The Mock LISA Data Challenges: from Challenge 3 to Challenge 4

The *Mock LISA Data Challenge Task Force*: Stanislav Babak[1], John G. Baker[2], Matthew J. Benacquista[3], Neil J. Cornish[4], Shane L. Larson[5], Ilya Mandel[6], Antoine Petiteau[1], Edward K. Porter[7], Emma L. Robinson[1], Michele Vallisneri[8,9], Alberto Vecchio[10], and *the Challenge 3 participants*: Matt Adams[4], Keith A. Arnaud[2], Arkadiusz Błaut[11], Michael Bridges[12], Michael Cohen[9], Curt Cutler[8,9], Farhan Feroz[12], Jonathan R. Gair[13], Philip Graff[12], Mike Hobson[12], Joey Shapiro Key[4], Andrzej Królak[14], Anthony Lasenby[12,15], Reinhard Prix[16], Yu Shang[1], Miquel Trias[17], John Veitch[10], John T. Whelan[18]

[1] Max-Planck-Institut für Gravitationsphysik (Albert-Einstein-Institut), Am Mühlenberg 1, D-14476 Golm bei Potsdam, Germany
[2] Gravitational Astrophysics Lab., NASA Goddard Space Flight Center, 8800 Greenbelt Rd., Greenbelt, MD 20771, USA
[3] Center for Gravitational Wave Astronomy, Univ. of Texas at Brownsville, Brownsville, TX 78520, USA
[4] Dept. of Physics, Montana State Univ., Bozeman, MT 59717, USA
[5] Dept. of Physics, Utah State Univ., Logan, UT 84322, USA
[6] Dept. of Physics and Astronomy, Northwestern Univ., Evanston, IL, USA
[7] APC, UMR 7164, Univ. Paris 7 Denis Diderot, 10, rue Alice Domon et Leonie Duquet, 75025 Paris Cedex 13, France
[8] Jet Propulsion Laboratory, California Inst. of Technology, Pasadena, CA 91109, USA
[9] Theoretical Astrophysics, California Inst. of Technology, Pasadena, CA 91125
[10] School of Physics and Astronomy, Univ. of Birmingham, Edgbaston, Birmingham B152TT, UK
[11] Inst. of Theoretical Physics, Univ. of Wrocław, Wrocław, Poland
[12] Astrophysics Group, Cavendish Laboratory, Univ. of Cambridge, Cambridge CB30HE, UK
[13] Inst. of Astronomy, Univ. of Cambridge, Cambridge CB30HA, UK
[14] Inst. of Mathematics, Polish Academy of Sciences, Warsaw, Poland
[15] Kavli Institute for Cosmology, Univ. of Cambridge, Cambridge CB30HA, UK
[16] Max-Planck-Institut für Gravitationsphysik (Albert-Einstein-Institut), D-30167 Hannover, Germany
[17] Departament de Física, Univ. de les Illes Balears, Cra. Valldemossa Km. 7.5, E-07122 Palma de Mallorca, Spain
[18] Center for Computational Relativity and Gravitation & School of Math. Sciences, Rochester Inst. of Technology, 85 Lomb Memorial Drive, Rochester, NY 14623, USA

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 one or more datasets containing simulated instrument noise and gravitational waves from sources of undisclosed parameters. Participants analyze the datasets and report best-fit solutions for the source parameters.

Here we present the results of the third challenge, issued in Apr 2008, which demonstrated the positive recovery of signals from chirping Galactic binaries, from spinning supermassive–black-hole binaries (with optimal SNRs between ~ 10 and 2000), from simultaneous extreme–mass-ratio inspirals (SNRs of 10–50), from cosmic-string–cusp bursts (SNRs of 10–100), and from a relatively loud isotropic background with $\Omega_{\mathrm{gw}}(f) \sim 10^{-11}$, slightly below the LISA instrument noise.

1. The Mock LISA Data Challenges

The Laser Interferometer Space Antenna (LISA) is a planned NASA–ESA gravitational-wave (GW) observatory sensitive in the 10^{-5}–10^{-1} Hz range [1]. LISA's data will contain superposed signals from millions of sources, including all the binaries in the Galaxies with orbital periods below five hours and massive–black-hole (MBH) binary coalescences out to $z \sim 20$. Thousands of sources will be resolvable individually. The potential for source confusion and the very complex dynamics and waveforms of sources such as extreme–mass-ratio inspirals (EMRIs), suggested the need for a coordinated effort to develop and demonstrate LISA data-analysis capabilities. The Mock LISA Data Challenges (MLDCs) began in early 2006 with this very purpose.

The complexity and ambition of the challenges has risen with each round: MLDC 1 [2, 3] focused on simple sources, isolated or moderately interfering; MLDC 2 [4, 5] introduced a Galactic *ensemble* of 26 million binaries (20,000 of which were successfully recovered), as well as the problem of detecting MBH binaries over the Galactic cacophony; MLDC 1B [6] reprised the first challenge for new research collaborations joining the effort, and saw the first successful detections of EMRI signals. MLDC 3, released in April 2008 and due in April 2009, consisted of five subchallenges that featured more realistic models of previously examined sources (chirping Galactic binaries in MLDC 3.1, spinning MBH binary inspirals in 3.2, superposed EMRIs in 3.3) and entirely new sources (GW bursts from cosmic-string cusps in 3.4, an isotropic stochastic background in 3.5); see [6] for a detailed discussion of the source models and GW content of each subchallenge. Fifteen collaborations, comprising all the participants listed in the byline and most task-force members, submitted a total of seventeen entries (all can be found at `www.tapir.caltech.edu/~mldc/results3` together with technical notes about search implementation).

In this paper we briefly report on the detection and parameter-estimation performance demonstrated by each entry. Altogether, MLDC 3 showed substantial progress in tackling increasingly difficult data-analysis problems, and introduced new search methods such as nested sampling and sophisticated genetic optimization–Markov Chain hybrids. However, there is certainly room for improvement and further work: fewer Galactic binaries were recovered by the searches employed here than by the multi-source MCMC demonstrated in MLDC 2; MBH binaries and EMRIs were detected with high confidence, but the accurate estimation of their parameters (beyond the dominant ones) was stymied by the complex global structure and the many local maxima of likelihood surfaces. On the bright side, searches for cosmic-string bursts and stochastic backgrounds (admittedly simple problems in the absence of nonstationary or non-Gaussian instrument noise) met no roadblocks.

Section 7 of this article introduces MLDC 4, which is being released as we write (November 2009), with entries due at the end of 2010.

2. Galactic Binaries (MLDC 3.1)

Challenge dataset 3.1 contained signals from over 60 million chirping Galactic binaries. The vast majority of these are too weak to be isolated, and the unresolved component forms a nonstationary confusion noise that adds to the overall noise level. Estimates based on self-consistent removal schemes [7] and Bayesian model selection [8] suggest that it should be possible to recover between 20,000 and 30,000 binaries. Three groups submitted source catalogs for MLDC 3.1:

- **BhamUIB** (a collaboration between the Universities of Birmingham and the Balearic Islands) implemented a delayed-rejection MCMC algorithm [9] to search three narrow frequency windows, $0.3\,\mathrm{mHz} \leq f \leq 0.4\,\mathrm{mHz}$, $0.9\,\mathrm{mHz} \leq f \leq 1.0\,\mathrm{mHz}$, and $1.6\,\mathrm{mHz} \leq f \leq 1.7\,\mathrm{mHz}$, using the MLDC waveform generator [10]. A total of 494 sources were reported.
- **AEIRIT** (researchers at the Albert Einstein Institute in Golm, Germany, and the Rochester Institute of Technology) set up a LIGO-style hierarchical, sequential search based on the $\mathcal{F}$-statistic and on frequency-domain rigid-adiabatic templates [11]. Triggers are generated for the individual TDI channels; those found in coincidence are analyzed coherently using noise-orthogonal TDI observables. A total of 1940 sources were reported.
- **PoWrWa** (a collaboration between the Albert Einstein Institute, the University of Wrocław, and the Polish Academy of Sciences) adopted a sequential matched-filtering search that used the $\mathcal{F}$-statistic and rigid-adiabatic templates [12], and analyzed a few 0.1-mHz wide frequency bands. The brightest source in each band is identified and removed, and the process repeated until a pre-set SNR threshold is reached. A total of 14,838 sources were reported.

The entries for this round fell short of the theoretical target for a variety of reasons. BhamUIB analyzed only a small fraction of the data, while AEIRIT and PoWrWa used sequential removal schemes, which are limited in how deep they can dig before source confusion degrades signal recovery. Previous studies [8] indicate that it should be possible to recover approximately 99% of the resolvable sources to an accuracy of better than 90%, as measured by the overlap between injected and recovered waveforms. We therefore adopt correlation as the metric by which the entries are measured.

For this we need to identify the injected signal that corresponds most closely to each recovered signal. To do this, we consider all injected signals with SNR > 3 within six frequency bins of the recovered signal, and we select the injected signal with the minimum $\chi^2 = (h - h'|h - h')$, where h and h' are the injected and recovered waveforms, and $(a|b)$ denotes the standard noise-weighted inner product, summed over the noise-orthogonal A, E, and T TDI channels. The correlation is then computed as $C = (h|h')/(h|h)^{1/2}(h'|h')^{1/2}$. Using as a figure of merit the percentage of recovered sources with $C > 0.9$, we find 30% for BhamUIB, 95% for AEIRIT, 33% for PoWrWa. PoWrWa have reported a bug in their code that affects sources with frequencies above 3 mHz. Results are better after correcting it [12]. If we consider only the 6,955 sources they report below 3 mHz, their figure of merit improves to 58%.

3. Massive black-hole binaries (MLDC 3.2)

Challenge dataset 3.2 contained five signals from MBH inspirals, embedded in instrument noise and a partially subtracted Galactic background (see [6] for a synopsis

of the waveform model and for the random selection of source parameters). Multi-TDI SNRs ranged from ~ 13 to ~ 1671, with the two weakest signals (MBH-2 and MBH-6) corresponding to mergers after the end of the dataset. Five groups submitted entries:

- **AEI** (researchers at the Albert Einstein Institute) used a genetic matched-filtering algorithm extended to a multi-modal search; as a merit function they took the A-statistic, a geometrical mean of the log-likelihood for the signal and for a low-frequency signal subset [13].

- **CambAEI** (a collaboration between the Cambridge Cavendish Laboratory and Institute of Astronomy with the the Albert Einstein Institute) adopted the MultiNest implementation of nested-sampling integration [14] to compute the evidence and produce posterior distributions. This method is intrinsically multimodal, and the A-statistic was used to identify modes.

- **GSFC** (researchers at Goddard Space Flight Center) employed the tempered–Metropolis-Hastings MCMC algorithm found in Xspec [15].

- **JPLCITNWU** (a collaboration between JPL/Caltech and Northwestern University) employed a two-stage method whereby they followed up a nonspinning MBH search [16] with MultiNest.

- **MTGWAGAPC** (researchers at Montana State University and at the Astro-Particle and Cosmology Institute, APC, in Paris) adopted a parallel-tempering MCMC algorithm, evolved from the thermostated, frequency-annealed algorithm employed in previous challenges [17].

All five groups recovered parameters for the loud signals; MBH-2 (SNR ~ 19) was recovered by AEI, CambAEI, JPLCITNWU and MTGWAGAPC; MBH-6 (SNR ~ 13) by AEI, CambAEI and MTGWAGAPC. Participants were encouraged to submit multiple modes of comparable probability, if present. For the best (highest-SNR) mode in each entry, Table 1 lists the fractional parameter errors, as well as the recovered SNR and individual TDI-channel fitting factors (FFs). Both the recovered SNR and the SNR_{true} of the true waveform were computed by filtering the noisy dataset (in its LISA Simulator version) with the appropriate template, while the FFs were computed between noiseless signals. Note that the reported modes had often SNR differences of less than one, and the highest-SNR mode did not always have the best parameters.

The binaries that merge within the duration of the dataset (MBH-1, 3, and 4) have significantly larger SNR and reach higher frequencies, and thus allow a much more accurate recovery of source parameters. For these binaries, the mass, time-of-coalescence, distance, and sky-position errors are comparable, if not better, to those of nonspinning-binary searches (see Tables 1 and 2 of [5]). The errors in the spin amplitudes are consistent with the Fisher-matrix results of [18]. However, despite the high SNR, the estimation of the initial direction of the spins and of the orbital angular momentum has proved to be very difficult. We observe a large number of the local likelihood maxima that have comparable, high SNR (and FF > 0.99), yet very different values of these parameters, as illustrated in Fig. 1 for the MBH-3 entries.

Parameter estimation was not as successful for the weaker signals with mergers beyond the end of the dataset. Still, for MBH-2 the errors in the masses and time of coalescence are comparable to the Fisher-matrix predictions. The errors in sky position are ~ 10 deg, with strong local likelihood maxima at the antipodal sky position. Spin amplitudes are determined very poorly; this reflects the fact that the spins are nearly degenerate with other parameters in the low-frequency part of the waveforms.

Table 1. Selected parameter errors, SNRs, and FFs for each group's highest-SNR entries to MLDC 3.2. The time of coalescence t_c, spin magnitudes $a_{1,2}$ and luminosity distance D are defined in Table 7 of [6]; in addition, the (redshifted) chirp mass $M_c \equiv (m_1 m_2)^{3/5}/(m_1 + m_2)^{1/5}$, and the symmetric mass ratio $\eta = m_1 m_2/(m_1 + m_2)^2$. Δsky is the angular geodesic distance between the estimated and true positions; values ~ 180 deg correspond to the antipodal sky location, a known quasi-degeneracy in the LISA response.

source (SNR_{true})	group	$\Delta M_c/M_c$ $\times 10^{-5}$	$\Delta\eta/\eta$ $\times 10^{-4}$	Δt_c (sec)	Δsky (deg)	Δa_1 $\times 10^{-3}$	Δa_2 $\times 10^{-3}$	$\Delta D/D$ $\times 10^{-2}$	SNR	FF_A	FF_E
MBH-1	AEI	2.4	6.1	62.9	11.6	7.6	47.4	8.0	1657.71	0.9936	0.9914
(1670.58)	CambAEI	3.4	40.7	24.8	2.0	8.5	79.6	0.7	1657.19	0.9925	0.9917
	MTAPC	24.8	41.2	619.2	171.0	13.3	28.7	4.0	1669.97	0.9996	0.9997
	JPL	40.5	186.6	23.0	26.9	39.4	66.1	6.9	1664.87	0.9972	0.9981
	GSFC	1904.0	593.2	183.9	82.5	5.7	124.3	94.9	267.04	0.1827	0.1426
MBH-3	AEI	9.0	5.2	100.8	175.9	6.2	18.6	2.7	846.96	0.9995	0.9989
(847.61)	CambAEI	13.5	57.4	138.9	179.0	21.3	7.2	1.5	847.04	0.9993	0.9993
	MTAPC	333.0	234.1	615.7	80.2	71.6	177.2	16.1	842.96	0.9943	0.9945
	JPL	153.0	51.4	356.8	11.2	187.7	414.9	2.7	835.73	0.9826	0.9898
	GSFC	8168.4	2489.9	3276.9	77.9	316.3	69.9	95.6	218.05	0.2815	0.2314
MBH-4	AEI	4.5	75.2	31.4	0.1	47.1	173.6	9.1	160.05	0.9989	0.9994
(160.05)	CambAEI	3.2	171.9	30.7	0.2	52.9	346.1	21.6	160.02	0.9991	0.9992
	MTAPC	48.6	2861.0	5.8	7.3	33.1	321.1	33.0	149.98	0.8766	0.9352
	JPL	302.6	262.0	289.3	4.0	47.6	184.5	28.3	158.34	0.8895	0.9925
	GSFC	831.3	1589.2	1597.6	94.4	59.8	566.7	95.4	-45.53	-0.1725	-0.2937
MBH-2	AEI	1114.1	952.2	38160.8	171.1	331.7	409.0	15.3	20.54	0.9399	0.9469
(18.95)	CambAEI	88.7	386.6	6139.7	172.4	210.8	130.7	24.4	20.36	0.9592	0.9697
	MTAPC	128.6	45.8	16612.0	8.9	321.4	242.4	13.1	20.27	0.9228	0.9260
	JPL	287.0	597.7	11015.7	11.8	375.3	146.3	9.9	18.69	0.9661	0.9709
MBH-6	AEI	1042.3	1235.6	82343.2	2.1	258.2	191.6	26.0	13.69	0.9288	0.9293
(12.82)	CambAEI	5253.2	1598.8	953108.0	158.3	350.8	215.4	29.4	10.17	0.4018	0.4399
	MTAPC	56608.7	296.7	180458.8	119.7	369.2	297.6	25.1	11.34	-0.0004	0.0016

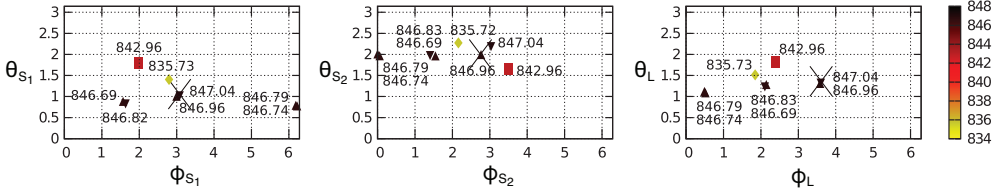**Figure 1.** Distribution of reported modes for MBH-3 along the spin and orbital–angular-momentum angles. Each mode is annotated with the SNR. ×: true value; ▲: AEI; ▼: CambAEI; ■: MTGWAGAPC; ♦: JPLCITNWU.

Lang and Hughes [18] report that spin-induced modulations remove correlations between parameters in Fisher-matrix computations, improving overall parameter determination. However, here we observe that spin interactions also cause nonlocal degeneracies in parameters space, especially so for spin and orbital–angular-momentum angles. Further investigations are needed to determine which phenomenon is stronger. Nevertheless, the entries to MLDC 3.2 demonstrate a solid detection capability for spinning-binary inspirals, and a good recovery of most source parameters.

4. Extreme–mass-ratio inspirals (MLDC 3.3)

Challenge dataset 3.3 contained five Barack–Cutler [19] EMRI signals immersed in instrument noise (see [6] for details about the waveforms and the random choice of parameters). In comparison to previous EMRI challenges, here participants had to contend with multiple simultaneous sources, as well as weaker signals—the injected SNRs varied between 20 and 37. Three groups submitted entries:

- **BabakGair** (a collaboration between the Albert Einstein Institute and Cambridge University) used stochastic sampling and MCMC to identify EMRI harmonics, then carried out an $\mathcal{F}$-statistic search in the space of harmonics, and performed a final MCMC fit in source parameter space. This search improved on the method described in [20], adding more sophisticated harmonic identification.

- **EtfAG** (researchers at Cambridge and Northwestern Universities) searched for harmonics in the time–frequency spectrogram using the Chirp-based Algorithm for Track Search (CATS, [21]) developed for earlier MLDCs [22], and improved for this search to deal with intersecting tracks from multiple sources.

- **MTAPCIOA** (a collaboration between Montana State University, APC–Paris, and Cambridge University) improved the MLDC algorithm used in previous rounds [23] to include parallel tempering as described in [24], and to enhance the implementation of "harmonic jumps" between secondary likelihood maxima.

Parameter-estimation errors are presented in Table 2. Altogether, all EMRIs were found by at least one group, and masses were estimated accurately; more work remains to be done on improving the estimates of EMRI parameters for relatively weak and overlapping signals.

MTAPCIOA recovered all five EMRI signals, generally with very good parameter-estimation accuracies: errors of a few tenths % in the masses of both bodies, and sky-position errors of a few deg (13 for EMRI-4). Their second solution for EMRI-1 was particularly impressive, with fractional errors of a few 10^{-5} in masses, initial eccentricity, and spin. However, their other solution exemplifies the difficulty of resolving secondary maxima: the SNR is almost the same (21.794 vs. 21.804), but parameter errors are two orders of magnitude greater.

The time–frequency analysis carried out by EtfAG was particularly hard-hit by the simultaneous lowering of the SNR and the presence of multiple overlapping signals in MLDC 3.3. The group was unable to find the low-frequency, low-SNR EMRI-1; while EtfAG did find a medium-frequency source, the relatively large parameter-estimation errors suggest that the time–frequency approach did not adequately resolve between the overlapping harmonics of EMRI-2 and EMRI-3.

BabakGair submitted three (relatively close) solutions for each of EMRI-2 and EMRI-3. For EMRI-3, their estimates are better than those of EtfAG and comparable to those of MTAPCIOA, albeit somewhat less accurate for the initial parameter values and distance. For EMRI-2, BabakGair had errors of a few % in the masses and initial eccentricity, and had significant errors in spin-orientation angle and in distance, although sky location was still found correctly, within 3.5 deg.

The high-frequency EMRI-4 and EMRI-5 presented a challenge for all groups. Although MTAPCIOA found them and estimated their masses fairly accurately, the errors in spin magnitude and orientation were significantly larger than for other sources, and the distance to both sources was overestimated by factors of 2 or 3 (vs. errors $\lesssim 10\%$ for the other EMRIs). Furthermore, the negative SNR for claimed

Table 2. Parameter-estimation errors for the EMRIs in MLDC 3.3. M and μ are the masses of the central and inspiraling bodies; ν_0 and e are the initial azimuthal orbital frequency and eccentricity; $|S|$ is the dimensionless central-body spin; λ_{SL} is the spin–orbit misalignment angle, and D the luminosity distance. Δspin and Δsky are the geodesic angular distances between the estimated and true spin direction and sky position. $\mathrm{SNR_{true}}$ is computed with the LISA Simulator; the SNR for each entry with the simulator used in that search (the LISA Simulator [25] for MTAPCIOA, Synthetic LISA [26] for EtfAG and BabakGair).

Source ($\mathrm{SNR_{true}}$)	Group	SNR	$\frac{\Delta M}{M}$ $\times 10^{-3}$	$\frac{\Delta \mu}{\mu}$ $\times 10^{-3}$	$\frac{\Delta \nu_0}{\nu_0}$ $\times 10^{-5}$	Δe_0 $\times 10^{-3}$	$\Delta |S|$ $\times 10^{-3}$	$\frac{\Delta \lambda_{\mathrm{SL}}}{\lambda_{\mathrm{SL}}}$ $\times 10^{-3}$	Δspin (deg)	Δsky (deg)	$\frac{\Delta D}{D}$
EMRI-1 (21.673)	MTAPCIOA	21.794	5.05	3.29	1.61	−5.1	−1.4	−19	23	2.0	0.07
	MTAPCIOA	21.804	−0.06	−0.01	−0.08	−0.05	0.02	0.54	3.5	1.0	0.13
EMRI-2 (32.935)	MTAPCIOA	32.387	−3.64	−2.61	−3.09	3.8	0.87	12	11	3.7	3×10^{-3}
	BabakGair	22.790	33.1	−19.7	10.1	−33	−7.3	250	47	3.5	−0.25
	BabakGair	22.850	32.7	−20.0	9.94	−32	−7.2	250	58	3.5	−0.24
	BabakGair	22.801	33.5	−19.5	10.5	−33	−7.4	240	40	3.5	−0.25
EMRI-3 (19.507)	MTAPCIOA	19.598	1.62	0.38	−0.10	−0.35	−0.94	−3.0	5.0	3.0	−0.04
	BabakGair	21.392	1.77	1.01	1.95	−1.2	−0.68	−2.3	116	4.5	0.13
	BabakGair	21.364	2.26	1.88	2.71	−2.0	−0.69	−2.5	65	6.1	0.14
	BabakGair	21.362	1.51	1.01	2.09	−1.3	−0.50	−1.7	7.6	6.2	0.14
	EtfAG	—	54.0	4.88	−7375	26	17	—	—	32	0.83
EMRI-4 (26.650)	MTAPCIOA	−0.441	−8.77	−10.1	−6.03	−3.7	144	950	99	13	−2.3
EMRI-5 (36.173)	MTAPCIOA	17.480	−3.32	5.00	−1.80	0.22	55	62	43	1.8	−1.3

EMRI-4 and the low FFs between the recovered and injected noiseless waveforms indicate that the MTAPCIOA search could not resolve these sources individually, but converged on two parameter sets that jointly fit the combination of the two injected sources.

5. Cosmic-string–cusp bursts (MLDC 3.4)

Challenge dataset 3.4 contained three burst signals from cosmic-string cusps, immersed in instrument noise with slightly randomized levels for each individual noise (i.e., from the six proof masses and photodetectors). The dataset was less than a month long (2^{21} s), with a higher sampling rate (1 s) than the others, to accommodate the potential high-frequency content of these signals, which have power-law spectrum up to an f_{max} determined by the characteristic length scale of the string and the viewing angle (see [6] for more details about the waveforms and the random choice of their parameters). Four collaborations submitted entries:

- **CAM** (a collaboration between Cambridge U. and APC–Paris) used MultiNest.
- **CaNoe** (researchers at Cambridge and Northwestern Universities) implemented a time–frequency algorithm, a modified version of CATS [21].
- **JPLCIT** (Caltech/JPL) experimented with MCMC and MultiNest, but only submitted entries based on the latter [27].
- **MTGWAG** (Montana State University) used a parallel-tempering MCMC [24].

All groups successfully recovered all three bursts. Table 3 shows the parameter-estimation errors, SNRs, and FFs for all entries. Although the accuracy of parameters is poor, the FFs are very high; this suggests that these results are not due to

Table 3. Select parameter errors, SNRs, and FFs for the MLDC 3.4 entries. Here Δsky is the geodesic angular distance between the recovered and true sky positions; t_D is the time of burst arrival at LISA; ψ and $\mathcal{A}$ are the GW polarization and amplitude. CaNoE reported only t_C and $\mathcal{A}$, so Δsky, Δt_D, $\Delta\psi$, SNR, and FFs cannot be computed.

source ($\mathrm{SNR_{true}}$)	group	Δsky (deg)	Δt_D (sec)	$\Delta\psi$ (rad)	$\Delta\mathcal{A}/\mathcal{A}$	SNR	FF_A	FF_E
String-1	CAM	106.9	1.462	0.501	0.904	43.706	0.99947	0.99797
(43.46)	CAM	49.4	2.331	1.065	1.128	43.520	0.99964	0.99591
	JPLCIT	34.2	1.585	3.726	0.413	43.506	0.99986	0.99844
	JPLCIT	113.7	1.574	3.739	0.431	43.497	0.99988	0.99847
	MTGWAG	106.6	2.071	2.600	0.745	43.287	0.99975	0.99565
String-2	CAM	82.0	3.683	4.846	0.062	33.690	0.99945	0.99986
(33.6)	JPLCIT	90.5	4.005	4.268	0.282	33.689	0.99949	0.99929
	JPLCIT	45.2	3.847	6.364	0.231	33.694	0.99939	0.99960
	MTGWAG	53.1	3.223	0.158	0.011	33.696	0.99926	0.99978
String-3	CAM	80.8	1.249	3.785	0.338	41.326	0.99073	0.99923
(41.42)	CAM	133.3	1.715	3.257	0.238	41.456	0.99388	0.99869
	CAM	44.5	0.763	3.202	0.066	41.142	0.99700	0.99883
	JPLCIT	59.0	1.546	3.129	0.317	41.315	0.99554	0.99848
	JPLCIT	157.7	1.226	5.614	0.220	41.316	0.99717	0.99864
	MTGWAG	137.9	0.980	0.110	0.161	41.418	0.99327	0.99948

shortcomings in the search methods, but to the very character of the waveforms. As a matter of fact, for relatively short signals such as these bursts, LISA can be considered as a static detector, and its response is not imprinted with any modulations from the LISA orbit or from the rotation of its constellation. Hence the sky position of burst sources can only be determined by triangulation between the spacecraft—a weaker effect, and one that vanishes in the limit of long wavelengths.

Thus, the determination of sky position is intrinsically harder for bursts, and it is further complicated by the presence of degeneracies [24] such as the reflection of sky position across the instantaneous LISA plane. The bursts' barycentric central time t_C and its polarization are strongly coupled with sky position, and therefore are also determined poorly. To compensate for this fact, we calculate the error in the arrival time of the burst at the center of the detector constellation t_D. This parameter has a weaker correlation with the sky position and constrained better by observations.

6. Stochastic background (MLDC 3.5)

Challenge dataset 3.5 contained an isotropic stochastic background signal immersed in instrument noise, with noise levels slightly randomized as in MLDC 3.4. To facilitate the use of the GW-canceling TDI combination T, the LISA orbits were approximated as those traced by a rigidly rotating triangle, with equal and constant arm lengths (except for the Sagnac effect). The background was realized by placing 192×2 linearly polarized, stochastic pseudosources at positions uniformly distributed across the sky. See [6] for more details. Now, isotropic stochastic backgrounds can be characterized by a single dimensionless quantity,

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

where $\rho_{\mathrm{gw}}(f)$ is the energy density in GWs, and $\rho_{\mathrm{crit}} = 3c^2 H_0^2/8\pi G$ is the energy density required to close the Universe. In this MLDC round, Ω_{gw} was taken to be

constant across frequencies, and in this dataset it was set equal to 1.123×10^{-11} (with $H_0 = 70$km/s/Mpc). Two groups submitted entries for this challenge, and both analyzed the same version of the dataset, generated by Synthetic LISA [26]:

- **AEIBham** (a collaboration between the Albert Einstein Institute and the University of Birmingham) submitted multiple entries using different applications of the same MCMC algorithm. Here we focus on the AEIBham analysis of the TDI A and E observables, which was run separately over frequency bands of 0.1–1 mHz (a) and 0.1–5 mHz (b).
- **MTGWAG** (Montana State University) analyzed the TDI A, E, and T spectra with parallel-tempering MCMC. They estimated the background level as well as the levels of select linear combinations of the individual LISA noises (other combinations are left essentially undetermined by TDI observations).

Figure 2 shows the reported posterior PDFs and the best-fit values for Ω_{gw}, along with their fractional error; both groups recovered the injected value within less than 10%. Figure 2 shows also the best-fit levels reported by MTGWAG for the determinable proof-mass and photodetector noise combinations. The AEIBham analyses used theoretical expressions for TDI A and E noise spectra, which left only their overall levels as unknowns. However, this treatment assumed that the secondary noises were symmetric, which was not the case in this dataset. In addition, because of limitations in the MLDC stochastic-background generation code, the actual GW strain spectrum differed from the nominal f^{-3} power law at higher frequencies. This may explain why the AEIBham (b) result was further from the injected value that either the AEIBham (a) result (in which band the GW spectrum was actually f^{-3}) or the MTGWAG result, which used all three TDI channel, solved for their unknown correlations, and calibrated the GW spectrum using training datasets.

7. Moving forward: a synopsis of MLDC 4

While the third round of the MLDCs was focused on increasing the complexity and variety of GW sources, we are devoting the next iteration to the *global-fit* problem of detecting and analyzing sources of different types superposed in the LISA data. Thus, MDLC 4 consists of a single, "whole enchilada" challenge that includes all the sources of MLDCs 3.1–3.5 *in the same dataset*, albeit with larger source numbers (for EMRIs and cosmic-string bursts) and parameter ranges (for MBH binaries and EMRIs). See Table 4 for details. The duration of the dataset is again approximately two years ($2^{22} \times 15$ s), but the data are released both as a high-cadence time series of 2^{25} samples with $\Delta t = 1.875$ s, and as a downsampled time series of 2^{22} samples with $\Delta t = 15$ s (however, only the cosmic-string bursts and instrument noise contribute spectral content above 33 mHz).

Three versions of both datasets are released: fractional-frequency–fluctuation data generated by Synthetic LISA [26] and LISACode [28], and equivalent-strain data generated by the LISA Simulator [25]. In contrast to MLDCs 3.4 and 3.5, all the LISA instrument noises are set to symmetric, nominal levels, and the LISA armlength are allowed to "breathe." See [4, 6] for more details about the MLDC models of LISA orbits, noises, and interferometric observables. The waveform models of MLDC 3 [6] are used unchanged in MLDC 4, except for the following: the spectra of cosmic-string bursts are truncated below 10^{-5} Hz; the stochastic-background spectrum is f^{-3} between 10^{-5} Hz and 33 mHz, and drops lower below and above that range; last, when
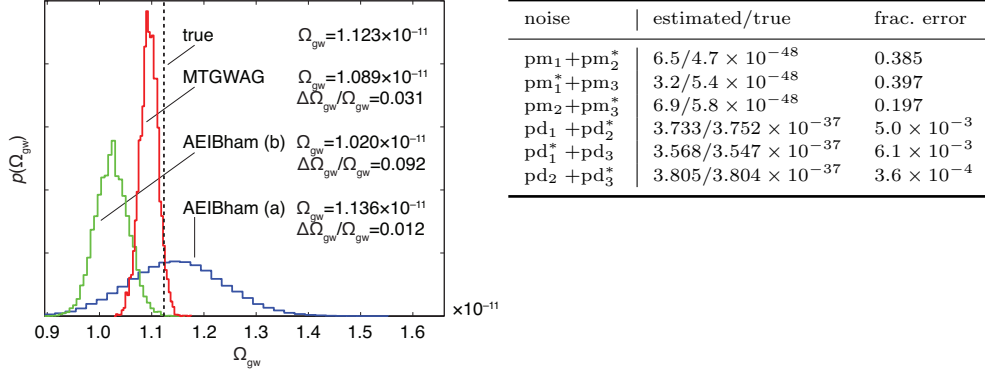
noise	estimated/true	frac. error
$pm_1 + pm_2^*$	$6.5/4.7 \times 10^{-48}$	0.385
$pm_1^* + pm_3$	$3.2/5.4 \times 10^{-48}$	0.397
$pm_2 + pm_3^*$	$6.9/5.8 \times 10^{-48}$	0.197
$pd_1 + pd_2^*$	$3.733/3.752 \times 10^{-37}$	5.0×10^{-3}
$pd_1^* + pd_3$	$3.568/3.547 \times 10^{-37}$	6.1×10^{-3}
$pd_2 + pd_3^*$	$3.805/3.804 \times 10^{-37}$	3.6×10^{-4}

Figure 2. Left: posterior PDFs, their modes, and corresponding fractional error w.r.t. the true Ω_{gw} reported by each analysis. The horizontal-axis range corresponds to the "prior" MLDC range. Right: combinations of LISA noise levels recovered by MTGWAG; the combinations correspond to the total acceleration noise (pm) or position noise (pd) in a single LISA arm.

Table 4. MLDC 4 source content and parameter selection (cf. Table of [6]; figures in **bold** have changed compared to MLDC 3). Parameters are sampled uniformly from the ranges given below; angular parameters are drawn from appropriate sphere-uniform distributions. Source distances are set from the randomly selected multi-TDI SNRs, approximated as $\sqrt{2} \times \max\{\mathrm{SNR}_X, \mathrm{SNR}_Y, \mathrm{SNR}_Z\}$. Poisson($\lambda$) denotes a Poisson-distributed deviate with mean λ.

Galactic-binary background	$\sim 34 \times 10^6$ interacting, $\sim 26 \times 10^6$ detached systems
4–6 MBH binaries	$m_1 = \mathbf{0.5}$–$5 \times 10^6\,M_\odot$, $m_1/m_2 = 1$–$\mathbf{10}$, $a_1/m_1 = 0$–1, $a_2/m_2 = 0$–1, with t_c and SNRs as in MLDC 3.2
an average of 6 *EMRIs* ...including	$\mu = 9.5$–$10.5\,M_\odot$, $S = 0.5$–$0.7\,M^2$, $e_{\mathrm{plunge}} = \mathbf{0.05}$–$0.25$, $t_{\mathrm{plunge}} = 2^{21}$–$2^{22} \times 15$ s, $\mathrm{SNR} = \mathbf{25}$–$\mathbf{50}$ **Poisson(2)** systems with $M = 0.95$–$1.05 \times 10^7 M_\odot$ **Poisson(2)** systems with $M = 4.75$–$5.25 \times 10^6 M_\odot$ **Poisson(2)** systems with $M = 0.95$–$1.05 \times 10^6 M_\odot$
Poisson(20) *cosmic-string bursts*	$f_{\mathrm{max}} = 10^{-3\text{–}1}$ Hz, $t_C = \mathbf{0}$–$\mathbf{2^{22}} \times \mathbf{15}$ s, $\mathrm{SNR} = 10$–100
isotropic stochastic background	$S_h^{\mathrm{tot}} = 0.7$–$1.3 \times 10^{-47}(f/\mathrm{Hz})^{-3}$ Hz^{-1}

MBH binaries end within the duration of the dataset, they are terminated at the time t_{max} when the frequency derivative changes sign, indicating that the PN expansion is failing. To reduce spectral leakage, a half Hann window $\cos^2[\pi(t - t_{\mathrm{max}} + \Delta t)/2\Delta t]$ is applied between $t_{\mathrm{max}} - \Delta t$ and t_{max}, with $\Delta t = Q/f_{\mathrm{max}}$, f_{max} the frequency at time t_{max}, and Q the quality factor (lisaXML parameter `Q`) set to 3 for MLDC 4 (setting `Q`=0 yields the old $r = 6M$ termination condition).

Challenges beyond MLDC 4 will feature ever more realistic datasets, including more sophisticated waveform models, such as MBH coalescences with merger and ringdown phases, and "real-mission" noise models and instrument events drawn from the ongoing experimental investigations of the LISA architecture and subsystem. To obtain more information and to participate in the MLDCs, see the official MLDC website, `astrogravs.nasa.gov/docs/mldc`, the Task Force wiki, `www.tapir.caltech.edu/listwg1b`, and the `lisatools` software repository [29].

Acknowledgments

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 (long version with lisaXML description: gr-qc/0609106)

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

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

[5] Babak S et al. (the MLDC Task Force and Challenge 2 Participants) 2008 *Class. Quant. Grav.* **25** 114037

[6] Babak et al. 2008 *Class. Quant. Grav.* **25** 184026

[7] Timpano S E, Rubbo L J and Cornish N J 2006 *Phys. Rev. D* **73** 122001

[8] Crowder J and Cornish N J 2007 *Class. Quant. Grav.* **24** S575

[9] Trias M, Vecchio A and Veitch J 2009 arXiv:0904.2207

[10] Cornish N J and Littenberg T B 2007 *Phys. Rev. D* **76** 083006

[11] Whelan J T, Prix R and Khurana D 2009 arXiv:0908.3766; 2008 *Class. Quantum Grav.* **25** 184029; Prix R and Whelan J T 2007 *Class. Quant. Grav.* **24** S565

[12] Błaut A, Babak S and Królak A 2009 arXiv:0911.3020

[13] Petiteau A, Shang Y, Babak B and Feroz F *in preparation*

[14] Bridges M, Feroz F, Hobson M P 2009 *MNRAS* **398** 1601

[15] Arnaud K A et al., Xspec website, `http://heasarc.gsfc.nasa.gov/xanadu/xspec`

[16] Brown D A et al. 2007 *Class. Quant. Grav.* **24** S595

[17] Cornish N J and Porter E K 2006 *Class. Quant. Grav.* **24** 5729

[18] Lang R N and Hughes S A 2006 Phys. Rev. D **74** 122001

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

[20] Babak S, Gair J R, Porter E 2009 *Class. Quant. Grav.* **26** 135004.

[21] Mandel I and Gair J R *in preparation*

[22] Gair J R, Mandel I, Wen L 2008 *Class. Quant. Grav.* **25** 184031

[23] Cornish N J 2008 arXiv:0804.3323

[24] Shapiro Key J and Cornish N J 2009 *Phys. Rev. D* **79** 043014

[25] Cornish N J and Rubbo L J 2003 *Phys. Rev. D* **67** 022001; erratum-ibid. 029905; LISA Simulator website, `www.physics.montana.edu/lisa`; included in [29]

[26] Vallisneri M 2005 *Phys. Rev. D* **71**; Synthetic LISA website, `www.vallis.org/syntheticlisa`; included in [29]

[27] Cohen M, Cutler C and Vallisneri M *in preparation*

[28] Petiteau A et al. 2008 *Phys. Rev. D* **77** 023002; LISACode website, `www.apc.univ-paris7.fr/LISA-France/analyse.phtml`; included in [29]

[29] LISAtools website and SVN repository, `lisatools.googlecode.com`