Methods for Detecting Gravitational Waves from mini-Extreme-Mass-Ratio Inspirals II: A Spectral-Leakage-Aware Framework

Zi-Xuan Wang[1,2],* Xing-Yu Chen[3],† Gong Cheng[1,2],‡ Huai-Ke Guo[1,2],§ Ju Chen[1,2],¶ and Andrew L. Miller[1,2,4,5]**
[1] *International Center for Theoretical Physics Asia-Pacific (ICTP-AP), University of Chinese Academy of Sciences, Beijing 100190, China*
[2] *Taiji Laboratory for Gravitational Wave Universe, University of Chinese Academy of Sciences, 100049 Beijing, China*
[3] *Department of Physics, University of Chinese Academy of Sciences, Beijing 100190, China*
[4] *Nikhef – National Institute for Subatomic Physics, Science Park 105, 1098 XG Amsterdam, The Netherlands and*
[5] *Institute for Gravitational and Subatomic Physics (GRASP), Utrecht University, Princetonplein 1, 3584 CC Utrecht, The Netherlands*
(Dated: January 8, 2026)

Mini-Extreme-Mass-Ratio Inspirals (mini-EMRIs), comprising a sub-solar exotic compact object (such as a primordial black hole or boson star) orbiting a much heavier stellar-origin or exotic compact object, represent key targets for ground-based gravitational-wave detectors to probe the early universe and the nature of dark matter. However, detecting such systems, which could spend hours to years in LIGO, Virgo and KAGRA data, poses a computational challenge to standard matched-filtering methods, which has motivated the development of methods that divide the data into coherent chunks, Fourier transform them, and combine their spectral power or related statistics incoherently. However, these pipelines typically require that the signal be quasi-monochromatic within each chunk, which imposes a fundamental limit on the coherence time, since spectral leakage becomes significant when the frequency evolution within the chunk becomes comparable to the frequency resolution. In this work, we extend the development of our method, `ΣTrack` [1], to the regime in which the quasi-monochromatic approximation is relaxed, in two ways. First, we establish an analytical model for the spectral leakage, extending the validity of conventional analyses beyond the quasi-monochromatic regime. Second, we propose the ΣR statistic—a novel detection metric formed by a weighted summation of power ratios—which effectively recovers the signal energy dispersed across adjacent frequency bins. Building on this framework, we further introduce an innovative frequency-layered search strategy that dynamically optimizes the coherence time across the observation band. We benchmark our method against a globally optimized Hough transform pipeline using a fiducial mini-EMRI signal from a binary with masses $(1.5, 10^{-5})\,M_\odot$. The results demonstrate that our framework achieves an order-of-magnitude enhancement in the effective detection volume, significantly expanding the horizon for discovering mini-EMRIs and sub-solar exotic compact objects with ground-based gravitational wave detectors, which can be similarly applied to EMRI searches for future space-based gravitational wave detectors.

I. INTRODUCTION

The advent of gravitational-wave (GW) astronomy, heralded by the first direct detection of a binary black hole merger [2] and enabled by the global network of interferometric detectors—Advanced LIGO [3], Advanced Virgo [4], and KAGRA [5]—has provided a pristine probe into the universe's most energetic phenomena. The rapidly expanding catalog of transient events now encompasses hundreds of compact binary coalescences (CBCs) [6–9]. While these observations have profoundly advanced our understanding of compact object formation, they have simultaneously exposed significant tensions with established astrophysical models. The observed population exhibits properties challenging to reconcile with canonical stellar evolution theories [10], ranging from detections within the presumed high-mass gap [11] to the discovery of enigmatic sources populating the "lower mass gap", whose existence obscures the conventional distinction between neutron stars and black holes [12–15]. These observational puzzles imply the existence of alternative formation pathways, beyond astrophysical processes [16–19]. Moreover, a smoking gun evidence for new physics is the detection of a sub-solar compact object, which has received increasingly rising attention in the community, resulting in many searches for such objects [20–32]. Popular candidates for the sub-solar exotic compact objects are primordial black holes [33–37], boson stars [38], etc, the detection of which can also shed new light in the understanding of the nature of dark matter.

For binaries consisting of increasingly light sub-solar exotic compact objects, the gravitational wave signal decreases significantly as the chirp mass becomes smaller. An optimal strategy is to pair a sub-solar exotic compact object up with a much heavier one (of stellar or compact

* wangzixuan243@mails.ucas.ac.cn
† chenxingyu223@mails.ucas.ac.cn
‡ chenggong@ucas.ac.cn
§ guohuaike@ucas.ac.cn
¶ chenju@ucas.ac.cn
** andrew.miller.ligo@ucas.ac.cn

origin) to increase the signal strength, leading to a sytem with an extreme mass ratio $q = m_{\rm lighter}/m_{\rm primary} \lesssim 10^{-5}$, thus called mini extreme mass ratio inspirals (mini-EMRIs) [39]. While comparable-mass binaries ($q \sim 1$) are now routine targets for ground-based detectors [40, 41] and classical EMRIs involving supermassive black holes are key objectives for future space-based missions such as LISA [42–45], Taiji [46–48] and TianQin [49–51], mini-EMRIs serve as a critical phenomenological bridge. They offer the unprecedented opportunity to study extreme mass-ratio dynamics within the sensitive frequency band of terrestrial interferometers, ultimately opening a novel window into fundamental physics.

The signal phenomenology of a mini-EMRI diverges fundamentally from that of canonical compact binary coalescences (CBCs). In contrast to the fleeting "chirps" characteristic of stellar-mass mergers, a mini-EMRI trajectory may linger within the detector's most sensitive frequency band for timescales spanning days to years. This "long-transient" character places mini-EMRIs in a hybrid signal category: they exhibit the deterministic frequency evolution typical of inspiraling binaries, yet possess the temporal persistence of continuous wave (CW) sources. This duality imposes severe constraints on standard data analysis strategies. On one hand, the computational cost of fully coherent matched filtering becomes prohibitive due to the rapid expansion of the template bank parameter space over such extended durations [52–56]. On the other hand, conventional CW search pipelines—optimized for quasi-monochromatic sources like spinning neutron stars—lack the dynamic range to accommodate the significant frequency drift inherent to mini-EMRI dynamics [57–64]. Consequently, the detection of these sources necessitates the development of specialized search algorithms designed to efficiently track rapid phase evolution over long observational baselines.

Analyses of such non-stationary signals typically rely on the Short-Time Fourier Transform (STFT) as a fundamental preprocessing step to track time-frequency evolution. However, a fundamental limitation arises when the signal frequency evolves significantly within a single analysis window, inducing spectral leakage that degrades detection efficiency. Standard STFT-based algorithms, such as the Hough transform [57, 58], operate under a "quasi-monochromatic" approximation, assuming the signal remains effectively monochromatic within a segment of duration T. This constraint is succinctly quantified by the dimensionless widening factor $w = |\dot{f}|T^2$, which is empirically required to be $w \lesssim 0.5$—implying that the signal's frequency drift over the segment duration T is confined to less than half of the frequency resolution $f_{\rm bin} = 1/T$. For a long-transient GW signal, this condition constrains the choice of the coherent time. While a longer T is desirable for enhancing the signal-to-noise ratio (SNR) through extended coherent integration, analysts are compelled to use a short T to satisfy the approximation. This creates a fundamental trade-off: maximizing detection sensitivity is sacrificed for model fidelity, thus curtailing the reach for more distant or fainter sources.

This work establishes a theoretical framework to rigorously model spectral leakage, thereby enabling the unconstrained selection of the coherence time T, independent of the traditional quasi-monochromatic restriction. Leveraged by the massive parallelism of modern Graphics Processing Unit (GPU) architectures [65, 66], which allows computationally intensive approaches to rival the efficiency of the Hough transform, we build upon the foundations of the PowerFlux method [67] to introduce a novel detection statistic, denoted as ΣR. We fully characterize the statistical properties of ΣR, providing a theoretical guide for the optimization of search parameters that effectively resolves the trade-off between coherence time and spectral leakage loss. Ultimately, our method achieves an order-of-magnitude improvement in detection volume compared to the conventional Hough transform pipeline.

The structure of this paper is as follows. Section II delineates the mini-EMRI signal morphology and the corresponding interferometric detector response. Confronting the limitations of standard time-frequency analysis, Section III develops a rigorous analytical framework within the STFT domain to precisely model the phenomenon of spectral leakage. Building on this foundation, Section IV formally introduces the ΣR statistic, deriving its statistical properties and establishing analytical metrics for both the maximum detectable distance and the effective parameter space volume. Finally, in Section V, we validate the performance of our algorithm through comprehensive simulations and present a comparative benchmark against established pipelines.

II. SIGNAL MODEL

As described in [68, 69], the response $s(t)$ of an interferometric gravitational-wave detector to an incoming signal is a linear combination of the two gravitational-wave polarizations, $h_+(t)$ and $h_\times(t)$:

$$s(t) = F_+(t)\, h_+(t) + F_\times(t)\, h_\times(t), \tag{1}$$

where $F_+(t)$ and $F_\times(t)$ are the detector antenna-pattern functions (they depend on the detector orientation, source orientation and the source sky location).

The polarization waveforms for a quasi-sinusoidal signal can be written as

$$\begin{aligned} h_+(t) &= h_0(t)\,\frac{1+\cos^2\iota}{2}\,\cos\Phi(t), \\ h_\times(t) &= h_0(t)\,\cos\iota\,\sin\Phi(t), \end{aligned} \tag{2}$$

where $h_0(t)$ denotes the intrinsic strain amplitude, $\Phi(t)$ the phase, and ι the inclination angle between the orbital angular momentum and the line of sight. The intrinsic amplitude $h_0(t)$ is in general a slowly varying function of time through its dependence on frequency $f(t)$.

For clarity of the subsequent analysis, it is convenient to combine the two polarization contributions into a single sinusoid:

$$s(t) = h_0(t)\, Q(t)\, \cos\left[\Phi(t) + \phi_p(t)\right], \tag{3}$$

where the dimensionless modulation factor $Q(t) \geq 0$ collects the time-dependent projection of the polarizations onto the detector:

$$Q(t) \equiv \sqrt{F_+^2(t)\left(\frac{1+\cos^2\iota}{2}\right)^2 + F_\times^2(t)\cos^2\iota}. \tag{4}$$

The detector-frame mixing of polarizations also produces a (slowly varying) polarization phase

$$\phi_p(t) \equiv -\arctan\left(\frac{2\cos\iota\, F_\times(t)}{(1+\cos^2\iota)\, F_+(t)}\right), \tag{5}$$

chosen so that Eq. (3) reproduces Eq. (1).

In this work we model a mini-EMRI as a compact object in a circular, equatorial orbit about a (possibly spinning) central body. Under this assumption the spin-up $\dot{f}$ and the instantaneous intrinsic amplitude h_0 can be written as power-law forms modified by relativistic correction factors [39, 70]:

$$\begin{aligned} \dot{f} &= \frac{96}{5}\pi^{8/3}\left(\frac{G\mathcal{M}_c}{c^3}\right)^{5/3} f^{11/3}\, C_f(a,f), \\ h_0(f) &= \frac{4}{d}\left(\frac{G\mathcal{M}_c}{c^2}\right)^{5/3}\left(\frac{\pi f}{c}\right)^{2/3} C_h(a,f), \end{aligned} \tag{6}$$

where $\mathcal{M}_c = (m_1 m_2)^{3/5}(m_1+m_2)^{-1/5}$ is the chirp mass, d the luminosity distance, and a the dimensionless spin of the central body.

The correction factors $C_f(a,f)$ and $C_h(a,f)$ encode relativistic (strong-field) effects introduced by the central body's spin and by higher-order post-Newtonian physics. In the weak-field, large-separation limits $C_f \to 1$ and $C_h \to 1$, so Eq. (6) reduce to the standard Newtonian power-law expressions. As the orbit approaches the innermost stable circular orbit (ISCO), these correction factors depart significantly from unity and capture the rapid relativistic evolution of both phase and amplitude.

Finally, to obtain the phase as measured in the detector's reference frame, we must account for the modulation induced by the detector's motion. The dominant effect is the Doppler shift, which modifies the signal's instantaneous frequency. Integrating this observed frequency yields the detector-frame phase $\Phi(t)$:

$$\Phi(t) = 2\pi \int^t f(t')\left(1 + \frac{\vec{v}(t')\cdot\hat{n}}{c}\right) dt' + \Phi_0, \tag{7}$$

with Φ_0 a reference phase and $\vec{v}(t')\cdot\hat{n}$ represents the component of the detector's velocity along the line of sight to the source. The combination of this phase model (Eq. (7)) with the amplitude evolution described by Eqs. (3) and (6) provides the complete time-domain signal model used to construct the time-frequency templates in our semi-coherent search.

III. SPECTRAL ANALYSIS

A. Short-time Fourier Transform

The strain data recorded by the interferometer is denoted as $x(t)$ and is modeled as the linear superposition of a deterministic gravitational-wave signal $s(t)$ and stochastic instrumental noise $n(t)$:

$$x(t) = s(t) + n(t). \tag{8}$$

Throughout this article, we approximate the noise component $n(t)$ as a stationary, Gaussian random process.

To analyze the time-frequency evolution of the signal, the discrete data stream is partitioned into overlapping segments of duration T_{DFT}, each comprising N samples. A window function, $W[n]$, is applied to each segment prior to the computation of the Discrete Fourier Transform (DFT) to mitigate spectral leakage at the boundaries. We enforce a power-normalization on the window function, a detailed derivation and discussion of this convention are provided in Appendix C. For the i-th time segment, the complex-valued STFT coefficient at the k-th frequency bin is given by:

$$\tilde{x}_i[k] = \frac{1}{N}\sum_{n=0}^{N-1} x_i[n] W[n] e^{-i2\pi nk/N}, \tag{9}$$

where $x_i[n]$ denotes the discrete time samples within the i-th segment. This procedure is standard in CW analyses, e.g. the construction of short FFTs and short fast Fourier transform databases [60].

To facilitate statistical detection, we define a normalized *power-ratio spectrogram*, $R[i,k]$, by rescaling the measured power in each time-frequency bin by the expected noise power:

$$R[i,k] \equiv \frac{|\tilde{x}_i[k]|^2}{\langle|\tilde{n}_i[k]|^2\rangle}. \tag{10}$$

Here, the denominator $\langle|\tilde{n}_i[k]|^2\rangle$ represents the estimated value of the power spectrum of the noise background in the (i,k)-th bin. In practice, this baseline is estimated from the data itself using robust averaging techniques, such as the auto-regressive average spectrum method, to exclude transient outliers [71].

Under the assumption of stationary Gaussian noise, the scaled statistic $2R[i,k]$ follows a non-central chi-squared distribution with two degrees of freedom:

$$2R[i,k] \sim \chi^2\left(2, \lambda[i,k]\right). \tag{11}$$

The non-centrality parameter $\lambda[i,k]$ corresponds to the power SNR for that specific bin, defined as:

$$\lambda[i,k] \equiv \frac{2|\tilde{s}_i[k]|^2}{\langle|\tilde{n}_i[k]|^2\rangle}. \tag{12}$$

Note that the factor of 2 in Eq. (11) and Eq. (12) arises from the properties of the complex Gaussian distribution governing the spectrum.

B. Analytical framework

Adopting the factorization strategy introduced in our previous work for `ΣTrack` [1], we decompose the noncentrality parameter $\lambda[i,k]$ into two physically distinct components: a total power term and a normalized spectral shape term:

$$\lambda[i,k] = \frac{2P_i}{\langle|\tilde{n}_i[k]|^2\rangle} \cdot \frac{|\tilde{s}_i[k]|^2}{P_i} \equiv \mathcal{L}_i \cdot \eta_i[k]. \tag{13}$$

The first factor, defined as the *total power SNR statistic* $\mathcal{L}_i \equiv 2P_i/\langle|\tilde{n}_i[k]|^2\rangle$, quantifies the signal's average power within the i-th segment relative to the noise power in a single frequency bin. Here, P_i represents the average power of the signal component after windowing, and a detailed discussion is in Appendix C.

The second factor, $\eta_i[k] \equiv |\tilde{s}_i[k]|^2/P_i$, is designated as the *normalized power spectrum distribution factor*—or equivalently, the *spectral leakage factor*. This normalized discrete function characterizes how the total signal power P_i is partitioned among the discrete frequency bins $\{k\}$. Crucially, $\eta_i[k]$ is intrinsic to the signal's frequency evolution and the windowing function, remaining invariant under changes in the noise background.

To systematically analyze the leakage pattern, it is advantageous to transition from the absolute frequency index k to a signal-centric relative coordinate system. We define the instantaneous central frequency of the signal in dimensionless units as $k_{c,i} \equiv f_{c,i}T_{\mathrm{DFT}}$. This continuous frequency coordinate is decomposed into an integer component and a fractional offset:

- The *anchor bin*, denoted by $\lfloor k_{c,i} \rceil$, represents the integer DFT bin index closest to the true signal frequency.
- The *fractional offset*, $o_{0,i} \equiv \lfloor k_{c,i} \rceil - k_{c,i}$, quantifies the displacement of the anchor bin center relative to the true frequency, bounded within the interval $o_{0,i} \in (-0.5, 0.5)$.

The relative frequency index κ is then defined as the integer distance from this segment-specific anchor:

$$\kappa \equiv k - \lfloor k_{c,i} \rceil. \tag{14}$$

In this moving frame, $\kappa = 0$ consistently refers to the peak (anchor) bin, while $\kappa = \pm 1$ denote the immediate sidebands. The effective offset for any relative bin κ is given by $o_{\kappa,i} = \kappa + o_{0,i}$.

Consequently, the discrete leakage factors $\{\eta_i[\kappa]\}$ can be interpreted as equidistant samples drawn from an underlying continuous spectral leakage function, $\eta(o)$. The relationship is expressed as:

$$\eta_i[\kappa] = \eta(o_{\kappa,i}) = \eta(\kappa + o_{0,i}). \tag{15}$$

Here, the initial offset $o_{0,i}$ acts as a "sampling phase", dictating the discrete sampling grid along the continuous function. The temporal variation of the leakage pattern is thus entirely governed by the evolution of the sampling phase $\{o_{0,i}\}$, which is driven by the signal's intrinsic chirp.

This decomposition is visualized in Fig. 1. For a direct comparison, we employ the identical signal configuration as in Fig. 8 of Ref. [1]. However, a key distinction lies in the coherent integration time: guided by the analysis in Fig. 5, we extend the duration to $T = 64\,\mathrm{s}$, relaxing the strict quasi-monochromatic limit ($T = 8\,\mathrm{s}$) adopted in the previous work. The upper panel illustrates the temporal evolution of the total power statistic $\mathcal{L}_i$ (solid curve). Due to the extended integration time, the overall amplitude of $\mathcal{L}_i$ is significantly elevated compared to Ref. [1], benefiting from the increased coherent gain. The curve's modulation reflects the secular evolution of the gravitational-wave amplitude coupled with the diurnal variation of the detector's antenna pattern. The scatter points depict the λ values distributed across the anchor bin ($\kappa = 0$, red) and its nearest neighbors ($\kappa = \pm 1$, blue/green). A notable feature appears in the high-frequency regime (later times): while the total power remains high, the extended T induces severe spectral leakage as the signal's chirp rate increases. Consequently, the power concentrated in individual bins (the dots) suffers greater loss relative to the total power. This phenomenon is quantified in the lower panel by the spectral leakage factor η, which exhibits rapid oscillatory behavior corresponding to the evolution of the sampling phase $o_{0,i}$.

The objective of the subsequent sections is to construct a precise analytical model for this universal leakage function $\eta(o)$.

C. Validity of Approximations

The extent to which a signal's frequency evolves within a single STFT segment is quantified by the dimensionless *widening factor*, w. It is defined as the total frequency drift across the segment duration T_{DFT}, normalized by the frequency resolution bin width $f_{\mathrm{bin}} = 1/T_{\mathrm{DFT}}$:

$$w \equiv |\dot{f}|T_{\mathrm{DFT}}^2 = \frac{|\dot{f}|T_{\mathrm{DFT}}}{f_{\mathrm{bin}}}. \tag{16}$$

This parameter serves as the governing metric for the validity of spectral leakage models $\eta(o)$. In the *monochromatic limit* ($w \to 0$), where frequency evolution is negligible, the leakage function simplifies to the squared modulus of the window function's Fourier transform [1]:

$$\eta(o) = \left|\widetilde{W}(o)\right|^2. \tag{17}$$

Previous studies [57, 61] indicate that this approximation remains highly accurate for signals satisfying the *quasi-monochromatic condition*, typically $w \lesssim 0.5$.

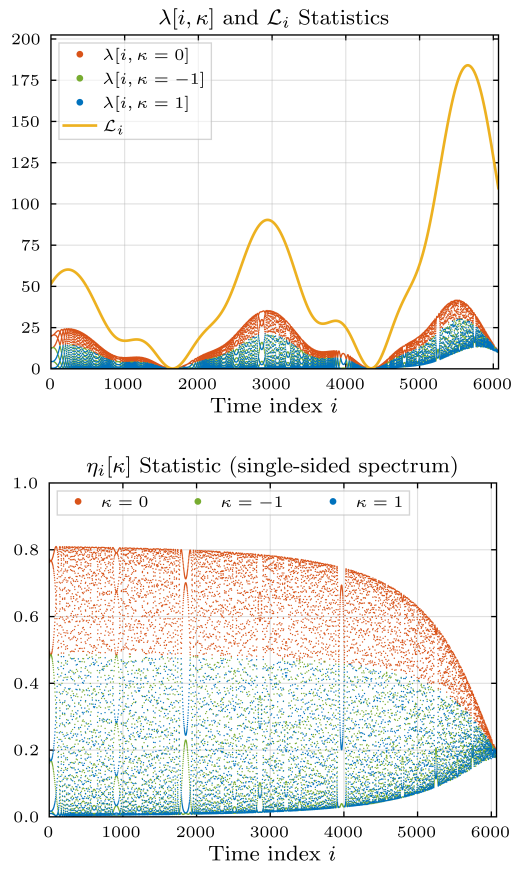

FIG. 1. **Evolution of $\lambda[i,\kappa]$ and $\mathcal{L}_i$ (upper) and the spectral leakage factor $\eta_i[\kappa]$ (lower) over STFT time segments for a simulated mini-EMRI signal**. In both panels, the red dots represent the value in the anchoring frequency bin ($\kappa = 0$), while the blue and green dots show the values in the adjacent bins ($\kappa = \pm 1$). The yellow curve in the upper panel shows the total power statistic $\mathcal{L}_i$, calculated using a smoothed LIGO-H1 O3 PSD. The simulation uses a signal with $m_1 = 1.5M_\odot$, $m_2 = 1 \times 10^{-5}M_\odot$, and $d = 8$ kpc, analyzed in the 100-200 Hz band. STFT parameters include $T_{\rm DFT} = 64$s, sampling frequency $f_s = 512$Hz, a Tukey window ($\alpha = 0.5$), and 50% overlap between segments.

However, when w is non-negligible, the monochromatic assumption breaks down. To rigorously assess the required model complexity, we analyze the signal phase $\Phi(t) = 2\pi \int f(t)dt$ by performing a Taylor series expansion of the instantaneous frequency $f(t)$ around the segment center. This expansion decomposes the phase into a polynomial in time: the linear term ($\propto t$) corresponds to the constant carrier frequency, the quadratic term ($\propto t^2$) arises from the linear frequency derivative $\dot{f}$ (linear chirp), and higher-order terms describe more complex frequency evolutions (e.g., $\ddot{f}$).

The validity of truncating this series is determined by the phase contribution of the neglected terms. We can define a critical widening factor $w_{\rm crit,m}$, as the value of w at which the contribution of the m-th order phase term becomes significant. Assuming the signal follows a local power-law evolution $\dot{f} = kf^n$, we derive a universal constant for the quadratic phase term—arising from the linear frequency drift $\dot{f}$—which is independent of specific signal parameters:

$$w_{\rm crit,2} = \frac{4}{\pi} \approx 1.2732. \tag{18}$$

This value rigorously demarcates the boundary of the monochromatic approximation. When $w \ll w_{\rm crit,2}$, the effects of linear frequency evolution are negligible. As w approaches or exceeds this critical value, the quadratic phase term induces significant spectral spreading, necessitating a linear chirp model.

For phase terms of order $m > 2$ (arising from $\ddot{f}$ and higher derivatives), the critical thresholds $w_{\rm crit,m}$ are not constant but depend on the specific signal parameters and frequency evolution. These are illustrated for two representative binary systems in Fig. 2.

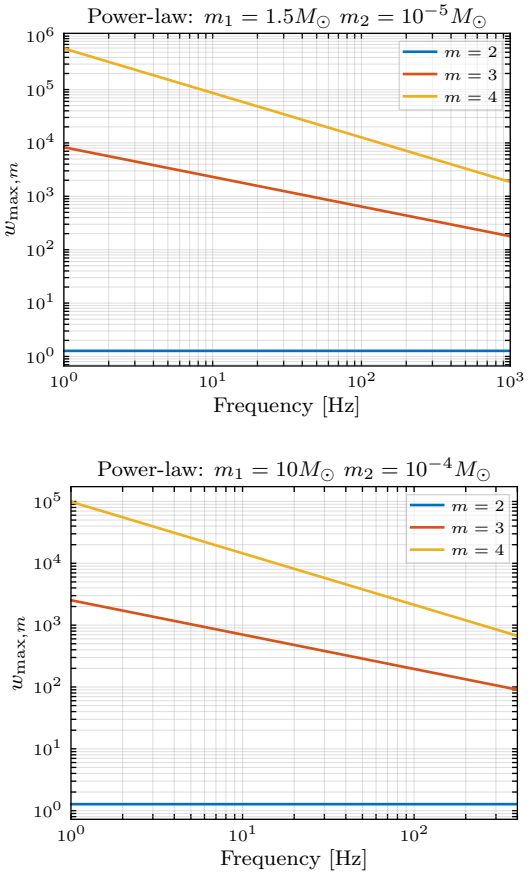

FIG. 2. **Frequency evolution of the critical widening (see text) factors $w_{\rm crit,m}$ (for orders $m = 3, 4$), assuming a zeroth-order post-Newtonian inspiral model ($n = 11/3$)**. The two panels correspond to different representative binary system parameters.

A key finding of this analysis is that the critical thresholds for higher-order terms are extremely high, with

$w_{\mathrm{crit},3}$ typically ranging from 10^2 to 10^5. This has profound practical implications. To prevent excessive spectral smearing and preserve viable SNR, STFT-based analyses are typically constrained to a regime where $w \lesssim 10$. Since the practical limit on w is orders of magnitude smaller than the critical values required for cubic phase terms to become relevant ($w \ll w_{\mathrm{crit},3}$), the effects of the second frequency derivative $\ddot{f}$—as well as all higher-order derivatives—can be safely neglected.

Consequently, this analysis rigorously justifies approximating the signal within each STFT segment as a *linear chirp* (i.e., constant $\dot{f}$). This simplification underpins the entire analysis framework presented in this paper. A detailed derivation is provided in Appendix A.

D. Quantitative Description of Spectral Leakage

Having justified the linear chirp approximation in Section III C, we now provide a detailed mathematical formulation of the resulting spectral leakage pattern. This serves as a high-fidelity local model for the signal's spectral shape.

The distribution of signal power within an STFT segment is completely characterized by the continuous *spectral leakage function*, $\eta(o, w)$. This function is parameterized by the dimensionless *widening factor* w and *offset factor* o, which quantify the intra-segment frequency drift and the proximity to the central frequency, respectively.

For a window function $W(\tau)$ defined on the normalized interval $\tau \in [-1/2, 1/2]$ and normalized to unit power (see Appendix C for details), the leakage function has the general form:

$$\eta(o, w) = \frac{1}{\pi w} \left| \int_{\sqrt{\pi}(o/\sqrt{w} - \sqrt{w}/2)}^{\sqrt{\pi}(o/\sqrt{w} + \sqrt{w}/2)} W\left(\frac{u}{\sqrt{\pi w}} - \frac{o}{w}\right) e^{iu^2} du \right|^2 . \tag{19}$$

Fig. 3 illustrates representative profiles of the spectral leakage function. Generally, the function exhibits a decay as the magnitude of the offset factor $|o|$ increases, indicating that power is primarily localized around the carrier frequency. However, a crucial phenomenon is observed at higher widening: as w increases, the signal power is smeared across a wider range of adjacent frequency bins.

This analytical model relies on the slow variation assumption: The signal's power is effectively constant over the segment duration T_{DFT}. This is valid because the timescales of astrophysical amplitude evolution and detector antenna pattern modulation are typically much longer than the STFT coherence time.

Finally, for real-valued strain data, the power is distributed symmetrically between positive and negative frequencies. The leakage function derived above describes the one-sided spectrum; thus, the effective leakage for the positive frequency components is approximately half that of the complex equivalent derived in Eq. (19). A rigorous treatment accounting for the real-valued nature of the signal is presented in Appendix B.

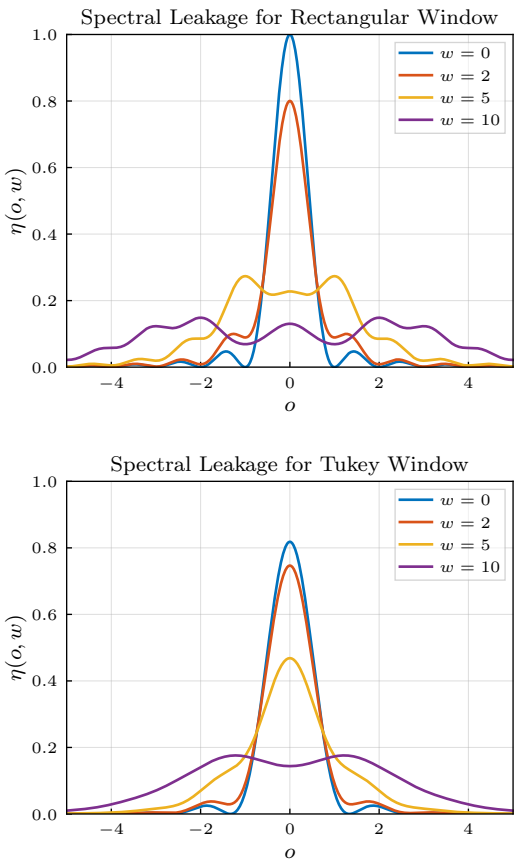

FIG. 3. **The spectral leakage function $\eta(o, w)$ plotted against the dimensionless frequency offset o for representative widening factors** $w \in \{0, 2, 5, 10\}$. The limiting case $w = 0$ represents a monochromatic signal. The upper panel illustrates the leakage pattern for a Rectangular window, while the lower panel shows the pattern for a Tukey window.

IV. DETECTION STATISTIC

A. Statistical method

1. *Distribution property*

The power ratio $R[i, k]$ in a pixel of the time-frequency map is a random variable that follows the probability density function (PDF):

$$\mathrm{PDF}\left\{R = x | \lambda\right\} = e^{-x - \frac{\lambda}{2}} I_0\left(\sqrt{2\lambda x}\right), \quad x \geq 0, \tag{20}$$

where λ is the non-centrality parameter, representing the expected power SNR in that bin, and $I_0(\cdot)$ is the modified Bessel function of the first kind of order zero.

The mean and variance of $R[i, k]$ can be calculated directly from this distribution. For a given non-centrality

parameter $\lambda[i,k]$, they are:

$$\begin{aligned}\mu_R[i,k] &= 1 + \frac{1}{2}\lambda[i,k],\\ \sigma_R^2[i,k] &= 1 + \lambda[i,k].\end{aligned} \tag{21}$$

It is crucial to note that these are exact results and do not rely on any weak-signal approximations.

Conventional search methodologies, such as the widely implemented Hough transform pipeline [1, 57, 58, 61–64], typically employ a peak-finding strategy for data reduction. In this paradigm, the full information content of the normalized power spectrogram, $R[i,k]$, is condensed into a sparse binary "peakmap" by identifying bins that exceed a significance threshold and constitute local maxima along the frequency axis. This discretized representation subsequently serves as the input for track-finding algorithms to identify candidate signal trajectories.

In this work, we adopt a more direct approach by operating on the power ratio map itself, which retains more information and possesses well-defined statistical properties. For a given point in the parameter space (i.e., a specific set of signal parameters), there is a corresponding trajectory on the time-frequency map. Due to spectral leakage, the signal power is not confined to a single pixel per time segment but is distributed across several frequency bins. The bins directly on the trajectory are expected to have the highest SNR, with the SNR decreasing in adjacent bins as a function of their distance from the central track.

We propose the ΣR statistic, computed by a weighted summation of the $R[i,k]$ values along a candidate trajectory. The weights $\omega_{i,k}$ are chosen to match the expected power SNR distribution with spectral leakage. To streamline the mathematical formulation, we define the *generalized track set* $\mathcal{T}$ as the collection of time-frequency pixels (i,k) associated with the signal. The detection statistic is then expressed as a unified weighted sum over this set:

$$\Sigma R = \sum_{\mathcal{T}} \omega_{i,k} R[i,k]. \tag{22}$$

Crucially, unlike the single-pixel tracks utilized in conventional methods, $\mathcal{T}$ can incorporate a cluster of frequency bins at each time step to recover the signal power dispersed by spectral leakage. Unless otherwise specified, we adopt a 3-pixel track width configuration (comprising the anchor bin and its two immediate sidebands, corresponding to relative indices $\kappa \in \{0, \pm 1\}$) as the default setting throughout this work.

Under the assumption that the power ratios $R[i,k]$ in different time-frequency pixels are statistically independent, the Central Limit Theorem implies that the aggregate statistic ΣR follows an approximately Gaussian distribution. By the linearity of expectation and variance, its mean and variance are derived as the weighted sums of the individual moments:

$$\begin{aligned}\mu_{\Sigma R} &= \sum_{\mathcal{T}} \omega_{i,k} \left(1 + \frac{1}{2}\lambda[i,k]\right),\\ \sigma_{\Sigma R}^2 &= \sum_{\mathcal{T}} \omega_{i,k}^2 \left(1 + \lambda[i,k]\right).\end{aligned} \tag{23}$$

It is important to acknowledge that in practice, the assumption of strict independence is not fully satisfied, primarily due to correlations induced by the overlap between time segments and spectral leakage between frequency bins. A detailed discussion on the impact of these correlations and the validity of this approximation is provided in Appendix H.

To create a standardized detection statistic, we define the *Critical Ratio* (CR). This quantity measures the significance of the observed ΣR by normalizing it with respect to its expected value and standard deviation under the null hypothesis (i.e., in the absence of a signal, where $\lambda[i,k] = 0$ for all pixels). The CR is therefore defined as the background-subtracted statistic divided by the background standard deviation:

$$\mathrm{CR} \equiv \frac{\Sigma R - \sum_{\mathcal{T}} \omega_{i,k}}{\sqrt{\sum_{\mathcal{T}} \omega_{i,k}^2}}. \tag{24}$$

By construction, the CR statistic for a track containing only noise is a random variable with a mean of 0 and a variance of 1. When a signal is present, the expected mean and variance of the CR statistic are shifted. Substituting the signal-dependent moments from Eq. (23) into the definition yields:

$$\begin{aligned}\mu_{\mathrm{CR}} &= \frac{1}{2}\frac{\sum_{\mathcal{T}} \omega_{i,k}\lambda[i,k]}{\sqrt{\sum_{\mathcal{T}} \omega_{i,k}^2}},\\ \sigma_{\mathrm{CR}}^2 &= 1 + \frac{\sum_{\mathcal{T}} \omega_{i,k}^2\lambda[i,k]}{\sum_{\mathcal{T}} \omega_{i,k}^2}.\end{aligned} \tag{25}$$

The quantity μ_{CR} can be interpreted as the expected statistical significance of the detection for a given candidate track. The variance, σ_{CR}^2, is close to unity for weak signals but increases with signal strength. The CR serves as the final, optimized statistic for identifying candidate signals in the data.

2. *Optimization of the weights*

The detection efficacy of the ΣR statistic is critically dependent on the optimal selection of the weights $\omega_{i,k}$. Maximizing the detection significance is equivalent to maximizing the expected Critical Ratio, μ_{CR}. According to the Cauchy-Schwarz inequality, this optimum is achieved when the weights are directly proportional to the expected power SNR in each time-frequency pixel: $\omega_{i,k} \propto \lambda[i,k]$. This aligns with the matched filter principle and was also a key component of the winning solution

in the Kaggle challenge [72]: weights should be proportional to the expected signal strength. We note that since the CR statistic is invariant under global scaling transformations (i.e., $\omega_{i,k} \to c \cdot \omega_{i,k}$ for any $c > 0$), the absolute normalization of the weights does not affect the detection significance.

In practice, however, constructing weights that are strictly proportional to the instantaneous $\lambda[i,k]$ is computationally intractable for blind searches. The difficulty arises from the sensitivity of the spectral leakage factor, $\eta_i[\kappa] = \eta(\kappa + o_{0,i})$, to the sub-bin sampling phase. As illustrated in the bottom panel of Fig. 1, η oscillates rapidly with variations in the offset factor $o_{0,i}$. Consequently, predicting the exact value of $\eta_i[\kappa]$ for every pixel would require a priori knowledge of the source parameters with a precision far exceeding that of the search grid, which is unrealistic for wide-parameter-space surveys.

To circumvent this limitation, we implement a *sampling phase-averaged strategy*, extending `ΣTrack` from [1]. Instead of relying on the highly sensitive instantaneous leakage, we define the *averaged leakage function*, $\hat{\eta}(o, w)$, by integrating over a unit bin interval:

$$\hat{\eta}(o, w) = \int_{o-1/2}^{o+1/2} \eta(o', w)\, do'. \tag{26}$$

From this, we define the discrete *averaged spectral leakage factor* for the i-th segment as $\hat{\eta}_i[\kappa] \equiv \hat{\eta}(\kappa, w_i)$. This formulation effectively marginalizes over the unknown sampling phase $o_{0,i}$, smoothing out the rapid oscillatory behavior of the instantaneous leakage. As demonstrated in Fig. 4, the averaged function captures the local mean spectral behavior over multiple segments, providing a deterministic template for weight construction.

Consequently, the practical weights are defined proportional to this averaged quantity:

$$\omega_{i,k} \propto \mathcal{L}_i \hat{\eta}_i[k]. \tag{27}$$

Substituting these weights into Eq. (23) and invoking the slow-variation approximations (see Appendix D), the expected mean and variance of the CR simplify to:

$$\begin{aligned} \mu_{\mathrm{CR}} &= \frac{1}{2}\sqrt{\sum_{\mathcal{T}} \left(\mathcal{L}_i \hat{\eta}_i[k]\right)^2}, \\ \sigma^2_{\mathrm{CR}} &= 1 + \frac{\sum_{\mathcal{T}} \left(\mathcal{L}_i \hat{\eta}_i[k]\right)^3}{\sum_{\mathcal{T}} \left(\mathcal{L}_i \hat{\eta}_i[k]\right)^2}. \end{aligned} \tag{28}$$

The adoption of this sampling phase-averaged weighting scheme yields a detection statistic that is significantly less sensitive to small parameter mismatches (specifically, errors in the sampling phase $o_{0,i}$). By effectively smoothing out the rapid fluctuations of the instantaneous value, this approach ensures that the statistic remains robust and computationally feasible for practical blind search campaigns.

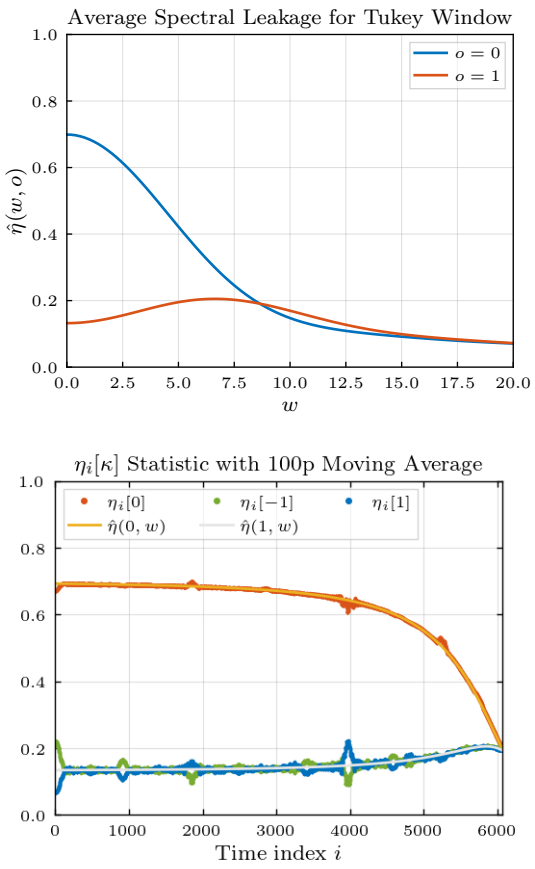

FIG. 4. **The averaged spectral leakage function and its comparison with smoothed simulation data.** (Upper panel) The average spectral leakage function $\hat{\eta}(w, o)$. The blue curve shows the power fraction in the central bin ($\kappa = 0$), while the red curve shows the power in the first adjacent bin ($\kappa = 1$). As widening increases, power clearly leaks from the central bin into its neighbors. (Lower panel) A direct comparison validating the averaged model. The rapidly oscillating data from Fig. 1 is smoothed using a 100-point moving average. This empirically smoothed result shows excellent agreement with the theoretical averaged leakage function, confirming its validity as a robust model for the expected power distribution.

B. Maximum detectable distance

The primary objective of this statistical framework is to map the abstract detection statistic into physically interpretable constraints on the universe. Specifically, we aim to translate the statistical confidence thresholds—defined by the false alarm probability (P_{fa}) and false dismissal probability (P_{fd})—into a metric of astrophysical sensitivity: the maximum distance, d_{max}, at which a specific source can be confidently detected.

The detection performance is governed by two fundamental parameters: the false alarm probability P_{fa}, which quantifies the risk of incorrectly classifying background noise as a signal, and the false dismissal probability P_{fd},

which quantifies the risk of overlooking a genuine signal present in the data.

First, the choice of $P_{\rm fa}$ determines the detection threshold. Assuming the CR in pure noise follows a standard normal distribution (as justified in Section IV), the detection threshold $\mathrm{CR}_{\rm thr}$ is given by the inverse complementary error function:

$$\mathrm{CR}_{\rm thr} = \sqrt{2}\ \mathrm{erfc}^{-1}(2P_{\rm fa}). \tag{29}$$

Any candidate event with a CR value falling below this threshold is deemed statistically insignificant and rejected.

Conversely, to ensure a high probability of detecting a true signal, its expected CR value must lie sufficiently above this noise floor. For a given false dismissal probability $P_{\rm fd}$, the condition for the minimum required mean statistic is:

$$\mu_{\rm CR} \geq \sqrt{2}\ \sigma_{\rm CR}\ \mathrm{erfc}^{-1}(2P_{\rm fd}) + \mathrm{CR}_{\rm thr}. \tag{30}$$

This inequality establishes the minimum statistical significance required to claim a detection with the specified confidence level.

To translate this statistical requirement into a physical distance, we isolate the distance dependence within the total power statistic $\mathcal{L}_i$. We define the *intrinsic strain amplitude*, $h_{e,i} \equiv h_{0,i}d$, which encapsulates the source's intrinsic evolution independent of its distance d. Consequently, the total power SNR statistic can be factored as:

$$\mathcal{L}_i = \frac{1}{d^2}\left(2h_{e,i}^2 Q_i^2 \frac{T_{\rm DFT}}{S_{n,i}}\right), \tag{31}$$

where Q_i represents the antenna pattern response and $S_{n,i}$ is the noise power spectral density. A detailed discussion is in Appendix C.

As detailed in Appendix E, when estimating the detectable distance, we are by definition operating in the *weak-signal limit*. In this regime, the signal's contribution to the variance is negligible compared to the noise background, implying $\sigma^2_{\rm CR} \approx 1$. Applying this approximation allows us to decouple the distance dependence and derive the central result of our sensitivity analysis:

$$d_{\rm max} = \left[\sqrt{2}\ \mathrm{erfc}^{-1}(2P_{\rm fa}) + \sqrt{2}\ \mathrm{erfc}^{-1}(2P_{\rm fd})\right]^{-1/2} \left[\frac{1}{4}\sum_{\mathcal{T}} h_{e,i}^4 Q_i^4 \frac{T_{\rm DFT}^2}{S_{n,i}^2}\hat{\eta}_i^2[\kappa]\right]^{1/4}. \tag{32}$$

This analytical formula provides a direct mapping from source parameters to the detection horizon. For any given astrophysical model, it enables the calculation of the maximum sensitive distance, thereby quantifying the effective volume of the universe accessible to our search.

C. Effective parameter sphere

Practical searches for gravitational-wave signals in noisy data necessitate the use of a discrete grid of waveforms, known as a *template bank*, to span the vast parameter space of potential sources. Since the parameters of astrophysical signals are continuous variables, a real signal will rarely align perfectly with any single discrete template. Consequently, it is imperative to quantify the degradation of the detection statistic caused by the parameter mismatch between the search template and the true signal.

To formalize this, we define the output of the search pipeline as the expected Critical Ratio, $\mu_{\rm CR}(\mathcal{P}_t|\mathcal{P}_r)$, obtained when probing for a signal with true parameters $\mathcal{P}_r$ using a template with parameters $\mathcal{P}_t$. While the analytical form of $\mu_{\rm CR}$ is derived in Appendix D, its value is fundamentally governed by the degree of overlap between the template and the signal waveforms.

To quantify the recovery efficiency, following the formalism in [1], we define the *fitting factor* (FF). This metric represents the ratio of the expected statistic recovered by the template to the maximum achievable statistic if the template were perfectly matched to the signal:

$$\mathrm{FF}(\mathcal{P}_t|\mathcal{P}_r) \equiv \frac{\mu_{\rm CR}(\mathcal{P}_t|\mathcal{P}_r)}{\mu_{\rm CR}(\mathcal{P}_r|\mathcal{P}_r)}. \tag{33}$$

By definition, the fitting factor is a normalized measure of effectualness, ranging from 0 (no match) to 1 (perfect match).

Building on this, we define the *mismatch* (MM) to represent the fractional loss in astrophysical sensitivity. Unlike standard conventions which often define mismatch in terms of power SNR loss, we adopt a definition directly tied to the cosmological reach. As established in Eq. (32) and Eq. (28), the detection distance as $d_{\rm max} \propto \sqrt{\mu_{\rm CR}}$. Consequently, the fractional loss in detectable distance is given by $1 - \sqrt{\mathrm{FF}}$. We therefore define the mismatch as:

$$\mathrm{MM} \equiv 1 - \sqrt{\mathrm{FF}}. \tag{34}$$

This physically motivated definition renders the mismatch an exceptionally intuitive metric: a value of $\mathrm{MM} = 0.01$ corresponds to a 1% reduction in the detector's reach distance.

This metric serves as the foundational criterion for template bank construction. The objective is to populate

the parameter space with templates such that the mismatch for any arbitrary signal remains below a predefined tolerance, $\mathrm{MM}_{\max}$, thereby ensuring uniform sensitivity across the target parameter space. This requirement can be formulated as a classic geometric covering problem. For a given template, the condition $\mathrm{MM} \leq \mathrm{MM}_{\max}$ defines a high-dimensional volume around the template center within which signals are detected with sufficient efficiency. In the limit of small parameter offsets, this iso-mismatch surface is well-approximated by a hyper-ellipsoid, often referred to as the *metric ellipse* or *effective detection sphere*. The task of constructing a template bank is thus transformed into the challenge of efficiently tiling the parameter space with these overlapping ellipsoids. The overarching goal is to minimize the number of templates to reduce computational cost while guaranteeing that no "holes" exist where the sensitivity drops below the required threshold. The specific algorithms implementing this stochastic geometric placement are detailed in subsequent work.

V. RESULTS

In this section, we apply the comprehensive theoretical framework developed in the preceding sections to a practical search scenario. We leverage our description of spectral leakage, which is valid beyond the traditional quasi-monochromatic approximation, to guide the optimization of key search parameters. By employing the new ΣR statistic, we demonstrate how this theoretically-grounded approach enables a significant enhancement in search sensitivity. The ultimate metric for this sensitivity is the effective detection volume, V_{eff}, which represents the astrophysical reach of our search averaged over all source sky locations and orientations, as formally defined in Eq. (F3) of Appendix F. This volume is computed via the effective distance, d_{eff}. The calculation is made tractable by incorporating the fully-averaged antenna-pattern factor $\langle Q \rangle$ (from Eq. (F6)) which encapsulates the complete geometric average over sky position, orientation, and sidereal time.

We now apply our framework to a concrete target: a fiducial mini-EMRI system consisting of a typical neutron star ($m_1 = 1.5 M_\odot$) and a strange compact object ($m_2 = 10^{-5} M_\odot$). Our analysis focuses on the 100-200 Hz frequency band, where the signal is expected to be most prominent for ground-based detectors. Fig. 1 illustrates the key statistics for a simulated signal from such a system.

Our primary goal is to determine the optimal coherent time T_{DFT} that maximizes search sensitivity. Previous searches were restricted by the quasi-monochromatic approximation, which strictly limited the coherent time to $T_{\mathrm{DFT}} \leq 8$ s to avoid excessive spectral leakage. Our new framework, which analytically models and accounts for this leakage, allows us to break free of this long-standing constraint. This enables a comprehensive optimization of T_{DFT} over a wide range to maximize search sensitivity.

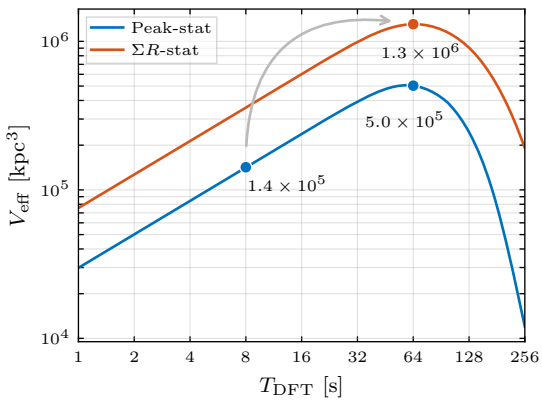

FIG. 5. **The effective detection volume V_{eff} as a function of coherent integration time, T_{DFT}, for the fiducial mini-EMRI system, computed using the O3 PSD for the LIGO Hanford detector**. The performance is evaluated for a set $P_{\mathrm{fd}} = 5\%$ and $P_{\mathrm{fa}} = 1\%$. STFT uses the Tukey window function ($\alpha = 0.5$) with a 50% overlap. The red curve illustrates the performance of our novel ΣR statistic. In comparison, the blue curve depicts the performance of a statistic based on the Hough transform strategy. Instead of a full pipeline execution, this peak statistic is computed by summing binary pixels along the signal trajectory, mimicking the standard Hough approach which selects peaks based on a power ratio threshold (here $\theta = 2.5$) and a spectral local maximum condition. The performance at $T = 8$s, the optimal choice under the quasi-monochromatic assumption, is marked in the panel. Our new method, which transcends this limitation, finds a true optimal coherent time at $T = 64$s.

Fig. 5 illustrates the direct consequence of this advancement, plotting the effective detection volume as a function of T_{DFT} for both our new method and a conventional approach. For both pipelines, the detection volume initially rises with longer integration times—a direct result of the coherent gain in SNR (see Appendix E). However, as T_{DFT} continues to increase, the signal's rapid frequency evolution begins to spread its power across many frequency bins. This spectral leakage eventually dominates, causing the sensitivity to decline. While the conventional method is confined to the quasi-monochromatic limit at $T_{\mathrm{DFT}} = 8\,\mathrm{s}$, our method identifies a true global optimum at $T_{\mathrm{DFT}} = 64\,\mathrm{s}$. At this optimal integration time, our method yields an approximately 10-fold increase in the effective detection volume compared to the previous state-of-the-art. This represents a dramatic, order-of-magnitude improvement in search sensitivity.

To demonstrate the practical impact of our method in a *realistic broadband search scenario*, we move beyond narrow-band analyses and consider a full signal trajectory evolving from 30 Hz to 1000 Hz over an observation period of 66 days. In this context, we allow both the conventional and the proposed pipelines to freely optimize their parameters to achieve their respective maximum performance.

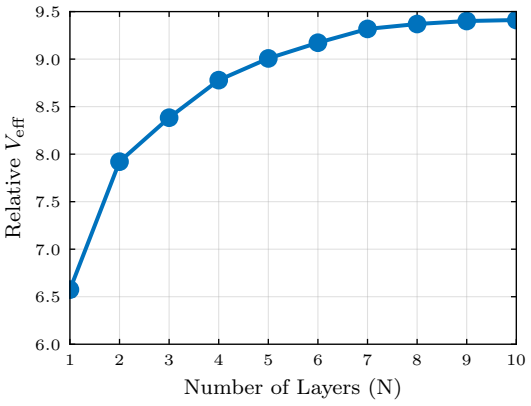

FIG. 6. **Relative effective detection volume as a function of the number of frequency layers** N. The y-axis represents the ratio of the effective detection volume to that of the reference case, which is the optimal configuration of the conventional Hough-transform pipeline.

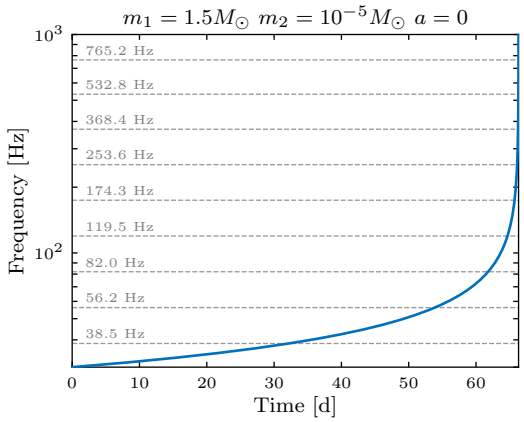

FIG. 7. **Visualization of the optimal frequency-layering strategy for the target mini-EMRI signal**. The blue curve represents the signal's frequency evolution over time. The entire search band is partitioned into $N = 10$ distinct layers by critical frequencies (horizontal dashed lines), with the coherent time $T_{\rm DFT}$ decreasing stepwise from 1024 s in the lowest layer to 2 s in the highest.

For the conventional Hough-transform pipeline, we perform a global optimization to select the single best coherent time $T_{\rm DFT}$ that maximizes the total detection volume over this wide band. The optimization yields a best-case scheme of $T_{\rm DFT} = 32$ s. Although this is the mathematical optimum for the standard pipeline, it effectively restricts the sensitive search window to a narrow frequency band of $f \in [80, 135]$ Hz—sacrificing the vast majority of the signal's time-frequency evolution.

In stark contrast, the ΣR framework is intrinsically capable of local optimization. Leveraging our rigorous leakage model, we are free to implement a frequency-layered strategy that adapts to the signal's evolution across the entire 30–1000 Hz band, which extends a framework presented for searches for compact binary inspirals in future ground-based GW interferometers [73] to mini-EMRIs. By partitioning the bandwidth into distinct layers and optimizing the coherent time locally, we maintain peak sensitivity throughout the signal's lifetime.

The result of this comparison, illustrated in Fig. 6, reveals a twofold advantage of our framework. First, even in the simplest single-layer configuration ($N = 1$), the ΣR method already yields a 6.5-fold increase in detection volume compared to the optimized Hough baseline. Second, as we exploit the freedom to implement the frequency-layered strategy, the performance improves further. As the number of layers increases from $N = 1$ to $N = 10$, the improved matching of the coherent time to the local signal evolution compounds the gain, ultimately achieving an effective detection volume that is approximately *an order of magnitude larger* than the conventional limit.

To elucidate the practical implementation of the multi-layer strategy, Fig. 7 presents the detailed configuration for the representative case of $N = 10$ layers. The plot visualizes the target signal's time-frequency trajectory overlaid with the optimized layer boundaries, defined by a set of critical frequencies. Details about the frequency layering strategy are discussed in Appendix G.

Enabled by this frequency-layered architecture, the coherent integration time $T_{\rm DFT}$ is dynamically adapted to the signal's local chirp rate, decreasing stepwise from a maximum of 1024 s in the lowest frequency layer to 2 s in the highest layer. This stepwise reduction is critical to counterbalance the rapidly accelerating frequency evolution of the inspiral. Consequently, the widening factor remains constrained within the optimal regime throughout the entire observation. By preventing excessive spectral leakage at high frequencies while preserving long coherent integration times at low frequencies, this configuration realizes the order-of-magnitude expansion in detection volume quantified in Fig. 6.

VI. CONCLUSIONS

In this paper, we have taken a major step forward in developing a search pipeline for mini-EMRI signals. We have extended our method, `ΣTrack`, to handle the case of significant spectral leakage via a new statistic, ΣR, that allows us to significantly increase the analysis coherence time with respect to that used in standard methods that only consider the signal as monochroamtic within each fast Fourier transform and do not consider the impact of spectral leakage. additionally, we have derived a new sensitivity estimate of our method, and shown that by accurately modeling spectral leakage and optimizing the coherent time, `ΣTrack` boosts the effective detection volume for this system by an order of magnitude. This dramatic enhancement in search sensitivity translates di-

rectly into a correspondingly higher probability of detecting these elusive mini-EMRI sources with current and future gravitational-wave observatories.

In future work, we plan to generalize our method further to handle non-stationary, non-Gaussian noise, which will be essenetial to apply `ΣTrack` to real data. Moreover, following [1], we will determine a discretization of the search parameter space based on a mismatch between signals in the time-frequency plane to generate a bank of mini-EMRI templates to ensure that a real search is tractable.

ACKNOWLEDGMENTS

We would like to thank Rodrigo Tenorio for carefully reading the manuscript and for valuable suggestions. This material is based upon work supported by NSF's LIGO Laboratory which is a major facility fully funded by the National Science Foundation. We acknowledge the Python scientific ecosystem, particularly `NumPy` [74], `SciPy` [75], and `Matplotlib` [76], which were used in the analysis and visualization. This work is supported by the National Natural Science Foundation of China (NSFC) under Grant No. 12347103 and 12547104.

Appendix A: Phase Expansion in the Fourier Domain

In this appendix, we provide a detailed derivation of the conditions under which a general frequency-evolving signal can be approximated by either a monochromatic or a linear-chirp model within a single coherent integration time, T. This analysis forms the mathematical basis for extending the coherent integration time beyond the traditional quasi-monochromatic limit.

We begin with a general, continuous frequency-modulated signal, which can be expressed in complex form as:

$$s(t) = A(t)e^{i\phi(t)}, \tag{A1}$$

where $A(t)$ is the time-varying amplitude and $\phi(t)$ is the phase. The phase is determined by the integral of the signal's instantaneous frequency $f(t)$:

$$\phi(t) = \phi_0 + 2\pi \int_0^t f(\tau)d\tau, \tag{A2}$$

where ϕ_0 is the initial phase at $t = 0$.

To analyze this signal in the frequency domain, we consider its STFT of duration T centered at $t = 0$. The Fourier transform of the windowed segment is:

$$\tilde{s}(f) = \int_{-T/2}^{T/2} A(t) \exp\left(i\phi_0 + i2\pi \int_0^t f(\tau)d\tau\right) e^{-i2\pi ft}dt. \tag{A3}$$

To make this integral tractable, we Taylor-expand the instantaneous frequency $f(t)$ around the center of the window, $t = 0$:

$$f(t) = f_0 + \dot{f}_0 t + \frac{1}{2}\ddot{f}_0 t^2 + \cdots, \tag{A4}$$

where $f_0, \dot{f}_0, \ddot{f}_0$ are the frequency and its time derivatives evaluated at $t = 0$. Integrating this series gives the full phase expansion:

$$\begin{aligned}\phi(t) &= \phi_0 + 2\pi \int_0^t \left(f_0 + \dot{f}_0\tau + \frac{1}{2}\ddot{f}_0\tau^2 + \cdots\right) d\tau \\ &= \phi_0 + 2\pi \left(f_0 t + \frac{1}{2}\dot{f}_0 t^2 + \frac{1}{6}\ddot{f}_0 t^3 + \cdots\right).\end{aligned} \tag{A5}$$

For many astrophysical signals, the amplitude $A(t)$ evolves on a much longer timescale than the coherent integration time T. We can therefore make the standard assumption that the amplitude is constant within the window, $A(t) \approx A(0)$. This allows us to simplify the Fourier integral significantly:

$$\begin{aligned}\tilde{s}(f) \approx& A(0) \int_{-T/2}^{T/2} e^{i\phi(t)} e^{-i2\pi ft} dt \\ =& A(0) \int_{-T/2}^{T/2} e^{\left(i\phi_0 + i2\pi f_0 t + i\pi \dot{f}_0 t^2 + \frac{i\pi}{3}\ddot{f}_0 t^3 + \cdots\right)} e^{-i2\pi ft} dt.\end{aligned} \tag{A6}$$

The central challenge in simplifying the integral is to determine at which order the phase expansion can be truncated. We can truncate the series after the quadratic term (t^2) if the combined phase contribution from all higher-order terms (cubic and beyond) is much less than one radian over the entire integration interval. When this condition, $|\sum_{j=3}^{\infty} 2\pi \frac{f_0^{\{j-1\}}}{j!} t^j| \ll 1$, is met, we can use the approximation $e^x = 1 + x + x^2/2 + \cdots \approx 1$ for the higher-order part of the exponent.

This allows us to separate the integral into a primary component and a small residual term, which can then be neglected:

$$\begin{aligned}&\int_{-T/2}^{T/2} e^{\left(i\phi_0 + i2\pi f_0 t + i\pi \dot{f}_0 t^2 + \frac{i}{3}\pi \ddot{f}_0 t^3 + \cdots\right)} e^{-i2\pi ft} dt \\ =& \int_{-T/2}^{T/2} e^{\left(i\phi_0 + i2\pi f_0 t + i\pi \dot{f}_0 t^2\right)} \left(1 + \frac{i}{3}\pi \ddot{f}_0 t^3 + \cdots\right) e^{-i2\pi ft} dt \\ \approx& \int_{-T/2}^{T/2} e^{\left(i\phi_0 + i2\pi f_0 t + i\pi \dot{f}_0 t^2\right)} e^{-i2\pi ft} dt.\end{aligned} \tag{A7}$$

To formalize this truncation, we must quantify the contribution of each higher-order term. We first adopt a general power-law signal model, $\dot{f} = kf^n$, which is representative of many astrophysical sources. Under this model, the m-th derivative of the frequency can be written in a compact form:

$$f_0^{\{m\}} = \left.\frac{d^m f(t)}{dt^m}\right|_{t=0} = C(m,n)k^m f_0^{m(n-1)+1}, \tag{A8}$$

where the coefficient is $C(m,n) = \prod_{j=0}^{m-1}[j(n-1)+1]$. The phase contribution corresponding to the t^m term in the expansion is:

$$\begin{aligned}\mathrm{PT}(m) &= 2\pi \int_0^t \frac{f_0^{\{m-1\}}}{(m-1)!}\tau^{m-1}d\tau \\ &= 2\pi \frac{f_0^{\{m-1\}}}{m!}t^m \\ &= 2\pi \frac{C(m-1,n)}{m!}k^{m-1}f_0^{(m-1)(n-1)+1}t^m.\end{aligned} \tag{A9}$$

The influence of this term is maximal at the edges of the window, $t = \pm T/2$. To create a dimensionless metric, we introduce the widening factor $w \equiv |\dot{f}_0|T^2$. By substituting $t = T/2$ into Eq. (A9) and re-expressing T in terms of w, we find the maximum phase contribution of the m-th order term:

$$\begin{aligned}\max|\mathrm{PT}(m)| &= 2\pi \frac{C(m-1,n)}{m!}k^{m-1}f_0^{(m-1)(n-1)+1}\left(\frac{T}{2}\right)^m \\ &= \frac{2\pi C(m-1,n)}{m!2^m}k^{\frac{m}{2}-1}f_0^{\frac{mn}{2}-m-n+2}w^{\frac{m}{2}}.\end{aligned} \tag{A10}$$

We now define a critical widening factor, $w_{\mathrm{crit},m}$, as the value of w for which this maximum phase contribution equals 1 radian. Setting the above expression to 1 and solving for w yields:

$$w_{\mathrm{crit},m} = \left(\frac{m!2^m}{2\pi C(m-1,n)}k^{1-\frac{m}{2}}f_0^{m+n-\frac{mn}{2}-2}\right)^{\frac{2}{m}}. \tag{A11}$$

This powerful result allows us to express the maximum phase contribution of any term in a very elegant form:

$$\max|\mathrm{PT}(m)| = \left(\frac{w}{w_{\mathrm{crit},m}}\right)^{\frac{m}{2}}. \tag{A12}$$

This inequality provides a clear condition for truncating the phase expansion. For a given order m, if the signal's widening factor w is much smaller than the critical factor $w_{\mathrm{crit},m}$, then the m-th order phase term is negligible. It is important to note that this linear approximation is only meaningful when the ratio in Eq. (A12) is small. When the phase contribution is of order 2π or larger, its periodic nature means its absolute value is no longer a good measure of its impact on the integral.

The framework provides a explanation for the validity of commonly used signal approximations. We summarize the key conclusions for the two most important cases.

- **The Quasi-Monochromatic Approximation** ($m = 2$): The phase expansion begins with the constant ($m = 0$) and linear ($m = 1$) terms, which represent the signal's central time frequency. The first term describing frequency evolution is the quadratic term ($m = 2$). The critical widening factor for this term is a universal constant for any power-law signal: $w_{\mathrm{max},2} = 4/\pi \approx 1.2732$. This result provides the formal justification for the widely-used quasi-monochromatic condition. To neglect all frequency evolution, the signal's widening factor w must be much less than $4/\pi$. The practical choice of $w < 0.5$ corresponds to the condition $w \ll w_{\mathrm{max},2}$, ensuring the signal can be accurately treated as monochromatic over the time duration T.

- **The Linear-Chirp Approximation** ($m = 3$): When the quasi-monochromatic condition is violated ($w \gtrsim 1$), we must include the quadratic phase term. The next question is whether the cubic term ($m = 3$) is also necessary. As shown in the examples in Fig. 2, the critical widening factor for the cubic term, $w_{\mathrm{max},3}$, is typically very large for astrophysical systems, often on the order of $10^3 - 10^4$.

This has a crucial implication: for any practical search where the widening factor is in the range of $0 < w \lesssim 10$, the condition $w \ll w_{\mathrm{max},3}$ is strongly satisfied. This means that the cubic and all higher-order phase terms can be safely ignored. Therefore, even when a signal evolves rapidly, a linear-chirp model (retaining terms up to $m = 2$) provides an extremely accurate representation of the signal's phase. This is the fundamental principle that enables our search method to extend the coherent integration time far beyond traditional limits.

Appendix B: Linear chirp signal model

We begin with the mathematical model for a discrete, complex linear-chirp signal sampled N times over a total duration T. The time-domain sequence is written as:

$$s[n] = \exp\left(i\left[\phi_0 + 2\pi f_0 \frac{T}{N}n + \pi \dot{f}\frac{T^2}{N^2}n^2\right]\right), \tag{B1}$$

where ϕ_0 is the initial phase at the start of the segment. The parameter f_0 represents the instantaneous frequency at $t = 0$ (the start of the interval), and $\dot{f}$ is the constant chirp rate. For this derivation, we assume $\dot{f} > 0$, such that the instantaneous frequency evolves as $f(t) = f_0 + \dot{f}t$ over the segment.

It is important to clarify why we specifically analyze a complex signal with unit amplitude, which can ensure that the time-domain power is constant at unit. This convention is analytically convenient because from the definition Eq. (13) the power spectrum value in a single bin $|\tilde{x}[k]|^2$, directly corresponds to our spectral leakage function, $\eta(w,o)$.

The DFT of this signal at frequency bin k is given by:

$$\begin{aligned}\tilde{s}[k] &= \frac{1}{N}\sum_{n=0}^{N-1} s[n]\, e^{-i2\pi\frac{k}{N}n} \\ &= \frac{e^{i\phi_0}}{N}\sum_{n=0}^{N-1}\exp\left\{i\left[2\pi\left(f_0T-k\right)\frac{n}{N}+\pi\dot{f}T^2\frac{n^2}{N^2}\right]\right\}.\end{aligned} \tag{B2}$$

For a sufficiently large number of samples N, where the normalized time variable $\mu = n/N$ becomes quasi-continuous, this discrete sum can be accurately approximated by its corresponding integral. By making the substitution $1/N\sum_{n=0}^{N-1} \to \int_0^1 d\mu$, we obtain:

$$\tilde{s}[k] \approx e^{i\phi_0}\int_0^1 \exp\left\{i\left[2\pi\left(f_0T-k\right)\mu+\pi\dot{f}T^2\mu^2\right]\right\}d\mu. \tag{B3}$$

This integral is of the Fresnel type. By completing the square in the exponent and performing the standard change of integration variable, the result can be expressed in terms of the complex Fresnel integral:

$$\tilde{s}[k] = \frac{1}{\sqrt{\pi\dot{f}T^2}}e^{i\phi_0-iv^2}\int_v^{v+\sqrt{\pi\dot{f}T^2}} e^{iu^2}\,du, \tag{B4}$$

where the lower integration limit v is a dimensionless quantity defined as:

$$v \equiv \sqrt{\pi}\,\frac{f_0T-k}{\sqrt{\dot{f}\,T^2}}. \tag{B5}$$

To generalize the analysis and make the spectral behavior more transparent, we introduce two convenient dimensionless quantities:

- The *widening factor*, w, which represents the dimensionless chirp strength over the segment duration:

$$w \equiv \dot{f}T^2. \tag{B6}$$

- The *offset factor*, o, which measures the frequency offset (in units of bins) between the bin frequency $f_k = k/T$ and the signal's central time frequency over the segment, $f_c = f_0 + \dot{f}T/2$:

$$\begin{aligned}o &\equiv (f_k - f_c)T \\ &= \left(\frac{k}{T} - f_0 - \frac{1}{2}\dot{f}T\right)T \\ &= (k - f_0T) - \frac{1}{2}w.\end{aligned} \tag{B7}$$

Expressed in terms of these dimensionless variables, the normalized power in a single frequency bin k, which we denote as the spectral leakage factor $\eta(w,o)$, is given by the squared modulus of the DFT coefficient:

$$\eta(o,w) \equiv \frac{1}{\pi w}\left|\int_{\sqrt{\pi}(o/\sqrt{w}-\sqrt{w}/2)}^{\sqrt{\pi}(o/\sqrt{w}+\sqrt{w}/2)} e^{iu^2}du\right|^2. \tag{B8}$$

For the case of a negative chirp rate ($\dot{f} < 0$), the result for $\eta(w,o)$ remains identical due to the symmetries of the complex exponential and the squared magnitude operation. The phase of $\tilde{s}[k]$ would change, but the power spectrum is unaffected.

It is crucial to recognize that the function $\eta(o,w)$ describes a continuous power spectrum as a function of the offset o. The power spectrum calculated from a DFT $|\tilde{s}[k]|^2$, is a discrete sampling of this continuous profile, evaluated at the specific offset values corresponding to each integer frequency bin k.

The preceding analysis was based on an implicit rectangular window, which has a constant value over the integration time T. However, in practical signal processing, other window functions (such as Hann, Hamming, or Tukey) are often applied to the time-domain data. It is therefore crucial to generalize our model for an arbitrary, power-normalized window function.

Let $W(\tau)$ be a window function defined over the normalized time interval $\tau \in [-1/2, 1/2]$, satisfying the power-normalization condition $\int_{-1/2}^{1/2}|W(\tau)|^2 d\tau = 1$. When this window is applied to the time-domain signal, the derivation for the spectral leakage factor η follows the same procedure, but with the window function carried through the integral.

After performing the same change of variables that leads to the Fresnel integral representation, the generalized spectral leakage function takes the form:

$$\eta(o,w) = \frac{1}{\pi w}\left|\int_{\sqrt{\pi}(o/\sqrt{w}-\sqrt{w}/2)}^{\sqrt{\pi}(o/\sqrt{w}+\sqrt{w}/2)} W\left(\frac{u}{\sqrt{\pi w}} - \frac{o}{w}\right)e^{iu^2}du\right|^2. \tag{B9}$$

This is a powerful and general result. It shows that the spectral power of a linear chirp signal, when windowed by an arbitrary function W, can be expressed as the squared magnitude of a generalized Fresnel integral, where the integrand is now modulated by the window function itself, mapped into the domain of the integration variable u.

Finally, we extend the analysis to a real-valued signal, which is the form encountered in physical measurements. A real-valued linear chirp with unit power can be constructed from its complex analytic representation, which we denote as $s[n]$:

$$\begin{aligned}s_r[n] &= \sqrt{2}\cos\left(\phi_0 + 2\pi f_0\frac{T}{N}n + \pi\dot{f}\frac{T^2}{N^2}n^2\right) \\ &= \frac{1}{\sqrt{2}}\left(s[n] + s[n]^*\right).\end{aligned} \tag{B10}$$

The factor of $\sqrt{2}$ ensures that the average power of the real signal is unity. Due to the linearity of the DFT, the transform of the real signal $\tilde{s}_r[k]$, can be expressed in terms of the transform of the complex signal $\tilde{s}[k]$:

$$\tilde{s}_r[k] = \frac{1}{\sqrt{2}}\left(\tilde{s}[k] + \tilde{s}[-k]^*\right). \tag{B11}$$

It is important to clarify the indexing notation $\tilde{s}[-k]^*$. For a DFT of length N, the index $-k$ is interpreted mod-

ulo N; that is, for $k \in \{0, \ldots, N-1\}$, the index corresponds to $N-1-k$. This term represents the contribution from the complex conjugate (negative frequency) component of the signal.

The power spectrum of the real signal $\eta_r = |\tilde{s}_r[k]|^2$, is therefore:

$$\eta_r(o,w) = \frac{1}{2}\left(|\tilde{s}[k]|^2 + |\tilde{s}[-k]|^2 + 2\mathrm{Re}\{\tilde{s}[k]\tilde{s}[-k]\}\right). \tag{B12}$$

Substituting our leakage function $\eta(w,o) = |\tilde{s}[k]|^2$, this becomes:

$$\eta_r(w,o) = \frac{1}{2}\eta(w,o) + \frac{1}{2}\eta(w, o+2f_cT) + \text{cross-term}. \tag{B13}$$

Here, the central time frequency of the signal is f_c, and the argument of the second term, $o + 2f_cT$, reflects the offset from the negative-frequency center at $-f_c$. This equation has a clear physical interpretation:

1. $\frac{1}{2}\eta(w,o)$: Half the power from the positive-frequency component, centered at $+f_c$.
2. $\frac{1}{2}\eta(w, o+2f_cT)$: Half the power from the negative-frequency "mirror" component, centered at $-f_c$, which leaks into the positive frequency bins.
3. cross-term: An interference term arising from the coherent sum of the positive and negative frequency components.

When the carrier frequency is well-separated from zero frequency (i.e., $f_c \gg \dot{f}T$, or in dimensionless terms, $f_cT \gg w$), the contribution from the negative-frequency image and the cross-term become negligible for positive frequency bins ($k > 0$).

$$\eta_r(w,o) \approx \frac{1}{2}\eta(w,o), \quad o > 0 \tag{B14}$$

This factor of 1/2 has a direct physical interpretation. A real-valued signal is mathematically the sum of a positive-frequency complex signal and its conjugate. The total power of the unit-power real signal is equally divided between these two components. Our leakage function $\eta(w,o)$ was defined for an idealized unit-power complex signal, which has all its power concentrated at the positive frequency.

Since the positive-frequency component of the real signal contains only half the total power, its contribution to the power spectrum is correspondingly scaled by a factor of 1/2. In essence, the power measured in the positive-frequency spectrum is approximately half that of the idealized complex signal because the other half of the power resides in the negative-frequency component, which we can safely ignore in this approximation.

Appendix C: Power-normalization of the window function

The DFT inherently applies an implicit rectangular window to the data segment. The relationship between the time-domain power and frequency-domain power is governed by Parseval's theorem, which states:

$$P = \frac{1}{N}\sum_{n=0}^{N-1}|s[n]|^2 = \sum_{k=0}^{N-1}|\tilde{s}[k]|^2. \tag{C1}$$

In practical STFT, however, various window functions, $W[n]$, are applied to the time series to mitigate spectral leakage. When a window is applied, the signal becomes $s[n] \cdot W[n]$, and Parseval's theorem must be applied to this new windowed sequence:

$$P_W = \frac{1}{N}\sum_{n=0}^{N-1}|W[n]s[n]|^2 = \sum_{k=0}^{N-1}|\tilde{s}[k]|^2. \tag{C2}$$

The application of a window function inherently alters the total power of the signal. To ensure that our power measurements remain consistent and physically meaningful regardless of the window chosen, we must adopt a specific normalization convention.

For this purpose, we define a power-normalized window. Any original window function, $W_{\mathrm{orig}}[n]$, can be converted to its power-normalized version, $W[n]$, by dividing it by its root mean square (RMS) value:

$$W[n] = \frac{W_{\mathrm{orig}}[n]}{C_{\mathrm{power}}}, \tag{C3}$$

where the normalization constant C_{power} is the RMS value of the original window:

$$C_{\mathrm{power}} = \sqrt{\frac{1}{T}\int_0^T |W_{\mathrm{orig}}(t)|^2 dt} \approx \sqrt{\frac{1}{N}\sum_{n=0}^{N-1}|W_{\mathrm{orig}}[n]|^2}. \tag{C4}$$

By its very definition, this normalization ensures that the power of the normalized window function itself is unity $\frac{1}{N}\sum|W[n]|^2 = 1$.

The benefit of this convention becomes clear when we consider the power of a physical signal. Let's assume the amplitude of our signal $s[n]$ evolves slowly and can be approximated as constant over a single short-time Fourier transform segment. If the underlying signal has an average power of P, we can approximate $|s[n]|^2 \approx P$. The power of the windowed signal then becomes:

$$P_W = \frac{1}{N}\sum_{n=0}^{N-1}|W[n]|^2|s[n]|^2 \approx P\left(\frac{1}{N}\sum_{n=0}^{N-1}|W[n]|^2\right). \tag{C5}$$

Since our power-normalized window ensures that the term in the parenthesis is equal to one, we arrive at the desired result:

$$P_W \approx P. \tag{C6}$$

This crucial result guarantees that the total power measured in the spectrum of the windowed signal P_W remains a consistent and unbiased estimator of the true average

power of the underlying physical signal P regardless of the specific shape of the window function applied.

Therefore, for the gravitational wave signal discussed in our main text, we can establish a specific, window-function-independent form for the signal's power P_i.

$$P_i \approx \frac{1}{2} h_{0,i}^2 Q_i^2. \tag{C7}$$

This provides a direct physical interpretation for the power measured in the spectrum.

Appendix D: Derivation of Critical Ratio Moments

In this appendix, we derive the analytical expressions for the mean ($\mu_{\rm CR}$) and variance ($\sigma^2_{\rm CR}$) of the Critical Ratio statistic, utilizing the sampling phase-averaged weighting scheme defined in Eq. (27).

1. Evaluation on the Real Signal Track (Perfect Match)

To simplify the exposition, we first consider the scenario of a *single-pixel track*, where for each time segment i, the generalized track $\mathcal{T}$ consists solely of the anchor bin (corresponding to the relative index $\kappa = 0$). The generalization to multi-pixel tracks is straightforward due to the linearity of the summation.

We focus first on the numerator of the expected mean $\mu_{\rm CR}$, denoted as $N_\mu = \sum_{\mathcal{T}} \omega_{i,k} \lambda[i,k]$. Substituting the weight definition $\omega_{i,k} = \mathcal{L}_i \hat{\eta}_i[\kappa]$ and the factorization $\lambda[i,k] = \mathcal{L}_i \eta_i[\kappa]$, the derivation proceeds as follows:

$$\begin{aligned} N_\mu &= \sum_{i=1}^{N} \mathcal{L}_i \hat{\eta}_i[\kappa] \left(\mathcal{L}_i \eta_i[\kappa]\right) \\ &= \sum_{i=1}^{N} \mathcal{L}_i^2 \hat{\eta}_i[\kappa] \eta_i[\kappa] \\ &= \sum_{i=1}^{N} \frac{1}{2L+1} \sum_{j=i-L}^{i+L} \mathcal{L}_j^2 \hat{\eta}_j[\kappa] \eta_j[\kappa] \\ &\approx \sum_{i=1}^{N} \frac{\mathcal{L}_i^2 \hat{\eta}_i[\kappa]}{2L+1} \sum_{j=i-L}^{i+L} \eta_j[\kappa] \\ &\approx \sum_{i=1}^{N} \mathcal{L}_i^2 \hat{\eta}_i[\kappa] \hat{\eta}_i[\kappa] = \sum_{i=1}^{N} \left(\mathcal{L}_i \hat{\eta}_i[\kappa]\right)^2. \end{aligned} \tag{D1}$$

Consequently, the mean of the Critical Ratio simplifies to:

$$\mu_{\rm CR} = \frac{1}{2} \frac{N_\mu}{\sqrt{\sum_{\mathcal{T}} \omega_{i,k}^2}} \approx \frac{1}{2} \sqrt{\sum_{\mathcal{T}} \left(\mathcal{L}_i \hat{\eta}_i[\kappa]\right)^2}. \tag{D2}$$

This derivation relies on three key mathematical steps:

- **Moving Average Identity (Line 2 → 3):** We rewrite the summation as a moving average over a window of $2L+1$ time segments. Neglecting boundary effects, this is a mathematical identity.
- **Slow-Varying Approximation (Line 3 → 4):** We assume that the macroscopic quantities—the total power statistic $\mathcal{L}_j$ and the averaged leakage factor $\hat{\eta}_j[\kappa]$—vary slowly compared to the window length L. They are treated as constant within the local window $[i-L, i+L]$ and factored out of the inner summation.
- **Ergodic Approximation (Line 4 → 5):** This is the core physical approximation. We posit that the time-average of the rapidly oscillating instantaneous leakage $\eta_j[\kappa]$ is equivalent to the phase-averaged leakage function $\hat{\eta}_i[\kappa]$ (which integrates over the offset parameter o). Mathematically, $\frac{1}{2L+1} \sum_{j=i-L}^{i+L} \eta_j[\kappa] \approx \hat{\eta}_i[\kappa]$. This assumes that the sequence of sampling phases $\{o_{0,j}\}$ provides a *dense and uniform sampling* of the continuous offset interval $[-0.5, 0.5]$.

Following an analogous logic, the variance $\sigma^2_{\rm CR}$ is derived by expanding its numerator $N_\sigma = \sum_{\mathcal{T}} \omega_{i,k}^2 \lambda[i,k]$:

$$\begin{aligned} N_\sigma &= \sum_{i=1}^{N} \left(\mathcal{L}_i \hat{\eta}_i[\kappa]\right)^2 \left(\mathcal{L}_i \eta_i[\kappa]\right) \\ &= \sum_{i=1}^{N} \frac{1}{2L+1} \sum_{j=i-L}^{i+L} \mathcal{L}_j^3 (\hat{\eta}_j[\kappa])^2 \eta_j[\kappa] \\ &\approx \sum_{i=1}^{N} \frac{\mathcal{L}_i^3 (\hat{\eta}_i[\kappa])^2}{2L+1} \sum_{j=i-L}^{i+L} \eta_j[\kappa] \\ &\approx \sum_{i=1}^{N} \mathcal{L}_i^3 (\hat{\eta}_i[\kappa])^2 \hat{\eta}_i[\kappa] = \sum_{i=1}^{N} \left(\mathcal{L}_i \hat{\eta}_i[\kappa]\right)^3. \end{aligned} \tag{D3}$$

Substituting this into the general variance formula yields:

$$\sigma^2_{\rm CR} = 1 + \frac{\sum_{\mathcal{T}} (\mathcal{L}_i \hat{\eta}_i[\kappa])^3}{\sum_{\mathcal{T}} (\mathcal{L}_i \hat{\eta}_i[\kappa])^2}. \tag{D4}$$

This result highlights the dependence of variance on signal strength. In the *weak-signal limit* ($\mathcal{L}_i \ll 1$), the second term vanishes, leading to $\sigma^2_{\rm CR} \approx 1$, consistent with the null hypothesis. For multi-pixel configurations, the derivation follows an identical logic, with the summation set $\mathcal{T}$ simply expanded to include the relevant sideband bins.

2. Rationale for the 3-Pixel Track

To determine the optimal spatial extent of the generalized track $\mathcal{T}$, we analyze the contribution of each relative frequency bin κ to the total detection sensitivity. Recall from Eq. (32) that the maximum detectable distance

scales with the square root of the collected power term. We expand the summation over the generalized track into contributions from the central anchor bin ($\kappa = 0$) and symmetric sideband pairs ($\kappa = \pm1, \pm2, \dots$):

$$\begin{aligned}\sum_{(i,k)\in\mathcal{T}} (\mathcal{L}_i \hat{\eta}_i[\kappa])^2 &= \sum_{i=1}^{N} \mathcal{L}_i^2 \sum_{\kappa} \hat{\eta}_i^2[\kappa] \\ &= \sum_{i=1}^{N} \mathcal{L}_i^2 \left(\underbrace{\hat{\eta}_i^2[0]}_{\text{Anchor Bin}} + \underbrace{\hat{\eta}_i^2[1] + \hat{\eta}_i^2[-1]}_{\text{1st Sidebands}} + \underbrace{\hat{\eta}_i^2[2] + \hat{\eta}_i^2[-2]}_{\text{2nd Sidebands}} + \cdots \right) \\ &\approx \sum_{i=1}^{N} \mathcal{L}_i^2 \underbrace{\left(\hat{\eta}^2(0, w_i) + 2\hat{\eta}^2(1, w_i) + 2\hat{\eta}^2(2, w_i) + \cdots \right)}_{\equiv \Psi(w_i)}.\end{aligned} \tag{D5}$$

Here, we invoke the *slowly-varying approximation*. We assume that the expected noise power spectral density varies negligibly across the narrow frequency bandwidth (typically spanning only a few bins) where the spectral leakage function $\hat{\eta}$ exhibits significant structure. Consequently, the term $\mathcal{L}_i$ is treated as effectively constant with respect to the relative frequency index κ and can be factored out of the summation. Proceeding with this simplification, and utilizing the symmetry of the averaged leakage function, $\hat{\eta}(o, w) = \hat{\eta}(-o, w)$, we introduce the *spectral efficiency factor*, $\Psi(w)$:

$$\Psi(w) \equiv \hat{\eta}^2(0, w) + 2\hat{\eta}^2(1, w) + \cdots . \tag{D6}$$

This factor serves as a fundamental metric for the efficiency of signal power recovery in the presence of spectral leakage. It is crucial to distinguish between energy conservation and statistical sensitivity:

- **Energy Conservation (Linear Sum):** The spectral leakage function itself is normalized, implying that the total signal energy is conserved regardless of how it is distributed across the frequency bins ($\sum_\kappa \hat{\eta}[\kappa] = 1$).
- **Statistical Dilution (Squared Sum):** However, the detection statistical significance scales with the *sum of squares* of these factors. As the signal power disperses (smears) into more sidebands due to an increasing widening factor w, the sum of squares inevitably drops below unity ($\Psi(w) < 1$).

Therefore, $\Psi(w)$ represents the *effective collection efficiency*. It reflects the inevitable penalty incurred by spreading signal power across multiple noise-filled frequency bins. Even if a search algorithm were to sum the signal over an infinite bandwidth to recover the power, the simultaneous accumulation of noise variance from these additional bins results in a net loss of statistical significance compared to a perfectly concentrated monochromatic signal (where $\Psi = 1$).

Practical search algorithms must truncate this summation to a finite track width. The decision to restrict the analysis to the anchor bin and its immediate neighbors ($|\kappa| \leq 1$, i.e., a 3-pixel track) is justified by the rapid spectral decay characteristic of standard window functions (e.g., Tukey or Hann). As illustrated in Fig. 3, the signal energy is highly localized: the central anchor bin ($\kappa = 0$) captures the dominant fraction of the power, while the first sidebands ($\kappa = \pm1$) account for the vast majority of the remaining leakage. Consequently, extending the track width beyond 3 pixels yields *diminishing returns*. The contribution from higher-order sidebands ($|\kappa| \geq 2$) provides a negligible improvement in the efficiency factor $\Psi(w)$, whereas including them linearly increases the computational overhead (specifically, memory bandwidth and storage). Thus, the 3-pixel configuration represents the optimal trade-off between maximizing signal recovery fidelity and maintaining computational efficiency.

3. Generalization to Template-Signal Mismatch

The preceding analysis assumed a perfect alignment where the integer bin index κ is defined relative to the true signal frequency. In a practical search utilizing a template bank, a continuous frequency offset inevitably exists between the template track, $f_{t,i}$, and the true signal track, $f_{r,i}$.

Let $f_{c,i}$ denote the instantaneous central frequency of the signal for the i-th segment. The normalized frequency offset between the true signal and the template, in units of frequency bins ($f_{\text{bin}} = 1/T_{\text{DFT}}$), is given by:

$$o_i = \frac{f_{r,i} - f_{t,i}}{f_{\text{bin}}}. \tag{D7}$$

To analyze this scenario, we distinguish between quantities derived from the *template* parameters (subscript t) and those intrinsic to the *real* signal (subscript r).

The detection weights, $\omega_{i,k}$, are pre-calculated based on the template trajectory:

$$(\omega_{i,k})_t \propto (\mathcal{L}_i)_t (\hat{\eta}_i[\kappa])_t, \tag{D8}$$

where $(\hat{\eta}_i[\kappa])_t$ is the averaged leakage factor evaluated at the template's integer bin offset κ.

The actual signal contribution, $\lambda[i,k]$, however, depends on the *true* signal's leakage into the bins defined by the template grid. The received power is governed by the continuous offset between the bin center and the real signal frequency. This is expressed as:

$$(\lambda[i,k])_r = (\mathcal{L}_i)_r(\eta(o_i,w_i))_r. \tag{D9}$$

Here, we conceptualize the signal contribution in terms of the mismatch offset o_i. Substituting these definitions into the fundamental moments, the generalized mean of the CR becomes:

$$\mu_{\rm CR} = \frac{1}{2}\frac{\sum_{\mathcal{T}}(\mathcal{L}_i\hat{\eta}_i[\kappa])_t\cdot(\mathcal{L}_i\hat{\eta}(o_i,w_i))_r}{\sqrt{\sum_{\mathcal{T}}(\mathcal{L}_i\hat{\eta}_i[\kappa])_t^2}}. \tag{D10}$$

Similarly, the variance generalizes to:

$$\sigma^2_{\rm CR} = 1 + \frac{\sum_{\mathcal{T}}(\mathcal{L}_i\hat{\eta}_i[\kappa])_t^2\cdot(\mathcal{L}_i\hat{\eta}(o_i,w_i))_r}{\sum_{\mathcal{T}}(\mathcal{L}_i\hat{\eta}_i[\kappa])_t^2}. \tag{D11}$$

These generalized expressions quantify the search performance for any arbitrary offset o_i. They serve as the theoretical basis for calculating the Fitting Factor (FF) and Mismatch (MM), enabling the rigorous construction of a template bank that guarantees high detection probability across the continuous parameter space.

Appendix E: Scaling Laws and the Maximum Detectable Distance

A critical characteristic of the proposed detection statistic is the scaling behavior of its moments—the mean ($\mu_{\rm CR}$) and variance ($\sigma^2_{\rm CR}$)—with respect to the number of analyzed time segments, N. This scaling directly governs how the search sensitivity improves with accumulating observation time.

To elucidate this behavior, we rewrite the cumulative sums over the track set $\mathcal{T}$ in terms of segment-averaged quantities. Let $\langle X\rangle_N \equiv \frac{1}{N}\sum_{i=1}^N X_i$ denote the arithmetic mean of a quantity over the N segments.

First, consider the mean of the Critical Ratio. Factoring out N, the expression becomes:

$$\mu_{\rm CR} = \frac{1}{2}\sqrt{\sum_{\mathcal{T}}(\mathcal{L}_i\hat{\eta}_i[\kappa])^2} = \frac{1}{2}\sqrt{N\left\langle(\mathcal{L}_i\hat{\eta}_i[\kappa])^2\right\rangle_N}. \tag{E1}$$

Assuming the average signal power per segment remains roughly constant over the observation, this implies that the expected statistical significance grows with the square root of the number of segments:

$$\mu_{\rm CR}\propto\sqrt{N}. \tag{E2}$$

This $\sqrt{N}$ growth is the hallmark of semi-coherent integration strategies (often referred to as incoherent summation of coherent powers), reflecting the stochastic accumulation of signal power.

In contrast, the variance of the CR statistic exhibits a fundamentally different scaling behavior. Expressing the sums in terms of averages yields:

$$\begin{aligned}\sigma^2_{\rm CR} &= 1 + \frac{\sum_{\mathcal{T}}(\mathcal{L}_i\hat{\eta}_i[\kappa])^3}{\sum_{\mathcal{T}}(\mathcal{L}_i\hat{\eta}_i[\kappa])^2}\\ &= 1 + \frac{N\left\langle(\mathcal{L}_i\hat{\eta}_i[\kappa])^3\right\rangle_N}{N\left\langle(\mathcal{L}_i\hat{\eta}_i[\kappa])^2\right\rangle_N}\\ &= 1 + \frac{\left\langle(\mathcal{L}_i\hat{\eta}_i[\kappa])^3\right\rangle_N}{\left\langle(\mathcal{L}_i\hat{\eta}_i[\kappa])^2\right\rangle_N}.\end{aligned} \tag{E3}$$

Crucially, the factor of N cancels out in the fractional term. Consequently, the variance is an intensive quantity with respect to observation time. For weak signals (where the higher-order terms are small), $\sigma^2_{\rm CR}$ remains close to unity, effectively independent of the observation duration.

This divergence in scaling behaviors—the unbounded growth of the mean ($\propto\sqrt{N}$) versus the asymptotic constancy of the variance ($\propto N^0$)—is the engine of the search's sensitivity. As the observation time N increases, the signal distribution shifts progressively away from the noise background without broadening significantly, thereby exponentially suppressing the false dismissal probability.

Finally, we connect this result to the astrophysical horizon. From the definition of the total power statistic, we know that $\mu_{\rm CR}$ scales quadratically with distance: $d\propto\sqrt{\mu_{\rm CR}}$. Combining this with Eq. (E2), we arrive at the fundamental scaling law for the maximum detectable distance:

$$d_{\rm max}\propto N^{1/4}. \tag{E4}$$

This fourth-root scaling law confirms that extending the observation duration is a robust strategy for expanding the search volume ($V\propto d^3\propto N^{3/4}$) for mini-EMRI systems.

Appendix F: Effective Detection Volume and Averaged Antenna Response

Eq. (32) defines the maximum detectable distance $d_{\rm max}$ under a specific set of extrinsic parameters (sky location and polarization). However, the astrophysical performance of a search pipeline is best characterized by its sensitivity to a population of sources distributed isotropically throughout the universe. To quantify this, we transition from the horizon distance to the *effective detection volume* $V_{\rm eff}$. This necessitates averaging the detection probability over all extrinsic variables: sky position (α,δ) and source orientation (ψ,ι).

The dependence on these geometric parameters is encapsulated within the detector's antenna pattern function, $Q(t;\alpha,\delta,\psi,\iota)$. The averaging procedure begins by exploiting the separation of timescales. Since the intrinsic signal parameters (frequency and amplitude) evolve on timescales much longer than the Earth's rotation period ($T_\oplus \approx 23.93\,\mathrm{h}$), we can perform a time-average over one sidereal day. This operation effectively decouples the rapid antenna pattern modulation from the secular signal evolution. A key consequence of this averaging is that the dependence on the source's right ascension (α) becomes degenerate, reducing the effective sky-dependence to a function of declination (δ) only.

Formally, the effective volume is constructed by integrating the detectable volume element over the full parameter space. We first define the volume for a fixed source orientation, $V(\psi,\iota)$, by integrating over the celestial sphere:

$$\begin{aligned} V(\psi,\iota) &= \int_0^{2\pi} d\alpha \int_{-1}^{1} d(\cos\delta) \int_0^{d_{\max}(\delta,\psi,\iota)} r^2 dr \\ &= \frac{1}{3}\int_0^{2\pi} d\alpha \int_{-1}^{1} d(\cos\delta)\, d^3_{\max}(\delta,\psi,\iota). \end{aligned} \tag{F1}$$

The final effective detection volume, $V_{\rm eff}$, is obtained by averaging this quantity over all possible polarization angles ψ and inclination angles ι:

$$V_{\rm eff} = \frac{1}{4\pi}\int_0^{2\pi} d\psi \int_{-1}^{1} d(\cos\iota)\, V(\psi,\iota). \tag{F2}$$

Combining these integrals yields the complete expression for the effective detection volume:

$$V_{\rm eff} = \frac{1}{12\pi}\int_0^{2\pi} d\psi \int_{-1}^{1} d(\cos\iota) \int_0^{2\pi} d\alpha \int_{-1}^{1} d(\cos\delta)\, d^3_{\max}(\delta,\psi,\iota). \tag{F3}$$

This scalar metric provides a robust quantification of the pipeline's global sensitivity. It is often convenient to express this volume in terms of an *effective distance*, $d_{\rm eff}$, defined as the radius of a Euclidean sphere with volume $V_{\rm eff}$:

$$d_{\rm eff} = \left(\frac{3V_{\rm eff}}{4\pi}\right)^{1/3}. \tag{F4}$$

While Eq. (F3) is exact, it is computationally intensive. We can simplify it conceptually and practically by encapsulating the entire geometric averaging process into a single, scalar factor: the *fully-averaged antenna pattern factor*, denoted as $\langle Q\rangle$. This factor represents the detector's effective response to an isotropic source population. For the ΣR statistic (where sensitivity scales as the fourth root of power accumulation, see Appendix E), this factor is defined as:

$$\langle Q\rangle_{\Sigma R} = \left\langle \langle Q(t)^4\rangle_t^{3/4} \right\rangle_{\rm sky,orient}^{1/3}. \tag{F5}$$

Numerical evaluation of this integral yields the following characteristic values for the current detector network:

$$\langle Q\rangle_{\Sigma R}^{\rm LIGO\ H} \approx 0.4590, \quad \langle Q\rangle_{\Sigma R}^{\rm LIGO\ L} \approx 0.4604, \quad \langle Q\rangle_{\Sigma R}^{\rm Virgo} \approx 0.4597. \tag{F6}$$

By factoring out this geometric term, the effective distance can be expressed in a modular form that separates geometry from signal strength:

$$d_{\rm eff} = \langle Q\rangle_{\Sigma R}\left[\sqrt{2}\,\mathrm{erfc}^{-1}(2P_{\rm fa}) + \sqrt{2}\,\mathrm{erfc}^{-1}(2P_{\rm fd})\right]^{-1/2}\left[\frac{1}{4}\sum_{\mathcal{T}} h_{e,i}^4 \frac{T_{\rm DFT}^2}{S_{n,i}^2}\hat{\eta}_i^2[\kappa]\right]^{1/4}. \tag{F7}$$

For comparison, the conventional Hough transform pipeline exhibits a different sensitivity scaling due to its binary thresholding nature. The corresponding averaged factor for the Hough transform is derived as:

$$\langle Q\rangle_{\rm Hough} = \left\langle \langle Q(t)^2\rangle_t^{3/2} \right\rangle_{\rm sky,orient}^{1/3}. \tag{F8}$$

Numerical integration yields effective response factors for the Hough pipeline: $\langle Q\rangle_{\rm Hough}^{\rm LIGO\ H} \approx 0.4207$, $\langle Q\rangle_{\rm Hough}^{\rm LIGO\ L} \approx 0.4191$, and $\langle Q\rangle_{\rm Hough}^{\rm Virgo} \approx 0.4198$.

Appendix G: Frequency layering strategy

To rigorously determine the optimal frequency layering configuration, we first establish a continuous metric for the accumulation of statistical significance. We observe

that the squared expectation of the Critical Ratio, $\mu^2_{\rm CR}$, behaves as an additive quantity. For a search trajectory $\mathcal{T}$ composed of disjoint sub-tracks $\mathcal{T}_1$ and $\mathcal{T}_2$, the total significance satisfies:

$$\mu^2_{\rm CR}(\mathcal{T}_1+\mathcal{T}_2)=\mu^2_{\rm CR}(\mathcal{T}_1)+\mu^2_{\rm CR}(\mathcal{T}_2). \tag{G1}$$

Substituting the definition of $\mu_{\rm CR}$ from Eq. (28) and expanding the terms for the 3-pixel track configuration, the total squared significance over N segments is:

$$\begin{aligned}\mu^2_{\rm CR}&=\frac{1}{4}\sum_{(i,k)\in\mathcal{T}}\left(\mathcal{L}_i\hat{\eta}_i[\kappa]\right)^2\\&=\frac{1}{4}\sum_{i=1}^{N}\mathcal{L}_i^2\left(\hat{\eta}_i^2[0]+2\hat{\eta}_i^2[1]\right)\\&=\frac{1}{4}\sum_{i=1}^{N}h_{0,i}^4Q_i^4\frac{T_{\rm DFT}^2}{S_{n,i}^2}\left[\hat{\eta}^2(0,w_i)+2\hat{\eta}^2(1,w_i)\right].\end{aligned} \tag{G2}$$

Here, we assume the standard condition of 50% temporal overlap between adjacent segments. In the **continuum limit**, the summation over discrete segments can be approximated by a time integral, where the segment density is given by $dN/dt=(T_{\rm DFT}/2)^{-1}$. Applying the fully-averaged antenna pattern factor $\langle Q\rangle_{\Sigma R}$, the expression transforms to:

$$\begin{aligned}\mu^2_{\rm CR}&\approx\frac{1}{4}\langle Q\rangle^4_{\Sigma R}\int\frac{h_0^4(t)T_{\rm DFT}^2}{S_n^2(f(t))}\Psi(w(t))\frac{2}{T_{\rm DFT}}dt\\&=2\langle Q\rangle^4_{\Sigma R}\int h_0^4(f(t))\frac{T_{\rm DFT}}{S_n^2(f(t))}\Psi(w(t))\,dt,\end{aligned} \tag{G3}$$

where $\Psi(w)\equiv\hat{\eta}^2(0,w)+2\hat{\eta}^2(1,w)$ represents the *spectral efficiency factor* for the default 3-pixel track configuration, as formally defined in Eq. (D6).

Finally, to facilitate optimization in the frequency domain, we perform a change of variables from time t to frequency f. Using the stationary phase relationship $dt=(df/dt)^{-1}df=\dot{f}^{-1}df$, we arrive at the spectral representation of the sensitivity:

$$\mu^2_{\rm CR}=\frac{1}{4}\langle Q\rangle^4_{\Sigma R}\int_{f_{\rm min}}^{f_{\rm max}}h_0^4(f)\frac{T_{\rm DFT}}{S_n^2(f)}\Psi\left(\dot{f}T_{\rm DFT}^2\right)\dot{f}^{-1}\,df. \tag{G4}$$

Based on this integral form, we can formally define the *Sensitivity Density* $\rho_{\rm CR}(f,T_{\rm DFT})$, which quantifies the contribution of each unit frequency interval to the total squared detection statistic:

$$\begin{aligned}\rho_{\rm CR}(f,T_{\rm DFT})&\equiv\frac{d(\mu^2_{\rm CR})}{df}\\&=\frac{1}{4}\langle Q\rangle^4_{\Sigma R}\,h_0^4(f)\frac{T_{\rm DFT}}{S_n^2(f)}\Psi\left(\dot{f}T_{\rm DFT}^2\right)\dot{f}^{-1}.\end{aligned} \tag{G5}$$

This density function reveals the fundamental trade-off: increasing $T_{\rm DFT}$ linearly increases the SNR accumulation (the linear $T_{\rm DFT}$ term), but eventually suppresses the efficiency factor $\Psi(w)$ as the widening $w=\dot{f}T_{\rm DFT}^2$ grows large. The optimal frequency layering strategy is thus equivalent to finding the function $T_{\rm DFT}(f)$ that maximizes this density at every frequency f.

For a monotonically evolving signal (where the frequency $f(t)$ is a strictly increasing function of time), the total time-frequency trajectory can be uniquely partitioned into N disjoint frequency layers. Consider a search configuration partitioned into N distinct frequency layers. This segmentation is defined by a set of $N-1$ critical frequencies, denoted as $\{f_{c,1},f_{c,2},\ldots,f_{c,N-1}\}$, which divide the total bandwidth $[f_{\rm min},f_{\rm max}]$ into contiguous intervals. Within the n-th layer, a specific coherent integration time T_n is applied.

The total squared statistical significance is obtained by summing the integral contributions from each layer:

$$\begin{aligned}\mu^2_{\rm CR}&=\sum_{n=1}^{N}\mu^2_{\rm CR}(\text{layer }n)\\&=\int_{f_{\rm min}}^{f_{c,1}}\rho_{\rm CR}(f,T_1)\,df+\int_{f_{c,1}}^{f_{c,2}}\rho_{\rm CR}(f,T_2)\,df+\cdots\\&\quad+\int_{f_{c,N-1}}^{f_{\rm max}}\rho_{\rm CR}(f,T_N)\,df.\end{aligned} \tag{G6}$$

This formulation transforms the optimization problem into finding the optimal set of transition frequencies $\{f_{c,n}\}$ and coherent times $\{T_n\}$ that maximize this global sum.

It is straightforward to show that the optimal critical frequency, $f_{c,n}$, separating two adjacent layers (with integration times T_n and T_{n+1}) is determined *solely* by the signal's frequency derivative $\dot{f}$.

The optimal transition point occurs where the sensitivity densities of the two configurations intersect, i.e., where switching from T_n to T_{n+1} yields no discontinuities in sensitivity:

$$\rho_{\rm CR}(f_{c,n},T_n)=\rho_{\rm CR}(f_{c,n},T_{n+1}). \tag{G7}$$

Substituting the definition of $\rho_{\rm CR}$ from Eq. (G5), we observe that the terms depending on the signal strength $h_0(f)$ and the detector noise $S_n(f)$ appear as common factors on both sides and explicitly cancel out. The condition simplifies to a transcendental equation involving only the integration times and the frequency derivative:

$$T_n\Psi\left(\dot{f}(f_{c,n})T_n^2\right)=T_{n+1}\Psi\left(\dot{f}(f_{c,n})T_{n+1}^2\right). \tag{G8}$$

This result reveals a fundamental universality in the search design: the optimal frequency layering is governed exclusively by the interplay between the coherent gain (linear in T) and the spectral leakage (encapsulated in Ψ), independent of the specific spectral shape of the noise or the source.

Appendix H: Statistical Independence and Pixel Correlations

In the derivation of the ΣR statistic (Eq. (23)), we posited the simplifying assumption that the power ratios $R[i,k]$ behave as mutually independent random variables across the entire time-frequency plane. While this hypothesis affords analytical tractability, it represents an idealization. In practice, the inherent mathematical structure of the STFT inevitably induces non-vanishing statistical correlations between pixels (i,k) and (j,l) with sufficiently small separation. These correlations arise from two primary mechanisms:

1. **Spectral Correlations (Frequency Axis):** The application of a window function $W[n]$ imposes an amplitude modulation on the time-domain data. In the frequency domain, this manifests as spectral leakage: the windowing process convolves the noise spectrum with the window's kernel, thereby smearing noise power across adjacent bins.
2. **Temporal Correlations (Time Axis):** These correlations arise primarily from the use of overlapping time segments (e.g., 50% overlap), where adjacent STFT columns share common time-domain data samples. Furthermore, even in the absence of physical overlap, the intrinsic temporal structure of the detector noise (i.e., its non-white nature and finite correlation length) introduces residual statistical coupling between consecutive time segments.

To rigorously characterize these effects, we construct the covariance structure of the spectrogram. For stationary Gaussian noise, the correlation in the time-frequency plane is mathematically governed by the properties of the window function and the segment overlap.

First, we examine the correlation between the complex DFT spectrum of the i-th and j-th time segments, denoted as $\tilde{x}_i[k]$ and $\tilde{x}_j[l]$. The expectation of their cross-product is:

$$\begin{aligned}\mathbb{E}\big[\tilde{x}_i[k]\tilde{x}_j^*[l]\big] &= \mathbb{E}\left[\left(\frac{1}{N}\sum_{n=0}^{N-1} W[n]x_i[n]e^{-i\frac{2\pi}{N}kn}\right)\left(\frac{1}{N}\sum_{m=0}^{N-1} W[m]x_j[m]e^{i\frac{2\pi}{N}lm}\right)^*\right] \\ &= \frac{1}{N^2}\sum_{n=0}^{N-1}\sum_{m=0}^{N-1} W[n]W[m]\mathbb{E}\big[x_i[n]x_j[m]\big]e^{-i\frac{2\pi}{N}(kn-lm)}.\end{aligned} \tag{H1}$$

Let $R_{ij}[n,m] \equiv \mathbb{E}[x_i[n]x_j[m]]$ denote the cross-correlation matrix of the time-domain noise between segments i and j. We further define a windowing mask matrix, $M[n,m] \equiv W[n]W[m]$. We can define two auxiliary spectral correlation matrices: C_{ij} (conjugate covariance) and D_{ij} (non-conjugate covariance/relation):

$$\begin{aligned}C_{ij}[k,l] &\equiv \mathbb{E}\big[\tilde{x}_i[k]\tilde{x}_j^*[l]\big], \\ D_{ij}[k,l] &\equiv \mathbb{E}\big[\tilde{x}_i[k]\tilde{x}_j[l]\big].\end{aligned} \tag{H2}$$

These terms can be elegantly expressed as spectrum of the 2D DFT of the windowed time-domain correlation matrix:

$$\begin{aligned}C_{ij}[k,l] &= \mathrm{DFT2}\left\{M \odot R_{ij}\right\}[k,-l], \\ D_{ij}[k,l] &= \mathrm{DFT2}\left\{M \odot R_{ij}\right\}[k,l],\end{aligned} \tag{H3}$$

where $\odot$ denotes the element-wise (Hadamard) product:

$$(M \odot R_{ij})[n,m] = M[n,m]\cdot R_{ij}[n,m]. \tag{H4}$$

We now derive the covariance of the power ratio R across distinct time-frequency bins. We first consider the case of pure noise. Our goal is to calculate the covariance $\mathrm{Cov}(R[i,k],R[j,l])$.

By definition, the covariance of the normalized variables is scaled by their respective normalization factors:

$$\mathrm{Cov}\left(R[i,k],R[j,l]\right) = \frac{\mathrm{Cov}\left(|\tilde{n}_i[k]|^2,|\tilde{n}_j[l]|^2\right)}{\langle|\tilde{n}_i[k]|^2\rangle\langle|\tilde{n}_j[l]|^2\rangle}. \tag{H5}$$

To evaluate the numerator, we recall the definition $\mathrm{Cov}(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$. We invoke *Wick's Theorem* for zero-mean complex Gaussian variables to decompose the fourth-order moment $\mathbb{E}[|\tilde{n}_i|^2|\tilde{n}_j|^2]$ into the sum of pairwise expectations:

$$\begin{aligned}\mathbb{E}\left[|\tilde{n}_i|^2|\tilde{n}_j|^2\right] &= \mathbb{E}[|\tilde{n}_i|^2]\mathbb{E}[|\tilde{n}_j|^2] \\ &\quad + \left|\mathbb{E}(\tilde{n}_i\tilde{n}_j^*)\right|^2 + \left|\mathbb{E}(\tilde{n}_i\tilde{n}_j)\right|^2.\end{aligned} \tag{H6}$$

When substituting this expansion back into the covariance expression, the first term on the right-hand side ($\mathbb{E}[|\tilde{n}_i|^2]\mathbb{E}[|\tilde{n}_j|^2]$) exactly cancels with the product of the means. This leaves only the cross-correlation terms. Using the C_{ij} and D_{ij} notation defined previously, we arrive at the final generalized covariance expression:

$$\mathrm{Cov}\left(R[i,k],R[j,l]\right) = \frac{|C_{ij}[k,l]|^2 + |D_{ij}[k,l]|^2}{\langle|\tilde{n}_i[k]|^2\rangle\langle|\tilde{n}_j[l]|^2\rangle}. \tag{H7}$$

The above derivation assumes pure noise. In the presence of a signal ($\tilde{s} \neq 0$), R follows a *non-central* distribution. While exact analytical expressions for the covariance in the signal-present regime can be derived following

an analogous methodological framework—specifically by expanding the fourth-order moments of the non-central variables—the resulting formulae are algebraically prohibitive to display. They involve multiple cross-terms representing the interaction between the deterministic signal amplitude and the stochastic noise background. Consequently, we omit these explicit forms for the sake of brevity. However, fundamentally, the correlation structure remains governed by the same windowing and overlap mechanisms identified in the noise-only analysis.

Finally, we explicitly quantify the impact of these correlations on the statistics of the aggregate metric ΣR. While the linearity of the expectation operator ensures that the mean $\mu_{\Sigma R}$ remains unbiased regardless of correlations, the variance $\sigma^2_{\Sigma R}$ requires a correction term to account for the covariance between contributing pixels:

$$\begin{aligned}\sigma^2_{\Sigma R} &= \sum_{(i,k)\in\mathcal{T}} \omega^2_{i,k}\sigma^2_R[i,k] + \sum_{\text{distinct pairs}} 2\omega_{i,k}\omega_{j,l}\mathrm{Cov}(R[i,k],R[j,l]) \\ &= \sigma^2_{\Sigma R,0} + \Delta\sigma^2_{\text{corr}}.\end{aligned} \tag{H8}$$

Here, the second summation runs over all distinct pairs of pixels in the track set. The factor of 2 explicitly accounts for the symmetry of the covariance matrix ($\mathrm{Cov}(X,Y) = \mathrm{Cov}(Y,X)$). The term $\sigma^2_{\Sigma R,0}$ denotes the variance derived under the independence assumption, while the second term captures the cumulative contribution of cross-correlations. Since the windowing-induced correlations are predominantly positive, this correction term leads to *variance inflation* (i.e., $\sigma^2_{\Sigma R} > \sigma^2_{\Sigma R,0}$).

In the regime of weak correlations (e.g., standard 50% temporal overlap), this inflation is minor. Consequently, for the calculation of the maximum detectable distance $d_{\max}$, where we adopt the approximation $\sigma_{\mathrm{CR}} \approx 1$, the impact of this correction is second-order and can be safely neglected. However, Eq. (H8) serves as a crucial cautionary principle: it dictates that one cannot arbitrarily enhance sensitivity by simply increasing the density of STFT segments within a fixed observation duration. Excessive temporal overlap (e.g., $\sim 100\%$) would drastically increase the number of correlated cross-terms. This would cause the background variance to grow faster than the signal accumulation, leading to diminishing returns or even a degradation in the statistical significance.

[1] Z.-X. Wang, G. Cheng, J. Chen, H.-K. Guo, and A. L. Miller, Methods for Detecting Gravitational Waves from mini-Extreme-Mass-Ratio Inspirals I: Statistics Based on Time-Frequency Signal Tracks (2025), arXiv:2512.21738 [gr-qc].

[2] B. P. Abbott *et al.* (LIGO Scientific, Virgo), Observation of Gravitational Waves from a Binary Black Hole Merger, Phys. Rev. Lett. **116**, 061102 (2016), arXiv:1602.03837 [gr-qc].

[3] J. Aasi *et al.* (LIGO Scientific), Advanced LIGO, Class. Quant. Grav. **32**, 074001 (2015), arXiv:1411.4547 [gr-qc].

[4] F. Acernese *et al.* (VIRGO), Advanced Virgo: a second-generation interferometric gravitational wave detector, Class. Quant. Grav. **32**, 024001 (2015), arXiv:1408.3978 [gr-qc].

[5] T. Akutsu *et al.* (KAGRA), KAGRA: 2.5 Generation Interferometric Gravitational Wave Detector, Nature Astron. **3**, 35 (2019), arXiv:1811.08079 [gr-qc].

[6] R. Abbott *et al.* (KAGRA, VIRGO, LIGO Scientific), GWTC-3: Compact Binary Coalescences Observed by LIGO and Virgo during the Second Part of the Third Observing Run, Phys. Rev. X **13**, 041039 (2023), arXiv:2111.03606 [gr-qc].

[7] A. G. Abac *et al.* (LIGO Scientific, VIRGO, KAGRA), GWTC-4.0: Updating the Gravitational-Wave Transient Catalog with Observations from the First Part of the Fourth LIGO-Virgo-KAGRA Observing Run (2025), arXiv:2508.18082 [gr-qc].

[8] A. H. Nitz, C. D. Capano, S. Kumar, Y.-F. Wang, S. Kastha, M. Schäfer, R. Dhurkunde, and M. Cabero, 3-OGC: Catalog of Gravitational Waves from Compact-binary Mergers, Astrophys. J. **922**, 76 (2021), arXiv:2105.09151 [astro-ph.HE].

[9] A. H. Nitz, S. Kumar, Y.-F. Wang, S. Kastha, S. Wu, M. Schäfer, R. Dhurkunde, and C. D. Capano, 4-OGC: Catalog of Gravitational Waves from Compact Binary Mergers, Astrophys. J. **946**, 59 (2023), arXiv:2112.06878 [astro-ph.HE].

[10] K. Belczynski, T. Bulik, C. L. Fryer, A. Ruiter, F. Valsecchi, J. S. Vink, and J. R. Hurley, On the maximum mass of stellar black holes, The Astrophysical Journal **714**, 1217–1226 (2010).

[11] R. Abbott *et al.* (LIGO Scientific, Virgo), GW190521: A Binary Black Hole Merger with a Total Mass of $150M_\odot$, Phys. Rev. Lett. **125**, 101102 (2020), arXiv:2009.01075 [gr-qc].

[12] B. P. Abbott *et al.* (LIGO Scientific, Virgo), GW190425: Observation of a Compact Binary Coalescence with Total Mass $\sim 3.4M_\odot$, Astrophys. J. Lett. **892**, L3 (2020), arXiv:2001.01761 [astro-ph.HE].

[13] R. Abbott *et al.* (LIGO Scientific, Virgo), GW190814: Gravitational Waves from the Coalescence of a 23 Solar Mass Black Hole with a 2.6 Solar Mass Compact Object, Astrophys. J. Lett. **896**, L44 (2020), arXiv:2006.12611

[astro-ph.HE].
[14] R. Abbott *et al.* (LIGO Scientific, KAGRA, VIRGO), Observation of Gravitational Waves from Two Neutron Star–Black Hole Coalescences, Astrophys. J. Lett. **915**, L5 (2021), arXiv:2106.15163 [astro-ph.HE].
[15] W. M. Farr, N. Sravan, A. Cantrell, L. Kreidberg, C. D. Bailyn, I. Mandel, and V. Kalogera, The mass distribution of stellar-mass black holes, The Astrophysical Journal **741**, 103 (2011).
[16] S. F. Portegies Zwart and S. McMillan, Black hole mergers in the universe, Astrophys. J. Lett. **528**, L17 (2000), arXiv:astro-ph/9910061.
[17] C. L. Rodriguez, M. Morscher, B. Pattabiraman, S. Chatterjee, C.-J. Haster, and F. A. Rasio, Binary Black Hole Mergers from Globular Clusters: Implications for Advanced LIGO, Phys. Rev. Lett. **115**, 051101 (2015), [Erratum: Phys.Rev.Lett. 116, 029901 (2016)], arXiv:1505.00792 [astro-ph.HE].
[18] D. Gerosa and M. Fishbach, Hierarchical mergers of stellar-mass black holes and their gravitational-wave signatures, Nature Astron. **5**, 749 (2021), arXiv:2105.03439 [astro-ph.HE].
[19] I. Mandel and F. S. Broekgaarden, Rates of compact object coalescences, Living Rev. Rel. **25**, 1 (2022), arXiv:2107.14239 [astro-ph.HE].
[20] B. P. Abbott *et al.* (LIGO Scientific Collaboration, Virgo), Search for Subsolar-Mass Ultracompact Binaries in Advanced LIGO's First Observing Run, Phys. Rev. Lett. **121**, 231103 (2018), arXiv:1808.04771 [astro-ph.CO].
[21] B. P. Abbott *et al.* (LIGO Scientific, Virgo), Search for Subsolar Mass Ultracompact Binaries in Advanced LIGO's Second Observing Run, Phys. Rev. Lett. **123**, 161102 (2019), arXiv:1904.08976 [astro-ph.CO].
[22] D. Singh, M. Ryan, R. Magee, T. Akhter, S. Shandera, D. Jeong, and C. Hanna, Gravitational-wave limit on the Chandrasekhar mass of dark matter, Phys. Rev. D **104**, 044015 (2021), arXiv:2009.05209 [astro-ph.CO].
[23] A. H. Nitz and Y.-F. Wang, Search for Gravitational Waves from High-Mass-Ratio Compact-Binary Mergers of Stellar Mass and Subsolar Mass Black Holes, Phys. Rev. Lett. **126**, 021103 (2021), arXiv:2007.03583 [astro-ph.HE].
[24] R. Abbott *et al.* (LIGO Scientific Collaboration, Virgo, KAGRA), Search for Subsolar-Mass Binaries in the First Half of Advanced LIGO's and Advanced Virgo's Third Observing Run, Phys. Rev. Lett. **129**, 061104 (2022), arXiv:2109.12197 [astro-ph.CO].
[25] K. S. Phukon *et al.*, The hunt for sub-solar primordial black holes in low mass ratio binaries is open (2021), arXiv:2105.11449 [astro-ph.CO].
[26] A. H. Nitz and Y.-F. Wang, Search for gravitational waves from the coalescence of subsolar mass and eccentric compact binaries, The Astrophysical Journal **915**, 54 (2021), arXiv:2102.00868.
[27] A. H. Nitz and Y.-F. Wang, Search for Gravitational Waves from the Coalescence of Subsolar-Mass Binaries in the First Half of Advanced LIGO and Virgo's Third Observing Run, Phys. Rev. Lett. **127**, 151101 (2021), arXiv:2106.08979 [astro-ph.HE].
[28] R. Abbott *et al.* (LVK), Search for subsolar-mass black hole binaries in the second part of Advanced LIGO's and Advanced Virgo's third observing run, Mon. Not. Roy. Astron. Soc. **524**, 5984 (2023), [Erratum: Mon.Not.Roy.Astron.Soc. 526, 6234 (2023)], arXiv:2212.01477 [astro-ph.HE].
[29] R. Abbott *et al.* (LIGO Scientific, VIRGO, KAGRA), Search for subsolar-mass black hole binaries in the second part of Advanced LIGO's and Advanced Virgo's third observing run, Mon. Not. Roy. Astron. Soc. **524**, 5984 (2023), [Erratum: Mon.Not.Roy.Astron.Soc. 526, 6234 (2023)], arXiv:2212.01477 [astro-ph.HE].
[30] A. H. Nitz and Y.-F. Wang, Broad search for gravitational waves from subsolar-mass binaries through LIGO and Virgo's third observing run, Phys. Rev. D **106**, 023024 (2022), arXiv:2202.11024 [astro-ph.HE].
[31] A. L. Miller, N. Aggarwal, S. Clesse, F. De Lillo, S. Sachdev, P. Astone, C. Palomba, O. J. Piccinni, and L. Pierini, Gravitational Wave Constraints on Planetary-Mass Primordial Black Holes Using LIGO O3a Data, Phys. Rev. Lett. **133**, 111401 (2024), arXiv:2402.19468 [gr-qc].
[32] A. G. Abac *et al.* (LIGO Scientific Collaboration, VIRGO, KAGRA), Search for planetary-mass ultracompact binaries using data from the first part of the LIGO–Virgo–KAGRA fourth observing run (2025), arXiv:2511.19911 [gr-qc].
[33] S. Hawking, Gravitationally collapsed objects of very low mass, Mon. Not. Roy. Astron. Soc. **152**, 75 (1971).
[34] M. Y. Khlopov, Primordial Black Holes, Res. Astron. Astrophys. **10**, 495 (2010), arXiv:0801.0116 [astro-ph].
[35] B. Carr, F. Kuhnel, and M. Sandstad, Primordial Black Holes as Dark Matter, Phys. Rev. D **94**, 083504 (2016), arXiv:1607.06077 [astro-ph.CO].
[36] B. Carr, S. Clesse, J. García-Bellido, and F. Kühnel, Cosmic conundra explained by thermal history and primordial black holes, Phys. Dark Univ. **31**, 100755 (2021), arXiv:1906.08217 [astro-ph.CO].
[37] A. Escrivà, F. Kuhnel, and Y. Tada, Primordial Black Holes (2022), arXiv:2211.05767 [astro-ph.CO].
[38] S. L. Liebling and C. Palenzuela, Dynamical boson stars, Living Rev. Rel. **26**, 1 (2023), arXiv:1202.5809 [gr-qc].
[39] H.-K. Guo and A. Miller, Searching for Mini Extreme Mass Ratio Inspirals with Gravitational-Wave Detectors (2022), arXiv:2205.10359 [astro-ph.IM].
[40] M. Maggiore, *Gravitational waves* (Oxford University Press, Oxford, 2008) oCLC: ocn180464569.
[41] S. A. Usman *et al.*, The PyCBC search for gravitational waves from compact binary coalescence, Class. Quant. Grav. **33**, 215004 (2016), arXiv:1508.02357 [gr-qc].
[42] P. Amaro-Seoane *et al.* (LISA), Laser Interferometer Space Antenna (2017), arXiv:1702.00786 [astro-ph.IM].
[43] C. P. L. Berry, S. A. Hughes, C. F. Sopuerta, A. J. K. Chua, A. Heffernan, K. Holley-Bockelmann, D. P. Mihaylov, M. C. Miller, and A. Sesana, The unique potential of extreme mass-ratio inspirals for gravitational-wave astronomy, Bull. Am. Astron. Soc. **51**, 42 (2019), arXiv:1903.03686 [astro-ph.HE].
[44] J. R. Gair, Probing black holes at low redshift using LISA EMRI observations, Class. Quant. Grav. **26**, 094034 (2009), arXiv:0811.0188 [gr-qc].
[45] J. R. Gair, S. Babak, A. Sesana, P. Amaro-Seoane, E. Barausse, C. P. L. Berry, E. Berti, and C. Sopuerta, Prospects for observing extreme-mass-ratio inspirals with LISA, J. Phys. Conf. Ser. **840**, 012021 (2017), arXiv:1704.00009 [astro-ph.GA].
[46] W.-R. Hu and Y.-L. Wu, The Taiji Program in Space for gravitational wave physics and the nature of gravity,

Natl. Sci. Rev. **4**, 685 (2017).
[47] W.-H. Ruan, Z.-K. Guo, R.-G. Cai, and Y.-Z. Zhang, Taiji program: Gravitational-wave sources, Int. J. Mod. Phys. A **35**, 2050075 (2020), arXiv:1807.09495 [gr-qc].
[48] Y.-L. Wu, Hyperunified field theory and Taiji program in space for GWD, Int. J. Mod. Phys. A **33**, 1844014 (2018), arXiv:1805.10119 [physics.gen-ph].
[49] J. Luo *et al.* (TianQin), TianQin: a space-borne gravitational wave detector, Class. Quant. Grav. **33**, 035010 (2016), arXiv:1512.02076 [astro-ph.IM].
[50] J. Mei *et al.* (TianQin), The TianQin project: current progress on science and technology, PTEP **2021**, 05A107 (2021), arXiv:2008.10332 [gr-qc].
[51] J. Luo *et al.*, The first round result from the TianQin-1 satellite, Class. Quant. Grav. **37**, 185013 (2020), arXiv:2008.09534 [physics.ins-det].
[52] B. J. Owen, Search templates for gravitational waves from inspiraling binaries: Choice of template spacing, Phys. Rev. D **53**, 6749 (1996), arXiv:gr-qc/9511032.
[53] P. R. Brady, T. Creighton, C. Cutler, and B. F. Schutz, Searching for periodic sources with LIGO, Phys. Rev. D **57**, 2101 (1998), arXiv:gr-qc/9702050.
[54] S. Walsh *et al.*, Comparison of methods for the detection of gravitational waves from unknown neutron stars, Phys. Rev. D **94**, 124010 (2016), arXiv:1606.00660 [gr-qc].
[55] D. Keitel, Robust semicoherent searches for continuous gravitational waves with noise and signal models including hours to days long transients, Phys. Rev. D **93**, 084024 (2016), arXiv:1509.02398 [gr-qc].
[56] A. H. Nitz and Y.-F. Wang, Search for gravitational waves from the coalescence of subsolar-mass binaries in the first half of advanced ligo and virgo's third observing run, Phys. Rev. Lett. **127**, 151101 (2021).
[57] B. Krishnan, A. M. Sintes, M. A. Papa, B. F. Schutz, S. Frasca, and C. Palomba, Hough transform search for continuous gravitational waves, Phys. Rev. D **70**, 082001 (2004).
[58] P. Astone, A. Colla, S. D'Antonio, S. Frasca, and C. Palomba, Method for all-sky searches of continuous gravitational wave signals using the frequency-hough transform, Physical Review D **90**, 10.1103/physrevd.90.042002 (2014).
[59] O. J. Piccinni, P. Astone, S. D'Antonio, S. Frasca, G. Intini, P. Leaci, S. Mastrogiovanni, A. Miller, C. Palomba, and A. Singhal, A new data analysis framework for the search of continuous gravitational wave signals, Class. Quant. Grav. **36**, 015008 (2019), arXiv:1811.04730 [gr-qc].
[60] P. Astone, S. Frasca, and C. Palomba, The short fft database and the peak map for the hierarchical search of periodic sources, Classical and Quantum Gravity **22**, S1197 (2005).
[61] A. Miller *et al.*, Method to search for long duration gravitational wave transients from isolated neutron stars using the generalized frequency-Hough transform, Phys. Rev. D **98**, 102004 (2018), arXiv:1810.09784 [astro-ph.IM].
[62] A. L. Miller, N. Aggarwal, S. Clesse, F. De Lillo, S. Sachdev, P. Astone, C. Palomba, O. J. Piccinni, and L. Pierini, Method to search for inspiraling planetary-mass ultracompact binaries using the generalized frequency-Hough transform in LIGO O3a data, Phys. Rev. D **110**, 082004 (2024), arXiv:2407.17052 [astro-ph.IM].
[63] S. S. Menon *et al.*, GFH-v2 Pipeline for Searches of Long-Transient Gravitational Waves from Newborn Magnetars (2025), arXiv:2512.09878 [astro-ph.IM].
[64] A. L. Miller and L. Pierini, BinaryGFH-v2: Improved method to search for gravitational waves from sub-solar-mass, ultra-compact binaries using the Generalized Frequency-Hough Transform (2025), arXiv:2512.10539 [gr-qc].
[65] R. Tenorio, J.-R. Mérou, and A. M. Sintes, One-stop strategy to search for long-duration gravitational-wave signals, Phys. Rev. D **111**, 104002 (2025), arXiv:2411.18370 [gr-qc].
[66] J.-R. Mérou, R. Tenorio, and A. M. Sintes, GPU-Accelerated Searches for Long-Transient Gravitational Waves from Newborn Neutron Stars (2025) arXiv:2507.07816 [gr-qc].
[67] B. Abbott *et al.* (LIGO Scientific), All-sky search for periodic gravitational waves in LIGO S4 data, Phys. Rev. D **77**, 022001 (2008), [Erratum: Phys.Rev.D 80, 129904 (2009)], arXiv:0708.3818 [gr-qc].
[68] P. Jaranowski, A. Krolak, and B. F. Schutz, Data analysis of gravitational - wave signals from spinning neutron stars. 1. The Signal and its detection, Phys. Rev. D **58**, 063001 (1998), arXiv:gr-qc/9804014.
[69] K. Riles, Searches for continuous-wave gravitational radiation, Living Rev. Rel. **26**, 3 (2023), arXiv:2206.06447 [astro-ph.HE].
[70] L. S. Finn and K. S. Thorne, Gravitational waves from a compact star in a circular, inspiral orbit, in the equatorial plane of a massive, spinning black hole, as observed by LISA, Phys. Rev. D **62**, 124021 (2000), arXiv:gr-qc/0007074.
[71] S. M. Kay, Modern spectral estimation: Theory and application (1988).
[72] R. Tenorio, M. J. Williams, J. Bayley, C. Messenger, M. Demkin, W. Reade, J. Koda, Y. Yamakawa, T. Yamaguchi, K. Abe, C. Achard, H. S. T. Bukhari, M. V. Shugaev, G. Sokolov, V. Debout, S. Goulet, J.-L. Tastet, I. Timiryasov, O. Ruchayskiy, D. Kanonik, S. Seferbekov, S. Saito, R. Sato, S. Segawa, A. Zhyvalkouski, Y. Uchida, S. Yokoi, A. Sayed, R.-Q. Xing, I. Yamashita, and Z. Wang, Learning to detect continuous gravitational waves: an open data-analysis competition, Machine Learning: Science and Technology **6**, 040702 (2025).
[73] R. Tenorio and D. Gerosa, Scalable data-analysis framework for long-duration gravitational waves from compact binaries using short Fourier transforms, Phys. Rev. D **111**, 104044 (2025), arXiv:2502.11823 [gr-qc].
[74] C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. van Kerkwijk, M. Brett, A. Haldane, J. F. del Río, M. Wiebe, P. Peterson, P. Gérard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke, and T. E. Oliphant, Array programming with NumPy, Nature **585**, 357 (2020).
[75] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, İ. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pe-

dregosa, P. van Mulbregt, and SciPy 1.0 Contributors, SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python, Nature Methods **17**, 261 (2020).

[76] J. D. Hunter, Matplotlib: A 2d graphics environment, Computing in Science & Engineering **9**, 90 (2007).