LISA data analysis: status and challenges

Stanislav Babak

Laboratoire Astroparticule & Cosmologie (APC), CNRS.

21-25 Nov. Toulouse, 2022

"Wanderer, your footsteps are the road, and nothing more; wanderer, there is no road, the road is made by walking. By walking one makes the road, and upon glancing behind one sees the path that never will be trod again. Wanderer, there is no road — Only wakes upon the sea." [Antonio Machado]

Overview

Bayesian data analysis

LISA data analysis

Galactic binaries (GBs)

Merging MBHBs

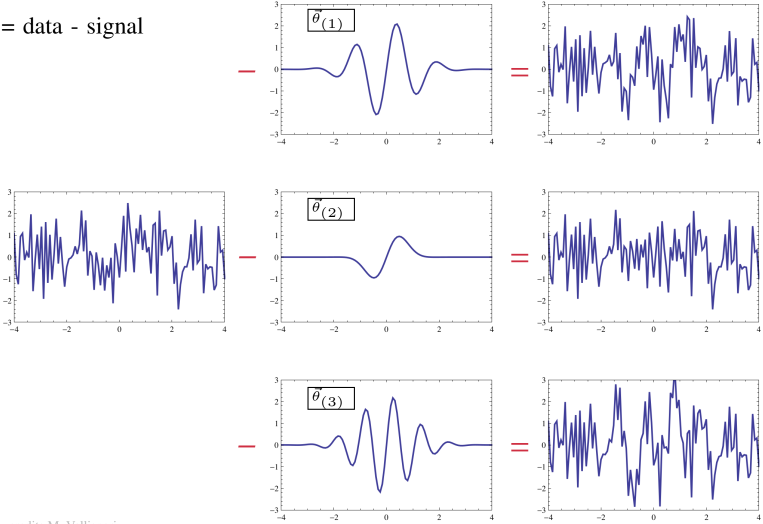
EMRIs

SBBHs

AOB

Matched filtering and parameter estimation

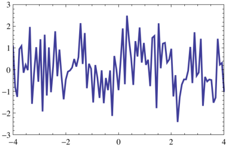
noise = data - signal

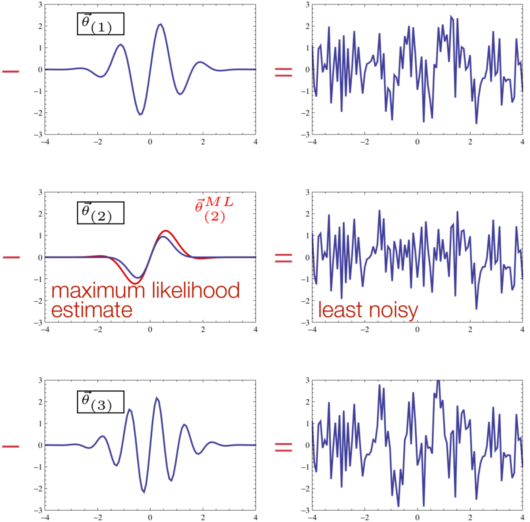


credits M. Vallisneri

Matched filtering and parameter estimation

noise = data - signal
p(signal parameters)
= p(noise residuals)





credits M. Vallisneri

Bayesian approach

- Expensive computationally: often used when the signal is detected, or the data volume is not very large:

$$data = noise + signal.$$

- Allows to test several models (different signal' models, non-GR theories).
- We have to assign the prior probability to our models and parameters of each model.
- We treat parameters describing a signal as random variables and estimating probability distribution function(s) for each parameter based on the observed data (*posterior*)

Bayesian approach

- Bayesian model selection. Consider several models:

$$P(M_i, d) = \frac{P(d|M_i)\pi(M_i)}{p(d)}$$

- Each model is assumed to be parametrized by a set of parameter $\vec{\theta}_i$:

Likelihood

prior

$$P(\vec{\theta}_i|d, M_i) = \frac{P(d|\vec{\theta}_i, M_i)\,\pi(\vec{\theta}_i)}{P(d|M_i)}$$

Posterior

evidence

Bayesian model selection

- Evidence of model M_i: important for the model selection

$$P(d|M_i) = \int d\vec{\theta}_i \, p(d|\vec{\theta}_i, M_i)\pi(\vec{\theta}_i)$$

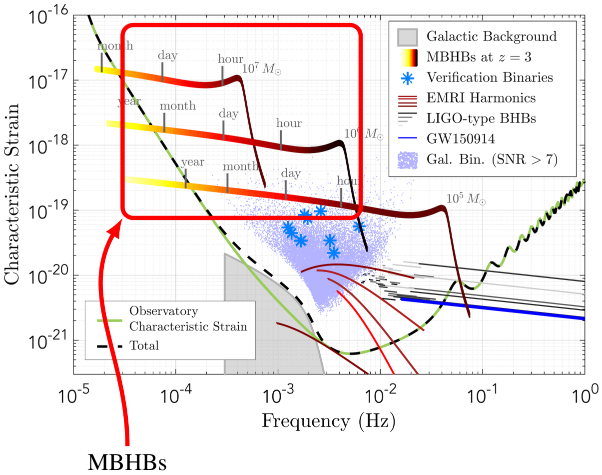
- Probability of the model M_i given a data

$$P(M_i|d) = \left[\int d\vec{\theta}_i \, p(d|\vec{\theta}_i, M_i)\pi(\vec{\theta}_i)\right] \frac{\pi(M_i)}{P(d)}$$

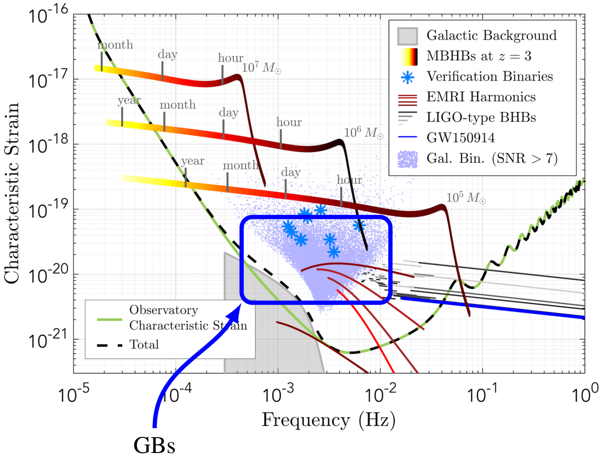
- *Odd ratio*: Cannot evaluate $P(d)$ - evaluate the ratio of probabilities instead:

$$O_{a,b} = \frac{P(M_a|d)}{P(M_b|d)} = \frac{P(d|M_a)}{P(d|M_b)} \frac{\pi(M_a)}{\pi(M_b)}$$

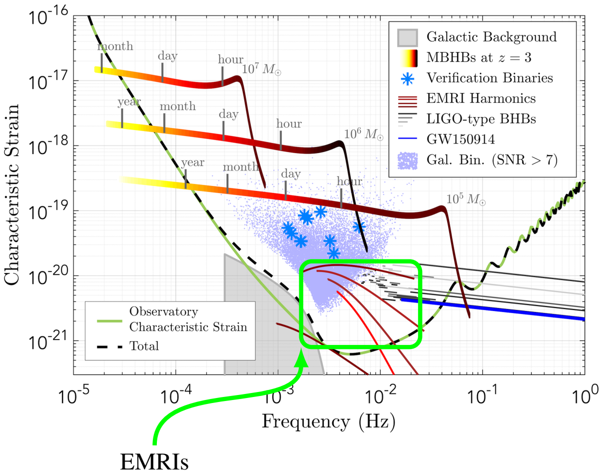
What is special about LISA data



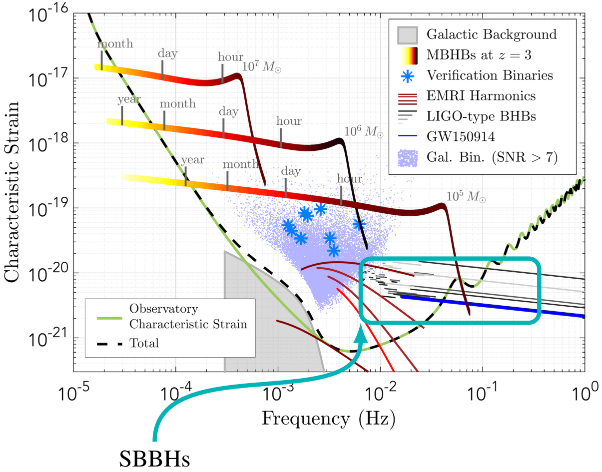
What is special about LISA data



What is special about LISA data



What is special about LISA data



What is special about LISA data

- Data size is small (does not require large data storage) But the data products are huge
- Data is dominated by signals (data contains many strong signals)
- TDI (time delay interferometry) (the main source of noise is the laser frequency noise in measuring $\delta\nu/\nu$, requires time delayed combination of measurements to cancel this noise)
- Non-trivial response function (many signal have $\lambda_{GW} \sim L$, it requires computation of response beyond the long wavelength approximation)
- GW signals are long lived (signals from the same and different kind of sources are present simultaneously in the data and could overlap in time and frequency)

What is special about LISA data

- Data size is small (does not require large data storage) But the data products are huge
- Data is dominated by signals (data contains many strong signals)
- TDI (time delay interferometry) (the main source of noise is the laser frequency noise in measuring $\delta\nu/\nu$, requires time delayed combination of measurements to cancel this noise)
- Non-trivial response function (many signal have $\lambda_{GW} \sim L$, it requires computation of response beyond the long wavelength approximation)
- GW signals are long lived (signals from the same and different kind of sources are present simultaneously in the data and could overlap in time and frequency)

What is special about LISA data

- Data size is small (does not require large data storage) But the data products are huge
- Data is dominated by signals (data contains many strong signals)
- TDI (time delay interferometry) (the main source of noise is the laser frequency noise in measuring $\delta\nu/\nu$, requires time delayed combination of measurements to cancel this noise)
- Non-trivial response function (many signal have $\lambda_{GW} \sim L$, it requires computation of response beyond the long wavelength approximation)
- GW signals are long lived (signals from the same and different kind of sources are present simultaneously in the data and could overlap in time and frequency)

What is special about LISA data

- Data size is small (does not require large data storage) But the data products are huge
- Data is dominated by signals (data contains many strong signals)
- TDI (time delay interferometry) (the main source of noise is the laser frequency noise in measuring $\delta\nu/\nu$, requires time delayed combination of measurements to cancel this noise)
- Non-trivial response function (many signal have $\lambda_{GW} \sim L$, it requires computation of response beyond the long wavelength approximation)
- GW signals are long lived (signals from the same and different kind of sources are present simultaneously in the data and could overlap in time and frequency)

What is special about LISA data

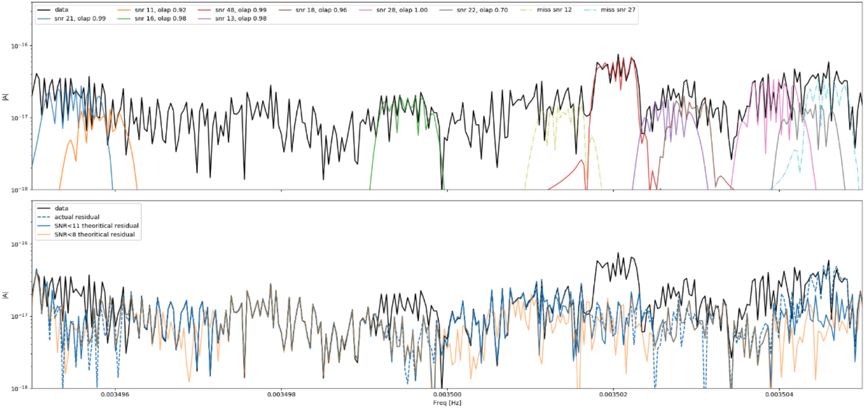
- Data size is small (does not require large data storage) But the data products are huge
- Data is dominated by signals (data contains many strong signals)
- TDI (time delay interferometry) (the main source of noise is the laser frequency noise in measuring $\delta\nu/\nu$, requires time delayed combination of measurements to cancel this noise)
- Non-trivial response function (many signal have $\lambda_{GW} \sim L$, it requires computation of response beyond the long wavelength approximation)
- GW signals are long lived (signals from the same and different kind of sources are present simultaneously in the data and could overlap in time and frequency)

Galactic binaries

- Sources are almost monochromatic, some have measurable change on frequency (+ve and -ve), signal is modulated due to LISA motion.
- Large number: resolving the individual sources. Trans-dimensional model selection.
- Model selection between stochastic GB foreground and overlapped individual signals.
- Re-analysing as more data becomes available: time evolving solution.

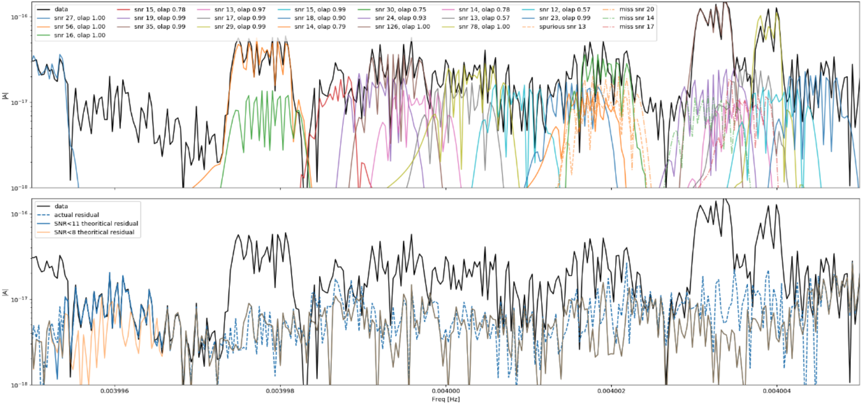
Galactic binaries: number of resolvable systems

Let us walk across frequency and look at the number of resolvable sources: resolution for the sources falling at (roughly) the same frequency comes from 1. frequency evolution 2. modulation of the signal (if sources are separated on the sky).



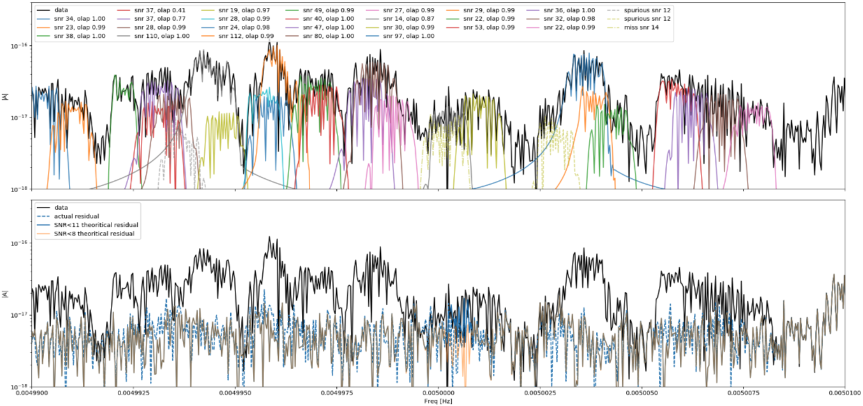
Galactic binaries: number of resolvable systems

Let us walk across frequency and look at the number of resolvable sources: resolution for the sources falling at (roughly) the same frequency comes from 1. frequency evolution 2. modulation of the signal (if sources are separated on the sky).



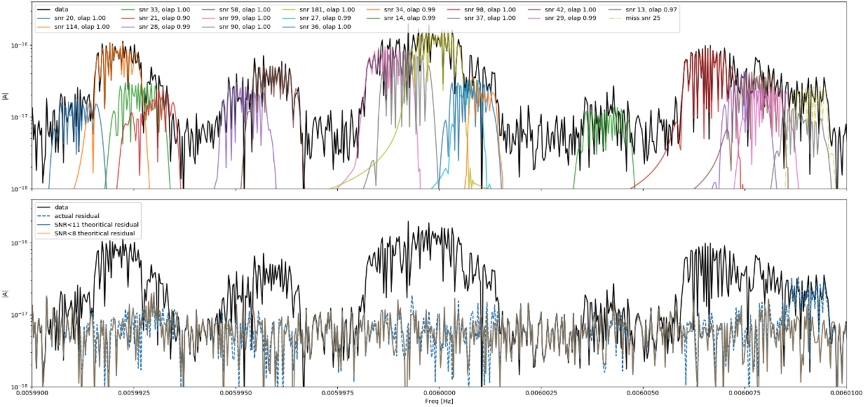
Galactic binaries: number of resolvable systems

Let us walk across frequency and look at the number of resolvable sources: resolution for the sources falling at (roughly) the same frequency comes from 1. frequency evolution 2. modulation of the signal (if sources are separated on the sky).



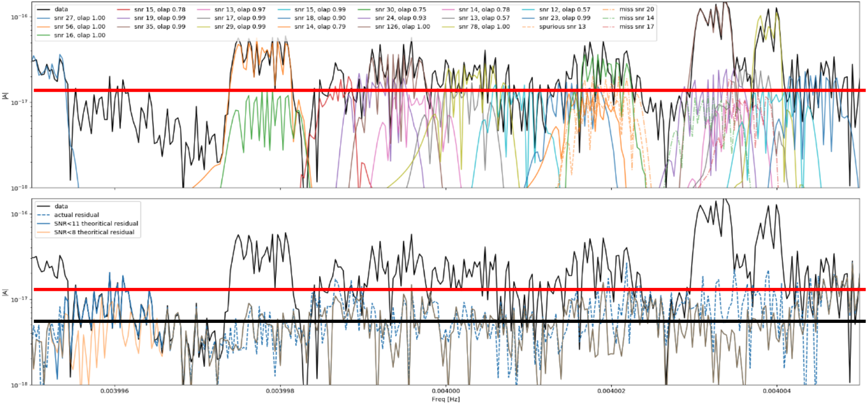
Galactic binaries: number of resolvable systems

Let us walk across frequency and look at the number of resolvable sources: resolution for the sources falling at (roughly) the same frequency comes from 1. frequency evolution 2. modulation of the signal (if sources are separated on the sky).



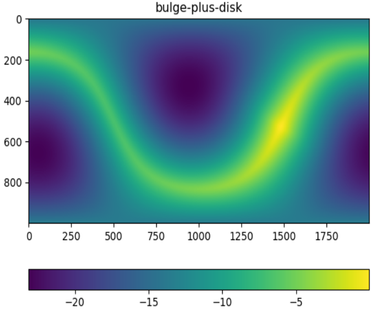
Confusion noise or resolvable sources

When there are too many overlapped signals, even though *theoretically* they are resolvable, but the model selection might prefer high level of the noise and only few brightest source. Model with fewer parameters often wins. Smooth broadband model of the noise?



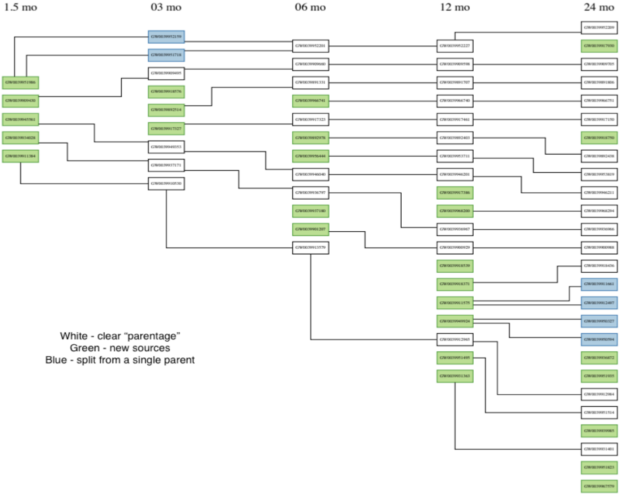
GBs (and in general) choice of priors

- Taking a wide prior will penalize you during the model selection procedure
- Taking narrow prior: possibility of loosing sources and interesting features not predicted by a "default" model
- Need to convert to dimensionless prior



GBs time-evolving solution

Analysis of LISA data with cadence 3 months: keep solution as a starting point (or as a proposal) for analysis of longer dataset.

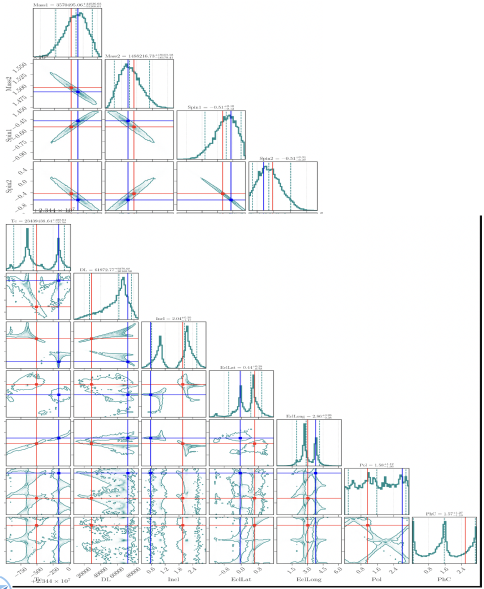


[Littenberg+ PRD, 2020]

Merging massive black hole binaries

- Relatively short duration: most of SNR comes from the last hours-days.
- Most likely the strongest sources: stringent requirements on modelling the GW signal (including LISA response)
- Very complex posterior (multimodality)
- we want to detect the signal with unbiased measurement of the sky position: pre-merger alert

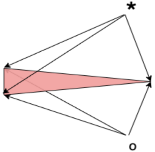
MBHB: posterior distribution

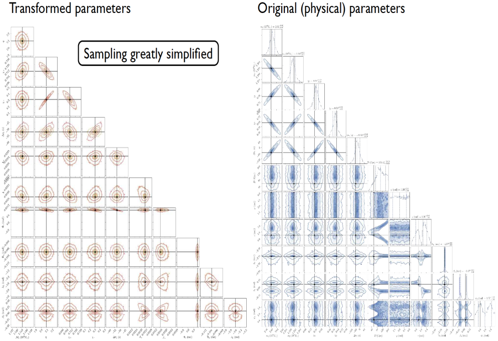


- Intrinsic parameters are usually well determined (modulo correlations)
- Extrinsic posteriors are multi-modal and show non-trivial structure (very hard to sample): especially if we consider only a dominant harmonic
- Need to re-parametrize: use LISA-based frame, use a smart combination of parameters

MBHB: reparametrized

- Short duration: LISA hardly moves → strong degeneracy in the sky: mirrored position
- Need to re-parametrize: use a smart combination of parameters where posteriror close to Gaussian distributions.





Extreme mass ratio inspirals

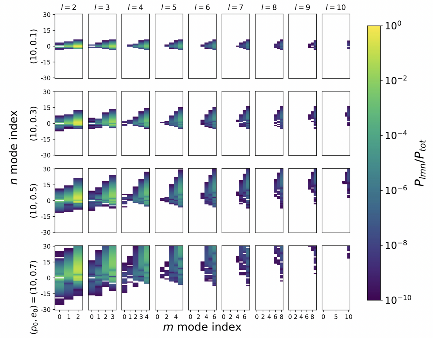
- Signal is long-lived
- Signal is present by a large number of harmonics: strength changes in time (intrinsic + LISA modulation). SNR of each harmonic is not large
- Large number of secondary maxima on the likelihood surface + small volume of the signal in a large prior volume: hard to find the signal.

EMRI Harmonics

- The GW signal from EMRIs can be seen as a superposition of harmonics of three quasi-periodic frequencies (f_r, f_θ, f_ϕ)

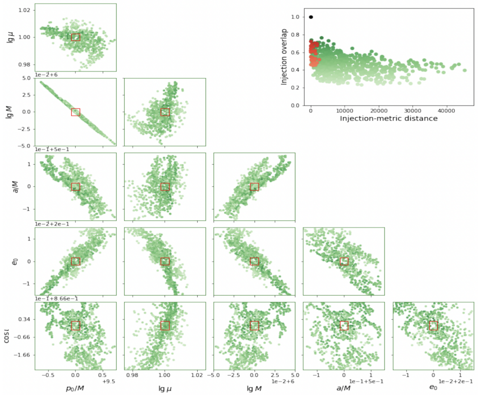
$$h(t) = \Re \sum_{k,l,n} A_{kln} e^{i(k\varphi_r(t) + l\varphi_\theta(t) + n\varphi_\phi(t))}$$

- For detection purposes one can use phenomenological waveform (simple analytic fit for phase(s) and amplitudes).
- How many harmonics we need? Answer is parameter dependent



[M. Katz+ PRD 2021]

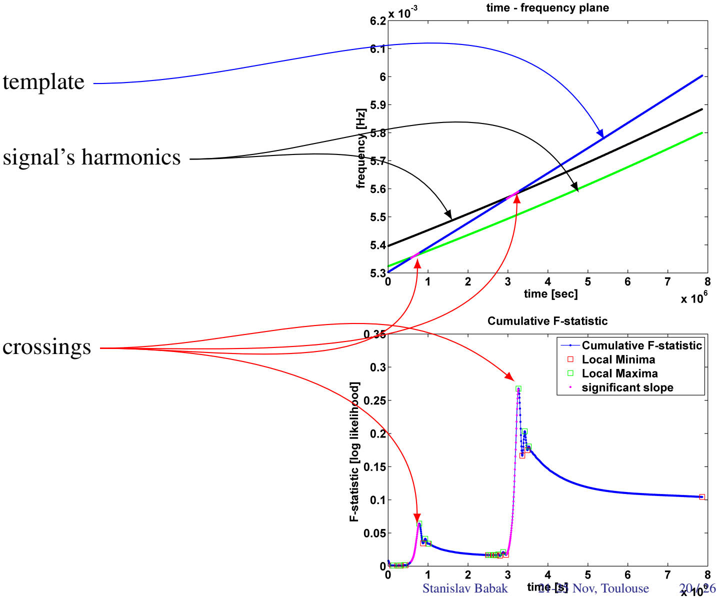
EMRI: secondary maxima



[A. Chua & C.Cutler 2021]

- Large number of maxima makes it hard to get accurate solution
- First approach: find as many maxima and use them as a constrain [Babak+ 2009]
- Second approach: Suppress them [A. Cua 2022]

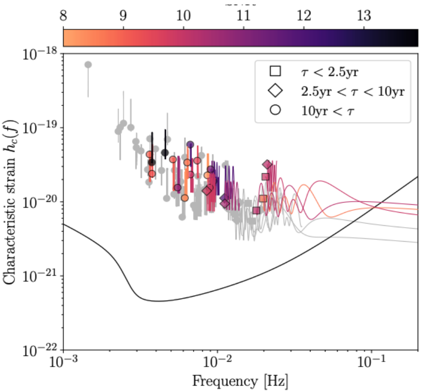
Multimodality of the likelihood surface



Stellar mass black hole binaries

- GW signal: from monochromatic to evolving outside the LISA band (multi-band sources), strongly depend on the orbital frequency (at $t = 0$)
- Similar to EMRIs: occupy a small volume compare to the total prior volume + low SNR = Hard to search
- Could be quite eccentric
- Possible archival search: first detected by 3G on the ground and then search in archival LISA data with a narrow prior.

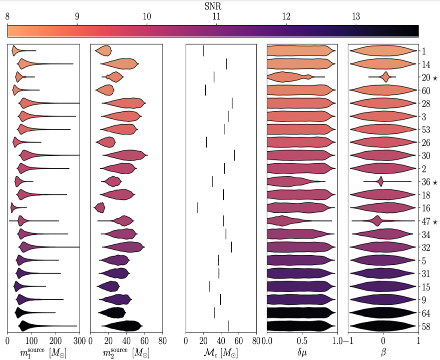
SBBH Frequency evolution

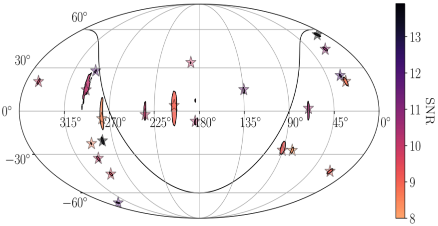


[R. Buscicchio+ PRD 2021]x

- We expect few sources during the LISA mission time to evolve out of the LISA sensitivity band.
- SNR for detection: 8-15
- Good sky position (long lived, high freq, LISA motion), chirp mass.

SBBH parameter estimation



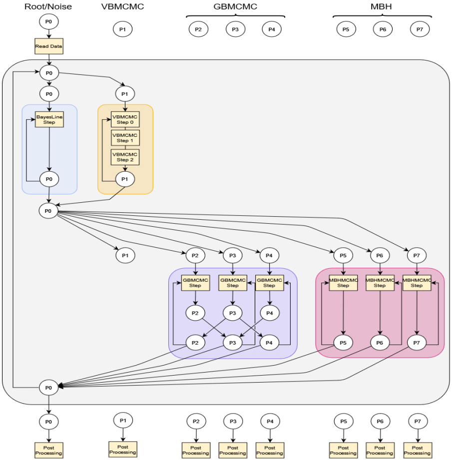


[R. Buscicchio+ PRD 2021]

Stochastic GW signal

- Stochastic signal from the early Universe: exploration. (e/w phase transition)
- Stochastic GW signal from astrophysical populations: highly uncertain.
- Stochastic GW signal could dominate over instrumental noise across a very large freq. band
- Challenging to disentangle from the instrumental noise and from the residuals of resolved sources.

Global fit: Sangria



[T. Littenberg]

- Need to fit Galactic binaries (∼ 10000 resolvable + stochastic + few dozen VGBs), MBHBs, Noise estimation
- The search/PE is done in an iterative manner.
- When is good enough (end of iterations)? How assess goodness of results? Load balance?

Conclusion

- LISA is fantastic (if we do not screw it up). But it will be challenging to pull all the science out of the data
- Some challenges are/will be common with ground based (3G) GW detectors, but most of them are unique
- Global fit: GW sources of the same and different type, instrumental noise (+ artifacts), unknown unknown. It is exploration mission, we need to be ready and flexible.
- What did I miss?