Search for Continuous Gravitational Waves: Optimal StackSlide method at fixed computing cost

R. Prix, M. Shaltev

Albert-Einstein-Institut, Callinstr. 38, 30167 Hannover, Germany

(Dated: Fri Jan 20 15:28:42 2012 +0100)

(LIGO-P1100156-v3)

Coherent wide parameter-space searches for continuous gravitational waves are typically limited in sensitivity by their prohibitive computing cost. Therefore semi-coherent methods (such as Stack-Slide) can often achieve a better sensitivity. We develop an analytical method for finding *optimal* StackSlide parameters at fixed computing cost under ideal conditions of gapless data with Gaussian stationary noise. This solution separates two regimes: an *unbounded* regime, where it is always optimal to use all the data, and a *bounded* regime with a finite optimal observation time. Our analysis of the sensitivity scaling reveals that both the fine- and coarse-grid mismatches contribute equally to the average StackSlide mismatch, an effect that had been overlooked in previous studies. We discuss various practical examples for the application of this optimization framework, illustrating the potential gains in sensitivity compared to previous searches.

I. INTRODUCTION

Motivation. The detection of continuous gravitational waves (CWs) from spinning neutron stars (NSs) in our galaxy remains an elusive goal, despite the global network of detectors LIGO [1], Virgo [2] and GEO 600 [3] having completed their initial and enhanced science runs (e.g. see [4–7]). The search for CWs will likely remain a difficult challenge with uncertain prospects even in the era of Advanced detectors [8–10] and third-generation detectors such as ET [11]. Two main reasons for this are (i) astrophysical priors on CWs and (ii) the large parameter space of signal parameters to explore (cf. [12] for a review and further references).

(i) Current astrophysical priors contain large uncertainties on the expected strength of CW emissions from spinning NSs, with a strong bias towards extremely weak signals, informed by the population of known pulsars as well as by a statistical analysis of a putative galactic "gravitar" population [13]. (ii) The required number of templates for a coherent matched-filter search over a range of unknown signal parameters typically grows very rapidly with increasing duration of data analyzed. Therefore only a fraction of the available data can be analyzed coherently (e.g. see [14–16]).

It was realized early on [17] that in situations where the total computing cost of the search is constrained, a *semi-coherent* approach could typically achieve better sensitivity than coherent matched filtering: shorter segments of data are analyzed coherently, then the statistics from these segments are combined incoherently. One method of incoherent combination simply consists of summing the statistics from the different segments, which is typically referred to as the "StackSlide" method in the CW context (also known as the Radon transform). The template bank used for the semi-coherent combination of coherent statistics is referred to as the *fine grid*, as it typically requires a higher resolution than the template banks of the per-segment coherent searches (referred to as *coarse grids*). Details of the respective template banks

will be discussed in Sec. III D.

There are a number of different semi-coherent methods: for example, recent work [18] has shown that Stack-Slide sensitivity can be improved by a sliding coherence-window approach. A closely related variant to StackSlide is the *Hough transform* [19], which counts the number of segments in which the statistic crosses a given threshold, instead of summing the statistics. This is generally less sensitive, but is designed to be more robust in the presence of strong non-stationarities. A somewhat different semi-coherent approach are cross-correlation methods, described in more detail in [20].

Related to the semi-coherent methods are the so-called *hierarchical* schemes, which consist of following up "promising" candidates from a (coherent or semi-coherent) search by subsequent, more sensitive searches, referred to as "stages". This procedure is iterated until the parameter space of surviving candidates is sufficiently narrowed down for a fully coherent follow-up using all the data. Work on implementing such schemes in practice is still ongoing.

Optimization problem. In this paper we focus on the standard single-stage StackSlide method, which was also used in previous optimization studies [17, 21], and is relatively straightforward to describe analytically.

Any search method contains a number of tuneable parameters, such as the template-bank mismatch, the data selection procedure, and the number and length of segments to analyze. Hierarchical schemes would further require specification of the number of stages and the respective distributions of computing costs and candidate thresholds. The sensitivity of a search generally depends on all these choices, and we therefore need to study how to maximize sensitivity as a function of these parameters.

This optimization problem has been studied previously by Brady and Creighton [17] (henceforth 'BC') and subsequently by Cutler, Gholami and Krishnan [21] (in the following 'CGK'). Both studies have focused on the wider problem of optimizing a multi-stage hierarchical scheme of StackSlide stages, and have directly resorted to fully

numerical exploration of the optimization problem. Here instead we focus on the simpler single-stage search, which allows us to fully analytically analyze the problem. In the next step this can be used as a building block to attack the optimization of hierarchical schemes.

Note that for a network of detectors with different noise-floors, the choice of detectors to use at fixed computing cost is be part of the optimization problem, but under the present assumption of "ideal data" the answer can be obtained independently [22]. More work is required to develop a practical algorithm to compute the optimal search parameters for given data from a network of detectors, including gaps, non-stationarities and various detector artifacts.

Summary of main results. Careful analysis of the sensitivity scaling shows that the average StackSlide mismatch is given by the *sum* of the average mismatches from the coarse- and fine-grid template banks, an effect that had previously been overlooked. Note that we allow for independent coarse- and fine-grid mismatches, while BC and CGK forced them to be equal as an *ad-hoc* constraint.

The analytic optimization is achieved by using local power-law approximations to the computing-cost and sensitivity functions. The results provide analytic self-consistency conditions for the optimal solution: if the initial power-law coefficients agree with those found at the analytic solution, then the solution is self-consistent and (locally) optimal. If this is not the case, one can iterate over successive solutions or scan a range of StackSlide parameters, in order to "bootstrap" into a self-consistent optimal solution.

We find that the analytic solution for StackSlide searches separates two different regimes depending on the power-law coefficients: a *bounded* regime in which there is a finite optimal observation time, and an *unbounded* regime in which the optimal solution always consists of using all of the available data, irrespective of the available computing-cost.

Plan of the paper. In Sec. II we discuss the general CW optimization problem, which includes the single-stage StackSlide search as the lowest-level building block. In Sec. III we derive the sensitivity estimate and computing-cost functions for StackSlide searches, and motivate their approximation as local power-laws. After deriving in Sec. IV the general analytical solution and discussing a few special cases, we provide examples for the practical application of this framework in Sec. V: directed searches, all-sky searches, and searches for CWs from NSs in binary systems.

II. MAXIMIZING PROBABILITY OF A CW DETECTION

The goal for wide parameter-space CW searches for unknown signals should be to maximize the probability of detection, given current astrophysical priors, detector data, and finite computing resources. Conceptually one can think of this problem as a hierarchy of two questions:

(i) What parameter-space $\mathbb{P} \subseteq \mathbb{P}^{(0)}$ to search? More generally: how to distribute the total available computing power C_0 over the space $\mathbb{P}^{(0)}$ of possible CW signals, given astrophysical priors, detector data and an (optimal) search method?

(ii) What is the optimal search method? Namely, which method yields the highest detection probability p_{det} on a parameter-space cell $\Delta \mathbb{P}$ if we spend computing-cost ΔC on it?

The answer to the first question relies on the second, but we can analyze the lower-level second question independently of the first. There has been surprisingly little work on this problem so far. The first question has hardly been addressed at all, except for recent work by Knispel [23]. The second question has been studied previously by BC [17] and CGK [21], assuming a specific type of hierarchical scheme, which we refer to as the *classical hierarchical scheme* (CHS).

In the CHS one performs a hierarchy of semi-coherent searches (called *stages*), starting with a relatively low-sensitivity search over the whole initial parameter space $\mathbb{P}^{(1)}$. Promising candidates crossing the first-stage threshold are selected and constitute the search subspace $\mathbb{P}^{(2)} \subset \mathbb{P}^{(1)}$ for the second, higher-sensitivity stage. This is iterated until eventually after m such stages a fully-coherent search over all the data can be performed on the surviving candidates. At this point one has reached the maximal possible sensitivity for a small portion $\mathbb{P}^{(m)} \subset \mathbb{P}^{(1)}$ of the initial parameter space.

Each stage (i) is characterized by its input parameter-space $\mathbb{P}^{(i)}$, a computing-cost constraint $C_0^{(i)}$ and a false-alarm probability $p_{\mathrm{fA}}^{(i)}$. Each stage selects a candidate subspace $\mathbb{P}^{(i+1)} \subset \mathbb{P}^{(i)}$ to follow up in the next stage. An optimal per-stage search would result in the highest detection probability $p_{\mathrm{det}}^{(i)}$ for given signal strength h_{rms} and constraints $\{p_{\mathrm{fA}}^{(i)}, \mathbb{P}^{(i)}, C_0^{(i)}\}$. The tuneable CHS parameters are therefore the number m of stages and the per-stage constraints $\{p_{\mathrm{fA}}^{(i)}, C_0^{(i)}\}$. These can be varied in order to maximize the overall detection probability $p_{\mathrm{det}}(h_{\mathrm{rms}})$ for the given total signal parameter-space $\mathbb{P}^{(0)}$, computing cost $C_0 = \sum_{i=1}^{m} C_0^{(i)}$ and false-alarm probability $p_{\mathrm{fA}} = \prod_{i=1}^{m} p_{\mathrm{fA}}^{(i)}$.

This formulation of the CHS suggests that each stage (i) could be considered an independent optimization problem for given external constraints $\{p_{\mathrm{fA}}^{(i)}, \mathbb{P}^{(i)}, C_0^{(i)}\}$, if none of its internal parameters interfere with the overall hierarchical scheme. One might contend that the parameter-space resolution of the search violates this clean factorization: the follow-up space $\mathbb{P}^{(i+1)}$ from stage (i) depends on its parameter-space resolution, which might impact the required computing-cost of the next stage. However, it is easy to see that (to first order) such a coupling is not expected. The number of candidates $\mathcal{N}_{\mathrm{cand}}$ returned from any stage (except for the last one)

will be dominated by the number $\mathcal{N}_{\mathrm{fA}}$ of false alarms. Therefore $\mathcal{N}_{\mathrm{cand}} \approx \mathcal{N}_{\mathrm{fA}} \approx p_{\mathrm{fA}} \mathcal{N}$, where $\mathcal{N}$ is the number of (approximately) independent templates searched in this stage. This can be estimated as $\mathcal{N} \approx V_{\mathbb{P}}/v_0$, in terms of the (metric) volume $V_{\mathbb{P}}$ of the parameter space $\mathbb{P}$, and the volume v_0 covered by one template. Therefore the *number* $\mathcal{N}_{\mathrm{cand}}$ of follow-up candidates from any stage does indeed depend on its parameter-space resolution, which depends on the internal stage parameters. However, the computing cost of the next stage depends primarily on the *volume* of the follow-up parameter-space, which is $V_{\mathrm{fA}} \approx \mathcal{N}_{\mathrm{fA}} v_0 \approx p_{\mathrm{fA}} V_{\mathbb{P}}$, and is therefore *independent* of internal stage parameters. It is interesting to note that each stage (i) in this scheme achieves a reduction of the input parameter-space volume by roughly a factor of the false-alarm probability $p_{\mathrm{fA}}^{(i)}$, irrespective of the internal details of that search.

The optimal per-stage search method is essentially unknown, but following BC and CGK we focus on a known good strategy, namely the StackSlide method. While different semi-coherent methods differ in the details and their exact sensitivity, they share the main characteristics of coherently searching N shorter segments of length ΔT, and combining them incoherently in some way. We roughly expect the *sensitivity per cost* of different methods to behave qualitatively similarly to the StackSlide method, but more work would be required to study this in detail.

III. PROPERTIES OF A SINGLE-STAGE STACKSLIDE SEARCH

The general StackSlide scheme consists of dividing the data (of total duration T) into N segments of duration $\Delta T = T/N$, then performing a coherent matched-filter search on each segment and combining these statistics *incoherently* to a new statistic Σ by summing them across segments. The coherent matched-filter statistic used is the $\mathcal{F}$-statistic, which was first derived in [24] and extended to multiple detectors in [25]. Using the same amount of data as a fully coherent search, the resulting semi-coherent statistic is less sensitive, but substantially cheaper to compute over a wide parameter space. At *fixed computing cost* a semi-coherent search is therefore generally more sensitive than a fully coherent $\mathcal{F}$-statistic search.

Notation: we distinguish quantities Q that can refer to either the coherent or the incoherent step in the following way: we use a tilde, i.e. $\widetilde{Q}$ when referring to the coherent step, and a hat, i.e. $\widehat{Q}$ when referring to the incoherent step. For the following derivations we restrict ourselves to a single-detector formalism for simplicity, but we state the (trivial) generalization to $N_{\mathrm{det}} \geq 1$ detectors of relevant results.

A. The StackSlide search method

Let $k = 1 \dots N$ be the index over segments, and $\lambda \in \mathbb{P}$ a point in the search space $\mathbb{P}$ of signal parameters. The "ideal" StackSlide statistic $\Sigma_0(\lambda)$ is defined as

$$\Sigma_0(\lambda) \equiv \sum_{k=1}^{N} 2\mathcal{F}_k(\lambda) \,, \tag{1}$$

i.e. a simple sum of $\mathcal{F}$-statistic values $\{2\mathcal{F}_k(\lambda)\}_{k=1}^{N}$ computed at *the same* template point λ across all N segments.

This would require computing the $\mathcal{F}$-statistic over the same template bank as Σ_0 in every segment. However, the metric resolution of Σ_0 is generally finer than that of the single-segment $\mathcal{F}$-statistics [26], and therefore this approach would spend unnecessary computing cost on the coherent $\mathcal{F}$-statistic. In practice $\mathcal{F}$ is therefore computed over a *coarse grid* of $\widetilde{\mathcal{N}}$ templates in each segment k, and is interpolated in order to sum $\mathcal{F}$ on the *fine grid* of $\widehat{\mathcal{N}} \geq \widetilde{\mathcal{N}}$ templates (e.g. see [27]).

Typically the interpolation consists of picking the *closest* (in the metric sense) coarse-grid point $\widetilde{\lambda}_k(\widehat{\lambda})$ to the fine-grid point $\widehat{\lambda}$ from every segment k, i.e. we approximate Eq. (1) by

$$\Sigma(\widehat{\lambda}) \equiv \sum_{k=1}^{N} 2\mathcal{F}_k\left(\widetilde{\lambda}_k(\widehat{\lambda})\right) \approx \Sigma_0(\widehat{\lambda}) \,, \tag{2}$$

which we refer to as the "interpolating" StackSlide statistic Σ. The following sensitivity optimization focuses exclusively on this interpolating StackSlide method, which is the most relevant approach for current practical applications. The subtle difference between *interpolating StackSlide* Σ and *ideal StackSlide* Σ_0 with respect to its sensitivity and mismatches has been overlooked in previous studies, and will be important for the optimization problem.

B. Mismatch and metric

1. $\mathcal{F}$-statistic mismatch

In the presence of a signal timeseries $s(t, \lambda_{\mathrm{s}})$ with phase parameters λ_{s}, the statistic $2\mathcal{F}_k(\widetilde{\lambda})$ in a point $\widetilde{\lambda}$ follows a non-central χ^2-distribution with four degrees of freedom and non-centrality parameter $\rho_k^2(\lambda_{\mathrm{s}}, \widetilde{\lambda})$. We denote this probability distribution as

$$P\left(2\mathcal{F}_k | \rho_k^2\right) = \chi_4^2(2\mathcal{F}_k; \rho_k^2) \,, \tag{3}$$

which has the expectation value

$$E\left[2\mathcal{F}_k(\lambda_{\mathrm{s}}, \widetilde{\lambda})\right] = 4 + \rho_k^2(\lambda_{\mathrm{s}}, \widetilde{\lambda}) \,. \tag{4}$$

The quantity ρ_k is often referred to as the coherent signal-to-noise ratio (SNR). In the case of a perfectly-matched

template $\widetilde{\lambda} = \lambda_{\mathrm{s}}$, the resulting "optimal" SNR [24] in segment k can be expressed as

$$\rho_k^2(\lambda_{\mathrm{s}}, \lambda_{\mathrm{s}}) = \frac{2}{S_{\mathrm{n}}} \int_{t_k}^{t_k + \Delta T} s^2(t, \lambda_{\mathrm{s}}) \, dt$$
$$\equiv \frac{2}{S_{\mathrm{n}}} h_{\mathrm{rms}, k}^2 \, \Delta T , \qquad (5)$$

where t_k is the start-time of the kth segment, S_{n} is the (single-sided) noise power spectral density at the signal frequency f_{s}. In the second equality we defined the rms signal strength $h_{\mathrm{rms}, k}$ [21] in segment k, which is a useful measure of the *intrinsic* signal strength in the detectors, independently of the quality and the amount of data used.

The signal strength $h_{\mathrm{rms}, k}$ depends on the intrinsic signal amplitude h_0, the sky-position, polarization angles, and detector orientation during segment k. One can show [24, 28] that averaging $h_{\mathrm{rms}, k}^2$ isotropically over sky-positions and polarization angles yields the relation $\langle h_{\mathrm{rms}, k}^2 \rangle_{\mathrm{sky, pol}} = (2/25) \, h_0^2$. Furthermore, for segment lengths of order $\Delta T \gtrsim \mathcal{O}(1 \, \mathrm{days})$, the averaging in Eq. (5) results in $h_{\mathrm{rms}, k}$ tending towards a constant over all segments. Therefore it will be convenient to approximate $h_{\mathrm{rms}, k} \approx h_{\mathrm{rms}}$, and so we can write

$$\rho_k^2(\lambda_{\mathrm{s}}, \lambda_{\mathrm{s}}) \approx \frac{2}{S_{\mathrm{n}}} h_{\mathrm{rms}}^2 \, \Delta T \equiv \rho_{\mathrm{opt}}^2(\Delta T) , \qquad (6)$$

which defines the average optimal SNR ρ_{opt} for given segment length ΔT.

Note that this approximation only applies to the perfectly-matched SNR $\rho_k(\lambda_{\mathrm{s}}, \lambda_{\mathrm{s}})$. The "mismatched" SNR $\rho_k(\lambda_{\mathrm{s}}, \widetilde{\lambda})$ in an offset template $\widetilde{\lambda} = \lambda_{\mathrm{s}} + \Delta\lambda$ is reduced with respect to the optimal SNR $\rho_{\mathrm{opt}}(\Delta T)$. The corresponding relative loss defines the (segment-specific) *mismatch* function $\widetilde{\mu}_k(\lambda_{\mathrm{s}}, \widetilde{\lambda})$, namely

$$\widetilde{\mu}_k(\lambda_{\mathrm{s}}, \widetilde{\lambda}) \equiv 1 - \frac{\rho_k^2(\lambda_{\mathrm{s}}, \widetilde{\lambda})}{\rho_{\mathrm{opt}}^2(\Delta T)}$$
$$= \widetilde{g}_{ij,k}(\lambda_{\mathrm{s}}) \, \Delta\lambda^i \, \Delta\lambda^j + \mathcal{O}(\Delta\lambda^3) , \qquad (7)$$

where Taylor-expansion for small offsets $\Delta\lambda$ defines the (coherent) metric tensor $\widetilde{g}_{ij,k}(\lambda)$ for segment k. The concept of the parameter-space metric was first introduced in [29, 30], and analyzed in the context of a simplified CW statistic [14] and the $\mathcal{F}$-statistic [31].

The per-segment *coarse-grid* template bank is constructed under the constraint that no signal point $\lambda_{\mathrm{s}} \in \mathbb{P}$ should exceed a given *maximal mismatch* $\widetilde{m}$ to its closest (i.e. with the smallest mismatch) coarse-grid template $\widetilde{\lambda}_k(\lambda_{\mathrm{s}})$, namely

$$\widetilde{\mu}_k\left(\lambda_{\mathrm{s}}, \widetilde{\lambda}_k(\lambda_{\mathrm{s}})\right) \leq \widetilde{m} \quad \text{for all} \quad \lambda_{\mathrm{s}} \in \mathbb{P} . \qquad (8)$$

2. Mismatch of "ideal" StackSlide

The "ideal" StackSlide statistic Σ_0 defined in Eq. (1) is the basis for the definition of the semi-coherent metric

[17, 26, 32]). The statistic Σ_0 follows a non-central χ^2-distribution with $4N$ degrees of freedom, denoted as

$$P\left(\Sigma_0 | \rho_{\Sigma_0}^2\right) = \chi_{4N}^2(\Sigma_0; \rho_{\Sigma_0}^2) , \qquad (9)$$

with non-centrality parameter

$$\rho_{\Sigma_0}^2(\lambda_{\mathrm{s}}, \widehat{\lambda}) \equiv \sum_{k=1}^{N} \rho_k^2(\lambda_{\mathrm{s}}, \widehat{\lambda}) , \qquad (10)$$

where λ_{s} are the signal parameters and $\widehat{\lambda}$ is the location of a fine-grid template. The corresponding expectation value is

$$E\left[\Sigma_0(\lambda_{\mathrm{s}}, \widehat{\lambda})\right] = 4N + \rho_{\Sigma_0}^2 . \qquad (11)$$

The perfectly-matched non-centrality parameter $\rho_{\Sigma_0}^2(\lambda_{\mathrm{s}}, \lambda_{\mathrm{s}})$ can be expressed as

$$\rho_{\Sigma_0}^2(\lambda_{\mathrm{s}}, \lambda_{\mathrm{s}}) = \sum_{k=1}^{N} \rho_k^2(\lambda_{\mathrm{s}}, \lambda_{\mathrm{s}})$$
$$= N \rho_{\mathrm{opt}}^2(\Delta T)$$
$$= \rho_{\mathrm{opt}}^2(T) , \qquad (12)$$

which is identical to that of a perfectly-matched $\mathcal{F}$-statistic over the same total duration T, as seen from Eq. (6). The reason why the StackSlide statistic Σ_0 is less sensitive than the $\mathcal{F}$-statistic for the same amount of data T stems from the different degrees of freedom of the respective distributions, namely $\chi_4^2(\rho^2)$ for the $\mathcal{F}$-statistic as opposed to $\chi_{4N}^2(\rho^2)$ for StackSlide Σ_0.

The mismatch function $\widehat{\mu}_0(\lambda_{\mathrm{s}}, \widehat{\lambda})$ of ideal StackSlide is defined in analogy to Eq. (7) as

$$\widehat{\mu}_0(\lambda_{\mathrm{s}}, \widehat{\lambda}) \equiv 1 - \frac{\rho_{\Sigma_0}^2(\lambda_{\mathrm{s}}, \widehat{\lambda})}{\rho_{\mathrm{opt}}^2(T)}$$
$$= \widehat{g}_{ij}(\lambda_{\mathrm{s}}) \, \Delta\lambda^i \, \Delta\lambda^j + \mathcal{O}(\Delta\lambda^3) , \qquad (13)$$

where $\Delta\lambda \equiv \widehat{\lambda} - \lambda_{\mathrm{s}}$ is the offset between the fine-grid template $\widehat{\lambda}$ and the signal location λ_{s}, and Taylor-expansion in small $\Delta\lambda$ defines the (semi-coherent) metric tensor $\widehat{g}_{ij}$. Using Eqs. (12) and (7), we can rearrange the expression for the mismatch as

$$\widehat{\mu}_0(\lambda_{\mathrm{s}}, \widehat{\lambda}) = \frac{1}{N} \sum_{k=1}^{N} \widetilde{\mu}_k(\lambda_{\mathrm{s}}, \widehat{\lambda})$$
$$\approx \left(\frac{1}{N} \sum_{k=1}^{N} \widetilde{g}_{ij,k}(\lambda_{\mathrm{s}})\right) \Delta\lambda^i \Delta\lambda^j , \qquad (14)$$

which shows that the ideal StackSlide mismatch $\widehat{\mu}_0$ and metric $\widehat{g}_{ij}$ are segment-averages of the coherent mismatches and metrics, respectively.

The *fine-grid* template bank of a StackSlide search is constructed under the constraint that no signal point $\lambda_{\mathrm{s}} \in \mathbb{P}$ should exceed a given *maximal mismatch* $\widehat{m}$ to its

5

closest (i.e. with the smallest mismatch) fine-grid template $\widehat{\lambda}(\lambda_{\rm s})$, namely

$$\widehat{\mu}_0\left(\lambda_{\rm s}, \widehat{\lambda}(\lambda_{\rm s})\right) \leq \widehat{m} \quad \text{for all} \quad \lambda_{\rm s} \in \mathbb{P}. \qquad (15)$$

3. Mismatch of "interpolating" StackSlide

We can now combine the above results to derive the mismatch of the interpolating StackSlide statistic Σ defined in Eq. (2). This statistic follows a non-central χ^2_{4N} distribution, namely

$$P\left(\Sigma|\rho_\Sigma^2\right) = \chi^2_{4N}(\Sigma; \rho_\Sigma^2), \qquad (16)$$

with non-centrality parameter

$$\rho_\Sigma^2(\lambda_{\rm s}, \widehat{\lambda}) \equiv \sum_{k=1}^{N} \rho_k^2\left(\lambda_{\rm s}, \widetilde{\lambda}_k(\widehat{\lambda})\right), \qquad (17)$$

where $\lambda_{\rm s}$ are the signal phase parameters, and $\widetilde{\lambda}_k(\widehat{\lambda})$ is the closest coarse-grid template in segment k to the fine-grid point $\widehat{\lambda}$.

The mismatch function $\widehat{\mu}(\lambda_{\rm s}, \widehat{\lambda})$ of interpolating StackSlide is therefore

$$\widehat{\mu}(\lambda_{\rm s}, \widehat{\lambda}) \equiv 1 - \frac{\rho_\Sigma^2(\lambda_{\rm s}, \widehat{\lambda})}{\rho_{\rm opt}^2(T)}$$
$$= \frac{1}{N}\sum_{k=1}^{N} \widetilde{\mu}_k\left(\lambda_{\rm s}, \widetilde{\lambda}_k(\widehat{\lambda})\right), \qquad (18)$$

which allows us to express the mismatched non-centrality parameter as

$$\rho_\Sigma^2(\lambda_{\rm s}, \widehat{\lambda}) = \left(1 - \widehat{\mu}(\lambda_{\rm s}, \widehat{\lambda})\right)\rho_{\rm opt}^2(T). \qquad (19)$$

The extra offset per-segment, $\delta\lambda_k \equiv \widetilde{\lambda}_k(\widehat{\lambda}) - \widehat{\lambda}$, incurred due to using the closest coarse-grid point $\widetilde{\lambda}_k(\widehat{\lambda})$ instead of the fine-grid point $\widehat{\lambda}$ tends to increase the mismatch with respect to the ideal mismatch function $\widehat{\mu}_0$ of Eq. (14). In order to quantify this effect, we write the effective per-segment offset from a signal as $\Delta\widetilde{\lambda}_k \equiv \widetilde{\lambda}_k(\widehat{\lambda}) - \lambda_{\rm s}$, while the ideal per-segment offset would be $\Delta\widehat{\lambda} \equiv \widehat{\lambda} - \lambda_{\rm s}$. We can write $\Delta\widetilde{\lambda}_k = \Delta\widehat{\lambda} + \delta\lambda_k$, and inserting this into the coherent-metric of Eq. (7) we obtain (neglecting higher-order terms $\mathcal{O}\left(\Delta\lambda^3\right)$):

$$\widetilde{\mu}_k\left(\lambda_{\rm s}, \widetilde{\lambda}_k(\widehat{\lambda})\right) = \widetilde{g}_{ij,k}\,\Delta\widetilde{\lambda}_k^i\Delta\widetilde{\lambda}_k^j$$
$$= \widetilde{\mu}_k(\lambda_{\rm s}, \widehat{\lambda}) + \widetilde{g}_{ij,k}\,\delta\lambda_k^i\,\delta\lambda_k^j + 2\widetilde{g}_{ij,k}\,\Delta\widehat{\lambda}^i\,\delta\lambda_k^j, \quad (20)$$

where in the first term we recover the ideal per-segment mismatch function of Eq. (14), the second term represents an extra mismatch due to the offset $\delta\lambda_k$, while the last term depends on the opening angle θ_k of the
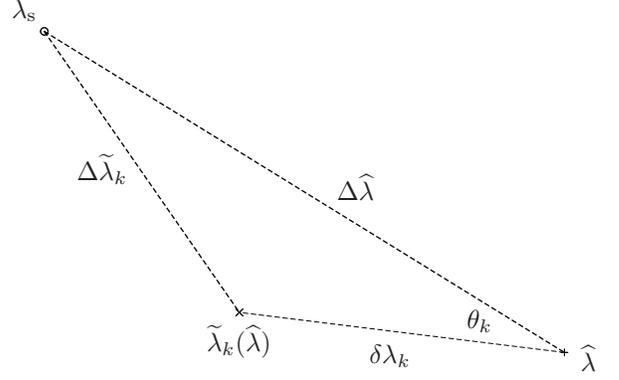

FIG. 1: Mismatch triangle formed by the signal point $\lambda_{\rm s}$, closest fine-grid template $\widehat{\lambda}$, and the coarse-grid template $\widetilde{\lambda}_k(\widehat{\lambda})$ closest to $\widehat{\lambda}$ in segment k.

mismatch triangle (see Fig. 1), namely $2|\Delta\widehat{\lambda}||\delta\lambda_k|\cos\theta_k$, with mismatch norm defined as $|x|^2 \equiv \widetilde{g}_{ij,k}\,x^i x^j$.

We assume that the fine-grid point $\widehat{\lambda}$ falls randomly into the Wigner-Seitz cell of the closest coarse-grid template $\widetilde{\lambda}_k(\widehat{\lambda})$ in segment k. Given that the coarse-grid metric $\widetilde{g}_{ij,k}$ generally varies across segments, we further assume that the offset $\delta\lambda_k$ approximates a uniform random sampling of the coarse-grid Wigner-Seitz cell. Inserting Eq. (20) into (18), we see that the average over the angle-term $|\Delta\widehat{\lambda}||\delta\lambda_k|\cos\theta_k$ will tend to zero, as any sign of $\cos\theta_k$ is equally likely, while the average norm $|\delta\lambda_k|^2$ will tend to the average mismatch $\langle\widetilde{\mu}\rangle$ of the coarse-grid template bank, and so we obtain

$$\widehat{\mu}(\lambda_{\rm s}, \widehat{\lambda}) \approx \widehat{\mu}_0(\lambda_{\rm s}, \widehat{\lambda}) + \langle\widetilde{\mu}\rangle. \qquad (21)$$

When estimating the sensitivity of the interpolating StackSlide statistic, we will further average this expression over randomly-chosen signal locations $\lambda_{\rm s}$, and therefore the above approximate averaging expressions will become exact, and we obtain

$$\langle\widehat{\mu}\rangle = \langle\widehat{\mu}_0\rangle + \langle\widetilde{\mu}\rangle, \qquad (22)$$

where averaging is performed over the coarse- and fine-grid template banks (i.e. the respective Wigner-Seitz cells).

The probability distribution of signal mismatches in a given template bank constructed with a certain maximal mismatch m depends on the structure and dimensionality of the template bank. The corresponding average mismatch can be expressed as $\langle\mu\rangle = \xi\,m$, where $\xi \in (0,1)$ is a characteristic geometric factor of the template bank. Such mismatch distributions were studied quantitatively, for example in [33]. For hyper-cubic lattices, the geometric relation is well known to be exactly $\langle\mu\rangle = m/3$, which was used in previous optimization studies [17, 21]. For the more efficient A_n^*-lattices this geometric factor is approximately $\xi \approx 0.5-0.6$ for low dimensions $n \lesssim 6$. Here we allow for general geometric factors ξ, but for simplicity

we assume it to be identical for the fine- and coarse-grid template banks, and so Eq. (22) can be written as

$$\langle \widehat{\mu} \rangle = \xi(\widehat{m} + \widetilde{m}), \quad \text{with} \quad \xi \in (0, 1), \qquad (23)$$

where $\widehat{m}$ and $\widetilde{m}$ are the maximal mismatch parameters of fine- and coarse-grid template banks, respectively.

Averaging the non-centrality parameter ρ_Σ^2 of Eq. (19) over random signal parameters λ_s at fixed signal strength h_{rms}, we can now obtain the expression

$$\langle \rho_\Sigma^2 \rangle = [1 - \xi(\widehat{m} + \widetilde{m})] \frac{2N_{\text{det}}}{S_{\text{n}}} h_{\text{rms}}^2 T, \qquad (24)$$

where we (trivially) generalized the result to the case of a network of N_{det} detectors. In this case S_{n} refers to the harmonic mean over individual-detector PSDs, and h_{rms} is a noise-weighted average over rms-amplitudes from different detectors (e.g. see [28]). The fact that *both* the coarse- and fine-grid mismatches enter this expression has been overlooked in previous studies [17, 21], where only the fine-grid mismatch $\widehat{m}$ had been included[1].

C. Sensitivity estimate

The false-alarm and false-dismissal probabilities for a given threshold Σ_{th} of the StackSlide statistic Σ of Eq. (2) are

$$p_{\text{fA}}(\Sigma_{\text{th}}) = \int_{\Sigma_{\text{th}}}^{\infty} \chi_{4N}^2(\Sigma; 0) \, d\Sigma, \qquad (25)$$

$$p_{\text{fD}}(\Sigma_{\text{th}}, \rho_\Sigma^2) = \int_{-\infty}^{\Sigma_{\text{th}}} \chi_{4N}^2(\Sigma; \rho_\Sigma^2) \, d\Sigma, \qquad (26)$$

where the special case of a coherent $\mathcal{F}$-statistic search corresponds to $N = 1$.

Sensitivity is often quantified in terms of the weakest (rms-) signal strength h_{th} required to obtain a given detection probability $p_{\text{det}}^* = 1 - p_{\text{fD}}^*$ at a given false-alarm probability p_{fA}^*. This requires inverting Eq. (25) to obtain the critical threshold $\Sigma_{\text{th}}^* = \Sigma_{\text{th}}(p_{\text{fA}}^*)$, then substituting this into Eq. (26) and inverting to find the critical non-centrality parameter

$$\rho_\Sigma^{*2} = \rho_\Sigma^2(p_{\text{fA}}^*, p_{\text{fD}}^*, N). \qquad (27)$$

The signal location λ_s is generally unknown, therefore the mismatch $\widehat{\mu}(\lambda_s, \widehat{\lambda})$ of the closest template $\widehat{\lambda}$ and the corresponding mismatched non-centrality parameter $\rho_\Sigma^2(\lambda_s, \widehat{\lambda})$ of Eq. (19) follow a random distribution. In order to estimate the threshold rms signal strength h_{th}, one would have to compute $p_{\text{fD}}(p_{\text{fA}}^*, h_{\text{th}})$ by *averaging* the right-hand side of Eq. (26) over the (known) mismatch distribution of $\widehat{\mu}$. Furthermore, for statements

about physical upper limits and sensitivity of a given search pipeline, it is often required to quantify the sensitivity in terms of the *intrinsic* GW amplitude h_0, instead of the rms detector strain h_{rms}, which would require further averaging of Eq. (26) over the (potentially) unknown sky-position and polarization parameters. This problem has recently been studied in detail in [34].

For our present purpose it will be sufficient to obtain the correct *scaling* of sensitivity with StackSlide parameters $\{N, T, \widetilde{m}, \widehat{m}\}$, while the absolute sensitivity level is less important. We will therefore employ the usual simplification of this problem, which consists in averaging ρ_Σ^2 instead of $p_{\text{fD}}(\rho_\Sigma^2)$ over the mismatch distribution of $\widehat{\mu}$, so we approximate

$$p_{\text{fD}}(p_{\text{fA}}^*, h_{\text{th}}) = \left\langle p_{\text{fD}}(p_{\text{fA}}^*, \rho_\Sigma^2)\big|_{h_{\text{th}}} \right\rangle_{\lambda_s}$$
$$\approx p_{\text{fD}}\left(p_{\text{fA}}^*, \langle \rho_\Sigma^2 \rangle_{\lambda_s}\right). \qquad (28)$$

The results of [34] indicate that this indeed approximately preserves the *scaling* of sensitivity as a function of StackSlide parameters.

We can now use Eq. (24) to translate the critical non-centrality parameter ρ_Σ^{*2} of Eq. (27) into a threshold rms signal-strength h_{th}, namely

$$h_{\text{th}}^{-2} = \frac{2N_{\text{det}}}{\rho_\Sigma^{*2}} [1 - \xi(\widehat{m} + \widetilde{m})] \frac{T}{S_{\text{n}}}. \qquad (29)$$

Following the Neyman-Pearson criterion we want to maximize detection probability $p_{\text{det}}^* = 1 - p_{\text{fD}}^*$ at fixed false-alarm probability p_{fA}^* and at fixed signal strength h_{rms}. Equivalently[2] we can fix the false-alarm and false-dismissal probabilities and *minimize* the required threshold rms signal strength h_{th}, which is the traditional optimization approach used in previous studies [17, 21].

1. Gauss approximation for large N

One approach (used in [19, 21]) to make further analytical progress consists in assuming a large number of segments, i.e. $N \gg 1$, and invoke the central limit theorem to approximate χ_{4N}^2 by a Gaussian distribution

$$P\left(\Sigma | \rho_\Sigma^2\right) \overset{N \gg 1}{\approx} \left(2\pi\sigma_\Sigma^2\right)^{-1/2} \exp\left[-\frac{(\Sigma - \bar{\Sigma})^2}{2\sigma_\Sigma^2}\right], \qquad (30)$$

with mean and variance of $\chi_{4N}^2(\rho_\Sigma^2)$ given by

$$\bar{\Sigma} = 4N + \rho_\Sigma^2,$$
$$\sigma_\Sigma^2 = 2(4N + 2\rho_\Sigma^2). \qquad (31)$$

---

[1] These studies additionally imposed the *ad-hoc* constraint of $\widetilde{m} = \widehat{m}$ in the computing-cost expressions

[2] Due the monotonicity of p_{fD} as a function of h_{rms}.

This allows us to analytically integrate Eqs. (25), (26), which yields

$$p_{\mathrm{fA}}(\Sigma_{\mathrm{th}}) = \frac{1}{2}\,\mathrm{erfc}\left(\frac{\Sigma_{\mathrm{th}} - 4N}{2\sqrt{4N}}\right)\,, \qquad (32)$$

$$p_{\mathrm{fD}}(\Sigma_{\mathrm{th}}, \rho_\Sigma^2) = \frac{1}{2}\,\mathrm{erfc}\left(\frac{\rho_\Sigma^2 - (\Sigma_{\mathrm{th}} - 4N)}{2\sqrt{4N + 2\rho_\Sigma^2}}\right)\,, \qquad (33)$$

where $\mathrm{erfc}(x) \equiv 1 - \mathrm{erf}(x)$ is the complementary error-function. Substituting Eq. (32) into Eq. (33), we obtain

$$\beta \equiv \frac{\rho_\Sigma^2 - 2\alpha\sqrt{4N}}{2\sqrt{4N + 2\,\rho_\Sigma^2}}\,, \qquad (34)$$

where we defined

$$\begin{aligned}
\beta &\equiv \mathrm{erfc}^{-1}(2\,p_{\mathrm{fD}}^*) = -\mathrm{erfc}^{-1}(2\,p_{\mathrm{det}}^*)\,, \\
\alpha &\equiv \mathrm{erfc}^{-1}(2\,p_{\mathrm{fA}}^*)\,.
\end{aligned} \qquad (35)$$

Solving Eq. (34) for the critical non-centrality parameter ρ_Σ^{*2}, we obtain[3]

$$\begin{aligned}
\rho_\Sigma^{*2}(\alpha, \beta, N) = {} & 2\alpha\sqrt{4N} + 4\beta^2 \\
& + 2\beta\sqrt{4N + 4\alpha\sqrt{4N} + 4\beta^2}\,, \quad (36)
\end{aligned}$$

which we refer to as the "Gauss approximation". Introducing the average per-segment SNR $\rho_{\mathcal{F}}$ as $\rho_{\mathcal{F}}^2 \equiv \langle \rho_\Sigma^2 \rangle / N$, one can consider two interesting limits of the false-dismissal equation (34):

(i) *strong-signal limit* ($\rho_{\mathcal{F}}^2 \gg 1$): the per-segment SNR of the signal is large, and we obtain

$$\rho_\Sigma^* \approx \sqrt{8}\,\beta\,, \qquad (37)$$

which is somewhat pathological, as $\beta \gg 1$ and therefore the detection probability is *extremely* close to $p_{\mathrm{det}} = 1$. Neither false-alarm threshold nor the number of segments N matter for detectability[4] in this case.

(ii) *weak-signal limit* ($\rho_{\mathcal{F}}^2 \ll 1$): the per-segment SNR of the signal is small, and using $N \gg 1$ we find

$$\rho_\Sigma^{*2} \approx 2\sqrt{4N}\,(\alpha + \beta)\,, \qquad (38)$$

which we refer to as the "weak-signal Gauss approximation" (WSG), which was first used in [19] to estimate the sensitivity of the Hough method. This approach results in the "classic" semi-coherent sensitivity scaling as a function of N, namely

$$h_{\mathrm{th},\mathrm{WSG}}^{-2} = \frac{N_{\mathrm{det}}}{2S_{\mathrm{n}}}\,\frac{[1 - \xi(\widehat{m} + \widetilde{m})]}{\alpha + \beta}\,\frac{T}{\sqrt{N}}\,. \qquad (39)$$

In practice we find that the WSG approximation is often not well satisfied, and the deviations of the N-scaling in Eq. (38) from the exact form of Eq. (27) can lead to dramatically different optimal solutions. Already the Gauss approximation of Eq. (36) is not well satisfied for small false-alarm probabilities $p_{\mathrm{fA}} \ll 1$ and segment numbers in the range $N \lesssim \mathcal{O}(1000)$, as can be seen in Fig. 2. A more reliable approximation was recently introduced in [34], namely using the Gaussian distribution only for the false-dismissal equation (26), while keeping the central χ^2-distribution for the false-alarm equation (25). For the present work this approach would not be well-suited, however, as we need the sensitivity equation in the form of a power-law in T and N, similarly to Eq. (39).

2. Local power-law approximation for ρ_Σ^*

We can incorporate the exact N-scaling of the critical non-centrality parameter ρ_Σ^{*2} of Eq. (27) by *locally* expressing it as a power-law in the form

$$\rho_\Sigma^{*2}(p_{\mathrm{fA}}^*, p_{\mathrm{fD}}^*,\,N) = r_0\,N^{1/(2w)}\,, \qquad (40)$$

where $w(p_{\mathrm{fA}}^*, p_{\mathrm{fD}}^*, N_0)$ is a parameter quantifying the relative deviation of the exact N-scaling from the WSG limit of Eq. (38), where $w = 1$. The power-law coefficients can be computed as

$$w = \left(2\,\frac{\partial \log \rho_\Sigma^{*2}}{\partial \log N}\right)^{-1}\,, \quad r_0 = \rho_\Sigma^{*2}\,N_0^{-1/(2w)}\,, \qquad (41)$$

evaluated at a point $\{p_{\mathrm{fA}}^*, p_{\mathrm{fD}}^*, N_0\}$.

The function $w(N)$ is shown in Fig. 2, for a reference false-dismissal probability of $p_{\mathrm{fD}}^* = 0.1$ and different choices of false-alarm probability p_{fA}^*, both for the exact solution Eq. (27) and for the Gauss approximation of Eq. (36). We see that the exact N-scaling w increasingly deviates from the WSG approximation ($w = 1$) at *lower* false-alarm probabilities and at smaller N. The Gauss approximation tends to agree better with the exact scaling at larger N (as expected), and at *higher* false-alarm probabilities.

Using the power-law approximation of Eq. (40), we can now express the threshold signal strength of Eq. (29) as

$$\begin{aligned}
L_0(N, T, \widetilde{m}, \widehat{m}) &\equiv \frac{r_0\,S_{\mathrm{n}}}{2N_{\mathrm{det}}}\,h_{\mathrm{th}}^{-2} \\
&= [1 - \xi(\widetilde{m} + \widehat{m})]\,T\,N^{-1/(2w)}\,, \quad (42)
\end{aligned}$$

which defines the objective function L_0 that we want to maximize as a function of the StackSlide parameters.

We see that, without further constraints the optimal solution would simply be $m \to 0$, $N \to 1$ and $T \to T_{\mathrm{max}}$, i.e. a fully coherent search over all the available data T_{max} with an infinitely fine template bank. This would obviously require infinite computing power, and we therefore need to extend the optimization problem by a computing-cost constraint.

---

[3] The second solution has $\beta < 0$, corresponding to $p_{\mathrm{fD}} > 0.5$.
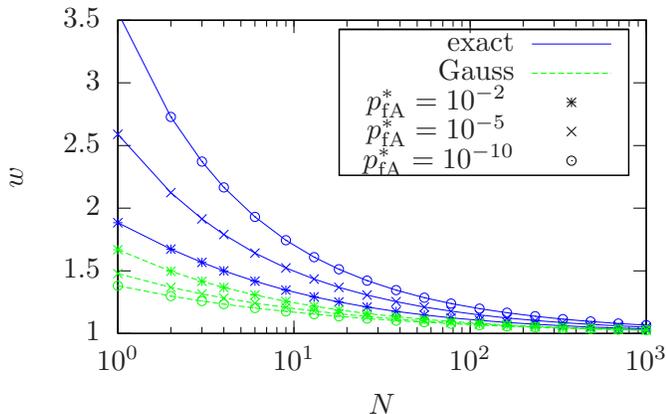[4] This has been noted previously for radio observations[35]

FIG. 2: N-scaling coefficient w defined in Eq. (40) as a function of N, for false-dismissal probability $p_{\text{fD}}^* = 0.1$, and different false-alarm probabilities $p_{\text{fA}}^* \in [10^{-10}, 10^{-5}, 10^{-2}]$. Solid lines show the scaling obtained from the exact solution Eq. (27), while dashed lines refer to the Gauss approximation of Eq. (36). The WSG approximation corresponds to $w = 1$.

D. Template counting

For both the coarse[5] and the fine grid, the respective number of templates[6] $\{\widetilde{\mathcal{N}}, \widehat{\mathcal{N}}\}$ covering the parameter space $\mathbb{P}$ is given [33, 36] by the general expression

$$\mathcal{N} = \theta_n \, m^{-n/2} \, \mathcal{V}_n, \quad \text{with} \quad \mathcal{V}_n \equiv \int_{\mathbb{T}_n} \sqrt{\det g} \, d^n\lambda, \quad (43)$$

where m is the maximal-mismatch parameter, $\det g$ is the determinant of the corresponding parameter-space metric g_{ij}, and $\mathcal{V}_n$ denotes the metric volume of the n-dimensional space $\mathbb{T}_n \subseteq \mathbb{P}$ spanned by the template-bank. The normalized thickness θ_n depends on the geometric structure of the covering, for example $\theta_{\mathbb{Z}_n} = n^{n/2} \, 2^{-n}$ for a hyper-cubic lattice $\mathbb{Z}_n$.

An important subtlety in Eq. (43) is the dimensionality n of the template-bank space $\mathbb{T}_n$, which can be smaller than the dimensionality of the parameter space $\mathbb{P}$, as previously discussed in [17, 21]. The template-bank dimensionality n is generally a (piece-wise constant) function of the StackSlide parameters $\{N, \Delta T, m\}$, which determine the metric resolution. The extent of $\mathbb{P}$ along certain directions can be "thin" compared to the metric resolution and would require only a single template along this direction, effectively not contributing to the template-bank dimensionality. For different StackSlide parameters, however, the resolution might be sufficient to require more

---

[5] We assume a roughly constant number of coarse-grid templates $\widetilde{\mathcal{N}}$ across all segments.
[6] The *templates* in this formulation are not to be confused with the "patches" used in BC [17] and CGK [21]. A "patch" in the BC/CGK framework corresponds to a *line* of templates along the frequency axis.

than one template along this direction, adding to the template-bank dimensionality n.

Following [14, 17, 21], the correct dimensionality for given StackSlide parameters can be determined by the condition that n should *maximize* the number of templates $\mathcal{N}_n$ computed via Eq. (43), i.e.

$$\widetilde{\mathcal{N}}_{\widetilde{n}} = \max_n \widetilde{\mathcal{N}}_n, \quad \text{and} \quad \widehat{\mathcal{N}}_{\widehat{n}} = \max_n \widehat{\mathcal{N}}_n. \quad (44)$$

This can be understood as follows: if $\mathcal{N}$ decreases when adding a template-bank dimension, then the corresponding parameter-space extent is thinner than the metric resolution and therefore adds "fractional" templates. On the other hand, if $\mathcal{N}$ decreases by removing a dimension, then its extent is thicker than the metric resolution and requires more than one template to cover it.

An interesting alternative formulation can be obtained by expressing Eq. (44) as the condition $\mathcal{N}_n/\mathcal{N}_{n-1} > 1$ for including an additional dimension n. For constant metrics and simple parameter-space shapes, i.e. $\mathcal{V}_n = \int \sqrt{g} \, d^n\lambda = \sqrt{g} \, \Delta\lambda_1 \times \Delta\lambda_2 \ldots \times \Delta\lambda_n$, this can be shown to be equivalent to

$$\frac{\theta_n}{\theta_{n-1}} \frac{\Delta\lambda_n}{d\lambda_n} > 1, \quad (45)$$

where $d\lambda_n \equiv \sqrt{m \, g^{nn}}$ is the metric *template extent* along dimension n, in terms of the diagonal element g^{nn} of the inverse metric g^{ij}. This shows that Eq. (44) boils down to (apart from the lattice-thickness ratio) the requirement that the parameter-space extent $\Delta\lambda_n$ along a given dimension n must exceed the corresponding metric template resolution $d\lambda_n$.

The coherent (coarse-grid) metric volume $\widetilde{\mathcal{V}}_{\widetilde{n}}$ is typically a steep function of the coherence time ΔT, and can often be well approximated (over some range of ΔT) by a power law, namely $\widetilde{\mathcal{V}}_{\widetilde{n}}(\Delta T) \propto \Delta T^{\widetilde{q}}$. We can therefore write Eq. (43) for $\widetilde{\mathcal{N}}$ in the power-law form

$$\widetilde{\mathcal{N}}_{\widetilde{n}}(\widetilde{m}, \Delta T) = \widetilde{k} \, \widetilde{m}^{-\widetilde{n}/2} \, \Delta T^{\widetilde{q}}, \quad (46)$$

where $\widetilde{k} = \theta_{\widetilde{n}} \widetilde{\mathcal{V}}_{\widetilde{n}}(\Delta T_0) \, \Delta T_0^{-\widetilde{q}}$ for some choice of segment length ΔT_0.

The semi-coherent (fine-grid) metric volume $\widehat{\mathcal{V}}_{\widehat{n}}$ generally depends on both ΔT and N and can typically [17, 26, 32] be factored in the form

$$\widehat{\mathcal{V}}_{\widehat{n}}(N, \Delta T) = \gamma_{\widehat{n}}(N) \, \widetilde{\mathcal{V}}_{\widehat{n}}(\Delta T), \quad (47)$$

in terms of the *refinement factor* $\gamma_{\widehat{n}}(N) \geq 1$ and the coherent-metric volume $\widetilde{\mathcal{V}}_{\widehat{n}}(\Delta T)$ of the fine-grid template space. Typically $\gamma(N)$ can be well approximated (over some range of N) by a power law, namely $\gamma(N) \propto N^{\widehat{p}}$. We can therefore write Eq. (43) for $\widehat{\mathcal{N}}$ in the power-law form

$$\widehat{\mathcal{N}}_{\widehat{n}}(\widehat{m}, \Delta T, N) = \widehat{k} \, \widehat{m}^{-\widehat{n}/2} \, \Delta T^{\widehat{q}} \, N^{\widehat{p}}, \quad (48)$$

where $\widehat{k} = \theta_{\widehat{n}} \widehat{\mathcal{V}}_{\widehat{n}}(\Delta T_0, N_0) \, \Delta T_0^{-\widehat{q}} N_0^{-\widehat{p}}$ for some choice of parameters $\{\Delta T_0, N_0\}$.

E. Computing-cost model

The total computing cost C_{tot} of the interpolating StackSlide statistic has two main contributions, namely

$$C_{\text{tot}}(\widetilde{m}, \widehat{m}, N, \Delta T) = \widetilde{C} + \widehat{C}, \qquad (49)$$

where $\widetilde{C}(\widetilde{m}, N, \Delta T)$ is the computing cost of the $\mathcal{F}$-statistic over the coarse grid of $\widetilde{\mathcal{N}}_{\widetilde{n}}$ templates for each of the N segments, and $\widehat{C}(\widehat{m}, N, \Delta T)$ is the cost of incoherently summing these $\mathcal{F}$-values across all segments on a fine grid of $\widehat{\mathcal{N}}_{\widehat{n}}$ templates. Note that we neglect all other costs such as data-IO etc, which for any computationally limited search will typically be much smaller than C_{tot}.

1. Computing cost $\widetilde{C}$ of the coherent step

The computing cost of the coherent step is

$$\widetilde{C}(\widetilde{m}, N, \Delta T) = N\,\widetilde{\mathcal{N}}_{\widetilde{n}}(\widetilde{m}, \Delta T)\,N_{\text{det}}\,\widetilde{c}_1(\Delta T), \qquad (50)$$

where $\widetilde{c}_1(\Delta T)$ is the $\mathcal{F}$-statistic computing cost of a single template for a single segment and a single detector. Here we used the fact that to first order [31] the number of detectors has no effect on the number of templates $\widetilde{\mathcal{N}}$.

As discussed previously in [21], there are two fundamentally different implementations of the $\mathcal{F}$-statistic calculation currently in use: a direct *SFT-method* [37], and a (generally far more efficient) *FFT-method* based on barycentric resampling [24, 38].

(i) The *SFT-method* consists in interpolating frequency bins of short Fourier transforms ("SFTs") of length T_{SFT}, using approximations described in [28, 37]. The resulting per-template cost $\widetilde{c}_1(\Delta T)$ is directly proportional to the segment length ΔT:

$$\widetilde{c}_1^{\text{SFT}}(\Delta T) = \widetilde{c}_0^{\text{SFT}}\,\frac{\Delta T}{T_{\text{SFT}}}, \qquad (51)$$

where $\widetilde{c}_0^{\text{SFT}}$ is an implementation- and hardware-dependent fundamental computing cost.

(ii) In the *FFT-method* the cost of searching a frequency band Δf using an (up-sampled by u) FFT frequency-resolution of $u/\Delta T$ is proportional to $\mathcal{N}_f\,\log 2\mathcal{N}_f$, where $\mathcal{N}_f = u\Delta f\,\Delta T$ is the number of frequency bins. We can therefore express the per-template $\mathcal{F}$-statistic cost $\widetilde{c}_1(\Delta T)$ as

$$\widetilde{c}_1^{\text{FFT}}(\Delta T) = \widetilde{c}_0^{\text{FFT}}\,\log(2u\Delta f\,\Delta T), \qquad (52)$$

where $\widetilde{c}_0^{\text{FFT}}$ is an implementation- and hardware-dependent fundamental computing cost.

Using the power-law model of Eq. (46) for $\widetilde{\mathcal{N}}$, we can write the coherent computing cost in the form

$$\widetilde{C}(\widetilde{m}, N, \Delta T) = \widetilde{\kappa}\,\widetilde{m}^{-\widetilde{n}/2}\,N^{\widetilde{\eta}}\,\Delta T^{\widetilde{\delta}}, \qquad (53)$$

where

$$\widetilde{\eta} = 1, \quad \widetilde{\delta} = \widetilde{q} + \Delta\widetilde{\delta}, \qquad (54)$$

and where $\Delta\widetilde{\delta}$ is either

$$\Delta\widetilde{\delta} = \begin{cases} \Delta\widetilde{\delta}_{\text{SFT}} \equiv 1, \\ \Delta\widetilde{\delta}_{\text{FFT}} \equiv [\log(2u\Delta f\Delta T_0)]^{-1}, \end{cases} \qquad (55)$$

depending on whether the $\mathcal{F}$-statistic is computed using the *SFT-* or *FFT-method*, respectively. The expression for $\Delta\widetilde{\delta}_{\text{FFT}}$ can be obtained via Eq. (62) and depends (albeit weakly) on the reference segment length ΔT_0. The corresponding proportionality factors $\widetilde{\kappa}$ are found as

$$\begin{aligned} \widetilde{\kappa}_{\text{SFT}} &= \theta_{\widetilde{n}}\,N_{\text{det}}\,\frac{\widetilde{c}_0^{\text{SFT}}}{T_{\text{SFT}}}\,\frac{\widetilde{\mathcal{V}}_{\widetilde{n}}(\Delta T_0)}{\Delta T_0^{\widetilde{q}}}, \\[2mm] \widetilde{\kappa}_{\text{FFT}} &= \theta_{\widetilde{n}}\,N_{\text{det}}\,\frac{\widetilde{c}_0^{\text{FFT}}}{\Delta\widetilde{\delta}_{\text{FFT}}}\,\frac{\widetilde{\mathcal{V}}_{\widetilde{n}}(\Delta T_0)}{\Delta T_0^{\widetilde{\delta}}}. \end{aligned} \qquad (56)$$

2. Computing cost $\widehat{C}$ of the incoherent step

The computing cost of the StackSlide step is

$$\widehat{C}(\widehat{m}, N, \Delta T) = N\,\widehat{\mathcal{N}}_{\widehat{n}}(\widehat{m}, \Delta T, N)\,\widehat{c}_0, \qquad (57)$$

where $\widehat{c}_0$ is the implementation- and hardware-dependent fundamental cost of adding one value of $2\mathcal{F}_k$ for one fine-grid point $\widehat{\lambda}$ in Eq. (2), including the cost of mapping the fine-grid point to its closest coarse-grid template $\widetilde{\lambda}_k(\widehat{\lambda})$. The incoherent step operates on coherent multi-detector $\mathcal{F}$-statistic values, and therefore does not depend on the number of detectors N_{det}.

Using the power-law model of Eq. (48) for $\widehat{\mathcal{N}}$, we can write the incoherent computing cost as

$$\widehat{C}(\widehat{m}, N, \Delta T) = \widehat{\kappa}\,\widehat{m}^{-\widehat{n}/2}\,N^{\widehat{\eta}}\,\Delta T^{\widehat{\delta}}, \qquad (58)$$

where

$$\widehat{\eta} = \widehat{p} + 1, \quad \widehat{\delta} = \widehat{q}, \qquad (59)$$

and the proportionality factor

$$\widehat{\kappa} = \theta_{\widehat{n}}\,\widehat{c}_0\,\frac{\widehat{\mathcal{V}}_{\widehat{n}}(N_0, \Delta T_0)}{N_0^{\widehat{p}}\,\Delta T_0^{\widehat{q}}}, \qquad (60)$$

for given reference values $\{N_0, \Delta T_0\}$.

3. General power-law computing-cost model

Combining Eqs. (53) and (58) we arrive at the following power-law model for the total computing cost, defined in Eq. (49), namely

$$C_{\text{tot}} = \widetilde{\kappa}\,\widetilde{m}^{-\widetilde{n}/2}\,N^{\widetilde{\eta}}\,\Delta T^{\widetilde{\delta}} + \widehat{\kappa}\,\widehat{m}^{-\widehat{n}/2}\,N^{\widehat{\eta}}\,\Delta T^{\widehat{\delta}}. \qquad (61)$$

If a given computing-cost function does not follow this model, we can always produce a local fit to Eq. (61), which should be valid over some range of parameters $\{\Delta T, N\}$, namely

$$\delta \equiv \frac{\partial \log C}{\partial \log \Delta T} \ , \quad \eta \equiv \frac{\partial \log C}{\partial \log N} \ , \quad (62)$$

$$\kappa \equiv \frac{C(m_0, N_0, \Delta T_0)}{m_0^{-n/2} N_0^\eta \Delta T_0^\delta} \ , \quad (63)$$

for reference values $\{m_0, N_0, \Delta T_0\}$. Note that δ generally depends only on ΔT_0, while η depends only on N_0, due to the way these dependencies typically factor (cf. Sec. III E). The mismatch dependency $\propto m^{-n/2}$ is exact according to Eq. (43), but a given computing-cost function might still deviate from this behaviour (e.g. the BC/CGK computing-cost function discussed in Sec. V C). In this case one can extend the power-law fit by extracting the "mismatch-dimension" n via

$$n \equiv -2 \ \frac{\partial \log C}{\partial \log m} \ . \quad (64)$$

It will be more convenient in the following to work in terms of $\{N, T\}$ instead of $\{N, \Delta T\}$, where $T = N \Delta T$ is the total time span of data used. Changing variables, we obtain the computing-cost model in the form

$$C_{\text{tot}} = \widetilde{\kappa} \, \widetilde{m}^{-\widetilde{n}/2} \, N^{-\widetilde{\varepsilon}} \, T^{\widetilde{\delta}} + \widehat{\kappa} \, \widehat{m}^{-\widehat{n}/2} \, N^{-\widehat{\varepsilon}} \, T^{\widehat{\delta}} , \quad (65)$$

where we defined

$$\varepsilon \equiv \delta - \eta , \quad (66)$$

generally satisfying $\varepsilon > 0$ for all realistic cases considered here. Note that m and N are dimensionless, therefore the respective units of $[C/\kappa]$ are $[T^\delta]$.

IV. MAXIMIZING SENSITIVITY AT FIXED COMPUTING COST

We want to maximize the objective function $L_0 \propto h_{\text{th}}^{-2}$ defined in Eq. (42) under the constraint of fixed computing cost, $C_{\text{tot}} = C_0$. We therefore need to find the stationary points of the Lagrange function

$$L(N, T, \widetilde{m}, \widehat{m}, \varpi) = L_0 + \varpi \left[\widetilde{C} + \widehat{C} - C_0 \right], \quad (67)$$

where stationarity with respect to the Lagrange multiplier, i.e. $\partial_\varpi L = 0$, returns the computing-cost constraint $\widetilde{C} + \widehat{C} = C_0$.

Before embarking on the full optimization problem, it is instructive to consider two special cases, namely (i) a fully coherent search, and (ii) searches where the computing cost is dominated by one contribution, either coherent $\widetilde{C}$ or incoherent $\widehat{C}$.

A. Special case (i): Fully coherent search

The fully coherent search is a special case of Eq. (67) with the additional constraint $N = 1$, and therefore $\Delta T = T$, $\widehat{m} = 0$, and $\widehat{C} = 0$. This leaves us with the reduced Lagrangian

$$L(T, \widetilde{m}, \varpi) = (1 - \xi \widetilde{m}) \, T + \varpi \left[\widetilde{\kappa} \, \widetilde{m}^{-\widetilde{n}/2} \, T^{\widetilde{\delta}} - C_0 \right]. \quad (68)$$

Requiring stationarity with respect to $\{T, \widetilde{m}, \varpi\}$ results in the optimal StackSlide parameters

$$\xi \, \widetilde{m}_{\text{opt}} = \left(1 + \frac{2\widetilde{\delta}}{\widetilde{n}} \right)^{-1} , \quad (69)$$

$$T_{\text{opt}} = \left(\frac{C_0}{\widetilde{\kappa}} \right)^{1/\widetilde{\delta}} \widetilde{m}_{\text{opt}}^{\widetilde{n}/(2\widetilde{\delta})} . \quad (70)$$

Interestingly the optimal mismatch $\widetilde{m}_{\text{opt}}$ is independent of both the computing-cost constraint C_0 and the observation time T. The scaling of the resulting threshold signal strength h_{th} with computing cost C_0 is therefore

$$h_{\text{th}}^{-1} \propto C_0^{1/(2\widetilde{\delta})} . \quad (71)$$

In practical applications we often find $\widetilde{\delta} \approx 3 - 7$, and so T_{opt} and h_{th}^{-1} will increase very slowly with increasing computing cost C_0. This indicates that a brute-force approach of throwing more computing power at a fully coherent search will typically yield meagre returns in sensitivity.

B. Special case (ii): Computing cost dominated by one contribution

If either the coherent $\widetilde{C}$ or incoherent $\widehat{C}$ contribution dominates the total computing cost (65), we can write

$$C_{\text{tot}} \approx \kappa \, m^{-n/2} \, N^{-\varepsilon} \, T^\delta , \quad (72)$$

where all StackSlide parameters now refer to dominant contribution only.

We assume that the negligible computing-cost contribution implies that we can also neglect the corresponding mismatch: if the respective step is cheap, one can easily increase sensitivity by reducing the corresponding mismatch until it is negligible, i.e. we assume $\langle \mu_{\text{sc}} \rangle \approx \xi \, m$. This qualitative argument will be confirmed by the general solution in the next section. We can therefore write the objective function Eq. (42) as

$$L_0(N, T, m) \approx (1 - \xi \, m) \, N^{-1/(2w)} \, T . \quad (73)$$

Using Eq. (72) we can obtain

$$N(C_0, m, T) = (C_0/\kappa)^{-1/\varepsilon} \, m^{-n/(2\varepsilon)} \, T^{\delta/\varepsilon} , \quad (74)$$

$$\Delta T(C_0, m, T) = (C_0/\kappa)^{1/\varepsilon} \, m^{n/(2\varepsilon)} \, T^{-\eta/\varepsilon} , \quad (75)$$

which shows that increasing T at fixed C_0 results in more and shorter segments, while increasing C_0 at fixed T results in fewer and longer segments (assuming $\varepsilon > 0$). Substituting this into Eq. (73) yields the threshold signal strength

$$h_{\mathrm{th}}^{-2} \propto (C_0/\kappa)^{1/(2w\varepsilon)} \left[(1 - \xi m)\, m^{n/(4w\varepsilon)}\right]\, T^{a/(2w\varepsilon)}, \tag{76}$$

where we introduced the parameter

$$a \equiv 2w\varepsilon - \delta, \tag{77}$$

which will be of critical importance in determining the character of the optimal solution.

The objective function $L_0 \propto h_{\mathrm{th}}^{-2}$ can be easily maximized over mismatch m, resulting in

$$\xi\, m_{\mathrm{opt}}^{(0)} = \left[1 + \frac{4w\varepsilon}{n}\right]^{-1}, \tag{78}$$

which is independent of both C_0 and T. This solution differs from Eq. (69) of the fully coherent case, even when the coherent cost dominates (where $\widetilde{\varepsilon} = \widetilde{\delta} - 1$).

We see in Eq. (76) that there is *no extremum* of h_{th} (at least in regions of approximately constant power-law exponents). Given that $w \geq 1$ and generally $\varepsilon > 0$, we can distinguish two different regimes depending on the sign of critical scaling exponent a defined in Eq. (77):

$a > 0$: sensitivity improves (i.e. h_{th}^{-1} increases) with *increasing* T (at fixed C_0). Therefore sensitivity is only limited by the total amount of data $T_{\max}$ available.

$a < 0$: sensitivity improves (i.e. h_{th}^{-1} decreases) with *decreasing* T, so one should use less data (until the assumptions change).

In practice these extreme conclusions will be modified, as the power-law exponents will vary (slowly) as functions of N and T, and the assumption of a dominating computing-cost contribution might also no longer be satisfied. The marginal case $a = 0$ marks a possible sensitivity maximum, namely if increasing T results in $a < 0$ and decreasing T leads to $a > 0$.

We can obtain a useful qualitative picture of the full optimization problem by considering the two extreme cases of dominating computing contribution $\widetilde{C}$ or $\widehat{C}$:

- if $\widetilde{C} \gg \widehat{C}$: we always have $\widetilde{a} > 0$ (for all cases of interest $\widetilde{\eta} = 1$, $\widetilde{\delta} > 2$ and $w \geq 1$). Therefore sensitivity improves with increasing T. As seen in Sec. IV C 3 this shifts computing cost to the incoherent contribution. Eventually one either uses all the data $T_{\max}$ or the coherent cost no longer dominates.

- if $\widehat{C} \gg \widetilde{C}$: the incoherent parameter $\widehat{a}$ can have any sign. If $\widehat{a} > 0$ one would *increase* T until all the data $T_{\max}$ is used (or we reach $\widehat{a} = 0$). If $\widehat{a} < 0$ one would *decrease* T until the incoherent cost no longer dominates.

These limiting cases show that the *type* of optimal solution will be determined solely by the *incoherent* critical exponent $\widehat{a} = 2w\widehat{\varepsilon} - \widehat{\delta}$, namely

$$T_{\mathrm{opt}}^{(0)} = \begin{cases} \text{finite} & \text{if } \widehat{a} \leq 0, \\ \infty & \text{otherwise}, \end{cases} \tag{79}$$

which we refer to as the *bounded* and the *unbounded* regime, respectively.

C. General optimality conditions

We now return to the full optimization problem of Eq. (67), namely:

$$L = L_0 + \varpi\left[\widetilde{C} + \widehat{C} - C_0\right], \tag{80}$$

$$L_0 = [1 - \xi(\widetilde{m} + \widehat{m})]\, T\, N^{-1/(2w)}, \tag{81}$$

$$\widetilde{C} = \widetilde{\kappa}\, \widetilde{m}^{-\widetilde{n}/2}\, N^{-\widetilde{\varepsilon}}\, T^{\widetilde{\delta}}, \tag{82}$$

$$\widehat{C} = \widehat{\kappa}\, \widehat{m}^{-\widehat{n}/2}\, N^{-\widehat{\varepsilon}}\, T^{\widehat{\delta}}. \tag{83}$$

It will be useful introduce the computing-cost *ratio*

$$\varkappa \equiv \widetilde{C}/\widehat{C}, \tag{84}$$

and express the respective contributions as

$$\widetilde{C} = \frac{C_0}{1 + \varkappa^{-1}}, \quad \widehat{C} = \frac{C_0}{1 + \varkappa}. \tag{85}$$

Using Eqs. (82), (83) to solve for T and N, respectively, we obtain

$$N^D = \frac{(C_0/\widehat{\kappa})^{\widetilde{\delta}}}{(C_0/\widetilde{\kappa})^{\widehat{\delta}}} \frac{\left[\widetilde{m}^{-\widetilde{n}/2}\left(1 + \varkappa^{-1}\right)\right]^{\widehat{\delta}}}{\left[\widehat{m}^{-\widehat{n}/2}\left(1 + \varkappa\right)\right]^{\widetilde{\delta}}}, \tag{86}$$

$$T^D = \frac{(C_0/\widehat{\kappa})^{\widetilde{\varepsilon}}}{(C_0/\widetilde{\kappa})^{\widehat{\varepsilon}}} \frac{\left[\widetilde{m}^{-\widetilde{n}/2}\left(1 + \varkappa^{-1}\right)\right]^{\widehat{\varepsilon}}}{\left[\widehat{m}^{-\widehat{n}/2}\left(1 + \varkappa\right)\right]^{\widetilde{\varepsilon}}}, \tag{87}$$

where D is the determinant of the matrix $[\widetilde{\delta}, \widetilde{\eta}; \widehat{\delta}, \widehat{\eta}]$, which for all cases of practical interest seems to be positive definite, namely

$$D \equiv \widetilde{\delta}\widehat{\eta} - \widehat{\delta}\,\widetilde{\eta} > 0. \tag{88}$$

The segment length $\Delta T = T/N$ can similarly be obtained as

$$\Delta T^D = \frac{(C_0/\widetilde{\kappa})^{\widehat{\eta}}}{(C_0/\widehat{\kappa})^{\widetilde{\eta}}} \frac{\left[\widehat{m}^{-\widehat{n}/2}\left(1 + \varkappa\right)\right]^{\widetilde{\eta}}}{\left[\widetilde{m}^{-\widetilde{n}/2}\left(1 + \varkappa^{-1}\right)\right]^{\widehat{\eta}}}. \tag{89}$$

1. Stationarity with respect to mismatches $\{\widetilde{m}, \widehat{m}\}$

Requiring stationary with respect to the mismatches, i.e. $\partial_{\widetilde{m}} L = \partial_{\widehat{m}} L = 0$, yields

$$
\begin{aligned}
\varpi \widetilde{C} &= -2\xi \, \frac{\widetilde{m}_{\mathrm{opt}}}{\widetilde{n}} \, T N^{-1/(2w)} \,, \\
\varpi \widehat{C} &= -2\xi \, \frac{\widehat{m}_{\mathrm{opt}}}{\widehat{n}} \, T N^{-1/(2w)} \,,
\end{aligned}
\tag{90}
$$

which results in the remarkable relation

$$
\frac{\widetilde{m}_{\mathrm{opt}}/\widetilde{n}}{\widehat{m}_{\mathrm{opt}}/\widehat{n}} = \varkappa \,.
\tag{91}
$$

The ratio of optimal mismatch per dimension is simply given by the computing-cost ratio $\varkappa$. This result confirms an assumption made in Sec. IV B about the optimal solution, namely that a negligible computing-cost contributions also implies that one can neglect the corresponding mismatch.

2. Stationarity with respect to number of segments N

Requiring stationarity with respect to N (treated as continuous), i.e. $\partial_N L = 0$ yields

$$
L_0 + 2w \, \varpi \left[\widetilde{\varepsilon} \, \widetilde{C} + \widehat{\varepsilon} \, \widehat{C} \right] = 0 \,,
\tag{92}
$$

and substituting Eqs. (90) and (81), we obtain

$$
\frac{\widetilde{m}_{\mathrm{opt}}}{\widetilde{m}_{\mathrm{opt}}^{(0)}} + \frac{\widehat{m}_{\mathrm{opt}}}{\widehat{m}_{\mathrm{opt}}^{(0)}} = 1 \,,
\tag{93}
$$

where we used the asymptotic optimal mismatches $m_{\mathrm{opt}}^{(0)}$ defined in Eq. (78) for the two limiting cases of dominating coherent or incoherent computing-cost, respectively. Equation (93) can be interpreted as defining a two-dimensional ellipse in $\sqrt{m}$ with semi-major axes $\sqrt{m_{\mathrm{opt}}^{(0)}}$. Combining this with Eq. (91) we obtain the optimal mismatches

$$
\begin{aligned}
\frac{\widetilde{n}}{\widetilde{m}_{\mathrm{opt}}} &= \frac{\widetilde{n}}{\widetilde{m}_{\mathrm{opt}}^{(0)}} + \frac{\widehat{n}}{\widehat{m}_{\mathrm{opt}}^{(0)}} \, \varkappa^{-1} \,, \\
\frac{\widehat{n}}{\widehat{m}_{\mathrm{opt}}} &= \frac{\widehat{n}}{\widehat{m}_{\mathrm{opt}}^{(0)}} + \frac{\widetilde{n}}{\widetilde{m}_{\mathrm{opt}}^{(0)}} \, \varkappa \,,
\end{aligned}
\tag{94}
$$

which reduces to the limiting cases of Eq. (78) when either computing cost dominates, i.e. when $\varkappa \ll 1$ or $\varkappa \gg 1$. We can express the optimal mismatch prefactor in Eq. (81) as

$$
[1 - \xi(\widetilde{m} + \widehat{m})]_{\mathrm{opt}} = \left[1 + \frac{1}{4w} \frac{\widetilde{n}\varkappa + \widehat{n}}{\widetilde{\varepsilon}\varkappa + \widehat{\varepsilon}} \right]^{-1} \,.
\tag{95}
$$

The optimal mismatches Eq. (94) only depend on the computing-cost ratio $\varkappa$. Substituting into Eq. (87) we

therefore obtain a relation of the form $T_0 = T(C_0, \varkappa_{\mathrm{opt}})$ for given observation time T_0, which can (numerically) be solved for $\varkappa_{\mathrm{opt}} = \varkappa(C_0, T_0)$. Similarly, one could specify N_0 and solve Eq. (86) for $\varkappa_{\mathrm{opt}} = \varkappa(C_0, N_0)$. In either case the optimal mismatches are obtained from Eq. (94) and the optimal number and length of segments from Eqs. (86) and (87), fully closing the optimal solution at fixed T.

3. Monotonicity relations with T

It is interesting to consider the behaviour of the optimal "fixed-T" solution of the previous section as a function of T. We see in Eq. (94) that $\widetilde{m}_{\mathrm{opt}}$ is monotonically increasing with $\varkappa$, while $\widehat{m}_{\mathrm{opt}}$ is decreasing, i.e.

$$
\partial_{\varkappa} \widetilde{m}_{\mathrm{opt}} > 0 \,, \quad \text{and} \quad \partial_{\varkappa} \widehat{m}_{\mathrm{opt}} < 0 \,.
\tag{96}
$$

We generally assume $D \equiv \widetilde{\delta}\widehat{\eta} - \widehat{\delta}\widetilde{\eta} > 0$ and $\varepsilon > 0$ which implies that the right-hand side of Eq. (87) is monotonically *decreasing* with $\varkappa$, while the left-hand side is monotonically increasing with T. Therefore $\varkappa$ must be montonically *decreasing* with T, i.e.

$$
\partial_T \varkappa < 0 \,.
\tag{97}
$$

Therefore the optimal solution shifts computing cost from the coherent to the incoherent step with increasing T, which had already been used in Sec. IV B. Combining this with Eq. (96) we find

$$
\partial_T \widetilde{m}_{\mathrm{opt}} < 0 \,, \quad \text{and} \quad \partial_T \widehat{m}_{\mathrm{opt}} > 0 \,,
\tag{98}
$$

and using this with Eqs. (86) and (89), we can further deduce

$$
\partial_T N_{\mathrm{opt}} > 0 \,, \quad \text{and} \quad \partial_T \Delta T_{\mathrm{opt}} < 0 \,,
\tag{99}
$$

namely increasing T results in more segments of shorter duration.

4. Stationarity with respect to observation time T

Requiring stationarity of L with respect to T, i.e. $\partial_T L = 0$, yields the final condition

$$
L_0 + \varpi \left[\widetilde{\delta} \, \widetilde{C} + \widehat{\delta} \, \widehat{C} \right] = 0 \,,
\tag{100}
$$

which combined with Eq. (92) results in

$$
\widetilde{a} \, \widetilde{C} + \widehat{a} \, \widehat{C} = 0 \,,
\tag{101}
$$

where the critical exponents a are defined in Eq. (77). We generally expect $\widetilde{a} > 0$, as discussed in Sec. IV B, and therefore the stationarity condition can only have a solution if

$$
\widehat{a} \equiv \widehat{\delta} - 2\widehat{\eta} + 2(w-1)\widehat{\varepsilon} < 0 \,.
\tag{102}
$$

This conclusion is consistent with the analysis of Sec. IV B: $\widehat{a} < 0$ characterizes a *bounded* regime with finite optimal $T_{\mathrm{opt}}^{(0)}$, while $\widehat{a} > 0$ characterizes an *unbounded* regime with $T_{\mathrm{opt}}^{(0)} \to \infty$.

If $T_{\mathrm{opt}}^{(0)}$ exceeds the available data T_{max}, then we simply apply the fixed-T solution of Sec. IV C 2. Otherwise Eq. (101) directly yields

$$\varkappa_{\mathrm{opt}} = -\frac{\widehat{a}}{\widetilde{a}}, \qquad (103)$$

closing the optimal solution via Eqs. (94), (86) and (87).

5. Monotonicity relations with C_0

For a bounded optimal solution with $T_{\mathrm{opt}}^{(0)} \leq T_{\mathrm{max}}$, we see from Eq. (103) that $\varkappa_{\mathrm{opt}}$ and $\{\widetilde{m}_{\mathrm{opt}}, \widehat{m}_{\mathrm{opt}}\}$ are independent of the computing-cost constraint C_0. Inserting Eqs. (86),(87) into Eq. (42), we can therefore read off the scaling

$$h_{\mathrm{th}}^{-1} \propto C_0^{(\widetilde{a}-\widehat{a})/(4wD)}, \qquad (104)$$

which shows that any "reasonable" search should satisfy

$$\widetilde{a} > \widehat{a}, \qquad (105)$$

in order for sensitivity to *improve* with increasing C_0 (assuming $D > 0$). Furthermore, from Eqs. (86), (87) and (89) we obtain the monotonicity relations:

$$\begin{aligned} \partial_{C_0} N_{\mathrm{opt}} &\propto \widetilde{\delta} - \widehat{\delta}, \\ \partial_{C_0} T_{\mathrm{opt}} &\propto \widetilde{\varepsilon} - \widehat{\varepsilon}, \\ \partial_{C_0} \Delta T_{\mathrm{opt}} &\propto \widehat{\eta} - \widetilde{\eta}. \end{aligned} \qquad (106)$$

We expect $\widehat{\eta} > \widetilde{\eta} = 1$, therefore the optimal segment length ΔT_{opt} will generally increase with C_0.

The behaviour of the optimal number of segments is less clear-cut: if $\widetilde{\delta} < \widehat{\delta}$ then N_{opt} *decreases* with C_0, which can result in a fully coherent search being optimal, despite $T_{\mathrm{opt}}^{(0)} < T_{\mathrm{max}}$. A StackSlide search is therefore not guaranteed to be more sensitive than a fully coherent search at the same computing power, even when computationally limited.

Similarly, T_{opt} can either increase with C_0 (if $\widetilde{\varepsilon} > \widehat{\varepsilon}$), or decrease: a more expensive and more sensitive search can be using less data.

V. EXAMPLES OF PRACTICAL APPLICATION

In order to illustrate the practical application of this analytical framework and its potential gains in sensitivity we consider a few different examples of CW searches.

A. Directed searches for isolated neutron stars

Directed searches target NSs with *known* sky-position but unknown frequency and frequency derivatives, i.e. $\{f, \dot{f}, \ddot{f}, \ldots\}$. The approximate phase metric of this parameter space for isolated NSs is known analytically and constant over the parameter space, e.g. see [Eq. (10) in [15]]. The number of coarse-grid templates scales as

$$\widetilde{\mathcal{N}} \propto \Delta T^{n(n+1)/2}, \qquad (107)$$

while the refinement of the semi-coherent metric [Eq. (92) in [26]] scales as

$$\gamma_n \propto N^{n(n-1)/2}. \qquad (108)$$

The coherent computing-cost exponents Eq. (54) are therefore

$$\widetilde{\delta} = \widetilde{n}(\widetilde{n}+1)/2 + \Delta\widetilde{\delta}, \quad \widetilde{\eta} = 1, \qquad (109)$$

where $\Delta\widetilde{\delta}$ depends on the $\mathcal{F}$-statistic implementation as given by Eq. (55). The incoherent computing-cost exponents Eq. (59) are

$$\widehat{\delta} = \widehat{n}(\widehat{n}+1)/2, \quad \widehat{\eta} = 1 + \widehat{n}(\widehat{n}-1)/2, \qquad (110)$$

which results in $\widehat{\varepsilon} = \widehat{n} - 1$.

For $\widetilde{n} \geq 2$ the condition $\widetilde{a} = \widetilde{\delta} - 2 + 2(w-1)\widetilde{\varepsilon} > 0$ holds as expected, while the critical boundedness parameter of Eq. (102) now reads as

$$\widehat{a} = \frac{\widehat{n}}{2}(3 - \widehat{n}) - 2 + 2(w-1)(\widehat{n}-1), \qquad (111)$$

which for the first few values of n evaluates to

$$\begin{aligned} \widehat{a}_1 &= -1, \\ \widehat{a}_2 &= -3 + 2w, \\ \widehat{a}_3 &= -6 + 4w, \\ \widehat{a}_4 &= -10 + 6w. \end{aligned} \qquad (112)$$

In the WSG limit (i.e. $w \to 1$) this is always $\widehat{a} < 0$, and therefore the search falls into the bounded regime. However, in general $w > 1$ and therefore directed StackSlide searches can be either bounded or unbounded.

Directed search for Cassiopeia-A

As a concrete example we consider the directed search for the compact object in Cassiopeia-A (CasA). This search has been performed using LIGO S5 data, and the resulting upper limits have been published in [7]. For the present example we use the search setup as originally proposed in [15], namely a fully coherent $\mathcal{F}$-statistic search (using the "SFT" method, i.e. $\Delta\widetilde{\delta} = 1$) using data spanning $T = 12\,$days, with a maximal template-bank mismatch of $\widetilde{m} = 0.2$. The setup assumed two detectors

with identical noise floor S_n and a 70% duty cycle, which we can formally incorporate as $N_{det} = 2 \times 0.7 = 1.4$ in Eqs. (50) and (29). The parameter space spanned a frequency range of $f \in [100, 300]$ Hz and spindown-ranges corresponding to a spindown-age of $\tau = 300$ y, see [15]. The template-bank dimension for the given StackSlide parameters was determined as $\widetilde{n} = 3$, resulting in a power-law scaling of $\widetilde{\delta} = 7$ according to Eq. (109).

In order to compare sensitivity estimates of different search setups, we use nominal (per-template) false-alarm and false-dismissal probabilities of

$$ p_{fA}^* = 10^{-10}, \qquad p_{fD}^* = 0.1 \,. \tag{113} $$

We use a rough estimate of $\xi = 0.5$ (e.g. see [Fig. 8 in [33]]) for the geometric average-mismatch factor of the A_3^*-lattice that was used in this search. Integrating Eqs. (25),(26) and solving for $\rho_{\mathcal{F}}^*$ yields $\rho_{\mathcal{F}}^* \approx 8.35$. Substituting this into Eq. (29) with $\widehat{m} = 0$, $\widetilde{m} = 0.2$ yields an estimate for the weakest detectable signal h_{th} of the original $\mathcal{F}$-statistic search:

$$ \left. \frac{h_{th}}{\sqrt{S_n}} \right|_{ref} \approx 5.2 \times 10^{-3} \sqrt{\text{Hz}}, \tag{114} $$

Timing a current StackSlide code using the SFT-method, one can extract approximate timing parameters

$$ \widehat{c}_0^{SFT} = 7 \times 10^{-8} \,\text{s}, \qquad \widehat{c}_0 = 6 \times 10^{-9} \,\text{s} \,, \tag{115} $$

which results in a total computing cost for the original search[7] of $C_0 \approx 472$ days on a single cluster node. This number is used as the computing-cost constraint C_0 for this example.

First we consider an optimal *coherent* search as described in Sec. IV A, namely using Eq. (69) we find

$$ \xi \, \widetilde{m}_{opt} \approx 0.18 \quad \Longrightarrow \widetilde{m}_{opt} \approx 0.36 \,, \tag{116} $$

and using Eq. (70) this results in $T_{opt} = 13.6$ days, which is only about $\sim 13\%$ longer then the original search proposal of [15]. The total improvement in the minimal signal strength h_{th} is less than 2% compared to Eq. (114), which shows that the original search proposal was remarkably close to an optimal coherent search.

Next we consider a StackSlide search over the same parameter space using the same computing cost C_0. Assuming the optimal solution will have segment lengths in the range 1 days $\lesssim \Delta T \lesssim 7$ days, and a total span of $T \lesssim 365$ days, the parameter-space dimensions would be $\widetilde{n} = 2$, $\widehat{n} = 3$ (see [15]). This results in power-law exponents $\widetilde{\delta} = 4$, $\widehat{\delta} = 6$, $\widehat{\eta} = 4$, and therefore $\widetilde{\varepsilon} = 3$, $\widehat{\varepsilon} = 2$, and $\widetilde{a} = 2 + 6(w-1)$, $\widehat{a} = -2 + 4(w-1)$. In order to simplify the example we use the WSG approximation, i.e.

---

[7] Using the original timing constant $\widetilde{c}_{SFT}^{(0)} = 6 \times 10^{-7}$ s of [15], we correctly recover the original estimate of $\widetilde{C} \approx 20 \times 200$ days

$w = 1$, which implies that the search would be bounded ($\widehat{a} < 0$). We can therefore use Eq. (103) to obtain the optimal computing-cost ratio as

$$ \varkappa_{opt} = 1 \,. \tag{117} $$

Note that when $w > 1.25$ we would have $\widehat{a} > 0$ and therefore this search would become unbounded. From Eq. (78) we obtain $\widetilde{m}_{opt}^{(0)} \approx 0.29$, $\widehat{m}_{opt}^{(0)} \approx 0.55$, and using Eq. (94) we find the respective optimal mismatches as

$$ \widetilde{m}_{opt} \approx 0.16 \,, \quad \widehat{m}_{opt} \approx 0.24 \,. \tag{118} $$

Using Eq. (63) we can extract the computing-cost coefficients $\widetilde{\kappa} \approx 3.14 \times 10^{-17}$ and $\widehat{\kappa} \approx 3.12 \times 10^{-34}$ (with time measured in seconds), and plugging this into Eqs. (86), (87) we find the optimal StackSlide parameters as

$$ N_{opt} = 139 \,, \quad \Delta T_{opt} \approx 1.9 \,\text{days} \,, \\ T_{opt} \approx 266.5 \,\text{days} \,, \tag{119} $$

which is self-consistent with the initially-assumed template-bank dimensions, as it falls into the assumed ranges for ΔT and T.

We can estimate the resulting sensitivity by solving Eqs. (25),(26), which yields $\rho_{\Sigma}^* \approx 17.3$, and substituting into Eq. (29) we find a weakest detectable signal strength h_{th} of

$$ \left. \frac{h_{th}}{\sqrt{S_n}} \right|_{opt} \approx 2.4 \times 10^{-3} \sqrt{\text{Hz}} \,, \tag{120} $$

which is an improvement on the optimal coherent sensitivity by more than a factor of two.

Figure 3 illustrates the behaviour of the optimal solution as a function of T without using the WSG approximation. This is obtained by numerically solving Eq. (87) for $\varkappa_{opt}(T)$, which yields $m_{opt}(T)$ via Eq. (94) and $N_{opt}(T)$ via Eq. (86). We see that the non-WSG approximated optimal solution results in somewhat different StackSlide parameters than the WSG solution of Eq. (119), but it hardly gains any further sensitivity.

Increasing the total computing cost C_0 would increase the relative advantage of the StackSlide method compared to a fully-coherent search: the coherent search would gain sensitivity as $h_{th}^{-1} \propto C_0^{1/14}$ according to Eq. (71), while the StackSlide search would gain sensitivity as $h_{th}^{-1} \propto C_0^{1/10}$ according to Eq. (104) (in the WSG approximation), so here the StackSlide search is more "efficient" at converting increases of computing-power into gains of sensitivity.

B. All-sky CW search using Einstein@Home

As an example for a wide parameter space all-sky search with massive computing power, we consider two recent CW searches performed on the Einstein@Home computing platform [6, 39, 40], namely the StackSlide
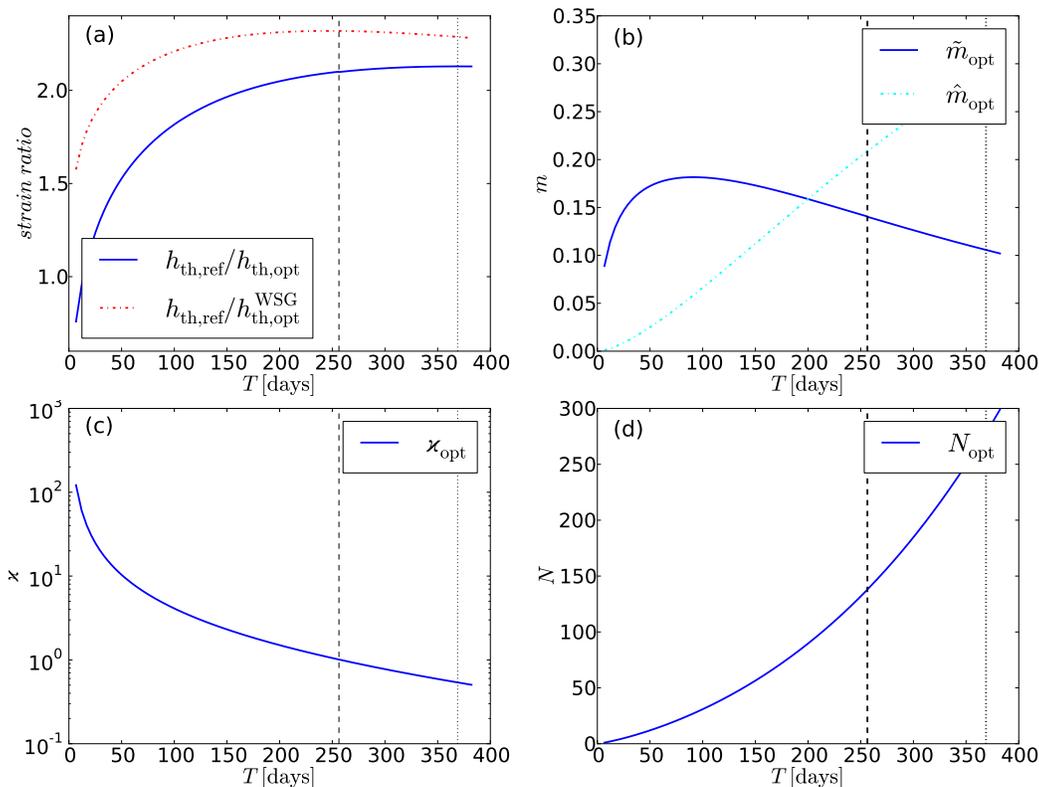
FIG. 3: Numerical optimal fixed-T solution for a directed CasA search as a function of T. The dashed vertical line indicates the analytical WSG-optimal solution of Eq. (118), while the dotted vertical line corresponds to the exact optimal solution. Panel (a) shows the weakest detectable signal strength compared to the reference value $h_{\rm th,ref}$ of Eq. (114), for the exact $h_{\rm th}$ and for the WSG-approximated $h_{\rm th}^{\rm WSG}$ (using $w = 1$). (b) shows the optimal mismatch parameters $\widetilde{m}_{\rm opt}(T)$ and $\widehat{m}_{\rm opt}(T)$, (c) shows the optimal computing-cost ratio $\varkappa_{\rm opt}(T)$ and (d) the optimal number of segments $N_{\rm opt}(T)$ (treated as continuous).

searches labelled 'S5GC1' and 'S6Bucket', which employed an efficient grid mapping implementation described in [27].

An Einstein@Home search divides the total workload into many small *workunits*, each of which covers a small fraction of the parameter space and requires only a couple of hours to complete on a host machine. These searches consisted of roughly 10^7 workunits. The E@H platform delivers a computing power of order 100 Tflop/s, and these searches ran for about 6 months each, so we can estimate their total respective computing cost is of order $C_{\rm tot} \sim 10^{21}$ flop (i.e. ~ 1 Zeta flop). Each E@H workunit is designed to require about the same computing cost, which allows us to base the present analysis on just a single workunit.

The detector data used in these searches contained non-stationarities and gaps, and the template banks were constructed in somewhat semi-empirical ways that are hard to model analytically. In order to simplify this analysis we assume gapless stationary Gaussian data, and we use the analytic metric expressions from [26] to estimate the number of templates. This example is therefore "inspired by" recent E@H searches, but does not represent a detailed description of their computing cost or sensitivity.

The two searches 'S5GC1' and 'S6Bucket' covered a

fixed spindown-range corresponding to a spindown age of $\tau = f_0/\dot{f} = 600$ y at a reference frequency of $f_0 = 50$ Hz. Each workunit covers a frequency-band of $\Delta f = 0.05\,Hz$, the spindown range of $\Delta\dot{f} = 2.7 \times 10^{-9}$ Hz/s and a (frequency-dependent) fraction q of the sky. We can incorporate the sky-fraction q by using template counts $q\mathcal{N}$ in the computing-cost expressions, where $\mathcal{N}$ are the all-sky expressions from [26]. For simplicity we fix the parameter-space dimension to $n = 4$, namely {sky, frequency, spindown}, and we use [Eq. (56),(50) in [26]][8] for the number of coarse-grid templates $\widetilde{\mathcal{N}}$ and the refinement factor $\gamma(N)$ of [Eq. (77) in [26]] (assuming gapless data).

For a workunit at 50 Hz, the reference StackSlide parameters are:

- S5GC1: $q = 1/3$, $N = 205$, $\Delta T = 25$ h

- S6Bucket: $q = 1/51$, $N = 90$, $\Delta T = 60$ h

For both searches the mismatch distributions of the

---

[8] There are missing terms in both [Eqs. (57) and (83) in [26]], but one can use their Eqs. (50) and (84) instead to compute $\det g$.

	$\widetilde{\delta}$	$\widehat{\delta}$	$\widehat{\eta}$	w	N	ΔT[d]	$\widetilde{m}$	$\widehat{m}$	C_0[h]	$\varkappa$	$\frac{h_{\mathrm{th}}}{\sqrt{S_{\mathrm{n}}}}[\sqrt{\mathrm{Hz}}]$
S5GC1	8.7	7.7	2.0	1.1	205	1.0	0.50	0.50	0.91	2.545	2.69×10^{-3}
$T_{\max} = 1\,\mathrm{y}$	10.0	9.0	2.0	1.1	528	0.7	0.14	0.17	0.91	0.869	2.19×10^{-3}
S6Bucket	4.6	3.6	2.0	1.2	90	2.5	0.50	0.50	2.54	13.914	2.20×10^{-3}
$T_{\max} = 1\,\mathrm{y}$	3.7	2.7	2.0	1.2	175	2.1	0.58	0.32	2.54	1.815	1.93×10^{-3}

TABLE I: Einstein@Home example setups 'S5GC1' and 'S6Bucket', with corresponding results from an iterative optimization at fixed computing power C_0, with assumed maximal observation time of $T_{\max} = 1\,\mathrm{y}$. The gains in weakest detectable signal strength h_{th} are $\sim 23\%$ and $\sim 14\%$, respectively.

coarse- and fine-grid template banks are not well quantified, so we simply assume hyper-cubic template banks ($\xi = 1/3$) with $\widetilde{m} = \widehat{m} = 0.5$, i.e. an average total mismatch of $\langle \mu \rangle = 1/3$. Plugging these parameters into the template-counting formulae of [26], together with the timing constants of Eq. (115) from the Cas-A example, we find a reference per-workunit computing cost of $C_0 \approx 0.91\,\mathrm{h}$ for S5GC1, and $C_0 \approx 2.5\,\mathrm{h}$ for S6Bucket[9]. Table I shows the estimated sensitivity for these reference searches assuming the same false-alarm and false-dismissal probabilities as in the previous section.

We can apply the analytical optimal solution from Sec. IV with the extracted power-law coefficients at the reference StackSlide parameters found in Table I. This initially places us into the *unbounded* regime (i.e. $\widehat{a} > 0$) for both 'S5GC1' and 'S6Bucket'. We therefore expect to improve sensitivity by increasing T until we hit the assumed upper bound of $T_{\max} = 1\,\mathrm{y}$, so we solve Eq. (87) for $\varkappa_{\mathrm{opt}}(T_{\max})$, substitute into Eq. (94) for $\{\widetilde{m}_{\mathrm{opt}}, \widehat{m}_{\mathrm{opt}}\}$ and obtain N_{opt} from Eq. (86).

In order to find a *self-consistent* solution, we need to iterate this procedure: we extract new power-law coefficients at the new solution, then re-solve until the parameters converge to better than 1% accuracy. In the case of the 'S5GC1' search, the converged solution falls into the unbounded regime. In the case of 'S6Bucket' the converged solution falls into the bounded regime, but with $T_{\mathrm{opt}}^{(0)} > T_{\max}$. The optimal observation time is therefore $T_{\mathrm{opt}} = 1\,\mathrm{y}$ in both cases, and the resulting converged solutions and power-law coefficients are given in Table I. We see that (under the present idealized conditions) we could gain $\sim 23\%$ in detectable signal strength h_{th} compared to the 'S5GC1' setup, and $\sim 14\%$ compared to the 'S6Bucket' setup.

C. All-sky search examples from CGK

The all-sky search examples studied in CGK [21] provide another interesting test case for our optimization

framework. CGK considered a multi-stage optimization, but we can treat their first-stage result as a single-stage optimization problem at fixed given computing cost. CGK discussed four different cases, namely a search for either "young" (Y) neutron stars ($\tau = f/\dot{f} = 40\,\mathrm{y}$) or "old" (O) neutron stars ($\tau = 10^6\,\mathrm{y}$), using either a "fresh-data" (f) or "data-recycling" (r) mode (a distinction that is irrelevant for our present purpose). The optimized CGK StackSlide parameters and computing-cost constraints are found in [Tables I-VIII in [21]], and are summarized in Table II. For the sensitivity estimates we use the same false-alarm and false-dismissal probabilities as in Sec. V A.

Note that we expect our results to improve on the sensitivity of the CGK solution, as they incorporated an *ad-hoc* constraint of $m = \widetilde{m} = \widehat{m}$, and the total average mismatch in [Eq.(46) in CGK] incorrectly included only the contribution from one template grid instead of both, as discussed in Sec. III B 3.

The functional form of the template-bank equations (originally from BC [17]) in the CGK computing-cost model [Eq.(53) in CGK] is not consistent with the generic form of Eq. (43) with respect to the mismatch scaling. We therefore resort to extracting (potentially fractional) "mismatch dimensions" $\{\widetilde{n}, \widehat{n}\}$ using Eq. (64), in order to fully reproduce their computing-cost function with the power-law model of Eq. (61). The scaling parameters $\{\delta, \eta\}$ are extracted via Eq. (62) and w from Eq. (41). The resulting values are given in Table II, assuming the FFT/resampling method for the $\mathcal{F}$-statistic calculations.

Using the extracted scaling coefficients to compute the optimal solution from Sec. IV results in a solution that is inconsistent with the initially extracted scaling coefficients. An iteration over solutions, allowing both N and T to vary, did not converge. We therefore solve a simpler problem by fixing the number of segments to the original CGK values, i.e. we constrain the solutions to $N = N_{\mathrm{CGK}}$. We proceed by solving Eq. (86) for $\varkappa_{\mathrm{opt}}(N)$, closing the solution via Eqs. (94) and (87). We then extract new power-law coefficients at this solution and iterate this procedure until convergence to better than 1% accuracy is achieved. The resulting fixed-N optimal solutions are given in Table II. The respective improvements of the weakest detectable signal strength h_{th} compared to the original CGK solutions are $\sim 15\%$ in the young (Y) pulsar case, and $\sim 5\%$ in the old (O) pulsar case.

---

[9] The actual E@H workunits take about 6 h to complete on a machine with these timings, but these setups included bigger refinement factors γ due to gaps in the data, and used rather different template-bank designs.

	$\widetilde{n}$	$\widehat{n}$	$\widetilde{\delta}$	$\widehat{\delta}$	$\widehat{\eta}$	w	N	$\Delta T[\mathrm{d}]$	$\widetilde{m}$	$\widehat{m}$	$C_0[\mathrm{Zf}]$	$\varkappa$	$\frac{h_{\mathrm{th}}}{\sqrt{S_{\mathrm{n}}}}[\sqrt{\mathrm{Hz}}]$
Y/r	3.0	4.0	3.1	6.0	4.2	1.7	10	2.6	0.78	0.78	0.94	0.071	7.31×10^{-3}
$N = N_{\mathrm{CGK}}$	3.0	4.0	3.1	6.0	4.2	1.7	10	2.1	0.17	0.48	0.94	0.482	6.38×10^{-3}
Y/f	3.0	4.0	3.1	6.0	4.2	1.7	9	2.7	0.78	0.78	0.82	0.086	7.42×10^{-3}
$N = N_{\mathrm{CGK}}$	3.0	4.0	3.1	6.0	4.2	1.7	9	2.2	0.18	0.46	0.82	0.526	6.50×10^{-3}
O/r	2.5	2.5	3.0	10.0	7.4	1.8	8	14.8	0.35	0.35	0.74	0.004	2.61×10^{-3}
$N = N_{\mathrm{CGK}}$	2.8	2.5	3.0	10.0	7.4	1.8	8	14.6	0.03	0.34	0.74	0.090	2.46×10^{-3}
O/f	2.5	2.6	3.0	10.0	7.3	1.7	9	11.8	0.21	0.21	0.35	0.009	2.65×10^{-3}
$N = N_{\mathrm{CGK}}$	2.7	2.5	3.0	10.0	7.3	1.7	9	12.4	0.04	0.33	0.35	0.107	2.56×10^{-3}

TABLE II: CGK example search setups for young ('Y') and old pulsars ('O'), using either fresh ('f') or recycling ('r') data-modes. The first line of each example gives the original CGK solution with respective extracted power-law coefficients, and the second line shows our optimal self-consistent solution with constraint $N = N_{\mathrm{CGK}}$. The computing cost C_0 is measured in Zeta-flop ($1\mathrm{Zf} = 10^{21}\mathrm{flop}$).

D. CWs from binary neutron stars

For CWs from NSs in binary systems with known sky-position (such as Sco-X1 and other LMXBs), the search parameter space typically consists of the intrinsic signal frequency and orbital parameters of the binary system, i.e. (projected) semi-major axis, orbital period P, periapse angle, eccentricity and eccentric anomaly. The corresponding template-counting formulae were initially studied in [41] for coherent searches. These have recently been extended to semi-coherent searches by Messenger [32], giving explicit template scalings in two limiting cases, namely (i) *short* coherent segments compared to the orbital period, i.e. $\Delta T \ll P$, and (ii) *long* coherent segments, i.e. $\Delta T \gg P$.

(i) Short coherent segments ($\Delta T \ll P$)

One can change parameter-space coordinates and Taylor-expand in small $\Delta T/P \ll 1$ to obtain the coherent template scaling [Eq. (24) in [32]]:

$$\widetilde{\mathcal{N}}_{\widetilde{n}} \propto \Delta T^{\widetilde{n}(\widetilde{n}+1)/2}, \tag{121}$$

where $\widetilde{n}$ is the effective coherent parameter-space dimension using the new coordinates. The coherent cost power-law coefficients are therefore $\widetilde{\delta} = \widetilde{n}(\widetilde{n}+1)/2 + \Delta\widetilde{\delta}$ and $\widetilde{\eta} = 1$.

The semi-coherent template scaling including eccentricity results in a 6-dimensional semi-coherent template bank, i.e. $\widehat{n} = 6$, and a template scaling [Eq. (28) in [32]] of $\widehat{\mathcal{N}}_{\mathrm{ecc}} \propto N\,\Delta T^7$. In the case of small eccentricity one has $\widehat{n} = 4$, and the template scaling given in [Eq. (29) in [32]] is $\widehat{\mathcal{N}}_{\mathrm{circ}} \propto N\,\Delta T^5$. In both cases the semi-coherent power-law exponents satisfy $\widehat{\delta} \geq 5$, and $\widehat{\eta} = 2$, resulting in the critical parameter $\widehat{a} > 0$. This implies that the boundedness-condition Eq. (102) is always violated, i.e. one should use all the available data.

(ii) Long coherent segments ($\Delta T \gg P$)

In this limit the template scalings in both the coherent and semi-coherent step are [Eqs. (32,33) in [32]]: $\widetilde{\mathcal{N}} \propto \widehat{\mathcal{N}} \propto \Delta T^2$, which is unusual as there is *no refinement*. Therefore $\widetilde{\eta} = \widehat{\eta} = 1$, and $\widetilde{\delta} = 2 + \Delta\widetilde{\delta}$, while $\widehat{\delta} = 2$, and therefore $\widehat{\varepsilon} = 1$. We see that always $\widetilde{a} > 0$ and $\widehat{a} = 2(w - 1) > 0$, and therefore binary-CW searches in the long-segment limit also fall into the *unbounded* regime, i.e. one should use all the data.

VI. DISCUSSION

We have derived an improved estimate of the Stack-Slide sensitivity scaling, correctly accounting for the mismatches from both coarse- and fine-grid template banks, which had been overlooked by previous studies. By locally fitting sensitivity and computing-cost functions to power laws we are able to derive fully analytical self-consistency relations for the optimal sensitivity at fixed computing cost. This solution separates two different regimes depending on the critical parameter $\widehat{a}$ of Eq. (102): a bounded regime with a finite optimal $T_{\mathrm{opt}}^{(0)}$, and an unbounded regime where $T_{\mathrm{opt}}^{(0)} \to \infty$.

Several practical examples are discussed in order to illustrate the application of this framework. The corresponding sensitivity gains in terms of the weakest detectable signal strength h_{th} are found to be $\sim 100\%$ compared to a fully coherent directed search for CasA, and about $5\% - 20\%$ compared to previous StackSlide searches such as Einstein@Home and the examples given in CGK [21]. We show that CW searches for binary neutron stars seem to generally fall into the unbounded regime where all the available data should be used irrespective of available computing power.

This study only considered single-stage StackSlide searches on Gaussian stationary gapless data from detectors with identical noise-floors. Further work is required to extend this analysis to more realistic data conditions.

VII.   ACKNOWLEDGMENTS

[1] Abbott B P *et al.* (LIGO Scientific Collaboration) 2009 *Reports on Progress in Physics* **72** 076901
[2] Accadia T *et al.* 2011 *Class. Quant. Grav.* **28** 114002
[3] Grote H (for the LIGO Scientific Collaboration) 2010 *Class. Quant. Grav.* **27** 084003
[4] Abadie J *et al.* (LIGO Scientific Collaboration and Virgo Collaboration) 2011 (*Preprint* arXiv:1110.0208)
[5] Abbott B *et al.* (LIGO Scientific Collaboration and Virgo Collaboration) 2010 *ApJ* **713** 671
[6] Abbott B *et al.* (LIGO Scientific Collaboration) 2009 *Phys. Rev. D.* **80** 042003
[7] Abadie J *et al.* (LIGO Scientific Collaboration) 2010 *ApJ* **722** 1504
[8] Harry G M (for the LIGO Scientific Collaboration) 2010 *Class. Quant. Grav.* **27** 084006
[9] The Virgo Collaboration 2009 (*Preprint* VIR–027A–09)
[10] Willke B *et al.* 2006 *Class. Quant. Grav.* **23** 207
[11] Punturo M *et al.* 2010 *Class. Quant. Grav.* **27** 084007
[12] Prix R (for the LIGO Scientific Collaboration) 2009 *Neutron Stars and Pulsars* (*Astrophysics and Space Science Library* vol 357) ed Becker W (Springer Berlin Heidelberg) p 651 (*Preprint* LIGO-P060039-v3)
[13] Knispel B and Allen B 2008 *Phys. Rev. D.* **78** 044031
[14] Brady P R, Creighton T, Cutler C and Schutz B F 1998 *Phys. Rev. D.* **57** 2101
[15] Wette K *et al.* 2008 *Class. Quant. Grav.* **25** 235011
[16] Abbott B *et al.* (LIGO Scientific Collaboration) 2007 *Phys. Rev. D.* **76** 082001
[17] Brady P R and Creighton T 2000 *Phys. Rev. D.* **61** 082001
[18] Pletsch H J 2011 *Phys. Rev. D.* **83** 122003
[19] Krishnan B *et al.* 2004 *Phys. Rev. D.* **70** 082001
[20] Dhurandhar S, Krishnan B, Mukhopadhyay H and Whelan J T 2008 *Phys. Rev. D.* **77** 082001
[21] Cutler C, Gholami I and Krishnan B 2005 *Phys. Rev. D.* **72** 042004
[22] Prix R 2008 *The Eleventh Marcel Grossmann Meeting On Recent Developments in Theoretical and Experimental General Relativity, Gravitation and Relativistic Field Theories* ed H Kleinert, R T Jantzen, & R Ruffini p 2441 (*Preprint* arXiv:gr-qc/0702068)
[23] Knispel B 2011 *Pulsar Discoveries by Volunteer Distributed Computing and the Strongest Continuous Gravitational Wave Signal* Ph.D. thesis Leibniz Universität Hannover
[24] Jaranowski P, Królak A and Schutz B F 1998 *Phys. Rev. D.* **58** 063001
[25] Cutler C and Schutz B F 2005 *Phys. Rev. D.* **72** 063006
[26] Pletsch H J 2010 *Phys. Rev. D.* **82** 042002
[27] Pletsch H J and Allen B 2009 *Phys. Rev. Lett.* **103** 181102
[28] Prix R 2010 The $\mathcal{F}$-statistic and its implementation in ComputeFStatistic_v2 Tech. rep. (*Preprint* LIGO-T0900149-v2)
[29] Balasubramanian R, Sathyaprakash B S and Dhurandhar S V 1996 *Phys. Rev. D.* **53** 3033
[30] Owen B J 1996 *Phys. Rev. D.* **53** 6749
[31] Prix R 2007 *Phys. Rev. D.* **75** 023004
[32] Messenger C 2011 (*Preprint* arXiv:1109.0501)
[33] Messenger C, Prix R and Papa M A 2009 *Phys. Rev. D.* **79** 104017
[34] Wette K 2011 (*Preprint* arXiv:1111.5650)
[35] Woan G (private communication)
[36] Prix R 2007 *Class. Quant. Grav.* **24** S481
[37] Williams P R and Schutz B F 1999 (*Preprint* arXiv:gr-qc/9912029)
[38] Patel P, Siemens X, Dupuis R and Betzwieser J 2010 *Phys. Rev. D.* **81** 084032
[39] Einstein@home project URL `http://einstein.phys.uwm.edu`
[40] Abbott B *et al.* (LIGO Scientific Collaboration) 2009 *Phys. Rev. D.* **79** 022001
[41] Dhurandhar S V and Vecchio A 2001 *Phys. Rev. D.* **63** 122001
[42] Watts A, Krishnan B, Bildsten L and Schutz B F 2008 *Mon. Not. Roy. Astron. Soc.* **389** 839
[43] Jaranowski P, Krolak A and Schutz B F 1998 *Phys. Rev. D.* **58** 063001
[44] Jaranowski P and Krolak A 2000 *Phys. Rev. D.* **61** 062001
[45] Abbott B *et al.* (LIGO Scientific Collaboration) 2004 *Phys. Rev. D.* **69** 082004
[46] Abbott B *et al.* (LIGO Scientific Collaboration) 2008 *Phys. Rev. D.* **77** 022001
[47] Conway J H and Sloane N J A 1999 *Sphere packings, lattices and groups* A Series of Comprehensive Studies in Mathematics (Springer)