Gravitational wave data analysis
an overview

Sylvain Marsat (L2IT, Toulouse)

Introduction

Analyzing GW signals

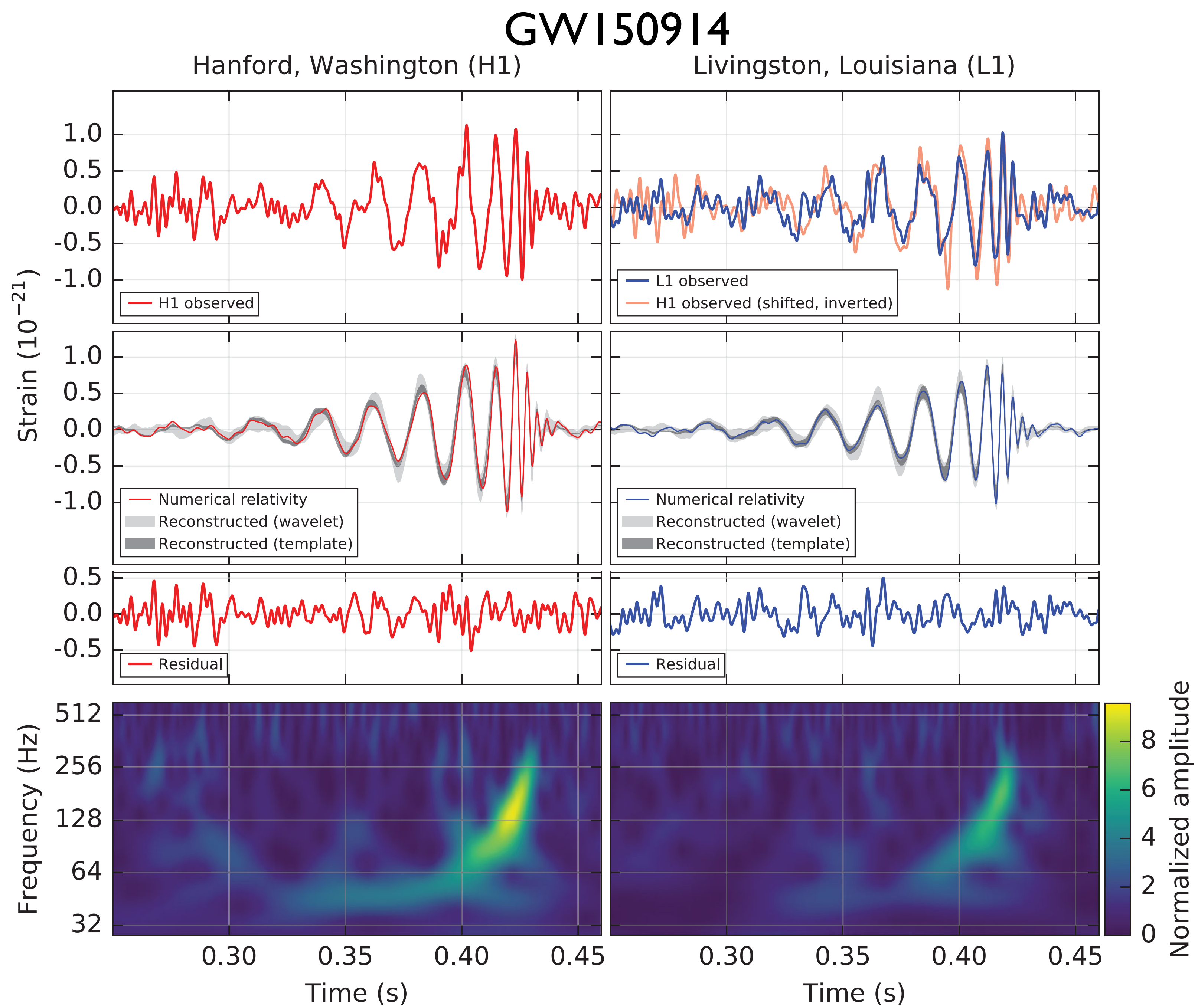
- Detection, low-latency alerts
- Bayesian parameter estimation (PE) for each source
- Hierarchical Bayesian analyses (population, cosmology, …)

Coalescence of compact binaries (BBH, NSBH, BNS): waveform templates for detection (digging in the noise) and PE

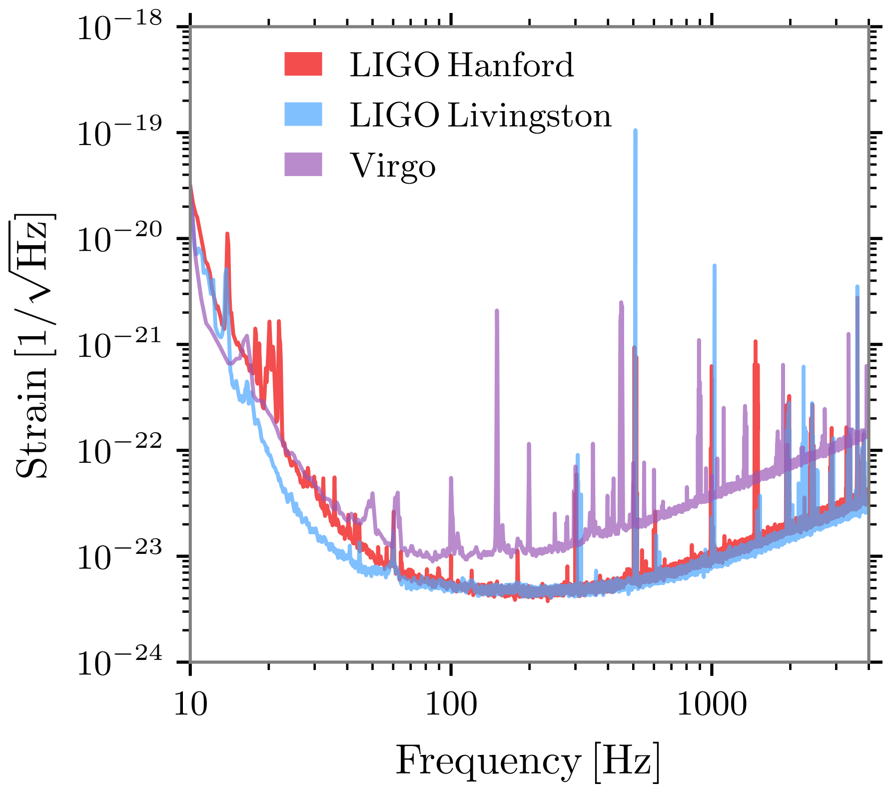
- Theoretical challenge
- Computational challenge

Left aside in this talk:

- unmodeled searches (for SN, and the unexpected)
- Stochastic backgrounds



Data analysis basics



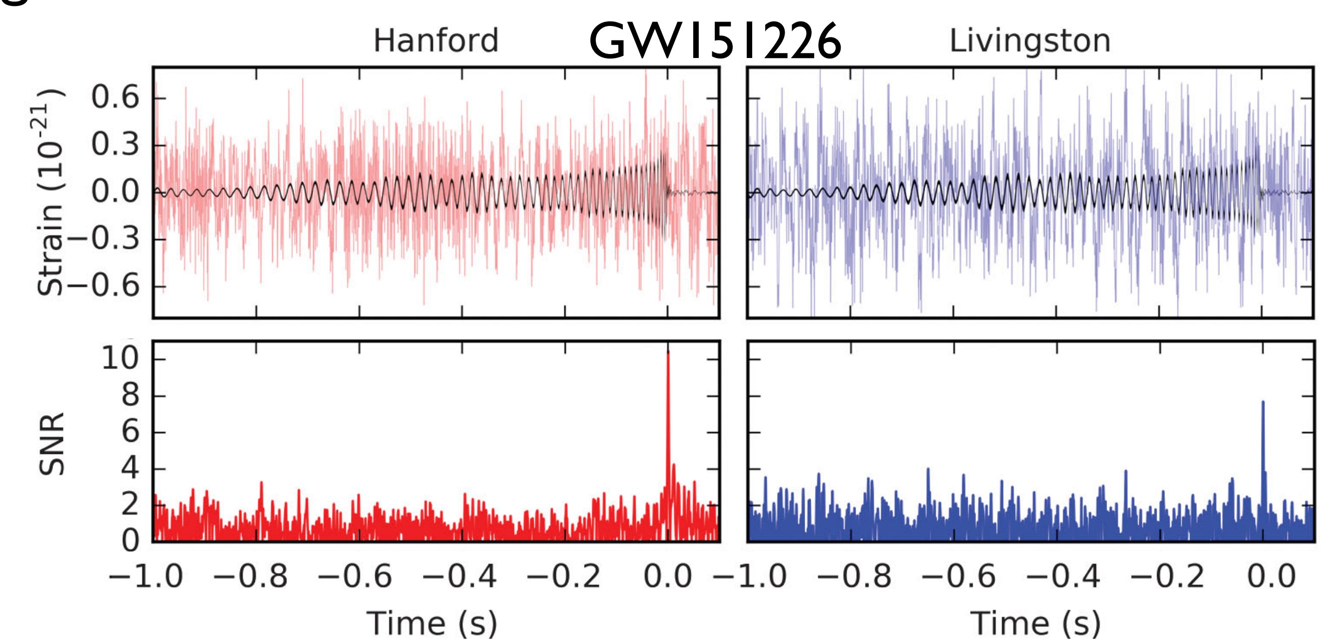
Instrument and noise

Noise PSD, idealized: $\langle \tilde{n}(f)\tilde{n}^*(f')\rangle = \frac{1}{2}S_n(f)\delta(f-f')$

- stationarity
- Gaussianity

Detector Characterization (DetChar):

- calibration, noise removal (lines)
- PSD estimation
- glitch identification/removal



Data analysis basic tools

Matched filter overlap:

$$(h_1|h_2) = 4\mathrm{Re}\int df\, \frac{\tilde{h}_1(f)\tilde{h}_2^*(f)}{S_n(f)}$$

Matched filter SNR: $(h|d)$

Likelihood (stationary, Gaussian):

$$\ln\mathcal{L}(d|\theta) = -\sum_{\mathrm{channels}} \frac{1}{2}(h(\theta)-d|h(\theta)-d)$$

$d = h(\theta_0) + n_0$ (+ calibration uncertain.)

Bayesian parameter estimation, posterior:

$$p(\theta|d) = \frac{\mathcal{L}(d|\theta)p_0(\theta)}{p(d)}$$

$p_0(\theta)$ prior
$p(d)$ evidence

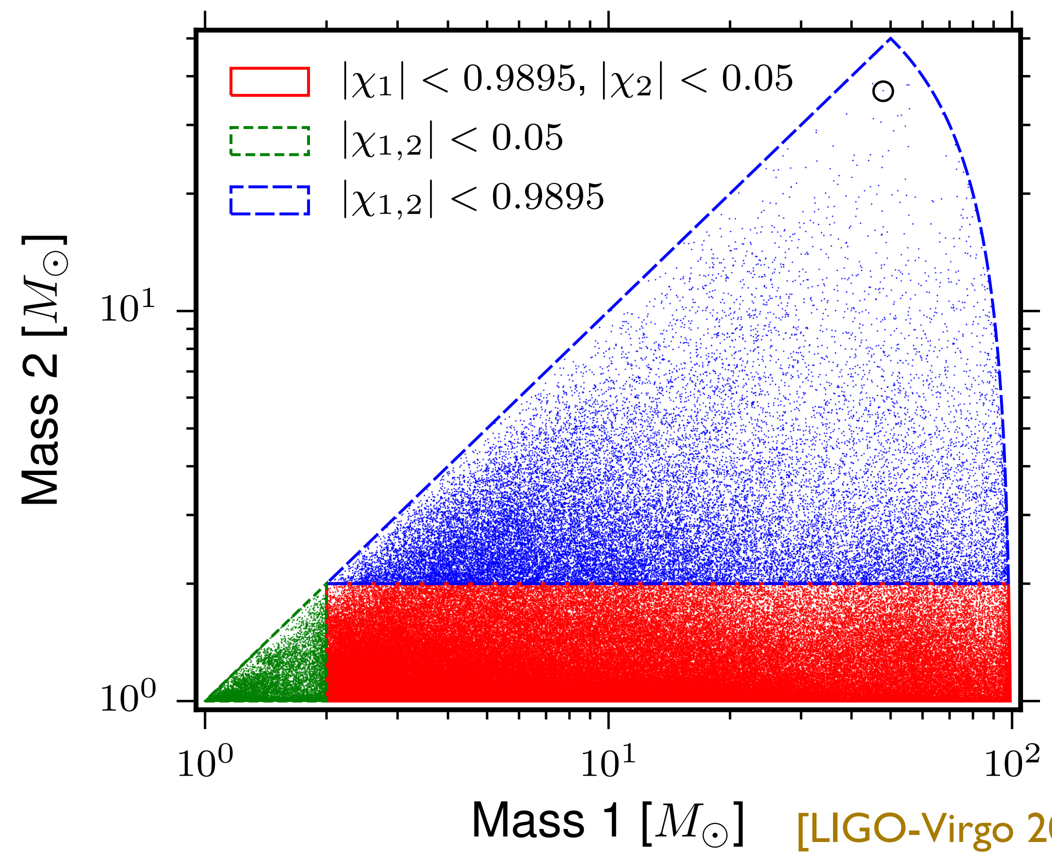
LIGO/Virgo detections: methods

Template bank searches

Precomputed bank of templates optimally placed in parameter space

Modified detection statistic: take into account SNR and chi-square of the residuals

Trade-off between sensitivity of the search and false alarm rate



[LIGO-Virgo 2016]

Signal / background, p_astro

Time slides: generate large amount of background data, defines false alarm rate (**FAR**)

Poisson model for rates (assume population): astrophysical or instrumental origin
Probability of astrophys. origin (**p_astro**)

Main pipelines

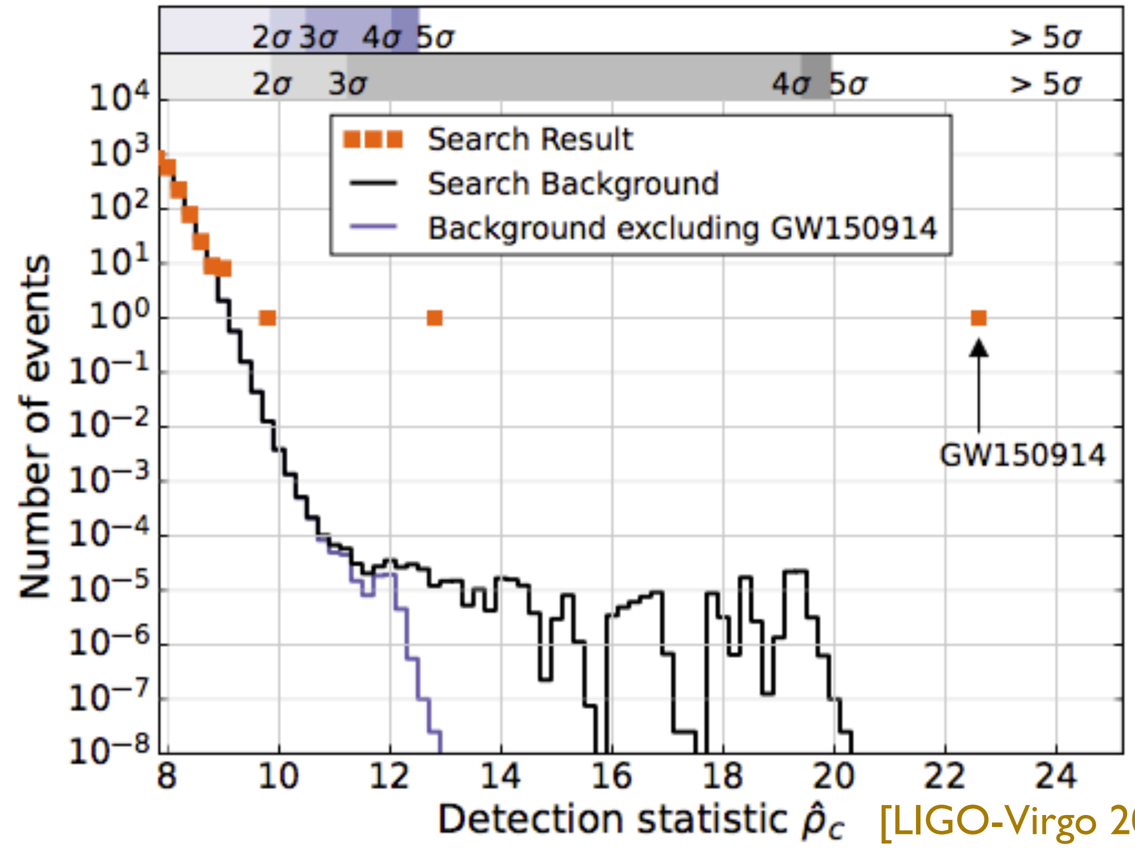
- **PyCBC**: [Usman&al 2016]
- **GstLAL**: [Cannon&al 2012]
- **MBTA**: [Adams&al 2016]
- **SPIIR**: [Chu&al 2020]

Online (low-latency) and offline versions

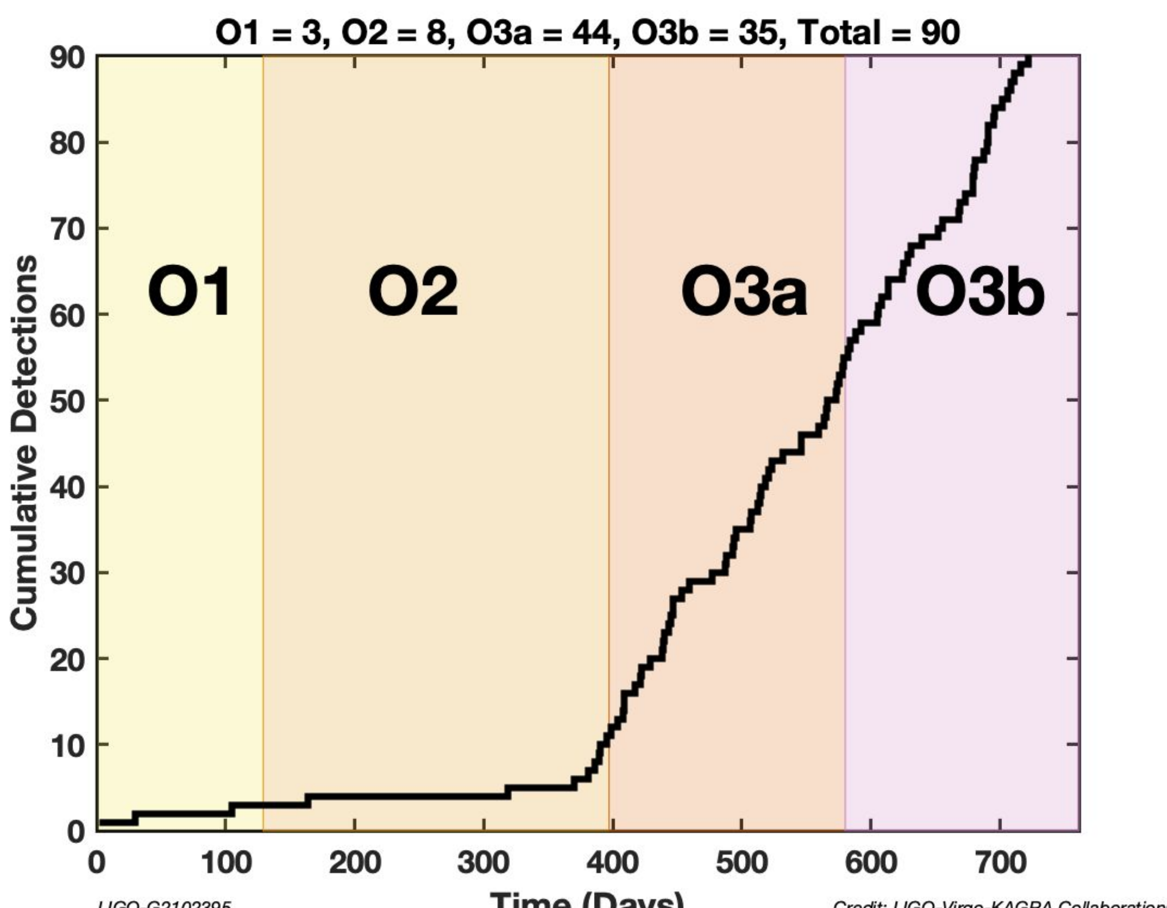
LIGO/Virgo detections: results

First detection: GW150914

Luckily, loud and clear ! $> 5\sigma$



[LIGO-Virgo 2016]



GWTC-3

- 90 detections in total
- 82 BBH
- 6 NSBH
- 2 BNS
- large number of sub-threshold triggers

LIGO/Virgo parameter estimation: methods

Sampling

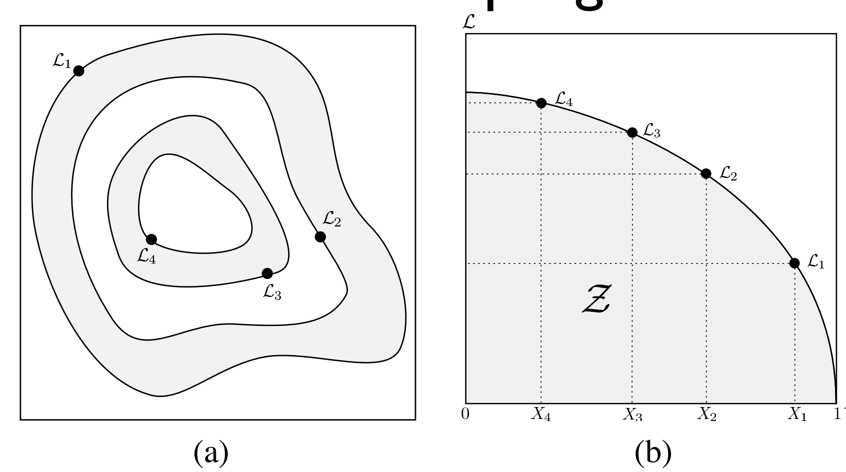
$$p(\theta|d) = \frac{\mathcal{L}(d|\theta)p_0(\theta)}{p(d)}$$

- MCMC: evolve a chain with proposal and Metropolis-Hastings acceptance — parallel tempering, tuned proposals

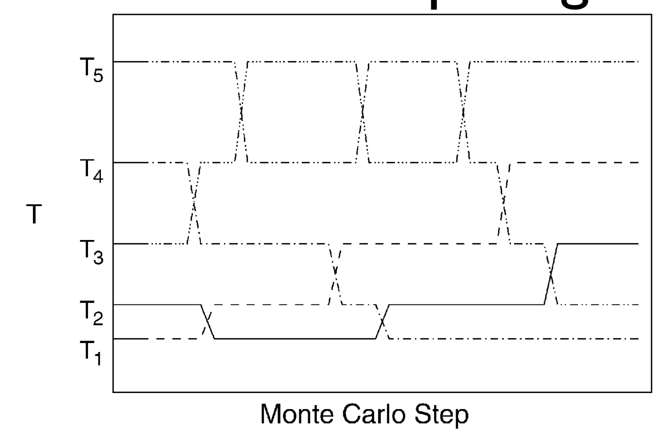
- Nested sampling: sample uniformly from inside isolikelihood contours (and get evidence)

\+ Machine learning ?

Nested sampling



Parallel tempering



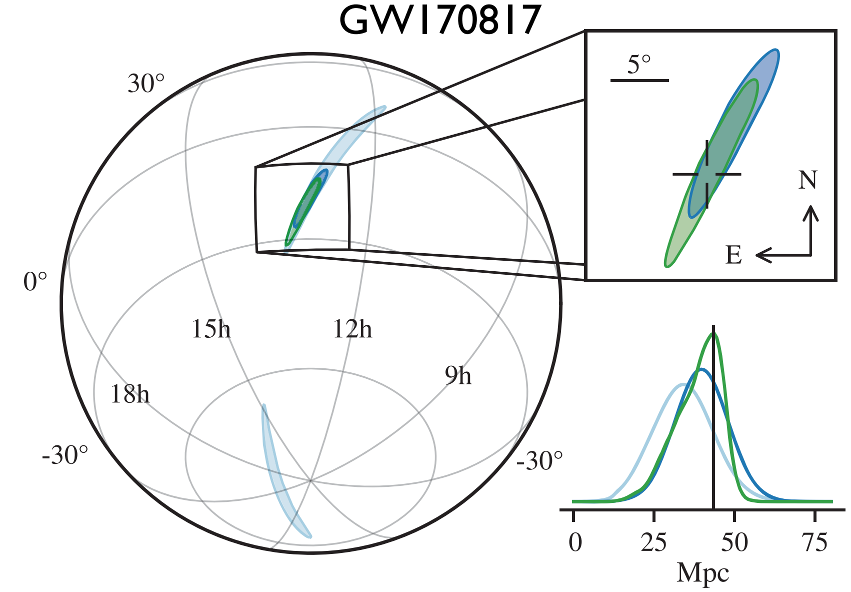
Pipelines

- **LALInference**: MCMC (PT), Nest (nested sampling chain proposals) [Veitch&al 2015]
- **Bilby**: nested sampling Dynesty [Singer&al 2015]
- **RIFT**: (expensive waveforms) two stages, first model marginal intrinsic likelihood [Singer&al 2015]

Fast localization

BayesStar [Singer&al 2015]: fast skymap in ~ seconds

GW170817

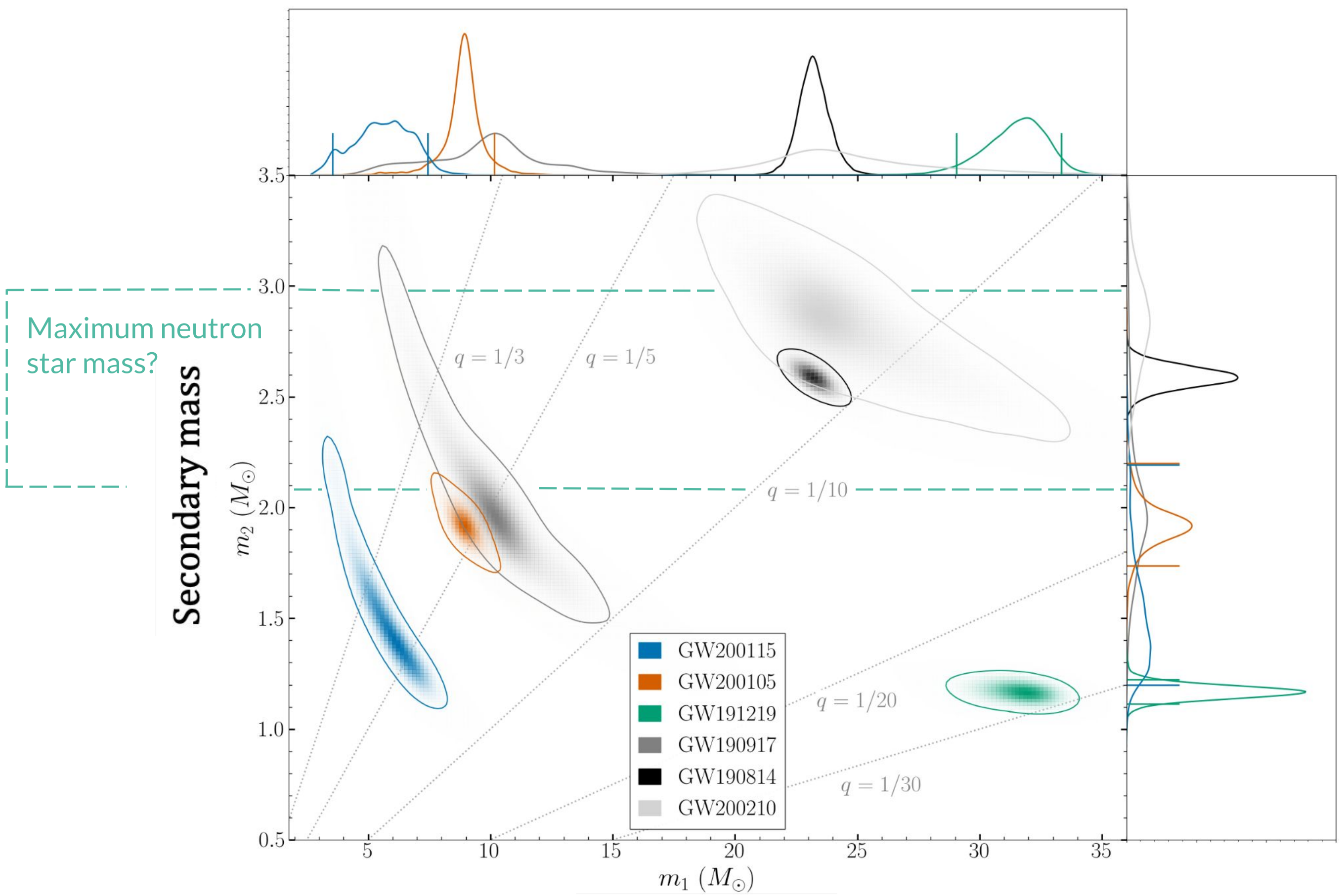


[LIGO-Virgo 2017]

LIGO/Virgo parameter estimation: results

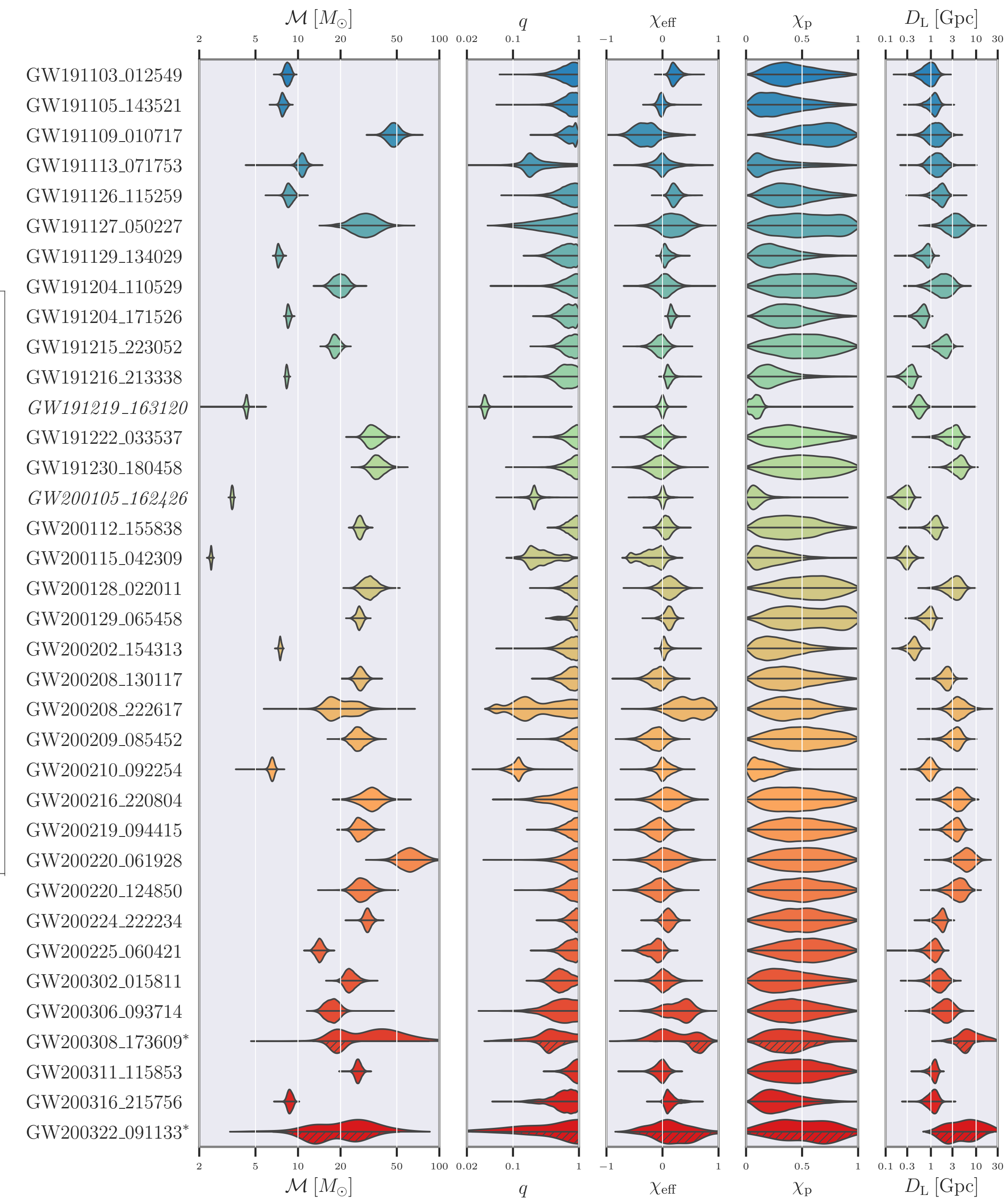
GWTC-3

NSBH candidates



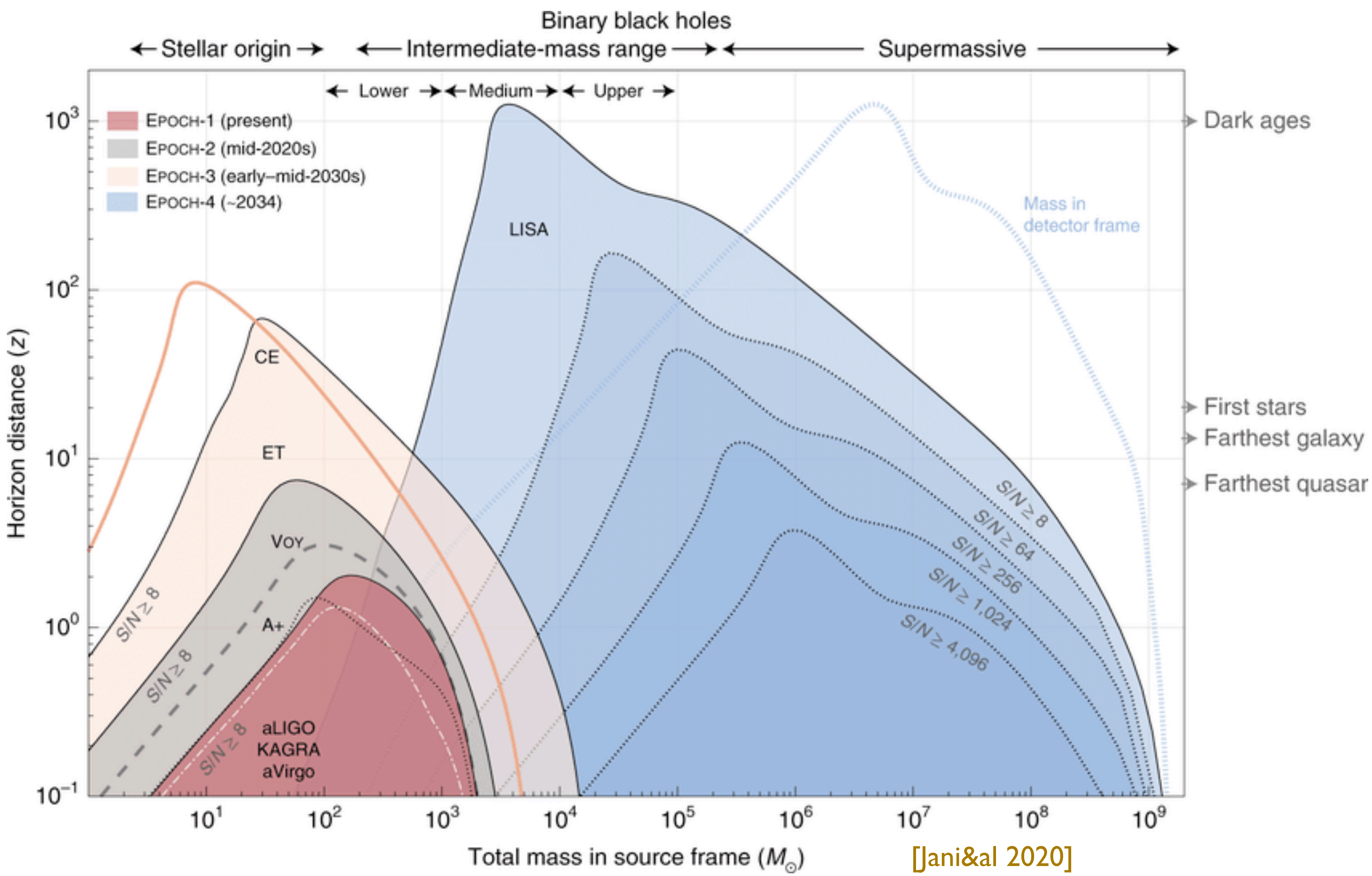
Primary mass [LVK 2021]

GWTC-3b



[LVK 2021]

The future of Gravitational Wave Astronomy

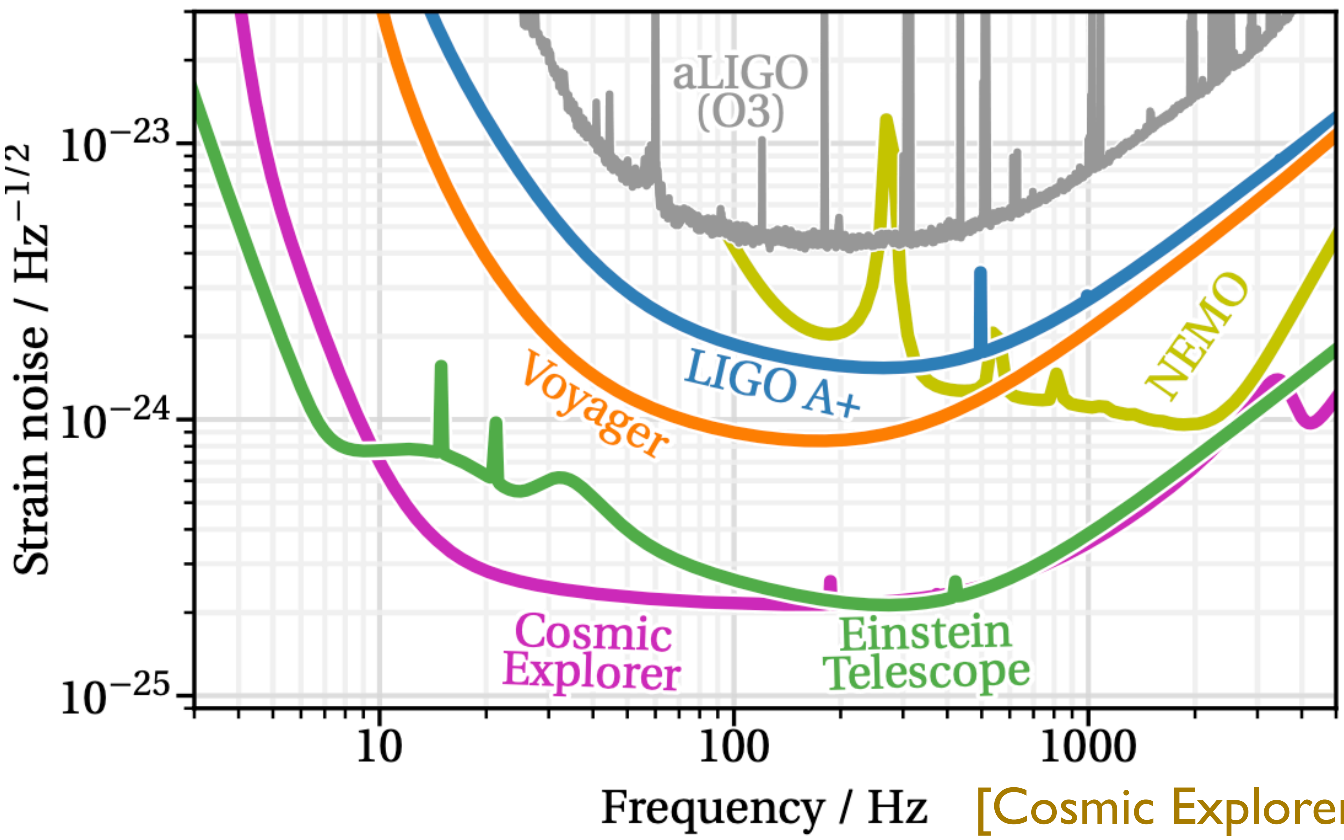


[Jani&al 2020]

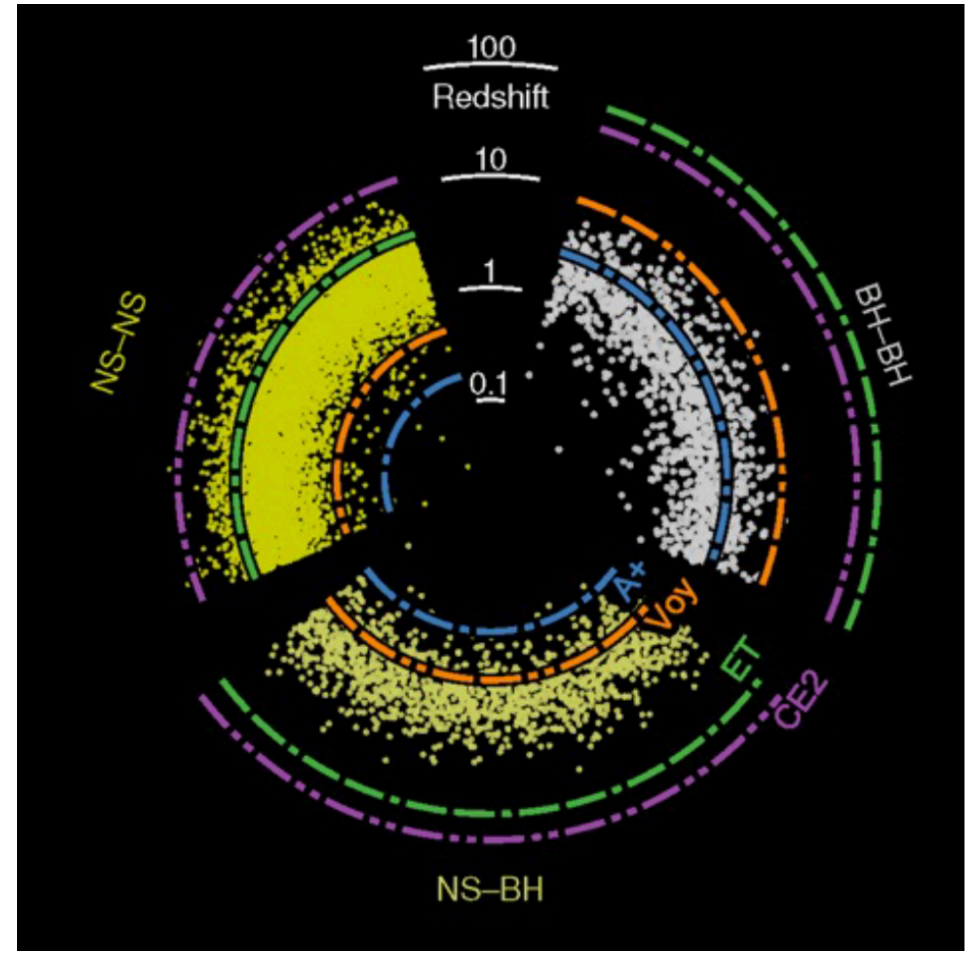
High SNRs → Great precision science !

High SNRs → Systematics become important…

Third-generation detectors: data analysis challenges



[Cosmic Explorer]



[Cosmic Explorer]

Number of detections

Events/yr (low-median-high):

- BBH: 60k-90k-150k
- BNS: 300k-1000k-3000k

Detections (2 CE + 1 ET):

- BBH: 93%
- BNS: 35%

Computational challenge !

(from [Samajdar&al 2021])

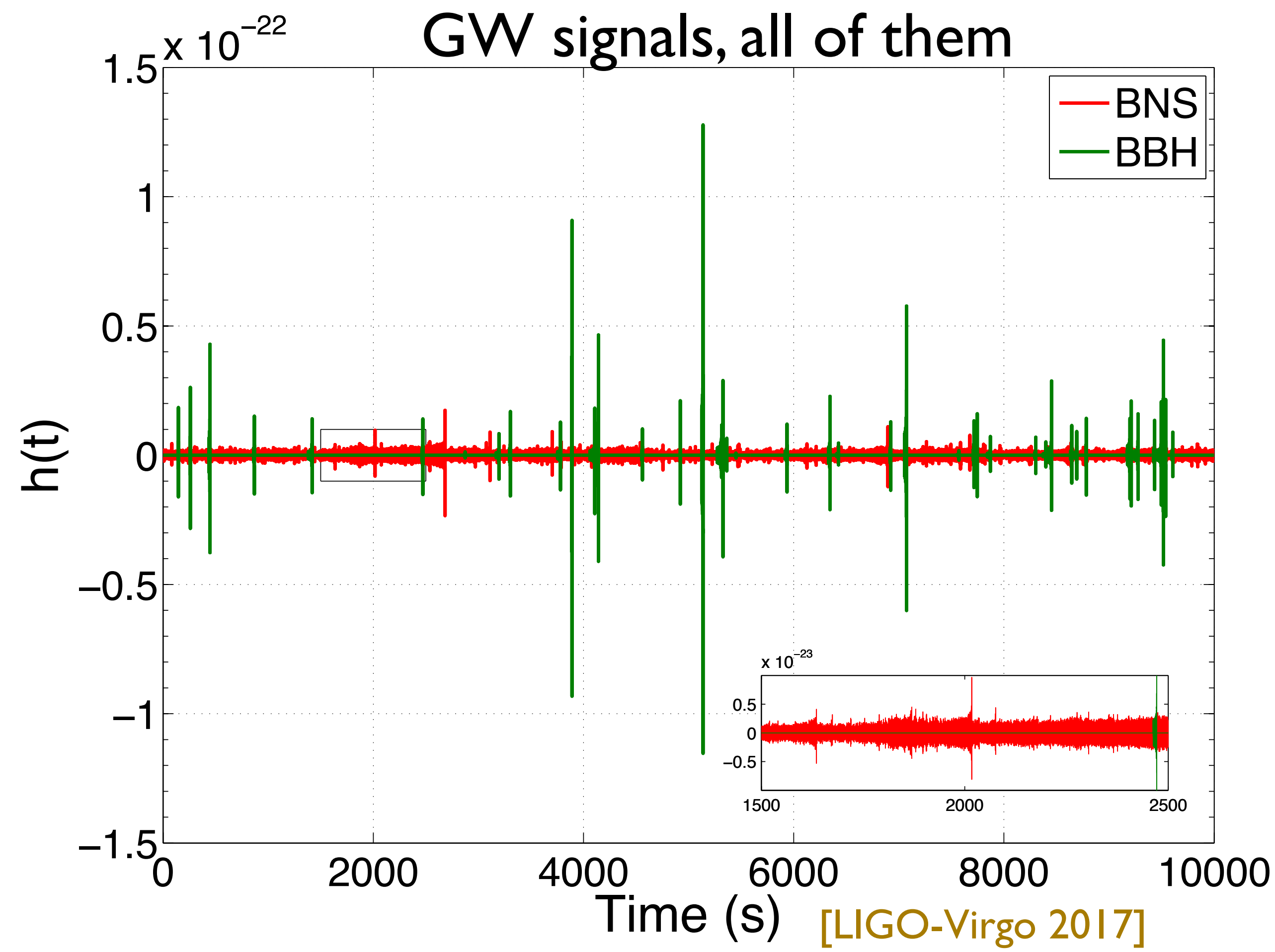
Third-generation detectors: data analysis challenges

Overlapping signals

BNS signals: several hours !

Up to ~200 overlapping mergers

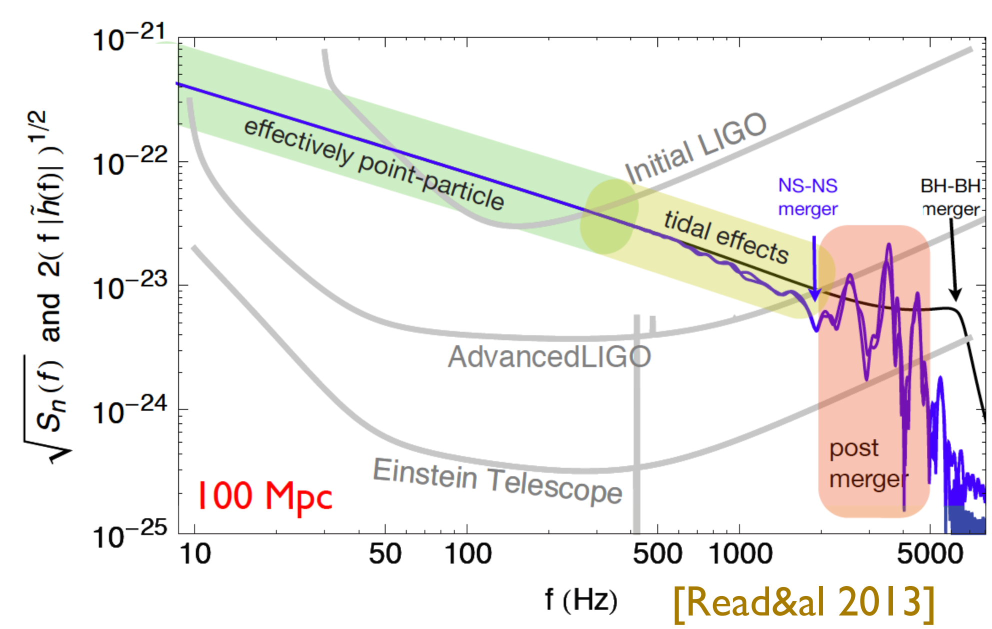
In most cases, only mild confusion



[LIGO-Virgo 2017]

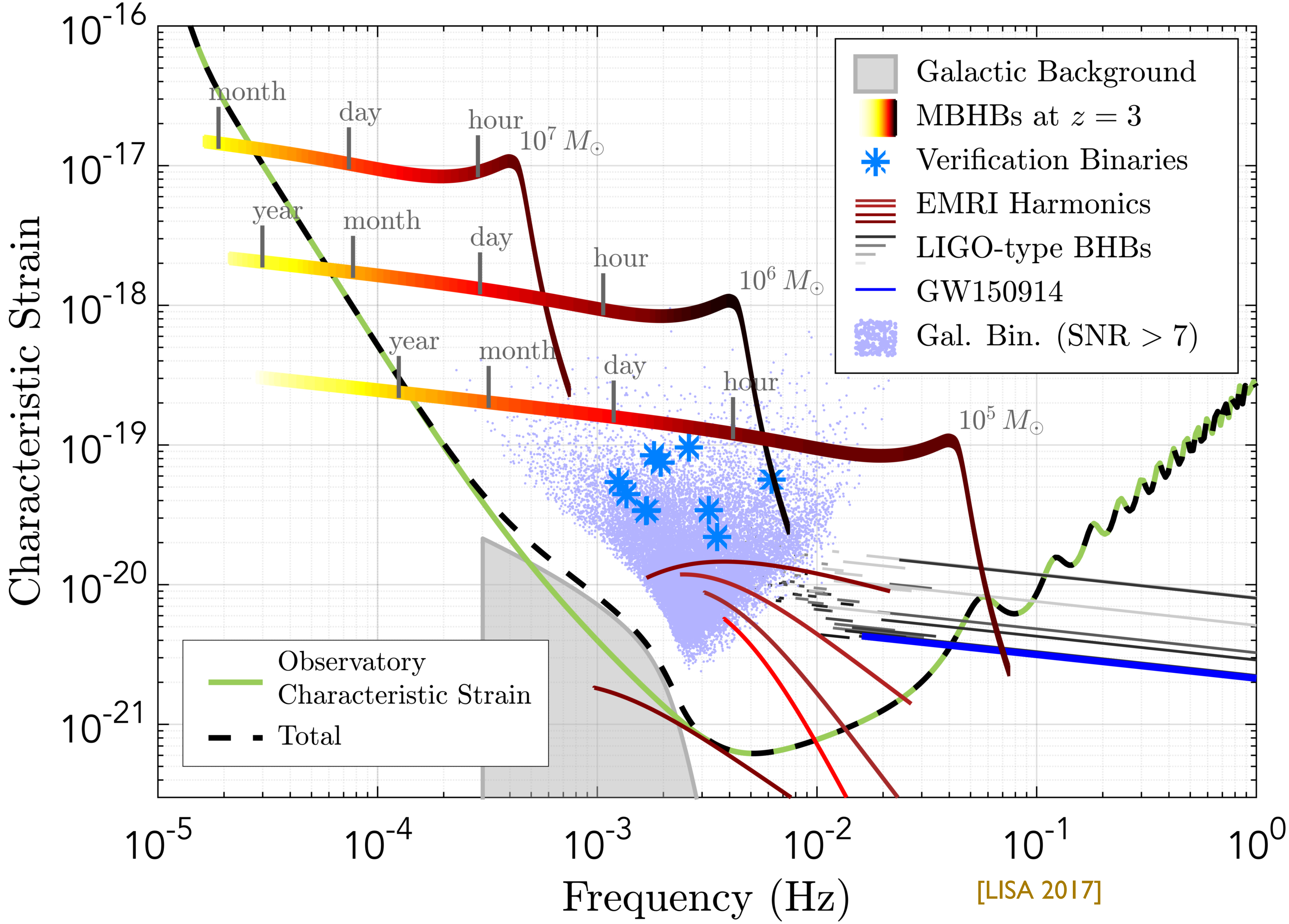
Modelling the signals

- Long BNS signals: motion of the Earth
- High SNR detections: systematics ?
- BNS post-merger



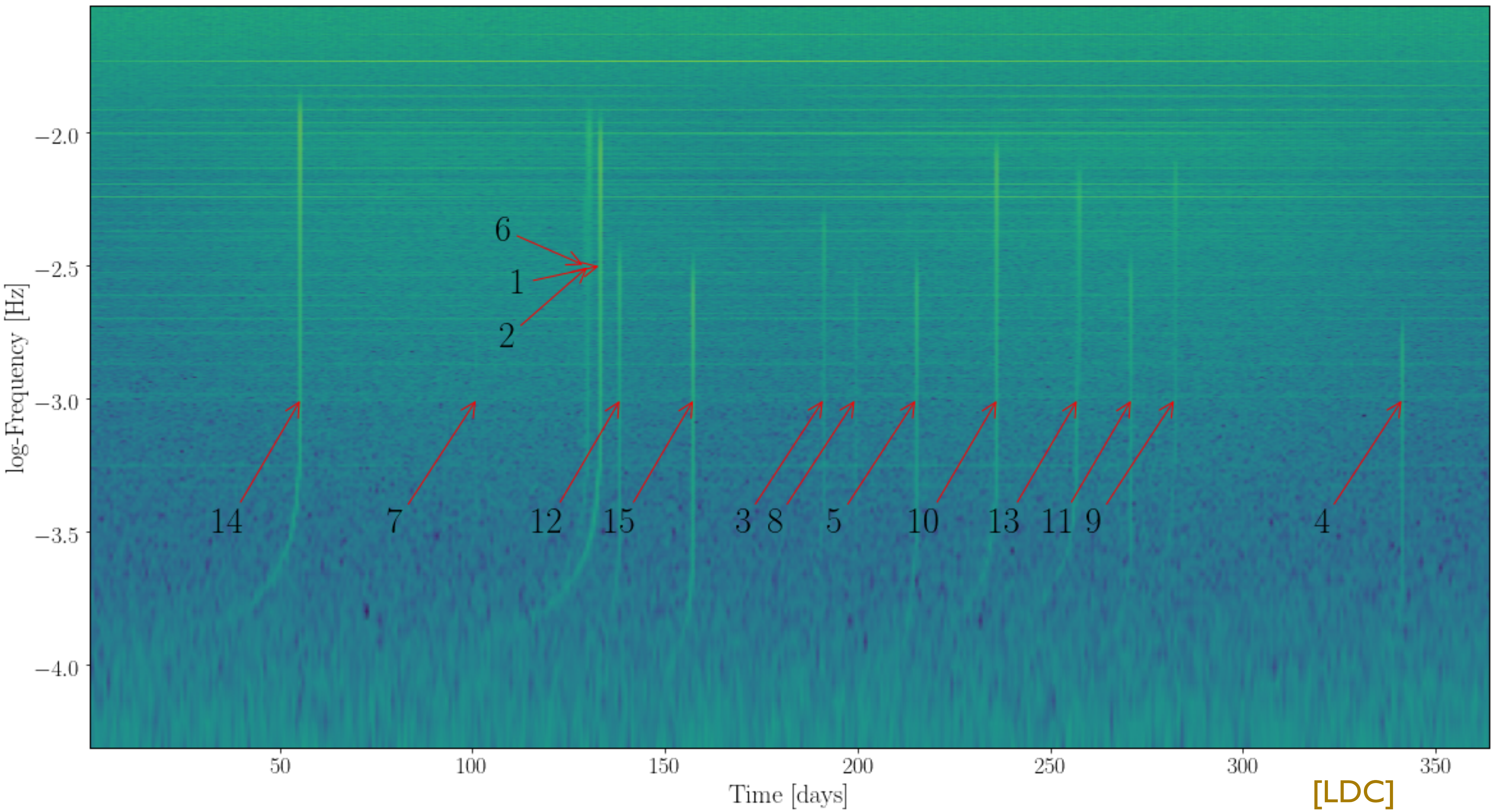
[Read&al 2013]

LISA sources



[LISA 2017]

LISA data - LDC-2 Sangria

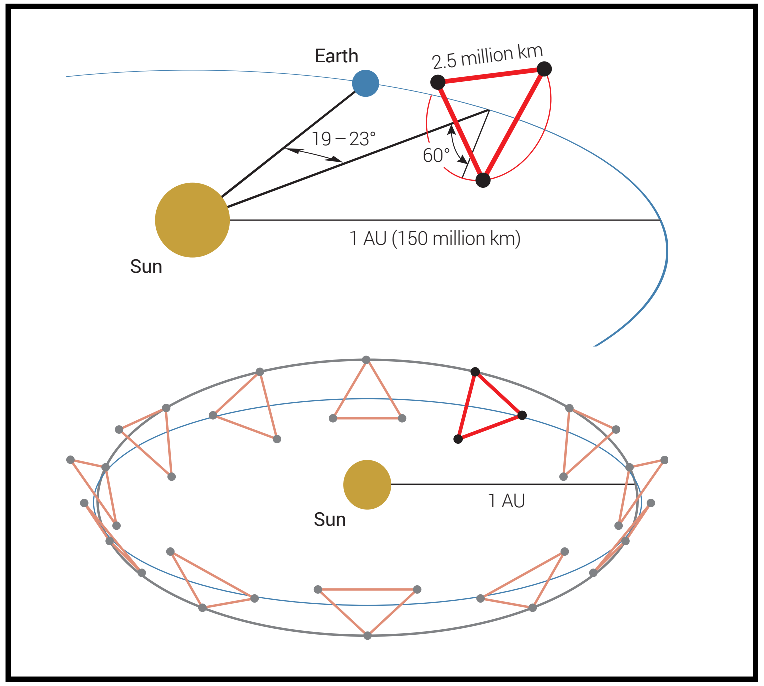


[LDC]

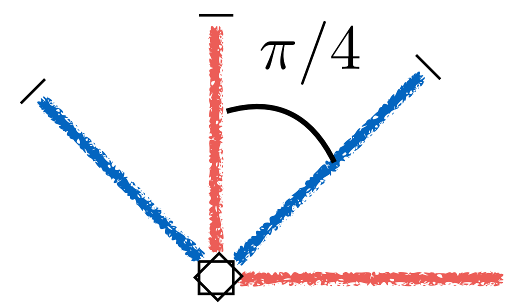
- **MBHBs**: chirping signals, emerging from low-f noise
- **GBs**: quasi-monochromatic, horizontal lines

Contrasting LIGO/Virgo and LISA

LISA orbits



Low-f approximation: **two LIGO-type detectors** in motion [Cutler 1997]



High-f: **three channels**
frequency-dependence

LISA response

Laser frequency shift, spacecrafts s to r through link l: $y = \Delta\nu/\nu$

$$y_{slr} = \frac{1}{2}\frac{1}{1-\hat{k}\cdot n_l} n_l \cdot (h(t_s) - h(t_r)) \cdot n_l$$

Chirping signals Fourier-domain:

$$\mathcal{T}_{slr} = \frac{i\pi f L}{2} \mathrm{sinc}\left[\pi f L \left(1 - k\cdot n_l\right)\right] \exp\left[i\pi f \left(L + k\cdot(p_r + p_s)\right)\right] n_l \cdot P \cdot n_l(t_f)$$

Time and frequency-dependency
Time: motion of LISA on its orbit
Frequency: departure from long-wavelength

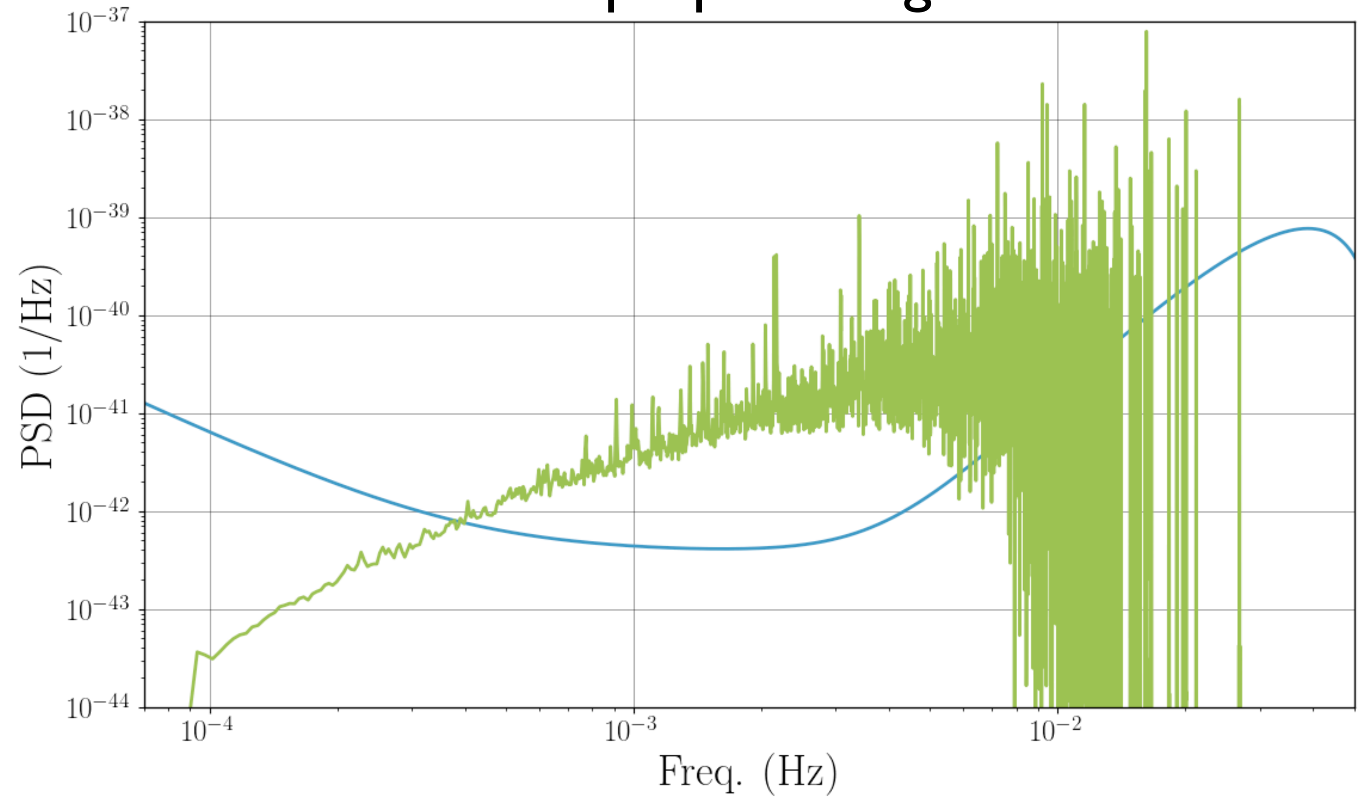
+ **Time-delay interferometry (TDI)**
linear combinations with more delays

Galactic binaries: signals and challenges

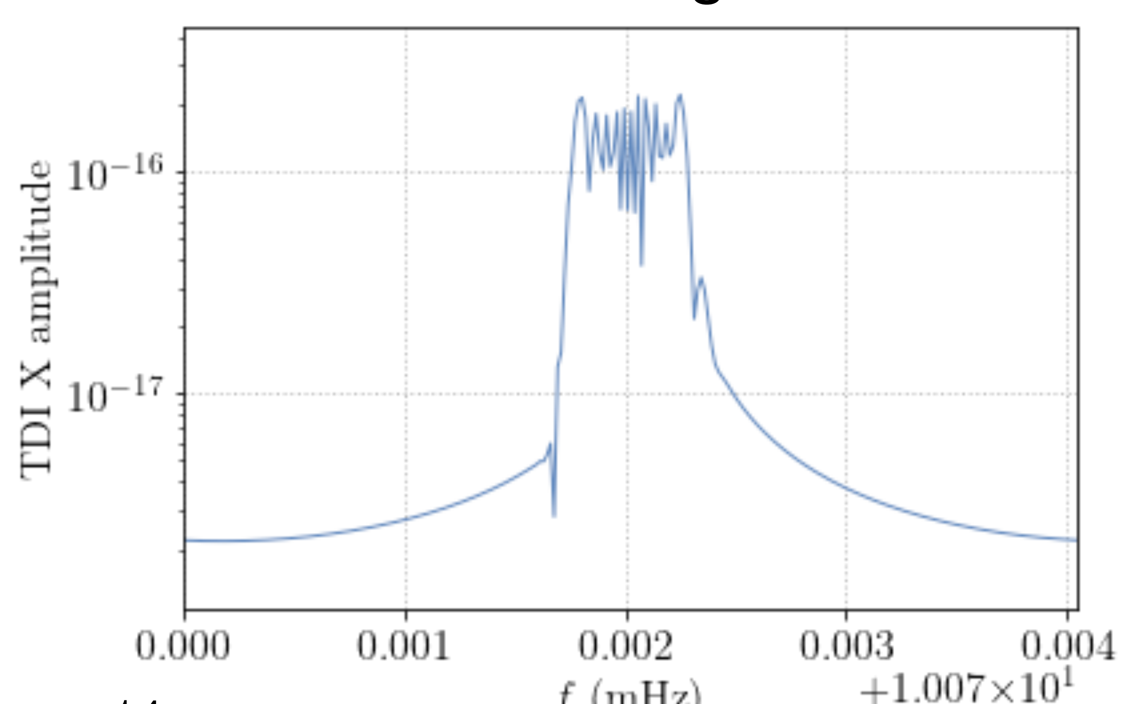
- Mostly WD-WD, some other compact objects
- Full galaxy: ~20 million systems !
- About ~20000 individually resolvable
- Form a (non-stationary) background
- Verification binaries

- Quasi-monochromatic GW emitters
- Modulation by LISA motion (sidebands in Fourier-domain)
- Superposition of signals in Fourier-domain

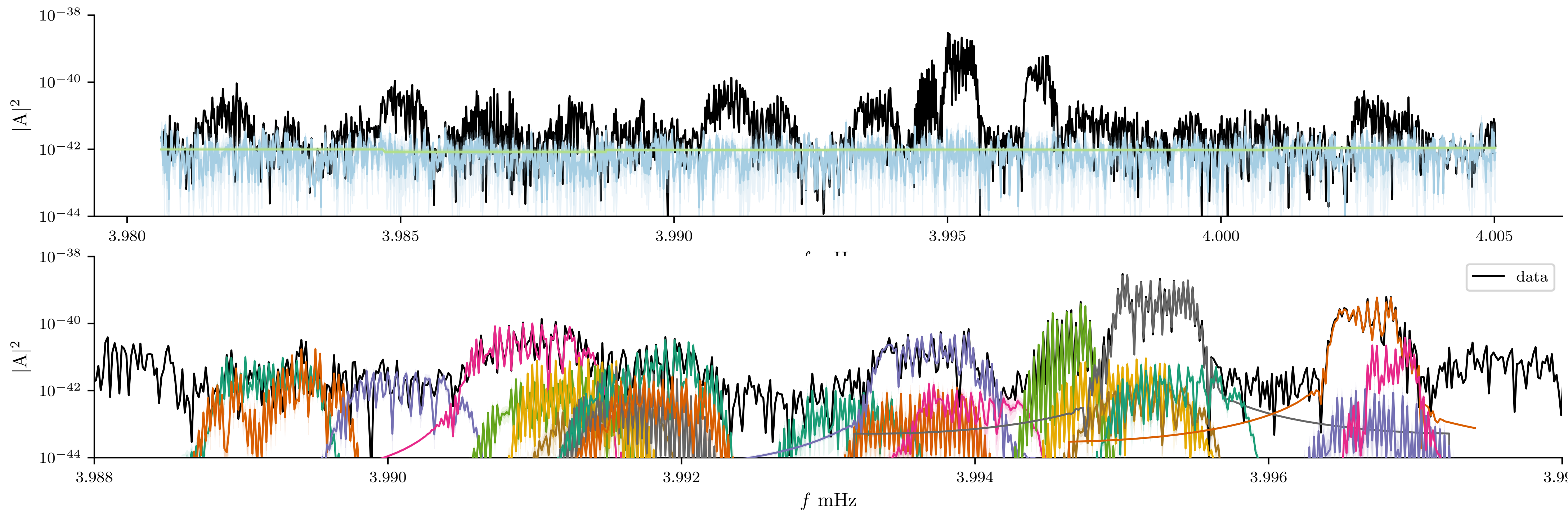
Superposed signal



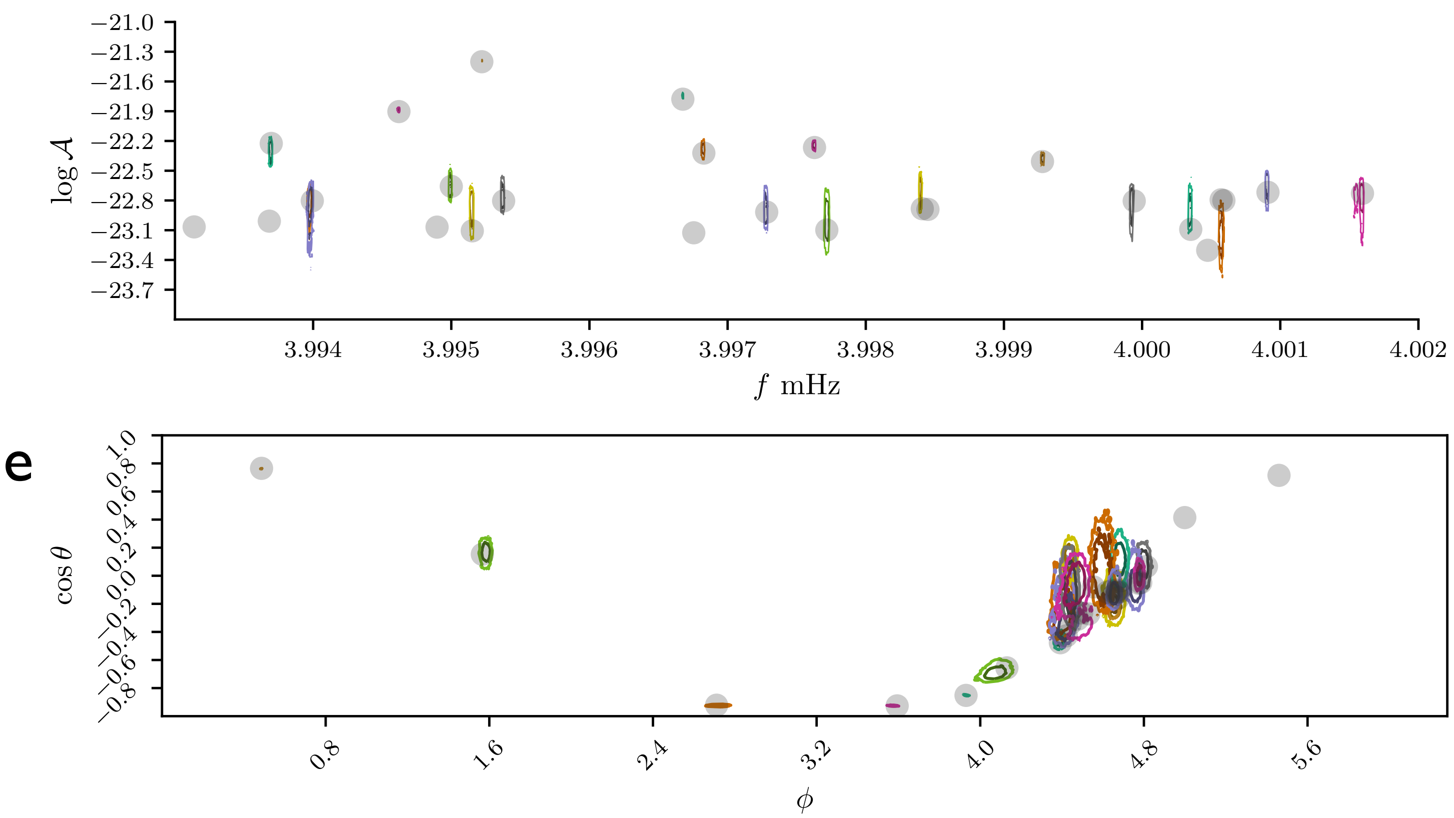
Individual signal



Galactic binaries: example results



[Littenberg&al 2020]

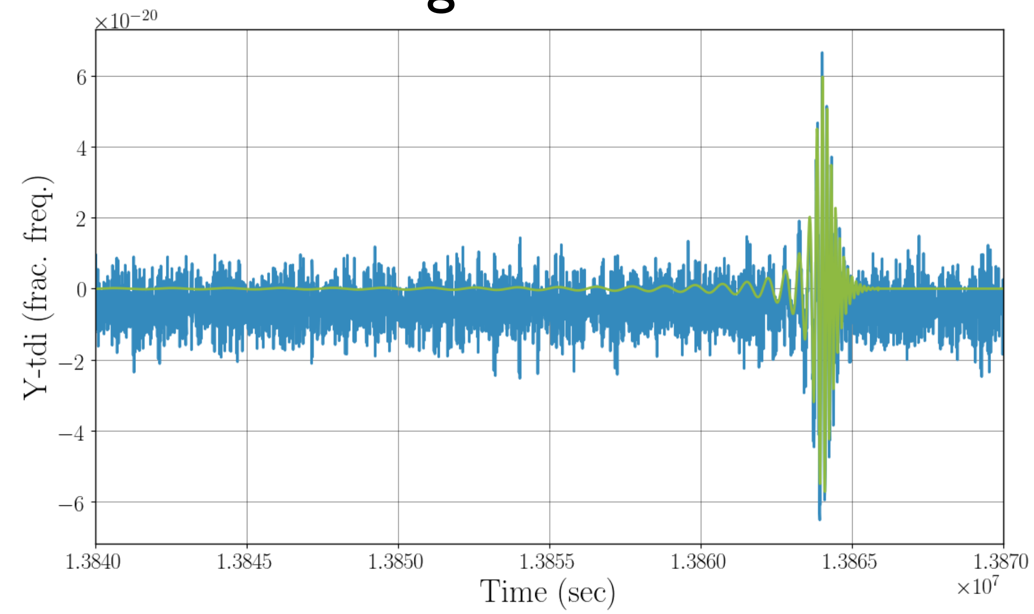


- Demonstration on a subset of the full Galaxy
- Reversible jumps MCMC to handle the superposition of sources

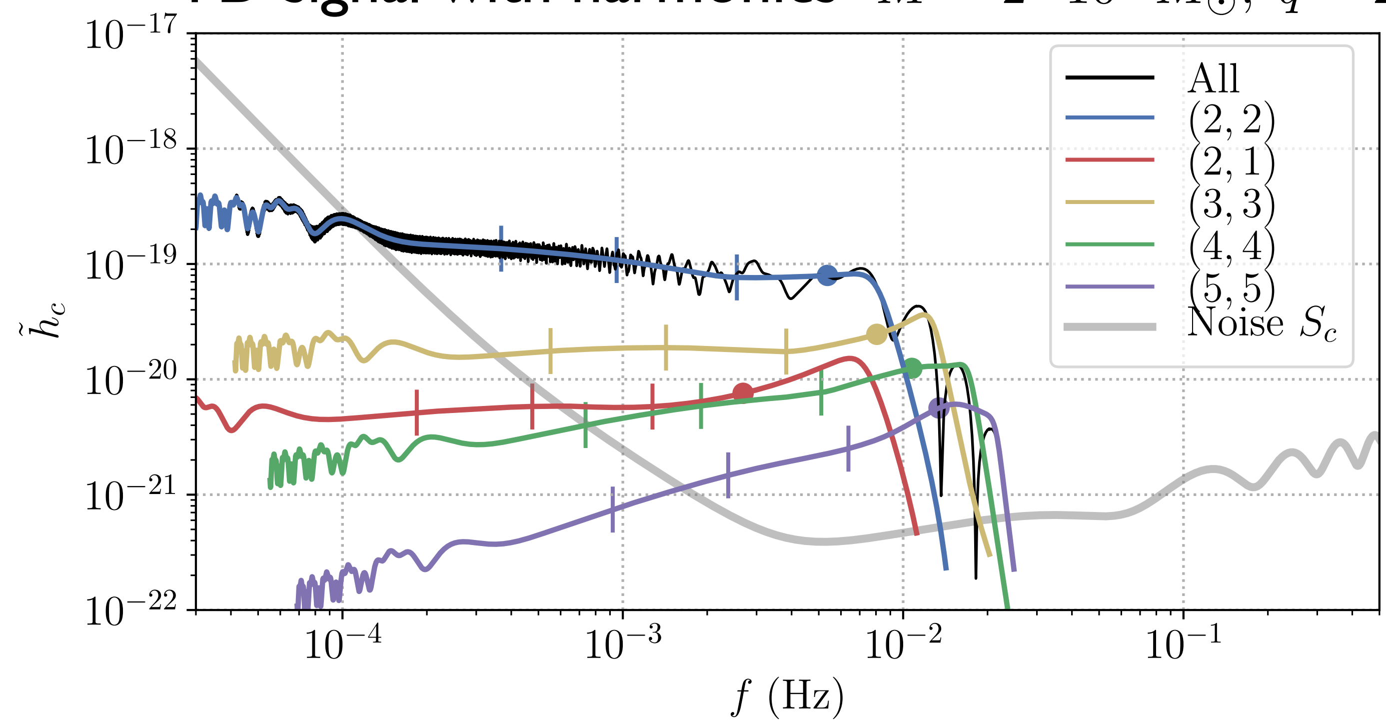
Massive black holes: signals and challenges

- Very loud sources, SNRs of several thousands !
- Detection of merger easy, but detection as early as possible ?
- Advance localization for multimessenger observations ?
- Signals can be short (< 1day) and degenerate
- Waveform model systematics for such loud signals ? Biases, residuals for other sources ?
- Subdominant features in the signal are important

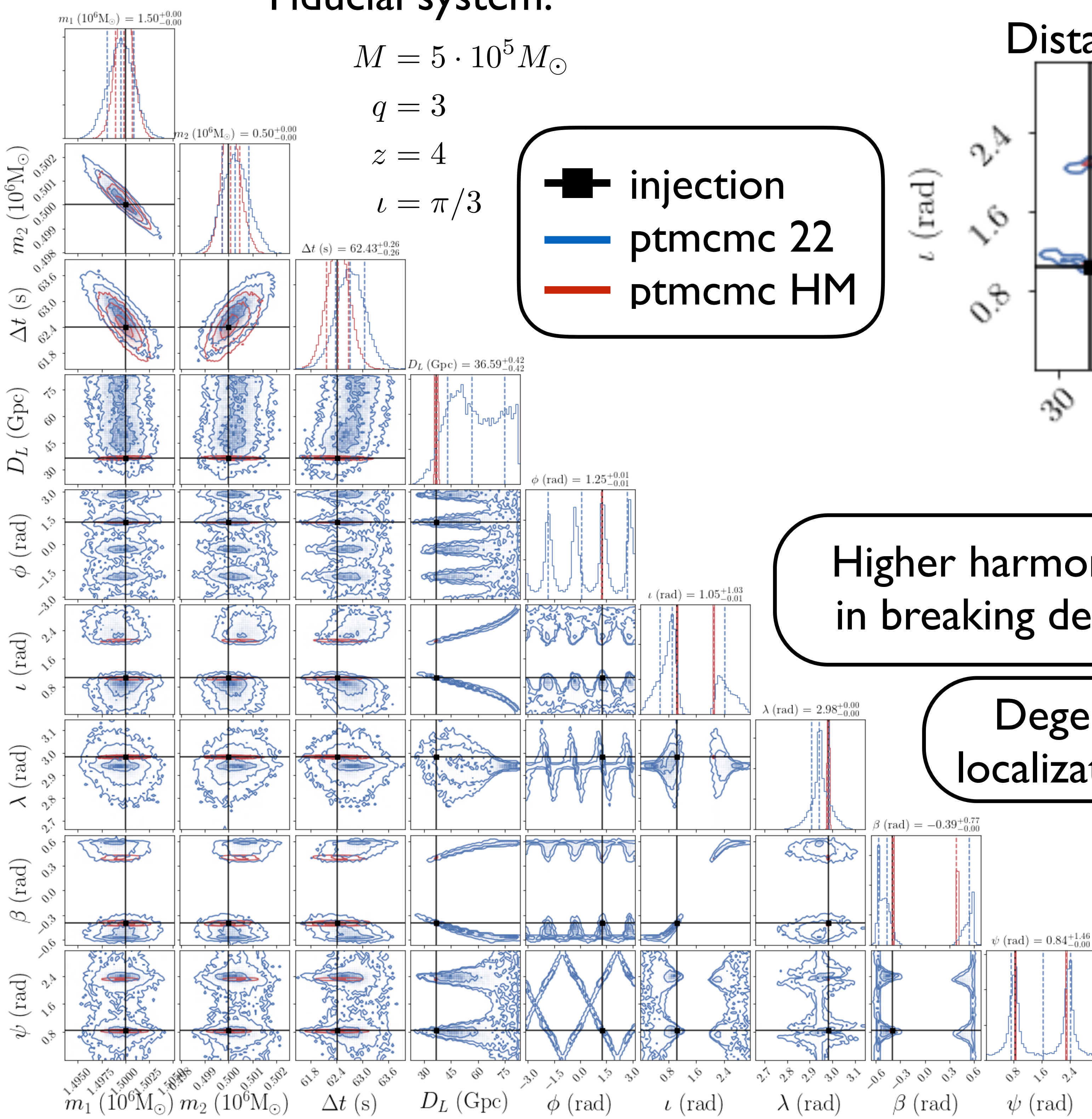
MBHB merger in time-domain



FD signal with harmonics $M = 2 \cdot 10^6\ M_\odot,\ q = 2$



Massive black holes: example results

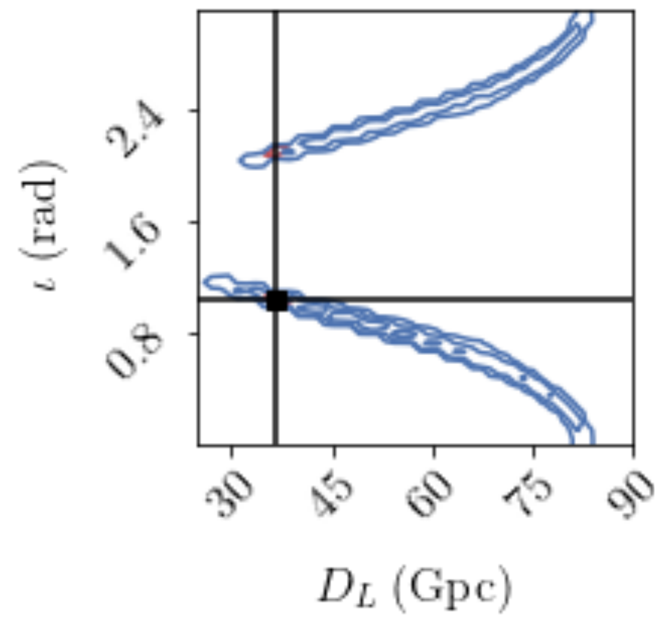


Fiducial system:

$$M = 5 \cdot 10^5 M_\odot$$
$$q = 3$$
$$z = 4$$
$$\iota = \pi/3$$

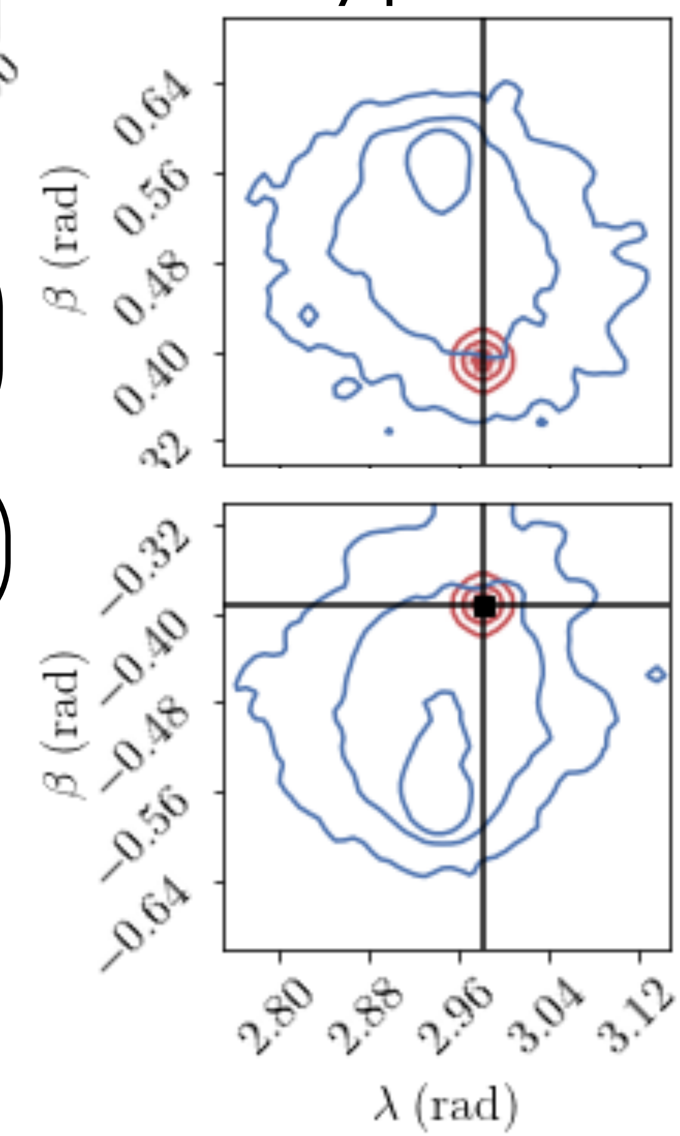
injection
ptmcmc 22
ptmcmc HM

Distance-inclination



[Marsat&al 2020]

Sky position



Higher harmonics crucial in breaking degeneracies

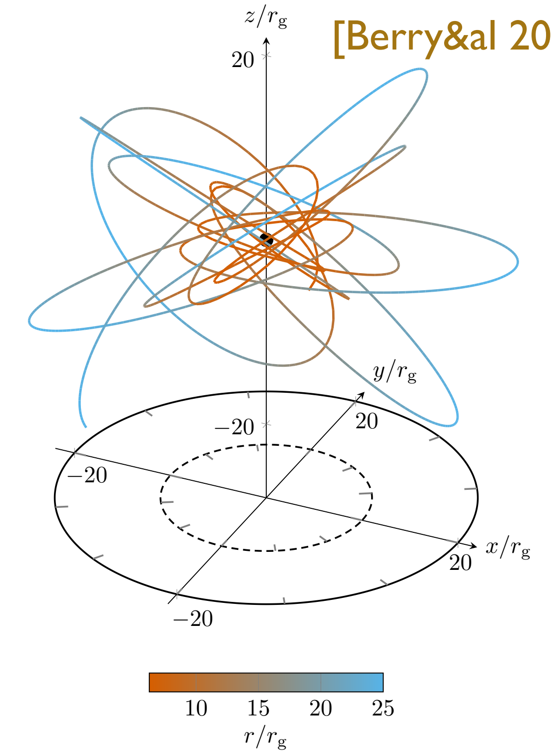
Degenerate sky localization possible

Extreme mass ratio inspirals, stellar-mass black holes

EMRIs

- Long-lived, complex signals, large number of wave cycles $(10^4 - 10^5)$
- Strong precession and eccentricity features, orbits in the relativistic regime around Kerr
- Exquisite determination of some parameters — also means that the signals are hard to find !
- Theoretical work on waveform models needed

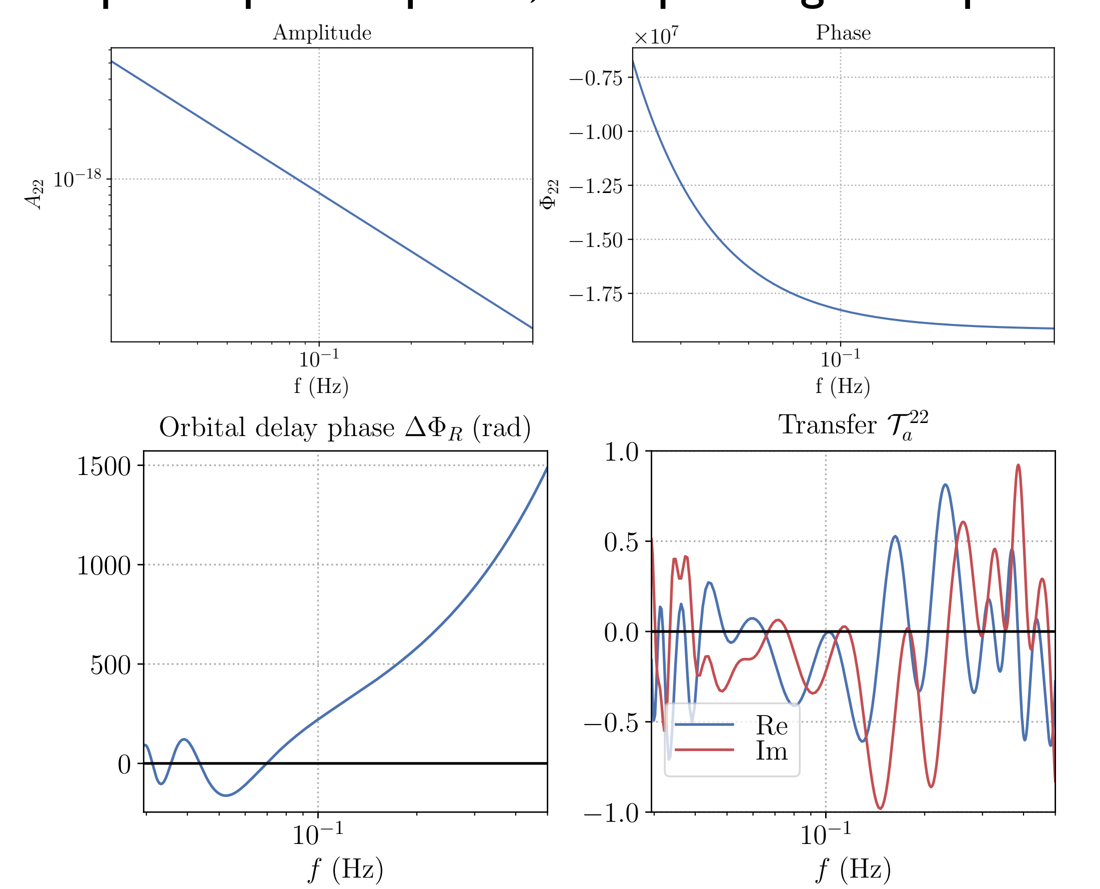
[Berry&al 2019]



Stellar-mass BHs

- Quiet signals: a few detections in the LISA band
- Inspiral regime far from merger, very large number of cycles $(10^5 - 10^6)$
- Challenge of detection: template banks impossible
- Multiband analysis, archival searches ?

Simple amplitude/phase, Complex high-f response



LISA Data Challenge and future challenges

LISA Data Challenge

https://lisa-ldc.lal.in2p3.fr/

LDC-1 (Radler):

- Noise known in advance
- Not blind, sources known
- Single class of sources

LDC-2 (Sangria):

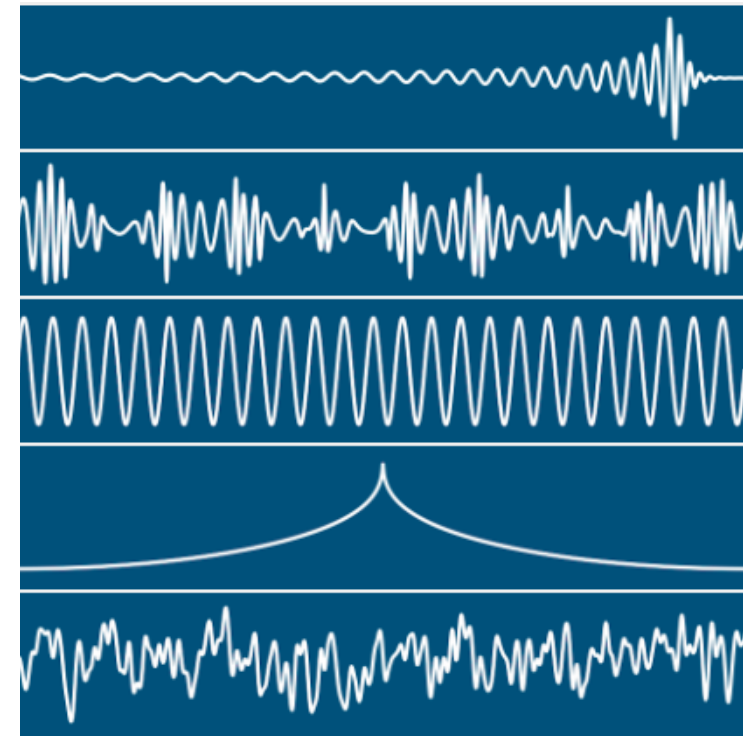
- Noise unknown
- Blind
- MBHBs + full GBs

LDC-2 (Spritz):

- MBHB + data gaps
- Glitches

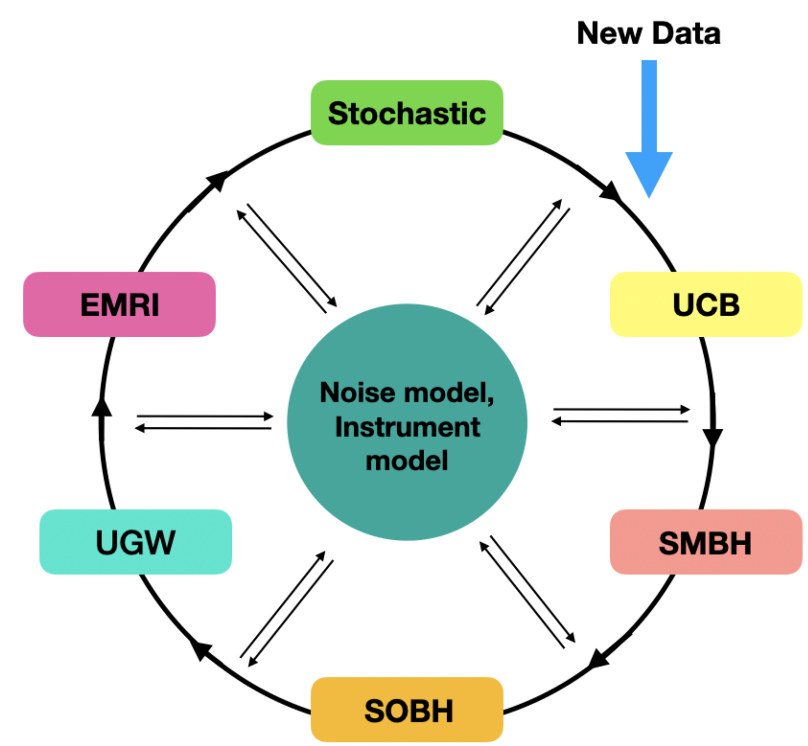
LDC-2 (Yorsh):

- EMRI
- SBHB



The full LISA problem

- Noise reduction (TDI)
- Noise properties must be inferred
- Data artifacts (gaps, …)
- Global fit required, updated regularly



Global fit

Machine learning applications

Reviews (rapidly changing!): [Cuoco&al 2019] https://iphysresearch.github.io/Survey4GWML/

DetChar, glitches

- Glitch classification (CNN, image recognition in time-frequency)
- Supervised Learning: GravitySpy, citizen science
- Denoising

Detection

- CNNs for detection

Low-latency

- Sky map generation

Waveform models

- Remnant properties
- Coefficients on a reduced basis
- Interpolating waveforms (GPRs)

Parameter estimation

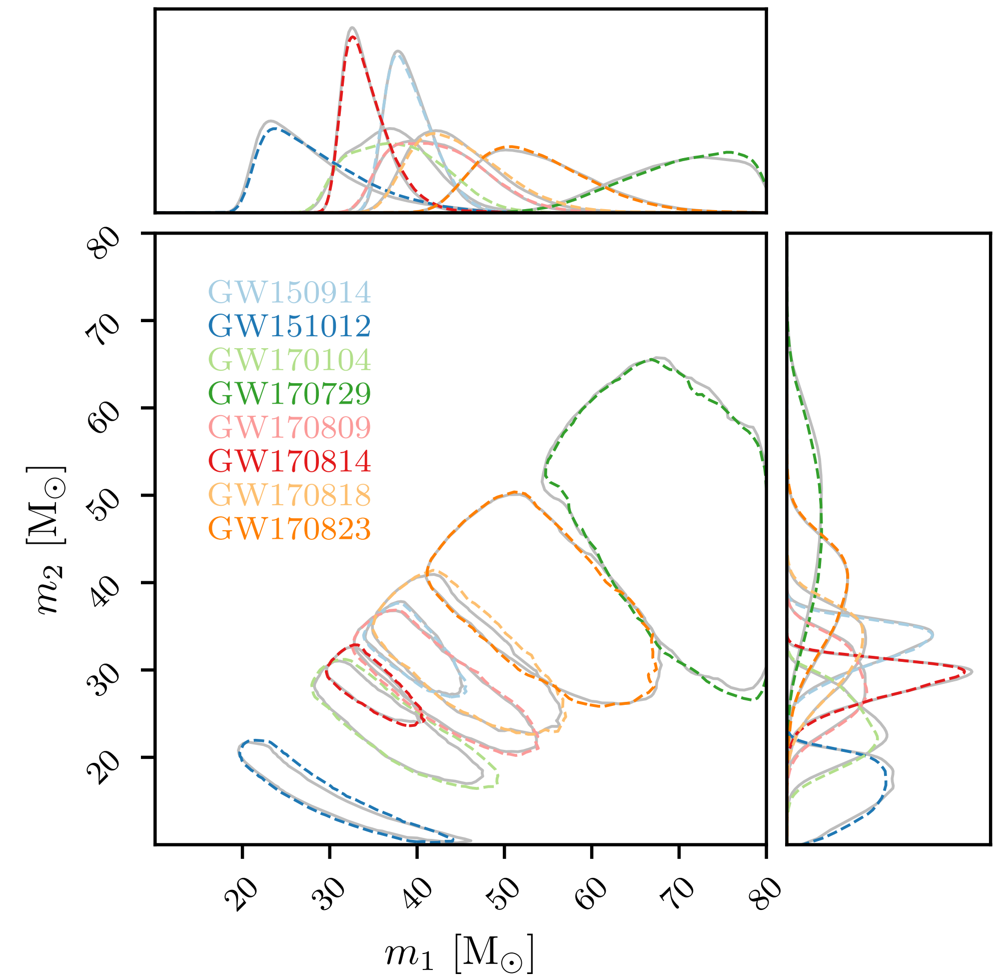
Variational inference, likelihood-free

- Variational auto-encoder (CVAE)
- Normalizing flows

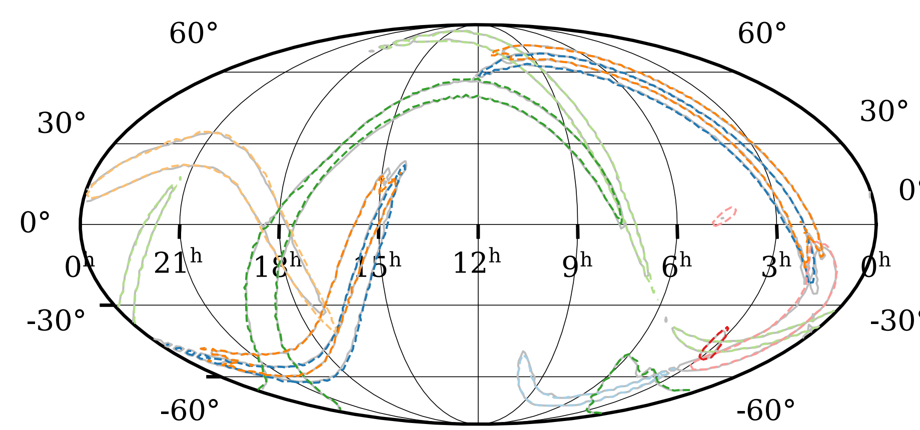
Machine learning example - fast PE

[Dax&al 2021]

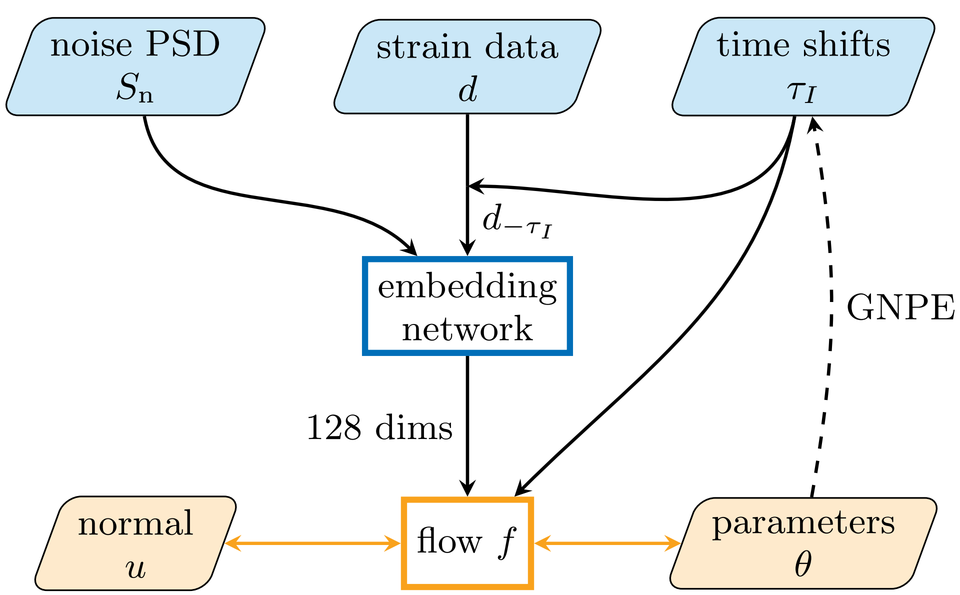
- Normalizing flows
- Embedding network
- Guided neural parameter estimation



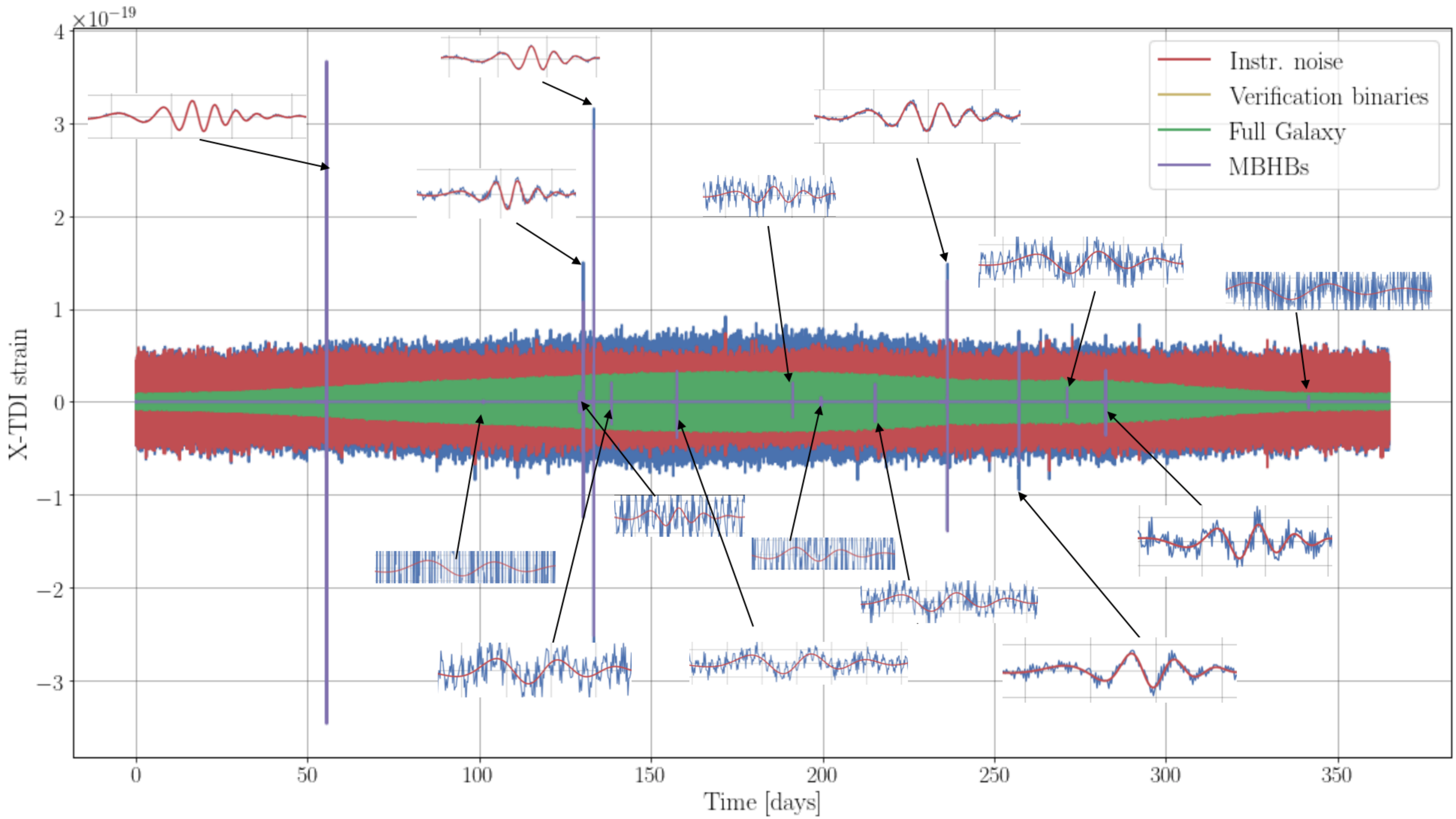
(a)



(b)

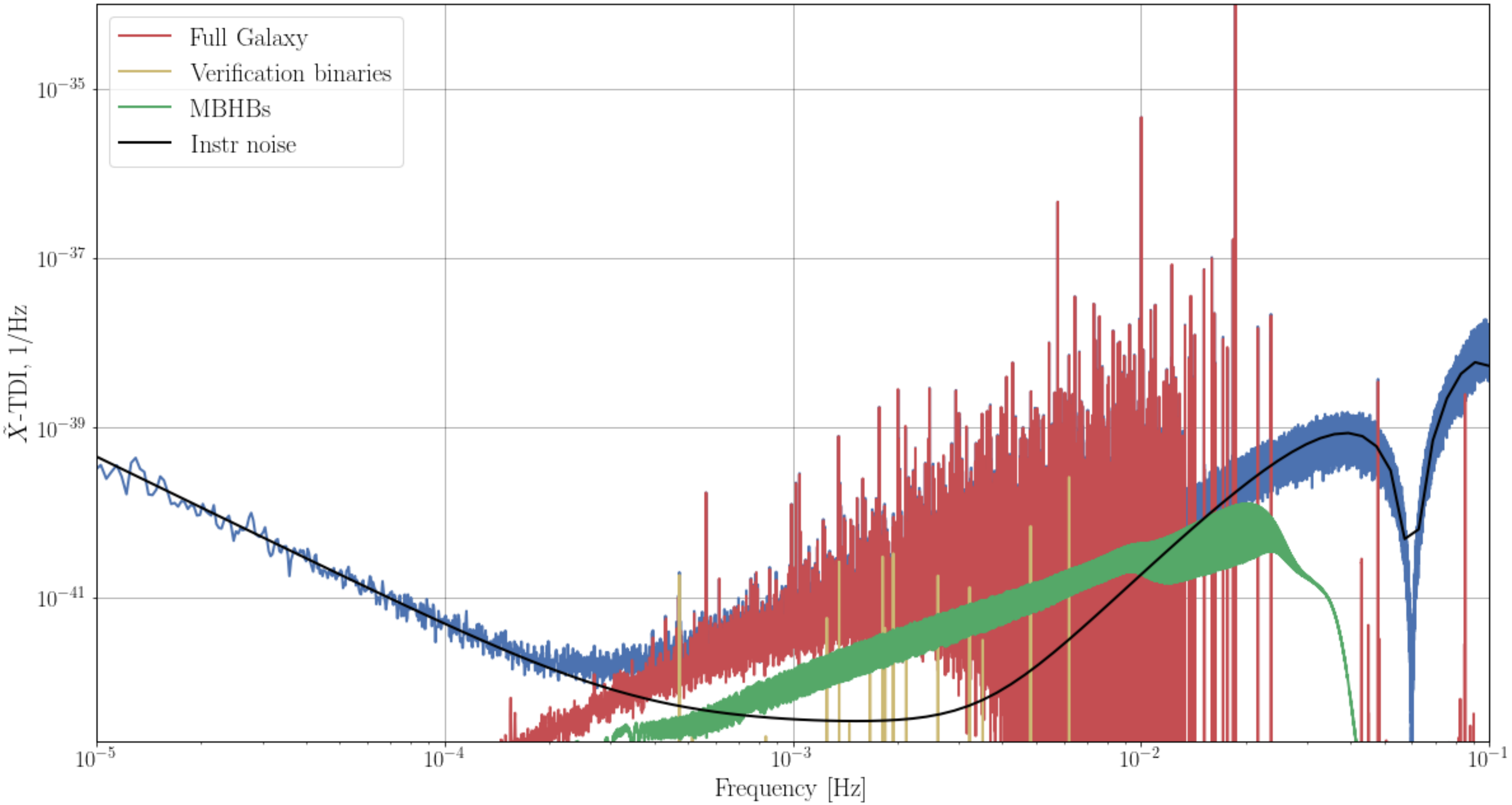


LISA data - LDC-2 Sangria Time-Domain



- **MBHBs**: loudest ones clearly visible by eye above the noise
- **GBs**: superposed signals, annual modulation due to the LISA motion

LISA data - LDC-2 Sangria Frequency-Domain



- **MBHBs**: loudest ones visible in the spectrum, subdominant
- **GBs**: signals local in frequency, both individually resolvable and building up a background

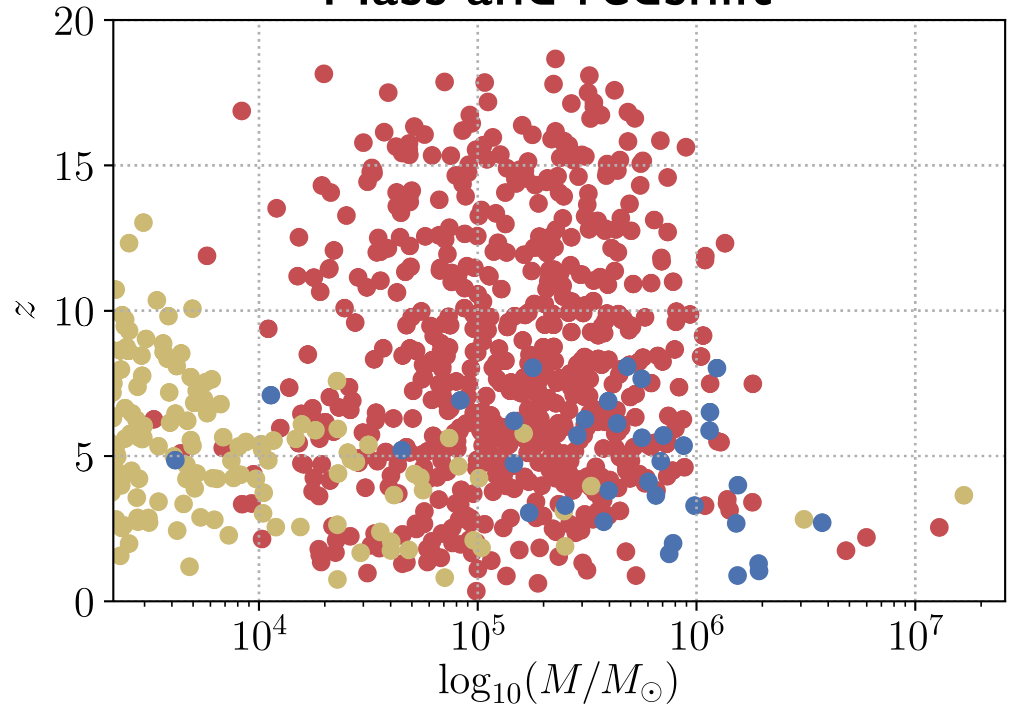
LISA: simulated catalog for MBHB astrophysical models

[Barausse 2012]

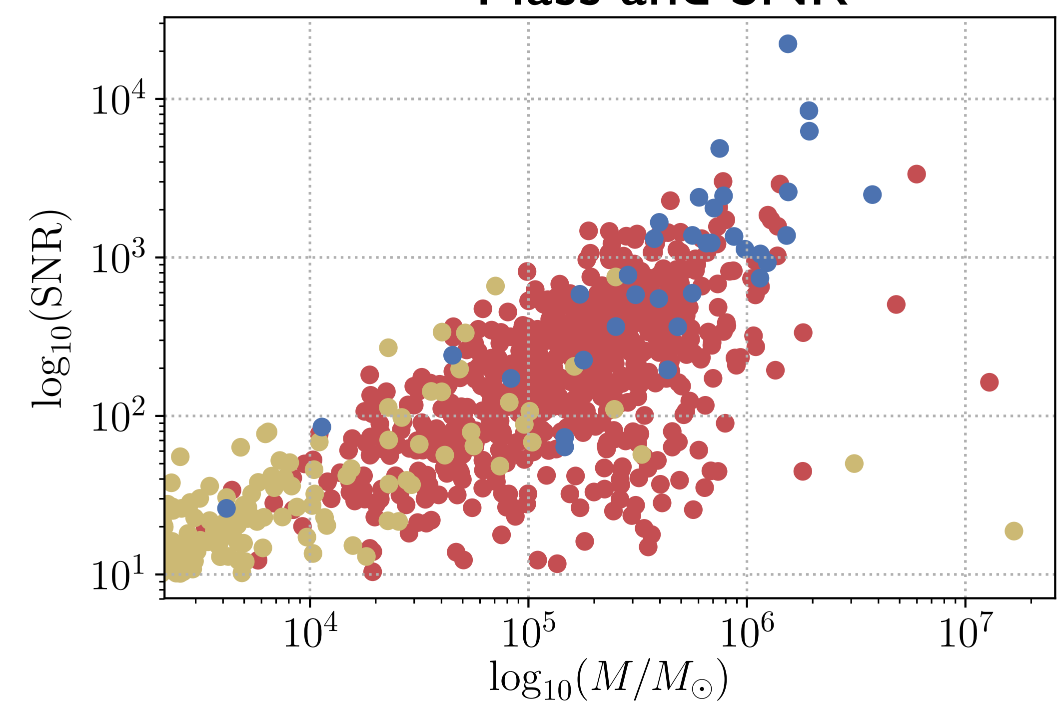
Astrophysical models:

- Heavy seeds - delay
- Light seeds - no delay
- PopIII seeds - delay

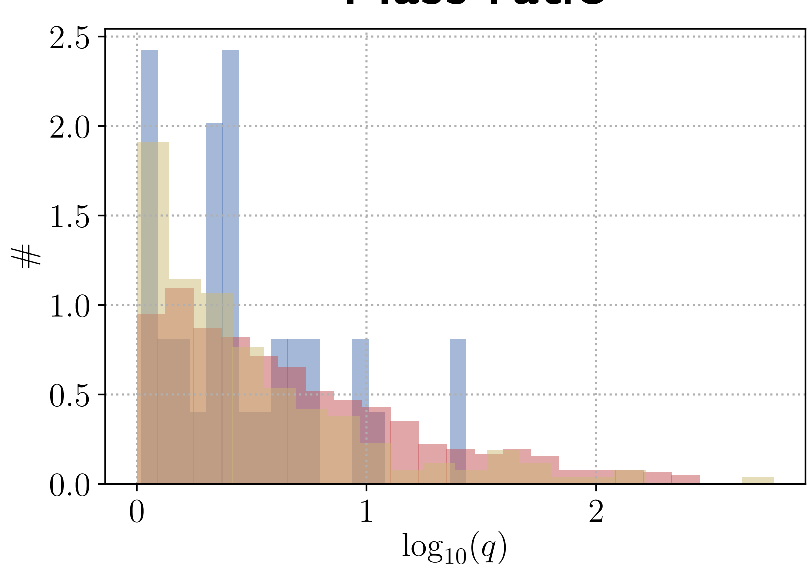
Mass and redshift



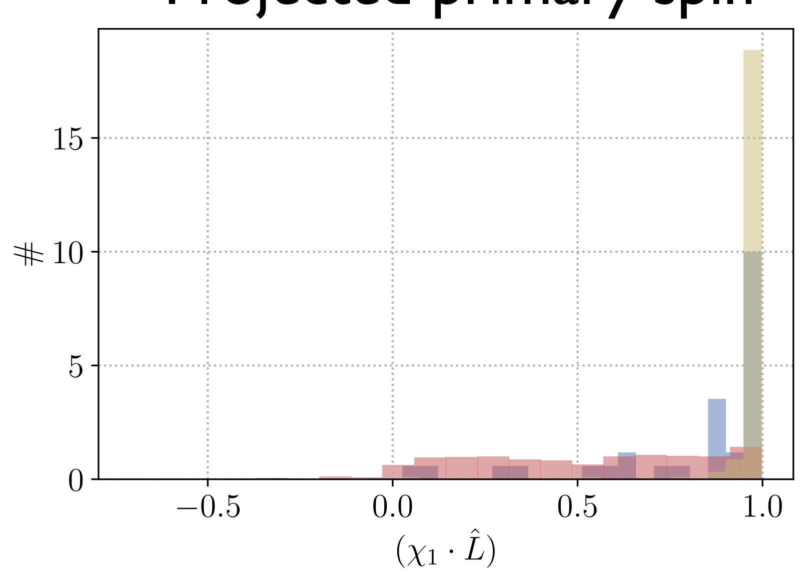
Mass and SNR



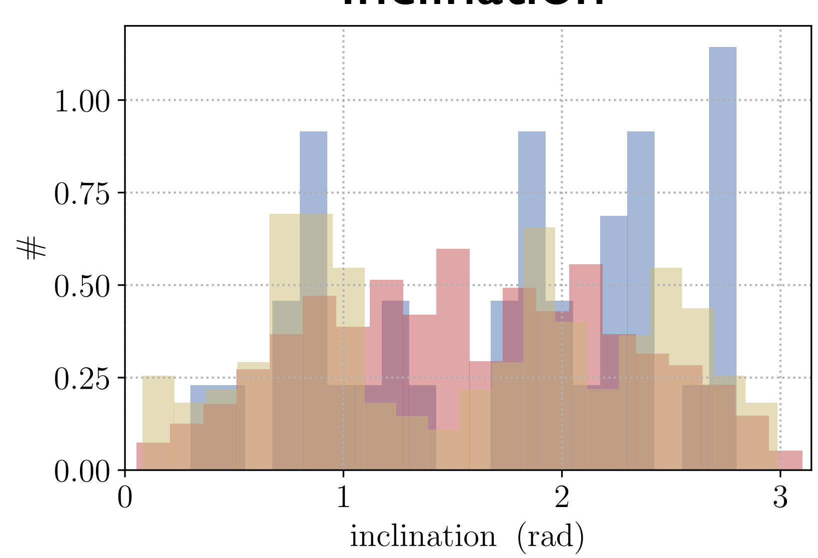
Mass ratio



Projected primary spin



Inclination



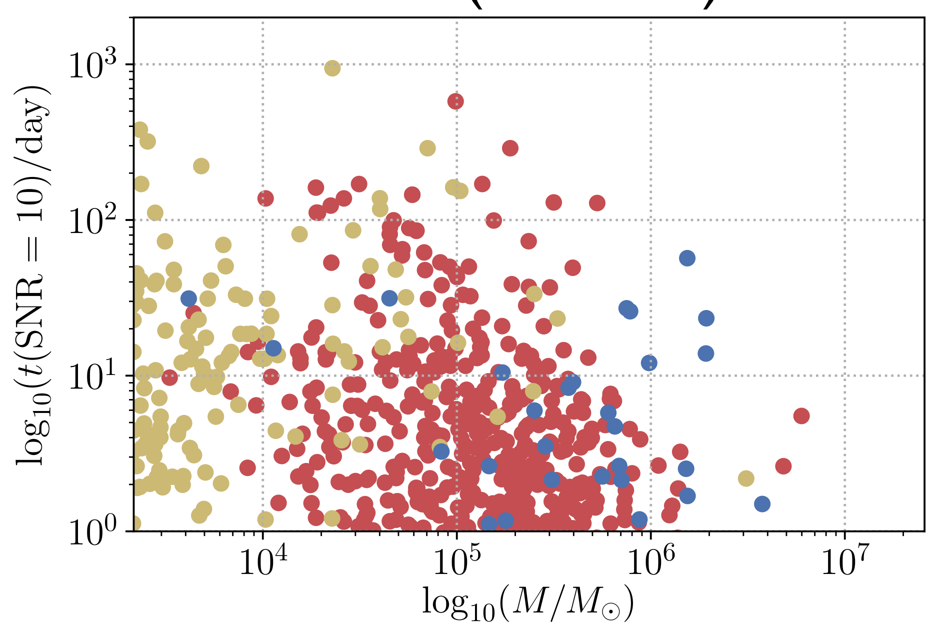
LISA: simulated catalog for MBHB astrophysical models

[Barausse 2012]

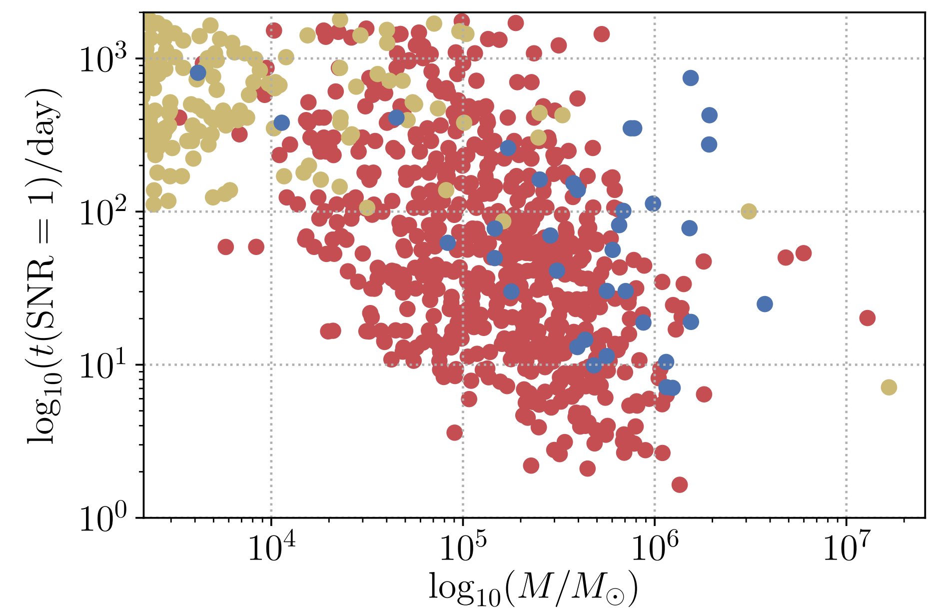
Astrophysical models:
- Heavy seeds - delay
- Light seeds - no delay
- PopIII seeds - delay

MBHB detected signals:
Bulk shorter than ~10days
Tail extending to ~3months

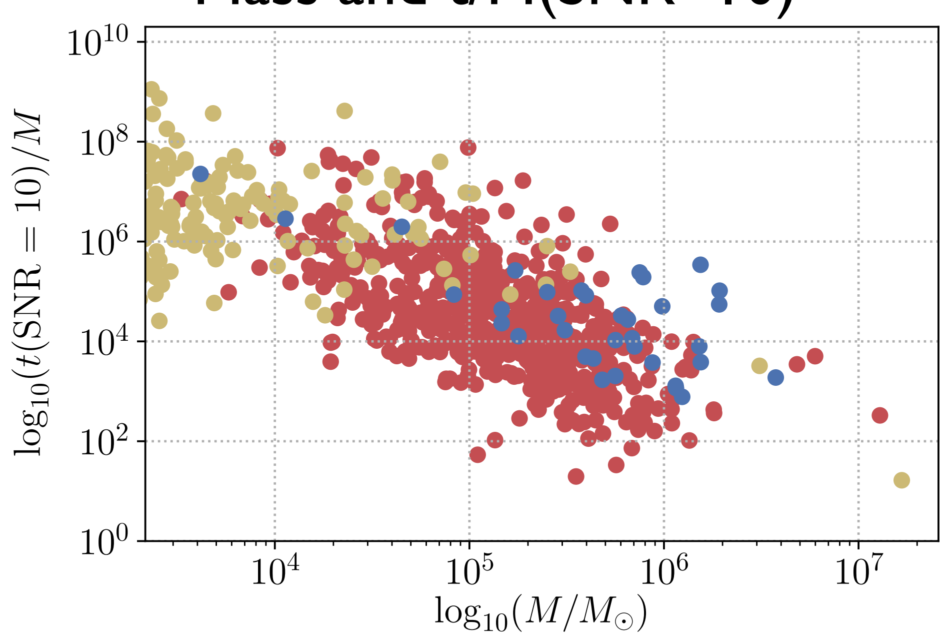
Mass and t(SNR=10)



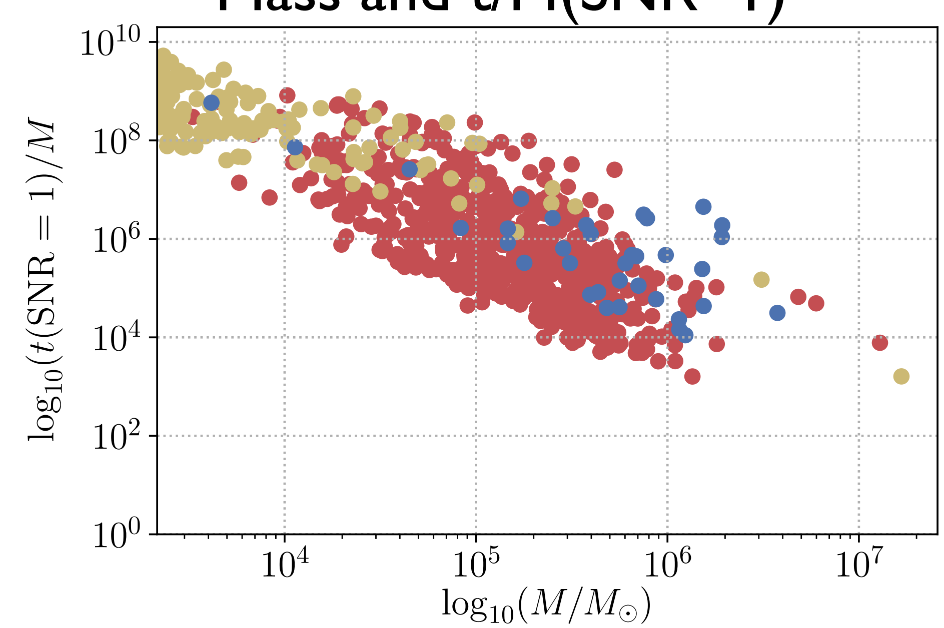
Mass and t(SNR=1)



Mass and t/M(SNR=10)



Mass and t/M(SNR=1)



SMBH PE: accumulation of information with time

Method

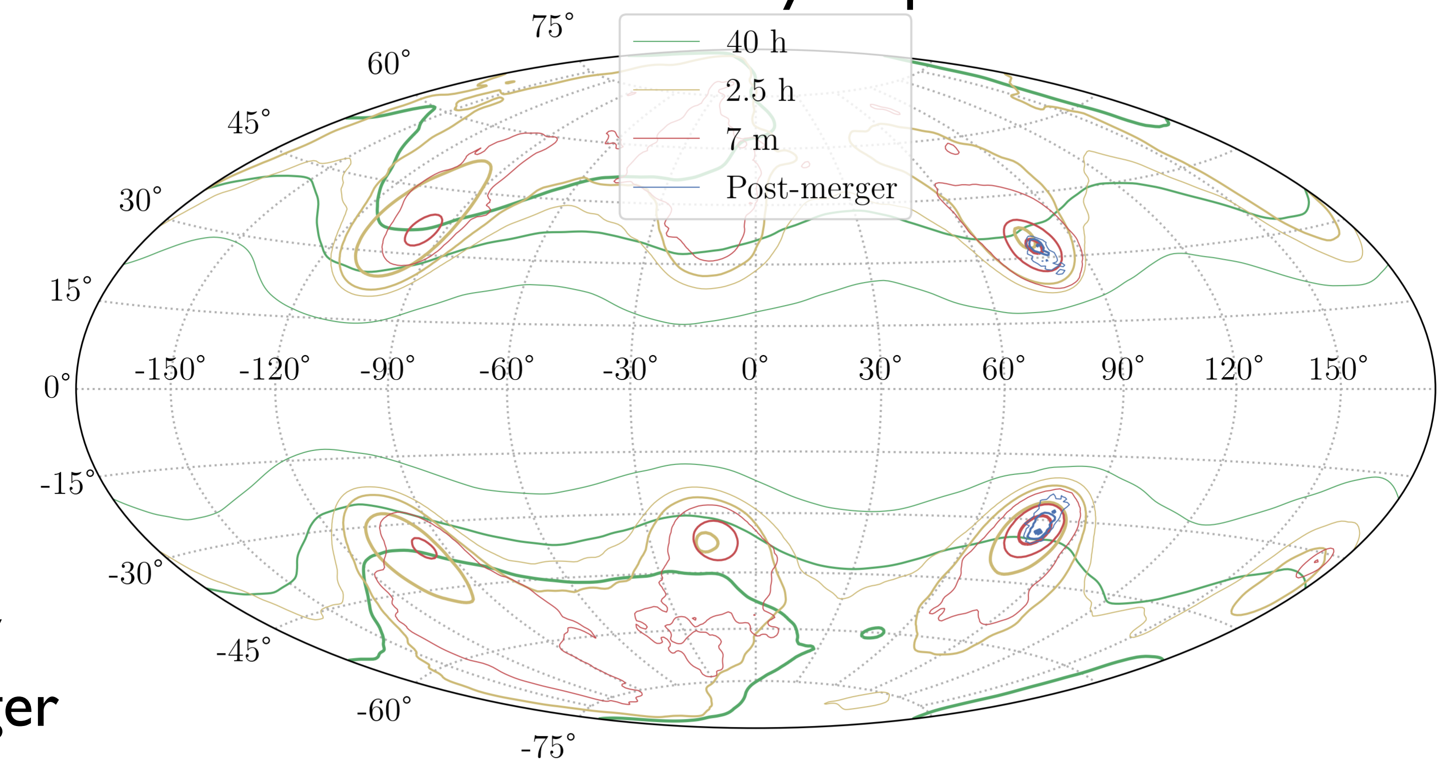
- Represent a cut in time-to-merger by a cut in frequency, becomes inaccurate at merger
- Use Multinest and PTMCMC with and without higher harmonics

8-maxima sky degeneracy
only broken shortly before merger
2-maxima sky degeneracy
survives after merger

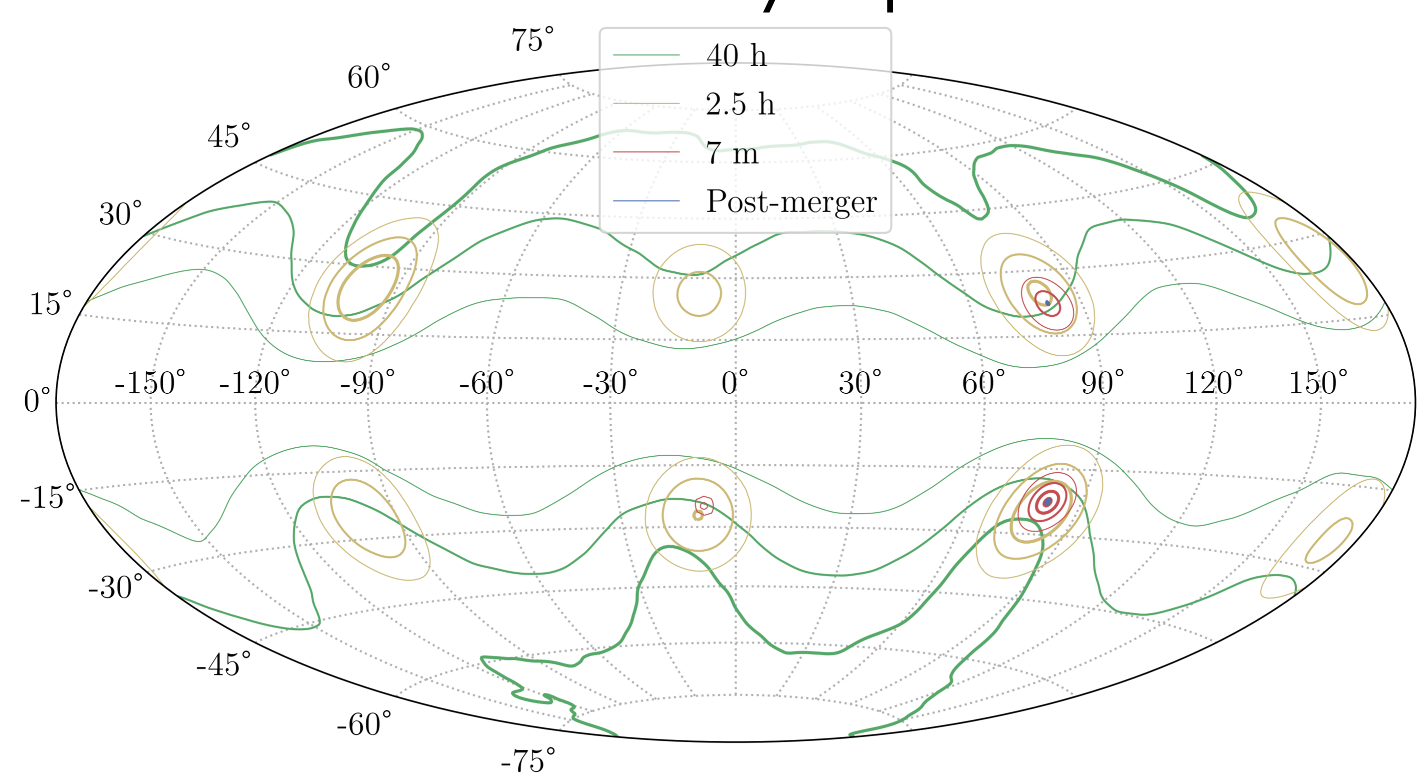
SNR-based time cuts:

SNR	DeltaT
10	40h
42	2.5h
167	7min
666	-

LISA-frame sky map 22



LISA-frame sky map hm



LISA instrument response

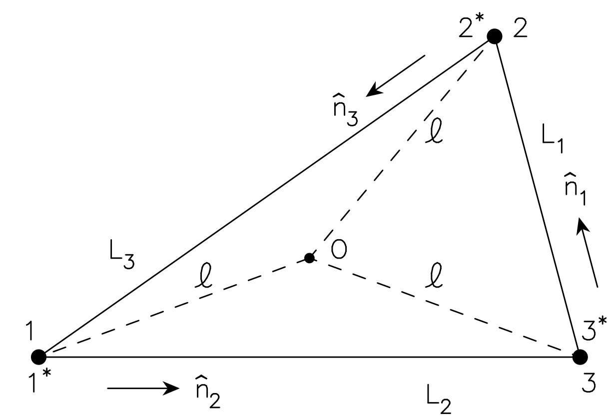
One-arm frequency observables

From spacecraft s to spacecraft r through link s: $y = \Delta\nu/\nu$

$$y_{slr} = \frac{1}{2}\frac{1}{1-\hat{k}\cdot n_l} n_l \cdot (h(t_s) - h(t_r)) \cdot n_l$$

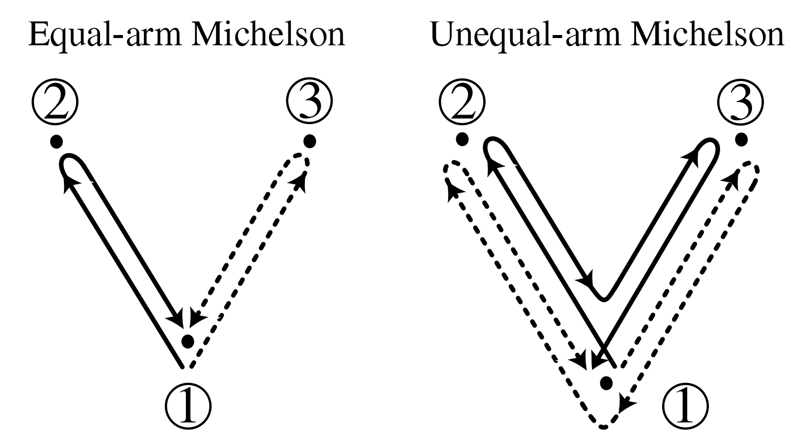
$$t_s = t - L - \hat{k}\cdot p_s, \quad t_r = t - \hat{k}\cdot p_r$$

$$h = h_+ P_+(\hat{k}) + h_\times P_\times(\hat{k})$$ GW at SSB



Time-delay interferometry (TDI)

- Crucial to cancel laser noise
- First generation: unequal arms
- Second generation: propagation and flexing
- Michelson X,Y,Z - Uncorrelated noises A,E,T

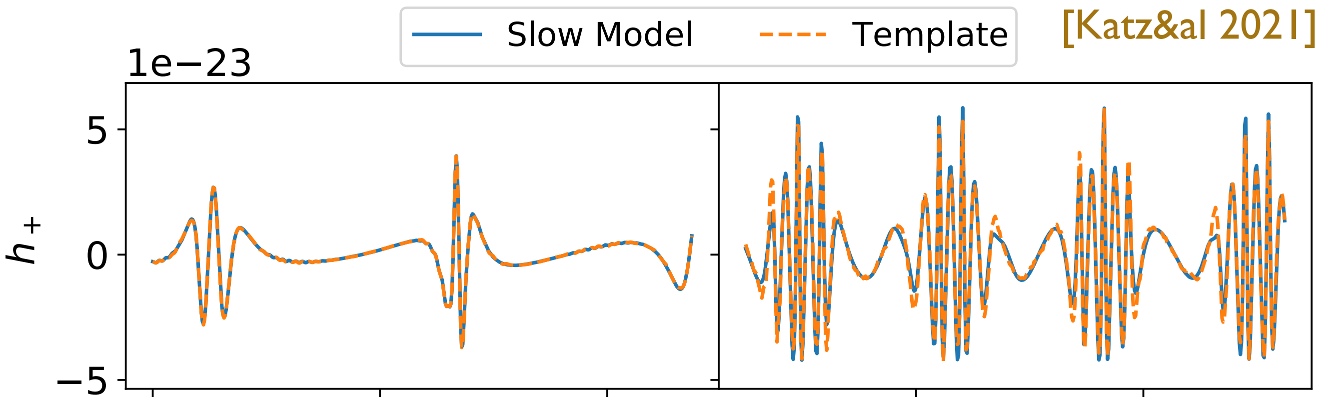


$$X_1^{\mathrm{GW}} = \underbrace{\left[(y_{31}^{\mathrm{GW}} + y_{13,2}^{\mathrm{GW}}) + (y_{21}^{\mathrm{GW}} + y_{12,3}^{\mathrm{GW}})_{,22} - (y_{21}^{\mathrm{GW}} + y_{12,3}^{\mathrm{GW}}) - (y_{31}^{\mathrm{GW}} + y_{13,2}^{\mathrm{GW}})_{,33}\right]}_{X^{\mathrm{GW}}(t)}$$
$$- \underbrace{\left[(y_{31}^{\mathrm{GW}} + y_{13,2}^{\mathrm{GW}}) + (y_{21}^{\mathrm{GW}} + y_{12,3}^{\mathrm{GW}})_{,22} - (y_{21}^{\mathrm{GW}} + y_{12,3}^{\mathrm{GW}}) - (y_{31}^{\mathrm{GW}} + y_{13,2}^{\mathrm{GW}})_{,33}\right]_{,2233}}_{X^{\mathrm{GW}}(t-2L_2-2L_3)\simeq X^{\mathrm{GW}}(t-4L)} .$$

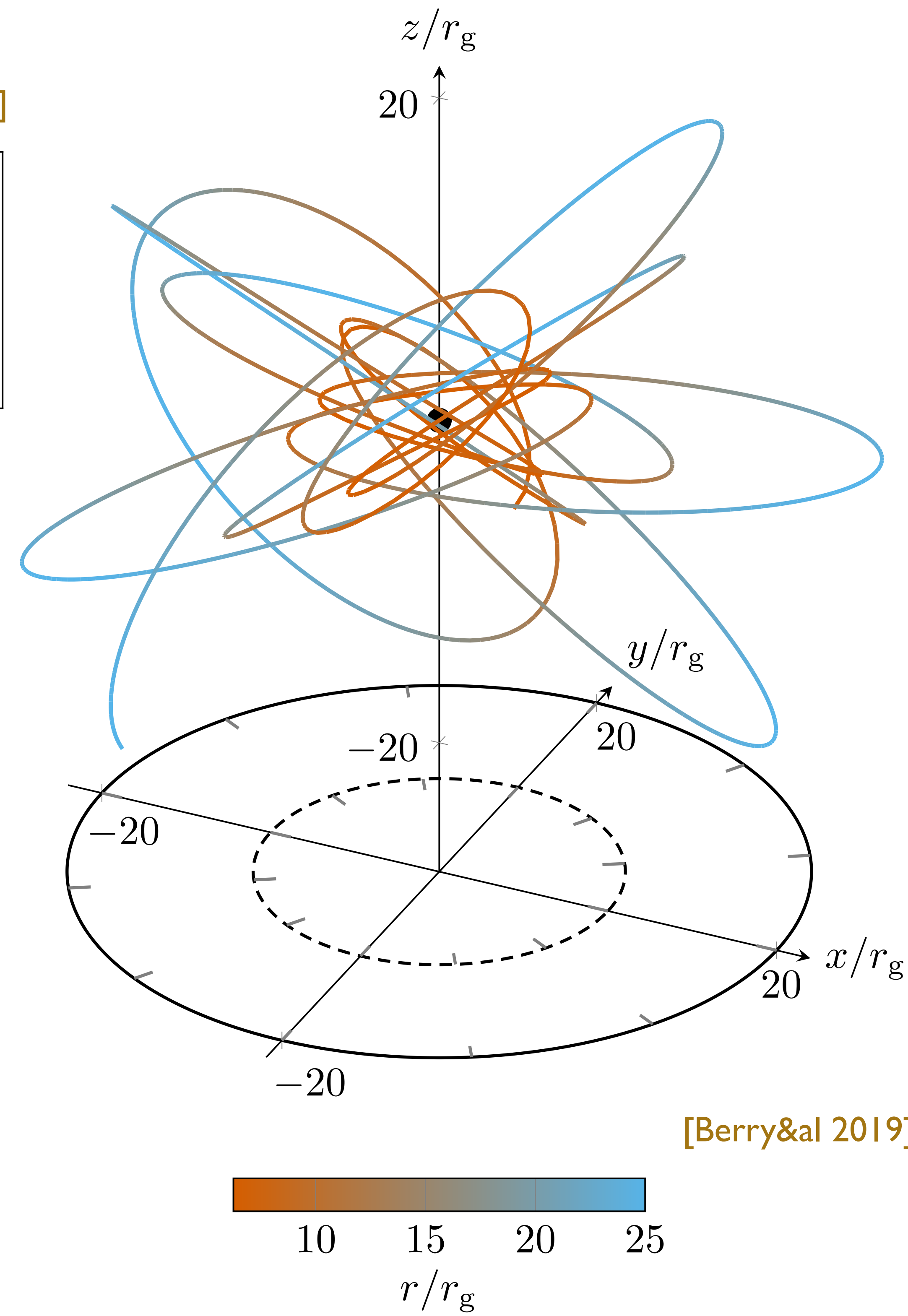
Approximations

- Long-wavelength approximation: two moving LIGOs rotated by $\pi/4$ + orbital delay
- **Rigid approximation** (order of the delays does not matter, delay=L simple in Fourier domain)

Extreme mass ratio inspirals: signals and challenges

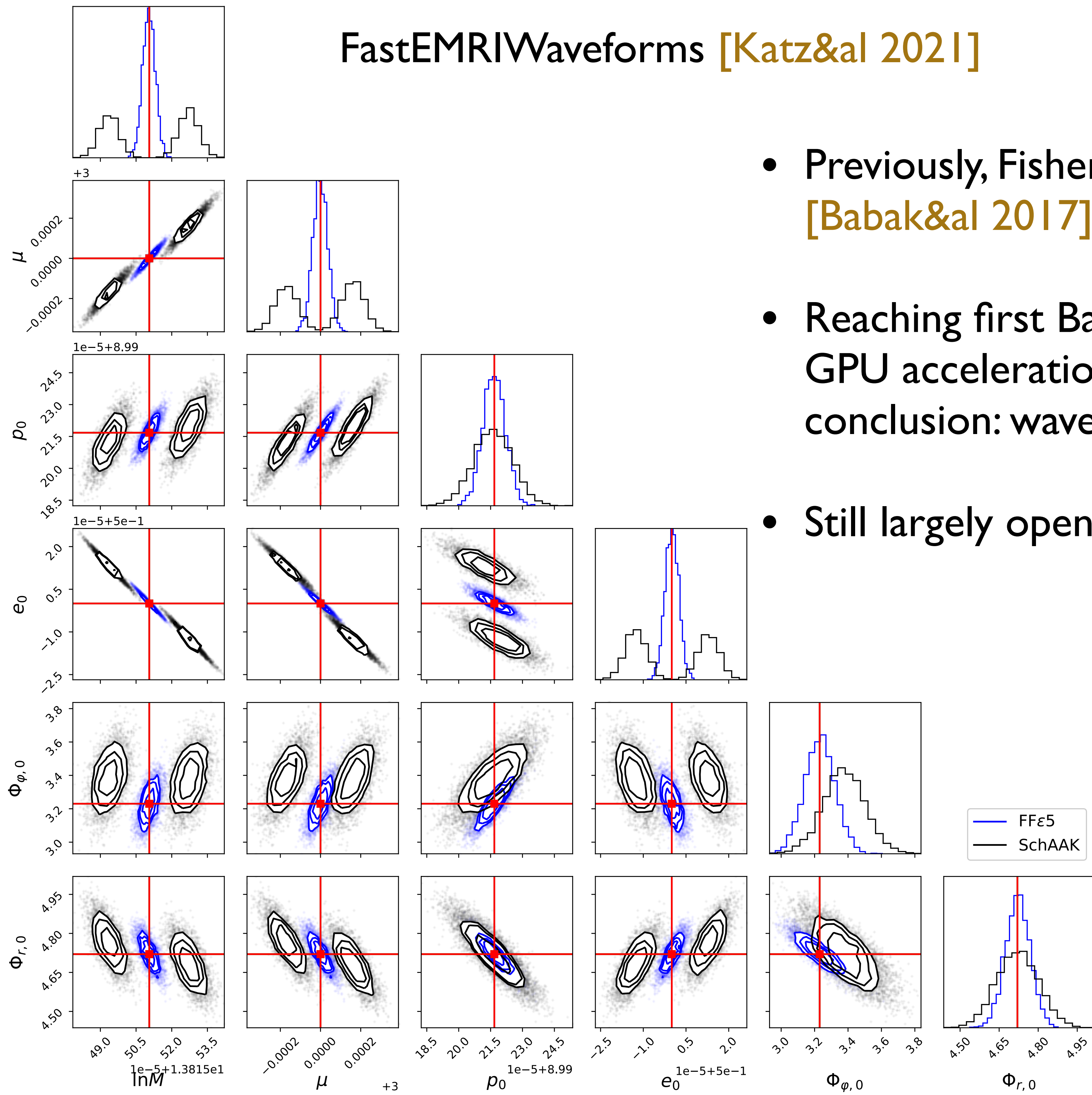


- Long-lived signals, large number of wave cycles
- Strong precession and eccentricity features, orbits in the relativistic regime around Kerr
- Decomposition in harmonics: ~100 modes
- Exquisite determination of some parameters — also means that the signals are hard to find !
- Theoretical work on waveform models needed (2nd-order Self-Force)



Extreme mass ratio inspirals: example results

FastEMRIWaveforms [Katz&al 2021]



- Previously, Fisher matrix estimates (e.g. [Babak&al 2017])
- Reaching first Bayesian PE results, thanks to GPU acceleration of the waveforms — conclusion: waveform models matter
- Still largely open problem !

SBHB signal: Fourier-domain signal and response

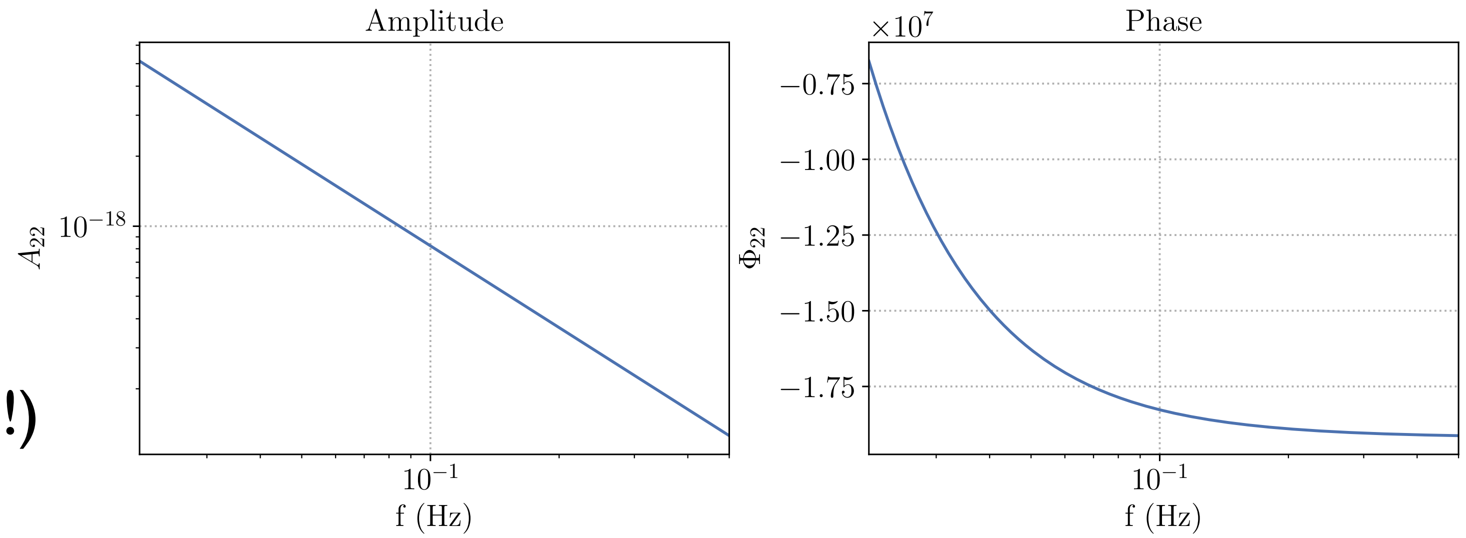
Early inspiral signal

Here simple amplitude and phase

$$A(f) \sim f^{-7/6}$$

$$\Phi(f) \sim f^{-5/3}$$

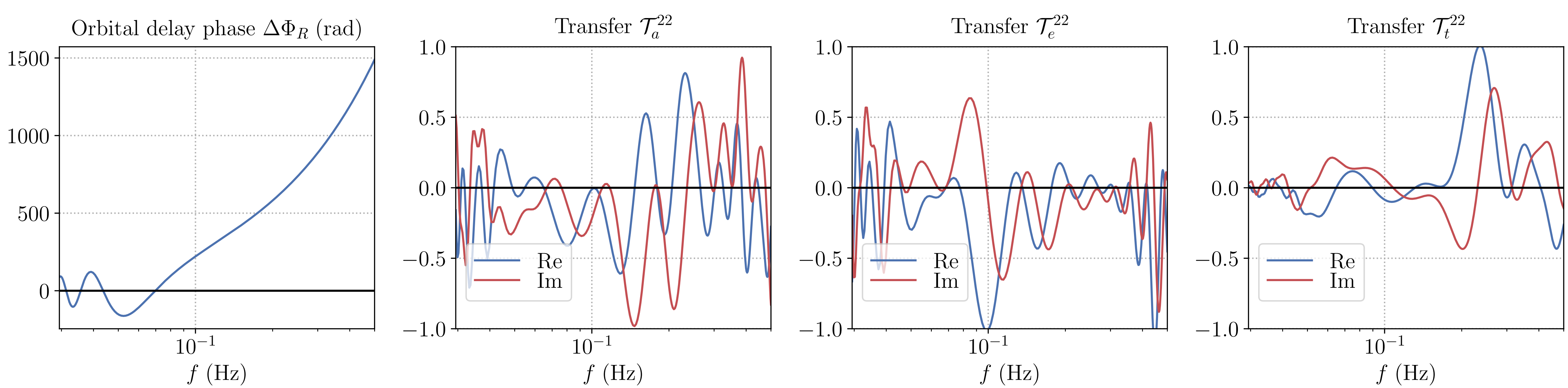
(but might have eccentricity+precession !)



Response decomposed

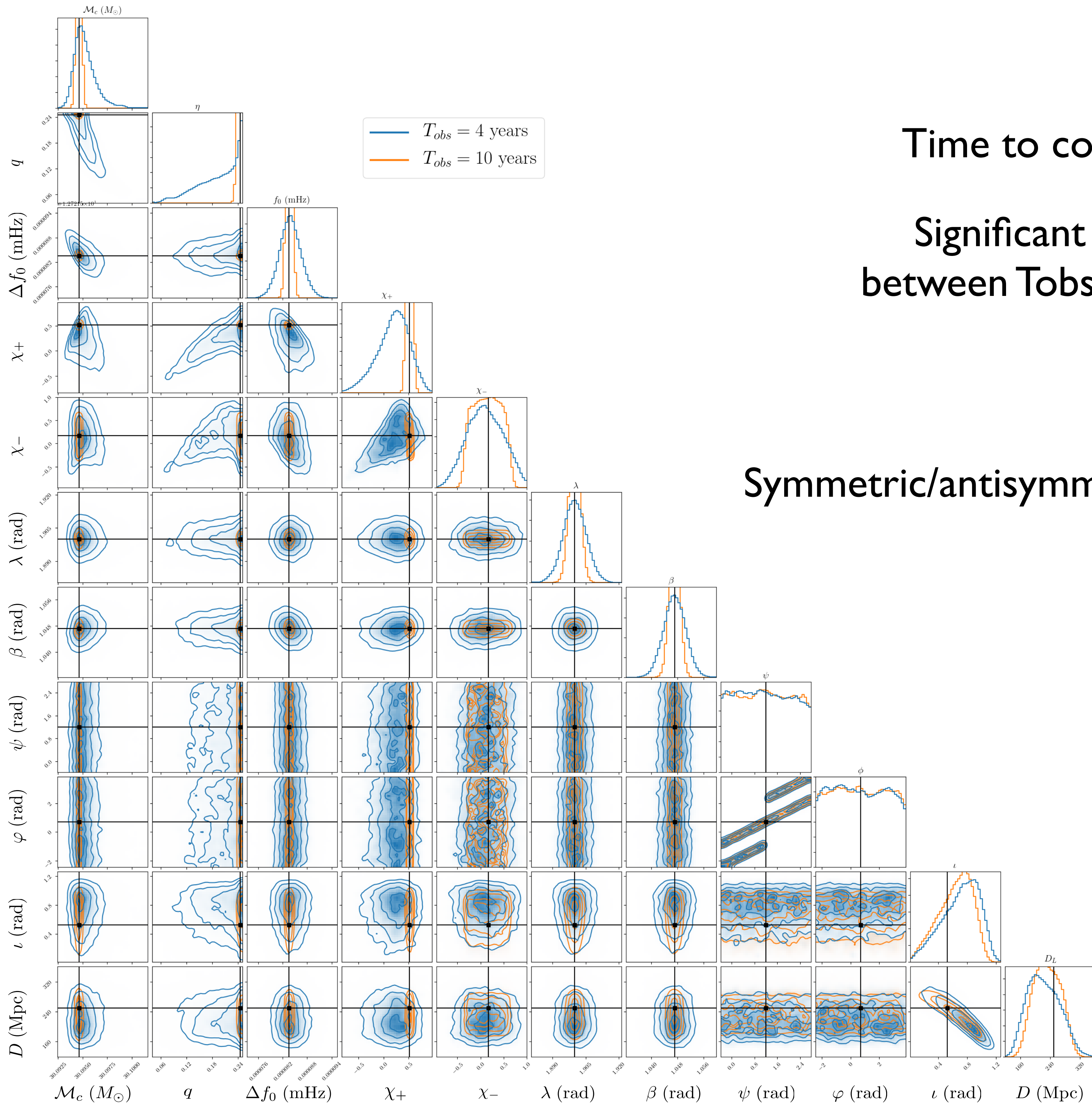
$$\mathcal{T}_{slr} = \frac{i\pi f L}{2} \mathrm{sinc}\left[\pi f L \left(1 - k\cdot n_l\right)\right] \exp\left[i\pi f \left(L + k\cdot (p_r + p_s)\right)\right] n_l \cdot P \cdot n_l(t_f)$$

+ Doppler phase (delay to the center of constellation): $\exp\left[2i\pi f k\cdot p_0(t_f)\right]$



Complicated modulations, long-lived signals
Departure from the low-frequency approx.
Large Doppler phase

PE results: SBHBs



Time to coalescence: Tc = 8yrs

Significant qualitative differences between Tobs = 4yrs and Tobs = 10yrs

Symmetric/antisymmetric spin combinations:

$$\chi_+ = \frac{m_1\chi_1 + m_2\chi_2}{m_1 + m_2}$$

$$\chi_- = \frac{m_1\chi_1 - m_2\chi_2}{m_1 + m_2}$$