JOURNAL OF COMMUNICATION AND INFORMATION SYSTEMS, VOL. 37, NO.1, 2022.

10

Spectrum Sensing: A Tutorial

Dayan Adionel Guimarães

Abstract—Spectrum sensing, combined or not with database information on radio-frequency (RF) spectrum occupation, is envisaged as part of the solution to the spectrum scarcity inherited by the fixed spectrum allocation policy currently adopted around the world. The solution is grounded on the premise of spectrum sharing between primary (incumbent) and secondary networks under a new dynamic spectrum access (DSA) paradigm. This tutorial presents basic concepts, fundamentals and state-of-the-art techniques related to spectrum sensing. Thought as a short course material, it covers from concepts regarding the forms of spectrum sensing and basic fundamentals on detection theory, to modern cooperative spectrum sensing techniques and research challenges, passing through the mathematical model for the signal and the sensing channel, the signal-to-noise ratio wall, practical issues regarding signal processing tasks, and performance metrics. A complete DSA framework is also addressed, which makes use of Internet of things devices equipped with spectrum sensing modules as a means to feed spectrum occupation databases.

Keywords—Cognitive radio, dynamic spectrum access, dynamic spectrum sharing, spectrum sensing.

I. INTRODUCTION

THE unprecedented demand for new telecommunications services has become the main research driver on new technologies, as can be noticed, for instance, in the recent advances related to the fifth generation (5G) of communication networks and to the Internet of things (IoT), as well as regarding the discussions and research already started about the sixth generation (6G) of these networks [1], [2].

To make many of the envisaged telecommunications services come true, especially in the case of wireless systems, the bottleneck of radio-frequency (RF) spectrum scarcity must be overcome. This scarcity is owed to the fact that the current fixed spectrum allocation policy grants to the incumbent (primary user, PU) network the exclusive usage right to frequency portions. Hence, with the growth of wireless communication systems in operation, only a few bands remain unused.

It is a widespread belief that the fixed spectrum allocation policy may not be capable of accommodating further growths in wireless communication systems. A new policy must be adopted, in which spectrum sharing becomes the novel paradigm. The associated technology is referred to as dynamic spectrum sharing (DSS) or dynamic spectrum access (DSA).

DSA makes use of the fact that many spectral bands already allocated to primary networks are underutilized [3],

Dayan A. Guimarães is a Researcher and Full Professor at the National Institute of Telecommunications (*Instituto Nacional de Telecomunicações*, Inatel), Santa Rita do Sapucaí, MG, Brazil. Phone: +55 35 3471-9227, e-mail: dayan@inatel.br.

[4], potentially allowing to be shared with secondary user (SU) networks.

In a shared-spectrum scenario, the SU transmissions can be made in two main ways: simultaneously to the PU transmissions, as long as no harmful interference is caused to the PU network, and in a non-interfering basis, opportunistically occupying unused licensed bands. The former is usually referred to as an underlay sharing approach, the latter is known as interweave approach.

The cognitive radio (CR) concept [5] arose in this spectrum sharing context. A CR transceiver is a device capable of acquiring cognition on the environment and network in which it is inserted, adapting its operational parameters to achieve optimized performance targets. Among the multitude of cognition-related attributes, a CR is also capable of identifying vacant bands for opportunistic access, applying a technique called spectrum sensing [1], [6]–[8].

Hence, in a first simplified definition, spectrum sensing is the technique that makes it possible the monitoring of the RF spectrum in search for vacant spectral bands, which are commonly referred to as spectrum holes or white-spaces.

A. Contribution and organization of the article

In this tutorial, basic concepts, fundamentals and state-of-the-art techniques related to spectrum sensing are addressed. The content herein is based on a graduate course material, and as such can be used as the reference text for a short course, for example spanning from 30 to 40 hours.

The tutorial covers from concepts regarding the forms of spectrum sensing and basic fundamentals on detection theory, to modern cooperative spectrum sensing techniques and research challenges, going through the mathematical model for the signal and the sensing channel, the signal-to-noise ratio wall concept, practical issues regarding signal processing tasks, and performance metrics. Several examples are given throughout the text, aiming at illustrating some theoretical content and helping the reader to understand more complex concepts. A complete DSA framework is also addressed as a closing material. The framework makes use of IoT devices equipped with spectrum sensors as a means to feed spectrum occupation databases for subsequent queries by a secondary network.

Owed to its particular scope and structure, this tutorial significantly differs from existing tutorials or surveys on spectrum sensing available in the literature (see, for instance, [1], [6]–[8] and references therein).

The remaining sections of the article are organized as follows. Section II addresses the basic initial concepts about spectrum sensing. Section III is devoted to notions on the signal detection theory applied to the design of test statistics

for spectrum sensing. The model for the signals and the sensing channel is covered in Section IV. Some well-know as well as recent test statistics are presented in Section V. Section VI addresses some signal processing aspects of moderns spectrum sensing techniques. The metrics commonly used to assess the spectrum sensing performance are covered in Section VII. Section VIII describes the idea behind a complete DSA framework in which a supporting IoT network is used as a means for sensing the spectrum and feed a database with spectrum occupation data. Challenges and research opportunities covering several aspects of the spectrum sensing context are discussed in Section IX. Section X summarizes the article and presents some final remarks.

B. Notation

Throughout the text, the notations x (italic lowercase letters), $\mathbf{x}$ (boldface lowercase letters) and $\mathbf{X}$ (boldface uppercase letters) represent a scalar, a vector and a matrix, respectively. The n-th element of a vector $\mathbf{x}$ is denoted by x_n. The element in the i-th row and j-th column of a matrix $\mathbf{X}$ is represented by $x_{i,j}$. The determinant and the trace of a matrix $\mathbf{X}$ are denoted by $\det(\mathbf{X})$ and $\text{tr}(\mathbf{X})$, respectively. The set of complex and real numbers are respectively denoted by $\mathbb{C}$ and $\mathbb{R}$. The absolute value of the scalar $x \in \mathbb{R}$ or the modulus of $x \in \mathbb{C}$ is represented by $|x|$. The smallest integer greater than or equal to x is denoted by $\lceil x \rceil$ (ceil operation). The largest integer smaller than or equal to x is denoted by $\lfloor x \rfloor$ (floor operation). The complex conjugate of $x \in \mathbb{C}$ is represented by x^*. The sets of matrices of order $x \times y$ having complex entries and real entries are respectively represented by $\mathbb{C}^{x \times y}$ and $\mathbb{R}^{x \times y}$. The operations of vector and matrix transposition are respectively denoted by $\mathbf{x}^{\text{T}}$ and $\mathbf{X}^{\text{T}}$. The Hermitian operator applied to the vector $\mathbf{x}$ or to the matrix $\mathbf{X}$ results in the transposition and complex conjugation of their elements, being represented by $\mathbf{x}^\dagger$ and $\mathbf{X}^\dagger$, respectively. The Euclidean norm of the vector $\mathbf{x}$ is denoted by $\|\mathbf{x}\| = \sqrt{\mathbf{x}^\dagger \mathbf{x}}$. The expected value (or expectation, or statistical average) of the random variable Z is represented by $\mathbb{E}\{Z\}$. The real and the imaginary parts of $r \in \mathbb{C}$ are represented by $\Re(r)$ and $\Im(r)$, respectively. Similarly, if the argument is a vector $\mathbf{x}$ or a matrix $\mathbf{X}$ with complex elements, $\Re(\mathbf{x})$ and $\Im(\mathbf{x})$, respectively $\Re(\mathbf{X})$ and $\Im(\mathbf{X})$, yield a vector (resp. matrix) formed by the real and the imaginary parts of the corresponding elements. The factorial of the scalar x is denoted by $x!$. A real (resp. complex) Gaussian random variable X with mean μ and variance σ^2 is denoted by $X \sim \mathcal{N}[\mu, \sigma^2]$ (resp. $X \sim \mathbb{C}\mathcal{N}[\mu, \sigma^2]$). A real-valued (resp. complex-valued) Gaussian random vector $\mathbf{y}$ with mean μ and covariance matrix Σ is denoted by $\mathbf{y} \sim \mathcal{N}[\mu, \Sigma]$ (resp. $\mathbf{y} \sim \mathbb{C}\mathcal{N}[\mu, \Sigma]$). A real uniform random variable X in the closed interval $[a, b]$ is denoted by $X \sim \mathcal{U}[a, b]$.

II. SPECTRUM SENSING

A. Basic concepts

Figure 1 depicts a didactic example of a primary network and a secondary network coexisting and having parts of their coverage areas in common. In this example, the primary network having two transmission (Tx) stations broadcasts

television (TV) signals to four TV set receivers (Rx). The secondary cognitive network contains a base station (BS) and three SUs connected to the BS via wireless links. The dashed lines represent the hypothetical coverage area limits. The example is consistent with the IEEE 802.22 standard [9], [10], which regulates the shared use of TV bands, in the context of wireless regional area networks (WRANs).
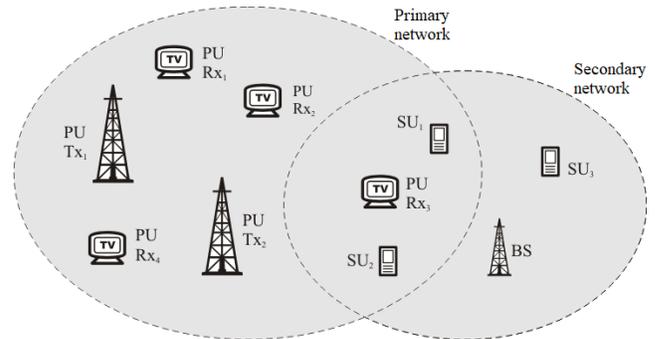
Fig. 1. Didactic example of a primary network and a secondary network coexisting and having a common coverage area.

The SUs shown in Figure 1 monitor the RF spectrum, seeking for unoccupied bands via spectrum sensing. In the example, SU_3 is beyond the reach of the primary network coverage and, as a consequence, it may assume that any sensed band is vacant. This situation is usually referred to as a hidden terminal problem. It potentially causes interference to the Rx_3 of the primary network, in case the secondary network decides to use that band while it is in use by the primary network. The terminals SU_1 and SU_2, though reachable by the primary network signal, may be incapable of detecting it, a situation that can happen, for instance, if the sensed signal is subjected to severe fading or is blocked by obstacles between the PU transmitters and the SU receivers. In this situation, SU_1 and SU_2 may start transmitting on a busy frequency band, producing interference in the primary network devices.

The hidden terminal problem illustrated by the situation of terminal SU_3, as well as multipath fading and signal shadowing[1] are capable of strongly penalizing the PU signal detection capability when independent spectrum sensing is made by each SU. To circumvent, or at least alleviate the hidden terminal problem and combat the deleterious effects of multipath fading and shadowing, a better choice in opposition to the independent spectrum sensing is the adoption of cooperative (or collaborative) spectrum sensing (CSS). As the name suggests, in CSS a group of SUs contribute to the decision on the occupation state of the sensed band, thus improving the PU detection capability.

---

[1] Multipath fading is the instantaneous variation of the received signal level due to changes, caused by the relative motion of the transmitter, the receiver or both, changing the condition of constructive and destructive combination of the signals arriving at the receive antenna through multiple propagation paths [11, p. 207]. Signal shadowing is caused by obstacles between the PU transmitters and the SU receivers, causing long-term signal level variations as the transmitter, the receiver or both move [11, p. 202].

B. Distributed and centralized CSS

There are two main forms of CSS: distributed and centralized. In distributed CSS, the SUs exchange spectrum sensing information (local decisions or received samples), then collectively make the global decision (by consensus, for example) on the occupation state of the sensed band. In centralized CSS, which is the focus of this tutorial, the spectrum sensing information gathered by the SUs in cooperation are transmitted to a fusion center (FC), which can be the BS of the secondary network or a special-purpose SU, where the global decision upon the sensed band state is made.

The exchange of sensing information among SUs (in the case of distributed CSS) or the transmissions from the SUs to the FC (in the case of centralized CSS) are made through low bandwidth dedicated control channels. In the latter situation, the control channels are sometimes referred to as report (or reporting) channels.

The global decision reached in the FC is subsequently broadcast to the SUs, also via control channels, and the access to the vacant band is made by means of some multiple access[2] technique appropriate to the secondary network.

The centralized CSS is classified according to how the spectrum sensing information is transmitted to and processed by the FC, which is referred to as fusion. A centralized CSS with decision fusion is realized if the local SUs decisions on the spectrum occupancy are sent to the FC. If the samples received by the SUs, or some quantity derived from these samples are sent to the FC, a centralized CSS with data fusion takes place [7].

In the case of decision fusion, a logical global decision rule is applied in the FC, called k-out-of-m rule. In this rule, the global decision is made in favor of a busy frequency band if at least k out of the m SUs in cooperation decide that the band is busy. Three special cases of the k-out-of-m rule are the OR rule (when $k = 1$), the AND rule (when $k = m$), and the majority voting (MAJ) rule (when $k = \lceil (m + 1)/2 \rceil$).

Both in data fusion, when the global decision is made by processing the quantities sent to the FC, and in decision fusion, when each local decision is made by processing the quantities gathered by each SU, the global decision and the local decisions, respectively, result from a binary hypothesis test. The test is made by comparing a decision variable (or test statistic) formed from the received samples (or from related quantities) with a decision threshold, as detailed in Section V.

C. Timing of the spectrum sensing

Fig. 2 depicts a typical frame structure of a secondary terminal having the spectrum sensing capability. Intervals are periodically allotted in each frame for spectrum sensing, for reporting the sensing information to the FC, for the decision and channel allocation processes, and for the secondary network data communication in the band considered vacant.

[2]Multiple access is a technique that allows multiple terminals share a given communication channel established in the frequency, time or code domains, combined or not. The most known multiple access techniques are the frequency division multiple access (FDMA), the time division multiple access (TDMA), and the code division multiple access (CDMA).

The total time dedicated to sensing and reporting, plus decision and channel allocation is much smaller than the one dedicated to data communication. On the other hand, the reporting and decision intervals are usually small compared to the sensing interval, except when adopting a large number of SUs in cooperation combined with time multiplexing of the reporting transmissions to the FC. In this case, the reporting interval can grow prohibitively because it has to accommodate the orthogonal transmissions of all cooperating SUs.
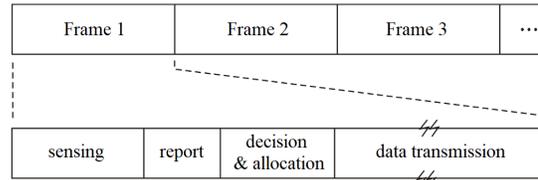
Fig. 2. A possible frame structure to coordinate the tasks of a SU terminal.

As formally explained later on in this tutorial, the performance of the spectrum sensing improves when the sensing interval is increased. This is an intuitively satisfying statement, since more samples of the received signal is collected, increasing the chance of a correct decision upon the presence of the PU signal. However, the increase of the sensing interval reduces the interval for data communication, also reducing the throughput of the secondary network. Thus, a trade-off solution must be adopted, giving to spectrum sensing the interval just enough for the target performance to be attained, aiming at saving resources for communication.

The timing process of the secondary network is not trivial, since it must guarantee that the sensing information from the various SUs reach the FC synchronized according to the adopted frame structure. However, this difficulty is not specific to CSS, since most wireless communication systems use frame structures that demand some kind of time coordination, especially when time division multiplexing (TDM) or time division duplexing (TDD) is applied.

III. NOTIONS OF SIGNAL DETECTION

A. Binary hypothesis test

Spectrum sensing is a binary hypothesis test in which are defined the null hypothesis, $\mathcal{H}_0$, denoting the absence of the PU signal in the sensed band (or an inactive PU transmitter), and the alternative hypothesis, $\mathcal{H}_1$, denoting the presence of the PU signal (or an active PU transmitter).

The main purpose of the detection process is to decide whether the received signal was generated under the $\mathcal{H}_0$ hypothesis or under the $\mathcal{H}_1$ hypothesis. To this end, a test statistic T is formed from the received signal samples $\mathbf{y}$ and compared with a decision threshold γ to decide in favor of $\mathcal{H}_1$ if $T > \gamma$, or $\mathcal{H}_0$ if $T < \gamma$.

In general, the binary hypothesis test theory postulates that the test statistic is actually used to decide whether the null hypothesis should be rejected or not. Thus, the decision made when comparing the test statistic with the decision threshold refers to a statement based on the null hypothesis: 'the null hypothesis is rejected' or 'the null hypothesis is not rejected'.

The alternative hypothesis (the claim) is considered valid if the null hypothesis (the counterclaim) is unlike to occur.

Associated with the decision is a reliability metric, commonly measured by the p-value [12]. In essence, this metric[3] quantifies the statistical significance of the evidence, which is the test statistic. For spectrum sensing, there are metrics more suitable than the p-value. These metrics are covered in Section VII.

B. Detector design criteria

The fundamental challenge in detector design is to develop the test statistic T and determine the decision threshold γ so that the desired spectrum sensing performance is achieved. Among several approaches to this design [13], [14], there are two main ones that deserve to be mentioned [14]: the classical approach and the Bayesian inference.

In the classical approach, the objective is to find T and γ so as to maximize the probability of detection, subjected to a limitation on the maximum probability of false alarm[4], giving rise to the so-called Neyman-Pearson criterion.

In Bayesian inference, costs (or weights) are assigned to each possible outcome of the decision process and, taking into account the prior probability of each hypothesis, the detector design criterion consists of minimizing the average cost of the possible decisions.

Despite the differences between the classical approach and the Bayesian inference, both can result in a test statistic that is formed by a ratio of two quantities [15], resulting in the test known as the likelihood ratio test (LRT), whose statistic is

$$T = \frac{p\left(\mathbf{y}|\mathcal{H}_1\right)}{p\left(\mathbf{y}|\mathcal{H}_0\right)}, \qquad (1)$$

or a scaling factor of it. In Bayesian inference, such a scaling factor is due to the presence of the weights and the prior probabilities of the hypotheses. In the above test statistic, $p\left(\mathbf{y}|\mathcal{H}_1\right)$ and $p\left(\mathbf{y}|\mathcal{H}_0\right)$ denote the probability density functions (PDFs) of the multidimensional (or multivariate) observed signal vector $\mathbf{y}$, respectively conditioned on the hypotheses $\mathcal{H}_1$ and $\mathcal{H}_0$.

The likelihood ratio is sometimes rewritten as the log-likelihood ratio (LLR)

$$T = \log\left[\frac{p\left(\mathbf{y}|\mathcal{H}_1\right)}{p\left(\mathbf{y}|\mathcal{H}_0\right)}\right]. \qquad (2)$$

The logarithm of the likelihood ratio does not change the decision rule, because the log function, whatever the basis, is monotonically increasing with the argument. However, the LLR is sometimes preferred because it facilitates the mathematical treatment of the resulting ratio, especially when it involves exponential functions (in this case, the logarithm in the natural base is used).

To compute the likelihood ratio T, the probability distribution of the observation $\mathbf{y}$ must be determined perfectly under both hypotheses, which means that all parameters involved in such a distribution must be known in advance. However, it is not uncommon for some of the parameters to be unknown, which sometimes makes it difficult to obtain optimal test statistics according to the aforementioned criteria.

Example 1: To illustrate the application of the Neyman-Pearson criterion in the design of a test statistic, consider that the vector $\mathbf{y}$ of samples received by an SU contains n signal samples plus noise, that is,

$$\mathbf{y} = \mathbf{x} + \mathbf{v}, \qquad (3)$$

where $\mathbf{x}$ e $\mathbf{v}$ are the vectors that represent the signal and the noise samples, respectively.

It is assumed that $\mathbf{v}$ is a vector of independent and identically distributed complex Gaussian random variables with zero mean and variance σ^2, that is, $\mathbf{v} \sim \mathbb{CN}[\mathbf{0}, \sigma^2\mathbf{I}]$, where $\mathbf{I}$ is the identity matrix of order $n \times n$ and $\sigma^2\mathbf{I}$ is the covariance[5] matrix of $\mathbf{v}$. Then, it follows that $\mathbf{y} = \mathbf{v}$ under the hypothesis $\mathcal{H}_0$, and $\mathbf{y} = \mathbf{x} + \mathbf{v}$ under $\mathcal{H}_1$.

Before going on, recall that the optimal Neyman-Pearson criterion consists of comparing an LLR, constructed from $\mathbf{y}$ according to (2), with a decision threshold.

It is assumed that the received signal does not have any structure that can be exploited to facilitate the detection process. In this case, it is considered that $\mathbf{x}$ is simply a vector of independent and identically distributed complex Gaussian random variables with zero mean and variance σ_x^2, that is, $\mathbf{y} \sim \mathbb{CN}[\mathbf{0}, \sigma^2\mathbf{I}]$ under $\mathcal{H}_0$, and $\mathbf{y} \sim \mathbb{CN}[\mathbf{0}, (\sigma^2 + \sigma_x^2)\mathbf{I}]$ under $\mathcal{H}_1$, with σ_x^2 denoting the variance of the signal samples.

It is known [16, p. 575] that the PDF of an n-dimensional random vector $\mathbf{z}$ having real Gaussian-distributed entries, mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, usually referred to as the multivariate Gaussian PDF, is given by

$$\begin{aligned}p(\mathbf{z}) &= p(z_1, z_2, \ldots, z_n) \\ &= (2\pi)^{-\frac{n}{2}} \det\left(\boldsymbol{\Sigma}\right)^{-\frac{1}{2}} e^{-\frac{1}{2}(\mathbf{z}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{z}-\boldsymbol{\mu})},\end{aligned} \qquad (4)$$

where $p(z_1, z_2, \ldots, z_n)$ denotes the joint PDF of the elements of $\mathbf{z}$. Since for the case in analysis $\boldsymbol{\Sigma} = \sigma^2\mathbf{I}$ or $\boldsymbol{\Sigma} = (\sigma^2 + \sigma_x^2)\mathbf{I}$, additionally knowing that the determinant of a diagonal matrix is the product of its diagonal elements, and that the inverse of a diagonal matrix is another diagonal matrix with elements equal to the reciprocal of the corresponding original elements [17],

---

[3]For details on the computation and interpretation of the p-value, besides [12], see also https://en.wikipedia.org/wiki/P-value.

[4]The probability of detection is the probability of declaring the PU signal present in the sensed band, when it is indeed present. The probability of false alarm is the probability of declaring the PU signal present in the sensed band, given that it is in fact absent. For more details on these probabilities, refer to the beginning of Section VII.

[5]A covariance matrix $\mathbf{R}$ is a square matrix that gives the covariance between each pair of elements of a random vector $\mathbf{y} = (y_1, y_2, \ldots, y_m)^{\mathrm{T}}$. The element of the i-th row and j-th column of $\mathbf{R}$ is $R_{i,j} = \mathrm{cov}[y_i, y_j] = \mathbb{E}[(y_i - \mathbb{E}[y_i])(y_j - \mathbb{E}[y_j])]$. If n realizations of the m-dimensional random vector are arranged in a matrix $\mathbf{Y}$, the covariance matrix is estimated as $\hat{\mathbf{R}} = \frac{1}{n}\mathbf{Y}\mathbf{Y}^{\mathrm{T}}$, which is called the sample covariance matrix. For more interpretations and explanations, see for example https://en.wikipedia.org/wiki/Covariance_matrix.

then the LLR is given by

$$
\begin{aligned}
T &= \ln\left[\frac{p\,(\mathbf{y}|\mathcal{H}_1)}{p\,(\mathbf{y}|\mathcal{H}_0)}\right] \\
&= \ln\left[\frac{\frac{1}{(2\pi)^{n/2}(\sigma^2+\sigma_{\mathrm{x}}^2)^{n/2}}\exp\left(-\frac{\|\mathbf{y}\|^2}{2(\sigma^2+\sigma_{\mathrm{x}}^2)}\right)}{\frac{1}{(2\pi)^{n/2}\sigma^n}\exp\left(-\frac{\|\mathbf{y}\|^2}{2\sigma^2}\right)}\right].
\end{aligned} \tag{5}
$$

Applying $\ln(a/b) = \ln(a) - \ln(b)$ and noticing that the multiplicative constants of the exponentials are independent of $\mathbf{y}$, it follows that

$$
\begin{aligned}
T &\propto \frac{\|\mathbf{y}\|^2}{2\sigma^2} - \frac{\|\mathbf{y}\|^2}{2(\sigma^2+\sigma_{\mathrm{x}}^2)} \\
&= \frac{1}{2}\left(\frac{1}{\sigma^2} - \frac{1}{\sigma^2+\sigma_{\mathrm{x}}^2}\right)\|\mathbf{y}\|^2,
\end{aligned} \tag{6}
$$

meaning that the resulting test statistic is proportional to $\|\mathbf{y}\|^2$, which is the energy of the samples that form $\mathbf{y}$. Since the proportionality constant is independent of $\mathbf{y}$, then

$$
\begin{aligned}
T &\propto \|\mathbf{y}\|^2 \\
&= \mathbf{y}^\dagger\mathbf{y} \\
&= \sum_{i=1}^{n} y_i^* y_i \\
&= \sum_{i=1}^{n} |y_i|^2.
\end{aligned} \tag{7}
$$

As any proportionality constant applied to a test statistic does not affect the performance of the hypothesis test, but only changes the decision threshold in the same proportion, it is convenient for the present analysis to apply the scale factor $1/n$ to the test statistic, yielding

$$
T = \frac{1}{n}\sum_{i=1}^{n}|y_i|^2. \tag{8}
$$

The test statistic (7) or, equivalently, (8), refers to the well-known energy detector (ED), which is therefore an optimal detector[6] according to the Neyman-Pearson criterion, under the conditions imposed by the signal model considered in the present example.

It is noteworthy that the determination of the decision threshold with which T is compared presupposes that the noise variance is known, so that the energies of $\mathbf{y}$ under the hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$ are discriminated. This is a major limitation of the energy detector, as explored later.

Mathematical analyzes of the energy detector are relatively simple due to the fact that the distributions of T are known under both hypotheses: for the model at hand, one has the sum of $2n$ zero-mean, squared Gaussian random variables (recall that $\mathbf{y}$ is composed of complex random variables). Then, the variable T has a central chi-square distribution [18] under both hypotheses.

---

[6]A detector considered optimal under a given design criterion may have lower performance with respect to another optimal detector developed under a different design criterion. Therefore, when one says that a detector is optimal, it means that there is no other better detector under the same design criterion, which does not prevent the existence of a detector that performs better when designed under another criterion.

One of the biggest obstacles to the development of test statistics is to determine the PDFs operated in the likelihood ratio, not only due to lack of knowledge of some parameter, but also when the observation is a quantity resulting from some operation on the values of the received signal samples, that may make the mathematical treatment of the resulting test statistic difficult. In other words, when the collected samples are processed to generate other quantities from which the likelihood ratio is defined, the analytical development of the test statistic may become extremely cumbersome or, in some cases, even prohibitive.

When some parameters are not known in advance, one of the strategies typically used in the classical approach to hypothesis testing is to estimate them. A standard technique is to perform maximum likelihood estimates, which replaces the actual parameters in the likelihood ratio. This strategy gives rise to the so-called generalized likelihood ratio test (GLRT). In summary, the GLRT is a likelihood ratio test, LRT, in which unknown parameters are replaced by their maximum likelihood estimates [14, p. 430].

It is worth stating that a model is always a simplification of reality. Consequently, probability distributions are never or rarely perfectly known. In other terms, even if the model is consistent with reality, there may still be parameters that are unknown or uncertain, for example the noise power, the signal strength and the coefficients (gains) of the sensing channel.

Fortunately, sometimes there is an escape from such design limitations. For example, a detector can be developed based on some existing discrepancy metric adopted in a different context. Indeed, this was the approach used to design the detectors proposed in [19], [20] and [21], which is also explored in this tutorial, specifically in Section V.

C. Signal-to-noise ratio wall

Cognitive secondary terminals must be able to detect very weak primary user signals. This is difficult as there are fundamental limits to the detection at low signal-to-noise ratio (SNR). Specifically, due to uncertainties about model parameters, accurate detection is impossible below a certain level of signal-to-noise ratio known as the SNR wall [22].

The SNR wall, which can be denoted by $\mathrm{SNR_w}$, is the upper limit below or equal to which it is impossible to control the probability of detection, P_{d}, and the probability of false alarm, P_{fa}, to yield $P_{\mathrm{fa}} \approx 0$ and $P_{\mathrm{d}} \approx 1$, or at least so that these probabilities can reach any desired values within their useful limits, which are $0 \leq P_{\mathrm{fa}} \leq 0.5$ and $0.5 \leq P_{\mathrm{d}} \leq 1$. Such control is performed through the number of samples, n, meaning that, if $\mathrm{SNR} \leq \mathrm{SNR_w}$, the increase of n does not bring improvement of the spectrum sensing performance.

Example 2: To better understand the $\mathrm{SNR_w}$, firstly it is convenient to interpret that, in the case of the energy detector, the objective is to discriminate the values σ^2 and $(\sigma^2+\sigma_{\mathrm{x}}^2)$, which are the powers of the received signal under the hypothesis $\mathcal{H}_0$ and $\mathcal{H}_1$, respectively. This conclusion can be easily drawn from (8), as explained in the sequel.

Observing the test statistic defined in (8), it can be noticed that it is nothing more than an estimate of the average power P of the received signal, since, by definition,

$$P = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} |y_i|^2. \qquad (9)$$

However, from (8) one realizes that T is a random variable due to the randomness of y_i. Hence, for sufficiently high n, the test statistic T approximates a Gaussian random variable, according to the central limit theorem (CLT). The mean of this random variable is σ^2 under $\mathcal{H}_0$ and $(\sigma^2 + \sigma_x^2)$ under $\mathcal{H}_1$, with variances inversely proportional to n, as detailed below.

Since the samples in $\mathbf{y}$ are mutually independent, due to the mutually independent samples in $\mathbf{x}$ and $\mathbf{v}$, the expectation of the ED test statistic given in (8) is $\mathbb{E}[T] = (1/n)n\mathbb{E}[|y_i|^2] = \mathbb{E}[|y_i|^2] = \mathbb{E}[y_i^* y_i] = \mathbb{E}\{[\mathfrak{R}(y_i)]^2 + [\mathfrak{I}(y_i)]^2\} = \mathbb{E}\{[\mathfrak{R}(y_i)]^2\} + \mathbb{E}\{[\mathfrak{I}(y_i)]^2\}$, whose result is $\mathbb{E}[T] = \sigma^2/2 + \sigma^2/2 = \sigma^2$ under $\mathcal{H}_0$, and $\mathbb{E}[T] = (\sigma^2 + \sigma_x^2)/2 + (\sigma^2 + \sigma_x^2)/2 = \sigma^2 + \sigma_x^2$ under $\mathcal{H}_1$. The variance is $\text{var}[T] = (1/n^2)n\text{var}[|y_i|^2]$. Taking into account that the variances of $[\mathfrak{R}(y_i)]^2$ and $[\mathfrak{I}(y_i)]^2$ are the same, and denoting $\mathfrak{R}(y_i)$ by X, it follows that $\text{var}[T] = (2/n)\text{var}[X^2] = (2/n)\{\mathbb{E}[X^4] - \mathbb{E}^2[X^2]\}$, where it has been used the fact that the variance of a random variable is its second moment minus its squared first moment. Moreover, the fourth central moment [18, p. 144] of a Gaussian random variable whose variance (second central moment) is $\sigma^2/2$ is $3\sigma^4/4$. Then, one obtains $\text{var}[T] = \sigma^4/n$ under $\mathcal{H}_0$, and $\text{var}[T] = (\sigma^4 + \sigma_x^4)/n$ under $\mathcal{H}_1$.

Fig. 3 illustrates the PDFs of T under both hypotheses, for an arbitrary high value of n and arbitrary values of signal and noise variances. This figure shows that it is possible to set the decision threshold somewhere in-between the PDFs so as to result in an almost perfect spectrum sensing, with $P_{\text{fa}} \approx 0$ and $P_{\text{d}} \approx 1$, because σ^2 is easily distinguishable from $\sigma^2 + \sigma_x^2$.

It is also clear from Fig. 3 that the reduction of the signal-to-noise ratio σ_x^2/σ^2 corresponds to the approximation of the means of the two PDFs.
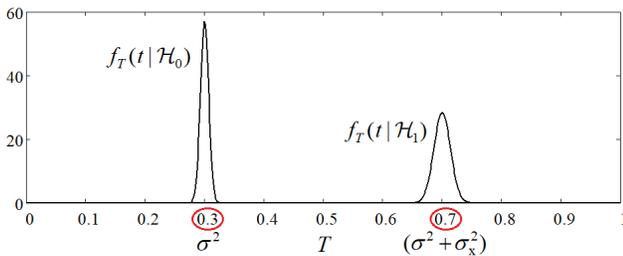
Fig. 3. Distributions of the test statistic T defined in (8) for a large n.

In the limit of $n \to \infty$, the variances of both PDFs tend to zero, still allowing the averages of T under $\mathcal{H}_0$ and $\mathcal{H}_1$ to be discriminated even if $\sigma_x^2/\sigma^2 \to 0$, which means $\text{SNR}_\text{w} \to 0$. However, the variance of the noise, σ^2, must be perfectly known, in such a way that the decision threshold is set with a value infinitesimally greater than such variance, thus making it possible to distinguish σ^2 from $\sigma^2 + \sigma_x^2$ with accuracy.

Now assume that σ^2 may not be known perfectly, with its value having an uncertainty between σ^2/ρ and $\rho\sigma^2$, with

$\rho \geq 1$. In other words, the estimated variance, $\hat{\sigma}^2$, to be used in the computation of the decision threshold lies in-between σ^2/ρ and $\rho\sigma^2$. Such uncertainty is mainly due to the fact that, in practice, when the noise variance is not known, it is estimated and, therefore, carries an inherent estimation error.

On the other hand, even if σ^2 is accurately determined in the detector design phase, and the decision threshold is set using such information, variations in σ^2 may occur due to natural miscalibration among receivers or due to unwanted signals entering the receiver as if they were noise.

The noise uncertainty has the potential to deteriorate the spectrum sensing performance, as the estimated noise variance, $\hat{\sigma}^2$, when applied in the computation of the energy detector test statistic, divides the right-side[7] of (8), changing the averages of this statistic under $\mathcal{H}_0$ and $\mathcal{H}_1$. The worst-case situations correspond to $\hat{\sigma}^2 = \sigma^2/\rho$ under the hypothesis $\mathcal{H}_0$ and $\hat{\sigma}^2 = \rho\sigma^2$ under $\mathcal{H}_1$, cases in which the test statistic has its mean increased under $\mathcal{H}_0$ and decreased under $\mathcal{H}_1$, meaning that the conditional PDFs of T become close to each other.

Taking the previous paragraph and Fig. 3 as references, assume that, in the limit, the signal-to-noise ratio σ_x^2/σ^2 is such that the means of the test statistic under $\mathcal{H}_0$ and $\mathcal{H}_1$ are the same. In light of the test statistic given in the footnote, it follows that $\sigma^2/\hat{\sigma}^2 = (\sigma^2 + \sigma_x^2)/\hat{\sigma}^2$. In this case, reliable detection is impossible even with $n \to \infty$, as it becomes impossible to set the decision threshold to have control over the false alarm and detection probabilities. Thus, the corresponding SNR is the SNR wall, whose value can be determined by operating with the above-mentioned equality under the worst-case situations, that is, $\sigma^2/\hat{\sigma}^2 = (\sigma^2 + \sigma_x^2)/\hat{\sigma}^2 \Rightarrow \sigma^2/(\sigma^2/\rho) = (\sigma^2 + \sigma_x^2)/(\rho\sigma^2) \Rightarrow \rho = (1 + \sigma_x^2/\sigma^2)/\rho \Rightarrow \sigma_x^2/\sigma^2 = \rho^2 - 1$.

Hence, for the energy detector with uncertainty between σ^2/ρ and $\rho\sigma^2$ in the noise variance, the signal-to-noise ratio wall is given by

$$\text{SNR}_\text{w} = \rho^2 - 1. \qquad (10)$$

Note that if there is no uncertainty in the noise variance, in which case $\rho = 1$, it follows that $\text{SNR}_\text{w} = 0 = -\infty$ dB, as expected.

The method explained in the previous example is intuitively satisfying to determine the SNR wall, and the reasoning behind it is formally grounded on a theorem from [22], stating that the existence of an SNR wall below which every detector is not capable of meeting useful performance metrics requires the test statistics to have overlapping medians under the two hypotheses. In the case of the ED, the medians are equal to the corresponding means, due to the Gaussian approximation of the distributions of its test statistic for sufficiently large n.

In fact, the method of searching for the SNR in which the medians of the test statistic under both hypotheses become the same, for sufficiently large number of samples, in thesis

---

[7]The same performance deterioration occurs if the noise uncertainty is applied to the computation of the decision threshold, that is, if the noise variance is embedded into this threshold. However, didactically it is more appropriate to consider that the noise variance is moved to the test statistic, dividing the right-side of (8), that is, $T = \frac{1}{n\hat{\sigma}^2} \sum_{i=1}^{n} |y_i|^2$.

can be applied to any detector, as indicated by the above-mentioned theorem. Another indication of this possibility can be seen in [23], where a coarse empirical version of the method has been applied to show the existence of such wall for the maximum-minimum eigenvalue (MME) detector.

IV. SIGNAL AND CHANNEL MODELS

A. Modeling for CSS with data fusion

The model for centralized CSS considered herein assumes m spectrum sensors, which suits to m SUs equipped with a single antenna each, or a single SU equipped with m antennas. Each sensor collects n complex samples[8] of the signal transmitted by s primary transmitters during each sensing interval. In CSS with data fusion, such samples form, at the fusion center, the matrix $\mathbf{Y} \in \mathbb{C}^{m \times n}$ given by

$$\mathbf{Y} = \mathbf{HX} + \mathbf{V}. \tag{11}$$

Under $\mathcal{H}_0$ the PU signal is absent in the sensed band, that is, $\mathbf{Y} = \mathbf{V}$. Under $\mathcal{H}_1$ the PU signal is present, that is, $\mathbf{Y} = \mathbf{HX} + \mathbf{V}$.

In (11), the samples associated to the signals transmitted by the s PUs are arranged in the matrix $\mathbf{X} \in \mathbb{C}^{s \times n}$. In order to model the envelope fluctuations of modulated and filtered signals [25], such samples are complex Gaussian random variables with zero mean and variance dependent on the average SNR at the input of the SUs receivers. Another option that can be considered is to adopt complex samples to model, for example, the signal of a baseband quaternary phase-shift keying (QPSK) modulation. In this case, it is common to set the number of samples per QPSK symbol as a fraction of the total number of samples n, so that the temporal correlation between samples of the modulated signal is controlled. This second option is particularly useful for analyzing the performance of detection techniques that assume that there is such temporal correlation to work properly, as is the case of the family of detectors based on the cooperative power spectral density split cancellation (CPSC) method [26].

The channel matrix $\mathbf{H} \in \mathbb{C}^{m \times s}$ in (11) is formed by elements $h_{i,j}$, $i = 1, 2, \ldots, m$, $j = 1, 2, \ldots, s$ that represent the gains of the sensing channels between the j-th PU and the i-th SU. The variation of these gains can model the fading effect produced in the signal due to multipath propagation of the electromagnetic wave, and other variations related to signal propagation, such as distance-dependent attenuation and signal shadowing due to blockage.

In the context of spectrum sensing, channel gains are normally considered constant during the sensing interval, being independent and identically distributed among consecutive sensing realizations. Constant gains mean that the sensing duration is smaller than the coherence time[9] of the sensing

channel, that is, the fading produced in the channel is considered slow during the sensing interval. The independence between successive sensing rounds means that the interval between them is greater than the channel coherence time. Both situations are quite plausible in practice. Moreover, the multiplication between $\mathbf{H}$ and $\mathbf{X}$ in (11) is meant to represent a flat fading channel [11, p. 222], meaning that the sensed signal bandwidth is smaller than the coherence bandwidth[10] of the multipath fading channel.

In the present model, the channel matrix is given by

$$\mathbf{H} = \mathbf{GA}, \tag{12}$$

where $\mathbf{G} \in \mathbb{R}^{m \times m}$ controls the powers of the signals received by the SUs, as detailed a little ahead, and $\mathbf{A} \in \mathbb{C}^{m \times s}$ models a multipath Ricean fading[11].

Matrix $\mathbf{A}$ is formed by elements $a_{i,j}$ that are complex Gaussian random variables with mean $\sqrt{\kappa_{i,j}/(2\kappa_{i,j}+2)}$ and variance $1/(\kappa_{i,j}+1)$, that is, $a_{i,j} \sim \mathbb{CN}[\sqrt{\kappa_{i,j}/(2\kappa_{i,j}+2)}, 1/(\kappa_{i,j}+1)]$, such that $\mathbb{E}\{|a_{i,j}|^2\} = 1$, with $\kappa_{i,j}$ being the Rice factor of the channel between the j-th PU and the i-th SU. The magnitude of $a_{i,j}$, i.e. $|a_{i,j}|$, represents the voltage gain of the channel between the j-th PU and the i-th SU. The unitary value for the second moment of this magnitude, i.e. $\mathbb{E}\{|a_{i,j}|^2\}$, is intended to model unit power gain, without losing the generality of the model, as other power gains could be used. However, there is no need of using $\mathbb{E}\{|a_{i,j}|^2\} \neq 1$, because different power gains are already being modeled by $\mathbf{G}$.

The values of the mean μ_κ and variance σ_κ^2 of both the real and imaginary parts of $a_{i,j}$ are obtained by applying the equalities: $\mathbb{E}\{|a_{i,j}|^2\} = 2\mu_\kappa^2 + 2\sigma_\kappa^2$ and $\kappa_{i,j} = \mu_\kappa^2/\sigma_\kappa^2$. Alternatively, one can model the Rice fading by applying the mean μ_κ only to the real part or to the imaginary part of $a_{i,j}$. In this case, the following relations are used: $\mathbb{E}\{|a_{i,j}|^2\} = \mu_\kappa^2 + 2\sigma_\kappa^2$ and $\kappa_{i,j} = \mu_\kappa^2/2\sigma_\kappa^2$.

In a channel with Ricean fading, the Rice factor is the ratio of the signal strength received via the dominant propagation path, for example in a line-of-sight (LoS) condition or specular reflection, and the power contained in the signals from the other paths. A larger Rice factor means less variability (less fading) in the instantaneous received signal strength. A null Rice factor means the absence of a dominant path, which corresponds to a channel with Rayleigh fading. In practical terms, a Rice factor greater than 10 already makes the channel approximately free of fading.

---

[8]The adoption of complex samples aligns with the usual quadrature receiver structure, in which the received signal is converted into baseband by quadrature carriers (sine and cosine functions). The resulting baseband signals compose the real and imaginary parts of the sampled signal [24].

[9]The coherence time is the time interval during which successively observed channel impulse responses exhibit high correlation. Alternatively, the coherence time is the interval during which the channel practically does not change its response [11, p. 223].

[10]The coherence bandwidth is a reference bandwidth value that determines if the signal undergoes frequency flat fading (approximately the same time-varying channel gain applied to all frequency components of the signal) or frequency selective fading (possibly different time-varying channel gains over the frequency components of the signal). The coherence bandwidth is inversely proportional to the time dispersion produced by the channel due to the echo-like effect of the multipath propagation [11, p. 221].

[11]The Ricean (or Rice-type) fading is a short-term signal level variation that occurs due to the already-mentioned interfering behavior of signals received via multiple paths. In this type of fading there is a level dominance of some path or cluster of paths with respect to the remaining ones, yielding less signal variations in comparison with the Rayleigh fading. In the latter, such domination does not exist, i.e. the signals coming from all paths and all directions have the same average level [11, p. 212].

In the model described herein, realistic sensing channels are taken into account by means of adopting realistic Rice factors $\kappa_{i,j} = 10^{K_{i,j}/10}$, being $K_{i,j}$ a Gaussian random variable with mean μ_K and standard deviation σ_K, in dB, that is, $K_{i,j} \sim \mathcal{N}[\mu_K, \sigma_K^2]$. According to a study published in [27], typical values of such mean and standard deviation are $\mu_K = 1.88$ dB and $\sigma_K = 4.13$ dB for urban areas, $\mu_K = 2.41$ dB and $\sigma_K = 3.84$ dB for suburban areas, and $\mu_K = 2.63$ dB and $\sigma_K = 3.82$ dB for rural or open areas.

This model for the Ricean fading is consistent with the fact that a moving SU experiences, along its path, different situations in terms of the LoS degree with the primary transmitter. Urban areas typically unveil more and larger obstacles between PUs and SUs, with varied characteristics, reducing the mean Rice factor and increasing its variability, which is measured by the standard deviation σ_K. On the other hand, rural or open areas tend to exhibit higher Rice factors, on average, with lower variability, due to the smaller quantity and lesser diversity of obstacles typically found in these regions.

When the possibility of inequality between the power levels of the signals received by the PUs is admitted, the matrix $\mathbf{G} \in \mathbb{R}^{m \times m}$ present in (12) is given by

$$\mathbf{G} = \operatorname{diag}\left(\sqrt{\frac{\mathbf{p}}{p_{\text{avg}}}}\right), \qquad (13)$$

where $\mathbf{p} = [p_1, p_2, \ldots, p_m]$ is the vector containing the power levels of the signals received by the SUs, and

$$p_{\text{avg}} = \frac{1}{m} \sum_{i=1}^{m} p_i \qquad (14)$$

is the average power of these signals. Since the average power gain of the channel is unitary by hypothesis and without loss of generality, also without loss of generality it is considered that each PU transmits with a constant power p_{avg}/s. Additionally, when the received signal powers are time-varying, at each sensing round the values of p_i are uniformly distributed in the range $[(1 - \rho_S)p_{\text{avg}}, (1 + \rho_S)p_{\text{avg}}]$, that is,

$$p_i \sim \mathcal{U}[(1 - \rho_S)p_{\text{avg}}, (1 + \rho_S)p_{\text{avg}}], \qquad (15)$$

where ρ_S, $0 \leq \rho_S < 1$, is the fraction of variation intended for the received signal power around the average.

The variation of the received signal power levels across the SUs is consistent with the signal level variations experienced by moving SUs, mainly due to signal shadowing caused by obstacles, and distance-dependent signal attenuation.

When taking into account the possibility of temporal variation of the power levels of the additive white Gaussian noise (AWGN) in the PUs, the i-th row of the matrix $\mathbf{V} \in \mathbb{C}^{m \times n}$ in (11) is composed of zero-mean Gaussian random variables with variance uniformly distributed in the interval $[(1 - \rho_N)\sigma_{\text{avg}}^2, (1 + \rho_N)\sigma_{\text{avg}}^2]$, that is,

$$\sigma_i^2 \sim \mathcal{U}[(1 - \rho_N)\sigma_{\text{avg}}^2, (1 + \rho_N)\sigma_{\text{avg}}^2] \qquad (16)$$

in each realization of the spectrum sensing, with

$$\sigma_{\text{avg}}^2 = \frac{1}{m} \sum_{i=1}^{m} \sigma_i^2 \qquad (17)$$

denoting the average noise power in the SUs, and ρ_N, $0 \leq \rho_N < 1$ denoting the desired fraction of variation for the noise power around the average.

Noise power variations can occur because of variations in the ambient temperatures to which the SUs are subjected, as well as owed to uncalibrated receiver front-end circuits or uneven gains and noise figures among low noise amplifiers (LNAs), or even due to the presence of unwanted signals in the sensed band, which can be considered as background noise added to thermal noise.

Finally, in the present model it follows that the average SNR, in dB, over the SUs, is given by

$$\text{SNR} = 10 \log_{10}\left(\frac{p_{\text{avg}}}{\sigma_{\text{avg}}^2}\right). \qquad (18)$$

B. Modeling for CSS with decision fusion

In centralized CSS with data fusion, it is known that the i-th row of the matrix $\mathbf{Y}$ generated in the i-th SU is sent to the fusion center, where $\mathbf{Y}$ is formed. Subsequently, the samples of $\mathbf{Y}$ are processed to give rise to the desired test statistic, according to the adopted detection technique. On the other hand, in centralized CSS with decision fusion, each row of the matrix $\mathbf{Y}$ is operated in each SU to generate a test statistic to allow the local decision on the occupancy state of the sensed band. The decisions of all SUs are then transmitted to the FC, where they are logically combined to yield the global decision. Note that when the decision fusion takes place, the matrix $\mathbf{Y}$ is not defined as in (11). In this case, the i-th SU forms its own test statistic by processing the sample vector

$$\mathbf{y}_i^{\mathsf{T}} = (\mathbf{h}_i \mathbf{X})^{\mathsf{T}} + \mathbf{v}_i^{\mathsf{T}}. \qquad (19)$$

where $\mathbf{y}_i$, $\mathbf{v}_i$ and $\mathbf{h}_i$ correspond to the i-th row of the matrices $\mathbf{Y}$, $\mathbf{V}$ and $\mathbf{H}$, respectively.

To illustrate the spectrum sensing scenario covered by the model discussed throughout this section, Fig. 4 highlights the sensing channel between $s = 2$ primary transmitters and $m = 3$ secondary receivers, as well as the control channel used for transmitting the decisions made by the secondary receivers to the fusion center of the secondary network (in the case of decision fusion), or for transmitting the samples collected by them, or quantities derived from these samples, to the fusion center (in the case of data fusion).
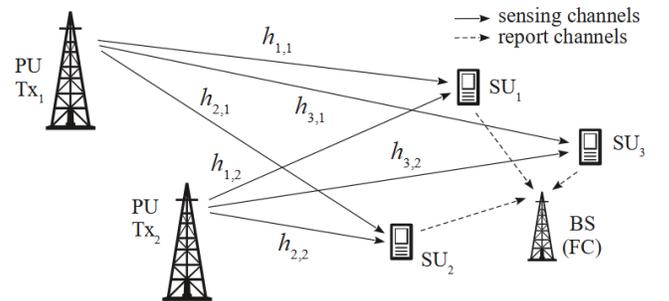
Fig. 4. Spectrum sensing scenario covered by the model discussed throughout this Section IV.

V. Some test statistics

There are several detection techniques for spectrum sensing, the most famous among them being the ED, the detection by matched filtering, the cyclostationary feature detection, and the eigenvalue-based detection [7], [28], [29].

ED is often adopted as a reference in studies on spectrum sensing because of its simple implementation and satisfactory performance in many situations. However, it has the big disadvantage of requiring knowledge of the noise power at the input of each receiver so that the appropriate decision threshold is determined. Furthermore, the ED is optimal only when samples of the primary signals are uncorrelated over time, as is the case when considering that such signals are represented by white noise samples. Additionally, the performance of the energy detector is affected by the inherent uncertainty about the noise power that is used in composing its test statistic or in determining the decision threshold, as discussed in Section III-C.

Matched filter detection is considered optimal, but it needs to know information about the waveform of the primary signal pulses, as well as the channel between PUs and SUs, which makes it quite complex to implement.

High complexity is also a striking characteristic of the detection by exploring cyclostationary properties of the primary signal (however, such complexity is lower than in the matched filter detection), which has an attractive performance, though inferior to the detection by matched filter. The high complexity is due to the need of knowing the waveform of the primary signal so that its cyclostationary properties can be extracted. These properties allow determining the presence of the primary signal, since noise is not a cyclostationary process, making it possible to differentiate it from a cyclostationary process.

Eigenvalue-based detectors can assume different forms in their test statistics, but they are all constructed from the eigenvalues of the received signal covariance matrix[12]. Its main advantages are high detection reliability and no need to know the characteristics of the sensed signal. In some eigenvalue sensing techniques it is also not necessary to know the noise power in the receiver, in which case such techniques are considered completely blind.

The main test statistics in eigenvalue spectrum sensing are the aforementioned GLRT, operating on the eigenvalues of the received signal covariance matrix, the detection by the ratio between the maximum and the minimum eigenvalue (MMED), and the maximum eigenvalue detection (MED), also known as Roy's largest root test (RLRT) [25]. The GLRT and the MMED are blind, while the MED is semi-blind due to the fact that it needs the knowledge of the noise power, likewise in the case of the ED.

Among the detectors mentioned above, the remainder of this section details a little more, as case studies, with the ED, with the GLRT based on eigenvalues, with the MMED and with the MED. Later, the article also focuses on other detectors not mentioned yet, which have low complexity and were recently proposed in the literature.

A. Energy detection and eigenvalue detection

The test statistic of the energy detector in the case of centralized CSS with data fusion is given by

$$T_{\text{ED}} = \sum_{i=1}^{m} \frac{1}{\sigma_i^2} \sum_{j=1}^{n} |y_{i,j}|^2, \tag{20}$$

where $y_{i,j}$ is the element on the i-th row and j-th column of $\mathbf{Y}$. Notice that, as previously highlighted, the energy detector needs to know the noise variance σ_i^2 in the i-th SU. This variance can be incorporated into the test statistic itself, as in (20), but not into the decision threshold.

The ED test statistic in the case of unequal noise variances across the SUs sometimes is found in the literature written with the average of the variances, that is, $T_{\text{ED}} = (1/\bar{\sigma}^2) \sum_{i=1}^{m} \sum_{j=1}^{n} |y_{i,j}|^2$, where $\bar{\sigma}^2 = \frac{1}{m} \sum_{i=1}^{m} \sigma_i^2$. This form of test statistic is not correct and leads to performance degradation in comparison with (20). Now, observe that the average noise variance $\bar{\sigma}^2$ can be embedded into the test statistic or into the decision threshold.

The reason for performance degradation of the ED in the case where the average noise variance $\bar{\sigma}^2$ is used instead of the correct variances σ_i^2, $i = 1, \ldots, m$, is that each row of $\mathbf{Y}$ is weighted equally in the formation of the test statistic, but the true noise power affecting the i-th row is σ_i^2.

In the literature, the most common situation, although unrealistic, considers identical noise variances across the SUs, that is, $\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_m^2 = \sigma^2$, yielding the ED test statistic

$$T_{\text{ED}} = \frac{1}{\sigma^2} \sum_{i=1}^{m} \sum_{j=1}^{n} |y_{i,j}|^2. \tag{21}$$

In centralized eigenvalue-based CSS with data fusion, spectral holes are detected based on the eigenvalues of the covariance matrix of the signal received by the m secondary terminals, which, as already mentioned, is defined by $\mathbf{R}_Y = \mathbb{E}\left[\mathbf{Y}\mathbf{Y}^\dagger\right]$. In practice, $\mathbf{R}_Y$ is not exactly known, instead being used its maximum likelihood estimate

$$\hat{\mathbf{R}}_Y = \frac{1}{n}\mathbf{Y}\mathbf{Y}^\dagger, \tag{22}$$

which is referred to as the sample covariance matrix (SCM). The matrix $\hat{\mathbf{R}}_Y$ approaches $\mathbf{R}_Y$ as the number n of samples collected by each SU increases. In the limit, it follows that

$$\mathbf{R}_Y = \lim_{n\to\infty} \left(\frac{1}{n}\mathbf{Y}\mathbf{Y}^\dagger\right). \tag{23}$$

When analyzing (22) one notices that the elements in the i-th row and k-th column of $\hat{\mathbf{R}}_Y$, for $i, k = 1, 2, \ldots, m$, which are given by

$$r_{i,k} = \frac{1}{n} \sum_{j=1}^{n} y_{i,j} y_{k,j}^*, \tag{24}$$

are estimates of the correlation between each of the possible pairs of the elements in the m-dimensional random vector whose realizations are the n columns of $\mathbf{Y}$.

---

[12]The covariance matrix (or autocovariance matrix, in this case) is $\mathbf{R}_Y = \mathbb{E}\{[\mathbf{Y} - \mathbb{E}(\mathbf{Y})][\mathbf{Y} - \mathbb{E}(\mathbf{Y})]^\dagger\}$, but since in the present context we have $\mathbb{E}(\mathbf{Y}) = \mathbf{0}$, then the covariance matrix reduces to $\mathbf{R}_Y = \mathbb{E}\left[\mathbf{Y}\mathbf{Y}^\dagger\right]$. In this case it is usually referred to as the correlation matrix, or autocorrelation matrix.

From the Linear Algebra theory, it is known that the eigenvalues $\{\lambda_i\}$, $i = 1, \ldots, m$, of a square matrix $\mathbf{Z}$ of order $m \times m$ are the solutions (roots) of the equation $\det(\mathbf{Z} - \lambda\mathbf{I}) = 0$, known as the characteristic equation or characteristic polynomial [17]. However, in practice, eigenvalues are not commonly computed by solving such an equation, but rather using more appropriate numerical algorithms, such as those described in [30], [31].

The eigenvalues of a square matrix $\mathbf{Z}$ are the scalars that, when multiplied by the eigenvectors $\mathbf{w}$ of the matrix, produce vectors in the same direction of the vectors produced by the linear operation $\mathbf{Z}\mathbf{w}$, that is, $\mathbf{Z}\mathbf{w} = \lambda\mathbf{w}$.

The eigenvalues and eigenvectors of a matrix have different interpretations and usages, depending on the application. The Appendix A demonstrates how the eigenvalues of $\mathbf{R}_Y$ (or, in practice, the eigenvalues of $\hat{\mathbf{R}}_Y$) are associated with the power of the received signal and how they can be used in the formation of test statistics for spectrum sensing. It is recommended to study this appendix carefully, as it contains interesting concepts, and from where can be extracted some important interpretations.

After the computation of the eigenvalues, it is convenient to put them in descending order, that is, $\{\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m\}$. Subsequently, the test statistics of the detectors GLRT, MMED and MED are respectively formed at the FC according to [25]

$$T_{\text{GLRT}} = \frac{\lambda_1}{\frac{1}{m}\sum_{i=1}^{m}\lambda_i}, \qquad (25)$$

$$T_{\text{MMED}} = \frac{\lambda_1}{\lambda_m}, \qquad (26)$$

$$T_{\text{MED}} = \frac{\lambda_1}{\sigma^2}. \qquad (27)$$

Notice that the detectors GLRT and MMED are blind, whereas the MED is semi-blind due to the need of the noise variance information. Notice also that if the SUs are subjected to unequal noise variances, there is no way of plugging these variances into the MED test statistic, unless in terms of their average $\bar{\sigma}^2$, yielding $T_{\text{MED}} = \lambda_1/\bar{\sigma}^2$. However, the use of $\bar{\sigma}^2$ induces performance loss, for a reason equivalent to the one related with the energy detector.

B. Low complexity robust detectors

Other recent test statistics that deserve to be highlighted, mainly because they are blind, robust and have low computational complexity, are the locally most powerful invariant test (LMPIT) [32], the Gerschgorin radii and centers ratio (GRCR) [19], the Gini index detector (GID) [20], and the Pietra-Ricci index detector (PRIDe) [21].

When it is said that a technique is robust, depending on the context it may refer to different interpretations. The most common interpretation states that a robust detector is one that has little sensitivity of a performance metric to variations in system parameters. In the context of spectrum sensing, the detection technique that has its average performance unaffected

(or at least slightly affected) by the temporal variation of the received signal power, the noise power at the sensor inputs or both, is considered robust, maintaining fixed the averages of these parameters.

The test statistic of the LMPIT is given by

$$T_{\text{LMPIT}} = \sum_{i=1}^{m}\sum_{j=1}^{m}|c_{i,j}|^2, \qquad (28)$$

where $c_{i,j}$ is the element on the i-th row and j-th column of the matrix $\mathbf{C} = \mathbf{D}^{-1/2}\hat{\mathbf{R}}_Y\mathbf{D}^{-1/2}$, for $i, j = 1, 2, \ldots, m$, where $\mathbf{D}$ is the diagonal matrix in which $d_{i,i} = r_{i,i}$, with $r_{i,i}$ denoting the elements of the main diagonal of the sample covariance matrix $\hat{\mathbf{R}}_Y$.

It can be noticed that the computation of T_{LMPIT} is quite simple, since $\mathbf{C}$ is formed by a matrix multiplication, with the operation $\mathbf{D}^{-1/2}$ being also simple due to the fact that $\mathbf{D}$ is a diagonal matrix, so that $\mathbf{D}^{-1/2}$ is formed simply by the reciprocal of the square root of the main diagonal elements of $\mathbf{D}$. The greatest implementation complexity of the LMPIT detector lies, therefore, in the computation of $\hat{\mathbf{R}}_Y$.

In the case of the GRCR, the test statistic is

$$T_{\text{GRCR}} = \frac{\sum_{i=1}^{m}\sum_{j=1, j\neq i}^{m}|r_{i,j}|}{\sum_{i=1}^{m}r_{i,i}}, \qquad (29)$$

where $r_{i,j}$ is the element on the i-th row and j-th column of of the sample covariance matrix $\hat{\mathbf{R}}_Y$.

Finally, the test statistics of the GID and the PRIDe are given respectively by

$$T_{\text{GID}} = \frac{\sum_{i=1}^{m^2}|r_i|}{\sum_{i=1}^{m^2}\sum_{j=1}^{m^2}|r_i - r_j|}, \qquad (30)$$

$$T_{\text{PRIDe}} = \frac{\sum_{i=1}^{m^2}|r_i|}{\sum_{i=1}^{m^2}|r_i - \bar{r}|}, \qquad (31)$$

where r_i, for $i = 1, 2, \ldots, m^2$, is the i-th element of the vector $\mathbf{r}$ formed by stacking all columns of $\hat{\mathbf{R}}_Y$, and $\bar{r} = (1/m^2)\sum_{i=1}^{m^2}r_i$.

Observe that the calculations of T_{GRCR}, T_{GID} and T_{PRIDe} are simple, even simpler than the calculation of T_{LMPIT}, since for the first three statistics the greatest complexity is associated with the computation of $\hat{\mathbf{R}}_Y$. These four detectors are, therefore, slightly more complex than the energy detector, being the blind detectors of less computational complexity known to date.

It must be kept in mind that, although the complexity of computing the test statistic T_{ED} is low if the noise variances are known in advance, the practical-appealing scenario of unknown noise variances should be considered instead, a case in which the complexity associated to the computation of T_{ED} increases due to the signal processing operations needed to estimate the noise variances.

It is informative to mention that the GRCR detector was recently implemented using a hardware-efficient very large scale integration (VLSI) architecture and an application-specific integrated circuit (ASIC) [33], adopting a simplified version of the test statistic (29). A similar implementation strategy applied to the PRIDe detector is now in ongoing development.

Finally, it is worth emphasizing that if centralized CSS with decision fusion is to be implemented, the test statistic chosen among those discussed herein, or among others, must be computed in each SU from the samples defined in (19). For the energy detector, it is enough for each SU to acquire a given number of samples, which have their squared modules added and divided by the variance of the Gaussian noise present in the SU's input, generating the test statistic T_{ED} analogously to (20). To form a local eigenvalue-based test statistic (MED, MMED or GLRT), or one of those for the LMPIT, GRCR, GID and PRIDe detectors, the SCM must be built in each SU through the samples collected by multiple antennas, thus fitting the signal model given in (11). Alternatively, the SCM can be computed in each SU by applying a technique called smoothing [34] in the case of using a single antenna (smoothing can be also applied to the multi-sensor approach).

VI. PRACTICAL ASPECTS OF THE SCM

A. Eigenvalue-equivalent matrix

In practice, in terms of signal processing circuits and even in some simulation software, a complex quantity is manipulated by operating with the real values corresponding to the real and imaginary parts of the quantity. This must be done because the imaginary value $j = \sqrt{-1}$ does not exist in the environment of certain simulators[13], nor in the hardware of real circuits.

This section describes how the eigenvalues of a complex SCM can be computed from a real matrix, which is an eigenvalue-equivalent form of the complex SCM. From this matrix, test statistics based on eigenvalues can be implemented, as is the case of the GLRT, the MED and the MMED detectors presented in the previous section, as well as other detectors that make use of the elements of the SCM, such as the LMPIT, the GRCR, the GID and the PRIDe.

Consider the matrix $\mathbf{Y}$ defined in (11), rewritten as

$$\mathbf{Y} = \Re(\mathbf{Y}) + j\Im(\mathbf{Y}) = \mathbf{Y}_R + j\,\mathbf{Y}_I, \qquad (32)$$

where $\mathbf{Y}_R = \Re(\mathbf{Y})$ and $\mathbf{Y}_I = \Im(\mathbf{Y})$ are real matrices formed by the real and imaginary parts of the elements in $\mathbf{Y}$, respectively. Hence, the product $\mathbf{YY}^\dagger$ in (22) can be written as

$$\begin{aligned}\mathbf{YY}^\dagger &= (\mathbf{Y}_R + j\,\mathbf{Y}_I)(\mathbf{Y}_R + j\,\mathbf{Y}_I)^\dagger \\ &= (\mathbf{Y}_R + j\,\mathbf{Y}_I)(\mathbf{Y}_R^T - j\,\mathbf{Y}_I^T) \\ &= \mathbf{Y}_R\mathbf{Y}_R^T + \mathbf{Y}_I\mathbf{Y}_I^T + j\,\mathbf{Y}_I\mathbf{Y}_R^T - j\mathbf{Y}_R\mathbf{Y}_I^T,\end{aligned} \qquad (33)$$

from where can be defined the matrices

$$\mathbf{S} = \mathbf{Y}_R\mathbf{Y}_R^T + \mathbf{Y}_I\mathbf{Y}_I^T, \qquad (34)$$

$$\mathbf{T} = \mathbf{Y}_I\mathbf{Y}_R^T - \mathbf{Y}_R\mathbf{Y}_I^T. \qquad (35)$$

[13]For example, the imaginary $j = \sqrt{-1}$ does not exist in VisSim/Comm, which is a graphical language for simulation and modeling of end-to-end communication systems at the signal or physical level [35].

Thus, the SCM $\hat{\mathbf{R}}_Y$ in (22) can be rewritten as

$$\begin{aligned}\hat{\mathbf{R}}_Y &= \frac{1}{n}\mathbf{YY}^\dagger \\ &= \frac{1}{n}(\mathbf{S} + j\mathbf{T}).\end{aligned} \qquad (36)$$

Recall that an eigenvector $\mathbf{w}$ of a square matrix $\mathbf{Z}$ is a nonzero vector that, when multiplied by $\mathbf{Z}$, produces a vector in the same direction as $\mathbf{w}$, multiplied by a real or complex scalar λ, that is,

$$\mathbf{Zw} = \lambda\mathbf{w}, \qquad (37)$$

where the scalar λ is an eigenvalue of $\mathbf{Z}$ associated with the eigenvector $\mathbf{w}$. Applying the definition (37) to the alternative representation of $\hat{\mathbf{R}}_Y$ in (36), it follows that

$$\frac{1}{n}(\mathbf{S} + j\mathbf{T})(\mathbf{u} + j\mathbf{v}) = \frac{1}{n}\lambda(\mathbf{u} + j\mathbf{v}), \qquad (38)$$

where $\mathbf{u}$ and $\mathbf{v}$ are the real and the imaginary part of any of the eigenvectors of $\mathbf{YY}^\dagger$, respectively, with the variables λ that satisfy the equation being the corresponding eigenvalues.

From (38) one can write

$$\mathbf{Su} - \mathbf{Tv} + j\mathbf{Tu} + j\mathbf{Sv} = \lambda\mathbf{u} + j\lambda\mathbf{v},$$

or as the system of equations

$$\mathbf{Su} - \mathbf{Tv} = \lambda\mathbf{u},$$

$$\mathbf{Tu} + \mathbf{Sv} = \lambda\mathbf{v},$$

from where it can be obtained the matrix representation

$$\begin{bmatrix} \mathbf{S} & -\mathbf{T} \\ \mathbf{T} & \mathbf{S} \end{bmatrix} \begin{bmatrix} \mathbf{u} \\ \mathbf{v} \end{bmatrix} = \lambda \begin{bmatrix} \mathbf{u} \\ \mathbf{v} \end{bmatrix}. \qquad (39)$$

Observe that this representation is analogous to the basic equation for relating eigenvectors and eigenvalues given in (37). Therefore, the eigenvalues of $\hat{\mathbf{R}}_Y$ can be obtained from the eigenvalues of the block matrix formed with the matrices $\mathbf{S}$ and $\mathbf{T}$ defined in (34) and (35), that is,

$$\begin{aligned}\text{eig}\left(\hat{\mathbf{R}}_Y\right) &= \frac{1}{n}\text{eig}\left(\mathbf{YY}^\dagger\right) \\ &= \frac{1}{n}\text{eig}\left(\begin{bmatrix} \mathbf{S} & -\mathbf{T} \\ \mathbf{T} & \mathbf{S} \end{bmatrix}\right),\end{aligned} \qquad (40)$$

where eig($\cdot$) are the eigenvalues of the underlying matrix. Note that the block matrix contains only real values, as desired. Note also that such matrix has order $2m$, that is, it has $2m$ eigenvalues, which appear duplicated (with multiplicity 2). The eigenvalue pairs are distinct from each other with probability 1 due to the random nature of $\mathbf{S}$ and $\mathbf{T}$. Therefore, only the distinct eigenvalues of the block matrix are used, which, after being multiplied by $1/n$, become equal to the eigenvalues of $\hat{\mathbf{R}}_Y$.

JOURNAL OF COMMUNICATION AND INFORMATION SYSTEMS, VOL. 37, NO.1, 2022.

21

B. Extended sample matrix

Another way to perform the SCM computation, operating with real matrices, is to calculate

$$\hat{\mathbf{R}}_Y \approx \frac{1}{n}\mathbf{Y}_e\mathbf{Y}_e^{\dagger}, \tag{41}$$

where $\mathbf{Y}_e$ is the real matrix generated by concatenating the matrices formed by the real and imaginary parts of $\mathbf{Y}$, that is,

$$\mathbf{Y}_e = [\mathbf{Y}_R \quad \mathbf{Y}_I]. \tag{42}$$

However, it is noteworthy that the approximation given in (41) is quite crude, that is, the SCM estimated by means of (41) differs considerably from the maximum likelihood estimate of $\mathbf{R}_Y$ obtained via (22).

It is also important to point out that the block matrix formed from the matrices $\mathbf{S}$ and $\mathbf{T}$ is not equal to $\hat{\mathbf{R}}_Y$, but it has eigenvalues and eigenvectors equal to those calculated from $\hat{\mathbf{R}}_Y$. Additionally, $\mathbf{S}$ and $\mathbf{T}$ have the real and imaginary parts of all elements of $\hat{\mathbf{R}}_Y$, which opens the possibility of using these matrices for computing any SCM-based test statistic. Thus, the performance of any spectrum sensing technique that uses $\hat{\mathbf{R}}_Y$ is the same as the one achieved using $\mathbf{S}$ and $\mathbf{T}$. However, one have to pay attention to the possible need of adapting the calculations that involve $\hat{\mathbf{R}}_Y$ to those that involve $\mathbf{S}$ and $\mathbf{T}$.

On the other hand, there is no way to predict how the performance of detectors that make use of the SCM will behave when this matrix is approximated by means of (41). Such performance may be better or worse than in the case of using $\hat{\mathbf{R}}_Y$ computed according to (22).

VII. PERFORMANCE METRICS

A. Probability of false alarm and probability of detection

The main spectrum sensing performance metrics have already been briefly covered in Section III, but in this section they are explored in more depth.

As we already know, the spectrum sensing performance is commonly measured through the probability of false alarm

$$P_{\text{fa}} = \text{Pr}\{T > \gamma | \mathcal{H}_0\}, \tag{43}$$

and the probability of detection

$$P_{\text{d}} = \text{Pr}\{T > \gamma | \mathcal{H}_1\}, \tag{44}$$

where $\text{Pr}\{\cdot\}$ denotes the probability of occurrence of the underlying event, T is the test statistic (or decision variable) from the adopted detection technique, and γ is the decision threshold.

In the more general context of binary hypothesis test, the probability of false alarm is also referred to as false positive rate, or type I error rate. The probability of detection is sometimes called true positive rate and the complement of it is called missed detection rate, or type II error rate.

Fig. 5 illustrates the PDFs of a hypothetical test statistic T, conditioned on the occupancy state of the sensed band, as well as the areas that define the probabilities P_{fa} and P_{d}, which are given by

$$P_{\text{fa}} = \int_{\gamma}^{\infty} f_T\{t|\mathcal{H}_0\}dt, \tag{45}$$

$$P_{\text{d}} = \int_{\gamma}^{\infty} f_T\{t|\mathcal{H}_1\}dt. \tag{46}$$
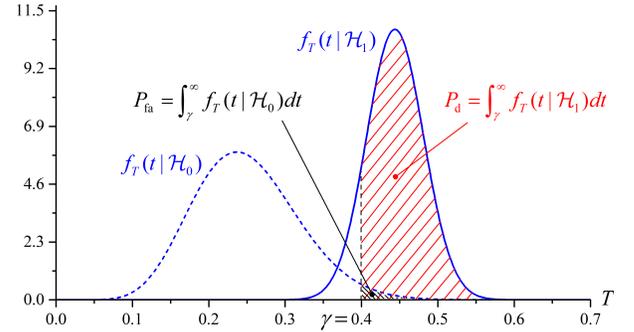
Fig. 5. Definitions of P_{fa} and P_{d} from the conditional PDFs (respectively conditioned on $\mathcal{H}_0$ and $\mathcal{H}_1$) of a hypothetical test statistic T.

The probabilities P_{fa} and P_{d} can also be calculated from the cumulative distribution functions (CDFs) of the test statistic, yielding

$$P_{\text{fa}} = 1 - F_T(\gamma|\mathcal{H}_0), \tag{47}$$

$$P_{\text{d}} = 1 - F_T(\gamma|\mathcal{H}_1), \tag{48}$$

where $F_T(\gamma|\mathcal{H}_0)$ and $F_T(\gamma|\mathcal{H}_1)$ are the CDFs of T under $\mathcal{H}_0$ and $\mathcal{H}_1$, respectively, in $t = \gamma$, that is, $F_T(\gamma|\mathcal{H}_i) = F_T(t|\mathcal{H}_i)|_{t=\gamma}$, for $i = 1$ and $i = 2$.

The decision threshold γ can also be calculated from the CDF of the test statistic under the hypothesis $\mathcal{H}_0$, having as input the desired P_{fa}, that is,

$$\gamma = F_T^{-1}(1 - P_{\text{fa}}|\mathcal{H}_0), \tag{49}$$

where $F_T^{-1}(\cdot)$ is the inverse function of $F_T(t|\mathcal{H}_0)$ for $t = \gamma$, that is, γ is the value of t for which $F_T(t|\mathcal{H}_0) = 1 - P_{\text{fa}}$.

It is desirable to have a high value of P_{d} and a low value of P_{fa}, which is justified as follows. The detection of primary signals with high probability translates into reduced interference in the primary network caused by an opportunistic transmission made in an occupied frequency band wrongly detected as unoccupied. On the other hand, a low probability of false alarm corresponds to a small probability of considering a band that is in fact unoccupied as occupied, increasing the chances of opportunistic use of the spectrum, which in turn increases the data throughput in the secondary network.

Unfortunately, as can be seen from Fig. 5, the goals of increasing P_{d} and reducing P_{fa} are concurrent (or competitors), meaning that increasing the first (by lowering the decision threshold) also causes the second to increase; and the reduction of the second (by increasing the decision threshold) also causes the reduction of the first. In practice, a trade-off typically governed by a standardization document is adopted. For example, the IEEE 802.22 standard [10] requires $P_{\text{d}} \geq 0.9$ and $P_{\text{fa}} \leq 0.1$.

It is also usual to specify a constant target false alarm probability, better known as constant false alarm rate (CFAR), ensuring that the detection probability (or detection rate) stays above of a minimum acceptable value, also recommended by

a standard. For example, one can set $P_{fa} = 0.1$ and configure the system parameters so that, in the worst case situation in terms of SNR, $P_d \geq 0.9$ to comply with the IEEE 802.22 standard.

B. The ROC curve

The probabilities P_{fa} and P_d are typically expressed through the receiver operating characteristic (ROC) curve, in which P_d is expressed as a function of P_{fa} according to the value of the decision threshold γ, as illustrated by[14] Fig. 6.

Each point on a ROC curve is equally optimal in comparison to any other point, as it establishes the intrinsic trade-off resulting from the attempt of optimizing competing objectives, that is, minimize P_{fa} and maximize P_d.

Each point on a ROC curve is associated with a decision threshold value γ, whose range of variation depends on the implemented test statistic. Usually such range corresponds to the union of the supports of both PDFs, beyond which they become practically zero. As an example, in Fig. 5 the range goes from ≈ 0.05 to ≈ 0.6.
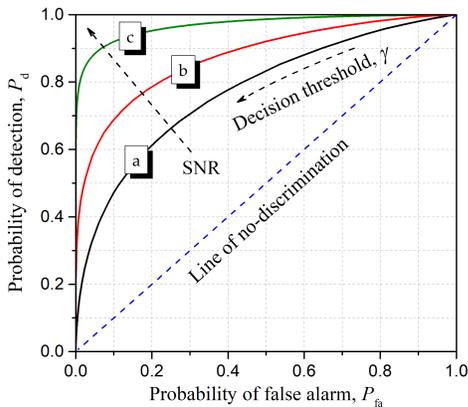
Fig. 6. Examples of ROC curves.

In Fig. 6, curve 'a' shows the worst performance, followed by curve 'b' and finally curve 'c'. Notice this by choosing P_{fa} at some value other than 0 and 1 and verifying that, for the chosen value, P_d is smaller for the curve 'a' than for curve 'b', which in turn has P_d smaller than in the case of curve 'c'.

Increasing the SNR is the most commonly adopted alternative for improving the performance of a given spectrum sensing technique. This improvement can also come from changes in other systemic parameters, such as an increase in the number of samples collected by the SUs or an increase in the number of SUs in cooperation, as discussed in Section IV. Different sensing techniques can also perform differently under the same system conditions, as discussed in Section V.

It is noteworthy that it is not always feasible or possible to change certain system parameters or even the sensing

---

[14] It is not uncommon to find ROC curves expressing the missed detection probability, which is the complement of the detection probability ($P_{md} = 1 - P_d$), as a function of P_{fa}. The P_{md} is the probability of not detecting the primary signal as present, if in fact it is present. There are many other concepts, interpretations and applications of ROC curves in different contexts. See, for example, https://en.wikipedia.org/wiki/Receiver_operating_characteristic.

technique during the operation (on the fly), due to an intrinsic design limitation. What is commonly done is to preset these parameters in order to meet the target performance in critical situations, for example in terms of the lowest acceptable SNR. Thus, the increase of the SNR is indeed the most commonly adopted alternative for improving the spectrum sensing performance.

Still referring to Fig. 6, a random decision on the occupancy status of the sensed band would lead to a point along the line identified as the 'line of no-discrimination'. A ROC curve below this line corresponds to useless performance, except when it is known that the ROC is there, in which case the performance would revert to useful just by reversing the decisions made.

C. The AUC and the decision error probability

Other spectrum sensing performance metrics are the area under the ROC curve (AUC) and the decision error probability. The former, as the name implies, is the area under a ROC curve. Its useful value is in-between 0.5 (corresponding to a ROC on the line of no-discrimination) and 1 (corresponding to the optimal performance with $P_{fa} = 0$ and $P_d = 1$). The decision error probability is calculated by the weighted average of the false alarm and missed detection probabilities, that is,

$$P_{error} = P_{fa} P_{\mathcal{H}_0} + (1 - P_d) P_{\mathcal{H}_1}, \qquad (50)$$

where $P_{\mathcal{H}_1}$ and $P_{\mathcal{H}_0}$ are the probabilities that the primary signals are active and inactive, respectively. The first term of (50) accounts the error probability associated with false alarm events, and the second term accounts for the error probability associated with missed detection events.

Both the AUC and the P_{error} are particularly useful metrics when it is desired to combine P_{fa} and P_d in a single metric, which is attractive, for instance: i) when a ROC curve cross another one, a situation that makes it difficult to establish performance comparisons; ii) when it is desired to reduce the amount of performance measurement values reported in an article or other equivalent scientific document, due to space constraints; iii) when looking for easier visualization and fast interpretation of results.

Example 3: Fig. 7 presents ROC curves obtained by Monte Carlo computer simulation of a CSS system with data fusion, for some of the detectors described in Section V, considering $s = 1$ primary signal having independent and zero mean complex Gaussian samples, $m = 6$ cooperating secondary receivers, $n = 50$ samples gathered by each secondary receiver, and complex AWGN sensing channel with SNR = −10 dB. Notice that, in this situation, which does not represent a general rule, the MED is the one that achieves the best performance, followed by the ED, the GLRT and MMED, in this order. Different models for the sensing channel, for the primary signal or for the noise, as well as different signal power and noise levels at the SU inputs, or the temporal variation of these quantities, may change both the absolute and the relative performance of these and any other sensing technique [19]–[21]. In alternative terms, it is expected that

different detectors perform unequally under a given system setup, and it is also expected that the variation of a given parameter produces different degrees of performance variation in different detectors, since their test statistics operate the received signal differently from each other. Moreover, the intrinsic characteristics of the system hardware contribute with such non-standard behavior of different detectors [36].
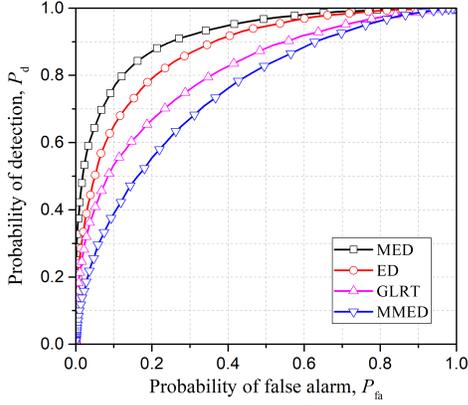
Fig. 7. ROC curves for the detectors MED, ED, GLRT (eigenvalu-based) and MMED, for $s = 1$, $m = 6$, $n = 50$ and SNR $= -10$ dB over the AWGN sensing channel.

The Matlab code used to generate the curves presented in Fig. 7 is available in [37]. The code is also capable of assessing the performance of several other detectors under the system model described in this tutorial, and is thoroughly commented to help the reader understand each line, as well as make modifications if desired.

It is already known that a point on a ROC curve, such as those curves shown in Fig. 7, is associated with a pair P_{fa} and P_d, in this case determined by each decision threshold value in the FC (recall that this figure refers to CSS with data fusion). Each pair of P_{fa} and P_d leads to a decision error probability P_{error}, which is calculated by means of (50) as a function of the probabilities $P_{\mathcal{H}_1}$ and $P_{\mathcal{H}_0}$. Although each pair of P_{fa} and P_d on a ROC curve is equally optimal relative to the other pairs, it is natural to conjecture that a pair yielding the minimum value of P_{error} for each ROC curve may exist. If there is such a minimum P_{error}, the pair P_{fa} and P_d associated with it now becomes global optimal and no longer equally optimal to the others. Note that this is a normal situation, because the performance metric has changed from P_{fa} and P_d to P_{error}, hence changing the spectrum sensing optimality analysis criterion: there are no longer two competing objectives (minimize P_{fa} and maximize P_d), but a single objective (minimize P_{error}).

D. Performance metrics for CSS with decision fusion

The probabilities P_{fa} and P_d discussed so far are global performance metrics achieved at the FC when using centralized cooperative spectrum sensing with data fusion. In the case of centralized cooperative spectrum sensing with decision fusion, there are also local performance metrics, which refer to the sensing performance in each of the SUs in cooperation.

Assume, for simplicity, that the sensing performances achieved by the SUs are equal to each other, and that the corresponding local performance metrics are denoted by P_{faSU} and P_{dSU}. Additionally, consider that transmissions through the control channels from the SUs to the FC are subject to bit errors that occur with probability P_b. Hence, decisions made by the SUs, when subjected to errors in these channels, are interpreted by the FC as if they were associated with equivalent [29] local performance metrics P'_{faSU} and P'_{dSU}, which are given by

$$P'_{faSU} = P_{faSU}(1 - P_b) + P_b(1 - P_{faSU}), \qquad (51)$$

$$P'_{dSU} = P_{dSU}(1 - P_b) + P_b(1 - P_{dSU}). \qquad (52)$$

Assuming that the primary transmitter is inactive, the first of the two terms added in (51) is the probability of maintaining a decision made in favor of $\mathcal{H}_1$ (which is equivalent to maintaining a false positive), and the second is the probability of changing a decision made in favor of $\mathcal{H}_0$ (which is equivalent to reversing a false negative). Now considering that the primary transmitter is active, the first of the terms added in (52) is the probability of maintaining a decision made in favor of $\mathcal{H}_1$ (which is equivalent to maintaining a PU signal detection), and the second is the probability of changing a decision made in favor of $\mathcal{H}_0$ (which is equivalent to reversing a missed detection).

The bit error probability P_b present in (51) and (52) depends on the modulation used for the transmissions over the reporting (control) channels and on the characteristics of these channels. For example, for binary phase-shift keying (BPSK) modulation with coherent detection over the AWGN channel [11], it follows that the bit error probability is

$$P_b = \frac{1}{2}\text{erfc}\left(\sqrt{\frac{E_b}{N_0}}\right), \qquad (53)$$

where $\text{erfc}(x) = \frac{2}{\sqrt{\pi}}\int_x^\infty e^{-z^2}dz$ is the complementary error function, E_b is the average energy per bit, measured in joules, at the receiver side, and N_0 is the power spectral density of the AWGN noise generated at the receiver, measured in watts/hertz.

On the other hand, in a channel with flat and slow Rayleigh fading and with unitary average power gain, the average bit error probability for BPSK modulation with coherent detection [11] is given by

$$P_b = \frac{1}{2}\left(1 - \sqrt{\frac{E_b/N_0}{1 + E_b/N_0}}\right). \qquad (54)$$

The global performance (at the FC) of CSS with fusion of uncorrelated decisions, under the k-out-of-m rule, considering the possibility of errors in the report channel, is related to the local performances (at the SUs) by means of

$$P_{fa} = \sum_{i=k}^{m}\binom{m}{i}(P'_{faSU})^i(1 - P'_{faSU})^{m-i}, \qquad (55)$$

$$P_d = \sum_{i=k}^{m}\binom{m}{i}(P'_{dSU})^i(1 - P'_{dSU})^{m-i}, \qquad (56)$$

where $\binom{m}{i} = \frac{m!}{(m-i)!i!}$ is the binomial coefficient.

Expressions (55) and (56) can be interpreted as the probabilities of occurrence of k or more successes in m Bernoulli trials, taking into account that the number of successes in a set of uncorrelated Bernoulli trials has a binomial distribution. Thus, P_{fa} in (55) is the probability that k or more SUs produce a false alarm and P_{d} in (56) is the probability that k or more SUs detect the primary signal [29].

It is important to call the reader's attention to the fact that expressions (55) and (56) assume that the local performance metrics are the same for all SUs. In case of possibly unequal performances [38], the global probabilities of false alarm and detection under the k-out-of-m rule are respectively given by

$$P_{\text{fa}} = \sum_{\ell=k}^{m} \sum_{j=1}^{|\mathbf{D}_\ell|} \prod_{i=1}^{m} P'_{\text{faSU}i}{}^{D_{\ell_{j,i}}} (1 - P'_{\text{faSU}i})^{1-D_{\ell_{j,i}}}, \quad (57)$$

$$P_{\text{d}} = \sum_{\ell=k}^{m} \sum_{j=1}^{|\mathbf{D}_\ell|} \prod_{i=1}^{m} P'_{\text{dSU}i}{}^{D_{\ell_{j,i}}} (1 - P'_{\text{dSU}i})^{1-D_{\ell_{j,i}}}, \quad (58)$$

where $|\mathbf{D}_\ell| = \binom{m}{\ell} = \frac{m!}{(m-\ell)!\ell!}$ is the cardinality (number of elements) of the set $\mathbf{D}_\ell$. This set is defined as follows: Let $d_i = 1$ or $d_i = 0$ denote the decision made by the i-th SU in favor of $\mathcal{H}_1$ or $\mathcal{H}_0$, respectively. The set $\mathbf{D}_\ell$ contains the binary m-uples that satisfy $\sum_{i=1}^{m} d_i = \ell$. Thus, $\mathbf{D}_\ell$ can be interpreted as a matrix of order $\binom{m}{\ell} \times m$, and $D_{\ell_{j,i}} \in \{0, 1\}$ is the element sitting on the j-th row and i-th column of this matrix.

Obviously, (57) and (58) specialize to (55) and (56) if the local performance metrics are the same for all SUs, that is, $P'_{\text{faSU}i} = P'_{\text{faSU}}$ and $P'_{\text{dSU}i} = P'_{\text{dSU}}$ for all SUs.

The decision error probability defined in (50) also applies to the decision fusion, in this case serving to calculate local or global P_{error}.

Example 4: A possible difficulty may arise when applying (57) and (58), owed mainly to a misunderstanding about the formation of the set $\mathbf{D}_\ell$. This example aims at exemplifying this formation. Assume $k = 2$ and $m = 3$ in the k-out-of-m rule. The set with all possible decisions of the $m = 3$ SUs has the $2^m = 2^3 = 8$ triplets

$$
\begin{matrix}
0 & 0 & 0 \\
0 & 0 & 1 \\
0 & 1 & 0 \\
0 & 1 & 1 \\
1 & 0 & 0 \\
1 & 0 & 1 \\
1 & 1 & 0 \\
1 & 1 & 1
\end{matrix}
$$

In this case, according to the leftmost summation in (57) and (58), ℓ goes from $k = 2$ to $m = 3$, yielding $m - k + 1 = 3 - 2 + 1 = 2$ sets $\mathbf{D}_\ell$, namely $\mathbf{D}_2$ and $\mathbf{D}_3$, which are

$$\mathbf{D}_2 = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix},$$

$$\mathbf{D}_3 = \begin{bmatrix} 1 & 1 & 1 \end{bmatrix}.$$

Notice that the sum of the elements in each row of $\mathbf{D}_2$ is $\ell = 2$, and the sum of the elements in the single row of $\mathbf{D}_3$ is $\ell = 3$, which is in agreement with $\sum_{i=1}^{m} d_i = \ell$.

Thus, the cardinality of $\mathbf{D}_2$ is $\binom{m}{\ell} = \binom{3}{2} = 3$ and the cardinality of $\mathbf{D}_3$ is $\binom{m}{\ell} = \binom{3}{3} = 1$. The elements $D_{\ell_{j,i}}$ of these sets are: $D_{2_{1,1}} = 0$, $D_{2_{1,2}} = 1$, $D_{2_{1,3}} = 1$, $D_{2_{2,1}} = 1$, $D_{2_{2,2}} = 0$, $D_{2_{2,3}} = 1$, $D_{2_{3,1}} = 1$, $D_{2_{3,2}} = 1$, and $D_{2_{3,3}} = 0$ in the case of $\mathbf{D}_2$. In the case of $\mathbf{D}_3$, the elements $D_{\ell_{j,i}}$ are $D_{3_{1,1}} = 1$, $D_{3_{1,2}} = 1$, and $D_{3_{1,3}} = 1$.

Example 5: Fig. 8 illustrates the relationship between local and global performances of centralized CSS with fusion of uncorrelated decisions, under the k-out-of-m rule, for $m = 5$ SUs, $k = 1$ (OR rule), $k = 3$ (majority voting rule, MAJ) and $k = 5$ (AND rule), for $P_{\text{b}} = 0$ ($E_{\text{b}}/N_0 = \infty$ dB on the Rayleigh or AWGN channel) and $P_{\text{b}} = 0.05$ ($E_{\text{b}}/N_0 = 6.25$ dB on the Rayleigh channel, $E_{\text{b}}/N_0 = 1.29$ dB on the AWGN channel). The ROC identified as 'local ROC' in this figure has been arbitrarily placed in its position, and represents equal performances of the SUs. The other ROCs were generated by applying the expressions (51), (52), (55) and (56).
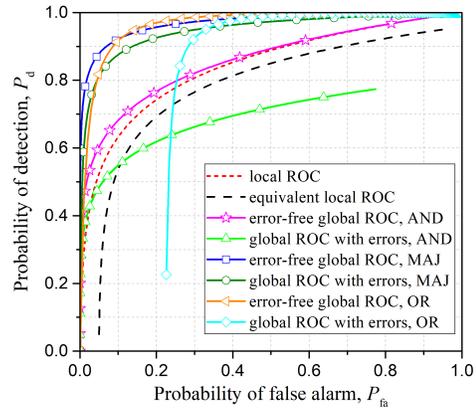
Fig. 8. Illustration of the relation between local and global performance metrics when the k-out-of-m rule is applied, for $m = 5$, $k = 1$ (OR rule), $k = 3$ (MAJ rule) and $k = 5$ (AND rule), in the error-free ($P_{\text{b}} = 0$) situation, and with errors ($P_{\text{b}} = 0.05$) caused in the report channel transmissions.

In the absence of errors in the report channel, it can be observed in Fig. 8 that the cooperation under the OR and MAJ rules significantly improves the sensing performance with respect to what would be obtained with a single SU, whereas the AND rule only produces a small improvement.

When errors in the report channel exist, with $P_{\text{b}} = 0.05$, only the MAJ rule produces cooperation gains in the whole extension of its ROC. The OR and AND rules have their performances strongly affected by errors in the report channel, and the OR rule still produces cooperation gains in some regions of its ROC, while the AND rule does not produce cooperation gains in practically any region of its ROC.

It can be also noticed in Fig. 8 that the occurrence of errors in the report channel produces a saturation or bounding [39] effect in the ROCs of the OR and AND rules. This saturation bounding to irreducible levels of P_{fa} in the OR rule, and insurmountable levels of P_{d} in the AND rule. Bounding occurs due to the fact that errors in the report channel never allow that,

in the OR rule, the global probability of false alarm is zeroed, even if the decisions of the SUs are always in favor of the $\mathcal{H}_0$ hypothesis. Similarly, in the AND rule, the errors do not allow the global probability of detection to be unitary even if the SUs' decisions are always in favor of $\mathcal{H}_1$. It is informative to mention that all ROC saturation endpoints coincide with the no-discrimination line, in which $P_d = P_{fa}$.

It is also important to emphasize that the relative performances of the OR, MAJ and AND rules are not always those shown in Fig. 8, being influenced in different ways for different system parameters and different system models and SNRs for the sensing and report channels. It is also interesting to conclude from this figure that the MAJ rule can be less sensitive to errors in the report channel, yielding performances very close to those obtained in the error-free reporting transmission situation, even with $P_b = 0.05$. Observe that the value of $P_b = 0.05$ is much higher than those bit error probabilities usually found in conventional digital transmission systems, demonstrating that the decision fusion process for spectrum sensing purposes is much less sensitive to bit errors than the conventional digital communication process.

An important concept associated with any decision process refers to the fact that its accuracy also depends on the way the input information is presented to it. Basically, there is the possibility of operating with soft information (such as the sample values sent by the SUs in the context of spectrum sensing) or hard information (such as the decisions sent by the SUs). Soft information is naturally richer, since, for example, its polarity can be associated with a hard decision and its absolute value can in some way represent the reliability of that hard decision. For example, the farther from the decision threshold the value of a received sample is, the more assertive the corresponding decision becomes. On the other hand, a sample value close to the threshold may be associated with a higher influence of noise or other factors that degrade the sensing performance. From what is exposed in this paragraph, it can be inferred that data fusion has the potential to provide more accuracy in spectrum sensing than decision fusion.

A simple experiment can be made to demonstrate that data fusion can outperform decision fusion in spectrum sensing, modifying the simulation code available in [37]. The steps that must be followed are: i) Enable only the energy detector and set $s = 1$ and $m = 1$, AWGN sensing channel, and no signal and noise power variation. ii) Adjust the simulation parameters (for example, the SNR, the number of samples or both) so that the ROC is located as close as possible to the local ROC shown in Fig. 8. iii) Make $m = 5$ and check the position of the global ROC: If it sits above the best ROC with decision fusion, it is demonstrated that data fusion can deliver better performance than decision fusion. The implementation of this experiment is left to the reader as an exercise.

Example 6: Aiming at further exploring the concepts about the error probabilities involved in the decision fusion scenario, Fig. 9 presents bit error probabilities and minimum decision error probabilities obtained by computational search, taking

the MAJ rule as an example. The figure shows five curves, described as follows:

1) The minimum local P_{error} is the minimum value of the decision error probability defined in (50), associated with the performance of the SUs. It is constant, as it is measured in the SUs and, therefore, it is not linked to bit errors in the report channel.
2) The equivalent minimum local P_{error} is the minimum value of the decision error probability of the SUs 'from the point of view' of the FC, due to errors induced by the report channel. It is influenced by the SNR of the report channel, tending to the constant value of the previous item as the SNR grows. This P_{error} is just an indirect measure, not actually occurring in the system.
3) The minimum error-free global P_{error} is the minimum value of the error probability of a global decision made in the FC on the occupation state of the sensed band, considering that there are no errors in the transmission of the decisions from the SUs to the FC.
4) The minimum global P_{error} with errors is the minimum value of the global decision error probability, taking into account the errors in the transmissions from the SUs to the FC. Notice that it tends to the P_{error} of the previous item as the report channel's SNR increases.
5) P_b for BPSK over Rayleigh fading is the bit error probability defined in (54).
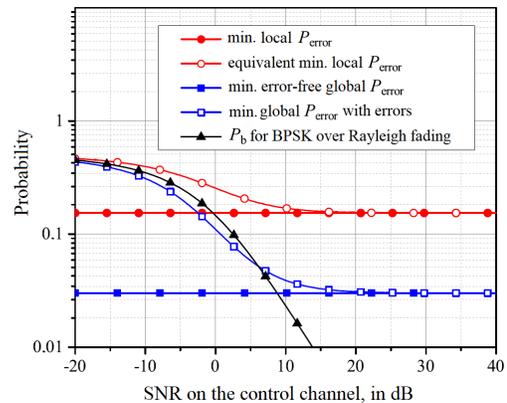
Fig. 9. Bit error probability and minimum local and global decision error probabilities under the k-out-of-m rule, for $k = 3$ and $m = 5$ (MAJ rule), as functions of the SNR in the reporting Rayleigh fading channels.

In addition to exemplifying how the minimum decision error probability is influenced by the SNR of the report channel, Fig. 9 shows the cooperation gain seen from another viewpoint. Note that the lowest value of local P_{error} is not able to reach 0.1, a value that would correspond, for example, to the target values $P_d = 0.9$ and $P_{fa} = 0.1$, if $P_{\mathcal{H}_1} = P_{\mathcal{H}_0} = 0.5$ in (50). By the effect of cooperation, the global values of P_{error} reach values even smaller than 0.1. Also note that the global value of $P_{error} = 0.1$ is reached for an SNR around 1 dB, for which the bit error probability of the BPSK modulation is approximately 0.14. In other words, it is again observed that it is possible to achieve target sensing performances even under very high bit error rates (or very low SNRs) in the control

channels, compared to those bit error rates typically required in conventional digital communication systems.

VIII. DATABASE-DRIVEN INTERNET OF THINGS-ENABLED DYNAMIC SPECTRUM ACCESS

It has been already mentioned that the opportunistic use of temporarily free radio-frequency bands constitutes the main driver of the dynamic spectrum access (DSA) policy, which is undeniably crucial to solve the problem of spectrum scarcity.

To carry out the DSA, two main approaches, combined or not, can be adopted. In the first, a spectrum sensing technique detects unused frequency bands before the opportunistic dynamic access by the SUs. In the second approach, the SUs access a spectrum occupation database, where a list of available channels is stored. This list is constantly updated, for example with the aid of spectrum sensing, coverage prediction, or both. Regardless of the adopted approach, as soon as a channel becomes vacant, it is desirable to promptly detect such an event in order to increase the chances of shared spectrum usage. On the other hand, as soon as the primary user re-establishes the occupation of a certain band, the secondary network must be able to detect the event and vacate the band before harmful interference is caused to the primary network. Hence, not only the accuracy regarding the spectrum occupancy state is important, but also the agility in detecting occupancy state changes.

The conventional CSS approach, the one that applies distributed detection and centralized decision, as illustrated in Fig. 10. This approach can be improved if the secondary network has access to a spectrum occupation database to refine the global decision compared to the one that would be made only using spectrum sensing.
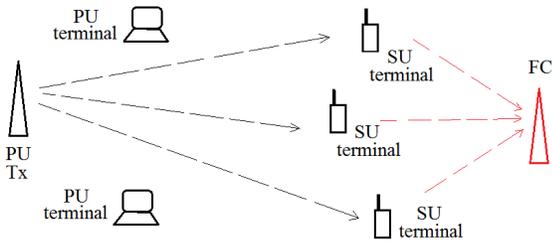
Fig. 10. Conventional CSS, not assisted by database information.

In the conventional approach, it is noteworthy that SU terminals, which can be as simple as sensor nodes or sophisticated as smartphones, need to be equipped with the spectrum sensing capability, which increases complexity and energy consumption of such devices, potentially increasing their cost and physical dimensions, and may also reduce their portability.

A DSA architecture enabled by a database fed by a support IoT network is proposed in [40], and is illustrated in Fig. 11, in a simplistic didactic version. In this architecture, the spectrum sensing task is shifted from the SU terminals to special IoT devices with spectrum sensing functions, hereafter named spectrum sensing IoT (SSIoT) devices. An SSIoT device, whose block diagram is depicted in Fig. 12, is formed by

connecting an ordinary IoT device to a spectrum sensing (SS) module, through a standard wired or wireless interface. Notice in Fig. 11 that not all IoT devices form an SSIoT.
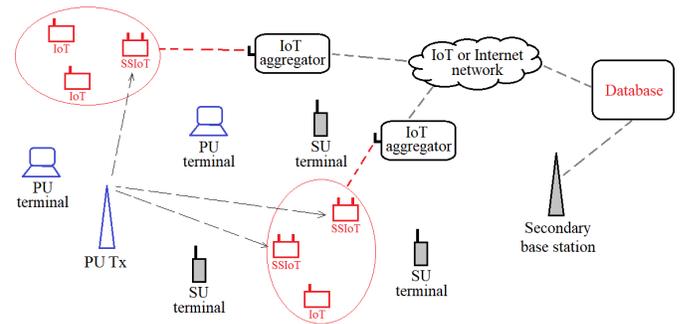
Fig. 11. Database-driven Internet of things-enabled dynamic spectrum access architecture.

The IoT device and the SS module have their own antennas, as their characteristics, mainly the bandwidth and the central operating frequency, are distinct in practice. For example, the IoT network operates in a frequency range, while spectrum sensing is performed in the secondary network's operating frequency range.
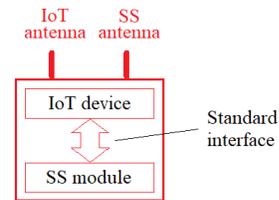
Fig. 12. SSIoT device: IoT device equipped with a spectrum sensing module.

The SSIoT devices monitor the RF spectrum and transmit the sensing information to an IoT aggregator, which also collects IoT-related data from a cluster of IoT devices that are conveniently close to each other. The existence of the aggregator is not mandatory, but it can remove specific tasks from the terminals, mainly those related to the security of IoT and SSIoT devices, making these devices simpler and protecting them from malicious attacks.

The aggregators communicate with the IoT network, which can be part of an Internet network or is simply connected to the Internet. The spectrum occupancy database regularly accesses this network to update the list of available channels for DSA. This database can also process current and past primary network activities to provide spectrum usage predictions and other relevant information that can improve the efficacy of the search for vacant bands.

The dynamic access to vacant bands is carried out through constant monitoring of the spectrum occupation database, what can be easily made during regular control communication between the secondary terminals and their base station.

It is worth mentioning that, in the present DSA architecture, the information from the spectrum sensing process carried out by the network of SSIoTs is used to update the spectrum occupation database, allowing dynamic access by the secondary network without the terminals of this network being equipped

with spectrum sensing circuits. However, SSIoTs can also be used to monitor the spectrum used by the IoT network itself, helping its terminals to access bands that are less congested or free from strong interference, for example.

IX. RESEARCH CHALLENGES AND OPPORTUNITIES REGARDING SPECTRUM SENSING

There are numerous challenges and, therefore, many research opportunities on spectrum sensing in general, and on centralized cooperative spectrum sensing in particular, the latter being the focus of this tutorial. Such challenges can be classified according to the steps of the process: the sensing itself, the fusion of the sensing information, the global decision at the fusion center, and the subsequent access to the spectrum. Listed in the next subsections are some of the research challenges in the context of spectrum sensing, regarding the physical layer aspects.

A. Sensing channel

Regarding the challenges inherent to the sensing channel, possible research could seek to develop or adapt models that mimic this channel as accurate as possible. The models that are welcome are those viable from the point of view of mathematical treatment and the implementation of simulations, or at least form the point of view of the implementation of simulations. As an example, the sensing channel models range from the simplest AWGN channel, passing through channels with flat, selective, slow or fast fading, with typical distributions like Rayleigh and Rice, or more elaborate ones like the generalized distributions, culminating with the addition of effects like the shadowing by obstacles, the attenuation by distance or other propagation mechanisms, and the correlation between fading, shadowing, or both, affecting the signals received by different secondary terminals.

B. Detection technique

With regard to local (in CSS with decision fusion) or centralized (in CSS with data fusion) detection technique, the challenges become even more relevant. There are several criteria for designing detectors for spectrum sensing purposes, but often the mathematical treatment associated with the design becomes an insurmountable obstacle because it is not possible to take into account all situations in terms of the sensing channel, the characteristics of the primary signal, the knowledge of the noise variance, the imprecision of parameters that must be estimated and the temporal variation of noise and signal levels, just to name a few examples.

As they do not consider all systemic variables and are developed under different criteria, local and global sensing techniques end up presenting huge variations in performance under varied circumstances, which leaves room for the development of new techniques that are suitable for typical applications in more specific scenarios. In addition, the various test statistics present different degrees of theoretical analysis complexity, mainly in obtaining and manipulating the probability density functions associated with them, under both test hypotheses, so that expressions for calculating the decision threshold and the detection and false alarm probabilities are derived.

There are also opportunities for developing detectors based on some existing discrepancy metric adopted in a context different of the spectrum sensing, an alternative that has been proved to work properly, for instance in the design of the detectors GRCR [19], GID [20] and PRIDe [21].

C. Report channel

It is very common in research efforts dealing with centralized CSS to consider that the sensing information, whether decisions, samples or quantities derived from these samples, are sent to the fusion center through orthogonal and degradation-free report channel, so that such information is available there as it was generated at the SUs in cooperation.

The main challenges concerning the report channels consist of considering them as not perfect and non-orthogonal, allowing the impact on the global sensing performance to be investigated, at the same time seeking to increase the efficiency in the use of the small available bandwidth, preferably with high energy efficiency.

The influence of correlated decisions or correlated channel imperfections is also an important scenario to consider in the report channel. The imperfections of the report channel are similar to those associated with the sensing channel and, therefore, the models of both may be similar. However, it is important to remember that the roles of these channels in the spectrum sensing process as a whole are very different, allowing to conclude that the impacts of similar imperfections in both may be very different. Hence, the need of treating them with equal importance in research is of paramount relevance.

Added to the effects of channel imperfections are those caused by imperfections and signal processing tasks performed in the receivers, especially in the case of data fusion, before the samples are sent to the fusion center. Nonlinearities, quantization errors and automatic gain control effects, noise figure and filtering are examples [41].

Also with regard to errors induced by the report channel, an important and challenging topic of investigation that has been little explored in the technical literature is the development of error-correcting codes or the simple application of existing codes to protect the transmissions to the fusion center. Such transmissions differ significantly from those carried out in conventional communication systems, as the bit error rates supported in the report channel are considerably higher than those typically aimed at in conventional systems. Furthermore, transmissions under extremely low SNR are expected, aiming at energy savings. Such regimes direct the use of error-correcting codes with high correction capability, which, in general, have long codewords. However, it is important to recall that the interval reserved for reporting transmissions is significantly limited, preventing long codes from being applied. On the other hand, the way in which transmissions are carried out in the report channel prevents the use of error control techniques via retransmissions, or at least significantly limits the degree of freedom in the design of such techniques, mainly due to the aforementioned time restrictions.

Digital communication over report channels under extremely low signal-to-noise ratios also motivates another topic of research associated with the one discussed in the previous paragraph, which is the study of the Information Theory applied to the reporting transmissions scenario, not only to address the limits of communication, but also to support the design of appropriate error-correcting codes.

X. Summary and final remarks

This tutorial presented basic concepts, fundamentals and state-of-the-art techniques related to spectrum sensing. The topics covered included concepts regarding the forms of spectrum sensing and basic fundamentals on detection theory, modern cooperative spectrum sensing techniques and research challenges, mathematical models for the signal and the sensing channel, the signal-to-noise ratio wall, practical issues regarding signal processing tasks and performance metrics. Several examples were also given throughout the text to illustrate some theoretical content and help the reader to understand more complex concepts. It has been also addressed a dynamic spectrum access framework that makes use of Internet of things devices equipped with spectrum sensing modules as a means to feed spectrum occupation databases to drive the dynamic spectrum access.

An attempt has been made to confer to this tutorial a significant difference from existing tutorials or surveys on spectrum sensing, thinking its content as a short course material capable of giving to the reader a first contact with the subject without the need of complex mathematical background. It is the author's hope that the material presented in this tutorial, along with the list of references, will be able to prepare the reader for further studies in this exciting area.

Appendix A

This appendix demonstrates that the total normalized (with respect to a 1 Ω load) power of the signal received at the fusion center is the sum of the eigenvalues of the received signal covariance matrix, and that it is possible to carry out the primary signal detection by comparing eigenvalues or quantities obtained from them.

In the CSS model considered in this tutorial, the sample matrix received by the FC is $\mathbf{Y} = \mathbf{HX} + \mathbf{V}$. The received signal covariance matrix is $\mathbf{R}_Y = \mathbb{E}\left[\mathbf{YY}^\dagger\right] = \mathbf{HR}_X\mathbf{H}^\dagger + \mathbf{R}_V$, where $\mathbf{R}_X$ is the covariance matrix of the transmitted signal. Using the property $\mathrm{tr}(\mathbf{A} + \mathbf{B}) = \mathrm{tr}(\mathbf{A}) + \mathrm{tr}(\mathbf{B})$, it follows that $\mathrm{tr}(\mathbf{R}_Y) = \mathrm{tr}(\mathbf{HR}_X\mathbf{H}^\dagger) + \mathrm{tr}(\mathbf{R}_V)$. Using the property $\mathrm{tr}(\mathbf{A}) = \sum_j \lambda_{\mathrm{A}j}$, being $\{\lambda_{\mathrm{A}_j}\}$ the eigenvalues of $\mathbf{A}$, then $\mathrm{tr}(\mathbf{R}_Y) = \sum_{j=1}^m \lambda_{\mathrm{Ry}_j}$, where $\{\lambda_{\mathrm{Ry}_j}\}$ are the eigenvalues of $\mathbf{R}_Y$.

Let $\mathbf{y}_j$ be the j-th row of the matrix $\mathbf{Y}$. Thus, $\mathrm{tr}(\mathbf{R}_Y) = \sum_{j=1}^m \mathbb{E}[\mathbf{y}_j\mathbf{y}_j^\dagger] = \sum_{j=1}^m \mathbb{E}[\sum_{i=1}^n |y_{j,i}|^2] = \sum_{j=1}^m \lim_{n \to \infty} [\frac{1}{n}\sum_{i=1}^n |y_{j,i}|^2]$, which is the total average power of the signal received by the FC, which can be denoted by P_{RX}. Since $\mathrm{tr}(\mathbf{R}_Y) = \sum_{j=1}^m \lambda_{\mathrm{Ry}_j}$, it can be concluded that $\sum_{j=1}^m \lambda_{\mathrm{Ry}_j} = P_{\mathrm{RX}}$, that is, the sum of the eigenvalues of the received signal covariance matrix is equal to the average normalized power of this signal.

Since $\sum_{j=1}^m \lambda_{\mathrm{V}j} = \mathrm{tr}(\mathbf{R}_V) = m\sigma^2$, with $m\sigma^2$ being the total noise power at the SUs, it follows that $\sum_{j=1}^m \lambda_{\mathrm{V}j} = P_{\mathrm{noise}}$.

Taking into account that $P_{\mathrm{RX}} = P_{\mathrm{signal}} + P_{\mathrm{noise}}$ and that $\mathrm{tr}(\mathbf{R}_Y) = \mathrm{tr}(\mathbf{HR}_X\mathbf{H}^\dagger) + \mathrm{tr}(\mathbf{R}_V)$, additionally it can be concluded that $\mathrm{tr}(\mathbf{HR}_X\mathbf{H}^\dagger)$ is the power P_{signal} associated with the signal part present in $\mathbf{Y}$.

Now the analysis turns to what happens in terms of the sum of the eigenvalues. For square matrices $\mathbf{A}$ and $\mathbf{B}$, it is known that, with exceptions, $\mathrm{eig}(\mathbf{A} + \mathbf{B}) \neq \mathrm{eig}(\mathbf{A}) + \mathrm{eig}(\mathbf{B})$. However, in the present context it is true that $\mathrm{eig}(\mathbf{R}_Y) = \mathrm{eig}(\mathbf{HR}_X\mathbf{H}^\dagger + \mathbf{R}_V) = \mathrm{eig}(\mathbf{HR}_X\mathbf{H}^\dagger) + \mathrm{eig}(\mathbf{R}_V)$, as demonstrated below.

The set of eigenvalues λ_{Ry} is the solution of the characteristic equation $\det(\mathbf{R}_Y - \lambda_{\mathrm{Ry}}\mathbf{I}) = 0$. But $\det(\mathbf{HR}_X\mathbf{H}^\dagger + \mathbf{R}_V - \lambda_{\mathrm{Ry}}\mathbf{I}) = \det(\mathbf{HR}_X\mathbf{H}^\dagger + \sigma^2\mathbf{I} - \lambda_{\mathrm{Ry}}\mathbf{I}) = \det[\mathbf{HR}_X\mathbf{H}^\dagger - (\lambda_{\mathrm{Ry}} - \sigma^2)\mathbf{I}]$, that is, $(\lambda_{\mathrm{Ry}} - \sigma^2)$ are the eigenvalues of $\mathbf{HR}_X\mathbf{H}^\dagger$. Then, it follows that $\mathrm{eig}(\mathbf{HR}_X\mathbf{H}^\dagger + \mathbf{R}_V) = \mathrm{eig}(\mathbf{HR}_X\mathbf{H}^\dagger) + \mathrm{eig}(\mathbf{R}_V)$, since the following equality must hold:

$$\underbrace{\lambda_{\mathrm{Ry}}}_{\mathrm{eig}(\mathbf{HR}_X\mathbf{H}^\dagger + \mathbf{R}_V)} = \underbrace{(\lambda_{\mathrm{Ry}} - \sigma^2)}_{\mathrm{eig}(\mathbf{HR}_X\mathbf{H}^\dagger)} + \underbrace{\sigma^2}_{\mathrm{eig}(\mathbf{R}_V)} .$$

Observe that, in the absence of the PU signal, the eigenvalues of $\mathbf{R}_Y$ are equal to one another and, for example, $\lambda_1/\lambda_m = 1$. When the PU signal is present, then $\lambda_1/\lambda_m > 1$. With this simple example, it can be seen that it is possible to use the eigenvalues of the received signal covariance matrix as a means to detect the presence of the PU in the sensed band.

Complementarily, the eigenvalues of $\mathbf{R}_Y$ also contain information about the number of transmitters (whose most general and usual term is number of sources), as the number of nonzero eigenvalues in $\mathrm{eig}(\mathbf{HR}_X\mathbf{H}^\dagger)$ is equal to the number of transmitters. This fact is used to estimate the number of signal sources from the eigenvalues of $\mathbf{R}_Y$: the smallest $(m - s)$ eigenvalues are equal to σ^2, where s is the number of sources [42]. However, in practice, such an estimate is performed through $\hat{\mathbf{R}}_Y$ and not through $\mathbf{R}_Y$. As a consequence, the classification of the eigenvalues of $\hat{\mathbf{R}}_Y$ into two groups with the largest s and the smallest $(m - s)$ is therefore not simple, constituting the hardest difficulty for estimating the number of signal sources in practice.

References

[1] A. Nasser, H. Al Haj Hassan, J. Abou Chaaya, A. Mansour, and K.-C. Yao, "Spectrum sensing for cognitive radio: Recent advances and future challenge," *Sensors*, vol. 21, no. 7, 2021, doi: 10.3390/s21072408.

[2] W. Dias, A. Ferreira, R. Kagami, J. S. Ferreira, D. Silva, and L. L. Mendes, "5G-Range: A transceiver for remote areas based on software-defined radio," in *2020 European Conference on Networks and Communications (EuCNC)*, 2020, pp. 100–104, doi: 10.1109/EuCNC48522.2020.9200925.

[3] Federal Communications Commission, FCC, *Spectrum Policy Task Force Report*. FCC, 2002. [Online]. Available: https://books.google.com.br/books?id=p9MQtwAACAAJ.

[4] D. Das and S. Das, "A survey on spectrum occupancy measurement for cognitive radio," *Wirel. Pers. Commun.*, vol. 85, no. 4, p. 2581–2598, Dec 2015, doi: 10.1007/s11277-015-2921-1.

[5] J. Mitola III and G. Q. Maguire Jr., "Cognitive radio: making software radios more personal," *IEEE Pers. Commun. Mag.*, vol. 6, no. 4, pp. 13–18, Aug. 1999, doi: 10.1109/98.788210.

[6] Y. Zeng, Y.-C. Liang, A. Hoang, and R. Zhang, "A review on spectrum sensing for cognitive radio: Challenges and solutions," *EURASIP Journal on Advances in Signal Processing*, vol. 2010, no. 1, p. 381465, 2010, doi: 10.1155/2010/381465.

[7] I. F. Akyildiz, B. F. Lo, and R. Balakrishnan, "Cooperative spectrum sensing in cognitive radio networks: A survey," *Elsevier Physical Comm.*, vol. 4, pp. 40–62, mar 2011, doi: 10.1016/j.phycom.2010.12.003.

[8] Y. Arjoune and N. Kaabouch, "A comprehensive survey on spectrum sensing in cognitive radio networks: Recent advances, new challenges, and future research directions," *Sensors*, vol. 19, no. 1, 2019, doi: 10.3390/s19010126.

[9] C. R. Stevenson, G. Chouinard, Z. Lei, W. Hu, S. J. Shellhammer, and W. Caldwell, "IEEE 802.22: The first cognitive radio wireless regional area network standard," *IEEE Commun. Mag.*, vol. 47, no. 1, pp. 130–138, January 2009, doi: 10.1109/MCOM.2009.4752688.

[10] The Institute of Electrical and Electronic Engineers, IEEE. (2011) IEEE 802 part 22: Cognitive wireless RAN medium access control (MAC) and physical layer (PHY) specifications: Policies and procedures for operation in the TV bands. [Online]. Available: http://standards.ieee.org/getieee802/download/802.22-2011.pdf.

[11] D. A. Guimarães, *Digital Transmission: A Simulation-Aided Introduction with VisSim/Comm*. Berlin Heidelberg, Germany: Springer-Verlag, 2009, doi: 10.1007/978-3-642-01359-1.

[12] L. Sachs and Z. Reynarowych, *Applied Statistics: A Handbook of Techniques*, ser. Springer Series in Statistics. Springer, New York, USA, 1984. [Online]. Available: https://books.google.com.br/books?id=apLuAAAAMAAJ.

[13] D. A. Guimarães and R. A. A. de Souza, "Exploring practical aspects of the FSK modulation with non-coherent matched filter detection," *Revista de Tecnologia da Informação e Comunicação*, vol. 5, no. 1, pp. 22–28, 2015. [Online]. Available: http://rtic.com.br/index.php/rtic/article/view/60.

[14] R. McDonough and A. Whalen, *Detection of Signals in Noise*. Elsevier Science, 1995. [Online]. Available: https://books.google.com.br/books?id=AyBFXKzbx_QC.

[15] H. V. Poor, *An Introduction to Signal Detection and Estimation*, 5th ed. Berlin, Heidelberg: Springer-Verlag, 1994. [Online]. Available: https://link.springer.com/book/10.1007/978-1-4757-2341-0.

[16] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1993. [Online]. Available: https://dl.acm.org/doi/10.5555/151045.

[17] G. Strang, *Introduction to Linear Algebra*, 4th ed. Wellesley, MA: Wellesley-Cambridge Press, 2009. [Online]. Available: https://www.bibsonomy.org/bibtex/2bc7873401f7c192b135b97133880f2e8/ytyoun.

[18] C. Forbes, M. Evans, N. Hastings, and B. Peacock, *Statistical Distributions*. Hoboken, NJ, USA: John Wiley & Sons, Ltd., 2010, doi: 10.1002/9780470627242.ch1.

[19] D. A. Guimarães, "Robust test statistic for cooperative spectrum sensing based on the Gerschgorin circle theorem," *IEEE Access*, vol. 6, pp. 2445–2456, 2018, doi: 10.1109/ACCESS.2017.2783443.

[20] ——, "Gini index inspired robust detector for spectrum sensing over Ricean channels," *Electronics Letters*, November 2018, doi: 10.1049/el.2018.7375.

[21] ——, "Pietra-Ricci index detector for centralized data fusion cooperative spectrum sensing," *IEEE Trans. Veh. Technol.*, vol. 69, no. 10, pp. 12 354–12 358, 2020, doi: 10.1109/TVT.2020.3009440.

[22] R. Tandra and A. Sahai, "SNR walls for signal detection," *IEEE J. Sel. Topics Signal Process.*, vol. 2, no. 1, pp. 4–17, 2008, doi: 10.1109/JSTSP.2007.914879.

[23] A. Bollig, C. Disch, M. Arts, and R. Mathar, "SNR walls in eigenvalue-based spectrum sensing," *EURASIP journal on wireless commun. and networking : EURASIP JWCN*, vol. 2017, no. 1, p. 109, 2017, doi: 10.1186/s13638-017-0899-y.

[24] D. A. Guimarães, "Complex envelope based modems: A tutorial," *Journal of Communication and Information Systems, JCIS*, vol. 35, no. 1, pp. 34–50, Feb. 2020, doi: 10.14209/jcis.2020.4.

[25] B. Nadler, F. Penna, and R. Garello, "Performance of eigenvalue-based signal detectors with known and unknown noise level," in *IEEE Int. Conf. Communications*, jun 2011, pp. 1–5, doi: 10.1109/icc.2011.5963473.

[26] L. S. Costa, D. A. Guimarães, E. P. Frigieri, and R. A. A. de Souza, "Weighted circular folding cooperative power spectral density split cancellation algorithm," *IEEE Trans. Veh. Technol.*, vol. 70, no. 1, pp. 1062–1066, 2021, doi: 10.1109/TVT.2021.3049707.

[27] S. Zhu, T. S. Ghazaany, S. M. R. Jones, R. A. Abd-Alhameed, J. M. Noras, T. Van Buren, J. Wilson, T. Suggett, and S. Marker, "Probability

[28] N. Reisi, M. Ahmadian, and S. Salari, "Performance analysis of energy detection-based spectrum sensing over fading channels," in *Wireless Communications Networking and Mobile Computing (WiCOM), 2010 6th Int. Conf. on*, Sep. 2010, pp. 1–4, doi: 10.1109/WICOM.2010.5600868.

[29] S. Atapattu, C. Tellambura, and H. Jiang, "Energy detection based cooperative spectrum sensing in cognitive radio networks," *IEEE Trans. Wireless Commun.*, vol. 10, no. 4, pp. 1232–1241, April 2011, doi: 10.1109/TWC.2011.012411.100611.

[30] G. H. Golub and H. A. van der Vorst, "Eigenvalue computation in the 20th century," *J. Comput. Appl. Math.*, vol. 123, no. 1-2, pp. 35–65, Nov. 2000, doi: 10.1016/S0377-0427(00)00413-1.

[31] D. A. Guimarães, G. P. Aquino, R. R. Santos, and A. V. dos Reis, "Performance of eigenvalue-based spectrum sensing with approximate eigenvalue estimation methods," *Journal of Communication and Information Systems, JCIS*, vol. 34, no. 1, pp. 239–247, Oct. 2019, doi: 10.14209/jcis.2019.25.

[32] D. Ramirez, J. Via, I. Santamaria, and L. L. Scharf, "Locally most powerful invariant tests for correlation and sphericity of Gaussian vectors," *IEEE Trans. Inf. Theory*, vol. 59, no. 4, pp. 2128–2141, April 2013, doi: 10.1109/TIT.2012.2232705.

[33] R. B. Chaurasiya and R. Shrestha, "Hardware-efficient VLSI architecture and ASIC implementation of GRCR-based cooperative spectrum sensor for cognitive-radio network," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, pp. 1–11, 2021, doi: 10.1109/TVLSI.2021.3114859.

[34] Y. Zeng and Y.-C. Liang, "Spectrum-sensing algorithms for cognitive radio based on statistical covariances," *IEEE Trans. Veh. Technol.*, vol. 58, no. 4, pp. 1804–1815, 2009, doi: 10.1109/TVT.2008.2005267.

[35] Altair Engineering, Inc. (former Visual Solutions, Inc.), "VisSim: A graphical language for simulation and model-based embedded development," Feb. 2022. [Online]. Available: http://www.vissim.com/products/vissim/comm.html.

[36] D. Guimarães and E. Pereira, "Influence of a direct-conversion receiver model on the performance of detectors for spectrum sensing," *Journal of Communication and Information Systems, JCIS*, vol. 36, no. 1, p. 173–183, Nov. 2021, doi: 10.14209/jcis.2021.19.

[37] D. A. Guimarães, "Matlab code for performance analysis of detectors for centralized data fusion cooperative spectrum sensing," Dec. 2021. [Online]. Available: https://www.inatel.br/docentes/dayan/easyfolder/TP553/SS_Performance_TP553.m.

[38] T. An, D. Kim, I. Song, M. S. Yeu, H.-K. Min, S. Lee, and W. Lee, "Cooperative spectrum sensing based on generalized likelihood ratio test under impulsive noise circumstances," in *MILCOM 2012 - 2012 IEEE Military Commun. Conf.*, 2012, pp. 1–6, doi: 10.1109/MILCOM.2012.6415621.

[39] K. B. Letaief and W. Zhang, "Cooperative communications for cognitive radio networks," *Proceedings of the IEEE*, vol. 97, no. 5, pp. 878–893, May 2009, doi: 10.1109/JPROC.2009.2015716.

[40] D. A. Guimarães, E. J. T. Pereira, A. M. Alberti, and J. V. Moreira, "Design guidelines for database-driven Internet of Things-enabled dynamic spectrum access," *Sensors*, vol. 21, no. 9, 2021, doi: 10.3390/s21093194.

[41] D. A. Guimarães and R. A. A. de Souza, "Implementation-oriented model for centralized data-fusion cooperative spectrum sensing," *IEEE Comm. Lett.*, vol. 16, no. 11, pp. 1804–1807, 2012, doi: 10.1109/LCOMM.2012.092112.121614.

[42] M. Wax, and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 33, no. 2, pp. 387–392, Apr. 1985, doi: 10.1109/TASSP.1985.1164557.

distribution of Rician K-factor in urban, suburban and rural areas using real-world captured data," *IEEE Trans. Antennas Propag.*, vol. 62, no. 7, pp. 3835–3839, Jul 2014, doi: 10.1109/TAP.2014.2318072.

Dayan Adionel Guimarães received the MSc and the PhD in Electrical Engineering from the State University of Campinas (Unicamp), Brazil, in 1998 and 2003, respectively. He is a Researcher and Senior Lecturer in the National Institute of Telecommunications (Inatel), Brazil. His research focuses the general aspects of wireless communications, specifically radio propagation, digital transmission, dynamic spectrum access, and convex optimization and signal processing applied to telecommunications.