Semicoherent DM-off Veto for Continuous Gravitational Wave Searches

IISER PUNE

A thesis submitted towards partial fulfilment of
BS-MS Dual Degree Programme

by

DIVYA SINGH
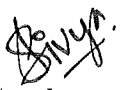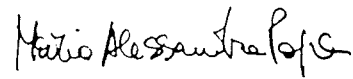20131107

under the guidance of

DR. MARIA ALESSANDRA PAPA

Max Planck Institute for Gravitational Physics
Hannover

Indian Institute of Science Education and Research
Pune

Certificate

This is to certify that this thesis entitled "Semicoherent DM-off veto for Continuous Gravitational Wave Searches" submitted towards the partial fulfilment of the BS-MS dual degree programme at the Indian Institute of Science Education and Research Pune represents original research carried out by Divya Singh at the Max Planck Institute for Gravitational Physics, Hannover, under the supervision of Dr. Maria Alessandra Papa and Dr. Sylvia Zhu during the academic year 2017-2018.

Supervisor
DR. MARIA
ALESSANDRA
PAPA

Co-Supervisor
DR. SYLVIA ZHU
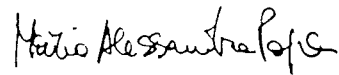
Student
DIVYA SINGH
20131107

Declaration

I hereby declare that the matter embodied in the report entitled "Semicoherent DM-off veto for Continuous Gravitational Wave Searches" are the results of the investigations carried out by me at the Max Planck Institute for Gravitational Physics, Hannover under the supervision of Dr. Maria Alessandra Papa and Dr. Sylvia Zhu, and the same has not been submitted elsewhere for any other degree.

Supervisor
DR. MARIA
ALESSANDRA
PAPA

Student
DIVYA SINGH
20131107

Co-Supervisor
DR. SYLVIA ZHU

Acknowledgements

Lastly, I thank my mother, and my sister for all their support and encouragement that has made it possible for me to leave the nest and pursue my academic goals.

Abstract

The recent all-sky low-frequency Einstein@Home search for continuous gravitational waves in the data from the first observing run (O1) of Advanced LIGO had > 6000 surviving candidates at the end of the last (most sensitive) follow-up stage [1]. In order to identify and rule out candidates due to detector artefacts, the DM-off veto was developed [2]. A coherent version of the DM-off veto rejected more than 99.9% of the surviving candidates. In this work, we investigate whether a semicoherent DM-off veto could exclude noise candidates at earlier stages of the search. We characterise the veto on this all-sky search and were able to exclude > 75% of the 36248 candidates from the first stage of the all-sky search using the semicoherent version of the veto. We find that we could save over 20000 hours of computational time by using the veto at this stage. We further optimise the DM-off search setups for the characterisation of the veto. It is observed that the DM-off setup can be made coarser while maintaining its effectiveness for this all-sky search. Lastly, we explore different DM-off search setups for characterising and applying the semicoherent DM-off veto to candidates from a recent directed search for continuous waves from three astrophysical sources — Vela Jr., Cassiopeia A, and G347.3 — in LIGO O1 data.

Contents

List of Figures

4

List of Tables

Chapter 1

Introduction

1.1 Gravitational Waves and their Detection

Gravitational waves can be described as ripples in the curvature of space that travel at the speed of light c. They manifest as solutions to Einstein's equations in the weak field limit, which makes their detection a successful test of Einstein's Theory of General Relativity. Gravitational waves are produced due to changes in the gravitational field of moving masses.

The potential sources for gravitational wave emission can be classified into short-lived sources and long-lived sources. In the case of short-lived sources, we have mergers of compact objects like BH-BH mergers, NS-NS mergers or BH-NS mergers (BH: Black Hole; NS: Neutron star). Other potential short-lived sources include supernova explosions. Long-lived gravitational wave sources include spinning neutron stars, and the stochastic gravitational wave background. [3].

The Laser Interferometer for Gravitational-Wave Observatory (LIGO) was built to detect these ripples in the curvature of spacetime. LIGO consists of two ground-based detectors which are essentially giant Michelson Interferometers. These detectors are located in Livingston, Louisiana (L1), and Hanford, Washington (H1) in the United States.

How do interferometers detect gravitational waves? In an interferometer, a laser beam passes through a beam splitter that transmits one half beam through a straight arm and reflects the other half through the perpendicular arm. The two beams have correlated phases and get reflected by mirrors at the end of the two arms. These reflected beams interfere when they return

to the photodetector. Any change in the phase results in a change in the interference pattern. In an interferometer, if the proper length of the two arms is equal, then the two split beams interfere constructively. On the other hand, if the arms differ by some length, the beams may interfere destructively. The interferometer uses one of the arms as the stable clock by acting as the reference for the return time for the other arm. A passing gravitational wave causes these arm lengths to change. This results in a change in the phase of the laser beams which causes a change in the interference pattern which is analysed to obtain the strain of the gravitational wave signal.

A gravitational wave causes a change in the separation between two particles. The relative change produced by the gravitational wave in the separation between the two particles is called the strain produced by the gravitational wave. This change is proportional to the initial separation of the particles, which implies that a wave will cause a greater change in separation if the initial separation is large. For this reason, the arm lengths of the existing ground based detectors are huge with LIGO detectors having arm lengths of 4 km.

LIGO detectors are highly specialised Michelson interferometers. Advanced LIGO will have 10 times better strain sensitivities over a broad frequency band than the initial detectors [4]. However, these strain sensitivities are limited by the presence of different noise sources. Sources of noise include quantum noise in the laser beam, thermal noise due to the suspensions, the test masses and the mirror coatings, and seismic noise. These sources of noise have been minimised using highly specialised detector components and highly coherent laser technology. All primary optical components in the vacuum cavities are suspended by pendulum systems to isolate the components from seismic vibrations and thermal noise. These detectors have Fabry-Perot cavities in both arms to increase the carrier power and phase shift for a given change in arm length. This makes the detectors highly sensitive so that they can detect changes in length of orders of magnitude less than 10^{-20} m. However, other unknown sources of disturbance also affect the detectors. Environmental factors, random fluctuations in atmospheric conditions in the cavities as well as glitches during data transfer also affect the detector data.

1.2   Continuous Gravitational Waves

Continuous gravitational waves are periodic, long-lived, and quasi-monochromatic signals. Rapidly rotating neutron stars that are non-axisymmetric are the

most probable sources for continuous waves that can be detected by ground-based detectors. These signals are expected to be orders of magnitude weaker than the gravitational wave signals from compact binary coalescences (for example, GW150914 [5], and GW151226 [6]). Hence, their detection by ground-based detectors is very challenging.

To put this in perspective, we look at the strain amplitudes of continuous waves expected from these sources. For a star that has quadrupole asymmetry and is at a distance r away, spinning about its symmetry axis of rotation (z) with frequency f_{rot}, the expected intrinsic strain amplitude of gravitational radiation, h_o is given by

$$h_o = \frac{4\pi^2 G I_{zz} f_{GW}^2}{c^4 \ r} \epsilon \tag{1.1}$$

where I_{zz} is the moment of inertia of the star along its symmetry axis of rotation, f_{GW} is the gravitational wave frequency, $f_{GW} = 2f_{rot}$ for quadrupole emission, and ϵ is the ellipticity of the star which is defined as

$$\epsilon = \frac{I_{xx} - I_{yy}}{I_{zz}} \tag{1.2}$$

For a neutron star with observed frequency f, and spindown (the first time derivative of frequency) $\dot{f}$, the maximum strain that could be detected on Earth can be calculated by

$$h = \frac{1}{r}\sqrt{-\frac{5}{2}\frac{G}{c^3}I_{zz}\frac{\dot{f}_{GW}}{f_{GW}}} = (2.5{\times}10^{-25})\left(\frac{1kpc}{r}\right)\sqrt{\left(\frac{1kHz}{f_{GW}}\right)\left(\frac{-\dot{f}_{GW}}{10^{-10}Hz/s}\right)\left(\frac{I_{zz}}{I_o}\right)} \tag{1.3}$$

Here, $I_o = 10^{38} kg \ m^2 (10^{45} g \ cm^2)$ is a nominal moment of inertia of a neutron star. The maximum limit on the strain given by equation 1.3 is of the order 10^{-25}. This strain would produce a change of the order 10^{-22} metres in the length of the arms of the ground-based detectors, which is 10^{12} times smaller than the radius of an atom. Hence, the detectors have to be extremely sensitive to be able to note such minuscule changes in length.

1.2.1 The expected signal waveform

Consider a wave reference-frame transverse to the direction of propagation of a gravitational wave. In this reference frame, the two polarizations of the

wave have the form given by equation 1.4.

$$h_+(t) = A_+ cos\Phi(t)$$
$$h_\times(t) = A_\times sin\Phi(t) \qquad (1.4)$$

In these equations, $A_{+,\times}$ are the amplitudes, while $\Phi(t)$ is the phase of the wave at time t. For signals that we expect to detect with LIGO, $A_{+,\times}$ are in general constant or slowly varying with time, while the phase $\Phi(t)$ varies rapidly with time.

Equation 1.5 gives the form for $\Phi(t)$, where t is the detector time when the wavefront that passed the solar system barycentre (SSB) at τ_{SSB} arrives at the detector.

$$\Phi(t) = \Phi_o + 2\pi[f_o(\tau_{SSB} - \tau_{oSSB}) + \frac{1}{2}\dot{f_o}(\tau_{SSB} - \tau_{oSSB})^2 + \frac{1}{6}\ddot{f_o}(\tau_{SSB} - \tau_{oSSB})^3 + ...]$$
$$(1.5)$$

We are interested in the phase of the signal when it reaches the detectors on Earth. For this we want to find out how the arrival time at the detector is related to τ_{SSB}. If the arrival time at the detector is given by t, the position vector of the detector in the SSB frame is given by $\mathbf{r}(t)$, and $\mathbf{n}$ is the unit vector pointing to the source, then the transformation between the detector arrival time t, and SSB time τ_{SSB} is given by

$$\tau_{SSB} = t + \frac{\mathbf{r}(t).\mathbf{n}}{c} + \Delta_{E\odot} - \Delta_{S\odot} \qquad (1.6)$$

where c is the speed of light, and $\Delta_{E\odot}$ and $\Delta_{S\odot}$ are Einstein and Shapiro time delays.

A rapidly rotating neutron star produces gravitational waves with amplitudes that are constant over time. If i is the angle between the star's total angular momentum and its direction to the Earth, then the amplitudes of gravitational waves generated by the stars is given by equation 1.7.

$$A_+ = \frac{1}{2}h_o(1 + cos^2 i)$$
$$A_\times = h_o cos\, i \qquad (1.7)$$

where h_o is the intrinsic gravitational wave strain (amplitude).

Now, the detectable continuous wave signal $h(t)$ is a superposition of the two polarizations of the gravitational wave and is given by

$$h(t) = F_+(\alpha, \delta, \psi; t)h_+(t) + F_\times(\alpha, \delta, \psi; t)h_\times(t) \qquad (1.8)$$

11

where F_+ and $F_\times$ are the beam pattern functions for the two detectors — these are functions of the sky position of the source given by the right ascension α and the declination δ, and the orientation of the wave-frame with respect to the detector frame ψ. Due to the Earth's orbital movement around the Sun, and its diurnal rotation on its axis, the relative orientation of the source with respect to the detector changes. Therefore, the beam pattern functions vary with time. Equation 1.8 gives the waveform of the continuous gravitational wave that we expect to detect with LIGO.

1.3   Searches for Continuous Wave signals

Continuous gravitational wave signals from rapidly rotating neutron stars in our galaxy are expected to be weaker than signals from compact binary coalescences. In order to obtain a significant signal to noise ratio to make a continuous wave detection, one has to integrate over very long timespans. This is possible because continuous wave signals are long-lived and must persist over long durations of time. Since we do not know the exact signal parameters that we expect to find, we need to carry out this search for a large number of points in the parameter space. This is a computationally expensive task. Hence, it can be challenging to look for these signals in the data collected from ground based detectors at their current sensitivities because any signal that we expect to find is buried in noise. The timespan over which the data is integrated is called the coherence time, T_{coh}, of the search. The search techniques and the coherence time depend strongly on the known parameters of the source. Based on the number of known source parameters, continuous wave searches can be classified into targeted, directed and all-sky searches.

In a targeted search, the sky position and the rotation frequency of the source are known from electromagnetic observations of the star. This provides the information required to demodulate the phase and frequency of the expected signal from the source.

In a directed search, the sky position of the source is known with little to no information about the rotation frequency of the source. In such a search, the signal parameter space is larger and the search parameter ranges depend on the age of the star. The parameter space usually includes frequency, spindown (the first time derivative of frequency), and even higher order time derivatives of frequency. For a directed search that includes the second order time derivative of frequency $\ddot{f}$s, the parameter space volume is proportional

to T^6_{obs} where T_{obs} is the observation span over which phase coherence is maintained. In such a search, the computational cost escalates rapidly with T_{obs} [7].

All-sky searches are blind searches for continuous wave signals from unknown neutron stars. These searches span the whole sky. In such a search, the parameter space volume increases by a further factor of $(T_{\rm coh})^2$, and also depends quadratically on the signal frequency f [7]. Hence a fully-coherent all-sky search can quickly become very expensive.

Searches for continuous wave signals are performed in the frequency domain. For example, for Einstein@Home searches [1, 8], the time series data obtained from the detectors is Fourier-transformed in chunks of 1800 seconds. These short Fourier transforms (SFTs) are then narrow-banded, and the analysis is performed on these narrow-banded SFTs.

For a particular search, search parameters and their ranges are chosen according to the target signal. For example, the search parameters for an all-sky search include the sky position, frequency and the spindown (the first time derivative of frequency). This defines the search parameter space. The search is then performed over a grid in this parameter space. Each grid-point in the search defines a particular signal waveform and is a candidate for being a potential continuous wave signal. The detection statistic is computed at each grid-point and measures the significance of a potential signal candidate. These candidates are then ranked based on their detection statistic values.

In order to overcome the huge computational cost associated with long integration times, *semicoherent* searches were developed. In a *semicoherent* search, the observation span T_{obs} is divided into multiple segments $N_{seg} \sim T_{obs}/T_{coh}$, which are analysed coherently and then combined incoherently at the end. In each segment, the detection statistic is computed over all the points in the parameter space. The detection statistics from all segments are then combined at the end by adding the detection statistics from the different segments or by taking their average.

Hierarchical search approaches have also been been developed where the search is conducted in multiple stages [8]. Each successive stage has a longer coherence time T_{coh} and at each stage, candidates below a predetermined significance are rejected. These significance thresholds are decided based on the false alarm and false dismissal rates that are deemed acceptable to claim that a candidate is potentially due to a continuous wave signal. The surviving candidates are then followed up in the later stages by performing searches over smaller regions around the candidate using finer grids and

longer coherence times. The coherent time for searches increases at each follow up stage with fewer and fewer surviving candidates. Hence, the searches become more sensitive with each follow-up step.

As an example, consider the recent all-sky search for continuous wave signals in the data from the first observing run of Advanced LIGO (O1) using Einstein@Home [1]. This search focussed on a low frequency range of 20 to 100 Hz. The search was conducted in four stages. At each stage, the coherence time of the search increased with the last stage being a fully-coherent search. Low-significance candidates were rejected at each stage and the surviving candidates were successively followed up in consequent stages. After the last follow-up stage, a new veto was applied to the surviving candidates which were over 6000 in number. This was the coherent DM-off veto.

1.3.1 F-statistic

Continuous wave searches employ a coherent matched filtering technique which utilises the known signal waveform and calculates the likelihood that a signal is present in the data at a particular template compared to having Gaussian noise at that template point. The method is known as the F-statistic in the case of continuous wave searches [9]. Using this statistic, we essentially build the signal-to-noise ratio by looking for signal shapes in the data that match the known waveform.

For gravitational wave data $x(t)$ containing a signal $h(t)$ and noise $n(t)$, the log of the likelihood ratio of having a signal at a given template to having Gaussian noise at that template is

$$ln \ \Lambda = ln \ \frac{p(x|signal + noise)}{p(x|noise)}$$

When the likelihood Λ is maximised over the amplitude parameters of the signal, we get the F-statistic. The amplitude parameters include the signal amplitudes, the polarisation angle, and the initial signal phase. The F-statistic then becomes a function of the Doppler parameters of the signal — sky position, frequency, and the time derivatives of frequency. In a continuous gravitational wave search, the F-statistic is computed over the parameter space chosen for a particular search. The grid-points defined by these parameters represent the search templates. At each point, the $2F$ value is computed and the templates are ranked based on the detection statistic.

In a coherent search, $2F$ is computed at each template by integrating over the whole observation time. On the other hand, in a semicoherent search, the

observation time is divided into multiple segments. $2F$ is computed at each template in each of the segments by integrating over the chosen coherence time. At the end, these values are combined to give the final $2\overline{F}$ value. One way to combine the $2F$ values is by taking their average over the number of segments [11].

1.4 The Semicoherent DM-off veto - Need and Concept

Continuous wave searches employ various statistical vetoes, for example vetoes based on the consistency of the detection statistic between the two detectors, as well as across segments in a semicoherent search (for more detail, see [12]). These vetoes identify and rule out candidates due to noise. A local disturbance, unlike a gravitational wave signal, is expected to affect the detection statistic for one of the detectors which causes the cumulative detection statistic to be high. In the case when the detection statistic for a candidate is higher for one of the detectors than the cumulative detection statistic, this candidate can be vetoed as a local disturbance. At the same time, transient disturbances must have high detection statistic values only across a few segments in a semicoherent search. These few high values can lead to a significant final detection statistic value. Comparing the detection statistics across the different segments can therefore help us identify such transient disturbances.

The upgrades from initial LIGO to Advanced LIGO have reduced the detector noise by a huge magnitude, increasing the strain sensitivities. However, new detector artefacts appear as well, some of which behave somewhat like astrophysical signals. They can be periodic, and persistent for long durations. Their behaviours and forms are diverse and it is non-trivial to predict their origin. So, how do we identify these detector artefacts? In the following subsections, I introduce a veto concept based on the Doppler modulation of frequencies that distinguishes artefacts from potential signal candidates by virtue of their origin.

1.4.1 Concept of the DM-off veto

Any potential astrophysical source of continuous waves is in relative motion with respect to the Earth due to the diurnal rotation of Earth and the orbital movement of the Earth around the Sun. A daily relative frequency

modulation of $v_{rot}/c \approx 10^{-6}$, and an annual relative frequency modulation of $v_{orb}/c \approx 10^{-4}$ takes place due to Earth's motion. Therefore, any continuous wave signal detected by a ground-based detector must be Doppler shifted in frequency. This implies that the power of an astrophysical signal is spread over a range of frequencies detected by the ground-based detector. This Doppler modulation of signal frequency is taken into account while developing algorithms for continuous wave searches. The detector frequencies are Doppler demodulated while computing the detection statistics at each signal template.

Detector artefacts are of terrestrial origin. The movement of Earth is not expected to modulate their frequency. Hence, they gain significance and become louder on turning off the demodulation in a search. On the other hand, astrophysical signal candidates lose significance when the demodulation is turned off because the power of the signal remains distributed across a range of detector frequencies. Fig 1.1 shows how a detector artefact becomes loud in a search with the demodulation turned off (DM-off), while a signal loses its significance in the same search. This difference in behaviour was exploited in designing a new veto for eliminating noise candidates that arise due to artefacts that persist for long durations and may be periodic like potential signal candidates. This veto is called the DM-off veto.

1.4.2 Need for the DM-off veto

The data used in continuous wave searches goes through multiple stages during which frequency bands with non-Gaussianities are identified and removed before running a search. However, disturbances and artefacts still persist in the data. Certain detector artefacts can mimic continuous gravitational wave signals. They are weak, periodic, and persist for long durations of time. Hence, they, like potential signal candidates, are not ruled out in follow-up stages and continue to gain significance at each stage.

In the O1 all-sky search, more than 6000 candidates survived at the end of the last follow-up stage [1]. This stage was fully-coherent and the most sensitive search possible given the O1 data. The resulting dilemma was the large number of surviving candidates and no way to follow them up using the same data. Hence, a new veto method was developed that could differentiate between detector artefacts and potential signal candidates. This veto was called the DM-off veto [2].

A coherent DM-off veto was applied to the 6000 plus surviving candidates of the O1 all-sky search which resulted in four surviving candidates. These
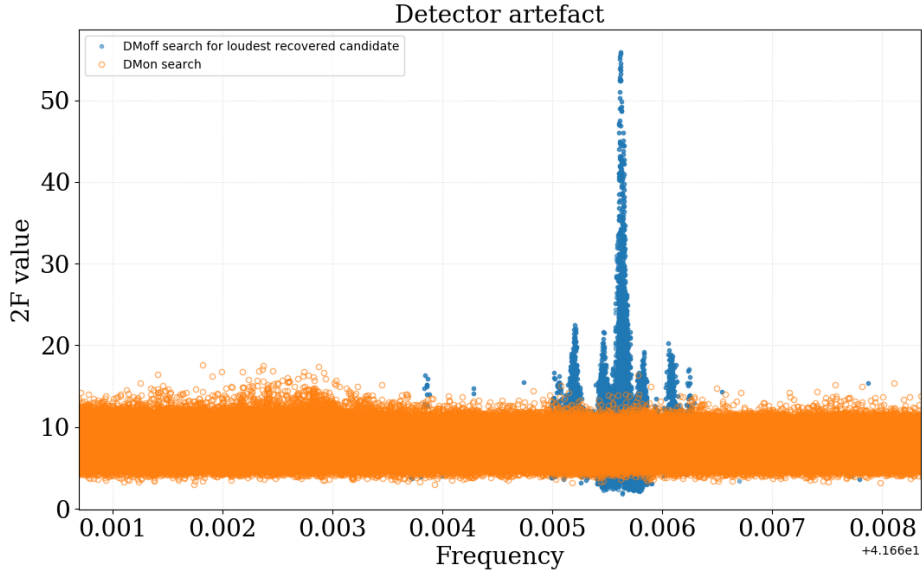
a. Detector artefact
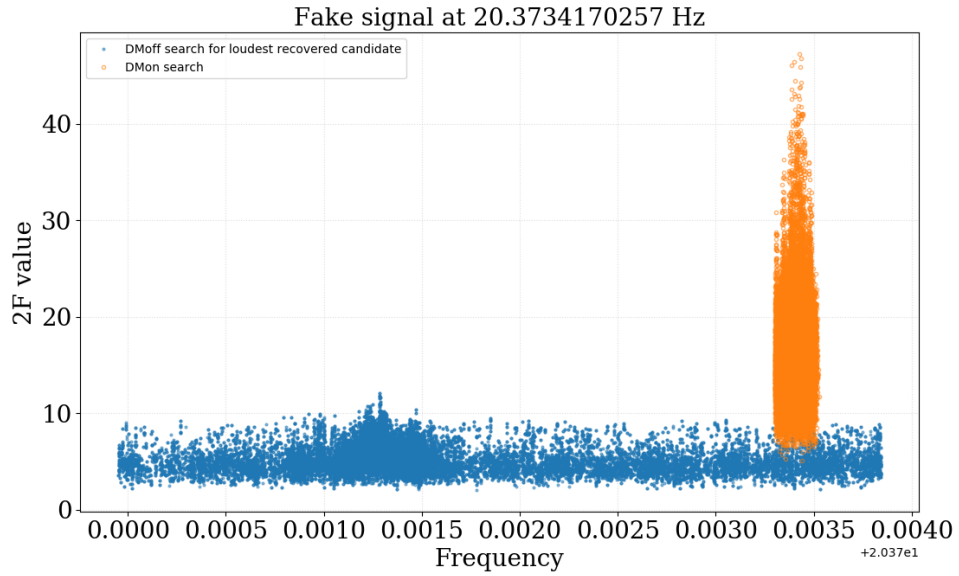
b. Simulated Signal

Figure 1.1: Difference in behaviour of a detector artefact and a simulated astrophysical signal in searches with demodulation turned on and turned off in a continuous wave search. The orange dots represent candidates from a demodulation on (DM-on) search. The blue dots represent candidates from a demodulation off (DM-off) search. A detector artefact gets louder in a DM-off search while a signal gets quieter.

17

candidates were later ruled out through other tests. However, the veto successfully eliminated more than 99% of the candidates when applied after the most sensitive follow-up stage. [1]

The next step in the development of the DM-off veto was to investigate the effectiveness of the veto if applied right after the initial stage of the search. If the veto could be applied at an earlier stage, more candidates caused by detector artefacts that passed the statistical thresholds at later stages could be identified and rejected early on in the search. This would reduce the number of candidates to be followed-up and save computational resources. Since the initial stages of the search are semicoherent, a semicoherent version of the DM-off veto was required. In this thesis, I present the development and characterisation of the semicoherent DM-off veto.

1.5 Scope of the thesis

I explore three broad questions in my work:

- How effective would a semicoherent version of the DM-off veto be when applied after an earlier stage of the hierarchical search?

- How much computational time could be saved by the application of the veto at that stage?

- How could the veto be optimised and modified for future continuous wave searches?

In order to answer the first question, the veto was characterised on the O1 all-sky search. We investigated what would happen if the veto had been applied right after Stage 0 (the first stage) of the O1 all-sky search. Using simulated signals and their behaviour, a rejection criteria for "non-signal-like" candidates was developed based on the difference between the DM-on (standard continuous wave search) and DM-off detection statistics. Simulated signals were used in the study because the expected waveforms for continuous waves are known, whereas the detector artefacts could arise due to any arbitrary causes and cannot be simulated. This rejection criterion was applied to Stage 0 candidates to see how many candidates would have to be followed-up in the later stages. Over 75% candidates were rejected using the semicoherent DM-off veto.

The second question was answered by computing the time taken by the semicoherent veto procedure and comparing that to the total computational times

of the searches in the follow-up stages of the O1 all-sky search. It was observed that a significant computational gain is possible by the application of the DM-off veto earlier in the search. If a greater number of candidates are rejected early on, the time taken during follow-up stages can reduce appreciably.

Search setup optimisation studies were then conducted to determine how coarse we could make the search grids in order to save computational time while still effectively identifying detector artefacts. For this purpose, different search setups were tested for DM-off searches on known artefacts. The changes in the detection statistic values as well as the computational runtimes were calculated for each setup to measure the gain of using a particular setup.

The semicoherent DM-off veto was then optimised for a directed search in O1 data. The search focussed on three astrophysical sources — Vela Jr., Cassiopaeia A, and G347.3. The search parameters for this search were different from the O1 all-sky search parameters. In this search, the sky position of the sources was known. However, the second time derivative of frequency $\ddot{f}$ was added to the search parameter space. Different search setups were tested for each source to ascertain how coarse the DM-off searches could be made with respect to the original search setups in order to effectively rule out artefacts but also minimise computational expense.

The proceeding chapters summarise the methods and results from studies conducted to answer the questions. The exploratory studies presented in the proceeding chapters were conducted on the supercomputing cluster, ATLAS at the Albert Einstein Institute, Hannover. Continuous wave searches were simulated using tools from the LAL/LALAPPS software suite [10].

Chapter 2

How effective is the semicoherent version of the DM-off veto?

In searches for continuous waves, new veto procedures are frequently developed to identify and exclude detector artefacts that mimic astrophysical signals. When large-scale changes are made to the detectors, for example the upgrade from Initial LIGO to Advanced LIGO, the overall performance of the detectors improves. However, new detector artefacts may get introduced in the data as well. The effect of these artefacts was observed in the recent Einstein@Home all-sky search in the data from the first observing run (O1) of Advanced LIGO [1]. Hence, the DM-off veto was devised as a tool to identify and eliminate these artefacts.

2.1 O1 all-sky search

The O1 all-sky search was performed in the low frequency range corresponding to 20 - 100 Hz. The search was conducted in four stages. The first stage, Stage 0 of the search, was conducted on the Einstein@Home volunteer-based computing system. This stage explored about 10^{17} parameter space points with a coherence time T_{coh} of 210 hours and the observation span divided into 12 segments. Stage 0 was followed by three follow-up stages which zoomed in on high-significance candidates using successively longer coherent times and greater sensitivities. At each stage, candidates that did not pass a certain detection threshold were rejected. Over 6000 candidates survived at the end of the third follow-up stage which was fully coherent and the most sensitive search possible with the given data. In order to further rule out spurious

candidates, the coherent DM-off veto was developed [2]. Only four candidates survived this veto. These candidates were later ruled out by repeating the search on a different data set.

Continuous wave searches become more sensitive, but also more expensive as the integration time increases. For this reason, coherent searches are more expensive than semicoherent searches. Therefore, a coherent DM-off search is computationally more expensive than a semicoherent DM-off veto. If we could reject candidates of non-astrophysical origins at an earlier stage using the semicoherent DM-off searches, fewer candidates would have to be followed-up in the later stages. This would result in saving computational time in the follow-up stages. Also, the coherent DM-off veto would have to be applied to a smaller number of candidates which would again save computational time.

In this chapter, the semicoherent DM-off veto is designed and characterised for the O1 low frequency all-sky search. We do this by simulating the DM-off and DM-on searches on fake signals, and developing a rejection criterion based on their behaviour. The study is performed on signals since the expected waveforms for continuous wave signals are known. An acceptance region is defined in the DM-on - DM-off detection statistic plane such that signals would not be dismissed. This rejection criterion is then applied to the surviving candidates of Stage 0 of the O1 all-sky search and the results are reported.

2.2 Characterisation of the veto for O1 all-sky search: Methods

2.2.1 Simulation of fake signals

The expected waveforms for continuous gravitational waves have been studied and are well-known. However, we do not know the behaviour and appearance of detector artefacts. These artefacts could arise due to various reasons which makes it difficult to simulate them. Therefore, we chose to study the behaviour of fake signals and characterise them to develop a criterion that would not reject signals.

The study was conducted on a thousand fake signals that were simulated in Gaussian noise. These signals were representative of the candidates from the O1 all-sky search. The frequency of these signals ranged from 20 to 100 Hz.

The other important parameters of a signal are the spindown (the first time derivative of frequency) $\dot{f}$, sky position of the source, and the strain produced by the signal h_o. These parameters were picked randomly and uniformly from their respective distributions over the parameter ranges specified in Table 2.1. For $\dot{f}$ and h_o, the values were picked log uniformly to get a uniform distribution of the order of magnitude of the two parameters. The fake signals were simulated in both Hanford (H1) and Livingston (L1) detectors.

Parameter	Value
Frequency, f	20 - 100 Hz
Spindown, $\dot{f}$	2.6×10^{-9} - 1.0×10^{-13} Hz/s
Right Ascension, α	0 to 2π rad
Declination, δ	$-\pi/2$ to $\pi/2$ rad
Strain, h_o	1.0×10^{-26} - 1.0×10^{-23}

Table 2.1: Parameters of the simulated signals.

The fake signals were created using `lalapps_Makefakedata` [10] in Gaussian noise.

2.2.2 Simulation of searches on fake signals

The main purpose of performing studies on fake signals is to simulate what happens in a standard continuous wave search in order to quantify the expected behaviour of the simulated signals. Any candidate that behaves unlike the simulated signals can then be rejected.

At each stage of a hierarchical search, a list of of the most significant signal candidates is generated with their parameters as well as their detection statistic(s). Our aim is to generate a detection statistic based on the veto concept which can be compared to the detection statistic from the standard search. For this, a modified search with the doppler demodulation turned off is performed on the fake signal. We choose to reject or accept a candidate depending on whether it behaves like a signal does, by comparing the two detection statistics. In our studies, we used standard analysis software for conducting hierarchical continuous wave searches, `lalapps_HierarchSearchGCT` [10].

2.2.2.1   DM-on search

An astrophysical signal detected by an Earth-based detector is Doppler modulated due to the rotation of the Earth around the Sun and its diurnal rotation. In a standard continuous wave search, we demodulate the data to take this effect into account (DM-on). To determine the expected behaviour of an astrophysical signal, we simulate fake signals and run a DM-on search on these simulated signals. For this, we use the same search setup as used in the original astrophysical search that produced the candidates. The DM-on search was semicoherent like the Stage 0 search, and used the same segments and search grid spacings as the original search. The DM-on search parameters for our simulations matched the search parameters of Stage 0 of the O1 all-sky search. This ensured that we treated the fake signals in the same way as real signals would be treated in Stage 0 of the O1 all-sky search.

The only difference between the Stage 0 search and the DM-on search on fake signals was the search parameter space. Stage 0 covered the whole search frequency and spindown range across the whole sky. Therefore, the data was packaged into smaller chunks that were searched over the Einstein@Home computing system. In such a system, volunteers donate idle computing hours for scientific studies that require extremely huge computational investments. The volunteers received small portions of the search space, called 'work-units', that were designed to take around 8 hours to run. For our signal injection studies, however, we only needed to search a significantly smaller region. Since we wanted our searches to recover the fake signals, the size of the DM-on search region was chosen to be comparable to that of the first follow up stage in the O1 all-sky search. This first follow-up search region had been chosen such that candidate signals in Stage 0 had a 99% chance of being recovered. The dimensions of the search region were sufficiently small for the searches to take not more than a few minutes on the supercomputing cluster. The search parameters for the DM-on searches are given in Table 2.2.

The parameters of the loudest recovered candidate from a DM-on search do not necessarily match the parameters of the simulated signal because the grid points do not always coincide with the signal frequency, spindown or the sky position. Hence, a mismatch arises between the expected detection statistic of a perfectly recovered signal and the actual observed detection statistic, which depends on the search grid parameters. Equation 2.1 gives the definition of this mismatch in the presence of noise. Here, $2F_{exact}$ is the detection statistic value that one would obtain at the exact signal parameters, and $2F_{search}$ is the detection statistic value obtained from an offset search

Parameter	Value
T_{coh}	210 hr
T_{ref}	1132729647.5 GPS s
N_{seg}	12
δf (Hz)	8.3×10^{-7}
$\delta \dot{f}_c$ (Hz/s)	1.3×10^{-11}
γ	100
Δf (Hz)	2.3×10^{-6} Hz
$\Delta \dot{f}$ (Hz/s)	1.1×10^{-10} Hz/s

Table 2.2: Search parameters for DM-on searches on simulated signals. δ refers to the grid spacings in frequency f, and spindown $\dot{f}$. Δ refers to the dimensions of the search parameter space in frequency f, and spindown $\dot{f}$. γ is the refinement factor for spindown $\dot{f}$; it quantifies the fineness of the grid in the spindown dimension used for recalculating the detection statistic of candidates with the highest significance. These parameters match the parameters of Stage 0 the O1 all-sky search [1].

template. The value of 4.0 in the denominator corresponds to the mean detection statistic value in stationary Gaussian noise.

$$mismatch = \frac{2F_{exact} - 2F_{search}}{2F_{exact} - 4.0} \qquad (2.1)$$

Mismatches are inherent in detection statistics obtained in actual continuous wave searches, and more generally, in any search performed on a discrete grid of parameter space points. Since the DM-on searches replicate a standard continuous wave search, the mismatches must be reproduced in the DM-on detection statistic values as well. Fig 2.1 shows the mismatch distribution obtained from the detection statistics for Stage 0 of the O1 all-sky search (in blue) and the mismatch distribution obtained from the DM-on searches on fake signals (in orange). These distributions are consistent with each other, indicating that our search and recovery procedure satisfactorily represents the procedure of Stage 0 of the O1 all-sky search.

2.2.2.2 DM-off search

Next, we performed searches on the fake signals with the Doppler demodulation turned off (DM-off). The DM-off semicoherent search was conducted
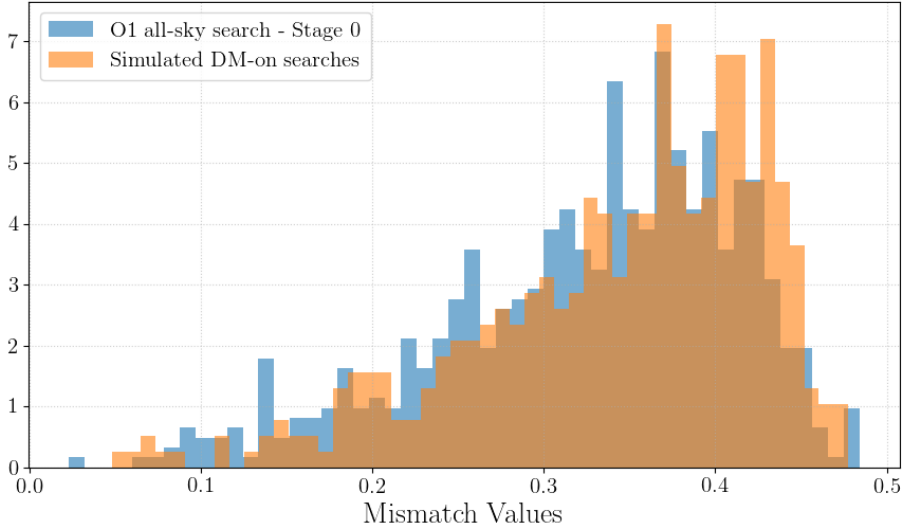
Figure 2.1: The distribution of mismatch values obtained for the O1 all-sky search and the DM-on searches simulated on the fake signals.

over a parameter space region centered around the loudest candidate recovered from the DM-on search for each simulated signal. This search was coarser in frequency, spindown, and sky position than the DM-on search because the detector artefacts are expected to be broader in frequency than astrophysical signals in general. Using the frequency, spindown and sky coordinates of the loudest candidate, we calculated the Doppler and spindown wings of each candidate. The Doppler and spindown wings represent all the detector frequencies that could have contributed to the candidate.

We use equation 2.2 to compute the spindown wings by substituting t with the start time t_{start} and end time t_{stop} of the data we are searching. For the Doppler wings, using equation 2.3, we calculate the range of detector frequencies f_d that result from the Doppler shift in the astrophysical signal frequency f_o due to the relative motion of the Earth with respect to the astrophysical source, given by v_{Earth}, in the part of its solar orbit that it covers between t_{start} and t_{stop}. The minimum and maximum frequencies from these ranges represent the bounds of detector frequencies that contributed to the search candidate. These detector frequencies represent the search frequency range for the DM-off searches.

$$f(t) = f(T_{ref}) + (t - T_{ref})\dot{f}(T_{ref}) \tag{2.2}$$

25

Parameter	Value
T_{coh}	210 hours
T_{ref}	1132729647.5 GPS s
N_{seg}	12
δf (Hz)	1.0×10^{-6}
$\delta \dot{f}_c$ (Hz/s)	1.0×10^{-10}
γ	100
n_{sky}	60

Table 2.3: Search parameters for DM-off searches on simulated signals. The search grid spacings are ten times coarser than the grid spacings used in the corresponding DM-on searches.

$$f_d = f_o \frac{(1 + v_{Earth}/c)}{(1 - v_{Earth}/c)} \tag{2.3}$$

2.2.3 Defining the rejection criterion

The rejection criterion was defined by studying how the fake signals behaved in the DM-on and DM-off searches. For each simulated signal, we examined the loudest DM-on and DM-off statistic values. The loudest candidate from a DM-on search on a fake signal represents the candidates from the continuous wave search. On the other hand, the loudest candidate from the DM-off searches is the new value of comparison and gives a measure of how likely it is that the candidate was caused by a detector artefact.

For both the DM-on and DM-off searches, we used the F-statistic — a standard statistic used in CW searches — to quantify a candidate's significance. The F-statistic ($2F$) values were used to define the rejection criterion. The DM-off $2F$ ($2F_{DMoff}$) values were plotted against the DM-on $2F$ ($2F_{DMon}$) values. Based on the dependence observed, contours were defined in the plane of $2F_{DMoff}$ and $2F_{DMon}$ values. These contours divided the detection statistic plane into "signal-like" and "non-signal-like" regions.

On plotting the recovered $2F$ values for the fake signals (see Fig 2.2), it was observed that the signals could be divided into two populations. The blue dots represent the weaker signals which have $2F_{DMon}$ values less than 26.71. The $2F_{DMoff}$ values showed no dependence on the $2F_{DMon}$ values for these weaker signals. Their $2F_{DMoff}$ values ranged from 10 to 12.25. For these signals, a flat threshold was set at the $2F_{DMoff}$ corresponding to 13. For

signals with $2F_{DMon}$ values greater than 26.71 (shown through orange dots in Fig 2.2), the $2F_{DMoff}$ values showed a dependence on the $2F_{DMon}$ values. For these louder signals, the threshold was defined according to the second line of equation 2.4.

$$2F_{DMoff} = \begin{cases} 13 & \text{if } 2F_{DMon} < 26.71 \\ 10^{-0.24}2F_{DMon}{}^{0.95} & \text{if } 2F_{DMon} \geqslant 26.71 \end{cases} \tag{2.4}$$

The $2F_{DMon}$ value is the independent variable here which grows with the square of the gravitational wave signal amplitude. For each DM-on F-statistic value, the threshold equations define a corresponding $2F_{DMoff}$ value as the loudest $2F_{DMoff}$ value expected for a potential signal candidate. Any candidate with a given $2F_{DMon}$ value can be safely rejected if its corresponding $2F_{DMoff}$ value exceeds this upper limit because it does not behave in a "signal-like" way. Hence, a rejection/acceptance region is established in the plane of $2F_{DMoff}$ and $2F_{DMon}$ values where the acceptance region signifies signal-like behaviour.

Fig 2.2 shows the acceptance region for signals established in the $2F_{DMon}$ and $2F_{DMoff}$ plane based on the searches conducted on the thousand simulated signals. The weaker signals (shown in blue) have a flat threshold. The red dotted line at the $2F_{DMon}$ 26.71 represents the cut-off $2F_{DMon}$ value for these signals. The orange dots represent the louder signals. The dotted black lines represent the contours of the rejection region given by equations 2.4.

2.2.4   Application of the semicoherent DM-off veto

We applied the rejection criterion to the surviving candidates of Stage 0 of the O1 all-sky search. For each candidate, the Doppler plus spindown range of detector frequencies was computed using equations 2.2 and 2.3 for the given candidate frequency, spindown, and sky position. In order to reject the candidate, the highest of the semicoherent $2F_{DMoff}$ values for the Doppler and spindown detector frequencies was compared to the threshold of equations 2.4. If this highest value was greater than the threshold, the candidate was rejected. Fig 2.7 shows the results of applying the veto to Stage 0 candidates.

With the veto presented here, we rejected 27518 of the original 36248 Stage 0 candidates. In the original search, candidates that survived Stage 0 were then passed through the three successive follow-up stages. We followed the
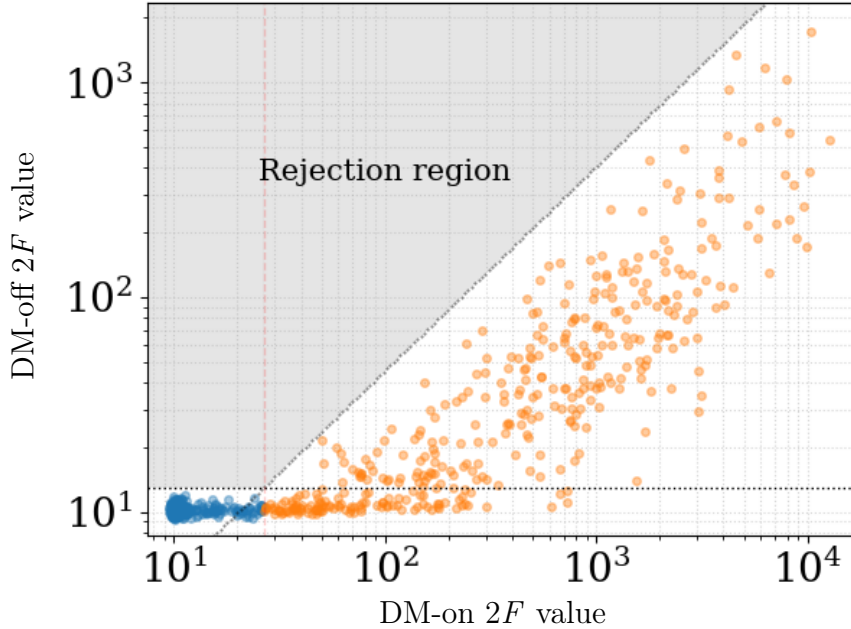
Figure 2.2: Detection statistic values for DM-on and DM-off searches on simulated signals. The grey area represents the rejection region. This is chosen so that no signal would be rejected.

remaining 8730 candidates that survived the DM-off veto to see how many would have dropped out at each follow-up stage. The results are summarised in Table 2.5. Fig 2.8 also shows the candidates that would have survived at each follow-up stage of the hierarchical search after the application of the DM-off veto to the Stage 0 candidates. The question is then whether a semicoherent DM-off veto at stage zero, a significantly reduced number of follow-ups at stages 1, 2, 3, and the fully coherent DM-off veto on the 20 surviving candidates after stage 3, would have resulted in significant savings in computational cost. I address this briefly in the next section and then more in-depth in section 2.3.2.2.

2.2.5 Calculating the computational time for searches

`lalapps_HierarchSearchGCT` [10] provides information to determine the runtime of the search. Using this information, the runtime of the DM-on
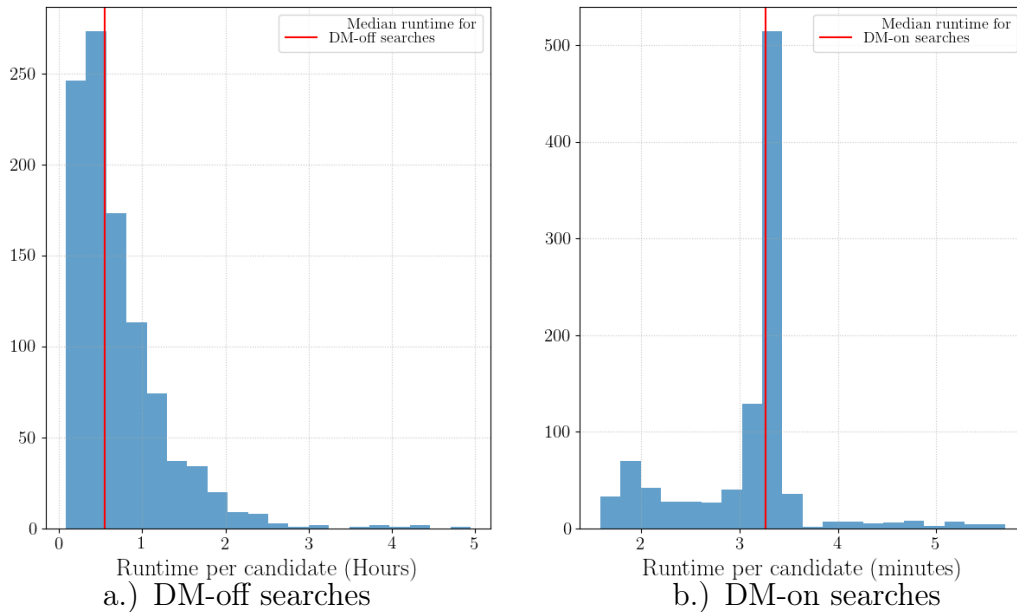
Figure 2.3: Distribution of total runtime for DM-off searches and DM-on searches around individual candidates. The red lines marks the median runtime for the searches. The average runtime for DM-off searches was 0.7 hours per candidate. The average runtime for DM-on searches was 3.08 minutes.

and DM-off searches was computed for each simulated signal. Figure 2.3 shows the distributions of runtimes for the two searches.

The average runtime for the follow-up stages in the O1 all-sky search was 4 hours per candidate, while the average runtime of a DM-off search around a candidate was 0.7 hours. Based on this, we can already say that the DM-off search is a cheaper way to eliminate candidates due to noise before they are followed up in the later stages of the hierarchical search. I will go back to this topic in section 2.3.2.2 where I discuss the computational savings in detail.

2.3 Results and Discussion

In this section, the thresholds defined in section 2.2.3, and reasons for the observed behaviour of weak signals (Fig 2.2) is discussed. Then, the results of applying the semicoherent DM-off veto to Stage 0 candidates of the O1 all-sky search are presented and discussed.

2.3.1   Understanding the thresholds

Fig 2.2 shows that the $2F_{DMoff}$ value was independent of the $2F_{DMon}$ value for the weaker signals (shown in blue). This suggests that, for each signal, the loudest recovered candidate was not actually associated with the signal but instead came from a random noise fluctuations. If this is true, then the observed $2F_{DMoff}$ values must be consistent with the expected distribution of the loudest $2F$ values in Gaussian noise over a certain number of independent trials.

$2F$ values in Gaussian noise follow a χ^2 distribution with 4 degrees of freedom for coherent searches [9]. The probability $p(N; 2F^*)$, of a given F-statistic value, $2F^*$, to be the highest of N trials in Gaussian noise is given by equation 2.5.

$$p(N; 2F^*) = Np(\chi_4^2; 2F^*)[cdf(\chi_4^2; 2F^*)]^{(N-1)} \tag{2.5}$$

where N is the total number of independent trials that generate F-statistic values, $p(\chi_4^2; 2F^*)$ is the χ^2 probability of getting an F-statistic value of $2F^*$, and $[cdf(\chi_4^2; 2F^*)]^{(N-1)}$ is the probability that the other (N-1) $2F$ values are smaller than $2F^*$.

In the case of semicoherent searches, the highest detection statistic becomes $N_{seg} \times 2F^*$. In this case, $N_{seg} \times 2F$ values follow a χ^2 squared distribution with $4 \times N_{seg}$ degrees of freedom.

If we repeatedly perform a search with N trials, and pick the maximum $2F$ each time, we expect to obtain a range of $2F^*$. This means that $2F^*$ can take any value which has a probability greater than 0. Here, $2F^*$ is now the random variable whose probability distribution function is given by equation 2.5.

In a gridded search with N number of total independent signal templates, the probability of recovering a particular $2F$ value as the highest must be equal to equation 2.5 when $2F^* = 2F$. However, the templates are spaced very close together to increase the chance of finding a signal. So, they are statistically correlated and the distribution function with N being the number of searched templates overestimates the probability. Due to this, the total number of templates N must be replaced with the effective number of templates N_{eff}, which are statistically independent, to get the probability distribution function that correctly describes the maxima from the search.
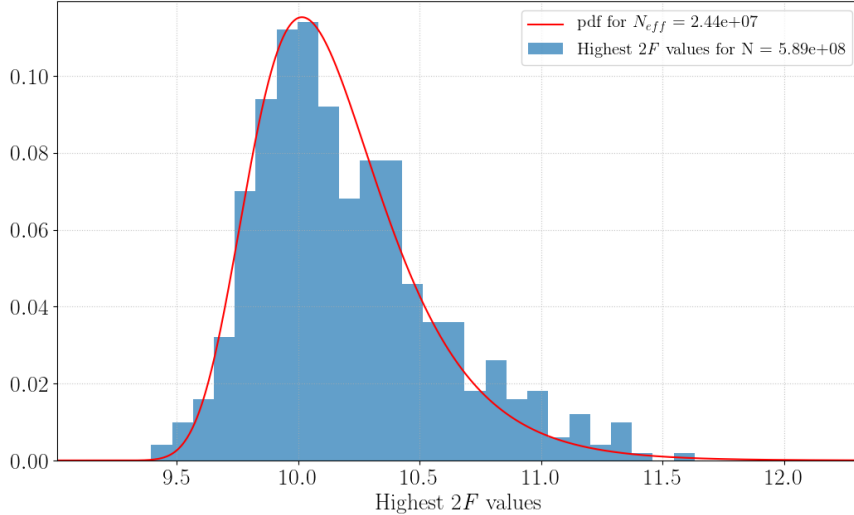
Figure 2.4: Fitting the probability distribution function of the highest 2F values to the distribution obtained from multiple trials for a 5 mHz frequency band

The probability distribution function given by equation 2.5 was used to estimate the independent number of templates for DM-off searches by fitting N to the observed distribution of maxima. DM-off searches were performed on Gaussian noise over varying search frequency ranges. For each frequency band, the same DM-off search was repeated multiple times and the highest $2F$ value was retrieved from each trial to get a set of the highest $2F$ values. To the histogram of these values, the probability distribution function (Eq 2.5) was fitted with N as the free parameter to get the N_{eff}. After obtaining N_{eff} for five different values of N (Table 2.4), I found that the two were related by the formula given in equation 2.6.

The DM-off searches that we performed on the simulated signals had different search frequency ranges, and the number of search templates in each search was therefore, not constant. So, the expected value of the maximum $2F$ value is not constant, and is in fact drawn from different distributions, such as the ones shown in Fig 2.5.

$$N_{eff} = 10^{0.92} \, N^{-0.65} \tag{2.6}$$

Using equations 2.6 and 2.5, we can determine the expected loudest in noise for all of our DM-off searches on weak signals and compare this to

31

Frequency Band (mHz)	N_{total}	$N_{\text{effective}}$	Ratio
0.5	5.89×10^7	2.83×10^6	0.048
5.0	5.89×10^8	2.44×10^7	0.041
50.0	5.89×10^9	2.12×10^8	0.036
100.0	1.18×10^{10}	3.67×10^8	0.031
500.0	5.89×10^{10}	1.56×10^9	0.026

Table 2.4: Results of fitting the probability distribution function for highest $2F$ values to their observed distribution for different search frequency bands.

the observed loudest candidates. Fig 2.5 shows the observed distribution of $2F_{DMoff}$ values for weaker signals along with the expected distributions for the smallest and largest DM-off search frequency ranges for these signals. The distribution function shifts towards the right for a higher number of templates. Hence, the distribution functions for intermediate search frequency ranges must lie between these two probability distribution functions. The figure shows that the recovered $2F_{DMoff}$ values fall within the expected range of highest $2F_{DMoff}$ values for Gaussian noise. This is consistent with the hypothesis that the loudest $2F_{DMoff}$ values for weak signals are due to random noise fluctuations.

Comparing the flat threshold at the $2F_{DMoff}$ value of 13 to the 5σ F-statistic values for weak signals (Fig 2.6), it is observed that 13 is greater than the 5σ values for these signals. Hence, by setting the flat threshold at 13, the probability of falsely rejecting a potential signal candidate with $2F_{DMon}$ value less than 26.71 is extremely low.

2.3.2   Application of the semicoherent DM-off veto

2.3.2.1   Rejection of candidates

The semicoherent DM-off veto was applied to the 36248 surviving candidates of Stage 0 of the O1 all-sky search. Table 2.5 summarises the results of the application of the semicoherent DM-off veto on the Stage 0 candidates. 27518 candidates were rejected based on the defined rejection criterion. This would have resulted in less than 25 percent of the candidates to be followed up in the next stage of the hierarchical search. On comparing the surviving candidates of the veto to the surviving candidates of each of the follow up stages of the O1 all-sky search, it is observed that only 20 candidates would have survived at the end of the third follow up stage. This is a significant
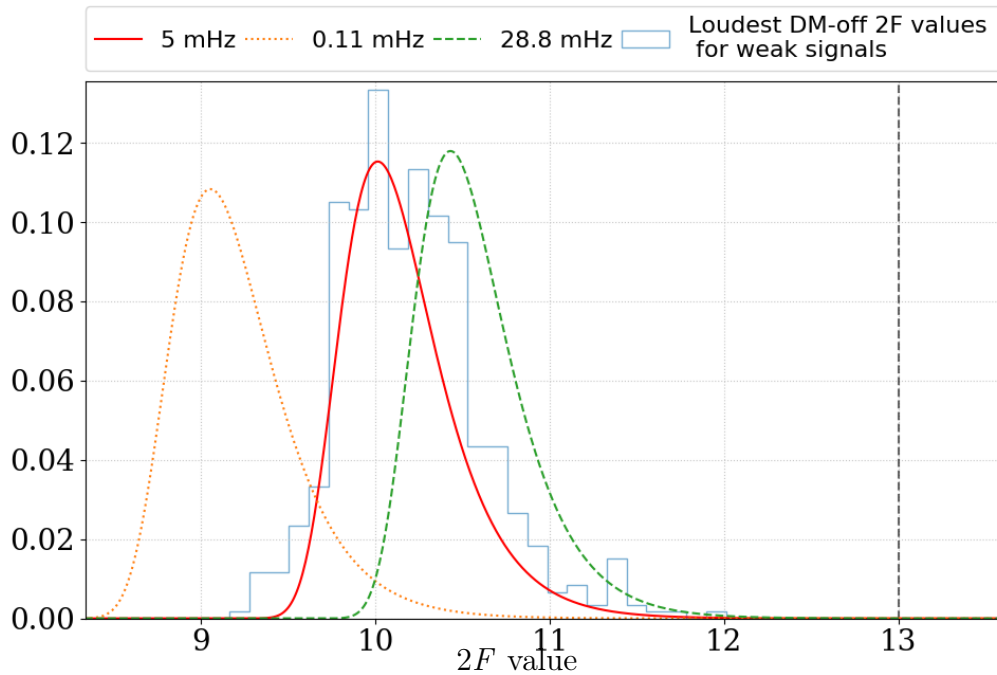
Figure 2.5: Distribution of loudest DM-off values recovered from DM-off searches on weak fake signals plotted along with the probability distribution functions of highest DM-off $2F$ values for the smallest and largest DM-off search frequency ranges. The probability distribution functions of highest DM-off $2F$ values for a 5 mHz band is plotted for reference.

improvement in rejecting spurious candidates given the fact that over 6000 candidates survived the third follow up stage in the O1 all-sky search. The 20 candidates that would survive the third follow-up stage also include the four candidates that survived the coherent DM-off veto, shown in Fig 2.8. This shows that the semicoherent DM-off veto is consistent with the coherent DM-off veto.

The number of surviving candidates reported in [1] do not include candidates due to a known fake signal that was injected in the hardware at 52.8 Hz in the detector. The numbers reported in Table 2.5 for the O1 all-sky search include these candidates and are therefore greater. The candidates due to to this hardware injection survive the semicoherent DM-off veto (see Fig 2.8). This is proof that the semicoherent DM-off veto predicts signal behaviour well as it did not falsely reject the candidates due to the hardware injection.

33
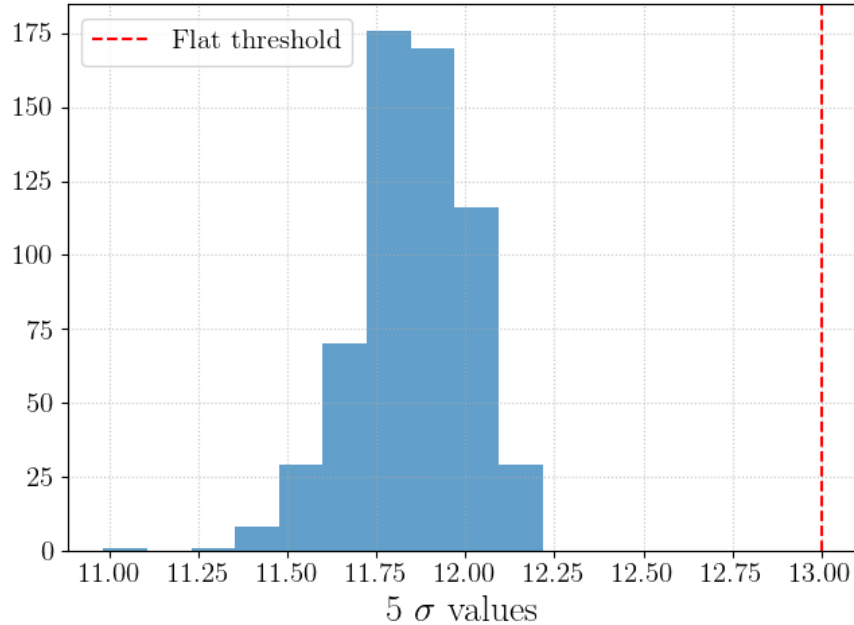
Figure 2.6: Distribution of 5σ values for the DM-off searches for weak simulated signals.

2.3.2.2 Computational gain

In the original O1 all-sky search, the average computational time invested for each follow up stage was 4 hours per candidate. All of these searches were performed on a supercomputing cluster. If instead the veto had been applied before the first follow up stage, only 8730 candidates would have to have been followed up. This would have resulted in a gain of over one hundred thousand hours in computational time and power at the first follow up stage itself. However, this does not account for the computational investment required to set up and apply the DM-off veto.

The runtime for a DM-off search was 0.7 hours per candidate on average, while that for a DM-on search was 3 minutes per simulated signal on average (see Fig 2.3). The number of fake signals used in the study was a thousand. Hence, the computing time for the DM-off searches that were used to define the rejection region was approximately 700 hours, while that for DM-on searches was 50 hours. So, approximately 750 hours of computational time were invested for defining the rejection region using a thousand simulated signals. The semicoherent DM-off search that was performed over the whole

34
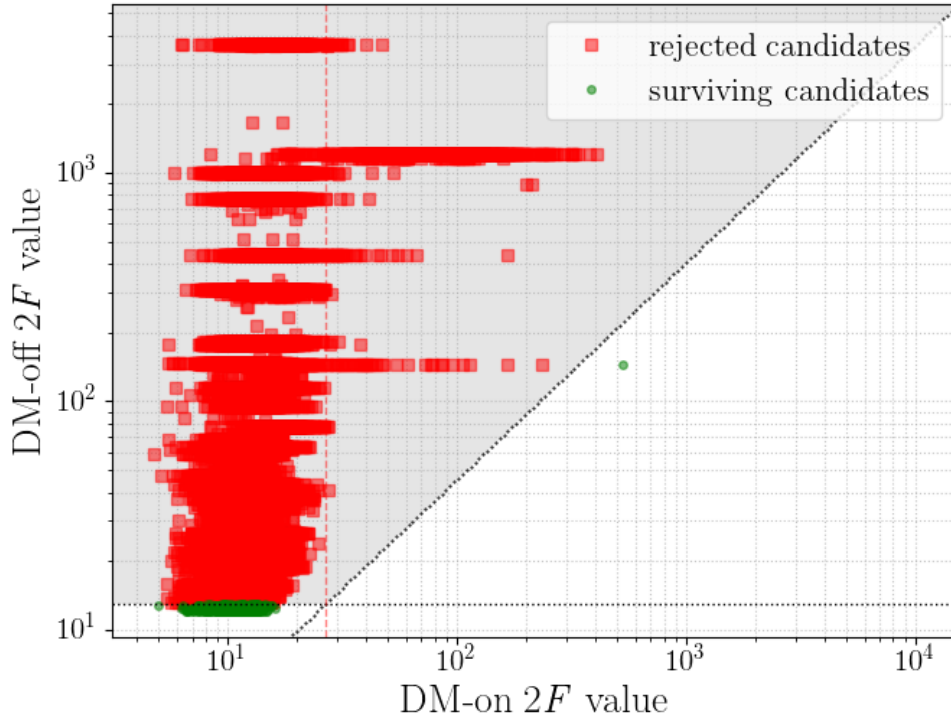
Figure 2.7: Application of the semicoherent DM-off veto on the surviving candidates of Stage 0 of the O1 all-sky search. The red squares represent the candidates due to their "non-signal-like" behaviour. The green circles represent candidates that survive because of their "signal-like" behaviour.

search frequency range took approximately 4000 hours. In total, 4750 hours were invested in setting up and applying the semicoherent DM-off veto. A computational investment of 4750 hours to set up the DM-off veto would have saved over two hundred thousand hours of computational time on the supercomputing cluster.

To summarise, the semicoherent DM-off veto applied to the candidates of Stage 0 of the O1 all-sky search rejected over 75% of the candidates by identifying them as detector artefacts. This would have resulted in 20 surviving candidates at the end of the last follow-up stage. This is a huge improvement considering over 6000 candidates had survived at the end of the last follow-up
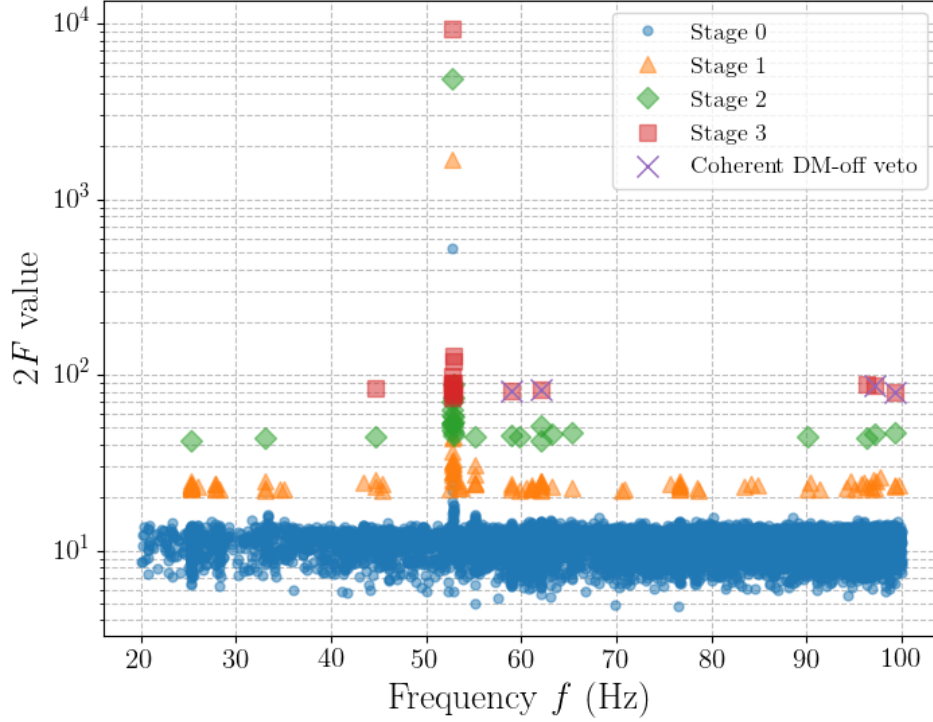
Figure 2.8: Surviving candidates from the follow up stages after the application of the semicoherent DM-off veto. The candidates that survived the coherent DM-off veto are also shown.

stage in the O1 all-sky search. The application of the semicoherent DM-off veto at this stage would have allowed us to save over two hundred thousand hours of computational time that could have been allocated to other steps or to other continuous wave searches. Based on these studies for the O1 all-sky search, we can say that the semicoherent DM-off veto is extremely effective in rejecting noise candidates and in saving computational time.

	$N_{\text{survivors}}$ O1 all-sky search	$N_{\text{survivors}}$ semicoherent DM-off veto
Stage 0	36248	8730
Stage 1	14695	105
Stage 2	8710	28
Stage 3	6551	20

Table 2.5: Results of applying the semicoherent DM-off veto to the surviving candidates of Stage 0. The numbers reported for the original O1 all-sky search include candidates from a known hardware injection. These candidates were removed from the numbers reported in [1]. Therefore, the numbers in the table are greater than the numbers reported in [1]

Chapter 3

Optimisation of the semicoherent DM-off veto - how much cheaper can we make the veto?

The DM-off search setup used for defining the rejection criterion for the O1 all-sky search was based on the search setup used for the original O1 all-sky search. The grid spacings used in DM-off searches were approximately ten times coarser than those used in Stage 0 of the O1 all-sky search [1]. At this coarseness, the computational time of the searches was 0.7 hours per candidate on average and the veto effectively rejected over 75% of Stage 0 candidates.

In this chapter, I explore how much coarser the DM-off searches could be made in order to save more computational power. However, when the searches are made coarser, they become less sensitive, hence, less effective. This results in greater mismatch values and the probability of not catching the disturbance increases as the grid spacings become wider. The challenge then lies in choosing a search setup that reduces the computational investment but at the same time also rejects a considerable number of spurious candidates. Keeping these two ideas in mind, a metric for determining the relative improvement in the performance (here, called the gain) was defined and computed for different setups.

3.1 Methods

The goal of using the DM-off veto is to identify detector artefacts in the data. We observe that detector artefacts are broader in frequency than signals.

Therefore, the DM-off searches used for the characterisation of the DM-off veto for the O1 all-sky search had approximately ten times coarser frequency and spindown grid spacings and we could reject over 75% of those candidates using the veto (Chapter 2). Any alternative DM-off search setup must also be able to identify these artefacts and reject a significant number of spurious candidates. For this reason, we look at what happens to known detector artefacts when new search setups are used.

The effect of making search grids coarser is quantified by studying the change in the largest F-statistic $(2F)$ values and computing the runtime for searches. For each artefact, the percentage change in the largest $2F$ value and the computational runtime of the search were computed. The DM-off search setup used for the characterisation of the semicoherent veto for the O1 low frequency all-sky search (Chapter 2) was used as the reference setup for measuring the change. The effectiveness of the veto with these search setups was also tested by estimating the number of rejected candidates for each setup.

3.1.1   DM-off search setups

Detector artefacts were chosen from the semicoherent DM-off search performed on the whole frequency search range of the O1 all-sky search. Fig 3.1 shows the three chosen artefacts. Artefact 1 is of low significance, while Artefact 3 is very loud. Artefact 2 is of intermediate significance. This choice was made to study how the search setups would affect a range of artefact types and strengths. We performed DM-off searches with a particular setup on each artefact multiple times to incorporate a range of possible mismatch values.

DM-off searches have multiple search parameters. In this study, three parameters were varied to study the effect on the F-statistic values and the runtime. These parameters were the frequency grid spacing δf, the refinement factor in spindown γ, and the number of sky positions n_{sky}. We could also make the spindown grid spacing $\delta\dot{f}$ larger, but the number of spindown grid points in the previous DM-off searches was not very high (approximately 20 per search). Hence, we decided to not make $\delta\dot{f}$ coarser, lest we increase the loss of search sensitivity. Table 3.1 summarises the different values for the three parameters that were used in different combinations. The new setups had coarser frequency grid spacings, smaller values of the refinement factor, and fewer sky grid points.
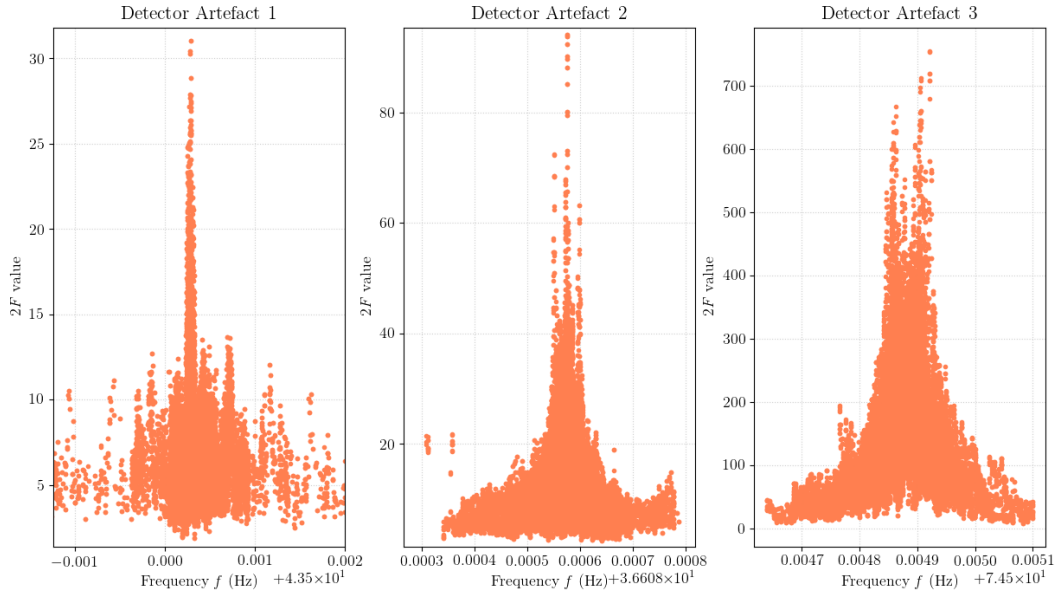
Figure 3.1: The chosen detector artefacts. The plots show the loudest candidates from a DM-off search with 60 sky-grid points, δf equal to 1.0×10^{-6} Hz, and γ equal to 100 (the original DM-off search setup) in the respective frequency ranges. The DM-off search makes these artefacts visible. $2F$ values (y-axis) are plotted against detector frequencies (x-axis).

The DM-off search setup used for characterising the O1 all-sky semicoherent DM-off veto had 60 sky grid points, δf equal to 1.0×10^{-6} Hz, and γ equal to 100. The change in detection statistics, computational time, and gain were computed for the new search setups with respect to this original search setup. We do this to estimate how much better the new setups are with respect to the setup that we already used for the veto which gave great results. We want to find out if there's any other setup that could be better than the original setup.

Parameter	Values				
δf	1.0×10^{-6}	4.0×10^{-6}	8.0×10^{-6}	9.5×10^{-6}	1.5×10^{-5}
γ	100	1			
n_{sky}	60	32			

Table 3.1: Values of parameters varied in DM-off search setups. Combinations of these values were used in the various search setups.

3.1.2   F-statistic values and Computational time

For each setup, DM-off searches were performed multiple times on each arte-fact. For example, the DM-off search with 60 sky-grid points, δf equal to 1.0×10^{-6} Hz, and γ equal to 1 was performed a hundred times on arte-fact 1 and the exact start frequency was randomized within a small range of possible values for each trial. The candidate with the highest 2F value was then recovered. The multiple trials ensure that a range of mismatches is sampled.

Fig 3.2 shows a single example: the distribution of the highest $2F$ values and their average for Artefact 2 plotted for the search setup with 60 sky-grid points, δf equal to 1.0×10^{-6} Hz, and γ equal to 1. Using the average of the highest $2F$ values for each setup, the percentage change was calculated with respect to the original search setup. Fig 3.3 shows the average highest $2F$ value for each setup and for each artefact. Fig 3.4 shows the percentage change in these values with respect to the highest $2F$ value recovered for the original setup.
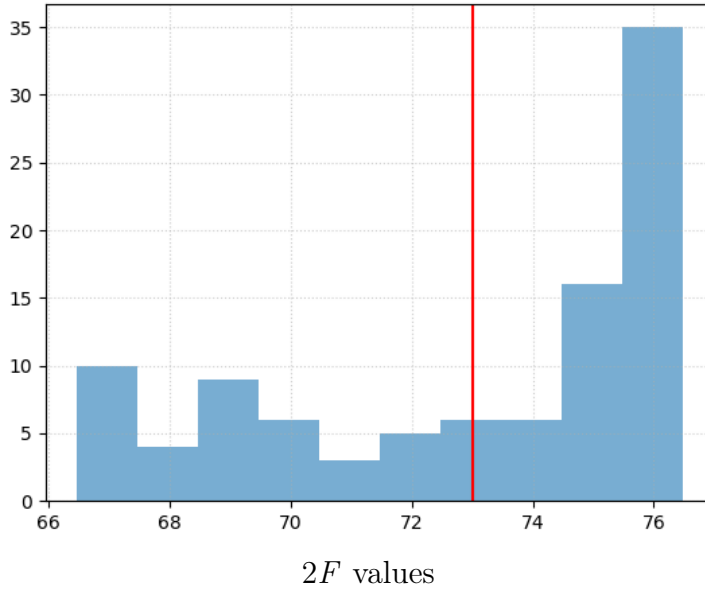
$2F$ values

Figure 3.2: Distribution of highest $2F$ values for a setup with 60 sky-grid points, δf equal to 1.0×10^{-6} Hz, and γ equal to 1. The red line marks the average $2F$ value for this setup.
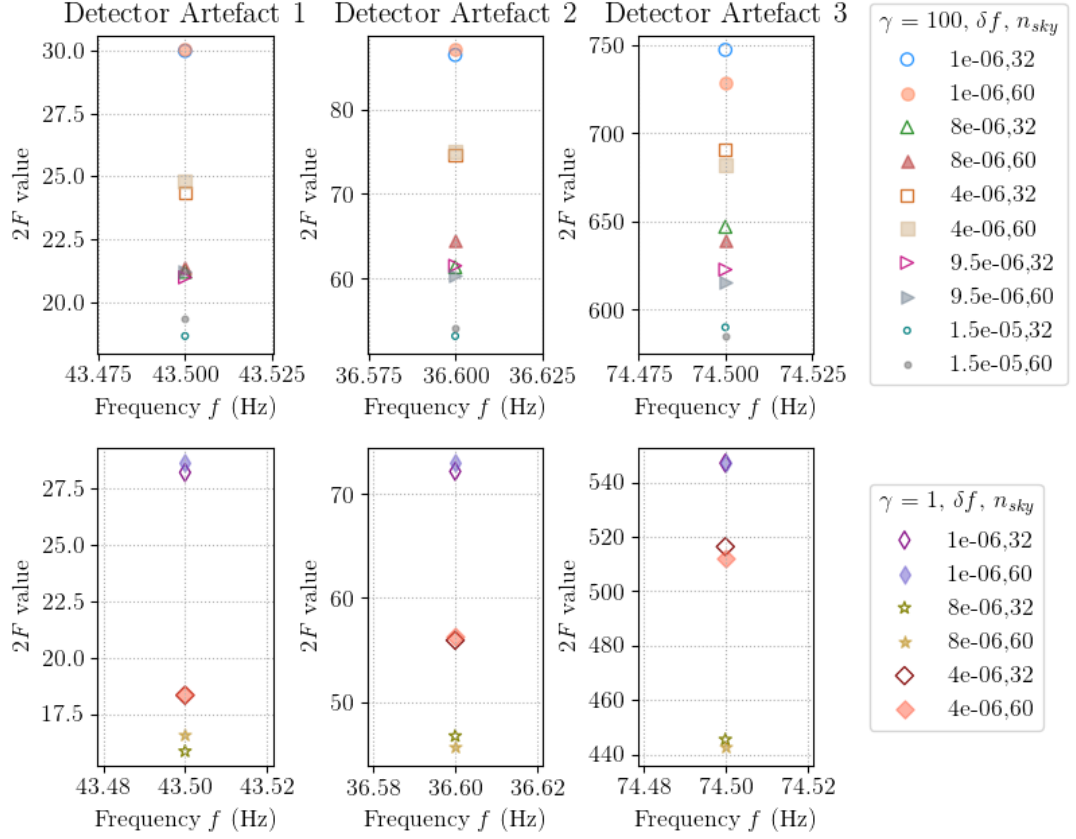
Figure 3.3: The average of recovered highest $2F$ values for different DM-off search setups.

Using the time information generated by `lalapps_HierarchSearchGCT`, the total runtime for the searches was computed. For each setup, the average runtime across the trials was chosen. In order to derive a comparable quantity between the artefacts and the different setups, the runtime per mHz was computed for each setup. Fig 3.5 shows the runtime per mHz for different search setups with refinement factor γ equal 100.

3.1.3 Application of veto with different setups

In order to test the effectiveness of the new search setups, each DM-off veto was applied to the Stage 0 candidates with $2F$ values that would be obtained with DM-off searches with these new setups. For each setup, we took the
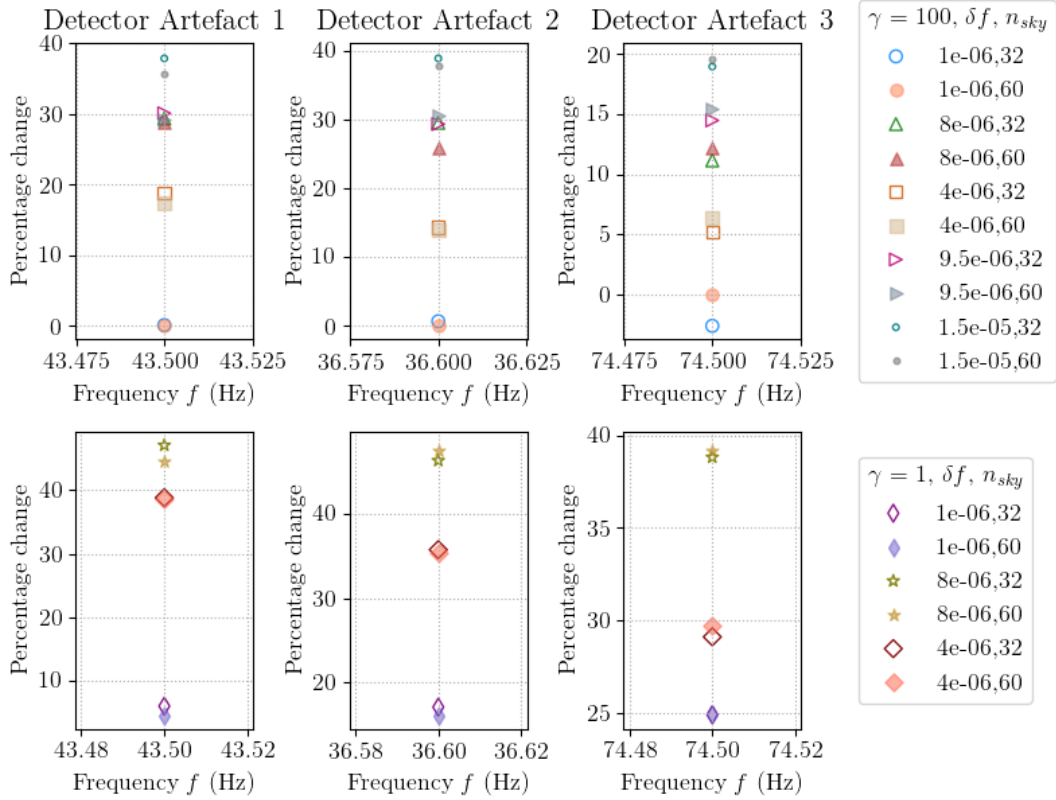
Figure 3.4: The percentage change in recovered highest $2F$ value for different DM-off search setups with respect to the highest recovered $2F$ value for the DM-off search setup of the semicoherent DM-off veto for the O1 all-sky search.

maximum percentage changes in the DM-off $2F$ values from the studies on the three artefacts as representative of how the $2F$ value would change for artefacts in general. Using these percentage changes, we reduced the semicoherent DM-off $2F$ values obtained from the DM-off search with the original setup. These reduced F-statistic values were used as representative DM-off $2F$ values for the new search setups.

The modified DM-off $2F$ values were then compared to the DM-on $2F$ values of the Stage 0 candidates. As a preliminary test, we used the thresholds defined previously using the original DM-off setup (see equation 2.4) for the veto. These thresholds were applied to the modified DM-off $2F$ values to accept or reject candidates in order to get an insight into the number of candidates that could be rejected using the new search setup. We use the maximum percentage changes in $2F$ values and use thresholds defined for
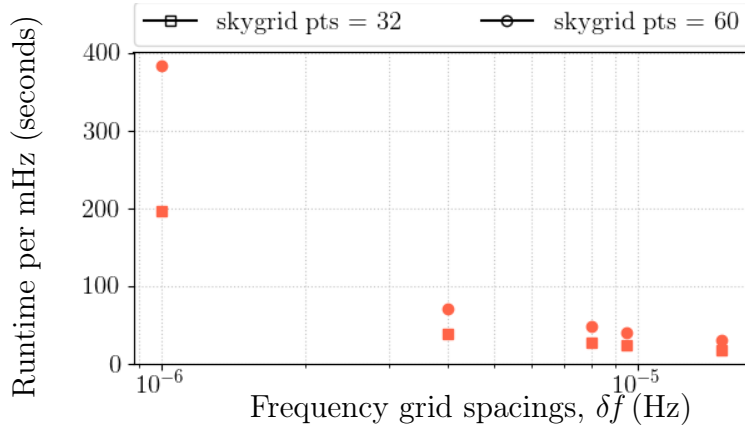
Figure 3.5: The computational runtime per mHz for different frequency grid spacings δf, and sky-grid points n_{sky}. The refinement factor γ for all of these setups is 100.

the finest and most sensitive setup to apply the DM-off veto. However, the new setups are coarser, less sensitive and include fewer templates. Hence, the thresholds for these setups would be lower. Therefore, this is a conservative approximation of the number of rejected candidates.

The underlying assumption for using the thresholds defined for the original DM-off veto (equation 2.4) was that the change in coarseness must not affect the thresholds drastically. A grid with four times coarser δf must have 4 times fewer templates. This change in the number of templates would shift the distribution function of highest $2F$ values (equation 2.5) to the left by a small amount. Hence, the flat threshold will not reduce significantly. Since most of the Stage 0 candidates had DM-on $2F$ values less than 26.71 (see Fig 2.7), the flat threshold is of more significance in this study.

Further investigations were conducted by estimating the location of the flat threshold from the probability distribution function of loudest $2F$ values (equation 2.5) for each search setup. The log-linear threshold (for louder signals) was also shifted down by the percentage change in the flat thresholds with respect to the flat threshold at DM-off $2F$ value of 13. The results obtained from these investigations were consistent with the results of the preliminary tests and are shown in Fig 3.6. With increasing coarseness in δf, the number of candidates rejected by the veto decreases. This is expected as the significance of candidates reduces with larger δf and the chances of missing the exact candidate parameters is greater for larger δf.

44
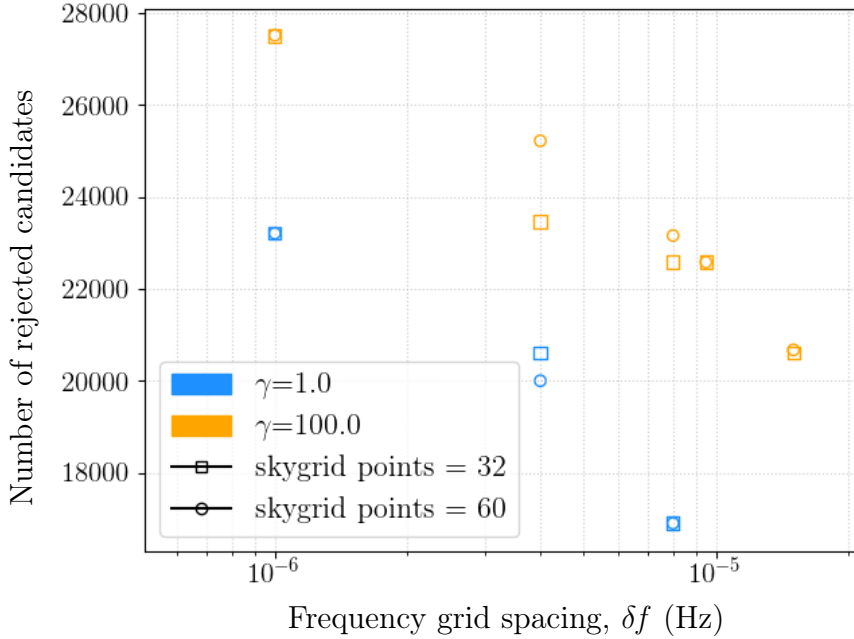
Figure 3.6: The number of Stage 0 candidates rejected by the semicoherent DM-off veto with different search setups using the thresholds given by equations 2.4. This is an estimate of how effectively each search setup would rule out spurious candidates.

3.1.4 Estimation of Gain for different setups

In order to quantify the expected gain associated with a setup, it is important to note not only the reduction in computational time but also account for the change in number of surviving candidates with respect to the original search setup. Although the computational time reduces as the search setup gets coarser, the number of surviving candidates is expected to increase because the $2F$ values drop as the search setup gets coarser. This contributes to the cost of using a particular setup with respect to the original setup. Hence, the total gain associated with the choice of a particular setup is quantified according to equation 3.1. The assumption while defining this metric is that we would have to apply the veto in 2 stages - first with the coarser grid i, then with the original (finest) grid f - in order to reject as many candidates as we did with the original DM-off search setup.

$$Gain = N_{vetoed,\ f} \times c_f - N_{vetoed,\ i} \times c_i - (N_{survive,\ i} - N_{survive,\ f}) \times c_f \quad (3.1)$$

45

Here, $N_{vetoed, \ f}$ is the number of candidates rejected by the original (finest), f grid, and c_f is the cost associated with rejecting a candidate using this finest grid. Similarly, $N_{vetoed, \ i}$ is the number of candidates rejected by the new grid, i, and c_i is the cost associated with rejecting a candidate using this new grid. $N_{survive, \ i}$, and $N_{survive, \ f}$ are the number of candidates that survive the veto when the new, i, and the finest, f grids are used respectively.

In equation 3.1, $(N_{vetoed, \ f} \times c_f - N_{vetoed, \ i} \times c_i)$ corresponds to the gain associated with choosing setup i with respect to the finest (original) f setup. The second difference corresponds to the cost of using the new setup. It quantifies the added cost due to the change in the number of candidates that survive the veto when the new setup is used.

In equation 3.1, the cost per candidate is essentially the runtime per candidate. However, it is non-trivial to estimate the runtime per candidate from the total runtime for a search, specifically because the DM-off search for the O1 all-sky search was performed over the whole search frequency range and not for each candidate. Therefore, the runtime per mHz is used as the cost per candidate under the assumption that the density of candidates is uniform across the search frequency range. Fig 3.7 shows the distribution of number of surviving candidates of stage 0 of the O1 all-sky search per mHz across the search frequency range. Over 98.75% of mHz bands have 0 to 9 candidates which suggests that the distribution of candidates is more or less uniform.

The gain, as defined by equation 3.1, was calculated for each setup using the runtime per mHz and the number of surviving candidates. These quantities are plotted in Fig 3.8.

3.2 Results and Discussion

3.2.1 Optimisation of the setup

The DM-off search setups were varied using three parameters — frequency grid spacing δf, sky grid n_{sky}, and the refinement factor for spindown γ. For each setup, the gain was computed with respect to the original setup used for characterising the veto for the O1 all-sky search described in the previous chapter. Based on the results of the studies described in the previous chapter, in this section, the questions that I look to answer are — which parameter affects the veto optimisation the most and at what search setup coarseness does the veto break down? The results after varying the three parameters are presented and discussed.
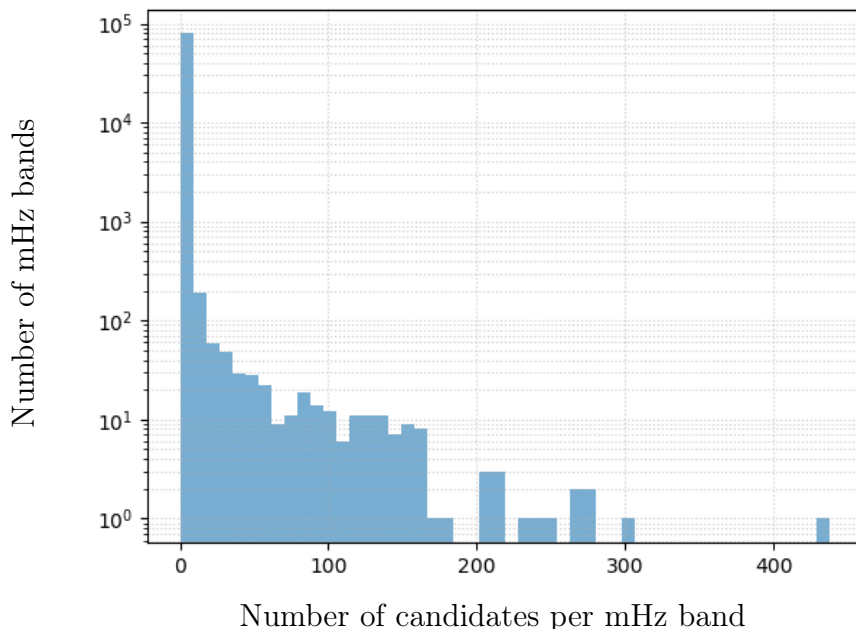
Figure 3.7: The distribution of number density of Stage 0 candidates per mHz band in the O1 all-sky search frequency range of 20 -100 Hz.

3.2.1.1 Frequency grid spacing

In Fig 3.8, it is observed that the gain is affected the most by changing the frequency grid spacing, δf. On increasing δf, the gain, as defined by equation 3.1, increases up to a particular value of δf beyond which it reduces again. This indicates that the setup is optimal for a range of δf only. This trend is observed because the number of candidates that the search can reject with greater δf reduces drastically beyond a particular value of δf. Even though the runtime reduces drastically (see Fig 3.5), the increase in the number of surviving candidates raises the cost of the next stages, and hence, the overall cost, reducing the gain (equation 3.1). An optimal setup balances this cost with a reduction in computational runtime. Hence, the gain with respect to the original setup is maximised over a range of δf which represents the optimised range of δf. Beyond this range, the gain starts to fall again which can be seen in Fig 3.8.
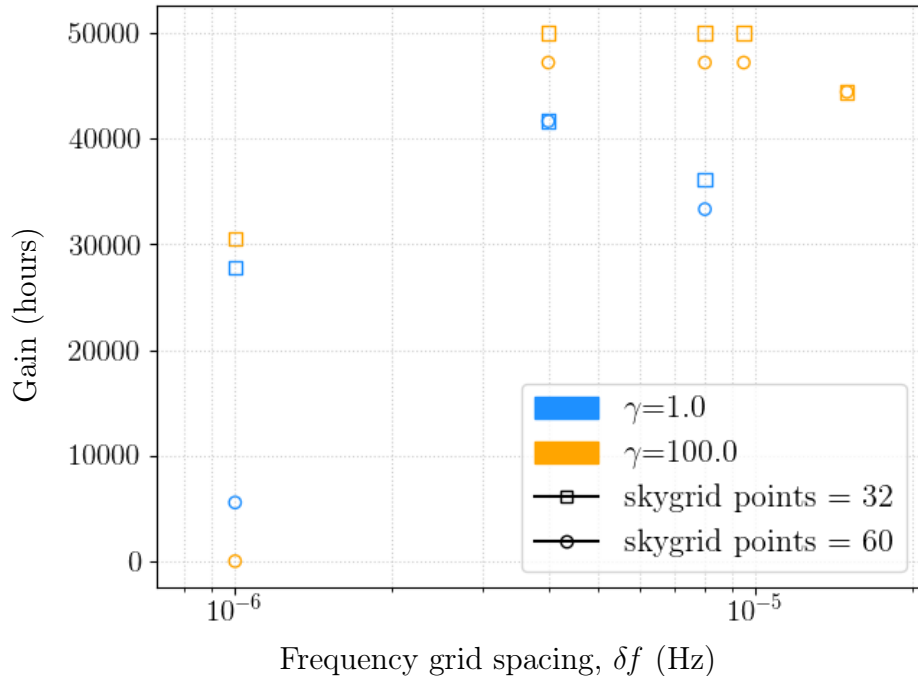
Figure 3.8: Computed gain for different search setups with respect to the DM-off search setup used for the characterisation of the semicoherent DM-off veto for the O1 all-sky search.

3.2.1.2 Sky grid

For DM-off searches, the sky positions are distributed uniformly across the whole sky to give the veto more freedom in finding the loudest $2F$ value. However, with Doppler demodulation turned off, sky localization is possible only through amplitude modulation which is set by the antenna pattern functions. So, we can determine how the sky positions must be distributed to optimally cover the whole sky by studying the antenna pattern functions of the detectors.

In the optimisation studies, two sky-grids were used - one had 32 sky points while the other had 60 sky positions. In general, one naively expects that the grid with a greater number of points must provide detection statistics that are higher by reducing the chances of mismatches. However, it is observed in Fig 3.4 that this is not always the case. For most setups with 32 sky-grid points, the percentage change in loudest F-statistic value is nearly equal or equal to the corresponding setups with 60 sky-grid points. This observation indicates

that reducing the number of sky points by half only negligibly affects the sensitivity of the DM-off searches.

Fig 3.6 shows that the sky-grid with 32 points is almost as effective as the sky-grid with 60 points with regards to the number of candidates rejected. At the same time, with the other two parameters kept constant, the runtime is halved for a sky-grid with 32 points, which can be seen in Fig 3.5. This is expected considering that the runtime is proportionate to the number of sky-grid points if the rest of the parameters are held constant.

In Fig 3.8, we observe that the gain is higher for search setups with 32 sky-grid points than those with 60 sky-grid points. The runtime of searches reduces appreciably when a 32 points sky-grid is used, while it rejects almost as many candidates as a 60 points sky-grid. The reduction in the runtimes increases the gain for search setups with 32 sky-grid points over search setups with 60 sky-grid points.

3.2.1.3 Refinement factor

It was observed that the gain was higher for a higher value of γ for a particular value of δf, and sky-grid points. For example, for setups with δf equal to 1.0×10^{-6} Hz, and 32 sky-grid points, we see that the setup with γ equal to 100 has higher gain than the setup with γ equal to 1. The refinement factor is used during the recalculation of the detection statistic for the most significant candidates returned by the search algorithm. The algorithm searches a small parameter space using a finer grid around each high-significance candidate. The fineness of this new grid is quantified by the refinement factor. For example, a refinement factor of 100 implies that the fine grid is 100 times finer in spindown $\dot{f}$ than the initial search grid. With this finer search, we obtain 2F values with smaller mismatches for these high-significance candidates. If the value of the refinement factor is high, the finer grid has a greater chance of coinciding with the actual parameters of the artefact. Hence, the chances of catching the artefact with the search setup with a higher refinement factor are greater. This must increase the gain as the setup can identify and reject more candidates due to detector artefacts. Fig 3.8 shows that for a particular δf and sky-grid, the gain increases as the refinement factor increases.

The recalculation step in the search algorithms takes less time than the computation steps. Therefore, the runtime of the search is not affected greatly by changing the refinement factor. Hence, the only major effect of changing the refinement factor is on the significance of the candidate.

49

3.2.1.4 Choice of optimised setup

The choice of refinement factor γ, and sky-grid becomes easy when considering the previous analysis. The setups with γ equal to 100 and 32 sky-grid points are considered as candidates for the optimised search setup. In the case of δf for this choice of γ and sky-grid, the gain is maximised over a wider range of δf. We have 3 choices of δf out of the search setups explored in this study. We choose the search setup with δf equal to 4×10^{-6} Hz. At this coarseness, we already save close to 50000 hours of computational time and the runtime per mHz is approximately 50 seconds. The maximum percentage change in 2F values is 20% for the search setup with γ equal to 100, 32 sky-grid points and δf equal to 4×10^{-6} Hz (see Fig 3.4). At greater values of δf, the runtime is smaller, the gain remains the same but the percentage change in $2F$ values is higher than 20%. In such cases, the probability of losing detector artefacts increases. Therefore, we choose δf equal to 4×10^{-6} Hz as the optimal δf.

With this setup — δf equal to 4×10^{-6} Hz, γ equal to 100 and 32 sky-grid points — we would have rejected over 20000 candidates out of the 36248 candidates that survived the first stage of the O1 all-sky search, while saving approximately 50000 hours of computational time even after the candidates had been followed up using the finer semicoherent DM-off veto after this DM-off veto.

Chapter 4

The semicoherent DM-off veto for a directed search

In chapter 2, I described the semicoherent DM-off veto which was applied to the O1 all-sky search as a proof of concept. The veto was originally developed according to the search parameters of the O1 all-sky search. However, the semicoherent DM-off veto is flexible — the veto can be applied to any continuous wave search simply by adjusting the parameters of the DM-off veto, and defining the rejection region accordingly. In this chapter, I discuss the challenges and possible solutions for setting up the semicoherent DM-off veto for a set of directed searches around three astrophysical sources in the data from the first observing run of Advanced LIGO. This search is referred to as O1MD1 from hereon.

4.1   O1 Multi-Directed Search

A directed search focusses on sources whose sky positions are known from electromagnetic observations, but whose rotation frequency is unknown. Unlike an all-sky search, only one sky position is included in a directed search. Since the most promising targets are typically young objects, it is necessary to extend the search to second order spindown values $\ddot{f}$ so that the phase evolution of the sources can be tracked more precisely.

The O1MD1 search focussed on three known sources that are supernova remnants and are believed to harbour neutron stars — Vela Jr., Cassiopeia A, and G347.3 [13]. The source parameters are summarised in Table 4.1. The

search frequency range for O1MD1 was 20-1500 Hz, which is much larger than the search frequency range for the O1 all-sky search. The search ranges for spindown $\dot{f}$, and $\ddot{f}$ for O1MD1 were defined according to equations 4.1 and 4.2. They depend on the frequency and the lifetime, τ of the source.

$$\dot{f} = [-f/\tau, 0] \qquad (4.1)$$

$$\ddot{f} = [0, \frac{5\,\dot{f}^2}{f}] \qquad (4.2)$$

For each source, the search used different grid spacings and search ranges for $\dot{f}$ and $\ddot{f}$. The grid parameters for each source are summarised in Table 4.2; these were determined using an optimisation method [13] that takes into account a source's age and distance. Both the $\dot{f}$ and $\ddot{f}$ increase linearly with frequency, so their combined parameter space increases quadratically with increasing frequency. Hence, the $\dot{f}$ and $\ddot{f}$ ranges quickly become very large at high frequencies.

DM-off searches are run over the Doppler and spindown frequencies associated with the candidate frequency, $\dot{f}$, $\ddot{f}$, and sky position. Since we now have $\ddot{f}$, the spindown frequency wings are calculated according to equation 4.3, which is the Taylor expansion of frequency up to the second order derivative of frequency. We use equation 2.2 to compute the spindown wings by substituting t with the start time t_{start} and end time t_{stop} of the data we are searching. The Doppler frequency wings are computed using equation 4.4. We calculate the range of detector frequencies f_d that result from the Doppler shift in the astrophysical signal frequency f_o due to the relative motion of the Earth with respect to the astrophysical source, given by v_{Earth}, in the part of its solar orbit that it covers between t_{start} and t_{stop}. The minimum and maximum frequencies from these ranges represent the bounds of detector frequencies that contributed to the search candidate. These detector frequencies represent the search frequency range for the DM-off searches.

$$f(t) = f(T_{ref}) + (t - T_{ref})\dot{f}(T_{ref}) + \frac{1}{2!}(t - T_{ref})^2\ddot{f}(T_{ref}) \qquad (4.3)$$

$$f_d = f_o\frac{(1 + v_{Earth}/c)}{(1 - v_{Earth}/c)} \qquad (4.4)$$

Fig 4.1 shows the range of $\dot{f}$ and $\ddot{f}$ over the entire search frequency range for Vela Jr. The spindown ranges are as large as 6.8×10^{-8} Hz/s at high frequencies, which is more than an order of magnitude larger than the frequency range in the O1 all-sky search. In DM-off searches, we search over

53

both the positive and negative $\dot{f}$ and $\ddot{f}$ to account for a larger possible range of detector artefact behaviour. Hence, the search ranges are twice the values plotted in Fig 4.1. For these values of $\dot{f}$ and $\ddot{f}$, the Doppler and spindown frequency wings of a DM-off search could be close to 1 Hz, compared to a few mHz for the O1 all-sky search. Therefore, the DM-off searches could become extremely expensive even with search grids that are 10 times coarser than the search grids of the original astrophysical searches.

	Vela Jr.	Cassiopeia A	G347.3
Lifetime,	700 years	300 years	1600 years
Right ascension, α (rad)	2.32	6.12	4.50
Declination, δ (rad)	-0.81	1.03	-0.69

Table 4.1: Parameters of the astrophysical sources explored in the O1MD1 search.

	Vela Jr.	Cassiopeia A	G347.3
δf (Hz)	3.2×10^{-7}	6.8×10^{-7}	2.4×10^{-7}
$\delta \dot{f}_c$ (Hz/s)	1.3×10^{-12}	3.9×10^{-12}	6.2×10^{-12}
$\delta \ddot{f}_c$ (Hz/s^2)	1.2×10^{-18}	4.0×10^{-18}	5.1×10^{-18}
γ_1	8	4	8
γ_2	20	20	10

Table 4.2: Search parameters for the three astrophysical sources in the O1MD1 search. δ refers to the grid spacings in f, $\dot{f}$, and $\ddot{f}$. γ_1 is the refinement factor for $\dot{f}$. γ_2 is the refinement factor for $\ddot{f}$.

Since the search setups used for the three sources were different, we treat these as three separate searches. We want to then determine the DM-off search setups for each search separately. In order to determine an appropriate coarseness of the DM-off search grids, the performance of different DM-off search setups were explored for each of the three sources. The $\dot{f}$ and $\ddot{f}$ ranges increase linearly with increasing frequency, causing the search parameter space to increase quadratically. Therefore, DM-off searches for frequencies closer to 1500 Hz will be in general more computationally expensive than for frequencies closer to 20 Hz. The O1MD1 astrophysical search was originally conducted in three frequency ranges and we decided to characterise the DM-off veto separately for each of these frequency ranges. Hence, the search frequency range for the semicoherent DM-off veto was also subdivided into three ranges:
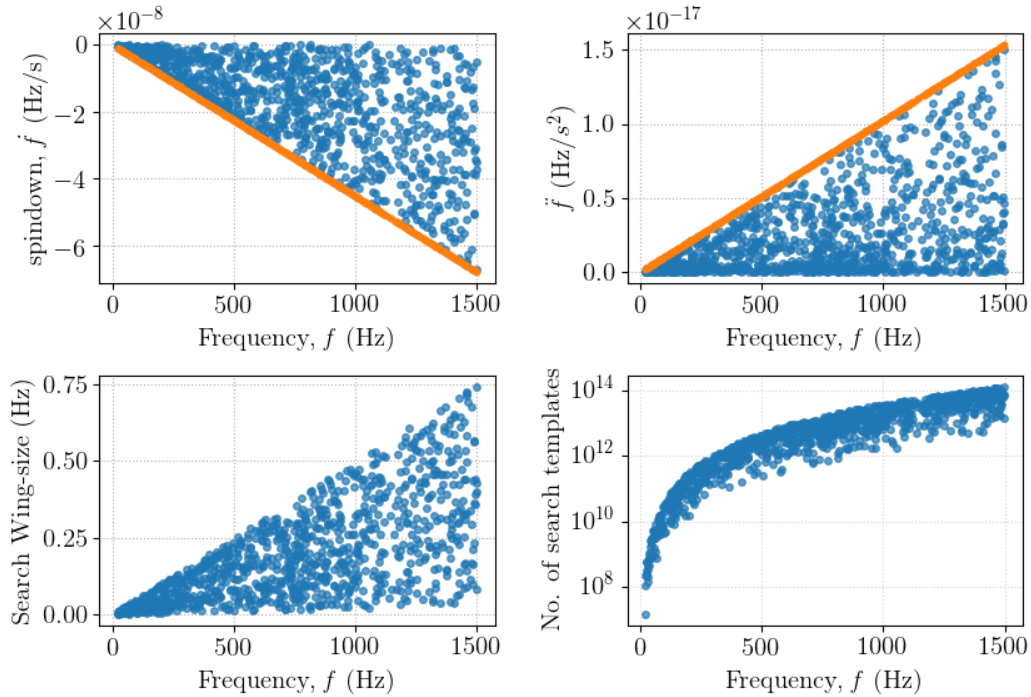
- Low frequency - 20 - 250 Hz

Figure 4.1: The DM-off search parameter space for Vela Jr. with ten times coarser grid spacings than the original search. These parameters were calculated for a thousand fake signals simulated according to the parameters of continuous wave signals expected from Vela Jr. The orange dots represent the maximum $\dot{f}$ and $\ddot{f}$ for the given frequency. The blue dots represent the search parameters for each fake signal.

- Mid frequency - 250 - 520 Hz

- High frequency - 520 - 1500 Hz

The setups were then explored for each range separately.

In this chapter, the method and results of optimising the DM-off search setup for the *low frequency* range are discussed. We conducted DM-off searches with different search setups on observed disturbances in the data from the three searches. The change in F-statistic (*2F*) values and the runtime were computed for each setup. These values were compared to find the optimum coarseness of the DM-off search grid out of the setups that we tested. The searches and analysis were conducted on a supercomputing cluster, AT-LAS.

4.2 Optimisation of the DM-off search setup

The goal of this study was to explore using coarser DM-off search grids to address the potentially enormous computational costs. However, the veto loses power in distinguishing between signals and artefacts if the grids are made too coarse. Therefore, different search setups were tested to pick an optimal search grid.

Like the previous optimisation studies (Chapter 3), we conducted the DM-off searches on known disturbances to observe how the highest value of the DM-off statistic changes when different search setups are used. These studies were conducted on frequency bands that were known to have non-Gaussianities in the astrophysical search. Fig 4.2 shows the disturbed band at 23.95 Hz in data from the Vela Jr. search.

Table 4.2 shows the search grid parameters for each of the sources in the O1MD1 astrophysical search. We varied three of the parameters to obtain the DM-off search grids that we used in our tests — frequency grid spacing δf, $\dot{f}$ grid spacing $\delta \dot{f}$, and $\ddot{f}$ grid spacing $\delta \ddot{f}$. Since detector artefacts have been observed to extend over a larger range of detector frequencies (they are broader) than astrophysical signals, the DM-off search setups had coarser frequency grid spacing δf, spindown grid spacing $\delta \dot{f}$, and $\ddot{f}$ grid spacing δf than the O1MD1 astrophysical searches. Table 4.3 summarises the coarseness of these setups with respect to the O1MD1 search setups. The DM-off $\delta \dot{f}$ is coarser than astrophysical search $\delta \dot{f}$ in all setups because the number of $\dot{f}$ points searched is enormous for the computed $\dot{f}$ ranges which contributes to long computational times. Hence, it was decided to make $\delta \dot{f}$ larger. It was also observed that decreasing the refinement factor does not drastically decrease the computational time but does lower the search sensitivity in the previous optimisation studies (see Chapter 3). Hence, we kept the same refinement factors that were used in the astrophysical searches. A sky-grid with 32 points was used for these searches.

Unlike the studies that were performed on the O1 all-sky search (Chapter 3), the DM-off searches for O1MD1 were conducted separately on the data from the two LIGO detectors - Hanford (H1) and Livingston (L1). The artefacts are assumed to be of terrestrial origin and are therefore expected to be uncorrelated, and each artefact should only appear in one detector. A preliminary simplified DM-off search with zero $\dot{f}$ and $\ddot{f}$ for Vela Jr. revealed that the disturbance (artefact) near 23.95 Hz was present only in L1 (see Fig 4.2). This shows that this disturbance is of terrestrial origin as it only affected one of the detectors.

a.)L1
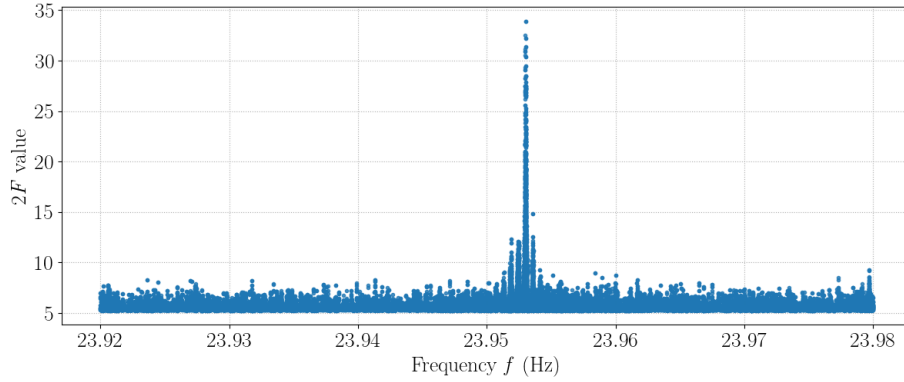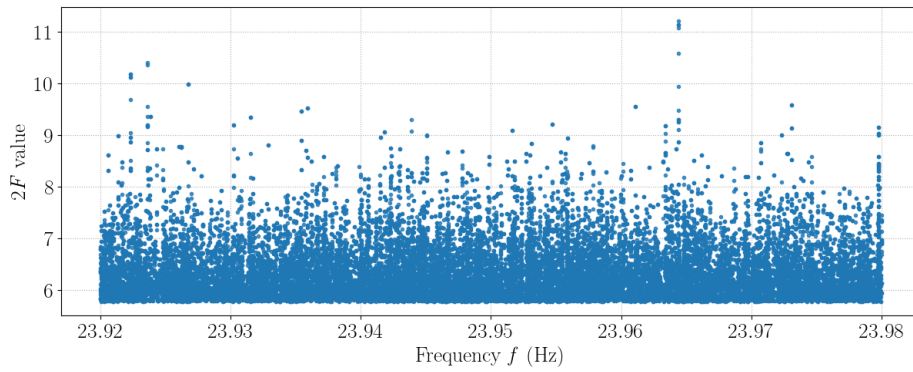
b.)H1
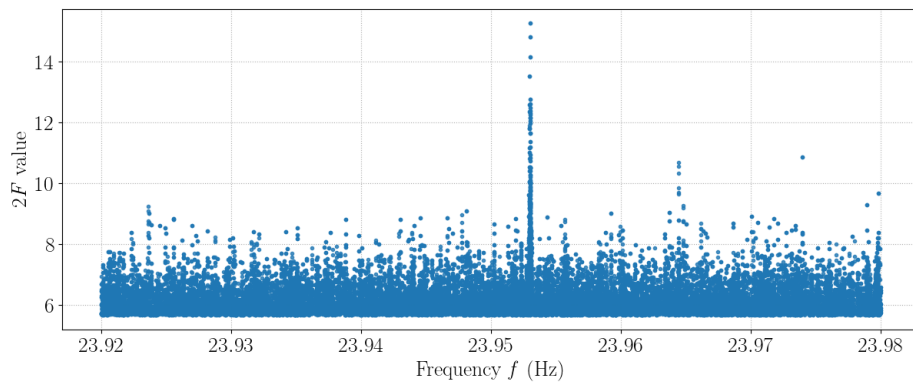
c.)Both detectors

Figure 4.2: Disturbance in the 23.95 Hz band in a.)L1, b.)H1 c.)Both detectors. DM-off searches with zero $\dot{f}$, and $\ddot{f}$ reveal that the disturbance is present only in L1. This band was noted as disturbed in the search results for Vela Jr.

Coarseness in δf	Coarseness in $\delta \dot{f}$	Coarseness in $\delta \ddot{f}$
20×	10×	20×
10×	100×	10×
10×	200×	10×
20×	100×	20×
20×	200×	20×
20×	500×	20×
20×	1000×	20×

Table 4.3: The DM-off search grids explored for the three astrophysical sources in the O1MD1 search. 10× coarseness in δf denotes a ten times coarser frequency grid spacing with respect to the original O1MD1 astrophysical search frequency grid spacings.

It was also observed that the computational runtimes for searches on data from single detectors was smaller than that for searches on data from both detectors. Also, the artefact is observed to be louder in a single detector search, if present in that detector, than in the search in data from both detectors (see Fig 4.2). The artefact loses significance in searches that include data from both detectors if the artefact is absent in one of the detectors, which is generally the case. Hence, it is a better option to perform DM-off searches in data from single detectors.

For the DM-off optimisation studies for the O1 all-sky search, we had a reference DM-off search setup that was used to characterise the semicoherent DM-off veto. Hence, we looked at the reduction in computational cost with respect to this setup, and we could also estimate the number of candidates that could be rejected by using a particular DM-off search setup. However, for the O1MD1 astrophysical search, we do not have any such reference DM-off search setup. Here, we are instead exploring multiple choices for a DM-off search setup to decide which setup would be appropriate for characterising the veto and then applying it to the surviving candidates. Hence, we chose the finest DM-off search grid in our set of grids (20 times coarser in f, 10 times coarser in $\dot{f}$, and 20 times coarser in $\ddot{f}$) as our reference search grid.

For each artefact, each DM-off search was repeated multiple times by varying the start frequency each time to account for a range of mismatch values. The highest $2F$ value was recovered from each trial, and the average of these values was computed for each search grid.

From the runtime information generated by `lalapps_HierarchSearchGCT` [10],

the average runtime was computed for each search grid. This process was repeated for each detector artefact picked from the three astrophysical searches.

4.3 Results and Discussion

For each grid, we calculated the loudest DM-off $2F$ as well as the runtime, averaged over all the trials. We used this information to select an optimal DM-off search grid. The study concerns the low frequency search range (20 to 250 Hz). By comparing the properties of the DM-off search grids, the optimal grids (from the choice of grids evaluated in the study) for each of the three astrophysical sources are chosen.

As expected, the coarser search grids produced lower $2F$ values. This was observed for all the detector artefacts studied here. Fig 4.3 shows the highest DM-off $2F$ values for each setup for the three sources, which correspond to a detector artefact near 23.5 Hz. The artefact was observed near 23.35 Hz in the DM-off search for Cassiopeia A, while it was around 23.98 Hz in the DM-off searches for Vela Jr., and G347.3. This disturbance is present only in L1 which is consistent with it being a detector artefact of terrestrial origin.

Since the detector artefact is absent in H1, the $2F$ values do not show a significant change across the different DM-off search setups. These $2F$ values correspond to random noise fluctuations. The slight drop in $2F$ values can be attributed to the shift toward smaller values of $2F$ in the distribution of highest $2F$ as the total number of trials reduces with coarser search grids.

The main reason for exploring coarse search grids was to find a solution for the potentially huge computational costs of conducting the DM-off studies. Compared to the previous semicoherent DM-off search (Chapter 2), the computational times increased with the addition of $\ddot{f}$ as a search parameter. The runtimes also depend on the search frequency as well as the lifetime τ of the source. Because the $\dot{f}$ range in the astrophysical search increases linearly with frequency, the runtime must increase with increasing frequencies due to a larger $\dot{f}$ search range. For example, the runtime per mHz is on average higher for a search at 50 Hz than for a search at 25 Hz. The spindown range is inversely proportional to the lifetime of the source. This means that the $\dot{f}$ range is smaller for G347.3 than for Cassiopeia A at a particular frequency. This causes the spindown wings to be smaller for G347.3 than for Cassiopeia A at that frequency. Vela Jr. is expected to have intermediate spindown wing sizes as it is younger than G347.3 and older than Cassiopeia A.

4.3.1 Vela Jr.

It can be observed in Fig 4.3a. that the artefact does not lose a lot of its significance until $\delta\dot{f}$ is made 1000 times coarser. Even when the coarseness of the grid is $20 \times 200 \times 20\times$ (20 times coarser δf, 200 times coarser $\delta\dot{f}$ and 20 times coarser $\delta\ddot{f}$ than the grid spacings used in the astrophysical search), the percentage change in $2F$ values is approximately 10% (see Fig 4.4a.).

The runtime per mHz showed a sharp drop when $\delta\dot{f}$ was made 100 times coarser (see Fig 4.5a.). For grids with the same coarseness in $\delta\dot{f}$ but different values of δf and $\delta\ddot{f}$, the runtimes per mHz were similar. This trend can be observed in Fig 4.5a. For example, the computational cost for the grid with $10 \times 200 \times 10\times$ coarseness is lower than that for the grid with $10 \times 100 \times 10\times$ coarseness. When δf and $\delta\ddot{f}$ were made larger, the computational cost remained unaffected. Fig 4.5a. shows that the runtime per mHz for a $20 \times 200 \times 20\times$ coarser grid is equal to that for a $10 \times 200 \times 10\times$ coarser search grid. These are expected results because we vary δf and $\delta\ddot{f}$ by a small proportion whereas we vary $\delta\dot{f}$ the most.

Based on these results, a search grid with 10 times coarser δf and $\delta\ddot{f}$ and 200 times coarser $\delta\dot{f}$ qualifies as the best choice for DM-off searches for Vela Jr.

4.3.2 Cassiopeia A

In the case of Cassiopeia A, the detector artefact is observed to lose very little significance for 20 times coarser δf and $\delta\ddot{f}$ (see Fig 4.3). The percentage change in $2F$ values is less than 5% for a grid that is $20 \times 200 \times 20\times$ coarser. For a grid with $10 \times 200 \times 10\times$ coarseness, the percentage change is close to 0%.

In case of runtimes, it was observed that the computational cost showed a slight reduction when the coarseness of $\delta\dot{f}$ changed from 100 times to 200 times. Fig 4.5b. shows that the runtime per mHz is slightly lower for a $20 \times 200 \times 20\times$ coarser grid than the runtime per mHz for a $20 \times 100 \times 20\times$ coarser grid. However, the computational cost was seen to increase when δf and $\delta\ddot{f}$ were made coarser. This is a counter-intuitive result as one would expect the cost to be higher for a finer grid. This change can be observed in Fig 4.5b. The runtime per mHz is higher for a $20 \times 200 \times 20\times$ coarser search grid than for a $10 \times 200 \times 10\times$ coarser grid.

From these results, a $10 \times 200 \times 10\times$ coarser search grid would be the best choice for DM-off searches in case of Cassiopeia A.

4.3.3 G347.3

Fig 4.3c. shows that the artefact does not lose a lot of significance until $\delta\dot{f}$ is made 500 times coarser. The percentage change in $2F$ values is less than 10% for a $20 \times 200 \times 20\times$ coarser grid (see Fig 4.4c.).

In this case, the computational cost reduced when $\delta\dot{f}$ was made coarser. This trend was observed even in the case of δf and $\delta\ddot{f}$. Fig 4.5c. shows that the runtime per mHz for a $20 \times 100 \times 20\times$ coarser grid is lower than that for a $10 \times 200 \times 10\times$ coarser grid. Similarly, a $20 \times 200 \times 20\times$ coarser search grid has a lower runtime per mHz compared to a $10 \times 200 \times 10\times$ coarser search grid. However, it was observed that grids with $20 \times 100 \times 20\times$ and $10 \times 200 \times 10\times$ coarseness had similar runtimes per mHz.
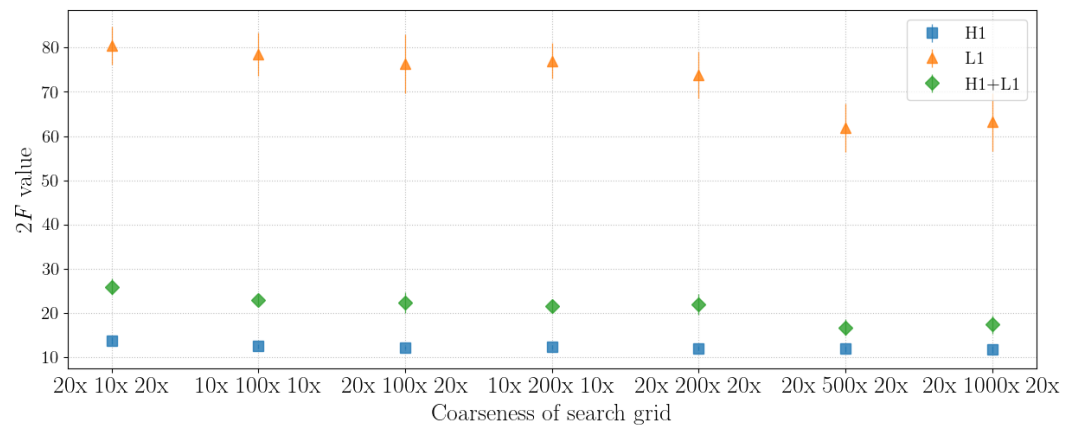
Based on these analyses, the best choice of search grid for G347.3 is the $20 \times 200 \times 20\times$ coarser search grid.
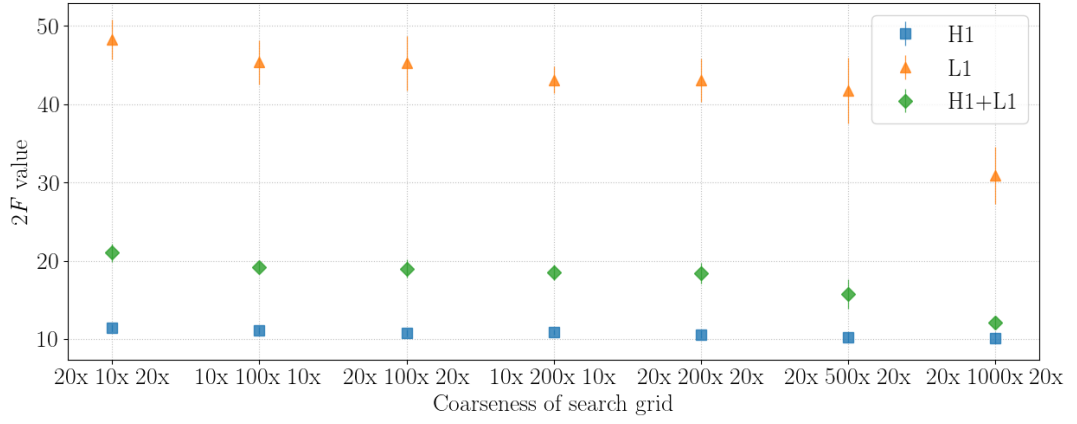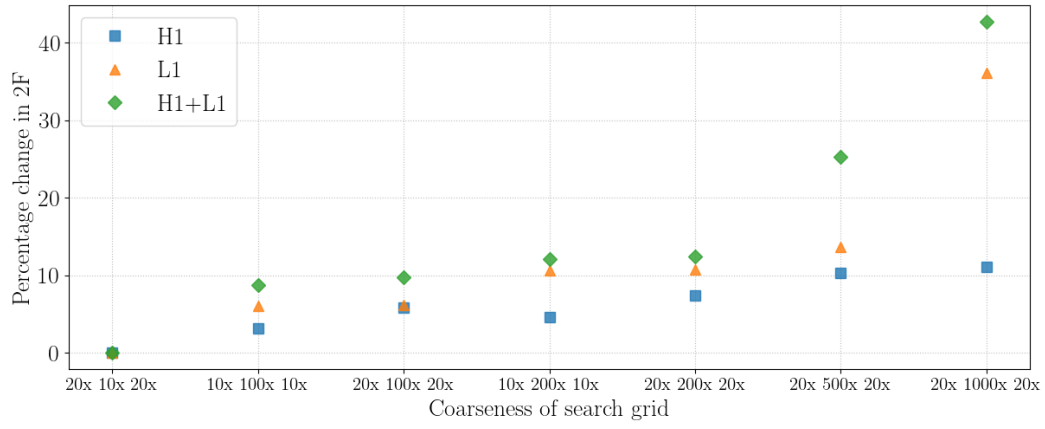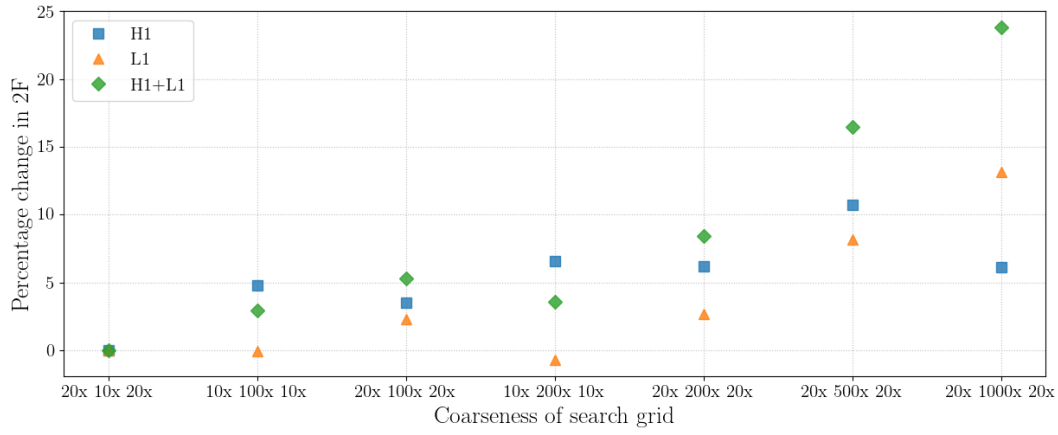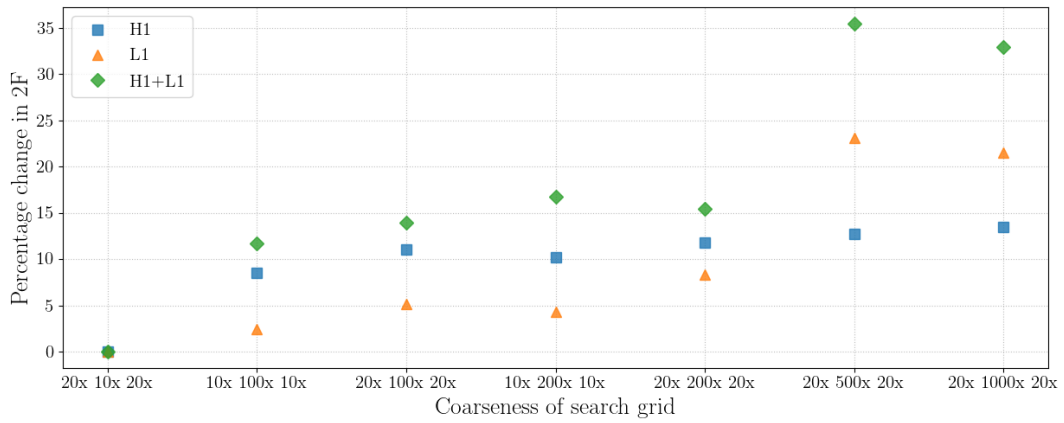
a. Vela Jr.
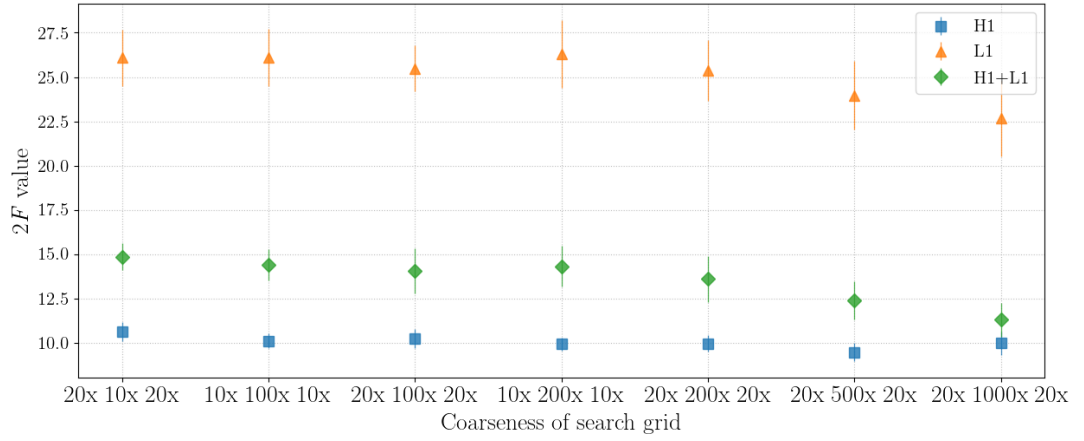
b. Cassiopeia A

c. G347.3

Figure 4.3: Highest $2F$ values recovered for each DM-off search setup for the three astrophysical sources. The artefact is present only in L1 for all three sources. The 2F values for the H1 and the both detector searches are also plotted to show that the the artefact is present in L1 only. This disturbance is present around 23.5 Hz in the three astrophysical searches.
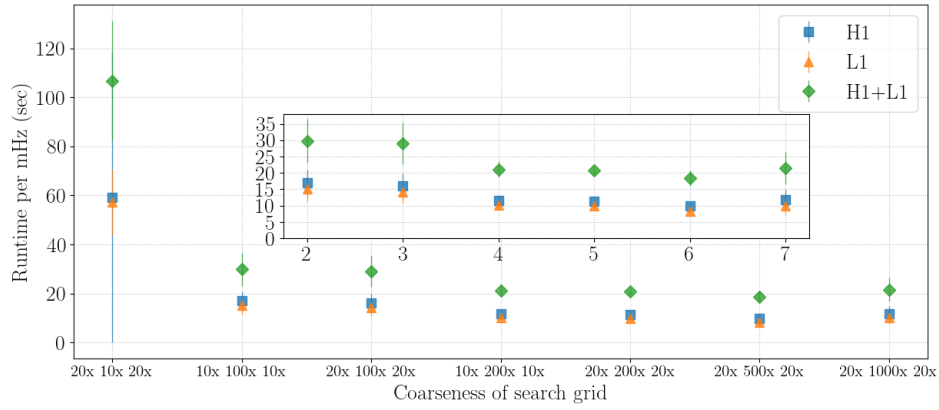
a. Vela Jr.

b. Cassiopeia A

c. G347.3

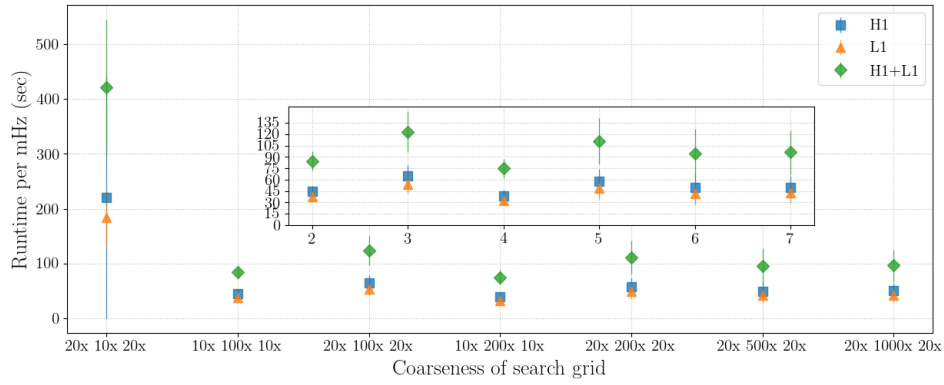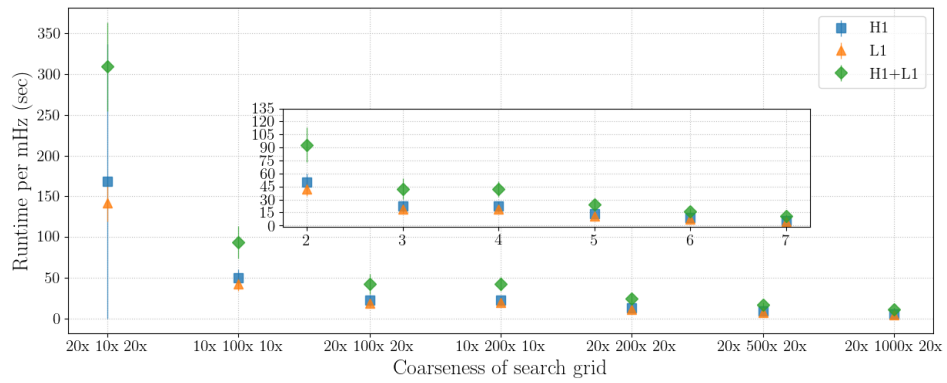Figure 4.4: Percentage change in the highest $2F$ values for different DM-off search setups for the three sources for the disturbance shown in Fig 4.3.

a. Vela Jr.

b. Cassiopeia A

c. G347.3

Figure 4.5: The computational cost associated with each DM-off search setup in terms of the runtime per mHz. The inset plots zoom in on the coarser search setups to clearly observe the trend in the runtimes.

Chapter 5

Conclusion

In the preceding chapters, we looked at the effectiveness of the semicoherent DM-off veto by characterising it for the O1 low-frequency all-sky search and applying the veto to the candidates that survived the first stage of that search. We then investigated whether this veto could be made even more computationally cheap by exploring coarser DM-off search setups for the O1 all-sky search. Finally, we looked at how the veto would have to be modified for the O1 multi-directed search. For this, we examined the DM-off search parameter space that must be searched for setting up the veto. Then, we explored different DM-off search setups to determine the which would be the best choice for characterising the semicoherent DM-off veto for the O1 multi-directed search. In this chapter, I summarise the results from the different studies and present the future plans for the DM-off veto.

5.1 How effective is the semicoherent version of the DM-off veto?

As a proof of concept, the semicoherent DM-off veto was characterised for and applied to Stage 0 candidates of the O1 all-sky search. The veto rejected more than 75% of the candidates. On following-up the surviving candidates through the later stages of the O1 all-sky search, we had only 20 surviving candidates after the last follow-up stage. This would have been a huge improvement for the O1 all-sky search since more than 6000 candidates had originally survived the last stage which was fully-coherent and the most sensitive search. Therefore, we demonstrated that the application of the veto at

an earlier stage of the hierarchical follow-up would have ruled out a majority of candidates arising due to noise and detector artefacts. The semicoherent DM-off veto could effectively reject a majority of spurious candidates at earlier stages of hierarchical continuous wave searches.

5.1.1 How much computational time was saved by the application of the veto at an earlier stage in the hierarchical follow-up of continuous wave candidates?

The semicoherent DM-off searches took 0.7 hours per candidate on average while characterising the veto. The semicoherent DM-off search over the whole frequency range of the O1 all-sky search took 4000 hours. A computational investment of 4750 hours helped us reject over 75% of Stage 0 candidates. The follow-up searches that were originally performed took 4 hours per candidate on an average. Had we with applied the DM-off search to each candidate instead of the standard follow-up search, we would have saved close to 3 hours per candidate, while rejecting a greater number of noise candidates that mimic signal behaviour and gain significance in DM-on follow-up searches. This would have saved more than 203160 hours of computational time that were invested in the follow-up stages performed on the supercomputing cluster.

5.1.2 How can we further improve the veto?

The DM-off searches used for the semicoherent DM-off veto were coarser than the Stage 0 search in frequency f, spindown $\dot{f}$, and sky positions. The f and spindown $\dot{f}$ grid spacings in DM-off searches were ten times coarser in than they were for the astrophysical search. A sky-grid of 60 points distributed across the whole sky was used in DM-off searches. At this coarseness itself, we showed that the semicoherent DM-off veto effectively rejects detector artefacts and saves computational time. The next question we investigated was if we could make the DM-off search grids coarser. We found that the grid spacings in frequency could be made forty times coarser than the Stage 0 frequency grid spacings and a sky-grid with 32 sky positions would be as effective as one with 60 sky-positions.

5.2 How can the veto be optimised and modified for future continuous wave searches?

The underlying concept used in the DM-off veto is that astrophysical signals and detector artefacts behave differently when the Doppler demodulation is turned off in continuous wave search algorithms. Using this concept, we perform a search with the demodulation off on the candidates from a continuous wave search to determine whether the candidate's behaviour is signal-like or non-signal-like. The semicoherent DM-off veto was characterised for the O1 all-sky search to check its effectiveness. In order to characterise the veto for another search, we need to optimise the DM-off search setup in order to define the acceptance/rejection regions. We examined the DM-off search parameter space for a directed search in O1 data for three astrophysical sources — Vela Jr., Cassiopeia A, and G347.3. This multi-directed search was different from the O1 all-sky search. For this astrophysical search, we had a single sky-position instead of the whole sky for each source. In addition, the second time derivative of frequency $\ddot{f}$ was added as a search parameter for the directed searches. These modifications were taken into account while determining the DM-off search parameter space. Consequently, we examined various DM-off search grids with varying coarseness in f, $\dot{f}$, and $\ddot{f}$ grid spacings. Out of the different DM-off search grids, we found that a grid with 10 times coarser f, 200 times coarser $\dot{f}$, and 10 times $\ddot{f}$ grid spacings was the best choice for Vela Jr. and Cassiopeia A, while a grid with 20 times coarser f, 200 times coarser $\dot{f}$, and 20 times $\ddot{f}$ grid spacings was the best choice for G347.3.

5.3 Future directions

We observe that the semicoherent DM-off veto is an effective tool to reject detector artefacts at the earliest stages of hierarchical continuous wave searches. In order to use this tool, we must modify this veto for different continuous wave searches in the future. The semicoherent DM-off veto will be used in the O1 multi-directed search for which the DM-off search setups were optimised. This veto could be used as a line-finding tool to find disturbances in the data arising due to noise in the detector. In the future, the DM-off search could be introduced in the pipelines used for continuous wave searches such that a "non-signal-like" candidate could be rejected at the initial stage of hierarchical searches.

References

[1] B. P. Abbott *et al.* [LIGO Scientific and Virgo Collaborations], *First low-frequency Einstein@Home all-sky search for continuous gravitational waves in Advanced LIGO data*, [arXiv:1707.02669 [gr-qc]].

[2] S. J. Zhu, M. A. Papa, S. Walsh, *A new veto for searches for continuous gravitational waves in LIGO data*, [arXiv:1707.05268 [gr-qc]].

[3] K. Riles, *Gravitational Waves: Sources, Detectors and Searches*, [arXiv:1209.0667v3 [hep-ex]]

[4] LIGO Scientific Collaboration, *Advanced LIGO*, [arXiv:1411.4547v1 [gr-qc]]

[5] B. P. Abbott *et al.* [LIGO Scientific Collaboration and Virgo Collaboration], *GW150914: The Advanced LIGO Detectors in the Era of First Discoveries*, Phys. Rev. Lett. 116 131103 (2016).

[6] B. P. Abbott *et al.* [LIGO Scientific Collaboration and Virgo Collaboration], *GW151226: Observation of Gravitational Waves from a 22-Solar-Mass Binary Black Hole Coalescence*, Phys. Rev. Lett. 116 241103 (2016).

[7] K. Riles, *Recent searches for continuous gravitational waves*, [arXiv:1712.05897v1 [gr-qc]]

[8] Papa *et al.*, *Hierarchical follow-up of sub-threshold candidates of an all-sky Einstein@Home search for continuous gravitational waves on LIGO sixth science run data*, Phys. Rev. D 94, 122006 (2016), [arXiv:1608.08928 [astro-ph.IM]].

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

[10] Lal/lalapps software suite, *http://www.lsc-group.phys.uwm.edu/daswg/projects/lalsuite.html*

[11] J. Aasi *et al.* [LIGO Scientific and Virgo Collaborations], *Directed search for continuous Gravitational Waves from the Galactic center*, Phys. Rev. D 88, 102002 (2013), [arXiv:1309.6221 [gr-qc]].

[12] B. Behnke, M. A. Papa, R. Prix, *Postprocessing methods used in the search for continuous gravitational-wave signals from the Galactic Center*, Phys. Rev. D 91, 064007 (2015), [arXiv:1410.5997 [gr-qc]]

[13] J. Ming *et al.*, *Optimally setting up directed searches for continuous gravitational waves in Advanced LIGO O1 data*, Phys. Rev. D 97, 024051 (2018) [arXiv:1708.02173v2 [gr-qc]]