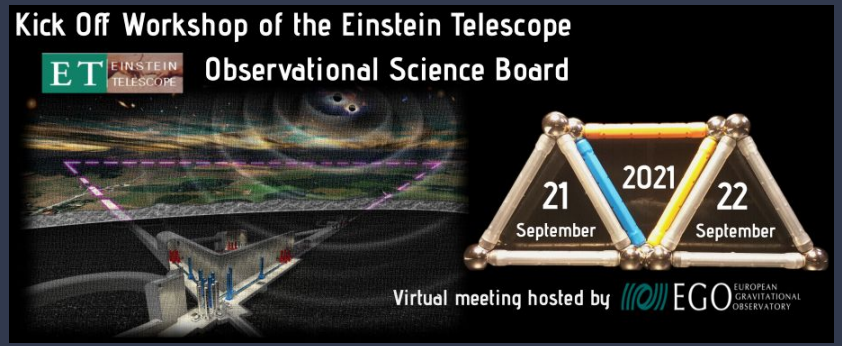
OSB Division 10
Data Analysis Platform

Chris Van Den Broeck, **Elena Cuoco**, Tania Regimbau, John Veitch



September 22th 2021, ET kick-off

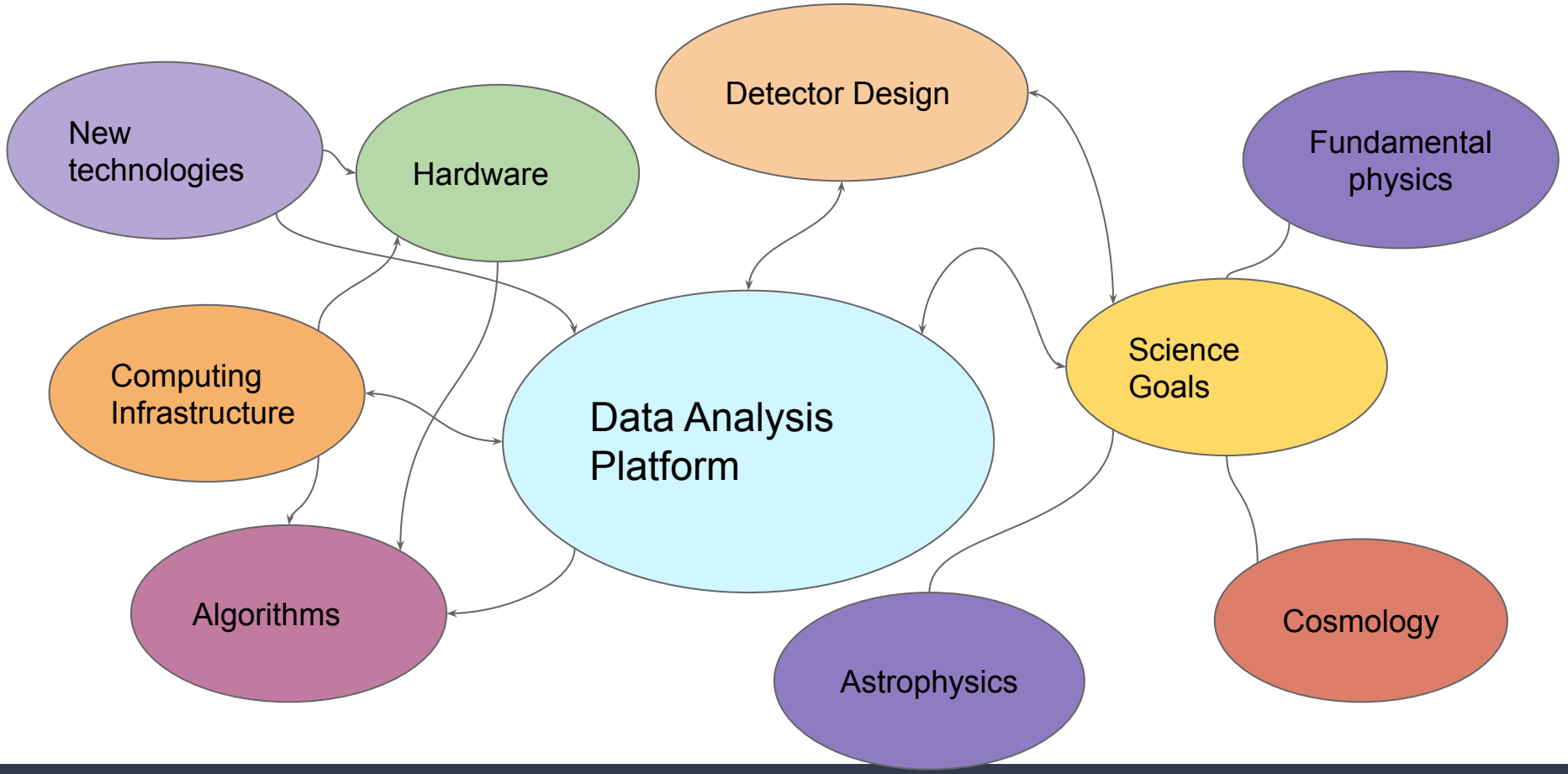
Outline

- Goals
- Methodologies
- Simulations and MDCs
- Work Package structure
- Interaction with other divisions
- DIV10 organization

Goals

Division 10 have identified following topics that cut across many science areas

01	What are the unique advantages of a null stream?	How will real instrumental noise impact the ideal case?
02	How can we address challenges of ET signals:	• Low frequency / Long duration CBCs • Very high SNR • Huge number of detections
03	What are the implications of overlapping signals?	• What impact will they have on science targets? • How to analyse overlapping signals?
04	What are the computational challenges?	• What new methods can be used? • Real time analysis?
05	What performance metrics can be used to guide instrument design?	• Do we have standard evaluation metric for different searches?



Data Analysis Landscape

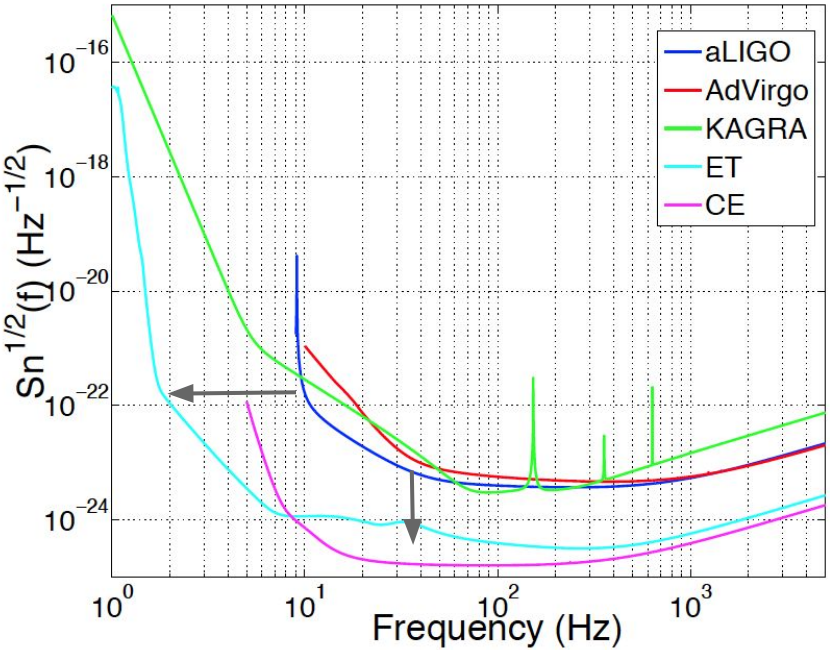
Methodologies and Technologies

- **Lessons from 2G**
  - Best practices and future extensions
- **Machine Learning and Deep Learning**
  - Waveform simulation
  - Signal detection and classification
  - Parameter estimation
- **Overlapping signals**
  - Multi-signal likelihoods + Reduced Order Quadratures (ROQ)s (global fit)
  - Hierarchical analyses (fit and subtract)
- **Computing Infrastructure**
  - European Open Science Cloud
  - Common data analysis platform also with other types of observatories (KM3Net, CTA, …)
  - Quantum computing algorithms in anticipation of hardware

What we have now

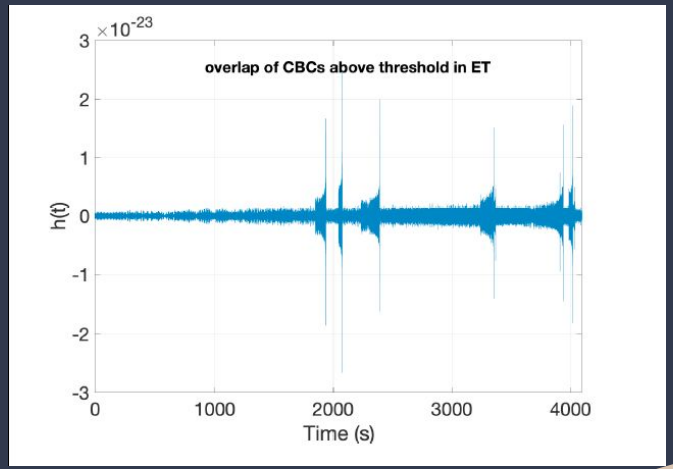
- **Division of searches by source type**
  - CBC, burst, stochastic, CW
- **Algorithms based on matched filtering, template banks, cross-correlation, Bayesian inference**
  - Problem of scalability
  - Problem of parallelizability
- **Computing hardware**
  - CPUs (for the most part)
  - Beowulf clusters that are not globally integrated
- **Data is distributed to compute sites**
  - Causes problems with synchronization
  - Not good for low latency

From 2G sensitivity to 3G



- Lower frequencies (10Hz -> 1-5Hz)
- Sensitivity ~ x10

New problems in 3G



- **Overlapping signals**
  - Want to detect all signals, not just the majority (e.g. for population studies)
  - Parameter estimation problem
  - Impact on testing GR (systematics)
- **Long duration waveform for CBCs**
  - Huge number of templates
  - Time-varying response functions
- **FAR estimate in the presence of a strong foreground**
- **Environmental correlated noise**
  - Can we treat detectors as independent?

New Opportunities

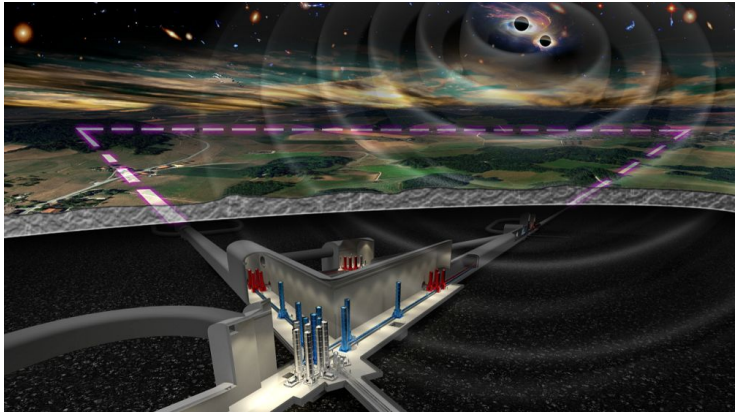
- **Null stream**
  - Polarisation
  - PSD estimation
  - New search methods for unmodelled sources
  - Astrophysical calibration
- **New computing technologies**
  - GPU/TPU/FPGA
  - Cloud computing
- **New algorithms**
  - Machine learning
  - 10+ years classical development
  - Quantum algorithms
- **New synergies**
  - Cosmic Explorer
  - EM observatories
  - Neutrino observatories
  - LISA

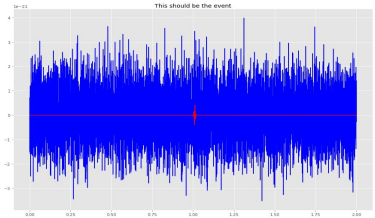
Innovative methodologies: AI techniques

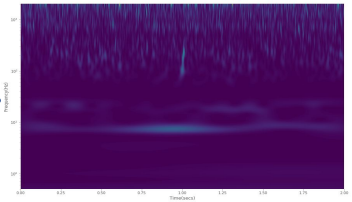
How Machine/Deep learning can help?

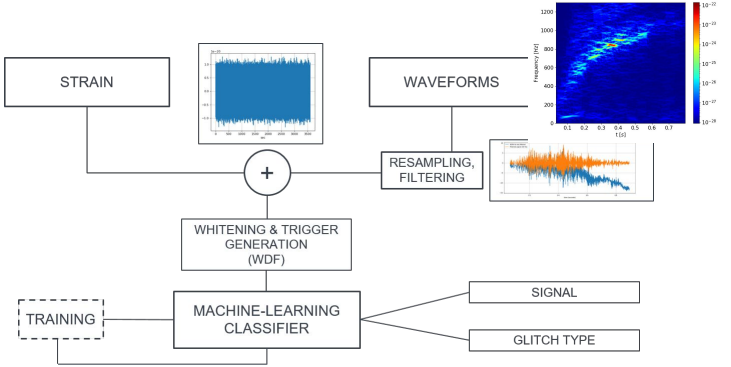
- Efficiency
  - Signal Classification
  - Parameter estimation
  - Waveform simulations
- Speediness
  - Use advanced hardware (GPU, TPU...)
  - FPGAs / custom hardware
- Automatization
  - Automatize standard procedure for Data Quality

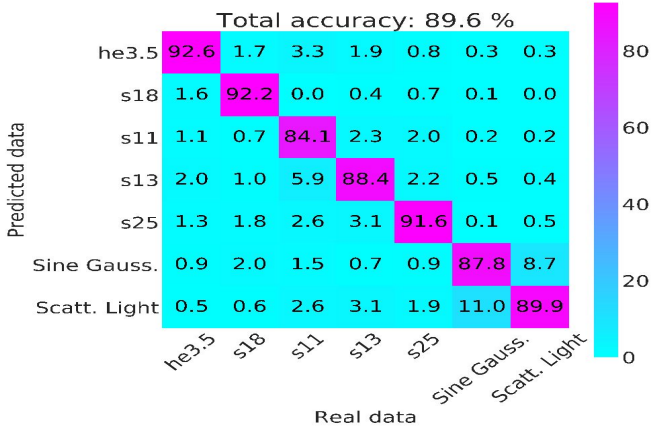
Quantum computing?



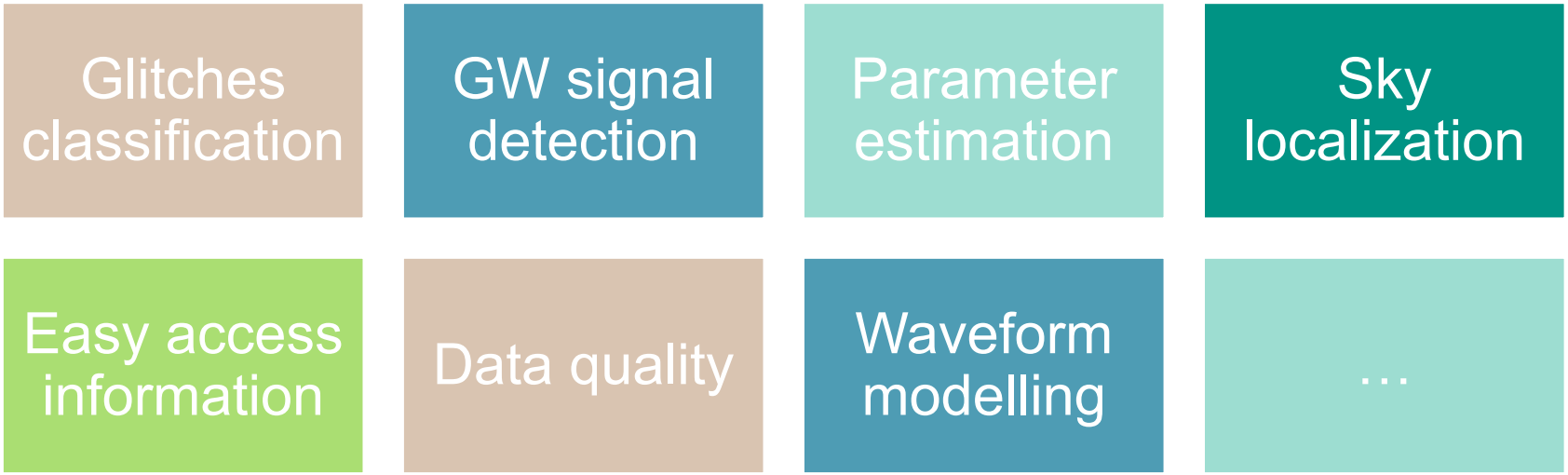






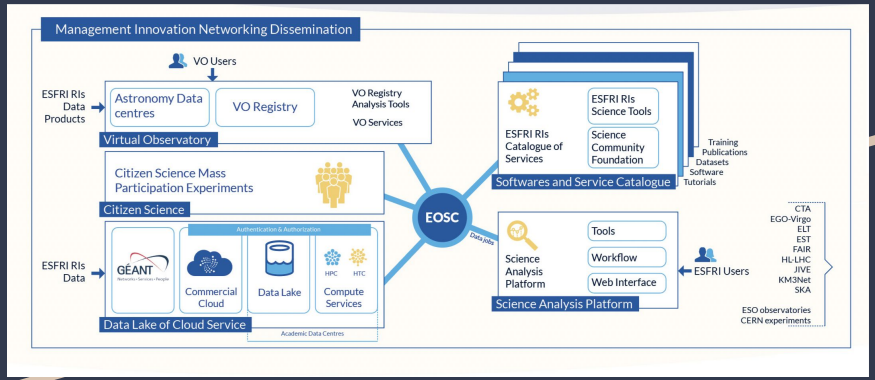


Machine learning for GW search



Review paper: Enhancing gravitational-wave science with machine learning Elena Cuoco *et al* 2021 *Mach. Learn.: Sci. Technol.* 2 011002

European Open Science Cloud (EOSC)



"The ambition of the European Open Science Cloud (EOSC) is to provide European researchers, innovators, companies and citizens with a federated and open multi-disciplinary environment where they can publish, find and re-use data, tools and services for research, innovation and educational purposes.

This environment will operate under well-defined conditions to ensure trust and safeguard the public interest." Citing [EOSC europa.eu]

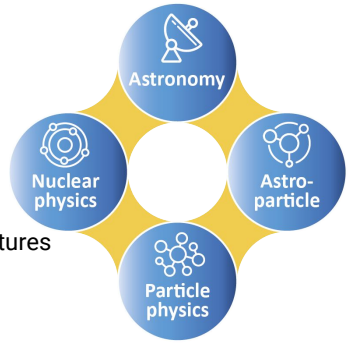
- **ESCAPE EOSC cell**
  - Data lake
  - Data analysis platform
  - Software catalogue
- Collaborate with other ESFRI and Synergy division
- Standardize best practices, software repo and deployment
- User friendly platform

Interaction with EIB E-Infrastructure Board

ESCAPE Science Projects – background

- **Two large Science Projects** will be deployed with a number of high-level objectives (Dark matter and Extreme Universe)
  - To demonstrate new cutting-edge science capabilities, in particular those involving inter-RI collaboration and science outcomes;
  - To validate, that the software, tools, services, and infrastructure developed within ESCAPE are what is required by the science use cases;
- The SP objectives are supported by the thematic consortia (of the national funding agencies):
  - ECFA, APPEC, ASTRONET, NuPECC, and the collaboration of those bodies within JENAA.
- The European Strategy for Particle Physics update in 2020 encouraged synergies between these research infrastructures, via ESCAPE.

European
Science
Cluster of
Astronomy &
Particle physics
ESFRI research infrastructures

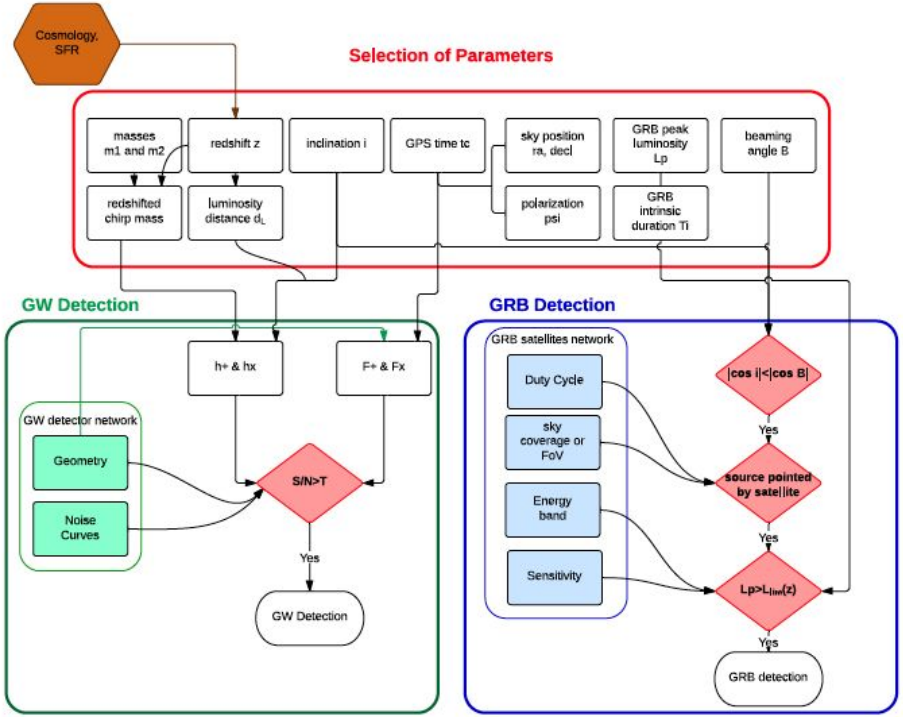


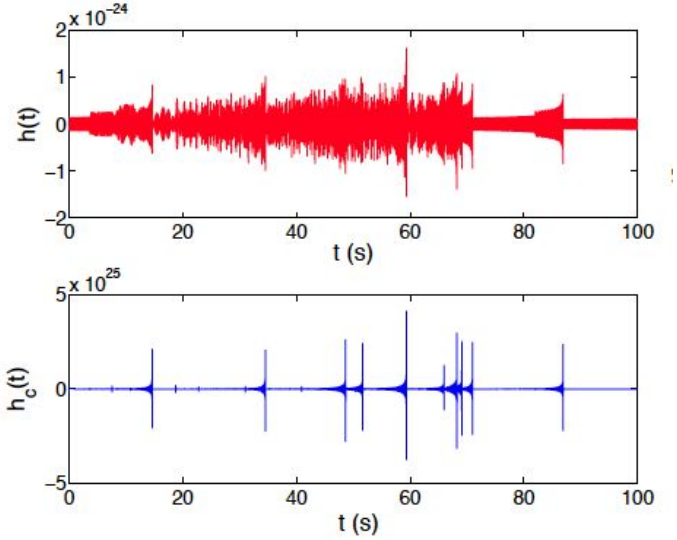
Simulations and Mock Data/Science Challenges

Exploring challenges and opportunities

- **Goal: develop/test/optimize data analysis methods and parameter estimation**
- **How can we create realistic 3G noise**?
  - Different design sensitivity curves
  - Correlations between IFO channels
  - New families of glitches?
  - Calibration uncertainties
- **Signals**
  - CBCs / CWs / SNae / Stochastic
  - Exotica
  - Need input from other divisions for specifications
- **Evaluation**
  - Evaluation criteria driven by science goals
  - Feedback to instrument science groups

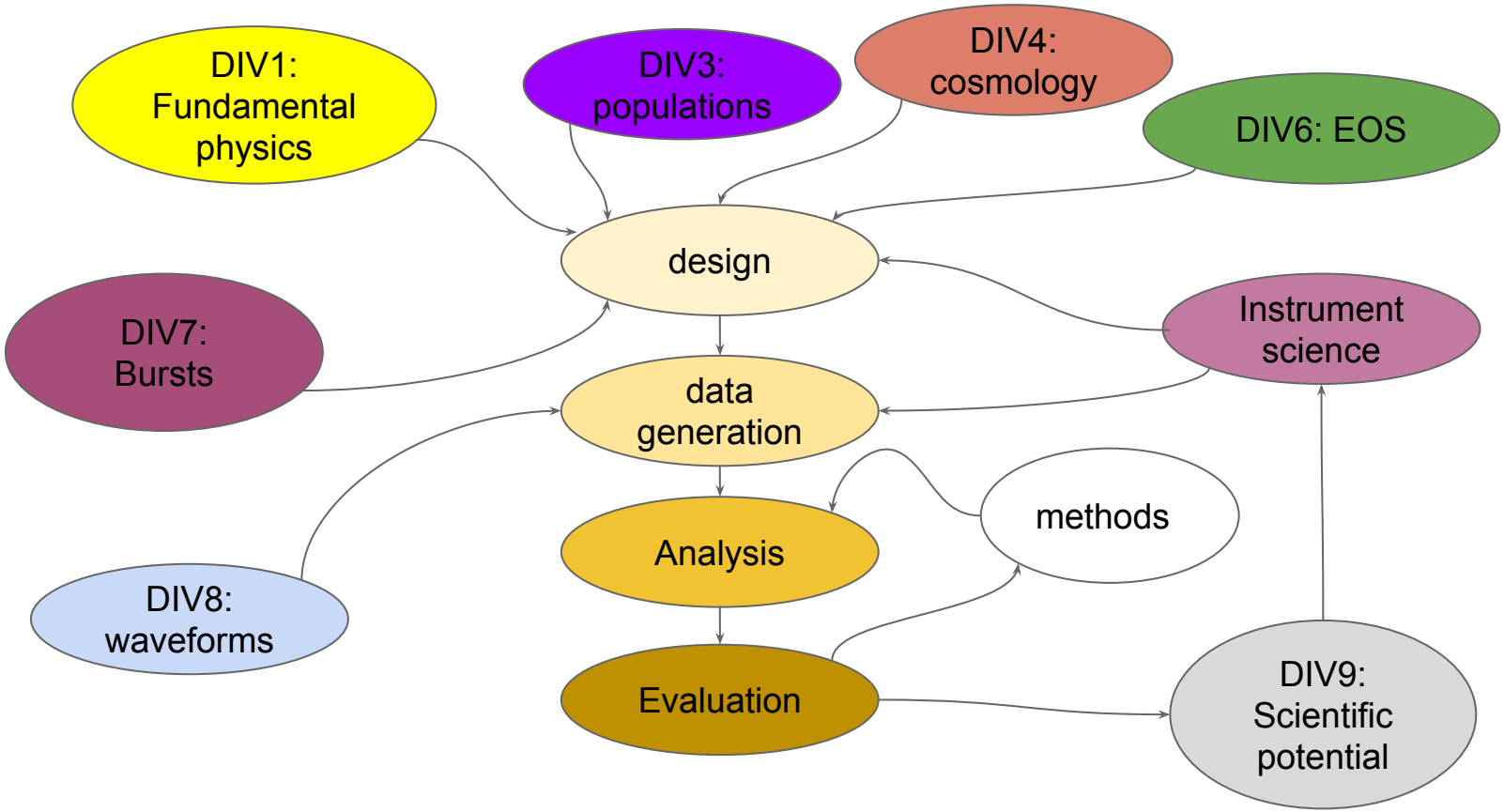
ET MDC&S_Generation package: using Monte Carlo techniques, generates time series on E1, E2, E3 and the null stream containing colored Gaussian noise and the GW waveforms from all the CBCs in the Universe, up to z~10.





"Mock Data Challenge for the Einstein GW Telescope", 2012PhRvD..86l2001R

Previous ET MDCs



MDCs: Interaction with other divisions

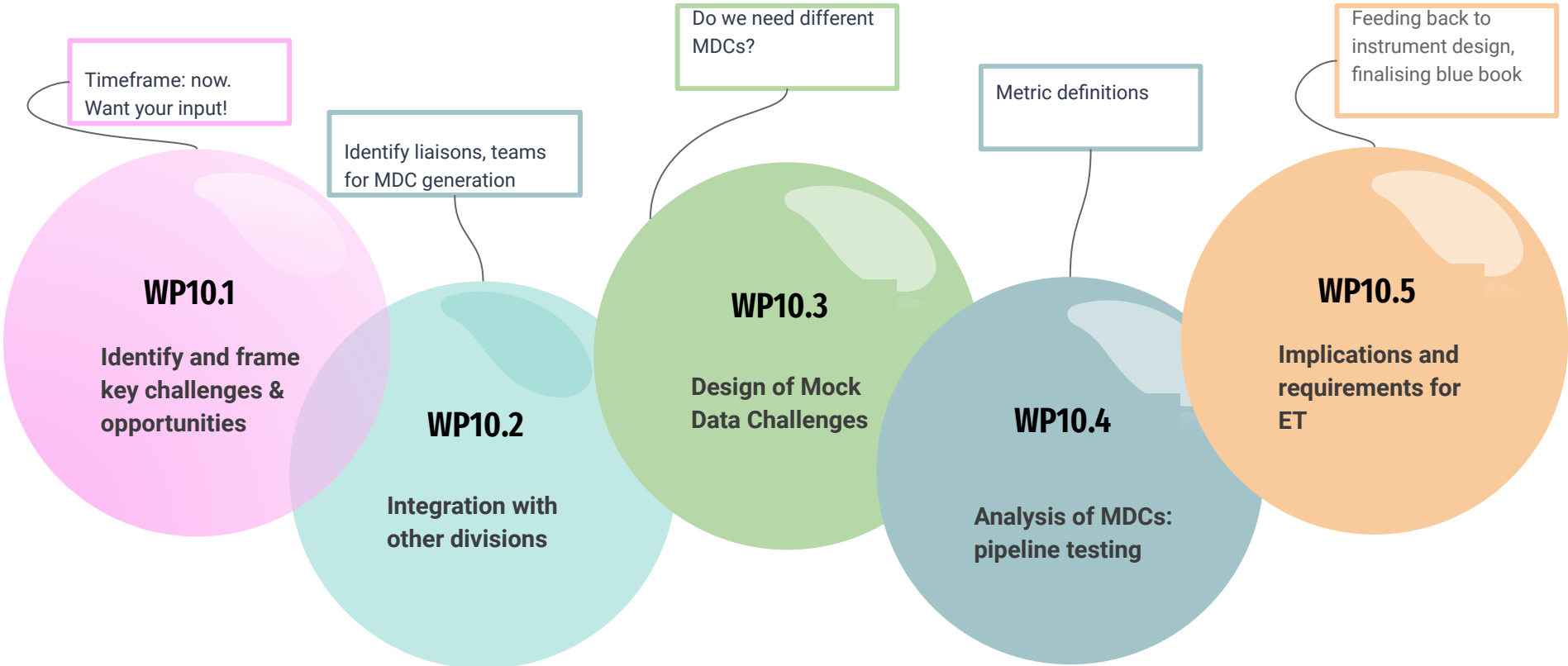
Interactions with other Divisions

- MDC design
  - Gathering requirements from observational science divisions - co-design problems.
- Initial request for Inputs:
  - **DIV1**: What kinds of exotic signals?
  - **DIV3**: Realistic signal rates, population
  - **DIV6**: Realistic EOS
  - **DIV8**: Waveform choices
  - **DIV9**: Detector configuration(s)
  - **DIV7**: Burst MDCs?
- Building analysis teams to solve MDCs
  - Timelines and deadlines to be discussed
  - Liaisons with other divisions
- Evaluation
- Interaction with Instrumental Design
- Interaction with Cosmic Explorer - co-MDC?

Analysis Platform

What are the final requirements for ET?

- Latency
  - Early warning requirements?
    - Liaise with **DIV4** for multimessenger scenarios
  - Computational throughput
    - signals / second? (observed rate)
    - Cost / signal (algorithmic requirements)
- Scalability
  - Cluster platform?
  - Cloud computing?
  - What training will be required for on-boarding researchers, students, etc?
- Data Archiving
  - Reproducibility
  - Bulk data storage requirements?
- Aim to produce minimum requirements estimate for blue book.



Work packages (in progress)

Div10 tools and plans

- Wiki page: https://wiki.et-gw.eu/OSB/DataAnalysisPlatform/WebHome
- Mailing list: to be setup with every contributors/ fun
- Gitlab repo: https://gitlab.et-gw.eu/
- Regular meeting: we will start biweekly meeting
- Many colleagues already contacted and accepted to contribute. Please send your interest to us if you want to contribute
- **Div10 general meeting** to organize the contribution to the different WPs

Discussion Topics

- MDC design
- DIV10 liaisons
- Science requirements
- Platform & pipeline ideas
- Low latency