Blind Adaptive Space-Time Processing for Cyclostationary Signals

Jie Zhang, K. Max Wong*, and Timothy N. Davidson
Department of Electrical and Computer Engineering,
McMaster University,
Hamilton, Ontario, Canada, L8S 4K1.

9 July 2001

Abstract

In this paper we present a blind adaptive space-time processing algorithm for separating signals which are spectrally and/or spatially overlapped. The algorithm exploits the cyclostationary nature of many communication signals, but does not require knowledge of the statistical properties of the desired signal. It merely requires knowledge of a (distinct) cycle frequency. It is shown that the performance of the algorithm converges at a rate $O(1/N)$, where N is the number of received samples, to the performance of the optimal (trained) receiver with the given structure. Furthermore, we provide an (algebraic) analysis of the performance of a multiuser communication system which employs our receiver, and confirm this result in simulation examples. These examples demonstrate that our cyclic adaptive space-time processor is an attractive alternative to beamforming or filtering alone.

*Corresponding author. Department of Electrical and Computer Engineering, McMaster University, 1280 Main Street West, Hamilton, Ontario, Canada, L8S 4K1. Telephone: +1-905-525-9140, Ext. 24098. Fax: +1-905-521-2922. Email: wongkm@mcmaster.ca

Number of pages: 32

Key words: blind signal separation; space-time signal processing; cyclostationarity; multiuser communications.

List of Figures

List of Tables

1 Introduction

A commonly encountered problem in multiuser communication systems is the extraction of the desired signal from co-channel interference that may overlap spectrally with the desired signal. Conventional filtering techniques are unable to perform this extraction, but a very useful property that can be exploited to accomplish this task is the cyclostationarity of the signals [1, 2]. (A cyclostationary signal has a mean and variance which vary periodically in time. A formal definition is given in Section 2.) The optimum frequency-shift (FRESH) filtering technique, called the cyclic Wiener filter [3], enables us to separate spectrally overlapped signals by using the cyclostationarity of the signals. The idea of retrieving cyclostationary signals in a multiuser environment has been studied [4, 5], and blind channel identification and equalization methods using induced cyclostationarity have also been proposed [6, 7, 8]. The advantage of such blind methods is that we do not need to have the statistics of the desired signal, nor a training signal, which, in practice, are often not available. Another important application of cyclostationarity is beamforming in a multiuser communication system. It has been shown [9, 10] that spatial re-use of allocated frequency slots by employing an antenna array with multiple beams increases the system capacity substantially. Cyclostationarity of signals can be exploited to arrive at blind adaptive array beamforming algorithms [11, 12, 13]. These beamforming methods have been proposed to be applied to cellular radio systems to increase the capacity of the system [14, 15]. The main advantages of such blind adaptive beamforming algorithms are: 1) No reference signal is required; 2) No advanced knowledge of the correlation properties of noise and interference is needed; 3) No antenna array calibration is necessary. For the above blind processing techniques, whether filtering or beamforming, signal selectivity is achieved using the knowledge of the cyclostationarity of the desired signal. However, if the spectral overlap between the desired signal and the interference is very large, the performance of blind adaptive FRESH filtering will deteriorate [5]. Likewise, if the desired signal and the interference are spatially too close to each other, the blind cyclostationary beamforming techniques will also deteriorate in performance [15].

In this paper, we propose to combine the above cyclostationarity-based filtering and the beamforming techniques to separate signals which are spectrally and/or spatially overlapped. The resulting methods will be called Cyclic Adaptive Space-Time (CAST) processing techniques. In particular, we introduce a blind CAST (B-CAST) technique which has the advantage that it does

not need the knowledge of the statistical properties of the signal nor does it need a training signal. The convergence of the CAST algorithms is analyzed, showing that the convergence rate is $O(1/N)$, where N is the number of samples received. For large N, it is also shown that B-CAST has the same performance as CAST methods which employ a training signal. The performance of a multiuser binary communication system employing a CAST processor is then analyzed. Computer simulation results confirm the derived probability of error, and show that the communication system with CAST processing at the receiver is superior in performance to that using cyclic adaptive beamforming alone.

2 Cyclostationarity and the CAST Processor

A stochastic sequence $x(n)$ is said to be cyclostationary in the wide sense [3] if its mean and variance are periodic in n with some period M; i.e., $E[x(n)] = \mu_x(n) = \mu_x(n+M)$, and $E[x(n_1)x(n_2)] = R_{xx}(n_1, n_2) = R_{xx}(n_1+M, n_2+M)$, for all n, n_1 and n_2. The Fourier expansions of $\mu_x(n)$ and $R_{xx}(n_1, n_2)$ at (angular) frequency α are denoted by μ_x^α and $R_{xx}^\alpha(n_1 - n_2)$, and are referred to as the cyclic mean and cyclic autocorrelation coefficients, respectively. They are given by

$$\mu_x^\alpha = \lim_{N\to\infty} \left\langle \mu_x(n) e^{-j\alpha n} \right\rangle_N , \qquad \text{and} \qquad R_{xx}^\alpha(k) = \lim_{N\to\infty} \left\langle R_{xx}(n+k, n) e^{-j\alpha n} \right\rangle_N ,$$

respectively, where $\langle \cdot \rangle_N$ denotes the time average over N samples. The set of α for which $R_{xx}^\alpha(k) \not\equiv 0$ is called the cycle spectrum, and any such α is called an (angular) cycle frequency. A wide sense cyclostationary process is said to be cycloergodic in the mean and autocorrelation if, with probability 1,

$$\hat{\mu}_x^\alpha(N) = \left\langle x(n) e^{-j\alpha n} \right\rangle_N \xrightarrow{\text{m.s.}} \mu_x^\alpha$$
$$\hat{R}_{xx}^\alpha(k, N) = \left\langle x(n+k) x(n) e^{-j\alpha n} \right\rangle_N \xrightarrow{\text{m.s.}} R_{xx}^\alpha(k),$$

where $\xrightarrow{\text{m.s.}}$ denotes convergence in the mean-square sense.

In this paper, we assume all the signals, whether desired or interfering are complex, cycloergodic in the mean and in autocorrelation [1, 2], and are independent from sample to sample. Furthermore, we assume that the cycle frequencies of the desired signal are different from those of the interference. This assumption is not restrictive since the desired signal and the interferences usually have different

features. Even when the users in a multiuser scheme have the same modulation formats, carrier frequencies, and symbol periods, different induced cyclostationarity for each user [6, 7, 8] can be achieved. It is the utilization of the set of cycle frequencies to distinguish the desired signal from the interferences that leads to the development of an effective B-CAST algorithm.

The general structure of a CAST processor is shown in Fig. 1, in which an antenna array of L sensors is employed. We assume that the signal arriving at the antenna consists of the desired signal $s(n)$ together with I interfering signals $u_i(n),\ i = 1 \cdots I$, which may overlap spectrally with the desired signal, and accompanied by white Gaussian noise $\nu(n)$. Thus, if we denote the signal arriving at ℓth sensor by $x_\ell(n)$, then we can write

$$\boldsymbol{x}(n) = \boldsymbol{d}(\theta_s)s(n) + \sum_{i=1}^{I} \boldsymbol{d}(\theta_{u_i})u_i(n) + \boldsymbol{\nu}(n) \tag{2.1}$$

where $\boldsymbol{x}(n) = [x_1(n) \ldots x_L(n)]^T$, $[\cdot]^T$ denotes the transpose, and $\boldsymbol{d}(\theta_s)$ and $\boldsymbol{d}(\theta_{u_i})$ are the array manifolds (steering vectors) [16] for the signal and the ith interference arriving from angles θ_s and θ_{u_i} respectively, and $\boldsymbol{\nu}(n)$ is an L dimensional noise vector, assumed to be temporally and spatially white.

The signal $x_\ell(n)$ arriving at the ℓth sensor is processed by a frequency-shift (FRESH) filter [3], the structure of which is shown in Fig. 2. Each of the parallel branches of the FRESH filter processes the input signal by shifting it, or its conjugate, in frequency by an (angular) cycle frequency $\alpha_{\ell m}$ of the desired signal and then passing it through an FIR filter of impulse response $h_{\ell m}(n)$. Without loss of generality, we assume in this paper that all the FIR filters are of length N_h and that the L FRESH filters in the CAST processor have the same number of branches M. To describe the operation of the CAST processor in a concise way, we will use the following sliding window of data samples received by the ℓth sensor:

$$\boldsymbol{x}_\ell(n) = [x_\ell(n),\ \ x_\ell(n-1), \ldots, x_\ell(n-N_h+1)]^T,\ \ \ell = 1, \ldots, L. \tag{2.2}$$

This vector and its conjugate are frequency shifted to a cycle frequency $\alpha_{\ell m}$ by the FRESH filter resulting in

$$\begin{bmatrix} \tilde{\boldsymbol{x}}_{\ell m}(n) \\ \tilde{\boldsymbol{x}}'_{\ell m}(n) \end{bmatrix} = \left[\begin{array}{c|c} \boldsymbol{A}_{\ell m} & \boldsymbol{A}_{\ell m} \end{array}\right] \begin{bmatrix} \boldsymbol{x}_\ell(n) \\ \hline \boldsymbol{x}^*_\ell(n) \end{bmatrix},\ \ m = 1, \ldots, M, \tag{2.3}$$

where

$$\boldsymbol{A}_{\ell m} = \mathrm{diag}(e^{j\alpha_{\ell m} n}, \ldots, e^{j\alpha_{\ell m}(n-N_h+1)}). \tag{2.4}$$

Let $\tilde{\boldsymbol{X}}_\ell(n)$ and $\boldsymbol{H}_\ell(n)$ be the frequency-shifted received data matrix and the filter coefficient matrix at instant n in the ℓth branch of the CAST processor, arranged as

$$\tilde{\boldsymbol{X}}_\ell(n) = \left[\tilde{\boldsymbol{x}}_{\ell 1}(n),\ \ \tilde{\boldsymbol{x}}'_{\ell 1}(n),\ \ \ldots,\ \ \tilde{\boldsymbol{x}}_{\ell M}(n),\ \ \tilde{\boldsymbol{x}}'_{\ell M}(n)\right], \tag{2.5}$$

$$\boldsymbol{H}_\ell(n) = \left[\boldsymbol{h}_{\ell_1}(n),\ \ \boldsymbol{h}'_{\ell_1}(n),\ \ \ldots,\ \ \boldsymbol{h}_{\ell M}(n),\ \ \boldsymbol{h}'_{\ell M}(n)\right], \tag{2.6}$$

where $\boldsymbol{h}_{\ell m}(n) = [h_{\ell m}(n), \ldots, h_{\ell m}(n-N_h+1)]^T$ and $\boldsymbol{h}'_{\ell m}(n) = [h'_{\ell m}(n), \ldots, h'_{\ell m}(n-N_h+1)]^T$ are the FIR filters respectively for the signals $\tilde{\boldsymbol{x}}_{\ell m}(n)$ and $\tilde{\boldsymbol{x}}'_{\ell m}(n)$ at the nth sampling instant. The output of the CAST processor can then be written as:

$$y(n) = \mathrm{vec}^\dagger\big(\boldsymbol{H}(n)\big)\ \mathrm{vec}\big(\tilde{\boldsymbol{X}}(n)\big) = \boldsymbol{\eta}^\dagger(n)\tilde{\boldsymbol{\chi}}(n), \tag{2.7}$$

where

$$\boldsymbol{H}(n) = [\boldsymbol{H}_1(n), \ldots, \boldsymbol{H}_L(n)], \qquad \tilde{\boldsymbol{X}}(n) = \left[\tilde{\boldsymbol{X}}_1(n), \ldots, \tilde{\boldsymbol{X}}_L(n)\right], \tag{2.8}$$

$\boldsymbol{\eta}(n) = \mathrm{vec}\big(\boldsymbol{H}(n)\big)$, and $\tilde{\boldsymbol{\chi}}(n) = \mathrm{vec}\big(\tilde{\boldsymbol{X}}(n)\big)$, with a superscript † denoting the conjugate transpose. The purpose of the adaptive CAST processor is to adjust the filter coefficients $\boldsymbol{\eta}(n)$ so that the desired signal $s(n)$ can be extracted from the interference $u_i(n)$ and noise $\nu(n)$.

To facilitate the representation of the processing of the components of the received data through the CAST processor, we use the following sliding window of samples of the desired signal component of the data received by the ℓth sensor:

$$\boldsymbol{s}_\ell(n) = [s_\ell(n),\ \ s_\ell(n-1),\ \ \ldots,\ \ s_\ell(n-N_h+1)]^T,\ \ \ell = 1, \ldots, L. \tag{2.9}$$

Using definitions analogous to those in Eqs (2.3)–(2.8), the desired signal component of the output of the CAST processor is $\boldsymbol{\eta}^\dagger(n)\tilde{\boldsymbol{\sigma}}(n)$, where

$$\tilde{\boldsymbol{\sigma}}(n) = \mathrm{vec}\big(\tilde{\boldsymbol{S}}(n)\big), \tag{2.10}$$

and $\tilde{\boldsymbol{S}}(n)$ is defined by analogy with $\tilde{\boldsymbol{X}}(n)$ in (2.8) Since the received data also contains the

interfering signals and noise, then the frequency-shifted interference vector $\tilde{\boldsymbol{\iota}}(n)$ and the frequency-shifted noise vector $\tilde{\boldsymbol{\nu}}(n)$ can be defined by analogy with Eq. (2.10).

The T-CAST Processor If a copy of the desired signal is available at the receiver, then $s(n)$ or its frequency shifted version can be used as a training signal (reference). Let this frequency-shifted version of the desired signal be

$$\tilde{s}_r(n) = s(n)e^{j\alpha_r n}. \tag{2.11}$$

An optimum CAST processor can then be designed to minimize the mean-square error of the output; i.e., we seek to

$$\min_{\boldsymbol{\eta}} J_T = \min_{\boldsymbol{\eta}} E\left[\left|\tilde{s}_r(n) - \boldsymbol{\eta}^\dagger(n)\tilde{\boldsymbol{\chi}}(n)\right|^2\right], \tag{2.12}$$

giving rise to the Trained Cyclic Adaptive Space-Time (T-CAST) processor. Using the principle of orthogonality [17], it is straightforward to see that the T-CAST filter coefficient vector is given by

$$\boldsymbol{\eta}_T = \boldsymbol{R}_{\tilde{\chi}\tilde{\chi}}^{-1}\boldsymbol{\rho}_{\tilde{\chi}\tilde{s}_r}, \tag{2.13}$$

where $\boldsymbol{R}_{\tilde{\chi}\tilde{\chi}}^{-1}$ is the correlation matrix of the frequency shifted data vector and $\boldsymbol{\rho}_{\tilde{\chi}\tilde{s}_r}$ is the cross-correlation vector between the frequency shifted data vector and the α_r-shifted desired signal. We note that $\boldsymbol{R}_{\tilde{\chi}\tilde{\chi}}$ is almost always positive definite and thus non-singular [17].

The B-CAST Processor In a communication system, the transmission of a known training signal may represent extra costs on bandwidth and overhead, and in some other applications, a copy of the desired signal may simply not be available. In such cases, a Blind Cyclic Adaptive Space-Time (B-CAST) processor will be desirable; i.e., we try to extract the desired signal from the interference without a training signal and without knowledge of the statistics of the desired signal. We blindly adapt the filter coefficients with prior knowledge of only the modulation type, the carrier frequency, and the baud rate of the desired signal.

The configuration of the B-CAST processor is the same as that shown in Fig. 1, while its input is given by Eq. (2.1). Our purpose is to extract the desired signal $s(n)$ by adjusting the filter coefficients so that the output closely approximates $s(n)$. Since there is no reference signal available, a suitable reference is created by taking the input $x(n)$ and modulating it by a cycle frequency α_r of the desired signal; i.e., the reference signal $r(n)$ of the B-CAST processor is given

by:

$$r(n) = x(n)e^{j\alpha_r n} = \tilde{s}_r(n) + \sum_{i=1}^{I} \tilde{u}_{r_i}(n) + \tilde{\nu}_r(n), \tag{2.14}$$

where $\tilde{s}_r(n)$ is given by Eq. (2.11) and $\tilde{u}_{r_i}(n)$ and $\tilde{\nu}_r(n)$ are respectively the α_r-shifted versions of $u_i(n)$ and $\nu(n)$, defined in an analogous way to $\tilde{s}_r(n)$. With $\alpha_r \neq \alpha_{\ell m}, \quad \ell = 1, \ldots, L, \quad m = 1, \ldots, M$, the B-CAST processor seeks to maximize the normalized correlation between $y(n)$ and $r(n)$ by adjusting the filter coefficients $\boldsymbol{\eta}$; i.e.,

$$\max_{\boldsymbol{\eta}} J_B = \max_{\boldsymbol{\eta}} \frac{|R_{yr}|^2}{|R_{yy}||R_{rr}|}, \tag{2.15}$$

where $R_{uv} = E[u(n)v^*(n)]$. The rationale behind this is that if the output $y(n)$ is a close approximate to desired signal $s(n)$ and is relatively free from interference $u_i(n)$, $i = 1, 2, \ldots, I$, and noise $\nu(n)$, then it must have high correlation with the α_r-shifted version of $s(n)$ and must have low correlation with the α_r-shifted versions of $u_i(n)$ and $\nu(n)$ respectively. Thus, the correlation of the two signals may provide a measure of the suppression of the interference and the closeness of the output to the desired signal. In using this measure, we must ensure that the FRESH filters in the B-CAST processor primary branch has no common frequency shift with the reference. The optimum filter coefficients $\boldsymbol{\eta}_B$ can be obtained by substituting Eqs (2.7) and (2.14) into (2.15) and applying the Schwarz inequality so that

$$\boldsymbol{\eta}_B = \boldsymbol{R}_{\tilde{\chi}\tilde{\chi}}^{-1} \boldsymbol{\rho}_{\tilde{\chi}r}, \tag{2.16}$$

where $\boldsymbol{\rho}_{\tilde{\chi}r} = E[\tilde{\boldsymbol{\chi}}(n)r^*(n)]$. We note that the optimum filter coefficients of both the T-CAST and B-CAST processors as expressed in Eqs (2.13) and (2.16) have the form of the Wiener-Hopf equation, being different only in the cross-correlator vectors used. A recursive method for updating the coefficients in both processors can be obtained following the Widrow-Hoff algorithm [17] by using $\tilde{s}_r(n)$ and $r(n)$ as the respective reference signals and choosing an appropriate step-size. Alternatively, we may window the frequency-shifted data vector $\tilde{\boldsymbol{\chi}}(n)$ to obtain a time-averaged estimate $\hat{\boldsymbol{R}}_{\tilde{\chi}\tilde{\chi}}(n)$ of the data correlation matrix $\boldsymbol{R}_{\tilde{\chi}\tilde{\chi}}$ and establish the corresponding correlation vectors with either $\tilde{s}_r(n)$ or $r(n)$ and apply the standard recursive least-square (RLS) algorithm [17] to up-date the filter coefficients. In the next section, we show that the RLS estimates of $\boldsymbol{\eta}_T$ and $\boldsymbol{\eta}_B$ both converge to the same value under large sample time-average realization.

3 Convergence Analysis of the CAST Processors

To realize the T-CAST and B-CAST signal processors of Eqs (2.13) and (2.16) respectively, we have to employ time-averaged approximations of the correlation matrix and cross-correlation vectors. To this end, we define the N-sample estimates of the correlation matrix of the signal vector $\tilde{\boldsymbol{\chi}}(n)$ and the cross-correlation vector of $\tilde{\boldsymbol{\chi}}(n)$ and another signal $q(n)$ as

$$\hat{\boldsymbol{R}}_{\tilde{\chi}\tilde{\chi}}(N) = \Big\langle \tilde{\boldsymbol{\chi}}(n)\tilde{\boldsymbol{\chi}}^{\dagger}(n) \Big\rangle_N \quad \text{and} \quad \hat{\boldsymbol{\rho}}_{\tilde{\chi}q}(N) = \langle \tilde{\boldsymbol{\chi}}(n) q^*(n) \rangle_N \,, \tag{3.1}$$

respectively. Thus, the finite-sample time-average realization of $\boldsymbol{\eta}_T$ and $\boldsymbol{\eta}_B$ in Eqs (2.13) and (2.16) are given, respectively, by

$$\hat{\boldsymbol{\eta}}_T(N) = \hat{\boldsymbol{R}}_{\tilde{\chi}\tilde{\chi}}(N)\hat{\boldsymbol{\rho}}_{\tilde{\chi}s_r}(N) \quad \text{and} \quad \hat{\boldsymbol{\eta}}_B(N) = \hat{\boldsymbol{R}}_{\tilde{\chi}\tilde{\chi}}(N)\hat{\boldsymbol{\rho}}_{\tilde{\chi}r}(N). \tag{3.2}$$

We will apply the 2-norms of a vector $\boldsymbol{v}$ and a matrix $\boldsymbol{A}$, which are defined as

$$\|\boldsymbol{v}\| = \big(\boldsymbol{v}^{\dagger}\boldsymbol{v}\big)^{1/2} \qquad \text{and} \qquad \|\boldsymbol{A}\| = \max_{\|\boldsymbol{v}\|=1} \big\|\boldsymbol{A}\boldsymbol{v}\big\|, \tag{3.3}$$

respectively. In the following theorem we state that the finite-sample realizations of the weight vectors converge to their optimal values at a rate which is $O\big(1/N\big)$.

Theorem 3.1 *Let Ψ_s and Ψ_u respectively denote the set of cycle frequencies of the desired signal $s(n)$ and interferences $u_i(n)$, $i = 1, 2, \ldots, I$, and let them be such that $\Psi_s \bigcap \Psi_u = \emptyset$. Let $\alpha_{\ell m}$, $\ell = 1, \ldots, L$, $m = 1, \ldots, M$, be the frequency shift parameters in the branches of the T-CAST and B-CAST processors. Let α_r be the frequency shift parameter in the reference branch of the B-CAST processor be such that*

$$\alpha_{\ell m} \neq \alpha_r, \qquad \textit{and} \qquad \alpha_{\ell m}, \;\; (\alpha_{\ell m} - \alpha_r) \in \Psi_s, \quad \forall \ell, m. \tag{3.4}$$

Suppose that $\hat{\boldsymbol{R}}_{\tilde{\chi}\tilde{\chi}}(N)$ is full rank for $N \geq 2MLN_h$ with probability 1, then both $\hat{\boldsymbol{\eta}}_T(N)$ and $\hat{\boldsymbol{\eta}}_B(N)$ converge in the mean-square sense to

$$\boldsymbol{\eta}_{opt} = \boldsymbol{R}^{-1}_{\tilde{\chi}\tilde{\chi}}\boldsymbol{\rho}_{\tilde{\sigma}\tilde{s}_r} \tag{3.5}$$

at rates which are $O\big(1/N\big)$.

Proof: As shown in Section 2, the frequency-shifted input data vector in a CAST processor contains the frequency-shifted desired signal, interference, and noise vectors such that

$$\tilde{\boldsymbol{\chi}}(n) = \tilde{\boldsymbol{\sigma}}(n) + \tilde{\boldsymbol{\imath}}(n) + \tilde{\boldsymbol{\nu}}(n), \tag{3.6}$$

where $\tilde{\boldsymbol{\sigma}}(n)$ was defined in Eq. (2.10) and $\tilde{\boldsymbol{\imath}}(n)$ and $\tilde{\boldsymbol{\nu}}(n)$ are similarly defined. Under the assumption that $\hat{\boldsymbol{R}}_{\tilde{\chi}\tilde{\chi}}(N)$ has full rank for $N \geq 2MLN_h$, we have

$$\begin{aligned} E\left[\left\|\boldsymbol{R}_{\tilde{\chi}\tilde{\chi}}^{-1} - \hat{\boldsymbol{R}}_{\tilde{\chi}\tilde{\chi}}^{-1}(N)\right\|^2\right] &= E\left[\left\|\boldsymbol{R}_{\tilde{\chi}\tilde{\chi}}^{-1}(\hat{\boldsymbol{R}}_{\tilde{\chi}\tilde{\chi}}(N) - \boldsymbol{R}_{\tilde{\chi}\tilde{\chi}})\hat{\boldsymbol{R}}_{\tilde{\chi}\tilde{\chi}}^{-1}(N)\right\|^2\right] \\ &\leq \left\|\boldsymbol{R}_{\tilde{\chi}\tilde{\chi}}^{-1}\right\|^2 \left[\left\|\hat{\boldsymbol{R}}_{\tilde{\chi}\tilde{\chi}}^{-1}(N)\right\|^2\right] E\left[\left\|\hat{\boldsymbol{R}}_{\tilde{\chi}\tilde{\chi}}(N) - \boldsymbol{R}_{\tilde{\chi}\tilde{\chi}}\right\|^2\right], \end{aligned} \tag{3.7}$$

where we have used the Schwarz inequality. From Eq. (3.1) and a property of matrix norms [18], we have

$$\begin{aligned} E\left[\left\|\hat{\boldsymbol{R}}_{\tilde{\chi}\tilde{\chi}}(N) - \boldsymbol{R}_{\tilde{\chi}\tilde{\chi}}\right\|^2\right] &\leq E\, tr\left[\left(\hat{\boldsymbol{R}}_{\tilde{\chi}\tilde{\chi}}(N) - \boldsymbol{R}_{\tilde{\chi}\tilde{\chi}}\right)\left(\hat{\boldsymbol{R}}_{\tilde{\chi}\tilde{\chi}}(N) - \boldsymbol{R}_{\tilde{\chi}\tilde{\chi}}\right)^\dagger\right] \\ &= trE\left[\frac{1}{N^2}\sum_{n=1}^{N}\left(\tilde{\boldsymbol{\chi}}(n)\tilde{\boldsymbol{\chi}}^\dagger(n) - \boldsymbol{R}_{\tilde{\chi}\tilde{\chi}}\right)\left(\tilde{\boldsymbol{\chi}}(n)\tilde{\boldsymbol{\chi}}^\dagger(n) - \boldsymbol{R}_{\tilde{\chi}\tilde{\chi}}\right)^\dagger\right] \\ &\quad + trE\left[\frac{1}{N^2}\sum_{m}\sum_{\substack{n \\ n\neq m}}\left(\tilde{\boldsymbol{\chi}}(m)\tilde{\boldsymbol{\chi}}^\dagger(m) - \boldsymbol{R}_{\tilde{\chi}\tilde{\chi}}\right)\left(\tilde{\boldsymbol{\chi}}(n)\tilde{\boldsymbol{\chi}}^\dagger(n) - \boldsymbol{R}_{\tilde{\chi}\tilde{\chi}}\right)^\dagger\right] \\ &= \frac{1}{N}trE\left[\left(\tilde{\boldsymbol{\chi}}(n)\tilde{\boldsymbol{\chi}}^\dagger(n) - \boldsymbol{R}_{\tilde{\chi}\tilde{\chi}}\right)\left(\tilde{\boldsymbol{\chi}}(n)\tilde{\boldsymbol{\chi}}^\dagger(n) - \boldsymbol{R}_{\tilde{\chi}\tilde{\chi}}\right)^\dagger\right], \end{aligned} \tag{3.8}$$

where tr denotes the trace of a matrix, and in the last step we have used the assumption that $\tilde{\boldsymbol{\chi}}(m)$ and $\tilde{\boldsymbol{\chi}}(n)$ are independent for $m \neq n$, and therefore, the expected value of the product is equal to the product of the expected values. Hence,

$$E\left[\left\|\hat{\boldsymbol{R}}_{\tilde{\chi}\tilde{\chi}}(N) - \boldsymbol{R}_{\tilde{\chi}\tilde{\chi}}\right\|^2\right] = O\big(1/N\big). \tag{3.9}$$

Using the boundedness of $\|\boldsymbol{R}_{\tilde{\chi}\tilde{\chi}}^{-1}\|^2$ and that of $\|\hat{\boldsymbol{R}}_{\tilde{\chi}\tilde{\chi}}^{-1}(N)\|^2$ with probability 1, together with Eqs (3.7) and (3.9) we obtain

$$E\left[\left\|\boldsymbol{R}_{\tilde{\chi}\tilde{\chi}}^{-1}-\hat{\boldsymbol{R}}_{\tilde{\chi}\tilde{\chi}}^{-1}(N)\right\|^2\right]=O\big(1/N\big),\qquad \text{for } N\geq 2MLN_h. \tag{3.10}$$

Following similar steps as in Eq. (3.9) and noting that $\boldsymbol{\rho}_{\tilde{\chi}r}=\boldsymbol{\rho}_{\tilde{\chi}\tilde{s}_r}=\boldsymbol{\rho}_{\tilde{\sigma}\tilde{s}_r}$, we can also see that

$$E\left[\left\|\hat{\boldsymbol{\rho}}_{\tilde{\chi}\tilde{s}_r}(N)-\boldsymbol{\rho}_{\tilde{\sigma}\tilde{s}_r}\right\|^2\right]=O\big(1/N\big)\quad\text{and}\quad E\left[\left\|\hat{\boldsymbol{\rho}}_{\tilde{\chi}r}(N)-\boldsymbol{\rho}_{\tilde{\sigma}\tilde{s}_r}\right\|^2\right]=O\big(1/N\big). \tag{3.11}$$

Now, from Eqs (3.2) and (3.5), we have

$$\begin{aligned}E\left[\left\|\hat{\boldsymbol{\eta}}_T(N)-\boldsymbol{\eta}_{opt}\right\|^2\right]&=E\left[\left\|\boldsymbol{R}_{\tilde{\chi}\tilde{\chi}}^{-1}\left(\hat{\boldsymbol{\rho}}_{\tilde{\chi}s_r}(N)-\boldsymbol{\rho}_{\tilde{\sigma}\tilde{s}_r}\right)+\left(\hat{\boldsymbol{R}}_{\tilde{\chi}\tilde{\chi}}^{-1}(N)-\boldsymbol{R}_{\tilde{\chi}\tilde{\chi}}^{-1}\right)\hat{\boldsymbol{\rho}}_{\tilde{\chi}\tilde{s}_r}(N)\right\|^2\right]\\&\leq 2E\left[\left\|\boldsymbol{R}_{\tilde{\chi}\tilde{\chi}}^{-1}\right\|^2\left(\left\|\hat{\boldsymbol{\rho}}_{\tilde{\chi}\tilde{s}_r}(N)-\boldsymbol{\rho}_{\tilde{\sigma}\tilde{s}_r}\right\|^2\right)+\left(\left\|\hat{\boldsymbol{R}}_{\tilde{\chi}\tilde{\chi}}^{-1}(N)-\boldsymbol{R}_{\tilde{\chi}\tilde{\chi}}^{-1}\right\|^2\right)\left\|\hat{\boldsymbol{\rho}}_{\tilde{\chi}\tilde{s}_r}(N)\right\|^2\right]\end{aligned} \tag{3.12}$$

where we have used the inequality $\|\boldsymbol{a}+\boldsymbol{b}\|^2\leq 2\left(\|\boldsymbol{a}\|^2+\|\boldsymbol{b}\|^2\right)$ followed by the Schwarz inequality. Similarly, we have

$$E\left[\left\|\hat{\boldsymbol{\eta}}_B(N)-\boldsymbol{\eta}_{opt}\right\|^2\right]\leq 2E\left[\left\|\boldsymbol{R}_{\tilde{\chi}\tilde{\chi}}^{-1}\right\|^2\left(\left\|\hat{\boldsymbol{\rho}}_{\tilde{\chi}r}(N)-\boldsymbol{\rho}_{\tilde{\sigma}\tilde{s}_r}\right\|^2\right)+\left(\left\|\hat{\boldsymbol{R}}_{\tilde{\chi}\tilde{\chi}}^{-1}(N)-\boldsymbol{R}_{\tilde{\chi}\tilde{\chi}}^{-1}\right\|^2\right)\left\|\hat{\boldsymbol{\rho}}_{\tilde{\chi}r}(N)\right\|^2\right]. \tag{3.13}$$

Using the boundedness of $\|\boldsymbol{R}_{\tilde{\chi}\tilde{\chi}}^{-1}\|$ together with Eqs. (3.10) and (3.11) in Eqs (3.12) and (3.13), we conclude that

$$E\left[\left\|\hat{\boldsymbol{\eta}}_T(N)-\boldsymbol{\eta}_{opt}\right\|^2\right]=O\big(1/N\big)\quad\text{and}\quad E\left[\left\|\hat{\boldsymbol{\eta}}_B(N)-\boldsymbol{\eta}_{opt}\right\|^2\right]=O\big(1/N\big). \tag{3.14}$$

From Eq. (3.14) we can conclude that as N gets large, both $\hat{\boldsymbol{\eta}}_T(N)$ and $\hat{\boldsymbol{\eta}}_B(N)$ converge in the mean-square sense to $\boldsymbol{\eta}_{opt}$ at a rate which is $O\big(1/N\big)$. □

4 Reception of BPSK Signals Using CAST Processors

In this section, we provide an algebraic performance analysis of a binary phase-shift-keying (BPSK) transmission system which employs CAST processors under the effect of co-channel inter-

ference. (This analysis will be confirmed by simulations in Section 5.) The analyses for both the T-CAST and B-CAST processors are similar, being different only in the values of the parameters. The analysis can also be extended to reception of other digital signals in a straightforward manner. A discrete-time baseband equivalent model for the transmission and reception of the signals is depicted in Fig. 3, where we have assumed that the received signal is sampled at N_b times the baud rate. The sampled desired signal is then given by

$$s(n) = \sum_{k=-\infty}^{\infty} b(k)p(n - kN_b), \tag{4.1}$$

where we have assumed that the centre frequency of the baseband model is the carrier frequency of the desired signal and that the carrier has been acquired by the receiver. Here, $p(n) = p_c(t)|_{t=nT_b/N_b}$, where $p_c(t)$ is the impulse response of the (continuous-time) pulse shaping filter and T_b is the baud period. We assume that $\alpha_{\ell m}/(2\pi)$ is a factor of the sampling frequency; i.e., $\alpha_{\ell m} = 2\pi k N_b/T_b$, for some integer k. The message data $b(k) = \pm 1$ is an independent binary sequence for which $+1$ and -1 occur with equal probability. If $\boldsymbol{d}(\theta_s)$ denotes the array manifold of the sensors for the desired signal arriving from an angle θ_s, then the received desired signal vector at the array containing L sensors is given by

$$\boldsymbol{s}(n) = [s_1(n), \ldots, s_L(n)]^T = s(n)\boldsymbol{d}(\theta_s). \tag{4.2}$$

Similarly the ith interfering binary signal overlapping in spectrum with the desired signal can be modeled as

$$u_i(n) = A_{u_i} \sum_{k=-\infty}^{\infty} b_{u_i}(k)p_i(n - kN_b)\cos(\Delta\omega_{u_i} n + \phi_i), \tag{4.3}$$

where $p_i(n) = p_c(t)|_{t=nT_b/N_b-\tau_i}$, and A_{u_i} is the amplitude, $\Delta\omega_{u_i}$ the frequency off-set, ϕ_i the phase off-set and τ_i the timing offset of the interference, and $b_{u_i}(k) = \pm 1$ is an independent sequence. Here the interference has the same baud rate, $1/T_b$, and pulse shape, $p_c(t)$, as the desired signal, but has a different carrier frequency and is not necessarily phase nor symbol synchronous with the desired signal. This model is common in multi-user mobile communication scenarios, and can be easily extended to signals having different baud rates and different pulse shapes without affecting the analysis. If θ_{u_i} is the direction of arrival (DOA) of the ith interference, then the interference

vector at the L-sensor array is given by

$$\boldsymbol{u}_i(n) \triangleq [u_{i_1}(n), \ldots, u_{i_L}(n)]^T = u_i(n)\boldsymbol{d}(\theta_{u_i}). \tag{4.4}$$

In general, there may be a total of I of such interfering signals. Thus, the vector of received samples at the input to the CAST processor is given by

$$\boldsymbol{x}(n) = \boldsymbol{s}(n) + \sum_{i=1}^{I} \boldsymbol{u}_i(n) + \boldsymbol{\nu}(n), \tag{4.5}$$

where $\boldsymbol{\nu}(n)$ is the vector of noise samples at the input of the processor. These samples are complex-valued and are assumed to be zero-mean, Gaussian, jointly stationary, and spatially white so that

$$E\left[\boldsymbol{\nu}(n)\right] \triangleq E\left\{[\nu_1(n)\cdots\nu_L(n)]^T\right\} = \mathbf{0} \quad \text{and} \quad E\left[\boldsymbol{\nu}(n_1)\boldsymbol{\nu}^\dagger(n_2)\right] = \text{diag}\{\sigma_{\nu_\ell}^2 r_\ell(n_1 - n_2)\}, \tag{4.6}$$

where $\nu_\ell(n)$ is the complex noise process at the input to the ℓth FRESH filter in the CAST processor, with $\sigma_{\nu_\ell}^2$ being its power, and $r_\ell(m)$ its normalized temporal correlation. The temporal correlation is induced by oversampling the output of the bandpass "radio-frequency (RF)" filter which is implicit in the baseband equivalent model in Fig. 3, [22]. It becomes negligible as the bandwidth of the RF filter approaches N_b/T_b.

To simplify the analysis, we invoke the usual assumption [17] that the filter coefficient vector $\hat{\boldsymbol{\eta}}(N)$ defined in Eq. (3.2) for the T-CAST and B-CAST processors is uncorrelated with the noise $\nu(n)$, the desired symbols $b(n)$, and the interfering symbols $b_{u_i}(n)$. This assumption has been verified to be valid by extensive computer simulation testing of the correlation coefficients [19].

When the input $\boldsymbol{x}(n)$ arrives at the CAST processor, it is frequency shifted by $\alpha_{\ell m}$ at the mth branch of the ℓth FRESH filter. Here, we choose the set of cycle frequencies for each FRESH filter to be the same, i.e., $\alpha_{\ell m} = \alpha_{\lambda m}$. Hence, from Fig. 2, the output of the ℓth FRESH filter is given by

$$y_\ell(n) = \sum_{m=1}^{M} \sum_{k} \left[h_{\ell m}(n-k)e^{j\alpha_{\ell m}k}x_\ell(k) + h'_{\ell m}(n-k)e^{j\alpha_{\ell m}k}x_\ell^*(k)\right] \tag{4.7}$$

and the output of the CAST processor (Fig. 1) is given by

$$y(n) = \sum_{\ell=1}^{L} y_\ell(n). \tag{4.8}$$

We now refer our discussion to the transmitter/receiver system shown in Fig. 3. After passing through the CAST processor at the receiver, the signal is demodulated and processed by a filter $p(-n)$ matched to the pulse shaping filter $p(n)$ at the desired transmitter, to produce

$$z(n) = \sum_k p(k-n)y(k). \tag{4.9}$$

Substituting Eqs.(4.7) and (4.8) into Eq. (4.9) and simplifying, we obtain

$$z(nN_b) = \sum_{\ell=1}^{L} \sum_{k=0}^{N_h+N_p-1} \left[g_\ell(k)x_\ell(nN_b-k) + g'_\ell(k)x^*_\ell(nN_b-k) \right], \tag{4.10}$$

where N_h and N_p are the lengths of the FIR filters and the receiver matched filter respectively,

$$g_\ell(n) = \sum_{m=1}^{M} \sum_{i=0}^{N_p-1} h_{\ell m}(n-i)e^{-j\alpha_{\ell m}n}\, p(-i), \tag{4.11}$$

and $g'_\ell(n)$ is defined similarly, with $h'_{\ell m}(k)$ replacing $h_{\ell m}(k)$. Using Eqs. (4.1), (4.2) and (4.5), we see that the input to the threshold can be written as

$$\hat{b}(n) = \gamma_o b(n) + \xi_{ISI}(n) + \xi_{CT}(n) + \xi_\nu(n) \triangleq \gamma_o b(n) + \zeta_e(n), \tag{4.12}$$

where

$$\xi_{ISI}(n) = \sum_{\substack{i=-\infty \\ i\neq n}}^{\infty} b(i)\gamma(n-i), \tag{4.13}$$

$$\gamma(n) = \sum_{\ell=1}^{L} \sum_{k=0}^{N_h+N_p-1} p(nN_b-k)\,\mathrm{Re}\left[g_\ell(n)d_\ell(\theta_s) + g'_\ell(n)d^*_\ell(\theta_s) \right], \tag{4.14}$$

$$\xi_{CT}(n) = \sum_{i=1}^{I} \mathrm{Re}\left[\sum_{\ell=1}^{L} \sum_{k=0}^{N_h+N_p-1} \left[g_\ell(k)u_{i\ell}(nN_b-k) + g'_\ell(k)u^*_{i\ell}(nN_b-k) \right] \right], \tag{4.15}$$

$$\xi_\nu(n) = \mathrm{Re}\left[\sum_{\ell=1}^{L} \sum_{k=0}^{N_h+N_p-1} \left[g_\ell(k)\nu_\ell(nN_b-k) + g'_\ell(k)\nu^*_\ell(nN_b-k) \right] \right], \tag{4.16}$$

Re[$\cdot$] denotes the real part, $\gamma_o \equiv \gamma(0)$ and $d_\ell(\theta_s)$ is the ℓth element of the array manifold at the DOA of the desired signal. We note that the length of $\gamma(n)$ is given by $K_\gamma = \lceil (N_h+2N_p-2)/N_b \rceil$, where $\lceil x \rceil$

denotes the least integer $\geq x$. We observe that input to the threshold device in Eq. (4.12) consists of four terms: (i) the desired bit $b(n)$ scaled by a random variable γ_o; (ii) the intersymbol interference $\xi_{ISI}(n)$ caused by the previous bits of the desired signal; (iii) the cross-talk $\xi_{CT}(n)$ caused by the co-channel signals $u_i(n)$ which arrive from angles θ_{u_i}; and (iv) the processed channel noise $\xi_\nu(n)$. We note that the scaled desired bit $\gamma_o b(n)$ and the intersymbol interference $\xi_{ISI}(n)$ depend on different bits. By the assumption of independence of bits, these two terms are uncorrelated. We further note that the last three terms of interference and noise are from different sources and are therefore independent of each other. We now examine each of the terms in Eq. (4.12). Taking the expected value of Eq. (4.16), since $E[\nu_\ell(n)] = 0$, we have $E\left[\xi_\nu(n)\right] = 0$. If the temporal correlation in (4.6) is negligible and if each sensor has the same noise power, $\sigma^2_{\nu_\ell} = \sigma^2_\nu$, then the variance of $\xi_\nu(n)$ can be calculated in a straightforward manner:

$$\sigma^2_{\xi\nu} = E\left[\xi^2_\nu(n)\right] = \frac{\sigma^2_\nu}{2}\sum_{\ell=1}^{L}\sum_{k=0}^{N_h+N_p-1} E\Big\{\mathrm{Re}^2\left[g_\ell(k)\right] + 2\mathrm{Re}\left[g_\ell(k)\right]\mathrm{Re}\left[g'_\ell(k)\right] + \mathrm{Re}^2\left[g'_\ell(k)\right] + \mathrm{Im}^2\left[g_\ell(k)\right] + 2\mathrm{Im}\left[g_\ell(k)\right]\mathrm{Im}\left[g'_\ell(k)\right] + \mathrm{Im}^2\left[g'_\ell(k)\right]\Big\}, \quad (4.17)$$

where Im[·] denotes the imaginary part. An equivalent expression for the case where the RF filter is narrow enough to induce significant temporal correlation in (4.6) can be calculated in a similar way, but is omitted for brevity.

Using Eqs. (4.13) and (4.15) and the independence assumption [17] we can conclude that $E\left[\xi_{ISI}(n)\right] = 0$ and $E\left[\xi_{CT}(n)\right] = 0$. The variance of the ISI component can be obtained using Eq. (4.13) and is given by

$$\sigma^2_{ISI} = E\left[\xi_{ISI}(n)\xi^*_{ISI}(n)\right] = \sum_{k=1}^{K_\gamma} E\left[\left|\gamma(k)\right|^2\right]. \quad (4.18)$$

The variance of the cross-talk interference can be obtained as follows: Using Eq. (4.3) in Eq. (4.15), and after much simplifying, we have

$$\xi_{CT}(n) = \sum_{i=1}^{I}\sum_{m} b_{u_i}(n-m)\beta_i(m,n), \quad (4.19)$$

where

$$\beta_i(m,n) \triangleq A_{u_i} \sum_{\ell=1}^{L} \sum_{k=1}^{N_h+N_p-1} p_i(mN_b-k)\mathrm{Re}\Big[g_\ell(k)\cos(\Delta\omega_{u_i}(nN_b-k)+\phi_i)d_\ell(\theta_{u_i}) + g'_\ell(k)\cos(\Delta\omega_{u_i}(nN_b-k)+\phi_i)d^*_\ell(\theta_{u_i})\Big], \tag{4.20}$$

with $d_\ell(\theta_{u_i})$ being the ℓth element of the array manifold of the sensors for the ith interfering signal. We note that $\beta_i(m,n)$ is periodic in n having a period $Q_i = \min\{n : nN_b\Delta\omega_{u_i}/(2\pi) \text{ is an integer}\}$ and its length in m is finite being equal to $N_{\beta_i} = \lceil (N_h + 2N_{p_i} - 2)/N_b \rceil$. From Eq. (4.19) we obtain

$$\sigma^2_{CT}(n) = E\left[\xi_{CT}(n)\xi^*_{CT}(n)\right] = \sum_{i=1}^{I} E\left[\sum_{m=0}^{N_{\beta_i}-1} \beta_i(m,n)\beta^*_i(m,n)\right], \tag{4.21}$$

where we have used the fact that $E[b_{u_{i_1}}(n-m_1)b^*_{u_{i_2}}(n-m_2)] = \delta_{m_1,m_2}\delta_{i_1,i_2}$. Since $\beta_i(m,n)$ is periodic in n, we can obtain an averaged value of the cross-talk over the period yielding

$$\bar{\sigma}^2_{CT} = \sum_{i=1}^{I} \frac{1}{Q_i} \sum_{n=0}^{Q_i-1} \sigma^2_{CT}(n). \tag{4.22}$$

Since $\xi_\nu(n)$, $\xi_{ISI}(n)$, $\xi_{CT}(n)$ are results of linear combinations of many random variables from different sources, we can approximate the sum of them by a Gaussian random variable by virtue of the Generalized Central Limit Theorem for a correlated sequence of random variables [20]. Each of these variables are of zero mean and their variances are given by Eqs. (4.17), (4.18) and (4.22), respectively. Thus the combined noise and interference term $\zeta_e(n)$ in Eq. (4.12) can be approximated by a zero-mean Gaussian random variable with variance given by

$$\sigma^2_{\zeta_e} = \sigma^2_{\xi_\nu} + \sigma^2_{ISI} + \bar{\sigma}^2_{CT}. \tag{4.23}$$

This approximation is common in the performance analysis of digital communication systems in the presence of ISI and cross-talk. For the current application, the Gaussianity of $\zeta_e(n)$ has be verified by extensive simulations [19].

We now turn our attention to the signal term $\gamma_o b(n)$ in Eq. (4.12) . This is the information bit

$b(n)$ multiply by a random variable γ_o. From Eq.(4.14) we obtain

$$E[\gamma_o] = \sum_{\ell=1}^{L} \sum_{k=1}^{N_o+N_2-1} p(-k)\mathrm{Re}\left\{E\left[g_\ell(0)d_\ell(\theta_s) + g'_\ell(0)d^*_\ell(\theta_s)\right]\right\}. \tag{4.24}$$

By definition, the variance of γ_o can be evaluated using $\sigma^2_{\gamma_o} = E[\gamma_o^2] - E^2[\gamma_o]$. From the Generalized Central Limit Theorem, we can approximate γ_o by a Gaussian random variable. To show that $E[\gamma_o] > 0$ theoretically may prove to be mathematically intractable. However, through many simulation examples with different pulse-shaping filters, DOAs, carrier frequencies, and FRESH filter lengths, we have observed that γ_o is indeed Gaussian with positive mean. Fig. 4 shows an example of the histograms of γ_o (for 100,000 runs) for different DOAs in a typical operating scenario. The scenario has a single interferer with 30% spectral overlap of the desired signal [as defined in Eq. (5.1) below], sample size $N = 15$, input signal-to-noise ratio (SNR) of 20 dB and input signal-to-interference ratio (SIR) of 0 dB. Since the amplitude of the desired signal has been normalized to one in our model, we have that $\mathrm{SNR} = 1/\sigma^2_\nu$ and $\mathrm{SIR} = 1/\left(\sum_{i=1}^{I} A^2_{u_i}\right)$. The mean square error d^2 between these histograms of γ_o and their Gaussian approximation is given in Table 1, along with the the sample means and variances of the histograms. It can be observed that the value of the sample mean is positive and increases as the difference in the DOAs of the signal and interference increases. From the squared error values, it can be seen that the assumption of Gaussian distribution for γ_o is very accurate. Many other simulations have been carried out under different conditions. It is observed that [19]:

i) The Gaussian approximation of γ_o is very accurate in all cases.

ii) The sample mean of γ_o is positive in all cases and increases with increasing data length, increasing DOA difference and decreasing frequency overlap.

iii) The sample variance of γ_o is essentially independent of the DOA difference and it decreases with increasing data length and decreasing frequency overlap.

From the assumption of independence [17] we can conclude that γ_o and $\zeta_e(n)$ are uncorrelated Gaussian random variables. From Eq. (4.12) we can see that both γ_o and $\zeta_e(n)$ constitute interference on the detected symbol $\hat{b}(n)$. We now examine the effect of these interferences on, the probability of error, P_e. An error occurs if $b(n) = 1$ and $\hat{b}(n) < 0$, or if $b(n) = -1$ and $\hat{b}(n) > 0$.

	0°	2°	5°	10°
$E[\gamma_o]$	0.4223	0.4435	0.4632	0.4696
$\mathrm{Var}[\gamma_o]$	0.8×10^{-4}	2.4×10^{-4}	2.3×10^{-4}	2.2×10^{-4}
d^2	6.9×10^{-5}	5.7×10^{-5}	3.2×10^{-5}	3.0×10^{-5}

Table 1: The mean and the variance of γ_o, and the mean square error, d^2, between the experimental histogram of γ_o and its Gaussian approximation, for different DOAs in the scenario described in Section 4.

Assuming equally likely transmission of $+1$ and -1, we have

$$\begin{aligned} P_e &= \frac{1}{2}\mathrm{Prob}\big(-\gamma_o + \zeta_e(n) > 0 \,\big|\, b(n) = -1\big) + \frac{1}{2}\mathrm{Prob}\big(\gamma_o + \zeta_e(n) < 0 \,\big|\, b(n) = 1\big) \\ &= \frac{1}{2}\left[\int_{-\infty}^{\infty} p(\gamma_o) \int_{\gamma_o}^{\infty} p(\zeta_e)\, d\zeta_e\, d\gamma_o + \int_{-\infty}^{\infty} p(\gamma_o) \int_{-\infty}^{-\gamma_o} p(\zeta_e)\, d\zeta_e\, d\gamma_o\right] \\ &= \int_{-\infty}^{\infty} p(\gamma_o) \int_{\gamma_o}^{\infty} \frac{1}{\sqrt{2\pi\sigma_{\zeta_e}^2}} \exp\left(\frac{-\zeta_e^2}{2\sigma_{\zeta_e}^2}\right) d\zeta_e\, d\gamma_o, \end{aligned} \tag{4.25}$$

where in the last step, we have use the symmetry of $p(\zeta_e)$ about $\zeta_e = 0$. After some changes of variables, it can be shown that [19]

$$P_e = \frac{1}{2}\mathrm{erfc}\left(\sqrt{\frac{(E[\gamma_o])^2}{2\sigma_{\gamma_o}^2 + 2\sigma_{\zeta_e}^2}}\right), \tag{4.26}$$

where $\mathrm{erfc}(x) = \frac{2}{\sqrt{\pi}} \int_x^{\infty} e^{-z^2} dz$. Eq. (4.26) indicates that the higher the mean value of γ_o, the lower the probability of error. On the other hand, the higher the power of the interference ζ_e or the higher the variance of γ_0, the higher the P_e. Since Eqs. (2.13) and (2.16) are identical in form, the above error analysis is applicable to both T-CAST and B-CAST processors (the differences appear in the definitions of γ_o and ζ_e).

5 Simulations

We now present some computer simulation examples in which the B-CAST and T-CAST processors are employed to extract the desired signal from interference using the receiver shown in Fig. 3. In all the examples shown here, we assume that there is a desired signal and a single interfering signal, both of the BPSK type, given by Eqs. (4.1) and (4.3), respectively. The signals have the same baud period and the same (continuous-time) square root raised cosine pulse shaping filter

with 100% roll-off factor [21]. For simplicity, we assume that the interference is phase and symbol synchronous with the desired user. In all cases we use a uniform linear array with $L = 3$ sensors for the receiver.

Example 5.1 We firstly present some computer simulation results illustrating the convergence properties of the T-CAST and B-CAST processors. Here, the baud rate is 5000 symbols-per-second, and the carrier frequencies of the desired signal and interference are 10 kHz and 17 kHz, respectively, so the spectral overlap is 30%. The spectral overlap is defined as

$$\frac{B_s + B_i - |f_i - f_s|}{2B_s} \times 100\% \tag{5.1}$$

where B_s and B_i are the baud rates and f_s and f_i are the carrier frequencies of the signal and interference, respectively. There are $M = 2$ branches in each FRESH filter, and each branch contains an FIR filter of length $N_h = 6$. The frequency shift parameters for the FRESH filters are chosen to be $\alpha_{\ell 1}/(2\pi) = 20$ kHz, $\alpha_{\ell 2}/(2\pi) = -20$ kHz, for $\ell = 1, 2, 3$. These are cycle frequencies of the desired signal. The frequency shift parameter for the reference signal for both the B-CAST and T-CAST processors is chosen to be $\alpha_r = 0$. Fig. 5 shows the convergence of $E\left[\|\hat{\boldsymbol{\eta}}_T - \boldsymbol{\eta}_{opt}\|^2\right] / \|\boldsymbol{\eta}_{opt}\|^2$ and $E\left[\|\hat{\boldsymbol{\eta}}_B - \boldsymbol{\eta}_{opt}\|^2\right] / \|\boldsymbol{\eta}_{opt}\|^2$, against the number of received data symbols, each being averaged over 20 realizations. The desired signal DOA in this scenario is 0° (i.e., normal to the array), the received SNR is 10 dB and the received SIR is 0 dB. (The optimal weight vector $\boldsymbol{\eta}_{opt}$, was calculated using Eq. (3.5) using a time-average of 500 symbols.) It is observed that both the T-CAST and B-CAST processors converge to the optimum processor when the number of samples is large, with the T-CAST processor being the faster in convergence. As a further comparison, the output signal-to-interference-and-noise (power) ratio (SINR) of both the T-CAST and B-CAST processors for various DOA differences are plotted in Fig. 6. We also plotted the output SINR of the corresponding processor using the C-CAB [13] and SCORE [12, 13] beamforming algorithms (not facilitated with FRESH filtering). It is observed that while the T-CAST and B-CAST algorithms both produce acceptable output SINR, the performance of C-CAB and SCORE are much inferior, especially when the signal and interference are close in space, a scenario in which C-CAB and SCORE fail completely. This example confirms the validity of Theorem 3.1 on the convergence of T-CAST and B-CAST.

Example 5.2 In this example, we examine the effect of data length on the performance of the receiver employing the B-CAST and T-CAST processors. The basic scenario is the same as that in Ex. 5.1, with the DOA of the interference being set to 2°. The bit error rate (BER) curves for both B-CAST and T-CAST processors using $15, 25, 50$ and 150 received data symbols were found by averaging over 10,000 simulation runs. These simulated curves, together with the analytical expression for the probability of error in Eq. (4.26) are shown in Fig. 7. It can be observed the averaged BER agrees very well with the analytical expression. It can also be seen that the larger the data length, the closer are the performance of the B-CAST processor to that of the T-CAST processors, again confirming the convergence analysis in Section 3.

Example 5.3 In this example, we examine the finite sample BER of the receiver employing the B-CAST and T-CAST processors, when the signal and interference vary in their DOA. Here, the basic scenario is the same as that in Ex. 5.1. The number of received data symbols is fixed at 15, while the DOA of the interference varies amongst $\theta_u = 0^\circ$, 2°, 5°, and 10°. The experiments were repeated 10,000 times and the averaged BERs at different SNRs are plotted in Fig. 8. The analytical expression for the probability of error in Eq. 4.26 is also shown in each case. It can be observed that when the interference moves from 0° to 5° the improvement in BER is marginal, but that it is more significant when the DOA of the interference is at 10°. The improvement will be increased if the number of sensors in the array is increased as confirmed by other simulations [19].

Example 5.4 We now examine the effect of spectral overlap between the signal and interference. The basic scenario is once again the same as that in Ex. 5.1. Since the baud rate of each signal is 5000 symbols-per-second, they each have a bandwidth of about 10 kHz. We fix the carrier frequency for the desired signal at 10 kHz while varying that for the interference so that the percentages of spectral overlap are 40%, 30%, 20% and 10%. The number of received data symbols is fixed at 15, and the DOA of the interference is either 2° (Fig. 9) or 10° (Fig. 10). The experiments were repeated 10,000 times and the average BER curves are plotted in Figs 9 and 10. It can be observed from these figures that decreasing the spectral overlap between signal and interference improves performance. Furthermore, increasing the physical separation between signal and interference improves performance, especially when the spectral overlap is relatively large. Other simulations have been carried out and it has been observed that if the spectral overlap between signal and interference is relatively large, an increase in number of sensors (L) and/or an increase in the FIR filter

length (N_h) will also improve the performance [19].

6 Conclusions

In this paper, we have proposed a class of adaptive space-time processors based on the cyclostationarity properties of signals. These processors can be used with or without (blind) a training signal leading to realizations called T-CAST and B-CAST, respectively. The advantage of employing the T-CAST or B-CAST processors over using beamforming or filtering alone is that when the spectral overlap between the signal and interference is large, or when they are close to each other in their direction of arrival, these processors can still extract the signal from the interference satisfactorily. Moreover, the B-CAST technique has the further advantage of not having to need a training signal.

Analysis showed that under moderate data lengths, the performance of B-CAST and T-CAST converge to that of the ideal Wiener solution. The space-time processors were then applied to a binary digital communication receiver and their performance analyzed. It has been shown that their performance depends on the data length, the DOA of the signal and inferences, as well as on the percentage of spectral overlap between the signal and interference.

In view of their ability to extract the desired signal from interference, especially when the signal and interferer have large spectral overlap or when their DOAs are close, the T-CAST and B-CAST processors offer an attractive alternative to beamforming or filtering alone.

References

[1] W. A. Gardner, "Exploitation of spectral redundancy in cyclostationary signals," *IEEE Acoust., Speech, Sig. Proc. Mag.*, pp. 14-36, April 1991.

[2] W. A. Gardner, *Statistical Spectral Analysis: A Non-probabilistic Theory.* Prentice-Hall, Englewood Cliffs, NJ, 1988.

[3] W. A. Gardner, "Cyclic Wiener filtering: Theory and method," *IEEE Trans. Commun.*, vol. 41, pp. 151-163, Jan. 1993.

[4] A. Chevreuil and P. Loubaton, "On the use of conjugate cyclostationarity: A blind second-order multi-user equalization method," *Proc. Int. Conf. Acoust., Speech, Sig. Proc.*, pp. 2439-2442, Atlanta, GA, May 1996.

[5] J. Zhang, K. M. Wong, Z. Q. Luo and P. C. Ching, "Blind adaptive FRESH filtering for signal extraction," *IEEE Trans. Sig. Proc.*, vol. 47, no. 5, pp. 1397-1402, May 1999.

[6] G. B. Giannakis, "Filter-banks for blind channel identification and equalization," *IEEE Sig. Proc. Lett.*, vol. 4, pp. 184-187, June 1997.

[7] A. Chevreuil and P. Loubaton, "Blind second-order identification of FIR channels: Forced cyclostationarity and structured subspace method," *IEEE Sig. Proc. Lett.*, vol. 4, pp. 204-206, July 1997.

[8] E. Serpedin and G. B. Giannakis, "Blind channel identification and equalization with modulation induced cyclostationarity," *IEEE Trans. Sig. Proc.*, vol. 46, no. 7, pp. 1930–1944, July 1998.

[9] J. H. Winters, "Optimum combining in digital mobile radio with co-channel interference," *IEEE Trans. Veh. Technol.*, vol. 33, pp. 144-155, Aug. 1984.

[10] S. C. Swales, M. A. Beach, D. J. Edwards and J. P. McGeehan, "The performance enhancement of multibeam adaptive base station antennas for cellular land mobile radio systems," *IEEE Trans. Veh. Technol.*, vol. 39, pp. 56-67, Feb. 1990.

[11] B. G. Agee, S. V. Schell, and W. A. Gardner, "Spectral self-coherence restoral: A new approach to blind adaptive signal extraction using antenna arrays," *Proc. IEEE,* vol. 78, pp. 753-767, April 1990.

[12] S. V. Schell and W. A. Gardner, "Maximum likelihood and common factor analysis-based blind adaptive spatial filtering for cyclostationary signals," *Proc. Int. Conf. Acoust., Speech, Sig. Proc.*, vol. IV, Minneapolis, MN, April 1993.

[13] Q. Wu and K. M. Wong, "Blind adaptive beamforming for cyclostationary signals," *IEEE Trans. Sig. Proc.*, vol. 44, no. 11, pp. 2757-2767, Nov. 1996.

[14] W. A. Gardner, S. V. Schell, and P. Murphy, "Multiplication of cellular capacity by blind adaptive spatial filtering," *Proc. Int. Conf. Selected Topics Wireless Commun.*, Vancouver, Canada, June 1992.

[15] R. Ho, Q. Wu, and K. M. Wong, "Implementation of cyclic beamforming techniques on mobile communication systems," *Proc. 2nd Workshop Cyclostationary Signals*, pp. 16.1-16.9, Monterey, CA, Aug. 1994.

[16] B. D. Van Veen and K. M. Buckley, "Beamforming: A versatile approach to spatial filtering," *IEEE Acoust., Speech, Sig. Proc. Mag.*, pp. 4–24, Apr. 1988.

[17] S. Haykin, *Adaptive Filter Theory*. 2nd Ed., John Wiley and Sons, NY, 1991.

[18] G. H. Golub and C. F. Van Loan, *Matrix Computations*. Johns Hopkins University Press, 1983.

[19] J. Zhang, "Blind adaptive space-time signal processing." PhD Dissertation, Dept. Elec. & Comp. Eng., McMaster University, 2001.

[20] W. Feller, *An Introduction to Probability Theory and its Applications*. John Wiley and Sons, New York, 1950.

[21] M. C. Jeruchim, P. Balaban, and K. S. Shanmugan, *Simulation of Communication Systems*. Plenum Press, NY, 1992.

[22] D. K. Borah, R. A. Kennedy, Z. Ding, and I. Fijalkow, "Sampling and prefiltering effects on blind equalizer design", *IEEE Trans. Sig. Proc.*, vol. 49, no. 1, pp. 209-218, Jan. 2001.

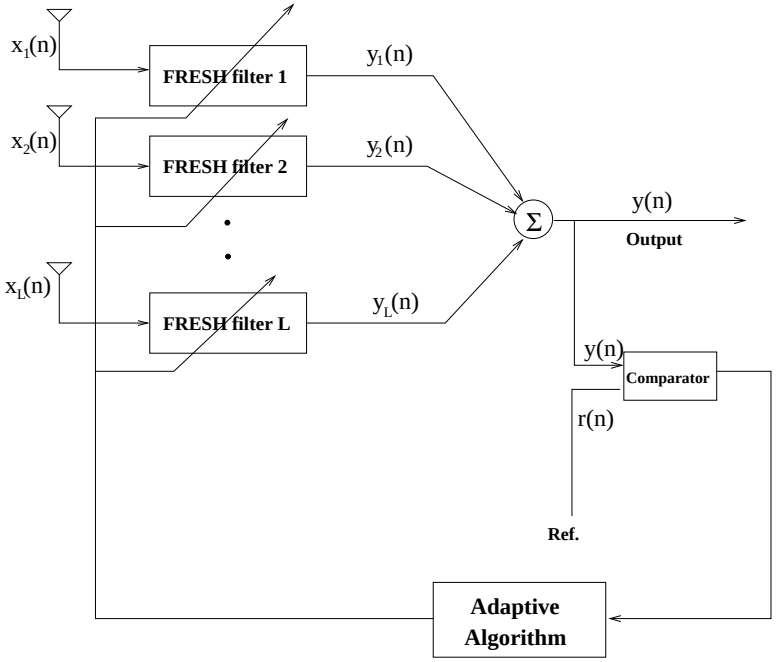

Figure 1: General structure of the CAST processor

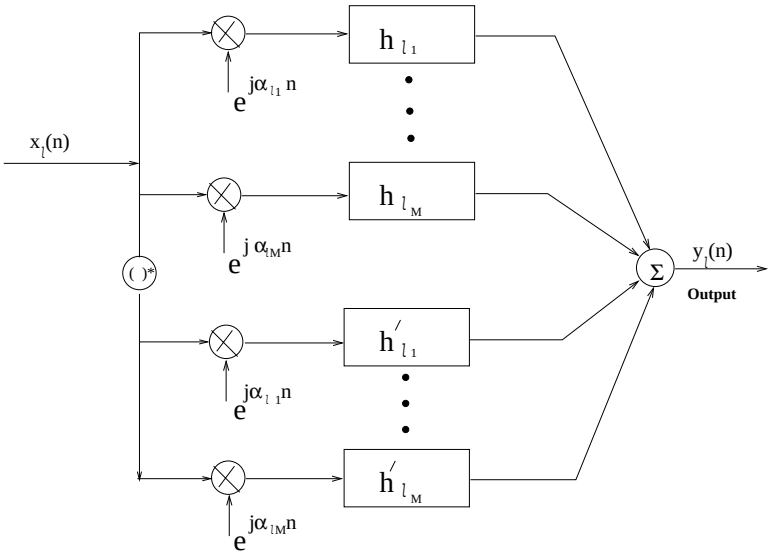

Figure 2: A FRESH filter

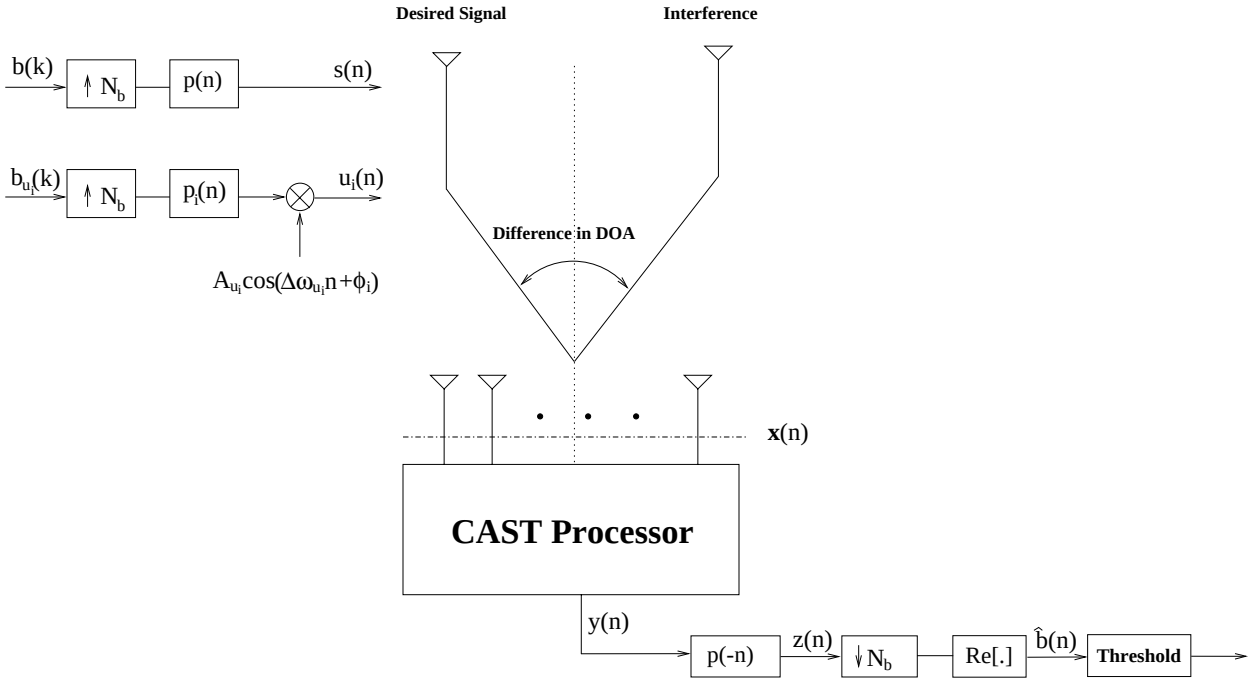

Figure 3: Discrete-time baseband equivalent model of a multi-user BPSK system with a CAST processor.

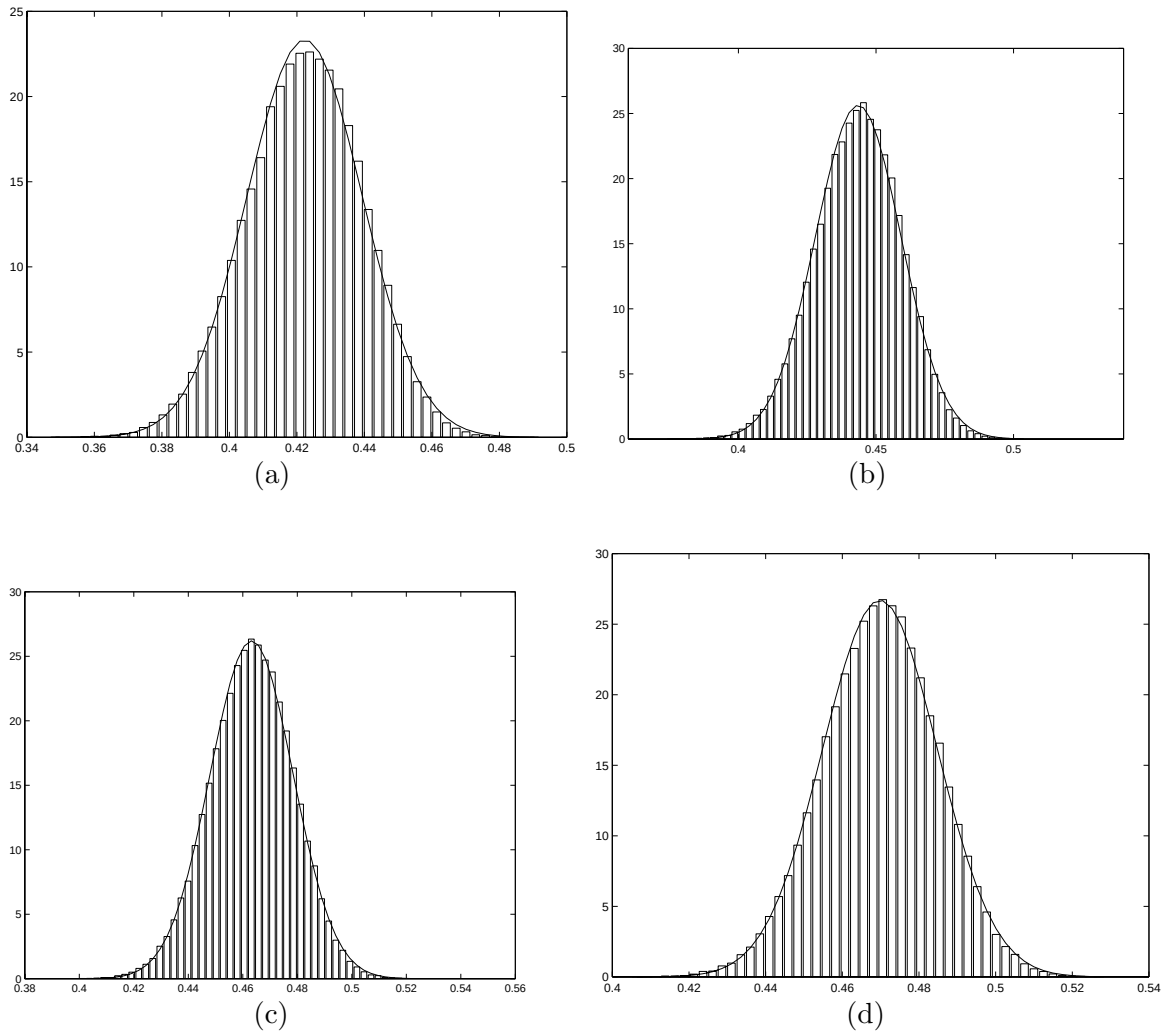

Figure 4: Comparison between the experimental histogram for γ_o and its Gaussian approximation for DOA differences of (a) 0°, (b) 2°, (c) 5°, (d) 10° in the scenario described in Section 4.

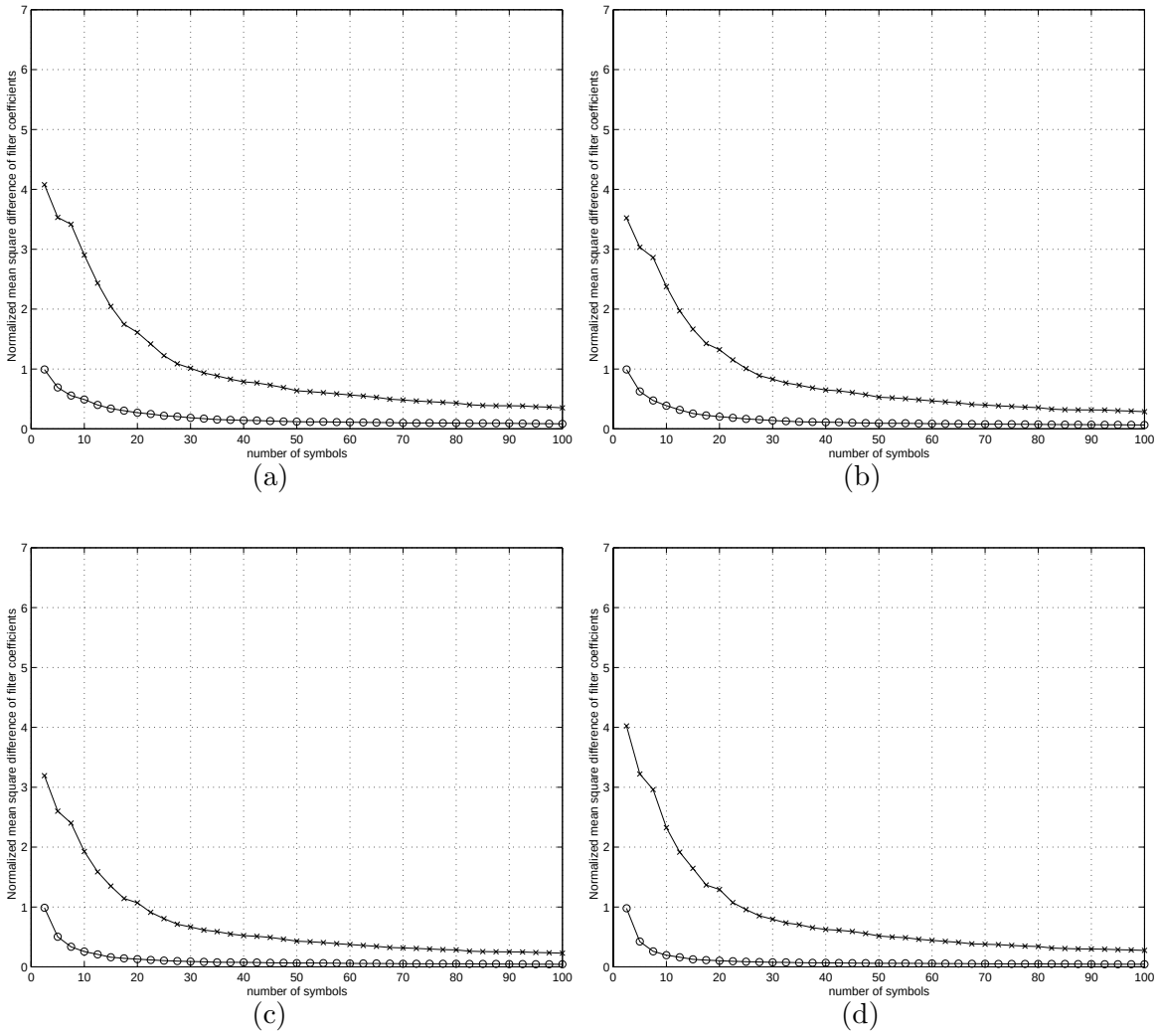

Figure 5: Normalized filter coefficient convergence of B-CAST and T-CAST for Ex. 5.1, with DOA differences of (a) 0°, (b) 2°, (c) 5°, and (d) 10°.

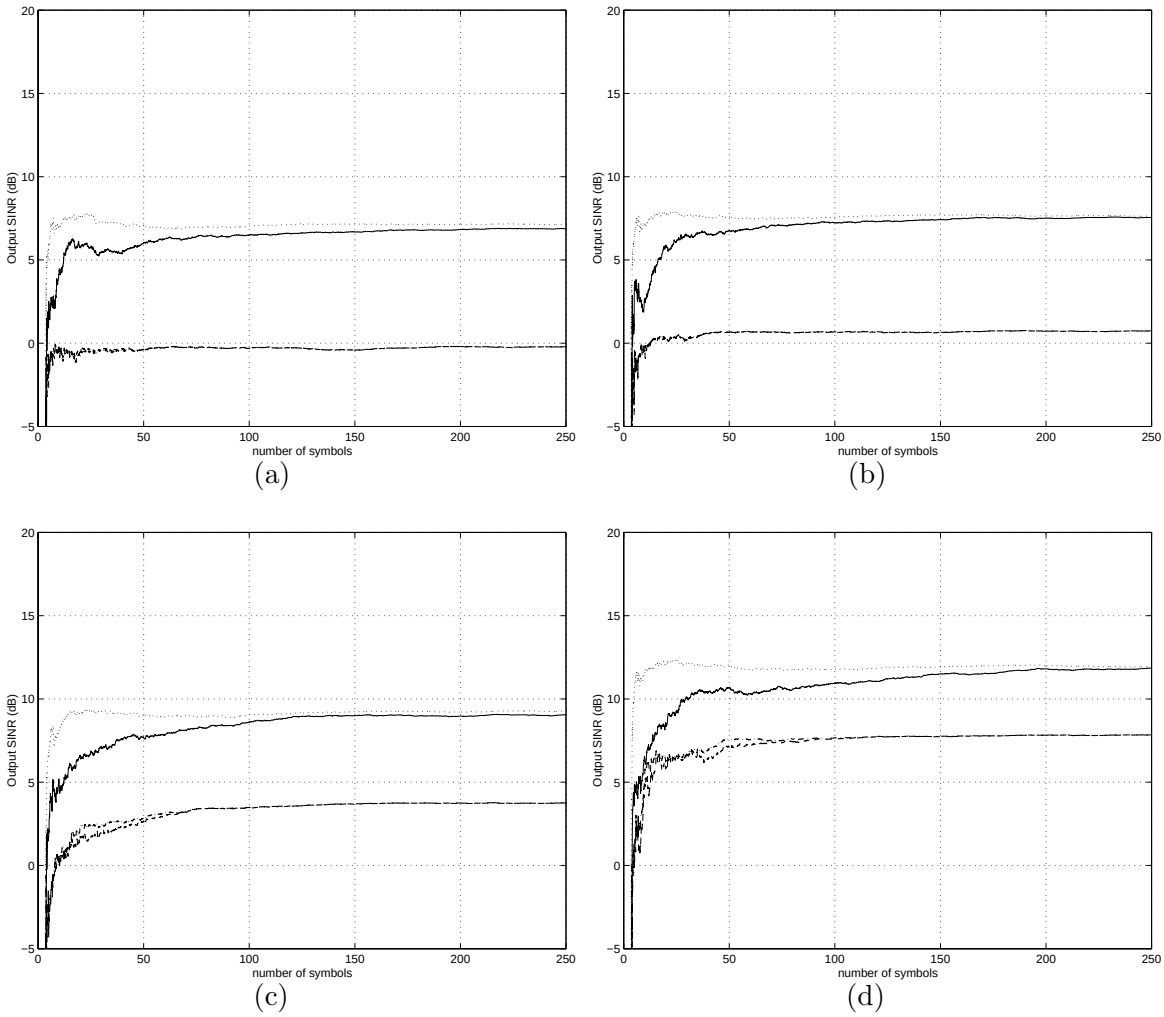

Figure 6: Output SINR of B-CAST (solid), T-CAST (dotted), C-CAB (dash-dot), and SCORE (dashed) algorithms against the number of received data symbols for Ex. 5.1, with DOA differences of (a) 0°, (b) 2°, (c) 5°, and (d) 10°.

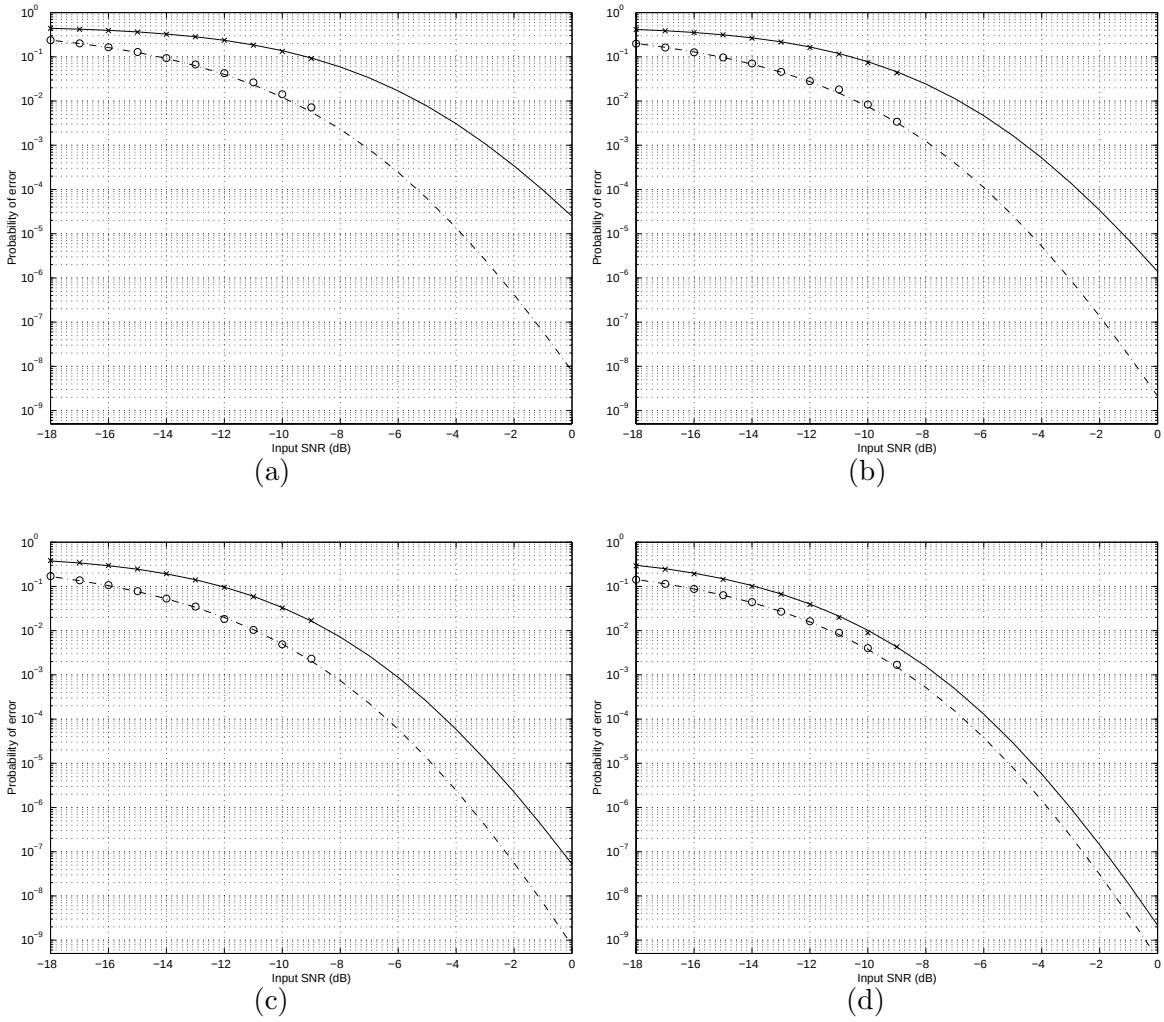

Figure 7: Probability of error (analytical and simulation) for B-CAST (solid and $\times$) and T-CAST (dash-dot and $\circ$) against SNR for Ex. 5.2, with the number of received data symbols being (a) 15, (b) 25, (c) 50, and (d) 150.

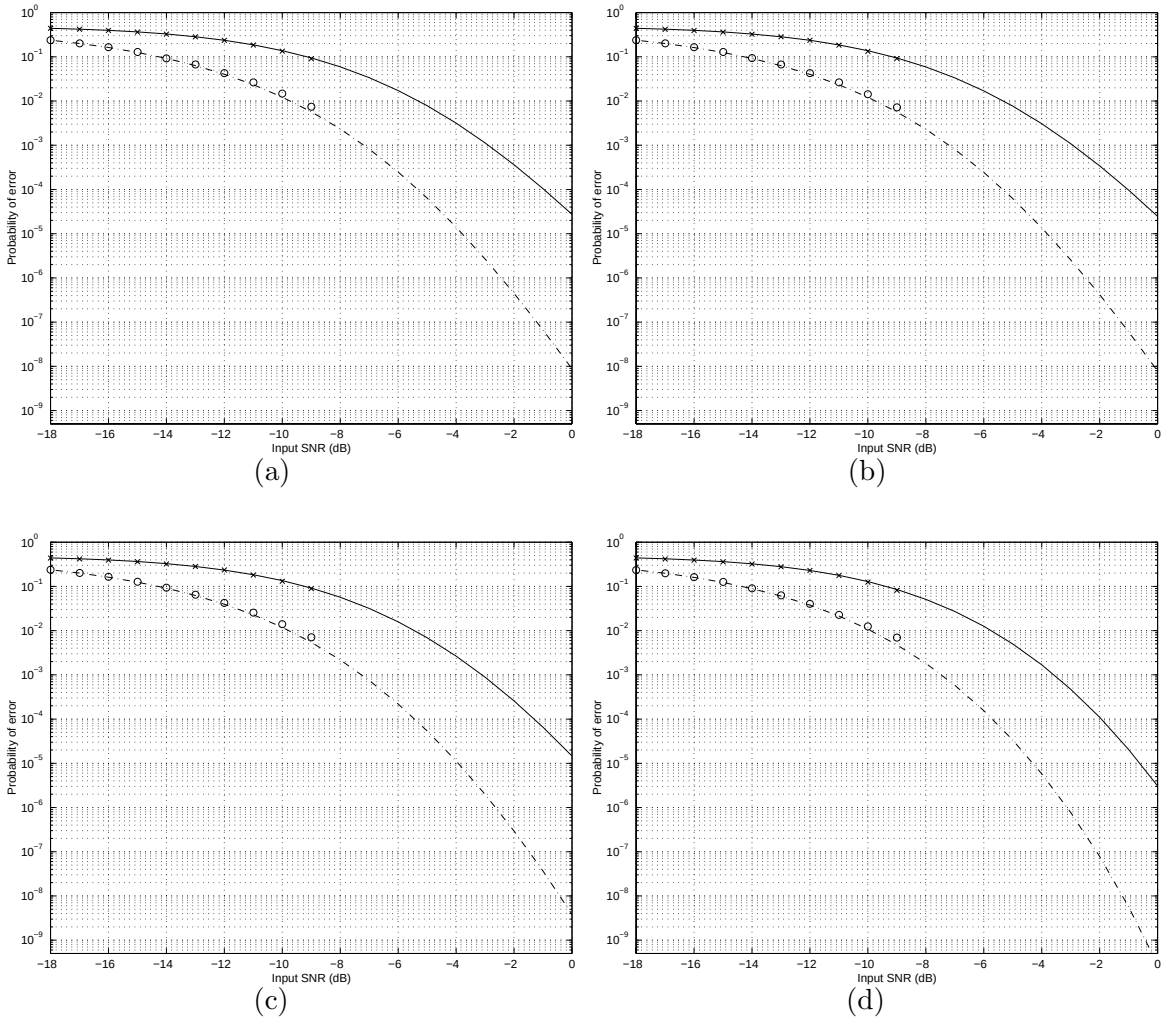

Figure 8: Probability of error (analytical and simulation) for B-CAST (solid and $\times$) and T-CAST (dash-dot and $\circ$) algorithms against SNR for Ex. 5.3, with DOA differences of (a) 0°, (b) 2°, (c) 5°, and (d) 10°.

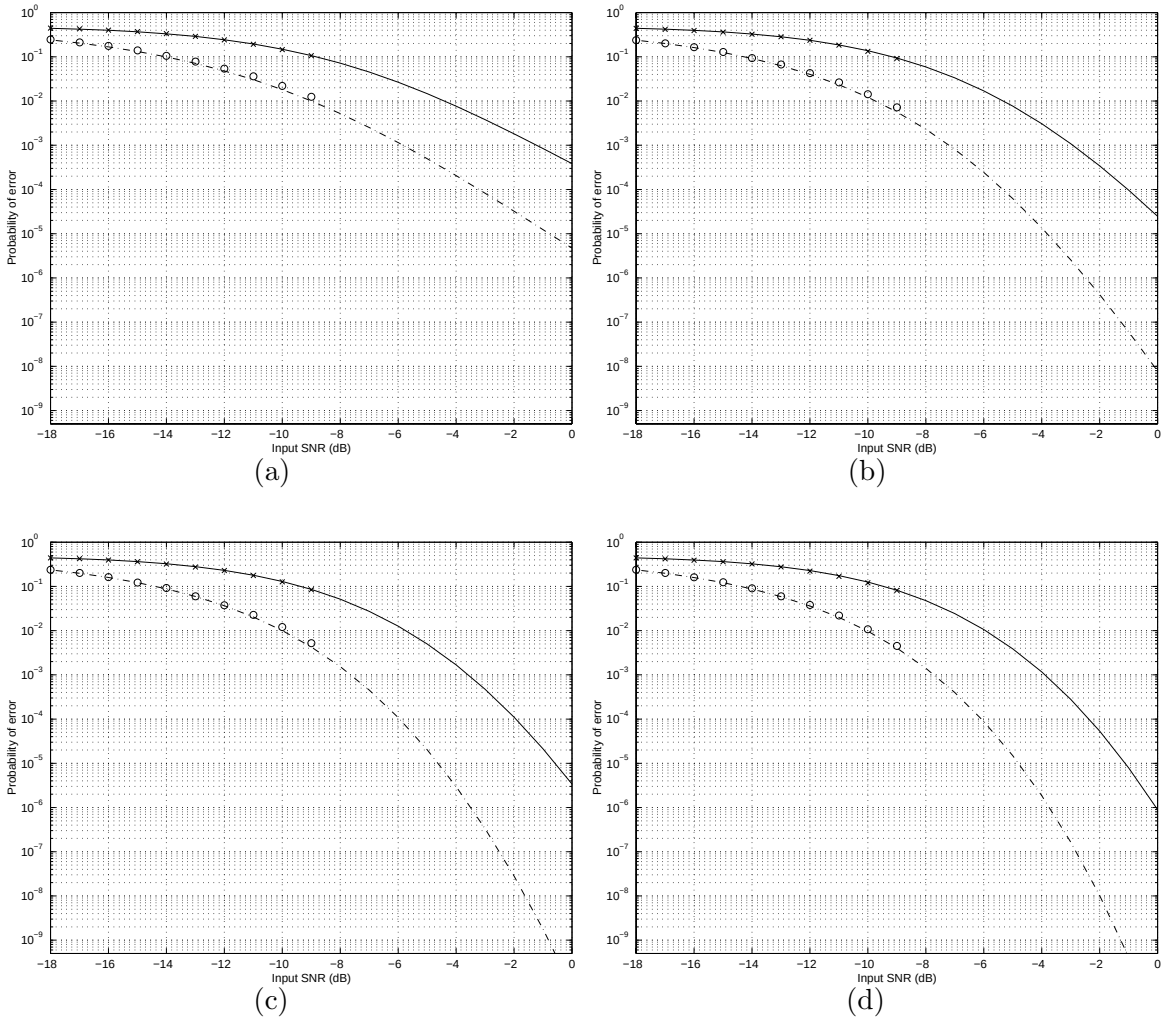

Figure 9: Probability of error (analytical and simulation) for B-CAST (solid and $\times$) and T-CAST (dash-dot and $\circ$) algorithms against SNR for Ex. 5.4 with a DOA difference of 2°, for frequency overlappings of (a) 40%, (b) 30%, (c) 20%, and (d) 10%.

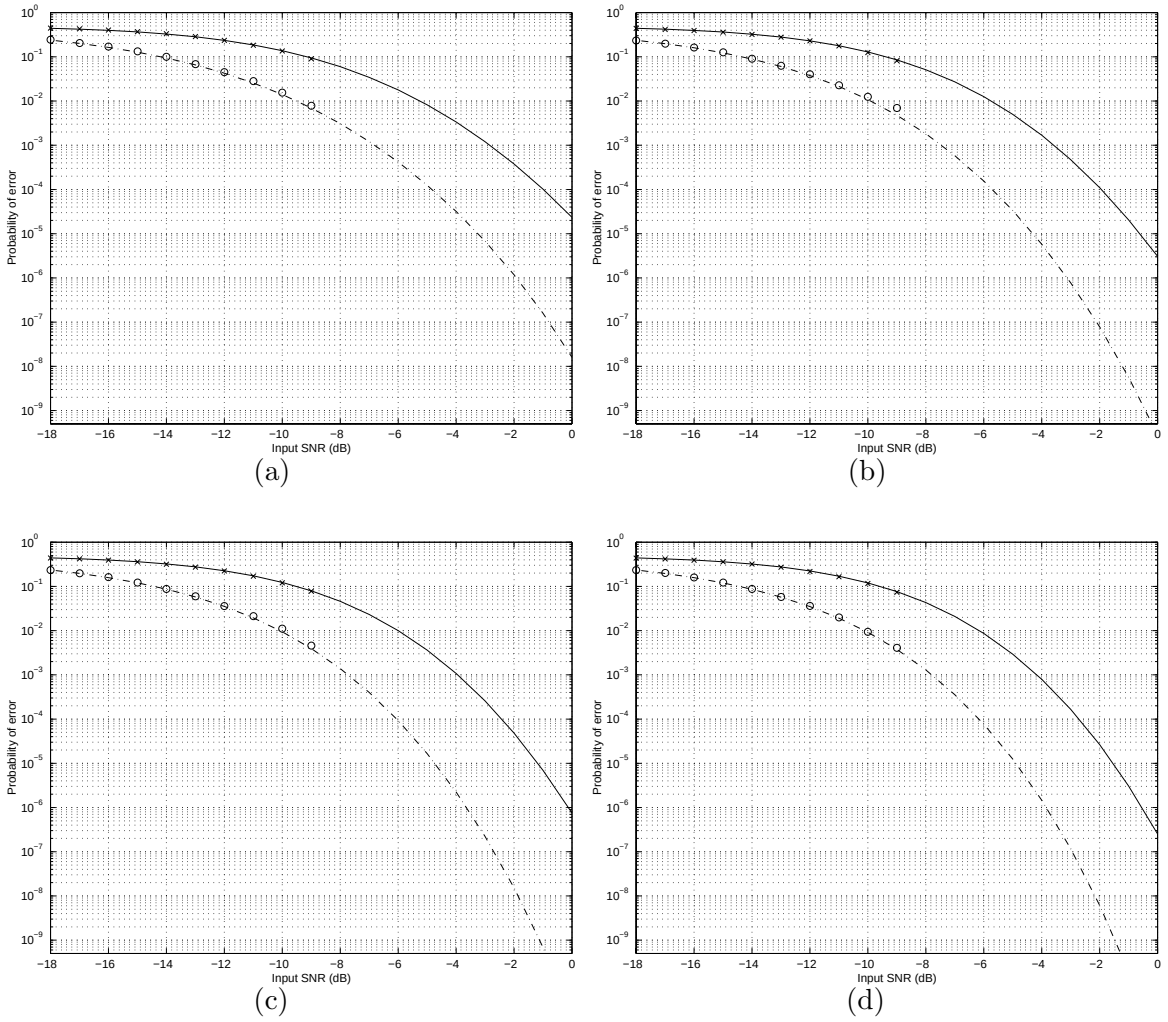

Figure 10: Probability of error (analytical and simulation) for B-CAST (solid and $\times$) and T-CAST (dash-dot and $\circ$) algorithms against SNR for Ex. 5.4 with a DOA difference of 10°, for frequency overlappings of (a) 40%, (b) 30%, (c) 20%, and (d) 10%.