RESEARCH   **Open Access**

Multi-antenna spectrum sensing by exploiting spatio-temporal correlation

Sadiq Ali[1,2*], David Ramírez[3], Magnus Jansson[4], Gonzalo Seco-Granados[1] and José A López-Salcedo[1]

Abstract

In this paper, we propose a novel mechanism for spectrum sensing that leads us to exploit the spatio-temporal correlation present in the received signal at a multi-antenna receiver. For the proposed mechanism, we formulate the spectrum sensing scheme by adopting the generalized likelihood ratio test (GLRT). However, the GLRT degenerates in the case of limited sample support. To circumvent this problem, several extensions are proposed that bring robustness to the GLRT in the case of high dimensionality and small sample size. In order to achieve these sample-efficient detection schemes, we modify the GLRT-based detector by exploiting the covariance structure and factoring the large spatio-temporal covariance matrix into spatial and temporal covariance matrices. The performance of the proposed detectors is evaluated by means of numerical simulations, showing important advantages over existing detectors.

Keywords: Cognitive radio; Generalized likelihood ratio test (GLRT); Multi-antenna; Spatio-temporal correlation

1 Introduction

For a cognitive radio system, which opportunistically accesses the wireless channel, spectrum sensing becomes a crucial task for detecting the presence of primary user transmissions [1]. Recently, sensors with multiple antennas have become an integral part of many cognitive receivers [2,3], thus giving us the chance to consider multi-antenna techniques to improve the performance of spectrum sensing. Multiple antennas can offer spatial diversity and improve the spectrum sensing performance [4,5]. Intuitively, the presence of any primary signal should result in some spatial correlation in the observations received at the multi-antenna receivers [5] and thus, exploiting this correlation improves the primary user detection performance. In addition to being spatially correlated, the received signal samples are usually correlated (wide-sense stationary) in time due to presence of a temporal dispersive channel, oversampling of the received signals or just because the originally transmitted signals are correlated in time

[5,6]. This spatio-temporal correlation is a feature that can be used for detection purposes, since the remaining (i.e. undesired) noise processes at different antennas can be safely assumed statistically independent, both in time and space.

Spectrum sensing methods that only exploit the spatial structure of the received signal covariance matrix have been of great interest in the recent years [5]. The majority of these schemes are based on multivariate statistical inference theory [7,8], and interested readers can find comprehensive details in ([8], Ch. 9-10), which discusses multivariate detectors for testing the independence of random observations with the help of the generalized likelihood ratio test (GLRT). These GLRT-based detectors typically end up in a simple quotient between the determinant of the sample covariance matrix and the determinant of its diagonal version, and these tests have been widely applied to the detection of signals especially in the context of cognitive radios [2,9].

Through careful study of various existing spectrum sensing techniques, one can conclude that the signal's temporal correlation is not fully exploited in most of these techniques. In fact, in most of them, temporal correlation is ignored or considered as a deleterious effect. In the very few works that exploit the temporal correlation, they usually assume some prior knowledge about it [10-14].

*Correspondence: engsat@gmail.com
[1]SPCOMNAV, Universitat Autònoma de Barcelona, 08193 Bellaterra, Barcelona, Spain
[2]Electrical Engineering, University of Engineering and Technology, Peshawar, 25000 Peshawar, Pakistan
Full list of author information is available at the end of the article

One of the reasons for ignoring or only partially exploiting temporal correlation is that it often makes it difficult to achieve tractable solutions. However, the exploitation of temporal correlation jointly with spatial correlation can provide us extra side information to enhance the detection performance. Hence, it will be interesting to find ways to devise detection mechanisms that exploit temporal correlation with tractable solutions. Taking this into account, the main focus of this work is to develop a multi-antenna detector that robustly exploits spatio-temporal correlation. In order to do so, we propose a signal model that leads us to tractable detection schemes while exploiting the temporal correlation jointly with spatial correlation.

In order to achieve the proposed goal, we start with our earlier work in [15], where we derived the GLRT. This detector basically tests whether the covariance matrix is block diagonal or not. To make the discussion and notation simple, for our case, we call this GLRT scheme the spatio-temporal GLRT (ST-GLRT). Compared to the traditional spatial covariance-based GLRT, the ST-GLRT provides some improved performance. The reason for this is that the ST-GLRT scheme exploits temporal correlation as an additional feature on top of spatial correlation and energy. However, since any GLRT involves the estimation of unknown parameters (e.g., covariance matrix), its performance depends on the sample size and the dimensionality of the signal model. In practice, the GLRT is used based on the assumption that the sample size is large compared to the model dimension. When this is not the case, the performance of the GLRT degenerates because the sample covariance matrix becomes singular, and the whole problem becomes ill-conditioned [16,17]. In the case of the ST-GLRT, we have to deal with both the spatial and temporal dimensions, and hence, the overall data dimension is even larger. Thus, the ST-GLRT has some further limitations when the detection process requires a quick decision, as it is in the case of the detection of primary signals in cognitive radio. Hence, although for the large sample support, the ST-GLRT can certainly achieve an improved detection performance; for small sample support, it has some limitations that deserve a detailed study.

In order to reduce the demand for large sample support and bring robustness to the ST-GLRT, one may assume existence of some underlying structure based on the spatial and temporal components of the covariance matrix. In our earlier work [15,18], by assuming wide sense stationary (WSS), we exploited the Toeplitz structure of the covariance matrix. Doing so, we proposed an approximated GLRT in the frequency domain that leads to robustness against the small sample support. Contrary to that work, in the present one, we rather focus on exploiting the covariance structures without the GLRT

approximation in the frequency domain. In particular, we approximate the block-Toeplitz structure of a multivariate WSS process as persymmetric. Doing so, we will take advantage of the result in [19] which states that by exploiting the persymmetric structure, the number of independent vector measurements required for the covariance matrix estimator can be decreased by up to a factor of two. This will certainly bring down the demand for high sample support required for the ST-GLRT to not degenerate and provide robustness against the repercussions of small sample support and large dimensional data.

Moving one step forward, we also approximate the block-Toeplitz structure of the spatio-temporal correlated multi-antenna measurements with a Kronecker product structure [20,21]. Recently, exploitation of the Kronecker structure in covariance matrices has received a lot of interest in statistics [22,23]. Moreover, the maximum likelihood (ML) method for estimating the covariance matrix based on the Kronecker product has been previously discussed in [22,23]. Similarly, in the cases where the correlation structure is not separable through the Kronecker product, [24] discusses some insightful details about the nearest Kronecker product approximations. In this paper, we use the Kronecker structure to reformulate the ST-GLRT by taking advantage of the inherent spatio-temporal structure of the received observations. In order to do so, we adopt and extend our earlier work [25] by using the Kronecker product to efficiently exploit the space-time correlation in a multi-antenna spectrum sensing scheme. In addition to the Kronecker product-based factorization, we also exploit the fact that the factored matrices could have additional persymmetric structure [21]. Therefore, by exploiting the Kronecker product structure jointly with the persymmetric structure, the performance of the proposed detection scheme can further be improved in terms of the required number of sample to estimate the covariance matrix (i.e. the detector efficiency).

To compare the proposed methods with traditional techniques, numerical simulations are conducted. These results illustrate that the proposed detection schemes indeed outperform the traditional approaches, especially in the case of small sample support.

The remainder of the paper is organized as follows. Section 2 introduces the proposed methodology and the signal model. In Section 3, we solve the problem by using the traditional GLRT formulations. The proposed detection schemes are derived in Sections 4 and 5. Numerical results are provided in Section 6. The conclusion is finally drawn in Section 7.
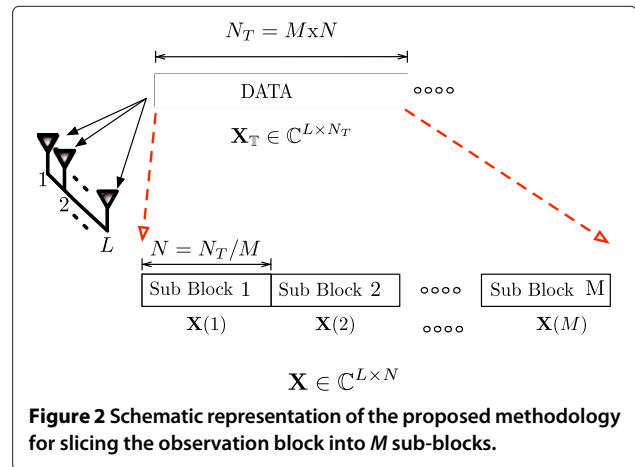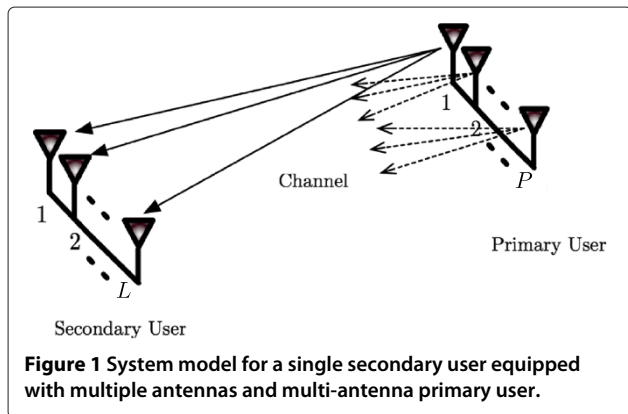
2 Proposed methodology and signal model

Herein, we address the problem of detecting the presence of a primary user by a single cognitive radio that

is equipped with L antennas as shown in Figure 1. We assume no prior knowledge about the primary transmission or the noise processes except that the noise is spatio-temporally independent. We focus on a practical scenario where due to the presence of a primary user (PU) signal, the received signals at the L antennas are correlated in space as well as time. The spatial correlation is due to the proximity between the receiving antennas and the temporal correlation is due to the presence of a temporal dispersive channel, oversampling of the received signal or the time correlation of the transmitted signal [6]. In this work, we consider the temporal correlation to be WSS. Hence, consecutive received samples of the $(L \times 1)$ vector $\mathbf{x}(n)$, $n = 1, \dots, N_T$ at L antennas of a user are temporally correlated, where $\mathbf{x}(n) \triangleq [x_1(n), x_2(n), \dots, x_L(n)]^T$ and N_T is the total number of received vector samples. Exploiting the temporal correlation, jointly with spatial correlation, can provide us extra information to enhance the detection performance. In order to devise detection mechanisms that exploit temporal correlation with tractable solutions, the proposed technique can be described as follows:

1. We split the received block of N_T vectors $\mathbf{x}(n)$ into M sub-blocks where each block contains N samples of vector $\mathbf{x}(n)$, as shown in Figure 2.
2. We assume that the consecutive vector-valued samples within each sub-block are temporally correlated with $N \times N$ temporal correlation matrix $\mathbf{C}_t$.
3. Independence is assumed between consecutive sub-blocks.

It is clear that the proposed approach does not exploit the correlation among different sub-blocks. However, as it will become clear in the following sections, exploiting such correlation would result in an ill-posed problem. This would make it intractable to derive the ML estimators of the covariance matrices, required for the generalized likelihood ratio test (GLRT). The proposed

Figure 1 System model for a single secondary user equipped with multiple antennas and multi-antenna primary user.

Figure 2 Schematic representation of the proposed methodology for slicing the observation block into M sub-blocks.

mechanism could also be motivated from the concept of correlated block-fading channel, which presents correlation in each N-samples block, but independence between consecutive blocks [6,26].

In order to proceed, we need the distribution of $\{\mathbf{x}(n)\}_{n=1}^{N_T}$, and we assume it to be zero-mean complex Gaussian[a]. This assumption is particularly reasonable if the primary network employs orthogonal frequency division multiplexing (OFDM) as modulation format [15,18]. Similarly, we also assume that the SU collects N_T consecutive samples of vector $\mathbf{x}(n)$. Based on the proposed three step approach discussed above, the received N_T vector samples are divided into M blocks and each block consists of N vector samples, such that $MN = N_T$. The m-th block is defined as:

$$
\mathbf{X}(m)
$$
$$
= \begin{bmatrix} x_1(1+(m-1)N) & x_1(2+(m-1)N) & \cdots & x_1(N+(m-1)N) \\ x_2(1+(m-1)N) & x_2(2+(m-1)N) & \cdots & x_2(N+(m-1)N) \\ \vdots & \vdots & \ddots & \vdots \\ x_L(1+(m-1)N) & x_L(2+(m-1)N) & \cdots & x_L(N+(m-1)N) \end{bmatrix}
$$
$$
= \begin{bmatrix} \mathbf{x}_1^T(m) \\ \mathbf{x}_2^T(m) \\ \vdots \\ \mathbf{x}_L^T(m) \end{bmatrix},
$$
$$(1)$$

where the i-th row, $\mathbf{x}_i^T(m)$ for $m = 1, \dots, M$, contains N-samples at the i-th antenna. Let us define a vector $\mathbf{z}(m) \triangleq$ vec $\left(\mathbf{X}^T(m)\right)$. The covariance matrix of the $LN \times 1$ vector $\mathbf{z}(m)$ under hypothesis $\mathcal{H}_1$ is

$$
\mathbf{\Sigma} = \mathrm{E}\left[\mathbf{z}\mathbf{z}^T\right] = \begin{bmatrix} \mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12} & \cdots & \mathbf{\Sigma}_{1L} \\ \mathbf{\Sigma}_{21} & \mathbf{\Sigma}_{22} & \cdots & \mathbf{\Sigma}_{2L} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{\Sigma}_{L1} & \mathbf{\Sigma}_{L2} & \cdots & \mathbf{\Sigma}_{LL} \end{bmatrix} \in \mathbb{C}^{LN \times LN},
$$
$$(2)$$

where the sub-block covariance matrices $\boldsymbol{\Sigma}_{il} = \mathrm{E}\left[\mathbf{x}_i \mathbf{x}_l^T\right] \in \mathbb{C}^{N \times N}$, $1 \le i, l \le L$ in $\boldsymbol{\Sigma}$ capture all space-time second-order information about the random vector $\{\mathbf{x}_i\}_{i=1}^L$. Thus, the hypothesis testing problem becomes

$$
\begin{aligned}
\mathcal{H}_0 &: \mathbf{z} \sim \mathcal{CN}\left(\mathbf{0}, \boldsymbol{\Sigma}_0\right), \\
\mathcal{H}_1 &: \mathbf{z} \sim \mathcal{CN}\left(\mathbf{0}, \boldsymbol{\Sigma}\right).
\end{aligned}
\tag{3}
$$

Now, considering that the noise powers at the L antennas of the user are different and spatially and temporally uncorrelated noise, the covariance matrix under noise-only hypothesis is $\boldsymbol{\Sigma}_0 = \boldsymbol{\Sigma}_{A,0} \otimes \mathbf{I}_N$, where

$$
\boldsymbol{\Sigma}_{A,0} = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & \sigma_L^2 \end{bmatrix},
\tag{4}
$$

and $\sigma_i^2, i = 1, 2, \ldots, L$ is noise power at i-th antenna of the array.

3 GLRT based on spatio-temporal correlation

In this section, by adopting the GLRT formulation, we review a spectrum sensing scheme that exploits both the spatial and temporal correlation without exploiting any a-priori structure. This will be done just to provide a reference benchmark to compare the proposed detectors in Sections 4.2 and 5. Since, the parameters $\{\boldsymbol{\Sigma}_0, \boldsymbol{\Sigma}\}$ are unknown, we need to adopt the GLRT approach and the test statistic of the ST-GLRT can be formulated as:

$$
\Lambda_{\mathrm{ST}}(\mathbf{Z}) = \frac{\max\limits_{\boldsymbol{\Sigma}_0} f_{\mathbf{z}}(\mathbf{Z}; \boldsymbol{\Sigma}_0)}{\max\limits_{\boldsymbol{\Sigma}} f_{\mathbf{z}}(\mathbf{Z}; \boldsymbol{\Sigma})} \gtrless_{\mathcal{H}_1}^{\mathcal{H}_0} \gamma,
\tag{5}
$$

where $f_{\mathbf{z}}(\mathbf{Z}, \boldsymbol{\Sigma}_0)$ and $f_{\mathbf{z}}(\mathbf{Z}; \boldsymbol{\Sigma})$ are the likelihood functions under hypothesis $\mathcal{H}_0$ and $\mathcal{H}_1$, respectively. We assume that we have M independent blocks of the data $\mathbf{X}$, or equivalently vector $\mathbf{z}$, $\mathbf{Z} \triangleq \left[\mathbf{z}(1) \, \mathbf{z}(2), \cdots, \mathbf{z}(M)\right]$ available. To solve the GLRT (5), we have to derive the ML estimators of the parameters for each of the hypotheses. Note that under weak conditions, the ML estimator is asymptotically an unbiased and efficient estimator [27]. The expression for the likelihood function $f_{\mathbf{z}}(\mathbf{Z}; \boldsymbol{\Sigma})$ under hypothesis $\mathcal{H}_1$ can then be written as:

$$
f_{\mathbf{z}}(\mathbf{Z}; \boldsymbol{\Sigma}) = \frac{1}{(\pi)^{\mathrm{MLN}} |\boldsymbol{\Sigma}|^M} \exp\left\{-M\mathrm{tr}\left(\boldsymbol{\Sigma}^{-1} \hat{\boldsymbol{\Sigma}}\right)\right\}.
\tag{6}
$$

Taking into account that $\boldsymbol{\Sigma}$ has no further structure beyond being positive semi-definite, it is easy to prove [7] that its ML estimate is given by the sample covariance matrix i.e. $\hat{\boldsymbol{\Sigma}} = \frac{1}{M} \sum_{m=1}^M \mathbf{z}(m) \mathbf{z}^H(m)$. Under the alternative hypothesis, the expression for $f_{\mathbf{z}}(\mathbf{Z}, \boldsymbol{\Sigma}_0)$, ignoring constant factors, is

$$
\begin{aligned}
f_{\mathbf{z}}(\mathbf{Z}, \boldsymbol{\Sigma}_0) &= f_{\mathbf{z}}\left(\mathbf{Z}; \boldsymbol{\Sigma}_{A,0} \otimes \mathbf{I}_N\right) \\
&= \left|\boldsymbol{\Sigma}_{A,0} \otimes \mathbf{I}_N\right|^{-M} \\
&\quad \times \exp\left[-\sum_{m=1}^M \mathbf{z}(m)\left(\boldsymbol{\Sigma}_{A,0}^{-1} \otimes \mathbf{I}_N\right) \mathbf{z}^H(m)\right].
\end{aligned}
\tag{7}
$$

Now, taking into account the Kronecker structure of the covariance matrix, (7) may be rewritten as [25]:

$$
\begin{aligned}
f_{\mathbf{z}}(\mathbf{Z}, \boldsymbol{\Sigma}_0) &= f_{\mathbf{x}}\left(\mathbf{X}_{\mathbb{T}}; \boldsymbol{\Sigma}_{A,0}, \mathbf{I}_N\right) \\
&= \left|\boldsymbol{\Sigma}_{A,0}\right|^{-MN} \left|\mathbf{I}_N\right|^{-ML} \\
&\quad \times \exp\left[-\sum_{m=1}^M \mathrm{tr}\left\{\boldsymbol{\Sigma}_{A,0}^{-1} \mathbf{X}(m) \mathbf{X}^H(m)\right\}\right],
\end{aligned}
\tag{8}
$$

where $\mathbf{X}_{\mathbb{T}} \triangleq \left[\mathbf{X}(1), \mathbf{X}(2), \cdots, \mathbf{X}(M)\right]$, as in Figure 2, and the ML estimate of the covariance matrix $\boldsymbol{\Sigma}_{A,0}$ becomes $\hat{\boldsymbol{\Sigma}}_{A,0} = \mathrm{diag}\left(\frac{1}{MN} \sum_{m=1}^M \mathbf{X}(m) \mathbf{X}^H(m)\right)$ [7] and thus $\hat{\boldsymbol{\Sigma}}_0 = \hat{\boldsymbol{\Sigma}}_{A,0} \otimes \mathbf{I}_N$. To solve (5), we have to replace $\boldsymbol{\Sigma}_0$ and $\boldsymbol{\Sigma}$ by $\hat{\boldsymbol{\Sigma}}_0$ and $\hat{\boldsymbol{\Sigma}}$, respectively. Therefore, the final expression of the ST-GLRT becomes:

$$
\Lambda_{\mathrm{ST}}(\mathbf{Z}) = \frac{\left|\hat{\boldsymbol{\Sigma}}\right|}{\left|\hat{\boldsymbol{\Sigma}}_0\right|} \gtrless_{\mathcal{H}_1}^{\mathcal{H}_0} \gamma
\tag{9}
$$

Note that the detection scheme in (9) assumes no structure for the covariance matrix, except that the covariance matrix is Hermitian and non-singular.

Before concluding the discussion in this section, as a reference, we present the GLRT that only exploits the spatial correlation, ignoring the temporal correlation. Assuming that the vector samples $\mathbf{x}(n), n = 1, \ldots, N_T$ are temporally uncorrelated, the GLRT scheme can be formulated as:

$$
\Lambda_{\mathrm{Tr}}(\mathbf{X}_{\mathbb{T}}) = \frac{\max\limits_{\boldsymbol{\Sigma}_{A,0}} f_{\mathbf{x}}\left(\mathbf{X}_{\mathbb{T}}; \boldsymbol{\Sigma}_{A,0}\right)}{\max\limits_{\boldsymbol{\Sigma}_{A,1}} f_{\mathbf{x}}\left(\mathbf{X}_{\mathbb{T}}; \boldsymbol{\Sigma}_{A,1}\right)} \gtrless_{\mathcal{H}_1}^{\mathcal{H}_0} \gamma,
\tag{10}
$$

where $f_{\mathbf{x}}\left(\mathbf{X}_{\mathbb{T}}; \boldsymbol{\Sigma}_{A,0}\right)$ and $f_{\mathbf{x}}\left(\mathbf{X}_{\mathbb{T}}; \boldsymbol{\Sigma}_{A,1}\right)$ are the likelihood functions, and $\boldsymbol{\Sigma}_{A,0}$ and $\boldsymbol{\Sigma}_{A,1}$ represent the covariance matrices under hypothesis $\mathcal{H}_0$ and $\mathcal{H}_1$, respectively.

Solving (10), the final expression of the detector that does not exploit the temporal structure is

$$\Lambda_{\mathrm{Tr}}(\mathbf{X}_{\mathbb{T}}) = \frac{|\mathbf{\Sigma}_{\mathrm{A},1}|}{|\mathbf{\Sigma}_{\mathrm{A},0}|} \gtrless_{\mathcal{H}_1}^{\mathcal{H}_0} \gamma, \tag{11}$$

where $\mathbf{\Sigma}_{\mathrm{A},1} = \frac{1}{N_T} \sum_{n=1}^{N_T} \mathbf{x}(n)\mathbf{x}^H(n)$ is a sample covariance matrix. Under the hypothesis $\mathcal{H}_0$, as we have previously shown, the covariance matrix may be estimated as: $\mathbf{\Sigma}_{\mathrm{A},0} = \mathrm{diag}\left(\mathbf{\Sigma}_{\mathrm{A},1}\right)$ [7]. The detector (11) only exploits the energy and the spatial correlation across the L antennas of the receiver and it wrongly assumes independence in time the information provided by the temporal correlation. Compared to (11), the ST-GLRT (9) provides improved detection performance since it uses temporal correlation as an additional source of information. However, in the case when $M < NL$, the ST-GLRT may completely collapse due to the ill-conditioned sample covariance matrix [17]. In order to circumvent this limitation, in the following sections, we propose some modifications in the ST-GLRT by exploiting the presence of some inherent structures in the space-time correlation.

4 Exploiting persymmetric structure

In order to solve the detection problem (3) with unknown covariance matrices, a critical requirement for the detectors based on the GLRT is that the sample covariance matrices must be non-singular [23]. To this end, we have to make sure that the number of available observations given by M, is not smaller than LN (i.e. $M \geq LN$). However, in quick spectrum sensing, a number of samples greater than LN is a requirement difficult to fulfill in practice [28]. Hence, the motivation of the remaining discussion is to bring robustness against this small sample support. Note that in (9) we assume no prior knowledge about the spatio-temporal structure of the covariance matrix except that it is positive definite. One way to achieve the robustness against the small sample support is to look for possible a-priori known patterns/structures in the large spatio-temporal covariance matrix.

4.1 Persymmetric structure

In this section, we use the fact that the spatio-temporal covariance matrix $\mathbf{\Sigma}$ has a persymmetric structure. From [15,18], we know that the multivariate WSS time series has a block-Toeplitz covariance matrix. We remark here that Toeplitz structured matrices belong to a subclass of the persymmetric matrices [19,21,29]. Furthermore, considering structured antenna array configurations (i.e., uniform linear arrays) at the SU, the spatio-temporal covariance matrix $\mathbf{\Sigma}$ can be modelled as persymmetric-block-Toeplitz, as shown in [30]. Therefore, we propose to approximate the spatio-temporal covariance matrix $\mathbf{\Sigma}$,

which is block-Toeplitz, as a persymmetric matrix. Obviously, this approximation results in a suboptimal detection scheme, but it is well known that it does not exist in closed-form ML estimators for Toeplitz matrices. Moreover, it provides better detection performance than the unconstrained ML estimate, particularly, in the case of small sample support. Persymmetric covariance matrices fulfill [31]

$$\mathbf{\Sigma} = \mathbf{J}_{LN}\mathbf{\Sigma}^T\mathbf{J}_{LN}, \tag{12}$$

where the $LN \times LN$ counter-identity matrix $\mathbf{J}_{LN}$ is

$$\mathbf{J}_{LN} = \begin{bmatrix} 0 & \cdots & 0 & 1 \\ \vdots & \cdots & 1 & 0 \\ 0 & \cdot^{\cdot^{\cdot}} & 0 & \vdots \\ 1 & 0 & \cdots & 0 \end{bmatrix}, \tag{13}$$

also known as exchange matrix [19]. Based on these ideas, in Section 4.2, we present a modified GLRT that exploits the persymmetric property of the block-Toeplitz covariance matrix.

4.2 Persymmetric GLRT (P-GLRT)

The difference in the formulation of the GLRT that exploits the persymmetry comes due to the constraint (12). Therefore, the formulation of the GLRT based on the persymmetric covariance matrix can be represented as:

$$\Lambda_{\mathrm{PS}}(\mathbf{Z}) = \frac{\max\limits_{\mathbf{\Sigma}_0} f_{\mathbf{z}}(\mathbf{Z}; \mathbf{\Sigma}_0)}{\max\limits_{\mathbf{\Sigma}} f_{\mathbf{z}}(\mathbf{Z}; \mathbf{\Sigma})} \gtrless_{\mathcal{H}_1}^{\mathcal{H}_0} \gamma. \tag{14}$$
$$\text{s.t. } \mathbf{\Sigma} = \mathbf{J}_{LN}\mathbf{\Sigma}^T\mathbf{J}_{LN}$$

Comparing (14) to (5), we can see that the difference lies only in the denominator. In order to exploit the persymmetry of $\mathbf{\Sigma}$, we need to use the forward-backward (FB) log likelihood [32], which is a combination of the forward-looking and the backward-looking log-likelihood functions. The forward-looking log-likelihood $\log f_{\mathbf{z}}(\mathbf{Z}; \mathbf{\Sigma})$ can be written as:

$$\log f_{\mathbf{z}}(\mathbf{Z}; \mathbf{\Sigma}) = -\log|\mathbf{\Sigma}| - \mathrm{tr}\left\{\mathbf{\Sigma}^{-1}\hat{\mathbf{\Sigma}}\right\}, \tag{15}$$

where we have ignored constant terms that do not depend on data. Similarly, using the constraint $\mathbf{\Sigma} = \mathbf{J}_{LN}\mathbf{\Sigma}^T\mathbf{J}_{LN}$, the backward-looking log-likelihood $\log f_{\mathbf{z}}^{(B)}(\mathbf{Z}; \mathbf{\Sigma})$ can be written as [32]:

$$\log f_{\mathbf{z}}^{(B)}(\mathbf{Z}; \mathbf{\Sigma}) = -\log|\mathbf{\Sigma}| - \mathrm{tr}\left\{\mathbf{\Sigma}^{-1}\mathbf{J}_{LN}\hat{\mathbf{\Sigma}}^T\mathbf{J}_{LN}\right\}. \tag{16}$$

Adding (15) and (16), it gives us the forward-backward log likelihood as:

$$\frac{1}{2}\log f_{\mathbf{z}}^{(FB)}(\mathbf{Z}; \mathbf{\Sigma}) = -\log|\mathbf{\Sigma}| - \mathrm{tr}\left\{\mathbf{\Sigma}^{-1}\hat{\mathbf{\Sigma}}_{\mathrm{PS}}\right\}, \tag{17}$$

where

$$\hat{\boldsymbol{\Sigma}}_{\mathrm{PS}} = \frac{1}{2}\left(\hat{\boldsymbol{\Sigma}} + \mathbf{J}_{LN}\hat{\boldsymbol{\Sigma}}^{T}\mathbf{J}_{LN}\right). \tag{18}$$

The covariance matrix estimator (18) is called forward-backward sample covariance matrix, which is the ML estimator of the persymmetric covariance matrix [32]. An exact theory indicating the performance of the estimator as a function of number of independent vector $\mathbf{z}(m)$, $m = 1, 2, \ldots, M$, is not available. However, a qualitative discussion reported in [19,29] shows that the required number of samples decreases by approximately a factor of two. Similarly, it is reported in [33] that $\hat{\boldsymbol{\Sigma}}_{\mathrm{PS}}$ has consistently lower variance than the variance of $\hat{\boldsymbol{\Sigma}}$. Finally, solving (14) and using (18), the GLRT becomes

$$\Lambda_{\mathrm{PS}}\left(\mathbf{Z}\right) = \frac{\left|\hat{\boldsymbol{\Sigma}}_{\mathrm{PS}}\right|}{\left|\hat{\boldsymbol{\Sigma}}_{0}\right|} \gtrless_{\mathcal{H}_{1}}^{\mathcal{H}_{0}} \gamma. \tag{19}$$

where $\hat{\boldsymbol{\Sigma}}_{0} = \hat{\boldsymbol{\Sigma}}_{\mathrm{A},0} \otimes \mathbf{I}_{N}$ as before. Compared to the detection scheme in (9), the new one in (19) offers improved performance at small sample support, as the number of independent vector measurements required for the covariance matrix estimator decreases by up to a factor of two [19]. Motivated by these facts, in Section 5, we go one step further and exploit the properties of the Kronecker product to decompose the large covariance matrix into smaller ones, reducing considerably the number of unknown parameters.

5 GLRT based on Kronecker factorization

In Section 3, we presented the ST-GLRT approach for the detection problem in (3) and argued that it performs poorly for $M < KN$. In Section 4.1, we have (partially) exploited the temporal structure by imposing persymmetry on the covariance matrix. We have also discussed that the spatio-temporal covariance matrix $\boldsymbol{\Sigma}$ has block-Toeplitz (with Toeplitz blocks) structure. In [34,35], it is reported that the block-Toeplitz structure can be approximated by the Kronecker product of two matrices. Taking this into account, we therefore approximate the covariance matrix $\boldsymbol{\Sigma}$ into a purely spatial and a purely temporal component as:

$$\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_{\mathrm{A}} \otimes \boldsymbol{\Sigma}_{\mathrm{T}}. \tag{20}$$

In (20), the matrix $\boldsymbol{\Sigma}_{\mathrm{A}}$ captures the spatial correlation between the observations received at different antennas and matrix $\boldsymbol{\Sigma}_{\mathrm{T}}$ captures the time correlation between N column vectors in $\mathbf{X}$. Herein, we remark that the covariance structure in (20) makes the implicit assumption that the temporal correlation structure remains the same at all

spatial locations. Similarly, the spatial correlation structure remains the same for the whole sub-block.

5.1 KR-GLRT

In this section, we consider the Kronecker product-based factorization and derive the GLRT that exploits this structure, which is given by

$$\Lambda_{\mathrm{KR}}(\mathbf{Z}) = \frac{\max\limits_{\boldsymbol{\Sigma}_{0}} f_{\mathbf{z}}\left(\mathbf{Z}; \boldsymbol{\Sigma}_{0}\right)}{\max\limits_{\boldsymbol{\Sigma}_{\mathrm{T}}, \boldsymbol{\Sigma}_{\mathrm{A}}} f_{\mathbf{z}}\left(\mathbf{Z}; \boldsymbol{\Sigma}_{\mathrm{A}} \otimes \boldsymbol{\Sigma}_{\mathrm{T}}\right)} \gtrless_{\mathcal{H}_{1}}^{\mathcal{H}_{0}} \gamma. \tag{21}$$

In order to solve (21), under hypothesis $\mathcal{H}_{1}$, we need to obtain the ML estimates of the unknown covariance matrices $\boldsymbol{\Sigma}_{\mathrm{A}}$ and $\boldsymbol{\Sigma}_{\mathrm{T}}$. Under $\mathcal{H}_{1}$, the likelihood function $f_{\mathbf{z}}\left(\mathbf{Z}; \boldsymbol{\Sigma}_{\mathrm{A}} \otimes \boldsymbol{\Sigma}_{\mathrm{T}}\right)$, ignoring constant factors, can be written as:

$$f_{\mathbf{z}}\left(\mathbf{Z}; \boldsymbol{\Sigma}_{\mathrm{A}} \otimes \boldsymbol{\Sigma}_{\mathrm{T}}\right) = |\boldsymbol{\Sigma}_{\mathrm{A}} \otimes \boldsymbol{\Sigma}_{\mathrm{T}}|^{-M}$$
$$\times \exp\left[-\sum_{m=1}^{M} \mathbf{z}^{H}(m)\left(\boldsymbol{\Sigma}_{\mathrm{A}}^{-1} \otimes \boldsymbol{\Sigma}_{\mathrm{T}}^{-1}\right)\mathbf{z}(m)\right]. \tag{22}$$

Now, we can write $f_{\mathbf{z}}\left(\mathbf{Z}; \boldsymbol{\Sigma}_{\mathrm{A}} \otimes \boldsymbol{\Sigma}_{\mathrm{T}}\right) = f_{\mathbf{x}}\left(\mathbf{X}_{\mathbb{T}}; \boldsymbol{\Sigma}_{\mathrm{A}}, \boldsymbol{\Sigma}_{\mathrm{T}}\right)$ [36], which yields

$$f_{\mathbf{x}}\left(\mathbf{X}_{\mathbb{T}}; \boldsymbol{\Sigma}_{\mathrm{A}}, \boldsymbol{\Sigma}_{\mathrm{T}}\right) = |\boldsymbol{\Sigma}_{\mathrm{A}}|^{-MN}|\boldsymbol{\Sigma}_{\mathrm{T}}|^{-ML}$$
$$\times \exp\left[-\sum_{m=1}^{M} \mathrm{tr}\left\{\boldsymbol{\Sigma}_{\mathrm{T}}^{-1}\mathbf{X}^{H}(m)\,\boldsymbol{\Sigma}_{\mathrm{A}}^{-1}\mathbf{X}(m)\right\}\right], \tag{23}$$

To find the ML estimates of $\boldsymbol{\Sigma}_{\mathrm{A}}$ and $\boldsymbol{\Sigma}_{\mathrm{T}}$, we need to take the derivative of $\log f_{\mathbf{X}}\left(\mathbf{x}_{\mathbb{T}}; \boldsymbol{\Sigma}_{\mathrm{A}}, \boldsymbol{\Sigma}_{\mathrm{T}}\right)$, given by

$$\log f_{\mathbf{x}}\left(\mathbf{X}_{\mathbb{T}}; \boldsymbol{\Sigma}_{A}, \boldsymbol{\Sigma}_{\mathrm{T}}\right) = -N\log|\boldsymbol{\Sigma}_{\mathrm{A}}| - L\log|\boldsymbol{\Sigma}_{\mathrm{T}}|$$
$$-\frac{1}{M}\sum_{m=1}^{M}\mathrm{tr}\left\{\boldsymbol{\Sigma}_{\mathrm{T}}^{-1}\mathbf{X}^{H}(m)\,\boldsymbol{\Sigma}_{\mathrm{A}}^{-1}\mathbf{X}(m)\right\} \tag{24}$$

with respect to $\boldsymbol{\Sigma}_{\mathrm{A}}$ ($\boldsymbol{\Sigma}_{\mathrm{T}}$), keeping $\boldsymbol{\Sigma}_{\mathrm{T}}$ ($\boldsymbol{\Sigma}_{\mathrm{A}}$) fixed. Equating the result of the derivative to zero and after simple mathematical operations, the estimators under $\mathcal{H}_{1}$ can be written as:

$$\hat{\boldsymbol{\Sigma}}_{\mathrm{T}} = \frac{1}{LM}\sum_{m=1}^{M}\mathbf{X}^{H}(m)\,\hat{\boldsymbol{\Sigma}}_{\mathrm{A}}^{-1}\mathbf{X}(m), \tag{25}$$

$$\hat{\boldsymbol{\Sigma}}_{\mathrm{A}} = \frac{1}{NM}\sum_{m=1}^{M}\mathbf{X}(m)\,\hat{\boldsymbol{\Sigma}}_{\mathrm{T}}^{-1}\mathbf{X}^{H}(m). \tag{26}$$

Equations (25) and (26) suggest that $\hat{\boldsymbol{\Sigma}}_{\mathrm{T}}$ and $\hat{\boldsymbol{\Sigma}}_{\mathrm{A}}$ can be obtained using an iterative method such as the Flip-Flop algorithm. The Flip-Flop algorithm maximizes $\log f_{\mathbf{x}}\left(\mathbf{X}_{\mathbb{T}}; \boldsymbol{\Sigma}_{\mathrm{A}}, \boldsymbol{\Sigma}_{\mathrm{T}}\right)$ w.r.t. $\boldsymbol{\Sigma}_{\mathrm{A}}$, keeping the last available

estimate of $\mathbf{\Sigma}_T$ fixed and vice versa. In [23], numerical experiments show that the Flip-Flop algorithm performs very well and is much faster than a more general purpose optimization algorithm such as Newton–Raphson [23]. In [22], it has been reported that for the case of large enough M, asymptotic efficiency of the ML approach can be achieved without iterating. Moreover, it is reported in [23,36] that if $(M-1)N \geq L$ and $(M-1)L \geq N$, or equivalently $M \geq \max(N/L; L/N) + 1$, then every iterate in the Flip-Flop algorithm results in positive definite $\hat{\mathbf{\Sigma}}_A \otimes \hat{\mathbf{\Sigma}}_T$ with probability one. Taking into account this fact, in order to get $\hat{\mathbf{\Sigma}}$, we adopt a non-iterative Flip-Flop approach and only perform the steps given in Algorithm 1, with an initial value of $\hat{\mathbf{\Sigma}}_A^0 = \mathbf{I}_{L \times L}$. On the other hand, under $\mathcal{H}_0$, we have the estimate of $\mathbf{\Sigma}_0$ as: $\hat{\mathbf{\Sigma}}_0 = \hat{\mathbf{\Sigma}}_{A,0} \otimes \mathbf{I}_{N \times N}$. Having all of the maximum likelihood-based estimates, and solving (21), we can get the expression:

Algorithm 1 ML based Non-Iterative Flip-Flop

- Choose a starting value for $\hat{\mathbf{\Sigma}}_A^0$ as $\mathbf{I}_{L \times L}$.
- Estimate $\hat{\mathbf{\Sigma}}_T^1$ from (25) with $\hat{\mathbf{\Sigma}}_A = \hat{\mathbf{\Sigma}}_A^0$.
- Find the following

  1. Estimate $\hat{\mathbf{\Sigma}}_A$ from (26) with $\hat{\mathbf{\Sigma}}_T^1$.
  2. Estimate $\hat{\mathbf{\Sigma}}_T$ from (25) with $\hat{\mathbf{\Sigma}}_A$ from previous step.

$$\Lambda_{\text{KR}}(\mathbf{Z}) = \frac{\left|\hat{\mathbf{\Sigma}}_T\right|^L \left|\hat{\mathbf{\Sigma}}_A\right|^N}{\left|\hat{\mathbf{\Sigma}}_0\right|} \underset{\mathcal{H}_1}{\overset{\mathcal{H}_0}{\gtrless}} \gamma. \tag{27}$$

The main advantage of the proposed GLRT (27) over the traditional is that under $\mathcal{H}_1$ instead of $\frac{1}{2}LN(LN+1)$ parameters, it has only $\frac{1}{2}L(L+1) + \frac{1}{2}N(N+1)$ parameters to estimate. Furthermore, the dimensions of these two covariance matrices $\mathbf{\Sigma}_T$ and $\mathbf{\Sigma}_A$ are much smaller than the dimension of full covariance matrix $\mathbf{\Sigma}$, that is why the computations are much less demanding. Hence, the Kronecker model is a good approximation that captures important information about the correlations, while it is positive definite for $M \geq \max(N/L; L/N) + 1$, i.e. a much smaller number of samples.

5.2 PK-GLRT

In Section 4, we have assumed that the covariance matrix $\mathbf{\Sigma}$ has a block-Toeplitz (with Toeplitz blocks) structure. Keeping this in mind, in this section, we assess the possible improvement in the detection performance of

(27) by exploiting the fact that the factored matrices $\mathbf{\Sigma}_T$ and $\mathbf{\Sigma}_A$ have persymmetric structures. We show that it is possible to account for the persymmetric structure by a simple modification of the Flip-Flop algorithm. Hence, as we did in Section 4.1, the persymmetric structures are exploited by imposing the constraints [21]:

$$\mathbf{\Sigma}_T = \mathbf{J}_N \mathbf{\Sigma}_T^T \mathbf{J}_N, \tag{28}$$

$$\mathbf{\Sigma}_A = \mathbf{J}_L \mathbf{\Sigma}_A^T \mathbf{J}_L, \tag{29}$$

where $\mathbf{J}_L$ and $\mathbf{J}_N$ are the reversal matrices of dimensions $(L \times L)$ and $(N \times N)$, respectively. The modified version of KR-GLRT (21) that we denoted as PK-GLRT can be written as:

$$\Lambda_{\text{PKR}}(\mathbf{Z}) = \frac{\max_{\mathbf{\Sigma}_0} f_{\mathbf{z}}(\mathbf{Z}; \mathbf{\Sigma}_0)}{\max_{\mathbf{\Sigma}_T, \mathbf{\Sigma}_A} f_{\mathbf{z}}(\mathbf{Z}; \mathbf{\Sigma}_A \otimes \mathbf{\Sigma}_T)}$$

$$\text{s.t} \quad \begin{aligned} \mathbf{\Sigma}_A &= \mathbf{J}_L \mathbf{\Sigma}_A^T \mathbf{J}_L \\ \mathbf{\Sigma}_T &= \mathbf{J}_N \mathbf{\Sigma}_T^T \mathbf{J}_N \end{aligned} \tag{30}$$

We see that in the expression (30), under hypothesis $\mathcal{H}_0$, the solution is the same as in (14) and (21). The difference lies in the case of hypothesis $\mathcal{H}_1$, where we need to solve

$$\max_{\mathbf{\Sigma}_A, \mathbf{\Sigma}_T} \log f_{\mathbf{Z}}(\mathbf{Z}; \mathbf{\Sigma}_A \otimes \mathbf{\Sigma}_T),$$

$$\text{s.t} \quad \begin{aligned} \mathbf{\Sigma}_A &= \mathbf{J}_L \mathbf{\Sigma}_A^T \mathbf{J}_L \\ \mathbf{\Sigma}_T &= \mathbf{J}_N \mathbf{\Sigma}_T^T \mathbf{J}_N \end{aligned} \tag{31}$$

As in Section 4.1, to obtain the ML estimate of a persymmetric covariance matrix, we need the forward-backward (FB) log likelihood, which is the combination of the forward-looking and the backward-looking log likelihoods. In this case, the forward-looking log-likelihood function for estimating $\mathbf{\Sigma}_T$ can be written as:

$$\log f_{\mathbf{z}}^F(\mathbf{Z}; \mathbf{\Sigma}_A \otimes \mathbf{\Sigma}_T) = -N \log|\mathbf{\Sigma}_A| - L \log|\mathbf{\Sigma}_T|$$
$$- \frac{1}{M} \sum_{m=1}^{M} \text{tr}\left\{ \mathbf{\Sigma}_T^{-1} \mathbf{X}^H(m) \mathbf{\Sigma}_A^{-1} \mathbf{X}(m) \right\}. \tag{32}$$

Similarly, the backward-looking log likelihood is

$$\log f_{\mathbf{z}}^B(\mathbf{Z}; \mathbf{\Sigma}_A \otimes \mathbf{\Sigma}_T) = -N \log|\mathbf{\Sigma}_A| - L \log|\mathbf{\Sigma}_T|$$
$$- \frac{1}{M} \text{tr}\left\{ \mathbf{\Sigma}_T^{-1} \mathbf{J}_N \mathbf{Q}^H \mathbf{J}_N \right\}, \tag{33}$$

where in (33), $\mathbf{Q} \triangleq \sum_{m=1}^{M} \mathbf{X}^H(m) \mathbf{\Sigma}_A^{-1} \mathbf{X}(m)$. Adding (32) and (33) gives us the Kronecker product-based forward-backward log-likelihood $\log f_{\mathbf{z}}^{BF}(\mathbf{Z}; \mathbf{\Sigma}_A \otimes \mathbf{\Sigma}_T)$, required for estimating the persymmetric estimate of $\mathbf{\Sigma}_T$. Following similar steps as in (32) and (33), we can find forward-

and backward-looking log-likelihood functions for finding the persymmetric estimate of $\boldsymbol{\Sigma}_A$. To find the ML estimates of the covariance matrices, in the first step, we fix $\boldsymbol{\Sigma}_A$ and find $\boldsymbol{\Sigma}_T$ that maximizes $\log f_{\mathbf{Z}}^{BF}(\mathbf{Z}; \boldsymbol{\Sigma}_A \otimes \boldsymbol{\Sigma}_T)$. Taking into account these results, the estimator of the persymmetric $\boldsymbol{\Sigma}_T$ can be found as:

$$
\hat{\boldsymbol{\Sigma}}_{\mathrm{PS,T}} = \frac{1}{2ML} \sum_{m=1}^{M} \mathbf{X}^H(m) \, \hat{\boldsymbol{\Sigma}}_{\mathrm{PS,A}}^{-1} \mathbf{X}(m)
$$
$$
+ \frac{1}{2ML} \sum_{m=1}^{M} \mathbf{J}_{\mathrm{N}} \left(\mathbf{X}^H(m) \, \hat{\boldsymbol{\Sigma}}_{\mathrm{PS,A}}^{-1} \mathbf{X}(m) \right)^T \mathbf{J}_{\mathrm{N}},
$$

(34)

Similarly, by following the same process with fixed $\boldsymbol{\Sigma}_T$, the estimator of the persymmetric $\boldsymbol{\Sigma}_A$ can be written as:

$$
\hat{\boldsymbol{\Sigma}}_{\mathrm{PS,A}} = \frac{1}{2MN} \sum_{m=1}^{M} \mathbf{X}(m) \, \hat{\boldsymbol{\Sigma}}_{\mathrm{PS,T}}^{-1} \mathbf{X}^H(m)
$$
$$
+ \frac{1}{2MN} \sum_{m=1}^{M} \mathbf{J}_{\mathrm{L}} \left(\mathbf{X}(m) \, \hat{\boldsymbol{\Sigma}}_{\mathrm{PS,T}}^{-1} \mathbf{X}^H(m) \right)^T \mathbf{J}_{\mathrm{L}},
$$

(35)

As it was in the case of expressions (25) and (26), both of the expressions (34) and (35) suggest that $\hat{\boldsymbol{\Sigma}}_{\mathrm{PS,T}}$ and $\hat{\boldsymbol{\Sigma}}_{\mathrm{PS,A}}$ can be estimated using an iterative method such as the Flip-Flop algorithm, as shown in Algorithm 2. Using (34) and (35), the final expression for the GLRT becomes

$$
\Lambda_{\mathrm{PKR}}(\mathbf{Z}) = \frac{\left| \hat{\boldsymbol{\Sigma}}_{\mathrm{PS,T}} \right|^L \left| \hat{\boldsymbol{\Sigma}}_{\mathrm{PS,A}} \right|^N}{\left| \hat{\boldsymbol{\Sigma}}_0 \right|} \gtrless_{\mathcal{H}_1}^{\mathcal{H}_0} \gamma.
$$

(36)

Compared to (27), (36) provides better detection performance in the small sample support regime. It is because, finding the estimates of $\boldsymbol{\Sigma}_T$ and $\boldsymbol{\Sigma}_A$, (34) and (35) require smaller M compared to (25) and (26). In conclusion, by exploiting the underlying structure of the covariance matrix $\boldsymbol{\Sigma}$ via the persymmetric ML estimates of the covariance matrices $\boldsymbol{\Sigma}_T$ and $\boldsymbol{\Sigma}_A$, it further increases the robustness of (27) at small sample support.

Algorithm 2 Non-Iterative Flip-Flop (persymmetric)

- Choose a starting value for $\hat{\boldsymbol{\Sigma}}_A^0$ as $\mathbf{I}_{L \times L}$
- Estimate $\hat{\boldsymbol{\Sigma}}_T^1$ from (25) with $\hat{\boldsymbol{\Sigma}}_A^1 = \hat{\boldsymbol{\Sigma}}_A^0$.
- Find the following

  1. Estimate $\hat{\boldsymbol{\Sigma}}_{\mathrm{PS,A}}$ from (35) with $\hat{\boldsymbol{\Sigma}}_T^1$.
  2. Estimate $\hat{\boldsymbol{\Sigma}}_{\mathrm{PS,T}}$ from (34) with $\hat{\boldsymbol{\Sigma}}_{\mathrm{PS,A}}$ from step 1.

---

6 Numerical results

In this section, we present numerical results to evaluate the performance of the proposed detection schemes, presented in the preceding sections. For the analysis to be conducted herein, we use the receiver operating characteristic (ROC) curve and the area under the ROC curve (AUC), which varies between 0.5 (poor performance) and 1 (good performance) [37], as the performance measures. We also evaluate the performance by the probability of detection vs the signal-to-noise ratio for a fixed false alarm probability. For the evaluation, we performed the following two experiments:

6.1 Experiment no. 1: detection of a WSS signal at multiple antennas

In this experiment, we assume that the signal received at the multi-antenna receiver is a vector WSS Gaussian process corrupted by uncorrelated noise. In order to assess the detectors with this assumption, the SNRs (expected) $\kappa_l, l = 1, 2, \cdots, L$ are allocated differently, with average SNR of all antennas is: $\bar{\kappa} = \frac{1}{L} \sum_{l=1}^{L} \kappa_l$. For a specified received signal power (equal at different antennas) and $\bar{\kappa}$, we find P_n, the mean noise power. Noise powers (expected[b]) at L different antennas $\sigma_{n,l}^2, l = 1, 2, \cdots, L$ are kept different while $P_n = \frac{1}{L} \sum_{i=1}^{L} \sigma_{n,l}^2$. Moreover, we assume $L = 4$ antennas and a separable spatio-temporal correlation. The spatial covariance matrix is generated as $[\boldsymbol{\Sigma}_A]_{i,j} = 0.3^{|i-j|}, i,j = 1,\ldots,L$, and the temporal covariance matrix is generated as $[\boldsymbol{\Sigma}_T]_{i,j} = 0.9^{|i-j|}, i,j = 1,\ldots,N_T$. The remaining parameters for each experiment are described in the captions of the corresponding diagrams.

In the first experiment, we obtain the ROC curves using Matlab to compare the performance of the proposed schemes to that of the traditional schemes (GLRT that ignores temporal correlation $\Lambda_{\mathrm{TR}}(\mathbf{X}_{\mathbb{T}})$, the energy detector $\Lambda_{\mathrm{Eng}}(\mathbf{X}_{\mathbb{T}})$) and the ST-GLRT and its frequency-domain approximation for WSS series. In particular, the energy detector can be expressed as:

$$
\Lambda_{\mathrm{Eng}}(\mathbf{X}_{\mathbb{T}}) = \sum_{m=1}^{M} \sum_{n=1}^{N} \sum_{l=1}^{L} \left| x_{l,n}(m) \right|^2 \gtrless_{\mathcal{H}_0}^{\mathcal{H}_1} \gamma,
$$

(37)

and the frequency-domain detector is given by (39) in Appendix A. It is to be noted that the energy detector assumes known noise power.

In this experiment, we plot ROC curves for two different cases. In the first case, we assume that the unknown noise powers at different antennas are perfectly estimated and there is no noise power uncertainty. The ROC plots are given in Figure 3. In the second part of this experiment, shown in Figure 4, we repeat the same experimental
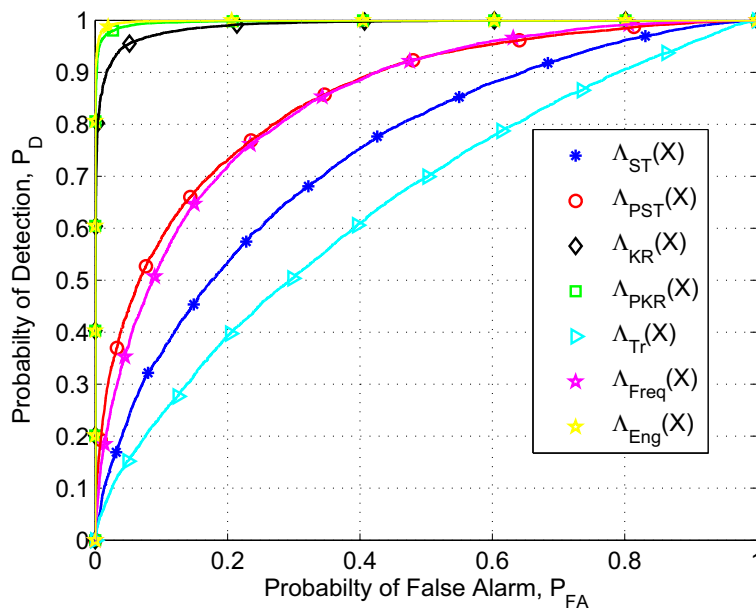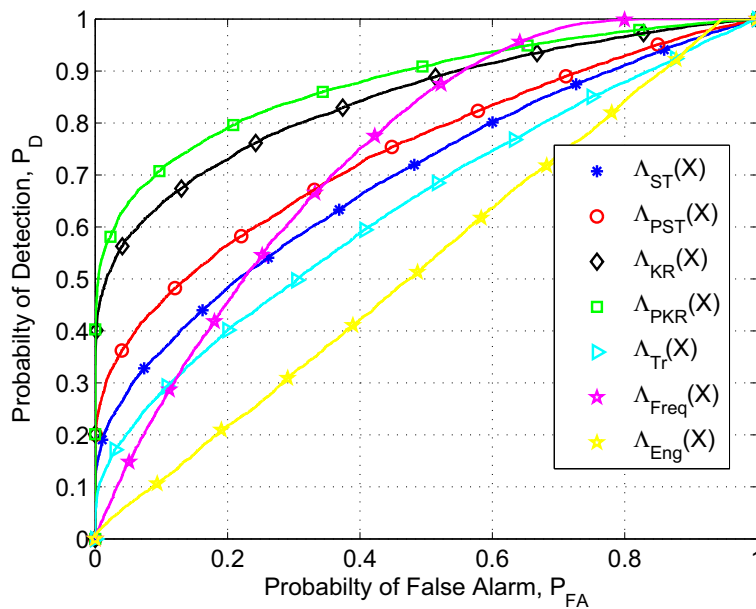
Figure 3 ROC curves for comparison of detection schemes: sample support size *M* = 70, number of vector samples per sub-block *N* = 10, number of antennas *L* = 4, noise uncertainty α_{nu} = 1, and average SNR $\bar{\kappa}$ = −10 dB.

setup for the case with noise power uncertainty. Note that we model the noise power uncertainty by generating the noise power at the *l*-th antenna as $\sigma_{w,l}^2 \sim \mathcal{U}\left(\frac{\sigma_{n,l}^2}{\alpha_{nu}}, \alpha_{nu}\sigma_{n,l}^2\right)$, where $\alpha_{nu} \geq 1$, and $\alpha_{nu} = 1$ means no noise uncertainty [1]. From the results, it is clear that the proposed

schemes clearly outperform traditional schemes. Moreover, from the experiment, we can also conclude that the noise power uncertainty slightly deteriorates the performance of all detectors. However, the energy detector completely collapses at $\alpha_{nu} = 2$. To further analyze the effects of the sample support, noise power uncertainty,

Figure 4 ROC curves for comparison of detection schemes: sample support size *M* = 70, number of vector samples per sub-block *N* = 10, number of antennas *L* = 4, noise uncertainty α_{nu} = 2, and average SNR $\bar{\kappa}$ = −10 dB.

and shadowing parameters, we need to have a single and quantitative figure of merit. This metric is the area under the ROC curve (AUC), which varies between 0.5 (poor performance) and 1 (good performance). Hence, next, we use AUC curves to see effects of the sample support, noise power uncertainty, and shadowing parameters on the detection performance of the spectrum sensing schemes. We model the shadowing effect by using log-normal random variable as: $p_{sh} = p_m 10^{x_\sigma/10}$, where p_m is the expected signal power at the receiver (equal for all antennas), p_{sh} is the signal power after shadowing effect, and x_σ is Gaussian random variable with 0 mean and standard deviation σ_{Sh}. The log-normal shadow fading is often characterized by its dB spread, σ_{dB}, which has the relationship $\sigma_{\text{Sh}} = 0.1\sigma_{\text{dB}} \log_{10}$ [38].

In Figure 5, we plot the AUC curves to analyze the effects of sample size in the presence of shadowing, with and without the effects of noise power uncertainty. In this figure, the curves with dashed lines represent the case where both shadowing and noise power uncertainty with $\alpha_{nu} = 1.5$. On the other hand, the curves with solid lines show the case with no noise power uncertainty (i.e. $\alpha_{nu} = 1$). From these plots, it can be concluded that the proposed schemes $\Lambda_{\text{KR}}(\mathbf{Z})$ and $\Lambda_{\text{PKR}}(\mathbf{Z})$ are robust against the small sample support both in the presence and absence of noise power uncertainty. Particularly, in the region $20 \leq M \leq 80$, the detectors $\Lambda_{\text{KR}}(\mathbf{Z})$ and $\Lambda_{\text{PKR}}(\mathbf{Z})$ clearly outperform the other detectors. As expected, we can also see that in the small sample regime, $\Lambda_{\text{PKR}}(\mathbf{Z})$ performs better than $\Lambda_{\text{KR}}(\mathbf{Z})$. The obvious reason for this is that under $\mathcal{H}_1$ instead of $\frac{1}{2}L(L+1) + \frac{1}{2}N(N+1)$ parameters

in the case of $\Lambda_{\text{KR}}(\mathbf{Z})$, the detection scheme $\Lambda_{\text{PKR}}(\mathbf{Z})$ has approximately only $(2L - 1) + (2N - 1)$ parameters to estimate. In order to confirm this, in Figure 6, we plot the normalized minimum square error (MSE) of the estimator of the covariance matrix under hypothesis $\mathcal{H}_1$, expressed as:

$$\text{RNMSE} = \sqrt{\frac{1}{J_{\text{avg}}} \sum_{j=1}^{J_{\text{avg}}} \frac{\left\| \hat{\boldsymbol{\Sigma}}_j - \boldsymbol{\Sigma} \right\|_F^2}{\|\boldsymbol{\Sigma}\|_F^2}} \tag{38}$$

where $\boldsymbol{\Sigma}$ is the true spatio-temporal covariance matrix, $\hat{\boldsymbol{\Sigma}}_j$ is the estimated covariance for each estimator in the Monte Carlo simulation j, $\|.\|_F^2$ is the Frobenius norm, and J_{avg} is the number of Monte Carlo simulations. The results confirm that for small sample support, estimators of $\boldsymbol{\Sigma}$ used in the case $\Lambda_{\text{KR}}(\mathbf{Z})$ and $\Lambda_{\text{PKR}}(\mathbf{Z})$ have smaller error.

In Figure 7, we show the AUC plots to analyze the effect of shadowing (i.e. σ_{dB}) both in the presence and absence of noise uncertainty, $\alpha_{nu} = 1.5$ and $\alpha_{nu} = 1$, respectively. It is clear from these results that the effects of shadowing are very small on the performance of the detection schemes. However, we can see that by incrementing σ_{dB}, a slight improvement occurs in the performance of the detection schemes $\Lambda_{\text{TR}}(\mathbf{Z})$, $\Lambda_{\text{ST}}(\mathbf{Z})$, and $\Lambda_{\text{PS}}(\mathbf{Z})$. The most obvious reason for this interesting outcome can be the heavy-tailed distribution of the primary signal strength due to the log-normally-distributed shadow fading that behaves in such a way at lower SNR [39].
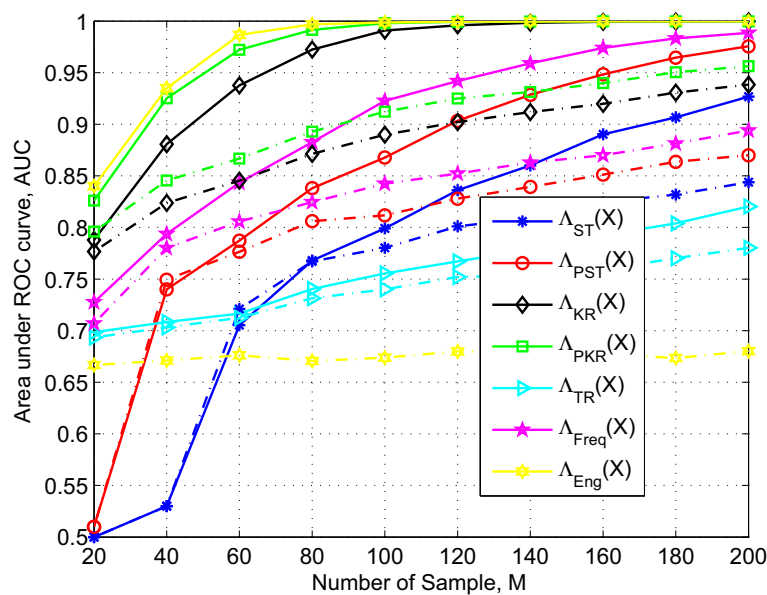

Figure 5 AUC curves (solid lines for $\alpha_{nu} = 1$ and dashed lines for $\alpha_{nu} = 2$) to assess the effects of number of samples *M*: number of vector samples per sub-block *N* = 15, number of antennas *L* = 4, shadowing $\sigma_{\text{dB-spread}} = 4$, and average SNR $\bar{\kappa} = -12$ dB.
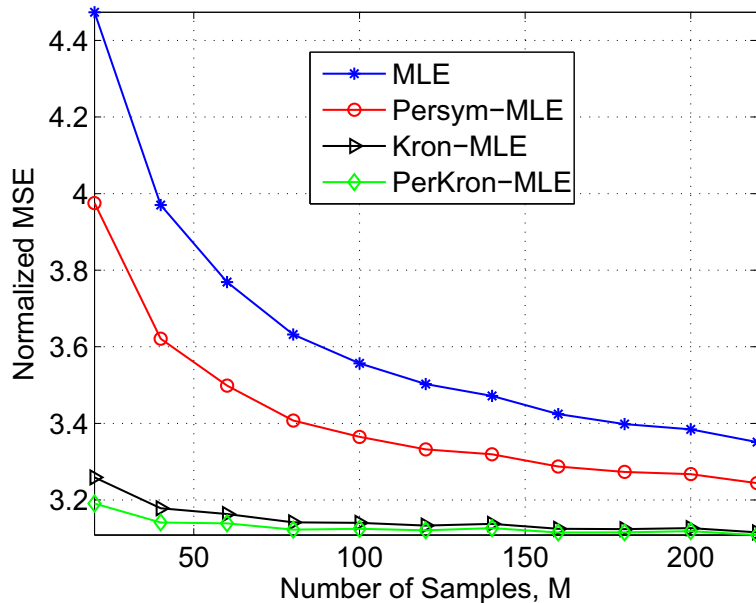
Figure 6 Normalized MSE of the estimator of covariance matrix under hypothesis $\mathcal{H}_1$.

In Figure 8, we show the AUC plots to analyze the effects of noise power uncertainty. The results show a robust behavior for the detection schemes against the noise power uncertainty. Once again, we observe that the performance of the proposed schemes $\Lambda_{\mathrm{KR}}(\mathbf{Z})$ and $\Lambda_{\mathrm{PKR}}(\mathbf{Z})$ is better than other schemes that do not exploit the underlying structure of the received signal.

Finally, keeping the probability of false alarm P_F fixed, we simulate the performance of the detection schemes by plotting P_D against different values of average SNR $\bar{\kappa}$. The simulation results are shown in Figure 9. The results clearly show that for different SNR values, the proposed schemes consistently perform better than the detection schemes that do not exploit the covariance structure.
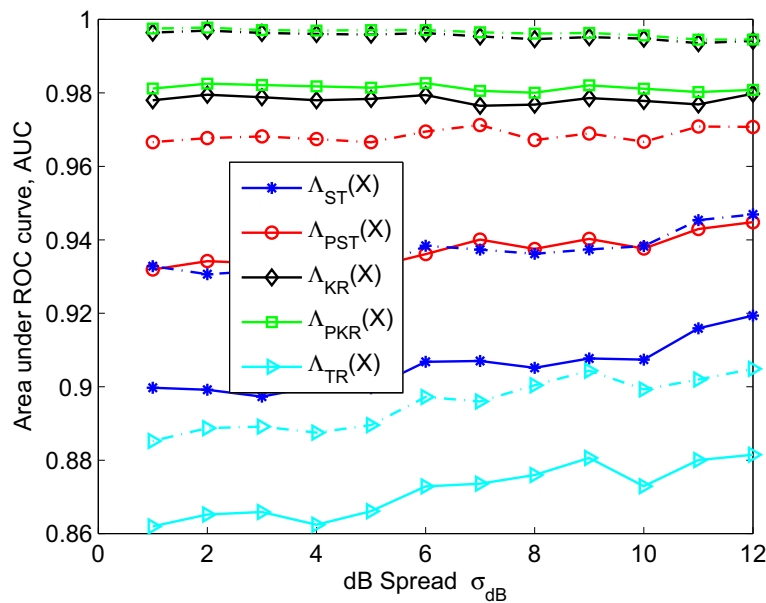
Figure 7 AUC curves (solid lines for $\alpha_{nu} = 1.5$ and dashed lines for $\alpha_{nu} = 1$) to assess the effects of shadowing σ_{dB}: sample size $M = 80$, number of vector samples per sub-block $N = 15$, number of antennas $L = 4$, and average SNR $\bar{\kappa} = -8$ dB.
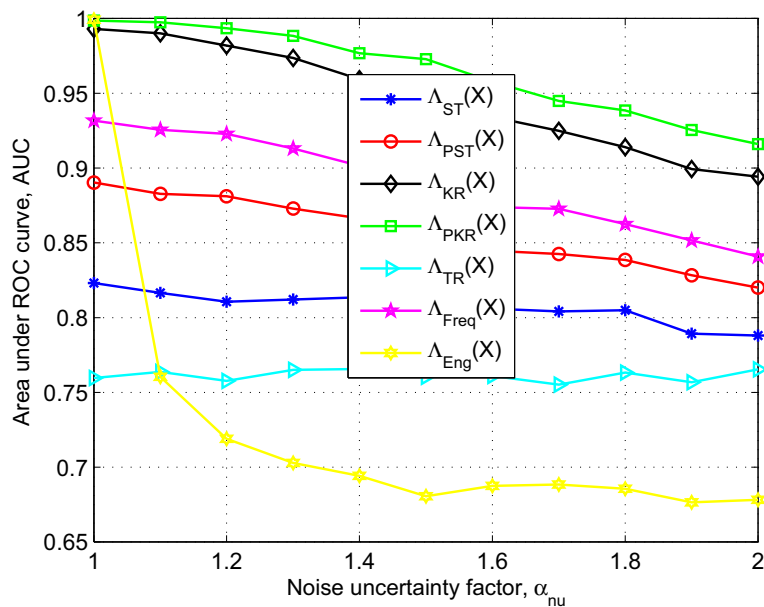
Figure 8 AUC curves to assess the effects of noise uncertainty α_{nu}: with sample size $M = 80$, number of vector samples per sub-block $N = 15$, number of antennas $L = 4$, effect of shadowing $\sigma_{dB-spread} = 4$, and average SNR $\bar{\kappa} = -8$ dB.

In conclusion, we can say that the exploitation of inherent structure of covariance matrix both in frequency and time domain leads us to robustness against the small sample support compared to the ST-GLRT in (9).

6.2 Experiment no. 2: cognitive radio

In the previous set of experiments, we analyzed the proposed schemes for detection of a Gaussian signal with separable spatio-temporal covariance matrix in unknown additive uncorrelated noise. In the present set of experiments, we perform simulations to illustrate the application of the proposed detection schemes in cognitive radio, i.e. with an actual communication signal instead of a Gaussian signal and with a non-separable spatio-temporal covariance matrix. For the simulations, we have used an ODFM-modulated DVB-T signal (8-K mode, $64 - $QAM,
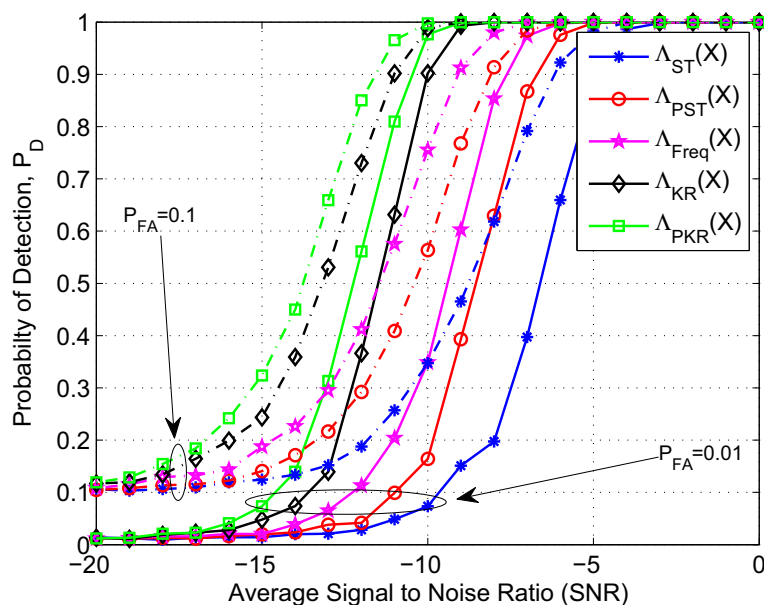
Figure 9 P_D vs average SNR: number of vector samples per sub-block $N = 15$, number of antennas $L = 4$, $M = 80$.
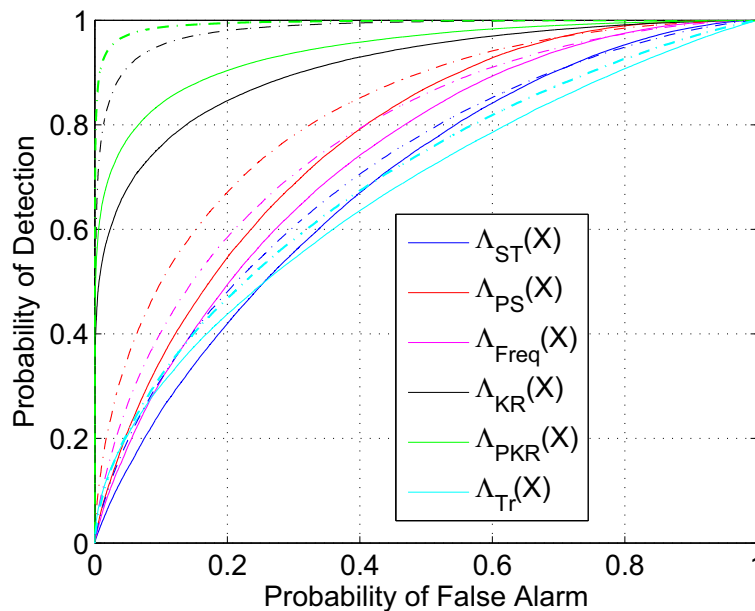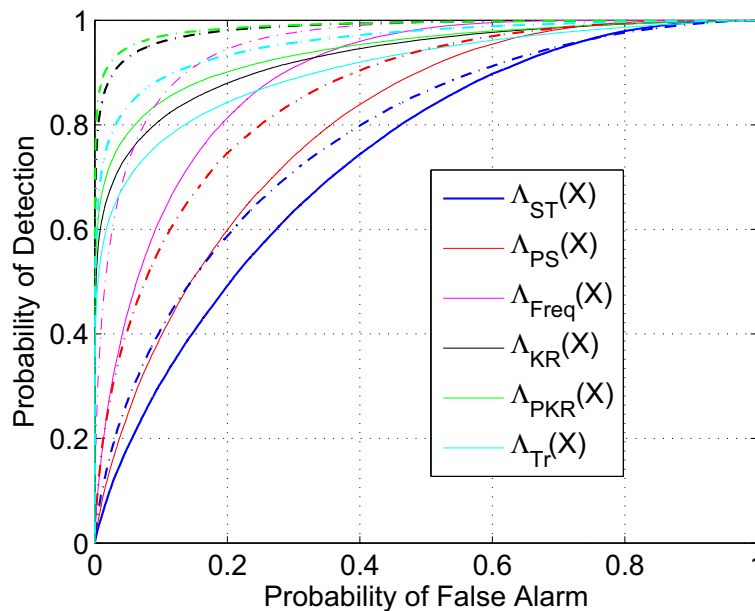
Figure 10 ROC curves (solid lines for $\alpha_{nu} = 2$ and dashed lines for $\alpha_{nu} = 1$): sample support size $M = 70$, number of vector samples per sub-block $N = 15$, number of antennas $L = 4$, channel delay spread 0.779 μs, and average SNR $\bar{\kappa} = -8$ dB.

guard interval 1/4, and inner code rate 2/3) with a bandwidth of 7.61 MHz. We have considered a 4×4 Rayleigh channel with unit power and an exponential power delay profile with length 64 samples (at a sampling frequency of 7.61 MHz). The additive noises at each antenna are generated as a zero-mean complex Gaussian process. We used a noise power different at each antenna with average

SNR $\bar{\kappa} = -8$ dB. To analyze the schemes, we plot ROC curves with $N = 15$ vector samples per sub-block. The rest of the parameters are given in the captions of the figures.

In Figure 10, we plot the ROC curves for the case when the channel delay spread is 0.779 μs with and without noise power uncertainty. In this figure, we can

Figure 11 ROC curves (solid lines for $\alpha_{nu} = 2$ and dashed lines for $\alpha_{nu} = 1$): sample support size $M = 70$, number of vector samples per sub-block $N = 15$, number of antennas $L = 4$, channel delay spread 0.097 μs, and average SNR $\bar{\kappa} = -8$ dB.

easily see that the detectors $\Lambda_{KR}(\mathbf{Z})$ and $\Lambda_{PKR}(\mathbf{Z})$ clearly outperform the other detectors, even for this realistic spatio-temporal correlation. In Figure 11, we repeat the experiment of Figure 10 for a channel delay spread 0.097μs (almost flat fading channel). Even for channels with low-frequency selectivity, we can see that the detection schemes $\Lambda_{KR}(\mathbf{Z})$ and $\Lambda_{PKR}(\mathbf{Z})$ consistently perform better compared to the rest of the detection schemes. However, compared to Figure 10, in Figure 11, we can see interesting results that the detector which ignores temporal correlation outperforms spatio-temporal correlation-based schemes due to the low selectivity of the channel.

Before commenting on these interesting results, for further confirmation, we need to have a single and quantitative figure of merit, so we plot AUC curves in Figure 12.

The AUC curves in Figure 12 demonstrate that the proposed detectors have better performance compared to the traditional detection schemes, for different values of sample support. In order to see the effect of delay spreads, we consider two types of channel delay spreads (0.097 and 0.779 μs). The results in Figure 12 confirm that the proposed schemes $\Lambda_{KR}(\mathbf{Z})$ and $\Lambda_{PKR}(\mathbf{Z})$ consistently outperform other schemes in small sample support regime. We can also see that in general, the performance of all schemes is degraded for different values of delay spreads; since the larger the delay spread, the larger the degradation incurred. However, we can see that compared to rest of the detection schemes, $\Lambda_{KR}(\mathbf{Z})$ and

$\Lambda_{PKR}(\mathbf{Z})$ show more robustness against changes in the delay spread. We can further observe an interesting outcome in the case of the detector that ignores temporal correlation and the frequency-based approximate GLRT, where we can see that the performances of these two detection schemes are quite different for the two channel delay spreads. This is also evident from comparison of ROC curves in Figures 10 and 11. The possible explanation for this can be that for a short channel delay spread (i.e. 0.097 μs), the multi-tap channel adds negligible temporal correlation. On the other hand, in the case of channel delay spread 0.779 μs, we have temporal correlation imposed by the channel on the temporally uncorrelated transmitted OFDM signal. Therefore, the inclusion of temporal correlation as an additional detection metric is not helping in the case of delay spread (i.e. 0.097 μs).

In order to further analyze the detection performance of the proposed schemes, in Figure 13, we compare the results by plotting P_D vs average SNR $\bar{\kappa}$. Once again, we can observe that the proposed schemes outperform the remaining detection schemes presented in this paper.

7 Conclusion

In this paper, we have proposed novel detection schemes that exploit the spatio-temporal correlation present in the received observations at a multi-antenna receiver. When exploiting the spatio-temporal correlation, we have observed that the GLRT performs poorly when the sample support is small. To cope with this problem, we
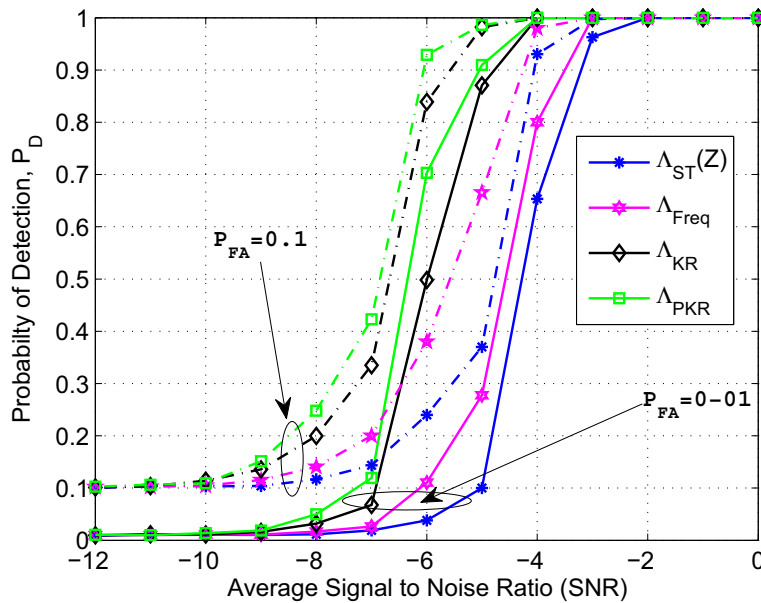

Figure 12 AUC curves (solid lines for channel delay spread 0.779 μs and dashed lines for channel delay spread 0.097 μs) to assess the effects of number of samples *M*: number of vector samples per sub-block $N = 15$, number of antennas $L = 4$, and average SNR $\bar{\kappa} = -8$ dB.
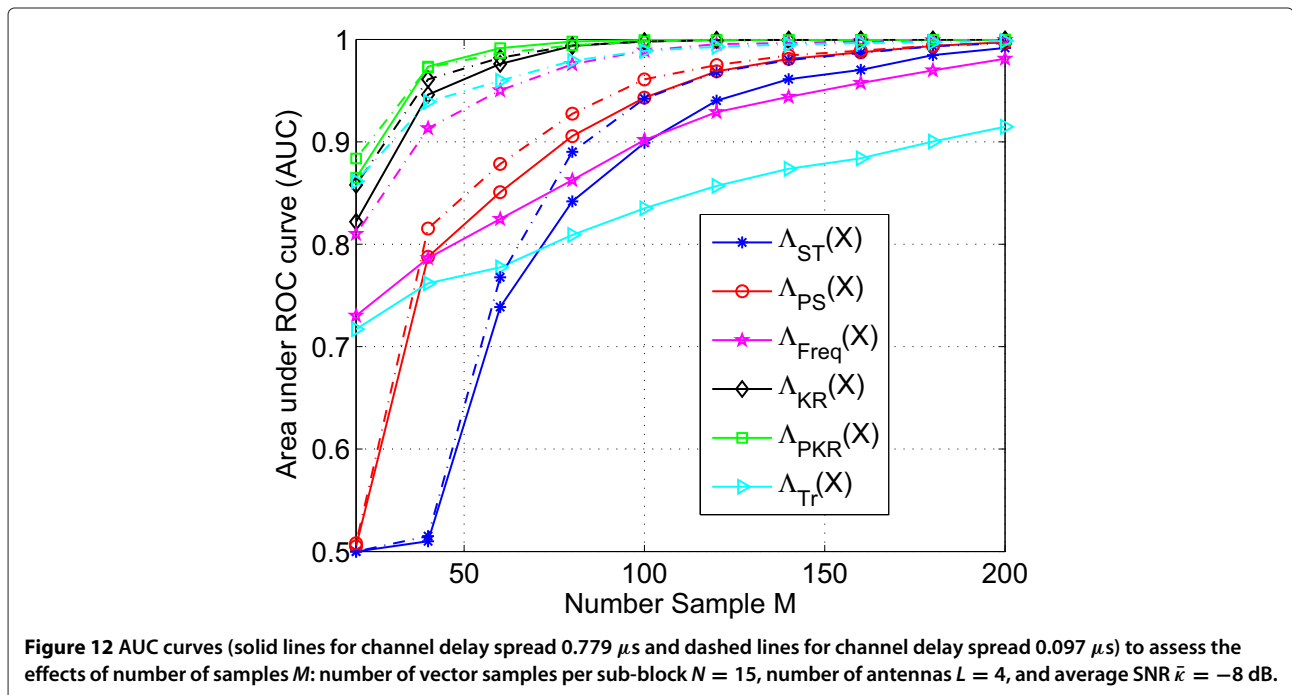
Figure 13 P_D vs average SNR: number of vector samples per sub-block $N = 15$, number of antennas $L = 4$, $M = 80$, channel delay spread 0.097 μs.

have proposed detectors that are robust against the sample support. The proposed detectors (approximately) exploit the inherent spatio-temporal structure of the received covariance matrix by using the properties of persymmetric matrices and the Kronecker product of the spatial and temporal covariance matrices. The performance of the proposed detectors has been evaluated with the help of numerical simulations, which show important improvements compared to the traditional schemes.

Endnotes

[a]We begin with the complex base-band signal sampled at the specific Nyquist rate.

[b]These values are maintained during Monte-Carlo process. In order to create noise power uncertainty, these values are affected by random uncertainty at each run.

Appendix

A Approximated GLRT in the frequency domain

For the interested readers, herein, we reproduce the approximated GLRT in the frequency domain, originally presented in [15]. While considering the case in which the received signals are jointly WSS, the limiting form (L fixed and $M, N \to \infty$) of the the frequency domain GLRT statistic is written as [15]:

$$
\Lambda_{\text{ST}}\left(\mathbf{Z}_f\right) \xrightarrow[M,N\to\infty]{} \exp\left\{\int_{-\pi}^{\pi} \log\left[\frac{\det\left(\hat{\mathbf{S}}\left(e^{j\theta_n}\right)\right)}{\prod_{i=1}^{L} \hat{s}_{ii}\left(e^{j\theta_n}\right)}\right]\frac{d\theta_n}{2\pi}\right\},
$$

(39)

where $\left[\hat{\mathbf{S}}\left(e^{j\theta_n}\right)\right]_{l,k} = \mathbf{f}^H\left(e^{j\theta_n}\right)\hat{\mathbf{\Sigma}}_{l,k}\mathbf{f}\left(e^{j\theta_n}\right)$ is a quadratic estimator and the Fourier vector is given by

$$
\mathbf{f}\left(e^{j\theta_n}\right) = \left[1, e^{-j\theta_n}, e^{-j2\theta_n}, e^{-j3\theta_n}, \ldots, e^{-j(N-1)\theta_n}\right]^T.
$$

(40)

Author details
[1]SPCOMNAV, Universitat Autònoma de Barcelona, 08193 Bellaterra, Barcelona, Spain. [2]Electrical Engineering, University of Engineering and Technology, Peshawar, 25000 Peshawar, Pakistan. [3]Department of Electrical Engineering and Information Technology (EIM-E), Universität Paderborn, Paderborn, Germany. [4]Signal Processing School of Electrical Engineering, KTH Royal Institute of Technology, SE-100 44 Stockholm, Sweden.

References
1. E Biglieri, AJ Goldsmith, LJ Greenstein, NB Mandayam, HV Poor, *Principles of Cognitive Radio* (Cambridge University Press, Cambridge, 2012)
2. AW Min, X Zhang, KG Shin, Detection of small-scale primary users in cognitive radio networks. IEEE J. Selected Areas Commun. **29**, 349–361 (2011)
3. R Zhang, TJ Lim, Y-C Liang, Y Zeng, Multi-antenna based spectrum sensing for cognitive radios: a GLRT approach. IEEE Trans. Commun. **58**(1), 84–88 (2010)
4. E Axell, G Leus, EG Larsson, HV Poor, Spectrum sensing for cognitive radio: state-of-the-art and recent advances. IEEE Signal Process. Mag. **29**(3), 101–116 (2012)

5. Y Zeng, Y-C Liang, AT Hoang, R Zhang, A review on spectrum sensing for cognitive radio: challenges and solutions. EURASIP J. Adv. Signal Process. (2010). doi:10.1155/2010/381465

6. W Yang, G Durisiand, VI Morgenshtern, E Riegler, in *Proc. ISWCS*. Capacity pre-log of SIMO correlated block-fading channels (Piscataway, 2011), pp. 869–873

7. KV Mardia, JT Kent, JM Bibby, *Multivariate Analysis*, 1st edn. (Academic Press, Waltham, 1979), p. 536

8. TW Anderson, *An Introduction to Multivariate Statistical Analysis* (Wiley-Interscience, Hoboken, 2003)

9. Y Zeng, Y-C Liang, Eigenvalue-based spectrum sensing algorithms for cognitive radio. IEEE Trans. Commun. **57**(6), 1784–1793 (2009)

10. C Vazquez-Vilar, R López-Valcarce, Spectrum sensing exploiting guard bands and weak channels. IEEE Trans. Signal Process. **59**(12), 6045–6057 (2011)

11. Z Quan, W Zhang, SJ Shellhammer, A Sayed, Optimal spectral feature detection for spectrum sensing at very low SNR. IEEE Trans. Commun. **59**(1), 201–212 (2011)

12. J Sala-Alvarez, G Vazquez-Vilar, R Lopez-Valcarce, Multiantenna GLR detection of rank-one signals with known power spectrum in white noise with unknown spatial correlation. IEEE Trans. Signal Process. **60**(6), 3065–3078 (2012)

13. A Perez-Neira, M-A Lagunas, MA Rojas, P Stoica, Correlation matching approach for spectrum sensing in open spectrum communications. IEEE Trans. Signal Process. **57**(12), 4823–4836 (2009)

14. G Vazquez-Vilar, R Lopez-Valcarce, J Sala, Multiantenna spectrum sensing exploiting spectral a priori information. IEEE Trans. Wireless Commun. **10**(12), 4345–4355 (2011)

15. D Ramirez, J Via, I Santamaria, LL Scharf, Detection of Spatially Correlated Gaussian Time Series. IEEE Trans. Signal Process. **58**(10), 5006–5015 (2010)

16. TF Ayoub, AR Haimovich, Modified GLRT signal detection algorithm. IEEE Trans. Aerospace Electron. Syst. **36**(3), 810–818 (2000)

17. Y Chen, A Wiesel, AO Hero, in *Proc. of IEEE ICASSP*. Shrinkage Estimation of High Dimensional Covariance Matrices (Piscataway, 2009), pp. 2937–2940

18. D Ramírez, G Vazquez-Vilar, R López-Valcarce, J Vía, I Santamaría, Detection of rank-p signals in cognitive radio networks with uncalibrated multiple antennas. IEEE Trans. Signal Process. **59**(8), 3764–3774 (2011)

19. R Nitzberg, Application of maximum likelihood estimation of persymmetric covariance matrices to adaptive processing. IEEE Trans. Aerospace Electron. Syst. **16**(1), 124–127 (1980)

20. D Akdemir, AK Gupta, Array variate random variables with multiway Kronecker delta covariance matrix structure. J. Algebraic Stat. **2**(1), 98–113 (2011)

21. P Wirfalt, M Jansson, in *Proc. of Sensor Array and Multichannel Signal Processing Workshop (SAM)*. On Toeplitz and Kronecker Structured Covariance Matrix Estimation (Piscataway, 2010), pp. 185–188

22. K Werner, M Jansson, P Stoica, On estimation of covariance matrices with Kronecker product structure. IEEE Trans. Signal Process. **56**(2), 478–491 (2008)

23. N Lu, DL Zimmerman, The likelihood ratio test for a separable covariance matrix. Stat. Probab. Lett. **73**(5), 449–457 (2005)

24. MG Genton, Separable approximations of space-time covariance matrices. Environmetrics **18**(7), 681–695 (2007)

25. S Ali, JA López-Salcedo, G Seco-Granados, in *Proc. of 20th European Signal Processing Conference*. Exploiting Structure of Spatio-temporal Correlation for Detection in Wireless Sensor Networks (Piscataway, 2012), pp. 774–778

26. Y Liang, VV Veeravalli, Capacity of noncoherent time-selective rayleigh-fading channels. IEEE Trans. Inform. Theory **50**(12), 3095–3110 (2004)

27. SM Kay, *Fundamentals of Statistical Signal Processing, Volume 2: Detection Theory* (Prentice Hall, Upper Saddle River, 1998)

28. L Lu, H-C Wu, SS Iyengar, A novel robust detection algorithm for spectrum sensing. IEEE J. Selected Areas Commun. **29**(2), 305–315 (2011)

29. M Wax, T Kailath, Efficient inversion of Toeplitz-block Toeplitz matrix. IEEE Trans. Acoustics, Speech Signal Process. **31**(5), 1218–1221 (1983)

30. J Ward, *Space-Time Adaptive Processing for Airborne Radar*. Technical Report 1015, Lincoln Laboratory, MIT, (Lexington, December 1994)

31. TK Huckle, K Waldherr, T Schulte-Herbrüggen, Exploiting matrix symmetries and physical symmetries in matrix product states and tensor trains. Linear Multilinear Algebra **61**(1), 91–122 (2013)

32. M Jansson, P Stoica, Forward-only and forward-backward sample covariances – a comparative study. Signal Process. **77**(3), 235–245 (1999)

33. B Armour, Structured Covariance Autoregressive Parameter Estimation. PhD thesis, Department of Electrical Engineering, McGill Canada, 1989

34. J Kamm, J Nagy, Optimal Kronecker Product Approximation of Block Toeplitz Matrices. SIAM J. Matrix Anal. Appl. **22**(1), 155–172 (2000)

35. MW Mitchell, MG Genton, ML Gumpertz, A likelihood ratio test for separability of covariances. J. Multivar. Anal. **97**(5), 1025–1043 (2006)

36. P Dutilleul, The MLE algorithm for the matrix normal distribution. J. Stat. Comput. Simul. **64**, 105–123 (1999)

37. T Fawcett, *ROC graphs: Notes and practical considerations for researchers.* Technical Report HPL-2003-4, HP Laboratories, (Palo Alto, 2003)

38. A Ghasemi, ES Souca, Asymptotic performance of collaborative spectrum sensing under correlated log-normal shadowing. IEEE Commun. Lett. **11**(1), 34–36 (2007)

39. T Muetze, P Stuedi, F Kuhn, G Alonso, in *Proc. of IEEE SECON,* San Francisco. Understanding radio irregularity in wireless networks (Piscataway, 2008), pp. 82–90