Detecting Gravitational Waves from space: Searching for noisy signals

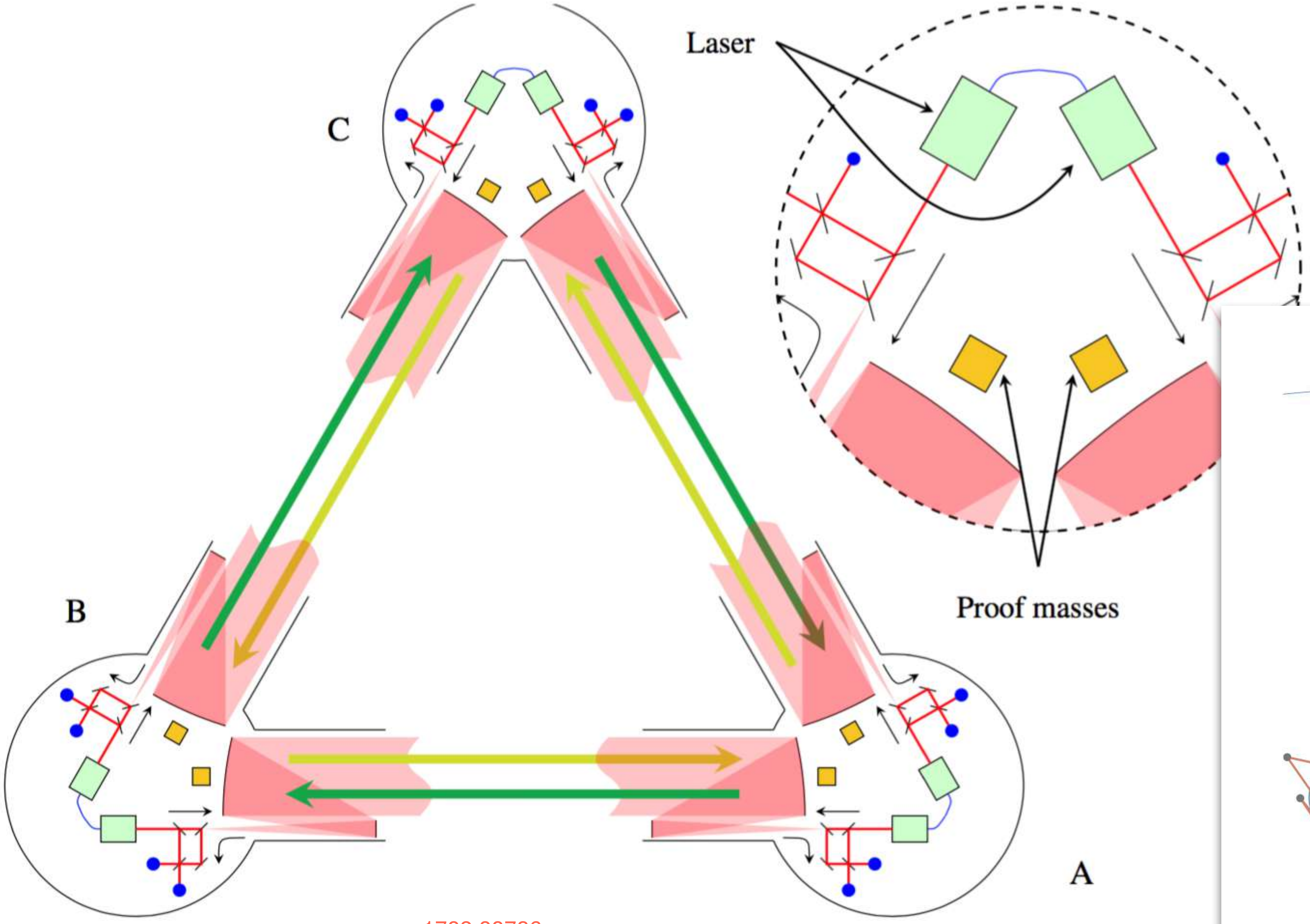
N Karnesis

Aristotle University of Thessaloniki

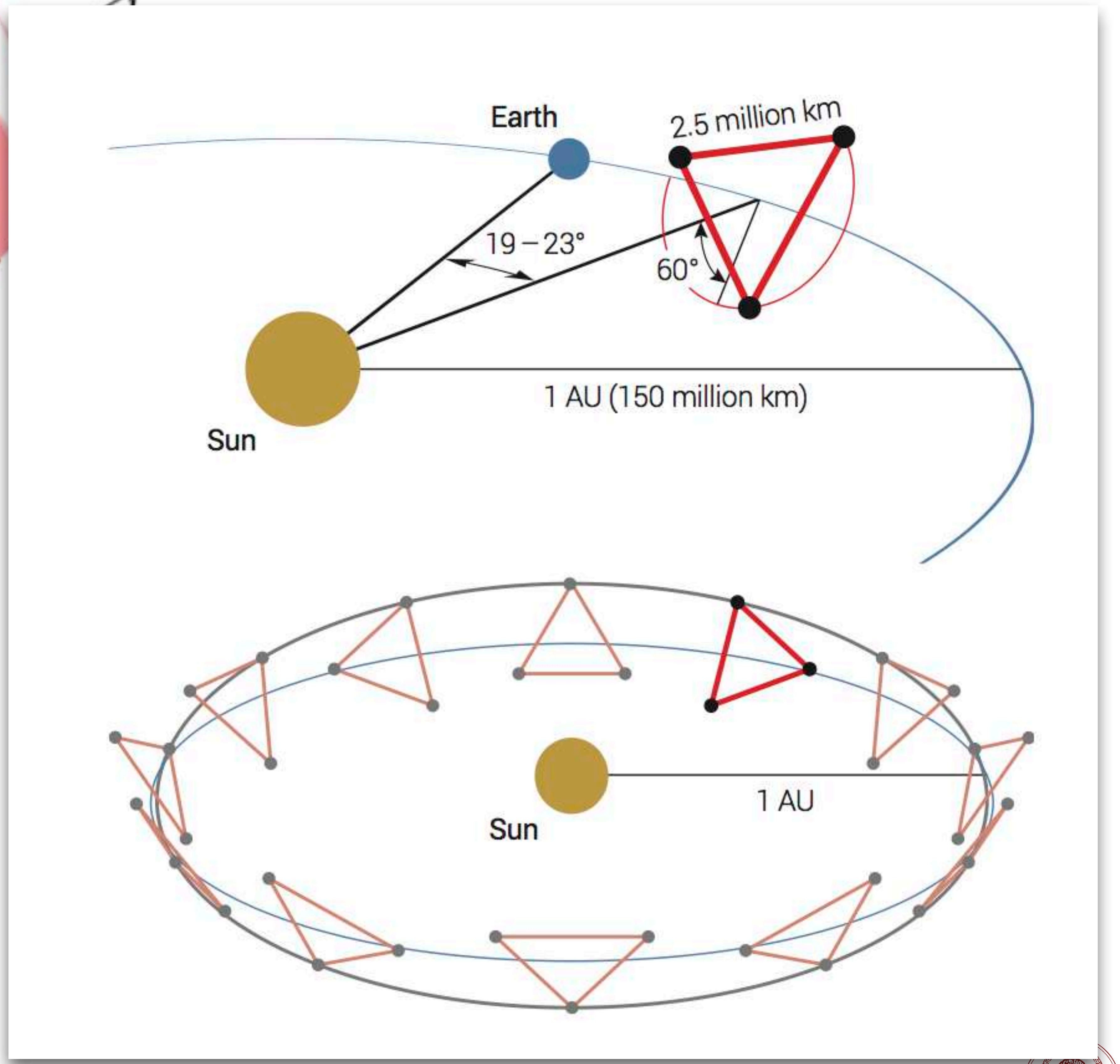
11th Iberian Gravitational Wave Meeting

09 June 2021

- Science with LISA & data analysis challenges
  - Suppose we subtract loud sources, how do we *characterize* astrophysical GW foregrounds?
  - Is there can easy way to tell if a stochastic GW signal is *detectable* or *not*? (remember: no real data yet!)
  - Take first assumption back, go for the real thing. How do we find sources in a *signal dominated observatory*?

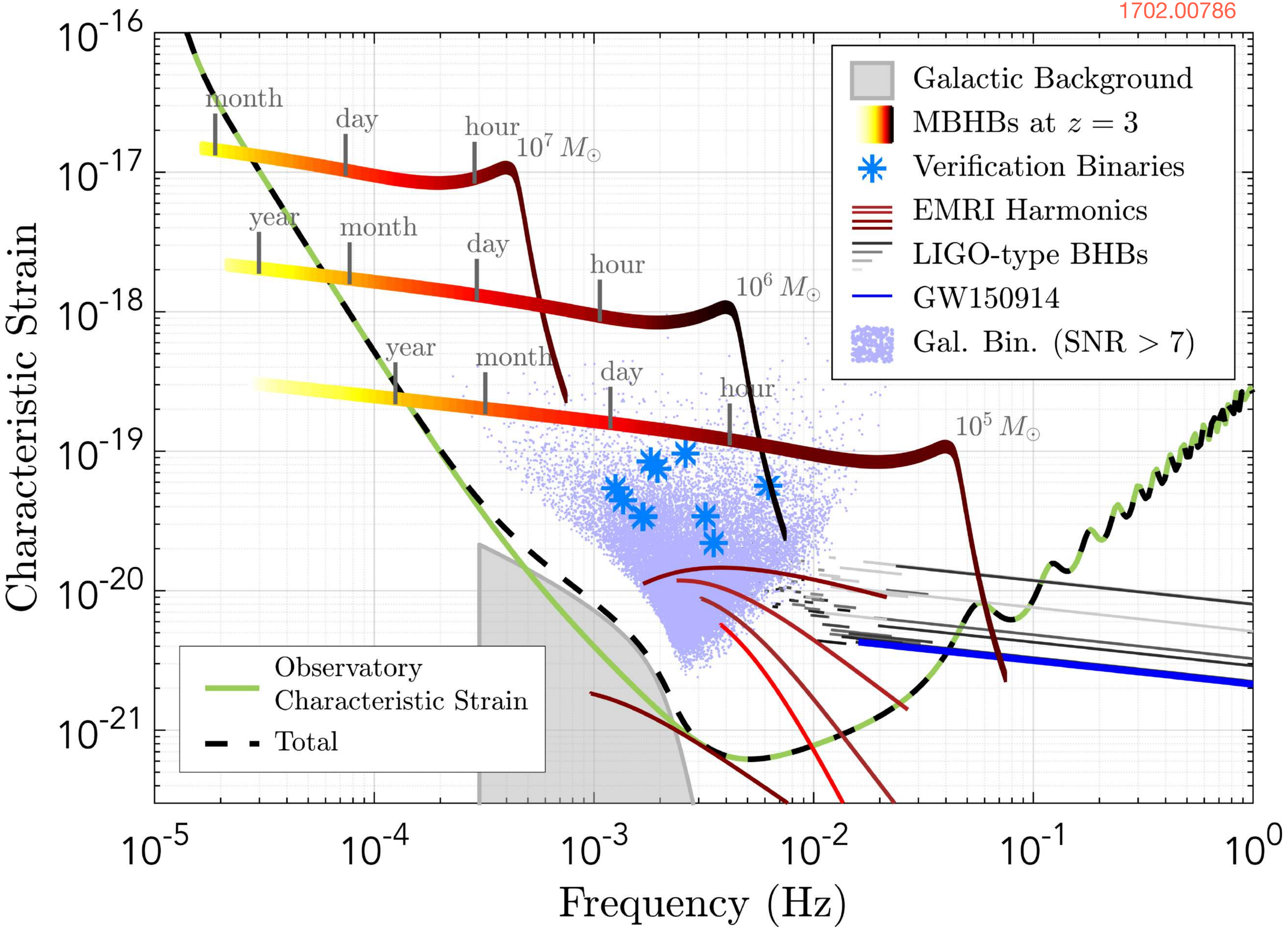

C
B
A
Laser
Proof masses

1702.00786


Earth
2.5 million km
19 – 23°
60°
Sun
1 AU (150 million km)
Sun
1 AU

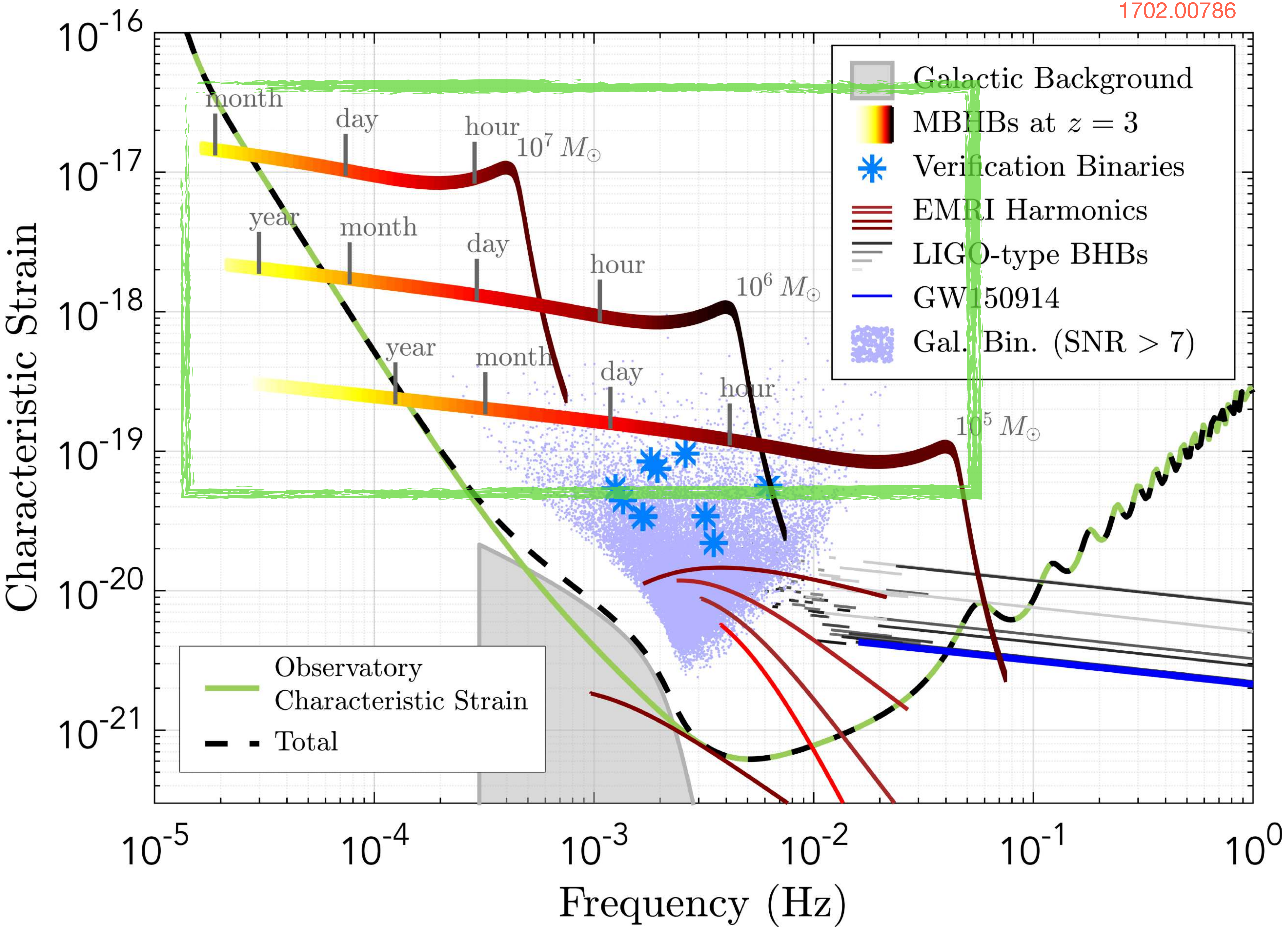
Science with LISA : Extracting the signals

1702.00786



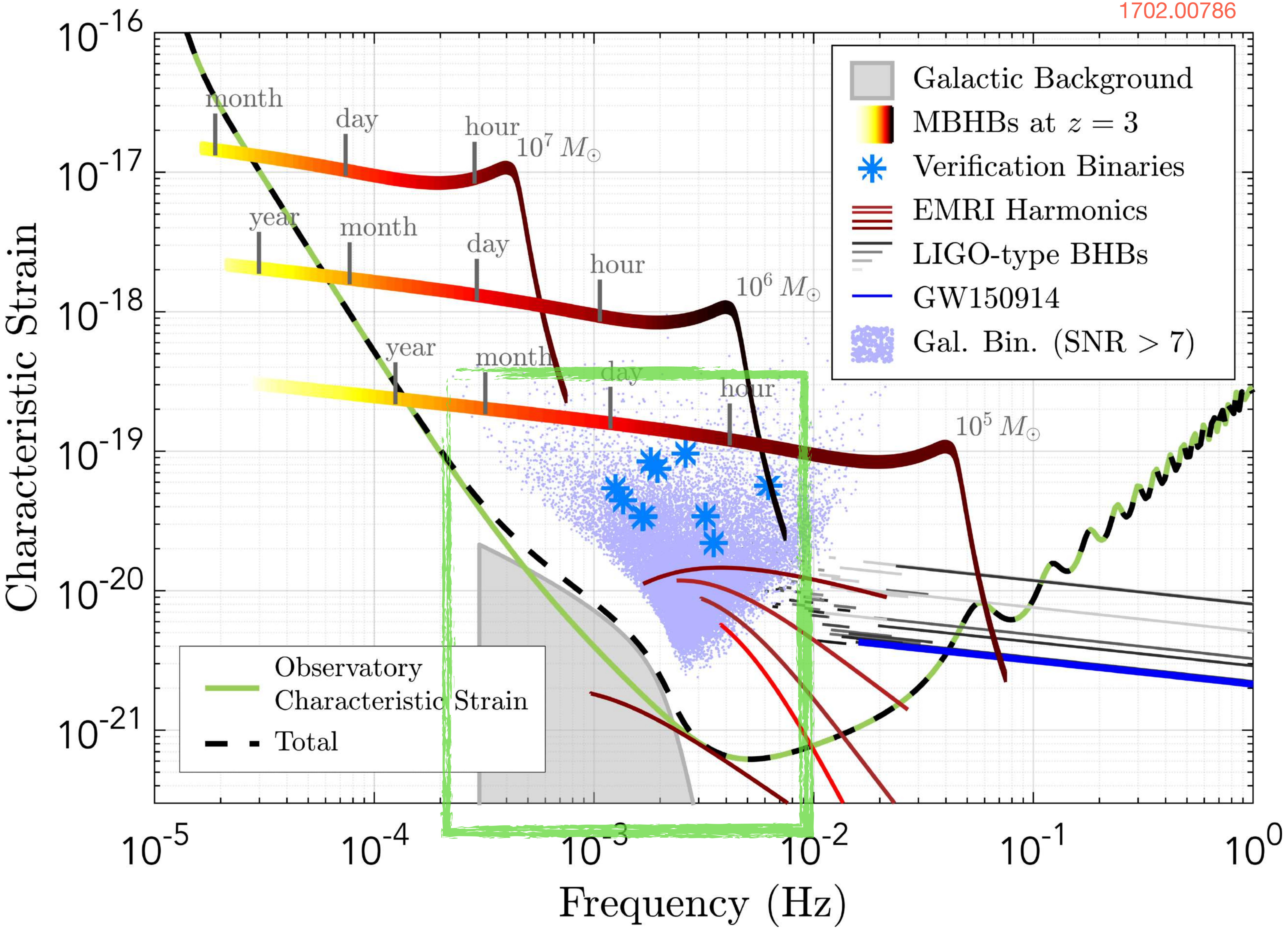
Characteristic Strain
Frequency (Hz)
10^{-16}
10^{-17}
10^{-18}
10^{-19}
10^{-20}
10^{-21}
10^{-5}
10^{-4}
10^{-3}
10^{-2}
10^{-1}
10^{0}
Galactic Background
MBHBs at $z = 3$
Verification Binaries
EMRI Harmonics
LIGO-type BHBs
GW150914
Gal. Bin. (SNR > 7)
Observatory Characteristic Strain
Total
$10^7\ M_\odot$
$10^6\ M_\odot$
$10^5\ M_\odot$
month
day
hour
year
month
day
hour
year
month
day
hour

1702.00786

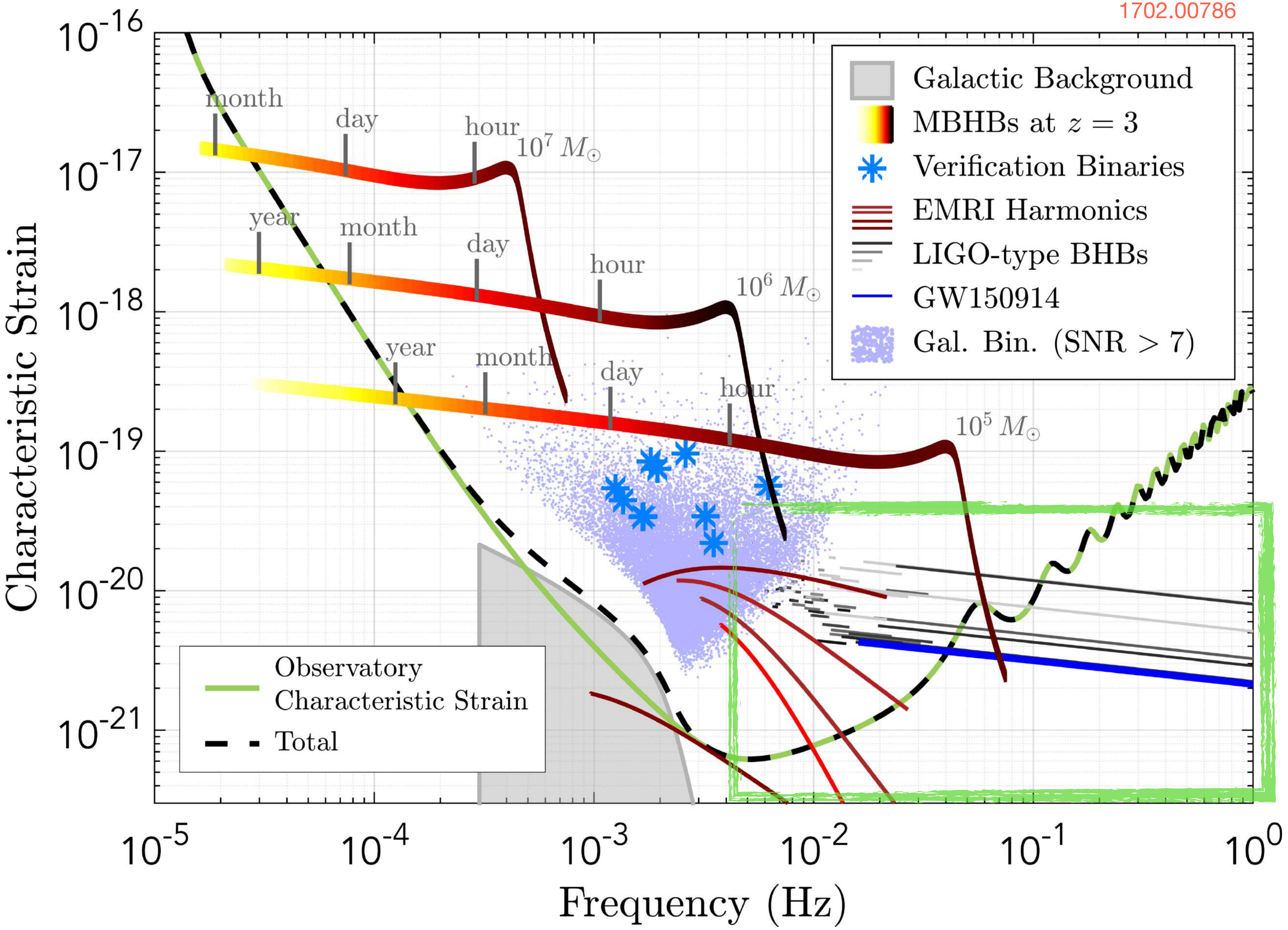


Characteristic Strain
Frequency (Hz)
10^-16
10^-17
10^-18
10^-19
10^-20
10^-21
10^-5
10^-4
10^-3
10^-2
10^-1
10^0
month
day
hour
10^7 M⊙
year
month
day
hour
10^6 M⊙
year
month
day
hour
10^5 M⊙
Galactic Background
MBHBs at z = 3
Verification Binaries
EMRI Harmonics
LIGO-type BHBs
GW150914
Gal. Bin. (SNR > 7)
Observatory Characteristic Strain
Total

1702.00786


Characteristic Strain
Frequency (Hz)
10^{-16}
10^{-17}
10^{-18}
10^{-19}
10^{-20}
10^{-21}
10^{-5}
10^{-4}
10^{-3}
10^{-2}
10^{-1}
10^{0}
month
day
hour
$10^7 M_\odot$
year
month
day
hour
$10^6 M_\odot$
year
month
day
hour
$10^5 M_\odot$
Galactic Background
MBHBs at $z = 3$
Verification Binaries
EMRI Harmonics
LIGO-type BHBs
GW150914
Gal. Bin. (SNR > 7)
Observatory Characteristic Strain
Total

1702.00786

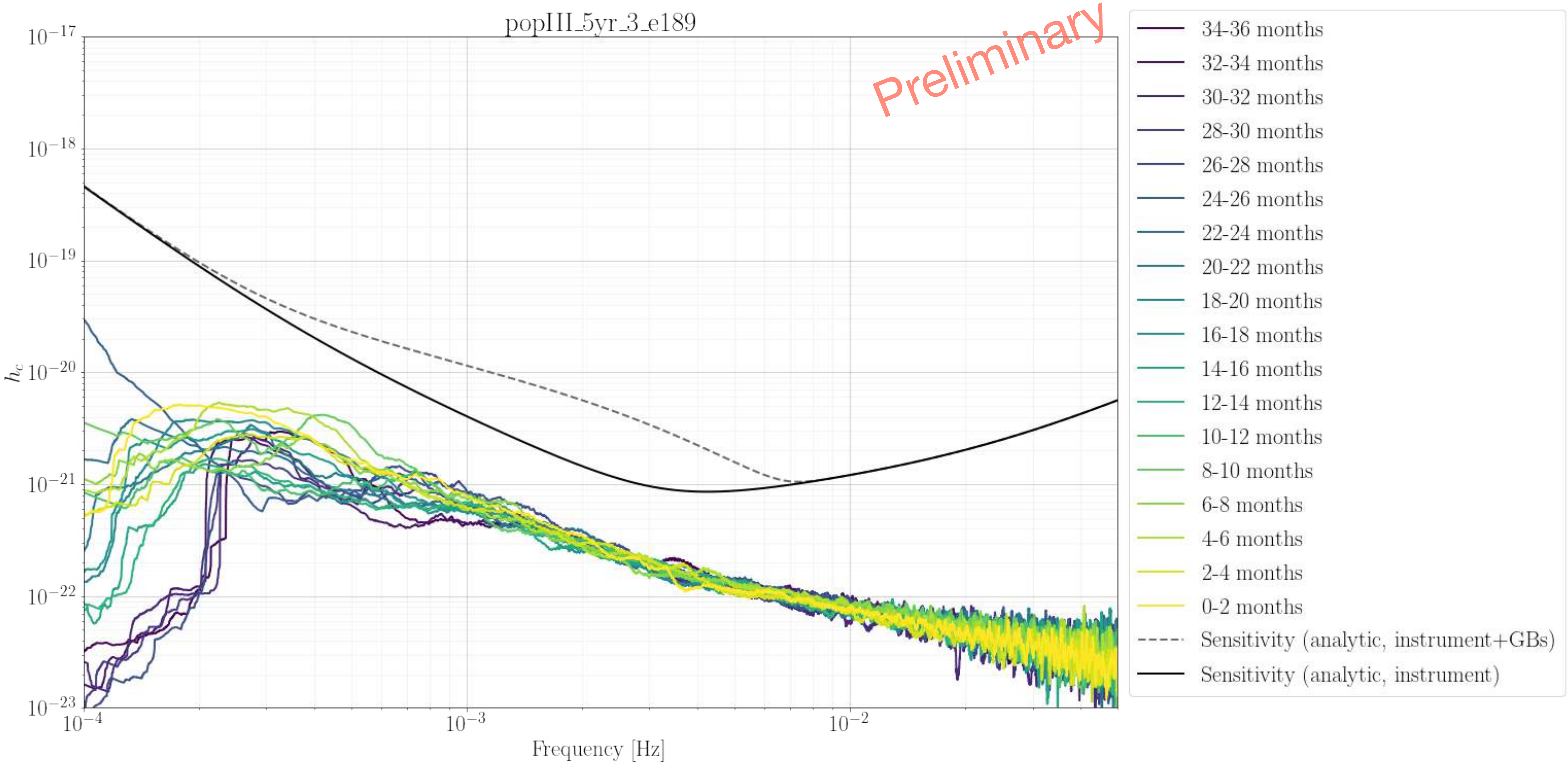

Galactic Background
MBHBs at $z = 3$
Verification Binaries
EMRI Harmonics
LIGO-type BHBs
GW150914
Gal. Bin. (SNR > 7)
Observatory Characteristic Strain
Total
month
day
hour
$10^7 M_\odot$
year
month
day
hour
$10^6 M_\odot$
year
month
day
hour
$10^5 M_\odot$
Characteristic Strain
Frequency (Hz)
10^{-16}
10^{-17}
10^{-18}
10^{-19}
10^{-20}
10^{-21}
10^{-5}
10^{-4}
10^{-3}
10^{-2}
10^{-1}
10^{0}

- Now, let us ignore the loud and relatively short GW signals.
- Suppose we build an effective pipeline that allows us to find them, and subtract them, almost perfectly.
- What is the remaining signal?
    - Astrophysical
    - Cosmological

- Now, let us ignore the loud and relatively short GW signals.
- Suppose we build an effective pipeline that allows us to find them, and subtract them, almost perfectly.
- What is the remaining signal?
    - Astrophysical
    - Cosmological

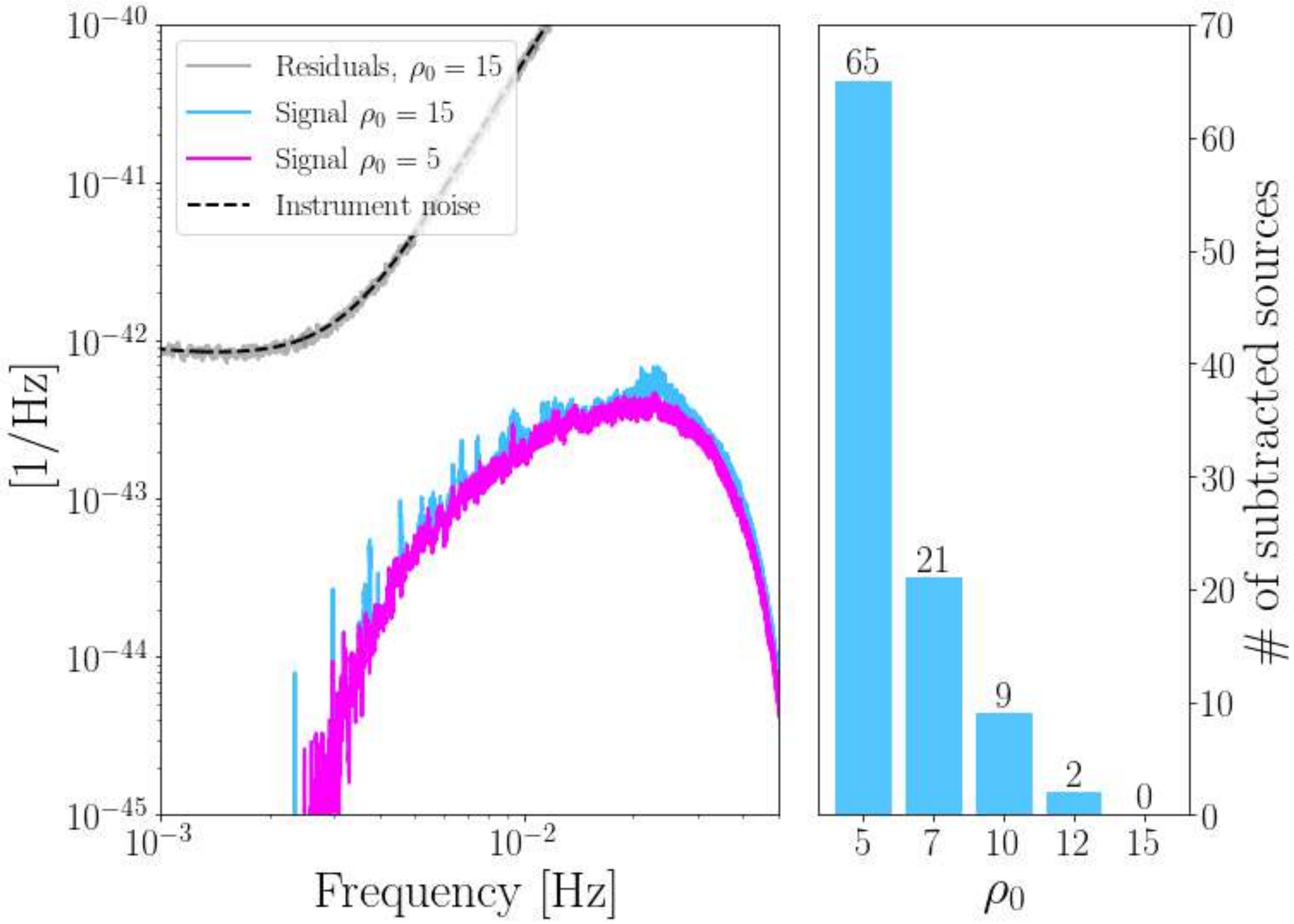
• Black Hole Binaries

- Some seeds might generate a stochastic "pop-corn"-like noisy signal! [D. Langeroodi, NK, *et al*, in preparation]



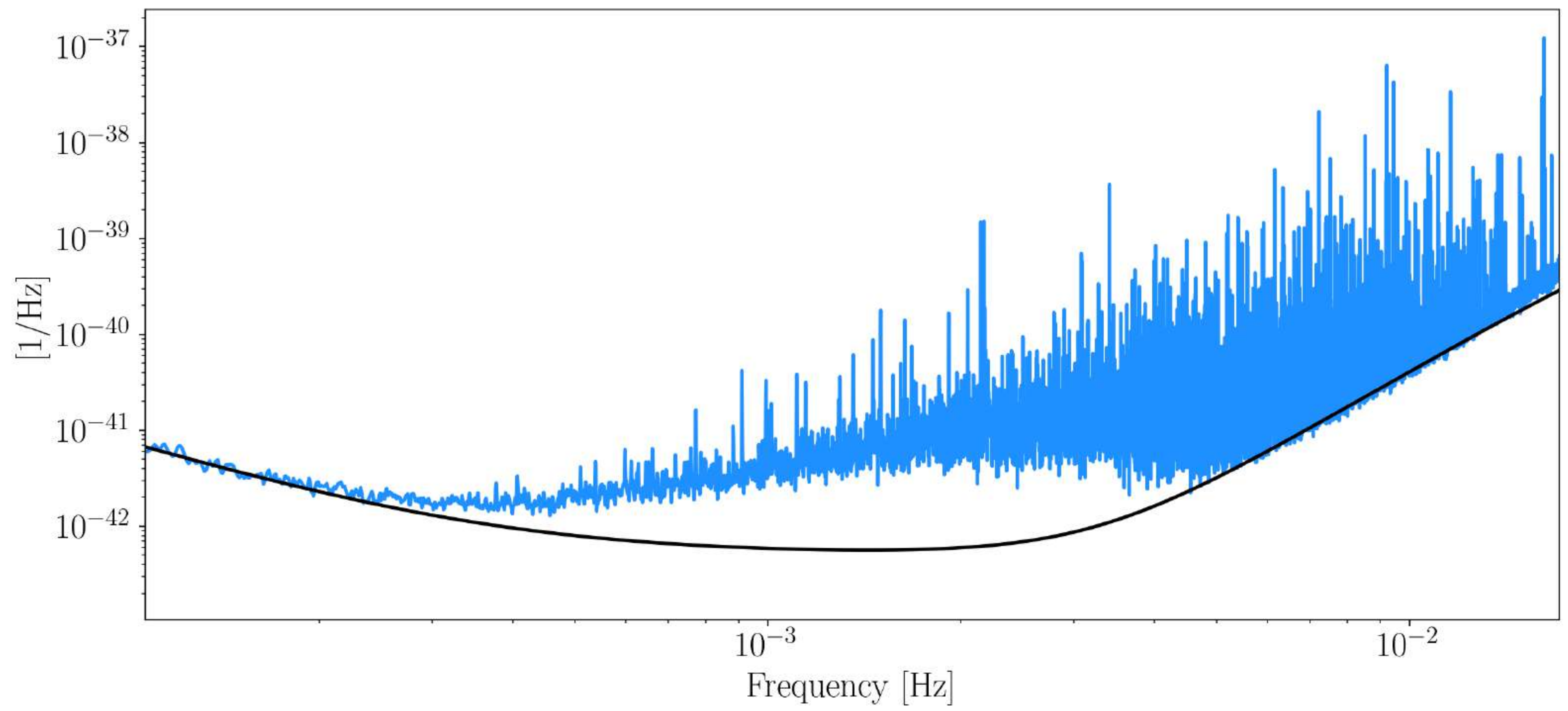
• Black Hole Binaries

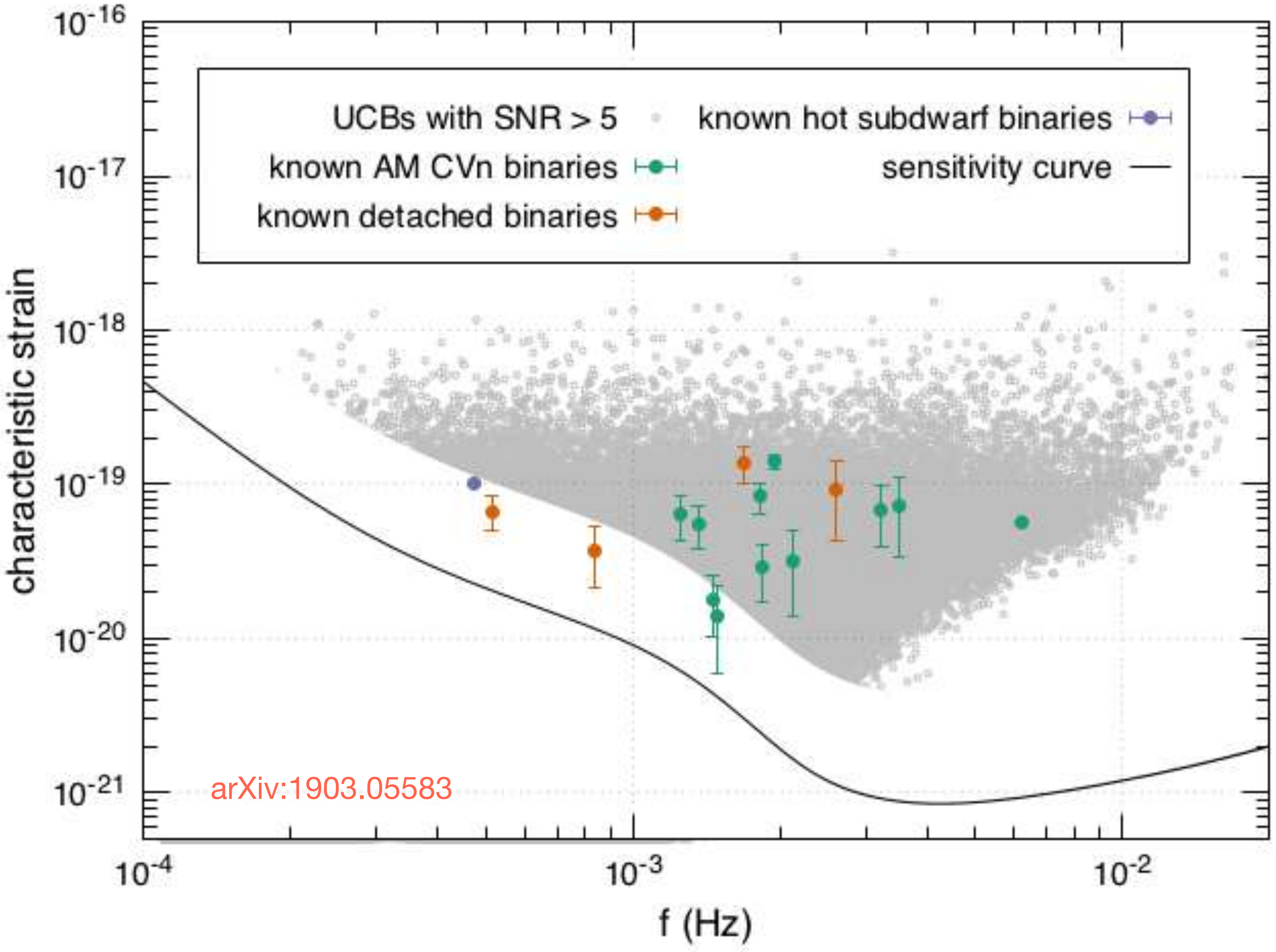
- *Lighter LIGO-like binaries.*
- Stellar Masses.
- Maybe detectable with lower SNR.
- Will generate a stochastic signal in the LISA band.
- Work on an accurate extrapolation of the LIGO population to the LISA band is under study. [P. Marcoccia, et al, in prep.]



• Compact Galactic Binaries

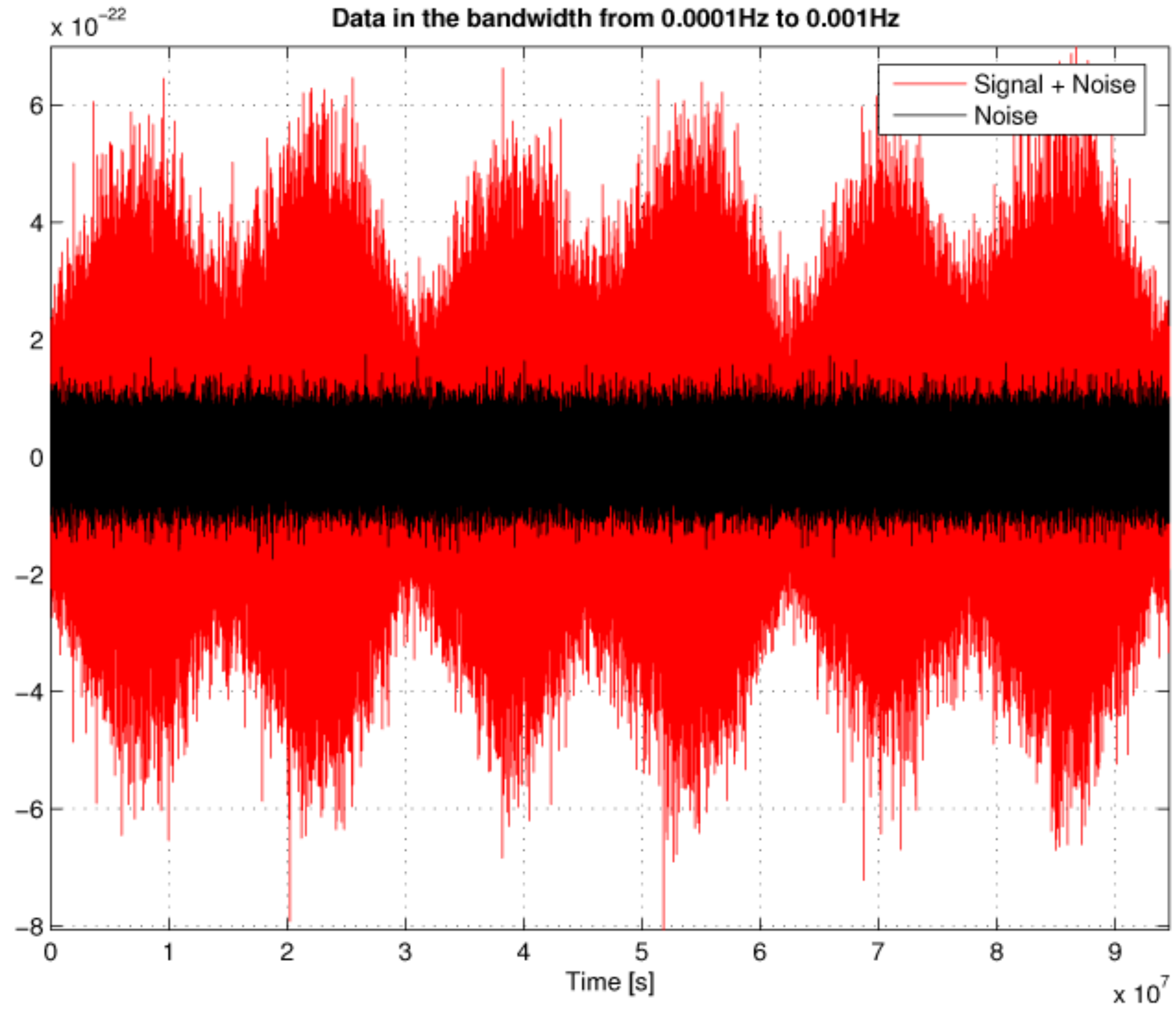
- Ultra-compact (White Dwarf) binaries in our neighborhood.
- Stellar-mass compact objects with orbital periods < 1 h.
- Almost monochromatic.
- Millions of sources measured in the LISA band.
- Guaranteed detection!





• Compact Galactic Binaries

- Ultra-compact (White Dwarf) binaries in our neighborhood.
- Stellar-mass compact objects with orbital periods < 1 h.
- Almost monochromatic.
- Millions of sources measured in the LISA band.
- Guaranteed detection!
- Source of non-stationarity in the data-steam!



PhysRevD.71.122003

See next talk by Ivan M. Vilchez!

- How do we analyze data with such particularities?
  - We need to perform **model selection** and **Parameter Estimation**.
  - For many types and high number of sources we may use stochastic algorithms (MCMC on steroids).
  - **Expensive** (computationally), challenging to **tune**, takes **time**.
  - We will discuss this in a bit.
- But for now, keeping things a little simple, how can we make estimates of the unresolved signal that originates from a population of binaries?

lisa

• Parenthesis: Estimating the total signal from unresolved binaries

- A practical method to get a zero-eth order of the *foreground* signals.
- *Iterative process*, based on more "loose" criteria about the detection of each source, i.e. a SNR limit.
- For example, we define a SNR_0, for which if a given source surpasses it, then we subtract it. Basically *loop over* the known catalogue.

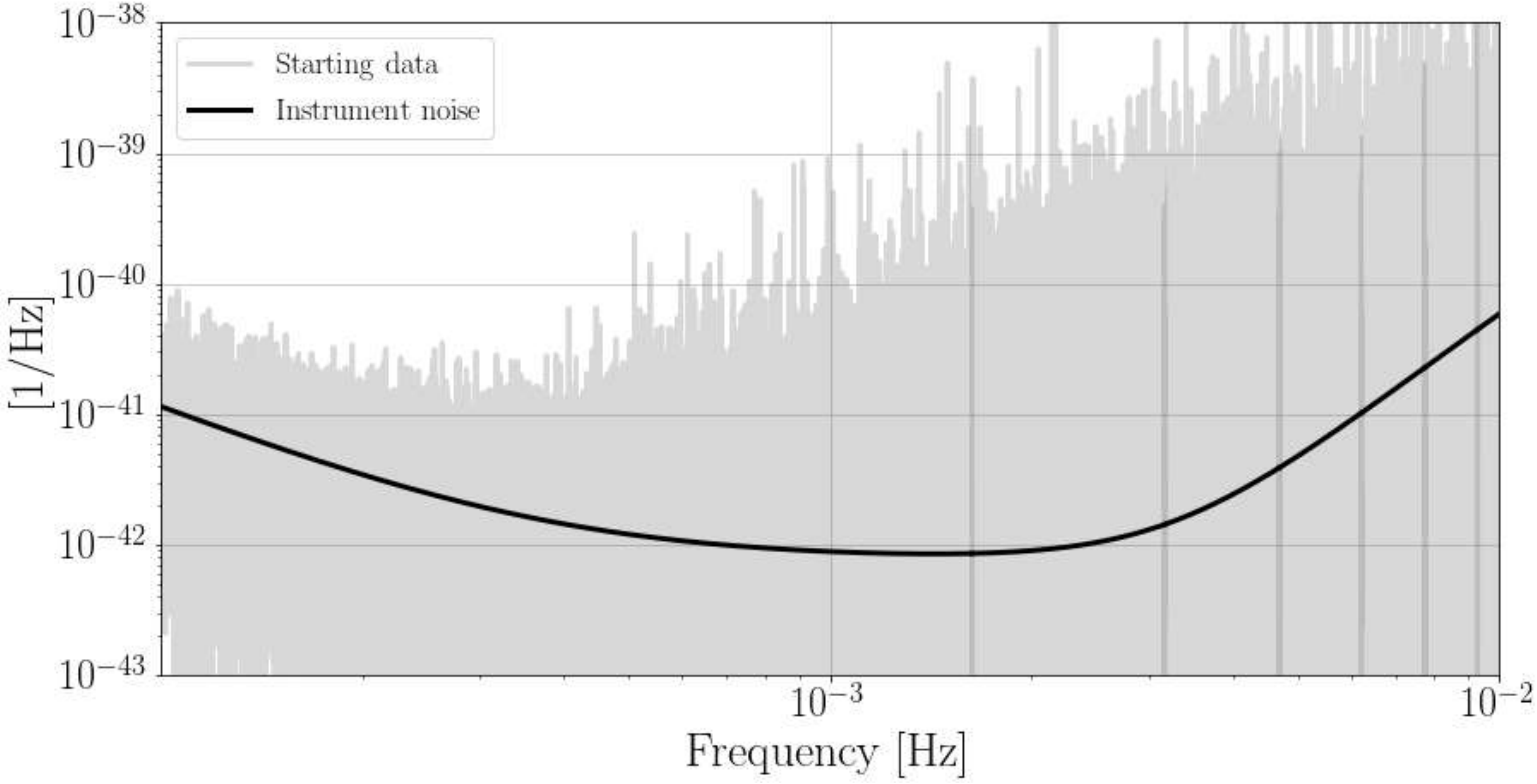
NK et al, arXiv:2103.14598

• Parenthesis: Estimating the total signal from unresolved binaries

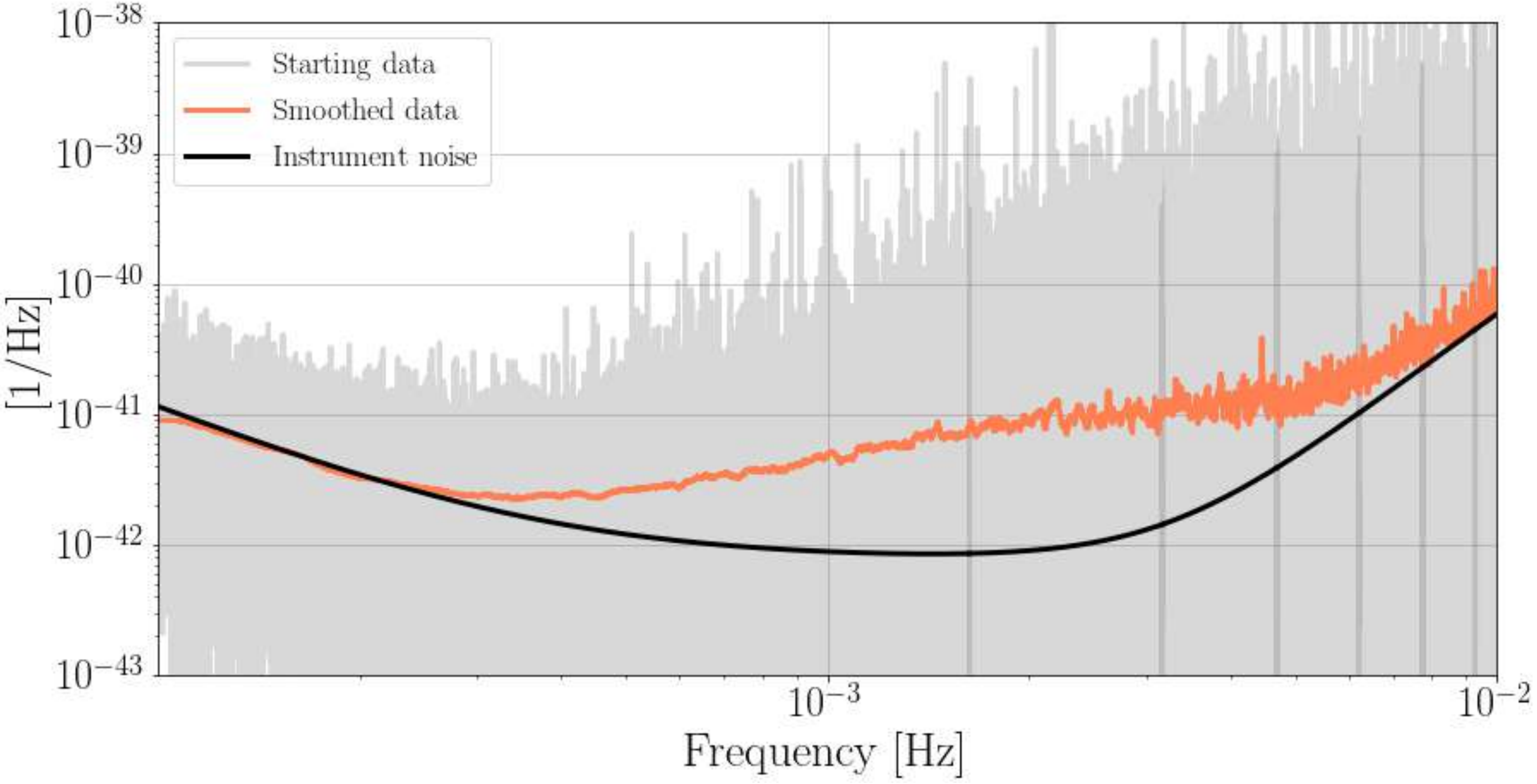
- A practical method to get a zero-eth order of the *foreground* signals.
- *Iterative process*, based on more "loose" criteria about the detection of each source, i.e. a SNR limit.
- For example, we define a SNR_0, for which if a given source surpasses it, then we subtract it. Basically *loop over* the known catalogue.
  - Fast
  - Generic
  - Idealized: no source overlap problem.
  - Idealized: perfect subtraction == perfect residuals.
  - Idealized: Noise.

NK et al, arXiv:2103.14598

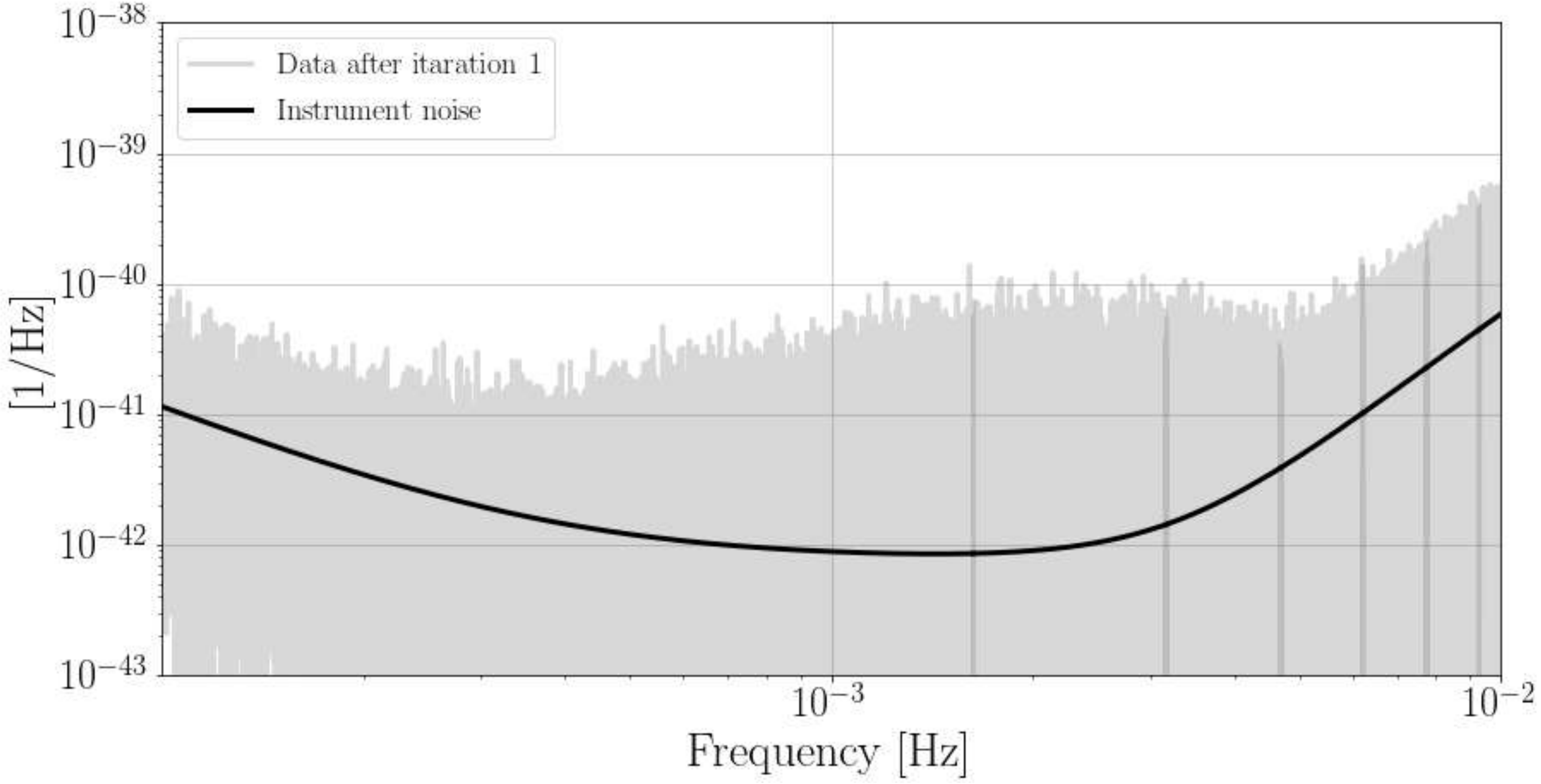
- In more details:



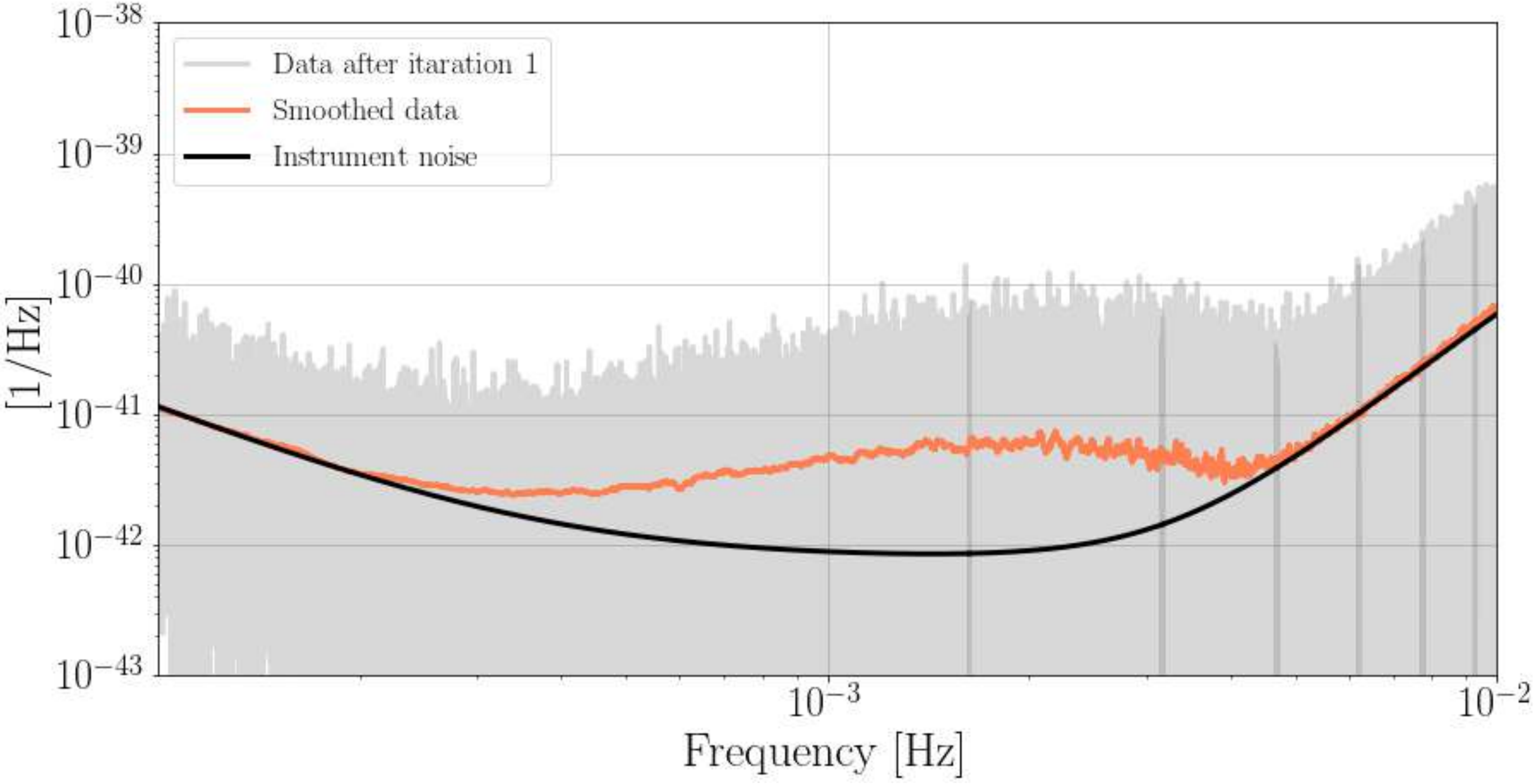
- In more details:



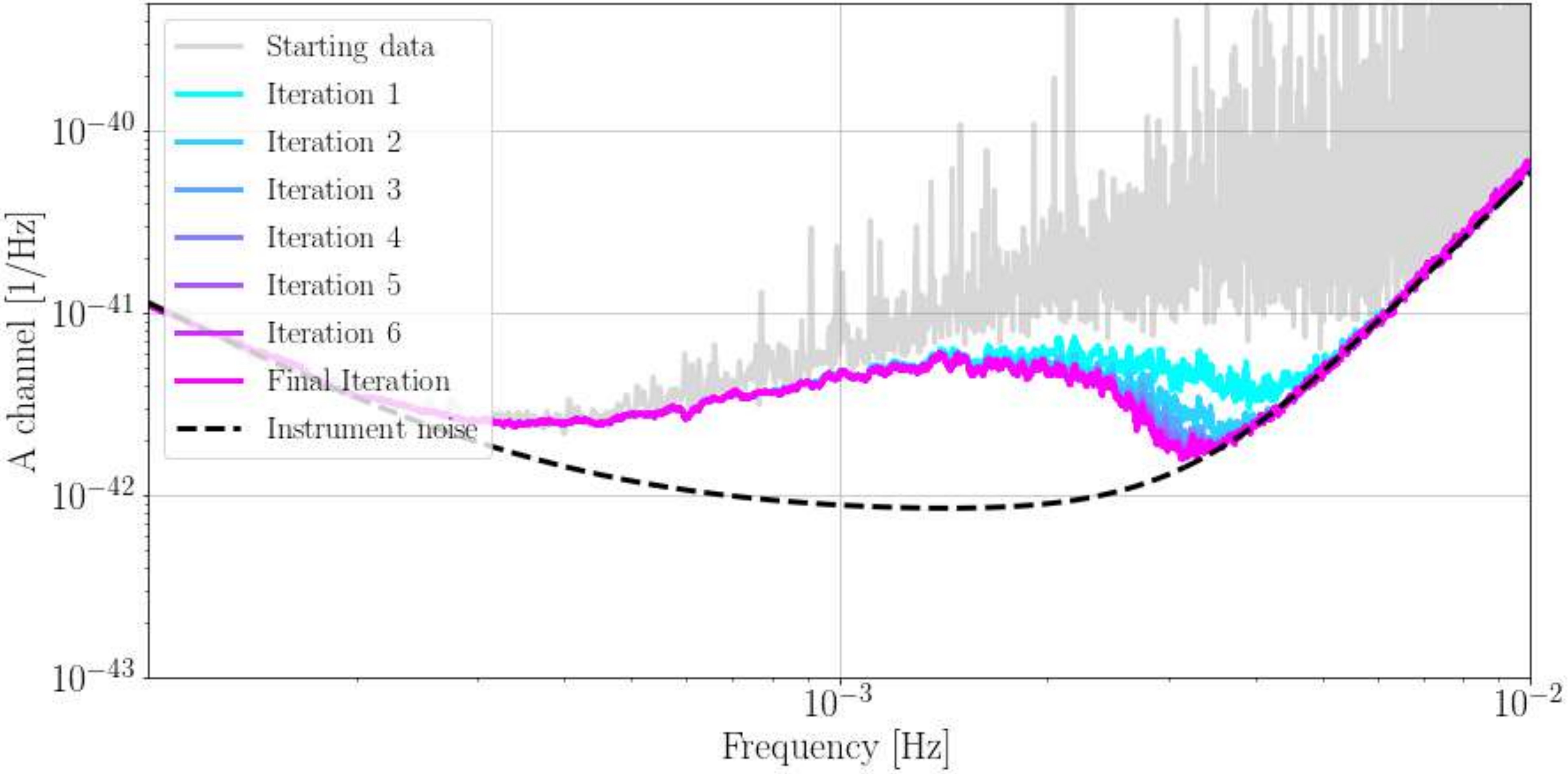
- In more details:



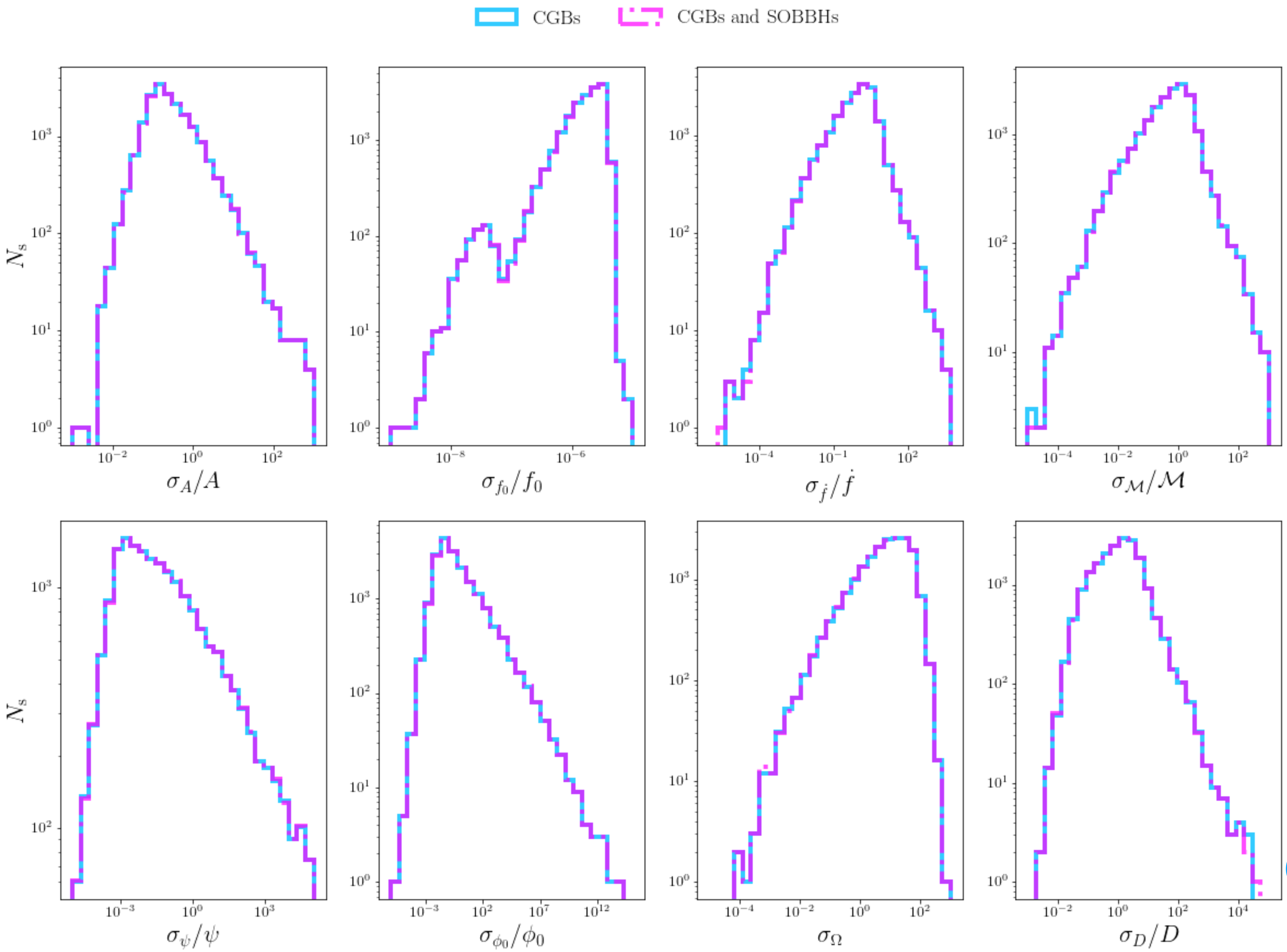
• In more details:

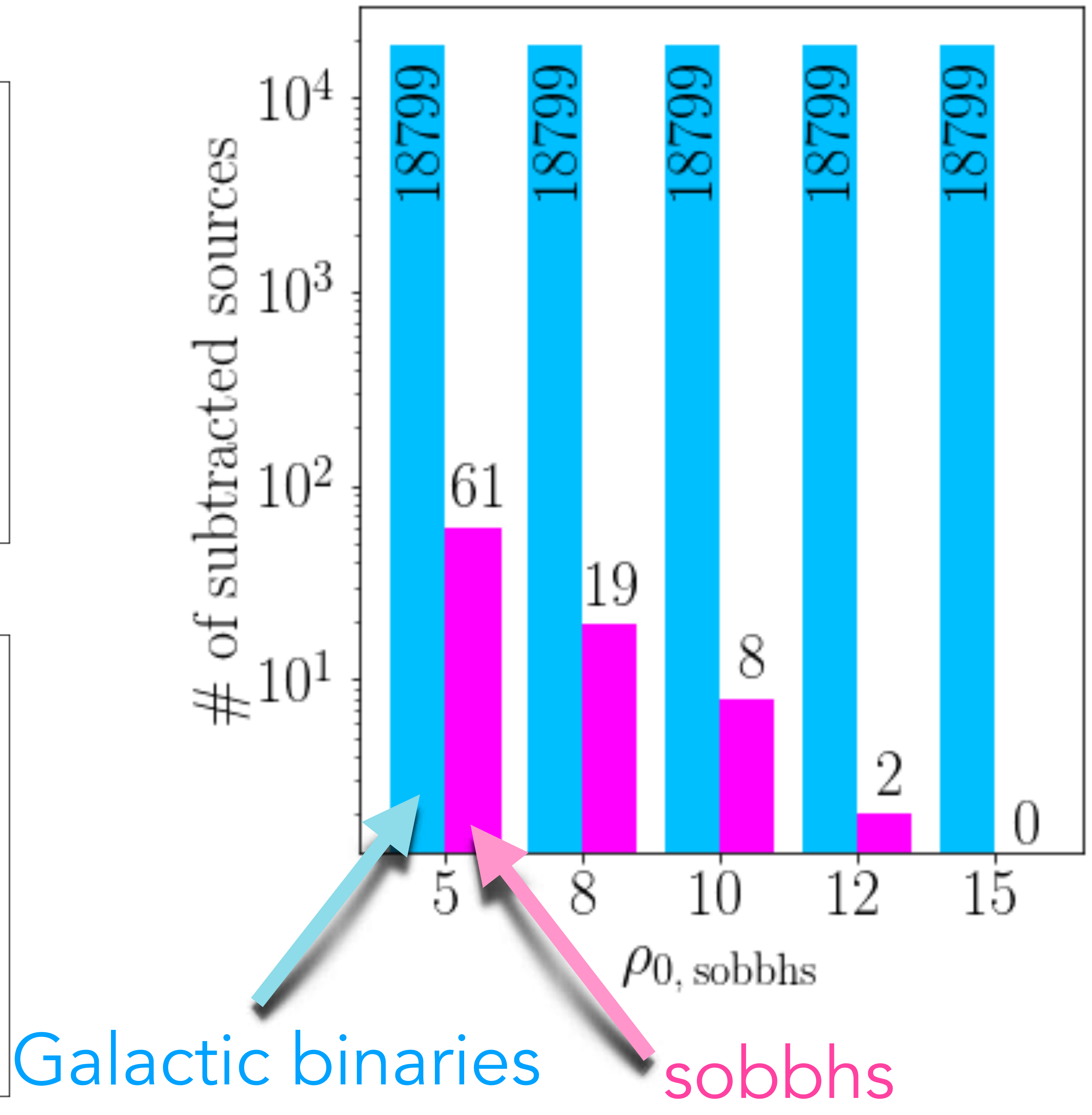


- In more details:



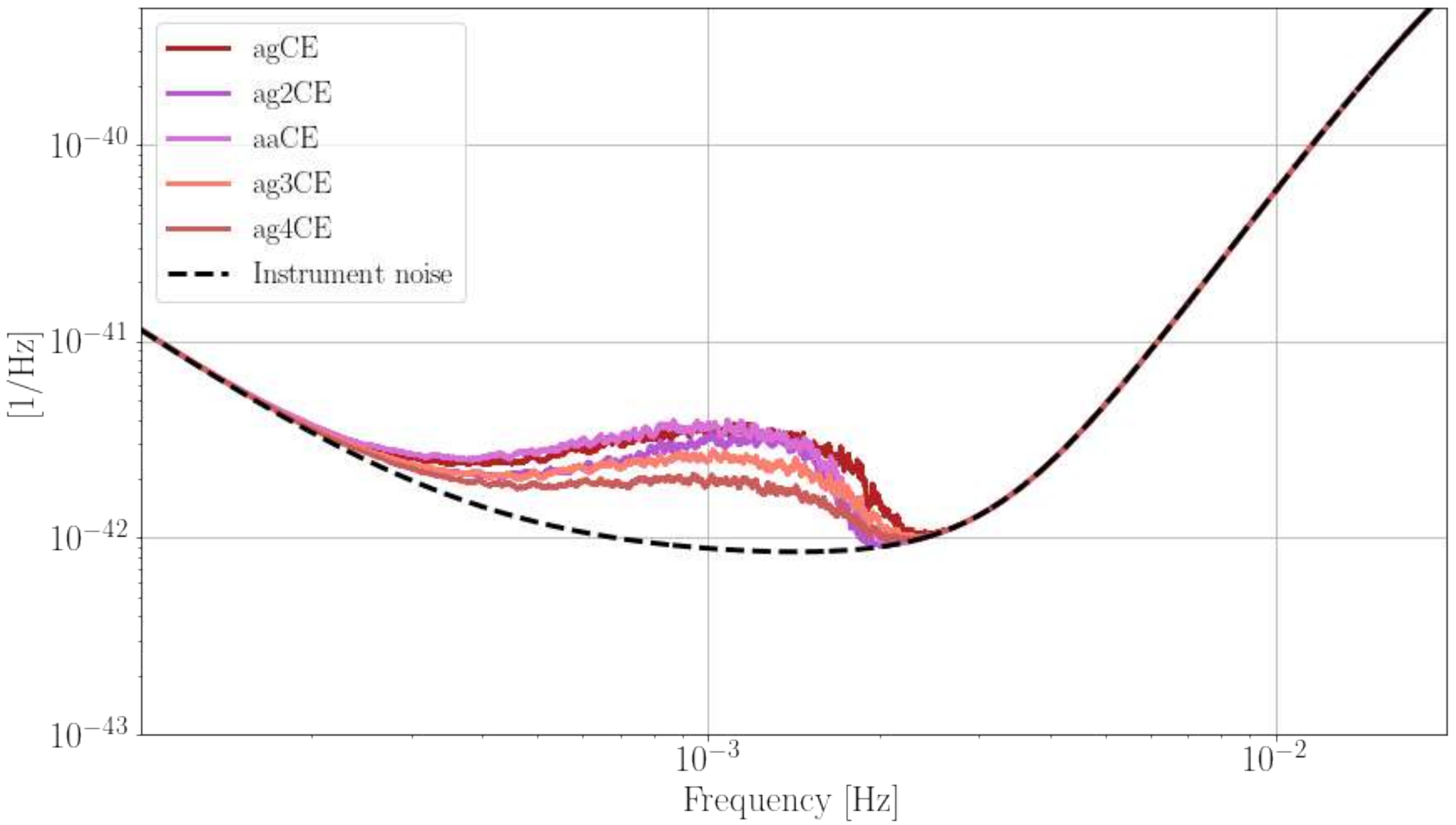
- Combining *more than one* binary population!
  - This method is generic enough to allow us to combine **any given population of sources**, of any type.

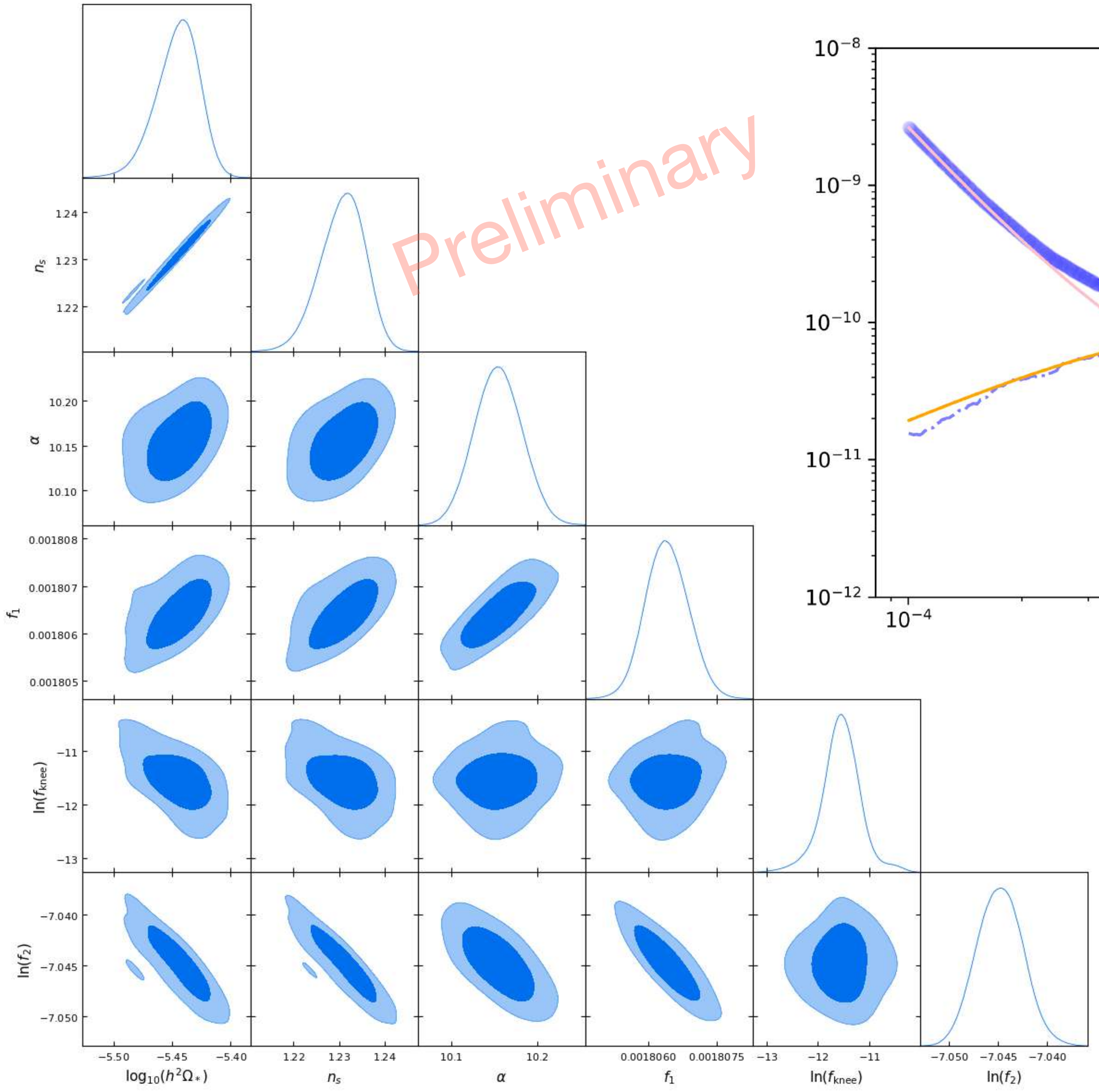


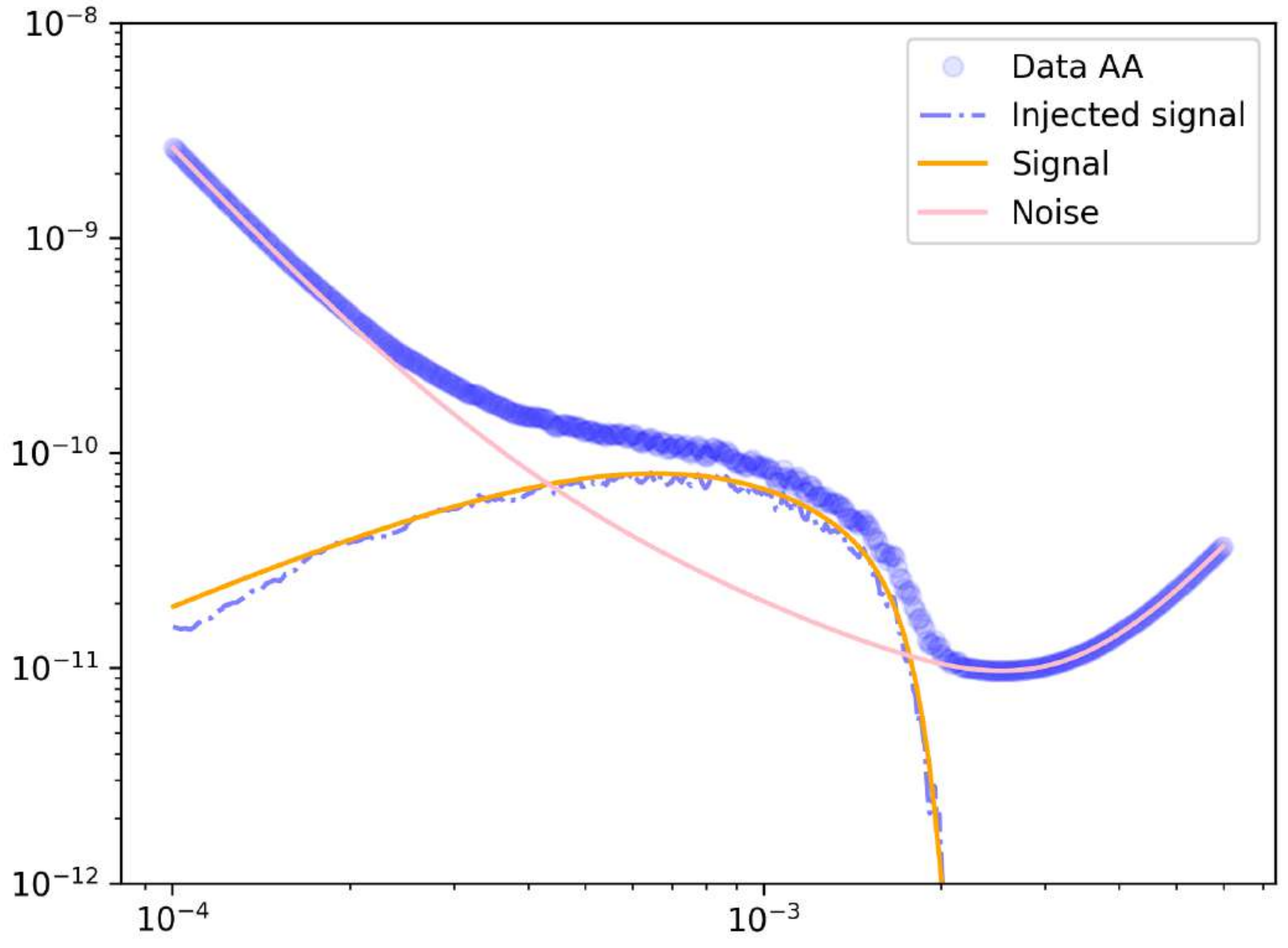


• Example of application

- M Georgousi, NK, V. Korol and M. Pieroni.
- Study the Galaxy properties by investigating the unresolved signal properties as measured by LISA.



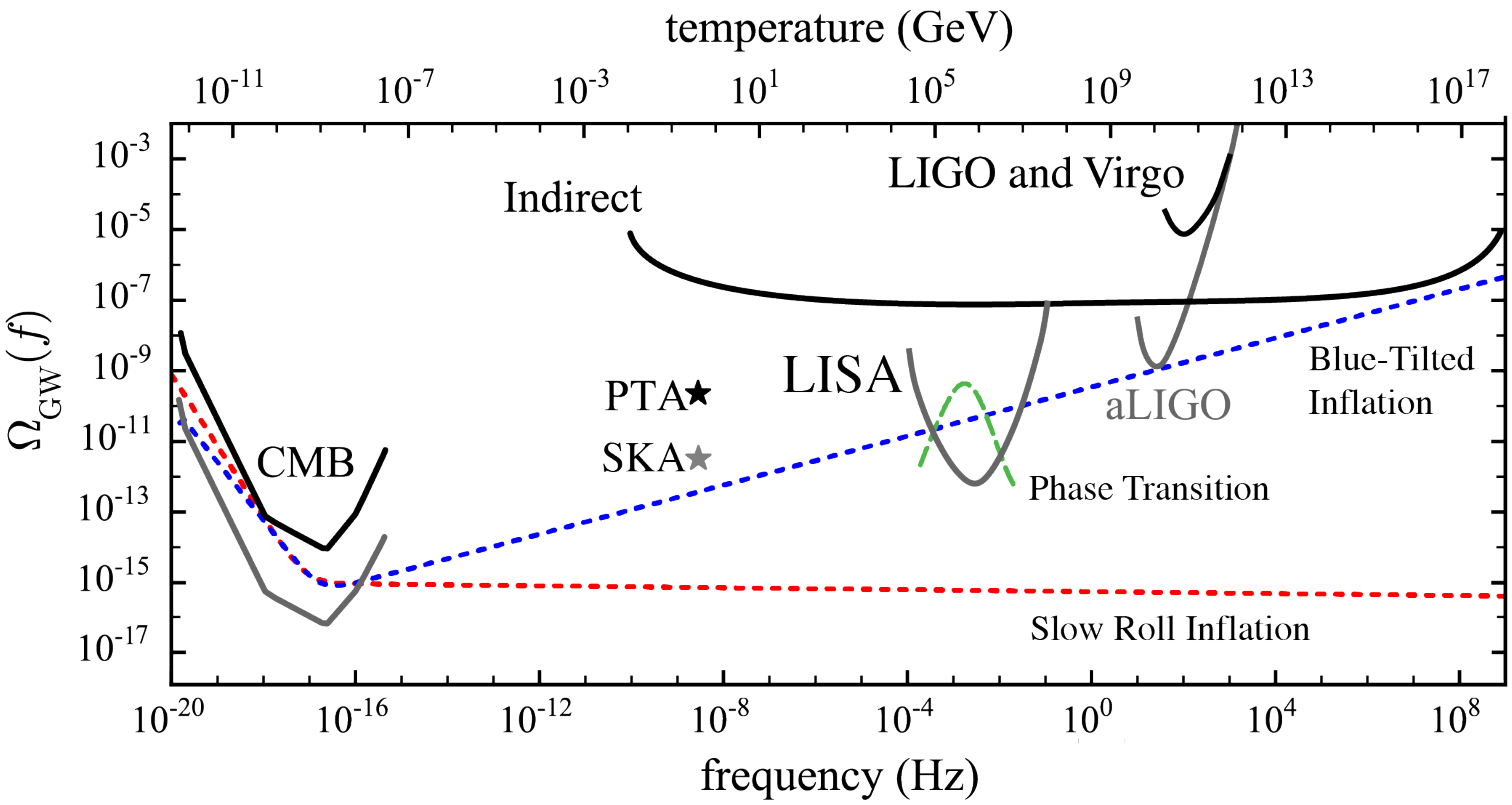




Toonen et al. 2012, 2017;
Korol et al. 2017
Korol et al. (inc.) NK 2021 *in prep*
NK et al, arXiv:2103.14598

- Now, let us ignore the loud and relatively short GW signals.
- Suppose we build an effective pipeline that allows us to find them, and subtract them, almost perfectly.
- What is the remaining signal?
    - Astrophysical
    - Cosmological

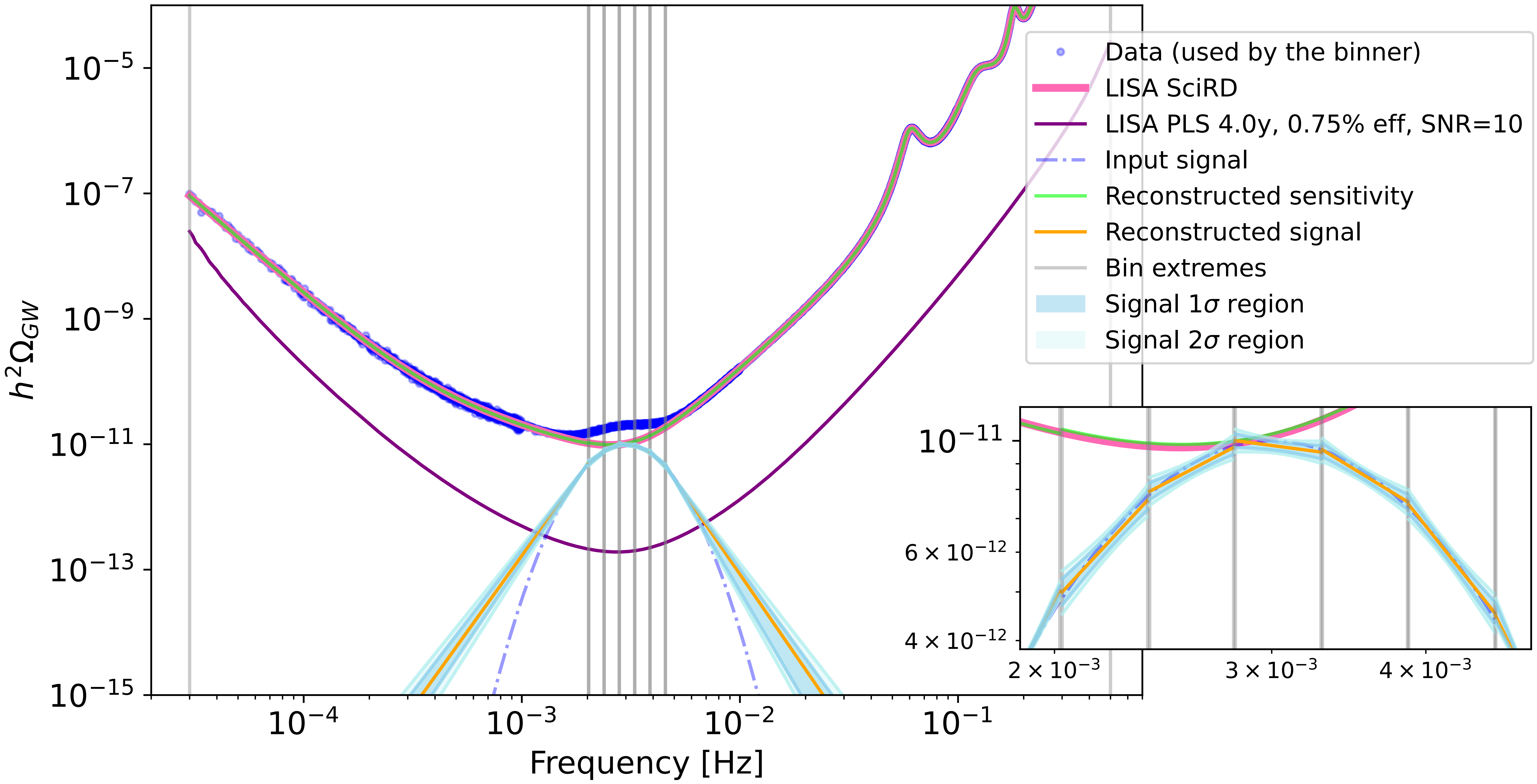
• Stochastic signals from cosmological sources



- Different mechanisms of the early Universe could produce a stochastic GW background measured by LISA.
- We need to extract it from the data - dig in the residuals!

• Stochastic signals from cosmological sources

- Again, take a step back, and ask:
  - If we assume we subtract loud sources successfully,
  - how good is LISA in detecting weak stochastic GW signals (both cosmo + astro)?



R Flauger et al, arXiv: 2009.11845

• Stochastic signals from cosmological sources

- Again, take a step back, and ask:
  - If we assume we subtract loud sources successfully,
  - how good is LISA in detecting weak stochastic GW signals (both cosmo + astro)?

- Normally, one would have to run simulations with different noise realizations, different signals, or perform a Fisher Matrix analysis …
- But there is another way to go forward, assuming again idealized conditions.

• Stochastic signals from cosmological sources

- Again, take a step back, and ask:
  - If we assume we subtract loud sources successfully,
  - how good is LISA in detecting weak stochastic GW signals (both cosmo + astro)?
- We can make analytical predictions, if we start with computing the power spectra

O M Solomon Jr, 10.2172/ 5688766

$$p(S_\mathrm{d}|S_\mathrm{t}) = \prod_i \frac{1}{S_\mathrm{t}[i]} \exp\left(-\frac{S_\mathrm{d}[i]}{S_\mathrm{t}[i]}\right)$$

- Where S_t is the theoretical power spectrum we are interested in. Then if we assume

$$S_\mathrm{t}[i] = S_\mathrm{o}[i] + S_\mathrm{n}[i]$$

- and that we have a prior knowledge of S_n around ε, we can try to marginalize it out, and then

$$p(S_\mathrm{o}|\overline{D}, S_\mathrm{n}) = \frac{1}{(\epsilon^+ + \epsilon^-)} \int_{\bar{S}_\mathrm{n}-\epsilon^-}^{\bar{S}_\mathrm{n}+\epsilon^+} \frac{e^{-N\frac{\overline{D}}{S_\mathrm{o}+S_\mathrm{n}}}}{(S_\mathrm{o}+S_\mathrm{n})^N} \mathrm{d}S_\mathrm{n}$$

NK et al, arXiv:1906.09027

• Stochastic signals from cosmological sources

- In a Bayesian framework we proceed by calculating the Bayes Factor between the
  - M_1: Instrumental noise + SGWB signal
  - M_0: Instrumental noise only

$$\mathcal{B}_{10}(\epsilon) = \frac{P(\overline{D}|\mathcal{M}_1)}{P(\overline{D}|\mathcal{M}_0)}.$$

- Since we have nice closed forms of the posteriors, we marginalize so that:

$$\mathcal{B}_{10}(\epsilon) = \frac{\overline{D}N\left(\Gamma^{\alpha^-} - \Gamma^{\beta^-} - \Gamma^{\alpha^+} + \Gamma^{\beta^+}\right)}{\kappa(N-2)\left(\Gamma^{\alpha^-} - \Gamma^{\alpha^+}\right)} + \frac{\left(\beta^-\Gamma^{\beta^-} - \alpha^-\Gamma^{\alpha^-} + \alpha^+\Gamma^{\alpha^+} - \beta^+\Gamma^{\beta^+}\right)}{\kappa\left(\Gamma^{\alpha^-} - \Gamma^{\alpha^+}\right)}$$

with

$$\alpha^\pm = \overline{S}_\mathrm{n} \pm \epsilon^\pm \text{ and } \beta^\pm = \overline{S}_\mathrm{n} \pm \epsilon^\pm + \kappa$$

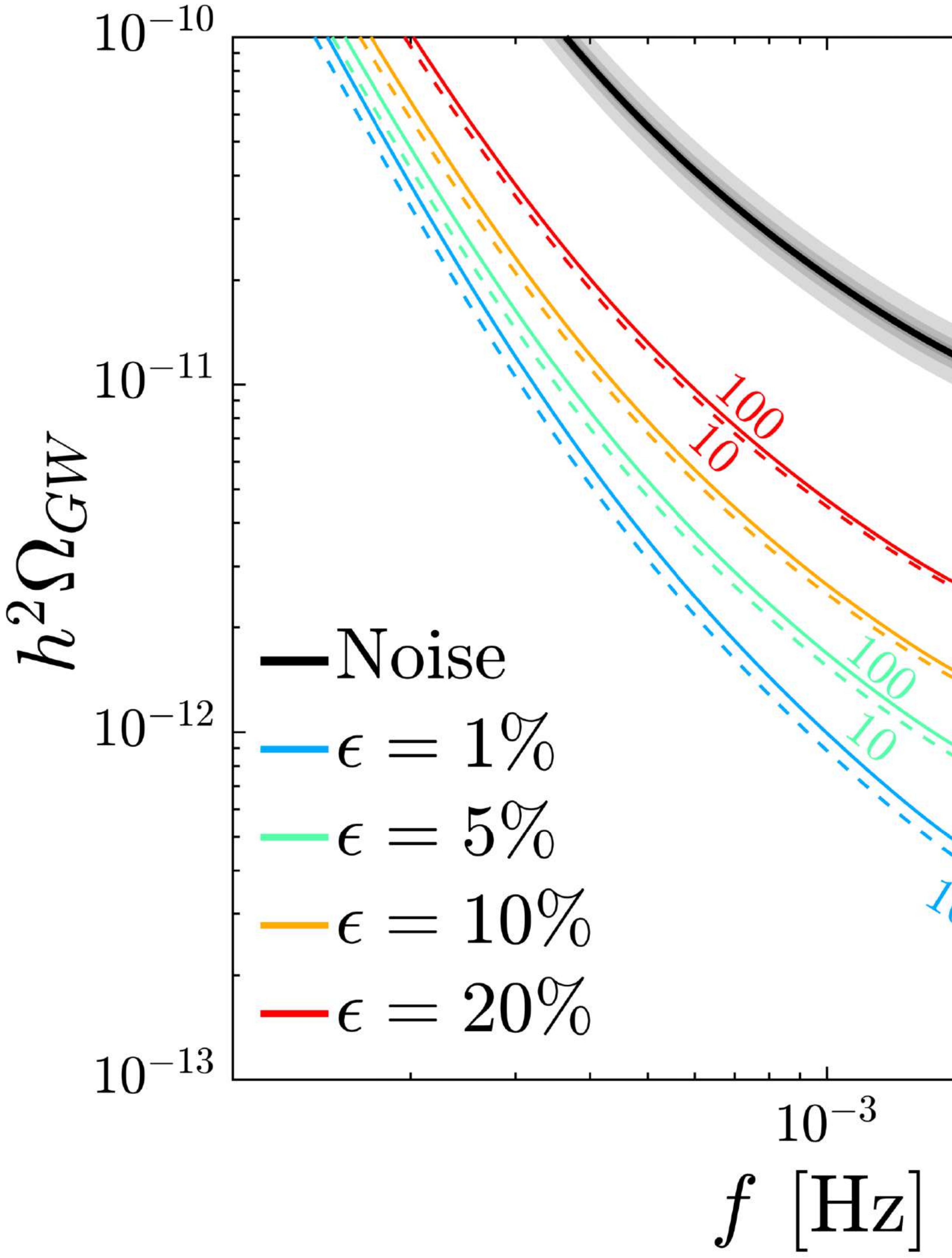
$$\Gamma^{\alpha^\pm} = \Gamma_{N-2}\left(N\overline{D}/\alpha^\pm\right)$$
$$\Gamma^{\beta^\pm} = \Gamma_{N-2}\left(N\overline{D}/\beta^\pm\right)$$

- Get BF(f) for a given spectrum model, without carrying about shapes, MCs, etc!

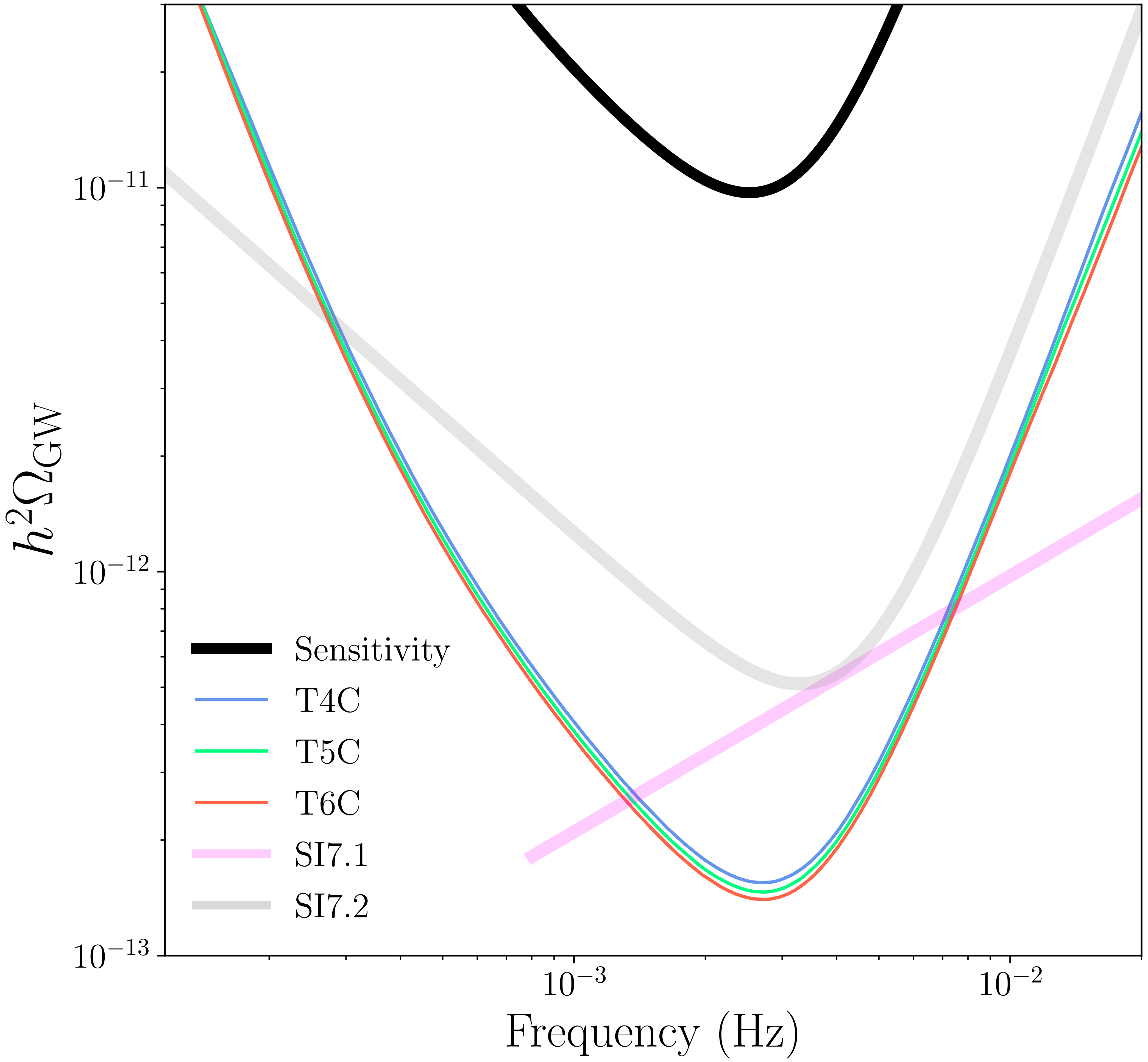
NK et al, arXiv:1906.09027

• Stochastic signals from cosmological sources

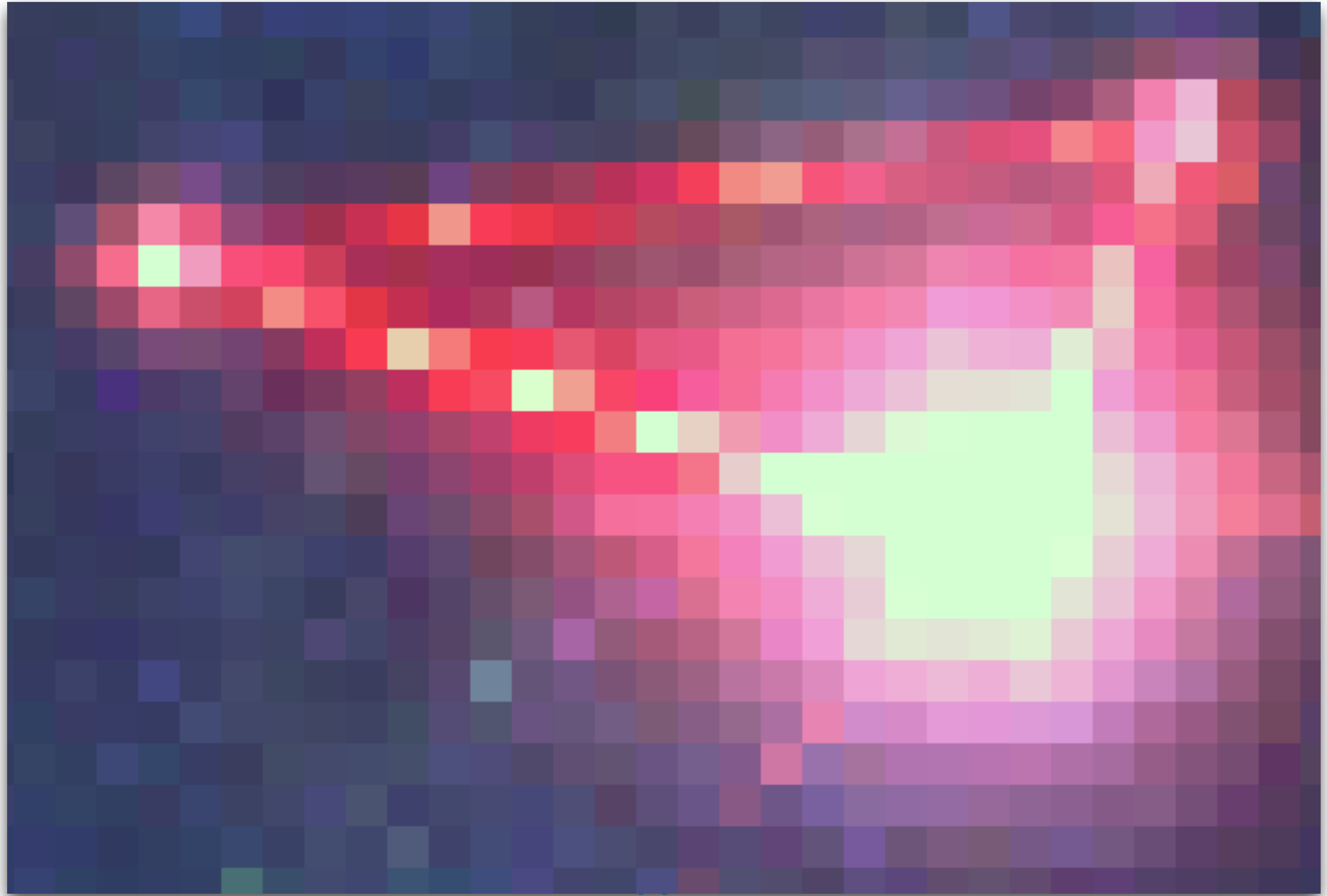
Then, what is left is to plot for example the BF > 100 (strong evidence==detection) for different values of ε.



Or plot signals with BF > 100 for different mission durations. Study was done for the LSG to study effects of duty. cycle to science.



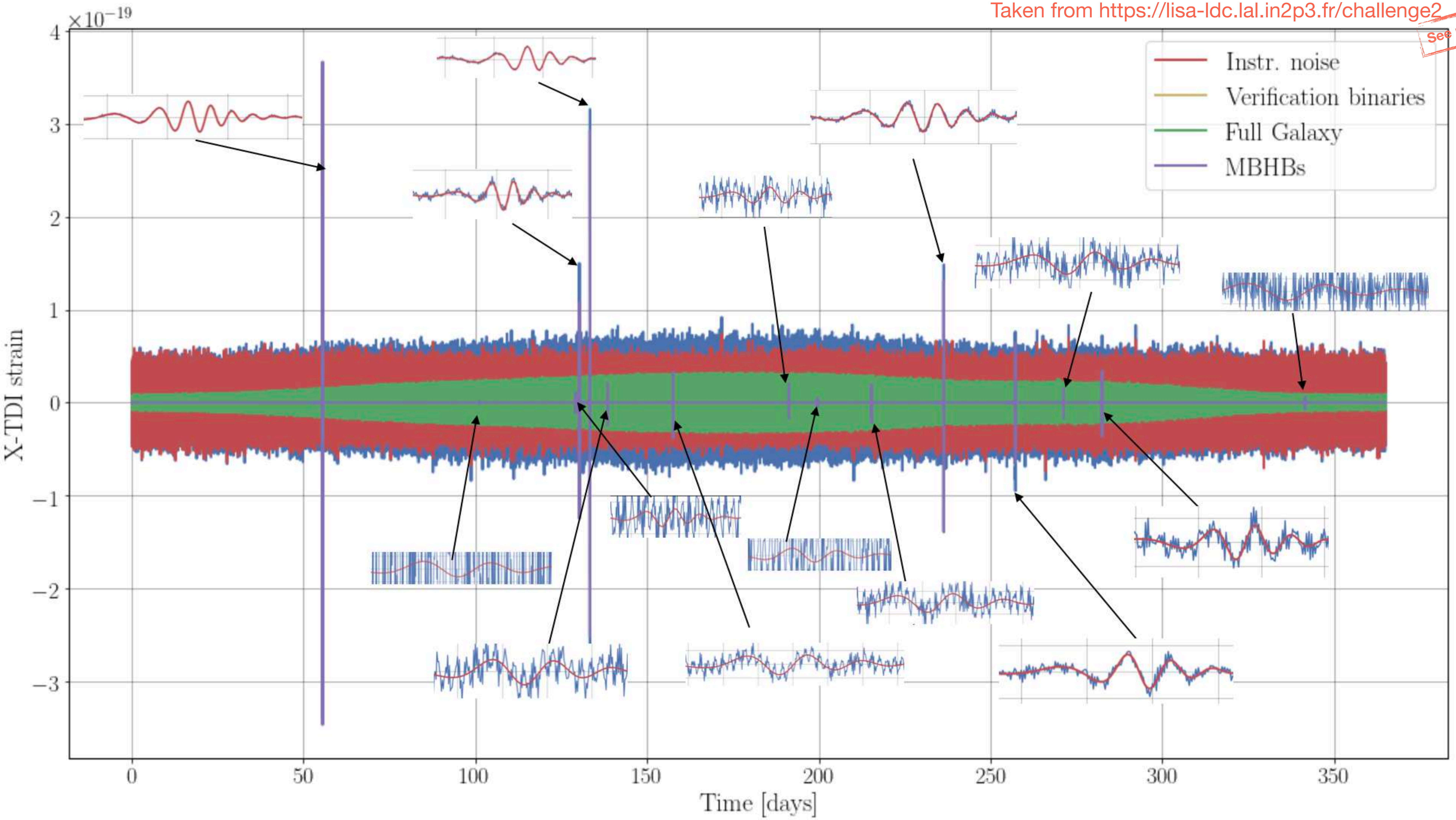
NK et al, arXiv:1906.09027



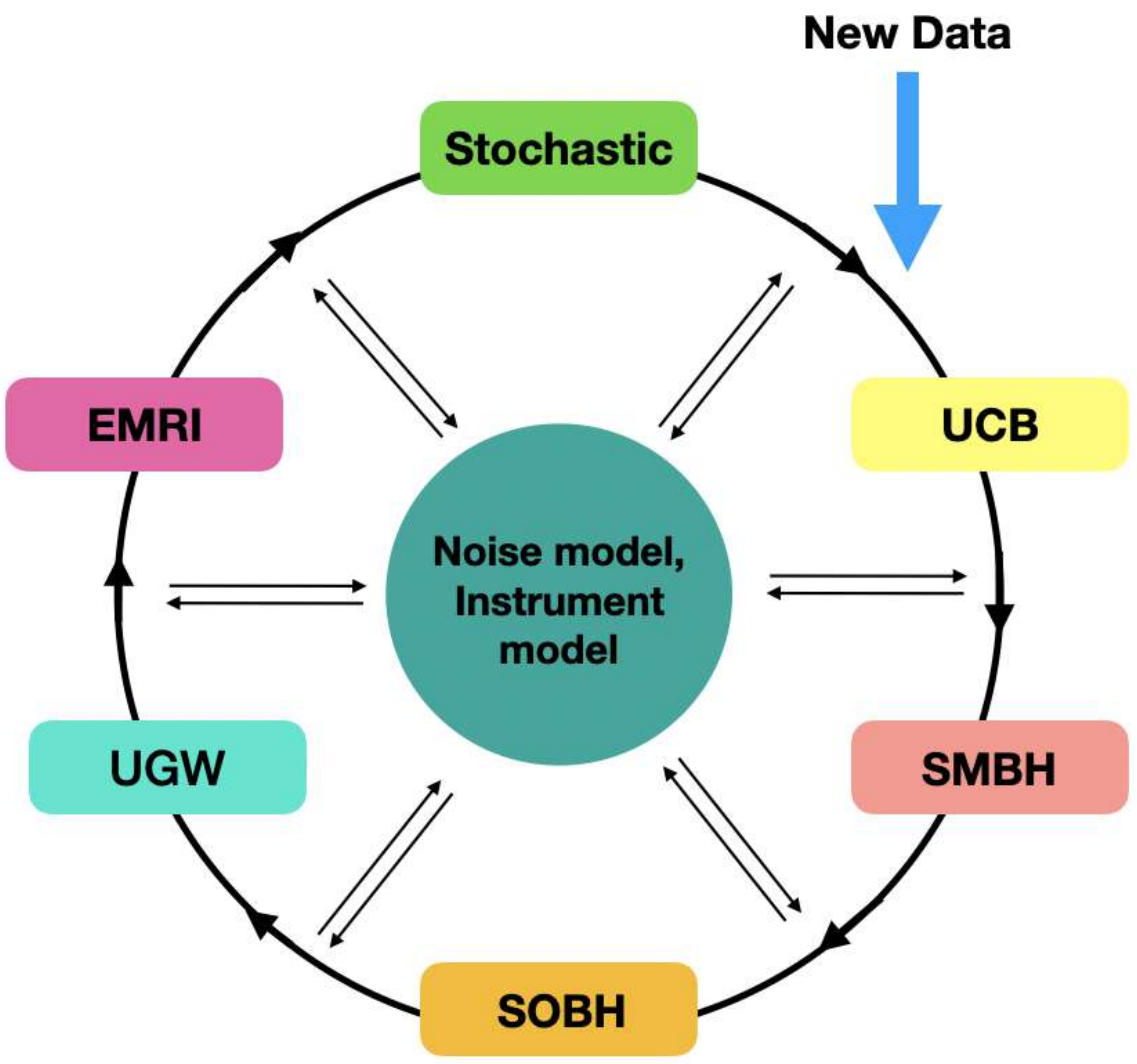
lisa

Taken from https://lisa-ldc.lal.in2p3.fr/challenge2

See next talk by Ivan M. Vilchez!



- This means that there is going to be signals overlapping in time and frequency.
- At the same time, given different population synthesis models, we expect contributions that would yield a confusion signal in certain frequency bands (.i.e the GBs case).
  - How to deal with them (analyze them)?



arXiv:2004.08464

- Many types of sources
- We have models and waveforms, but the problem is that the total number of sources is *unknown*.
- Search for them in the noisy data, a noise which will contain unresolvable confusion GW signals.
- Tackle the issue with a “Global fit” scheme.

- 10/2021 - 09/2023
- LISA Group @ Aristotle University of Thessaloniki
- Collaboration with M. Katz & N. Korsakova
- Full members of the LISA Consortium: Nikolaos Stergioulas, George Pappas, Nikolaos Karnesis, Lazaros Souvaitzis
- Diploma student: Mary Georgousi

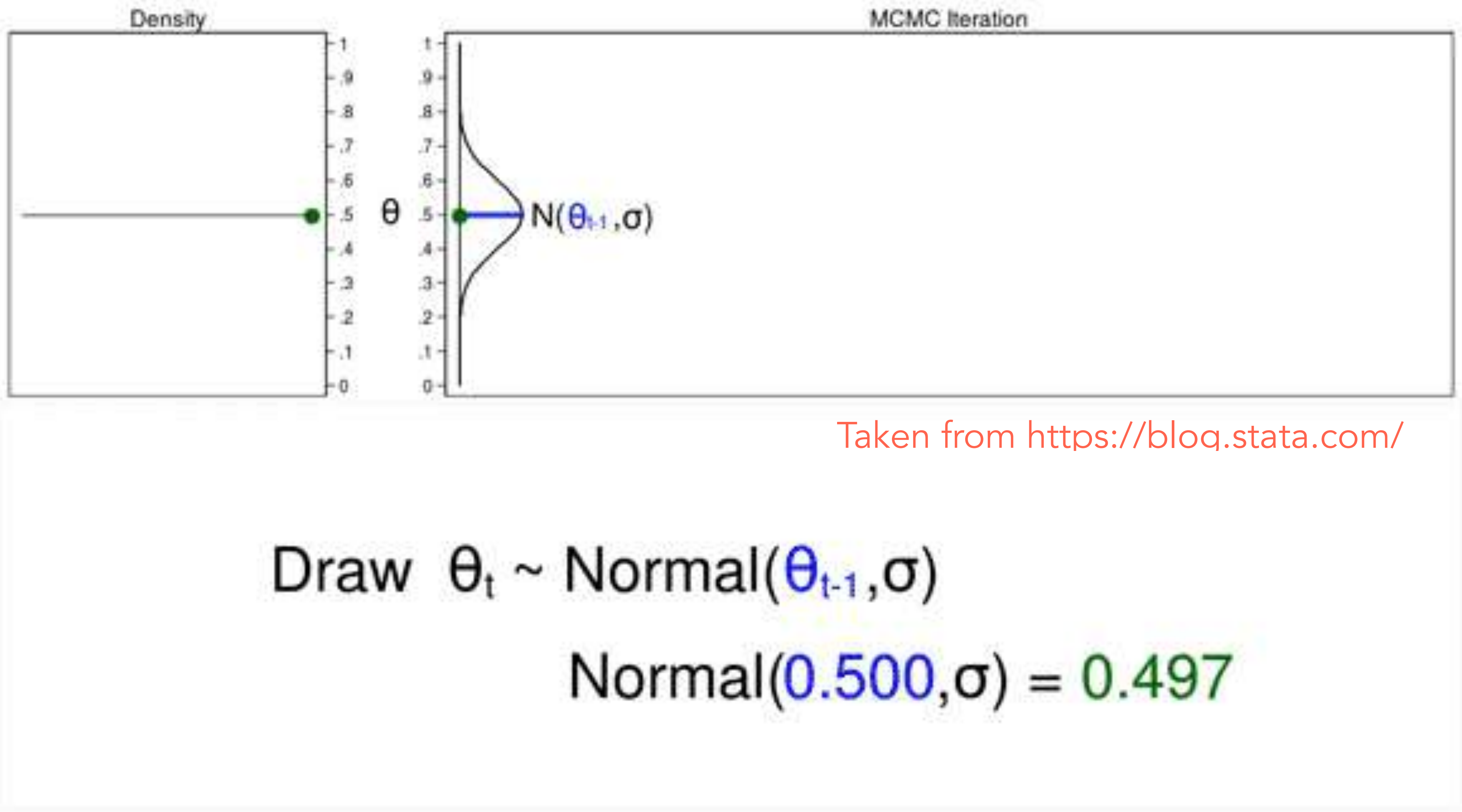
LISAπ

prodex

esa

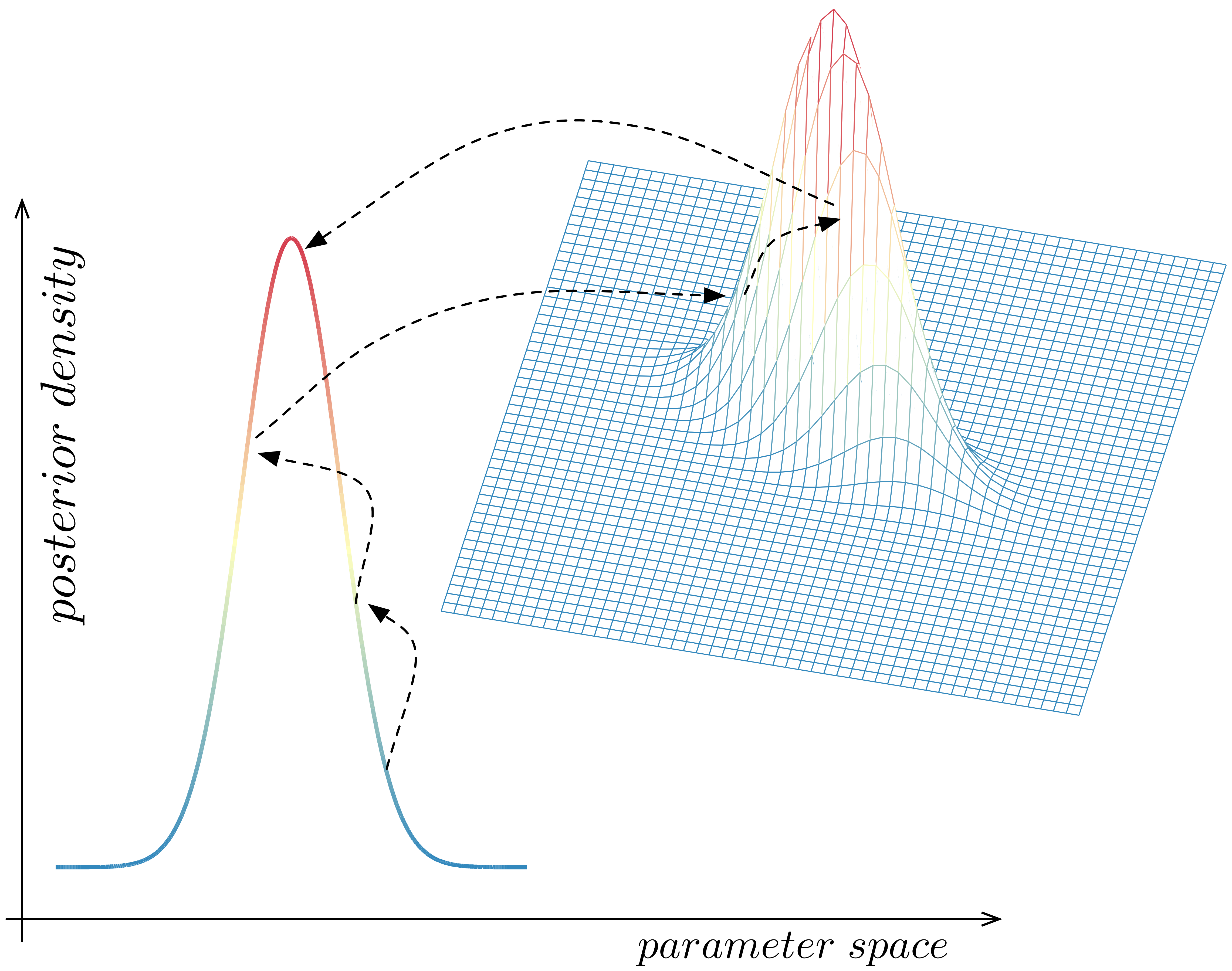
- We base our analysis on Markov Chain Monte Carlo methods.
  - Suppose we want to do parameter estimation and search.
  - We define a likelihood function

- We base our analysis on Markov Chain Monte Carlo methods.
  - Suppose we want to do parameter estimation and search.
  - We define a likelihood function
  - The surface can be "bumpy", and we need to explore the parameter space.
  - MCMC: Start from a point.
  - Propose a new based on a proposal distribution.
  - Accept it based on a probability.

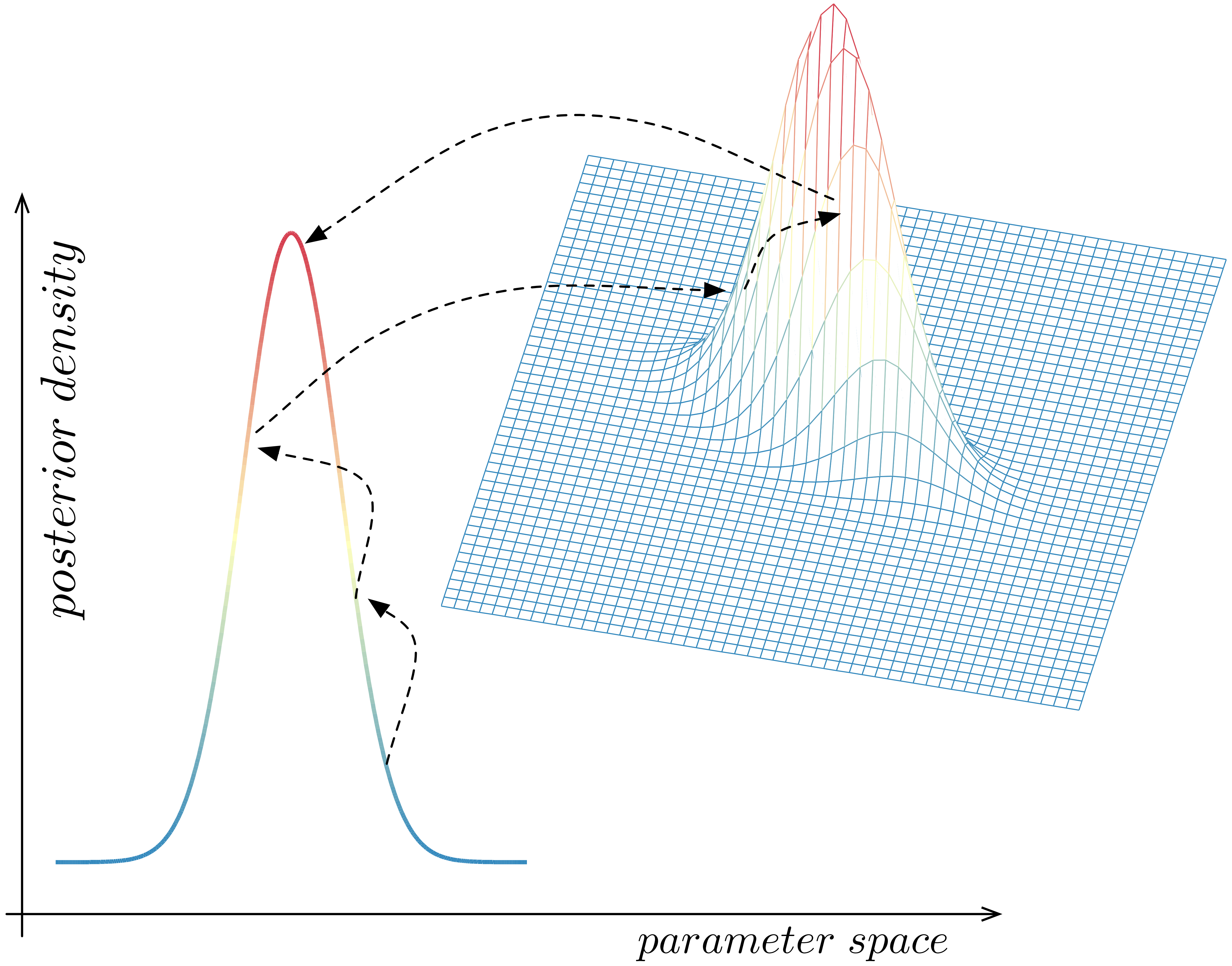


Taken from https://blog.stata.com/

- What happens if we do not know how many sources are there? More dimensions.
  - Now we have to move across models!
  - Occam's Razor: *Simpler models that explain the observations are always preferred*!
  - Challenging to tune.
  - This is where the LISAπ proposes novel solutions.
  - Employ improvements that will allow smooth dimension transitions and efficient sampling.

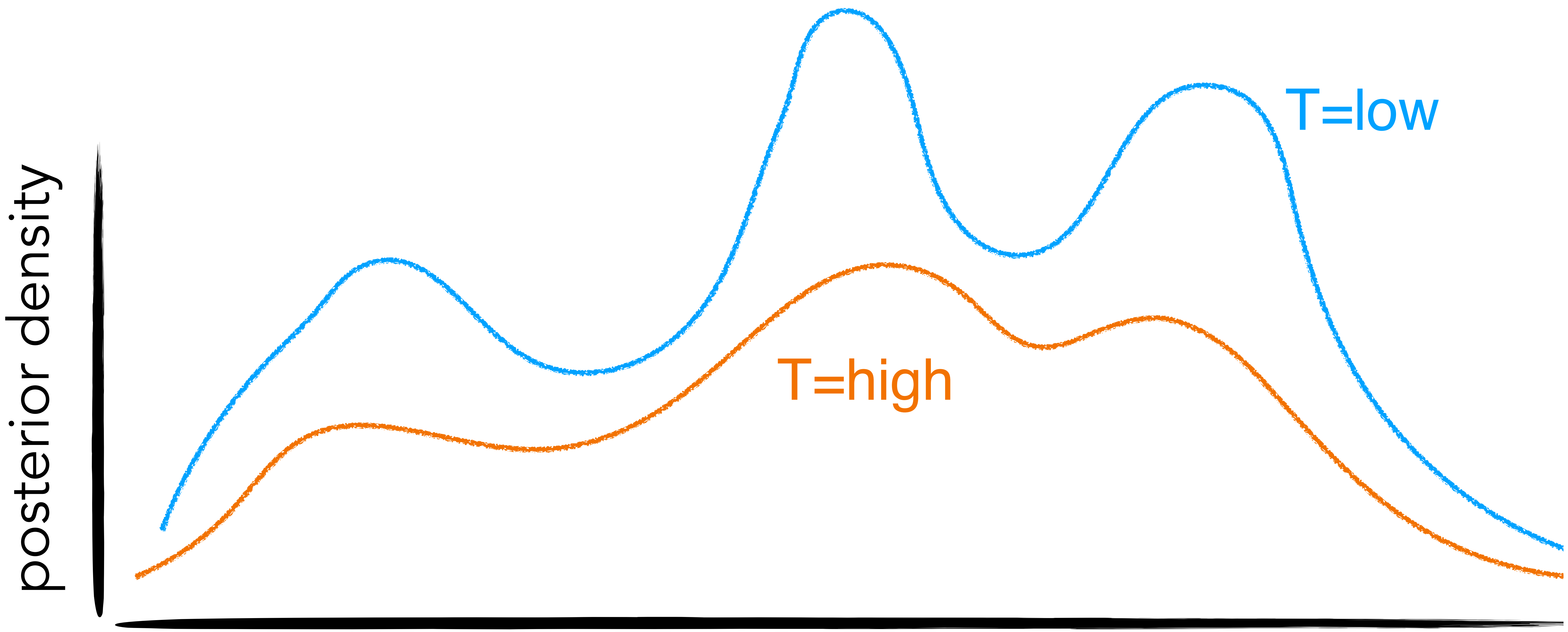


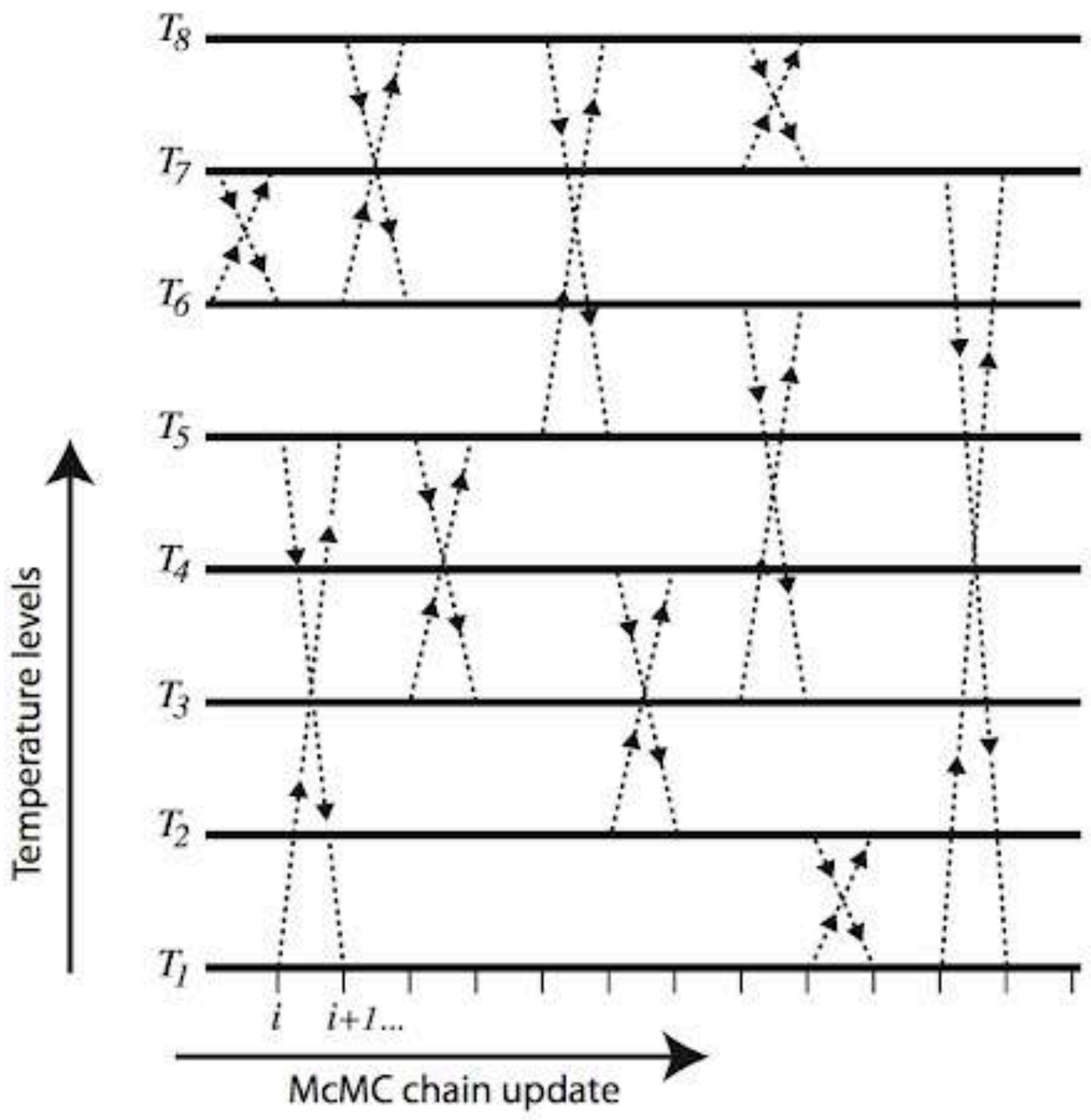
- What happens if we do not know how many sources are there? More dimensions.
  - Now we have to move across models!
  - Occam's Razor: *Simpler models that explain the observations are always preferred*!
  - Challenging to tune.
  - This is where the LISAπ proposes novel solutions.
  - Employ improvements that will allow smooth dimension transitions and efficient sampling.



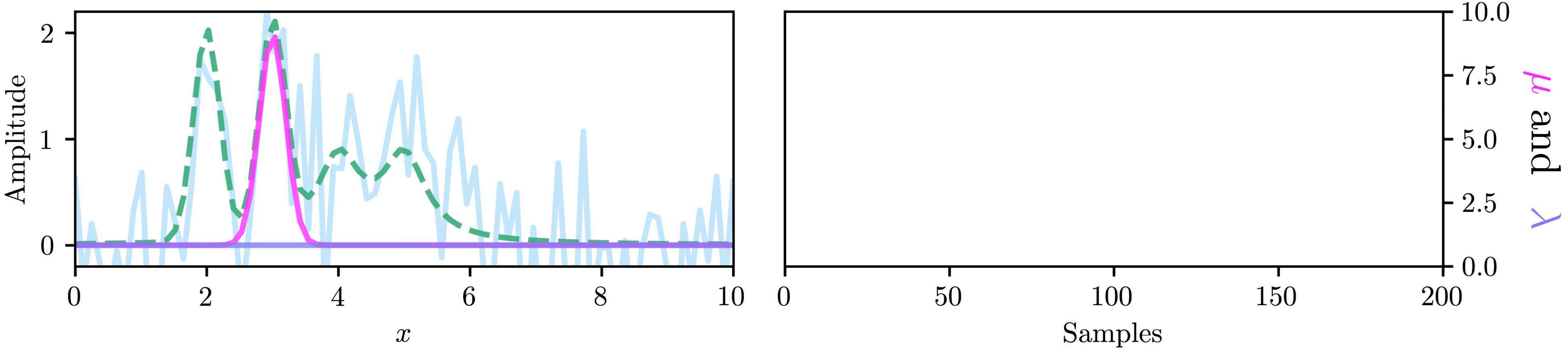
Littenberg+ 2020 - *PRD 101, 123021*

• Novel ideas tested with LISAπ : *Adaptive Parallel Tempering*

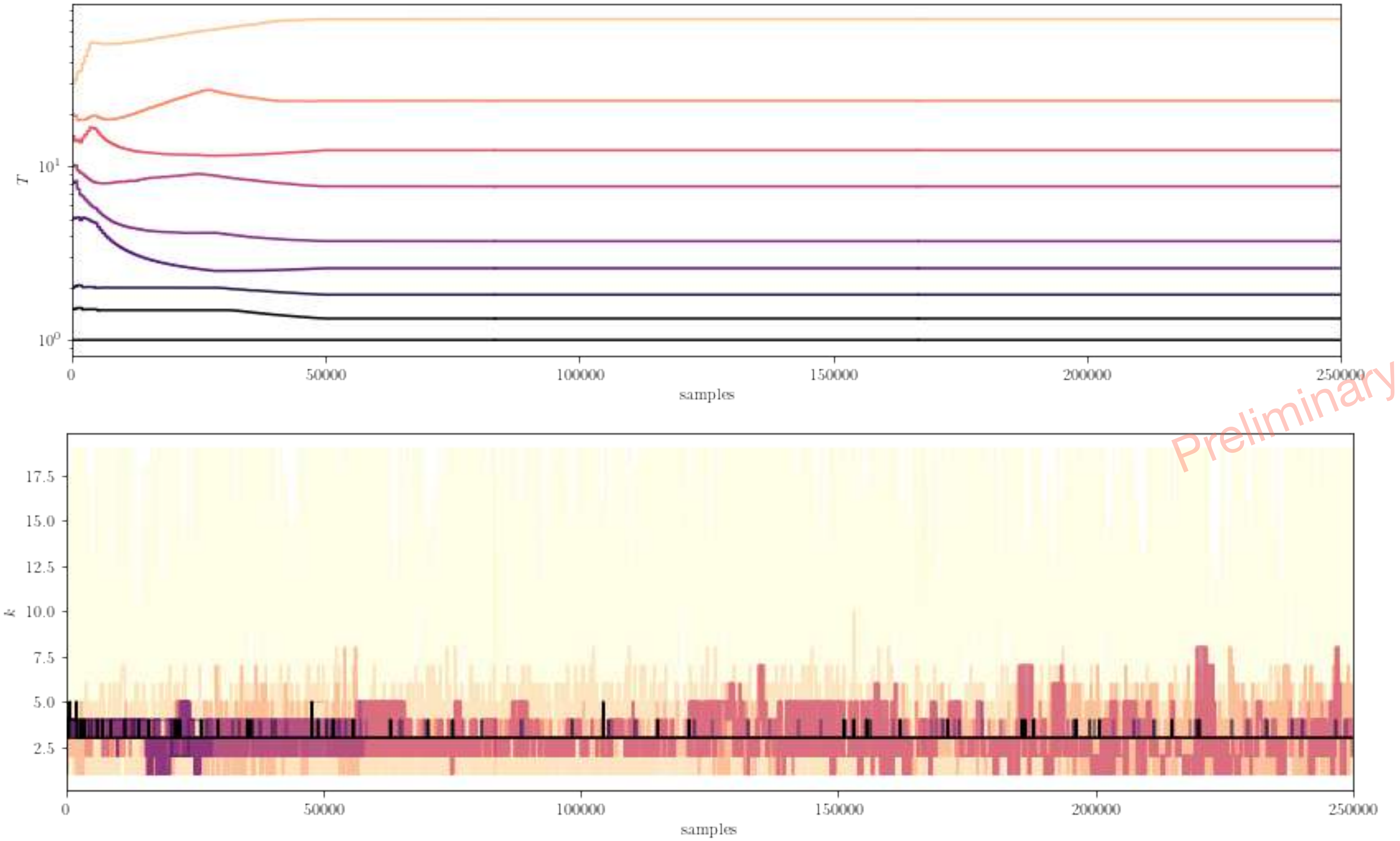


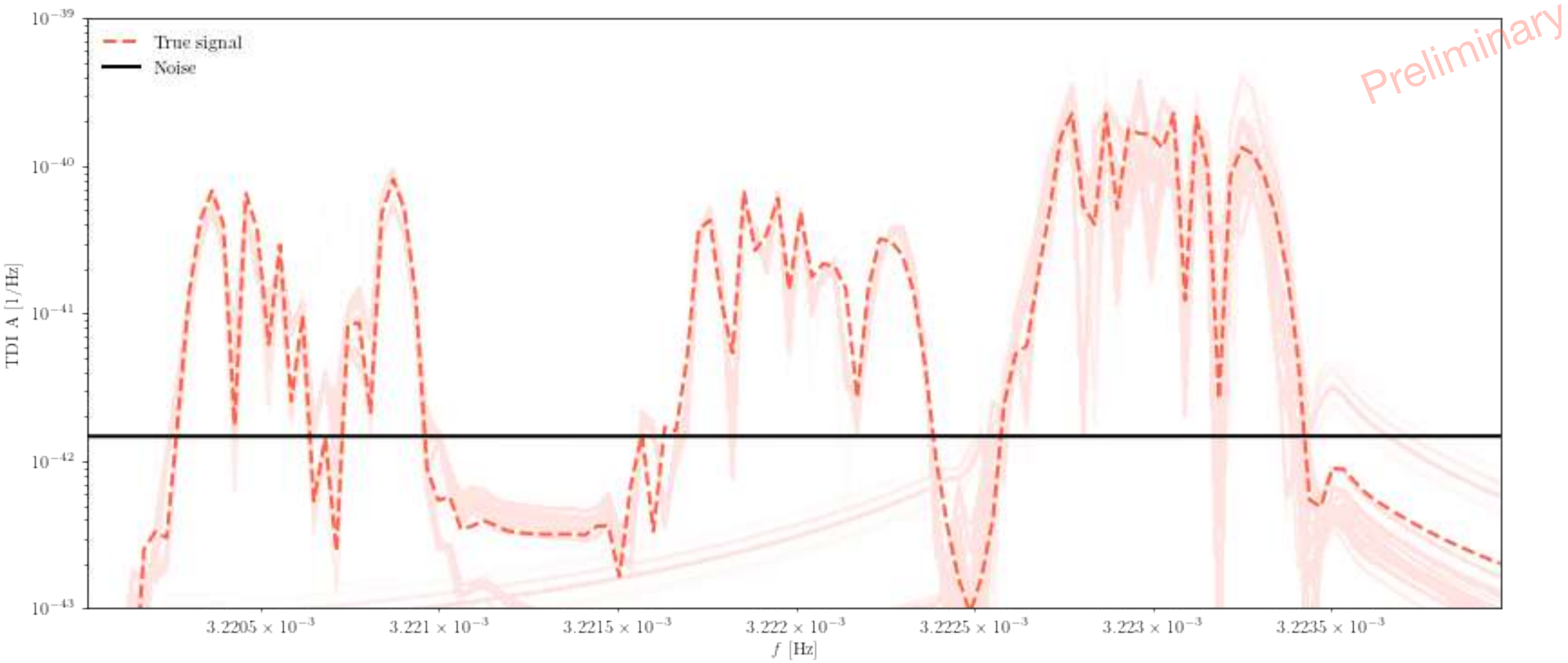


- A simple example / illustration
  - Assume a mixture of Gaussian and Cauchy.
  - Run in order to decide the dimensionality and the component types!



• Output example





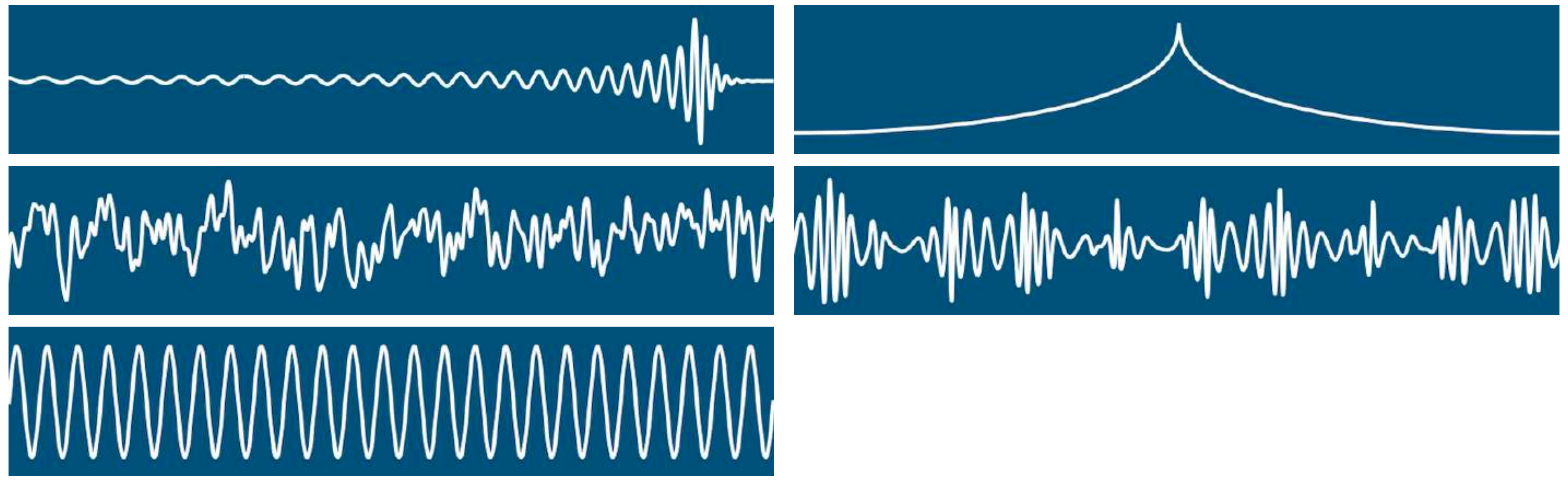
True signal
Noise
Preliminary
TDI A [1/Hz]
10^{-39}
10^{-40}
10^{-41}
10^{-42}
10^{-43}
3.2205×10^{-3}
3.221×10^{-3}
3.2215×10^{-3}
3.222×10^{-3}
3.2225×10^{-3}
3.223×10^{-3}
3.2235×10^{-3}
f [Hz]

- Plans for LISAπ
  - With N Korsakova (SYRTE)
    - Make use of NNs : build efficient proposal distributions.
  - With M. Katz (AEI Potsdam)
    - GPU acceleration : efficiency + prototyping
  - Part of the analysis of the second LISA Data Challenge (LDC2):
    - Collaboration of AEI, AUTh, APC, UBirmingham, Caltech.
    - Analyze the data with a single pipeline.

- Maybe a few words on the LISA Data Challenges:
  - A common language for the LISA community.
  - Share ideas, codes, methods.
  - Prototyping
  - Test realistic scenarios.
    - Both from instrument and nature point of view.

See next talk by Ivan M. Vilchez!

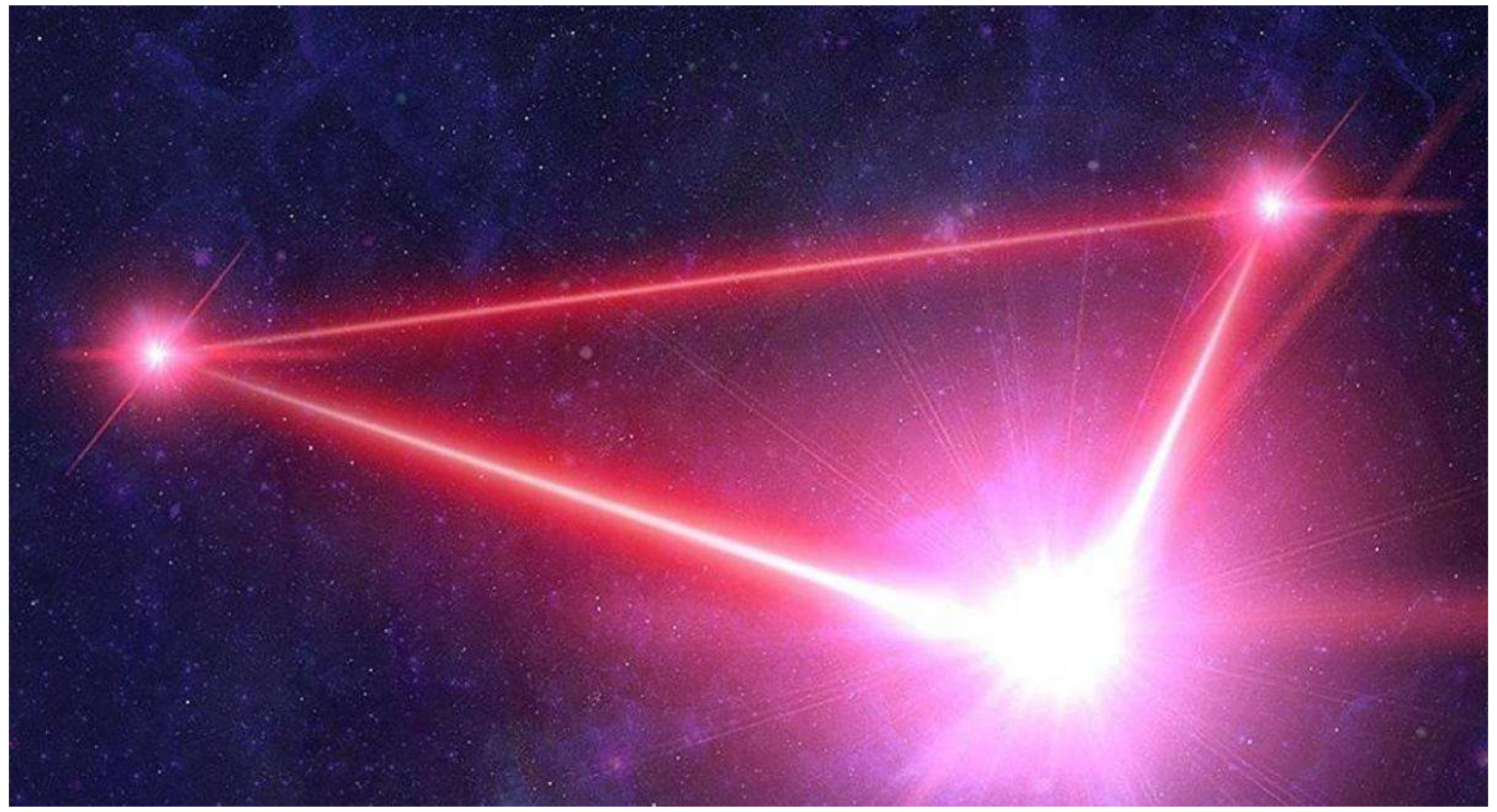
- LDC1: **Radler** is already completed - publication will follow.
- LDC2: **Sangria** is under way (+ **Yorsh** + **Spritz**)



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

lisa

- Searching for signals in LISA, requires to look for all types of sources simultaneously, model the noise, repeat as data comes to ground.
- Test ideas like the *Global Fit*.
- Test algorithm ideas (MCMCs, Nested Sampling, ML, …)
- With the LDCs we focus on answering some of these questions.



Έξτρα Ματέριαλ

• Stochastic signals from cosmological sources

- Again, take a step back, and ask:
  - If we assume we subtract loud sources successfully,
  - how good is LISA in detecting weak stochastic GW signals (both cosmo + astro)?
- We can make analytical predictions, if we start with computing the power spectra

O M Solomon Jr, 10.2172/ 5688766

$$p(\overline{D}[i]|S_\mathrm{t}) = \frac{e^{-\frac{\sum_{j=1}^{N} D_j[i]}{S_\mathrm{t}[i]}}}{S_\mathrm{t}[i]^N} = \frac{e^{-N\frac{\overline{D}[i]}{S_\mathrm{t}[i]}}}{S_\mathrm{t}[i]^N},$$

- Where S_m is the theoretical power spectrum we are interested in. Then if we assume

$$S_\mathrm{t}[i] = S_\mathrm{o}[i] + S_\mathrm{n}[i]$$

- and that we have a prior knowledge of S_n around ε, we can try to marginalize it out, and then

$$p(S_\mathrm{o}|\overline{D}, S_\mathrm{n}) = \frac{1}{(\epsilon^+ + \epsilon^-)} \int_{\bar{S}_\mathrm{n}-\epsilon^-}^{\bar{S}_\mathrm{n}+\epsilon^+} \frac{e^{-N\frac{\overline{D}}{S_\mathrm{o}+S_\mathrm{n}}}}{(S_\mathrm{o}+S_\mathrm{n})^N} \mathrm{d}S_\mathrm{n}$$

NK et al, arXiv:1906.09027