



LISA Data Challenge: *Sangria*

N/Ref :	LISA-LCST-SGS-MAN-001
Title	LISA Data Challenge: <i>Sangria</i>
Abstract	Description of the data content for LISA Data Challenge (LDC)-2

	Name	Date	Signature
Prepared by	LDC WG	2020/10/12	
Checked by			
Checked by (QA)			
Approved by			



Document Change Record

Contributor List

Author's name	Institute	Location
Babak Stas	APC	Paris
Le Jeune Maude	APC	Paris
Petiteau Antoine	APC	Paris
Vallisneri Michele	JPL/Caltech	Pasadena

Ver.	Date	Author	Description	Pages
0.1	2020-10-08	S. Babak,	Initial version	all
0.1	2020/10/12	M. Le Jeune (APC)	Current Version	

Distribution list

Recipient	Restricted	Not restricted
LISA Consortium		X
LDC Participants		X



Contents

1	Acronyms and Glossary	4
1.1	Acronyms	4
1.2	Acronyms	4
1.3	Glossary	4
2	Purpose and Scope	4
3	Short description of LDC-2a <i>Sangria</i>	4
4	Description of GW sources	4
4.1	Convention for binary systems	4
4.1.1	Source frame	4
4.1.2	From the source frame to the SSB frame	6
4.1.3	Mode decomposition	7
4.1.4	Frequency domain	9
4.1.5	Link to the LAL conventions	9
4.2	MBHB: Massive black hole binaries	10
4.2.1	IMRPhenomD	10
4.3	EMRI	10
4.3.1	Analytic Kludge model	10
4.3.2	Augmented Analytic Kludge model	12
4.4	Galactic Binaries	12
4.4.1	The signal description in the source frame	12
5	Description of datasets	12
5.1	Training dataset	12
5.2	Blind dataset	13
6	Implementation	13
6.1	Link response	13
6.2	LISA orbits and travel times	14
6.3	LISA measurements and TDI	14
6.4	Software	15



1 Acronyms and Glossary

1.1 Acronyms

1.2 Acronyms

GBs Galactic Binaries

GW Gravitational Wave

LDC LISA Data Challenge

MBHB Massive Black Hole Binary

LISA-ScRD-004 Science Requirement Document

TDI Time Delay Interferometry

1.3 Glossary

2 Purpose and Scope

LDC is more than just fun, it is really critical for the LISA project!

3 Short description of LDC-2a *Sangria*

The dataset LDC-2a nicknamed *Sangria* is aiming at solving mild source confusion problem (sometimes also referred as "Enchilada"). The data set contains full simulated Galaxy (about 30mln white dwarf binaries) and a population of merging massive black hole binaries. The instrumental noise is simplified: we simulate Gaussian noise without gaps, however we allow deviation in the level of the noise from values adopted in the science requirement document.

We issue two datasets: (i) training data where all parameters are known, moreover we also provide noise less data sets for different sources (ii) blind challenge, which should be analyzed and the results to be submitted by deadline.

4 Description of GW sources

4.1 Convention for binary systems

4.1.1 Source frame

We start by introducing a source frame defined by unit vectors $(\mathbf{x}_S, \mathbf{y}_S, \mathbf{z}_S)$. The orientation of this source frame with respect to the binary system considered is left to the waveform model, as different choices might be preferable for different systems. An example is given by the NR convention of [8, 7], where the unit separation vector is $\mathbf{n} = \mathbf{x}_S$ and the normal to the orbital plane is $\hat{\mathbf{L}}_N = \mathbf{z}_S$ at the start of the waveform.

In this source frame, we introduce standard spherical coordinates (θ_S, ϕ_S) , and the associated spherical orthonormal basis vectors $(\mathbf{e}_r^S, \mathbf{e}_\theta^S, \mathbf{e}_\phi^S)$. The unit vector \mathbf{k} defines the direction of



propagation of the gravitational waves, from the source towards the observer. Expressed in spherical coordinates,

$$\mathbf{k} = \mathbf{e}_r^S = \{\sin \theta_S \cos \phi_S, \sin \theta_S \sin \phi_S, \cos \theta_S\}. \quad (1)$$

The explicit expressions for the other vectors of the spherical basis are

$$\mathbf{e}_\theta^S = \frac{\partial \mathbf{e}_r^S}{\partial \theta_S} = \{\cos \theta_S \cos \phi_S, \cos \theta_S \sin \phi_S, -\sin \theta_S\}, \quad (2a)$$

$$\mathbf{e}_\phi^S = \frac{1}{\sin \theta_S} \frac{\partial \mathbf{e}_r^S}{\partial \phi_S} = \{-\sin \phi_S, \cos \phi_S, 0\}. \quad (2b)$$

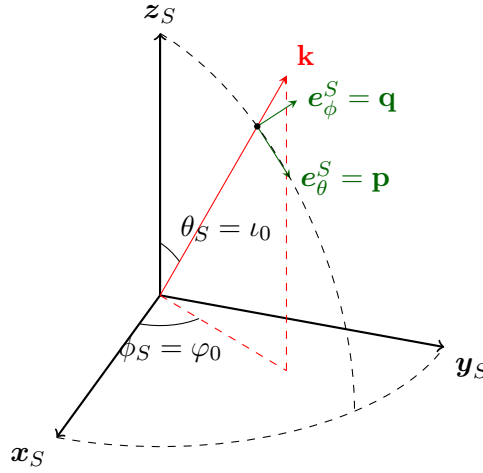


Figure 1: Source frame.

Next, introduce the polarization basis vectors:

$$\mathbf{p} = \mathbf{e}_\theta^S, \quad \mathbf{q} = \mathbf{e}_\phi^S \quad (3)$$

which form together with \mathbf{k} the radiation frame or wave-frame: $(\mathbf{p}, \mathbf{q}, \mathbf{k})$. They can be defined from \mathbf{z}_S only as

$$\mathbf{q} = \frac{\mathbf{z}_S \times \mathbf{k}}{|\mathbf{z}_S \times \mathbf{k}|}, \quad \mathbf{p} = \mathbf{q} \times \mathbf{k}. \quad (4)$$

The source frame and the polarization vectors are shown in the fig. 1. Defining the polarization tensors

$$\mathbf{e}_{ij}^+ = (\mathbf{p} \otimes \mathbf{p} - \mathbf{q} \otimes \mathbf{q})_{ij}, \quad \mathbf{e}_{ij}^\times = (\mathbf{p} \otimes \mathbf{q} + \mathbf{q} \otimes \mathbf{p})_{ij}, \quad (5)$$

the GW strain in transverse-traceless gauge takes the form

$$h_{ij}^{\text{TT}} = \mathbf{e}_{ij}^+ h_+ + \mathbf{e}_{ij}^\times h_\times. \quad (6)$$

This defines the polarizations h_+ and h_\times , functions of time. They are also given by the inverse relations

$$h_+ = \frac{1}{2} h_{ij}^{\text{TT}} \mathbf{e}_{ij}^+, \quad h_\times = \frac{1}{2} h_{ij}^{\text{TT}} \mathbf{e}_{ij}^\times. \quad (7)$$

The terminology used in the code tries to stay close to usual names and notation. The 0 subscript reminds us that these angles are constants.

$$\text{Inclination : } \iota_0 \equiv \theta_S, \quad (8)$$

$$\text{Observer phase : } \varphi_0 \equiv \phi_S. \quad (9)$$



4.1.2 From the source frame to the SSB frame

We now introduce as a detector frame a Solar System Barycenter (SSB) frame $(\mathbf{x}, \mathbf{y}, \mathbf{z})$, based on the ecliptic plane. Note that although we parallel here the geometric definitions of LAL [1], in their case this detector frame is a different frame, geocentric and based on the celestial equator.

Like for the source frame, we introduce standard spherical coordinates in the SSB-frame (θ, ϕ) , and the associated spherical orthonormal basis vectors $(\mathbf{e}_r, \mathbf{e}_\theta, \mathbf{e}_\phi)$. The position of the source in the sky will be parametrized by the ecliptic latitude $\beta = \pi/2 - \theta$ and the ecliptic longitude $\lambda = \phi$. The GW propagation vector \mathbf{k} in spherical coordinates is now

$$\mathbf{k} = -\mathbf{e}_r = \{-\cos \beta \cos \lambda, -\cos \beta \sin \lambda, -\sin \beta\}. \quad (10)$$

Again, the explicit expressions for the other vectors of the spherical basis are

$$\mathbf{e}_\theta = \{\sin \beta \cos \lambda, \sin \beta \sin \lambda, -\cos \beta\}, \quad (11a)$$

$$\mathbf{e}_\phi = \{-\sin \lambda, \cos \lambda, 0\}. \quad (11b)$$

Introduce reference polarization vectors as

$$\mathbf{u} = -\mathbf{e}_\phi = \{\sin \lambda, -\cos \lambda, 0\}, \quad (12a)$$

$$\mathbf{v} = -\mathbf{e}_\theta = \{-\sin \beta \cos \lambda, -\sin \beta \sin \lambda, \cos \beta\}, \quad (12b)$$

so that $(\mathbf{u}, \mathbf{v}, \mathbf{k})$ form a direct orthonormal triad. Equivalent expressions directly in terms of \mathbf{k} and \mathbf{z} are

$$\mathbf{u} = \frac{\mathbf{z} \times \mathbf{k}}{|\mathbf{z} \times \mathbf{k}|}, \quad \mathbf{v} = \mathbf{k} \times \mathbf{u}. \quad (13)$$

The SSB frame and its reference polarization vectors are given schematically in the fig 2.

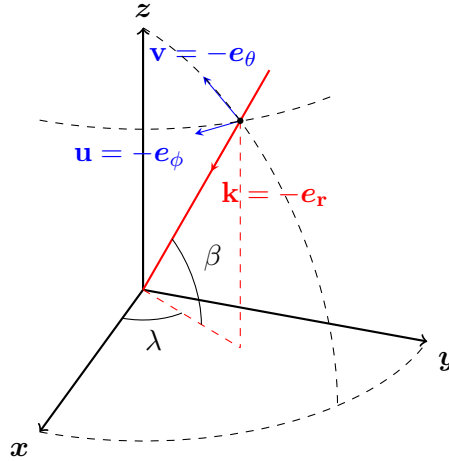


Figure 2: SSB frame

The last degree of freedom between the frames corresponds to a rotation around the line of sight, and is represented by the polarization angle ψ . We define the polarization to be the angle of the the rotation around \mathbf{k} that maps \mathbf{u} to \mathbf{p} (see figure 3):

$$\mathbf{p} = \mathbf{u} \cos \psi + \mathbf{v} \sin \psi, \quad (14a)$$

$$\mathbf{q} = -\mathbf{u} \sin \psi + \mathbf{v} \cos \psi. \quad (14b)$$

The polarization angle can be computed as¹

$$\psi = \arctan_2 [\mathbf{p} \cdot \mathbf{u}, \mathbf{p} \cdot \mathbf{v}]. \quad (15)$$

¹With the convention that $\arctan_2[x, y]$ is the polar angle of the point of coordinates (x, y) . Note that the function `numpy.arctan2` uses a reverse ordering $[y, x]$.



Given the direction of $z_S = \{\cos \phi_{z_S} \sin \theta_{z_S}, \sin \phi_{z_S} \sin \theta_{z_S}, \cos \theta_{z_S}\}$ in SSB we can compute the inclination and polarization as

$$\iota_o = \arccos [-\cos \theta_{z_S} \sin \beta - \cos \beta \sin \theta_{z_S} \cos (\lambda - \phi_{z_S})] \quad (16)$$

$$\tan \psi = \frac{-\sin \beta \sin \theta_{z_S} \cos (\lambda - \phi_{z_S}) + \cos \theta_{z_S} \cos \beta}{\sin \theta_{z_S} \sin (\lambda - \phi_{z_S})} \quad (17)$$

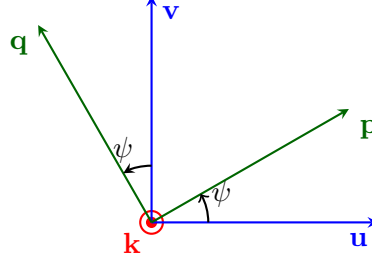


Figure 3: Polarization angle.

Defining as in (18) polarization basis tensors associated to (\mathbf{u}, \mathbf{v}) as

$$\epsilon_{ij}^+ = (\mathbf{u} \otimes \mathbf{u} - \mathbf{v} \otimes \mathbf{v})_{ij}, \quad \epsilon_{ij}^\times = (\mathbf{u} \otimes \mathbf{v} + \mathbf{v} \otimes \mathbf{u})_{ij}, \quad (18)$$

the relation between the polarization tensors is

$$\mathbf{e}^+ = \epsilon^+ \cos 2\psi + \epsilon^\times \sin 2\psi, \quad (19a)$$

$$\mathbf{e}^\times = -\epsilon^+ \sin 2\psi + \epsilon^\times \cos 2\psi. \quad (19b)$$

and corresponding representation of the strain in the SSB frame is

$$h_{ij}^{SSB} = (h_+ \cos 2\psi - h_\times \sin 2\psi) \epsilon_{ij}^+ + (h_+ \sin 2\psi + h_\times \cos 2\psi) \epsilon_{ij}^\times. \quad (20)$$

The terminology used in the code is:

Ecliptic longitude : $\lambda \equiv \phi$,	(21)
Ecliptic latitude : $\beta \equiv \pi/2 - \theta$,	(22)
Polarization : ψ .	(23)

4.1.3 Mode decomposition

Those two polarizations can be decomposed in the spin-weighted (-2) spherical harmonics:

$$h_+(t) - ih_\times(t) = \sum_{\ell=2}^{+\infty} \sum_{m=-\ell}^{\ell} h_{\ell m}(t) {}_{-2}Y_{\ell m}(\iota_0, \varphi_0). \quad (24)$$

Our convention for the spin-weighted spherical harmonics is the same as the one used in LAL [1], in [2] and [5]:

$${}_{-2}Y_{\ell m}(\iota_0, \varphi_0) = \sqrt{\frac{2\ell+1}{4\pi}} d_{m,2}^\ell(\iota_0) e^{im\varphi_0}, \quad (25a)$$

$$d_{m,2}^\ell(\iota_0) = \sum_{k=k_1}^{k_2} \frac{(-1)^k}{k!} \frac{\sqrt{(\ell+m)!(\ell-m)!(\ell+2)!(\ell-2)!}}{(k-m+2)!(\ell+m-k)!(\ell-k-2)!} \left(\cos \frac{\iota_0}{2}\right)^{2\ell+m-2k-2} \left(\sin \frac{\iota_0}{2}\right)^{2k-m+2}, \quad (25b)$$



with $k_1 = \max(0, m - 2)$ and $k_2 = \min(\ell + m, \ell - 2)$. The polarizations can be expressed as

$$h_+ = \frac{1}{2} \sum_{\ell, m} (-2Y_{\ell m} h_{\ell, m} + -2Y_{\ell, -m}^* h_{\ell, m}^*) , \quad (26a)$$

$$h_\times = \frac{i}{2} \sum_{\ell, m} (-2Y_{\ell m} h_{\ell, m} - -2Y_{\ell, -m}^* h_{\ell, m}^*) , \quad (26b)$$

Note that our polarization vectors (4) differ from the PN convention of [5] by a rotation of $\pi/2$, which translates into an overall factor (-1) in the polarizations $h_{+, \times}$ and in the modes $h_{\ell m}$.

For non-precessing binary systems, with a fixed equatorial plane of orbit, an exact symmetry relation between modes holds:

$$h_{\ell, -m} = (-1)^\ell h_{\ell m}^* . \quad (27)$$

When this symmetry is verified, we can write

$$h_{+, \times} = \sum_{\ell, m} K_{\ell m}^{+, \times} h_{\ell m} , \quad (28)$$

with

$$K_{\ell m}^+ = \frac{1}{2} \left(-2Y_{\ell m} + (-1)^\ell -2Y_{\ell, -m}^* \right) , \quad (29a)$$

$$K_{\ell m}^\times = \frac{i}{2} \left(-2Y_{\ell m} - (-1)^\ell -2Y_{\ell, -m}^* \right) . \quad (29b)$$

As an example of the connection between modes and polarizations, let us consider the emission from a non-precessing binary system on a circular orbit, at the leading post-Newtonian order. The source frame is such that $\mathbf{z}_S = \hat{\mathbf{L}}_N$ is the normal to the orbital plane, and Φ denotes the phase of the binary separation vector $\mathbf{y}_1 - \mathbf{y}_2 = r\mathbf{n}$ (with $\mathbf{y}_{1,2}$ the position of the two bodies) in the plane $(\mathbf{x}_S, \mathbf{y}_S)$. We also use the orbital frequency $\omega = \dot{\Phi}$.

At leading PN order, only the dominant $\{\ell, m\} = \{2, \pm 2\}$ mode enters. From Sec. (9.5) of [5], taking into account the sign change caused by the difference in the definition of \mathbf{p}, \mathbf{q} ,

$$h_{22} = -\frac{2Gm\nu x}{Rc^2} \sqrt{\frac{16\pi}{5}} e^{-2i\Phi} , \quad (30a)$$

$$h_{2, -2} = h_{22}^* , \quad (30b)$$

where we introduced $m = m_1 + m_2$ the total mass, $\nu = m_1 m_2 / m^2$ the symmetric mass ratio, $x = (Gm\omega/c^3)^{2/3}$ is the PN parameter and R is the distance to the source. The expressions of the spin-weighted spherical harmonics (25a) for $\{\ell, m\} = \{2, \pm 2\}$ are

$$-2Y_{22}(\iota_0, \varphi_0) = \frac{1}{2} \sqrt{\frac{5}{\pi}} \cos^4 \frac{\iota_0}{2} e^{2i\varphi_0} , \quad (31a)$$

$$-2Y_{2, -2}(\iota_0, \varphi_0) = \frac{1}{2} \sqrt{\frac{5}{\pi}} \sin^4 \frac{\iota_0}{2} e^{-2i\varphi_0} . \quad (31b)$$

Using the combinations (29),

$$\begin{aligned} h_+ &= K_{22}^+ h_{22} + K_{2, -2}^+ h_{2, -2} \\ &= -\frac{4Gm\nu x}{Rc^2} \frac{1 + \cos^2 \iota_0}{2} \cos(2(\Phi - \varphi_0)) , \end{aligned} \quad (32a)$$

$$\begin{aligned} h_\times &= K_{22}^\times h_{22} + K_{2, -2}^\times h_{2, -2} \\ &= -\frac{4Gm\nu x}{Rc^2} \cos \iota_0 \sin(2(\Phi - \varphi_0)) . \end{aligned} \quad (32b)$$



4.1.4 Frequency domain

For the definitions of the Fourier transform (FT), we use the same convention as in [1], used also `numpy` (and up to normalization in `FFTW`): for a function of time $F(t)$

$$\tilde{F}(f) = \int dt F(t)e^{-2i\pi ft}, \quad F(t) = \int df \tilde{f}(f)e^{2i\pi ft}. \quad (33)$$

The discrete Fourier transform for N time samples F_j and frequency samples \tilde{F}_k is then

$$\tilde{F}_k = \sum_{j=0}^{N-1} F_j e^{-2i\pi \frac{jk}{N}}, \quad F_j = \frac{1}{N} \sum_{k=0}^{N-1} \tilde{F}_k e^{2i\pi \frac{jk}{N}}. \quad (34)$$

For a real time-domain signal,

$$F(t) \in \mathbb{R} \implies \tilde{F}(-f) = \tilde{F}(f)^*, \quad (35)$$

and we recall the following useful relation valid in general:

$$\tilde{F}^*(f) = \tilde{F}(-f)^*. \quad (36)$$

For non-precussing systems, (27) implies

$$\tilde{h}_{\ell, -m}(f) = (-1)^\ell \tilde{h}_{\ell m}(-f)^*. \quad (37)$$

Since a given mode has a phase dependency $h_{\ell m} \propto \exp[-im\phi_{\text{orb}}]$, with the orbital phase verifying $\dot{\phi}_{\text{orb}} > 0$, an approximation often used for non-precussing systems (or in the precussing frame for a binary with misaligned spins) is

$$\tilde{h}_{\ell m}(f) \simeq 0 \text{ for } m > 0, f > 0, \quad (38a)$$

$$\tilde{h}_{\ell m}(f) \simeq 0 \text{ for } m < 0, f < 0, \quad (38b)$$

$$\tilde{h}_{\ell 0}(f) \simeq 0. \quad (38c)$$

Note that in the Fourier convention (33), this approximation means that for positive frequencies $f > 0$ the mode $\tilde{h}_{2, -2}(f)$ has support while $\tilde{h}_{22}(f)$ is negligible². When using the approximation (38), (39) becomes for $f > 0$

$$\tilde{h}_{+, \times}(f) = \sum_{\ell} \sum_{m < 0} K_{\ell m}^{+, \times} \tilde{h}_{\ell m}. \quad (39)$$

4.1.5 Link to the LAL conventions

In conclusion of this section we explain of our conventions relate to those used in LAL (software library used by LIGO/Virgo collaborations [1]), which are standardized in the documents [8, 7]. First, in LAL the source frame is constructed by convention [8] setting the unit separation vector $\mathbf{n} = \mathbf{x}_S$ and the normal to the orbital plane $\hat{\mathbf{L}}_N = \mathbf{z}_S$, at the start of the waveform. In our case we have not specified how this source-frame is built, as this construction might differ for different physical systems (SMBHs, EMRIs, ...).

Translated in the notations we used above, the choice of polarization vectors for LAL is

$$\mathbf{p}_{\text{LAL}} = -\mathbf{p}, \quad (40a)$$

$$\mathbf{q}_{\text{LAL}} = -\mathbf{q}, \quad (40b)$$

²Note that because of this, some parts of the LDC code internally use the opposite sign convention in the exponentials of (33), so as to have support for the modes $m > 0$ for $f > 0$. This amounts to a mapping $f \leftrightarrow -f$, which can be undone at the end of the computation by conjugating the FT of the observables, which are real signals (see (35)).



which amounts to a rotation of π of the polarization basis. Because the polarizations $h_{+, \times}$ transform with $\cos 2\psi, \sin 2\psi$ under a change of polarization, this means that there is no difference here, provided the source-frame is identical:

$$h_+^{\text{LAL}} = h_+, \quad (41a)$$

$$h_\times^{\text{LAL}} = h_\times. \quad (41b)$$

The phase quantity used in LAL differs from our definition by

$$\Phi_{\text{LAL}} = \frac{\pi}{2} - \varphi_0. \quad (42)$$

Although the geometric definition of the reference polarization vectors \mathbf{u}, \mathbf{v} is the same, the detector frame $(\mathbf{x}, \mathbf{y}, \mathbf{z})$ is different. In the case of LAL, it is a geocentric frame based on the celestial equator, while in our case it is the SSB frame based on the ecliptic plane. For the polarization, a difference of π in the definition also comes from (40):

$$\psi_{\text{LAL}} \leftrightarrow \psi : \text{diff. of } \pi \text{ and different frame.} \quad (43)$$

For the sky position, LAL uses ra, dec instead of our ecliptic λ, β :

$$(\alpha_{\text{LAL}}, \beta_{\text{LAL}}) \leftrightarrow (\lambda, \beta) : \text{different frame.} \quad (44)$$

We leave for future work this exact map, relating the SSB frame to the geocentric frame.

4.2 MBHB: Massive black hole binaries

4.2.1 IMRPhenomD

IMRPhenomD was introduced in [6] and we will not repeat here its description. Instead we describe the parametrization used in the generated data. We have assume the merger as a reference point to set the initial phase and the time of coalescence. The particular implementation can be found in <https://gitlab.in2p3.fr/LISA/LDC/-/tree/master/ldc/waveform/imrphenomD/>. We generate h_+ and h_\times for each source which we project on the links (see below) and use LISANode software package to produce Time Delay Interferometry (TDI)s: those are Michelson-like channels X, Y, Z generated in the time domain with cadence 5 sec.

Parameters for the source together with units are stored in the training hdf5 file. In the table 1 we give parameters describing the MBHB signal.

Let us also give some necessary explanations. We create those parameters to cover also precessing binaries (eccentric binaries will require an extension of this table), therefore some parameters (ϕ_{S1}, ϕ_{S2}) are redundant. The spin used for the IMRPhenomD model are computed as $a_i = s_i \cos \theta_{Si}$. Masses quoted in the hdf5 files are always redshifted masses. The time of coalescence roughly corresponds to the merger of two BHs.

We have given redundant information: redshift z and the luminosity distance D_L are connected by adopted cosmological model (based on Planck2015 parameters).

4.3 EMRI

EMRIs **DO NOT** enter the Sangria dataset, however following the popular request we re-issue the Radler-EMRI data.

4.3.1 Analytic Kludge model

The AK model used in the Radler challenge is based on [4] for generation of h_+, h_\times in the SSB frame. The major modification is that we fix the number of harmonics (similar to how it was done in the old MLDC [3]). The TDI response is applied using LISACode.



Parameter	Description	units
β	EclipticLatitude	Radian
λ	EclipticLongitude	Radian
θ_{S1}	PolarAngleOfSpin1	Radian
θ_{S2}	PolarAngleOfSpin2	Radian
ϕ_{S1}	AzimuthalAngleOfSpin1	Radian
ϕ_{S2}	AzimuthalAngleOfSpin2	Radian
s_1	magnitude Spin1	MassSquared
s_2	magnitude Spin2	MassSquared
m_1	Mass1	SolarMass
m_2	Mass2	SolarMass
t_c	CoalescenceTime	Second
ϕ_{ref}	PhaseAtCoalescence	Radian
θ_L	InitialPolarAngleL	Radian
ϕ_L	InitialAzimuthalAngleL	Radian
-	Approximant	ModelName
z	Redshift	dimensionless
D_L	Distance (luminosity)	Gpc

Table 1: Parametrization used for MBHBs.

Description	Parameter	Notation	units
Sky position (SSB)	β	EclipticLatitude	Radian
Sky position (SSB)	λ	EclipticLongitude	Radian
Mass of SMBH	M	MassOfSMBH	SolarMass
Mass of compact object	μ	MassOfCompactObject	SolarMass
SMBH spin	S	SMBHspin	MassSquared
SMBH spin orient. (in SSB)	θ_K, ϕ_K	PolarAngleOfSpin	Radian
Radial orb. freq. ($t = 0$)	ν_0	InitialAzimuthalOrbitalFrequency	Hertz
Orb. mean anom. ($t = 0$)	Φ_0	InitialAzimuthalOrbitalPhase	Radian
Eccentricity ($t = 0$)	e_0	InitialEccentricity	1
Dir. of pericenter ($t = 0$)	$\tilde{\gamma}_0$	InitialTildeGamma	Radian
Azimuthal angle of orb. ($t = 0$)	α_0	InitialAlphaAngle	Radian
Inclination of orbit	Λ	LambdaAngle	Radian
Luminosity distance	D_L	Distance	Gpc
time of plunge	t_{pl}	PlungeTime	Second
-	Approximant	AK	ModelName

Table 2: Parametrization used for EMRI (AK).



We summarize in the table 2 the waveform parametrization.

One remark, the plunge time and ν_0 are related, so either of them could be used to parametrize the waveform.

4.3.2 Augmented Analytic Kludge model

4.4 Galactic Binaries

4.4.1 The signal description in the source frame

The GW signal from Galactic white dwarf binaries in the source frame is defined as:

$$h_+^S = \mathcal{A}(1 + \cos^2 \iota) \cos(\Phi(t)) \quad (45)$$

$$h_\times^S = 2\mathcal{A} \cos \iota \sin(\Phi(t)) \quad (46)$$

The phase is described as

$$\Phi(t) = -\phi_0 + 2\pi f_0 t + \pi \dot{f}_0 t^2, \quad (47)$$

where we have used the fact that the frequency evolves slowly and we retained only the first derivative of frequency. The subscript 0 implies that those are constants and given at the initial moment of time $t = 0$, and ϕ_0 is initial GW phase. The expression for the GW strain in the SSB frame is described in the section 4.1.

In the table 3 we summarize the parametrization of the GB signal.

Parameter	Notation	units
β	EclipticLatitude	Radian
λ	EclipticLongitude	Radian
\mathcal{A}	Amplitude	strain
f_0	Frequency	Hz
\dot{f}_0	FrequencyDerivative	Hz^2
ι	Inclination	Radian
ψ	Polarization	Radian
ϕ_0	InitialPhase	Radian
T_{obs}	ObservationDuration	Seconds
Δt	Cadence	Seconds

Table 3: Parametrization used for GB signal.

5 Description of datasets

As we have already mentioned that we issue training data set (all parameters used in the data generation are known) and blind data set. In the following we will give more details on each data set. Note that all data is generated in fractional frequency.

5.1 Training dataset

The training dataset is provided to (i) check the waveform generator, to make sure that you agree with the signals embedded in the data (ii) to train the data analysis algorithm before jumping on the blind dataset.



The instrumental noise was simulated as stationary Gaussian. Its level is given in the hdf5 file (in terms of the (elementary noise components):

$$S_X = 16 \sin^2 \omega L [2(1 + \cos^2 \omega L)S^{acc} + S^{OMS}]. \quad (48)$$

Note that the noise is correlated between X, Y, Z with the cross-spectrum:

$$S_{XY} = -8 \sin^2 \omega L \cos \omega L (S^{OMS} + 4S^{acc}) \quad (49)$$

where the values used in the Science Requirement Document (LISA-ScRD-004):

$$\sqrt{S_{d\nu/\nu}^{OMS}(f)} = 15 \times 10^{-12} \frac{2\pi f}{c} \left[\frac{1}{\sqrt{\text{Hz}}} \right] \sqrt{1 + \left(\frac{2 \times 10^{-3}}{f} \right)^4} \quad (50)$$

$$\sqrt{S_{d\nu/\nu}^{acc}(f)} = \frac{3 \times 10^{-15}}{2\pi f c} \left[\frac{1}{\sqrt{\text{Hz}}} \right] \sqrt{1 + \left(\frac{0.4 \times 10^{-3}}{f} \right)^2} \sqrt{1 + \left(\frac{f}{8 \times 10^{-3}} \right)^4}. \quad (51)$$

We have used the variations in noise noise about that level S^{OMS} (-20%) and for S^{acc} (-25%).

Note that we can construct the noise independent data combinations, called here A, E, T . Note that these combinations are not uniquely defined as

$$E = \frac{X - 2Y + Z}{\sqrt{6}}, \quad A = \frac{Z - X}{\sqrt{2}}, \quad T = \frac{X + Y + Z}{\sqrt{3}} \quad (52)$$

As mentioned the data contains about 30mln Galactic Binaries (GBs). On top of those signals we have added 17 known verification binaries. For this source we have also provide the noiseless datasets for: (i) only detached GBs, evolving under gravitational radiation reaction (ii) only interacting binaries, those are the binaries with the *steady* mass transfer, note that some of them could be outspiralng (iii) 17 verification binaries.

Parameters of each Massive Black Hole Binary (MBHB) are given in the hdf5 file. They were drawn randomly from one of the astrophysical catalogues provided by AstroWG. It happens that the training data set contains 15 MBHBs. We provide the noiseless data set which contains all those binaries together. Most of the time they do not have a strong overlap.

5.2 Blind dataset

The blind data set was generated in an identical way to the training data set. Of course the parameters and number of sources (besides verification binaries) were randomized. In addition, we have randomized each level of noise as 20% around the mean values $S^{OMS} = 10 \times 10^{-12}$ and $S^{acc} = 2.4 \times 10^{-15}$ replacing the numerical factors in from of the eqns. 50, 51. We can only guarantee that the number of MBHBs is less than 50 (for purposes of keeping computational cost reasonable).

6 Implementation

6.1 Link response

For each source we have generated h_+ and h_\times in the time domain. Before applying TDI we have projected Gravitational Wave (GW)s on the detector's links.

The response of each link (single arm) to the GW signal is given as

$$y_{slr}^{GW} = \frac{\Phi_l(t_s - \mathbf{k}\mathbf{R}_s(t_s)) - \Phi_l(t - \mathbf{k}\mathbf{R}_r(t))}{2(1 - \mathbf{k}\mathbf{n}_l)} \quad (53)$$



where the subscripts slr mean “s”ender, “l”ink, “r”eciever and

$$\Phi_l = \mathbf{n}_l h_{ij} \mathbf{n}_l \quad (54)$$

is the projection of the GW strain on the link’s unit vector (\mathbf{n}_l). The time t_s could be approximated as follows $t_s = t - |\mathbf{R}_r(t) - \mathbf{R}_s(t_s)| \approx t - L_l$, and $L_l \mathbf{n}_l \approx \mathbf{R}_r(t) - \mathbf{R}_s(t_s)$. The vectors \mathbf{R}_i define the position of “i”-th spacecraft.

We have computed 6 projections y_{slr}^{GW} for each source in parallel using up to 1000 cores and then summed them up. Then the noise was added to each link and TDIs (Michelson-like combinations X, Y, Z) were produced using LISANode.

6.2 LISA orbits and travel times

We used simple equal arm analytic orbits. The orbital motion for each LISA spacecraft:

$$x_n = a \cos \alpha + ae \left(\sin \alpha \cos \alpha \sin \beta_n - (1 + \sin^2 \alpha) \cos \beta_n \right) \quad (55)$$

$$y_n = a \sin \alpha + ae \left(\sin \alpha \cos \alpha \cos \beta_n - (1 + \cos^2 \alpha) \sin \beta_n \right) \quad (56)$$

$$z_n = -\sqrt{3}ae \cos(\alpha - \beta_n) \quad (57)$$

where those are the coordinates in the solar system barycentric (SSB) frame and the phases are

$$\beta_n = (n - 1) \frac{2\pi}{3} + \lambda \quad (58)$$

$$\alpha(t) = \frac{2\pi}{1 \text{ year}} t + \kappa \quad (59)$$

and $\kappa = 0$, $\lambda = 0$ define the initial conditions. The parameter a is equal to 1 AU (Astronomical Unit). The orbital eccentricity is computed based on the armlength: $e = L/(2a\sqrt{3})$.

The velocity is given by:

$$\dot{x}_n = -ab \sin \alpha + abe \left((\cos \alpha^2 - \sin \alpha^2) \sin \beta_n - (2 \cos \alpha \sin \alpha \cos \beta_n) \right) \quad (60)$$

$$\dot{y}_n = ab \cos \alpha + abe \left((\cos \alpha^2 - \sin \alpha^2) \cos \beta_n - (2 \cos \alpha \sin \alpha \sin \beta_n) \right) \quad (61)$$

$$\dot{z}_n = \sqrt{3}abe \sin(\alpha - \beta_n) \quad (62)$$

The parameter b is equal to $\frac{2\pi}{1 \text{ year}}$.

We used half order travel time equation (or first order following LISACode convention):

$$\tau_{AB} = \tau_{AB}^0 + \tau_{AB}^{(\frac{1}{2})} \quad (63)$$

$$\tau_{AB}^0 = \frac{\|\mathbf{r}_{AB}\|}{c} \quad (64)$$

$$\tau_{AB}^{(\frac{1}{2})} = \tau_{AB}^0 \frac{\mathbf{v}_B \hat{\mathbf{n}}}{c} \quad (65)$$

where $\mathbf{R}_A, \mathbf{R}_B$ are positions of spacecraft A and B respectively, \mathbf{v}_B is the velocity of spacecraft B, \mathbf{r}_{AB} is the separation vector, and $\hat{\mathbf{n}} = \frac{\mathbf{r}_{AB}}{\|\mathbf{r}_{AB}\|}$ the link unit vector.

6.3 LISA measurements and TDI

We have used the following convention for X -tdi: The $X_{1.5}$ tdi is given as

$$X_{1.5}^{GW} = D_3 D_2 D_{2'} y_{1-32} + D_{2'} D_2 y_{231} + D_{2'} y_{123} + y_{32'1} - [D_3 D_{3'} D_{2'} y_{123} + D_3 D_{3'} y_{32'1} + D_3 y_{13'2,3} + y_{231}], \quad (66)$$



where we have used the delay operator $D_{2'}y = y(t - L_{2'})$.

We have used the LISA instrument model and TDI engine provided by the LISANode simulator, in its \sim v1.1 version. As the software is in an active development phase, we have incorporated all changes which happened until March 2020, then freeze it in a dedicated branch. A complete description of the LISANode simulation model is in preparation, and reference will be given as soon as it is released by the simulation WG.

We have made additional configuration changes with respect to the baseline LISA instrument model:

- We have used equal *physical* sampling frequency and *measurement* sampling frequency (3Hz), and deactivated the associated anti aliasing decimation step;
- We have added the GW signal y_{slr} to the list of signals propagated through LISA links, including laser, laser frequency and time (no side-band);
- We have set custom values for acceleration, readout and optical noise amplitude, and deactivated the laser, modulation, USO and ranging noises;
- We have used the same orbits and travel times computation as the one used to get the y_{slr} ;
- We have used an additional post processing elliptic filter to down-sample the TDI signal at the LDC frequency sampling (0.2Hz).

6.4 Software

The code and pipeline used to generate the datasets are released as the first version of the new LDC toolbox, accessible in the LISA Consortium GitLab: <https://gitlab.in2p3.fr/LISA/LDC>

This Python toolbox will be the foundation for a set of LISA common tools, to be used to generate and analyze the data (including physical constants, orbits, I/O utilities, waveform generation, noise characterization, and so on).

The data-generation pipeline https://gitlab.in2p3.fr/LISA/LDC/data_generation is built on top of the toolbox, using the Snakemake workflow manager to orchestrate the multiple steps needed to build the dataset. Documentation (<https://lisa.pages.in2p3.fr/LDC/>) is provided so that anyone may reproduce the dataset, or generate custom datasets for other studies. The `singularity` image used to build the data are available on the GitLab registry, for traceability and reproducibility purpose (tag `sangria-prod-v1-0`).

Notebook tutorials are also provided to help gain familiarity with the LDC toolbox and data format.

Note that the LISANode software repository is only available to LISANode software project members for now. As stated in the LISANode readme: "LISANode can only be shared with internal LISA collaborators for LISA project-oriented studies. LISA members involved in such a project should inform the Simulation Working Group leaders of their project. If the work leads to an external publication, the main LISANode developers should be at least appropriately cited in the acknowledgment section of the paper, as well as the Simulation Working Group and the LISA project."

References

- [1] LSC algorithm library. <http://www.lsc-group.phys.uwm.edu/lal>.
- [2] P. Ajith, M. Boyle, D. A. Brown, S. Fairhurst, M. Hannam, I. Hinder, S. Husa, B. Krishnan, R. A. Mercer, F. Ohme, C. D. Ott, J. S. Read, L. Santamaria, and J. T. Whelan. Data



formats for numerical relativity waves. *ArXiv e-prints*, page arXiv:0709.0093, September 2007.

- [3] K. A. Arnaud et al. An Overview of the second round of the Mock LISA Data Challenges. *Class. Quant. Grav.*, 24:S551–S564, 2007.
- [4] Leor Barack and Curt Cutler. LISA capture sources: Approximate waveforms, signal-to-noise ratios, and parameter estimation accuracy. *Phys. Rev.*, D69:082005, 2004.
- [5] Luc Blanchet. Gravitational Radiation from Post-Newtonian Sources and Inspiralling Compact Binaries. *Living Rev. Rel.*, 17:2, 2014.
- [6] Sebastian Khan, Sascha Husa, Mark Hannam, Frank Ohme, Michael Pürrer, Xisco Jiménez Forteza, and Alejandro Bohé. Frequency-domain gravitational waves from non-precessing black-hole binaries. II. A phenomenological model for the advanced detector era. *Phys. Rev.*, D93(4):044007, 2016.
- [7] Harald Pfeiffer. Overview lal gravitational wave frame definitions. Technical Report LIGO-T1800226, LIGO Scientific Collaboration and Virgo Collaboration, 2018.
- [8] P. Schmidt, I. W. Harry, and H. P. Pfeiffer. Numerical Relativity Injection Infrastructure. *ArXiv e-prints*, March 2017.