IOP PUBLISHING CLASSICAL AND QUANTUM GRAVITY

Class. Quantum Grav. **25** (2008) 184029 (12pp) doi:10.1088/0264-9381/25/18/184029

Improved search for galactic white-dwarf binaries in Mock LISA Data Challenge 1B using an $\mathcal{F}$-statistic template bank

John T Whelan[1], **Reinhard Prix**[2] **and Deepak Khurana**[3]

[1] Max-Planck-Institut für Gravitationsphysik (Albert-Einstein-Institut), D-14476 Potsdam, Germany
[2] Max-Planck-Institut für Gravitationsphysik (Albert-Einstein-Institut), D-30167 Hannover, Germany
[3] Indian Institute of Technology, Kharagpur, West Bengal-721 302, India

E-mail: John.Whelan@aei.mpg.de and Reinhard.Prix@aei.mpg.de

Received 20 April 2008, in final form 30 July 2008
Published 2 September 2008
Online at stacks.iop.org/CQG/25/184029

Abstract

We report on our $\mathcal{F}$-statistic search for white-dwarf binary signals in the Mock LISA Data Challenge 1B (MLDC1B). We focus in particular on the improvements in our search pipeline since MLDC1, namely refinements in the search pipeline and the use of a more accurate detector response (rigid adiabatic approximation). The search method employs a hierarchical template-grid-based exploration of the parameter space, using a coincidence step to distinguish between primary ('true') and secondary maxima, followed by a final (multi-TDI) 'zoom' stage to provide an accurate parameter estimation of the final candidates.

PACS numbers: 95.85.Sz, 95.75.Pq, 95.75.−z, 97.80.−d

1. Introduction

The Mock LISA Data Challenges (MLDCs) [1] have the purpose of encouraging the development of LISA data-analysis tools and assessing the technical readiness of the community to perform gravitational-wave (GW) astronomy with LISA. The rounds so far have been labelled MLDC1 [2], MLDC2 [3] and MLDC1B [4]. The challenges have consisted of several datasets containing different types of simulated sources and LISA noise. The three types of sources are white-dwarf binaries (WDBs), coalescing supermassive black holes (SMBHs) and extreme mass-ratio inspirals (EMRIs). GW signals from WDBs will be long-lasting and (quasi-)monochromatic with slowly-varying[4] intrinsic frequency $f(\tau)$; in this

[4] In fact, the signals in the MLDCs so far have been strictly monochromatic. Frequency evolution is being introduced for the first time in MLDC3.

sense they belong to the class of continuous GWs. In the case of ground-based detectors the typical sources of continuous GWs are spinning neutron stars with non-axisymmetric deformations. One of the standard tools developed for these searches is the $\mathcal{F}$-statistic [5], which corresponds to the generalized log-likelihood ratio.

We have applied this method in our MLDC searches, adapting the LAL/LALApps [6] search code `ComputeFStatistic_v2` used within the LIGO Scientific Collaboration to search for periodic GW signals in data from ground-based detectors such as LIGO and GEO 600. We have previously conducted searches for WDBs on data from MLDC1 [7] and MLDC2 [8, 9]. MLDC1B is a rerun of MLDC1 with different source parameters, and gives us a chance to evaluate improvements in our pipeline since MLDC1. Among the issues encountered in our original MLDC1 analysis were inaccurate determination of a subset of signal parameters due to the use of the long-wavelength (LW) limit in modelling the LISA response, and the lack of a method to distinguish secondary maxima in parameter space from primary peaks of true signals. Both of these aspects have been improved in our MLDC1B pipeline.

2. Continuous gravitational-wave signals

A system with an oscillating mass quadrupole moment emits GWs described, far from the source, by a metric perturbation $\overleftrightarrow{h}$. The WDB signals in the MLDCs 1, 2 and 1B have been restricted to monochromatic signals with constant intrinsic frequencies f, so in an inertial reference frame such as the solar-system barycentre (SSB), the phase of this signal can be written as $\phi(\tau) = 2\pi f(\tau - \tau_{\mathrm{ref}})$, where τ_{ref} is a reference time. We refer the reader to [7] for a more complete discussion of the formalism; here, we only introduce the notation and key results used in the following derivation. The GW tensor can be expressed as

$$\overleftrightarrow{h}(\tau) = \mathcal{A}^{\mu}\overleftrightarrow{h}_{\mu}(\tau), \tag{1}$$

where we introduce the convention of an implicit sum $\sum_{\mu=1}^{4}$ over repeated indices μ, ν. The four *amplitude parameters* $\{\mathcal{A}^{\mu}\}$ are determined by the overall GW amplitude h_0, the inclination angle ι of the orbital plane, the polarization angle ψ and the initial phase ϕ_0. The explicit relations $\mathcal{A}^{\mu} = \mathcal{A}^{\mu}(h_0, \iota, \psi, \phi_0)$ can be found in equation (4) of [7]. The tensor wave components $\{\overleftrightarrow{h}_{\mu}\}$ depend on the frequency f and the propagation direction $\widehat{k}$ of the GW (determined by the sky position of the source), namely

$$\begin{aligned} \overleftrightarrow{h}_1(\tau;\theta) &= \overleftrightarrow{\varepsilon}_{+}(\widehat{k})\cos\phi(\tau), & \overleftrightarrow{h}_2(\tau;\theta) &= \overleftrightarrow{\varepsilon}_{\times}(\widehat{k})\cos\phi(\tau), \\ \overleftrightarrow{h}_3(\tau;\theta) &= \overleftrightarrow{\varepsilon}_{+}(\widehat{k})\sin\phi(\tau), & \overleftrightarrow{h}_4(\tau;\theta) &= \overleftrightarrow{\varepsilon}_{\times}(\widehat{k})\sin\phi(\tau), \end{aligned} \tag{2}$$

where we denote the set of *Doppler parameters* $\theta \equiv \{f, \widehat{k}\}$. The polarization basis $\overleftrightarrow{\varepsilon}_{+,\times}$ associated with the sky position is defined in terms of the right-handed orthonormal basis $\{\widehat{\xi}, \widehat{\eta}, \widehat{k}\}$ with $\widehat{\xi}$ lying in the ecliptic plane and $\widehat{\eta}$ in the northern hemisphere, namely $\overleftrightarrow{\varepsilon}_{+} = \widehat{\xi}\otimes\widehat{\xi} - \widehat{\eta}\otimes\widehat{\eta}$ and $\overleftrightarrow{\varepsilon}_{\times} = \widehat{\xi}\otimes\widehat{\eta} + \widehat{\eta}\otimes\widehat{\xi}$.

3. The $\mathcal{F}$-statistic method

The $\mathcal{F}$-statistic was originally developed in [5] and extended to the multi-detector case in [10]. A generalization to the full TDI framework for LISA was developed in [11]. In this work, we present an approach that unifies the method for space- and ground-based detector data, allowing for a more direct application of existing LIGO/GEO 600 codes to LISA data analysis.

A 'detector' I (here a TDI observable) provides a linear transformation of the tensor metric perturbation $\overleftrightarrow{h}_\mu(\tau)$ into a scalar 'signal' $h^I_\mu(t)$ as a function of detector time, so the detector output can be written as

$$x^I(t) = n^I(t) + \mathcal{A}^\mu h^I_\mu(t;\theta), \tag{3}$$

where $n^I(t)$ is the instrumental noise in detector I. Following the notation of [10, 12], we write the different data streams $x^I(t)$ as a vector $\boldsymbol{x}(t)$, and we define the standard multi-detector (with uncorrelated noise) scalar product as

$$(\boldsymbol{x}|\boldsymbol{y}) = \sum_\alpha \sum_I \int_{-\infty}^{\infty} \tilde{x}^{I*}_\alpha(f)[S_{\alpha I}(f)]^{-1}\tilde{y}^I_\alpha(f)\,\mathrm{d}f. \tag{4}$$

Here we have broken up the observation time into intervals labelled by α, $\tilde{x}_\alpha$ is the Fourier transform of the data in the αth time interval, x^* denotes complex conjugation and $\{S_{\alpha I}(f)\}$ is the two-sided noise power spectral density appropriate to the αth time interval. We search for a signal $\{\mathcal{A}_\mathrm{s}, \theta_\mathrm{s}\}$ by seeking the parameters $\{\mathcal{A}_\mathrm{c}, \theta_\mathrm{c}\}$ which maximize the log-likelihood ratio

$$L(\boldsymbol{x};\mathcal{A},\theta) = (\boldsymbol{x}|\boldsymbol{h}) - \tfrac{1}{2}(\boldsymbol{h}|\boldsymbol{h}) = \mathcal{A}^\mu(\boldsymbol{x}|\boldsymbol{h}_\mu) - \tfrac{1}{2}\mathcal{A}^\mu(\boldsymbol{h}_\mu|\boldsymbol{h}_\nu)\mathcal{A}^\nu, \tag{5}$$

with automatic summation over repeated amplitude indices μ, ν. Defining

$$x_\mu(\theta) \equiv (\boldsymbol{x}|\boldsymbol{h}_\mu), \qquad \text{and} \qquad \mathcal{M}_{\mu\nu}(\theta) \equiv (\boldsymbol{h}_\mu|\boldsymbol{h}_\nu), \tag{6}$$

we see that L is maximized for given θ by the amplitude estimator $\mathcal{A}^\mu_\mathrm{c} = \mathcal{M}^{\mu\nu}x_\nu$, where $\mathcal{M}^{\mu\nu}$ is the inverse matrix of $\mathcal{M}_{\mu\nu}$. Thus the detection statistic L, maximized over the amplitude parameters $\mathcal{A}$, is $\mathcal{F}$, where

$$2\mathcal{F}(\boldsymbol{x};\theta) \equiv x_\mu\mathcal{M}^{\mu\nu}x_\nu. \tag{7}$$

This defines the (multi-detector) $\mathcal{F}$-statistic. One can show that the expectation in the perfect-match case $\theta = \theta_\mathrm{s}$ is $E[2\mathcal{F}(\theta_\mathrm{s})] = 4 + |\mathcal{A}_\mathrm{s}|^2$, where we used the definition

$$|\mathcal{A}|^2 \equiv \mathcal{A}^\mu\mathcal{M}_{\mu\nu}(\theta_\mathrm{s})\mathcal{A}^\nu, \tag{8}$$

for the norm of a 4-vector $\mathcal{A}^\mu$, using $\mathcal{M}_{\mu\nu}$ as a *metric* on the amplitude-parameter space. Note that $|\mathcal{A}_\mathrm{s}|$ is the (optimal) signal-to-noise ratio (SNR) of the true signal $\{\mathcal{A}_\mathrm{s}, \theta_\mathrm{s}\}$.

4. Modelling the LISA response

The MLDC data were generated by two different programs: synthetic LISA [13] simulates a detector output consisting of Doppler shifts of the LISA lasers due to relative motion of the spacecraft, while LISA Simulator [14] simulates the phase differences between laser light following different paths between the spacecraft. In both cases, the underlying variables are combined with appropriate time shifts to form TDI observables which cancel the (otherwise dominating) laser frequency noise [11, 15, 16]. One choice of such TDI quantities is the set of three observables $\{X, Y, Z\}$, which were used to publish the data of the first and second MLDCs. These observables, which can be thought of as representing the output of three virtual 'detectors' (which we label with the index I), are related to the gravitational wave tensor $\overleftrightarrow{h}$ through the detector 'response', which can be modelled at different levels of accuracy. In the following we discuss two such approximations for the response, the simple 'long-wavelength limit' and the more accurate 'rigid adiabatic approximation'.

4.1. Long-wavelength limit (LWL) response

In the LWL approximation the reduced wavelength $c/(2\pi f)$ is assumed to be large compared to the distance L between the spacecraft, which corresponds to a light-travel time of $T = L/c \sim 17$ s (assuming equal arm lengths), and so this approximation requires $f \ll 10$ mHz. In this approximation the GW contribution to each observable can be modelled as

$$X^{\text{synthLISA}} \approx -4T^2 \overleftrightarrow{d}^X_{\text{LWL}} : \frac{\mathrm{d}^2 \overleftrightarrow{h}}{\mathrm{d}t^2}, \tag{9a}$$

$$X^{\text{LISAsim}} \approx -2T \overleftrightarrow{d}^X_{\text{LWL}} : \frac{\mathrm{d} \overleftrightarrow{h}}{\mathrm{d}t}, \tag{9b}$$

where $:$ denotes the contraction of both tensor indices, and $\overleftrightarrow{d}^X_{\text{LWL}} \equiv (\hat{n}_2 \otimes \hat{n}_2 - \hat{n}_3 \otimes \hat{n}_3)/2$ is the usual LWL response tensor for a GW interferometer with arms $\hat{n}_2$ and $\hat{n}_3$. The analogous expressions for Y and Z are obtained by cyclic permutations of the indices $1 \to 2 \to 3 \to 1$. In the remainder of this section we will give explicit expressions associated with the X variable, with the understanding that the formulae related to Y and Z can be constructed by analogy.

It is convenient to describe the 'response' of a gravitational wave detector in the frequency domain in terms of a response function $R(f)$, relating the detector output to a 'strain' more closely connected to the metric perturbation tensor $\overleftrightarrow{h}$, so that

$$\widetilde{X}(f) = \frac{\widetilde{h}^X(f)}{R(f)} = \frac{\overleftrightarrow{d}^X : \widetilde{\overleftrightarrow{h}}(f)}{R(f)}. \tag{10}$$

In the long-wavelength limit, $\overleftrightarrow{d}^X \approx \overleftrightarrow{d}^X_{\text{LWL}}$ and

$$R^{\text{synthLISA}}(f) \approx R^{\text{synthLISA}}_{\text{LWL}}(f) = \left(\frac{1}{4\pi f T}\right)^2 \tag{11a}$$

$$R^{\text{LISAsim}}(f) \approx R^{\text{LISAsim}}_{\text{LWL}}(f) = \mathrm{i}\frac{1}{4\pi f T}. \tag{11b}$$

This formalism is valid in the regime where the finite lengths of data used to approximate the idealized Fourier transforms are short enough that the geometry and orientation of the detector does not change significantly during this time.

4.2. Rigid adiabatic approximation

A more accurate approximation to the TDI response is the so-called rigid adiabatic approximation [17], which for a wave propagating along the unit vector $\hat{k}$ results in

$$\frac{\overleftrightarrow{d}^X(f,\hat{k})}{R(f)} = \frac{\mathrm{e}^{-\mathrm{i}4\pi f T}}{R_{\text{LWL}}(f)} \operatorname{sinc}(2\pi f T) \left\{ \mathfrak{T}_{\hat{n}_2}(f,\hat{k}) \frac{\hat{n}_2 \otimes \hat{n}_2}{2} - \mathfrak{T}_{-\hat{n}_3}(f,\hat{k}) \frac{\hat{n}_3 \otimes \hat{n}_3}{2} \right\}, \tag{12}$$

where (defining $\xi(\hat{k}) \equiv 1 - \hat{k} \cdot \hat{n}$)

$$\mathfrak{T}_{\hat{n}}(f,\hat{k}) = \frac{\mathrm{e}^{\mathrm{i}2\pi f T \hat{k}\cdot\hat{n}/3}}{2} \{ \mathrm{e}^{\mathrm{i}\pi f T \xi(\hat{k})} \operatorname{sinc}[\pi f T \xi(-\hat{k})] + \mathrm{e}^{-\mathrm{i}\pi f T \xi(-\hat{k})} \operatorname{sinc}[\pi f T \xi(\hat{k})] \} \tag{13}$$

is a transfer function associated with the arm along $\hat{n}$. Note that this is related to the $\mathcal{T}_{\hat{n}}(f,\hat{k})$ defined in [17] by an overall phase, and also that $\mathfrak{T}_{\hat{n}}(f,\hat{k})$ reduces to unity in the LWL

Table 1. Definitions of the long-wavelength (LW), partial rigid adiabatic (pR) and full rigid adiabatic (RA) formalisms, in terms of the response function $R(f)$ (used to calibrate SFTs) and the detector tensor $\overleftrightarrow{d}^I(f,\hat{k})$.

Full name	Label	Response	Detector tensor
Long wavelength	LW	$R_{\mathrm{LWL}}(f)$	$\overleftrightarrow{d}^I_{\mathrm{LWL}}$
Partial rigid adiabatic	pR	$R(f)$	$\overleftrightarrow{d}^I_{\mathrm{LWL}}$
Full rigid adiabatic	RA	$R(f)$	$\overleftrightarrow{d}^I(f,\hat{k})$

$f \ll 1/(\pi T)$. For the separation of (12) into a response function $R(f)$ and a detector tensor $\overleftrightarrow{d}^X(f,\hat{k})$, we choose

$$\overleftrightarrow{d}^X(f,\hat{k}) = \left\{ \mathfrak{T}_{\hat{n}_2}(f,\hat{k}) \frac{\hat{n}_2 \otimes \hat{n}_2}{2} - \mathfrak{T}_{-\hat{n}_3}(f,\hat{k}) \frac{\hat{n}_3 \otimes \hat{n}_3}{2} \right\} \tag{14}$$

$$R(f) = \frac{R_{\mathrm{LWL}}(f)\, \mathrm{e}^{\mathrm{i}4\pi f T}}{\mathrm{sinc}(2\pi f T)}. \tag{15}$$

4.3. Calibrated SFTs

The input to the LAL/LALApps search code consists of Fourier-transformed data stretches of duration T_{SFT}, referred to as short Fourier transforms (SFTs). This is a common data format used within the LIGO Scientific Collaboration for continuous-wave searches (e.g., see [18]). The time baseline T_{SFT} has to be chosen sufficiently short such that the noise floor can be approximated as stationary and the rotation and acceleration of the LISA detector can be neglected, and we chose $T_{\mathrm{SFT}} = 7$ days.

We produce 'calibrated SFTs' by Fourier transforming the raw TDI data and applying a frequency-domain response function to produce a Fourier-transformed strain (including noise) of

$$\tilde{x}^X(f) \equiv R(f)\tilde{X}(f). \tag{16}$$

For our MLDC1 analysis [7] and MLDC2 submission [8] we used the long-wavelength approximation $R_{\mathrm{LWL}}(f)$ for calibrating SFTs, but for subsequent analyses we have produced 'rigid adiabatic' SFTs, which use the full form of $R(f)$ defined in (15).

4.4. Modelling of detector response in different analyses

Our MLDC1B pipeline includes modifications to implement the full form of $\overleftrightarrow{d}^I(f,\hat{k})$. However, a logistically simpler intermediate approximation was also used in the initial followup to our MLDC2 work. In this 'partial rigid adiabatic' formalism, the more precise form of $R(f)$ from (15) is used to construct the SFTs, but the further analysis proceeds with the simpler form of $\overleftrightarrow{d}^I_{\mathrm{LWL}}$. See table 1 for a summary of the three different levels of response approximation considered in this analysis.

5. Signal templates in the rigid adiabatic formalism

5.1. Amplitude modulation coefficients

In the long-wavelength and *partial* rigid adiabatic approximations, the strain $h^I(t)$ at the detector is taken to be the contraction of the metric perturbation $\overleftrightarrow{h}(t(\tau))$ with a detector tensor $\overleftrightarrow{d}^I_{\rm LWL}(t)$ which is independent of the frequency and sky direction of the signal, but which varies slowly with time due to the change of orientation of the detector, in this case as LISA orbits the sun: $h^I(t) = \overleftrightarrow{d}^I_{\rm LWL}(t) : \overleftrightarrow{h}(t(\tau))$. The template 'basis functions' (2) therefore read as

$$\begin{aligned} h^I_1(t;\theta) &= a^I_{\rm LWL}(t,\widehat{k})\cos\phi\left(\tau(t;\theta)\right), & h^I_2(t;\theta) &= b^I_{\rm LWL}(t,\widehat{k})\cos\phi\left(\tau(t;\theta)\right), \\ h^I_3(t;\theta) &= a^I_{\rm LWL}(t,\widehat{k})\sin\phi\left(\tau(t;\theta)\right), & h^I_4(t;\theta) &= b^I_{\rm LWL}(t,\widehat{k})\sin\phi\left(\tau(t;\theta)\right), \end{aligned} \tag{17}$$

where we have defined the usual amplitude-modulation factors $a^I_{\rm LWL}(t,\widehat{k}) \equiv \overleftrightarrow{d}^I_{\rm LWL}(t) : \overleftrightarrow{\varepsilon}_+(\widehat{k})$ and $b^I_{\rm LWL}(t,\widehat{k}) \equiv \overleftrightarrow{d}^I_{\rm LWL}(t) : \overleftrightarrow{\varepsilon}_\times(\widehat{k})$. In the full rigid adiabatic analysis, however, we need to perform this conversion in the frequency domain, because the response tensor $\overleftrightarrow{d}^I(f,\widehat{k})$ depends on the frequency of the incoming waves. In the rigid adiabatic approximation, we model the slow time dependence of the orientation of the detector by using a detector tensor $\overleftrightarrow{d}^I_\alpha(f,\widehat{k})$ appropriate for the time of the αth SFT. This yields the templates

$$\begin{aligned} h^I_{\alpha,1}(f;\theta) &= a^I_\alpha(f,\widehat{k})\,\widetilde{\cos\phi_\alpha}(f;\theta), & h^I_{\alpha,2}(f;\theta) &= b^I_\alpha(f,\widehat{k})\,\widetilde{\cos\phi_\alpha}(f;\theta), \\ h^I_{\alpha,3}(f;\theta) &= a^I_\alpha(f,\widehat{k})\,\widetilde{\sin\phi_\alpha}(f;\theta), & h^I_{\alpha,4}(f;\theta) &= b^I_\alpha(f,\widehat{k})\,\widetilde{\sin\phi_\alpha}(f;\theta). \end{aligned} \tag{18}$$

Now the amplitude modulation (AM) coefficients are constructed from the response tensor according to

$$a^I_\alpha(f,\widehat{k}) = \overleftrightarrow{d}^I_\alpha(f,\widehat{k}) : \overleftrightarrow{\varepsilon}_+(\widehat{k}), \qquad b^I_\alpha(f,\widehat{k}) = \overleftrightarrow{d}^I_\alpha(f,\widehat{k}) : \overleftrightarrow{\varepsilon}_\times(\widehat{k}). \tag{19}$$

Note that the amplitude-modulation functions a and b now depend on frequency as well as sky position, contrary to the LWL case, and they are now complex, due to the complex response tensor (12).

5.2. Amplitude metric

Using the rigid adiabatic forms of the templates (18) gives the following form for the amplitude metric $\mathcal{M}_{\mu\nu}$ defined in (6) :

$$\{\mathcal{M}_{\mu\nu}\} = \begin{pmatrix} A & C & 0 & E \\ C & B & -E & 0 \\ 0 & -E & A & C \\ E & 0 & C & B \end{pmatrix}. \tag{20}$$

This form was first exhibited in [11]. In our notation, the nonzero amplitude metric elements for a signal with intrinsic frequency f_0 are

$$\begin{aligned} A &= \sum_\alpha \sum_I \frac{T_{\rm SFT}}{2S^I_\alpha(f_0)}\left|a^I_\alpha(f_0)\right|^2, & C &= \sum_\alpha \sum_I \frac{T_{\rm SFT}}{2S^I_\alpha(f_0)}\,{\rm Re}\left[a^I_\alpha(f_0)^* b^I_\alpha(f_0)\right], \\ B &= \sum_\alpha \sum_I \frac{T_{\rm SFT}}{2S^I_\alpha(f_0)}\left|b^I_\alpha(f_0)\right|^2, & E &= \sum_\alpha \sum_I \frac{T_{\rm SFT}}{2S^I_\alpha(f_0)}\,{\rm Im}\left[a^I_\alpha(f_0)^* b^I_\alpha(f_0)\right]. \end{aligned} \tag{21}$$

Note that the off-block-diagonal element E vanishes in the long-wavelength (and partial rigid adiabatic) case because the detector tensor $\overleftrightarrow{d}^{I}_{\mathrm{LWL}}$ and thus the AM coefficients are real.

For this project, we have enhanced the $\mathcal{F}$-statistic search in the LAL and LALApps libraries to allow for AM coefficients which depend on frequency and sky direction, and for the more complicated form of the amplitude metric (20) arising from the complex AM coefficients.

6. Results and evaluation

As in the first MLDC, the WDB portion of MLDC1B consisted of seven challenges, labelled 1.1.1a–c and 1.1.2–5. For each challenge, 'training' and 'blind challenge' datasets were provided, generated with different randomly chosen sources. We performed the analysis for our challenge entry on the LISA Simulator data, but discovered subsequently that inconsistent metadata in this dataset led to a 7.5 s time offset and associated systematic error in the initial phase. We are therefore presenting here the results of an analysis with the same pipeline on Synthetic LISA data.

6.1. MLDC1B pipeline

A major limitation of our MLDC1 analysis [7] was the lack of a robust method for distinguishing secondary maxima in Doppler parameter space from additional true signals with lower $\mathcal{F}$-statistic values. Starting with our MLDC2 analysis [8], we implemented a coincidence condition, where consistent signals are required in searches performed with all three TDI variables (X, Y, and Z). This is described in detail in a forthcoming paper [9]. Figure 1 illustrates the pipeline used for this analysis. The first stage consisted of wide-parameter searches over the Doppler space $\theta = \{f, \widehat{k}\}$, using each of the three TDI variables X, Y and Z independently. The template bank used in this first pipeline stage used an isotropic grid on the sky, with angular mesh size $d\alpha = \sqrt{2m_0}/(2\pi f R_{\mathrm{orb}}/c)$, with the orbital radius $R_{\mathrm{orb}} = 1$ AU. The frequency spacing used was $df = \sqrt{12m_0}/(\pi T)$, where $T = 1$ y is the observation time. The mismatch m_0 used in this first stage template bank was $m_0 = 0.25$. The expressions for these step sizes were computed from the orbital metric g_{ij} [12]. The mismatches used in the local-maxima finder and coincidence steps were computed from this metric, using the definition $m \equiv g_{ij}\Delta\theta^i\Delta\theta^j$, where $\Delta\theta^i$ are the Doppler coordinate differences between two candidates. Local maxima are defined as the loudest candidate within a metric sphere of mismatch $m < m_{\mathrm{LM}}$. Coincidence between X, Y and Z is defined as having at least one local maximum from each TDI variable within a coincidence window of mismatch $m < m_{c1}$ in the first coincidence step, and m_{c2} in the second (zoomed) coincidence step. This coincidence scheme was found to be effective in eliminating candidates related to secondary maxima. Zooming of candidates was achieved by running a search covering four neighbouring template points in each dimension with a template-grid resolution increased by a factor `zoomFactor`, i.e. $m_0' = m_0/\mathtt{zoomFactor}$. The final zoom-stage serves only to increase the parameter-estimation accuracy of the final candidates, and is using a coherent TDI combination of the noise-independent variables X and $Y - Z$.

6.2. Evaluation

In order to evaluate our errors in parameter estimation, we compare our estimates to the injected parameters (provided in the 'key', and denoted by a subscript 's'). In Doppler space, $\Delta f = f_{\mathrm{c}} - f_{\mathrm{s}}$ denotes the frequency error, and ϕ_{sky} is the angle between recovered and true sky

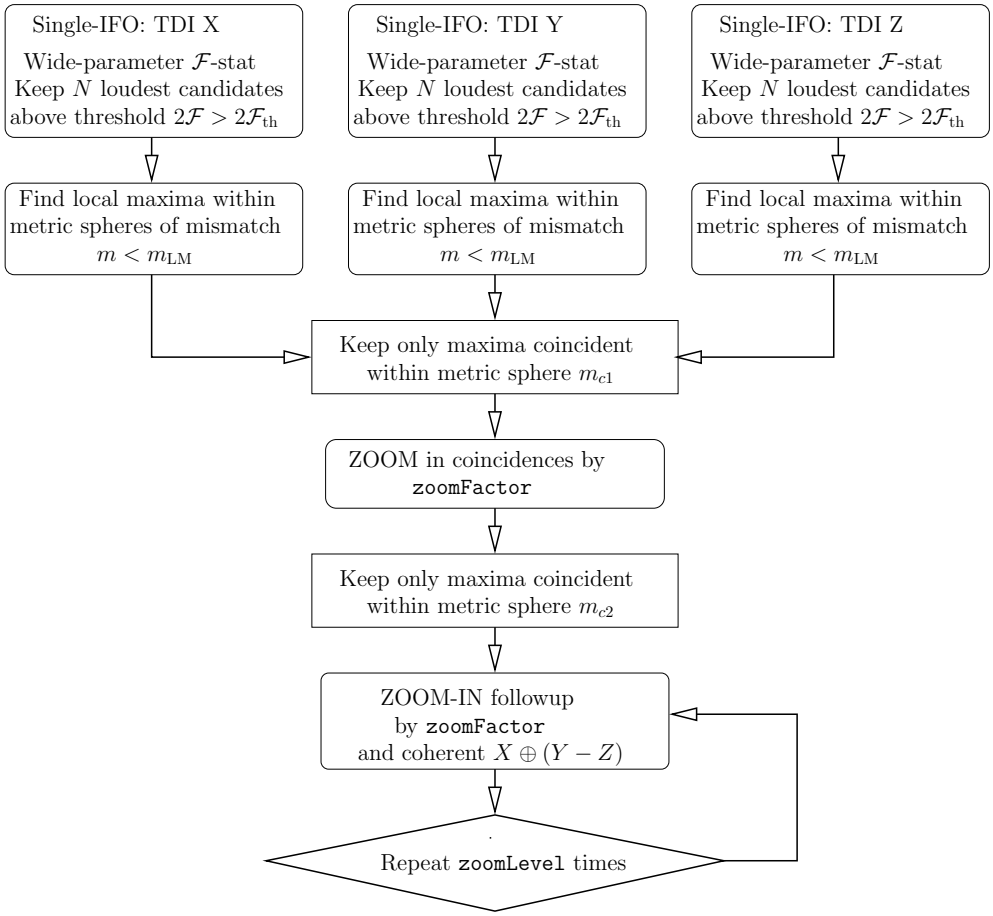

Figure 1. The pipeline used for MLDC1B. The pipeline settings used in this search were: initial template-grid mismatch $m_0 = 0.25$, local-maximum sphere of mismatch $m_{\rm LM} = 4.0$, first coincidence-stage mismatch $m_{c1} = 0.8$, followup coincidence mismatch $m_{c2} = 0.35$. The zoom-stage used zoomLevels = 2 and `zoomFactor` = 10.

position; they can be combined into $\epsilon_\theta^2 \equiv \frac{1}{3} m\, |\mathcal{A}_{\rm s}|^2$, where ϵ_θ measures the number of 'sigmas' error in the Doppler mismatch between the recovered parameters and the injected ones. This is based on the Fisher matrix $\bar{\Gamma}_{ij}$, which is related to the Doppler metric g_{ij} via $\bar{\Gamma}_{ij} = g_{ij}\, |\mathcal{A}_{\rm s}|^2$, e.g., see [12], and where $m = g_{ij}\, \Delta\theta^i\, \Delta\theta^j$, with $\Delta\theta^i \equiv \theta^i_{\rm c} - \theta^i_{\rm s}$. This error measure is chosen in such a way that in Gaussian noise it should satisfy $E\left[\epsilon_\theta^2\right] = 1$. The Doppler metric g_{ij} used here is a simplified 'phase metric'; ideally the full $\mathcal{F}$-statistic metric should be used instead [12], so this should only be considered a rough estimate of the statistical significance of the Doppler errors $\Delta\theta$. For challenges with multiple signals, the Doppler mismatch was also used to distinguish found signals (where the candidate parameters were within $m \leqslant 1$ of the true signal) from false alarms (candidates having no true signal within $m \leqslant 1$).

The amplitude parameter errors are characterized by $\delta_{\mathcal{A}} \equiv \frac{|\mathcal{A}_{\rm c}|^2 - |\mathcal{A}_{\rm s}|^2 - 4}{2|\mathcal{A}_{\rm s}|^2}$, which measures the error in the length of the 4D amplitude parameter vector, and $\phi_{\mathcal{A}}$, which is the angle between the recovered and true amplitude parameter vectors[5]. These two measures of amplitude error

[5] All lengths and angles are calculated using the amplitude metric $\mathcal{M}_{\mu\nu}(\theta_{\rm s})$ corresponding to the *signal* Doppler parameters as in [7], i.e., Doppler parameter errors are ignored when calculating amplitude parameter errors.

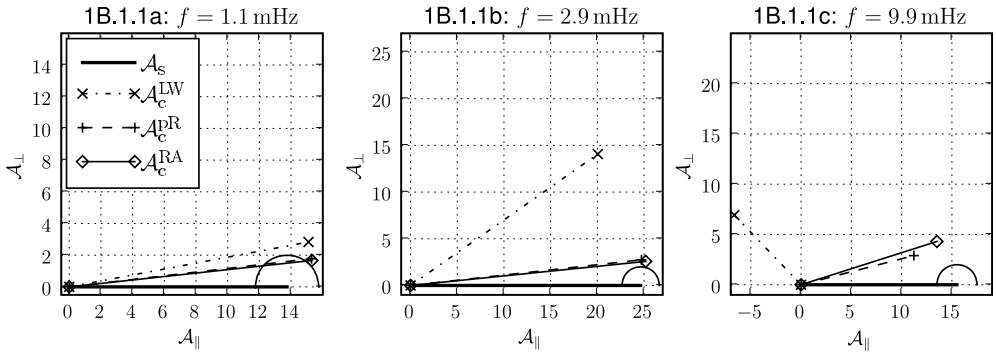

Figure 2. Recovery of amplitude parameters in challenges 1B.1.1a (*left*), 1B.1.1b (*middle*) and 1B.1.1c (*right*), using LW, pR and RA response models (see table 1). Each plot compares the recovered amplitude 4-vector $\mathcal{A}_\mathrm{c}$ to the injected signal 4-vector $\mathcal{A}_\mathrm{s}$, shown in the plane defined by the two vectors. Gaussian fluctuations would lead to a separation of the endpoints of the order $|\Delta\mathcal{A}| \equiv 2\epsilon_\mathcal{A} \sim 2$. (Note that the two-dimensional projection is slightly misleading: since all vectors do not lie in the same plane, the recovered amplitude vectors should be compared only with the key and not with each other.)

are defined so that for statistical errors due to Gaussian noise, the expectation is $E[\delta_\mathcal{A}] = 0$ and the standard deviation of both $\delta_\mathcal{A}$ and $\phi_\mathcal{A}$ is $|\mathcal{A}_\mathrm{s}|^{-1}$. The amplitude parameter errors can be combined into a 'sigma' error, which is $\epsilon_\mathcal{A} = |\Delta\mathcal{A}|/2$, with $\Delta\mathcal{A} \equiv \mathcal{A}_\mathrm{c} - \mathcal{A}_\mathrm{s}$, defined so that $E\left[\epsilon_\mathcal{A}^2\right] = 1$. (Note that the use of $\delta_\mathcal{A}$ for the error in the magnitude and $\epsilon_\mathcal{A}$ for the overall amplitude error is a change in notation from [7].)

6.3. Isolated binaries (challenge 1.1.1)

This challenge consisted of three separate datasets, each containing one WDB signal at an unspecified sky position and within a given frequency band: in 1.1.1a at ∼1 mHz, in 1.1.1b at ∼3 mHz and in 1.1.1c at ∼10 mHz. In each case we were able to recover the injected signal, regardless of the model used for the LISA response, but the recovered amplitude parameters were considerably better, especially at high frequencies, using the full rigid adiabatic response than the long-wavelength limit. The recovery of amplitude and phase parameters is illustrated in figure 2 and table 2.

6.4. Verification binaries (challenge 1.1.2)

In challenge 1.1.2, the sky position and frequency of 25 'verification binaries' was given, while the amplitude parameters of the injected signals were unknown. We therefore performed a targeted $\mathcal{F}$-statistic search at each of the specified sets of Doppler parameters, and found the maximum-likelihood estimators $\mathcal{A}_\mathrm{c}$ for the amplitude parameters. These results are shown in figure 3 and illustrate the performance of the various LISA response models as seen already in challenge 1.1.1.

6.5. Recovery of multiple signals (challenges 1.1.3–5)

Challenges 1B.1.3–5 were supposed to contain multiple binaries, increasingly crowded in Doppler space. Unfortunately, 1B.1.3 was generated with no detectable signals; even the loudest had $|\mathcal{A}_\mathrm{s}|^2 \approx 0.6$. To check performance for resolvable, detectable binaries,

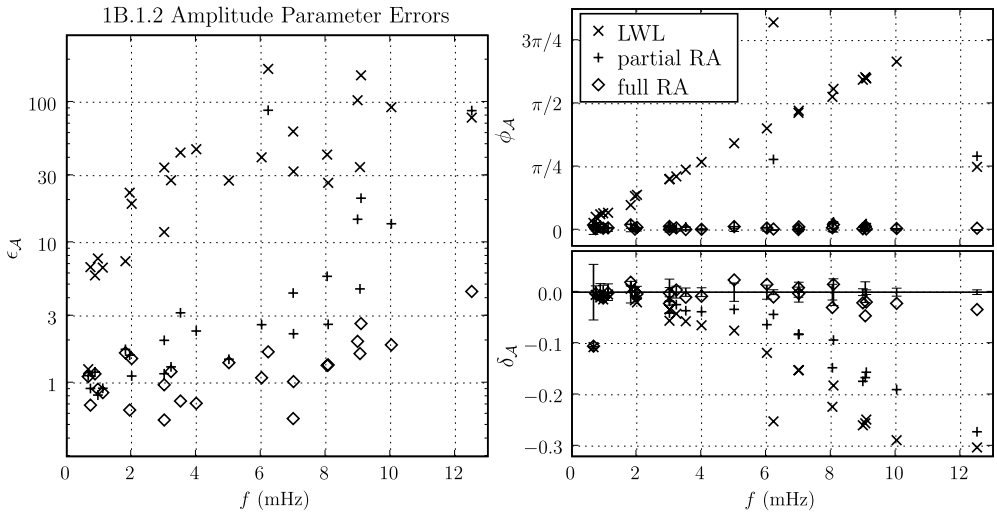

Figure 3. Amplitude parameter errors in challenge 1B.1.2. The quantities plotted are defined in section 6.2. We see a continuation of the trends observed for isolated binaries: the LW response has phase errors ($\phi_{\mathcal{A}}$) which grow with frequency; at higher frequencies, both LW and partial RA exhibit a loss of SNR (negative $\delta_{\mathcal{A}}$), which is much less pronounced with the full RA response. The latter nonetheless displays a slight systematic excess in overall error, i.e. $\epsilon_{\mathcal{A}} \gtrsim 1$.

Table 2. Recovery of Doppler and amplitude parameters in challenges 1B.1.1a–c. The results indicate adequate Doppler-parameter recovery regardless of the response model used. As seen previously in figure 2, recovery of *amplitude* parameters improves using the full RA response, especially at higher frequencies. Note, however, that the errors $\epsilon_{\mathcal{A}}$ in amplitude parameters with the full RA response still appear to be a bit larger than the statistical expectation.

Challenge	Resp.	Δf(nHz)	ϕ_{sky} (mrad)	ϵ_θ	$\delta_{\mathcal{A}}$	$\phi_{\mathcal{A}}$	$\epsilon_{\mathcal{A}}$
1B.1.1a	RA	−0.68	46.06	0.54	0.12	0.11	1.14
($f = 1.1$ mHz;	pRA	−0.68	61.88	0.70	0.12	0.12	1.18
$\lvert\mathcal{A}_{\text{s}}\rvert^{-1} = 0.07$)	LW	−0.68	61.88	0.70	0.11	0.19	1.57
1B.1.1b	RA	0.95	7.71	0.94	0.02	0.10	1.32
($f = 2.9$ mHz;	pRA	0.95	12.30	1.03	0.01	0.11	1.37
$\lvert\mathcal{A}_{\text{s}}\rvert^{-1} = 0.04$)	LW	0.95	12.30	1.03	−0.01	0.61	7.40
1B.1.1c	RA	1.84	7.49	0.72	−0.09	0.31	2.37
($f = 9.9$ mHz;	pRA	1.84	5.12	0.67	−0.23	0.25	2.59
$\lvert\mathcal{A}_{\text{s}}\rvert^{-1} = 0.06$)	LW	1.84	5.12	0.67	−0.32	2.34	11.60

we consider the application of our MLDC1B search pipeline both to MLDC1B data, and to the original MLDC1 datasets. The latter search we refer to MLDC1A, to distinguish it from the MLDC1 search we reported in [7]. The results are summarized in table 3. The corresponding Doppler parameter recovery for 1A.1.3 is shown in figure 4. Note that our signal recovery in challenges 1.1.4 and 1.1.5 is still limited by source confusion, where the primary maximum of a weaker signal can be modified substantially by interference with the secondary peak structure of stronger signals.

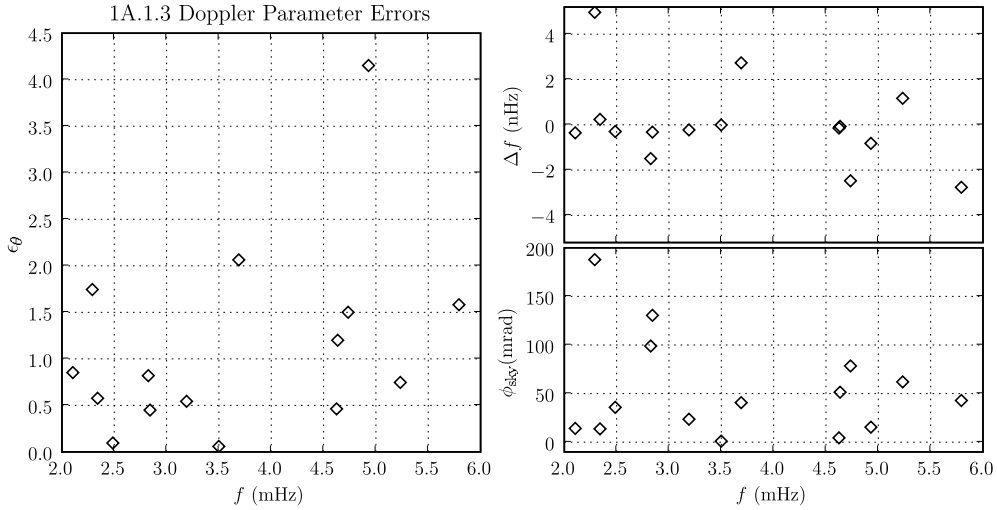

Figure 4. Doppler parameter recovery in 1A.1.3. The quantities plotted are defined in section 6.2. Most signals were recovered with errors compatible with statistical expectations ($\epsilon_\theta \sim 1$); the few outliers may be due to source confusion.

Table 3. Summary of the number of signals found, signals missed and false alarms in challenges 1A.1.3–5 and 1B.1.3–5. We distinguish missed signals with $|\mathcal{A}_s|^2 > 40$, which should in principle be detectable with our current pipeline, and those with $|\mathcal{A}_s|^2 < 40$, which are likely to be too weak to pass our detection threshold. Challenge 1B.1.3 contained no detectable signal, and the only candidate returned by our search was one low-significance false alarm with $2\mathcal{F} \sim 30$.

				Missed					
		Found		$\lvert\mathcal{A}_s\rvert^2 > 40$		$\lvert\mathcal{A}_s\rvert^2 < 40$		False	
Chal	f range (mHz)	1A	1B	1A	1B	1A	1B	1A	1B
1.1.3	2–7	15	0	5	0	0	20	1	1
1.1.4	3.000–3.015	15	13	27	22	3	17	5	1
1.1.5	2.9985–3.0015	3	3	30	29	0	12	0	0

7. Conclusions and outlook

We implemented an $\mathcal{F}$-statistic search for white-dwarf binaries in the mock LISA data challenge based on the LAL/LALApps code developed to search for spinning neutron stars in LIGO/GEO 600 data. MLDC1B has given us the opportunity to improve our MLDC1 search, replacing the long-wavelength response with the rigid adiabatic formalism more appropriate to LISA data analysis. We see that our amplitude parameter recovery is markedly improved by this enhancement, especially at frequencies above 5 mHz.

Another difference from our MLDC1 search is an improved discrimination between secondary maxima and true signals, which is accomplished using a coincidence condition between searches on different TDI variables. This is most relevant in the case of a large number of signals, such as in MLDC2, and a separate paper will address our findings in detail [9].

More work remains to improve our handling of source confusion and deal with multiple interfering sources, as illustrated by the limited number of sources recovered in challenges 1.1.4 and 1.1.5.

Acknowledgments

This work was supported by the Max-Planck-Society, by DFG grant SFB/TR 7 and by the German Aerospace Center (DLR). DK would like to thank the Max Planck Institute for Gravitational Physics (Albert Einstein Institute) for support and hospitality. This paper has been assigned LIGO document no LIGO-P080037-02-Z.

References

[1] MLDC homepage http://astrogravs.nasa.gov/docs/mldc/
[2] Arnaud K A *et al* 2007 *Class. Quantum Grav.* **24** S529–40 (*Preprint* gr-qc/0701139)
[3] Babak S *et al* (Mock LISA Data Challenge Task Force) 2008 *Class. Quantum Grav.* **25** 114037 (*Preprint* arXiv:0711.2667)
[4] Babak S *et al* (Mock LISA Data Challenge Task Force) 2008 *Class. Quantum Grav.* **25** 184026 (*Preprint* arXiv:0806.2110)
[5] Jaranowski P, Krolak A and Schutz B F 1998 *Phys. Rev.* D **58** 063001 (*Preprint* gr-qc/9804014)
[6] LIGO Scientific Collaboration LAL/LALApps: FreeSoftware (GPL) tools for data-analysis http://www.lsc-group.phys.uwm.edu/daswg/
[7] Prix R and Whelan J T 2007 *Class. Quantum Grav.* **24** S565–74 (*Preprint* arXiv:0707.0128)
[8] Prix R and Whelan J T 2007 F-statistic search on the second Mock LISA Data Challenge *LIGO document number* LIGO-G070462-00-Z
[9] Whelan J T, Prix R and Khurana D 2008 Searching for galactic white-dwarf binaries in Mock LISA data using an F-statistic template bank *LIGO document number* LIGO-P080087-00-Z
[10] Cutler C and Schutz B F 2005 *Phys. Rev.* D **72** 063006 (*Preprint* gr-qc/0504011)
[11] Krolak A, Tinto M and Vallisneri M 2004 *Phys. Rev.* D **70** 022003 (*Preprint* gr-qc/0401108)
[12] Prix R 2007 *Phys. Rev.* D **75** 023004 (*Preprint* gr-qc/0606088)
[13] Vallisneri M Synthetic LISA Software http://www.vallis.org/syntheticlisa/
[14] Cornish N J and Rubbo L The LISA Simulator http://www.physics.montana.edu/lisa/
[15] Armstrong J, Estabrook F and Tinto M 1999 *Astrophys. J.* **527** 814–26
[16] Tinto M, Estabrook F B and Armstrong J W 2004 *Phys. Rev.* D **69** 082001 (*Preprint* gr-qc/0310017)
[17] Rubbo L J, Cornish N J and Poujade O 2004 *Phys. Rev.* D **69** 082003 (*Preprint* gr-qc/0311069)
[18] Abbott B *et al* (LIGO Scientific) 2007 *Phys. Rev.* D **76** 082001 (*Preprint* gr-qc/0605028)