LISA Data Challenge.

Stas Babak

Astroparticule et Cosmologie (APC)
Paris, France

18 October, 2018, Mainz

Aim of LDC

- To foster the data analysis development: improve performance of existing algorithms, try new algorithms
- To make a common platform for evaluation and performance comparison of various algorithms
- To address the science requirements: Project oriented challenges
- To introduce the software development standards for the data analysis pipeline
- To prototype and develop the end-to-end data analysis pipeline

Aim of LDC

- To foster the data analysis development: improve performance of existing algorithms, try new algorithms
- To make a common platform for evaluation and performance comparison of various algorithms
- To address the science requirements: Project oriented challenges
- To introduce the software development standards for the data analysis pipeline
- To prototype and develop the end-to-end data analysis pipeline

Aim of LDC

- To foster the data analysis development: improve performance of existing algorithms, try new algorithms
- To make a common platform for evaluation and performance comparison of various algorithms
- To address the science requirements: Project oriented challenges
- To introduce the software development standards for the data analysis pipeline
- To prototype and develop the end-to-end data analysis pipeline

Aim of LDC

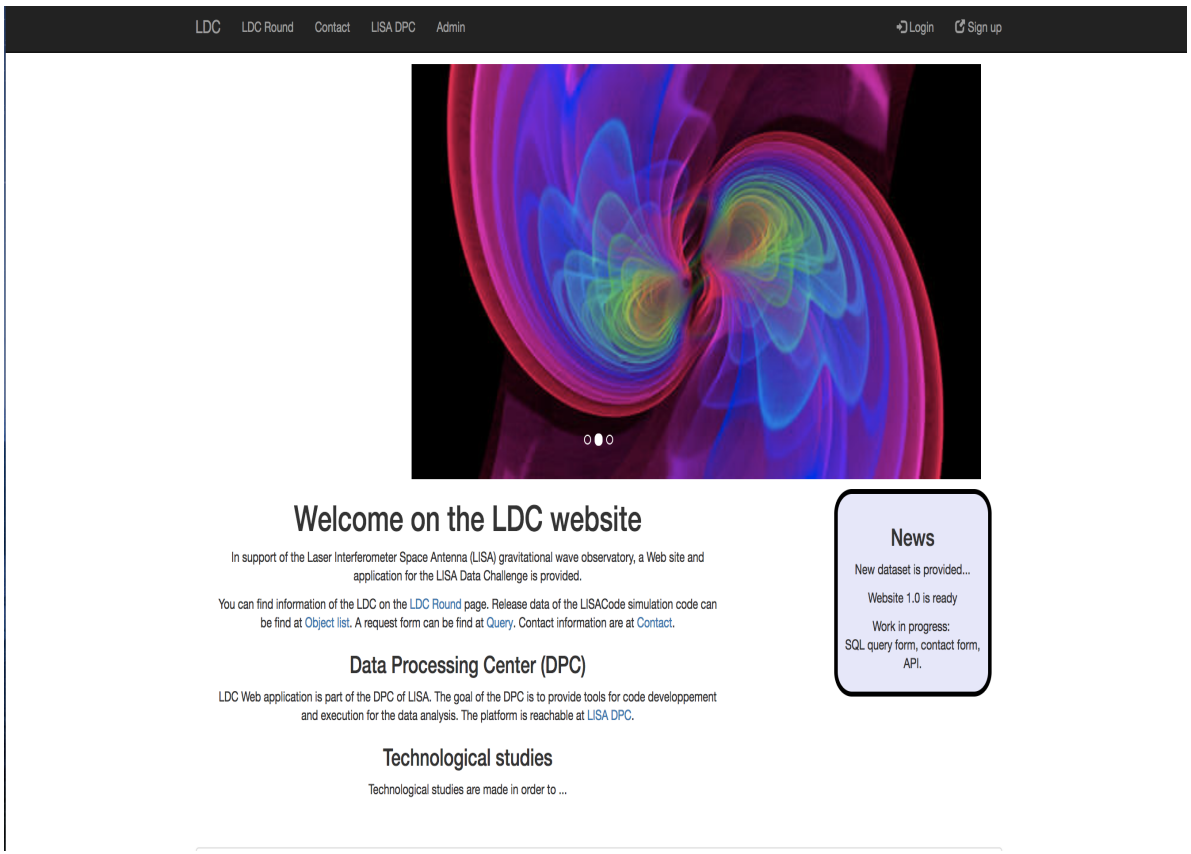
- To foster the data analysis development: improve performance of existing algorithms, try new algorithms
- To make a common platform for evaluation and performance comparison of various algorithms
- To address the science requirements: Project oriented challenges
- To introduce the software development standards for the data analysis pipeline
- To prototype and develop the end-to-end data analysis pipeline

Aim of LDC

- To foster the data analysis development: improve performance of existing algorithms, try new algorithms
- To make a common platform for evaluation and performance comparison of various algorithms
- To address the science requirements: Project oriented challenges
- To introduce the software development standards for the data analysis pipeline
- To prototype and develop the end-to-end data analysis pipeline

LDC web page

- The project is hosted under git-lab https://gitlab.in2p3.fr/stas/MLDC (sign up is required)
- There are bi-weekly teleconferences. The web-page will be open shortly for everyone to sign up for the challenge and download the simulated data set.



History: MLDC 2006-2011

	MLDC 1	MLDC 2	MLDC 1B	MLDC 3	MLDC 4
Galactic binaries	• Verification • Unknown isolated • Unknown interfering	• Galaxy 3×10^6	• Verification • Unknown isolated • Unknown interfering	• Galaxy 6×10^7 chirping	• Galaxy 6×10^7 chirping
Massive BH binaries	• Isolated	• 4-6x, over "Galaxy" & EMRIs	• Isolated	• 4-6x spinning & precessing over "Galaxy"	• 4-6x spinning & precessing, extended to low-mass
EMRI		• Isolated • 4-6x, over "Galaxy" & MBHs	• Isolated	• 5 together, weaker	• 3 x Poisson(2)
Bursts				• Cosmic string cusp	• Poisson(20) cosmic string cusp
Stochastic background				• Isotropic	• Isotropic

Generating LDC data set

What do we need to create the simulated data set

- We need to decide on the GW sources (and number of sources) which we want to put in the data
- We need to decide on the parameters of each signal (we will use catalogues of sources based on several astrophysical models)
- We need to decide on the theoretical model of the GW signal to be used ("state-of-art" models are usually computationally expensive)
- We need to apply the response function to the GW signal : requires LISA orbit.
- We need to decide on the noise (simplistic: equal noise in each measurement, uncorrelated, Gaussian, or ....)
- We need to produce the noise and add the signal(s): input from LPF

Generating LDC data set

What do we need to create the simulated data set

- We need to decide on the GW sources (and number of sources) which we want to put in the data
- We need to decide on the parameters of each signal (we will use catalogues of sources based on several astrophysical models)
- We need to decide on the theoretical model of the GW signal to be used ("state-of-art" models are usually computationally expensive)
- We need to apply the response function to the GW signal : requires LISA orbit.
- We need to decide on the noise (simplistic: equal noise in each measurement, uncorrelated, Gaussian, or ....)
- We need to produce the noise and add the signal(s): input from LPF

Generating LDC data set

What do we need to create the simulated data set

- We need to decide on the GW sources (and number of sources) which we want to put in the data
- We need to decide on the parameters of each signal (we will use catalogues of sources based on several astrophysical models)
- We need to decide on the theoretical model of the GW signal to be used ("state-of-art" models are usually computationally expensive)
- We need to apply the response function to the GW signal : requires LISA orbit.
- We need to decide on the noise (simplistic: equal noise in each measurement, uncorrelated, Gaussian, or ....)
- We need to produce the noise and add the signal(s): input from LPF

Generating LDC data set

What do we need to create the simulated data set

- We need to decide on the GW sources (and number of sources) which we want to put in the data
- We need to decide on the parameters of each signal (we will use catalogues of sources based on several astrophysical models)
- We need to decide on the theoretical model of the GW signal to be used ("state-of-art" models are usually computationally expensive)
- We need to apply the response function to the GW signal : requires LISA orbit.
- We need to decide on the noise (simplistic: equal noise in each measurement, uncorrelated, Gaussian, or ....)
- We need to produce the noise and add the signal(s): input from LPF

Generating LDC data set

What do we need to create the simulated data set

- We need to decide on the GW sources (and number of sources) which we want to put in the data
- We need to decide on the parameters of each signal (we will use catalogues of sources based on several astrophysical models)
- We need to decide on the theoretical model of the GW signal to be used ("state-of-art" models are usually computationally expensive)
- We need to apply the response function to the GW signal : requires LISA orbit.
- We need to decide on the noise (simplistic: equal noise in each measurement, uncorrelated, Gaussian, or ....)
- We need to produce the noise and add the signal(s): input from LPF

Generating LDC data set

What do we need to create the simulated data set

- We need to decide on the GW sources (and number of sources) which we want to put in the data
- We need to decide on the parameters of each signal (we will use catalogues of sources based on several astrophysical models)
- We need to decide on the theoretical model of the GW signal to be used ("state-of-art" models are usually computationally expensive)
- We need to apply the response function to the GW signal : requires LISA orbit.
- We need to decide on the noise (simplistic: equal noise in each measurement, uncorrelated, Gaussian, or ....)
- We need to produce the noise and add the signal(s): input from LPF

The data sets in the "Radler"-data release

The aim is to resurrect the existing tools and get up to speed for real challenges. The noise will be **very** simple (Gaussian, uncorrelated A, E, T channels), analytic LISA orbit, 1.5 generation TDI (rigid LISA). Duration of each simulated data is 1 year.

- Galaxy: Gaussian noise + 60 mln. Galactic white dwarf binaries (using the new catalogue for the detached binaries)
- Binary massive black holes (V1): Gaussian noise + 1 MBH binary system, SNR $\sim 150 - 300$, spinning non-precessing, includes inspiral, merger and ringdown (IMRPhenomD model)

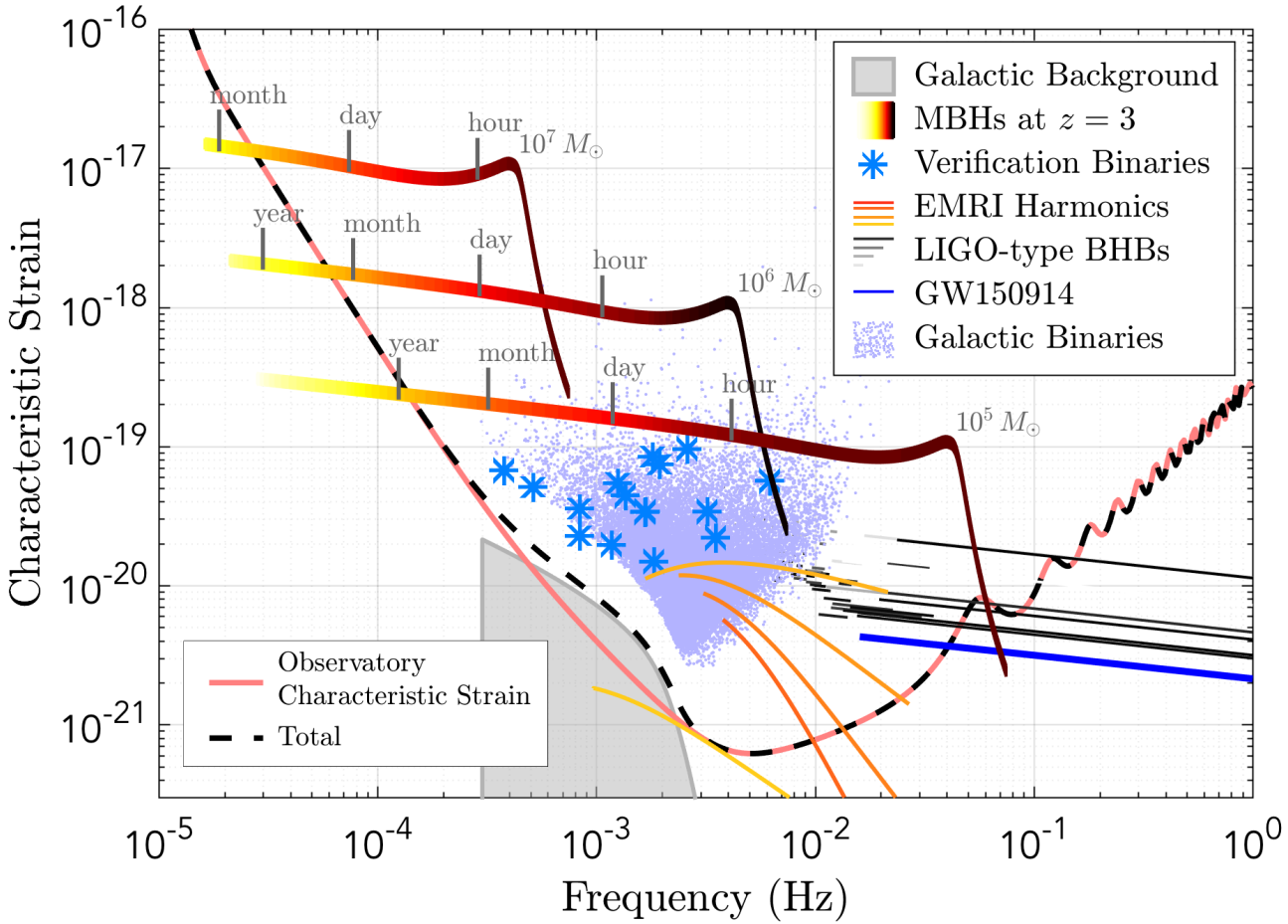
The data sets in th "Radler"-data release

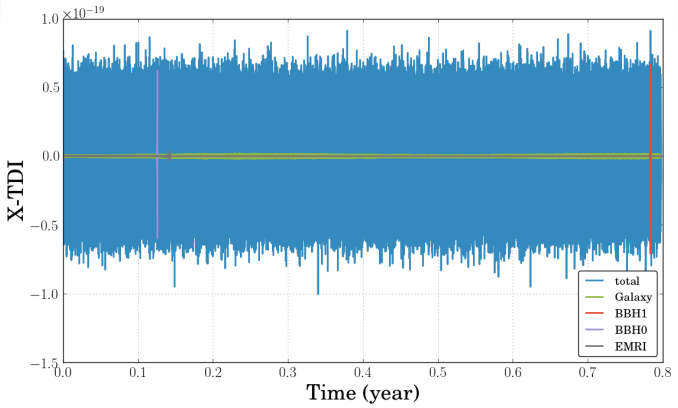
- Binary massive black holes (V2): Gaussian noise but unequal in each link + 1 MBH binary system, SNR $\sim 150 - 300$, spinning non-precessing, includes inspiral, merger and ringdown (IMRPhenomD model)
- Extreme mass ratio inspirals (EMRIs): Gaussian noise + 1 EMRI system SNR $\sim 40 - 60$, generic orbit. Relatively narrow priors on the source parameters. Using AK model (Barack & Cutler 2006)

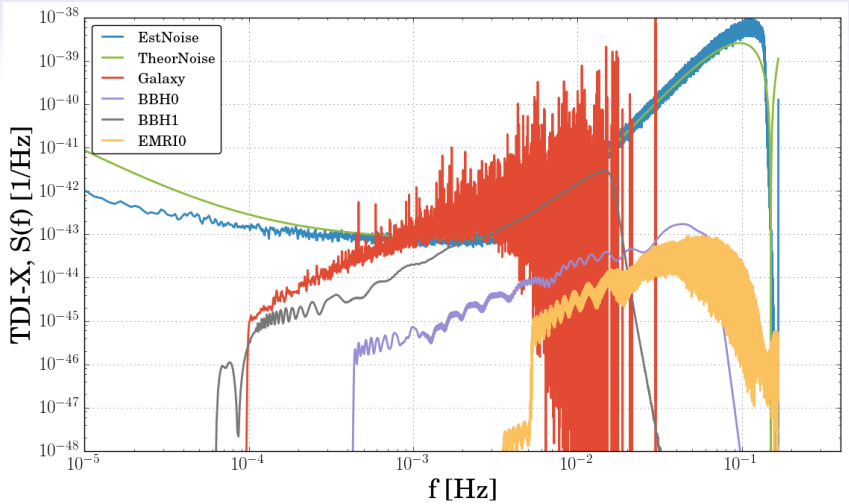
The data sets in th "Radler"-data release

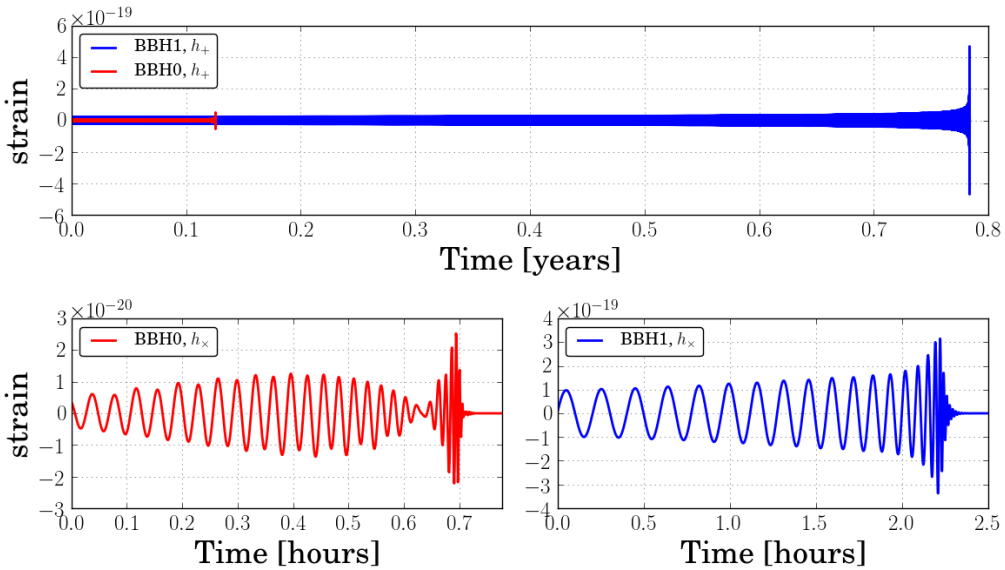
- LIGO sources: binary black holes of stellar origin. Those are BBHs observed in LIGO band. Gaussian noise + population of BBHs. There could be additional data set with only "bright", individually detectable sources. Problem of data generation (chirping, high sampling rate, long duration).
- Stochastic, isotropic GW signal + Gaussian noise. Detectable level. Realized as a superposition of uniformly distributed on the sky sources.

Overview of sources in the LISA band









Comments: (bottom figure) Cartoon showing GW signal from two spinning (precessing) MBH binaries. (top figure) The simulated X-TDI data with instrumental noise, Galaxy, 1 EMRI and 2 MBH binaries in time and in frequency domain.

Beyond the first data set

- We need to move away from the simplistic assumption about the noise
- Develop the pipelines to produce L1 data (TDI) from raw data (L0): calibrations, remove / reduce noises, gaps, frequency planning, non-stationarity, unexpected events
- We will utilize LPF results to mimic non-stationary noise in LISA simulations
- Work together with the simulation group: end-to-end simulation of the data
- For each astrophysical source we need to revisit the detection (Gaussian) algorithms with realistic noise

Summary

The LDC production is underway...

- LDC webpage is almost there
- Waveform generation code: almost there need some tests.
- The data and metadata will be stored in hdf5 format and distributed via web interface (stored in the database)
- LISACode – the simulator which will be used to apply response function and to produce the noise.
- First prototype of simplistic pipeline, data flow, data products, data formats, code integration, ...

We learn through doing...