The Mock LISA Data Challenges: from Challenge 1B to Challenge 3

The *Mock LISA Data Challenge Task Force*: Stanislav Babak[aei], John G. Baker[gsfc], Matthew J. Benacquista[utb], Neil J. Cornish[montana], Jeff Crowder[jpl], Shane L. Larson[weber], Eric Plagnol[apc], Edward K. Porter[aei], Michele Vallisneri[jpl,cit], Alberto Vecchio[bgham,north] and the *Challenge-1B participants*: Keith Arnaud[gsfc], Leor Barack[south], Arkadiusz Błaut[wroclaw], Curt Cutler[jpl,cit], Stephen Fairhurst[cardiff], Jonathan Gair[cambridge,aei], Xuefei Gong[cas], Ian Harry[cardiff], Deepak Khurana[iit], Andrzej Królak[warsaw], Ilya Mandel[cit,north], Reinhard Prix[hannover], B. S. Sathyaprakash[cardiff], Pavlin Savov[cit], Yu Shang[cas], Miquel Trias[mallorca], John Veitch[bgham], Yan Wang[nanjing], Linqing Wen[aei,cit,uwa], John T. Whelan[aei]

[aei] Max-Planck-Institut für Gravitationsphysik (Albert-Einstein-Institut), Am Mühlenberg 1, D-14476 Golm bei Potsdam, Germany
[apc] APC, UMR 7164, Univ. Paris 7 Denis Diderot, 10, rue Alice Domon et Leonie Duquet, 75025 Paris Cedex 13, France
[bgham] School of Physics and Astronomy, Univ. of Birmingham, Edgbaston, Birmingham B152TT, UK
[cambridge] Inst. of Astronomy, Univ. of Cambridge, Madingley Rd., Cambridge, CB30HA, UK
[cardiff] School of Physics and Astronomy, Cardiff Univ., 5, The Parade, Cardiff, CF243YB, UK
[cas] Inst. of Appl. Math, Acad. of Math. and System Sci., Chinese Academy of Sciences, 55 Zhongguancun Donglu, Beijing, 100080, China
[cit] Theoretical Astrophysics, California Inst. of Technology, Pasadena, CA 91125
[gsfc] Gravitational Astrophysics Lab., NASA Goddard Space Flight Center, 8800 Greenbelt Rd., Greenbelt, MD 20771, USA
[hannover] Max-Planck-Institut fr Gravitationsphysik (Albert-Einstein-Institut), D-30167 Hannover, Germany
[iit] Indian Institute of Technology, Kharagpur, India
[jpl] Jet Propulsion Laboratory, California Inst. of Technology, Pasadena, CA 91109, USA
[mallorca] Dept. de Física, Universitat de les Illes Balears, Cra. Valldemossa km. 7.5, E-07122 Palma de Mallorca, Spain
[montana] Dept. of Physics, Montana State Univ., Bozeman, MT 59717, USA
[nanjing] Dept. of Astronomy, Nanjing Univ., 22 Hankou Road, Nanjing, 210093, China
[north] Dept. of Physics and Astronomy, Northwestern Univ., Evanston, IL, USA
[south] School of Mathematics, Univ. of Southampton, Southampton, SO171BJ, UK
[utb] Center for Gravitational Wave Astronomy, Univ. of Texas at Brownsville, Brownsville, TX 78520, USA
[warsaw] Inst. of Mathematics, Polish Academy of Sciences, Warsaw, Poland
[weber] Dept. of Physics, Weber State Univ., 2508 University Circle, Ogden, UT 84408, USA
[wroclaw] Inst. of Theoretical Physics, Univ. of Wrocław, Wrocław, Poland

[uwa] School of Physics, M013, Univ. of Western Australia, 35 Stirling Highway, Crawley, WA 6009, Australia

E-mail: Michele.Vallisneri@jpl.nasa.gov

Abstract. The Mock LISA Data Challenges are a program to demonstrate LISA data-analysis capabilities and to encourage their development. Each round of challenges consists of several data sets containing simulated instrument noise and gravitational waves from sources of undisclosed parameters. Participants are asked to analyze the data sets and report the maximum information about the source parameters. The challenges are being released in rounds of increasing complexity and realism: here we present the results of Challenge 1B [issued... new groups...] and we describe Challenge 3 [which includes...].

PACS numbers: 04.80.Nn, 95.55.Ym

1. Introduction

The Laser Interferometer Space Antenna (LISA), a NASA and ESA space mission to detect gravitational waves (GWs) in the 10^{-5}–10^{-1} Hz range [1], will produce time series consisting of the superposition of the signals from millions of sources, many in our Galaxy, some as far as the edge of the observable universe. Some of the signals, such as those from extreme–mass-ratio inspirals (EMRIs), are very complex functions of the source parameters; others, such as those from Galactic white-dwarf binaries, are simpler, but their resolution will be confused by the presence of many other similar signals that overlap in frequency space. Thus, data analysis is integral to the LISA measurement concept, because no source can be observed without first carefully teasing out its individual voice in the noisy party of the LISA data. Indeed, it is important to understand data analysis in order to demonstrate that LISA can meet its science requirements, and to translate these into decisions on instrument design.

The idea of the Mock LISA Data Challenges (MLDCs) arose in late 2005 from this very realization. The MLDCs have the purpose of encouraging and tracking progress in LISA data-analysis development, and (as a useful byproduct) of producing a prototype of the LISA computational infrastructure, including common data formats, standard models of the LISA orbits, noises and measurements, software tools to generate waveforms and to simulate the LISA response, and more. The MLDCs are a coordinated (but voluntary) effort in the GW community, whereby a task force chartered by the LISA International Science Team periodically issues data sets containing synthetic noise and GW signals from sources of undisclosed parameters; challenge participants return detection candidates and parameter estimates, together with descriptions of their search methods. These results are then compiled and compared to the previously secret challenge "key." **[These two paragraphs are verbatim from the MLDC-2 report. Need to rephrase, add news, more interesting comments.]**

Challenge 1, issued in Jun 2006 with results due in Dec 2006 (see [2, 3]), tackled the detection and parameter characterization of *verification binaries* (Galactic binaries of known frequency and position); of loud unknown Galactic binaries, either alone or in small, moderately interfering groups; and of relatively loud inspirals of nonspinning supermassive–black-hole (MBH) binaries. All sources were represented by somewhat idealized waveforms, and they were staged on simulated instrument noise alone. Ten collaborations submitted entries, adopting a variety of methods (template-bank, stochastic- and genetic-optimization matched filtering; time–frequency; tomography; Hilbert transform). Despite the short timescale, each challenge was "solved" by at least one group, although some searches locked on strong secondary probability maxima for the source parameters. More important, Challenge 1 helped set the playing field and assemble the computational tools for the more realistic Challenge 2.

Challenge 2, issued in Jan 2007 with results due at the end of Jun 2007, raised the bar by proposing three complex subchallenges. Data set 2.1 contained signals from a full population of Galactic binary systems (about 26 million sources). Data set 2.2 contained signals from a similar (but distinct) Galactic-binary population, plus an undisclosed number (between 4 and 6) of signals from nonspinning-MBH binary inspirals with optimal single-interferometer signal-to-noise ratios (SNRs) between 10 and 2000 and a variety of coalescence times, and plus five EMRI signals with optimal SNRs between 30 and 100. The EMRIs were modeled as Barack and Cutler's "analytic kludges" [4]: adiabatic sequences of elliptical orbits emitting Peters–Mathews waveforms, with separation, precession and eccentricity evolving according to post-Newtonian equations. Last, five more data sets (denoted 1.3.1–5, since they were released at the time of Challenge 1) contained single EMRI signals over instrument noise alone, with optimal SNRs between 40 and 110. See [5] for more details about the signal models and the ranges from which the source parameters were drawn. Altogether, Challenge 2 successfully demonstrated the identification of $\sim 20,000$ Galactic binaries, the accurate estimation of nonspinning-MBH inspiral parameters, and the positive detection of EMRIs. [**These two paragraphs are verbatim from the MLDC-2 report. Need to shorten, rephrase, add more about results.**]

The very steep increase in complexity introduced by Challenge 2 over a short time-scale, the need to consolidate analysis techniques and to involve a larger number of groups before moving to even more taxing challenges motivated a second round of Challenge 1- type MLDCs. This concept, together with the need of a fresh exercise including EMRIs – that due to their complexity were only partially tackled for Challenge 2 – inspired Challenge 1B, with data sets distributed in late Summer 2007 with results due at the beginning of the following December. Challenge 1B is in essence a new round of challenge 1 data sets with add-on of EMRI data sets where a single signal is included into Gaussian and stationary instrumental noise. Ten collaborations took part to this challenge. Results from this challenge are described in detail in Sections 2 and more comprehensive accounts of the analysis techniques, and the associated results are given in Refs XXX[**include GWDAW proceedings where available**]

The Challenge 3.1 galactic foreground data sets are direct descendants of Challenge 2.1. The only modifications are that the underlying population synthesis model now includes interacting AM CVn systems in addition to the detached systems used in the earlier challenges, and the leading term in the frequency evolution is now included in the signal model.

[**A short introduction to Challenge3. What's the focus? What's new? Point to section later in this paper.**]

Table 1. The performance of challenge entries on the single WDWD binary challenges as calculated using SNR, and C. Data in which the extrinsic parameters were corrected using the $\mathcal{F}$-statistic are denoted with an asterisk (*); numbers are not reported where the frequency is well off, and the $\mathcal{F}$-statistic is only fitting to noise.

Group	SNR	C	SNR	C	SNR	C
	Challenge 1b.1.1a ($\mathrm{SNR}_{\mathrm{key}} = 13.819$)		Challenge 1b.1.1b ($\mathrm{SNR}_{\mathrm{key}} = 24.629$)		Challenge 1b.1.1c ($\mathrm{SNR}_{\mathrm{key}} = 15.237$)	
AEI	20.435	0.108	18.652	0.922	1.949	-0.190
AEI*	14.199	0.984	23.266	0.996	14.770	0.989
GSFC	13.805	0.992	20.310	0.807	4.827	-0.138
GSFC*			20.987	0.814		
IMPAN	14.427	0.988	25.235	0.981	16.465	0.925
IMPAN*			23.152	0.997		
MCMNJU	13.524	0.952	22.641	0.906	6.830	0.033
MCMNJU*	14.193	0.996	23.270	0.994		
UIBBham	13.577	0.992	23.479	0.996		

2. Report on Challenge 1B

2.1. Challenges 1B.1.X: Galactic binaries

If the equation for correlation has not been given, please include it.

$$C = \frac{(h_{\mathrm{key}}|h_{\mathrm{rec}})}{\sqrt{(h_{\mathrm{key}}|h_{\mathrm{key}})\,(h_{\mathrm{rec}}|h_{\mathrm{rec}})}}. \tag{1}$$

Similarly, if the equation for the parameter differences has not been given please include it.

$$\Delta\lambda = \lambda_{\mathrm{key}} - \lambda_{\mathrm{rec}}. \tag{2}$$

For white dwarf binary systems 7 parameters are required to fully characterise each source. These parameters are: the amplitude $\mathcal{A}$, the frequency f, the sky location θ, ϕ, the inclination angle ι, the polarisation angle ψ, and the initial phase ϕ_0. In Table 1, we give the values of the correlations for each challenge entry for Challenges 1b.1.1a-c. As some of the entries had close fits to the intrinsic parameters (f, θ, and ϕ), but had issues matching the remaining (extrinsic) parameters, these parameters were recalculated using the $\mathcal{F}$-statistic [**?**]. Results for these recalculated parameters are also provided. Table 2 provides the deviations of the parameter values from the values in the key files of Challenges 1b.1.1a-c.

The Challenge 1b.1.2 data set contained 25 "verification" binaries. The frequency and sky location of each of the binaries is given to the participants, who then searched for the extrinsic parameters using their algorithms. This simulates the search for signals from binaries that are already known to exist. In fact, five of these binaries are taken from a list of known binaries available on Gijs Nelemans' website [**?**], while the remaining twenty were simulated binaries. Table 3 contains the global SNRs and correlations between all sources characterized by the parameters in the key file and the all sources recovered by the three groups that participated in this challenge. As in Challenges 1b.1.1a-c the issue with the extrinsic parameters caused the correlations to be less than ideal. However, as the intrinsic parameters were provided to the participants an $\mathcal{F}$-statistic calculations were not performed.

Table 2. The performance of challenge entries on the single binary challenges as calculated using recovered parameter differences. Data in which the extrinsic parameters were corrected using the $\mathcal{F}$-statistic are denoted with an asterisk (*).

Group	$\Delta\theta$	$\Delta\phi$	Δf (nHz)	$\Delta\psi$	$\Delta\iota$	$\Delta\varphi$	$\Delta\mathcal{A}(\times 10^{-23})$
			Challenge 1b.1.1a				
AEI	0.0318	-0.120	-2.43	0.217	-0.454	1.17	1.22
AEI*	0.0318	-0.120	-2.43	0.700	0.215	-1.23	1.18
GSFC	0.00412	-0.0715	-1.81	0.708	0.252	1.33	1.20
IMPAN	0.0311	0.0185	2.13	0.454	0.212	-1.06	1.25
MCMNJU	0.0170	-0.0424	-0.534	0.662	0.426	-1.57	2.34
MCMNJU*	0.0170	-0.0424	-0.534	0.746	0.248	-1.44	1.37
UIBBham	-0.00540	-0.0790	-1.51	0.708	0.173	4.96	0.647
			Challenge 1b.1.1b				
AEI	0.0558	-0.00899	0.946	-1.05	0.283	1.63	-0.0664
AEI*	0.0558	-0.00899	0.946	-0.761	0.0776	1.46	0.00307
GSFC	0.462	0.0606	-30.9	2.56	0.182	0.516	-0.0245
GSFC*	0.462	0.0606	-30.9	-0.534	0.198	0.624	0.0665
IMPAN:	-0.0203	0.000708	0.852	0.333	0.339	-0.603	0.713
MCMNJU	0.0670	-0.00627	2.069530	-0.732	-0.0643	0.844	-0.223
MCMNJU*	0.0670	-0.00627	2.069530	-0.739	0.0633	1.30	-0.0165
UIBBham	0.0436	-0.00817	1.777530	-0.636	0.0428	1.13	-0.0293
			Challenge 1b.1.1c				
AEI	0.0261	0.00530	1.84	-0.499	-1.12	3.02	0.124
AEI*	0.0261	0.00530	1.84	0.0410	-0.0937	-0.123	0.106
GSFC	0.452	-1.48	140	1.82	-0.471	-0.663	-0.695
IMPAN	0.0158	0.0248	3.72	-1.51	-0.197	2.68	0.478
IMPAN*	0.0158	0.0248	3.72	-0.0287	-0.113	0.356	0.0792
MCMNJU	0.555	-0.368	359	-1.59	-0.250	-0.943	-0.532

Table 3. The performance of challenge entries on the verification binaries challenge as calculated using SNR, and C.

Group	SNR	C	# Recovered
Challenge 1b.1.2 ($\mathrm{SNR}_{\mathrm{key}} = 634.918$, 25 Sources)			
AEI	891.677	-0.822	25
GSFC	807.012	0.006	25
MCMNJU	603.805	0.267	25

The data set for Challenge 1b.1.3 contained the signals from 20 binaries distributed across the LISA band. Unfortunately, there was an issue with the generation of the signals that caused their SNRs to be too small for detection (all were below 1). So no evaluation of this challenge is given. As a silver lining to this cloud, the participating groups responding to this challenge reported that they were unable to find any sources.

Challenge 1b.1.4 was a test of the search algorithms in the presence of mild source confusion. Fifty-one sources were spread across a 15μHz band starting at 3mHz; a source density of 0.108 sources per frequency bin. Challenge 1b.1.5 tested the algorithms in the presence of a high level of source confusion. Forty-four sources were spread across a 3μHz band centered on 3mHz; a source density of 0.465 sources

Table 4. The performance of challenge entries on the mildly confused binaries challenge as calculated using SNR, and C. Data in which the extrinsic parameters were corrected using the $\mathcal{F}$-statistic are denoted with an asterisk (*).

Group	SNR	C	# Recovered	False Positives
Challenge 1b.1.4 ($\mathrm{SNR_{key}}$ = 340.233, 51 Sources)				
AEI	375.366	0.774	13	2
AEI*	329.344	0.966	13	2
GSFC	209.411	0.003	6	1
GSFC*	90.506	0.282	6	1

Table 5. The performance of challenge entries on the highly confused binaries challenge as calculated using SNR, and C. Data in which the extrinsic parameters were corrected using the $\mathcal{F}$-statistic are denoted with an asterisk (*).

Group	SNR	C	# Recovered	False Positives
Challenge 1b.1.5 ($\mathrm{SNR_{key}}$ = 273.206, 44 Sources)				
AEI	208.273	0.453	3	0
AEI*	251.985	0.929	3	0

per frequency bin. The presence of multiple sources in the data streams of these challenges introduces the possibility of the search algorithms missing sources (false negatives) as well as returning false positives. To determine the presence of false positives we first matched the frequencies of the recovered sources and those from the key file to within one frequency bin ($\frac{1}{\mathrm{year}}$) of each other. If the correlation between the pair was less than 0.7 (after correcting with the $\mathcal{F}$-statistic) the recovered source was considered a false positive. The two measures for the matches in these challenges, as in Challenge 1b.1.2, are the SNR and correlation. Again, we use the combined signal from all recovered sources compared to the signal from the sources in the key file. Table 4 gives the results of the two entries to Challenge 1b.1.4, and Table 5 gives the results of the single entry to Challenge 1b.1.5.

2.2. Challenges 1B.2.X: MBH binaries

The challenge for massive black holes consisted of isolated binaries over a Gaussian stationary instrument noise. For the massive black hole search, groups were required to return the nine parameters of the systems : the two masses $m_{1,2}$, the time to coalescence of the wave t_c, the sky position (β, λ), the luminosity distance D_L, the orbital inclination angle ι, the gravitational polarization angle ψ and the initial gravitational wave phase φ_0. The groups were given the following limited prior information on the parameters of the systems. One of the individual masses in each case would be in the range $1 \leq m_1/10^6\,M_\odot \leq 5$, while the other mass would be in the range $m_2 = m_1/x$ where $1 \leq x \leq 4$. The time to coalescence would be $t_c = 6 \pm 1$ months for Challenge 2.1 and $t_c = 400 \pm 40$ days for Challenge 2.2.

Two groups submitted results for the massive black hole search :

JPL : used a three step strategy combining a time-frequency track search analysis, followed by a template bank matched filter search, finishing with a Metropolis-Hastings

Table 6. Relative/absolute errors for MBH in Challenges 1B 2.1 and 2.2. The optimal key SNRs are 531.84 and 80.67 for each challenge respectively.

Group	SNR	$\Delta m_1/m_1$ $(\times 10^{-2})$	$\Delta m_2/m_2$ $(\times 10^{-2})$	$\Delta t_c/t_c$ $(\times 10^{-5})$	$\Delta\beta$	$\Delta\lambda$	$\Delta D_L/D_L$ $(\times 10^{-1})$	$\Delta\iota$ $(\times 10^{-1})$	$\Delta\psi$	$\Delta\varphi_0$
					2.1					
JPL	531.57	0.61	0.52	1.37	2.43	3.133	1.22	7.13	5.719	-2.846
Cardiff	511.78	12.1	10.01	3.601	1.374	0.549	5.89	6.87	4.835	-2.389
					2.2					
JPL	79.86	1.39	1.18	7.296	5.86	-1.462	4.803	-6.96	1.522	-4.725

Monte Carlo refinement. The JPL group submitted entries for both 2.1 and 2.2.
Cardiff : used a stochastic template bank matched filtering search. The Cardiff group submitted an entry for 2.1.

We can see from Table 6 that both groups suceeded in finding the injected binary systems. In all cases the groups recovered almost all of the key SNR. It is interesting to note that all entries ended up on secondary sky positions that affected the precision of the other extracted parameters.

2.3. Challenges 1B.3.X: EMRIs

Three groups participated in this challenge and those are the same groups which submitted results for the challenge 2 [5]. The basic underlying techniques used in the extracting the parameters remained the same, but were further improved and tuned (see more detailed article on EMRIs in this issue). The challenge for EMRIs had five data sets with a single signal in each, for details on the parameter distribution and SNRs we refer reader to [5]. The evaluation of the results was performed along the same line as for the challenge 2 and similar to the MBH binaries evaluation. In particular we have computed overlaps of the signals generated with recovered parameters with the true signal for each A and E channels. We have also evaluated combined SNR between the data set and the recovered signal (see Table 7). For the results submitted by Gair, Mandel, and Wen (EtfAG, [28]) we could not evaluate overlaps and SNR since their method (based on the time-frequency analysis) targeted only intrinsic parameters and was insensitive to the initial phases. Therefore we have produced also the table of the relative errors in the parameter estimation and the results are summarized in the Table 8.

From booth tables one can see clear detection and excellent estimation of the parameters for 1.3.1 by MT2 group (N. Cornish). Other submissions seem to suffer by the same problem as before (in challenge 1): the search algorithm got stack at the secondary maxima.

3. Synopsis of Challenge 3

Data sets 3.1, 3.2, and 3.3 consist of approximately two years of data (2^{22} samples with 15 s sampling time) for the TDI-1.5 observables X, Y, and Z (definition), with simulated GW signals and secondary instrument noises. Noise levels. LISA orbits. Two varieties (synthlisa, lisasim); conversion.

Data sets 3.4 and 3.5 are different...

Table 7. Errors for challenge 1.3.X

Entry	$\frac{d\beta}{\Delta\beta}$	$\frac{d\lambda}{\Delta\lambda}$	$\frac{d\theta_K}{\Delta\theta_K}$	$\frac{d\phi_K}{\Delta\phi_K}$	$\frac{da}{\Delta a}$	$\frac{d\mu}{\Delta\mu}$	$\frac{dM}{\Delta M}$	$\frac{d\nu_0}{\nu_0}$	$\frac{de_0}{0.15}$	$\frac{d\lambda_{SL}}{\Delta\lambda_{SL}}$
BBGP-1B.3.1	-0.03	-0.0059	-0.14	0.053	0.31	-0.20	-0.84	0.026	0.37	-0.022
EtfAG-1B.3.1	0.019	-0.0045	0.56	0.33	0.16	-0.11	-0.27	-9.3e-05	0.17	0.078
MT2-1B.3.1	0.0058	0.0027	0.00044	0.0051	-0.0022	0.0065	0.014	3.2e-06	-0.0085	-0.0020
BBGP-1B.3.2	-0.16	-0.43	0.46	-0.33	-0.0088	-0.0040	0.016	0.00014	-0.010	-0.0013
EtfAG-1B.3.2	-0.014	0.0042	0.97	-0.36	0.0043	-0.046	-0.069	-6.5e-05	0.041	0.0041
MT2-1B.3.2	0.0040	-0.0086	0.79	0.41	0.093	-0.064	0.35	-0.035	0.068	0.092
BBGP-1B.3.3	0.091	0.50	-0.23	0.045	-0.32	-0.49	-0.029	0.00061	0.019	0.054
EtfAG-1B.3.3	-0.01	-0.004	0.49	-0.34	0.0073	-0.059	-0.061	-7.8e-05	0.038	0.0061
MT2-1B.3.3	0.045	-0.019	-0.1	0.077	-0.066	0.13	0.59	0.00036	-0.33	0.010
BBGP-1B.3.4	-0.57	-0.37	0.37	-0.31	-0.025	0.020	-0.88	0.066	0.065	-0.16
EtfAG-1B.3.4	-0.56	0.49	0.56	-0.34	0.059	0.12	0.04	0.00028	-0.039	0.0040
BBGP-1B.3.5	-0.48	-0.14	-0.35	0.1	-0.094	-0.094	0.55	-0.0021	-0.017	-0.060
EtfAG-1B.3.5	-0.58	0.46	0.27	-0.084	0.20	-0.7	0.83	-0.066	0.066	0.27

Table 8. Overlaps for challenge 1.3.X, the true SNR are: 1.3.1 - 123.7, 1.3.2 - 133.46, 1.3.3 - 81.0, 1.3.4 - 104.5, 1.3.5 - 57.6.

Entry	overlap (A)	SNR_A	overlap (E)	SNR_E	SNR
BBGP-1B.3.1	0.57	51 .0	0.58	51.6	72.5
MT2-1B.3.1	0.998	86.1	0.997	88.3	123.4
BBGP-1B.3.2	0.07	6.6	0.18	18.2	17.6
BBGP-1B.3.2[a]	0.39	37.6	0.41	39.8	54.7
MT2-1B.3.2	0.54	49.5	0.54	50.8	70.9
BBGP-1B.3.3	-0.06	-4.2	-0.0003	-0.05	-3.0
BBGP-1B.3.3[a]	-0.2	-11.5	-0.32	-19.0	-21.5
MT2-1B.3.3	22.0	22.0	20.5	20.9	30.4
BBGP-1B.3.4	0.0007	2.1	-0.0002	0.8	2.1
BBGP-1B.3.4[b]	0.16	13.9	0.04	6.7	14.6
BBGP-1B.3.5	0.09	3.4	0.1	4.2	5.3

[a] corrected sign of the latitude
[b] corrected phases at $t = 0$

One "MLDC month"-long dataset (2^{21} samples with 1 s sampling time) with Poisson-distributed cosmic-string–cusp bursts, defined as in the accompanying document [**To be included as a chapter here**]. The Poisson mean rate is 5 per "MLDC month". SNRs will be uniformly distributed between 10 and 100. The logarithm of the maximum frequency (see accompanying document) will be uniformly distributed as $\log_{10} f_{\max} \in [-3, 1.0]$. Sky position and polarization are random over spheres.

Instrument noise is Gaussian and stationary; it includes secondary noise, where the levels of proof-mass and photo-detector noises are randomized independently on each optical bench by a uniform draw in $[-20, +20]\%$ w.r.t. their nominal value; it also includes laser noise, also randomized, with nominal amplitude reduced to $10\times$ the nominal amplitude of secondary noises at 1 mHz [**Michele: check**.] The datasets include the standard X, Y, Z TDI observables *and* the 12 inter-spacecraft and inter-spacecraft raw phase measurements. The datasets will be distributed only

Table 9. Summary of data-set content and source-parameter selection in Challenge 3. Parameters are sampled randomly from uniform distributions across the ranges given below, and all angular parameters (including spin and orbital–angular-momentum directions for MBH binaries) are drawn randomly from uniform distributions over the whole appropriate ranges. Source distances are set from the selected SNRs (in Challenge 3, "SNR" refers to the multiple–TDI-observable SNR approximated as $\sqrt{2} \times \max\{\mathrm{SNR}_X, \mathrm{SNR}_Y, \mathrm{SNR}_Z\}$). The MBH time of coalescence t_c and the cosmic-string–cusp burst central time t_C are given relative to the beginning of the relevant data sets. **[What else is new?]**

Dataset	Sources	Parameters
3.1	*Galactic-binary background*	randomized population (see section 3.1) $\sim 34 \times 10^6$ interacting, $\sim 26 \times 10^6$ detached
	plus 20 *verification binaries*	known parameters (see section 3.1)
3.2	4–6 *MBH binaries*	for each: $m_1 = 1$–$5 \times 10^6\, M_\odot$, $m_1/m_2 = 1$–4, $a_1/m_1 = 0$–$1, a_2/m_2 = 0$–1
	. . . including	MBH$_1$: $t_c = 90 \pm 30$ days, SNR ~ 2000 MBH$_2$: $t_c = 765 \pm 15$ days, SNR ~ 20
	. . . and 2–4 chosen from	MBH$_3$: $t_c = 450 \pm 270$ days, SNR ~ 1000 MBH$_4$: $t_c = 450 \pm 270$ days, SNR ~ 200 MBH$_5$: $t_c = 540 \pm 45$ days, SNR ~ 100 MBH$_6$: $t_c = 825 \pm 15$ days, SNR ~ 10
	plus *Galactic confusion*	randomized population with approx. SNR < 5 $\sim 26 \times 10^6$ binaries; no verification
3.3	5 *EMRIs*	for each: $\mu = 9.5$–$10.5\, M_\odot$, $S = 0.5$–$0.7\, M^2$, time at plunge $= 2^{21}$–$2^{22} \times 15$ s, ecc. at plunge $= 0.15$–0.25, SNR $= 10$–50
	. . . including	EMRI$_1$: $M = 0.95$–$1.05 \times 10^7 M_\odot$ EMRI$_2$ and EMRI$_3$: $M = 4.75$–$5.25 \times 10^6 M_\odot$ EMRI$_4$ and EMRI$_5$: $M = 0.95$–$1.05 \times 10^6 M_\odot$
3.4	*n Cosmic-string–cusp bursts*	(with n Poisson-distributed with mean 5) $f_{\max} = 10^{-3-1}$ Hz, $t_C = 0$–2^{21} s, SNR $= 10$–100 all instrument noise levels randomized $\pm 20\%$
3.5	*Isotropic stochastic background*	2×192 incoherent h_+ and $h_\times$ sources over sky $S_h^{\mathrm{tot}} = 0.7$–$1.3 \times 10^{-47} (f/\mathrm{Hz})^{-3}\,\mathrm{Hz}^{-1}$ all instrument noise levels randomized $\pm 20\%$

as fractional-frequency–fluctuation time series (i.e., as *Synthetic LISA* datasets).

The GW content of the data sets is summarized in table 9; in the next few subsection we give precise definition of the waveform and parameters for each source class.

Do a parameter table?

Each data set in blind and training variants.

3.1. Chirping Galactic binaries

Data set 3.1 contains GWs from a population of $\sim 26 \times 10^6$ detached and $\sim 34 \times 10^6$ interacting Galactic binaries. Each binary is modelled as a system of two point masses

Table 10. Source parameters in Challenge 3. 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 table **??** are those measured at the SSB, which are red/blue-shifted by factors $(1+z)^{\pm 1}$ with respect to those measured locally near the sources.

Parameter	Symbol	Standard parameter name (lisaXML descriptor)	Standard unit (lisaXML descriptor)
		Common parameters	
Ecliptic latitude	β	`EclipticLatitude`	`Radian`
Ecliptic longitude	λ	`EclipticLongitude`	`Radian`
Polarization angle	ψ	`Polarization`	`Radian`
Inclination	ι	`Inclination`	`Radian`
Luminosity distance[a]	D	`Distance`	`Parsec`
		Galactic binaries	
Amplitude[b]	$\mathcal{A}$	`Amplitude`	`1` (GW strain)
Frequency	f	`Frequency`	`Hertz`
Frequency derivative	$\dot{f}$	`FrequencyDerivative`	`Hertz/Second`
Initial GW phase	ϕ_0	`InitialPhase`	`Radian`
		Spinning massive black-hole binaries	
Initial polar angle of the 1-st spin	θ_{S1}	`PolarAngleOfSpin1`	`Radian`
Initial polar angle of the 2-nd spin	θ_{S2}	`PolarAngleOfSpin2`	`Radian`
Initial azimuthal angle of 1-st spin	ϕ_{S1}	`AzimuthalAngleOfSpin1`	`Radian`
Initial azimuthal angle of 2-nd spin	ϕ_{S2}	`AzimuthalAngleOfSpin2`	`Radian`
Magnitude of the 1-st spin	a_1	`Spin1`	`MassSquared`
Magnitude of the 2-nd spin	a_2	`Spin2`	`MassSquared`
Mass of 1-st MBH[c]	m_1	`Mass1`	`SolarMass`
Mass of 2-nd MBH[c]	m_2	`Mass2`	`SolarMass`
Time to coalescence	T_c	`CoalescenceTime`	`Second`
Phase at coalescence	Φ_c	`PhaseAtCoalescence`	`Radian`
Initial polar angle of orbital momentum	θ_L	`InitialPolarAngleL`	`Radian`
Initial azimuthal angle of orbital momentum	ϕ_L	`InitialAzimuthalAngleL`	`Radian`
		EMRIs: see table 5 of [5]	
		Cosmic string cusp bursts	
Amplitude[b] (Fourier)	$\mathcal{A}$	`Amplitude`	`Hertz^(1/3)`
Central time of arrival	t_C	`CentralTime`	`Second`
Maximum frequency[d]	$f_{\max}$	`MaximumFrequency`	`Hertz`
		Isotropic stochastic background	
PSD[b] at 1 Hz	S_h	`PowerSpectralDensity`	`(f/Hz)^-3/Hz`
note: $S_h = S_h^{\mathrm{tot}}/384$; ψ is set to 0, and ι not used			

[a] We do not deal explicitly with the redshifting of sources at cosmological distances, so D is a *luminosity* distance, and all masses and frequencies are measured at the SSB and red/blue-shifted by factors $(1+z)^{\pm 1}$ with respect to those measured locally near the sources.
[b] Replaces D for Galactic binaries, cosmic-string–cusp bursts, and stochastic-background pseudosources.
[c] Here are red-shifted masses.
[d] Effectively replaces ι for cosmic-string–cusp bursts.

m_1 and m_2 in circular orbit with linearly increasing or decreasing frequency (depending on whether gravitational radiation or equilibrium mass transfer is dominant). The polarization amplitudes at the Solar-system barycenter, expressed in the source frame, are given by

$$\begin{aligned} h_+^S(t) &= \mathcal{A}\left(1+\cos^2\iota\right)\cos[2\pi(ft+\dot{f}t^2/2)+\phi_0], \\ h_\times^S(t) &= -2\mathcal{A}(\cos\iota)\sin[2\pi(ft+\dot{f}t^2/2)+\phi_0], \end{aligned} \tag{3}$$

where the amplitude is derived from the physical parameters of the source as $\mathcal{A} = (2\mu/D)(\pi M f)^{2/3}$, with $M = m_1 + m_2$ the total mass, $\mu = m_1 m_2/M$ the reduced mass, and D the distance; $\dot{f}$ is the (constant) frequency derivative, and ϕ_0 is the phase at $t = 0$.

Since it would be unfeasible to process millions of barycentric binary waveforms individually through the LISA simulators to compute the TDI-observable time series, we adopt a fast frequency-domain method [15] that rewrites the LISA phase measurements as the fast–slow decomposition

$$y_{ij}(t) = C(t)\cos(2\pi f_0 t) + S(t)\sin(2\pi f_0 t) \tag{4}$$

the functions $C(t)$ and $S(t)$ describe slowly varying effects such as the rotation of the LISA arms, the Doppler shift induced by orbital motion, and the intrinsic frequency evolution of the source. These "slow" terms can be sampled very sparsely and Fourier-transformed numerically, while the "fast" sine and cosine terms can be Fourier-transformed analytically. The results are then convolved to produce the LISA phase measurements, and these are assembled into the desired TDI variables. This algorithm is three to four orders of magnitude faster than the time-domain LISA simulators, although it effectively approximates LISA as a rigidly rotating triangle with equal and constant armlengths. See [15] for full details, and `lisatools/MLDCwaveforms/Galaxy3` for the source code.

The starting point for each realization of data set 3.1 are two large catalogs provided by Gijs Nelemans (files `MLDCwaveforms/Galaxy3/Data/AMCVn_GWR_MLDC.dat` and `dwd_GWR_MLDC.dat` in the `lisatools` installation), which contain the parameters of 26.1×10^6 detached and 34.2×10^6 interacting systems produced by the population synthesis codes described in [13, 14]. Recent work by Roelofs, Nelemans and Groot [16] suggests that the model in [14] overpredicts the number of (AM CVn) interacting systems by a factor of 5–10, but we did not implement this correction for Challenge 3.

The parameters of each binary in the catalogs are modified by randomly tweaking f by $\pm 1\%$, A by $\pm 10\%$, β and λ by $\pm 0.5\,^\circ$, and by randomly assigning ψ, ι, and ϕ_0 ($\dot{f}$ is computed from the catalog's binary-period derivative and from the tweaked f). These random perturbations are large enough to render the original population files useless as answer keys, but small enough to preserve the overall parameter distributions. Binaries with approximate single-Michelson SNR > 10 are regarded as "bright" and listed in a separate table in the challenge keys. Data set 3.1 includes also 20 verification binaries of known parameters (specified in `lisatools` file `MLDCwaveforms/Galaxy3/Data/Verification.dat` as rows of f, $\dot{f}$, β, λ, A).

3.2. Spinning MBH binaries

Data set 3.2 contains...

A two-year dataset (2^{22} samples with 15 s sampling time) with signals from U[4,6] spinning MBH binary inspirals.

The spinning-binary signals are modeled as 2PN circular adiabatic inspirals with uncoupled orbital frequency evolution and spin and orbital precession. No higher-PN harmonics are included. Both phase and orbital frequency are explicit functions of time (TaylorT3 in classification presented in [24]):

$$M\omega = \frac{1}{8}\tau^{-3/8}\left[1 + \left(\frac{743}{2688} + \frac{11}{32}\eta\right)\tau^{-1/4} - \frac{3}{10}\left(\pi - \frac{\beta}{4}\right)\tau^{-3/8} + \right.$$
$$\left. + \left(\frac{1855099}{14450688} + \frac{56975}{258048}\eta + \frac{371}{2048}\eta^2 - \frac{3}{64}\sigma\right)\tau^{-1/2}\right] \tag{5}$$

where $\eta = \mu/M$ is symmetric mass ratio and

$$\tau = \frac{\eta}{5M}(T_c - t), \tag{6}$$

$$\beta = \frac{1}{12}\sum_{i=1,2}\left[\chi_i\left(\hat{\mathbf{L}}_{\mathbf{N}}.\hat{\mathbf{S}}_{\mathbf{i}}\right)\left(113\frac{m_i^2}{M^2} + 75\eta\right)\right] \tag{7}$$

$$\sigma = -\frac{1}{48}\eta\chi_1\chi_2\left[247(\hat{\mathbf{S}}_{\mathbf{1}}.\hat{\mathbf{S}}_{\mathbf{2}}) - 721(\hat{\mathbf{L}}_{\mathbf{N}}.\hat{\mathbf{S}}_{\mathbf{1}})(\hat{\mathbf{L}}_{\mathbf{N}}.\hat{\mathbf{S}}_{\mathbf{2}})\right] \tag{8}$$

Here $\hat{\mathbf{L}}_{\mathbf{N}},\ \hat{\mathbf{S}}_1,\ \hat{\mathbf{S}}_2$ are unit vectors along leading order angular orbital momentum and hole's spins. The intrinsic orbital phase is

$$\Phi_{orb} = \Phi_C - \frac{\tau^{5/8}}{\eta}\left[1 + \left(\frac{3715}{8064} + \frac{55}{96}\eta\right)\tau^{-1/4} - \frac{3}{16}(4\pi - \beta) + \right.$$
$$\left. \left(\frac{9275495}{14450688} + \frac{284875}{258048}\eta + \frac{1855}{2048}\eta^2 - \frac{15}{64}\sigma\right)\tau^{-1/2}\right]. \tag{9}$$

Due to the spin-orbital coupling the spins and orbital angular momentum are precessing around total angular momentum, therefor we need to add precessional correction to the orbital phase (see [25]):

$$\dot{\Phi} = \omega + \frac{(\hat{\mathbf{L}}_{\mathbf{N}}.\hat{\mathbf{n}})[\hat{\mathbf{L}}_{\mathbf{N}} \times \hat{\mathbf{n}}]\dot{\hat{\mathbf{L}}}_N}{1 - (\hat{\mathbf{L}}_{\mathbf{N}}.\hat{\mathbf{n}})^2} \equiv \omega + \delta\dot{\Phi}, \tag{10}$$

where $\hat{\mathbf{n}}$ is direction to the source in SSB. The constant of integration of (10), Φ_C, can be redefined so that $\delta\dot{\Phi} = 0$ at $t = 0$. The precession equations for $\hat{\mathbf{L}}_{\mathbf{N}},\ \hat{\mathbf{S}}_1,\ \hat{\mathbf{S}}_2$ are given by eqn. (2.9)-(2.11) in [26]. As mentioned above we use restricted waveform (only leading order amplitude) and in the source frame (with the phasing formulae above) it takes the following form

$$h_+ = -2\frac{\mu}{D}(1 + \cos(i)^2)(M\omega)^{2/3}\cos 2\Phi \tag{11}$$

$$h_\times = 4\frac{\mu}{D}\cos(i)(M\omega)^{2/3}\sin 2\Phi. \tag{12}$$

The inclination angle i is defined by initial position of the orbital momentum and by the direction to the source: $\cos i = (\hat{\mathbf{L}}_{\mathbf{N}}.\hat{\mathbf{n}})$. Note that if one uses the approach given in [27], the amplitudes will be more complicated and the precession part of the phase at $t = 0$ should be $\delta\dot{\Phi}^K = -\gamma^K$, where superscript K stands for Kidder and

$$\gamma^K = \frac{\mathbf{e}_z \times \hat{\mathbf{L}}_{\mathbf{N}}}{|\mathbf{e}_z \times \hat{\mathbf{L}}_{\mathbf{N}}|}\frac{\hat{\mathbf{L}}_{\mathbf{N}} \times \hat{\mathbf{n}}}{\sqrt{1 - (\hat{\mathbf{L}}_{\mathbf{N}}.\hat{\mathbf{n}})^2}}.$$

The end of the inspiral is handled with an exponential taper, as in Challenge 2. The expressions for h_+ and $h_\times$ are given in the time varying polarization basis, but to generate the LISA responses it is necessary to re-express it in terms of fixed polarization tensors. This is achieved through a rotation by the polarization angle ψ:

$$\tan\psi = \frac{\cos\beta\cos(\lambda-\phi_L)\sin\theta_L - \cos\theta_L\sin\beta}{\sin\beta\sin(\lambda-\phi_L)}. \tag{13}$$

and the waveform in SSB becomes

$$h_+^{\mathrm{SSB}} = -\,h_+\cos 2\psi - h_\times \sin 2\psi \tag{14}$$

$$h_\times^{\mathrm{SSB}} = h_+\sin 2\psi - h_\times\cos 2\psi. \tag{15}$$

Masses, SNRs, and times of coalescence are chosen as in Challenge 2 (see also table 9); the spin magnitudes S_1/M_1^2 and S_2/M_2^2 are drawn uniformly in $[0,1]$; all the angles (spin directions, initial orbital angular momentum, sky position) are drawn uniformly over spheres.

The dataset includes also a Galactic confusion background generated from the same detached-binary population as Challenge 3.1, but withholding all binaries with individual SNR > 5 relative to the combined instrument plus galactic confusion noise **[I think we use only instrument noise! Neil?]**, approximated as

$$S_{X,\mathrm{gal}} = 16\,(2\pi f L)^2 \sin^2(2\pi f L) \times \begin{cases} 10^{-44.62} f^{-2.3} & \text{for } f \in [10^{-4}, 10^{-3}] \quad \text{Hz},\\ 10^{-50.92} f^{-4.4} & \text{for } f \in [10^{-3}, 10^{-2.7}] \quad \text{Hz},\\ 10^{-62.8} f^{-8.8} & \text{for } f \in [10^{-2.7}, 10^{-2.4}] \ \text{Hz},\\ 10^{-89.68} f^{-20.0} & \text{for } f \in [10^{-2.4}, 10^{-2.0}] \ \text{Hz}, \end{cases} \tag{16}$$

This estimate was derived using the BIC analysis code used in the evaluation of Challenge 2, and it does not include interacting systems; these should not make a significant contribution to confusion noise, since they typically have very small chirp masses, and hence amplitudes, compared to detached systems.

Approximation formula for confusion noise.

Instrument noise is secondary-only, Gaussian, stationary, and equal on all spacecraft.

3.3. EMRIs

A single two-year datasets (2^{22} samples with 15 s sampling time) containing the signals from 5 "MLDC EMRIs", with parameters drawn as in Challenge 1B: compact object mass m in $\mathrm{U}[9.5, 10.5] \times M_\odot$, spin of central BH S/M in U[0.5, 0.7], time at plunge in $\mathrm{U}[2^{21}, 2^{22}] \times 15$ sec, and eccentricity at plunge in U[0.15, 0.25]. The central black holes masses are chosen so that one system has M in $\mathrm{U}[0.95, 1.05] \times 10^7 M_\odot$, two systems have M in $\mathrm{U}[4.75, 5.25] \times 10^6 M_\odot$ and two systems have M in $\mathrm{U}[0.95, 1.05] \times 10^6 M_\odot$. In addition to having mutliple EMRIs in a single data set, the other new wrinkle for this challenge is that the SNRs are much lower: SNR in U[10,50]. The number of eccentric-orbit harmonics does not vary with eccentricity along the evolution (as was the case in Challenge 2 and 1B), but is fixed at 5. Sky position, polarization, extrinsic parameters are random over spheres.

Instrument noise is secondary-only, Gaussian, stationary, and equal on all spacecraft.

3.4. Cosmic string cusps

Data set 3.4 contains a number of bursts from cosmic strings, the first of two new GW sources introduced with Challenge 3. Cosmic strings are linear topological defects that may be formed in early Universe at the phase transitions predicted in many elementary-particle and superstring models. Cosmic-string oscillations emit gravitational radiation, with a substantial part of the emission from *cusps*, which can achieve very large Lorentz boosts [17]. In the limit where the tip of a cusp is moving directly toward the observer, the observed metric perturbation is a linearly polarized GW with [18]

$$h(t) = A|t - t_C|^{1/3}, \quad A \sim \frac{G\mu L^{2/3}}{D_L}; \tag{17}$$

here t_C is the burst's central time of arrival, G is Newton's constant, μ is the string's mass per unit length, D is the luminosity distance to the source, and L is the size of the feature that produces the cusp (e.g., the length of a cosmic string loop). If the observer's line of sight does not coincide with the cusp's direction of motion, the waveform becomes a much more complicated mixture of polarizations [19]. If the viewing angle α departs only slightly from zero, the waveform remains dominantly linearly polarized, and the sharp spike in (17) is rounded off, introducing an exponential suppression of Fourier-domain power for frequencies above $f_{\max} = 2/(\alpha^3 L)$.

Following the model used by the LIGO Science Collaboration, we define our cusp waveforms in the Fourier domain according to

$$|h_+(f)| = \mathcal{A} f^{-4/3} \left(1 + (f_{\text{low}}/f)^2\right)^{-4}, \quad h_\times = 0, \tag{18}$$

with $\exp(1 - f/f_{\max})$ suppression above $f_{\max}$. The amplitude $\mathcal{A}$ has dimensions $\mathrm{Hz}^{1/3}$; f_{low} sets the low-frequency cut-off of what is effectively a fourth-order Butterworth filter, which prevents dynamic-range issues with the inverse Fourier transforms (for Challenge 3 we set $f_{\text{low}} = 1 \times 10^{-5}$ Hz). The phase of the waveform is set to $\exp \mathrm{i}(\pi - 2\pi f t_C)$ before inverse-Fourier transforming to the time domain. See `lisatools/MLDCwaveforms/CosmicStringCusp` for the source code.

3.5. Stochastic background

Data set 3.5 contains the second GW source new to Challenge 3: an isotropic, unpolarized, Gaussian and stationary stochastic background. Allen and Romano [20] define a stochastic background as the "gravitational radiation produced by an extremely large number of weak, independent, and unresolved gravity-wave sources, [...] stochastic in the sense that it can be characterized only statistically." Such backgrounds are usually characterized by the dimensionless quantity

$$\Omega_{\text{gw}}(f) = \frac{1}{\rho_{\text{crit}}} \frac{\mathrm{d}\rho_{\text{gw}}}{\mathrm{d}\log f}, \tag{19}$$

with ρ_{gw} the energy density in GWs, and $\rho_{\text{crit}} = 3c^2 H_0^2/(8\pi G)$ the closure energy density of the Universe, and they are idealized as the collective, incoherent radiation of uncorrelated infinitesimal emitters distributed across the sky. If the background is isotropic, unpolarized, Gaussian and stationary, the Fourier amplitude $\tilde{h}_A(f, \hat{\Omega})$ of each emitter (with A indexing the $+$ and $\times$ polarizations, and $\hat{\Omega}$ the direction on the two sphere) is completely characterized by the power-spectral-density relation [20]

$$\left\langle \tilde{h}^*_A(f,\hat{\Omega})\tilde{h}_{A'}(f',\hat{\Omega}')\right\rangle = \frac{3H_0^2}{32\pi^3}|f|^{-3}\Omega_{\rm gw}(|f|) \times \delta_{AA'}\delta(f-f')\delta^2(\hat{\Omega},\hat{\Omega}'). \tag{20}$$

In Challenge 3, we assume a constant $\Omega_{\rm gw}(f)$, as appropriate for the primordial background predicted in many cosmological scenarios. We implement the uncorrelated emitters as a collection of 192 pseudosources distributed at *HEALpix* pixel centers across the sky. HEALpix (the Hierarchical Equal Area isoLatitude Pixelization of spherical surfaces [21]) is often used to represent cosmic microwave background data sets; 192 pixels correspond to a twice-refined HEALpix grid with $N_{\rm side} = 2^2$ pixels.

Each pseudosource consists of uncorrelated pseudorandom processes for h_+ and $h_\times$, generated as white noise in the time domain, and filtered to achieve the f^{-3} spectrum of (20), using the the recursive $1/f^2$ filtering algorithm proposed by Plaszczynski [22], extended to spectral slope -3. The algorithm employs a chain of $1/f^2$ infinite–impulse-response filters to reshape the white noise spectrum between minimum and maximum frequencies $f_{\rm low}$ and $f_{\rm knee}$, set to 10^{-5} and 10^{-2} Hz in this Challenge (see source code `lisatools/MLDCwaveforms/Stochastic.py` for the Synthetic LISA implementation).

The *one-sided* PSD of each single-polarization random process is then $S_h(f) = S_h^{\rm tot}(f)/(2\times 192)$, with $(1/2)S_h^{\rm tot}(f) = 3H_0^2/(4\pi)f^{-3}\Omega_{\rm gw}$. In data set 3.5, we set $S_h^{\rm tot}$ so that, in the TDI observables, the GW background is a few times stronger than LISA's secondary instrument noise. Namely,

$$S_h^{\rm tot}(f) = 0.7\text{–}1.3\times 10^{-47}(f/\mathrm{Hz})^{-3}\,\mathrm{Hz}^{-1} \tag{21}$$

(taking $H_0 = 70\,\mathrm{km/s/Mpc}$, this corresponds to $\Omega_{\rm gw} = 2.8\text{–}5.3\times 10^{-12}$). **[Check the 1/2 for the one-sided spectrum.]**

One of the more promising approaches to detect GW backgrounds with LISA relies on estimating instrument noise levels by way of symmetrized TDI observables that is insensitive to GWs at low frequencies in the LISA band [23]. For realistic LISA orbits, however, the low-frequency behavior of such observables becomes more complicated than discussed in the literature. To simplify the initial investigation of the background-detection problem in data set 3.5, we have therefore approximated LISA as a rigidly rotating triangle with equal and constant armlengths (i.e., Synthetic LISA's `CircularRotating`).

3.6. LISACode

LISACode[29] is a new simulator for LISA that has been develloped at the APC (Paris). Its ambition is to map, as closely as possible, the impact of the different sub-systems on the measurements. LISACode is not a detailed simulator at the engineering level but rather a tool whose purpose is to bridge the gap between the basic principles of LISA and a future, sophisticated end-to-end simulator. This is achieved by introducing, in a realistic manner, most of the ingredients that will influence LISA's sensitivity (orbits, instrumental noise functions, time stamping through the use of Ultra Stable Oscillator functions, phasemeter response functions,...). LISACode also include various internal generators of GW waveforms generators (monochromatic, binary systems, stochastic,...) as well as the application of various TDI combinations. The outputs of LISACode are time series which can be requested either at the phasemeter output level and/or after TDI processing. The inputs can be given both in

ASCII and XML form and the outputs are either ASCII or binary. Many user-defined parameters allow the code to study different configurations of LISA thus helping to analyse the impact of these on its sensitivity.

The conventions used by LISACode follow closely those of the MLDC and of SyntheticLisa. For the MLDC 3.5, LISACode is run by using Python scripts that are included in the lisatools framework, see `lisatools`. LISACode can be obtained either as part of lisatools or by contacting A. Petiteau (`antoine.petiteau@apc.univ-paris7.fr`).

4. Conclusion

We are very excited about the outcome of the first two MLDCs, which have given a convincing demonstration that a significant portion of the LISA science objectives could already be achieved with techniques that are currently in hand. Most of the research groups that participated in Challenge 1 have successfully made the transition to the greater complexity of Challenge 2. Challenge 3 (with data sets released in Jan 2008 and results due in Dec 2008) will continue to move in the direction of more realistic signals, featuring chirping Galactic binaries and precessing binaries of spinning MBHs. It will also include two new classes of signals: an isotropic primordial GW background and bursts from the cusps of cosmic strings. In addition, Challenge 1B took place between July and Dec 2007. This was a repeat of Challenge 1, conceived to provide a softer entry point for research groups new to the MLDCs. Ten collaborations (including five new institutions) participated, demonstrating increasing sophistication and proficiency in a range of LISA data-analysis techniques.

Furthermore, the MLDC conventions, file formats, and software tools (see `lisatools.googlecode.com`) have matured to the point where interested parties can use them to generate a variety of data sets. This enables a wealth of interesting side investigations, such as the studies of the LISA science reach that are now being undertaken by the LISA Science Team. To obtain more information and to participate in the MLDCs, see the official MLDC website (`astrogravs.nasa.gov/docs/mldc`) and the task force wiki (`www.tapir.caltech.edu/listwg1b`). [**These two paragraphs are verbatim from the MLDC-2 report. Need to shorten, rephrase, add more about results.**]

Acknowledgments

MV: the LISA Mission Science Office and by JPL's HRDF. CC's, JC's and MV's work was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration. SB's and EP's work was supported by the Deutsches Zentrum für Luft- und Raumfahrt. JG thanks the Royal Society for support and the Albert Einstein Institute for hospitality and support while part of this work was being completed. IM was partially supported by NASA ATP Grant NNX07AH22G to Northwestern University. LW's work is supported by the Alexander von Humboldt Foundation's Sofja Kovalevskaja Programme funded by the German Federal Ministry of Education and Research. MT acknowledges the support of the Spanish Ministerio de Eduación y Ciencia Research projects FPA-2007-60220, CSD207-00042 and the Govern de les Illes Balears, Conselleria d'Economia, Hisenda i Innovació.

References

[1] Bender P, Danzmann P and the LISA Study Team 1998 "Laser Interferometer Space Antenna for the Detection of Gravitational Waves, Pre-Phase A Report" **MPQ 233** (Garching: Max-Planck-Institüt für Quantenoptik)

[2] Arnaud K A et al. (the MLDC Task Force) 2006 *Laser Interferometer Space Antenna: 6th International LISA Symp. (Greenbelt, MD, 19–23 Jun 2006)* ed Merkowitz S M and Livas J C (Melville, NY: AIP) p 619; *ibid.* p 625

[3] Arnaud K A et al. (the MLDC Task Force and Challenge 1 participants) 2007 *Class. Quant. Grav.* **24** S529

[4] Barack L and Cutler C 2004 *Phys. Rev.* D **69** 082005

[5] Arnaud K A et al. (the MLDC Task Force) 2007 *Class. Quant. Grav.* **24** S551

[6] P. Jaranowski, A. Królak, and B. F. Schutz, Phys. Rev. D **5**8, 063001 (1998).

[7] A. Królak, M. Tinto, and M. Vallisneri, *Phys. Rev. D*, **70**, 022003 (2004).

[8] Prix R and Whelan J T 2007 *Class. Quant. Grav.* **24** S565; *Poster* `www.ligo.caltech.edu/docs/G/G070462-00.pdf`

[9] Cornish N J and Porter E K 2007 *Class. Quantum Grav.* **24** 5729–5755

[10] Brown D A, Crowder J, Cutler C, Mandel I and Vallisneri M 2007 *Class. Quant. Grav.* **24** S595

[11] Vallisneri M, Crowder J and Tinto M 2007 "Sensitivity and parameter-estimation precision for alternate LISA configurations" *Preprint* arxiv.org/0710.4369

[12] Gair J R, Mandel I and Wen L 2007 Proc. 7th Amaldi Conf. on Gravitational Waves (Sydney, 8–14 July 2007), submitted. *Preprint* arXiv:0710.5250

[13] G. Nelemans, L. R. Yungelson & S. F. Portegies Zwart, A&A **375**, 890 (2001).

[14] G. Nelemans, L. R. Yungelson & S. F. Portegies Zwart, Mon. Not. Roy. Astron. Soc. **349** 181, (2004).

[15] N. J. Cornish and T. B. Littenberg, Phys. Rev. D **76**, 083006 (2007).

[16] G. H. A. Roelofs, G. Nelemans and P. J. Groot, "The population of AM CVn stars from the Sloan Digital Sky Survey," arXiv:0709.2951 [astro-ph].

[17] T. Damour and A. Vilenkin, Phys. Rev. Lett. **85**, 3761 (2000).

[18] X. Siemens, J. Creighton, I. Maor, S. Ray Majumder, K. Cannon, and J. Read, Phys. Rev. D **73** 105001 (2006).

[19] X. Siemens and K. D. Olum, Phys. Rev. D **68**, 085017 (2003).

[20] B. Allen and J. D. Romano, Phys. Rev. D **59**, 102001 (1999).

[21] K. M. Górski et al., Astrophys. J. **622**, 759 (2005).

[22] S. Plaszczynski, Fluctuation and Noise Letters **7**, R1 (2007) [arXiv:astro-ph/0510081].

[23] M. Tinto, J. W. Armstrong and F. B. Estabrook, Phys. Rev. D 63, 021101 (2000).

[24] T. Damour, B. Iyer and B. Sathyaprakash, Phys. Rev. D **63**, 044023 (2001).

[25] T. A. Apostolatos, C. Cutler, G. J. Sussman and K. S. Thorne, Phys. Rev. D **49**, 6274 (1994)

[26] R. N. Lang and S. A. Hughes, Phys. Rev. D **74**, 122001 (2006)

[27] L. Kidder, Phys. Rev. D **52** 821 (1995)

[28] J. R. Gair, I. Mandel and L. Wen "Improved time-frequency analysis of extreme-mass-ratio inspiral signals in mock LISA data," proceedings of GWDAW-12, Dec 13–16 2007, Cambridge MA.

[29] M. Trias, A. Vecchio and J. Veitch, "Markov Chain Monte Carlo searches for Galactic binaries in MLDC-1B data sets,"

[30] A. Petiteau, G. Auger, H. Halloin, O. Jeannin, E. Plagnol, S. Pireaux, T. Regimbeau and J.-Y. Vinet, Phys. Rev. D **77**, 023002 (2008).