University of Nevada, Reno

Department of Physics

Applying the matched-filter technique to the search for dark matter transients with networks of quantum sensors

Author:

Guglielmo Panelli

Supervisor:

Prof. Andrei Derevianko

A thesis submitted for the degree of

Bachelor of Science in Physics

November 9, 2020

Preface

This thesis is submitted for the degree of Bachelor of Science at the University of Nevada, Reno. The research described in this manuscript was conducted under the supervision of Professor Andrei Derevianko in the Department of Physics, University of Nevada, Reno, between December 2017 and November 2019.

This manuscript contains original work except where acknowledgements and references are made.

Work from this thesis has been submitted for publication and is available in pre-print:

- G. Panelli, B.M. Roberts, and A. Derevianko, *Applying the matched-filter technique to the search for dark matter transients with networks of quantum sensors* (2019). arXiv:1908.03320

Elements of this thesis have been presented at the following conferences and symposia:

- *Applying the matched-filter technique to search for dark matter with networks of precision measurement devices.* Talk at the APS Far West Section Meeting, Stanford, CA; November 01, 2019

- *Applying the matched-filter technique to search for dark matter with networks of precision measurement devices.* Poster at the University of Nevada, Reno Undergraduate Research Symposium, Reno, NV; May 06, 2019

- *Applying the matched-filter technique to search for dark matter with networks of precision measurement devices.* Poster at the APS April Meeting, Denver, CO; April 12, 2019

- *The Ubiquitous Matched-Filter and its Application to Dark Matter Detection.* Talk at the University of Nevada, Reno - Department of Physics, Reno, NV; April 01, 2019

- *Matched Filter Technique in the Detection of Dark Matter with the GPS Atomic Clocks.* Poster at the University of Nevada, Reno College of Science Research Symposium, Reno, NV; November 26, 2018

- *Data Analysis in the Search for Dark Matter with GPS.* Poster at the University of Nevada, Reno Summer Undergraduate Research Symposium, Reno, NV; August 09, 2018

- *Data Analysis in the Search for Dark Matter with a Network of Precision Measurement Devices.* Talk at the University of Nevada, Reno McNair Scholars Research Symposium, Reno, NV; August 03, 2018

- *Data Analysis in the Search for Dark Matter with a Network of Precision Measurement Tools.* Talk at the McNair Scholars National Conference, Los Angeles, California; July 26, 2018

Guglielmo

November 9, 2020

Acknowledgements

Abstract

There are several networks of precision quantum sensors in existence, including networks of atomic clocks, magnetometers, and gravitational wave detectors. These networks can be re-purposed for searches of exotic physics, such as direct dark matter searches. Here we explore a detection strategy for macroscopic dark matter objects with such networks using the matched-filter technique. Such "clumpy" dark matter objects would register as transients sweeping through the network at galactic velocities. As a specific example, we consider a network of atomic clocks aboard the Global Positioning System (GPS) satellites. We apply the matched-filter technique to simulated GPS atomic clock data and study its utility and performance. The analysis and the developed methodology have a wide applicability to other networks of quantum sensors.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Astrophysical observations on the galactic scale indicate that dark matter (DM) constitutes 85% of all matter in the universe, leaving only 15% to ordinary matter [5]. These galactic scale observations (e.g., rotation curves, cosmic microwave background, gravitational lensing) characterize solely the gravitational interaction of dark matter and ordinary matter, leaving the microscopic composition of DM and its non-gravitational interactions with standard model particles and fields a mystery. Among the current DM searches, weakly interacting massive particles (WIMPs) have been the primary target with no success to date, thereby partially motivating alternative DM candidates [5]. One such alternative is ultralight fields where the DM candidate may take the form of a macroscopic object coherent over large scales. Such objects would exert gentle minute perturbations detectable by quantum sensors [6]. Quantum sensors are typically well protected from environmental perturbations which makes them uniquely sensitive to new physics. Examples include atomic clocks, magnetometers, atom interferometers, and microwave and optical cavities. The particular hypothesized form of coupling of DM fields to baryonic matter determines the type of the sensor to be used in the DM search [7].

Here we focus exclusively on ultralight fields. In general, one may consider free non-interacting and self-interacting fields. We are interested in models with self-interaction that can form macroscopic dark-matter objects, such as topological defects or Q-balls [8, 9, 10, 11]. Interactions of such DM constituents with standard model (SM) particles and fields can induce variations in fundamental constants of nature. Such variations may be detected by observing atomic frequencies in atomic clocks [12, 13, 14]. As the DM constituents sweep through a clock, the variation would register as a transient perturbation.

Geographically-distributed networks can resolve the velocities of the transients and provide a powerful vetoing of the potential events as the sweep velocity must be consistent with standard halo model priors. These ideas form the basis of dark matter searches with distributed networks [15, 16, 17, 2, 4].

Our GPS.DM group focuses on searching for DM-induced transients in GPS atomic clock data [16, 17]. Here the advantage is the public availability of nearly two decades of archival data enabling relatively inexpensive data mining. A dark matter signature would consist of a correlated propagation of atomic clock perturbations through the GPS constellation at galactic ($\sim 300\,\mathrm{km/s}$) velocities. Previously, our GPS.DM collaboration performed an analysis of the archival GPS data in search for domain walls (a particular type of topological defect) [17]. Although no DM signatures were found, new limits were placed on certain DM couplings to atoms that were several orders of magnitude more stringent than prior astrophysical limits. The original search [17] focused on finding large DM signals well above the instrument noise. In Ref. [2], we have shown that the application of more sophisticated Bayesian search techniques can extend the discovery reach by several orders of magnitude both in terms of sensitivity and size/geometry of the DM objects. Here we study the performance of the matched-filter technique (MFT) as an alternative frequentist search method. In addition, we develop analytic results for idealized network of white-noise sensors with cross-node correlation.

The MFT is a relatively ubiquitous technique, utilized, for example, by the Laser Interferometer Gravitational Wave Observatory (LIGO) in gravitational wave detection (see, e.g. [18]). It is also used in a variety of applications, such as astrophysics [19], geophysics [20], and searches for exotic physics [21]. Of a special interest to us is the performance of the MFT for large networks, as the GPS.DM sensor array may include over a hundred instruments if we take into account all GNSS constellations and terrestrial clocks. A technical complication in applying the MFT to these networks is that the noise is correlated between spatially separated nodes of the network. Understanding the consequences of this cross-node correlation is of importance to our analysis.

The structure of this paper is as follows. Sec. 2 reviews the MFT for data analysis and previous applications. Sec. 3 formalizes requirements for a network detector of clumpy DM. Sec. 4 reviews the relevant theory and how the GPS network can be re-purposed for DM searches of this sort. A summary of processing the GPS data is provided in

Sec. 5. In Sec. 6, we describe our methodology, including the formulation of our detection statistic and the search algorithm. The detection threshold, detection probability and parameter estimation capabilities are provided in Secs. 8, 9, and 10, respectively. Lastly, the projected discovery reach for this method is provided in Sec. 11 and Sec 12 draws conclusions. The paper also contains appendices where we discuss network covariance matrix and its inversion, inverse transform sampling, and present supporting derivations. Since the intended audience includes both atomic and particle physics communities, we restore $\hbar$ and c in the formulas in favor of using natural or atomic units. We also use the rationalized Heaviside-Lorentz units for electromagnetism.

Chapter 2

The matched-filter technique

The matched-filter technique is often used to search for hidden signals within data streams in cases where the signal's "shape" is known but the signal's strength is not. In this case, the general shape can be compared to the data stream to search for an underlying correlation that is not immediately evident to the un-aided eye. The matched-filter itself is the best estimate of the unknown signal strength by employing an optimal filter. In the most general sense, an optimal filter is a particular combination of data that optimizes a quantity deemed to be significant, usually relating to signal detection within data sets [18]. Usual quantities of interest include the detection probability for a given signal strength and the signal-to-noise ratio (SNR), though many other application-specific statistics can be devised.

Since the MFT requires a predefined signal shape, this approach cannot be used for un-modeled signals. When a hypothesized signal signature is able to be modeled, there often exist many candidate shapes along with the unknown signal strength. Thus, one forms a collection of signal shape templates that approximately spans the range of possible shapes. One could think of the MFT as a technique that maximizes an overlap between the templates and the data stream. This maximization is done with the help of a matched-filter statistic, such as a SNR. However, it is not usually the value of the SNR alone that determines the level of overlap, but rather the value of the SNR compared to a threshold. Additionally, one of the most promising aspects of the MFT is the ability to align the shape of the detected signal with the template that produced an SNR above the detection threshold, leading to immediate signal parameter estimation.

The efficacy of the MFT depends on how well one can distinguish a weak signal

from intrinsic device noise. Thus, a network of devices can offer a better sensitivity and higher confidence in the event of a positive detection since all network sensors would experience a signal from the same event.[1] If a hypothesized signal shape can be modeled for a network of devices, the MFT becomes a powerful tool for weak signal detection and signal parameter estimation.

2.1 Examples of the MFT in practice

Perhaps the most well-known application of the MFT comes from the gravitational wave detection by LIGO (see e.g., [22]). A detailed outline of search techniques and matched-filtering is provided in Ref. [18] and guided much of our development of the method discussed in this paper. However, LIGO's use of the MFT involved a small network of devices that exhibit uncorrelated noise. The black hole merger gravitational wave detection from 2015, for instance, used only two spatially separated interferometers in the waveform template matching analysis [22]. Another previous application of the MFT used 15-20 station from the International Deployment of Accelerometers (IDA), where geophysicists were been able to identify previously undetected global seismic events in archival IDA data [20]. The method has also found use in galaxy cluster identification [19] and in the search for neutrino-less beta decay [21].

All of these examples consider networks of devices far smaller than that available ($\sim$ 100) to our DM search. Furthermore, an essential feature of our GPS.DM network search is a cross-node correlated noise (due to a reference clock common to all the nodes, see Sec. 5), which, to the best of our knowledge, has not been addressed in the literature. A related theoretical development for ultralight (non self-interacting) DM fields is presented in Ref. [23]. There a quantum sensor network was considered for detecting DM waves and a SNR statistic was developed in the frequency domain. By contrast, here we focus on network detection of dark matter transients and develop a SNR statistic in the time domain.

---

[1]This is technically only true for the domain-wall DM object as domain walls sweep through all the network nodes. Other DM objects such as monopoles may only affect a fraction of the nodes.

Chapter 3

Network desiderata

Here we are devising a strategy for detecting macroscopic DM objects that sweep through a distributed network of N_{D} sensors. We use the words "sensor" and "node" interchangeably. In particular, a single geographical location may host several instruments, yet each individual sensor is referred to as a distinct node for our purposes. We will assume that DM objects interact (non-gravitationally) with the instruments and the interaction only occurs when the bulk of the DM object overlaps with a sensor.

There are several criteria for such networks:

(i) The network should be sufficiently dense so that the finite-size DM object can overlap with at least several geographically-distinct nodes.

(ii) The network volume should be sufficiently large to increase the rate of encounters with DM objects.

(iii) As per the standard halo model (SHM) [24] the DM objects sweep through the network at galactic velocities ($v_g \sim 300\,\mathrm{km/s}$), the sampling rate should be sufficiently high to enable tracking the propagation of the DM object through the network. The tracking enables reconstructing the geometry of the encounter.

(iv) Although not necessary, it is desired that the encounters of DM objects with the network are sufficiently rare, so that only a single DM object interacts with the network at any given time.

While most of these requirements are apparent, criterion (iii) deserves further discussion. For example, while a setup [3] of two co-located clocks with a shared optical

cavity can be considered as a rudimentary two-node network, such network does not satisfy criterion (iii) since even if both clocks were to register a DM signal, there would be no galactic velocity/direction signature to support the signal's DM origin. Thus, such low-sampling rate networks can only be used to constrain couplings to the DM sector.

Chapter 4

Clumpy dark matter models

Stable macroscopic objects may form from ultralight DM fields due to their self-interaction in the dark sector [25, 26, 27, 28, 29, 30]. Topological defects (TDs) are an example of such macroscopic "clumpy" DM objects, though they can also contribute to dark energy depending on their cosmological fluid equation of state [2]. Monopoles (0D), strings (1D), and domain walls (2D) are all examples of TDs of various dimensionalities. Other examples of macroscopic DM candidates include Q-balls [9, 10, 11], solitons [31, 32], and axion stars [33, 34]. A special case of clumpy DM are DM "blobs" [35] – particle-like DM objects sourcing long-range Yukawa-type interactions with the SM sector.

For concreteness, we focus on topological defects. Inside the defect, the amplitude of the DM field A and the average energy density of the defect is related by $\rho_{\mathrm{inside}} = A^2/(\hbar c\, d^2)$, where d is the width or spatial scope of the defect (we use the convention where the field has units of energy). The DM object width d is treated as a free observational parameter and, for TD models, may be linked to the mass of the DM field particles m_ϕ through the healing length which is on the order of the Compton wavelength $d = \hbar/(m_\phi c)$. Further, the local DM energy density ρ_{DM} may be linked to d and A by assuming that these objects saturate the local DM energy density,

$$A^2 = (\hbar c)\, \rho_{\mathrm{DM}} d^2 \frac{\mathcal{T}}{\tau_{\mathrm{avg}}}\,, \tag{4.1}$$

where $\tau_{\mathrm{avg}} \sim d/v_g$ is the average duration of crossing through a point-like instrument and $\mathcal{T}$ is the average time between subsequent encounters of the DM objects with the device [17].

15

As for the non-gravitational interactions, to be specific and consistent with our earlier work [2], we assume the quadratic scalar portal,

$$-\mathcal{L}_{\text{int}} = \left(\Gamma_f m_f c^2 \overline{\psi}_f \psi_f + \Gamma_\alpha \frac{F_{\mu\eta}^2}{4} + \dots\right)\phi\phi^*,\tag{4.2}$$

where m_f are the fermion masses, ϕ is the scalar DM field (measured in units of energy), Γ_X are coupling constants that quantify the DM interaction strength, and ψ_f and $F_{\mu\eta}$ are the SM fermion fields and the electromagnetic Faraday tensor, respectively. The SM fermions f in the above equation are summed over implicitly. Such interactions appear naturally for DM fields possessing either Z_2 or $U(1)$ intrinsic symmetries. The above Lagrangian leads to an effective redefinition of fundamental masses and coupling constants,

$$\alpha^{\text{eff}}(\boldsymbol{r}, t) = \left[1 + \Gamma_\alpha \left|\phi(\boldsymbol{r}, t)\right|^2\right]\alpha,\tag{4.3}$$

$$m_f^{\text{eff}}(\boldsymbol{r}, t) = \left[1 + \Gamma_f \left|\phi(\boldsymbol{r}, t)\right|^2\right]m_f,\tag{4.4}$$

where m_f are the fermion (electron/proton $m_{e/p}$ and light quark $m_q \equiv [m_u + m_d]/2$) masses and $\alpha \approx 1/137$ is the electromagnetic fine-structure constant. The coupling constants Γ have units of $[\text{Energy}]^{-2}$ and we also define the effective energy scales as $\Lambda_X \equiv 1/\sqrt{|\Gamma_X|}$ with $X = \alpha,\, m_{e/p},\, m_q$ to aid the comparison with previous literature.

The observable atomic frequency shift induced by the DM objects can be linked to the transient variation of fundamental constants, and thus the DM field parameter, from Eqs. (4.3) and (4.4). For a particular clock transition,

$$\frac{\delta\omega(\boldsymbol{r}, t)}{\omega_0} = \sum_X \kappa_X \Gamma_X |\phi(\boldsymbol{r}, t)|^2 \equiv \Gamma_{\text{eff}}\,|\phi(\boldsymbol{r}, t)|^2,\tag{4.5}$$

where ω_0 is the nominal clock frequency, X runs over relevant fundamental constants, and κ_X are dimensionless sensitivity coefficients. For convenience, we introduced the effective constant, $\Gamma_{\text{eff}} \equiv \sum_X \kappa_X \Gamma_X$, which depends on the specific clock.

The effective coupling constants for the GPS network microwave atomic clocks (Rb, Cs and H) read (using computations [36, 37], see [2] for details, and Ref. [38] for illucidating

16

the underlying logic)

$$\Gamma_{\text{eff}}(^{87}\text{Rb}) = 4.34\,\Gamma_\alpha - 0.069\,\Gamma_{m_q} + 2\,\Gamma_{m_e}, \tag{4.6}$$

$$\Gamma_{\text{eff}}(^{133}\text{Cs}) = 4.83\,\Gamma_\alpha - 0.048\,\Gamma_{m_q} + 2\,\Gamma_{m_e}, \tag{4.7}$$

$$\Gamma_{\text{eff}}(^{1}\text{H}) = 4\,\Gamma_\alpha - 0.150\,\Gamma_{m_q} + 2\,\Gamma_{m_e}. \tag{4.8}$$

Although linear combinations of the coupling constants differ for each type of clock in the GPS network, individual coupling constants Γ_X (or, equivalently, individual energy scales Λ_X) can be obtained by combining the results from different clock types. Laboratory optical Sr clock has provided the most stringent constrains on Λ_α for specific regions of the $(d, \mathcal{T})$ parameter space (see, e.g. [3]). More recently, new constraints have been placed on $\Lambda_\alpha, \Lambda_{m_e}$, and Λ_{m_q} by our GPS.DM collaboration [17] and on Λ_α by a global network of optical laboratory clocks [39]. These two papers reported null results for domain wall searches.

4.1   Thin domain walls

In this paper we will focus on a specific type of DM signal – "thin" domain walls. While retaining the main features of more complicated signals from other types of DM "clumps", this signal offers a sufficiently simple, analytically treatable signature. Domain wall-like signatures can appear naturally in the context of bubbles, i.e., domain walls closed on themselves [17]. Locally, one can neglect the bubble curvature as long as the bubble radius is much larger than the spatial extent of the sensor network. Another example are Q-balls that couple to SM fields through derivative couplings, $\phi\phi^* \to \partial_\mu\phi\partial^\mu\phi^*$ in Eq.(4.2). Q-balls are spherically symmetric objects with a nearly flat density profile in the bulk. Thus the dominant part of the interaction would occur at the Q-ball walls. Again one needs to require the radii of these objects to be much larger then the network size. Since bubbles and Q-balls are spherically symmetric, gravitationally interacting ensembles of these DM objects are a subject to the equation of state for pressureless cosmological fluid as required by the ΛCDM paradigm.

We distinguish between "thin" and "thick" walls based on the sampling rate, which is finite for any realistic device. If the interaction time with the device d/v_g is shorter

than the sampling interval τ_0, the exact arrival time of the DM clump is not resolved, and neither its shape. Thus the DM object is "thin" for observational purposes if its size $d \ll v_g \tau_0$. For domain walls, strictly speaking, the relevant velocity is its component normal to the wall, $v_\perp$.

For the GPS sampling interval of $\tau_0 = 30$ s, the above arguments translate into domain walls of thicknesses below the Earth size, $d \ll 300 \, \mathrm{km \, s^{-1}} \times 30 \, \mathrm{s} \approx 10^4 \, \mathrm{km}$. Any domain wall with a thickness larger than this value is characterized as "thick" and is discussed in more detail in Ref. [2].

The thin wall network signature is formalized in Sec. 6.2. For thin walls, the value of the effective coupling relates to the maximum DM-induced accumulated clock phase (time) signal $h = \delta\omega_{\mathrm{max}}/\omega_0 \times \tau$ by

$$h = A^2 \Gamma_{\mathrm{eff}} \tau \, , \tag{4.9}$$

$\tau = d/v_\perp$ being the interaction time between the wall moving at velocity v and an individual device. Again, for the wall to be "thin", we require τ to be less than the sampling time interval τ_0.

With the theoretical background established, we now review GPS data and characterize the utility of applying the matched-filter technique for DM search in network data streams. This includes establishing a signal-to-noise ratio test statistic and benchmarking the method via simulation.

Chapter 5

Overview of GPS data

A detailed description of modern GPS data acquisition and processing techniques and their application in precision geodesy can be found in Ref. [40]. Details relevant to DM searches with GPS constellation are given in Ref. [17]. Here, we briefly review the main aspects of GPS atomic clock data and introduce relevant concepts and terminology.

In our search, we analyze the GPS data generated by the Jet Propulsion Laboratory (JPL) [41]. This data consists of clock biases, the difference in clock phases (i.e., the operational "time" as counted by the clocks) between a given satellite clock and a fixed reference clock, and are sampled at $\tau_0 = 30\,\mathrm{s}$ intervals. The same reference clock is used for the entire network of satellite clocks on any given day. The data set also provides the satellite orbits, so we know the locations of the networks nodes (satellites). The JPL performs the initial GPS data processing [41]. In their processing, they do not limit clock bias behavior, meaning that real transient signals are not removed as outliers.

Each clock's bias data, denoted $d_j^{(0)}$, is non-stationary and is dominated by random walk process. Prior to our analysis, we "whitten" the data by performing the first-order differencing and define a new data stream

$$d_j^{(1)} = d_j^{(0)} - d_{j-1}^{(0)}. \tag{5.1}$$

This differencing procedure is sufficient for the Rb satellite clocks while a second-order differencing procedure $(d_j^{(2)} = d_j^{(0)} - 2d_{j-1}^{(0)} + d_{j-2}^{(0)})$ is often preferred for Cs clocks. We refer to the differenced data $d_j^{(1)}$ as the pseudo-frequency due to its proportionality to the discrete clock bias derivative. The units of pseudo-frequency $d_j^{(1)}$ are nanoseconds. As

19

shown in Ref. [2] the pseudo-frequency noise is dominated by the Gaussian white noise. To streamline notation, for the rest of the paper, $d_j \equiv d_j^{(1)}$. Such data standard deviation σ is related to the commonly used Allan deviation $\sigma_y(\tau_0)$ as $\sigma = \tau_0 \sigma_y(\tau_0)$.

An important aspect of the GPS time series data is that it consists of the individual clock noise and the noise of the reference clock. So, for each clock a, the noise component can be represented as

$$n_j^a = e_j^a - c_j \,, \tag{5.2}$$

where e_j is the individual clock noise and c_j is the contribution from the reference clock noise common to all data streams. Here and below the subscript enumerates data points (epochs) and the superscript enumerates sensors. While both sources of noise in pseudo-frequencies are dominated by the Gaussian white noise, in our simulations we will include realistic auto-correlation functions for the GPS clocks computed in [2].

5.1 Simulating GPS data

Characterizing the efficacy of the MFT is contingent on our ability to simulate the GPS atomic clock data. A detailed description of GPS data simulation along with a direct comparison to the GPS archival data is provided in Ref. [2]. The essence of the simulation method comes from utilizing the known power spectral densities for each clock (from JPL) to "color" pseudo-random white noise [42]. Moreover, we are able to simulate cross-clock correlation by adding an extra set of simulated white noise with standard deviation equal to that of a typical reference clock ($\sigma_\times \approx 0.006$ ns) to all of the simulated satellite clock data streams. This effectively acts as the common reference clock noise component to the satellite data in Eq. (5.2).

With the necessary background provided, we pivot to a description of our methodology in next section.

Chapter 6

Methodology

We wish to determine whether there is significant evidence that a thin wall DM signal is present in the GPS archival atomic clock data or not. This can be formalized by the following two-sided hypothesis test:

$$H_0 : h = 0, \quad \text{vs.} \quad H_1 : h \neq 0 \,,$$

where h represents the strength of the possible hidden DM signal. Note that the alternative hypothesis H_1 can be thought of a union of all possible alternatives H_h where h may be any non-zero real number. Furthermore, note that we allow the signal strength h to be both positive and negative, which is different than a typical signal strength search that treats h as an amplitude (and therefore non-negative). This is because the sign of the DM interaction coupling Γ is not known *a priori*. That is, we do not know if the DM-induced perturbations will cause the GPS atomic clocks to tick faster or slower.

To perform the aforementioned hypothesis test, we must formulate a detection statistic ρ, which in this case will be a signal-to-noise ratio and is formulated in the following section. Then, in order to claim detection, we must first establish a detection threshold ρ^* (provided in Sec 8) to compare to our observed statistic ρ.

The cases in which our search produces an observed SNR that exceeds the established threshold or not are treated differently. If our search results in a detection statistic larger than the threshold, we then wish to estimate the parameters relating to the DM interaction event. This is discussed in Sec 10. On the other hand, if our search does not result in a detection statistic indicating an event, we wish to place limits on the

signal strength h which translates into into limits on the DM coupling Γ via Eq. (4.9), see Sec 11.

6.1 Formulation of test statistic

Consider a candidate DM model M that would leave a coherent signal in network sensor data set D. The data set may or may not include a candidate model-prescribed signal $\boldsymbol{s}(\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is the specific set of parameters that define the signal signature, such as the DM object's velocity, orientation, arrival time and strength. If there is no signal within the data, then $\boldsymbol{s} = 0$. Taking a frequentist approach, all one needs is a likelihood function for the data set. This is given by the following Gaussian:

$$\mathcal{L}(D|M) = K \exp\left[-\frac{1}{2}\chi^2(\boldsymbol{s})\right],\tag{6.1}$$

where K is the normalization factor and

$$\chi^2(\boldsymbol{s}) = \sum_{ab}^{N_\mathrm{D}}\sum_{jl}^{J_W}\left(d_j^a - s_j^a\right)\left(E^{-1}\right)_{jl}^{ab}\left(d_l^b - s_l^b\right).\tag{6.2}$$

Here d_j^a is data for the ath clock at epoch j and $\boldsymbol{E}$ is the covariance matrix for the network. To streamline notation, we dropped the explicit reference to template parameters $\boldsymbol{\theta}$. The indices a, b run over N_D sensors (excluding the reference clock for GPS) and j, l range over the epochs (data points) in the observation time window of length J_W. In an equivalent vector notation,

$$\chi^2(\boldsymbol{s}) = (\boldsymbol{d} - \boldsymbol{s})^T \boldsymbol{E}^{-1}(\boldsymbol{d} - \boldsymbol{s}),\tag{6.3}$$

where $\boldsymbol{d}$ is the data stream and $\boldsymbol{s}$ is the signal stream.

The likelihood in Eq. (6.1) is a multivariate function of the signal parameters $\boldsymbol{\theta}$. An important aspect of our method is the signal linearity with respect to its strength h so that we can define a "unit" signal that is scaled by its strength

$$\boldsymbol{s}(\boldsymbol{\theta}) = h\overline{\boldsymbol{s}}(\boldsymbol{\theta}|_{h=1}).\tag{6.4}$$

This way, our likelihood becomes

$$\mathcal{L}(D|M) = K \exp\left[-\frac{1}{2}\Big((\boldsymbol{d} - h\overline{\boldsymbol{s}})^T \boldsymbol{E}^{-1} (\boldsymbol{d} - h\overline{\boldsymbol{s}}) \Big) \right]. \tag{6.5}$$

Since we do not know the shape or parameters of the unit signal $\overline{\boldsymbol{s}}$ (or if there exists a signal within the data at all), we span the space of all possible signals by forming a large repository of unit signal templates (a discussion of how we form the repository of templates is provided in Sec. 6.2). Suppose we form a unit signal template $\overline{\boldsymbol{s}}_i$ with a set of randomly (but strategically) chosen parameters. We then compare this template to the data stream via the likelihood function in Eq. (6.5). The template specific likelihood is then a function of only the signal strength h since each of the other signal parameters have been fixed to form $\overline{\boldsymbol{s}}_i$. In this case, the template-specific likelihood can be re-cast as a function of h alone

$$\mathcal{L}_i(D|M) \propto \exp\left[-\frac{1}{2}\Big(\frac{h - \hat{h}}{\sigma_h} \Big)^2 \right], \tag{6.6}$$

where

$$\hat{h} = \frac{\boldsymbol{d}^T \boldsymbol{E}^{-1} \overline{\boldsymbol{s}}_i}{\overline{\boldsymbol{s}}_i^T \boldsymbol{E}^{-1} \overline{\boldsymbol{s}}_i}, \tag{6.7}$$

$$\sigma_h = \frac{1}{\sqrt{\overline{\boldsymbol{s}}_i^T \boldsymbol{E}^{-1} \overline{\boldsymbol{s}}_i}}. \tag{6.8}$$

Here $\hat{h}$ is the signal strength that maximizes the template-specific likelihood and σ_h is the template-specific likelihood standard deviation. We quantify how well the signal template $\overline{\boldsymbol{s}}_i$ matches the signal within the data via a signal-to-noise ratio statistic defined as

$$\rho_i \equiv \frac{\hat{h}}{\sigma_h} = \frac{\boldsymbol{d}^T \boldsymbol{E}^{-1} \overline{\boldsymbol{s}}_i}{\sqrt{\overline{\boldsymbol{s}}_i^T \boldsymbol{E}^{-1} \overline{\boldsymbol{s}}_i}}. \tag{6.9}$$

We emphasize that our use of the SNR as a statistic rather than its square is to retain the dependence on the sign of the DM coupling. The SNR statistic depends on the inverse of the covariance matrix $\boldsymbol{E}$. Properties of the covariance matrix and its inversion techniques are discussed in Appendix A. Note that Eq. (6.9) is general in that it applies to any modelled signal (monopoles, strings, walls, etc.), though here we treat the case of thin walls only.

In general, due to the central limit theorem, we can assume that the intrinsic noise of the network sensors is Gaussian (the noise may be colored). This was the underlying assumption in writing the likelihood (6.1). Then the SNR statistic (6.9) is a linear combination of Gaussian random variables, and as such, the SNR is also a random Gaussian variable.

Now, recall that we wish to span the space of possible model-prescribed signals with a repository of unit signal templates. Then, given a repository of M randomly generated templates, we define our detection statistic ρ as the template-specific SNR with the largest magnitude

$$\rho = \rho_j \quad \text{such that} \quad |\rho_j| = \max\{|\rho_i|\}_{i=1}^M \,. \tag{6.10}$$

Choosing this as the detection statistic results in finding the signal template that maximizes the multivariate likelihood (6.1).

With our detection statistic defined, in the next sections we outline our method of unit template generation and provide an overview of our procedure for searching the GPS datastreams for DM events.

6.2 Template generation

Each signal template is determined by the DM model used (in this case, the thin domain wall) and the necessary parameters associated with the event: velocity (and incident direction), time of the event and thickness of the DM object. Since it would take an infinite number of model-prescribed signal templates to cover such a continuous parameter space, we strategically generate our finite repositories (template banks) of signals with a Monte-Carlo approach using prior distributions for individual parameters. When generating signals for a template bank, we employ importance sampling for each parameter according to these prior distributions in an effort to approximately span the continuous parameter space with a finite sample. This approach is formalized in appendix D.

We use the SHM to generate necessary parameter prior distributions. The velocity distribution for DM objects in the halo is quasi-Maxwellian and isotropic with a dispersion around $v \approx 300\,\text{km s}^{-1}$ [24]. In addition, there is a motion of the Solar system through the halo at galactic velocities of $v_g \approx 220\,\text{km s}^{-1}$. The resulting most probable incident direction of a DM object is along the path of the Sun's orbit in the galaxy, toward the

Cygnus constellation. This implies that over 90% of DM events would come from the forward facing hemisphere [17]. Further discussion of priors for event parameters such as domain wall width and event rate is provided in Ref. [2]. The event parameters (velocity, incident direction, etc.) determine in which order the nodes are swept, as well as how quick the sweep is. These characteristics distinguish the templates within the template bank.

After generating the template parameters, we form the signal templates. The thin domain wall is characterized by an interaction time with each device (clock) of less than the sampling time interval; thereby, the profile contains delta-functions of time. The DM-induced clock bias (phase) of a given clock a is proportional to an integral of the frequency shift (4.5). Further, the bias data stream is given with respect to a fixed reference clock R, which is also affected by the domain wall. We thus distinguish between maximum signals h^a and h^R as shown in Eq. (4.9). The signal in the differenced data stream $s_j^{a(1)}$ then reads (for the case when the wall encounters the clock a prior to the reference clock R)

$$
s_j^{a(1)} = \begin{cases} 0, & j < j_a \,, \\ h^a, & j = j_a \,, \\ 0, & j_a < j < j_R \,, \\ -h^R, & j = j_R \,, \\ 0, & j > j_R \,, \end{cases} \tag{6.11}
$$

where the time at epoch j is $j \times \tau_0$, a discrete time on the sampling grid, and j_a, j_R are the epochs in which the satellite clock and reference clock interact with the DM object, respectively. This template is shown graphically in Fig. 6.1.

In a homogeneous network of clocks, the values of h^a and h^R are the same, thereby allowing us to split a single h from the differenced signal to form the templates with unit spikes at j_a and j_R. This is also true for any network under the assumption that the Γ_{m_e} coupling dominates over other couplings (i.e., $|\Gamma_{m_e}| \gg |\Gamma_{\alpha,q}|$) since the Γ_{m_e} contribution in Eqs. (4.6 - 4.8) is the same for all clock types. Assuming that any other coupling dominates, we may still split a single h from the differenced signal but the unit templates will contain a unit spike at j_a and a spike of magnitude $\eta \equiv h^R/h^a = \Gamma_{\text{eff}}^R/\Gamma_{\text{eff}}^a$ and of opposite sign at j_R.
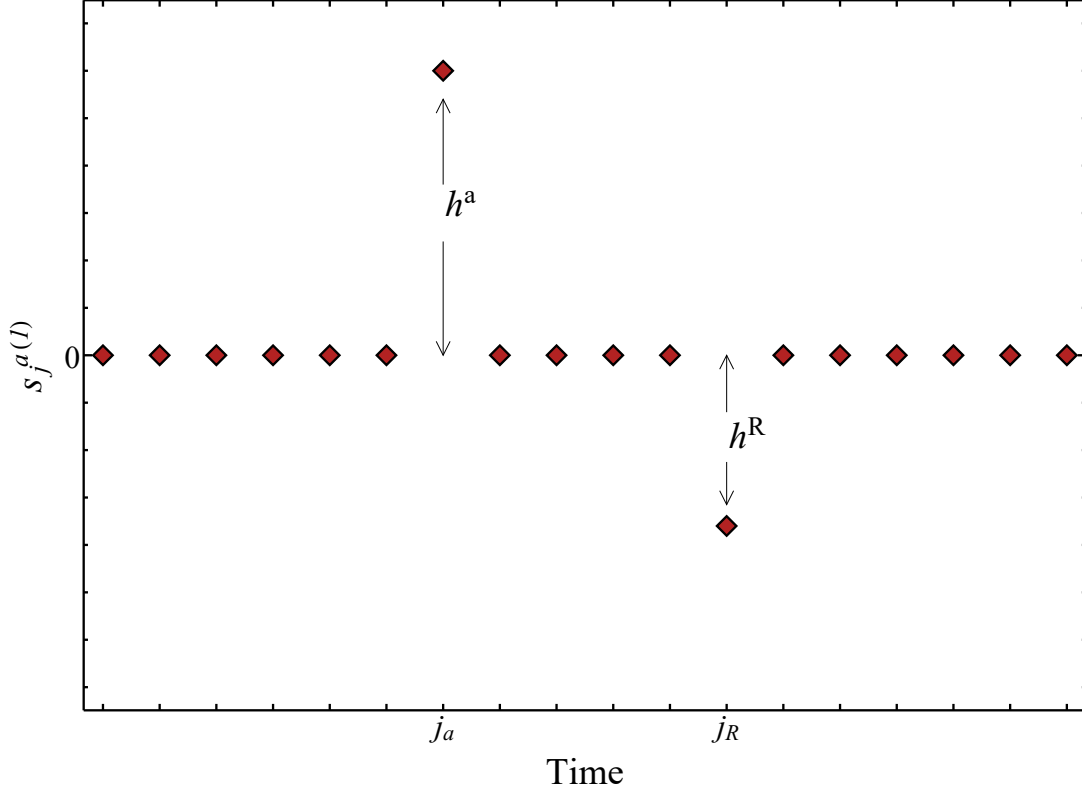
Figure 6.1: Time series for a differenced thin domain wall signal. Here clock a is affected by the DM object prior to the reference clock R. When the thin wall object interacts with clock a between epochs $j_a - 1$ and j_a, a spike of magnitude h^a is seen at epoch j_a. Then, as the thin wall object sweeps the reference clock between epochs $j_R - 1$ and j_R, a spike of opposite sign and magnitude h^R is seen at epoch j_R.

We would also like to highlight the importance of a well-spaced network for the MFT approach and template generation. Well-spaced meaning that few (if any) satellite clocks are affected by the DM wall within the same $30\,\mathrm{s}$ time period as the reference device (an approximation of the number of degenerate devices is provide in B). If the network nodes were not sufficiently spatially separated, the signal templates from Eq. (6.11) would collapse into "null" templates, where all elements of the signal stream are zero. This is due to the node devices and reference device de-synchronizing and re-synchronizing all within the time period of one epoch, effectively eliminating detectable DM interaction effects on the data stream.

Beyond individual template generation, we must choose an appropriate number of templates in the repository to accurately span the parameter space. In Sec. 9 we gauge how the number of templates affects the DM signal detection capabilities.

26

Chapter 7

Analytic results for idealized network

Now we turn to the general SNR (6.9) and determine the statistical properties of SNR for thin domain wall signals (6.11). In this section, we consider an analytically treatable case of an idealized network comprised of N_D identical white noise sensors. We additionally incorporate a white noise reference sensor common to all the sensors. This common noise reference sensor is especially relevant to GPS clock network, where it arises due to all clock biases reported with respect to a single reference clock. We will denote the intrinsic noise variance of the network sensors and the reference sensor as σ^2 and $\sigma^2_\times$, respectively. Both the sensors and the reference can be affected by dark matter transients.

We will determine the expected distribution of the template-specific SNR values ρ_i given that there is no signal in the data stream, $\mathrm{Prob}\,(\rho_i|H_0)$, as well as the distribution of the detection statistic given that there is a signal of strength h present, $\mathrm{Prob}\,(\rho|H_h)$, for this idealized sensor network.

As discussed in Sec. 2, the central quantity of interest, SNR (6.9), is a Gaussian random variable and as such its probability distribution is fully characterized by its mean value and variance. Because it is random, the SNR can fluctuate. Due to these fluctuation, even in the absence of the DM signal, the SNR may attain large values that can be falsely misinterpreted as the presence of the DM signal. The larger the SNR variance, the larger the fluctuations are, and the larger detection threshold must be.

For the idealized network, the inverse of the covariance matrix needed to compute the SNR statistic can be found analytically (see Appendix A)

$$\left(E^{-1}\right)^{ab}_{jl} = \frac{1}{\sigma^2}\delta_{jl}\left(\delta^{ab} - \frac{1}{N_\mathrm{D}}\frac{\xi}{1+\xi}\right), \tag{7.1}$$

27

where $\xi \equiv N_{\mathrm{D}}\sigma_\times^2/\sigma^2$.

Now, if there is a signal present in the data stream, each individual sensor's data is given by $d_j^a = e_j^a - c_j + h\bar{s}_j^a$ (a sum of an individual sensor noise, reference noise and a signal term). When the signal is absent, one can simply set $h \to 0$. Our explicit computation using Eq. (6.9) with $\bar{s}_i = \bar{s}$ (see Appendix C for derivation) results in a Gaussian distribution for ρ with a mean of

$$\mu_\rho = \frac{h\sqrt{N_{\mathrm{D}}}}{\sigma}\sqrt{\frac{1+\eta^2+\xi}{1+\xi}} \tag{7.2}$$

and variance of

$$\sigma_\rho^2 = 1\,. \tag{7.3}$$

Here $\eta \equiv h^R/h^a = \Gamma_{\mathrm{eff}}^R/\Gamma_{\mathrm{eff}}^a$ is the ratio between the strength of the signal experienced for the reference clock to that of the satellite clocks. Note that we assumed that device degeneracy (multiple sensors experiencing a signal at the same epoch) can be ignored. Note that when the DM signal is absent ($h = 0$), $\mu_\rho = 0$ while σ_ρ remains constant. Moreover, the standard deviation of template-specific SNR σ_{ρ_i} is also constant at 1. The probability density distribution for SNR (for a fixed, matching template) is given by

$$\mathrm{Prob}(\rho|H_h) = \frac{1}{\sqrt{2\pi}}\exp\left[-\frac{(\rho-\mu_\rho)^2}{2}\right]\,. \tag{7.4}$$

Assuming that none of the couplings Γ_X dominates the DM interaction with GPS devices, $\eta \approx 1$ for any of the satellite-reference clock combination (see Eqs. (4.6-4.8)). This remains true if either Γ_α or Γ_{m_e} dominates the interaction. The only major deviation of η from a value near 1 is when Γ_{m_q} is the dominate coupling, for which the use of a network of Rb clocks with an H-maser reference clock will result in $\eta \approx 2$. For the following analysis, we will assume that $\eta = 1$.

In the limit $\xi \ll 1$, i.e., $\sigma_\times \ll \sigma/\sqrt{N_{\mathrm{D}}}$, we arrive at $\mu_\rho = h\sqrt{2N_{\mathrm{D}}}/\sigma$, recovering the known result for a network of uncorrelated devices (see e.g., Ref [18]). For networks with large cross-correlation or large number of devices ($\xi \gg 1$), we arrive at $\mu_\rho = h\sqrt{N_{\mathrm{D}}}/\sigma$, a factor of $\sqrt{2}$ less than the uncorrelated network. Regardless of the level of cross-correlation, the network sensitivity grows with the sensor number as $\sqrt{N_{\mathrm{D}}}$.

In our search, we do not use the exact inversion of the covariance matrix as was used

to derive the expressions in this section. Instead we incorporate a perturbative inversion (see Appendix A) which assumes that the reference clock noise is small compared to the noise of the satellite clocks.

7.1 Multiple events

The Bayesian technique outlined in [2] assumed there to be at most one DM interaction event in any particular time window of the archival GPS data. So far, here we have also only treated the case of a single thin wall interaction event occurring in a time period of J_W epochs. However, if we consider dark matter encounters to be Poisson distributed in time, with an average time between consecutive events $\mathcal{T}$, over the 20 years of archival data we would expect there to be $N_E = (20\,\text{years})/\mathcal{T}$ events. Then, extending our search window J_W to contain the total number of epochs in the entire two-decade window of GPS data, and assuming that consecutive events are non-overlapping, we find that the mean of our detection statistic (7.2) increases by a factor of $\sqrt{N_E}$, while the variance of the statistic remains unchanged. This ultimately improves our sensitivity by $\sqrt{N_E}$.

While this section analyzed an idealized network, simplifying assumptions will be lifted in full numerical simulations in later sections. We will use real colored noise auto-correlation functions for heterogeneous networks of GPS clocks.

Chapter 8

Determining a detection threshold

In order to find a detection threshold, we must determine how the detection statistic behaves in the case when there is no signal present in the data. This way we can determine whether the computed statistic provides a significant evidence for rejecting the null hypothesis and claiming DM detection. Rather than obtaining the distribution for our test statistic given the null hypothesis is true, $\text{Prob}(\rho|H_0)$, we determine our detection threshold in a different yet equivalent fashion given the nature of our test statistic.

Recall that we define our detection statistic as the template-specific SNR with the *maximum magnitude* out of a repository of M templates. Instead of assessing the distribution of the maximum magnitude ρ_i, we simply assess the distribution of ρ_i. To this end, we performed Monte-Carlo simulations consisting of $\approx 10^6$ SNR calculations on event free simulated data and confirmed that the distributions for ρ_i given the null hypothesis is true are Gaussian with a mean of zero and standard deviation σ_{ρ_i}. The results of the simulations for various simulated clock networks (clock networks for the years 2000, 2005, 2010, and 2015; see Table 8.1) are provided in Fig. 8.1. Given that the template-specific SNR behaves in this fashion, the probability that any of the templates in the repository produces an SNR value larger in magnitude than some SNR threshold $\rho^* = n^* \sigma_{\rho_i}$, is

$$\text{Prob}(|\rho_i| > n^* \sigma_{\rho_i}|H_0) = 1 - \left[\text{erf}\left(\frac{n^*}{\sqrt{2}}\right)\right]^M, \tag{8.1}$$

where M is the number of templates used in the template repository. This is a false positive rate per epoch.

A reliable SNR threshold will ensure that we can expect less that 1 false positive in
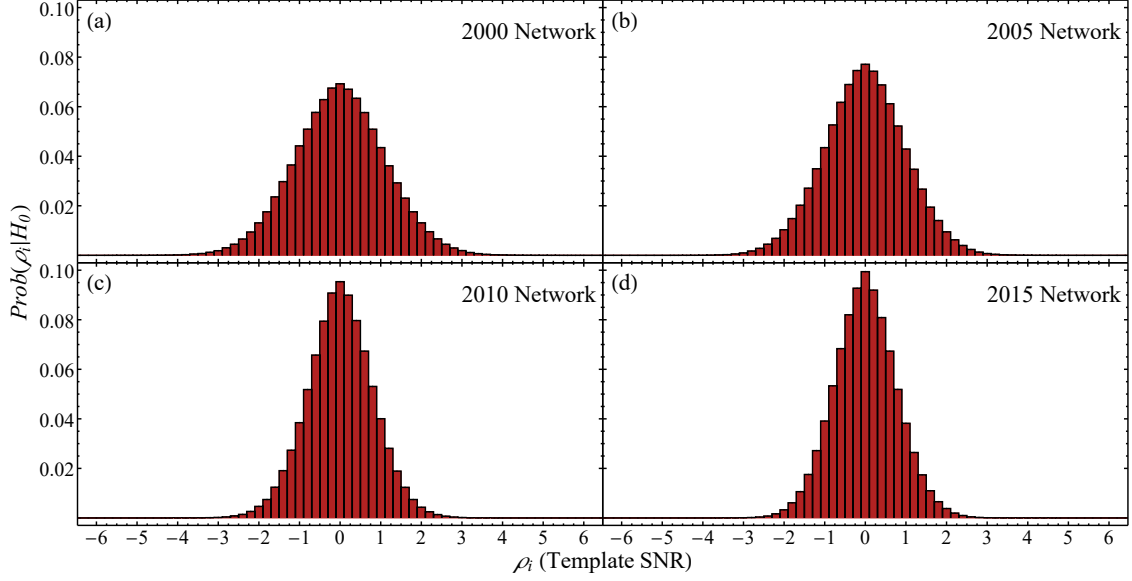
Figure 8.1: Histograms of template specific SNR calculations (ρ_i) on signal-free simulated data for various clock networks. The network composition for indicated years is given in Table 8.1. Standard deviations for template-specific SNR values are: 1.15 (2000), 1.03 (2005), 0.86 (2010), and 0.81 (2015).

Table 8.1: GPS network clock types used for simulation with a white noise reference clock.

Year	Cs-II	Cs-IIA	Cs-IIF	Rb-II	Rb-IIA	Rb-IIR	Rb-IIF	H[1]
2000	5	11	0	1	7	3	0	0
2005	1	8	0	1	8	12	0	0
2010	0	5	0	0	5	19	2	2
2015	0	1	1	0	3	19	3	0

Z epochs. The value of n^* that meets this criterion is

$$n^* = \sqrt{2}\,\mathrm{erf}^{-1}\left[\left(1 - \frac{1}{Z}\right)^{\frac{1}{M}}\right] \approx \sqrt{\mathrm{Log}\left(\frac{2M^2 Z^2}{\pi}\right)}. \tag{8.2}$$

The most reliable threshold would allow less than one false positive in the entire span of data. For 20 years of archival 30-second GPS data, $Z = 2.1 \times 10^7$ epochs. With $M = 1024$ templates in the repository, the value of n^* is 6.57, corresponding to a threshold SNR of $\rho^* = 6.57\sigma_{\rho_i}$. However, less strict detection thresholds may be used to identify possible weak candidate events for further investigation. Note that the value for n^* depends weakly (logarithmically) on the number of templates M, thereby it does not vary significantly for different sized template repositories. Ultimately, our detection threshold is given by

$\rho^* = n^* \sigma_{\rho_i}$. Using the distributions in Fig. 8.1, we calculate the detection thresholds for each network allowing for 1 false positive per day, 10 false positives per year, and 1 false positive in 20 years. Table 8.2 summarizes the results.

Table 8.2: Detection thresholds ρ^* as a function of various generations of GPS constellation and false positive (f.p.) rates. The network compositions for indicated years are given in Table 8.1.

Year	1 f.p./day	10 f.p./year	1 f.p./20 years
2000	5.87	6.61	7.58
2005	5.26	5.92	6.79
2010	4.40	4.95	5.67
2015	4.11	4.62	5.30

8.1 Future networks and alternative data processing

As mentioned in Sec. 7, in our simulations we use a perturbative inverse of the covariance matrix; see appendix A. This approximation relies on the noise level of the satellite clocks being sufficiently larger than the noise of the reference clock. The GPS clock networks for the years 2000 through 2015 satisfy the quiet reference clock requirement. In recent years, however, more stable Rb-IIF clocks with noise comparable to that of the reference clocks have been added to the GPS constellation thereby weakening the justification for the perturbative approximation for $\boldsymbol{E}^{-1}$. Moreover, future GNSS networks will contain a plethora of ground- and satellite- based H-maser clocks to be exploited in our searches (Galileo satellites already house stable H-masers [43]). Switching our method to use the exact inversion mitigates, but with the trade-off of computational overhead. We wish to avoid using an exact inversion of the covariance matrix due to the fact that it drastically increases computation time and would add a considerable amount of time to a search through the GPS data.

Considering the insufficiency of the perturbative approximation for more recent clock networks, here we offer a possible mitigation technique. As more accurate satellite clocks are being placed in orbit, more reference clocks are being placed around the globe. We propose pairing each of the satellite clocks with their own reference clock, thereby eliminating the cross-correlation caused by the use of a single reference clock that is inhibiting current search techniques. Suppose there are N_D atomic clocks, satellite- and Earth-based, at our

disposal with half of them being Earth-based. The large level of cross-correlation that restricts the perturbative inversion may be eliminated by using data from $N_D/2$ satellite-Earth clock pairs. The application of the matched-filter technique can be reformulated for a network of device pairs and is left for future work when such networks become a reality.

Chapter 9

Detection Probability

9.1 Detecting events

In the event of a weak DM signal presence in the data stream, it may not be immediately noticeable in the atomic clock data due to the randomness of the clock noise. The advantage of the SNR (and a detection statistic in general) is to provide a clear gauge of the signal presence. We verify that the SNR statistic is capable of detecting sought DM signals via simulation.

To this end, we simulated 2 hours of data for a network of $N_D = 30$ clocks that exhibit Gaussian white noise with a standard deviation of $\sigma = 0.05$ ns along with a white noise reference clock contribution with noise level $\sigma_\times = 0.006$ ns. We then injected a signal of strength $h = 0.1$ ns in the middle of the data stream with normal velocity of $v_\perp \approx 300\,\mathrm{km\,s^{-1}}$ and incident direction angles $\theta \approx 1.7\pi\,\mathrm{rad}$ and $\phi \approx 0.2\pi\,\mathrm{rad}$ (this in the Earth-centered Inertial (ECI) J2000 frame), which are the most probable event parameters according to the SHM and our previous calculations (see [2]). The results of performing our search technique on this data set is shown in Fig. 9.1, where we show the time series data (with the injected signal) for 6 of the 30 clocks. The upper panel shows magnitude of the calculated detection statistic [Eq. (6.10)] for each epoch in the two hour window. This simulation used a template repository of size $M = 1024$ and the calculated detection threshold for this type of network is given by $\rho^* = 6.57$ (allowing for no false-positives in 20 years). While the injected signal is not recognizable by the eye in the simulated data streams, a spike in the detection statistic at the time of the injected event is apparent.

Note that the search method is not aware of the event's strength, speed, direction, time of occurrence, or the fact that there was an injected signal at all. The injected signals were generated independently of the search routine and the template bank.
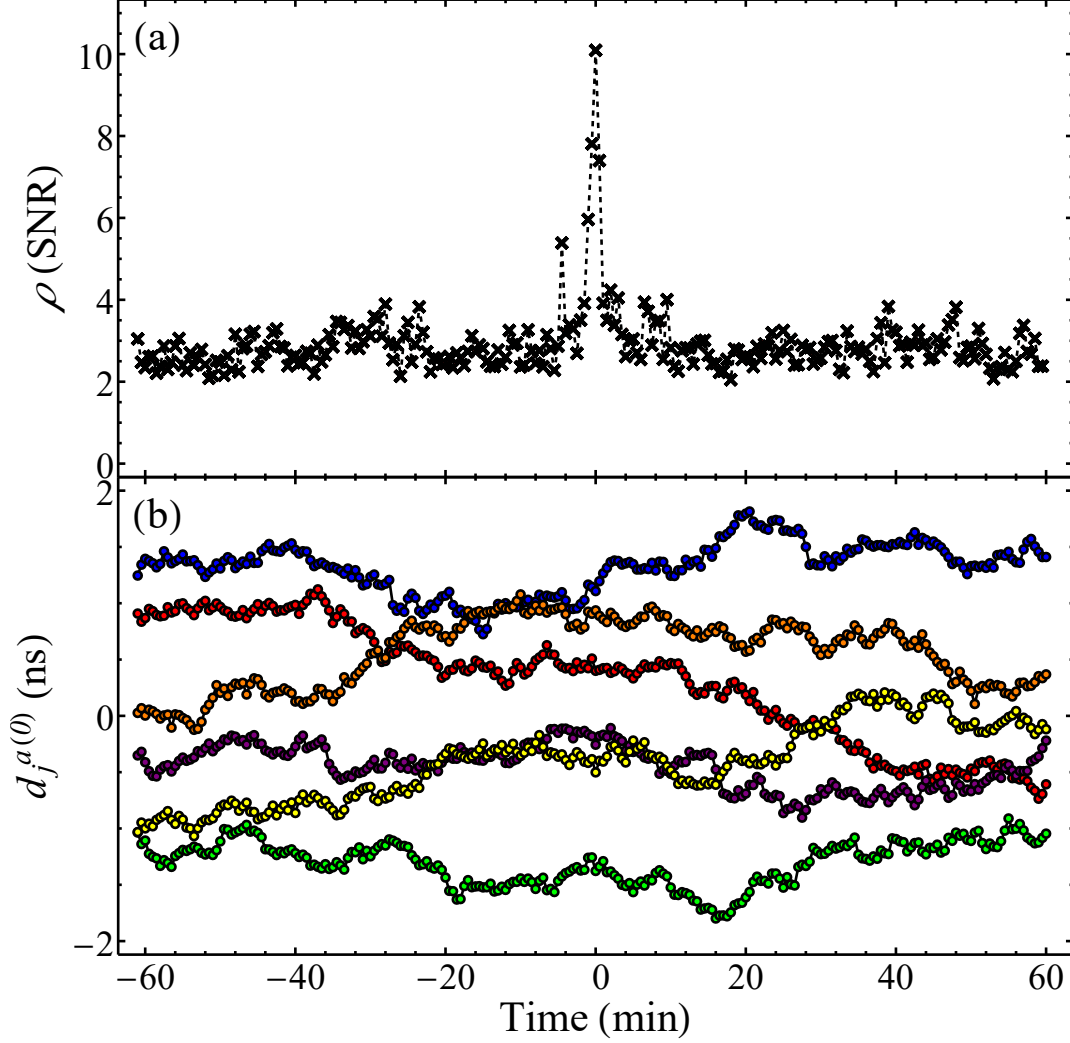
Figure 9.1: Frequentist detection of a simulated thin domain wall signal. The thin wall with strength $h = 0.10$ ns interacts with a network of 30 white noise clocks, each with $\sigma = 0.05$ ns, at time $t_0 = 0$. The common reference clock contribution is white noise with $\sigma_\times = 0.006$ ns. *Bottom panel*: simulated clock bias data for 6 of the 30 clocks with the injected signal included. For visual clarity, each data stream is shifted by a constant. *Top panel*: magnitude of the corresponding SNR statistic ρ for the same time scale.

9.2 Detection probability

The main figure of merit of the MFT algorithm is a detection probability curve for the various clock networks that have been collecting data for the past two decades. The detection probability is defined as the probability that the observed detection statistic exceeds the detection threshold given that the alternative hypothesis is true: $\text{Prob}(\rho > \rho^*|H_h)$. We wish to determine the detection probability for our various clock networks as a function of the signal strength h and obtain a 95% detection probability signal strength, denoted $h^{95\%, \text{D.P.}}$.

Monte-Carlo simulations produced the detection probability curves shown in Fig. 9.2. The simulation scheme consisted of 128 trials where a randomly-generated thin-wall signal of strength h was injected into a data stream for a particular clock network and the detection statistic was calculated for every epoch within the simulated data stream. An event was considered found if the computed SNR exceeded the network's threshold within one epoch of the injected event time. We compared the calculated SNR values with two different detection thresholds: one that allows for 10 false positive events per year and another one that allows for less than 1 false positive event in the 20 year span of the GPS data. The number of found events divided by the number of iterations gave us the detection probability. This detection probability was computed for a range of injected signal strengths.

Along with the detection probability curves, we plot the average of the 128 SNR calculations at the epoch where the signal was injected as a function of the signal strength. This is also shown in Fig. 9.2. We can see that the SNR is a linear function of h, as expected. This fact helps form a confidence interval for the signal strength h in the event that a DM signal is found.

To verify that our detection probability curve provides us with the correct value for $h^{95\%, \text{D.P.}}$, we performed an auxiliary simulation. Here we injected signals of strength $h^{95\%, \text{D.P.}} = 0.045\,\text{ns}$ for the 2010 clock network into a simulated data stream. For each signal injection with random parameters, we calculate the detection statistic ρ. The histogram of the detection statistic for 10^5 of these simulations is provided in Fig. 9.3. The resulting histogram confirms that the distribution for $\text{Prob}(\rho|H_h, h = h^{95\%, \text{D.P.}})$ is indeed Gaussian. Moreover, using a Gaussian distribution with the same mean and standard deviation as calculated, we find that $\text{Prob}(\rho > \rho^*|H_h, h = h^{95\%, \text{D.P.}}) = 0.946$,

Figure 9.2: *Top panel:* Detection probability curves as a function of the injected signal strength h for various simulated clock networks. Each point represents the percent of injected events of strength h found for 128 trials. Each curve contains 42 points. *Bottom panel:* The average detection statistic value ρ of the 128 trials at the time the event was injected. Though not plotted here, the 2005 network maintains a similar sensitivity to the 2000 clock network while that of the 2015 clock network sensitivity is similar to the 2010 network.

almost exactly as expected.

A major factor associated with detection probability is the number of devices in the network N_{D}. A more complete discussion of this using the analytic results from Sec. 7 is provided in Sec. 11.1. Our analysis of detection probability using simulated GPS data was continued by the varying the number of devices in the network N_D. We injected signals of varying strength into simulated homogeneous networks of $20, 30$, and 50 white noise devices. The percentage of events found as a function of the injected signal strength for these networks is shown in Fig 9.4. The average value of the detection statistics for each signal strength and clock network is also provided in the same figure. It is clear
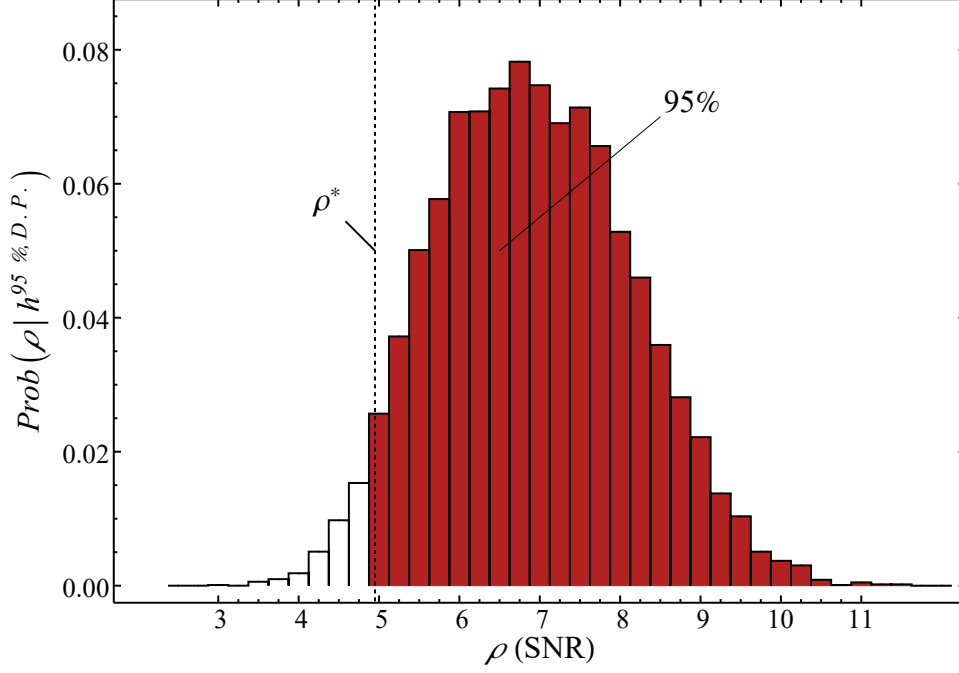
Figure 9.3: Histogram of the detection statistic for a simulated network replicating the clock types of the 2010 GPS constellation with an injected signal of strength equal $h^{95\%,\,\mathrm{D.P.}}$ ($h = 0.045\,\mathrm{ns}$). The dotted black line is the detection threshold for this network ($\rho^* = 4.95$). The shaded red region encompasses all iterations that resulted in a detection statistic larger than the threshold. 94.6% of the iterations fall into this region.

that our sensitivity to weaker signals improves as the number of devices in the network increases. We have found that $h^{95\%,\,\mathrm{D.P.}} \propto 1/\sqrt{N_{\mathrm{D}}}$, as expected.

To complete our analysis of factors affecting detection probability, we tested the effect of the template repository size M. To this end, we simulated a network of 30 homogeneous devices with standard deviation $\sigma = 0.05\,\mathrm{ns}$ and injected events of varying strengths into the data streams. The simulated reference clock had a standard deviation of $\sigma_\times = 0.006\,\mathrm{ns}$. We then calculated detection statistics for the event using repositories of 256, 1024, and 4096 templates. The effect of template size on sensitivity in provided in Fig. 9.5 along with the corresponding detection statistic. Notice that an increased template repository size results in better sensitivity and larger SNR values. However, increasing the number of templates results in an increase in the false positive rate [Eq. (8.1)] along with an increase in computation time. To balance the false positive rate, detection probability and computation time, we typically use 1024 templates in our template repositories.
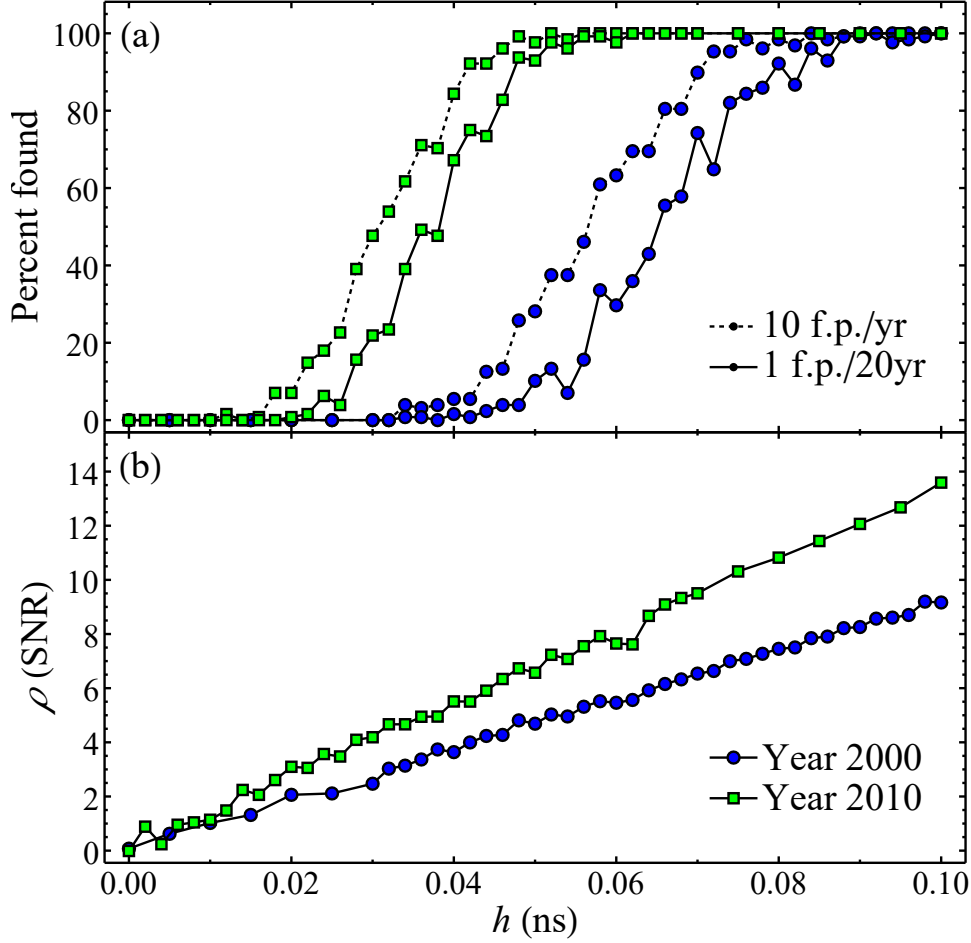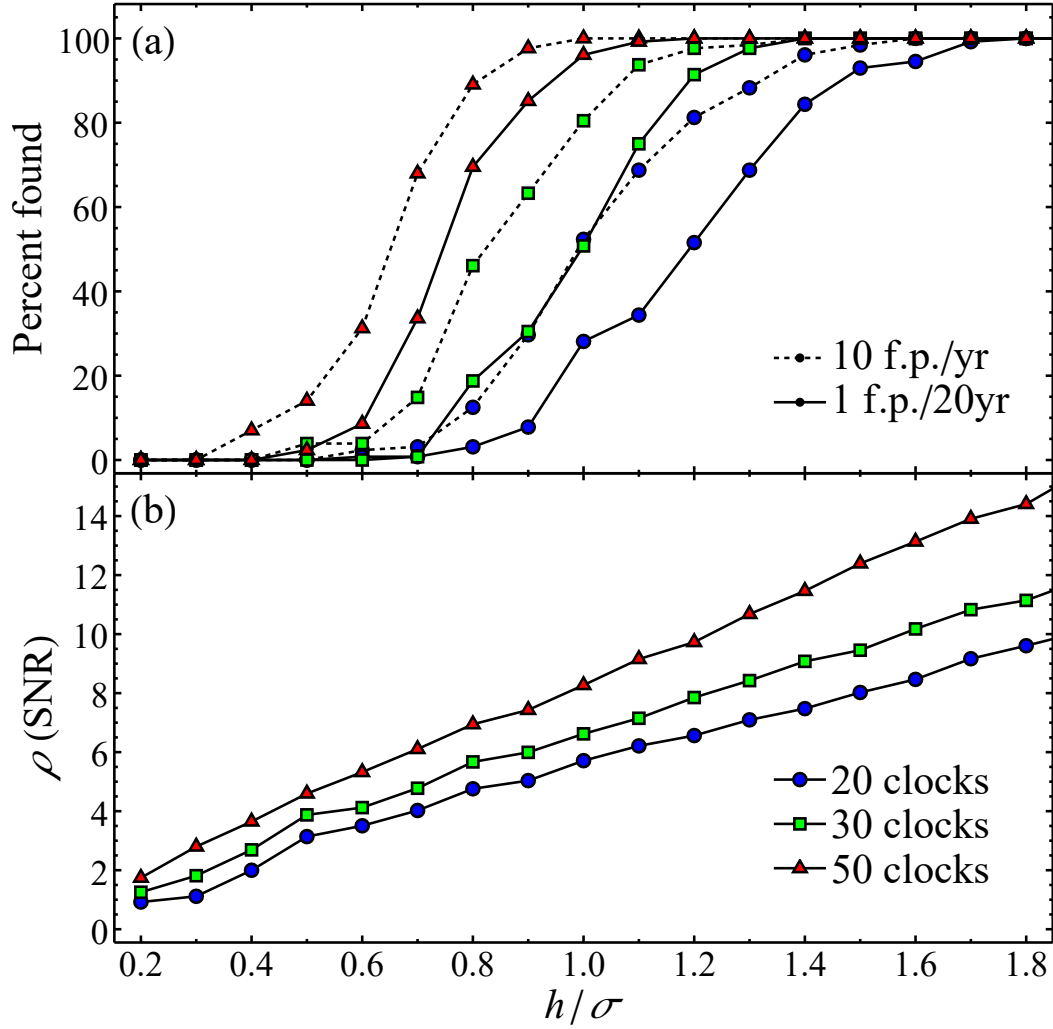
Figure 9.4: *Top panel:* Detection probability curves as a function of the injected signal strength h for simulated clock networks of varying size. Each point represents the percent of injected events of strength h found for 128 trials. Here σ represents the standard deviation of the individual white noise devices that comprise the simulated network. We use $\sigma_\times = 0$ in this simulation. *Bottom panel:* The average detection statistic value ρ of the 128 trials at the time the event was injected.
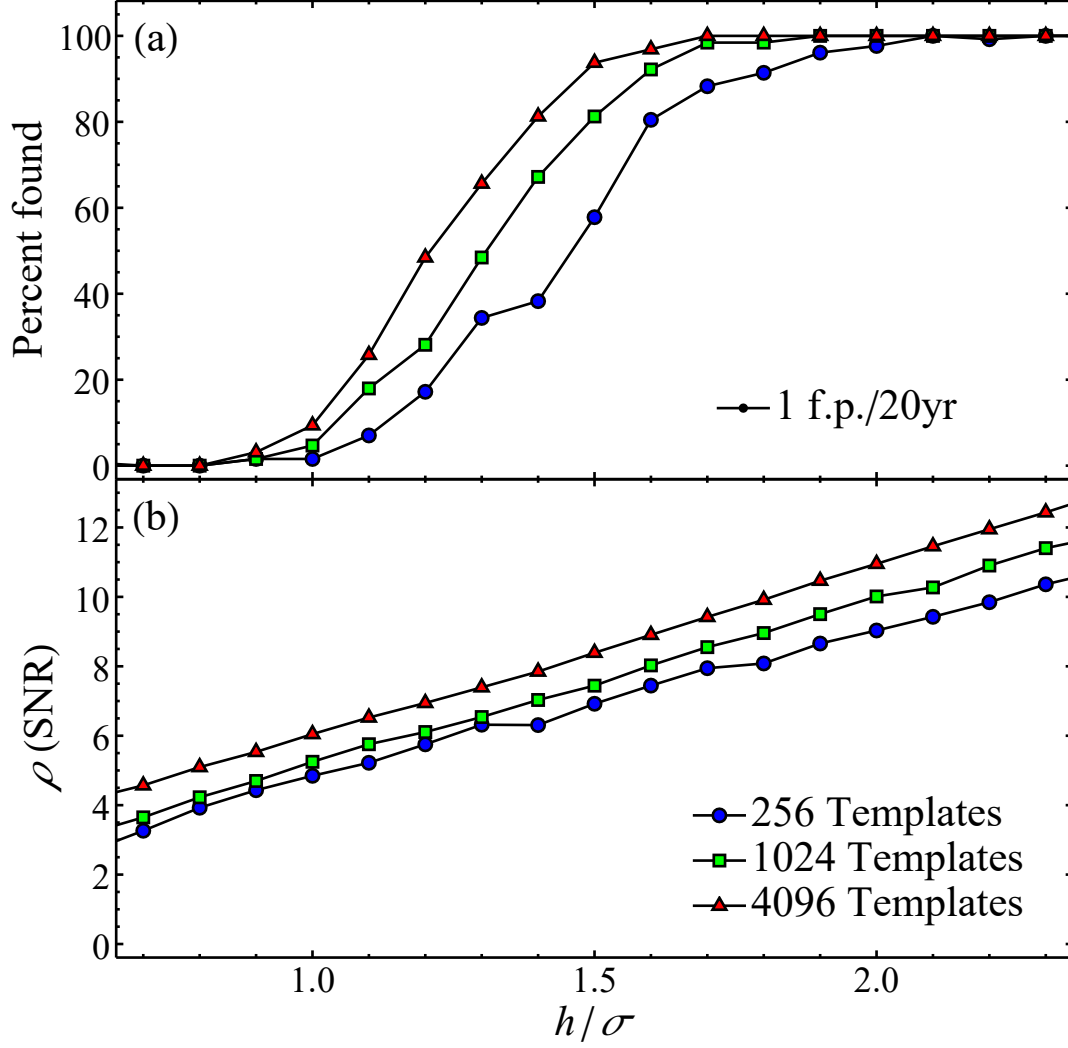
Figure 9.5: *Top panel:* Detection probability curves as a function of the injected signal strength h for varying template repository sizes. Each point represents the percentage of injected events of strength h found for 128 trials. Here σ represents the standard deviation of the individual white noise devices that comprise the simulated network. *Bottom panel:* The average detection statistic value ρ of the 128 trials at the time the event was injected.

Chapter 10

Parameter estimation

In the event that we find a DM signal in the data stream, our main goal is to estimate the parameters associated with the DM object that caused the signal. Among the parameters of interest are the incident speed, incident direction, event time, and signal strength. The estimates we provide on these parameters correspond to the parameters associated with the model-prescribed signal template that results in an SNR above the detection threshold.

In order to test the efficacy of our parameter estimation, we performed $\approx 20,000$ iterations of injecting a DM signal of considerable strength (twice the level of the noise standard deviation, $\sigma = 0.05\,\mathrm{ns}$) with random parameters into a stream of simulated white-noise data for $N_D = 30$ clocks. For each iteration, we calculate the SNR for every epoch in the simulated time window and store the event parameters that resulted in an SNR above the detection threshold. These extracted parameters are then compared to the injected parameters to check the precision of our parameter extraction. Histograms depicting our precision are shown in Fig. 10.1. Our resulting resolution was the following: $\pm 27\,\mathrm{km\,s^{-1}}$ for velocity, $\pm 0.05\pi$ radians for incident angle θ, and $\pm 5\,\mathrm{s}$ for the event time.
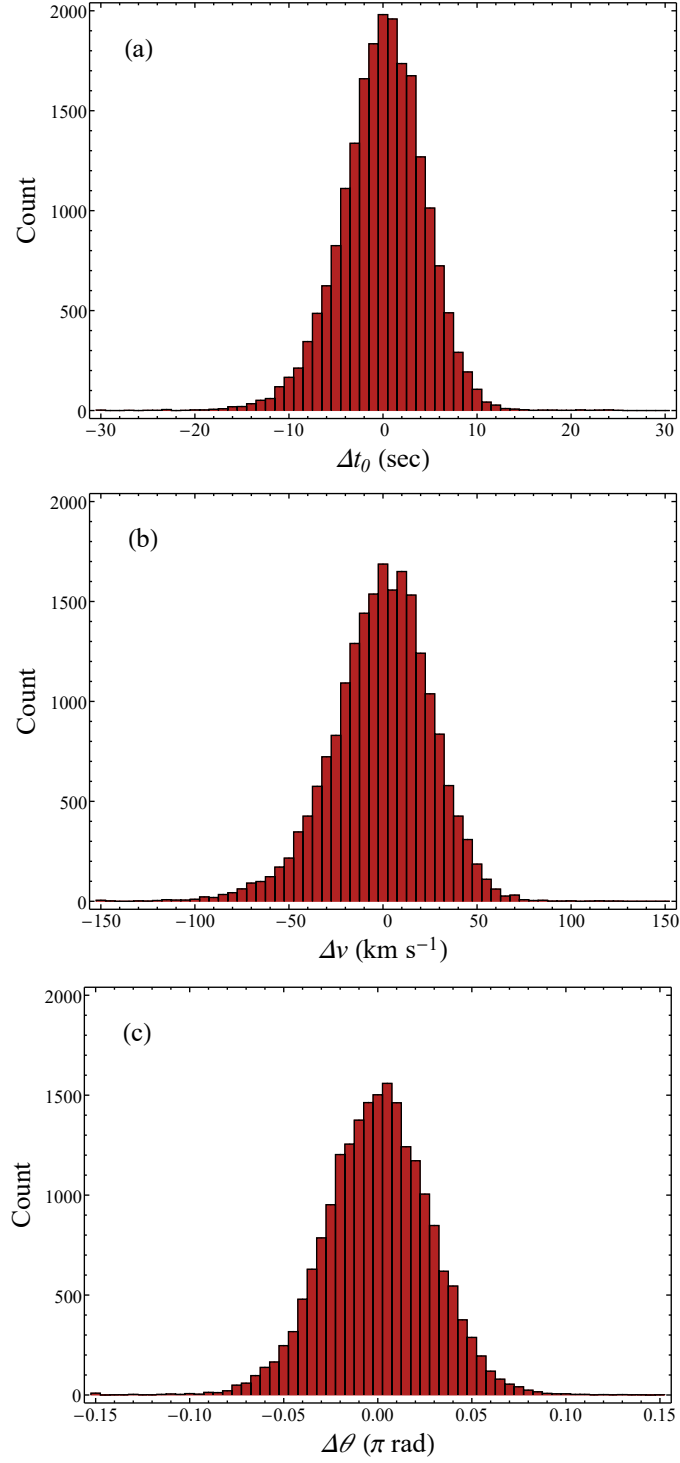
Figure 10.1: Histograms of the difference in extracted and injected event parameters from parameter estimation routine. *Top*: Event time t_0, 1σ resolution of $\pm 5\,\mathrm{s}$. *Middle*: Event velocity v, 1σ resolution of $\pm 27\,\mathrm{km\,s^{-1}}$. *Bottom*: Incident direction angle θ, resolution of $\pm 0.05\pi$ radians.

Chapter 11

Placing limits

Suppose we do not observe a DM interaction signature in the GPS atomic clock data stream. This means that there were no SNR values with a magnitude above the detection threshold ρ^*. We may then establish the lower and upper limits on the DM signal strength h. For the upper limit, suppose the largest SNR value we observed was ρ_{obs}. We define the 95% confidence upper limit $h^{95\%,\,\text{U.L.}}$ as the minimum value of h for which

$$\text{Prob}(\rho > \rho_{obs} | H_h) = 0.95\,. \tag{11.1}$$

That is, we find the minimum value of h for which we would observe an SNR value larger than ρ_{obs} 95% of the time if there was in fact a signal of strength h in the data stream. The 95% confidence lower limit, $h^{95\%,\,\text{L.L.}}$, is defined similarly. Since the SNR ρ is an odd function in h, $h^{95\%,\,\text{L.L.}} = -h^{95\%,\,\text{U.L.}}$.

11.1 Maximum and minimum sensitivity given clock network characteristics

Before analyzing the data, we can project a minimum upper limit and maximum lower limit on h by replacing $\rho_{\text{obs}} \to \langle \rho_{\text{obs}} \rangle$. Then, the minimum 95% confidence upper limit for h is the minimum value of h for which

$$\text{Prob}(\rho > \langle \rho_{\text{obs}} \rangle \,| h = h^*, H_h) = 0.95\,. \tag{11.2}$$

We will denote the value of h that satisfies this requirement by h^*. Assuming that events are weak, i.e., well below the noise floor, it is clear by the nature of the SNR $\langle \rho_{\mathrm{obs}} \rangle \to 0$. Once again, the maximum lower limit is defined in a similar fashion, resulting in $-h^*$ serving as the maximum lower limit for the signal strength. The maximum possible exclusion limits can be placed on the magnitude of h with 95% confidence by bounding it by h^*: $|h| < |h^*|$.

For the idealized sensor network of Sec. 7, we are able to find the exact relation between h^* and the network characteristics ($N_{\mathrm{D}}, \sigma, \xi,$ and η) using probability distribution (7.4). Ultimately, we find that

$$h^* = K \frac{\sigma}{\sqrt{N_{\mathrm{D}} N_E}} \sqrt{\frac{1 + \xi}{1 + \xi + \eta^2}}, \tag{11.3}$$

where K is determined by the level of confidence (for 95% confidence $K = 1.64$). Notice that when $\xi = N_{\mathrm{D}} \sigma_\times^2 / \sigma^2 \ll 1$ (i.e., when cross-correlation is negligible) our sensitivity is $K\sigma/\sqrt{N_{\mathrm{D}} N_E (1 + \eta^2)}$. However, when cross-correlation is considerable, or the network is large ($\xi \gg 1$), the sensitivity becomes $K\sigma/\sqrt{N_{\mathrm{D}} N_E}$. Thus, when the reference sensor is noisy, its sensitivity encoded by the constant $\eta = \Gamma_{\mathrm{eff}}^R / \Gamma_{\mathrm{eff}}^a$ is effectively suppressed.

We can also estimate our minimum sensitivity by projecting a maximum upper limit and minimum lower limit on h by using our detection threshold value as the maximum possible observed SNR value for which we do not claim a detection. Then, the minimum 95% confidence lower limit is given by the minimum value of h for which $\mathrm{Prob}(\rho > \rho^* | H_h) = 0.95$. It should be clear that the maximum upper limit above is the same as the 95% detection probability signal strength $h^{95\%, \mathrm{D.P.}}$. In this case, $h^{95\%, \mathrm{D.P.}}$ scales in the same fashion as h^* above, but the constant of proportionality K increases as a function of the false positive rate and template repository size (for less than 1 false positive in 20 years and 1024 templates in the repository, $K = 8.2$). Ultimately, this makes the minimum sensitivity reach nearly an order of magnitude below the maximum predicted reach.

11.2 Projected sensitivity and discovery reach

Given a DM model type (domain wall, monopole, etc.), the signal strength h links to the specific field parameters for those models. For a thin domain wall, the average strength h is related to the effective coupling and DM object parameters by

$$h_{\mathrm{avg}} = \hbar c \Gamma_{\mathrm{eff}} \sqrt{\pi} \rho_{\mathrm{DM}} d^2 \mathcal{T} \,, \tag{11.4}$$

in agreement with Ref. [17].

If we assume that a specific coupling Γ_X dominates the effective coupling then $\Gamma_{\mathrm{eff}} \rightarrow \kappa_X \Gamma_X$ in Eq. (11.4). The limit on h, $|h_{\mathrm{avg}}| \leq h^*$ translates into a limit on the coupling constant for a particular coupling to a fundamental constant Γ_X

$$|\Gamma_X| \leq \frac{|h^*|}{\hbar c \rho_{\mathrm{DM}} \mathcal{T} \sqrt{\pi} d^2 \kappa_X} \,, \tag{11.5}$$

where for idealized network h^* is given by Eq. (11.3). This provides projected exclusion limits on the effective energy scale $\Lambda_X = 1/\sqrt{|\Gamma_X|}$,

$$\Lambda_X > d \sqrt{\frac{\hbar c \rho_{\mathrm{DM}} \mathcal{T} \sqrt{\pi} \kappa_X}{|h^*|}} \,. \tag{11.6}$$

Our projected discovery reach using h^* for the 2010 GPS network ($\sigma \approx 0.02$, $\sigma_\times \approx 0.006$, $N_{\mathrm{D}} = 33$, and $\eta \approx 1$), is plotted in Fig. 11.1 along with existing constraints. This discovery reach includes the possibility of multiple DM interaction events occurring within the time window of the search, which results in a sensitivity that is comparable to that of optical clocks [3, 39]. Notice that the projected sensitivity reach in Fig. 11.1 exhibits a sharp cutoff for domain-walls of thickness larger than 10^4 km and for average times between events larger than $\mathcal{T} = 20$ years. This is due to the fact that DM objects of size larger than 10^4 km will affect satellite clocks and the reference clock simultaneously, resulting in a no detectable signal is the data stream. Moreover, for thin domain walls, we require that the signal be present for just a single epoch. The regime for $d \geq 10^4$ km belongs to "thick" domain walls (see Ref. [2]). The sharp cutoff for average time between events larger than 20 years comes from the fact that only two decades of archival data exists.
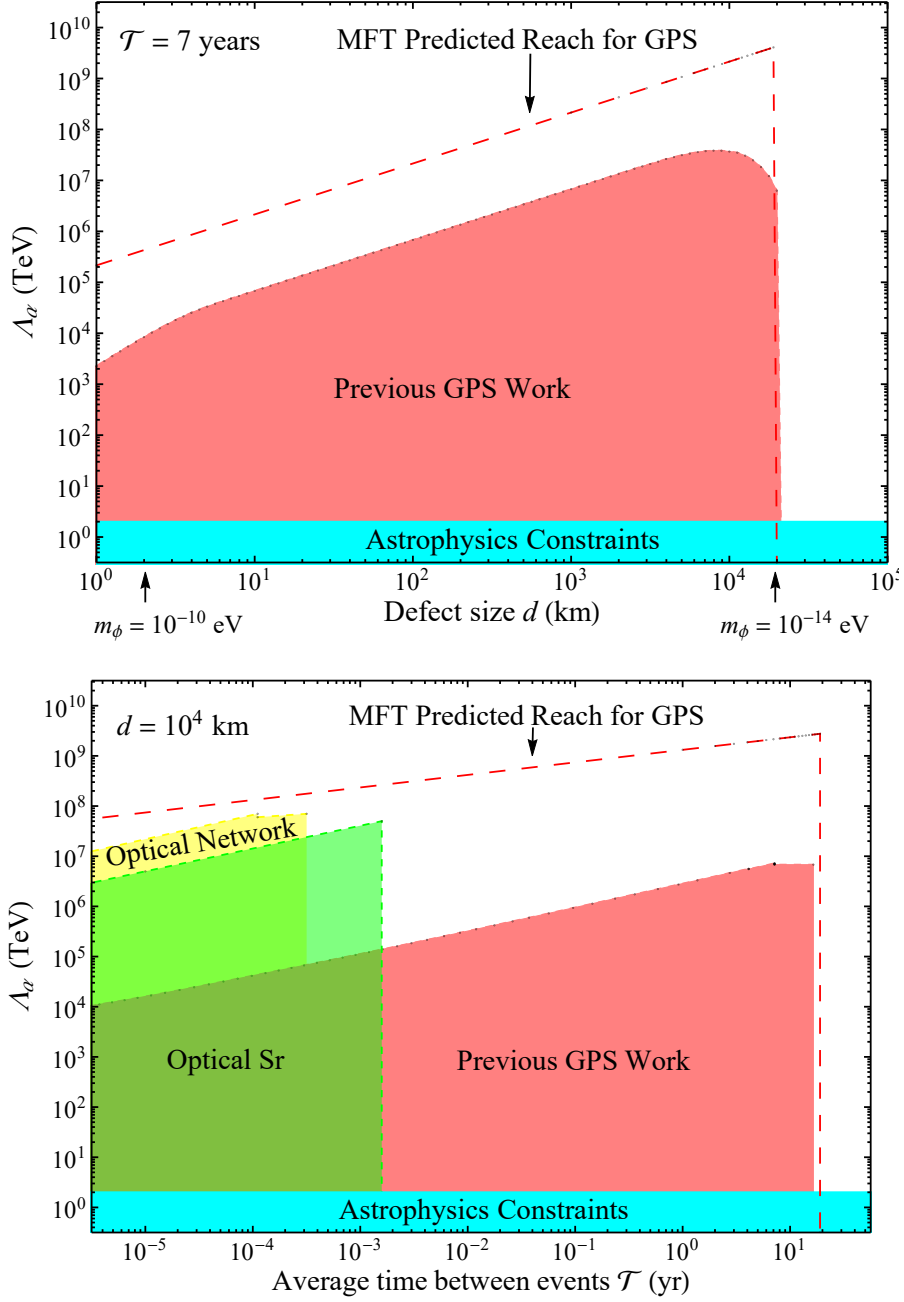
Figure 11.1: Projected discovery reach for thin wall dark matter objects using the matched-filter technique along with existing constraints. The red dotted lines represent the least stringent and most stringent discovery reaches for the 2010 GPS atomic clock network. The shaded blue regions are the constraints coming from astrophysics [1] while the salmon shaded regions are the constraints placed by previous work from the GPS.DM collaboration [2]. The green shaded region contains the constraints placed by optical clock experiments [3], while the yellow region contains the constraints from a global terrestrial network of laboratory clocks [4].

46

Chapter 12

Conclusion

In this paper we focused on detecting dark matter transients with networks of atomic sensors. We formalized the desired characteristics of such networks and developed applications of matched-filter technique in the network settings. We extended the previous literature to the practically important case of a network with cross-node correlations. This setting is especially relevant to GPS atomic clock network. Our simulations have proved the method's signal detection and event parameter estimation capabilities.

While our paper deals with classical networks of quantum sensors, it is worth noting recent proposals [44, 45] for massively entangled networks of atomic clocks. In these networks, entanglement is spread not only over an atomic ensemble at a single node, but also over nodes. We leave generalization of our paper to entangled networks for the future work when such entangled networks become a reality.

Appendix A

The network covariance matrix and its perturbative inversion

A.1 Properties of the network covariance matrix

The covariance matrix $\boldsymbol{E}$ is given by the ensemble average

$$E_{jl}^{ab} = \langle n_j^a n_l^b \rangle \equiv E_{(aj)(bl)} \,, \tag{A.1}$$

where n_j^a is the noise in the datastream of the a-th device at the temporal grid point (epoch) j, and $\langle n_j^a \rangle = 0$ is assumed. Here subscripts j and l range over epochs and the superscripts a and b span network sensors. When $a = b$ the covariance refers to a single instrument, while cross-node correlations are given by the $a \neq b$ elements. The matrix E_{jl}^{ab} can be visualized as a 2D matrix with super-indexes (aj) and (bl): $E_{(aj)(bl)}$. The dimension of the matrix is determined by the number of devices in the network (excluding the reference clock) and the number of points in data window, $N_{\mathrm{D}} \times J_W$. Because the data streams are stationary, the covariance matrix only depends on the lag $|j - l|$. From the definition (A.1), it is apparent that the covariance matrix is symmetric with respect to swapping the (aj) and (bl) super-indexes. Further, the covariance matrix is positive (semi-) definite.

For the GPS constellation, as discussed in Sec. 5, the noise component entering the

definition (A.1) can be represented as

$$n_j^a = e_j^a - c_j \,, \tag{A.2}$$

where e_j^a is the individual clock noise and c_j is the contribution from the reference clock noise common to all data streams. Then

$$
\begin{aligned}
E_{jl}^{ab} &= \langle e_j^a e_l^b \rangle - \langle e_j^a c_l \rangle - \langle c_j e_l^b \rangle + \langle c_j c_l \rangle \\
&= \langle e_j^a e_l^b \rangle + \langle c_j c_l \rangle \,,
\end{aligned} \tag{A.3}
$$

as the reference and the node clock noises are uncorrelated.

While the definition of the covariance matrix (A.1) explicitly refers to noise, in practice [2] we use data d_j^a to compute this matrix (this assumes that DM events are exceedingly rare, so that most of contributions into the above values comes from the intrinsic noise of the network). Notice that in our approximation, the covariance matrix does not depend on the spatial geometry of the network - however it does depend on the network composition. For example, for GPS, if the reference clock is switched to a different clock or a satellite clock is swapped, the covariance matrix is affected and needs to be recomputed.

To gain an insight into the structure of the covariance matrix, consider a simplifying case: suppose the network is comprised from white-noise devices (including reference clock). Then

$$E_{jl}^{ab} = \sigma_a^2 \delta_{jl} \delta^{ab} + \sigma_\times^2 \delta_{jl} \,, \tag{A.4}$$

where σ_a^2 and $\sigma_\times^2$ are variances for the individual nodes and the reference clock respectively. The common noise source contributes to all the clocks. For example, for $N_\mathrm{D} = 2$

nodes and $J_W = 3$ time window, the covariance matrix reads

$$
\boldsymbol{E} =
\begin{pmatrix}
\begin{array}{|ccc|}
\hline
\sigma_1^2 & 0 & 0 \\
0 & \sigma_1^2 & 0 \\
0 & 0 & \sigma_1^2 \\
\hline
\end{array}
& \mathbf{0} \\[2ex]
\mathbf{0} &
\begin{array}{|ccc|}
\hline
\sigma_2^2 & 0 & 0 \\
0 & \sigma_2^2 & 0 \\
0 & 0 & \sigma_2^2 \\
\hline
\end{array}
\end{pmatrix}
+
\begin{pmatrix}
\begin{array}{|ccc|}
\hline
\sigma_\times^2 & 0 & 0 \\
0 & \sigma_\times^2 & 0 \\
0 & 0 & \sigma_\times^2 \\
\hline
\end{array}
&
\begin{array}{|ccc|}
\hline
\sigma_\times^2 & 0 & 0 \\
0 & \sigma_\times^2 & 0 \\
0 & 0 & \sigma_\times^2 \\
\hline
\end{array}
\\
\begin{array}{|ccc|}
\hline
\sigma_\times^2 & 0 & 0 \\
0 & \sigma_\times^2 & 0 \\
0 & 0 & \sigma_\times^2 \\
\hline
\end{array}
&
\begin{array}{|ccc|}
\hline
\sigma_\times^2 & 0 & 0 \\
0 & \sigma_\times^2 & 0 \\
0 & 0 & \sigma_\times^2 \\
\hline
\end{array}
\end{pmatrix}.
$$

$$(A.5)$$

The block structure of the covariance matrix is apparent. Each block corresponds to individual sensors and the elements inside each block refer to epochs.

For colored noise sensors each block is assembled from elements of the auto-correlation functions $A_a(|j-l|) = \langle e_j^a e_l^a \rangle / \sigma_a^2$ and $A_\times(|j-l|) = \langle c_j c_l \rangle / \sigma_\times^2$, so that Eq. (A.3) becomes

$$
E_{jl}^{ab} = \sigma_a^2 A_a(|j-l|)\delta^{ab} + \sigma_\times^2 A_\times(|j-l|)\,. \tag{A.6}
$$

Thereby, the covariance matrix is a block matrix: the block diagonals are composed of the sum of cross-node and individual device auto-correlation functions while the off-diagonal blocks contain the cross-node correlation. By the definition of the auto-correlation function, $|A_{a,\times}(|j-l|)| \leq 1$, and typically inside each block the elements with larger lag (further away from diagonals) become smaller. $A_{a,\times}(0) = 1$ by definition.

Based on these observations, and to aid in computer implementation, we can introduce blocks $\boldsymbol{A}^{ab}$ and $\boldsymbol{X}^{ab}$, so that the corresponding blocks of the covariance matrix $\boldsymbol{E}^{ab} = \boldsymbol{A}^{ab} + \boldsymbol{X}^{ab}$, with each block internally assembled as (c.f., Eq. (A.6))

$$
\left(\boldsymbol{A}^{ab}\right)_{ij} = \sigma_a^2 A_a(|j-i|)\delta^{ab}\,, \tag{A.7}
$$

$$
\left(\boldsymbol{X}^{ab}\right)_{ij} = \sigma_\times^2 A_\times(|j-i|)\,. \tag{A.8}
$$

Because $\boldsymbol{X}^{ab}$ does not depend on the particular sensor, we simply refer to all such blocks as $\overline{\boldsymbol{X}}$, i.e., $\boldsymbol{X}^{ab} \equiv \overline{\boldsymbol{X}}$.

We are interested in the inverse of the covariance matrix required for computing

the SNR statistic (6.9). For our white noise example (A.5), the inversion of the first matrix is trivial as it is diagonal. The second (×-node covariance matrix) contribution introduces off-diagonal matrix elements making the inversion difficult. Moreover, the ×-node covariance matrix is singular.

It is instructive to rewrite the definition of the inverse, $\boldsymbol{E}\boldsymbol{E}^{-1} = \boldsymbol{I}$, in our notation,

$$\sum_{bj} E_{(ai)(bj)} \left(E^{-1}\right)_{(bj)(a'i')} = \sum_{bj} E_{ij}^{ab} \left(E^{-1}\right)_{ji'}^{ba'} = \delta^{aa'}\delta_{ii'} \,. \tag{A.9}$$

We derived the covariance matrix inverse in a closed form for a special case of white noise devices (A.4),

$$\left(E^{-1}\right)_{jl}^{ab} = \frac{1}{\sigma^2}\delta_{jl}\Big(\delta^{ab} - \frac{1}{N_{\mathrm{D}}}\frac{\xi}{1+\xi}\Big) \,, \tag{A.10}$$

where $\xi \equiv N_{\mathrm{D}}\sigma_\times^2/\sigma^2$. One can verify directly that Eq. (A.9) is satisfied. The inverse retains the same block structure as the original matrix (A.4). Its derivation is outlined in the following section.

Additionally, certain simplifications can be obtained using discrete Fourier transformation (DFT) (see, e.g., Ref. [23], where DFT for a network covariance matrix was carried out). In DFT, the transformed matrix becomes block-diagonal, each block being of dimension N_{D}, thus simplifying the inversion procedure. However, this approach also requires DFT of the sought DM signal. In our work, this signal is non-oscillatory making the interpretation of the DFT procedure non-transparent; we leave the DFT implementation for future work if the computational speed-up is needed. In this work, the full covariance matrix (A.6) is inverted numerically using Cholesky decomposition. Below we outline a perturbative method which holds in the limit when the noise of the reference sensor is well below that of the network sensors. We used this perturbative inversion for older (pre-2015) GPS data, where this approximation remains valid.

A.2 Perturbative inversion

This approximation to inverting the network covariance matrix was used in our earlier GPS.DM work [2] and we will detail it below. It relies on the von Neumann series

expansion,

$$(\boldsymbol{G} + \lambda \boldsymbol{F})^{-1} = \boldsymbol{G}^{-1} \sum_{n=0}^{\infty} (-1)^n \lambda^n \left(\boldsymbol{F} \boldsymbol{G}^{-1} \right)^n$$

$$\approx \boldsymbol{G}^{-1} - \lambda \boldsymbol{G}^{-1} \boldsymbol{F} \boldsymbol{G}^{-1} \,, \tag{A.11}$$

where $\boldsymbol{G}$ and $\boldsymbol{F}$ are matrices and λ is the expansion (book-keeping) parameter. This identity can be proven by matching terms of the same power of λ in the definition $(\boldsymbol{G} + \lambda \boldsymbol{F})^{-1} (\boldsymbol{G} + \lambda \boldsymbol{F}) = \boldsymbol{I}$. The series converges as long as the absolute values of eigenvalues of the product matrix $\boldsymbol{F} \boldsymbol{G}^{-1}$ are smaller than $1/|\lambda|$.

Returning to our covariance matrix $\boldsymbol{E}$, we make the following identification using our block decomposition (A.8)

$$\boldsymbol{G}^{ab} = \boldsymbol{A}^{aa} \delta^{ab} \,, \tag{A.12}$$

$$\boldsymbol{F}^{ab} = \overline{\boldsymbol{X}} \,, \tag{A.13}$$

and $\lambda = 1$. Notice that $\boldsymbol{G}$ is a block-diagonal matrix (inverse of such a matrix is again a block-diagonal matrix composed of inverses of original blocks).

Now we illustrate this perturbative technique for a network of white-noise devices. Here the covariance matrix is given by Eq. (A.4). Then the above decomposition leads to $G_{jl}^{ab} = \sigma_a^2 \delta^{ab} \delta_{jl}$, $F_{jl}^{ab} = \sigma_\times^2 \delta_{jl}$. Then

$$\left(E^{-1} \right)_{ij}^{ab} = \left(\frac{1}{\sigma_a^2} \delta^{ab} - \frac{\sigma_\times^2}{\sigma_a^2 \sigma_b^2} \right) \delta_{ij} \,, \tag{A.14}$$

which are the two leading terms in the expansion of the exact result (A.10) in ξ. Nominally, the contribution of the second, perturbative, term is suppressed when $\sigma_\times \ll \min_a(\sigma_a)$, i.e., this approximation can only be used for networks of sensors that have noise levels far greater than that of the reference sensor. In cases when the reference sensor noise is comparable to that of the network sensors, the approximation breaks down and either an exact inversion must be used or mitigation techniques must be implemented to eliminate or minimize the reference sensor contribution to the individual sensor data streams (see Sec. 8.1).

For the general case of colored noise sensors, we return to our block decomposi-

tion (A.8) of the network covariance matrix and focus on a single block,

$$(\boldsymbol{E}^{-1})^{ab} \approx (\boldsymbol{G}^{-1})^{ab} - (\boldsymbol{G}^{-1}\boldsymbol{F}\boldsymbol{G}^{-1})^{ab}.$$

Because $\boldsymbol{G}$ is block-diagonal, $(\boldsymbol{G}^{-1})^{ab} = \delta^{ab}(\boldsymbol{A}^{aa})^{-1}$. To simplify the second term in the expansion, recall the product rule for block matrices,

$$(\boldsymbol{A}\boldsymbol{B})^{ab} = \sum_{c=1}^{N_{\mathrm{D}}} \boldsymbol{A}^{ac}\boldsymbol{B}^{cb}. \tag{A.15}$$

This rule parallels the conventional matrix multiplication with individual matrix elements replaced with blocks. Then

$$(\boldsymbol{E}^{-1})^{ab} \approx \delta^{ab}(\boldsymbol{A}^{aa})^{-1} - (\boldsymbol{A}^{aa})^{-1}\,\overline{\boldsymbol{X}}\,(\boldsymbol{A}^{bb})^{-1}. \tag{A.16}$$

Further approximation may consist in neglecting off-diagonal matrix elements of the $\times$-node correlation function in the above expression [2], $\overline{\boldsymbol{X}} \equiv \sigma_\times^2 \boldsymbol{I}$. In this secondary approximation,

$$(\boldsymbol{E}^{-1})^{ab} \approx \delta^{ab}(\boldsymbol{A}^{aa})^{-1} - \sigma_\times^2 (\boldsymbol{A}^{aa})^{-1}(\boldsymbol{A}^{bb})^{-1}. \tag{A.17}$$

This is the approximation used in our calculations for pre-2015 GPS network generations. We also point out that the exact covariance matrix inversion (A.10) for our idealized network of white noise sensors can be derived by following these block-matrix steps and summing the von Neumann series (A.11) to all orders.

A.3   Performance comparison between exact and perturbative inversion

In order to utilize the perturbative inversion outlines in the above section, we require that the reference device noise be sufficiently smaller than that of the node devices. To verify the inadequacy of the perturbative inversion for networks with a noisy reference sensor, we simulated signal-free data for a homogeneous network of 30 white noise node devices for various reference device noise levels ($\sigma_\times/\sigma = 0$, 0.1, 0.5, and 1). We then calculated standard deviation of $\approx 25,000$ template-specific SNR values for the simulated

data streams using both the exact inversion and the approximate inversion of $\boldsymbol{E}$. The results of these simulations are provided in Table A.1. We find that the results using the perturbative inversion are nearly identical to the exact inversion for small levels of cross-correlation ($\sigma_\times/\sigma \leq 0.1$). However, when the noise of the reference sensor is large, the approximate inversion behaves poorly compared to the expected value of σ_ρ. Note that the deviations from $\sigma_\rho = 1$ in the exact inversion column can be attributed to sampling error.

Table A.1: Comparison of performance of exact numerical inversion of the covariance matrix and the perturbative inversion for various degrees of cross-correlation with simulated white noise data. The computed values for σ_ρ are to be compared with the exact analytic result of $\sigma_\rho = 1$, see Sec. C.1.

$\sigma_\times/\sigma$	σ_ρ (Exact $\boldsymbol{E}^{-1}$)	σ_ρ (Approx. $\boldsymbol{E}^{-1}$)
0.0	0.99	1.00
0.1	1.04	0.99
0.5	1.00	11.15
0.9	0.99	64.07

Appendix B

Ratio of Degenerate Clocks

To approximate the ratio of clocks that experience a degenerate signal (i.e., interact with a DM object within the same epoch), we will assume that the network of N clocks are spaced out homogeniously on the surface of a sphere of radius r. For DM objects traveling at galactic velocities, $r \approx 3$ epochs is the time it takes for a DM object to traverse the GPS constellation. Any clocks within a 1 epoch distance of each other will experience a degenerate signal. We approximate the ratio of degenerate clocks by the ratio of the volume cylindrical slice of the spherical GPS constellation swept by the DM object over a period of 1 epoch to the total volume of the spherical GPS constellation (equal to $\frac{4}{3}\pi r^3$). The volume of the cylindrical slice is given by $V_s = \pi r_s^2 h$ where r_s is the average radius a circle created by a plane intersecting a sphere of radius r and h is equal to 1 epoch. The average radius of this circle is the same as the average value of x for a quarter-circle of radius r, that is, the average value of the function $f(x) = \sqrt{r^2 - x^2}$ from $x = 0$ to $x = r$. Using the mean value theorem for integrals, we find that

$$r_s = \frac{1}{r} \int_0^r \sqrt{r^2 - x^2} dx = \frac{\pi}{4} r \,. \tag{B.1}$$

This leads to the expected volume of the swept slice to be

$$V_s = \pi \left(\frac{\pi}{4} r\right)^2 h \tag{B.2}$$

It follows that λ is given by

$$\lambda = \frac{V_s}{V} = \frac{\pi^2}{4^3} \approx 0.1542 \,. \tag{B.3}$$

Appendix C

Derivation of SNR for idealized network

Consider a homogeneous network of N_D devices each with Gaussian white noise profiles with zero mean and standard deviation σ, along with a reference sensor also with a Gaussian white noise profile with a standard deviation of $\sigma_\times$ and also with zero mean. That is,

$$e_j \sim \text{Normal}(0, \sigma^2) \qquad \text{and} \qquad c_j \sim \text{Normal}(0, \sigma_\times^2) \,.$$

Data that contains a dark matter transient signal will be of the form

$$d_j = e_j - c_j + h\overline{s}_j \,, \tag{C.1}$$

where e_j is the node sensor Gaussian noise at epoch j (with variance σ^2), c_j is the reference sensor Gaussian noise at epoch j (with variance $\sigma_\times^2$), and $\overline{s}_j$ is the "unit-ized" DM signal which is scaled by the DM signal strength h (the strength of the signal felt by the network sensors). Since the network is assumed to be homogeneous, h is the same for all network sensors, though we allow for the possibility of the strength of the DM interaction with the reference device to be different than that of the satellite nodes. In this case, the reference sensor experiences a signal of strength h^R from the same event in which the network sensors experience a signal of strength h. Then, the unit signal for a sensor a will be of the form (in the case that $h > 0$ and the network sensor interacts with

the DM wall prior to the reference device)

$$\bar{\boldsymbol{s}}^a = \left\{0, ..., 0, 1, 0, ...0, -\eta, 0, ..., 0\right\}. \tag{C.2}$$

where $\eta = h^R/h$. In order to calculate the detection statistic mean [Eq. (7.2)] and its variance [Eq. (7.3)], we must utilize the inverse of the covariance matrix from Appendix A. This is given by

$$\left(E^{-1}\right)^{ab}_{jl} = \frac{1}{\sigma^2}\delta_{jl}\left(\delta^{ab} - \frac{1}{N_D}\frac{\xi}{1+\xi}\right), \tag{C.3}$$

where $\xi = N_D\sigma_\times^2/\sigma^2$.

Recall the definition of the template-specific SNR from Eq. (6.9), and suppose that the template repository contains the exact signal that lies in the data stream, i.e., $\bar{\boldsymbol{s}}_i = \bar{\boldsymbol{s}}$ for some signal template in the repository. Then, the detection statistic is given by $\rho = \boldsymbol{d}^T\boldsymbol{E}^{-1}\bar{\boldsymbol{s}}/\sqrt{\bar{\boldsymbol{s}}^T\boldsymbol{E}^{-1}\bar{\boldsymbol{s}}}$. Thus, calculating the expectation value of ρ and its variance will consist of calculating $\bar{\boldsymbol{s}}^T\boldsymbol{E}^{-1}\bar{\boldsymbol{s}}$ as well as $\boldsymbol{d}^T\boldsymbol{E}^{-1}\bar{\boldsymbol{s}}$.

Using the thin wall template from Eq. (C.2), the vector-matrix-vector product is computed as the sum

$$\bar{\boldsymbol{s}}^T\boldsymbol{E}^{-1}\bar{\boldsymbol{s}} = \sum_{a,b}^{N_D}\sum_{j,l}^{J_W}\bar{s}^a_j\frac{1}{\sigma^2}\delta_{jl}\left(\delta^{ab} - \frac{1}{N_D}\frac{\xi}{1+\xi}\right)\bar{s}^b_l,$$

where N_D is the number of network sensors and J_W is the number of epochs in the given time window. Factoring out $1/\sigma^2$, one can sum over the terms separated by the subtraction independently: (1) sum over $\bar{s}^a_j\delta_{jl}\delta^{ab}\bar{s}^b_l$ and (2) sum over $-\bar{s}^a_j\delta_{jl}\frac{1}{N_D}\frac{\xi}{1+\xi}\bar{s}^b_l$. The sum in (1) is just the sum of the squares of all the signal terms from Eq. (C.2) multiplied by the number of devices. Since the signal terms are all zero except at epochs where the satellite clock a and the reference clock R are affected, the sum simplifies immensely. Now, to compute the sum in (2), it must be dissected further. The Kronecker delta δ_{jl} in (2) collapses the sum over l and we split the sum in (2) based on whether the signal terms come from distinct sensors or not

$$-\frac{1}{N_D}\frac{\xi}{1+\xi}\sum_{a,b}^{N_D}\sum_j^{J_W}\bar{s}^a_j\bar{s}^b_j = -\frac{1}{N_D}\frac{\xi}{1+\xi}\left(\sum_a^{N_D}\sum_j^{J_W}(\bar{s}^a_j)^2 + \sum_{a,b\neq a}^{N_D}\sum_j^{J_W}\bar{s}^a_j\bar{s}^b_j\right). \tag{C.4}$$

Notice that the first term on the right side of this equation is the same as (1) above.

Now, since every clock experiences the same signal term at the epoch when the reference sensor interacts with the DM wall (epoch j_R), the second term on the right can undergo yet another dissection

$$\sum_{a,b\neq a}^{N_D}\sum_{j}^{J_W} \overline{s}_j^a \overline{s}_j^b = \sum_{a,b\neq a}^{N_D} \left(\overline{s}_{j_R}^a \overline{s}_{j_R}^b + \sum_{j\neq j_R}^{J_W} \overline{s}_j^a \overline{s}_j^b \right). \tag{C.5}$$

The first term in parentheses on the right side is simply η^2. Since the signal template values at all epochs consists of entirely null values besides the epochs where the individual sensors and the reference sensor interact with the DM wall, the second term on the right is only non-zero when there are disparate sensors that are affected by the DM object at the same epoch. We denote the ratio of network sensors that are close enough spatially to interact with the DM wall within the sampling time interval τ_0 as λ. For GPS, $\tau_0 = 30\,\mathrm{s}$ and at galactic velocities, this "fractional degeneracy" factor $\lambda \approx 0.2$. Ultimately, we arrive at

$$\overline{s}^T E^{-1} \overline{s} = \frac{1}{\sigma^2}\left[(1+\eta^2)N_D - \frac{\xi}{1+\xi}\Big((1+\eta^2) + (N_D - 1)(\eta^2 + \lambda) \Big) \right]. \tag{C.6}$$

For a time window large enough to contain multiple non-overlapping events, it is clear that $\overline{s}^T E^{-1} \overline{s} \to N_E\, \overline{s}^T E^{-1} \overline{s}$, where N_E is the number of events contained within the J_W time window.

To derive $d^T E^{-1} \overline{s}$, we recall that each network sensor data term is the sum of noise (comprised of node sensor noise and reference sensor noise) and a signal [Eq. (C.1)]. Then,

$$d^T E^{-1} \overline{s} = h\overline{s}^T E^{-1} \overline{s} + n^T E^{-1} \overline{s}, \tag{C.7}$$

where the elements of n are given by Eq. (A.2). Since the elements of n are Gaussian random variables with zero mean, the quantity $n^T E^{-1} \overline{s}$ will also be Gaussian distributed with a mean of zero. Furthermore, since $\overline{s}^T E^{-1} \overline{s}$ is a constant given by Eq. (C.6), we find that $d^T E^{-1} \overline{s}$ is Gaussian distributed with a mean of $h\overline{s}^T E^{-1} \overline{s}$ and variance equal to the variance of $n^T E^{-1} \overline{s}$. This implies that the nature of the SNR detection statistic from Eq. (6.9) is Gaussian as well, with the same mean and standard deviation as $d^T E^{-1} \overline{s}$ scaled by $1/\sqrt{\overline{s}^T E^{-1} \overline{s}}$.

Ultimately, using Eq. (C.6), the mean of the SNR when a signal is present is given by

$$\mu_\rho = \frac{h}{\sigma} \sqrt{\frac{N_\mathrm{D}(1 + \eta^2 + \xi) - \xi}{1 + \xi}}$$

$$\approx \frac{h\sqrt{N_\mathrm{D}}}{\sigma} \sqrt{\frac{1 + \eta^2 + \xi}{1 + \xi}}, \tag{C.8}$$

where we dropped the dependence on the fractional degeneracy factor λ. The variance of the SNR,

$$\sigma_\rho^2 = \frac{\mathrm{Var}\{\boldsymbol{d}^T \boldsymbol{E}^{-1} \overline{\boldsymbol{s}}\}}{\overline{\boldsymbol{s}}^T \boldsymbol{E}^{-1} \overline{\boldsymbol{s}}}, \tag{C.9}$$

is computed in Sec. C.1, where we prove that $\sigma_\rho^2 = 1$ in the most general case; this holds, in particular, for our idealized network.

C.1 SNR variance

The goal of this section is to prove that

$$\mathrm{Var}\{\boldsymbol{d}^T \boldsymbol{E}^{-1} \overline{\boldsymbol{s}}\} = \overline{\boldsymbol{s}}^T \boldsymbol{E}^{-1} \overline{\boldsymbol{s}},$$

implying that the SNR variance (C.9) is $\sigma_\rho^2 = 1$. The proof holds regardless of the nature of the covariance matrix - for example it applies to colored noise network with arbitrary cross-node correlations.

Explicitly,

$$\mathrm{Var}\{\boldsymbol{d}^T \boldsymbol{E}^{-1} \overline{\boldsymbol{s}}\} = \left\langle \left(\boldsymbol{n}^T \boldsymbol{E}^{-1} \overline{\boldsymbol{s}}\right) \left(\boldsymbol{n}^T \boldsymbol{E}^{-1} \overline{\boldsymbol{s}}\right) \right\rangle,$$

where $\boldsymbol{n}$ is the intrinsic noise. To streamline notation, we will use Greek letters to index combinations (a, i), where as in Sec. A the first index (a) enumerates the sensors and the second index (i) spans epochs. Then,

$$\left\langle \left(\boldsymbol{n}^T \boldsymbol{E}^{-1} \overline{\boldsymbol{s}}\right) \left(\boldsymbol{n}^T \boldsymbol{E}^{-1} \overline{\boldsymbol{s}}\right) \right\rangle = \left\langle \left(\sum_{\alpha\alpha'} n_\alpha \left(E^{-1}\right)_{\alpha\alpha'} \overline{s}_{\alpha'}\right) \left(\sum_{\beta\beta'} n_\beta \left(E^{-1}\right)_{\beta\beta'} \overline{s}_{\beta'}\right) \right\rangle$$

$$= \sum_{\alpha\alpha'} \sum_{\beta\beta'} \left(E^{-1}\right)_{\alpha\alpha'} \overline{s}_{\alpha'} \left\langle n_\alpha n_\beta \right\rangle \left(E^{-1}\right)_{\beta\beta'} \overline{s}_{\beta'}.$$

By the definition of the covariance matrix, $\langle n_\alpha n_\beta \rangle = E_{\alpha\beta}$. Further, $\sum_\beta E_{\alpha\beta} \left(E^{-1}\right)_{\beta\beta'} =$

$\delta_{\alpha\beta'}$, which reduces

$$\mathrm{Var}\{\boldsymbol{d}^T \boldsymbol{E}^{-1}\overline{\boldsymbol{s}}\} = \sum_{\alpha\alpha'} \overline{s}_\alpha \left(E^{-1}\right)_{\alpha\alpha'} \overline{s}_{\alpha'} = \overline{\boldsymbol{s}}^T \boldsymbol{E}^{-1}\overline{\boldsymbol{s}}\,,$$

as we intended to prove. From Eq. (C.9), it follows that $\sigma_\rho^2 = 1$.

Appendix D

Inverse transform sampling (importance sampling)

Consider a prior probability density function on one of the DM model parameters $p(\theta)$ (e.g., the standard halo model velocity distribution). The cumulative distribution function (CDF) for the prior is defined as

$$C(u) = \int_{-\infty}^{u} p(\theta)d\theta \,. \tag{D.1}$$

We can then define $g(u) = \theta = C^{-1}(u)$. This is particularly useful for sampling from known probability distributions: if u is randomly drawn from a uniform [0:1] distribution, then the $\theta = g(u)$ values will be drawn from the prior $p(\theta)$ distribution. This has the affect of concentrating the sampled points in the regions where $p(\theta)$ is large (and thus naturally reducing the probability of false-positives where $p(\theta)$ is small). Thereby, the priors are taken into account implicitly in the template generation procedure. Note that this just the standard method of inverse transform sampling.

Bibliography

[1] Keith A. Olive and Maxim Pospelov. Environmental dependence of masses and coupling constants. *Phys. Rev. D*, 77:043524, Feb 2008.

[2] B. M. Roberts, G. Blewitt, C. Dailey, and A. Derevianko. Search for transient ultralight dark matter signatures with networks of precision measurement devices using a Bayesian statistics method. *Phys. Rev. D*, 97(8):83009, 2018.

[3] P. Wcisło, P. Morzyński, M. Bober, A. Cygan, D. Lisak, R. Ciuryło, and M. Zawada. Experimental constraint on dark matter detection with optical atomic clocks. *Nature Astronomy*, 1(1):0009, dec 2016.

[4] P. Wcisło, P. Ablewski, et al. New bounds on dark matter coupling from a global network of optical atomic clocks. *Science Advances*, 4(12):eaau4869, 2018.

[5] Gianfranco Bertone, Dan Hooper, and Joseph Silk. Particle dark matter: Evidence, candidates and constraints. *Phys. Rep.*, 405(5–6):279–390, jan 2005.

[6] C. L. Degen, F. Reinhard, and P. Cappellaro. Quantum sensing. *Rev. Mod. Phys.*, 89(3):035002, jul 2017.

[7] M. S. Safronova, D. Budker, D. DeMille, Derek F. Jackson Kimball, A. Derevianko, and Charles W. Clark. Search for new physics with atoms and molecules. *Rev. Mod. Phys.*, 90(2):025008, jun 2018.

[8] Alexander Vilenkin. Cosmic strings and domain walls. *Phys. Rep.*, 121(5):263–315, may 1985.

[9] Sidney Coleman. Q-balls. *Nucl. Phys. B*, 262(2):263 – 283, 1985.

[10] A. Kusenko. Q-ball candidates for self-interacting dark matter. *Phys. Rev. Lett.*, 87:141301, 2001.

[11] Kimyeong Lee, Jaime A. Stein-Schabes, Richard Watkins, and Lawrence M. Widrow. Gauged q balls. *Phys. Rev. D*, 39:1665–1673, Mar 1989.

[12] Jianglai Liu, Xun Chen, and Xiangdong Ji. Current status of direct dark matter detection experiments. *Nature Physics*, 13:212, 2017.

[13] N. Huntemann, B. Lipphardt, Chr. Tamm, V. Gerginov, S. Weyers, and E. Peik. Improved limit on a temporal variation of m_p/m_e from comparisons of yb$^+$ and cs atomic clocks. *Phys. Rev. Lett.*, 113:210802, Nov 2014.

[14] R. M. Godun, P. B. R. Nisbet-Jones, J. M. Jones, S. A. King, L. A. M. Johnson, H. S. Margolis, K. Szymaniec, S. N. Lea, K. Bongs, and P. Gill. Frequency ratio of two optical clock transitions in ^{171}yb$^+$ and constraints on the time variation of fundamental constants. *Phys. Rev. Lett.*, 113:210801, Nov 2014.

[15] 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, Jan 2013.

[16] A. Derevianko and M. Pospelov. Hunting for topological dark matter with atomic clocks. *Nature Physics*, 10(12):933–936, 2014.

[17] Benjamin M. Roberts, Geoffrey Blewitt, Conner Dailey, Mac Murphy, Maxim Pospelov, Alex Rollings, Jeff Sherman, Wyatt Williams, and Andrei Derevianko. Search for domain wall dark matter with atomic clocks on board global positioning system satellites. *Nature Communications*, 8(1):1195, apr 2017.

[18] Joseph D. Romano and Neil J. Cornish. Detection methods for stochastic gravitational-wave backgrounds: A unified treatment. *Living Reviews in Relativity*, 20(1):1–223, 2017.

[19] F. Dong, E. Pierpaoli, J. Gunn, and R. Wechsler. Optical cluster finding with an adaptive matched-filter technique: algorithm and comparison with simulations. *Astrophys. J.*, 676:868–879, April 2008.

[20] P. Shaerer. Global seismic event detection using a matched filter on long-period seismograms. *J. of Geophys. Research*, 99, July 1994.

[21] M. Auger, D.J. Auty, et al. Search for neutrinoless double-beta decay in 136xe with exo-200. *Phys. Rev. Lett.*, 109, July 2012.

[22] B.P. Abbott et al. Observation of gravitational waves from a binary black hole merger. *Phys. Rev. Lett.*, 116, Feb 2016.

[23] Andrei Derevianko. Detecting dark-matter waves with a network of precision-measurement tools. *Phys. Rev. A*, 97(4):042506, apr 2018.

[24] Jo Bovy and Scott Tremaine. On the local dark matter density. *Astrophys. J.*, 756(1):89, 2012.

[25] P. Sikivie. Axions, domain walls, and the early universe. *Phys. Rev. Lett.*, 48:1156–1159, Apr 1982.

[26] W.H. Press, B.S. Ryden, and D.N. Spergel. Dynamical evolution of domain walls in an expanding universe. *Astrophys. J.*, 347:590–604, Dec 1989.

[27] A. Vilenkin and E. P. S. Shellard. *Cosmic Strings and Other Topological Defects*. Cambridge University Press, Cambridge, United Kingdom, 2000.

[28] R. Durrer, M. Kunz, and A. Melchiorri. Cosmic structure formation with topological defects. *Phys. Rep.*, 364(1):1 – 81, 2002.

[29] Alexander Friedland, Hitoshi Murayama, and Maxim Perelstein. Domain walls as dark energy. *Phys. Rev. D*, 67:043519, Feb 2003.

[30] P. P. Avelino, C. J. A. P. Martins, J. Menezes, R. Menezes, and J. C. R. E. Oliveira. Dynamics of domain wall networks with junctions. *Phys. Rev. D*, 78:103508, Nov 2008.

[31] D. Marsh and A. Pop. Axion dark matter, solitons and the cusp–core problem. *Mon. Not. Roy. Astron. Soc.*, 451(3):2479–2492, 2015.

[32] Hsi-Yu Schive, Tzihong Chiueh, and Tom Broadhurst. Cosmic structure as the quantum interference of a coherent dark wave. *Nature Physics*, 10:496, 2014.

[33] C.J. Hogan and M.J. Rees. Axion miniclusters. *Phys. Lett. B*, 205(2):228 – 230, 1988.

[34] Edward W. Kolb and Igor I. Tkachev. Axion miniclusters and bose stars. *Phys. Rev. Lett.*, 71:3051–3054, Nov 1993.

[35] Dorota M Grabowska, Tom Melia, and Surjeet Rajendran. Detecting dark blobs. *Phys. Rev. D*, 98(11):115020, dec 2018.

[36] V. A. Dzuba, V. V. Flambaum, and M. V. Marchenko. Relativistic effects in sr, dy, yb ii, and yb iii and search for variation of the fine-structure constant. *Phys. Rev. A*, 68:022506, Aug 2003.

[37] V. V. Flambaum and A. F. Tedesco. Dependence of nuclear magnetic moments on quark masses and limits on temporal variation of fundamental constants from atomic clock experiments. *Phys. Rev. C*, 73:055501, May 2006.

[38] Etienne Savalle, Benjamin M. Roberts, Florian Frank, Paul-Eric Pottie, Ben T. McAllister, Conner B. Dailey, Andrei Derevianko, and Peter Wolf. Novel approaches to dark-matter detection using space-time separated clocks. feb 2019.

[39] P. Wcisło, P. Ablewski, K. Beloy, S. Bilicki, M. Bober, R. Brown, R. Fasano, R. Ciuryło, H. Hachisu, T. Ido, J. Lodewyck, A. Ludlow, W. McGrew, P. Morzyński, D. Nicolodi, M. Schioppo, M. Sekido, R. Le Targat, P. Wolf, X. Zhang, B. Zjawin, and M. Zawada. New bounds on dark matter coupling from a global network of optical atomic clocks. *Science Advances*, 4(12):eaau4869, 2018.

[40] G. Blewitt. *Treatise on Geophysics*. Elsevier,B.V., Amsterdam, Netherlands, 2015.

[41] Data freely available online: Jet Propulsion Laboratory `ftp://sideshow.jpl.nasa.gov/pub/jpligsac/`.

[42] A. Rollings. Undergraduate Thesis, 2016.

[43] GNSS constellation status online: European GNSS Service Centre `https://www.gsc-europa.eu/system-status/Constellation-Information`.

[44] P Kómár, EM Kessler, and M Bishof. A quantum network of clocks. *Nature Physics*, 10:582, 2013.

[45] P Kómár, T Topcu, E M Kessler, A Derevianko, V Vuletić, J Ye, and M D Lukin. Quantum network of atom clocks: a possible implementation with neutral atoms. *Phys. Rev. Lett.*, 117(6):60506, aug 2016.