

Article

Not peer-reviewed version

Swarm intelligence methods for Extreme Mass Ratio Inspiral search: First application of Particle Swarm Optimization

Xiao-Bo Zou , [Soumya D. Mohanty](#) * , Honggang Luo , [Yu-Xiao Liu](#)

Posted Date: 22 January 2024

doi: [10.20944/preprints202401.1586.v1](https://doi.org/10.20944/preprints202401.1586.v1)

Keywords: LISA; Gravitational Waves; EMRI; PSO; Likelihood Ratio



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Swarm intelligence methods for Extreme Mass Ratio Inspiral search: First application of Particle Swarm Optimization

Xiao-Bo Zou ^{1,3,4,5} , Soumya D. Mohanty ^{2,3,*} , Hong-Gang Luo ^{1,4,5}  and Yu-Xiao Liu ^{1,4,5} 

¹ School of Physical Science and Technology, Lanzhou University, Lanzhou 730000, China

² Department of Physics and Astronomy, The University of Texas Rio Grande Valley, Brownsville, TX 78520, USA

³ Morningside Center of Mathematics, Academy of Mathematics and System Science, Chinese Academy of Sciences, 55, Zhong Guan Cun Donglu, Beijing 100190, China

⁴ Institute of Theoretical Physics & Research Center of Gravitation, Lanzhou University, Lanzhou 730000, China

⁵ Lanzhou Center for Theoretical Physics and Key Laboratory of Theoretical Physics of Gansu Province, Lanzhou University, Lanzhou 730000, China

* Correspondence: soumya.mohanty@utrgv.edu

Abstract: Swarm intelligence (SI) methods are nature-inspired metaheuristics for global optimization that exploit a coordinated stochastic search strategy by a group of agents. Particle Swarm Optimization (PSO) is an established SI method that has been applied successfully to the optimization of rugged high-dimensional likelihood functions, a problem that presents the main bottleneck across a variety of gravitational wave (GW) data analysis challenges. We present results from a first application of PSO to one of the most difficult of these challenges, namely, the search for Extreme Mass Ratio Inspiral (EMRI) in data from future spaceborne GW detectors such as LISA, Taiji, or Tianqin. We use the standard Generalized Likelihood Ratio Test formalism, with minimal use of restrictive approximations, to search 6 months of simulated LISA data and quantify the search depth, in signal-to-noise ratio (SNR), and breadth, in the ranges of the EMRI parameters, that PSO can handle. Our results demonstrate that a PSO-based EMRI search is successful over a search region that is in the ballpark of the one that current hierarchical schemes can identify. Directions for future improvements, including computational bottlenecks to be overcome, are identified.

Keywords: LISA; Gravitational Waves; EMRI; PSO; Likelihood Ratio

1. Introduction

Spaceborne gravitational wave (GW) detectors, namely LISA [1] and the planned Taiji [2,3] and Tianqin [4], are expected to observe long-lived signals from a variety of sources in the mHz regime. Among these would be the large population of compact object binaries [5–7] in the Milky Way, binaries of massive black holes [8–10], and Extreme Mass Ratio Inspirals (EMRI) [50–54] consisting of a massive black hole orbited by a much smaller one. Extracting individual GW signals from this crowded superposition poses a huge data analysis challenge that has mostly been addressed only for specific types of sources. For example, much work has been done to address the extraction of individual signals from the Galactic Binary population, where both the global fit [11,12] and iterative subtraction approaches [13–16] have been developed to a mature level. On the other hand, progress on the detection and estimation of even single EMRI signals has been much more modest given the extreme challenges that it involves: it is estimated that a deterministic matched filter based search for a single EMRI signal would require $\sim 10^{40}$ distinct template waveform [50]. Given that the correlation of ~ 2 years of data with an equally long template involves $\sim 10^6$ floating point operations, the computational cost of the straightforward approach is insurmountable even with projected advances in computing hardware.

The sheer complexity of the data analysis challenge confronting space-based GW detection necessitates the exploration and development of diverse methods. To provide a common baseline for the comparison of these methods, a series of public data analysis challenges have been provided by the community. Among these, the Mock LISA Data Challenges (MLDCs) [22,23], now called the LISA Data Challenges (LDCs) [26], and the recent Taiji Data Challenge [34] have fostered the development of a number of new data analysis methods. These include a number of different lines of attack on the EMRI problem that may be roughly divided into parametric and nonparametric approaches, with some overlap between the two. Parametric methods use parametrized waveform models, either physical or empirical, while nonparametric methods [68] avoid their use. It is important to note here that the calculation of the exact EMRI waveform is itself an open problem that is being actively addressed by several groups. Hence, all parametric approaches use surrogate models that are assumed to be sufficiently realistic while being computationally cheap to calculate. Among the main surrogate waveform used in the literature, and also the one used in the mock data challenges, is the Analytical Kludge (AK) developed in [55] that is based on the quadrupolar waveform for an eccentric orbit [57,58] with post-Newtonian corrections to the procession of the pericenter and orbital plane as well as the GW radiation reaction driven inspirals. The AK waveform can be expressed as a sum over sinusoidal signals with harmonically separated frequencies that change over time following orbital evolution.

The end goal for all parametric methods is to find the global maximum of a fitness function defined over the high-dimensional space of EMRI signal parameters, with the fitness function in most cases being the log-likelihood ratio (LLR) [20] or the closely related Bayesian posterior probability [21]. The LLR at a given point in signal parameter space for the case of Gaussian noise is simply the projection (modulo terms that take care of signal normalization) of the given data on the signal waveform, called the *template*, corresponding to that point. As mentioned earlier, the global maximization task is computationally infeasible if addressed using a brute force deterministic strategy, which leaves a stochastic global optimization method or a hierarchical search strategy (or their combination) as the only viable way forward. The trade off here is between computational feasibility and lack of certainty in locating the global optimum since stochastic or hierarchical methods are not provably convergent in a time-limited search (but may be asymptotically convergent).

The most popular class of stochastic global optimization methods for the EMRI problem so far has been those based on modifications of the Markov Chain Monte Carlo (MCMC) sampler [59,61–64,66,67]. Hierarchical searches [59,61,63,64,66,67] use several levels of progressively higher fidelity approximations to the fitness function to find, with manageable computational cost, promising search regions in the parameter space that can be handed off to a computationally expensive stochastic or deterministic global optimizer. The fitness approximations are often implemented through waveform restrictions such as limiting the number of harmonics or duration. Most methods proposed so far need to combine elements of both the stochastic and hierarchical approaches to work well. Due to the various possibilities in how the fitness function approximations can be implemented, as well as the large number of tunable parameters they create, the methods proposed so far tend to be quite complex in nature and involve many ad hoc design choices.

In addition to reducing the computational cost of the fitness evaluation, one of the objectives of the approximations is to mitigate the problem of secondary maxima in the exact fitness function [59,64,65] arising from complex degeneracies between EMRI parameters and the overlap of multiple harmonics. These secondary maxima are numerous and widely dispersed in parameter space with values that are comparable to the global maximum. While their presence makes the problem of detection fairly easy, since locating any strong peak is an indication that the data is not pure noise, they have the adverse effect of creating large errors in the estimated signal parameters. In a multiple signal resolution problem, such as the one that all space-based detectors will face, the parameter estimation problem is equally important to that of detection because individual sources must be fitted out from the data in order to reveal weaker sources.

It is commonly assumed, although by no means proven, that a hierarchical search should narrow down the search range to $\approx 100\sigma$ in each signal parameter [64], where σ is the Cramer-Rao lower bound (CRLB) on estimation error [20] obtained from the Fisher information matrix (FIM) at a certain signal-to-noise ratio (SNR). While serving well as a rule of thumb, we note that the search ranges suggested in the Mock LISA Data Challenge (MLDC) [22–24] do not obey this rule (when scaled to the same data duration) and deviate significantly from 100σ in all directions, ranging from $\sim 20\sigma$ to $\sim 500\sigma$ for some of the signal parameters at an SNR = 50 over 0.5 year data.

In this paper, we present a novel approach to the challenge of EMRI data analysis using the Particle Swarm Optimization (PSO) [17–19] algorithm for the global optimization of the LLR. While PSO is a well-known stochastic optimization method that has been highly successful in a wide variety of GW data analysis problems [38–47], as well as related problems in high-dimensional statistical regression [35–37], it has not been applied to the EMRI problem yet. (The application of PSO to the EMRI problem has been proposed in [63] but not actually implemented.) The key features of PSO that sets it apart from MCMC-based searches and makes its application to the EMRI problem attractive is its small set of tunable parameters (only two in most cases) and the significantly smaller number of LLR evaluations that it typically needs for successful convergence to the global optimum. For example, in the search for binary inspiral signals using a network of ground-based detectors, PSO has been demonstrated to need ≈ 10 times fewer LLR evaluations [39–41] than MCMC searches.

Our main objective in this paper is to conduct an *ab initio* investigation that removes as many ad hoc choices from the search method as possible and focus on the end stage of any hierarchical search in which the global optimization is over the approximation-free LLR. In particular, we want to empirically establish the baseline that a hierarchical search should target when narrowing down the search space range. We lift as many restrictions on the fitness function as computationally feasible, such as using template waveform that include the 10 loudest harmonics instead of 3 [59]. Similarly, we do not use approximations that attempt to subdivide the set of signal parameters into analytically maximized *extrinsic* and numerically maximized *intrinsic* since, strictly speaking, such a division does not exist for the EMRI waveform except for the overall distance to the source. Instead, we numerically maximize over all the parameters (except the distance), albeit using different maximization algorithms for the different subsets. Thus, our study provides a good foundation on which to build a PSO-based hierarchical search method in the future that could provide another competitive and promising approach to address the EMRI data analysis challenge.

Within the present computational resource constraints on our analysis, we find that the PSO-based search can successfully handle the global optimization problem for EMRI signals in $\gtrsim 0.5$ years of data over search ranges of $\gtrsim 10\sigma$ for the majority of parameters and up to $\gtrsim 200\sigma$ for one parameter. This is demonstrated over progressively weaker signal-to-noise ratios (SNRs) ranging from 50 to 30, with σ dependent on the SNR. We also report on the parameter estimation accuracy achieved in each case. We note that all our results establish lower bounds on the search ranges since our code is not yet fully optimized to take advantage of Graphics Processing Units (GPUs) and we have not been able to extend the number of PSO iterations or runs to the levels that may be needed at the lowest SNR. Nonetheless, our results show that the plain vanilla PSO is already able to handle search ranges that are in the ballpark of rules-of-thumb values without requiring overly restrictive approximations.

The rest of the paper is organized as follows. Section 2 sets the stage for the EMRI data analysis problem with a review of the TDI combinations, noise, and signal models used in this paper. Section 3 describes the search method along with implementation and computational details. We provide a brief review of the PSO algorithm in Section 4 that is not comprehensive, but suffices for the purposes of this paper. We present our results in Section 5 followed by a discussion in Section 6.

2. Data Description

The design of all space-based GW detectors consists of a constellation of spacecrafts on near-Keplerian orbits, with continuous measurement of the inter-spacecraft distances through the

exchange of laser light signals. The technique of time-delay interferometry (TDI) [56] is used to reduce the dominant noise source, namely laser frequency noise, to a level where the distance fluctuations caused by incident GW signals become detectable. In TDI, the measured time-dependent inter-spacecraft distances along each arm and direction of laser light propagation are combined after being shifted by known time delays relative to each other. The combinations of delays and single arm measurements must take a number of physical effects into account, with an increasing number of effects incorporated in successively better generations of TDI. We discuss below the TDI combinations used in this paper followed by the description of the noise and signal models. Throughout the paper, we follow the coordinate conventions defined in [27].

2.1. TDI Combinations

In this paper, we will use first generation TDI and the combinations called A , E , and T that have mutually independent noise. They are constructed out of the Michelson TDI combinations X , Y , and Z using

$$\begin{aligned} A &= \frac{Z - X}{\sqrt{2}}, \\ E &= \frac{X - 2Y + Z}{\sqrt{6}}, \\ T &= \frac{X + Y + Z}{\sqrt{3}}, \end{aligned} \quad (1)$$

where the combinations X , Y , and Z are obtained from the single arm length measurements, y_{slr} , following

$$\begin{aligned} X &= y_{1-32,32-2} + y_{231,2-2} + y_{123,-2} + y_{3-21} - y_{123,-2-33} - y_{3-21,-33} - y_{1-32,3} - y_{231}, \\ Y &= y_{2-13,13-3} + y_{312,3-3} + y_{231,-3} + y_{1-32} - y_{231,-3-11} - y_{1-32,-22} - y_{2-13,1} - y_{312}, \\ Z &= y_{3-21,21-1} + y_{123,1-1} + y_{312,-1} + y_{2-13} - y_{311,-1-22} - y_{2-13,-11} - y_{3-21,2} - y_{123}. \end{aligned} \quad (2)$$

with s and r labeling the sender and receiver spacecrafcts, respectively, and l labeling the direction of laser light propagation. The expression for the single-arm response to GW, $y_{slr}^{\text{GW}}(t)$, is given by

$$y_{slr}^{\text{GW}}(t) = \frac{\Phi_l(t - \hat{k} \cdot \hat{R}_s - L_l) - \Phi_l(t - \hat{k} \cdot \hat{R}_r)}{2(1 - \hat{k} \cdot \hat{n}_l)}, \quad (3)$$

where the vectors \hat{R}_i , \hat{n}_l , \hat{k} denote the position of the i -th spacecraft, the unit vector along l -th arm and the direction of GW propagation, respectively. The length, L_i , of the i -th arm is approximated as constant in the first generation TDI. The quantity Φ_l , given by

$$\Phi_l = F_l^+ h_+ + F_l^\times h_\times, \quad (4)$$

is the strain response of arm l in the long-wavelength approximation with $F_l^{+,\times}$ being antenna pattern functions and $h_{+,\times}$ being the two GW polarization waveforms of a plane wave in the Transverse-Traceless gauge. The antenna patters are defined by

$$\begin{bmatrix} F_l^+ \\ F_l^\times \end{bmatrix} = \begin{bmatrix} \cos(2\psi) & -\sin(2\psi) \\ \sin(2\psi) & \cos(2\psi) \end{bmatrix} \begin{bmatrix} U_l^+ \\ U_l^\times \end{bmatrix}. \quad (5)$$

with ψ being the polarization angle and the $U_l^{+,\times}$ defined by

$$U_l^+ = (\hat{n}_l \otimes \hat{n}_l) : \epsilon^+, \quad (6)$$

$$U_l^\times = (\hat{n}_l \otimes \hat{n}_l) : \epsilon^+ , \quad (7)$$

where $\epsilon^{+,\times}$ are the polarization tensors in the fiducial wave frame (as defined in [27]) that depend on the sky location of the GW source. Here, the symbol $U : V = \sum_{i,j} U_{ij} V_{ij}$ denotes the contraction operation on arbitrary tensors U and V and, for arbitrary vectors a and b , $(a \otimes b)_{ij} = a_i b_j$. Of the three TDI combinations A , E , and T , the first two carry a significantly larger GW strain response from a given source than the last. Therefore, the combination T is generally left out of GW search considerations. When generating the TDI combinations for a given source, we compute the 24 time-delayed single arm responses, y_{slr} , before linearly combining them, following Equation (2), to get X , Y , Z .

In data analysis, one works with a finite length uniformly sampled time series represented as a row matrix $\bar{d} = (d_0, d_1, \dots, d_{N-1})$, where $d_k = d(t_k)$ and $t_k = k\Delta$ with Δ being the sampling interval. Typically, mock LISA data is generated with $\Delta = 15$ sec, which corresponds to a maximum frequency bandwidth of $1/30 \approx 33$ mHz following the Nyquist sampling theorem. In searches for specific types of signals, computational savings may be obtained by resampling the data to have a higher Δ if the maximum frequency bandwidth of the signals involved is lower. However, directly undersampling a TDI combination by computing the samples of the single arm responses, y_{slr}^{GW} , with a larger Δ causes significant numerical error when introducing the time shifts required in Equation (3) since they are smaller than even the sampling interval of 15 sec. Therefore, in this paper, we keep the original LDC sampling interval when generating the GW strain responses and the TDI combinations, although the implementation of downsampling is clearly a low hanging fruit for speeding up our codes in the future.

Each TDI combination can be described by the data model,

$$\bar{d}^I = \bar{h}^I + \bar{n}^I , \quad (8)$$

where \bar{d}^I denotes the TDI combination I , with $I \in \{A, E, T\}$, \bar{h}^I denotes the GW strain signal, and \bar{n}^I the noise realization. Note that in the case of multiple overlapping GW signals, \bar{h}^I can be split into a sum of resolvable signals and unresolved signals. The latter contribute to the noise realization along with sources of instrumental noise. In this paper, in line with most other studies of the EMRI problem, we consider the simplified problem where \bar{n}^I is purely instrumental noise and \bar{h}^I is the signal from a single GW source.

2.2. Noise Model and Signal to Noise Ratio

Due to the lack of real data, the noise realization in most studies of LISA data analysis is assumed to be from a stationary Gaussian noise process. The LDC manual [27] provides theoretically derived analytical expressions for the power spectral densities (PSDs) of the noise processes in the TDI combinations. For the PSDs of the A and E combinations, the expressions are

$$S_n^A(f) = S_n^E(f) = 8 \sin^2 \omega L [4(1 + \cos \omega L + \cos^2 \omega L) S^{\text{Acc}} + (2 + \cos \omega L) S^{\text{IMS}}] . \quad (9)$$

where f is the Fourier frequency, $w = 2\pi f$ is the corresponding angular frequency, and L is the (assumed constant) arm length. Under the specific noise model called SciRDv1, the acceleration noise S^{Acc} and the Instrumental/Optical Metrology System noise S^{IMS} are given by [32]

$$\begin{aligned} S^{\text{Acc}}(f) &= 9.0 \times 10^{-30} \left[1 + \left(\frac{0.4 \text{mHz}}{f} \right)^2 \right] \left[1 + \left(\frac{f}{8 \text{mHz}} \right)^4 \right] , \\ S^{\text{IMS}}(f) &= 2.25 \times 10^{-22} \left[1 + \left(\frac{2 \text{mHz}}{f} \right)^4 \right] . \end{aligned} \quad (10)$$

Given a PSD $S_n(f)$ and the assumption of Gaussianity, a natural inner product can be defined on the vector space of equal length finite time series as

$$(\bar{a}|\bar{b}) = \frac{1}{T} \sum_{k=0}^{N-1} \frac{\tilde{a}_k \tilde{b}_k^*}{S_n(f_k)}, \quad (11)$$

where \tilde{x} denotes the DFT of a time series $\bar{x} = (x_0, x_1, \dots, x_{N-1})$,

$$\tilde{x} = \mathbf{F} \bar{x}^T, \quad (12)$$

$$F_{lm} = e^{-2\pi i l m / N}, \quad (13)$$

and $f_k = k/T$, $k = 0, 1, \dots, N-1$, with T being the data duration in seconds. We note that, logically, the inner product notation above should also carry the TDI index $I \in \{A, E, T\}$ due to the dependence on the respective PSD. However, we only deal with the combinations A and E in this paper and the two have identical PSDs. Therefore, for simplicity of notation, we have dropped the index on the inner product symbol. In terms of the inner product, one can define the SNR of a signal as,

$$\text{SNR}^2 = (\bar{h}^A|\bar{h}^A) + (\bar{h}^E|\bar{h}^E). \quad (14)$$

It should be noted that, due to the presence of noise, the SNR of an estimated signal will differ by a random amount from that of the true signal in the data. It is also useful to define the *overlap* between two signals, \bar{h}_1^I, \bar{h}_2^I as,

$$\text{ff}_{AE} = \frac{(\bar{h}_1^A|\bar{h}_2^A) + (\bar{h}_1^E|\bar{h}_2^E)}{\sqrt{(\bar{h}_1^A|\bar{h}_1^A) + (\bar{h}_1^E|\bar{h}_1^E)} \sqrt{(\bar{h}_2^A|\bar{h}_2^A) + (\bar{h}_2^E|\bar{h}_2^E)}}. \quad (15)$$

The overlap plays a key role [22] in the context of resolving multiple sources for deciding if an estimated signal matches any of a set of true signals. In the case of a single EMRI signal, the overlap between the estimated and true signals (in simulated data) can serve to diagnose the presence of degeneracies in the parameter space that show up as a large overlap despite large parameter estimation errors. In addition to using both TDI combinations for overlap, one can also define overlap between individual combinations by setting the other combination to zero. These will be designated as f_A and f_E , corresponding to setting $h_{1,2}^E = 0$ and $h_{1,2}^A = 0$, respectively, in Equation (15).

2.3. Signal Model: EMRI Waveform

As mentioned earlier, the current standard EMRI waveform used in studies of data analysis methods as well as the LDCs is the AK waveform. In total, the AK waveform is characterized by 14 parameters: $\{\mu, M, \lambda, S/M^2, e_0, \nu_0, \theta_s, \phi_s, \theta_k, \phi_k, \phi_0, \tilde{\gamma}_0, \alpha_0, D\}$. Here, μ and M represent the masses of the compact objects (COs) and the massive black hole (MBH), respectively. The parameter λ corresponds to the inclination angle between the angular momentum of the COs and the spin direction of MBH. S/M^2 denotes the magnitude of the MBH's spin. The parameters $e_0, \nu_0, \phi_0, \tilde{\gamma}_0$, and α_0 refer to the initial eccentricity, the initial orbital frequency, and the initial angles associated with the orbit rotation, the pericenter precession, and the Lense-Thirring precession, respectively. The angles θ_s and ϕ_s represent the ecliptic colatitude and longitude, respectively, of the sky location of the source in the SSB frame and θ_k, ϕ_k represent the polar and azimuthal angles, respectively, of the MBH spin in the SSB frame. Finally, D represents the distance between the source and the SSB center. The polarization angle ψ depends on $\theta_s, \phi_s, \theta_k$, and ϕ_k and stays constant in a static source frame [59].

To obtain the AK waveform for a given set of signal parameters, the ordinary differential equations (ODEs) given below need to be solved for the mean anomaly ϕ , the orbital frequency ν , the azimuthal

angle of the pericenter precession $\tilde{\gamma}$, the eccentricity of the orbit e , and the azimuthal angle of the Lense-Thirring precession α .

$$\frac{d\phi}{dt} = 2\pi\nu, \quad (16)$$

$$\begin{aligned} \frac{dv}{dt} = & \frac{96}{10\pi}(\mu/M^3)(2\pi M\nu)^{11/3}(1-e^2)^{-9/2}\left\{\left[1+(73/24)e^2+(37/96)e^4\right](1-e^2)\right. \\ & +(2\pi M\nu)^{2/3}\left[(1273/336)-(2561/224)e^2-(3885/128)e^4-(13147/5376)e^6\right] \\ & -(2\pi M\nu)(S/M^2)\cos\lambda(1-e^2)^{-1/2}\left[(73/12)+(1211/24)e^2\right. \\ & \left.\left.+(3143/96)e^4+(65/64)e^6\right]\right\}, \end{aligned} \quad (17)$$

$$\begin{aligned} \frac{d\tilde{\gamma}}{dt} = & 6\pi\nu(2\pi\nu M)^{2/3}(1-e^2)^{-1}\left[1+\frac{1}{4}(2\pi\nu M)^{2/3}(1-e^2)^{-1}(26-15e^2)\right] \\ & -12\pi\nu\cos\lambda(S/M^2)(2\pi M\nu)(1-e^2)^{-3/2}, \end{aligned} \quad (18)$$

$$\begin{aligned} \frac{de}{dt} = & -\frac{e}{15}(\mu/M^2)(1-e^2)^{-7/2}(2\pi M\nu)^{8/3}\left[(304+121e^2)(1-e^2)(1+12(2\pi M\nu)^{2/3})\right. \\ & -\frac{1}{56}(2\pi M\nu)^{2/3}\left((8)(16705)+(12)(9082)e^2-25211e^4\right)] \\ & +e(\mu/M^2)(S/M^2)\cos\lambda(2\pi M\nu)^{11/3}(1-e^2)^{-4}\left[(1364/5)+(5032/15)e^2\right. \\ & \left.\left.+(263/10)e^4\right], \end{aligned} \quad (19)$$

$$\frac{d\alpha}{dt} = 4\pi\nu(S/M^2)(2\pi M\nu)(1-e^2)^{-3/2}. \quad (20)$$

Since the ODEs generally evolve slowly, we follow the solution described in [27] of using a fifth-order Cash-Karp Runge-Kutta ODE solver [28] with a cadence of 15360 seconds, corresponding to a timescale of a few hours, followed by interpolating the solution to a cadence of 15 seconds. Having acquired the solutions to the ODEs, one proceeds to compute the GW polarization waveforms as described schematically below.

Denoting the combined initial phase of the i -th harmonic as $\Phi_0^i = n\phi_0 + 2\tilde{\gamma}_0 + m\alpha_0$, $i = 1, 2, \dots$, (the map from n and m to i being unimportant here) and absorbing amplitude factors such as $1/D$ that are common to all harmonics in an overall parameter A , we get

$$\bar{h}_{+,\times}^i(\Theta) = A\bar{s}_{+,\times}^i(\theta') = A \operatorname{Re}(e^{i\Phi_0^i}\bar{x}_{+,\times}^i(\theta)), \quad (21)$$

where Θ is the set of 14 EMRI parameters, θ' denotes the 13 parameters other than A , and θ denotes the 10 parameters excluding A , ϕ_0 , $\tilde{\gamma}_0$, and α_0 . Thus, $\Theta = \theta' \cup \{A\}$ and $\theta' = \theta \cup \{\phi_0, \tilde{\gamma}_0, \alpha_0\}$. In agreement with the number of parameters they depend on, we call $\bar{h}_{+,\times}^i(\Theta)$, $\bar{s}_{+,\times}^i(\theta')$ and $\bar{x}_{+,\times}^i(\theta)$ the 14, 13, and 10-dimensional polarization waveforms, respectively. (The expression for $\bar{x}_{+,\times}^i(\theta)$ is provided in [55]). As with the calculation of single arm responses, the calculation of the harmonics is parallelized in our code (using OpenMP [30]) before they are summed, which speeds up the calculation of the polarization waveforms substantially.

3. Generalized Likelihood Ratio Test

3.1. LLR

Under the assumption of Gaussian stationary noise, the log-likelihood (LLR) of the data, \bar{d}^I , $I \in \{A, E\}$, described in Section 2 is given by

$$\Lambda(\Theta) = \sum_{I \in \{A, E\}} \left[-(\bar{h}^I(\Theta)|\bar{h}^I(\Theta)) + 2(\bar{d}^I|\bar{h}^I(\Theta)) \right], \quad (22)$$

To distinguish the true but unknown signal present in the data from $\bar{h}^I(\Theta)$, which is used in the evaluation of the LLR at a given point in parameter space, the latter is called a *template*.

In both the Generalized Likelihood Ratio Test (GLRT) and Maximum Likelihood Estimation (MLE), the LLR is maximized over Θ ,

$$L_G = \Lambda(\hat{\Theta}), \quad (23)$$

$$\hat{\Theta} = \underset{\Theta}{\operatorname{argmax}} \Lambda(\Theta), \quad (24)$$

with the value, L_G , of the global maximum serving as the detection statistic and its location, $\hat{\Theta}$, providing the estimated values of the parameters. The parameter A can always be maximized over analytically by solving for $\partial\Lambda(\theta', A)/\partial A = 0$, which gives

$$L_G = \max_{\theta'} \rho(\theta'), \quad (25)$$

$$\rho(\theta') = \max_A \Lambda(\Theta) = \frac{[\sum_{I \in \{A, E\}} (\bar{d}^I | \bar{s}^I(\theta'))]^2}{[\sum_{I \in \{A, E\}} (\bar{s}^I(\theta') | \bar{s}^I(\theta'))]}, \quad (26)$$

with the maximizer being

$$\hat{A} = \underset{A}{\operatorname{argmax}} \Lambda(\Theta) = \frac{[\sum_{I \in \{A, E\}} (\bar{d}^I | \bar{s}^I(\theta'))]}{[\sum_{I \in \{A, E\}} (\bar{s}^I(\theta') | \bar{s}^I(\theta'))]}. \quad (27)$$

In the following, we call $\rho(\theta')$ the 13-dimensional log-likelihood, with further layers of maximization producing functions that will be referred to by the corresponding number of parameters that are not maximized over. The search for the global maximum of $\rho(\theta')$ over the 13 parameters, θ' , in Equation (25) is the main challenge in EMRI data analysis since the number of evaluation points required in a grid-based search is incredibly large.

A dimensionality reduction method was proposed in [59] that attempts to split the maximization of $\rho(\theta')$ into a nested analytical inner maximization over the three initial angles, ϕ_0 , $\tilde{\gamma}_0$, α_0 , and an outer numerical maximization over the remaining parameters θ . Briefly, this method is based on the fact that by combining the antenna patterns $F_l^{+, \times}$ of a given arm l with the 13-dimensional polarization waveform (c.f., Equation (21)), the corresponding 13-dimensional single-arm strain response $\bar{s}_l^i(\theta')$ arising from the i -th harmonic is expressed as

$$\begin{aligned} \bar{s}_l^i(\theta') &= F_l^+ \bar{s}_+^i(\theta') + F_l^\times \bar{s}_\times^i(\theta'), \\ &= \cos(\Phi_0^i) [F_l^+ \operatorname{Re}(\bar{x}_+^i(\theta)) + F_l^\times \operatorname{Re}(\bar{x}_\times^i(\theta))] \\ &\quad - \sin(\Phi_0^i) [F_l^+ \operatorname{Im}(\bar{x}_+^i(\theta)) + F_l^\times \operatorname{Im}(\bar{x}_\times^i(\theta))], \\ &= \cos(\Phi_0^i) \bar{x}_{l,1}^i(\theta) - \sin(\Phi_0^i) \bar{x}_{l,2}^i(\theta), \\ &= \sum_{p=1}^2 a_p^i \bar{x}_{l,p}^i(\theta). \end{aligned} \quad (28)$$

Due to linearity, it follows that the TDI signal generated by the i -th harmonic for combination I is given by

$$\begin{aligned} \bar{s}^{I,i}(\theta') &= \cos(\Phi_0^i) \bar{x}_1^{I,i}(\theta) - \sin(\Phi_0^i) \bar{x}_2^{I,i}(\theta), \\ &= \sum_{p=1}^2 a_p^i \bar{x}_p^{I,i}(\theta), \end{aligned} \quad (29)$$

$$(a_1^i)^2 + (a_2^i)^2 = 1, \quad (30)$$

where the time delays involved in the generation of TDI combinations appear only in the time-dependent $x_p^{I,i}(\theta)$ above. Thus, the initial angles are absorbed in the set of linear coefficients a_j^i , much like the set of 4 linear coefficients that appear in the consideration of a single continuous wave source [60]. In the latter, these linear parameters are mutually independent and can easily be maximized analytically, which considerably simplifies the subsequent maximization of the resultant function, called the F -statistic, over the remaining parameters.

It is apparent from simply counting degrees of freedom that implementing the same trick for EMRI requires the number of harmonics to be restricted to 3 in order to treat the 6 linear coefficients a_j^i , $i = 1, 2, 3$, subject to the 3 equality constraints (c.f., Equation (30)), as mutually independent and to invert the 3 initial angles from them. In [59], this restriction is used to perform the maximization over the a_j^i , $i = 1, 2, 3$ analytically, producing a 10 dimensional fitness function that is then maximized numerically. To fix the three harmonics, a physically motivated assumption about the dominance of their contribution to the total SNR of the signal is used. However, as illustrated in Table 1, the restriction to 3 harmonics has the shortcoming that (a) the dominant harmonics depend on the true parameters of a signal, (b) the SNR contribution of the top three harmonics can fluctuate significantly, and (c) the estimated \hat{A} , hence the estimated D , depends on the choice of the three harmonics and different choices of harmonics may get inconsistent estimates of D . Hence, it is not safe to assume that the same reduced set of 3 harmonics will work the best for every EMRI source.

Table 1. Illustration of variation in the order of contributions of harmonics to the total SNR of an EMRI signal as a function of its parameters. The harmonics are labeled by a single index i , varying from 1 to 25 in LDC, with the mapping of i to n and m given by $n = \text{floor}((i-1)/5) + 1$ and $m = \text{mod}(i-1, 5) + 1$. Each column, besides the leftmost, corresponds to one set of EMRI signal parameters and lists the indices of harmonics in descending order of their contribution to the total signal SNR (comprised of 25 harmonics). Each entry in the table is of the form C/F , where C is the index based on the SNR defined in Equation (14) and F is the index based on a computationally cheaper surrogate defined as the SNR of the strain signal [55], $\Phi_1 - \Phi_2$, in the long-wavelength approximation where Φ_l is defined in Equation (4). (The noise PSD used in the latter is given in [33].) The two ways of computing the SNR order of the harmonics agree well with each other, especially for the moderately eccentric (~ 0.228) systems in column 2, 3, 4. The column labeled by LDC parameters corresponds to the EMRI parameters given as non-blind injection in LDC-1.4 [27] and also listed in Table 2. For each of the other columns, only one parameter was changed and this is noted in the heading of the column. We also provide the total SNR contributed by the top 3 and the top 10 harmonics, as fractions of the total SNR contributed by 25 harmonics, in the rows labeled as SNR fraction.

SNR order (descending)	LDC parameters	$\mu = 10M_\odot$	$\mu = 100M_\odot$	$e_0 = 0.5$	$e_0 = 0.6$
1	7/7	7/7	7/7	7/7	17/12
2	8/8	8/8	8/8	12/12	22/17
3	12/12	12/12	12/12	8/8	18/22
SNR fraction	0.849/ 0.887	0.826/ 0.876	0.906/ 0.904	0.702/ 0.736	0.673/ 0.746
4	13/13	13/13	6/6	13/13	23/7
5	6/6	17/6	9/9	17/17	12/2
6	9/9	6/9	13/13	18/2	13/12
7	17/17	9/17	10/17	22/18	7/18
8	18/11	18/11	17/10	23/22	16/23
9	11/14	11/14	11/14	6/6	21/8
10	14/2	14/18	14/11	11/3	8/3
SNR fraction	0.985/ 0.987	0.981/ 0.985	0.992/ 0.991	0.945/ 0.943	0.945/ 0.943

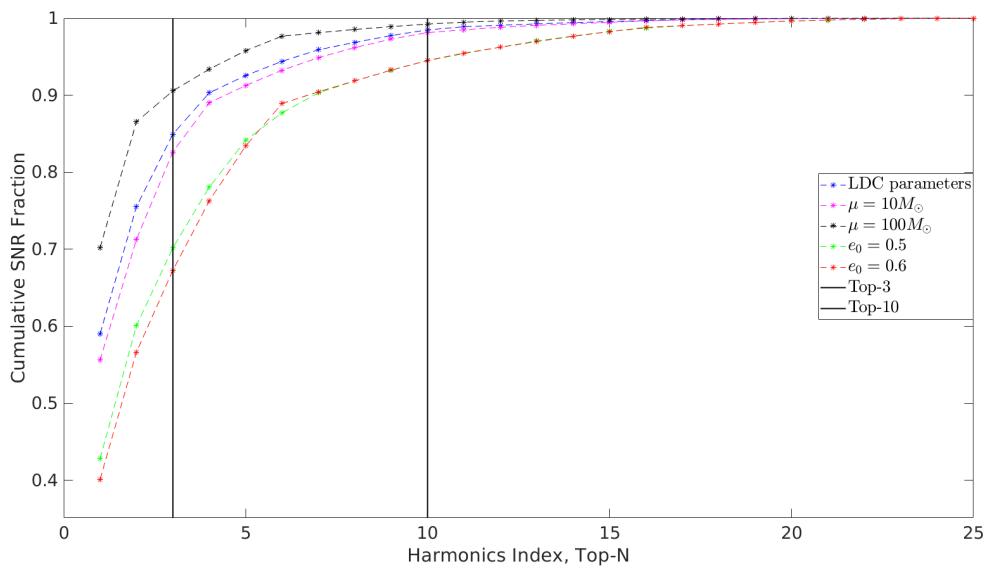


Figure 1. Cumulative SNR fractions over harmonics in descending order for the 5 signal parameters in Table 1.

3.2. The 10-Dimensional Search

In our method, we avoid the above issues associated with restricting the number of harmonics to 3 by keeping a larger number of harmonics in the templates and carrying out the inner maximization over the initial angles numerically. Switching to numerical maximization over the initial angles not only obviates the need for restricting harmonics but also confers some benefits. This is illustrated in Table 1 and Figure 1, where we have compared the effect of retaining the top 10 harmonics with the top 3 on the SNR. First, retaining the top 10 harmonics in the templates considerably stabilizes their fractional SNR contribution compared to 3 harmonics. Secondly, considering the case of high eccentricity, the set of top 10 harmonics, unlike the top 3, remains almost the same across a wider range of signal parameters despite variations in the order of their SNR contributions. For a given location in the 13-dimensional parameter space, the 10 loudest harmonics are obtained in our method using the computationally cheap SNR calculation as outlined in the caption of Table 1. We see that this agrees quite well with the exact SNR calculation (c.f., Equation (14)) as far as identifying the dominant harmonics is concerned.

The nested inner maximization over the initial angles, $\phi_0, \tilde{\gamma}_0, \alpha_0$, is carried out using the Simplex algorithm of Nelder and Mead [29], which has guaranteed convergence to a local maximum. This turns out to be quite effective because the 13-dimensional LLR varies much more slowly over the initial angles compared to, for example, the six parameters related to the ODEs. Consequently, as illustrated in Figure 2, the 13-dimensional LLR usually contains only a few peaks over the 3-dimensional space of the initial angles that are, moreover, equal or comparable in magnitude to the highest peak. Thus, to find the maximum value over the initial angles, it is sufficient to use a local maximizer to converge to any one of these local peaks. To ensure that we get the value of the global maximum within the 3-dimensional space, we use a grid of 32 different starting points for the local maximization (e.g., $[0, 0.5\pi, 1.0\pi, 1.5\pi]$ for ϕ_0, α_0 and $[0, 0.25\pi]$ for $\tilde{\gamma}_0$) and select the best solution.

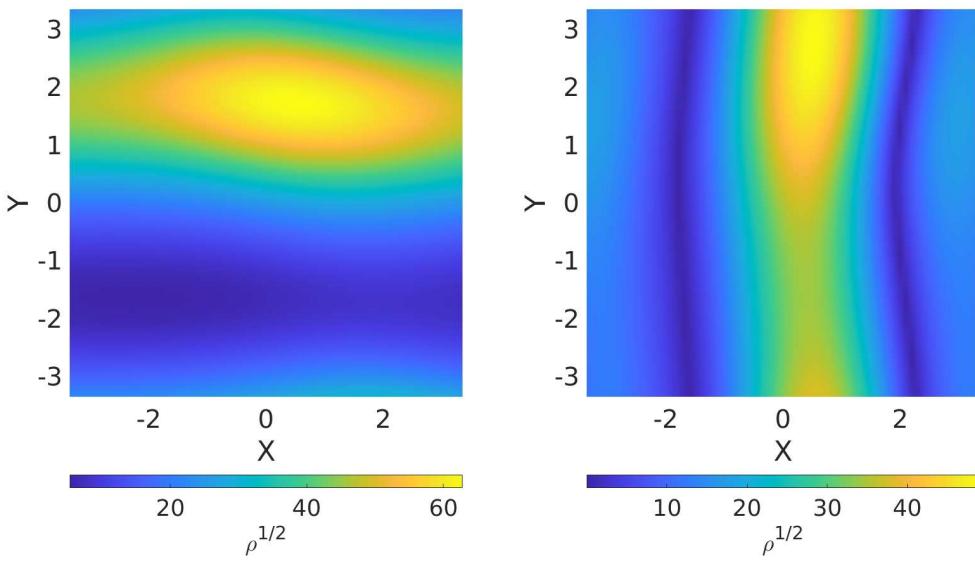


Figure 2. Illustrations of the peak structure of the function $\rho(\theta')$ using two 2-dimensional planar cross-sections of the 3-dimensional space formed by the three initial angles. The plane on the left is $2\phi_0 + 2\tilde{\gamma}_0 - \alpha_0 = 5.1051$ and the plane on the right is $2\phi_0 + 2\tilde{\gamma}_0 = 1.9635$. The X and Y axes lie in these planes and the range along both is $[-\pi, \pi]$. Other planes exhibit similar patterns as above. The figure is produced using the noise realization in LDC-1.4 [27] data and an arbitrary injected signal.

Although numerical maximization over the initial angles incurs additional computation compared to the more restrictive analytical approach outlined in Section 3, its cost is reduced considerably by optimizing the code implementation as follows. The key idea is to apply the TDI linear decomposition in Equation (29) to separate the parameters $\phi_0, \tilde{\gamma}_0, \alpha_0$ from the inner product $(\bar{d}^I | \bar{s}^I(\theta'))$ and $(\bar{s}^I(\theta') | \bar{s}^I(\theta'))$. Thus, we have

$$\begin{aligned} (\bar{d}^I(\theta') | \bar{s}^I(\theta')) &= \sum_{i=1}^N \sum_{p=1}^2 a_p^i (\bar{d}^I | \bar{x}_p^{I,i}(\theta)) , \\ (\bar{s}^I(\theta') | \bar{s}^I(\theta')) &= \sum_{i=1}^N \sum_{j=1}^N \sum_{p=1}^2 \sum_{q=1}^2 a_p^i a_q^j (\bar{x}_p^{I,i}(\theta) | \bar{x}_q^{I,j}(\theta)) . \end{aligned} \quad (31)$$

where N is the number of harmonics involved and the parameters $\phi_0, \tilde{\gamma}_0, \alpha_0$ are absorbed in the linear coefficients a_p^i, a_q^j . For a given θ , precomputing the inner products $(\bar{d}^I | \bar{x}_p^{I,i}(\theta))$ and $(\bar{x}_p^{I,i}(\theta) | \bar{x}_q^{I,j}(\theta))$ in the above expressions allows significant savings in computational cost when using a local maximizer over the three initial angles since they appear only in the set of coefficients $\{a_p^i\}$. Further savings are obtained by neglecting the off-diagonal inner products above between different harmonics, which tend to be very small. In addition, OpenMP is used to parallelize and speed up the evaluation of the inner products. As a result of the above optimizations, the computational cost of each run of the local maximization is on the millisecond scale using a 1.6 GHz 8-core processor, which is negligible compared to the time (on the order of seconds) taken by the other steps in the algorithm.

4. Particle Swarm Optimization

The maximization of the LLR over the 10 parameters left over after (a) analytical maximization over A , and (b) local maximization over the initial angles, must be carried out numerically. As discussed earlier, we use PSO for this step. A brief review of the PSO algorithm is provided below.

PSO is a nature-inspired optimization algorithm inspired by the social behavior of birds and fish, where individuals in a group coordinate their movements to find the best solution to an optimization

problem. PSO has been successfully applied in various fields, including engineering, finance, and machine learning.

The PSO algorithm solves the optimization problem

$$\bar{x}_* = \underset{\bar{x} \in \mathbb{D} \subset \mathbb{R}^M}{\operatorname{argmax}} f(\bar{x}), \quad (32)$$

$$f(\bar{x}_*) \geq f(\bar{x}), \forall \bar{x} \in \mathbb{D}, \quad (33)$$

where the function $f(\bar{x})$ to be maximized is called the *fitness function* and the space \mathbb{D} is called the *search space*. (The fitness function in our case is the 10-dimensional LLR.) This is accomplished in PSO using a set of particles exploring the search space iteratively, where a particle is simply a location in \mathbb{R}^M . Each particle represents a potential solution to the optimization problem, and its movement is influenced by both its own experience and the experiences of its neighboring particles.

The dynamics of a particle in PSO is governed by the following equations (with t representing an iteration). The position of a particle is updated following the rule:

$$\bar{x}_i(t+1) = \bar{x}_i(t) + \bar{v}_i(t+1), \quad (34)$$

where $\bar{x}_i(t)$ is the position of particle i at time t , and $\bar{v}_i(t+1)$ is the displacement (called *velocity* in PSO) update at time $t+1$. The starting positions and velocities of the particles are commonly picked randomly from a uniform distribution over \mathbb{D} . Using x^j to denote the j -th component of a vector $\bar{x} = (x^0, x^1, \dots, x^{M-1})$, the velocity is updated following the rule:

$$v_i^j(t+1) = wv_i^j(t) + c_1r_1(p_i^j(t) - x_i^j(t)) + c_2r_2(g_j(t) - x_i^j(t)), \quad (35)$$

where w is the inertia weight, c_1 and c_2 are acceleration coefficients, r_1 and r_2 are uniformly distributed random variables between 0 and 1, $\bar{p}_i(t)$ is the best position of particle i so far (called its *personal best* or *pbest*), and $\bar{g}(t)$ is the best position in the entire swarm so far (called the *global best* or *gbest*). Here, a position is better than another if its corresponding fitness value is higher. The inertia weight determines the influence of the previous velocity on the current velocity. The corresponding term in the dynamical equation promotes exploration of the search space by a particle. A common choice is to use a linearly decreasing inertia weight over iterations. The acceleration coefficients control the attraction strengths of personal best (c_1) and global best (c_2) on the movement of the particle. A typical setting is $c_1 = c_2 = 2.0$. These terms promote the convergence of the algorithm towards previously identified good solutions. In addition to these parameters, it is necessary to constrain the velocities of particles in order to prevent the entire swarm from quickly leaving the search space. The velocity is commonly constrained by a parameter called the maximum velocity, V_{\max} , with $-V_{\max} \leq v_i^j(t) \leq V_{\max}$ enforced for all iterations.

Before updating the velocity of a particle, the pbest of all particles and the gbest are updated following

$$\text{if } f(\bar{x}_i(t)) > f(\bar{p}_i(t)), \text{ then } \bar{p}_i(t+1) = \bar{x}_i(t+1), \quad (36)$$

and

$$\text{if } f(\bar{x}_i(t)) > f(\bar{g}(t)), \text{ then } \bar{g}(t+1) = \bar{x}_i(t+1). \quad (37)$$

In addition to the parameters listed above, one must set some hyperparameters in the PSO algorithm. These include the number of particles N_p , and the form of the initial and boundary conditions. We use the so-called let-them-fly boundary condition in which no changes are made to the position or velocity of a particle that leaves the search space but its fitness is set to $-\infty$. Gradually, since the pbest and gbest are always located inside the search space, the particle is drawn back in by their attractive forces. The number of particles in the swarm influences the algorithm's exploration and convergence. A small swarm cannot efficiently explore a high-dimensional space while a larger swarm size may prematurely

converge to a local maximum before the particles have had time to fully explore the search space. Empirical evidence suggests that having about 40 particles works well in most cases and this is the number adopted in this paper for obtaining our main results. Finally, one must set up a termination criterion and a common one that is also adopted in this paper is reaching a predefined number, N_{iter} , of iterations.

There are several variations around the basic PSO algorithm described above that seek to achieve different trade-offs between exploration and exploitation (i.e., convergence) phases of the search. Delaying the onset of the exploitation phase leads to a higher computational cost for the search but also improves the chances of finding the global maximum due to a more thorough exploration of the search space. One such variant of PSO, and the one that is used in this paper, is Local Best PSO, which introduces the concept of *local best* (*lbest*) positions, $\bar{p}_{\text{local},i}(t)$, to replace *gbest* in the velocity update equation. The *lbest* position is defined by,

$$f(\bar{p}_{\text{local},i}(t)) = \max_{j \in \mathcal{N}_i} f(\bar{p}_j(t)) , \quad (38)$$

where \mathcal{N}_i is a set of particle indices called the *neighborhood* of particle i . The scheme used for setting up \mathcal{N}_i is called the *swarm topology*. Note that if \mathcal{N}_i is set to be the entire swarm for each i , one recovers the *gbest* PSO described earlier. In this paper, we use the so-called *ring* topology in which the particle indices are arranged on a closed circle and the neighborhood of each particle is the set of adjacent indices on both sides. For example, in our case, we set the neighborhood size to 3, which means that only the particles immediately adjacent on either side constitute the neighborhood of a given particle. By using *lbest* instead of *gbest* in the velocity update equation, the information about the latter propagates only indirectly, through neighbors, and more slowly through the swarm. This extends the exploration phase of the search, generally leading to better performance for more rugged fitness functions.

To boost the performance of PSO, or any stochastic optimization algorithm for that matter, a straightforward approach is to carry out $N_{\text{runs}} \geq 1$ parallel runs of PSO with independent pseudorandom number streams, where N_{runs} depends on the specific computational resources at hand, and pick the final solution to be the one from the run that finds the best fitness value. If the probability of any one run succeeding in a specified region around the global maximum is p , the probability that all independent runs will fail to converge, $(1 - p)^{N_{\text{runs}}}$, drops exponentially fast with increasing N_{runs} . Thus, one can tune PSO to perform moderately well in any one run and simply use the parallel runs strategy to obtain a significantly better overall performance.

For parametric inference problems, such as the one being considered here, there exists [38] an objective strategy for tuning PSO. This is carried out using simulated data realizations, each containing the same true signal. If PSO is successful in locating the global maximum of the fitness function in a given realization, the fitness value it finds must be higher than the value at the location of the true signal in the search space. Hence, one can measure the fraction of realizations in which this condition is satisfied. The higher this fraction, the better tuned PSO is. In most cases, the tuning process involves searching over combinations of N_{iter} and N_{runs} alone, keeping all other PSO parameters fixed. We provide more details about the choice of N_{iter} and N_{runs} in Section 5.

5. Results

Given the computational resources available to us and the current level of parallelization used in our code, we have limited our analysis to 0.5 year data in this paper. For similar reasons, the same duration has been used widely in other studies [64,69,70] for exploring EMRI data analysis methods. We have used the LDC-1.4 [27] signal as our injected signal but adjust the SNR to have three different values, $\text{SNR} \in \{50, 40, 30\}$ by setting the distance D to 1.535300 Gpc, 1.919125 Gpc, 2.558834 Gpc, respectively. As a result, our injected signal has the same source parameters, given in Table 2, as used in LDC-1.4 [27] except distance. The noise realization in the data analyzed here was obtained as the

difference between the data and the signal provided in LDC-1.4 [27]. Thus, both our injected signal and the noise realization are identical to the ones used in LDC-1.4 [27] except for (a) distance scaling, and (b) reduction of data duration from 2 years to 6 months. Figure 3 shows the data and the injected signal with $\text{SNR} = 30$ in the Fourier domain for TDI A and E combinations. We see that, at this SNR, the signal is quite weak in the Fourier domain relative to the noise.

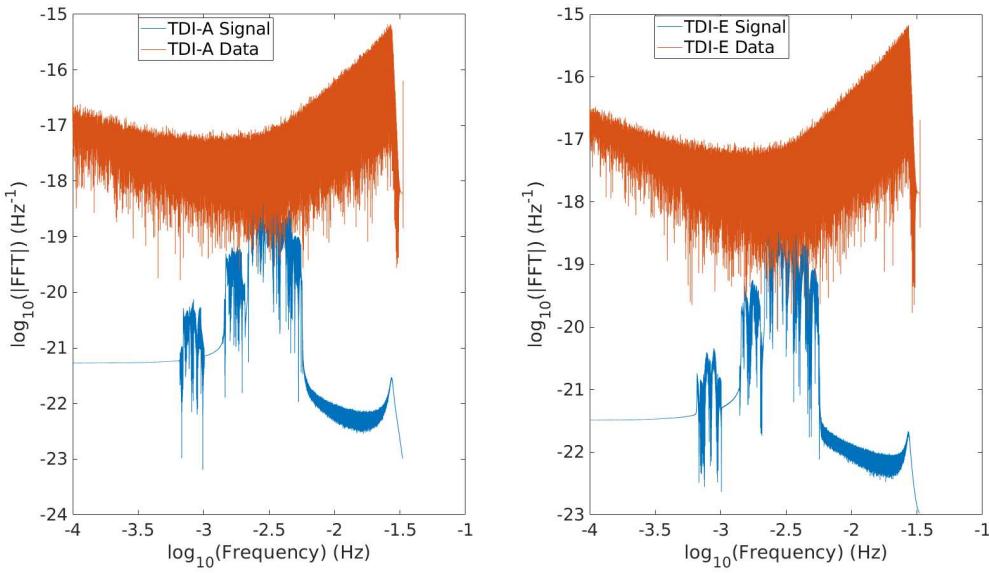


Figure 3. Magnitudes of the FFTs of the injected signal with $\text{SNR} = 30$ in blue and the corresponding data in red where the TDI A combination is illustrated in the left panel and the TDI E combination is displayed in the right panel. See their definitions in Equation (1).

We carry out the 10-dimensional PSO search with the search ranges for the signal parameters set to the values given in Table 2. The search range for a given parameter is expressed as a multiple of σ , where σ is the CRLB on the estimation error for that parameter at the specified signal SNR. Thus, it is important to note that the search range expands with the lowering of SNR since this increases σ . In effect, since a lower SNR and larger search range makes it harder to locate the global maximum, the strength of the challenge posed to the data analysis method is controlled in this study using SNR as a parameter.

Table 2. The injected source parameters and range width used in our search. Currently, the location of the injected signal is set at the center of the given range. We leave a more general search, with injected signals placed non-centrally in the search space, to future work.

Parameters	Values	Search range
$\mu(M_\odot)$	2.9490000e+01	10σ
$M(M_\odot)$	1.1349449e+06	10σ
$\lambda(\text{rad})$	2.1422000e+00	10σ
S/M^2	9.6970000e-01	10σ
e_0	2.2865665e-01	10σ
$\nu_0(\text{Hz})$	7.3804631e-04	200σ
$\theta_s(\text{rad})$	4.989445e-01	π
$\phi_s(\text{rad})$	2.232797e+00	2π
$\theta_k(\text{rad})$	1.522100e+00	π
$\phi_k(\text{rad})$	3.946698e+00	2π

Following the strategy outlined in Section 4, there are two primary tunable hyperparameters for our PSO-based search, the number N_{runs} of independent runs of PSO and the number of iterations,

N_{iter} , per run. However, due to limitations of computational resources, we carried out a very coarse tuning of these parameters based on a small set of runs and set $N_{\text{iter}} = 10000$ and $N_{\text{runs}} = 6$. Moreover, we were unable to run all 6 runs in parallel and had to carry them out serially. In order to save computational resources, we terminated the sequence of runs once, as explained in Section 4, a successful one appeared in which the fitness value exceeded that at the true location. As a result, the actual N_{runs} was 3, 1, 3 for $\text{SNR} \in \{50, 40, 30\}$, respectively. It is important to note that, since we did not pick the best of 6 runs, we may not have obtained the best possible fitness values.

The results obtained from the 10-dimensional search applied to the three data realizations described above are summarized in Table 3. We also report the square root of the fitness values, which provide the estimated SNRs, at the true signal location for both the 13-dimensional fitness function, given by Equation (26), and the 10-dimensional one (c.f., Section 3.2) in which the three initial angles are maximized over numerically using the Nelder-Mead method. Since the global maximum over the three initial angles always shifts from their true values due to noise in the data, the 10-dimensional search always finds a better estimated SNR as observed. Further details about Table 3 are noted below.

1. As noted above, one expects a successful PSO search to find a 10-dimensional fitness value that is larger than the one at the true location. We show the corresponding estimated SNR from the successful run in bold.
2. The parameter estimation errors listed in the table are evaluated based on the best-fit locations of the successful PSO search. For the parameters μ , M , λ , S/M^2 , e_0 , and ν_0 we show their estimation errors, defined as the difference between the true and best-fit values, relative to their respective CRLB errors (evaluated at the true location), while the error is shown relative to its true value for D . For the parameters θ_s , ϕ_s , θ_k , and ϕ_k , we simply show the error itself.
3. Having obtained the 10-dimensional best-fit location of the successful PSO search, which uses templates restricted to the 10 loudest harmonics, we rerun the 3-dimensional local maximization at this location using all 25 harmonics to estimate the three initial angles, ϕ_0 , $\tilde{\gamma}_0$, α_0 . This is done to reveal the influence of the weak harmonics beyond the loudest 10 on the initial angles. The estimated initial angles are then used in the estimation of the distance D using Equation (27).
4. With the 14-dimensional recovered parameters in hand, the estimated TDI A and E signals can be obtained by rerunning Equations (2) and (1). The overlap between the estimated A and E signals and the corresponding true signals, computed separately and in combination as defined in Equation (15), are reported as the quantities ff_A , ff_E , and ff_{AE} .

Table 3. Outputs from the PSO searches for different injected signal SNRs. For each of the 6 ODE related parameters, we show two numbers (with σ denoting the standard deviation): (top) difference between the estimated and true parameter values relative to the 1σ FIM error for that parameter, and (bottom) the corresponding 1σ FIM error. Further details about the table are discussed in Section 5.

	SNR 50	SNR 40	SNR 30
Square root of fitness values			
True 13-dimensional location	48.737520	38.994759	29.251997
True 10-dimensional location	48.794305	39.02358	29.260434
Best location from PSO	47.468231 39.266858 48.888190	39.176120	24.556344 29.525760 23.467734
Parameter estimation errors			
$\mu(M_\odot)$	-1.060 4.872139e-02	1.048 6.090174e-02	0.311 8.120229e-02
$M(M_\odot)$	0.875 3.582834e+03	-1.406 4.478545e+03	-0.376 5.971391e+03
$\lambda(\text{rad})$	-0.905 9.471417e-03	1.349 1.183927e-02	0.368 1.578570e-02
S/M^2	-0.915 3.153740e-03	1.334 3.942175e-03	0.363 5.256233e-03
e_0	1.534 1.842612e-04	0.604 2.303266e-04	-0.057 3.071021e-04
$\nu_0(\text{mHz})$	0.117 3.202842e-06	-2.215 4.003554e-06	-0.150 5.338071e-06
$D(\text{Gpc})$	-0.015	-0.008	-0.006
$\theta_s(\text{rad})$	0.059	0.045	0.065
$\phi_s(\text{rad})$	0.037	0.093	0.137
$\theta_k(\text{rad})$	0.004	-0.186	0.983
$\phi_k(\text{rad})$	0.044	3.426	-1.482
Overlap between the estimated and true signals			
ff_A	-0.991000	0.983481	-0.982155
ff_E	-0.981498	0.968413	-0.954917
ff_{AE}	-0.987463	0.977902	-0.972034

We see from Table 3 that the estimated errors of the parameters $\mu, M, \lambda, S/M^2, e_0, \nu_0$ are within $\sim 1\sigma$ level, parameter D within $\sim 1\sigma$ level and the angles of sky locations θ_s, ϕ_s are within $\sim 10^{-2}$ radians level for all three SNRs. The errors in the parameters θ_k, ϕ_k for $\text{SNR} \in \{40, 30\}$ are larger, which may result from secondary peaks in the fitness as described in Section 1. As a complementary measure of performance, we see that the overlaps, both individual and combined, between the estimated and injected signals are $\gtrsim 97\%$, which indicates that the signal waveform was estimated well. We observe that the errors in the parameters related to the ODEs are smaller for the SNR 30 case, despite the higher difficulty of this search, compared to $\text{SNR} \in \{50, 40\}$. If this is not a random fluctuation specific to the data realization used, it either indicates the effect of degenerate secondary peaks in the fitness function, which may be more prominent for higher SNR signals and could have attracted PSO, or indicates that our current settings for PSO, governed mainly by computational constraints, may need to be changed to higher N_{iter} or N_{runs} to stabilize its performance.

6. Discussion

We have demonstrated the first application of PSO to the EMRI search problem and shown that, in the context of a limited search space and reduced data length, it performs well at reasonable SNR levels. Using three different SNRs ($\{50, 40, 30\}$) and a search space with coordinate ranges limited to 10σ for all parameters except 200σ for ν_0 , with σ being the CRLB standard deviation, PSO was able to successfully find the signal as shown by the small estimation errors for the parameters that affect the intrinsic phase evolution of the GW signal and the overlaps between the estimated and true signals. This sets the stage for further exploration of the EMRI search problem using PSO-based strategies. Our study, which is ongoing, also provides useful guidance for hierarchical strategies in terms of the coordinate ranges to which they should seek to constrain the main search.

The main innovation in our approach is the lifting of several restrictions and approximations adopted in other proposed strategies. In particular, we use local numerical maximization over the three initial angles $\phi_0, \tilde{\gamma}_0, \alpha_0$, which obviates the need to severely restrict the number of harmonics used in the templates and avoids the issues associated with such restrictions. We only make the reasonable restriction of keeping the 10 loudest harmonics based on their contribution to the total SNR. Allowing 10 harmonics in the templates leads to a more stable SNR contribution as well as the order of the dominant harmonics across a wider range of parameter values. Implementing certain optimizations in the implementation of the local maximization makes its computational cost insignificant compared to the rest of the code.

Besides the high-dimensionality of the search space, the main challenge in the EMRI search is the high degeneracy of the fitness function, with many comparable secondary peaks due to the superposition of multiple harmonics in the waveform. As discussed in [65], characterizing this degeneracy is quite challenging and difficult to exploit in a search. One possible approach is to design less degenerate fitness functions by employing [64] surrogate phenomenological EMRI waveform [62]. Another approach would be to use the fact that, like the initial angles, the parameters θ_k and ϕ_k only contribute to the time-independent amplitude of the waveform and use local maximization over them also, following the code optimization discussed in Section 3.2, leaving a potentially easier 8-dimensional search to PSO. Such a lower-dimensional search will be considered in our future work.

A key limitation of our study was the inability to test a much larger number of PSO iterations or runs due to computational resource constraints. While successful in all the cases tested here, PSO exhibited potentially unstable performance by showing a better performance at the lowest SNR that was used. While this could have also been caused by degeneracies in the EMRI fitness function, a better understanding requires much larger PSO runs. Computational resources also limited us from studying the case of signals with weaker SNRs. We have conducted preliminary studies for a much lower SNR of 20 in 0.5 year of data and find that PSO can still detect the signal by obtaining the square root of the fitness value (e.g., ≈ 17) that are well above the ones obtained for pure noise (e.g., ≈ 10). However, due to the lack of a sufficient number of iterations and runs, PSO tends to converge to secondary maxima most of the time and incurs very large parameter estimation errors. In the future, we will include GPU-based hardware acceleration in our code to alleviate this problem and push the PSO-based search further using larger runs.

Author Contributions: Conceptualization, S.D.M. and X.-B.Z.; software X.-B.Z. and S.D.M. (for PSO); validation X.-B.Z., S.D.M., H.-G.L. and Y.-X.L.; writing—original draft preparation, X.-B.Z.; writing—review and editing, S.D.M., H.-G.L and Y.-X.L.; supervision, S.D.M. and H.-G.L; project administration, Y.-X.L.; funding acquisition H.-G.L and Y.-X.L. All authors have read and agreed to the published version of the manuscript.

Funding: The work of X.-B.Z., H.-G.L., and Y.-X.L. was supported by the following grants. (1)National Key Research and Development Program of China (grant No. 2021YFC2203003); (2) The National Key Research and Development Program of China (Grant No. 2022YFA1402704); (3) The National Natural Science Foundation of China (NSFC) (Grant No. 11834005 and No. 12247101).

Data Availability Statement: The LDC data used in this study is publicly available from <https://lisa-ldc.lal.in2p3.fr>. The PSO code used in this study is based on one provided in the GitHub repository RAAPTR available

at <https://github.com/yanwang2012/RAAPTR>. The rest of the code simply implements the mathematical formalism for GLRT and MLE described in this paper, with all necessary details provided to aid reproducibility.

Acknowledgments: The computations were performed on the high-performance computers of the State Key Laboratory of Scientific and Engineering Computing, Chinese Academy of Sciences. We express our gratitude to RunQiu Liu for granting us access to the cluster and providing invaluable suggestions for our research. We also extend our appreciation to the administrator, Yin Qian, for assisting us in utilizing the cluster effectively. Furthermore, we would like to acknowledge the insightful discussions we had with Xian Chen, Wen-biao Han, Peng Xu, Wen-lin Tang, Qun-ying Xie, Xue-hao Zhang, Shao-dong Zhao, Yi-yang Guo, Han-zhi Wang, Shu-zhu Jin, Qian-yun Yun, and Hou-qiang Teng.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

AK	Analytical Kludge
CO	Compact Object
CRLB	Cramer-Rao Lower Bound
DFT	Discrete Fourier Transform
EMRI	Extreme Mass Ratio Inspiral
FIM	Fisher Information Matrix
GW	Gravitational Wave
GLRT	Generalized Likelihood Ratio Test
GPU	Graphics Processing Unit
LDC	LISA Data Challenge
LLR	Log-Likelihood Ratio
LISA	Laser Interferometer Space Antenna
MLE	Maximum Likelihood Estimation
MCMC	Markov Chain Monte Carlo
MLDC	Mock LISA Data Challenge
MBH	Massive Black Hole
MPI	Message Passing Interface
ODE	Ordinary Differential Equation
PSD	Power Spectral Density
PSO	Particle Swarm Optimization
SNR	Signal-to-Noise Ratio
SSB	Solar System Barycenter
SI	Swarm Intelligence
TDI	Time Delay Interferometry

References

1. P. Amaro-Seoane *et al.* [LISA], "Laser Interferometer Space Antenna," [arXiv:1702.00786 [astro-ph.IM]].
2. W. R. Hu and Y. L. Wu, "The Taiji Program in Space for gravitational wave physics and the nature of gravity," *Natl. Sci. Rev.* **4**, no.5, 685-686 (2017) doi:10.1093/nsr/nwx116
3. Z. Luo, Z. Guo, G. Jin, Y. Wu and W. Hu, "A brief analysis to Taiji: Science and technology," *Results Phys.* **16**, 102918 (2020) doi:10.1016/j.rinp.2019.102918
4. J. Luo *et al.* [TianQin], "TianQin: A space-borne gravitational wave detector," *Class. Quant. Grav.* **33**, no.3, 035010 (2016) doi:10.1088/0264-9381/33/3/035010
5. G. Nelemans, L. R. Yungelson and S. F. Portegies Zwart, "The gravitational wave signal from the galactic disk population of binaries containing two compact objects," *Astron. Astrophys.* **375**, 890-898 (2001) doi:10.1051/0004-6361:20010683
6. V. Korol, E. M. Rossi, P. J. Groot, G. Nelemans, S. Toonen and A. G. A. Brown, "Prospects for detection of detached double white dwarf binaries with Gaia, LSST and LISA," *Mon. Not. Roy. Astron. Soc.* **470**, no.2, 1894-1910 (2017) doi:10.1093/mnras/stx1285

7. V. Korol, N. Hallakoun, S. Toonen and N. Karnesis, "Observationally driven Galactic double white dwarf population for LISA," *Mon. Not. Roy. Astron. Soc.* **511**, no.4, 5936-5947 (2022) doi:10.1093/mnras/stac415
8. E. Barausse, J. Bellovary, E. Berti, K. Holley-Bockelmann, B. Farris, B. Sathyaprakash and A. Sesana, "Massive Black Hole Science with eLISA," *J. Phys. Conf. Ser.* **610**, no.1, 012001 (2015) doi:10.1088/1742-6596/610/1/012001
9. A. Klein, E. Barausse, A. Sesana, A. Petiteau, E. Berti, S. Babak, J. Gair, S. Aoudia, I. Hinder and F. Ohme, *et al.* "Science with the space-based interferometer eLISA: Supermassive black hole binaries," *Phys. Rev. D* **93**, no.2, 024003 (2016) doi:10.1103/PhysRevD.93.024003
10. H. T. Wang, Z. Jiang, A. Sesana, E. Barausse, S. J. Huang, Y. F. Wang, W. F. Feng, Y. Wang, Y. M. Hu and J. Mei, *et al.* "Science with the TianQin observatory: Preliminary results on massive black hole binaries," *Phys. Rev. D* **100**, no.4, 043003 (2019) doi:10.1103/PhysRevD.100.043003
11. T. Littenberg, N. Cornish, K. Lackeos and T. Robson, "Global Analysis of the Gravitational Wave Signal from Galactic Binaries," *Phys. Rev. D* **101**, no.12, 123021 (2020) doi:10.1103/PhysRevD.101.123021
12. T. B. Littenberg and N. J. Cornish, "Prototype global analysis of LISA data with multiple source types," *Phys. Rev. D* **107**, no.6, 063004 (2023) doi:10.1103/PhysRevD.107.063004
13. X. Zhang, S. D. Mohanty, X. Zou and Y. Liu, "Resolving Galactic binaries in LISA data using particle swarm optimization and cross-validation," *Phys. Rev. D* **104**, 024023 (2021) doi:10.1103/PhysRevD.104.024023
14. X. H. Zhang, S. D. Zhao, S. D. Mohanty and Y. X. Liu, "Resolving Galactic binaries using a network of space-borne gravitational wave detectors," *Phys. Rev. D* **106**, no.10, 102004 (2022) doi:10.1103/PhysRevD.106.102004
15. S. H. Strub, L. Ferraioli, C. Schmelzbach, S. C. Stähler and D. Giardini, "Bayesian parameter estimation of Galactic binaries in LISA data with Gaussian process regression," *Phys. Rev. D* **106**, no.6, 062003 (2022) doi:10.1103/PhysRevD.106.062003
16. S. H. Strub, L. Ferraioli, C. Schmelzbach, S. C. Stähler and D. Giardini, "Accelerating global parameter estimation of gravitational waves from Galactic binaries using a genetic algorithm and GPUs," *Phys. Rev. D* **108**, no.10, 103018 (2023) doi:10.1103/PhysRevD.108.103018
17. J. Kennedy and R. C. Eberhart, Particle swarm optimization, in *Proceedings International Conference on Neural Networks* (IEEE, Perth, 1995), pp. 1942–1948.
18. D. Bratton and J. Kennedy, "Defining a Standard for Particle Swarm Optimization," *2007 IEEE Swarm Intelligence Symposium*, Honolulu, HI, USA, 2007, pp. 120-127, doi: 10.1109/SIS.2007.368035.
19. S. D. Mohanty, "Swarm Intelligence Methods for Statistical Regression", CRC Press, December 2018, DOI:10.1201/b22461
20. S. M. Kay, *Fundamentals of Statistical Signal Processing*, (Prentice Hall, 1993), Vol. 1&2
21. Jun S. Liu, *Monte Carlo Strategies in Scientific Computing*, (Springer Verlag, 2008)
22. S. Babak, J. G. Baker, M. J. Benacquista, N. J. Cornish, J. Crowder, S. L. Larson, E. Plagnol, E. K. Porter, M. Vallisneri and A. Vecchio, *et al.* "The Mock LISA Data Challenges: From Challenge 1B to Challenge 3," *Class. Quant. Grav.* **25**, 184026 (2008) doi:10.1088/0264-9381/25/18/184026
23. S. Babak *et al.* [Mock LISA Data Challenge Task Force], "The Mock LISA Data Challenges: From Challenge 3 to Challenge 4," *Class. Quant. Grav.* **27**, 084009 (2010) doi:10.1088/0264-9381/27/8/084009
24. K. A. Arnaud, S. Babak, J. G. Baker, M. J. Benacquista, N. J. Cornish, C. Cutler, L. S. Finn, S. L. Larson, T. Littenberg and E. K. Porter, *et al.* "An Overview of the second round of the Mock LISA Data Challenges," *Class. Quant. Grav.* **24**, S551-S564 (2007) doi:10.1088/0264-9381/24/19/S18
25. E. K. Porter, "An Overview of LISA Data Analysis Algorithms," [arXiv:0910.0373 [gr-qc]].
26. Q. Baghi [LDC Working Group], "The LISA Data Challenges," [arXiv:2204.12142 [gr-qc]].
27. LISA Data Challenge, code and maunal, <https://lisa-ldc.lal.in2p3.fr/code> and <https://lisa-ldc.lal.in2p3.fr/static/data/pdf/LDC-manual-002.pdf>
28. William H. Press, Saul A. Teukolsky, William T. Vetterling, Brian P. Flannery, "Numerical Recipes in Fortran 90: The Art of Parallel Scientific Computing, 2nd ed. (Fortran Numerical Recipes 2)", Cambridge University Press, Year: 1996
29. The gsl library, <https://www.gnu.org/software/gsl/doc/html/multimin.html>
30. OpenMP, <https://www.openmp.org/>
31. OpenMPI, <https://www.open-mpi.org/>

32. S. Babak, A. Petiteau and M. Hewitson, "LISA Sensitivity and SNR Calculations," [arXiv:2108.01167 [astro-ph.IM]].
33. T. Robson, N. J. Cornish and C. Liu, "The construction and use of LISA sensitivity curves," *Class. Quant. Grav.* **36**, no.10, 105011 (2019) doi:10.1088/1361-6382/ab1101
34. Z. Ren, T. Zhao, Z. Cao, Z. K. Guo, W. B. Han, H. B. Jin and Y. L. Wu, "Taiji data challenge for exploring gravitational wave universe," *Front. Phys. (Beijing)* **18**, no.6, 64302 (2023) doi:10.1007/s11467-023-1318-y
35. S. D. Mohanty, "Spline Based Search Method For Unmodeled Transient Gravitational Wave Chirps," *Phys. Rev. D* **96**, no.10, 102008 (2017) doi:10.1103/PhysRevD.96.102008
36. Mohanty, S.D., Fahnstock, E. "Adaptive spline fitting with particle swarm optimization," *Comput Stat* **36**, 155–191 (2021). doi:10.1007/s00180-020-01022-x
37. S. D. Mohanty and M. A. T. Chowdhury, "Glitch subtraction from gravitational wave data using adaptive spline fitting," *Class. Quant. Grav.* **40**, no.12, 125001 (2023) doi:10.1088/1361-6382/acd0fe
38. Y. Wang and S. D. Mohanty, "Particle Swarm Optimization and gravitational wave data analysis: Performance on a binary inspiral testbed," *Phys. Rev. D* **81**, 063002 (2010) doi:10.1103/PhysRevD.81.063002
39. T. S. Weerathunga and S. D. Mohanty, "Performance of Particle Swarm Optimization on the fully-coherent all-sky search for gravitational waves from compact binary coalescences," *Phys. Rev. D* **95**, no.12, 124030 (2017) doi:10.1103/PhysRevD.95.124030
40. M. E. Normandin, S. D. Mohanty and T. S. Weerathunga, "Particle Swarm Optimization based search for gravitational waves from compact binary coalescences: Performance improvements," *Phys. Rev. D* **98**, no.4, 044029 (2018) doi:10.1103/PhysRevD.98.044029
41. M. E. Normandin and S. D. Mohanty, "Towards a real-time fully-coherent all-sky search for gravitational waves from compact binary coalescences using particle swarm optimization," *Phys. Rev. D* **101**, no.8, 082001 (2020) doi:10.1103/PhysRevD.101.082001
42. Y. Wang, S. D. Mohanty and F. A. Jenet, "A coherent method for the detection and estimation of continuous gravitational wave signals using a pulsar timing array," *Astrophys. J.* **795**, no.1, 96 (2014) doi:10.1088/0004-637X/795/1/96
43. Y. Wang, S. D. Mohanty and F. A. Jenet, "Coherent network analysis for continuous gravitational wave signals in a pulsar timing array: Pulsar phases as extrinsic parameters," *Astrophys. J.* **815**, no.2, 125 (2015) doi:10.1088/0004-637X/815/2/125
44. X. Zhu, L. Wen, J. Xiong, Y. Xu, Y. Wang, S. D. Mohanty, G. Hobbs and R. N. Manchester, "Detection and localization of continuous gravitational waves with pulsar timing arrays: The role of pulsar terms," *Mon. Not. Roy. Astron. Soc.* **461**, no.2, 1317-1327 (2016) doi:10.1093/mnras/stw1446
45. Y. Wang and S. D. Mohanty, "Pulsar Timing Array Based Search for Supermassive Black Hole Binaries in the Square Kilometer Array Era," *Phys. Rev. Lett.* **118**, no.15, 151104 (2017) [erratum: *Phys. Rev. Lett.* **124**, no.16, 169901 (2020)] doi:10.1103/PhysRevLett.118.151104
46. Y. Wang, S. D. Mohanty and Y. Q. Qian, "Continuous gravitational wave searches with pulsar timing arrays: Maximization versus marginalization over pulsar phase parameters," *J. Phys. Conf. Ser.* **840**, no.1, 012058 (2017) doi:10.1088/1742-6596/840/1/012058
47. Y. Q. Qian, S. D. Mohanty and Y. Wang, "Iterative time-domain method for resolving multiple gravitational wave sources in pulsar timing array data," *Phys. Rev. D* **106**, no.2, 023016 (2022) doi:10.1103/PhysRevD.106.023016
48. P. Amaro-Seoane, J. R. Gair, M. Freitag, M. Coleman Miller, I. Mandel, C. J. Cutler and S. Babak, "Astrophysics, detection and science applications of intermediate- and extreme mass-ratio inspirals," *Class. Quant. Grav.* **24**, R113-R169 (2007) doi:10.1088/0264-9381/24/17/R01
49. P. Amaro-Seoane, "The gravitational capture of compact objects by massive black holes," doi:10.1007/978-981-15-4702-7_17-1 [arXiv:2011.03059 [gr-qc]].
50. J. R. Gair, L. Barack, T. Creighton, C. Cutler, S. L. Larson, E. S. Phinney and M. Vallisneri, "Event rate estimates for LISA extreme mass ratio capture sources," *Class. Quant. Grav.* **21**, S1595-S1606 (2004) doi:10.1088/0264-9381/21/20/003
51. S. Babak, J. Gair, A. Sesana, E. Barausse, C. F. Sopuerta, C. P. L. Berry, E. Berti, P. Amaro-Seoane, A. Petiteau and A. Klein, "Science with the space-based interferometer LISA. V: Extreme mass-ratio inspirals," *Phys. Rev. D* **95**, no.10, 103012 (2017) doi:10.1103/PhysRevD.95.103012

52. J. R. Gair, S. Babak, A. Sesana, P. Amaro-Seoane, E. Barausse, C. P. L. Berry, E. Berti and C. Sopuerta, "Prospects for observing extreme-mass-ratio inspirals with LISA," *J. Phys. Conf. Ser.* **840**, no.1, 012021 (2017) doi:10.1088/1742-6596/840/1/012021
53. H. M. Fan, Y. M. Hu, E. Barausse, A. Sesana, J. d. Zhang, X. Zhang, T. G. Zi and J. Mei, "Science with the TianQin observatory: Preliminary result on extreme-mass-ratio inspirals," *Phys. Rev. D* **102**, no.6, 063016 (2020) doi:10.1103/PhysRevD.102.063016
54. V. Vázquez-Aceves, L. Zwick, E. Bortolas, P. R. Capelo, P. Amaro-Seoane, L. Mayer and X. Chen, "Revised event rates for extreme and extremely large mass-ratio inspirals," *Mon. Not. Roy. Astron. Soc.* **510**, no.2, 2379-2390 (2022) doi:10.1093/mnras/stab3485
55. L. Barack and C. Cutler, "LISA capture sources: Approximate waveform, signal-to-noise ratios, and parameter estimation accuracy," *Phys. Rev. D* **69**, 082005 (2004) doi:10.1103/PhysRevD.69.082005
56. M. Tinto and S. V. Dhurandhar, "TIME DELAY," *Living Rev. Rel.* **8**, 4 (2005) doi:10.12942/lrr-2005-4
57. P. C. Peters and J. Mathews, "Gravitational radiation from point masses in a Keplerian orbit," *Phys. Rev.* **131**, 435-439 (1963) doi:10.1103/PhysRev.131.435
58. P. C. Peters, "Gravitational Radiation and the Motion of Two Point Masses," *Phys. Rev.* **136**, B1224-B1232 (1964) doi:10.1103/PhysRev.136.B1224
59. S. Babak, J. R. Gair and E. K. Porter, "An Algorithm for detection of extreme mass ratio inspirals in LISA data," *Class. Quant. Grav.* **26**, 135004 (2009) doi:10.1088/0264-9381/26/13/135004
60. P. Jaranowski, A. Krolak and B. F. Schutz, "Data analysis of gravitational - wave signals from spinning neutron stars. 1. The Signal and its detection," *Phys. Rev. D* **58**, 063001 (1998) doi:10.1103/PhysRevD.58.063001
61. N. J. Cornish, "Detection Strategies for Extreme Mass Ratio Inspirals," *Class. Quant. Grav.* **28**, 094016 (2011) doi:10.1088/0264-9381/28/9/094016
62. Y. Wang, Y. Shang, S. Babak, Y. Shang and S. Babak, "EMRI data analysis with a phenomenological waveform," *Phys. Rev. D* **86**, 104050 (2012) doi:10.1103/PhysRevD.86.104050
63. D. Bandopadhyay and C. J. Moore, "LISA stellar-mass black hole searches with semicoherent and particle-swarm methods," *Phys. Rev. D* **108**, no.8, 084014 (2023) doi:10.1103/PhysRevD.108.084014
64. C. Q. Ye, H. M. Fan, A. Torres-Orjuela, J. d. Zhang and Y. M. Hu, "Identification of Gravitational-waves from Extreme Mass Ratio Inspirals," [arXiv:2310.03520 [gr-qc]].
65. A. J. K. Chua and C. J. Cutler, "Nonlocal parameter degeneracy in the intrinsic space of gravitational-wave signals from extreme-mass-ratio inspirals," *Phys. Rev. D* **106**, no.12, 124046 (2022) doi:10.1103/PhysRevD.106.124046
66. A. Ali, thesis 2011, "Bayesian Inference on EMRI Signals in LISA Data", <http://hdl.handle.net/2292/7123>
67. A. Ali, N. Christensen, R. Meyer and C. Rover, "Bayesian inference on EMRI signals using low frequency approximations," *Class. Quant. Grav.* **29**, 145014 (2012) doi:10.1088/0264-9381/29/14/145014
68. J. R. Gair, I. Mandel and L. Wen, "Improved time-frequency analysis of extreme-mass-ratio inspiral signals in mock LISA data," *Class. Quant. Grav.* **25**, 184031 (2008) doi:10.1088/0264-9381/25/18/184031
69. Q. Yun, W. B. Han, Y. Y. Guo, H. Wang and M. Du, "Detecting extreme-mass-ratio inspirals for space-borne detectors with deep learning," [arXiv:2309.06694 [gr-qc]].
70. Q. Yun, W. B. Han, Y. Y. Guo, H. Wang and M. Du, "The detection, extraction and parameter estimation of extreme-mass-ratio inspirals with deep learning," [arXiv:2311.18640 [gr-qc]].
71. A. J. K. Chua, *Phys. Rev. D* **106**, no.10, 104051 (2022) doi:10.1103/PhysRevD.106.104051 [arXiv:2205.08702 [gr-qc]].

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.