An astrometric search method for individually resolvable gravitational wave sources with Gaia

Christopher J. Moore,[1,*] Deyan P. Mihaylov,[2] Anthony Lasenby,[3,4] and Gerard Gilmore[2]

[1] *Department of Applied Mathematics and Theoretical Physics, Centre for Mathematical Sciences, University of Cambridge, Wilberforce Road, Cambridge CB3 0WA, UK*
[2] *Intitute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK*
[3] *Astrophysics Group, Cavendish Laboratory, J J Thomson Avenue, Cambridge CB3 0HE, UK*
[4] *Kavli Institute for Cosmology, Madingley Road, Cambridge CB3 0HA, UK*

(Dated: February 12, 2018)

Gravitational waves (GWs) cause the apparent position of distant stars to oscillate with a characteristic pattern on the sky. Astrometric measurements (e.g. those made by Gaia) provide a new way to search for GWs. The main difficulty facing such a search is the large size of the data set; Gaia observes more than one billion stars. In this letter the problem of searching for GWs from individually resolvable supermassive black hole binaries using astrometry is addressed for the first time; it is demonstrated how the data set can be compressed by a factor of more than 10^6, with a loss of sensitivity of less than 1%. This technique was successfully used to recover artificially injected GW signals from mock Gaia data and to assess the GW sensitivity of Gaia. Throughout the letter the complementarity of Gaia and pulsar timing searches for GWs is highlighted.

Introduction – The first detection of gravitational waves (GWs) from merging black holes has recently been achieved by LIGO [1]. LIGO can detect binaries with total masses $\lesssim 160M_\odot$ [2]; however, more massive supermassive black hole binaries radiate at lower frequencies, inaccessible to ground-based instruments. Observing GWs from these massive systems would shed light on the black hole mass function and the coalescence process of the host galaxies and is a target for current and future searches. The planned space-based detector LISA will detect merging binary black holes in the mass range $(10^5$–$10^7)M_\odot$ out to redshifts $z\lesssim 20$ [3]. Pulsar timing arrays (PTAs) use the precise timing of millisecond pulsars to search for GWs with frequencies $10^{-9}\lesssim f/\mathrm{Hz}\lesssim 10^{-7}$. Such GWs may be generated in the early inspiral of a binary in the mass range $(10^7$–$10^{10})M_\odot$. A GW passing over the Earth–pulsar system induces a Doppler shift which affects the pulse arrival times at the Earth. By making a number of time-of-arrival measurements over a timespan T PTAs achieve sensitivity to GWs with frequencies $1/T\lesssim f$ [4]. Current PTAs include NANOGrav [5], EPTA [6], PPTA [7], and the combined IPTA [8].

It is also possible to detect GWs using astrometry [9–11]. The passage of a GW over the Earth–star system changes the apparent position of the star. By making repeated astrometric measurements of many objects and recording their changing positions it is possible to turn an astrometric data set into a nHz GW observatory. The ESA mission Gaia [12] is providing an all-sky astrometric map of $>10^9$ stars. Gaia will operate for 5–10 years, making $\sim$80 observations (in 5 years) per source, delivering proper motion accuracy of $20\,\mu\mathrm{as\,yr}^{-1}$ at magnitude 15, degrading to $300\,\mu\mathrm{as\,yr}^{-1}$ at magnitude 20.7.

The sensitivity bandwidth of Gaia is set by the measurement timings (similar to PTAs); Gaia is sensitive to $1/T\lesssim f$. Gaia and PTAs can search for monochromatic GWs from resolvable circular binaries, stochastic backgrounds of GWs from the superposition of many binaries (or from cosmic string networks [13] or early universe perturbations [14]), or GW bursts with memory [15, 16]. The astrometric analysis of a nearly monochromatic GW is considered here; for example, from a supermassive black hole binary in the early inspiral stage of its evolution.

The astrometric response to GWs – Astrometric measurements of distant objects may be used to detect GWs; the term "star" is used to refer to any such object. It is assumed that the necessary corrections for Gaia's orbital motion have been made, and the term "Earth" is used to refer to an idealised stationary observer.

The possibility of detecting GWs via astrometry was first suggested in [9]; the astrometric deflection of a distant star was derived in [10] (also see [11]) and is summarised here. The Earth and star are assumed to be at rest in flat space. The coordinate components of the photon's four-momentum are not directly observable; instead an observer on Earth measures the *tetrad* components of the photon's four-momentum and from these is able to deduce the star's astrometric position (the unit vector $\vec{n}$), and the frequency of the starlight.

A plane monochromatic GW from the direction of the unit vector[1] $\vec{q}$ has metric perturbation $h_{\mu\nu}=\Re\{H_{\mu\nu}\exp(\mathrm{i}k_\rho x^\rho)\}$, where $H_{\mu\nu}$ are small complex constants satisfying the usual transverse-traceless gauge conditions, and the wavevector, $k^\mu=(\omega,-\omega\vec{q})$, is null. The observed photons follow null geodesics from the star

[1] When working with astrometry it is natural to define the sky position of the GW source, $\vec{q}$; this differs from the usual PTA convention where the GW propagation direction, $\vec{\Omega}=-\vec{q}$, is used.

to the Earth; integrating the geodesic equations gives the change in the *coordinate* components of the photon four-momentum. The GW also changes the observer's tetrad: an orthonormal set of vectors parallel transported along the Earth's worldline. Combining these gives the change in the tetrad components of the four-momentum, and hence the measured frequency and astrometric position.

The frequency perturbation is described by the redshift, $1+z \equiv \Omega_{\rm emit}/\Omega_{\rm obs}$, where

$$z = \frac{n^i n^j}{2(1-\vec{q}\cdot\vec{n})}\left[h_{ij}(\mathrm{E}) - h_{ij}(\mathrm{S})\right] ; \qquad (1)$$

this is the foundation of PTA efforts to detect GWs [17, 18]. The redshift depends (anti)symmetrically on the metric perturbations at the "emission" and "absorption" events at the star (S) and Earth (E) respectively (i.e. z depends only on the difference $[h_{ij}(\mathrm{S}) - h_{ij}(\mathrm{E})]$). This symmetry arises from the endpoints of the integral along the null geodesic from the star to the Earth. The redshift can be integrated to give the timing residual signal searched for by PTAs.

The astrometric perturbation also depends on the Earth and star metric perturbations, although not symmetrically because the perturbation to the spatial vectors in the observer's tetrad depends only on the metric at the Earth. The expression for the astrometric deflection is lengthy, however it simplifies in the limit where the star is many gravitational wavelengths away from Earth [10];

$$\delta n_i = \frac{n_i - q_i}{2(1-\vec{q}\cdot\vec{n})} h_{j\hat{k}}(\mathrm{E}) n^{\hat{j}} n^k - \frac{1}{2} h_{ij}(\mathrm{E}) n^j \, . \qquad (2)$$

In this limit the astrometric deflection depends only on the "Earth term". The "star term" (or "pulsar term") is also sometimes dropped in PTA searches for individually resolvable sources because each pulsar is at a different (generally poorly constrained) distance from Earth, so the "pulsar terms" have different frequencies and phases and may be treated as an effective noise source. Recent searches have tended to include the "pulsar terms" (see, e.g. searches for individual binaries from the three PTAs [19–21], as well as [22–25]) which has the benefit of increasing the observed signal-to-noise at the expense of fitting for the distance to each pulsar.

Gaia's GW sensitivity comes from the large number of stars it observes. Stars are typically separated by many gravitational wavelengths, therefore each "star term" will be different (as well as being suppressed by the distance to the star) whereas the "Earth term" is dominant and common to all stars. It is this common "Earth term" that Gaia aims to detect. Including the "star term" marginally increases the signal-to-noise ratio for the closest few stars but makes a negligible difference for the majority (e.g. a GW with wavelength $\lambda = 10^{16}\,\mathrm{m}$ deflecting a typical star at $d = 10\,\mathrm{kpc}$ gives a "star term" suppressed by $\lambda/d \approx 10^{-5}$). Fig. 1 shows the "Earth term" astrometric deflection pattern for a field of distant stars.

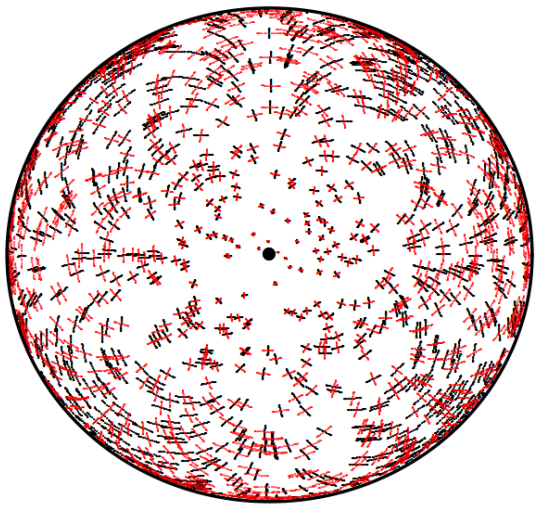

FIG. 1. Orthographic projection of the Northern hemisphere with 10^3 stars. A GW from the North pole (black dot) causes stars to oscillate at the GW frequency. The black (red) lines show movement tracks for a linearly plus (cross) polarised GW. For clarity, the GW has an unphysically large strain amplitude of $A = 0.1$. The four-fold rotational symmetry of the transverse–traceless GWs is clearly imprinted on the sky.

Data analysis – This section describes how to search for a monochromatic GW in an astrometric data set. The likely astrophysical source of such a GW is a circular supermassive black hole binary with total mass in the range $(10^7$–$10^{10})M_\odot$. Such systems spend most of their lifetime in the relatively weak gravitational field where they can be safely assumed to be non-evolving over the observation period[2]. Points on the sky are denoted as $\vec{n}$, and vectors tangent to the sky are denoted as $\mathbf{h}$. For small vectors $|\mathbf{h}| \ll 1$, e.g. the GW astrometric deflection, the sum $\vec{n}\,' = \vec{n} + \mathbf{h}$ gives a nearby point on the sphere.

The GW metric perturbation may be written as

$$h_{ij}\left(\overline{\Psi}\right) = \left(A_+ H^+_{ij}(\vec{q}) e^{\mathrm{i}\phi_+} + A_\times H^\times_{ij}(\vec{q}) e^{\mathrm{i}\phi_\times}\right) e^{2\pi \mathrm{i} f t} \, , \qquad (3)$$

where H^+_{ij}, $H^\times_{ij}$ are the usual GW basis tensors, and $\overline{\Psi}$ is a 7-dimensional parameter vector: two amplitudes A_+, $A_\times$, two phases ϕ_+, $\phi_\times$, the GW frequency f, and two angles describing the direction $\vec{q}$ to the GW source.

The data set, $\mathcal{S}$, consists of N separate astrometric measurements of M stars. The different stars (and measurements) are indexed by I (and J). The observations

[2] For a binary to be considered monochromatic for Gaia analysis, the timescale, τ, over which the GW frequency, $f_{\rm GW}$, evolves must exceed the mission lifetime of $\approx 10\,$years. This timescale can be estimated via $\tau \approx f_{\rm GW}/\dot{f}_{\rm GW}$ using leading order post-Newtonian expressions (see, e.g. [26]). All binaries satisfy $\tau > 10\,$years up to $\approx 3.5\,$years before merger, independent of the component masses. In contrast, these systems cannot always be considered monochromatic for PTA analysis because the "pulsar terms" provide snapshots of the $f_{\rm GW}$ at widely separated times allowing the frequency evolution to be measured (see, e.g. [27]).

are made at times t_J (for simplicity the t_J are assumed to be the same for all stars);

$$\mathcal{S} = \{\vec{s}_{I,J} | I = 1, 2, \ldots, M;\ J = 1, 2, \ldots, N\} \ . \quad (4)$$

Each individual measurement is a combination of the background star position, $\vec{n}_I(t_J)$, instrumental noise, $\mathbf{r}_{I,J}$, and (possibly) a GW;

$$\vec{s}_{I,J} = \vec{n}_I(t_J) + \mathbf{r}_{I,J} + \mathbf{h}\left(\overline{\Psi}; \vec{n}_I(t_J), t_J\right) \ . \quad (5)$$

The background positions vary due to the star's proper motion. For each star the function $\vec{n}_I(t_J)$ is modelled as a quadratic, $\underline{\vec{n}}_I(t_J)$ and subtracted from the data;

$$\mathbf{s}_{I,J} = \vec{s}_{I,J} - \underline{\vec{n}}_I(t_J) \, . \quad (6)$$

Thereby the background positions, proper motions, and accelerations are fit out of the data. This is the astrometric equivalent of the pulsar *timing model* and sets the low frequency sensitivity [4]. The position model can be marginalised over (see [28] in the PTA context), however, here the maximum likelihood parameters are used.

For simplicity the noise in each measurement is assumed to be identical and independent ($\sigma \equiv \sigma_{I,J}$),

$$\mathrm{E}\left[\mathbf{r}_{I,J} \cdot \mathbf{r}_{I',J'}\right] = \sigma^2 \delta_{II'} \delta_{JJ'} \, . \quad (7)$$

The likelihood of $\mathcal{S}$ given the parameters $\overline{\Psi}$, assuming the star's motion has been correctly modelled and under the noise assumptions described, may be written as

$$P\left(\mathcal{S}|\overline{\Psi}\right) \propto \exp\left(\sum_{I=1}^{M}\sum_{J=1}^{N} \frac{-\left|\mathbf{s}_{I,J} - \mathbf{h}\left(\overline{\Psi}; \tilde{\mathbf{n}}_I(t_J), t_J\right)\right|^2}{2\sigma^2}\right), \quad (8)$$

where $|\cdot|$ denotes the norm of a vector on the sphere. The posterior probability follows from Bayes' theorem,

$$P\left(\overline{\Psi}|\mathcal{S}\right) = \frac{\Pi\left(\overline{\Psi}\right) P\left(\mathcal{S}|\overline{\Psi}\right)}{\mathcal{Z}_{\mathrm{s}}} \, , \quad (9)$$

with prior $\Pi\left(\overline{\Psi}\right)$. Throughout this letter uniform periodic priors for the phase angles ϕ_+, $\phi_\times$, uniform in log priors for the amplitudes A_+, $A_\times$, uniform in log prior for the frequency in the range $f \sim \mathcal{U}\left[1/T, N/2T\right]$, and a uniform prior on the sphere for $\vec{q}$ are used.

The Bayesian signal evidence normalises the distribution in Eq. 9 and is given by

$$\mathcal{Z}_{\mathrm{signal}} = \int \mathrm{d}\overline{\Psi}\, \Pi\left(\overline{\Psi}\right) P\left(\mathcal{S}|\overline{\Psi}\right) \ . \quad (10)$$

The noise evidence $\mathcal{Z}_{\mathrm{noise}}$, is simply given by the likelihood in Eq. 8 evaluated with no GW signal. The Bayes' factor $\mathcal{B} \equiv \mathcal{Z}_{\mathrm{signal}}/\mathcal{Z}_{\mathrm{noise}}$ is used as a detection statistic; it is assumed that any signal with $\mathcal{B} > \mathcal{B}_{\mathrm{threshold}} = 10^{1.5}$ can be confidently detected. This is generally a conservative choice, and corresponds to Jeffrey's [29] criterion for detection with "very strong" evidence (the threshold choice is discussed in the supplement to this letter); the precise detection threshold is problem specific and will depend on the details of the final Gaia data release.

The MultiNest [30] implementation of nested sampling [31] was used to sample the posterior (Eq. 9) and evaluate the evidence (Eq. 10).

A mock Gaia data set was constructed with $M = 10^5$ stars (approximately a factor of 10^4 less than the full Gaia catalog for computational necessity) each measured $N = 75$ times evenly spaced over a $T = 5\,\mathrm{year}$ mission (the effect of non-uniform sampling is explored below). The simulated noise in each measurement was $\sigma = 100\,\mu\mathrm{as}/\sqrt{10^4}$, reflecting an estimate of the errors in each measurement in Gaia's final data release and the reduced number of stars (the validity of this scaling and our ability to achieve the compression is established below). For each star the position model $\underline{\vec{n}}_I(t_J)$ was fitted, and subtracted according to Eq. 6.

The sensitivity is largely determined by N, M, T and σ; the values of N, M, T used are pessimistic estimates for the final Gaia values, while the value of σ is slightly optimistic. In particular, Gaia errors vary strongly with magnitude (see Fig. 1 of [32]); a simple estimate of the appropriate error in each measurement derived by averaging over the full magnitude range, using fits to the histogram of mean G magnitude [33], yielded a conservative estimate of $200\,\mu\mathrm{as}$. The data set used here reflects our current best guess of Gaia's ultimate sensitivity but should be updated following future Gaia data releases.

A GW from a high mass, non-spinning binary was injected into this data set; black holes with masses $m_1 = m_2 = 5 \times 10^8 M_\odot$ on a circular orbit of radius 1500 au at a distance of 20 Mpc (orientated with the angular momentum along the line-of-sight) give a circularly polarised GW with frequency $2\pi f = 2 \times 10^{-7}\mathrm{s}^{-1}$ and amplitude $A_+ = A_\times = 3 \times 10^{-14}$. The GW was confidently recovered with $\mathcal{B} = 10^{4.2} > \mathcal{B}_{\mathrm{threshold}}$ and the 1–dimensional marginalised posterior distributions are shown in Fig. 2.

Compressing the GAIA dataset – Searches with $M = 10^5$ stars take days to run; the full Gaia data set with $M > 10^9$ stars is impractically large to search using the Bayesian techniques described. Here we show how the data can be greatly compressed with little loss in sensitivity. The need for compression is even greater when performing an astrometric search for a stochastic GW background because the likelihood involves the inverse of a $M \times M$ correlation matrix [11] (compression for stochastic searches will be addressed in a future publication).

A small number $\tilde{M} (\ll M)$ of points on the sky, called *virtual stars*, are selected. Each virtual star defines a *Voronoi cell* [34] consisting of the points nearest that virtual star. Each real star is identified with the nearest virtual star. Virtual stars are indexed by $\tilde{I} = 1, 2, \ldots, \tilde{M}$ and the Voronoi cells are denoted $\mathcal{V}_{\tilde{I}}$.

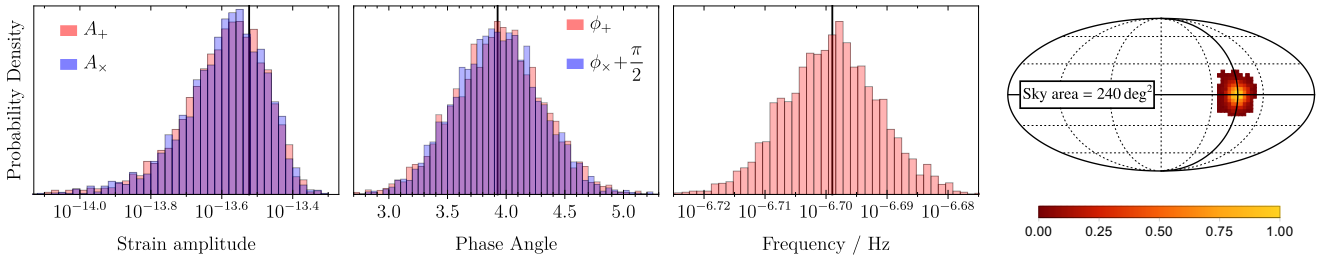

FIG. 2. 1–dimensional marginalised posteriors on $\overline{\Psi}$ (black lines indicate injected values). The injected GW was circularly polarised (i.e $\phi_+ - \phi_\times = \pi/2$) so the $\phi_\times$ posterior is shifted such that it overlaps with ϕ_+. The Mollweide sky map is shown with the area of the 68% credible region given.

The astrometric data set is compressed into a smaller *virtual data set* (quantities associated with the virtual data are denoted with a tilde). All astrometric deflections in a set time interval for stars in a given cell are averaged;

$$\tilde{\mathbf{s}}_{\tilde{I},J} = \frac{1}{|\mathcal{V}_{\tilde{I}}|}\sum_{I\in\mathcal{V}_{\tilde{I}}}\mathbf{s}_{I,J}\,, \quad \frac{1}{\tilde{\sigma}^2_{\tilde{I},J}} = \sum_{I\in\mathcal{V}_{\tilde{I}}}\frac{1}{\sigma^2_{I,J}}\,, \tag{11}$$

where $|\mathcal{V}_{\tilde{I}}|$ denotes the number of real stars in $\mathcal{V}_{\tilde{I}}$. The virtual data $\tilde{\mathcal{S}} = \{\tilde{\mathbf{s}}_{\tilde{I},J} | \tilde{I} = 1,\ldots,\tilde{M};\ J = 1,\ldots,N\}$ (c.f. Eq. 4) may be analysed using the techniques described above for the original data, $\mathcal{S}$.

This compression would be lossless if (i) the noise was described by Eq. 7, and (ii) the astrometric deflections of all stars in a cell were parallel. The deflections vary smoothly across the sky (see Fig. 1) so as $\tilde{M}$ is increased condition (ii) becomes satisfied. In fact, for a given $\tilde{M}$ the sensitivity loss can be estimated by considering the angle between deflections of stars in the same Voronoi cell (see Fig. 3).

While condition (i) cannot be expected to hold perfectly, correlations are not expected to significantly degrade the sensitivity. Temporal correlations will be mitigated against by the fact that between measurements the spacecraft rotates into a different orientation and the starlight strikes a different part of the CCD. Spatial correlations exist, but only at the level of 3% for colocated stars, dropping to 0% for stars separated by 0.7°. As the mission proceeds correlations are expected to reduce [35]. In this first analysis we do not consider correlated errors.

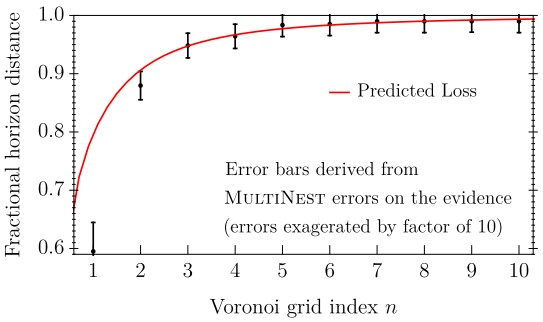

FIG. 3. The horizon distance is reduced (relative to the uncompressed data) during compression onto grid $n = 1, 2, \ldots, 10$. Shown in red is loss estimate obtained by considering the maximum angle between deflections in the same Voronoi cell.

The virtual star locations may be freely specified, e.g. they could be randomly generated. Here they are taken to be the midpoints of the faces of certain polyhedra. The base polyhedron was an icosahedron (the resulting Voronoi cells are called "grid 1"). Successive polyhedra were formed by constructing geodesic domes from the icosahedron — subdividing great circles between vertices into $n = 2, 3, \ldots$ smaller arcs, and constructing n^2 triangles on each face. The midpoints of the faces of the resulting polyhedra give a set of virtual stars and the resulting Voronoi cells are called "grid n". The n^{th} grid has $\tilde{M} = 20 \times n^2$ virtual stars; grids up to $n = 10$ were used. The level of compression can be controlled by varying n.

The mock data described above was compressed onto each of the grids $n = 10, 9, \ldots, 1$ and the virtual data sets searched as before. The Bayes' factor recovered from smaller grids is reduced because stars in the larger Voronoi cells have astrometric deflections which are not parallel and partially cancel each other out in the compression (Eq. 11). This lower Bayes' factor reduces the maximum distance at which the source can be detected; this reduction in *horizon distance* is shown in Fig. 3. The compression loss is independent of the number of real stars. Provided grids with $n \geq 7$ are used the sensitivity loss is less than 1%. The $n = 7$ grid contains $\tilde{M} = 980$ virtual stars; therefore the full Gaia data containing $M > 10^9$ stars can be compressed onto the $n = 7$ grid (a compression factor of $10^9/980 \approx 10^6$) with a sensitivity loss below 1%. The averaging in Eq. 11 gives these impressive compressions because of the smooth, large angle (approximately quadrupolar) pattern in Fig. 1.

GAIA's sensitivity – Here the frequency dependence of Gaia's sensitivity is quantified, along with the effect of nonuniform time sampling (the directional sensitivity variation is quantified in the supplementary material). Multiple mock data sets, similar to those used above, were constructed. The astrometric position of each star was measured $N = 75$ times over a $T = 5$ year pe-

riod; data sets were constructed assuming both uniform time sampling (T_0), and several realistic Gaia samplings, constructed using https://gaia.esac.esa.int/gost/ (these are labeled T_1, T_2, and T_3).

Circularly polarised GWs were injected with different amplitudes and frequencies and the data compressed onto the $n = 10$ grid for analysis. For fixed frequencies in $(10^{-8.5}$–$10^{-6})$Hz, multiple injections were used to find the minimum amplitude where the Bayes' factor exceeds $\mathcal{B}_{\rm threshold}$ for at least 50% of noise realisations (i.e. a detection probability of $> 50\%$). The resulting sensitivity curves are shown in Fig. 4 for each T_α; the variability in the sampling has a small effect on the GW sensitivity.

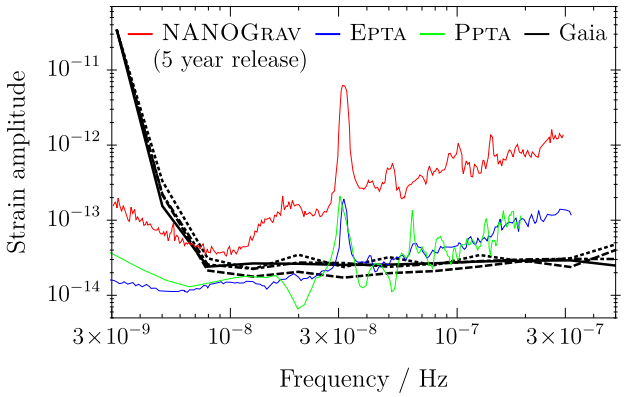

FIG. 4. The black curves show the strain sensitivity of the final Gaia data release using different time samplings (T_0, solid; T_1, dotted; T_2, dashed; T_3 dot-dashed). The coloured lines show 95% PTA upper limits: NANOGrav ([19] red), EPTA ([20] blue) and PPTA ([21] green). These curves show different quantities and are only intended for approximate comparison; the NANOGrav curve is a Bayesian upper limit, the EPTA and PPTA curves are frequentist upper limits, while the Gaia curves show the amplitude necessary to achieve a (conservative) threshold Bayes' factor. It should be noted that these PTA limits are several years old and improve over time; Gaia's sensitivity will not improve further.

The strain sensitivity of Gaia is flat above $f \gtrsim 1/T$ (where $T = 5$ years is the mission lifetime). This is in contrast to the sensitivity of PTAs, which degrade at high frequencies. This discrepancy comes from the fact that GWs cause redshifts (Eq. 1) and PTAs measure *timing residuals* which are the time-integral of redshifts. In the frequency domain, integration over time corresponds to division by frequency; this suppresses the sensitivity of PTAs for frequencies above $f \approx 1/T$. In contrast, astrometric deflections are directly proportional to the GW strain (Eq. 2). This difference in slopes means that it is at mid to high frequencies, $f \gtrsim 10^{-7.5}$Hz, where Gaia will best complement PTA efforts.

Conclusions – GWs cause the apparent astrometric positions of stars to oscillate with a characteristic pattern on the sky (see Fig. 1 and Eq. 2). Gaia is the ideal observatory to make the large number of accurate astrometric measurements necessary to search for low frequency GWs using this effect. This letter summarises recent progress towards a practicable GW search algorithm for the fast approaching final Gaia data release. It has been shown how a large astrometric data set may be greatly compressed with little loss in sensitivity, and the GW sensitivity of Gaia to monochromatic GWs has been quantified and shown to be at a level which is potentially interesting and complementary to that from current PTA searches..

Acknowledgments – This work was supported by Grants MSCA-RISE-2015 690904, STFC ST/L000636/1, ST/H008586/1, ST/K00333X/1, and BIS ST/J005673/1. We thank Jonathan Gair, Stephen Taylor, Michalis Agathos, Ulrich Sperhake and the anonymous referees for their helpful comments.

* cjm96@cam.ac.uk

[1] B. P. Abbott *et al.* (LIGO Scientific- and Virgo Collaborations), Phys. Rev. Lett. **116**, 061102 (2016).
[2] B. P. Abbott *et al.* (LIGO Scientific- and Virgo Collaborations), ApJ **818**, L22 (2016).
[3] P. A. Seoane *et al.* (The eLISA Collaboration), ArXiv e-prints (2013), arXiv:1305.5720 [astro-ph.CO].
[4] C. J. Moore *et al.*, Class. Quantum Grav. **32**, 055004 (2015).
[5] M. A. McLaughlin, Class. Quantum Grav. **30**, 224008 (2013).
[6] M. Kramer and D. J. Champion, Class. Quantum Grav. **30**, 224009 (2013).
[7] G. Hobbs, Class. Quantum Grav. **30**, 224007 (2013).
[8] R. N. Manchester *et al.*, Class. Quantum Grav. **30**, 224010 (2013).
[9] V. B. Braginsky, N. S. Kardashev, A. G. Polnarev, and I. D. Novikov, Nuovo Cimento B Serie **105**, 1141 (1990).
[10] T. Pyne *et al.*, ApJ **465**, 566 (1996).
[11] L. G. Book and É. É. Flanagan, Phys. Rev. D **83**, 024024 (2011).
[12] Prusti, T. *et al.*, A&A **595**, A1 (2016).
[13] A. Vilenkin, Phys. Rev. D **24**, 2082 (1981).
[14] L. P. Grishchuk, Pisma v Zhurnal Eksperimentalnoi i Teoreticheskoi Fiziki **23**, 326 (1976).
[15] J. B. Wang *et al.*, MNRAS **446**, 1657 (2015).
[16] Z. Arzoumanian *et al.*, ApJ **810**, 150 (2015).
[17] W. J. Kaufmann, Nature **227**, 157 (1970).
[18] F. B. Estabrook and H. D. Wahlquist, Gen. Rel. Gravit. **6**, 439 (1975).
[19] Z. Arzoumanian *et al.*, ApJ **794**, 141 (2014).
[20] S. Babak *et al.*, MNRAS **455**, 1665 (2016).
[21] X.-J. Zhu *et al.*, MNRAS **444**, 3709 (2014).
[22] K. J. Lee *et al.*, MNRAS **414**, 3251 (2011).
[23] C. M. F. Mingarelli *et al.*, Phys. Rev. Lett. **109**, 081104 (2012).
[24] L. Boyle and U.-L. Pen, Phys. Rev. D **86**, 124028 (2012).
[25] X.-J. Zhu *et al.*, MNRAS **461**, 1317 (2016).
[26] A. Buonanno, (2007), arXiv:0709.4682 [gr-qc].
[27] S. Taylor, J. Ellis, and J. Gair, Phys. Rev. D **90**, 104028 (2014).

[28] R. van Haasteren *et al.*, MNRAS **395**, 1005 (2009).
[29] H. Jeffreys, *Theory of probability* (Clarendon Press New York, Oxford, 1983).
[30] F. Feroz and M. P. Hobson, MNRAS **384**, 449 (2008).
[31] J. Skilling, AIP Conference Proceedings **735**, 395 (2004), http://aip.scitation.org/doi/pdf/10.1063/1.1835238.
[32] F. Mignard and S. A. Klioner, IAU Symposium **261**, 306 (2010).
[33] https://gaia.esac.esa.int/documentation/GDR1/Catalogue_consolidation/sec_cu1cva/,.
[34] J.-R. Sack and J. Urrutia, in *Handbook of Computational Geometry* (North-Holland Publishing Co., 2000) Chap. 5.
[35] B. Holl *et al.*, A&A **543**, A15 (2012).

Supplemental Material

Significance of the Bayesian detection – Throughout this letter the Bayes' factor is used as a detection statistic with a detection threshold of $\mathcal{B}_{\rm threshold}=10^{1.5}$. In this section of the supplement the suitability of this threshold and the impact on the estimates of the GW sensitivity of Gaia are discussed.

The false alarm probability of a candidate signal with a given Bayes' factor can be established by looking at the distribution of Bayes' factors in the absence of a GW signal. Mock astrometric data sets similar to those used in Fig. 4 in the main letter were generated, but without injecting a mock GW signal. For each of the Gaia sampling functions (T_0, T_1, T_2, and T_3) 100 mock data sets were generated and for each data set MultiNest was used to evaluate the Bayes' factor; histograms of the results are shown in Fig. S1. None of the 4×100 mock data realisations comes close to exceeding the detection threshold; therefore, under the noise assumptions used, this threshold corresponds to a false alarm probability of well below 1%.

In Fig. 4 of the main letter the detection threshold is used to estimate Gaia's GW sensitivity as a function of frequency; i.e. the critical GW strain amplitude, $h_{\rm Gaia}(f)$, at which the detection probability exceeds 50%. In fact the strain sensitivity depends only weakly on the threshold Bayes' factor; it can be shown analytically that in the limit of large signal strength $h_{\rm Gaia}(f)\propto\sqrt{\log(\mathcal{B}_{\rm threshold})}$ (see [4] for a discussion in the context of estimating the sensitivity of PTAs) and the applicability of this scaling has been checked numerically in this case.

The results in Fig. S1 were derived with the idealised, stationary Gaussian noise described by Eq. 7 in the main letter. The noise properties of the final Gaia data release will be more complicated, and in particular may contain systematic effects affecting the distribution of Bayes' factors in Fig. S1. Therefore, Fig. S1 cannot be used to accurately assess the false alarm probability of a candidate signal in real Gaia data. However, the results in Fig. S1 do further motivate our conservative choice of threshold (corresponding to Jeffrey's criterion for "very strong" evidence) by showing it is likely to yield a very low false alarm probability. Additionally, the estimates of Gaia's GW strain sensitivity in Fig. 4 of the main letter depend only weakly on the exact choice of the threshold. The results in Fig. 4 of the main letter are a robust estimate of Gaia's ultimate strain sensitivity but should be updated as more information on the noise properties of the final Gaia data release become available.

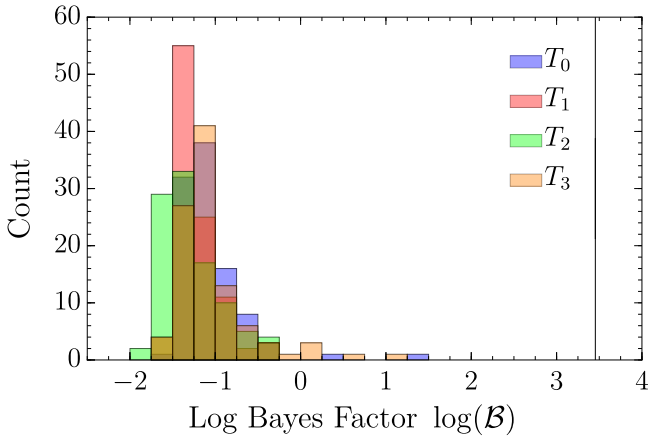

FIG. S1. Histograms of the Bayes' factor for mock data realisations *not* containing GW signals. The vertical line indicates the detection threshold $\mathcal{B}_{\rm threshold}=10^{1.5}$. The four different histograms, each containing 100 samples, show the distributions for the different Gaia samplings used in the main letter; changing the assumed sampling does not have a significant effect on the distribution of the Bayes' factor. No noise realisation exceeds the detection threshold.

GAIA's directional sensitivity – In the main letter the frequency dependence of Gaia's GW sensitivity was established using multiple mock injections and recovery (see Fig. 4 of main letter). In this section of the supplement to the letter similar techniques are used to assess how this sensitivity varies with direction; i.e. the *antenna pattern* response to GWs.

The distribution of stars on the sky is not uniform (as was assumed for simplicity in the main letter), therefore astrometric measurements are not uniformly sensitive to GWs from all directions. In this supplement the directional dependence of Gaia's GW sensitivity is quantified.

For the purpose of this discussion, and without loss of generality, let the GW source lie on the positive z–axis ($\vec{q}=\{0,0,1\}$) and a star lie at angular distance γ from the GW source, in the x–z plane ($\vec{n}=\{\sin\gamma,0,\cos\gamma\}$). Using the general plane wave metric perturbation (Eq. 4 of the main letter),

$$h_{ij}\left(\overline{\Psi}\right)=\left(A_{+}H^{+}_{ij}(\vec{q})e^{\mathrm{i}\phi_{+}}+A_{\times}H^{\times}_{ij}(\vec{q})e^{\mathrm{i}\phi_{\times}}\right)e^{2\pi\mathrm{i}ft}\,,$$

the magnitude of the astrometric deflection vector (Eq. 3 of the main letter),

$$\delta n_i=\frac{n_i-q_i}{2(1-\vec{q}\cdot\vec{n})}h_{jk}(\mathrm{E})n^{\hat{j}}n^k-\frac{1}{2}h_{ij}(\mathrm{E})n^j\,.$$

is given by

$$|\delta\vec{n}|=\frac{1}{2}\sqrt{A_{+}^2+A_{\times}^2}\,\sin\gamma\,.$$

The largest astrometric deflections occur for stars which lie orthogonal to the GW source direction (i.e. $\gamma=\pi/2$). Therefore, it is expected that Gaia's peak sensitivity will occur orthogonal to regions of high stellar density (i.e. at the galactic poles).

Mock data sets, similar to those used in the main letter to quantify Gaia's frequency sensitivity, were constructed, but now using the $M=1.1\times10^9$ real stars in the first Gaia data release (`https://www.cosmos.esa.int/web/gaia/dr1`). The astrometric positions were sampled $N=75$ times uniformly over a $T=5$ year mission. Into these mock data sets were injected circularly polarised GWs from 500 sky locations approximately uniformly spaced on the sky. Each mock data set was compressed onto the $n=5$ Voronoi grid to be efficiently searched. The variation in *horizon distance* (the maximum distance at which the source can be detected) with the source's sky location is plotted in Fig. S2. The results in the figure verify the expectation that peak sensitivity occurs at the galactic poles.

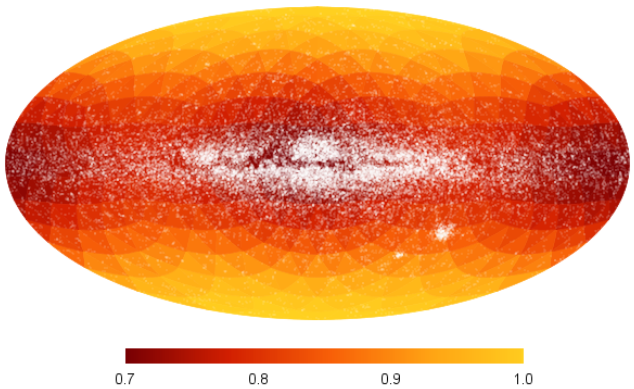

FIG. S2. The variation in Gaia's sensitivity over the sky, $F(\theta,\phi)$. A sample of 60,000 stars drawn randomly from the Gaia catalogue are shown as white dots. The sensitivity varies by $\sim30\%$ across the sky with minima at (and antipodal to) the galactic centre, and maxima at the galactic poles.