

A How-To for the Mock LISA Data Challenges

*The Mock LISA Data Challenge Task Force*¹

Abstract. The LISA International Science Team Working Group on Data Analysis (LIST- WG1B) is sponsoring several rounds of mock data challenges, with the purpose of fostering development of LISA data-analysis capabilities, and of demonstrating technical readiness for the maximum science exploitation of the LISA data. The first round of challenge data sets were released at this Symposium. We describe the models and conventions (for LISA and for gravitational-wave sources) used to prepare the data sets, the file format used to encode them, and the tools and resources available to support challenge participants.

Keywords: gravitational waves, LISA, data analysis, galactic binaries, black holes, XML

PACS: 04.80.Nn, 07.60.Ly, 95.55.Ym

The objectives, structure, and timeline of the Mock LISA Data Challenges (MLDCs) are discussed in the other contribution in this volume by the MLDC Task Force. Here we concentrate on the technical side of the challenges, and in particular on the theoretical models used to embody LISA and gravitational-wave (GW) sources in the first Challenge, and on the file format used to distribute the Challenge datasets. More details can be found on the official MLDC website [1], in the living *Omnibus* document for Challenge 1 [2], and on the MLDC Task Force wiki [3].

1. MODELING LISA

The analysis of real LISA data will necessarily involve a detailed modeling of instrument response and noise; for the purposes of the MLDCs, it is desirable to decouple this aspect from the inherent complexity of GW analysis, by distilling the LISA measurements into a standard idealized model. Thus, the MLDC task force has developed a set of *pseudo-LISA* assumptions and conventions, which we lay out in this section (see also Ref. [2]). Both of the LISA response simulators used in Challenge 1 (the LISA Simulator [4] and Synthetic LISA [5]) comply with these assumptions, and adhere to these conventions. In later challenges, as the craft of LISA data-analysis matures, so will the pseudo-LISA model, becoming increasingly realistic.

1.1. The pseudo-LISA orbits

The pseudo-LISA orbits are obtained by truncating exact Keplerian orbits for a small mass orbiting the Sun to first order in the eccentricity (see the Appendix of Ref. [4]).

¹ See the preceding contribution by the Task Force in this volume for the full author list, as well as acknowledgments.

In Solar-System Barycentric (SSB) coordinates (with the x axis aligned with the vernal point), we set

$$\begin{aligned}x_n &= a \cos \alpha + a e (\sin \alpha \cos \alpha \sin \beta_n - (1 + \sin^2 \alpha) \cos \beta_n), \\y_n &= a \sin \alpha + a e (\sin \alpha \cos \alpha \cos \beta_n - (1 + \cos^2 \alpha) \sin \beta_n), \\z_n &= -\sqrt{3} a e \cos(\alpha - \beta_n),\end{aligned}\tag{1}$$

where $\beta_n = (n - 1) \times 2\pi/3 + \lambda$ ($n = 1, 2, 3$) is the relative orbital phase of each spacecraft, $a = 1$ AU is the semi-major axis of the guiding center, and $\alpha(t) = 2\pi t/(1\text{year}) + \kappa$ is its orbital phase. In this approximation, the spacecraft form a rigid equilateral triangle with side length $L = 2\sqrt{3}ae = 5 \times 10^6$ km for $e = 0.00965$. (In fact, the LISA Simulator and Synthetic LISA implement e^2 -accurate orbits, but the additional terms make very little difference to the instrument response.)

The parameters κ and λ (InitialPosition and InitialRotation in lisaXML, see Sec. 3.2) set the initial location and orientation of the LISA constellation; in Challenge 1, $\kappa = \lambda = 0$. This choice places LISA at the vernal point, with spacecraft 1 directly below the guiding center in the southern ecliptic hemisphere. See Ref. [2] for expressions to convert to other LISA orbit specifications.

All times are measured by an ideal clock at the SSB.

1.2. The LISA response

The basic (individual-link) LISA response to GWs is taken to be the *phase response* Φ_{ij} used in the LISA Simulator and discussed in Sec. II of Ref. [4], or the *fractional frequency response* y_{slr}^{gw} used in Synthetic LISA and discussed in Sec. II B of Ref. [5]. (See the TDI Rosetta Stone [6] for translations between index notations.) The phase and fractional-frequency formalisms are equivalent, and are related by a simple time integration. The former has the advantage of representing more closely the actual output of the LISA phasemeters; the latter of being directly proportional to (differences of) the gravitational strains at the spacecraft. (In fact the *LISA Simulator* produces *equivalent-strain* data, with a nominal length of $L_n = 10^{10}$ m. To convert equivalent strain to fractional frequency, differentiate and multiply by $2\pi L_n/c$.)

LISA will employ Time-Delay Interferometry (TDI; see Refs. [7, 8, 9]) to cancel the otherwise overwhelming laser phase noise. In essence, TDI observables are constructed from time-delayed linear combinations of individual-link measurements, and they represent *synthesized* interferometers where laser phase fluctuations move in closed paths across the LISA arms. More complicated paths are required to deal with the variations of the armlengths due to the finer details of the LISA orbits, giving rise to the three TDI “generations.”

It is expected that high-level LISA data-analysis tasks (such as those targeted in the challenges) will be performed directly on TDI observables, and not on the underlying phase measurements. Thus, for the initial challenges we elect to represent the LISA output as *TDI 1.5* observables [8, 9], and in particular as the *unequal-arm Michelson* observables X , Y , and Z defined in Refs. [9]. Strictly speaking, TDI 2.0 would be required to cancel laser noise completely in a rotating and flexing LISA array such

as our pseudo-LISA; however, the upgrade from TDI 1.5 to 2.0 changes little in the response to GW signals, but it requires more careful numerical treatments and adds to the complexity of analysis codes. Thus, the initial Challenge data sets contain TDI 1.5 observables *without* laser noise.

1.3. The pseudo-LISA noises

The model of LISA instrument noise adopted in Challenge 1 includes only contributions from optical noise (assumed white in phase, with one-sided spectral density $S_{\text{opt}}^{1/2}(f) = 20 \times 10^{-12} \text{ mHz}^{-1/2}$), and from acceleration noise (assumed white in acceleration, but increasing as $1/f$ below 10^{-4} Hz , with one-sided spectral density $S_{\text{acc}}^{1/2}(f) = 3 \times 10^{-15} [1 + (10^{-4} \text{ Hz}/f)^2]^{1/2} \text{ ms}^{-2} \text{ Hz}^{-1/2}$), but not from laser phase noise, as discussed above.

The six optical noises and six acceleration noises (for the two optical benches on each spacecraft) are modeled as independent Gaussian random processes, and are realized in practice with sequences of pseudo-random numbers. Specifically, Synthetic LISA generates independent Gaussian deviates (i.e., white noise) in the time domain, and then filters them digitally to obtain the desired spectral shape; the LISA Simulator generates independent Gaussian deviates in the frequency domain, multiplies them by $S^{1/2}(f)$, and FFTs to the time domain.

2. MODELING GW SOURCES

Another source of complexity that we wish to exclude from the initial challenges is the uncertainty about the true shape of the gravitational waveforms that Nature will provide to LISA. However, we can already begin prepare for their detection and analysis, while waiting for theory to provide more and more accurate models, by working with fully known waveforms of comparable structure and increasing complexity. This section describes the standard simplified waveforms used in Challenge 1 to embody the signals emitted by the three kinds of GW sources under consideration: galactic binaries, massive black-hole binaries, and extreme-mass-ratio inspirals. A special care was devoted to choosing standard source parametrizations that could be used by MLDC participants to report their analysis results and compare them easily.

2.1. Conventions

The sky location of a GW source is described by its J2000 *ecliptic latitude* β and *longitude* λ , the latter measured from the vernal point, aligned with the \hat{x} axis in our convention. We model gravitational radiation from the source as a plane wave traveling along the direction $\hat{k} = -(\cos \beta \cos \lambda, \cos \beta \sin \lambda, \sin \beta)$, with surfaces of constant phase given by $\xi = t - \hat{k} \cdot x$. As written in the transverse-traceless gauge, the gravitational strain tensor can be decomposed in two standard polarization states,

$$\mathbf{h}(\xi) = h_+(\xi) [\hat{u} \otimes \hat{u} - \hat{v} \otimes \hat{v}] + h_\times(\xi) [\hat{u} \otimes \hat{v} + \hat{v} \otimes \hat{u}], \quad (2)$$

TABLE 1. Common source parameters. Note that in the initial challenges we do not deal explicitly with the redshifting of sources at cosmological distances; thus, D is a *luminosity* distance, and the masses and frequencies of Tables 3 are those measured at the SSB, which are red/blue-shifted by factors $(1+z)^{\pm 1}$ w.r.t. to those measured locally near the sources.

Parameter	Symbol	Standard parameter name (lisaXML descr.)	Standard unit (lisaXML descr.)
Ecliptic latitude	β	EclipticLatitude	Radian
Ecliptic longitude	λ	EclipticLongitude	Radian
Polarization angle	ψ	Polarization	Radian
Inclination	ι	Inclination	Radian
Luminosity distance	D	Distance	Parsec

where $h_+(\xi)$ and $h_\times(\xi)$ multiply the polarization tensors \mathbf{e}^+ and \mathbf{e}^\times formed from $\hat{u} = \partial \hat{k} / \partial \beta$, $\hat{v} \propto \partial \hat{k} / \partial \lambda$. Thus, GWs from any MLDC source are completely specified by β , λ , and by the two functions $h_+(\xi)$ and $h_\times(\xi)$ for the source's GW polarization amplitudes, measured at the SSB.

The orbital orientation of nonprecessing binaries is described by the inclination ι (the angle between the line of sight $-\hat{k}$ and the orbital angular momentum of the binary), and by their polarization angle ψ : specifically, if $h_+^S(\xi)$ and $h_\times^S(\xi)$ are the binary's GW polarizations in the source frame (i.e., defined with respect to the binary's *principal polarization axes* \hat{p} and \hat{q}) then

$$h_+(\xi) + ih_\times(\xi) = e^{-2i\psi} \left[h_+^S(\xi) + ih_\times^S(\xi) \right], \quad (3)$$

with $\psi = -\arctan(\hat{v} \cdot \hat{p} / \hat{u} \cdot \hat{p})$. Together with β , λ , and with the luminosity distance D , ι and ψ form a set of common standard parameters, listed in Tab. 1 with their standard lisaXML (see Sec. 3.2) descriptors and units.

2.2. Galactic binaries

Challenge 1 includes only searches for individually resolvable galactic binaries, as opposed to quasi-stochastic signals from populations of unresolvable sources. As an added simplification, all binaries are taken to be circular and monochromatic. Consequently, a Challenge-1 GalacticBinary source is completely determined by the parameters of Tables 1 and 2 together. The source-frame polarization amplitudes are computed in the restricted post-Newtonian approximation, and they are given by

$$\begin{aligned} h_+^S(\xi) &= \mathcal{A} (1 + \cos^2 \iota) \cos(2\pi f \xi + \phi_0), \\ h_\times^S(\xi) &= -2\mathcal{A} (\cos \iota) \sin(2\pi f \xi + \phi_0). \end{aligned} \quad (4)$$

The amplitude is specified explicitly among the source parameters; it is given in terms of the underlying physical parameters by $\mathcal{A} = (2\mu/D)(\pi M f)^{2/3}$, with $M = m_1 + m_2$ the total mass, and $\mu = m_1 m_2 / (m_1 + m_2)$ the reduced mass.

TABLE 2. GalacticBinary source parameters. Note that Amplitude effectively replaces the standard Distance parameter.

Parameter	Symbol	Standard parameter name (lisaXML descr.)	Standard unit (lisaXML descr.)
Amplitude	\mathcal{A}	Amplitude	1 (GW strain)
Frequency	f	Frequency	Hertz
Initial GW phase	ϕ_0	InitialPhase	Radian

2.3. Massive–black-hole binaries

For the sake of simplicity, all the massive–black-hole binaries GW sources considered in Challenge 1 are taken to be circular; black-hole spins are ignored, as are the final plunge and merger phases. In such spin-less, circular, adiabatic binary inspirals, the Taylor-expanded post-Newtonian equations for energy balance can be integrated analytically, yielding expressions [10, 11] for the orbital phase Φ as a function of the instantaneous orbital frequency ω , and for the time to coalescence $t_c - t$ as a function of ω . Truncating the two expressions to 2PN order, and inverting the second and substituting (numerically) in the first, we write the restricted post-Newtonian waveform for the inspiral as

$$\begin{aligned} h_+^S(\xi) &= \frac{2\mu}{D} [M\omega(\xi)]^{2/3} (1 + \cos^2 \iota) \cos[2\Phi(\xi)], \\ h_\times^S(\xi) &= -\frac{2\mu}{D} [M\omega(\xi)]^{2/3} (2 \cos \iota) \sin[2\Phi(\xi)]. \end{aligned} \quad (5)$$

We end the waveform when one of the following conditions is realized: i) the (Schwarzschild) last stable orbit is reached *or* ii) the “MECO” condition [12] is fulfilled *or* iii) $\dot{\omega}$ becomes negative. Such a termination engenders ringing in the Fourier domain. In reality this would not happen, because the inspiral waveform flows smoothly into the plunge and merger waveforms (which we do not model). Thus, we smooth out the waveform, beginning at an orbital separation $R_{\text{taper}} \in [7, 9]M$, by multiplying it by the *ad hoc* taper

$$w(t) = (1 + \tanh [A(M/R - M/R_{\text{taper}})]) / 2, \quad (6)$$

where R is approximated with Kepler’s law ($R = M^{1/3} \omega^{-2/3}$), and where the dimensionless coefficient $A = 150$ was determined empirically to produce smooth damping.

The lisaXML standard parameters of these Challenge-1 BlackHoleBinary sources are listed in Tables 1 and 3.

2.4. Extreme–Mass-Ratio Inspirals

The Extreme–Mass-Ratio Inspiral (EMRI) waveforms adopted in Challenge 1 are the Barack–Cutler “analytic kludge” waveforms [13], whereby orbits are instantaneously approximated as Newtonian ellipses (and gravitational radiation is given by the corresponding Peters–Matthews formula [14]), but perihelion direction, orbital plane, semi-major axis, and eccentricity evolve according to post-Newtonian equations. While these

TABLE 3. BlackHoleBinary source parameters.

Parameter	Symbol	Standard parameter name (lisaXML descr.)	Standard unit (lisaXML descr.)
Mass of first BH	m_1	Mass1	SolarMass
Mass of second BH	m_2	Mass2	SolarMass
Time of coalescence	t_c	CoalescenceTime	Second
Angular orbital phase at time $t = 0$	Φ_0	InitialAngularOrbitalPhase	Radian
Tapering radius	R	TaperApplied	TotalMass

TABLE 4. EMRI source parameters. Note that EMRIs do not use the nonprecessing-binary inclination ι , and be aware of the collision between the symbols for the EMRI compact-object mass (μ) and opening angle (λ), the binary reduced mass (again μ), and the ecliptic longitude (again λ).

Parameter	Symbol	Standard parameter name (lisaXML descr.)	Standard unit (lisaXML descr.)
Mass of central BH	M	MassOfSMBH	SolarMass
Mass of compact object	μ	MassOfCompactObject	SolarMass
Central-BH spin	$ S /M^2$	CoalescenceTime	Second
Central-BH spin orientation w.r.t. SSB frame	θ_K, ϕ_K	PolarAngleOfSpin, AzimuthalAngleOfSpin	Radian
Azimuthal orbital freq. at $t = 0$	ν_0	InitialAzimuthal... ...OrbitalFrequency	Hertz
Azimuthal orbital phase at $t = 0$	Φ_0	...OrbitalPhase	Radian
Eccentricity at $t = 0$	e_0	InitialEccentricity	1
Direction of pericenter at $t = 0$	$\tilde{\gamma}_0$	InitialTildeGamma	Radian
Direction of orbital angular momentum w.r.t. S at $t = 0$	α_0	InitialAlphaAngle	Radian
Opening angle	λ	LambdaAngle	Radian

waveforms are not particularly accurate in the highly relativistic regime of interest for real EMRI searches, they do exhibit the main qualitative features of the true waveforms, and they are considerably simpler to generate. It is expected that any search strategy that works for them could be modified fairly easily to deal with the true general-relativistic waveforms, once these become available.

The “analytic kludge” waveforms are too complex to describe in this restricted space, so we refer the reader to Refs. [13] and [2], and we content ourselves with presenting a complete table of Challenge-1 EMRI parameters in Tab. 4.

3. ENCODING CHALLENGE DATA SETS

All MLDC training and challenge data sets are distributed from the MLDC website [1] in a standard file format, developed by the MLDC Task Force with the goal of facilitating the use of the data sets on different computing platforms, and of enabling their identification, perusal, tracking, and archival. The MLDC file format is also used

internally by the MLDC Task Force within the workflow that leads from the choice of random source parameters to the generation of TDI data sets.

3.1. File structure

Depending on its use, a MLDC file will contain one or more of the following sections:

- A *prolog* including file metadata such as author, generation date, the name and version of the computer code used to create the data, and any other relevant comments.
- A *LISA data* section describing the model of the LISA orbits used in the simulations; for the initial challenges, this amounts to the `InitialPosition` and `InitialRotation` needed to fully specify the pseudo-LISA orbits of Sec. 1.1.
- A *noise data* section describing the models of the LISA noises used in the simulations; for the initial challenges, this amounts to the power spectral densities and generation timesteps of the pseudo-random sequences described in Sec. 1.3.
- A *source data* section describing the gravitational waveform(s) included in the simulation, whether specified in terms of the standard source parameters of Sec. 2, or represented as explicit time series of the h_+ and h_\times at the SSB, plus a few parameters such as the source's sky position and the time offset, cadence, and length of the time series.
- A *TDI data* section containing one or more time series of TDI observables, assembled from LISA's response to noises and sources, as described in Sec. 1.2; for the initial challenges, the observables of choice are X , Y , and Z of TDI 1.5, plus (trivially) the SSB time. The standard names of these observables are X_p , Y_p , and Z_p for the equivalent-strain version of the data sets (generated with the LISA Simulator), and X_f , Y_f , and Z_f for the fractional-frequency-fluctuation version (generated with Synthetic LISA).

Thus, training data sets will be represented by MLDC files including all types of sections; on the other hand, challenge data sets (which must be “blind”) will omit the source-data section.

3.2. Implementation

The MLDC file format is implemented using XML (the eXtensible Markup Language), a simple, flexible text format related to HTML, and widely used in the exchange, manipulation, and storage of many kinds of data, especially on the world-wide web [15]. Software libraries to handle XML are readily available for most computer languages. Specifically, the MLDC file format (known as *lisaXML*) is based on XSIL (the eXtensible Scientific Interchange Language), an XML dialect developed at Caltech to represent scientific data in multiple applications [16].

XSIL data structures consist of few simple building elements: a hierarchical container (XSIL), used, for instance, to enclose the LISA-data and source-data sections of an MLDC file; a parameter element (Param), used to describe parameter values and their units (for GW sources, some of the allowed parameter Names and Units are found in Tables 1–4); and an array element (Array), used to specify sequences of homogeneous

data, such as the time series of TDI observables. The actual array elements are stored in separate binary files, which can be saved, stored, and loaded very efficiently, while the information about the location, organization, and encoding of the binary files, as well as the file metadata and the LISA, noise, and source parameters, are contained in the textual XML files, which are easily parseable and editable by humans (or by powerful XML libraries). In fact, lisaXML files can be viewed, pleasantly formatted, in standard-compliant web browsers such as Firefox and Safari.

The MLDC Task Force has developed dedicated software tools to read and write lisaXML files from different computing environments, including C/C++ (see `lisaXML/io-C` in [17]), Python (by way of the routines build into Synthetic LISA [5]), MATLAB (see `xml2matlab.c` in `lisaXML/C-examples` [17]), and all applications that read ASCII files (see `xml2ascii.c`, also in `lisaXML/C-examples`). For more detailed information on the structure and usage of lisaXML files, see the MLDC omnibus documents [2] and the LISA Tools SourceForge website [17]; these include also the instructions and scripts necessary to reproduce the MLDC workflow and create additional training sets.

REFERENCES

1. Mock LISA Data Challenge Homepage, astrograves.nasa.gov/docs/mldc.
2. Mock LISA Data Challenge Task Force, "Document for Challenge 1," svn.sourceforge.net/viewvc/lisatools/Docs/challenge1.pdf.
3. Mock LISA Data Challenge Task Force wiki, www.tapir.caltech.edu/dokuwiki/listwglb:home.
4. N. J. Cornish and L. J. Rubbo, *Phys. Rev. D* **67**, 022001 (2003); erratum, *ibid.*, 029905 (2003); LISA Simulator website, www.physics.montana.edu/lisa; the canonical MLDC releases of Synthetic LISA are also available from the SourceForge MLDC site [17].
5. M. Vallisneri, *Phys. Rev. D* **71**, 022001 (2005); Synthetic LISA website, www.vallis.org/syntheticlisa; the canonical MLDC releases of Synthetic LISA are also available from the SourceForge MLDC site [17].
6. M. Vallisneri, the *TDI Rosetta Stone* v. 1/5/2005, www.vallis.org/tdi.
7. (TDI 1.0) J. W. Armstrong, F. B. Estabrook, and M. Tinto, *Astrophys. J.* **527**, 814 (1999); F. B. Estabrook, M. Tinto, and J. W. Armstrong, *Phys. Rev. D* **62**, 042002 (2000).
8. (TDI 1.5) N. J. Cornish & R. W. Hellings, *Clas. Quant. Grav.* **20**, 4851 (2003).
9. (TDI 2.0) D. A. Shaddock, M. Tinto, F. B. Estabrook, and J. W. Armstrong, *Phys. Rev. D* **68**, 061303(R) (2003); M. Tinto, F. B. Estabrook, and J. W. Armstrong, *Phys. Rev. D* **69**, 082001 (2004).
10. L. Blanchet, G. Faye, B. R. Iyer, and B. Joguet, *Phys. Rev. D* **65**, 061501 (2002); erratum, *Phys. Rev. D* **71**, 129902(E) (2005).
11. T. Damour, B. R. Iyer, and B. S. Sathyaprakash, *Phys. Rev. D* **63**, 044023 (2001).
12. A. Buonanno, Y. Chen, and M. Vallisneri, *Phys. Rev. D* **67**, 104025 (2003); erratum, *Phys. Rev. D* **74**, 029904(E) (2006).
13. L. Barack and C. Cutler, *Phys. Rev. D* **69**, 082005 (2004).
14. P. C. Peters and J. Mathews, *Phys. Rev.* **131**, 435 (1963); P. C. Peters, *ibid.* **136**, B1224 (1964).
15. World Wide Web Consortium (W3C), "eXtensible Markup Language (XML) activity and information," www.w3.org/XML.
16. Caltech Center for Advanced Computing Research, "XSIL: Extensible Scientific Interchange Language," www.cacr.caltech.edu/SDA/xsil.
17. LISA Tools SourceForge Project, sourceforge.net/projects/lisatools; the LISA Tools Subversion code archive is accessible from sourceforge.net/svn/?group_id=161442; the Subversion version control system is distributed and documented at subversion.tigris.org.