All-sky search for continuous gravitational waves from isolated neutron stars using Advanced LIGO and Advanced Virgo O3 data

The LIGO Scientific Collaboration, The Virgo Collaboration, and The KAGRA Collaboration∗
(Dated: June 17, 2022)

We present results of an all-sky search for continuous gravitational waves which can be produced by spinning neutron stars with an asymmetry around their rotation axis, using data from the third observing run of the Advanced LIGO and Advanced Virgo detectors. Four different analysis methods are used to search in a gravitational-wave frequency band from 10 to 2048 Hz and a first frequency derivative from $-10^{-8}$ to $10^{-5}$ Hz/s. No statistically-significant periodic gravitational-wave signal is observed by any of the four searches. As a result, upper limits on the gravitational-wave strain amplitude $h_0$ are calculated. The best upper limits are obtained in the frequency range of 100 to 200 Hz and they are $\sim 1.1 \times 10^{-25}$ at 95% confidence-level. The minimum upper limit of $1.10 \times 10^{-25}$ is achieved at a frequency 111.5 Hz. We also place constraints on the rates and abundances of nearby planetary- and asteroid-mass primordial black holes that could give rise to continuous gravitational-wave signals.

I. INTRODUCTION

The Advanced LIGO [1] and Advanced Virgo [2] detectors have made numerous detections of gravitational waves (GW), to date consisting of short-duration (transient) GW emitted during the inspirals and mergers of compact binary systems of black holes (BH), neutron stars (NS), [3, 4], as well as mixed NS-BH binaries [5]. Among still undiscovered types of GW radiation are long-lasting, almost-monochromatic continuous waves (CW), whose amplitudes and frequencies change much more slowly compared to those of transient sources (on the timescale of years rather than seconds). Astrophysically, promising sources of CW are rotating, non-axisymmetric NS, emitting GW at a frequency close to, or related to, their spin frequency. Deviations from the symmetry (a NS ‘deformation’) may be caused by fluid instabilities, such as in the case of r-modes, or by elastic, thermal or magnetic stresses in the crust and/or core of NS, and may be acquired at various stages of stars’ isolated evolution, or during an interaction with a companion in a binary system (for recent reviews on sources of CW, see e.g., [6–8]). Discovery of CW emitted by NS would allow to probe their still mysterious interiors, study properties of dense matter in conditions distinct from those occurring in inspirals and mergers of binary NS systems, as well as carry out additional tests of the theory of gravity [9]. Due to intrinsically smaller GW amplitude of CW in comparison to the already-detected transient sources, searches for CW from rotating non-axisymmetric NS are essentially limited to the Galaxy.

The search presented here is not limited to gravitational-wave signals from deformed rotating neutron stars. Another source of quasi-monochromatic, persistent GWs are very light, planetary- and asteroid-mass, inspiraling primordial black holes (PBHs), which could comprise a fraction or the totality of dark matter [10]. Such signals would arise from inspiraling PBHs whose chirp masses are less than $O(10^{-5}) M_\odot$ and whose GW frequencies are less than $\sim 250$ Hz, and would be indistinguishable from those arising from non-axisymmetric rotating NSs spinning up.

Recent detections of black holes made by the LIGO-Virgo-KAGRA Collaboration have revived interest in PBHs: low spin measurements and the rate inferences are consistent with those expected for BHs that formed in the early universe [11]. Existence of light PBHs is well-motivated theoretically and experimentally: recent detections of star and quasar microlensing events [12–14] suggest compact objects or PBHs with masses between $10^{-6}$ and $10^{-5} M_\odot$ could constitute a fraction of dark matter of order $f_{PBH} \sim 0.01$, which is consistent within the unified scenario for PBH formation presented in [15], but greater than expected for free-floating (i.e. not bound to an orbit) planets [16] (e.g. the hypothetical Planet 9 could be a PBH with a mass of $10^{-6} M_\odot$ that was captured by the solar system [17]). PBHs may also collide with NS and be responsible for the origin of NS-mass BHs, potentially detectable in the LIGO-Virgo-KAGRA searches [18]. However, constraints arising from such observations [10], even those that come from the LIGO-Virgo merging rate inferences [19, 20] and stochastic background searches [21, 22], rely on modelling assumptions, and can be evaded if, for example, PBHs formed in clusters [23–28]. It is therefore important to develop complementary probes of these mass regimes to test different PBH formation models [29, 30], which is possible by searching for continuous GWs.

Searches for continuous waves are usually split in three different domains: targeted searches look for signals from known pulsars; directed searches look for signals from known sky locations; all-sky searches look for signals from unknown sources. All-sky searches for a priori unknown CW sources have been carried out in the Advanced LIGO and Advanced Virgo data previously [31–33]. A recent review on pipelines for wide parameter-space searches can be found in [44].

Here we report on results from an all-sky, broad fre-
frequency range search using the most-sensitive data to date, the LIGO-Virgo O3 observing run, employing four different search pipelines: the FrequencyHough [45], SkyHough [46], Time-Domain $F$-statistic [47, 48], and SOAP [49]. Each pipeline uses different data analysis methods and covers different regions of the frequency and frequency time derivative parameter space, although there exist overlaps between them (see Table I and Fig. 1 for details). The search is performed for frequencies between 10 Hz and 2048 Hz and for a range of frequency time derivative between $-10^{-8}$ Hz/s and $10^{-6}$ Hz/s, covering the whole sky. We note here that the search is generally-agnostic to the type of the GW source, so the results are not actually limited to signals from non-axisymmetric rotating NS in our Galaxy. A comprehensive multi-stage analysis of the signal outliers obtained by the four pipelines has not revealed any viable candidate for a continuous GW signal. However we improve the broad-range frequency upper limits with respect to previous O1 and O2 observing run and also with respect to the recent analysis of the first half of the O3 run [39]. This is also the first all-sky search for CW sources that uses the Advanced Virgo detector’s data.

The article is organized as follows: in Section II we describe the O3 observing run and provide details about the data used. Section III we present an overview of the pipelines used in the search. Section IV, details of the data-analysis pipelines are described. Section V, we describe the results obtained by each pipeline, namely the signal candidates and the sensitivity of the search whereas Section VI contains a discussion of the astrophysical implications of our results.

II. DATA SETS USED

The data set used in this analysis was the third observing run (O3) of the Advanced LIGO and Advanced Virgo GW detectors [1, 2]. LIGO is made up of two laser interferometers, both with 4 km long arms. One is at the LIGO Livingston Observatory (L1) in Louisiana, USA and the other is at the LIGO Hanford Observatory (H1) in Washington, USA. Virgo (V1) consists of one interferometer with 3 km arms located at European Gravitational Observatory (EGO) in Cascina, Italy. The O3 run took place between the 2019 April 1 and the 2020 March 27. The run was divided into two parts, O3a and O3b, separated by one month commissioning break that took place in October 2019. The duty factors for this run were $\sim 76\%$, $\sim 71\%$, $\sim 76\%$ for L1, H1, V1 respectively. The maximum uncertainties (68% confidence interval) on the calibration of the LIGO data were of 7%/11% in magnitude and 4 deg/9 deg in phase for O3a/O3b data ([50, 51]). For Virgo, it amounted to 5% in amplitude and 2 deg in phase, with the exception of the band 46 - 51 Hz, for which the maximum uncertainty was estimated as 40% in amplitude and 34 deg in phase during O3b. For the smaller range 49.5 - 50.5 Hz, the calibration was unreliable during the whole run [52].

III. OVERVIEW OF SEARCH PIPELINES

In this section we provide a broad overview of the four pipelines used in the search. The three pipelines: FrequencyHough, SkyHough, and Time-Domain $F$-statistic have been used before in several all-sky searches of the LIGO data. The SOAP pipeline is a new pipeline applied for the first time to an all-sky search. It uses novel algorithms. SOAP aims at a fast, preliminary search of the data before more sensitive but much more time consuming methods are applied (see [44] for a review on pipelines for wide parameter-space searches). The individual pipelines are described in more detail in the following section.

A. Signal model

The GW signal in the detector frame from an isolated, asymmetric NS spinning around one of its principal axis of inertia is given by [47]:

$$h(t) = h_0[F_+(t, \alpha, \delta, \psi) \frac{1 + \cos^2 t}{2} \cos \phi(t) + F_{\times}(t, \alpha, \delta, \psi) \cos t \sin \phi(t)],$$

(1)

where $F_+$ and $F_{\times}$ are the antenna patterns of the detectors dependent on right ascension $\alpha$, declination $\delta$ of the source and polarization angle $\psi$, $h_0$ is the amplitude of the signal, $t$ is the angle between the total angular momentum vector of the star and the direction from the star to the Earth, and $\phi(t)$ is the phase of the signal. The amplitude of the signal is given by:

$$h_0 = \frac{4\pi^2 G e I_{zz} f^2}{c^4 d} \approx 1.06 \times 10^{-26} \left(\frac{e}{10^{-6}}\right) \left(\frac{I_{zz}}{10^{48} \text{ kg m}^2}\right) \left(\frac{f}{100 \text{ Hz}}\right)^2 \left(\frac{1 \text{ kpc}}{d}\right),$$

(2)

where $d$ is the distance from the detector to the source, $f$ is the GW frequency (assumed to be twice the rotation frequency of the NS), $e$ is the ellipticity or asymmetry of the star, given by $(I_{xx} - I_{yy})/I_{zz}$, and $I_{zz}$ is the moment of inertia of the star with respect to the principal axis aligned with the rotation axis.

We assume that the phase evolution of the GW signal can be approximated with a second order Taylor expansion around a fiducial reference time $\tau_r$:

$$\phi(\tau) = \phi_0 + 2\pi[f(\tau - \tau_r) + \frac{f}{2!}(\tau - \tau_r)^2],$$

(3)

where $\phi_0$ is an initial phase and $f$ and $\dot{f}$ are the frequency and first frequency derivative at the reference time. The
relation between the time at the source \( \tau \) and the time at the detector \( t \) is given by:

\[
\tau(t) = t + \frac{\vec{r}(t) \cdot \vec{n}}{c} + \Delta E_\odot - \Delta S_\odot ,
\]

where \( \vec{r}(t) \) is the position vector of the detector in the Solar System Barycenter (SSB) frame, and \( \vec{n} \) is the unit vector pointing to the NS; \( \Delta E_\odot \) and \( \Delta S_\odot \) are respectively the relativistic Einstein and Shapiro time delays. In standard equatorial coordinates with right ascension \( \alpha \) and declination \( \delta \), the components of the unit vector \( \vec{n} \) are given by \((\cos \alpha \cos \delta, \sin \alpha \cos \delta, \sin \delta)\).

### B. Parameter space analyzed

All the four pipelines perform an all-sky search, however the frequency and frequency derivative ranges analyzed are different for each pipeline. The detailed ranges analyzed by the four pipelines are summarized in Table I and presented in Fig. 1. The FrequencyHough pipeline analyzes a broad frequency range between 10 Hz and 2048 Hz and a broad frequency time derivative range between \(-10^{-8}\) Hz/s and \(10^{-9}\) Hz/s. A very similar range of \( f \) and \( \dot{f} \) is analyzed by SOAP pipeline. The SkyHough pipeline analyzes a narrower frequency range where the detectors are most sensitive whereas Time-Domain \( F \)-statistic pipeline analyzes \( f \) and \( \dot{f} \) ranges of the bulk of the observed pulsar population (see Fig. 2 in Sect. IV C).

### C. Detection statistics

As all-sky searches cover a large parameter space they are computationally very expensive and it is computationally prohibitive to analyze coherently the data from the full observing run using optimal matched-filtering. As a result each of the pipelines developed for the analysis uses a semi-coherent method. Moreover to reduce the computer memory and to parallelize the searches the data are divided into narrow bands. Each analysis begins with sets of short Fourier transforms (SFTs) that span the observation period, with coherence times ranging from 1024s to 8192s. The FrequencyHough, SkyHough and SOAP pipelines compute measures of strain power directly from the SFTs and create detection statistics by stacking those powers with corrections for frequency evolution applied. The FrequencyHough and SkyHough pipelines use Hough transform to do the stacking whereas SOAP pipeline uses the Viterbi algorithm. The Time-Domain \( F \)-statistic pipeline extracts band-limited 6-day long time-domain data segments from the SFT sets and applies frequency evolution corrections coherently to obtain the \( F \)-statistic ([47]). Coincidences are then required among multiple data segments with no stacking.

### D. Outlier follow-up

All four pipelines perform a follow-up analysis of the statistically significant candidates (outliers) obtained during the search. All pipelines perform vetoing of the outliers corresponding to narrow, instrumental artifacts (lines) in the advanced LIGO detectors ([53]). Several other consistency vetoes are also applied to eliminate outliers. The FrequencyHough, SkyHough, and Time-Domain \( F \)-statistic pipelines perform follow-up of the candidates by processing the data with increasing long coherence times whereas SOAP pipeline use convolutional neural networks to do the post processing.

### E. Upper limits

No periodic gravitational wave signals were observed by any of the four pipelines and all the pipelines obtain upper limits on their strength. The three pipelines SkyHough, Time-Domain \( F \)-statistic and SOAP obtain the upper limits by injections of the signals according to the model given in Section III A above for an array of signal amplitudes \( h_0 \) and randomly choosing the remaining parameters. The FrequencyHough pipeline obtains upper limits using an analytic formula (see Eq. 6) that depends on the spectral density of the noise of the detector. The formula was validated by a number of tests consisting of injecting signals to the data.
The frequency bin width is the inverse of the time duration, in seconds of the data chunks on which the FFT is computed. The time duration $T_{\text{obs}}$ is the total run duration.

TABLE II. Properties of the FFTs used in the FrequencyHough pipeline. The time duration $T_{\text{FFT}}$ refers to the length in seconds of the data chunks on which the FFT is computed. The frequency bin width is the inverse of the time duration, while the spin-down bin width is computed as $\delta f = \delta f/T_{\text{obs}}$, where $T_{\text{obs}}$ is the total run duration.

<table>
<thead>
<tr>
<th>Band [Hz]</th>
<th>$T_{\text{FFT}}$ [s]</th>
<th>$\delta f$ [Hz]</th>
<th>$\delta f$ [Hz/s]</th>
</tr>
</thead>
<tbody>
<tr>
<td>10–128</td>
<td>1.22 $\times 10^{-4}$</td>
<td>3.92 $\times 10^{-12}$</td>
<td></td>
</tr>
<tr>
<td>128–512</td>
<td>4.4 $\times 10^{-4}$</td>
<td>7.83 $\times 10^{-12}$</td>
<td></td>
</tr>
<tr>
<td>512–1024</td>
<td>4.88 $\times 10^{-4}$</td>
<td>1.57 $\times 10^{-11}$</td>
<td></td>
</tr>
<tr>
<td>1024–2048</td>
<td>9.76 $\times 10^{-4}$</td>
<td>3.13 $\times 10^{-11}$</td>
<td></td>
</tr>
</tbody>
</table>

IV. DETAILS OF SEARCH METHODS

A. FrequencyHough

The FrequencyHough pipeline is a semi-coherent procedure in which interesting points (i.e. outliers) are selected in the signal parameter space, and then are followed-up in order to confirm or reject them. This method has been used in several past all-sky searches of Virgo and LIGO data [31, 34, 35, 54]. A detailed description of the methodology can be found in [45]. In the following, we briefly describe the main analysis steps and specific choices used in the search.

Calibrated detector data are used to build “short duration” and cleaned [55] Fast Fourier Transform (FFTs), with duration $T_{\text{FFT}}$ which depends on the frequency band, as shown in Table II, and are defined, respectively, as

$$\delta f = 1/T_{\text{FFT}}$$

and

$$\delta \dot{f} = \dot{f}/T_{\text{obs}},$$

where $T_{\text{obs}}$ is the total run duration.

The peakmap is cleaned of the strongest disturbances using a line persistency veto [45].

The time-frequency peaks of the peakmap are properly shifted, for each sky position, to compensate for the Doppler effect due to the detector motion [45]. The shifted peaks are then fed to the FrequencyHough algorithm [45], which transforms each peak to the frequency/spin-down plane of the source. The frequency and spin-down bins (which we will refer to as coarse bins in the following) depend on the frequency band, as shown in Table II, and are defined, respectively, as

$$\delta f = 1/T_{\text{FFT}}$$

and

$$\delta \dot{f} = \dot{f}/T_{\text{obs}},$$

where $T_{\text{obs}}$ is the total run duration.

In practice, the nominal frequency resolution has been increased by a factor of 10 [45], as the FrequencyHough is not computationally bounded by the width of the frequency bin. The algorithm, moreover, adaptively weights any noise non-stationarity and the time-varying detector response [56].

The whole analysis is split into tens of thousands of independent jobs, each of which covers a small portion of the parameter space. Moreover, for frequencies above 512 Hz a GPU-optimized implementation of the FrequencyHough transform has been used [57].

The output of a FrequencyHough transform is a 2-D histogram in the frequency/spin-down plane of the source.

Outliers, that is significant points in this plane, are selected by dividing each 1 Hz band of the corresponding histogram into 20 intervals and taking, for each interval, and for each sky location, the one or (in most cases) two candidates with the highest histogram number count [45]. All the steps described so far are applied separately to the data of each detector involved in the analysis.

As in past analyses [31, 34], candidates from each detector are clustered and then coincident candidates among the clusters of a pair of detectors are found using a distance metric $d_{\text{FH}}$ built in the four-dimensional parameter space of sky position $(\lambda, \beta)$ (in ecliptic coordinates), frequency $f$ and spin-down $\dot{f}$. Pairs of candidates with distance $d_{\text{FH}} \leq 3$ are considered coincident. In the current O3 analysis, coincidences have been done only among pairs of candidates of the two detectors, and $\delta f$, $\delta \dot{f}$, $\delta \lambda$, and $\delta \beta$ are the corresponding bin widths.

1 The metric is defined as

$$d_{\text{FH}} = \sqrt{\left(\frac{\Delta f}{\delta f}\right)^2 + \left(\frac{\Delta \dot{f}}{\delta \dot{f}}\right)^2 + \left(\frac{\Delta \lambda}{\delta \lambda}\right)^2 + \left(\frac{\Delta \beta}{\delta \beta}\right)^2},$$

where $\Delta f$, $\Delta \dot{f}$, $\Delta \lambda$, and $\Delta \beta$ are the differences, for each parameter, among pairs of candidates of the two detectors, and $\delta f$, $\delta \dot{f}$, $\delta \lambda$, and $\delta \beta$ are the corresponding bin widths.
among the two LIGO detectors for frequencies above 128 Hz, while also coincidences H1 - Virgo and L1 - Virgo have been considered for frequencies below 128 Hz, where the difference in sensitivity (especially in the very low frequency band) is less pronounced.

Coincident candidates are ranked according to the value of a statistic built using the distance and the FrequencyHough histogram weighted number count of the coincident candidates [45]. After the ranking, the eight outliers in each 0.1 Hz band with the highest values of the statistic are selected and subject to the follow-up.

1. Follow-up

The FrequencyHough follow-up runs on each outlier of each coincident pair. It is based on the construction of a new peakmap, over ±3 coarse bins around the frequency of the outlier, with a longer $T_{FFT}$. This new peakmap is built after the removal of the signal frequency variation due to the Doppler effect for a source located at the outlier sky position. A new refined grid on the sky is built around this point, covering ±3 coarse bins, in order to take into account the uncertainty on the outlier parameters. For each point of this grid we remove the residual Doppler shift from the peakmap by properly shifting the frequency peaks. Each new corrected peakmap is the input for the FrequencyHough transform to explore the frequency and the spin-down range of interest (±3 coarse bins for the frequency and the spin-down). The most significant peak among all the FrequencyHough histograms, characterized by a set of refined parameters, is selected and subject to further post-processing steps.

First, the significance veto (V1) is applied. It consists in building a new peakmap over 0.2 Hz around the outlier refined frequency, after correcting the data with its refined parameters. The corrected peakmap is then projected on the frequency axis. Its frequency range is divided in sub-bands, each covering ±2 coarse frequency bins. The maximum of the projection in the sub-band containing the outlier is compared with the maxima selected in the remaining off-source intervals. The outlier is kept if it ranks as first or second for both detectors. Second, a noise line veto (V2) is used, which discards outliers whose frequency, after the removal of the Doppler and spin-down corrections, overlaps a band polluted by known instrumental disturbances.

The consistency test (V3) discards pairs of coincident outliers if their Critical Ratios (CRs), properly weighted by the detector noise level, differ by more than a factor of 5. The CR is defined as

$$CR = \frac{x - \mu}{\sigma},$$  \hspace{1cm} (5)$$

where $x$ is the value of the peakmap projection in a given frequency bin, $\mu$ is the average value and $\sigma$ the standard deviation of the peakmap projection.

The distance veto (V4) consists in removing pairs of coincident outliers with distance $d_{FH} > 6$ after the follow-up. Finally, outliers with distance $d_{FH} < 3$ from hardware injections are also vetoed (V5). Outliers which survive all these vetoes are scrutinized more deeply, by applying a further follow-up step, based on the same procedures just described, but further increasing the segment duration $T_{FFT}$.

2. Parameter space

The FrequencyHough search covers the frequency range [10, 2048] Hz, a spin-down range between $-10^{-8}$ Hz/s to $10^{-9}$ Hz/s and the whole sky. The frequency and spin-down resolutions are given in Tab. II. The sky resolution, on the other hand, is a function of the frequency and of the sky position and is defined in such a way that for two nearby sky cells the maximum frequency variation, due to the Doppler effect, is within one frequency bin, see [45] for more details.

3. Upper limits

“Population average” upper limits are computed for every 1 Hz sub-band in the range of 20–2048 Hz, considering only the LIGO detectors, as Virgo sensitivity is worse for most of the analyzed frequency band. First, for each detector we use the analytical relation [45]

$$h_{UL,95\%} \approx \frac{4.97}{N^{1/4}} \sqrt{\frac{S_n(f)}{T_{FFT}}} \sqrt{CR_{max} + 1.6449},$$  \hspace{1cm} (6)$$

where $N$ is the actual number of data segments used in the analysis, $S_n(f)$ is the detector average noise power spectrum, computed through a weighted mean over time segments of duration $T_{FFT}$ (in order to take into account noise non-stationarity), and $CR_{max}$ is the maximum outlier CR, in the given 1 Hz band. For each 1 Hz band, the final upper limit is the worse among those computed separately for Hanford and Livingston. Such upper limits implicitly assume an average over the source population parameters. In order to compute upper limits which hold for specific source parameters, a scaling factor must be applied as discussed in the Appendix.

As verified through a detailed comparison based on LIGO and Virgo O2 and O3 data, this procedure produces conservative upper limits with respect to those obtained through the injection of simulated signals, which is computationally much heavier [58].

---

4 Although the search starts at 10 Hz, we decided to compute upper limit starting from 20 Hz, due to the unreliable calibration at lower frequency.

5 Defined by Eq. 5 and where in this case the various quantities are computed over the Frequency-Hough map.
Moreover, it has been shown that the upper limits obtained through injections are always above those based on Eq. 6 when the minimum CR in each 1 Hz sub-band is used. The two curves based, respectively, on the highest and the smallest CR delimit a region containing both a more stringent upper limit estimate and the search sensitivity estimate, that is the minimum strain of a detectable signal. Any astrophysical implication of our results, discussed in Sec. V will be always based on the most conservative estimate.

B. SkyHough

SkyHough [46, 59] is a semicoherent pipeline based on the Hough transform to look for CW signals from isolated neutron stars. Several versions of this pipeline have been used throughout the initial [60, 61] and advanced [31, 32] detector era, as well as to look for different kinds of signals such as CW from neutron stars in binary systems [40, 41, 62] or long-duration GW transients [63]. The current implementation of SkyHough closely follows that of [32] and includes an improved suite of post-processing and follow-up stages [64–66].

1. Parameter space

The SkyHough pipeline searches over the standard four parameters describing a CW signal from isolated NS: frequency $f$, spin-down $\dot{f}$ and sky position, parametrized using equatorial coordinates $\alpha, \delta$.

Parameter-space resolutions are given in [46]

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Resolution</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\delta f$</td>
<td>$1.4 \times 10^{-4}$ Hz</td>
</tr>
<tr>
<td>$\delta \dot{f}$</td>
<td>$5 \times 10^{-12}$ Hz/s</td>
</tr>
<tr>
<td>$\delta \theta$</td>
<td>0.69 Hz/f</td>
</tr>
</tbody>
</table>

TABLE III. Parameter-space resolutions employed by the SkyHough pipeline.

The first stage of the SkyHough pipeline performs a multi-detector search using H1 and L1 SFTs with $T_{\text{SFT}} = 7200$s. Each 0.025 Hz sub-band is analyzed separately using the same two step strategy as in [32, 41]: parameter-space is efficiently analyzed using SkyHough’s look-up table approach; the top 0.1% most significant candidates are further analyzed using a more sensitive statistic. The result for each frequency sub-band is a toplist containing the $10^5$ most significant candidates across the sky and spin-down parameter-space.

Each toplist is then clustered using a novel approach presented in [64] and firstly applied in [41]. A parameter-space distance is defined using the average mismatch in frequency evolution between two different parameter-space templates

$$d(\vec{x}, \vec{x}_s) = \frac{T_{\text{SFT}}}{N_{\text{SFT}}} \sum_{\alpha=0}^{N_{\text{SFT}}} \left| f(t_\alpha; \vec{x}) - f(t_\alpha; \vec{x}_s) \right| ,$$  \hspace{1cm} (8)

where $f(t; \vec{x})$ is defined as

$$f(t; \vec{x}) = \left[ f + (t-t_{\text{ref}}) \cdot \dot{f} \right] \cdot \left[ 1 + \frac{\vec{v}(t) \cdot \vec{n}}{c} \right]$$  \hspace{1cm} (9)

and $\vec{x} = \{ f, \dot{f}, \alpha, \delta \}$ refers to the phase-evolution parameters of the template.

Clusters are constructed by pairing together templates in consecutive frequency bins such that $d(\vec{x}, \vec{x}_s) \leq 1$. Each cluster is characterized by its most significant element (the loudest element). From each 0.025 Hz sub-band, we retrieve the forty most significant clusters for further analysis. This results in a total of 456000 candidates to follow-up.

The loudest cluster elements are first sieved through the line veto, a standard tool to discard clear instrumental artifacts using the list of known, narrow, instrumental artifacts (lines) in the advanced LIGO detectors [53]: If the instantaneous frequency of a candidate overlaps with a frequency band containing an instrumental line of known origin, the candidate is ascribed an instrumental origin and consequently ruled out.

Surviving candidates are then followed-up using PyFstat, a Python package implementing a Markov-chain Monte Carlo (MCMC) search for CW signals [65, 68]. The follow-up uses the $F$-statistic as a (log) Bayes factor to sample the posterior probability distribution of the phase-evolution parameters around a certain parameter-space region

$$P(\vec{x}|x) \propto e^{F(\vec{x}|x)} \cdot P(\vec{x}) ,$$  \hspace{1cm} (10)

where $P(\vec{x})$ represents the prior probability distribution of the phase-evolution parameters. The $F$-statistic, as opposed to the SkyHough number count, allows us to use longer coherence times, increasing the sensitivity of the follow-up with respect to the main search stage.
TABLE IV. Coherence-time configuration of the multi-stage follow-up employed by the *SkyHough* pipeline. The data stream is divided into a fixed number of segments of the same length; the reported coherence time is an approximate value obtained by dividing the observation time by the number of segments at each stage.

<table>
<thead>
<tr>
<th>Stage</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N_{\text{seg}}$</td>
<td>660</td>
<td>330</td>
<td>92</td>
<td>24</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>$T_{\text{coh}}$ 0.5 day</td>
<td>1 day</td>
<td>4 days</td>
<td>15 days</td>
<td>90 days</td>
<td>360 days</td>
<td></td>
</tr>
</tbody>
</table>

As initially described in [68], the effectiveness of an MCMC follow-up is tied to the number of templates covered by the initial prior volume, suggesting a hierarchical approach: coherence time should be increased following a ladder so that the follow-up is able to converge to the true signal parameters at each stage. We follow the proposal in [66] and compute a coherence-time ladder using $N^* = 10^3$ (see Eq. (31) of [68]) starting from $T_{\text{coh}} = 1$ day including an initial stage of $T_{\text{coh}} = 0.5$ days. The resulting configuration is collected in Table IV.

The first follow-up stage is similar to that employed in [40, 41]: an MCMC search around the loudest candidate of the selected clusters is performed using a coherence time of $T_{\text{coh}} = 0.5$ days. Uniform priors containing 4 parameter-space bins in each dimension are centered around the loudest candidate. A threshold is calibrated using an injection campaign: any candidate whose loudest 2F value over the MCMC run is lower than 2F = 3450 is deemed inconsistent with CW signal.

The second follow-up stage is a variation of the method described in [66], previously applied to [69, 70]. For each outlier surviving the initial follow-up stage (stage 0 in Table IV), we construct a Gaussian prior using the median and inter-quartile range of the posterior samples and run the next-stage MCMC follow-up. The resulting maximum 2F is then compared to the expected 2F inferred from the previous MCMC follow-up stage. Highly-discrepant candidates are deemed inconsistent with a CW signal and hence discarded.

Given an MCMC stage using $\hat{N}$ segments from which a value of 2F is recovered, the distribution of 2F values using $N$ segments is well approximated by

$$P(2F|N, 2\hat{F}, \hat{N}) = \text{Gauss}(2F; \mu, \sigma),$$  \hspace{1cm} (11)

where

$$\mu = \rho_0^2 + 4N,$$

$$\sigma^2 = 8 \cdot (N + \hat{N} + \rho_0^2),$$

and $\rho_0^2 = 2\hat{F} - 4\hat{N}$ is a proxy for the (squared) SNR [71]. Equation (11) is exact in the limit of $N, \hat{N} \gg 1$ or $\rho_0^2 \gg 1$. In this search, however, we calibrate a bracket on $(2F - \mu)/\sigma$ for each follow-up stage using an injection campaign, shown in Table V. Candidates outside of the bracket are deemed inconsistent with a CW signal.

<table>
<thead>
<tr>
<th>Comparing stages</th>
<th>$(2F - \mu)/\sigma$ bracket</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stage 0 vs. Stage 1</td>
<td>(-1.79, 1.69)</td>
</tr>
<tr>
<td>Stage 1 vs. Stage 2</td>
<td>(-1.47, 1.35)</td>
</tr>
<tr>
<td>Stage 2 vs. Stage 3</td>
<td>(-0.94, 0.80)</td>
</tr>
<tr>
<td>Stage 3 vs. Stage 4</td>
<td>(-0.63, 0.42)</td>
</tr>
<tr>
<td>Stage 4 vs. Stage 5</td>
<td>(-0.34, 0.11)</td>
</tr>
</tbody>
</table>

TABLE V. 2F consistency brackets employed in the multi-stage follow-up of the *SkyHough* pipeline. Brackets were computed using a campaign of 500 software-injected signals representing an isotropic population of uniformly sky-distributed NS at 150 representative frequency bands with an amplitude corresponding to the $h_0^{95\%}$ sensitivity estimation. The implied false dismissal probability is $\lesssim 1/(150 \times 500) \approx 1.3 \times 10^{-5}$. Stages correspond to those described in Table IV.

Any surviving candidates are subject to manual inspection in search for obvious instrumental causes such as hardware-injected artificial signals or narrow instrumental artifacts.

C. Time-Domain $F$-statistic

The *Time-Domain $F$-statistic* search method has been applied to an all-sky search of VSR1 data [48] and all-sky searches of the LIGO O1 and O2 data [31, 32, 34]. The main tool of the pipeline is the $F$-statistic [47] with which one can coherently search the data over a reduced parameter space consisting of signal frequency, its derivatives, and the sky position of the source. However, a coherent all-sky search over the long data set like the whole data of O3 run is computationally prohibitive. Thus the data are divided into shorter time domain segments. Moreover, to reduce the computer memory required to do the search, the data are divided into narrow-band segments that are analyzed separately. As a result the *Time-Domain $F$-statistic* pipeline consists of two parts. The first part is the coherent search of narrowband, time-domain segments. The second part is the search for coincidences among the parameters of the candidates obtained from the coherent search of all the time domain segments.

The algorithms to calculate the $F$-statistic in the coherent search are described in Sec. 6.2 of [48]. The time series is divided into segments, called frames, of six sidereal days long each. Moreover the data are divided into sub-bands of 0.25 Hz overlapped by 0.025 Hz. The O3 data has a number of non-science data segments. The values of these bad data are set to zero. For our analysis, we choose only segments that have a fraction of bad data less than 60% both in H1 and L1 data and there is an overlap of more than 50% between the data in the two detectors. This requirement results in forty-one 6-day-long data segments for each sub-band. For the search we use a four-dimensional grid of templates (parameterized by frequency, spin down rate, and two more parameters related to the position of the source in the sky) constructed in Sec. 4 of [72] with grid’s minimal match parameter MM.
chosen to be $\sqrt{3}/2$. This choice of the grid spacing led to the following resolution for the four parameters of the space that we search

\[
\begin{align*}
\Delta f &\simeq 1.9 \times 10^{-6} \text{ Hz}, \\
\Delta \dot{f} &\simeq 1.1 \times 10^{-11} \text{ Hz/s}, \\
\Delta \alpha &\simeq 7.4 \times 10^{-2} \left(\frac{100 \text{ Hz}}{f}\right) \text{ rad}, \\
\Delta \delta &\simeq 1.5 \times 10^{-2} \left(\frac{100 \text{ Hz}}{f}\right) \text{ rad}.
\end{align*}
\]

We set a fixed threshold of 15.5 for the $F$-statistic and record the parameters of all threshold crossings, together with the corresponding values of the $F$-statistic. In the second stage of the analysis we use exactly the same coincidence search algorithm as in the analysis of VSR1 data and described in detail in Sec. 8 of [48] with only one change. We use a different coincidence cell from that described in [48]. In [48] the coincidence cell was constructed from Taylor expansion of the autocorrelation function of the $F$-statistic. In the search performed here the chosen coincidence cell is a suitably scaled grid cell used in the coherent part of the pipeline. We scale the four dimensions of the grid cell by different factors given by \([16 8 2 2]\) corresponding to frequency, spin down rate (frequency derivative), and two more parameters related to the position of the source in the sky respectively. This choice of scaling gives optimal sensitivity of the search. We search for coincidences in each of the bands analyzed. Before identifying coincidences we veto candidate signals overlapping with the instrumental lines identified by independent analysis of the detector data. To estimate the significance of a given coincidence, we use the formula for the false alarm probability derived in the appendix of [48]. Sufficiently significant coincidences are called outliers and are subject to a further investigation.

1. Parameter space

Our Time-Domain $F$-statistic analysis is a search over a 4-dimensional space consisting of four parameters: frequency, spin-down rate and sky position. As we search over the whole sky the search is very computationally intensive. Given that our computing resources are limited, to achieve a satisfactory sensitivity we have restricted the range of frequency and spin-down rates analyzed to cover the frequency and spin-down ranges of the bulk of the observed pulsars. Thus we have searched the gravitational frequency band from 20 Hz to 750 Hz. The lower frequency of 20 Hz is chosen due to the low sensitivity of the interferometers below 20 Hz. In the frequency 20 Hz to 130 Hz range, assuming that the GW frequency is twice the spin frequency, we cover young and energetic pulsars, such as Crab and Vela. In the frequency range from 80 Hz to 160 Hz we can expect GW signal due to r-mode instabilities [73, 74]. In the frequency range from 160 Hz to 750 Hz we can expect signals from most of the recycled millisecond pulsars, see Fig. 3 of [75].

For the GW frequency derivative $\dot{f}$ we have chosen a frequency dependent range. Namely, for frequencies less than 200 Hz we have chosen $\dot{f}$ to be in the range $[-f/\tau_{\text{min}}, 0]$, where $\tau_{\text{min}}$ is a limit on pulsar’s characteristic age, and we have taken $\tau_{\text{min}} = 1000$ yr. For frequencies greater than 200 Hz we have chosen a fixed range for the spin-down rate. As a result, the following ranges of $\dot{f}$ were searched in our analysis:

\[
\begin{align*}
0 &< \dot{f} > -3.2 \times 10^{-9} \frac{f}{100 \text{ Hz}} \text{ Hz/s}, \\
\text{for } f < 200 \text{ Hz}, \\
2 \times 10^{-11} \text{ Hz/s} &> \dot{f} > -2 \times 10^{-10} \text{ Hz/s}, \\
\text{for } f > 200 \text{ Hz}. 
\end{align*}
\]

In Fig. 2 we plot GW frequency derivatives against GW frequencies (assuming the GW frequency is twice the spin frequency of the pulsar) for the observed pulsars from the ATNF catalogue [76]. We show the range of the GW frequency derivative selected in our search, and one can see that the expected frequency derivatives of the observed pulsars are well within this range. Note, finally, that we have made the conservative choice of including positive values of the frequency derivative (‘spin-up’), in order to search as wide a range as possible. In most cases, however, the pulsars that appear to spin-up are in globular clusters, for which the local forces make the measurement unreliable [77].

2. Sensitivity of the search

In order to assess the sensitivity of the $F$-statistic search, we set upper limits on the intrinsic GW amplitude $h_0$ in each 0.25 Hz bands. To do so, we generate signals for an array of 8 amplitudes $h_0$ and for randomly selected sky positions (samples drawn uniformly from the sphere). For each amplitude, we generate 100 signals with $f$, $\dot{f}$, the polarization angle $\psi$ and cosine of the inclination angle $\iota$ are chosen from uniform random distributions in their respective ranges. The signals are added to the real data segments, and searches are performed with the same grids and search set-up as for the real data search, in the neighbourhood of injected signal parameters. We search $\pm 6$ grid points for $\dot{f}$ and $\pm 1$ grid points for the sky positions away from the true values of the signal’s parameters. We consider a signal detected if coincidence multiplicity for the injected signal is higher than the highest signal multiplicity in a given sub-band and in a given hemisphere in the real data search. The detection efficiency is the fraction of recovered signals. We estimate the $h_0^{95\%}$, i.e., 95% confidence upper limit on
the GW amplitude $h_0$, by fitting a sigmoid function to a range of detection efficiencies $E$ as a function of injected amplitudes $h_0$, $E(h_0) = (1 + e^{k(x_0-h_0)})^{-1}$, with $k$ and $x_0$ being the parameters of the fit. Figure 3 presents an example fit to the simulated data with $1\sigma$ errors on the $h_0^{95\%}$ estimate marked in red.

D. SOAP

SOAP [49] is a fast, model-agnostic search for long duration signals based on the Viterbi algorithm [81]. It is intended as both a rapid initial search for isolated NSs, quickly providing candidates for other search methods to investigate further, as well as a method to identify long duration signals which may not follow the standard Continuous Wave (CW) frequency evolution. In its most simple form SOAP analyzes a spectrogram to find the continuous time-frequency track which gives the highest sum of fast Fourier transform power. If there is a signal present within the data then this track is the most likely to correspond to that signal. The search pipeline consists of three main stages, the initial SOAP search [49], the post processing step using convolutional neural networks [82] and a parameter estimation stage.

1. Data preparation

The data used for this search starts as calibrated detector data which is used to create a set of fast Fourier transforms (FFTs) with a coherence time of 1800 s. The power spectrum of these FFTs are then summed over one day, i.e. every 48 FFTs. Assuming that the signal remains within a single bin over the day, this averages out the antenna pattern modulation and increases the SNR in a given frequency bin. As the frequency of a CW signal increases, the magnitude of the daily Doppler modulation also increases, therefore the assumption that a signal remains in a single frequency bin within one day no longer holds. Therefore, the analysis is split into 4 separate bands (40-500 Hz, 500-1000 Hz, 1000-1500 Hz, 1500-2000 Hz) where for each band the Doppler modulations are accounted for by taking the sum of the power in adjacent frequency bins. For the bands starting at
At this stage there is a set of Viterbi statistics and CNN statistics for each sub-band that is analysed, from which a set of candidate signals need to be selected for followup. Before doing this, any sub-bands which contain known instrumental artefacts are removed from the analysis. The sub-bands corresponding to the top 1% of the Viterbi statistics from each of the four analysis bands are then combined with the sub-bands corresponding to the top 1% of CNN statistics, leaving us with a maximum of 2% of the sub-bands as candidates. It is at this point where we begin to reject candidates by manually removing sub-bands which contain clear instrumental artefacts and still crossed the detection threshold for either the Viterbi or CNN statistic. There are a number of features we use to reject candidates including: strong detector artefacts which only appear in a single detectors spectrogram, broad ( $> 1/5$ sub-band width) long duration signals, individual time-frequency bins which contribute large amounts to the final statistic and very high power signals in both detectors. Examples of these features can be seen in section 6.3 of [83]. Any remaining candidates are then passed on for parameter estimation.
5. Parameter estimation

The parameter estimation stage uses the Viterbi track to estimate the Doppler parameters of the potential source. Due to the complicated and correlated noise which appears in the Viterbi tracks, defining a likelihood is challenging. To avoid this difficulty, likelihood-free methods are used, in particular a machine learning method known as a conditional variational auto-encoder. This technique was originally developed for parameter estimation of compact binary coalescence signals [84], and can return Bayesian posteriors rapidly (<1s). In our implementation, the conditional variational auto-encoder is trained on isolated NS signals injected into many sub-bands, and returns an estimate of the Bayesian posterior in the frequency, frequency derivative and sky position [85]. This acts both as a further check that the track is consistent with that of an isolated NS, and provides a smaller parameter space for a follow-up search.

V. RESULTS

In this section we summarize the results of the search obtained by the four pipelines. Each pipelines presents candidates obtained during the analysis and the results of the follow-up of the promising candidates. The upper limits on the GW strain are determined for each of the search procedures. There is also a study of the hardware injections of continuous wave signals added to the data. During the O3 run 18 hardware injections were added to the LIGO data. The injections are denoted by ipN where N is the consecutive number of the injection. The amplitudes of the injections added in the O3 run were significantly lower than those added in previous observing runs. Consequently the injections were more difficult to detect.

A. FrequencyHough

Outliers produced by the FrequencyHough search are followed-up by the procedure described in Sec. IV A 1. The increase in FFT duration sets the sensitivity gain of the follow-up step and it is mainly limited by the resulting computational load, which increases with the fourth power of $T_{\text{FFT}}$ for a fixed follow-up volume. Moreover, $T_{\text{FFT}}$ cannot be longer than about one sidereal day, because the current procedure is not able to properly deal with the sidereal splitting of the signal power, which would cause a sensitivity loss.

All the coincident outliers produced by the FrequencyHough transform stage in the first frequency band, 10-128 Hz, have been followed-up. On the remaining frequency bands, from 128 Hz up to 2048 Hz, only outliers with $CR \geq 5$ (computed over the FrequencyHough map) in both detectors have been followed-up. This selection was also applied for pairs of coincident outliers produced in the L1 - Virgo and H1 - Virgo detectors in the frequency band 10-128 Hz.

Table VI summarize the results of the first follow-up stage over coincident H1 - L1 outliers, for each of the four analyzed frequency bands, given in the first column. The second columns is the value of $T_{\text{FFT}}$ used at this stage, $N_i$ the initial number of outliers to which the follow-up is applied. Subsequent columns indicate the number of candidates removed by the various vetoes, indicated as $V_i, i = 1,..5$ and discussed in the section IV A 1. The last column shows the number of outliers surviving the first follow-up stage. As it can be seen from the last column, 29 outliers survive this follow-up stage. Tab. VII shows the same quantities for the follow-up of coincident H1 - Virgo and L1 - Virgo outliers, which have been selected in the lowest frequency band, from 10 to 128 Hz. In this case, all the outliers have been discarded. Outliers which survived the first follow-up stage have been analyzed with a second step based on the same procedure as before but with a further increase in the FFT duration, which has been roughly doubled. The main quantities for the second follow-up stage are shown in Tab. VIII. The eight outliers in the band 512 - 1024 Hz are due to hardware injection ip1. An example is shown in Fig. 4, where the
peakmap after Doppler correction is plotted for a small frequency range around the outlier frequency. Although the outlier parameters are relatively far from those of ip1, it is expected, especially in the case of a strong signal like this that - due to parameter correlations - outliers can spread over a rather large portion of the parameter space around the exact signal.

1. Upper limits

Having concluded that no candidate has a likely astrophysical origin, we have computed upper limits following the method described in Sec. IV A 3. Results are shown in Fig. 5. Although the search has been carried with a minimum frequency of 10 Hz, due to the unreliable calibration below 20 Hz, upper limits are given starting from this minimum frequency. The bold continuous curve represents our conservative upper limit estimation, computed on 1 Hz sub-bands and based on the maximum CR, while the lighter dashed curve is a (non-conservative) lower bound, obtained using the minimum CR in each sub-band. We expect the search sensitivity, defined as the minimum detectable strain amplitude, to be comprised among the two curves. The minimum upper limit is about $1.1 \times 10^{-25}$, at 116.5 Hz.

The search distance reach, expressed as a relation between the absolute value of the first frequency derivative and the frequency of detectable sources for various source distances, under the assumption the GW emission is the only spin-down mechanism (NSs in this case are often dubbed as gravitars [86]), is shown in Fig. 16.

2. Hardware Injections

Table IX shows the error of the recovered signal with respect to the hardware injections. The reported values have been obtained at the end of the first follow-up stage, which was enough to confidently detect the reported signals. The second column gives the total distance metric, defined in Sec. IV A, among the injection and the corresponding strongest analysis candidate. Columns 3-6 give the error values for the individual parameters. Column 7 indicates the CR of the strongest candidate corresponding to each injection, and the last column gives the expected number of candidates due to noise, having the same (or bigger) CR value, after taking into account the trial factor. As shown in the Table, we have been able to detect 5 injections in the analyzed parameter space and the estimated parameters do show a good agreement with the injected ones. All reported values are the mean of the values obtained separately for the Livingston and Hanford detectors, with the exception of the CR and $N_n$ for ip3, for which the reported values refer to Livingston alone. This hardware injection is in fact very weak and it was confidently detected, after the first follow-up stage, only in Livingston detector, which has a better sensitivity at the injection frequency.

B. SkyHough

1. Candidate follow-up

Table X summarizes the number of outliers discarded by each of the veto and follow-up stages employed in this search. A total of 36 candidates survive the complete suite of veto and follow-up stages of the SkyHough pipeline. Candidates can be grouped into two sets according to their corresponding $F$-statistic value: 31 candidates present a value of $2F \sim O(10^3)$, while the remaining 5 candidate only achieve $2F \sim O(30)$. Their corresponding parameters are collected in Table XI.

The 31 strong candidates present consistent values with the only two hardware injections within the SkyHough search range: 24 candidates are ascribed to the hardware injection ip0, while 7 candidates are ascribed to the hardware injection ip3. Parameter deviation of the loudest candidate associated to each injection are reported in Table XII.

The five weaker candidates are manually inspected using the segment-wise $F$-statistic on 660 coherent segments, in a similar manner to that in [39, 66].

The first pair of candidates is found around 89.850 Hz, where the H1 detector presents a broad spectral feature. As shown in Fig 6, their single-detector $F$-statistic is more prominent in the H1 detector rather than the L1 detector, and scores over the multi-detector $F$-statistic. These characteristics point towards an instrumental, rather than astrophysical, origin.

A second pair of candidates is found around 95.7 Hz. This frequency band is populated by narrow spectral artifacts of unknown origin in the H1 detector. Correspondingly, as shown in Fig. 7, the single-detector $F$ statistic is prominent in the H1 detector rather than the L1 detector. Due to the narrowness of the feature, in this case the accumulation is better localized around a fraction of the run. As in the previous case, the single-detector $F$-statistic scores over the multi-detector $F$-statistic. These characteristics point towards an instrumental origin.

The last weak candidate in the vicinity of 246.275 Hz, where the H1 detector presents another narrow spectral artifact of unknown origin. The single-detector $F$-statistic is more prominent in the H1 detector than in the L1 detector, and accumulates rapidly the beginning of the run. As in the previous cases, this behavior is consistent with that of an instrumental artifact.

This concludes the analysis of surviving candidates of the SkyHough pipeline. Every single one of them could be related to an instrumental feature.
FIG. 4. Peakmap of H1 (left) and L1 (right) data showing one of the eight outliers removed by veto V5 in the second follow-up step, see VI. All the 8 outliers were generated by the hardware injection ip1. A Doppler correction, with parameters not exactly equal with those of the signal, nevertheless aligns some of the signal peaks, thus producing an excess of counts in the FrequencyHough map.

TABLE IX. Hardware injection recovery by the FrequencyHough pipeline. The second column indicates the total distance metric among the injection and the corresponding strongest analysis candidate. Columns 3-6 give the error values for the individual parameters (frequency, spin-down, ecliptical longitude and latitude). Column 7 indicate the CR of the strongest candidate corresponding to each injection, and the last column gives the expected number of candidates due to noise, having the same (or bigger) CR value, after taking into account the trial factor. All the reported values are the mean of individual values found separately in Livingston and Hanford detectors, with the exception of the CR and \( N_n \) for ip3, indicated by an asterisk, for which the reported values refer to Livingston alone. This hardware injection is very weak and it was confidently detected, after the first follow-up stage, only in Livingston, which has a better sensitivity at the injection frequency.

<table>
<thead>
<tr>
<th>Injection</th>
<th>( d_{FH} )</th>
<th>( \Delta f ) [Hz]</th>
<th>( \Delta f ) [nHz/s]</th>
<th>( \Delta \lambda ) [deg]</th>
<th>( \Delta \beta ) [deg]</th>
<th>CR</th>
<th>( N_n )</th>
</tr>
</thead>
<tbody>
<tr>
<td>ip1</td>
<td>0.77 ( 1.15 \times 10^{-4} )</td>
<td>15.11 ( 10^{-3} )</td>
<td>0.015</td>
<td>-0.027</td>
<td>51.76</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>ip3</td>
<td>1.05 ( 4.88 \times 10^{-5} )</td>
<td>3.52 ( 10^{-3} )</td>
<td>0.088</td>
<td>-0.377</td>
<td>6.34*</td>
<td>0.04</td>
<td></td>
</tr>
<tr>
<td>ip5</td>
<td>1.92 ( 2.65 \times 10^{-5} )</td>
<td>9.64 ( 10^{-3} )</td>
<td>0.615</td>
<td>-0.130</td>
<td>41.58</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>ip6</td>
<td>0.16 ( 9.27 \times 10^{-6} )</td>
<td>1.31 ( 10^{-3} )</td>
<td>0.009</td>
<td>0.045</td>
<td>56.05</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>ip14</td>
<td>1.52 ( 5.77 \times 10^{-4} )</td>
<td>52.27 ( 10^{-3} )</td>
<td>0.054</td>
<td>0.521</td>
<td>20.58</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

FIG. 5. O3 conservative upper limit estimation (bold continuous curve) and sensitivity lower bound (light dashed curve) for the FrequencyHough search.

TABLE X. Summary of candidates processed by each of the veto and follow-up stages of the SkyHough search.

<table>
<thead>
<tr>
<th>Search stage</th>
<th>Candidates</th>
<th>% removed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clustering</td>
<td>456000</td>
<td></td>
</tr>
<tr>
<td>Line veto</td>
<td>414459</td>
<td>9%</td>
</tr>
<tr>
<td>2( F ) threshold</td>
<td>3767</td>
<td>99%</td>
</tr>
<tr>
<td>Stage 0 v.s. Stage 1</td>
<td>697</td>
<td>18%</td>
</tr>
<tr>
<td>Stage 1 v.s. Stage 2</td>
<td>172</td>
<td>75%</td>
</tr>
<tr>
<td>Stage 3 v.s. Stage 3</td>
<td>90</td>
<td>48%</td>
</tr>
<tr>
<td>Stage 3 v.s. Stage 4</td>
<td>48</td>
<td>47%</td>
</tr>
<tr>
<td>Stage 4 v.s. Stage 5</td>
<td>36</td>
<td>25%</td>
</tr>
</tbody>
</table>

2. Sensitivity estimation

We estimate the search sensitivity following the same procedure as previous searches \([31, 32, 34, 40, 41]\). Search
TABLE XI. Surviving candidates of the SkyHough multi-stage MCMC follow-up using PyFstat. $2\hat{F}$ corresponds to the loudest fully-coherent $F$-statistic value of the MCMC run. Band index corresponds to a frequency of $(65 + 0.025 \times \text{Band})$ Hz. Reference time is GPS 1238166018.

<table>
<thead>
<tr>
<th>Injection</th>
<th>$2\hat{F}$</th>
<th>$\Delta f$ [Hz]</th>
<th>$\Delta f$ [nHz/s]</th>
<th>$\Delta \alpha$ [rad]</th>
<th>$\Delta \delta$ [rad]</th>
<th>$\Delta \alpha$ [deg]</th>
<th>$\Delta \delta$ [deg]</th>
</tr>
</thead>
<tbody>
<tr>
<td>ip0</td>
<td>1543.72</td>
<td>-4.80 x 10^{-9}</td>
<td>3.52 x 10^{-7}</td>
<td>-2.82 x 10^{-7}</td>
<td>-2.49 x 10^{-8}</td>
<td>-1.62 x 10^{-5}</td>
<td>-1.43 x 10^{-6}</td>
</tr>
<tr>
<td>ip3</td>
<td>1055.71</td>
<td>1.16 x 10^{-8}</td>
<td>-7.29 x 10^{-7}</td>
<td>9.35 x 10^{-7}</td>
<td>1.53 x 10^{-6}</td>
<td>5.35 x 10^{-5}</td>
<td>8.74 x 10^{-5}</td>
</tr>
</tbody>
</table>

TABLE XII. Hardware injection recovery by the SkyHough pipeline. For each hardware injection within search range we report the dimension-wise errors with respect to loudest surviving candidate of the follow-up.

sensitivity is quantified using the sensitivity depth \cite{87, 88}.

\[
D = \frac{\sqrt{\sum_\alpha}}{h_0}, \tag{17}
\]

where $S_n$ represents the power spectral density (PSD) of the data, computed as the inverse squared average of the individual SFT’s running-median PSD \cite{41, 60}.

\[
S_n(f) = \sqrt{\frac{N_\alpha}{\sum_\alpha [S_\alpha(f)]^{-2}}}. \tag{18}
\]

where $S_\alpha$ represents the running-median noise floor estimation using 101 bins from the SFT labeled by starting time $t_\alpha$ (including SFTs from both the H1 and L1 detect-
FIG. 6. *SkyHough* candidates consistent with a broad spectral artifact in the H1 detector. Upper panel shows the cumulative semicoherent $F$-statistic using 660 coherent segments ($T_{\text{coh}} = 0.5$ days). Lower panel shows the segment-wise $F$-statistic. Dashed red line represents the single-detector $F$-statistic using H1-only data; dot-dashed blue line represents the single-detector $F$-statistic using L1-only data. Solid gray line represents the multi-detector $F$-statistic. Dotted horizontal line represents the threshold of $2F = 3450$ set at the initial follow-up stage.

FIG. 7. *SkyHough* candidate consistent with two narrow spectral artifacts of unknown origin in the H1 detector. The legend is equivalent to that of Fig. 6.

FIG. 8. *SkyHough* candidate consistent with a narrow spectral artifact of unknown origin in the H1 detector. The legend is equivalent to that of Fig. 6.

The sensitivity depth $D^{95\%}$ corresponding to a 95% average detection rate is characterized by adding a campaign of software-simulated signals into the data. Simulated signals are added into 150 representative frequency bands at several sensitivity depth values bracketing the $D^{95\%}$ value in each band, as represented in Fig. 10. For each sensitivity depth, 200 simulated signals drawn from uniform distribution in phase and amplitude parameters are added into the data. The *SkyHough* is run on each of these signals in order to evaluate how many of them are detected, and the resulting toplists are clustered using the same configuration as in the main stage of the search.

For each simulated signal, we retrieve the best forty resulting clusters. The following two criteria must be fulfilled in order to label a simulated signal as “detected”. First, the loudest significance of at least one of the selected clusters must be higher than the minimum significance recovered by the corresponding all-sky clustering; this ensures the signal is significant enough to be selected for a follow-up stage. Second, the parameters of the loudest candidate in said clusters must be closer than two parameter-space bins (see Eq. (7) and Table III) from the simulated-signal’s parameter, as otherwise the follow-up would have missed the signal.

The efficiency associated to each sensitivity depth $E$ is computed as the fraction of simulated signals labeled as detected. A binomial uncertainty $\delta E$ is associated to each efficiency

$$\delta E = \sqrt{\frac{E \cdot (1 - E)}{N_I}}, \quad (19)$$
FIG. 9. ASD employed by the SkyHough pipeline to estimate the sensitivity of the search. ASD is computed as the square root of the single-sided inverse-square averaged PSD using data from both the H1 and L1 advanced LIGO detectors, as explained in the text surrounding Eq. (18).

FIG. 10. Example computation of \( D_{95\%} \) (white star) at a frequency band by fitting a sigmoid function (blue solid line) to a set of efficiencies (blue dots) computed using 200 injections at each sensitivity depth for the SkyHough search. Shaded regions represent 1, 2, and 3 sigma envelopes of the sigmoid fit. Error bars are computed as discussed in the main text.

FIG. 11. Wide-band interpolation \( D_{95\%}(f) \) of the results obtained by the SkyHough pipeline. Each dot represents a \( D_{95\%} \) at a particular frequency band computed using the procedure exemplified in Fig. 10. The red solid line represents a non-parametric interpolation using a Gaussian process regression, as discussed in the main text. The shaded region represents a 3% relative error with respect to the interpolation and corresponds to the 98% credible interval.

We compute the average wide-band \( D_{95\%}(f) \) value using Gaussian process regression, as shown in Fig. 11. We fit a Gaussian process using to the ensemble of \( D_{95\%} \) obtained from the injection campaign using scikit-learn’s GaussianProcessRegressor with an RBF kernel [89]. The uncertainty associated to the fit is computed as the 98% credible region of the deviations with respect to the Gaussian process regression, which corresponds to a 3% relative uncertainty. Equation (17) allows us to translate \( D_{95\%}(f) \) into a corresponding CW amplitude \( h_{95\%}(f) \), shown in Fig. 12.

C. Time-Domain \( F \)-statistic

In the frequency bandwidth of [20, 750] Hz that we analyze we have 3245 sub-bands that are 0.25 Hz wide and that are overlapped by 0.025 Hz. 104 sub-bands were not analyzed because of the excessive noise originating mainly from the 1st harmonic of the violin mode, 1st and 2nd harmonics of the beam splitter violin mode, and

where \( N_1 = 200 \) represents the number of signals. Then, we use scipy’s curve_fit function [79] to fit a sigmoid curve to the data given by

\[
S(D; a, b) = 1 - \frac{1}{1 + \exp(-aD + b)} \tag{20}
\]

where \( a, b \) represent the parameters to adjust. After fitting, this expression can be numerically inverted to obtain \( D_{95\%} \). The uncertainty associated to the fit is computed through the covariance matrix \( C \) as

\[
\delta D_{95\%} = \sqrt{\left( \frac{\partial S}{\partial a} \right)^2 C_{aa} + 2 \left( \frac{\partial S}{\partial a} \right) \left( \frac{\partial S}{\partial b} \right) C_{ab} + \left( \frac{\partial S}{\partial b} \right)^2 C_{bb}} \tag{21}
\]

This procedure is exemplified in Fig. 10.

We compute the average wide-band \( D_{95\%}(f) \) value using Gaussian process regression, as shown in Fig. 11. We fit a Gaussian process using to the ensemble of \( D_{95\%} \) obtained from the injection campaign using scikit-learn’s GaussianProcessRegressor with an RBF kernel [89]. The uncertainty associated to the fit is computed as the 98% credible region of the deviations with respect to the Gaussian process regression, which corresponds to a 3% relative uncertainty. Equation (17) allows us to translate \( D_{95\%}(f) \) into a corresponding CW amplitude \( h_{95\%}(f) \), shown in Fig. 12.

7 This method is akin to that employed by the SkyHough search in [34]. We note that Eq. (19) in [41] is incorrect and should be equivalent to Eq. (21) in this document. This is just a typographical error, as the analysis was performed using the correct formulae.
60 Hz mains line and its harmonics. This leads to the loss of around 23.50 Hz of the band. Moreover, we have vetoed lines identified by the detector characterization group. This leads to an additional 34.18 Hz band loss. Thus altogether 57.68 Hz of the band was vetoed, which constitutes 7.9% of the 730 Hz band analyzed. Consequently we searched 3141 sub-bands. For each sub-band we analyzed coherently 41 six-day time segments with the $F$-statistic. As a result with our $F$-statistic threshold of 15.5 we obtained $5.47 \times 10^{10}$ candidates.

In the second stage of the analysis for each sub-band we search for coincidences among the candidates from the 41 time-domain segments. For each sub-band and each hemisphere we find the candidate with the smallest coincidence false alarm probability, i.e. the most significant candidate. As a result we have 6282 top candidates from our search. Among the top candidates we consider a candidate to be statistically significant if the coincidence false alarm probability is less than 1%. This leads to the selection of 311 candidates that we call outliers. The outliers were subject to further investigation to determine whether they can be considered as true GW events. Three of the outliers were determined to be ‘artificial’ GW signals injected in hardware to the LIGO detectors data.

1. Hardware injections

In the parameter space analysed by Time-Domain $F$-statistic only six hardware injections were present. These are injections ip0, ip2, ip3, ip5, ip10, and ip11. In Table XIII we have compared the parameters of the top candidates obtained in our search in the frequency sub-bands, where the injections were made, with the parameters of the injections. In the table we show the false alarm probability of coincidence of the top candidates and the difference between the parameters of the candidate and the parameters of the injections. We see that the two injections ip5 and ip10 are detected with a very high confidence. Their false alarm probability is close to 0 and the errors in the parameter estimation are small. The top candidate in the band where injection ip11 is located has a very small false alarm probability; however, the right ascension of the candidate differs very much from the true value the right ascension of the injection. A close analysis shows that this candidate is associated with a strong line present in the Hanford detector. The line frequency is different from the hardware injection frequency by only around 10 mHz. The amplitude of the injection ip11 is very low. Its SNR in the 6-day segments that we analyse coherently with $F$-statistic is around 4. This is considerably lower than our threshold SNR of around 5.2 and it is not surprising that the injection is not recovered. For the remaining 3 bands we see that the top candidates have parameters very close to the parameters of the hardware injections ip0, ip2 and ip3, however their false alarm probabilities are greater than 1% and we cannot consider these injections as detected. The SNRs of the two detected injections ip5 and ip10 is considerably above our threshold of 5.2, whereas SNRs of the 3 remaining injections ip0, ip2, and ip3 are close to our threshold and they could not be detected.

2. Outliers

We have identified 311 outliers in our search. For these outliers the probability of being due to accidental coincidence between the candidates from the 41 time segments is less than 1%.

In our search we have vetoed the lines of known origin identified in LIGO detectors. However, the LIGO data contained additional lines and interferences. In order to identify the origin of the outliers in our search we have performed three independent investigations. Firstly we compared our outliers with the lines of unknown origin identified by the LIGO data characterization group. Secondly we have performed an independent search for strictly periodic signals in all the 6-day time-domain segments that we analyzed in our search. We have searched for periodic signals separately in the data from the Hanford and the Livingston LIGO detector. Thirdly we have performed a visual inspection of the outliers by searching the data with $F$-statistic around the outliers separately in the two LIGO detectors. In addition we have checked whether outliers are around the frequencies associated with the suspension violin mode 1st harmonic around 500 Hz and the beam splitter violin mode 1st and 2nd harmonics around 300 Hz and 600 Hz respectively. As a result of the above study 204 outliers were found to be associated with lines and interferences present in the detector. They were classified as follows. 146 originated from the Hanford detector, 21 were associated with the Livingston detector. One line that appeared in both de-
detectors was the 20 Hz tooth of the 1 Hz comb known to be present in both detectors. 36 outliers were associated with the two violin mode resonances. 2 outliers are pulsar injections ip5 and ip10 that were confidently detected and they are described in Sec. V C 1.

One of the outliers was associated with the pulsar injection ip6. The frequency of the outlier was only 15 mHz from the frequency of the injection. The injected signal ip6 has a spin-down of $-6.73 \times 10^{-9}$ Hz/s, which is outside our search range. However, the SNR of the injection was around 17 for each of the 6-day segments that we analyzed. This resulted in a sufficiently strong correlation to give a significant signal; however, with the spin-down and the sky position of the outlier very much displaced from the true values (see Table XIV).

The 102 outliers that could not be associated with interferences in the detector or hardware injections appeared with frequencies on the left edges of the 0.25 Hz sub-bands of the narrowband segments that we analyzed. To determine whether these are artifacts or they warrant a further detailed follow-up, we regenerated the narrowband data where the artefacts occurred, however with the offset frequencies decreased by 0.125 Hz (half of the width of the sub-band). Consequently the outliers that appeared at the left edges of the sub-bands, should now be present approximately in the middle of the sub-band. We have then performed a search with our pipeline around the parameters of the outliers. None of the outliers were found to be significant. The smallest false probability was found to be around 59%.

As a result we were left with 2 outliers for a more detailed study, with parameters given in Table XV. We followed up the outliers in the data segments that are twice as long as the original segments. For each sub-band where the outliers are present we divided the data into 12-day segments and we performed the search around the position of the outliers. A two-fold increase of the coherence times would result in the increase of the signal-to-noise ratio of a true GW signal by a factor of $\sqrt{2}$. We performed a coherent search $\pm 16$ grid points in spin down and $\pm 4$ points in the sky position around the point of the parameter space where the outliers should be present. We then performed a coincidence search. For the two cases we did not find a significant coincidence. The probability that the best coincidence was accidental was close to 1.

3. Upper limits

The analysis of the outliers described in Secs. V C 1 and V C 2 has not revealed a viable candidate for a GW event. We therefore proceeded to establish upper limits on the amplitude of GW signals in our search. We establish upper limits in each sub-band analyzed and for each hemisphere by using the procedure described in Sec. IV C 2 (as a result periodic interferences in the data for 201 sub-bands out of 3141 that we analyzed we were not able to establish upper limits). The 95% confidence upper limits $h_0^{95\%}$ for analysis of LIGO O3 data presented in the paper are plotted in Fig. 13 in comparison with upper limits obtained with our pipeline in O1 and O2 data. We see a considerable improvement which is more than the

<table>
<thead>
<tr>
<th>Injection</th>
<th>FAP</th>
<th>$\Delta f$ [Hz]</th>
<th>$\Delta f$ [nHz/s]</th>
<th>$\Delta \delta$ [deg]</th>
<th>$\Delta \alpha$ [deg]</th>
</tr>
</thead>
<tbody>
<tr>
<td>ip10</td>
<td>$&lt; 10^{-8}$</td>
<td>$7.73 \times 10^{-4}$</td>
<td>$3.91 \times 10^{-3}$</td>
<td>0.74</td>
<td>1.51</td>
</tr>
<tr>
<td>ip11</td>
<td>$&lt; 10^{-8}$</td>
<td>$6.20 \times 10^{-3}$</td>
<td>$7.22 \times 10^{-2}$</td>
<td>14.78</td>
<td>248.13</td>
</tr>
<tr>
<td>ip5</td>
<td>$&lt; 10^{-8}$</td>
<td>$1.62 \times 10^{-4}$</td>
<td>$6.64 \times 10^{-2}$</td>
<td>30.24</td>
<td>24.83</td>
</tr>
<tr>
<td>ip3</td>
<td>0.997</td>
<td>$5.21 \times 10^{-2}$</td>
<td>$5.21 \times 10^{-2}$</td>
<td>3.85</td>
<td>1.56</td>
</tr>
<tr>
<td>ip0</td>
<td>0.055</td>
<td>$1.84 \times 10^{-1}$</td>
<td>$1.15 \times 10^{-1}$</td>
<td>18.94</td>
<td>20.98</td>
</tr>
<tr>
<td>ip2</td>
<td>0.275</td>
<td>$4.41 \times 10^{-5}$</td>
<td>$9.28 \times 10^{-3}$</td>
<td>1.12</td>
<td>0.074</td>
</tr>
</tbody>
</table>

TABLE XIII. Hardware injection recovery with the Time-Domain $F$-statistic pipeline. The first column is the injection index ipN, where N is the injection number. The last 4 columns are the differences between the true values of the parameters of the injected signal ipN and the parameters of the most significant candidate in the sub-band where injection is added. The second column is the false alarm probability associated with the topmost candidate.

FIG. 13. Comparison of 95% confidence upper limits on GW amplitude $h_0$ obtained with the Time-Domain $F$-statistic pipeline in the analysis of Advanced LIGO data. The magenta circles, green triangles, and blue squares represent the $h_0^{95\%}$ upper limits in 0.25 Hz sub-bands of the O1, O2, and O3 data, respectively.
up to 1000 Hz and ±∼ summed respective to the analysis band.

SOAP frequency axis, and every 1 (2, 3, 4) frequency bins are the time axis, 48 FFTs (1 day) are summed along the

and frequency bins are summed together, where along

of their width. For each of the sub-bands, time segments

0.1 (0.2, 0.3, 0.4) Hz wide sub-bands overlapping by half

running median of width 100 bins before being split into
described in Sec. IV D 1, the FFTs are normalised to the

for when values are outside this range we lose sensitivity.

We start from a set of 1800s long FFTs of cleaned time-

series data from the two LIGO detectors H1 and L1. As

described in Sec. IVD1, the FFTs are normalised to the

running median of width 100 bins before being split into

0.1 (0.2, 0.3, 0.4) Hz wide sub-bands overlapping by half

of their width. For each of the sub-bands, time segments

and frequency bins are summed together, where along

the time axis, 48 FFTs (1 day) are summed along the

frequency axis, and every 1 (2, 3, 4) frequency bins are

summed respective to the analysis band. SOAP is then

run on each of these sub-bands, returning the Viterbi

statistic, Viterbi map and Viterbi tracks, which can be

input to the CNN to return a second statistic. The num-

ber of sub-bands searched totals to 19 868 across all four

analysis bands, where for each band (40-500, 500-1000,

1000-1500, 1500-2000 ) Hz the respective total is (9200,

5040, 3263, 2362). Sub-bands which contained known

instrumental lines identified by the calibration group are

then removed from the analysis leaving a total number of

sub-bands as 17 929, with each separate band containing

(8297, 4494, 2952, 2186) sub-bands. Candidates are then

selected by taking the sub-bands which contribute to the
top 1% of both the remaining Viterbi and CNN statist-
sics. These candidates can then be investigated further
to identify whether a real GW signal is present. Sub-

bands which contain an instrumental line identified by the
calibration group but also cross the 1% threshold are

also investigated to check whether it is the instrumental
line which causes the high statistic value. There were 293
sub-bands which were in this category, and in 291 sub-

bands the Viterbi tracks closely follow the instrumental
line, and the remaining two contained both an instru-
mental line and a hardware injection (ip5). These were
then reintroduced into the analysis as the Viterbi tracks
did not follow that of the instrumental line. From the
total of the 17 929 sub-bands, 248 were selected for a fol-
low-up investigation where 107 of these sub-bands cross
the thresholds of both the Viterbi and CNN statistics.

D. SOAP

SOAP was run on the O3 dataset from 40-2000 Hz
where we are sensitive to a broad range of signals from
the entire sky. To contain an entire signal within a single
sub-band, its spin-down must be within ± ∼ 10⁻⁵ Hz/s
up to 1000 Hz and ± ∼ 10⁻⁶ Hz/s above 1000 Hz, there-
fore we are sensitive to a broad range of signals from

the entire sky. To contain an entire signal within a single
sub-band, its spin-down must be within ± ∼ 10⁻⁵ Hz/s
up to 1000 Hz and ± ∼ 10⁻⁶ Hz/s above 1000 Hz, there-
fore when values are outside this range we lose sensitivity.

We start from a set of 1800s long FFTs of cleaned time-

series data from the two LIGO detectors H1 and L1. As

described in Sec. IVD1, the FFTs are normalised to the

running median of width 100 bins before being split into

0.1 (0.2, 0.3, 0.4) Hz wide sub-bands overlapping by half

of their width. For each of the sub-bands, time segments

and frequency bins are summed together, where along

the time axis, 48 FFTs (1 day) are summed along the

frequency axis, and every 1 (2, 3, 4) frequency bins are

summed respective to the analysis band. SOAP is then

run on each of these sub-bands, returning the Viterbi

statistic, Viterbi map and Viterbi tracks, which can be

input to the CNN to return a second statistic. The num-

ber of sub-bands searched totals to 19 868 across all four

analysis bands, where for each band (40-500, 500-1000,

1000-1500, 1500-2000 ) Hz the respective total is (9200,

5040, 3263, 2362). Sub-bands which contained known

instrumental lines identified by the calibration group are

then removed from the analysis leaving a total number of

sub-bands as 17 929, with each separate band containing

(8297, 4494, 2952, 2186) sub-bands. Candidates are then

selected by taking the sub-bands which contribute to the
top 1% of both the remaining Viterbi and CNN statist-
sics. These candidates can then be investigated further
to identify whether a real GW signal is present. Sub-

bands which contain an instrumental line identified by the
calibration group but also cross the 1% threshold are

also investigated to check whether it is the instrumental
line which causes the high statistic value. There were 293
sub-bands which were in this category, and in 291 sub-

bands the Viterbi tracks closely follow the instrumental
line, and the remaining two contained both an instru-
mental line and a hardware injection (ip5). These were
then reintroduced into the analysis as the Viterbi tracks
did not follow that of the instrumental line. From the
total of the 17 929 sub-bands, 248 were selected for a fol-
low-up investigation where 107 of these sub-bands cross
the thresholds of both the Viterbi and CNN statistics.

D. SOAP

SOAP was run on the O3 dataset from 40-2000 Hz
where we are sensitive to a broad range of signals from
the entire sky. To contain an entire signal within a single
sub-band, its spin-down must be within ± ∼ 10⁻⁵ Hz/s
up to 1000 Hz and ± ∼ 10⁻⁶ Hz/s above 1000 Hz, there-
fore when values are outside this range we lose sensitivity.

We start from a set of 1800s long FFTs of cleaned time-

series data from the two LIGO detectors H1 and L1. As

described in Sec. IVD1, the FFTs are normalised to the

running median of width 100 bins before being split into

0.1 (0.2, 0.3, 0.4) Hz wide sub-bands overlapping by half

of their width. For each of the sub-bands, time segments

and frequency bins are summed together, where along

the time axis, 48 FFTs (1 day) are summed along the

frequency axis, and every 1 (2, 3, 4) frequency bins are

summed respective to the analysis band. SOAP is then

run on each of these sub-bands, returning the Viterbi

statistic, Viterbi map and Viterbi tracks, which can be

input to the CNN to return a second statistic. The num-

ber of sub-bands searched totals to 19 868 across all four

analysis bands, where for each band (40-500, 500-1000,

1000-1500, 1500-2000 ) Hz the respective total is (9200,

5040, 3263, 2362). Sub-bands which contained known

instrumental lines identified by the calibration group are

then removed from the analysis leaving a total number of

sub-bands as 17 929, with each separate band containing

(8297, 4494, 2952, 2186) sub-bands. Candidates are then

selected by taking the sub-bands which contribute to the
top 1% of both the remaining Viterbi and CNN statist-
sics. These candidates can then be investigated further
to identify whether a real GW signal is present. Sub-

bands which contain an instrumental line identified by the
calibration group but also cross the 1% threshold are

also investigated to check whether it is the instrumental
line which causes the high statistic value. There were 293
sub-bands which were in this category, and in 291 sub-

bands the Viterbi tracks closely follow the instrumental
line, and the remaining two contained both an instru-
mental line and a hardware injection (ip5). These were
then reintroduced into the analysis as the Viterbi tracks
did not follow that of the instrumental line. From the
total of the 17 929 sub-bands, 248 were selected for a fol-
low-up investigation where 107 of these sub-bands cross
the thresholds of both the Viterbi and CNN statistics.

TABLE XIV. Outlier associated with the hardware injection ip6.

<table>
<thead>
<tr>
<th>Injection</th>
<th>FAP</th>
<th>∆f [Hz]</th>
<th>∆f [nHz/s]</th>
<th>∆δ [deg]</th>
<th>∆α [deg]</th>
<th>FAP</th>
</tr>
</thead>
<tbody>
<tr>
<td>ip6</td>
<td>4.02 × 10⁻⁸</td>
<td>1.54 × 10⁻²</td>
<td>2.26</td>
<td>36.59</td>
<td>314.54</td>
<td></td>
</tr>
</tbody>
</table>

TABLE XV. Outliers of unknown origin from the Time-Domain F-statistic analysis.

<table>
<thead>
<tr>
<th>f [Hz]</th>
<th>f [nHz/s]</th>
<th>δ [deg]</th>
<th>α [deg]</th>
<th>FAP</th>
</tr>
</thead>
<tbody>
<tr>
<td>83.52</td>
<td>-6.58 × 10⁻³</td>
<td>-18.08</td>
<td>179.16</td>
<td>0.0094</td>
</tr>
<tr>
<td>726.07</td>
<td>-3.30 × 10⁻²</td>
<td>58.07</td>
<td>190.79</td>
<td>0.0034</td>
</tr>
</tbody>
</table>
In O3 there are a total of 18 hardware injections, where 9 fall within our search parameter space and two of these (ip1 and ip5) appear in sub-bands which cross our detection threshold without being excluded. These signals appear in multiple sub-bands due to the 50% overlap, therefore the sub-band containing a larger fraction of the signal is used for followup. Two additional injections outside of our ‘sensitive’ range for \( f \) also crossed our detection thresholds (ip4 and ip6) as SOAP identified the part of the signal which crossed the search band. Of the 7 injections we did not detect, two are in binary systems which we are less likely to detect as this search was optimised for isolated NSs. The remaining missed injections have SNRs which are below our expected sensitivity for isolated NSs, therefore would not be expected to cross our threshold. The two remaining hardware injections crossed the detection threshold for both the Viterbi statistic or the CNN statistic. These candidates were then followed up using the parameter estimation method described in Sec. IV D 5, where we correctly recover the injected parameters of the injections.

3. Sensitivity

The sensitivity of SOAP can be tested by running the search on a set of CW signals injected into real O3 data. A total of \( 3.3 \times 10^4 \) signals are injected across each of the four frequency bands described in Sec. IV D 1, where the signals have Doppler parameters which are drawn uniformly on the sky, uniformly within the respective frequency range and uniformly in the range \([-10^{-9}, 0]\) Hz s\(^{-1}\) for the frequency derivative. The other amplitude parameters varied in the same ranges as described in Sec. IV D 3. A false alarm value of 1% can be set for each of the odd and even data-sets within the four analysis bands by taking the corresponding statistic value at which 1% of the noise only bands exceed. Both the Viterbi and CNN statistics are calculated separately for each of the odd and even bands. Each of the bands containing injected signals can then be classified as detected or not depending on if a statistic crossed its respective false alarm value. These classified statistics can then be combined together to produce an efficiency curve shown in Fig. 14, which show the fraction of detected signals at a given sensitivity depth, defined in Eq. 17. At a false alarm value of 1% and a detection efficiency of 95% we are sensitive to signals with a depth of 9.9, 8.0, 6.5 and 5.3 Hz\(^{-1/2}\) for the frequency bands 40-500, 500-1000, 1000-1500 and 1500-2000 Hz respectively. To further investigate our sensitivity, we split each of the four analysis bands into smaller bands ranging from 20 Hz wide at lower frequency to 100 Hz wide at higher frequencies. For each of these bands a detection efficiency curve is generated in the same way as for the sensitivity depth above, however, they are now generated for values of \( h_0 \) the false alarm values for each band are set based on which of the four larger analysis bands that it falls within. Our false alarm values are then contaminated by the strongest artefacts within each 500 Hz wide analysis band, meaning that this is a conservative estimate of our sensitivity. The error on these curves is found using the binomial error on each of the points as defined in Eq. 19, giving two bounds on our efficiency curves. Values of \( h_0 \) for each frequency band can then be selected where the detection efficiency reaches 95%, defining our sensitivity shown in Fig. 15.

VI. CONCLUSIONS

In Fig. 15 we summarize 95% confidence-level upper limits on strain amplitude \( h_0 \) for the four pipelines used in this search. The upper limits obtained improve on those obtained using the PowerFlux method in early O3 LIGO data [39]. Our results constitute the most sensitive all-sky search to date for continuous GWs in the range 20-2000 Hz while probing spin-down magnitudes as high as \( 1 \times 10^{-8} \) Hz/s. Only the O2 Falcon search [37, 38, 90] provides a better sensitivity in the frequency range 20-2000 Hz; however it does so with a dramatically reduced frequency derivative range. In the frequency range of \([20, 500]\) Hz Falcon searches a \( f \) range from \(-3 \times 10^{-13}\) Hz/s to \(3 \times 10^{-13}\) Hz/s and \( f \) range up to \([-7.5 \times 10^{-12}, 3 \times 10^{-12}]\) Hz/s for frequencies above 500 Hz. Thus the Falcon search parameter space is smaller than ours by factor of \( \sim 1.8 \times 10^4 \) below 500 Hz and factor of \( 10^3 \) above 500 Hz. A recent search for persistent narrowband gravitational waves using radiometer analysis of combined O1, O2, and O3 LIGO and Virgo data in the frequency range...
of 20 - 1726 Hz [91] has not revealed any significant signals and has reported upper limits on an equivalent strain amplitude in the range of (0.030 – 9.6) × 10^{-24}. As briefly discussed in [91], the radiometer search is expected to be significantly less sensitive than our CW searches for two reasons. First, the former uses frequency bins much larger than the latter (1/32 Hz vs O(mHz)), thus collecting more noise in each bin. Second, it does not take into account the Doppler effect due to the Earth motion, which causes a spread of the signal power over several bins (especially at higher frequencies), thus producing a further sensitivity loss.

We can use the amplitude $h_0$ given by Eq. (2) to calculate star’s ellipticity $\epsilon$,

$$
\epsilon = \frac{c^4}{4\pi^2 G} \frac{h_0 d}{I_{zz} f^2} \approx 9.46 \times 10^{-6} \left( \frac{h_0}{10^{-25}} \right)
\times \left( \frac{10^{38} \text{ kg m}^2}{I_{zz}} \right) \left( \frac{100 \text{ Hz}}{f} \right)^2 \left( \frac{\text{1 kpc}}{d} \right). \tag{22}
$$

Using the above equation the upper limits on the GW strain amplitude $h_0$ can be converted to upper limits on the ellipticity $\epsilon$. The results are plotted in Fig. 16 (left panel) for four representative values of the distance $d$ and they provide astrophysically interesting results. The NSs with ellipticities above a given trace and distance value corresponding to the trace in the left panel of Fig. 16 would be detectable by our searches. For instance, at frequency 200 Hz we would be able to detect a GW signal from a NS within a distance of 100 pc if its ellipticity were at least $3 \times 10^{-7}$. Similarly, in the middle frequency range, around 550 Hz, we would be able to detect the CW signal up to a distance of 1 kpc, with $\epsilon > 5 \times 10^{-7}$. Finally at higher frequencies, around 1550 Hz, the same signal would be detectable up to a distance of 10 kpc if $\epsilon > 2 \times 10^{-6}$. These levels of ellipticity are below the maximum value of the ellipticity that may be supported by the crust of a NS described by a standard equation of state reported in [92–94]. However they are above the most recent estimates in general relativity by [95, 96]. The latter do not, however, exclude larger values of ellipticity when additional physical processes, such as plastic flow in the crust, are taken into account. Our upper limits are starting to probe the range predicted for pulsars by the models of [97], which predict ellipticities up to $\epsilon \approx 10^{-7} - 10^{-6}$ for younger stars in which the deformation is not supported by crustal rigidity, but by a non-axisymmetric magnetic field at the end of its Hall driven evolution in the crust. Note however that for known pulsars at a distance of a few kpc, such as the Crab, the signal would be at frequencies $f \lesssim 100$ Hz, so still beyond the reach of our searches.

Another way of representing limits on ellipticity is shown in the right panel of Fig. 16. Assuming that the emission of gravitational radiation is the sole energy loss mechanism for a rotating NS, we obtain the so-called spin-down limit $h_{0}^{\text{sd}}$ on the amplitude $h_0$, see Eqs. (7)–(9) of [98]:

$$
h_{0}^{\text{sd}} = \frac{1}{d} \left( \frac{5 \ G I_{zz} |f|}{2 \ c^5} \right)^{1/2} \approx 2.55 \times 10^{-25} \left( \frac{1 \text{ kpc}}{d} \right)
\times \left( \frac{I_{zz}}{10^{38} \text{ kg m}^2} \right)^{1/2} \left( \frac{100 \text{ Hz}}{f} \right)^{1/2} \left( \frac{|f|}{10^{-11} \text{ Hz s}^{-1}} \right)^{1/2}. \tag{23}
$$

Inverting the above equation and replacing the spin-down limit amplitude $h_{0}^{\text{sd}}$ with our upper limit amplitudes $h_0$ we have the following relation between the frequency derivative and frequency:

$$
|f| = \frac{2c^3 (h_{0}^{5\text{\%}} d)^2 f}{5G I_{zz}} \approx 1.54 \times 10^{-16} \left( \frac{h_0^{95\text{\%}}}{10^{-24}} \right)^2
\times \left( \frac{I_{zz}}{10^{38} \text{ kg m}^2} \right) \left( \frac{f}{100 \text{ Hz}} \right) \left( \frac{d}{\text{1 kpc}} \right)^2. \tag{24}
$$

In the right panel of Fig. 16 we have plotted $|f|$ as a function of frequency $f$ for several representative values of the distance $d$ and for a canonical value of the moment of inertia. The NSs with $|f|$ above a given trace and distance value corresponding to the trace in the right panel of Fig. 16 would be detectable by our searches.

By equating Eq. (2) for the amplitude $h_0$ and Eq. (23) for the spin-down limit, we obtain the following equation for $f$:

$$
|f| = \frac{32\pi^4 G}{5c^9} c^2 I_{zz} f^5 \approx 1.72 \times 10^{-14} \left( \frac{\epsilon}{10^{-6}} \right)^2
\times \left( \frac{I_{zz}}{10^{38} \text{ kg m}^2} \right) \left( \frac{f}{100 \text{ Hz}} \right)^5. \tag{25}
$$

The dashed lines in the right panel of Fig. 16 are constant ellipticity curves from Eq. (25) above. These lines are independent of the distance $d$.

In addition to constraints on ellipticities of isolated NSs, we can make statements about the rate and abundance of inspiraling planetary-mass and asteroid-mass PBHs [29]. The upper limits presented in Fig. 15 are generic: they can be applied to any quasi-monochromatic, persistent GW that follows a linear frequency evolution over time and whose frequency derivative lies within the search range. Based on these all-sky searches, GW signals from inspiralling PBH binaries with chirp masses less than $\mathcal{O}(10^{-5}) M_\odot$ and GW frequencies less than $\sim 250$ Hz would be identical to those arising from non-axisymmetric rotating NSs. Following the procedure presented in [30], and using the FrequencyHough upper limits in Fig. 5, which cover the widest range of spin-down/spin-up, we obtain constraints on highly asymmetric mass ratio binary systems, assuming that one object in the binary has a mass $m_1 = 2.5 M_\odot$, motivated by the QCD phase transition [11, 26, 99]. In Fig. 17, we plot constraints on the merging rates and an effective parameter,
FIG. 15. Comparison of broadband search sensitivities obtained by the FrequencyHough pipeline (black triangles), the SkyHough pipeline (red squares), the Time-Domain $F$-statistic pipeline (blue circles), and the SOAP pipeline (magenta diamonds). Vertical bars mark errors of $h_0$ obtained in the procedures described in Sects. IV and V. Population-averaged upper limits obtained in [102] using the O3a data are marked with dark-green crosses.

The “population average” upper limit formula given in Eq. 6 has been derived in [45]. It assumes an underlying population of sources randomly distributed in the sky, with a uniform distribution of the polarization angle $\psi$ and of the cosine of the star’s rotation axis inclination angle, $\iota$, w.r.t. the line of sight. We show here how to obtain the scaling factor to be applied for a specific set of source parameters. The relevant term, which contains the dependence on the source parameters, is - see Eq. (B15) in [45]:

$$S^2 = (A_+ F_+ + A_x F_x)^2.$$ (A.1)

$F_+$, $F_x$ are the time-dependent detector beam pattern functions, which can be expressed as

$$F_+(t) = a(t) \cos 2\psi + b(t) \sin 2\psi$$

$$F_x(t) = b(t) \cos 2\psi - a(t) \sin 2\psi$$ (A.2)
FIG. 16. Left panel: detectable ellipticity, given by Eq. (22), as a function of the GW frequency for neutron stars with the ‘canonical’ moment of inertia $I_{zz} = 10^{38}$ kg m$^2$ at a distance of 10 kpc, 1 kpc, 100 pc, and 10 pc (from top to bottom). Results for the FrequencyHough pipeline are marked in black, SkyHough in red and for Time-Domain $F$-statistic in blue. The right panel shows the relation between the absolute value of the first GW frequency derivative $|\dot{f}|$ and the GW frequency $f = 2f_{\text{rot}}$ (with $f_{\text{rot}}$ the rotational frequency) of detectable sources as a function of the distance, assuming their spin-down is due solely to the emission of GWs. Constant spin-down ellipticities $\epsilon_{sd}$ corresponding to this condition, are denoted by dashed green curves. The magenta horizontal line marks the maximum spin down searched.

FIG. 17. Constraints on $\tilde{f}$, a quantity that, if less than one, indicates the sensitivity to a given $f_{\text{phb}}$, and inspiraling rate (color) as a function of the secondary mass, with a primary mass $m_1 = 2.5M_\odot$, assuming a monochromatic mass function for $m_1$, no rate suppression, and $f_{\text{phb}} = 1$. These constraints are valid at distances of $\mathcal{O}(pc)$.

where $a(t), b(t)$ explicit expressions are given e.g. in [47]. The terms $A_+, A_\times$ are given by

$$A_+ = \frac{1 + \cos^2 \iota}{2}, \quad A_\times = \cos \iota.$$  \hspace{1cm} (A.3)

Taking the average over all the source parameters, it can be found

$$\mathcal{S}^2_{\alpha, \delta, \psi, \iota} = <F^2>_{\alpha, \delta, \psi, \iota} \simeq \frac{4}{25},$$ \hspace{1cm} (A.4)

nearly independent of the specific detector being considered. If we consider a specific set of source parameters $(\alpha, \delta, \psi, \iota)$, the only average we need to compute is over the time and we can write

$$\mathcal{S}^2_{t} = <F^2_t>_{t} \left( \frac{1 + \cos^2 \iota}{2} \right)^2 + <F^2_\times>_{t} \cos^2 \iota$$ \hspace{1cm} (A.5)

The scaling factor by which the population average upper limit must be multiplied, in order to refer it to a specific set of source parameters, is

$$C = \sqrt{\frac{\mathcal{S}^2_{\alpha, \delta, \psi, \iota}}{\mathcal{S}^2_{t}}}$$ \hspace{1cm} (A.6)

ACKNOWLEDGMENTS

This material is based upon work supported by NSF’s LIGO Laboratory which is a major facility fully funded by the National Science Foundation. The authors also gratefully acknowledge the support of the Science and Technology Facilities Council (STFC) of the United Kingdom, the Max-Planck-Society (MPS), and the State of Niedersachsen/Germany for support of the construction of Advanced LIGO and construction and operation of the GEO 600 detector. Additional support for Advanced LIGO was provided by the Australian Research Council. The authors gratefully acknowledge the Italian Istituto Nazionale di Fisica Nucleare (INFN), the French Centre National de la Recherche Scientifique (CNRS) and the Netherlands Organization for Scientific Research (NWO), for the construction and operation of the Virgo
detector and the creation and support of the EGO consortium. The authors also gratefully acknowledge research support from these agencies as well as by the Council of Scientific and Industrial Research of India, the Department of Science and Technology, India, the Science & Engineering Research Board (SERB), India, the Ministry of Human Resource Development, India, the Spanish Agencia Estatal de Investigación (AEI), the Spanish Ministerio de Ciencia e Innovación and Ministerio de Universidades, the Conselleria de Fons Europeus, Universitat i Cultura and the Direcció General de Política Universitaria i Recerca del Govern de les Illes Balears, the Conselleria d’Innovació, Universitats, Ciència i Societat Digital de la Generalitat Valenciana and the CERCA Programme Generalitat de Catalunya, Spain, the National Science Centre of Poland and the European Union – European Regional Development Fund; Foundation for Polish Science (FNP), the Swiss National Science Foundation (SNF), the Russian Foundation for Basic Research, the Russian Science Foundation, the European Commission, the European Social Funds (ESF), the European Regional Development Funds (ERDF), the Royal Society, the Scottish Funding Council, the Scottish Universities Physics Alliance, the Hungarian Scientific Research Fund (OTKA), the French Lyon Institute of Origins (LIO), the Belgian Fonds de la Recherche Scientifique (FRS-FNRS), Actions de Recherche Concertées (ARC) and Fonds Wetenschappelijk Onderzoek – Vlaanderen (FWO), Belgium, the Paris Ile-de-France Region, the National Research, Development and Innovation Office Hungary (NKFIH), the National Research Foundation of Korea, the Natural Science and Engineering Research Council Canada, Canadian Foundation for Innovation (CFI), the Brazilian Ministry of Science, Technology, and Innovations, the International Center for Theoretical Physics South American Institute for Fundamental Research (ICTP-SAIFR), the Research Grants Council of Hong Kong, the National Natural Science Foundation of China (NSFC), the Leverhulme Trust, the Research Corporation, the Ministry of Science and Technology (MOST), Taiwan, the United States Department of Energy, and the Kavli Foundation. The authors gratefully acknowledge the support of the NSF, STFC, INFN, CNRS and PL-Grid for provision of computational resources.

This work was supported by MEXT, JSPS Leading-edge Research Infrastructure Program, JSPS Grant-in-Aid for Specially Promoted Research 26000005, JSPS Grant-in-Aid for Scientific Research on Innovative Areas 2905: JP17H06358, JP17H06361 and JP17H06364, JSPS Core-to-Core Program A. Advanced Research Networks, JSPS Grant-in-Aid for Scientific Research (S) 17H06133 and 20H05639 , JSPS Grant-in-Aid for Transformative Research Areas (A) 20A203: JP20H05854, the joint research program of the Institute for Cosmic Ray Research, University of Tokyo, National Research Foundation (NRF), Computing Infrastructure Project of KISTI-GSDC, Korea Astronomy and Space Science Institute (KASI), and Ministry of Science and ICT (MSIT) in Korea, Academia Sinica (AS), AS Grid Center (ASGC) and the Ministry of Science and Technology (MoST) in Taiwan under grants including AS-CDA-105-M06, Advanced Technology Center (ATC) of NAOJ, and Mechanical Engineering Center of KEK.

We would like to thank all of the essential workers who put their health at risk during the COVID-19 pandemic, without whom we would not have been able to complete this work.


F. Acernese et al. Calibration of Advanced Virgo and reconstruction of detector strain h(t) during the Observing Run O3. submitted to CQG, 2021.


Concordia University Wisconsin, Mequon, WI 53097, USA
Department of Electrophysics, National Yang Ming Chiao Tung University, Hsinchu, Taiwan
Department of Physics, Rikkyo University, Toshima-ku, Tokyo 171-8501, Japan

(Dated: June 17, 2022)

* Deceased, August 2020.
† Deceased, April 2021.