LISA Data Challenge: *Spritz*

N/Ref :	LISA-LCST-XXX-TN-001
Title	**LISA Data Challenge: *Spritz***
Abstract	Description of the data content for LISA Data Challenge (LDC)-2b.

	Name	**Date**	**Signature**
Prepared by	LDC WG	2025/05/14	
Checked by			
Checked by (QA)			
Approved by			

Document Change Record

Contributor List

Author's name	Institute	Location
Eleonora Castelli	Unversity of Trento	Trento
Stas Babak	APC	Paris
Maude Le Jeune	APC	Paris
Natalia Korsakova	Syrte	Paris
Quentin Baghi	CEA	Saclay

Ver.	Date	Author	Description	Pages
0.1	2021/11/06	Quentin Baghi (CEA)	Initial version	all
1.0	2022/02/09	Natalia Korsakova (APC)	Version for release	all
2.0	2025/05/14	Quentin Baghi (CEA)	Current Version	

Distribution list

Recipient	Restricted	Not restricted
LISA Consortium		×

Reference documents

Ref. No.	Document No.	Document Title	Date	Issue No.
RD1	ESA-L3-EST-SCI-RS-001	LISA Science Requirements Document	14/05/2018	1

Contents

Purpose and Scope

This documents aims at describing the content of the LISA Data Challenge LDC-2b dataset named *Spritz*.

1 Short description of LDC-2b *Spritz*

The purpose of this challenge is to address for the first time the realistic instrumental noise.The datasets are rather easy in the astrophysical content, moreover, the training data contain the sources that were used in the Sangria training dataset. If you are new to the Challenges, we strongly advise that you first complete the first data challenge *Radler* before moving to *Spritz*, specifically, challenges containing the merger of Massive Black Hole Binary (MBHB) systems and Verification Binary (VB) sources.

The training Spritz data contain three sets. All three datasets were generated with the extended pipeline used in the Sangria production. All Gravitational Wave (GW)s sources follow conventions described in the Sangria documentation and the same models (PhenomD for MBHB and Taylor-expanded model for Galactic Binaries (GBs)). Besides, we have varied the noise level (assuming the same acceleration and optical metrology noise in each spacecraft) within the same prior as in Sangria. Noise levels are given in the training data.

2 Description of GW sources

Gravitational wave sources contained in the Spritz challenge are GBs and MBHB, whose waveforms are described in the LDC-2a *Sangria* documentation [1].

3 Description of data artefacts

3.1 Glitches

Glitches injected in the Spritz data were generated with two different physically equivalent models. The two glitches present in the two MBHB datasets are events detected and fitted during the LISA Pathfinder (LPF) operations [2], with the so-called LPF-legacy model. A new treatment based on the phenomenological shapelet model has been recently introduced [3]. This new analysis allowed for the implementation of a generator of synthetic glitches, used to sample the glitches injected in the VB dataset.

Glitches in verification Galactic binaries dataset. The glitch is modeled as a finite linear combination of the shapelet functions

$$g(t) = \sum_{i=0}^{P-1} \alpha_i \psi_n \left(\frac{t - \tau_i}{\beta_i} \right), \tag{1}$$

where shapelet function is defined as

$$\psi_n(t) = c_n \frac{2t}{n} e^{-\frac{t}{n}} L^1_{n-1} \left(\frac{2t}{n} \right) h_+(t), \tag{2}$$

where L^1_{n-1} is the generalized Laguerre polynomial of order $n-1$, $c_n = (-1)^{n-1} n^{-\frac{3}{2}}$ is a normalizing factor ensuring that the quadratic sum of the components is equal to 1, and $h_+(t)$ is the Heaviside function which is equal to 1 for $t \geq 0$, 0 otherwise.

The glitches injected in the dataset are parametrized as single-component $n = 1$ shapelets in Eq. (1), and are integrated over time. The single-component shapelet glitch of Eq. (1) for $n = 1$ becomes

$$g(t) = 2\alpha \frac{t - \tau}{\beta} e^{\frac{-(t-\tau)}{\beta}} h_+ \left(\frac{t - \tau}{\beta} \right), \tag{3}$$

which is then integrated in time, resulting in

$$v_g(t) = 2\alpha\beta \left[1 - \left(1 + \frac{t - \tau}{\beta} \right) e^{\frac{-(t-\tau)}{\beta}} \right] h_+ \left(\frac{t - \tau}{\beta} \right). \tag{4}$$

The shapelet function is parameterized by damping time β which effectively correspond to the glitch duration, arrival time τ and glitch acceleration amplitude α. Further details on the model are available in Appendix B of Ref. [3]. In addition to that, this reference contain the information on the glitch description which was estimated from LPF data and the description of how one can sample the LPF-like glitches that will follow this distribution. The glitches were injected at random on different test mass interferometers `tm_ij`.

Caution!

Because of a typo in the code, the implementation of Eqs. (3), (4) in the Spritz generation code was instead

$$g(t) = 2\Delta v \frac{t-\tau}{\sqrt{\beta}} e^{\frac{-(t-\tau)}{\beta}} h_+\left(\frac{t-\tau}{\beta}\right), \qquad \Delta v = 2\alpha\beta \tag{5}$$

$$v_g(t) = \frac{2\Delta v}{\beta\sqrt{\beta}}\left[1 - \left(1 + \frac{t-\tau}{\beta}\right) e^{\frac{-(t-\tau)}{\beta}}\right] h_+\left(\frac{t-\tau}{\beta}\right), \qquad \Delta v = 2\alpha\beta, \tag{6}$$

where Δv is the glitch injection amplitude parameter (input parameter `level` in the code), in units of [m/s], and is equivalent to $2\alpha\beta$, in units of [m/s^2][s]. Please take this into account when analyzing Spritz data (generated with LISAGlitch version < 1.5). From LISAGlitch v.1.5 on, the shapelet models implemented in the code will be the corrected ones in Eqs. (3),(4), with $\alpha = \frac{\Delta v}{2\beta}$.

Glitches in MBHB datasets. For the MBHB datasets we have chosen to take the real glitches that appeared in LPF. One set contains the glitch which had highest amplitude and the other the most long lasting glitch. They were generated using the LPF legacy model $h(t)$, given by two double decaying exponentials

$$h(t) = \frac{\Delta v}{\tau_{\text{rise}} - \tau_{\text{fall}}}\left(e^{-\frac{t-t_0}{\tau_{\text{rise}}}} - e^{-\frac{t-t_0}{\tau_{\text{fall}}}}\right) h_+(t-t_0), \tag{7}$$

and integrating it in time

$$v_h(t) = \frac{\Delta v}{\tau_{\text{rise}} - \tau_{\text{fall}}}\left(1 - \tau_{\text{rise}} e^{-\frac{t-\tau}{\tau_{\text{rise}}}} - \tau_{\text{fall}} e^{-\frac{t-\tau}{\tau_{\text{fall}}}}\right) h_+(t-\tau), \tag{8}$$

where τ is the glitch arrival time, Δv is the impulse transferred by the glitch (the glitch injection amplitude), and $\tau_{\text{rise,fall}}$ are the time constants of the exponentials. The legacy models in Eq. (7),(8) are equivalent to the shapelet models Eq. (3),(4) for $\tau_{\text{rise}} \to \tau_{\text{fall}} = \beta$ and $\Delta v = 2\alpha\beta$. The injection point for glitches in both datasets is `tm_12`.

3.2 Data gaps

We introduce gaps in *Spritz* data to mock the measurement dropouts that will arise in LISA due to the periodic high-gain antenna repointing, an operation that ensures regular communication with Earth. We adopt a pattern of biweekly repointings lasting 7 hours each. While the gap durations are fixed, the interval between them ranges from 10 to 15 days. Gaps are identified in Time Delay Interferometry (TDI) data by NaN values.

4 Description of the pipeline

The data generation pipeline is made of several steps involving different tools developed by the LISA Consortium:

4.1 Generation of orbit

Using LISA Orbits[1] (v1.0.1), we generate an orbit file containing in particular time series of spacecraft positions, velocities, and inter-spacecraft light travel times computed according to a given orbit model, at a low sampling rate (below 1×10^{-4} Hz). We then upsample the time series to the simulator cadence (4 Hz), to generate a second orbit file to be used as input for LISANode.

For *Spritz*, we used the Keplerian orbits model, corresponding to nearly analytical solutions of the 2-body problem in Newtonian gravity, optimized to minimize the light travel time variations to second order in the orbital parameter. The model is based on [4] and is described in details in the LISA Orbit Simulation Model document[2].

4.2 Generation of GW signal and their projection on LISA arms (LDC Toolbox)

For each GW signal, we generate h_+, $h_\times$ and project the corresponding signal on LISA arms, using the LDC toolbox. The orbit file is used to get inter-spacecraft light travel times and spacecraft positions. Projected strains for each GW source are generated at low sampling rate (0.4 Hz) and are summed up and saved to a file to be used as input by LISANode.

4.3 Generation of glitches

Using LISA Glitch[3] (v1.0), we generate a glitch file at the simulator cadence (4 Hz) containing the total glitch time series for each injection point. The file is then used as input for LISANode.

4.4 Generation of L0 and TDI

We use LISANode[4] (v1.3) to generate noisy raw phasemeter data, and derive the TDI combinations from them. The physics is simulated at the same sampling rate 4 Hz as the telemetry data to limit the ratio between the lowest (GW signal) and highest (L0 data) sampling rates, which has a big impact on LISANode performance (especially on compilation memory and CPU time). For the VB dataset, we simulate a full year of data in a single shot. We use a subset of the possible instrumental effects or noise sources that can be simulated by LISANode, namely: laser, backlink, acceleration and readout noises.

Keeping the noise seeds fixed, we run LISANode for the following configurations: with and without noise, with and without glitches.

4.5 TDI downsampling

Each TDI time series output by LISANode is downsampled to 5-second cadence using a Kaiser filter, and shifted back to its original time origin.

[1] https://gitlab.in2p3.fr/lisa-simulation/orbits
[2] https://atrium.in2p3.fr/9dc7bc5d-4ba8-49f8-8b8f-aa33ae537eb9
[3] https://gitlab.in2p3.fr/lisa-simulation/glitch
[4] https://gitlab.in2p3.fr/j2b.bayle/LISANode

4.6 Gaps and non-stationary noise

Non-stationary noise is generated from a dedicated pipeline, which performs an iterative subtraction of all GBs signals above SNR = 7, in Fourier domain, from a population of 40 millions Galactic sources. The remaining signals are transformed into TDI combinations in the time domain and added to the downsampled TDI combinations from the previous step. We then replace part of the TDI X, Y, Z vectors by NaN values to simulate gaps. Each gap has a fixed duration of 7 hours. Starting times are randomly drawn between 10 and 15 days after the previous gap, and we iteratively add new gaps until we reach the end of the time series.

4.7 Dataset generation

Each TDI combination and all configuration files used for all the steps above are gathered into a single HDF5 file. We separately provide orbit and glitch files, as they follow a specific file format which can be read by the simulator.

The full pipeline is orchestrated by a workflow manager named Snakemake, which allows to apply the same sequence of "rules" to each dataset seamlessly and orchestrate the computation in an efficient way by saving recomputation and parallelizing the execution.

5 Description of datasets

5.1 Features common to all datasets

For the first time, we have used second-generation Michelson TDI X, Y, Z combinations, expressed as fractional frequency deviations, downsampled to 5-second cadence. We have used a Keplerian model for the LISA orbits. The data contain scheduled gaps of 7 hours duration each, distributed randomly with intervals between 10 and 15 days. We have included the non-stationary noise from the unresolved population of Galactic binaries (including all binaries with SNR < 7 with respect to the total noise budget). For the first time, we have included laser frequency noise, which is strongly suppressed to sub-dominant levels in the TDI combinations. In addition to the total data, the training datasets also contain partial versions of the data, which can be summed to retrieve the total signal: with and without noise, with and without artifacts (including gaps, glitches and non-stationary noise). Orbit and glitch files used as input to the simulator are given, too.

5.2 Massive black hole binary Spritz dataset

We have two datasets with merging MBHBs.

1. Dataset with a loud (Signal-to-Noise Ratio (SNR) of about 2000) GW signal, lasting for about 31 days. The signal is expected to be detectable a few weeks before the merger and, therefore, is suitable for testing low-latency algorithms. We have added three identical short loud glitches distributed in the inspiral, late inspiral and near merger parts of the signal.

   The glitch model is given in Sec. 3.1, with the following parameters:

Δv	τ_{rise}	τ_{fall}
2.20 pm/s	10 s	11 s

   ,

2. Dataset with a quiet (SNR 100) GW signal lasting for one week, with a several-hour-long glitch placed near the merger. Parameters of both MBHBs are available in the training data, as well as the information about glitches.

   In this dataset, the glitch parameters are

Δv	τ_{rise}	τ_{fall}
1.18 pm/s	5661.65 s	5661.71 s

   .

5.3 Verification GBs Spritz dataset

A one-year long dataset contains 36 verification binaries, with parameters available in the same data file. We have placed glitches according to a Poisson distribution with a rate of 4 glitches per day, whose model is described in Sec 3.1. The parameter data file includes one entry per each injected glitch, including the `generator` glitch class used for the injection, the `size` size of the simulation file, the `dt` sampling time of the simulation, the `t0` starting time of the simulation, the `t_inj` glitch injection time τ in seconds, the `inj_point` glitch injection point `tm_ij`, the `beta` glitch damping time β in seconds and the `level` glitch injection amplitude Δv in meters/second.

Acronyms and Glossary

GBs Galactic Binaries

GW Gravitational Wave

LPF LISA Pathfinder

MBHB Massive Black Hole Binary

SNR Signal-to-Noise Ratio

TDI Time Delay Interferometry

VB Verification Binary

References

[1] S. Babak, M. Le Jeune, A. Petiteau and M. Vallisneri. LISA Data Challenge: Sangria. *LISA Technical Note*, (LISA-LCST-SGS-MAN-001), 2020.

[2] M. Armano, H. Audley, J. Baird, P. Binetruy, M. Born, D. Bortoluzzi et al. Beyond the Required LISA Free-Fall Performance: New LISA Pathfinder Results down to 20 μHz. *Phys. Rev. Lett.*, 120:061101, Feb 2018.

[3] Q. Baghi, N. Korsakova, J. Slutsky, E. Castelli, N. Karnesis and J.B. Bayle. Detection and characterization of instrumental transients in LISA Pathfinder and their projection to LISA. *In preparation*, 2021.

[4] K.R. Nayak, S. Koshti, S.V. Dhurandhar and J.Y. Vinet. Reducing the flexing of the arms of LISA. , 2005.