PHYSICAL REVIEW D **84,** 023007 (2011)

Search method for long-duration gravitational-wave transients from neutron stars

R. Prix,* S. Giampanis,† and C. Messenger‡

Albert-Einstein-Institut, Callinstrasse 38, 30167 Hannover, Germany

(Received 9 April 2011; published 14 July 2011)

We introduce a search method for a new class of gravitational-wave signals, namely, long-duration $\mathcal{O}$(hours-weeks) transients from spinning neutron stars. We discuss the astrophysical motivation from glitch relaxation models and we derive a rough estimate for the maximal expected signal strength based on the superfluid excess rotational energy. The transient signal model considered here extends the traditional class of infinite-duration *continuous-wave* signals by a finite start-time and duration. We derive a multidetector Bayes factor for these signals in Gaussian noise using $\mathcal{F}$-statistic amplitude priors, which simplifies the detection statistic and allows for an efficient implementation. We consider both a fully coherent statistic, which is computationally limited to directed searches for known pulsars, and a cheaper semicoherent variant, suitable for wide parameter-space searches for transients from unknown neutron stars. We have tested our method by Monte-Carlo simulation, and we find that it outperforms orthodox maximum-likelihood approaches both in sensitivity and in parameter-estimation quality.

DOI: 10.1103/PhysRevD.84.023007 PACS numbers: 04.30.−w, 04.30.Tv, 04.80.Nn, 07.05.Kf

I. INTRODUCTION

Gravitational waves (GWs), predicted by Einstein's general theory of relativity, are expected to be emitted by spinning neutron stars (NSs) with nonaxisymmetric deformations or currents. Traditionally we distinguish two categories of GW emission from NSs:

(i) short ''burstlike'' GWs from NS oscillations, e.g. fundamental (f-), pressure (p-), Rossby (r-) or w-modes, which could be excited by a NS glitch. These signals would typically be in the kHz frequency range and would be damped on time scales of milliseconds (see [1,2]). Various data-analysis methods for such high-frequency bursts have been developed (e.g. see [3,4]) and recently a search for a GW burst from a glitch in the Vela pulsar has been performed on LIGO data [5].

(ii) long-duration ''continuous waves'' (CWs) from nonaxisymmetric deformations or currents in spinning NSs. These CWs are quasisinusoidal with a well-defined, slowly varying frequency f, which is typically of order of the NS spin frequency ν: in particular $f = 2\nu$ for ''mountains'' or precession, additionally $f \approx \nu$ for precessing NSs, while $f \approx 4\nu/3$ for r-mode oscillations. In the past decade a number of data-analysis methods have been developed and applied to searches for CW signals in the data of ground-based detectors. These searches typically come in two flavors, either fully coherently targeting known pulsars at $f = 2\nu$, or semicoherently searching for CWs from unknown NSs in a large parameter space of frequencies and sky-positions. See [6] for a review of the astrophysical models, data-analysis methods and results of searches for CWs, and for further references.

The traditional CW model assumes that these signals are truly *continuous* in the sense of a quasi-infinite duration, or at least of longer duration than the available observation time, typically $T_{\mathrm{obs}} \sim 1$–2 yr. This postulates a spinning neutron star with a quasistable nonaxisymmetric deformation, such that the relaxation to the axisymmetric thermodynamic equilibrium happens on long time scales $\gg T_{\mathrm{obs}}$. Alternatively, the nonaxisymmetry can be driven by external influences such as accretion in low-mass X-ray binary systems, or by GW-driven instabilities such as unstable r-mode oscillations. Both are complex dynamical processes that need to be perfectly stable in order to produce a traditional CW signal.

In the wide gap between burstlike $\mathcal{O}$(ms) and truly continuous $\mathcal{O}(\infty)$ signals, spinning NSs can reasonably be assumed to emit ''CW'' signals of intermediate duration of $\mathcal{O}$(hours-weeks). We refer to this third category of GWs from spinning NSs by the oxymoron ''transient CWs.'' We can give three plausibility arguments for why it would be worthwhile to explore this new parameter space: (a) Young NSs exhibit enigmatic sudden spin-up events called ''glitches'', followed by a relaxation phase with time scale between days to months. This shows that internal dynamical processes on these time scales do exist in NSs. (b) Equilibrium NS configurations are axisymmetric and GWs from quasistable deviations are expected to be weak: deformations strong enough to produce detectable GWs on Earth would therefore seem more likely to be associated with ''catastrophic'' transient events. (c) Our astrophysical understanding of the universe is incomplete. Independently of any astrophysical models, this range of parameter space is currently not covered by any other searches and should therefore be investigated.

The transient-CW signal model is very similar to the standard CW model, namely, quasisinusoidal emission with a well-defined, slowly varying frequency of the order

*Reinhard.Prix@aei.mpg.de
†gstef@gravity.phys.uwm.edu
‡chris.messenger@astro.cf.ac.uk

of the NS spin frequency, $f \sim \mathcal{O}(\nu)$. Contrary to CWs, however, the signal has a definite start-time t^0 and a finite effective lifetime τ_s, and the signal amplitude h_0 can be modulated by a window-function, for example, an exponential decay.

While the present study is concerned with extending the traditional $\mathcal{O}(\infty)$ CW search to finite durations, there is an independent and complementary effort underway (called "STAMP") to extend the traditional $\mathcal{O}$(ms) "burst" searches to longer durations [7]. This "long burst" method does not assume a parametrized signal model, but instead tries to find connected time-frequency patterns in the cross-correlation power between multiple detectors. These transients can span any duration from seconds to weeks. This approach will be computationally cheaper and more robust towards a wider class of unmodeled GW signals, but would therefore also be expected to be less sensitive towards the particular class of transient-CW signals considered here.

The plan of this paper is as follows: in Sec. II we discuss the astrophysical motivation for transient CWs, including a simple signal-to-noise ratio (SNR) estimate based on the superfluid excess rotation energy in NSs. Sec. III introduces the parametrized transient-CW signal model, and Sec. IV develops the coherent and semicoherent Bayesian search methods for these signals. Sec. V presents numerical Monte-Carlo results on the detection efficiency, and Sec. VI illustrates the performance on parameter estimation. Sec. VII gives a concluding summary. Details on implementation and computing cost estimates are given in the appendix A.

II. ASTROPHYSICAL MOTIVATION FOR TRANSIENT CWS

A. Concrete astrophysical predictions

There are currently very few concrete predictions in the literature regarding the existence and properties of potential transient-CW signals from spinning NSs. Notable exceptions are the recent studies [8,9] of GW emission from nonaxisymmetric Ekman flow during the post-glitch relaxation phase, which typically lasts for days to months (e.g. see [10]). The authors conclude that GWs from this mechanism could be detectable with second- or third-generation ground-based detectors. Another interesting recent idea [11] suggests that giant magnetic flares in magnetars could trigger polar Alfvén oscillations, emitting GWs at around ~100 Hz and lasting for days to months, although the details of the underlying physics are uncertain [12].

One can further speculate on a number of potential transient-CW emission mechanisms from spinning NSs. Many "classical" CW mechanisms discussed in the literature (e.g. see [6] for references) typically have large uncertainties on the lifetime of the emission, and are therefore potential sources of transient-CW signals. For example, free precession of NSs occurs when the spin-axis is misaligned with the axis of symmetry, and has long been considered a possible mechanism for CW emission (e.g. [13]). A more detailed analysis [14] concluded that the emission from free precession could be damped on a (highly uncertain) time scale of a few weeks to years. Young NSs could be deformed by extreme magnetic fields and result in stronger transient CWs from damped free precession [15]. Newly-born magnetars with strong toroidal fields could be subject to a magnetic instability producing strong GWs on the time scale of several days [16]. Similarly, the strength and time scale of the r-mode GW instability remain highly uncertain (e.g. see [17] for a review), despite a number of studies over more than a decade. This instability could operate on time scales of days to months in newly-born NSs. As an example, see [18] for recent work on the large variety of possible scenarios for the r-mode instability and spin-down-evolution taking into account nonlinear mode couplings.

B. Energetics and SNR of transient CWs

It is instructive to consider the general relation between GW energy emitted and the average expected SNR for transient CWs. This can be useful to derive an order-of-magnitude estimate of the expected transient-CW SNR in a simple toy-model.

In the following estimate we assume that the CWs are emitted by a nonaxisymmetric deformation $\epsilon(t)$ of the quadrupole moment, emitting GWs at frequency f. See also [19] for a discussion of the relevant relations in the case of r-modes. The total energy emitted in GWs is $E_{\mathrm{GW}} = \int L_{\mathrm{GW}} dt$, in terms of the GW luminosity L_{GW} (e.g. see [6]), which is

$$L_{\mathrm{GW}}(t) = \frac{1}{10}\frac{G}{c^5}(2\pi f)^6 I^2 \epsilon^2(t), \tag{1}$$

where I is the axial moment of inertia, and $\epsilon(t)$ is the dimensionless deviation from axisymmetry of the spinning NS. G is Newton's gravitational constant, and c is the speed of light. We can write the corresponding signal amplitude $h_0(t)$ at the observer as

$$h_0(t) = \frac{4\pi^2 G}{c^4}\frac{I f^2}{d}\epsilon(t), \tag{2}$$

where d is the distance to the NS. Combining Eqs. (1) and (2), one can eliminate $\epsilon(t)$ and write the total GW energy emitted during a time span T as

$$E_{\mathrm{GW}} = \frac{2\pi^2 c^3}{5G} f^2 d^2 \int^T h_0^2(t) dt, \tag{3}$$

assuming a roughly constant average frequency f. The expected optimal signal-to-noise ratio (SNR), which will be defined in Eq. (42), can be averaged over sky-position and signal polarizations [13,20], to yield

$$\langle \rho_o^2 \rangle = \frac{4}{25} \frac{1}{S(f)} \int^T h_0^2(t) dt, \tag{4}$$

where $S(f)$ is the (single-sided) noise power spectral density (at the signal frequency f) of the detector (or multi-detector combination, see [20,21]). By combining Eqs. (3) and (4), we obtain the average optimal SNR of a transient CW in terms of the total emitted GW energy $E_{\rm GW}$, namely

$$\langle \rho_o^2 \rangle = \frac{2G}{5\pi^2 c^3} \frac{E_{\rm GW}}{S(f) f^2 d^2}, \tag{5}$$

which agrees with the result given in [22]. An analogous expression was derived in [23] for the case of r-modes. This expression is interesting for two reasons: (i) at fixed frequency f, the expected SNR only depends on the total GW energy emitted, and not on the time scale of the emission, i.e. short-strong or long-weak transient CWs of the same total energy typically result in the same SNR, and (ii) at fixed emitted $E_{\rm GW}$, the optimal SNR ρ_o *decreases* linearly with increasing GW frequency f, i.e. for the same transient GW energy, transients from slowly rotating NS would be easier to detect than from fast rotators.

For convenience we define the dimensionless root-mean-square amplitude $\hat{h}_0$ over the time scale T as

$$\hat{h}_0^2 T \equiv \int^T h_0^2(t) dt. \tag{6}$$

Note that this quantity is not to be confused with the "root-sum-squared" amplitude $h_{\rm rss}$ often used to characterize signal strength in the context of burst searches (e.g. see [3]), which is defined as $h_{\rm rss}^2 \equiv \int h^2(t) dt$. The difference is that $h_{\rm rss}$ has dimension $\sqrt{t}$ and refers to the *measured strain* $h(t)$ in a given detector (see Eq. (16)), which is rapidly oscillating with frequency f. The dimensionless $\hat{h}_0$, on the other hand, refers to the intrinsic signal amplitude $h_0(t)$, which for a transient CW would be slowly varying on a time scale T. From the definition (42) of the (optimal) SNR, one can see that $\rho_o^2 = (2/S) h_{\rm rss}^2$, and therefore Eq. (4) also yields the relation

$$\langle h_{\rm rss}^2 \rangle = \frac{2}{25} \hat{h}_0^2 T. \tag{7}$$

The quantity $\hat{h}_0^2 T$ is proportional to the GW energy (3), as well as the average optimal SNR (4), namely

$$\hat{h}_0^2 T \propto \langle \rho_o^2 \rangle S \propto \frac{E_{\rm GW}}{f^2 d^2}. \tag{8}$$

C. Transient CWs from superfluid excess energy

As a simple toy-model for the energy available *in principle* for transient CWs from a spinning NS, we consider the standard two-fluid model, which is at the core of current attempts to understand pulsar glitches. See [10] for an overview of the phenomenology of observed pulsar glitches and the basic two-fluid model.

The observed pulses (arriving with frequency ν) are commonly associated with the rotating NS magnetic field, which is attached to the crust and the *normal* fluid interior, both rotating with angular velocity $\Omega = 2\pi\nu$. This normal component has a moment of inertia $I_{\rm c}$, believed to form the bulk of the total moment of inertia, $I \approx I_{\rm c}$. The basic two-fluid model assumes that some of the interior neutrons are superfluid, forming an independent component that rotates at an (unobserved) angular velocity $\Omega_{\rm s}$, and which has a moment of inertia $I_{\rm s}$, typically believed to be of order $I_{\rm s}/I_{\rm c} \sim 10^{-2}$ (e.g. see [10]). The normal NS components slow down at an observed rate $\dot{\Omega} = 2\pi\dot{\nu}$ due to losses of angular momentum from the electromagnetic emission or interactions with the surrounding medium. The superfluid, on the other hand, is believed to be weakly coupled to the normal component (therefore $\dot{\Omega}_{\rm s} \approx 0$) and would continue to spin with angular velocity $\Omega_{\rm s}$, until the "lag" $\Delta\Omega = \Omega_{\rm s} - \Omega = -\dot{\Omega}\Delta t$ between the two components reaches a critical level (see [24] for a more detailed study including superfluid coupling effects). The time scale Δt here could correspond to the interglitch period, if one assumes that every glitch restores perfect corotation between the fluids. At this "breaking point" some type of instability is believed to occur, resulting in the transfer of angular momentum from the superfluid to the normal fluid. This would produce the visible "glitch", i.e. a sudden spin-up $\delta\Omega = 2\pi\delta\nu$ of the observed pulse frequency ν. The details of this instability are poorly understood and various models have been suggested in the literature, such as the crust breaking due to the strain exerted by pinned vortices (e.g. see [25]), the vortex array becoming unpinned due to the Magnus force (e.g. see [26]), or a two-stream instability developing due to the dynamical coupling of the two fluids [27,28]. See also [29] for an alternative mechanism to vortex pinning that can produce crust strain. Similarly unclear are the physical mechanisms involved in the ensuing "relaxation" back to a state of preglitch steady-state spin-down, which typically occurs on time scales of days to months (see [10]).

During a glitch, the observed normal fluid would spin up by $\delta\Omega$, while the superfluid would spin down by $\delta\Omega_{\rm s}$, such that angular momentum is conserved. Following [24] and assuming that the moments of inertia are constant during a glitch, we have

$$I_{\rm c}\delta\Omega + I_{\rm s}\delta\Omega_{\rm s} = 0, \tag{9}$$

where typically the observed spin-up is up to $\delta\Omega/\Omega \sim 10^{-6}$ [10]. Note that for fiducial values of $I_{\rm s}/I_{\rm c} \sim 10^{-2}$ this implies that the superfluid spins down by $\delta\Omega_{\rm s}/\Omega \sim -10^{-4}$ during a glitch. Besides this angular-momentum transfer in the glitch, there is some excess energy $E_{\rm glitch}$ left. Assuming that corotation between the two fluids is restored after the glitch, namely $\delta\Omega_{\rm s} = -\Delta\Omega + \delta\Omega$, the glitch excess energy can be shown to be

$$E_{\rm glitch} = \frac{1}{2} I_{\rm s} \delta\Omega_{\rm s}^2 + \frac{1}{2} I_{\rm c} \delta\Omega^2 \approx \frac{1}{2} I_{\rm s} \delta\Omega_{\rm s}^2, \qquad (10)$$

which agrees with the result in [24], where in the last step we assumed the fiducial scales $I_{\rm s}/I_{\rm c} \sim 10^{-2}$ and therefore $\delta\Omega/\delta\Omega_{\rm s} \sim 10^{-2}$, which makes the second term negligible. Using Eq. (9), this can also be expressed as $E_{\rm glitch} \approx \frac{1}{2} I_{\rm c} \delta\Omega \Delta\Omega$. This energy would have to be dissipated directly in the glitch, for example, by exciting oscillation modes, which would radiate GWs on short time scales, as considered in [3,5], or by heating up the NS (e.g. see [30]).

However, after a glitch the NS generally relaxes back to a steady-state spin-down on a time scale of days to months, and therefore some (if not all) of the initial excess energy stored in the faster-rotating superfluid could be driving GW emission on this *relaxation* time scale of $\mathcal{O}$(days-months). The recent studies of post-glitch spin-up of the fluid NS core by nonaxisymmetric Ekman flow [8,9] provide one concrete example for exactly such a mechanism. It is also conceivable that this hidden energy in the superfluid ''flywheel'' can lead to transient-CW emission directly, via an internal instability, without transferring its angular-momentum to the crust first, i.e. without a visible pulsar glitch.

The well-known spin-down upper limit on CW emission from known pulsars [6] assumes that the total steady-state spin-down energy of the pulsar is emitted as GWs. Similarly we can consider a ''superfluid upper limit'' on transient CWs, where the total superfluid excess energy is converted into transient CWs, for example, by sustaining some nonaxisymmetric process that leads to emission at frequencies $f \sim \mathcal{O}(\nu)$. The excess rotational energy E_s of the superfluid is

$$E_s = \frac{1}{2} I_{\rm s} (\Omega_{\rm s}^2 - \Omega^2) \approx 4\pi^2 I_{\rm s} \nu \Delta\nu, \qquad (11)$$

where in the last expression we dropped the second-order energy term (10), which is smaller by $\Delta\Omega/\Omega \approx 10^{-4}$. In the following we return to using the spin frequency ν instead of the angular velocity $\Omega = 2\pi\nu$. Note that this expression happens to be numerically very similar to the ''starquake'' glitch-energy derived under the assumption of a spin-up caused by a reduction in the crust's moment of inertia (e.g. see [5]), which results in $E_{\rm quake} = 2\pi^2 I \nu \delta\nu$, and substituting $\delta\nu = (I_s/I)\Delta\nu$ from (9), we obtain $2E_{\rm quake} = E_{\rm s}$. For fiducial values of $I \sim 10^{38}$ kg m^2, Vela spin frequency $\nu \sim 10$ Hz and a large glitch of $\delta\nu/\nu \sim 10^{-6}$, this yields $E_{\rm s} \sim 4 \times 10^{42}$ erg.

We have no direct observational evidence about the size of the lag $\Delta\Omega = 2\pi\Delta\nu$ or the superfluid angular velocity Ω_s. In the simplest models one typically assumes the lag to be reset to zero after every glitch, starting to build up again due to spin-down $\dot{\nu}$. However, this is not necessarily the case and the lag could accumulate over longer time scales and be correspondingly larger.

To obtain an upper limit estimate on the signal strength, we can equate $E_{\rm GW}$ of Eq. (3) with E_s given in Eq. (11), assuming that the GW is emitted at a frequency $f = 2\nu$ (corresponding to a nonaxisymmetric deformation). This corresponds to the extreme case where the built-up excess superfluid rotational energy drives emission of a transient CW. Combining this with Eq. (6), we can obtain a superfluid transient-CW upper limit estimate in the form

$$\hat{h}_0 \sqrt{T} = \frac{1}{d} \sqrt{\frac{5G}{2c^3} I_{\rm s} \frac{\Delta\nu}{\nu}}. \qquad (12)$$

Alternatively, if we assume that the total excess angular momentum is transferred in a glitch, we can use Eq. (9) to rewrite this in terms of more directly observed glitch quantities, namely

$$\hat{h}_0 \sqrt{T} = \frac{1}{d} \sqrt{\frac{5G}{2c^3} I \frac{\delta\nu}{\nu}}, \qquad (13)$$

where I is the axial NS moment of inertia, and $\delta\nu/\nu$ is the observed glitch spin-up. As a third alternative, we can use the simple two-fluid model for the buildup of the lag, namely $\Delta\nu = |\dot{\nu}|\Delta t$, and parametrize the unknown superfluid moment of inertia via $I_s = \gamma I$ (with fiducial value $\gamma \sim 10^{-2}$), and obtain the relation

$$\hat{h}_0 = \sqrt{\frac{\gamma \Delta t}{T}} h_{\rm sd}, \qquad (14)$$

in terms of the well-known ''spin-down limit'' $h_{\rm sd}$, given by

$$h_{\rm sd} = \frac{1}{d} \sqrt{\frac{5G}{2c^3} I \frac{|\dot{\nu}|}{\nu}}, \qquad (15)$$

which is derived from the assumption that the NS spin-down energy is completely converted into GWs. The relation (14) shows that the superfluid transient-CW upper limit can be similar or even higher than the usual pulsar spin-down limit: a fraction γ of the steady-state spin-down energy is accumulated in the superfluid over an interglitch time scale Δt, and is released on a short time scale T. Assuming $\gamma \sim 10^{-2}$, interglitch periods of $\Delta t \sim 1$ yr and transient-CW time scale of $T \sim 2$ weeks, we see from Eq. (14) that for these values $\hat{h}_0 \approx h_{\rm sd}$. This is interesting as it indicates that the young (and glitching) pulsars with the most promising spin-down upper limits on CW emission might also be the most attractive targets for directed transient-CW searches, for example, the Crab pulsar, Vela, and J0537-69 (e.g. see Fig. 4 in [6]).

III. SIGNAL MODEL FOR TRANSIENT CWS

The family of transient-CW signals considered here is a straightforward generalization of the traditional infinite-duration CW model [13,31], allowing for a finite duration and nontrivial time-evolution of the overall amplitude h_0.

The set $\mathcal{T}$ of extra transient-CW parameters therefore consists of the start-time t^0, characteristic time scale τ, and the type of window function ϖ, i.e. $\mathcal{T} \equiv \{\varpi, t^0, \tau\}$. The transient-CW signal family therefore simply consists of a window function $g_\varpi(t; t^0, \tau)$ applied to the standard CW signal model, namely

$$h^X(t; \mathcal{A}, \lambda, \mathcal{T}) = g_\varpi(t; t^0, \tau)\mathcal{A}^\mu h^X_\mu(t; \lambda), \tag{16}$$

were we use implicit summation over the four amplitudes, $\mu = 1, \ldots 4$, and X is an index over different detectors. The four canonical amplitudes $\mathcal{A}^\mu$ are functions of the CW amplitude h_0, polarization angles $\cos\iota$, ψ and the initial phase ϕ_0, i.e. $\mathcal{A}^\mu = \mathcal{A}^\mu(h_0, \cos\iota, \psi, \phi_0)$. The corresponding basis functions $h^X_\mu(t; \lambda)$ are found, for example, in [32], but their explicit form is not relevant to our discussion here. The set of Doppler parameters λ determines the time evolution of the signal phase, for example, the source sky-position $\hat{n}$ and the GW frequency $f(t)$, which is generally allowed to be slowly varying with time. If the CW source is a neutron star in a binary system, λ would also include the relevant orbital parameters of the system.

In the following we restrict our attention to two simple types of transient-window functions $g_\varpi(t)$, namely, either *rectangular*, denoted as $\varpi = \mathrm{r}$, i.e.

$$g_\mathrm{r}(t; t^0, \tau) \equiv \begin{cases} 1 & \text{if } t \in [t^0, t^0 + \tau] \\ 0 & \text{otherwise,} \end{cases} \tag{17}$$

or *exponentially decaying*, denoted as $\varpi = \mathrm{e}$, namely

$$g_\mathrm{e}(t; t^0, \tau) \equiv \begin{cases} e^{-(t-t^0)/\tau} & \text{if } t \in [t^0, t^0 + 3\tau] \\ 0 & \text{otherwise,} \end{cases} \tag{18}$$

where we somewhat arbitrarily truncated the exponential window at an e-folding of 3, in order to simplify the practical implementation of this window. This truncation gives the window a finite duration of 3τ, and at the truncation point the amplitude has decreased by more than 95% and we can neglect the corresponding loss of SNR.

Following the notation of [21,33], we use boldface to denote multidetector vectors, i.e. $\{\boldsymbol{x}\}^X = x^X$ denotes the set of data-streams $x(t)$ from different detectors X. We can now conveniently absorb the window function $g(t)$ in Eq. (16) into the definition of new transient basis functions $\boldsymbol{h}'_\mu$, namely

$$\boldsymbol{h}'_\mu(t; \lambda, \mathcal{T}) \equiv g_\varpi(t; t^0, \tau)\boldsymbol{h}_\mu(t; \lambda). \tag{19}$$

If we denote θ the set of *all* signal parameters of our search, i.e.

$$\theta \equiv \{\mathcal{A}, \lambda, \mathcal{T}\}, \tag{20}$$

then we can write the transient signal model (16) now more compactly as

$$\boldsymbol{h}(t; \theta) = \mathcal{A}^\mu \boldsymbol{h}'_\mu(t; \lambda, \mathcal{T}). \tag{21}$$

IV. DETECTION METHOD: ODDS RATIO

A. Hypothesis testing framework

Based on observed data $\boldsymbol{x}$, we want to decide between two hypotheses: under the noise hypothesis $\mathcal{H}_\mathrm{G}$ the observed data consists only of Gaussian stationary noise $\boldsymbol{n}$, and under the signal hypothesis $\mathcal{H}_\mathrm{S}$ the data contains in addition a transient-CW signal $\boldsymbol{h}(t; \theta)$ of Eq. (21), namely

$$\mathcal{H}_\mathrm{G}\colon \boldsymbol{x}(t) = \boldsymbol{n}(t), \tag{22}$$

$$\mathcal{H}_\mathrm{S}\colon \boldsymbol{x}(t) = \boldsymbol{n}(t) + \boldsymbol{h}(t; \theta) \quad \text{for any } \theta \in \mathbb{P}, \tag{23}$$

where $\mathbb{P}$ denotes the signal parameter space. Note that the signal hypothesis (23) is incomplete without the specification of a probability distribution for the unknown signal parameters θ over their parameter space $\mathbb{P}$, i.e. a *prior* probability $P(\theta|\mathcal{H}_\mathrm{S})$. For simplicity our notation does not explicitly distinguish between proper probabilities and probability densities. This difference is implicit in the type of argument, namely, whether it is discrete, such that $\sum_i P(x_i|\mathcal{I}) = 1$, or continuous, where $\int P(x|\mathcal{I})dx = 1$. Furthermore, we sometimes (but not always) explicitly state $\mathcal{I}$ as a conditional in probability statements, denoting the full set of remaining relevant prior model assumptions that the probability statement depends on.

B. Gaussian noise and scalar product

For quasisinusoidal CWs (16) in stationary Gaussian noise we can define a multidetector scalar product [21], using the notation of [33], as

$$(\boldsymbol{x}|\boldsymbol{y}) \equiv 2\sum_X S_X^{-1}(f) \int x^X(t)y^X(t)dt, \tag{24}$$

where $S_X(f)$ is the (stationary) single-sided noise power spectral density in detector X, which is assumed constant over a narrow frequency band around the signal frequency f. This allows us to write the likelihood for the data $\boldsymbol{x}$ in the Gaussian noise case (22) as

$$P(\boldsymbol{x}|\mathcal{H}_\mathrm{G}) = \kappa e^{-(1/2)(\boldsymbol{x}|\boldsymbol{x})}, \tag{25}$$

where κ is a normalization constant. In the presence of a signal $\boldsymbol{h}(t; \theta)$ with parameters θ, subtracting this signal from the data $\boldsymbol{x}$ results again in pure Gaussian noise $\boldsymbol{n}$, i.e. $\boldsymbol{n} = \boldsymbol{x} - \boldsymbol{h}(\theta)$, and therefore

$$P(\boldsymbol{x}|\mathcal{H}_\mathrm{S}, \theta) = \kappa e^{-(1/2)(\boldsymbol{x}-\boldsymbol{h}(\theta)|\boldsymbol{x}-\boldsymbol{h}(\theta))}. \tag{26}$$

The likelihood for the data $\boldsymbol{x}$ containing *any* signal $\boldsymbol{h}(t; \theta)$ with $\theta \in \mathbb{P}$ drawn from the prior $P(\theta|\mathcal{H}_\mathrm{S})$ can easily be obtained (e.g. see [34]) as

$$P(\boldsymbol{x}|\mathcal{H}_\mathrm{S}) = \int_\mathbb{P} P(\boldsymbol{x}|\mathcal{H}_\mathrm{S}, \theta)P(\theta|\mathcal{H}_\mathrm{S})d\theta, \tag{27}$$

which is often referred to as the *marginal likelihood* (and sometimes as *evidence*).

C. Odds ratio and Bayes factor

We can express the *odds ratio* O_{SG} between signal- and Gaussian-noise hypothesis for the observed data $\boldsymbol{x}$ as

$$O_{\mathrm{SG}}(\boldsymbol{x}) \equiv \frac{P(\mathcal{H}_{\mathrm{S}}|\boldsymbol{x})}{P(\mathcal{H}_{\mathrm{G}}|\boldsymbol{x})} = B_{\mathrm{SG}}(\boldsymbol{x})\frac{P(\mathcal{H}_{\mathrm{S}}|\mathcal{I})}{P(\mathcal{H}_{\mathrm{G}}|\mathcal{I})}, \tag{28}$$

where we used Bayes' theorem, namely $P(a|b)P(b|\mathcal{I}) = P(b|a)P(a|\mathcal{I})$, to express $O_{\mathrm{SG}}(\boldsymbol{x})$ as a product of the prior hypothesis odds and the so-called *Bayes factor* $B_{\mathrm{SG}}(\boldsymbol{x})$, defined as

$$B_{\mathrm{SG}}(\boldsymbol{x}) \equiv \frac{P(\boldsymbol{x}|\mathcal{H}_{\mathrm{S}})}{P(\boldsymbol{x}|\mathcal{H}_{\mathrm{G}})} = \int_{\mathbb{P}} \mathcal{L}(\boldsymbol{x};\theta)P(\theta|\mathcal{H}_{\mathrm{S}})d\theta, \tag{29}$$

where we used Eqs. (25) and (26) and defined the standard *likelihood ratio* as

$$\mathcal{L}(\boldsymbol{x};\theta) \equiv \frac{P(\boldsymbol{x}|\mathcal{H}_{\mathrm{S}},\theta)}{P(\boldsymbol{x}|\mathcal{H}_{\mathrm{G}})} = e^{(\boldsymbol{x}|\boldsymbol{h}(\theta))-(1/2)(\boldsymbol{h}(\theta)|\boldsymbol{h}(\theta))}. \tag{30}$$

Inserting the transient-CW signal model (21), we can write this as

$$\log\mathcal{L}(\boldsymbol{x};\theta) = \mathcal{A}^{\mu}x'_{\mu} - \frac{1}{2}\mathcal{A}^{\mu}\mathcal{M}'_{\mu\nu}\mathcal{A}^{\nu}, \tag{31}$$

where we defined

$$x'_{\mu}(\lambda,\mathcal{T}) \equiv (\boldsymbol{x}|\boldsymbol{h}'_{\mu}), \tag{32}$$

$$\mathcal{M}'_{\mu\nu}(\lambda,\mathcal{T}) \equiv (\boldsymbol{h}'_{\mu}|\boldsymbol{h}'_{\nu}), \tag{33}$$

generalizing the corresponding CW quantities, e.g. see [34], to the transient signal model.

D. Maximum likelihood: the $\mathcal{F}$-statistic

Contrary to the marginalization in Eq. (29) over unknown parameters θ, which follows from the axioms of probability, the orthodox "maximum-likelihood" approach consists of an *ad-hoc maximization* of the likelihood ratio $\mathcal{L}$ over the unknown parameters θ, i.e. we define the maximum-likelihood statistic as

$$\mathcal{L}_{\mathrm{ML}}(\boldsymbol{x}) \equiv \max_{\theta}\mathcal{L}(\boldsymbol{x};\theta). \tag{34}$$

Given the explicitly quadratic dependency on $\mathcal{A}^{\mu}$ in Eq. (31), this maximization can be performed explicitly, which results in

$$\ln\mathcal{L}_{\mathrm{ML}}(\boldsymbol{x}) = \max_{\{\lambda,\mathcal{T}\}}\mathcal{F}(\boldsymbol{x};\lambda,\mathcal{T}), \tag{35}$$

were we encounter the well-known "$\mathcal{F}$-statistic", which was first derived in [13] for CW signals. In the present transient-CW case, the transient $\mathcal{F}$-statistic is obtained explicitly as

$$2\mathcal{F}(\boldsymbol{x};\lambda,\mathcal{T}) \equiv x'_{\mu}\mathcal{M}'^{\mu\nu}x'_{\nu}, \tag{36}$$

where $\mathcal{M}'^{\mu\nu}$ is defined as the inverse matrix of $\mathcal{M}'_{\mu\nu}$ of Eq. (33).

If the data $\boldsymbol{x}(t)$ contains a signal $\boldsymbol{s}(t;\theta_s)$, such that $\boldsymbol{x} = \boldsymbol{n} + \boldsymbol{s}(\theta_s)$, then we can write Eq. (32) as

$$x'_{\mu} = n'_{\mu} + s'_{\mu}, \tag{37}$$

with the obvious definitions $n'_{\mu}(\lambda,\mathcal{T}) \equiv (\boldsymbol{n}|\boldsymbol{h}'_{\mu})$ and

$$s'_{\mu}(\theta_s;\lambda,\mathcal{T}) \equiv (\boldsymbol{s}(\theta_s)|\boldsymbol{h}'_{\mu}(\lambda,\mathcal{T})), \tag{38}$$

which depends both on the signal parameters $\theta_s = \{\mathcal{A}_s,\lambda_s,\mathcal{T}_s\}$ and the matched-filter parameters $\{\lambda,\mathcal{T}\}$ of the "template".

Gaussian detector noise $\boldsymbol{n}(t)$ has zero mean, i.e. $E[\boldsymbol{n}] = 0$, and therefore $E[n'_{\mu}] = 0$ and $E[x'_{\mu}] = s'_{\mu}$. One can also show that the corresponding covariance is $E[n'_{\mu}n'_{\nu}] = \mathcal{M}'_{\mu\nu}(\lambda,\mathcal{T})$, and therefore

$$E[x'_{\mu}x'_{\nu}] = \mathcal{M}'_{\mu\nu} + s'_{\mu}s'_{\nu}. \tag{39}$$

Using this together with Eq. (36) one can further show that $2\mathcal{F}$ follows a χ^2-distribution with four degrees of freedom and noncentrality parameter ρ^2, i.e.

$$E[2\mathcal{F}] = 4 + \rho^2, \tag{40}$$

where the signal-to-noise ratio (SNR) ρ is expressible as

$$\rho^2(\theta_s;\lambda,\mathcal{T}) = s'_{\mu}\mathcal{M}'^{\mu\nu}s'_{\nu}. \tag{41}$$

We see from Eq. (38) that the SNR will depend in a complicated way on the offset between signal parameters θ_s and template parameters $\{\lambda,\mathcal{T}\}$ (see [33] for the non-transient CW case). In the special case of perfectly matched template parameters, i.e. $\lambda = \lambda_s$ and $\mathcal{T} = \mathcal{T}_s$, we obtain the so-called "optimal SNR" ρ_o, which can be expressed as

$$\rho_o^2(\theta_s) = \mathcal{A}_s^{\mu}\mathcal{M}'_{\mu\nu}(\lambda_s,\mathcal{T}_s)\mathcal{A}_s^{\nu} = (\boldsymbol{s}|\boldsymbol{s}). \tag{42}$$

E. Choice of signal priors

In order to fully define the Bayes factor (29), we need to provide a complete signal hypothesis including the prior probabilities $P(\theta|\mathcal{H}_{\mathrm{S}})$ for the signal parameters $\theta = \{\mathcal{A},\lambda,\mathcal{T}\}$.

1. Prior on Doppler-parameters λ

For simplicity we assume that the Doppler-parameters λ are independent of amplitude- and transient parameters $\{\mathcal{A},\mathcal{T}\}$, so we can factor the full parameter prior $P(\theta|\mathcal{H}_{\mathrm{S}})$ into

$$P(\theta|\mathcal{H}_{\mathrm{S}}) = P(\lambda|\mathcal{H}_{\mathrm{S}})P(\mathcal{A},\mathcal{T}|\mathcal{H}_{\mathrm{S}}), \tag{43}$$

and so the Bayes factor (29) now reads as

$$B_{\mathrm{SG}}(\boldsymbol{x}) = \int B_{\mathrm{SG}}(\boldsymbol{x};\lambda)P(\lambda|\mathcal{H}_{\mathrm{S}})d^n\lambda, \tag{44}$$

in terms of a "targeted" Bayes factor $B_{\mathrm{SG}}(\boldsymbol{x};\lambda)$ for a single Doppler point λ, namely

$$B_{\mathrm{SG}}(\boldsymbol{x};\lambda) \equiv \int \mathcal{L}(\boldsymbol{x};\mathcal{A},\mathcal{T},\lambda) P(\mathcal{A},\mathcal{T}|\mathcal{H}_{\mathrm{S}}) d\mathcal{A}\, d\mathcal{T}. \tag{45}$$

In the following we will focus exclusively on the *targeted* Bayes factor, and sometimes drop λ for simplicity of notation. The generalization to parameter searches over λ is straightforward as given by Eq. (44).

2. Prior on transient parameters $\mathcal{T}$

Astrophysically it would make sense to assume that the amplitude h_0 of a transient CW is related to its time scale τ. For example, it might be reasonable to suspect that stronger transient GWs have shorter duration and vice-versa, according to some prior on the total transient-CW energy emitted and on the distance of such sources, cf. Sec. II. However, given the current lack of concrete astrophysical predictions to base such priors on, we assume a simple independent prior $P(\mathcal{T}|\mathcal{H}_{\mathrm{S}})$ for the transient signal parameters $\mathcal{T}$. A naturally simple choice consists of independent uniform priors within some appropriate time windows, i.e.

$$\begin{aligned} P(t^0|\mathcal{H}_{\mathrm{S}}) &= U(t^0_{\mathrm{min}}, t^0_{\mathrm{min}} + \Delta t^0), \\ P(\tau|\mathcal{H}_{\mathrm{S}}) &= U(\tau_{\mathrm{min}}, \tau_{\mathrm{min}} + \Delta\tau), \end{aligned} \tag{46}$$

where we write uniform probability densities as $U(a,b) \equiv 1/(b-a)$ for the parameter falling inside $[a,b]$ and zero otherwise. We would also need priors on the window *type*, e.g. $P(\varpi = \mathrm{r}|\mathcal{H}_{\mathrm{S}}) = P(\varpi = \mathrm{e}|\mathcal{H}_{\mathrm{S}}) = \frac{1}{2}$, in order to marginalize over ϖ. However, for simplicity we will often assume a particular window-type ϖ as given and only marginalize over t^0, τ. The effect of assuming an incorrect window-function within $\varpi \in \{\mathrm{r},\mathrm{e}\}$ is numerically studied in Sec. V B, and appears to entail only mild losses of detection power.

3. Prior on amplitude-parameters $\mathcal{A}$

Physically reasonable priors on the angle parameters $\{\cos\iota, \psi, \phi_0\}$ are relatively easy to obtain if we assume ignorance about the orientation of the spinning neutron star (e.g. see [34] for a more detailed discussion), namely, by symmetry one can obtain

$$\begin{aligned} P(\cos\iota|\mathcal{H}_{\mathrm{S}}) &= U(-1,1), \\ P(\psi|\mathcal{H}_{\mathrm{S}}) &= U\left(-\frac{\pi}{4},\frac{\pi}{4}\right) \\ P(\phi_0|\mathcal{H}_{\mathrm{S}}) &= U(0,2\pi). \end{aligned} \tag{47}$$

The choice of prior for the overall amplitude parameter h_0 is less obvious: one could choose a scale-free Jeffrey's prior, or a simpler uniform prior on some physically meaningful domain.

The downside of the isotropic amplitude priors (47) is that the resulting marginalization over $\mathcal{A}$ cannot be performed analytically, and computing the Bayes factor would require a numerical integration over $\mathcal{A}$ in every point $\lambda, \mathcal{T}$, which will be computationally prohibitive. However, as shown in [34], by using an unphysical *uniform prior* on the 4-*vector* $\mathcal{A}^\mu$, one can *analytically* marginalize over $\mathcal{A}$ and obtain a Bayes factor (45) expressed in terms of the well-known $\mathcal{F}$-statistic (36).

This choice has the major advantage of simplicity and computational efficiency, while incurring only small losses in detection power compared to the more physical prior (47), as shown in [34]. Given that fast and mature codes exist to compute $\mathcal{F}$ on real detector data (cf. [20,35]), we choose this prior as a convenient practical approximation.

However, as will be seen in the following, the original formulation of the uniform $\mathcal{A}^\mu$-prior in [34] leads to a somewhat weak detection statistic for transient-CW signals and can be improved by a minor modification. The original "constant-h_{max}" prior was defined as

$$P(\mathcal{A}^\mu|\mathcal{H}_{\mathrm{S}}) = \begin{cases} C & \text{if } h_0(\mathcal{A}) < h_{\mathrm{max}}, \\ 0 & \text{otherwise}, \end{cases} \tag{48}$$

where h_{max} is a maximum cutoff amplitude needed in order to normalize the prior.

The $\mathcal{A}$-integration in Eq. (45) can be performed analytically with this prior, provided the data does not cause the likelihood to peak close to the upper cutoff h_{max}. Namely, if the value of the integrand $\mathcal{L}(x;\mathcal{A})$ is already negligible at the cutoff boundary, we can extend the domain to infinity and obtain a four-dimensional Gaussian integral, namely

$$\int^{h_0<h_{\mathrm{max}}} \mathcal{L}(\boldsymbol{x};\mathcal{A},\mathcal{T}) C d^4\mathcal{A} \approx \frac{(2\pi)^2 C}{\sqrt{|\mathcal{M}|}} e^{\mathcal{F}(x;\mathcal{T})}, \tag{49}$$

where $|\mathcal{M}| \equiv \det\mathcal{M}$ is the determinant of the matrix $\mathcal{M}_{\mu\nu}$ of Eq. (33). Using this approximation, we can therefore write the Bayes factor (45) as

$$B_{\mathrm{SG}}(\boldsymbol{x};\lambda) \approx \int \frac{(2\pi)^2 C}{|\mathcal{M}|^{1/2}} e^{\mathcal{F}(x;\mathcal{T})} P(\mathcal{T}|\mathcal{H}_{\mathrm{S}}) d\mathcal{T}. \tag{50}$$

The Jacobian J of the coordinate transformation $\mathcal{A}^\mu = \mathcal{A}^\mu(h_0,\cos\iota,\psi,\phi_0)$ is $J = \frac{1}{4}h_0^3(1-\cos^2\iota)^3$ (see [34]), and therefore $d^4\mathcal{A} = J dh_0 d\cos\iota d\psi d\phi_0$. We can now determine the normalization constant C as

$$1 = \int P(\mathcal{A}|\mathcal{H}_{\mathrm{S}}) d^4\mathcal{A} = C\frac{2\pi^2}{35} h_{\mathrm{max}}^4, \tag{51}$$

and obtain the constant-h_{max} Bayes factor explicitly as

$$B_{\mathrm{SG}}|_{h_{\mathrm{max}}} = \frac{70}{h_{\mathrm{max}}^4 \Delta t^0 \Delta\tau} \int \frac{e^{\mathcal{F}}}{|\mathcal{M}|^{1/2}} dt^0 d\tau, \tag{52}$$

where we assumed a fixed window type ϖ. The antenna-pattern weighting factor $|\mathcal{M}|^{-1/2}$ generally depends on the sky-position $\hat{n}$, the transient-window type ϖ, start-time t^0

and time scale τ, as seen from Eq. (33). In the case of a fully targeted search with fixed t^0, τ and λ, as discussed in [34], the weighting factor is constant and plays no role for the power of the detection statistic. The fully targeted Bayes factor is therefore strictly equivalent to the detection power of the $\mathcal{F}$-statistic.

Interestingly, in the case of transient signals we find that the presence of this antenna-pattern weighting factor in Eq. (52) seems to *degrade* the detection power, and for some choices of parameter ranges $B_{\mathrm{SG}}|_{h_{\max}}$ performs *worse* than the maximum-likelihood statistic (35), namely

$$\mathcal{F}_{\max}(\boldsymbol{x};\boldsymbol{\varpi}) \equiv \max_{\{t^0,\tau\}} \mathcal{F}(\boldsymbol{x};\boldsymbol{\varpi},t^0,\tau). \tag{53}$$

However, a simple modification of the cutoff boundary in Eq. (48) allows us to eliminate the antenna-pattern weighting factor $|\mathcal{M}|^{-1/2}$ in Eq. (50). Namely, by introducing an "SNR-scale" $\hat{\rho} \equiv h_0|\mathcal{M}|^{1/8}$, and using a cutoff $\hat{\rho} < \hat{\rho}_{\max}$ as the outer boundary of the domain instead of the amplitude-cutoff $h_0 < h_{\max}$, the modified *ad-hoc* "$\mathcal{F}$-statistic prior" is now

$$P(\mathcal{A}^\mu|\mathcal{H}_{\mathrm{S}},\mathcal{T}) = \begin{cases} \hat{C} & \text{if } \hat{\rho}(\mathcal{A}) < \hat{\rho}_{\max}, \\ 0 & \text{otherwise,} \end{cases} \tag{54}$$

which results in the new transient Bayes factor

$$B_{\mathcal{F}}(\boldsymbol{x};\lambda,\boldsymbol{\varpi}) = \frac{70}{\hat{\rho}_{\max}^4 \Delta t^0 \Delta\tau} \int e^{\mathcal{F}(\boldsymbol{x};\lambda,\mathcal{T})} dt^0 d\tau, \tag{55}$$

where in the following we denote $B_{\mathcal{F}} \equiv B_{\mathrm{SG}}|_{\hat{\rho}_{\max}}$ as our detection statistic of choice.

As discussed in Sec. VA, numerical simulations show that this detection statistic is more powerful than both $B_{\mathrm{SG}}|_{h_{\max}}$ as well as the orthodox maximum-likelihood statistic. More work would be required to study these priors in more detail, and to develop a more physical choice of priors that would be equally practical.

F. SNR loss due to rectangular-window offsets

Let us consider the effect of an offset in transient-window parameters $\mathcal{T}$ from the signal parameters $\mathcal{T}_s$, assuming perfectly-matched Doppler parameters, i.e. $\lambda = \lambda_s$. The dependence of the matched-filter SNR (41) on Doppler offsets $\lambda_s - \lambda$ has already been studied in great detail, e.g. see [31,33,36].

In the following we drop λ and write Eq. (38) more explicitly as

$$s'_\mu(\mathcal{A}_s,\mathcal{T}_s,\mathcal{T}) = \mathcal{A}_s^\nu(\boldsymbol{h}'_\nu(\mathcal{T}_s)|\boldsymbol{h}'_\mu(\mathcal{T})), \tag{56}$$

and using Eqs. (19) and (24), we can further expand this as

$$s'_\mu = \frac{2\mathcal{A}_s^\nu}{S} \int h_\nu(t) h_\mu(t) g_{\varpi_s}(t;t_s^0,\tau_s) g_{\varpi}(t;t^0,\tau) dt, \tag{57}$$

where we have omitted multidetector summation for simplicity of notation. It will be difficult to make further analytic progress with this expression in the general case, but we can analyze the interesting special case of the rectangular-window function (17), i.e. $\varpi_s = \varpi = \mathrm{r}$. In this case the window functions simply truncate the integral, and so we obtain

$$s'_\mu = \frac{2\mathcal{A}_s^\nu}{S} \int_{t_0}^{t_1} h_\nu(t) h_\mu(t) dt, \tag{58}$$

where we defined $t_0 \equiv \max(t_s^0, t^0)$, and $t_1 \equiv \min((t_s^0 + \tau_s), (t^0 + \tau))$. Note that $[t_0, t_1]$ denotes the rectangular overlap between $\mathcal{T}_s$ and $\mathcal{T}$, and in the above integral we assumed $t_1 \geq t_0$, otherwise the expression is zero.

In order to simplify this even further, we note that the antenna-pattern matrix in Eq. (33) can be written as

$$\mathcal{M}'_{\mu\nu} = \frac{2\tau}{S} m_{\mu\nu}, \tag{59}$$

where we defined $m_{\mu\nu} = \langle h_\mu h_\nu \rangle_S$, in terms of the multi-detector time-average $\langle \ldots \rangle_S$ introduced in [Eq. (59)] in [33]. The antenna-pattern functions are periodic with period of a sidereal day, and so the average $m_{\mu\nu}$ will be weakly oscillatory and converges to a constant for $\tau \gg 1\mathrm{d}$. Let us therefore approximate $m_{\mu\nu}$ as constant for fixed λ, i.e. $m_{\mu\nu} \approx \bar{m}_{\mu\nu}$, which allows us to write Eq. (58) as

$$s'_\mu \approx \frac{2\tau_\Delta}{S} \mathcal{A}_s^\nu \bar{m}_{\nu\mu}, \quad \text{with} \quad \tau_\Delta \equiv [t_1 - t_0]^+, \tag{60}$$

where $[\ldots]^+$ is the positivity operator, defined as $[x]^+ = x$ for $x > 0$ and zero otherwise. Therefore τ_Δ is the length of overlap between signal and template windows. Using $\mathcal{M}'_{\mu\nu} \approx (2\tau/S)\bar{m}_{\mu\nu}$, we obtain the mismatched SNR after substituting into Eq. (41), namely

$$\rho^2 \approx \frac{\tau_\Delta^2}{\tau} \frac{2}{S} \mathcal{A}_s^\mu \bar{m}_{\mu\nu} \mathcal{A}_s^\nu = \frac{\tau_\Delta^2}{\tau\tau_s} \rho_o^2, \tag{61}$$

in terms of the perfectly-matched "optimal SNR" ρ_o, defined in Eq. (42). Note that always $\tau_\Delta \leq \min(\tau, \tau_s)$, and equality only holds in the perfect-match case. Expressing this in terms of the usual definition of mismatch m, we obtain

$$m(t_s^0,\tau_s;t^0,\tau) \equiv \frac{\rho_o^2 - \rho^2}{\rho_o^2} \approx 1 - \frac{\tau_\Delta^2}{\tau\tau_s}. \tag{62}$$

The behavior of this approximate mismatch function and the corresponding measured SNR loss is illustrated in Fig. 1 for a start-time $t^0 = 5$ days and duration $\tau = 5$ days [37]. Contrary to the well-known mismatch behavior in Doppler parameters λ, the transient mismatch metric is not differentiable at the perfect-match point $\tau = \tau_s = \tau_\Delta$, as seen in Eq. (62) and Fig. 1, where the mismatch has a kink. Therefore we cannot Taylor-expand the mismatch around the signal location and define a metric tensor from the second-order quadratic form. Furthermore, the mismatch (62) depends not only on the parameter offsets, but also on the *actual value* of the signal duration τ_s. Therefore the mismatch function is not constant over the

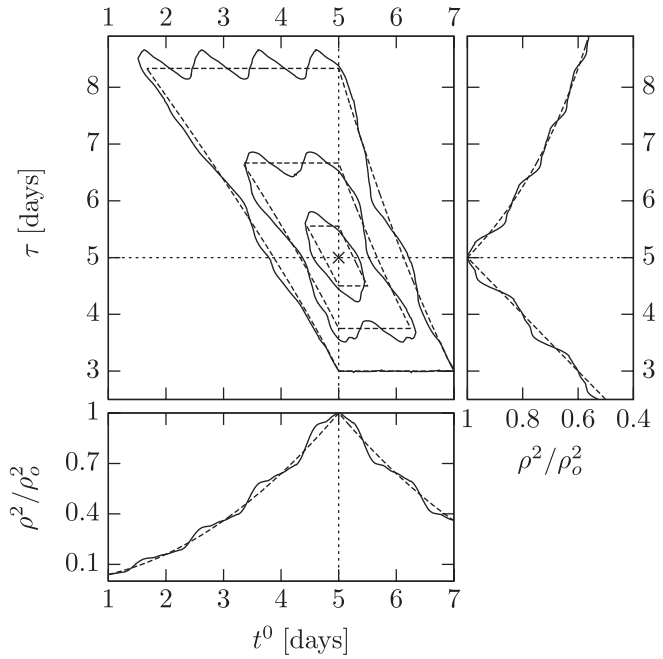

FIG. 1. Approximate mismatch (62) (dashed lines) and measured SNR loss (solid lines) as functions of the rectangular-window offsets $\{t^0, \tau\}$. The top left plot shows mismatch contour lines at $m = 0.90, 0.75, 0.60$, respectively, and the side-panels show cross-sections of ρ^2/ρ_o^2 at fixed $t^0 = t_s^0$ (right panel), and at fixed $\tau = \tau_s$ (bottom panel).

parameter space. However, it is interesting to note that the mismatch behaves close to linearly around the point of perfect match, and we can obtain the first-order variation of the mismatch (62) as

$$dm(\mathcal{T}_s) = 2\frac{|dt^0|}{\tau_s} + \frac{|d\tau|}{\tau_s}, \tag{63}$$

which shows that the mismatch increases twice as fast for offsets in start-time t^0 than for offsets in time scale τ, which is also seen in Fig. 1. Also, as seen in Fig. 1, the parameters t^0 and τ are correlated, and the isomismatch curves close to the signal are straight lines with steepness $|d\tau/dt^0| = -2$, as seen from Eq. (63).

These properties are important for covering the transient-parameter space with a template bank, where one tries to use the smallest number of templates while guaranteeing a certain minimum match for any signal within the template-space. For our present purpose it will be sufficient to make sure that the finite step-sizes in t^0, τ are *fine enough* to ensure a reasonable approximation to the integral (55). The worst-case mismatch (62) occurs for the shortest time scale τ_s, so for $\tau_{\min} = 0.5$ days, say, and a time-sampling in steps of $T_{\mathrm{SFT}} = 1800$ sec, the worst-case mismatch will be bounded by $m \lesssim T_{\mathrm{SFT}}/\tau_{\min} \approx 4\%$.

G. Semicoherent Bayes factor

Increasing the coherent integration time (or in our case, the maximal time scale τ) in wide parameter-space CW searches over unknown Doppler parameters λ typically results in a dramatic increase in computing cost. The reason is that the likelihood function $\mathcal{F}$ becomes increasingly finely structured over Doppler parameters λ, such that more and more templates need to be sampled in order to cover the parameter space (see [31,33,36]). This feature severely limits the feasible coherence time to $\sim\mathcal{O}(\mathrm{days})$ for fully coherent searches over unknown Doppler parameters.

The usual approach to this problem is to abandon fully coherent integration of Eq. (24) over the whole observation time and instead adopt a "semicoherent" approach: this is typically achieved by relaxing the constraint of a consistent signal phase over the whole lifetime. Namely, one splits the observation time T_{obs} into N *segments* of duration ΔT, such that $T_{\mathrm{obs}} = N\Delta T$, and requires phase-coherence only over each segment ΔT. The initial phase ϕ_0 is part of the set of four amplitude-parameters $\mathcal{A}$, but for simplicity one relaxes the consistency-constraints for all four amplitude-parameters $\mathcal{A}^\mu$ across different data-segments $\boldsymbol{x}_{(i)}$ with $i = 1 \ldots N$. This corresponds to replacing the four unknown amplitude parameters $\mathcal{A}^\mu$ by $N \times 4$ unknown amplitude parameters $\{\mathcal{A}^\mu_{(i)}\}_{i=1}^N$. Using the product rule for joint probabilities of independent events, namely $P(\{\boldsymbol{x}_{(i)}\}|\ldots) = \prod_{i=1}^N P(\boldsymbol{x}_{(i)}|\ldots)$, we can express the corresponding semicoherent transient Bayes factor (29) as

$$B^{\mathrm{sc}}_{\mathrm{SG}}(\boldsymbol{x}) = \int d\lambda \int d\mathcal{T}\, P(\lambda, \mathcal{T}|\mathcal{H}_{\mathrm{S}}) \times \prod_{i=1}^{N} \int \mathcal{L}(\boldsymbol{x}_{(i)}; \mathcal{A}_{(i)}, \lambda, \mathcal{T}) P(\mathcal{A}_{(i)}|\mathcal{H}_{\mathrm{S}}) d^4\mathcal{A}_{(i)}. \tag{64}$$

Using the $\mathcal{F}$-statistic prior (54), we obtain

$$B^{\mathrm{sc}}_{\mathcal{F}}(\boldsymbol{x}; \lambda, \boldsymbol{\varpi}) = \frac{1}{\Delta t^0 \Delta\tau}\left(\frac{70}{\hat{\rho}^4_{\max}}\right)^N \int e^{\mathcal{S}_N(\boldsymbol{x};\lambda,\mathcal{T})} dt^0 d\tau, \tag{65}$$

where we defined the semicoherent sum $\mathcal{S}_N$ as

$$\mathcal{S}_N(\boldsymbol{x}; \lambda, \mathcal{T}) \equiv \sum_{i=1}^{N} \mathcal{F}(\boldsymbol{x}_{(i)}; \lambda, \mathcal{T}). \tag{66}$$

The semicoherent sum can be shown to require substantially fewer templates in parameter space for the same amount of data analyzed e.g. see [38,39], and can in many cases achieve a higher sensitivity for wide parameter-space CW searches at fixed computing cost. In order to maximize the available computing power, such searches are performed on large computer clusters and on Einstein@Home [40], a large public distributed computing project (e.g. see [41] for an Einstein@Home search on LIGO data).

Therefore the semicoherent Bayes factor (65) allows for an efficient and computationally feasible transient-CW

search over a wide Doppler parameter space λ of sources with unknown sky-position, frequency and spin-down. More work is required to fully develop and study this approach, as our present analysis is mostly focused on the coherent method. A semicoherent search for transient CWs would be very suitable to be performed on the Einstein@Home computing platform.

V. DETECTION EFFICIENCY

In order to quantify the detection efficiency of different statistics we perform Monte-Carlo studies with simulated Gaussian noise and injected signals with parameters drawn according to their priors. Note that in order to implement this in the most efficient way, we do not perform end-to-end simulations from generated artificial data processed through the whole pipeline. Rather, we more directly synthesize draws of the different statistics from their known distributions, or compute them from draws of intermediate data-products. This is possible because we know the distribution of intermediate data-products in the case of Gaussian noise. The details of this procedure are described in Appendix A 4.

A useful method to compare different detection statistics is to compare their "receiver-operator-characteristic" (ROC), namely, the expected detection probability p_{det} versus the false-alarm probability p_{FA} for some signal population. This follows the spirit of the Neyman-Pearson criterion, which compares the detection probability of different statistics at fixed false-alarm probability.

Note, however, that the traditional way of plotting ROC curves, namely p_{det} versus p_{FA} is somewhat unfortunate. We know *a priori* that any statistic of positive detection power satisfies $p_{\mathrm{det}} > p_{\mathrm{FA}}$, and $p_{\mathrm{det}} = p_{\mathrm{FA}}$ corresponds to a complete random guess. Therefore only the upper triangle of the plot is of any interest, and it is more informative to plot p_{det}-p_{FA} versus p_{FA} instead, which quantifies how much better any statistic performs compared to a random guess.

A. Comparing different amplitude priors

We first consider the detection efficiency of the two Bayes factors $B_{\mathrm{SG}}|_{\hat{\rho}_{\mathrm{max}}}$ of Eq. (55), and $B_{\mathrm{SG}}|_{h_{\mathrm{max}}}$ of Eq. (52). These only differ by their cutoff boundary on the uniform-$\mathcal{A}^{\mu}$ amplitude priors, as discussed in Sec. IV E 3. For simplicity we consistently used a rectangular transient window (17) for both the injections and the search. We performed $N = 10^4$ random draws in the noise case and in the signal case. We estimated the errors on p_{det} using a jackknife estimator on 100 subsets, the resulting estimated 1σ errors in the following ROC curves are always less than $d(p_{\mathrm{det}}) \lesssim 0.02$.

The amplitude parameters are drawn according to their *physical priors* (47). For the first simulation we fixed the optimal SNR (42) to $\rho_o = 3$. This is achieved by rescaling h_0 according to the resulting SNR for drawn values of $\{\cos\iota, \psi, \phi_0\}$. We always use the same ranges for the search and for drawing signal parameters from, and we consider two sets of transient-window parameter ranges, namely

$$\begin{aligned} &\text{range I:} \quad \tau \in [0.1, 0.6]\ \text{days}, \quad t^0 \in [0, 9]\ \text{days}, \\ &\text{range II:} \quad \tau \in [0.5, 2.5]\ \text{days}, \quad t^0 \in [0, 6]\ \text{days}. \end{aligned} \tag{67}$$

The results of this Monte-Carlo simulation are shown in Fig. 2, and we see that the statistic $B_{\mathcal{F}} \equiv B_{\mathrm{SG}}|_{\hat{\rho}_{\mathrm{max}}}$ seems to generally perform better than $B_{\mathrm{SG}}|_{h_{\mathrm{max}}}$, as discussed in Sec. IV E 3. For some choices of transient-window ranges (such as *range I*), the latter actually performs worse than the orthodox maximum-likelihood statistic $\mathcal{F}_{\mathrm{max}}$, as seen in the upper plot in Fig. 2. We also note that the detection probability for signals of equal optimal SNR ρ_o is lower for *range I* than for *range II*. This can be understood from the substantially larger parameter space associated with

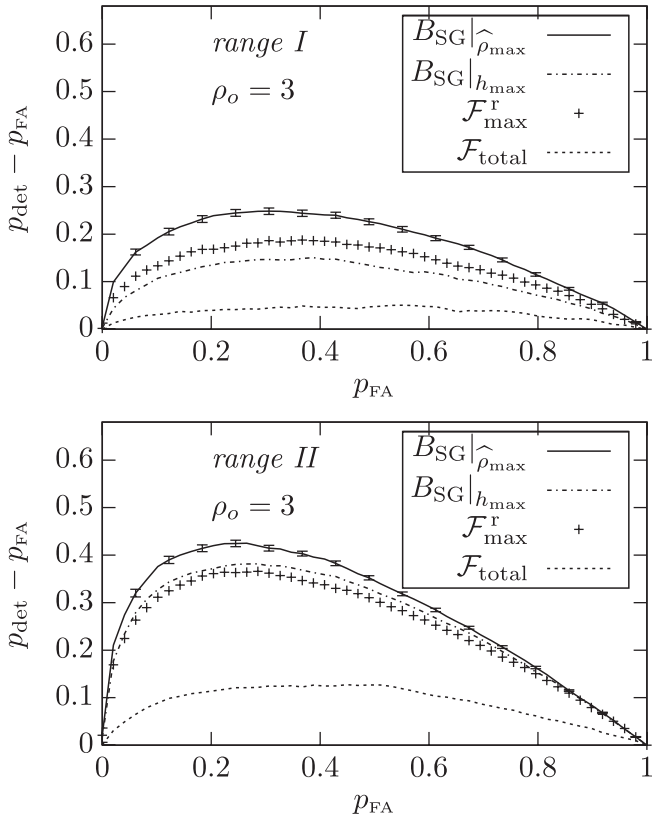

FIG. 2. Detection efficiency of $B_{\mathrm{SG}}|_{\hat{\rho}_{\mathrm{max}}}$, $B_{\mathrm{SG}}|_{h_{\mathrm{max}}}$ and $\mathcal{F}_{\mathrm{max}}$ for injected signals at fixed optimal SNR of $\rho_o = 3$, using a rectangular transient window ($\varpi_s = \mathrm{r}$). The curve labeled '$\mathcal{F}_{\mathrm{total}}$' refers to a standard CW $\mathcal{F}$-statistic search over the full data span. The upper plot corresponds to transient-window parameters drawn from *range I*, while the lower plot is for *range II*, see Eq. (67). We only plot 1σ error-bars for $B_{\mathrm{SG}}|_{\hat{\rho}_{\mathrm{max}}}$, which are representative for the size of the errors on the other curves.

range I, namely $\tau_{\max} = 0.6$ days in $T_{\text{obs}} = 9$ days, as compared to $\tau_{\max} = 2.5$ days in $T_{\text{obs}} = 6$ days for *range II*. Therefore we can fit $T_{\text{obs}}/\tau = 15$ independent rectangular windows into the observation time in *range I*, while for *range II* this factor is only 2.4. This entails more independent trials and therefore a higher false-alarm probability for *range I*, which reduces the detection power for signals of the same SNR.

In order to verify that these qualitative conclusions are not restricted to injections at constant SNR ρ_o, we also performed these Monte-Carlo simulations for injected signals at constant amplitude $h_0/\sqrt{S}$. Note that this results in a wide range of injected signal SNRs ρ_o due to the varying signal durations τ. The results of these simulations are shown in Fig. 3, which qualitatively agree with Fig. 2.

As expected, these results also show that the transient statistics are substantially more sensitive to transient-CW signals than a standard "infinite-duration" CW $\mathcal{F}$-statistic search over the full span T_{obs}. The quantitative advantage in recovered SNR depends on the details of the transient parameter-space, as seen from the mismatch (62), namely, setting $\tau = T_{\text{obs}}$ and $\tau_\Delta = \tau_s$, we find the recovered fraction of ρ_o^2 is roughly proportional to the "duty cycle" τ_s/T_{obs} of the transient signal with respect to the observation time T_{obs}.

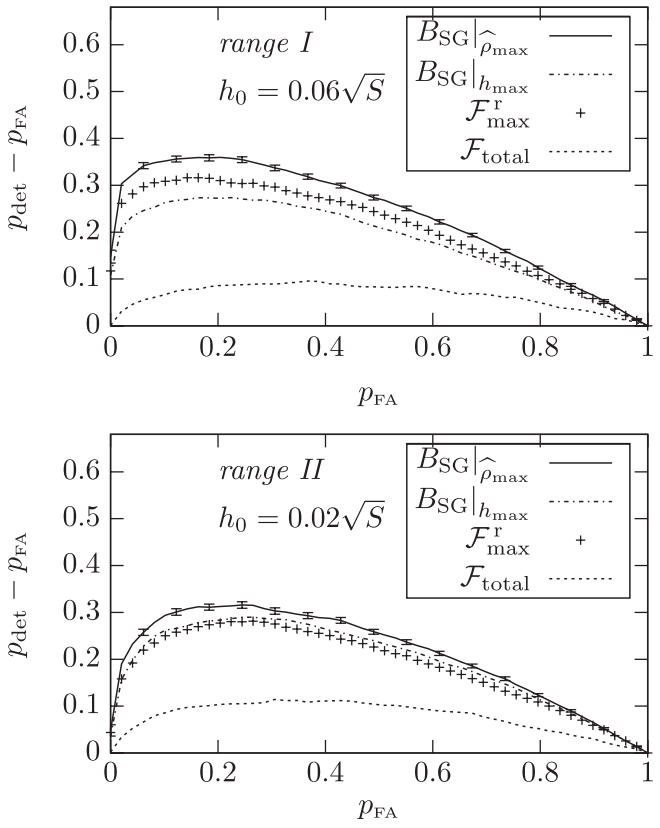

FIG. 3. Same as Fig. 2, with signals injected at fixed amplitude $h_0 = 0.06\sqrt{S}$ and $h_0 = 0.02\sqrt{S}$, respectively.

B. Comparing rectangular and exponential windows

Another question of interest is how robust the detection statistic $B^{\varpi}_{\mathcal{F}}$ is, which assumes a particular transient-window type ϖ, if the transient signal actually has a different window type ϖ_s. We cannot answer this question in general, but it is instructive to study the simple case of type-mismatch between the rectangular and exponential transient-window types. We inject signals with either rectangular (Eq. (17)) or exponential (Eq. (18)) transient window, and we perform the search using both rectangular and exponential transient windows, respectively. The Monte-Carlo parameters are the same as in the previous section, and we only show the results for the transient parameters *range II* of Eq. (67), the qualitative conclusions are the same for *range I*. The results of these Monte-Carlo simulations are shown in Fig. 4. We see that using the wrong window-type, i.e. $\varpi \neq \varpi_s$, results in a loss of detection power in $B_{\mathcal{F}}$, as expected. However, the loss is quite moderate, which in practice would favor a search

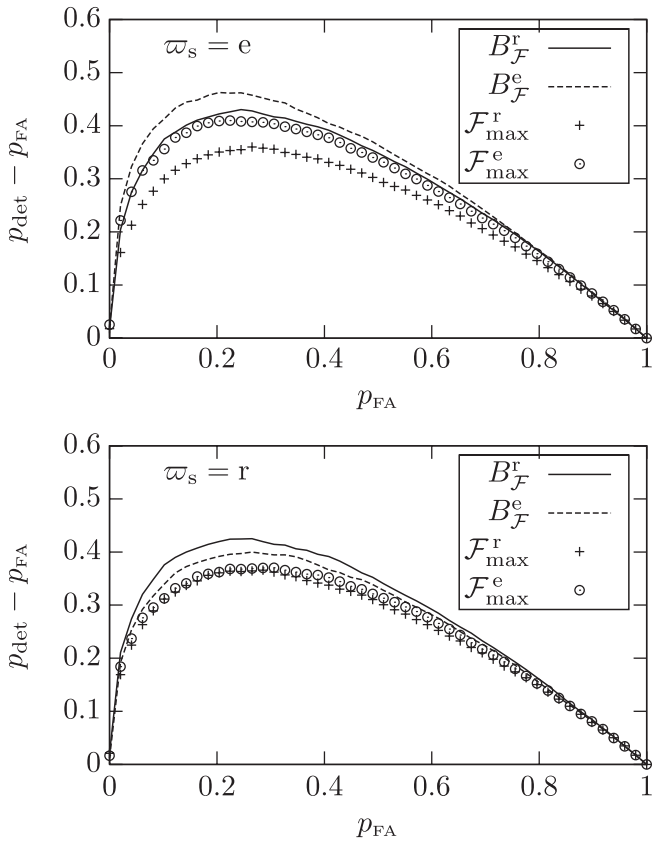

FIG. 4. Detection efficiency for different transient-window *types*. The plot shows detection power of the Bayes factor $B^{\varpi}_{\mathcal{F}}$ and maximum-likelihood statistic $\mathcal{F}^{\varpi}_{\max}$ for assumed window-type ϖ in the case of injected signals of window type ϖ_s. Transient parameters were drawn from *range II* of Eq. (67) with a fixed SNR of $\rho = 3$. The upper plot is for *rectangular* injected transients ($\varpi_s = \text{r}$), while the lower plot is for exponential injected transients ($\varpi_s = \text{e}$).

assuming the simpler and much more computationally efficient rectangular-window type, $\varpi = \mathrm{r}$ (cf. Sec. A 3), if the search is computationally limited. If computing cost is not an issue, for example, in a targeted search for known pulsars, one could perform both searches and then marginalize over ϖ. More importantly, however, these results suggest that we can hope to be reasonably sensitive also to transient signals with *different* time evolutions not contained in $\varpi \in \{\mathrm{r}, \mathrm{e}\}$.

It is interesting to note in Fig. 4 that in the case of injecting rectangular-window signals, i.e. $\varpi_s = \mathrm{r}$, the maximum-likelihood $\mathcal{F}^{\mathrm{e}}_{\mathrm{max}}$ assuming an exponential window, can outperform the more correct $\mathcal{F}^{\mathrm{r}}_{\mathrm{max}}$-statistic. A similar effect is seen in the corresponding simulations for *range I*, which are not shown here. The origin of this ''anomaly'' can be traced to the fact that for the same time scale parameter τ, the exponential-window waveform (18) lasts three times longer than the rectangular-window waveform (17). Therefore, the parameter spaces *range I* and *range II* contain substantially more independent trials for the rectangular-window waveform compared to the exponential one. This results in lower false-alarm probabilities at fixed threshold for $\mathcal{F}^{\mathrm{e}}_{\mathrm{max}}$ compared to $\mathcal{F}^{\mathrm{r}}_{\mathrm{max}}$. Although there is a loss in SNR due to window-type mismatch, this effect is partly weaker than the difference in false-alarm probabilities, resulting in a partly more powerful detection statistic. To test this explanation, we have repeated the Monte-Carlo simulation with empty ranges (i.e. a fixed transient window), and with ranges where $\tau_{\mathrm{min}} > \Delta t^0$, such that all waveforms overlap in the transient start-time range, independently of window-function type. In both cases the anomaly disappears. Ranges I and II both have the feature that $\tau_{\mathrm{min}} < \Delta t^0$, such that the waveform-overlap will be substantially lower for the rectangular-window compared to the exponential-window waveforms, resulting in a large difference in independent trials in the noise case. These observations are consistent with the explanation that the anomaly is caused by the difference in independent trials. Interestingly, however, this effect is not observed for the marginalized Bayes factors, which always seem to correctly take into account the effective size of the parameter space.

C. Bayes factor self-consistency condition

One can derive an interesting self-consistency condition from the general definition (29) of the Bayes factor, which provides a useful Monte-Carlo test of our implementation: the probability of obtaining a Bayes factor $B_{\mathrm{SG}}(\boldsymbol{x}) \in [B_0, B_0 + dB]$ under any hypothesis $\mathcal{H}_i$ is given by the probability of obtaining a measurement $\boldsymbol{x}$ in the infinitesimal volume slice $\triangle\mathcal{V}_0 \equiv \{\boldsymbol{x}\colon B_{\mathrm{SG}}(\boldsymbol{x}) \in [B_0, B_0 + dB]\}$ under that hypothesis, i.e.

$$P(B_0|\mathcal{H}_i)\, dB = \int_{\Delta\mathcal{V}_0} P(\boldsymbol{x}|\mathcal{H}_i) d^n\boldsymbol{x}. \tag{68}$$

Changing local coordinates from $\boldsymbol{x}$ to $\mathbf{y} = \{B(\boldsymbol{x}), \mathbf{y}_\perp\}$, where $\mathbf{y}_\perp$ denotes $n-1$ coordinates on the constant-B_0 hypersurface $\mathbf{S_0} \equiv \{\boldsymbol{x}\colon B_{\mathrm{SG}}(\boldsymbol{x}) = B_0\}$, this can be written as

$$P(B_0|\mathcal{H}_i) = \int_{\Delta_{\mathbf{S_0}}} P(\boldsymbol{x}|\mathcal{H}_i) d\mathbf{S}, \tag{69}$$

where we defined the surface element $d\mathbf{S} \equiv J d^{n-1}\mathbf{y}_\perp$ and we assume that the Jacobian J is non-singular, i.e. $J \equiv |\partial\mathbf{y}/\partial\boldsymbol{x}| \neq 0$ everywhere inside $\Delta\mathcal{V}_0$. Using the definition (29) of the Bayes factor, we can substitute $P(\boldsymbol{x}|\mathcal{H}_{\mathrm{S}}) = B_{\mathrm{SG}}(\boldsymbol{x}) P(\boldsymbol{x}|\mathcal{H}_{\mathrm{G}})$, and obtain

$$\begin{aligned} P(B_0|\mathcal{H}_{\mathrm{S}}) &= \int_{\mathbf{S_0}} B_{\mathrm{SG}}(\boldsymbol{x}) P(\boldsymbol{x}|\mathcal{H}_{\mathrm{G}}) d\mathbf{S} \\ &= B_0 \int_{\mathbf{S_0}} P(\boldsymbol{x}|\mathcal{H}_{\mathrm{G}}) d\mathbf{S} = B_0 P(B_0|\mathcal{H}_{\mathrm{G}}), \end{aligned} \tag{70}$$

where we used the fact that $B_{\mathrm{SG}}(\boldsymbol{x})|_{\mathbf{S}_0} = B_0$ by definition of $\mathbf{S_0}$. We therefore find a general self-consistency relation for any Bayes factor, namely

$$B_{\mathrm{SG}}(\boldsymbol{x}) \equiv \frac{P(\boldsymbol{x}|\mathcal{H}_{\mathrm{S}})}{P(\boldsymbol{x}|\mathcal{H}_{\mathrm{G}})} = \frac{P(B_{\mathrm{SG}}|\mathcal{H}_{\mathrm{S}})}{P(B_{\mathrm{SG}}|\mathcal{H}_{\mathrm{G}})}. \tag{71}$$

In the framework of Monte-Carlo simulations [42], this implies that if we draw random data x according to the priors assumed in the Bayes factor, the ratio of probability-densities of obtaining $B_{\mathrm{SG}} = B_0$ in the signal- and the noise-case is identical to B_0. If the assumptions are satisfied, the signal-distribution of B_{SG} is therefore not independent of the noise-distribution, but is uniquely determined by it. If we know $P(B_{\mathrm{SG}}|\mathcal{H}_{\mathrm{G}})$, then we also know the signal distribution, and vice versa.

The resulting self-consistency relation for the odds ratio (28) is

$$O_{\mathrm{SG}} = \frac{P(O_{\mathrm{SG}}|\mathcal{H}_{\mathrm{S}})}{P(O_{\mathrm{SG}}|\mathcal{H}_{\mathrm{G}})} \frac{P(\mathcal{H}_{\mathrm{S}}|\mathcal{I})}{P(\mathcal{H}_{\mathrm{G}}|\mathcal{I})}, \tag{72}$$

where the prior odds ratio determines the probability of drawing a sample $\boldsymbol{x}$ from the signal- or noise-population, respectively. Therefore the odds ratio predicts the ratio of event densities at any value O_{SG}, rather than the ratio of normalized probability densities.

Note that in order for Eq. (71) to hold for the transient-CW Bayes factor $B_{\mathcal{F}}$ defined in Eq. (55), one must not draw signal amplitude parameters $\mathcal{A}^\mu$ according to the physical priors (47), but according to the (unphysical) $\mathcal{F}$-statistic priors (54) that went into the construction of $B_{\mathcal{F}}$. The self-consistency relation (71) can equivalently be expressed as

$$\log B_{\mathcal{F}} = \log P(\log B_{\mathcal{F}}|\mathcal{H}_{\mathrm{S}}) - \log P(\log B_{\mathcal{F}}|\mathcal{H}_{\mathrm{G}}), \tag{73}$$

which is more directly suitable for numerically testing this relation in a Monte-Carlo simulation. Defining the shortcut $p_i \equiv P(\log B_{\mathcal{F}}|\mathcal{H}_i)$, this can also be written as

$$\log\frac{p_S}{p_G} = \log B_{\mathcal{F}}. \tag{74}$$

We have performed a Monte-Carlo simulation generating values of $B_{\mathcal{F}}$ in the noise- and signal-cases, with amplitude parameters $\mathcal{A}^\mu$ drawn according to the (unphysical) $\mathcal{F}$-statistic priors (54). Figure 5 shows the resulting distributions of $p_S$and p_G, and the plots of $\log(p_S/p_G)$ versus $\log B_{\mathcal{F}}$, which should fall on a straight line of unit slope according to the self-consistency relation (74). These results show that the self-consistency relation is increasingly well satisfied with increasing prior cutoff $\hat{\rho}_{\max}$, in particular we find good agreement for cutoff values above $\hat{\rho}_{\max} \gtrsim 10$, as illustrated in Fig. 5 . This can be understood as follows: for smaller values of $\hat{\rho}_{\max}$, the noise population $P(B_{\mathcal{F}}|\mathcal{H}_G)$ is biased towards larger values of $B_{\mathcal{F}}$, because the approximation in Eq. (49) is increasingly violated. In the noise-only case, the likelihood ratio $\mathcal{L}$ will peak somewhere around $\mathcal{A}^\mu = 0$ and fall off according to a Gaussian (31) with characteristic width of order $\sigma(\hat{\rho}) \sim \mathcal{O}(1)$, modulo geometric factors or order unity. Therefore the value of $\mathcal{L}$ will not be negligible at the cutoff boundary $\hat{\rho}_{\max} \sim \mathcal{O}(1)$, and the extension to infinity will overestimate the integral. Therefore $\hat{\rho}_{\max} \gg 1$ is necessary for the integral to be well approximated by Eq. (49).

VI. PARAMETER ESTIMATION

Parameter estimation simply consists of computing the posterior probability for the signal parameters θ, given the observed data $\boldsymbol{x}$, namely

$$P(\theta|\boldsymbol{x}, \mathcal{H}_S) = cP(\boldsymbol{x}|\mathcal{H}_S, \theta)P(\theta|\mathcal{H}_S), \tag{75}$$

where $c = 1/P(\boldsymbol{x}|\mathcal{H}_S)$ is a normalization constant independent of θ. Often one is not interested in a simultaneous estimate of the full set of parameters θ, but only in a subset $\theta_1 \subset \theta$ without regard for the "nuisance" parameters θ_2, where $\theta = \{\theta_1, \theta_2\}$. From the general expression $P(\theta_1|\ldots) = \int P(\theta_1, \theta_2|\ldots)d\theta_2$ and Bayes' theorem (75), we obtain the marginalized posterior

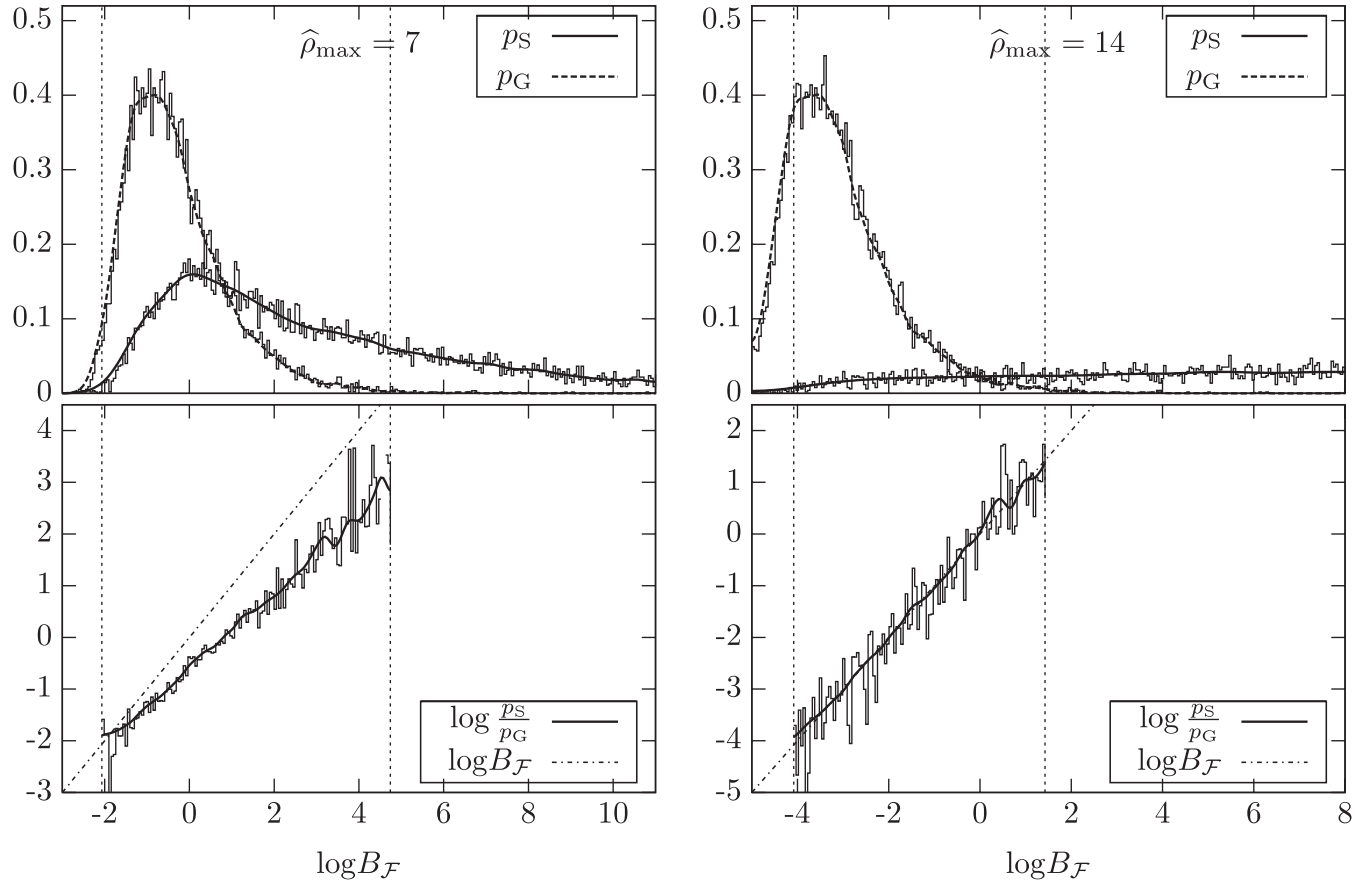

FIG. 5. Results of Monte-Carlo simulation of the Bayes-factor consistency relation (74), using $N = 10^4$ draws, transient *range II*, and amplitude priors (54) with cutoff $\hat{\rho}_{\max} = 7$ (*left panel*), and $\hat{\rho}_{\max} = 14$ (*right panel*), respectively. The upper plots show the distributions of $\log B_{\mathcal{F}}$ in the noise case (dashed), and in the signal case (solid), and the lower plots show $\log(p_S/p_G)$, which should coincide with $\log B_{\mathcal{F}}$ (dot-dashed line) according to (74). We show the binned histogram values (thin lines) as well as kernel-smoothed fits (thick lines), and we have restricted the consistency test to a region of good overlap (indicated by vertical dashed lines) between both distributions in order to avoid numerical problems.

$$P(\theta_1|\boldsymbol{x}, \mathcal{H}_S) \propto \int P(\boldsymbol{x}|\mathcal{H}_S, \theta_1, \theta_2) P(\theta_1, \theta_2|\mathcal{H}_S) d\theta_2. \tag{76}$$

For our present model the likelihood function (26) can be expressed as $P(\boldsymbol{x}|\theta) = \kappa e^{-(\boldsymbol{x}|\boldsymbol{x})/2} \mathcal{L}(\boldsymbol{x}; \theta)$ in terms of the likelihood ratio $\mathcal{L}$ of Eq. (31). Using $\mathcal{F}$-statistic priors (54), we can perform the $\mathcal{A}^\mu$-integration explicitly and obtain

$$P(\mathcal{T}, \lambda|\boldsymbol{x}, \mathcal{H}_S) \propto e^{\mathcal{F}(\boldsymbol{x};\lambda,\mathcal{T})} P(\mathcal{T}, \lambda|\mathcal{H}_S), \tag{77}$$

which is a useful starting point for further marginalization. If we consider a targeted search in Doppler parameter, i.e. $P(\lambda|\mathcal{H}_S) = \delta(\lambda - \lambda_s)$, with an assumed window-function type ϖ, we can write the posterior probability for $\{t^0, \tau\}$ as

$$P(t^0, \tau|\boldsymbol{x}, \mathcal{H}_S, \lambda, \varpi) \propto e^{\mathcal{F}(\boldsymbol{x};\lambda,\mathcal{T})} P(t^0, \tau|\mathcal{H}_S), \tag{78}$$

and the respective marginal posteriors on the transient parameters are simply

$$\begin{aligned} P(t^0|\boldsymbol{x}, \mathcal{H}_S, \lambda, \varpi) &\propto \int e^{\mathcal{F}(\boldsymbol{x};\lambda,\mathcal{T})} d\tau, \\ P(\tau|\boldsymbol{x}, \mathcal{H}_S, \lambda, \varpi) &\propto \int e^{\mathcal{F}(\boldsymbol{x};\lambda,\mathcal{T})} dt^0, \end{aligned} \tag{79}$$

where we assumed uniform priors (46) for $\{t^0, \tau\}$. The generalization to marginalization over the window type ϖ is straightforward, and yields a weighted sum of these posteriors with relative prior probabilities of the different window types, e.g.

$$P(t^0|\boldsymbol{x}, \mathcal{H}_S, \lambda) \propto \sum_{\varpi} P(t^0|\boldsymbol{x}, \mathcal{H}_S, \lambda, \varpi) P(\varpi|\mathcal{H}_S). \tag{80}$$

Similarly, parameter-estimation on the window-type ϖ itself can be expressed as

$$\begin{aligned} P(\varpi|\boldsymbol{x}, \mathcal{H}_S, \lambda) &\propto P(\varpi|\mathcal{H}_S) \int e^{\mathcal{F}(\boldsymbol{x};\lambda,\mathcal{T})} dt^0 d\tau \\ &\propto P(\varpi|\mathcal{H}_S) B_{\mathcal{F}}(\boldsymbol{x}; \lambda, \varpi), \end{aligned} \tag{81}$$

so the window-type specific Bayes factor (55) is proportional to the relative likelihood of different window-types ϖ.

In the frequentist framework one often uses maximum-likelihood estimators for parameter estimation, i.e. $\{t^0_{\mathrm{ML}}, \tau_{\mathrm{ML}}\}$ such that

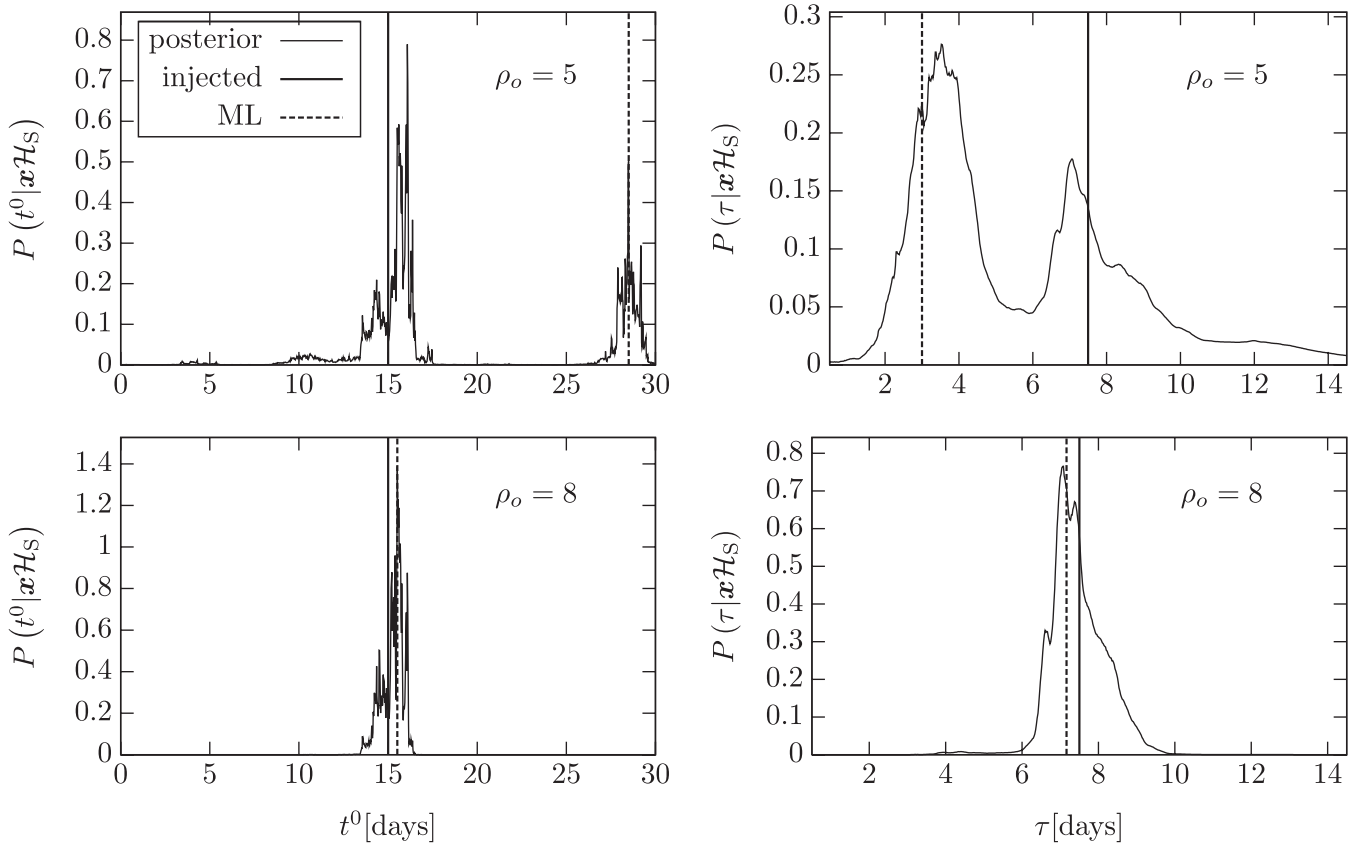

FIG. 6. Example posteriors (79) on t^0 (left column) and τ (right column), for an injected rectangular transient-CW signal within *range III*, with randomly drawn amplitude parameters at fixed optimal SNR $\rho_o = 5$ (upper row) and $\rho_o = 8$ (lower row). The solid vertical line indicates the injected parameter value, and the dashed vertical line ("ML") indicates the maximum-likelihood estimate.

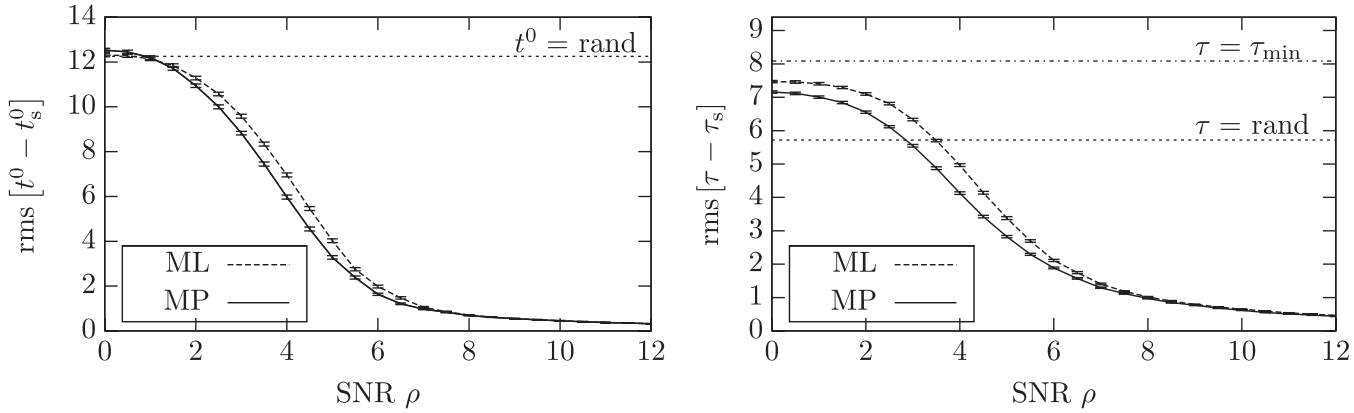

FIG. 7. Parameter-estimation accuracy: rms-errors on t^0 (*left panel*) and τ (*right panel*) as a function of SNR for maximum-likelihood (ML) and maximum-posterior (MP) estimates. Each value of SNR corresponds to $N = 10^4$ random parameter draws, and the error bars represent 1σ jackknife error estimates. The dotted horizontal line ("rand") corresponds to a uniform random guess (85), and the dot-dashed horizontal line corresponds to a maximally biased $\tau = \tau_{\min}$ guess (86).

$$\mathcal{F}(\boldsymbol{x}; t^0_{\mathrm{ML}}, \tau_{\mathrm{ML}}) = \max_{\{t^0,\tau\}} \mathcal{F}(\boldsymbol{x}; t^0, \tau), \tag{82}$$

for fixed Doppler-point λ and window-type $\boldsymbol{\varpi}$.

The following Monte-Carlo studies use rectangular transient windows within a wider range than (67), namely

$$\text{range III: } \tau \in [0.5, 14.5] \text{ days}, \qquad t^0 \in [0, 30] \text{ days}. \tag{83}$$

Using the transient-parameter *range III*, Fig. 6 shows one example of parameter posteriors (79) and maximum-likelihood (ML) estimators (82) on t^0 and τ for injected signals with SNR $\rho_o = 5$ and $\rho_o = 8$, respectively. We see in Fig. 6 that the time scale of variations in $P(t^0|\boldsymbol{x})$ is shorter than in $P(\tau|\boldsymbol{x})$. This is due to the combined effect of the wider plotted range in t^0 and the twice smaller characteristic correlation time scale in t^0, as derived in Eq. (63).

In order to study the quality of transient parameter estimation as a function of SNR, we performed Monte-Carlo simulations comparing maximum-posterior estimators (MP), defined as

$$\begin{aligned} P(t^0_{\mathrm{MP}}|\boldsymbol{x}, \mathcal{H}_{\mathrm{S}}) &= \max_{t^0} P(t^0|\boldsymbol{x}, \mathcal{H}_{\mathrm{S}}), \\ P(\tau_{\mathrm{MP}}|\boldsymbol{x}, \mathcal{H}_{\mathrm{S}}) &= \max_{\tau} P(\tau|\boldsymbol{x}, \mathcal{H}_{\mathrm{S}}), \end{aligned} \tag{84}$$

and maximum-likelihood estimators (ML) of Eq. (82), using uniform priors on $\{t^0, \tau\}$ within *range III*, and physical priors (47) on amplitude parameters. For different fixed values of SNR we perform $N = 10^4$ simulated parameter estimates, and compute the rms errors from $t^0 - t^0_s$ and $\tau - \tau_s$, for ML- and MP- estimators, respectively. We also compare the results to the error of a pure random guess within the range $[t_1, t_1 + \Delta t]$, where $\Delta t \equiv t_2 - t_1$. For uniform priors this is

$$(\mathrm{rms}[t - t_s])^2 = \frac{1}{\Delta t^2} \int_{t_1}^{t_2} dt_s \int_{t_1}^{t_2} dt (t - t_s)^2 = \frac{\Delta t^2}{6}. \tag{85}$$

For the transient *range III* this yields random-guess errors of $\mathrm{rms}[t^0 - t^0_s] \approx 12.25$ days, and $\mathrm{rms}[\tau - \tau_s] \approx 5.72$ days, which are shown in Fig. 7. On the other hand, for a maximally biased "guess" of $\tau = \tau_{\min}$, one finds

$$\mathrm{rms}[\tau_{\min} - \tau_s] = \Delta\tau/\sqrt{3} \approx 8.08 \text{ days}. \tag{86}$$

The results of the parameter-estimation Monte-Carlo simulation are shown in Fig. 7. For low SNRs of $\rho \lesssim 7$, we see that the MP estimators perform better than the ML estimators, while for higher SNR the estimation quality of both estimators converges. We note that for $\rho \to 0$ the parameter estimation on t^0 converges to a random guess as expected, but in the case of τ_{ML} we notice a substantial deviation, and to a lesser degree, also for τ_{MP}. These estimates fall closer to a maximally biased $\tau = \tau_{\min}$ guess, which indicates an increasing bias in the τ estimators for low SNR, strongly favoring values close to $\tau_{\min}$. This surprising effect will be studied in some more detail in the following section.

Estimation bias on time scale τ in pure noise

Figure 8 shows normalized histograms of the parameter estimates on τ in pure Gaussian noise, i.e. for $\rho = 0$. These results confirm the estimation bias towards $\tau_{\min}$ previously seen in Fig. 7. We have been able to trace this effect to a surprising fundamental feature of the Gaussian random walk underlying matched filtering. If we discretize the integration time as $T_j = j\Delta T$ in steps of ΔT, then linearity of the scalar product (32) implies

$$x'_\mu(T_j) = x'_\mu(T_{j-1}) + \Delta x'_{\mu,j}, \tag{87}$$

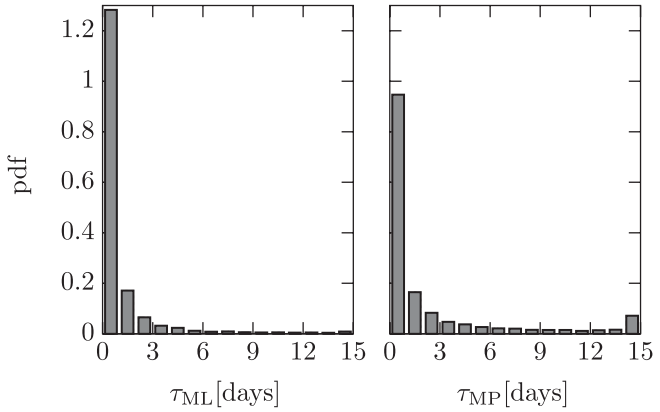

FIG. 8. Parameter-estimation bias in τ_{ML} (*left panel*) and τ_{MP} (*right panel*) in pure noise data (SNR = 0). The plots show normalized histograms of $N = 10^4$ parameter estimates of τ in Gaussian random noise data.

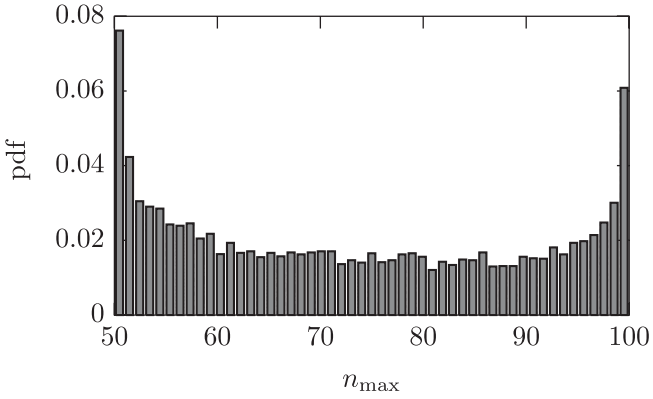

FIG. 9. Distribution of random-walk index n_{max} at which the maximum of s_n^2 occurs ($N = 10^4$ trials).

where $\Delta x'_{\mu,j}$ is independent of $x'_\mu(T_{j-1})$ and follows a Gaussian distribution with zero mean (in the noise case $\mathcal{H}_{\mathrm{G}}$). We see from this expression that the amplitudes x'_μ can be interpreted as Gaussian random walks over finite steps ΔT in the integration time. The four Gaussian random walks $\{x'_\mu\}$ are combined in the quadratic form (36) to yield $\mathcal{F}(T_j)$, where they are normalized by $\mathcal{M}^{\mu\nu}$ such that $\mathcal{F}$ follows a χ^2-distribution with 4 degrees of freedom *for any* T_j, independently of the random-walk step j.

We can therefore consider a simpler toy-model, namely, a one-dimensional normalized random walk, defined as

$$s_n = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \Delta x_j, \tag{88}$$

where the Δx_j are n independent Gaussian random variables with zero mean and unit variance, i.e. $\Delta x_j \sim$ Gauss(0, 1). The random walk (88) is normalized in such a way that it follows *exactly* the same distribution at every step n, i.e. $s_n \sim$ Gauss(0, 1). We denote as $s^2_{n_{\mathrm{max}}}$ the maximum of s_n^2 over a walk-step window $[n_1, n_2]$, i.e.

$$s^2_{n_{\mathrm{max}}} \equiv \max_{n\in[n_1,n_2]} s_n^2, \tag{89}$$

and plot the distribution of $n_{\mathrm{max}} \in [n_1, n_2]$ in repeated trials of such normalized random walks. For example, if we consider $n_1 = 50$ and $n_2 = 100$, and $N = 10^4$ trials, we obtain the distribution of maxima shown in Fig. 9, which illustrates a qualitative bias towards n_{min}, and to a lesser extent n_{max}, similar to what was seen in Fig. 8 for the physical parameter estimation of τ. This seems to be a manifestation of a well-known property of random walks, namely, Lévy's arcsine law, which is discussed, for example, in Sec. 4.2 of [43].

VII. CONCLUSIONS

We have discussed the case for a transient-CW search, from the point of view of its astrophysical motivation, and in order to bridge the gap between short burstlike signals and traditional infinite-duration CW signals.

We have introduced a simple transient signal model based on the classical CW signal, modulated by a window function of finite support. The corresponding Bayes factor has been derived and implemented, and we have performed Monte-Carlo simulations to compare its efficiency to the orthodox maximum-likelihood detection method. These results show that the Bayes factor is both more sensitive and more robust than a maximum-likelihood statistic, and it yields better parameter estimates, at similar computing cost.

The Monte-Carlo studies presented in this work have been limited to pure Gaussian noise, and it would be important to test this method in practice on actual detector data. In particular one needs to address the question of how to assign meaningful false-alarm probabilities to detection candidates in real detector data (see also [5] for an example of these difficulties).

The search-method discussed here is currently restricted to nonrepeating transient CWs, and more work would be required to generalize this approach to allow for repeating transient-CW signals.

In addition to the fully coherent search method, we have derived the necessary formalism for a semicoherent transient search, which could be used to perform an all-sky, all-frequency wide parameter-space transient search, for example, running on Einstein@Home. More work is required to fully develop and implement this approach.

ACKNOWLEDGMENTS

We thank Badri Krishnan, Curt Cutler, Christian Röver, and Vladimir Dergachev for interesting discussions. We further thank Holger Pletsch, Ben Owen, Nils Andersson,

Ian Jones, Trevor Sidery and Eric Thrane for careful reading of the manuscript and useful comments.

APPENDIX A: TRANSIENT SEARCH IMPLEMENTATION

Our numerical implementation of the Bayes factor $B_{\mathcal{F}}(\boldsymbol{x};\lambda)$ of Eq. (55) consists of two steps:

(1) calculate a discretized $\mathcal{F}$-statistic map $\mathcal{F}(\boldsymbol{x};t^0,\tau)$ over the search ranges in t^0 and τ,
(2) compute $B_{\mathcal{F}}(\boldsymbol{x})$ by discretizing the marginalization integral in Eq. (55) as a sum.

The following two sections provide more details about these two steps, respectively.

1. Atoms-based $\mathcal{F}$-statistic computation

We discretize the two-dimensional $\mathcal{F}$-statistic map over the search ranges $t^0 \in [t^0_{\min}, t^0_{\min} + \Delta t^0]$ and $\tau \in [\tau_{\min}, \tau_{\min} + \Delta\tau]$ in steps dt^0 and $d\tau$, respectively. Namely, we compute the $\mathcal{F}$-matrix

$$\mathcal{F}_{mn} \equiv \mathcal{F}(\boldsymbol{x};t^0_m,\tau_n), \tag{A1}$$

over the $N_{t^0} \times N_\tau$ rectangular grid

$$t^0_m = t^0_{\min} + m\,dt^0, \qquad \tau_n = \tau_{\min} + n\,d\tau, \tag{A2}$$

where $N_{t^0} = \Delta t^0/dt^0$ and $N_\tau = \Delta\tau/d\tau$.

The default for these step-sizes used here are $dt^0 = d\tau = 1800$ s. If the shortest signals considered are $\tau_{\min} \sim 0.5$ days long, then according to Eq. (62) the worst-case mismatch is $m \lesssim d\tau/\tau_{\min} \approx 0.04$, i.e. a 4% loss of squared SNR.

In the current implementation of the transient-CW search we use the underlying discretization of the $\mathcal{F}$-statistic computation in COMPUTEFSTATISTIC_V2 (CFSV2), as described in more detail in [20]. Namely the $\mathcal{F}$-statistic (36) is computed from $\mathcal{M}'_{\mu\nu}$ and x'_μ of Eqs. (32) and (33), which are approximated as sums

$$x_\mu \approx \sum_{i=1}^{N_{\mathrm{SFT}}} g_i x_{\mu,i}, \qquad \mathcal{M}_{\mu\nu} \approx \sum_{i=1}^{N_{\mathrm{SFT}}} g_i^2 \mathcal{M}_{\mu\nu,i}, \tag{A3}$$

in terms of the $\mathcal{F}$-statistic "atoms"

$$\begin{aligned} x_{\mu,i} &\equiv 2\sum_X S_X^{-1} \int_{t_i}^{t_i+T_{\mathrm{SFT}}} x^X(t) h^X_\mu(t)\,dt, \\ \mathcal{M}_{\mu\nu,i} &\equiv 2\sum_X S_X^{-1} \int_{t_i}^{t_i+T_{\mathrm{SFT}}} h^X_\mu(t) h^X_\nu(t)\,dt, \end{aligned} \tag{A4}$$

where T_{SFT} is the length of the short Fourier transforms (SFTs) that are used as input data, typically $T_{\mathrm{SFT}} = 1800$ s. In the above expressions we implicitly assumed that the transient-window function $g(t)$ varies slowly and can be approximated as constant over the time scale T_{SFT}. For any chosen Doppler position λ, the code first computes the N_{SFT} atoms $\{x_{\mu,i}, \mathcal{M}_{\mu\nu,i}\}_{i=1}^{N_{\mathrm{SFT}}}$ over the whole observation time of interest, which are also the primary input to this implementation of the standard CW $\mathcal{F}$-statistic. The $\mathcal{F}$-statistic value $\mathcal{F}_{mn}$ for any particular transient parameters $\{t^0_m, \tau_n\}$ is then computed from the corresponding partial sums in Eq. (A3). This approach allows for an efficient computation of the $\mathcal{F}_{mn}$ map in the case of a rectangular-window function g_{r} of Eq. (17): going from τ_n to τ_{n+1} can be achieved by a single extra addition, namely

$$x_\mu(t^0_m,\tau_n) = x_\mu(t^0_m,\tau_{n-1}) + x_{\mu,i_1}, \tag{A5}$$

(and similarly for $\mathcal{M}_{\mu\nu}$), where i_1 is the atom-index corresponding to the time-step $t^0_m + \tau_n$ (assuming for simplicity that $d\tau = T_{\mathrm{SFT}}$).

In the case of the exponential transient window g_{e} of Eq. (18), the whole sum in Eq. (A3) needs to be recomputed for every matrix-element m, n, as the window-function provides different weights at every point.

2. Transient marginalization integrals

Given the transient matrix $\mathcal{F}_{mn}$, we can now turn the integrals in Eqs. (55) and (79) into simple sums. A minor subtlety arises because of the potential numerical problems of expressions like $e^{\mathcal{F}}$, which overflow in double precision for values $\mathcal{F} > 709$. Such values are easily possible for noisy non-Gaussian data with line-artifacts or for strong injected signals. It is therefore numerically safer to rewrite these sums using the discretized maximum-likelihood statistic value (53), namely

$$\mathcal{F}_{\max} \approx \max_{mn} \mathcal{F}_{mn}, \tag{A6}$$

and write $B_{\mathcal{F}}$ of Eq. (55) as a discretized sum in the form

$$\log B_{\mathcal{F}}(\boldsymbol{x};\lambda) \approx \mathcal{F}_{\max} + c_0 + \log \sum_{m=1}^{N_{t^0}} \sum_{n=1}^{N_\tau} e^{\Delta\mathcal{F}_{mn}}, \tag{A7}$$

where we defined $c_0 \equiv \log(70/(\hat{\rho}^4_{\max} N_{t^0} N_\tau))$, and

$$\Delta\mathcal{F}_{mn} \equiv \mathcal{F}_{mn} - \mathcal{F}_{\max} \le 0. \tag{A8}$$

For large values of $\mathcal{F}_{\max}$, some terms $e^{\Delta\mathcal{F}_{mn}}$ can now *underflow* to zero, but this poses no numerical problems because these contributions were negligible anyway. Similarly, we can compute the parameter posteriors (79) as

$$P(t^0_m|\boldsymbol{x},\mathcal{H}_{\mathrm{S}}) \propto \sum_{n=1}^{N_\tau} e^{\Delta\mathcal{F}_{mn}}, \tag{A9}$$

$$P(\tau_n|\boldsymbol{x},\mathcal{H}_{\mathrm{S}}) \propto \sum_{m=1}^{N_{t^0}} e^{\Delta\mathcal{F}_{mn}}, \tag{A10}$$

where $\mathcal{F}_{\max}$ only affects the normalization and has been dropped.

3. Computing cost

In order to be able to plan which types of searches can be performed with a reasonable investment of computing cost, it is useful to have a rough computing-cost model that allows one to predict the expected run time of any search.

Let us start with the underlying $\mathcal{F}$-statistic implementation `CFSv2` which is used in our current coherent transient-search implementation described in Sec. A 1. Note that more efficient $\mathcal{F}$-statistic algorithms do exist, based on resampling and FFT techniques [44,45]. However, the "Williams-Schutz" method [46] currently implemented in `CFSv2` is well suited to our purpose, as it is already based on computing the atoms $\{x_{\mu,i}, \mathcal{M}_{\mu\nu,i}\}$ over SFTs. Nevertheless it will be interesting to study possibly more efficient transient-CW implementations based on resampling and FFT techniques.

We can break up the total computing time $\mathfrak{c}_B^{(1)}$ for $B_{\mathcal{F}}(\boldsymbol{x};\lambda)$ at one Doppler point λ as follows:

(1) the time $\mathfrak{c}_{\text{atoms}}$ to compute the atoms (A4),
(2) the time $\mathfrak{c}^{\varpi}_{\mathcal{F}\text{map}}$ to compute the $\mathcal{F}$-map (A1) over $\{t^0, \tau\}$ for given window-type ϖ, and
(3) the time $\mathfrak{c}_{\text{marg}}$ to marginalize over the $\mathcal{F}$-map to obtain the Bayes factor $B_{\mathcal{F}}(\boldsymbol{x};\lambda)$, using Eq. (A7)

Atoms cost: the time $\mathfrak{c}_{\text{atoms}}$ to compute for one Doppler position λ the $N_{\text{SFT}} = T_{\text{data}}/T_{\text{SFT}}$ atoms using `CFSV2` is simply

$$\mathfrak{c}_{\text{atoms}} = \mathfrak{c}_0 \frac{T_{\text{data}}}{T_{\text{SFT}}}, \tag{A11}$$

where $\mathfrak{c}_0$ is a machine-dependent timing constant.

$\mathcal{F}$-map cost: the time $\mathfrak{c}^{\varpi}_{\mathcal{F}\text{map}}$ to compute the $N_{t^0} \times N_\tau$ matrix of values $\{\mathcal{F}_{mn}\}$, where $N_{t^0} = \Delta t^0/dt^0$, and $N_\tau = \Delta\tau/d\tau$. This time can be expressed as

$$\mathfrak{c}^{\varpi}_{\mathcal{F}\text{map}} = \sum_{m=1}^{N_{t^0}} \sum_{n=1}^{N_\tau} \mathfrak{c}^{\varpi}_{mn}, \tag{A12}$$

where $\mathfrak{c}^{\varpi}_{mn}$ is the time required to compute $\mathcal{F}_{mn}$ for particular values m, n, which depends crucially on the type ϖ of window-function.

Let us first consider the exponential transient window (18), which corresponds to the generic case where no special optimizations can be used. Note that at the smallest time scale index, $n = 1$, we already had to sum all atoms corresponding to a time scale τ_{min}. However, for every matrix element we need to recompute this sum, as the window-weights will be different every time. This introduces a constant computing-time offset $\mathfrak{c}_{n=0} = \mathfrak{c}_{\text{e}}\tau_{\text{min}}$, where $\mathfrak{c}_{\text{e}}$ is the machine-dependent time to compute all atom-additions and window-weights within one unit-time. From this we can express the cost for computing $\mathcal{F}_{mn}$ as $\mathfrak{c}^{\text{e}}_{mn} = \mathfrak{c}_{\text{e}}(\tau_{\text{min}} + nd\tau)$, which does not depend on the start-time index m. Therefore we obtain the total $\mathcal{F}$-map cost as

$$\mathfrak{c}^{\text{e}}_{\mathcal{F}\text{map}} \approx \mathfrak{c}_{\text{e}} \frac{\Delta t^0}{dt^0} \frac{\Delta\tau}{d\tau} \frac{(\tau_{\text{min}} + \Delta\tau/2)}{T_{\text{SFT}}}, \tag{A13}$$

which is quadratic in $\Delta\tau$ (and where we assumed $N_\tau \gg 1$).

In the case of a rectangular transient-window, various sums are closely related and we can use the optimization of Eq. (A5) to reduce the computing cost. Namely, every step $n \to n + 1$ in the timescale adds just the cost of one extra time-step $d\tau$. This cost is $\mathfrak{c}_{\text{r}}d\tau$, where $\mathfrak{c}_{\text{r}}$is the machine-dependent time to do all sums for a unit timescale in the case of a rectangular window. For every start-time index m, we need to compute the sums up to τ_{min} first, costing $\mathfrak{c}_{\text{r}}\tau_{\text{min}}$. Summarizing, we can express the accumulated computing cost $\mathfrak{c}^{\text{r}}_{\text{m}} = \sum_n \mathfrak{c}^{\text{r}}_m$ *per line* of the $\mathcal{F}_{mn}$ matrix as $\mathfrak{c}^{\text{r}}_{\text{m}} = \mathfrak{c}_{\text{r}}\tau_{\text{min}} + \mathfrak{c}_{\text{r}}N_\tau d\tau = \mathfrak{c}_{\text{r}}\tau_{\text{max}}$. The total $\mathcal{F}$-map cost is therefore

$$\mathfrak{c}^{\text{r}}_{\mathcal{F}\text{map}} = \mathfrak{c}_{\text{r}} \frac{\Delta t^0}{dt^0} \frac{(\tau_{\text{min}} + \Delta\tau)}{T_{\text{SFT}}}. \tag{A14}$$

Bayes-factor marginalization: the marginalization (A7) is a simple sum over the exponentiated $\mathcal{F}$-map matrix, and so we can directly write the marginalization cost as

$$\mathfrak{c}_{\text{marg}} = \mathfrak{c}_1 \frac{\Delta t^0}{dt^0} \frac{\Delta\tau}{d\tau}, \tag{A15}$$

where $\mathfrak{c}_1$ is the machine-dependent cost of exponentiation and summation of real numbers. Note that exponentiation is a very costly operation, and so we use a lookup-table approximation to reduce this cost.

The total computing cost $\mathfrak{c}_B^{(1)}$ for the transient Bayes factor at one Doppler point λ is now expressible as

$$\mathfrak{c}_B^{(1)} = \mathfrak{c}_{\text{atoms}} + \mathfrak{c}^{\varpi}_{\mathcal{F}\text{map}} + \mathfrak{c}_{\text{marg}}, \tag{A16}$$

and for a search over N_λ Doppler points, this would simply extend as

$$\mathfrak{c}_B^{(N_\lambda)} = N_\lambda \mathfrak{c}_B^{(1)}, \tag{A17}$$

where the extra cost of summing these Bayes factors will be negligible.

We have verified that these timing models are a good description of the actual performance of the code [47] by varying the parameters N_{SFT}, Δt^0 and $\Delta\tau$, and by fitting the measured times to the model. This yields the following timing constants on an Intel Core2 Duo CPU with 2.60 GHz (Lenovo T61p):

$$\begin{aligned} \mathfrak{c}_0 &= 1.4 \times 10^{-6}\ \text{sec}, & \mathfrak{c}_1 &= 2.8 \times 10^{-8}\ \text{sec}. \\ \mathfrak{c}_{\text{r}} &= 4.2 \times 10^{-8}\ \text{sec}, & \mathfrak{c}_{\text{e}} &= 1.3 \times 10^{-7}\ \text{sec} \end{aligned} \tag{A18}$$

If we target a known pulsar with a transient-CW search using data from 2 detectors spanning one year ($N_{\text{SFT}} = 35,000$), with a time scale range of $\tau \in [0.5, 14.5]$ days, using step-sizes $dt^0 = d\tau = T_{\text{SFT}} = 1800$ sec , we obtain an estimated computing costs of

$$\begin{aligned} \mathfrak{c}_{\text{atoms}} &\approx 0.05 \text{ sec}, & \mathfrak{c}_{\text{marg}} &\approx 0.3 \text{ sec} \\ \mathfrak{c}^{\text{r}}_{\mathcal{F}\text{map}} &\approx 0.5 \text{ sec}, & \mathfrak{c}^{\text{e}}_{\mathcal{F}\text{map}} &\approx 540 \text{ sec}. \end{aligned} \tag{A19}$$

We see that such a targeted search would be easy to perform for all interesting pulsars, even using the much slower transient exponential window. Comparing this transient-CW search to a coherent CW search over one year, we notice that the cost per template λ is about 18 times higher than the CW search for a rectangular window, and about 10^4 times higher for the exponential window. Wide parameter-space transient-CW searches will therefore be severely limited by computing resources, and a semicoherent transient-search method as discussed in Sec. IV G will be required.

4. Synthesizing Monte-Carlo draws

Following the method used in [34], we have implemented an efficient Monte-Carlo simulation method, by avoiding the generation of the primary data-input of the search code (time-series or short Fourier transforms), and instead synthesizing higher-level secondary data-input to the transient-search functions directly. In our case, we synthesize the N_{SFT} *atoms* $\{x_{\mu,i}, \mathcal{M}_{\mu\nu,i}\}$, which are the intermediate input-data to the transient-search function. This approach is very economical in computing resources and allows us to generate large numbers of Monte-Carlo draws in a very short time on a single machine. We draw signal parameters $\{\mathcal{A}_s, \lambda_s, \mathcal{T}_s\}$ according to the priors, and from this we can compute the deterministic signal-atoms $s_{\mu,i}$ of Eq. (38) and the antenna-pattern atoms $\mathcal{M}_{\mu\nu,i}$ of Eq. (A4). The noise-atoms $n_{\mu,i}$ are Gaussian random variables with zero mean and covariance matrix $\mathcal{M}_{\mu\nu,i}$. These can be generated from uncorrelated Gaussian variates, e.g. by using a Cholesky decomposition on $\mathcal{M}_{\mu\nu,i}$. The data-atoms are then simply $x_{\mu,i} = n_{\mu,i} + s_{\mu,i}$, according to Eq. (37). Note that we have assumed the transient-window function to be constant on the atoms' time scale T_{SFT}, and therefore we can synthesize standard non-transient CW atoms. The transient-window function is applied when computing the $\mathcal{F}$-statistic map, cf. Sec. A 1. All Monte-Carlo simulations used the "Mersenne Twister" random-number generator GSL RNG MT19937 from GSL [48].

[1] N. Andersson and K. D. Kokkotas, Mon. Not. R. Astron. Soc. **299**, 1059 (1998).
[2] K. D. Kokkotas, T. A. Apostolatos, and N. Andersson, Mon. Not. R. Astron. Soc. **320**, 307 (2001).
[3] J. Clark, I. S. Heng, M. Pitkin, and G. Woan, Phys. Rev. D **76**, 043003 (2007).
[4] K. Hayama, S. Desai, S. D. Mohanty, M. Rakhmanov, T. Summerscales, and S. Yoshida, Classical Quantum Gravity **25**, 184016 (2008).
[5] J. Abadie *et al.* (LIGO Scientific Collaboration), Phys. Rev. D **83**, 042001 (2011).
[6] R. Prix (for the LIGO Scientific Collaboration), in *Neutron Stars and Pulsars*, edited by W. Becker, *Astrophysics and Space Science Library* Vol. 357 (Springer, Berlin Heidelberg, 2009), p. 651, (LIGO-P060039-v3), http://dcc.ligo.org/cgi-bin/DocDB/ShowDocument?docid=635.
[7] E. Thrane, S. Kandhasamy, C. D. Ott, W. G. Anderson, N. L. Christensen, M. W. Coughlin, S. Dorsher, S. Giampanis, V. Mandic, A. Mytidis *et al.*, Phys. Rev. D **83**, 083004 (2011).
[8] C. A. Van Eysden and A. Melatos, Classical Quantum Gravity **25**, 225020 (2008).
[9] M. F. Bennett, C. A. Van Eysden, and A. Melatos, Mon. Not. R. Astron. Soc. **409**, 1705 (2010).
[10] A. G. Lyne, S. L. Shemar, and F. G. Smith, Mon. Not. R. Astron. Soc. **315**, 534 (2000).
[11] K. Kashiyama and K. Ioka, Phys. Rev. D **83**, 081302 (2011).
[12] Y. Levin and M. Van Hoven, arXiv:1103.0880.
[13] P. Jaranowski, A. Królak, and B. F. Schutz, Phys. Rev. D **58**, 063001 (1998).
[14] D. I. Jones and N. Andersson, Mon. Not. R. Astron. Soc. **331**, 203 (2002).
[15] L. Gualtieri, R. Ciolfi, and V. Ferrari, arXiv:1011.2778.
[16] S. Dall'Osso, S. N. Shore, and L. Stella, Mon. Not. R. Astron. Soc. **398**, 1869 (2009).
[17] N. Andersson and K. D. Kokkotas, Int. J. Mod. Phys. D **10**, 381 (2001).
[18] R. Bondarescu, S. A. Teukolsky, and I. Wasserman, Phys. Rev. D **79**, 104003 (2009).
[19] B. J. Owen, Phys. Rev. D **82**, 104002 (2010).
[20] R. Prix, Tech. Rep. No. LIGO-T0900149-v2, 2010, https://dcc.ligo.org/cgi-bin/DocDB/ShowDocument?docid=1665 .
[21] C. Cutler and B. F. Schutz, Phys. Rev. D **72**, 063006 (2005).
[22] C. Cutler and K. S. Thorne, in *General Relativity and Gravitation*, edited by N. T. Bishop and S. D. Maharaj (World Scientific, Singapore, 2002), pp. 72–111.
[23] B. J. Owen and L. Lindblom, Classical Quantum Gravity **19**, 1247 (2002).
[24] T. Sidery, A. Passamonti, and N. Andersson, Mon. Not. R. Astron. Soc. **405**, 1061 (2010).
[25] M. Ruderman, Astrophys. J. **382**, 587 (1991).
[26] B. K. Link and R. I. Epstein, Astrophys. J. **373**, 592 (1991).
[27] N. Andersson, G. L. Comer, and R. Prix, Phys. Rev. Lett. **90**, 091101 (2003).
[28] N. Andersson, G. L. Comer, and R. Prix, Mon. Not. R. Astron. Soc. **354**, 101 (2004).

[29] B. Carter, D. Langlois, and D. M. Sedrakian, Astron. Astrophys. **361**, 795 (2000) http://aa.springer.de/bibs/0361002/2300795/small.htm.
[30] K. A. Van Riper, R. I. Epstein, and G. S. Miller, Astrophys. J. **381**, L47 (1991).
[31] P. R. Brady, T. Creighton, C. Cutler, and B. F. Schutz, Phys. Rev. D **57**, 2101 (1998).
[32] R. Prix and J. T. Whelan, Classical Quantum Gravity **24**, S565 (2007).
[33] R. Prix, Phys. Rev. D **75**, 023004 (2007).
[34] R. Prix and B. Krishnan, Classical Quantum Gravity **26**, 204013 (2009).
[35] LIGO Scientific Collaboration, LALSuite: FreeSoftware (GPL) Tools for Data-Analysis https://www.lsc-group.phys.uwm.edu/daswg/projects/lalsuite.html.
[36] D. I. Jones, B. J. Owen, and D. Whitbeck, Tech. Rep. No. LIGO-T0900500-v1, 2005 https://dcc.ligo.org/cgi-bin/DocDB/ShowDocument?docid=6339.
[37] All numerical results refer to a detector the Hanford site (H1), sky-location of the Crab pulsar, RA= 05h34m31.973s, DEC=22:00:52.06, and start-time $t^0 = 0$ with respect to a reference time of $t^0_{\mathrm{ref}} = 814838413$ sec in GPS seconds (Nov 1, 2005).
[38] P. R. Brady and T. Creighton, Phys. Rev. D **61**, 082001 (2000).
[39] H. J. Pletsch, Phys. Rev. D **82**, 042002 (2010).
[40] http://www.einsteinathome.org.
[41] B. Abbott *et al.* (LIGO Scientific Collaboration), Phys. Rev. D **80**, 042003 (2009).
[42] A. C. Searle, arXiv:0804.1161.
[43] S. N. Majumdar, Physica A (Amsterdam) **389**, 4299 (2010).
[44] P. Astone, K. M. Borkowski, P. Jaranowski, and A. Królak, Phys. Rev. D **65**, 042003 (2002).
[45] P. Patel, X. Siemens, R. Dupuis, and J. Betzwieser, Phys. Rev. D **81**, 084032 (2010).
[46] P. R. Williams and B. F. Schutz, arXiv:gr-qc/9912029.
[47] Note that the atoms-cost c_0 per SFT reported here refers to a standard (-O2) build, while highly optimized versions of this code running on Einstein@Home can achieve more than 10x faster performance.
[48] M. Galassi *et al.*, *GNU Scientific Library Reference Manual* (Network Theory Ltd., Godalming, UK, 2009), 3rd ed.) ISBN 0-9546120-7-8. http://www.gnu.org/software/gsl/