Traitement des données, simulations et DPC

Antoine Petiteau
(APC – Université Paris-Diderot)

Réunion d'informations sur la mission LISA
CEA - Saclay
18 Mai 2017

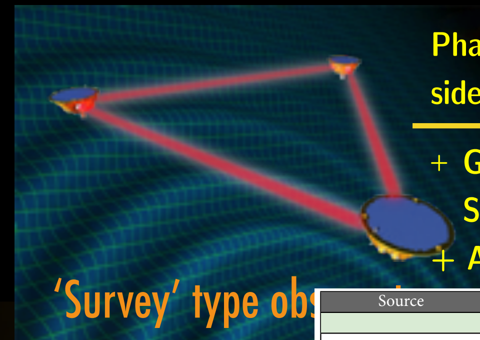
Overview

- LISA data analysis and data processing
  - Overview and data level
  - Calibration, TDI, etc
  - Sources and waveforms
  - Extracting sources and their parameters
- LISA Simulation
- Data Processing Center
  - Tasks
  - Organisation

Overview

- LISA data analysis and data processing
  - Overview and data level
  - Calibration, TDI, etc
  - Sources and waveforms
  - Extracting sources and their parameters
- LISA Simulation
- Data Processing Center
  - Tasks
  - Organisation

LISA data



Phasemeters (carrier, sidebands, distance)

+ Gravitational Reference Sensor

+ Auxiliary channels

Calibrations corrections

'Survey' type obs

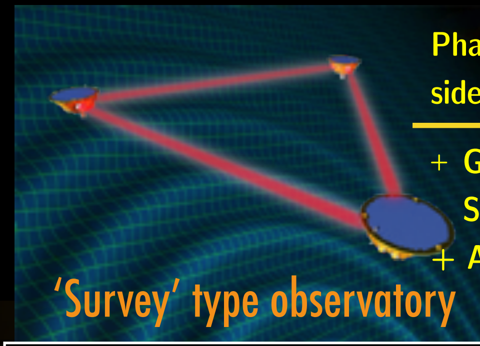
Source	Class	Measurement	Count	Sampling Rate [Hz]	Bits / channel	Rate [bits/s]
Payload						
Phasemeter	IFO Longitudinal	Science IFO	2	3.3	32	213.3
		Test Mass IFO	2	3.3	32	213.3
		Reference IFO	2	3.3	32	213.3
		Clock Sidebands	2	3.3	32	213.3
	IFO Angular	S/C θ,η	4	3.3	32	426.6
		TM θ,η	4	3.3	32	426.6
	Anciliary	Time Semaphores	2	3.3	96	639.9
	Optical Monitoring	PAAM Longitudinal	2	3.3	32	213.3
		PAAM Angular	4	3.3	32	426.6
		Optical Truss	6	3.3	32	639.9
GRS FEE	GRS Cap. Sensing	TM x,y,z	6	3.3	24	480.0
		TM θ,η,ϕ	6	3.3	24	480.0
Payload Computer	DFACS	TM applied torques	6	3.3	24	480.0
		TM applied forces	6	3.3	24	480.0
		S/C applied torques	3	3.3	24	240.0
		S/C applied forces	3	3.3	24	240.0
	Payload HK	e.g. Temperature, Power Monitors *etc.*				2613
	Total Payload					8639
Platform						
	Housekeeping (based on LPF)					1189
	Total Platform					1189
Totals						
	Raw rate per S/C					9828
	Paketisation overhead [10%]					983
	Packaged rate per S/C					10811
	Packaged rate for Constellation					32433

Gravitational w
emitting betwee
and 100

LISA data level

- **Level L0 data**: raw science telemetry and housekeeping data.
- **Level L1 data**: TDI variables, all calibrated science data streams and auxiliary data.
- **Level L2**: intermediate waveform products such as partially regressed observable series (i.e., dataset obtained by progressively deeper subtraction of identified signals).
- **Level L3**: catalogs of identified sources, with faithful representations of posterior parameter distributions.

LISA data



'Survey' type observatory

Phasemeters (carrier, sidebands, distance)

+ Gravitational Refer Sensor

+ Auxiliary channels

L0

Calibrations corrections

Resynchronisation (clock)

Time-Delay Interferometry reduction of laser noise

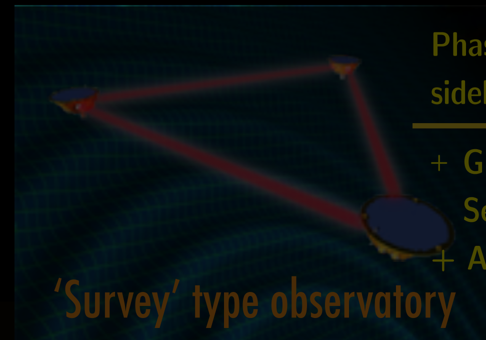
L1 — 2 data channels TDI non-correlated

L2 — Data Analysis of GWs

L3 — Catalogs of GWs sources with their waveform

Gravitational wave sources emitting between 0.02mHz and 100 mHz

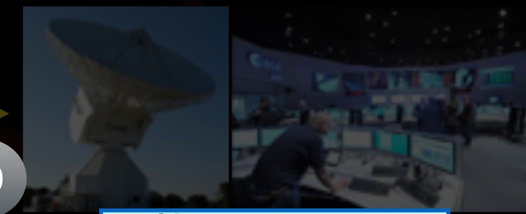
LISA data



Phasemeters (carrier, sidebands, distance)

+ Gravitational Reference Sensor
+ Auxiliary channels

'Survey' type observatory



Gravitational wave sources emitting between 0.02mHz and 100 mHz

L0

Calibrations corrections

Resynchronisation (clock)

Time-Delay Interferometry reduction of laser noise

L1 — 2 data channels TDI non-correlated

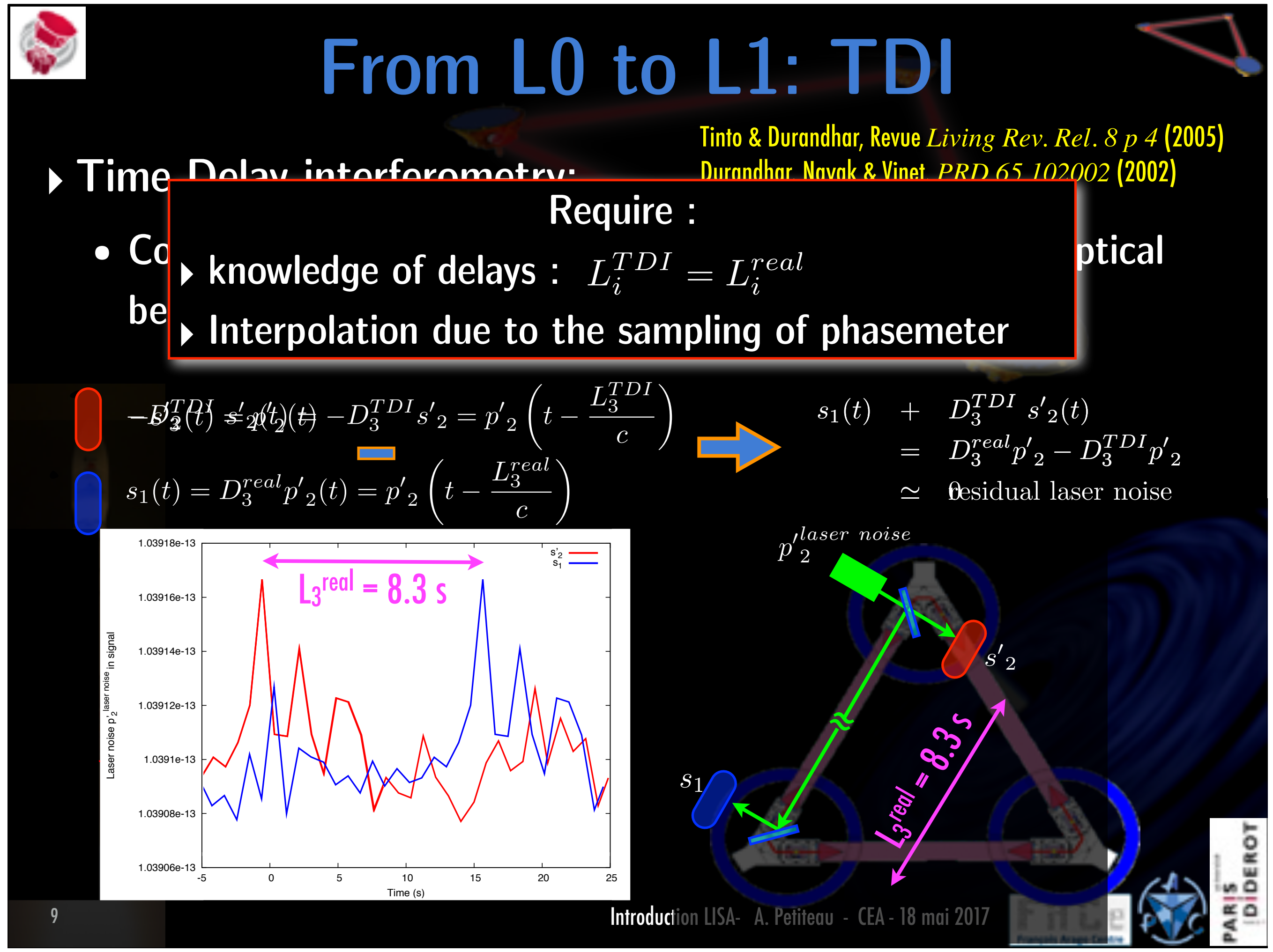
L2 — Data Analysis of GWs

L3 — Catalogs of GWs sources with their waveform

From L0 to L1

- Consolidate the data
- Check data quality
- Calibrations and correction of data (amplitude & time):

  => convert data in usable measurements
- Correct the main measurements by subtracting various effects measured using other channels (a la LISAPathfinder):
  - ex: subtract cross-talk effects
- Synchronise time references (clock) between the 3 spacecrafts

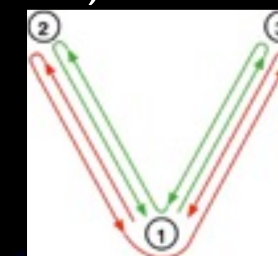


Nécessite :
- une connaissance des temps de parcours
- interpolation du fait de l'echantillonnage du signal

From L0 to L1: TDI

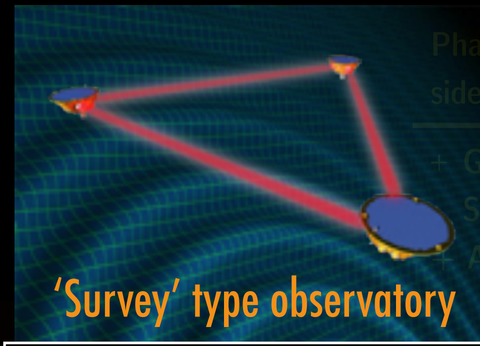
- Time Delay Interferometry:
  - Combine delayed measurements to reduce laser noises, optical bench noises, ... ?
  - Algebraic development : many combinations (generators)

$$
\begin{aligned}
X &= -s_1 - D_3\, s'_2 - D_3 D_{3'}\, s'_1 - D_3 D_{3'} D_{2'}\, s_3 \\
&\quad +s'_1 + D_{2'}\, s_3 - D_{2'} D_2\, s_1 - D_{2'} D_2 D_3\, s_3 \\
&\simeq 0
\end{aligned}
$$



  - Different precisions level
    - 1st generation : rigid formation of LISA : $D_{i'}\, s = D_i\, s$,
    - 1.5 generation : Sagnac effect : $D_{i'}\, s \neq D_i\, s$ but $D_j D_i\, s = D_i D_j\, s$,
    - 2nd generation : flexing and Sagnac effect : $D_j D_i\, s \neq D_i\ D_j\, s$

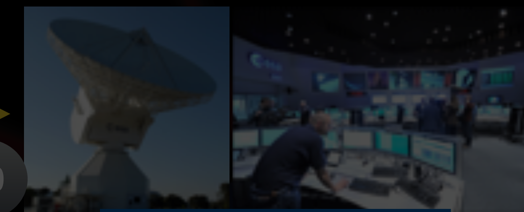
LISA data



Phasemeters (carrier, sidebands, distance)

+ Gravitational Refere
Sensor

+ Auxiliary channels



L0

Calibrations corrections

Resynchronisation (clock)

Time-Delay Interferometry
reduction of laser noise

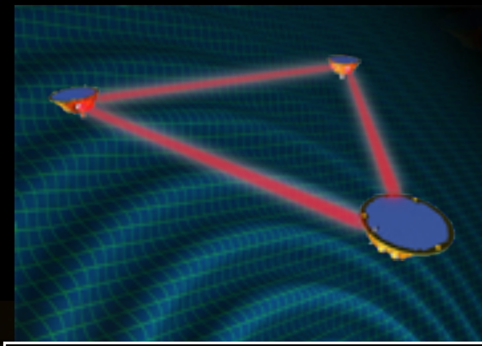
L1 2 data channels TDI non-correlated

L2 Data Analysis of GWs

L3 Catalogs of GWs sources with their waveform

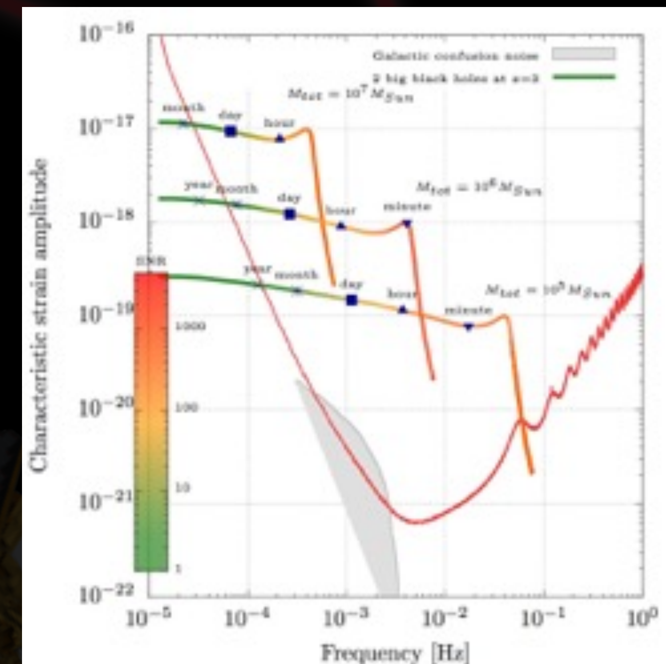
Gravitational wave sources emitting between 0.02mHz and 100 mHz

Galactic binaries



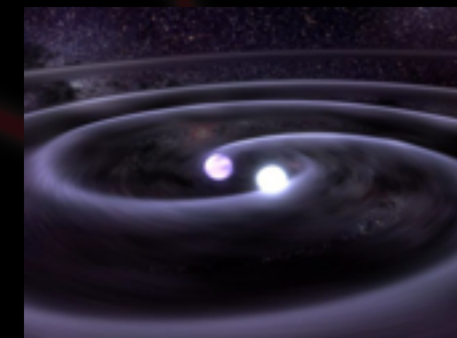
GW sources

- 6 x10^7 galactic binaries

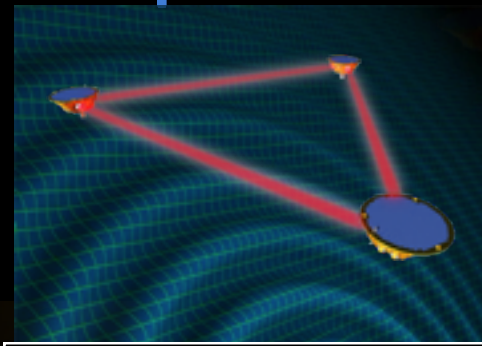


Galactic binaries

- Gravitational wave:
  - quasi monochromatic
- Duration: permanent
- Signal to noise ratio:
  - detected sources: 7 - 1000
  - confusion noise from non-detected sources
- Event rate:
  - 25 000 detected sources
  - more than 10 guarantied sources (verification binaries)

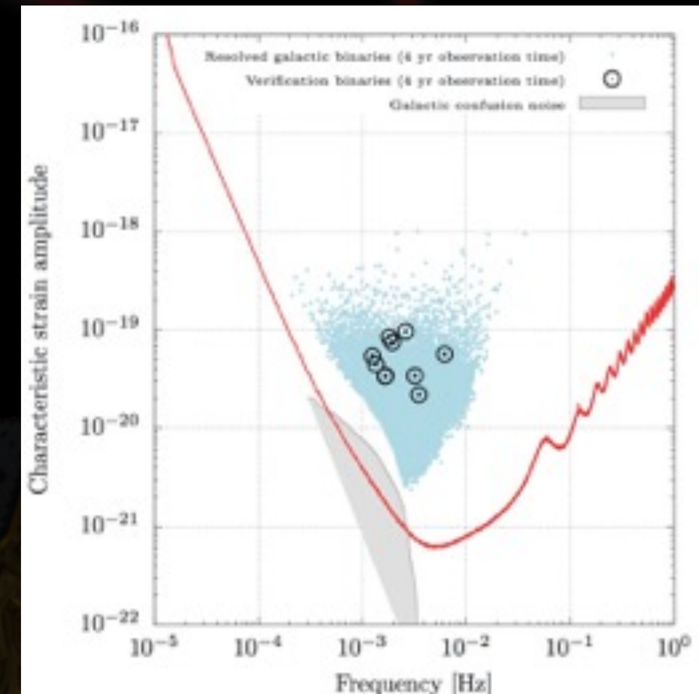


Super Massive Black Hole Binaries



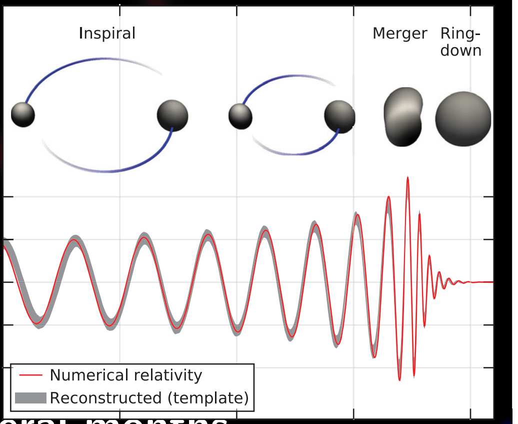
OG sources

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



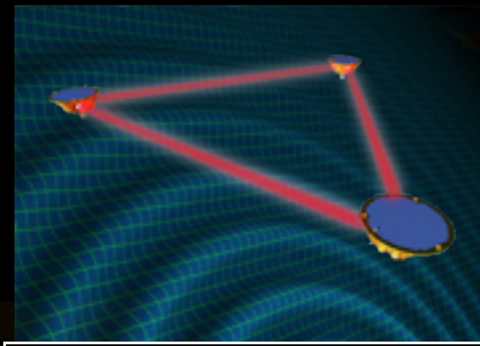
Super Massive Black Hole Binaries

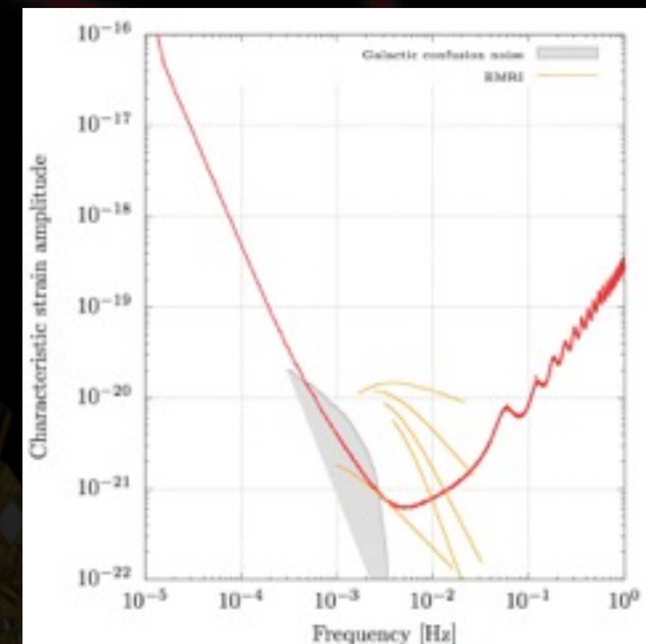
- Gravitational wave:
  - Inspiral: Post-Newtonian,
  - Merger: Numerical relativity,
  - Ringdown: Oscillation of the resulting MBH.



- Duration: between few hours and several months
- Signal to noise ratio: until few thousands
- Event rate: 10-100/year

EMRIs



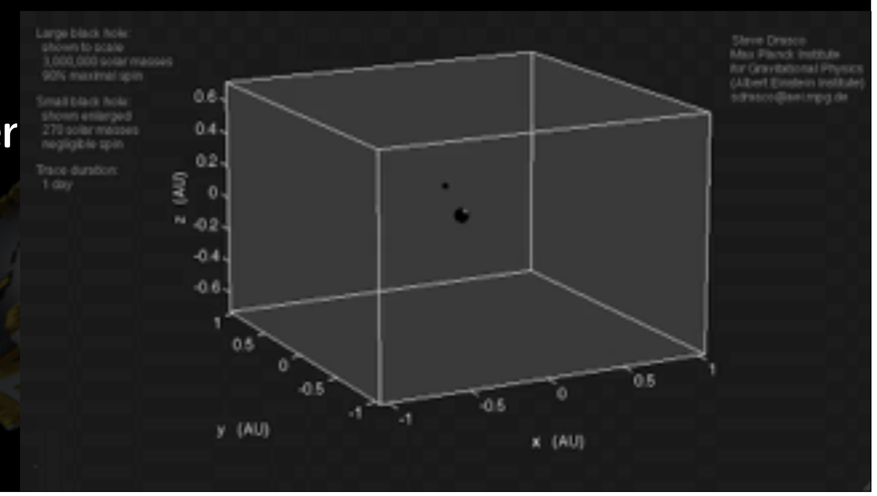


OG sources

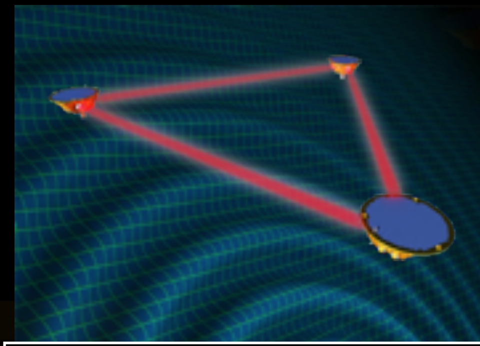
- 6 x10^7 galactic binariess
- 10-100/year SMBHBs
- 10-1000/years EMRIs

EMRIs

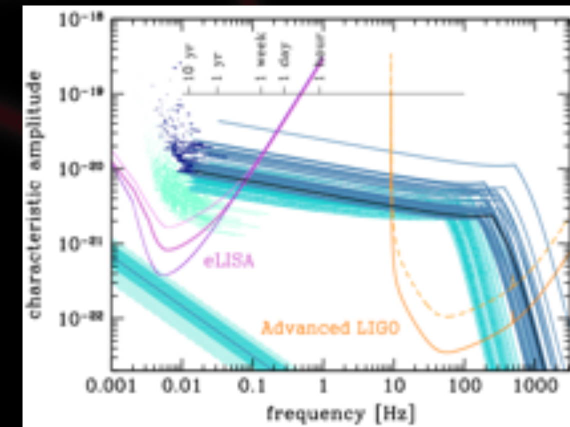
- Gravitational wave:
  - very complex waveform
  - No precise simulation at the moment
- Duration: about 1 year
- Signal to Noise Ratio: from tens to few hundreds
- Event rate:
  from few events per year to few hundreds



Others sources

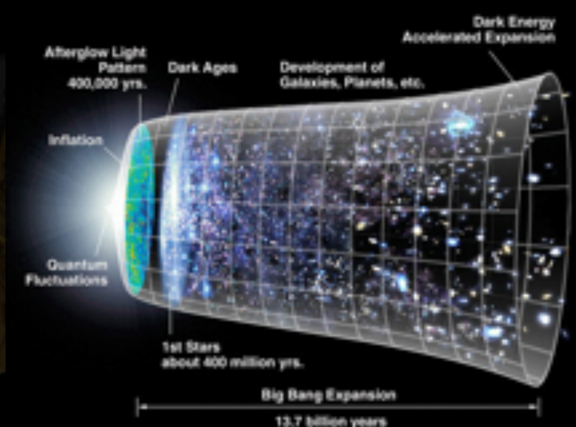


?



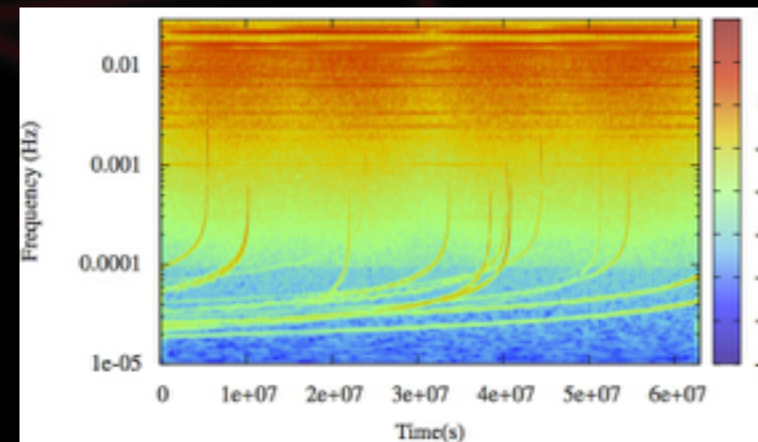
GW sources

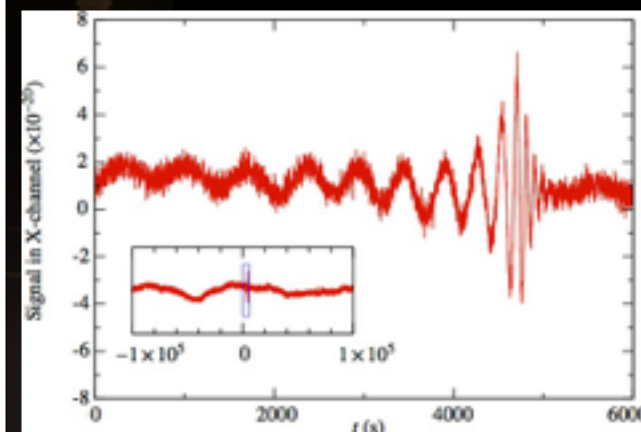
- 6×10^7 galactic binaries
- 10-100/year SMBHBs
- 10-1000/year EMRIs
- large number of Stellar Origin BH binaries (LIGO/Virgo)
- Cosmological backgrounds
- Unknown sources

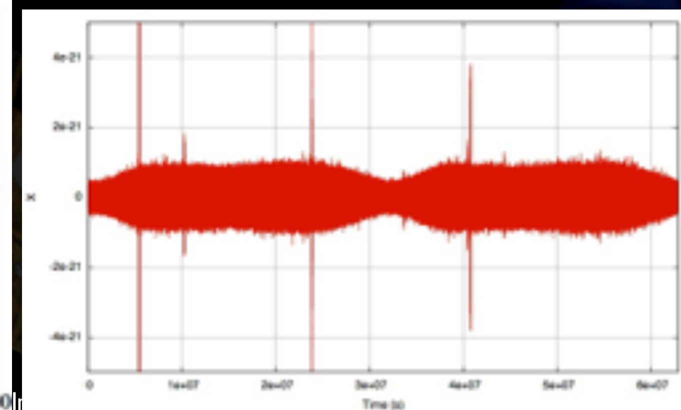


GWs in LISA data

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







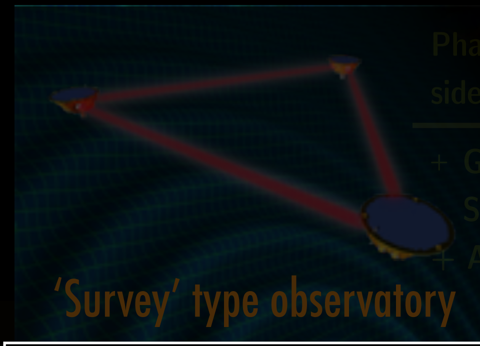
LISA data volume

- Data volume to be stored:
  - Level L0: about 300 Mo per day
  - Level L1: about 600 Mo per day
  - Sub-product of the analysis: fews Go per day
  - Level L2 and L3: about 6 Go per day

  => Storages and archives are not problematic

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

LISA data

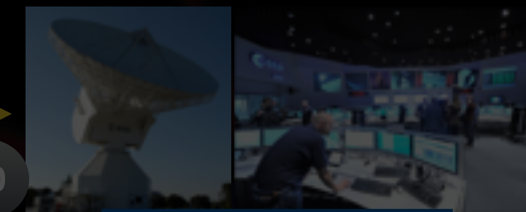


Phasemeters (carrier, sidebands, distance)

+ 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

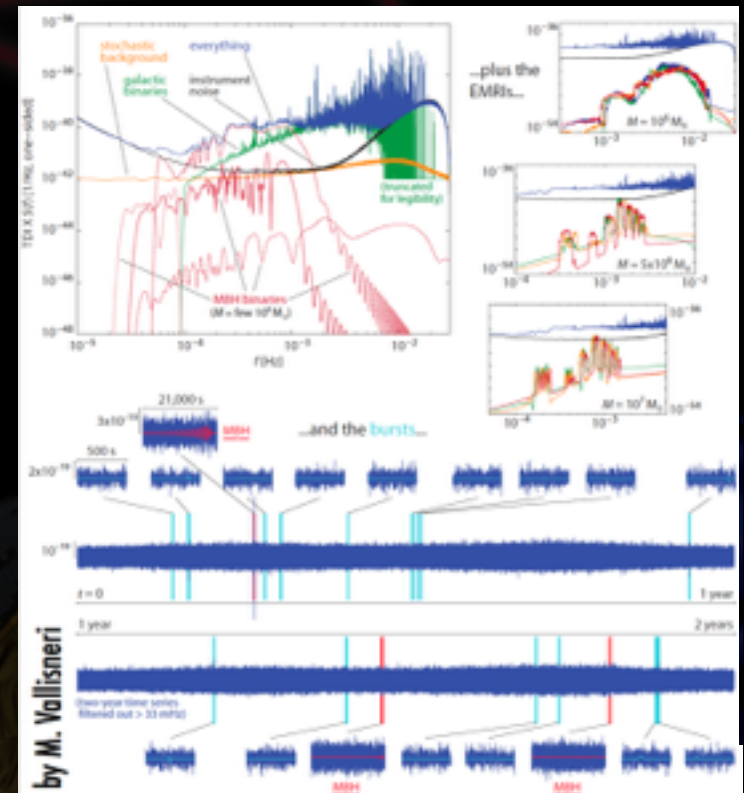
L1 — 2 data channels TDI non-correlated

L2 — Data Analysis of GWs

L3 — Catalogs of GWs sources with their waveform

(Old) Mock LISA Data Challenges

- Old MLDC : 2005 → 2011
- Data: few sources + simplified noises
- Challenges of increasing complexities
- Training & blind dataset
- Goals :
  - Check the feasibility of LISA data analysis
  - Develop « new » data analysis methods



LISA data analysis

- LISA data analysis is tractable
- Main category of methods (same as ground based detector):
  matched filtering : fit the best model to the data by exploring a large parameter space
  - Template bank: check all sets of parameters (too heavy)
  - Frequentist analysis: maximized likelihood: Genetic Algorithm, ...
  - Bayesian analysis: Metropolis Hasting Markov chain, MultiNest: sample posterior distribution => probability distribution on each parameters
- Time-frequency analysis (wavelet), un-modeled waveform analysis, ...

New MLD(C)s

- 2017 → launch: restart MLDC type activities
- Goal: build the pipeline of the mission
- The main challenges:
  - Large number of sources all together and complex waveform
  - Realistic instrument data: complex noises, gaps, glitches, ...
  - « Cumulative analysis » : analyse data accumulating the segments and not directly for the full duration of the mission
- 2 parallel ways that have to converge:
  - Simplified noises but more realistic waveform and number of sources growth toward realistic even rates
  - Few simple sources + complex noises based on LISAPathfinder and technical developments

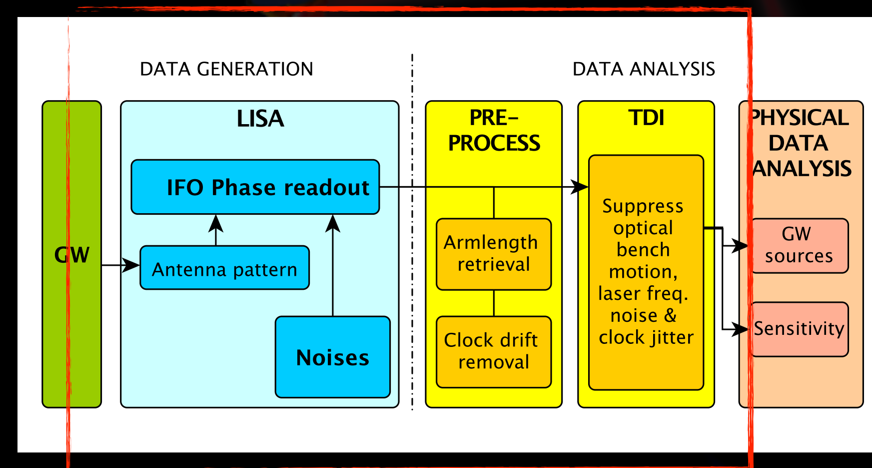
Overview

- LISA data analysis and data processing
  - Overview and data level
  - Calibration, TDI, etc
  - Sources and waveforms
  - Extracting sources and their parameters
- LISA Simulation
- Data Processing Center
  - Tasks
  - Organisation

Simulation

- Several types of simulation required:
  - Generate data for the MLDC
  - Study hardware performance of subsystem and interface
  - Develop preprocessing pipelines
  - Simulate the 3D dynamics of the 9 bodies with sensing and actuation
  - Study signal from gravitational wave source after the detector
- Several existing simulator of various types:
  - LISACode, LISADyn, TDISim, LISASimulator, Synthetic LISA, ...

Simulation



- LISACode (new modular version in development)
- 2 complementary simulators:
  - TDISim (check TDI)
  - LISADyn (3D dynamic)

Mission simulator

- **Goals**:
  - End-to-end simulation → the mission simulator
  - "Quick performance" study for various configurations → final design (required for phase A)
  - Accompany the hardware developments (industries & labs.)
  - Tool(s) for performance controls
- **First requirements**:
  - Close modeling of the instrument subsystems
  - Waveform generation for various GW sources
  - Noise generation using various types of representation
  - Data pre-processing (distinct from simulation)
  - Modularity
  - Computation speed (> 10-100 times faster than reality)
  - Open-source

Overview

- LISA data analysis and data processing
  - Overview and data level
  - Calibration, TDI, etc
  - Sources and waveforms
  - Extracting sources and their parameters
- LISA Simulation
- Data Processing Center
  - Tasks
  - Organisation

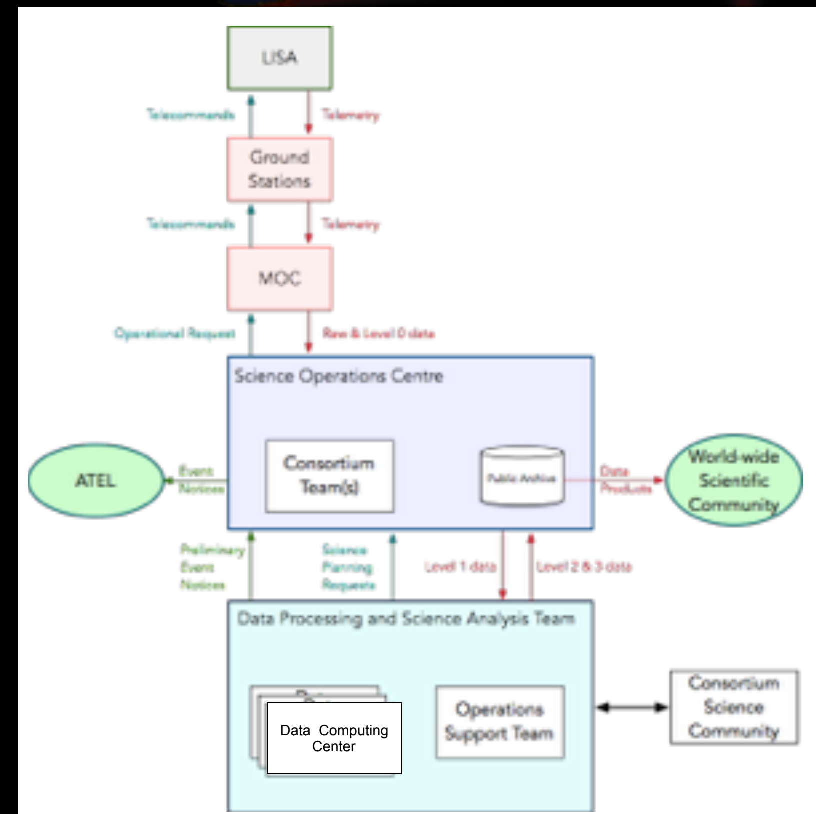
Particularities LISA data

- **First** data of this kind
  - Discovery mission; no previous expertise on this kind of data
- Event rate is **uncertain**
  - Depending on the type of sources but typically from few tens to few thousands per year
- Potential **unknown** sources
- **Transient** sources + continuous sources

=> Constrains on data processing:

- **Large fluctuation** of computation needs
- **Continuous evolution** of the pipelines

LISA DPC



LISA
Telecommands
Telemetry
Ground Stations
Telecommands
Telemetry
MOC
Operational Request
Raw & Level 0 data
Science Operations Centre
ATEL
Event Notices
Consortium Team(s)
Public Archive
Data Products
World-wide Scientific Community
Preliminary Event Notices
Science Planning Requests
Level 1 data
Level 2 & 3 data
Data Processing and Science Analysis Team
Data Computing Center
Operations Support Team
Consortium Science Community

DPC in LISA proposal

- DPC activities:
  - Receive L1 data from the SOC;
  - Identify and extract waveforms;
  - Build the catalogs of sources;
  - Create L2 et L3 science products;
  - Analyse the quality of science data products;
  - Distribute data to SOC and to the scientific community of the Consortium
  - Produce periodic releases of science data products
  - Generate alerts for upcoming transients, such as mergers

DPC CNES Phase 0: conclusions

- Infrastructure: fluctuation of computing load :
  - constant sources + transitoires
  - Regular full reprocessing:
    - More data accumulated
    - Continuous evolution of pipelines
- Proposed solution: standard cluster + cloud(s)

Current vision of the DPC

- DPC: unique entity responsable for the data processing (driving, integration of software block, ...)
- DPC 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 DPC software « suite » can run on any DCC.
  - Software: codes (DA & Simu.) + services (LDAP, wiki, database) + OS.
- First solutions:
  - Separation of hardware and software: ligth virtualization, ...
  - Collaborative development: continuous integration, ...
  - Fluctuations of computing load: hybrids cluster/cloud

Current vision of the DPC

- The DPC is a tools for the consortium
- To avoid reintegration in the pipeline of blocks developed separately (ex: difficult in GAIA), our idea is:
  - to develop with scientists the tools that fulfill their needs,
  - to adapt « DPC tools » (i.e. required tools for having consolidated DA pipelines,) to the scientists,
  - to make scientists and developers used to all these tools and to the way of working with them.
- In 2015, we started at APC with the support of CNES the development of a proto-DPC:
  - continuous integration, technological watch, virtualization, docker, ...

LISA proto-DPC

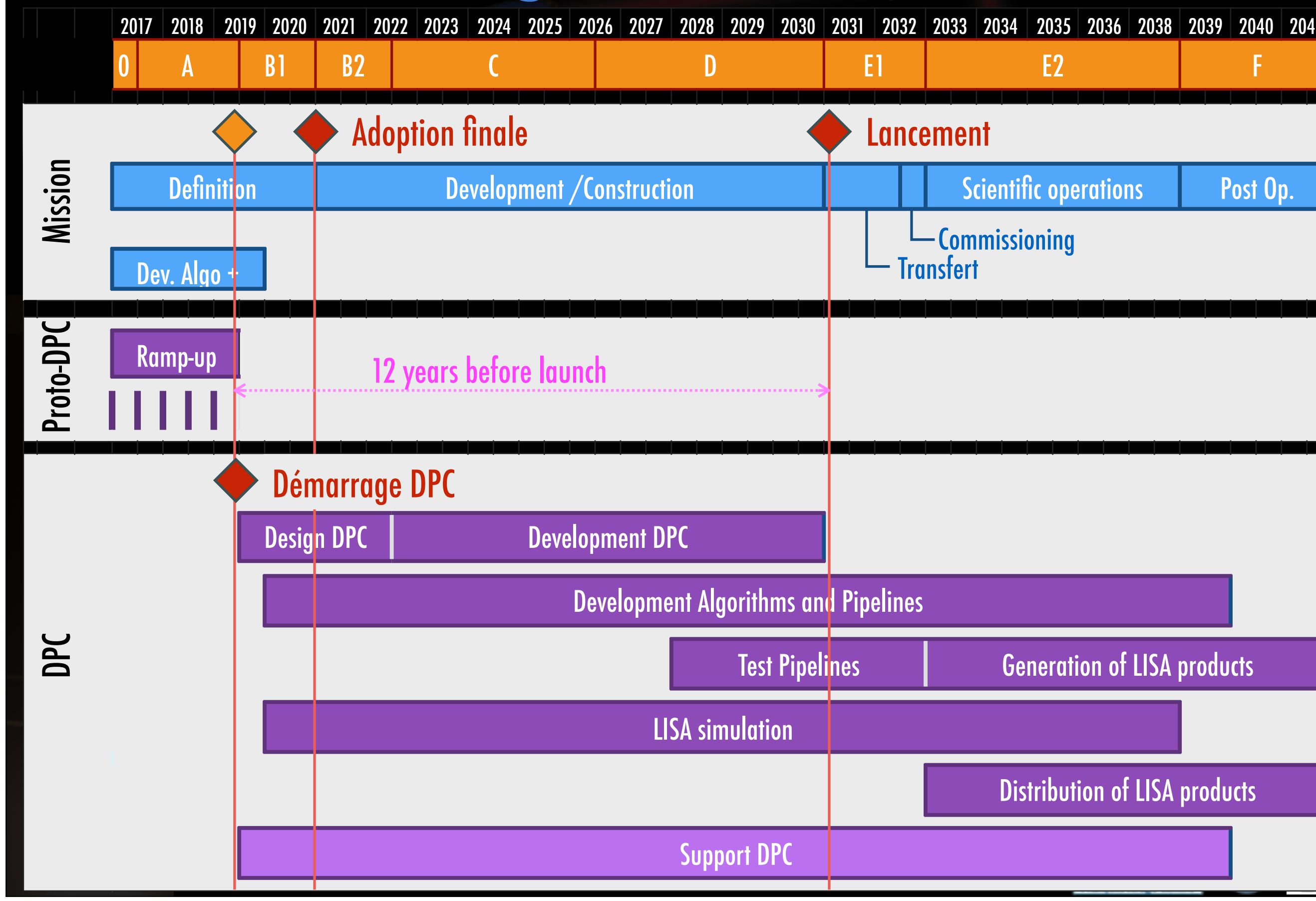
- https://elisadpc.in2p3.fr/home
- Already used by consortium (simulation, proposal, …)

Conclusion

- LISA data analysis is challenging but possible
- Low data volume / high computing
- Required developments:
  - Data analysis:
    - now: essentially based on matched filtering.
    - needs: improve samplers and estimators for matched filtering, global strategies, « burst type » analysis, new ideas !
  - Infrastructure: service/tool supplying commonalities, development environment, devops, virtualisation, database, collaboration enablers, ...
- Important role of France (DPC): first proto-DPC in place ... open meeting DPC-France soon.


Thank you

Planning of development



Continuous integration