LISA

Data analysis for observing gravitational wave sources from space

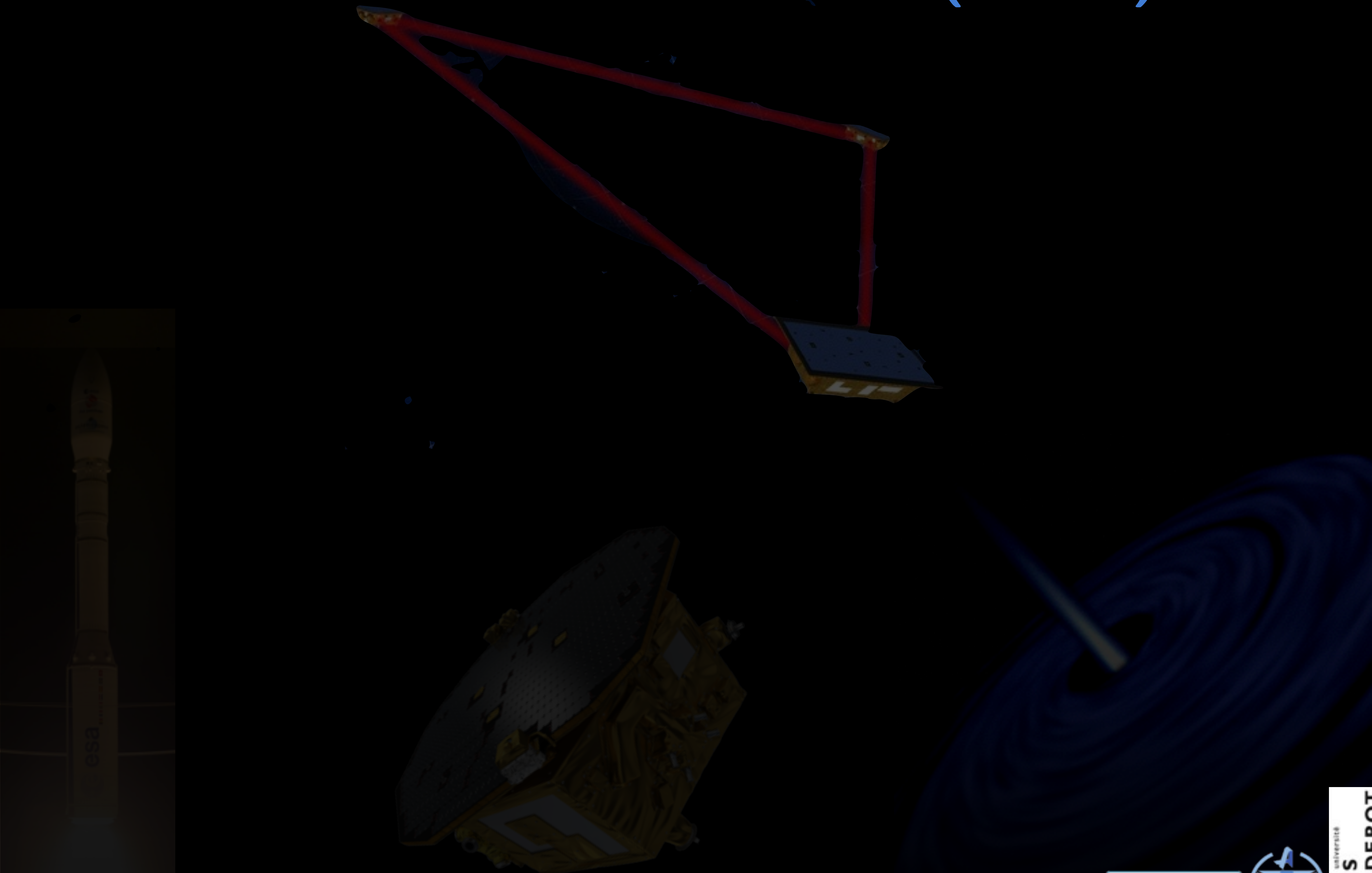
Antoine Petiteau

(APC – Université de Paris / CNRS)

Workshop GPU @CC-IN2P3 - CCIN2P3 - Lyon

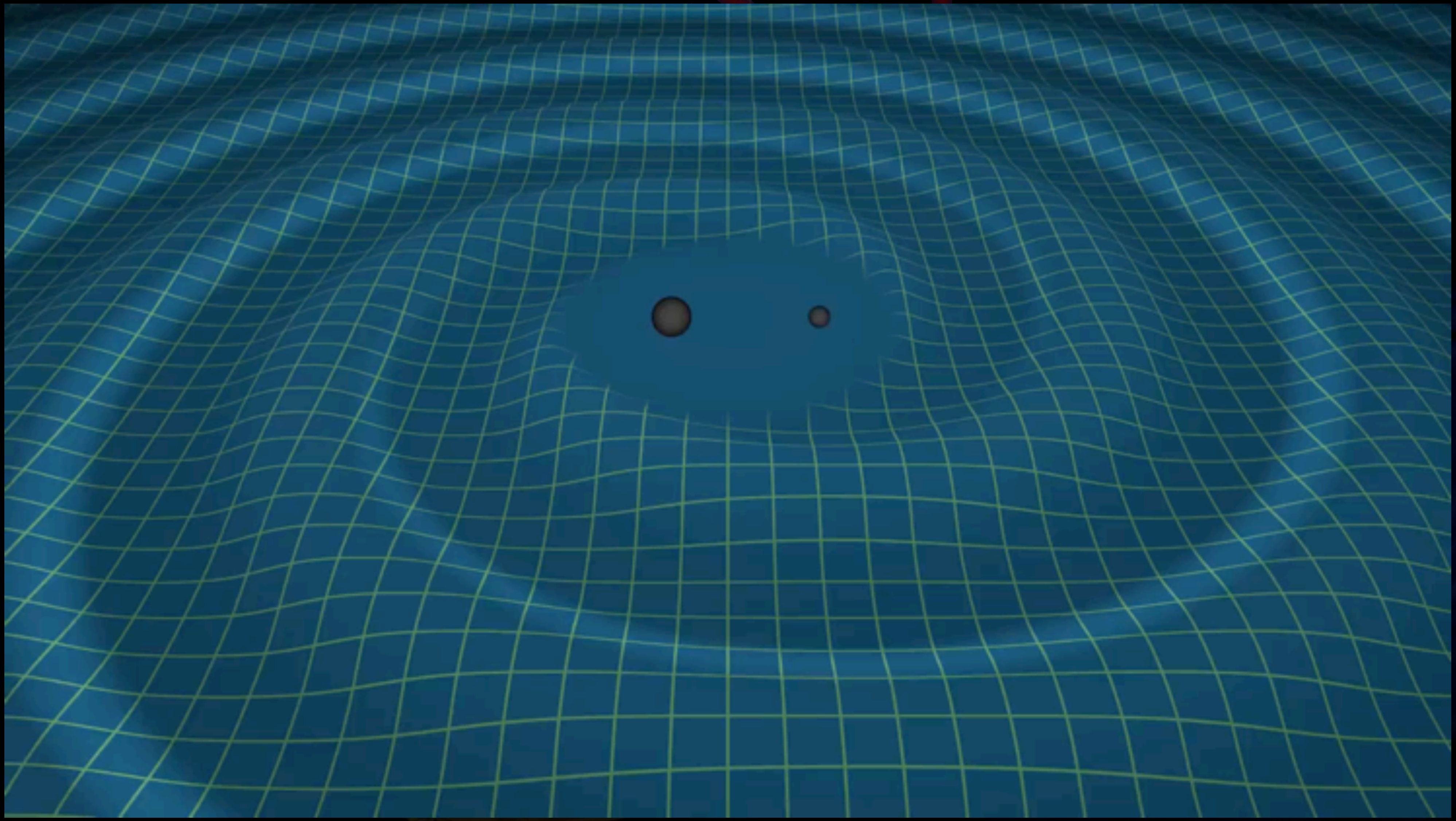
3rd April 2018

Gravitational waves (GWs)



Gravitational waves (GWs)

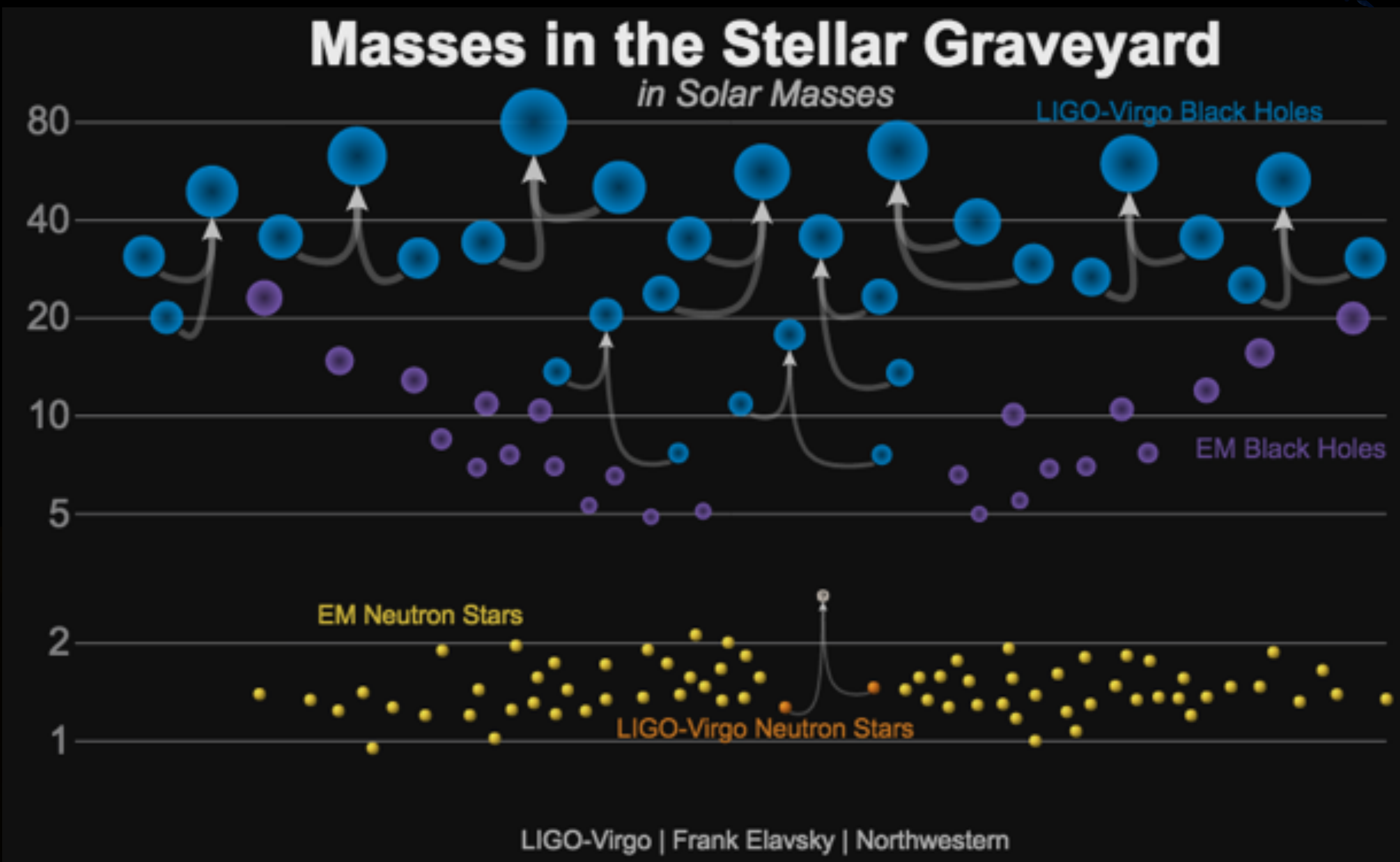
- General relativity: GW are created by non-spherical acceleration of one or several massive objects (asymetric collapse, bodies in orbits or coalescing)

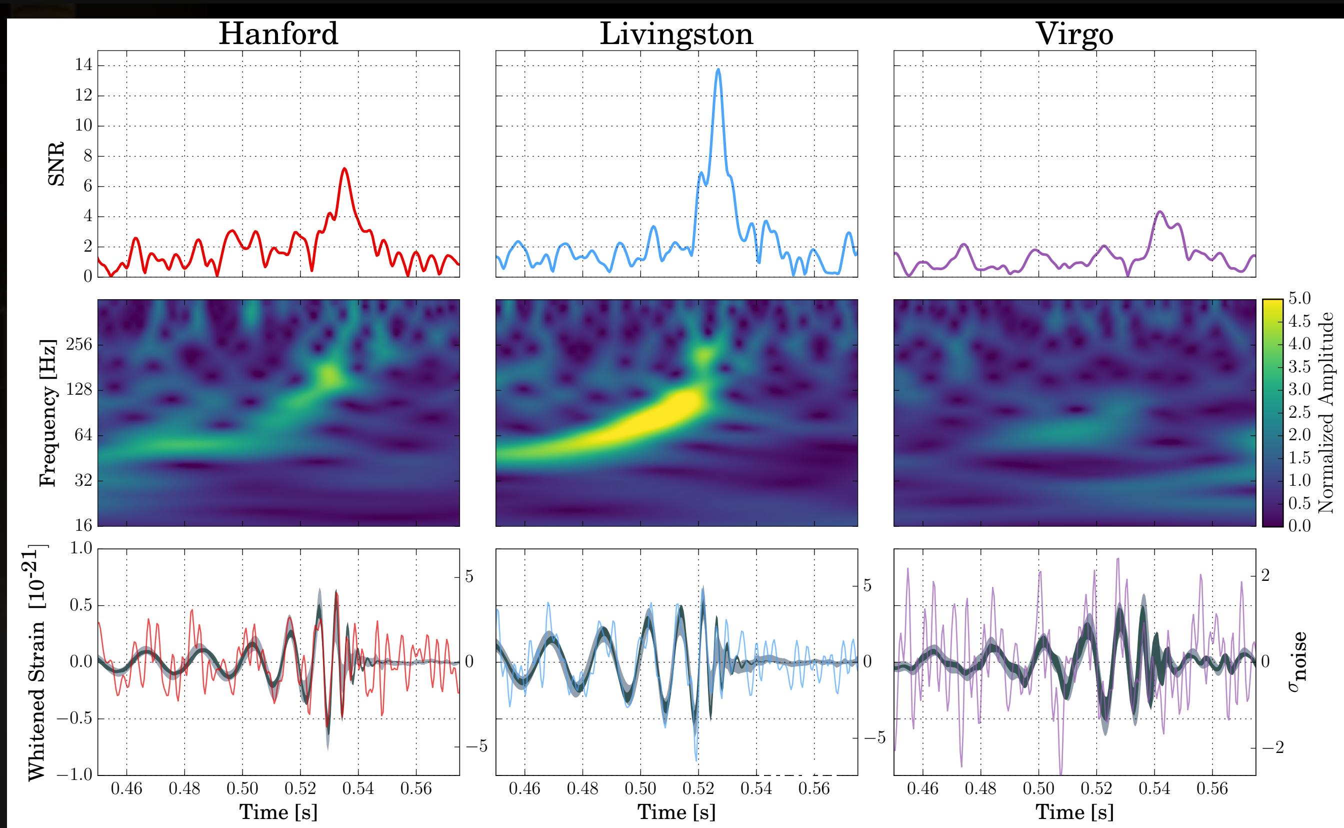


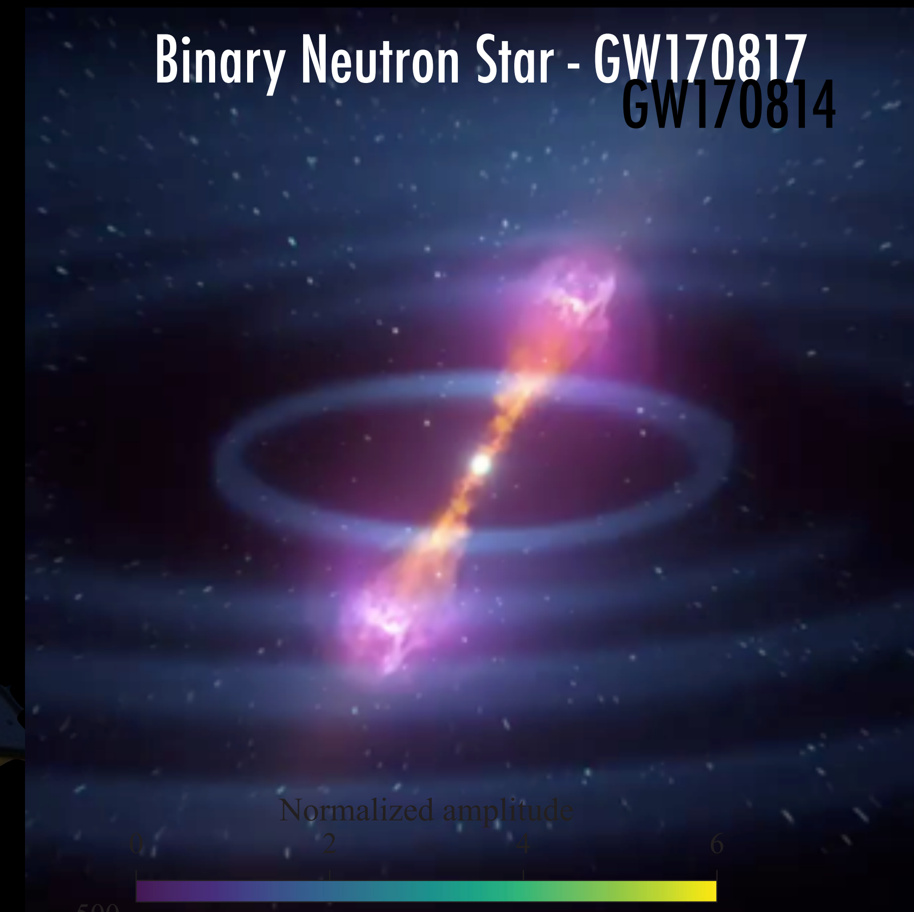
Gravitational waves (GWs)

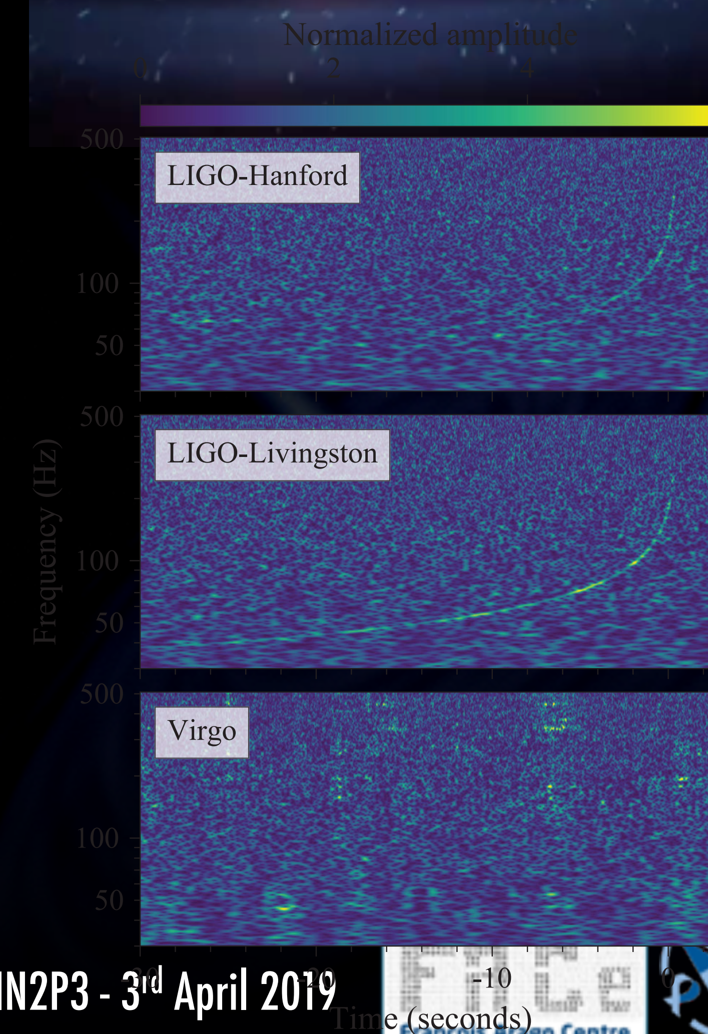
- General relativity: GW are created by non-spherical acceleration of one or several massive objects (asymetric collapse, bodies in orbits or coalescing)
- Modification of distance between 2 objects:
  - Elastic deformation proportional to the distance between the 2 obj.,
  - Transverse deformation: perpendicular to the direction of propagation (different from ripples on water !),
  - Two components of polarisation : h_+ and $h_\times$

Ground-based obs.: GWs detected

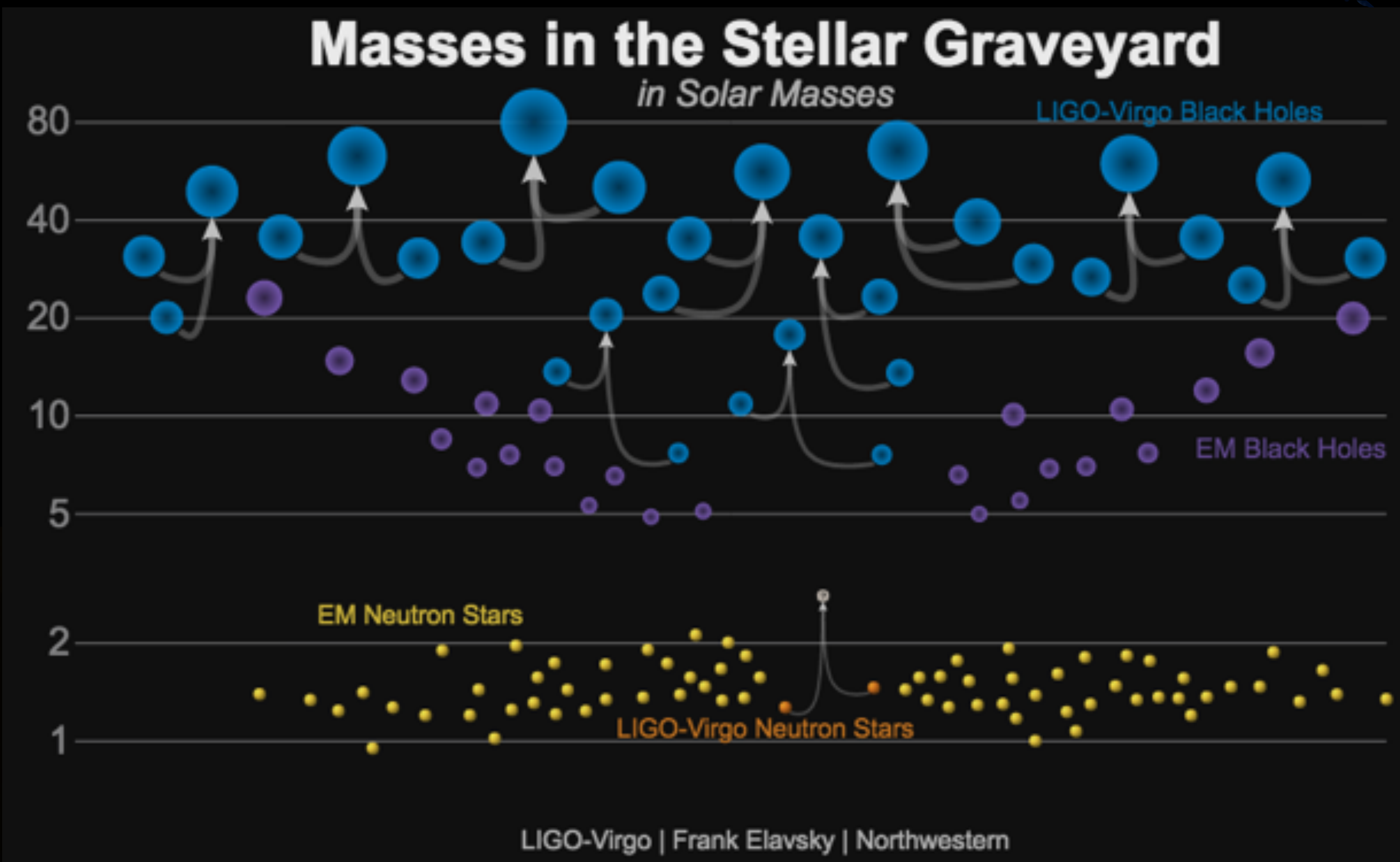


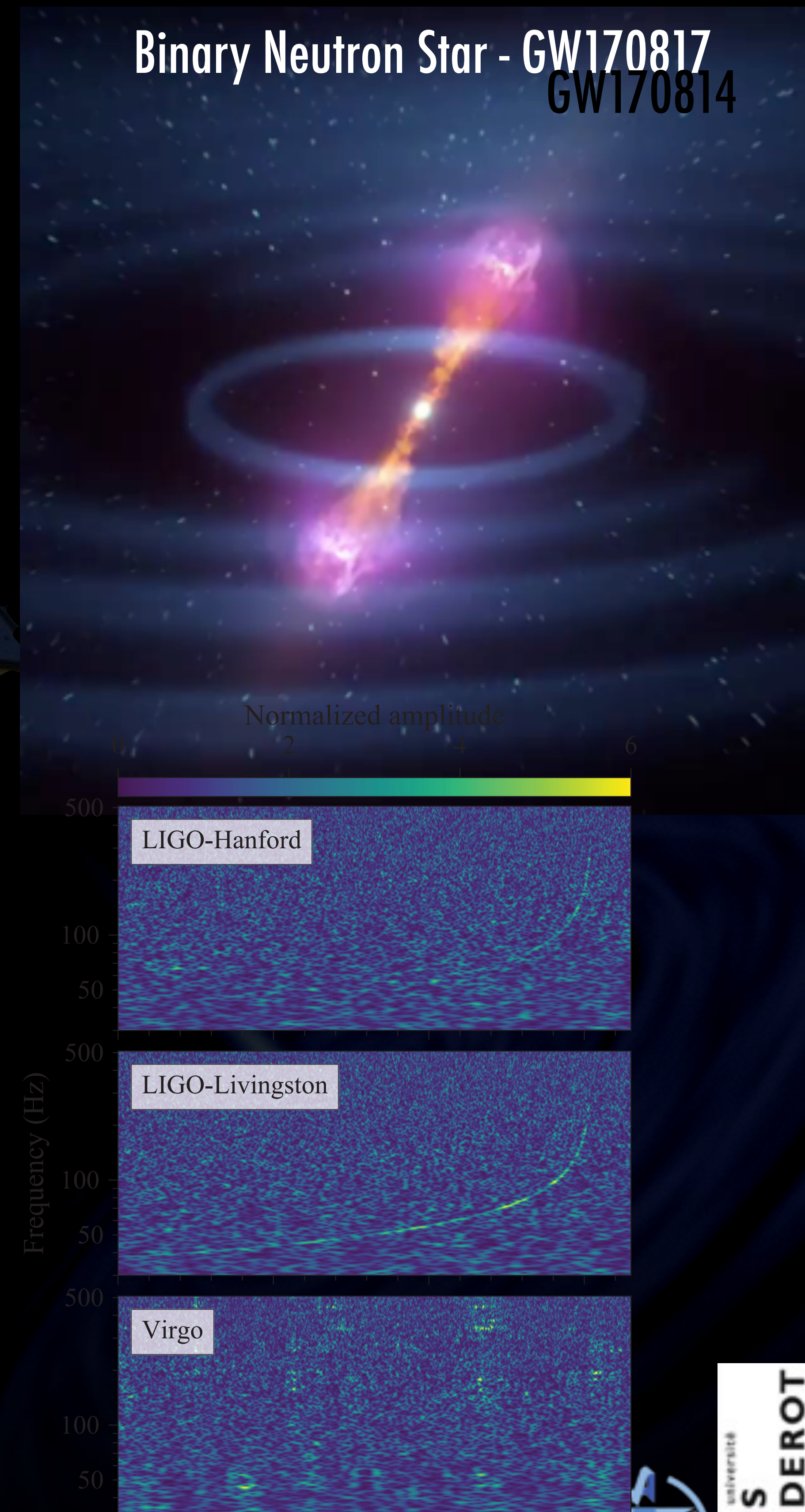


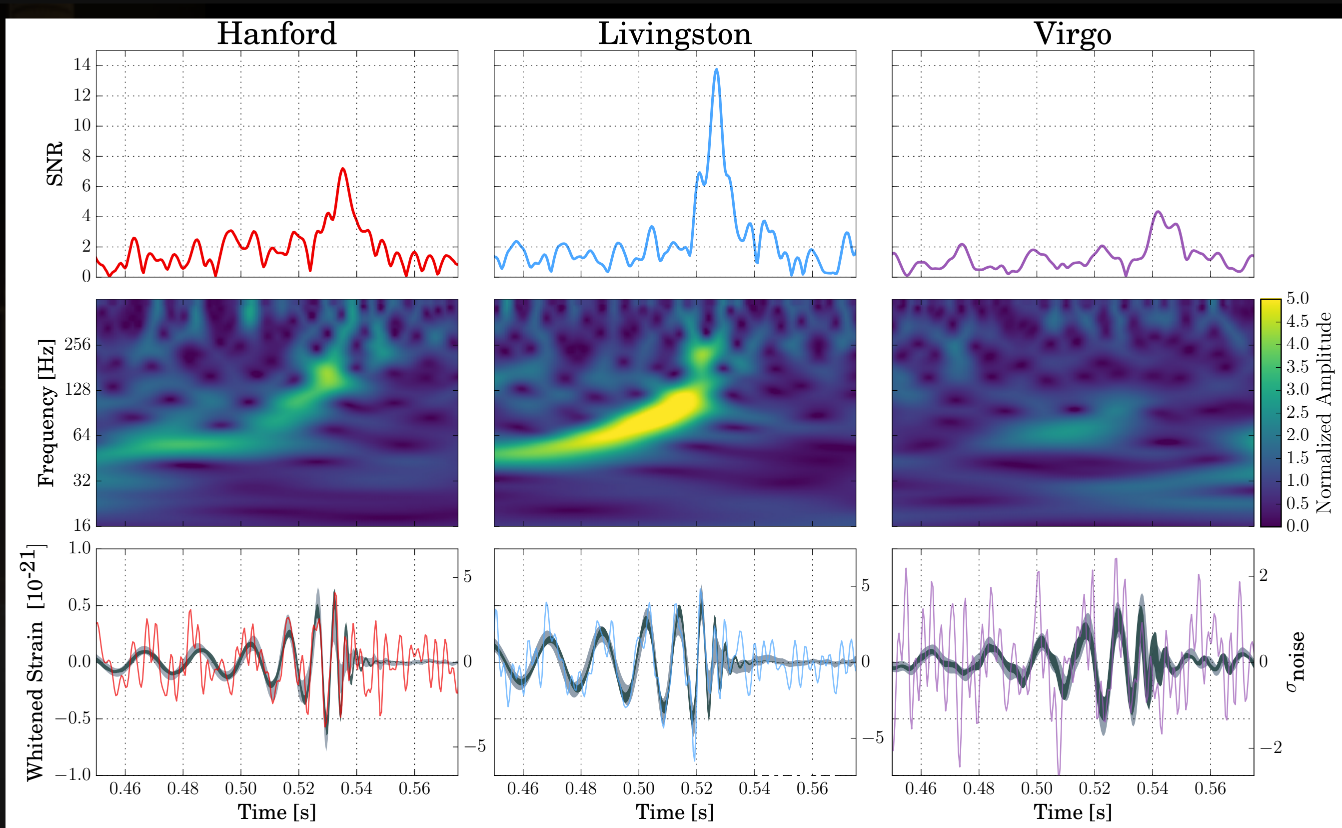




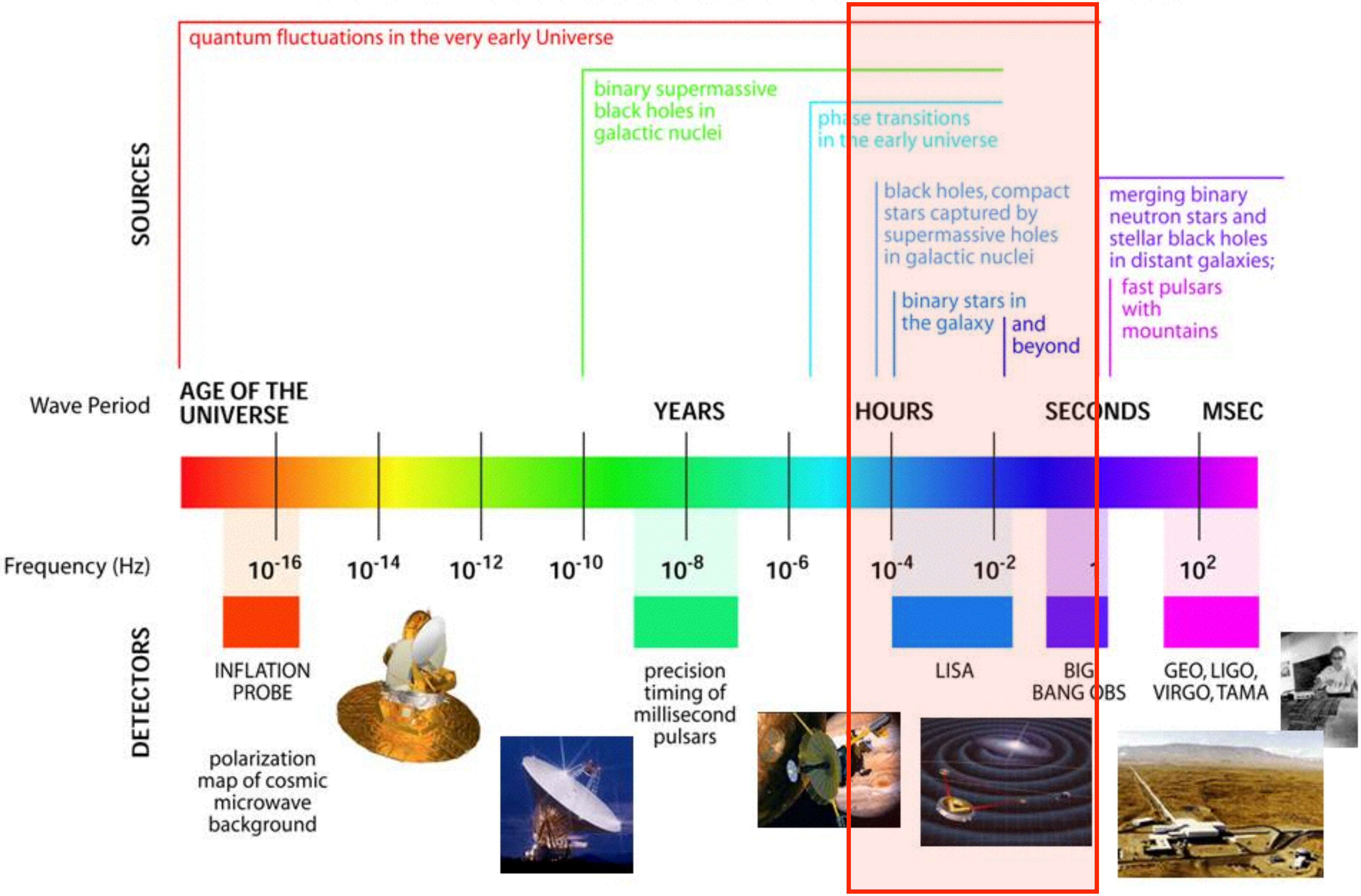
Ground-based obs.: GWs detected





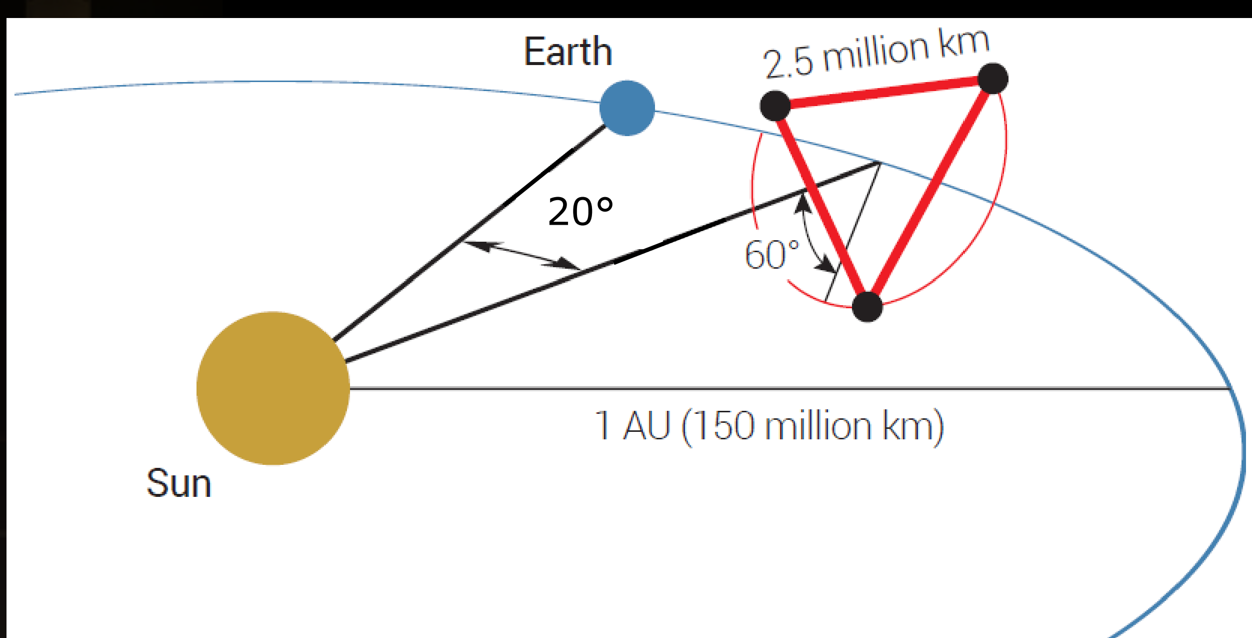


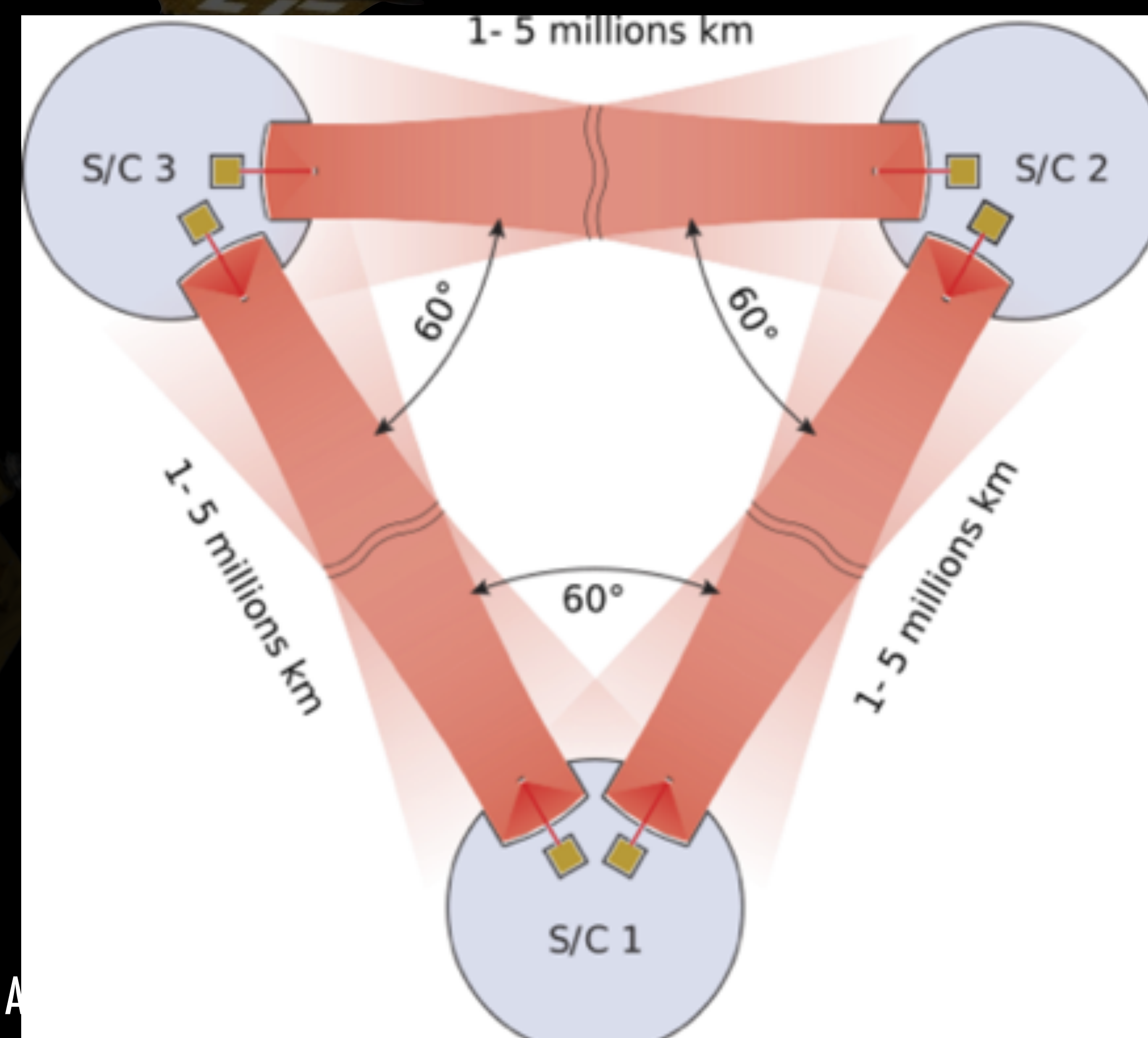
THE GRAVITATIONAL WAVE SPECTRUM



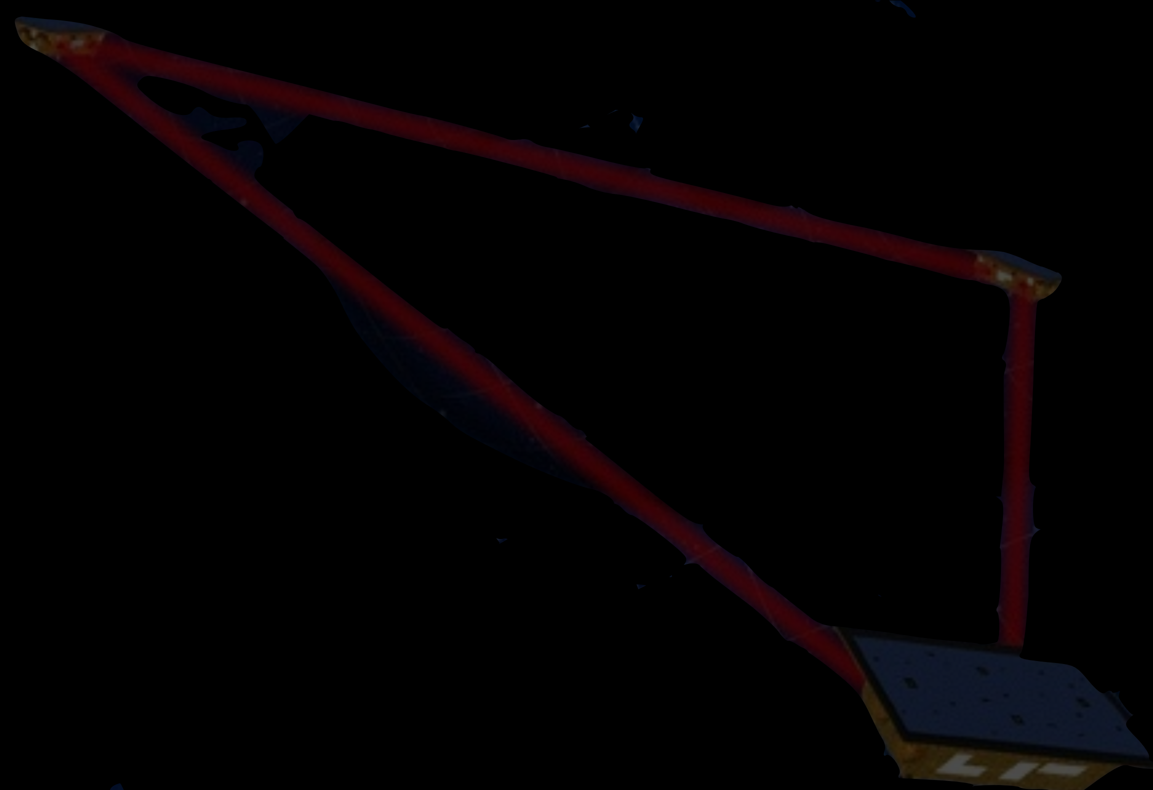
LISA

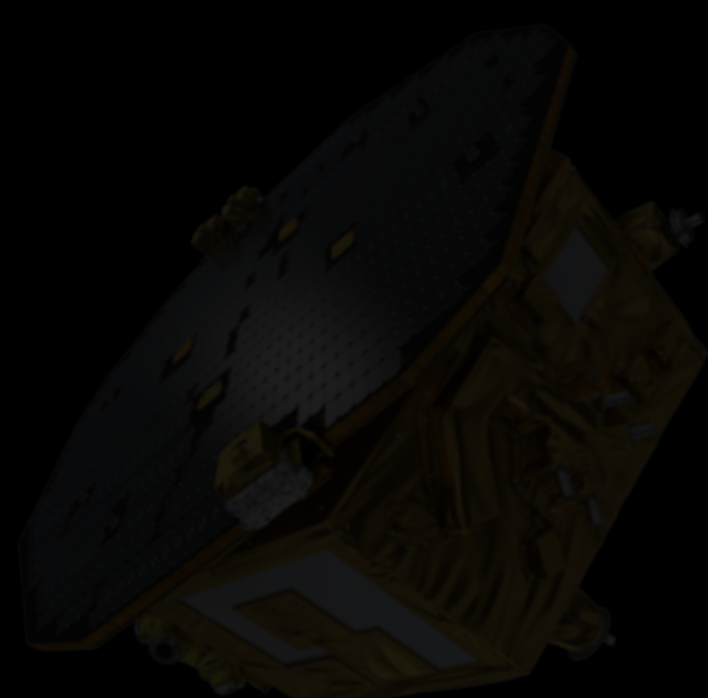
- Laser Interferometer Space Antenna
- 3 spacecrafts on heliocentric orbits and distant from 2.5 millions kilometers
- Goal: detect relative distance changes of 10^{-21}: few picometers

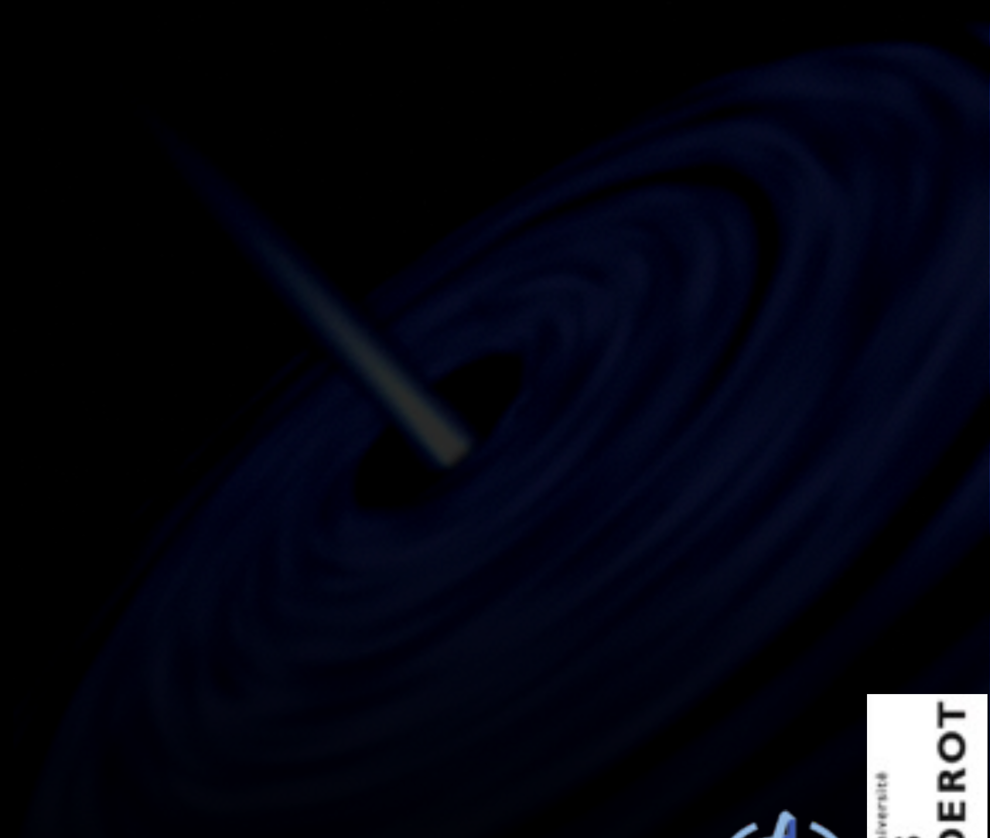




LISA

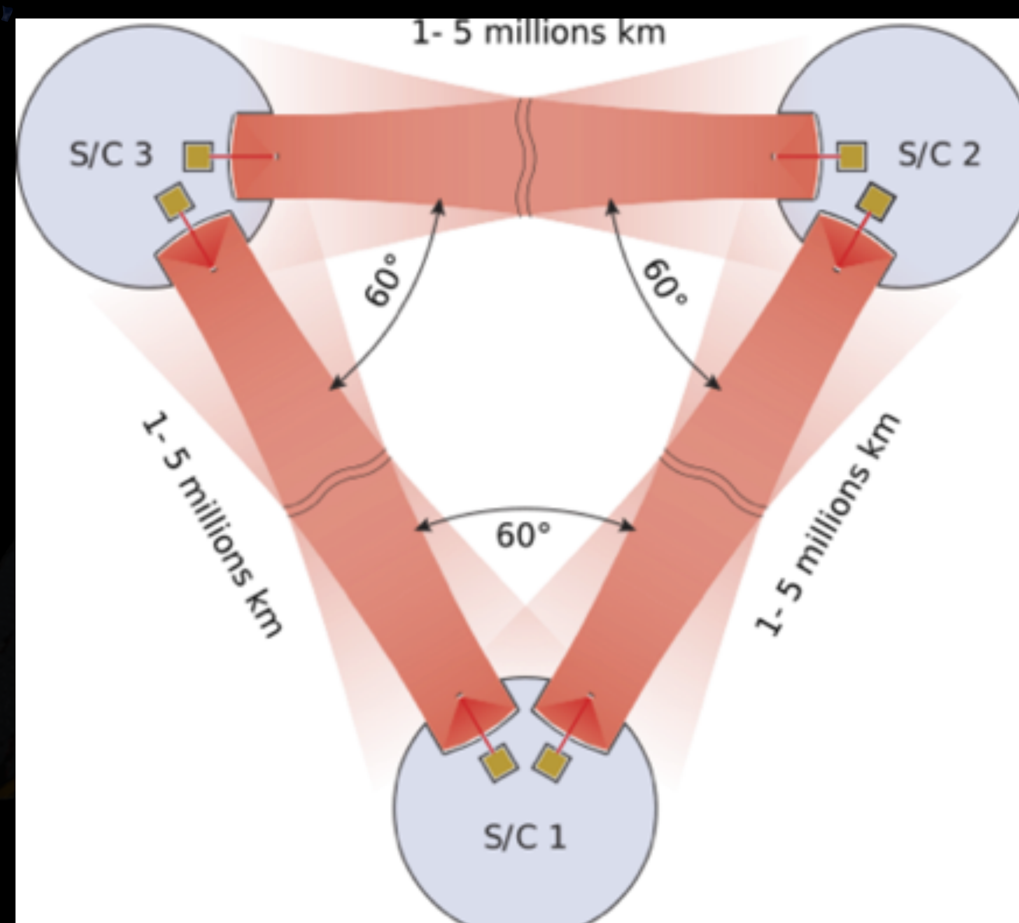






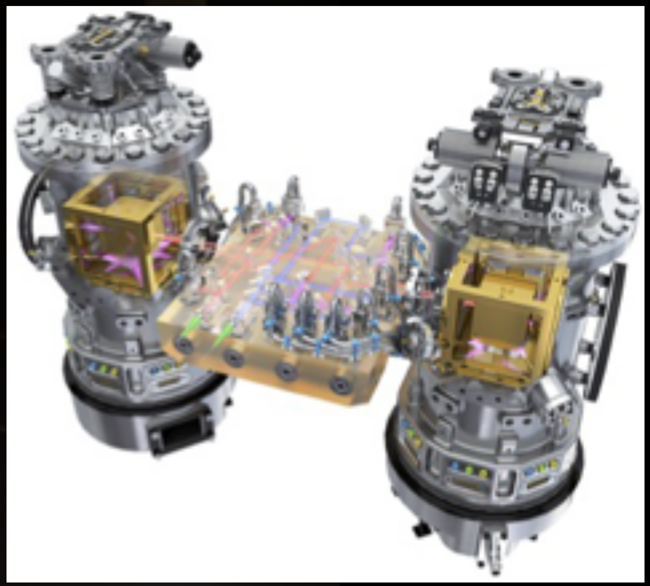
LISA

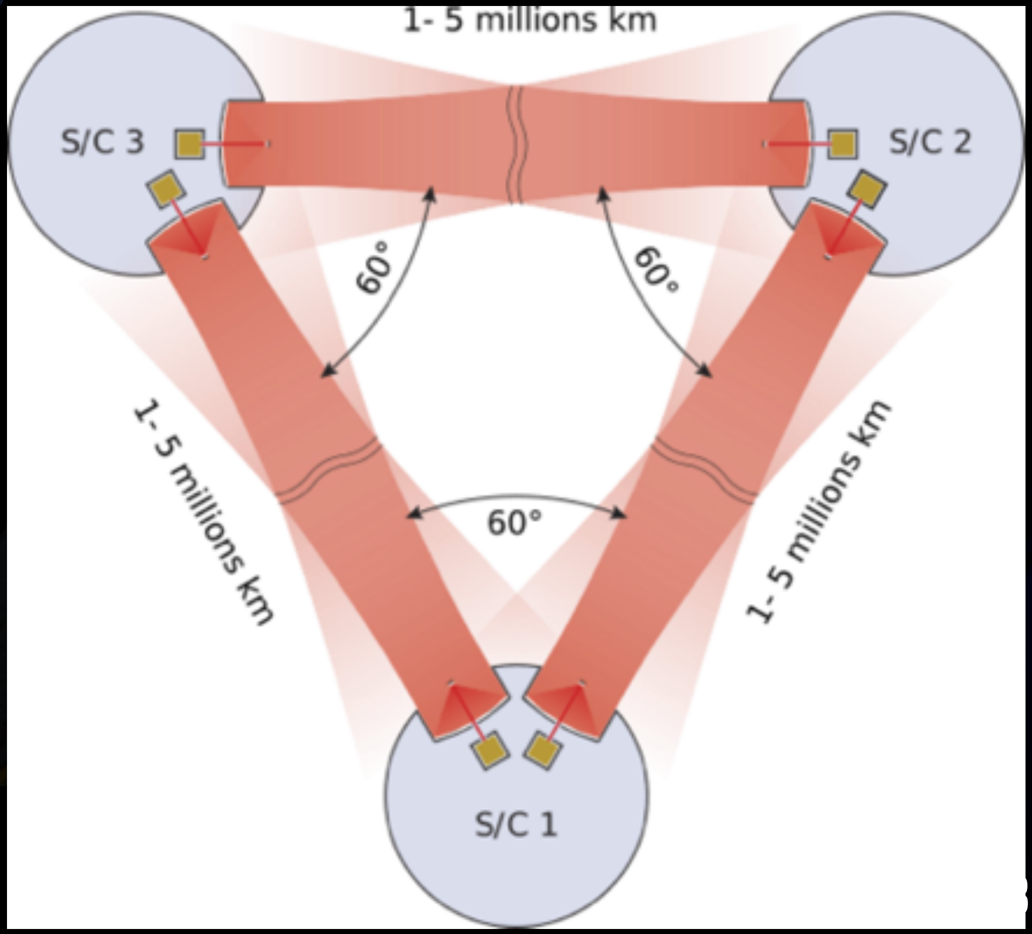
- Photon flight time measurement between free-floating objects:



LISA

- Photon flight time measurement between free-floating objects:
  - Reference masses in each spacecraft only sensitive to gravity along measurement axis (follow geodesics)



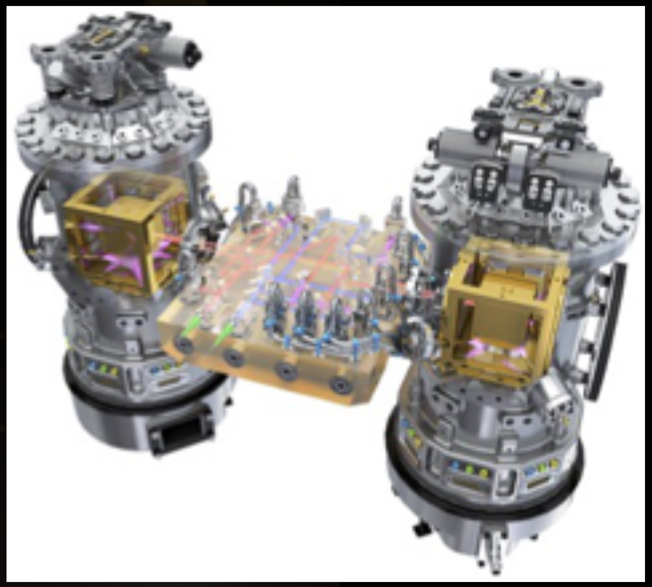


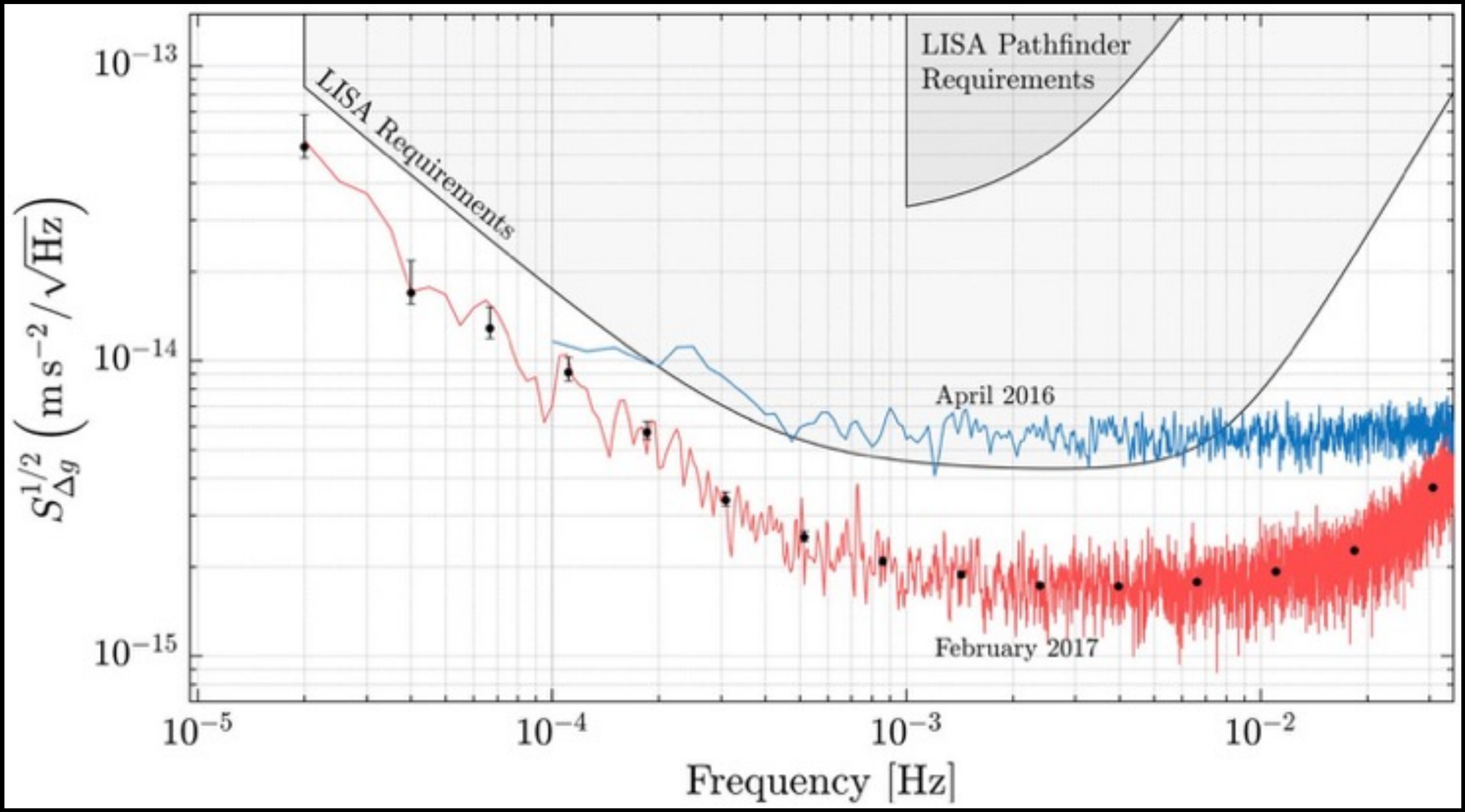
LISA

- Photon flight time measurement between free-floating objects:
  - Reference masses in each spacecraft only sensitive to gravity along measurement axis (follow geodesics)

LISAPathfinder success:

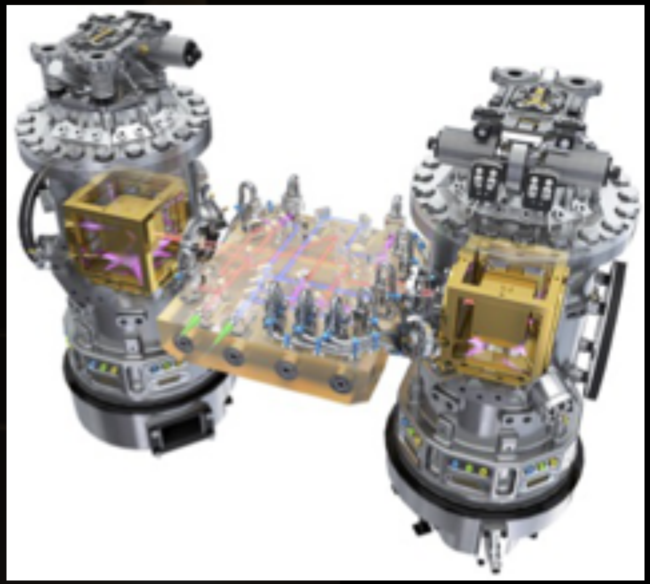
M. Armano et al. PRL 120, 061101 (2018)

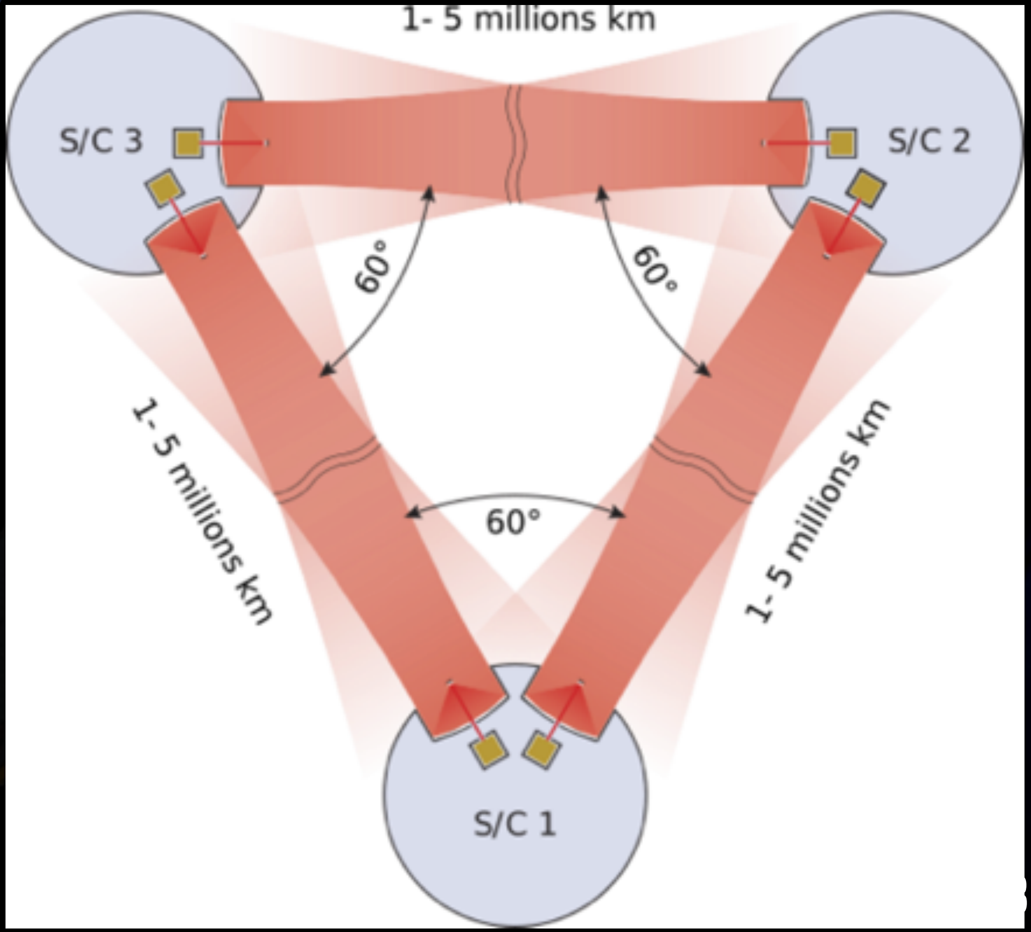


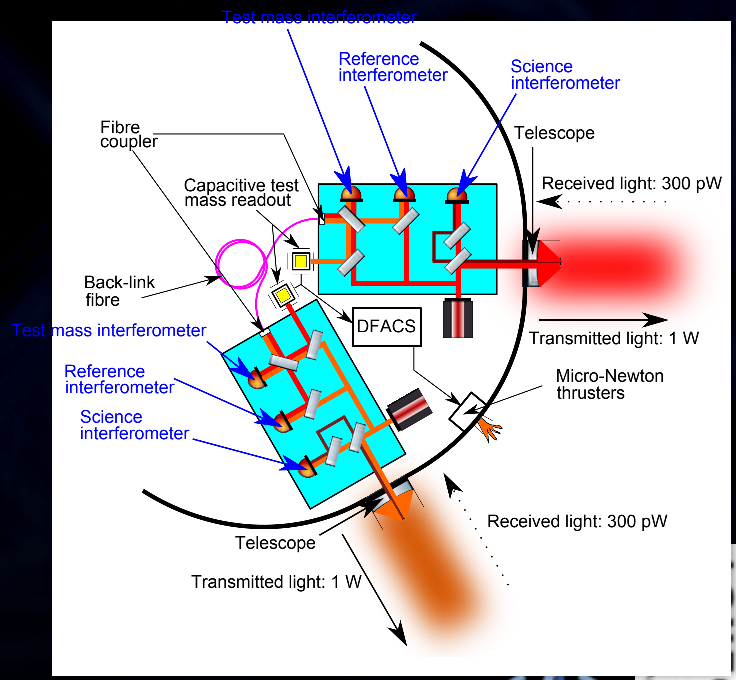


LISA

- Photon flight time measurement between free-floating objects:
  - Reference masses in each spacecraft only sensitive to gravity along measurement axis (follow geodesics)
  - Exchange of laser beam between spacecraft
  - Interferometry at the picometer precision

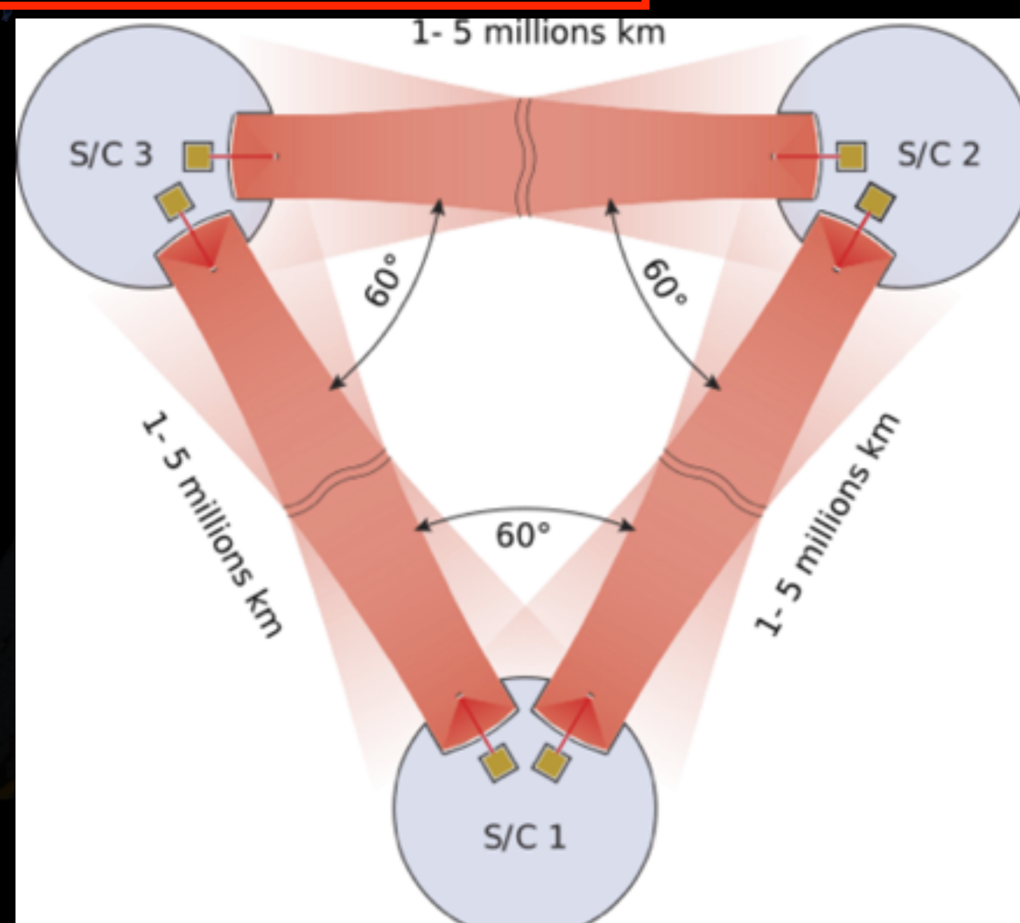


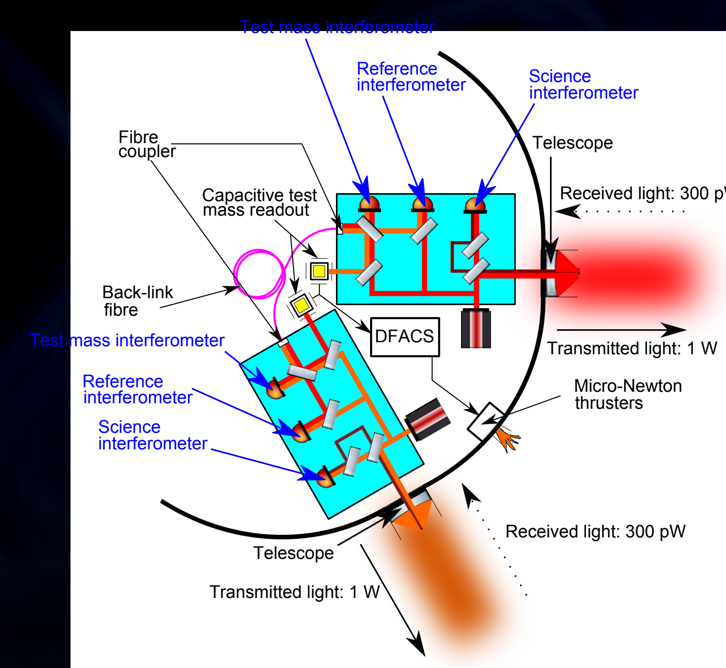




LISA

- Photon flight time measurement between free-floating objects:
  - Reference masses in each spacecraft only sensitive to gravity along measurement axis (follow geodesics)
  - Exchange of laser beam between spacecraft
  - Interferometry at the picometer precision
  - Extract GW signals from the data





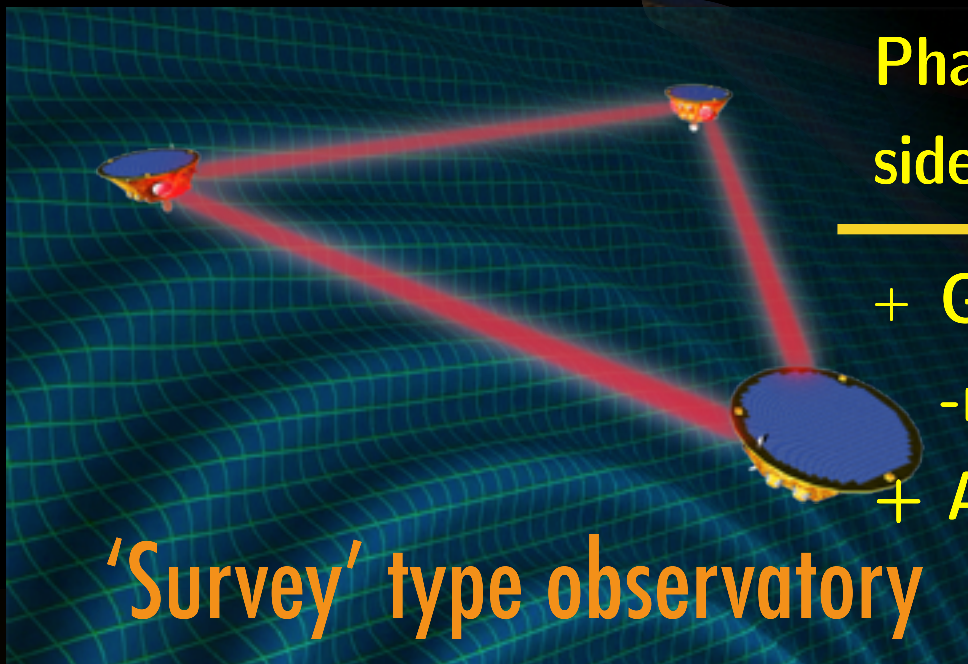
LISA at ESA

- 25/10/2016 : Call for mission
- 13/01/2017 : submission of «LISA proposal» (LISA consortium)
- 8/3/2017 : Phase 0 mission (CDF 8/3/17 → 5/5/17)
- 20/06/2017 : LISA mission approved by SPC
- 8/3/2017 : Phase 0 payload (CDF June → November 2017)
- 2018→2020 : competitive phase A: 2 companies compete
- 2020→2022 : B1: start industrial implementation
- 2023 : mission adoption
- During about 8.5 years : construction
- 2030-2034 : launch Ariane 6.4
- 1.5 years for transfert
- 4 years of nominal mission → GW observations !
- Possible extension to 10 years

LISA science objectives

- SO1: Study the formation and evolution of compact binary stars in the Milky Way Galaxy.
- SO2: Trace the origin, growth and merger history of massive black holes across cosmic ages
- SO3: Probe the dynamics of dense nuclear clusters using EMRIs
- SO4: Understand the astrophysics of stellar origin black holes
- SO5: Explore the fundamental nature of gravity and black holes
- SO6: Probe the rate of expansion of the Universe
- SO7: Understand stochastic GW backgrounds and their implications for the early Universe and TeV-scale particle physics
- SO8: Search for GW bursts and unforeseen sources

LISA data



Phasemeters (carrier, sidebands, distance)

+ Gravitational Refe- -rence Sensor
+ Auxiliary channels

Calibrations corrections

Resynchronisation (clock)

Time-Delay Interferometry reduction of laser noise

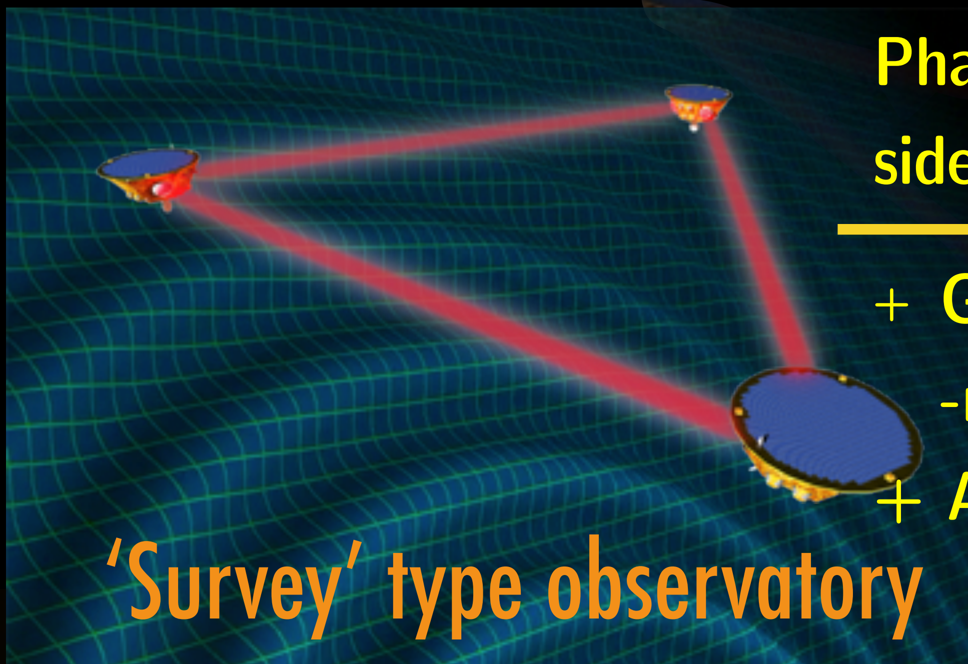
3 TDI channels with 2 "~independents"

Data Analysis of GWs

Catalogs of GWs sources with their waveform

Gravitational wave sources emitting between 0.02mHz and 1 Hz

LISA data



Phasemeters (carrier, sidebands, distance)

+ Gravitational Reference Sensor
+ Auxiliary channels

'Survey' type observatory

Gravitational wave sources emitting between 0.02mHz and 1 Hz

L0

Calibrations corrections

Resynchronisation (clock)

Time-Delay Interferometry reduction of laser noise

L1

3 TDI channels with 2 "~independents"

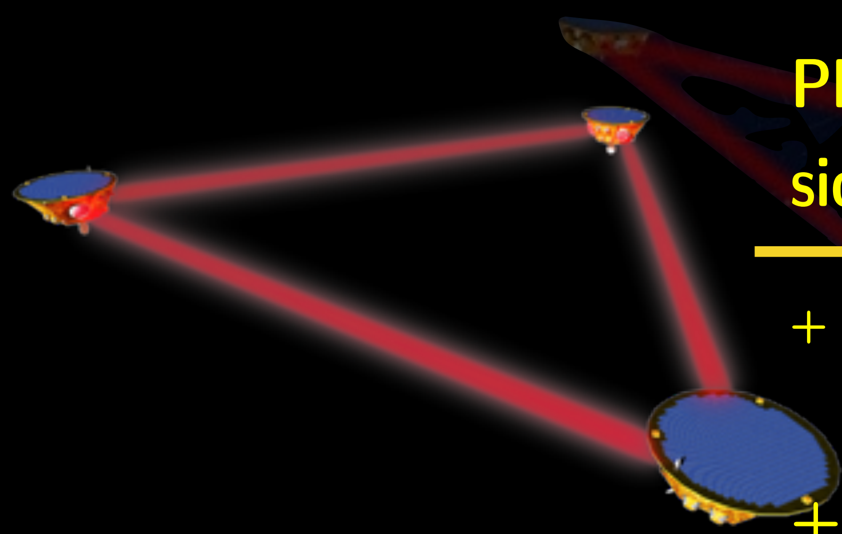
L2

Data Analysis of GWs

L3

Catalogs of GWs sources with their waveform

LISA data flow



Phasemeters (carrier, sidebands, distance)

+ Gravitational Reference Sensor

+ Auxiliary channels

‘Survey’ type observatory

GW sources

- 6 x10^7 galactic binaries
- 10-100/year SMBHBs
- 10-1000/year EMRIs
- large number of Stellar Origin BH binaries (LIGO/Virgo)
- Cosmological backgrounds
- Unknown sources

L0

Calibrations corrections

Resynchronisation (clock)

Time-Delay Interferometry reduction of laser noise

L1

3 TDI channels with 2 “~independents”

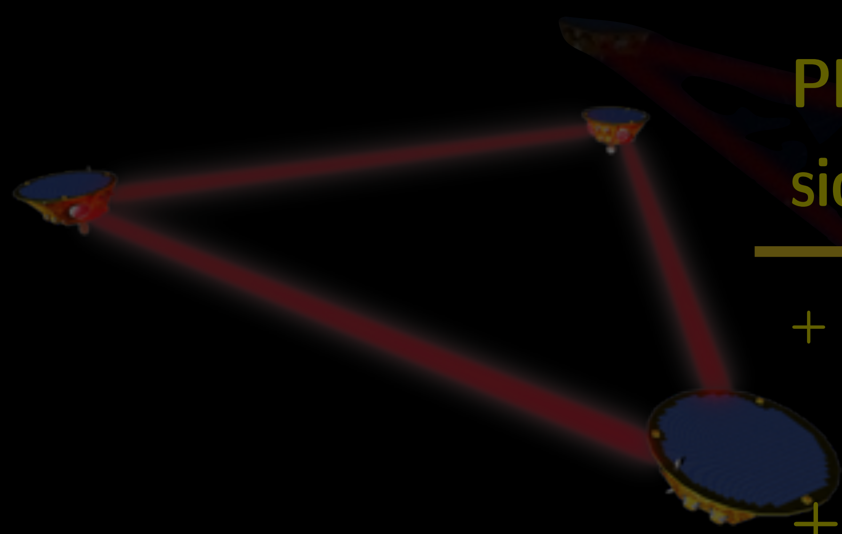
L2

Data Analysis of GWs

L3

Catalogs of GWs sources with their waveform

LISA data flow



Phasemeters (carrier, sidebands, distance)

+ Gravitational Reference Sensor

+ Auxiliary channels

'Survey' type observatory

L0

Calibrations corrections

Resynchronisation (clock)

Time-Delay Interferometry reduction of laser noise

L1 — 3 TDI channels with 2 "~independents"

L2 — Data Analysis of GWs

L3 — Catalogs of GWs sources with their waveform

GW sources

- 6×10^7 galactic binaries
- 10-100/year SMBHBs
- 10-1000/year EMRIs
- large number of Stellar Origin BH binaries (LIGO/Virgo)
- Cosmological backgrounds
- Unknown sources

LISA data flow

Mission Operation Centre

Phasemeters (carrier, ... stance)

\+ Gravitational Reference Sensor

\+ Auxiliary channels

‘Survey’ type observatory

Science Operation Centre

- 6 x10^7 galactic binaries
- 10-100/year SMBHBs
- 10-1000/year EMRIs
- large number of Stellar Origin BH binaries (LIGO/Virgo)
- Cosmological backgrounds
- Unknown sources

L0

Calibrations corrections

Resynchronisation (clock)

Time-Delay Interferometry reduction of laser noise

L1

3 TDI channels with 2 “~independents”

L2

Data Analysis of GWs

L3

Catalogs of GWs sources with their waveform

LISA data flow

Mission Operation Centre

(carrier, ...stance)

+ Gravitational Reference Sensor

+ Auxiliary channels

'Survey' type observatory

Science Operation Centre

- 6×10^7 galactic binaries
- 10-100/year SMBHBs

Distributed Data Processing Centre: LISA Consortium

French responsibility

L0

Calibrations corrections

Resynchronisation (clock)

Time-Delay Interferometry reduction of laser noise

L1

3 TDI channels with 2 "~independents"

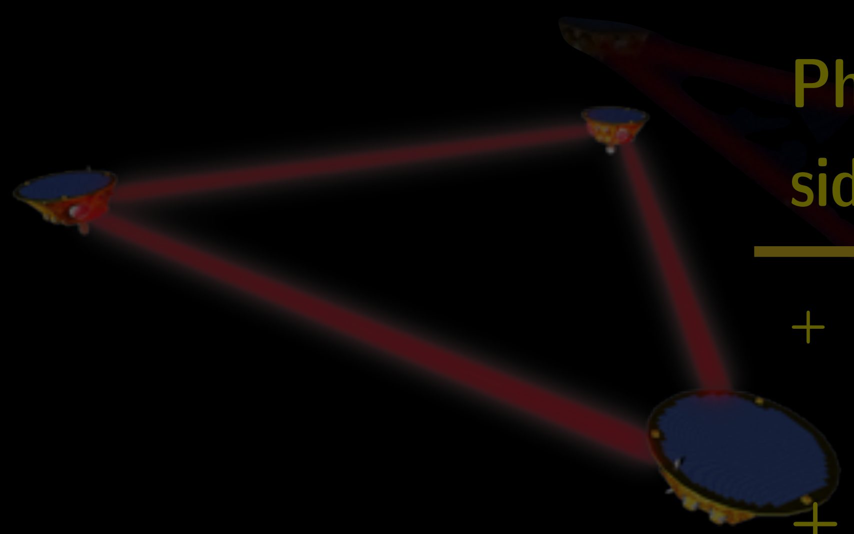
L2

Data Analysis of GWs

L3

Catalogs of GWs sources with their waveform

LISA data flow



Phasemeters (carrier, sidebands, distance)

+ Gravitational Reference Sensor

+ Auxiliary channels

'Survey' type observatory

L0

Calibrations corrections

Resynchronisation (clock)

Time-Delay Interferometry reduction of laser noise

L1

3 TDI channels with 2 "~independents"

L2

Data Analysis of GWs

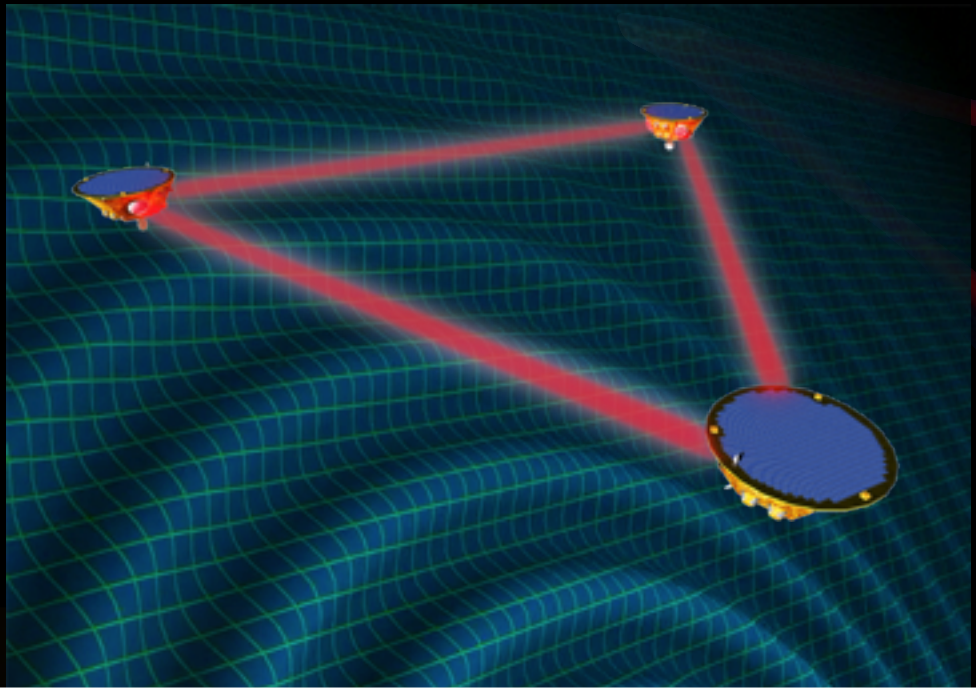
L3

Catalogs of GWs sources with their waveform

GW sources

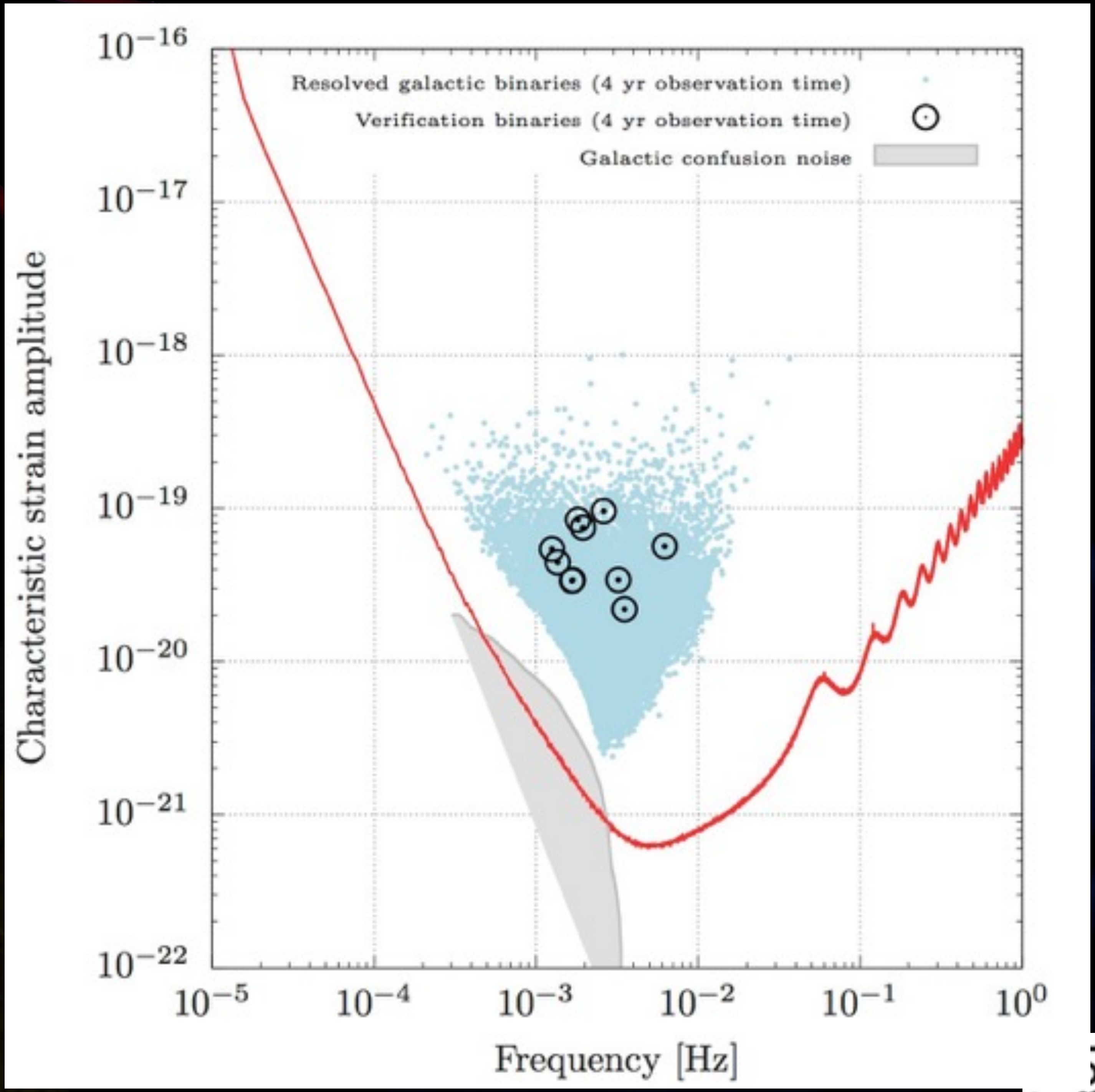
- 6×10^7 galactic binaries
- 10-100/year SMBHBs
- 10-1000/year EMRIs
- large number of Stellar Origin BH binaries (LIGO/Virgo)
- Cosmological backgrounds
- Unknown sources

Galactic binaries

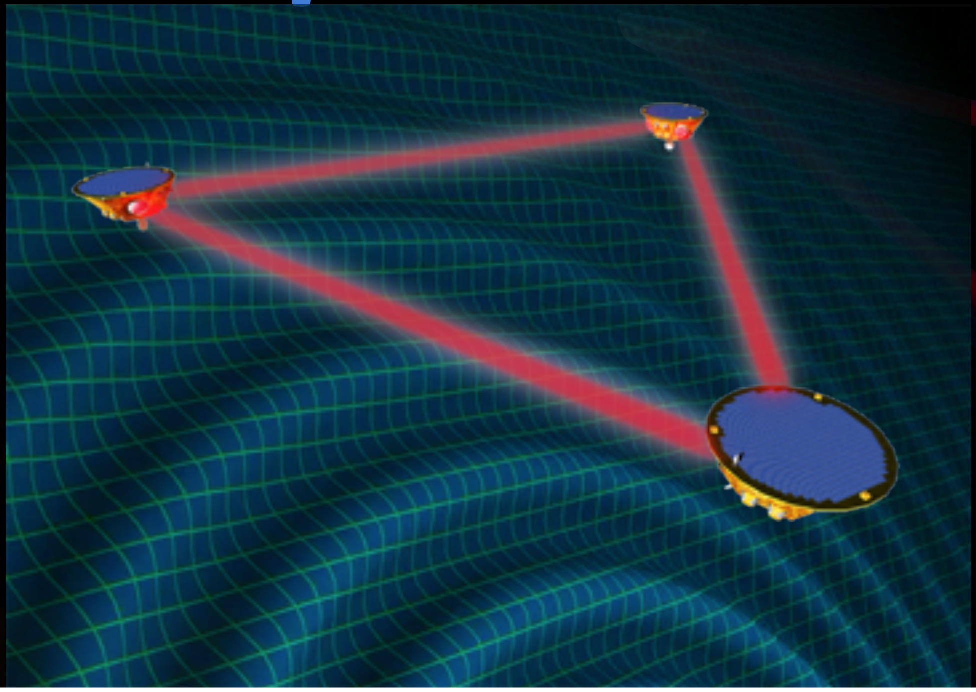


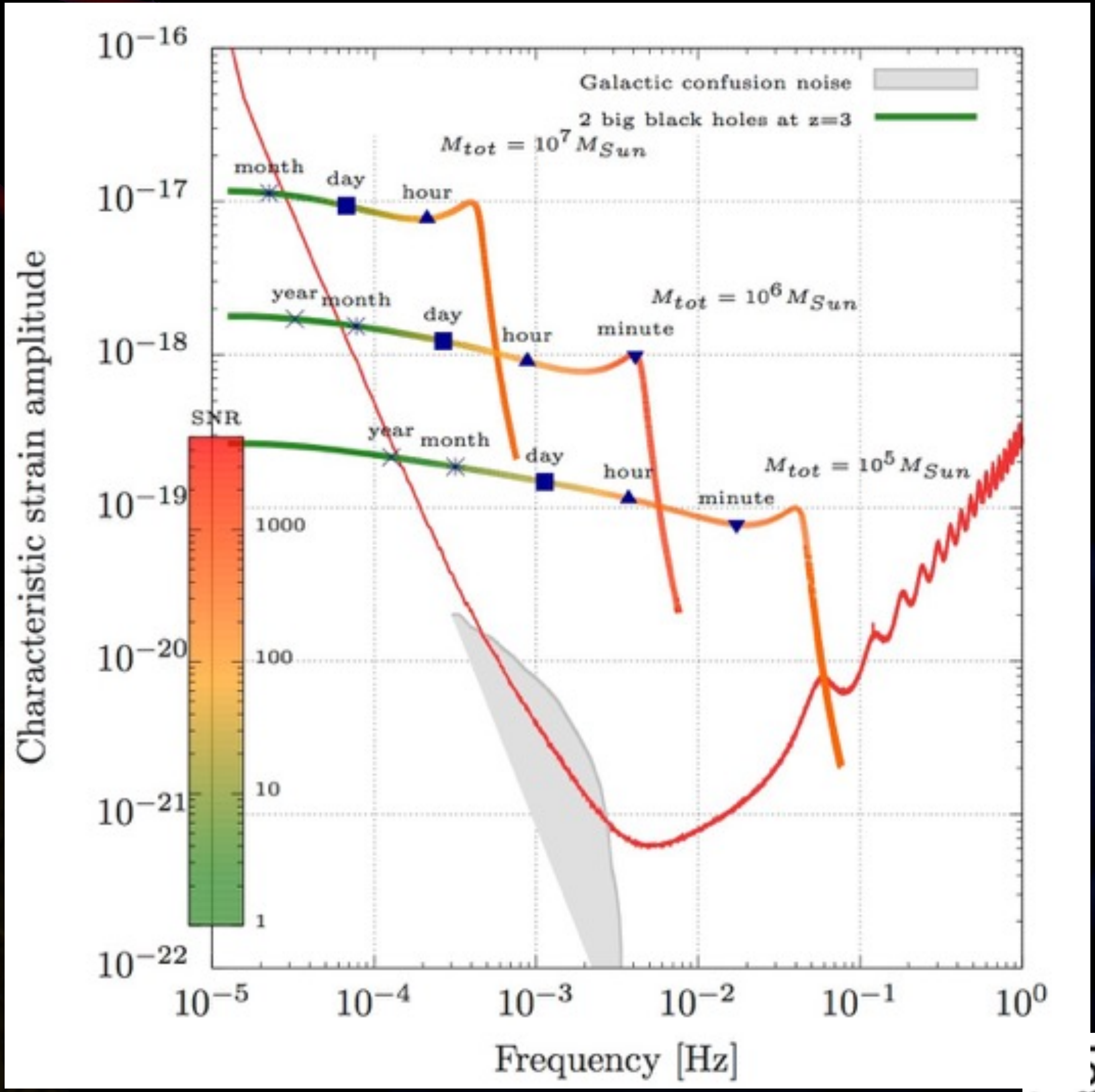
GW sources

- 6 x10^7 galactic binaries



Super Massive Black Hole Binaries

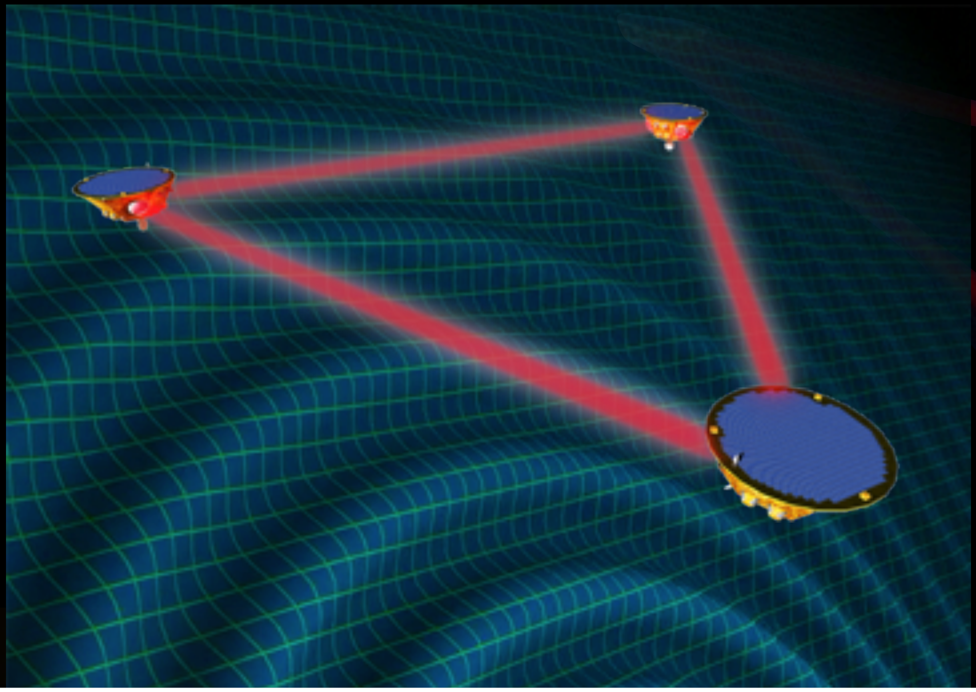




GW sources

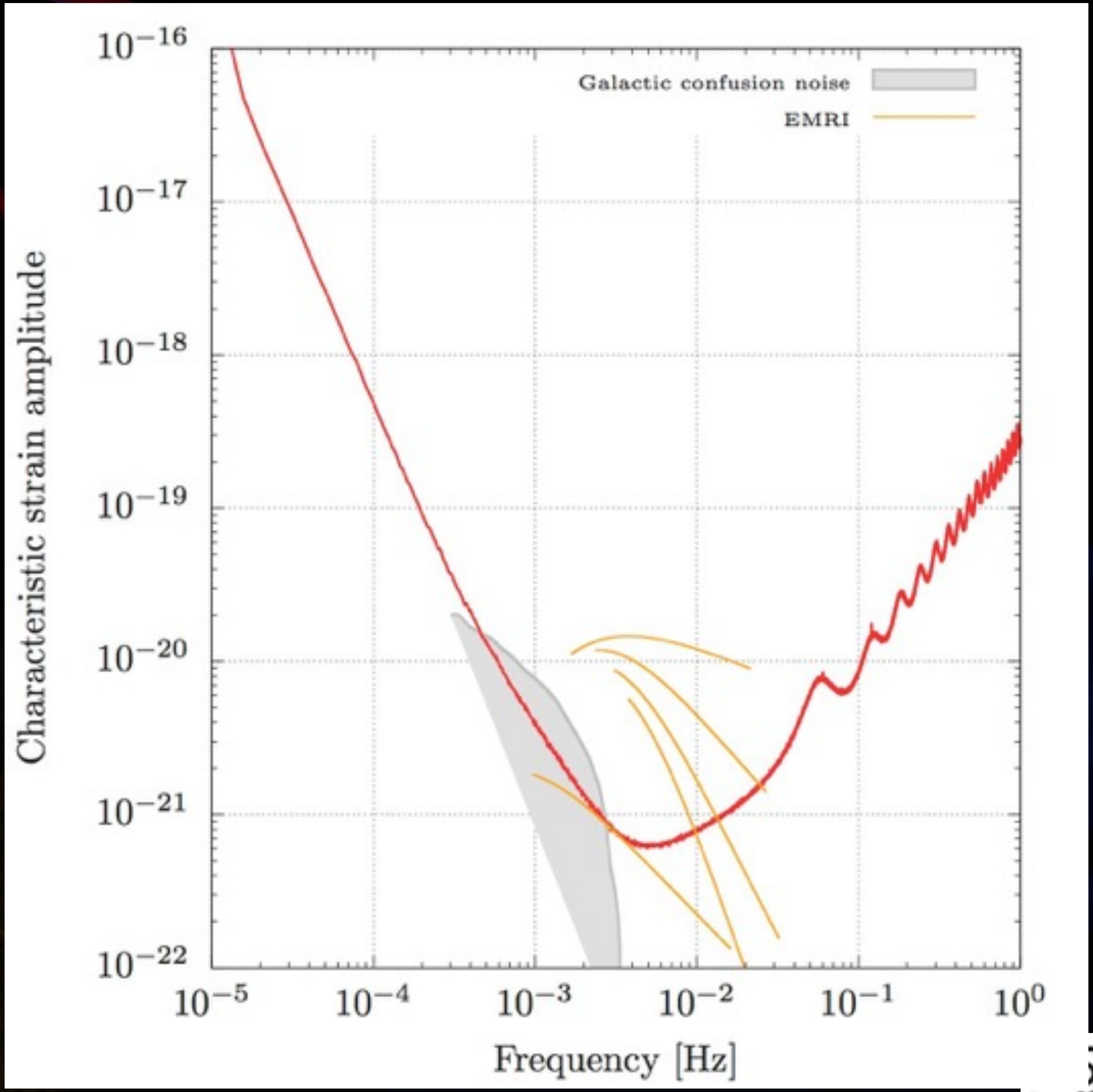
- 6×10^7 galactic binaries
- 10-100/year SMBHBs

EMRIs

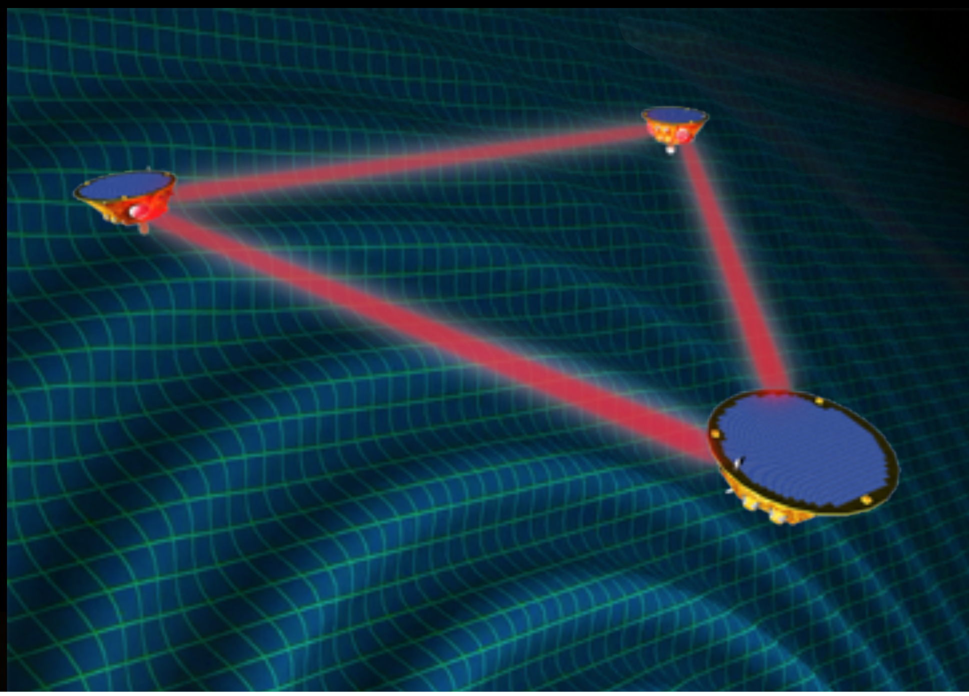


GW sources

- 6 x10^7 galactic binaries
- 10-100/year SMBHBs
- 10-1000/year EMRIs

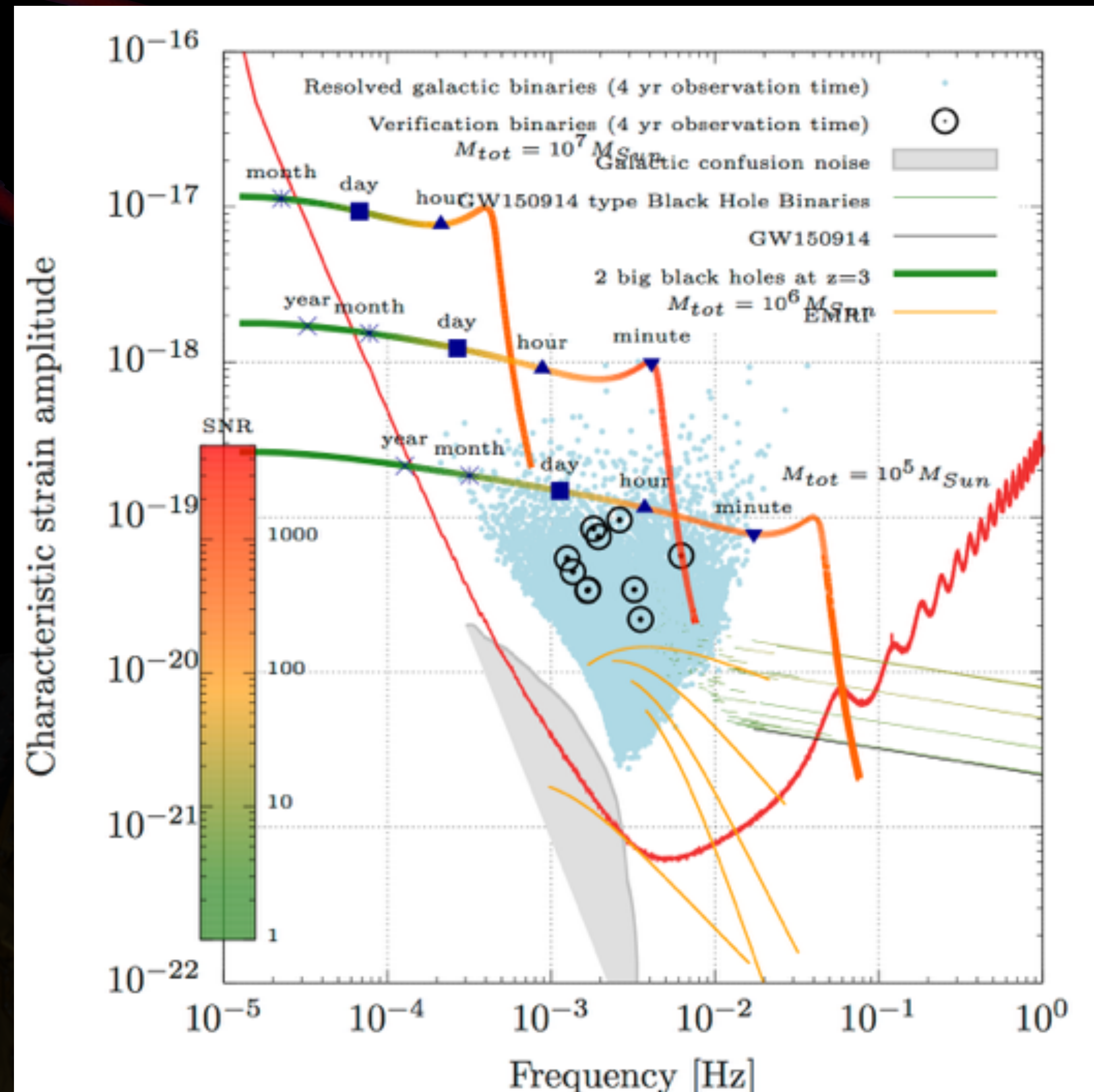


LISA data

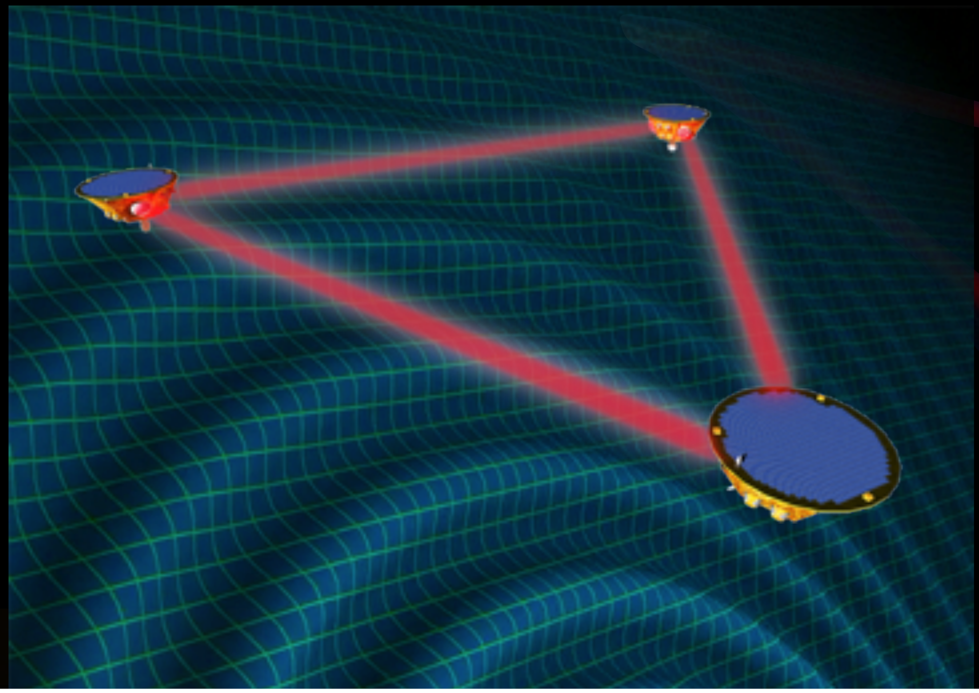


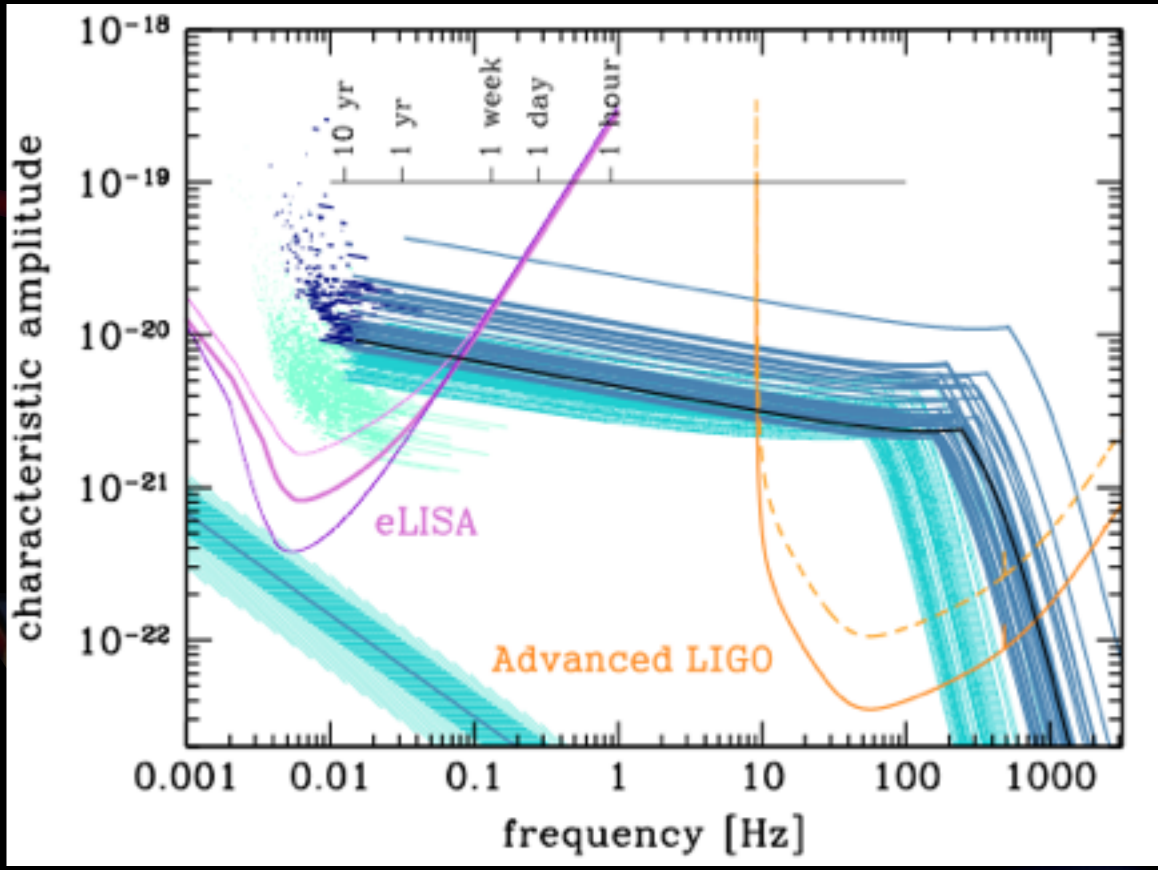
GW sources

- 6×10^7 galactic binaries
- 10-100/year SMBHBs
- 10-1000/year EMRIs
- large number of Stellar Origin BH binaries (LIGO/Virgo)
- Cosmological backgrounds
- Unknown sources



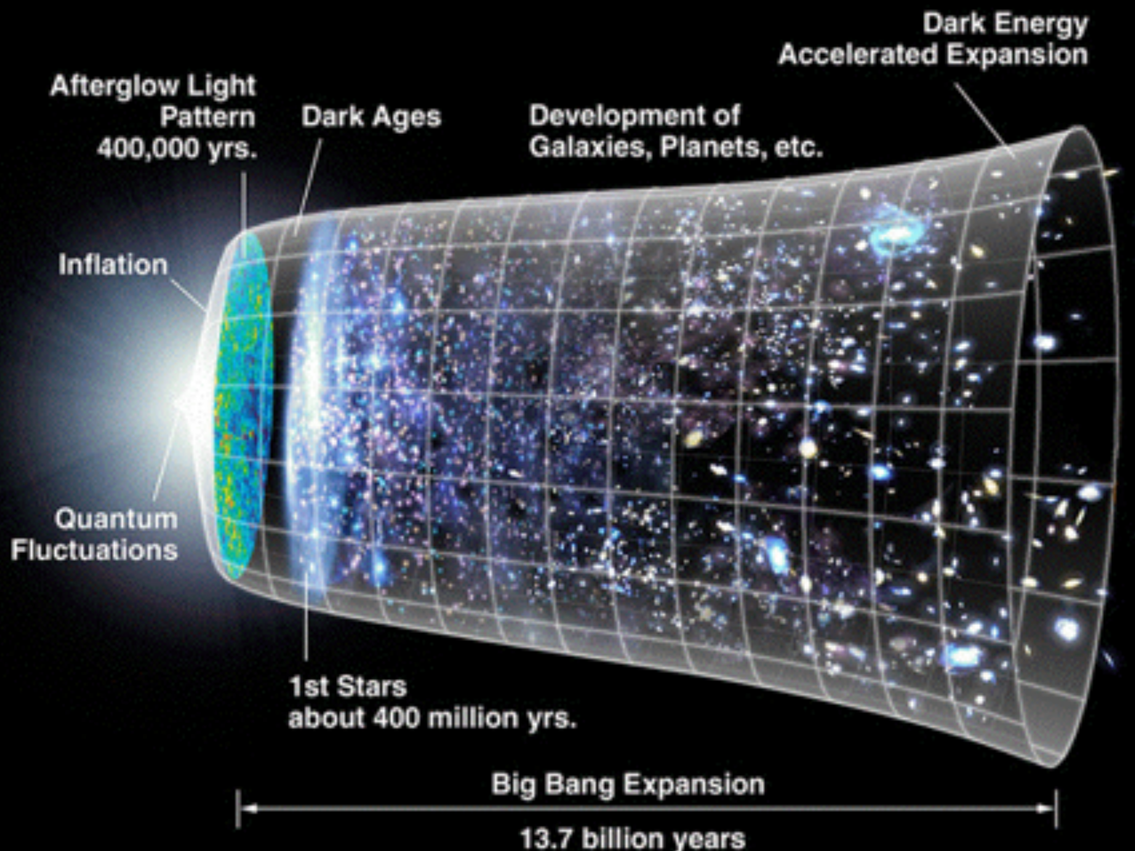
Others sources



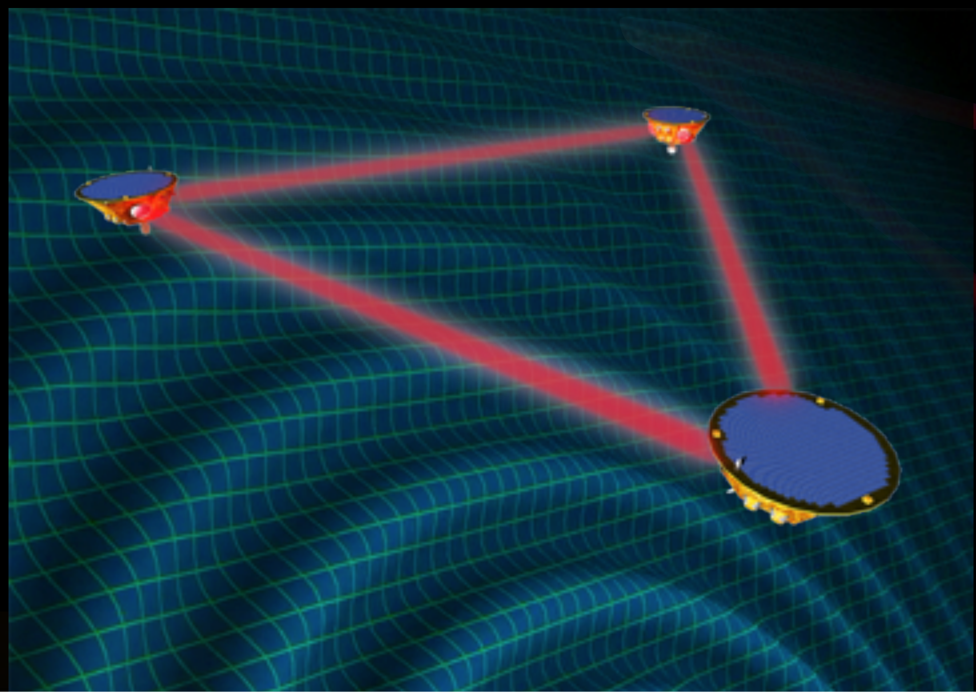


GW sources

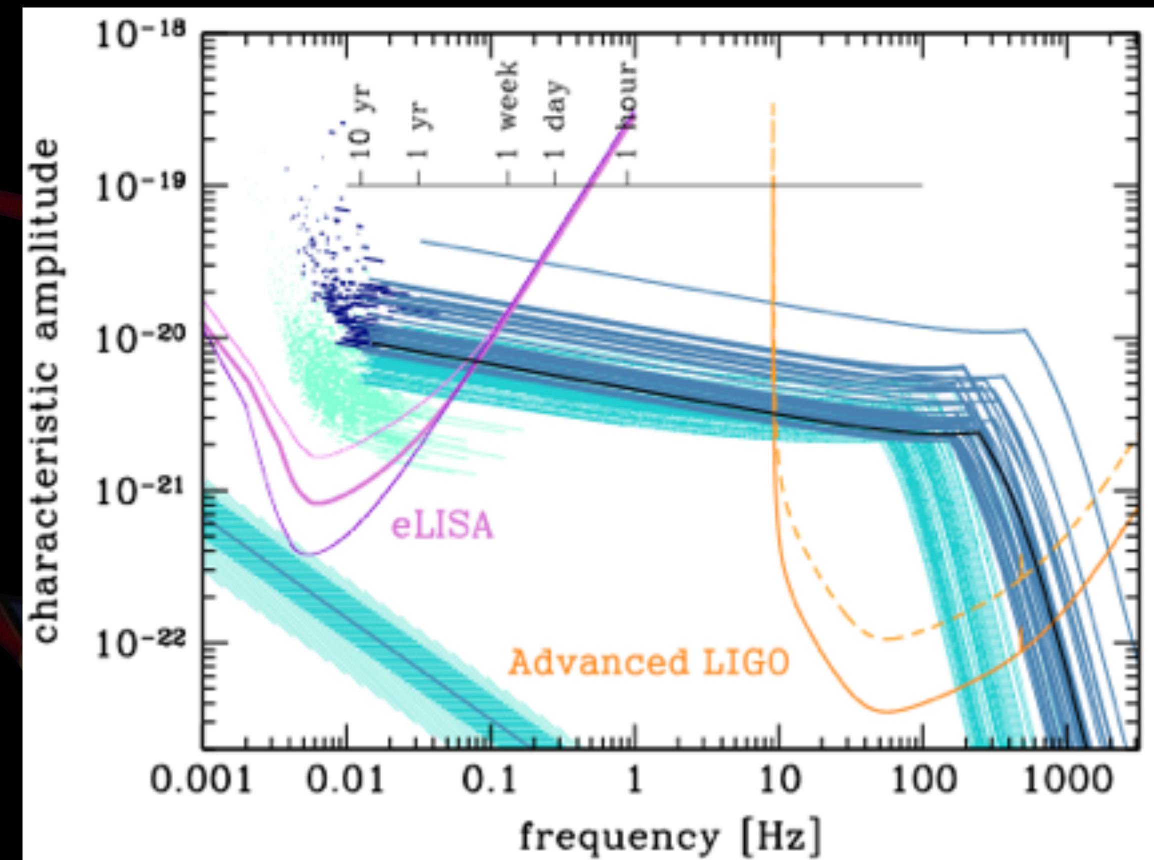
- 6×10^7 galactic binaries
- 10-100/year SMBHBs
- 10-1000/year EMRIs
- large number of Stellar Origin BH binaries (LIGO/Virgo)
- Cosmological backgrounds
- Unknown sources



Others sources

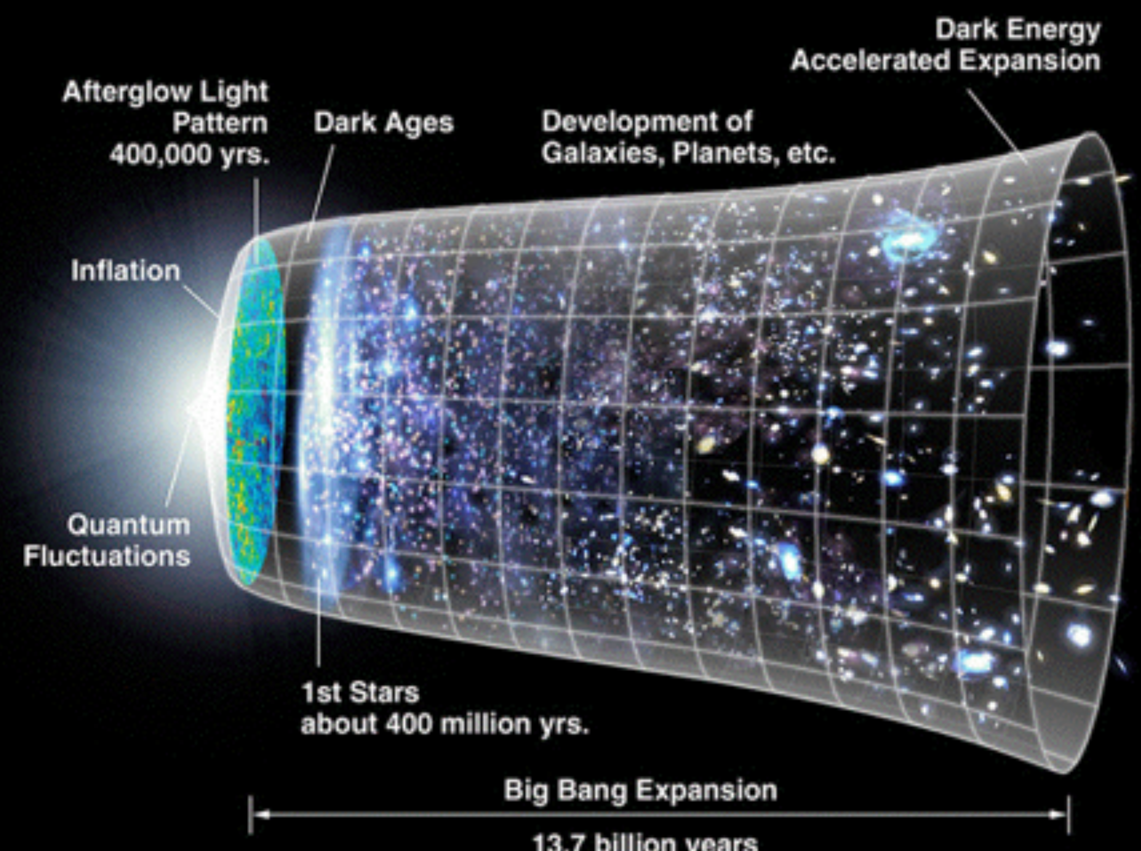


?

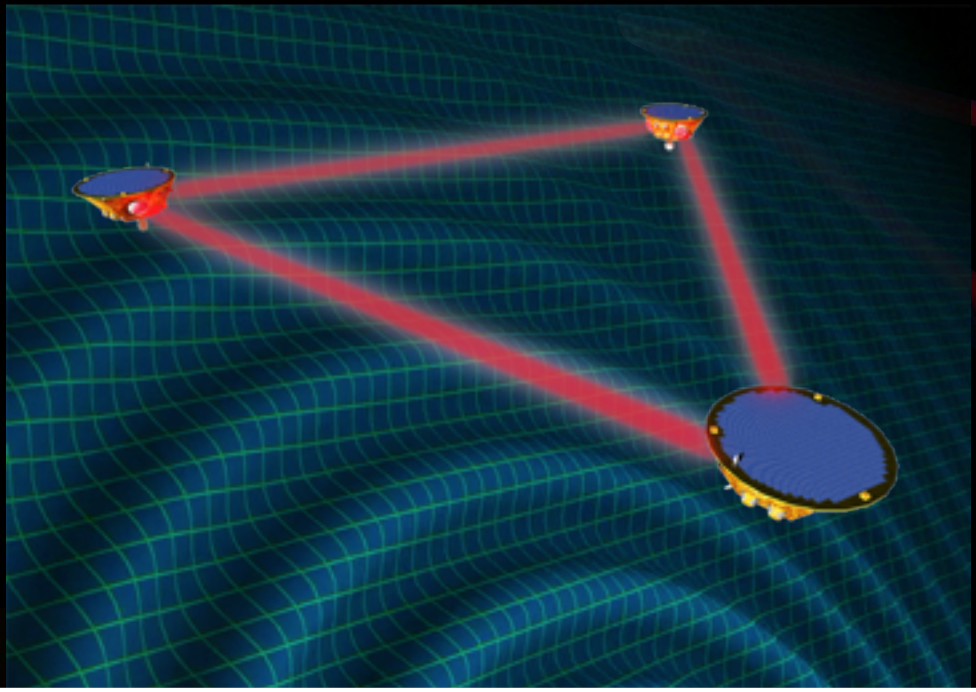


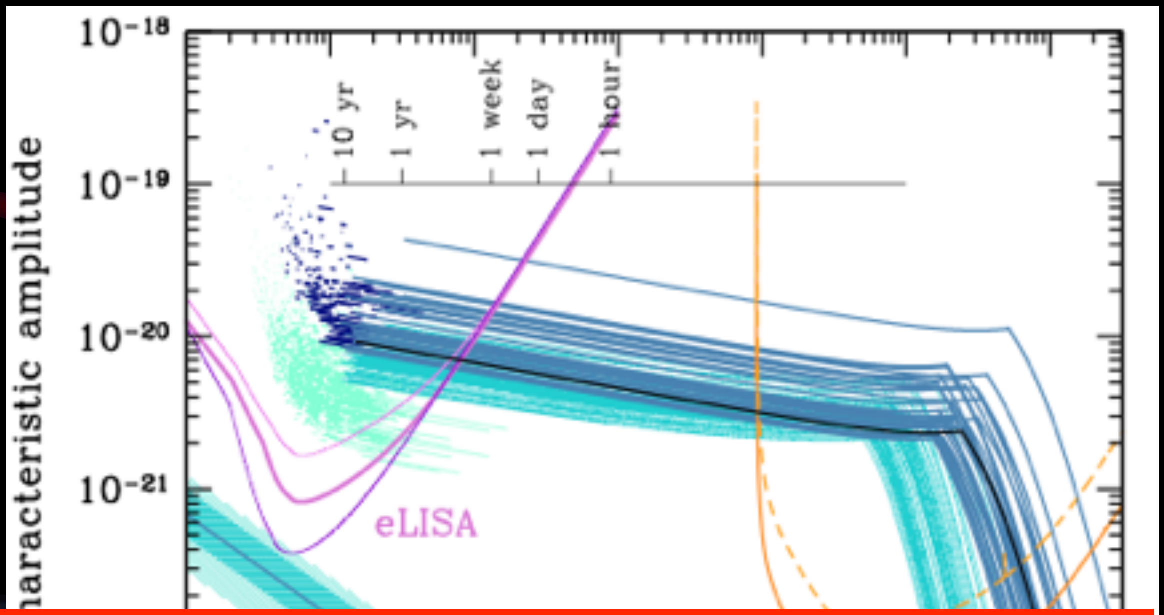
GW sources

- 6 x10[7] galactic binaries
- 10-100/year SMBHBs
- 10-1000/year EMRIs
- large number of Stellar Origin BH binaries (LIGO/Virgo)
- Cosmological backgrounds
- Unknown sources



Others sources





?

GW sources

- 6 x10^7 galactic binaries
- 10-100/year SMBHBs
- 10-1000/year EMRIs
- large number of Stellar Origin BH binaries (LIGO/Virgo)
- Cosmological backgrounds
- Unknown sources

Challenges of the LISA Data Analysis:

- Measuring more than 200 000 sources parameters from 2-3 time series,
- Overlapping sources,
- Taking into account the complexity of the instrument (noises subtraction, artefacts, ...)

13.7 billion years

LISA Data Processing

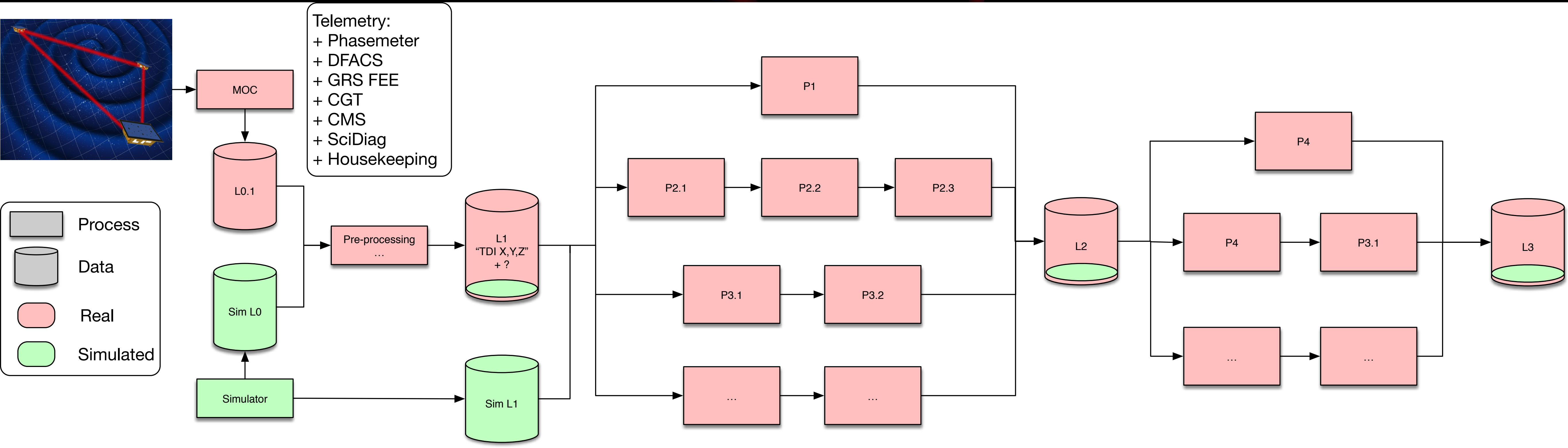
- Data volume to be stored:
  - Level L0: about 1 GB per day
  - Level L1: about 2-3 GB per day
  - Sub-product of the analysis: about few tens GB per day or more
  - Level L2 and L3: about few 10 GB per day

=> Storages and archives are not problematic

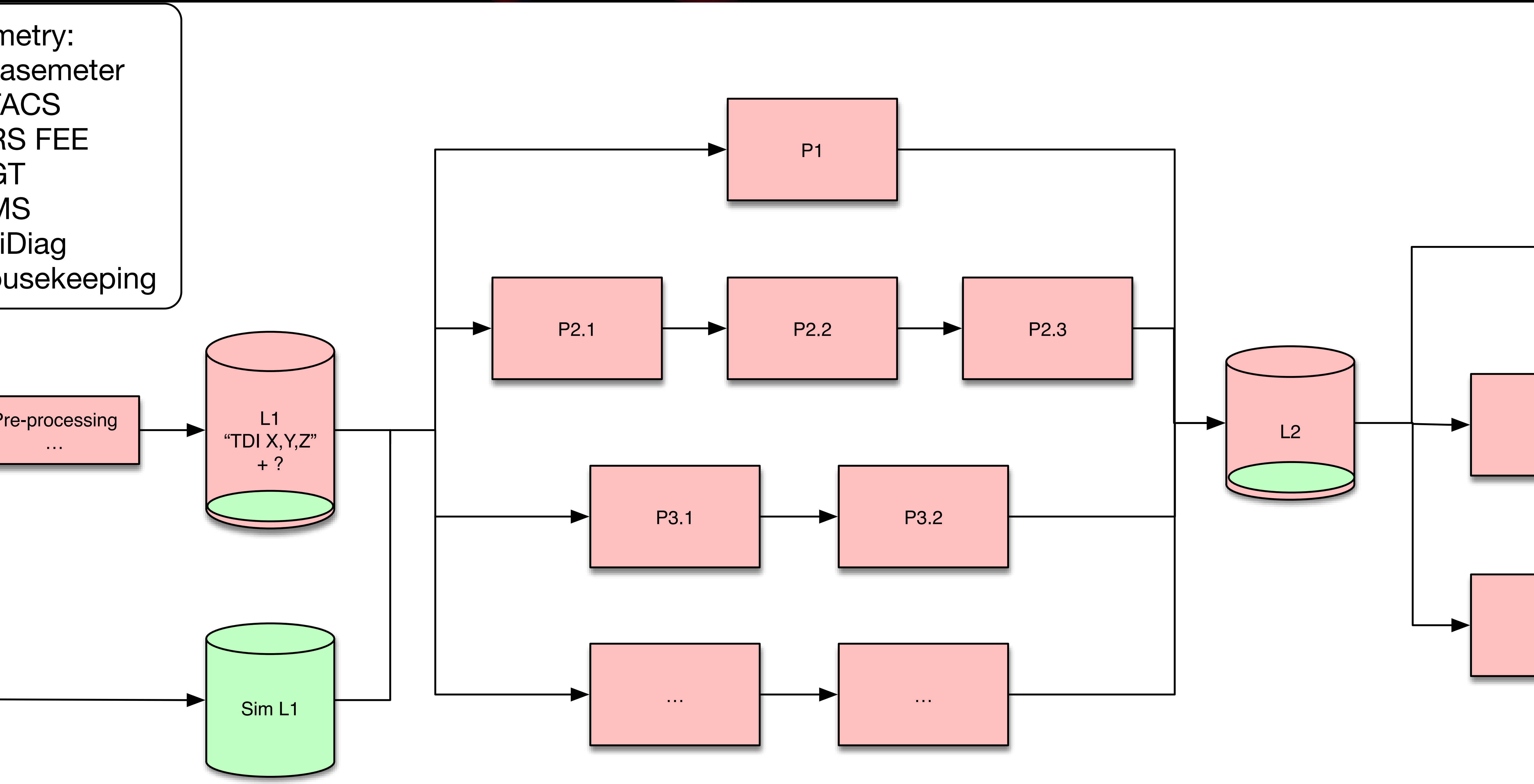
- But simulations will require some storage to be properly sized

- Complexity for the DDPC is mainly in data analysis because the goal is to extract the parameters for a maximum number of sources.

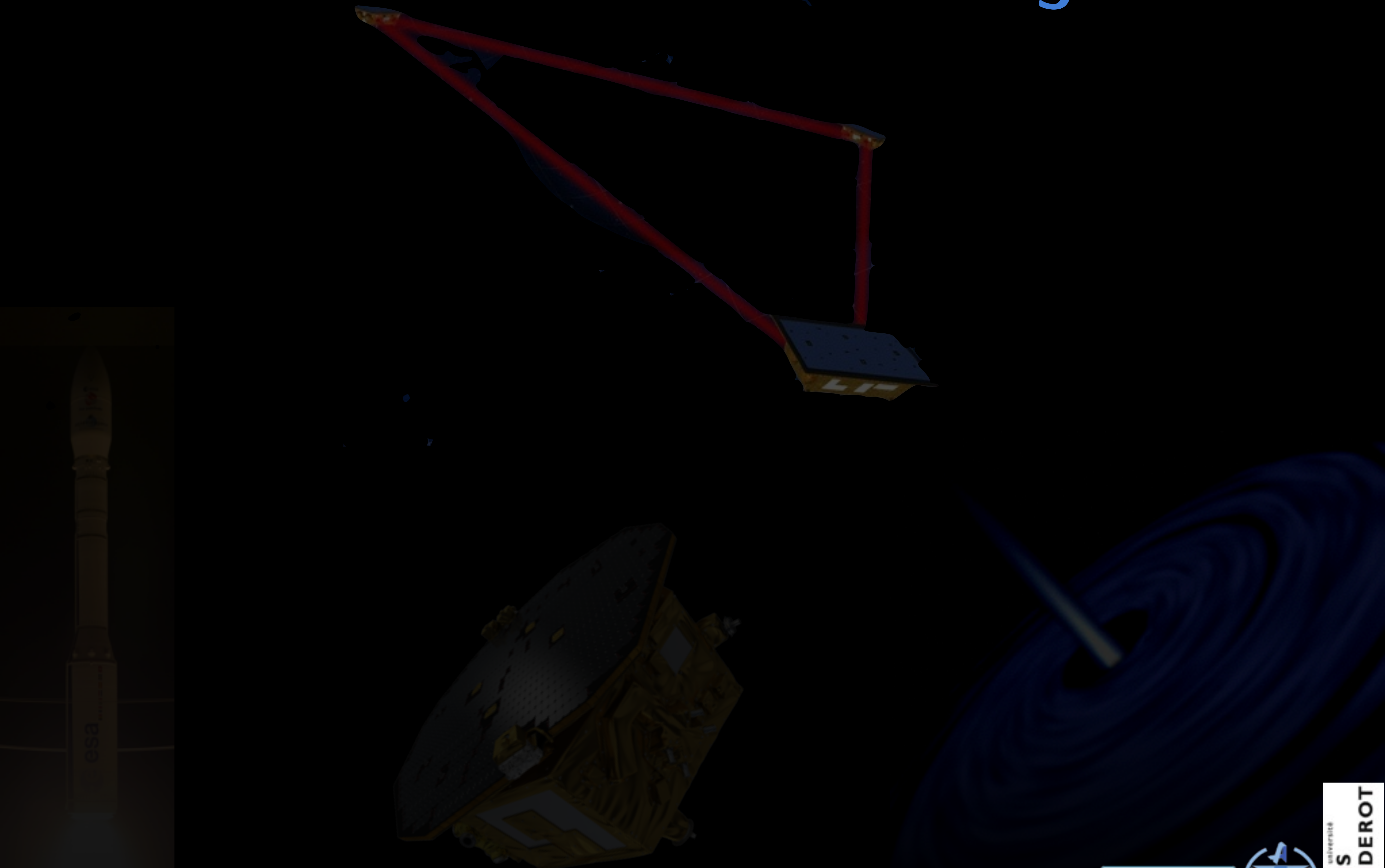
LISA Data Processing



LISA Data Processing



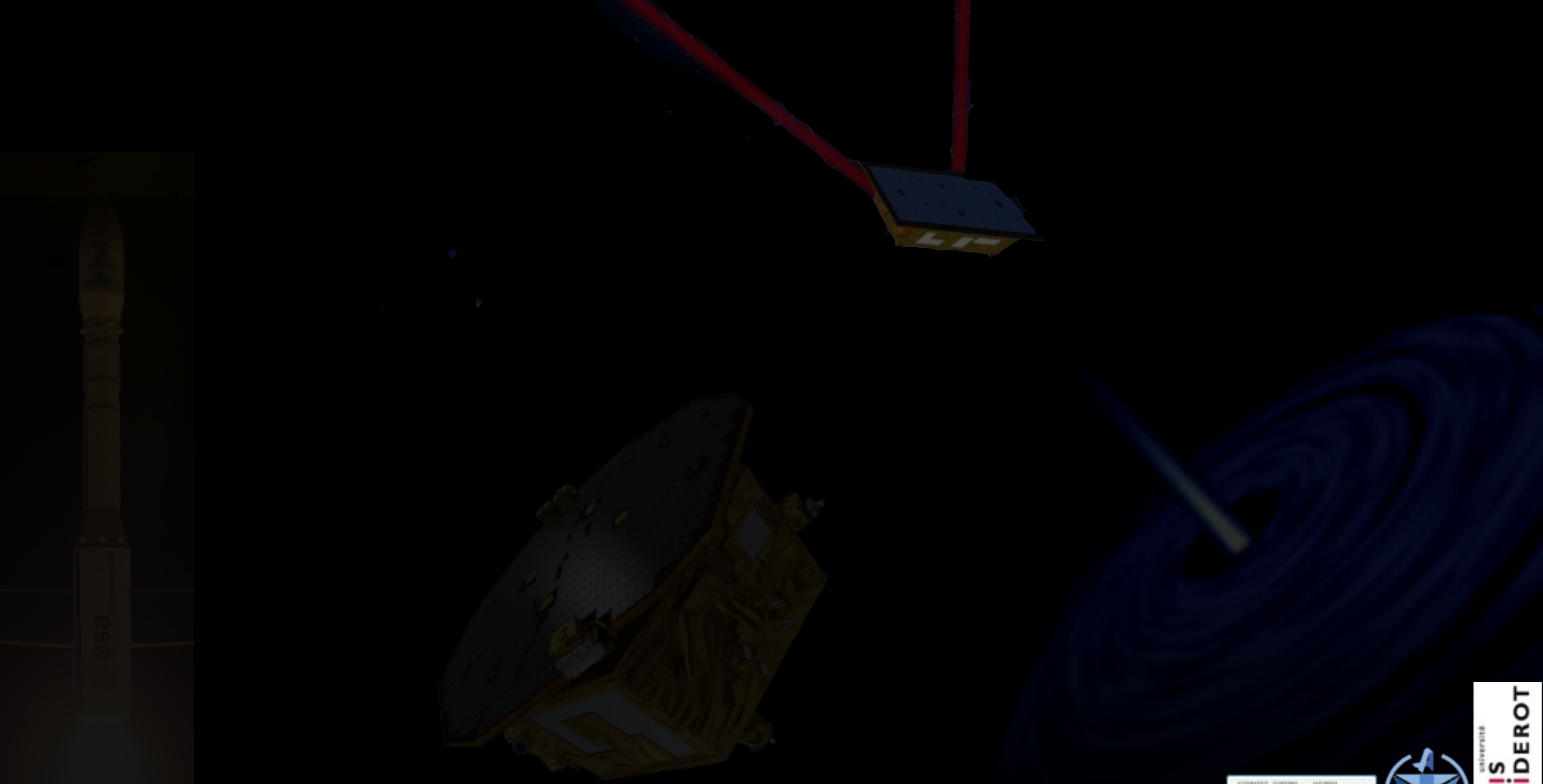
LISA Data Processing



LISA Data Processing

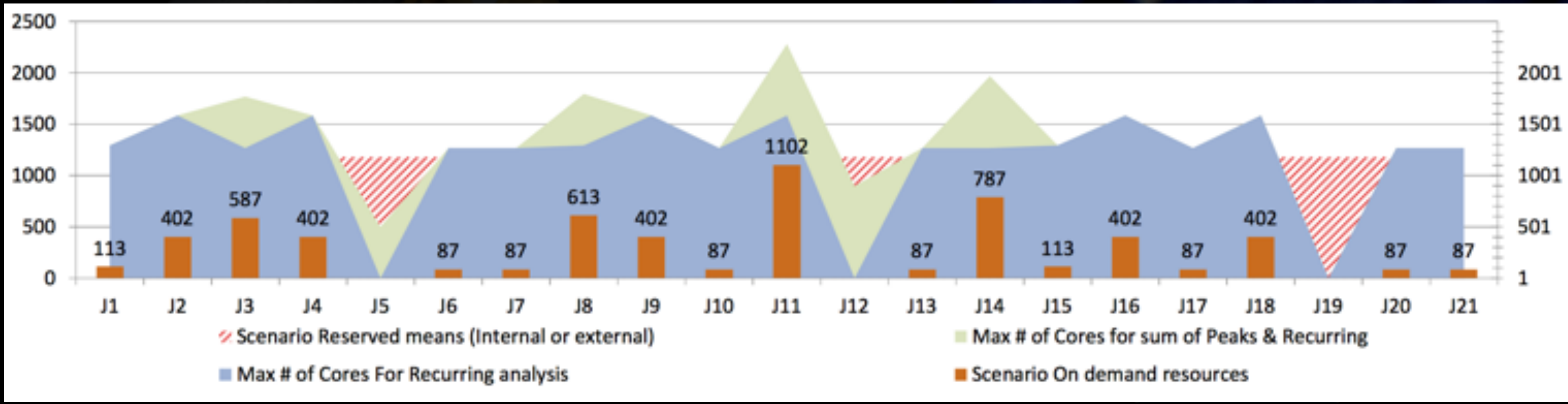
- **First** data and analysis of this kind + potential **unknown** sources

  => Keep **flexibility** + **continuous evolution**



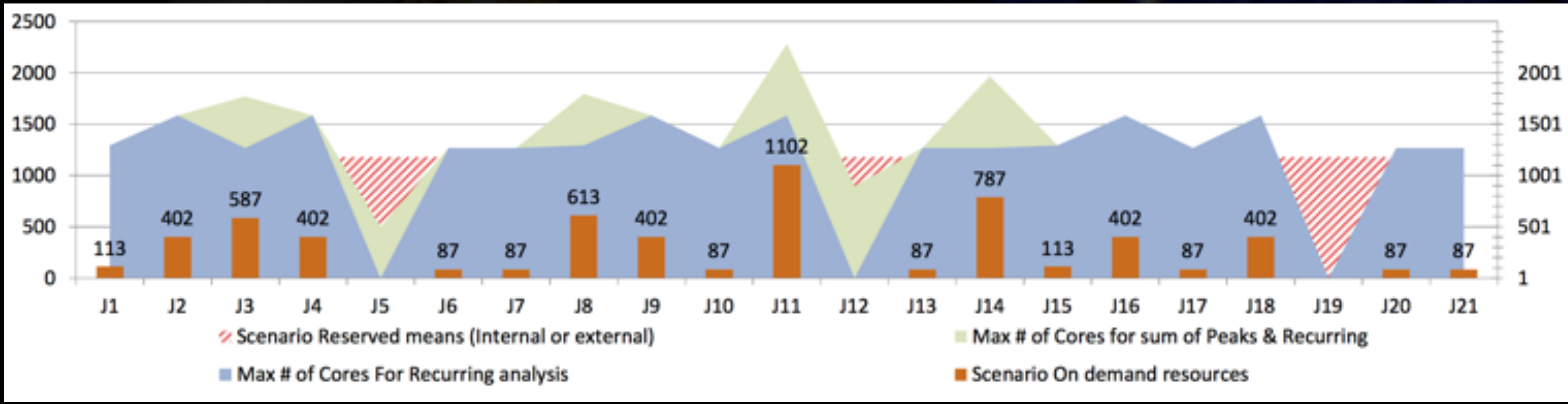
LISA Data Processing

- First data and analysis of this kind + potential unknown sources
  => Keep flexibility + continuous evolution
- Permanent sources + transient sources + continuous evolution of codes, i.e. full reprocessing phase
  => fluctuations of the computational charge: mixed infrastructure (standard clusters + on demand, i.e. Cloud)



LISA Data Processing

- First data and analysis of this kind + potential unknown sources
  => Keep flexibility + continuous evolution
- Permanent sources + transient sources + continuous evolution of codes, i.e. full reprocessing phase
  => fluctuations of the computational charge: mixed infrastructure (standard clusters + on demand, i.e. Cloud)
- Data analysis challenges: large number of mixed sources + no direct calibration of instrument
  => need to start the studies now!
  - Simulations
  - LISA Data Challenge

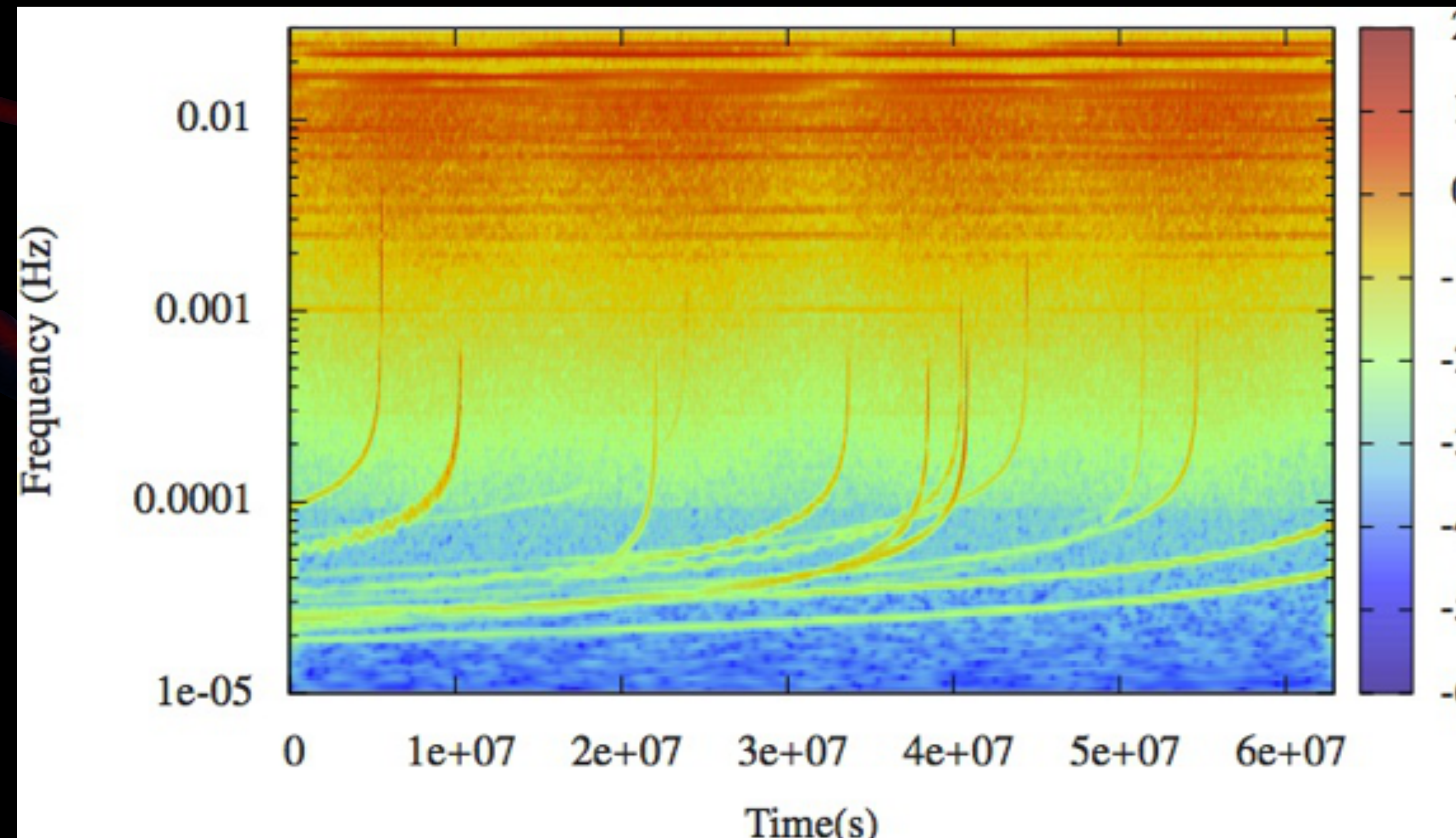


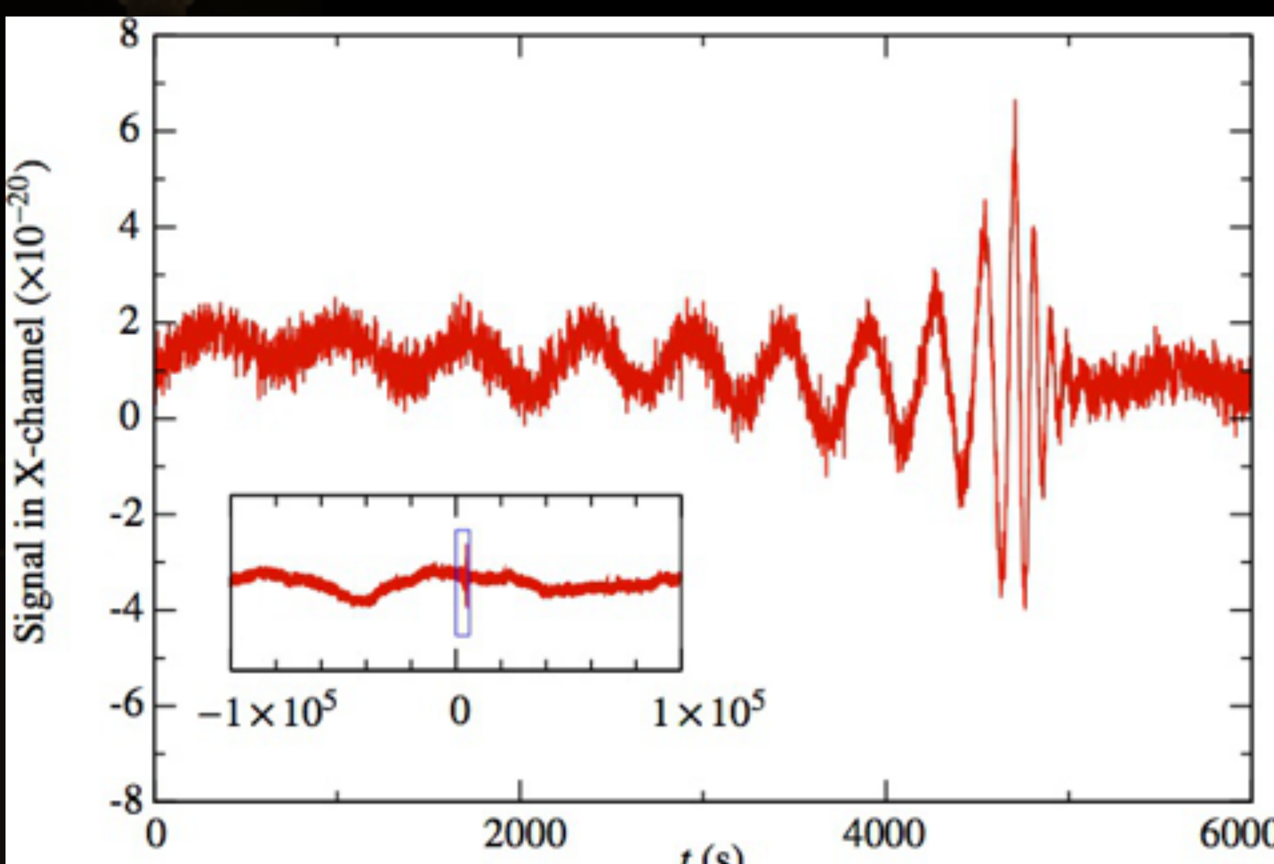
Data analysis & simulations

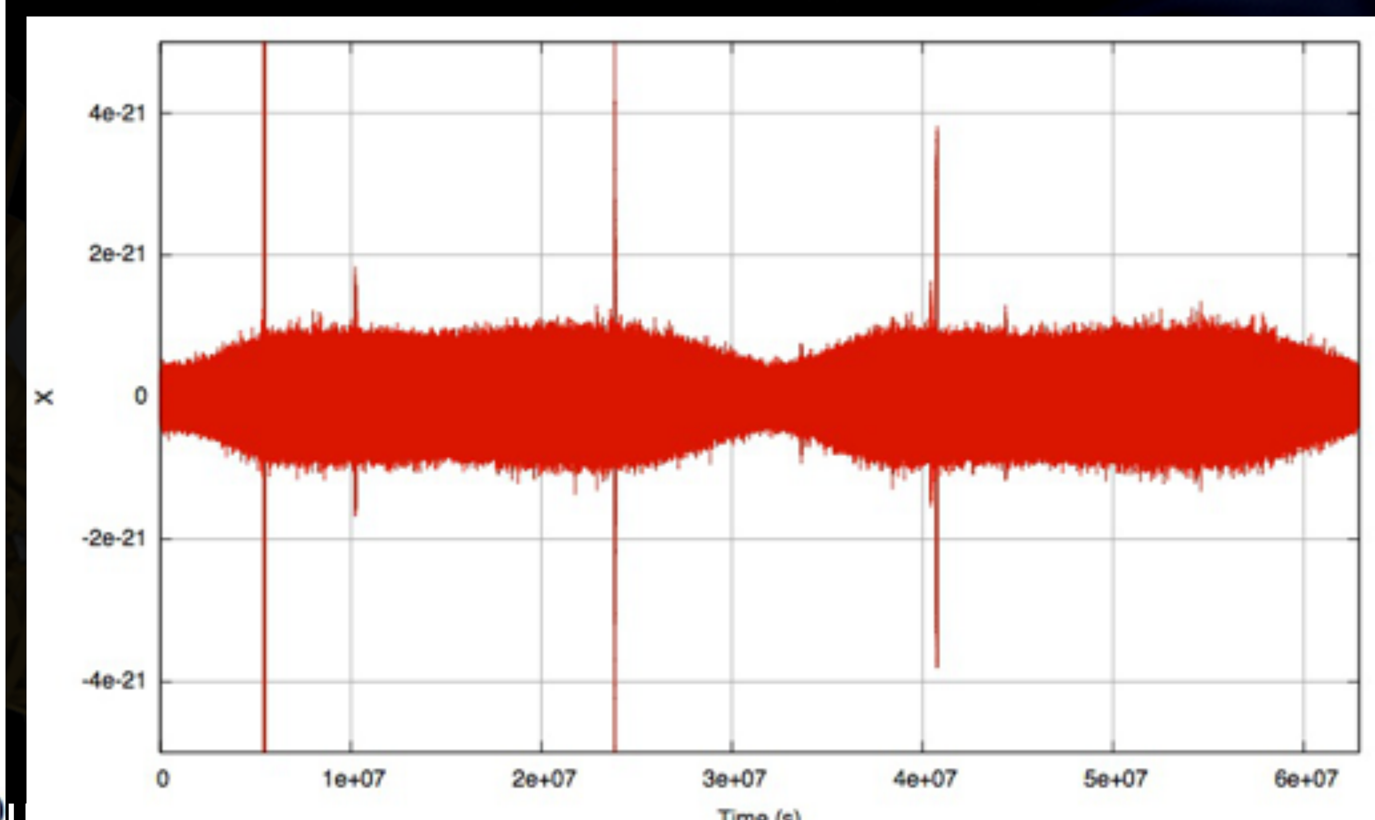
- **Simulations**:
  - Simulations at different scales: micro-sec to years in reasonable time
  - Coherently simulate control loops, integrate discretization/ interpolation, precisions, …
- **Data pre-processing**: clock, ranging, TDI
- **Data processing**: extracting science
  - For the matched filtering: optimisation of likelihood computation, variety of samplers, possibly large number of parameters, evolving number of parameters, …
  - Orchestration of multiple pipelines in parallel
  - Keep track of all produced data
  - Incremental data: new data to integrate every day
  - Fast pipeline for alerts, …

GWs in LISA data

- Example of simulated data (LISACode):
  - about 100 SMBHs,
  - Galactic binaries

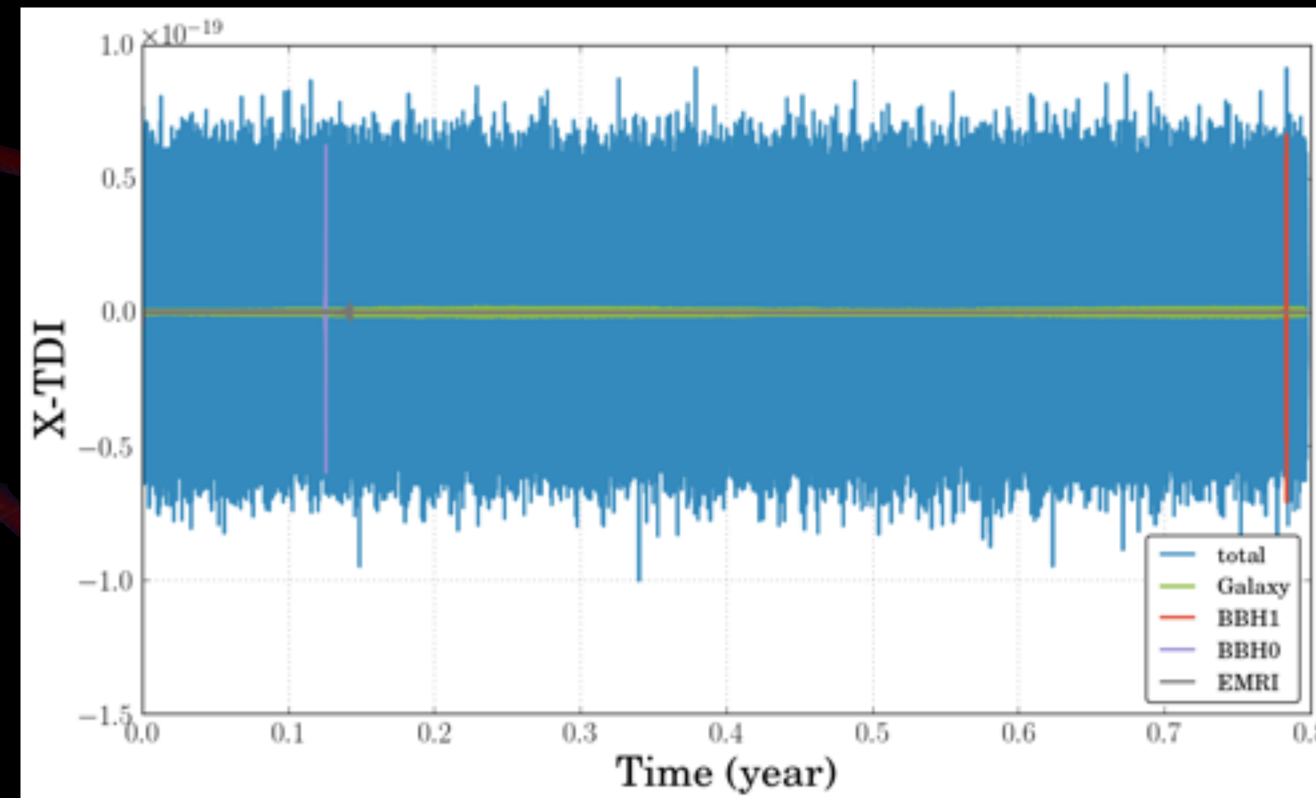


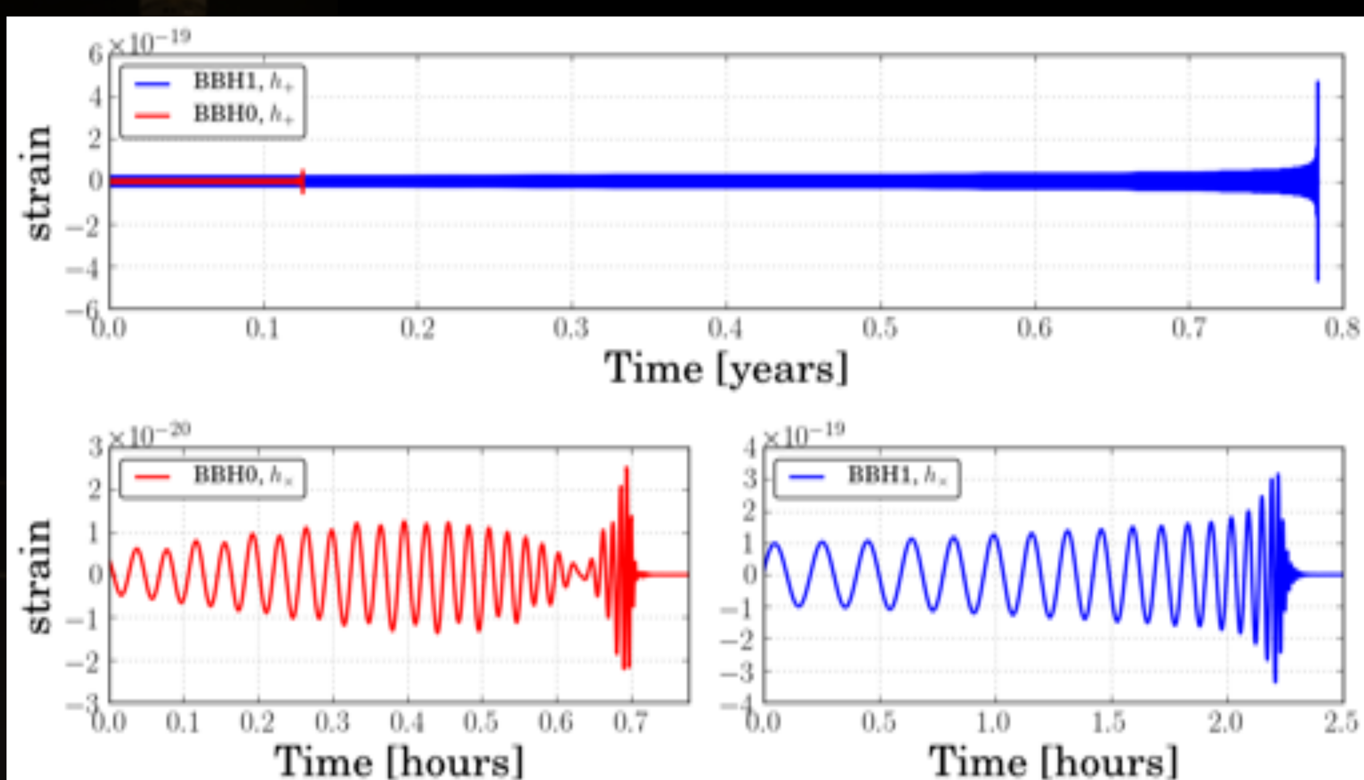


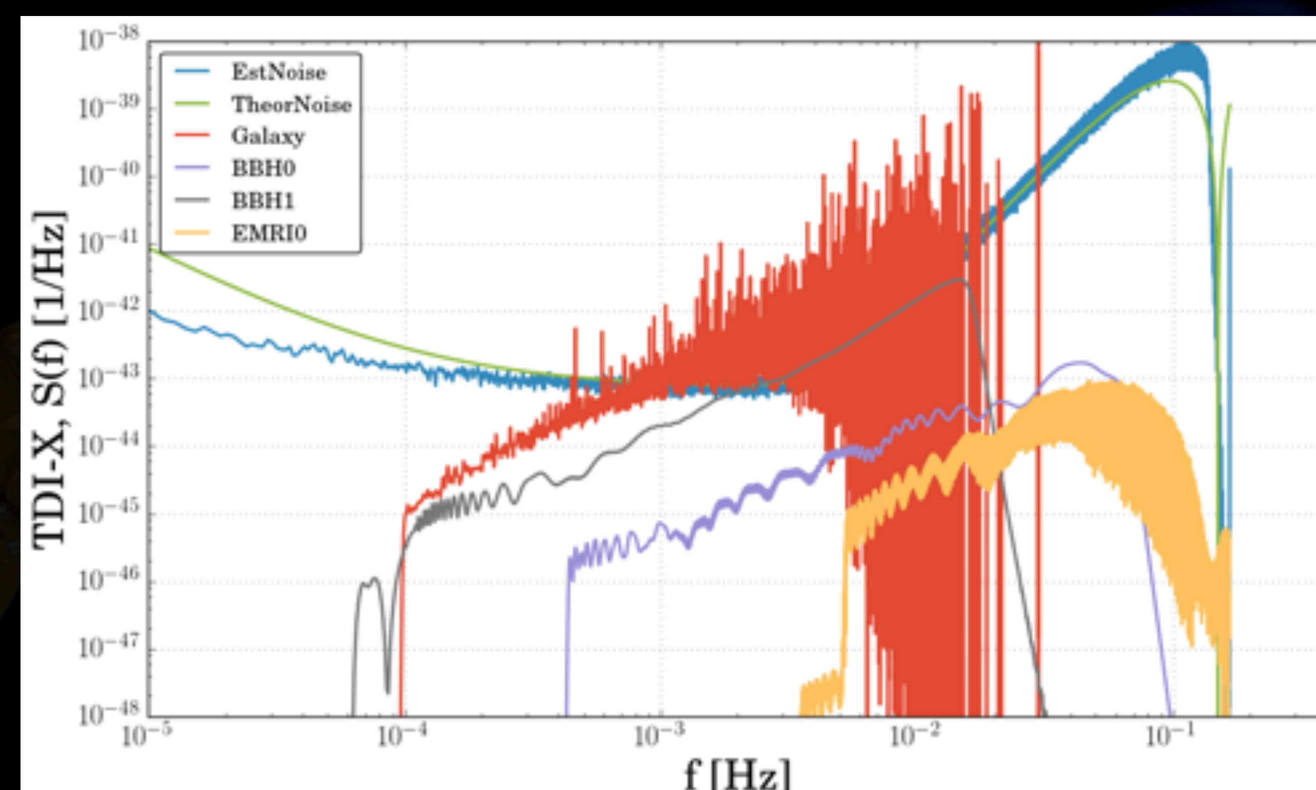


LISA Data Challenges

- Mock LDC: 2005→2011
- 2017: start of the LDC
  - Develop data analysis
  - Design the pipelines of the mission
- Example of the potential data for LDC1







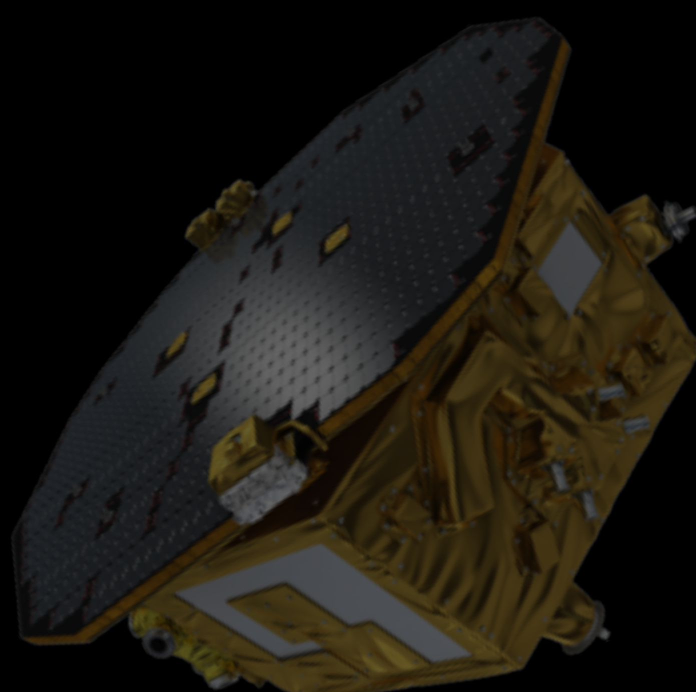
LISA / GPU

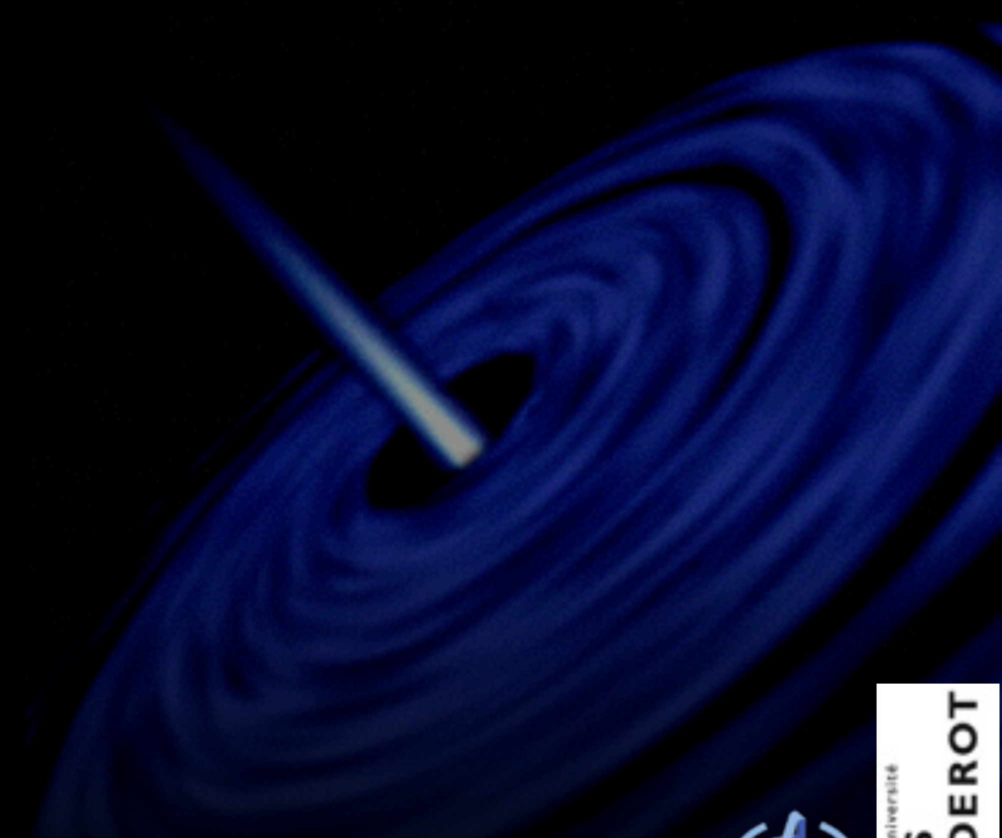
- Exploration phase:
  - Speed-up of the computation of template (GW waveform + instrument transfert function: core part of most of the search algorithm (i.e. Bayesian samplers ≃ "fit")
    - Successful first test using numba & cudapy
  - Mapping between GW sources parameters and likelihood
  - Analysis and classification of short transient using Machine Learning: instrument artefacts vs. GW "burst"
  - "Solving source separation problem for LISA data analysis with autoencoders" => Natalia Korsakova's talk
  - …
  - Any idea is welcome !

Conclusion

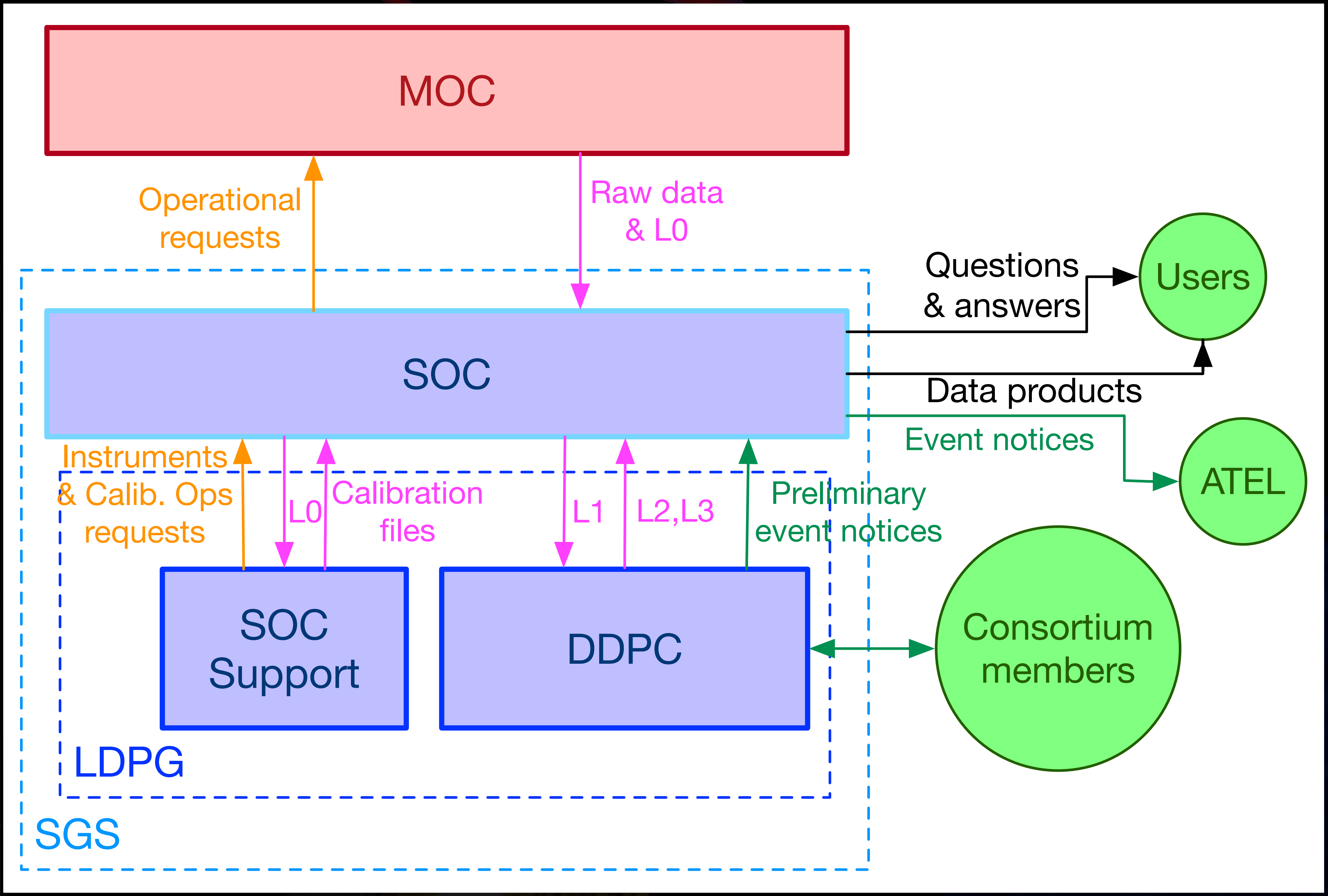
- LISA started: phase A for launch early 2030s
- First mission of this kind => some uncertainties (number of sources, data quality, unknown sources ...) => flexibility + continuous evolution + computation load fluctuations
- Distributed Ground Segment: MOC + SOC + Distributed DPC
  - SOC: L0 → L1: calibration, pre-processing reducing noises
  - DDPC: L1 → L2,L3 : extract GW sources from TDI data (L1) to produce catalogs and science products (L2 & L3)
- Challenge of the LISA data analysis: measure large number of parameters of overlapping sources in 3 times series !
  - Very active field, multiple pipelines, heavy computation
  - GPUs: exploratory phase

Thank you





LISA Ground Segment



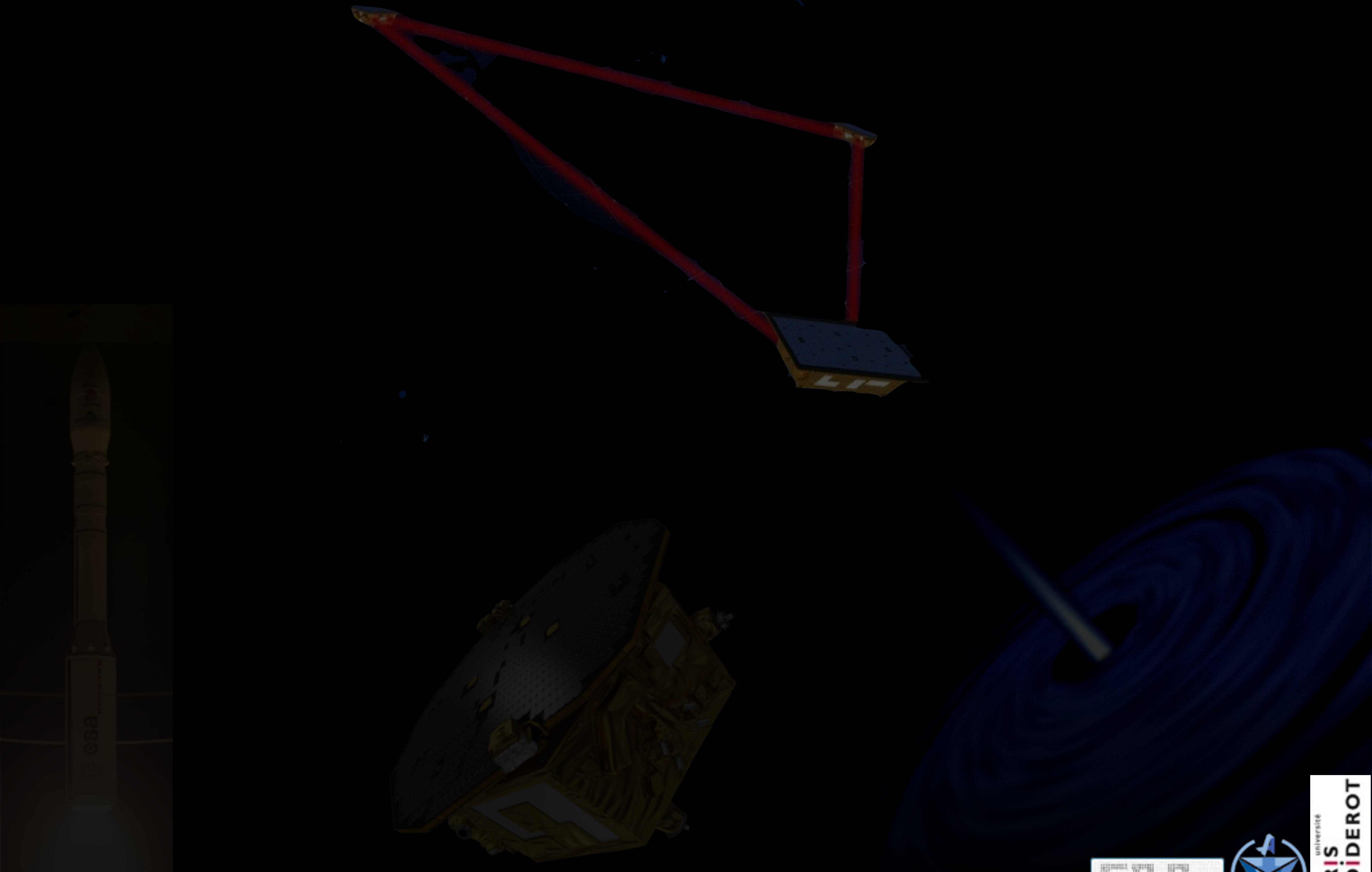
From L0 to L1

- Input (L0): "raw" data from the MOC
- Output (L1): TDI + all data "cleaned"
- Responsibility: SOC (ESA)
- With Consortium support => SOC Support group
- Activities / Challenges:
  - Processing ——————>
    - Calibration
    - Clock synchronisation
    - Ranging (estimation of delays)
    - TDI
  - Hardware monitoring
  - Quick-look of instrument data
  - …

From L1 to L3

- Inputs: TDI + all data “cleaned”
- Outputs: final science products (catalogs, …)
- Responsibility: Consortium => DDPC
- Activities:
  - Data analysis pipelines and simulation:
    - Prepare, Implement, Operate;
    - Support (LSG, SimWG, LDC) design and prototyping;
  - Define, coordinate and implement software framework and management structure for data and products
  - Coordinate and operate the DCCs
  - Define, implement and maintain dev. and op. environment

Current vision of the DDPC



Current vision of the DDPC

- DDPC: **unique entity** responsable for the **data processing** (driving, integration of software blocks, ...)

Current vision of the DDPC

- DDPC: unique entity responsable for the data processing (driving, integration of software blocks, ...)
- DDPC in charge of delivering L2 & L3 products + what's necessary to reproduce/refine the analysis (i.e. input data + software + its running environment + some CPU to run it).

Current vision of the DDPC

- DDPC: unique entity responsable for the data processing (driving, integration of software blocks, ...)
- DDPC in charge of delivering L2 & L3 products + what's necessary to reproduce/refine the analysis (i.e. input data + software + its running environment + some CPU to run it).
- Data Computing Centres (DCC): hardware, computer rooms (computing and storage) taking part to the data processing activities.

Current vision of the DDPC

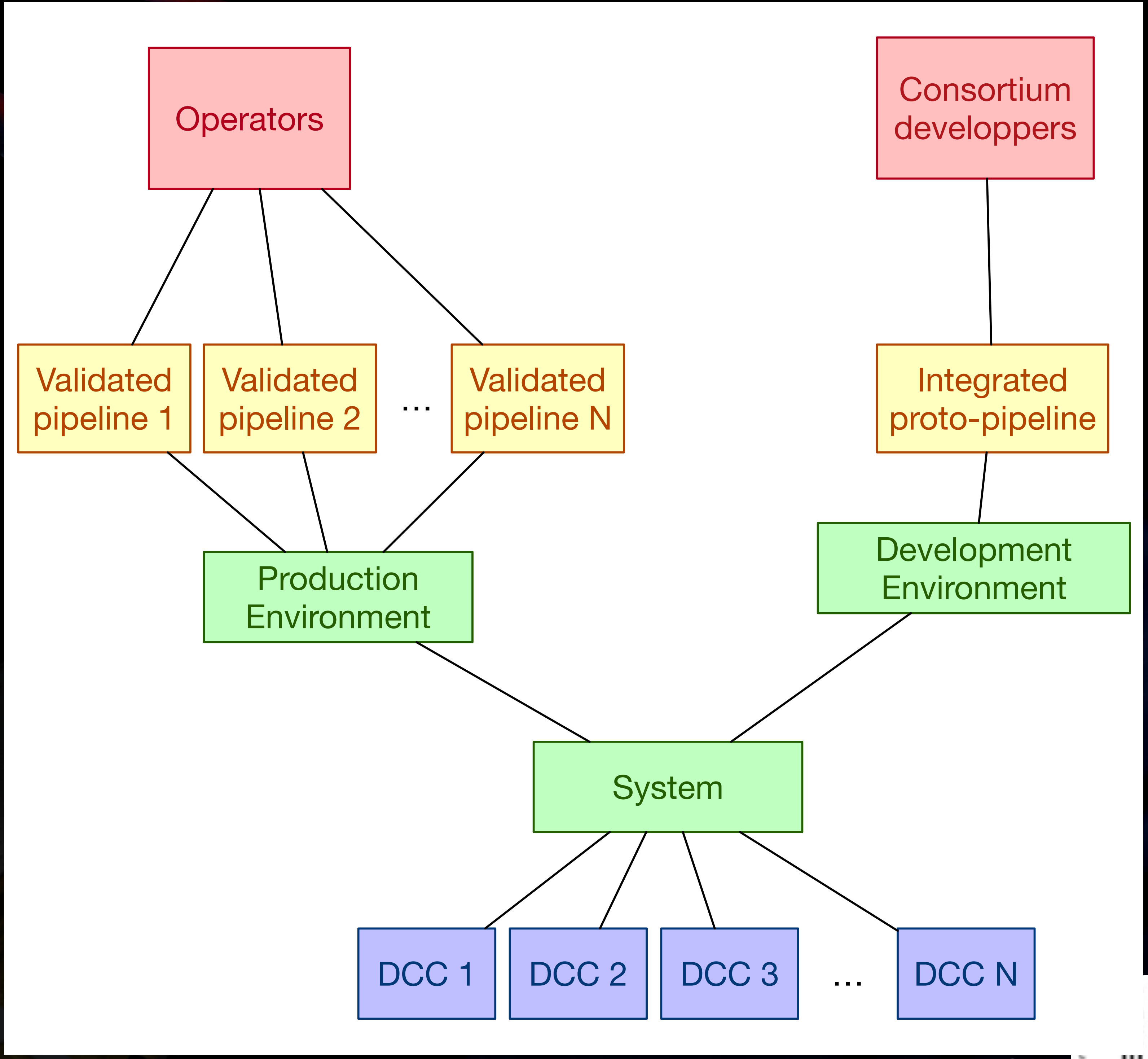
- DDPC: unique entity responsable for the data processing (driving, integration of software blocks, ...)
- DDPC in charge of delivering L2 & L3 products + what's necessary to reproduce/refine the analysis (i.e. input data + software + its running environment + some CPU to run it).
- Data Computing Centres (DCC): hardware, computer rooms (computing and storage) taking part to the data processing activities.
- The DDPC software « suite » can run on "any" DCC.
  - Software: codes (DA & Simu.) + services (LDAP, wiki, database) + OS.

Current vision of the DDPC

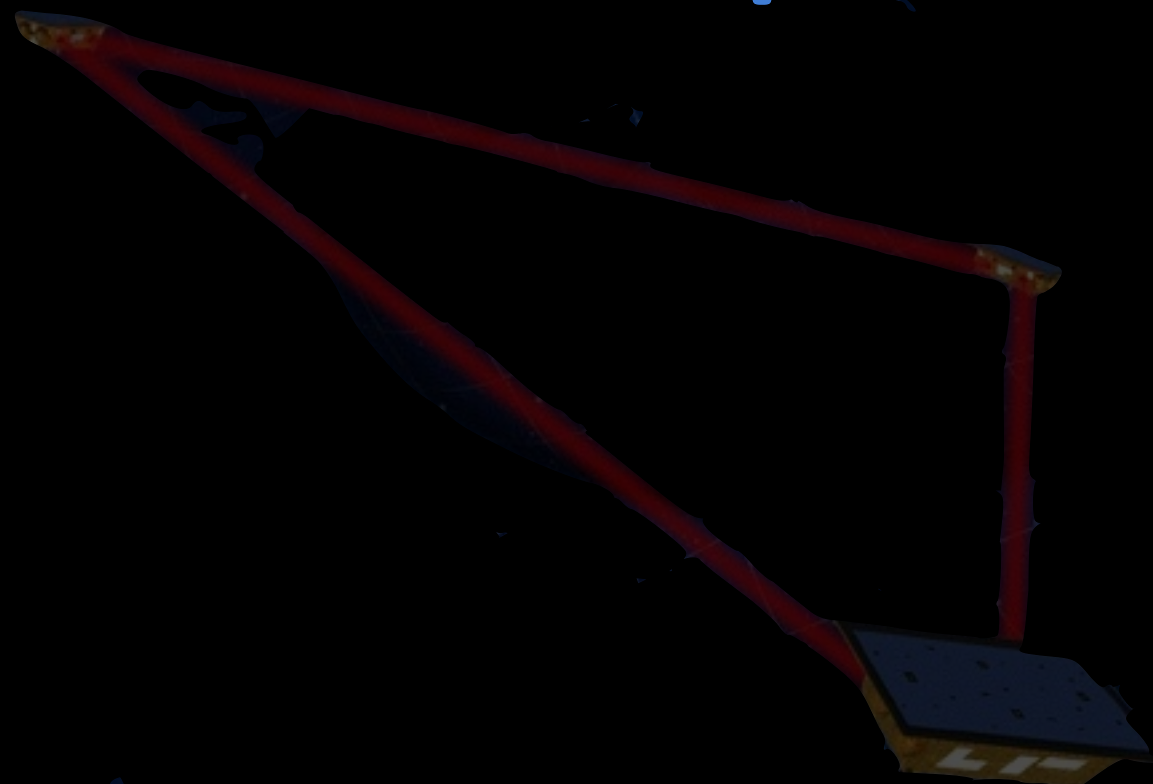
- DDPC: unique entity responsable for the data processing (driving, integration of software blocks, ...)
- DDPC in charge of delivering L2 & L3 products + what's necessary to reproduce/refine the analysis (i.e. input data + software + its running environment + some CPU to run it).
- Data Computing Centres (DCC): hardware, computer rooms (computing and storage) taking part to the data processing activities.
- The DDPC software « suite » can run on “any” DCC.
  - Software: codes (DA & Simu.) + services (LDAP, wiki, database) + OS.
- First solutions:
  - Separation of hardware and software: light virtualization, ...
  - Collaborative development: continuous integration, ...
  - Fluctuations of computing load: hybrids cluster/cloud

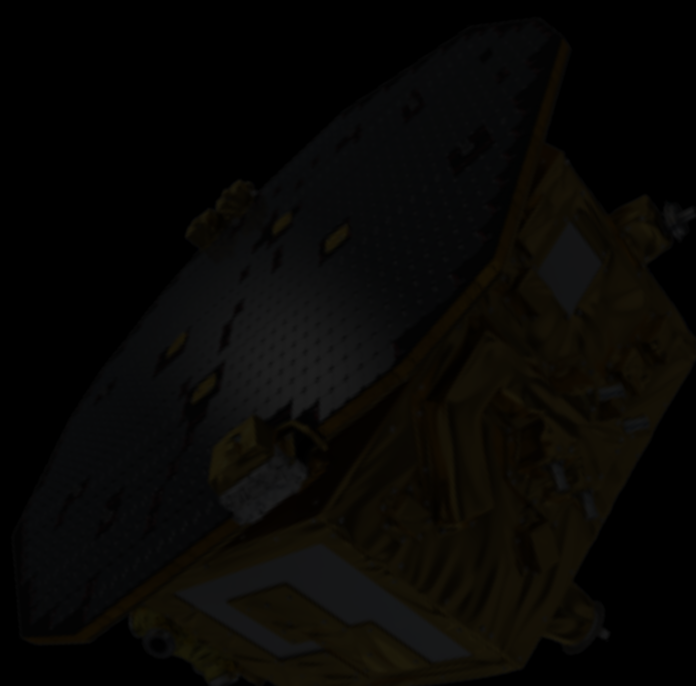
Common system: dev./prod.

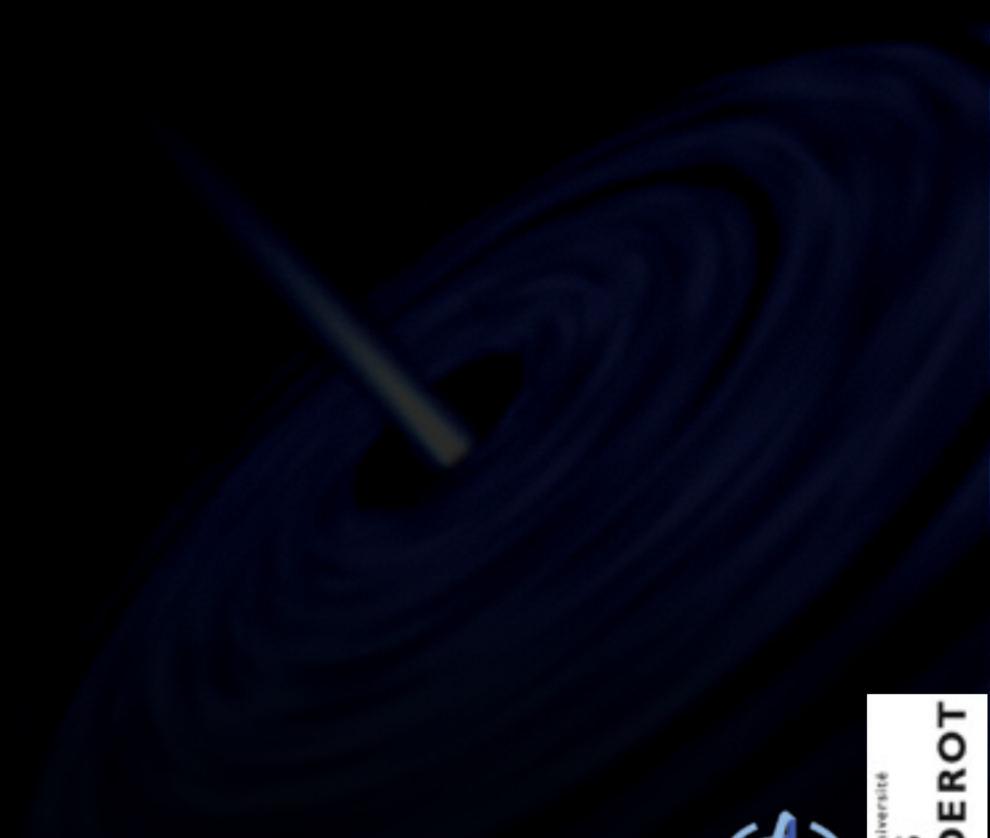
- First ideas based on a common system:
  - short cycle between dev. & prod.
  - distributed hardware on DCCs (Data Computing Centres)
  - cloud compatibility



Tasks for phase A







Tasks for phase A

1. Definition and detailed design of the Ground Segment part delivered by the Consortium => start: QUESTIONNAIRE

Tasks for phase A

1. Definition and detailed design of the Ground Segment part delivered by the Consortium => start: QUESTIONNAIRE

2. Support/Contribute to the definition of LISA Ground Segment with ESA => contribution to documents

Tasks for phase A

1. Definition and detailed design of the Ground Segment part delivered by the Consortium => start: QUESTIONNAIRE

2. Support/Contribute to the definition of LISA Ground Segment with ESA => contribution to documents

3. Prototyping (benchmarking)

Tasks for phase A

1. Definition and detailed design of the Ground Segment part delivered by the Consortium => start: QUESTIONNAIRE
2. Support/Contribute to the definition of LISA Ground Segment with ESA => contribution to documents
3. Prototyping (benchmarking)
4. Support LISA activities (Consortium & ESA): LDC, SimWG, …

Tasks for phase A

1. Definition and detailed design of the Ground Segment part delivered by the Consortium => start: QUESTIONNAIRE

2. Support/Contribute to the definition of LISA Ground Segment with ESA => contribution to documents

3. Prototyping (benchmarking)

4. Support LISA activities (Consortium & ESA): LDC, SimWG, …

5. Provide services to the Consortium: Doc. Management, repositories, wiki, computing facilities

Proto-DPC: basics

Proto-DPC: basics

- Development environment: in production
  - Collaborative work, reproducibility of a rapidly evolving & composite DA pipeline; Keep control of performance, precision, readability, etc
  - Use existing standard tools (version control, Continuous Integration, Docker)

Proto-DPC: basics

- Development environment: in production
  - Collaborative work, reproducibility of a rapidly evolving & composite DA pipeline; Keep control of performance, precision, readability, etc
  - Use existing standard tools (version control, Continuous Integration, Docker)
- Data basis & data model: in R&D
  - Data sharing, a lot of information (search engine, DB request, tree view);
    - Context: Not very big data volume for data itself but large number of sub-products, simulations, … => LDC, simulations, LPF data

Proto-DPC: basics

- Development environment: in production
  - Collaborative work, reproducibility of a rapidly evolving & composite DA pipeline; Keep control of performance, precision, readability, etc
  - Use existing standard tools (version control, Continuous Integration, Docker)
- Data basis & data model: in R&D
  - Data sharing, a lot of information (search engine, DB request, tree view);
    - Context: Not very big data volume for data itself but large number of sub-products, simulations, ... => LDC, simulations, LPF data
- Execution environment: in R&D (singularity, ...)

Support LISA Consortium today

- **Simulation**:
  - LISACode and LISANode: git with continuous integration, docker image, singularity, documentation, …
    => realistic data used for ex for performance, pre-processing, …
- **Exchange**: LDC database, Virtual Machine on demand
- IT: Repositories, Document Management System, wikis
- Coming soon:
  - Jupyter hub available soon: share scripts
  - Singularity hub: share image containing all LDC tools
  - Computing facilities (prototyping DCCs)
  - Integration of LDC DA methods submitted with responses