

State-of-the-art and recent advances Spectrum Sensing for Cognitive Radio State-of-the-art and recent advances

Erik Axell, Geert Leus, Erik G. Larsson and H. Vincent Poor

Linköping University Post Print

N.B.: When citing this work, cite the original article.

©2012 IEEE. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE.

Erik Axell, Geert Leus, Erik G. Larsson and H. Vincent Poor, State-of-the-art and recent advances Spectrum Sensing for Cognitive Radio State-of-the-art and recent advances, 2012, IEEE signal processing magazine (Print), (29), 3, 101-116.

<http://dx.doi.org/10.1109/MSP.2012.2183771>

Postprint available at: Linköping University Electronic Press

<http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-70151>

Spectrum Sensing for Cognitive Radio: State-of-the-Art and Recent Advances

Erik Axell, Geert Leus, Erik G. Larsson and H. Vincent Poor

I. INTRODUCTION TO SPECTRUM SENSING AND PROBLEM FORMULATION

The ever increasing demand for higher data rates in wireless communications in the face of limited or under-utilized spectral resources has motivated the introduction of cognitive radio. Traditionally, licensed spectrum is allocated over relatively long time periods, and is intended to be used only by licensees. Various measurements of spectrum utilization have shown substantial unused resources in frequency, time and space [1], [2]. The concept behind cognitive radio is to exploit these under-utilized spectral resources by reusing unused spectrum in an opportunistic manner [3], [4]. The phrase “cognitive radio” is usually attributed to Mitola [4], but the idea of using learning and sensing machines to probe the radio spectrum was envisioned several decades earlier (cf. [5]).

Cognitive radio systems typically involve primary users of the spectrum, who are incumbent licensees, and secondary users who seek to opportunistically use the spectrum when the primary users are idle¹. The introduction of cognitive radios inevitably creates increased interference and thus can degrade the quality-of-service of the primary system. The impact on the primary system, for example in terms of increased

Geert Leus is supported in part by the NWO-STW under the VICI program (project 10382).

The research leading to these results has received funding from the European Community’s Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 216076. This work was also supported in part by the Swedish Research Council (VR), the Swedish Foundation for Strategic Research (SSF) and the ELLIIT. E. Larsson is a Royal Swedish Academy of Sciences (KVA) Research Fellow supported by a grant from the Knut and Alice Wallenberg Foundation.

This paper was prepared in part under the support of the Qatar National Research Fund under Grant NPRP 08-522-2-211

¹Note that here we are describing and addressing so-called “interweave” cognitive radio systems. Other methods of spectrum sharing have also been envisioned. These include overlay and underlay systems, which make use of techniques such as spread-spectrum or dirty-paper coding, to avoid excessive interference. Such systems are not addressed here except to the extent that they may also rely on spectrum sensing.

interference, must be kept at a minimal level. Therefore, cognitive radios must sense the spectrum to detect whether it is available or not, and must be able to detect very weak primary user signals [6], [7]. Thus spectrum sensing is one of the most essential components of cognitive radio.

The problem of spectrum sensing is to decide whether a particular slice of the spectrum is “available” or not. That is, in its simplest form we want to discriminate between the two hypotheses

$$\begin{aligned} H_0 : \mathbf{y}[n] &= \mathbf{w}[n], \quad n = 1, \dots, N \\ H_1 : \mathbf{y}[n] &= \mathbf{x}[n] + \mathbf{w}[n], \quad n = 1, \dots, N, \end{aligned} \tag{1}$$

where $\mathbf{x}[n]$ represents a primary user’s signal, $\mathbf{w}[n]$ is noise and n represents time. The received signal $\mathbf{y}[n]$ is vectorial, of length L . Each element of the vector $\mathbf{y}[n]$ could represent, for example, the received signal at a different antenna. Note that (1) is a classical detection problem, which is treated in detection theory textbooks. Detection of very weak signals $\mathbf{x}[n]$, in the setting of (1) is also a traditional topic, dealt with in depth in [8, Ch. II-III], for example. The novel aspect of the spectrum sensing when related to the long-established detection theory literature is that the signal $\mathbf{x}[n]$ has a specific structure that stems from the use of modern modulation and coding techniques in contemporary wireless systems. Clearly, since such a structure may not be trivial to represent, this has resulted in substantial research efforts. At the same time, this structure offers the opportunity to design very efficient spectrum sensing algorithms.

In the sequel, we will use bold-face lowercase letters to denote vectors and bold-face capital letters to denote matrices. A discrete-time index is denoted with square brackets and the m th user is denoted with a subscript. That is, $\mathbf{y}_m[n]$ is the vectorial observation for user m at time n . When considering a single user, we will omit the subscript for simplicity. Moreover, if the sequence is scalar, we use the convention $y[n]$ for the time sequence. The l th scalar element of a vector is denoted by $y_l[n]$, not to be confused with the vectorial observation $\mathbf{y}_m[n]$ for user m .

For simplicity of notation, let the vector $\mathbf{y} \triangleq [\mathbf{y}[1]^T, \mathbf{y}[2]^T, \dots, \mathbf{y}[N]^T]^T$ of length LN contain all observations stacked in one vector. In the same way, denote the total stacked signal by \mathbf{x} and the noise by \mathbf{w} . The hypothesis test (1) can then be rewritten as

$$\begin{aligned} H_0 : \mathbf{y} &= \mathbf{w}, \\ H_1 : \mathbf{y} &= \mathbf{x} + \mathbf{w}. \end{aligned} \tag{2}$$

A standard assumption in the literature, which we also make throughout this paper, is that the additive noise \mathbf{w} is zero-mean, white, and circularly symmetric complex Gaussian. We write this as $\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$, where σ^2 is the noise variance.

II. FUNDAMENTALS OF SIGNAL DETECTION

In signal detection, the task of interest is to decide whether the observation \mathbf{y} was generated under H_0 or H_1 . Typically, this is accomplished by first forming a test statistic $\Lambda(\mathbf{y})$ from the received data \mathbf{y} , and then comparing $\Lambda(\mathbf{y})$ with a predetermined threshold η :

$$\Lambda(\mathbf{y}) \underset{H_0}{\overset{H_1}{\gtrless}} \eta. \quad (3)$$

The performance of a detector is quantified in terms of its *receiver operating characteristics* (ROC), which gives the probability of detection $P_D = \Pr(\Lambda(\mathbf{y}) > \eta | H_1)$ as a function of the probability of false alarm $P_{FA} = \Pr(\Lambda(\mathbf{y}) > \eta | H_0)$. By varying the threshold η , the operating point of a detector can be chosen anywhere along its ROC curve.

Clearly, the fundamental problem of detector design is to choose the test statistic $\Lambda(\mathbf{y})$, and to set the decision threshold η in order to achieve good detection performance. These matters are treated in detail in many books on detection theory (e.g. [8]). Detection algorithms are either designed in the framework of classical statistics, or in the framework of Bayesian statistics. In the classical (also known as deterministic) framework, either H_0 or H_1 is deterministically true, and the objective is to choose $\Lambda(\mathbf{y})$ and η so as to maximize P_D subject to a constraint on P_{FA} : $P_{FA} \leq \alpha$. In the Bayesian framework, by contrast, it is assumed that the source selects the true hypothesis at random, according to some *a priori* probabilities $\Pr(H_0)$ and $\Pr(H_1)$. The objective in this framework is to minimize the so-called *Bayesian cost*. Interestingly, although the difference in philosophy between these two approaches is substantial, both result in a test of the form (3) where the test statistic is the likelihood-ratio [8][Ch. II]

$$\Lambda(\mathbf{y}) = \frac{p(\mathbf{y}|H_1)}{p(\mathbf{y}|H_0)}. \quad (4)$$

A. Unknown Parameters

To compute the likelihood ratio $\Lambda(\mathbf{y})$ in (4), the probability distribution of the observation \mathbf{y} must be perfectly known under both hypotheses. This means that one must know all parameters, such as noise

variance, signal variance and channel coefficients. If the signal to be detected, \mathbf{x} , is perfectly known, then², $\mathbf{y} \sim \mathcal{N}(\mathbf{x}, \sigma^2 \mathbf{I})$ under H_1 , and it is easy to show that the optimal test statistic is the output of a matched filter [8][Sec. III.B]:

$$\operatorname{Re}(\mathbf{x}^H \mathbf{y}) \underset{H_0}{\overset{H_1}{\gtrless}} \eta.$$

In practice, the signal and noise parameters are not known. In the following, we will discuss two standard techniques that are used to deal with unknown parameters in hypothesis testing problems.

In the Bayesian framework, the optimal strategy is to marginalize the likelihood function to eliminate the unknown parameters. More precisely, if the vector $\boldsymbol{\theta}$ contains the unknown parameters, then one computes

$$p(\mathbf{y}|H_i) = \int p(\mathbf{y}|H_i, \boldsymbol{\theta})p(\boldsymbol{\theta}|H_i)d\boldsymbol{\theta},$$

where $p(\mathbf{y}|H_i, \boldsymbol{\theta})$ denotes the conditional PDF of \mathbf{y} under H_i and conditioned on $\boldsymbol{\theta}$, and $p(\boldsymbol{\theta}|H_i)$ denotes the *a priori* probability density of the parameter vector given hypothesis H_i . In practice, the actual *a priori* parameter density $p(\boldsymbol{\theta}|H_i)$ often is not perfectly known, but rather is chosen to provide a meaningful result. How to make such a choice, is far from clear in many cases. One alternative is to choose a non-informative distribution in order to model a lack of *a priori* knowledge of the parameters. One example of a non-informative prior is the gamma distribution, which was used in [9] to model an unknown noise power. Another option is to choose the prior distribution via the so-called *maximum entropy principle*. According to this principle, the prior distribution of the unknown parameters that maximizes the entropy given some statistical constraints (e.g. limited expected power or second-order moment) should be chosen. The maximum entropy principle was used in the context of spectrum sensing for cognitive radio in [10].

In the classical hypothesis testing framework, the unknown parameters must be estimated somehow. A standard technique is to use maximum-likelihood (ML) estimates of the unknown parameters, which gives rise to the well-known generalized likelihood-ratio test (GLRT):

$$\frac{\max_{\boldsymbol{\theta}} p(\mathbf{y}|H_1, \boldsymbol{\theta})}{\max_{\boldsymbol{\theta}} p(\mathbf{y}|H_0, \boldsymbol{\theta})} \underset{H_0}{\overset{H_1}{\gtrless}} \eta.$$

This is a technique that usually works quite well, although it does not necessarily guarantee optimality.

Other estimates than the ML estimate may also be used.

²Recall that we assume circularly symmetric Gaussian noise throughout.

B. Constant False-Alarm Rate (CFAR) Detectors

A detector is said to have the property of constant-false alarm rate (CFAR), if its false alarm probability is independent of parameters such as noise or signal powers. In particular, the CFAR property means that the decision threshold can be set to achieve a pre-specified P_{FA} without knowing the noise power. The CFAR property is normally revealed by the equations that define the test (3): if the test statistic $\Lambda(\mathbf{y})$ and the optimal threshold are unaffected by a scaling of the problem (such as multiplying the received data by a constant), then the detector is CFAR. CFAR is a very desired property in many applications, especially when one has to deal with noise of unknown power, as we will see later.

C. Energy detection

As an example of a very basic detection technique, we present the well known energy detector, also known as the radiometer [11]. The energy detector measures the received energy during a finite time interval, and compares it to a predetermined threshold. It should be noted that the energy detector works well also for other cases than the one we will present, although it might not be optimal.

To derive this detector, assume that the signal to be detected does not have any known structure that could be exploited, and model it via a zero-mean circularly symmetric complex Gaussian $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \gamma^2 \mathbf{I})$. Then, $\mathbf{y}|H_0 \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$ and $\mathbf{y}|H_1 \sim \mathcal{N}(\mathbf{0}, (\sigma^2 + \gamma^2) \mathbf{I})$. After removing irrelevant constants, the optimal (Neyman-Pearson) test can be written as

$$\Lambda(\mathbf{y}) = \frac{\|\mathbf{y}\|^2}{\sigma^2} = \frac{\sum_{i=1}^{LN} |y_i|^2}{\sigma^2} \underset{H_0}{\overset{H_1}{\gtrless}} \eta. \quad (5)$$

The operational meaning of (5) is to compare the energy of the received signal against a threshold and this is why (5) is called the energy detector. Its performance is well known, cf. [8][Sec. III.C], and is given by

$$P_D = \Pr(\Lambda(\mathbf{y}) > \eta | H_1) = 1 - F_{\chi_{2NL}^2} \left(\frac{2\eta}{\sigma^2 + \gamma^2} \right) = 1 - F_{\chi_{2NL}^2} \left(\frac{F_{\chi_{2NL}^2}^{-1}(1 - P_{\text{FA}})}{1 + \frac{\gamma^2}{\sigma^2}} \right).$$

Clearly, P_D is a function of P_{FA} , NL and the SNR $\triangleq \gamma^2/\sigma^2$. Note that for a fixed P_{FA} , $P_D \rightarrow 1$ as $NL \rightarrow \infty$ at any SNR. That is, ideally any pair (P_D, P_{FA}) can be achieved if sensing can be done for an arbitrarily long time. This is typically not the case in practice, as we will see in the following section.

It has been argued that for several models, and if the probability density functions under both hypotheses are perfectly known, energy detection performs close to the optimal detector [7], [12]. For example, it

was shown in [7] that the performance of the energy detector is asymptotically equivalent, at low SNR, to that of the optimal detector when the signal is modulated with a zero-mean finite signal constellation, assuming that the symbols are independent of each other and that all probability distributions are perfectly known. A similar result was shown numerically in [12] for the detection of an orthogonal frequency-division multiplexing (OFDM) signal. These results hold if all probability density functions, including that of the noise, are perfectly known. By contrast, if for example the noise variance is unknown, the energy detector cannot be used because knowledge of σ^2 is needed to set the threshold. If an incorrect (“estimated”) value of σ^2 is used in (5) then the resulting detector may perform rather poorly. We discuss this matter in more depth in the following section.

D. Fundamental limits for sensing: SNR wall

Cognitive radios must be able to detect very weak primary user signals [6], [7]. This is difficult, because there are fundamental limits on detection at low SNR. Specifically, due to uncertainties in the model assumptions, accurate detection is impossible below a certain SNR level, known as the *SNR wall* [13], [14]. The reason is that to compute the likelihood ratio $\Lambda(\mathbf{y})$, the probability distribution of the observation \mathbf{y} must be perfectly known under both hypotheses. In any case, the signal and noise in (2) must be modeled with some known distributions. Of course, a model is always a simplification of reality, and the true probability distributions are never perfectly known. Even if the model would be perfectly consistent with reality, there will be some parameters that are unknown such as the noise power, the signal power and the channel coefficients, as noted above.

To exemplify the SNR wall phenomenon, consider the energy detector. To set its decision threshold, the noise variance σ^2 must be known. If the knowledge of the noise variance is imperfect, the threshold cannot be correctly set. Setting the threshold based on an incorrect noise variance will not result in the desired value of false-alarm probability. In fact, the performance of the energy detector quickly deteriorates if the noise variance is imperfectly known [7], [13]. Let σ^2 denote the imperfect estimate of the noise variance, and let σ_t^2 be the true noise variance. Assume that the estimated noise variance is known only to lie in a given interval, such that $\frac{1}{\rho}\sigma_t^2 \leq \sigma^2 \leq \rho\sigma_t^2$ for some $\rho > 1$. To guarantee that the probability of false alarm is always below a required level, the threshold must be set to fulfill the requirement in the worst

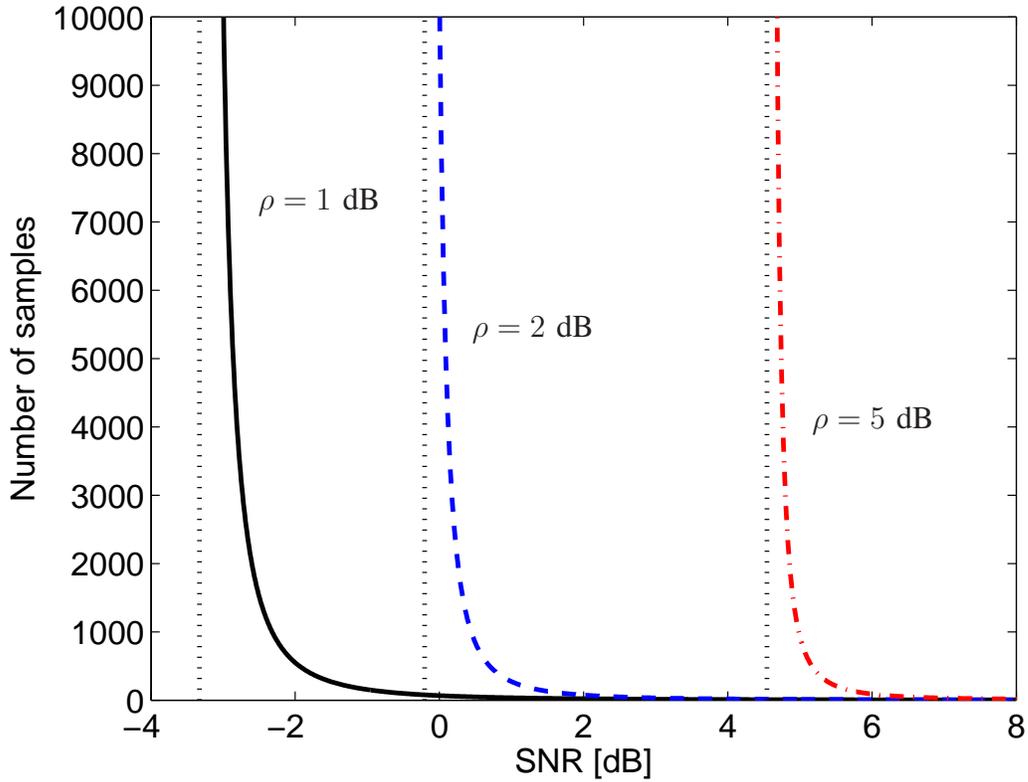


Fig. 1. The number of samples required to meet $P_{\text{FA}} = 0.05$ and $P_{\text{D}} = 0.9$ using energy detection under noise uncertainty.

case. That is, we need to make sure that

$$\max_{\sigma^2 \in [\frac{1}{\rho}\sigma_t^2, \rho\sigma_t^2]} P_{\text{FA}}$$

is below the required level. The worst case occurs when the noise power is at the upper end of the interval, that is when $\sigma^2 = \rho\sigma_t^2$. It was shown in [14] that under this model, the number of samples LN that are required to meet a P_{D} requirement, tends to infinity as the $\text{SNR} = \gamma^2/\sigma_t^2 \rightarrow \frac{\rho^2-1}{\rho}$. That is, even with an infinite measurement duration, it would be impossible to meet the P_{D} requirement when the SNR is below the SNR wall $\frac{\rho^2-1}{\rho}$. This effect occurs only because of the uncertainty in the noise level. The effect of the SNR wall for energy detection is shown in Figure 1. The figure shows the number of samples that are needed to meet the requirements $P_{\text{FA}} = 0.05$ and $P_{\text{D}} = 0.9$ for different levels of the noise uncertainty.

It was shown in [14] that errors in the noise power assumption introduce SNR walls to any moment-based detector, not only to the energy detector. This result was further extended in [14] to any model

uncertainties, such as color and stationarity of the noise, simplified fading models, ideality of filters and quantization errors introduced by finite-precision analog-to-digital (A/D) converters. It is possible to mitigate the problem of SNR walls by taking the imperfections into account, in the sense that the SNR wall can be moved to a lower SNR level. For example, it was shown in [14] that noise calibration can improve the detector robustness. Exploiting known features of the signal to be detected can also improve the detector performance and robustness. Known features can be exploited to deal with unknown parameters using marginalization or estimation as discussed before. It is also known that fast fading effects can somewhat alleviate the requirement of accurately knowing the noise variance in some cases [15]. Note also that a CFAR detector is not exposed to the SNR wall phenomenon, since the decision threshold is set independently of any potentially unknown signal and noise power parameters.

Other recent work has shown that similar limits arise based on other parameters in cooperative spectrum sensing techniques [16].

III. FEATURE DETECTION

Information theory teaches us that communication signals with maximal information content (entropy) are statistically white and Gaussian and hence, we would expect signals used in communication systems to be nearly white Gaussian. If this were the case, then no spectrum sensing algorithm could do better than the energy detector. However, signals used in practical communication systems always contain distinctive features that can be exploited for detection and that enable us to achieve a detection performance that substantially surpasses that of the energy detector. Perhaps even more importantly, known signal features can be exploited to estimate unknown parameters such as the noise power. Therefore, making use of known signal features effectively can circumvent the problem of SNR walls discussed in the previous section. The specific properties that originate from modern modulation and coding techniques have aided in the design of efficient spectrum sensing algorithms.

The term feature detection is commonly used in the context of spectrum sensing and usually refers to exploitation of known statistical properties of the signal. The signal features referred to may be manifested both in time and space. Features of the transmitted signal are the result of redundancy added by coding, and of the modulation and burst formatting schemes used at the transmitter. For example, OFDM modulation adds a cyclic prefix which manifests itself through a linear relationship between the

transmitted samples. Also, most communication systems multiplex known pilots into the transmitted data stream or superimpose pilots on top of the transmitted signals, and doing so results in very distinctive signal features. A further example is given by space-time coded signals, in which the space-time code correlates the transmitted signals. The received signals may also have specific features that occur due to characteristics of the propagation channel. For example, in a MIMO (multiantenna) system, if the receiver array has more antennas than the transmitter array, then samples taken by the receiver array at any given point in time must necessarily be correlated.

In this section, we will review a number of state-of-the-art detectors that exploit signal features and which are suitable for spectrum sensing applications. Most of the presented methods are very recent advances in spectrum sensing, and there is still much ongoing research in these areas.

A. Detectors Based on Second-order Statistics

A very popular and useful approach to feature detection is to estimate the second-order statistics of the received signals and make decisions based on these estimates. Clearly, in this way we may distinguish a perfectly white signal from a colored one. This basic observation is important, because typically, the redundancy added to transmitted signals in a communication system results in its samples becoming correlated. The correlation structure incurred this way does not necessarily have to be stationary; in fact, typically it is not as we shall see. Since $\text{cov}(\mathbf{Ax}) = \mathbf{A}\text{cov}(\mathbf{x})\mathbf{A}^H$ for any \mathbf{A} and \mathbf{x} , the correlation structure incurred by the addition of redundancy at the transmitter is usually straightforward to analyze if the transmit processing consists of a linear operation. Moreover, we know that the distribution of a Gaussian signal is fully determined by its first and second-order moments. Therefore, provided that the communication signals in question are sufficiently near to Gaussian and that enough samples are collected, we expect that estimated first and second-order moments are sufficient statistics to within practical accuracy. Since communication signals are almost always of zero-mean (in order to minimize the power spent at the transmitter), just looking at the second-order moment is adequate. Taken together, these arguments tell us that in many cases we can design near-optimal spectrum sensing algorithms by estimating second-order statistics from the data, and making decisions based on these estimates.

We explain detection based on second-order-statistics using OFDM signals as an example. OFDM signals have a very explicit correlation structure imposed by the insertion of a cyclic prefix (CP) at the

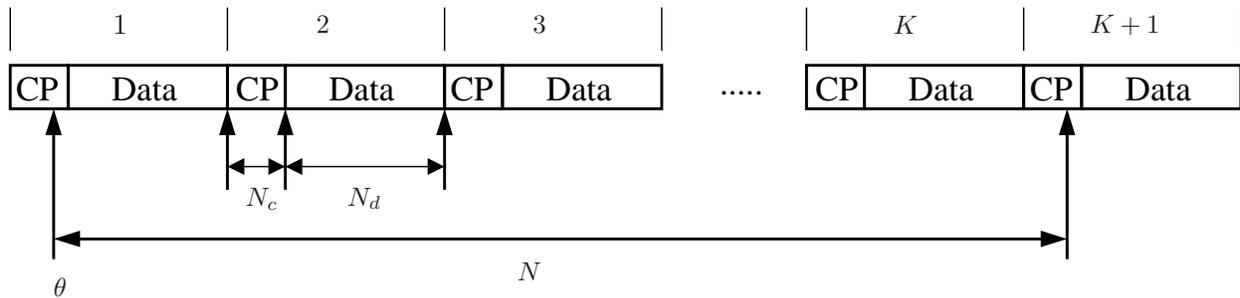


Fig. 2. Model for the N samples of a received OFDM signal.

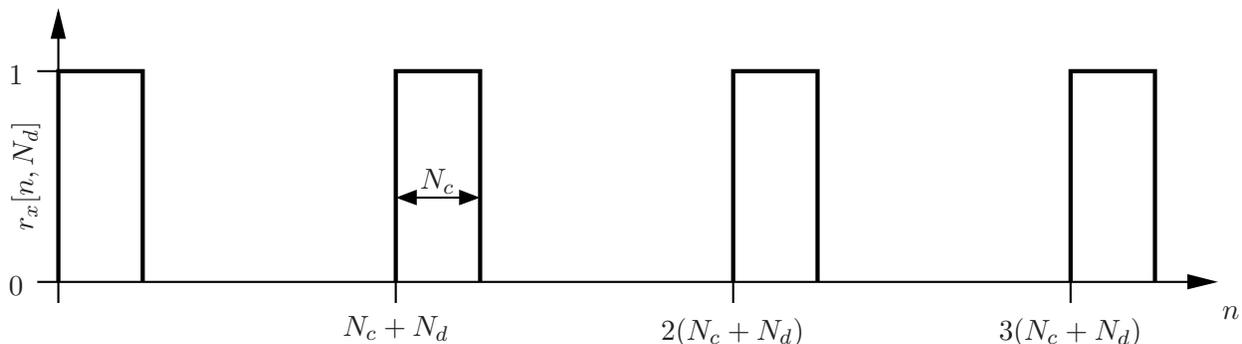


Fig. 3. Example of a periodic autocorrelation function for an OFDM signal with a cyclic prefix.

transmitter. Moreover, OFDM is a popular modulation method in modern wireless standards. Consequently a sequence of papers have proposed detectors that exploit the correlation structure of OFDM signals [12], [17]–[19]. We will briefly describe those detectors in the following. These detectors can be used for any signal with a CP structure, for example single-carrier transmission with a CP and repeated training or so-called known symbol padding, but in what follows we assume that we deal with a conventional OFDM signal.

Consider an OFDM signal with a CP, as shown in Figure 2. Let N_d be the number of data symbols, that is, the block size of the inverse fast Fourier transform (IFFT) used at the transmitter or equivalently the number of subcarriers. The CP has length N_c , and it is a repetition of the last N_c samples of the data. Assume that the transmitted data symbols are independent and identically distributed (i.i.d.), zero-mean and have unit variance, and consider the autocorrelation function (ACF)

$$r_x[n, \tau] \triangleq E[x[n]x^*[n + \tau]]. \quad (6)$$

Owing to the insertion of the CP, the OFDM signal is nonstationary and therefore the ACF $r_x[n, \tau]$ in

(6) is time-varying. In particular, it is non-zero at time lag $\tau = N_d$ for some time instances n , and zero for others. This is illustrated in Figure 3. The non-zero values of the ACF occur due to the repetition of symbols in the CP. This non-stationary property of the ACF can be exploited in different ways by the detectors, as we will see in what follows. Of course, the more knowledge we have of the parameters that determine the shape of the ACF (N_c and N_d specifically, and σ^2), the better performance we can obtain.

For simplicity of notation, assume that the receiver has observed K consecutive OFDM signals out of an endless stream of OFDM modulated data, so that the received signal $y[n]$ contains $N = K(N_c + N_d) + N_d$ samples. Furthermore, for simplicity we consider an additive white Gaussian noise (AWGN) channel. The quantitative second-order statistics will be the same in a multipath fading channel, but the exact ACF may be smeared out due to the time dispersiveness. However, averaging the second-order statistics over multiple OFDM symbols mitigates the impact of multipath fading, and the detection performance is close to the performance in an AWGN channel in many cases (cf. [17]). We are interested in estimating $r_x[n, N_d]$, and we form the following estimate of it:

$$\hat{r}[n] \triangleq y[n]y^*[n + N_d], \quad n = 1, \dots, K(N_c + N_d).$$

Note that $r_w[n, \tau] = 0$ for any $\tau \neq 0$, since the noise is white and zero-mean. Here $r_w[n, \tau]$ and $r_y[n, \tau]$ are defined based on $w[n]$ and $y[n]$ similarly to (6). Hence, $r_y[n, N_d] = r_x[n, N_d]$ whenever $N_d \neq 0$. By construction $E[\hat{r}[n]] = r_y[n, N_d] = r_x[n, N_d]$ is the ACF of the OFDM signal at time lag N_d for $N_d \neq 0$. We know from the above discussion (see Figure 3) that $\hat{r}[n]$ and $\hat{r}[n + k(N_c + N_d)]$ have identical statistics and that they are independent. Therefore, it is useful to define

$$\hat{R}[n] \triangleq \frac{1}{K} \sum_{k=0}^{K-1} \hat{r}[n + k(N_c + N_d)], \quad n = 1, \dots, N_c + N_d.$$

What is the best way of making decisions on signal presence versus absence based on $\hat{r}[n]$? We know that the mean of $\hat{r}[n]$ is nonzero for some n and zero for others and this is the basic observation that we would like to exploit. It is clear that the design of an optimal detector would involve an accurate analysis of the statistical distribution of $\hat{r}[n]$. This is a nontrivial matter, since $\hat{r}[n]$ is a nonlinear function of $y[n]$; moreover, this is difficult if there are unknown parameters such as the noise power. The recent literature has proposed several ways forward.

- One of the first papers on the topic was [18], in which the following statistical test was proposed:

$$\max_{\theta} \left| \sum_{n=\theta+1}^{\theta+N_c} \hat{r}[n] \right| \underset{H_0}{\overset{H_1}{\gtrless}} \eta. \quad (7)$$

The test in (7) exploits the non-stationarity of the OFDM signal. The variable θ in (7) has the interpretation of synchronization mismatch. The intuition behind this detection is therefore to catch the “optimal” value of θ and then measure, for that θ , how large is the correlation between values of $y[n]$ spaced N_d samples apart. For this to work, the detector must know N_c and N_d . Perhaps more importantly, in order to set the threshold one also needs to know σ^2 and hence the detector in (7) is susceptible to the SNR wall phenomenon. This is so for the same reasons as previously discussed for the energy detector: the test statistic in (7) is not dimensionless and hence the test is not CFAR. The original test in [18] looks only at one received OFDM symbol but it can be extended in a straightforward manner to use all K symbols. The resulting statistic then sums the variables $\hat{R}[n]$ instead of $\hat{r}[n]$ and we have

$$\max_{\theta \in \{0, \dots, N_c + N_d - 1\}} \left| \sum_{n \in S_{\theta}} \hat{R}[n] \right| \underset{H_0}{\overset{H_1}{\gtrless}} \eta, \quad (8)$$

where $S_{\theta} \subset \{1, 2, \dots, N_c + N_d\}$ denotes the set of N_c (cyclic) consecutive indices for which $E[\hat{R}[n]] \neq 0$, given the synchronization error θ .

- A different path was taken in [17]. The detector proposed therein uses the empirical mean of the autocorrelation normalized by the received power, as the test statistic. More precisely, the test is

$$\frac{\sum_{n=1}^{N-N_d} \text{Re}(\hat{r}[n])}{\sum_{n=1}^N |y[n]|^2} \underset{H_0}{\overset{H_1}{\gtrless}} \eta. \quad (9)$$

The advantage of (9) is that in order to use this test, one needs to know only N_d , but not N_c . This is useful if N_c is unknown, or if there is substantial uncertainty regarding N_c ; think for example, of a system that alternates between CPs of different lengths or that uses different CPs on different component carriers. On the other hand, a potential disadvantage of (9) is that it does not exploit the fact that the OFDM signal is non-stationary. This is evident from (9) as all samples of $\hat{r}[n]$ are weighed equally when forming the test statistic; hence, the time-variation of the ACF is not reflected in the detection criterion. Not surprisingly, one can obtain better performance if this time-variation is exploited.

Ref.	Detector	Test	σ^2	γ^2	N_d	N_c
[11]	Energy	(5)	×	–	–	–
[17]	Chaudhari et al.	(9)	–	–	×	–
[12]	Axell, Larsson	(10)	–	–	×	×
[18]	Huawei, UESTC	(8)	×	–	×	×
[19]	Lei, Chin		×	×	×	×

TABLE I

SUMMARY OF OFDM DETECTION ALGORITHMS BASED ON SECOND-ORDER-STATISTICS, AND THE SIGNAL PARAMETERS THAT DETERMINE THEIR PERFORMANCE. FOR EACH PARAMETER, “–” MEANS THAT THE DETECTOR DOES NOT NEED TO KNOW THE PARAMETER, AND “×” MEANS THAT IT DOES NEED TO KNOW IT.

By construction, (9) is a CFAR test. Hence, it requires no knowledge of the noise power σ^2 . We note in passing that a detector similar to [17], but without the power normalization, was proposed in [19].

- A more recently proposed test is the following [12]:

$$\max_{\theta \in \{0, \dots, N_c + N_d - 1\}} \frac{\sum_{n=1}^{N_c + N_d} |\hat{R}[n]|^2}{\sum_{n \in S_\theta} \left| \hat{R}[n] - \frac{1}{N_c} \sum_{i \in S_\theta} \text{Re}(\hat{R}[i]) \right|^2 + \sum_{n \notin S_\theta} |\hat{R}[n]|^2} \underset{H_0}{\overset{H_1}{\gtrless}} \eta. \quad (10)$$

Equation (10) is essentially an approximation of the GLRT, treating the synchronization mismatch between the transmitter and the receiver, and the signal and noise variances, as unknown parameters. It needs no knowledge of σ_n^2 and this is directly also evident from (10) as this test statistic is CFAR. It differs from the detectors in [17] and [19] in that it explicitly takes the non-stationarity of $x[n]$ into account. This results in better performance for most scenarios of interest. Of course, the cost for this increased performance is that in contrast to (9), the test in (10) needs to know the CP length, N_c .

The ACF detectors described above are summarized in Table I and a numerical performance comparison between them is shown in Figure 4. This comparison uses an AWGN channel, and parameters as follows: $P_{\text{FA}} = 0.05$, $N_d = 32$, $N_c = 8$ and $K = 50$. The performance of the energy detector is also included as a baseline, both with perfectly known noise variance and with a 1 dB mismatch. It is clear that knowing the

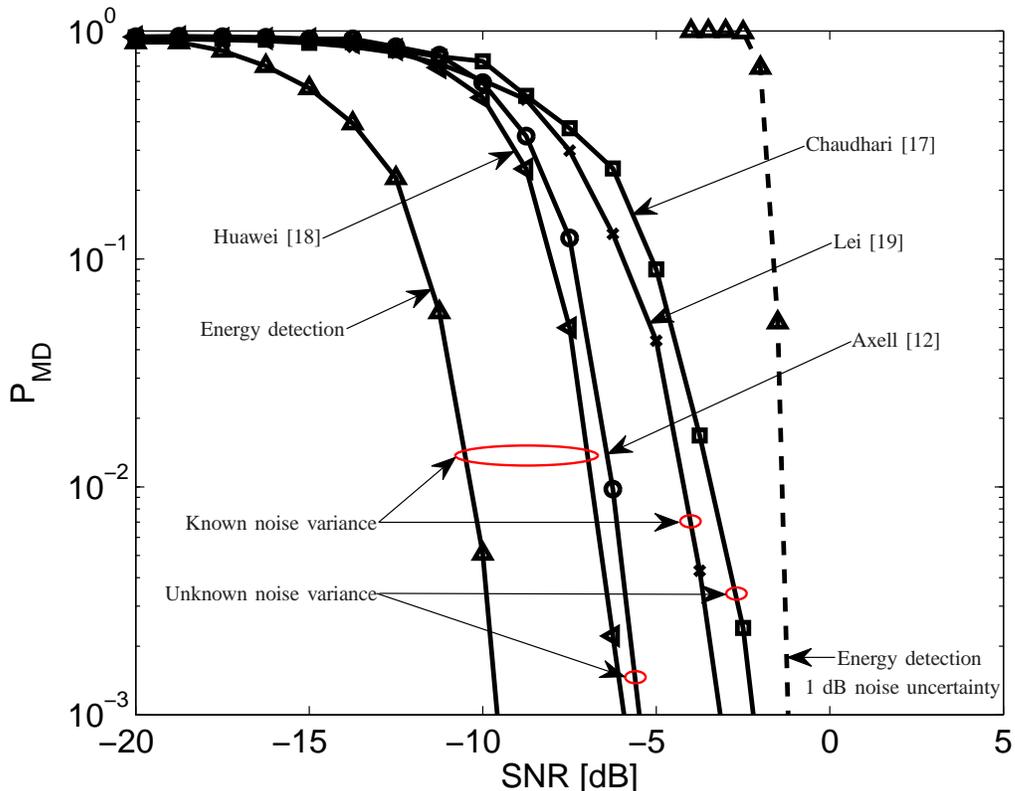


Fig. 4. Comparison of the autocorrelation-based detection schemes. $P_{FA} = 0.05$, $N_d = 32$, $N_c = 8$, $K = 50$.

noise variance significantly improves the detector performance. Interestingly, here, the energy detector has the best performance when the noise variance is known, and the worst performance when the noise variance is uncertain with as little as 1 dB. When the noise power is not known, more sophisticated detectors such as those of [17] and [12] must be used.

B. Detectors Based on Cyclostationarity

In many cases, the ACF of the signal is not only non-stationary, but is also periodic. Most man-made signals show periodic patterns related to symbol rate, chip rate, channel code or cyclic prefix. Such second-order periodic signals can be appropriately modeled as second-order cyclostationary random processes [20]. As an example, consider again the OFDM signal shown in Figure 2. The autocorrelation function of this OFDM signal, shown in Figure 3, is periodic. The fundamental period is the length of the OFDM symbol, $N_c + N_d$. Knowing some of the cyclic characteristics of a signal, one can construct detectors that exploit the cyclostationarity [21], [22] and benefit from the spectral correlation.

A discrete-time zero-mean stochastic process $y[n]$ is said to be *second-order cyclostationary* if its time-varying ACF $r_y[n, \tau] = E[y[n]y^*[n + \tau]]$ is periodic in n [20], [21]. Hence, $r_y[n, \tau]$ can be expressed by a Fourier series

$$r_y[n, \tau] = \sum_{\alpha} R_y(\alpha, \tau) e^{j\alpha n},$$

where the sum is over integer multiples of fundamental frequencies and their sums and differences. The Fourier coefficients depend on the time lag τ and are given by

$$R_y(\alpha, \tau) = \frac{1}{N} \sum_{n=0}^{N-1} r_y[n, \tau] e^{-j\alpha n}.$$

The Fourier coefficients $R_y(\alpha, \tau)$ are also known as the *cyclic autocorrelation* at cyclic frequency α . The process $y[n]$ is second-order cyclostationary when there exists an $\alpha \neq 0$ such that $R_y(\alpha, \tau) > 0$, because $r_y[n, \tau]$ is periodic in n precisely in this case. The *cyclic spectrum* of the signal $y[n]$ is the Fourier coefficient

$$S_y(\alpha, \omega) = \sum_{\tau} R_y(\alpha, \tau) e^{-j\omega\tau}.$$

The cyclic spectrum represents the density of correlation for the cyclic frequency α .

Knowing some of the cyclic characteristics of a signal, one can construct detectors that exploit the cyclostationarity and thus benefit from the spectral correlation (see, e.g., [21]–[23]). Note that the inherent cyclostationarity property appears both in the cyclic ACF $R_y(\alpha, \tau)$ and in the cyclic spectral density function $S_y(\alpha, \omega)$. Thus, detection of the cyclostationarity can be performed both in the time domain, and in the frequency domain. The paper [21] proposed detectors that exploit cyclostationarity based on one cyclic frequency, either from estimates of the cyclic autocorrelation or of the cyclic spectrum. The detector of [21] based on cyclic autocorrelation was extended in [22] to use multiple cyclic frequencies. The cyclic autocorrelation is estimated in [21] and [22] by

$$\hat{R}_y(\alpha, \tau) \triangleq \frac{1}{N} \sum_{n=0}^{N-1} y[n]y^*[n + \tau] e^{-j\alpha n}.$$

The cyclic autocorrelation $\hat{R}_y(\alpha_i, \tau_{i, N_i})$ can be estimated for the cyclic frequencies of interest α_i , $i = 1, \dots, p$, at time lags $\tau_{i,1}, \dots, \tau_{i, N_i}$. The detectors of [21] and [22] are then based on the limiting probability distribution of $\hat{R}_y(\alpha_i, \tau_{i, N_i})$, $i = 1, \dots, p$.

In practice only one or a few cyclic frequencies are used for detection, and this is usually sufficient to achieve a good detection performance. Note however that this is an approximation. For example, a perfect Fourier series representation of the signal shown in Figure 3 requires infinitely many Fourier coefficients. The autocorrelation-based detector of [17] and the cyclostationarity detector of [22] are compared in [24], for detection of an OFDM signal in AWGN. The results show that the cyclostationarity detector using two cyclic frequencies outperforms the autocorrelation detector, but that the autocorrelation detector is superior when only one cyclic frequency is used.

C. Detectors that Rely on a Specific Structure of the Sample Covariance Matrix

Signal structure, or correlation, is also inherent in the covariance matrix of the received signal. Some communication signals impart a specific known structure to the covariance matrix. This is the case for example when the signal is received by multiple antennas [25]–[27] (single-input/multiple-output - SIMO), [10] (multiple-input/multiple-output - MIMO), when the signal is encoded with an orthogonal space-time block code (OSTBC) [28], or if the signal is an OFDM signal [12]. In these cases, the covariance matrix has a known eigenvalue structure, as shown in [29].

Consider again the vectorial discrete-time representation (1). For better understanding we will start with the example of a single symbol received by multiple antennas (SIMO). This case was dealt with, for example, in [10], [25], [26] and [27]. Suppose that there are $L > 1$ receive antennas at the detector. Then, under H_1 , the received signal can be written as

$$\mathbf{y}[n] = \mathbf{h}s[n] + \mathbf{w}[n], \quad n = 1, \dots, N, \quad (11)$$

where \mathbf{h} is the $L \times 1$ channel vector and $s[n]$ is the transmitted symbol sequence. Assume further that the signal is zero-mean Gaussian, i.e. $s[n] \sim \mathcal{N}(0, \gamma^2)$, and as before $\mathbf{w}[n] \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$. Then, the covariance matrix under H_1 is $\mathbf{\Psi} \triangleq E[\mathbf{y}[n]\mathbf{y}[n]^H | H_1] = \gamma^2 \mathbf{h}\mathbf{h}^H + \sigma^2 \mathbf{I}$. Let $\lambda_1, \lambda_2, \dots, \lambda_L$ be the eigenvalues of $\mathbf{\Psi}$ sorted in descending order. Since $\mathbf{h}\mathbf{h}^H$ has rank one, then $\lambda_1 = \gamma^2 \|\mathbf{h}\|^2 + \sigma^2$ and $\lambda_2 = \dots = \lambda_L = \sigma^2$. In other words, $\mathbf{\Psi}$ has two distinct eigenvalues with multiplicities one and $L - 1$ respectively. Denote the sample covariance matrix by

$$\hat{\mathbf{R}} \triangleq \frac{1}{N} \sum_{n=1}^N \mathbf{y}[n]\mathbf{y}[n]^H.$$

Moreover, let $\nu_1, \nu_2, \dots, \nu_L$ denote the eigenvalues of $\hat{\mathbf{R}}$ sorted in descending order. An example of the

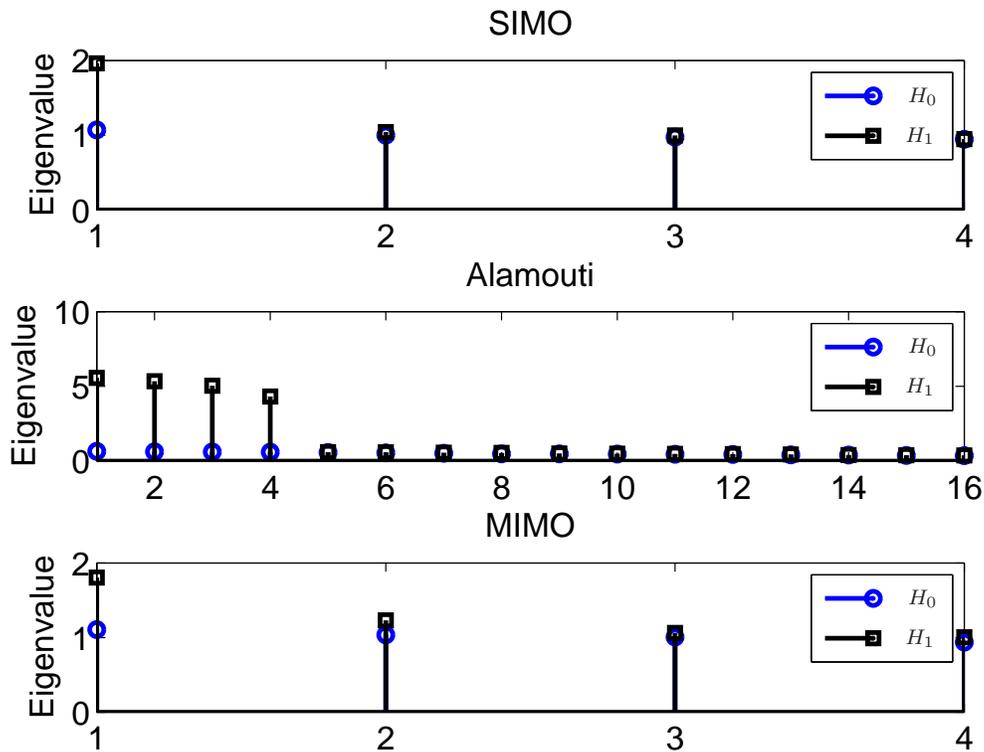


Fig. 5. Example of the sorted eigenvalues of the sample covariance matrix $\hat{\mathbf{R}}$ with four receive antennas, for $N = 1000$ and $\text{SNR} = 10 \log_{10}(\gamma^2/\sigma^2) = 0$ dB. The Alamouti scheme codes two complex symbols over two time intervals and two antennas.

eigenvalues $\nu_1, \nu_2, \dots, \nu_L$ in this case, with four receive antennas, $N = 1000$ and $\text{SNR} = 10 \log_{10}(\gamma^2/\sigma^2) = 0$ dB, is shown at the top of Figure 5. It is clear that there is one dominant eigenvalue under H_1 due to the rank-one channel vector. It can be shown (cf. [25] and [26]) that the GLRT when the channel \mathbf{h} and the powers σ^2 and γ^2 are unknown is given by

$$\frac{\nu_1}{\text{trace}(\hat{\mathbf{R}})} = \frac{\nu_1}{\sum_{i=1}^L \nu_i} \underset{H_0}{\overset{H_1}{\geq}} \eta. \quad (12)$$

Here, we have considered independent observations $\mathbf{y}[n]$ at multiple antennas. A similar covariance structure could of course also occur for a time-series. Then, we could construct the sample covariance matrix by considering a scalar time-series $y[n]$, $n = 1, 2, \dots, N$ as in [30] and [31], and letting $\mathbf{y}[n] = [y[n], y[n+1], \dots, y[n+L-1]]^T$ for some integer $L > 0$. This can be seen as a windowing of the sequence $y[n]$ with a rectangular window of length L . The choice of the window length L will of course affect the performance of the detectors. The reader is referred to the original papers [30] and [31] for

discussions of this issue.

Now, consider more generally that the received signal under H_1 can be written as

$$\mathbf{y}[n] = \mathbf{G}\mathbf{s}[n] + \mathbf{w}[n], \quad n = 1, \dots, N, \quad (13)$$

where \mathbf{G} is a low rank matrix, and $\mathbf{s}[n] \sim \mathcal{N}(\mathbf{0}, \gamma^2 \mathbf{I})$ is an i.i.d. sequence. Then, the covariance matrix is $\mathbf{\Psi} = \gamma^2 \mathbf{G}\mathbf{G}^H + \sigma^2 \mathbf{I}$ under H_1 , which is a “low-rank-plus-identity” structure. Suppose that $\mathbf{\Psi}$ has d distinct eigenvalues with multiplicities q_1, q_2, \dots, q_d respectively. This can happen if the signal has some specific structure, for example in a multiple antenna (MIMO) system [10], when the signal is encoded with an orthogonal space-time block code [28], or if the signal is an OFDM signal [12], [29]. Examples of the sorted eigenvalues of $\hat{\mathbf{R}}$ for an orthogonal space-time block code (Alamouti) [28], and for a general MIMO system [10], with two transmit and four receive antennas, are shown in Figure 5 (middle and bottom respectively). The reason that the number of eigenvalues for the Alamouti case is four times higher than for the general MIMO system is that the space-time code is coded over two time intervals, and the observation is divided into real and imaginary parts (see [28] for details). For the Alamouti code, the four largest eigenvalues are significantly larger than the others. In fact, the expected values of the four largest eigenvalues are equal, due to the orthogonality of the code. For the general MIMO case, we note that two of the eigenvalues are significantly larger than the others, because the channel matrix has rank two (there are two transmit antennas). In this case, however, the expectations of the two largest eigenvalues are different in general. Define the set of indices $\mathcal{S}_i \triangleq \{(\sum_{j=1}^{i-1} q_j) + 1, \dots, \sum_{l=1}^i q_l\}$, $i = 1, 2, \dots, d$. For example, if there are two distinct eigenvalues with multiplicities q_1 and $q_2 (= L - q_1)$ respectively, then $\mathcal{S}_1 = \{1, \dots, q_1\}$ and $\mathcal{S}_2 = \{q_1 + 1, \dots, L\}$. It was shown in [29] that the GLRT when the eigenvalues are unknown, but have known multiplicities and order, is

$$\frac{\left(\frac{1}{L} \text{trace}(\hat{\mathbf{R}})\right)^L}{\prod_{i=1}^d \left(\frac{1}{q_i} \sum_{j \in \mathcal{S}_i} \nu_j\right)^{q_i}} \underset{H_0}{\overset{H_1}{\gtrless}} \eta. \quad (14)$$

It can be shown that in the special case when $q_1 = 1$ and $q_2 = L - 1$, this test is equivalent to the test (12).

Properties of the covariance matrix are also exploited for detection in [30] and [31], without knowing the structure. Detection without any knowledge of the transmitted signal is usually referred to as *blind detection*, and will be discussed further in the following section.

D. Blind Detection

Even though a primary user's signal is correlated or has some other structure, this structure might not be perfectly known. An example of this is shown at the bottom of Figure 5. This eigenvalue structure occurs in a general MIMO system, when the number of receive antennas is larger than the number of transmit antennas. In general, the number of antennas and the coding scheme used at the transmitter might not be known. The transmit antennas could of course also belong to an (unknown) number of users that transmit simultaneously [32], [33]. If the transmitted signals have a completely unknown structure, we must consider *blind detectors*. Blind detectors are blind in the sense that they exploit structure of the signal without any knowledge of signal parameters. We saw in the previous section that the eigenvalues of the covariance matrix behave differently under H_0 and H_1 if the signal is correlated. This is still true, even if the exact structure of the eigenvalues is not known. Blind eigenvalue-based tests, similar to those described in the previous section, have been proposed recently in [30] and [31].

We will begin by describing the blind detectors of [30] and [31] based on the eigenvalues of the sample covariance matrix. The presentation here will be slightly different from the ones in [30] and [31], in order to include complex-valued data and be consistent with the notation used above. The paper [30] proposes two detectors based on the eigenvalues of $\hat{\mathbf{R}}$, similar to the detectors of the previous section. The detectors proposed in [30] are

$$\frac{\nu_1}{\nu_L} \underset{H_0}{\overset{H_1}{\gtrless}} \eta, \quad \text{and} \quad \frac{\text{trace}(\hat{\mathbf{R}})}{\nu_L} \underset{H_0}{\overset{H_1}{\gtrless}} \eta,$$

where $\nu_i, i = 1, 2, \dots, L$ are the sorted eigenvalues of $\hat{\mathbf{R}}$, as before. Thus, ν_1 is the maximum eigenvalue and ν_L is the minimum eigenvalue. The motivation for these tests is based on properties similar to those discussed in Section III-C. If the received sequence contains a (correlated) signal, the expectation of the largest eigenvalues will be larger than if there is only noise, but the expectation of the smallest eigenvalues will be the same in both cases.

Blind detectors are commonly also based on information theoretic criteria, such as Akaike's Information Criterion (AIC) or the Minimum Description Length (MDL) [32]–[35]. These information theoretic criteria typically result in eigenvalue tests similar to those of the previous section. The aims of [32] and [33] are not only to decide whether a signal has been transmitted or not, but rather also to estimate

the number of signals transmitted. Assume as in the previous section that the received signal under H_1 is $\mathbf{y}[n] = \mathbf{G}\mathbf{s}[n] + \mathbf{w}[n]$. The number of uncorrelated transmitters is the rank d of the matrix \mathbf{G} . The problem of [32] and [33] is then to determine the rank of \mathbf{G} by minimizing the AIC or MDL, which are functions of d . The result of [32] is applied in [34] and [35] to the problem of spectrum sensing. More specifically, the estimator of [32] is used in [34] to determine whether the number of signals transmitted is zero or non-zero. This idea is further simplified in [35] to that of using only the difference $\text{AIC}(0) - \text{AIC}(1)$ as a test statistic. Note that these detectors are very similar to the detectors of the previous section and to the detectors described in the beginning of this section. They all exploit properties of the eigenvalues of the sample covariance matrix, and use functions of the eigenvalues as test statistics. The detectors of this section use only the assumption that the received signal is correlated. They are all blind detectors, in the sense that they do not require any more knowledge.

E. Filterbank-based Detectors and Multitaper Methods

If the spectral properties of the signal to be detected are known, but the signal has otherwise no usable features that can be efficiently exploited, then spectrum estimation techniques like filterbank-based detectors may be preferable [3], [36]–[38]. In addition, if the cognitive radio system exploits a filter bank multicarrier technique, the same filter bank can be used for both transmission and spectrum sensing [36]. Hence, the sensing can be done without any additional cost. In the following, we briefly describe spectrum estimation based on filterbanks and multitaper methods.

Suppose that we are interested in estimating the spectrum in the frequency band from $f - B/2$ to $f + B/2$. The standard periodogram estimates the spectrum of the random process $y[n]$ based on N samples as

$$\hat{S}(f) = \left| \sum_{n=1}^N v[n] e^{-j2\pi f n} y[n] \right|^2,$$

where $v[n]$ is a window function. The window function $v[n]$ is a finite-impulse-response (FIR) low pass filter with bandwidth B , usually called a prototype filter. In this case, $v[n]e^{-j2\pi f n}$ is a bandpass filter centered at frequency f . The filterbank spectral estimator improves the estimate by using multiple prototype filters $v_k[n]$ and by averaging the energy of the filter outputs. This leads to a k th output spectrum

of the form

$$\hat{S}_k(f) = \left| \sum_{n=1}^N v_k[n] e^{-j2\pi f n y[n]} \right|^2.$$

The prototype filters $v_k[n]$ must be chosen properly. The multitaper method (cf. [37] or [38]), uses the so-called *Slepian sequences* also known as *discrete prolate spheroidal wave functions* as prototype filter coefficients. The Slepian sequences are characterized by two important properties: i) they have maximal energy in the main lobe, and ii) they are orthonormal. The orthogonality assures that the outputs from the prototype filters are uncorrelated, as long as the variation over each subband is negligible. After estimating the spectrum of the frequency band of interest, one can perform spectrum sensing using, for example, energy detection. Moreover, [38] analyzes the space-time and time-frequency properties of the multitaper estimates, for exploitation of signal features for spectrum sensing as discussed in the previous sections. The cyclostationarity property is given particular emphasis. For more details on spectrum sensing using filterbanks and multitaper methods, we refer the reader to [36] and [38].

IV. WIDEBAND SPECTRUM SENSING

In many cognitive radio applications, a wide band of spectrum must be sensed, which requires high sampling rates and thus high power consumption in the A/D converters. One solution to this problem is to divide the wideband channel into multiple parallel narrowband channels, and to jointly sense transmission opportunities on those channels. This technique is called *multiband sensing*. Another approach argues that the interference from the primary users can often be interpreted as being sparse in some particular domain, e.g., in the spectrum or in the edge spectrum (the derivative of the spectrum). In that case, subsampling methods or *compressive sensing* (see [39] and [40] and the references therein) can be used to lower the burden on the A/D converters.

A. Multiband Sensing

A simple, and sometimes most natural, way of dealing with a wideband channel is to divide it into multiple subchannels as shown in Figure 6. Think for example of a number of digital TV bands. Together, they constitute a wideband spectrum, but are naturally divided into subchannels. In general, the subchannels do not even have to be contiguous. Some of the subchannels may be occupied and some may

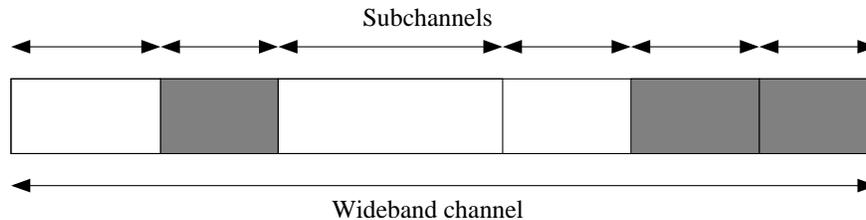


Fig. 6. Example of a wideband channel divided into multiple subchannels. The white subchannels represent white spaces, or spectrum holes, and the shaded subchannels represent occupied channels.

be available. The problem of multiband sensing is of course to decide upon which of the subchannels are occupied and which are available.

The simplest approach to the multiband sensing problem is to assume that all subchannels (and unknown parameters) are independent. Then, the multiband sensing problem reduces to a binary hypothesis test of the type (2) for each subchannel. However, in practice the subchannels are not independent. For example, the primary user occupancy can be correlated [41], or the noise variance can be unknown but correlated between the bands [9]. Then, the detection problem becomes a composite hypothesis test, that grows exponentially with the number of subchannels. The huge complexity of the optimal detector, then leads to the need for approximations or simplifications of the detection algorithm (cf. [9] and [41]).

Many papers on multiband sensing, have also considered joint spectrum sensing and efficient resource utilization. For example, we may wish to maximize the communication rate or allocate other resources within constraints on the detection probability [42], [43]. The opportunistic sum-rate over all subchannels is maximized in [42] and [43], with constraints on the detection probabilities. Multiple cooperating sensors are used in [42] to improve the detection performance and robustness. However, only one secondary transmitter is considered in [42], whereas multiple secondary users, and allocation of them to the available subchannels, are dealt with in [43]. This may lead to non-convex and potentially NP-hard optimization problems.

B. Compressive Sensing

The basic idea of compressive spectrum sensing is to exploit the fact that the original observed analog signal $y(t)$ with double-sided bandwidth or Nyquist rate $1/T$ can often be sampled below the Nyquist rate within an interval $t \in [0, N_b T)$ through a special linear sampling process, sometimes referred to as

an analog-to-information (A/I) converter. The resulting $M_b \times 1$ vector of samples $\mathbf{z} = [z[1], \dots, z[M_b]]^T$ can then be expressed as

$$\mathbf{z} = \Phi \mathbf{y}, \quad (15)$$

where $\mathbf{y} = [y[1], \dots, y[N_b]]^T$ is the $N_b \times 1$ vector obtained by Nyquist rate sampling $y(t)$ within the interval $t \in [0, N_b T)$, and Φ is the $M_b \times N_b$ measurement matrix, where $M_b \ll N_b$. We remark that (15) is used only for representation purposes. It represents an operation that is carried out in the analog domain, and not in the digital domain. So the compression ratio compared to Nyquist rate sampling is given by M_b/N_b . Depending on the type of A/I converter, the measurement matrix can take different forms. In wideband spectrum sensing, one often resorts to a non-uniform sampler (Φ consists of M_b randomly selected rows from the $N_b \times N_b$ identity matrix) or a random demodulator (Φ consists of random entries, uniformly, normally, or ± 1 distributed). Now since (15) has more unknowns than equations, it has infinitely many solutions and in order to reduce the feasible set, additional constraints are introduced. In compressive sensing, these constraints are based on sparsity considerations for \mathbf{y} . More specifically, it is assumed that \mathbf{y} is sparse in some basis Ψ , meaning that we can write $\mathbf{y} = \Psi \mathbf{s}$, where \mathbf{s} has only a few non-zero elements. For instance, if primary user presence is not very likely, sparsity reveals itself in the spectrum, i.e., $\Psi = \mathbf{F}^{-1}$, with \mathbf{F} the $N_b \times N_b$ discrete Fourier transform (DFT) matrix, whereas if primary users occupy only flat frequency bands, the edge spectrum (the derivative of the spectrum) can be viewed as being sparse, i.e., $\Psi = (\mathbf{\Gamma} \mathbf{F})^{-1}$ with $\mathbf{\Gamma}$ the $N_b \times N_b$ differentiation matrix³ [44]. Under such sparsity constraints (possibly relaxed), we can then solve

$$\mathbf{z} = \Phi \Psi \mathbf{s} = \mathbf{A} \mathbf{s}, \quad (16)$$

using any existing sparse reconstruction method such as orthogonal matching pursuit (OMP), basis pursuit (BP), or the least-absolute shrinkage and selection operator (LASSO) (see [40] and references therein).

It is also possible to carry out the above sampling process in every consecutive interval of length $N_b T$, resulting in a periodic sampling device, e.g., a periodic non-uniform sampler (also known as a multi-coset sampler) or a periodic random demodulator (also known as a modulated wideband converter). For the

³In practice, spectral smoothing is required to obtain improved sparsity in the spectrum or edge spectrum. However, we abstract this operation in this work.

k th interval, we then obtain $\mathbf{z}[k] = \mathbf{A}\mathbf{s}[k]$, and stacking K such vectors in a matrix, we obtain

$$\mathbf{Z} = \mathbf{A}\mathbf{S}, \quad (17)$$

where the $M_b \times K$ matrix \mathbf{Z} and $N_b \times K$ matrix \mathbf{S} are respectively given by $\mathbf{Z} = [\mathbf{z}[1], \dots, \mathbf{z}[K]]$ and $\mathbf{S} = [\mathbf{s}[1], \dots, \mathbf{s}[K]]$. In that case, we can resort to so-called multiple measurement vector (MMV) approaches to sparse reconstruction, thereby exploiting the fact that all the columns of \mathbf{S} enjoy the same sparsity pattern [45]. However, in this MMV case, also more traditional sparse reconstruction methods can be employed, such as multiple signal classification (MUSIC) or the minimum variance distortionless response (MVDR) method. It is interesting to observe that this MMV set-up is very closely related to *spectrum-blind sampling*, in which the goal is to enable minimum-rate sampling and reconstruction given that the spectrum is sparse yet unknown [46].

Cooperative versions of compressive wideband sensing have also been developed [47], [48]. Here, individual radios can make a local decision about the presence or absence of a primary user, and these results can then be fused in a centralized or decentralized manner. However, a greater cooperation gain can be achieved by fusing all the compressed measurements, again in a centralized or decentralized manner. In general, such measurement fusion requires that each cognitive radio knows the channel state information (CSI) from all primary users to itself [47], which is cumbersome. But recent extensions show that measurement fusion can also be carried out without CSI knowledge [49].

V. COOPERATIVE SPECTRUM SENSING

Spectrum sensing using a single cognitive radio has a number of limitations. First of all, the sensitivity of a single sensing device might be limited because of energy constraints. Furthermore, the cognitive radio might be located in a deep fade of the primary user signal, and as such might miss the detection of this primary user. Moreover, although the cognitive radio might be blocked from the primary user's transmitter, this does not mean it is also blocked from the primary user's receiver, an effect that is known as the hidden terminal problem. As a result, the primary user is not detected but the secondary transmission could still significantly interfere at the primary user's receiver. To improve the sensitivity of cognitive radio spectrum sensing, and to make it more robust against fading and the hidden terminal problem, cooperative sensing can be used. The concept of cooperative sensing is to use multiple sensors and

combine their measurements into one common decision. In this section, we will consider this approach, including both soft combining and hard combining, where for the latter we will also look at the influence of fading of the reporting channels to the fusion center. Throughout this and other sections on cooperative sensing, we will indicate the local probabilities of detection, missed detection, and false alarm as P_d , P_{md} , and P_{fa} , respectively, whereas their global representatives will be denoted as P_D , P_{MD} , and P_{FA} .

A. Soft Combining

Assume that there are M sensors. Then, the hypothesis test (2) becomes

$$H_0 : \mathbf{y}_m = \mathbf{w}_m, m = 1, \dots, M,$$

$$H_1 : \mathbf{y}_m = \mathbf{x}_m + \mathbf{w}_m, m = 1, \dots, M.$$

Suppose that the received signals at different sensors are independent of one another, and let $\mathbf{y} = [\mathbf{y}_1^T, \mathbf{y}_2^T, \dots, \mathbf{y}_M^T]^T$. Then, the log-likelihood ratio is

$$\log \left(\frac{p(\mathbf{y}|H_1)}{p(\mathbf{y}|H_0)} \right) = \log \left(\prod_{m=1}^M \frac{p(\mathbf{y}_m|H_1)}{p(\mathbf{y}_m|H_0)} \right) = \sum_{m=1}^M \log \left(\frac{p(\mathbf{y}_m|H_1)}{p(\mathbf{y}_m|H_0)} \right) = \sum_{m=1}^M \Lambda^{(m)}, \quad (18)$$

where $\Lambda^{(m)} = \log \left(\frac{p(\mathbf{y}_m|H_1)}{p(\mathbf{y}_m|H_0)} \right)$ is the log-likelihood ratio for the m th sensor. That is, if the received signals for all sensors are independent, the optimal fusion rule is to sum the local log-likelihood ratios.

Consider the case in which the noise vectors \mathbf{w}_m are independent $\mathbf{w}_m \sim \mathcal{N}(\mathbf{0}, \sigma_m^2 \mathbf{I})$, and the signal vectors \mathbf{x}_m are independent $\mathbf{x}_m \sim \mathcal{N}(\mathbf{0}, \gamma_m^2 \mathbf{I})$. After removal of irrelevant constants, the log-likelihood ratio (18) can be written as

$$\sum_{m=1}^M \frac{\|\mathbf{y}_m\|^2}{\sigma_m^2} \frac{\gamma_m^2}{(\sigma_m^2 + \gamma_m^2)}. \quad (19)$$

The statistic $\frac{\|\mathbf{y}_m\|^2}{\sigma_m^2}$ is the soft decision from an energy detector at the m th sensor, as shown in (5). Thus, the optimal cooperative detection scheme is to use energy detection for the individual sensors, and combine the soft decisions by the weighted sum (19). This result is also shown in [50], for the case when $\sigma_m^2 = 1$, and thus γ_m^2 is equivalent to the SNR experienced by the m th sensor. The cooperative gain and the effect of untrusted users, under the assumption that the noise and signal powers are equal for all sensors, are analyzed in [51]. It is shown in [51] that correlation between the sensors severely decreases the cooperation gain and that if one out of M sensors is untrustworthy, then the sensitivity of each individual sensor must be as good as that achieved with M trusted users.

B. Hard Combining

So far we have considered optimal cooperative detection. That is, all users transmit soft decisions to a fusion center, which combines the soft values to one common decision. This is equivalent to the case in which the fusion center has access to the received data for all sensors, and performs optimal detection based on all data. This potentially requires a very large amount of data to be transmitted to the fusion center. The other extreme case of cooperative detection is that each sensor makes its own individual decision, and transmits only a binary value to the fusion center. Then, the fusion center combines the hard decisions into one common decision, for instance using a voting rule (c.f. [52]).

Suppose that the individual statistics $\Lambda^{(m)}$ are quantized to one bit, such that $\Lambda^{(m)} \in \{0, 1\}$ is the hard decision from the m th sensor. Here, 1 means that a signal is detected and 0 means that the channel is deemed to be available. The voting rule then decides that a signal is present if at least C of the M sensors have detected a signal, for $1 \leq C \leq M$. The test decides on H_1 if

$$\sum_{m=0}^{M-1} \Lambda^{(m)} \geq C.$$

A majority decision is a special case of the voting rule when $C = M/2$, whereas the AND-logic and OR-logic are obtained for $C = M$ and $C = 1$, respectively. In [53], hard combining is studied for energy detection with equal SNR for all cognitive radios. In particular, the optimal voting rule, optimal local decision threshold, and minimal number of cognitive radios are derived, where optimality is defined in terms of the (unweighted) global probability of error $P_{\text{FA}} + P_{\text{MD}}$ (note that this is different from the true global probability of error). It turns out that when the local probability of false alarm P_{fa} and missed detection P_{md} are of the same order, the majority rule is optimal, whereas the optimal voting rule leans towards the OR rule if $P_{\text{fa}} \ll P_{\text{md}}$ and to the AND rule if $P_{\text{fa}} \gg P_{\text{md}}$.

There are also some works that consider BPSK signaling of the hard local decisions to the fusion center over fading reporting channels, and assuming phase coherent reception. Such a scenario is investigated in [54], in which the corresponding optimal fusion rule of the received signals is derived. This fusion rule requires the knowledge of the reporting channel SNRs as well as the local probabilities of false alarm $\{P_{\text{fa}}^{(m)}\}_m$ and detection $\{P_{\text{d}}^{(m)}\}_m$. At high SNR, this fusion rule corresponds to the Chair-Varshney rule [55], in which knowledge of only $\{P_{\text{fa}}^{(m)}\}_m$ and $\{P_{\text{d}}^{(m)}\}_m$ is required, whereas at low SNR, it

becomes the maximal ratio combiner (if $P_d^{(m)} = P_d$, $P_{fa}^{(m)} = P_{fa}$, and $P_d > P_{fa}$), for which only the reporting channel SNRs are needed. As a robust alternative, equal gain combining is also suggested, which does not require any prior knowledge. In [56], the above optimal fusion rule is extended to the case in which the channel is rapidly Rayleigh fading, such that only the channel statistics can be obtained, and as before phase coherent reception is assumed. In this case, at high SNR, the optimal fusion rule corresponds again to the Chair-Varshney rule, but at low SNR, it now becomes the equal gain combiner (if $P_d^{(m)} = P_d$, $P_{fa}^{(m)} = P_{fa}$, and $P_d > P_{fa}$). When on/off signaling is assumed with non-coherent reception at the fusion center, the optimal decision rule is derived in [57], with either the knowledge of the reporting channel envelopes or the knowledge of the channel statistics. And as before, also the local probabilities of false alarm $\{P_{fa}^{(m)}\}_m$ and detection $\{P_d^{(m)}\}_m$ are required for these optimal fusion rules. At low SNR, both rules lead to a weighted energy detector (if $P_d^{(m)} = P_d$, $P_{fa}^{(m)} = P_{fa}$, and $P_d > P_{fa}$). If the channel envelopes are known, the weights are given by the channel powers, and if the channel statistics are known, the weights are all the same for Rayleigh or Nakagami fading channels (the weighted energy detector then reduces to an energy detector), whereas they are given by the powers of the line-of-sight components for Rician fading channels.

VI. ENERGY EFFICIENCY IN COOPERATIVE SPECTRUM SENSING

When using techniques such as those described in the preceding section, as the number of cooperating users grows, the energy consumption of the cognitive radio network increases, but the performance generally saturates. Hence, techniques have been developed to improve the energy efficiency in cooperative cognitive radio networks. In this section, we will review some of these briefly.

A. Cooperative Sequential Sensing

In classical sequential detection, the basic idea is to minimize the sensing energy by minimizing the average sensing time, subject to constraints on the probability of false alarm and missed detection, i.e., $P_{FA} \leq \alpha$ and $P_{MD} \leq \beta$. These two constraints are important in a cognitive radio system, since P_{FA} is related to the throughput of the cognitive radio system, whereas P_{MD} is related to the interference to the primary system. Under i.i.d. observations, this leads to the so-called sequential probability ratio test (SPRT) [58], in which sensing is continued as long as the likelihood ratio Λ satisfies $\eta_1 \leq \Lambda < \eta_2$

and a decision is made otherwise, with $\eta_1 = \beta/(1 - \alpha)$ and $\eta_2 = (1 - \beta)/\alpha$. Note that one can also consider minimizing the average Bayesian cost of sensing and making a wrong decision, but this also leads to an SPRT. Sequential detection has been adopted to reduce the sensing time in single-radio spectrum sensing, see e.g. [59]. However, multi-sensor versions of sequential detection, i.e., cooperative or distributed sequential detection (see [60] and references therein), are encountered more frequently in the field of spectrum sensing, since they provide the ability to significantly reduce the energy consumption of the overall system. In the following, we briefly discuss a few of these approaches.

In [61], all the radios send their most current local log-likelihood ratios (LLRs) to the fusion center, where an SPRT will be carried out. If the test is positive, a decision can be made and the radios can stop sensing and transmitting, thereby saving not only sensing energy but also transmission energy. If not, all the radios gather new information and send their corresponding new LLRs to the fusion center. Unknown modeling parameters are also taken into account in [61], following an approach similar to Section II-A. In [17], on the other hand, the radios will not send their LLRs in parallel to the fusion center, as done in [61], but they do it sequentially. If the SPRT performed at the fusion center is negative, only one radio that did not yet participate in the fusion gathers new information and sends its LLR to the fusion center. Note that the LLRs in [17] are based on second-order statistics.

B. Censoring

Another popular energy-aware cooperative sensing technique is *censoring*. In such a system a cognitive radio m will send a sensing result only if it is deemed informative, and it will censor those sensing results that are uninformative. In [62], optimal censoring has been considered in terms of the global probability of error $P_E = Pr(H_0)P_{FA} + Pr(H_1)(1 - P_D)$ (Bayesian framework), the global probability of detection P_D subject to a global probability of false alarm constraint $P_{FA} \leq \alpha$ (Neyman-Pearson framework), or any Ali-Silvey distance between the two hypotheses (such as the J-divergence). If we interpret this for a cognitive radio system, the Bayesian approach basically minimizes the difference between the interference to the primary system and the throughput of the cognitive radio system. The Neyman-Pearson approach minimizes the interference to the primary system subject to a minimal throughput of the cognitive radio system. The Ali-Silvey distance provides a generalization, which we simply mention here for completeness. In addition, a global communication constraint is adopted, which is given by a

constraint on the true global rate

$$\Pr(H_0) \sum_{m=0}^{M-1} \Pr(\Lambda^{(m)} \text{ is sent} | H_0) + \Pr(H_1) \sum_{m=0}^{M-1} \Pr(\Lambda^{(m)} \text{ is sent} | H_1) \leq \kappa,$$

for the Bayesian case (this case generally assumes that $\Pr(H_0) \approx \Pr(H_1)$), a constraint on the global rate under H_0

$$\sum_{m=0}^{M-1} \Pr(\Lambda^{(m)} \text{ is sent} | H_0) \leq \kappa,$$

for the Neyman-Pearson case (this case generally assumes that $\Pr(H_0) \gg \Pr(H_1)$), and either one of them for an Ali-Silvey distance. Under such a constraint, [62] shows that the optimal local decision rule is a censored local log-likelihood ratio $\Lambda^{(m)}$ where the censoring region consists of a single interval. More specifically, a radio will not send anything when $\eta_1^{(m)} \leq \Lambda^{(m)} < \eta_2^{(m)}$ and it will send $\Lambda^{(m)}$ otherwise. Furthermore, it is proven in [62] that if the communication rate constraint κ is sufficiently small and either $\Pr(H_1)$ (in the Bayesian framework) or the probability of false alarm constraint α (in the Neyman-Pearson framework) is small enough, then the optimal lower threshold $\eta_1^{(m)}$ is given by $\eta_1^{(m)} = 0$. This result has also been generalized in [63] for a communication rate constraint *per radio*, in which case the upper threshold $\eta_2^{(m)}$ can be directly determined from κ_m and no joint optimization of the set of upper thresholds $\{\eta_2^{(m)}\}_{m=0}^{M-1}$ is required.

In addition to communication rate constraints, other cost functions have been considered, such as the global cost of sensing and transmission:

$$C = \sum_{m=0}^{M-1} C_{s,m} + C_{t,m} \Pr(\Lambda^{(m)} \text{ is sent}),$$

where $C_{s,m}$ and $C_{t,m}$ are respectively the cost of sensing and transmission for cognitive radio m . Under a constraint on C , it can again be shown that the optimal local decision rule is a censored local log-likelihood ratio $\Lambda^{(m)}$ with a censoring region consisting of a single interval, where optimality can be in the Bayesian, Neyman-Pearson, or Ali-Silvey sense [64]. Furthermore, even if a digital transmission is considered, the optimal local decision rule is a quantized local log-likelihood ratio $\Lambda^{(m)}$, where every quantization level corresponds to a single interval and where one of the quantization levels is censored [64]. An extreme case of such a quantization is considered in [57] and [65] with only two quantization levels for $\Lambda^{(m)}$ (on/off signaling). A Bayesian approach is considered there with a communication rate constraint per

radio. Under different fading reporting channels, the optimal non-coherent combining rule and optimal local threshold have been determined in those papers.

In censored cooperative spectrum sensing, energy detection is often considered. In other words, the local decision is based on the locally collected energy $\Lambda^{(m)} = \|\mathbf{y}_m\|^2$, and the radio will not send anything when $\eta_1^{(m)} \leq \Lambda^{(m)} < \eta_2^{(m)}$. Outside this region, we can basically distinguish between two cases. When a soft decision rule is used at the fusion center, the radio will send $\Lambda^{(m)}$ when $\Lambda^{(m)} < \eta_1^{(m)}$ or $\Lambda^{(m)} \geq \eta_2^{(m)}$. When a hard decision rule is used on the other hand (such as the OR or AND rule), the radio will send a 0 when $\Lambda^{(m)} < \eta_1^{(m)}$ and a 1 when $\Lambda^{(m)} \geq \eta_2^{(m)}$. Such cases are investigated and analyzed in [66] for the hard decision OR rule and in [67] for the soft decision rule as well as the hard decision OR and AND rule. Note that [66] also takes reporting errors to the fusion sensor into account.

In addition to energy detection, autocorrelation-based and cyclostationarity detection have also been used in combination with censoring [22], [24]. These works consider a soft decision rule, under a Neyman-Pearson setting with a communication rate constraint per radio.

To conclude this subsection on censoring, note that censoring can also be combined with ordered transmissions to further improve the energy efficiency [68].

C. Sleeping

Sleeping or on/off sensing is a power saving mechanism in which every cognitive radio randomly turns off its sensing device with a probability μ , the sleeping rate. The advantage of sleeping over censoring is that the cognitive radios that are asleep do not waste any sensing or transmission power, whereas in censoring all the cognitive radios have to spend energy on sensing. Sleeping has generally been applied in combination with censoring [69], [70]. The combination of sleeping and censoring is studied in [69], with the goal of maximizing the mutual information between the state of signal occupancy and the decision state of the fusion center. In [70], sleeping is combined with the approach of [66] where energy detection and a hard decision OR rule is considered. More specifically, in [70], the global cost of sensing and transmission (C from above multiplied by $1 - \mu$) is optimized with respect to the sleeping rate μ and the thresholds $\eta_1^{(m)}$ and $\eta_2^{(m)}$, subject to a global probability of false alarm constraint $P_{\text{FA}} \leq \alpha$ and a global probability of detection constraint $P_{\text{D}} \geq \beta$. An interesting result from [70] is that the optimal lower threshold is again given by $\eta_1^{(m)} = 0$ if the feasible set is not empty.

D. Clustering

Finally, clustering has been proposed in networks to improve the energy efficiency [71], and it can easily be used in cognitive radio systems as well. Such an approach basically groups the cognitive radios into different clusters, where in each cluster a cluster head is assigned that reports to the fusion center (also more than two layers can be considered). For cooperative spectrum sensing specifically, this method reduces the average communication range to pass on information to the fusion center, and thus diminishes the average transmission energy, but it also allows for taking intermediate decisions about the presence or absence of the primary user (soft or hard) at the cluster heads [72], [73]. In [72], every cluster selects the radio with the best link to the fusion center as cluster head, in order to exploit selection diversity and improve the performance. In [73], confidence voting is proposed as a kind of censoring mechanism that can be used within every cluster to reduce the transmission energy even more. The idea is that a radio sends results to the cluster head only if it is confident, and it gains confidence when its result accords with the cluster consensus, and loses confidence otherwise.

VII. OTHER TOPICS AND OPEN PROBLEMS

We have reviewed some of the state-of-the-art methods and recent advances in spectrum sensing for cognitive radio. In doing so, we have necessarily had to make choices and cover only selected parts of existing work. There are several other topics worth mentioning, that also have been subject to recent research efforts:

- We have only considered spectrum sensing when the conditions are static, so that a primary signal is either present or absent. *Quickest detection* is a research area that addresses situations when the conditions are more dynamic. The problem of quickest detection is to detect the beginning of a primary user's transmission as quickly as possible after it happens. Similar issues with unknown parameters also occur in quickest detection problems, and tools such as the GLRT and marginalization that we have discussed here, can be used [74]. Likewise, collaboration can be applied to quickest detection problems [75]. A comprehensive treatment of quickest detection is provided in [76].
- *Adaptive sensing and learning* are other related topics that we did not treat. These topics also focus more on dynamic situations, and are an important part of the overall philosophy of cognitive radio.

Some recent work in this area is described in [77]. The problem of dynamic spectrum sensing (and channel access) as a partially observed Markov process is studied in [78]. The analysis of [78] considers cooperative dynamic spectrum sensing of primary channels whose occupancies are assumed to follow a Markovian evolution.

- *Joint spectrum sensing and efficient resource utilization* is a large field. Other examples where spectrum sensing and resource allocation are merged can be found in [79] and [80]. The optimization problems posed there are often non-convex and potentially NP-hard. Some formulations lead to *multi-armed bandit problems*. One example of such a multi-armed bandit problem is the allocation of M users to L ($> M$) available channels where the users may get different rewards (e.g. rate) for different channels [79]. Another example is the optimal selection of the sensing order when one out of L channels is sensed at a time [81].

In addition, there is a substantial amount of research being conducted on the specific topics that we have dealt with in this article. A key example is in the area of feature detection in which there is still considerable research. Exploitation of certain features in all dimensions (time, space and frequency) simultaneously is quite challenging, and leads to very complex detectors. Complexity is a major issue in many cases, for example when it comes to real-time implementations or energy saving. Taken together, while considerable research progress has been made on the problem of spectrum sensing, there are still many challenges and open problems to solve.

REFERENCES

- [1] FCC, "Spectrum policy task force report," Federal Communications Commission, Tech. Rep. 02-135, Nov. 2002, available: http://hraunfoss.fcc.gov/edocs_public/attachmatch/DOC-228542A1.pdf.
- [2] M. McHenry, "NSF spectrum occupancy measurements project summary," SSC, Tech. Rep., Aug. 2005, available: <http://www.sharespectrum.com/>.
- [3] S. Haykin, "Cognitive radio: Brain-empowered wireless communications," *IEEE J. Sel. Areas Commun.*, vol. 23, pp. 201–220, Feb. 2005.
- [4] I. J. Mitola, "Software radios: Survey, critical evaluation and future directions," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 8, pp. 25–36, Apr. 1993.
- [5] C. E. Shannon, "Computers and automata," *Proc. IRE*, vol. 41, pp. 1234–1241, Oct. 1953.
- [6] E. G. Larsson and M. Skoglund, "Cognitive radio in a frequency-planned environment: some basic limits," *IEEE Trans. Wireless Commun.*, vol. 7, pp. 4800–4806, Dec. 2008.

- [7] A. Sahai, N. Hoven, and R. Tandra, "Some fundamental limits on cognitive radio," in *Proc. 42nd Annual Allerton Conf. on Commun., Cont., and Comp.*, Monticello, Illinois, USA, Oct. 2004, pp. 1662–1671.
- [8] H. V. Poor, *An Introduction to Signal Detection and Estimation*. Springer Verlag, New York, 1994.
- [9] E. Axell and E. G. Larsson, "A Bayesian approach to spectrum sensing, denoising and anomaly detection," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Process. (ICASSP)*, Taipei, Taiwan, Apr. 2009, pp. 2333–2336.
- [10] R. Couillet and M. Debbah, "A Bayesian framework for collaborative multi-source signal sensing," *IEEE Trans. Signal Process.*, vol. 58, pp. 5186–5195, Oct. 2010.
- [11] H. Urkowitz, "Energy detection of unknown deterministic signals," *Proc. IEEE*, vol. 55, pp. 523–531, Apr. 1967.
- [12] E. Axell and E. G. Larsson, "Optimal and sub-optimal spectrum sensing of OFDM signals in known and unknown noise variance," *IEEE J. Sel. Areas Commun.*, vol. 29, pp. 290–304, Feb 2011.
- [13] A. Sonnenschein and P. Fishman, "Radiometric detection of spread-spectrum signals in noise of uncertain power," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 28, no. 3, pp. 654–660, Jul. 1992.
- [14] R. Tandra and A. Sahai, "SNR walls for signal detection," *IEEE J. Sel. Topics Signal Process.*, vol. 2, pp. 4–17, Feb. 2008.
- [15] E. G. Larsson and G. Regnoli, "Primary system detection for cognitive radio: Does small-scale fading help?" *IEEE Commun. Lett.*, vol. 11, no. 10, pp. 799–801, 2007.
- [16] S. Chaudhari, J. Lundén, V. Koivunen, and H. V. Poor, "BEP walls for hard decision-based cooperative sensing in cognitive radios," *IEEE Trans. Signal Process.*, 2011, undergoing review.
- [17] S. Chaudhari, V. Koivunen, and H. V. Poor, "Autocorrelation-based decentralized sequential detection of OFDM signals in cognitive radios," *IEEE Trans. Signal Process.*, vol. 57, pp. 2690–2700, Jul. 2009.
- [18] Huawei Technologies and UESTC, "Sensing scheme for DVB-T," Jul. 2006, IEEE Std.802.22-06/0127r1.
- [19] Z. Lei and F. Chin, "OFDM signal sensing for cognitive radios," in *PIMRC*, Sep. 2008, pp. 1–5.
- [20] W. A. Gardner, "Cyclostationarity: Half a century of research," *Signal Processing*, vol. 86, pp. 639–697, Apr. 2006.
- [21] A. V. Dandawaté and G. B. Giannakis, "Statistical tests for presence of cyclostationarity," *IEEE Trans. Signal Process.*, vol. 42, pp. 2355–2369, Sep. 1994.
- [22] J. Lundén, V. Koivunen, A. Huttunen, and H. V. Poor, "Collaborative cyclostationary spectrum sensing for cognitive radio systems," *IEEE Trans. Signal Process.*, vol. 57, pp. 4182–4195, Nov. 2009.
- [23] S. Enserink and D. Cochran, "A cyclostationary feature detector," in *Proc. IEEE Asilomar Conf. Signals, Systems, and Computers*, Pacific Grove, CA, Oct-Nov 1994, pp. 806–810.
- [24] S. Chaudhari, J. Lundén, and V. Koivunen, "Collaborative autocorrelation-based spectrum sensing of OFDM signals in cognitive radios," in *Proc. Annual Conf. on Information Sciences and Systems (CISS)*, Princeton, NJ, Mar. 2008, pp. 191–196.
- [25] P. Bianchi, M. Debbah, M. Maida, and J. Najim, "Performance of statistical tests for single source detection using random matrix theory," *IEEE Trans. Inf. Theory*, vol. 57, no. 4, pp. 2400–2419, Apr 2011.
- [26] P. Wang, J. Fang, N. Han, and H. Li, "Multiantenna-assisted spectrum sensing for cognitive radio," *IEEE Trans. Veh. Technol.*, vol. 59, pp. 1791–1800, May 2010.

- [27] A. Taherpour, M. Nasiri-Kenari, and S. Gazor, "Multiple antenna spectrum sensing in cognitive radios," *IEEE Trans. Wireless Commun.*, vol. 9, pp. 814–823, Feb. 2010.
- [28] E. Axell and E. G. Larsson, "Spectrum sensing of orthogonal space-time block coded signals with multiple receive antennas," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Process. (ICASSP)*, Dallas, TX, USA, Mar. 2010, pp. 3110–3113.
- [29] —, "A unified framework for GLRT-based spectrum sensing of signals with covariance matrices with known eigenvalue multiplicities," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Process. (ICASSP)*, Prague, Czech Republic, 2011.
- [30] Y. Zeng and Y.-C. Liang, "Eigenvalue-based spectrum sensing algorithms for cognitive radio," *IEEE Trans. Commun.*, vol. 57, pp. 1784–1793, Jun. 2009.
- [31] —, "Spectrum-sensing algorithms for cognitive radio based on statistical covariances," *IEEE Trans. Veh. Technol.*, vol. 58, no. 4, pp. 1804–1815, May 2009.
- [32] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Trans. Acoustics, Speech, and Signal Process.*, vol. 33, no. 2, pp. 387–392, Apr. 1985.
- [33] M. Chiani and M. Win, "Estimating the number of signals observed by multiple sensors," in *Proc. IEEE Int. Workshop on Cognitive Information Process. (CIP)*, Elba Island, Italy, Jun. 2010, pp. 156–161.
- [34] M. Haddad, A. M. Hayar, M. Debbah, and H. M. Fetoui, "Cognitive radio sensing information-theoretic criteria based," in *Proc. Int. Conf. on Cognitive Radio Oriented Wireless Networks and Commun. (CrownCom)*, Orlando, FL, USA, Aug. 2007, pp. 241–244.
- [35] R. Wang and M. Tao, "Blind spectrum sensing by information theoretic criteria for cognitive radios," *IEEE Trans. Veh. Technol.*, vol. 59, no. 8, pp. 3806–3817, Oct. 2010.
- [36] B. Farhang-Boroujeny, "Filter bank spectrum sensing for cognitive radios," *IEEE Trans. Signal Process.*, vol. 56, pp. 1801–1811, May 2008.
- [37] D. Thomson, "Spectrum estimation and harmonic analysis," *Proc. IEEE*, vol. 70, no. 9, pp. 1055 – 1096, Sep. 1982.
- [38] S. Haykin, D. Thomson, and J. Reed, "Spectrum sensing for cognitive radio," *Proc. IEEE*, vol. 97, pp. 849–877, May 2009.
- [39] D. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.
- [40] E. Candes and M. Wakin, "An introduction to compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 21–30, Mar. 2008.
- [41] K. Hossain and B. Champagne, "Wideband spectrum sensing for cognitive radios with correlated subband occupancy," *IEEE Signal Process. Lett.*, vol. 18, no. 1, pp. 35 –38, Jan. 2011.
- [42] Z. Quan, S. Cui, A. 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, Mar. 2009.
- [43] R. Fan, H. Jiang, Q. Guo, and Z. Zhang, "Joint optimal cooperative sensing and resource allocation in multi-channel cognitive radio networks," *IEEE Trans. Veh. Technol.*, vol. 60, no. 2, pp. 722–729, Feb. 2011.
- [44] Z. Tian and G. B. Giannakis, "Compressed sensing for wideband cognitive radios," in *Proc. IEEE Int. Conf. on Acoustics,*

- Speech, and Signal Process. (ICASSP)*, Honolulu, HI, USA, Apr. 2007, pp. 1357–1360.
- [45] S. Cotter, B. Rao, K. Engan, and K. Kreutz-Delgado, “Sparse solutions to linear inverse problems with multiple measurement vectors,” *IEEE Trans. Signal Process.*, vol. 53, no. 7, pp. 2477–2488, Jul. 2005.
- [46] M. Mishali and Y. Eldar, “From theory to practice: Sub-Nyquist sampling of sparse wideband analog signals,” *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 2, pp. 375–391, Apr. 2010.
- [47] Z. Tian, “Compressed wideband sensing in cooperative cognitive radio networks,” in *Proc. IEEE Global Telecommun. Conf. (GLOBECOM)*, Nov.–Dec. 2008, pp. 1–5.
- [48] Y. Wang, A. Pandharipande, Y. Polo, and G. Leus, “Distributed compressive wide-band spectrum sensing,” in *Proc. Information Theory and Applications Workshop*, Feb. 2009, pp. 178–183.
- [49] F. Zeng, C. Li, and Z. Tian, “Distributed compressive spectrum sensing in cooperative multihop cognitive networks,” *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 1, pp. 37–48, Feb. 2011.
- [50] J. Ma, G. Zhao, and Y. Li, “Soft combination and detection for cooperative spectrum sensing in cognitive radio networks,” *IEEE Trans. Wireless Commun.*, vol. 7, pp. 4502–4507, Nov. 2009.
- [51] S. M. Mishra, A. Sahai, and R. W. Brodersen, “Cooperative sensing among cognitive radios,” in *Proc. IEEE Int. Conf. on Commun. (ICC)*, Istanbul, Turkey, Jun. 2006, pp. 1658–1663.
- [52] Z. Quan, S. Cui, H. V. Poor, and A. Sayed, “Collaborative wideband sensing for cognitive radios,” *IEEE Signal Process. Mag.*, vol. 25, pp. 60–73, Nov. 2008.
- [53] W. Zhang, R. Mallik, and K. Letaief, “Optimization of cooperative spectrum sensing with energy detection in cognitive radio networks,” *IEEE Trans. Wireless Commun.*, vol. 8, no. 12, pp. 5761–5766, Dec. 2009.
- [54] B. Chen, R. Jiang, T. Kasetkasem, and P. Varshney, “Channel aware decision fusion in wireless sensor networks,” *IEEE Trans. Signal Process.*, vol. 52, no. 12, pp. 3454–3458, Dec. 2004.
- [55] Z. Chair and P. Varshney, “Optimal data fusion in multiple sensor detection systems,” *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-22, no. 1, pp. 98–101, Jan. 1986.
- [56] R. Niu, B. Chen, and P. Varshney, “Fusion of decisions transmitted over Rayleigh fading channels in wireless sensor networks,” *IEEE Trans. Signal Process.*, vol. 54, no. 3, pp. 1018–1027, Mar. 2006.
- [57] R. Jiang and B. Chen, “Fusion of censored decisions in wireless sensor networks,” *IEEE Trans. Wireless Commun.*, vol. 4, no. 6, pp. 2668–2673, Nov. 2005.
- [58] A. Wald, *Sequential Analysis*. John Wiley & Sons, New York, 1947.
- [59] K. W. Choi, W. S. Jeon, and D. G. Jeong, “Sequential detection of cyclostationary signal for cognitive radio systems,” *IEEE Trans. Wireless Commun.*, vol. 8, no. 9, pp. 4480–4485, Sep. 2009.
- [60] V. V. Veeravalli, T. Basar, and H. V. Poor, “Decentralized sequential detection with a fusion center performing the sequential test,” *IEEE Trans. Inf. Theory*, vol. 39, no. 2, pp. 433–442, Mar. 1993.
- [61] Q. Zou, S. Zheng, and A. Sayed, “Cooperative sensing via sequential detection,” *IEEE Trans. Signal Process.*, vol. 58, no. 12, pp. 6266–6283, Dec. 2010.

- [62] C. Rago, P. Willett, and Y. Bar-Shalom, "Censoring sensors: A low-communication-rate scheme for distributed detection," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 32, pp. 554–568, Apr. 1996.
- [63] S. Appadwedula, V. V. Veeravalli, and D. L. Jones, "Decentralized detection with censoring sensors," *IEEE Trans. Signal Process.*, vol. 56, pp. 1362–1373, Apr. 2008.
- [64] —, "Energy-efficient detection in sensor networks," *IEEE J. Sel. Areas Commun.*, vol. 23, pp. 693–702, Apr. 2005.
- [65] R. Jiang, Y. Lin, B. Chen, and B. Suter, "Distributed sensor censoring for detection in sensor networks under communication constraints," in *Proc. IEEE Asilomar Conf. Signals, Systems, and Computers*, Pacific Grove, CA, Oct.–Nov. 2005, pp. 946–950.
- [66] C. Sun, W. Zhang, and K. Ben, "Cooperative spectrum sensing for cognitive radios under bandwidth constraints," in *Proc. IEEE Wireless Commun. Network. Conf. (WCNC)*, Hong Kong, Mar. 2007, pp. 1–5.
- [67] Y. Chen, "Analytical performance of collaborative spectrum sensing using censored energy detection," *IEEE Trans. Wireless Commun.*, vol. 9, no. 12, pp. 3856–3865, Dec. 2010.
- [68] R. Blum and B. Sadler, "Energy efficient signal detection in sensor networks using ordered transmissions," *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 3229–3235, Jul. 2008.
- [69] K. Yamasaki and T. Ohtsuki, "Design of energy-efficient wireless sensor networks with censoring, on-off, and censoring and on-off sensors based on mutual information," in *Proc. IEEE Vehicular Technology Conf. (VTC)*, vol. 2, May–Jun. 2005, pp. 1312–1316.
- [70] S. Maleki, A. Pandharipande, and G. Leus, "Energy-efficient distributed spectrum sensing for cognitive sensor networks," *IEEE Sensors J.*, vol. 11, no. 3, pp. 565–573, Mar. 2011.
- [71] O. Younis, M. Krunz, and S. Ramasubramanian, "Node clustering in wireless sensor networks: recent developments and deployment challenges," *IEEE Network*, vol. 20, no. 3, pp. 20–25, May–Jun. 2006.
- [72] C. Sun, W. Zhang, and K. Ben, "Cluster-based cooperative spectrum sensing in cognitive radio systems," in *Proc. IEEE Int. Conf. on Commun. (ICC)*, Jun. 2007, pp. 2511–2515.
- [73] C.-H. Lee and W. Wolf, "Energy efficient techniques for cooperative spectrum sensing in cognitive radios," in *Proc. IEEE Consumer Communications and Networking Conference*, Las Vegas, NV, Jan. 2008, pp. 968–972.
- [74] L. Lai, Y. Fan, and H. V. Poor, "Quickest detection in cognitive radio: A sequential change detection framework," in *Proc. IEEE Global Telecommun. Conf. (GLOBECOM)*, New Orleans, LA, Nov.–Dec. 2008, pp. 1–5.
- [75] H. Li, H. Dai, and C. Li, "Collaborative quickest spectrum sensing via random broadcast in cognitive radio systems," *IEEE Trans. Wireless Commun.*, vol. 9, no. 7, pp. 2338–2348, Jul. 2010.
- [76] H. V. Poor and O. Hadjilias, *Quickest Detection*. Cambridge University Press, Cambridge, UK, 2009.
- [77] J. Lundén, V. Koivunen, S. R. Kulkarni, and H. V. Poor, "Reinforcement learning based distributed multiagent sensing policy for cognitive radio ad hoc networks," in *Proc. IEEE Int. Dynamic Spectrum Access Networks (DySPAN) symposium*, Aachen, Germany, May 2011.
- [78] J. Unnikrishnan and V. V. Veeravalli, "Algorithms for dynamic spectrum access with learning for cognitive radio," *IEEE Trans. Signal Process.*, vol. 58, no. 2, pp. 750–760, Feb. 2010.

- [79] K. Liu and Q. Zhao, "Distributed learning in multi-armed bandit with multiple players," *IEEE Trans. Signal Process.*, vol. 58, no. 11, pp. 5667–5681, Nov. 2010.
- [80] S.-J. Kim and G. B. Giannakis, "Sequential and cooperative sensing for multi-channel cognitive radios," *IEEE Trans. Signal Process.*, vol. 58, no. 8, pp. 4239–4253, Aug. 2010.
- [81] L. Lai, H. El Gamal, H. Jiang, and H. V. Poor, "Cognitive medium access: Exploration, exploitation, and competition," *IEEE Trans. Mobile Comput.*, vol. 10, no. 2, pp. 239–253, Feb. 2011.

Erik Axell (axell@isy.liu.se) is a Ph.D. student with the Communication Systems Division of the Department of Electrical Engineering, Linköping University, Sweden, since March 2006. He received his Lic. Eng. degree in December 2009 and his M.Sc. degree in 2005, from Linköping University, Sweden. From February 2005 to December 2005 he was with the Ericsson Research group in Linköping, Sweden. His main research interests are within the areas of wireless communications and signal processing, and in particular spectrum sensing and system aspects of cognitive radio.

Geert Leus (g.j.t.leus@tudelft.nl) is an Associate Professor at the Faculty of Electrical Engineering, Mathematics and Computer Science of the Delft University of Technology, The Netherlands. Geert Leus received a 2002 Young Author and a 2005 Best Paper Award from the IEEE Signal Processing Society. He was the Chair of the IEEE Signal Processing for Communications and Networking Technical Committee, and an Associate Editor of the IEEE Transactions on Signal Processing, IEEE Signal Processing Letters, and the IEEE Transactions on Wireless Communications. Currently, he is on the Editorial Board of the EURASIP Journal on Advances in Signal Processing.

Erik G. Larsson (erik.larsson@isy.liu.se) is Professor and Head of the Division for Communication Systems in the Department of Electrical Engineering (ISY) at Linköping University (LiU) in Linköping, Sweden. His main professional interests are wireless communications and signal processing. He has published some 70 journal papers on these topics, he is co-author of the textbook *Space-Time Block Coding for Wireless Communications* (Cambridge Univ. Press, 2003) and he has 10 patents on wireless technology. He is currently Associate Editor for the *IEEE Transactions on Communications*.

H. Vincent Poor (poor@princeton.edu) is the Dean of Engineering and Applied Science at Princeton University, where he is also the Michael Henry Strater University Professor of Electrical Engineering. His interests include the areas of statistical signal processing and stochastic analysis, with applications in wireless networks and related fields. Among his publications is the recent book *Quickest Detection* (Cambridge, 2009). Dr. Poor is an IEEE Fellow and a member of the NAE, the NAS and the RAE. Recent recognition includes the 2009 Armstrong Award of the IEEE Communications Society, the 2010 IET Fleming Medal and the 2011 IEEE Sumner Award.