A how to for the Mock LISA Data Challenges

Michele Vallisneri
Jet Propulsion Laboratory

for the **Mock LISA Data Challenge Task Force**:
Keith Arnaud, Stas Babak, John Baker, Matt Benacquista, Neil Cornish, Curt Cutler, Shane Larson, Tyson Littenberg, Edward Porter, M.V., Alberto Vecchio

The Mock Data Challenges (MLDCs) have the dual purpose of fostering the development of LISA data-analysis tools and capabilities and of demonstrating the technical readiness already achieved by the gravitational-wave community in distilling a rich science payoff from the LISA data. The second round of MLDCs has just been completed and is being discussed at this Amaldi7 conference. A third round, including new classes of LISA sources, is planned for release at the end of this year. Here we describe the science and technology underlying the MLDCs, provide details about their implementation, and present the simulation and modeling software tools made available to challengers.

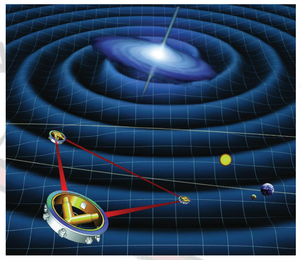
Introduction

The **Laser Interferometer Space Antenna** (LISA) is a spaceborne gravitational-wave (GW) observatory planned jointly by NASA and ESA, and designed for detailed studies of a wide variety of GW sources throughout the Universe, in the frequency range 0.1 mHz–0.1 Hz.

LISA is an **all-sky monitor** that can measure source parameters such as masses, spins, and distances with **unprecedented precision**. The LISA data set is expected to contain a large number (~10^4) of resolvable overlapping sources (ranging from galactic subsolar-mass binary systems to high-redshift massive black-hole binaries), on top of **strong GW backgrounds** generated by abundant populations of (mainly) Galactic and extra-Galactic white-dwarf binary systems. In addition, LISA has the potential of discovering radically new classes of GW sources, such as primordial stochastic backgrounds, cosmic strings, and exotic compact objects.

Although the vigorous ground-based GW-detection effort has taught us much about GW data analysis, the novel features and the very richness of the LISA data make it crucial that we start early in developing the special tools and methods necessary to extract the maximum scientific payoff.

The **Mock LISA Data Challenges** are a program sponsored by the LISA International Science Team (LIST) with the purpose of tracking progress, comparing methods, and encouraging research efforts related to the scientific analysis of LISA data. In each challenge, several **simulated LISA data sets** are distributed to challenge participants: the data including GW signals from sources of undisclosed parameters, as well as realistic instrument noise with known characteristics. Challenge participants are asked to analyze the data and report their estimates for the GW source parameters. These challenges are meant to be **blind tests**, but not really contests: their greatest benefit comes from the quantitive comparison of results, analysis methods, and implementations.



LISA is a constellation of three drag-free spacecraft, separated by 5x10^6 km, *and flying on an Earth-trailing solar orbit.*

Gravitational waves are detected in the 10^{-4}–10^{-1} *band as a modulation of the distance between spacecraft, by way of picometer interferometry.*

Challenge 1

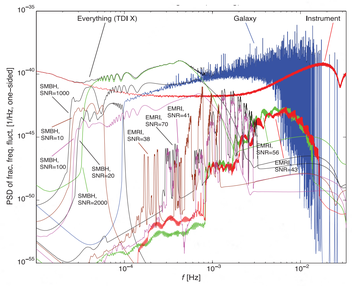
Challenge 1 was issued[1] in Jun 2006, and results were reported[2] at GWDAW-11 in Dec 2006. The data sets included the sources listed in the LISA minimum science requirements: **Galactic binaries** (including **verification** binaries), and **MBHB** systems. The sources were either isolated or moderately interfering. For simplicity, no background radiation was included from the Galactic binary population.

Entries were submitted by **ten research collaborations** among international institutions. Many different techniques were employed, and for each data set, at least one entry (and usually more) produced correct results. More importantly, Challenge 1 created a common field for all data-analysis players, and developed tools useful for the meaningful comparison of future results.

Challenge 2

Challenge 2 was issued[3] in Jan 2007, and entries[4] were due on June 15. The data sets included a **full-scale Galactic model** (30 million white-dwarf binaries, drawn from a Nelemans–Yungelsen–Zwart population model), together with overlapping **MBHB**s and **EMRI**s (EMRIs were also featured as isolated signals in five additional data sets).

The goal was to build on the standard tools and experience of Challenge 1 to tackle the **global-analysis problem** of LISA data analysis, whereby **sources overlap and interfere** in the time and frequency domains, so it can become difficult or impossible to distinguish them without skewing the estimates of their parameters, or missing signals altogether. **Thirteen collaborations** submitted a total of 17 entries, covering all data sets: results are being reported at this (Amaldi-7) conference.



A breakdown of the Challenge-2 data set into individual GW and noise components.

Challenge 1B and Challenge 3

Challenge 1B is being issued in Jul 2007 to provide a **second entry point** to the MLDC program for research groups that did not participate previously; results will be due in December 2007 (or possibly later, depending on the progress of the participants). This challenge will include the same data sets as Challenge 1 (plus EMRIs).

Challenge 3 will be issued in Dec 2007, and it will include **new LISA source classes**, such as cosmic-string bursts and cosmological stochastic backgrounds, and **increasingly accurate waveforms** (e.g., chirping galactic binaries, SMBH binaries with spins and numerical-relativity–inspired mergers and ringdowns) with parameters drawn from more extensive ranges than in Challenges 1 and 2.

Pseudo-LISA (orbits, TDI, noises)

The MLDC data sets are based on a conventional model (a *pseudo-LISA*) of the LISA orbits, GW response, and noise.[3]

- The pseudo-LISA **orbits** are analytical Keplerian orbits for point masses orbiting the Sun, truncated to 2nd order in the eccentricity; the initial location and orientation of the LISA constellation are set to standard values.[3]
- The one-way **phasemeter measurements** between the spacecraft are modeled as either the LISA Simulator's[5] equivalent-strain responses Φ_{ij} or Synthetic LISA's[6] fractional-frequency responses y_{ij}. The two responses are equivalent, and are related by a simple time integration.[3]
- The real LISA science data will consist of time series of **Time-Delay Interferometry** (TDI) observables[7] that suppress the otherwise overwhelming laser phase noise. In the MLDC we adopt the "TDI 1.5" *unequal-arm Michelson* observables *X*, *Y*, and *Z* of Ref. [8].
- The pseudo-LISA TDI time series include contributions from **optical noise** (assumed white in phase), with one-sided spectral density $S^{1/2}(f) = 20 \times 10^{-12}\,\mathrm{Hz}^{-1/2}$, and from **acceleration noise** (assumed white in acceleration, but increasing as 1/*f* below 10^{-4} Hz), with one-sided spectral density $S^{1/2}(f) = 3{\times}10^{-15}\,[1 + 10^{-4}\,(\mathrm{Hz}/f)^2]^{1/2}\,\mathrm{m\,s^{-2}\,Hz^{-1/2}}$. Laser phase noise, assumed to be cancelled by TDI, was not included.

Model waveforms

All GW signals included in the MLDC data sets are modeled as plane gravitational waves defined by their TT-gauge **polarization states** $h_+(t)$ and $h_\times(t)$ at the Solar-System Barycenter. The sky location of GW sources is described by **J2000 ecliptic coordinates** β (*latitude*) and λ (*longitude*); if $\boldsymbol{k}$ is the direction to the source, the (initial) orientation of the source is described by the *inclination* ι (the angle between $\boldsymbol{k}$ and the normal to the orbital plane of the source) and *polarization* ψ (the angle between the *principal axes*[3] of the source and the unit vectors $\partial\boldsymbol{k}/\partial\beta$ and $\partial\boldsymbol{k}/\partial\lambda$).

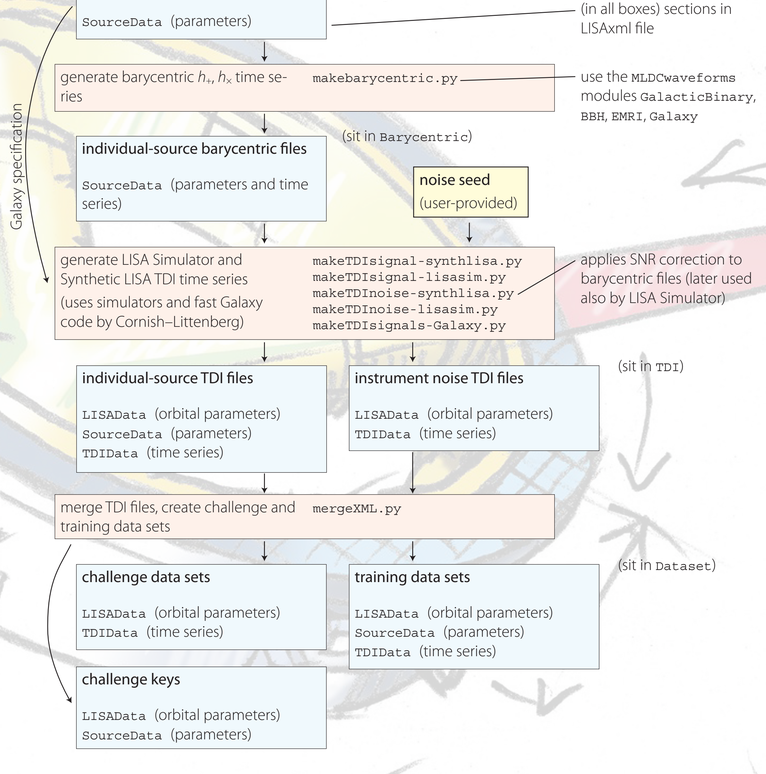
Challenges 1 and 2 included signals from Galactic binaries, massive–black-hole binaries, and extreme–mass-ratio inspirals. To reduce the complexity of the initial data-analysis task, the waveforms were distilled to very simplified forms with a minimum of parameters:

- for **Galactic binaries**, to the monochromatic quadrupolar emission from two point masses in circular orbit;[3]
- for **MBHB**s, to the quadrupolar emission from the adiabatic, quasi-circular inspiral of two nonspinning point masses, modeled by 2PN t2/t3 equations;[3,9]
- for **EMRI**s, to the quadrupolar emission from the PN/Peters–Matthews "analytical kludge" waveforms by Barack and Cutler.[3,10]

LISA and GW simulators

The Challenge-1 and -2 data sets were created with **the LISA Simulator**[5] (by N. Cornish and L. Rubbo) and **Synthetic LISA**[6] (by M. Vallisneri), used together with a set of source-specific **GW generators**[3] (by S. Babak, M. Vallisneri, N. Cornish, and T. Littenberg). These tools, as well other scripts and utilities, interoperate through the **lisaXML** file format, and are available in the public domain as part of the *LISAtools* suite.[11] They can be used to reproduce the **MLDC pipeline** and generate more challenge and training sets.

The MLDC challenge-generation pipeline



The MLDC challenge-generation pipeline is run from the main LISAtools script `challenge.py`. For instance, to generate a training set for Challenge 2 data set 2.2 with source parameters and noise derived from seed (say) 31415, one would run

```
$ challenge.py --training --seed=31415 challenge2.2
```

MLDC file formats

The MLDC data sets and pipeline working files are encoded in *lisaXML* [1], a format based on **XSIL** (the eXtensible Scientific Interchange Language) [12]. XSIL itself is a dialect of the widely-used **XML**, a simple, flexible and widely used text format related to HTML.

In lisaXML/XSIL, **large data sequences** such as the TDI time series are stored in separate **binary files** that are linked from the XML, and that can be saved, stored, and loaded very efficiently; on the other hand, more granular data items such as source parameters are contained in the text-based XML, and are easily parseable and editable by humans (or by powerful XML libraries).

A number of **dedicated software tools** to read and write lisaXML files (for C/C++, Python, MATLAB, and for conversion to ASCII) were developed by the MLDC Task Force and are available in the LISAtools suite [11]. For instance, to load the TDI data from a challenge file:

in C, use the LISAtools library `lisaXML/io-C`

```
TimeSeries *ts;
ts = getTDIdata("challenge.xml");

Then the length and cadence of the data are in ts->Length and
ts->Cadence, and the time series for t, X, Y, and Z are in
timeseries->Data[0]->data[],
timeseries->Data[1]->data[], etc.
```

To use the XSIL files in MATLAB or extract the TDI data as text, use `lisaXML/io-C/C-examples/xml2matlab.c` and `xml2ascii.c`

```
>> array = xml2matlab('challenge.xml');

$ xml2ascii challenge.xml output.txt
```

The Python interface is especially expressive; in the MLDC pipelines, it is also used to easily manipulate source parameters

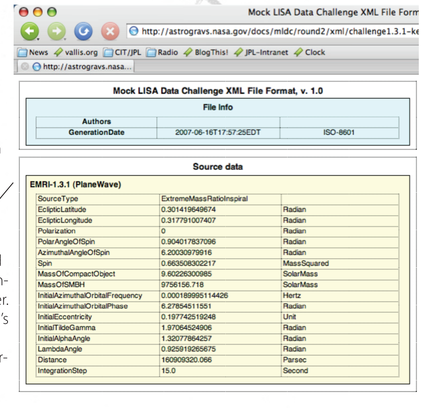
```
xml = lisaxml.readXML("challenge.xml")
tdidata = xml.TDIData[0]

Then the length and cadence of the data are in tdidata.Time-
Series.Length and Cadence, and the time series for t, X, Y, and Z
are in tdidata.t, tdidata.X, etc.
```

XML files consist of simple hierarchical structures of elements endowed with **named** attributes and containing data (or more elements)

```
<XSIL>
 <Param Name="Author">
  MLDC Task Force
 </Param>
 <Param Name="GenerationDate" Type="ISO-8601">
  2007-06-16T17:57:25EDT
 </Param>
 ...
 <XSIL Type="SourceData">
  <XSIL Name="EMRI-1.3.1" Type="PlaneWave">
   <Param Name="SourceType">
    ExtremeMassRatioInspiral
   </Param>
   <Param Name="EclipticLatitude" Unit="Radian">
    0.301419649674
   </Param>
   <Param Name="EclipticLongitude" Unit="Radian">
    0.317791007407
   </Param>
   ...
  </XSIL>
 ...
 </XSIL>
</XSIL>
```

XSIL consists of few simple building blocks, including the container `XSIL` (used in lisaXML to delimit sections such as `SourceData`), the physical-variable element `Param`, and the tabular-data descriptor `Array` (not seen in this example)

lisaXML files can be displayed in intuitive fashion in any standards-compliant web browser. This is achieved by the MLDC's *XSLT* transformation rules of the original XML into browser-friendly HTML



Web resources

NASA LISA site: **lisa.nasa.gov**
ESA SITE: **lisa.esa.int**
LISA International Science Community: **www.lisascience.org**
MLDC official site: **astrogravs.nasa.gov/docs/mldc**
MLDC taskforce wiki: **www.tapir.caltech.edu/listwg1b**
MLDC software repository: **sourceforge.net/projects/lisatools**

Image credits

ESA–C. Vijoux (background); NASA (LISA concept); MLDC Task Force (spectra)

Bibliography

[1] K. A. Arnaud et al., "The Mock LISA Data Challenges: An overview" and "A how-to for the Mock LISA Data Challenges," in Laser Interferometer Space Antenna: 6th International LISA Symposium (Greenbelt, MD, 19-23 Jun 2006), S. M. Merkowitz and J. C. Livas, eds. (AIP, Melville, NY, 2006), p. 619, 625.

[2] K. A. Arnaud et al., "Report on the first round of the Mock LISA Data Challenges," proceedings of the 11th Gravitational Wave Data Analysis Workshop, Potsdam, Germany, Dec 18-21, 2006. gr-qc/0701139. These proceedings include several other contributions by individual challenge participants.

[3] K. A. Arnaud et al. (the LISA Mock Data Challenge Task Force), "An overview of the second round of the Mock LISA Data Challenges," proceedings of the 11th Gravitational Wave Data Analysis Workshop, Potsdam, Germany, Dec 18-21, 2006. gr-qc/0701170.

[4] See www.tapir.caltech.edu/~mldc/results2.

[5] N. J. Cornish and L. J. Rubbo, Phys. Rev. D 67, 022001 (2005). See also www.physics.montana.edu/lisa.

[6] M. Vallisneri, Phys. Rev. D 71, 022001 (2005). See also www.vallis.org/syntheticlisa.

[7] See, e.g., M. Vallisneri, Phys. Rev. D 72, 042003 (2005).

[8] D. A. Shaddock et al., Phys. Rev. D 68, 061303(R) (2003); M. Tinto, F. B. Estabrook, and J. W. Armstrong Phys. Rev. D 69, 082001 (2004).

[9] T. Damour et al., Phys. Rev. D 63, 044023 (2001).

[10] L. Barack and C. Cutler Phys. Rev. D 69, 082005 (2004).

[11] See sourceforge.net/projects/lisatools.

[12] Caltech Center for Advanced Computing Research, whitepaper on "XSIL: Extensible Scientific Interchange Language" (2000). www.cacr.caltech.edu/SDA/xsil.