Finanziato dall'Unione europea
NextGenerationEU

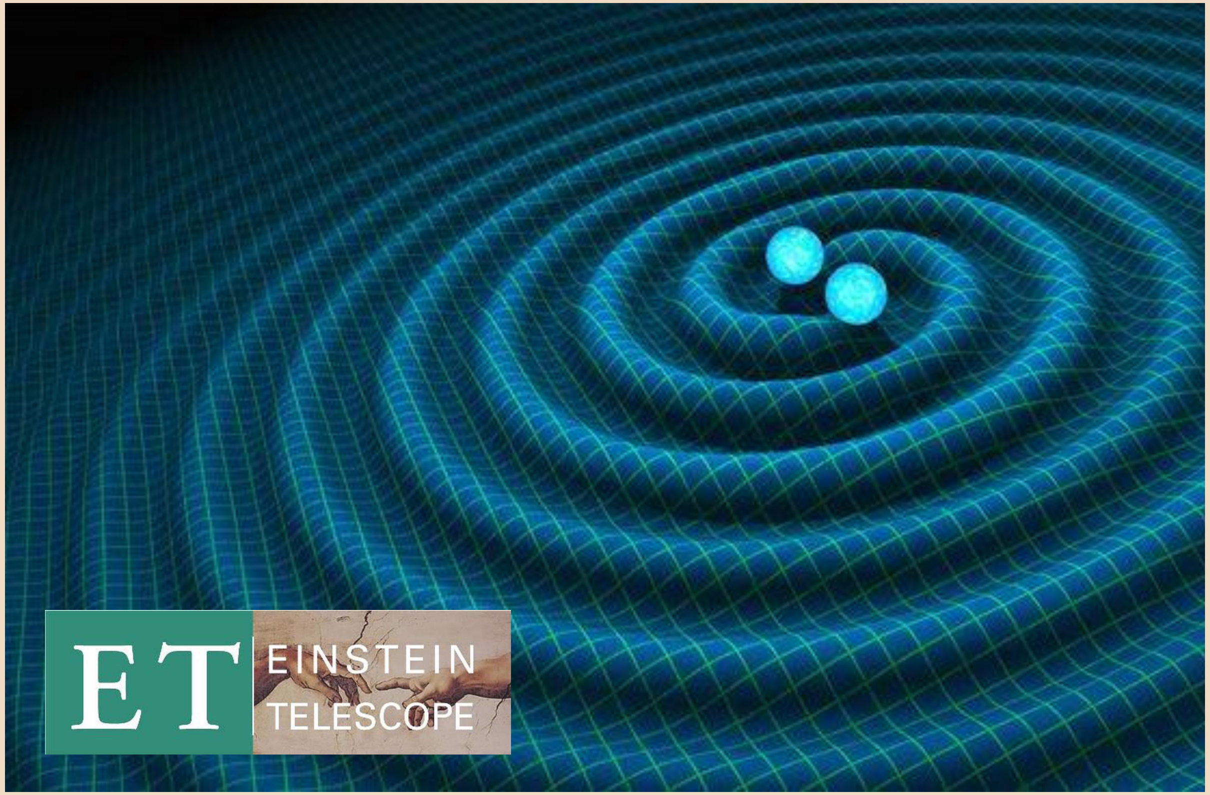
Ministero dell'Università e della Ricerca

Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA

INFN

HTCondor in Einstein Telescope

Evolving Computing for Gravitational Wave Research



Luca Tabasso, et al.

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
INFN

Introduction: What is the Einstein Telescope?

The **Einstein Telescope (ET)** is a 3rd gen gravitational wave observatory, designed to be **ten times more sensitive** than current detectors like LIGO and Virgo, that will delve into **frequencies below 10 Hz** expanding the observable spectrum.

Goals:

- Detection of around 10^5 Compact Binary Coalescences systems per year.
- **Early warnings** of detections, **~ hours to minutes before** events.
- Accurate **sky localization** facilitating multi-messenger astronomy.

Computational Domains

Online (on-site infrastructure):

- Real-time data acquisition
- Instrument control, environmental monitoring, …

Offline (HTC and some HPC):

- deep searches
- offline **parameter estimation**
- **template** bank generation

Low Latency:

- Sky **localization**
- LL **parameter estimation**
- **Alert** generation and distribution

Technological Challenges

Volume of Raw Data:

- Raw interferometer data **does not grow significantly** with increased sensitivity.
- Expectation of **a few tens of petabytes (PB)** of raw data per year (today: ~2 PB/year).

Growth of Scientific Information:

- The amount of **useful scientific information** extracted from the data grows substantially.
- Requires increased **computing power**, especially for CBC parameter estimation.

ET Project Timeline

- **2008**: Conception of the Einstein Telescope idea.
- **2011**: Publication of the Conceptual Design Report (CDR).
- **2020**: CDR update with new technical and scientific insights.
- **2021**: ESFRI approval, recognizing ET as a strategic project.
- **2022**: Establishment of the ET Collaboration.

Today

- **2026**: Expected site-selection.
- **2028**: Expected start of construction.

(...)

- **2035**: Expected start of operations.

Current computing status: embryonic development phase

Current official services:

- Data distribution for **Mock Data Challenges** (using the Open Science Data Federation and CVMFS-for-data)
- **GitLab** for code management

Future data distribution infrastructure:

- **Data Lake** managed by **Rucio**

Future data processing:

- **Not yet defined**
- **Not in the Dark**: we have ideas but no urgent need to define now
- Opportunity: **time to explore** different solutions and tools

Lessons from LIGO and Virgo - A Bit of History

LIGO and Virgo Beginnings:

- Started with modest computing needs
- Used **HTCondor from the start**, at least for some workflows
- Data processing on **dedicated clusters** (LIGO, CIT, EGO, CNAF, IN2P3), pipelines often tied to a **specific site configuration**

HTC:

- In GW and HEP, we use High Throughput Computing
- Run the **same job on multiple input data files**
- No need for tightly coupled parallel jobs as in High Performance Computing (HPC)

Finanziato dall'Unione europea
NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA
INFN

Scaling Up - The Need for Distributed Computing

Increased Sensitivity and Interest

- After the discovery of the **first GW event**

Growth in Computing Needs:

- Substantial increase in data processing requirements and in the size of the collaborations
- A clear necessity to move towards **shared resources** wherever possible

Adoption of Distributed Computing:

- Formation of the IGWN computing initiative
- **Imported grid computing model** (but not all the tools) from LHC experiments

The Grid Computing Model and HTCondor in GW

HTCondor:

- Used as a **distributed batch system** to run jobs on the grid

Cluster Diversity:

- Each cluster can have its own batch system (Slurm, PBS, HTCondor)

glideinWMS:

- Submits 'glideins' or 'pilot jobs' via Computing Element (CE)
- Pilot jobs start HTCondor startd processes
- Users submit jobs to the grid using this HTCondor system

Challenges Faced by LIGO-Virgo

Pipeline Design Issues:

- Analysis **pipelines not designed to run on the grid**
- E.g., relied on a **shared file system between cluster and submission node**

Resulting Limitations:

- Only ~22% of LIGO-Virgo offline jobs run on the IGWN grid today, but the effort is ongoing and this is growing

ET's Approach - Starting from IGWN Model and Evolving

Exploring New Tools: investigating Snakemake on top of HTCondor

- framework for **reproducible data analysis**
- ensure that our workflows are **grid-compatible**
- used in REANA (a tool that is part of the Virtual Research Environment developed within ESCAPE / EOSC for open science)
- runs **jobs on various computing resources**:
  a. HTCondor clusters (e.g., IGWN grid)
  b. Slurm clusters (e.g., EU HPC clusters)
  c. Cloud

Conclusion

ET's Opportunities

- Learn from LIGO-Virgo experiences and errors
- Explore and test different solutions and tools
- Ensure grid-compatibility of jobs.

For any questions or further information, please contact

Stefano Bagnasco
(e-Infrastructure Board co-chair)

Andres Tanasijczuk
(e-Infrastructure Board - Division1: software, frameworks, and data challenge support chair)