Institut national de physique nucléaire
et de physique des particules

Einstein Telescope: Data & Computing

Réunion presentation ET à l'IP2I

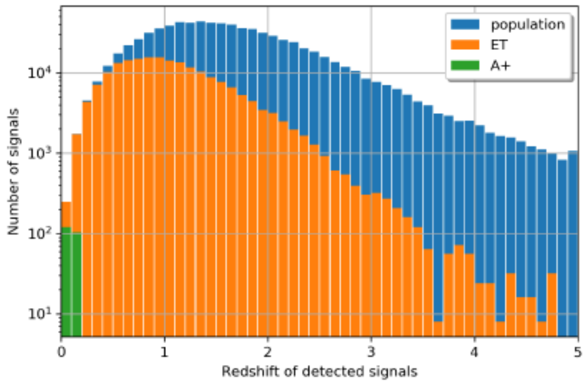
Patrice Verdier – patrice.verdier@in2p3.fr

Lyon – July 4th, 2022

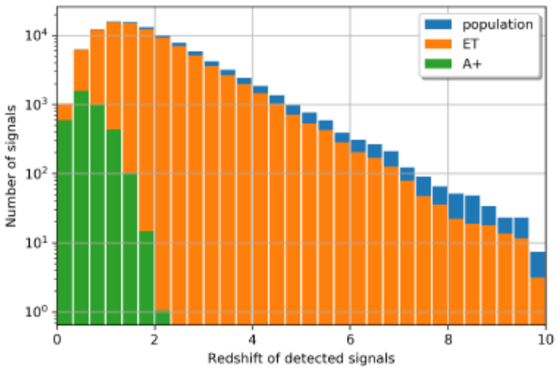
ET sensitivity

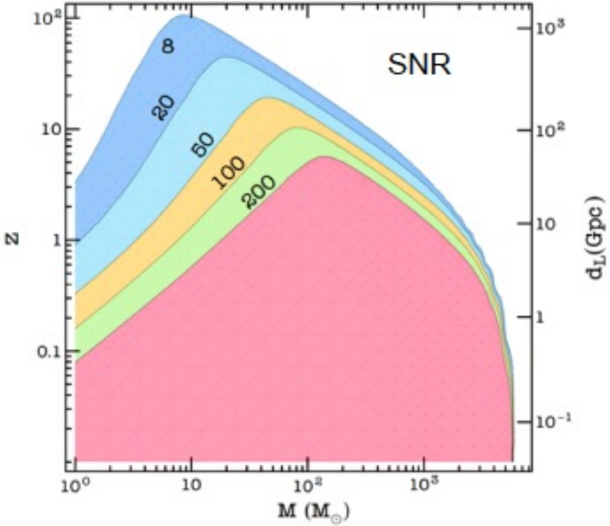
Institut national de physique nucléaire
et de physique des particules

BINARY NEUTRON-STAR MERGERS



BINARY BLACK-HOLE MERGERS





- 10^5-10^6 BBH detections per year
- 10^5 BNS detections per year among which ~100-1000 with EM counterparts
- High SNR events
- Overlapping events

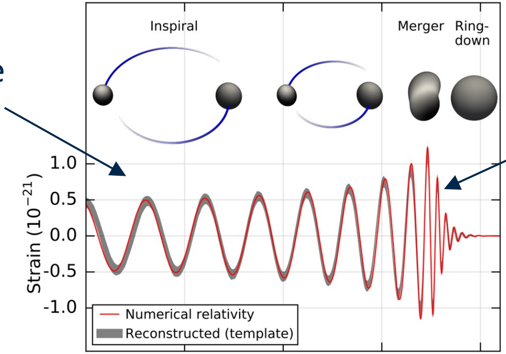
~1 detection every 30s

ET sensitivity

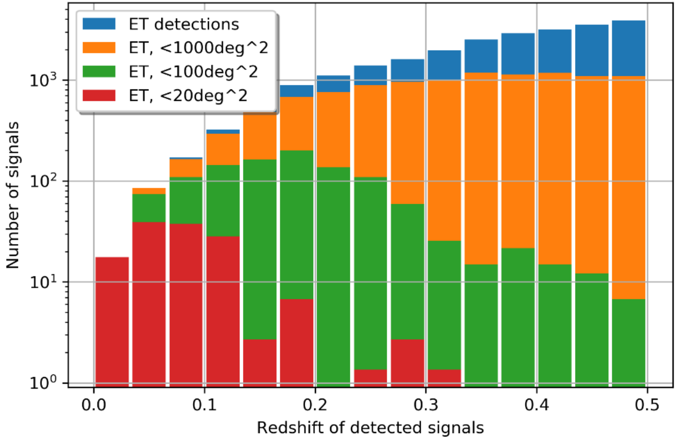
Institut national de physique nucléaire
et de physique des particules

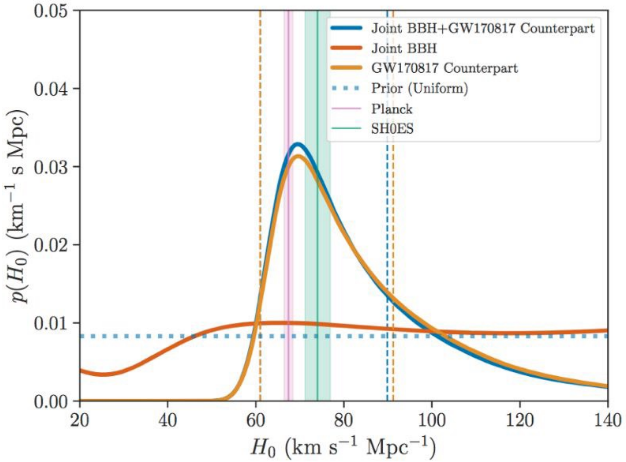
- **BNS detection with EM counterparts and localization precision < 20 deg²** : o(100) per year
- Overlap with many BBH signals
- Potentially, very long signals
- ET will be able to provide alerts few hours before the merger

Identify early the inspiral …



… and provide alert before the merger phase



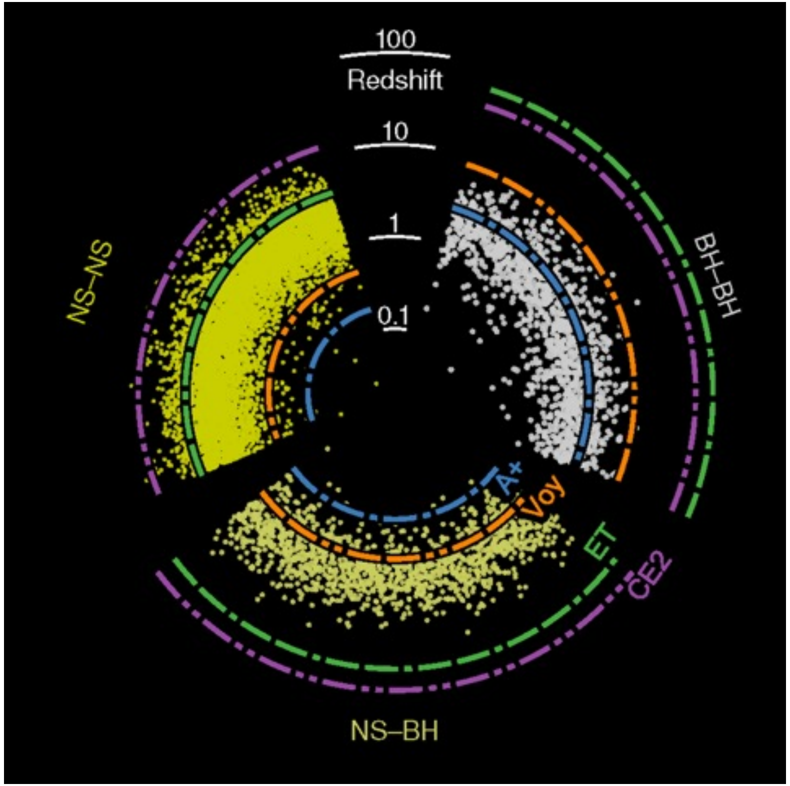


- And with ~600 BNS-EM detection, we can reach Planck resolution on H0 measurement

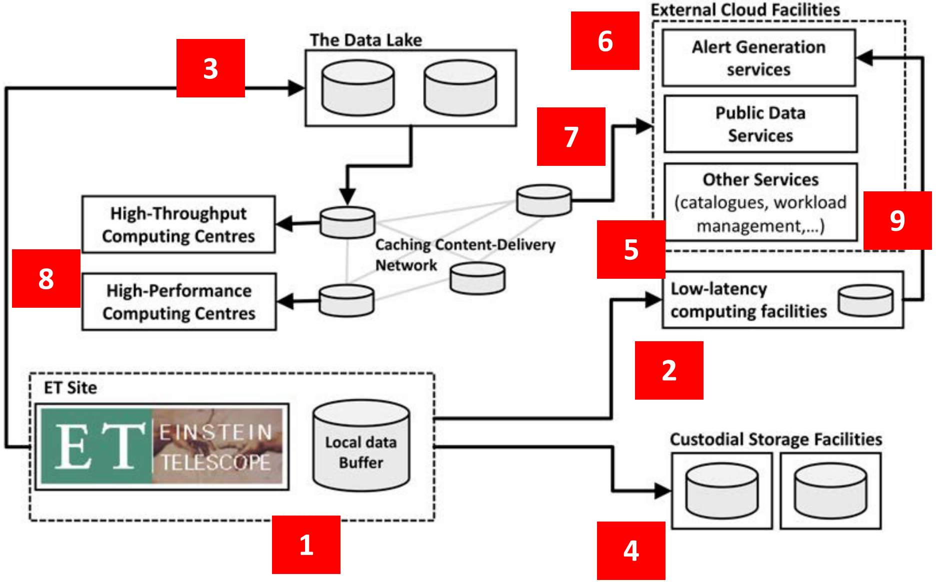
ET data

- Despite the large increase in the expected number of events (10^5-10^6 events/year) compare to 2G interferometers, ET raw data will represent **only few tens of PB per year**. ET data will however need to be distributed to a much larger number of users
- Data processing in ET is the highest challenge, **especially parameter estimation of CBC** (ET CPU will represent ~10% of ATLAS in the HL-LHC era) : plenty of margin for improvements with R&D !

=> ET will use a **distributed computing infrastructure largely based on existing infrastructures**: network and national computing centers, mainly HTC, but also HPC (GPU, FPGA) for numerical relativity simulations & Deep Learning techniques



ET will use a distributed computing infrastructure largely based on existing infrastructures: most of the computing (data processing and services) might run off-site



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
[It may be possible for the low-latency processing sites to exploit the data lake, reducing the complexity of the computing infrastructure]

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 data policy and computing model definition: ET will follow the path created by LIGO and Virgo (c.f. today's talk by Franco Carbognani)

- Open and Public Access to alerts
- Limited period of property data (18 months)
  - Calibration, data quality, internal analysis
- Releasing short chunk of data around published gold events
- Open Access to data implementing the F.A.I.R. indications

And when ET will be taking data:

- EOSC will be operational
- Infrastructures, services and tools for Low Latency Alerts and multi-messenger analysis will be available
- And several experiments will be already using these : CTA, Vera Rubin Observatory, DUNE, Hyperkamiokande, KM3Net, SKA …

➢ **ET is willing to contribute from the beginning**

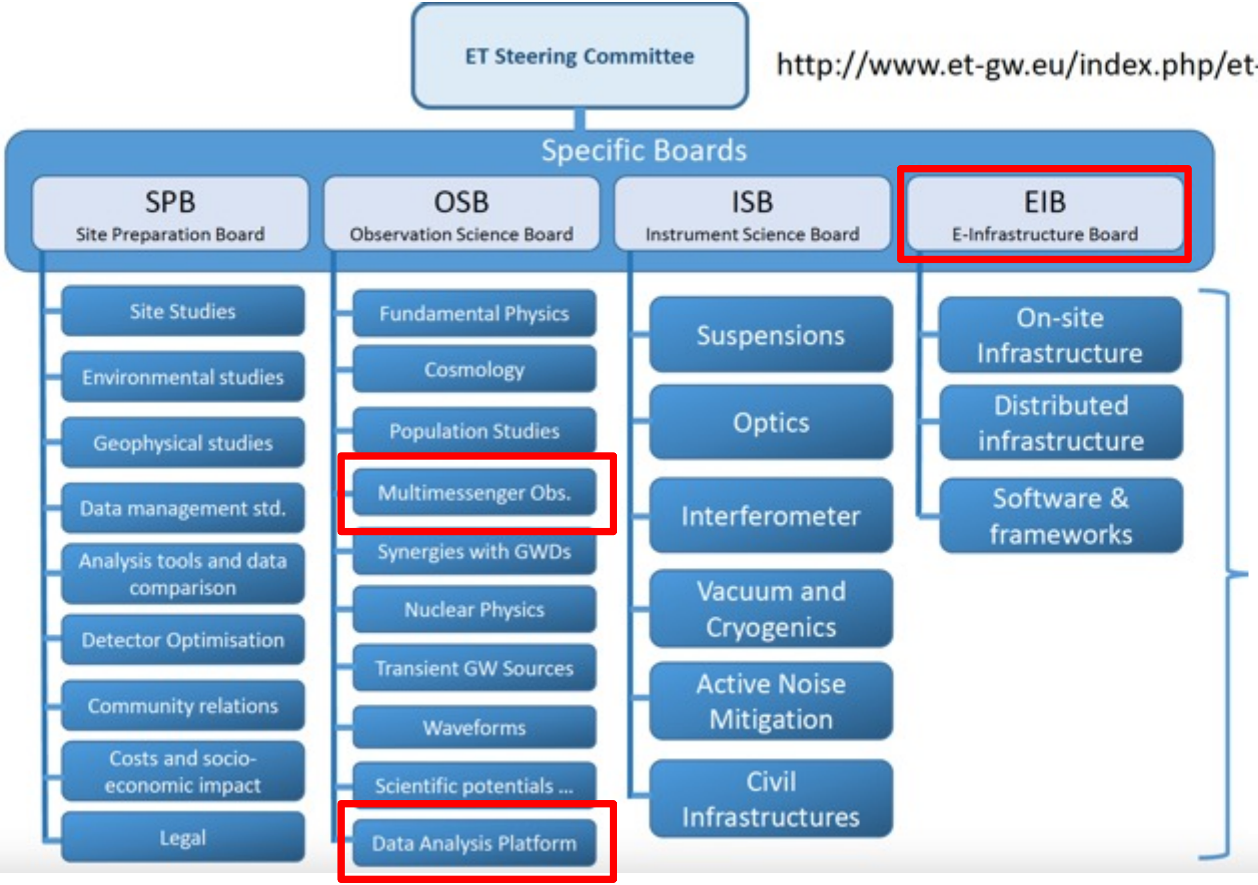
➢ **Possibilities to develop next generation solutions using advanced technologies**

Advanced Computing Technologies

- **Handling ET data volume (~10 PB/year) should be under control, but what will grow compared to 2G GW detectors is**
  - the amount of scientific information encoded in the data
  - the complexity of this information
  - the number of users accessing this data (including for MMA)

- Huge work and improvements are expected from **using ML/AI and new techniques** which will be developed during the next ~10 years
- Most ET needs can be addressed with High-Throughput Computing, however **the role of HPC is expected to increase** (numerical relativity, templates production, parameter estimation) : the **use of GPU and FPGA architectures** will also be considered (for online/offline)
- Most of the computing aspects of ET will have to be organized in order to be able to run on various computing centres and in an automated way
- ET foresees technology tracking of leading-edge computing technologies
- ET computing model will be tested with early data challenges

ET proto-collaboration



http://www.et-gw.eu/index.php/et-steering-committee

ET is now an ESFRI project:

ESFRI ROADMAP 2021

Renewed interests and sharp increase in participation

The ET proto-collaboration is currently in construction

Works towards the definition of the Infrastructure and Project governance is starting

EIB organization

Division 1: Software, frameworks, and data challenge support
Division 2: Services and Collaboration Support
Division 3: Computing and data model, Resource Estimation
Division 4: Multimessenger alerts infrastructure

TTG: Technology Tracking working Group

	T8.1 T0 Data center	T8.2 Computing and Data Model	T8.3 Resources	T8.4 Data Access Implementation
D1 Software, frameworks, data challenges		Computing frameworks computing domains data formats	Resources for frameworks execution and data storage availability	Data availability Data releases format
D2 Services and Collaboration Support			Resources and infrastructure for services and collaborative tools	Tools for monitoring, AAI (IAM), data access
D3 Computing and data model, Resource Estimation	T0 storage and computing resources estimation	Computing model data model	Resources estimation	
D4 Multimessenger alerts and low-latency infrastructure	ET site infrastructure for alerts and low-latency	Low latency computing model	Resources for low-latency infrastructure	Tools for low latency results sharing
TTC technology tracking committee				

nadia.tonello@bsc.es

ET-PP WP6:

- Task 8.1: “T0 data center” Conceptual design of the center in close collaboration with the instrument science board. Definition of the services provided by the center, delimitation against services realized with distributed computing.
- Task 8.2: “Computing and Data Model” Development of the computing and data model in close cooperation with the instrument science board and observational science board of ET. Definition of the workflow from the instrument to the publication.
- Task 8.3: “Resources” Estimate of the computing resources (computing power and data storage), the personnel, and the operational cost required for all aspects of ET computing. The potential for mitigation must be addressed.
- Task 8.4: “Data Access Implementation” Guidelines for the data policy compliance, relevant to the data storage, access, process and distribution, on all relevant time scales, respecting the EU policies on open data.

Infratech: M2TECH

M2Tech (INFRA-Tech) WP6

- M2Tech proposal submitted together with CTA, KM3NeT, Virgo, focused on multi-messenger astronomy
- WP6 “Efficient computing and algorithms”

Task 6.1 – Sustainable fast data reduction and processing

AI/ML-based processing, GPU and hardware accelerators,…

Task 6.2 – Operational Intelligence for instrument and data flows management

- ML-based tools and methods

Task 6.3 – Efficient tools for multimessenger alerts

- Common tools, standards, formats for alert management. Dynamic alerts database.

- Also: “Digital twin” concept in WP4
- See Edwige’s talk tomorrow

Conclusion

- **Einstein Telescope** will provide enhanced sensitivities to GW detection compared to 2G experiments, 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 the needs for analyses, in particular multi-messengers

- ET is willing to contribute to current developments of tools and services for Low Latency Alerts & Multi-Messenger Analysis

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