Noise in Radio Pulsar Timing Array Data

Pulsar timing is the comparison between modeled and observed pulse times of arrival (TOAs). Astrophysical parameters leave a unique imprint on the timing *residuals* between data and model: a spin frequency error $\delta\nu$ produces linear residuals, an error in $\delta\dot{\nu}$ leaves quadratic residuals, a position error induces an annual sinusoid, etc. (*S1*). Pulsar timing can characterize a wide range of astrophysical phenomena and constrain fundamental physics (*S2, S3, S4, S5*), particularly when using millisecond pulsars (*S6*), whose stable rotation and high pulse rate facilitate high-precision timing. Pulsar timing arrays—PTAs, collections of many timed pulsars (*S7, S8*)—extend sensitivity to global signals, such as the effects of weak, low-frequency gravitational waves (GW). The imprint of these subtle signals can be obscured by noise, and many noise sources are carefully mitigated from radio PTA data to reveal signals as faint as the $A_{\mathrm{gwb}} = 2 - 3 \times 10^{-15}$ one (whose typical amplitude in residuals is $< 100\,\mathrm{ns}$) discussed in the main text. To place the *Fermi* PTA results in context, we give an overview of the radio PTA noise budget below. These main noise sources are summarized in Table S1.

White Noise

White noise ($<1\,\mathrm{d}$ time scales) cannot produce GWB-like signals, but it must be modeled to reliably estimate the significance of such signals and the precision of timing model parameters. White noise models are typically implemented through the parameters `EFAC` and `EQUAD` (*S9*), which modify the measurement uncertainty σ_i of a TOA as $\sigma_i^2 \rightarrow \mathtt{EFAC}^2\sigma_i^2 + \mathtt{EQUAD}^2$. Additional `ECORR` parameters can describe correlated noise, e.g. jitter. Because the noise sources depend on observing frequency, pulsar, and observing system, tens to hundreds of such parameters may be needed to represent the excess white noise in a radio PTA data set. The mapping of these parameters onto the actual noise sources may be imperfect, leaving un- or over-modeled white noise in the residuals.

Table S1: Sources of noise in radio and γ-ray pulsar timing array data. The list is incomplete and qualitative: for each source, we attempt to give an estimate of the importance (amplitude in residuals, or difficulty of complete mitigation) and complexity (degrees of freedom in existing models for a typical pulsar).

| Noise Source | Radio | | γ ray | | Note |
	Impact	d.o.f.	Impact	d.o.f.	
White Noise					
Measurement	moderate	–	serious	–	Major limiting factor for γ rays.
RFI	minor	?	–	–	Impact of RFI varies widely.
Calibration	minor	?	–	–	Affects certain pulsars/observing systems.
Jitter	moderate	10s	–	–	Jitter is important for high S/N observations.
Red Noise					
DM variation	serious	100s	–	–	DM(t) drives radio PTA observing strategies.
Solar wind	moderate	10s?	–	–	Solar wind mitigation poorly supported.
Scattering	moderate	100s	–	–	Affects some pulsars/low radio frequencies.
Pulse variation	moderate	0–10s	–	–	No γ-ray MSP pulse profile changes known.
Discontinuities	moderate	10s	–	–	Continuity specific to *Fermi*-LAT data set.
Spin noise	serious	10s–100s	serious	10s	Fewer d.o.f. needed for less precise LAT data.

Measurement Error Radio PTAs use large, sensitive radio telescopes. Light collected from the antenna is transduced, amplified, digitized, dedispersed, and filtered into frequency channels. Pulses are stacked into a single pulse profile by "folding" the data at the instantaneous pulsar spin period. The observation time is noted against a precise clock, and the offset of the recorded pulse to this time yields the pulse TOA. The precision of a TOA can be estimated from the random noise in the pulse profile which arises from electronics noise, background radiation from astrophysical sources, and terrestrial spillover into the antenna.

Jitter "Jitter" (*S10, S11*) is a TOA bias caused by the finite number of pulses received in an observation. It requires a dedicated noise model because its impact depends on pulsar brightness, which varies due to IISM scintillation (below), and because it is correlated over observing bandwidth (*S10*).

Radio-frequency interference   RFI affects all radio observatories but is less severe for those in remote areas. Strong RFI can render an observation unusable. Fainter RFI can often be characterized as narrowband—contaminating portions of the observing band, intermittently or continuously—or impulsive—affecting much of the band briefly. RFI biases TOA measurements (*S12*), with impulsive RFI altering the observed pulse profile, while narrowband RFI changes the effective observing frequency, inducing residuals from DM(t) corrections (below). RFI noise is effectively white, but a changing RFI environment can introduce biases on longer time scales.

Polarization   Pulsars are highly polarized radio sources (*S13*), so recording a stable pulse profile requires accurate polarization calibration. Since most radio telescopes use an altitude-azimuth mount, the pulsar polarization angle rotates as the source tracks across the sky. If a radio receiver has non-zero cross-polarization, this alters the pulse profile (*S14*). Using bright pulsars to characterize the receiver response as a function of parallactic angle (*S15*) can eliminate much of the apparent TOA variation. As with RFI, observational bias can couple this white noise to longer time scales.

Red Noise

Red noise operates on longer time scales and includes signals with GWB-like power spectral densities, therefore it must be carefully modeled.

Effects of the Ionized Interstellar Medium   The ionized interstellar medium (IISM) is turbulent (*S16*) and contains lens-like structures (*S17*). Radio waves from a pulsar encounter a continually changing electron density and bend to new trajectories with a degree proportional to the inverse square of the radio frequency ν^{-2}. The wide range of observational consequences includes dispersive delays, strong intensity variations (scintillation), broadening of pulses due

to multipath scattering, and higher-order effects, such as apparent position shifts (*S18,S19,S20*).

The line of sight through the IISM changes constantly due to relative motion, and the most noticeable effect is scintillation—frequency- and time-dependent variations in the received intensity. Scintillation can render some observations useless if the pulsar is too faint, and it shifts the effective observing frequency by enhancing or depressing the signal over the observational band, magnifying the effects of DM uncertainty discussed below.

The dominant effect of the IISM on residuals is the dispersive delay $\tau(t) \propto \mathrm{DM}(t)/\nu^2$, which changes measurably on timescales of days to weeks (*S21*). Low-frequency measurements yield higher DM precision, and since their inception, PTAs have monitored DM(t) using multiple receivers covering a wide range of frequencies (e.g. 350 MHz to 4 GHz). New wideband receivers (*S22*) further improve DM measurements, as do high-cadence observations by low-frequency observatories like CHIME (*S23*) and LOFAR (*S24*). However, the apparent DM also depends on ν because low-frequency waves scatter through a larger volume of the IISM, so even precise DM(t) measurements cannot perfectly correct $\tau(t)$ at other radio frequencies (*S25,S24*).

DM evolution contributes potentially hundreds of degrees of freedom to a pulsar timing model. One approach uses `DMX` parameters (*S26*) to tabulate DM measurements from TOAs within discrete time segments. Another makes use of a constrained stochastic model for DM(t) (see "spin noise" below), which has fewer effective degrees of freedom but can only model stationary DM variations, which is generally insufficient to capture observed variations (*S21, S17, S27*). These different approaches to DM modeling can produce discrepant results: NANOGrav, using `DMX` (*S28*), and the PPTA, using the stochastic model (*S29*), find marginally different values for spin noise (below) for pulsars common to the two PTAs.

Other IISM effects can be approximated using corrections of the form ν^n, e.g. ν^{-4} for multipath scattering, and other powers for higher-order effects (*S30*), though only ν^{-2} corrections are in widespread use today. Scattering can be corrected exactly for a few bright MSPs using cyclic

4

spectroscopy (*S31*), but a more universal method to incorporate such corrections is "pulse portraiture", a model of a pulse profile over typically an octave or more of bandwidth (*S32*). This portrait attempts to track the intrinsic variation in the pulse shape with frequency, and then is convolved with models of the IISM, e.g. a ν^{-2} kernel to account for a changing DM, a ν^{-4} kernel for altered scattering, etc.

In summary, the monitoring, measurement, and interpolation of IISM-induced residuals is an observationally and computationally expensive endeavor and a source of unmodeled error, and the reduction of these effects to the $<\approx 1\,\mu$s level is a major achievement of radio PTAs.

The Solar Wind The solar wind makes a small contribution (typically 10^{-4}–10^{-3} DM units) to the total DM for any pulsar, but it varies daily due to solar activity (*S33, S34*) and annually as the apparent pulsar-sun angular separation changes. Pulsars near the ecliptic, e.g. PSR J0030+0451, are the most strongly affected. Current models and observational strategies are insufficient to capture the contribution of the solar wind to pulsar timing residuals (*S33*), and most PTA data analyses employ a static model. Because the coherent structure of the solar wind subtends large portions of the sky, uncorrected solar wind variations can contribute correlated noise to pulsars.

Pulse profile variations Pulse profile variations induce apparent variations in TOAs and cannot be easily mitigated. Observed variations in PSRs J1643$-$1224 (*S30*) and J0437$-$4715 (*S12*) have been attributed to a transient magnetospheric reconfiguration, while alterations in the pulse profile of PSR J1713+0747 are associated with rapid DM variations and recovery (*S35*). No pulse profile variations have been observed in γ-ray MSP profiles, though correlated variations in intensity and pulse shape have been observed for the young PSR J2021+4026 (*S36*).

Discontinuities Each observation of a pulsar is referenced to international time standards, a well-established process with typical errors expected to be <10 ns. However, instrumentation

changes alter the signal propagation time through the system which must be measured or fit from the data, incurring potential systematic error or additional degrees of freedom in timing models (*S12*). Fitting the 10–20 `JUMP` parameters for a typical pulsar acts as a high-pass filter, reducing sensitivity to low-frequency signals.

Spin Noise Spin noise, also called timing noise or intrinsic noise, is most well known in young, more slowly spinning pulsars (*S37*), but is also present in MSPs with an amplitude that is related to the pulsar spin-down power (*S38*). This intrinsic noise is generally well modeled with a power-law spectrum of the same form as Equation 2, and it is represented in data either as long-term correlations between the TOAs, with a non-diagonal covariance matrix (*S39, S40*), or, as in this work, using a Fourier expansion with the coefficients constrained to follow the assumed power spectrum (*S41*). The prescriptions both assume that spin noise originates from a stationary process. This assumption may work well if such spin noise originates from superfluid turbulence within the neutron star (*S42*) or fluctuations within the pulsar magnetosphere, but it may be a poor approximation if such noise arises from e.g. switches between metastable equilibria of the magnetosphere (*S43, S44, S45*). Most candidate mechanisms predict identical spin noise in the radio and γ-ray bands.

Gamma-ray Pulsar Timing

Relatively short ($<1\,\mathrm{hr}$) radio observations yield a direct measurement of the pulsar spin phase at each epoch (the TOA). Conversely, the LAT collects individual γ-rays, and it is impossible to interpret such an event as a TOA, since it could originate from any phase of the neutron star rotation or from a background source. Pulsar timing with the LAT thus depends on photon-by-photon analysis.

Photon-by-photon analysis When a pulsar is in the LAT field of view, photons arrive with a rate that depends on $\phi(t)$, the pulsar spin phase, with $\delta\phi = 1$ being a complete rotation. The distribution of photons over rotational phase, i.e. the pulsar light curve, is denoted $f(\phi)$. This pulse profile can be normalized such that $\int_0^1 f(\phi)\,d\phi = 1$, and $f(\phi)$ is then also a probability density function for observing a photon at a given ϕ. The LAT has a relatively broad, energy-dependent angular resolution which causes photons from sources to overlap. We can estimate the probability that a given photon originated from the pulsar by building a model of the γ-ray sky with the position and spectrum of each source. We fold this model through the instrument response, and the probability, or photon weight, is simply the ratio of the pulsar intensity to the total intensity at the observed energy and position. Using photon weights optimizes sensitivity (*S46, S47, S48*).

Given these weights and the pulse profile model, the data follow the Poisson likelihood,

$$\log \mathcal{L}_{data} = \sum_i \log\Big(w_i f(\phi[t_i, \lambda, \beta]) + (1 - w_i)\Big), \tag{S1}$$

with w_i and t_i the weight and time of the ith photon and λ the timing model parameters (constants have been suppressed). The parameters β are Fourier coefficients describing an arbitrary additional component, e.g. a random noise process such as spin noise or GWB perturbations. This method is fully general, and can measure or constrain any conceivable timing model parameter by maximizing the likelihood.

In addition to determining λ, we can use Equation S1 to make inferences on the GWB amplitude. We assume its signal is a Gaussian random process and that the full likelihood can be written

$$\log \mathcal{L} = \log \mathcal{L}_{data} - 0.5 \sum_{kl} \mathcal{C}_{kl}\beta_k\beta_l - 0.5 \det \mathcal{C}, \tag{S2}$$

with the entries of the diagonal covariance matrix C given as $C_{kl} = \delta_{kl}P(f_k, A_{\mathrm{gwb}})/2$, i.e. the covariance is given by the power spectrum evaluated at the frequency $f_k = k/t_{\mathrm{obs}}$, with t_{obs} the

151   length of the data; see (*S41*). This same formulation also describes spin noise, *mutatis mutandis.*

152   **TOA estimation**   In some cases, it is also possible to form TOAs from LAT data. In the ab-

153   sence of pulsation, $f(\phi) = 1$ and $\log \mathcal{L}_{data} = 0$. Thus, the total log likelihood can be interpreted

154   as a measure of pulsation significance. The log likelihood grows as data are accumulated, and

155   once it surpasses a threshold (roughly $\delta \log \mathcal{L} > 20$) there is enough information to measure an

156   offset $\delta\phi$ between the template $f(\phi)$ and the observed data. To convert $\delta\phi$ to a TOA, we choose

157   a reference time near the mid-point of the integration t_0 such that $\phi(t_0) \bmod 1 = 0$ according

158   to an initial timing model. We iteratively determine δt such that $\phi(t_0 + \delta_t) \bmod 1 = \delta\phi$. Be-

159   cause the precise photon timestamps are referenced to the on-board GPS clock, the resulting

160   TOA, $t_0 + \delta t$ is in the UTC(GPS) time system (*S49*). As the log likelihood increases, the shape

161   of $\mathcal{L}(\delta\phi)$ becomes more Gaussian. Thus, a higher $\log \mathcal{L}$ threshold both reduces the systematic

162   uncertainty of a TOA and improves the error estimation process itself (*S50*), at the expense of a

163   longer integration period.

164   This tradeoff between efficacy and integration time determines whether the TOA technique

165   is suitable for a particular purpose. The faintest MSPs in our sample require more than six

166   months to generate a TOA, so they would be insensitive to the annual sinusoidal residual from

167   a position error. The brightest MSPs yield a TOA every two weeks, easily resolving it. Similar

168   arguments apply to *a priori* unknown signals, such as residuals from spin noise or the GWB.

169   Care must be taken to ensure self-consistency between the timing model and the TOA estima-

170   tion, in stark constrast to radio pulsar timing, where TOAs can generally be obtained in $< 1\,\mathrm{hr}$,

171   independently of the detailed pulsar timing model.

172   **Advantages of γ rays**   The LAT has low sensitivity compared to most radio timing campaigns,

173   but properties of the γ-ray data and its acquisition are well suited for pulsar timing, and to PTA

174   work specifically:

1. None of the IISM effects described above affect γ-ray timing, and they are also immune to the eclipses and DM variations affecting compact interacting binaries (*S51*).

2. γ-ray data have simple noise models, requiring at most spin noise, easing computational requirements and reducing systematic uncertainty.

3. The LAT continually monitors every MSP in the sky, such that when a new MSP is discovered, archival LAT data can provide a full pulse timing history.

4. The data span for each pulsar is uniform. This ensures that each pulsar is sensitive to the same spectrum of gravitational waves and enables simpler computational approaches.

5. There are no discontinuities in the data. The instrument is essentially unchanged since its launch, with no interruptions to its clock, admitting no potential of systematic error or loss of sensitivity from fitting `JUMP` parameters.

6. The 12.5 yr span of LAT data is comparable to those of modern radio PTAs. Such long data sets are critical in characterizing steep spectrum processes like the GWB.

7. Polarization has a negligible effect on γ-ray pulsar observations, eliminating potential calibration errors.

8. γ-ray data are free from the effects of radio-frequency interference (RFI) and there is no analogous counterpart in the γ-ray band.

9. The origin of pulsed γ-ray emission is more well understood, likely originating from synchro-curvature radiation precipitated by large-scale and stable electric fields in the outer magnetosphere or current sheet (*S52*). In contrast, the origin of the radio emission in MSPs is less certain and is subject to vagaries of conditions close to the star surface, and variations of the pulse profile have been observed in PTA pulsars (*S30, S35*).

9

10. In comparison to other high-energy observations, e.g. soft X-rays, there are no gaps due to solar avoidance.

Data Preparation

Photon data   Properties of each γ-ray are determined by reconstructing the particle interactions in the LAT detector into measured quantities of incident direction, energy, and arrival time (*S53*), with the timestamping precision having been verified as <300 ns (*S54*). We begin with a set of 127 MSPs with γ-ray counterparts in the 4FGL (*S55*) DR2 (*S56*) catalog. For each pulsar, we download data from the *Fermi* Science Support Center[1] and process it using the *Fermi* Science Tools v2.0.0. We select all data between 2008 Aug 04 (MJD 54682) and 2021 Jan 28 (MJD 59242) with reconstructed energy 0.1 GeV$<$E<10 GeV, measured zenith angle $<100°$, and a reconstructed incidence direction placing the photon within $3°$ of the pulsar position. Using the 4FGL-DR2 sky model, we assign each photon a weight (as in Equation S1) using *gtsrcprob* and restrict attention to events with $w > 0.05$. This selection retains the great majority of the pulsar signal while eliminating background photons that increase computational costs. We use the PINT software package (*S57*) to assign pulsar spin phase to photons.

Ephemerides   For both the TOA-based and photon-by-photon analyses described below, we require timing models for each pulsar. These have been produced using data from: the Nançay Radio Telescope (NRT), Green Bank Observatory (GBT), Arecibo Observatory (AO), the Parkes telescope/Murriyang (PKS), Jodrell Bank Observatory (JBO), and the Giant Metre-wave Radio Telescope (GMRT). From these initial ephemerides, we (re)fit some parameters with LAT data using PINT to map timing model parameters to pulse phase and directly optimizing the parameters of Equation S1. We always fit for ν and $\dot{\nu}$, the spin frequency and spin-down rate. Beyond

---

[1]https://fermi.gsfc.nasa.gov/ssc/data/access/

this, for most pulsars, the timing model parameters are very precisely constrained by the radio data, and we keep them fixed during our analysis, equivalent to taking a strong prior. Some parameters are more precisely determined with LAT data, and in such cases, the LAT-optimized parameters are also allowed to vary during the GWB analysis. An overview of these parameters for the 35 MSPs in the GWB analysis sample is provided in Table S2, and the timing models accompanying this work provide the exact degrees of freedom for each pulsar.

To gauge the effect of the additional degrees of freedom on the GWB limits for MSPs with LAT-constrained timing models, we computed limits with those values held fixed to their maximum likelihood values. In all cases, this improved the resulting limit, typically by 5–10% (see Table S2), and up to 15–25% for a few binaries. In the cases of PSRs J1858−2216 (21%) and J2034+3632 (13%), the relatively large changes can be attributed to improved numerical stability resulting from the reduced degrees of freedom. The other marked differences occur for pulsars with strong variations in the orbital period, J1810+1744 and J1959+2048, indicating possible degeneracy between these parameters and the GWB parameters. Because the time scales associated with these processes are widely separated, this improvement could also result from the substantial reduction in degrees of freedom, viz. 8 and 10 orbital frequency derivatives. In this case, substantial improvement could be had by using a constrained representation of the orbital period variations (*S58*).

Solar system ephemeris   In all cases, we have used the DE421 solar system ephemeris (*S59*), whose widespread use in the pulsar timing community makes it a natural choice when collecting timing solutions from multiple observatories and teams. It has been shown (though see (*S60*)) that the choice of ephemeris can alter residuals at the $\sim$100 ns level (*S61, S62*), but the LAT data are not yet sensitive to such small effects.

Table S2: Properties of the 35 MSP ephemerides used for GWB analyses, with the observatory column indicating the primary source of data. "LAT" entries jointly fit the indicated parameter family with the GWB signal, e.g. position and proper motion ("astrometry") or "binary" parameters such as the orbital period/period derivative (FBX). "Change" indicates the fraction of the original limit obtained when these parameters are instead fixed. The final column indicates pulsars which use fewer than five harmonics to represent stochastic signals (see text).

Name	Observatory	LAT Parameters	Change	Harmonics
J0030+0451	NRT	–	–	–
J0034−0534	NRT	–	–	–
J0101−6422	LAT	binary,astrometry	0.95	–
J0102+4839	NRT	–	–	–
J0312−0921	LAT	binary,astrometry	0.86	3
J0340+4130	GBT	–	–	–
J0418+6635	LAT	astrometry	0.95	–
J0533+6759	LAT	astrometry	0.95	–
J0613−0200	NRT	–	–	2
J0614−3329	NRT	–	–	–
J0740+6620	NRT	–	–	–
J1124−3653	LAT	binary,astrometry,FB1	0.92	–
J1231−1411	NRT	–	–	–
J1513−2550	NRT	–	–	2
J1514−4946	LAT	binary,astrometry	0.98	–
J1536−4948	LAT	binary,astrometry	0.94	–
J1543−5149	PKS	–	–	1
J1614−2230	NRT	–	–	–
J1625−0021	LAT	–	–	–
J1630+3734	JBO	–	–	–
J1741+1351	NRT	–	–	2
J1810+1744	LAT	binary,astrometry,FB1−8	0.73	–
J1816+4510	LAT	binary	0.97	–
J1858−2216	LAT	binary,astrometry	0.79	1
J1902−5105	LAT	binary,astrometry	0.94	–
J1908+2105	LAT	binary,astrometry	0.89	–
J1939+2134	NRT	–	–	–
J1959+2048	LAT	position	0.88	–
see text	LAT	position,binary,FB1−10	0.46	–
J2017+0603	NRT	–	–	–
J2034+3632	LAT	astrometry	0.87	2
J2043+1711	NRT	–	–	–
J2214+3000	NRT	–	–	–
J2241−5236	PKS+LAT	binary	0.97	–
J2256−1024	GBT+NRT+LAT	binary	0.91	–
J2302+4442	NRT	–	–	–

Pulse profile templates   We model the pulse profile, $f(\phi)$, using 1–8 wrapped Gaussian components—i.e., forced to be periodic—which are also fit using maximum likelihood, with brighter pulsars having more components. We hold these templates fixed in the GWB studies described below.

GWB Analysis

MSP sample selection   Of the 127 MSPs with good γ-ray spectral models, we have timing solutions for 114, of which only some are suitable for a GWB analysis. For each pulsar, we estimate the white noise level, i.e. the typical amplitude of random fluctuations from the timing model, by averaging the measured amplitudes of 5 harmonics (β_k in Equation S2). In the absence of other noise, this value is a good proxy for sensitivity. We also calculate the total significance of the pulsed signal via $\log \mathcal{L}$, which is a good proxy for the number of independent TOAs that can be obtained for a pulsar. These two values, shown for the full sample in Figure S1, are correlated but with substantial scatter. At a given intensity, a pulsar with a narrower pulse or faster spin frequency has better timing precision.

We eliminate faint MSPs, which cannot constrain the GWB but would increase computational complexity, by requiring a white noise level $< 500\,\mu\mathrm{s}^2\,\mathrm{yr}^{-1}$, corresponding to a typical one-year TOA precision of $\sim 16\,\mu\mathrm{s}$. We further require a total $\log \mathcal{L} > 300$ for pulsars in the TOA-based analysis, ensuring each TOA is significantly detected. We set $\log \mathcal{L} > 150$ for the photon-by-photon analysis to reduce problems with underconstrained parameters (see below). In addition to these requirements, we have eliminated PSRs J1311$-$3430, J1555$-$2908, J2215+5135, and J2339$-$0533. The first two have strong spin noise, far in excess of a possible GWB signal, and the extreme amplitude causes numerical issues in the algorithms described below. The latter two pulsars have strong orbital period variations that cannot be modeled adequately with current pulsar timing software. From the 114 initial pulsars we thus arrive at 35
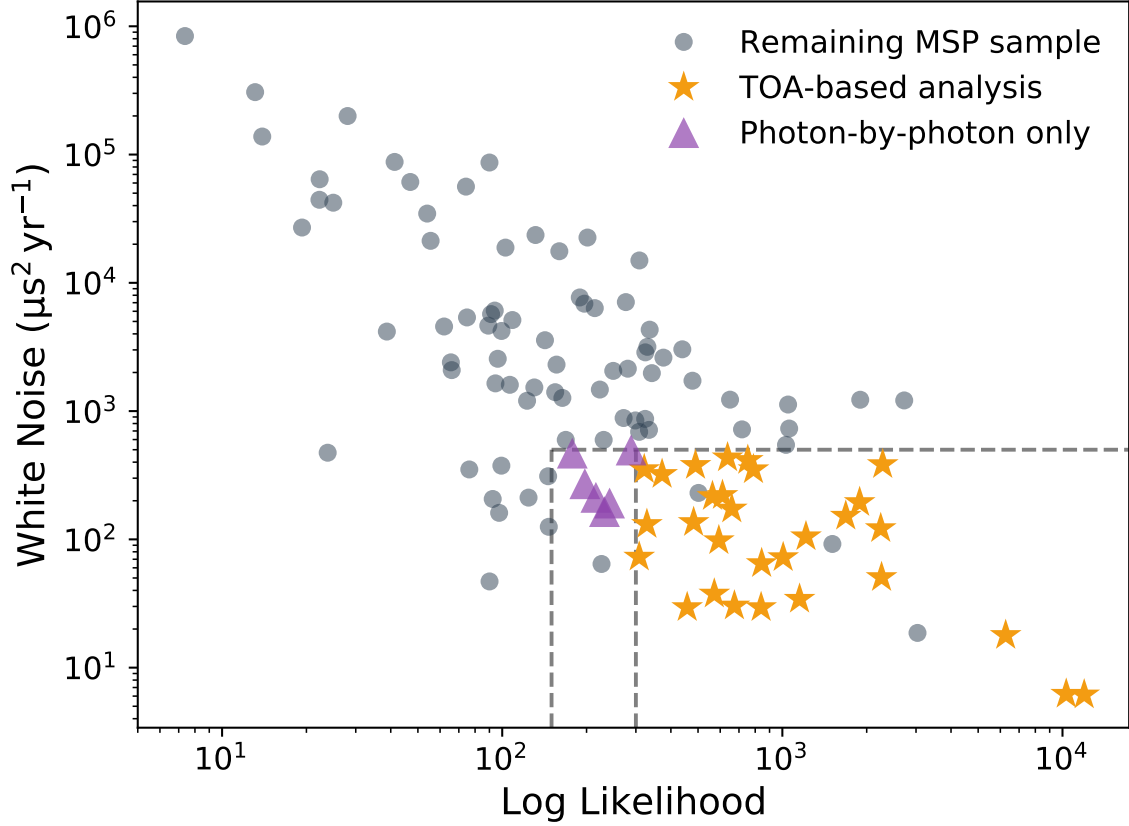
13

Figure S1: Key properties for the 114 MSPs in the parent sample. The y-axis shows the sensitivity, as estimated by the limiting amplitude of a white noise process, while the x-axis gives the source significance, or brightness, as estimated by its total log likelihood, $\log \mathcal{L}$. Gray dashed lines indicate sample divisions, with the 29 orange stars indicating pulsars used in both analyses and the 6 purple triangles those MSPs suitable for photon-by-photon analysis only. Gray circles indicate remaining MSPs.

pulsars in total and 29 suitable for TOA-based analysis.

TOA estimation   We obtain TOAs following the method of (*S50*), finding them generally to be of good quality, though those pulsars near the log-likelihood threshold show asymmetric error bars and occasional non-detections, indicating a modest breakdown in the Gaussian approximation. For each pulsar, we computed TOAs with a cadence of 2, 1.5, and 1 TOAs per year, yielding a total of 25, 19, and 12 TOAs, and analyzed each of the data sets before choosing a single representative cadence for the final GWB analysis.

TOA-based GWB analysis   For each pulsar, we analyzed the three data sets with varying cadences using TEMPONEST (*S63*), which can constrain a variety of single-pulsar noise models and produce single-pulsar GWB limits via nested sampling methods. In particular, we tested for the presence of an intrinsic noise process ("red noise", RN) and for excess white noise (WN). RN models, like the GWB, are assumed to be stationary processes with a power-law power spectral density. WN models are implemented through the parameters EFAC and EQUAD (see above). In general, we expect no WN in the γ-ray data. However, information is lost when reducing the full photon data to a single TOA and (assumed Gaussian) uncertainty. This could take the form of increased scatter of TOAs and appear as a WN process.

Table S3 reports the priors on the noise model parameters used in both TEMPONEST and ENTERPRISE. We use 12 frequency components to model low-frequency noise processes. In all cases, we compute the Bayesian evidence for models with and without WN and RN, and the results are presented in Table S4. As expected, we observe that the preferred noise model for each pulsar excludes RN and WN processes for all pulsars save PSR J1959+2048 and J2241−5236. Both of these pulsars exhibit orbital period variations, and the excess white noise (detected with very modest Bayes factor $\sim$7), could indicate unmodeled period variations. The improved treatment of such pulsars is an important avenue for future work.

15

Table S3: Prior ranges for the noise models used in TEMPONEST and ENTERPRISE. The amplitudes are scaled such that the resulting power spectral density follows Equation 2.

Parameter	Prior ranges
EFAC	$0.1, 5$
$\log_{10} (\text{EQUAD s}^{-1})$	$-9, -5$
$\log_{10}$ RN Amplitude	$-18, -10$
RN Spectral index	$0, 7$
GWB Amplitude (LinearExp)	$-18, -9$

In general, we find that the results, including the inferred limit on the GWB, are consistent over the three cadences. For the faintest pulsars, we expect to observe more systematic errors associated with the (poor) Gaussian approximation when using the 2-per-year cadence as compared to the sparser cadences, and this is the case for PSRs J0533+6759, J0740+6620, J1939+2134, and J2034+3632, for which more complicated noise models were modestly preferred. With longer integrations, the preferred model has no noise. On the other hand, using the highest cadence provides the most information about noise processes, particularly at higher frequencies. Therefore, when possible we adopt the 2-per-year cadence. The cadence chosen for each pulsar is indicated in Table S4.

We next employ ENTERPRISE (*S64*), which is capable of both single-pulsar and multiple-pulsar analyses. Figure 2 shows the TEMPONEST and ENTERPRISE limits for individual pulsars to be in excellent agreement. In the multiple-pulsar analyses, we have focused on the simpler "common mode" process, for which the analysis produces a joint posterior for a fixed-amplitude, identical-spectrum noise process with an independent realization in each pulsar data set. We have also performed a true correlation analysis, with quadrupolar cross-correlation amplitudes given by (*S65*). These results (see Table S5) are essentially identical to the common mode analysis. Because the correlation amplitudes are typically ∼0.1, most of the constraining power for a GWB signal near threshold is in the common mode channel. We discuss these

multiple-pulsar results below, in the context of our results and comparison of methods.

Photon-by-photon GWB Analysis TOA-based methods are fast and well-tested, but as discussed above the integration time to produce a TOA can be prohibitive. We have therefore implemented a fully unbinned, photon-by-photon approach which directly maximizes the likelihood of Equation S2. This approach eliminates any errors made by assuming that TOA uncertainties are Gaussian. Furthermore, it allows simultaneous optimization of the timing model parameters λ with the GWB, though most timing model parameters are held fixed due to the strong radio prior.

For each pulsar, we represent an unknown GWB process ($\Gamma = 13/3$) with the Fourier coefficients β with frequencies $1/t_{\mathrm{obs}}$ to $5/t_{\mathrm{obs}}$. We have verified that using higher frequencies does not contribute significantly to GWB constraints (*S28*) and makes the numerical analysis described below less robust. Even so, for some faint pulsars—typically those with additional free timing model parameters—the data were insufficient to constrain all five harmonics, and we used fewer frequencies as indicated in Table S2.

In order to directly characterize A_{gwb}, we marginalize the nuisance parameters λ (ν, $\dot{\nu}$, and any LAT-optimized parameters in Table S2) and β, the Fourier coefficients of the GWB signal for each pulsar. This can be done approximately analytically by expanding the log likelihood as a quadratic form and solving the resulting Gaussian integrals; (*S66*) show that the likelihood surface for β is generally Gaussian. We do this for each pulsar, scanning A_{gwb} over the range 10^{-20} to 10^{-10}. The marginal log likelihoods obtained in this way, as a function of A_{gwb}, are shown in Figure S2. With the uniform prior, they are equivalent to the logarithm of the posterior distribution.

To obtain single pulsar limits, we simply integrate the posterior density function until we have accumulated 95% of it. The resulting limits are tabulated in Table S4. In no single pulsar is

Table S4: The single pulsar limits on A_{gwb} for 35 pulsars with TEMPONEST (TN in column 4), ENTERPRISE (ENT. in column 5) and the photon-by-photon method (columns 6 and 7). Pulsars with only photon-by-photon limits are highlighted. Data for PSRs J1959+2048 and J2241−5236 favor a model with white noise, while all others favor no additional noise. Most pulsars can be analyzed with a 2/yr (182 d) cadence, but the indicated six pulsars require longer integrations (243 d) to produce reliable TOAs. The single-pulsar A_{gwb} limits are all 95% confidence levels. The first set of photon-based limits include only the GWB in the noise model, while the limits in column 7 also include an intrinsic noise model (RN) for each pulsar, and the parameters of this noise model are marginalized.

Pulsar	Cadence TOA/yr	Noise model (favored)	TN $\times 10^{-14}$	ENT. $\times 10^{-14}$	Photon $\times 10^{-14}$	Photon+RN $\times 10^{-14}$
J0030+0451	2	None	7.54	7.77	7.61	8.44
J0034−0534	2	None	13.40	13.39	15.56	18.00
J0101−6422	2	None	18.63	18.94	18.80	22.16
J0102+4839	2	None	39.29	38.90	38.90	38.76
J0312−0921	−	−	−	−	21.57	27.86
J0340+4130	2	None	26.13	26.54	48.26	58.21
J0418+6635	−	−	−	−	32.59	36.41
J0533+6759	1.5	None	21.66	22.14	21.83	26.23
J0613−0200	2	None	22.57	21.59	21.66	25.80
J0614−3329	2	None	4.15	4.20	3.94	4.36
J0740+6620	1.5	None	15.76	16.62	17.24	20.55
J1124−3653	2	None	15.84	15.58	14.34	16.77
J1231−1411	2	None	2.19	2.30	2.71	3.54
J1513−2550	−	−	−	−	25.57	43.22
J1514−4946	2	None	38.89	38.54	41.85	34.90
J1536−4948	2	None	12.30	11.69	14.36	15.38
J1543−5149	−	−	−	−	98.02	1356.47
J1614−2230	2	None	9.08	9.23	8.14	9.76
J1625−0021	1.5	None	30.70	30.74	28.01	31.07
J1630+3734	2	None	7.28	7.27	7.91	9.08
J1741+1351	−	−	−	−	84.24	120.30
J1810+1744	2	None	14.31	14.95	18.54	21.35
J1816+4510	2	None	34.91	35.61	39.09	41.31
J1858−2216	2	None	31.51	30.46	110.83	1417.63
J1902−5105	2	None	11.50	11.38	11.87	15.06
J1908+2105	−	−	−	−	41.80	47.97
J1939+2134	1.5	None	12.86	13.04	10.24	12.99
J1959+2048	2	WN	8.25	8.12	6.15	7.84
J2017+0603	2	None	16.63	16.60	17.59	20.25
J2034+3632	1.5	None	21.62	22.21	38.82	74.75
J2043+1711	2	None	13.38	13.80	13.96	15.07
J2214+3000	2	None	29.67	30.44	28.74	38.60
J2241−5236	2	WN	13.80	13.66	14.05	16.39
J2256−1024	1.5	None	12.98	13.23	13.84	13.31
J2302+4442	2	None	11.74	11.67	12.26	14.82

there a strong detection of a GWB-like timing noise process, and in many the posterior plateaus at $A_{\mathrm{gwb}} = 0$. Joint limits on a common mode are obtained by simply multiplying the posteriors and integrating the resulting distribution. (This is the "factorized likelihood" approach of (*S28*).) The models for pulsars J2043+1711 and J2256−1024 are somewhat improved with $A_{\mathrm{gwb}} > 0$, and when included in the common mode limit, they substantially increase it. We discuss these two pulsars in more detail below.

Finally, we have considered intrinsic timing noise. For each pulsar, we introduce A_{rn} and Γ_{rn}, defined analogously to Equation 2, but with Γ_{rn} allowed to take values other than $13/3$. To marginalize over these additional nuisance parameters, we perform the same calculation described above but constraining the Fourier coefficients β by the summed power spectral density $P_{\mathrm{gwb}}(f) + P_{\mathrm{rn}}(f)$. We scan over a grid of A_{rn} and $1 < \Gamma_{\mathrm{rn}} < 7$ and tabulate the resulting marginalized log likelihood again as a function of A_{gwb}, again using a uniform prior on both A_{gwb} and A_{rn}. The log likelihoods in this case are shown in Figure S3. The apparent peaks in Figure S2 have largely vanished, indicating that they are likely caused either by statistical fluctuations (they are low significance) or by timing noise processes with $\Gamma_{\mathrm{rn}} \neq 13/3$. The resulting constraints on A_{gwb} (Table S4) are generally poorer, except in cases where the log likelihood with an intrinsic RN model peaked at $A_{\mathrm{gwb}} > 0$. The most conservative limit is obtained using this approach.

Comparison of methods   The two codes ENTERPRISE and TEMPONEST provide single pulsar limits in almost exact agreement, so the main comparison to be made is between the TOA-based and photon-by-photon approaches. Aside from computational aspects—the TOA-based methods are sampled, while the photon-by-photon method is analytic—there are two primary differences: the photon-based approach avoids the assumption of Gaussianity on TOA uncertainties, and it retains sensitivity to all timescales. We generally expect the photon-based ap-
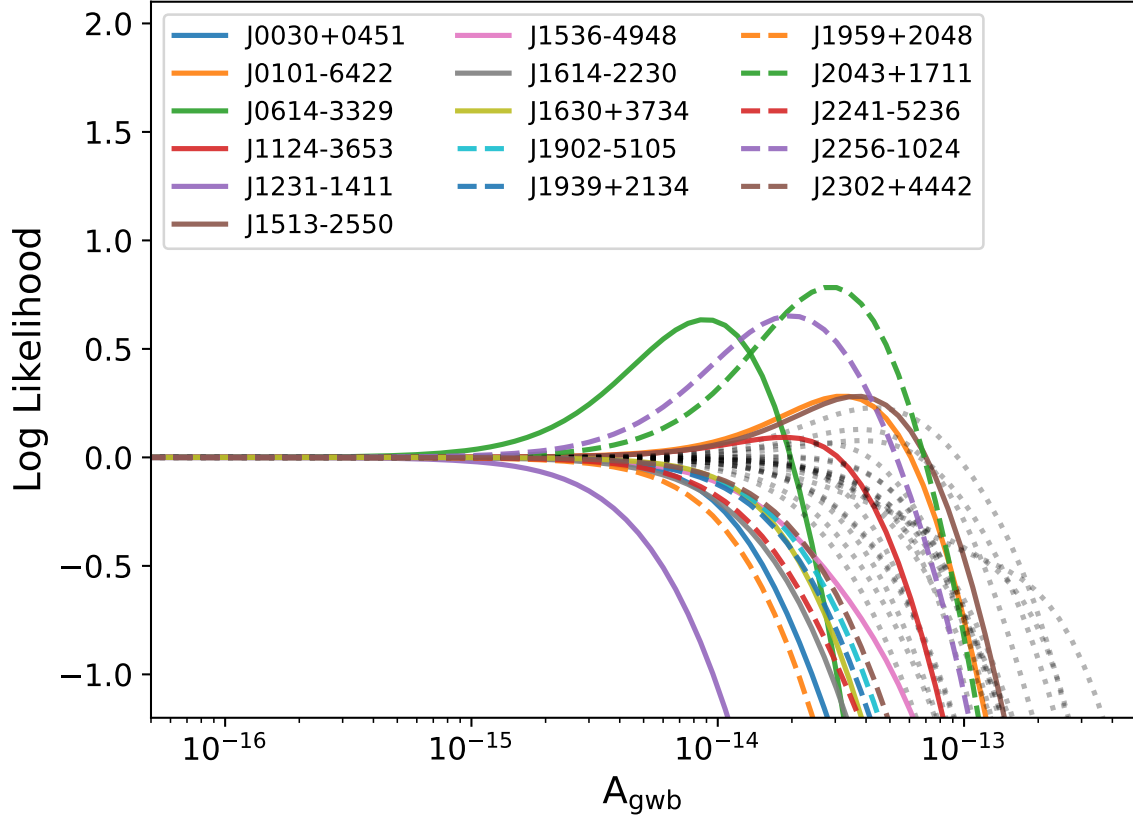
19

Figure S2: The log likelihood produced using the photon-by-photon method, marginalized analytically over the timing model parameters and Fourier coefficients of the noise process. The highlighted pulsars have an individual limit $A_{\mathrm{gwb}} < 1.5 \times 10^{-13}$ or a peak value at $A_{\mathrm{gwb}} > 0$ along with a limit $A_{\mathrm{gwb}} < 3.0 \times 10^{-13}$. These are the pulsars with greatest influence. The others are shown in gray.
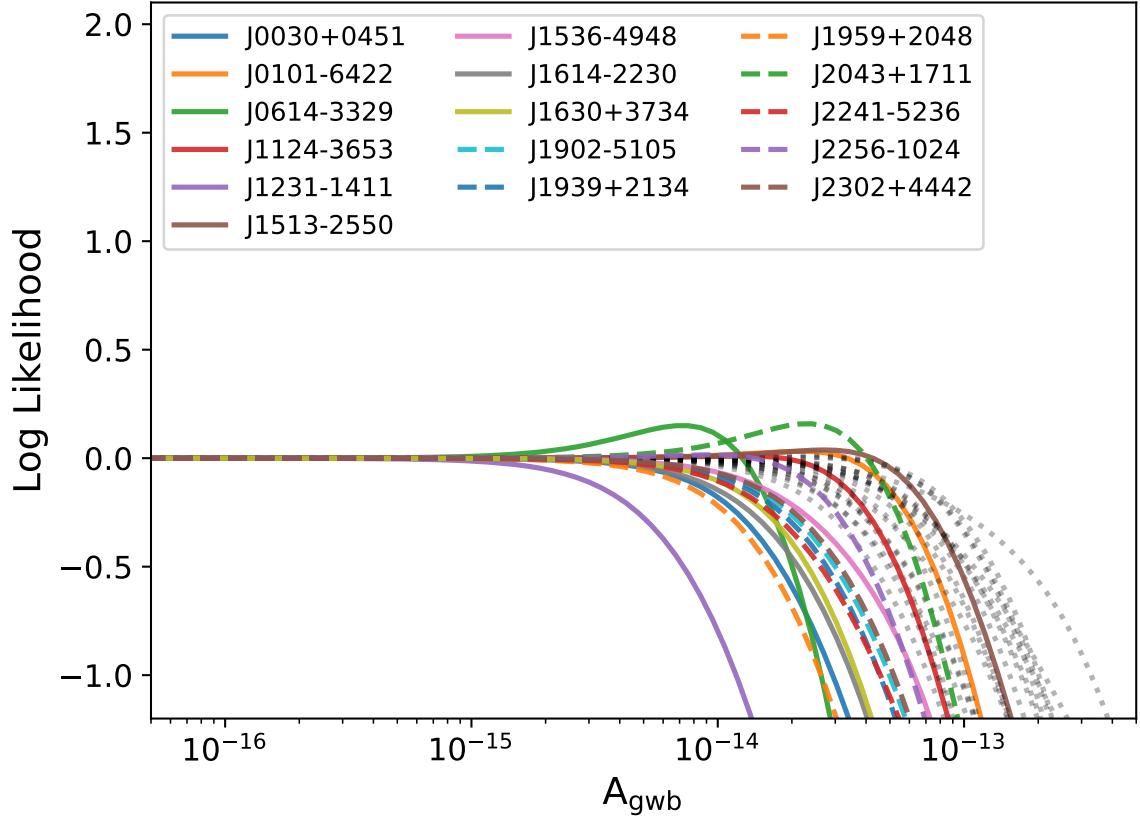
Figure S3: As Figure S2, with an additional, intrinsic noise process which is marginalized numerically.

proach to be more precise due to these advantages. The agreement between these two methods is largely excellent, giving confidence in the newly developed photon-by-photon capability. We discuss the exceptions below and argue that they stem from the fundamental differences in the methods. None of these discrepancies has a large impact on the common mode limits discussed below, since the total number is small compared to the full sample and because none of these pulsars contribute particularly strongly to the timing array.

PSR J0340+4130 is the most striking outlier, with a photon-by-photon limit nearly twice that of the TOA-based methods. Its timing solution is well constrained by radio observations, so there are no additional degrees of freedom in the photon-based fit. However, the photon-by-photon log likelihood peaks at $A_{\mathrm{gwb}} > 0$, causing the greater limit. We speculate that in this case, the Gaussianity assumption may play a role, distorting a TOA in such a way that the TOA-based limit is reduced.

For PSR J1536−4948, the photon-based method also delivers a higher limit. This is almost certainly due to the additional degrees of freedom that are fit with the photon-by-photon method, including both binary parameters and astrometry.

PSR J1810+1744 is a compact binary with extensive orbital period variations, and these degrees of freedom are likewise inaccessible to the TOA-based method. When we computed the photon-by-photon limit with the binary parameters held fixed, it dropped to 73% of its previous value, in reasonable agreement with the TOA-based results.

PSRs J1858−2215 and J2034+3632 are both faint pulsars and use a restricted number of harmonics for numerical stability (see above), and this degrades the sensitivity of the photon-based analysis. Additional LAT data, or improved numerical techniques, can ameliorate this problem.

Table S5: 95% confidence upper limits on $A_{\mathrm{gwb}}/10^{-14}$ from the combined samples. The pulsars corresponding to the subsets are given in the text and are ranked by their single-pulsar GWB limits. The "Full" rows indicate the total sample for the two methods, 29 pulsars common to TOA-based and photon-by-photon, and 35 to photon-by-photon only. Only the last row includes the "Hellings and Downs" (HD) spatial correlations predicted for quadrupolar gravitational radiation (*S65*).

Subset	ENTERPRISE	ENTERPRISE with RN	Photon	Photon with RN
Best 2	1.89	1.94	1.84	2.02
Best 3	1.71	1.74	1.50	1.66
Best 9	1.15	1.19	1.06	1.16
Full 29	1.12	1.04	1.14	1.06
Full 35	–	–	1.14	1.05
Full 29 w/HD	1.06	–	–	–

Combined limits and scaling   We can now combine the pulsars into a PTA and obtain limits. To assess the dependence of the overall limit on particular pulsars, we present limits for both the full array and some subsets. The best nine pulsars in ascending order of single pulsar limits are J1231−1411, J0614−3329, J1959+2048, J0030+0451, J1630+3734, J1614−2230, J1939+2134, J1902−5105 and J2302+4442. These pulsars form the subsets defined in Table S5. For both methods, the limits steadily improve with additional pulsars, and in the case of the TOA-based approach, with the inclusion of the spatial correlation information predicted for a GWB (*S65*). In both cases, the best limits are $A_{\mathrm{gwb}} < 1.0 \times 10^{-14}$.

As noted above, the limit in the photon-by-photon case can be degraded by including pulsars whose A_{gwb} posteriors peak at $A_{\mathrm{gwb}} > 0$, which can happen even in the absence of a GWB signal due to statistical fluctuations. For completeness, however, we consider limits computed when removing those with the strongest signal, PSR J2043+1711, followed by PSR J2256−1024, which exhibit modest orbital period variations. This is motivated if the log likelihood peak is in fact due to some deficiency in the timing solution. When removing these two pulsars, we obtain limits of 9.8×10^{-15} and 10.3×10^{-15}.

Table S6: 95% confidence upper limits on $A_{\mathrm{gwb}}/10^{-14}$ from the combined samples. Here, we compare results obtained with ENTERPRISE, in both cases using spatial correlations (*S65*) but with and without BAYESEPHEM.

Subset	ENTERPRISE	ENTERPRISE with BAYESEPHEM
Best 2	1.90	1.81
Best 3	1.70	1.64
Best 9	1.17	1.16
Full 29	1.06	1.08

We have additionally considered any systematic effects from our choice of the DE421 solar system ephemeris, by performing the same limit calculation using a perturbative Bayesian modeling software, BAYESEPHEM (*S67*). This allows us to model uncertainties in the solar system planetary masses and orbital parameters while simultaneously constraining other pulsar noise parameters and the GWB. We compare two sets of limits obtained with ENTERPRISE, with and without BAYESEPHEM, reported in Table S6. The agreement is excellent. The limits with BAYESEPHEM are slightly tighter for the few-pulsar subsets, which follows from the increased degeneracy between the GWB quadrupolar signature and dipolar-like effects from solar system ephemeris errors (*S68*). The full array results, on the other hand, contain enough information to separate the two signals. The agreement between our limits fixed at DE421 and those obtained with BAYESEPHEM indicates there are no issues with its use at the current *Fermi* PTA sensitivity.

In summary, we have obtained a robust limit of $A_{\mathrm{gwb}} < 10^{-14}$, with two independent methods yielding both expected scalings and values in good agreement. The limit is robust to the solar system ephemeris (DE421) we have chosen, and to particular realizations of noise modeling, e.g. the inclusion of per-pulsar spin noise.

Table S7: Central values and references for published upper limits (u.l.) and detections of the GWB.

Label	Reference	A_{gwb} $\times 10^{-15}$	Range $\times 10^{-15}$	Note
PPTA 2006	(*S69*)	11	–	95% u.l.
PPTA 2013	(*S70*)	2.4	–	95% u.l.
PPTA 2015	(*S71*)	1.0	–	95% u.l.
PPTA 2021	(*S60*)	2.2	1.9–2.6	68% range
EPTA 2011	(*S39*)	6.0	–	95% u.l.
EPTA 2015	(*S72*)	3.0	–	95% u.l.
EPTA DR2	Chen et al., (submitted)	–	2.3–3.7	5–95% confidence
NANOGrav 5-yr	(*S73*)	7.0	–	95% u.l.
NANOGrav 11-yr	(*S62*)	1.45	–	95% u.l.
NANOGrav 12.5-yr	(*S28*)	1.92	1.4–2.7	5–95% confidence
IPTA DR1	(*S9*)	1.7	–	95% u.l.
IPTA DR2	priv. comm.	–	2.0–3.6	5–95% confidence

Other GWB models We have calculated 95% upper limits on more general power-law GWB models, shown in Figure 3. For each value of α, we computed a photon-based limit as described above, without including any per-pulsar spin noise. The resulting scaling is approximately exponential, and the scaling largely reflects the fact that the limit is dominated by frequencies around $2/t_{\mathrm{obs}}$. Limits are slightly less constraining for models with more high-frequency power due to the limited number of Fourier coefficients used in the analysis.

Spin noise There is little evidence for intrinsic spin noise in any of the pulsars in our sample (Table S4 and Figure S2). PTAs have measured such intrinsic timing noise for some of them, but the amplitudes are generally below the single-pulsar sensitivity. The most interesting case to consider is PSR J1939+2134, in which intrinsic timing noise is obvious over the decades-long data sets (*S9*). We detect no spin noise and obtain limits somewhat below the values reported by radio PTAs (see below).

Even though we see no strong evidence for spin noise in any pulsar, two pulsars, J2043+1711

Table S8: Comparison between measured radio PTA spin noise amplitudes and 95% *Fermi* PTA upper limits. The spin noise power spectra are given in $\mathrm{s^2\,yr^{-1}}$ evaluated at $f = 1/\mathrm{yr}$. *Fermi* limits are computed similarly to GWB amplitude limits save using the value of Γ measured in each row rather than $\Gamma = 13/3$. The ratio of PTA power to *Fermi* power is given, as is the ratio if the limit were to scale to twice the length of the data set ($2\times$).

Pulsar	PTA	Γ	P(f) PTA	P(f) *Fermi*	Ratio	Ratio $2\times$
J0030+0451	NANOGrav	6.3	9.0×10^{-18}	8.3×10^{-18}	1.1	86
J0613$-$0200	NANOGrav	2.1	1.5×10^{-14}	2.0×10^{-13}	0.1	0.3
J0613$-$0200	PPTA	4.2	2.5×10^{-16}	5.4×10^{-15}	0.1	0.8
J1939+2134	NANOGrav	3.3	9.8×10^{-15}	6.6×10^{-15}	1.5	15
J1939+2134	PPTA	5.4	1.8×10^{-16}	1.1×10^{-16}	1.7	72

and J2256$-$1024, have posteriors which peak at $A_{\mathrm{gwb}} > 0$, but at sufficiently low values of A_{gwb} that the combined photon-by-photon limit increases. The posterior peaks decrease when we include a spin noise model and marginalize over its parameters, indicating the modest noise is likely more consistent with a different spectral index. In general, the single-pulsar A_{gwb} limits increase when including these extra degrees of freedom (Table S4), but decrease for those pulsars with modest RN. The total limit on A_{gwb} is slightly lower when including spin noise, but both values are within 10% of each other.

Comparison to radio measurements Radio PTAs have published spin noise models for many MSPs, particularly two recent sets based on the NANOGrav 12.5-yr data set (*S26*) and the PPTA data release 2 (*S12,S29*) containing measurements for 14 and 10 MSPs, respectively. Both contain treatments for the time-varying DM (see above; NANOGrav use `DMX` and the PPTA a stationary process model), so these measurements are in principle reflective of the true spin noise. Of these pulsars, only 3 are in common with our γ-ray MSP sample, PSRs J0030+0451, J0613$-$0200, and J1939+2134. For these pulsars, the radio estimates for spin noise do differ, with the PPTA analysis finding somewhat steeper spectra for the two common pulsars in our sample.

We have computed 95% upper limits on a spin-noise process for each pulsar using the radio-PTA measured spectral index Γ, with results listed in Table S8. Aside from the changed value of Γ, the procedure is identical to the photon-by-photon method we have used to compute single-pulsar GWB limits. The *Fermi* upper limits for J0030+0451 and J1939+2134 are lower than the measured radio values and modestly incompatible. This could be evidence for uncorrected IISM/solar wind effects leaking into the spin noise estimation. As noted above, radio emission from PSR J0030+0451 is particularly affected by the solar wind because its position passes relatively near to the sun each year.

If this is a true difference, then it will become stark with accumulation of additional γ-ray data. Using the spectral index to scale to a data set twice as long as the 12.5 yr set used for the limits, the 95% limit would be at least $10\times$ smaller than the radio-PTA measurements (in terms of the power spectral density, P(f)), providing a robust measurement of the "true" spin noise model and a touchstone for testing radio data analysis techniques.

There are an additional six MSPs from our parent sample (114 MSPs) with measured spin noise parameters. While we do not have single-pulsar GWB upper limits for these, we can compare the measured spin noise against the γ-ray white noise level and estimate if and when the spin noise could be detectable. We find that only PSR J1747$-$4036, with a NANOGrav-measured $\Gamma = 3.3$, is promising. Its spin noise could be detectable now, and will almost certainly be detectable with high significance in a doubled data set.

Although the number of MSPs here is too small to draw firm conclusions, there is some evidence for contamination of spin noise models (and, ultimately, GWB-like signals) by uncorrected IISM effects. The most sensitive γ-ray MSPs, J1231$-$1411, J0614$-$3329, and others are not among the pulsars routinely monitored by radio PTAs. Dedicated radio timing campaigns capable of detecting their intrinsic spin noise would be invaluable for better understanding its measurement and the effects of the IISM.

Sensitivity improvements for the *Fermi* PTA

Discoveries We have estimated the number of potentially-useful γ-ray MSPs that remain to be discovered. Using a preliminary version of the third release of the fourth *Fermi* catalog (4FGL-DR3) (*S55, S56*), with its spectral fits and source association information, we have tabulated and plotted (Figure S4) the source significance and the peak spectral energy ("Epeak") of all known MSPs; all known young pulsars; and unassociated sources, those without a known counterpart at some other wavelength. The emitted power of γ-ray pulsars peaks at a few GeV, evident as the cloud of pulsars reaching to high significance with this value. We can thus estimate the total number of undiscovered pulsars simply by counting the unassociated sources within this cluster. We apply the simple prescription of a significance $>25\sigma$ and $0.178 < \mathrm{Epeak(GeV)} < 5.620$. The resulting 23 sources may be binary MSPs, since extensive blind searches of the data are likely to have detected most young pulsars (*S74*) and many isolated MSPs (*S75*). Young pulsars with very strong timing noise or frequent glitches may have been missed, but such systems, with high spin-down luminosities, are rare and tend to be very close to the Galactic plane, near their parent star-forming regions. The median absolute Galactic latitude of the selected unassociated sources is $3.8°$.

Assuming then that most of these sources represent potential MSPs, we can estimate how many of them could be good timers. To do this, we have classified the pulsars in our sample as "excellent", "good", or "fair" timers according to their single-pulsar limit on the GWB, quantitatively $A_{\mathrm{gwb}} < 10^{-13}$, $10^{-13} < A_{\mathrm{gwb}} < 2 \times 10^{-13}$, and $2 \times 10^{-13} < A_{\mathrm{gwb}} < 5 \times 10^{-13}$. In the 25–100σ range, there are 3, 8, and 8 members of each class, while there are 52 known MSPs. There are 23 unassociated sources and 2 associated MSPs which are likely to become LAT MSPs with refined timing solutions. Thus we estimate there are 1.44 excellent, 3.85 good, and 3.85 fair timers to discover.
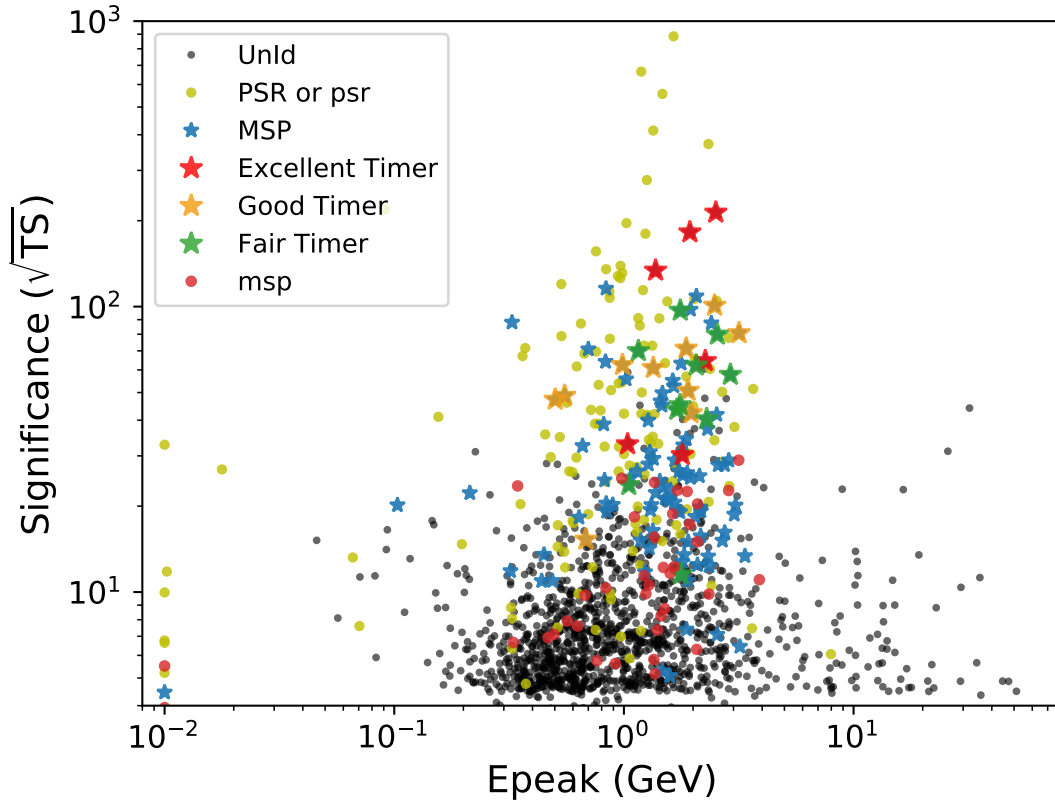
Figure S4: The population of known and potential pulsars along with *Fermi*-LAT sources without known counterparts. The significance of each source (*S55*), as realized by the "Test Statistic" (TS) is shown as a function of the peak energy of its spectral energy distribution. The spectra of pulsars peak at a few GeV, a trend clearly illustrated. Pulsars fixed at 0.01 GeV have unreliable peak energy measurements.

To estimate how these could improve sensitivity, we apply a simple bootstrapping procedure using the posterior distributions from the photon-based analysis. For each iteration, we simulate three Poisson variables from distributions with means of 1.44, 3.85, and 3.85 to represent N new pulsars from each class. We then randomly choose N members (with replacement) from matching class and duplicate the posteriors in the upper limit calculation. To avoid undue influence from any one pulsar, we use the likelihood marginalized over an intrinsic timing noise process. We do this 10,000 times and evaluate the limit on A_{gwb} with the new, bootstrap sample. Doing this, we find that the limit is reduced to 81–96% of its current value (central 68% percentile range).

Technical developments We had to exclude pulsars like PSRs J2129$-$0429 and J2215+5135, compact binary systems whose orbital period variations are too complex to represent with current software. More sophisticated methods, like Gaussian process regression (*S58*), would enable the inclusion of these otherwise-excellent timers. Such methods may also sharpen GWB sensitivity by reducing the degeneracy between period variations and the GWB signal. Likewise, improved radio constraints on the timing model parameters we must currently fit with LAT data can also yield a typical improvement of 5–10% (Table S2 and discussion). New capabilities with telescopes like CHIME (*S23*) and MeerKAT (*S76*) promise substantial improvement in radio timing capabilities.

The performance for all pulsars can be modestly improved by using energy-dependent templates, analogous to the "portrait" methods used in radio pulsar timing (*S32*), which account for the evolution of the pulse profile with radio frequency. Such evolution is more modest in γ rays (*S77*) but is nevertheless notable over the very wide 0.1–30 GeV band, and many γ-ray pulse profiles have two peaks whose relative ratios change with energy. We have estimated the impact of accounting for this evolution by dividing the data for J1231$-$1411 and J0614$-$3329

30

into two energy bands with approximately equal detection significance, at 1.2 GeV and 1.5 GeV respectively. When using a template re-optimized for the specific bands, we find the overall timing precision improves by 1.5% and 4.2%, respectively. We expect similar improvements of a few percent from tuning the probability weight cut for each pulsar, and/or refining zenith angle cuts to maximize signal.

Combined improvements Considering the $\sim$5–20% improvement from new pulsars and a 5–10% improvement from technical development, we estimate a total 10–25% sensitivity enhancement over the next ten years. We have indicated this 10–25% range, and the corresponding 2.1 and 2.5×10^{-15} projections, via conservative and optimistic tracks in Figure 1. Doubling the data could produce a limit as low as 1.7×10^{-15}, a true test of the origin of the GWB-like signal.

References

S1. D. R. Lorimer, M. Kramer, *Handbook of Pulsar Astronomy* (Cambridge, UK, 2004).

S2. J. H. Taylor, L. A. Fowler, P. M. McCulloch, *Nature* **277**, 437 (1979).

S3. M. Kramer, *et al.*, *Science* **314**, 97 (2006).

S4. M. Kramer, N. Wex, *Classical and Quantum Gravity* **26**, 073001 (2009).

S5. C. M. Will, *Living Reviews in Relativity* **17** (2014).

S6. D. C. Backer, S. R. Kulkarni, C. Heiles, M. M. Davis, W. M. Goss, *Nature* **300**, 615 (1982).

S7. M. V. Sazhin, *Soviet Ast.* **22**, 36 (1978).

532   S8. S. Detweiler, *ApJ* **234**, 1100 (1979).

533   S9. J. P. W. Verbiest, *et al.*, *MNRAS* **458**, 1267 (2016).

534   S10. R. M. Shannon, *et al.*, *MNRAS* **443**, 1463 (2014).

535   S11. A. Parthasarathy, *et al.*, *MNRAS* **502**, 407 (2021).

536   S12. M. Kerr, *et al.*, *PASA* **37**, e020 (2020).

537   S13. S. Dai, *et al.*, *MNRAS* **449**, 3223 (2015).

538   S14. W. van Straten, R. N. Manchester, S. Johnston, J. E. Reynolds, *PASA* **27**, 104 (2010).

539   S15. W. van Straten, *ApJS* **204**, 13 (2013).

540   S16. J. M. Cordes, J. M. Weisberg, V. Boriakoff, *ApJ* **288**, 221 (1985).

541   S17. W. A. Coles, *et al.*, *ApJ* **808**, 113 (2015).

542   S18. W. A. Coles, B. J. Rickett, J. J. Gao, G. Hobbs, J. P. W. Verbiest, *ApJ* **717**, 1206 (2010).

543   S19. R. M. Shannon, J. M. Cordes, *MNRAS* **464**, 2075 (2017).

544   S20. D. Stinebring, *Classical and Quantum Gravity* **30**, 224006 (2013).

545   S21. M. J. Keith, *et al.*, *MNRAS* **429**, 2161 (2013).

546   S22. G. Hobbs, *et al.*, *PASA* **37**, e012 (2020).

547   S23. CHIME/Pulsar Collaboration, *et al.*, *arXiv e-prints* p. arXiv:2008.05681 (2020).

548   S24. J. Y. Donner, *et al.*, *A&A* **624**, A22 (2019).

549   S25. J. M. Cordes, R. M. Shannon, D. R. Stinebring, *ApJ* **817**, 16 (2016).

550 S26. M. F. Alam, *et al.*, *ApJS* **252**, 4 (2021).

551 S27. M. T. Lam, *et al.*, *ApJ* **861**, 132 (2018).

552 S28. Z. Arzoumanian, *et al.*, *ApJL* **905**, L34 (2020).

553 S29. B. Goncharov, *et al.*, *MNRAS* **502**, 478 (2021).

554 S30. R. M. Shannon, *et al.*, *ApJL* **828**, L1 (2016).

555 S31. P. B. Demorest, *MNRAS* **416**, 2821 (2011).

556 S32. T. T. Pennucci, P. B. Demorest, S. M. Ransom, *ApJ* **790**, 93 (2014).

557 S33. C. Tiburzi, *et al.*, *MNRAS* **487**, 394 (2019).

558 S34. C. Tiburzi, *et al.*, *A&A* **647**, A84 (2021).

559 S35. F. X. Lin, *et al.*, *arXiv e-prints* p. arXiv:2106.09851 (2021).

560 S36. A. Allafort, *et al.*, *ApJL* **777**, L2 (2013).

561 S37. A. Parthasarathy, *et al.*, *MNRAS* **489**, 3810 (2019).

562 S38. R. M. Shannon, J. M. Cordes, *ApJ* **725**, 1607 (2010).

563 S39. R. van Haasteren, *et al.*, *MNRAS* **414**, 3117 (2011).

564 S40. W. Coles, G. Hobbs, D. J. Champion, R. N. Manchester, J. P. W. Verbiest, *MNRAS* **418**,
565   561 (2011).

566 S41. L. Lentati, *et al.*, *PhRvD* **87**, 104021 (2013).

567 S42. A. Melatos, B. Link, *MNRAS* **437**, 21 (2014).

S43. M. Kramer, A. G. Lyne, J. T. O'Brien, C. A. Jordan, D. R. Lorimer, *Science* **312**, 549 (2006).

S44. A. Lyne, G. Hobbs, M. Kramer, I. Stairs, B. Stappers, *Science* **329**, 408 (2010).

S45. W. Hermsen, *et al.*, *Science* **339**, 436 (2013).

S46. P. Bickel, B. Kleijn, J. Rice, *ApJ* **685**, 384 (2008).

S47. M. Kerr, *ApJ* **732**, 38 (2011).

S48. P. Bruel, *A&A* **622**, A108 (2019).

S49. P. S. Ray, *et al.*, *ApJS* **194**, 17 (2011).

S50. M. Kerr, P. S. Ray, S. Johnston, R. M. Shannon, F. Camilo, *ApJ* **814**, 128 (2015).

S51. R. Main, *et al.*, *Nature* **557**, 522 (2018).

S52. C. Kalapotharakos, A. K. Harding, D. Kazanas, Z. Wadiasingh, *ApJL* **883**, L4 (2019).

S53. W. B. Atwood, *et al.*, *ApJ* **697**, 1071 (2009).

S54. The Fermi LAT Collaboration, *et al.*, *arXiv e-prints* p. arXiv:2106.12203 (2021).

S55. S. Abdollahi, *et al.*, *ApJS* **247**, 33 (2020).

S56. J. Ballet, T. H. Burnett, S. W. Digel, B. Lott, *arXiv e-prints* p. arXiv:2005.11208 (2020).

S57. J. Luo, *et al.*, *ApJ* **911**, 45 (2021).

S58. C. J. Clark, *et al.*, *MNRAS* **502**, 915 (2021).

S59. W. Folkner, J. Williams, D. Boggs, *Interplanetary Network Progress Report* **42–178**, 1 (2009).

S60. B. Goncharov, *et al.*, *arXiv e-prints* p. arXiv:2107.12112 (2021).

S61. R. N. Caballero, *et al.*, *MNRAS* **481**, 5501 (2018).

S62. Z. Arzoumanian, *et al.*, *ApJ* **859**, 47 (2018).

S63. L. Lentati, *et al.*, *MNRAS* **437**, 3004 (2014).

S64. J. A. Ellis, M. Vallisneri, S. R. Taylor, P. T. Baker, Enterprise: Enhanced numerical tool-box enabling a robust pulsar inference suite, Zenodo (2020).

S65. R. W. Hellings, G. S. Downs, *ApJL* **265**, L39 (1983).

S66. M. Kerr, *ApJ* **885**, 92 (2019).

S67. M. Vallisneri, *et al.*, *ApJ* **893**, 112 (2020).

S68. C. Tiburzi, *et al.*, *MNRAS* **455**, 4339 (2016).

S69. F. A. Jenet, *et al.*, *ApJ* **653**, 1571 (2006).

S70. R. M. Shannon, *et al.*, *Science* **342**, 334 (2013).

S71. R. M. Shannon, *et al.*, *Science* **349**, 1522 (2015).

S72. L. Lentati, *et al.*, *MNRAS* **453**, 2576 (2015).

S73. P. B. Demorest, *et al.*, *ApJ* **762**, 94 (2013).

S74. C. J. Clark, *et al.*, *ApJ* **834**, 106 (2017).

S75. C. J. Clark, *et al.*, *Science Advances* **4**, eaao7228 (2018).

S76. M. Bailes, *et al.*, *PASA* **37**, e028 (2020).

S77. A. A. Abdo, *et al.*, *ApJS* **208**, 17 (2013).