# Soliton Star Analysis for the Global Network of Optical Magnetometers for Exotic Physics (GNOME)

Perrin Claire Segura

April 13, 2020

# Contents

# Acknowledgements

I would like to thank my research advisor Jason Stalnaker for his valuable insights and endless patience. A special thanks as well to the laser team, the members and former members of Stalnaker Lab at Oberlin College, who have offered their support and advice on this project: current members Dhruv Tandon, Heather Pearson, Eleda Fernald, Jay McClendon, Norah Han, and Sammy Siegel, and former member Sunyool Park.

Of course, many thanks to the entire GNOME collaboration, and to all of those members who have offered their advice and contributed directly or indirectly to this analysis project.

In particular, I would like to thank Derek Jackson Kimball and his group at California State East Bay, especially former member Madeline Monroy for her work developing the early iterations of the consistency check. I would also like to thank Ibrahim Sulai at Bucknell University for his important contributions to the development of the excess power code, and Dmitry Budker and his group at the Helmholtz institute in Mainz. Special thanks to Hector Masia-Roig and Joey Smiga, whose suggestions for the consistency check have proven invaluable.

Last but not least, thanks to my family and friends for their support and encouragement.

This work is supported by NSF grant PHY-1707803.

# Chapter 1

# Introduction

In the mid 20th century, prominent physics professor Philippe Von Jolly told his student Max Planck that physics was nearing its last days as an active area of fundamental discovery, saying, "In this field, almost everything is already discovered, and all that remains is to fill a few holes" [1]. In this regard, he was dramatically mistaken. One of these "holes" he was referring to was the problem known as the "ultraviolet catastrophe." Max Planck was undaunted by Jolly's warning, and continued to pursue a solution to this problem. He eventually discovered that this hole could be "filled" if one assumed that energy is quantized in discrete increments. This assumption was the first step along the path that has eventually led to the development of modern quantum mechanics.

Today, much of what we know about physics is encompassed by the Standard Model. It successfully encompasses the known fundamental particles and (with the exception of gravity) the interactions that govern their behavior in a single mathematical framework [2]. The success of the Standard Model at describing known particles and the electroweak and strong interactions is undeniable, however the questions it leaves open are far from trivial. Learning from the past, we can recognize that these unanswered questions are indicators of directions we can look toward to uncover unexplored physics, potentially accessing entirely new areas of research.

## 1.1 Dark matter

With this in mind, let us focus on a current question that represents a major hole in modern physics knowledge: the nature of dark matter. There are several sources of evidence supporting the existence of a large amount of mass in the universe unaccounted for by known particles. One of the strongest and most well known evidence comes from galaxy rotation curves, which have also been the source of much of our knowledge about the density and distribution of dark matter around galaxies or galaxy clusters [2]. Experimental constraints on formerly promising theoretical candidates such as Weakly Interacting Particles (WIMPS) have now ruled out significant portions of the well motivated parameter, leaving a large range

of potential dark matter candidates (see Ref. [2] for a review of many of these candidates and current constraints).

One of these proposed candidates is ultralight *axion-like particles* (ALPs). This hypothesis places the majority of dark matter mass in the form of spin-0 bosons with mass $m_a c^2 \lesssim 10$ eV [3]. At such a low mass, these particles would function like a classical field, which could in turn form topological defects (e.g. domain walls) that could couple to the spins of known elementary particles, causing spin precession as if the particle was experiencing a magnetic field [3] [4] . However unlike a true magnetic field, the relative coupling would be different for different particles [5]. This "pseudomagnetic" effect would lead to spin precession detectable via standard atomic physics techniques in a low-energy, tabletop experiment [4], [6].

## 1.2 The Global Network of Optical Magnetometers for Exotic physics (GNOME)

The Global Network of Optical Magnetometers for Exotic physics (GNOME) is an experiment designed to be capable of detecting this proposed "exotic spin coupling" using a network of optical magnetometers.

In an optical magnetometer, magnetic field measurements are based on observing the precession of atomic magnetic moments. The magnetic moments due to atomic spins experience a torque in the presence of a magnetic field, causing the spins to precess around the perpendicular field component [7]. By measuring this precession using standard atomic physics techniques, a precise magnetic field measurement can be obtained. Detailed information on the principles of optical magnetometry can be found in reference [7]. There are several different types of magnetometers currently active in the GNOME network, each of which uses one of three different types of atomic species: Rubidium, Potassium, and Cesium. Though the specific methods vary from station to station, all magnetometers in the GNOME network do share an one important trait: they are all vector magnetometers which are only sensitive to the projection of the magnetic field along a sensitive axis [8].

Such an instrument is well suited for detecting topological defects in dark matter, because the spin coupling to the exotic ALP field would result in a precession of the alkali atoms that mimics the effects of a magnetic field [6]. By surrounding the vapor cell in multiple layers of magnetic shielding to reduce the effects of actual magnetic fields, we can search for transient "pseudo-magnetic" signals that may represent evidence of exotic spin interactions. The cells are additionally surrounded by multiple sets of compensation coils allowing control over the components of the magnetic field and magnetic field gradient withing the shields [8]. These shields do not impact the the exotic nuclear couplings for which GNOME is searching [9].

Even with the magnetic shielding and compensation coils, the data from a single magnetometer are not free of "transient events" of perfectly mundane origin. For example, changes in laser characteristics, fluctuations in lab temperature, physical vibrations, and electronic noise all can produce fluctuations in the magnetic field that would not be readily distinguishable from a transient exotic interaction. For this reason,

the GNOME experiment uses a network of such magnetometer stations distributed at various locations. Searching for correlations between multiple distant stations allows legitimate signals to be distinguished from false positives. Employing a network of multiple magnetometers not only allows for vetoing false positives based on coincident, but enables the collaboration to use the vector characteristic of the magnetometers (each one is sensitive along a different direction) to gain additional information about the direction, magnitude, and velocity of the triggering event.

The data are collected and time-stamped via GPS. The magnetic field data are sampled at a rate of 512 Hz, and the detector bandwidths are all $\approx 100$ Hz. These data are collected using a box designed by S. Pustlney's group at Jagiellonian University in Krakow, Poland, and uploaded to a server at the Helmholtz Institute in Mainz, Germany for analysis. The stations have sensitivity on the order of $10 - 100$ fT$/\sqrt{\text{Hz}}$ [8]. In order to further improve the ability of the network to distinguish potentially relevant deviations in the magnetic field, each station also collects "sanity" data at a 1 Hz sampling rate. This "sanity" value is a binary value that is used to flag certain data as invalid based on known external characteristics. A "sanity box" designed and built by a group from the University of Fribourg in Switzerland uses information from a built-in accelerometer, magnetometer, and thermometer as well as output from important experimental components to automatically flag the data as "insane" [8]. The data can also be manually flagged during modifications such as laser alignment or magnetic field trimming. An important attribute of the analysis process is the ability to successfully ignore flagged or missing data. The operating characteristics of GNOME impose several important constraints on the characteristics of detectable events. First, the GNOME network is searching for transient signals, present for a "brief" period as the earth passes through an astrophysical structure of finite size. This sets GNOME apart from experiments like Cosmic Axion Spin Precession Experiment (CASPEr) (Ref. [10]), which searches for persistent oscillatory signals. The second constraint is the 100 Hz bandwidth and 512 Hz sampling rate of the GNOME detectors, which further limits range of ALP masses to which GNOME is sensitive to $m_a c^2 \leq 0.01$ neV [8]. Finally, there is an additional constraint on the scale of the object necessary to make use of the spatially distributed network. If the structure with finite size in all directions (e.g. soliton stars) and is smaller than the Earth, then we cannot assume that every station will be triggered by the event and the broadly distributed network is no longer as effective. Therefore, we assume that the structures GNOME is sensitive to are present at scales much larger than the size of the Earth

More detailed information on the motivation and scientific goals of GNOME can be found in Ref. [6], and a detailed description of the six magnetometers active in the early network can be found in reference [8]. At the time of writing, an additional five magnetometers have been constructed, and one of the original six was relocated completely. One of these new magnetometers is located at Oberlin College, details on its operation and construction can be found in reference [11].

One of the strengths of the GNOME network is its potential for generating constraints for any type of transient exotic spin-coupling, which encompasses a variety of theories including ultralight dark matter,

potentially locating signs of new physics without requiring prior theoretical prediction. Since so little is known about where the answers to major questions of particle physics will come from or what form they will take, this generality is certainly advantageous. However, preserving this generality in the analysis mechanisms is exceedingly difficult, and the dominant strategy among the GNOME collaboration so far has been to develop multiple analysis methods tailored to specific theories, including Refs. [12] and [13]. By imposing further assumptions to narrow the scope of the analysis to a particular category of theoretically motivated structures and adjusting the analysis method to suit these assumptions, we can improve the speed and sensitivity to make a more careful search for types of signals backed by theoretical motivation. In light of this, we at Oberlin have developed an analysis method aimed at a specific but important category of proposed dark matter structures: soliton stars.

## 1.3 Soliton stars

An important subcategory of the structures that could be formed by ultra-light dark matter are soliton stars, compact astronomical objects held together by self interactions [3]. Q-balls (or Q-stars) are a further subset of soliton structures with promising theoretical motivation and several particularly convenient features. One such feature is that Q-ball formation and stability is theoretically feasible for a comparatively wide range of particle masses [14]. Another important feature is Q-balls formed from ultra-light mass particles can potentially interact with standard model particles, particularly protons, but have not yet been ruled out by prior experiments [14]. Another similar type of self-interacting soliton star is the "axion star", which is a more specific structure formed from coalescence of ALP dark matter in the early universe into bound "clumps" as the result of a specific production mechanism [3]. Together, these hypothetical structures are known as soliton stars, and they form an important theoretically motivated exotic search target for the GNOME magnetometer network.

If such a structure exists, it is useful to estimate the possible duration of such a signal, which will inform our decisions when constructing a search method. Assuming the simplest case that all dark matter is contained in soliton stars and that the Earth travels the full diameter of the axion star, the time during which the entire earth is within the axion start is given by:

$$\Delta t = \frac{2R - 2R_E}{v} \tag{1.1}$$

where $v \sim 10^{-3}$ c is the the relative velocity between the solar system and virialized DM [8], $R_E = 6400$ km is the radius of the Earth, and $R$ is the radius of the axion star. Ref. [3] estimates the upper bound for $R$ based on the constraint that GNOME is able to detect at least 1 event within a 1 year observation period, $R_{max} \sim 10^6 R_E$. Plugging this value into equation 1.1, we obtain $\delta t_{max} = 42 \times 10^6$ s or 1.4 yrs, which is clearly not feasible as a "transient" signal in a 1 yr observation period. For a lower bound, GNOME is sensitive to structures that exist in least one direction on a scale much larger than the earth, so the approximately spherical soliton stars must have $R >> R_E$. For $R = nR_E$, we have $\Delta t \sim 40(n-1)$ s. Note that if the

Earth were to actually pass through an axion star, the distance traversed would be along some chord of a sphere with length $k$. If one considers $k$ as a function of the perpendicular offset $l$ from the parallel diameter (i.e. how far off the actual path is from our "ideal" path), the derivative is inversely proportional to $R$. In intuitive terms this simply means that as you consider larger soliton stars, there are more ways in which the Earth can cross along a path with length $k \approx 2R$. We can further quantify this estimation using Dirac's chord method for estimating the average chord length of a convex solid [15]

$$k_{avg} = \frac{4V}{S} \tag{1.2}$$

Where $V$ is the volume and $S$ is the surface area. With some caveats, this equation provides the expected path length for a particle traveling through a sphere. For a spherical axion star, this reduces to $k_{avg} = \frac{4}{3}R$, where $R$ is the radius of the axion star. We can estimate some additional values to find a lower bound on $R$ using this relationship. Ultimately, for this purposes of this analysis, a reasonable concrete estimate for the lower bound is $n = 4$, with $\Delta t \sim 100$ s. This is small compared to the estimate in [3], which puts the lower bound at $10R_E$, resulting in $\Delta t \sim 400$ s.

Ref. [3] describes two main methods through which the soliton stars could to couple to the spins of elementary particles: the linear coupling and the quadratic coupling. The following is a summary of the conducted in this reference, with the goal of understanding what type of signals we can look for, as well as the likelihood that a coupling of this form could be detected by GNOME.

In the linear coupling model, the oscillating ALP scalar field $a(\vec{r}, t)$ is real-valued, and couples to the spin of particle $j$ though the Hamiltonian

$$H_{l,j} = \frac{\hbar c}{f_{l,j}} \vec{S}_j \cdot \vec{\nabla} a \tag{1.3}$$

where $j = n$, $p$, or $e$ corresponds respectively to neutron, proton, or electron, $f_{l,j}$ is related to the coupling constant with particle $j$, and $\vec{S}_j$ is the spin of particle $j$. In this model, the time dependence of the oscillatory field $a(\vec{r}, t)$ is maintained in the Hamiltonian 1.3, the result of which is that atomic spins with the axion star will experience an oscillating energy shift. This detected signal would be represented in the form of an oscillating burst detectable while the station is within the interior of the soliton star, which should make this model a strong candidate for detection by GNOME. However, under the conditions considered in [3] (mean free path for terrestrial encounters $L = 10^{-3}$ ly, relative velocity $v = 10^{-3}c$, and radius $R = 10R_E$), the region of linear coupling parameter space accessible to the current version of the GNOME network is ruled out by astrophysical constraints (assuming the number density is such that the encounter rate with the Earth is one per year). If the network is fully upgraded to the "Advanced GNOME" co-magnetometers sensitive to neutron coupling, then there remains regions of the accessible parameter space that are still unconstrained.

The second model considered in Ref. [3] is the quadratic coupling model, which is a necessary modification to predict a feasible Hamiltonian in the case of a complex scalar axion field (e.g. of the form $a(\vec{r}, t) = \phi(\vec{r})e^{i\omega t}$, such as that proposed in Ref. [14]. There are two important versions of this quadratic Hamiltonian described

in Ref. [3]. The first one is

$$H_{q,j} = \frac{\hbar c}{f_{q,j}^2} \vec{S_j} \cdot i \left( a^* \vec{\nabla} a - (\vec{\nabla} a^*) a \right) \tag{1.4}$$

and the second is

$$H_{q,j} = \frac{\hbar c}{f_{q,j}^2} \vec{S_j} \cdot i |\vec{\nabla} a|^2 \tag{1.5}$$

where $f_{q,j}$ is related to the quadratic coupling constant with particle $j$ [3]. However, if the axion scalar field $a$ is complex, then the time-varying component of $a$ and $\vec{\nabla} a$, $e^{i\omega t}$, will cancel out due to the factor $a^* \vec{\nabla} a - (\vec{\nabla} a^*)$ in equation 1.4 and $|\vec{\nabla} a|^2$ in equation 1.5, resulting in a signal which only couples to the spin as the Earth crosses the boundaries entering and exiting the wall, making this type of analysis better suited for short transient detection methods such as the one developed to search the GNOME data for domain walls described in Ref. [12]. On the other hand, if the complex field in the quadratic Hamiltonian is replaced with the real-valued oscillating field used in the linear Hamiltonian, equation 1.5 takes the form

$$H_{q,j} = \frac{\hbar c}{f_{q,j}^2} \vec{S_j} \cdot i \vec{\nabla} a^2 \ . \tag{1.6}$$

However, this circumstance is not as well motivated and has not yet been subjected to additional analysis. Nevertheless, with so little known about these hypothetical structures, it can be argued that the important characteristics of this signal (an oscillatory burst with duration on the order of 100 s or longer) are still sufficiently generic to motivate a specific analysis program, which can be adapted to meet the specific needs of other models as needed in the future.

# Chapter 2

# The analysis method

The analysis method is divided into two stages: The coincidence and the consistency check. In addition to the assumptions imposed by the network, this analysis method imposes an additional four key assumptions. First, that the signal will specifically take the form of an oscillatory burst, as predicted for a model of from a quadratic coupling with a real coefficient [3]. The most important implication of this assumption is that we expect the burst to endure for the entire period that the Earth is inside the soliton star. This is related to the second main assumption, that a soliton star is much larger than the Earth, so that the entire Earth is inside the soliton star at some point, and oscillating transient signal is "long" enough that we do not have to worry about taking into account delayed signal times due to the finite velocity of the Earth passing through the soliton star. At some point, if the entire earth is contained in the soliton star, all signals will be experiencing a signal at the same time, for a duration $\gtrsim 100$ s which from equation 1.1 corresponds roughly to a soliton star radius $R \gtrsim 4R_E$, a reasonable lower bound for satisfaction of the condition $R >> R_E$. The third assumption is that the exotic spin couplings of the different atomic species used in different magnetometers are approximately the same, and are not taken into account when predicting the relative magnetometer amplitudes. This is justified for the electron coupling based on the predictions in Ref. [8], where the electron coupling scaling factor is expected to be $\approx 0.5$ the same for all of the listed magnetometer types. However, the proton coupling is not similarly constant, and with the addition of more magnetometer types to the network since [8] was published, this should be be reevaluated and accounted for in subsequent improvements. The fourth and final assumption states that while the different atomic species do not impact the relative amplitudes we expect to measure in each magnetometer, the direction from which the event propagates does. Recall that each of the GNOME vector magnetometers can only detect events along a specific sensitive axis. This means that the expected signal measurement in each detector depends on the relative orientation between this sensitive axis and the propagation direction of the event, an important idea that is explored in more detail in section 2.2.2.

The analysis method was designed with these assumptions in mind. Further restrictions are imposed as

necessary. The result incorporates aspects of analysis methods described in Refs. [13] and [12]. The analysis in Ref. [13] searches specifically for Exotic Light Fields (ELFs) accompanying known astrophysical events. Since the times of interest are localized in specific a window around the observed event, this analysis can search for signals by comparing the time series within a specific window to portions of the time series outside of the window, where it is assumed that there are no true events. The analysis in Ref. [12] is geared toward shorter transient signals such as those caused by a domain wall or soliton star envelope, and therefore cannot make use of the assumption that the signal will be present in every station at once, and instead must take into account the time delay as structure propagates a finite velocity to interact with the different stations. The search technique developed here can be applied more broadly than the single window search of the former, but is faster than the more general latter.

## 2.1 Stage 1: the excess power coincidence search

The goal of the first stage is to locate in each individual station any transient oscillatory signals that could represent an "event" where the Earth passes through a soliton star. Potential events then undergo a coincidence check to determine if a similar signal was located at the same time and frequency in at least three other stations in the network. Events that pass the coincidence check are sent to the second stage.

For each magnetometer in the network, the analysis must first locate transient signals that deviate from the typical noise of the detector. For a single detector, the range of accessible signals is constrained by the network (a transient signal $\lesssim 100$ Hz), and the soliton star assumptions (an oscillatory burst with duration $\gtrsim 50$ s). The first stage analysis aims to effectively locate signals while limiting the number of additional assumptions which must be imposed. For this reason, we have implemented a search based on the Excess Power (EP) statistic. Our implementation of this method was inspired by the LIGO analysis described in Ref. [16]. This statistic can be calculated relatively easily (section 2.1.1), is well characterized statistically (section 2.1.1), and was demonstrated in Ref.[16] to be the optimal search method in the case of an unknown waveform, making it a strong choice for our initial analysis stage.

### 2.1.1 Calculating the Excess Power Statistic

Consider a time series of $n$ magnetic field samples

$$\tilde{a}_j = \tilde{c_j} + \tilde{h_j}, \tag{2.1}$$

where $\tilde{a}_j$ is the measured value in pT, $\tilde{c_j}$ is a transient oscillatory signal due to the earth passing through a soliton star, and $\tilde{h}_j$ is noise. Note that the notation of using the tilde for the time series data is opposite from what is typically used in literature. This was done for the purpose of visual clarity, since the frequency domain values are used almost exclusively for the remainder of the analysis. For the purposes of breaking down the analysis method, assume that this noise is a random variable from a Gaussian distribution with

mean $\langle \tilde{h}_j \rangle = 0$ and standard deviation $\sigma = 1$. We first divide the time series into $n_t$ time segments consisting of $N$ samples each. Fourier transforming each segments gives a set of frequency channels, with value $a_{kj}$ at time segment $j$ and frequency channel $k$ given by the equation

$$a_{kj} = \sum_{m=jN}^{2jN-1} a_m e^{\frac{-2\pi i m k}{N}} \ . \tag{2.2}$$

From this, we define the (one-sided) power spectral density (PSD) for a single time frequency tile $(j, k)$ as:

$$x_{kj} = \frac{1}{N^2}\left(|a_{kj}|^2 + |a_{(N-k),j}|^2\right) \qquad \text{for k ranging from 0 to } \frac{N}{2}+1 \tag{2.3}$$

$$= \frac{2}{N^2}|a_{kj}|^2 \qquad \text{since the time series is real valued} \tag{2.4}$$

where the leading term $\frac{1}{N^2}$ normalizes the sum according to the convention suggested from in Ref. [17] so that for each time segment the sum over all the frequency values is equal to the mean-square amplitude of the time series (within that segment of time), i.e. for time segment $j$,

$$\frac{1}{N}\sum_{i=jN}^{2jN-1} \tilde{a}_i^2 = \frac{2}{N^2}\sum_{k=0}^{N/2+1} |a_{kj}|^2 \ . \tag{2.5}$$

Note that this value has units of $\mathrm{pT}^2$ rather than $\mathrm{pT}^2/\mathrm{Hz}$, and so by some conventions (for example the python spectrogram function documentation) should actually be referred to as a *Power Spectrum* (PS) rather than a true Power Spectral Density. However, the terms are often used interchangeably elsewhere (for example in Refs. [16] and [18], and other pages of the scipy documentation), and since the difference lies in a normalization factor that cancels out in calculating the excess power statistic, the distinction is not relevant at this stage of the analysis. For each frequency band $k$, we calculate the time averaged PSD $\langle x_j \rangle_k$. Now all the pieces are in place to calculate the excess power statistic $\epsilon_{kj}$ for each (frequency, time) tile $(k, j)$:

$$\epsilon_{kj} = \frac{2x_{kj}}{\langle x_j \rangle_k} \ . \tag{2.6}$$

By "normalizing" the PSD value of each time-frequency block based on the average PSD at that frequency, we can eliminate the effects of signals that we consider part of the "typical" noise of the detector, such as the 50 or 60 Hz line frequency present on all of the GNOME detectors. Since these signals are present for the entire time duration, they will be canceled out. Transient signals of the type we are searching for will have smaller effects on the average value (the shorter the signal compared to the total time, the less be effected), and so transient signals that are not part of typical detector noise will retain their "excess power." This process is illustrated in an exaggerated manner in figure 2.1.

To understand how the value of $\epsilon$ can be interpreted quantitatively, it is illuminating to first expand this statistic into the individual signal and noise components. Since the Fourier transform of the signal plus noise

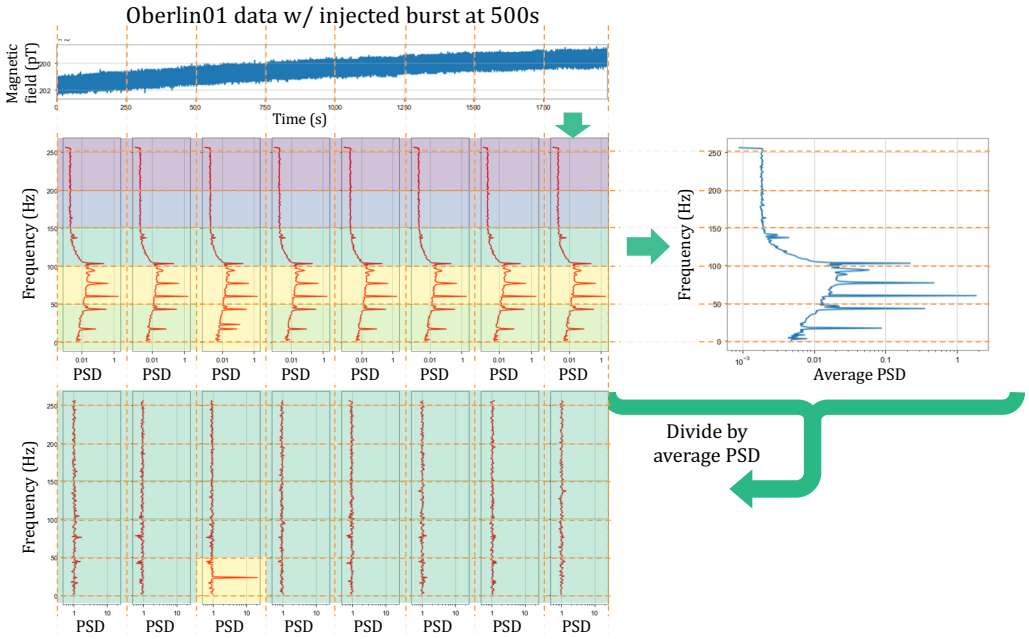


Figure 2.1: A "cartoon" deconstruction of a time series with an injected 30 Hz oscillating signal beginning at 500 s and lasting 100 s total, demonstrating how the excess power statistic is calculated. First, the magnetic field time series is broken into segments of width $\delta t$ (vertical dashed lines). Each of these segments is decomposed in to its constituent frequency components (dashed horizontal lines). Together this process creates a spectrogram, a grid of time-frequency tiles with a PSD value for each tile. To normalize the spectrogram, the average PSD is calculated for each frequency channel, then each time tile within the corresponding frequency band is divided by this value. This eliminates signals that otherwise have large PSD values (such as the 60 Hz line frequency), while leaving the power in transient signals such as the injected 30 Hz burst starting at 500 seconds shown in this example.

is simply the sum of the two components, we can reduce equation 2.2 to

$$a_{kj} = c_{kj} + n_{kj} \tag{2.7}$$

$$\Rightarrow |a_{kj}|^2 = |c_{kj}|^2 + |h_{kj}|^2 + 2\Re\left(h^*_{kj} c_{kj}\right) \tag{2.8}$$

$$\Rightarrow \epsilon_{kj} = \frac{2\left(|c_{kj}|^2 + |h_{kj}|^2 + 2\Re(h^*_{kj} c_{kj})\right)}{\frac{1}{n_t}\left(\sum_{j=0}^{n_t-1} |c_{kj}|^2 + \sum_{j=0}^{n_t-1} |h_{kj}|^2 + 2\sum_{j=0}^{n_t-1} \Re(h^*_{kj} c_{kj})\right)} \,. \tag{2.9}$$

Assume that in a given frequency channel, there is a constant signal with complex constant Fourier components $c_k = u + iv$ present in $n_s < n_t$ segments, and the signal has an amplitude $|c_{kj}| = C$ that is equal in all segments $j$ where the signal is present. The noise Fourier components $h_{kj}$ can be similarly represented as a complex number of the form $h_{kj} = p_{kj} + iq_{kj}$, however instead of constants, $p_{kj}$ and $q_{kj}$ are Gaussian distributed random variables, with means $\langle p_j \rangle_k = \langle q_j \rangle_k = 0$ [19]. We can rewrite 2.9 as

$$\epsilon_{kj} \approx \frac{2\left(C^2 + |h_{kj}|^2 + 2(up_{kj} + vq_{kj})\right)}{\frac{n_s}{n_t}\left(C^2 + 2u(\frac{1}{n_s}\sum_{j=0}^{n_s-1} p_{kj}) + 2v(\frac{1}{n_s}\sum_{j=0}^{n_s-1} q_{kj})\right) + \frac{1}{n_t}\sum_{j=0}^{n_t-1}|h_{j,k}|^2} \tag{2.10}$$

$$\approx \frac{2(C^2 + |h_{kj}|^2 + 2(up_{kj} + vq_{kj}))}{\frac{n_s}{n_t}\left(C^2 + 2(u\langle p_j\rangle_k + v\langle q_j\rangle_k)\right) + \langle |h_j|^2\rangle_k} \tag{2.11}$$

$$\approx \frac{2\left(C^2 + |h_{kj}|^2 + 2(up_{kj} + vq_{kj})\right)}{\frac{n_s}{n_t}C^2 + \langle |h_j|^2\rangle_k} \ . \tag{2.12}$$

If no signal is present, then $c_k = 0$ for all time segments $j$ and frequency channels $k$, and Eqn. 2.9 reduces to

$$\epsilon_{kj} = \frac{2|h_{kj}|^2}{\langle |h_j|^2\rangle_k} \tag{2.13}$$

note that for this case, we expect that the average value $\langle \epsilon_{kj} \rangle = 2$. Assuming that our original time series values $\tilde{h}_j$ are Gaussian random noise with mean value 0 and standard deviation $\sigma = 1$, then the their values Fourier transformed keep these properties [17]. The normalized PSD values are therefore the squares of Gaussian random variables, which means that they are distributed as a $\chi^2$ distribution, in this case with 2 degrees of freedom [20] [16]. Note that in the description in Ref. [20], the 'sum' of squares with one Gaussian random variable should give a distribution with 1 degree of freedom. The extra factor of 2 is added to the value of $\epsilon$ in equation 2.6, and is related to the original time series: for a 'tile' with time segment width $\delta t$ s and frequency channel width $\delta f = 1/\delta t$ Hz, the information in the tile can be minimally represented by a time series with sampling rate $2\delta f$, so the minimum number of time series points needed to represent a single tile in this case is $(2\delta f)\delta t = 2$ [16]. To be consistent with the derived statistic in Ref. [16], we have included the extra factor of 2 in Eqn. 2.6. Expanding this idea to summing over multiple time segments or frequency channels, the average value becomes

$$\langle \epsilon \rangle = 2\left\langle \sum_{k=1}^{N_f} \left( \frac{\sum_{j=1}^{N_t} |h_{j,k}|^2}{\langle |h_j|^2\rangle_k} \right) \right\rangle \tag{2.14}$$

$$= \frac{2N_f N_t \langle |h_{j,k}|^2\rangle}{\langle |h_{j,k}|^2\rangle} \tag{2.15}$$

$$= \frac{2\Delta T \Delta F}{\delta t \delta f} \tag{2.16}$$

$$= 2\Delta T \Delta F \ , \tag{2.17}$$

where $N_f$ is the number of frequency channels summed to make one tile, $N_t$ is the number of time segments summed in one tile, $\Delta T$ is the time length of each summed tile, and $\Delta F$ is the frequency width of each tile. This is once again a sum of squares of $N_f N_t$ Gaussian random variables (times a factor of 2), and thus also follows a $\chi^2$ distribution with $2N_f N_t = 2\Delta F \Delta T$ degrees of freedom.

We have now shown that, if we assume that the noise on our detector is Gaussian, then the results of equation 2.6 from each tile should follow a $\chi^2$ distribution with $2V = 2\Delta F \Delta T$ degrees of freedom. The resulting probability distribution function (from Ref. [20]) is

$$f_{\chi^2}(x, V) = \frac{x^{V-1}e^{-x/2}}{2^V \Gamma(V)} \ , \tag{2.18}$$

where

$$\Gamma(V) = \int_0^{\inf} x^{V-1}e^{-x}dx \ .$$

From this, we can assign for a given measured value of $\epsilon$, a numerical probability that that value was measured from a time series only containing noise. The probability of measuring excess power $\epsilon$ that is greater or equal to some value $\epsilon^*$ is (from Ref. [16])

$$P(\epsilon \geq \epsilon^*) = \int_{\epsilon^*}^{\inf} f_{\chi^2}(x, V)dx \tag{2.19}$$

$$= \int_{\epsilon^*}^{\inf} \frac{x^{V-1}e^{-x/2}}{2^V \Gamma(V)}dx \ . \tag{2.20}$$

This value provides a quantitative method of determining how much power qualifies as "excess power:" in a single station. We set a threshold value $\epsilon^*$ based on the desired false positive rate, and pass only events with excess power $\epsilon \geq \epsilon^*$. The theoretical and actual curves are compared in figures 2.2 and 3.2. Now that we know the probability of measuring a particular excess power in a tile, we can set an excess power threshold $\epsilon^*$ and flag only events with excess power above the threshold.

Now that that we have a quantitative methods of estimating the false positives probability for a single station, we need a method of estimating the false negative probability. Define the *sensitivity level* as the minimum signal size that produces significant excess power, i.e. the signal power $|c_{kj}|^2$ needed to produce a value of $\epsilon_{kj} \geq \epsilon^*$ for a given value of $\epsilon^*$. In order maximize the sensitivity, we must minimize this level by ensuring that for a given signal size, the value of $\epsilon$ resulting from the signal is as large as possible. Returning to equation 2.12, notice that one clear way by which this can be accomplished is by minimizing the ratio $\frac{n_s}{n_f}$. approaches The method used to accomplish this is discussed in section 2.1.1, but for now, note that as $\frac{n_s}{n_t} \to 0$, Eqn. 2.12 approaches

$$\epsilon_{j,k} \to \frac{2(C^2 + |h_{j,k}|^2 + \Re(h^*_{j,k}c_{j,k}))}{\langle |h_j|^2 \rangle_k} \tag{2.21}$$

$$= \frac{2(C^2 + \Re(h^*_{j,k}c_{j,k}))}{\langle |h_j|^2 \rangle_k} + \frac{2|h_{kj}|^2}{\langle |h_j|^2 \rangle_k} \ . \tag{2.22}$$

From which we can calculate the mean value $\langle \epsilon_{kj} \rangle$ as

$$\langle \epsilon_{kj} \rangle = \frac{2C^2}{\langle |h_{kj}|^2 \rangle} + 2 \tag{2.23}$$

$$= 2(A^2 + 1) \ , \tag{2.24}$$

where $A^2$ is the signal to noise ratio of the PSD values. While a formal proof is beyond the scope of this thesis, it is explained in Refs. [16] and [19] that the values of this distribution, with a signal present, are

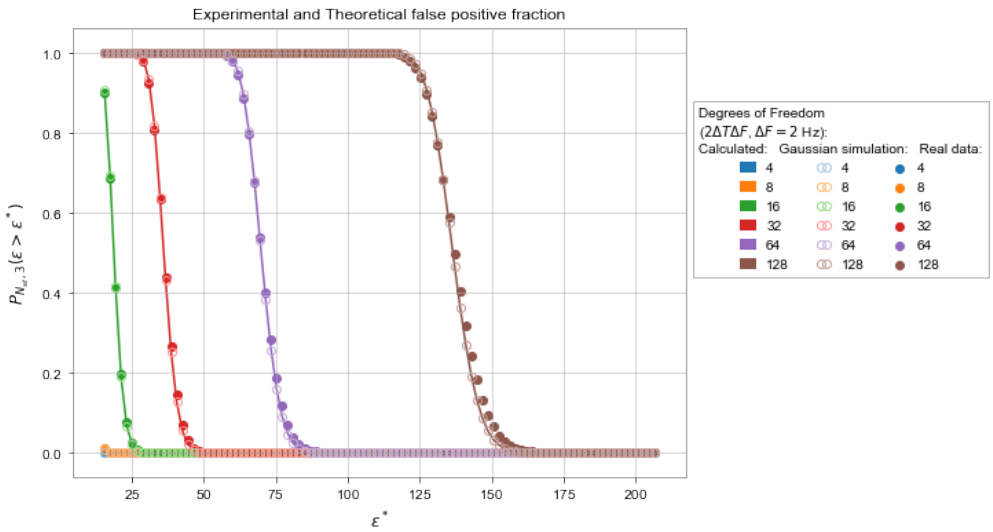


Figure 2.2: Theoretical and experimental false positive rates vs. excess power threshold for a single iteration of the excess power coincidence search. The solid curves are the theoretical values from combining equations 2.20 and 2.27, assuming, with 9 active stations and a 3-fold coincidence. The solid points are the average fractional false positives for 5 consecutive 2000 s windows of real time-shifted data. The unfilled circles are the average fractional false positives for 5 consecutive 2000 s windows of simulated Gaussian data with mean 0 and standard deviation 1, emulating the ideal case from which equation 2.20 was derived. This graph is the same figure as 3.3, plotted on a linear scale to highlight the overall agreement between the real, Gaussian, and theoretical data.

distributed as a non-central $\chi^2$ distribution, with non-centrality parameter $A^2$ [16]. This distribution has an approximate closed form solution [16]

$$f_{\tilde{\chi}^2}(\epsilon, A^2, V) = \frac{1}{2} e^{-(\epsilon+2A^2)/2} \left(\sqrt{\frac{\epsilon}{2A^2}}\right)^{V-1} I_{V-1}(\sqrt{2A^2\epsilon}) \tag{2.25}$$

where $I_{V-1}$ is a modified Bessel function of the first kind described in Ref. [19]. The probability of measuring $\epsilon \geq \epsilon^*$ with a signal with power spectrum SNR $A^2$ is given by integrating equation 2.25

$$P(\epsilon \geq \epsilon^* | A^2) = \int_{\epsilon^*}^{\inf} f_{\tilde{\chi}^2}(x, A^2, V) dx \ . \tag{2.26}$$

Both theoretical and experimental (using Gaussian data) plots of $A^2$ vs. detection probability are shown in figure 3.6.

This method only becomes practical once we expand the scope of the search to multiple detectors. This introduces a new element to the analysis: the coincidence check. After calculating excess power for each station individually, we look for coincidences were at least $N_c$ stations demonstrate excess power $\epsilon > \epsilon^*$ in the same time/frequency tile. Assumption 2 comes into play at this stage; because we are looking for "long" signals on the order of 100 s, we can assume that there will be at least one tile's worth of time where

the signals experienced by the different stations overlap, and so it is possible to ignore the finite velocity with which the event would propagate across the Earth. This in turn allows this analysis method to be simpler and faster in comparison to the more general search technique described in Ref. [12], which requires scanning over the space of possible velocities. By including the requirement of $N_c$ fold coincidence out of $N_{st}$ stations, we must reevaluate the calculation of false positive probability associated with excess power cutoff $\epsilon^*$. For each station, the probability of measuring excess power $\epsilon > \epsilon^*$ is given by $P$ from Eqn. 2.20, so the probability $P_{N_c}(\epsilon^*)$ in measuring excess power in at least $N_c$ stations out of $N_{st}$ is given by:

$$P_{N_c}(\epsilon^*) = \sum_{k=N_c}^{N_{st}} \binom{N_{st}}{k} P^k (1-P)^{N_{st}-k} \ . \tag{2.27}$$

Using this value, we can set a threshold $\epsilon^*$ so that we obtain the desired false positive and false negative rate.

**Implementation**

Implementing this theoretical statistic requires a few additional tricks and considerations. The actual excess power process is outlined in the flow chart in figure 2.3. To begin the excess power test, we create a spectrogram with tiles of dimension $\delta f \delta t$. The 0th channel is then discarded, since frequencies in this channel have oscillation period longer than the width of a segment, and thus the PSD is not accurately estimated. In this initial spectrogram (Fig 2.7), signals such as 60 Hz noise (in the American stations) are distinctly visible. The next step is to "normalize" the spectrogram by averaging over all time segments in each frequency channel, and dividing each tile by the average, producing an array of tiles with time frequency volume $\delta t \delta f$, with each tile described by:

$$\epsilon_{kj} = 2 \sum_{l=kN_f}^{2kN_f-1} \left( \frac{\sum_{m=jN_t}^{2jN_t-1} |a_{ml}|^2}{\frac{1}{n_t} \sum_{m=0}^{n_t} |a_{ml}|^2} \right) \tag{2.28}$$

where $N_f$ is the number of frequency channels per tile, $N_t$ is the number of segments per tile, and $n_t$ is the total number of time segments.

The next step is to sum together $N_t$ time segments and $N_f$ frequency channels to produce an array of tiles with time-frequency dimensions $N_t * \delta t = \Delta T$ by $N_f * \delta f = \Delta F$. In the time domain, creating a tile by summing multiple time segments is advantageous in reducing the number of tiles over which the signal is spread out. Averaging these tiles after the excess power also reduces the uncertainly in the measurement of un-normalized tile power $x_{kj}$, which is essential during the second stage (see section 3.2).

It is not immediately evident why one would want to sum frequency channels (not average, since we need to preserve the definition of PSD from equation 2.5 [17]), since this reduces the frequency resolution without the apparent corresponding improvement in signal accuracy (since the signal is expected to be different in adjacent frequencies, but roughly the same in adjacent tiles). The benefit of summing adjacent frequency tiles is the resulting reduction in the effects of spectral leakage. Spectral leakage is described in detail in Ref

[17], the essential issue is that when estimating a PSD of a finite set of data, the power produced in a given frequency channel is actually based on the convolution of the ideal, continuous PSD that would be obtained from an infinitely long time series and the PSD of the window used to truncate the 'infinite' data into the finite time segment.

The implication of this process is that for each frequency bin $k$, the primary contribution to the PSD value is from frequency $f_k = k\delta f$, with additional significant contributions from frequencies in the range $f_k - \delta f/2$ through $f_k + \delta f/2$, and much smaller contributions from all other frequencies $f$ that are not integer multiples of $\delta f$, with this contribution decreasing as the difference $|f_k - f|$ increases. As suggested in Ref. [17], spectral leakage from a particular frequency into "far away" tiles can be reduced by changing the window function, which broadens the central peak while reducing external leakage. This was deemed unnecessary for our current analysis, because the issue we have primarily encountered is not the excess leakage from signals outside of the desired bandwidth (which would still qualify as a valid coincidence), but loss of signals that span multiple tiles. To correct for this issue, we sum frequency tiles in two separate spectrograms: the "up" spectrogram (throwing out channel 0 and summing tiles 1 and 2, 3 and 4, etc.) and the "down" spectrogram (summing tiles 0 and 1, 2 and 3, 4 and 5, etc.).

We then set the desired false positive probability, and calculate the corresponding excess power cutoff $\epsilon^*$ from equation 2.27. Using this cutoff, we create a 2 new "binary spectrograms" for each station (one for "up" and one for "down"), with the excess power values replaced with 1's and 0's corresponding to values over or under the threshold, respectively. The next step is to simply add the spectrograms from each station together, as in the two bottom left graph in figure 2.8. For any tile with time frequency value greater than or equal to 3, we know that at least 3 stations had excess power at that same time and frequency, and the potential event is flagged and ready to be passed to the next stage.

Once this is complete, the result is a collection of normalized time-frequency tiles with values given by equation 2.12. The next step, therefore, is to maximize the sensitivity by reducing this equation to equation 2.22. Note that for the first excess power calculation described above, the sensitivity is limited by the length of the signal. The shorter the signal duration is compared to the total length of the time series, the smaller the ratio $\frac{n_s}{n_t}$ becomes. The simplest way to improve sensitivity, therefore, is to increase the total duration of the time series window, assuming you are not including any new signals by doing so. This upper limit would depend on how frequently we can expect to see signals, which is estimated in Ref. [3]. However, there is another important factor that places an upper limit on the total length of time that can be analyzed at once. The excess power analysis hinges on the idea that the station noise is *stationary*. If this is not true, and the noise changes with time, then the average value over the whole time series is no longer a good approximation of the noise value at any single point, and so the noise will not be properly normalized in equation 2.6. On short time scales of $\lesssim 2000$ s, the noise on the stations can be reasonably approximated as stationary, however on longer time scales, this assumption begins to break down. This trade-off between access to longer signals and maintaining stationary noise imposes limits on how we can improve the sensitivity by changing

the total time analyzed.

To partially circumvent this trade-off, we perform a trick that reduces the value of $\frac{n_s}{n_f}$ without increasing the value of $n$: instead of increasing $n_t$ (by increasing $n$), we *decrease* $n_s$. This may seem strange, since, if we fix $N$, $n_s$ is not a value that is actually under the experimenter's direct control. This is true for the first time we run the excess power check, where we are limited by equation 2.9. However, we then implement a series of loops. After going through the full coincidence check once, we can flag all tiles with $\epsilon \geq \epsilon^*$ coincidence in $N_c$ stations and repeat the process, ignoring these flagged tiles when calculating the average value. As long as at least one signal tile was flagged with excess power on the first try, $n_s$ will decrease, and the sensitivity for the next loop is improved. We can also set a more relaxed false positive rate for the early tests to ensure that some tiles are flagged, then push the cutoff back up to $5\sigma$ for the last loop. The process is repeated until there is no chance in the number of tiles flagged, ideally indicating that the value of $\frac{n_s}{n_t}$ has gone to 0, and leaving a grid of tiles with excess power that is reasonably approximated by equation 2.22.

Once the final loop has finished, we are left with two sets of flagged events, one for "up" and one for "down". We then average the summed time tiles to produce a final set of $2N_{st}$ spectrograms with tile values given by:

$$x_{kj} = \frac{4}{N_t N^2} \sum_{m=kN_f}^{(k+1)N_f-1} \left( \sum_{i=jN_t}^{(j+1)N_t-1} |a_{mi}|^2 \right) . \tag{2.29}$$

We then use the events to recalculate the masked average one final time. However this time, instead of dividing by the masked average, we subtract the average values in preparation for the consistency check, resulting in another set of $2N_st$ spectrograms with tiles described by

$$s_{kj} = x_{kj} - \frac{4}{N_t N^2} \sum_{m=kN_f}^{(k+1)N_f-1} \left( \frac{1}{(n_t - n_s)} \sum_{l \notin J_k} |a_{lm}|^2 \right) \tag{2.30}$$

$$= x_{kj} - \frac{4}{N_t N^2} \sum_{m=kN_f}^{(k+1)N_f-1} \langle |h_j|^2 \rangle_m \tag{2.31}$$

$$\approx x_{kj} - N_f H \tag{2.32}$$

where $J_k$ is the set of time indices $j$ in row $k$ where potential events were flagged by the final coincidence check, and H is the resulting average PSD value of the tile noise. On the real stations, the noise power will differ significantly based on the station. The extra factor of two is introduced so that the result is converted from rms amplitude squared to amplitude squared (the square of 1/2 the peak to peak difference of the oscillating signal) so that when simulated bursts are injected, the output $\vec{x}$ and $\vec{s}$ values can be compared at a glance to the expected values from the injected signals (the rms values differ by a factor of 2, so this decision has no impact on the actual experimental results). By subtracting the average value (calculated excluding tiles with flagged events), we can remove these differing offsets from each station value. This allows us to safely assume that these PSD values measured from every station are measured from the same 0 baseline, an assumption that is critical to the validity of the upcoming consistency check. Any values with $s_{kj} > 0$ pT

are set to $s_{kj} = 0$ pT, which is permissible since values with power smaller than the average power could not contain detectable events with $C^2 > 0$. This process is summarized in the flow chart in figure 2.3, and an example product of each stage shown in figures 2.7 through 2.9. The resulting spectrograms and a record of which tiles contain potential events is then passed on to the seconds analysis stage.

### 2.1.2 Calculating uncertainties

The final step of the first stage is to associate an uncertainty value with each PSD measurement. Note that because of the cross term in equation 2.8, the component of measured value that fluctuates due to noise (and thus the desired uncertainty value) depends on the strength of the signal $C^2$. If we truly has Gaussian noise that was identical in every frequency band, then when no tiles are summed, then the equation for the best estimate in the PSD tile value is derived in Ref. [18] as

$$\delta s_{kj} = \sqrt{2Nb\sigma^2 x_{kj} - (Nb\sigma^2)^2} \tag{2.33}$$

where $N$ is the PSD length, $b = \frac{4}{N^2}$ is the PSD normalization term, and $\sigma$ is the standard deviation of the original time series. To understand the structure of this equation, note that for a normally distributed random variable $h$, the expected value $\langle |h|^2 \rangle = \sigma^2$. So for tile with no signal present, the average value is approximately $Nb\sigma^2$ [18]. So for $x_{kj} = b|h_{kj}|^2 \approx b\langle |h_{jk}|^2 \rangle \approx Nb\sigma^2$, equation 2.33 reduces to

$$\delta s_i = Nb\sigma^2 \ . \tag{2.34}$$

We can generalize this estimate to frequency dependent non-Gaussian noise by using the same structure, but estimating an experimental value of $Nb\sigma^2$ using the tiles in the same frequency band that do not contain a flagged signal. So we assume that $Nb\sigma^2 \approx \sigma_k^{(PSD)}$, where $\sigma^{(PSD)}$ is the standard deviation of the PSD values $x_{kj}$ in the frequency band $k$, excluding any tiles flagged by the excess power coincidence check. The new estimate becomes

$$\delta s_{kj} = \sqrt{2\left(\sigma_k^{(PSD)}\right) x_{kj} - \left(\sigma_k^{(PSD)}\right)^2} \ . \tag{2.35}$$

When including the effects of summing time and frequency tiles, this equation becomes

$$\delta s_{kj} = \sqrt{2\left(\sigma_k^{(PSD)}\right) \frac{x_{kj}}{\sqrt{N_f N_t}} - \left(\sigma_k^{(PSD)}\right)^2} \ . \tag{2.36}$$

Note that equating the expected uncertainty with the expected average value in the absence of a signal is only valid when no tiles are summed. As the number of time tiles summed increases, the uncertainty $\delta_{kj}$ decreases by a factor of $\frac{1}{\sqrt{N_t}}$, but the overall average value remains constant.

## 2.2 Stage 2: the consistency check

The second stage of the analysis is made possible due to the 4th soliton star assumption, that the soliton star signal can be described (up to on overall sign) as a three dimensional vector $\vec{m} = (m_x, m_y, m_z)$, with

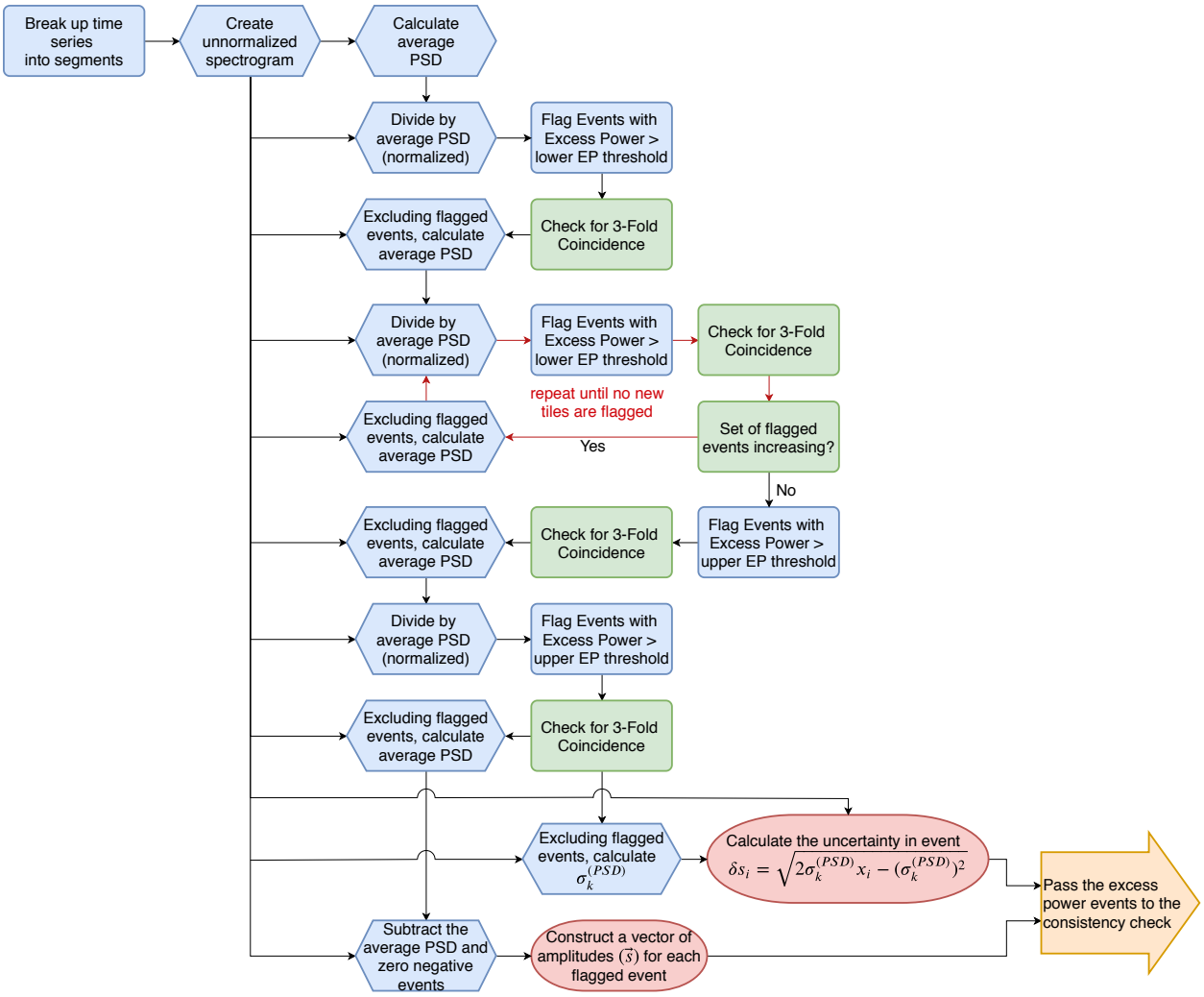


Figure 2.3: A flow chart representing the first stage of the analysis method, the excess power coincidence check.

magnitude equal to the oscillation amplitude ($C$) and direction relative to a coordinate axis fixed relative to the center of the earth (see figure 2.4). The important implication of this assumption is that the PSD value measured from each station depends on both the magnitude and the direction of the incident signal. Recall that the optical magnetometers that make up the GNOME network are vector magnetometers, sensitive only to magnetic field components along a particular axis.

### 2.2.1 Constructing the potential event vectors

The results from the first analysis stage provide three important sets of values. For each station, there are two $S = [s_{k,j}]$ spectrograms, which contain the PSD values with the mean subtracted and negative values set to 0 (from Eqn. 2.32, and two $X = [x_{kj}]$ spectrograms, with the raw PSD values (from Eqn. 2.29). The third is a list of $(j, k)$ coordinate pairs of the (time, frequency) tiles that were flagged by the excess power

coincidence check as potential events.

Consider the case of a single event flagged in tile $(j, k)$. Using the above values, we can construct a vector of the relative event amplitudes from each station:

$$\vec{s} = \begin{pmatrix} s_1 \pm \delta s_1 \\ \vdots \\ s_i \pm \delta s_i \\ \vdots \\ s_{N_{st}} \pm \delta s_{N_{st}} \end{pmatrix}$$

where $s_i$ is the subtracted PSD value from station $i$, and $\delta s_i$ is the uncertainty calculated via equation 2.35.

### 2.2.2 Checking for directional consistency

To check for directional consistency, we use a $\chi^2$ test adapted from the the suggestion proposed by GNOME collaborator Joey Smiga in Ref. [21]. Each vector magnetometer $i$ has a single sensitive direction, which can be represented by 3 dimensional unit vector $\hat{d}_i = (d_i^{(x)}, d_i^{(y)}, d_i^{(z)})$. The components of this vector are in turn measured with respect to the fixed coordinate system shown in figure 2.4, which is defined relative to the center of the earth with the the $+\hat{z}$ direction pointing up through the north pole. For an event propagating as $\vec{m}$, the expected value $\mu_i$ measured in station $i$ can be represented by the linear equation

$$\mu_i = \hat{d}_i \cdot \vec{m} = d_i^{(x)} m_x + d_i^{(y)} m_y + d_i^{(z)} m_z \ . \tag{2.37}$$

We can combine the unit vectors $\hat{d}_i$ into the rows of a single matrix $D$, and the expected values $\mu_i$ into a single vector $\vec{\mu}$. The resulting equation for $\vec{\mu}$ as a function of the event $\vec{m}$ is

$$\vec{\mu}(\vec{m}) = D\vec{m} = \begin{pmatrix} d_1^x & d_1^y & d_1^z \\ \vdots & \vdots & \vdots \\ d_i^x & d_i^y & d_i^z \\ \vdots & \vdots & \vdots \\ d_N^x & d_N^y & d_N^z \end{pmatrix} \begin{pmatrix} m_x \\ m_y \\ m_z \end{pmatrix} \ . \tag{2.38}$$

Now consider the vector of $N_{st}$ *measured* values $\vec{a}$. If the event described by $\vec{a}$ is a true event, then there must exist some 3-d vector $\vec{m}$ such that $\vec{\mu}(\vec{m}) \approx \vec{a}$. For a measured value of $\vec{a}$, we can solve for the value of $\vec{m}$ that gives the closest approximation $\vec{\mu} \approx \vec{a}$ by minimizing the $\chi^2$ value [21]

$$\chi^2(\vec{m}) = \sum_{i=1}^{N_{st}} \frac{(\mu_i(\vec{m}) - a_i)^2}{(\delta a)_i^2} \ . \tag{2.39}$$

This equation has a simple analytic solution which can be found by zeroing the gradient equation 2.39 with

respect to the components of $m$. Setting $\vec{\nabla}_{m_x,m_y,m_z}\chi^2(\vec{m}) = \vec{0}$ and simplifying results in the equation

$$D^T\left(\text{diag}\left\{\frac{1}{(\delta a)_i^2}\right\}\right)D\vec{m} = D^T\left(\text{diag}\left\{\frac{1}{(\delta a)_i^2}\right\}\right)\vec{a}$$
$$\Rightarrow D^T\Delta^{-2}D\vec{m} = D^T\Delta^{-2}\vec{a}\ , \tag{2.40}$$

where the $N_{st} \times N_{st}$ diagonal matrix $\Delta^{-2} \equiv \text{diag}\left\{\frac{1}{(\delta a)_i^2}\right\}$. Rearrange equation 2.40 to get the final result

$$\vec{m} = (D^T\Delta^{-2}D)^{-1}D^T\Delta^{-2}\vec{a}\ . \tag{2.41}$$

Note that if we group the elements of the matrix $D$ by column rather than row, we can obtain a different geometric interpretation of our original equation for $\vec{\mu}$,

$$\vec{\mu}(\vec{m}) = D\vec{m} = \begin{pmatrix} d_1^x & d_1^y & d_1^z \\ \vdots & \vdots & \vdots \\ d_N^x & d_N^y & d_N^z \end{pmatrix}\begin{pmatrix} m_x \\ m_y \\ m_z \end{pmatrix} = \begin{pmatrix} \vec{X} & \vec{Y} & \vec{Z} \end{pmatrix}\vec{m}\ .$$

Illustrated in figure 2.4, these columns correspond to the $N_{st}$ dimensional vectors representing expected events incident along the $\hat{x}$, $\hat{y}$, and $\hat{z}$ directions, forming a basis for the 3 dimensional subspace $A \subset \mathbb{R}^{N_{st}}$ of possible events. If we assume that $(\delta a_i)^2$ is constant for all $i$, then equation 2.40 becomes:

$$D^TD\vec{m} = D^T\vec{a}$$
$$\Rightarrow \vec{m} = (D^TD)^{-1}D^T\vec{a}$$
$$\Rightarrow D\vec{m} = D(D^TD)^{-1}D^T\vec{a}$$
$$\Rightarrow \vec{\mu} = P_D\vec{a}\ ,$$

where $P_D \equiv D(D^TD)^{-1}D^T$ is the projection matrix for the subspace $A$. This means that the value of $\vec{\mu}(\vec{m})$ produces as a result of the $\chi^2$ minimization procedure can be intuitively interpreted as the *weighted projection* of the measured event vector into the space of possible events. The square 3 matrix $(D^TD)$ is also symmetric and invertible, "nice" properties that allow for computational shortcuts [22].

There is one significant problem with this method. A vector $\vec{s}$ of measured values passed down from the first analysis stage consists of PSD values, which are related to the *square* of the desired value (i.e. $s_i \approx (\hat{d}_i \cdot \vec{m})^2$ for $1 \leq i \leq N_{st}$). If we are limited to the PSD vector $\vec{s}$ as defined in equation 2.32, we can construct a vector with components $a_i = \sqrt{s}$, which are correct up to a sign for each component. Since the overall vector is defined up to a sign, for each event we must calculate $2^{N_{st}-1}$ $\chi^2$ values for each of the distinct sign combinations for the components of $\vec{a}$. It is likely possible to use nonlinear optimization methods to minimize equation 2.39 using $s_i$ instead of $a_i$ and redefining $\mu_i(\vec{m}) = (\hat{d}_i \cdot \vec{m})^2$, however such a solution no longer presents the intuitive interpretation presented by the linear version, and so is left for future work.

### 2.2.3 The p value check

To assess the agreement between the measured vector $\vec{a}$ and the "closest" possible signal $\vec{\mu}(\vec{m})$, we can plug the value of $\vec{m}$ calculated from equation 2.41 back into the $\chi^2$ equation 2.39, to obtain a $\chi^2$ value for the

event. These values should follow a $\chi^2$ distribution with $v = N_{st} - 3$ degrees of freedom. From this value, we can use the cumulative $\chi^2$ distribution (as in equation 2.19) to calculate the probability $p_{\chi^2}$ of obtaining a $\chi^2$ value greater than or equal to the value obtained given $N_{st} - 3$ degrees of freedom. In other words, it is the probability of measuring a signal $\vec{a}$ due to noise with standard deviation $\vec{\delta a}$ on the vector $\vec{\mu}(\vec{m})$. We can then set a cutoff value $p^*$, accepting only events with $p_{\chi^2} \geq p^*$. This leaves us with a false negative probability of $p^*$ [21].

### 2.2.4 The magnitude check

The p value helps us quantitatively veto inconsistent events, however there is a hole in this process that must be closed by a final additional check. The event $\vec{\mu}(\vec{m}) = \vec{0}$ is still technically a consistent event, and so any events that managed to pass through the excess power check that are still consistent with the noise on the station (which is more likely to be an issue when the noise is non-stationary) would not be caught by the excess power check or the p value check. The third check imposes a threshold on the magnitude $m = \|\vec{m}\|$ of the calculated m vector. The uncertainty on $m$ is given by [21]

$$\delta m = \frac{1}{m}\sqrt{\vec{m}^T \left(D^T \Delta^{-2} D\right)^{-1} \vec{m}} \ . \tag{2.42}$$

With the value $m \pm \delta m$ for each event, we can set a threshold value $m^*$ to veto events with magnitude less than $m^*(\delta m)$ from 0. i.e. accept only events with

$$m - m^*(\delta m) > 0 \ , \tag{2.43}$$

where $n > 1$. Following the same example injected event from figures 2.7 through 2.9, the results of the consistency check are represented in figures 2.5 and 2.6.

In figure 2.5, the amplitude of the best fit $\vec{m}$ is plotted vs. flagged event index, for the full set of events passed from stage one (figure 2.9). Since the injected frequency of 30.5 Hz is half way between the center frequencies of the two adjacent channels, there is a clear difference in amplitude between the events from the "up" spectrogram (the first $\sim$ 100 points), and the "down" spectrogram (the last $\sim$ 200 points). In the first set of points, the adjacent frequency channels (30 Hz and 31 Hz) were correctly summed together to form a single event tile, although the original injected event amplitude of 1.0 pT was not completely recovered, (the average value of $m$ for these events is 0.9 pT). In the last set, the channels containing power from the injected signal were not summed together (i.e. 30 Hz was summed with 29 Hz and 31 Hz was summed with 32 Hz), resulting in twice as many tiles containing signals of a smaller amplitude (the average value of $m$ of these tiles is 0.7 pT). The following figure 2.6 shows the $p$ values for the same events plotted vs. event index. Note that since all the flagged tiles are due to a real event, those that did not pass the $p$ threshold check are false negatives. However, given the threshold $p^* = 0.01$ used for this example, we expect a false negative rate of 0.01. Two false positives out of 294 events gives $2/294 = 0.007$, which is still consistent with this expectation.

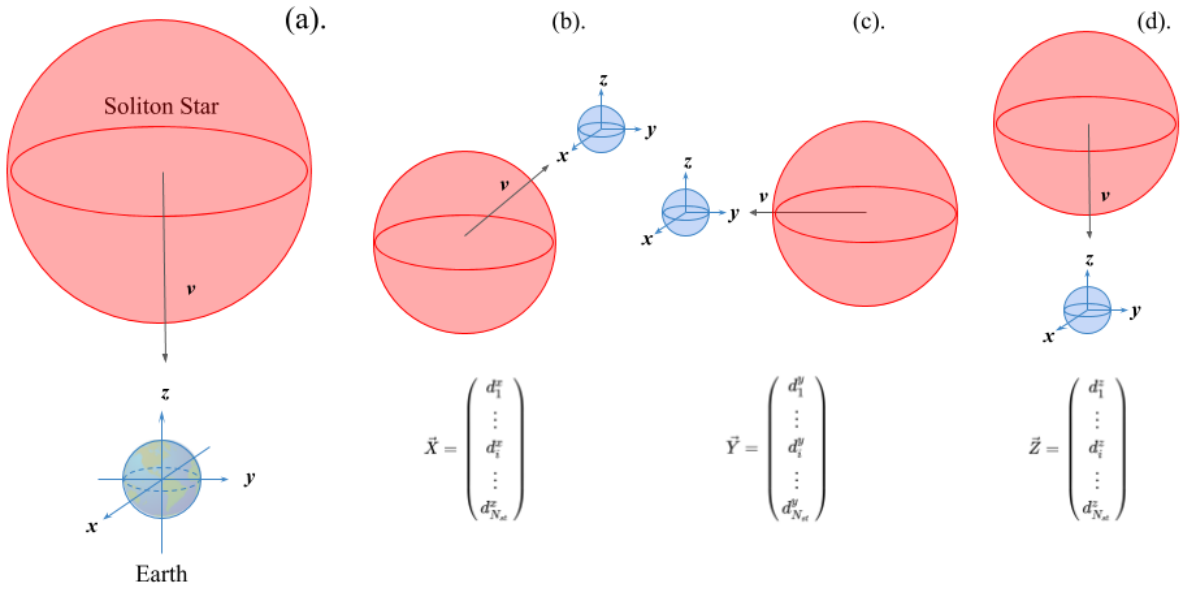


Figure 2.4: By defining a coordinate system with respect to the center of the earth, we can represent a soliton star event as a vector $\vec{m}$, where the direction $\vec{m}$ corresponds to the direction of propagation (represented in this figure as a velocity vector $v$), and the magnitude represents the amplitude of the "pseudomagnetic" field that would be measured by a detector with sensitive axis parallel to $\vec{m}$. The vector $\vec{m}$ of the event, and each vector $\vec{d_i}$ representing the sensitive axis of station $i$, are represented in a coordinate system fixed relative to the center of the earth (a). For an event $\vec{m}$, the signed amplitude of the oscillation expected in station $i$ is $\vec{d_i} \cdot \vec{m}$. Given the arrangement of magnetometers and their sensitive axes, a basis for the set of possible combinations of measured amplitudes is given by three $N_{st}$ dimensional vectors: $\vec{X}$, $\vec{Y}$, and $\vec{Z}$, where $\vec{X}$ (b) is the set of relative amplitudes obtained if the soliton star is moving directly along the x axis in the Earth's frame of reference, $\vec{Y}$ (c) is the vector of expected measurements for propagation along the y axis, and $\vec{Z}$ (d) for propagation along the z axis.

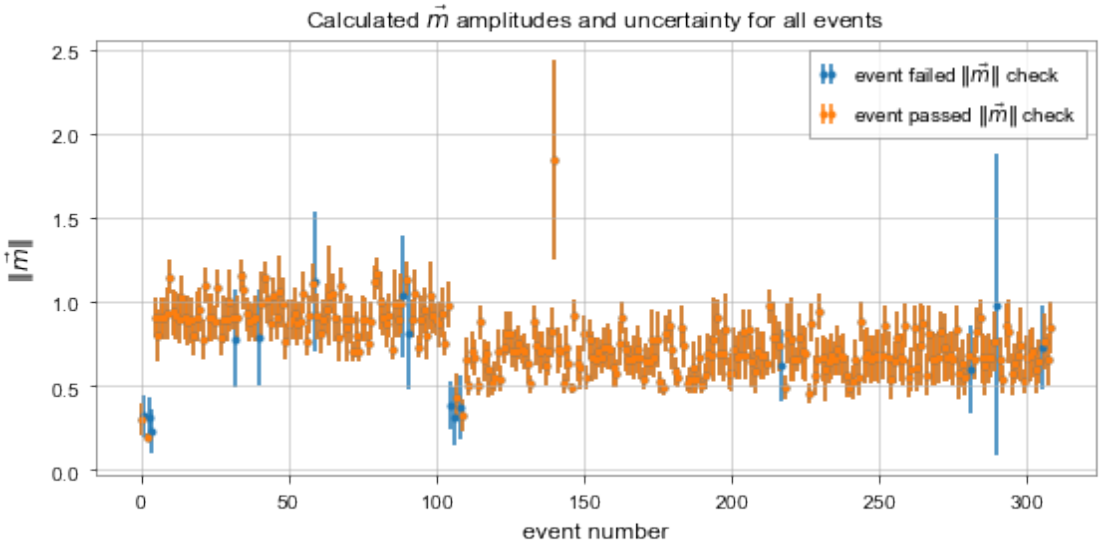


Figure 2.5: Scatter plot depicting the $m$ values calculated from equation 2.41 and associated uncertainty $\delta m$ from equation 2.42. The data are passed down from the excess power check with $\epsilon^*$ set for a false positive probability of $2.7 \times 10^{-7}$. All detected events are due (either directly or indirectly due to spectral leakage into adjacent tiles) to an injected event with $\vec{m} = (0, 0, 1)$. The data are clearly distinguishable into two collections. The injected signal was at 30.5 Hz, so the first $\sim 100$ events come from the case where the 30 Hz and 31 Hz tiles were summed together, and the last $\sim 200$ events are from the case where they were summed with the 29 Hz and 31 Hz tiles, respectively. The points highlighted in blue failed the $m$ threshold check at $m^* = 3$.


Figure 2.6: Scatter plot of the same events as shown in figure 2.5, plotted with $p$ value on the y axis. Points colored blue were rejected by the m check with $m^* = 3$ , and points colored orange were rejected by the $p$ check at $p^* = 0.01$

.

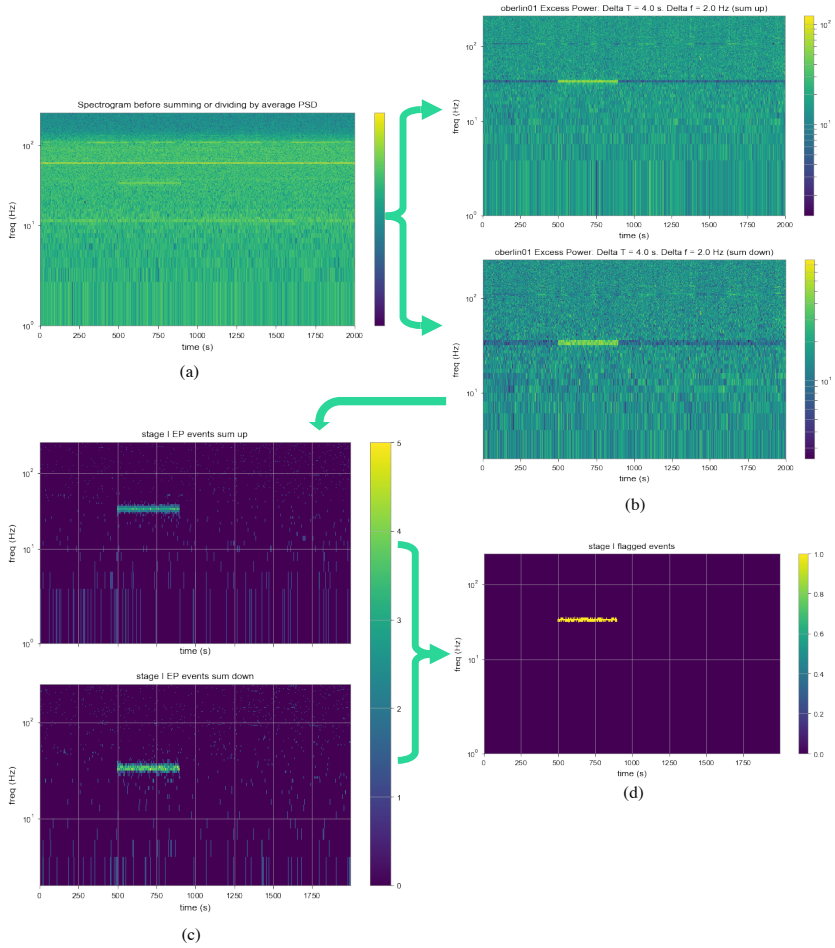


Figure 2.7: Example demonstrating the first loop of the excess power process, using 2000 s of real time-shifted data. A signal was injected with $\vec{m} = (0, 0, 1)$. The initial 1 s by 1 Hz tiles were summed to create a spectrogram with volume determined by $\Delta T = 4$ s, and $\Delta F = 2$ Hz. Figure (a) shows the raw spectrogram for station oberlin01, with time tiles summed but not frequency tiles. The 30.5 Hz signal is barely visible, and the 60 Hz line frequency is prominent. The two figures in (b) represent the "up" and "down" spectrograms normalized by the average PSD. Applying a threshold of $\epsilon^* = 38.1$ and summing together the results from all 9 active stations results in figure (c). Tiles in figure (c) with values $\geq 3$ are flagged, and expanded to fix the original 1 Hz frequency channels, forming the binary mask in figure *d*, in which many of the tiles within the 400 s signal are successfully flagged, but some still remain undetected. These events are passed onto the second excess power iteration in figure 2.8.

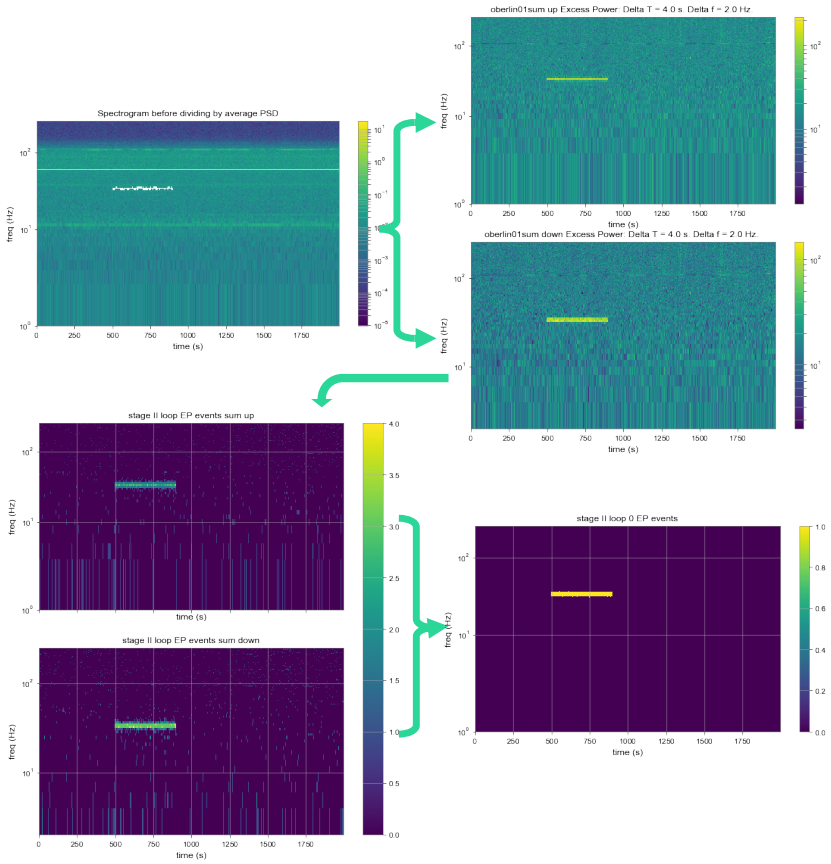


Figure 2.8: This example demonstrates the second loop of the excess power process, using 2000 s of real time-shifted data. The figure on the top left again depicts the the raw spectrogram for station oberlin01, with time tiles summed but not frequency tiles. However the tiles flagged by the first loop shown in figure 2.7 are masked (represented by white "missing" tiles) and not included in the calculation of the average PSD. After dividing by this average and summing 2 frequency tiles, the sum "up" and sum "down" normalized spectrograms on the upper right clearly show the injected signal.Applying a threshold of $\epsilon^* = 38.1$ and summing together the results from all 9 active stations results in the figure on the bottom left. After the three-fold coincidence check, the resulting flagged events on the bottom right now encompass the entirety of the injected signal. These events are passed onto the third loop excess power iteration in figure 2.9.

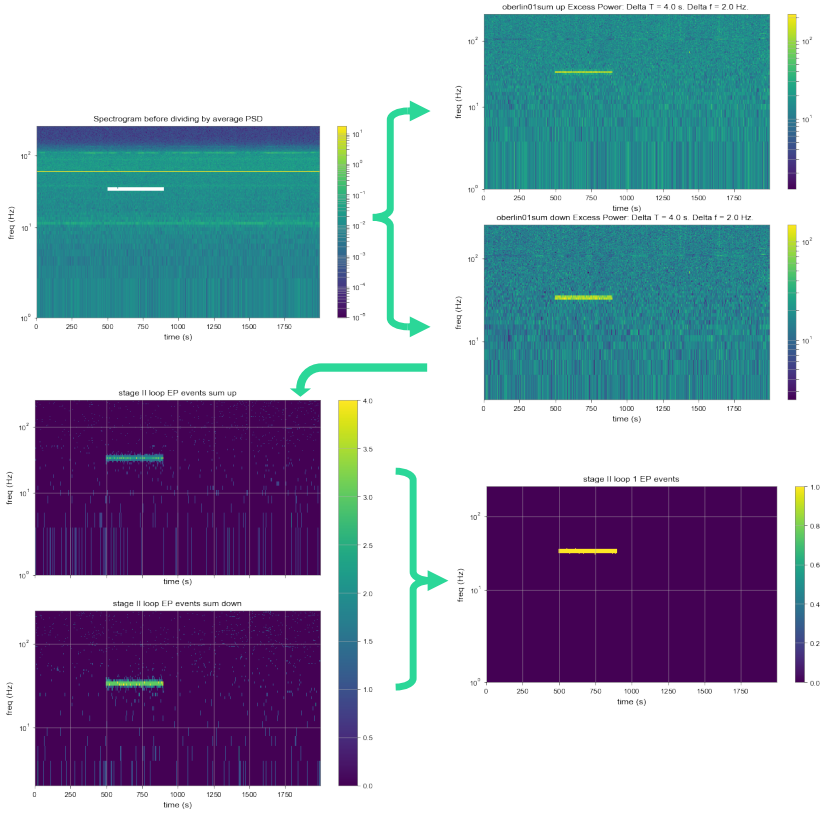


Figure 2.9: This example demonstrates the second loop of the excess power process. The figure on the top left again depicts the the raw spectrogram for station oberlin01, with time tiles summed but not frequency tiles. The tiles flagged by the second loop in figure 2.8 are masked and not included in the calculation of the average PSD. After dividing by this average and summing 2 frequency tiles, the sum "up" and sum "down" normalized spectrograms on the upper right once again clearly show the expected signal. Applying a threshold of $\epsilon^* = 38.1$ and summing the results from all 9 active stations results in the figure on the bottom left. Applying the three-fold coincidence check results in an identical set of flagged events as the previous check, indicating that the iteration should end. The resulting flagged events are used to calculate the subtracted spectrogram and uncertainties, which are passed onto the second stage.

# Chapter 3

# Characterizing the analysis method

In this section, we focus on a particular section of data, and prepare to analyze this time period by determining experimentally the relationship between the three thresholds (excess power threshold, $m$ threshold, and $p$ thresholds) and the false positive and false negative probabilities. We expect that the relationships derived in this section should hold for any analyzed time period, however we do not test this assumption here. For these preliminary tests, the retrieved magnetic field data are time shifted to prevent the possibility of real events. When the magnetic field data are imported from the GNOME server, the start time for each station is 500 s later than the start time for the previous station. So when we conduct preliminary characterization analysis (a false positive or false negative check) on 2000 s worth of data starting at $T_0$, the 2000 s from the first station starts at $T_0$, the 2000 s from the second station starts at $T_0 + 500$ s, the 2000 s from the third station starts at $T_0 + 1000$ s, and so on.

The earliest time series data analyzed are from March 13, 2020 starting 11:30 am, UTC. This particular time was chosen because 9 stations are active for approximately 14000 subsequent seconds, so when time shifting the data, we can analyze 5 consecutive 2000 s windows with consistent false positive and false negative probabilities. These 9 active stations are oberlin01, hayward01, krakow01, lewisburg01, berkeley01, berkeley02, daejeon01, moxa01, and losangeles01. The (non-time shifted) time series for each station for the entire 14000 s time are shown in figure 3.1. Not visible in this image are gaps in the time series representing missing or "insane" data (data flagged as invalid). These data are masked for the entire analysis duration and do not effect any of the calculations. If an event is detected in a time frequency tile where one station is insane, that station is still masked in the event vector and the consistency check (for that particular event) proceeds as if the "insane" station was not present.

In these preliminary tests, there are three experimentally controlled values that can be adjusted to achieve the desired false positive and false negative probabilities (not including the "lower threshold", which will be considered in future analyses). These parameters are the excess power threshold $\epsilon^*$ from the first stage coincidence check, the $m$ a $p$ thresholds from the consistency check, and the time-frequency volume $V$.

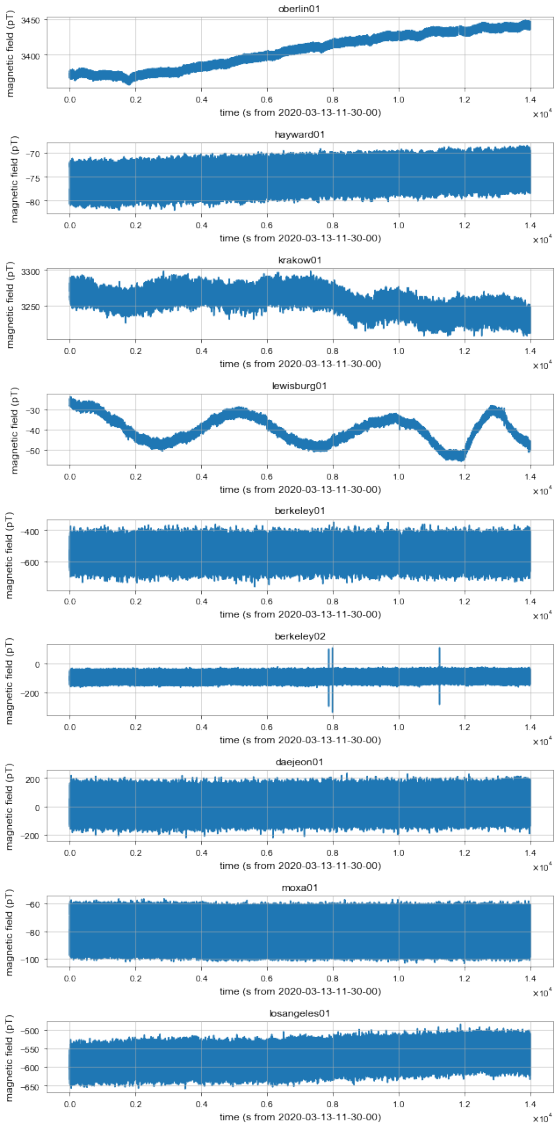


Figure 3.1: Time series of 14000 s of magnetic field data from the nine stations active on March 13, 2020 starting 11:30 am, UTC.

The following tests only varied the time frequency volume from 2 (2 frequency channels summed, no time segments summed) to 64 (2 frequency channels summed, 32 time segments summed) in powers of 2. However, 32 time tiles is by no means a desirable upper limit, and in principle the tile width can increase up to 1/3 of the total time, assuming that in the uncertainty calculations, $\sigma_{PSD}$ is calculated based on the standard deviation the un-summed tiles divided by $\sqrt{N_t}$. Further analysis is underway to push the tile sums further to determine the point at which increasing the tile time width no longer offers an advantage in detecting longer signals (i.e. at what point does the signal duration take up "too much" of the total time series and no longer behaves like a transient event in the excess power search).

## 3.1 False positive analysis

The goal of the false positive analysis is to experimentally observe the relationships between the actual probability that any tile with no signal (from time shifted data) will be accepted as a real event by the analysis and the four primary controlled quantities: the excess power threshold, the $m$ threshold, the $p$ threshold, and the time frequency volume $V$. Once these relationship are clear, we can use them. To set the threshold values so that there is a probability of $2.7 \times 10^{-}7$ ($5\sigma$) that a false positive event is flagged by the entire analysis process.

**Stage 1 only**

Of the three thresholds, the one with by far the greatest impact on the false positive rate is the first excess power threshold, which plays the largest role in reducing the number of candidate tiles. The expected false positive statistics of the excess power coincidence check are also thoroughly understood for Gaussian data, providing a clear theoretical expectation to which we can easily compare the experimental results.

To best replicate the conditions of the theoretical probability $P(\epsilon > \epsilon^*)$, we begin by eliminating the repeating loop in which tiles flagged by the previous coincidence check are excluded from the calculation of the average PSD in the following check. The results of this analysis are displayed in figures 2.2 and 3.2, which are different views of the same plot of $P(\epsilon > \epsilon^*)$ vs $\epsilon^*$, while summing time tiles in powers of 2 from 1 to 32 tiles. All nine stations were included, and the three-fold coincidence was performed after applying the excess power cutoff.

As seen in figure 2.2 and 3.2, for a single loop of 2000 seconds, the false positive probabilities for simulated Gaussian noise (unfilled circles) appear to follow the theoretical curve. This curve is calculated by integrating the $\chi^2$ distribution in equation 2.18 then applying the binomial sum (equation 2.27) to include the contribution of the 9 station 3-fold coincidence check. For time shifted real data during the chosen 2000 second window, the false positive rate agrees well for low threshold values (probability $> 0.1$) but deviates from the Gaussian and theoretical examples at high threshold values. This is due to the non-Gaussianity of the real data from the detectors; a relative abundance of high excess power outliers in the real station data results in values that average over time to represent an increase in false positive probability. The size of the data set limits the experimentally determined false positive rate. In particular, if we consider five 2000 second windows worth of data, the minimum possible non-zero experimental probability is

$$\begin{aligned} \frac{1}{N_{tiles\ total}} &= \frac{1}{n_T(n_F^{(u)} + n_F^{(d)})} \\ &= \frac{1}{\frac{2000}{\Delta T}\frac{(255+254)}{\Delta F}} \qquad \text{for total time 2000 s and Nyquist frequency 256 Hz} \\ &= \frac{V}{1018000} \end{aligned}$$

where $n_T$ is the total number of tiles along the time axis and $n_F^{(u)}$ and $n_F^{(d)}$, are the total number of tiles along the "up" and "down" frequency axis, respectively. This value is $2 \times 10^{-6}$ for $V = 2$, the smallest

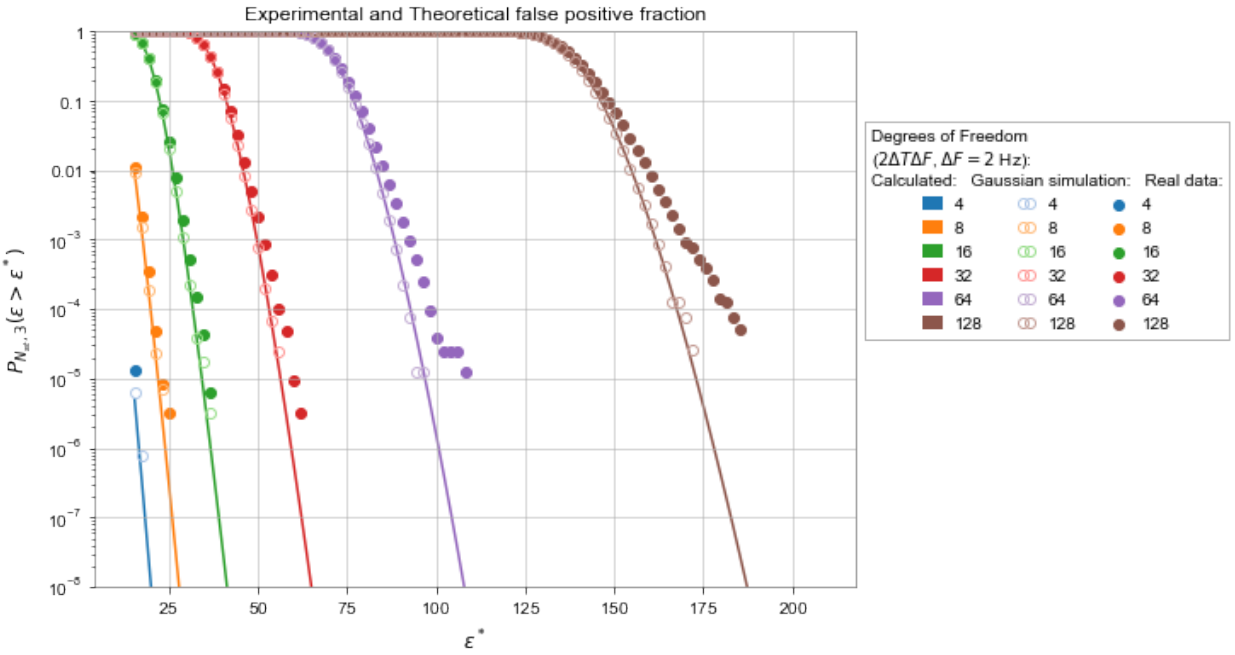


Figure 3.2: Theoretical and experimental false positive rates v.s. excess power threshold for a single iteration of the excess power coincidence search. The solid curves are the theoretical values from combining equations 2.20 and 2.27, assuming, with 9 active stations and a 3-fold coincidence. The solid points are the average fractional false positives for 5 consecutive 2000 s windows of real time-shifted data. The unfilled circles are the average fractional false positives for 5 consecutive 2000 s windows of simulated Gaussian data with mean 0 and standard deviation 1, emulating the ideal case from which equation 2.20 was derived. his graph is the same figure as 3.4, plotted on a log scale to highlight discrepancy between the real and Gaussian data at low probabilities.

time frequency volume tested. This means that with maximum of 5 2000 s windows, we must extend the experimental curve to determine the value of $\epsilon^*$ that gives $5\sigma$ value $p = 2.7 \times 10^{-7}$ for this particular data set. Including the entire excess power coincidence check produces the results shown in figure 3.4. Note that there is a "jump" in the false positive rate at $P \sim 0.9$. This can be understood as a result of the correction method for flagging entire rows of tiles. When an entire row of frequency tiles is flagged (which is likely to occur at excess power values threshold $\epsilon^* < 2\Delta T\Delta F$), then the entire row is unmasked and the average calculation proceeds in exactly the same way as if no tiles had been flagged, so that for that particular row, the analysis is no different than for the first loop. As the excess power threshold increases (so that $P(\epsilon > \epsilon^*) \sim> 0.9$), the probability of this effect occurring at some point in the loops for a given frequency channel decreases, however a significant fraction of the tiles are still below the threshold, and so the masking has an effect on the excess power probability. By masking tiles, we are eliminating some fraction of the upper tail of the distribution in the average. By dividing by this average, the resulting distribution

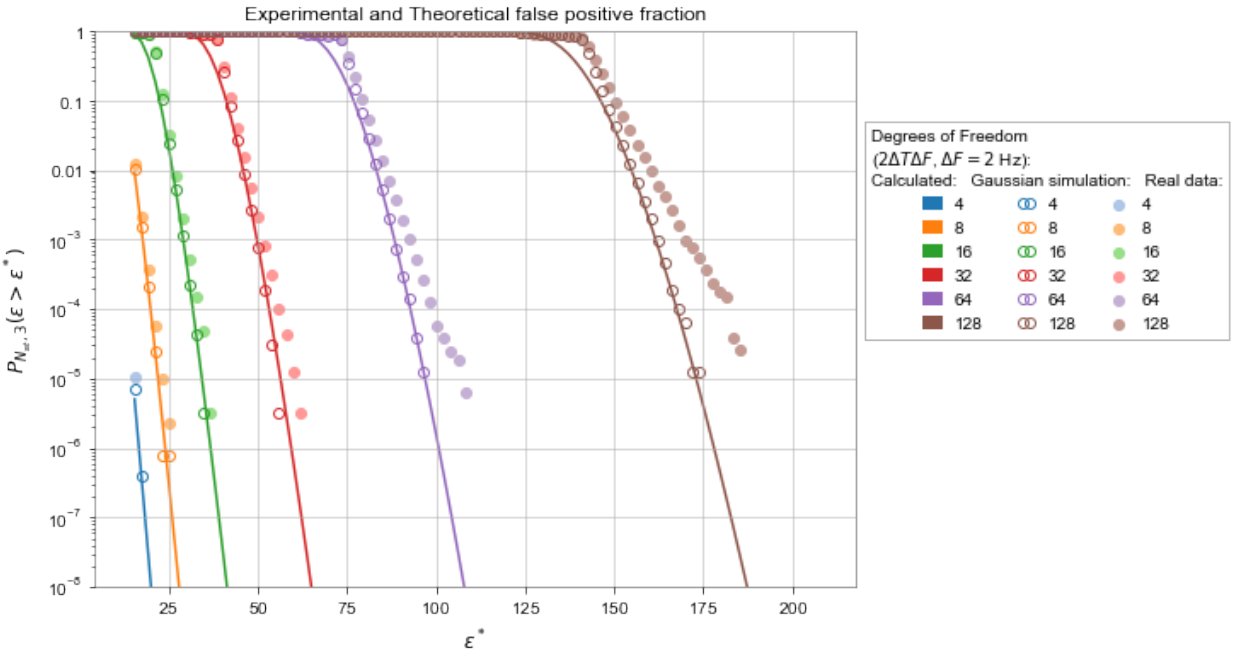


Figure 3.3: Theoretical and experimental false positive rates v.s. excess power threshold for the full excess power coincidence search. The solid curves are the theoretical values from combining equations 2.20 and 2.27, assuming 9 active stations and a 3-fold coincidence. The solid points are the average fractional false positives for 5 consecutive 2000 s windows of real time-shifted data. The unfilled circles are the average fractional false positives for 5 consecutive 2000 s windows of simulated Gaussian data with mean 0 and standard deviation 1, emulating the ideal case from which equation 2.20 was derived. This graph is the same figure as 3.4, plotted on a log scale to highlight the discrepancy between the real and Gaussian data at low probabilities.

no longer has mean $2\Delta T\Delta F$, and so represents a $\chi^2$ distribution with mean $\frac{2V}{\langle \epsilon < \epsilon^* \rangle}$. The lower the value of $\epsilon^*$, the larger the fraction of tiles that are eliminated, and the farther from $2V$ this value is shifted, leading to the visible "jump" in figures 3.3 and 3.4. This effect diminishes as the probability approaches 0 (i.e. the threshold $\epsilon^* >> 2V$), and for Gaussian data, the experimental and theoretical curves begin to overlap again as fewer tiles are flagged and the mean of the fraction of the distribution with $\epsilon < \epsilon^*$ approaches $2V$. Note that the effect is more pronounced for larger degrees of freedom, this is because each spectrogram has fewer tiles, so flagging a single tile has a larger effect on the average calculation (for the same excess power trigger, summing more time tiles causes that value to impact a larger fraction of the time series). While it is important to be aware of this effect, because the actual analysis process will operate in the regime of the third case (the limit of very low false positive probability), the effects of the repeating loop on the false positive probability can be ignored. Further tests are necessary to determine whether this is still the case if the initial value of $\epsilon^*$ is lowered for the initial iterations, until the set of newly flagged events is no longer

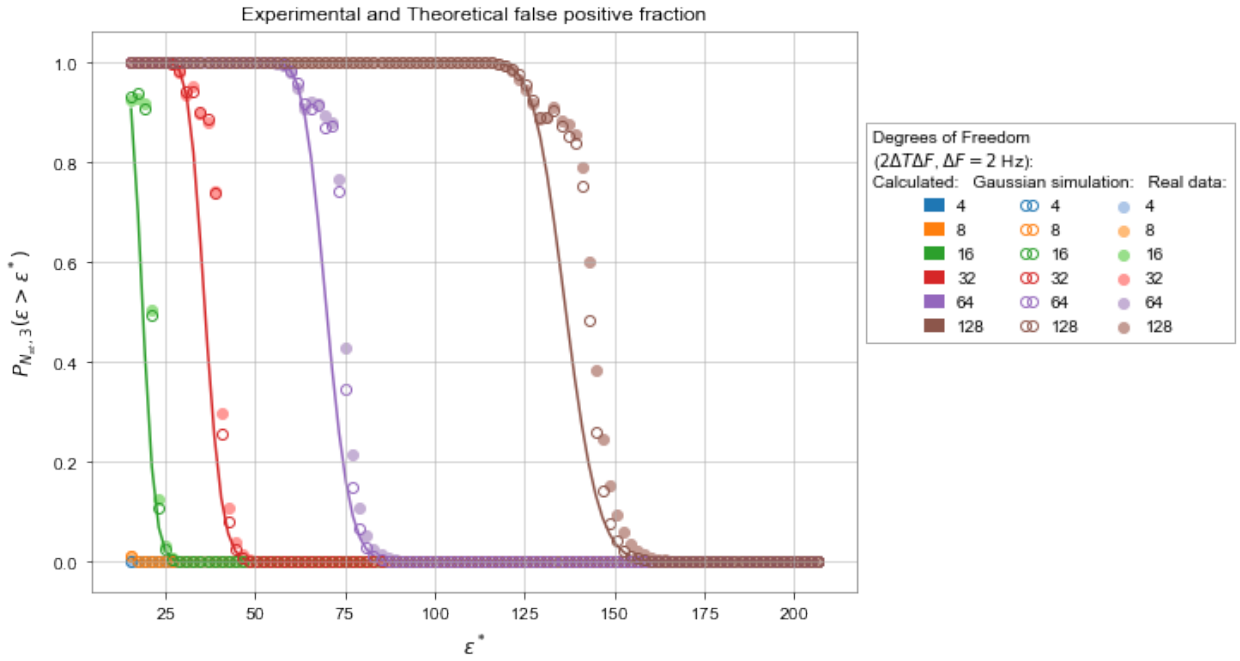


Figure 3.4: Theoretical and experimental false positive rates v.s. excess power threshold for the full excess power coincidence search. The solid curves are the theoretical values from combining equations 2.20 and 2.27, assuming 9 active stations and a 3-fold coincidence. The solid points are the average fractional false positives for 5 consecutive 2000 s windows of real time-shifted data. The unfilled circles are the average fractional false positives for 5 consecutive 2000 s windows of simulated Gaussian data with mean 0 and standard deviation 1, emulating the ideal case from which equation 2.20 was derived. This graph is the same figure as 3.4, plotted on a linear scale to highlight the "jump" that occurs at $P \sim 0.9$.

changing compared to the previous set of events, at which point the the threshold is increased to the full desired value. This is the intended route for the final analysis process represented in the figure 2.3 flow chart, but has not been implemented for the current analyses.

### 3.1.1 Stages 1 and 2

Since "events" consistent with detector noise are mathematically considered consistent, there is not a straightforward method of quantifying the effect of the $p$ value threshold on the false positive probability. The vectors rejected by the $p$ threshold must be inconsistent with both true events and the zero vector in the presence of Gaussian noise, whcih makes the development of a theoretical model difficult and impractical. We will focus instead, then, on the theoretical probability of accepting a false positive (due to Gaussian noise) in the second stage is primarily determined by the $m$ threshold, a value specifying the number of standard deviations from 0 the magnitude of the "best fit" event vector $\|\vec{m}\|$ must in order to consider the event inconsistent with noise. In the case of Gaussian data, the relationship between false positive probability and

the $m$ threshold $m^*$ is simply the the probability associated with the number of standard deviations. Due to constraints on time, the false positive analysis of for the second stage has not yet been fully completed, however we propose several tests which could verify the above relationship and provide other insights.

Although we cannot quantify and compare the relationship between false positive probability and the $p$ threshold for Gaussian noise, this threshold still has important consequences for the false positive rate in real data. The reason for this is that the real data are not Gaussian. The time series shown in figure 3.1 demonstrate the presence of the type of non-stationary noise that could emulate a transient signal in the excess power search, particularly the sharp spikes present in station "berkeley02". While the 3-fold coincidence requirement helps eliminate some of these events, the consistency serves as an additional method for vetoing this type of false positive that is not consistent with persistent noise, but nevertheless do not represent true events.

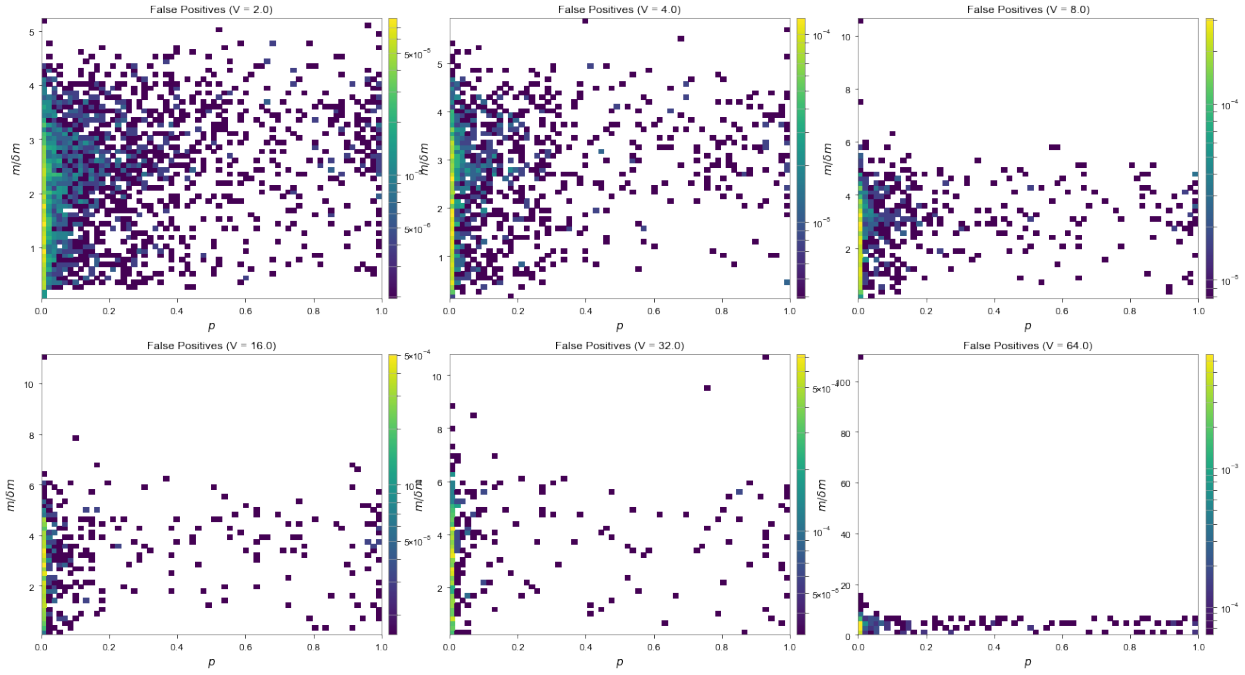


Figure 3.5: Heatmaps depicting fractional false positives passed from the first stage analysis with $\epsilon^* = 0.001$ ($\sim 3\sigma$) for each time frequency volume. The values are binned according $p$ value along the x axis and $m/\delta m$ along the y axis, for 5 consecutive 2000 s windows of real data. Note that the y axis scale differs for each figure, as does the range displayed on the colorbar. As expected, the majority of the values are clustered to the left, at low $p$ values.

To understand how the $m$ threshold and $p$ threshold together impact the false positive rates, consider the 2-dimensional histograms displayed in figure 3.5. With the value of $\epsilon^*$ set at the $5\sigma$ level, the resulting false positive event values were run through the $\chi^2$ and $m$ calculations. The color map is plotted on a log scale, so any $(p, m/\delta m)$ bins containing exactly 0 events are masked as white, and any colored tiles contain

a non-zero number of false positives within the pictured range. With $\epsilon^*$ set to a value with a a Gaussian false probability of 0.001 ($\sim 3\sigma$), a pattern starts to emerge in which the majority of values are clustered at the right (low consistency), and a "band" of false positive tiles connects the low consistency and high consistency sides of the figure. In the future, repeating this analysis at many different values of $\epsilon^*$ will allow more specific patterns to be revealed and enable us to draw experimental conclusions about how the $m^*$ and $p^*$ values relate to the false positive probability.

## 3.2 False negative analysis

Once we have set the values of $\epsilon^*$ and $m^*$ to achieve the desired false positive rate, we can conduct a false negative analysis to determine how these threshold values effect the sensitivity. This is done by injection simulated oscillatory bursts into the data and recording the fraction of tiles expected to contain a signal that were actually detected by the analysis program, as a function of injected signal amplitude. This analysis enables us to convert the threshold values set by the false positive analysis into a probability of detection for a signal of a amplitude. This quantitative description of the sensitivity is an important step in the process of converting the full analysis results into meaningful soliton star constraints.

### 3.2.1 Stage 1 only

Once again, the effect of the initial excess power threshold is worth examining separately. The results of the first stage not only have the largest impact on the overall sensitivity, but also have important impacts on the quality of the second stage consistency check. False negatives from this first analysis stage are not masked during the calculation of the final subtracted average or the uncertainty calculation, which inaccurately inflating the standard deviation estimation and preventing both the $m$ and $p$ checks from functioning as expected. It is thus particularly important to understand and control the false negative rate of the initial stage.

The experimental data collected in each true detection test (figures 3.6 through 3.7) represent the average number of tiles flagged out of tiles expected to contain an injected signal for 5 2000 s tests. For each test, a random frequency was chosen using numpy's random float generator, which selects from a "uniform continuous" distribution [22]. The frequencies are chosen from the range 10 Hz to 100 Hz. Unless otherwise specified, the signal duration is similarly chosen from between 100 and 600 seconds. A random start time is also chosen between the first data point and 10 seconds before the last time at which the signal can begin without truncating the duration. This signal is injected into the magnetic field time series at each station as a sine function oscillating at the desired frequency, and the first stage of the analysis proceeds as described in chapter 2 section 2.1 unless otherwise indicated.

When the threshold $\epsilon^*$ is fixed, the probability of detecting an injected signal with the signal component of the excess power statistic $A^2$ is given by a plugging the probability from equation 2.25 into the binomial

sum in equation 2.27 for $N_{st}$ active stations and an $N_c$ fold coincidence. Figure 3.6 shows depicts the

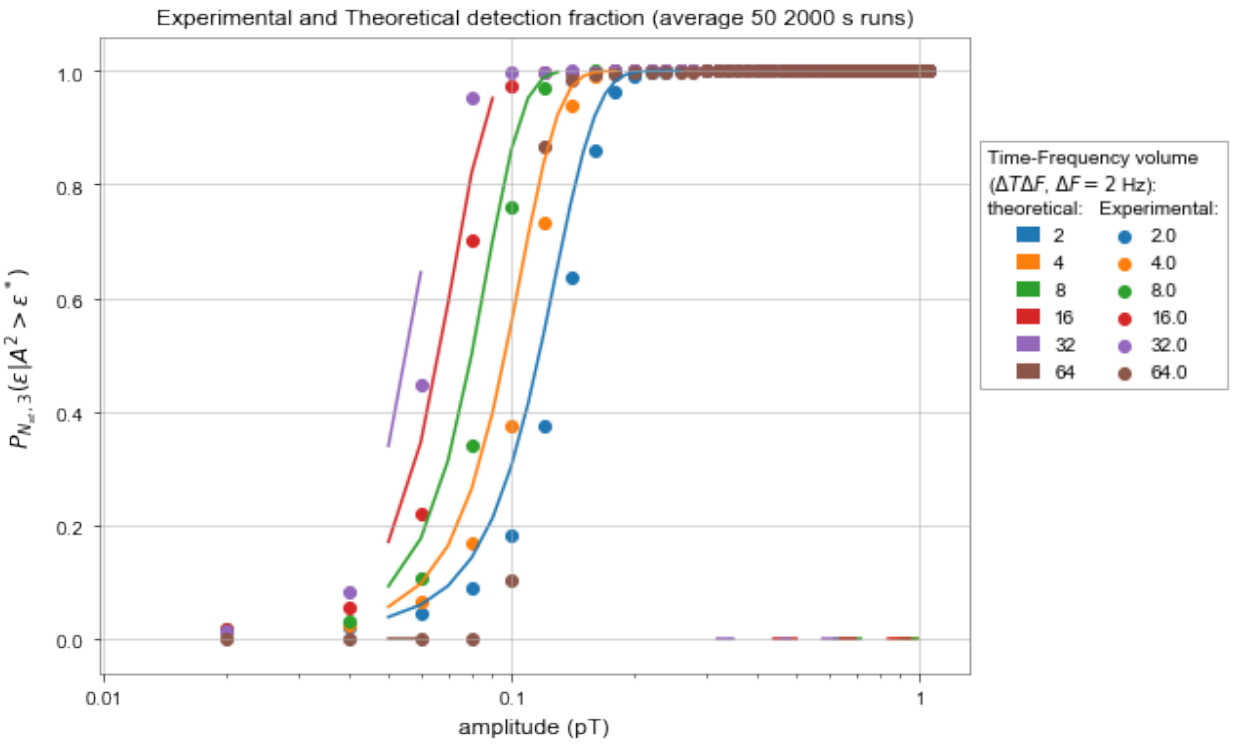


Figure 3.6: Theoretical and experimental true detection rate $P(\epsilon|A^2 > \epsilon^*)$ vs. injected signal amplitude for the excess power coincidence search, with $\epsilon^* = 0.001$. The solid curves are the theoretical values from combining equations the probability distribution 2.25 with the binomial sum in equation 2.27 with 9 active stations and a 3-fold coincidence. The points are the average fractional false positives for 5 2000 s windows of simulated Gaussian data with mean 0 and standard deviation 1, emulating the ideal case from which equation 2.25 was derived. Note that the theoretical curves drop suddenly to zero at large amplitudes, this is simply a numerical integration error; at large amplitudes, the theoretical detection probability should continue to approach 1.0.

theoretical "true detection" probability (1 - the false negative probability) for Gaussian data. The injected signal magnitude $C$ is related to the variable $A^2$ in equation 2.25 through

$$A^2 = \frac{2N_F N_T C^2}{\langle |h|^2 \rangle} = \frac{N_F N_T b C^2}{\frac{b}{2}\langle |h|^2 \rangle} \tag{3.1}$$

$$= \frac{2N_F N_T \alpha^2}{H} \tag{3.2}$$

where $C = |c_{kj}|$ the rms amplitude as defined in the previous sections, $\alpha = C\sqrt{b}$ is the amplitude of the injected signal (where $b = \frac{4}{N^2}$), and $H$ is the average noise PSD. Compared to the false positive analysis, the true detection graph does not agree nearly as closely. This is likely an indication that implementing the the multiple loop strategy with the same threshold value still does not completely replicate the conditions

assumed in generating the theoretical curve.

In order to accurately verify whether the experimental true detection probability (for both Gaussian and real data) agrees with the theoretical curve, we propose to repeat this initial process with some additional simplifications to the stage one coincidence check. First, in order to ensure that $H$ is truly calculated based on only the noise component, we can calculate the average PSD for each station before dividing by the average PSD, to ensure that the noise component did not include any part of the signal. Second, in order to simplify the interpretation of the injected amplitude, the signals can be injected into each station at a fixed amplitude, not from a projected direction. After projecting the events onto the sensitive axis of each detector, the actual measured amplitude is attenuated to an effective amplitude proportional to the projection of the event direction onto the sensitive axis of the detector.

In figure 3.7, the process is repeated with real data, including and the full loop analysis is completed.

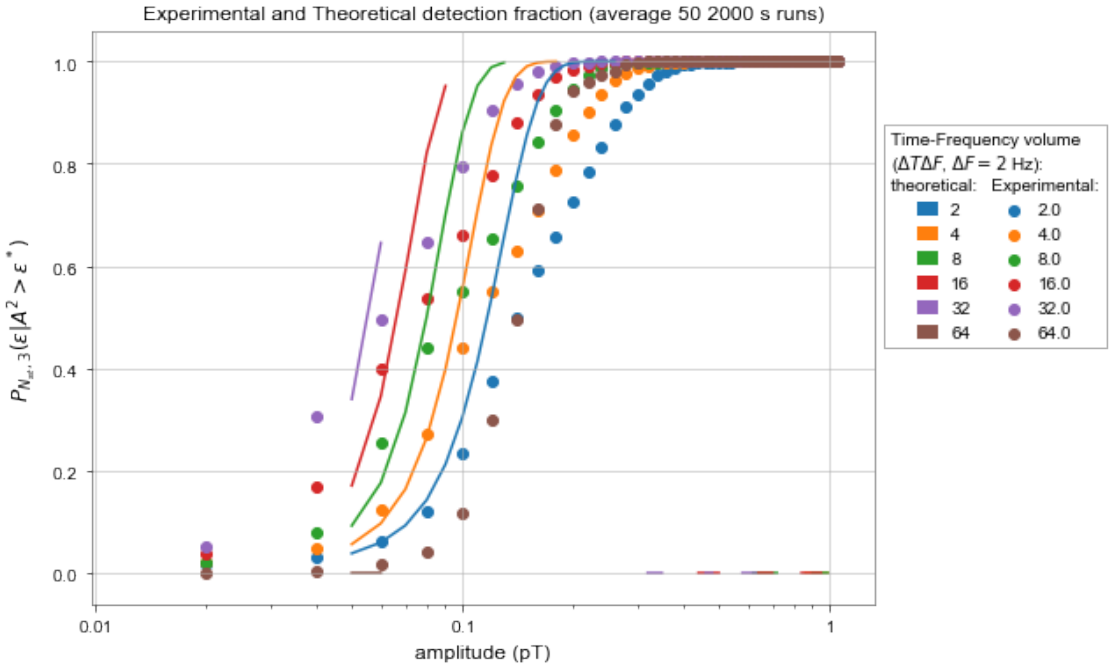


Figure 3.7: Theoretical and experimental true detection rate $P(\epsilon|A^2 > \epsilon^*)$ vs. injected signal amplitude for the excess power coincidence search, with $\epsilon^* = 0.001$. The solid curves are the theoretical values from combining equations the probability distribution 2.25 with the binomial sum in equation 2.27 with 9 active stations and a 3-fold coincidence. The solid points are the average fractional false positives for 5 consecutive 2000 s windows of real time shifted data. Note that the theoretical curves drop suddenly to zero at large amplitudes, this is simply a numerical integration error; at large amplitudes, the theoretical detection probability should continue to approach 1.0.

With the inclusion of these additional parameters, it is no longer helpful to compare these values to the theoretical predictions, but the specific true detection probability for a particular time span can be extracted

from the experimental curves. Note however that the sample of points used to calculate this probability value varies with degrees of freedom and signal duration, since the more tiles present contain a signal, the greater the probability that within the frame, at least one tile will contain a signal, which would account for why the curve with $V = 64$ in figures 3.6 and 3.7 breaks the observed pattern and demonstrates a higher than expected false positive rate. The simplest proposed solution is to change the injected duration for each value of the degrees of freedom to ensure that the same number of tiles contain a signal every time. However, this approach has several important drawbacks. First and most importantly, holding the number of tiles constant severely limits the range of possible durations. Injecting a signal that flags 100 1 s tiles is simple, and the 100 s resulting signal is adequately small compared to the total 2000 s window. Injecting a signal that flags 100 32 s tiles is actually impossible in a 2000 s time series. The second drawback is that this method requires the basic spectrogram to be regenerated for every value of the degrees of freedom in addition to every amplitude and test, drastically increasing the time needed to run the full false negative test.

An alternative method of interpreting these same data which could provide a different form of insight requires adjusting the initial definition of what constitutes a successful detection. Rather than considering the probability of true detection for any given tile, we can redefine "detection" based on the entire signal rather than each individual tile. If the successfully detected fraction of tiles containing an injected signal is greater than some new threshold (for example 20%) then the event is considered to be "detected successfully", regardless of the number of tiles missed. Since for each trial, the same injected signal is analysed for all six degrees of freedom, this will serve to remove the discrepancy in the number of tiles summed per degrees of freedom (which may be the source of the lower than expected true detection rate for the curve with $V = 64$), as well as assist in reducing the effects of inaccurately estimating the expected number of flagged tiles. While several tricks were implemented in the program to help prevent this second issue, there will always be an issue if the randomly selected start time does not fall at the start of a tile, in which case the amplitude is attenuated in that tile, resulting in a "effective amplitude" that is the average of the the signal and noise PSD and is in general much smaller than the initially injected value. "Off by one errors" may also occur when estimating start and end tiles, depending on the number of tiles summed and when the signal is injected. This process has already been used successfully in other characterization circumstances, can be implemented in future analysis.

Another important proposed check involves repeating the process once again, this time using real time-shifted data from the same time set as the false positive analysis in section 3.1 while fixing the signal duration at 400 s (representing passage through the center of an axion star with radius $10R_E$). This would allows the expected improvement in sensitivity with number of tiles summed to be experimentally verified.

**Stages 1 and 2**

As is the case for the false positive analysis, a full analysis has not yet been implemented to characterize the relationship between the $p$ value threshold $p^*$ and the false negative rate. The $p$ value can be considered

as the probability that the closest constant signal $\vec{\mu}$ could have been obtained from the measured vector $\vec{a}$ given noise of standard deviation $\vec{\delta a}$, thus the successful detection probability is given by $1 - p^*$. One simple method of testing this assumption is to inject many events, and record the fraction of successful detections as a function of $p^*$, for multiple amplitude and tile durations. One could then additionally choose a specific desired value representing the the minimum desired true detection rate (e.g. 90%) and plot the minimum amplitude that can be detected at this rate as a function of $p^*$ for each degree of freedom (which also allows the information gained from this test to be consolidated in a single graph). If the goal is simply to verify the relationship between $p*$ on its own and the false negative rate, then the effects of the first stage analysis could be bypassed entirely, and list of events, the subtracted spectrogram, and and the associated uncertainty values can be calculated from the list of expected events, ensuring that the true detection probability is due entirely to the effects of the $p$ threshold. Alternatively, a fixed $\epsilon^*$ and $m^*$ value could also be chosen based on the false positive probability, and the graphs can be used to assess the sensitivity of the entire analysis.

## 3.3 Proposal for application on real data

Once the method is fully characterized, we will have gained two crucial pieces of information. First, from the false positive tests, we know where to set the three thresholds $\epsilon^*$, $m^*$ and $p^*$ so that the overall probability of a false positive event being detected by the analysis is $2.7 \times 10^{-7}$, or $5\sigma$. Once these values have been set, the false negative tests provide a measure of the sensitivity level, the smallest detectable signal amplitude that can be successfully extracted with the analysis program at the chosen thresholds given the particular combination of active stations. This value can be calculated as average over all possible directions, or considered as a value that varies depending on the direction. All of these values will differ according to the time period under analysis and the degrees of freedom, so some of the tests demonstrated above will need to be repeated for time shifted data around the desired time period in order to get a reliable estimation of the noise characteristics. Alternatively, future efforts could focus on determining to what extent the thresholds from one section of data generalize to other time periods.

Once this information is obtained, the next step of the process is to break up the desired time range into 2000 s windows, and run the non-time-shifted data from each window through the excess power process. Repeating the process for multiple time tile widths maximizes the sensitivity for different durations. If no events are detected, then the next step is to convert the calculated sensitivities into constraints on for soliton star parameters. An example of how this process might be accomplished is shown in Ref. [3], where the relationship between the expected energy shift, the ALP mass, and the coupling constant are constructed for the Hamiltonians in equations 1.3 through 1.5, which are in turn used to estimate the sensitive parameter space accessible to GNOME. To accomplish this, in addition to choosing a particular model, concrete assumptions were made about factors including the particular soliton star model, the minimum detectable radius, and the rate of event encounters. These factors in turn were used to estimate other

necessary values and ultimately set constraints. Taking this approach, more work is needed to determine what further assumptions we will impose to interpret the analysis results.

# Chapter 4

# Conclusion

We have presented here a description of the excess power analysis developed specifically to search the data provided by the magnetometers of GNOME for transient oscillatory bursts. The analysis process checks that such signals are detected in at least 3 different stations within the same time and frequency bin, that the amplitude of the signals detected in each station are consistent with the known sensitive axis direction of the stations, and that the event is not consistent with noise. For Gaussian data, the first and most important of these checks can be quantitatively represented as a function of the threshold value $\epsilon^*$, which is then used to calculate the false positive probability and successful detection probability as a function of signal amplitude for various degrees of freedom. Given a specific time range of real data, we can adjust these probabilities based on time shifting the data during that period (so that we are confident any detected signal is a false positive) and either running the analysis to search for false positives or injecting simulated signals to investigate successful detections. This process was demonstrated in detail for the first stage of the analysis (the consistency check), and methods for characterizing the two combined stages have been proposed that should be implemented to fully characterize the analysis process.

## 4.1 Suggestions for Future Work

The next step for the immediate future is to implement and interpret the results of the proposed tests from chapter 3. It will also likely be useful to develop more sophisticated methods of estimating the false positive and false negative rates of the non-Gaussian data in order to avoid the need to constantly re-implement these preliminary tests, which are specific to the given time and set of magnetometers. A test that would likely still be useful to repeat for each specific set of active detectors is a generate map of sensitivity based on the propagation vector direction, which could be used to develop directionally dependent sensitivity values which could potentially enable us to place more accurate constraints. Once we understand how the program works, we need to understand how we can interpret our results to constrain soliton star parameters, based on calculations similar to the examples from Ref. [3]. When this process is prepared, the analysis can finally

be run on real data, dividing the full length of analyzed data into 2000 second (potentially overlapping) windows and running each window through the program. The longer the total time, the more effectively we can estimate the the constraints on the rate of encounters, which helps place additional constraints on other factors such as the size and number density of the soliton stars in the galaxy, as well as the mass and coupling constant for the individual ALPs.

Once this particular version of the axion star analysis is fully characterized and ready to use, additional modifications can be implemented to apply nearly the same method to different categories of signals. For example, by looking for correlations within a certain range of adjacent time bins rather than coincidences would allow for the potential detection of events with significant delay times between stations.

# Bibliography


[1] Editorial "A few holes to fill," *Nature Phys* **4**, 257 (2008).

[2] G. Bertone, D. Hooper, and J. Silk, "Particle Dark Matter: Evidence, Candidates and Constraints," *Phys. Reports* **405**, 279 (2005).

[3] D. F. Jackson Kimball, D. Budker, J. Eby, M. Pospelov, S. Pustelny, T. Scholtes, Y.V. Stadnik, A. Weis, and A. Wickenbrock, "Searching for axion stars and Q-balls with a terrestrial magnetometer network," *Phys. Rev. D* **97**, 043002 (2018).

[4] M. Pospelov, S. Pustelny, M. P. Ledbetter, D. F. Jackson Kimball, W. Gawlik, and D. Budker, "Detecting Domain Walls of Axionlike Models Using Terrestrial Experiments," Phys. Rev. Lett. 110, 021803 (2013).

[5] D. F. Jackson Kimball, "Nuclear spin content and constraints on exotic spin-dependent couplings," *New J. Phys.* **17**, 073008 (2015).

[6] S. Pustelny, D. F. Jackson Kimball, C. Pankow, M. P. Ledbetter, P. Wlodarczyk, P. Wcislo, M. Pospelov, J. R. Smith, J. Read, W. Gawlik, and D. Budker, "The Global Network of Optical Magnetometers for Exotic physics (GNOME): A novel scheme to search for physics beyond the Standard Model," *Ann. Phys.* **525**, 659 (2013).

[7] D. Budker and D.F. Jackson Kimball eds., *Optical Magnetomety*, Cambridge University Press, New York (2013).

[8] S. Afach, D. Budker, G. DeCamp, V. Dumont, Z. D. Grujic, H. Guo, D. F. Jackson Kimball, T. W. Kornack, V. Lebedev, W. Li, H. Masia-Roig, S. Nix, M. Padniuk, C. A. Palm, C. Pankow, A. Penaflor, X. Peng, S. Pustelny, T. Scholtes, J. A. Smiga, J. E. Stalnaker, A. Weis, A. Wickenbrock, and D. Wurm, "Characterization of the Global Network of Optical Magnetometers to search for Exotic Physics (GNOME)," *Physics of the Dark Universe* **22**, 162 (2018).

[9] D. F. Jackson Kimball, J. Dudley, Y. Li, and S. Thulasi, "Magnetic shielding and exotic spin-dependent interactions," *Phys. Rev. D* **94**, 082005 (2016).

[10] D. Budker, P. W. Graham, M. Ledbetter, S. Rajendran, A. O. Sushkov, "Proposal for a Cosmic Axion Spin Precession Experiment (CASPEr)," *Phys. Rev. X* 4, 021030 (2014).

[11] S. Park, "Construction of a Single Beam SERF Magnetometer using Potassium Atoms for GNOME," undergraduate thesis, Oberlin College (2019).

[12] H. Masia-Roig, J. A. Smiga, D. Budker, V. Dumont, Z. Grujic, D. Kim, D. F. Jackson Kimball, V. Lebedev, M. Monroy, S. Pustelny, T. Scholtes, P. C. Segura, Y. K. Semertzidis, Y. C. Shin, J. E. Stalnaker, I. Sulai, A. Weis, A. Wickenbrock "Analysis method for detecting topological defect dark matter with a global magnetometer network," *Physics of the Dark Universe* **28**, 100494 (2020).

[13] C. Daily, C. Bradley, D. F. Jackson Kimball, I. Sulai, S. Pustelny, A. Wickenbrock, and A. Derevianko, "Quantum sensor networks as exotic field telescopes for multi-messenger astronomy," arXiv:2002.04352 [astro-ph.IM], (2020).

[14] A. Kusenko and P. J. Steinhardt, "Q-Ball Candidates for Self-Interacting Dark Matter," *Phys. Rev. Lett.* **87**, 14 (2001).

[15] W.J.M. de Kruijf, J.L. Kloosterman, "On the average chord length in reactor physics," *Annals of Nuclear Energy* **30**, 549 (2003).

[16] W. G. Anderson, P. R. Brady, J. D. E. Creighton, and E. E. Flanagan, "An excess power statistic for detection of burst sources of gravitational radiation" *Phys. Rev. D* **63**, 042003 (2001).

[17] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes: the Art of Scientific Computing Third Edition*, Cambridge University Press, Cambridge UK, (2007).

[18] P. Hoyng, "An Error Analysis of Power Spectra," *Astronomy and Astrophysics* **47**, 449 (1976).

[19] E. J. Groth, "Probability Distributions Related to Power Spectra," *Ap. J., Suppl. Ser.* **29**, 285 (1975).

[20] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions*, Dover, New York, (1972).

[21] J. A. Smiga, "Magnitude consistency," private communication (2019).

[22] "Numpy and Scipy Documentation," *scipy.org*, https://docs.scipy.org/doc/ (accessed April 2020).