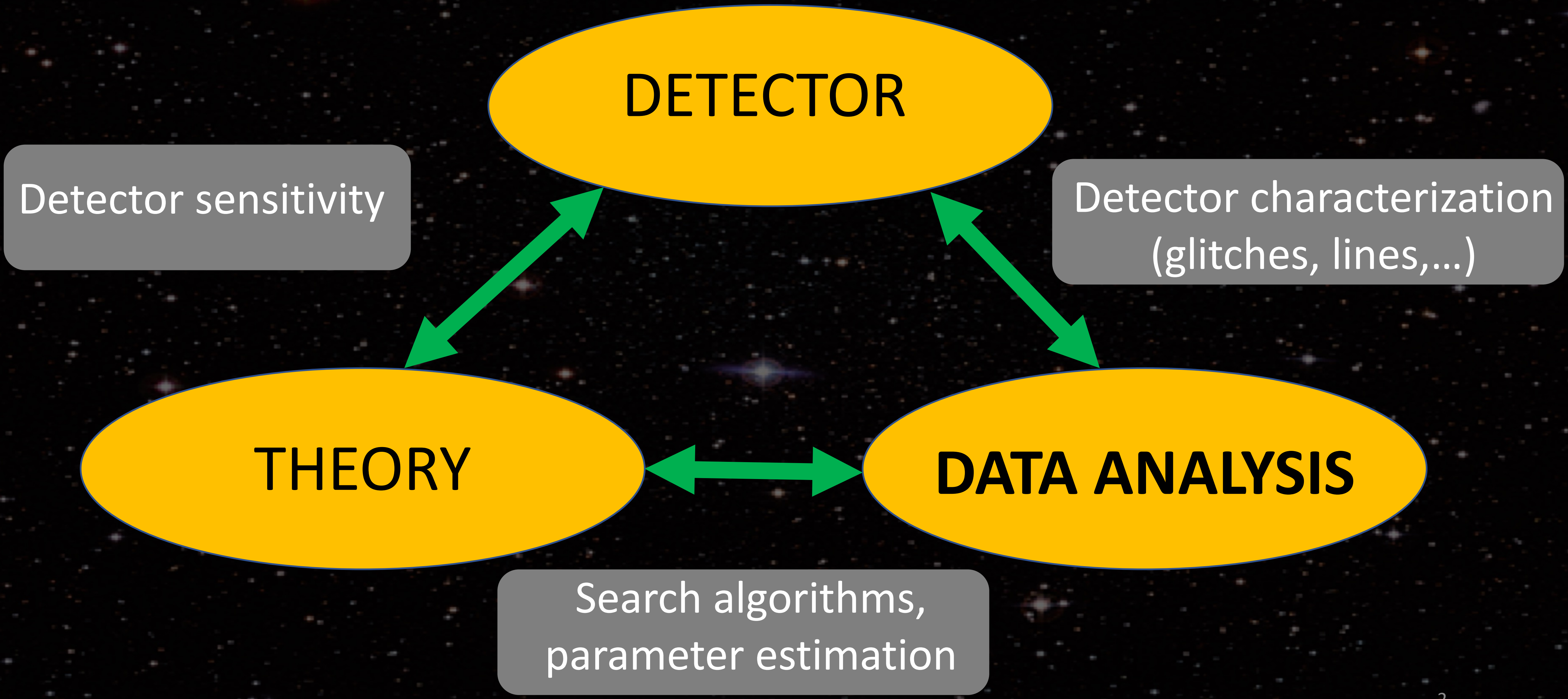
Data analysis challenges with Einstein Telescope

Cristiano Palomba – INFN Roma

Outline:

- ❑ Why is data analysis important?
- ❑ What are the difficulties of data analysis for gravitational waves?
- ❑ Data analysis challenges in ET

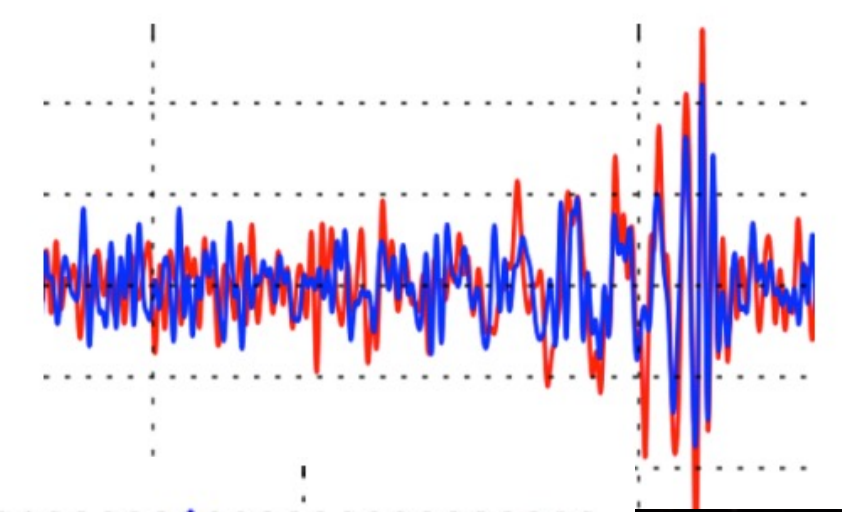
The “magic” triangle of GW detection



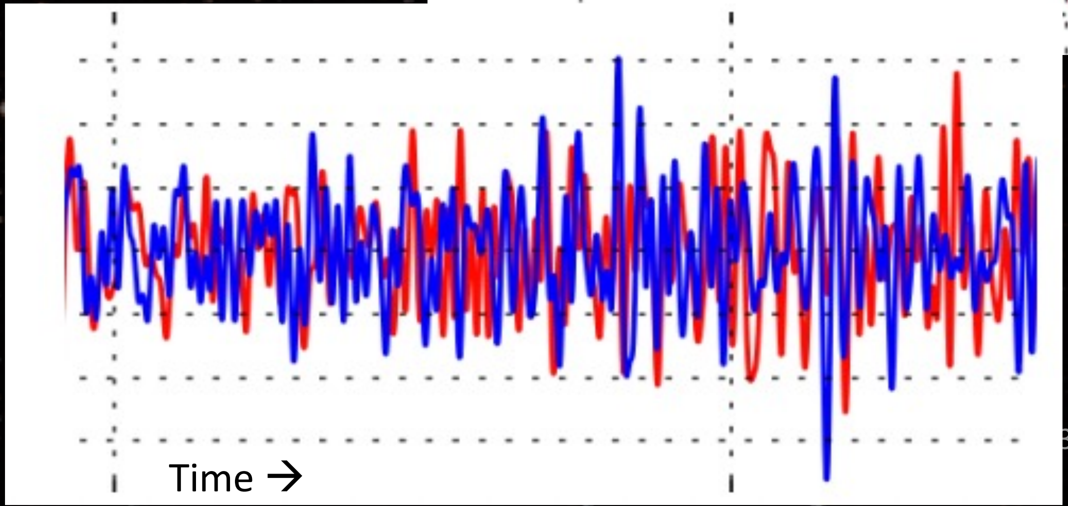
The goal of analysing the data is

* trying to detect signals which are, typically, buried in noise
* estimating signal parameters as accurately/precisely as possible
* inferring source properties

❖ In exceptional cases the signal is so strong it can be seen after simple processing (e.g. a band–pass filter in the case of GW150914)



❖ In most cases the signal is not visible 'by eye'

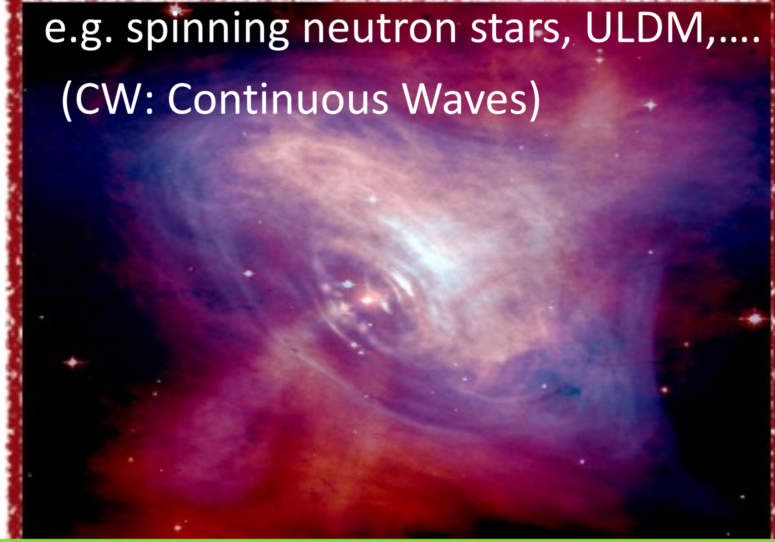


Sources of GWs: rough classification

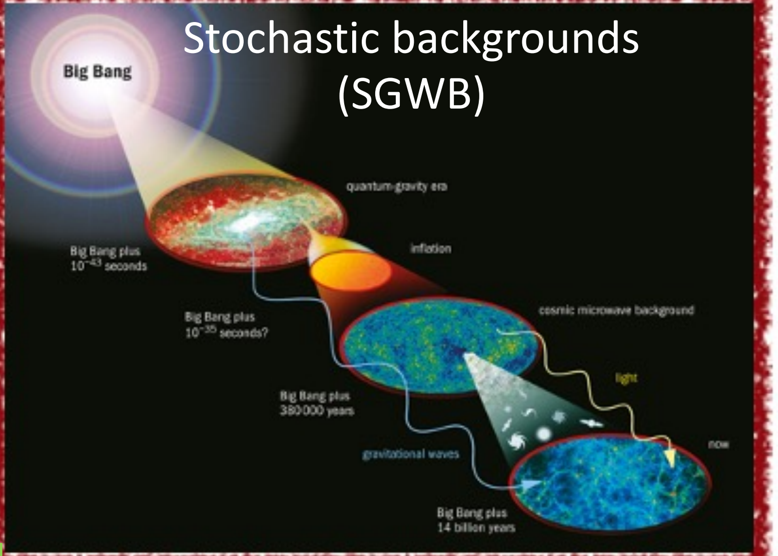
Long-lived

T ~ months or years

e.g. spinning neutron stars, ULDM,….
(CW: Continuous Waves)



Stochastic backgrounds
(SGWB)



Transients

T ~ up to minutes

BH and/or NS binary systems
(CBC: Compact binary coalescence)

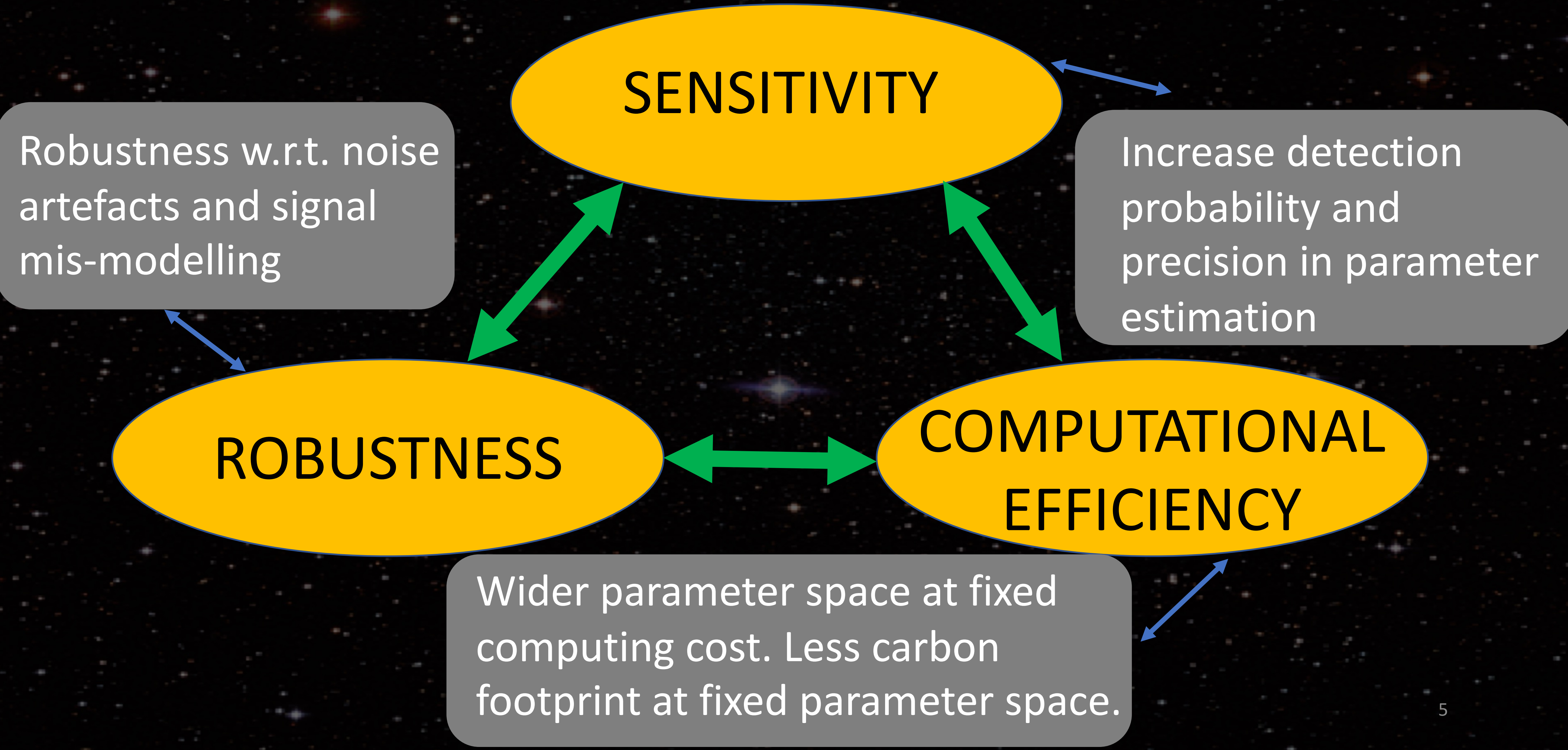
e.g. core collapse supernovae,
post-merger remnants,…
(Burst)

← Long-duration/short duration →

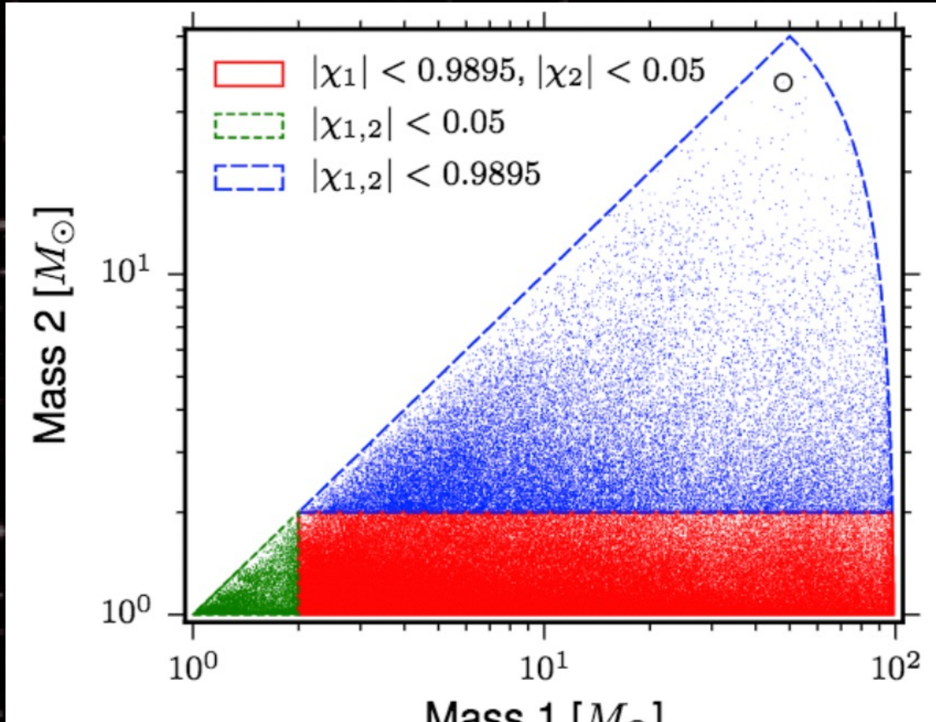
← modeled/unmodeled →

Different sources ←→ different data analysis techniques

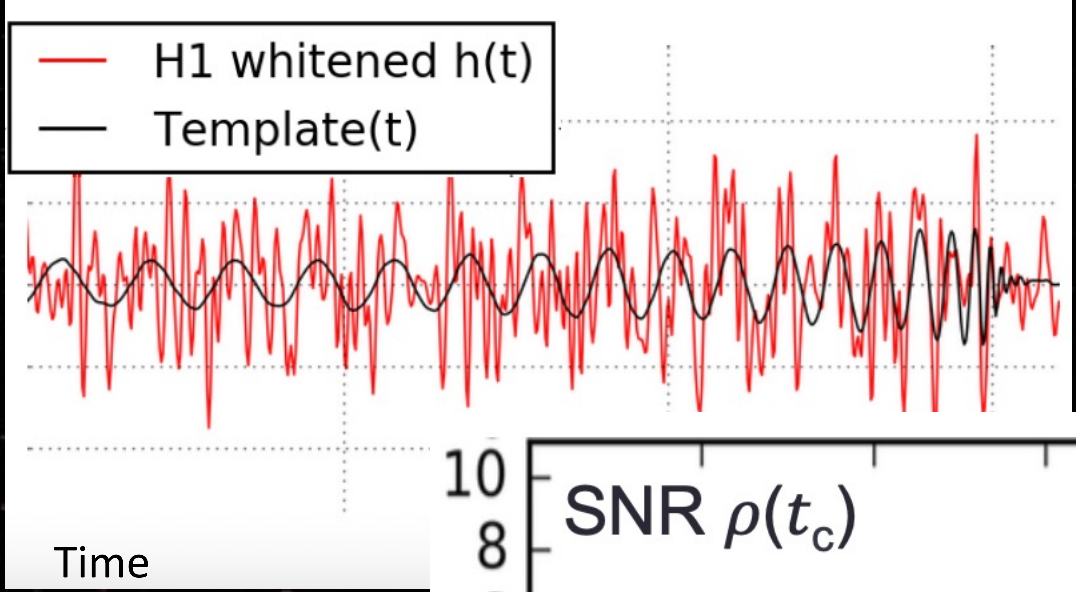
Data analysis cornerstones

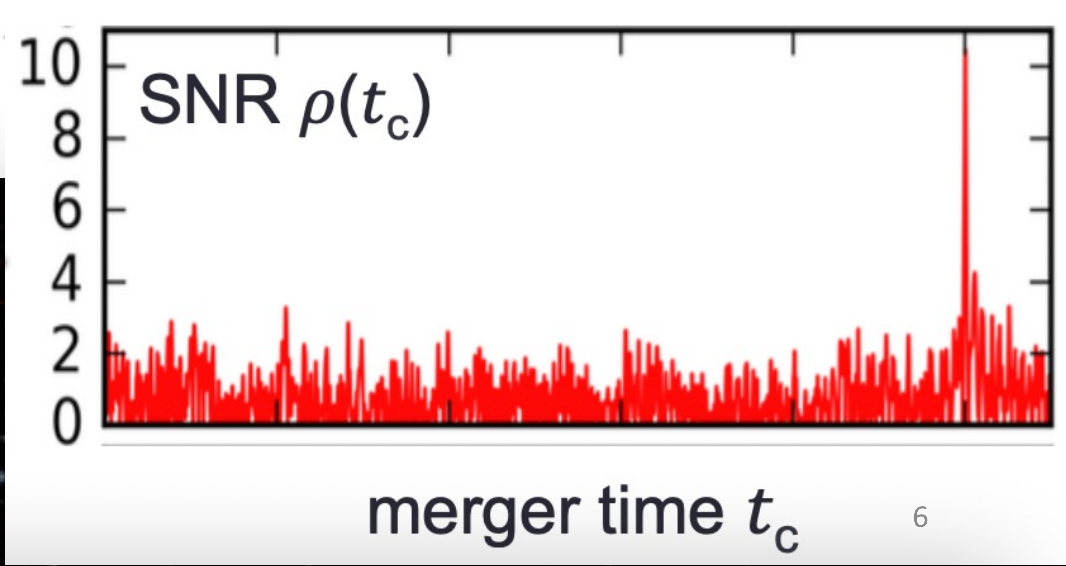


- For modeled signals, **matched filtering** is a standard technique
  - Template: anticipated gravitational waveform as a function of source parameters
  - Template bank: collection of templates



LV, PRX6 041015 (2016)



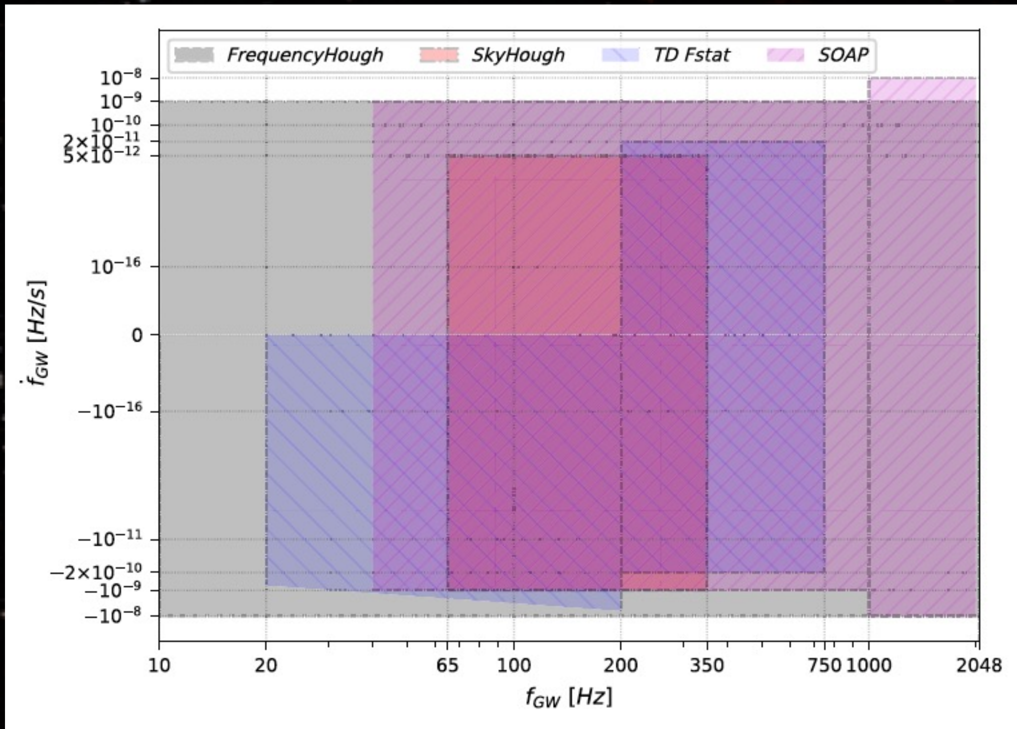


- Find template within a template bank that fits data best to provide confident detection

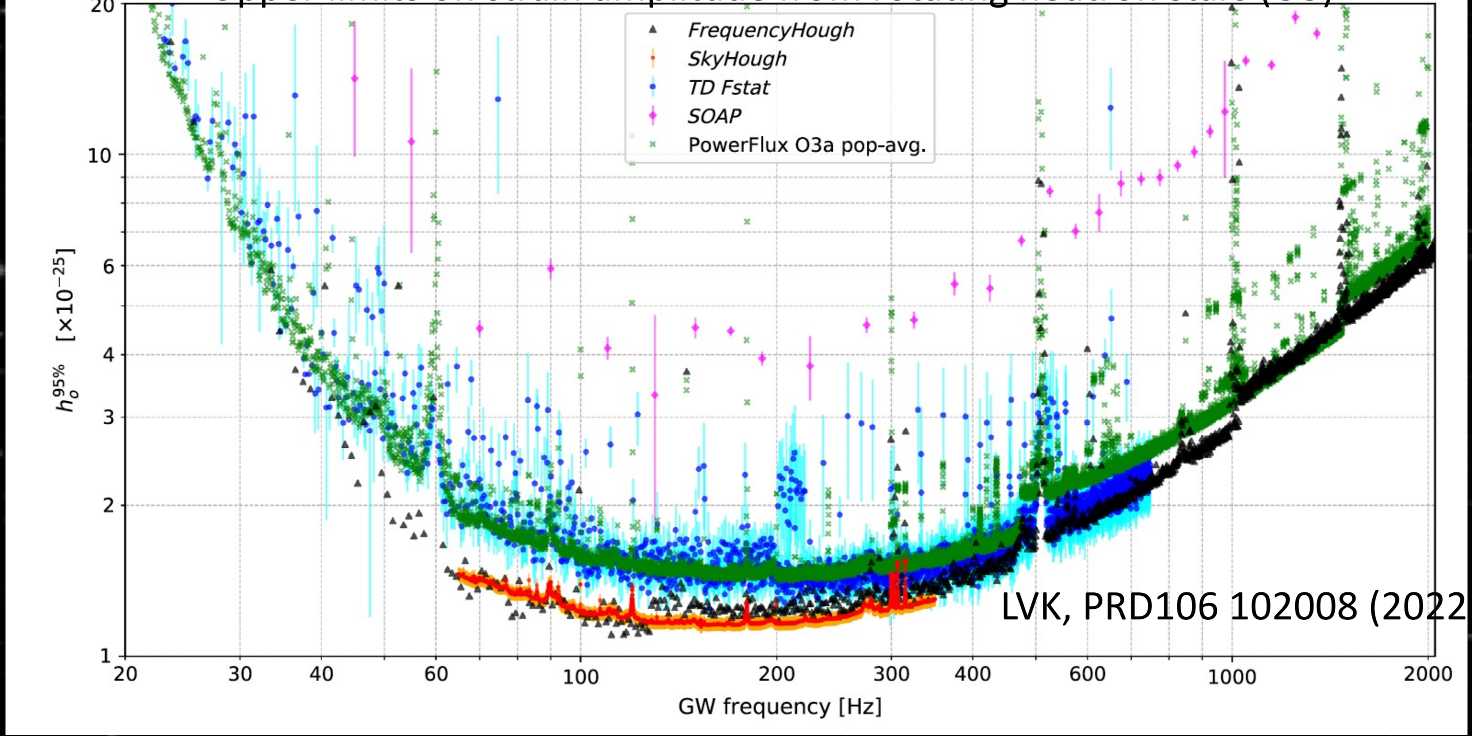
- Matched filtering can be computationally too expensive, if the parameter space to be explored is very large
- Matched filtering is not optimal at all if the signal cannot be accurately modeled
- When matched filtering is not feasible, other methods are used
  - Lower theoretical sensitivity
  - Lower computational cost

Be careful when, in literature, matched filtering is assumed to predict the detectability of some new source

- All-sky searches for CWs are an example of computationally-bound analysis
  - Matched filtering would imply to search over O(10^{31}) points in the parameter space →unfeasible!
  - Use semi-coherent methods to reduce computing cost (O(10^{14}) points)



Upper limits on strain amplitude from rotating neutron stars (O3)



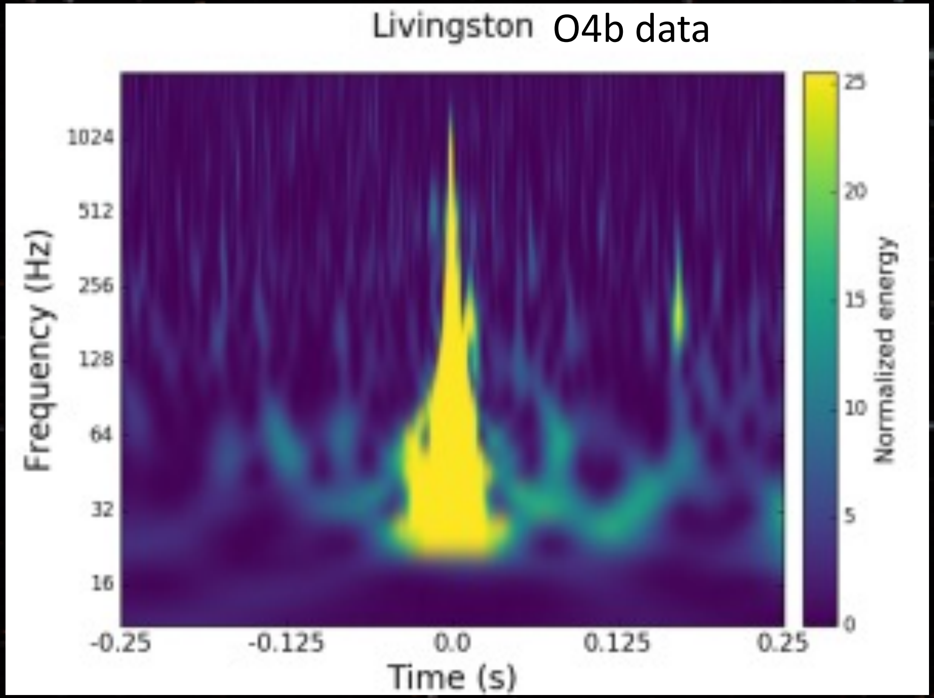
LVK, PRD106 102008 (2022

- Code optimization, use GPU,…

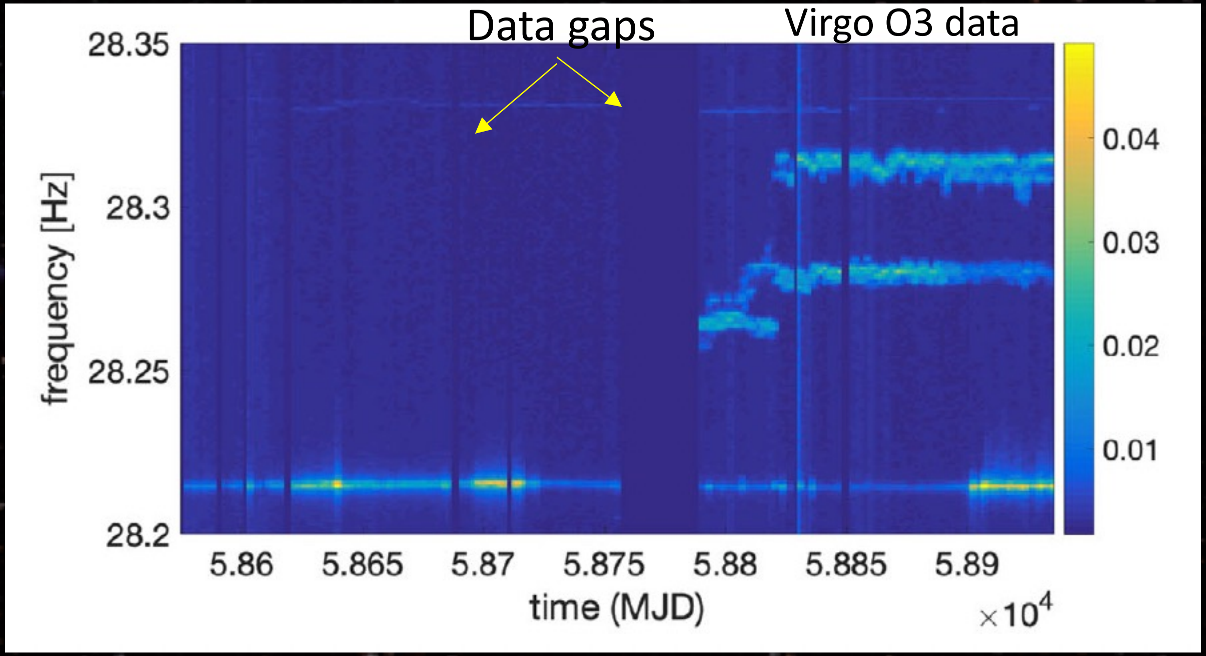
Noise is often non-Gaussian, non-stationary and may have weird features

Moreover, data have a lot of gaps (due to the detector not being in science mode or not properly working in a given time period)

glitches: short duration disturbances

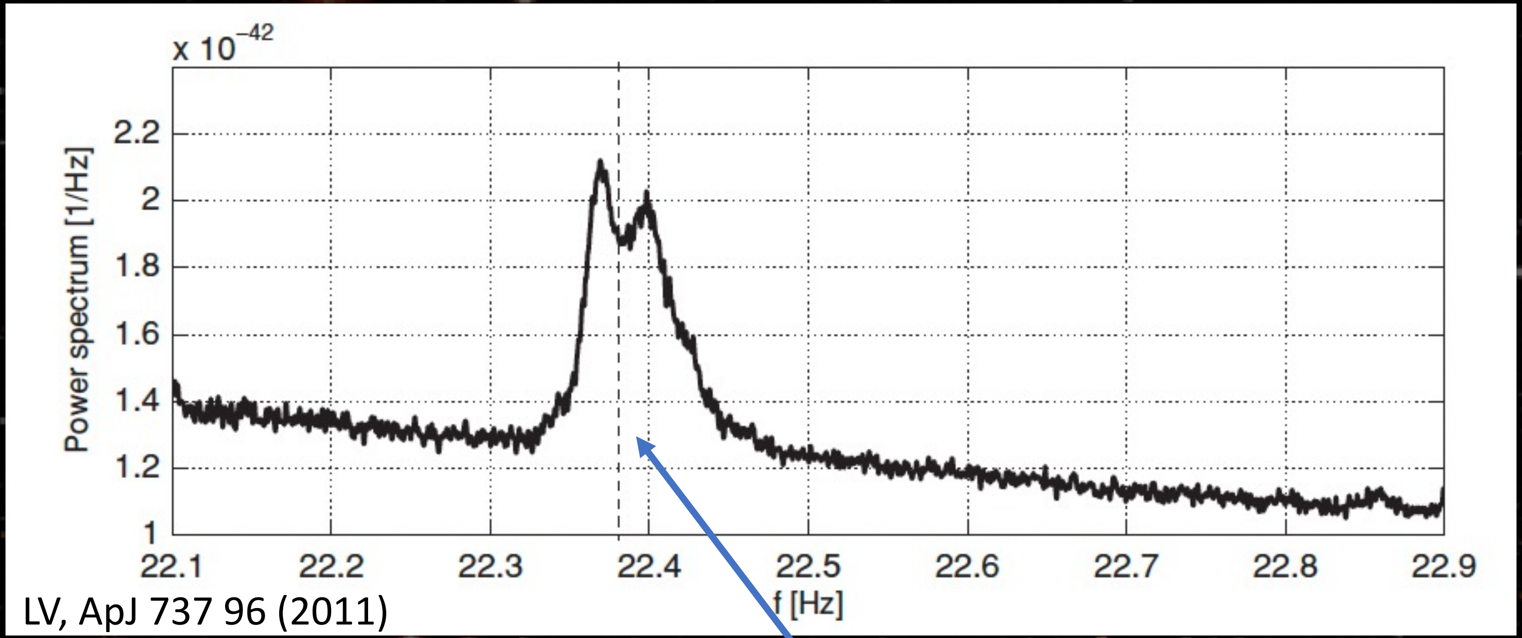


Spectral disturbances

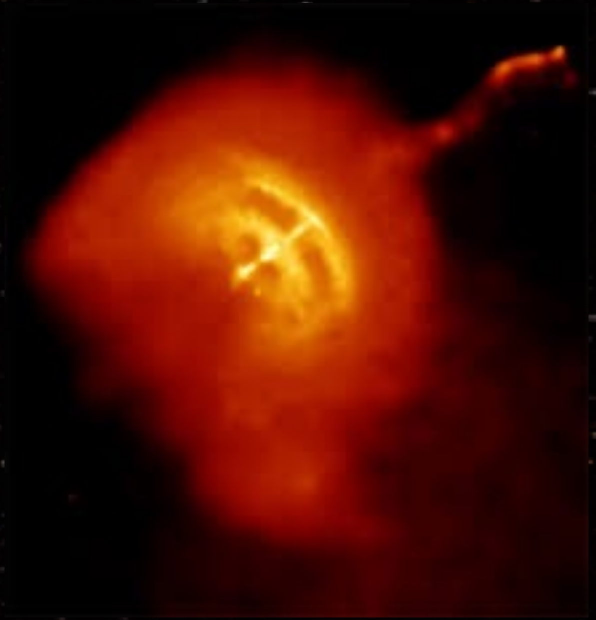


Identification of the noise source (and, possibly, its removal) is a critical task

An example of impact on searches: the “Vela killer” in Virgo VSR2 data



LV, ApJ 737 96 (2011)



Expected frequency of the GW emission from Vela pulsar

The source of this disturbance was seismic noise produced by the engine of the chiller pumps that circulate coolant fluid for the laser of the mirror thermal compensation system.

Detchar and noise hunting are crucial activities in support of data analysis

Analysis pipelines must be tested in real detector noise

Mock Data Challenges (MDC) are used to compare different analysis pipelines

But interpretation of the results is not trivial (brute force comparisons are often misleading):

- dependence on the specific dataset
- dependence on the parameter space

Develop your algorithm

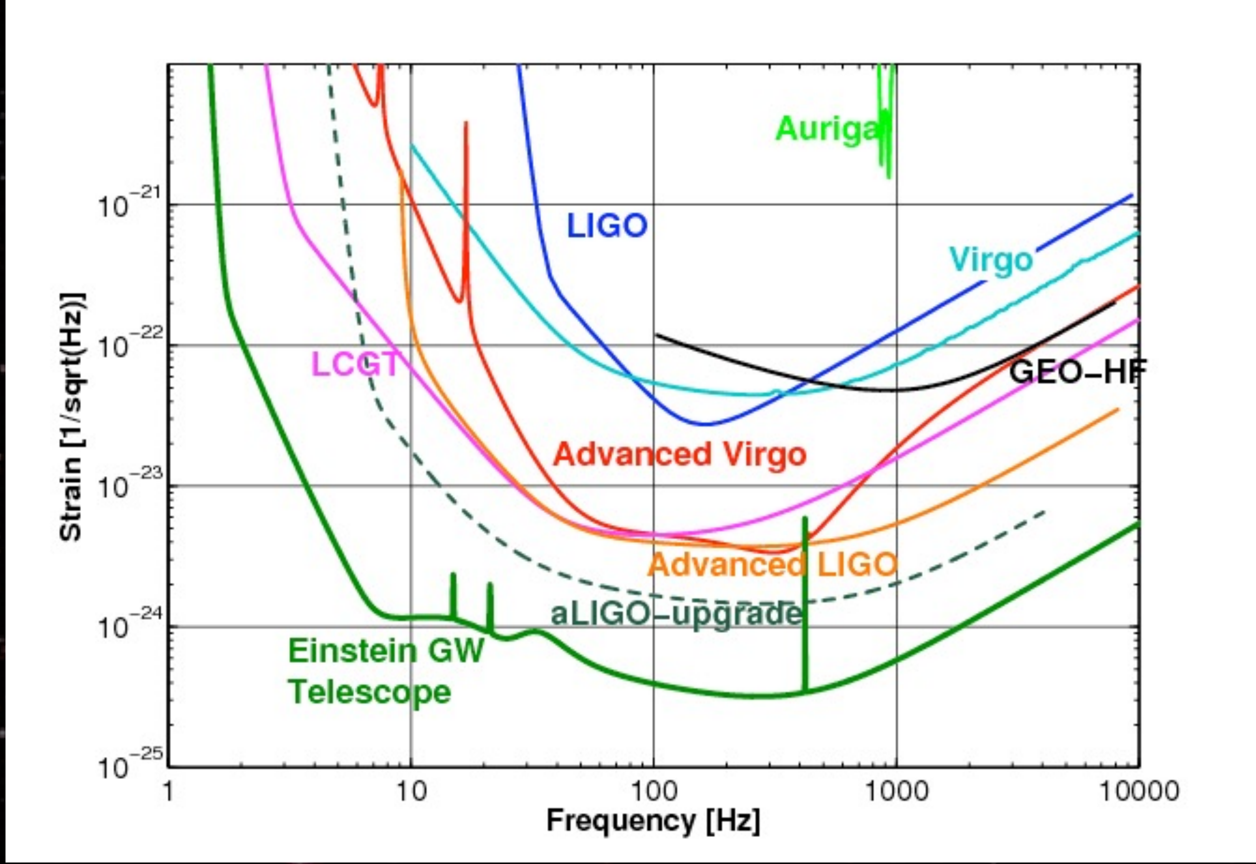
Test your algorithm in simulated noise

Test your algorithm in **REAL** noise

Benefits of ET (and of 3G detectors in general)

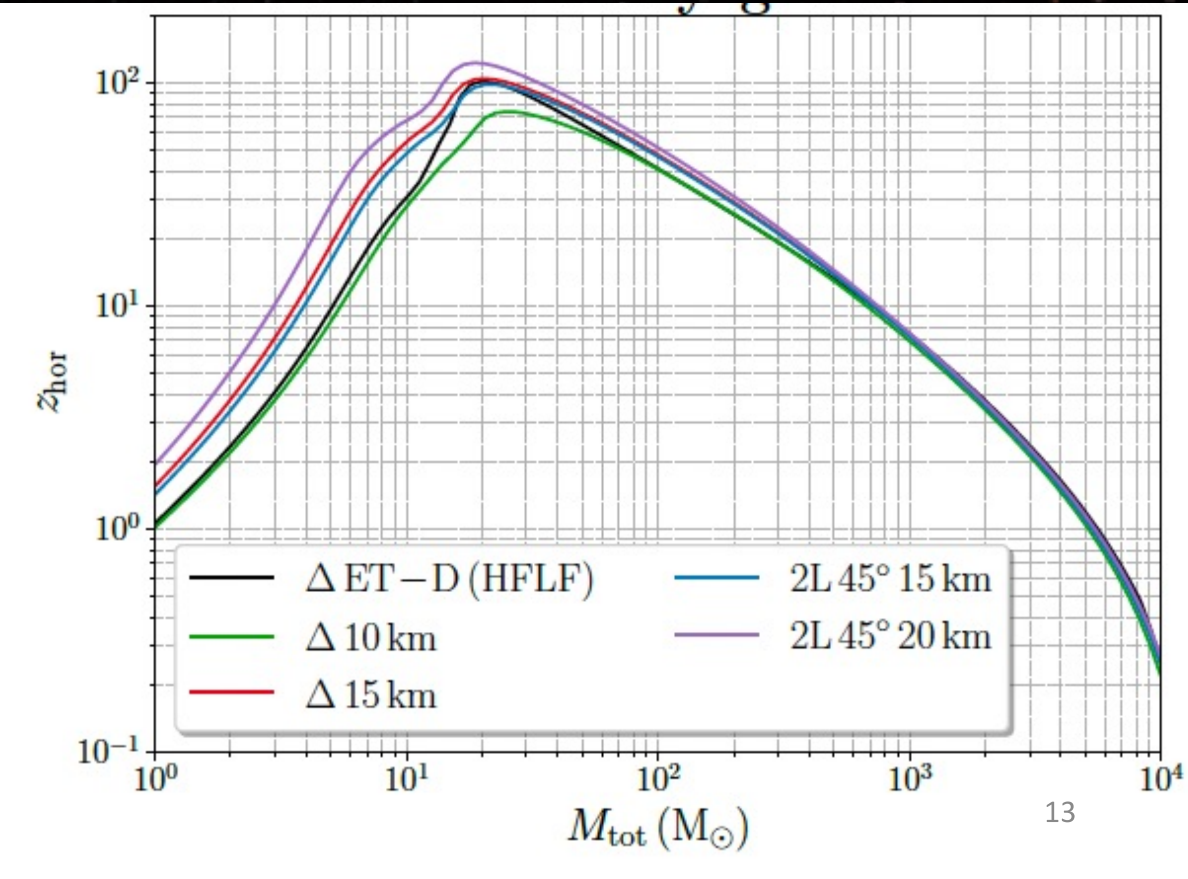
- **Deeper***: observe more distant sources (population studies, cosmology,...)
- **Wider***: increase accessible parameter space (new sources, wider study of known sources,...)
- **Sharper***: detect more subtle effects (new sources, test of GR, NS equation of state)

*: from the GWIC 3G white paper



Factor O(10) sensitivity improvement over a large frequency band

Several orders of magnitude below 10 Hz



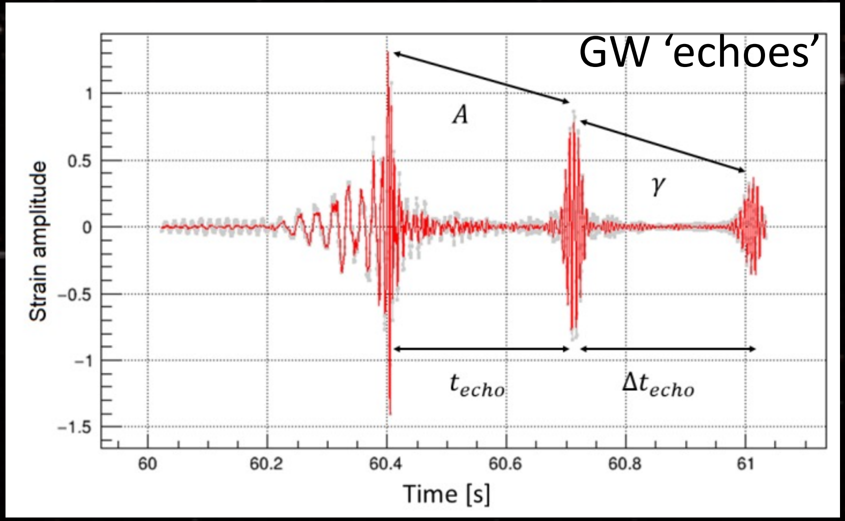
ET will detect all BH binary coalescence in the Universe and NS binary coalescence up to z~2

Developments in **source modeling**, **data analysis techniques** and **computing** are of paramount importance to exploit detector potentialities.

- Development of DA methods for current detectors paves the way to their application in the 3G era

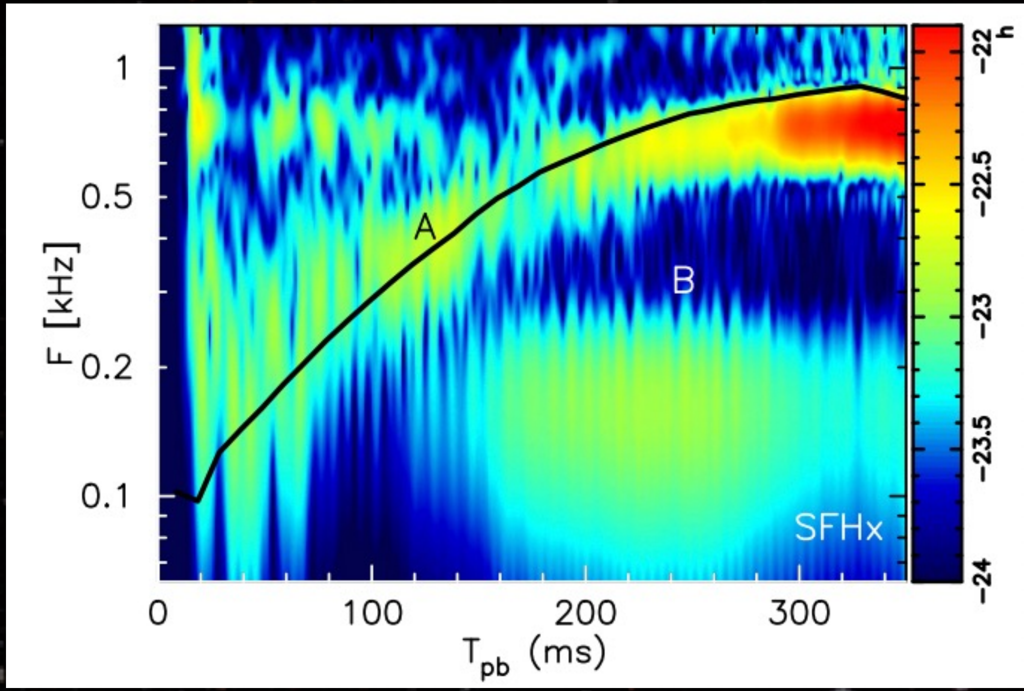
- Always trying to improve sensitivity/robustness/computational efficiency

- Start developing algorithms for sources currently not detectable



Miani+, PRD 108, 064018 (2023)

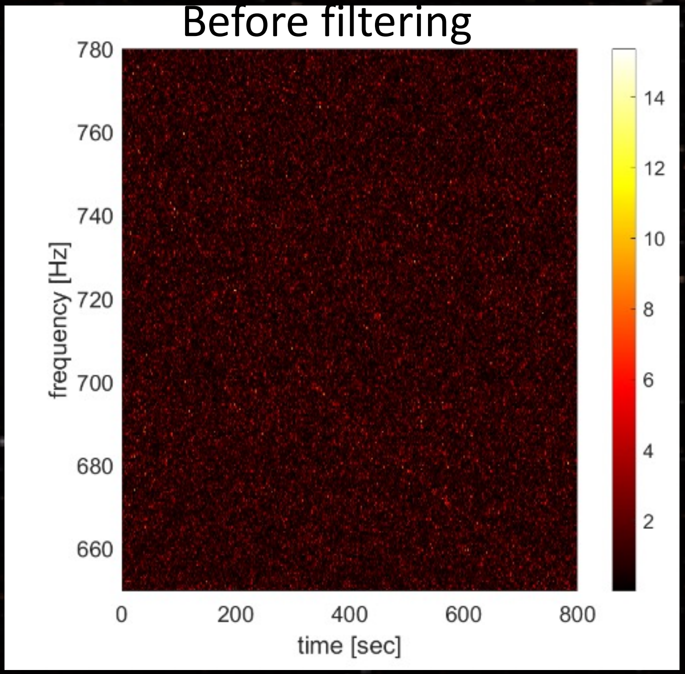
Innovative techniques are being developed. E.g:



CCSN, T. Kuroda+, ApJ 829 L14 (2016)

Image processing

* E.g. 2D filters to reduce the noise not compatible with the signal



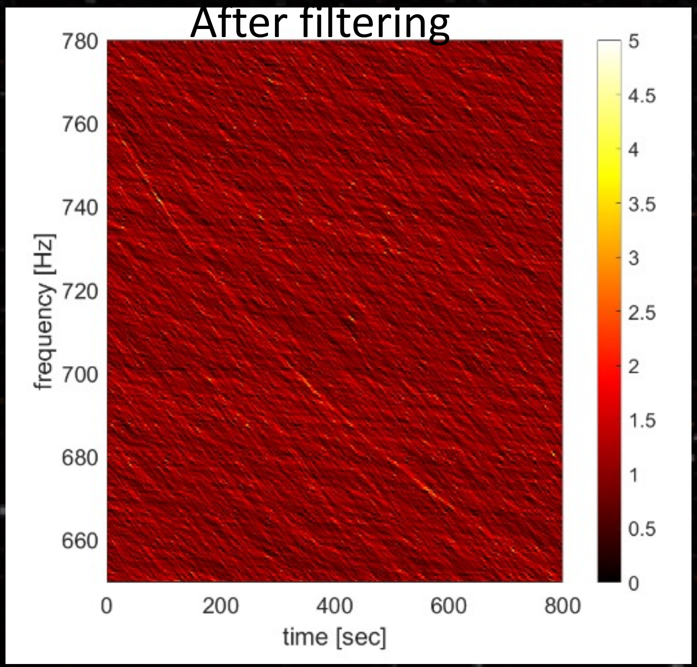


Image processing for long-transient signal searches.
Pierini+, subm. to PRD

real-time detection

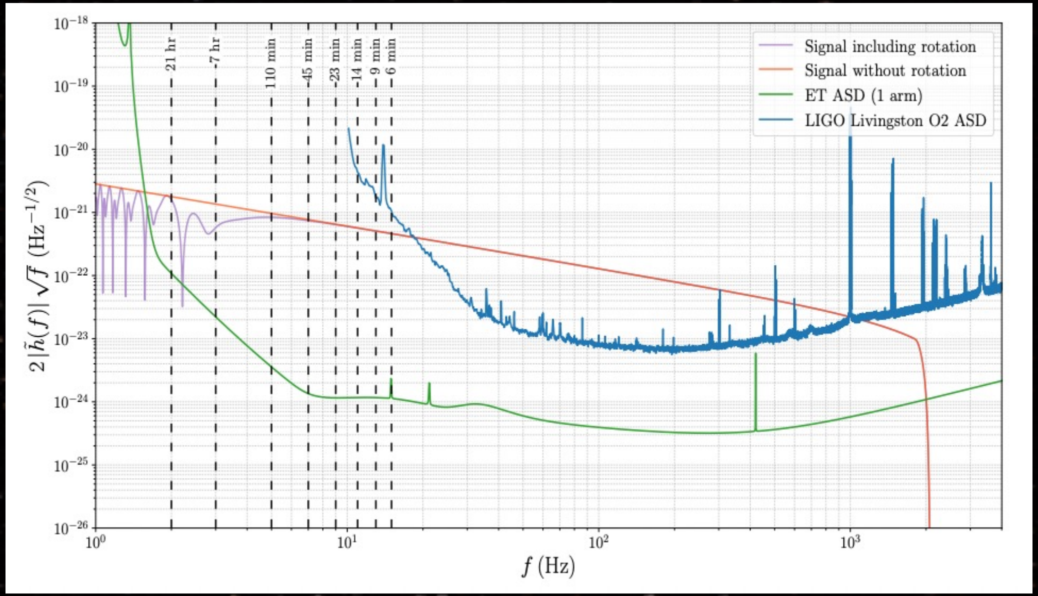
denoising/glitch classification

parameter estimation

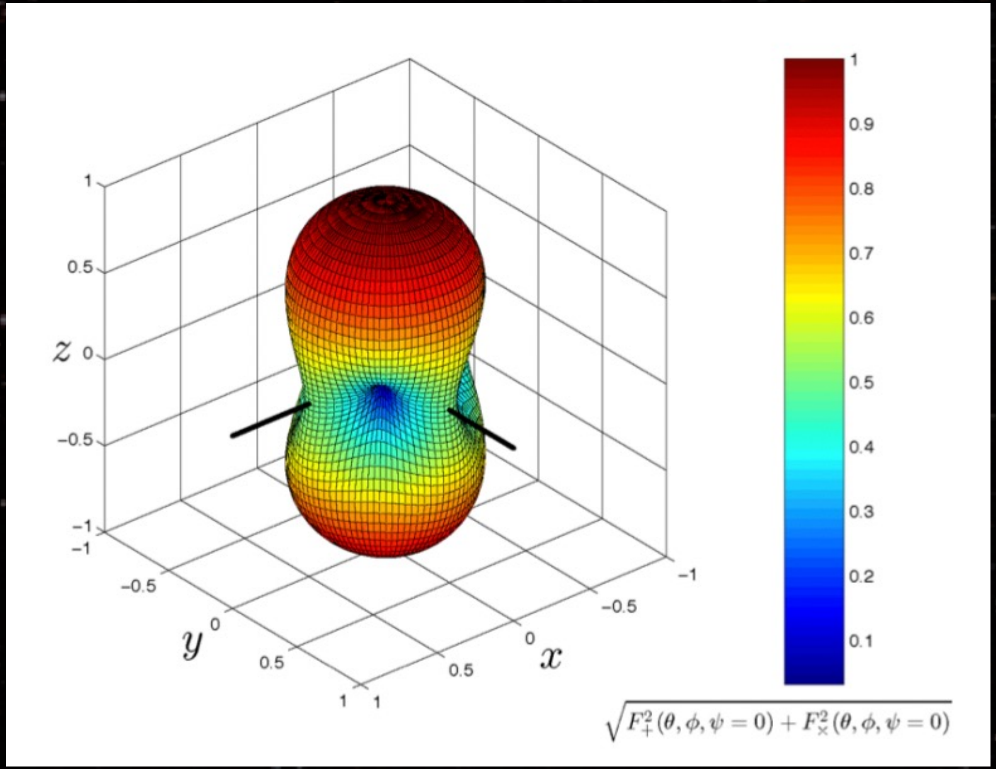
Machine Learning

However, new aspects to be taken into account with ET

- **CBC signals will last up to hours**:
  - need for longer templates → computationally more expensive
  - signal modulations due to the detector motion → templates will depend on the sky position



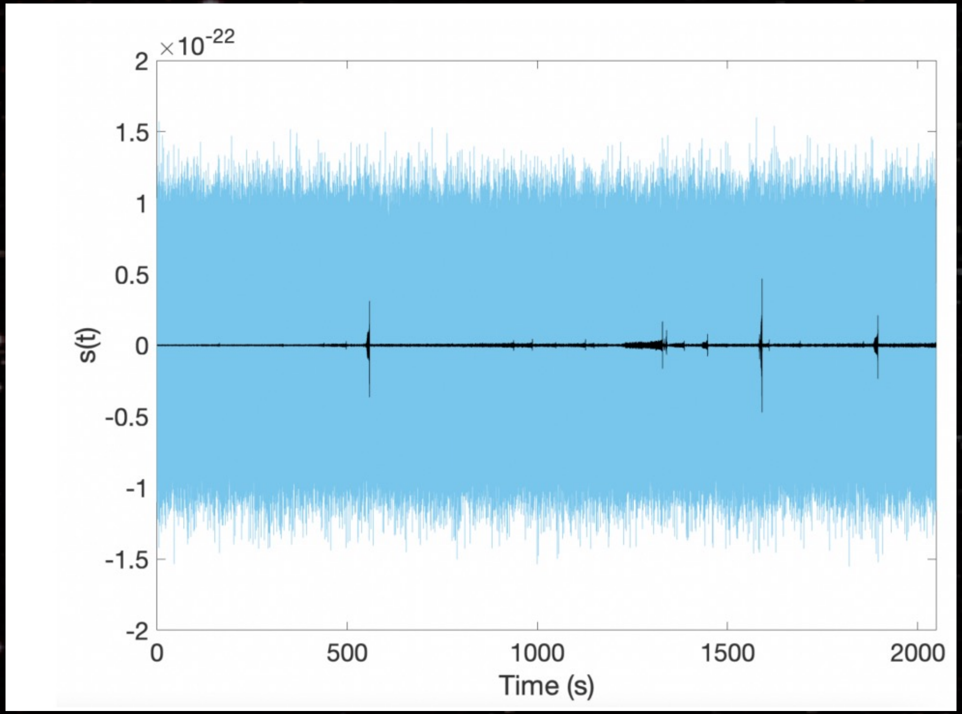
ET 'bluebook', arxiv:2503.12263

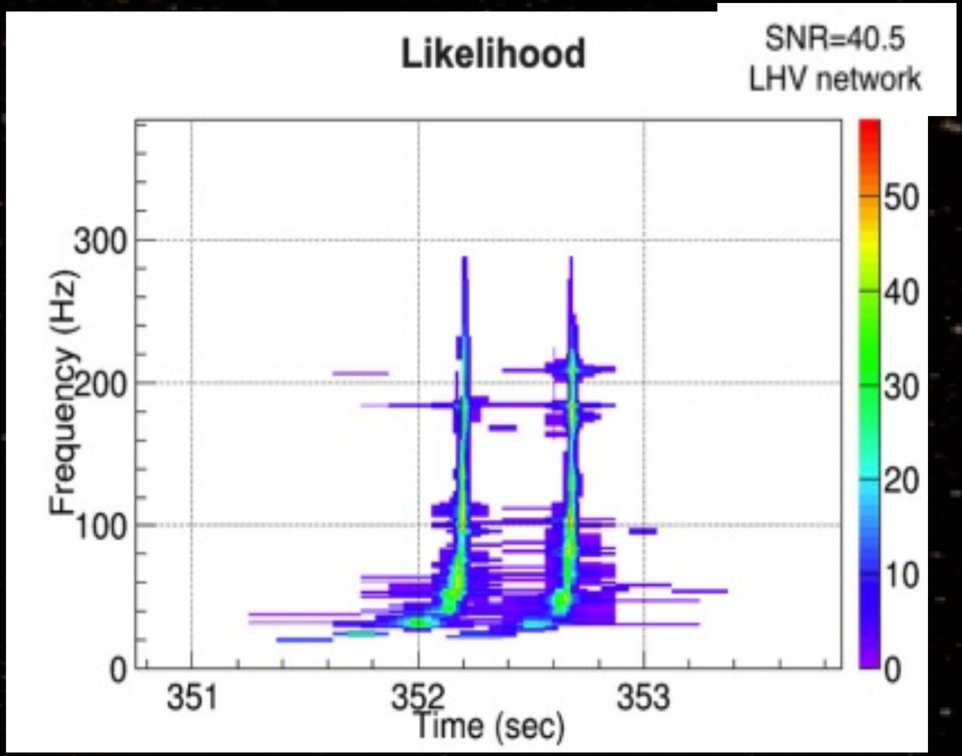


- **O(10^5) signals/yr**. Overall analysis will be harder matched filtering could become unfeasible

❖ Signal superposition:

- ❖ At any instant of time there will > 1 CBC signal overlapping
  - This may affect current analysis methods
  - We need to disentangle overlapping signals





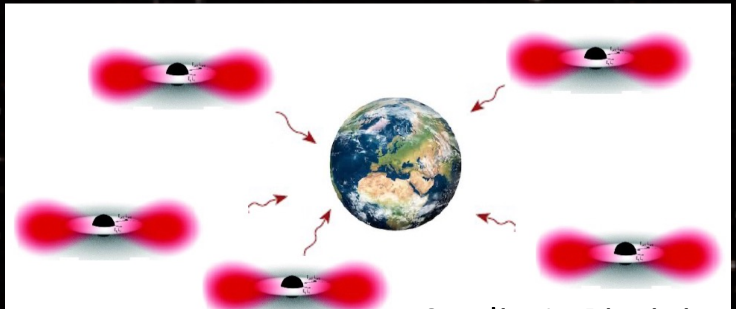
CBC overlap as seen by cWB algorithm
[Relton+, PRD106 104045 (2022)]

❖ It will be particularly relevant for the low frequency part

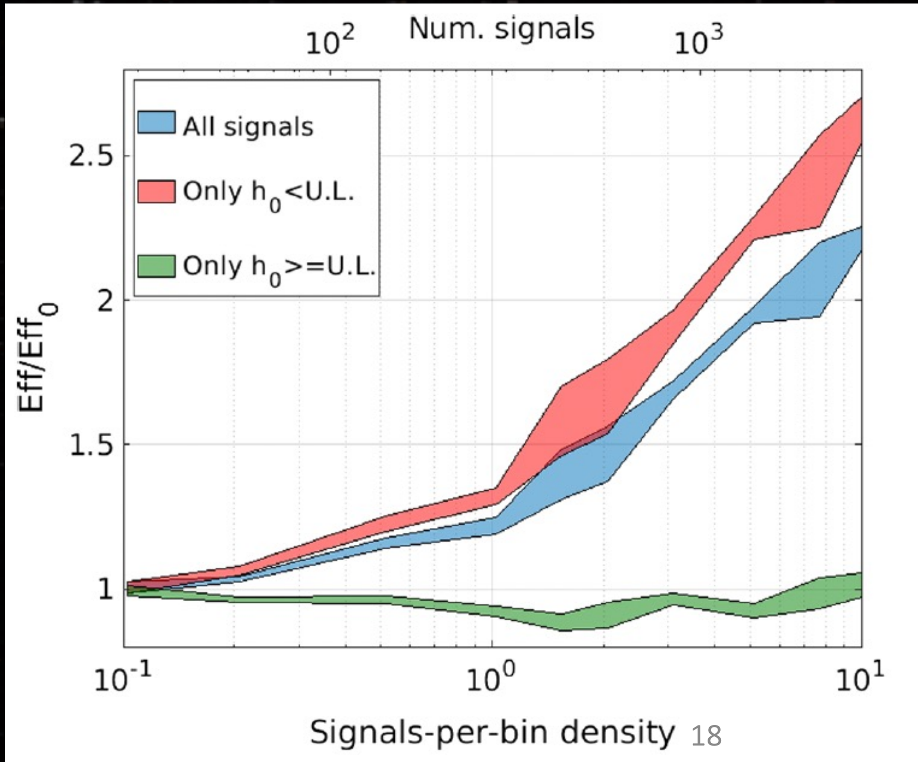
- CBC signal overlapping has also other consequences:
  - may impact on searches for other signals (additional noise background?)
  - may impact on noise power spectrum estimation (no signal-free segment)

- Other - still not detected – sources may produce overlapping signals as well

- E.g. CWs from ultra-light boson clouds around rotating black holes



Credit: L. Pierini



Detection efficiency for overlapping CW signals
[Lorenzo Pierini+, PRD106 042009 (2022)]

Conclusions

Obviously, exciting times are in front of us!

Keep analyzing LVK current and future data (also) in preparation for ET

Analysis results do not come out for magic: we have to invest in young people now in order to have skilled data analysists in the ET era

BACKUP SLIDES

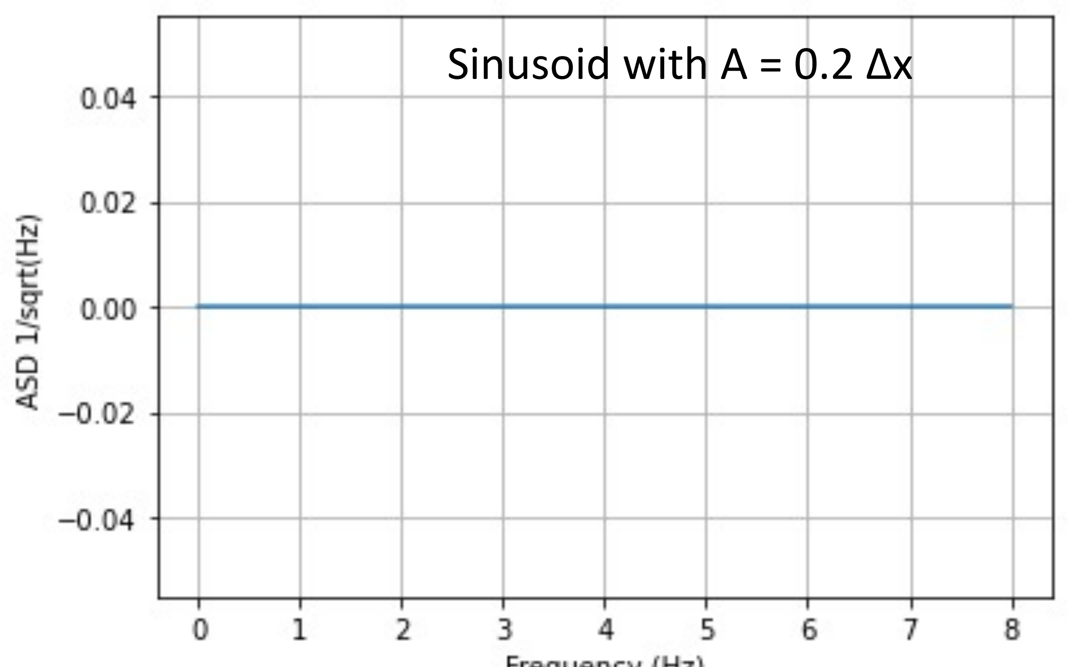
Rome ET RU significantly involved in the **ET Observational Science Board**

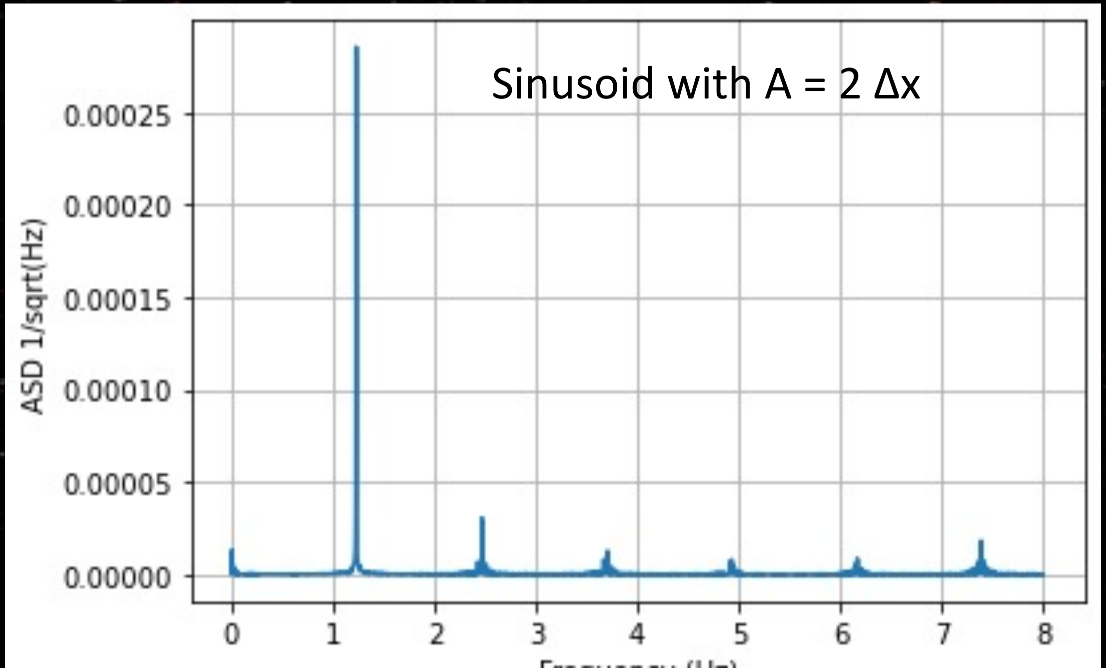
- Coordination roles: Division 1 – fundamental physics (Pani) - and Division 7 - Stellar collapse and rotating neutron stars (Palomba)
- Contributions to **2023 "CoBA" (COst and Benefit Analysis) paper**
- Contributions to the **ET 'bluebook'** (~880 pages, arxiv:2503.12263)
  - Chapter 1: fundamental physics
  - Chapter 2: cosmology
  - Chapter 3: Populations
  - Chapter 6: Subatomic Physics with ET
  - Chapter 7: Stellar collapse and rotating neutron stars
  - Chapter 10: Data analysis
- Now in the process of defining future activities/milestones

Quantization error and ‘dither’ effect

When a continuous signal is digited by an ADC, a quantization error is introduced given by $\Delta x = \frac{V_{max}}{2^n}$

- ✓ If the signal remains within a quantum it will appear as zero in output
- ✓ If the signal amplitude is a bit above the quantum, a spectral distortion appears (due to signal clipping)





The impact of quantization error is reduced by the presence of random noise (dither effect)

…[O]ne of the earliest [applications] of dither came in World War II. Airplane bombers used mechanical computers to perform navigation and bomb trajectory calculations. Curiously, **these computers** (boxes filled with hundreds of gears and cogs) **performed more accurately when flying on board the aircraft, and less well on ground**. Engineers realized that the vibration from the aircraft reduced the error from sticky moving parts. Instead of moving in short jerks, they moved more continuously. Small vibrating motors were built into the computers, and their vibration was called dither from the Middle English verb "didderen," meaning "to tremble." (From Wikipedia – K. Pohlmann, *Principles of Digital Audio*)