Data analysis Platform

ET Symposium, Bologna, 2025

Division 10

(chairs: Elena Cuoco, Gianluca Guidi, Tania Regimbau, Anuradha Samajdar)

Organization

- 4 chairs: Anuradha Samajdar , Elena Cuoco, Gianluca M Guidi, Tania Regimbau
- Liaison persons with other divisions and EIB (Elena Cuoco)
- 162 subscriptions to https://mlist.et-gw.eu/web/info/et-osb-da
- Monthly meetings
- Wiki page: https://wiki.et-gw.eu/OSB/DataAnalysisPlatform

Goal of division 10

- Prepare the analysis and parameter estimation with 3G data (methods and software) in order to exploit the full potential of 3G detectors.
- New challenges: signal regime, long waveforms, overlapping events, new sources.
- New geometry (null stream, correlated noise), new network (+CE).
- New computing technologies (GPU/TPU/FPGA, cloud computing) and algorithms (classical development, IA, Quantum algorithm).
- Synergies with other divisions and EIB

Blue book chapter

Contents

Einstein Telescope data analysis workshop

18–19 Feb 2025
Europe/Rome timezone

Enter your search term

- Overview
- Timetable
- Contribution List
- Places for lunch time
- Registration
- Participant List

Elena Cuoco
elena.cuoco@unibo.it

(image created with chatgpt)

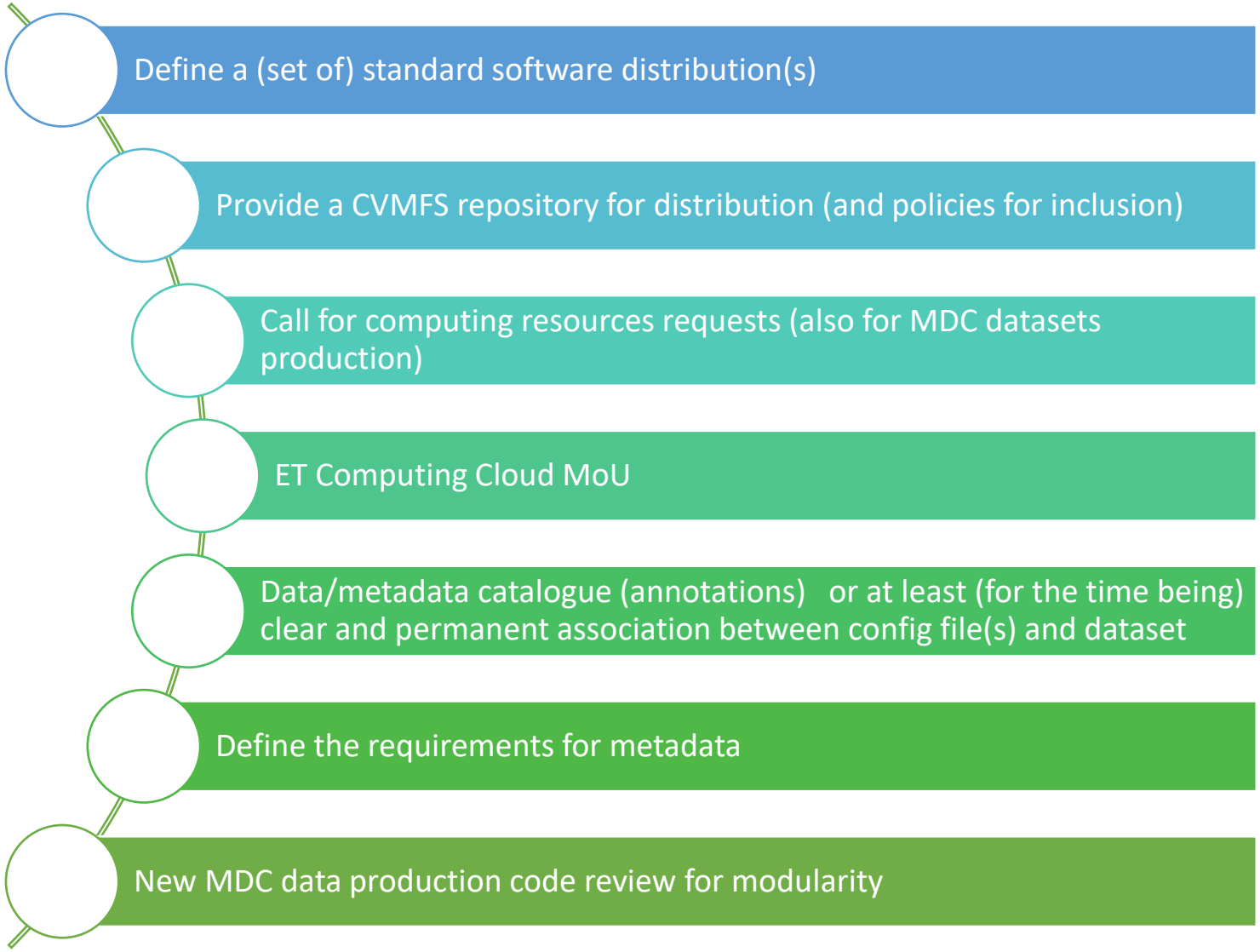
This event is designed as a collaborative meeting for individuals involved at various levels in the data analysis of the Einstein Telescope, as well as in the development of the infrastructure for data analysis and generation.

Einstein Telescope data analysis workshop – Bologna 18-19/02

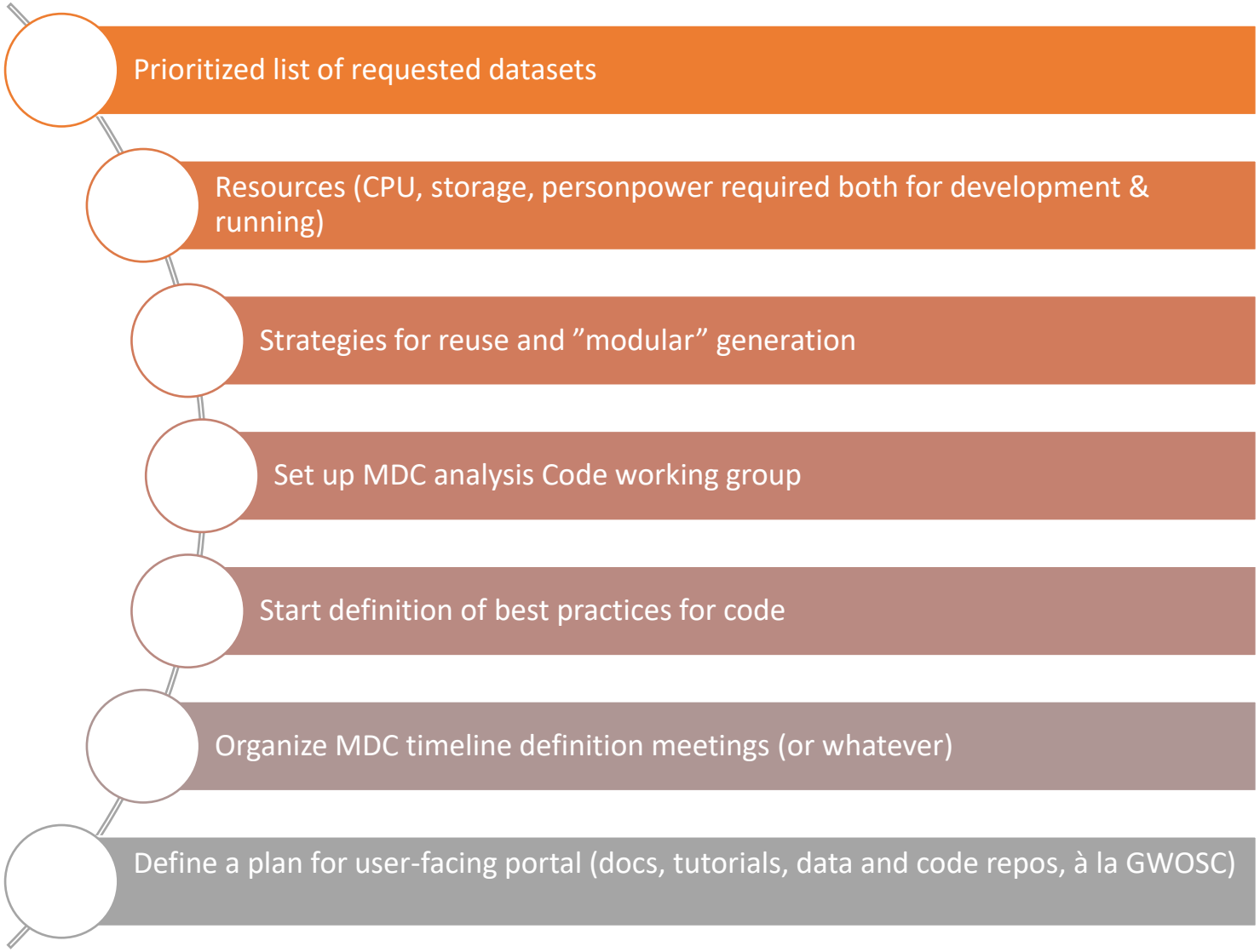
- MDC and EIB
- MDC task force for simulation purposes
- Open Discussion

- OSB division presentations
  - F. Santoliquido:Fast and accurate parameter estimation of high-redshift sources with the ET
  - A. Iess: Running a GW analysis on the ESCAPE VRE
  - L. Lavezzi et al.: CTLab4ET, MADDEN-Multi-RI Access and Discovery of Data for Experiment Networking
- Seminar. :
  - G. Inguglia: Gravitational Waves: A New Frontier Across Scales and Disciplines
- MDC hands on – Snakemake, EIB
- Open Discussion

BOLOGNA f2f


Define a (set of) standard software distribution(s)
Provide a CVMFS repository for distribution (and policies for inclusion)
Call for computing resources requests (also for MDC datasets production)
ET Computing Cloud MoU
Data/metadata catalogue (annotations) or at least (for the time being) clear and permanent association between config file(s) and dataset
Define the requirements for metadata
New MDC data production code review for modularity

BOLOGNA f2f


Prioritized list of requested datasets
Resources (CPU, storage, personpower required both for development & running)
Strategies for reuse and ”modular” generation
Set up MDC analysis Code working group
Start definition of best practices for code
Organize MDC timeline definition meetings (or whatever)
Define a plan for user-facing portal (docs, tutorials, data and code repos, à la GWOSC)

MDC1

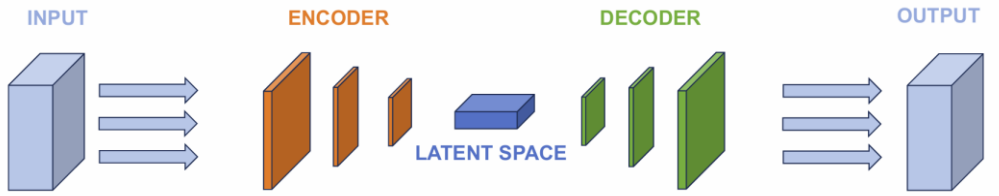
- Data produced by OSB-div 10 with a revised version of a code developped in 2012
  - ET-0148A-25*: A mock data challenge for next generation detectors*
- Data contains colored Gaussian noise+ GW signal from CBCs for ET triangle.
- Challenges include detection and parameter estimation in the presence of a large number of long and overlapping sources.

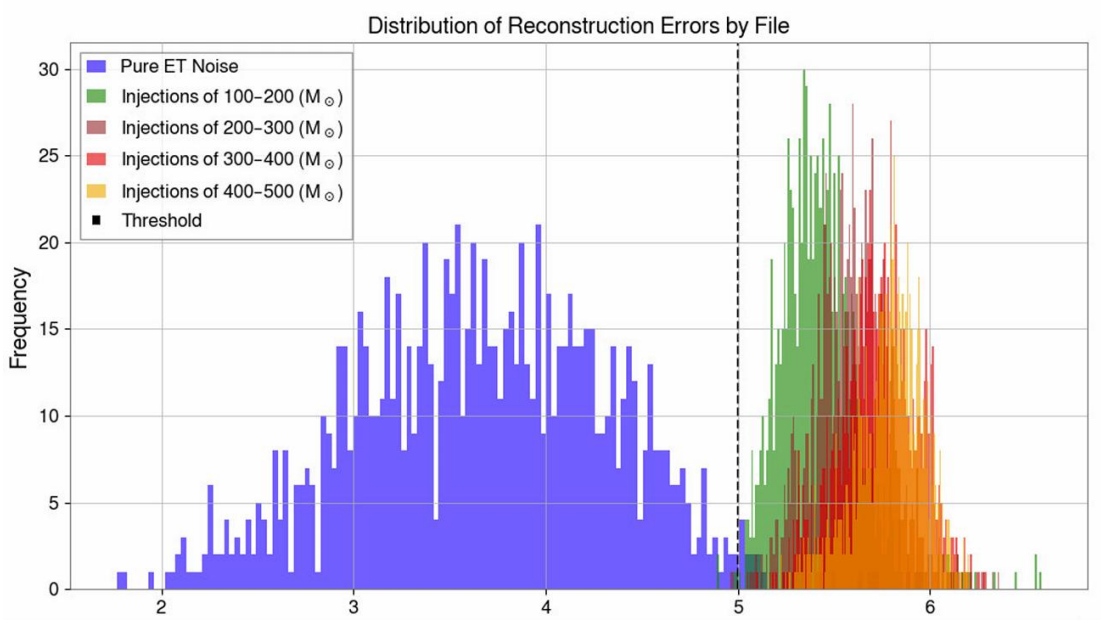
- **In fieri**:
  - *EIB is rewriting the existing code, we are waiting for access to the code after which we will start review.*
    - *Div10 chairs have a checklist to test the new code.*
    - *A review team will be in charge of comparing the outputs of the two codes.*

MDC 1 analysis

- *On the use of convolutional autoencoders to search for GWs via anomaly detection* Vitulova, Inguglia

A GW signal can be seen as an anomaly in the noise; hence, anomaly-detection algorithms might help in detecting the signal, irrespective of the signal shape



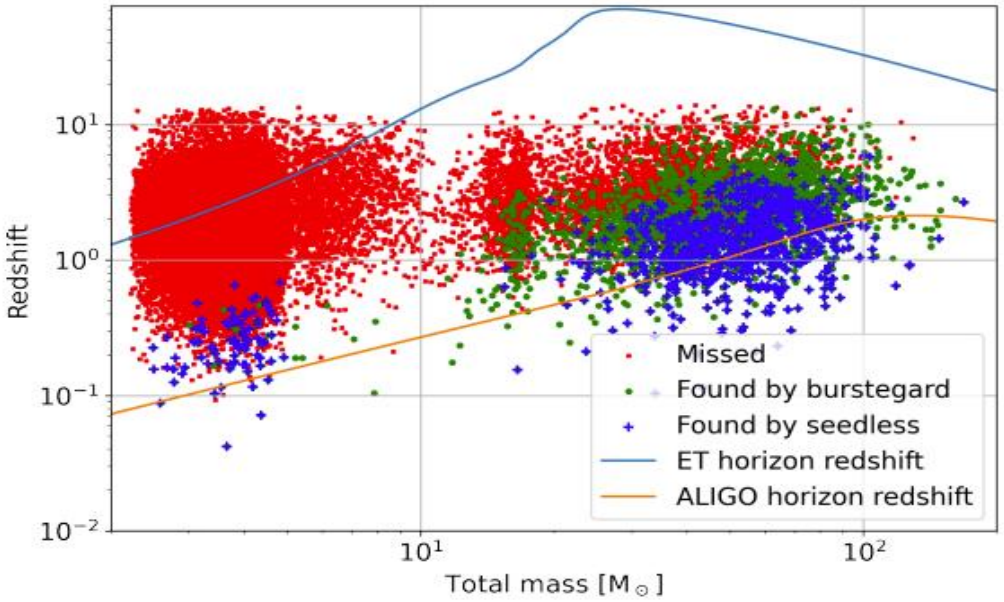


MDC 1 analysis

Weakly modeled search for compact binary coalescences in the Einstein Telescope
Macquet, Dal Canton, Regimbau - https://arxiv.org/abs/2408.13007

- Use a detection algorithm that does not rely on the waveform of the signal searched (PySTAMPAS)
  - less sensitive than matched filtering techniques, but computationally much cheaper

- 38% of the total number of BBH, including 89% of the systems with a total mass above 100 $M_\odot$
- the majority of BNS mergers closer than 850 Mpc (z = 0.17).
- PSD estimated via *null-stream*



MDC2 plans from Div10

- **In fieri** …..
  - Different PSDs and networks (triangle and 2 Ls)
  - Glitches, correlated noise, calibration errors, missing data, bursts, CWs (poster), cosmological sources.
  - Form groups of experts for each new signal.
  - Div 10 will proceed with the actual code if the new one is not ready.
  - The data for MDC2 will be possibly released at the Annual meeting in November 2025.

- EXTRA SLIDES

Challenges

With a better sensitivity and a lower frequency bound, new problems will appear:

- How do we deal with long duration signals?
- How do we separate overlaping signals?
- How do we estimate the noise background in the continuous presence of sources?
- How do we subtract the correlated noise and the astrophysical background?
- …

Limitations of actual methods

- **CBC**: based on match filtering

- Large template bank, Earth modulation, glitches, noise stationarity, calibration

- ML and GPUs, classification noise/noise+signal

- **GW Background:** based on cross-correlation

- XG break the assumption that noise>>signal

- account for the signal variance, null stream

- **CW**: semi coherent searches

Same problems as now but also superposition of a large number of CBCs (and CWs)

- **Burst:** coherent unmodelled searches + MLA (CNN, XGBoost)

Parameter Estimation

With an increase number of signals and their duration, speed up of PE bayesian analysis is crucial. Another challenge is to deal with the superposition of sources (multiple likelhood).

New routes are being explored:

- Hierarchical subtraction (error residual)
- Speed up of waveform calculation: reduced-order-methods, likelihoof-free inference, relative binning
- Machine learning: Neural posterior estimation, truncated marginal neural ratio estimator, nested sampling with AI, normalizing flows, MCMC, deep neural network driven hamiltonian MC…

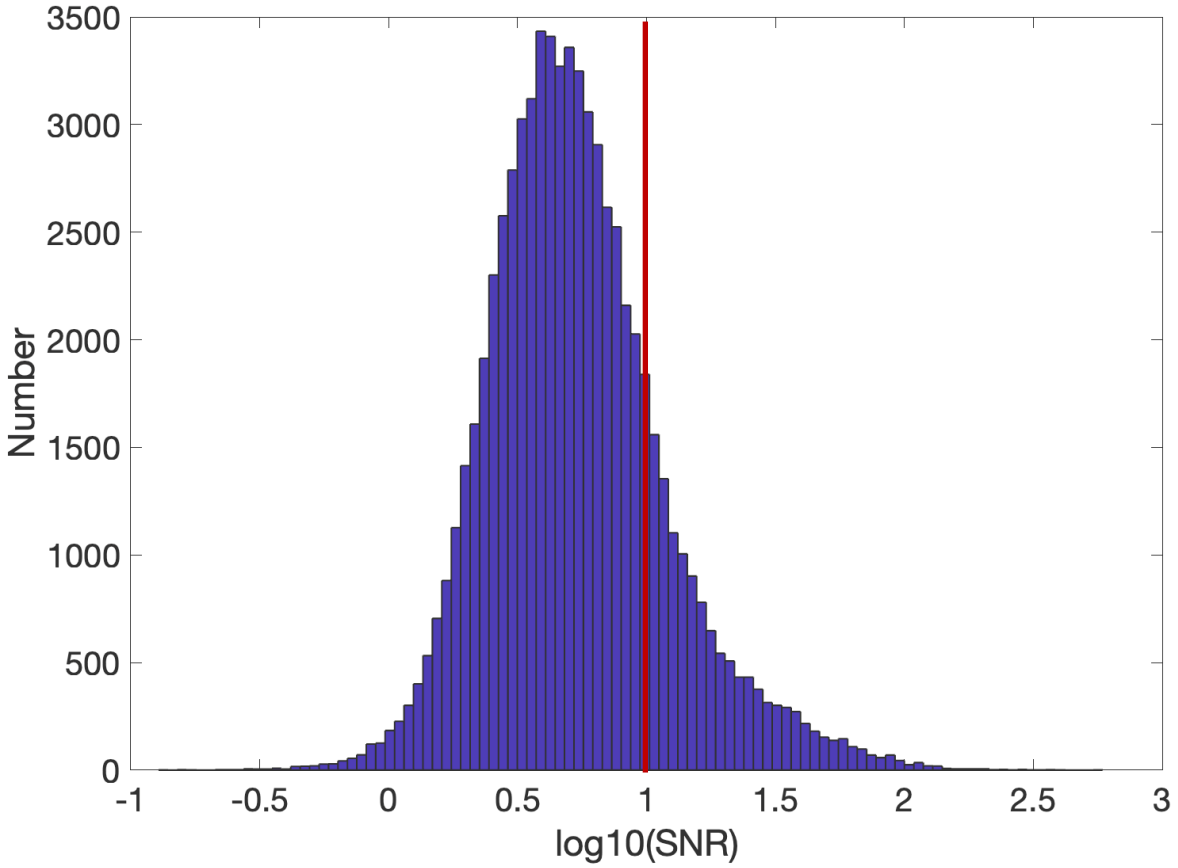
Null stream

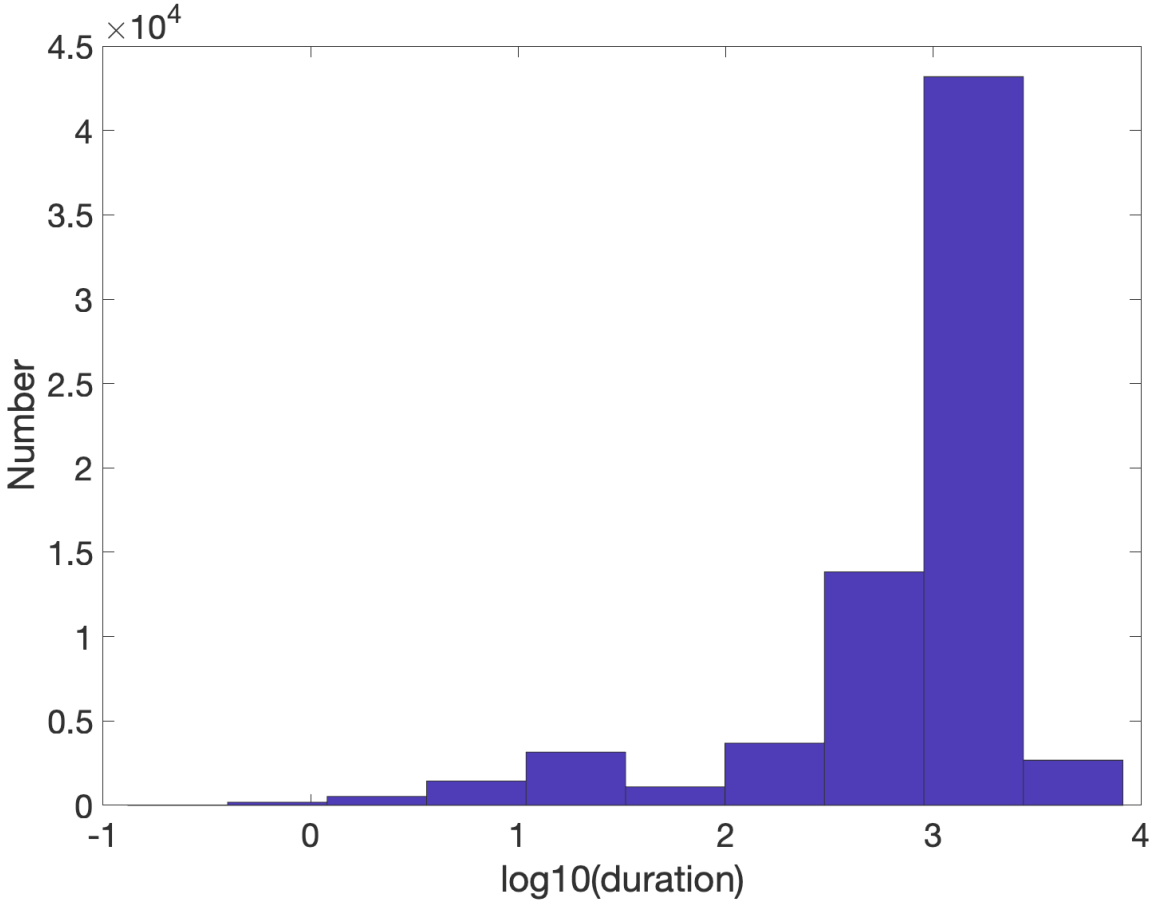
The Einstein Telescope in its triangle configuration provides a channel that cancels the GW signals for all directions:

$$ns = s_{E1} + s_{E2} + s_{E3} = n_{E1} + n_{E2} + n_{E3}$$

- unbiased noise PSD estimation
- GW background
- glitch identification
- self calibration (leakage of signal in the null stream)
- correlated noise

Some statistics (MDC1)





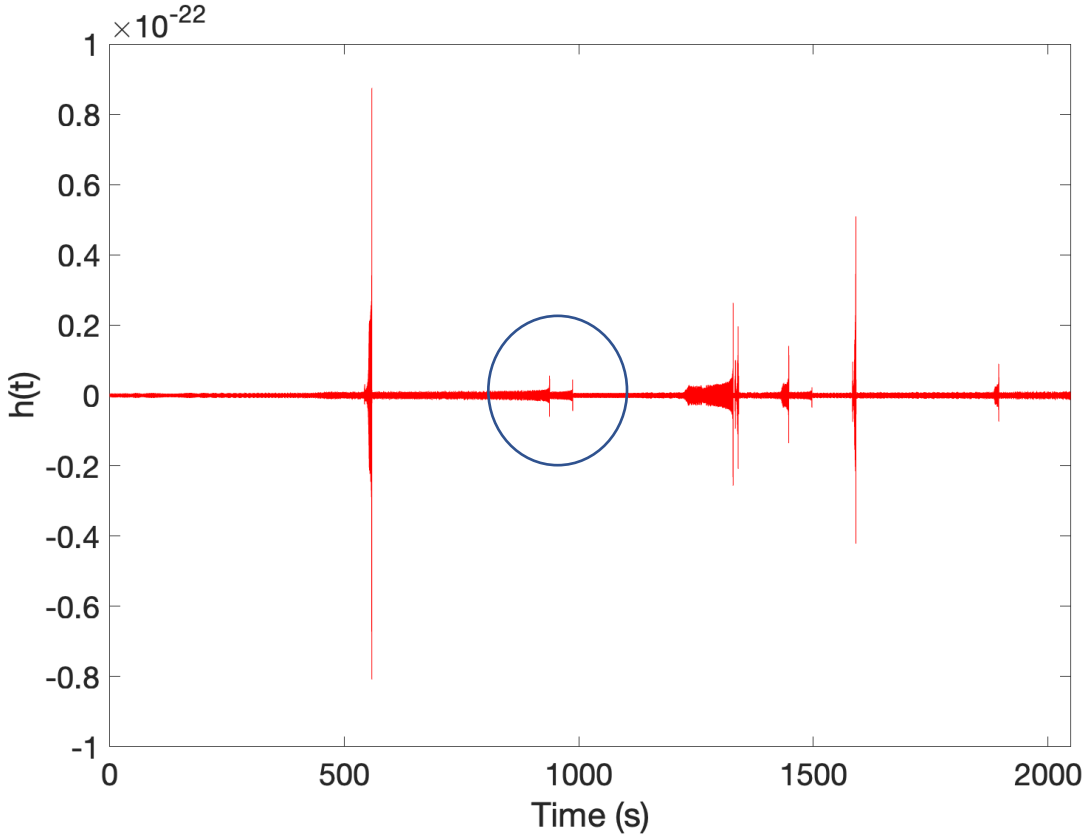
SNR>8: 11551 BNS , 537 BHNS, 6119 BBH

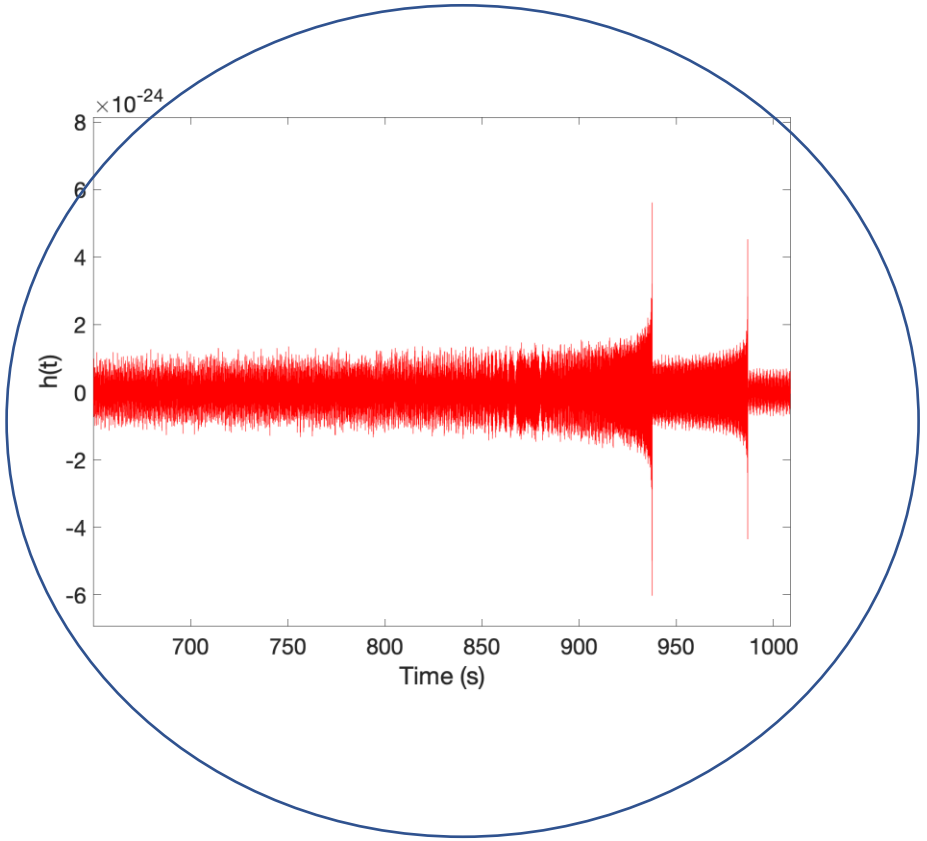
SNR>12: 4048 BNS, 238 BHNS, 5228 BBH

Example of the GW data (MDC1)

28 signals with SNR>6, largest at SNR=85

BNSs merging at 937s and 986s, that are both long duration and overlapping





Challenges (MDC1)

Beginner

- Recovery of high-SNR signals within given time windows

SNR = 597, 386, 383 (BNS), 374, 343, 306

Expert

- Parameter estimation of ultra-high SNR BBH signals
- Long duration binary neutron stars
- Overlapping signals