l}\right], \qquad \big(h(\theta), h(\theta)\big) = \frac{4}{T}\Re\left[\sum_{l=1}^{\lfloor (N-1)/2 \rfloor} \frac{\left|\tilde{h}(f_l;\theta)\right|^2}{S_l}\right].$$

S_l: Power spectral density at the l-th frequency bin

Why is PE costly?

Typically, the evaluations of $\tilde{h}(f_l;\theta)$ are computationally expensive.

The cost is proportional to the number of frequency samples, which is ~ **(duration of signal) × (frequency range of signal).**

→ It is huge for BNS or NSBH signal, which is long and high-frequency.

Two complementary methods to reduce the number of frequency samples where waveforms are evaluated:

1. Focused Reduced Order Quadrature (FROQ)

    SM and Vivien Raymond, Phys. Rev. D **102**, 104020 (2020).

2. Multi-banding

    SM, arXiv:2104.07813 (2021).

1. Focused Reduced Order Quadrature (FROQ)

   SM and Vivien Raymond, Phys. Rev. D **102**, 104020 (2020).

2. Multi-banding

   SM, arXiv:2104.07813 (2021).

Reduced Order Quadrature (ROQ)

Waveforms = vectors in L-dimensional space L: number of frequency samples

Subspace of waveforms is much lower dimensional.

Approximate waveforms with K ($\ll L$) basis vectors $B_k(f_l)$ ($k = 1,2, \dots, K$) :

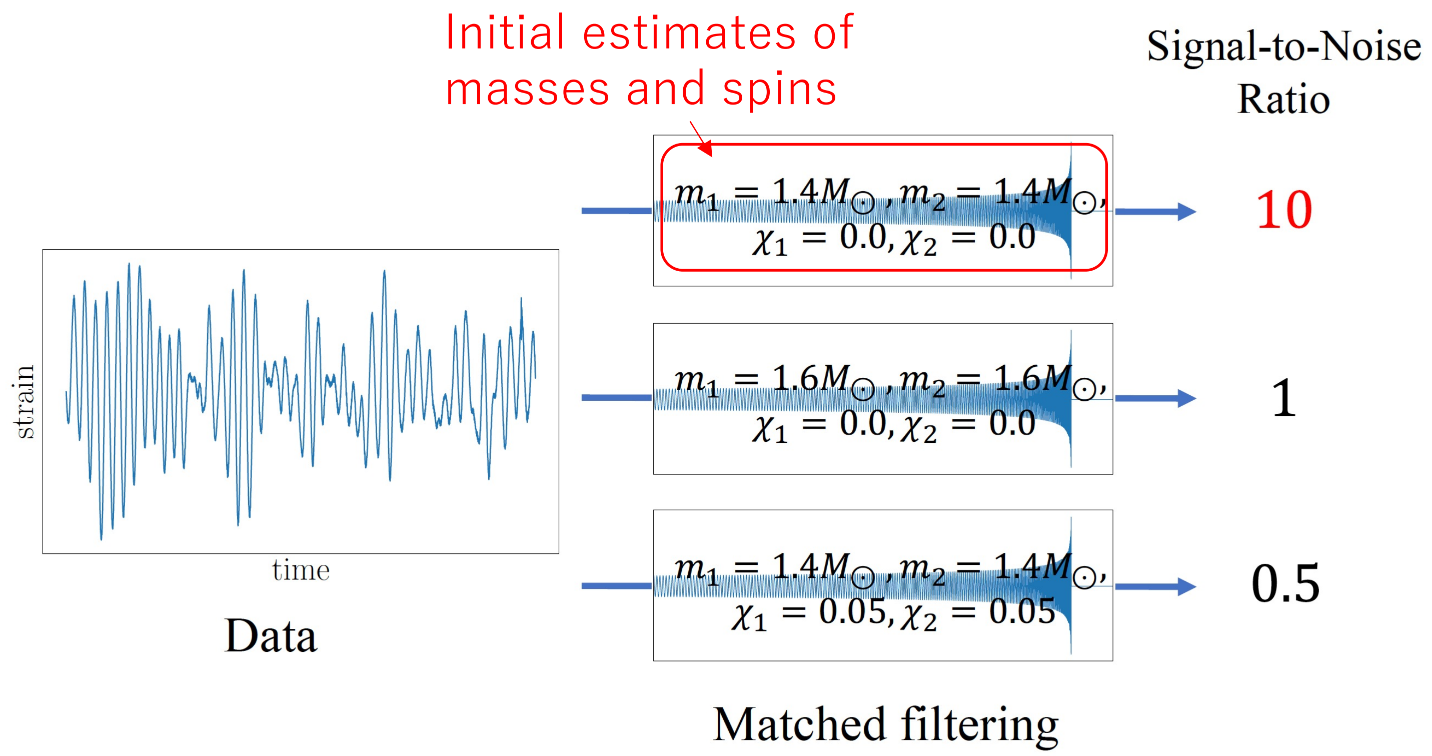
$$\tilde{h}(f_l) \simeq \sum_{k=1}^{K} \tilde{h}(F_k) B_k(f_l) \,.$$

The number of waveform evaluations is reduced to K $\rightarrow$ $\sim L/K$ speed up

Speed-up gains of $\mathcal{O}(10^2)$ for BNS, reducing the run time of PE to ~hours.

P. Canizares *et al.*, Phys. Rev. Lett. **114**, 071104 (2015).
R. Smith *et al.*, Phys. Rev. D **94**, 044031 (2016).

Focused Reduced Order Quadrature (FROQ)



The parameter space can be significantly restricted from the initial estimates.

Waveforms in the restricted space have similar morphologies.

→ The subspace of waveforms is extremely low dimensional.

→ Much smaller K, and much larger speed-up gains of L/K.

$\mu^1 - \mu^2$

Combinations whose initial estimates are reliable:

$$\mu^1 = 0.974\psi^0 + 0.209\psi^2 + 0.0840\psi^3,$$

$$\mu^2 = -0.221\psi^0 + 0.823\psi^2 + 0.524\psi^3.$$

ψ^0, ψ^2, ψ^3 are the coefficients in the Post-Newtonian expansion of GW's phase depending on masses (m_1, m_2) and spin components in parallel with the orbital angular momentum (χ_1, χ_2) ,

$$\Psi(f) = \psi^0 \left(\frac{f}{f_{\mathrm{ref}}}\right)^{-\frac{5}{3}} + \psi^2 \left(\frac{f}{f_{\mathrm{ref}}}\right)^{-1} + \psi^3 \left(\frac{f}{f_{\mathrm{ref}}}\right)^{-\frac{2}{3}},$$

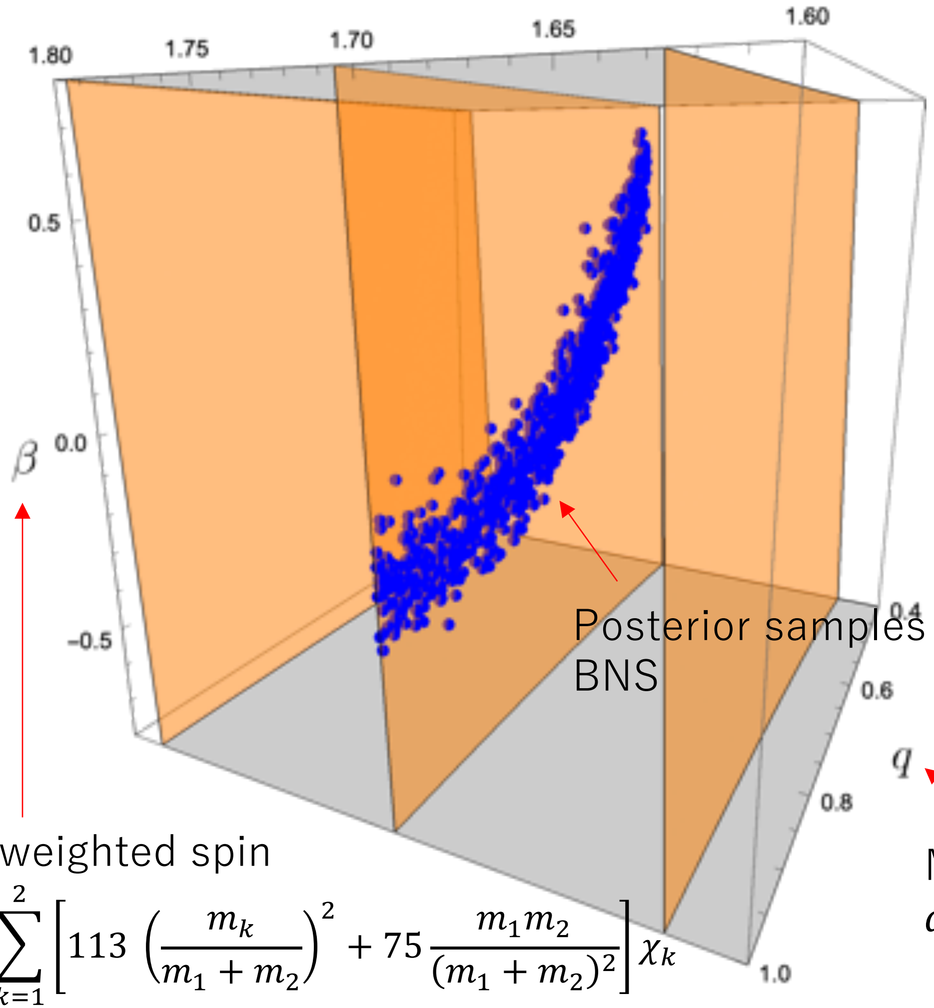
where $f_{\mathrm{ref}} = 200\ \mathrm{Hz}$.

Their range is determined based on search biases and statistical errors.

$\mu^1 - \mu^2$

Chirp mass

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



Posterior samples of BNS

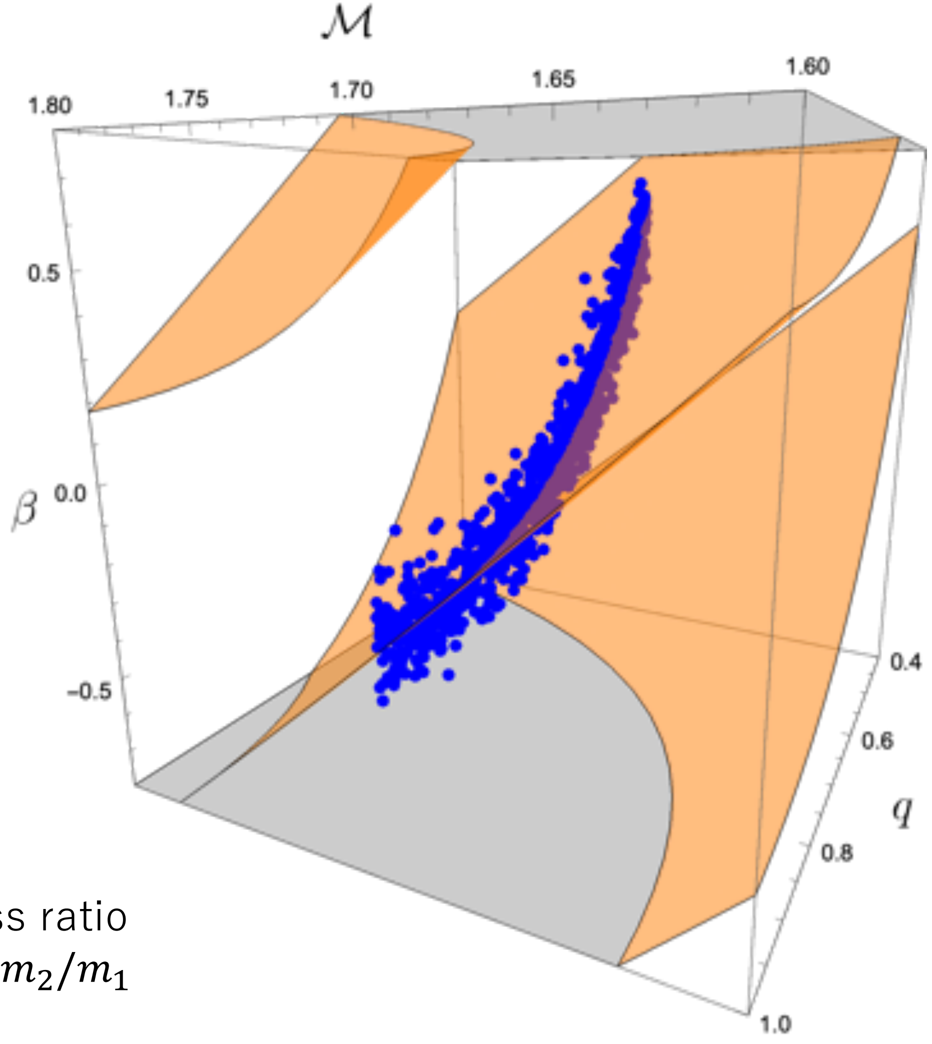
Mass-weighted spin

$$\beta = \frac{1}{12} \sum_{k=1}^{2} \left[113 \left(\frac{m_k}{m_1 + m_2} \right)^2 + 75 \frac{m_1 m_2}{(m_1 + m_2)^2} \right] \chi_k$$

Mass ratio

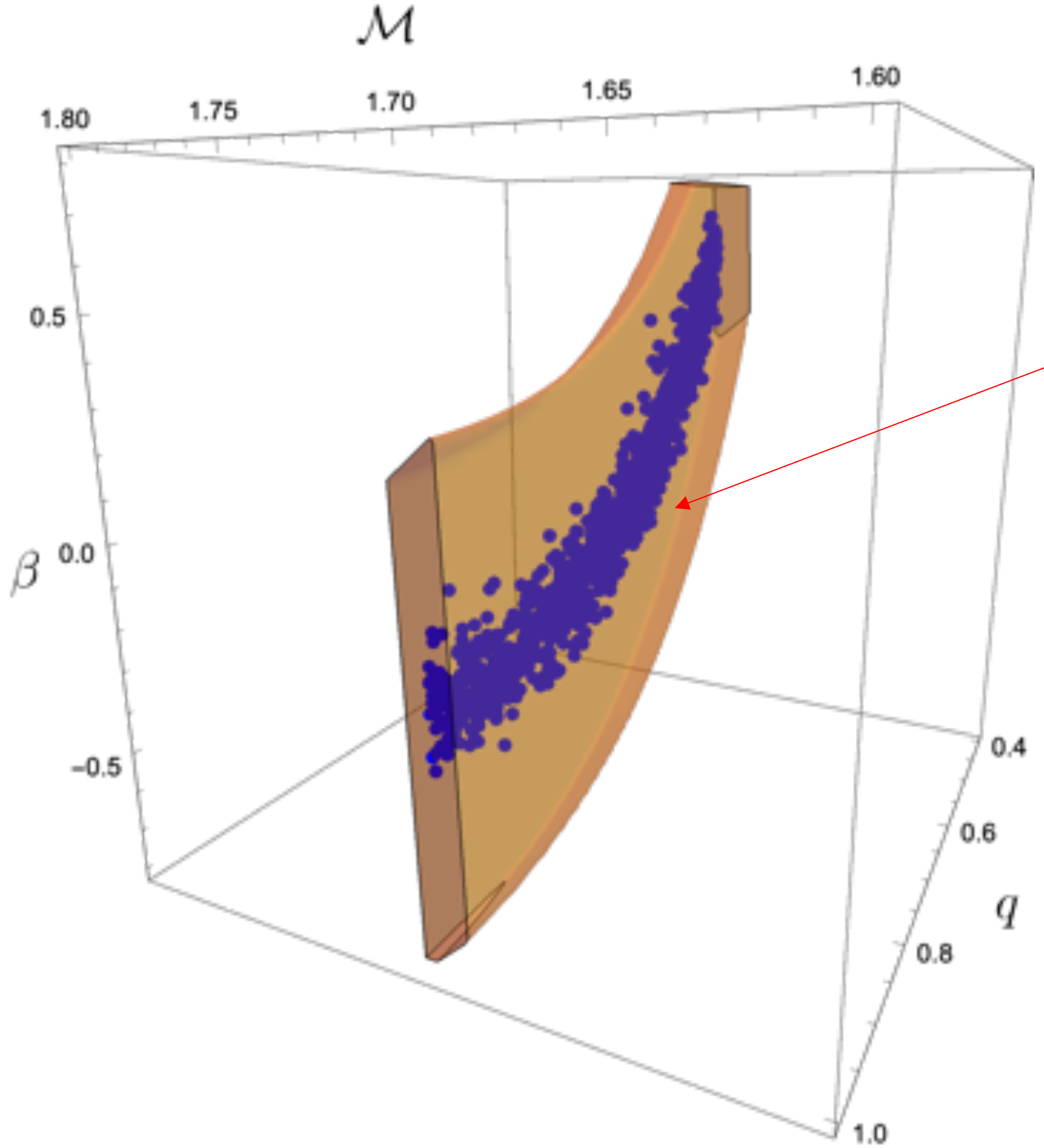
$q = m_2 / m_1$

μ^1-constant planes



μ^2-constant planes

$\mu^1 - \mu^2$



Narrow, but efficiently encompass samples

Restricted parameter space

Basis sizes

Table: Basis sizes and speed-up gains

$m_{1,\mathrm{t}} = m_{2,\mathrm{t}}$	Low-spin Prior range	Low-spin Size	Low-spin Speedup	High-spin Prior range	High-spin Size	High-spin Speedup
$1M_\odot$	$446.10 \leq \mu^1 \leq 446.38$ $-95.0 \leq \mu^2 \leq -89.9$	21	28000	$445.41 \leq \mu^1 \leq 446.75$ $-98.5 \leq \mu^2 \leq -86.5$	63	9800
$1.4M_\odot$	$254.79 \leq \mu^1 \leq 255.09$ $-60.8 \leq \mu^2 \leq -55.5$	21	12000	$254.31 \leq \mu^1 \leq 255.36$ $-62.9 \leq \mu^2 \leq -52.9$	43	5800
$2M_\odot$	$140.49 \leq \mu^1 \leq 140.76$ $-39.9 \leq \mu^2 \leq -35.5$	21	4600	$140.19 \leq \mu^1 \leq 141.05$ $-41.4 \leq \mu^2 \leq -32.6$	37	2600

- Waveform model：TaylorF2
- $m_1, m_2 < 3M_\odot$
- Low-spin: $|\chi| < 0.05$,   High-spin: $|\chi| < 0.7$
- Basis sizes are $\mathcal{O}(10)$, and speed-up gains are $\mathcal{O}(10^3) - \mathcal{O}(10^4)$.

Performance

- Thousands of simulated BNS signals in O2 data.
- The run time is
  < 16 minutes for 50%,
  < 29 minutes for 90% of the signals.
- No statistically significant biases from the restriction of parameter space.

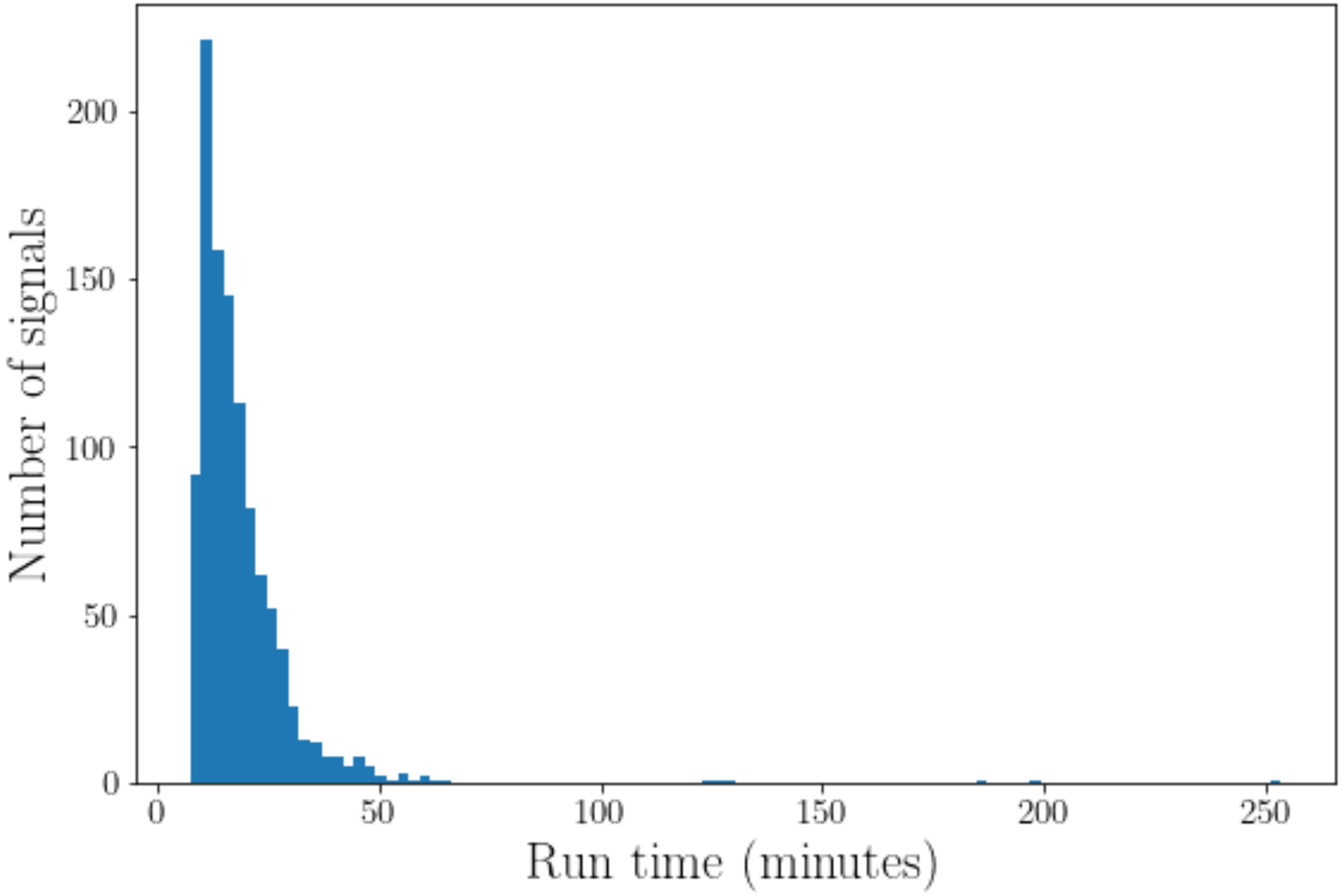


Figure: Histogram of PE run times

Searched area

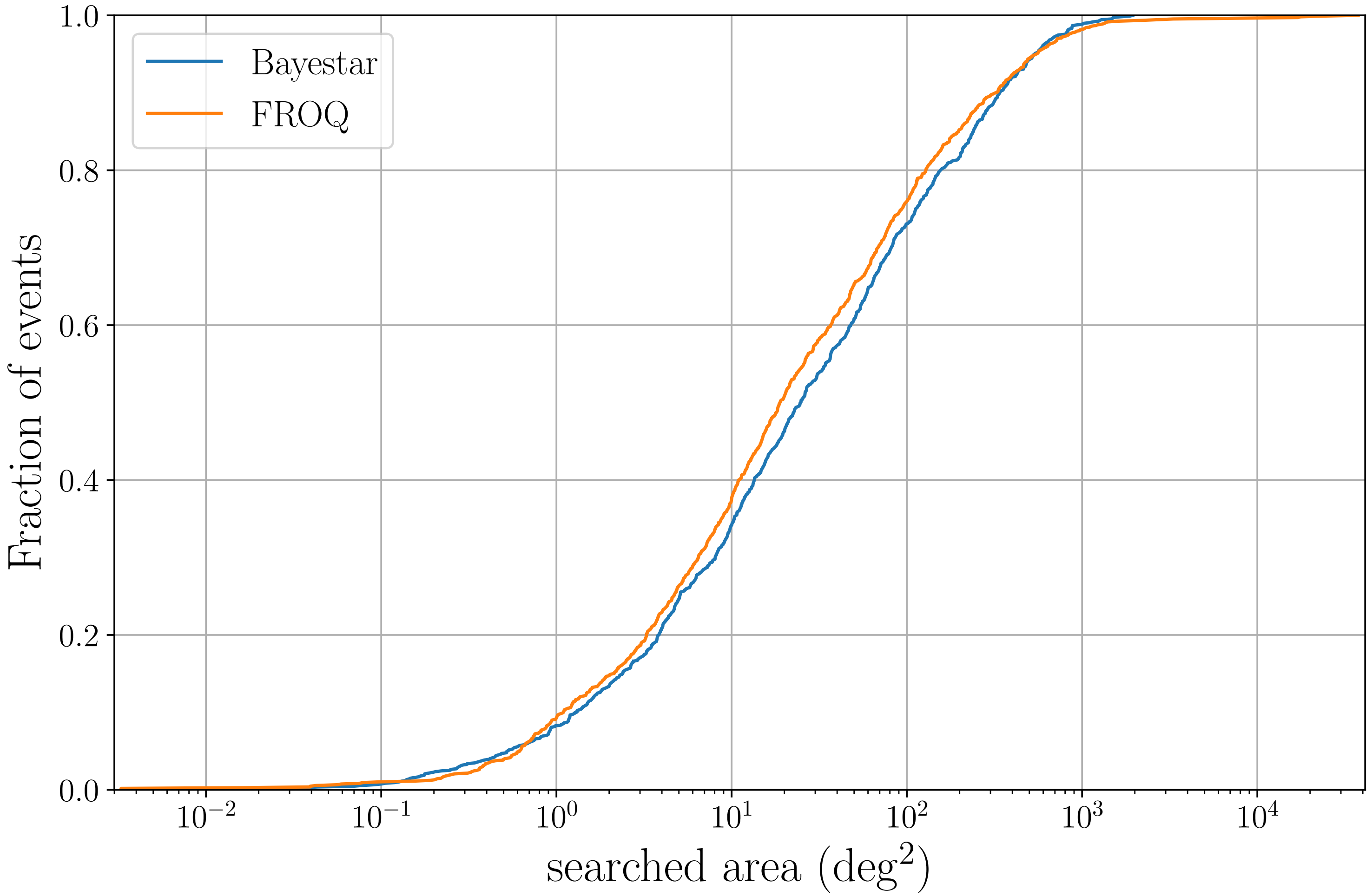


Figure: Cumulative distribution of searched areas

25 deg^2 (Bayestar) ⇨ 19 deg^2 (FROQ) at median

Short summary of FROQ

- FROQ builds a compact ROQ basis in narrow parameter space, whose range is restricted based on the initial estimates of masses and spins.

- FROQ relies on the two combinations of masses and spins, μ^1 and μ^2, whose initial estimates are reliable.

- FROQ speeds up PE of BNS signal by a factor of $\mathcal{O}(10^3) - \mathcal{O}(10^4)$, reducing its run time to a few tens of minutes.

- FROQ PE is accurate enough to update the initial localization and classification.

- μ^1 and μ^2 can also be used to make stochastic sampling more efficient (See **Eunsub's poster (P09)** for more detail). E. Lee, SM, and T. Hideyuki in preparation.

1. Focused Reduced Order Quadrature (FROQ)

   SM and Vivien Raymond, Phys. Rev. D **102**, 104020 (2020).

2. Multi-banding

   SM, arXiv:2104.07813 (2021).

ROQ is not easy-to-use

- Basis vectors $B_k(f_l)$ need to be computed offline with the greedy algorithm,

$$\tilde{h}(f_l) \simeq \sum_{k=1}^{K} \tilde{h}(F_k) B_k(f_l)\,.$$

- Computationally costly for long waveforms. Numerically very tough for the third-generation detectors (c.f. 1.4Msun-1.4Msun BNS signal is ~1.8 hours from 5 Hz).

- It needs to be done for each new waveform model.

Multi-banding

- The frequency interval is $1/T$, where T is the duration of data.
- Frequency increases with time.
→ Time-to-merger decreases with frequency.
- An increasing frequency interval can be used, reducing the number of waveform evaluations at high frequencies.

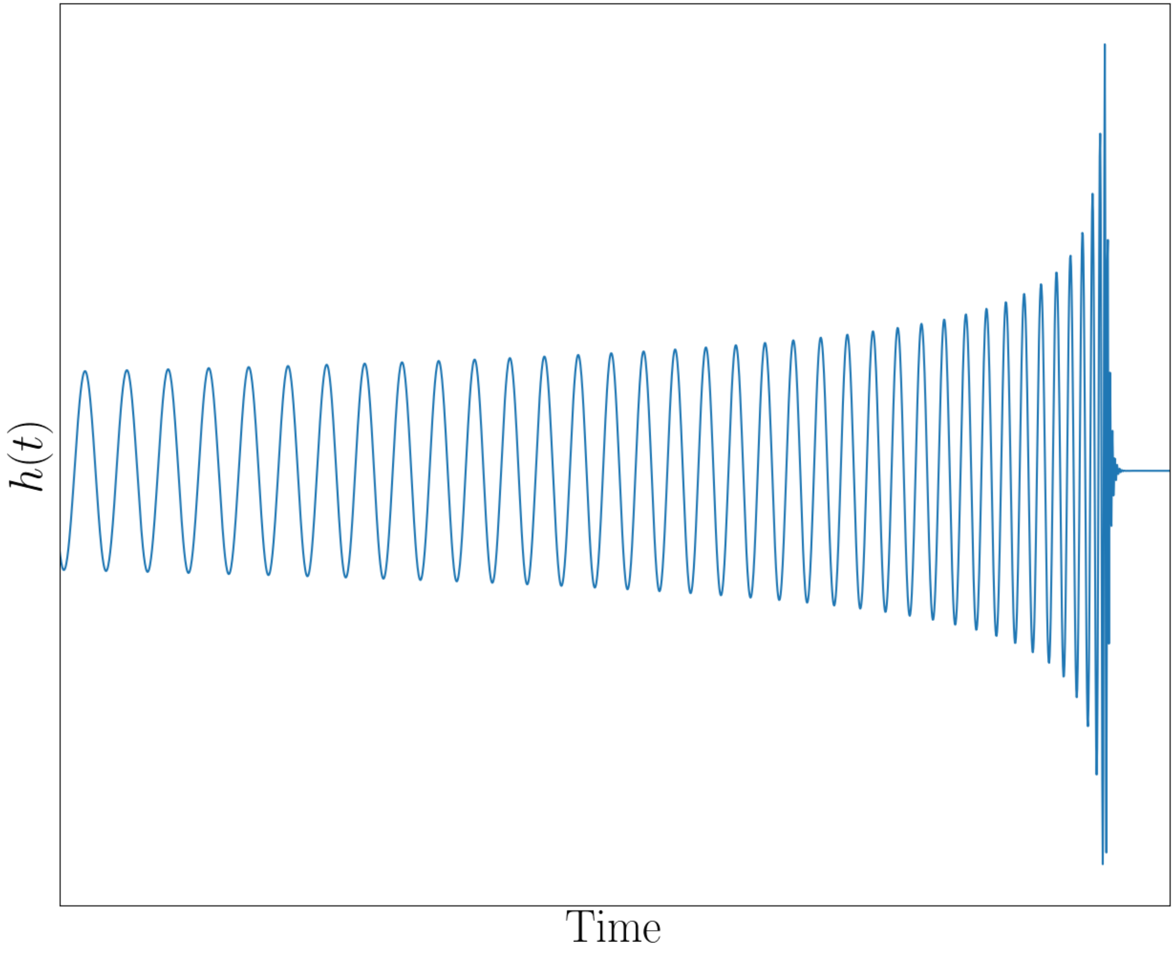


Figure: gravitational waves from compact binary coalescence

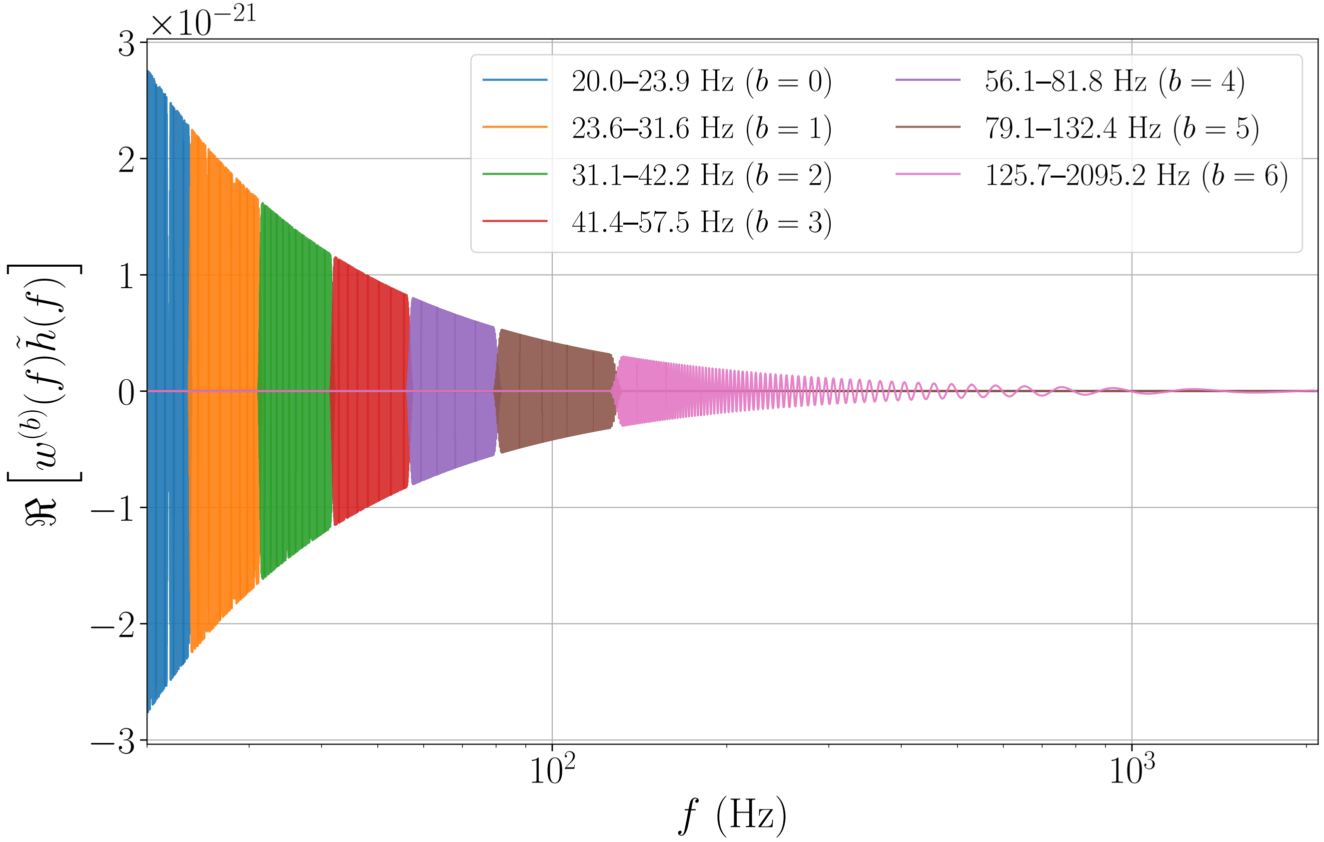
Multi-banding

Number of frequency bands

$$(d, h) = \sum_{b=0}^{B-1} \frac{4}{T} \Re \left[\sum_{l=1}^{\lfloor (N-1)/2 \rfloor} \underline{w^{(b)}(f_l)} \frac{\tilde{d}_l^* \tilde{h}(f_l)}{S_l} \right],$$

Window function of the b-th band

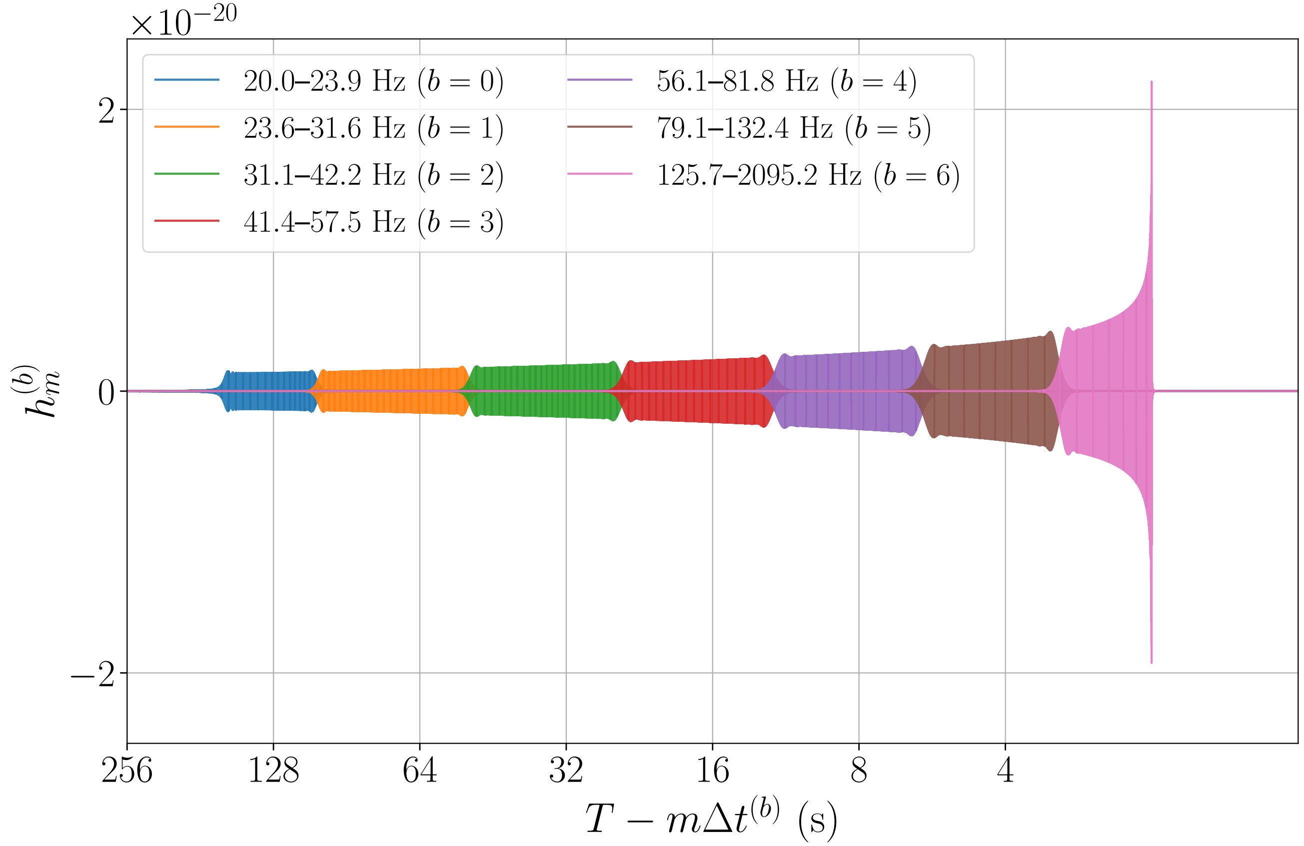
where $f_l = l/T$.



Multi-banding

$$\frac{4}{T}\Re\left[\sum_{l=1}^{\lfloor (N-1)/2 \rfloor} w^{(b)}(f_l)\frac{\tilde{d}_l^*\tilde{h}(f_l)}{S_l}\right] = 2\Delta t\sum_{m=0}^{N-1} D_m h_m^{(b)},$$

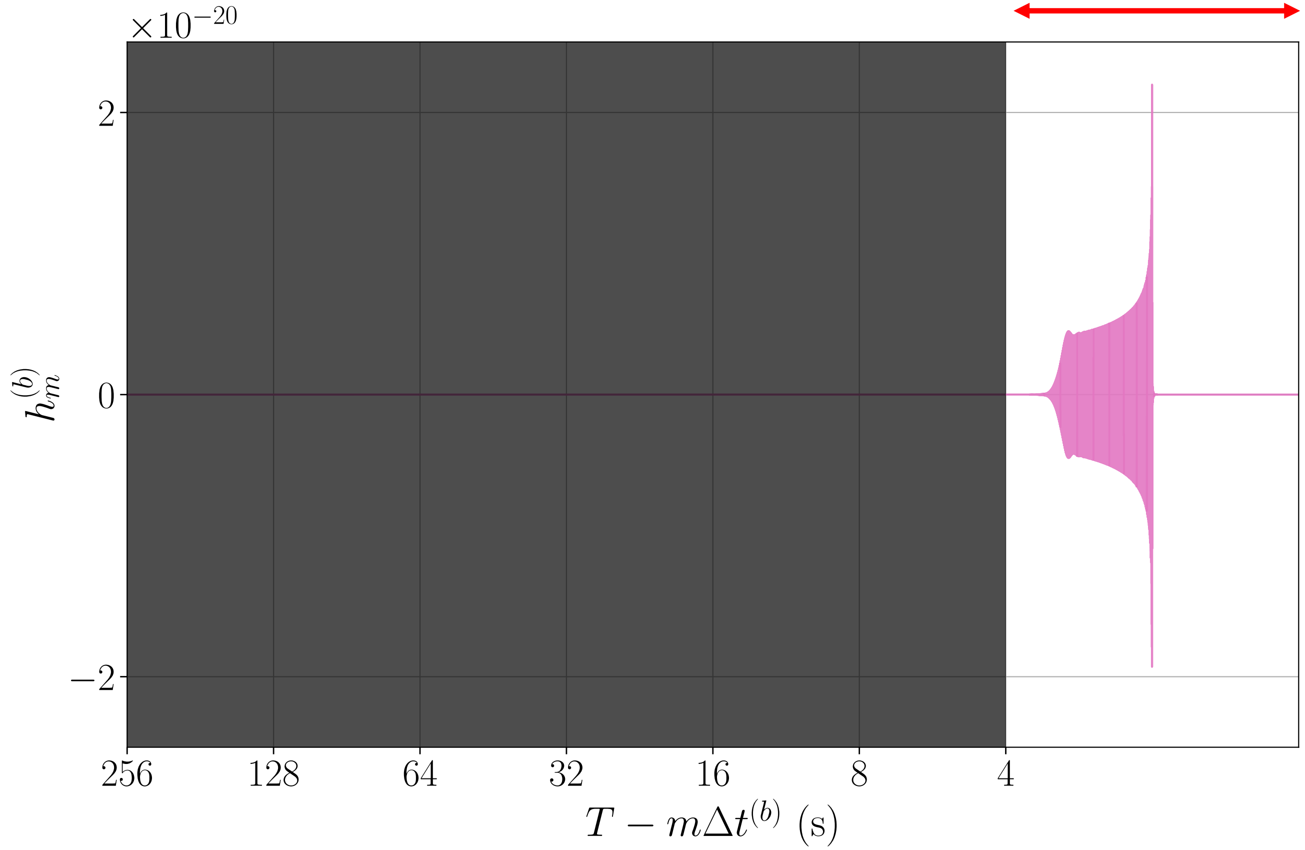
where D_m and $h_m^{(b)}$ are the inverse Fourier transform of $\tilde{d}_l/S_l$ and $w^{(b)}(f_l)\tilde{h}(f_l)$.



Multi-banding

$$2\Delta t \sum_{m=0}^{N-1} D_m h_m^{(b)} \simeq 2\Delta t \sum_{m=N-M^{(b)}}^{N-1} D_m h_m^{(b)}.$$

$T^{(b)} = M^{(b)}\Delta t$

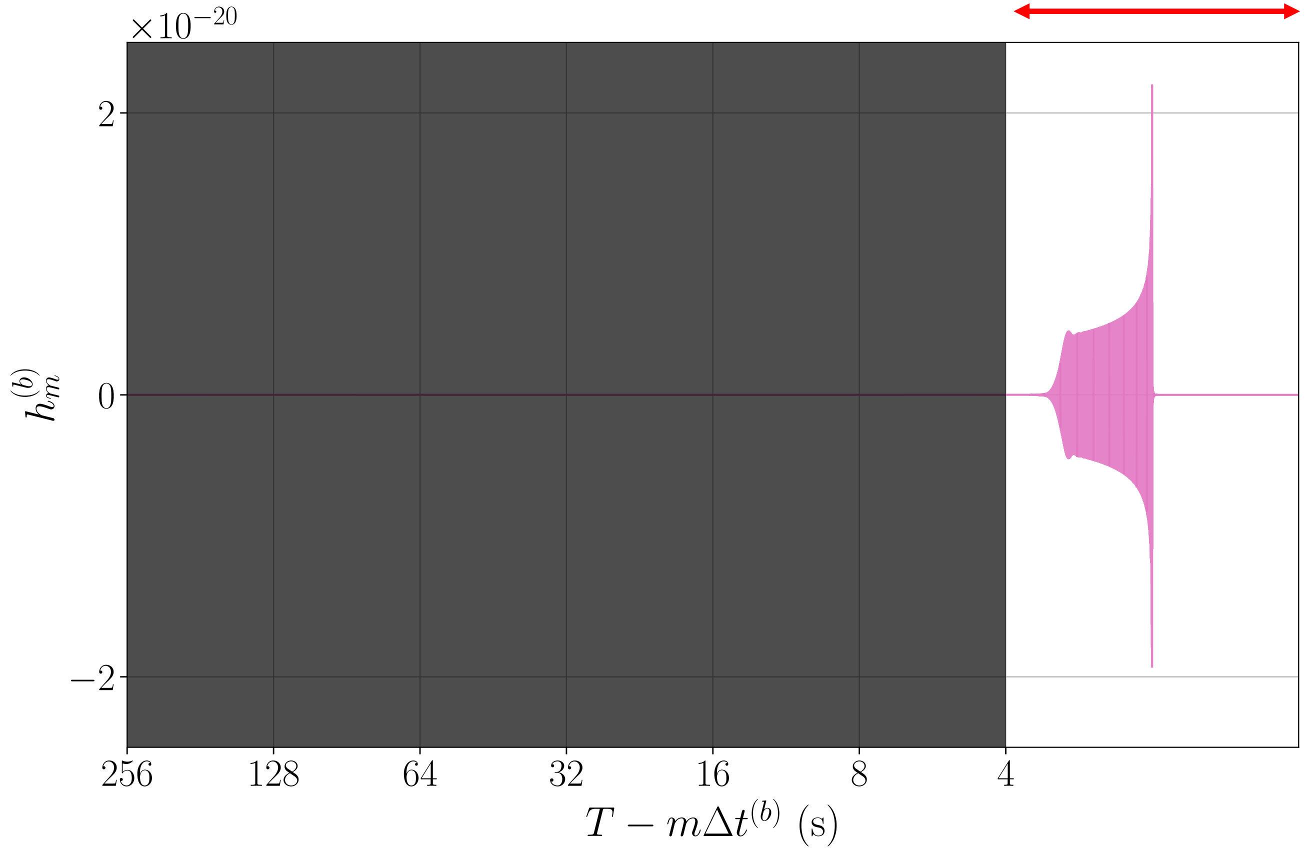


Multi-banding

$$2\Delta t \sum_{m=N-M^{(b)}}^{N-1} D_m h_m^{(b)} = \frac{4}{T^{(b)}} \Re\left[\sum_l w^{(b)}\left(f_l^{(b)}\right) \tilde{D}_l^{(b)*} \tilde{h}\left(f_l^{(b)}\right)\right],$$

where $T^{(b)} = M^{(b)}\Delta\mathrm{t}$ and $f_l^{(b)} = l/T^{(b)}$.

$T^{(b)} = M^{(b)}\Delta t$



How to determine frequency bands

Given $T^{(b)}$, the starting frequency $f^{(b)}$ is determined so that the waveform is vanishing at $t < T - T^{(b)}$.

$$\tau\left(f^{(b)}\right) + L\sqrt{-\tau'\left(f^{(b)}\right)} = T^{(b)} + t_{\mathrm{c,min}} - T,$$

where $\tau(f)$ is the time-to-merger, $t_{\mathrm{c,min}}$ is the minimum of coalescence time, and $L \gg 1$ ($L = 5$ is sufficient).

The frequency bands can be constructed on the fly from the minimum chirp mass in the prior range.

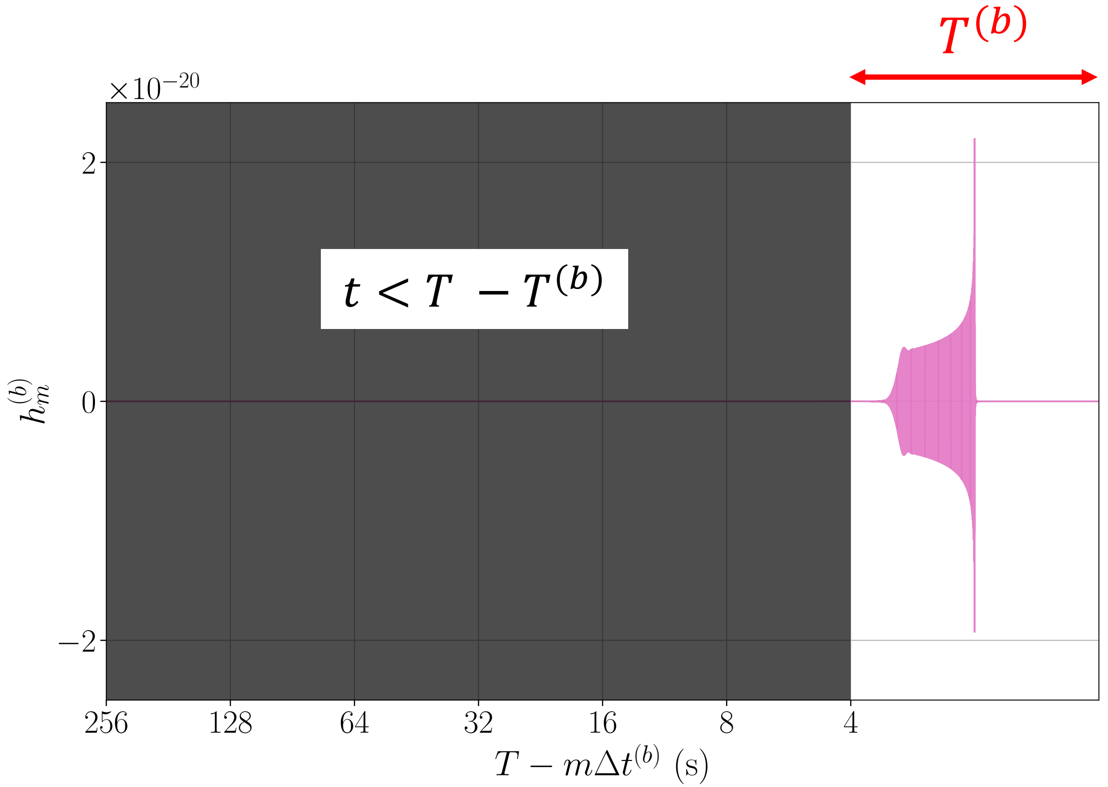


Figure: Windowed waveform in the time domain

Speed-up gain

Table: Speed-up gains for 1.4Msun-1.4Msun BNS signal for various low-frequency cutoffs f_{low}

f_{low} (Hz)	T (s)	Number of original samples	Reduction of samples	Speed-up gain
20	256	5.2×10^5	4.5×10	5.1×10
10	1024	2.1×10^6	1.2×10^2	1.5×10^2
5	8192	1.7×10^7	4.4×10^2	4.9×10^2

LVK

Third-generation

- The speed-up gain is larger for lower f_{low}, whose duration T is longer.
- The speed-up gain is ~50 for $f_{\mathrm{low}} = 20\ \mathrm{Hz}$ (used for LVK) and ~500 for $f_{\mathrm{low}} = 5\ \mathrm{Hz}$ (appropriate for the third-generation detectors).

Short summary of multi-banding

- Multi-banding exploits the chirping nature of signal, whose frequency increases with time.

- For BNS signal, the speed-up gain is ~ 50 for $f_{\mathrm{low}} = 20~\mathrm{Hz}$ and ~ 500 for $f_{\mathrm{low}} = 5~\mathrm{Hz}$.

- It is not as fast as FROQ, but it does not require any offline preparations.

Summary

- Parameter estimation of compact binary coalescence is computationally costly, and can take ~years for BNS signal.
- We developed two complementary methods to reduce its cost: Focused Reduced Order Quadrature (FROQ) and Multi-banding.
- FROQ speeds up PE of BNS signal by a factor of $\mathcal{O}(10^3) - \mathcal{O}(10^4)$, reducing its run time to a few tens of minutes.
- Multi-banding is not as fast as FROQ, but it is easy-to-use as it does not require any offline preparations.

Extra slides

Range of $\mu^1 - \mu^2$

- Mismatch between trigger and true waveforms

$$\mathbf{Mismatch} \simeq \frac{1}{2}\tilde{\Gamma}_{\alpha\beta}\left(\hat{\psi}^{\alpha} - \psi_t^{\alpha}\right)\left(\hat{\psi}^{\beta} - \psi_t^{\beta}\right) < 0.03$$

$\tilde{\Gamma}_{\alpha\beta}$: Fisher matrix $\hat{\psi}^{\alpha}$: True values ψ_t^{α}: Trigger values

- Statistical errors

$$\tilde{\Gamma}_{\alpha\beta}\left(\psi^{\alpha} - \hat{\psi}^{\alpha}\right)\left(\psi^{\beta} - \hat{\psi}^{\beta}\right) < \left(\frac{N}{\rho_{\mathrm{net}}}\right)^2$$

N: 99.9% upper limit of χ^2 with 4 d.o.f. ρ_{net}: Network SNR (12 is applied)

- Prior of masses and spins

Consistency

- X% credible intervals should encompass the true values for X% of the simulated signals.

- The fraction of signals as a function of credible level should be a diagonal line.

- The deviations from a diagonal line are consistent with statistical errors (p-values are 0.13 – 0.84).

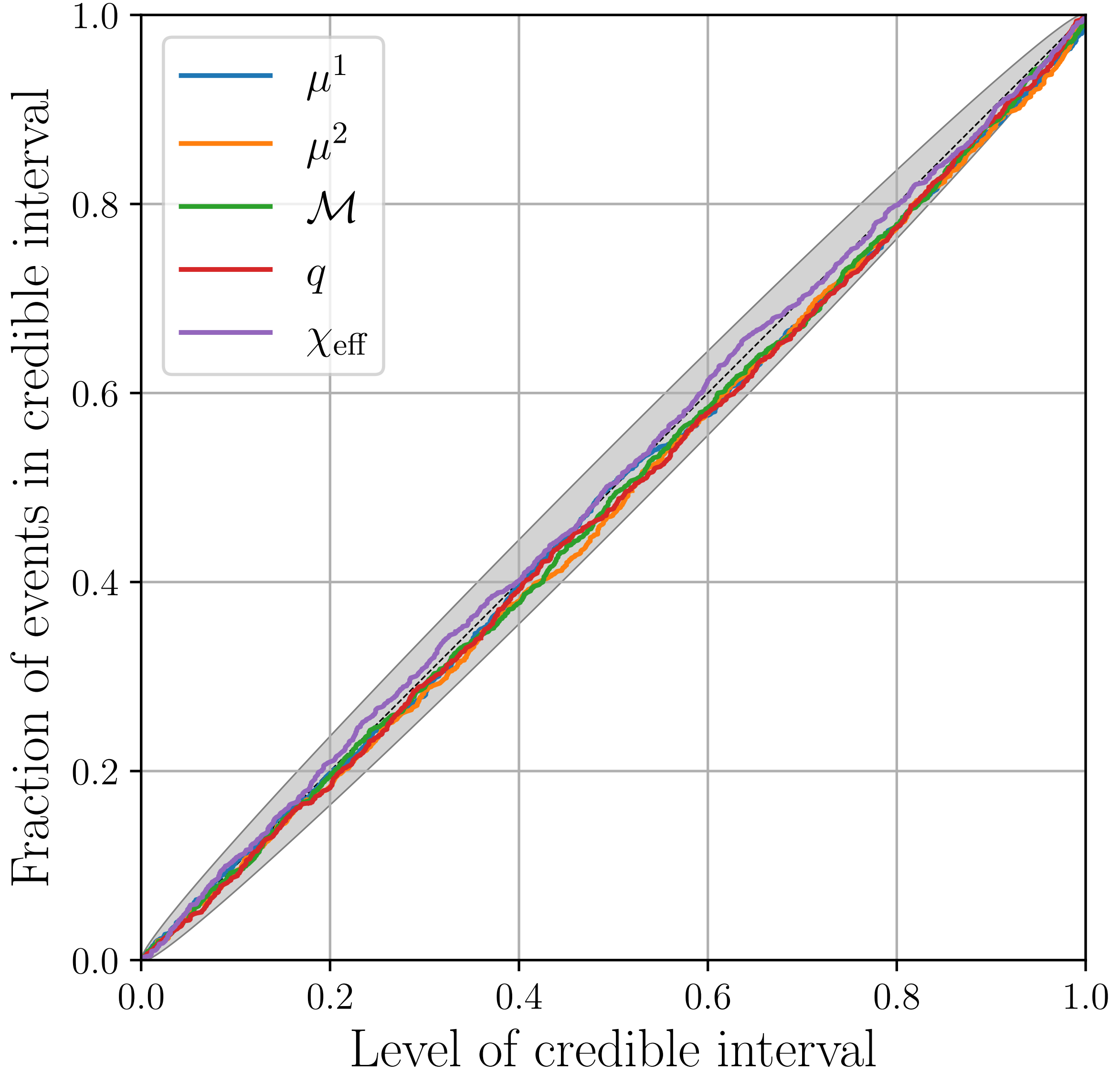
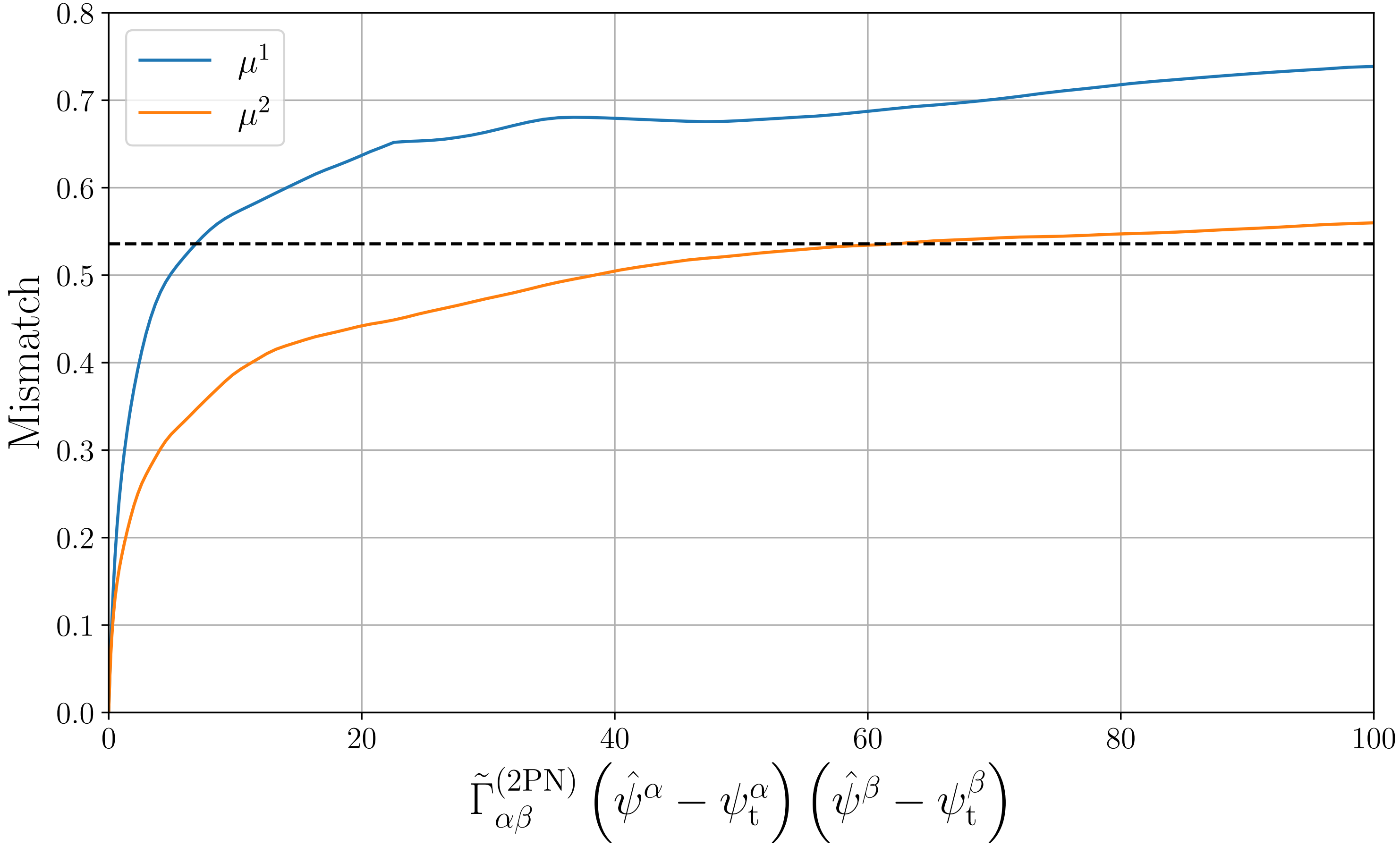


Figure: P-P plot for FROQ parameter estimation

Mismatch threshold for broader spin

$$\tilde{\Gamma}_{\alpha\beta}\left(\hat{\psi}^{\alpha}-\psi_t^{\alpha}\right)\left(\hat{\psi}^{\beta}-\psi_t^{\beta}\right)<63.7$$



Basis size for broader spin

High-spin Prior range	Size	Speedup	Broader-spin Prior range	Size	Speedup
$445.41 \leq \mu^1 \leq 446.75$ $-98.5 \leq \mu^2 \leq -86.5$	63	9800	$444.34 \leq \mu^1 \leq 447.95$ $-126.9 \leq \mu^2 \leq -58.8$	153	4100
$254.31 \leq \mu^1 \leq 255.36$ $-62.9 \leq \mu^2 \leq -52.9$	43	5800	$253.15 \leq \mu^1 \leq 256.53$ $-95.0 \leq \mu^2 \leq -25.8$	129	2000
$140.19 \leq \mu^1 \leq 141.05$ $-41.4 \leq \mu^2 \leq -32.6$	37	2600	$138.94 \leq \mu^1 \leq 142.14$ $-64.1 \leq \mu^2 \leq -14.2$	107	930

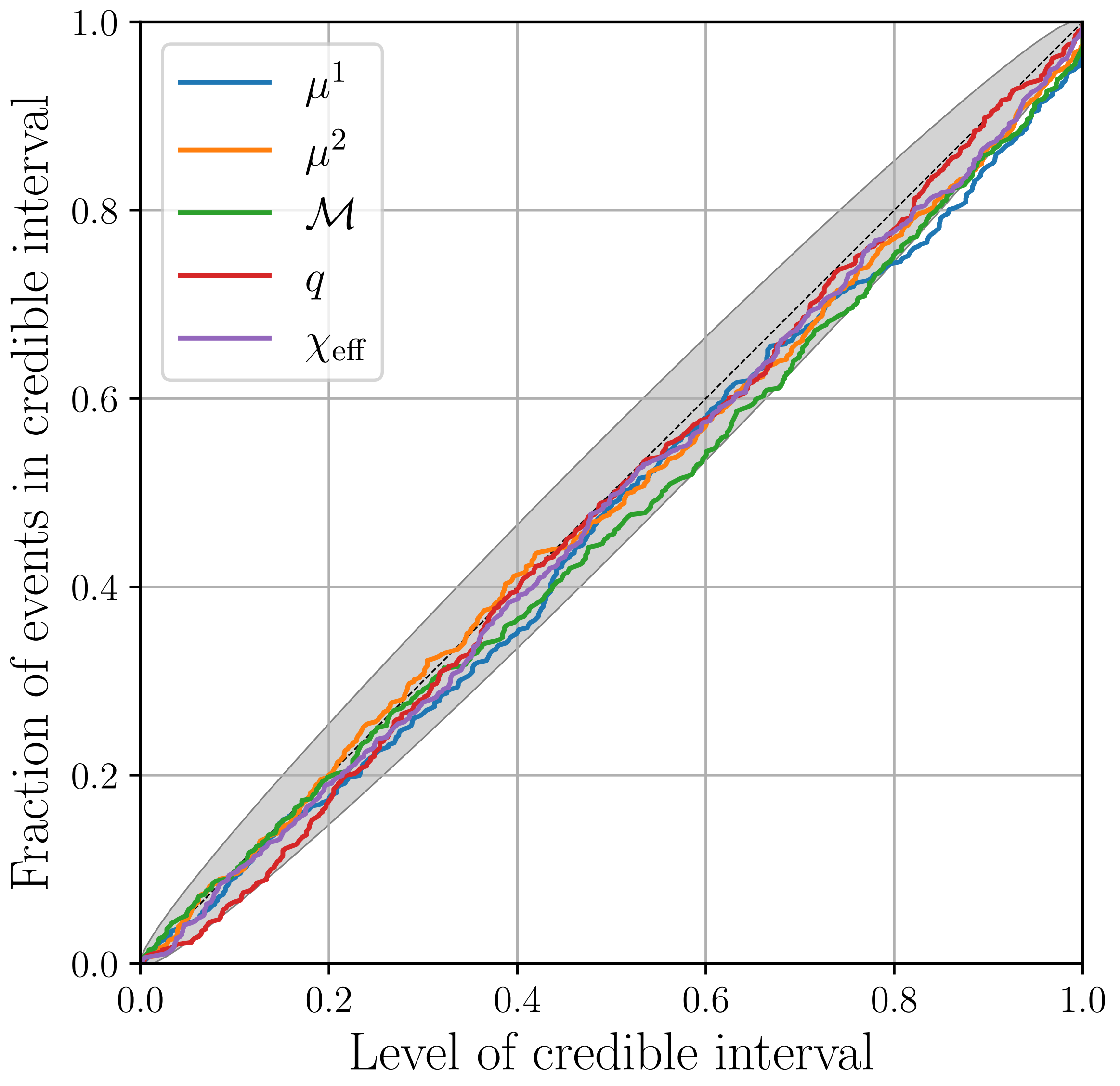
- $m_1, m_2 < 3M_\odot$
- Template bank covers $|\chi| < 0.05$.
- The spin prior of PE is $|\chi| < 0.7$.

Injection study

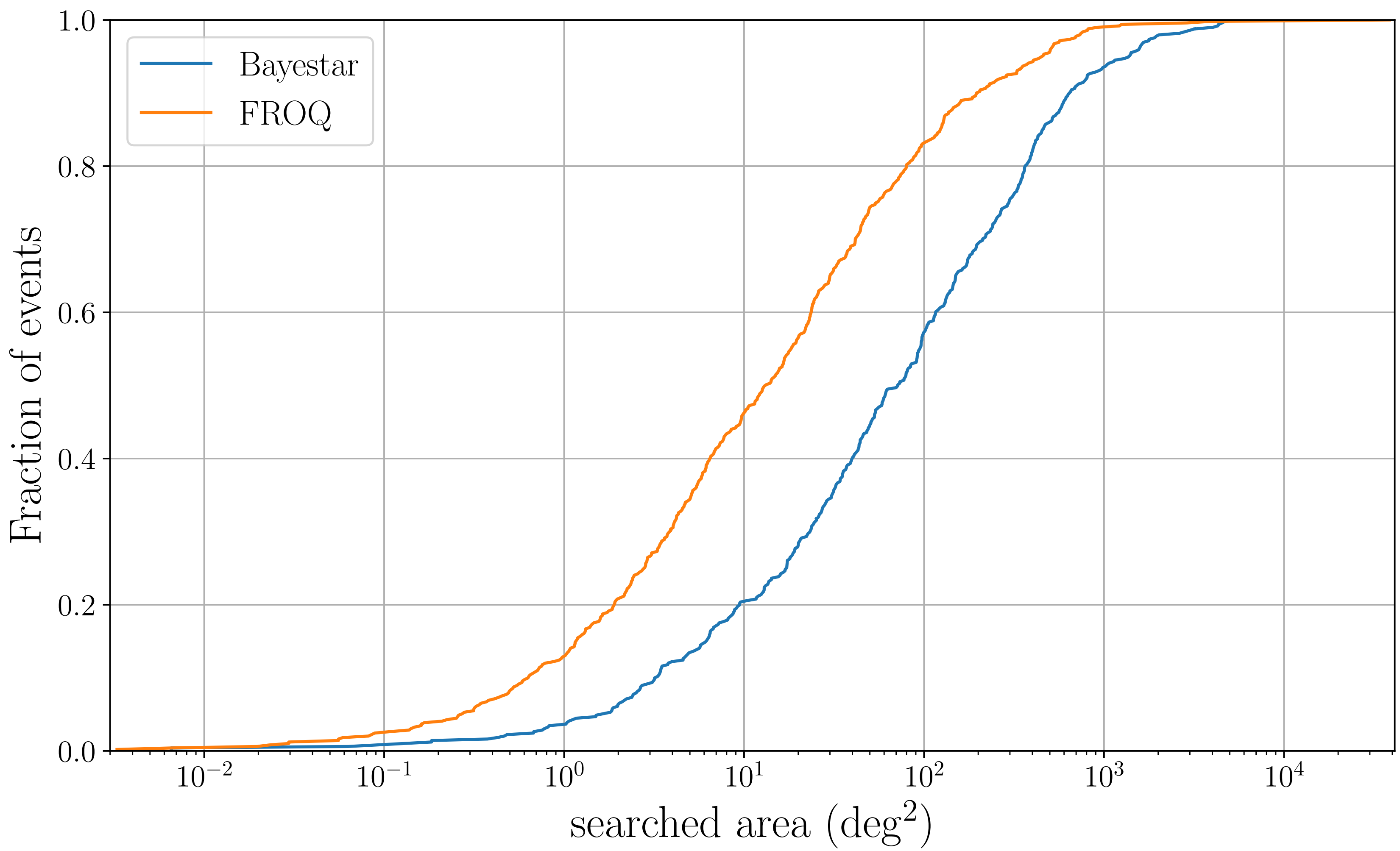
- 07:00:00 UTC ~ 23:39:00 UTC on August 19, 2017
- 2000 injections with the intervals of 30 seconds
- $m_1, m_2 < 2M_\odot$, $|\chi_1|, |\chi_2| < 0.05$
- Matched filter with GstLAL software [1]
- Detected if Network SNR >12, the second largest SNR > 5
- Low-spin FROQ with LALInference [2] for the detected injections

[1]: C. Messick *et al.*, Phys. Rev. D 95, 042001 (2017)
[2]: J. Veitch *et al.*, Phys. Rev. D 91, 042003 (2015)

P-P plot (broader spin)



Search area (broader spin)



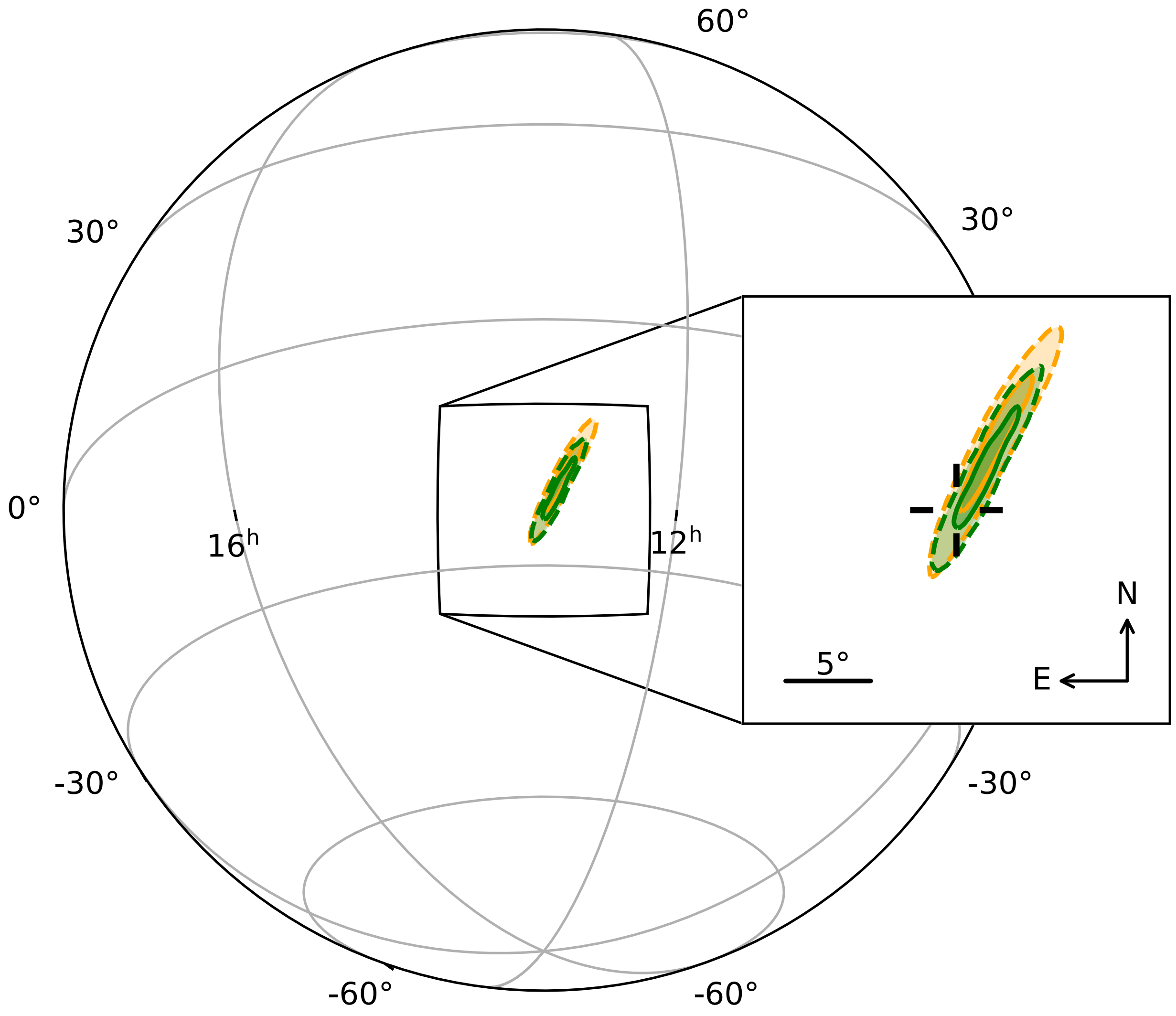
72 deg^2 (Bayestar) ⇨ 13 deg^2 (FROQ) at median

Application to GW170817

- Available in 13 minutes

- Searched area:

  10.4 deg^2 ⇨ 6.8 deg^2



Orange: Bayestar, green: FROQ, the tick represents the location of the host galaxy

Efficient sampling with $\mu^1 - \mu^2$

- The standard parameters used for stochastic sampling are $(\mathcal{M}, q, \chi_1, \chi_2)$.
- The likelihood is simpler in the $\mu^1 - \mu^2$ space.
- Stochastic sampling with $(\mu^1, \mu^2, q, \chi_2)$ is $\mathcal{O}(10-100)$ times faster.

  E. Lee, SM, and T. Hideyuki in preparation.
- See the Eunsub's poster for more detail.

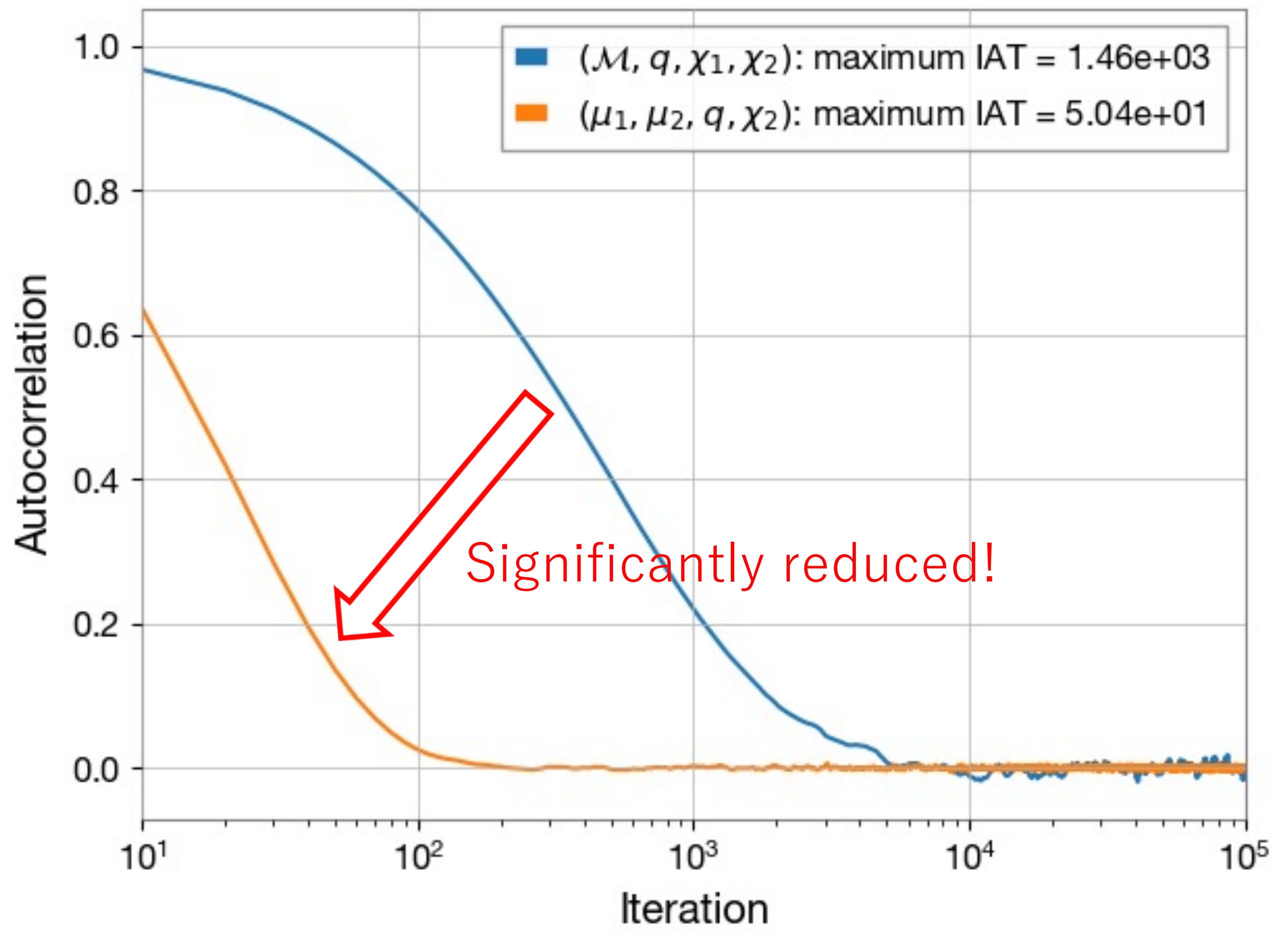


Figure: Autocorrelation functions for the two sets of sampling parameters

Multi-banding

Divide the total frequency range into B bands using overlapping smooth window functions $\{w^{(b)}(f)\}_{b=0}^{B-1}$.

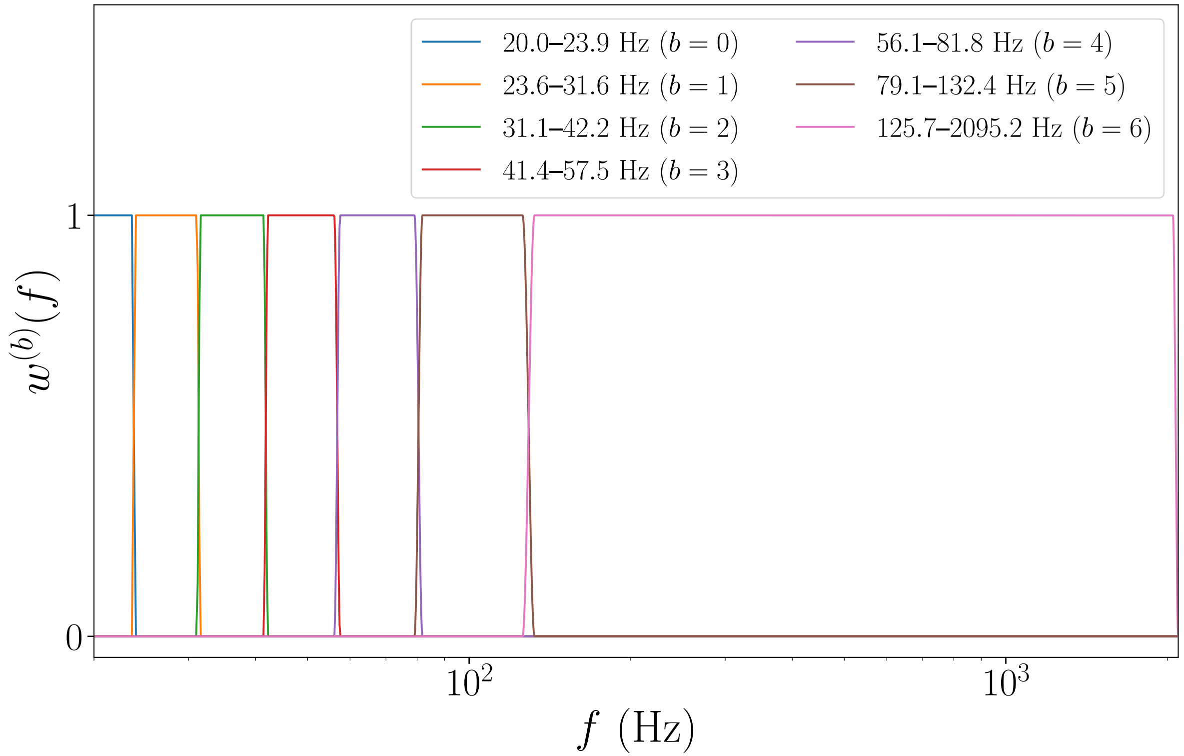


Figure: Overlapping smooth windows

They are constructed so that their sum becomes unity,

$$\sum_{b=0}^{B-1} w^{(b)}(f) = 1.$$

Speed-up gain (including higher-order moments)

Table: Speed-up gains for IMRPhenomD (only 2-2 mode) and IMRPhenomHM (including higher-order moments)

$f_{\rm low}$ (Hz)	T (s)	$K_{\rm orig}$	IMRPhenomD $K_{\rm orig}/K_{\rm MB}$	IMRPhenomD Speed up	IMRPhenomHM $K_{\rm orig}/K_{\rm MB}$	IMRPhenomHM Speed up
20	256	5.2×10^5	4.5×10	5.1×10	2.7×10	2.1×10
10	1024	2.1×10^6	1.2×10^2	1.5×10^2	5.7×10	4.6×10
5	8192	1.7×10^7	4.4×10^2	4.9×10^2	1.6×10^2	1.2×10^2

- Smaller speed-up gains for higher-order moments due to the longer time-to-merger.
- For either case, the speed-up gain is $\mathcal{O}(10)$ for $f_{\rm low} = 20\ \mathrm{Hz}$ (used for LVK) and $\mathcal{O}(100)$ for $f_{\rm low} = 5\ \mathrm{Hz}$ (appropriate for the third-generation detectors).

Consistency

- 200 simulated BNS signals
- LIGO-Hanford, LIGO-Livingston, and Virgo with their design sensitivities
- Uniformly distributed up to 100Mpc. The median SNR is 24.
- The deviations from a diagonal line are consistent with statistical errors.

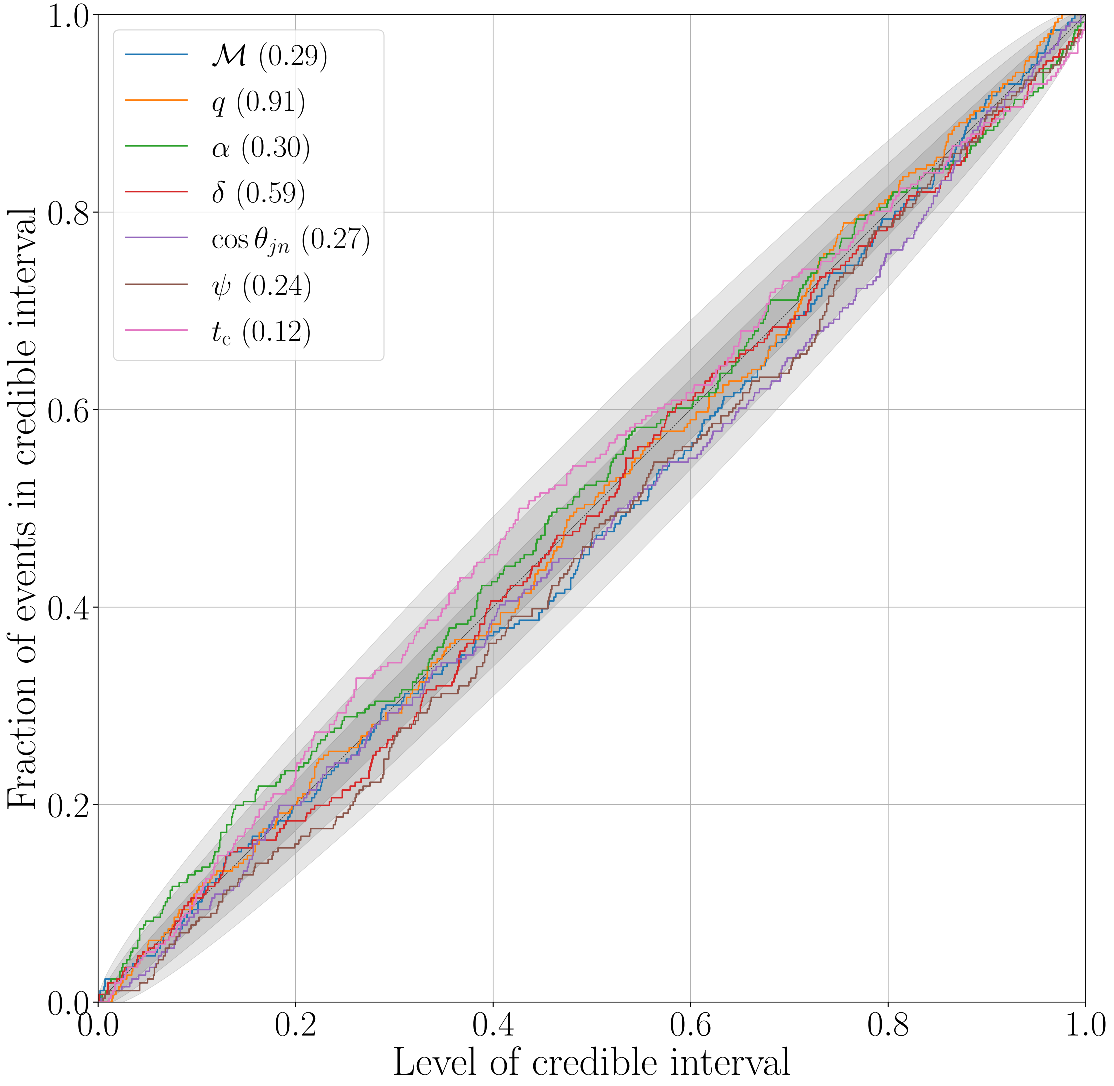


Figure: P-P plot for multi-banding PE

Likelihood errors for GW190814

Likelihood errors due to multi-banding are much smaller than statistical errors.