



Consulting Agreement Number 08-SC-1042

**Evaluation of
Symbol-Wise Autocorrelation
for the setiQuest Project**

Final Report

August 2011

**Ian S. Morrison
Astro Signal Pty Ltd**



<http://www.astrosignal.com/>

Evaluation of Symbol-Wise Autocorrelation for the setiQuest Project

Final Report

| | |
|------------------------|--|
| Customer: | SETI Institute |
| Customer Reference: | Consulting Agreement No 08-SC-1042 |
| AstroSignal Reference: | ASPL-SETIQUEST-FR |
| Document Name: | Evaluation of Symbol-Wise Autocorrelation for the setiQuest Project – Final Report |
| Submission Date: | 24 August, 2011 |
| Author: | Dr Ian S. Morrison Astro Signal Pty Ltd 23 Lewis Way Newington NSW 2127 Australia |
| Authorised: | Dr Ian S. Morrison [Director] |

© Astro Signal Pty Ltd, 2011

All rights reserved. No part of this document or any information appertaining to its content may be used, stored, reproduced or transmitted in any form or by any means, including photocopying, recording, taping, information storage systems, without the prior permission of Astro Signal Pty Ltd.

Revision History

| Issue | Date | Author | Changes |
|--------------|-------------|---------------|---|
| Draft a | 10-Jun-2011 | Ian Morrison | First complete draft. Ready for internal review. |
| Draft b | 13-Jun-2011 | Ian Morrison | Incorporated internal review comments. Ready for SETI Institute review. |
| V1.0 | 24-Aug-2011 | Ian Morrison | Responded to Jill Tarter's review comments. First release. |
| | | | |

Table of Contents

| | |
|---|-----------|
| REVISION HISTORY | II |
| TABLE OF CONTENTS | III |
| ACRONYMS..... | IV |
| 1. INTRODUCTION | 1 |
| 2. WIDEBAND SETI..... | 3 |
| 2.1 WHY WIDEBAND SETI IS CONSIDERED IMPORTANT | 3 |
| 2.2 CHALLENGES OF WIDEBAND SETI..... | 3 |
| 2.3 DETECTION OF WIDEBAND SIGNALS | 4 |
| 2.3.1 <i>Power Spectral Density</i> | 4 |
| 2.3.2 <i>Energy Detection</i> | 4 |
| 2.3.3 <i>Cyclic Spectral Analysis</i> | 5 |
| 2.3.4 <i>Autocorrelation</i> | 5 |
| 2.3.5 <i>Karhunen-Loève Transform</i> | 6 |
| 2.4 CYCLOSTATIONARITY, AUTOCORRELATION AND ANTIPODAL SIGNALLING | 6 |
| 3. SYMBOL-WISE AUTOCORRELATION..... | 10 |
| 3.1 ALGORITHM DEFINITION | 10 |
| 3.2 EXAMPLES | 11 |
| 3.3 COMMENTS ON DETECTION SENSITIVITY | 11 |
| 3.4 DERIVATION OF SWAC SENSITIVITY | 16 |
| 3.5 SENSITIVITY COMPARISON | 21 |
| 3.6 SWAC REFINEMENTS | 22 |
| 3.7 EFFECT OF CHANNEL IMPAIRMENTS..... | 23 |
| 3.7.1 <i>Doppler</i> | 23 |
| 3.7.2 <i>Interstellar Medium</i> | 24 |
| 4. SWAC COMPUTATIONAL COMPLEXITY..... | 25 |
| 4.1 CONVENTIONAL AUTOCORRELATION – FREQUENCY DOMAIN | 25 |
| 4.2 CONVENTIONAL AUTOCORRELATION – TIME DOMAIN..... | 25 |
| 4.3 SWAC..... | 26 |
| 4.4 COMPARISON..... | 27 |
| 5. CONCLUSIONS AND RECOMMENDATIONS..... | 29 |
| 5.1 CONCLUSIONS | 29 |
| 5.2 RECOMMENDATIONS | 29 |
| 6. REFERENCES | 31 |

Acronyms

| | |
|-----------|---|
| AC | AutoCorrelation |
| ATA | Allen Telescope Array |
| BPSK | Binary Phase Shift Keying |
| E_s/N_0 | Symbol Energy to Noise Density Ratio |
| ET | ExtraTerrestrial |
| GOES | Geostationary Operational Environmental Satellite |
| GPS | Global Positioning System |
| ISI | InterSymbol Interference |
| ISM | InterStellar Medium |
| KLT | Karhunen-Loève Transform |
| METI | Messaging to ExtraTerrestrial Intelligence |
| PSD | Power Spectral Density |
| SETI | Search for ExtraTerrestrial Intelligence |
| SKA | Square Kilometre Array |
| SNR | Signal-to-Noise Ratio |
| SWAC | Symbol-Wise AutoCorrelation |

1. Introduction

This report is prepared by Astro Signal Pty Ltd for the SETI Institute under Consulting Agreement Number 08-SC-1042. It documents the signal processing investigations carried out by Dr Ian S. Morrison for the setiQuest project during the period October 2010 to March 2011.

The setiQuest project is a citizen science initiative of the SETI Institute that aims to engage the public in the analysis and interpretation of radio telescope data obtained from the Allen Telescope Array (ATA). The SETI Institute already has powerful and efficient algorithms to search for and validate narrowband signals of possible extraterrestrial (ET) origin. Such signals appear distinctively on frequency versus time ‘waterfall plots’ as lines with a characteristic slope due to the Doppler effect of the Earth’s motion. However, signals of this type can generally be detected reliably in software without ‘manual’ human intervention. The greater potential benefit of setiQuest was perceived to be able to come from considering more complex wideband signals that could not currently be detected automatically in software. There are also many frequency sub-bands within the ATA data that are congested with terrestrial interference, making the task of discovering potential ET signals (narrowband or wideband) much more difficult. An important objective for setiQuest was therefore to investigate new methods for signal processing and visualisation that would facilitate a human contribution to the detection of ET signals under the more complex scenarios described above.

Research on methods for analysing wideband signals was already underway within the SETI Institute during 2010 with the work of Gerry Harp and Rob Ackermann on autocorrelation detection. Prototype software had been implemented that could demonstrate the ability of autocorrelation to reveal characteristic features of modulated wideband signals, including a range of terrestrial interferers and downlink signals from satellites such as the GPS satellite constellation. Although clearly able to detect the presence of strong signals (even multiple overlaid signals), questions remained about the sensitivity of the autocorrelation method in relation to the very weak signals anticipated for any ET sources that might exist.

In 2010 Jill Tarter became aware of the author’s research on wideband SETI and alternative signal detection algorithms. The potential to incorporate some of the author’s ideas in the setiQuest project and also explore new detection methods more generally was the motivation for engaging the author on a short-term consulting contract.

Working with Jill, Gerry and Rob, the author considered various alternative signal processing techniques that might be considered for detection of arbitrary wideband signals and interference sources. The techniques were examined in conjunction with novel ideas for data visualisation, with a view to developing methods that could form the basis for an advanced setiQuest client visualisation tool.

A short way into the contract it was concluded that, after frequency/time waterfall plots, the next natural type of visualisation that should be provided for setiQuest was some form of autocorrelation (AC) waterfall plot, i.e. a two-dimensional plot of the autocorrelation spectrum versus time. In this type of plot each horizontal line represents the autocorrelation spectrum for a given segment of data (shown by intensity) as a function of autocorrelation time delay. The vertical axis reveals changes in the autocorrelation spectrum over time, allowing both static and transient features to be identified. Prototyping of an autocorrelation waterfall client is being pursued by Rob Ackermann and others and is not described in this report.

This report focuses on the underlying signal processing algorithms that generate autocorrelation spectra suitable for display using the AC waterfall tool. In particular it presents and analyses a variation of the autocorrelation method that is more sensitive than conventional autocorrelation for many classes of signal of interest to SETI, and therefore considered to be a valuable complement to conventional autocorrelation for setiQuest. This new method, known as “symbol-wise autocorrelation” (SWAC) was first proposed by the author as part of his PhD research, and its

development and refinement is an area of ongoing research by the author. Some of the content of this report has been taken from a paper by the author that first introduced the SWAC concept [1].

For this specific setiQuest contract the focus was on two activities:

1. Analysing the detection sensitivity of SWAC and comparing against conventional autocorrelation to justify a further investment in its implementation;
2. Developing a prototype high-level language implementation of SWAC for evaluation purposes and to serve as the basis for code to be run in conjunction with the setiQuest AC waterfall visualisation tool.

The SWAC software provided under this contract implements what the author refers to as the “basic SWAC algorithm”. During 2011/12 the author expects to develop several refinements to the algorithm that will increase the detection sensitivity and further extend its advantages over other detection approaches. Once validated, these refinements will be made known to SETI Institute staff so that they can implement performance optimisations by way of incremental software updates.

Beyond conventional autocorrelation and SWAC variations there are numerous other signal processing techniques that could in the future be considered for incorporation into the setiQuest project. Every new visualisation has the potential to help reveal signal features not observable with the current tools. The development and incorporation of SWAC into setiQuest serves as one example, but the hope is that many other techniques will be added over time to further enhance setiQuest’s potential for success.

2. Wideband SETI

This section provides some background to the work described in subsequent sections.

2.1 Why Wideband SETI is Considered Important

There is increasing interest in broadening traditional narrowband SETI to also consider wideband signal formats, such as modulated information-bearing carriers. The possibility of detecting such signals arises in two scenarios:

1. Detection by ‘eavesdropping’ of an ET civilization’s ‘internal’ communications or radar signals;
2. Detection of a deliberately transmitted information-bearing interstellar beacon.

The relative likelihood of success for each of these scenarios will not be discussed here, other than to note that eavesdropping is unlikely to be fruitful beyond distances of a few hundred light-years at best [2]. The much larger number of potentially habitable planets at distances beyond this (such as in the vicinity of the galactic centre) would suggest that a search for beacons emanating from near the galactic centre should have a higher probability of success.

Beacon searches to date have focussed on monochromatic signals. These are assumed easier to generate at high powers, are less affected by dispersion from the interstellar medium (ISM), and highly sensitive receivers on Earth are easier to construct. However, amongst other concerns, a monochromatic beacon has low information content. It conveys just one bit: “you are not alone”. If we assume there may be many ETIs and beacons, any given beacon will typically not be the first detected by any given recipient – in which case there is no incremental value in being told “you are not alone”. Given the high cost to build and operate a galactic-scale beacon, the question must be asked...why would ET invest in a beacon and not send information?

Once we accept there is merit in searching for galactic scale information-bearing beacons, we then need to address the questions: what might such beacons signals look like, and what is the best way to search for them?

2.2 Challenges of Wideband SETI

Searching for wideband signals is perceived to be inherently more difficult than narrowband signals due to the increased dimensionality of the problem. Along with unknown carrier frequency there are additional unknowns of modulation type, rate and alphabet (the waveforms that represent the different symbols in a transmitted sequence). That is, there are more degrees of freedom in signal structure, and it is inherently more difficult to build an optimum receiver when you don’t know the structure of the signal you are looking for.

In his work on applying communications engineering principles to interstellar signalling [3], Messerschmitt has shown that a form of spread-spectrum signalling in which the signal appears like white noise represents a compelling choice of waveform type, since it can be shown that this will maximise immunity to unknown sources of noise and interference at the receiver. This form of signalling can potentially make it extremely challenging to detect the signal without having a knowledge of the spreading sequences involved (which define the modulation alphabet). This is because, without knowing the alphabet, there is no way to benefit from the processing gain associated with de-spreading – and prior to de-spreading the signal could well be below the noise floor of the receiver, or potentially even below the sky noise floor due to the Cosmic Microwave Background.

The temporal dimension represents another degree of freedom. Benford et al [4] have proposed that beacon signals operating on a galactic scale will be observed by receivers as transient sources. This follows from a consideration of galactic beacons as cost-constrained resource-limited systems, which

suggests they would utilise a narrow transmission beam that is spatially swept and would illuminate any given target star system for a limited time (the ‘dwell time’).

Another feature of wideband signals is that they are affected in a more complex way by Doppler and ISM degradations (dispersion, scattering), which further complicates the discovery process.

For all of these reasons there has long been a common perception that wideband SETI has too many degrees of freedom for practical searches. Some have postulated that an information-bearing wideband beacon signal would need to be accompanied by a more easily detectable “attractor beacon” signal (e.g. narrowband or pulsed) to draw attention to the information-bearing signal.

However, the concerns raised about the detectability of wideband signals have often overlooked the fact that wideband signals can be deliberately selected (or constructed) so as to aid detection. Specifically, signals that possess the property of *cyclostationarity* are amenable to signal analysis techniques that can make detectable features more visible. Cyclostationarity is discussed further in Section 2.4.

Another misconception concerning wideband signal detection is that there must be redundancy present in the signal, i.e. that there needs to be repetition of signal waveforms and/or information content. As will be shown in subsequent sections, this is not the case. We show that for cyclostationary signals it is enough for the signal to possess repetition in *structure* and not content. The SWAC algorithm that we describe in Section 3 exploits the ‘autocorrelation signature’ inherent to cyclostationary signals to simplify the detection problem. It obviates the need for knowledge of the specific modulation method and can be used to detect a wide range of modulation alphabets, including high-dimensionality spread-spectrum alphabets. SWAC reduces the search problem to essentially a single search dimension – the same dimensionality as searching for a narrowband signal of unknown frequency. Reliable detection can be achieved even at low signal-to-noise ratio (SNR), meaning there may be no need for a separate attractor beacon; the information-bearing signal itself serves as the beacon.

2.3 Detection of Wideband Signals

2.3.1 Power Spectral Density

Perhaps the most obvious method to detect wideband signals is to examine the power spectral density (PSD) of the band of interest, in the same way that is employed for the detection of narrowband signals. The PSD shows the distribution in the frequency domain of the measured power of [signal+noise]. If the target signal is significantly more powerful than the surrounding noise, the spectrum of the modulated carrier will be clearly visible above the noise level in the PSD. However, a more reasonable assumption for wideband SETI is that the target signal’s power density is low and that its spectrum will not appear clearly above the noise level, as illustrated in Figure 1.

2.3.2 Energy Detection

Similar to the PSD approach, energy detection involves integrating the [signal+noise] power over a given timespan to obtain a low-variance measure of the total energy. If the target signal is transient in nature, the measured energy will vary between [noise-only] (when the signal is not present) and [signal+noise] (when the signal is present). If the measurement variance is low enough, even very low signal energies may induce a sufficiently large measurement variation to determine if a signal is present, with an acceptable probability of false alarm. However, at least for key wideband signal classes of interest for beacons, we will see in a later section that the detection sensitivity of an energy detector is fundamentally lower than that of an appropriately designed autocorrelation detector.

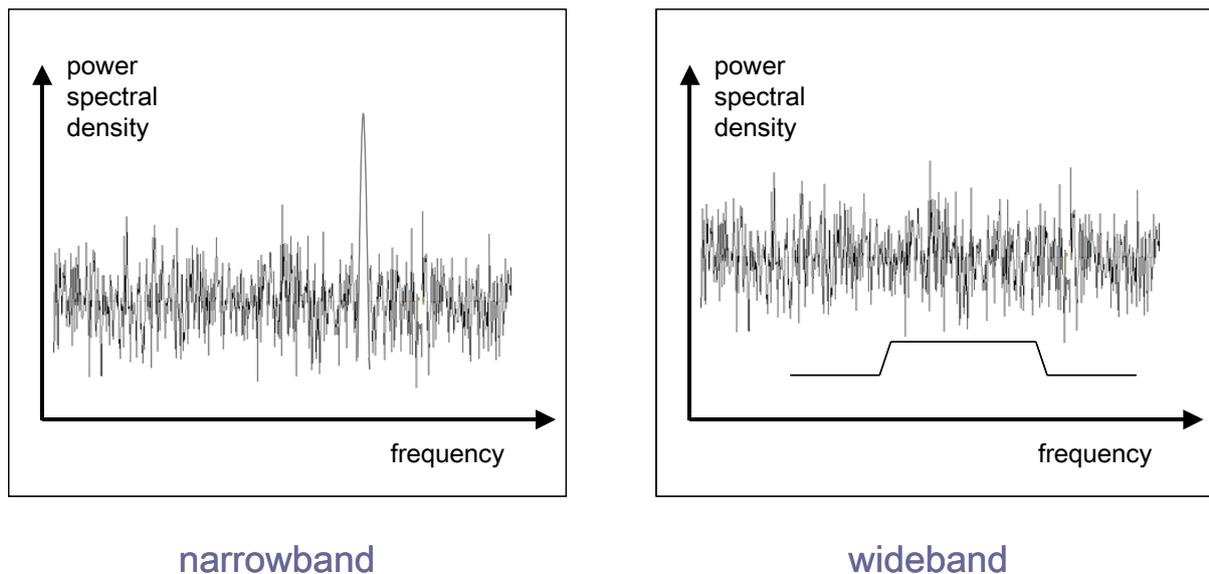


Figure 1: Illustrative power spectral densities for signals embedded in noise. For the wideband case the signal level is below the noise and will not be directly visible.

It is worth noting here that energy detection is an efficient way to detect narrowband signals because the energy is concentrated in a narrow bandwidth, allowing the noise entering the detector to be minimised. It is not efficient for spread-spectrum signals because of the much higher detector noise bandwidth. The optimum detector for any signal (narrowband or wideband) is a ‘matched filter’, which by definition requires knowledge of the transmitted waveform. In the case of a spread-spectrum receiver, the processing gain from de-spreading overcomes the high noise bandwidth of the receiver – but you need to know the spreading codes to perform de-spreading. If you know the spreading codes, the detector SNR is given by $SNR_{MF} = E_s/N_0$ where E_s is the average energy per modulation symbol and N_0 is the noise power density. If you don’t know the spreading codes and rely on energy detection, the detector SNR is given by $SNR_{ED} = (1/2WT_s)(E_s/N_0)^2$ where T_s is the symbol period and W is the measurement bandwidth. Detector performance will be poor for low E_s/N_0 (due to the squaring operation) and large W . However, it is possible to improve SNR_{ED} by averaging over a longer measurement time T , giving an improvement proportional to T .

2.3.3 Cyclic Spectral Analysis

The detection of a wideband signal is made more difficult if its PSD is relatively flat across the signal bandwidth and contains no discrete spectral lines. This is generally the case when carriers are modulated using power-efficient modulation schemes, to avoid wasting energy on signal components that do not carry information. In his work on cyclostationarity, Gardner [5] points out that signals not having discrete spectral lines in their PSD may, however, possess a second-order periodicity, which means that if a nonlinear process is applied to the signal, discrete spectral lines will be regenerated. In a generalisation of Fourier spectral analysis for periodic signals, Gardner has developed the concept of the ‘Cyclic Spectrum’ of a cyclostationary signal [6]. This method produces a 2-dimensional spectrum (in the axes of frequency and delay) that preserves the phase information of a signal. It can be used to obtain discrete spectral features that are not evident in a Fourier-generated spectrum. These features may, under some circumstances, be more easily distinguished from the noise than with a traditional PSD.

2.3.4 Autocorrelation

Autocorrelation can simply be defined as the correlation of a waveform with a delayed version of itself. The potential to apply autocorrelation methods in SETI was recognised as early as 1965 by Drake [7]. More recently Harp et al [8] have discussed a signalling method that can be effectively detected by means of autocorrelation. Both these examples consider a scenario where more than one

signal is superimposed in either time or frequency, with autocorrelation used to detect the presence of repetition.

The SWAC algorithm is also autocorrelation-based but it differs from conventional autocorrelation by taking account of assumed symbol boundaries in a modulated signal – hence the name *symbol-wise autocorrelation*. With SWAC, successful detection is not conditional on explicit repetition of (or within) target signals, so it can be used to detect a single signal in isolation. It exploits structural properties of a cyclostationary signal and there is no requirement for any repetition of the *information content* of the signal. This feature is attractive in a SETI context because it allows the possibility of signal discovery from any captured segment of an extraterrestrial signal without there needing to be explicit repetition of that segment or the content within it.

SWAC has a certain similarity to Cyclic Spectral Analysis in the way that time-segments of the signal waveform are repeatedly ‘folded’ on themselves as part of the process. However, SWAC generates discoverable features in a more direct way than Cyclic Spectral Analysis and, for the types of signals of interest to SETI, is expected to provide superior detection sensitivity¹.

It is difficult to compare the detection sensitivity of autocorrelation with energy detection because the behaviour of an autocorrelation detector is waveform-dependent. For example, for many randomly modulated signals conventional autocorrelation may yield worse sensitivity than energy detection. However, as we shall see in Section 3.4, SWAC can be shown to be superior to energy detection by some margin for signals of this type. The performance of SWAC is still sub-optimal compared to matched filtering, but unlike matched filtering it is not necessary to know the spreading codes. So for situations where the spreading codes are unavailable or cannot be deduced, SWAC could be the next best option. With enough measurement time, a SWAC detector may be sensitive enough to get the job done.

2.3.5 Karhunen-Loève Transform

The Karhunen-Loève Transform (KLT) is an algorithm capable of being used to detect the presence of signals of arbitrary unknown structure that are embedded within noise. It performs an orthogonal linear transformation of the [signal+noise], using an eigenvalue/eigenfunction computation to determine the optimal transformation axes for bringing the signal component ‘into view’. The KLT is computationally very complex but has moved from a theoretical curiosity to a potentially practical signal processing tool in recent years as real-time computing capabilities have increased. A good description of the KLT and the recent advances in its implementation can be found in Maccone’s text [9]. However, despite these developments there remain concerns about the ability to process radio telescope data in real time with the KLT to make an initial detection. Perhaps more realistic in the near-term is that the KLT will prove to be a powerful tool to analyse candidate ‘events’ once they have initially been discovered by other means.

The potential of the KLT for discovering wideband ET signals is left to others to explore. Instead, this report focuses on the SWAC autocorrelation-based algorithm, which is conceptually and computationally less complex than the KLT. While SWAC may not be applicable to any arbitrary signal, it should be effective for all cyclostationary signals; a class that includes most modulation schemes that are considered attractive for interstellar communications.

2.4 Cyclostationarity, Autocorrelation and Antipodal Signalling

As noted earlier, autocorrelation involves correlating a waveform with a delayed version of itself. By computing the degree of correlation over a range of delay values, one can generate an ‘autocorrelation spectrum’ that essentially shows how ‘self-similar’ a waveform is over time. If there exist any repeating patterns in the waveform, the autocorrelation spectrum will display peaks at the delays corresponding to the time separations between the repeated elements. As an example, if a waveform

¹ A detailed comparison of SWAC and Cyclic Spectral Analysis is the subject of ongoing research by the author.

happened to consist of a contiguous sequence of duplicate waveform segments of length T_w , its autocorrelation would display a peak at delay T_w (and integer multiples of T_w).

While it is easy to see how a repetitive redundant signal can be detected with autocorrelation, it is less obvious that a recognisable autocorrelation signature can also be exhibited by signals where there is no redundant repetition. Autocorrelation can also reveal the presence of signals that contain some form of periodicity in their *structure*, with no requirement for repetition of the *content* of the signal².

Signals that exhibit structural periodicity (or whose statistical properties vary cyclically with time) are referred to as *cyclostationary*. This class encompasses virtually all digital modulation methods used in terrestrial communications systems. Any modulation approach that involves sending a sequence of symbols with a common symbol period T_s and chosen from a finite symbol set (alphabet) will display some degree of cyclostationarity. Even when specific symbol values in the sequence are selected randomly, over a sufficient length of time the finite alphabet ensures cyclostationarity. Specifically there exists periodicity in time T_s and so the autocorrelation of such a signal will display peaks at delay T_s and its multiples. However, the strength of the autocorrelation will depend on the size of the symbol alphabet and the distribution of symbol values in the waveform sample being analysed.

Consider an arbitrary sequence of contiguous symbols, $s_k(t)$, as depicted in Figure 2. We assume the $s_k(t)$ are chosen from a finite alphabet.

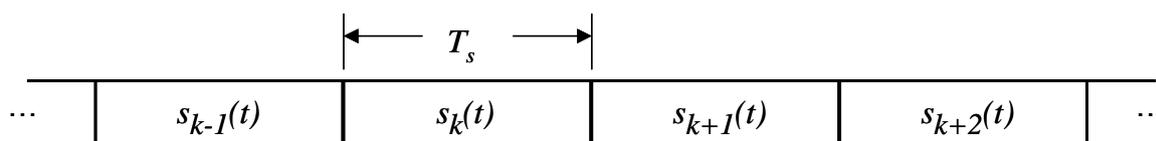


Figure 2: A Sequence of contiguous modulation symbols, $s_k(t)$

At delay T_s the autocorrelation process is effectively measuring the degree of similarity between *consecutive* symbols. It is easy to see that the larger the symbol alphabet, the lower will be the average degree of similarity between consecutive symbols. Conversely, the average degree of similarity will be maximised with the smallest symbol alphabet: a binary alphabet. Even if the two possible symbol values are completely orthogonal (zero cross-correlation), for randomly selected symbol values there will be maximum correlation 50% of the time, when the adjacent symbols happen to be the same value. In general the cross-correlation between different members of a symbol set will not be zero, so this represents just one particular scenario. However, it illustrates how autocorrelation sensitivity is, in general, maximised with a binary alphabet.

Antipodal signalling is a subclass of binary cyclostationary signalling in which the two members of the symbol alphabet are the inverse of one another. This alphabet can be represented as $[A, -A]$. Binary data can be mapped to this signal set by assigning one symbol type to represent 0 and the other to represent 1.

The simplest example of an antipodal signal set is Binary Phase Shift Keying (BPSK), which can be described as signalling with the alphabet $[1, -1]$. The BPSK signal set is illustrated in Figure 3.

² By “content” here we mean the information that selects the specific sequence of symbols that makes up the transmitted waveform, i.e. the information that controls the modulation of the signal.

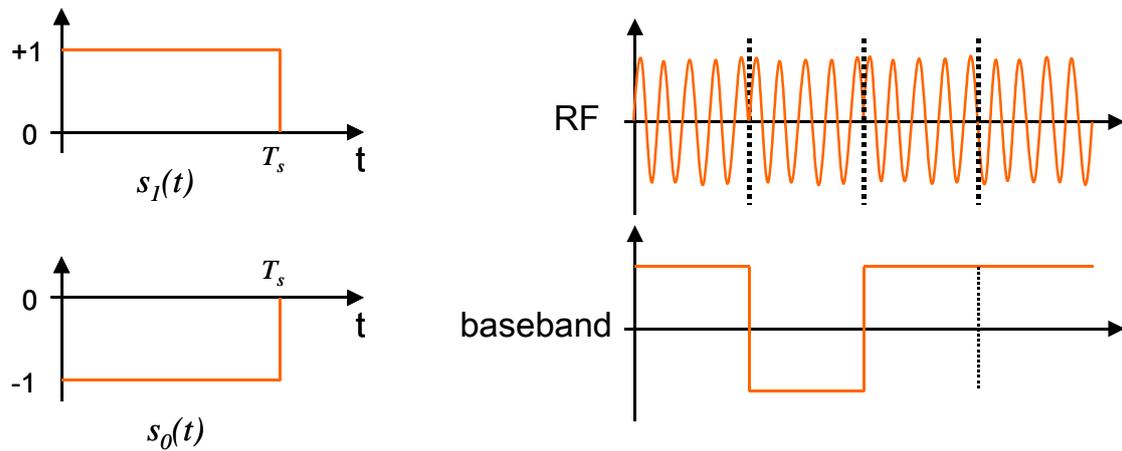


Figure 3: BPSK signal set, with example waveforms at RF and baseband

However, an antipodal signal set can potentially be much more complex than BPSK, with symbols having higher dimensionality (e.g. multiple chips per symbol³) and with chips varying in amplitude, phase or even width. The antipodal constraint means that the waveform representing one alphabet member is precisely the negative of the other member at every point in the complex waveform representation (i.e. a point-wise 180° complex rotation of the first waveform). An example is shown in Figure 4; in this case a form of spread-spectrum BPSK.

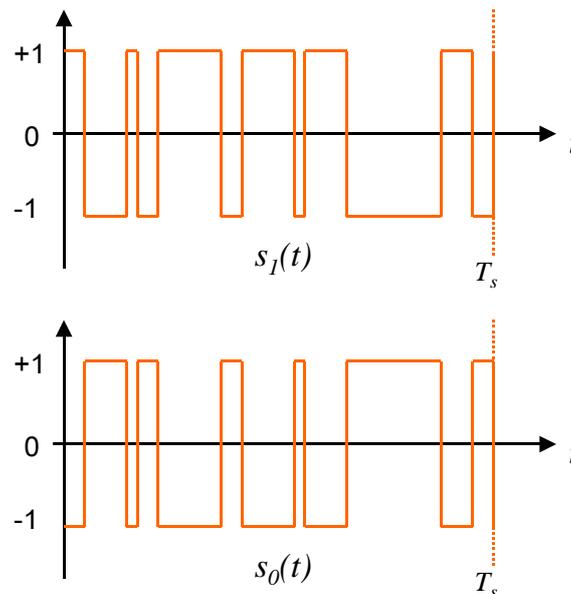


Figure 4: An illustrative spread-spectrum BPSK signal set – an example of a high-dimensionality binary antipodal alphabet

The key feature of an antipodal signal set is that, regardless of the specific waveform shape for one alphabet member, the correlation of any one symbol with any other will always give either 1 (if they are the same symbol) or -1 if they are different symbols. This also means that when successive symbols are transmitted on a channel, the correlation between adjacent symbols at the receiver will be 1 or -1 (following normalisation and ignoring noise and phase rotations for the present). This

³ For our purposes we define a ‘chip’ as a single signalling unit in the two-dimensional complex plane, i.e. a single phase/amplitude over the smallest modulation interval; the chip interval T_c . A higher dimensional symbol can be obtained by transmitting multiple chips over the symbol period T_s .

immediately suggests an autocorrelation process on the received signal could be used to reveal the presence of the modulation.

If we consider the autocorrelation of a sequence of randomly selected antipodal symbols, then at delay T_s we see interesting behaviour. Over time the autocorrelation score averages to zero. This is because on average 50% of adjacent symbols are fully correlated and the other 50% are fully uncorrelated (i.e. produce a negative correlation score). On the face of it, this suggests autocorrelation would not be an effective method for detecting antipodal signals in noise, which would be disappointing given the known power efficiency advantages of antipodal modulation. However, the autocorrelation behaviour of antipodal signal sets can be turned to an advantage when one recognises that *every* adjacent symbol pair produces a maximum magnitude correlation score, albeit a mix of positive and negative values. If the *absolute value* of each symbol pair's correlation score is accumulated over the symbol sequence then it can be seen that in fact antipodal signalling maximises the autocorrelation peak produced by the signal at delay T_s . This is the motivation behind the SWAC algorithm presented in Section 3.

As mentioned in Section 2.2, there are persuasive arguments in favour of utilising spread-spectrum modulation for interstellar signalling. We have already shown an example of such a signal set, for the antipodal case, in Figure 4. How does this type of signal set behave with autocorrelation? As explained above, any antipodal signal set will result in ± 1 correlation scores between adjacent symbols, i.e. for an autocorrelation delay of T_s . What does change for the spread-spectrum case is the behaviour at other values of delay. Assuming the pseudo-noise sequence used for the spreading process has been selected appropriately⁴, there will be a large reduction in the correlation score for sample delays that result in one or more chip intervals of time offset. This means that an autocorrelation spectrum for a spread-spectrum signal will display a sharper peak at T_s than the non-spread case. This helps to make the autocorrelation peak easier to distinguish amidst high levels of noise. In this way the use of spread-spectrum is highly beneficial for signal detection using autocorrelation methods.

The idea of using autocorrelation to detect unknown spread-spectrum signals is not new (e.g. [10]). However, previous approaches have not taken account of symbol boundaries and do not gain the benefits that arise from a symbol-wise approach to detecting antipodal spread-spectrum signals. The SWAC algorithm provides superior detection sensitivity in comparison with previous approaches, with a manageable increase in computational complexity (as will be seen in Section 4).

⁴ A 'good' pseudo-noise bit sequence in this context is one for which there is low autocorrelation at all non-zero bit delays.

3. Symbol-Wise Autocorrelation

This section describes the SWAC algorithm and derives its detection sensitivity. Comparisons are made with matched filtering and energy detection.

3.1 Algorithm Definition

As explained in Section 2.4, the effectiveness of conventional autocorrelation over multiple symbols in a random symbol sequence is limited by the averaging of positive and negative correlation scores. However, if the absolute value of the correlation score from each symbol pair is accumulated⁵, the signal component of each correlation adds coherently, and the correlation peak at T_s is maximised.

The challenge in SETI is that, for a received waveform $y(t)$, we do not know the symbol boundaries (if indeed a modulated signal is present), nor the symbol period, T_s . Furthermore, we do not know the signal alphabet, $[S, -S]$, nor the centre frequency of the modulated carrier.

However, if there was a signal component in $y(t)$ that happened to be modulated using an antipodal signal set, we can exploit the characteristics of antipodal signalling to reduce the search space dramatically. For a given segment of $y(t)$ we can perform a search over the symbol-period dimension (over variable τ) without knowing the carrier frequency or signal alphabet. We can make progressive calculations of the autocorrelation across a range of delays corresponding to the minimum symbol period τ_1 to maximum symbol period τ_2 under consideration in the search. For each trial τ we correlate an assumed sequence of noisy symbols $y(t)$ with a one-symbol-delayed version of $y(t)$ (i.e. $y(t+\tau)$), accumulating the absolute value of each symbol-by-symbol correlation score. If a signal is present then at τ close to T_s the autocorrelation score will peak at value D_{peak} . At other τ values the misalignment of symbol periods will produce a low average autocorrelation score. A signal is deemed to be present if D_{peak} exceeds a specified threshold, which is set relative to the mean autocorrelation score.

We call this algorithm *symbol-wise autocorrelation* (SWAC), which, in its discrete-time form, is expressed mathematically in equations (1), (2) and (3).

$$SWAC(\tau) = \sum_{n=1}^M \left| \sum_{k=k_0+(n-1)\tau}^{k_0+n\tau} (y_k \cdot \bar{y}_{k+\tau}) \right| \quad (1)$$

$$D = \max_{\tau \in [\tau_1, \tau_2], k_0 \in [0, \tau]} SWAC(\tau) \quad (2)$$

$$\hat{T}_s = \arg \max_{\tau \in [\tau_1, \tau_2], k_0 \in [0, \tau]} SWAC(\tau) \quad (3)$$

In Equation (1), y_k are the complex samples of waveform $y(t)$, M is the number of symbols processed, τ is the trial symbol period, and k_0 is the sample index corresponding to the first sample of each symbol. Equation (3) gives us the estimated symbol period of the signal embedded in $y(t)$, which will be useful for any subsequent processing to extract the information content of the signal.

⁵ It is undesirable to take the absolute value of every *sample-wise* correlation because this will result in the noise energy adding incoherently. It is better to accumulate complex sample-wise correlation scores over the duration of a complete symbol - producing a *symbol-wise correlation score* - then take the absolute value before combining with the scores from other symbol-wise correlations. This will provide the maximum degree of noise averaging without compromising the signal component, thus maximising detection sensitivity.

It is worth emphasising that the search is over τ and k_0 . One does not need to know the centre frequency, chip rate (bandwidth) or symbol alphabet (spreading codes).

A variation on Equation (1) that provides a worthwhile gain in detection sensitivity can be obtained by taking the absolute value of just the real component of each complex symbol-pair correlation score, as shown in Equation (4). This optimisation is only possible if the arbitrary phase shift between symbols in passband is successfully estimated and removed⁶, which should be possible when a sufficient number of symbols are available to process (i.e. $M > \sim 20$).

$$SWAC(\tau) = \sum_{n=1}^M \left| \operatorname{Re} \left\{ \sum_{k=k_0+(n-1)\tau}^{k_0+n\tau} (y_k \cdot \bar{y}_{k+\tau}) \right\} \right| \quad (4)$$

Note that SWAC can be used to detect any cyclostationary signal, but the algorithm achieves maximum sensitivity when the alphabet is of the binary antipodal form. An expression for the detection sensitivity is derived in Section 3.4 for the binary antipodal case and assuming the optimised formulation of Equation (4). This represents the best-case scenario, which is useful to understand. It is also mathematically tractable, unlike cases where the cross-correlations between alphabet members are unknown. The degradation in sensitivity when detecting other modulation alphabets varies on a case-by-case basis and cannot easily be generalised. However, if we restrict our attention to binary spread-spectrum alphabets, then it can be shown that the sensitivity of SWAC will fall somewhere between the best-case figure and 6 dB below that, depending on the specifics of the alphabet⁷.

3.2 Examples

We illustrate the SWAC algorithm by way of an example. Assume an antipodal spread-spectrum BPSK waveform with a symbol rate of 2 symbol/s and a chip rate of 1000 chips per symbol. The PSD of a passband representation of this signal (centred at approximately 2.5 kHz) is shown in Figure 5. Applying the SWAC algorithm to a 50 second burst of this waveform generates the autocorrelation spectrum of Figure 6, which shows a very strong peak at the correct symbol period of 500 ms.

Ignoring channel impairments, we now assume this signal is received embedded in noise at a low SNR such that the noise masks the presence of the signal in the PSD, as shown in Figure 7. Applying the SWAC algorithm to a 50 second burst of the received waveform generates the autocorrelation spectrum of Figure 8, which still shows a clear peak at the symbol period 500 ms. The SWAC process has achieved this detection without knowledge of the centre frequency, symbol rate, chip rate, signal bandwidth, modulation method or symbol alphabet.

3.3 Comments on Detection Sensitivity

In Section 3.4 we derive a precise mathematical formulation for the detection sensitivity of SWAC. It is worth noting here that the sensitivity increases proportionally with the time-span of signal processed. At a given symbol period this is the same as saying the sensitivity is proportional to the number of symbols processed, M .

⁶ In the general case on a passband channel there will be an arbitrary number of cycles of the carrier during a symbol interval, hence the complex correlation between successive symbols will result in a complex score with the maximum magnitude and arbitrary unknown phase P (for like-valued symbols) or $(P+180^\circ)$ (for dissimilar symbols). Given a sufficient number of correlations in relation to the level of noise present, it should be possible to obtain a reasonably accurate estimate of P . If so, it can be removed by appropriately rotating each complex correlation score, placing all results on the Real axis (plus complex noise).

⁷ The worst-case performance for a binary alphabet occurs when detecting binary orthogonal signalling, where the cross-correlation between the two alphabet members is zero. This example is discussed further in Section 3.4 where it is shown that the detection sensitivity is 6 dB worse than the binary antipodal case. Alphabets having non-zero cross-correlation between members will experience less sensitivity loss.

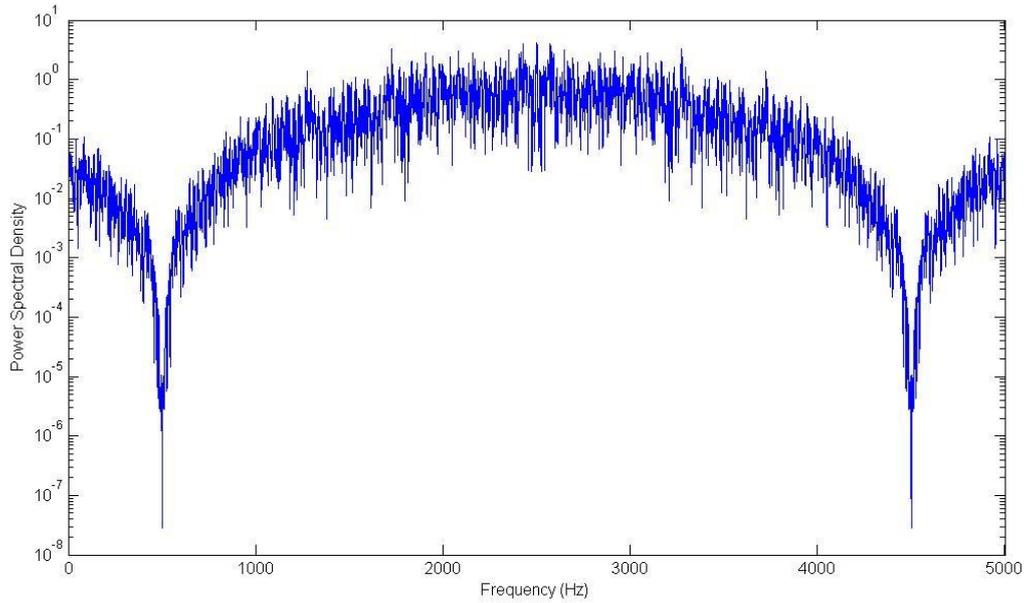


Figure 5: PSD for an illustrative antipodal spread-spectrum BPSK signal (2 symbol/s, 1000 chips/symbol, no noise)

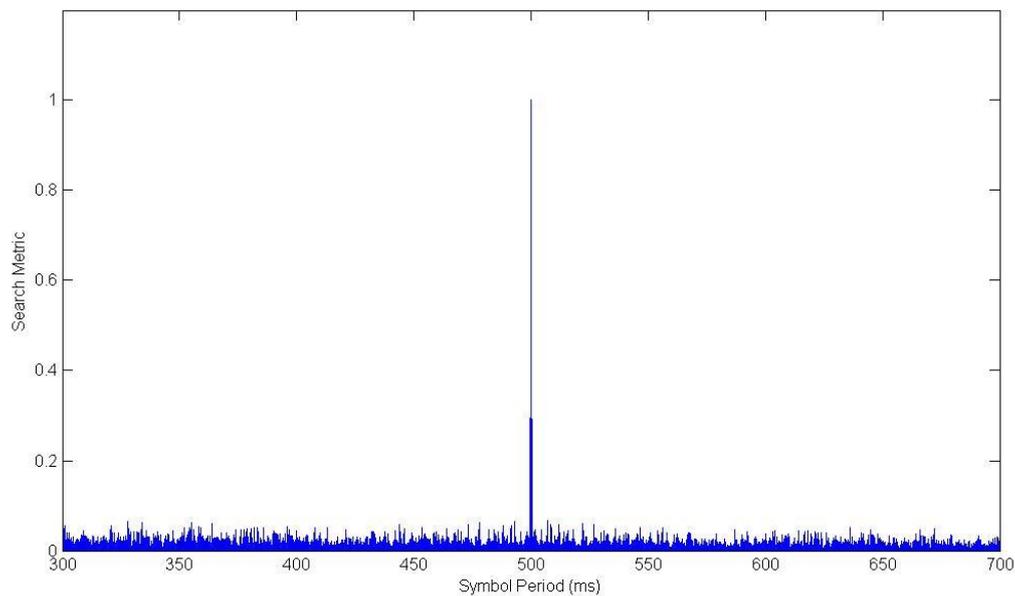


Figure 6: SWAC output as a function of assumed symbol period τ , with the waveform of Figure 5 as input

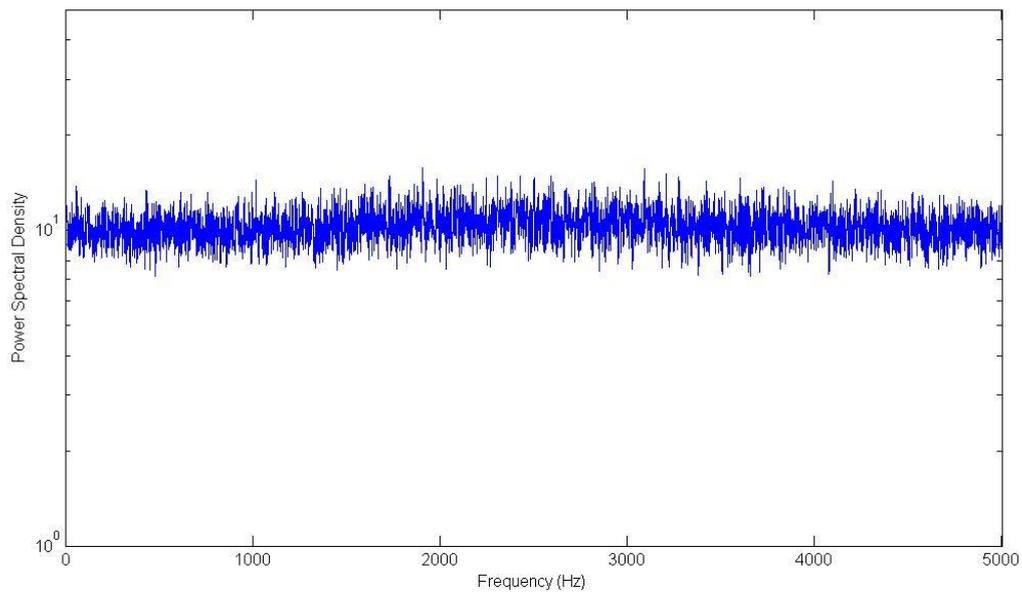


Figure 7: PSD for the illustrative signal of Figure 5 embedded in Gaussian noise (white across the measurement bandwidth)

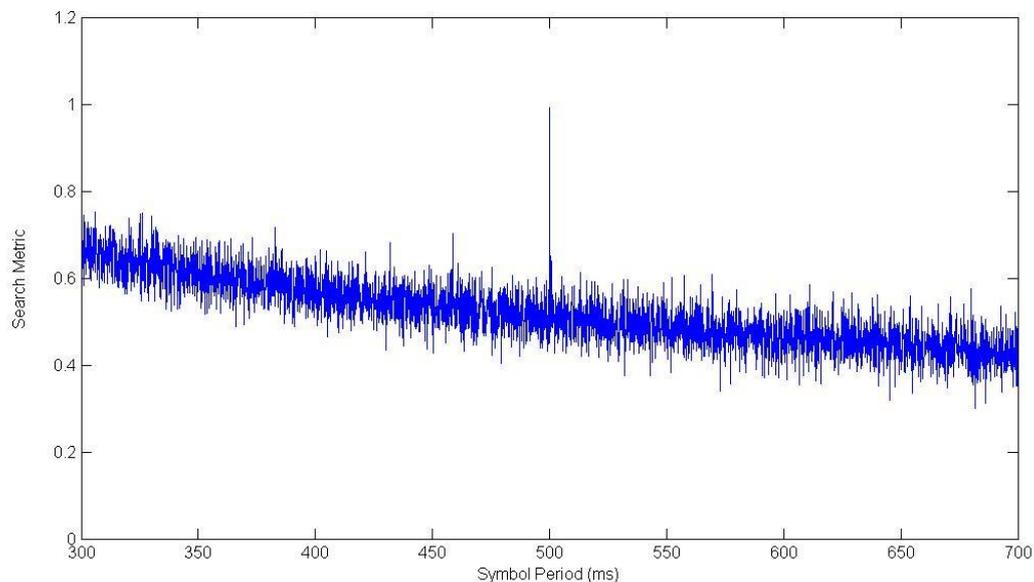


Figure 8: SWAC output as a function of assumed symbol period τ , with the waveform of Figure 7 as input

The SWAC plots of Figure 6 and Figure 8 were obtained using the optimum value of k_0 in Equation (4). In addition to the search in the τ dimension, a search was also conducted over different k_0 . It was found that the SWAC score is relatively insensitive to the k_0 assumption. This is seen in Figure 9, which is a pseudo-three-dimensional plot of the SWAC score as a function of both τ and k_0 (here shown as t_0 ; the starting time offset as a percentage of τ). Ten values of t_0 were tried at each trial τ , in

steps of 10% of τ . There is an optimum value of t_0 (in this case 50%) but, regardless of the value of t_0 there is in all cases a distinct autocorrelation peak at $\tau = 500$ ms.

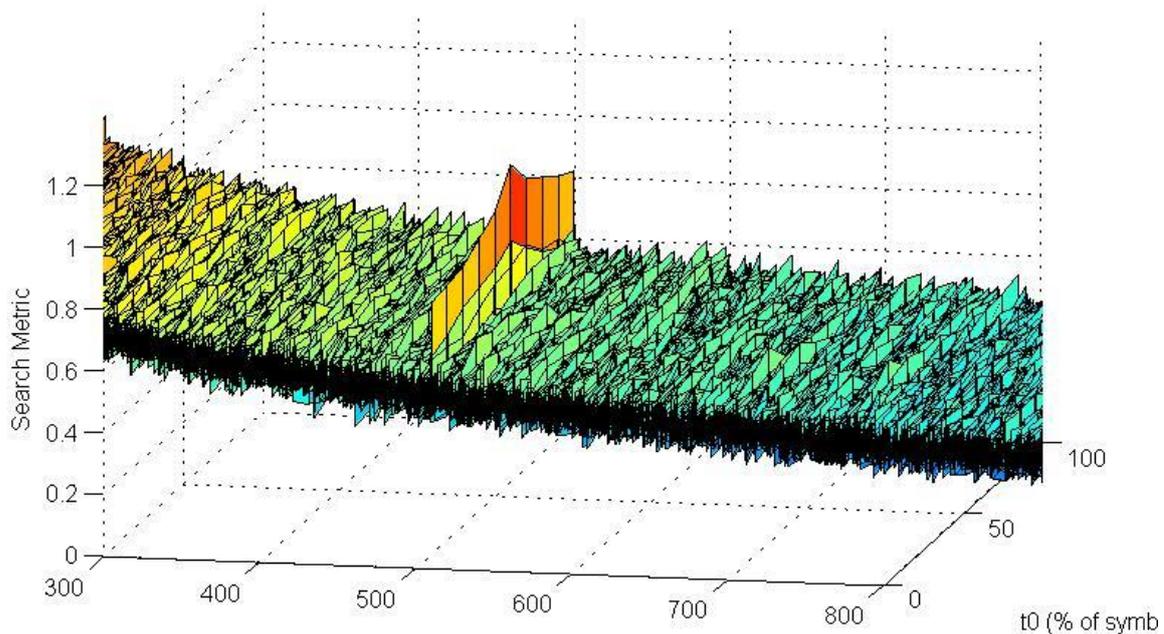


Figure 9: SWAC output as a function of assumed symbol period τ and t_0

Since the SWAC score is relatively insensitive to incorrect k_0 , rather than increase the computational complexity by a factor of 10 to search over k_0 it is actually more productive to set k_0 to zero and increase the amount of data processed to compensate for the incorrect k_0 assumption, i.e. to overcome the reduced level of the SWAC peak at sub-optimal k_0 . This is particularly useful when one realises that, in general, there will not be an exact integer multiple of waveform samples per symbol. Hence over a large M the assumed symbol boundaries will drift with respect to the actual boundaries, regardless of the initial choice for k_0 . Processing a larger M will provide a higher detector SNR and compensate for this effect. It has been found that a factor of two increase in M will overcome most of the loss due to incorrect k_0 , with only a doubling of computational complexity.

Another aspect that affects detection sensitivity is the width of the autocorrelation peak. Spread-spectrum modulations produce a peak that is narrow in the τ axis whereas non-spread modulations produce broader peaks, making the discrimination from noise more difficult, and also the ability to quantify the precise symbol rate at which the peak occurs. This is seen clearly in the example plots shown in Figure 10 for non-spread and spread BPSK signals taken from GOES and GPS satellites respectively. This shows the benefit of using a spread-spectrum form of modulation as far as detection is concerned. This benefit is welcomed to help overcome the noise-bandwidth issue with spread-spectrum modulation (when matched filtering is not able to be used), as will become more clear following the analysis in Section 3.4.

It was explained previously how SWAC is better able to detect randomly modulated signals than conventional autocorrelation. This can be seen clearly in the example shown in Figure 11. Here the same noisy signal of length 20 symbols is analysed using conventional autocorrelation and SWAC. Both methods show a peak at the correct symbol period, but the SWAC output is significantly stronger. Seen another way, SWAC is able to achieve successful detection at lower input SNR values than conventional autocorrelation.

It is interesting to note that if the number of symbols processed is increased, the variance of the noise seen off-peak will reduce, for both the top and bottom plots in Figure 11. However, the peak level in the top plot can be expected to reduce in size as the autocorrelation of the signal component averages closer to zero. By contrast the peak in the bottom plot will not be affected, meaning that SWAC provides increased detection sensitivity when the number of symbols processed is increased.

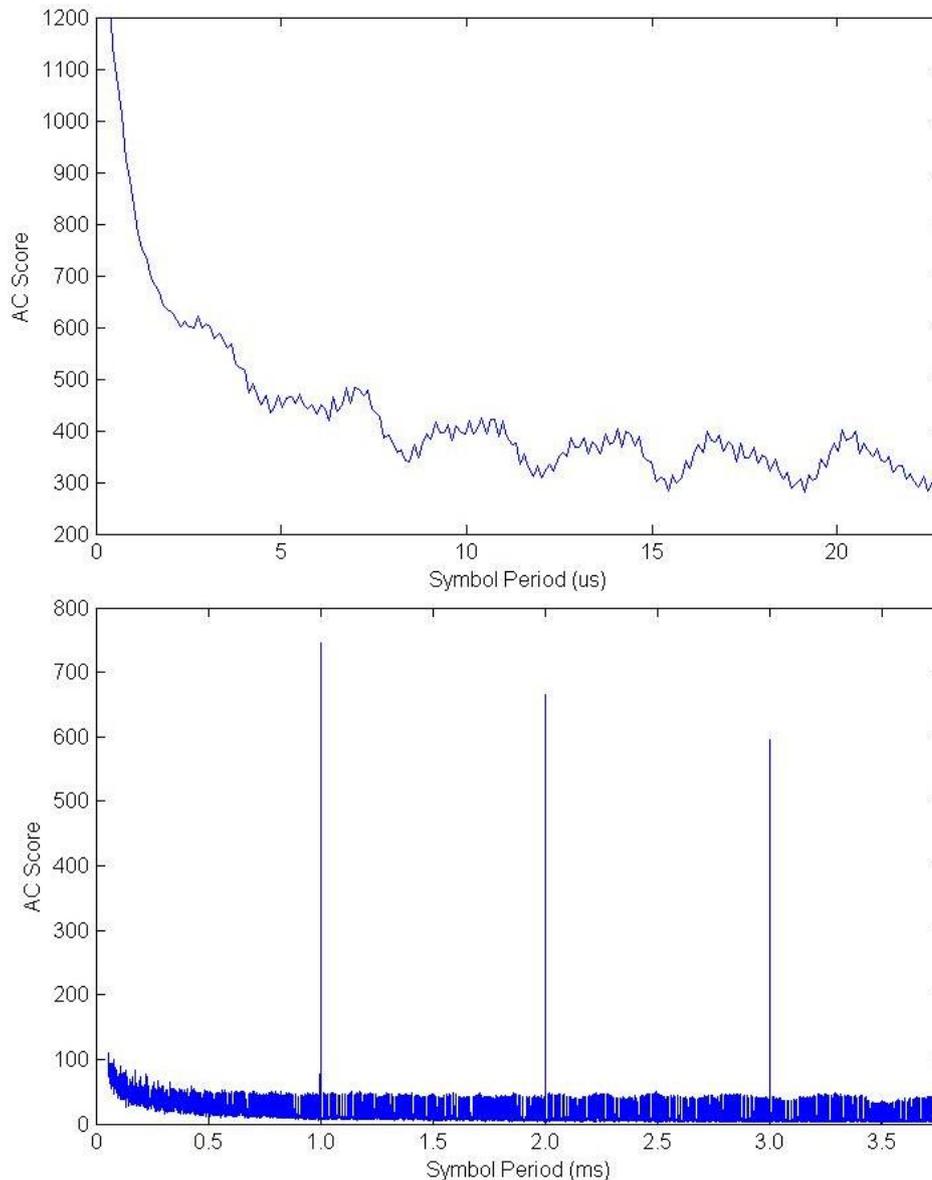


Figure 10: Comparison of SWAC plots for a non-spread BPSK modulation from the GOES satellite (top) and a spread BPSK modulation from a GPS satellite (bottom)

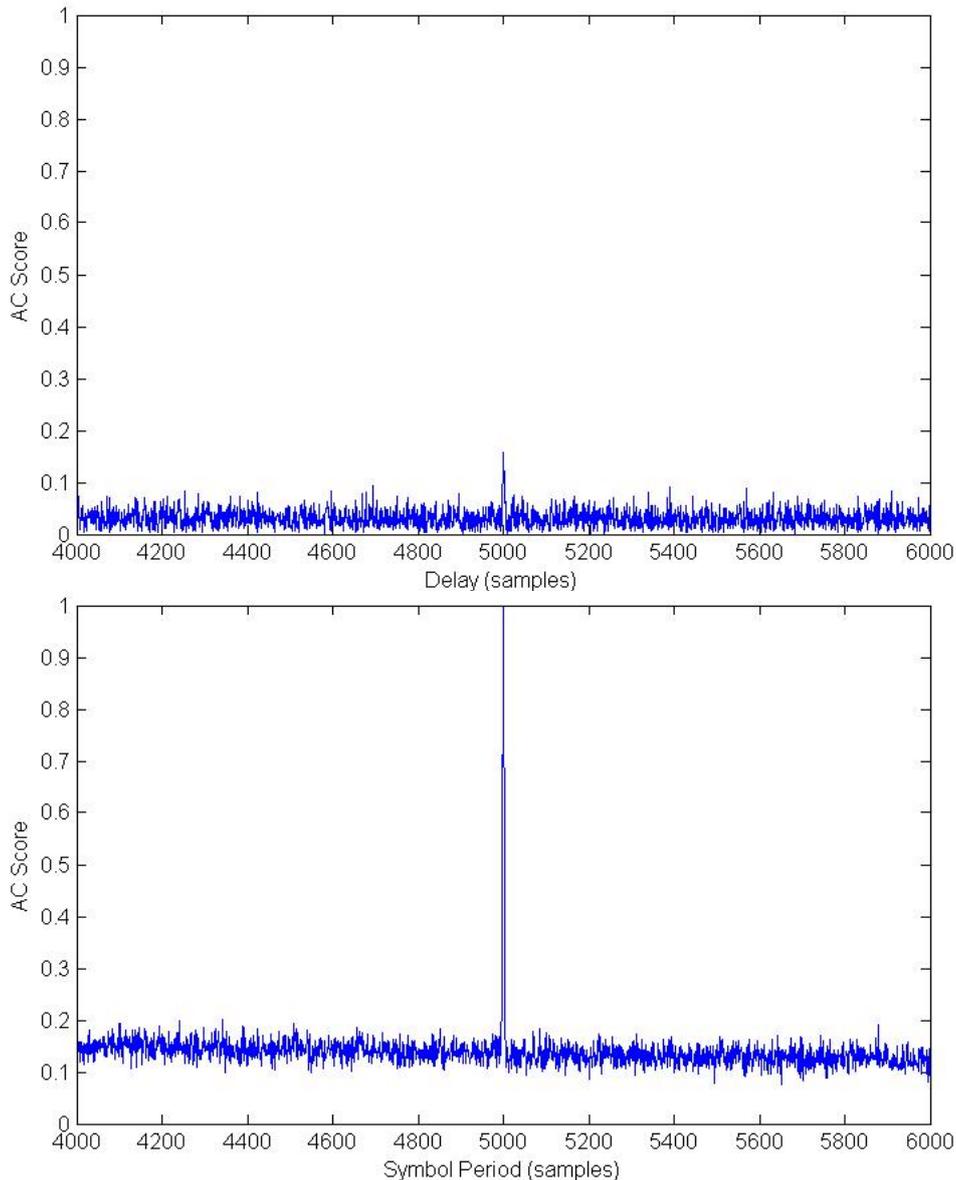


Figure 11: Comparison of conventional autocorrelation (top) with SWAC (bottom) for a spread-spectrum BPSK signal embedded in Gaussian noise with SNR of -10 dB

3.4 Derivation of SWAC Sensitivity

The sensitivity of a detector can be characterised by its output signal-to-noise ratio, SNR_{out} . In terms of evaluating detection and false alarm probabilities, the appropriate definition is given in Equation (5).

$$SNR_{out} = \frac{(E[D_{signal+noise}] - E[SWAC_{noise}])^2}{Var(SWAC_{noise})} \quad (5)$$

Equation (5) holds when the SWAC output has Gaussian statistics, which we show later in this section to be a good approximation. That being the case, it can be shown that the miss and false alarm probabilities are completely determined by the number of standard deviations between the detector output threshold and the expected detector outputs when signal is, respectively, present and not

present. That is, the required miss and false alarm probabilities will be achieved if the difference between $E[D_{\text{signal}}]$ and $E[D_{\text{signal+noise}}]$ equals or exceeds the appropriate multiple of standard deviations.

As an example, consider the case where the desired miss and false alarm probabilities are both to be a maximum of 10^{-3} . To achieve this, $(E[D_{\text{signal}}] - E[D_{\text{signal+noise}}])$ needs to exceed 6.2 standard deviations, with the detector threshold set mid-way between the two expected values⁸. Squaring this figure provides the corresponding SNR, so in this example the required SNR_{out} is ~ 38 (or ~ 16 dB).

It is appropriate to describe the ratio of Equation (5) as an SNR because it takes the form of power over variance, which is consistent with how the SNR of the detector input is defined, as we will see below. A more detailed derivation for Equation (5) can be found in [11].

Note that for spread-spectrum signals we are able to assume that the noise power is much greater than the signal power, so then $\text{Var}(\text{SWAC}_{\text{signal+noise}}) \approx \text{Var}(\text{SWAC}_{\text{noise}})$. This is why we can use $\text{Var}(\text{SWAC}_{\text{noise}})$ for both the miss and false alarm cases.

Note also that the (signal*noise) terms in the correlations in the SWAC algorithm are insignificant compared to (noise*noise) terms and can therefore be ignored.

It is important to recognise that the detector sensitivity is signal dependent. $E[D_{\text{signal}}]$ varies depending on the signal structure, the modulation type and the symbol alphabet. It may also be data-dependent, i.e. dependent on the specific data pattern with which the signal was modulated during the measurement interval.

However, with antipodal modulation, at the value of τ corresponding to T_s , $E[D_{\text{signal}}]$ is time-invariant and known (ignoring channel impairments). We therefore can proceed to formulate an expression for the detection sensitivity for that class of signal, noting that it represents a best-case scenario. The sensitivity for other classes will be below that of antipodal modulation, depending on the degree of correlation between members of the signalling alphabet. For example, consider binary orthogonal signalling, for which the two alphabet members have zero cross-correlation. With random modulation, consecutive symbols will, on average, be the same 50% of the time and different 50% of the time. When they are the same, the SWAC output will be +1; when different it will be zero. Hence, over a span of many symbols $E[D_{\text{signal}}]$ will be half that of the binary antipodal alphabet case. Therefore SNR_{out} will be reduced by a factor of four, i.e. the detection sensitivity will be 6 dB worse than the binary antipodal case.

We proceed with the derivation of the best-case sensitivity, as experienced when detecting binary antipodal signalling. We begin by assuming a signal $s(t)$ that is sampled at rate W and where:

- each sample has the same signal power: s^2
- the total energy in one symbol is: $E_s = s^2 \cdot T_s$

Assume that $s(t)$ is combined with Gaussian noise (white across bandwidth W) of variance σ^2 , i.e. the noise power in each sample is σ^2 . We take [signal+noise] as the input to the SWAC detector. The detector input SNR is given by $\text{SNR}_{\text{in}} = s^2 / \sigma^2$.

The number of samples delay when τ corresponds to T_s is $\tau = WT_s$.

In the numerator of Equation (5) we seek $(E[D_{\text{signal+noise}}] - E[\text{SWAC}_{\text{noise}}])$. However, for the range of useful SNRs (where a signal detection can be successfully accomplished), it happens that

⁸ The Q-function, $Q(x)$, provides the probability that a standard normal random variable will obtain a value greater than x . Since $Q(3.1) \approx 10^{-3}$, the detection threshold should be set at 3.1 standard deviations above the mean detector output when noise only is present for a 10^{-3} false alarm probability. The expected detector output when signal and noise are present will need to be a further 3.1 standard deviations above this threshold to achieve a 10^{-3} miss probability. Hence, in this example, $E[D_{\text{signal+noise}}]$ needs to equal or exceed 6.2 standard deviations.

$(E[D_{\text{signal+noise}}] - E[\text{SWAC}_{\text{noise}}])$ is closely approximated by $E[D_{\text{signal}}]$. This can be verified by conducting a rather complicated derivation of $E[D_{\text{signal+noise}}]$ involving the calculation of ‘raw absolute moments’ and evaluated using ‘generalised hypergeometric functions’. The full derivation is not provided here, but is planned to be published in a future paper by the author. However, in Figure 12 we illustrate the point for an example binary antipodal spread-spectrum waveform. The green curve is $E[D_{\text{signal+noise}}]$ as a function of input SNR; in this case flat at the value of $E[D_{\text{signal}}]$ down to around -20 dB. Below that it starts to rise due to the high noise power present. The blue curve is what happens with the same amount of noise but with the signal removed. What we are concerned with is the height of the green curve above the blue curve, which is plotted with the red curve. At higher SNRs the red and green curves are essentially matched and approximately equal to $E[D_{\text{signal}}]$. But as the SNR drops the gap between the curves reduces. It has dropped to ~ 0.7 of maximum by the time the SNR is down at -15 dB. After squaring, this will mean about 3 dB loss of sensitivity. The sensitivity loss increases quickly as the input SNR drops below -15 dB.

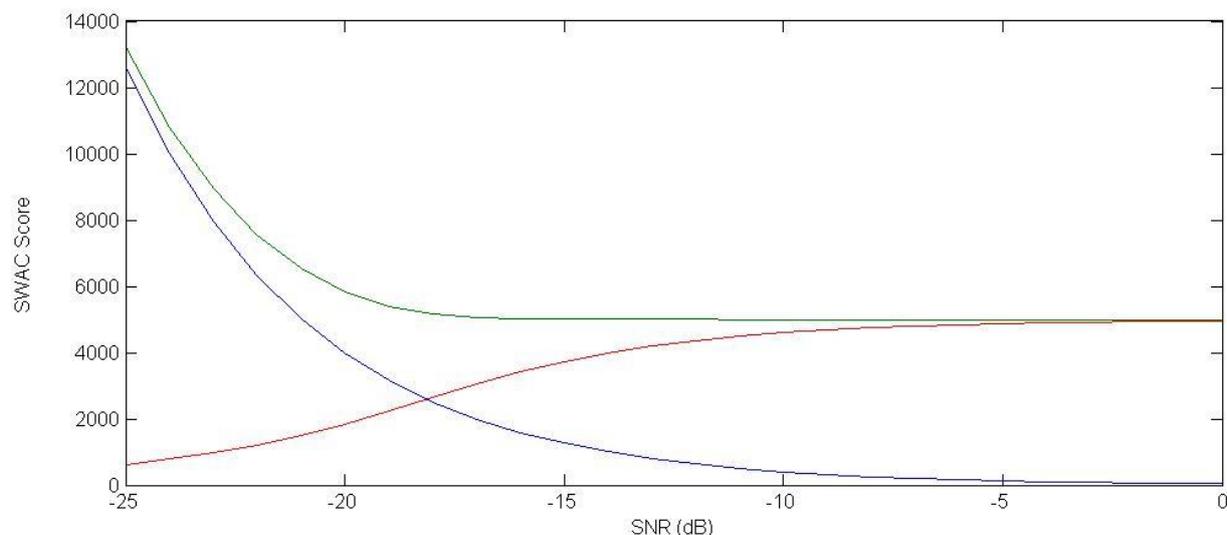


Figure 12: Variation in SWAC score versus input SNR for an illustrative spread-spectrum input signal in Gaussian noise. The green curve is for [signal+noise], the blue curve is for [noise only] and the red curve is the difference between green and blue.

We would, of course, prefer SWAC to maintain its sensitivity down to very low SNRs. The author is currently investigating a refinement to the algorithm that will extend the useful SNR range downwards by several dB. However, it is important to note that at very low input SNRs one starts to run into another limitation anyway: the limit of acceptable miss and false alarm probabilities due to the noise variance of the detector output. That limit is likely to dictate in practical terms the effective sensitivity of the detector, and reduce the significance of the effect discussed above. This is an area for further investigation. For present purposes we will assume we wish to operate the detector in the range where $(E[D_{\text{signal+noise}}] - E[\text{SWAC}_{\text{noise}}])$ can reasonably be approximated by $E[D_{\text{signal}}]$.

Now to find an expression for $E[D_{\text{signal}}]$. For a total measurement time of $T = MT_s$, the expected value of D_{signal} is simply equal to s^2 multiplied by the total number of samples correlated, i.e.

$$E[D_{\text{signal}}] = MWT_s s^2 = WT s^2 \quad (6)$$

Hence

$$(E[D_{\text{signal}}])^2 = (WT)^2 s^4 \quad (7)$$

We define random variable U as the real part of the sample-wise complex correlation of one adjacent symbol pair, i.e.

$$U = \text{Re} \left\{ \sum_{k=k_0+(n-1)\tau}^{k_0+n\tau} (y_k \cdot \bar{y}_{k+\tau}) \right\} \quad (8)$$

With only noise input to the detector, each y_k and $y_{k+\tau}$ are independent complex Gaussian noise values N_1 and N_2 .

$$\text{Var}(N_1 \cdot N_2) = \text{Var}(N_1) \cdot \text{Var}(N_2) = \sigma^4 \quad (9)$$

U is the sum of many (τ) such terms, each with normal-product distribution, with zero mean and variance σ^4 . Hence the summation will approach a Gaussian distribution for large τ , with zero mean and variance $\tau \sigma^4$. Since U is the real component only of the summation, the variance of U is given by

$$\text{Var}(U) = \frac{\tau \sigma^4}{2} \quad (10)$$

The question we need to address is... what is $\text{Var}(|U|)$? To find this we start by obtaining the “first central absolute moment” of U . Definitions of the various moments of Gaussian distributions can be found in [12].

$$\begin{aligned} E[|U|] &= \sigma_U (0!!) \cdot \sqrt{\frac{2}{\pi}} \\ &= \sqrt{\frac{2}{\pi}} \sqrt{\frac{\tau \sigma^4}{2}} = \sqrt{\frac{\tau}{\pi}} \sigma^2 \end{aligned} \quad (11)$$

Then find the “second central absolute moment” of U :

$$\begin{aligned} E[(|U|)^2] &= \sigma_U^2 (1!!) \\ &= \sigma_U^2 = \text{Var}(U) \\ &= \frac{\tau \sigma^4}{2} \end{aligned} \quad (12)$$

Then we find $\text{Var}(|U|)$ as follows:

$$\begin{aligned}
\text{Var}(|U|) &= E[(|U|)^2] - (E[|U|])^2 \\
&= \frac{\tau\sigma^4}{2} - \frac{\tau}{\pi}\sigma^4 \\
&= \tau\sigma^4\left(\frac{1}{2} - \frac{1}{\pi}\right) \\
&\approx (0.182)\tau\sigma^4
\end{aligned} \tag{13}$$

If we compare Equation (13) with the expression for $\text{Var}(U)$ in Equation (10) we see that the effect of taking the absolute value of U is to reduce the variance by a factor of ~ 2.75 .

$\text{SWAC}_{\text{noise}}$ is the sum of M independent random variables, each of variance $\text{Var}(|U|)$. Hence

$$\begin{aligned}
\text{Var}(\text{SWAC}_{\text{noise}}) &= M \cdot \text{Var}(|U|) \\
&\approx (0.182)M\tau\sigma^4 \\
&\approx (0.182)MWT_s\sigma^4
\end{aligned} \tag{14}$$

In the vicinity of $\tau = WT_s$

$$\text{Var}(\text{SWAC}_{\text{noise}}) \approx (0.182)MWT_s\sigma^4 \tag{15}$$

Finally we obtain the following expression for SNR_{out} :

$$\begin{aligned}
\text{SNR}_{\text{out}} &= \frac{(E[D_{\text{signal}}])^2}{\text{Var}(\text{SWAC}_{\text{noise}})} \\
&\approx \frac{(MWT_s)^2 s^4}{(0.182)MWT_s\sigma^4} \\
&\approx (5.5)MWT_s \left(\frac{s^2}{\sigma^2}\right)^2 \\
&\approx (5.5)MWT_s (\text{SNR}_{\text{in}})^2
\end{aligned} \tag{16}$$

This formulation is useful for comparing against simulations but is deceptive in the way it suggests SNR_{out} increases with W . In fact SNR_{in} decreases with W , because noise power = $W.N_0$ (where N_0 is the noise spectral density). Hence SNR_{out} actually decreases with W .

We re-formulate Equation (16) using the following substitution:

$$SNR_{in} = \frac{s^2}{\sigma^2} = \frac{\left(\frac{E_s}{T_s}\right)}{WN_0} = \frac{E_s}{WT_s N_0} \quad (17)$$

Substituting into Equation (16) gives the following expression for SNR_{out} :

$$\begin{aligned} SNR_{out} &\approx (5.5)MWT_s (SNR_{in})^2 \\ &\approx (5.5)MWT_s \left(\frac{E_s}{WT_s N_0}\right)^2 \\ &\approx \frac{(5.5)M}{WT_s} \left(\frac{E_s}{N_0}\right)^2 \end{aligned} \quad (18)$$

First we should note that this is the optimum detector sensitivity in the absence of Doppler acceleration or ISM impairments. Each of these will degrade detector sensitivity and place limits on practical measurement time.

As expected for an autocorrelation-based detector, SNR_{out} is proportional to the square of SNR_{in} . This results in a significant penalty when SNR_{in} is low.

The denominator term WT_s is the factor by which the noise bandwidth exceeds the symbol rate. The smallest this can get is when W is equal to the spread-spectrum signal bandwidth. We see that there is a detection benefit if the spreading factor is reduced because it allows smaller W . There is a trade-off when choosing the spreading factor between interference immunity and noise sensitivity.

Importantly SNR_{out} **increases with T** (directly proportional to the number of symbols processed, M). The significance of this is that the sensitivity loss due to the squaring of SNR_{in} (and any other causes) can be recovered by increasing M . One can process as many symbols as needed to achieve a target detection reliability in terms of miss/false-alarm probabilities. The required M to achieve acceptable miss/false-alarm probabilities will be longer with lower input SNR. However, if a sufficiently long segment of signal is available, detection can be made arbitrarily reliable (subject to the deterioration of detection sensitivity experienced at extremely low SNRs, as mentioned earlier in this section). This makes it possible to achieve signal discovery at SNRs well below what is needed to extract the information content of the signal. This is significant for SETI because it permits early discovery of a signal using a radio telescope that is smaller than what will later be needed for the data extraction phase.

3.5 Sensitivity Comparison

Using antipodal spread-spectrum modulation as an example, the sensitivity expressions for three detector types are given below.

Matched Filter

$$SNR_{MF} = M \left(\frac{E_s}{N_0} \right)$$

SWAC

$$SNR_{SWAC} \approx \frac{(5.5)M}{WT_s} \left(\frac{E_s}{N_0} \right)^2$$

Energy Detector

$$SNR_{ED} = \frac{M}{2WT_s} \left(\frac{E_s}{N_0} \right)^2$$



SWAC is ~10 dB more sensitive than energy detection

Figure 13 plots detector output SNR versus input SNR for the matched-filter, SWAC and energy-detector cases, for an illustrative signal format as described in the figure caption. As expected, the matched filter is superior at all input SNRs, given that it represents the theoretical optimum receiver type. However in a SETI context where the modulation waveforms are unknown, a matched filter is not available⁹. SWAC and energy detection track each other with the same slope (due to the squaring of SNR_{in}) but SWAC maintains a consistent 10 dB advantage. Note that the formulas derived for SWAC and energy detection are only valid for low SNR_{in} due to assumptions made about the statistics of [signal+noise] approaching those of [noise only]. Above approximately 10 dB SNR_{in} the output SNR is expected to tail off and asymptotically approach the matched filter performance.

Note that Figure 13 relates to the performance of differing detection methods, with no consideration of the differing computational complexities inherent to each method. The computational requirements of SWAC are discussed in detail in Section 4.

3.6 SWAC Refinements

Note that the SWAC algorithm assumed in Figure 13 is the “basic SWAC” approach that involves correlations between *consecutive* symbols. The concept can be generalised to include other pair-wise correlations of symbols in close vicinity. We refer to the generalised case as *near-neighbour SWAC* to differentiate it from the basic algorithm, which we call *adjacent-symbol SWAC*.

It is preferable to limit the time separation of correlated symbols to reduce the degradations due to Doppler drift and time-varying channel characteristics – as discussed in Section 3.7. However, additional correlation pairs will increase the effective M for a given length of waveform. As an example, correlating each symbol with $1T_s$, $2T_s$ and $3T_s$ delayed versions of the signal will yield an effective 3-fold increase in M, which should translate to an almost 5 dB improvement in detection sensitivity. The gap between SWAC and matched filter performance can therefore be closed to a significant extent by performing additional processing of the same M-symbol span of received waveform. This improvement is possible because it exploits the independence of the noise present in the samples of each symbol – an improvement that is not possible with an energy detector.

⁹ With unlimited computational capacity it may be possible to perform a search over the modulation parameter space to find a matched filter by trial-and-error. However, the number of parameters in the search space suggests this approach will remain impractical for the foreseeable future unless many of the parameters can be constrained somehow, e.g. deduced through ‘implicit coordination’.

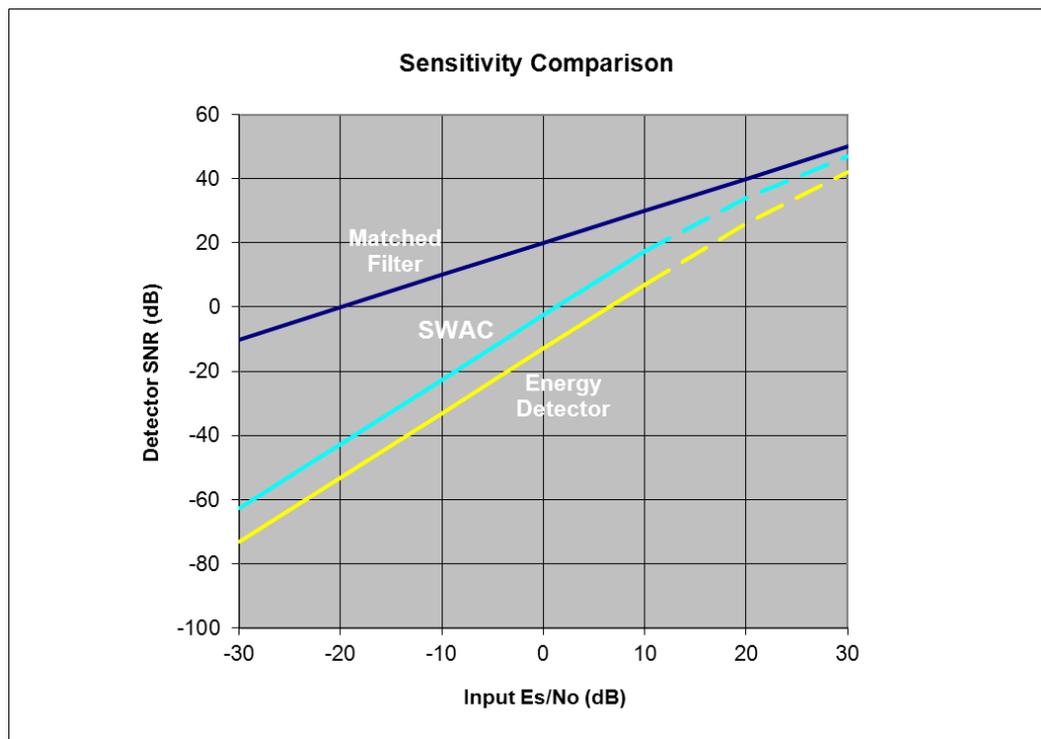


Figure 13: Sensitivity comparison of three detector types, assuming an antipodal spread-spectrum BPSK signal in Gaussian noise (symbol rate 1 symbol/s, spreading factor 500, sampling rate (=noise bandwidth) 1kHz, measurement interval 100 seconds ($M=100$))

All the approaches discussed so far assume pair-wise cross-correlations between two nearby symbols. Another way to improve SWAC performance, particularly at the low end of the operating SNR range, is to perform cross-correlations of multiple-symbol groupings with other multiple-symbol groupings. This will have the effect of increasing the effective value of τ , and hence increase the degree of noise averaging before the absolute value is taken. For the same noise variance this will result in a lower mean noise output, thus reducing the sensitivity loss effect illustrated in Figure 12.

The investigation of these and other refinements to SWAC is the subject of ongoing research by the author.

3.7 Effect of Channel Impairments

3.7.1 Doppler

Interstellar signalling involves the propagation of a signal from a transmitter, through the interstellar medium (ISM), to one or more receivers. The relative motions of the transmitter, ISM and receiver give rise to Doppler effects observed by the receiver. Motions with constant velocity will result in a static time dilation, whereas if the motions involve any acceleration components there will be a dynamic time dilation, referred to as ‘Doppler drift’.

SWAC is insensitive to static Doppler effects because consecutive symbols are affected equally. The only consequence is that the value of τ at which the SWAC peak occurs will move very slightly, because of the lengthening (or shortening) of the symbol interval due to the time dilation. SWAC is, however, sensitive to Doppler *drift* because consecutive symbols experience slightly different degrees of time dilation. Over longer processing timespans there may be a ‘smearing’ of the SWAC peak across multiple delay bins. The frequency offset and phase shift from one symbol to the next will also vary, which will reduce sensitivity and preclude the use of the optimisation described in footnote 6. The effect is less significant for shorter symbol periods because consecutive symbols are closer together in time. For example, the effect of the Earth’s rotation on a signal centred on 10 GHz will be

insignificant for symbol rates greater than 100 symbol/s. However, such a constraint can be avoided completely if Doppler compensation is employed. The transmitter and receiver are both aware of the component of their own acceleration along the line of sight. They can each therefore correct for this acceleration by appropriate frequency shifting processes synchronised to their known accelerations¹⁰. There is a compelling case for Doppler compensation to be routinely employed for both METI and SETI. This would reduce the difficulties a receiver will face when attempting to detect signals of low symbol rate – and low symbol rates are arguably more desirable for interstellar signalling to constrain transmitter power requirements.

3.7.2 Interstellar Medium

The ISM will also introduce time-varying *dispersion* and *scattering* effects that will degrade any signal propagating through it and make detection at a receiver more challenging [13]. Together dispersion and scattering cause complicated delay-spread behaviour for wideband signals, which results in waveform distortion and intersymbol interference (ISI) effects that are difficult to mitigate prior to discovery of the signal¹¹. These effects can seriously compromise the performance of a matched-filter detector. However, autocorrelation detection is relatively immune to the effects of dispersion and scattering. Successive symbols experience similar distortion to their waveforms, hence retaining high cross-correlation. ISI can be more problematic but its significance is reduced when operating with longer symbol periods and/or at higher carrier frequencies.

¹⁰ Doppler compensation at the receiver can most simply be performed by chirping the local oscillator used for down-conversion. However, this method can only be used in a single-beam observing mode. In a multi-beam mode there will be different de-drift requirements in each beamformer. In this case it is necessary to perform the Doppler compensation separately for each beam, implemented digitally as part of the signal detection process – which increases the computational complexity.

¹¹ After discovery it becomes possible to compensate for many of the distortions introduced by the ISM because the channel parameters can be estimated by the receiver and reversed [13].

4. SWAC Computational Complexity

This section analyses the computational complexity of SWAC and compares it with conventional autocorrelation.

4.1 Conventional Autocorrelation – Frequency Domain

It is well known that a computationally efficient way to compute the autocorrelation of a sequence of discrete samples is to take advantage of the Wiener-Khinchin theorem, which involves taking the Fast Fourier Transform (FFT) of the time samples, taking the magnitude-squared of the resultant frequency domain samples, then performing an Inverse FFT to obtain the autocorrelation function.

A commonly employed method of obtaining the autocorrelation of a long sequence of samples is to partition the sequence into a number of smaller sub-blocks and perform FFTs on these sub-blocks. To deal with edge effects, typically there will be twice as many sub-blocks used, each 50% overlapping with its neighbours, as shown in Figure 14.

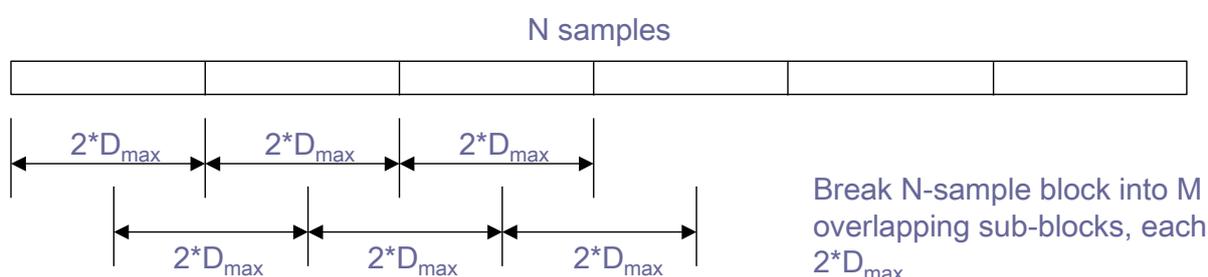


Figure 14: Overlapped sub-blocks for FFT based computation of autocorrelation

Each FFT is of length $(2 \cdot D_{\max})$, which produces an autocorrelation delay range of $\pm D_{\max}$.

The number of FFTs required is $M \approx 2 \cdot N / (2 \cdot D_{\max}) = N / D_{\max}$. Each FFT requires $(2 \cdot D_{\max} / 2) \log_2(2 \cdot D_{\max}) = D_{\max} \log_2(2 \cdot D_{\max})$ complex multiplications and $(2 \cdot D_{\max}) \log_2(2 \cdot D_{\max})$ complex additions. Each FFT also requires windowing, hence a further $(2 \cdot D_{\max})$ complex multiplications. Then each point in each FFT output needs to be squared, hence another $(2 \cdot D_{\max})$ complex multiplications. Finally, the squared output points of each FFT needs to be added, requiring $(N / D_{\max}) \cdot (2 \cdot D_{\max}) = 2N$ complex additions.

The Inverse FFT to produce the autocorrelation spectrum requires $D_{\max} \log_2(2 \cdot D_{\max})$ complex multiplications and $(2 \cdot D_{\max}) \log_2(2 \cdot D_{\max})$ complex additions.

Ignoring normalisation, this gives totals of approximately:

- $(N + D_{\max}) \log_2(2 \cdot D_{\max}) + 4N$ complex multiplications, and
- $2(N + D_{\max}) \log_2(2 \cdot D_{\max}) + 2N$ complex additions.

If we assume an arithmetic unit that performs complex multiplications and additions in the same number of clock cycles, then the total complexity is $C \approx 3(N + D_{\max}) \log_2(2 \cdot D_{\max}) + 6N$. This expression is approximately proportional to $N \log_2 D_{\max}$, a well-known result for autocorrelation using the Wiener-Khinchin algorithm.

4.2 Conventional Autocorrelation – Time Domain

Autocorrelation can also be computed entirely in the time domain using the so-called ‘brute force’ approach. Here the correlation score is calculated for each individual delay value by accumulating the product of each sample with the sample at the given delay value, as depicted in Figure 15. While this

is conceptually much simpler than the Wiener-Khinchin algorithm, it results in much larger computational complexity, particularly for large N .

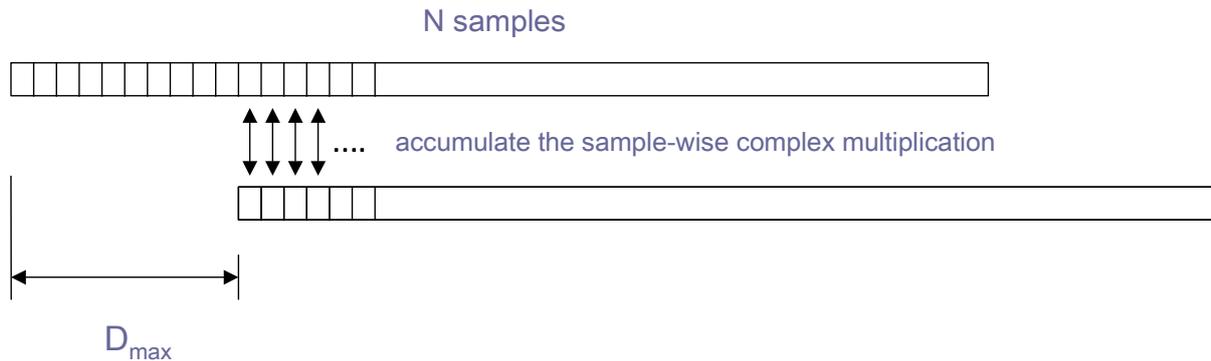


Figure 15: Sample-wise multiply-accumulate operations for time-domain computation of autocorrelation

Firstly, let us assume that all delays are to be evaluated from 1 to D_{max} (which will provide us with the same set of output data as the Wiener-Khinchin algorithm). Normally $D_{max} \ll N$ so it is a close approximation to say that for each autocorrelation delay value there are $\sim N$ complex multiplications and N complex additions. Since there are D_{max} delays to evaluate, this gives totals of approximately:

- $N \cdot D_{max}$ complex multiplications, and
- $N \cdot D_{max}$ complex additions.

Again, if we assume an arithmetic unit that performs complex multiplications and additions in the same number of clock cycles, then the total complexity is $C \approx 2 \cdot N \cdot D_{max}$. This expression is proportional to $N \cdot D_{max}$, a well-known result for autocorrelation using the brute force approach.

4.3 SWAC

SWAC is similar to conventional autocorrelation but with the difference that (assumed) symbols are effectively being ‘folded’ onto one another, and the absolute value of each symbol’s contribution to the autocorrelation is taken before the accumulation. The sample sequence is partitioned into M sub-blocks of length D_x , as depicted in Figure 16. Each D_x from 1 to D_{max} is trialled, calculating the SWAC score according to Equation (4) in Section 3.1, with τ set to D_x for each trial.

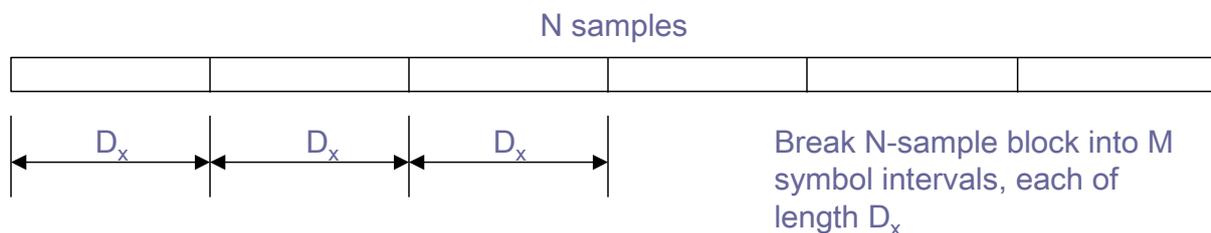


Figure 16: Partitioning into sub-blocks for SWAC processing

The method for implementing this in the time domain is conceptually simple and requires just minor modifications to conventional time-domain autocorrelation. For a frequency-domain implementation of SWAC the time-folding operation is problematic, as is the nonlinear $\text{abs}()$ function. It may be possible to derive a frequency-domain implementation of SWAC, but this has not yet been accomplished by the author.

As with conventional autocorrelation performed in the time-domain, let us assume that all delays are to be evaluated from 1 to D_{\max} . We assume $D_{\max} \ll N$ so the number of symbols to process in a given trial is $M \approx N/D_x$. Each cross-correlation between adjacent symbols requires D_x complex multiplications and D_x complex additions. The total computations for the whole block of samples at a given D_x is $\approx N$ complex multiplications and $\approx N$ complex additions. Since there are D_{\max} delays to evaluate, this gives totals of approximately:

- $N \cdot D_{\max}$ complex multiplications, and
- $N \cdot D_{\max}$ complex additions.

Again, if we assume an arithmetic unit that performs complex multiplications and additions in the same number of clock cycles, then the total complexity is $C \approx 2 \cdot N \cdot D_{\max}$. This is in fact the same as for conventional autocorrelation in the time domain.

At this point we note an important feature of both time-domain conventional autocorrelation and SWAC, which is that the computational complexity is directly proportional to the number of delay values to be evaluated. By contrast, the Wiener-Khinchin algorithm evaluates *all* delay values at one fell swoop. If the entire range of delay values is not needed, the time-domain approach becomes comparatively more efficient. For small delay ranges the time-domain approach may in fact require fewer total computations than the frequency-domain approach.

There are two ways that the range of delay values to be evaluated can be constrained. The first assumes the full sample resolution is maintained and that a decision has been made to limit the delay range between selected D_{\min} and D_{\max} values (on the basis of some other information constraining the range of symbol rates under consideration for the search). The second way is to search the full range from 1 to D_{\max} but more coarsely, i.e. by not evaluating every delay value in this range. This approach is possible without penalty if the bandwidth of the target signal is less than the sampling rate. A similar reduction in complexity can be obtained in this scenario by decimating the sample stream such that the sampling rate is closer to the minimum needed to represent the signal bandwidth, prior to performing the SWAC processing. However, when the target signal bandwidth is an unknown, the appropriate degree of decimation is also unknown. Therefore, some trial-and-error could be needed.

One possible strategy to reduce the overall computational requirement would be to search initially over a coarse sample resolution, and then where the SWAC output suggests that some interesting feature may be present, search in that vicinity using progressively higher sample resolutions until the strength of the feature is maximised. However, care must be taken to ensure that the initial resolution is not too coarse for the feature to ‘slip through the net’ completely.

4.4 Comparison

To illustrate relative complexities we analyse the following example scenario. We wish to process data with delays up to a maximum of $D_{\max} = 16,384$ and for data segments of length ranging from $(10^1 \cdot D_{\max})$ to $(10^4 \cdot D_{\max})$. We consider the total computational complexity C as the sum of the number of complex multiplications and additions. Figure 17 plots C versus N for three cases:

- (1) conventional autocorrelation using FFTs (evaluating 16,384 delays);
- (2) conventional autocorrelation or SWAC in the time-domain (evaluating 16,384 delays); and
- (3) conventional autocorrelation or SWAC in the time-domain (evaluating 50 delays).

The computational efficiency of frequency-domain autocorrelation is clear to see when the entire range of delay values is required. However, if the delay range of interest happens to be narrower (in this case 50 values), the complexity of the time-domain algorithms is seen to be very similar to that of the frequency-domain approach.

At the time of writing, FFT-based autocorrelation is generally considered to be feasible at real-time speeds, since it is similar in complexity to ordinary Fourier spectral analysis that is routinely implemented in radio telescope back-ends. SWAC initially appears to be significantly more complex, because it appears necessary to compute it in the time domain – raising concerns about the ability to run SWAC in real-time. However, this assumes all delay values from one to the maximum are required, which may not necessarily be the case. This gives us hope that real-time SWAC processing may be practical even with current processor technology. As processing power increases over time, and/or more efficient methods for computing SWAC are developed, the need to constrain the evaluated delay range with SWAC will diminish.

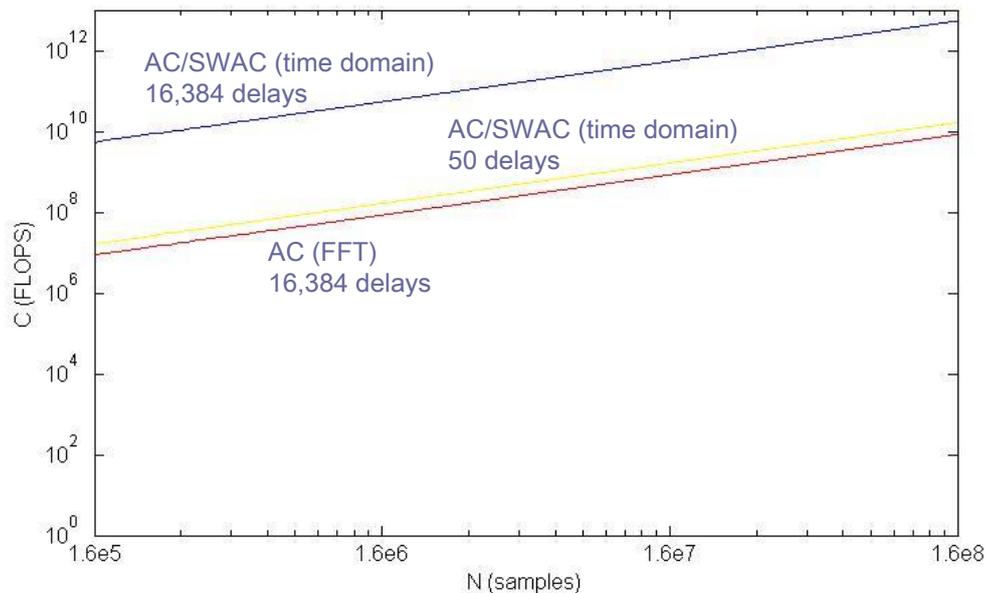


Figure 17: Computational complexity versus N for different autocorrelation-based algorithms

The overall SWAC computational complexity is a function of both the time sample resolution (which determines the number of delay values to evaluate) and the total time-span of signal being processed (number of assumed symbols M) – both of which influence the detection sensitivity. So there is a trade-off (as with all searches) between computational requirements and the parameter space that can be searched, including threshold SNR. For any practical implementation of the SWAC algorithm this trade-off should be given careful attention so as to optimise the use of the available resources.

5. Conclusions and Recommendations

This section summarises the findings of this study and makes some recommendations for future work.

5.1 Conclusions

- Wideband cyclostationary signals can be detected using autocorrelation without a priori knowledge of the signal's structure or carrier frequency. Detection is possible by searching over one dimension: symbol period.
- Symbol-wise autocorrelation (SWAC) is more effective than conventional autocorrelation for detecting randomly modulated signals.
- SWAC detector sensitivity is maximised with binary antipodal spread-spectrum signalling. A mathematical expression for sensitivity was derived for this scenario. It is shown that the basic SWAC algorithm is ~10 dB better than energy detection. Further gains of 3 to 5 dB can be expected from ongoing SWAC refinements.
- SWAC sensitivity is inferior to a matched filter detector, but in a SETI scenario the waveform parameters needed to implement a matched filter are not known. Therefore SWAC appears to represent the next best option. Its sub-optimality can be overcome by processing longer time-segments of received waveform data.
- The scaling of sensitivity with time allows the detection sensitivity to be made arbitrarily reliable, given access to sufficient signal. This allows the possibility of discovering the presence of a wideband beacon signal at an SNR well below what is needed to extract the information content of the signal. This is significant for SETI because it permits early discovery of a signal using a radio telescope that is smaller than what will later be needed for the data extraction phase.
- The computational complexity of SWAC is the same as conventional autocorrelation performed in the time domain. Whilst the frequency-domain computation of autocorrelation is more efficient than time-domain computation (at least for evaluating wide delay ranges), it is unclear whether frequency-domain methods can be applied to SWAC. To achieve real-time SWAC processing at the present time, it may be necessary to limit the number of delays evaluated, either by constraining the symbol period search range or by using coarser time resolution. This concern will diminish over time as available processing power increases, and/or more efficient implementations of SWAC are developed.

In summary: the outcomes of this study suggest there is a strong argument to be made for including autocorrelation detection (both conventional autocorrelation and SWAC) in future SETI programmes on the ATA and other radio telescopes.

5.2 Recommendations

- The effects of Doppler acceleration and the ISM must be taken into account when determining the sensitivity of a practical SWAC detector. These effects require further study and should be included in future simulation environments for evaluating SWAC.
- Extension of the basic adjacent-symbol SWAC algorithm to more powerful near-neighbour algorithm variants should be pursued and the performance gains quantified through analysis and simulation.
- Techniques to more efficiently compute the SWAC algorithm should be explored.
- Both conventional autocorrelation and SWAC variations should be incorporated into the setiQuest project, including the ability to display output results in the form of AC waterfall plots.
- In addition to setiQuest, autocorrelation and SWAC should be considered for real-time implementation in the ATA signal processing back-end.

- SWAC is just one example of a new detection method that can broaden the SETI search space. Investigation of other techniques offering different detection capabilities should continue generally within the SETI Institute and elsewhere to identify new approaches for incorporation into setiQuest and/or the signal processing back-ends of the ATA or other radio telescopes.
- New algorithms arising from the setiQuest project are particularly applicable to the forthcoming Square-Kilometre-Array (SKA) radio telescope. An opportunity exists now to begin to educate the SKA community on emerging capabilities to conduct wideband SETI. The ATA can serve as the ideal platform to validate and refine these techniques ahead of implementation on the SKA.

6. References

- [1] I. S. Morrison, “Detection of Antipodal Signalling and its Application to Wideband SETI”, *Acta Astronautica*, Special Issue: Life Signatures 2, awaiting publication.
- [2] D. H. Forgan and R. C. Nichol, “A failure of serendipity: the Square Kilometre Array will struggle to eavesdrop on human-like extraterrestrial intelligence”, *International Journal of Astrobiology*, Volume 10 Issue 02, 2011.
- [3] D. G. Messerschmitt, “The Case for Spread Spectrum Signalling in Interstellar Messaging”, draft awaiting publication, available at www.eecs.berkeley.edu/~messer.
- [4] J. Benford, G. Benford and D. Benford, “Messaging with Cost-Optimized Interstellar Beacons”, *Astrobiology Journal*, Volume 10 Number 5, 2010.
- [5] W. A. Gardner, “The spectral correlation theory of cyclostationary time-series”, *Journal of Signal Processing*, Volume 11 Number 1, 1986.
- [6] W. A. Gardner, “Exploitation of spectral redundancy in cyclostationary signals”, *IEEE Signal Processing Magazine*, Volume 8 Number 2, 1991.
- [7] F. D. Drake, “Current Aspects of Exobiology: The Radio Search for Intelligent Extraterrestrial Life, Chapter IX”, Pergamon Press, 1965.
- [8] G. R. Harp, R. F. Ackermann, S. K. Blair, J. Arbutich, P. R. Backus and J. C. Tarter, “A new class of SETI beacons that contain information”, *Astrobiology Journal*, in press, awaiting publication.
- [9] C. Maccone, “Deep Space Flight and Communications: Exploiting the Sun as a Gravitational Lens”, Praxis Publishing, 2009.
- [10] G. Burel, “Detection Of Spread Spectrum Transmissions Using Fluctuations Of Correlation Estimators”, *Proc. IEEE Intelligent Signal Processing and Communication Systems*, Hawaii, USA, 2000.
- [11] D. G. Messerschmitt, “Interstellar Spread Spectrum Communications: Receiver Design for Plasma Dispersion”, draft awaiting publication, available at www.eecs.berkeley.edu/~messer.
- [12] http://en.wikipedia.org/wiki/Normal_distribution.
- [13] D. G. Messerschmitt I. S. Morrison, “Design of Interstellar Digital Communication Links: Some Insights from Communication Engineering”, *Acta Astronautica*, Special Issue: Life Signatures 2, awaiting publication.