LISA Data Challenge.

Stas Babak and Antoine Petiteau (for LDC working group)

Astroparticule et Cosmologie (APC)
Paris, France

1-2 March, 2018, Padova
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 (integration into DDPC – Distributed Data Processing Center).

Stas Babak and Antoine Petiteau (for LDC working group)
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 (integration into DDPC – Distributed Data Processing Center).
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 (integration into DDPC – Distributed Data Processing Center).
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 (integration into DDPC – Distributed Data Processing Center).

Stas Babak and Antoine Petiteau (for LDC working group)

1-2 March 2018
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 (integration into DDPC – Distributed Data Processing Center).
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 is open to public (just requires registration to access data).
  https://lisa-ldc.lal.in2p3.fr/home
## History: MLDC 2006-2011

<table>
<thead>
<tr>
<th></th>
<th>MLDC 1</th>
<th>MLDC 2</th>
<th>MLDC 1B</th>
<th>MLDC 3</th>
<th>MLDC 4</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Galactic binaries</strong></td>
<td>Verification, Unknown isolated, Unknown interfering</td>
<td>Galaxy $3 \times 10^6$</td>
<td>Verification, Unknown isolated, Unknown interfering</td>
<td>Galaxy $6 \times 10^7$ chirping</td>
<td>Galaxy $6 \times 10^7$ chirping</td>
</tr>
<tr>
<td><strong>Massive BH binaries</strong></td>
<td>Isolated</td>
<td>4-6x, over “Galaxy” &amp; EMRIs</td>
<td>Isolated</td>
<td>4-6x spinning &amp; precessing over “Galaxy”</td>
<td>4-6x spinning &amp; precessing, extended to low-mass</td>
</tr>
<tr>
<td><strong>EMRI</strong></td>
<td>Isolated</td>
<td>4-6x, over “Galaxy” &amp; MBHs</td>
<td>Isolated</td>
<td>5 together, weaker</td>
<td>3 x Poisson(2)</td>
</tr>
<tr>
<td><strong>Bursts</strong></td>
<td></td>
<td></td>
<td></td>
<td>Cosmic string cusp</td>
<td>Poisson(20) cosmic string cusp</td>
</tr>
<tr>
<td><strong>Stochastic background</strong></td>
<td></td>
<td></td>
<td></td>
<td>Isotropic</td>
<td>Isotropic</td>
</tr>
</tbody>
</table>

Stas Babak and Antoine Petiteau (for LDC working group)
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 with 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 with 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 with 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 with 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 with 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 with the signal(s): input from LPF

Stas Babak and Antoine Petiteau (for LDC working group)
Generating LDC data set: first challenge

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 with the signal(s): input from LPF
LDC data production pipeline

**USER INTERFACE**
- Source type
- number of sources
- param. range
- SNR request
- Model/approximation
- time/frequency (dt, df)
- Observation duration
- LISA model
  - (orbit, configuration, TDIs, noise specification)

**HDF Object**
(data container)

- Choose the sources
- Generate h+, hx
- Generate TDIs
- Data Analysis pipeline
- Catalogues of sources
- Waveform repository
- Response generation LISACode, fast response
- require SNR
- No
- Publish data on LDC web (database)
- Yes

*Stas Babak and Antoine Petiteau (for LDC working group)*

*1-2 March 2018*
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 the "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 2004)
The data sets in the "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
What is LISA Data Challenge

Generation of LDC

Start up: "Radler" challenge

Stochastic GW Background example

Stas Babak and Antoine Petiteau (for LDC working group) 1-2 March 2018
Massive BH Binaries example
What is LISA Data Challenge

Generation of LDC

Start up: "Radler" challenge

EMRI example

Stas Babak and Antoine Petiteau (for LDC working group)

1-2 March 2018
Galactic binaries example

- Frequency (Hz)
- Power Spectral Density (1/Hz)
- Time (sec)
- Fractional Frequency

Stas Babak and Antoine Petiteau (for LDC working group)

1-2 March 2018
SOBBH example
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 instrumental artefacts in LISA simulations (gaps, glitches, non-stationarity)
- Work together with the simulation group: **end-to-end simulation** of the data
- Work on the estimation effect of gaps is under way
- For each astrophysical source we need to revisit the detection (Gaussian) algorithms with realistic noise
- Set of tutorials...
Summary

The LDC production is underway...

- LDC webpage is there
- Waveform generation code: needs further 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, uploading the data to the database and retrieving - done. Requires extensive testing.

We learn through doing...