MCMC parameter estimation methods for LISA massive black holes

Sylvain Marsat (L2IT, Toulouse)

in collaboration with S. Babak (APC), S. Deng (APC), Maude Lejeune (APC), Éric Plagnol (APC), Ollie Burke (L2IT), J. Baker (NASA), T. Dal Canton (IJCLab), A. Mangiagli (APC), R.Cotesta (J. Hopkins), A. Toubiana (APC), M. Katz (AEI), H. Inchauspé (APC), …

Outline

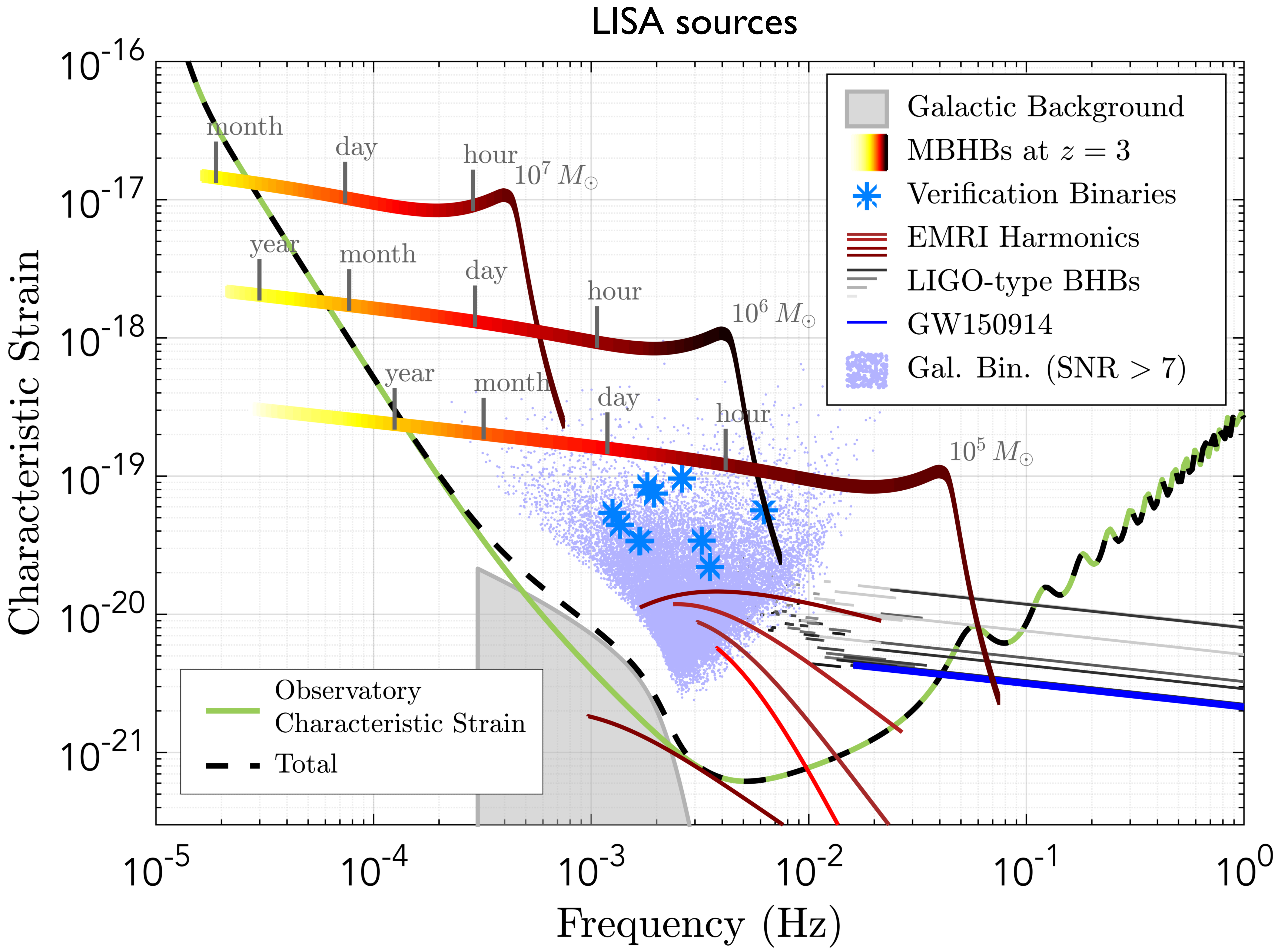
- **MBHB signals in LISA**
- Parameter space degeneracies
- Tools for Bayesian parameter estimation
- LISA Data Challenge: Sangria
- MBHB results for Sangria

The LISA sources and MBHBs

Massive black hole binaries

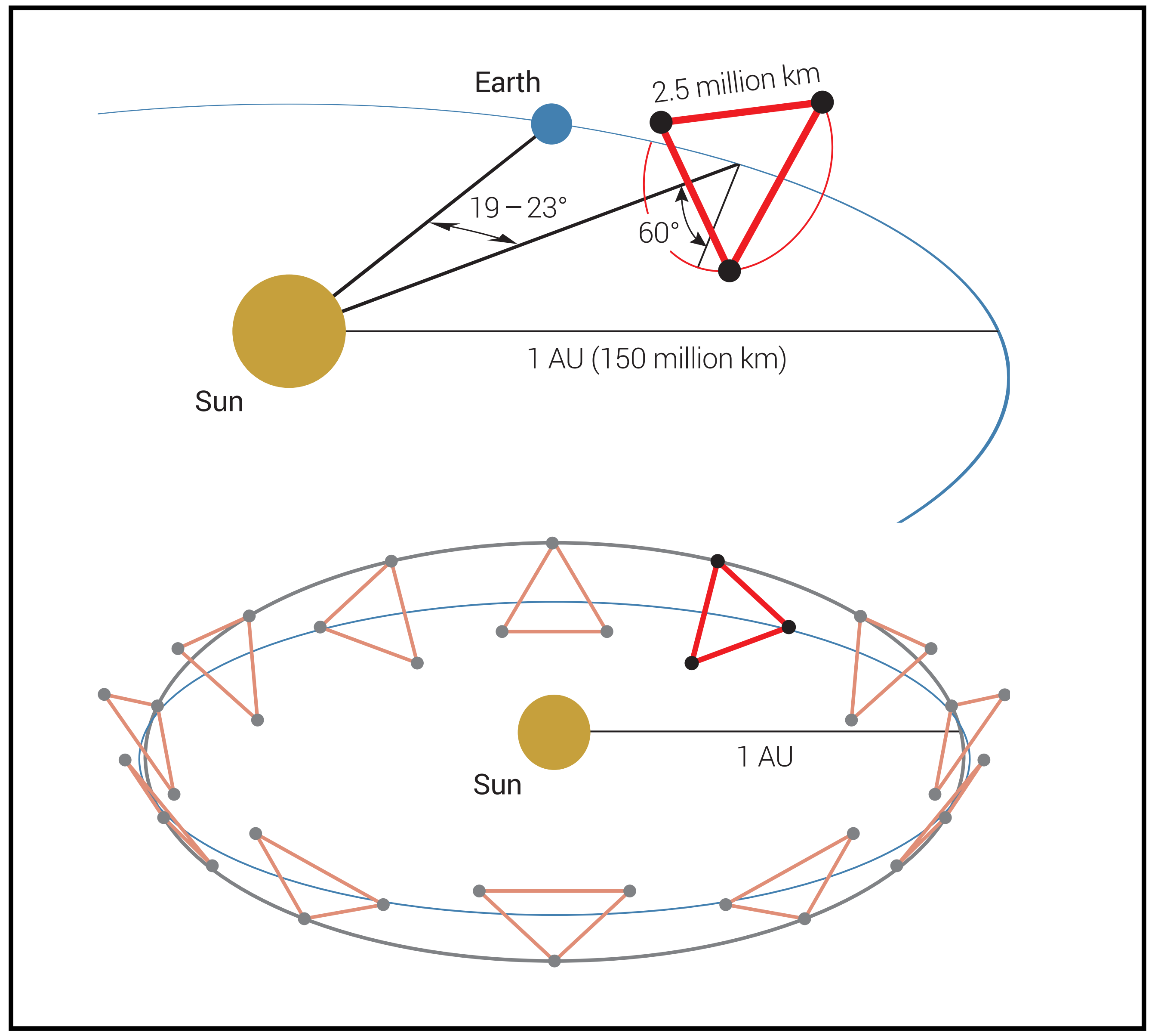
- **Primary sources** for LISA: very loud signals, signal dominated regime
- Precision GW science: details matter, waveform systematics important
- Primary candidates for EM counterparts, crucial for astrophysics and cosmology
- Primary candidates for TGR (with EMRIs): controlling biases and residuals crucial, need tools for extended waveform models

Not fully explored yet:
- IMR waveforms with precession, eccentricity
- Realistic instrument (gaps&glitches), global fit



LISA instrumental response

LISA orbits



Response

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

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

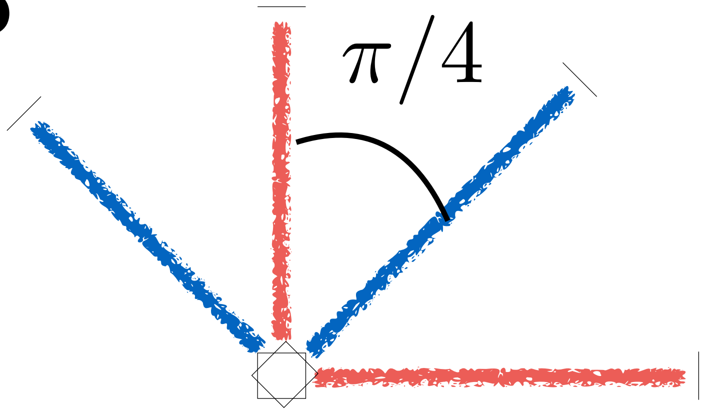
+ Time-delay interferometry (TDI)

Fourier-domain (separation of timescales [Marsat-Baker 2018])

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

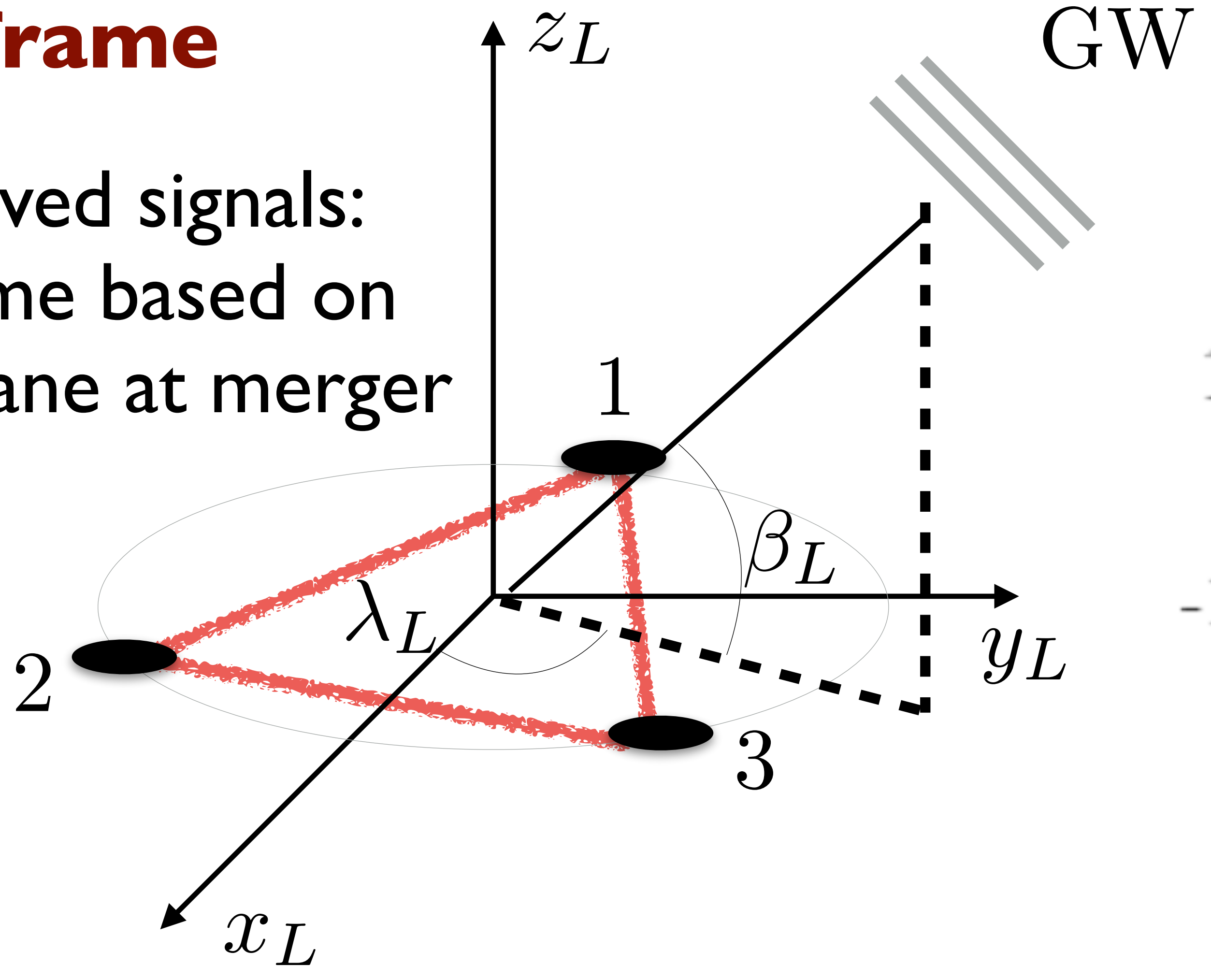
Time and frequency-dependency
Time: motion of LISA on its orbit
Frequency: departure from long-wavelength

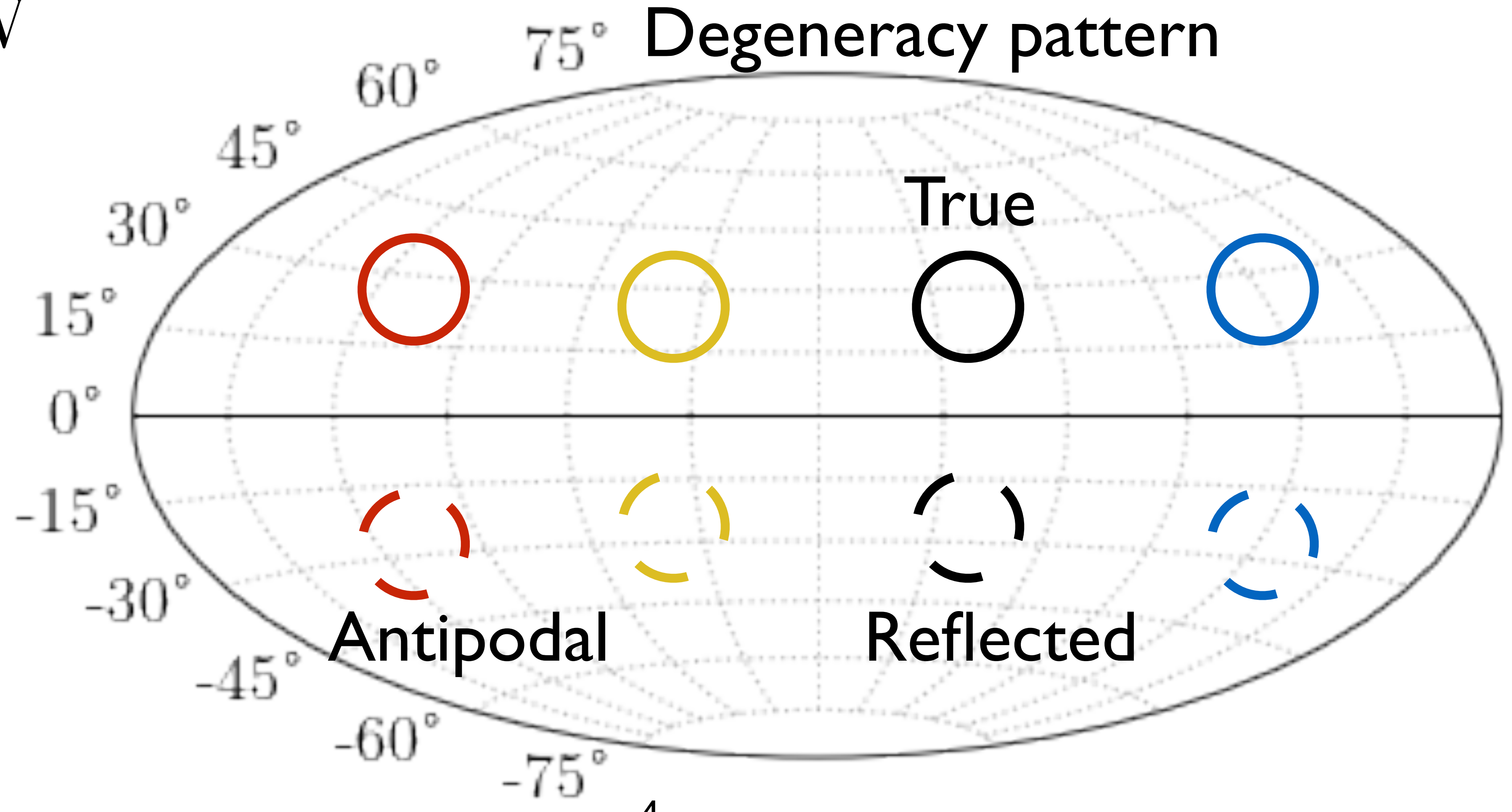
Low-f approximation: **two LIGO-type detectors** in motion [Cutler 1997]
High-f: more complicated



LISA frame

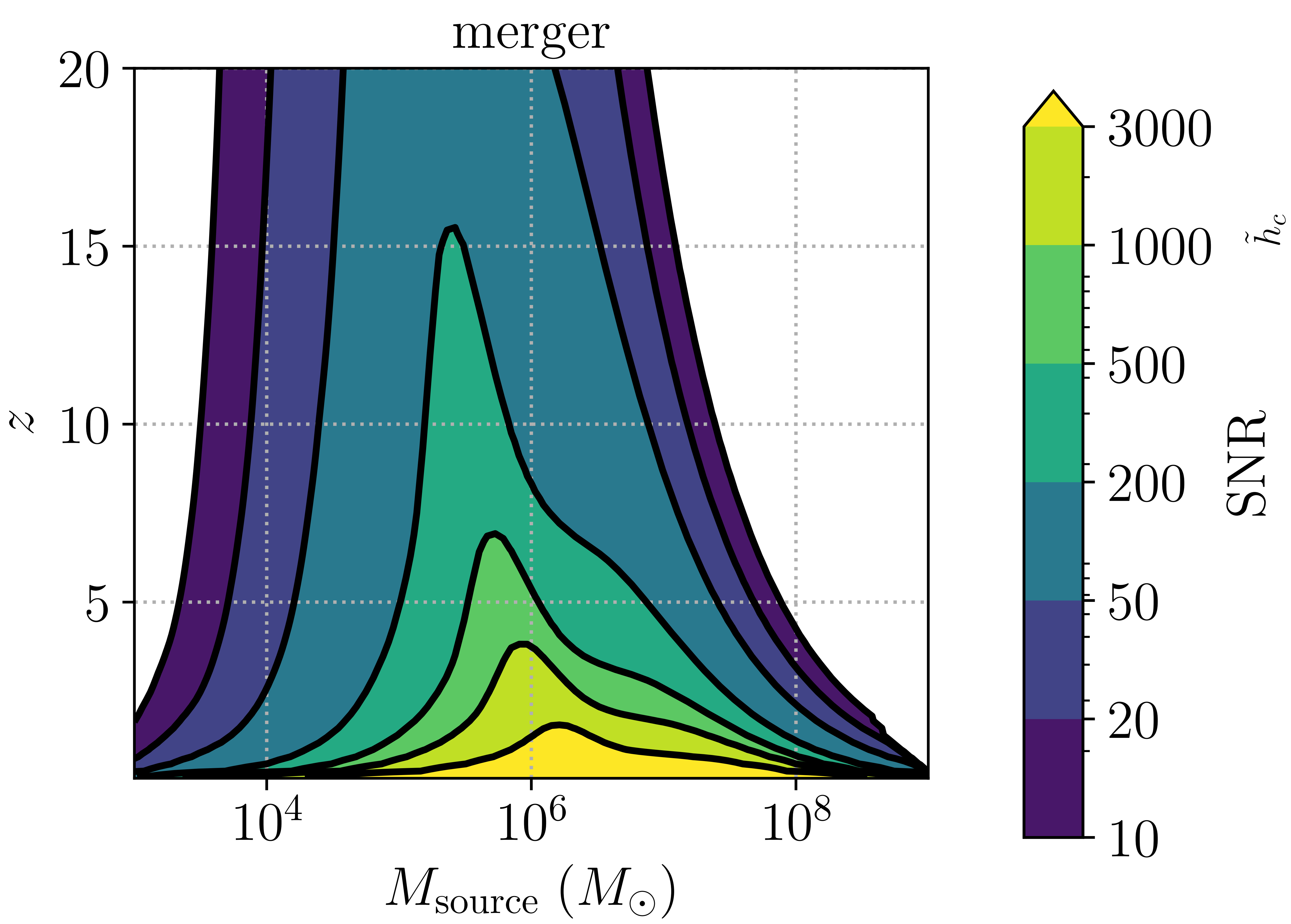
Short-lived signals: use frame based on LISA plane at merger





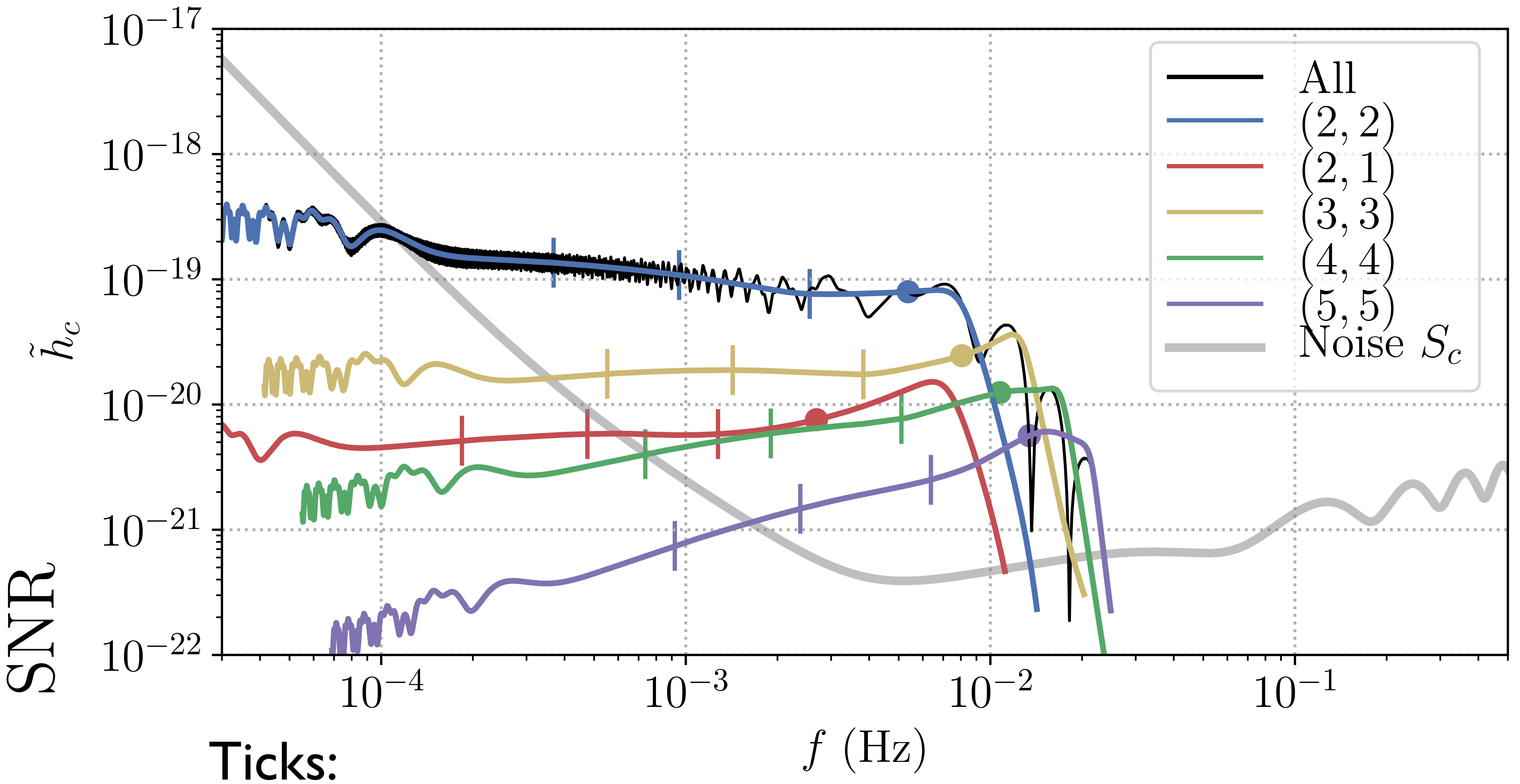
Massive black hole binaries

MBHB SNR contours post-merger



extremely loud signals…

Example MBHB GW signal with higher harmonics



Ticks:

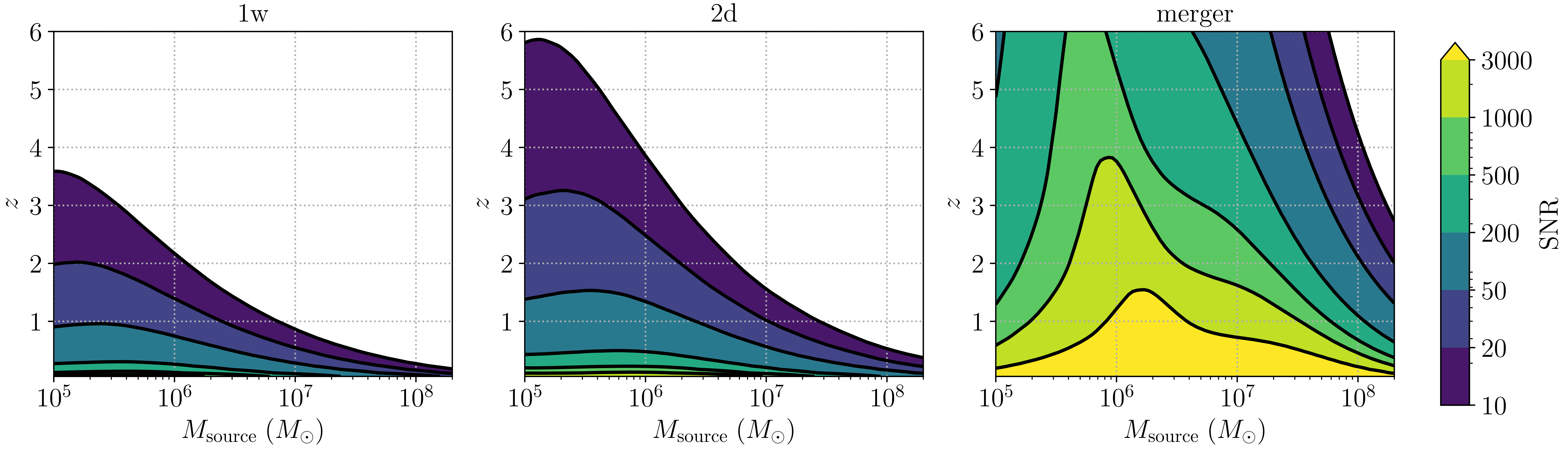
- SNR/64 (40h)
- SNR/16 (2.5h)
- SNR/4 (7min)
- merger

Higher harmonics strong at merger (break degeneracies)

Data analysis simulations still missing precession, eccentricity

MBHB signals are merger-dominated in SNR

MBHB SNR contours pre-merger and post-merger



Most of the SNR accumulates in the last hours before coalescence

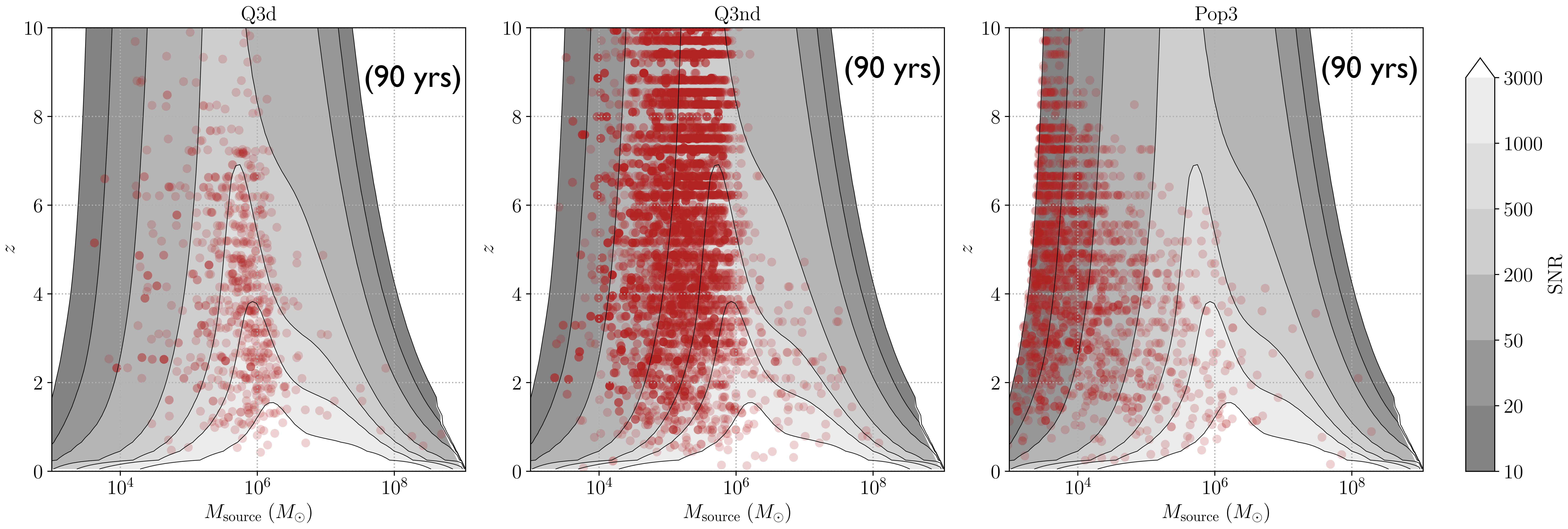
MBHB catalogs

Astrophysical models [Barausse 2012]:

- Heavy seeds - delay (Q3d)
- Heavy seeds - no delay (Q3nd)
- PopIII seeds - delay (Pop3)

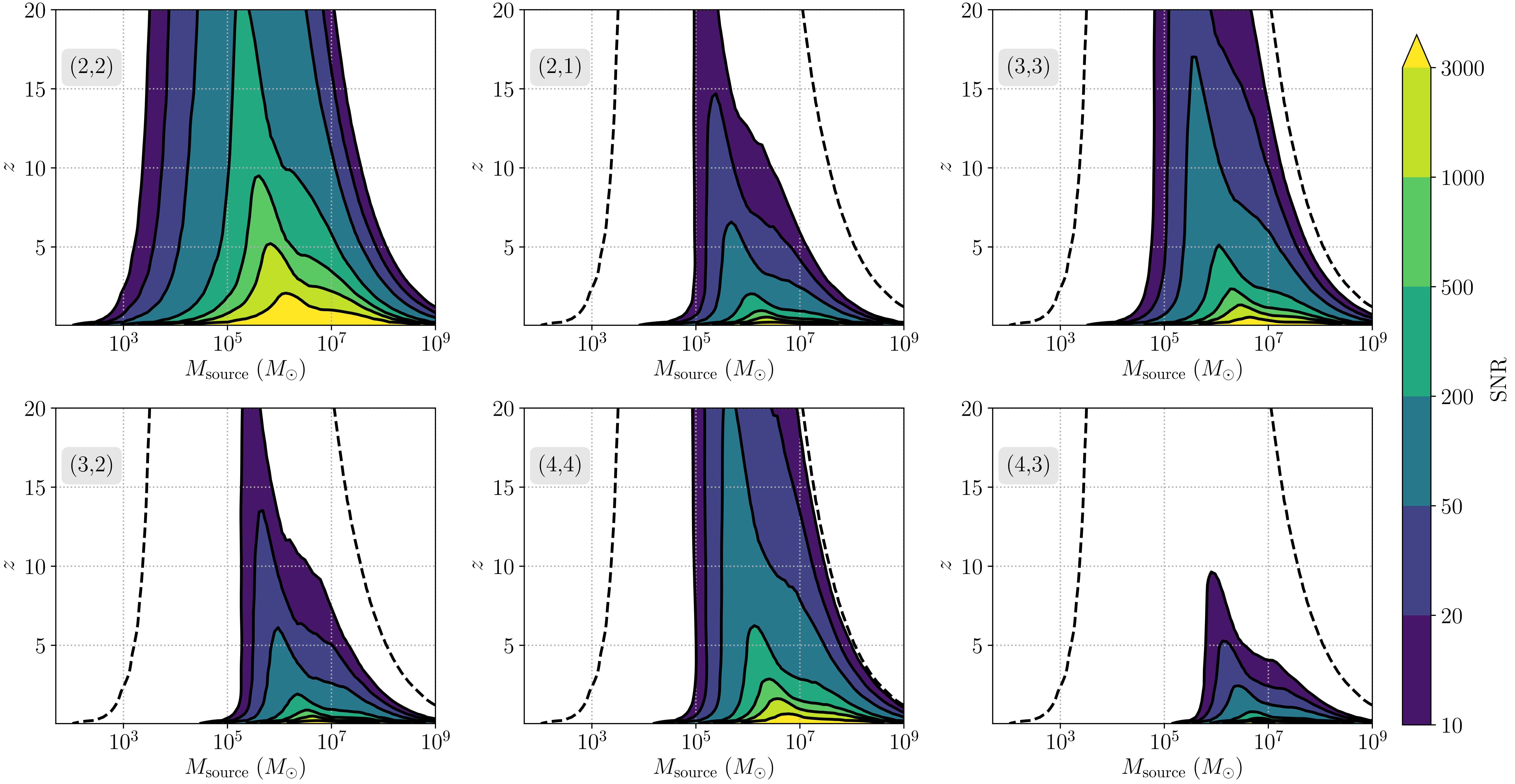
LISA detection rates from 90 yrs simulated:

- Q3d: 30 / 4yrs
- Q3nd: 471 / 4 yrs
- Pop3: 129 / 4yrs



MBHBs: SNR of higher harmonics

PhenomHM, q=5, averaging over spins and orientations



'Mode SNR':

$$\sqrt{(s_{\ell m}|s_{\ell m})}$$

Cross-terms not negligible

$$\sum_{(\ell m)\neq(\ell' m')} (s_{\ell m}|s_{\ell' m'})$$

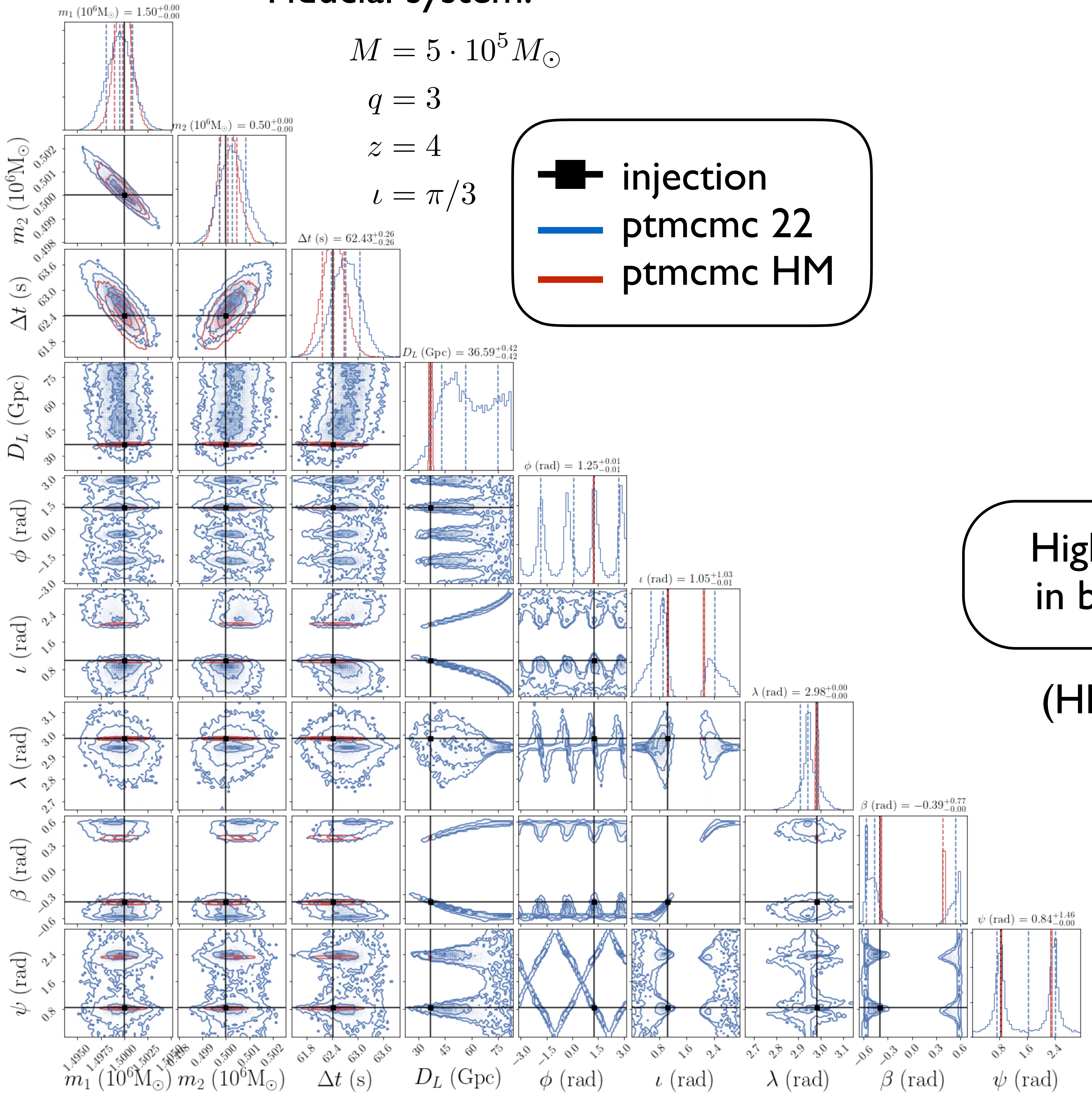
Outline

- MBHB signals in LISA
- **Parameter space degeneracies**
- Tools for Bayesian parameter estimation
- LISA Data Challenge: Sangria
- MBHB results for Sangria

MBHBs: importance of higher modes in parameter estimation

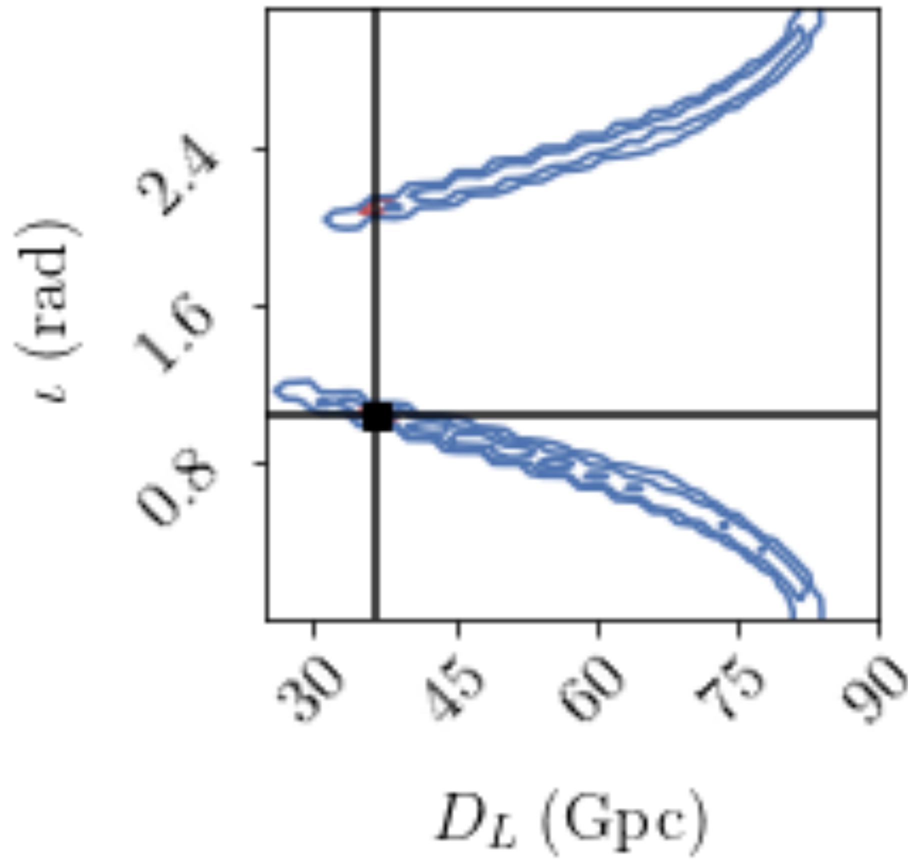
Fiducial system:

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



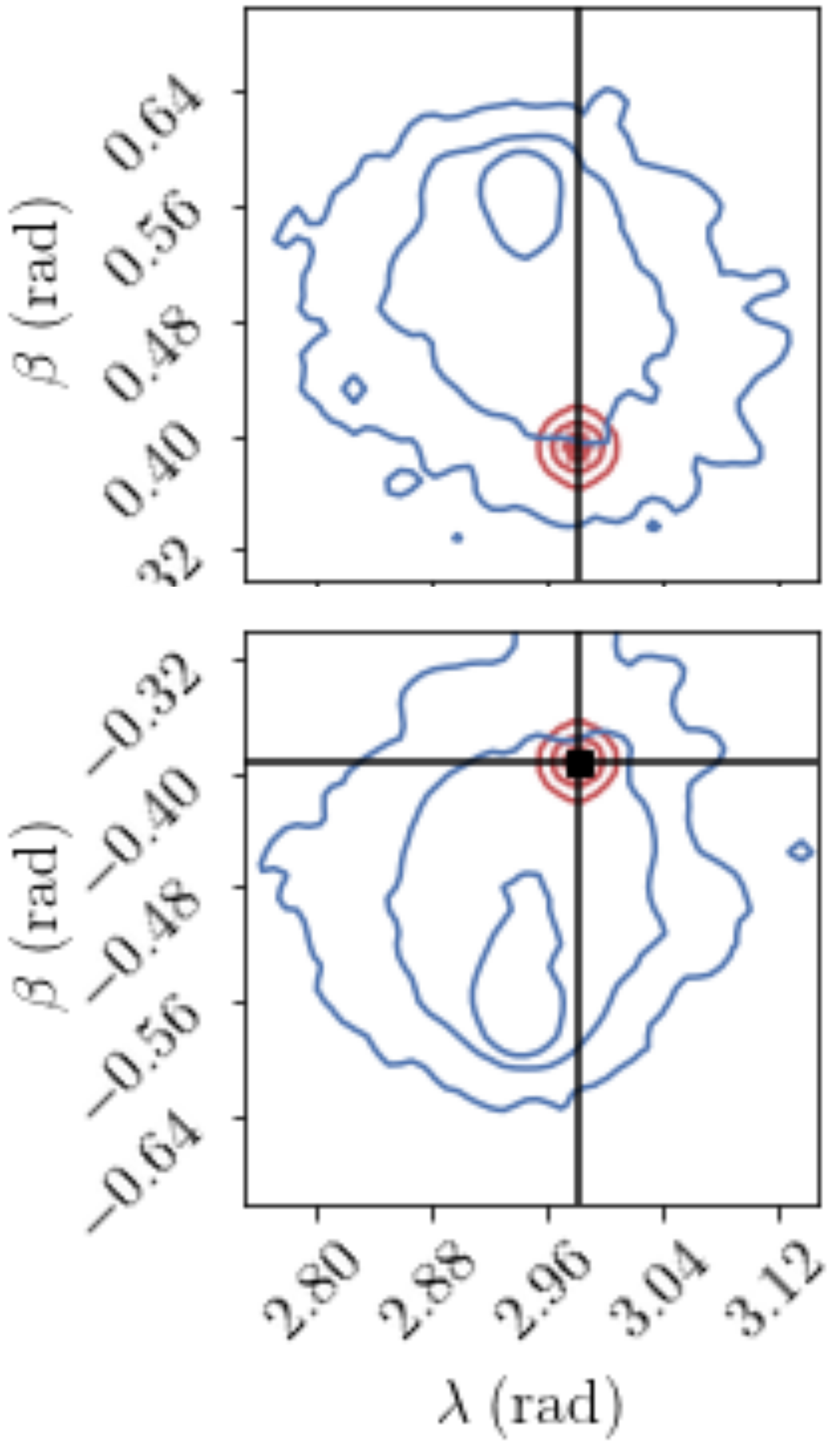
- injection
- ptmcmc 22
- ptmcmc HM

Distance-inclination



[Marsat&al 2020]

Sky position

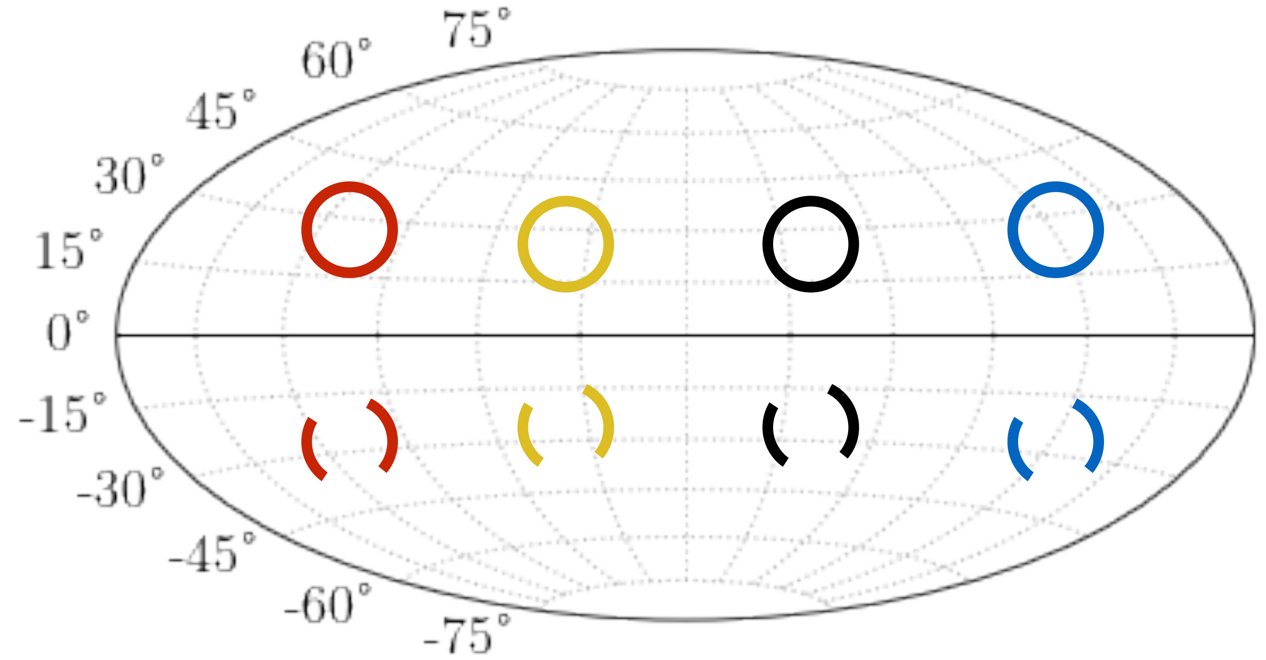


Higher harmonics crucial in breaking degeneracies

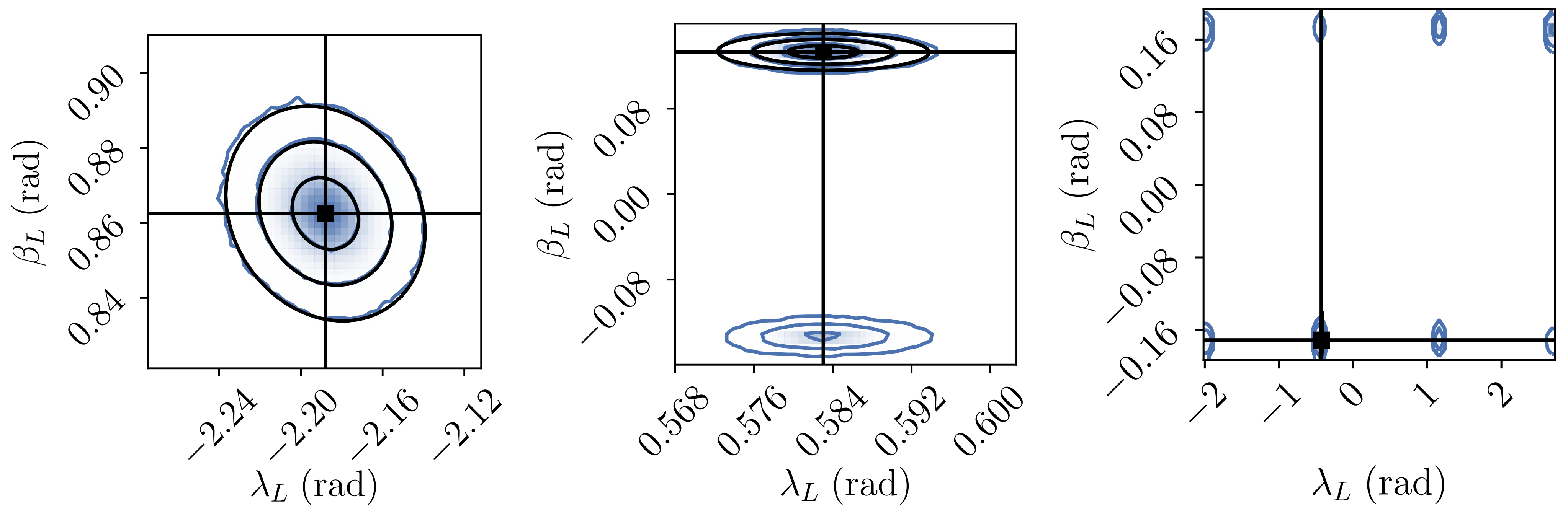
(HM weaker premerger and at low-mass)

Degenerate sky localization possible even with HM

MBHB catalogs: sky multimodality

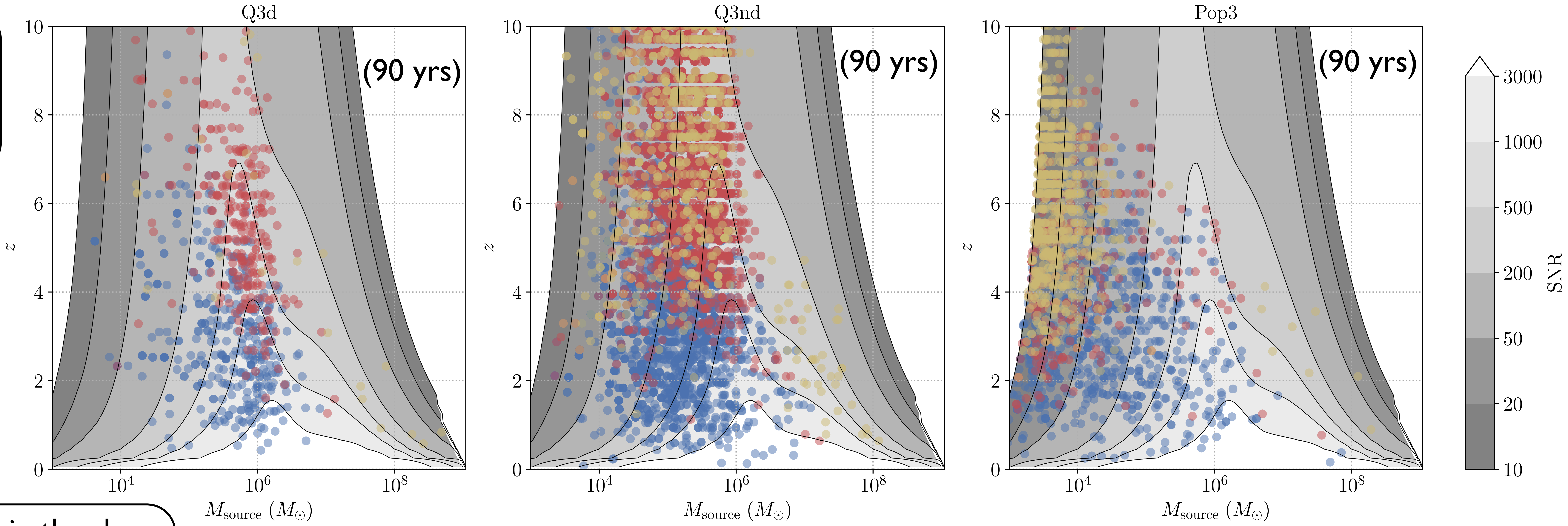


- Bayesian PE required to explore multimodal posteriors
- Simulation of 90yrs catalogs
- Custom proposals for degeneracies



- 1 mode
- 2 modes
- >2 modes

Threshold: 5% probability in the sky mode



Multimodality in the sky present, but rare for counterpart candidates post-merger

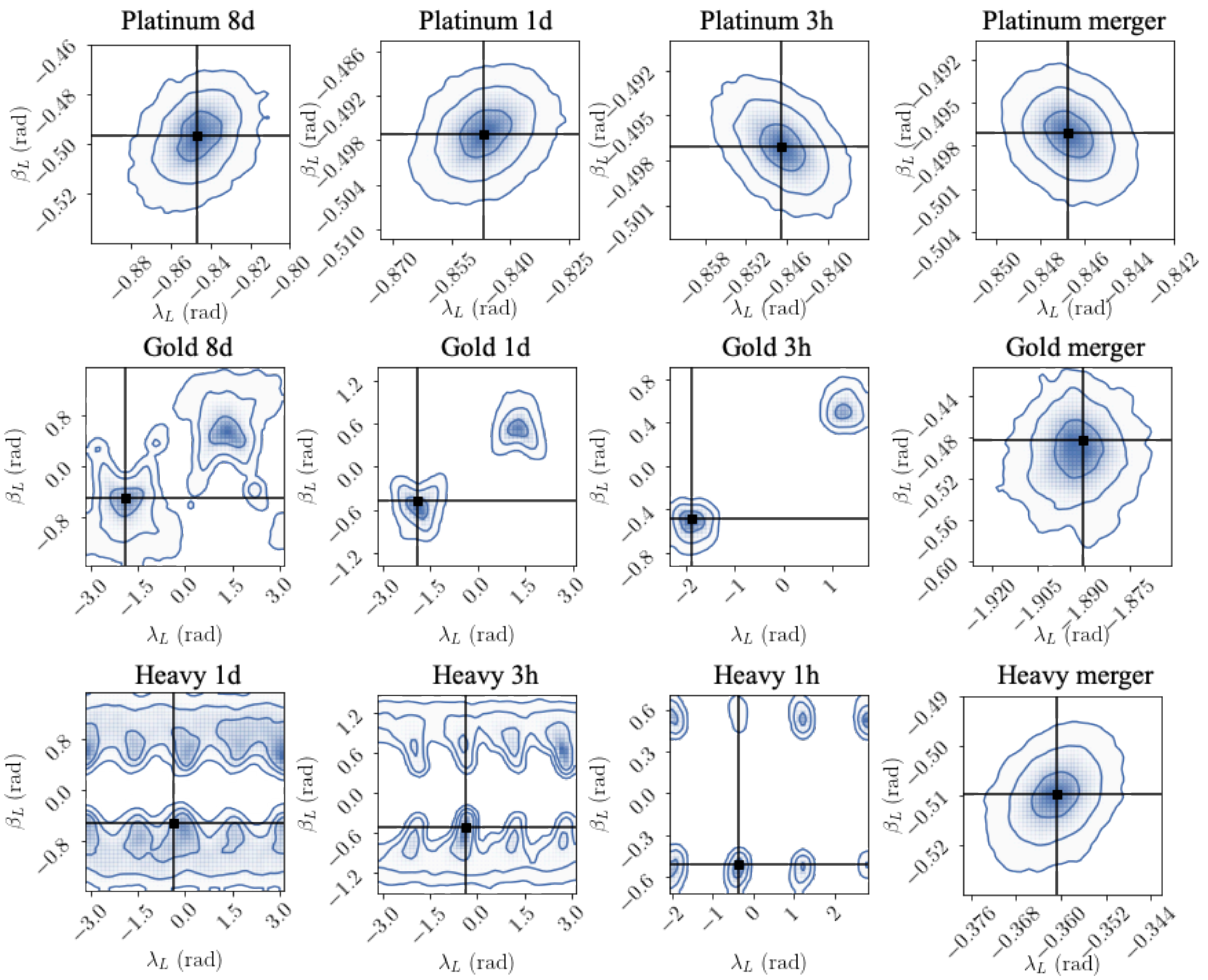
[see also Mangiagli&al 2022, A. Mangiagli's talk]

Localization of ‘golden’ MBHB sources: degeneracies

Bayesian sky localization
cutting at different times

- ‘Gold’: M3e6, z=1
- ‘Heavy’: M1e7, z=1
- ‘Platinum’: M3e5, z=0.3

- Wide range of multimodalities dep. on parameters
- Post-merger localization unimodal for ‘golden’ MBHBs



[Piro&al, in prep]

Outline

- MBHB signals in LISA
- Parameter space degeneracies
- **Tools for Bayesian parameter estimation**
- LISA Data Challenge: Sangria
- MBHB results for Sangria

Parameter estimation tool: lisabeta

Bayesian analysis

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

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

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

Data: signal+noise $d = s + n$

Producing samples from the posterior takes millions of evaluations of lnL !

Approximation levels

- Fisher matrix: local Gaussian approx. for lnL for high may work in high SNR limit, but misses degeneracies
- Simplified PE: MCMC initialized from Fisher, 0-noise
- Full simulation with unknown signal, noise (LDC)
- Superposition of sources, unknown noise, noise artifacts…

lisabeta package

- Science prospective
- Prototyping real analysis (LDC)
- Source types: MBHBs, SBHBs for now — GBs soon
- Consortium-available (full members, public soon)

https://gitlab.in2p3.fr/marsat/lisabeta_release

Features

- MBHB waveforms: PhenomD, PhenomHM, aligned spins with HM
- Fast Fourier-domain response
- SNR computations
- Fisher matrices
- MCMC: ensemble sampler with parallel tempering (ptemcee, [Vousden&al 2015])
- Informed proposals to deal with sky degeneracies
- Accelerated likelihoods (few ms)

Accelerating likelihoods: heterodyning

Overview

- Structure of the likelihood

$$\ln \mathcal{L} = -\frac{1}{2}(h-d|h-d) \qquad (a|b) = 4\mathrm{Re}\int df\, \frac{\tilde{a}(f)\tilde{b}^*(f)}{S_n(f)}$$

$h = Ae^{i\Phi}$ smooth amp/phase $\quad d$ numerical data

- Introduce a reference waveform $\quad \overline{h}(f)$

$\zeta(f) \equiv h(f)/\overline{h}(f)$ now **slowly variable**

in the vicinity of reference parameters

- Separate integrand in slowly and rapidly variable parts

$$(h|d) \sim \int df\, \frac{\overline{h}d^*}{S_n} \times \zeta \qquad (h|h) \sim \int df\, \frac{\overline{h}\overline{h}*}{S_n} \times \zeta\zeta^*$$

- Interpolate and precompute

$\zeta \quad$ interpolated on a coarse, reduced grid

$$(h|d) \sim \sum_i \int_{f_i}^{f_{i+1}} df\, \frac{\overline{h}d^*}{S_n} \times (a_i + b_i f)$$

$$(h|h) \sim \sum_i \int_{f_i}^{f_{i+1}} df\, \frac{\overline{h}\overline{h}*}{S_n} \times (a_i + b_i f + c_i f^2)$$

- Evaluate

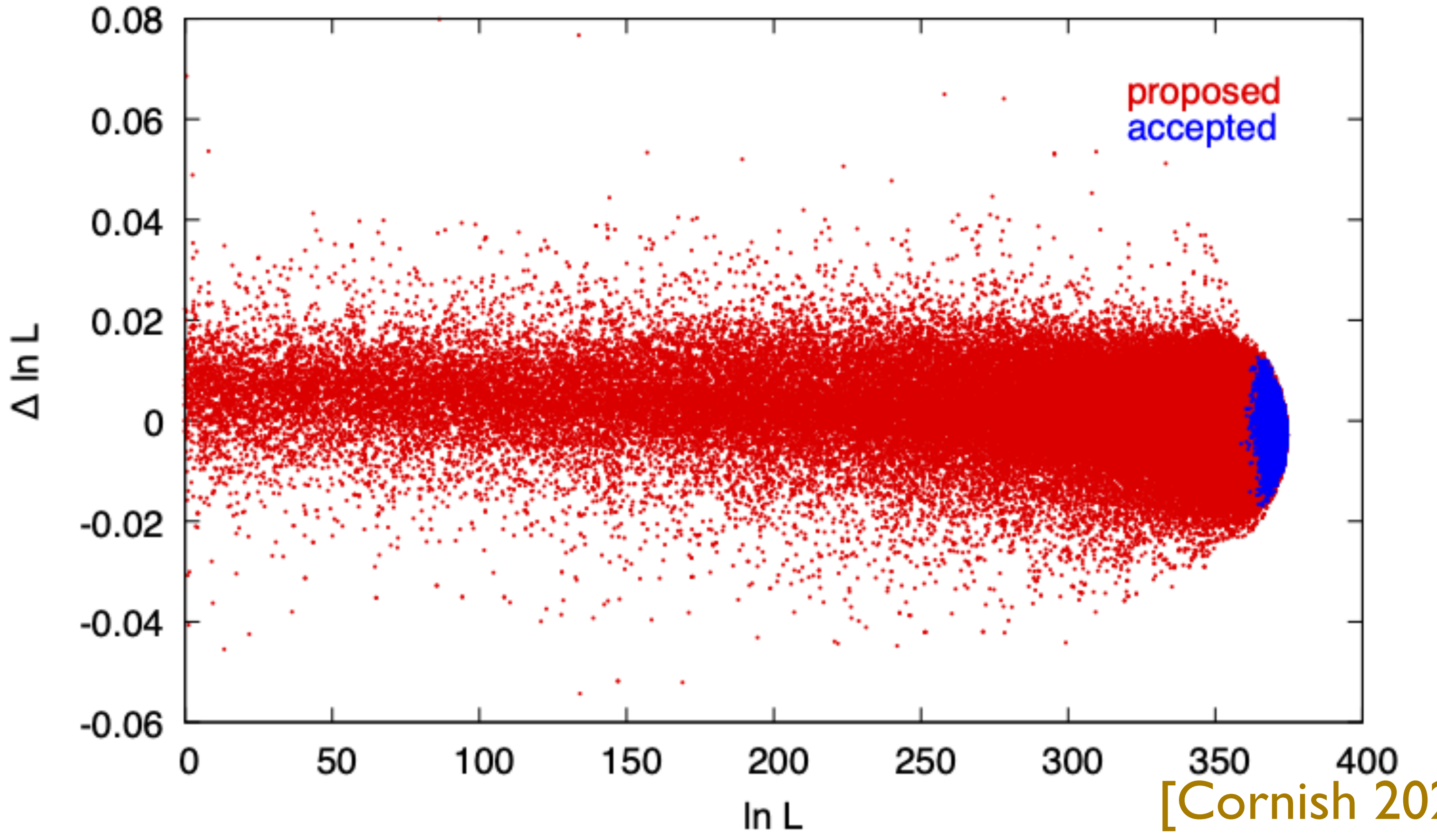
$h \quad$ on coarse grid, then sum weights and coeffs

[Cornish 2010, Cornish 2021]

[Zackay+ 2018] (relative binning)

Usage in practice

- Small reduced grid (N~100): cost <~ ms
- Different interpolation methods (linear, polynomial)
- Requires reference waveform (first guess for signal parameters) — can be updated on the way
- Distinguish burn-in from actual sampling, the latter happens close to the true signal



[Cornish 2021]

Accelerating likelihoods: heterodyning example for MBHB

Decomposing the likelihood:

$$\ln \mathcal{L} = -\frac{1}{2}(s-d|s-d)$$
$$= -\frac{1}{2}(s-s_0|s-s_0) + (s-s_0|d-s_0) - \frac{1}{2}(s_0-d|s_0-d)$$

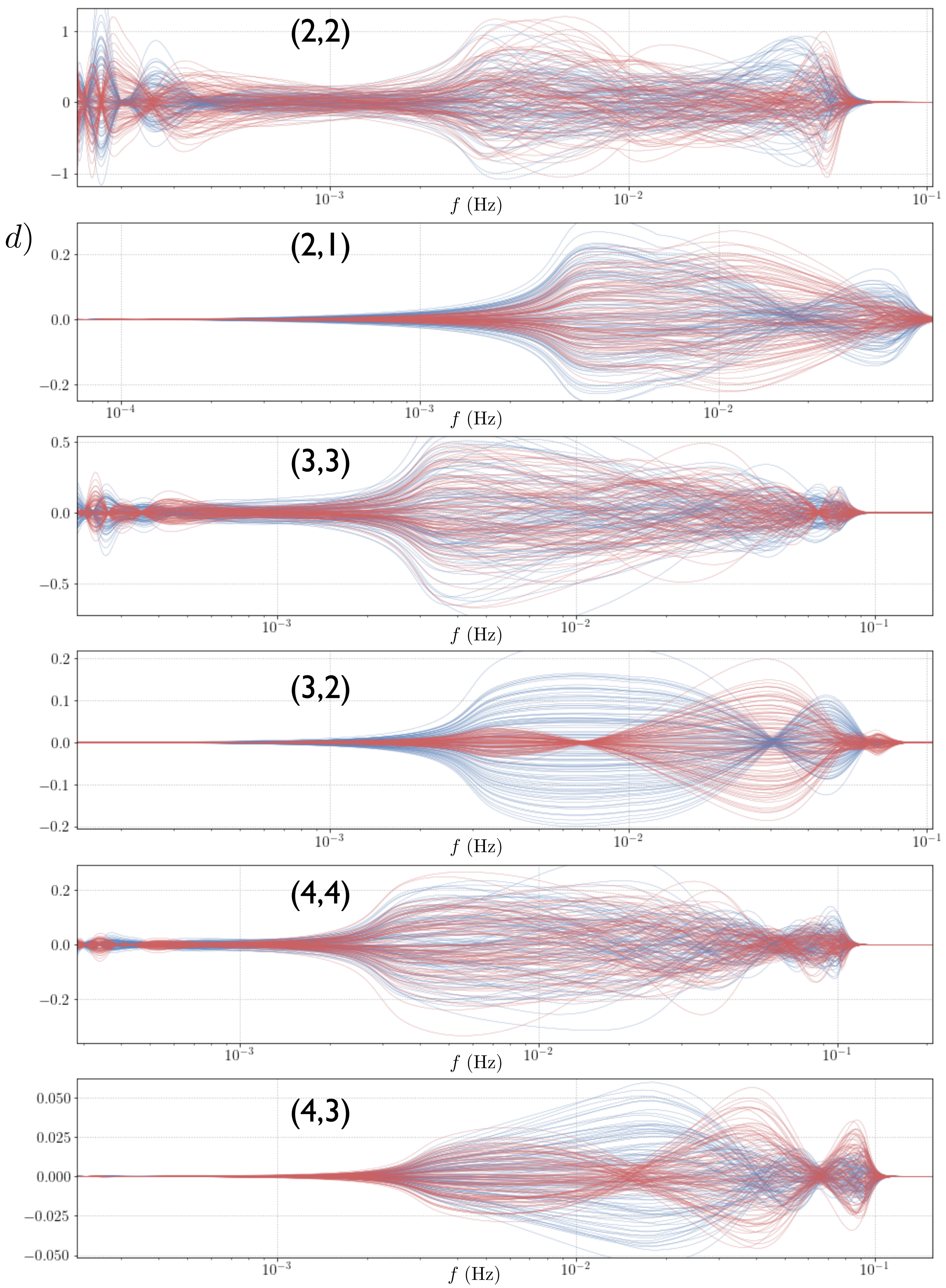
Residuals from reference waveform:

$$s_{\ell m} - s^0_{\ell m} = r_{\ell m} e^{i\Phi^0_{\ell m}}$$

Implementation:

$$(s-s_0|s-s_0) = \sum_{\ell m}\sum_{\ell' m'} (r_{\ell m} r^*_{\ell' m'} | e^{i(\Phi^0_{\ell' m'} - \Phi^0_{\ell m})})$$
$$(s-s_0|d-s_0) = \sum_{\ell m} (r_{\ell m} | e^{-i\Phi^0_{\ell m}}(d-s_0))$$

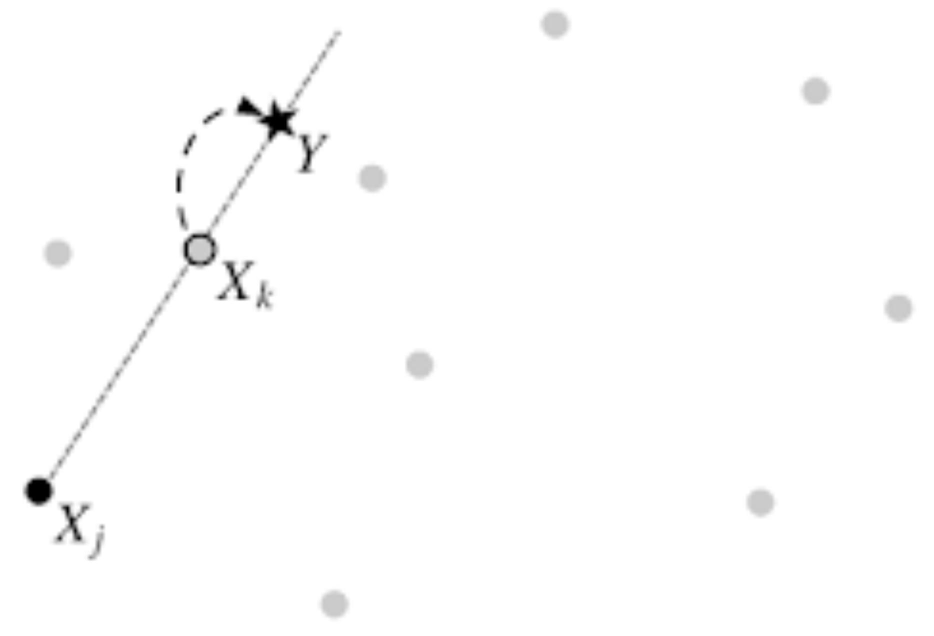
- Fix a sparse frequency grid (~128)
- Linear interpolation of the residuals, mode-by-mode
- Precompute 0-th and 1st polynomial inner products against phase and data terms, with a fine resolution



Dealing with degeneracies

Ensemble sampling

- Evolve a population of walker in parallel
- Self-tuning proposal based on the other walkers



Parallel tempering

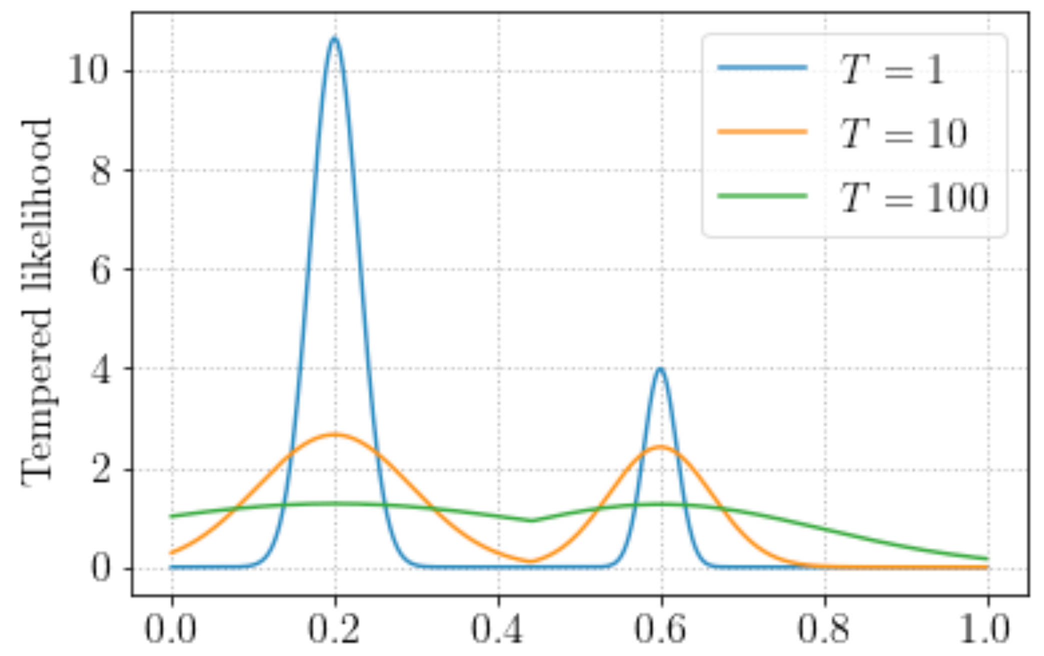
- Introduce parallel chains with temperatures, posterior: $p(\theta)^{\beta_i}$

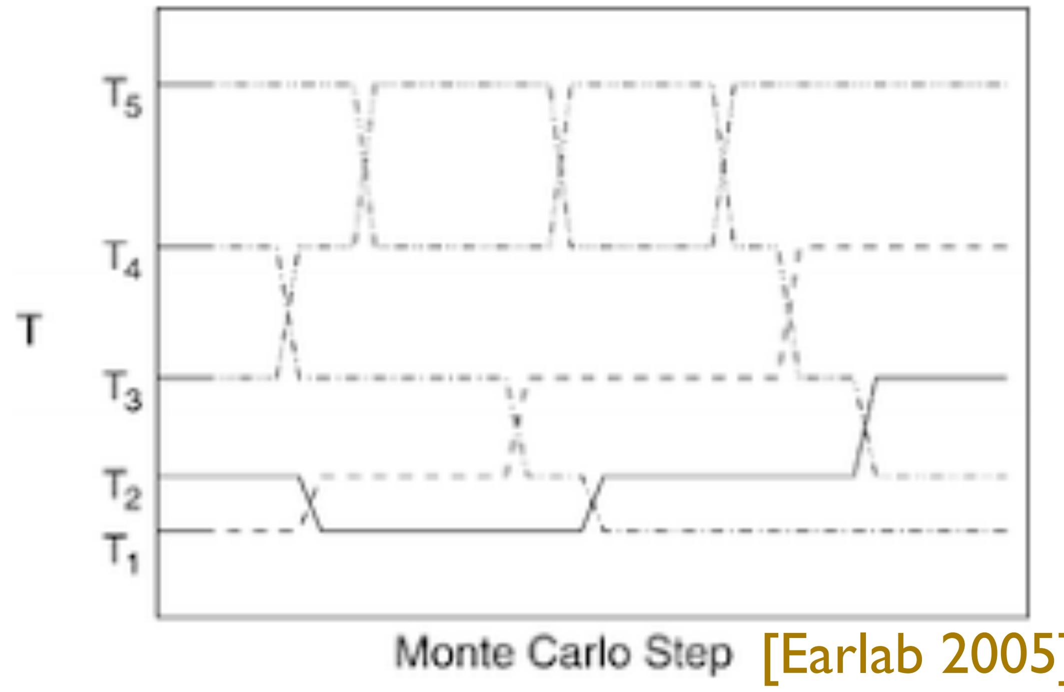
$$\beta_i = 1/T_i$$

- Propose swaps with acceptance:

$$p_{\rm swap} = \min \left[1, \left(\frac{p(\theta_i)}{p(\theta_j)} \right)^{\beta_j - \beta_i} \right]$$

- Crucial for robustness, avoids being stuck in a local maximum



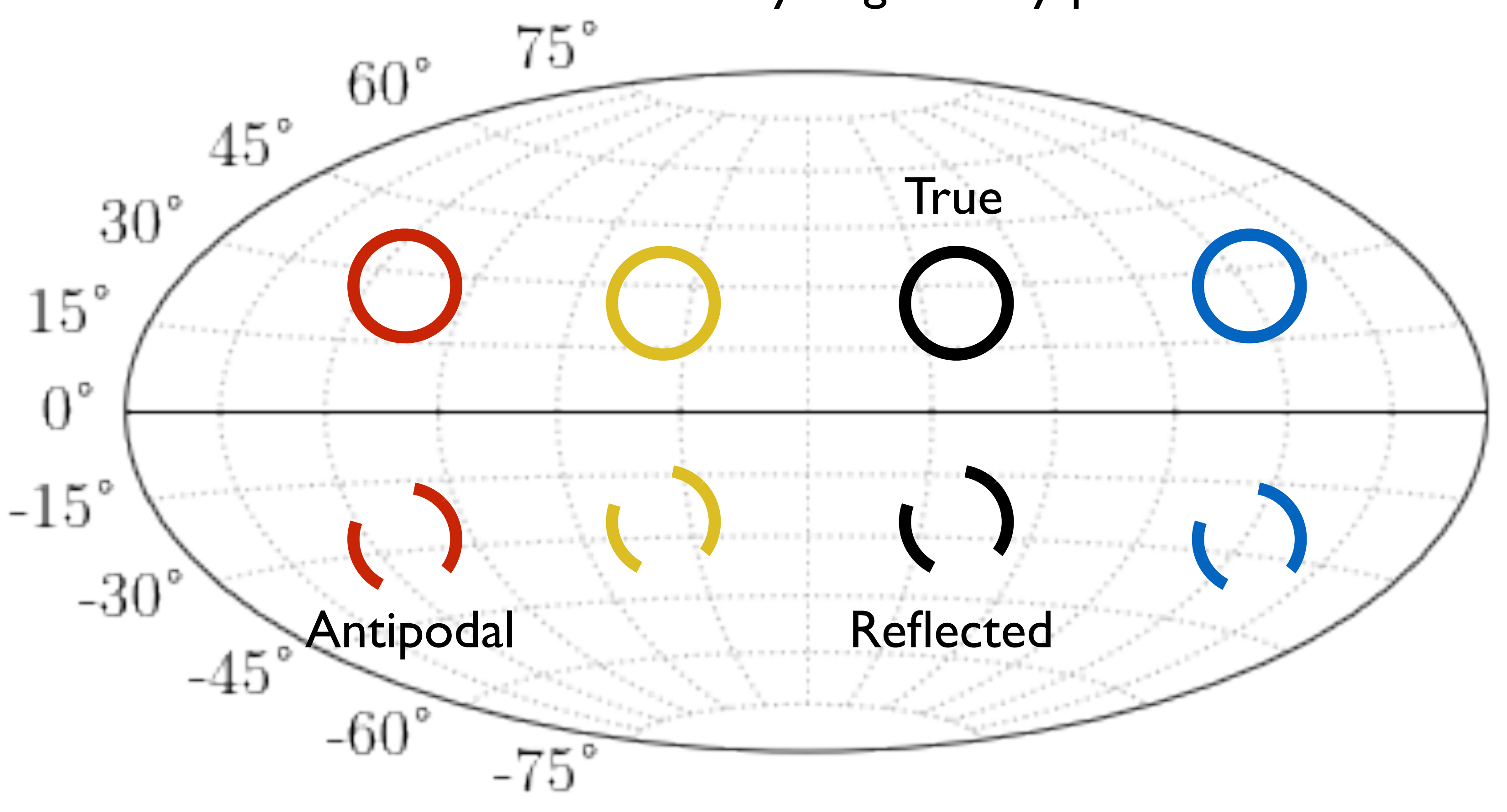


[Earlab 2005]

Tailored jump proposals

- In presence of known degeneracies, include jumps in proposal
- Very efficient for very disconnected multimodal posteriors

LISA MBHB sky degeneracy pattern

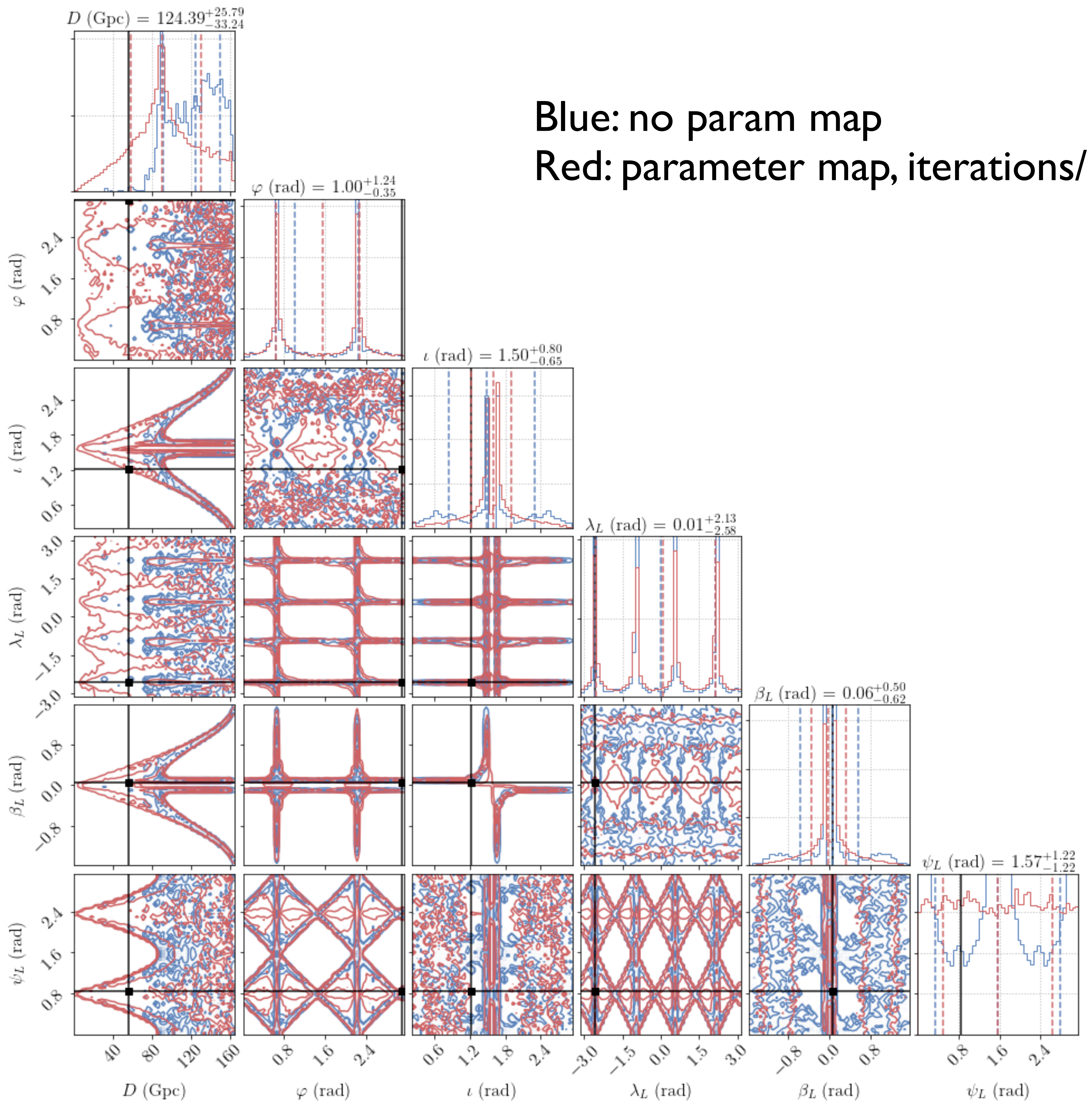


Tailored parameter map

- Transform to new variables (close to observables) in which the posterior is close to a Gaussian
- Easy to implement if transformation and Jacobian analytic

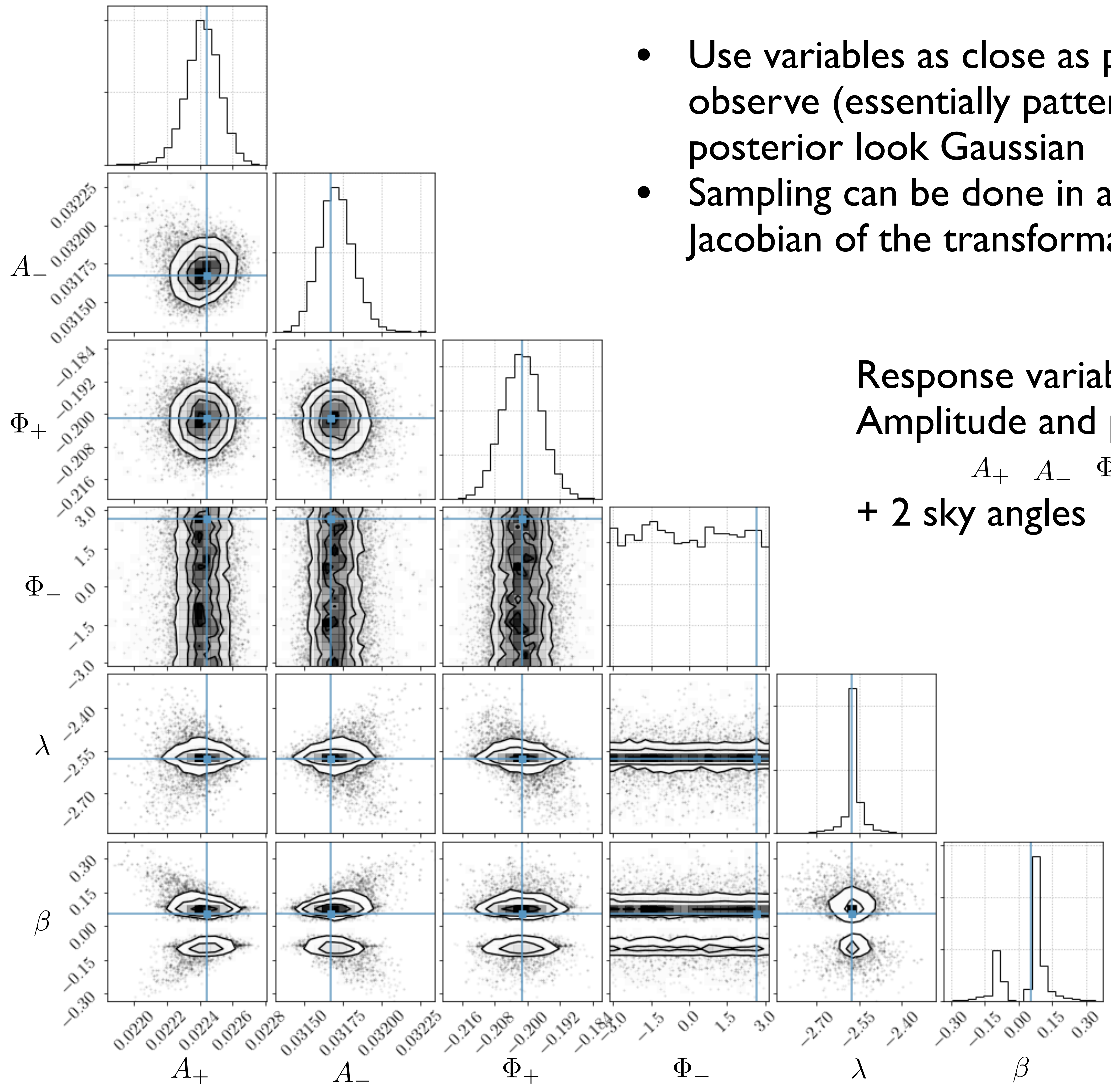
Dealing with degeneracies: maximally degenerate case

Toy problem, completely degenerate extrinsic 22 likelihood without motion and high-f effects



Blue: no param map
Red: parameter map, iterations/10

Dealing with degeneracies: parameter map



- Use variables as close as possible to what we really observe (essentially pattern functions), to make the posterior look Gaussian
- Sampling can be done in any set of parameters, with Jacobian of the transformation analytic here

Response variables: 2 complex pattern functions
Amplitude and phase

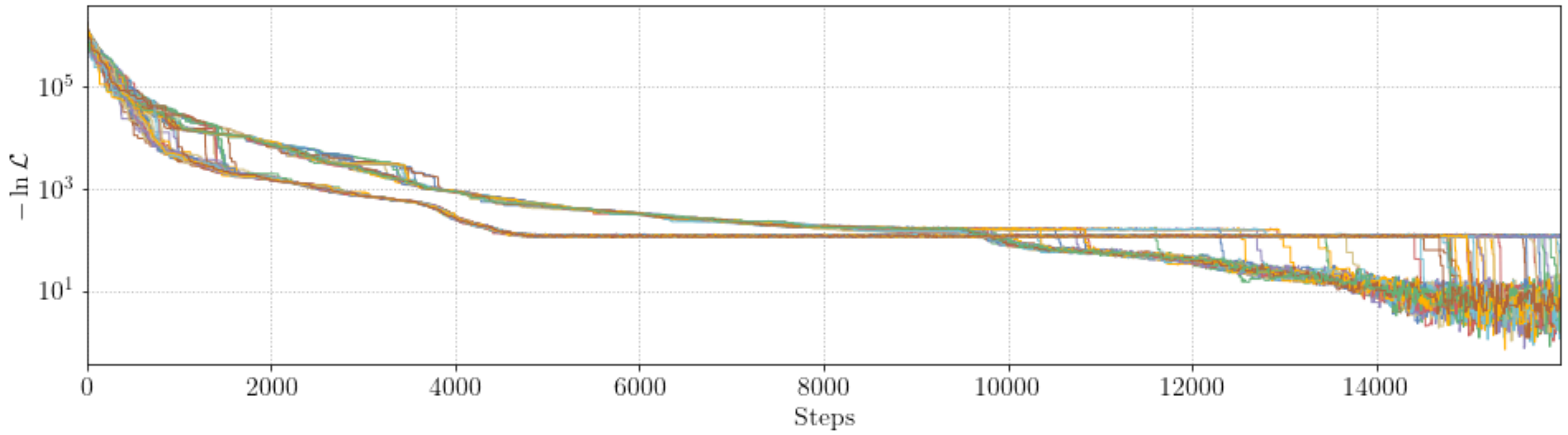
$A_+ \quad A_- \quad \Phi_+ \quad \Phi_-$

+ 2 sky angles $\quad \lambda \quad \beta$

Analytical transformation, no extra cost

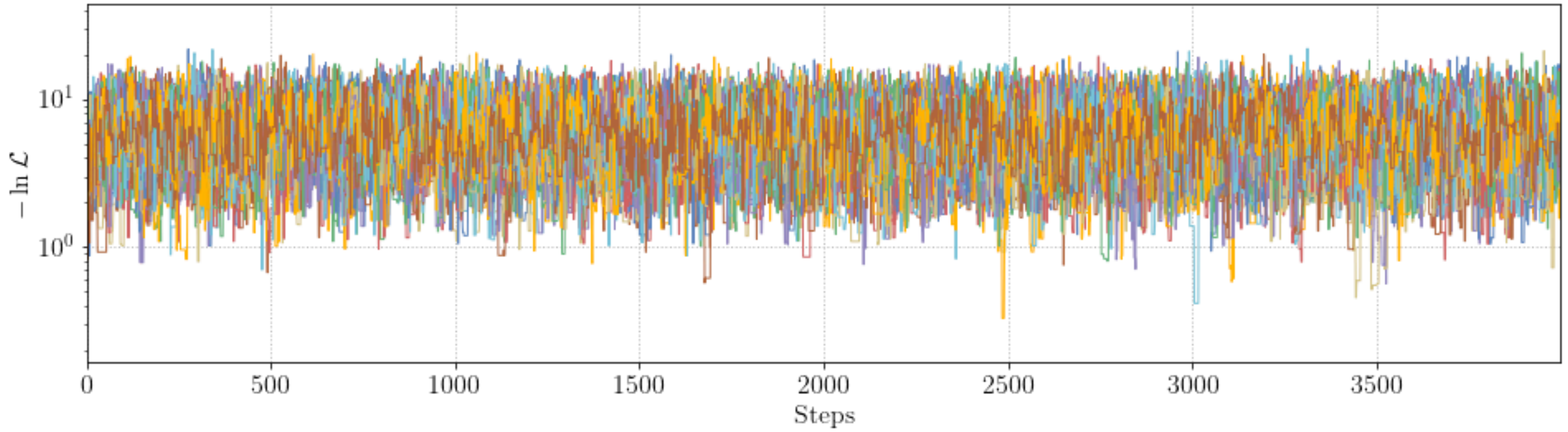
Accelerating PE: burn-in vs sampling

Burn-in (here struggling to find the signal !)



Scale of likelihood with completely wrong signal: $\ln \mathcal{L}_{\rm bad} \sim -{\rm SNR}^2$

Sampling (not moving much in likelihood)



Search and burn-in are different:

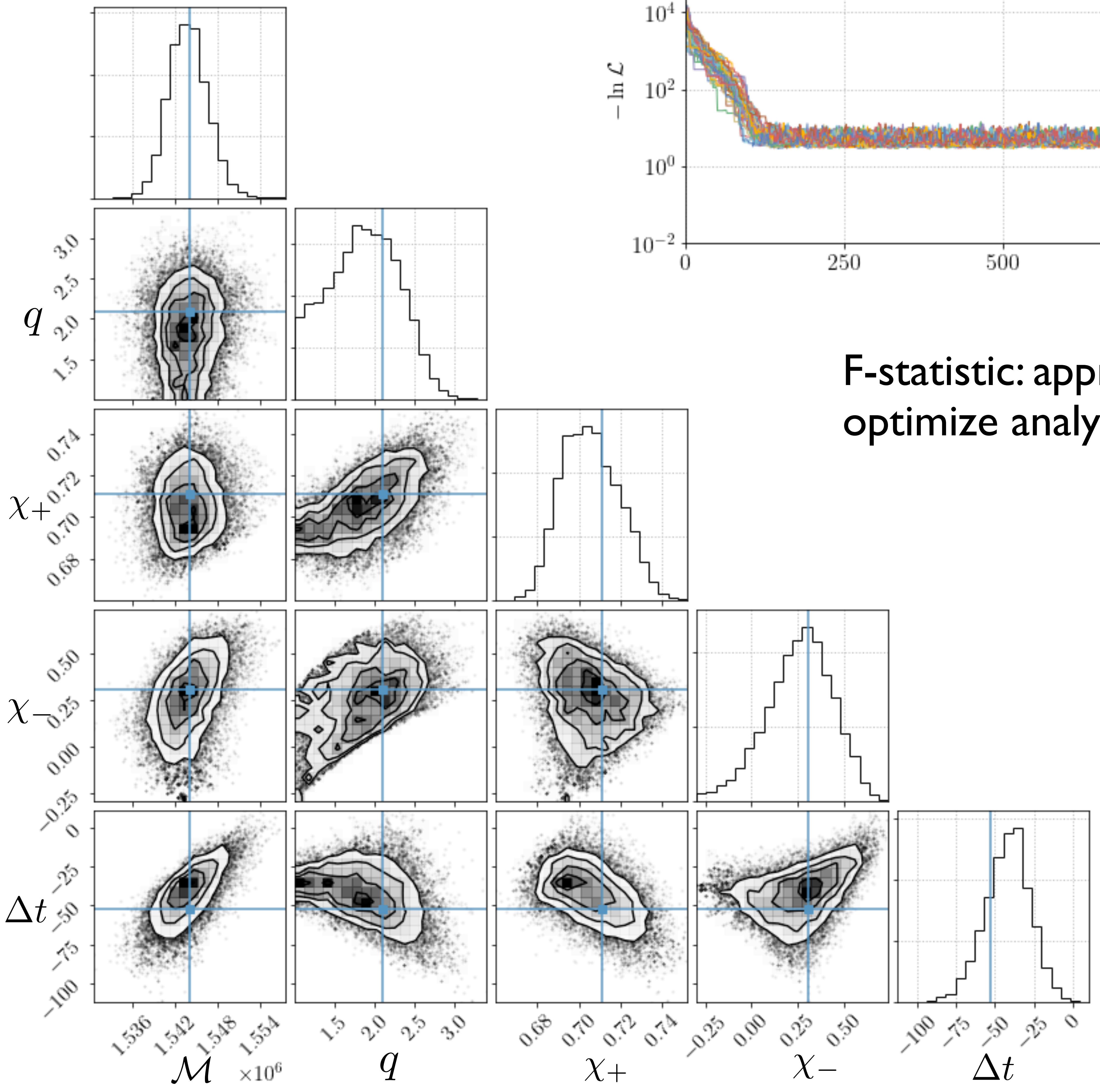
- Sampling algorithm can be inefficient to search for a signal
- First guess of the signal's parameters allows to accelerate likelihoods with heterodyning

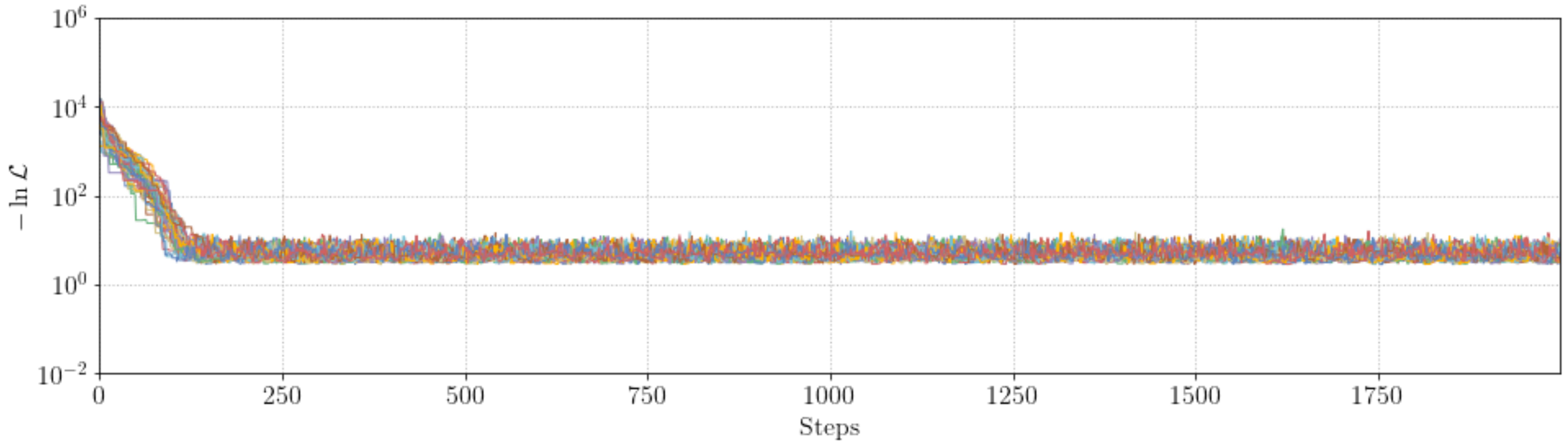
Different use cases for PE:

- Simulating realistic PE: start from prior
- Prospective parameter estimation, only interested in final result: cheat with initialization

Techniques for burn-in (search) or sampling can differ !

MBHB example: I) F-statistic search on small data segments



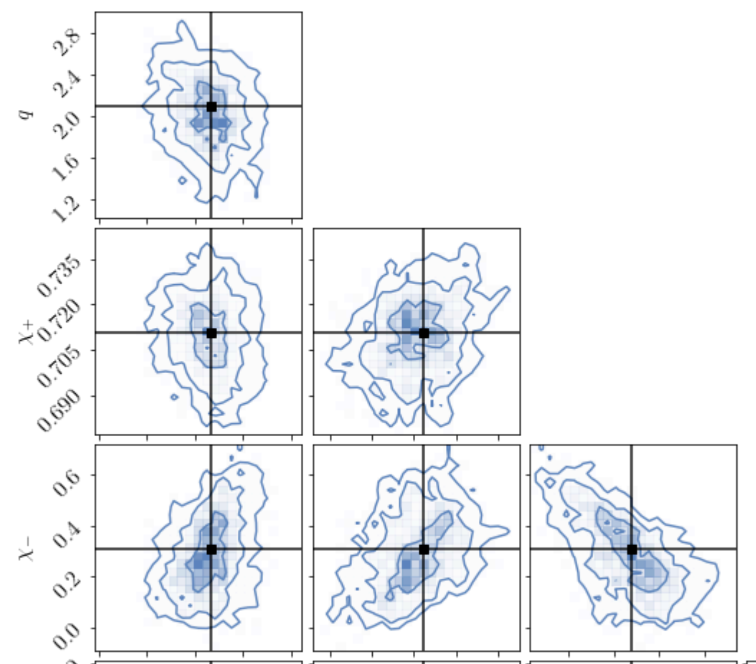


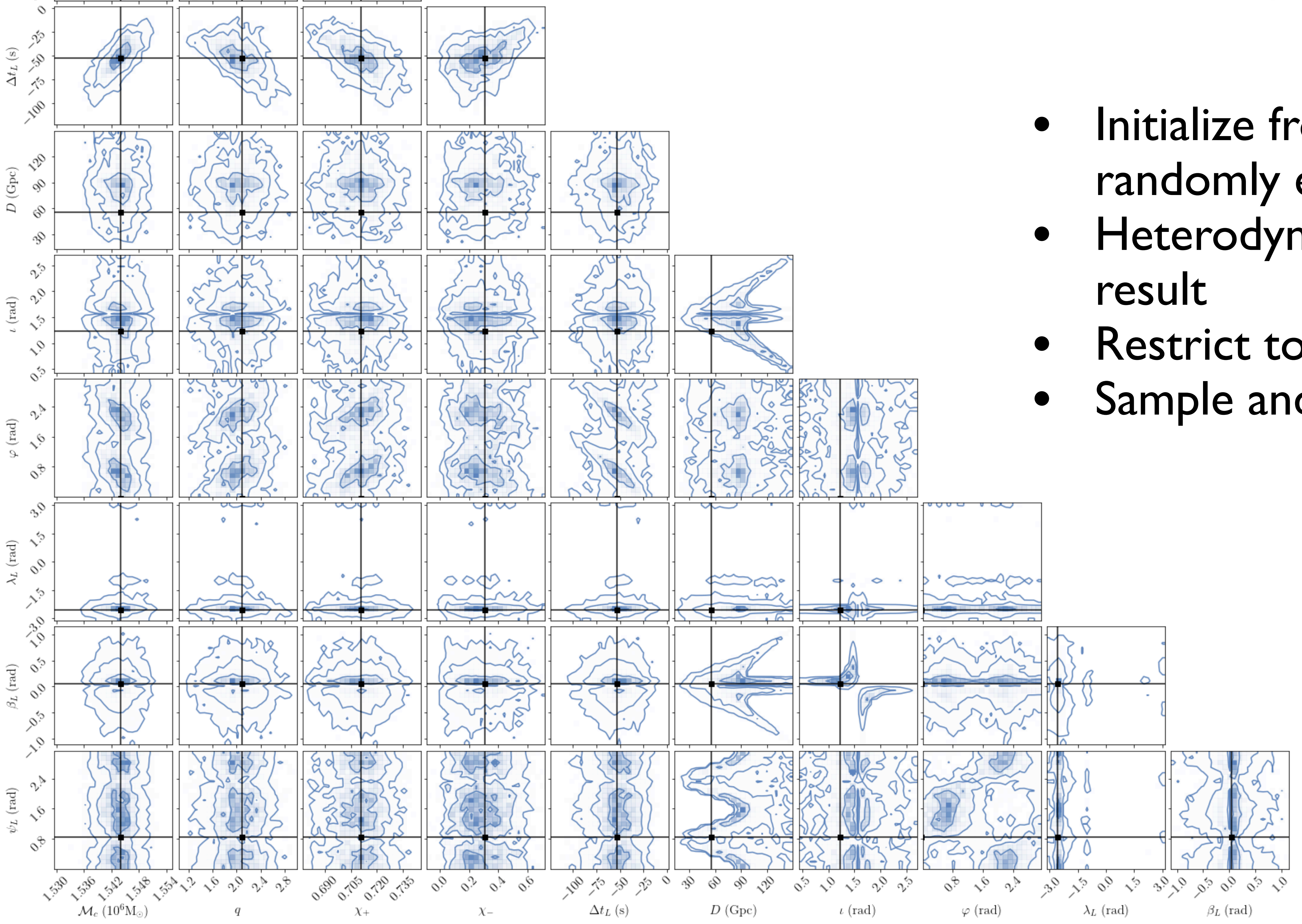
F-statistic: approximate response (low-frequency, no motion, 22 mode), optimize analytically over distance, inclination, phase, polarization, sky

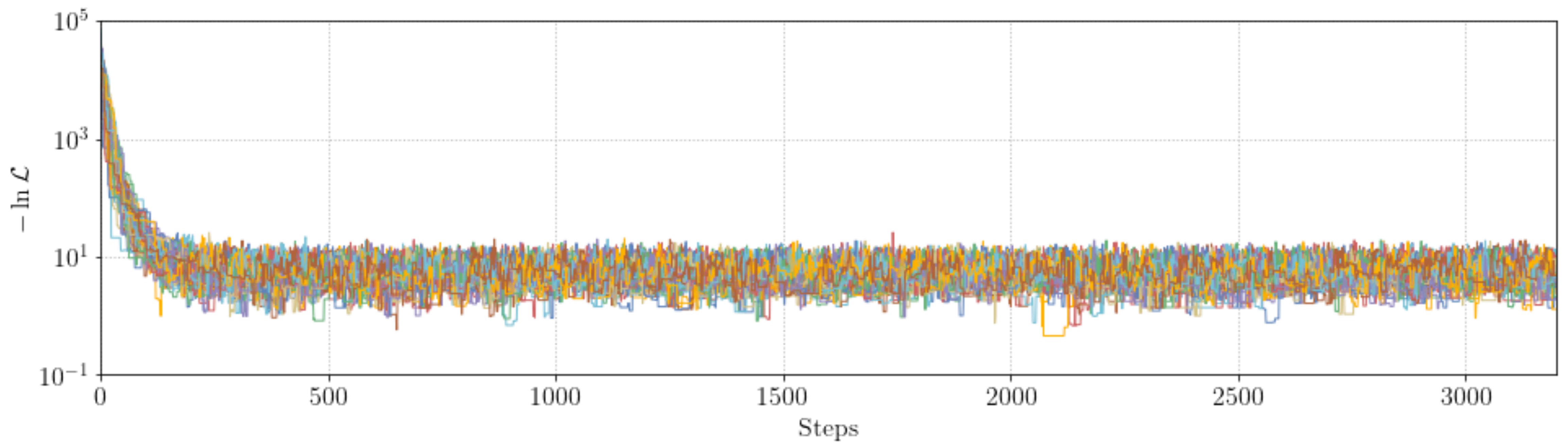
$$\mathcal{F} = \max_\lambda \ln \mathcal{L}(\theta, \lambda)$$

- Select short data segment
- F-statistic as pseudo-likelihood
- Sampling easier for a lower dimensionality
- Get a first guess of intrinsic parameters + time

MBHB example: II) initial PE with low frequencies

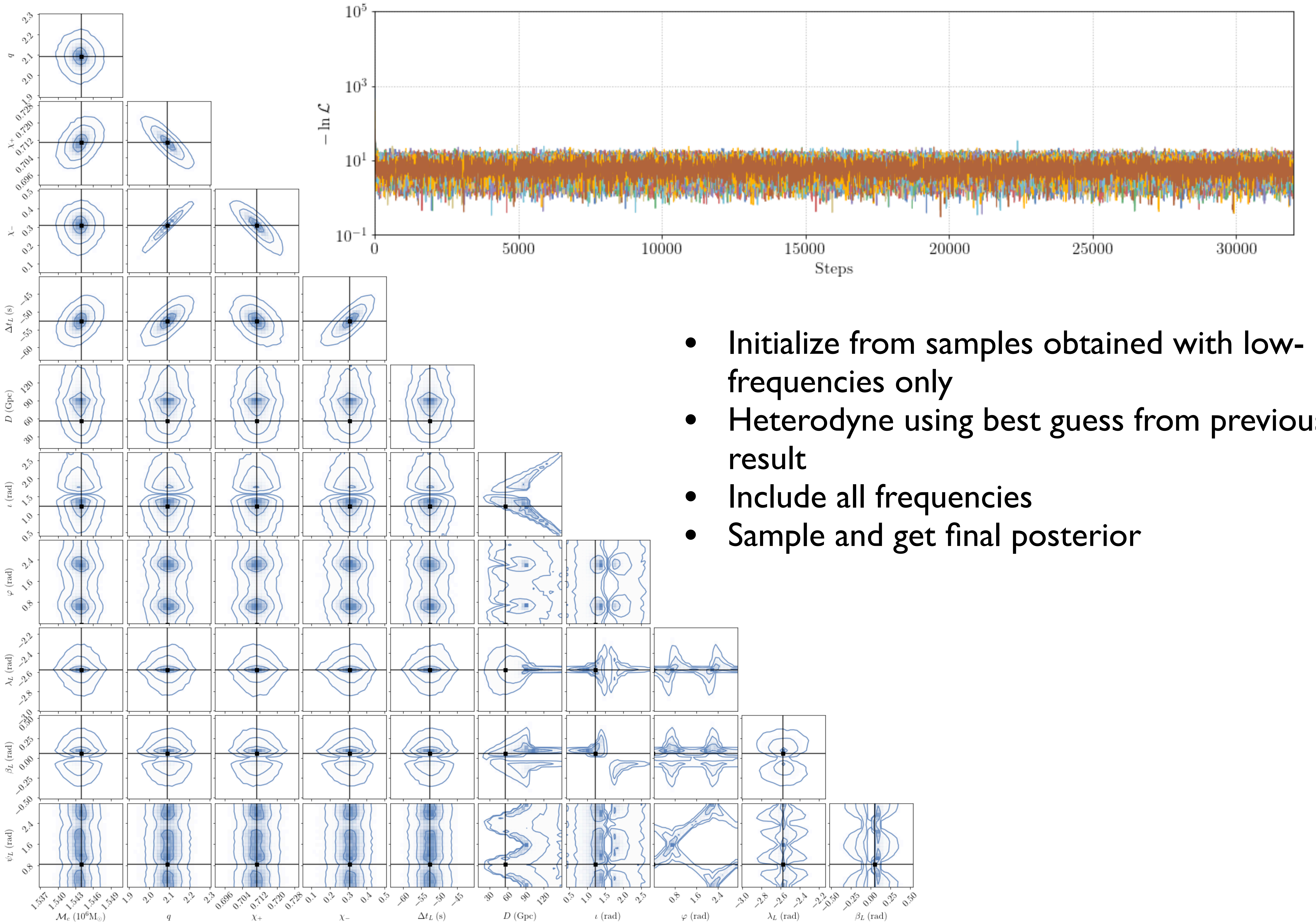






- Initialize from F-statistic samples, draw randomly extrinsic parameters
- Heterodyne using best guess from previous result
- Restrict to low frequencies
- Sample and get a first guess of all params

MBHB example: III) sampling with all frequencies

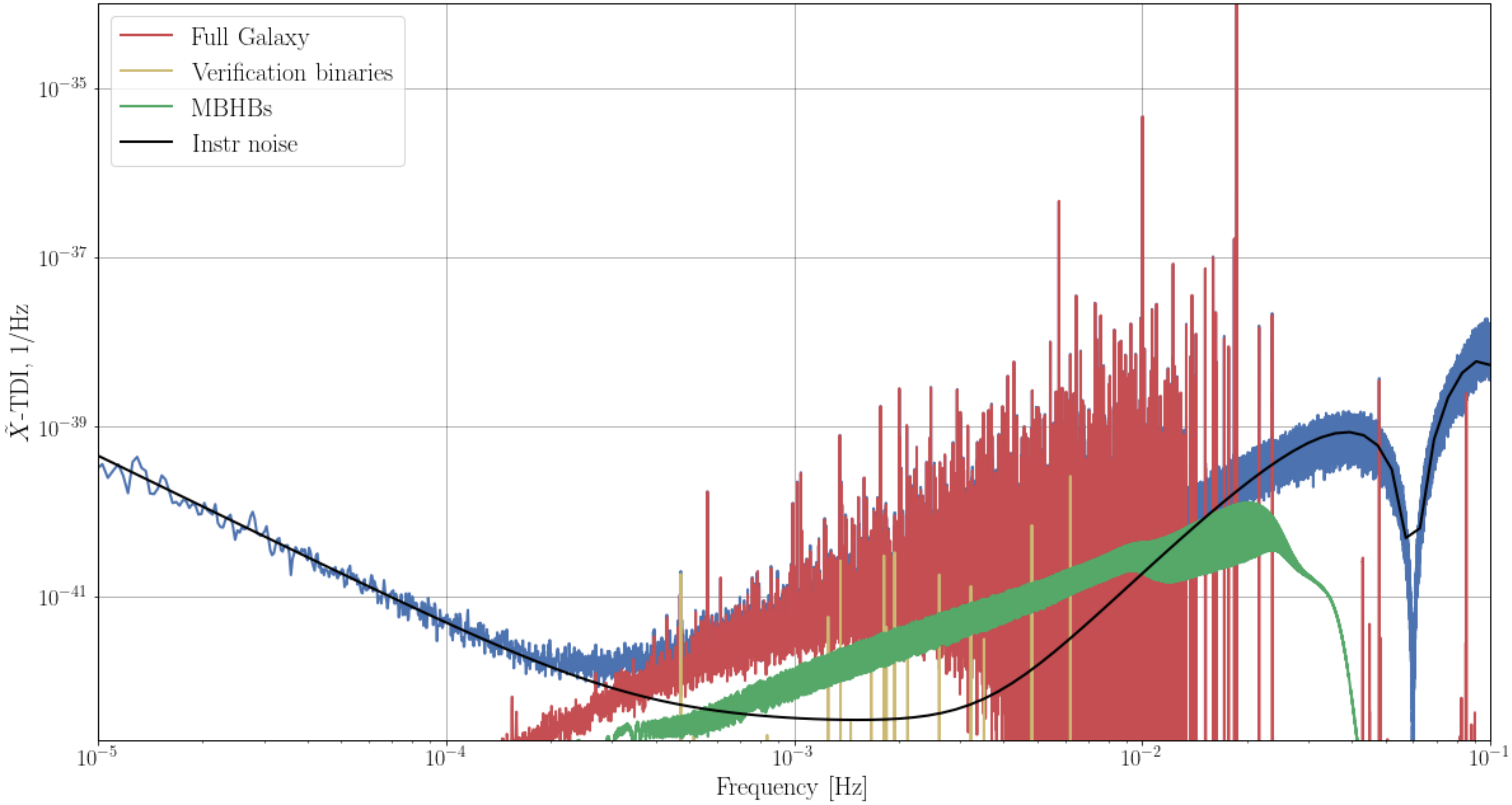


- Initialize from samples obtained with low-frequencies only
- Heterodyne using best guess from previous result
- Include all frequencies
- Sample and get final posterior

Outline

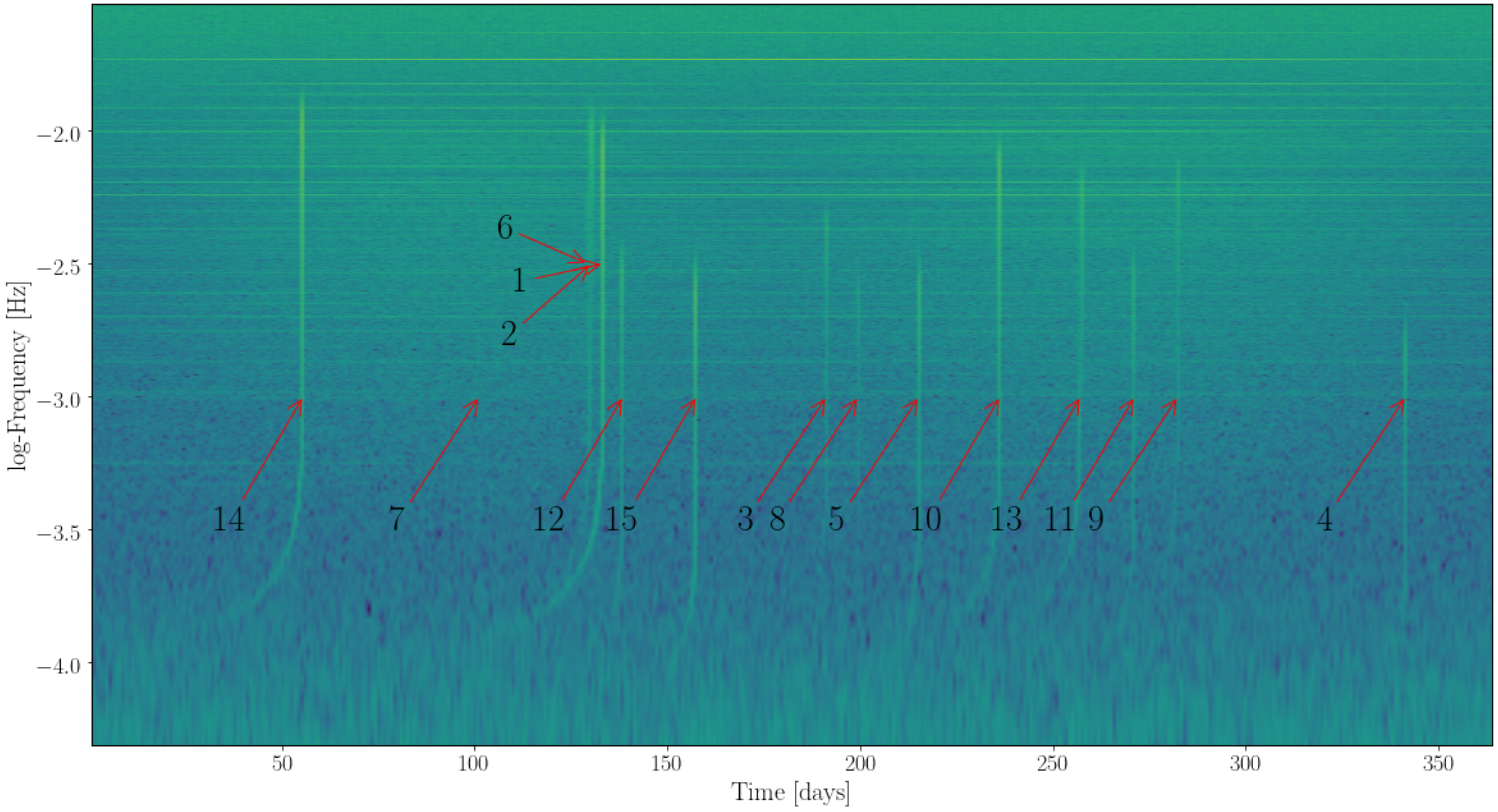
- MBHB signals in LISA
- Parameter space degeneracies
- Tools for Bayesian parameter estimation
- **LISA Data Challenge: Sangria**
- MBHB results for Sangria

LISA Data Challenge: Sangria



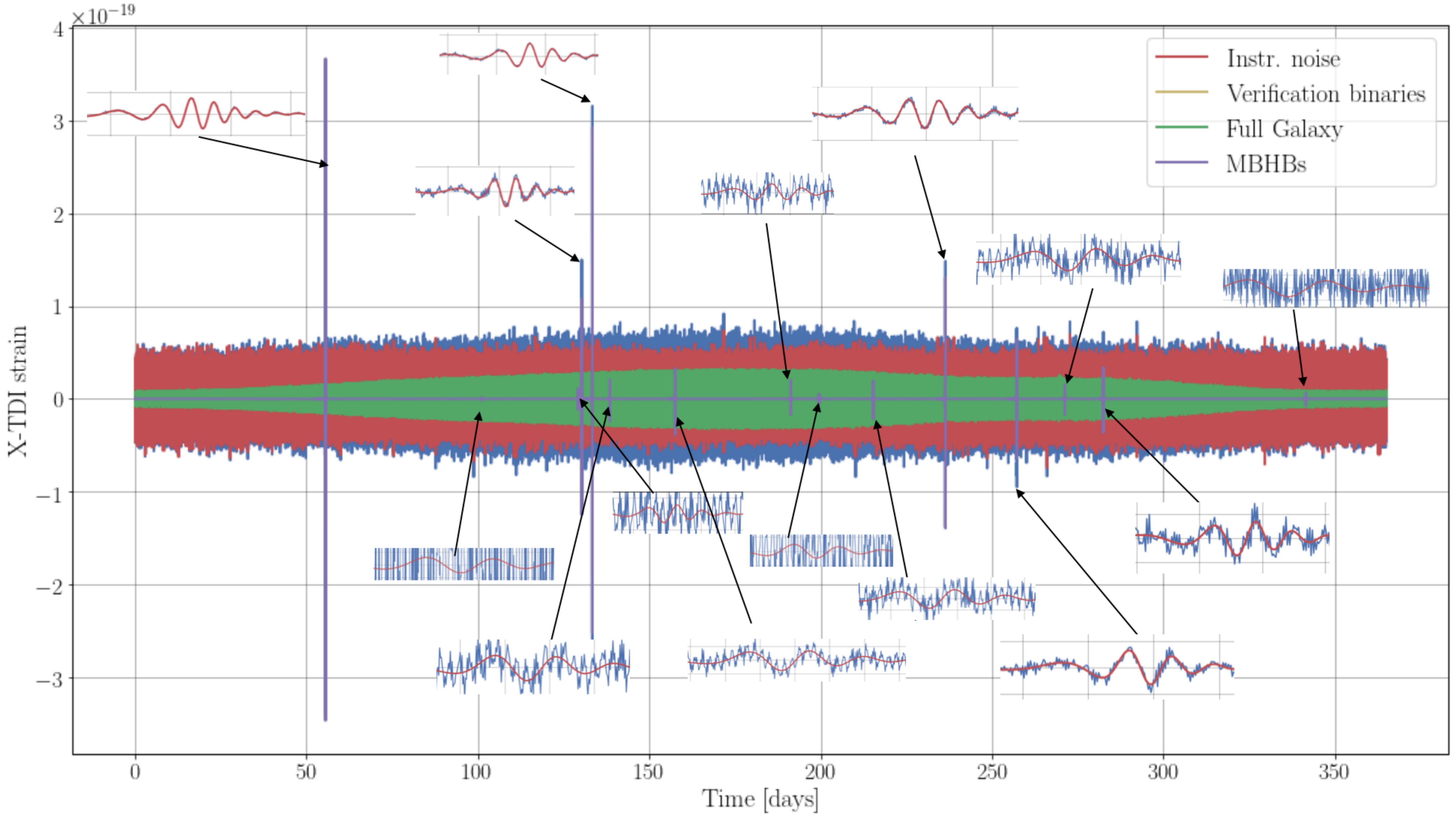
- **MBHBs**: loud and merger-dominated, localized in time but extended in frequency
- **GBs**: continuous signals very local in frequency, both individually resolvable and building up a background

LISA Data Challenge: Sangria

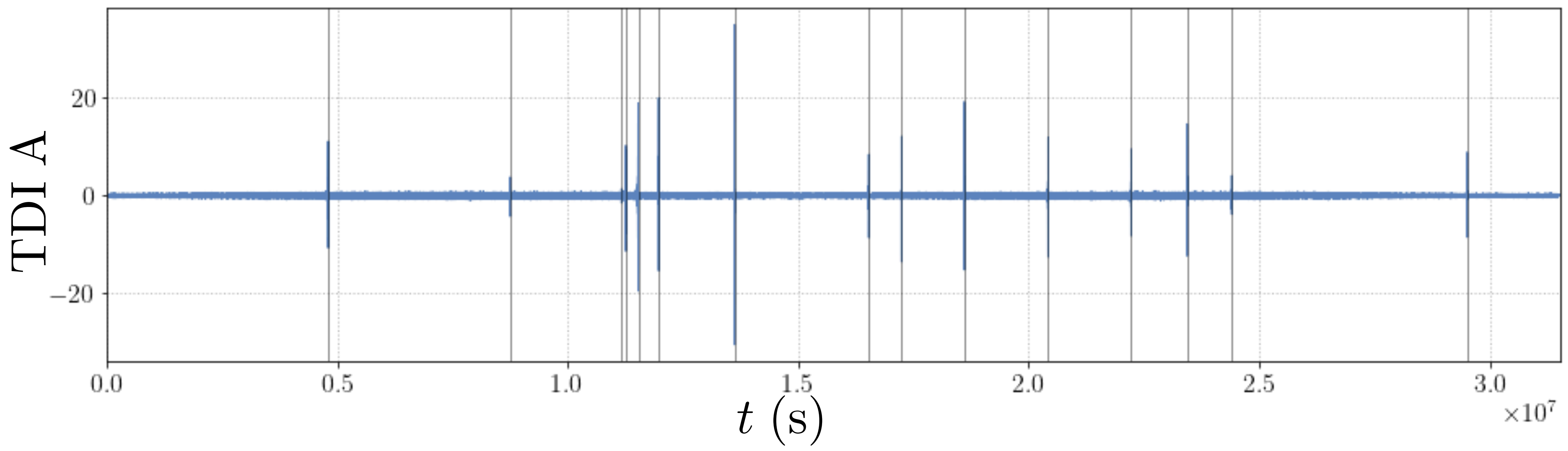


- **MBHBs**: loud and merger-dominated, localized in time but extended in frequency
- **GBs**: continuous signals very local in frequency, both individually resolvable and building up a background

LISA data - band-passed, whitened in time domain

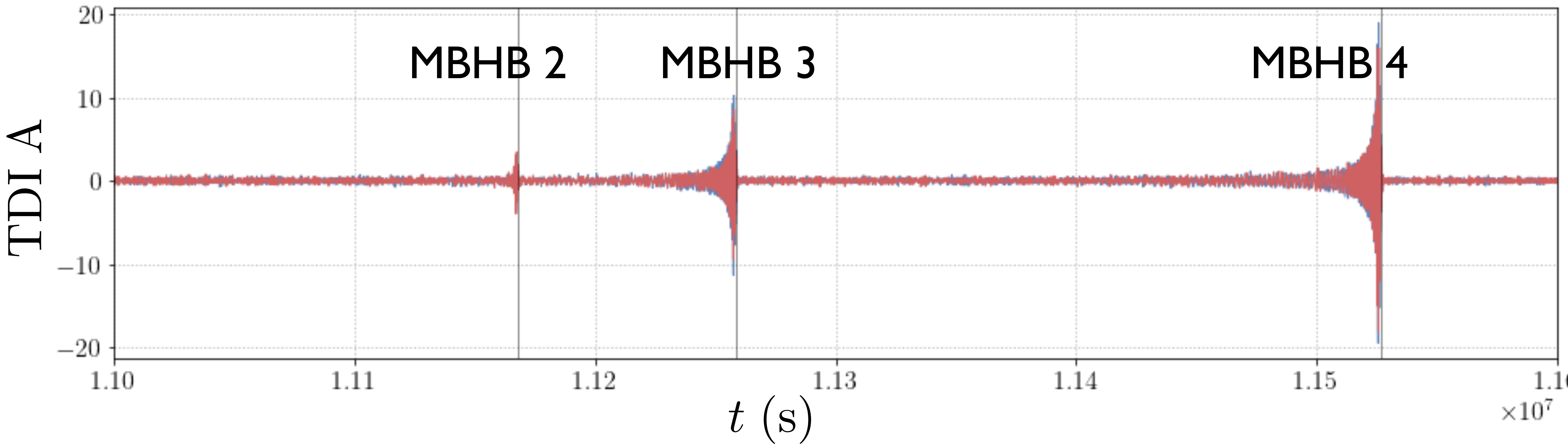


Whitened, band-passed data

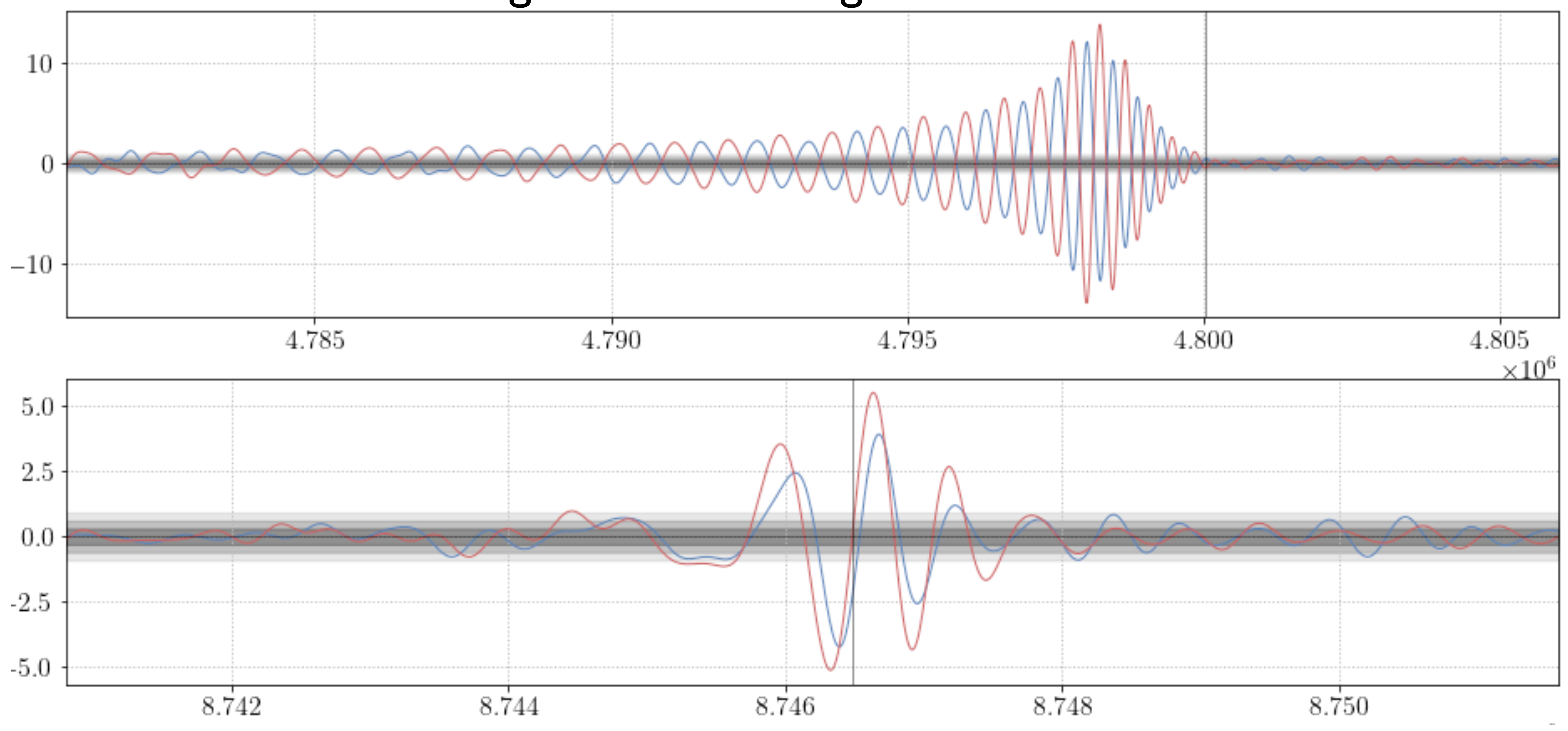


LISA data - band-passed, whitened in time domain

Whitened, band-passed data



sigma-thresholding for detection

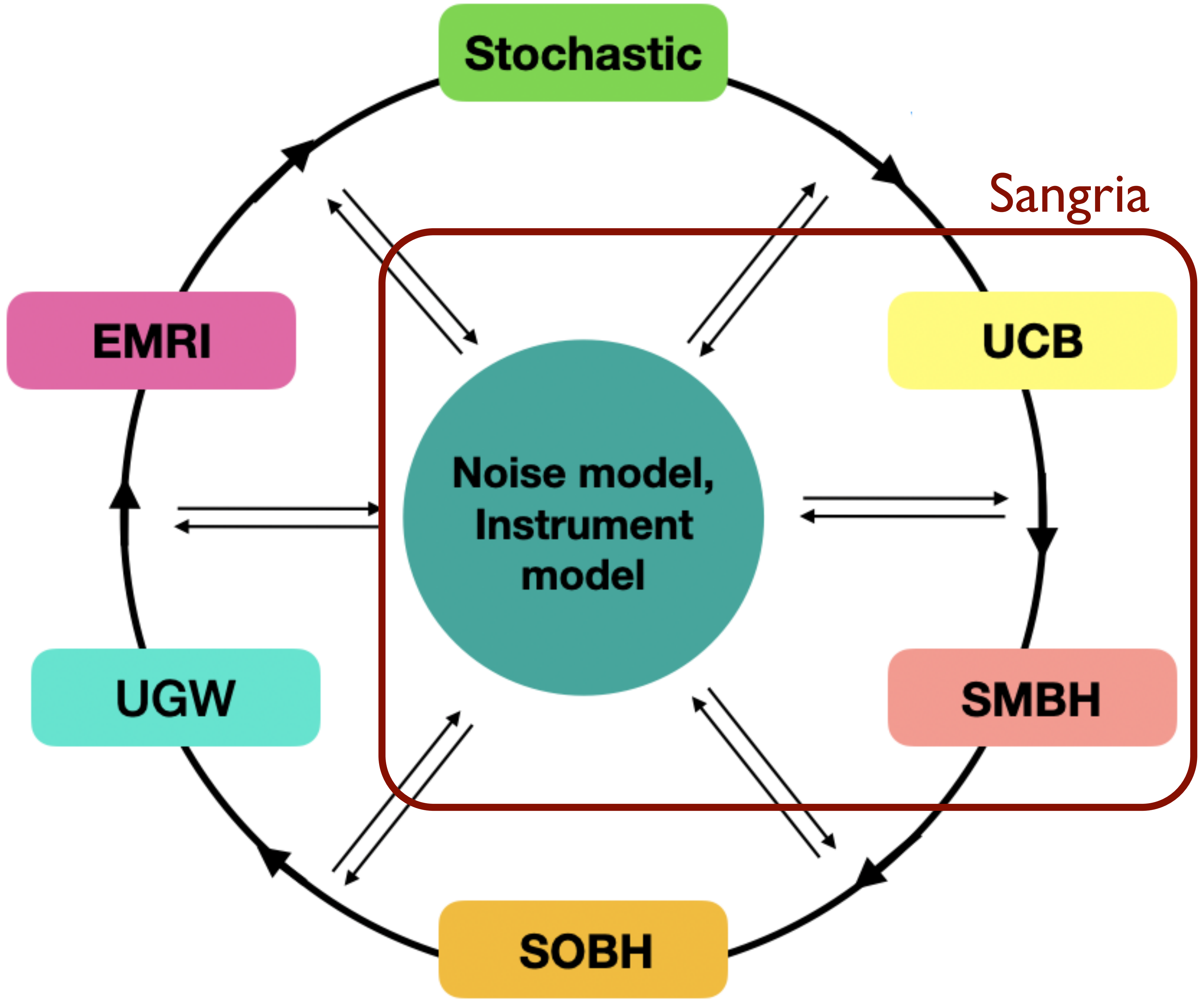


Detecting these MBHBs and getting tc is simple
Might be different at low masses

Outline

- MBHB signals in LISA
- Parameter space degeneracies
- Tools for Bayesian parameter estimation
- LISA Data Challenge: Sangria
- **MBHB results for Sangria [Preliminary]**

LDC Sangria: first steps of a global fit



[Cornish]

Chicken-and-egg problem

- MBHB analysis with full galaxy / GB analysis with full MBHBs are typically biased
- Some form of signal subtraction seems to be required

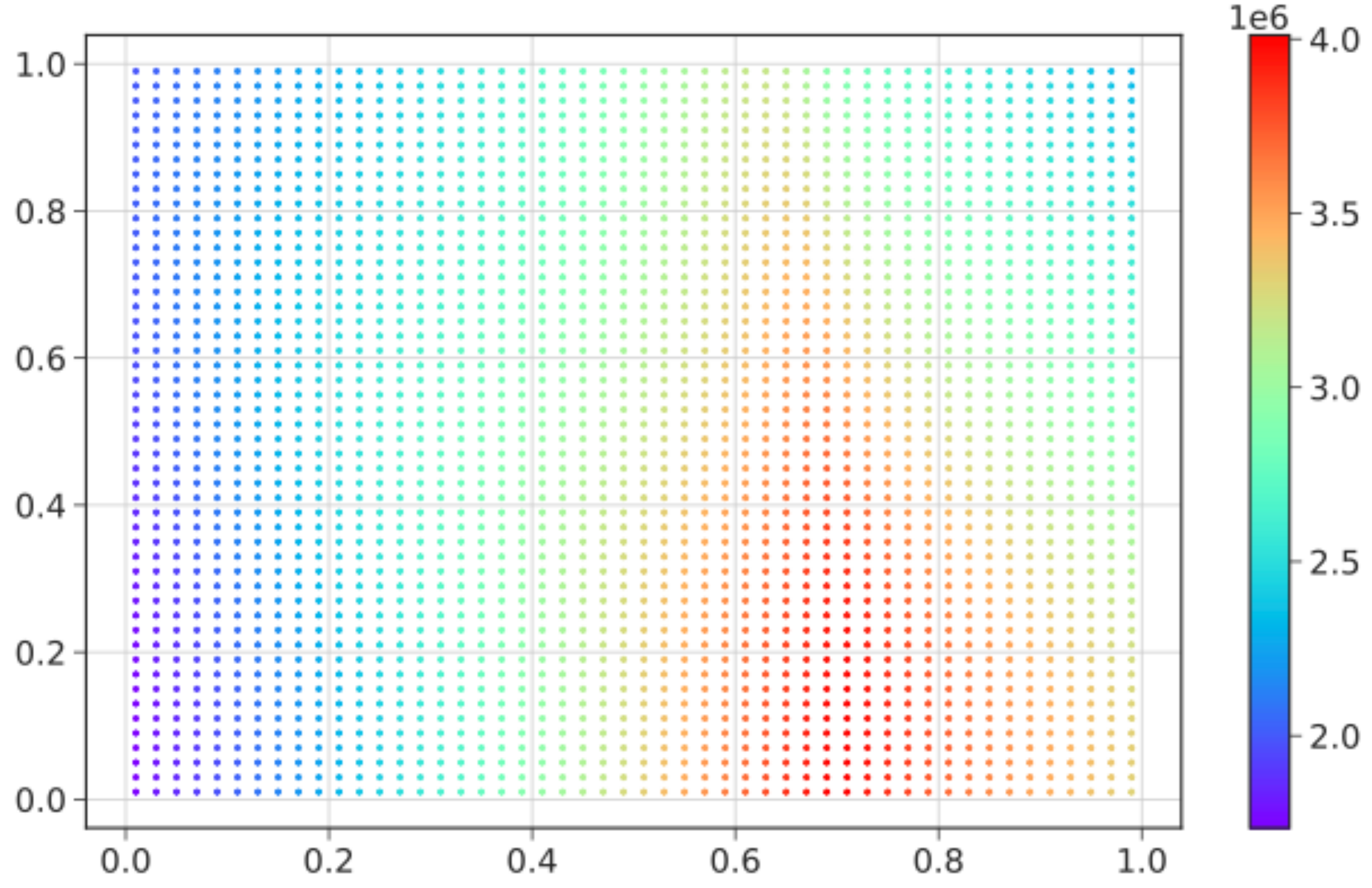
Global fit: a first approach

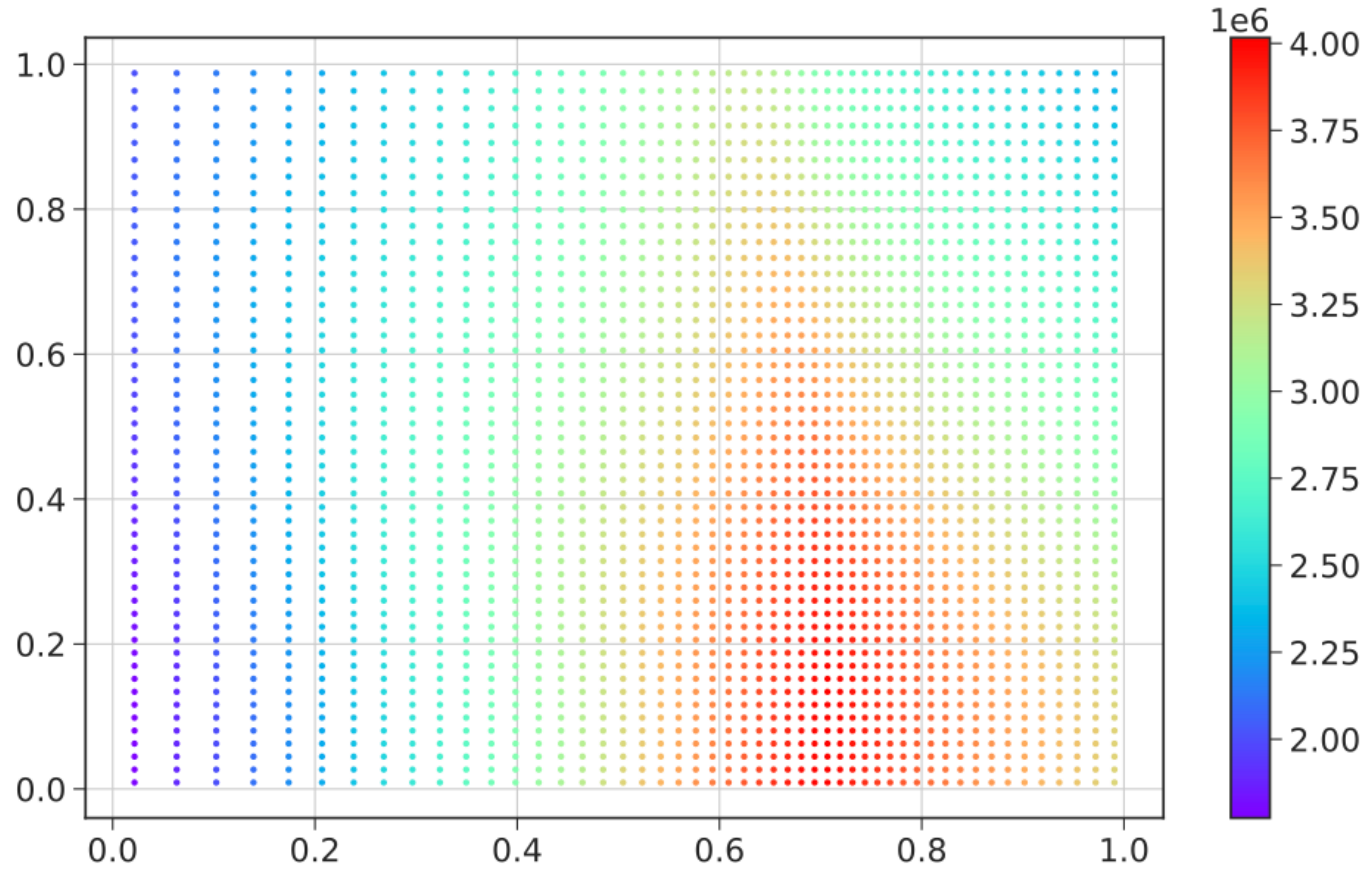
- First detection of MBHB
- Signal subtraction for MBHBs (no PE yet)
- First analysis of GBs, and noise estimation
- First PE (3-stage search/PE) for MBHBs
- Second analysis of GBs and noise ongoing…
- …

No unique approach ! See low-latency analysis of [Cornish 2021]

LDC Sangria: a first subtraction of MBHBs

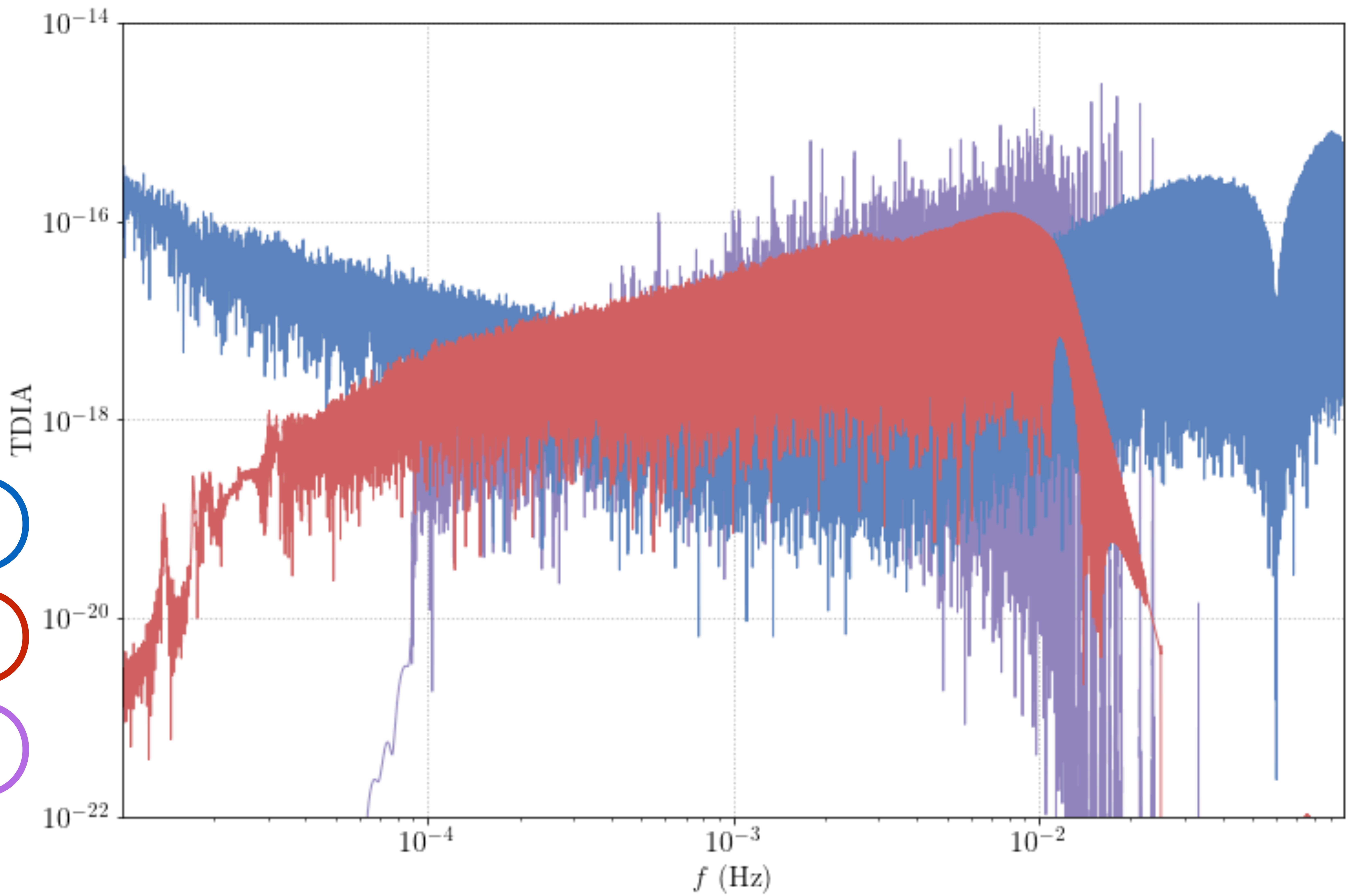
Analysis by [Senwen Deng, Stas Babak]





- Vegas: grid-based method with adaptive mesh refinement [Lepage 79]
- Restrict to low dimensions: masses+primary spin
- Produces best-fit estimate for MBHB signal subtraction

LDC Sangria: first steps of a global fit

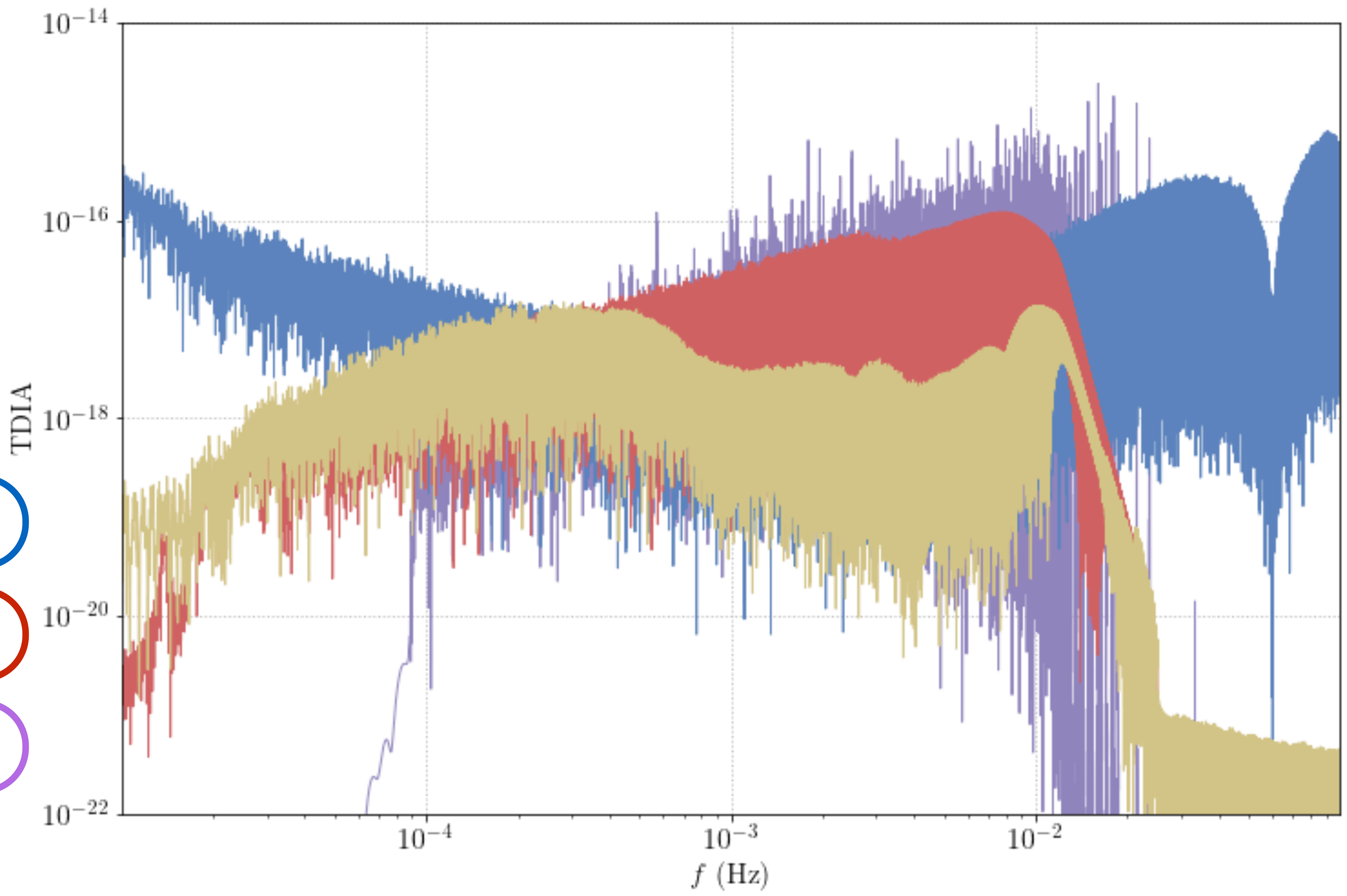


Instr. noise

MBHBs

GBs

LDC Sangria: first steps of a global fit



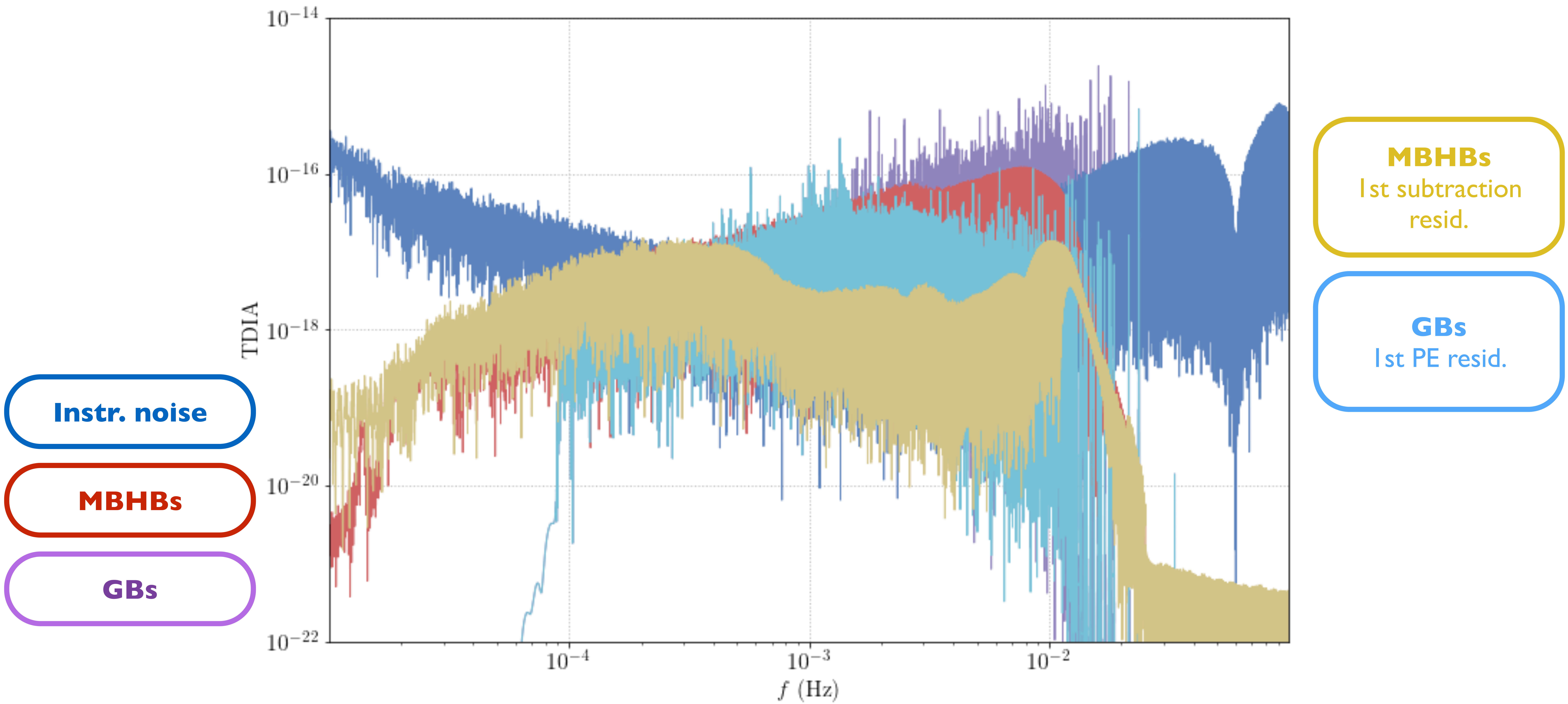
Instr. noise

MBHBs

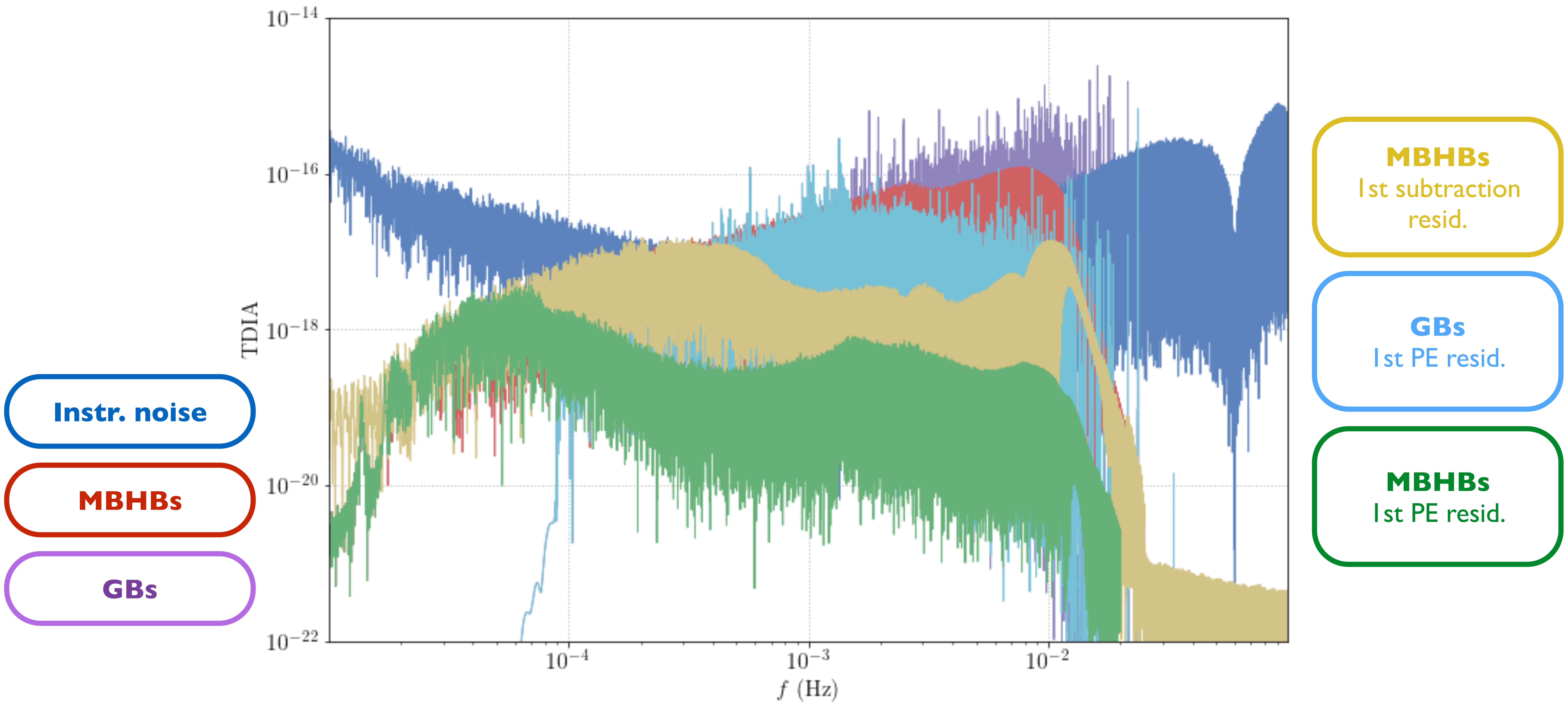
GBs

MBHBs
1st subtraction
resid.

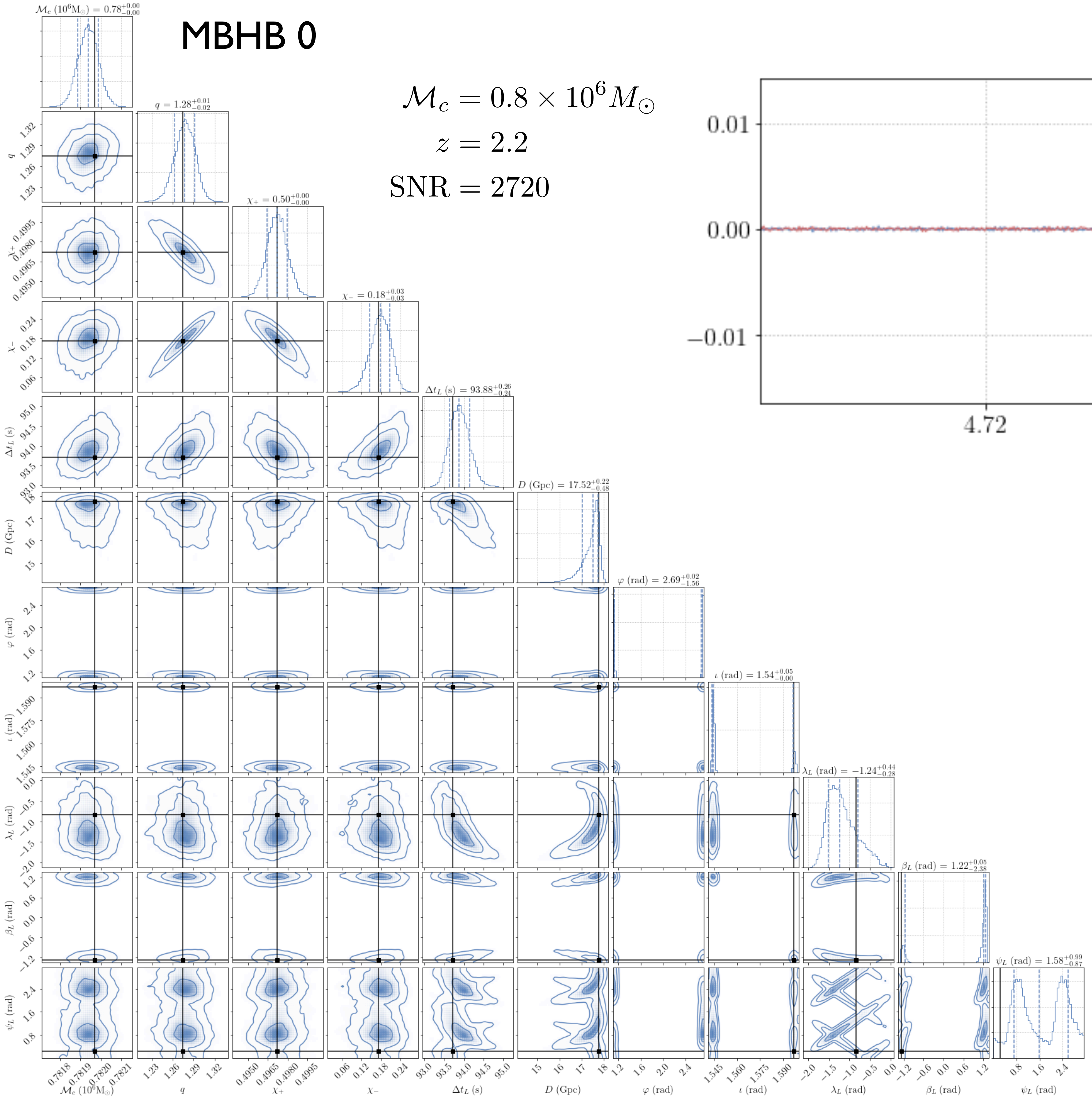
LDC Sangria: first steps of a global fit



LDC Sangria: first steps of a global fit

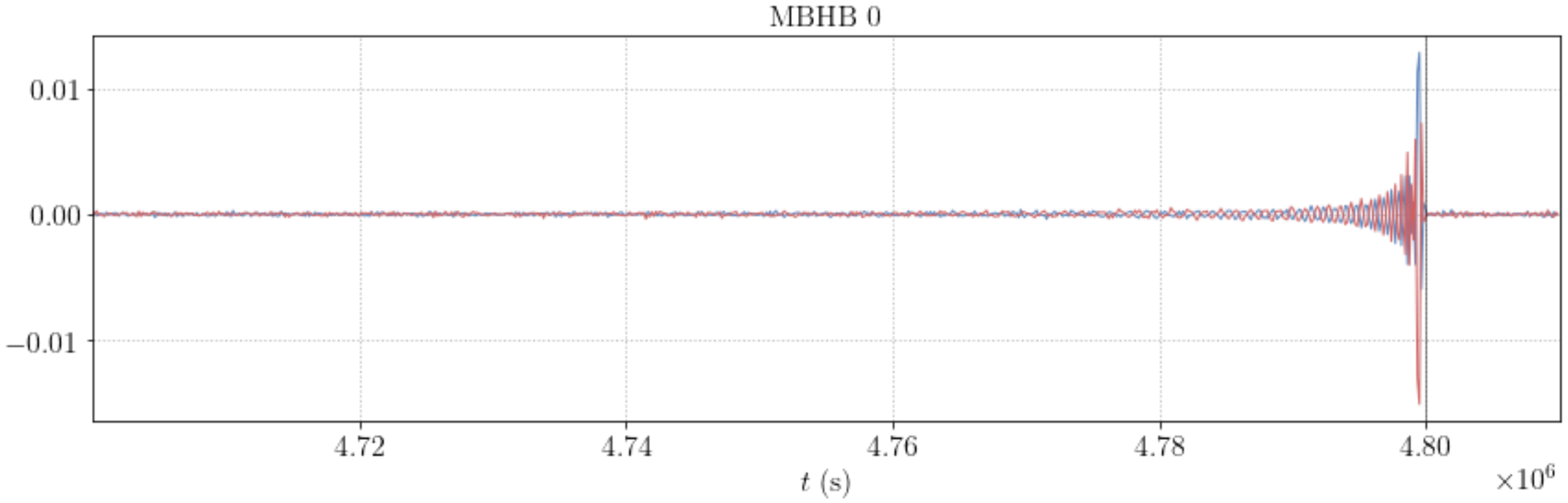


LDC Sangria: individual MBHB posteriors



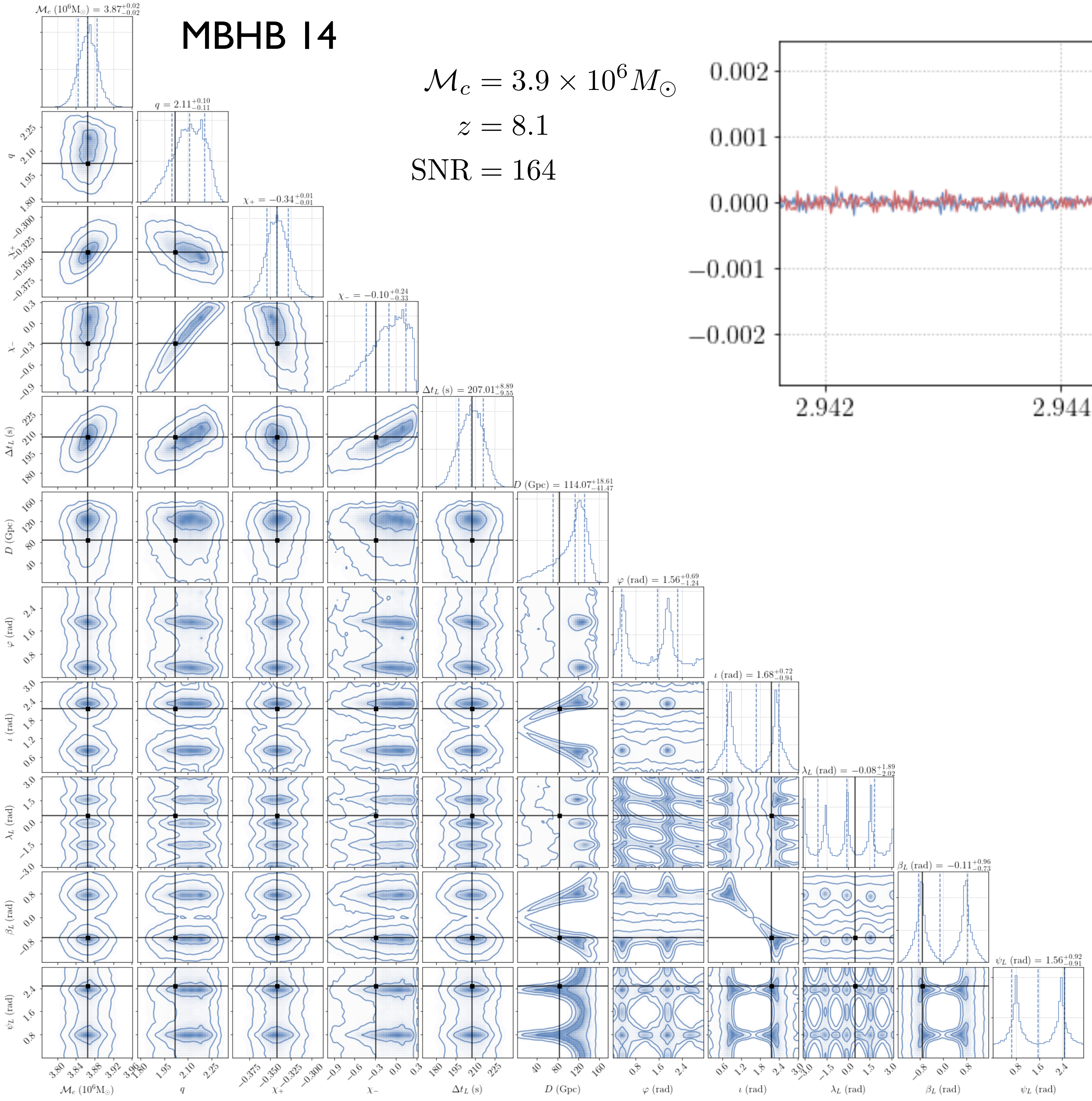
MBHB 0

$$\mathcal{M}_c = 0.8 \times 10^6 M_\odot$$
$$z = 2.2$$
$$\text{SNR} = 2720$$



LDC Sangria: individual MBHB posteriors

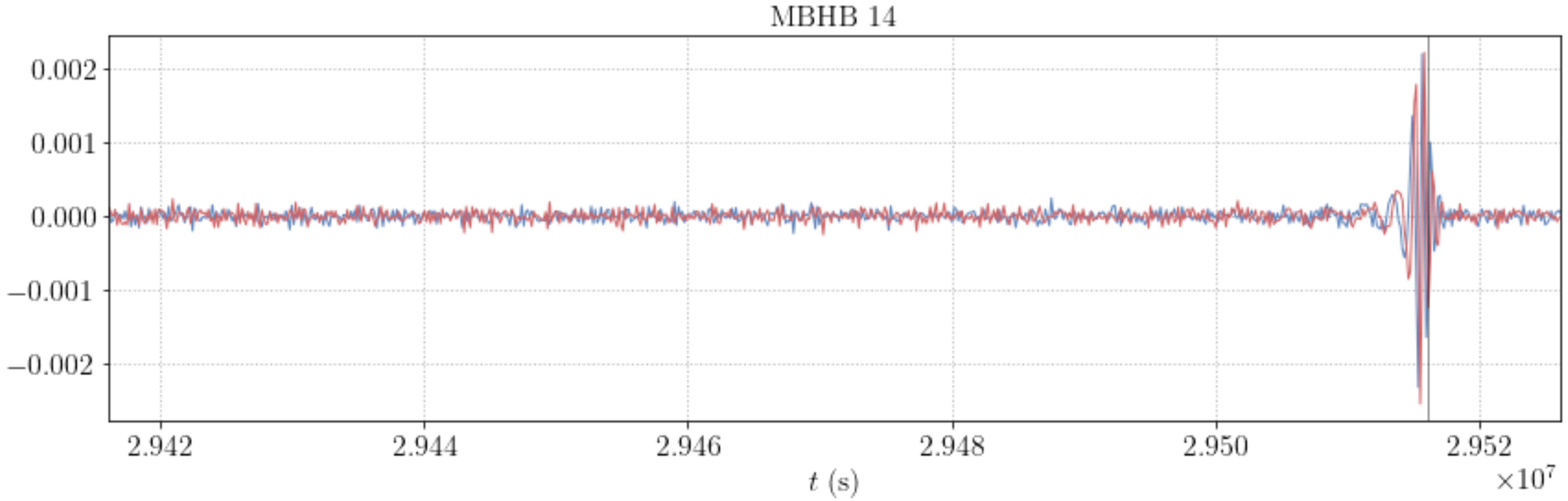
MBHB 14



$$\mathcal{M}_c = 3.9 \times 10^6 M_\odot$$

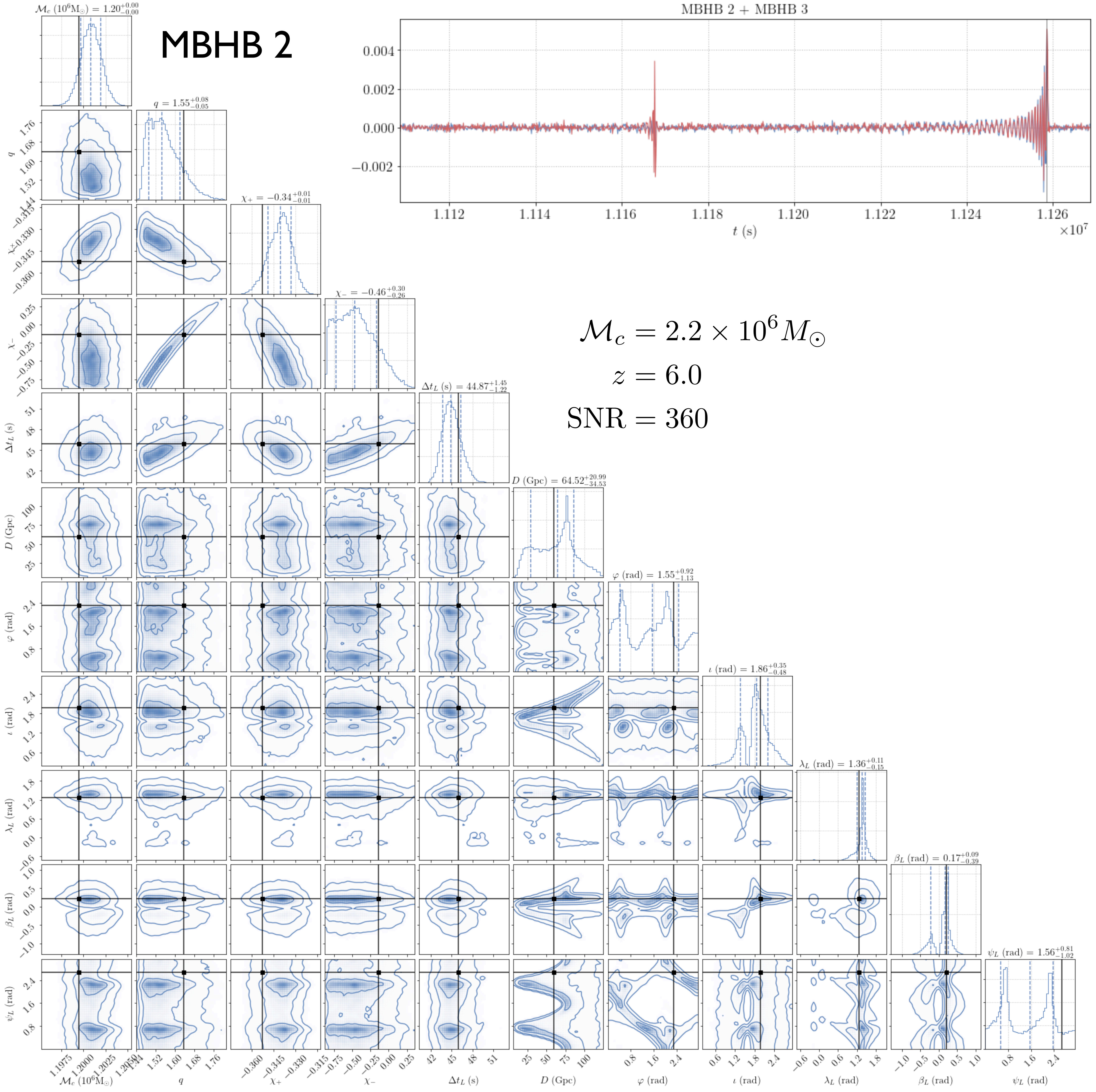
$$z = 8.1$$

$$\mathrm{SNR} = 164$$



LDC Sangria: individual MBHB posteriors

MBHB 2

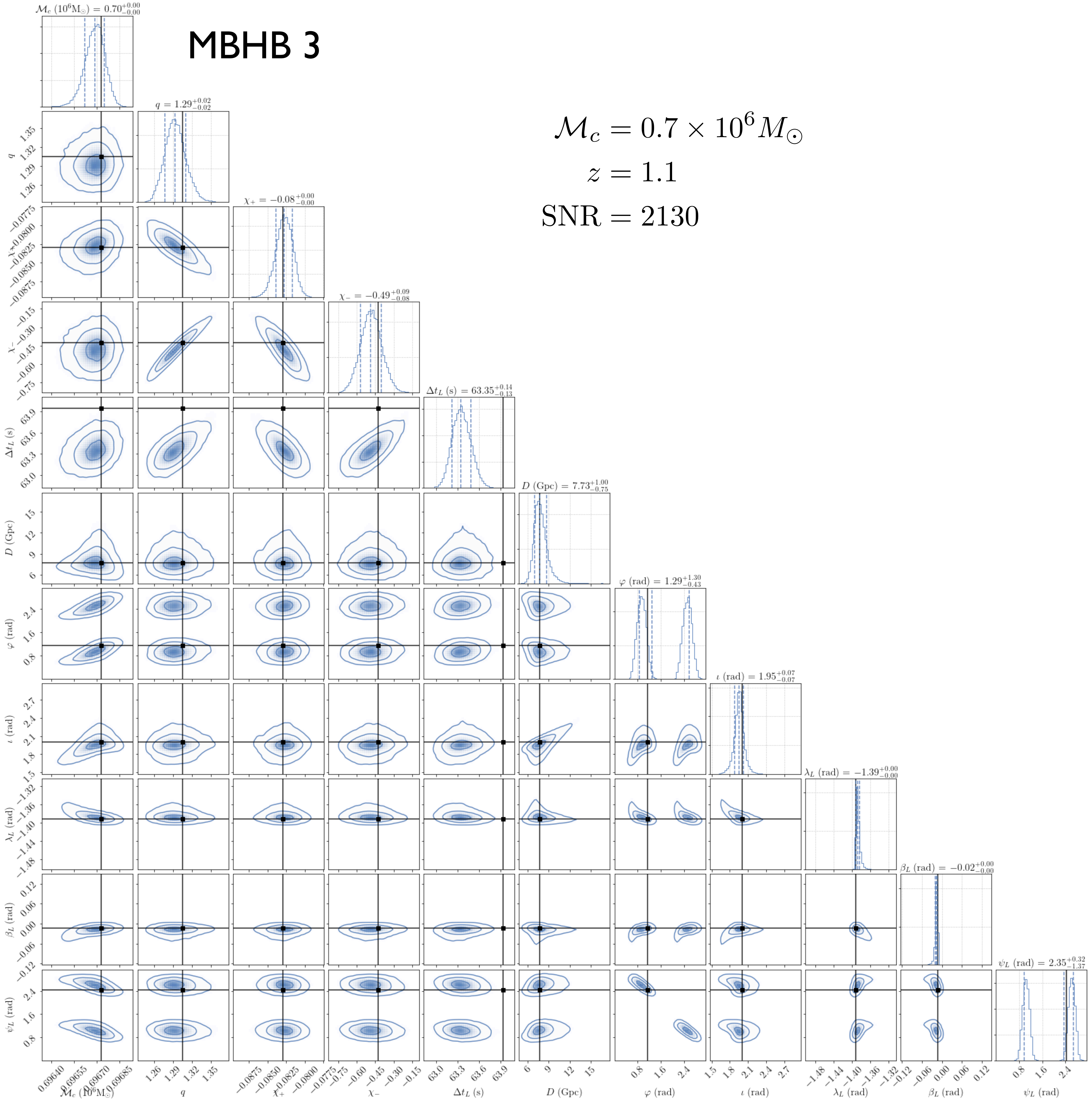


$\mathcal{M}_c = 2.2 \times 10^6 M_\odot$

$z = 6.0$

$\mathrm{SNR} = 360$

MBHB 3



$\mathcal{M}_c = 0.7 \times 10^6 M_\odot$

$z = 1.1$

$\mathrm{SNR} = 2130$

LDC Sangria: work in progress

- Resolve outstanding issues: biases in coalescence time (data generation ?), biases in intrinsic parameters for one source
- 2nd analysis of GBs (ongoing…) and noise
- 2nd analysis of MBHBs
- Confusion problem: do we have to analyze MBHBs jointly if they are correlated ?
- Multiple Gibbs iterations

LDC Sangria: work in progress

- Resolve outstanding issues: biases in coalescence time (data generation ?), biases in intrinsic parameters for one source
- 2nd analysis of GBs (ongoing…) and noise
- 2nd analysis of MBHBs
- Confusion problem: do we have to analyze MBHBs jointly if they are correlated ?
- Multiple Gibbs iterations

Thank you for your attention