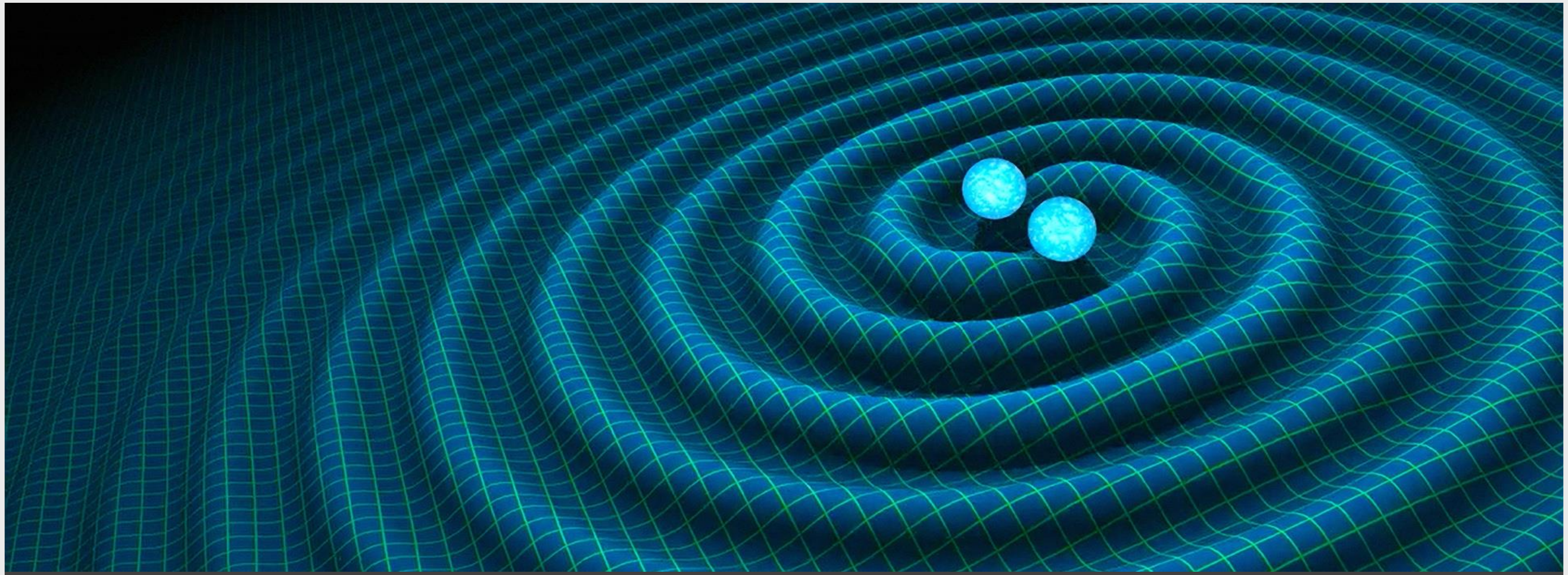


Towards the einstein telescope computing model

Stefano Bagnasco, INFN Torino

CHEP2024

Kraków
Jun 19, 2024

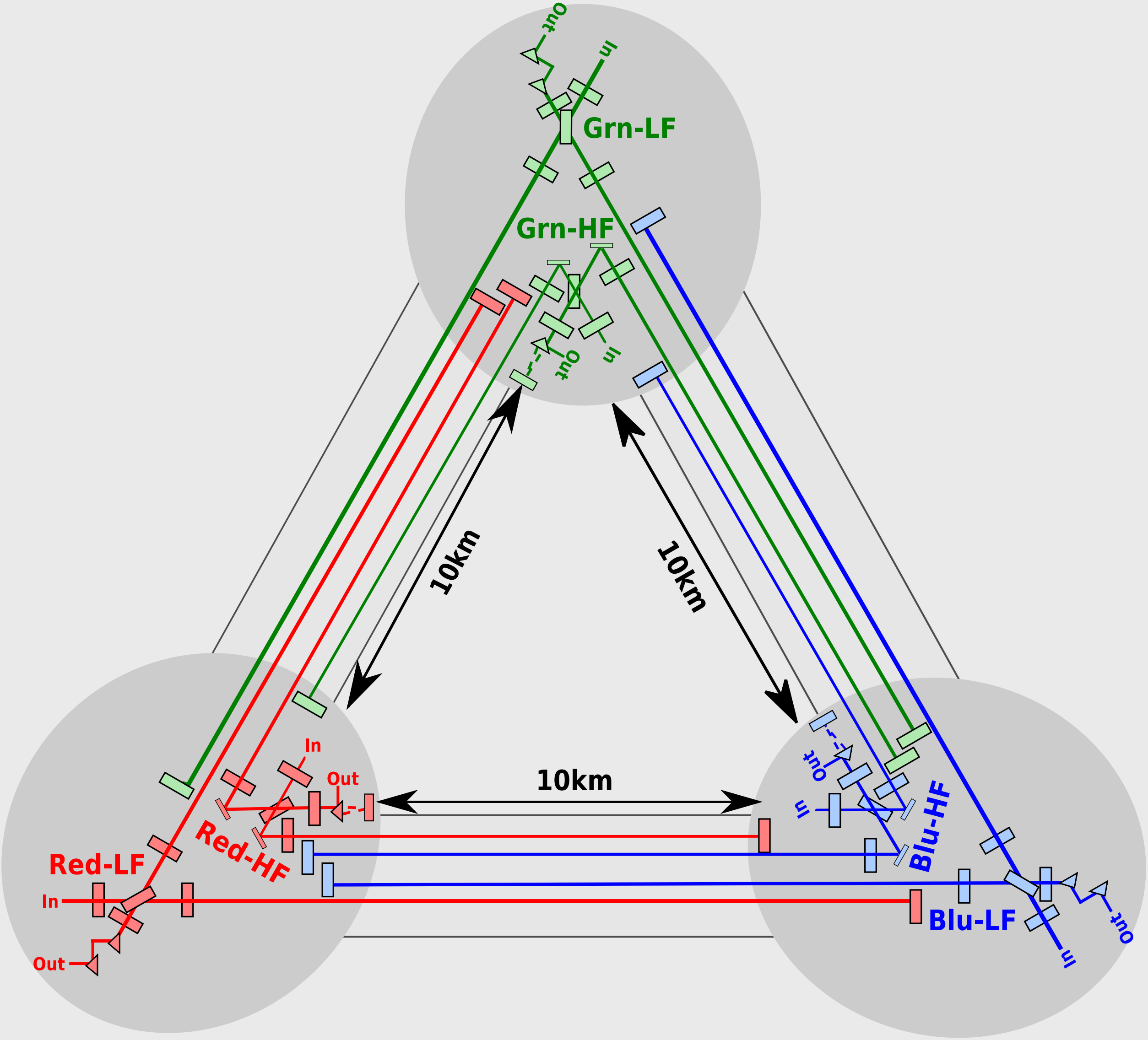
The Einstein Telescope

The **European 3rd Generation Gravitational Wave observatory**

- ET has been a pioneer idea that defined the concept of 3rd generation GW observatory:
    - A sensitivity at least 10 times better than the (nominal) advanced detectors on a large fraction of the detection frequency band
    - Wideband (possibly wider than the current detectors) accessing the low frequency band below 10Hz
    - High reliability and improved independent observation capability

ET: how

- Three detectors in a **triangular** structure (baseline design)
  - Closed geometry allows the use of the null data stream
  - Extra detector adds redundancy and makes up for the non-right 60° angle
- Each detector (“red”, “green” and “blue”) consists of **two** Michelson interferometers
  - High-frequency and Low-frequency
  - Underground, cryogenic
- Alternative 2L (15km, misaligned, non-colocated)



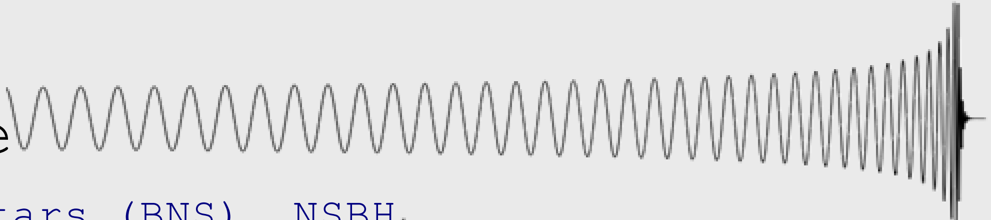
ET: where

- Currently there are two candidate sites being characterized to host ET:
    - The Sardinia site, close to the Sos Enattos mine
    - The Euregio Meuse-Rhine site, close to the NL-B-D border
    - A third option in Saxony (Germany) was also proposed and is under discussion

Looking for...

Transient sources:

- CBC: Compact Binary Coalescence



  - Binary Black Holes (BBH), Binary Neutron Stars (BNS), NSBH
- Burst: Unmodeled transient bursts

  - E.g., Core Collapse Supernovae (CCS, and anything else)

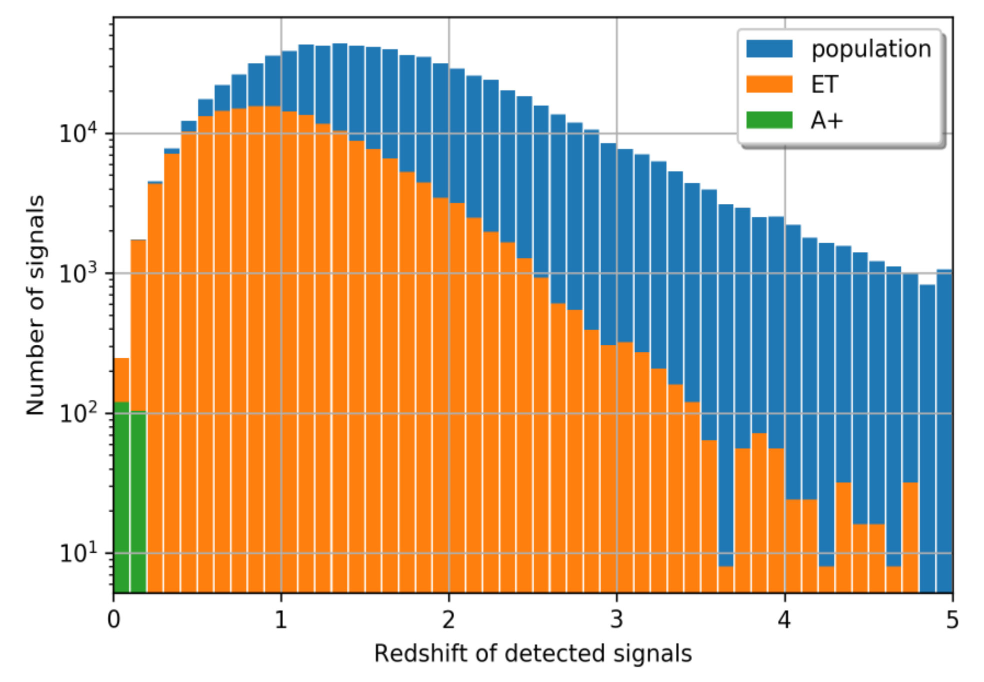
Continuous sources:

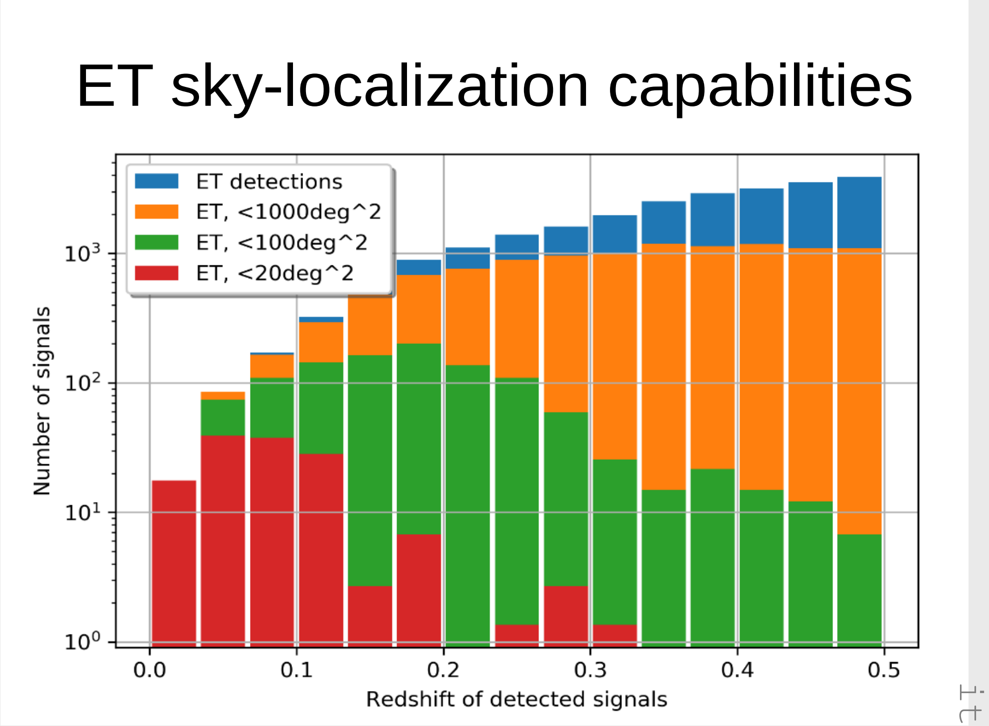
- CW: Continuous waves

  - E.g., Asymmetric spinning neutron stars
- SGWB: Continuous stochastic backgro

  - Cosmological & astrophysical

Enhanced sensitivity



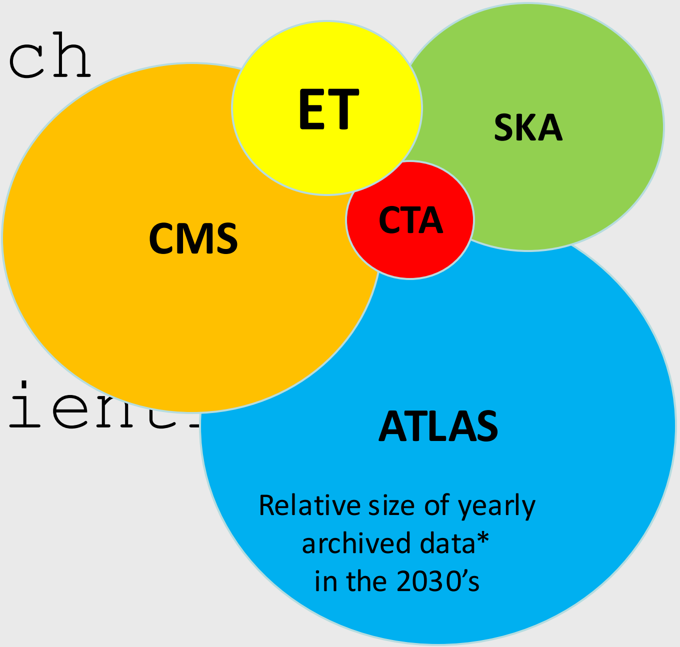


- 10^5 BBH detections per year
- 10^5 BNS detections per year

- o(100) detections per year with <20 deg^2
- Early warning by minutes (hours)

data and event rate

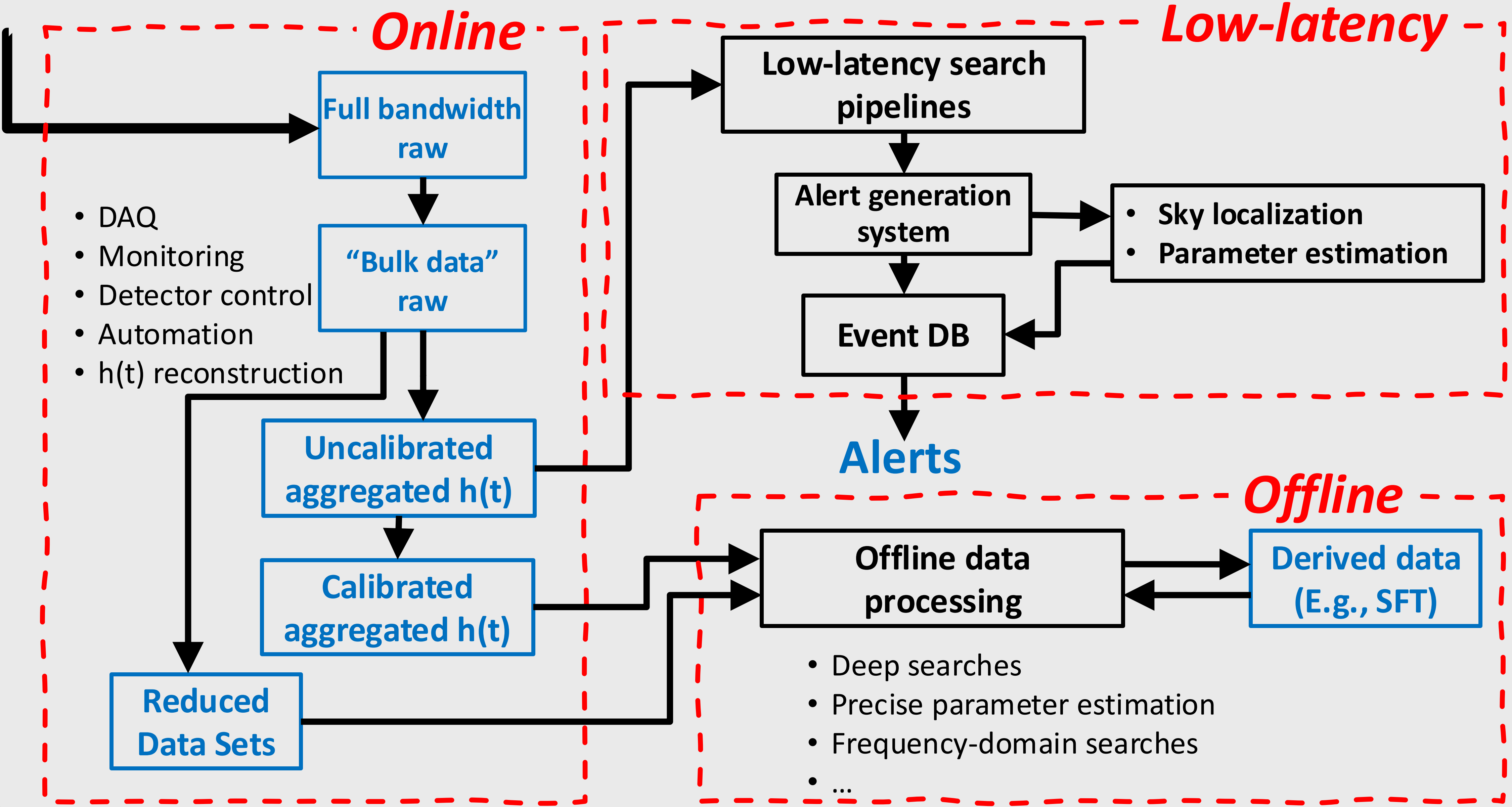
- Raw interferometer data don’t grow much with increasing sensitivity, but…
  - 6 × (and larger) interferometers
  - Cryogenics, huge vacuum volume,…
- What grows is the amount of useful scient information embedded in data
  - And the computing power needed to wring it out
- The event rate scales with the third power of the range
  - Today: 1 every couple of days during O4
  - ET: one every few minutes, and staying much longer in the sensitivity band
- Standard techniques and strategies will not scale
  - In particular, CBC Parameter Estimation
  - Estimating the needs is a task in itself



Relative size of yearly archived data*
in the 2030’s

*to say nothing of weather forecast, genomics, Earth observation, oil industry, GAFAM and everybody else

Three computing domains



- DAQ
- Monitoring
- Detector control
- Automation
- h(t) reconstruction

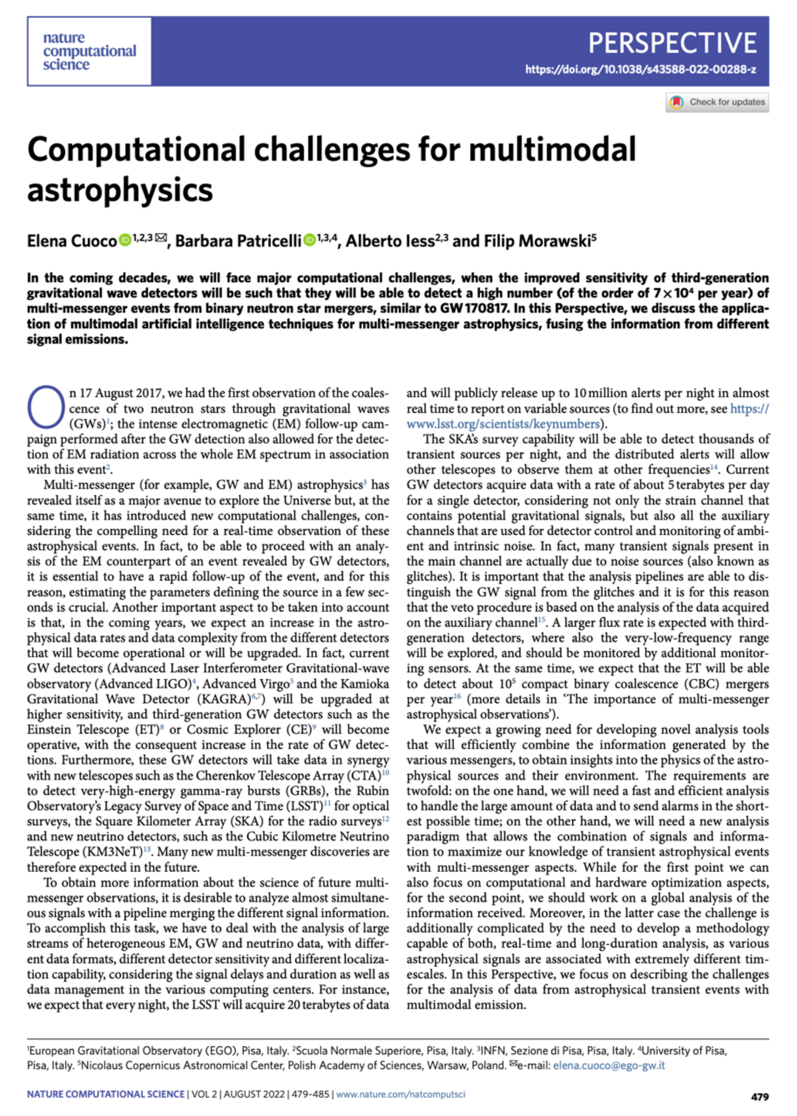
- Sky localization
- Parameter estimation

- Deep searches
- Precise parameter estimation
- Frequency-domain searches
- …

Boundary conditions

- We need a (reasonably) functional infrastructure from the beginning
  - To run Mock Data Challenges and develop algorithms, more about this shortly
- ET will not work in a vacuum
  - Cosmic Explorer project in the US: there will be a 3G network
  - Other “multimessenger” facilities (Vera Rubin, LISA, SKA, CTA, ELT,…)
- Distributed computing will be dominant
  - For low-latency as well (offline/LL infrastructures boundary is becoming blurry)
  - Maybe on a common infrastructure with other MM players, see next slide
- Alert generation, management and distribution will be a challenge
  - Much work already ongoing by Vera Rubin (brokers,…)

An example of an extra challenge



nature computational science
PERSPECTIVE
https://doi.org/10.1038/s43588-022-00288-z

Check for updates

Computational challenges for multimodal astrophysics

Elena Cuoco[1,2,3] ✉, Barbara Patricelli[1,3,4], Alberto Iess[2,3] and Filip Morawski[5]

In the coming decades, we will face major computational challenges, when the improved sensitivity of third-generation gravitational wave detectors will be such that they will be able to detect a high number (of the order of 7×10^4 per year) of multi-messenger events from binary neutron star mergers, similar to GW170817. In this Perspective, we discuss the application of multimodal artificial intelligence techniques for multi-messenger astrophysics, fusing the information from different signal emissions.

On 17 August 2017, we had the first observation of the coalescence of two neutron stars through gravitational waves (GWs)[1]; the intense electromagnetic (EM) follow-up campaign performed after the GW detection also allowed for the detection of EM radiation across the whole EM spectrum in association with this event[2].

Multi-messenger (for example, GW and EM) astrophysics[3] has revealed itself as a major avenue to explore the Universe but, at the same time, it has introduced new computational challenges, considering the compelling need for a real-time observation of these astrophysical events. In fact, to be able to proceed with an analysis of the EM counterpart of an event revealed by GW detectors, it is essential to have a rapid follow-up of the event, and for this reason, estimating the parameters defining the source in a few seconds is crucial. Another important aspect to be taken into account is that, in the coming years, we expect an increase in the astrophysical data rates and data complexity from the different detectors that will become operational or will be upgraded. In fact, current GW detectors (Advanced Laser Interferometer Gravitational-wave observatory (Advanced LIGO)[4], Advanced Virgo[5] and the Kamioka Gravitational Wave Detector (KAGRA)[6,7]) will be upgraded at higher sensitivity, and third-generation GW detectors such as the Einstein Telescope (ET)[8] or Cosmic Explorer (CE)[9] will become operative, with the consequent increase in the rate of GW detections. Furthermore, these GW detectors will take data in synergy with new telescopes such as the Cherenkov Telescope Array (CTA)[10] to detect very-high-energy gamma-ray bursts (GRBs), the Rubin Observatory's Legacy Survey of Space and Time (LSST)[11] for optical surveys, the Square Kilometer Array (SKA) for the radio surveys[12] and new neutrino detectors, such as the Cubic Kilometre Neutrino Telescope (KM3NeT)[13]. Many new multi-messenger discoveries are therefore expected in the future.

To obtain more information about the science of future multi-messenger observations, it is desirable to analyze almost simultaneous signals with a pipeline merging the different signal information. To accomplish this task, we have to deal with the analysis of large streams of heterogeneous EM, GW and neutrino data, with different data formats, different detector sensitivity and different localization capability, considering the signal delays and duration as well as data management in the various computing centers. For instance, we expect that every night, the LSST will acquire 20 terabytes of data and will publicly release up to 10 million alerts per night in almost real time to report on variable sources (to find out more, see https://www.lsst.org/scientists/keynumbers).

The SKA's survey capability will be able to detect thousands of transient sources per night, and the distributed alerts will allow other telescopes to observe them at other frequencies[14]. Current GW detectors acquire data with a rate of about 5 terabytes per day for a single detector, considering not only the strain channel that contains potential gravitational signals, but also all the auxiliary channels that are used for detector control and monitoring of ambient and intrinsic noise. In fact, many transient signals present in the main channel are actually due to noise sources (also known as glitches). It is important that the analysis pipelines are able to distinguish the GW signal from the glitches and it is for this reason that the veto procedure is based on the analysis of the data acquired on the auxiliary channel[15]. A larger flux rate is expected with third-generation detectors, where also the very-low-frequency range will be explored, and should be monitored by additional monitoring sensors. At the same time, we expect that the ET will be able to detect about 10^5 compact binary coalescence (CBC) mergers per year[16] (more details in 'The importance of multi-messenger astrophysical observations').

We expect a growing need for developing novel analysis tools that will efficiently combine the information generated by the various messengers, to obtain insights into the physics of the astrophysical sources and their environment. The requirements are twofold: on the one hand, we will need a fast and efficient analysis to handle the large amount of data and to send alarms in the shortest possible time; on the other hand, we will need a new analysis paradigm that allows the combination of signals and information to maximize our knowledge of transient astrophysical events with multi-messenger aspects. While for the first point we can also focus on computational and hardware optimization aspects, for the second point, we should work on a global analysis of the information received. Moreover, in the latter case the challenge is additionally complicated by the need to develop a methodology capable of both, real-time and long-duration analysis, as various astrophysical signals are associated with extremely different timescales. In this Perspective, we focus on describing the challenges for the analysis of data from astrophysical transient events with multimodal emission.

---

[1]European Gravitational Observatory (EGO), Pisa, Italy. [2]Scuola Normale Superiore, Pisa, Italy. [3]INFN, Sezione di Pisa, Pisa, Italy. [4]University of Pisa, Pisa, Italy. [5]Nicolaus Copernicus Astronomical Center, Polish Academy of Sciences, Warsaw, Poland. ✉e-mail: elena.cuoco@ego-gw.it

NATURE COMPUTATIONAL SCIENCE | VOL 2 | AUGUST 2022 | 479–485 | www.nature.com/natcomputsci

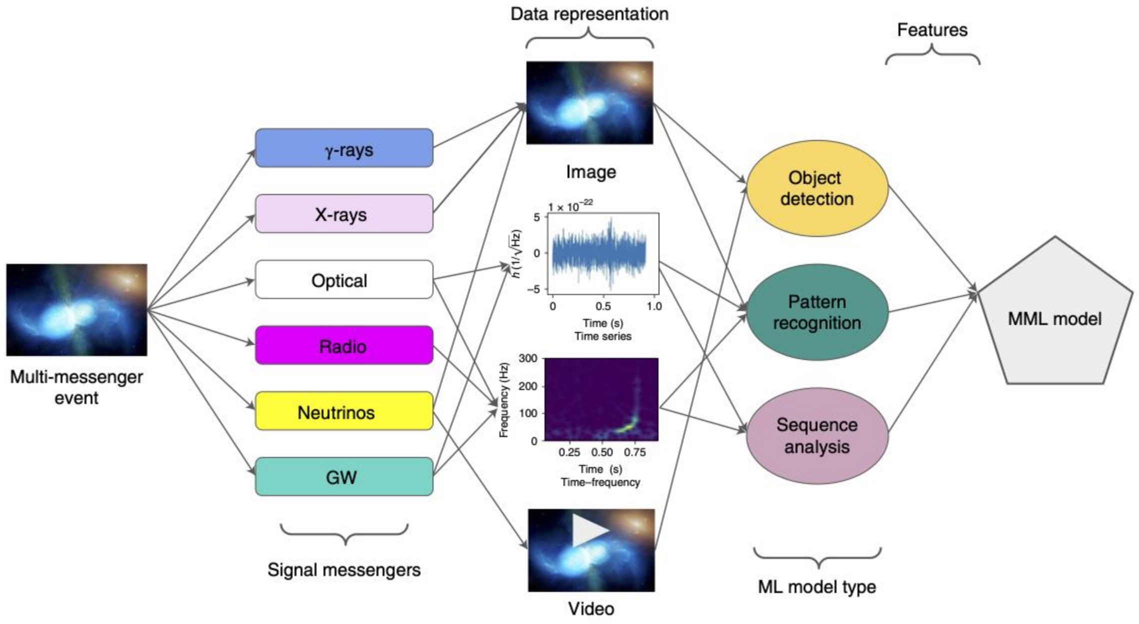
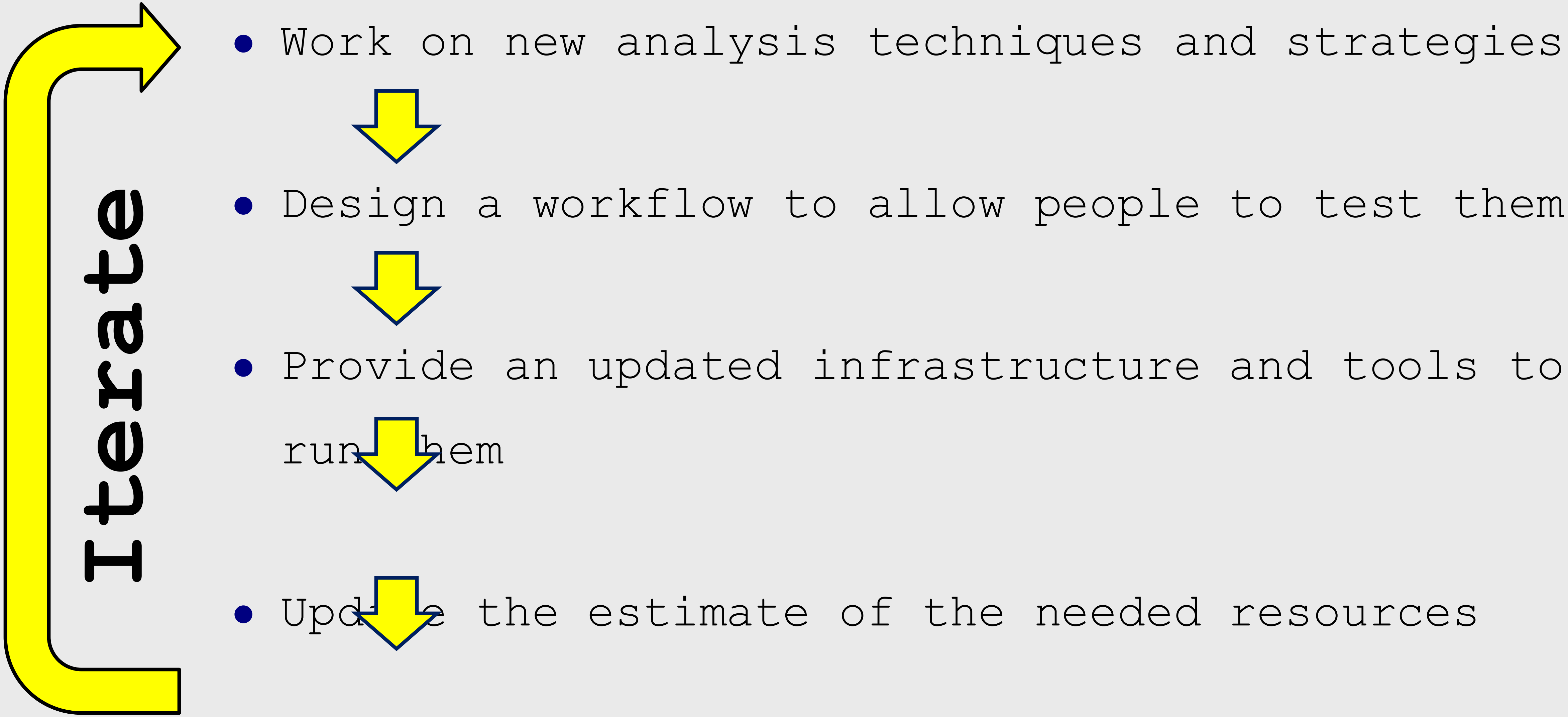


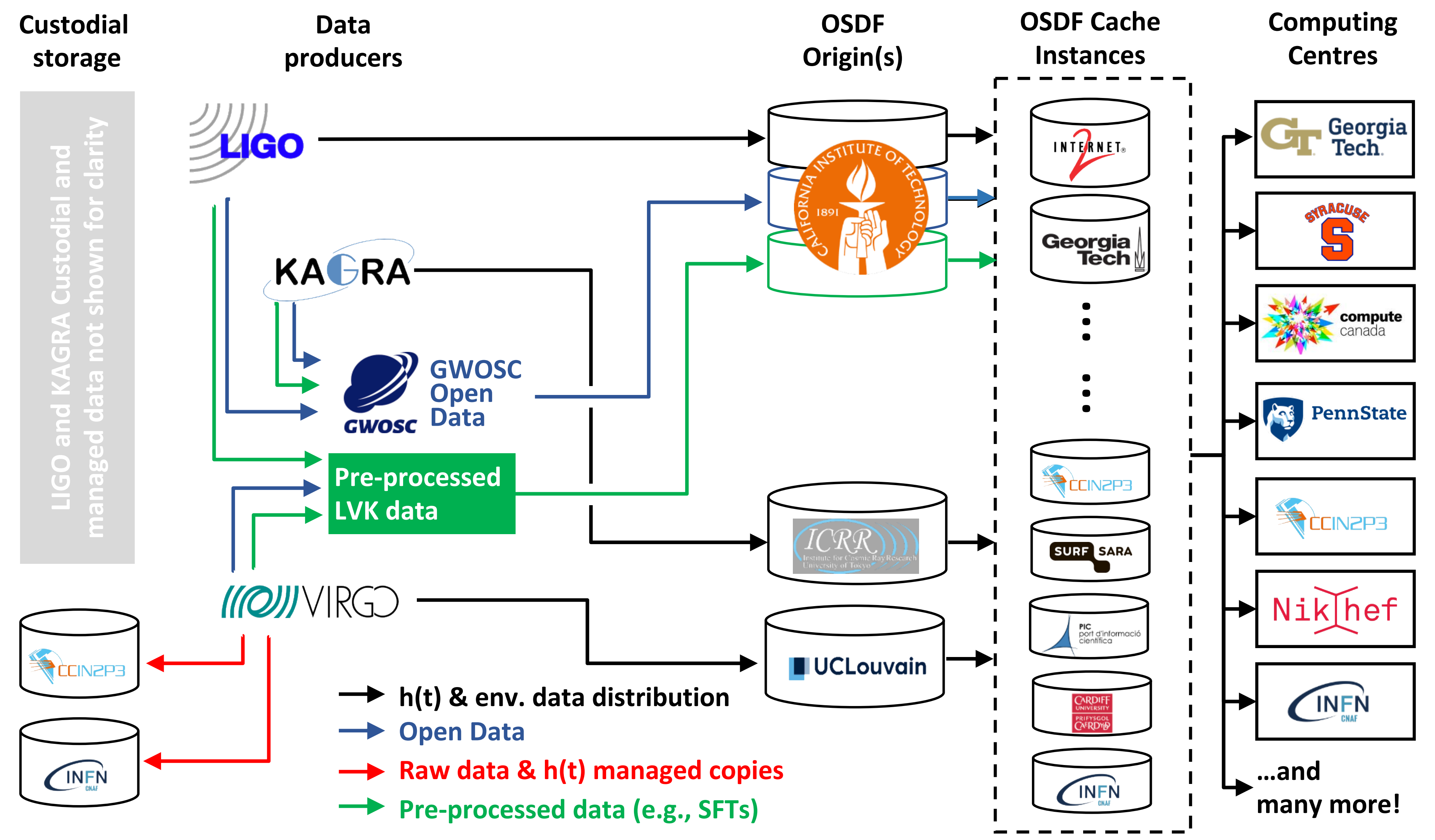
Fig. 3 | The astrophysical event's conceptual MML workflow. A multi-messenger astrophysical transient event can manifest itself via various signal types, including GW, γ-rays, X-rays, optical and radio emission, and neutrinos. Different modalities have their own representation in various domains. We can use the extracted features to perform model prediction in the first stage by using deep learning and ML models. Furthermore, we can analyze the predictions later by combining them in the global MML model.

What we need to do



- Work on new analysis techniques and strategies
- Design a workflow to allow people to test them
- Provide an updated infrastructure and tools to run them
- Update the estimate of the needed resources
- Evaluate the performance of the infrastructure

The LVK "IGWN" infrastructure



mock data challenges

- MDC as multipurpose tools
  - Develop and exercise analysis code and strategies with increasingly complex datasets
  - Build the data analysis community and bootstrap new groups
  - Educate the community in the use of common distributed computing tools and best practices
  - Iteratively test the distributed computing infrastructure

- Mock Data Challenge
  - **MDC1:** provide data distribution layer, adopt "all-comers, bring-your-own-workflow" model and survey the activities
  - **MDC2:** provide (a set of) prototype tools for workload management etc. and impose some (more) structure
  - **MDC3…n:** iterate

Workflow evaluation kits

- Independent packaged parts of the final architecture
  - Providing limited functionalities, possibly some as mere demonstrators
  - But actually to be released to users (i.e., they MUST be functional)
  - Different implementations may exist, with different tools/technologies used to provide same functionality
  - Integration of existing tools, with little bespoke developments, to map "kits" onto small(ish) projects
- Examples:
  - ESCAPE Datalake + RucioFS for data distribution
  - SnakeMake evaluation
  - ESCAPE Datalake + VRE interactive data analysis environment
  - OSDF + INFNCloud interactive data analysis environment
  - Multi-architecture evaluation platform @ETIC
  - HSE rich metadata tool integrated with data distribution tool

Oscars projects

OSCARS

Open Science Clusters' Action for Research & Society

- ETAP (Université de Genève)
  - Access to multiple ESCAPE Data Lakes (managed by MADDEN).
  - Rich metadata service integration
  - Access to multiple rich metadata instances
  - A lightweight CRM service monitoring the VRE

- MADDEN (INFN-Torino & Université Catholique de Louvain)
  - Multi-RI Data Lake managed with Rucio.
  - Development and test of RucioFS (a POSIX view of the Rucio catalog)
  - Extend RucioFS to support advanced metadata (provided by ETAP)

See Lia's talk on Wednesday: https://indico.cern.ch/event/1338689/contributions/6010882/

conclusions

- The path to a working Einstein Telescope detector is still very long
  - But the multimessenger community is already moving
- Many parameters of the computing model are still unknown
- We, however, need an infrastructure right away
  - To support Mock Data Challenges and develop analysis tools and strategies
  - To evaluate the missing parameters
  - To gradually produce the final infrastructure both for offlina and low-latency activities
- We have a strategy based on starting from what's used now by the 2G GW community
  - And evolve it starting from other existing tools and ESCAPE

Next: Spare slides

Thanks!