Computing Challenges for the Einstein Telescope project

Dr. Silvio Pardi on the behalf of ET Computing Group

CHEP 2023 - Norfolk - USA

8 May 2023

The Einstein Telescope

The Einstein Telescope (ET) will realise the European 3rd Generation Gravitational Wave observatory

- A new infrastructure capable to host future upgrades for decades without limiting the observation capabilities
- A sensitivity at least 10 times better than the (nominal) detectors of second generation (i.e. aLIGO and AdV) on a large fraction of the (detection) frequency band
- Large improvement in sensitivity in the low frequency (few Hz – 10Hz) range

The ET Collaboration

- The ET Collaboration was formed on 8.6.2022 @ XII ET Symposium Budapest
- 80 Research Units
- Ca. 1250 members

https://agenda.infn.it/event/30175/contributions/179015/attachments/97051/134056/ET_VulcanoWorkshop_Sept2022-Punturo.pdf

Detection distance of GWD

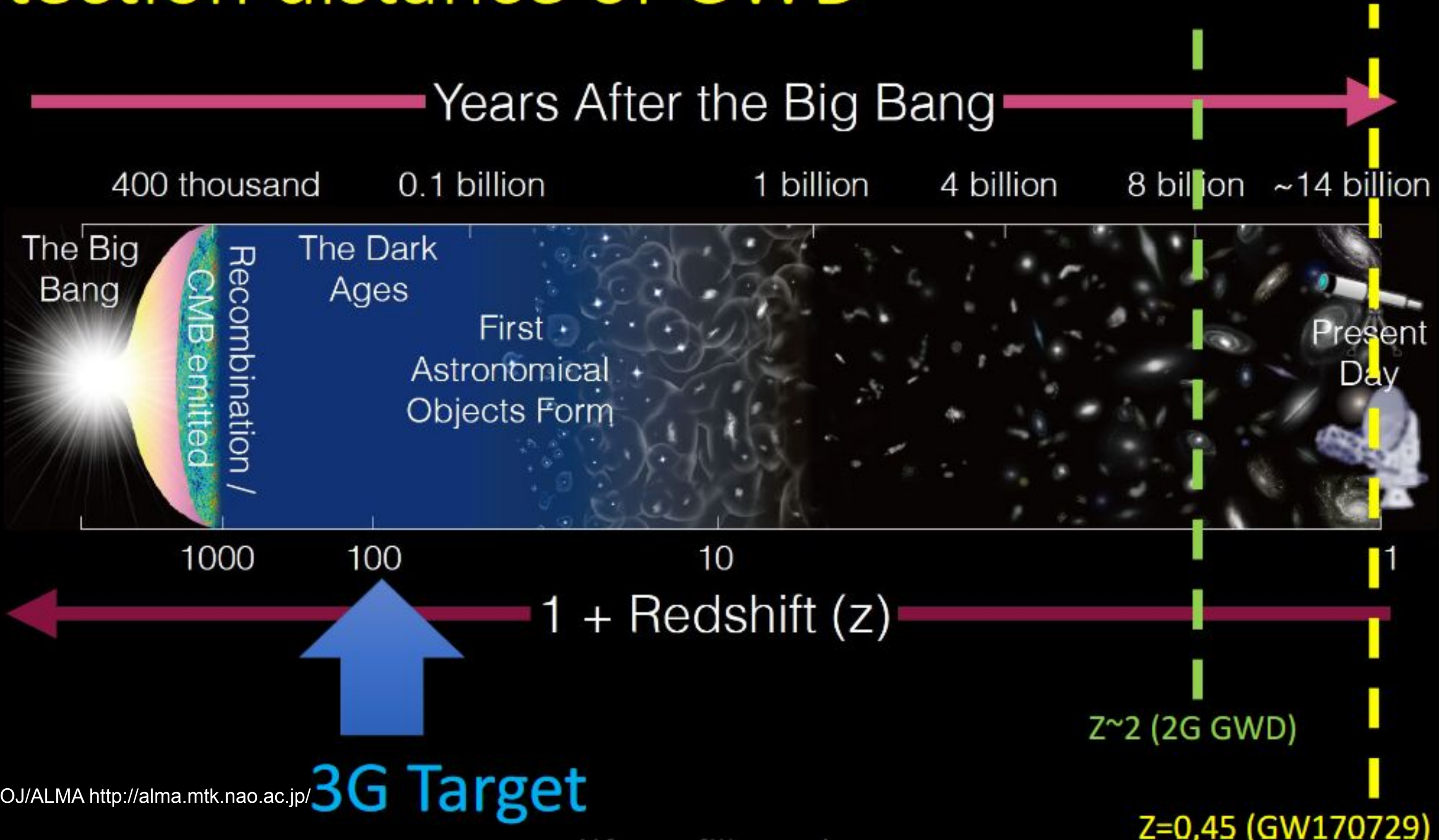


Image credit: NAOJ/ALMA http://alma.mtk.nao.ac.jp/

Main sources of Gravitational Wave

- Burst sources:
  - CBC: Compact Binary Coalescence
  - Burst: Unmodeled transient bursts
- Continuous sources:
  - CW: Continuous waves
  - SGWB: Continuous stochastic background

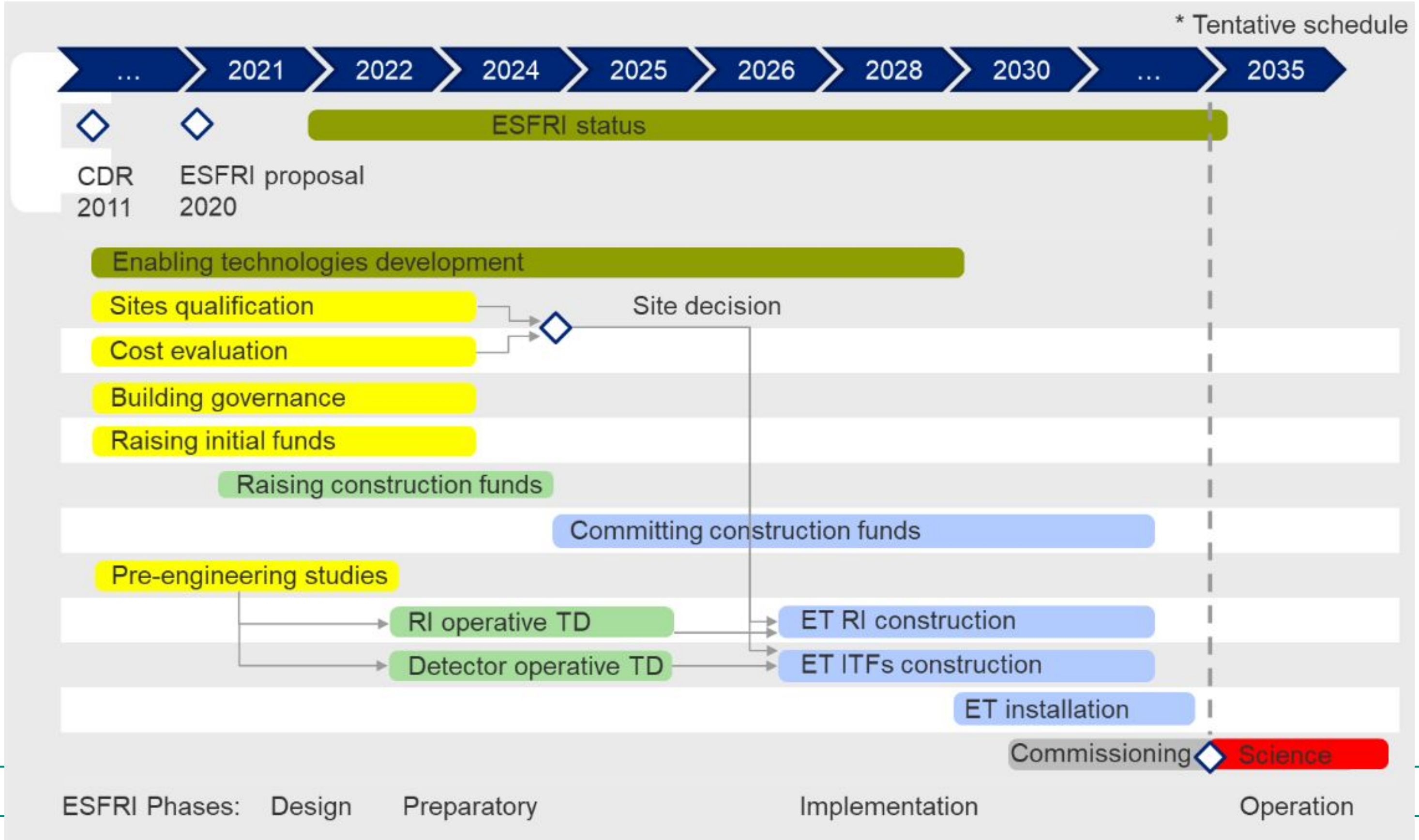
The Einstein Telescope

Two candidate site to host ET:
• The EU Regio Rhine-Meusse site, close to the NL-B-D border
• The Sardinia site, close to the Sos Enattos mine

• A third option in Saxony (Germany) is under discussion

Ongoing site characterization in term of:
• Seismic noise measurements
• Magnetic and ambient noises
• Geophysical and geotechnical characterizations

ET Time Schedule



ET Computing Domains

- Online
  - Data acquisition and pre-processing, Instrument control, Environmental monitoring
- Offline
  - Deep searches, Offline parameter estimation, (Template bank generation)
- Low-latency
  - Candidate search, Sky localization, LL parameter estimation, Alert generation and distribution

ET Data

The source of Data that will be collected at the ET site can be classified in two main classes:

- Interferometers data
  - sapling rate, resolution, channels per interferometer, number of interferometers data headers (rel. to data)
- Monitoring data
  - Seismic sensors, Vacuum sensors, Temperature sensors, Position & tilt, Cryogenics, Lasers

The expected amount of RAW data collected by ET is on the order of few tens of PB of raw data per year (baseline 6- interferometer design, more control channels)

However, in order to extract the increasing amount of useful scientific information encoded in the data, large computing power are needed.

In the 3G era the challenge will be in the computational power needed to extract the science content from the data.

Computing Power

The demand for data analysis computing in the 3G era will be driven by the high number of detections (up to hundreds per day), the expanded search parameter space (0.1 solar masses to 1000+ solar masses) and the subsequent parameter estimation (PE).
In addition given that GW signals may be visible for hours or days in a 3G detector (as opposed to seconds to minutes in the 2G era).

Approximately **1 million CPU core-hours** of computing were used for parameter estimation of each signal in the Ligo Virgo Collaboration (second observing run O2)

A naive scaling of current matched-filter CBC searches and PE to 3G design sensitivities will require many orders of magnitude more CPU and RAM than 2G the order of:

O(1 million CPU hours) * O(10^5 signals/year)

without considering the increasing complexity of overlapping signals, higher precision waveforms, and other factors that can increase the needed CPU power.

ET Software and Technologies

For ET in 10-15 years from now several technologies will be needed to design the telescope itself and to reduce computational costs for signal detection and parameter estimation and win the challenge to extrapolate information from the collected data. Among them:

- Machine Learning/AI algorithms.
- High Performance Computing.
- GPU Computing.
- Digital twins to describe a fully digitised copy of the planned facility.

Multimessenger Analysis

ET will also be part of the Multimessenger Ecosystem which includes a large number of existing and future facilities that will produce and consume alert triggers.

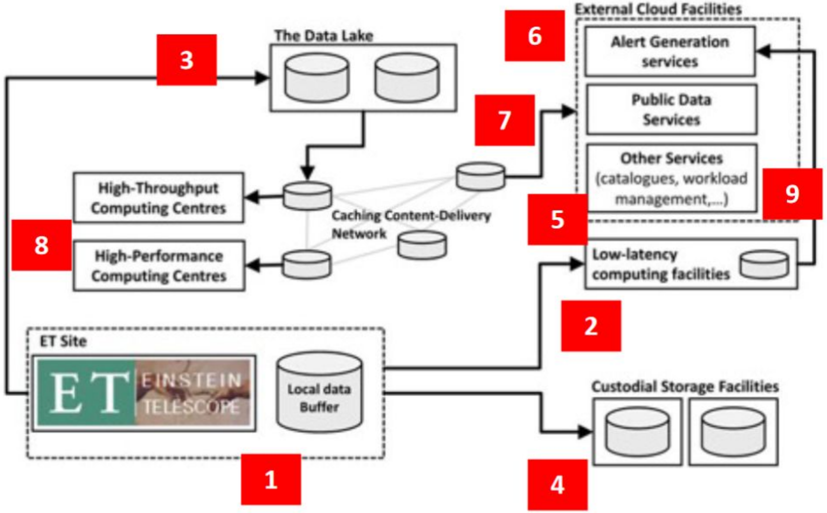
- Ground-based optical telescopes (like the Extremely Large Telescope- ESO on Cerro Armazones, Chile)
- Large radiotelescope arrays, like the SKA,
- Facilities for cosmic ray astronomy, like the Cherenkov Telescope Array:
- Cherenkov telescopes for highest-energy gamma-ray astronomy, in the Canary Islands and Chile (CTA, AUGER)
- Neutrino detectors, like KM3NeT

ET Computing Model

To accomplish all the task, ET community plans to define a computing model following the paradigm of distributed system since the beginning, taking advantage of the experience of large international experiments.

In particular the usage of middleware components developed within HEP experiments, and in the context of WLCG, OSG, EGI, EOSC, ESCAPE guaranteed the highest level of sustainability, helping to have a stable and reliable infrastructure reducing costs.

ET Computing Model

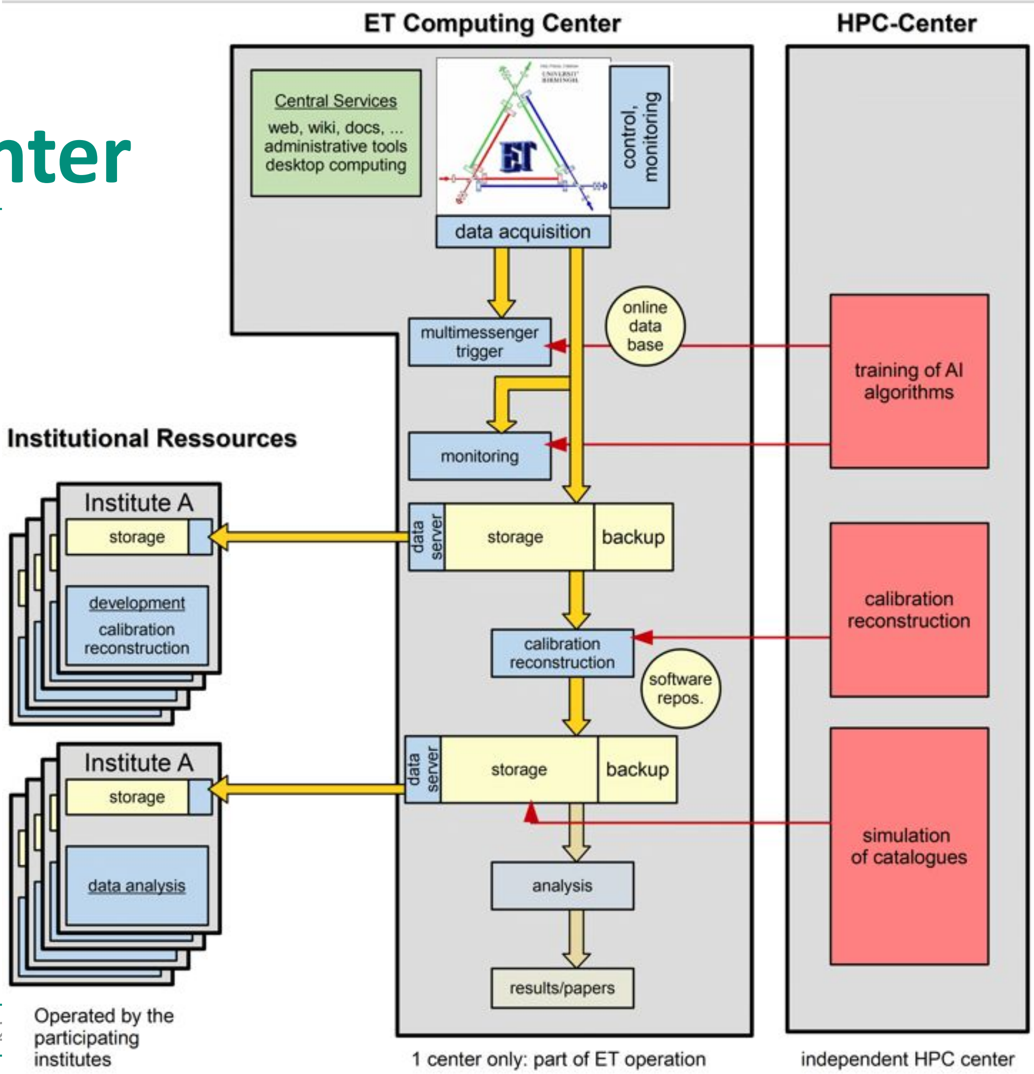


1. Data collection in a local circular data buffer. A local computing infrastructure is used to pre-process and reduce the data to the format used for low-latency and offline analyses.
2. Data are transferred to the low-latency search facilities, where search pipelines are automatically run.
3. Data are also shipped to the Data Lake for subsequent offline analyses.
4. A reduced version of raw data are transferred to archival sites for safekeeping.
5. All data (raw, processed, public) is registered in a general catalogue database that functions as a single front-end both for data discovery and access
6. Low-latency processing facilities run search pipelines and send triggers and candidate information to the low-latency alert generation and distribution services
7. Candidate event alerts are generated and distributed by the relevant services. Data segments to be distributed with the alert are not copied again but tagged in the database as “public”
8. Offline analyses (parameter estimation, deep searches etc.) and all scientific computing (numerical relativity simulation, Machine Learning model training, etc.) are run on available HTC, HPC or even “Big Data” facilities optimised for Machine Learning, depending upon the optimal type of technology.
9. Publicly released data are not copied again, but tagged in the database and made available through public discovery and access services.

ET local computing center

The ET local computing Center will be responsible for data acquisition, calibration, on-line analysis and data distribution.

It will operate on the ET site, and will be connected to the other computing facilities.



The International Gravitational-Wave Observatory Network (IGWN) Computing Grid

ET will develop the computing model also in connection with the International Gravitational-Wave Observatory Network (IGWN) initiative.

IGWN Computing Grid represent the current distributed, high-throughput computing (DHTC) infrastructure for Gravitational Wave community, providing dedicated and opportunistic computing resources in partnership with the Open Science Grid.
The Grid infrastructure is underpinned by these key technologies:

- HTCondor
- The Open Science Grid
- The Open Science Data Federation
- and many more.

ET Mock Data and Science Challenge

The design of the future computing model of ET is integrated with “The ET Mock Data and Science Challenge” a simulation activity with the following goals:

- produce realistic simulated data containing instrumental noise + GW signal, with increasing complexity: first with Gaussian colored noise and a cosmological population of CBCs (more details in the next slide) next with all type of sources, glitches, correlated noise.
- Test, develop, optimize data analysis pipelines and parameter estimation
- Stress-testing the current computational infrastructure

Conclusions

The Einstein Telescope (ET) will realise the European 3rd Generation Gravitational Wave observatory providing high enhanced sensitivities to GW detection, enabling a very rich scientific program in astrophysics and fundamental physics.

ET has to tackle many challenges in the fields of computing and algorithms development in order to meet its analysis needs.

To approach those challenge a new Computing model is going to be designed in order to use distributed computing paradigm since the beginning.

The project timescale also allows to start innovative R&D to develop advanced technologies for the 2030's