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).

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

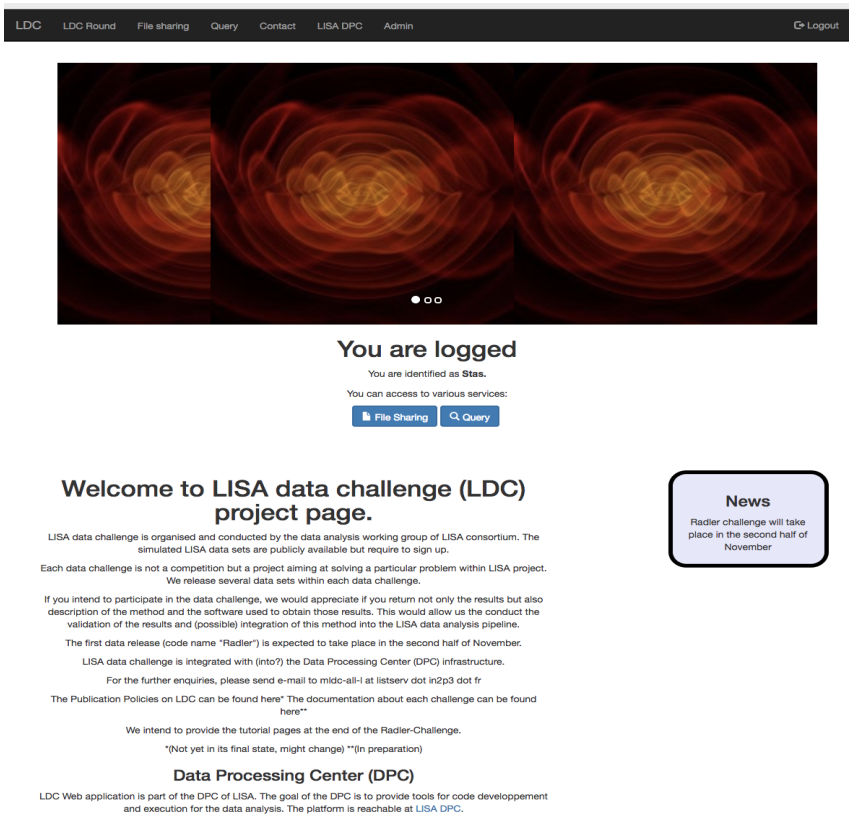
- To foster the data analysis development: improve performance of existing algorithms, try new algorithms
- To make a common platform for evaluation and performance comparison of various algorithms
- To address the science requirements: project oriented challenges
- To introduce the software development standards for the data analysis pipeline
- To prototype and develop the end-to-end data analysis pipeline (integration into DDPC – Distributed Data Processing Center).

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

	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 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

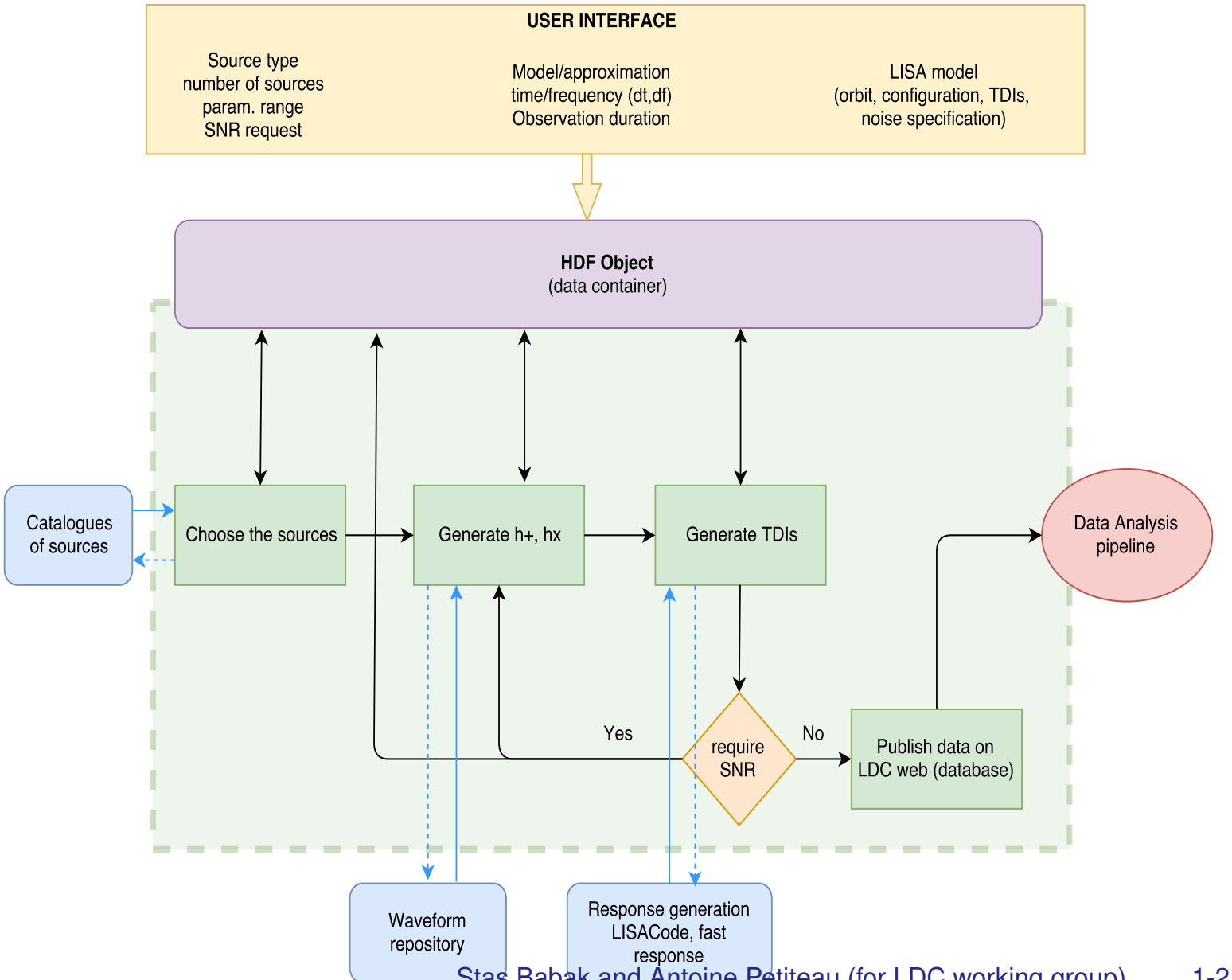
- We need to decide on the GW sources (and number of sources) which we want to put in the data
- We need to decide on the parameters of each signal (we will use catalogues of sources based on several astrophysical models)
- We need to decide on the theoretical model of the GW signal to be used ("state-of-art" 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

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



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)

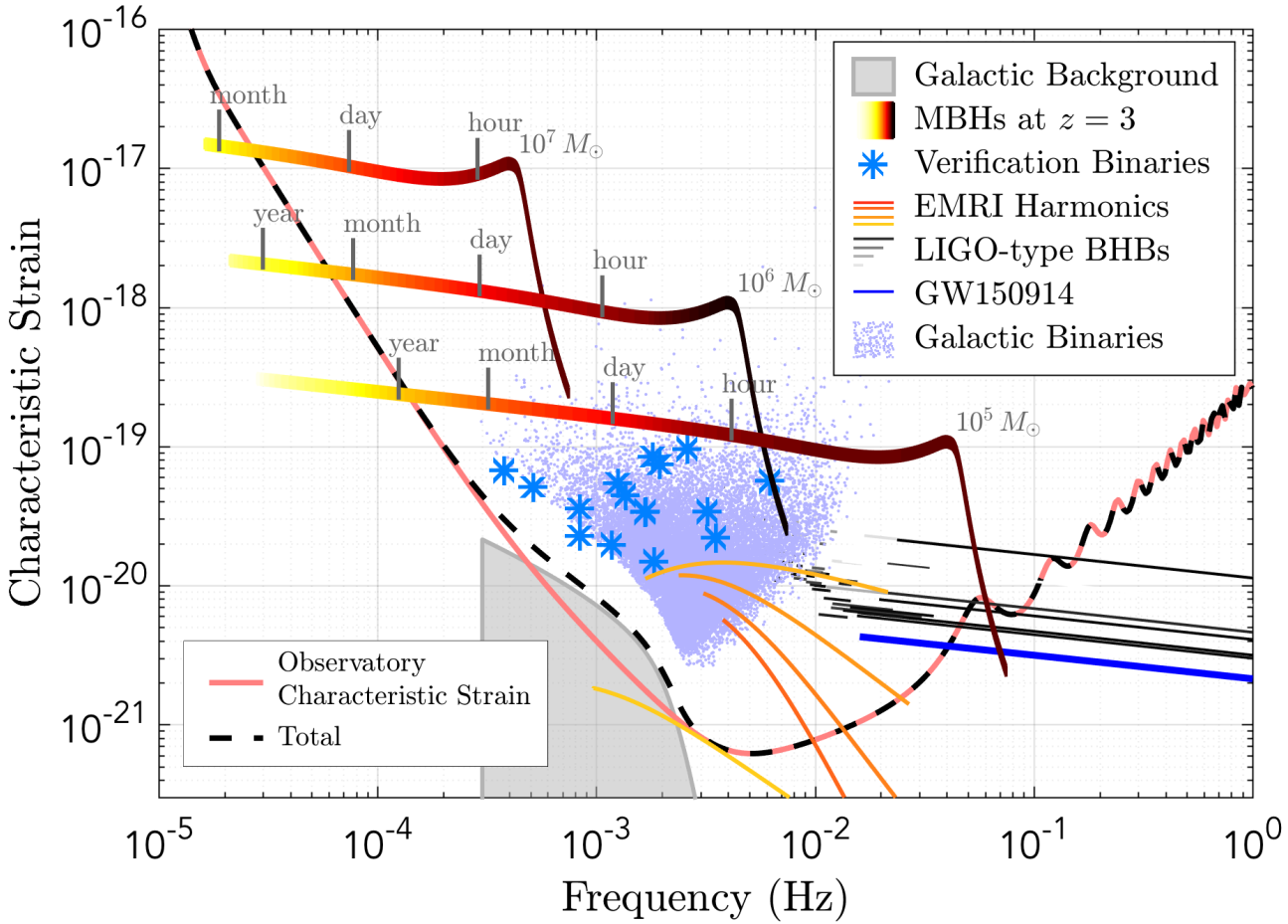
The data sets in the "Radler"-data release

- 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)

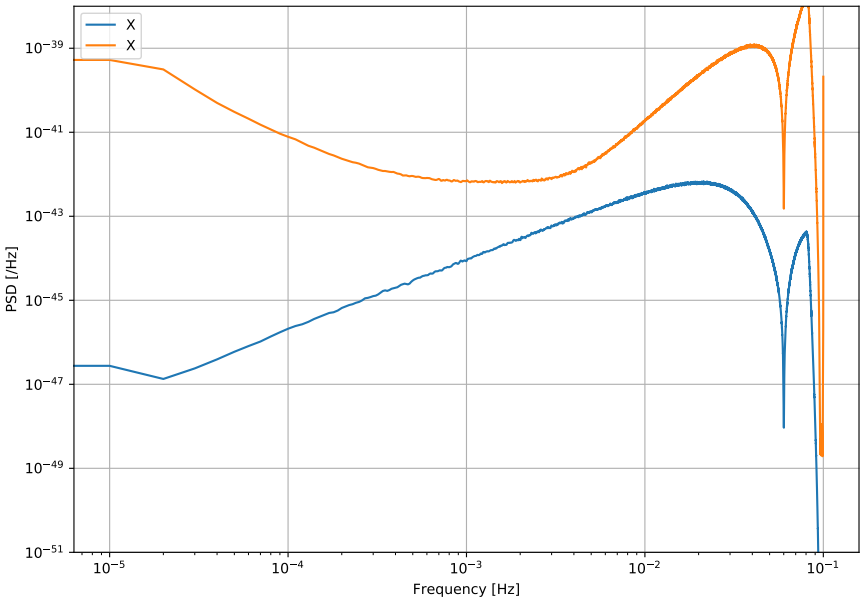
The data sets in the "Radler"-data release

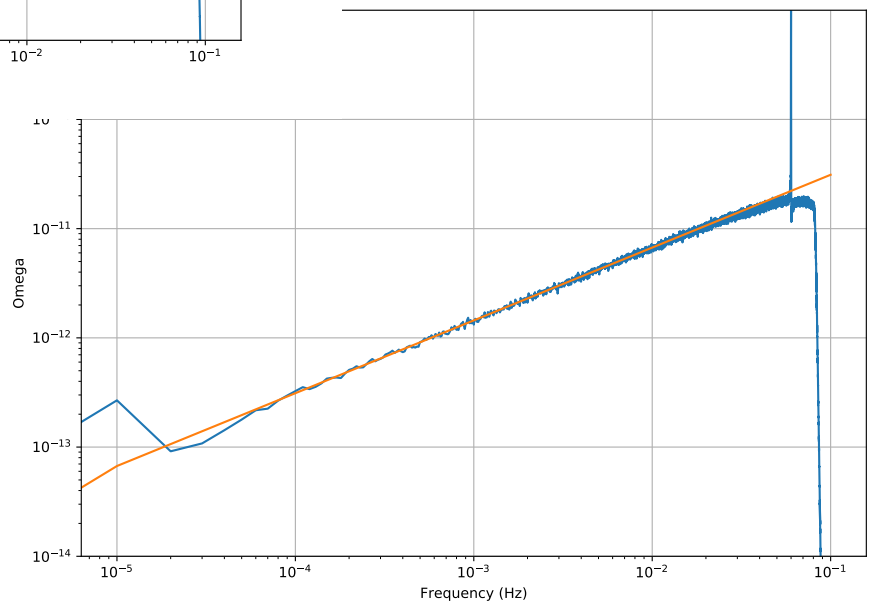
- LIGO sources: binary black holes of stellar origin. Those are BBHs observed in LIGO band. Gaussian noise + population of BBHs. There could be additional data set with only "bright", individually detectable sources. Problem of data generation (chirping, high sampling rate, long duration).
- Stochastic, 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

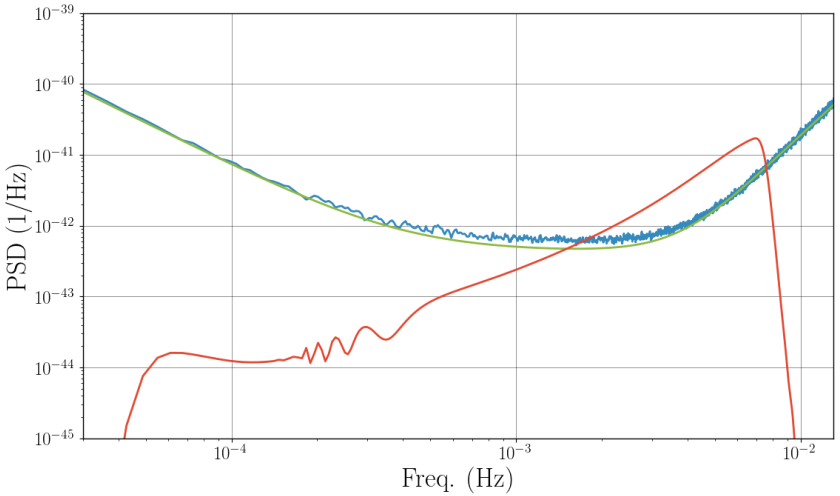


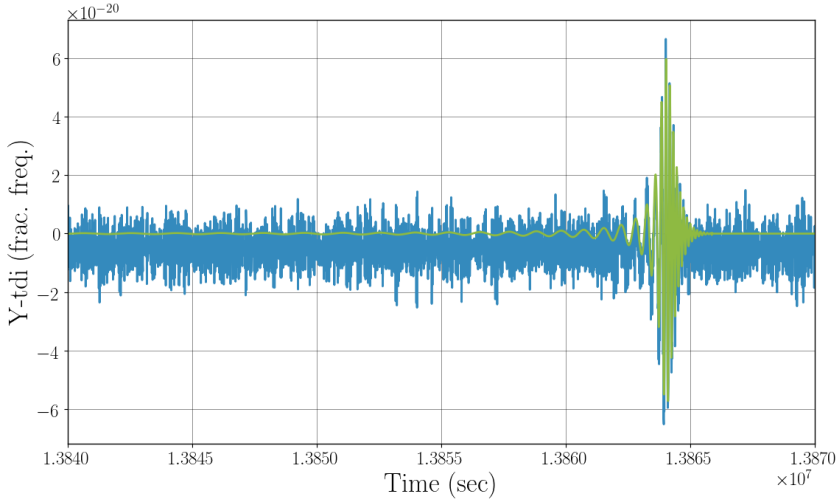
Stochastic GW Background example

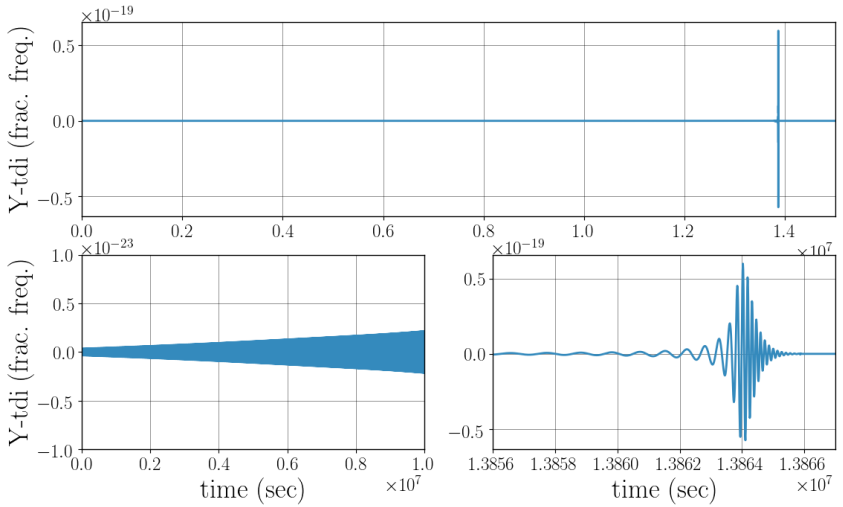




Massive BH Binaries example



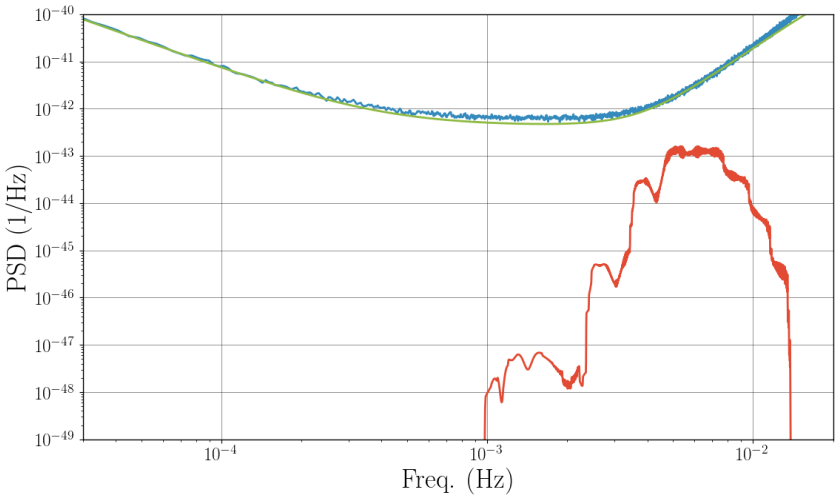


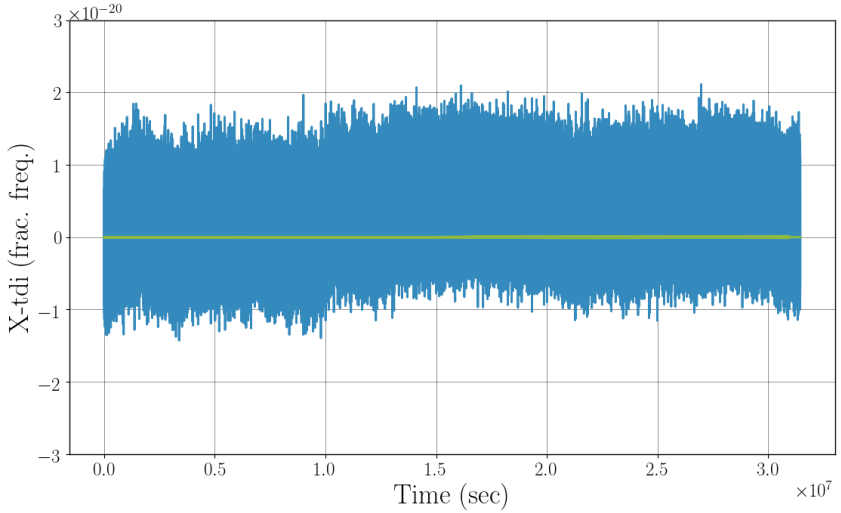


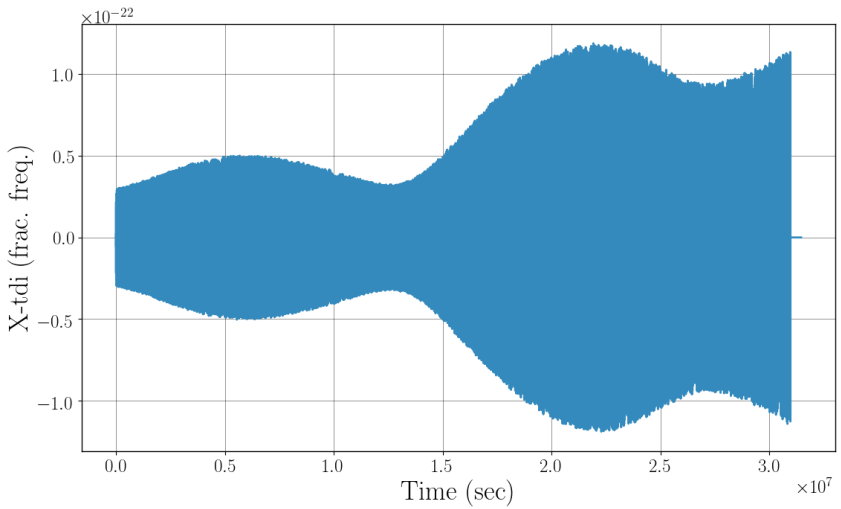
tk

PhaseAtCoalescence	1.0	Radian
PolarAngleOfSpin2	0.0	Radian
PolarAngleOfSpin1	0.0	Radian
Mass1	2057518.931	SolarMass
EclipticLongitude	4.4	Radian
CoalescenceTime	13864000.0	Second
Redshift	3.6111	dimensionless
Mass2	1440415.418	SolarMass
Approximant	PhenomD	ModelName
InitialAzimuthalAngleL	0.0	Radian
EclipticLatitude	0.7	Radian
AzimuthalAngleOfSpin1		Radian
AzimuthalAngleOfSpin2		Radian
Spin2	0.894516563931	MassSquared
Spin1	0.941100469122	MassSquared
Cadence	10.0	Seconds
InitialPolarAngleL	2.12	Radian
TDI	Y	name
armlength	2500000000.0	m
orbit	Analytic	name
Instrument	ESACallv1-2	name
LISACode_version	25.04.2016	string
Duration	20971520.0	sec
Model	Eccentric-Rigid-LISA_L3	
Cadence	10.0	sec

EMRI example



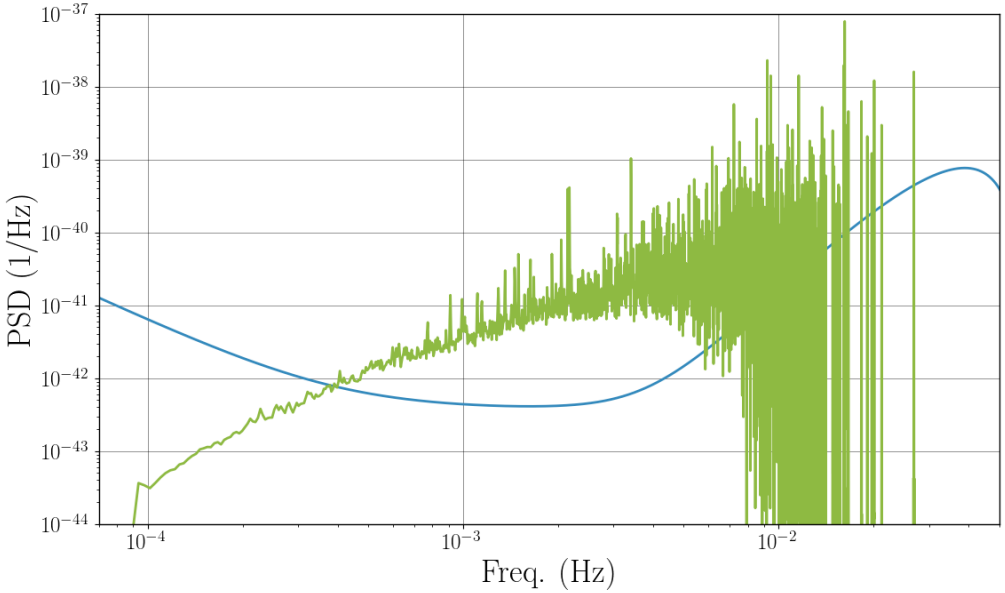


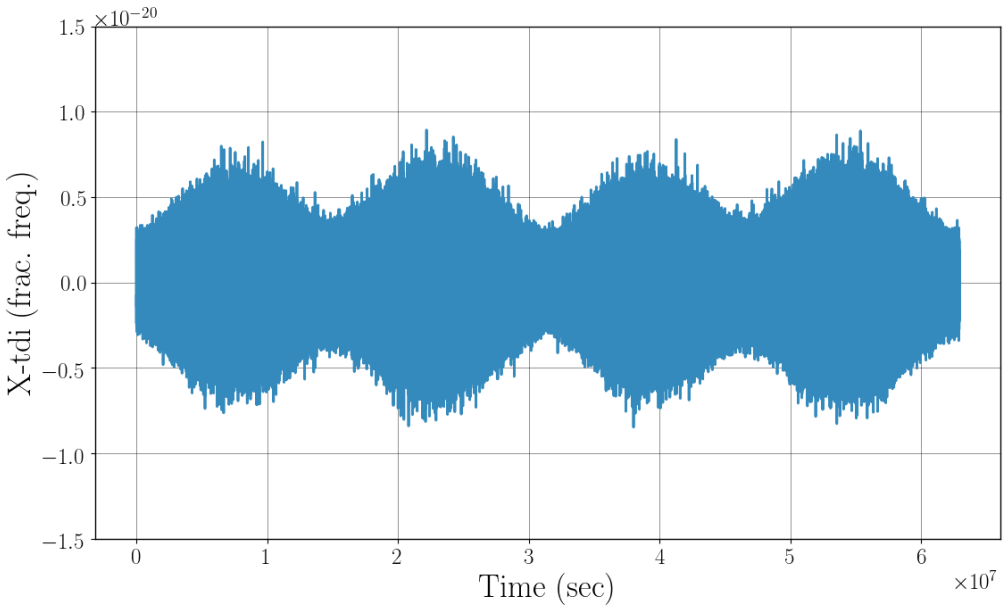


tk

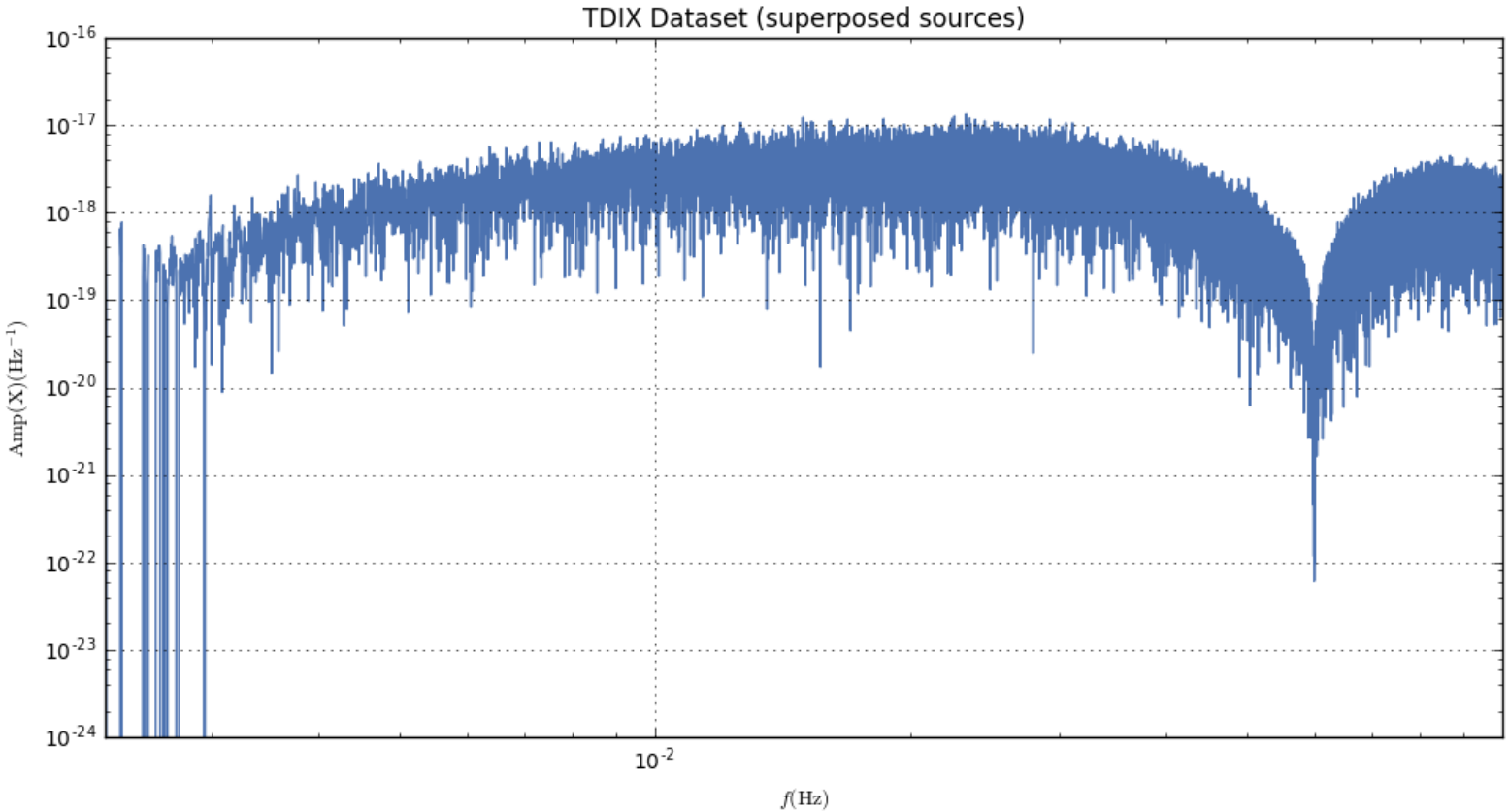
PlungeTime	30966857.3465	Second
FixHarmonics	5.0	1
PolarAngleOfSpin	0.767298246689	Radian
FinalAzimuthalOrbitalPhase	0.0698457861078	Radian
Polarization	0.0	Radian
InitialAzimuthalOrbitalPhase	4.05414493173	Radian
InitialTildeGamma	0.444268526709	Radian
MassOfSMBH	1011447.91955	solar_mass
InitialAzimuthalOrbitalFrequency	0.00108799353956	Hz
MassOfCompactObject	9.93422723681	solar_mass
LambdaAngle	1.85504384935	Radian
InitialEccentricity	0.292739358582	1
Spin	0.641082509452	mass_square
FinalEccentricity	0.168821357279	1
IntegrationStep	15.0	Second
Distance	1000000000.0	parsec
FinalAlphaAngle	0.717422837137	Radian
EclipticLongitude	4.11620153048	Radian
InitialAlphaAngle	1.67647060222	Radian
EclipticLatitude	-0.253027331648	Radian
FinalTildeGamma	2.50000775415	Radian
AzimuthalAngleOfSpin	0.0409959528889	Radian
TDI	X	name
armlength	2500000000.0	m
orbit	Analytic	name
Instrument	ESACallv1-2	name
LISACode_version	25.04.2016	string
Duration	31457280.0	sec
Model	Eccentric-Rigid-LISA_L3	
Cadence	15.0	sec

Galactic binaries example





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...