Fast and accurate sensitivity estimation for continuous-gravitational-wave searches

Christoph Dreissigacker,[1,2,*] Reinhard Prix,[1,2] and Karl Wette[3,1,2]

[1]*Max-Planck-Institut für Gravitationsphysik (Albert-Einstein-Institut), D-30167 Hannover, Germany*
[2]*Institut für Gravitationsphysik, Leibniz Universität Hannover, D-30167 Hannover, Germany*
[3]*ARC Centre of Excellence for Gravitational Wave Discovery (OzGrav) and Centre for Gravitational Physics, Research School of Physics and Engineering, The Australian National University, Acton ACT 2601, Australia*

This paper presents an efficient numerical sensitivity-estimation method and implementation for continuous-gravitational-wave searches, extending and generalizing an earlier analytic approach by Wette [1]. This estimation framework applies to a broad class of $\mathcal{F}$-statistic-based search methods, namely (i) semi-coherent StackSlide $\mathcal{F}$-statistic (single-stage and hierarchical multistage), (ii) Hough number count on $\mathcal{F}$-statistics, as well as (iii) Bayesian upper limits on $\mathcal{F}$-statistic search results (coherent or semi-coherent). We test this estimate against results from Monte-Carlo simulations assuming Gaussian noise. We find the agreement to be within a few % at high detection (i.e., low false-alarm) thresholds, with increasing deviations at decreasing detection (i.e., higher false-alarm) thresholds, which can be understood in terms of the approximations used in the estimate. We also provide an extensive summary of sensitivity depths achieved in past continuous-gravitational-wave searches (derived from the published upper limits). For the $\mathcal{F}$-statistic-based searches where our sensitivity estimate is applicable, we find an average relative deviation to the published upper limits of less than 10%, which in most cases includes systematic uncertainty about the noise-floor estimate used in the published upper limits.

I. INTRODUCTION

The recent detections of gravitational waves from merging binary-black-hole and double neutron-star systems [2–4] have opened a whole new observational window for astronomy, allowing for new tests of general relativity [5], new constraints on neutron star physics [6] and new measurements of the Hubble constant [7], to mention just a few highlights.

Continuous gravitational waves (CWs) from spinning nonaxisymmetric neutron stars represent a different class of potentially observable signals [8,9], which have yet to be detected [10]. These signals are expected to be long-lasting (at least several days to years) and quasi monochromatic, with slowly changing (intrinsic) frequency. The signal amplitude depends on the rich (and largely not yet well-understood) internal physics of neutron stars [11], as well as their population characteristics [12,13]. A detection

(and even nondetection) of CWs could therefore help us better understand these fascinating astrophysical objects, and may allow for new tests of general relativity [14,15].

A. Overview of search categories

We can categorize CW searches in two different ways: either based on the *search method*, or on the type of *explored parameter space*.

The *search methods* fall into two broad categories: *coherent* and *semicoherent* (sometimes also referred to as incoherent). Roughly speaking, a coherent search is based on signal templates with coherent phase evolution over the whole observation time, while semicoherent searches typically break the data into shorter coherent segments and combine the resulting statistics from these segments incoherently (i.e., without requiring a consistent phase evolution across segments). However, there are many different approaches and variations, which are beyond the scope of this paper, see, e.g., [10] for a more detailed overview. Here we will exclusively focus on coherent and semicoherent methods based on the $\mathcal{F}$-statistic, which will be introduced in Sec. II.

Coherent search methods are the more sensitive in principle, but in practice they suffer from severe computing-cost limitations: for finite search parameter spaces the

---

*christoph.dreissigacker@aei.mpg.de

required number of signal templates grows as a steep power-law of the observation time, making such searches infeasible except when the search region is sufficiently small. For larger signal parameter spaces the observation time needs to be kept short enough for the search to be computationally feasible, which limits the attainable coherent sensitivity. This is where semi-coherent searches yield substantially better sensitivity at fixed computing cost (e.g., see [16,17]).

Based on the *explored parameter space*, we distinguish the following search categories (referencing a recent example for each case):

(i) *Targeted searches* are looking for CW emission from known pulsars [18]. The pulsar spin evolution, the sky position and possibly the binary orbital parameters of these systems are known very accurately. Assuming a fixed relationship between pulsar spin and CW frequency, these searches therefore only need to target a single point in parameter space for each pulsar. Hence these searches are done with an optimal fully coherent search [19].

(ii) *Narrow-band searches* for known pulsars assume a small uncertainty in the relationship between CW frequency and the measured pulsar spin rates. This finite search parameter space requires a template bank with (typically) many millions of templates, still allowing for optimal fully coherent search methods to be used [20].

(iii) *Directed (isolated) searches* aim at isolated neutron stars with known sky-position and unknown spin frequency. The search parameter space covers the unknown frequency and spindowns of the neutron star signal within an astrophysically motivated range [21,22].

(iv) *(Directed) binary searches* aim at binary systems with known sky-position and parameter-space uncertainties in the frequency and binary-orbital parameters. Typically these sources would be in low-mass x-ray binaries, with the most prominent example being Scorpius X-1 (Sco X-1) [23,24].

(v) *All-sky (isolated) searches* search the whole sky over a large frequency (and spindown) band for unknown isolated neutron stars [25,26].

(vi) *All-sky binary searches* are the most extreme case, covering the whole sky for unknown neutron stars in binary systems [27,28].

B. Sensitivity estimation

In this work we use the term *sensitivity* to refer to the upper limit on *signal amplitude* h_0 (or equivalently *sensitivity depth* $\mathcal{D} \equiv \sqrt{S_n}/h_0$, see Sec. II E). This can be either the frequentist upper limit for a given detection probability at a fixed false-alarm level (p-value), or the Bayesian upper limit at a given credible level for the given data.

Sensitivity therefore only captures one aspect of a search, namely how "deep" into the noise-floor it can detect signals, without accounting for how "wide" a region in parameter space is covered, how much prior weight this region contains, or how robust the search is to deviations from the signal model. Comparing sensitivity depth therefore only makes sense for searches over very similar parameter spaces. A more complete measure characterizing searches would be their respective detection probability, which folds in sensitivity depth, breadth in parameter space, as well as the prior weight contained in that space [29,30].

However, it is often useful to be able to reliably and cheaply estimate the sensitivity of a search setup without needing expensive Monte-Carlo simulations:

(i) In order to determine optimal search parameters for a semicoherent search (i.e., the number and semicoherent segments and template-bank mismatch parameters), it is important to be able to quickly assess the projected sensitivity for any given search-parameter combination (e.g., see [17,29–31]).

(ii) For setting upper limits for a given search, one typically has to repeatedly add software-generated CW signals to the data and perform a search, in order to measure how often these signals are recovered above a given threshold. By iterating this procedure one tries to find the weakest signal amplitude that can be recovered at the desired detection probability (or "confidence"). This can be very computationally expensive, and a quick and reasonably-reliable estimate for the expected upper-limit amplitude can therefore substantially cut down on the cost of this iterative process, which can also improve the accuracy of the upper limit.

(iii) The estimate can also serve as a sanity check for determining upper limits.[1]

A number of theoretical sensitivity estimates have been developed over the past decades. One of the first estimates was obtained for a coherent $\mathcal{F}$-statistic search [32], yielding

$$h_0 = 11.4 \sqrt{\frac{S_n}{T_{\text{data}}}}, \qquad (1)$$

for a 90% confidence upper limit at 1% false-alarm (per template). S_n denotes the (single-sided) noise power spectral density, and T_{data} is the total duration of data from all detectors, e.g., for a search spanning one day (i.e., $T_{\text{span}} = 24$ h), one detector could have yielded 18 h of (possibly non-contiguous) usable data and another detector 16 h, giving a total of $T_{\text{data}} = 34$ h.

---

[1]In fact, in the course of this work we have identified a bug in the upper-limit script of a published result, while trying to understand the discrepancy between the estimate and the published value, see Sec. VI C.

This first estimate was later generalized to the semi-coherent Hough [33] and StackSlide method [34,35], yielding an expression of the form

$$h_0 = \kappa N_{\text{seg}}^{1/4} \sqrt{\frac{S_{\text{n}}}{T_{\text{data}}}}, \quad \text{with} \quad \kappa \sim 7\text{–}9, \qquad (2)$$

for the same confidence and false-alarm level as Eq. (1), and where N_{seg} denotes the number of semicoherent segments.

These latter results suggested the inaccurate idea that the sensitivity of semi-coherent searches follows an exact $N_{\text{seg}}^{1/4}$ scaling. However, this was later shown [1,17] to not be generally a good approximation except asymptotically in the limit of a large number of segments ($N_{\text{seg}} \gtrsim 100\text{–}1000$).

Furthermore, these past sensitivity estimates relied on the assumption of a "constant signal-to-noise ratio (SNR)" population of signals. While this approximation substantially simplifies the problem, it introduces a noticeable bias into the estimate, as discussed in more detail in [1]. For example, the constant-SNR bias combined with the incorrect $N_{\text{seg}}^{1/4}$ scaling in Eq. (2) would result in an overestimate by a *factor of two* of the sensitivity of the first Einstein@Home search on LIGO S5 data [36].

These limitations of previous sensitivity estimates were eventually overcome by the analytic sensitivity-estimation method developed by Wette [1] for semicoherent StackSlide $\mathcal{F}$-statistic searches. In this work we simplify and extend this framework by employing a simpler direct numerical implementation. This further improves the estimation accuracy by requiring fewer approximations. It also allows us to generalize the framework to multistage hierarchical StackSlide-$\mathcal{F}$ searches, Hough-$\mathcal{F}$ searches (such as [36]), as well as to Bayesian upper limits based on $\mathcal{F}$-statistic searches.

C. Plan of this paper

Section II provides a description of the CW signal model and introduces different $\mathcal{F}$-statistic-based search methods. In Sec. III we present the sensitivity-estimation framework and its implementation, for both frequentist and Bayesian upper limits. Section IV discusses how (frequentist) upper limits are typically measured using Monte-Carlo injection-recovery simulations. Section V provides comparisons of our sensitivity estimates to simulated upper limits in Gaussian noise, while in Sec. VI we provide a comprehensive summary of published sensitivities of past CW searches (translated into sensitivity depth), and a comparison to our sensitivity estimates where applicable. We summarize and discuss the results in Sec. VII. Further details on the referenced searches and upper limits are given in Appendix A. More technical details on the signal model can be found in Appendix B. Finally, Appendix D

contains a discussion of the distribution of the maximum $\mathcal{F}$-statistic over correlated templates.

II. $\mathcal{F}$-STATISTIC-BASED SEARCH METHODS

This section provides an overview of the $\mathcal{F}$-statistic-based search methods for which sensitivity estimates are derived in Sec. III. Further technical details about the signal model and the $\mathcal{F}$-statistic are given in Appendix B. For a broader review of the CW signal model, assumptions and search methods, see e.g., [8–10].

A. Signal model

For the purpose of sensitivity estimation we assume the data time series $x^X(t)$ from each detector X to be described by Gaussian noise, i.e., $n^X(t) \sim \text{Gauss}(0, S_{\text{n}}^X)$ with zero mean and (single-sided) power-spectral density (PSD) S_{n}^X. A gravitational-wave signal creates an additional strain $h^X(t)$ in the detector, resulting in a time series

$$x^X(t) = n^X(t) + h^X(t). \qquad (3)$$

For continuous gravitational waves the two polarization components can be written as

$$\begin{aligned} h_+(\tau) &= A_+ \cos{(\phi(\tau) + \phi_0)}, \\ h_\times(\tau) &= A_\times \sin{(\phi(\tau) + \phi_0)}, \end{aligned} \qquad (4)$$

where $\phi(\tau)$ describes the phase evolution of the signal in the source frame. For the quasiperiodic signals expected from rotating neutron stars, this can be expressed as a Taylor series expansion around a chosen reference time (here $\tau_{\text{ref}} = 0$ for simplicity) as

$$\phi(\tau) = 2\pi \left(f\tau + \frac{1}{2} \dot{f} \tau^2 + \cdots \right), \qquad (5)$$

in terms of derivatives of the slowly-varying intrinsic CW frequency $f(\tau)$. For a triaxial neutron star spinning about a principal axis, the two polarization amplitudes are given by

$$A_+ = \frac{1}{2} h_0 (1 + \cos^2 \iota), \qquad A_\times = h_0 \cos \iota, \qquad (6)$$

in terms of the angle ι between the line of sight and the neutron star rotation axis and the overall signal amplitude h_0. This definition uses the gauge condition of $A_+ \geq |A_\times|$. After translating the source-frame signal into the detector frame (see Appendix B for details), the strain signal $h^X(t)$ at each detector X can be expressed in the factored form

$$h^X(t; \mathcal{A}, \lambda) = \sum_{\mu=1}^{4} \mathcal{A}^\mu h_\mu^X(t; \lambda), \qquad (7)$$

which was first shown in [37], and where the four amplitudes $\mathcal{A}^\mu$ depend on the *amplitude parameters*

$\{h_0, \cos\iota, \psi, \phi_0\}$ as given in Eq. (B5). The four basis functions $h_\mu^X(t; \lambda)$, which are given explicitly in Eq. (B6), depend on the *phase-evolution parameters* $\lambda = \{\hat{n}, f, \dot{f}, \ldots\}$, namely sky position $\hat{n}$, frequency f and its derivatives $f^{(k)} = d^k f / d\tau^k|_{\tau_{\text{ref}}}$, and binary-orbital parameters in the case of a neutron star in a binary.

B. Coherent $\mathcal{F}$-statistic

For pure Gaussian-noise time series $\{n^X(t)\}$ in all detectors X, the likelihood can be written as (e.g., see [38–40]):

$$P(x = n | S_n) = \kappa e^{-\frac{1}{2}(n, n)}, \tag{8}$$

in terms of the multidetector scalar product

$$(x, y) \equiv 4\text{Re} \sum_X \int_0^\infty \frac{\tilde{x}^X(f) \tilde{y}^{X*}(f)}{S_n^X(f)} df, \tag{9}$$

where $\tilde{x}(f)$ denotes the Fourier transform of $x(t)$, and x^* denotes complex conjugation of x. Using the additivity of noise and signals [cf. Eq. (3)], we can express the likelihood for data x, assuming Gaussian noise plus a signal $h(\mathcal{A}, \lambda)$ as

$$P(x | S_n, \mathcal{A}, \lambda) = P(x - h(\mathcal{A}, \lambda) | S_n)$$
$$= \kappa e^{-\frac{1}{2}((x-h), (x-h))}. \tag{10}$$

From this we obtain the log-likelihood ratio between the signal and noise hypotheses as

$$\ln \Lambda(x; \mathcal{A}, \lambda) \equiv \ln \frac{P(x | S_n, \mathcal{A}, \lambda)}{P(x | S_n)}$$
$$= (x, h) - \frac{1}{2}(h, h). \tag{11}$$

Analytically maximizing the log-likelihood ratio over $\mathcal{A}$ (cf. Appendix B) yields the $\mathcal{F}$-statistic [37]:

$$\mathcal{F}(x; \lambda) \equiv \max_{\mathcal{A}} \ln \Lambda(x; \mathcal{A}, \lambda). \tag{12}$$

The statistic $2\mathcal{F}$ follows a χ^2-distribution with four degrees of freedom (d.o.f.) and non-centrality ρ^2,

$$P(2\mathcal{F} | \rho^2) = \chi_4^2(2\mathcal{F}; \rho^2), \tag{13}$$

with expectation and variance

$$E[2\mathcal{F}] = 4 + \rho^2, \qquad \text{var}[2\mathcal{F}] = 8 + 4\rho^2, \tag{14}$$

where ρ corresponds to the signal-to-noise ratio (SNR) of coherent matched filtering.

In the *perfect-match* case, where the template phase-evolution parameters λ coincide with the parameters λ_s of a signal in the data x, the SNR can be explicitly expressed as

$$\rho_0^2 \equiv (h, h) = \frac{4}{25} \frac{h_0^2}{S_n} T_{\text{data}} R^2(\theta), \tag{15}$$

where T_{data} is the total duration of data from all detectors,[2] S_n denotes the multidetector noise floor [defined in Eqs. (19) and (20)], and $R(\theta)$ is a geometric factor quantifying the detector response.

The response function $R(\theta)$ (following the definition in [1]) depends on the subset of signal parameters

$$\theta \equiv \{\hat{n}, \cos\iota, \psi\}, \tag{16}$$

and is defined with the normalization:

$$\langle R^2 \rangle_\theta = 1. \tag{17}$$

The explicit expression of R^2 can be found in Appendix C. Using this normalization with Eq. (15) we can recover the well-known sky- and polarization-averaged squared-SNR expression (e.g., see [37]):

$$\langle \rho_0^2 \rangle_\theta = \frac{4}{25} \frac{h_0^2}{S_n} T_{\text{data}}. \tag{18}$$

The multidetector noise-floor S_n is defined as the *harmonic mean* over the per-detector PSDs S_n^X, namely

$$\frac{1}{S_n} \equiv \frac{1}{N} \sum_X \frac{1}{S_n^X}. \tag{19}$$

Note that in practice CW searches do not assume stationary noise over the whole observation time T_{span}, but only over short durations of order ~ 30 mins. This corresponds to the length of the short Fourier transforms (SFTs) that are typically used as input data. The present formalism can straightforwardly be extended to allow for this type of nonstationarity [41]. In this case the definition Eq. (19) of the multidetector noise-PSD S_n generalizes to the *harmonic mean over all SFTs*,

$$\frac{1}{S_n} \equiv \frac{1}{N_{\text{SFT}}} \sum_\alpha \frac{1}{S_n^\alpha}, \tag{20}$$

where α is an index enumerating all SFTs (over all detectors), and S_n^α is the corresponding noise PSD estimated for SFT α.

---

[2]Not to be confused with the observation time T_{span}, denoting the total time between the first sample of the data and the last.

C. Semicoherent $\mathcal{F}$-statistic methods

Semicoherent methods [16] divide the data into N_{seg} segments each spanning a duration $T_{\mathrm{seg}} < T_{\mathrm{span}}$. The segments are analyzed coherently, and the per-segment detection statistics are combined incoherently. Generally this yields lower sensitivity for the same amount of data analyzed than a fully-coherent search. However, the computational cost for a fully-coherent search over the same amount of data is often impossibly large, while the semicoherent cost can be tuned to be affordable and hence ends up being more sensitive at fixed computing cost [16,17,42].

There are a number of different semicoherent methods currently in use, such as PowerFlux, FrequencyHough, SkyHough, TwoSpect, CrossCorr, Viterbi, Sideband, loosely-coherent statistics and others (e.g., see [10] and references therein). Many of these methods work on short segments, typically of length $T_{\mathrm{seg}} \sim 30$ min, and directly use the power in the frequency bins of short Fourier transforms (SFTs) as the coherent base statistic.

In this work we focus exclusively on sensitivity estimation of $\mathcal{F}$-statistic-based methods, namely StackSlide-$\mathcal{F}$ (e.g., see [17]) and Hough-$\mathcal{F}$ introduced in [33]. Here the length of segments is only constrained by the available computing cost, and segments will typically span many hours to days, which yields better sensitivity, but also requires higher computational cost. Therefore, many of the computationally expensive semicoherent $\mathcal{F}$-statistic searches are run on the distributed Einstein@Home computing platform [43].

Note that these methods, which use multiple SFTs for every segment, are not to be confused with the (albeit closely related) "classical" StackSlide and Hough methods, which use single SFTs directly as coherent segments, as described e.g., in [35].

1. StackSlide-$\mathcal{F}$: Summing $\mathcal{F}$-statistics

The StackSlide-$\mathcal{F}$ method uses the sum of the coherent per-segment $\tilde{\mathcal{F}}$-statistic values in a given parameter-space point λ as the detection statistic, namely

$$2\hat{\mathcal{F}} \equiv \sum_{\ell=1}^{N_{\mathrm{seg}}} 2\tilde{\mathcal{F}}_\ell, \qquad (21)$$

where $\tilde{\mathcal{F}}_\ell$ is the coherent $\mathcal{F}$-statistic of Eq. (12) in segment ℓ. This statistic follows a χ^2-distribution with $4N_{\mathrm{seg}}$ d.o.f. and noncentrality ρ^2, i.e.,

$$P(2\hat{\mathcal{F}}|\rho^2) = \chi^2_{4N_{\mathrm{seg}}}(2\hat{\mathcal{F}};\rho^2), \qquad (22)$$

where the noncentrality ρ^2 is identical to the expression for the *coherent squared SNR* of Eq. (15), with T_{data} referring to the whole data set used, and S_{n} is the corresponding noise floor. However, the noncentrality in the semicoherent

case cannot be considered a "signal to noise ratio," due to the larger N_{seg}-dependent d.o.f. of the χ^2 distribution compared to Eq. (13), which increases the false-alarm level at fixed threshold and reduces the "effective" semicoherent $\widehat{\mathrm{SNR}}$ to $\widehat{\mathrm{SNR}}^2 = \rho^2/\sqrt{N_{\mathrm{seg}}}$ (e.g., see [Eq. (14)] in [44]).

The expectation and variance for $2\hat{\mathcal{F}}$ are

$$E[2\mathcal{F}] = 4N_{\mathrm{seg}} + \rho^2, \qquad \mathrm{var}[2\mathcal{F}] = 8N_{\mathrm{seg}} + 4\rho^2. \qquad (23)$$

We note that StackSlide-$\mathcal{F}$ searches often quote the average $\overline{\mathcal{F}}$ over segments instead of the sum $\hat{\mathcal{F}}$, i.e.,

$$\overline{\mathcal{F}} \equiv \frac{1}{N_{\mathrm{seg}}} \hat{\mathcal{F}}. \qquad (24)$$

2. Hough-$\mathcal{F}$: Summing threshold crossings

The Hough-$\mathcal{F}$ method [33] sets a threshold $\tilde{\mathcal{F}}_{\mathrm{th}}$ on the per-segment coherent $\tilde{\mathcal{F}}$-statistics and uses the number of threshold-crossings over segments as the detection statistic, the so-called *Hough number count* n_{c}, i.e.,

$$n_{\mathrm{c}} \equiv \sum_{\ell=1}^{N_{\mathrm{seg}}} \Theta(\tilde{\mathcal{F}}_\ell - \tilde{\mathcal{F}}_{\mathrm{th}}), \qquad (25)$$

where $\Theta(x)$ is the Heaviside step function.

D. Mismatch and template banks

In wide-parameter-space searches the unknown signal parameters $\lambda \in \mathbb{P}$ are assumed to fall somewhere within a given search space $\mathbb{P}$. In this case one needs to compute a statistic (such as those defined in the previous sections) over a whole "bank" of templates $\mathbb{T} \equiv \{\lambda_i\}_{i=1}^{\mathcal{N}}$. This template bank has to be chosen in such a way that any putative signal $\lambda_{\mathrm{s}} \in \mathbb{P}$ would suffer only an acceptable level of loss of SNR. This can be quantified in terms of the so-called *mismatch* μ, defined as the relative loss of $\rho^2(\lambda_{\mathrm{s}};\lambda)$ in a template λ with respect to the perfect-match $\rho^2(\lambda_{\mathrm{s}};\lambda_{\mathrm{s}}) = \rho_0^2$ [of Eq. (15)], namely

$$\mu(\lambda_{\mathrm{s}};\lambda) \equiv \frac{\rho^2(\lambda_{\mathrm{s}};\lambda_{\mathrm{s}}) - \rho^2(\lambda_{\mathrm{s}};\lambda)}{\rho^2(\lambda_{\mathrm{s}};\lambda_{\mathrm{s}})}. \qquad (26)$$

We can therefore express the "effective" noncentrality parameter ρ_{eff}^2 in a template point λ in the $\mathcal{F}$-statistic χ^2-distribution of Eqs. (13), (22) as

$$\rho_{\mathrm{eff}}^2 \equiv \rho^2(\lambda_{\mathrm{s}};\lambda) = (1 - \mu)\rho_0^2. \qquad (27)$$

E. Sensitivity depth

The $\mathcal{F}$-statistic noncentrality parameter ρ^2 depends on signal amplitude h_0 and overall noise floor S_{n} [cf. Eq. (20)]

only through the combination $h_0/\sqrt{S_n}$, as seen in Eq. (15). The sensitivity of a search is therefore most naturally characterized in terms of the so-called *sensitivity depth* [45], defined as

$$\mathcal{D} \equiv \frac{\sqrt{S_n}}{h_0}, \qquad (28)$$

in terms of the overall noise PSD S_n defined as the harmonic mean over all SFTs used in the search, see Eq. (20).

A particular choice of search parameters (N_{seg}, T_{data}, template bank) in general yields a frequency-dependent upper limit $h_0(f)$, due to the frequency-dependent noise floor $S_n(f)$. However, for fixed search parameters this will correspond to a *constant* sensitivity depth $\mathcal{D}$, which is therefore often a more practical and natural way to characterize the performance of a search, independently of the noise floor.

III. SENSITIVITY ESTIMATE

As discussed in more detail in the Introduction, by *sensitivity* we mean the (measured or expected) upper limit on h_0 for a given search (or equivalently, the sensitivity depth $\mathcal{D} = \sqrt{S_n}/h_0$), which can either refer to the frequentist or Bayesian upper limit.

A. Frequentist upper limits

The frequentist upper limit is defined as the weakest signal amplitude h_0 that can be detected at a given detection probability p_{det}[3] (typically chosen as 90% or 95%) above a threshold d_{th} on a statistic $d(x)$. The threshold can be chosen as the loudest candidate obtained in the search, or it can be set corresponding to a desired false-alarm level p_{fa} (or p-value), defined as

$$p_{fa}(d_{th}) \equiv P(d > d_{th}|h_0 = 0), \qquad (29)$$

which can be inverted to yield $d_{th} = d_{th}(p_{fa})$. The detection probability for signals of amplitude h_0 is

$$p_{det}(d_{th}; h_0) \equiv P(d > d_{th}|h_0), \qquad (30)$$

which can be inverted to yield the upper limit $h_0(d_{th}, p_{det})$.

We can write $p_{fa}(d_{th}) = p_{det}(d_{th}; h_0 = 0)$, and so we can express both in terms of the general threshold-crossing probability as

$$P(d > d_{th}|h_0) = \int_{d_{th}}^{\infty} P(d|h_0)\mathrm{d}d. \qquad (31)$$

---

[3]or equivalently, false-dismissal probability $p_{fd} = 1 - p_{det}$

B. Approximating wide-parameter-space statistics

As discussed in Sec. II D, a wide parameter-space search for unknown signals $\lambda \in \mathbb{P}$ normally proceeds by computing a (single-template) statistic over a bank of templates $\mathbb{T} \equiv \{\lambda_i\}_{i=1}^{\mathcal{N}}$ covering the parameter space $\mathbb{P}$. This results in a corresponding set of (single-template) statistic values $\{d^1(x; \lambda_i)\}$, which need to be combined to form the overall wide-parameter-space statistic $d(x)$. This would naturally be obtained via marginalization (i.e., integrating the likelihood over $\mathbb{P}$), but in practice is mostly done by maximizing the single-template statistic over $\mathbb{T}$, i.e.,

$$d(x) \equiv d^*(x) \equiv \max_{\lambda_i \in \mathbb{T}} d^1(x; \lambda_i). \qquad (32)$$

1. Noise case

For the pure noise case of Eq. (29), it is difficult to determine a reliable expression for $P(d^*|h_0 = 0)$, even if the single-template statistic $P(d^1|h_0 = 0)$ follows a known distribution (such as for the $\mathcal{F}$-based statistics discussed in Sec. II). The reason for this difficulty lies in the correlations that generally exist between "nearby" templates in $\lambda_i \in \mathbb{T}$.

If all $\mathcal{N}$ templates were strictly uncorrelated, one could use the well-known expression Eq. (D1) [1,46] for the distribution of the maximum. In this case one can also relate the single-trial p-value $p_{fa}^1 \approx p_{fa}/\mathcal{N}$ to the wide-parameter-space p-value p_{fa} (for $p_{fa}^1 \ll 1$).

Although it is a common assumption in the literature, template correlations do not simply modify the "effective" number of independent templates to use in Eq. (D1), but they generally also affect the functional form of the underlying distribution for the maximum d^*, as illustrated in Appendix D with a simple toy model.

In this work we assume that the upper limit refers to a known detection threshold in Eq. (30). This can be obtained either from (i) the loudest observed candidate (the most common situation in real searches), or from (ii) setting a single-template p-value p_{fa}^1 and inverting the known single-template distribution Eq. (29), or from (iii) a numerically obtained relation between p_{fa} and the threshold d_{th}, e.g., via Monte-Carlo simulation.

2. Signal case

We assume that the highest value of d^1 will be realized near the signal location, i.e.,

$$d^*(x) \approx d^1(x; \lambda^*), \qquad (33)$$

where λ^* is the "closest" template $\in \mathbb{T}$ to the signal location λ_s, defined in terms of the mismatch Eq. (26). This template yields the highest effective noncentrality parameter over the template bank, namely

$$\rho_{eff}^2 \equiv \rho^2(\lambda_s; \lambda^*) = (1 - \mu)\rho_0^2(\lambda_s). \qquad (34)$$

This assumption turns out to be valid as long as the p-value p_{fa} is low (typically $p_{\text{fa}} \lesssim 1\%$) and the signals have relatively high detection probability (typically $p_{\text{det}} \sim 90\%$ or 95%). However, in Sec. V we also encounter deviations from the predictions that can be traced to violations of this assumption.

C. StackSlide-$\mathcal{F}$ sensitivity

We first consider a semi-coherent StackSlide-$\mathcal{F}$ search using the summed $\hat{\mathcal{F}}$-statistic of Eq. (21), i.e., $d^1(x;\lambda) = 2\hat{\mathcal{F}}(x;\lambda)$. This case also includes fully coherent $\mathcal{F}$-statistic searches, which simply correspond to the special case $N_{\text{seg}} = 1$.

We see from Eq. (31) that in order to estimate the sensitivity, we need to know $P(2\hat{\mathcal{F}}|h_0)$. This can be obtained via marginalization (at fixed h_0) of the known distribution $P(2\hat{\mathcal{F}}|\rho^2)$ of Eq. (22), combined with the assumption Eq. (34) that the highest statistic value will occur in the "closest" template, with mismatch distribution $P(\mu)$:

$$P(2\hat{\mathcal{F}}|h_0) = \int P(2\hat{\mathcal{F}}, \theta, \mu|h_0)\mathrm{d}^4\theta\mathrm{d}\mu$$

$$= \int P(2\hat{\mathcal{F}}|h_0, \theta, \mu)P(\theta)P(\mu)\mathrm{d}^4\theta\mathrm{d}\mu$$

$$= \int P(2\hat{\mathcal{F}}|\rho^2_{\text{eff}})P(\theta)P(\mu)\mathrm{d}^4\theta\mathrm{d}\mu, \quad (35)$$

where $\rho^2_{\text{eff}}(h_0, \theta, \mu) = \rho_0^2(h_0, \theta)(1 - \mu)$ in terms of the perfect-match noncentrality ρ_0^2 defined in Eq. (15), and in the last step we used the fact that the distribution for $2\hat{\mathcal{F}}$ is fully specified in terms of the noncentrality parameter ρ^2 of the χ^2-distribution with $4N_{\text{seg}}$ d.o.f., as given in Eq. (22).

Equation (35) requires five-dimensional integration for each sensitivity estimation, which would be slow and cumbersome. One of the key insights in [1] was to notice that the perfect-match SNR ρ_0 of Eq. (15) depends on the four parameters θ only through the scalar $R^2(\theta)$, and we can therefore use a reparametrization

$$\int_{\theta(R^2)} P(\theta)\mathrm{d}^4\theta = P(R^2)\mathrm{d}R^2, \quad (36)$$

where $\theta(R^2)$ denotes the subspace of θ values yielding a particular R^2 from Eq. (C1).

The one-dimensional distribution $P(R^2)$ can be obtained by Monte-Carlo sampling over the priors of sky-position $\hat{n}$ (typically either isotropically over the whole sky, or a single sky-position in case of a directed search) and polarization angles $\cos\iota$ (uniform in $[-1, 1]$) and ψ (uniform in $[-\pi/4, \pi/4]$). The resulting values of $R^2(\theta)$ are histogrammed and used as an approximation for $P(R^2)$, which can be reused for repeated sensitivity estimations with the same θ-priors. We can thus rewrite Eq. (35) as

$$P(2\hat{\mathcal{F}}|h_0) = \int P(2\hat{\mathcal{F}}|\rho^2_{\text{eff}})P(R^2)P(\mu)\mathrm{d}R^2\mathrm{d}\mu, \quad (37)$$

with

$$P(2\hat{\mathcal{F}}|\rho^2_{\text{eff}}) = \chi^2_{4N_{\text{seg}}}(2\hat{\mathcal{F}}; \rho^2_{\text{eff}}), \quad (38)$$

$$\rho^2_{\text{eff}}(h_0, R^2, \mu) = \frac{4}{25}\frac{h_0^2}{S_{\text{n}}}T_{\text{data}}R^2(1 - \mu). \quad (39)$$

The mismatch distribution $P(\mu)$ for any given search can be obtained via injection-recovery Monte-Carlo simulation, where signals are repeatedly generated (without noise) and searched for over the template bank, obtaining the corresponding mismatch μ for each injection. This is a common step in validating a search and template-bank setup. Alternatively, for some search methods precomputed estimates for the mismatch distributions exist as a function of the template-bank parameters, e.g., for the WEAVE search code [47].

Inserting Eq. (37) into the detection probability of Eq. (31), we obtain

$$p_{\text{det}}(2\hat{\mathcal{F}}_{\text{th}}; h_0) = \int p_{\text{det}}(2\hat{\mathcal{F}}_{\text{th}}; \rho^2_{\text{eff}})P(R^2)P(\mu)\mathrm{d}R^2\mathrm{d}\mu, \quad (40)$$

where

$$p_{\text{det}}(2\hat{\mathcal{F}}_{\text{th}}; \rho^2_{\text{eff}}) \equiv \int_{2\hat{\mathcal{F}}_{\text{th}}}^{\infty} \chi^2_{4N_{\text{seg}}}(2\hat{\mathcal{F}}; \rho^2_{\text{eff}})\mathrm{d}2\hat{\mathcal{F}}. \quad (41)$$

Equation (40) can be easily and efficiently computed numerically, and simple inversion (via 1-D root-finding) yields the sensitivity (i.e., upper limit) h_0 for given detection probability p_{det} and threshold $2\hat{\mathcal{F}}_{\text{th}}$.

D. Multistage StackSlide-$\mathcal{F}$ sensitivity

The sensitivity estimate for a single StackSlide-$\mathcal{F}$ search can be generalized to hierarchical multistage searches, where threshold-crossing candidates of one search stage are followed up by deeper subsequent searches in order to increase the overall sensitivity (e.g., see [16,26,42,48,49]). We denote the n stages with an index $i = 1...n$. Each stage i is characterized by the number $N_{\text{seg}}^{(i)}$ of segments, the amount of data $T_{\text{data}}^{(i)}$, the noise PSD $S_{\text{n}}^{(i)}$, a mismatch distribution $P(\mu^{(i)})$, and a threshold $2\hat{\mathcal{F}}_{\text{th}}^{(i)}$ (corresponding to a false-alarm level $p_{\text{fa}}^{(i)}$ at that stage).

The initial wide-parameter-space search (stage $i = 1$) yields candidates that cross the threshold $2\mathcal{F}_{\text{th}}^{(1)}$ in certain templates $\{\lambda\}$. The next stage follows up these candidates with a more sensitive search, which can be achieved by reducing the mismatch $\mu^{(i)}$ (choosing a finer template bank grid), or by increasing the coherent segment length

(and reducing the number of segments $N_{\text{seg}}^{(i)}$). Often the final stage $i = n$ in such a follow-up hierarchy would be fully coherent, i.e., $N_{\text{seg}}^{(n)} = 1$.

In order for any given candidate (which can be either due to noise or a signal) to cross the final threshold $2\mathcal{F}^{(n)}$, it has to cross *all* previous thresholds as well, in other words Eq. (29), (30) now generalize to

$$p_{\text{det}}^{(\text{tot})}(h_0) = P(\{2\hat{\mathcal{F}}^{(i)} > 2\hat{\mathcal{F}}_{\text{th}}^{(i)}\}_{i=1}^{n}|h_0). \quad (42)$$

In order to make progress at this point we need to assume that the threshold-crossing probabilities in different stages are *independent* of each other, so for $j \neq i$ we assume

$$P(2\hat{\mathcal{F}}^{(i)} > 2\hat{\mathcal{F}}_{\text{th}}^{(i)}|\rho^2, 2\hat{\mathcal{F}}^{(j)} > 2\hat{\mathcal{F}}_{\text{th}}^{(j)}) = P(2\hat{\mathcal{F}}^{(i)} > 2\hat{\mathcal{F}}_{\text{th}}^{(i)}|\rho^2), \quad (43)$$

which would be exactly true if the different stages used different data (see also [42]). In the case where the same data is used in different stages, this approximation corresponds to an *uninformative* approach, in the sense that we do not know how to quantify and take into account the correlations between the statistics in different stages. We proceed without using this potential information, which could in principle be used to improve the estimate. It is not clear if and how much of an overall bias this approximation would introduce. A detailed study of this question is beyond the scope of this work and will be left for future study.

Using the assumption of independent stages we write

$$p_{\text{det}}^{(\text{tot})}(h_0) = \int \prod_{i=1}^{n} p_{\text{det}}^{(i)}(2\hat{\mathcal{F}}_{\text{th}}^{(i)}; h_0, R^2)P(R^2)dR^2, \quad (44)$$

$$p_{\text{fa}}^{(\text{tot})} = \prod_{i=1}^{n} p_{\text{fa}}^{(i)}(2\hat{\mathcal{F}}_{\text{th}}^{(i)}), \quad (45)$$

where now the R^2-marginalization needs to happen over all stages combined, as the signal parameters $R^2(\theta)$ are intrinsic to the signal and hence independent of the stage. On the other hand, the mismatch distribution $P(\mu^{(i)})$ depends on the stage, as each stage will in general use a different template grid, and so we have

$$p_{\text{det}}^{(i)}(2\hat{\mathcal{F}}_{\text{th}}^{(i)}; h_0, R^2) = \int_0^1 p_{\text{det}}^{(i)}(2\hat{\mathcal{F}}_{\text{th}}^{(i)}; \rho_{\text{eff}}^{2(i)})P(\mu^{(i)})d\mu^{(i)}, \quad (46)$$

where $p_{\text{det}}(2\hat{\mathcal{F}}_{\text{th}}; \rho_{\text{eff}}^2)$ is given by Eq. (41) using the respective per-stage values.

Equation (44) can easily be solved numerically and inverted for the sensitivity h_0 at given $p_{\text{det}}^{(\text{tot})}$ and a set of thresholds $\{2\hat{\mathcal{F}}_{\text{th}}^{(i)}\}$.

Note that in practice (e.g., [49]) one would want to choose the thresholds in such a way that a signal that passed the 1st-stage threshold $2\hat{\mathcal{F}}_{\text{th}}^{(1)}$ should have a very low probability of being discarded by subsequent stages, in other words $p_{\text{det}}^{(i>1)} \approx 1$, and therefore $p_{\text{det}}^{(\text{tot})}(h_0) \approx p_{\text{det}}^{(1)}(2\hat{\mathcal{F}}_{\text{th}}^{(1)}; h_0)$. Therefore subsequent stages mostly serve to reduce the total false-alarm level $p_{\text{fa}}^{(\text{tot})}$, allowing one to increase the first-stage $p_{\text{fa}}^{(1)}$ by lowering the corresponding threshold $\hat{\mathcal{F}}^{(1)}$, resulting in an overall increased sensitivity.

E. Hough-$\mathcal{F}$ sensitivity

Here we apply the sensitivity-estimation framework to the Hough-$\mathcal{F}$ statistic introduced in Sec. II C 2. We define the per-segment threshold-crossing probability as

$$\begin{aligned} p_{\text{th}}^{\ell}(h_0, R^2) &\equiv P(2\tilde{\mathcal{F}}_{\ell} > 2\tilde{\mathcal{F}}_{\text{th}}|h_0, R^2) \\ &= p_{\text{det}}^{\ell}(2\tilde{\mathcal{F}}_{\text{th}}; h_0, R^2) \\ &= \int_0^1 p_{\text{det}}(2\tilde{\mathcal{F}}_{\text{th}}; \rho_{\text{eff},\ell}^2)P(\tilde{\mu})d\tilde{\mu}, \quad (47) \end{aligned}$$

where the per-segment effective SNR $\rho_{\text{eff},\ell}$ is given by replacing T_{data} and S_{n} in Eq. (39) with the per-segment quantities T_{data}^{ℓ} and S_{n}^{ℓ}. For the per-segment mismatch $\tilde{\mu}^{\ell}$ we assume that the distribution $P(\tilde{\mu})$ is the same for all segments.

The key approximation for the estimate is that for a given signal $\{h_0, R^2(\theta)\}$, the coherent per-segment $\tilde{\mathcal{F}}_{\ell}$-statistic has the same threshold-crossing probability p_{th} in every segment ℓ, i.e., $p_{\text{th}}^{\ell} = p_{\text{th}}$ for all $\ell = 1 \dots N_{\text{seg}}$. This allows us to write the probability for the Hough number count n_{c} of Eq. (25) for a fixed signal $\{h_0, R^2\}$ as a binomial distribution:

$$P(n_{\text{c}}|h_0, R^2) = \binom{N_{\text{seg}}}{n_{\text{c}}} p_{\text{th}}^{n_{\text{c}}}(1 - p_{\text{th}})^{N_{\text{seg}}-n_{\text{c}}}, \quad (48)$$

with $p_{\text{th}}(h_0, R^2)$ given by Eq. (47). For a given threshold $n_{\text{c,th}}$ on the number count the detection probability is

$$p_{\text{det}}(n_{\text{c,th}}; h_0, R^2) = \sum_{n_{\text{c}}=n_{\text{c,th}}}^{N_{\text{seg}}} P(n_{\text{c}}|h_0, R^2). \quad (49)$$

Marginalization over R^2 yields the detection probability at fixed amplitude h_0:

$$p_{\text{det}}(n_{\text{c,th}}; h_0) = \int p_{\text{det}}(n_{\text{c,th}}; h_0, R^2)P(R^2)dR^2. \quad (50)$$

We can numerically solve this for h_0 at given p_{det} and number-count threshold $n_{\text{c,th}}$ yielding the desired sensitivity estimate.

F. Bayesian Upper Limits

Bayesian upper limits are conceptually quite different [50] from the frequentist ones discussed up to this point. A Bayesian upper limit h_0^C of given confidence (or "credible level") C corresponds to the interval $[0, h_0^C]$ that contains the true value of h_0 with probability C. We can compute this from the posterior distribution $P(h_0|x)$ for the signal-amplitude h_0 given data x, namely

$$C = P(h_0 < h_0^C|x) = \int_0^{h_0^C} P(h_0|x)\mathrm{d}h_0. \tag{51}$$

The Bayesian targeted searches (here referred to as *BayesPE*) for known pulsars (see Table V and Sec. A 5) compute the posterior $P(h_0|x)$ directly from the data x, using a time-domain method introduced in [51].

Here we focus instead on $\mathcal{F}$-statistic-based searches over a template bank. As discussed in [50], to a very good approximation we can compute the posterior from the loudest candidate $2\mathcal{F}^*(x)$ found in such a search, using this as a proxy for the data x, i.e.,

$$P(h_0|x) \approx P(h_0|2\mathcal{F}^*(x)) \tag{52}$$

$$\propto P(2\mathcal{F}^*(x)|h_0)P(h_0), \tag{53}$$

where we used Bayes' theorem. The proportionality constant is determined by the normalization condition $\int P(h_0|x)\mathrm{d}h_0 = 1$.

We have already derived the expression for $P(2\mathcal{F}|h_0)$ in Eq. (37). Hence for any choice of prior $P(h_0)$ we can now compute the Bayesian upper limit $h_0^C(2\mathcal{F}^*)$ for given loudest candidate $2\mathcal{F}^*$ by inverting Eq. (51).

It is common for Bayesian upper limits on the amplitude to choose a uniform prior in h_0 (e.g., see [52]), which has the benefit of simplicity, and also puts relatively more weight on larger values of h_0 than might be physically expected (weaker signals should be more likely than stronger ones). This prior therefore results in larger, i.e., "more conservative," upper limits than a more physical prior would.

Note that the Bayesian ULs of targeted searches for known pulsars (see Sec. A 5) compute the h_0-posterior directly from the data rather than from an $\mathcal{F}$-statistic. Therefore we cannot use a known threshold or loudest candidate $2\mathcal{F}^*$ for inverting Eq. (51) and hence we cannot apply the above framework directly. We instead compute an *expected depth* by calculating estimates for $2\mathcal{F}^*$-values drawn randomly from the central χ_4^2-distribution and averaging the results.

G. Numerical implementation

The expressions for the various different sensitivity estimates of the previous sections have been implemented

in GNU OCTAVE [53], and are available as part of the OctApps [54] data-analysis package for continuous gravitational waves.

The function to estimate (and cache for later reuse) the distribution $P(R^2)$ of Eq. (36) is implemented in SqrSNRGeometricFactorHist().

The sensitivity-depth estimate for StackSlide-$\mathcal{F}$-searches is implemented in SensitivityDepthStackSlide(), both for the single-stage case of Eq. (40) and for the general multistage case of Eq. (44). For single-stage StackSlide-$\mathcal{F}$ there is also a function DetectionProbabilityStackSlide() estimating the detection probability for a given signal depth $\mathcal{D}$ and detection threshold.

The Hough-$\mathcal{F}$ sensitivity estimate of Eq. (50) is implemented in SensitivityDepthHoughF(). An earlier version of this function had been used for the theoretical sensitivity comparison in [36] (Sec. V B, and also [55]), where it was found to agree within an rms error of 7% with the measured upper limits.

The Bayesian $\mathcal{F}$-based upper limit expression Eq. (51) is implemented in SensitivityDepthBayesian().

Typical input parameters are the number of segments N_{seg}, the total amount of data T_{data}, the mismatch distribution $P(\mu)$, name of detectors used, single-template false-alarm level p_{fa}^1 (or alternatively, the $\mathcal{F}$-statistic threshold), and the confidence level p_{det}. The default prior on sky-position is isotropic (suitable for an all-sky search), but this can be restricted to any sky-region (suitable for directed or targeted searches).

The typical runtime on a ThinkPad P51 with 3 GHz Intel Xeon E3 for a sensitivity estimate including computing $P(R^2)$ (which is the most expensive part) is about 25 seconds per detector. When reusing the same θ-prior on subsequent calls, a cached $P(R^2)$ is used and the runtime is reduced to about 10 seconds total, independently of the number of detectors used.

IV. DETERMINING FREQUENTIST UPPER LIMITS

In order to determine the frequentist upper limit (UL) on the signal amplitude h_0 defined in Eq. (30), one needs to quantify the probability that a putative signal with fixed amplitude h_0 (and all other signal parameters drawn randomly from their priors) would produce a statistic value exceeding the threshold (corresponding to a certain false-alarm level, or p-value). The upper limit on h_0 is then defined as the value $h_0^{p_{\mathrm{det}}}$ for which the detection probability is exactly p_{det}, typically chosen as 90% or 95%, which is often referred to as the *confidence level* of the UL.

Note that here and in the following it will often be convenient to use the sensitivity depth $\mathcal{D} \equiv \sqrt{S_{\mathrm{n}}}/h_0$ introduced in Sec. II E instead of the amplitude h_0. We denote $\mathcal{D}^{p_{\mathrm{det}}}$ as the sensitivity depth corresponding to the upper

limit $h_0^{p_{\det}}$ (note that this corresponds to a *lower limit* on depth).

The UL procedure is often implemented via a Monte-Carlo injection-and-recovery method: a signal of fixed amplitude $h_0 = \sqrt{S_n}/\mathcal{D}$ and randomly-drawn remaining parameters is generated in software and added to the data (either to real detector data or to simulated Gaussian noise). This step is referred to as a *signal injection*. A search is then performed on this data, and the loudest statistic value $\mathcal{F}^*$ is recorded and compared against the detection threshold $\mathcal{F}_{\text{th}}$. Repeating this injection and recovery step many times and recording the fraction of times the threshold is exceeded yields approximation for $p_{\det}(\mathcal{F}_{\text{th}}; \mathcal{D})$. By repeating this procedure over different $\mathcal{D}$ values and interpolating one can find $\mathcal{D}^{p_{\det}}$ corresponding to the desired detection probability (and therefore also $h_0^{p_{\det}}$).

We distinguish in the following between *measured* and *simulated* upper limits:

(i) *Measured ULs* refer to the published UL results obtained on real detector data. These typically use an identical search procedure for the ULs as in the actual search, often using the loudest candidate (over some range of the parameter space) from the original search as the corresponding detection threshold for setting the UL. The injections are done in real detector data, and normally the various vetoes, data-cleaning and follow-up procedures of the original search are also applied in the UL procedure.

(ii) *Simulated ULs* are used in this work to verify the accuracy of the sensitivity estimates. They are obtained using injections in simulated Gaussian noise, and searching only a small box in parameter space around the injected signal locations. The box size is empirically determined to ensure that the loudest signal candidates are always recovered within the box. Only the original search statistic is used in the search without any further vetoes or cleaning.

A key difference between (most) published (measured) ULs and our simulated ULs concerns the method of interpolation used to obtain $\mathcal{D}^{p_{\det}}$: in practice this is often obtained via a sigmoid $p_{\det}$-interpolation approach (Sec. IV A), while we use (and advocate for) a (piecewise) linear threshold interpolation (Sec. IV B) instead.

A. Sigmoid $p_{\det}$ interpolation

In this approach one fixes the detection threshold $\mathcal{F}_{\text{th}}$ and determines the corresponding $p_{\det}$ for any given fixed-$\mathcal{D}$ injection set. The corresponding functional form of $p_{\det}(\mathcal{D})$ has a qualitative "sigmoid" shape as illustrated in Fig. 1. An actual sigmoid function of the form

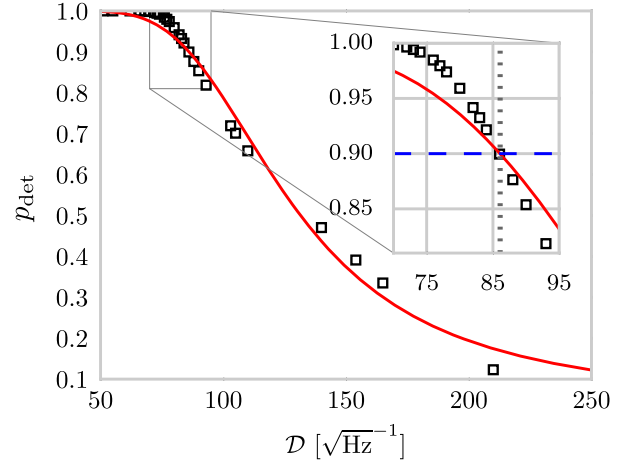$$y(\mathcal{D}) = \frac{1}{1 + e^{-k(\mathcal{D} - \mathcal{D}_0)}}, \tag{54}$$

FIG. 1. Detection probability $p_{\det}$ versus sensitivity depth $\mathcal{D}$ for the `S6-CasA-StackSlide-`$\mathcal{F}$ search (cf. Table II and Sec. A 3), using a detection threshold of $2\overline{\mathcal{F}}_{\text{th}} = 8$. The squares indicate the results from a simulation in Gaussian noise, while the solid line gives the best-fit sigmoid of Eq. (54).

is then fit to the data by adjusting the free parameters k and $\mathcal{D}_0$, and from this one can obtain an interpolation value for $\mathcal{D}^{p_{\det}}$.

One problem with this method is that the actual functional form of $p_{\det}(\mathcal{D})$ is not analytically known, and does not actually seem to be well described by the sigmoid of Eq. (54), as seen in Fig. 1. In this particular example the true value at $p_{\det} = 90\%$ just so happens to lie very close to the sigmoid fit, but the deviation is quite noticeable at $p_{\det} = 95\%$ (see the zoomed inset in Fig. 1).

Another problem with this method is that the range of depths required to sample the relation $p_{\det}(\mathcal{D})$ often needs to be quite wide, due to initial uncertainties about where the UL value would be found, which can compound the above-mentioned sigmoid-fitting problem. Furthermore, the injection-recovery step can be quite computationally expensive, limiting the number of trials and further increasing the statistical uncertainty on the $p_{\det}$ measurements.

Both of these problems can be mitigated to some extent by using the sensitivity-estimation method described in this paper (Sec. III) to obtain a fairly accurate initial guess about the expected UL value, and then sample only in a small region around this estimate, in which case even a linear fit would probably yield good accuracy.

B. Piecewise-linear threshold interpolation

An alternative approach is used in this work to obtain the simulated ULs: for each set of fixed-$\mathcal{D}$ injections and recoveries, we determine the threshold on the statistic required in order to obtain the desired detection fraction $p_{\det}$. This is illustrated in Fig. 2, which shows a histogram of the observed loudest $2\overline{\mathcal{F}}$ candidates obtained in each of $N = 10^4$ injection and recovery runs at a fixed signal depth
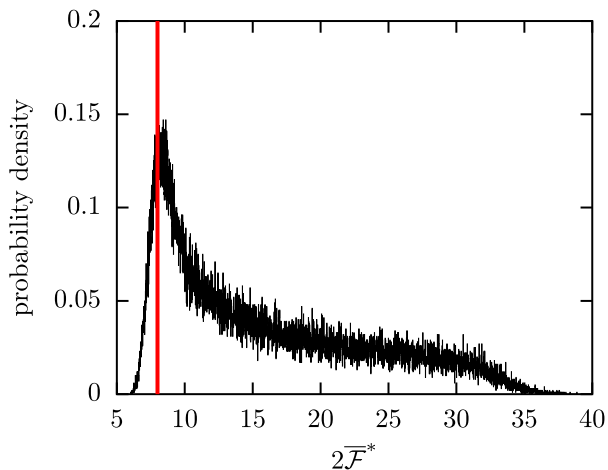
FIG. 2. Histogram of recovered loudest $2\overline{\mathcal{F}}$ values for repeated searches on signal injections at fixed sensitivity depth $\mathcal{D} = 86~\text{Hz}^{-1/2}$ (with all other signal parameters randomized), using the search setup of the `S6-CasA-StackSlide-`$\mathcal{F}$ directed search. The vertical line indicates the resulting threshold value $2\overline{\mathcal{F}}_{\text{th}} = 7.995$ corresponding to $p_{\text{det}} = 90\%$ for this injection set.

of $\mathcal{D} = 86~\text{Hz}^{-1/2}$, using the `S6-CasA-StackSlide-`$\mathcal{F}$ search setup (cf. Sec. A 3). By integrating the probability density from $2\overline{\mathcal{F}} = 0$ until we reach the desired value $1 - p_{\text{det}}$, we find the detection threshold $2\overline{\mathcal{F}}_{\text{th}}$ at this signal depth $\mathcal{D}$. Repeating this procedure at different depths therefore generates a sampling of the function $\mathcal{D}^{p_{\text{det}}}(2\overline{\mathcal{F}}_{\text{th}})$, illustrated in Fig. 3. These points can be interpolated to the required detection threshold, which yields the desired upper-limit depth $\mathcal{D}^{p_{\text{det}}}$.

We see in Fig. 3 that this function appears to be less "curvy" in the region of interest compared to $p_{\text{det}}(\mathcal{D})$
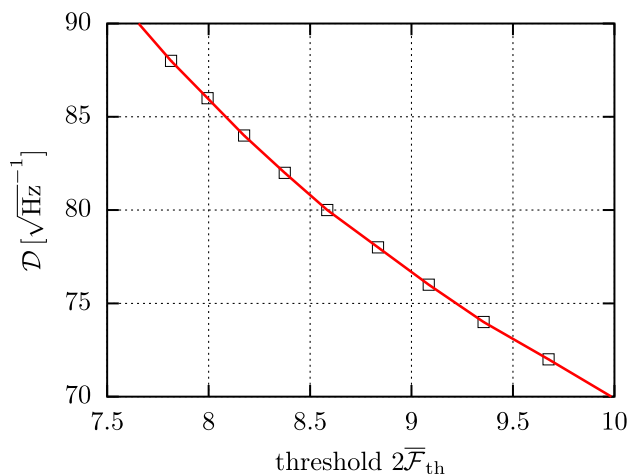

FIG. 3. Sensitivity depth versus detection threshold. Boxes and solid lines indicate the piecewise-linear interpolation through the obtained thresholds at different depths of an injection-recovery simulation, using the `S6-CasA-StackSlide-`$\mathcal{F}$ search setup ([22] and Sec. A 3).

shown in Fig. 1. This allows for easier fitting and interpolation, e.g., a linear or quadratic fit should work quite well. In fact, here we have simply used piecewise-linear interpolation, which is sufficient given our relatively fine sampling of signal depths.

As already mentioned in the previous section, using the sensitivity estimate of Sec. III one can determine the most relevant region of interest beforehand and focus the Monte-Carlo injection-recoveries on this region, which will help ensure that any simple interpolation method will work well.

Alternatively, for either the $p_{\text{det}}(\mathcal{D})$- or the $\mathcal{D}(2\mathcal{F}_{\text{th}})$-sampling approach, one could also use an iterative root-finding method to approach the desired p_{det} or $2\mathcal{F}_{\text{th}}$, respectively.

V. COMPARING ESTIMATES AGAINST SIMULATED UPPER LIMITS

In this section we compare the sensitivity estimates from Sec. III against simulated ULs for two example cases (an all-sky search and a directed search), in order to quantify the accuracy and reliability of the estimation method and implementation. This comparison shows generally good agreement, and also some instructive deviations.

Both examples are wide-parameter-space searches using a template bank over the unknown signal parameter dimensions (namely, {sky, frequency and spindown} in the all-sky case, and {frequency and first and second derivatives} in the directed-search case).

The simulated-UL procedure (see Sec. IV) performs a template-bank search over a box in parameter space containing the injected signal (at a randomized location) in Gaussian noise. On the other hand, the sensitivity estimate [cf. Eq. (40)] uses the mismatch distribution $P(\mu)$ obtained for this template bank via injection-recovery box searches on signals without noise. We refer to this in the following as the *box search*.

It will be instructive to also consider the (unrealistic) case of a perfectly-matched search, using only a single template that matches the signal parameters perfectly for every injection, corresponding to zero mismatch $\mu = 0$ in Eq. (40). We refer to this as the *zero-mismatch search*.

A. Example: S6-AllSky-StackSlide-$\mathcal{F}$ search

In this example we use the setup of the all-sky search `S6-AllSky-StackSlide-`$\mathcal{F}$ [56], which was using the GCT implementation [57] of the StackSlide-$\mathcal{F}$ statistic and was performed on the volunteer-computing project Einstein@Home [43], see Table I and Sec. A 2 for more details.

Figure 4 shows the comparison between simulated ULs and estimated sensitivity depths $\mathcal{D}^{90\%}$ versus threshold $2\overline{\mathcal{F}}_{\text{th}}$, for the *box search* (squares and solid line), as well as for the *zero-mismatch search* (crosses and dashed line). We see excellent agreement between estimated and simulated

TABLE I. All-sky searches: estimated $\mathcal{D}_{\mathrm{est}}$ and measured sensitivity depth $\mathcal{D}_{\mathrm{meas}}$ (median and standard deviation, see Sec. VI A 1). The columns labeled f and $\dot{f}$ give the frequency and spindown ranges covered by each search. Sensitivity depths in *italics* refer to 90%-confidence upper limits, while normal font refers to 95%-confidence. See Appendix A 2 for further details on the individual results.

Data	Search method	f [Hz]	$\dot{f}$ [nHz s^{-1}]	$\mathcal{D}_{\mathrm{est}}$ [Hz$^{-1/2}$]	$\mathcal{D}_{\mathrm{meas}}^{\mathrm{med}}$ [Hz$^{-1/2}$]	$\hat{\sigma}_{\mathcal{D}_{\mathrm{meas}}}$ [Hz$^{-1/2}$]	Ref., Sec.
S2	Hough	[200, 400]	[−1.1, 0]	...	11.3	1.5	[58], A 2 a
S2	$\mathcal{F}$	[160, 728.8]	0	6.5	5.5	1.6	[59], A 2 b
S4	StackSlide	[50, 1000]	[−10, 0]	...	10.5	1.1	[35], A 2 c
S4	Hough	[50, 1000]	[−2.2, 0]	...	13.4	0.7	[35], A 2 c
S4	PowerFlux	[50, 1000]	[−10, 0]	...	$\{6.1, 21.3\}^{\mathrm{a}}$	$\{0.7, 2.3\}$	[35], A 2 c
S4	$\mathcal{F}$ + Coinc	[50, 1500]	[−9.5, 1]	...	8.5	0.5	[60], A 2 d
earlyS5	PowerFlux	[50, 1100]	[−5, 0]	...	$\{16.1, 47.9\}^{\mathrm{a}}$	$\{2.4, 5.9\}$	[61], A 2 e
earlyS5	$\mathcal{F}$ + Coinc	[50, 1500]	[−12.7, 1.3]	...	*10.9*	0.2	[62], A 2 f
S5	PowerFlux	[50, 800]	[−6, 0]	...	$\{25.7, 71.3\}^{\mathrm{a}}$	$\{0.7, 2.2\}$	[63], A 2 g
S5	Hough-$\mathcal{F}$	[50, 1190]	[−2, 0.1]	*30.5*	*30.0*	1.4	[36], A 2 h
S5	Hough	[50, 1000]	[−0.9, 0]	...	28.1	0.6	[64], A 2 i
S5	StackSlide-$\mathcal{F}$	[1249.7, 1499.7]	[−2.9, 0.6]	*27.0*	*30.7*	...	[65], A 2 j
VSR1	$\mathcal{F}_{\mathrm{TD}}$ + Coinc	[100, 1000]	[−16, 0]	...	22.6	6.0	[66], A 2 k
VSR2,4	FreqHough + FUP	[20, 128]	[−0.1, 0.015]	...	35.5	11.1	[67], A 2 l
S6	StackSlide-$\mathcal{F}$	[50, 510]	[−2.7, 0.3]	*34.4*	*37.0*	...	[56], A 2 m
S6	StackSlide-$\mathcal{F}$ + FUP	[50, 510]	[−2.7, 0.3]	*38.3*	*46.9*	...	[49], A 2 n
S6	PowerFlux	[100, 1500]	[−11.8, 10]	...	$\{17.9, 52.8\}^{\mathrm{a}}$	$\{1.4, 3.4\}$	[68], A 2 o
O1	StackSlide-$\mathcal{F}$	[20, 100]	[−2.7, 0.3]	*46.4*	*48.7*	...	[26], A 2 p
O1	PowerFlux	[20, 200]	[−10, 1]	...	28.9	2.2	[26], A 2 q
O1	PowerFlux	[20, 475]	[−10, 1]	...	$\{19.9, 54.6\}^{\mathrm{a}}$	$\{1.3, 3.2\}$	[25], A 2 q
O1	SkyHough	[20, 475]	[−10, 1]	...	22.4	1.1	[25], A 2 q
O1	$\mathcal{F}_{\mathrm{TD}}$ + Coinc	[20, 475]	[−10, 1]	...	23.7	2.1	[25], A 2 q
O1	FreqHough	[20, 475]	[−10, 1]	...	21.4	10.6	[25], A 2 q
O1	PowerFlux	[475, 2000]	[−10, 1]	...	$\{18.6, 50.9\}^{\mathrm{a}}$	$\{1.3, 3.4\}$	[69], A 2 q
O1	SkyHough	[475, 2000]	[−10, 1]	...	16.8	3.0	[69], A 2 q
O1	$\mathcal{F}_{\mathrm{TD}}$ + Coinc	[475, 2000]	[−10, 1]	...	10.9	0.6	[69], A 2 q

$^{\mathrm{a}}$Sensitivity depths corresponding to worst linear and circular polarization, respectively, cf. Sec. A 1.

ULs for the zero-mismatch search. We also find very good agreement for the box-search at higher thresholds, while we see an increasing divergence $\mathcal{D} \to \infty$ of the simulated ULs at decreasing thresholds, which is not captured by the estimate.

This discrepancy can be understood as the effect of noise fluctuations, which can enter in two different ways (that are not completely independent of each other):

(i) For decreasing thresholds the corresponding false-alarm level Eq. (29) grows, as it becomes increasingly likely that a "pure noise" candidate (i.e., unrelated to a signal) crosses the threshold. In the extreme case where p_{fa} approaches p_{det}, the frequentist upper limit would tend to $h_0 \to 0$, corresponding to $\mathcal{D} \to \infty$.[4] This is illustrated in Fig. 5 showing the distribution of the loudest $2\mathcal{F}$ in a box search on pure Gaussian noise, which can be compared to the diverging depth of the simulated box search around $2\overline{\mathcal{F}}_{\mathrm{th}} \lesssim 6$ in Fig. 4.

FIG. 4. Comparison of estimated and simulated sensitivity depth $\mathcal{D}^{90\%}$ as a function of threshold $2\overline{\mathcal{F}}_{\mathrm{th}}$ for the S6-AllSky-StackSlide-$\mathcal{F}$ search [56]. The solid line shows the UL estimate for the *box search*, and the squares ($\square$) show the corresponding simulated ULs. The dashed line indicates the estimate for the *zero-mismatch case*, and the crosses ($\times$) are for the simulated zero-mismatch ULs. In the box search we observe an increasing divergence at decreasing thresholds due to noise effects, discussed in Sec. V A.

---

[4]Bayesian upper limits do not have this property, e.g., see [50] for more detailed analysis of these different types of upper limits.
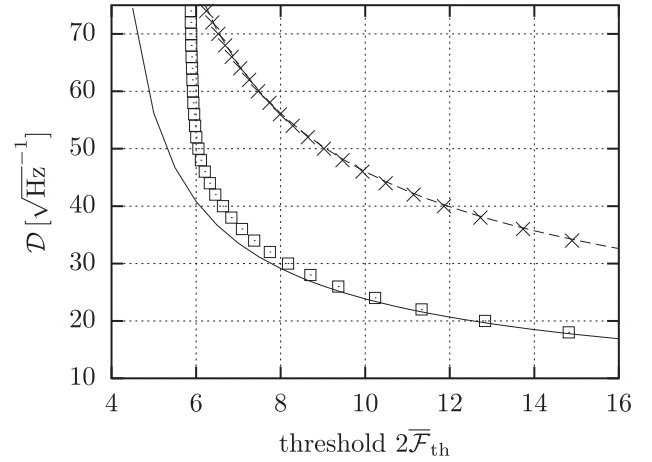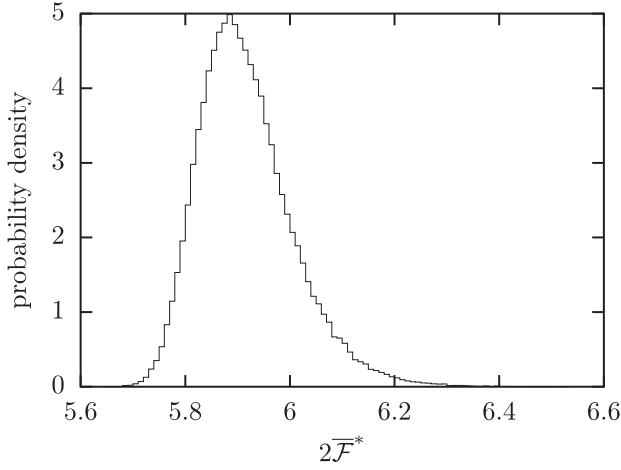
FIG. 5. Distribution of the loudest $2\overline{\mathcal{F}}$ for a box search on pure Gaussian noise, using the `S6-AllSky-StackSlide-`$\mathcal{F}$ search setup.

We note that the procedures used for *measured ULs* in CW searches typically make sure that the detection threshold has a very small false-alarm level, and we thus expect this effect to have a negligible impact in cases of practical interest.

(ii) The sensitivity estimate for wide-parameter-space searches makes the assumption that the loudest candidate $2\overline{\mathcal{F}}^*$ is always found in the *closest* template to the signal (i.e., with the smallest mismatch μ), as discussed in Sec. III B. However, while the closest template has the highest *expected* statistic value (by definition), other templates can actually produce the loudest statistic value in any given noise realization. How likely that is to happen depends on the details of the parameter space, the template bank and the threshold. Generally it is more likely at lower thresholds, as more templates further away

from the signal are given a chance to cross the threshold (despite their larger mismatch).

The *true* distribution $P(2\overline{\mathcal{F}}^*|h_0)$ of a box search will therefore be shifted to higher values compared to the approximate distribution used in Eq. (37). This implies that an actual search can have a higher detection probability than predicted by the estimate (corresponding to a larger sensitivity depth). Both of these effects contribute to different extents to the box-search discrepancy in Fig. 4 at lower thresholds:

The sampling distribution for $2\overline{\mathcal{F}}^*$ in the presence of relatively strong signals at $\mathcal{D} = 20$ Hz$^{-1/2}$ is shown in the left plot of Fig. 6, both for a simulated box search as well for the assumed distribution in the estimate. We see that most of the loudest candidates obtained in the simulation are above $2\overline{\mathcal{F}}^* > 9$, and are therefore extremely unlikely to be due to noise alone, as seen from Fig. 5. The difference between the two distributions in the left plot of Fig. 6 is therefore solely due to effect (ii). However, we see in Fig. 4 that the resulting discrepancy in the sensitivity estimate at $\mathcal{D} = 20$ Hz$^{-1/2}$ is still very small.

For weaker signals at $\mathcal{D} = 46$ Hz$^{-1/2}$, we see in the right plot of Fig. 6 that the corresponding distribution now overlaps with the pure-noise distribution of Fig. 5. The sensitivity depth therefore increasingly diverges for thresholds in the range $2\overline{\mathcal{F}}_{\text{th}} \sim [5.8, 6.1]$ due to the increasing impact of effect (i).

B. Example: Multidirected O1-MD-StackSlide-$\mathcal{F}$

In this example we use the search setup of the directed search `O1-MD-StackSlide-`$\mathcal{F}$ [30] currently running on Einstein@Home. This search consists of several directed searches for different targets on the sky, including Vela Jr. and Cas-A.
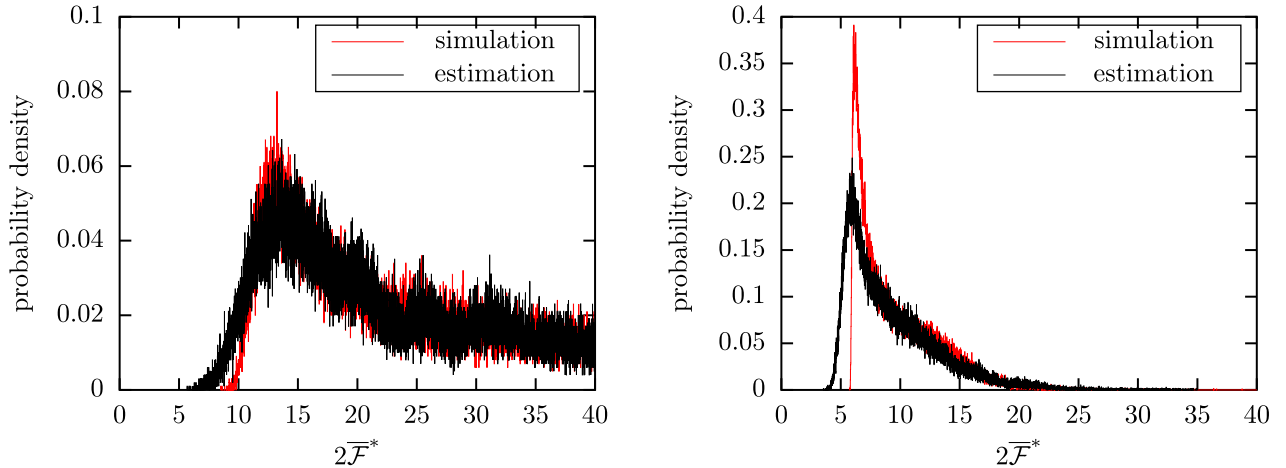

FIG. 6. Loudest $2\overline{\mathcal{F}}$ distribution for a box-search (using the `S6-AllSky-StackSlide-`$\mathcal{F}$ setup) with signals at a depth of $\mathcal{D} = 20$ Hz$^{-1/2}$ (left plot) and $\mathcal{D} = 46$ Hz$^{-1/2}$ (right plot). The black histogram shows the assumed distribution for sensitivity *estimation* in Eq. (37), and the lighter color shows the histogram obtained in a Monte-Carlo *simulation* with signals injected in Gaussian noise.
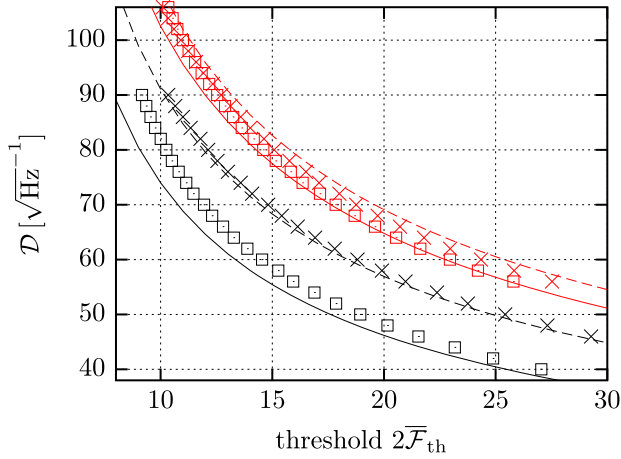
FIG. 7. Comparison of estimated and simulated sensitivity depth $\mathcal{D}^{90\%}$ as a function of the threshold $2\overline{\mathcal{F}}_{\mathrm{th}}$ for two targets of the multidirected search setup O1-MD-StackSlide-$\mathcal{F}$. The solid lines show the UL estimate for a *box search*, while the squares ($\square$) show the corresponding simulated ULs. The dashed lines indicate the estimate for the *zero-mismatch case*, and the crosses ($\times$) are for the simulated zero-mismatch ULs. The upper group of curves are for the target Vela Jr., while the lower group of curves are for Cas A.

The comparison between simulated and estimated UL depths $\mathcal{D}^{90\%}$ for these two targets is shown in Fig. 7. We see again very good agreement (relative deviations $\lesssim 3\%$) in the zero-mismatch case. However, these deviations are larger than in the all-sky case shown in Fig. 4. We suspect that this is due to the different antenna-pattern implementations of Eq. (B10) between the search code and the estimation scripts: we see different signs of the deviation for different sky positions (Vela Jr. versus Cas-A), and the effect disappears when averaging over the whole sky (as seen in Fig. 4). However, the small size of the deviations did not warrant further efforts to try to mitigate this.

For the box-search case we see good agreement at higher thresholds, with again increasing deviations at lower thresholds due to the noise effects discussed in the previous all-sky example Sec. V A.

VI. COMPARING ESTIMATES AGAINST MEASURED UPPER LIMITS

In this section we present a general overview of measured sensitivity depths $\mathcal{D}_{\mathrm{meas}}$ derived from the published upper limits of various past CW searches. For the subset of searches where an $\mathcal{F}$-statistic-based method was used (and for Bayesian targeted ULs), we provide the sensitivity estimate for comparison.

The results are summarized in Tables I–IV for the different search categories (all-sky, directed and narrowband, binary, and targeted), and more details about each search are found in Appendix A.
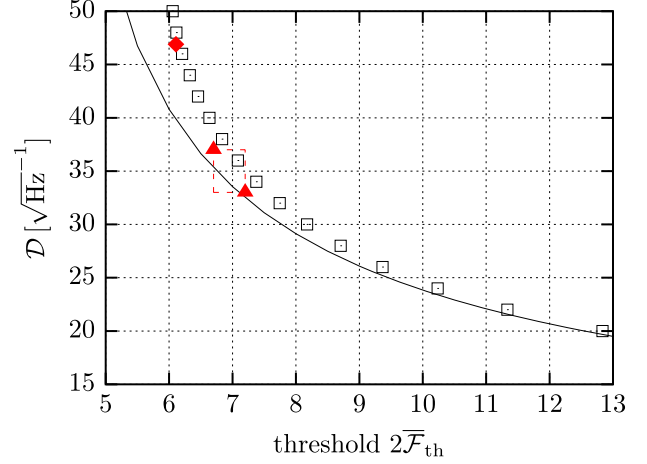

FIG. 8. Estimated (–) and simulated ($\square$) sensitivity depth versus threshold $2\overline{\mathcal{F}}_{\mathrm{th}}$ for the S6-AllSky-StackSlide-$\mathcal{F}$ (+FUP) search setup, illustrating the effect of the template-maximization in the estimate (discussed in Sec. V A). The triangles (Δ) and dashed lines show the measured upper-limit depth $\mathcal{D}_{\mathrm{meas}}^{\mathrm{med}}$ in the initial S6-AllSky-StackSlide-$\mathcal{F}$ search [56], and the diamond ($\diamond$) shows the corresponding result from the follow-up (FUP) search [49] (threshold $2\overline{\mathcal{F}}_{\mathrm{th}} = 6.1$).

A. General remarks and caveats

1. Converting published h_0 ULs into depths $\mathcal{D}$

Some searches already provide their upper limits in the form of a sensitivity depth $\mathcal{D}^{p_{\mathrm{det}}}$, but in most cases only the amplitude upper-limits $h_0^{p_{\mathrm{det}}}$ are given. For these latter cases we try to use a reasonable PSD estimate $S_{\mathrm{n}}(f)$ for the data used in the search in order to convert the quoted amplitude upper limits into sensitivity depths according to Eq. (28). This PSD estimate introduces a systematic uncertainty in the converted depth values, as in most cases we do not have access to the "original" PSD estimate used for the h_0 UL calculation.

In particular, even small differences in windowing or the type of frequency averaging can results in large differences in the PSD estimate near spectral disturbances. This can translate into large differences in the resulting converted depth values. In order to mitigate outliers due to such noise artifacts we quote the *median* over the converted measured depth values $\{\mathcal{D}_k\}$ (where k either runs over multiple frequencies, targets or detectors) and estimate the corresponding standard deviation using the *median absolute deviation* (MAD) [82], namely

$$\mathcal{D}^{\mathrm{med}} \equiv \mathrm{median}[\mathcal{D}_k],$$
$$\hat{\sigma} \equiv 1.4826\,\mathrm{median}[|\mathcal{D}_k - \mathcal{D}^{\mathrm{med}}|]. \qquad (55)$$

2. Comparing different searches by sensitivity depth $\mathcal{D}$

We can see in the Tables I–IV that searches within the same search category often show roughly comparable sensitivity depths. At one end of the spectrum are the

TABLE II.   Directed and narrow-band searches: estimated $\mathcal{D}_{\mathrm{est}}$ and measured sensitivity depth $\mathcal{D}_{\mathrm{meas}}$ (median and standard deviation, see Sec. VI A 1). The column labeled f gives the frequency range covered by each search (omitting $\dot{f}$ and $\ddot{f}$ search ranges). Sensitivity depths in *italics* refer to 90%-confidence upper limits, while normal font refers to 95%-confidence. See Appendix A 3 for further details on the individual results.

Science run	Search method	Target	f [Hz]	$\mathcal{D}_{\mathrm{est}}$ [Hz$^{-1/2}$]	$\mathcal{D}_{\mathrm{meas}}^{\mathrm{med}}$ [Hz$^{-1/2}$]	$\hat{\sigma}_{\mathcal{D}_{\mathrm{meas}}}$ [Hz$^{-1/2}$]	Ref., Sec.
earlyS5	$\mathcal{F}$	Crab	59.56 ± 0.006	221.3	223.1	…	[70], A 3 a
S5	$\mathcal{F}$	CasA	[100, 300]	35.9	35.5	0.8	[46], A 3 b
S5	StackSlide-$\mathcal{F}$	GalacticCenter	[78, 496]	*58.2*	*72.1*	4.5	[71], A 3 c
VSR4	5-vector	Vela	22.384 ± 0.02	…	100.5	…	[72], A 3 d
VSR4	5-vector	Crab	59.445 ± 0.02	…	90.1	…	[72], A 3 d
S6	$\mathcal{F}$	NineYoung (Table III)	[46, 2034]	37.8	37.7	0.3	[21], A 3 e
S6	StackSlide-$\mathcal{F}$	CasA	[50, 1000]	*79.6*	*72.9*	0.4	[22], A 3 f
S6	LooselyCoherent	OrionSpur	[50, 1500]	…	$\{30.2, 85.7\}^{\mathrm{b}}$	$\{2.3, 4.3\}$	[73], A 3 g
S6	$\mathcal{F}$	NGC6544	[92.5, 675]	29.3	29.6	1.7	[74], A 3 h
O1	5-vector	11 pulsars	$<\pm 0.1^{\mathrm{a}}$	…	111.6	12.2	[20], A 3 i
O1	Radiometer	SN1987A	[25, 1726]	…	*11.1*	4.3	[75], A 4 g
O1	Radiometer	GalacticCenter	[25, 1726]	…	*7.7*	2.9	[75], A 4 g

[a]search band around twice the pulsar spin frequency.
[b]Sensitivity depths corresponding to worst linear and circular polarization, respectively, cf. Sec. A 1.

TABLE III.   `S6-NineYoung-`$\mathcal{F}$ search: estimated $\mathcal{D}_{\mathrm{est}}$ and measured sensitivity depth $\mathcal{D}_{\mathrm{meas}}$ (median and standard deviation, see Sec. VI A 1). for nine young supernova remnants [21]. All sensitivity depths refer to 95%-confidence. See Appendix A 3 e for further details.

SN remnant	G1.9	G18.9	G93.3	G111.7	G189.1	G266.2deep	G266.2wide	G291.0	G347.3	G350.1
Name			DA 530	Cas A	IC 443	Vela Jr.	Vela Jr.	MSH 11-62		
$\mathcal{D}_{\mathrm{est}}$ [Hz$^{-1/2}$]	29.0	43.9	46.8	29.3	40.1	38.3	24.2	41.1	32.8	37.3
$\mathcal{D}_{\mathrm{meas}}^{\mathrm{med}}$ [Hz$^{-1/2}$]	28.3	44.4	49.6	31.5	39.2	40.8	26.1	44.0	32.1	36.1
$\hat{\sigma}_{\mathcal{D}_{\mathrm{meas}}}$ [Hz$^{-1/2}$]	0.8	1.3	1.5	0.9	1.2	1.0	0.7	1.2	0.8	1.1
T_{data} [10^6 s]	1.2	3.1	2.8	1.1	2.3	1.9	0.7	2.2	1.4	1.9
$2\mathcal{F}_{\mathrm{th}}$	58.0	56.3	55.6	55.6	55.3	53.7	52.8	56.6	54.1	57.6

TABLE IV.   Binary searches: measured sensitivity depth $\mathcal{D}_{\mathrm{meas}}$ (median and standard deviation, see Sec. VI A 1). All sensitivity depths refer to 95%-confidence. See Appendix A 4 for further details on the individual results.

Science run	Search method	Target	f [Hz]	$\mathcal{D}_{\mathrm{meas}}^{\mathrm{med}}$ [Hz$^{-1/2}$]	$\hat{\sigma}_{\mathcal{D}_{\mathrm{meas}}}$ [Hz$^{-1/2}$]	Ref., Sec.
S2	$\mathcal{F}$	ScoX1	[464, 484],[604, 624]	4.1	0.1	[59], A 4 a
S5	Sideband	ScoX1	[50, 550]	8.1	1.0	[76], A 4 b
S6,VSR2,3	TwoSpect	AllSky	[20, 520]	3.2	0.4	[28], A 4 c
S6,VSR2,3	TwoSpect	ScoX1	[20, 57.25]	8.2	4.0	[28], A 4 c
S6	TwoSpect	ScoX1	[40, 2040]	5.7	1.6	[77], A 4 d
S6	TwoSpect	J1751	$\{435.5, 621.5, 870.5\} \pm 1$	9.4	1.2	[77], A 4 d
O1	Viterbi	ScoX1	[60, 650]	7.6	1.0	[24], A 4 e
O1	CrossCorr	ScoX1	[25, 2000]	24.0	2.0	[23], A 4 f
O1	Radiometer	ScoX1	[25, 1726]	5.8	1.0	[75], A 4 g

fully-targeted searches, for which the parameter space (for each pulsar) is a single point, and one can achieve the maximal possible sensitivity for the available data, namely $\mathcal{D} \sim \mathcal{O}(500 \text{ Hz}^{-1/2})$ (see Table V). At the other end of the spectrum lies the all-sky binary search with a sensitivity depth of $\mathcal{D} \sim 3 \text{ Hz}^{-1/2}$ (see Table IV), which covers the largest parameter space of any search to date.

One cannot directly compare searches on sensitivity depth alone, even within the same search category. Other key aspects of a search are the parameter-space volume covered,

TABLE V. Targeted searches for known pulsars: estimated $\mathcal{D}_{\mathrm{est}}$ and measured sensitivity depth $\mathcal{D}_{\mathrm{meas}}$ (with respectively, median and standard deviation, see Sec. VI A 1). All sensitivity depths refer to 95%-confidence. See Appendix A 5 for further details on the individual results.

Science run	Search method	Targets	$\mathcal{D}_{\mathrm{est}}^{\mathrm{med}}$ [Hz$^{-1/2}$]	$\hat{\sigma}_{\mathcal{D}_{\mathrm{est}}}$ [Hz$^{-1/2}$]	$\mathcal{D}_{\mathrm{meas}}^{\mathrm{med}}$ [Hz$^{-1/2}$]	$\hat{\sigma}_{\mathcal{D}_{\mathrm{meas}}}$ [Hz$^{-1/2}$]	Ref., Sec.
S1	$\mathcal{F}$ (worst-orientation)	J1939 + 21	70.8	39.8	64.2	38.1	[32], A 5 a
S1	$\mathcal{F}$	J1939 + 21	110.4	66.7	101.8	61.8	[32], A 5 a
S1	BayesPE	J1939 + 21	81.5	19.8	85.2	14.3	[32], A 5 a
S2	BayesPE	28 pulsars	243.5	54.3	156.4	42.2	[78], A 5 b
S3,4	BayesPE	78 pulsars	337.8	81.2	299.5	79.0	[79], A 5 c
earlyS5	BayesPE	Crab	621.3	129.7	774.1	…	[70], A 5 d
S5	BayesPE	116 pulsars	997.8	210.4	932.1	317.1	[80], A 5 e
VSR2	BayesPE,$\mathcal{F}$,5-vector	Vela	351.9	78.5	408.5	20.8	[81], A 5 f
S6,VSR2,4	BayesPE,$\mathcal{F}$,5-vector	195 pulsars	555.7	116.2	514.7	171.0	[52], A 5 g
O1	BayesPE,$\mathcal{F}$,5-vector	200 pulsars	321.6	74.0	355.8	95.4	[19], A 5 h

the total computing power used, and the robustness of the search to deviations from the assumed signal- or noise-model.

Is it intuitively obvious that the more computing power spent on a fixed parameter-space volume, the more sensitive the search will tend to be, although the increase in sensitivity is typically very weak, often of order the 10th-14th root of the computing power [17].

It is also evident that the larger the parameter space covered by a search, the less sensitivity depth can be achieved due to the increased spending of computing power on "breadth" rather than depth. Ultimately the most directly relevant characteristic of a search would be its *total detection probability* [29,30], which factors in both breadth and depth as well as the underlying astrophysical prior on signal amplitudes over the parameter space searched.

B. All-sky searches

Estimated and measured sensitivity depths for all-sky searches are given in Table I, and further details about individual searches can be found in Appendix A 2.

The mean relative error between measured and estimated depths is 9%, while the median error is 7%.

One case of interest is the surprisingly large discrepancy of ∼18% observed for the S6-AllSky-StackSlide-$\mathcal{F}$+FUP search, shown in Fig. 8, were we see a significantly higher measured depth ($\mathcal{D}_{\mathrm{meas}}^{\mathrm{med}} = 46.9 \mathrm{Hz}^{-1/2}$) than estimated ($\mathcal{D}_{\mathrm{est}} = 38.3 \mathrm{Hz}^{-1/2}$). This can be traced back to the template-maximization approximation used in the estimate, namely effect (ii) discussed in Sec. V A. The low threshold used in the search ($2\overline{\mathcal{F}}_{\mathrm{th}} = 6.1$) appears to be at the cusp of becoming affected by pure-noise candidates (effect (i) in Sec. V A), but this effect is still small and does not account for the discrepancy. Furthermore, the upper limit procedure used a multi-stage follow-up, which ensures the final false-alarm level (p-value) is very small, which rules out contamination from pure-noise candidates.

C. Directed and narrow-band searches

Estimated and measured sensitivity depths for directed and narrow-band searches are given in Tables II and III, and further details about individual searches can be found in Appendix A 3.

The mean relative error between measured and estimated depths is 5%, and the median error is 1%.

For the S6-NineYoung-$\mathcal{F}$ search for nine young supernova remnants shown in Table III, the mean relative error between measured and estimated depths is 4% (median error 4%).

For two cases of interest we investigated more closely to understand the origin of the observed deviation:

S5-GalacticCenter-StackSlide-$\mathcal{F}$ search [71]: the reason for the relatively large deviation of 19% in this case between $\mathcal{D}_{\mathrm{est}} = 58.2 \mathrm{Hz}^{-1/2}$ and $\mathcal{D}_{\mathrm{meas}}^{\mathrm{med}} = 72.1 \mathrm{Hz}^{-1/2}$ can be understood by looking at the details of this search setup: contrary to the assumed uniform averaging of antenna-pattern functions over time (cf. Sec. III C, this search setup was specifically optimized by choosing the relatively short segments of $T_{\mathrm{seg}} = 11.5$ hours in such a way as to maximize sensitivity, by selecting times of maximal antenna-pattern sensitivity towards the particular sky direction of the galactic center. This is described in more detail in [45], and is quoted there as yielding a sensitivity improvement of about 20%, consistent with the observed enhancement of measured sensitivity compared to our estimate.

S6-CasA-StackSlide-$\mathcal{F}$ search [22]: the deviation between $\mathcal{D}_{\mathrm{est}} = 79.6 \mathrm{Hz}^{-1/2}$ versus $\mathcal{D}_{\mathrm{meas}}^{\mathrm{med}} = 72.9 \mathrm{Hz}^{-1/2}$ does not seem very large *per se*, but is unusual for the estimate typically does not tend to *overestimate* sensitivity by that much. A detailed investigation led us to discover a bug in the original upper-limit script used in [22], which resulted in the injection-recovery procedure to sometimes search the wrong box in parameter space, missing the
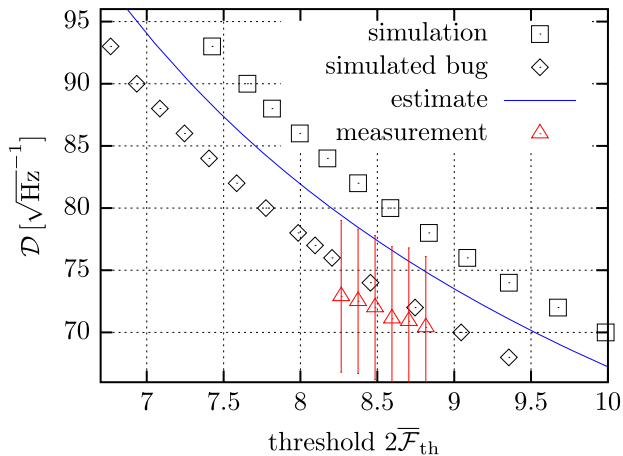
FIG. 9. Estimated (–) and simulated (□) sensitivity depth versus threshold $2\overline{\mathcal{F}}_{\rm th}$ for the `S6-CasA-StackSlide-`$\mathcal{F}$ search setup [22]. The published upper limits are plotted as triangles (△), while the diamonds (◇) show the simulated depths if we incorporate the bug found in the original UL procedure.

injected signal. By artificially reproducing the bug in our upper limit simulation we are able to confirm that this bug does account for a decrease in detection probability of about 7%, resulting in an underestimate of the upper-limit depth as shown in Fig. 9.

D. Searches for neutron stars in binaries

Estimated and measured sensitivity depths for searches for CWs from neutron stars in binary systems are given in Table IV, and further details about individual searches can be found in Appendix A 4. In this case the only $\mathcal{F}$-statistic-based search is `S2-ScoX1-`$\mathcal{F}$, for which we obtain an estimate of $\mathcal{D}_{\rm est} = 4.4 \; {\rm Hz}^{-1/2}$ (assuming an average mismatch of $\mu \sim 0.1/3$ corresponding to a cubic lattice with maximal mismatch of 0.1 [59]). The relative error is between measured and estimated sensitivity depth is therefore 8%.

E. Targeted searches for known pulsars

Estimated and measured sensitivity depths for targeted searches are given in Table V, and further details about individual searches can be found in Appendix A 5.

Note that the quoted upper limits of the BayesPE-method are obtained by Bayesian parameter-estimation [51] of $P(h_0|x)$ directly on the data x. Therefore we cannot directly apply the Bayesian sensitivity estimate derived in Sec. III F, which assumes an initial $\mathcal{F}(x)$-statistic computed on the data, from which the Bayesian upper limit would be derived. We therefore provide an approximate comparison with the *expected* sensitivity estimate, which we compute by estimating depths using $2\mathcal{F}^*$ drawn from a central χ_4^2 distribution (given each target corresponds to a single template) and averaging the resulting estimated $\mathcal{D}$ values.

In cases where several targets are covered by the search, we assume for simplicity that the targets are isotropically distributed over the sky and compute a single all-sky sensitivity estimate. For single-target searches the exact sky position is used for the estimate. The mean relative error between measured and estimated depths is 16%, and the median error is 10%.

VII. DISCUSSION

In this paper we presented a fast and accurate sensitivity-estimation framework and implementation for $\mathcal{F}$-statistic-based search methods for continuous gravitational waves, extending and generalizing an earlier analytic estimate derived by Wette [1]. In particular the new method is more direct and uses fewer approximations for single-stage StackSlide-$\mathcal{F}$ searches, and is also applicable to multistage StackSlide-$\mathcal{F}$ searches, Hough-$\mathcal{F}$ searches, and Bayesian upper limits (based on $\mathcal{F}$-statistic searches).

The typical runtime per sensitivity estimate is about 10 seconds with cached $P(R^2)$ distribution, and about 25 seconds per detector for the first call with a new parameter prior. The accuracy compared to simulated Monte-Carlo upper limits in Gaussian noise is within a few % (provided the threshold corresponds to a low false-alarm level), and we find generally good agreement (of less than $\sim$10% average error) compared to published upper limits in the literature. Several factors leading to the observed deviations in various cases are discussed in detail.

We also provided a comprehensive overview of published CW upper limit results, converting the quoted h_0 upper limits into sensitivity depths. This introduces some systematic uncertainties, as we often do not have access to the original PSD estimate used for the upper limits. We therefore advocate for future searches to directly provide their upper-limit results also in terms of the sensitivity depth of Eq. (28), in order to allow easier direct comparison between searches and to sensitivity estimates.

APPENDIX A: DETAILS ON REFERENCED CW SEARCHES

1. General remarks

In this Appendix we will refer to the different detectors as G for GEO600 [83], V for VIRGO [84,85], H1 and H2 for the two LIGO detectors in Hanford (4 km, 2 km) and L1 for LIGO Livingston [86,87].

We will use the common abbreviations CW for continuous gravitational waves, SFT for short Fourier transform, PSD for power spectral density and UL for upper limits.

The quoted sensitivity depths in Tables I–V can correspond to different confidence levels, as some searches use 90%- and others 95%-confidence upper limits. The applicable confidence level is denoted by using regular versus italic font in the tables, respectively.

For searches over many frequencies, multiple targets or for upper limits reported separately for different detectors, we use a consistent averaging procedure using the median and median absolute deviation of Eq. (55) in order to estimate the mean and standard deviation in an outlier-robust way.

PowerFlux and loosely-coherent searches typically give separate upper limits for circular (best) polarization and for the worst linear polarization, but not the more common type of population-averaged upper limits. There has been some work estimating conversion factors for these upper limits into polarization-averaged sensitivity, writing $\mathcal{D}^{\mathrm{PF}} \sim w_{\mathrm{worst}} \mathcal{D}^{\mathrm{PF}}_{\mathrm{worst}}$ and $\mathcal{D}^{\mathrm{PF}} \sim w_{\mathrm{best}} \mathcal{D}^{\mathrm{PF}}_{\mathrm{best}}$. For example [1] obtains the conversion factors in the ranges $w_{\mathrm{worst}} \sim 1.1\text{–}1.3$ and $w_{\mathrm{best}} \sim 0.39\text{–}0.46$. More recent work estimating these conversion factors on O1 data (cf. Ref. [26]) for 90%-confidence upper limits yields [88] $w_{\mathrm{worst}} = 1.51 \pm 0.13$ and $w_{\mathrm{best}} = 0.52 \pm 0.02$. However, these conversion factors were obtained by treating the set of upper limits as a whole, they should not be used to derive a proxy of population average upper limits in individual frequency bands. Furthermore, PowerFlux strict upper limits are derived by taking the highest upper limits over regions of parameter space. This procedure has the advantage of the upper limits retaining validity over any subset of parameter space, such as a particular frequency and or particular sky location. However, the maximization procedure makes it difficult to convert the data into population average upper limits which are more robust to small spikes in the data. Given that there is currently some uncertainty on the detailed values of the conversion factors to use for different PowerFlux searches, here we report the best/worst upper limits converted into sensitivity depths separately in Tables I and II.

Generally, for converting h_0 upper limits into depths according to Eq. (28), we need to use an estimate for the corresponding noise PSD S_{n}, for which we either use a corresponding PSD over the data used in the search, where available, or a 'generic' PSD estimate from LIGO for the given science run [89,90] otherwise. This adds another level of uncertainty in the conversions, which could easily be in the range 10%–20% due to different calibrations and different types of averaging over time.

2. All-sky searches, see Table I

a. S2-AllSky-Hough [58]

The first all-sky search for CWs from isolated neutron stars, using a semi-coherent Hough transform method applied on short Fourier transforms (SFTs) of the data of length $T_{\mathrm{seg}} = 30$ min. The search used data from the second LIGO Science Run (S2), and the number of SFTs used in the search was 687 from L1, 1761 from H1 and 1384 from H2.

The UL sensitivity depth for this search is calculated as the mean over the three depths for H1, L1, and H2, where each depth is computed from the respective quoted best upper-limit value $h_0^{95\%}$ and the corresponding PSD S_{n} in Table III of [58].

b. S2-AllSky-$\mathcal{F}$ [59]

A matched-filtering search based on the coherent (single-detector) $\mathcal{F}$-statistics, using 20 SFTs from H1 and 20 SFTs from L1 (SFT length $T_{\mathrm{SFT}} = 30$ min). The per-detector $\mathcal{F}$-statistic values were combined via a coincidence scheme, determining the most significant candidate in each ~ 1 Hz band, which was then used for measuring the upper limits.

The sensitivity depth for this search is calculated from the given (combined multidetector) upper limits $h_0^{95\%}(f)$ over the search frequency range, combined with the harmonic mean over generic H1- and L1- PSDs for the LIGO S2 data.

The estimate was calculated with the mean loudest templates of the search given in the paper as $\mathcal{F}_{\mathrm{th}} = (39.5, 32.2)$ for the L1 and H1 detector, respectively, and we used an average mismatch of 0.5% in the H1 search and 1% in the L1 search, estimated from Figs. 27,28 in [59].

c. S4-AllSky-{**StackSlide**,**Hough**,**PowerFlux**} [35]

Three semicoherent all-sky searches using different search methods, all based on incoherently combining SFTs of length $T_{\mathrm{seg}} = 30$ min. The StackSlide and the Hough search used 1004 SFTs from H1 and 899 from L1 and the Hough search additionally included 1063 SFTs from H2. The PowerFlux search used 1925 and 1628 SFTs from H1 and L1, respectively.

The sensitivity depths are calculated from the quoted upper limits $h_0^{95\%}(f)$ from each of the three searches over the search frequency range, combined with the PSDs for two (H1 and L1) detectors (as a common reference) from the S4 science run. Note that the Hough depth corresponds to the quoted multidetector UL, while the other searches reported only per-detector ULs.

d. S4-AllSky-$\mathcal{F}$ + Coinc [60]

A search which used the distributed computing project Einstein@Home [43] to analyse 300 h of H1 data and 210 h of L1 data from the S4 run. The data was split into 30 h long segments coherently analysed with the multidetector $\mathcal{F}$-statistic followed by a coincidence-step. The measured sensitivity depth $\mathcal{D}_{\mathrm{meas}}^{90\%}$ is calculated by converting the quoted sensitivity factors $R_{90\%} = \{31.8, 33.2\}$ (for frequencies below and above 300 Hz, respectively) into sensitivity depths. However, given these were computed with respect to an (arithmetic) averaged PSD estimate (given in Fig. 1 in the paper), we first converted these factors back into equivalent h_0 values using the mean-PSD, and then computed the Depth with respect to the harmonic-mean (over detectors) generic noise PSD for S4.

e. earlyS5-AllSky-PowerFlux [61]

An all-sky search with PowerFlux over the first eight months of S5 data. The search in total used roughly 4077 h of H1 data and 3070 h L1 data, divided into SFT segments of $T_{\mathrm{seg}} = 30$ min.

The sensitivity depth is calculated from the quoted per-detector upper limits $h_0^{95\%}(f)$ over the search frequency range and the corresponding S5 noise PSDs.

f. earlyS5-AllSky-$\mathcal{F}$ + Coinc [62]

An all-sky search run on Einstein@Home [43], using 660 h of data from H1 and 180 h of L1 data, taken from the first 66 days of the LIGO S5 science run. The data was divided into 28 segments of $T_{\mathrm{seg}} = 30$ h duration, and each segment was searched using the fully coherent multi-detector $\mathcal{F}$-statistic. These per-segment $\mathcal{F}$-statistics were combined across segments using a coincidence scheme.

The measured sensitivity depth $\mathcal{D}_{\mathrm{meas}}^{90\%}$ is calculated as the median over the converted sensitivity depths converted from the quoted sensitivity factors $R_{90\%} = \{29.4, 30.3\}$ in the paper for the frequencies below and above 400 Hz, respectively.

g. S5-AllSky-PowerFlux [63]

An all-sky search using PowerFlux analyzing the whole of LIGO S5 data, broken into more than 80000 50%-overlapping 30-minute SFTs from both H1 and L1.

The sensitivity depth is calculated from the quoted upper limits $h_0^{95\%}$ and the S5 noise PSD.

h. S5-AllSky-Hough-$\mathcal{F}$ [36]

An all-sky search using the Hough-$\mathcal{F}$ variant of the semi-coherent Hough method described in Sec. II C 2, which was run on Einstein@Home. The analyzed data consisted of 5550 and 5010 SFTs from the LIGO H1 and L1 interferometers, respectively, taken from the second year of the S5 science run. The data was divided into 121 segments of length $T_{\mathrm{seg}} = 25$ h, and the coherent per-segment $\mathcal{F}$-statistic was combined via the Hough method to compute the Hough number count of Eq. (25).

The sensitivity depth of the search is calculated from the quoted $h_0^{90\%}$ upper limits and the corresponding S5 noise PSD.

The estimated sensitivity depth uses the generalization of the estimator described in Sec. III E with a number-count threshold of $n_{\mathrm{c,th}} = 70$, a per segment threshold of $\tilde{\mathcal{F}}_{\mathrm{th}} = 2.6$ and a mismatch histogram obtained from an injection-recovery simulation (with an average mismatch of $\tilde{\mu} = 0.61$).

i. S5-AllSky-Hough [64]

An SFT-based Hough all-sky search on S5 data. The search was split into the first and the second year of S5, which were searched separately. The first year used 11402 SFTs from H1, 12195 SFTs from H2 and 8698 SFTs from L1, of length $T_{\mathrm{SFT}} = 30$ min. The analysis of the second year used 12590 H1-SFTs, 12178 H2-SFTs and 10633 L1-SFTs.

The sensitivity depth is calculated from the quoted $h_0^{90\%}$ upper limits of the second year search found in the paper and from the S5 noise PSD.

j. S5-AllSky-StackSlide-$\mathcal{F}$ [65]

A high frequency all-sky search to complement previous lower-frequency all-sky searches on S5 data. The search used the so-called GCT method [57] implementing the StackSlide-$\mathcal{F}$ statistic and was run on the distributed Einstein@Home platform. The search used a total of 17797 SFTs spanning the whole two years of S5 data from H1 and L1, divided into 205 segments of length $T_{\mathrm{seg}} = 30$ h.

The measured sensitivity depth $\mathcal{D}_{\mathrm{meas}}^{90\%}$ is determined by extrapolating the depth values given in the paper for critical ratios of 0 and 3.5 to the median critical ratio over all frequency bands of -0.15 according to Fig. 6 of [65].

For the estimate we determined the median threshold over all frequency bands from Fig. 4 of [65] to $2\overline{\mathcal{F}}_{\mathrm{th}} = 5.72$. Two mismatch histograms at 1255 Hz and 1495 Hz generated with injection-recovery studies were used. The average mismatch for both was $\mu \approx 0.82$. The quoted value is the mean of the two estimates with different mismatch histograms.

k. VSR1-AllSky-$\mathcal{F}_{\mathrm{TD}}$ + Coinc [66]

An all-sky search using data from the first Virgo science run, VSR1. The search method uses a time-domain implementation of the coherent $\mathcal{F}$-statistic, computed over 2-day coherent segments, which are combined using coincidences. In total the search used 134 days of data.

The measured sensitivity depth $\mathcal{D}_{\mathrm{meas}}^{90\%}$ is calculated as median of the given sensitivity factors of 15.6 and 22.4.

l. {VSR2,4}-AllSky-FreqHough + FUP [67]

This all-sky search was performed using data from initial Virgos second (VSR2) and forth (VSR4) science run. It used the FrequencyHough transform as incoherent step with 149 days of data of VSR2 and 476 days of data of VSR4 using segments of length 8192 seconds. The initial candidates were followed-up using 10 times longer segments.

The measured sensitivity depth was calculated from upper limits $h_0^{90\%}$ extracted from Fig. 12 of [67] and the harmonic mean of the PSD estimates of VSR2 and VSR4 in 0.1 Hz frequency bands.

m. S6-AllSky-StackSlide-$\mathcal{F}$ [56]

This search used 12080 SFTs from L1 and H1 data to perform a StackSlide-$\mathcal{F}$ search based on the GCT implementation, and was run on Einstein@Home. The search used 90 coherent segments of length $T_{seg} = 60$ h.

The measured sensitivity depth $\mathcal{D}_{meas}^{90\%}$ is determined by extrapolating the depth from the given critical ratios 0 and 6 to the median critical ratio of -0.07 according to Fig. 5 of [56].

The estimated depth is given for a threshold of $2\overline{\mathcal{F}}_{th} = 6.694$ which is the median of the thresholds given for the frequency bands in Fig. 4 of [56]. For the estimate two mismatch histogram created with injection-recovery studies for 55 and 505 Hz was used. The average mismatch of the grid in the parameter space was at both frequencies found to be $\mu = 0.72$. The quoted value is the mean of the two estimates with different mismatch histograms.

n. S6-AllSky-StackSlide-$\mathcal{F}$ + FUP [49]

A multistage follow-up on candidates from the S6-AllSky-StackSlide-$\mathcal{F}$ search described in the previous paragraph, zooming in on candidates using increasingly finer grid resolution and longer segments. Every candidate from the initial stage with $2\overline{\mathcal{F}} \geq 6.109$ was used as the center of a new search box for the first-stage follow-up, continuing for a total of four semi-coherent follow-up stages. The sensitivity of the search is dominated by the initial-stage threshold, because the later stages are designed to have a very low probability of dismissing a real signal. The measured sensitivity depth $\mathcal{D}_{meas}^{90\%} = 46.9$ Hz$^{-1/2}$ of this search is directly taken from the quoted value in the paper.

The estimated multistage sensitivity of Sec. III D using the thresholds given in the paper, namely $\{2\overline{\mathcal{F}}_{th}^{(i)}\} = (6.109, 6.109, 7.38, 8.82, 15)$ and a mismatch histogram generated by recovery injection studies for the main search and mismatch histograms provided by the original authors for every stage with average mismatches $\{\mu^{(i)}\} = (0.72, 0.55, 0.54, 0.29, 0.14)$, yields a value of $\mathcal{D}^{90\%} = 38.3$ Hz$^{-1/2}$, which differs significantly from the quoted measured sensitivity depth. As discussed in Sec. V, we trace this discrepancy to the low threshold used, which significantly affects the loudest-candidate mismatch approximation used in the theoretical estimate.

o. S6-AllSky-PowerFlux [68]

The data used by this search span a time of 232.5 d with duty factor of the detectors of 53% for H1 and 51% for L1.

The measured sensitivity depth is calculated from the quoted upper limits $h_0^{95\%}$ in the paper and the S6 noise PSD.

p. O1-AllSky-StackSlide-$\mathcal{F}$ [26]

A low-frequency all-sky search for gravitational waves from isolated neutron stars using the distributed computing project Einstein@Home on data from Advanced LIGO's first observing run (O1). This search used the GCT implementation of the semi-coherent StackSlide-$\mathcal{F}$ method with $N_{seg} = 12$ segments of length $T_{seg} = 210$ h in the initial search stage. The analyzed data consisted of 4744 SFTs from the H1 and the L1 detector. The search also included a hierarchical follow-up similar to the S6Bucket follow-up search [49].

The measured sensitivity depth $\mathcal{D}_{meas}^{90\%} = 48.7$ Hz$^{-1/2}$ of this search is directly taken from the quoted value in the paper.

The sensitivity estimate used a threshold $2\overline{\mathcal{F}}_{th} = 14.5$ which we inferred from Fig. 4 in [26] and we obtained the mismatch histograms of the template grid at different frequencies using an injection-recovery study, which yielded an average mismatch of $\mu = 0.35$ and $\mu = 0.37$ at 20 Hz and 100 Hz respectively. The quoted depth is the average of the two different estimates resulting for each mismatch histogram. Note that the contrary to the measured sensitivity, the estimate only uses the first-stage parameters in this case, as we currently cannot model the line-robust statistic used in the follow-up stages. However, as mentioned in Sec. III D, the overall detection probability is dominated by the first stage, while subsequent stages mostly serve to reduce the false-alarm level.

q. O1-AllSky-{PowerFlux,Hough,$\mathcal{F}_{TD}$ + Coinc} [25,69]

Two papers detailing the results of all-sky searches on O1 data using four different search methods.

The first paper [25] searched the lower frequency range [20,475] Hz, using four methods: PowerFlux, FrequencyHough, SkyHough, and a time-domain $\mathcal{F}$-statistic search with segment-coincidences (denoted as $\mathcal{F}_{TD}$ + Coinc). The PowerFlux, FrequencyHough, and SkyHough search used SFT lengths in the range 1800–7200 s as coherent segments while the time-domain $\mathcal{F}$-statistic used a coherence time of $T_{seg} = 6$ d. The total amount of analyzed data was about 77 d of H1 data and 66 d of L1 data.

In the second paper [69] three of these searches were extended up to 2000 Hz, namely PowerFlux, SkyHough, and a time-domain $\mathcal{F}$-statistic search with segment-coincidences (denoted as $\mathcal{F}_{\mathrm{TD}} + \mathrm{Coinc}$), using the same data.

The sensitivity depths for the four searches are calculated from the quoted $h_0^{95\%}$ amplitude upper limits and the noise PSD for the O1 science run.

Note that for the SkyHough method a sensitivity depth of 24.2 Hz$^{-1/2}$ is quoted in the paper. However, this value is based on a slightly different convention for the multi-detector noise PSD S_{n} (maximum over detectors instead of the harmonic mean) than used here. For consistency with the other searches in Table I we therefore compute the sensitivity depth by converting from the quoted $h_0^{95\%}$ upper limits instead.

A comparison of *PowerFlux* 90%-confidence upper limits for an *isotropic polarization* population were provided for the O1 Einstein@Home paper [26], with a frequency spacing of 0.0625 Hz, which are converted into sensitivity depth using the O1 noise PSD.

3. Directed searches, see Tables II, III

a. earlyS5-Crab-$\mathcal{F}$ [70]

This search aimed at the Crab pulsar and used the first nine month of initial LIGO's fifth science run (S5). It consisted of both a targeted (described in Sec. A 5 d) and a directed $\mathcal{F}$-statistic search described here. The directed search used 182, 206 and 141 days of data from the H1, H2, and L1 LIGO detectors, respectively. The measured depth value is calculated from the given upper limits $h_0^{95\%}$ and the PSD estimate of the S5 data at the search frequency.

The estimated depth uses the StackSlide estimator for a coherent search with $N_{\mathrm{seg}} = 1$ segment, a threshold of $\mathcal{F}_{\mathrm{th}} = 37$ and a maximal template bank mismatch of 5% (given in the paper), from which we estimate the average mismatch as $\tilde{\mu} \sim \frac{1}{3} 5\%$ (assuming a square lattice).

b. S5-CasA-$\mathcal{F}$ [46,91]

The first search for continuous gravitational waves from the Cassiopeia A supernova remnant using data from initial LIGO's fifth science run (S5). The search coherently analyzed data in an interval of 12 days (934 SFTs of length 30 min) using the $\mathcal{F}$-statistic.

The measured sensitivity depth is obtained from the quoted upper limits $h_0^{95\%}$ in the paper and the S5 noise PSD.

The estimate is calculated using the StackSlide estimator for a coherent search ($N_{\mathrm{seg}} = 1$ segment), with the mismatch histogram for an A_n^* lattice with maximal mismatch of $\mu = 0.2$ (obtained from LATTICEMISMATCHHIST() in [54]), and the average threshold of $2\mathcal{F}_{\mathrm{th}} = 55.8$ (averaged over the respective loudest $2\mathcal{F}$-candidates found in each of the upper-limit bands).

c. S5-GalacticCenter-StackSlide-$\mathcal{F}$ [45,71]

The first search for continuous gravitational waves directed at the Galactic center. The search used LIGO S5 data and the GCT implementation of the StackSlide-$\mathcal{F}$ semi-coherent search algorithm with 630 segments, each spanning 11.5h, for total data set of 21463 SFTs of length 30 min.

The segments of the search were selected from the whole S5 science run in such a way as to maximize the SNR for fixed-strength GW signals at the skyposition of the galactic center. Therefore the selected segments fall at times where the antenna patterns of the LIGO detectors are better than average for this particular sky position. As discussed in Sec. VI C, the sensitivity-estimation method presented in this work assumes the antenna patterns are averaged over multiple days, which causes a unusually large deviation between the estimate and the measured sensitivity depth from the $h_0^{90\%}$ upper limits.

The estimate is calculated using the mismatch histogram (with mean $\mu = 0.13$) obtained from an injection-recovery study on the template bank of this search, and a detection threshold of $2\overline{\mathcal{F}}_{\mathrm{th}} = 4.77$.

d. VSR4-{Vela,Crab}-5-vector [72]

This coherent narrow-band search on the data from initial Virgo's forth science run (VSR4) was directed at the Vela and the Crab pulsars. This search used the 5-vector method, and covers a range of ± 0.02 Hz the twice the known frequencies of Vela and Crab. The total amount of data used is 76 d.

The measured sensitivity depth for this search was obtained from the published $h_0^{95\%}$ upper limits and the noise PSD estimate for VSR4.

e. S6-NineYoung-$\mathcal{F}$ [21]

This search was directed at nine different targets, listed in Table III, each corresponding to a (confirmed or suspected) compact object in a young supernova remnant. The search uses a fully coherent $\mathcal{F}$-statistic. The amount of data used for every target varies between 7.3×10^5 s and 3.1×10^6 s (cf. Table III).

The measured depth is calculated for each of the targets from the quoted upper limits $h_0^{95\%}$ and the corresponding PSD for the actual data used in the search.

The estimate for each target is calculated using the StackSlide estimator for a coherent search ($N_{\mathrm{seg}} = 1$ segment), with the mismatch histogram for an A_n^* lattice with maximal mismatch of $\mu = 0.2$ (obtained from LATTICEMISMATCHHIST() in [54]), and the average $2\mathcal{F}_{\mathrm{th}}$ threshold found for each target (averaged over the respective loudest $2\mathcal{F}$-candidates found in each of the upper-limit bands) are given in Table III.

The "NineYoung" entry in Table II presents the median depth over all targets for the measured and estimated depths, respectively.

f. S6-CasA-StackSlide-$\mathcal{F}$ [22]

A search directed at Cassiopeia A, which was run on the distributed computing project Einstein@Home using data from the LIGO S6 science run. The search was based on the GCT implementation of the semi-coherent StackSlide-$\mathcal{F}$ statistic, with $N_{\text{seg}} = 44$ segments of length $T_{\text{seg}} = 140$ h, and a total amount of data of 13143 SFTs of length 30 min from the two LIGO detectors in Hanford (H1) and Livingston (L1). The measured sensitivity depth given in Table II is computed from the $h_0^{90\%}$ upper limits quoted the paper [22] combined with the corresponding PSD estimates. However, as discussed in VI C, this measurement suffered from a bug in the upper-limit script and as a result is somewhat too conservative (i.e., too high).

The estimated sensitivity is calculated assuming an average threshold of $\overline{\mathcal{F}}_{\text{th}} = 8.25$ (estimated from Fig. 4 in [22]) using the mean over estimates with different mismatch histograms generated by injection-recovery studies at different frequencies (spanning 50–1000 Hz, average mismatch ∼9%).

g. S6-OrionSpur-LooselyCoherent [73]

This was a search employing the so-called loosely-coherent method, aimed at the Orion spur towards both the inner and outer regions of our Galaxy. The explored sky regions are disks with 6.87° diameter around $20^{\text{h}}10^{\text{m}}54.71^{\text{s}} + 33°33'25.29''$ and 7.45° diameter around $8^{\text{h}}35^{\text{m}}20.61^{\text{s}} - 46°49'25.151''$. The data used in this search spanned 20 085 802 s with duty factors of 53% and 51% for LIGO Hanford and Livingston, respectively. Due to weighting of the data the effective amount of data used was only ∼12.5% of the available S6 data. For the analysis data segments of length 30 min were searched coherently.

The measured sensitivity depth was calculated from the quoted upper limits $h_0^{95\%}$ and a PSD estimate for the LIGO S6 data.

h. S6-NGC6544-$\mathcal{F}$ [74]

This was the first search directed at the nearby globular cluster NGC 6544. The search coherently analyzed data from the two LIGO detectors S6 science run with the $\mathcal{F}$-statistic, using a single coherent segment with $T_{\text{seg}} = 9.2$ d. The search analyzed two different data stretches separately. The first one contained 374 SFTs while the second contained 642 SFTs, with SFT duration of 30 min.

The measured depth was determined from the upper limits $h_0^{95\%}$ given in Fig. 2 of [74] and a PSD estimate for the LIGO S6 run.

The estimate used the StackSlide estimator with one segment, a threshold of $2\mathcal{F}_{\text{th}} = 55$ (quoted in the paper) and an average mismatch of 0.2/3 (assuming a roughly square lattice).

i. O1-Narrow-band-5-vector [20]

A narrow-band search aiming at 11 known pulsars using the fully-coherent 5-vector method on data from Advanced LIGO's first observing run (O1). The search used a total of 121 days of data from the Hanford (H1) and Livingston (L1) detectors.

The sensitivity depth in the table is calculated from the median over the single-target depths, which are converted from the upper-limits $h^{95\%}$ quoted in the paper and the corresponding noise PSD of the data used.

j. O1-{SN1987,GalacticCenter}-Radiometer [75]

Described in Sec. A 4 g.

4. Searches for neutron stars in binary systems, see Table IV

a. S2-ScoX1-$\mathcal{F}$ [59]

This first search designed specifically aimed at the NS in the LMXB system Scorpius X-1, using a coherent single-detector $\mathcal{F}$-statistic and a coincidence check on a 6 h long stretch of S2 dat.

The measured sensitivity depth was calculated from the quoted upper limits $h_0^{95\%}$ in the paper (for the zero-eccentricity case $e = 0$) and the PSD estimate of the corresponding S2 data.

b. S5-ScoX1-Sideband [76]

A search aimed at Scorpius X-1 by incoherently combining sidebands of a coherent $\mathcal{F}$-statistic search that only demodulates the signal for the sky-position but not its binary-orbital Doppler modulation. This method used a stretch of 10 days of data selected from the S5 science run for maximal sensitivity. Two searches were performed, one with no prior assumptions about the orientation of Sco-X1, and one using more restrictive angle-priors based on electromagnetic observations.

Bayesian upper limits $h_0^{95\%}$ were computed over the search frequency range, which we convert into sensitivity depths (for the unknown-polarization case, see Fig. 5(a) in [76]) using the noise PSD for the data given in the paper. In each 1 Hz-band, 2×10^6 upper limit values were quoted, of which we use the maximum value in each 1 Hz-band in order to be consistent with the usual "loudest-candidate" approach of setting upper limits in a given frequency band.

c. {S6,VSR2,3}-{AllSky,ScoX1}-TwoSpect [28]

A TwoSpect search for unknown binary signals from any sky-position, and a directed TwoSpect search for Scorpius X-1 specifically. This search used data from LIGO S6 science run, as well as from Virgo VSR2 and VSR3 runs, spanning 40 551 300 s from each detector.

The quoted upper limits $h_0^{95\%}$ for the all-sky search and the Scorpius X-1 search were converted into Depths using a

combined (generic) PSD for the S6, VSR2 and VSR3 science runs.

d. S6-{ScoX1,J1751}-TwoSpect [77]

A search for CW from the low-mass X-ray binaries Scorpius X-1 and XTE J1751-305 using the TwoSpect algorithm. It used about 4×10^7 s from each of the two detector in the S6 science run. It used two different length of the SFTs 840 and 360 s which also where the length of the coherently analysed segments.

The given sensitivity depth $\mathcal{D}_0^{95\%}$ is obtained from the quoted $h_0^{95\%}$ upper limits combined with the corresponding noise PSD for S6 data.

e. O1-ScoX1-Viterbi [24]

A search aimed at Scorpius X-1 using the Viterbi search method performed on 130 days of data from Advanced LIGO's first observational run (O1), segmented into coherent segments of length $T_{\rm seg} = 10$ days.

The measured sensitivity depth is converted from the quoted upper limits $h_0^{95\%}$ (for unknown polarization) and the noise PSD of the corresponding O1 data.

Note that contrary to many other search methods, this search setup appears to result in a frequency-dependent sensitivity depth, namely $\mathcal{D}(f) \propto f^{-1/4}$ (see Eq. (9) in [24]). For consistency with other searches, we quote the median and (MAD) standard-deviation over frequencies in Table IV, and note that the total range of sensitivity depths of this search is found as $\mathcal{D}(f) \sim 11(f/f_0)^{-1/4}$ Hz$^{-1/2}$ $\in$ [4.6, 11.2] Hz$^{-1/2}$ with $f_0 = 60.5$ Hz.

f. O1-ScoX1-CrossCorr [23]

This search aimed at Scorpius X-1 using the CrossCorr search algorithm using data from Advanced LIGO's first observational run (O1). The data was split into coherently analyzed segments (SFTs) with a (frequency-dependent) length between 240 and 1400 s.

The measured sensitivity depth is obtained from the quoted (isotropic-prior) upper limits $h_0^{95\%}$ and the noise PSD of the O1 data. Note, however, that the search ULs are given per 0.05 Hz bands, which is unusually small compared to most other upper-limit bands (typically 0.25–1 Hz), and therefore they display more variability. In order to make these ULs more comparable to other searches, we use the 95th-percentile highest upper limits per 1 Hz-bands (as recommended in Fig. 5 of [23]). This "binning" procedure only has a small effect on the resulting sensitivity depth, which is reduced from 25.3 Hz$^{-1/2}$ to 24.0 Hz$^{-1/2}$.

Note that this search has a frequency-dependent sensitivity depth, which starts at around $\mathcal{D}(25 \text{ Hz}) \sim 45$ Hz$^{-1/2}$ for low frequencies, asymptoting down to $\mathcal{D} \sim 23$ Hz$^{-1/2}$ above $f \gtrsim 800$ Hz. However, in order to be consistent with other searches, we quote the median and (MAD) standard deviation over all frequencies in Table. IV.

g. O1-{ScoX1 and others}-Radiometer [75]

The "radiometer" search method, which was developed mainly for stochastic background searches, can also be used for directed CW searches at particular sky-positions. This method does not use a particular signal model, which allows it to be sensitive to a wide range of possible signal families, at the cost of somewhat lower sensitivity to "regular" CW signals. This search aimed at the sky-positions of Sco-X1, as well as at the supernova remnant 1987A and the Galactic center.

The search reported $h_0^{90\%}$ (and $h_0^{95\%}$ for Sco-X1, reported in [23]) upper limits in narrow frequency bands of $1/32$ Hz $= 0.03125$ Hz bands, which is unusually small compared to most other upper-limit bands (typically 0.25–1 Hz), and therefore they display more variability. In order to make these ULs more comparable to other searches, we use the 95th-percentile highest upper limits per 1 Hz-bands (as recommended in Fig. 5 of [23]), and following the same procedure as used for the CrossCorr results (discussed in Sec. A 4 f).

5. Targeted searches, see Table V
a. S1-J1939+21-{$\mathcal{F}$,BayesPE} [32]

This first CW search on data from GEO 600 and LIGO's first science run (S1). It used (16.7,5.73,8.73,8.9) days of data from four detectors, GEO 600 (G1), LIGO Livingston (L1), LIGO Hanford-4 km (H1), and LIGO Hanford-2 km (H2), respectively. Two types of searches were performed, a coherent $\mathcal{F}$-statistic search as well as direct Bayesian parameter estimation (BayesPE).

Table V gives the mean and standard deviation for the sensitivity depths over the four detectors. The measured sensitivity depth for the $\mathcal{F}$-search was determined from the quoted upper limits $h_0^{95\%}$ in Table IV [32] for the most pessimistic ι ($\cos \iota = 0$) and ψ, and from the quoted numbers in the conclusion for the (standard) population-averaged orientation. The noise PSD values are taken from Table III in [32]. The corresponding estimate is calculated with the StackSlide estimator for $N_{\rm seg} = 1$ and quoted threshold values $2\mathcal{F}_{\rm th} = (1.5, 3.6, 6.0, 3.4)$ for the four detectors from Table III in the paper. For the "worst-case" estimate we use the prior $\cos \iota = 0$ and minimise the sensitivity depth over $\psi \in [-\pi/4, \pi/4]$ in order to reflect the "conservative" ULs quoted in the paper. Note, however, that contrary to the typically small false-alarm level (p-value) of the UL thresholds used (typically 1%), the loudest candidates used here as thresholds here had relatively high p-values of 83%, 46%, 20%, and 49%, respectively, as seen in Table III of [32].

b. S2-Known pulsars-BayesPE [78]

A coherent targeted search for 28 known isolated radio pulsars was performed using the Bayesian parameter-estimation pipline (BayesPE) on data from the second LIGO Science Run (S2), using 910 h of data from H1, 691 h from H2 and 342 h of L1 data from the S2 data set.

The measured sensitivity depth is calculated from the quoted Bayesian upper limits $h_0^{95\%}$ and corresponding noise PSD estimates for the S2 science run.

The sensitivity estimate is performed using the Bayesian sensitivity estimator, for simplicity assuming the sources are distributed isotropically over the sky.

c. {S3,4}-Known pulsars-BayesPE [79]

This search targeted 78 known radio pulsars by analysing (45.5,42.1,13.4) days of data from the three detectors (H1, H2, L1) from the third science run (S3) of LIGO and GEO 600, and (19.4,22.5,17.1) days of data from the three detectors from the S4 science run. The analysis used the Bayesian parameter-estimation pipeline (BayesPE).

The measured sensitivity depth was determined from the quoted Bayesian upper limits $h_0^{95\%}$ combined with the noise PSD of the S3 and S4 science runs combined (using harmonic mean).

The sensitivity estimate is calculated using the Bayesian sensitivity estimate, for simplicity assuming the sources to be isotropically distributed on the sky.

d. earlyS5-Crab-BayesPE [70]

This search on 9 months of data from the early LIGO S5 science run targeted only the Crab pulsar at twice its rotation rate, using the Bayesian parameter-estimation pipeline. A corresponding narrow-band search using the $\mathcal{F}$-statistic is described in Sec. A 3 a. The targeted search used 201, 222, and 158 days of data of the H1, H2, and L1 LIGO detectors.

The measured depth is determined from the quoted (i.e., the corrected value in the Erratum) upper limit $h_0^{95\%}$ assuming an isotropic polarization prior, and the corresponding noise PSD of the detectors for the early S5 science run data.

e. S5-Known pulsars-BayesPE [80]

A search targeting 116 known pulsars using 525 days of H1 data, 532 days of H2 data, and 437 days of L1 data from LIGO's fifth science run (S5). The search employed the Bayesian parameter-estimation pipeline.

The measured sensitivity depth is calculated from the quoted Bayesian upper limits $h_0^{95\%}$ and the noise PSD of the S5 data.

The estimate is calculated with the Bayesian sensitivity estimator under the assumption that the targets are distributed isotropically over the sky.

f. VSR2-Vela-{BayesPE,$\mathcal{F}$,5-vector} [81]

A targeted search for the Vela pulsar using Virgo's second science-run (VSR2) data, using three different methods: Bayesian parameter estimation, the $\mathcal{F}$-statistic (and $\mathcal{G}$-statistic) and the 5-vector method. The data set consisted of 149 days of Virgo data.

Two types of searches and upper limits were computed, namely (i) using uninformative (isotropic) priors on the pulsar orientation, and (ii) using angle priors on $\cos \iota$ and ψ from electromagnetic observations.

In Table V we only give the measured depth corresponding to the isotropic prior, averaged over the three methods, which obtained very similar results. This was computed from the quoted upper limits $h_0^{95\%}$ and the noise PSD for the Vela VSR2 run. The measured sensitivity depth obtained when using the angle priors is found as 462.1 ± 35.0 Hz$^{-1/2}$.

The estimated sensitivity depth is calculated using the Bayesian sensitivity estimator.

g. {S6,VSR2,4}-Known pulsars-{BayesPE,$\mathcal{F}$,5-vector} [52]

This search targeted 195 known pulsars, using 149 days of VSR2 and 76 days of VSR4 data for pulsars with a CW frequency lower than $f < 40$ Hz and an additional 238 days of S6 data from H1 and 225 days from L1 for faster spinning pulsars with $f > 40$ Hz. The analysis was done using three different methods: Bayesian parameter estimation, the $\mathcal{F}$-statistic (or $\mathcal{G}$-statistic for restricted angle priors), and the 5-vector method.

The given measured sensitivity depth in Table V is the median and MAD standard deviation over the sensitivity depths for the different targets (averaged over high- and low-frequency targets). The sensitivity depths are obtained from the quoted upper limits $h_0^{95\%}$ and the corresponding noise PSD estimate of the data used (which is either S6 and VSR2 and VSR4 for high-frequency targets $f > 40$ Hz, or only VSR2 and VSR4 for low-frequency targets).

The estimated sensitivity is obtained from the Bayesian sensitivity estimator assuming an isotropic prior over the sky, averaged over high- and low-frequency depths results.

h. O1-Known pulsars-{BayesPE,$\mathcal{F}$,5-vector} [19]

In this search 200 known pulsars were targeted using three different methods: Bayesian parameter estimation, the $\mathcal{F}$-statistic (or $\mathcal{G}$-statistic for restricted angle priors), and the 5-vector method. The searches used 78 and 66 days of H1 and O1 data from the first observational run of advanced LIGO (O1), respectively.

The measured sensitivity depth is obtained from the quoted Bayesian upper limits $h_0^{95\%}$ over all targets and the corresponding noise PSD for the LIGO detectors during O1.

The estimated sensitivity depth is determined from the Bayesian estimator as an all-sky estimate assuming the targets are isotropically uniformly distributed over the sky.

APPENDIX B: CW SIGNAL MODEL AND $\mathcal{F}$-STATISTIC

A plane gravitational wave arriving from a direction $\hat{n}$ (unit vector) can be written [92] in TT gauge (in the notation of [93]) as a purely spatial strain tensor $\overleftrightarrow{h}$ with two polarizations $+, \times$, namely

$$\overleftrightarrow{h}(\tau) = h_+(\tau)\overleftrightarrow{e}_+ + h_\times(\tau)\overleftrightarrow{e}_\times, \tag{B1}$$

where τ is the emission time of the signal in the source frame, and $\overleftrightarrow{e}_+$ and $\overleftrightarrow{e}_\times$ are the two polarization basis tensors, which can be constructed from a right-handed orthonormal basis $\{\hat{\ell}, \hat{m}, -\hat{n}\}$ as $\overleftrightarrow{e}_+ = \hat{\ell} \otimes \hat{\ell} - \hat{m} \otimes \hat{m}$ and $\overleftrightarrow{e}_\times = \hat{\ell} \otimes \hat{m} + \hat{m} \otimes \hat{\ell}$.

The measured scalar CW signal $h^X(t)$ at time t by detector X is the response of the detector to the GW tensor $\overleftrightarrow{h}(\tau^X(t))$, where $\tau^X(t)$ denotes the emission time of a wavefront that reaches detector X at time t. This timing relationship depends on the sky-position $\hat{n}$ of the source as well as any binary-orbital parameters in case of a CW from a neutron star in a binary system, as it describes the time-dependent light-travel time from the source to the detector. In the long-wavelength limit we assume the GW wavelength to be much larger than the detector armlength, which is a good approximation for current ground-based detectors up to kHz frequencies. This allows us to write the detector response as a tensor contraction (in both tensor indices):

$$h^X(t) = \overleftrightarrow{d}^X(t) : \overleftrightarrow{h}(\tau^X(t)), \tag{B2}$$

where $\overleftrightarrow{d}^X = \hat{u} \otimes \hat{u} - \hat{v} \otimes \hat{v}$ for interferometer arms along unit vectors $\hat{u}$ and $\hat{v}$.

It is helpful to define a *source-independent* orthonormal polarization basis $\{\hat{i}, \hat{j}, -\hat{n}\}$ instead, where for any sky position $\hat{n}$, the unit vector $\hat{i}$ is chosen to lie in Earth's equatorial plane (pointing West) and $\hat{j}$ is pointing in the northern hemisphere. This defines the (sky-position dependent) alternative polarization basis as $\overleftrightarrow{\varepsilon}_+(\hat{n}) \equiv \hat{i} \otimes \hat{i} - \hat{j} \otimes \hat{j}$ and $\overleftrightarrow{\varepsilon}_\times(\hat{n}) \equiv \hat{i} \otimes \hat{j} + \hat{j} \otimes \hat{i}$. The rotation between these two basis systems defines the *polarization angle* ψ, which is measured counterclockwise from $\hat{i}$ to $\hat{\ell}$, and relates the two polarization basis tensors as

$$\overleftrightarrow{e}_+ = \overleftrightarrow{\varepsilon}_+ \cos 2\psi + \overleftrightarrow{\varepsilon}_\times \sin 2\psi \tag{B3}$$

$$\overleftrightarrow{e}_\times = -\overleftrightarrow{\varepsilon}_+ \sin 2\psi + \overleftrightarrow{\varepsilon}_\times \cos 2\psi. \tag{B4}$$

Combining these expression, we can obtain the factored signal form $h^X(t; \mathcal{A}, \lambda) = \mathcal{A}^\mu h_\mu^X(t; \lambda)$ of Eq. (7), which was first derived in [37]. The four amplitudes $\{\mathcal{A}^\mu\}_{\mu=1}^4$ depend

on the signal amplitude h_0, the inclination angle ι via $A_+(h_0, \iota)$ and $A_\times(h_0, \iota)$ given in Eq. (6). They also depend on the polarization angle ψ, and the reference-time phase ϕ_0, namely

$$\mathcal{A}^1 = A_+ \cos\phi_0 \cos 2\psi - A_\times \sin\phi_0 \sin 2\psi,$$
$$\mathcal{A}^2 = A_+ \cos\phi_0 \sin 2\psi + A_\times \sin\phi_0 \cos 2\psi,$$
$$\mathcal{A}^3 = -A_+ \sin\phi_0 \cos 2\psi - A_\times \cos\phi_0 \sin 2\psi,$$
$$\mathcal{A}^4 = -A_+ \sin\phi_0 \sin 2\psi + A_\times \cos\phi_0 \cos 2\psi, \tag{B5}$$

and the four (detector-dependent) basis functions $h_\mu^X(t; \lambda)$ are

$$h_1^X(t) = a^X(t) \cos\phi(\tau^X(t)),$$
$$h_2^X(t) = b^X(t) \cos\phi(\tau^X(t)),$$
$$h_3^X(t) = a^X(t) \sin\phi(\tau^X(t)),$$
$$h_4^X(t) = b^X(t) \sin\phi(\tau^X(t)), \tag{B6}$$

in terms of the *antenna-pattern functions* $a^X(t)$, $b^X(t)$ given by the contractions

$$a^X(t; \hat{n}) = \overleftrightarrow{d}^X(t) : \overleftrightarrow{\varepsilon}_+(\hat{n}),$$
$$b^X(t; \hat{n}) = \overleftrightarrow{d}^X(t) : \overleftrightarrow{\varepsilon}_\times(\hat{n}). \tag{B7}$$

Using the factored signal form of Eq. (7), the log-likelihood ratio Eq. (B8) now takes the form

$$\ln \Lambda(x; \mathcal{A}, \lambda) = \mathcal{A}^\mu x_\mu - \frac{1}{2} \mathcal{A}^\mu \mathcal{M}_{\mu\nu} \mathcal{A}^\nu, \tag{B8}$$

where we defined

$$x_\mu(\lambda) \equiv (x, h_\mu), \quad \text{and} \quad \mathcal{M}_{\mu\nu}(\lambda) \equiv (h_\mu, h_\nu), \tag{B9}$$

in terms of the four basis function $h_\mu(t; \lambda)$ defined in Eq. (B6). The 4×4 *antenna-pattern matrix* $\mathcal{M}$ can be shown to be well approximated by the block-diagonal form

$$\mathcal{M} = S_n^{-1} T_{\text{data}} \begin{pmatrix} M & 0 \\ 0 & M \end{pmatrix} \quad \text{with} \quad M \equiv \begin{pmatrix} A & C \\ C & B \end{pmatrix}, \tag{B10}$$

defining the antenna-pattern coefficients A, B, C, which depend on the sky-position $\hat{n}$.

We see in Eq. (B8) that the log-likelihood ratio is a quadratic function of the amplitudes $\mathcal{A}^\mu$, and can therefore be analytically maximized [37] (or marginalized [94]) to yield the well-known $\mathcal{F}$-statistic:

$$\mathcal{F}(x; \lambda) \equiv \max_{\mathcal{A}} \ln \Lambda(x; \mathcal{A}, \lambda)$$

$$= \frac{1}{2} x_\mu \mathcal{M}^{\mu\nu} x_\nu, \tag{B11}$$

with $\mathcal{M}^{\mu\nu}$ defined as the inverse matrix to $\mathcal{M}_{\mu\nu}$ of Eq. (B10).

APPENDIX C: DEFINITION OF THE GEOMETRIC FACTOR R^2

The geometric factor R^2 can be explicitly expressed [95] as

$$R^2(\theta) = \frac{25}{4}[\alpha_1 A(\hat{n}) + \alpha_2 B(\hat{n}) + 2\alpha_3 C(\hat{n})], \tag{C1}$$

with the sky-dependent antenna-pattern coefficients $\{A, B, C\}$ of Eq. (B10), and

$$\alpha_1 \equiv \frac{1}{4}(1 + \cos^2\iota)^2 \cos^2 2\psi + \cos^2\iota \sin^2 2\psi, \tag{C2}$$

$$\alpha_2 \equiv \frac{1}{4}(1 + \cos^2\iota)^2 \sin^2 2\psi + \cos^2\iota \cos^2 2\psi, \tag{C3}$$

$$\alpha_3 \equiv \frac{1}{4}(1 - \cos^2\iota)^2 \sin 2\psi \cos 2\psi. \tag{C4}$$

One can show that R^2 averaged over $\psi \in [-\pi/4, \pi/4]$ and $\cos\iota \in [-1, 1]$ yields

$$\langle R^2 \rangle_{\cos\iota, \psi} = \frac{5}{2}(A(\hat{n}) + B(\hat{n})), \tag{C5}$$

and further averaging $\hat{n}$ isotropically over the sky yields

$$\langle R^2 \rangle_\theta = 1. \tag{C6}$$

APPENDIX D: DISTRIBUTION OF $\mathcal{F}$-STATISTIC MAXIMIZED OVER CORRELATED TEMPLATES

It has been a long-standing assumption (e.g., [1,46] that the distribution of the statistic $2\mathcal{F}^*(x) \equiv \max_{\lambda_i} 2\mathcal{F}(x; \lambda_i)$ in Gaussian noise x, maximized over a template bank $\lambda_i \in \mathbb{T}$ of $i = 1 \dots \mathcal{N}$ (generally correlated) templates can be modeled by assuming maximization over an "effective" number of uncorrelated trials $\mathcal{N}'$ instead, namely

$$P(2\mathcal{F}^*|\mathcal{N}') = \mathcal{N}' \mathrm{cdf}_0(2\mathcal{F}^*)^{\mathcal{N}'-1} \mathrm{pdf}_0(2\mathcal{F}^*), \tag{D1}$$

where

$$\mathrm{pdf}_0(2\mathcal{F}) = P(2\mathcal{F}|\rho = 0), \tag{D2}$$

$$\mathrm{cdf}_0(2\mathcal{F}) = \int_0^{2\mathcal{F}} \mathrm{pdf}_0(2\mathcal{F}') \mathrm{d}2\mathcal{F}', \tag{D3}$$

where the (single-template) $\mathcal{F}$-statistic in pure Gaussian noise follows a central χ^2 distribution [with four d.o.f. in the fully-coherent case Eq. (13), or $4N_{\mathrm{seg}}$ d.o.f. for a semicoherent $\mathcal{F}$-statistic over N_{seg} segments, Eq. (22)].

We show here by counterexample that the model of Eq. (D1) is not generally accurate, as correlations between templates do not simply modify $\mathcal{N}'$ but also change the functional *form* of the distribution. It has been hypothesized previously [1] that these (already-observed) deviations might be due to certain approximations (cf. [95]) used in the numerical implementation of the $\mathcal{F}$ statistic. While such effects will account for some amount of deviation, one can show this effect to be quite small overall.

We demonstrate the fundamental statistical nature of this discrepancy by using a simpler example: we generate a time-series $\{x_j\}_{j=0}^{N-1}$ of $N = 200$ samples drawn from a Gaussian distribution and compute the Fourier transform $\tilde{x}_k$ normalized to $E[|\tilde{x}_k|^2] = 2$, such that $2\mathcal{F}_2(x, f) \equiv |\tilde{x}(f)|^2$ follows a central χ^2 distribution with two d.o.f. in every frequency bin f. We can therefore set $\mathrm{pdf}_0(2\mathcal{F}_2) = \chi_2^2(2\mathcal{F}_2; 0)$ and use the corresponding cdf in Eq. (D1).

We consider different cases of *oversampling* by zero-padding the time series to a multiple (denoted as the oversampling factor in Fig. 10) of the original N time samples: the $N/2 - 1 = 99$ (positive) frequency bins without oversampling are strictly uncorrelated (and we also know that there can be at most $N = 200$ independent templates in total, given the length of the initial time series). With increasing oversampling, the correlations between frequency bins increase. We repeat this process 10^6 times for different noise realizations, and in each case we compute $2\mathcal{F}_2^*(x)$ over all the (positive) frequency bins of the Fourier power, and histogram these values. We then fit the number of effective templates $\mathcal{N}'$ in the theoretical distribution of Eq. (D1) by minimizing the (symmetric) Jensen–Shannon divergence between the measured and theoretical distributions. The results are shown in Fig. 10 for different cases of oversampling. We see that for increased oversampling, i.e., more correlations between "templates" (i.e., frequency bins), the functional form of the histogram agrees less with the theoretical distribution assuming independent templates. The effect seems to saturate for oversampling $\gtrsim 10$, with $\mathcal{N} \sim 230$ greater than the known maximal number (i.e., $N = 200$) of (strictly) independent templates in this vector space.

There is no simple or intuitive explanation for this effect that we are aware of, but it is reminiscent of a similarly surprising result found in the localization of the maximum over different assumed signal durations of transient CW signals, see Figs. 8 and 9 in [96]. The distribution of the statistic is identical in each time step, but the steps are correlated, resulting in a peculiar nonuniform distribution of the location of the maximum.
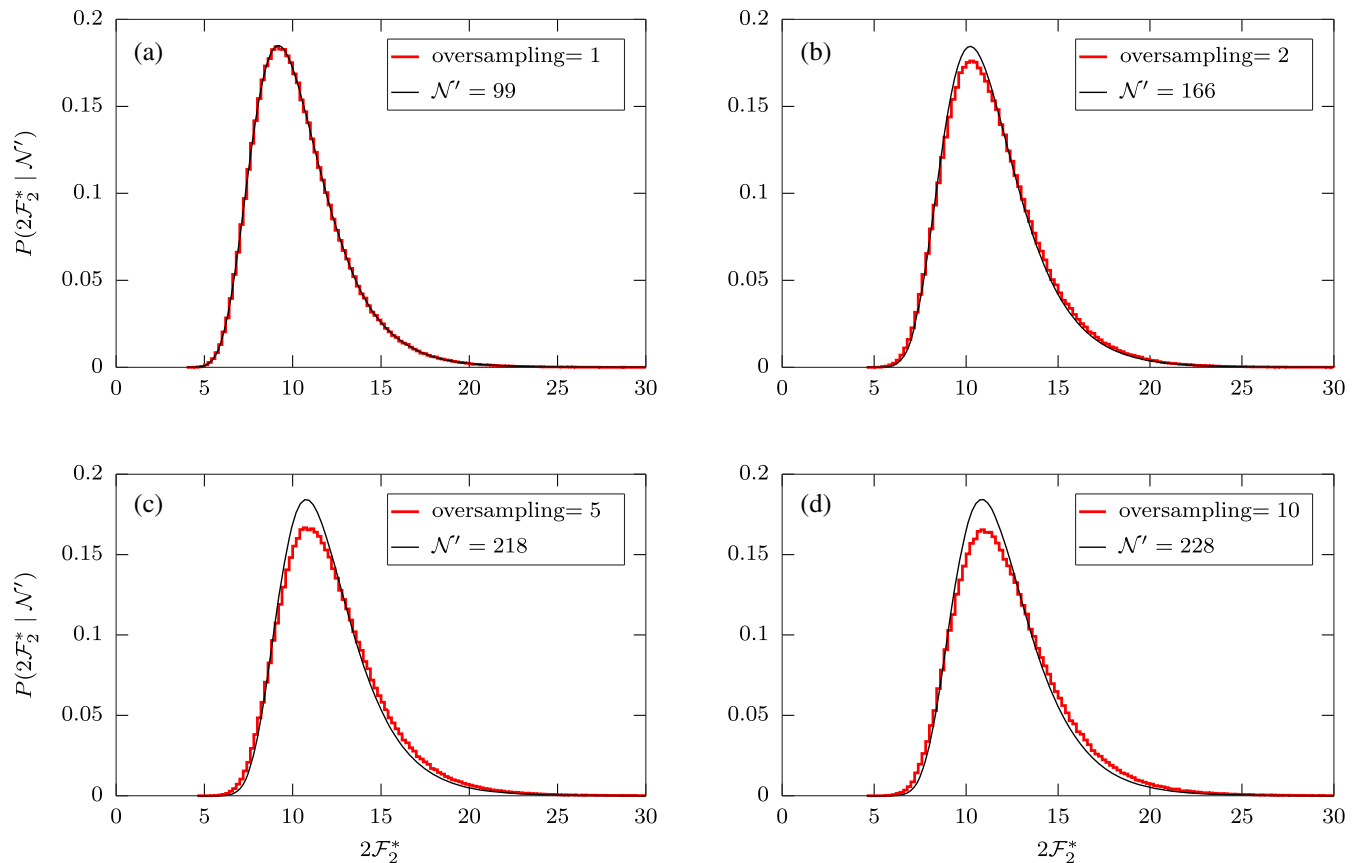
FIG. 10.   Stair-case plot: histogram (over 10^6 repeated trials) of $2\mathcal{F}_2^* = \max_k |\tilde{x}_k|^2$ for Fourier transforms of Gaussian-noise timeseries, using different oversampling factors (a)–(d), where oversampling $= 1$ corresponds to the original FFT frequency resolution. Solid thin line: corresponding best-fit theoretical model Eq. (D1) with an effective number of templates $\mathcal{N}'$.

[1] K. Wette, Estimating the sensitivity of wide-parameter-space searches for gravitational-wave pulsars, Phys. Rev. D **85,** 042003 (2012).

[2] B. P. Abbott, R. Abbott, T. D. Abbott, M. R. Abernathy, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari *et al.*, Observation of Gravitational Waves from a Binary Black Hole Merger, Phys. Rev. Lett. **116,** 061102 (2016).

[3] B. P. Abbott *et al.* (LIGO Scientific Collaboration; Virgo Collaboration), Binary Black Hole Mergers in the first Advanced LIGO Observing Run, Phys. Rev. X **6,** 041015 (2016).

[4] B. P. Abbott, R. Abbott, T. D. Abbott, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, V. B. Adya *et al.*, GW170817: Observation of Gravitational Waves from a Binary Neutron Star Inspiral, Phys. Rev. Lett. **119,** 161101 (2017).

[5] B. P. Abbott *et al.* (LIGO Scientific Collaboration; Virgo Collaboration), Tests of General Relativity with GW150914, Phys. Rev. Lett. **116,** 221101 (2016).

[6] B. P. Abbott *et al.* (LIGO Scientific Collaboration; Virgo Collaboration), GW170817: Measurements of neutron star radii and equation of state, arXiv:1805.11581.

[7] B. P. Abbott, R. Abbott, T. D. Abbott, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, V. B. Adya *et al.*, A gravitational-wave standard siren measurement of the Hubble constant, Nature (London) **551,** 85 (2017).

[8] R. Prix, Gravitational waves from spinning neutron stars, in *Neutron Stars and Pulsars* (Springer, New York, 2009), p. 651, https://dcc.ligo.org/LIGO-P060039/public.

[9] P. D. Lasky, Gravitational waves from neutron stars: A review, PASA **32,** e034 (2015).

[10] K. Riles, Recent searches for continuous gravitational waves, Mod. Phys. Lett. A **32,** 1730035 (2017).

[11] N. K. Johnson-McDaniel and B. J. Owen, Maximum elastic deformations of relativistic stars, Phys. Rev. D **88,** 044004 (2013).

[12] M. Camenzind, *Compact Objects in Astrophysics: White Dwarfs, Neutron Stars and Black Holes* (Springer Science & Business Media, New York, 2007).

[13] B. Knispel and B. Allen, Blandford's argument: The strongest continuous gravitational wave signal, Phys. Rev. D **78,** 044031 (2008).

[14] M. Isi, M. Pitkin, and A. J. Weinstein, Probing dynamical gravity with the polarization of continuous gravitational waves, Phys. Rev. D **96,** 042001 (2017).

[15] B. P. Abbott, R. Abbott, T. D. Abbott, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, V. B. Adya *et al.*, First Search for Nontensorial Gravitational Waves from Known Pulsars, Phys. Rev. Lett. **120,** 031104 (2018).

[16] P. R. Brady and T. Creighton, Searching for periodic sources with LIGO. II. Hierarchical searches, Phys. Rev. D **61,** 082001 (2000).

[17] R. Prix and M. Shaltev, Search for continuous gravitational waves: Optimal StackSlide method at fixed computing cost, Phys. Rev. D **85,** 084010 (2012).

[18] R. N. Manchester, G. B. Hobbs, A. Teoh, and M. Hobbs, The australia telescope national facility pulsar catalogue, Astron. J. **129,** 1993 (2005).

[19] B. P. Abbott, R. Abbott, T. D. Abbott, M. R. Abernathy, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari *et al.*, First search for gravitational waves from known pulsars with advanced LIGO, Astrophys. J. **839,** 12 (2017).

[20] B. P. Abbott, R. Abbott, T. D. Abbott, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, V. B. Adya *et al.*, First narrow-band search for continuous gravitational waves from known pulsars in advanced detector data, Phys. Rev. D **96,** 122006 (2017).

[21] J. Aasi, B. P. Abbott, R. Abbott, T. Abbott, M. R. Abernathy, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso *et al.*, Searches for continuous gravitational waves from nine young supernova remnants, Astrophys. J. **813,** 39 (2015).

[22] S. J Zhu, M. A. Papa, H.-B. Eggenstein, R. Prix, K. Wette, B. Allen, O. Bock, D. Keitel, B. Krishnan, B. Machenschalk *et al.*, Einstein@Home search for continuous gravitational waves from cassiopeia a, Phys. Rev. D **94,** 082008 (2016).

[23] B. P. Abbott, R. Abbott, T. D. Abbott, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, V. B. Adya *et al.*, Upper limits on gravitational waves from Scorpius X-1 from a model-based cross-correlation search in advanced LIGO data, Astrophys. J. **847,** 47 (2017).

[24] B. P. Abbott, R. Abbott, T. D. Abbott, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, V. B. Adya *et al.*, Search for gravitational waves from Scorpius X-1 in the first Advanced LIGO observing run with a hidden Markov model, Phys. Rev. D **95,** 122003 (2017).

[25] B. P. Abbott, R. Abbott, T. D. Abbott, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, V. B. Adya *et al.*, All-sky search for periodic gravitational waves in the O1 LIGO data, Phys. Rev. D **96,** 062002 (2017).

[26] B. P. Abbott, R. Abbott, T. D. Abbott, M. R. Abernathy, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari *et al.*, First low-frequency Einstein@Home all-sky search for continuous gravitational waves in advanced LIGO data, Phys. Rev. D **96,** 122004 (2017).

[27] E. Goetz and K. Riles, An all-sky search algorithm for continuous gravitational waves from spinning neutron stars in binary systems, Classical Quantum Gravity **28,** 215006 (2011).

[28] J. Aasi, B. P. Abbott, R. Abbott, T. Abbott, M. R. Abernathy, T. Accadia, F. Acernese, K. Ackley, C. Adams, T. Adams *et al.*, First all-sky search for continuous gravitational waves from unknown sources in binary systems, Phys. Rev. D **90,** 062010 (2014).

[29] J. Ming, B. Krishnan, M. A. Papa, C. Aulbert, and H. Fehrmann, Optimal directed searches for continuous gravitational waves, Phys. Rev. D **93,** 064011 (2016).

[30] J. Ming, M. Alessandra Papa, B. Krishnan, R. Prix, C. Beer, S. J. Zhu, H.-B. Eggenstein, O. Bock, and B. Machenschalk, Optimally setting up directed searches for continuous gravitational waves in Advanced LIGO O1 data, Phys. Rev. D **97,** 024051 (2018).

[31] P. Leaci and R. Prix, Directed searches for continuous gravitational waves from binary systems: Parameter-space metrics and optimal Scorpius X-1 sensitivity, Phys. Rev. D **91,** 102003 (2015).

[32] B. Abbott, R. Abbott, R. Adhikari, A. Ageev, B. Allen, R. Amin, S. B. Anderson, W. G. Anderson, M. Araya, H. Armandula *et al.*, Setting upper limits on the strength of periodic gravitational waves from PSR J1939+2134 using the first science data from the GEO 600 and LIGO detectors, Phys. Rev. D **69,** 082004 (2004).

[33] B. Krishnan, A. M. Sintes, M. A. Papa, B. F. Schutz, S. Frasca, and C. Palomba, Hough transform search for continuous gravitational waves, Phys. Rev. D **70,** 082001 (2004).

[34] G. Mendell and M. Landry, StackSlide and Hough search SNR and statistics, LIGO Technical Document No. LIGO-T050003-x0, 2005, https://dcc.ligo.org/LIGO-T050003/public.

[35] B. Abbott, R. Abbott, R. Adhikari, J. Agresti, P. Ajith, B. Allen, R. Amin, S. B. Anderson, W. G. Anderson, M. Arain *et al.*, All-sky search for periodic gravitational waves in LIGO S4 data, Phys. Rev. D **77,** 022001 (2008).

[36] J. Aasi, J. Abadie, B. P. Abbott, R. Abbott, T. D. Abbott, M. Abernathy, T. Accadia, F. Acernese, C. Adams, T. Adams *et al.*, Einstein@Home all-sky search for periodic gravitational waves in LIGO S5 data, Phys. Rev. D **87,** 042001 (2013).

[37] P. Jaranowski, A. Krolak, and B. F. Schutz, Data analysis of gravitational-wave signals from spinning neutron stars: The signal and its detection, Phys. Rev. D **58,** 063001 (1998).

[38] L. S. Finn, Detection, measurement, and gravitational radiation, Phys. Rev. D **46,** 5236 (1992).

[39] C. Cutler and B. F. Schutz, Generalized F-statistic: Multiple detectors and multiple gravitational wave pulsars, Phys. Rev. D **72,** 063006 (2005).

[40] R. Prix, Search for continuous gravitational waves: Metric of the multidetector F-statistic, Phys. Rev. D **75,** 023004 (2007).

[41] R. Prix, The $\mathcal{F}$-statistic and its implementation in ComputeFstatistic_v2, LIGO Technical Report No. LIGO-T0900149-v5, 2015, https://dcc.ligo.org/LIGO-T0900149/public.

[42] C. Cutler, I. Gholami, and B. Krishnan, Improved stack-slide searches for gravitational-wave pulsars, Phys. Rev. D **72,** 042004 (2005).

[43] Einstein@Home project page, https://einsteinathome.org/.

[44] K. Wette, Parameter-space metric for all-sky semicoherent searches for gravitational-wave pulsars, Phys. Rev. D **92,** 082003 (2015).

[45] B. Behnke, M. A. Papa, and R. Prix, Postprocessing methods used in the search for continuous gravitational-wave signals from the Galactic Center, Phys. Rev. D **91,** 064007 (2015).

[46] J. Abadie, B. P. Abbott, R. Abbott, M. Abernathy, C. Adams, R. Adhikari, P. Ajith, B. Allen, G. Allen, E. A. Ceron et al., First search for gravitational waves from the youngest known neutron star, Astrophys. J. **722,** 1504 (2010).

[47] K. Wette, S. Walsh, R. Prix, and M. A. Papa, Implementing a semicoherent search for continuous gravitational waves using optimally-constructed template banks, Phys. Rev. D **97,** 123016 (2018).

[48] R. Prix and M. Shaltev, Fully coherent follow-up of continuous gravitational-wave candidates, Phys. Rev. D **87,** 084057 (2013).

[49] M. A. Papa, H.-B. Eggenstein, S. Walsh, I. Di Palma, B. Allen, P. Astone, O. Bock, T. D. Creighton, D. Keitel, B. Machenschalk, R. Prix, X. Siemens, A. Singh, S. J. Zhu, and B. F. Schutz, Hierarchical follow-up of subthreshold candidates of an all-sky Einstein@Home search for continuous gravitational waves on LIGO sixth science run data, Phys. Rev. D **94,** 122006 (2016).

[50] C. Röver, C. Messenger, and R. Prix, Bayesian versus frequentist upper limits, in Proceedings of the PHYSTAT 2011 Workshop on Statistical Issues Related to Discovery Claims in Search Experiments and Unfolding, edited by H. B. Prosper and L. Lyons (CERN, Geneva, Switzerland, 2011), p. 158, arXiv:1103.2987.

[51] R. J. Dupuis and G. Woan, Bayesian estimation of pulsar parameters from gravitational wave data, Phys. Rev. D **72,** 102002 (2005).

[52] J. Aasi, J. Abadie, B. P. Abbott, R. Abbott, T. Abbott, M. R. Abernathy, T. Accadia, F. Acernese, C. Adams, T. Adams et al., Gravitational waves from known pulsars: Results from the initial detector era, Astrophys. J. **785,** 119 (2014).

[53] J. W. Eaton, D. Bateman, S. Hauberg, and R. Wehbring, GNU Octave version 4.0.0 manual: A high-level interactive language for numerical computations (2015), http://www.gnu.org/software/octave/doc/interpreter.

[54] K. Wette, R. Prix, D. Keitel, M. Pitkin, C. Dreissigacker, J. T. Whelan, and P. Leaci, OctApps: A library of Octave functions for continuous gravitational-wave data analysis, J. Open Source Software **3,** 707 (2018).

[55] R. Prix and K. Wette, Estimating sensitivity of the Einstein@Home search S5R5, Report No. LIGO-T1200272, 2012, https://dcc.ligo.org/LIGO-T1200272/public.

[56] B. P. Abbott, R. Abbott, T. D. Abbott, M. R. Abernathy, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari et al., Results of the deepest all-sky survey for continuous gravitational waves on LIGO S6 data running on the Einstein@Home volunteer distributed computing project, Phys. Rev. D **94,** 102002 (2016).

[57] H. J. Pletsch and B. Allen, Exploiting Large-Scale Correlations to Detect Continuous Gravitational Waves, Phys. Rev. Lett. **103,** 181102 (2009).

[58] B. Abbott, R. Abbott, R. Adhikari, A. Ageev, J. Agresti, B. Allen, J. Allen, R. Amin, S. B. Anderson, W. G. Anderson et al., First all-sky upper limits from LIGO on the strength of periodic gravitational waves using the hough transform, Phys. Rev. D **72,** 102004 (2005).

[59] B. Abbott, R. Abbott, R. Adhikari, J. Agresti, P. Ajith, B. Allen, R. Amin, S. B. Anderson, W. G. Anderson, M. Arain et al., Searches for periodic gravitational waves from unknown isolated sources and Scorpius X-1: Results from the second LIGO science run, Phys. Rev. D **76,** 082001 (2007).

[60] B. Abbott, R. Abbott, R. Adhikari, P. Ajith, B. Allen, G. Allen, R. Amin, D. P. Anderson, S. B. Anderson, W. G. Anderson et al., Einstein@Home search for periodic gravitational waves in LIGO S4 data, Phys. Rev. D **79,** 022001 (2009).

[61] B. P. Abbott, R. Abbott, R. Adhikari, P. Ajith, B. Allen, G. Allen, R. S. Amin, S. B. Anderson, W. G. Anderson, M. A. Arain et al., All-Sky LIGO Search for Periodic Gravitational Waves in the Early Fifth-Science-Run Data, Phys. Rev. Lett. **102,** 111102 (2009).

[62] B. P. Abbott, R. Abbott, R. Adhikari, P. Ajith, B. Allen, G. Allen, R. S. Amin, S. B. Anderson, W. G. Anderson, M. A. Arain et al., Einstein@Home search for periodic gravitational waves in early S5 LIGO data, Phys. Rev. D **80,** 042003 (2009).

[63] J. Abadie, B. P. Abbott, R. Abbott, T. D. Abbott, M. Abernathy, T. Accadia, F. Acernese, C. Adams, R. Adhikari, C. Affeldt et al., All-sky search for periodic gravitational waves in the full S5 LIGO data, Phys. Rev. D **85,** 022001 (2012).

[64] J. Aasi, J. Abadie, B. P. Abbott, R. Abbott, T. Abbott, M. R. Abernathy, T. Accadia, F. Acernese, C. Adams, T. Adams et al., Application of a hough search for continuous gravitational waves on data from the fifth LIGO science run, Classical Quantum Gravity **31,** 085014 (2014).

[65] A. Singh, M. A. Papa, H.-B. Eggenstein, S. Zhu, H. Pletsch, B. Allen, O. Bock, B. Maschenchalk, R. Prix, and X. Siemens, Results of an all-sky high-frequency Einstein@Home search for continuous gravitational waves in LIGO's fifth science run, Phys. Rev. D **94,** 064061 (2016).

[66] J. Aasi, B. P. Abbott, R. Abbott, T. Abbott, M. R. Abernathy, T. Accadia, F. Acernese, K. Ackley, C. Adams, T. Adams et al., Implementation of an F-statistic all-sky search for continuous gravitational waves in Virgo VSR1 data, Classical Quantum Gravity **31,** 165014 (2014).

[67] J. Aasi, B. P. Abbott, R. Abbott, T. D. Abbott, M. R. Abernathy, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso et al., First low frequency all-sky search for continuous gravitational wave signals, Phys. Rev. D **93,** 042007 (2016).

[68] B. P. Abbott et al., Comprehensive all-sky search for periodic gravitational waves in the sixth science run LIGO data, Phys. Rev. D **94,** 042002 (2016).

[69] B. P. Abbott, R. Abbott, T. D. Abbott, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso et al. (The LIGO Scientific Collaboration, the Virgo Collaboration), Full band all-sky search for periodic gravitational waves in the O1 LIGO Data, Phys. Rev. D **97,** 102003 (2018).

[70] B. Abbott, R. Abbott, R. Adhikari, P. Ajith, B. Allen, G. Allen, R. Amin, S. B. Anderson, W. G. Anderson, M. A. Arain *et al.*, Beating the spin-down limit on gravitational wave emission from the crab pulsar, Astrophys. J. Lett. **683**, L45 (2008).

[71] J. Aasi, J. Abadie, B. P. Abbott, R. Abbott, T. Abbott, M. R. Abernathy, T. Accadia, F. Acernese, C. Adams, T. Adams *et al.*, Directed search for continuous gravitational waves from the galactic center, Phys. Rev. D **88**, 102002 (2013).

[72] J. Aasi, B. P. Abbott, R. Abbott, T. Abbott, M. R. Abernathy, F. Acernese, K. Ackley, C. Adams, T. Adams, T. Adams *et al.*, Narrow-band search of continuous gravitational-wave signals from Crab and Vela pulsars in Virgo VSR4 data, Phys. Rev. D **91**, 022004 (2015).

[73] J. Aasi, B. P. Abbott, R. Abbott, T. D. Abbott, M. R. Abernathy, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso *et al.*, Search of the Orion spur for continuous gravitational waves using a loosely coherent algorithm on data from LIGO interferometers, Phys. Rev. D **93**, 042006 (2016).

[74] B. P. Abbott, R. Abbott, T. D. Abbott, M. R. Abernathy, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari *et al.*, Search for continuous gravitational waves from neutron stars in globular cluster NGC 6544, Phys. Rev. D **95**, 082005 (2017).

[75] B. P. Abbott, R. Abbott, T. D. Abbott, M. R. Abernathy, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari *et al.*, Directional Limits on Persistent Gravitational Waves from Advanced LIGO's First Observing Run, Phys. Rev. Lett. **118**, 121102 (2017).

[76] J. Aasi, B. P. Abbott, R. Abbott, T. Abbott, M. R. Abernathy, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso *et al.*, A directed search for gravitational waves from scorpius x-1 with initial LIGO data, Phys. Rev. D **91**, 062008 (2015).

[77] G. D. Meadors, E. Goetz, K. Riles, T. Creighton, and F. Robinet, Searches for continuous gravitational waves from Scorpius X-1 and XTE J1751-305 in LIGO's sixth science run, Phys. Rev. D **95**, 042005 (2017).

[78] B. Abbott, R. Abbott, R. Adhikari, A. Ageev, B. Allen, R. Amin, S. B. Anderson, W. G. Anderson, M. Araya, H. Armandula *et al.*, Limits on Gravitational-Wave Emission from Selected Pulsars Using LIGO Data, Phys. Rev. Lett. **94**, 181103 (2005).

[79] B. Abbott, R. Abbott, R. Adhikari, J. Agresti, P. Ajith, B. Allen, R. Amin, S. B. Anderson, W. G. Anderson, M. Arain *et al.*, Upper limits on gravitational wave emission from 78 radio pulsars, Phys. Rev. D **76**, 042001 (2007).

[80] B. P. Abbott, R. Abbott, F. Acernese, R. Adhikari, P. Ajith, B. Allen, G. Allen, M. Alshourbagy, R. S. Amin, S. B.

Anderson *et al.*, Searches for gravitational waves from known pulsars with Science Run 5 LIGO data, Astrophys. J. **713**, 671 (2010).

[81] J. Abadie, B. P. Abbott, R. Abbott, M. Abernathy, T. Accadia, F. Acernese, C. Adams, R. Adhikari, C. Affeldt, B. Allen *et al.*, Beating the spin-down limit on gravitational wave emission from the vela pulsar, Astrophys. J. **737**, 93 (2011).

[82] P. J. Rousseeuw and C. Croux, Alternatives to the median absolute deviation, J. Am. Stat. Assoc. **88**, 1273. (1993).

[83] H. Grote (for the LIGO Scientific Collaboration), The GEO 600 status, Classical Quantum Gravity **27**, 084003 (2010).

[84] T. Accadia *et al.* (VIRGO), Virgo: A laser interferometer to detect gravitational waves, J. Instrum. **7**, P03012 (2012).

[85] F. Acernese *et al.* (VIRGO), Advanced Virgo: A second-generation interferometric gravitational wave detector, Classical Quantum Gravity **32**, 024001 (2015).

[86] B. P. Abbott *et al.* (LIGO Scientific Collaboration), LIGO: The Laser interferometer gravitational-wave observatory, Rept. Prog. Phys. **72**, 076901 (2009).

[87] J. Aasi *et al.* (LIGO Scientific Collaboration), Advanced LIGO, Classical Quantum Gravity **32**, 074001 (2015).

[88] V. Dergachev (private communication).

[89] Initial LIGO strain noise data (psds), https://labcit.ligo .caltech.edu/~jzweizig/distribution/LSC_Data/.

[90] Representative O1 noise PSDs, H1 LIGO-G1600150, L1 LIGO-G1600151, https://dcc.ligo.org/LIGO-G1600150/ public for the H1 detector, https://dcc.ligo.org/LIGO-G1600151/public for the L1 detector.

[91] K. Wette *et al.*, Searching for gravitational waves from Cassiopeia A with LIGO, Classical Quantum Gravity **25**, 235011 (2008).

[92] S. Bonazzola and E. Gourgoulhon, Gravitational waves from pulsars: emission by the magnetic-field-induced distortion, Astron. Astrophys. **312**, 675 (1996), https://ui .adsabs.harvard.edu/#abs/1996A&A...312..675B.

[93] J. T. Whelan, R. Prix, C. J. Cutler, and J. L. Willis, New coordinates for the amplitude parameter space of continuous gravitational waves, Classical Quantum Gravity **31**, 065002 (2014).

[94] R. Prix and B. Krishnan, Targeted search for continuous gravitational waves: Bayesian versus maximum-likelihood statistics, Classical Quantum Gravity **26**, 204013 (2009).

[95] R. Prix, The F-statistic and its implementation in ComputeFStatistic v2, Report No. LIGO-T0900149, 2010, https://dcc.ligo.org/LIGO-T0900149/public.

[96] R. Prix, S. Giampanis, and C. Messenger, Search method for long-duration gravitational-wave transients from neutron stars, Phys. Rev. D **84**, 023007 (2011).