Technical Note on CAM entry for MLDC Round 3.4

Farhan Feroz[1], Jonathan R. Gair[2], Philip Graff[1],
and Michael Bridges[1], Michael P. Hobson[1], Anthony Lasenby[1], Edward K. Porter[3]
[1]*Astrophysics Group, Cavendish Laboratory, JJ Thomson Avenue, Cambridge CB3 0HE, UK*
[2]*Institute of Astronomy, Madingley Road, Cambridge CB3 0HA, UK and*
[3]*APC, UMR 7164, Université Paris 7 Denis Diderot, 10,*
*rue Alice Domon et Léonie Duquet, 75205 Paris Cedex 13, France**

We carried out an analysis of the round 3.4 data set, which contained an unknown number of burst sources from cosmic string cusps. Our search was a matched filtering search that employed the MultiNest algorithm to simultaneously recover multiple modes of the posterior. We identified three burst sources in the data set, and recovered possible parameters for each of them, but we found no other promising candidates in the released data.

I. SEARCH TECHNIQUE

The cosmic string bursts are short in duration (typical duration is 1000s), and so we analysed the data set in shorter segments. The search was performed in the Fourier domain, so we divided the data up into sections of length 32768s, with overlaps of 2000s, applied a Welch window and then transformed into the Fourier domain. The search was then repeated with segments offset from the first set of segments by 16384s. We also ran a search on the whole data set as a further check, but this did not identify any additional candidates than we had found in the searches of individual segments.

For the search we constructed the A, E and T optimal detection channels from the X, Y and Z channels in the data release. In the end, only the A and E channels were used in the search since the constructed T channel was quite noisy.

A. Waveform model

The waveform model was as described in the Challenge 3 specification paper [1]. The burst was prescribed in the Fourier domain by the magnitude

$$|h_+| = \mathcal{A} f^{-\frac{4}{3}} \left(1 + \left(\frac{f_{\rm low}}{f}\right)^2\right)^{-4} \exp\left(1 - \frac{f}{f_{\rm max}}\right), \qquad h_\times = 0 \tag{1}$$

and the phase was given by $\exp(2\pi {\rm i}\, f\, t_c)$, where t_c was the burst time. To implement the LISA response, we used the static LISA response as described in [2]. This is a valid approximation for these sources since they are such short duration that LISA is effectively stationary for the burst duration. This was also the approach adopted in [3]. We verified the code by comparing our time-domain waveforms to the noise free training data files.

Using this waveform model for a template h, we were able to define the usual likelihood function for the search of the data stream s, namely $\log \mathcal{L} = -\langle s - h | s - h \rangle /2$, where $< \cdot | \cdot >$ denotes the standard inner product

$$\langle a\, | b \rangle = 4 {\rm Re} \int_0^\infty \frac{\tilde{a}^*(f) \tilde{b}(f)}{S_n(f)}\, {\rm d}f. \tag{2}$$

B. MultiNest algorithm

The search on each segment was carried out using the MultiNest algorithm [4]. This is a search algorithm designed to efficiently compute the evidence associated with different modes of a multi-modal likelihood function. It is well suited to searching parameter spaces that contain multiple secondary modes, which is the case for the cosmic string

*Electronic address: F.Feroz@mrao.cam.ac.uk,jgair@ast.cam.ac.uk,pg338@cam.ac.uk

cusp sources. The algorithm is based on nested sampling, in which a set of live points is evolved through a sequence of iso-likelihood contours of increasing likelihood. The algorithm achieves sampling efficiency by decomposing the live point set into multiple overlapping ellipsoids at each stage. In addition, when distinct modes are identified these are separated and separately evolved, which allows the algorithm to efficiently identify and resolve multi-modal posteriors. We have previously applied this algorithm to a toy problem searching for non-spinning supermassive black hole binaries in simulated data [5], but this was the first round of the MLDC for which we attempted to use it (see also CamAEI entry to Challenge 3.2).

The code is fully parallel, and publicly available. For this search we ran the code on the Darwin supercomputing facility at the University of Cambridge. The search on a 32768s segment typically took $\sim$ 5minutes running in parallel on 8 3GHz Intel Woodcrest processors when the segment did not contain a signal, and approximately double this when a signal was present.

C. Candidate evaluation

The posteriors for these cosmic string cusps are highly multi-modal (see Section II A) and each search returned multiple possible solutions. Rather than submit many alternatives, we took just the few best candidates for submission. The MultiNest algorithm computes the evidence for each identified mode as it analyses the data set. The evidence gives us a natural way to evaluate each candidate. For each submission, we took the modes of highest evidence and assigned them probabilities according to their evidence ratios. In each case, we submitted all modes whose evidence was not less than 10% of the evidence of the dominant mode.

II. RESULTS

A. Analysis of training data

To test the search code, we analysed the noisy training data. We were able to successfully recover all five of the training data sources and no false sources. The recovered posteriors were highly multimodal, which corresponded to real degeneracies in the parameter space. This is illustrated in Figure 1 where we show the true signal for training source 1, and the reconstructed signals for the 9 modes with highest evidence in our recovered posterior. We see clearly that the algorithm is successfully recovering the waveform of the injected source, but the parameter determination is uncertain due to the parameter space degeneracies. For most (although not all) of the training sources the maximum likelihood mode was the closest to the true parameters.

B. Analysis of blind data

Our search identified three source candidates in the data set, at times of $t \sim 6 \times 10^5$s (source candidate 1), $t \sim 1.07 \times 10^6$s (source candidate 2) and $t \sim 1.6 \times 10^6$s (source candidate 3). Each source was followed up with searches on (windowed) segments of data with length 8192s, 16384s and 32768s, centred on the source candidate. The consistency of results between the different segment lengths was checked, but the final reported results were taken from the shortest segment length.

We ran the public evaluation scripts on all of our candidates prior to submission and found that they gave combined A plus E SNRs of 41.4, 33.7 and 43.7 respectively.

1. Source candidates 1 and 3

For both of these candidates, the recovered value of the $f_{\max}$ parameter was consistent with a flat distribution and therefore we concluded it was above the Nyquist frequency of 0.5Hz. The parameter space is less degenerate for higher frequency sources, but for both candidates we identified several possible modes of the solution. We submitted three modes for source candidate 1 and two for source candidate 3, assigning relative probabilities using the evidence ratio as described above. When we constructed time-domain waveforms for the three source 1 candidates one of these (1c) appeared to be a reflection of the other two about the mid-point of the burst. At present we have not yet resolved whether this was a bug in the waveform plotting code or a real degeneracy in the waveform model. The candidate gave equally high SNR when evaluated using the public evaluation scripts which was consistent with our likelihood results, so we included it in the submission.

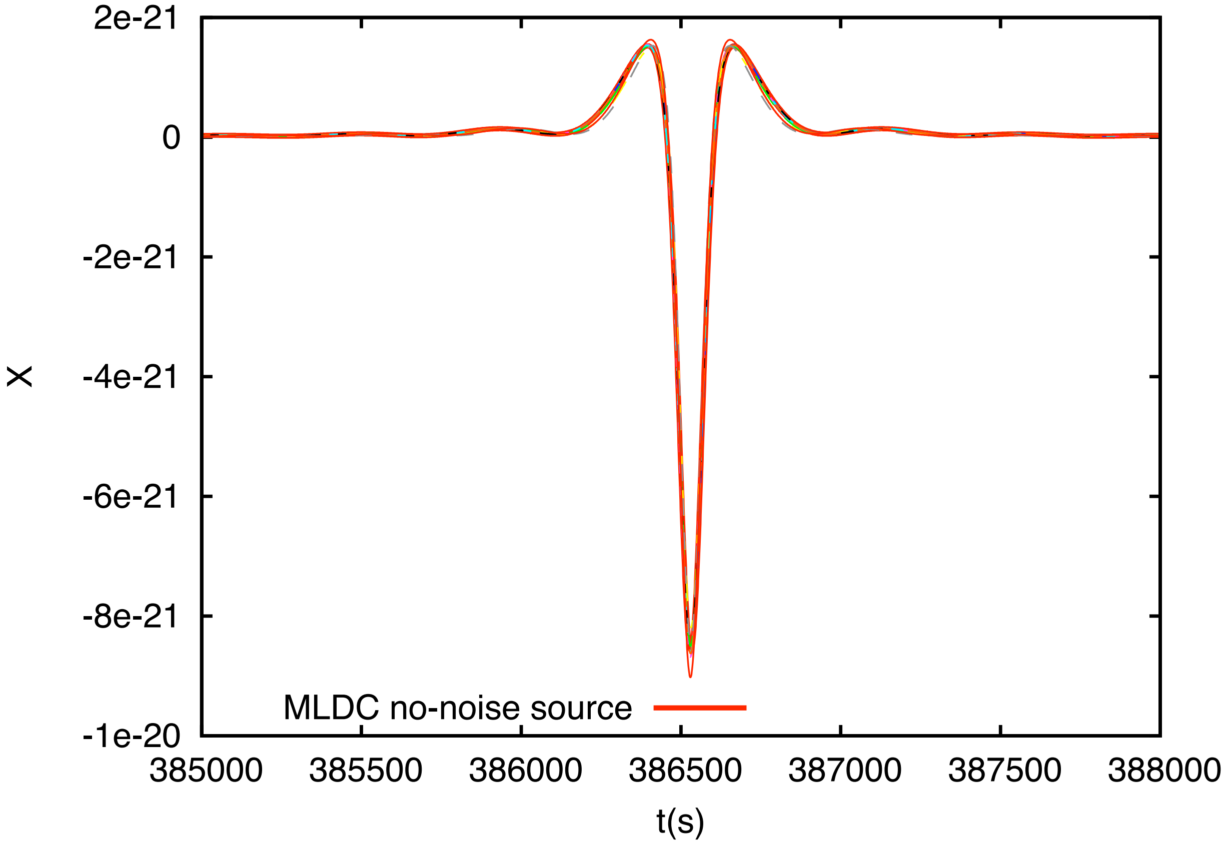

FIG. 1: True signal and the nine modes of highest evidence identified by the algorithm in our search of the training data for the first training source.

2. Source candidate 2

This candidate was consistent with a very low value of $f_{\max}$, very close to the lower end of the prior at 0.001Hz. At such low frequencies, we do not expect LISA to be able to resolve the sky position of a burst source, and as this is correlated with the other parameters we would expect to see a very degenerate posterior. Our results were consistent with this understanding — the posteriors were very flat for all of the parameters except $f_{\max}$. We submitted the maximum likelihood parameter values, but these may be far from the true parameters due to the parameter space degeneracies.

Acknowledgments

We thank Stanislav Babak for useful discussions regarding the implementation of the static LISA response, and for running the public evaluation scripts on our entry prior to the submission. This work was performed using the Darwin Supercomputer of the University of Cambridge High Performance Computing Service (`http://www.hpc.cam.ac.uk/`), provided by Dell Inc. using Strategic Research Infrastructure Funding from the Higher Education Funding Council for England and the authors would like to thank Dr. Stuart Rankin for computational assistance. FF is supported by the Cambridge Commonwealth Trust, Isaac Newton and the Pakistan Higher Education Commission Fellowships.

JG's work is supported by the Royal Society. PG is supported by the Gates Cambridge Trust.

---

[1] Babak, S. et al., *Class. Quantum Grav.***25** 184026 (2008)
[2] Rubbo, L. J., Cornish, N. J. & Poujade, O., Phys. Rev. D **69** 082003 (2004).
[3] Shapiro Key J. & Cornish N. J., Phys. Rev. D **79** 043014 (2009).
[4] Feroz F., Hobson M. P. & Bridges M., preprint arxiv:0809.3437 (2008).
[5] Feroz F., Gair J. R., Hobson M. P. & Porter E. K., preprint arxiv:0904.1544 (2009)