Spectrum sensing exploiting guard bands and weak channels

Gonzalo Vazquez-Vilar, *Student Member, IEEE*, Roberto López-Valcarce, *Member, IEEE*

***Abstract*—We address the problem of primary user detection in Cognitive Radio from a wideband signal comprising multiple primary channels, exploiting *a priori* knowledge about the primary network: channelization and spectral shape of primary transmissions. Using this second-order statistical information, a multichannel Gaussian model is formulated. In order to obtain a Generalized Likelihood Ratio Test, we first address Maximum Likelihood (ML) estimation of the power levels at the different channels, as well as of the noise variance. The ML conditions suggest a suboptimal closed-form estimate, which takes the form of a constrained Least Squares estimator whose asymptotic efficiency is shown for flat bandpass spectra in white noise, a case of practical importance. The resulting detectors exploit those frequency bins corresponding to guard bands and to primary channels perceived as weak to improve noise variance estimation. Analytical expressions for the probabilities of detection and false alarm are presented. Performance is evaluated via simulations in the setting of a terrestrial TV primary network with realistic channelization parameters.**

I. Introduction

Cognitive Radio [1] is receiving considerable interest as a means for wireless systems to improve spectrum usage [2]. The key idea of opportunistically accessing temporally and/or spatially unused licensed bands requires powerful spectrum monitoring techniques, since the interference produced to licensed (primary) users must be kept at sufficiently low levels [3]. The wireless medium makes reliable detection of primary users a challenging task: due to fading and shadowing phenomena, the received primary signal may be very weak, resulting in very low Signal-to-Noise Ratio (SNR) operation conditions. Although collaborative sensing has the potential to overcome the effects of shadowing [4], [5], it still relies on standalone detectors, whose performance should be optimized.

Certain properties of the primary signal, such as the presence of any pilots or cyclostationary features, could in principle be exploited in order to obtain powerful detectors. However, such approaches often rely on some level of synchronization with the primary signal, and with very low SNR the synchronization loops of the monitoring system cannot be expected to provide a high accuracy for the carrier frequency and/or clock rate estimates, with the associated performance loss of this class of detectors [6]. Among asynchronous schemes, Energy Detection is a popular choice due to its simplicity. However, it requires knowledge of the background noise variance in order to set the threshold to which the received power is compared. If the actual noise variance is different from its nominal value, performance is severely degraded so that the detection/false alarm requirements may not be satisfied [7]. This serious drawback motivates the search for asynchronous detectors robust to noise uncertainty. Our interest is in settings in which some information is available about the primary network, specifically its channelization and emission mask employed. This means that the power spectral density (psd) of the signal transmitted by primary users can be assumed known, up to a scaling. This knowledge should be exploited in order to improve detection performance. Availability of such information is reasonable when operating in bands licensed to wireless systems that make such parameters public. Terrestrial broadcasting (TV or radio) services constitute a prime example of interest, taking into account that cognitive operation in TV "white spaces" is being keenly promoted by regulators [8]–[11] and is likely to constitute the first practical deployment of cognitive radio technology. Other scenarios for application include primary services such as cellular telephony, specialized mobile radio (SMR) and private land mobile radio (PLMR) [12], [13].

If a wide frequency band comprising a large number of primary channels (say N) is to be sensed, different strategies are possible. While it is desirable to capture the whole bandwidth of interest using a wideband analog front end and analog to digital converter (ADC), this is not generally feasible due to stringent requirements on the analog stage, as well as on the sampling rate and resolution of the ADC. On the other hand, selecting one channel at a time poses significant challenges for the voltage controlled oscillator (VCO) in the mixing stage of the front end. The VCO must maintain a minimum performance in terms of phase noise and settling time while being able to sweep over a large frequency range, and these two requirements cannot be met simultaneously [14]. Dividing the band of interest into subbands comprising $M < N$ primary channels allows a trade-off between ADC and VCO complexity. In addition, the wideband approach provides more information about the background noise level, a parameter that must be estimated in practice.

Wideband spectrum sensing has been previously considered by several authors. In [15], knowledge of the noise variance is assumed, but the bandwidths and central frequencies of primary transmissions, as well as their number, are assumed unknown and estimated in turn. In the setting of [16], [17] pri-

The authors are with the Department of Signal Theory and Communications, Universidade de Vigo, 36310 Vigo, Spain. E-mail: {gvazquez,valcarce}@gts.uvigo.es

Research supported by the European Regional Development Fund (ERDF) and the Spanish Government under projects SPROACTIVE (ref. TEC2007-68094-C02-01/TCM), DYNACS (TEC2010-21245-C02-02/TCM) and COMONSENS (CONSOLIDER-INGENIO 2010 CSD2008-00010). Parts of this paper were presented at the 10th IEEE Workshop on Signal Processing Advances for Wireless Communications, Perugia, Italy, June 2009.

mary system channelization is known, and the noise variance is regarded as unknown. However, these methods do not exploit *a priori* information about the psd of primary transmissions, and they assume that a minimum number of unused channels exist in the subband under examination. In [18], the goal is to maximize the throughput of the unlicensed user under suitable interference constraints; to this end, the noise variance and the channel gains from potential primary transmitters to the spectrum monitor are assumed known.

In our setting, no knowledge about these parameters is assumed. Since operation in low SNR conditions is mandatory, we focus on asynchronous strategies. A Gaussian model is adopted for the primary signals. Under these assumptions we pose the composite Hypothesis Testing Problem for the detection of primary signals in the subband of interest. Uniformly Most Powerful Tests [19] do not exist in general for this problem, and a Generalized Likelihood Ratio Test (GLRT) appears as a reasonable approach. To this end, we first address joint ML estimation of the powers of the M primary signals and the background noise variance under the Gaussian model. While there is no closed-form general solution for the ML estimate, its derivation suggests a suboptimum, Least Squares (LS) estimate which can be shown to be asymptotically efficient in certain cases of practical interest, namely for bandlimited signals with flat psd within the passband. We use this estimator to derive a closed-form detection test for this scenario which is robust to noise uncertainty.

The paper is organized as follows. Section II presents the system model and the Hypothesis Testing Problem. The ML and LS estimators are considered in Section III. The LS estimate is used in Section IV to derive a detector for flat bandpass signals in white noise; detection performance is also analyzed in this section. Section VI presents some numerical examples, and final conclusions are drawn in Section VII. The proofs of all results are given in the Appendix.

Notation: lower and uppercase boldface symbols denote vectors and matrices, respectively. For a vector $\boldsymbol{x}$, $\mathrm{diag}(\boldsymbol{x})$ is a diagonal matrix with the elements of $\boldsymbol{x}$ along the diagonal. The $n\times 1$ vector of all ones (zeros) is denoted by $\mathbf{1}_n$ ($\mathbf{0}_n$); the subscript will be dropped whenever vector size is clear from the context. For real-valued $\boldsymbol{x}$, $\boldsymbol{x} > \mathbf{0}$ ($\boldsymbol{x} < \mathbf{0}$) denotes that all elements of $\boldsymbol{x}$ are positive (negative). The pseudoinverse of $\boldsymbol{A}$ is denoted by $\boldsymbol{A}^\dagger$.

II. Problem formulation

The primary network uses Frequency Division Multiple Access (FDMA), with fixed channelization known to the spectrum monitor. A subband comprising M primary channels is selected and the resulting wideband signal is downconverted to baseband. This baseband signal is sampled at its Nyquist rate f_s, thus obtaining K complex-valued samples:

$$\begin{aligned} r_k &= \sum_{i=1}^{M} \sigma_i \tilde{s}_k^{(i)} e^{j(\omega_i k+\phi_i)} + \sigma_0 n_k \\ &= \sum_{i=1}^{M} \sigma_i s_k^{(i)} + \sigma_0 n_k, \qquad 0\le k\le K-1, \end{aligned} \tag{1}$$

where r_k are the observations; $\tilde{s}_k^{(i)}$ are the samples of the i-th baseband equivalent primary signal; ω_i is the relative carrier frequency corresponding to the i-th channel after downconversion to baseband, with ϕ_i the corresponding relative phase offset; n_k are samples of zero mean, circular complex white Gaussian noise with unit variance; $\sigma_i^2 \ge 0$ is the power of the primary signal in the i-th channel ($\sigma_i^2 = 0$ if the i-th channel is idle); and $\sigma_0^2 > 0$ is the noise power. We assume that $\tilde{s}_k^{(i)}$ (and therefore $s_k^{(i)} \doteq \tilde{s}_k^{(i)} e^{j(\omega_i k+\phi_i)}$ as well) is circular complex Gaussian, normalized to unit variance: $E\{|\tilde{s}_k^{(i)}|^2\} = E\{|s_k^{(i)}|^2\} = 1$. The reasons for adopting a Gaussian model for the primary signal are as follows. First, under asynchronous sampling, the actual distribution is unknown; and since the noise is assumed Gaussian as well, the Gaussian pdf for the signal is the least informative one for the detection problem. Second, if the primary system uses multicarrier modulation with a sufficiently large number of subcarriers (which is the case in e.g. broadcasting applications), the Gaussian model is accurate [20]. Third, this model is tractable and leads to useful detectors under other distributions: note that Gaussianity is a common assumption in the development of signal detectors, either explicitly [17], [21] or implicitly, as many *ad hoc* methods that limit themselves to the use of second-order statistics of the observations can often be derived from a Gaussian model (the Energy Detector is the most prominent example).

Define $s_k^{(0)} = n_k$, so that (1) can be compactly written as

$$\boldsymbol{r} = \sum_{i=0}^{M} \sigma_i \boldsymbol{s}_i \tag{2}$$

with $\boldsymbol{r} \doteq [r_0\, r_1\, \cdots\, r_{K-1}]^T$ and $\boldsymbol{s}_i \doteq [s_0^{(i)}\, s_1^{(i)}\, \cdots\, s_{K-1}^{(i)}]^T$. Hence, each $\boldsymbol{s}_i$ is zero-mean circular Gaussian with covariance matrix $\boldsymbol{C}_i \doteq E\{\boldsymbol{s}_i \boldsymbol{s}_i^H\}$. Regarding the asynchronously sampled discrete-time processes $s_k^{(i)}$ as wide-sense stationary, $\boldsymbol{C}_i$ is Toeplitz with ones on the diagonal; and $\boldsymbol{C}_0 = \boldsymbol{I}$ since the noise is assumed white. If the channels from the primary transmitters to the monitor are frequency flat[1], then the $\boldsymbol{C}_i$ are known, and they summarize the knowledge about the primary network (channelization and spectral shape of transmissions) available to the spectrum monitor. Since $\boldsymbol{s}_i$, $\boldsymbol{s}_j$ with $1 \le i \ne j \le M$ correspond to different primary transmissions, they are regarded as statistically independent, and also independent of the background noise $\boldsymbol{s}_0$. Hence, under this model the observation $\boldsymbol{r}$ is zero-mean circular Gaussian with covariance

$$\boldsymbol{R}(\boldsymbol{\theta}) \doteq E\{\boldsymbol{r}\boldsymbol{r}^H\} = \sum_{i=0}^{M} \sigma_i^2 \boldsymbol{C}_i, \tag{3}$$

where we make explicit the dependence of $\boldsymbol{R}$ with the vector of unknown parameters $\boldsymbol{\theta} \doteq [\sigma_0^2\ \sigma_1^2\ \cdots\ \sigma_M^2]^T$.

Under the Gaussian model, second-order statistics capture all relevant information about the problem. In order to ensure identifiability of the parameter vector $\boldsymbol{\theta}$ from (3), it is assumed that the $\{\boldsymbol{C}_i\}_{i=0}^M$ are linearly independent. (As it turns out, this

[1]The effect of *unknown* frequency selective channels on the proposed detectors will be considered in Section VI.

condition amounts to requiring that the psds of the signals $\{s_k^{(i)}\}$, denoted by $\{\mathsf{S}_i(e^{j\omega})\}_{i=0}^M$, be linearly independent). Were this not the case, it would be impossible to distinguish among primary users with linearly dependent emission masks. This assumption is clearly valid in FDMA scenarios, which are the focus of our work[2].

The problem is to determine the subset of idle channels in $\{1, \ldots, M\}$. This could be cast as an Hypothesis Testing Problem with 2^M different hypotheses; however, multiple hypothesis testing in the presence of unknown parameters is a difficult problem [19], so we consider instead succesive detection of the M channels, one by one. For the m-th channel, the problem becomes:

$$\mathcal{H}_0^m: \quad \sigma_m^2 = 0 \qquad \text{(primary is absent in channel } m\text{)}, \quad (4)$$

$$\mathcal{H}_1^m: \quad \sigma_m^2 > 0 \qquad \text{(primary is present in channel } m\text{)}. \quad (5)$$

This is a composite problem, since the probability density function (pdf) f of the observations under the two hypotheses depends on the vector of unknown parameters $\boldsymbol{\theta}$. We consider the Generalized Likelihood Ratio Test (GLRT), see e.g. [19]:

$$T_{\text{GLRT}} \doteq \frac{f(\boldsymbol{r} \,|\, \hat{\boldsymbol{\theta}}_{\text{ML},0})}{f(\boldsymbol{r} \,|\, \hat{\boldsymbol{\theta}}_{\text{ML},1})} \underset{\mathcal{H}_1^m}{\overset{\mathcal{H}_0^m}{\gtrless}} \gamma', \quad (6)$$

with γ' a threshold, and $\hat{\boldsymbol{\theta}}_{\text{ML},j}$ the ML estimate of $\boldsymbol{\theta}$ under $\mathcal{H}_j^m$. Conditioned on $\boldsymbol{\theta}$, the observations are Gaussian distributed:

$$f(\boldsymbol{r} \,|\, \boldsymbol{\theta}) = \frac{\exp\left\{-\boldsymbol{r}^H \boldsymbol{R}^{-1}(\boldsymbol{\theta}) \boldsymbol{r}\right\}}{\pi^K \det \boldsymbol{R}(\boldsymbol{\theta})}. \quad (7)$$

Note that $\hat{\boldsymbol{\theta}}_{\text{ML},1}$ is the maximizer of (7) w.r.t. $\boldsymbol{\theta}$ subject to $\sigma_i^2 \geq 0$, whereas $\hat{\boldsymbol{\theta}}_{\text{ML},0}$ is obtained by fixing $\sigma_m^2 = 0$ and maximizing (7) w.r.t. the remaining parameters in $\boldsymbol{\theta}$ under the same constraint. Consequently, one has $f(\boldsymbol{r} \,|\, \hat{\boldsymbol{\theta}}_{\text{ML},1}) \geq f(\boldsymbol{r} \,|\, \hat{\boldsymbol{\theta}}_{\text{ML},0})$, so that the test statistic in (6) satisfies $0 \leq T_{\text{GLRT}} \leq 1$.

III. Parameter Estimation

It is seen in (7) that the unknown parameter $\boldsymbol{\theta}$ appears in the pdf through the covariance matrix $\boldsymbol{R}(\boldsymbol{\theta})$ only. Therefore the problem reduces to the estimation of a covariance matrix with structure given by (3) with $\sigma_i^2 \geq 0$ for $i = 0, \ldots, M$ and thus it fits in the framework addresed by Burg *et al.* in [22]. Here we follow a slightly different approach to derive the conditions on the unconstrained ML estimate, which leads to a simplified closed-form estimator which is asymptotically efficient for certain cases of practical interest.

A. Unconstrained ML estimation

ML estimation amounts to minimizing the negative of the log-likelihood function

$$L(\boldsymbol{r}; \boldsymbol{\theta}) \doteq \ln \det \boldsymbol{R}(\boldsymbol{\theta}) + \boldsymbol{r}^H \boldsymbol{R}^{-1}(\boldsymbol{\theta}) \boldsymbol{r}. \quad (8)$$

The partial derivatives of $L(\boldsymbol{r}; \boldsymbol{\theta})$ w.r.t. σ_i^2 are

$$\frac{\partial L(\boldsymbol{r}; \boldsymbol{\theta})}{\partial \sigma_i^2} = -\operatorname{Tr}\{\boldsymbol{R}^{-1}(\boldsymbol{\theta}) \boldsymbol{C}_i\} + \boldsymbol{r}^H \boldsymbol{R}^{-1}(\boldsymbol{\theta}) \boldsymbol{C}_i \boldsymbol{R}^{-1}(\boldsymbol{\theta}) \boldsymbol{r}. \quad (9)$$

[2]Linear independence of emission masks may not hold if, for example, different primary users share the same bandwidth using Code Division Multiple Access.

Neglecting the positivity constraints $\sigma_j^2 \geq 0$, the unconstrained ML estimate of $\boldsymbol{\theta}$ satisfies $\partial L / \partial \sigma_i^2 = 0$ for $0 \leq i \leq M$. In view of (9), the natural approach to solving these equations seems to be the diagonalization of the matrices involved. While in general the $K \times K$ covariance matrices $\boldsymbol{C}_i$ need not share common eigenvectors, we can exploit a well-known result regarding the asymptotic diagonalization of Toeplitz matrices [19]: as $K \to \infty$, the following approximation holds:

$$\boldsymbol{C}_i \approx \boldsymbol{W} \boldsymbol{\Lambda}_i \boldsymbol{W}^H, \qquad i = 0, 1, \ldots, M, \quad (10)$$

where $\boldsymbol{W}$ denotes the $K \times K$ orthonormal IDFT matrix, $\boldsymbol{\Lambda}_i \doteq \operatorname{diag}(\boldsymbol{\lambda}_i)$, and $\boldsymbol{\lambda}_i \doteq [\lambda_0^{(i)} \; \lambda_1^{(i)} \; \cdots \; \lambda_{K-1}^{(i)}]^T$ with

$$\lambda_k^{(i)} \doteq \mathsf{S}_i(e^{j\frac{2\pi k}{K}}), \quad 0 \leq k \leq K-1. \quad (11)$$

The approximation in (10) is justified by the asymptotic equivalence of the sequences of matrices $\{\boldsymbol{C}_i\}$ and $\{\boldsymbol{W} \boldsymbol{\Lambda}_i \boldsymbol{W}^H\}$ for $K = 1, 2, \ldots$ [23], and has been exploited extensively in the literature; as shown in [24], the loss in detection performance when adopting this approximation often becomes negligible for very moderate values of K.

Substituting now (10) into (3), it follows that, as $K \to \infty$,

$$\boldsymbol{R} \approx \boldsymbol{W} \boldsymbol{\Delta}(\boldsymbol{\theta}) \boldsymbol{W}^H, \qquad \text{with} \qquad \boldsymbol{\Delta}(\boldsymbol{\theta}) \doteq \sum_{i=0}^{M} \sigma_i^2 \boldsymbol{\Lambda}_i. \quad (12)$$

Note that $\boldsymbol{\Delta}(\boldsymbol{\theta}) = \operatorname{diag}\{[\; \delta_0(\boldsymbol{\theta}) \;\; \delta_1(\boldsymbol{\theta}) \;\; \cdots \;\; \delta_{K-1}(\boldsymbol{\theta}) \;]\}$ contains uniformly spaced samples of the psd of the observations, given by

$$\delta_k(\boldsymbol{\theta}) \doteq \sum_{j=0}^{M} \sigma_j^2 \lambda_k^{(j)}, \qquad 0 \leq k \leq K-1. \quad (13)$$

With this asymptotic diagonalization of $\boldsymbol{R}$, we can substitute (12) back into (9) to obtain

$$\frac{\partial L(\boldsymbol{r}; \boldsymbol{\theta})}{\partial \sigma_i^2} \approx -\operatorname{Tr}\left\{\boldsymbol{\Delta}^{-1}(\boldsymbol{\theta}) \boldsymbol{\Lambda}_i\right\} + \boldsymbol{v}^H \boldsymbol{\Delta}^{-1}(\boldsymbol{\theta}) \boldsymbol{\Lambda}_i \boldsymbol{\Delta}^{-1}(\boldsymbol{\theta}) \boldsymbol{v}, \quad (14)$$

where $\boldsymbol{v} \doteq \boldsymbol{W}^H \boldsymbol{r} = [v_0 \, v_1 \, \cdots \, v_{K-1}]^T$ is the DFT of the observations. Then, equating (14) to zero, we find that as $K \to \infty$ the unconstrained ML estimate $\hat{\boldsymbol{\theta}}_{\text{ML}}$ will satisfy

$$\sum_{k=0}^{K-1} \frac{\lambda_k^{(i)}}{\delta_k(\hat{\boldsymbol{\theta}}_{\text{ML}})} = \sum_{k=0}^{K-1} \frac{|v_k|^2 \lambda_k^{(i)}}{\delta_k^2(\hat{\boldsymbol{\theta}}_{\text{ML}})}, \quad 0 \leq i \leq M. \quad (15)$$

While in general it is not possible to obtain $\hat{\boldsymbol{\theta}}_{\text{ML}}$ in closed form[3] from the conditions (15), it is possible to obtain approximate closed form solutions as we will see next.

B. Unconstrained Least Squares estimation

The left-hand side of (15) can be rewritten as

$$\sum_{k=0}^{K-1} \frac{\lambda_k^{(i)}}{\delta_k} = \sum_{k=0}^{K-1} \frac{\delta_k \lambda_k^{(i)}}{\delta_k^2} \quad (16)$$

$$= \sum_{j=0}^{M} \sigma_j^2 \left(\sum_{k=0}^{K-1} \frac{\lambda_k^{(j)} \lambda_k^{(i)}}{\delta_k^2} \right). \quad (17)$$

[3]These nonlinear equations can be solved numerically by efficient fixed-point iterative algorithms [22], [25].

Substituting (17) into (15), one obtains, in matrix form,

$$\boldsymbol{L}^H\boldsymbol{\Delta}^{-2}(\hat{\boldsymbol{\theta}}_{\text{ML}})\boldsymbol{L}\hat{\boldsymbol{\theta}}_{\text{ML}} = \boldsymbol{L}^H\boldsymbol{\Delta}^{-2}(\hat{\boldsymbol{\theta}}_{\text{ML}})\boldsymbol{p}, \tag{18}$$

where the $K\times(M+1)$ matrix $\boldsymbol{L}$ and the $K\times 1$ vector $\boldsymbol{p}$ (the periodogram) are respectively defined as

$$\boldsymbol{L} \doteq [\ \boldsymbol{\lambda}_0 \quad \boldsymbol{\lambda}_1 \quad \cdots \quad \boldsymbol{\lambda}_M\ ], \tag{19}$$

$$\boldsymbol{p} \doteq [\ |v_0|^2 \quad |v_1|^2 \quad \cdots \quad |v_{K-1}|^2\ ]^T. \tag{20}$$

Note that the periodogram $\boldsymbol{p}$ is an asymptotically unbiased estimate of the psd of the observations [26], and therefore $\boldsymbol{p}_\star \doteq \lim_{K\to\infty} E\{\boldsymbol{p}\} = \boldsymbol{L}\boldsymbol{\theta}$, with $\boldsymbol{\theta}$ the vector of true parameters. Thus, asymptotically, the expected value of $\boldsymbol{p}$ lies in the subspace spanned by the columns of $\boldsymbol{L}$. The linear independence assumption on the psds $\{\mathsf{S}_i(e^{j\omega}),\ 0\le i\le M\}$ implies that $\boldsymbol{L}$ has full column rank, so that $\boldsymbol{L}^\dagger\boldsymbol{L} = \boldsymbol{I}_{M+1}$. Then it holds that

$$\boldsymbol{L}\boldsymbol{L}^\dagger\boldsymbol{p}_\star = \boldsymbol{L}\boldsymbol{L}^\dagger\boldsymbol{L}\boldsymbol{\theta} = \boldsymbol{L}\boldsymbol{\theta} = \boldsymbol{p}_\star, \tag{21}$$

which suggests the approximation $\boldsymbol{p}\approx\boldsymbol{L}\boldsymbol{L}^\dagger\boldsymbol{p}$. Substituting this in (18),

$$\begin{aligned}\hat{\boldsymbol{\theta}}_{\text{ML}} &\approx \left[\boldsymbol{L}^H\boldsymbol{\Delta}^{-2}(\hat{\boldsymbol{\theta}}_{\text{ML}})\boldsymbol{L}\right]^{-1}\boldsymbol{L}^H\boldsymbol{\Delta}^{-2}(\hat{\boldsymbol{\theta}}_{\text{ML}})\boldsymbol{L}\boldsymbol{L}^\dagger\boldsymbol{p}\\ &= \boldsymbol{L}^\dagger\boldsymbol{p} \doteq \hat{\boldsymbol{\theta}}_{\text{LS}}.\end{aligned} \tag{22}$$

The LS subscript refers to the fact that this estimate is the solution to the unconstrained Least Squares problem $\min_{\hat{\boldsymbol{\theta}}}\|\boldsymbol{L}\hat{\boldsymbol{\theta}}-\boldsymbol{p}\|_2^2$. The rows of $\boldsymbol{L}^\dagger$ can be interpreted as *matched filters* that combine the power in the different frequency bins (the entries of $\boldsymbol{p}$) in order to estimate the variances in each channel. The LS estimate is asymptotically unbiased, with covariance given by $\text{Cov}(\hat{\boldsymbol{\theta}}_{\text{LS}}) = \boldsymbol{L}^\dagger\,\text{Cov}(\boldsymbol{p})(\boldsymbol{L}^\dagger)^H$. Since the asymptotic covariance of $\boldsymbol{p}$ is given by $\lim_{K\to\infty}\boldsymbol{\Delta}^2(\boldsymbol{\theta})$ [26], one finds that

$$\lim_{K\to\infty}\text{Cov}(\hat{\boldsymbol{\theta}}_{\text{LS}}) = \lim_{K\to\infty}\boldsymbol{L}^\dagger\boldsymbol{\Delta}^2(\boldsymbol{\theta})(\boldsymbol{L}^\dagger)^H. \tag{23}$$

C. Cramér-Rao Lower Bound

Under the Gaussian model, the elements of the Fisher information matrix (FIM) $\boldsymbol{F}(\boldsymbol{\theta})$ are given by (see e.g. [27]):

$$[\boldsymbol{F}(\boldsymbol{\theta})]_{ij} = \text{Tr}\left\{\boldsymbol{R}^{-1}(\boldsymbol{\theta})\frac{\partial\boldsymbol{R}(\boldsymbol{\theta})}{\partial\sigma_i^2}\boldsymbol{R}^{-1}(\boldsymbol{\theta})\frac{\partial\boldsymbol{R}(\boldsymbol{\theta})}{\partial\sigma_j^2}\right\}. \tag{24}$$

The Cramér-Rao Lower Bound (CRLB) for any unbiased estimator of $\boldsymbol{\theta}$ is then given by $\text{var}(\hat{\sigma}_i^2)\ge[\boldsymbol{F}^{-1}(\boldsymbol{\theta})]_{ii}$. In our case, $\partial\boldsymbol{R}(\boldsymbol{\theta})/\partial\sigma_i^2 = \boldsymbol{C}_i$. Then, using the asymptotic approximations (10) and (12),

$$[\boldsymbol{F}(\boldsymbol{\theta})]_{ij} \approx \text{Tr}\left\{\boldsymbol{\Delta}^{-1}(\boldsymbol{\theta})\boldsymbol{\Lambda}_i\boldsymbol{\Delta}^{-1}(\boldsymbol{\theta})\boldsymbol{\Lambda}_j\right\} \tag{25}$$

$$= \sum_{k=0}^{K-1}\frac{\lambda_k^{(i)}\lambda_k^{(j)}}{\delta_k^2(\boldsymbol{\theta})}. \tag{26}$$

Thus, the asymptotic FIM is given by

$$\lim_{K\to\infty}\boldsymbol{F}(\boldsymbol{\theta}) = \lim_{K\to\infty}\boldsymbol{L}^H\boldsymbol{\Delta}^{-2}(\boldsymbol{\theta})\boldsymbol{L}. \tag{27}$$

D. Quasi-GLRT detection

Let $\hat{\boldsymbol{\theta}}_j = [\ \hat{\sigma}_{0j}^2 \quad \hat{\sigma}_{1j}^2 \quad \cdots \quad \hat{\sigma}_{Mj}^2\ ]^T$ denote an estimate (ML, LS, or other) of $\boldsymbol{\theta}$ under $\mathcal{H}_j^m$, $j\in\{0,1\}$. Using these estimates in the detection test, we obtain an approximation to the GLRT. Using the asymptotic diagonalization (12), we can write

$$\det\boldsymbol{R}(\hat{\boldsymbol{\theta}}_j)\approx\det\boldsymbol{\Delta}(\hat{\boldsymbol{\theta}}_j) = \prod_{k=0}^{K-1}\left[\sum_{i=0}^{M}\hat{\sigma}_{ij}^2\lambda_k^{(i)}\right], \tag{28}$$

$$\boldsymbol{r}^H\boldsymbol{R}^{-1}(\hat{\boldsymbol{\theta}}_j)\boldsymbol{r}\approx\boldsymbol{v}^H\boldsymbol{\Delta}^{-1}(\hat{\boldsymbol{\theta}}_j)\boldsymbol{v} = \sum_{k=0}^{K-1}\frac{p_k}{\sum_{i=0}^{M}\hat{\sigma}_{ij}^2\lambda_k^{(i)}}. \tag{29}$$

The resulting "Quasi-GLRT" (QGLRT) can be written as

$$\begin{aligned}\log\frac{f(\boldsymbol{r}\,|\,\hat{\boldsymbol{\theta}}_0)}{f(\boldsymbol{r}\,|\,\hat{\boldsymbol{\theta}}_1)}\approx\log T\doteq&\sum_{k=0}^{K-1}\log\left[\frac{\sum_{i=0}^{M}\hat{\sigma}_{i1}^2\lambda_k^{(i)}}{\sum_{i=0}^{M}\hat{\sigma}_{i0}^2\lambda_k^{(i)}}\right]\\ &+\sum_{k=0}^{K-1}\left[\frac{p_k}{\sum_{i=0}^{M}\hat{\sigma}_{i1}^2\lambda_k^{(i)}}-\frac{p_k}{\sum_{i=0}^{M}\hat{\sigma}_{i0}^2\lambda_k^{(i)}}\right]\end{aligned} \tag{30}$$

This detector can be implemented once $\hat{\boldsymbol{\theta}}_j$ are available, either by numerical means (ML) or in closed form (LS). However, it is difficult in general to evaluate the performance of this detector or to obtain some intuition about its operation. In the next section we focus on a particular scenario whose structure will allow further simplification of (30).

IV. Detection of Orthogonal Frequency-flat Signals in White Noise

For FDMA-based primary networks, the signals in different channels are orthogonal, i.e. their psds have disjoint supports. In addition, the psd of a multicarrier signal is approximately constant within its support. The QGLRT (30) will be particularized to this setting.

Definition 1: A signal is *frequency-flat bandpass* if its psd takes only two levels: zero or a given constant value.

Definition 2: Two signals $s_k^{(i)}$, $s_k^{(j)}$ are *non-partially overlapping* if either their psds have disjoint supports, or the support of one of them contains that of the other.

For this class of signals, it turns out that the LS estimate (22) is asymptotically efficient:

Theorem 1: If the signals $s_k^{(i)}$ and $s_k^{(j)}$ are frequency-flat bandpass and non-partially overlapping for any $0\le i,j\le M$, then the asymptotic covariance matrix (23) of the unconstrained LS estimate equals the inverse of the asymptotic FIM (27).

In the following we will assume that $s_k^{(i)}$, $i=1,\ldots,M$, are frequency-flat bandpass with disjoint frequency supports. Since $s_k^{(0)}$ (white noise) is frequency-flat bandpass covering the whole bandwidth, it follows that $s_k^{(i)}$, $s_k^{(j)}$ are non-partially overlapping for any $0\le i,j\le M$. This is a special case of the broader family of non-partially overlapping frequency-flat bandpass signals, and will be denoted here as orthogonal frequency-flat signals in white noise. For this class of signals, Theorem 1 motivates the use of LS estimates in the QGLRT.

A. QGLRT with unconstrained LS estimates

Let $\mathcal{W}_i$ denote the set of frequency bins within the support of $\mathsf{S}_i(e^{j\omega})$, $i = 1,\ldots, M$, and let the set of "noise-only" frequency bins (comprising all guard bands in the captured bandwidth) be

$$\mathcal{W}_0 \doteq \{k : k \in \{0, 1, \ldots, K-1\} \text{ and } k \notin \cup_{i=1}^{M} \mathcal{W}_i\}. \quad (31)$$

We also define the *fractional bandwidths* $w_i \doteq |\mathcal{W}_i|/K$, $0 \leq i \leq M$, such that $0 < w_i < 1$ and $\sum_{i=0}^{M} w_i = 1$. Since the signals are normalized to unit variance, it follows that

$$[\boldsymbol{L}]_{ki} = \lambda_k^{(i)} = \begin{cases} 1, & i = 0, \\ \frac{1}{w_i}, & k \in \mathcal{W}_i,\ i = 1, \ldots, M, \\ 0, & \text{otherwise.} \end{cases} \quad (32)$$

The pseudoinverse of $\boldsymbol{L}$ is given in this case by

$$[\boldsymbol{L}^{\dagger}]_{ik} = \begin{cases} \frac{1}{Kw_0}, & k \in \mathcal{W}_0,\ i = 0, \\ \frac{-w_i}{Kw_0}, & k \in \mathcal{W}_0,\ i \neq 0, \\ \frac{1}{K}, & k \in \mathcal{W}_i,\ i \neq 0, \\ 0, & k \in \mathcal{W}_j,\ j \neq i,\ j \neq 0. \end{cases} \quad (33)$$

Denote the averaged periodogram over $\mathcal{W}_i$ by $q_i \doteq \frac{1}{Kw_i} \sum_{k \in \mathcal{W}_i} p_k$, $0 \leq i \leq M$. The unconstrained LS estimate of $\boldsymbol{\theta}$ under $\mathcal{H}_1^m$ is just $\boldsymbol{L}^{\dagger}\boldsymbol{p}$, and is given by

$$\hat{\sigma}_{i1}^2 = \begin{cases} q_0, & i = 0, \\ w_i(q_i - q_0), & i \neq 0. \end{cases} \quad (34)$$

On the other hand, the unconstrained LS estimate under $\mathcal{H}_0^m$ is such that the subband corresponding to channel m is consolidated into the "noise-only" set:

$$\hat{\sigma}_{i0}^2 = \begin{cases} q_{0m}, & i = 0, \\ 0, & i = m, \\ w_i(q_i - q_{0m}), & i \neq 0, i \neq m, \end{cases} \quad (35)$$

where $q_{0m} \doteq (w_0 q_0 + w_m q_m)/(w_0 + w_m)$. If the estimates (34)-(35) are used in the QGLRT, then some straightforward algebra shows that (30) reduces to

$$\frac{1}{K} \log T = (w_0 + w_m) \log \frac{(q_0^{w_0} q_m^{w_m})^{\frac{1}{w_0 + w_m}}}{q_{0m}}. \quad (36)$$

The argument of the log in (36) is the *weighted geometric to arithmetic mean ratio* of q_0 and q_m, with respective weights w_0, w_m. This ratio, which is a function of q_m/q_0 alone, is always less than or equal to one, with equality iff $q_0 = q_m$. It is monotonically increasing for $q_m/q_0 < 1$, and decreasing for $q_m/q_0 > 1$. Thus, the QGLRT with unconstrained LS estimates decides that channel m is idle if $q_m/q_0 \in [\alpha, \beta]$, for some $\alpha < 1 < \beta$ depending on the threshold, and busy otherwise. This is against intuition, since $q_m < q_0$ is always a reasonable indicator of an idle channel. As the number of samples K increases, one has $q_m/q_0 \to 1 + \sigma_m^2/(\sigma_0^2 w_m) \geq 1$, which suggests the use of a simplified test

$$\frac{q_m}{q_0} \underset{\mathcal{H}_0^m}{\overset{\mathcal{H}_1^m}{\gtrless}} \gamma \qquad \Leftrightarrow \qquad q_m \underset{\mathcal{H}_0^m}{\overset{\mathcal{H}_1^m}{\gtrless}} \gamma q_0, \quad (37)$$

in which the power measured in channel m is compared to a threshold that now depends on the noise power measured in the guard bands. Were this noise power known, the threshold could be set independently of the observations, resulting in the commonly employed energy detector.

Intuition suggests that the test statistic should be monotonic in q_m/q_0. Note that if $q_m < q_0$, then the unconstrained LS estimate of σ_m^2 under $\mathcal{H}_1^m$ is $\hat{\sigma}_{m1}^2 = w_m(q_m - q_0) < 0$, which is against prior knowledge. This motivates the use of *constrained* estimators in the QGLRT.

B. QGLRT with constrained LS estimates

The constrained LS estimate of $\boldsymbol{\theta}$ for orthogonal frequency-flat signals in white noise is given next.

Theorem 2: Let $\boldsymbol{L}$ be given by (32). The minimizer of $\|\boldsymbol{L}\hat{\boldsymbol{\theta}} - \boldsymbol{p}\|_2^2$ subject to $\hat{\sigma}_i^2 \geq 0$, $0 \leq i \leq M$ (i.e. the constrained LS estimate under $\mathcal{H}_1^m$), takes the following form:

$$\hat{\sigma}_{j1}^2 = \begin{cases} \frac{w_0 q_0 + \sum_{l \in \mathcal{U}_1} w_l q_l}{w_0 + \sum_{l \in \mathcal{U}_1} w_l}, & j = 0, \\ 0, & j \in \mathcal{U}_1, \\ w_j(q_j - \hat{\sigma}_{01}^2), & \text{otherwise,} \end{cases} \quad (38)$$

where $\mathcal{U}_1 \doteq \{j : q_j < \hat{\sigma}_{01}^2,\ j \neq 0\}$.

Analogously, the minimizer of $\|\boldsymbol{L}\hat{\boldsymbol{\theta}} - \boldsymbol{p}\|_2^2$ subject to $\hat{\sigma}_m = 0$, $\hat{\sigma}_i^2 \geq 0$, $i \neq m$ (i.e. the constrained LS estimate under $\mathcal{H}_0^m$), is given by:

$$\hat{\sigma}_{j0}^2 = \begin{cases} \frac{w_0 q_0 + w_m q_m + \sum_{l \in \mathcal{U}_0} w_l q_l}{w_0 + w_m + \sum_{l \in \mathcal{U}_0} w_l}, & j = 0, \\ 0, & j \in \mathcal{U}_0 \cup \{m\}, \\ w_j(q_j - \hat{\sigma}_{00}^2), & \text{otherwise,} \end{cases} \quad (39)$$

where $\mathcal{U}_0 \doteq \{j : q_j < \hat{\sigma}_{00}^2,\ j \neq 0,\ j \neq m\}$.

Note that (38)-(39) are implicit expressions, since they depend on the sets $\mathcal{U}_1$, $\mathcal{U}_0$ whose definitions are in terms of $\hat{\sigma}_{01}^2$ and $\hat{\sigma}_{00}^2$ respectively. Nevertheless, these estimates and sets can be easily obtained using the algorithm in Table I. It is straightforward to verify that this algorithm outputs sets $\mathcal{U}_1$, $\mathcal{U}_0$ and estimates $\{\hat{\sigma}_{j1}^2\}$, $\{\hat{\sigma}_{j0}^2\}$ satisfying (38) and (39) respectively. This scheme successively includes the "weakest" channel (i.e., the channel with smallest averaged periodogram over the corresponding frequency support) into the computation of the noise variance estimate, until this estimate falls below the estimated power levels of the remaining channels.

In order to obtain the QGLRT based on the constrained LS estimates above, we distinguish two cases, depending on the strength with which channel m is perceived relative to the noise level. Note that by construction, $m \notin \mathcal{U}_0$, whereas m may or may not belong in $\mathcal{U}_1$.

Case 1: $m \in \mathcal{U}_1$, so that channel m is perceived as "weak" under $\mathcal{H}_1^m$. Then we have the following.

Proposition 1: If $m \in \mathcal{U}_1$, then $\mathcal{U}_1 = \mathcal{U}_0 \cup \{m\}$, so that the constrained LS estimates under $\mathcal{H}_1^m$ and $\mathcal{H}_0^m$ are the same.

Therefore, if $m \in \mathcal{U}_1$, then from (30) we have $\log T = 0$, i.e. the QGLRT declares channel m as idle.

Case 2: $m \notin \mathcal{U}_1$, so that channel m is perceived as "not weak" under $\mathcal{H}_1^m$. Then one has:

Proposition 2: If $m \notin \mathcal{U}_1$, then $\mathcal{U}_1 \subseteq \mathcal{U}_0$.

Hence, if $m \notin \mathcal{U}_1$, then $\mathcal{U}_0 = \mathcal{U}_1 \cup \mathcal{S}$ for some set $\mathcal{S}$ with $\mathcal{S} \cap \mathcal{U}_1 = \emptyset$. For $j \notin \mathcal{U}_0 \cup \{0, m\}$, it turns out that $w_j \hat{\sigma}_{01}^2 + \hat{\sigma}_{j1}^2 = w_j \hat{\sigma}_{00}^2 + \hat{\sigma}_{j0}^2$, and thus these indices do not

TABLE I
COMPUTATION OF THE CONSTRAINED LS ESTIMATE.

Under $\mathcal{H}_1^m$:	Under $\mathcal{H}_0^m$:
1) Set $\mathcal{U}_1 = \emptyset$.	1) Set $\mathcal{U}_0 = \emptyset$.
2) Set $\hat{\sigma}^2_{j1} = 0$ for all $j \in \mathcal{U}_1$.	2) Set $\hat{\sigma}^2_{j0} = 0$ for all $j \in \mathcal{U}_0 \cup \{m\}$.
3) Obtain the unconstrained estimates $\hat{\sigma}^2_{j1}$ for $j \notin \mathcal{U}_1$:	3) Obtain the unconstrained estimates $\hat{\sigma}^2_{j0}$ for $j \notin \mathcal{U}_0 \cup \{m\}$:
$\hat{\sigma}^2_{01} = \frac{w_0 q_0 + \sum_{l\in\mathcal{U}_1} w_l q_l}{w_0 + \sum_{l\in\mathcal{U}_1} w_l}$	$\hat{\sigma}^2_{00} = \frac{w_0 q_0 + w_m q_m + \sum_{l\in\mathcal{U}_0} w_l q_l}{w_0 + w_m + \sum_{l\in\mathcal{U}_0} w_l}$
$\hat{\sigma}^2_{j1} = w_j(q_j - \hat{\sigma}^2_{01}), \quad j \notin \mathcal{U}_1 \cup \{0\}$	$\hat{\sigma}^2_{j0} = w_j(q_j - \hat{\sigma}^2_{00}), \quad j \notin \mathcal{U}_0 \cup \{0, m\}$
4) If the obtained estimate is feasible, stop.	4) If the obtained estimate is feasible, stop.
5) Else, let $j_\star = \arg\min_{j\notin\mathcal{U}_1\cup\{0\}}\{q_j\}$, set $\mathcal{U}_1 \leftarrow \mathcal{U}_1 \cup \{j_\star\}$, and go back to Step 2.	5) Else, let $j_\star = \arg\min_{j\notin\mathcal{U}_0\cup\{0,m\}}\{q_j\}$, set $\mathcal{U}_0 \leftarrow \mathcal{U}_0 \cup \{j_\star\}$, and go back to Step 2.

contribute to the QGLRT (30), which after some algebra is found to yield

$$\frac{1}{K}\log T = \log \frac{(\hat{\sigma}^2_{01})^{w_0+\sum_{j\in\mathcal{U}_1} w_j} q_m^{w_m} \prod_{j\in\mathcal{S}} q_j^{w_j}}{(\hat{\sigma}^2_{00})^{w}}, \qquad (40)$$

where $w \doteq w_0 + w_m + \sum_{j\in\mathcal{U}_0} w_j$. Note that if $\mathcal{S} = \emptyset$, then this ratio becomes a monotonically decreasing function of $q_m/\hat{\sigma}^2_{01} \geq 1$. Hence, if $\mathcal{U}_1 = \mathcal{U}_0$, the QGLRT can be written as

$$\frac{q_m}{\hat{\sigma}^2_{01}} \underset{\mathcal{H}_0^m}{\overset{\mathcal{H}_1^m}{\gtrless}} \gamma \qquad \text{(Test 1)}. \qquad (41)$$

Similarly to (37), Test 1 compares the power measured in channel m to a threshold $\gamma\hat{\sigma}^2_{01}$ proportional to the measured noise power. The main difference is that now this noise power is estimated using not only the guard bands, but also those channels perceived as weak (i.e. channels for which the constrained LS power estimate yields a zero value).

If $\mathcal{S} \neq \emptyset$, then it is not possible to reduce (40) to a simple ratio of averaged powers. A possible approach is to disregard the influence of channels with indexes $j \in \mathcal{S}$, obtaining (41). Another possibility is to take these channels into account in order to obtained a new estimate of the noise power

$$\hat{\sigma}^2_{02} \doteq \frac{w_0 q_0 + \sum_{j\in\mathcal{U}_0} w_j q_j}{w_0 + \sum_{j\in\mathcal{U}_0} w_j}, \qquad (42)$$

and then use (42) in the following test:

$$\frac{q_m}{\hat{\sigma}^2_{02}} \underset{\mathcal{H}_0^m}{\overset{\mathcal{H}_1^m}{\gtrless}} \gamma \qquad \text{(Test 2)}, \qquad (43)$$

which reduces to (41) if $\mathcal{S} = \emptyset$.

V. STATISTICAL ANALYSIS

The detectors from Section IV-B are based on the random variables q_i, $0 \leq i \leq M$. Note that $\boldsymbol{v} = \boldsymbol{W}^H \boldsymbol{r}$ is Gaussian with a diagonal (asymptotic) covariance matrix $\boldsymbol{\Delta}(\boldsymbol{\theta})$. For orthogonal frequency-flat signals in white noise, the diagonal of $\boldsymbol{\Delta}(\boldsymbol{\theta})$ is piecewise constant, and in particular, its elements are constant over the bins corresponding to the i-th channel (this also applies to the set of "noise-only" bins). Hence, q_i is the sum of square magnitudes of zero-mean Gaussian random variables, asymptotically uncorrelated and with the same variance. Thus, for large K, q_i becomes chi-squared distributed with Kw_i degrees of freedom; in turn, as $K \to \infty$, this distribution converges to a Gaussian distribution: $q_i \sim \mathcal{N}(\mu_i, \alpha_i^2)$. Moreover, q_i, q_j are asymptotically uncorrelated for $i \neq j$, since the two sets of bins used for their computation are disjoint. In terms of the SNR in channel i, defined as $\rho_i \doteq \sigma_i^2/(w_i\sigma_0^2)$, the mean and variance of q_i are given by

$$\mu_i \doteq \begin{cases} \sigma_0^2, & i = 0, \\ \sigma_0^2(1+\rho_i), & i > 0, \end{cases} \qquad (44)$$

$$\alpha_i^2 \doteq \begin{cases} \frac{\sigma_0^4}{Kw_0}, & i = 0, \\ \frac{\sigma_0^4}{Kw_i}(1+\rho_i)^2, & i > 0. \end{cases} \qquad (45)$$

A. Single-channel detection with guard bands

As a first step, we analyze the case $M = 1$, for which all of the proposed detectors boil down to the same test. This test can be expressed as $z \doteq q_1 - \gamma q_0 \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} 0$, where $\gamma > 1$ is a threshold and the statistic z follows a Gaussian distribution:

$$z \sim \mathcal{N}\left(\sigma_0^2(1+\rho_1-\gamma),\, \frac{\sigma_0^4}{K}\left[\frac{(1+\rho_1)^2}{w_1} + \frac{\gamma^2}{w_0}\right]\right). \qquad (46)$$

Since $\sigma_0^2 > 0$, the probabilities of false alarm and detection can be respectively written as $P_{\text{FA}} = \Pr\{(z/\sigma_0^2) > 0 \,|\, \rho_1 = 0\}$ and $P_{\text{D}} = \Pr\{(z/\sigma_0^2) > 0 \,|\, \rho_1 > 0\}$. These probabilities do not depend on the noise power σ_0^2, as expected. In order to set the threshold γ, two approaches are possible:

1) Threshold for fixed P_{FA}: In view of (46), it is readily found that, for $P_{\text{FA}} \leq 0.5$,

$$\gamma(P_{\text{FA}}) = \frac{1 + \sqrt{1 - \left(1 - \frac{[Q^{-1}(P_{\text{FA}})]^2}{Kw_0}\right)\left(1 - \frac{[Q^{-1}(P_{\text{FA}})]^2}{Kw_1}\right)}}{1 - \frac{[Q^{-1}(P_{\text{FA}})]^2}{Kw_0}}, \qquad (47)$$

where $Q(\cdot)$ is the complementary Gaussian cumulative distribution function, and $Q^{-1}(\cdot)$ denotes its inverse. The resulting probability of detection for an SNR ρ_1 is then given by

$$P_{\text{D}} = Q\left(\sqrt{Kw_0}\frac{\gamma(P_{\text{FA}}) - (1+\rho_1)}{\sqrt{\gamma^2(P_{\text{FA}}) + \frac{w_0}{w_1}(1+\rho_1)^2}}\right). \qquad (48)$$

2) Threshold for fixed P_D: In the context of cognitive radio systems, a false alarm results in a missed opportunity of using an idle channel, and therefore P_{FA} is related to the throughput efficiency of the secondary system. However, this parameter is irrelevant to the primary network. On the other hand, a missed detection may result in the secondary user accessing a channel in use, thus producing interference to

the primary system. Regulatory bodies are likely to require a minimum detection performance to avoid collisions with primary (licensed) users [8], i.e. $P_{\rm D} \geq P_{\rm D}^*$ at some target SNR $\rho_1^\star$. The threshold γ is then determined for $P_{\rm D}^* \geq 0.5$ as

$$\gamma(P_{\rm D}^*;\rho_1^\star) = (1+\rho_1^\star)\frac{1-\sqrt{1-\left(1-\frac{\kappa}{Kw_0}\right)\left(1-\frac{\kappa}{Kw_1}\right)}}{1-\frac{\kappa}{Kw_0}}, \tag{49}$$

with $\kappa \doteq [Q^{-1}(P_{\rm D}^*)]^2$. This yields

$$P_{\rm FA} = Q\left(\sqrt{Kw_0}\frac{\gamma(P_{\rm D}^*;\rho_1^\star)-1}{\sqrt{\gamma^2(P_{\rm D}^*;\rho_1^\star)+\frac{w_0}{w_1}}}\right). \tag{50}$$

Comparison with previous detectors: In order to study the benefits of the proposed approach in the single-channel setting, we compare its performance with that of two state-of-the-art detectors that exploit prior knowledge on the second-order statistics of primary signal. This is a fair comparison, since this is also the information available to the proposed detector.

- In [21] Quan *et al.* propose a detector which correlates the (known) spectral shape of the primary signal with the periodogram of the observations; the resulting statistic is then compared with a noise-dependent threshold. In the case of flat bandpass signals, this amounts to using q_1 as test statistic.
  For a fixed threshold γ the distribution of q_1 yields $P_{\rm FA} = Q(\sqrt{Kw_1}(\frac{\gamma}{\sigma_0^2}-1))$ and $P_{\rm D} = Q(\sqrt{Kw_1}(\frac{\gamma}{\sigma_0^2(1+\rho_1)}-1))$. If the threshold is set for a given value $P_{\rm D}^*$ at a target SNR ρ_1^* using a nominal value $\hat{\sigma}_0^2$ of the noise power, the resulting performance is given by

$$P_{\rm FA} = Q\left(\sqrt{Kw_1}\left(\frac{\hat{\sigma}_0^2}{\sigma_0^2}\left(1+\frac{Q^{-1}(P_{\rm D}^*)}{\sqrt{Kw_1}}\right)(1+\rho_1^*)-1\right)\right), \tag{51}$$

$$P_{\rm D} = Q\left(\sqrt{Kw_1}\left(\frac{\hat{\sigma}_0^2}{\sigma_0^2}\left(1+\frac{Q^{-1}(P_{\rm D}^*)}{\sqrt{Kw_1}}\right)\frac{(1+\rho_1^*)}{(1+\rho_1)}-1\right)\right). \tag{52}$$

- Chaudhari *et al.* propose in [28] a detector exploiting the presence of the cyclic prefix in multicarrier transmissions. The test statistic is given by the empirical autocorrelation coefficient at the lag corresponding to the useful duration T_d of the multicarrier symbol (normalized by the estimated received power).

Fig. 1 shows the results obtained by the proposed detector and by those of Quan *et al.* and Chaudhari *et al.* in a single-channel scenario in which the channel and signal bandwidths are 8 MHz and 7.61 MHz respectively ($w_1 \approx 0.95$, $w_0 \approx 0.05$). Multicarrier modulation is assumed, with useful symbol duration $T_d = 896$ μs and cyclic prefix length $T_c = 224$ μs. These parameters are consistent with those of the DVB-T standard for terrestrial digital TV broadcasting for 8 MHz channels with 8K mode and guard interval 1/4 [29]. The sampling rate is $f_s = 8$ Msps and the observation interval is 2.5 ms, thus $K = 20 \cdot 10^3$ samples. The detection thresholds

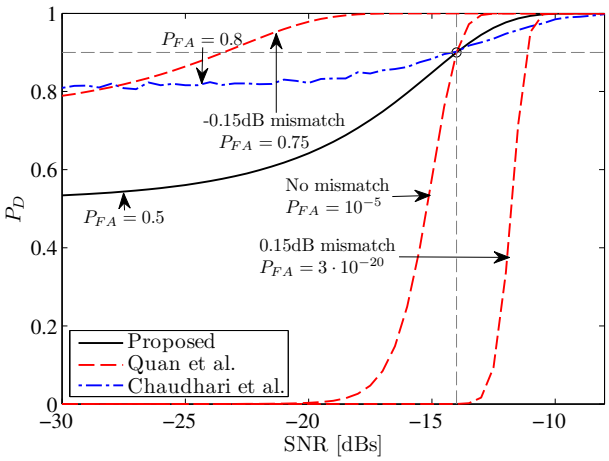


Fig. 1. Performance comparison of different detectors (single-channel case). Thresholds are set for a probability of detection of 0.9 at a target SNR of -14 dB. $K = 20 \cdot 10^3$, $w_0 = 0.05$, $w_1 = 0.95$. Noise mismatch is defined as the ratio of nominal to actual noise powers, $\hat{\sigma}_0^2/\sigma_0^2$.

are tuned in order to satisfy the requirement that $P_{\rm D} \geq 0.9$ for SNR ≥ -14 dB.

In Fig. 1 analytical results are shown for the proposed detector and that of Quan *et al.*, whereas the performance of the test of Chaudhari *et al.* was evaluated by Monte Carlo simulation using a synthetic multicarrier signal with the parameters given above. When σ_0^2 is exactly known, the detector of Quan *et al.* yields the best performance, satisfying the constraints with $P_{\rm FA} \approx 10^{-5}$. However, this scheme is severely affected by noise uncertainty: underestimating the noise power greatly increases $P_{\rm FA}$, with the subsequent throughput penalty for the secondary system; whereas overestimating this parameter results in the detection requirements not being satisfied. Although the other two tests are robust to noise uncertainty, the detector of Chaudhari *et al.* results in $P_{\rm FA} \approx 0.8$ and is clearly outperformed by the proposed scheme ($P_{\rm FA} \approx 0.5$). Note that among the possible values for the cyclic prefix length in the DVB-T standard (1/4, 1/8, 1/16 and 1/32), the 1/4 choice is the most favorable one for the detector of Chaudhari *et al.* In terms of computational load, the proposed detector and that of Quan *et al.* are based on the periodogram of the observations, and hence their complexity is basically $\mathcal{O}(K \log K)$ due to the FFT operation. The detector of Chaudhari *et al.* requires the computation of a single autocorrelation coefficient, and thus its complexity is $\mathcal{O}(K)$.

B. Multichannel detection

Single-channel spectrum sensing, as described in Section V-A, exploits the presence of upper and lower guard bands to estimate the noise power. In practice these guard bands may appear distorted due to the transition bands of the analog filter used for channel extraction. This may preclude the use of those frequency components for noise variance estimation. When $M > 1$ channels are simultaneously captured, the guard bands between adjacent channels remain undistorted, and therefore this problem is alleviated.

Without loss of generality, let $m = M$, so that the channel under scrutiny is the M-th one. In order to simplify the presentation, we restrict our analysis to the case in which all channels have the same bandwidth[4]: $w_1 = w_2 = \cdots = w_M$. Whereas finding analytical expressions for the performance of the QGLRT detector (40) seems intractable, an analysis of the simplified tests (41) and (43) is given next. Up to this point, no assumptions have been made about the occupancy of the band in the derivation of these detectors but, as it turns out, in a multichannel scenario their performance depends on the *a priori* probability of any given channel being in use by the primary network. This probability, or *activity factor,* will be denoted by a in what follows.

1) Test 1 from (41): This test is given by $z_1 \gtrless_{\mathcal{H}_0^M}^{\mathcal{H}_1^M} 0$, with $z_1 \doteq q_M - \gamma\hat{\sigma}_{01}^2$, $\gamma > 1$, and $\hat{\sigma}_{01}^2$ a linear combination of q_0 and $\{q_j,\, j \in \mathcal{U}_1\}$, as in (38). Denote by I_t the event of having t of the channels 1,…,$M-1$ not in use by the primary network, so that $\Pr\{I_t\} = \binom{M-1}{t} a^{M-1-t}(1-a)^t$. In addition, denote by U_n the event $|\mathcal{U}_1| = n$. Then we can write

$$\begin{aligned}\Pr\{z_1 > 0\} &= \sum_{t=0}^{M-1} \Pr\{I_t\} \Pr\{z_1 > 0 \,|\, I_t\} \\ &= \sum_{t=0}^{M-1} \Pr\{I_t\} \sum_{n=0}^{M-1} \Pr\{z_1 > 0, U_n \,|\, I_t\} \\ &\approx \sum_{t=0}^{M-1} \Pr\{I_t\} \sum_{n=0}^{t} \Pr\{z_1 > 0, U_n \,|\, I_t\}, \end{aligned} \tag{53}$$

where in the last step we have neglected the probability of a *busy* channel $j \neq M$ being included in the set $\mathcal{U}_1$ by the constrained LS estimate. This amounts to assuming that busy channels have sufficiently high SNRs. Without this approximation, $\Pr\{z_1 > 0\}$ would depend on the SNRs of the (busy) channels other than channel M, which is clearly undesirable. The accuracy of this assumption will be validated by the simulation results.

Let us define the vectors

$$\boldsymbol{x}_n \doteq \begin{bmatrix} q_1 \\ \vdots \\ q_n \end{bmatrix} - \hat{\sigma}_{01}^2 \mathbf{1}_n, \quad \boldsymbol{y}_{t-n} \doteq \begin{bmatrix} q_{n+1} \\ \vdots \\ q_t \end{bmatrix} - \hat{\sigma}_{01}^2 \mathbf{1}_{t-n}. \tag{54}$$

Now, when computing (53), we can assume that the t idle channels are channels 1 through t (due to the equal bandwidth assumption), so that

$$\Pr\{z_1 > 0\} \approx \sum_{t=0}^{M-1} \Pr\{I_t\} \sum_{n=0}^{t} \binom{t}{n} \Pr\{z_1 > 0,\, \boldsymbol{x}_n < \mathbf{0},\, \boldsymbol{y}_{t-n} > \mathbf{0} \,|\, I_t\}. \tag{55}$$

Note that $\boldsymbol{x}_n < \mathbf{0}$, $\boldsymbol{y}_{t-n} > \mathbf{0}$ imply that $\mathcal{U}_1 = \{1, \cdots, n\}$, so that $\hat{\sigma}_{01}^2 = (\sum_{l=0}^{n} w_l q_l)/(\sum_{l=0}^{n} w_l)$. Now one has that

$$\Pr\{z_1 > 0,\, \boldsymbol{x}_n < \mathbf{0},\, \boldsymbol{y}_{t-n} > \mathbf{0} \,|\, I_t\} = \Pr\{[z_1 \;\; -\boldsymbol{x}_n^T \;\; \boldsymbol{y}_{t-n}^T]^T/\sigma_0^2 > \mathbf{0} \,|\, I_t\}, \tag{56}$$

which is the integral of a $(t+1)$-variate Gaussian distribution over the positive orthant. The mean of this distribution is $\boldsymbol{\mu}_1 = [\,(1-\gamma+\rho_M)\;\mathbf{0}_t^T\,]^T$, and the covariance matrix is found blockwise from the following, where $\bar{w}_n \doteq w_0 + n w_M$:

$$\mathrm{Cov}(z_1, z_1) = \frac{\sigma_0^4}{K}\left(\frac{(1+\rho_M)^2}{w_M} + \frac{\gamma^2}{\bar{w}_n}\right), \tag{57}$$

$$\mathrm{Cov}(\boldsymbol{x}_n, \boldsymbol{x}_n) = \frac{\sigma_0^4}{K}\left(\frac{1}{w_M}\boldsymbol{I} - \frac{1}{\bar{w}_n}\mathbf{1}_n\mathbf{1}_n^T\right), \tag{58}$$

$$\mathrm{Cov}(\boldsymbol{y}_{t-n}, \boldsymbol{y}_{t-n}) = \frac{\sigma_0^4}{K}\left(\frac{1}{w_M}\boldsymbol{I} + \frac{1}{\bar{w}_n}\mathbf{1}_{t-n}\mathbf{1}_{t-n}^T\right), \tag{59}$$

$$\mathrm{Cov}(z_1, \boldsymbol{y}_{t-n}) = \frac{\sigma_0^4}{K}\frac{\gamma}{\bar{w}_n}\mathbf{1}_{t-n}, \tag{60}$$

$$\mathrm{Cov}(z_1, x_n) = \mathbf{0},\;\; \mathrm{Cov}(\boldsymbol{x}_n, \boldsymbol{y}_{t-n}) = \mathbf{0}. \tag{61}$$

Thus, $\Pr\{z_1 > 0\}$ is independent of σ_0^2 and can be computed numerically using any multivariate Gaussian integration package, such as Matlab's `mvncdf`. Note that $P_{\mathrm{FA}} = \Pr\{z_1 > 0 \,|\, \rho_M = 0\}$, whereas $P_{\mathrm{D}} = \Pr\{z_1 > 0 \,|\, \rho_M > 0\}$.

2) Test 2 from (43): This test is given by $z_2 \gtrless_{\mathcal{H}_0^M}^{\mathcal{H}_1^M} 0$, with $z_2 \doteq q_M - \gamma\hat{\sigma}_{02}^2$, $\gamma > 1$, and $\hat{\sigma}_{02}^2$ a linear combination of q_0 and $\{q_j,\, j \in \mathcal{U}_0\}$, as in (42). Denote by $\tilde{U}_n$ the event $|\mathcal{U}_0| = n$. Then, similarly to (53),

$$\Pr\{z_2 > 0\} \approx \sum_{t=0}^{M-1} \Pr\{I_t\} \sum_{n=0}^{t} \Pr\{z_2 > 0, \tilde{U}_n \,|\, I_t\} \tag{62}$$

$$= \sum_{t=0}^{M-1} \Pr\{I_t\} \sum_{n=0}^{t} \binom{t}{n} \Pr\{z_2 > 0, \tilde{\boldsymbol{x}}_n < \mathbf{0}, \tilde{\boldsymbol{y}}_{t-n} > \mathbf{0} \,|\, I_t\}, \tag{63}$$

where now

$$\tilde{\boldsymbol{x}}_n \doteq \begin{bmatrix} q_1 \\ \vdots \\ q_n \end{bmatrix} - \hat{\sigma}_{00}^2 \mathbf{1}, \qquad \tilde{\boldsymbol{y}}_{t-n} \doteq \begin{bmatrix} q_{n+1} \\ \vdots \\ q_t \end{bmatrix} - \hat{\sigma}_{00}^2 \mathbf{1}. \tag{64}$$

In this case, $\tilde{\boldsymbol{x}}_n < \mathbf{0}$, $\tilde{\boldsymbol{y}}_{t-n} > \mathbf{0}$ imply that $\mathcal{U}_0 = \{1, \cdots, n\}$, and thus $\hat{\sigma}_{00}^2 = (w_M q_M + \sum_{l=0}^{n} w_l q_l)/(w_M + \sum_{l=0}^{n} w_l)$. The probability in (63) can be written again as the integral of a $(t+1)$-variate Gaussian distribution over the positive orthant:

$$\Pr\{z_2 > 0,\, \tilde{\boldsymbol{x}}_n < \mathbf{0},\, \tilde{\boldsymbol{y}}_{t-n} > \mathbf{0} \,|\, I_t\} = \Pr\{[z_2 \;\; -\tilde{\boldsymbol{x}}_n^T \;\; \tilde{\boldsymbol{y}}_{t-n}^T]^T/\sigma_0^2 > \mathbf{0} \,|\, I_t\}. \tag{65}$$

The mean of this distribution is in this case

$$\boldsymbol{\mu}_2 = \begin{bmatrix} (1-\gamma+\rho_M) & \frac{w_M\rho_M}{\bar{w}_{n+1}}\mathbf{1}_n^T & \frac{-w_M\rho_M}{\bar{w}_{n+1}}\mathbf{1}_{t-n}^T \end{bmatrix}^T, \tag{66}$$

[4]The analysis can be readily modified in order to account for channels with different bandwidths, although the notation becomes somewhat cumbersome.

whereas the covariance matrix can be found from

$$\mathrm{Cov}(z_2, z_2) = \frac{\sigma_0^4}{K}\left(\frac{(1+\rho_M)^2}{w_M} + \frac{\gamma^2}{\bar{w}_n}\right), \quad (67)$$

$$\mathrm{Cov}(\tilde{\boldsymbol{x}}_n, \tilde{\boldsymbol{x}}_n) = \frac{\sigma_0^4}{K}\left[\frac{1}{w_M}\boldsymbol{I} + \left(\frac{w_M((1+\rho_M)^2-1)}{\bar{w}_{n+1}^2} - \frac{1}{\bar{w}_{n+1}}\right)\mathbf{1}_n\mathbf{1}_n^T\right], \quad (68)$$

$$\mathrm{Cov}(\tilde{\boldsymbol{y}}_{t-n}, \tilde{\boldsymbol{y}}_{t-n}) = \frac{\sigma_0^4}{K}\left[\frac{1}{w_M}\boldsymbol{I} + \left(\frac{w_M((1+\rho_M)^2-1)}{\bar{w}_{n+1}^2} + \frac{1}{\bar{w}_{n+1}}\right)\mathbf{1}_{t-n}\mathbf{1}_{t-n}^T\right], \quad (69)$$

$$\mathrm{Cov}(z_2, \tilde{\boldsymbol{x}}_n) = \frac{\sigma_0^4}{K}\left(\frac{\gamma-(1+\rho_M)^2}{\bar{w}_{n+1}} - \frac{\gamma}{\bar{w}_n}\right)\mathbf{1}_n, \quad (70)$$

$$\mathrm{Cov}(\tilde{\boldsymbol{x}}_n, \tilde{\boldsymbol{y}}_{t-n}) = \frac{\sigma_0^4}{K}\frac{w_M((1+\rho_M)^2-1)}{\bar{w}_{n+1}^2}\mathbf{1}_n\mathbf{1}_{t-n}^T, \quad (71)$$

$$\mathrm{Cov}(z_2, \tilde{\boldsymbol{y}}_{t-n}) = \frac{\sigma_0^4}{K}\frac{\gamma-(1+\rho_M)^2}{\bar{w}_{n+1}}\mathbf{1}_{t-n}. \quad (72)$$

Therefore, for Tests 1 and 2 P_{FA} and P_{D} can be found for a given scenario without resorting to Monte Carlo simulation.

VI. Numerical results

We evaluate now the performance of the proposed detectors (QGLRT (40), Test 1 (41) and Test 2 (43)), both theoretically and via Monte Carlo simulations. For the primary system we consider a terrestrial digital TV broadcast network using 8K-mode DVB-T modulation[5]. The channel spacing is 8 MHz with a 7.61 MHz signal bandwidth, which is one of the options considered in the DVB-T standard [29] resulting in $w_1 = \cdots = w_M = 0.95125/M$, $w_0 = 0.04875$.

A. Influence of channel occupancy

In the first experiment we consider a setting with $M = 4$ channels and $K = 2048$ samples. The SNR of the channel to detect is set to -5 dB. The detectors were analyzed for activity factors of $a = 0.1$, 0.5 and 0.9. In the simulations, the SNRs of the active channels (other than that under scrutiny) were generated following a log-normal distribution with mean 0 dB and dB-spread equal to 1 dB.

In order to investigate the issue of threshold selection, we plot in Fig. 2 the missed detection and false alarm probabilities of the three schemes, as a function of the detection threshold. For Tests 1 and 2, the analytical method of Sec. V-B was used, whereas for the QGLRT (40), P_{D} and P_{FA} were obtained empirically. It is seen that, in the region of interest (small probability of missed detection), and for fixed thresholds, the detection performance of the three tests improves as a decreases. This is reasonable, since lower primary activity results in more channels perceived as weak and this can be exploited in order to improve the noise variance estimates. Hence, in order to satisfy $P_{\mathrm{D}} \geq P_{\mathrm{D}}^*$ for a given target P_{D}^* when the activity factor is unknown, the threshold must be set assuming the worst case $a = 1$.

[5]For Monte Carlo simulation, the modulation parameters of the DVB-T signals were: 64-QAM, guard interval 1/4, inner code rate 2/3.

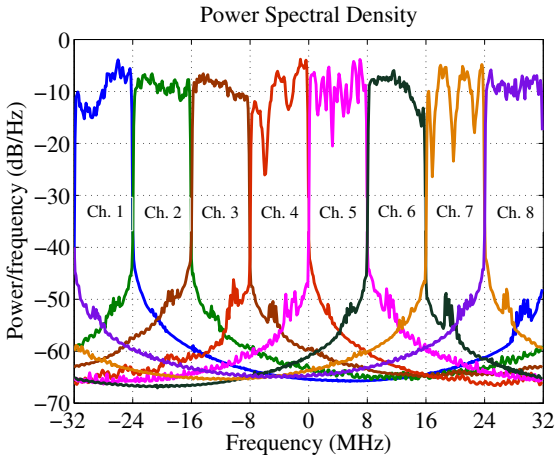


Fig. 4. Power spectral densities of the DVB-T signals affected by multipath propagation.

Once the threshold has been fixed in order to satisfy the detection requirements, the behavior of P_{FA} in terms of a is different for the three schemes. Whereas the false alarm rate of the QGLRT worsens as a decreases, for Test 1 P_{FA} is almost insensitive to variations in a. Interestingly, for Test 2 a region exists for which *both* P_{D} and P_{FA} improve with decreasing a. Thus, by setting the threshold for a given target $P_{\mathrm{D}} \geq P_{\mathrm{D}}^*$ assuming $a = 1$, performance guarantees in terms of P_{FA} (missed opportunities for transmission) can be given for Tests 1 and 2.

Next we consider a setting with $M = 8$ channels, with the remaining parameters kept at the same values as in the previous experiment. Fig. 3 shows the complementary Receiver Operating Characteristics (ROC) curves for the three detectors and different activity factors. As expected, the QGLRT-based detector outperforms the other two suboptimal schemes. Tests 1 and 2 perform similarly for high activity factors, although Test 2 presents an advantage as a decreases. Note that in the extreme case of $a = 1$ there are no idle channels, and thus $\mathcal{S} = \emptyset$ in the context of Sec. IV-B, which in turn implies that the three tests become approximately equivalent for $a \to 1$.

In Fig. 3 a good agreement is observed between analytical and empirical results for Tests 1 and 2, with just a slight mismatch for high activity settings ($a = 0.9$) which can be explained as follows. Recall that in the derivation of the analytical expressions it was assumed that busy channels do not affect the distribution of the statistics for these detectors. This assumption is more likely to be violated as the percentage of busy channels (i.e. the activity factor a) increases.

B. Impact of frequency selectivity

Simulations were carried out in order to gauge the effect of unknown multipath propagation conditions in the performance of the proposed detectors. The multipath channels were generated according to the WINNER Phase II Model [30] with Profile C1 (Suburban). The central frequency is 800 MHz, and it is assumed that each of the signals at the $M = 8$ different

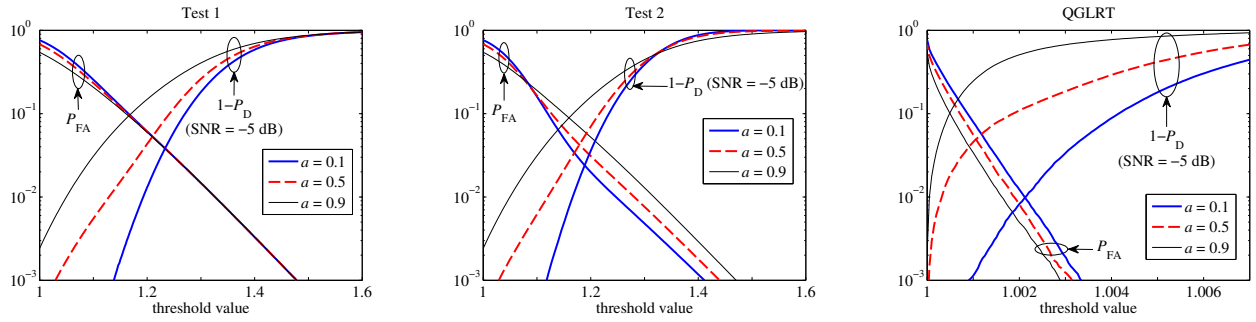


Fig. 2. False alarm and missed detection performance in a setting with $M = 4$ channels. (a) Test 1 (analytical). (b) Test 2 (analytical). (c) QGLRT (empirical).

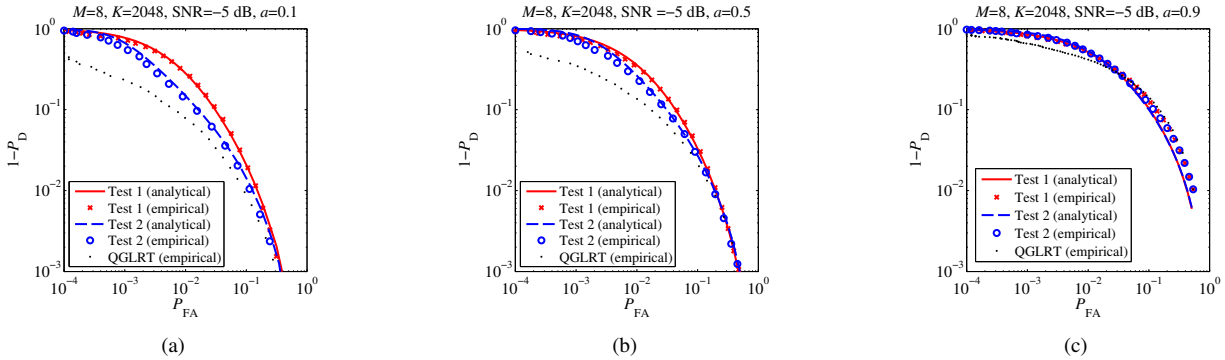


Fig. 3. Complementary ROC curves in a setting with $M = 8$ channels, for different values of the activity factor a.

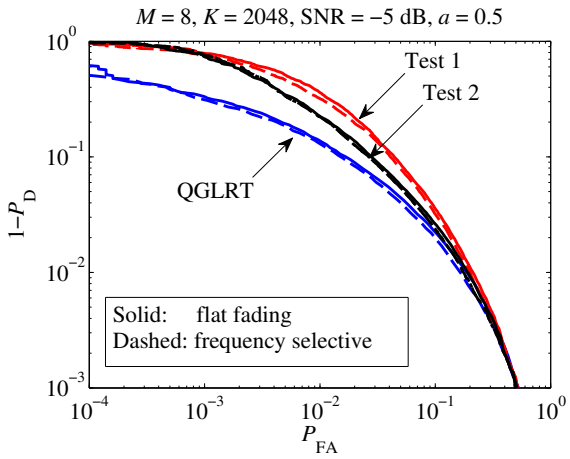


Fig. 5. Detector performance with frequency selective channels.

channels arrives from a different transmitter. The locations of the transmitters and of the spectrum monitor were randomly selected on a square of dimension 15×15 km. Fig. 4 shows the psds of the normalized signals used in this experiment.

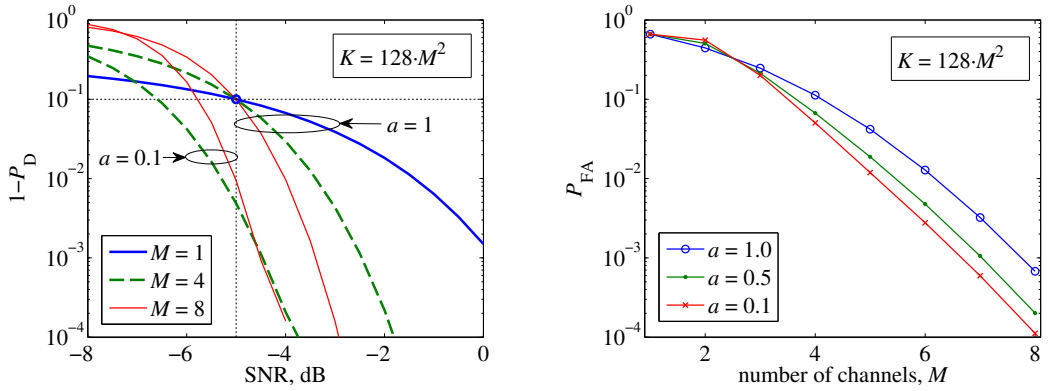
Fig. 5 shows the ROC curves of the three detectors under frequency-flat and frequency-selective channels, for a setting with $K = 2048$, SNR $= -5$ dB and $a = 0.5$. As can be seen, performance remains essentially unaltered under multipath conditions. This can be explained by the structure of the proposed detectors: the linear combinations of different frequency bins effectively averages out the effects of frequency-selective channels. Although in this scenario it is seen that the change in the statistics' distributions due to frequency selectivity even achieves some performance improvement, this effect is so slight that it can be safely neglected.

C. Influence of the number of channels

Consider a setting in which the operating band consists of N channels of B Hz each, which the spectrum monitor must scan in a total time of T s. To this end, the band is subdivided into subbands of $M < N$ channels each, which are sequentially analyzed. The observation time for each of the MB Hz-wide subbands is thus MT/N s. Hence, sampling at the Nyquist rate $f_s = MB$ Hz, the number of samples available for processing each subband of M channels is $K = M^2(BT/N)$. Thus, at the expense of a linear increase of f_s in terms of M, a *quadratic* increase of K is obtained, so that a favorable trade-off between detection performance and ADC cost/resolution can be achieved.

Assuming $BT/N = 128$, Fig. 6(a) shows the analytical probability of missed detection of Test 2 versus SNR for different values of M. For each M, the thresholds are computed in order to achieve $P_D = 0.9$ at a target SNR $= -5$ dB assuming full occupancy (worst case). With this design, having more channels in the subband is seen to improve detection performance for SNRs at and above the target value, for all values of a, thus offering additional interference guarantees to the primary system.

Fig. 6(b) shows the corresponding false alarm rate in terms of M. It is seen that P_{FA} decreases *exponentially* with the number of channels included in the subband. A reduction in



(a) (b)

Fig. 6. Analytical performance of Test 2 as a function of the number of channels M. The sample size is given by $K = 128M^2$.

P_{FA} increases the opportunities of accessing the spectrum and therefore the global throughput of the secondary system.

VII. Conclusions

The wideband approach to spectrum sensing provides a means to trade off detection performance and ADC complexity. In practice, primary networks using FDMA exhibit guard bands between adjacent channels which can be used to estimate the noise power. In addition, the frequency bins of those channels perceived as weak can be used for this task as well. These ideas are exploited by the proposed detectors, which were developed under the GLRT approach. In this way, the noise uncertainty problem that plagues the popular Energy Detector is largely alleviated. Further research will address the impact of noise coloring, a phenomenon to take into account as the processing bandwidth is increased.

Appendix

A. Proof of Theorem 1

For finite K, consider the matrix

$$\boldsymbol{A}(\boldsymbol{\theta}) \doteq [\boldsymbol{L}^H \boldsymbol{\Delta}^{-2}(\boldsymbol{\theta}) \boldsymbol{L}] \cdot [\boldsymbol{L}^\dagger \boldsymbol{\Delta}^2(\boldsymbol{\theta}) (\boldsymbol{L}^\dagger)^H]. \tag{73}$$

We will prove that, under the conditions of the Theorem, $\boldsymbol{A}(\boldsymbol{\theta}) = \boldsymbol{I}$. Then, taking the limit as $K \to \infty$ on both sides of (73), the desired result will be obtained, in view of (23) and (27). To this end, consider the singular-value decomposition $\boldsymbol{L} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{V}^H$, where $\boldsymbol{U}$ is $K \times (M+1)$ with orthogonal columns, $\boldsymbol{D}$ is $(M+1) \times (M+1)$ diagonal with the nonzero singular values, and $\boldsymbol{V}$ is $(M+1) \times (M+1)$ unitary. The pseudoinverse is thus $\boldsymbol{L}^\dagger = \boldsymbol{V}\boldsymbol{D}^{-1}\boldsymbol{U}^H$. Then

$$\boldsymbol{A}(\boldsymbol{\theta}) = \boldsymbol{V}\boldsymbol{D}\boldsymbol{U}^H \boldsymbol{\Delta}^{-2}(\boldsymbol{\theta}) \boldsymbol{U}\boldsymbol{U}^H \boldsymbol{\Delta}^2(\boldsymbol{\theta}) \boldsymbol{U}\boldsymbol{D}^{-1}\boldsymbol{V}^H. \tag{74}$$

As seen from (74), a sufficient condition for $\boldsymbol{A}(\boldsymbol{\theta}) = \boldsymbol{I}$ is that $\boldsymbol{U}\boldsymbol{U}^H$ and $\boldsymbol{\Delta}(\boldsymbol{\theta})$ commute. This we will show now.

Note that the columns of $\boldsymbol{U}$ constitute an orthonormal basis of $\mathcal{R}\{\boldsymbol{L}\}$, the subspace spanned by the columns of $\boldsymbol{L}$. Without loss of generality (since channel indexing is arbitrary), assume that the columns of $\boldsymbol{L}$ are sorted such that if the set of indices of nonzero entries of column j contains that of column i, then $i < j$. Additionally, we assume that the rows of $\boldsymbol{L}$ are arranged such that these sets of indices of nonzero entries contain only contiguous indices (frequency bins). This is also without loss of generality, since one can always apply a permutation to the rows of $\boldsymbol{L}$ to achieve this.

An orthogonal basis for $\mathcal{R}\{\boldsymbol{L}\}$ can also be obtained by applying the Gram-Schmidt orthogonalization procedure to the columns of $\boldsymbol{L}$. It is straightforward to show that this results in a basis $\tilde{\boldsymbol{U}} = [\ \tilde{\boldsymbol{u}}_0 \quad \tilde{\boldsymbol{u}}_1 \quad \cdots \quad \tilde{\boldsymbol{u}}_M\ ]$ such that (i) any nonzero entries of a given vector $\tilde{\boldsymbol{u}}_i$ are constant and in contiguous positions, and (ii) for any $\tilde{\boldsymbol{u}}_i$, $\tilde{\boldsymbol{u}}_j$ with $i \neq j$, the two sets of indices of their nonzero entries are disjoint. These properties imply that $\tilde{\boldsymbol{U}}\tilde{\boldsymbol{U}}^H$ is a block diagonal matrix, with each block on the diagonal having all of its elements equal:

$$\tilde{\boldsymbol{U}}\tilde{\boldsymbol{U}}^H = \begin{bmatrix} \alpha_0 \mathbf{1}_{K_0} \mathbf{1}_{K_0}^T & & \\ & \ddots & \\ & & \alpha_M \mathbf{1}_{K_M} \mathbf{1}_{K_M}^T \end{bmatrix} = \boldsymbol{U}\boldsymbol{U}^H, \tag{75}$$

where α_i are scalars, K_i is the number of nonzero entries in $\tilde{\boldsymbol{u}}_i$, and the last equality in (75) stems from the fact that both $\boldsymbol{U}\boldsymbol{U}^H$ and $\tilde{\boldsymbol{U}}\tilde{\boldsymbol{U}}^H$ are projection matrices onto $\mathcal{R}\{\boldsymbol{L}\}$.

On the other hand, since the diagonal of $\boldsymbol{\Delta}(\boldsymbol{\theta})$ is a linear combination of the columns of $\boldsymbol{L}$, see (12), one has that

$$\boldsymbol{\Delta}(\boldsymbol{\theta}) = \begin{bmatrix} \beta_0 \boldsymbol{I}_{K_0} & & \\ & \ddots & \\ & & \beta_M \boldsymbol{I}_{K_M} \end{bmatrix}, \tag{76}$$

for some scalars β_i. Given the structure of $\boldsymbol{U}\boldsymbol{U}^H$ and $\boldsymbol{\Delta}(\theta)$, it is readily seen that they do commute. ■

B. Proof of Theorem 2

We shall prove (38), as the proof for (39) is analogous. The cost $f(\hat{\boldsymbol{\theta}}) = \|\boldsymbol{L}\hat{\boldsymbol{\theta}} - \boldsymbol{p}\|_2^2$ is convex, and its gradient is given

by

$$\frac{1}{K}\nabla f(\hat{\boldsymbol{\theta}}) = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & w_1^{-1} & 0 & \cdots & 0 \\ 1 & 0 & w_2^{-1} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & \cdots & w_M^{-1} \end{bmatrix} \hat{\boldsymbol{\theta}} - \begin{bmatrix} w_0 q_0 + \cdots + w_M q_M \\ q_1 \\ q_2 \\ \vdots \\ q_M \end{bmatrix}. \quad (77)$$

In addition, we have the linear inequality constraints $g_j(\hat{\boldsymbol{\theta}}) = -\hat{\sigma}_j^2 \leq 0$, $0 \leq j \leq M$, whose gradient is $\nabla g_j(\hat{\boldsymbol{\theta}}) = -\boldsymbol{e}_j$, where $\boldsymbol{e}_j$ is the j-th unit vector. A sufficient condition for $\hat{\boldsymbol{\theta}}_1 = [\hat{\sigma}_{01}^2 \cdots \hat{\sigma}_{M1}^2]^T$ to be the global optimum is that there exist scalars $\mu_j \geq 0$, $0 \leq j \leq M$, such that

$$\nabla f(\hat{\boldsymbol{\theta}}_1) + \sum_{j=0}^{M} \mu_j \nabla g_j(\hat{\boldsymbol{\theta}}_1) = 0, \quad (78)$$

$$\mu_j g_j(\hat{\boldsymbol{\theta}}_1) = 0, \quad 0 \leq j \leq M, \quad (79)$$

which for this case amounts to saying that $[\nabla f(\hat{\boldsymbol{\theta}}_1)]_i = 0$ if $\hat{\sigma}_{i1}^2 > 0$ and $[\nabla f(\hat{\boldsymbol{\theta}}_1)]_i \geq 0$ if $\hat{\sigma}_{i1}^2 = 0$. Now we show that the vector $\hat{\boldsymbol{\theta}}_1$ given by (38) satisfies these conditions. Note that $\hat{\sigma}_{01}^2 > 0$, and that $\mathcal{U}_1 = \{j : \hat{\sigma}_{j1}^2 = 0\}$. In view of (77),

$$\frac{1}{K}[\nabla f(\hat{\boldsymbol{\theta}}_1)]_0 = \hat{\sigma}_{01}^2 + \sum_{i=1}^{M} \hat{\sigma}_{i1}^2 - w_0 q_0 - \sum_{i=1}^{M} w_i q_i \quad (80)$$

$$= -w_0(q_0 - \hat{\sigma}_{01}^2) + \sum_{i=1}^{M} [\hat{\sigma}_{i1}^2 - w_i(q_i - \hat{\sigma}_{01}^2)] \quad (81)$$

$$= -w_0(q_0 - \hat{\sigma}_{01}^2) - \sum_{i \in \mathcal{U}_1} w_i(q_i - \hat{\sigma}_{01}^2) \quad (82)$$

$$= -(w_0 q_0 + \sum_{i \in \mathcal{U}_1}^{M} w_i q_i) + (w_0 + \sum_{i \in \mathcal{U}_1}^{M} w_i)\hat{\sigma}_{01}^2 = 0, \quad (83)$$

where the second line follows from $w_0 + \cdots + w_M = 1$; the third, from the definitions of $\mathcal{U}_1$ and $\hat{\sigma}_{j1}^2$, and the last step, from the definition of $\hat{\sigma}_{01}^2$. On the other hand, for $1 \leq j \leq M$, using again the definitions of $\mathcal{U}_1$ and $\hat{\sigma}_{j1}^2$,

$$\frac{1}{K}[\nabla f(\hat{\boldsymbol{\theta}}_1)]_j = \hat{\sigma}_{01}^2 + w_j^{-1}\hat{\sigma}_{j1}^2 - q_j \quad (84)$$

$$= \begin{cases} \hat{\sigma}_{01}^2 - q_j \geq 0, & j \in \mathcal{U}_1, \\ 0, & j \notin \mathcal{U}_1, \end{cases} \quad (85)$$

as was to be shown. ■

C. Proof of Proposition 1

If the constrained LS estimate under $\mathcal{H}_1^m$ results in $\hat{\sigma}_{m1}^2 = 0$ (i.e. $m \in \mathcal{U}_1$), then it is clear that imposing $\hat{\sigma}_{m0}^2 = 0$ and then minimizing the LS cost under the same constraints for the remaining variables will yield the same result. But this is exactly the constrained LS estimate under $\mathcal{H}_0^m$. ■

D. Proof of Proposition 2

Let $\mathcal{U}_1 = \{l(1), l(2), \cdots, l(s)\}$ such that $q_{l(1)} \leq q_{l(2)} \leq \cdots \leq q_{l(s)}$. The proof is by induction, and is based on the constructive algorithm given in Table 2. Note that:

- $q_l \leq \hat{\sigma}_{01}^2$ for all $l \in \mathcal{U}_1$ (by definition of $\mathcal{U}_1$);
- $\hat{\sigma}_{01}^2 \leq q_0$ (since $\hat{\sigma}_{01}^2$ is a convex combination of q_0 and $\{q_l,\, l \in \mathcal{U}_1\}$);
- $\hat{\sigma}_{01}^2 \leq q_m$ (since $m \notin \mathcal{U}_1$).

The last two facts imply that $\hat{\sigma}_{01}^2 \leq q_{0m} \doteq (w_0 q_0 + w_m q_m)/(w_0 + w_m)$.

Now consider $q_{l(1)}$. In the process of constructing $\mathcal{U}_0$ given in Table 2, the first iteration results in $\hat{\sigma}_{00}^2 = q_{0m}$. The unconstrained estimate with respect to the remaining variables is not feasible, since $q_l < q_{0m}$ for $l \in \mathcal{U}_1$. Therefore, index $l(1)$ is picked so that $l(1) \in \mathcal{U}_0$.

Suppose now that $l(1),\ldots, l(i) \in \mathcal{U}_0$ for $i < s$. This means that the $(i+1)$-th iteration of the procedure from Table 2 results in

$$\hat{\sigma}_{00}^2 = \frac{w_0 q_0 + w_m q_m + \sum_{t=1}^{i} w_{l(t)} q_{l(t)}}{w_0 + w_m + \sum_{t=1}^{i} w_{l(t)}}. \quad (86)$$

Note that, since $l(i+1) \in \mathcal{U}_1$, it holds that

$$q_{l(i+1)} < \frac{w_0 q_0 + \sum_{t=1}^{i} w_{l(t)} q_{l(t)}}{w_0 + \sum_{t=1}^{i} w_{l(t)}} \leq \frac{w_0 q_0 + \sum_{t=1}^{s} w_{l(t)} q_{l(t)}}{w_0 + \sum_{t=1}^{s} w_{l(t)}} < q_m. \quad (87)$$

But (87) implies that $q_{l(i+1)}$ is smaller than the right-hand side of (86). Hence, index $l(i+1)$ is picked so that $l(i+1) \in \mathcal{U}_0$. By induction, it follows that $\mathcal{U}_1 \subset \mathcal{U}_0$. ■

REFERENCES

[1] J. Mitola and G. Q. Maguire Jr., "Cognitive radio: Making software radios more personal," *IEEE Pers. Commun.*, vol. 6, pp. 13–18, Aug 1999.

[2] J. M. Peha, "Sharing spectrum through spectrum policy reform and cognitive radio," *Proc. IEEE*, vol. 97, no. 4, pp. 708–719, Apr 2009.

[3] I. Akyildiz, W.-Y. Lee, M. Vuran, and S. Mohanty, "A survey on spectrum management in cognitive radio networks," *IEEE Commun. Mag.*, vol. 46, no. 4, pp. 40–48, April 2008.

[4] G. Ganesan and Y. Li, "Cooperative spectrum sensing in cognitive radio, part I: Two user networks," *IEEE Trans. Wireless Commun.*, vol. 6, no. 6, pp. 2204–2213, Jun. 2007.

[5] ——, "Cooperative spectrum sensing in cognitive radio, part II: Multiuser networks," *IEEE Trans. Wireless Commun.*, vol. 6, no. 6, pp. 2214–2222, Jun. 2007.

[6] D. Cabric, "Addressing the feasibility of cognitive radios," *IEEE Signal Processing Mag.*, vol. 25, no. 6, pp. 85–93, Nov. 2008.

[7] R. Tandra and A. Sahai, "SNR walls for signal detection," *IEEE J. Sel. Top. Signal Process.*, vol. 2, pp. 4–17, Feb 2008.

[8] FCC, "FCC 08-260, second report and order and memorandum opinion and order, in the matter of unlicensed operation in the TV broadcast bands and additional spectrum for unlicensed devices below 900 MHz and in the 3 GHz band," ET Docket 08-260, Nov. 2008, http://hraunfoss.fcc.gov/edocs_public/attachmatch/FCC-08-260A1.pdf.

[9] ——, "FCC 10-174, second memorandum opinion and order, in the matter of unlicensed operation in the TV broadcast bands and additional spectrum for unlicensed devices below 900 MHz and in the 3 GHz band," ET Docket 10-174, Sep. 2010, http://hraunfoss.fcc.gov/edocs_public/attachmatch/FCC-10-174A1.pdf.

[10] Ofcom, "Statement on cognitive access to interleaved spectrum," Jul. 2009, http://stakeholders.ofcom.org.uk/consultations/cognitive/statement/.

[11] ECC, "Technical and operational requirements for the possible operation of Cognitive Radio systems in the 'white spaces' of frequency band 470-790 MHz," ECC Group SE43 Draft Report, Sep. 2010, http://www.cept.org/0B322E6B-375D-4B8F-868B-3F9E5153CF72.W5Doc.

[12] B. Fette, Ed., *Cognitive Radio Technology*. Burlington, MA: Newnes, 2006, ch. 2: Communications policy and spectrum management.

[13] L. E. Doyle, *Essentials of Cognitive Radio*. Cambridge, UK.: Cambridge University Press, 2009.

[14] H. Zamat and B. Natarajan, "Use of dedicated broadband sensing receiver in cognitive radio," *Proc. 9th IEEE Conf. Commun. Workshops*, pp. 508–512, May 2008.

[15] C.-H. Hwang, G.-L. Lai, and S.-C. Chen, "Spectrum sensing in wideband OFDM cognitive radios," *IEEE Trans. Signal Process.*, vol. 58, no. 2, pp. 709–719, Feb 2010.

[16] A. Taherpour, S. Gazor, and M. Nasiri-Kenari, "Wideband spectrum sensing in unknown white Gaussian noise," *IET Commun.*, vol. 2, no. 6, pp. 763–771, July 2008.

[17] A. Taherpour, M. Nasiri-Kenari, and S. Gazor, "Invariant wideband spectrum sensing under unknown variances," *IEEE Trans. Wireless Commun.*, vol. 8, no. 5, pp. 2182–2186, May 2009.

[18] Z. Quan, S. Cui, A. H. Sayed, and H. V. Poor, "Optimal multiband joint detection for spectrum sensing in cognitive radio networks," *IEEE Trans. Signal Process.*, vol. 57, no. 3, pp. 1128–1140, March 2009.

[19] S. M. Kay, *Fundamentals of statistical signal processing: detection theory*. Englewood Cliffs, NJ: Prentice-Hall, 1998.

[20] J. Tellado, *Multicarrier modulation with low PAR: Applications to DSL and wireless*. Norwell, MA: Kluwer Academic, 2000.

[21] Z. Quan, W. Zhang, S. Shellhammer, and A. Sayed, "Optimal spectral feature detection for spectrum sensing at very low SNR," *IEEE Trans. Commun.*, vol. 59, no. 1, pp. 201 –212, Jan. 2011.

[22] J. Burg, D. Luenberger, and D. Wenger, "Estimation of structured covariance matrices," *Proc. IEEE*, vol. 70, no. 9, pp. 963–974, Sep. 1982.

[23] R. M. Gray, *Toeplitz and circulant matrices: a review*. New York: Hanover/Now, 2006.

[24] W. Zhang, H. V. Poor, and Z. Quan, "Frequency-domain correlation: An asymptotically optimum approximation of quadratic likelihood ratio detectors," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 969–979, Mar. 2010.

[25] R. López-Valcarce and G. Vazquez-Vilar, "Wideband spectrum sensing in cognitive radio: joint estimation of noise variance and multiple signal levels," *Proc. 10th IEEE Workshop on Signal Proc. Adv. in Wireless Commun.*, June 2009.

[26] P. Stoica and R. L. Moses, *Spectral analysis of signals*. Englewood Cliffs, NJ: Prentice-Hall, 2005.

[27] S. M. Kay, *Fundamentals of statistical signal processing: estimation theory*. Englewood Cliffs, NJ: Prentice-Hall, 1993.

[28] S. Chaudhari, V. Koivunen, and H. Poor, "Autocorrelation-based decentralized sequential detection of OFDM signals in cognitive radios," *IEEE Trans. Signal Process.*, vol. 57, no. 7, pp. 2690 –2700, Jul. 2009.

[29] EN 300 744 V1.5.1, *Digital Video Broadcasting (DVB); Framing structure, channel coding and modulation for terrestrial television*. ETSI, Nov. 2004.

[30] L. Hentilä, P. Kyösti, M. Käske, M. Narandzic, and M. Alatossava, "Matlab implementation of the WINNER Phase II channel model v1.1," Online: https://www.ist-winner.org/phase_2_model.html, Dec. 2007.