Wideband Spectrum Sensing technique based on Goodness-of-Fit testing

Bart Scheers, Djamel Teguig and Vincent Le Nir Dept. Communication, Information Systems & Sensors (CISS) Royal Military Academy Brussels BELGIUM.

E-mail: {bart.scheers, djamel.teguig, vincent.lenir}@rma.ac.be

Abstract—Recently, Goodness-of-Fit (GoF) based spectrum sensing has been proposed as a method for blind narrowband spectrum sensing. GoF based spectrum sensing has the nice feature that, compared to Energy Detection, it needs fewer samples to achieve the same sensing performance. In this paper, we present how GoF based spectrum sensing can be integrated in a conventional wideband spectrum sensing scheme, resulting in an accurate technique with a very short sensing time, which makes this method attractive for cognitive radio applications.

Keywords—Wideband spectrum sensing, Goodness-of-Fit based spectrum sensing, cognitive radio.

I. INTRODUCTION

Wideband spectrum sensing is about efficiently sensing across a very large frequency band, typically much larger than the bandwidth of the waveforms that are used or that are to be detected. Wideband spectrum sensing can be categorized into two types: Nyquist wideband sensing, also referred to as conventional wideband spectrum sensing, and sub-Nyquist wideband sensing, depending whether the sampling rate at which the signals are acquired is above or below the Nyquist rate. Recent surveys of both approaches can be found in [1], [2] and [3]. The method described in this paper is to be classified as a conventional wideband spectrum sensing method. Typically in conventional wideband spectrum sensing schemes, a time-frequency domain transformation (e.g. filter-bank, DFT or wavelet transformation) is used to divide a wide frequency band into a number of small frequency bins. Then, Energy Detection (ED) or other sensing algorithm can be used to determine the presence of a signal within these frequency bins. In practical cases, the spectrum sensing algorithm following the DFT is a non-coherent or blind sensing method as it is not straightforward to have a priori knowledge of the all potentially present signals over the spectrum band or to align the frequency bins with all possible signals.

Wideband spectrum sensing is challenging. The problem to solve is how to perform - accurate spectrum sensing - over a very large frequency band - in a reasonably short sensing time, as those requirements are contradictory.

Recently, GoF based spectrum sensing has been proposed as a method for blind spectrum sensing [4]. The major advantage of GoF based spectrum sensing over other blind sensing methods, like ED, is that it needs fewer sample for the same sensing performance, resulting in a short sensing time. Further, the proposed method can easily be adapted for any type of noise distributions, other than AWGN, as long as the noise distribution is known. GoF based sensing methods are also less influenced by typical spectrum sensing impairments like non-Gaussian noise and noise uncertainty [5].

In this paper, we are mainly interested in the short sensing time property of the method. We will describe how we can apply GoF based spectrum sensing on the Fourier coefficients of the power spectrum of the received samples, leading to a very accurate and time efficient wideband sensing method.

The remainder of this paper is organized as follows. In Section II, the GoF based spectrum sensing method is presented. In section III, the distribution of the Fourier Coefficients of AWGN is studied, as we will need it for the GoF testing. The wideband spectrum sensing method integrating the GoF based sensing is described in section IV. Finally we will discuss some examples emphasising the strengths of the proposed method.

II. GOF BASED SPECTRUM SENSING

The GoF test is a blind nonparametric hypothesis test problem which can be used to detect the presence of signals in noise by determining whether the received samples are (are not) drawn from a distribution with a Cumulative Distribution Function (CDF) F_0 , representing the noise CDF. The hypothesis to be tested can be formulated as follows:

$$\begin{cases} H_0: F_n(x) = F_0(x) \\ H_1: F_n(x) \neq F_0(x), \end{cases}$$
(1)

where F_0 represents the hypothesized CDF (as already mentioned, in the application of sensing this will be the noise CDF). $F_n(x)$ is the empirical CDF of the received sample and can be calculated by:

$$F_n(x) = |\{i : x_{(i)} \le x, 1 \le i \le n\}/n|,$$
(2)

where |.| indicates cardinality, $x_{(i)}$ are the ordered samples under test $(x_1 \leq x_2 \leq ... \leq x_n)$ and *n* represents the total number of samples.

In literature, there are different GoF based tests proposed for spectrum sensing. The most important ones are the Kolmogorov- Smirnov test [7], the Cramer-Von Mises test [9], the Anderson-Darling test [6] and order statistics [8]. All these tests are based on the hypothesis test as formulated in (1), but differ in the way the test statistic, i.e. the distance between the empirical cumulative distribution of the observations made locally at the CR user and the noise CDF $F_0(x)$, is calculated. The calculated distance is compared with a threshold to decide

whether the signal is present or not, given a certain probability of false alarm.

The GoF test based spectrum sensing was first presented in [6]. It is based on the Anderson-Darling (AD) GoF test to decide whether the received samples are drawn from the noise CDF F_0 (Gaussian CDF) or an alternative CDF. Authors in [6], show by simulations that AD-sensing outperforms the ED-sensing at low SNR. The sensing method takes as a normal distribution noise CDF F_0 for the GoF test, meaning that they assume that the samples of the received signal are real valued. However, in cognitive radios (CR), based on the SDR technology, this is a limitation, as the radio receives complex valued baseband IQ samples in the digital domain. In [4], this limitation is overcome by considering the energy of the received samples and test them against a chi-square distribution with 2 degrees of freedom. The sensing method can be summarized as follows: Consider the classical binary hypothesis test

$$\begin{cases} H_0 : x_i = n_i \\ H_1 : x_i = s_i + n_i, \end{cases}$$
(3)

where s_i are the received complex samples of the transmitted signal and n_i is the complex Gaussian noise. We now consider the random variable $Y_i = |x_i|^2$ which corresponds to the received energy. It is known that, if the real and the imaginary part of x_i are normally distributed, which is the case under H_0 hypothesis, the variable $Y_i = |x_i|^2$ is chi-squared distributed with 2 degrees of freedom. The spectrum sensing problem can now be reformulated as an hypothesis represented in (1) where we will test whether the received energy $Y_i = |x_i|^2$ are drawn from a chi-square distribution with 2 degrees of freedom or not.

Without loss of generality, we will use from now on the Anderson-Darling (AD) test to perform the GoF testing, as this test seems to perform better than other tests proposed in literature [4]. The AD test statistic was proposed by T. W. Anderson and D. A. Darling in 1952 [10].

$$\frac{1}{n}A_n^2 = \int_0^1 [F_n(x) - F_0(x)]^2 \psi[F_0(x)] dF_0(x), \qquad (4)$$

This test statistic is based on the Von Mises criterion, which is an average of the squared discrepancy $[F_n(x) - F_0(x)]^2$, weighted by the increase in $F_0(x)$. The authors in [10] proposed an extra weighting function $\psi(F_0(x))$ to give more importance to the tails of the CDF.

$$\psi[F_0(x)] = \frac{1}{F_0(x)[1 - F_0(x)]}$$
(5)

To numerically calculate the test statistic given in (4), one can to break down the integral as

$$\frac{1}{n}A_n^2 = \int_0^{z_1} \frac{F_0^2(x)}{F_0(x)[1-F_0(x)]} dF_0(x) + \int_{z_1}^{z_2} \frac{[F_n(x) - F_0(x)]^2}{F_0(x)[1-F_0(x)]} dF_0(x) + \dots + \int_{z_n}^1 \frac{[1-F_0(x)]^2}{F_0(x)[1-F_0(x)]} dF_0(x),$$
(6)

with $z_i = F_0(x_i), i = 1 ... n$.

By straightforward integration and collection of the terms, equation (6) can be written as [11]

$$A_n^2 = -n - \frac{\sum_{i=1}^n (2i-1)(\ln F_0(x_{(i)}) + \ln(1 - F_0(x_{(n+1-i)})))}{n}$$
(7)

with $x_{(i)}$ the ordered samples and F_0 the hypothetical noise energy distribution. In case of complex Gaussian noise, F_0 represents the CDF of a chi-square distribution given by:

$$F_0(y) = 1 - e^{-y/2\sigma_n^2} \sum_{k=0}^{m-1} \frac{1}{k!} (\frac{y}{2\sigma_n^2})^k, y > 0, \qquad (8)$$

where m is the degree of freedom (in our case m=2) and σ_n^2 is the noise power of the real and the imaginary part. It is assumed that the noise power of both parts equal 1.

Once the test statistic A_n^2 is computed, it will be compared to a predefined threshold λ :

$$\begin{cases} H_0: A_n^2 \le \lambda\\ H_1: A_n^2 > \lambda, \end{cases}$$
(9)

The value of λ is determined for a specific value of P_{fa} . A table listing values of λ corresponding to different false alarm probabilities P_{fa} is given in [12]. Otherwise, these values can be computed in advance by Monte Carlo approach.

As mentioned in the introduction, one of the nice features of GoF based spectrum sensing is that it need fewer samples than ED to achieve the same sensing performance in terms of probability of detection versus probability of false alarm. To illustrate this, we present in figure 1 the detection probability P_d versus SNR for the AD detector and the ED and this for different values of n, the total number of samples. The P_{fa} for both detectors is set to 0.01 and the SNR varies from -15dB to 10dB. The curves are obtained by 10000 Monte-Carlo simulations. As one can see, the detection performance of the AD detector for n = 40 is comparable to the detection performance of the ED for n = 100. In other words, for the same detection performance, the AD detector will yield a gain of a factor 2.5 in sensing time, compared to ED.

In this paper, we will integrate the GoF based spectrum sensing as described above in a conventional wideband sensing scheme. To this end, we will not perform the test on the energy of the received samples, but on the Fourier Coefficients of the power spectrum of the received samples. Therefore we need first to describe the distribution of the Fourier Coefficients of an AWGN.

III. FOURIER COEFFICIENTS DISTRIBUTION OF AWGN

Consider a complex Gaussian noise vector $\mathbf{x} = \{x_n\}$ of length N, with x_n normal distributed iid complex samples of



Fig. 1. Detection probability versus SNR for AD detector and ED, $P_{fa} = 0.01, n = 40$ and 100.

zero means and variance σ^2 . The Discrete Fourier coefficients of the noise vector are given by

$$X_k = \sum_{n=0}^{N-1} x_n e^{-j2\pi \frac{kn}{N}} \ k = 0 \cdots N - 1 \tag{10}$$

From probability theory, it is known that a weighted sum of Gaussian random variables remains Gaussian. In other words, the Fourier coefficient X_k for any given frequency bin k will be Gaussian distributed. As x_n has a zero means, the mean value of X_k equals zero. It is also known from probability theory that if X and Y are two independent random variables with variance σ^2 and τ^2 respectively, then the random variable X + Y has variance $\sigma^2 + \tau^2$. Further, if X is a random variable with variance σ^2 , then the random variable sX, with s some scalar constant, has variance $s^2\sigma^2$. Hence, the variance of X_k can be calculated as:

$$var(X_k) = \sum_{n=0}^{N-1} |e^{-j2\pi\frac{kn}{N}}|^2 \sigma^2$$
(11)

$$= N\sigma^2, \tag{12}$$

with σ^2 the power of the complex noise. In summary, a complex Gaussian noise process $x_n \sim N(0, \sigma^2)$ produces complex Gaussian Fourier coefficients $X_k \sim N(0, N\sigma^2)$, with N the length of the DFT.

Consequently, we can also state that the k^{th} power spectrum coefficient $|X_k|^2$, normalized by $var(X_k)/2$ follows a χ_2^2 distribution,

$$\frac{2|X_k|^2}{N\sigma^2} \sim \chi_2^2.$$
 (13)

The factor $2/N\sigma^2$ in (13) comes from the fact that we need to normalize the real and the imaginary part of the Gaussian Fourier coefficients X_k to a ~ N(0,1) distribution (i.e. the power of the complex Gaussian Fourier coefficients needs to be normalized to 2), as a χ_2^2 distribution is defined as the distribution of the sum of the squares of 2 independent standard normal variables.

IV. WIDEBAND SENSING METHOD BASED ON GOF

The observation in (13) can now be used to integrate the narrowband spectrum sensing based on GoF in a conventional wideband spectrum sensing scheme. The wideband sensing method is represented in figure 2. The received signal is sampled at the Nyquist rate, leading to a sequence of complex samples x_i . In the most general case, the received signal can be represented as the sum of some unknown narrowband signals plus noise.

$$x_i = \sum_j s_{j,i} + n_i, \ j = 0 \cdots J,$$
 (14)

with $s_{j,i}$ the samples of a narrowband signal j, present in the frequency band of interest and n_i an AWGN with variance σ^2 .



Fig. 2. Wideband sensing method block diagram.

An N-point DFT is calculated on K consecutive segments, resulting in a sequence $\{X_k\}$ of length K for every frequency bin. We assume that in every frequency bin, there is either only noise present (the H_0 hypothesis) or either there is a signals with zero-mean and unknown variances plus noise present (the H_1 hypothesis). Under the H_0 hypothesis, the normalised power spectrum coefficient $\frac{2|X_k|^2}{N\sigma^2}$ will follow a χ_2^2 distribution, which can be tested by the narrowband GoF based Spectrum Sensing as presented in section 2, using a chi-square CDF as F_0 . Under the H_1 hypothesis, the sequence X_k of length Kwill fail the GoF test.

The strength of this method is that only a few DFTs (e.g. K = 20 - 50) are needed to result in an accurate wideband spectrum sensing as the GoF test only needs few samples [4]. This nice feature of the GoF test is amplified in a conventional wideband spectrum sensing scheme, resulting in a total sensing time equivalent to KN samples.

Just like in other DFT based wideband spectrum sensing methods, windowing and overlapping can be used prior to the DFT to improve selectivity between the frequency bins and to decrease further the total number of samples for sensing. Note however that windowing will affect the noise power, hence we need to adapt the normalisation factor of the power spectrum coefficients. Consider a discrete-time window $\{w_n\}$ of length N. Following an equivalent reasoning as for (12), the variance of the Fourier coefficients X_k of a Gaussian noise vector $\{x_n\}$, point by point multiplied by a window $\{w_n\}$ can be written as:

$$var(X_k) = \sum_{n=0}^{N-1} w_n^2 \sigma^2,$$
 (15)

with σ^2 the power of the complex noise. This means that if a window $\{w_n\}$ is used, the normalisation factor, prior to the GoF test, needs to be taken

$$\frac{2|.|^2}{\sum_{n=0}^{N-1} w_n^2 \sigma^2}.$$
(16)

V. DISCUSSION

In this section, we will discuss some results of the proposed method on synthetic data. The aim of this section is not to compare the performance of the proposed method to a classic Energy Detection (ED) method. A comparison between GoF based spectrum sensing and ED can be found in [4]. Although [4] is about narrowband sensing, the results can be generalizes to wideband spectrum sensing.

The first simulation is mainly to illustrate the proposed method. For this, only one narrow-band signal, with high SNR, will be present in a 10MHz frequency band of interest. The incoming signal, $\{x_i\}$ as represented in figure 2, is a complex baseband signal, sampled at 10 MHz. The complex noise in the band is consider AWGN and has a noise power density of 0dBm/Hz. The parameters for the wideband sensing algorithm are the following: the number of consecutive segments K = 40 and the number of points for the DFT N = 1024, meaning that the wideband sensing is performed on K.N = 40960 samples, corresponding to a sensing time of 4.096ms. With these values, the width of the frequency bins equals approximatively 10kHz. The P_{fa} for the sensing is set to 1%, which corresponds to a threshold $\lambda = 3.89$ [12]. The signal to detect is a BPSK modulated signal at 3 MHz and a bandwidth of 25kHz. The modulated symbols are shaped using a RRC pulse shape with $\alpha = 0.5$. The power of the modulated signal is set to obtain an SNR of 10dB. Note that the SNR-value is calculated by only taking into account the noise power within the noise bandwidth of the modulated signal. As the BPSK signal has a 3dB-bandwidth of 25kHz, it can be expected that the signal will be detected in 3 frequency bins.

In figure 3 we show the empirical CDF for every frequency bin. The empirical CDFs corresponding to a bin in the H_0 hypothesis are in blue. The empirical CDFs corresponding to a bin in the H_1 hypothesis are in red. The reference CDF F_0 for the GoF test corresponding to a χ_2^2 CDF is represented in green. One can clearly distinguish the 3 empirical CDFs corresponding to the BPSK signal. The empirical CDFs in red, close to the F_0 CDF are false alarms.

The probability of detection P_d for the BPSK signal in this scenario equals almost 1. This can been seen in figure 1, on the curve that represents the P_d of the AD detector for n = 40 and $P_{fa} = 1\%$, i.e. the same parameters that are used in the first simulation. For SNR > 1dB the P_d equals almost 1, meaning that it is almost certain that we will detect the BPSK signal



Fig. 3. Empirical CDF for every frequency bin: in blue the CDFs in the H_0 hypothesis, in red the CDFs in the H_1 hypothesis. The CDF F_0 is represented in green.

of the first scenario in the frequency bins where the signal is present with SNR > 1dB.

In a second scenario, we have the same set-up, but now two modulated signals, at a much lower SNR, are present in the 10MHz band of interest. The BPSK modulated signal, still at 3 MHz, is now at SNR = 0dB. The second signal is a DAB mode-I signal, centred around 7 MHz and at an SNR = -5dB. The DAB mode-I signal is based on an OFDM modulation with 1536 sub-carriers and a sub-carrier spacing of 1kHz. The bandwidth equals approximately 1.54MHz. The parameters for the wideband sensing algorithm are kept identical to the previous simulation. Figure 4 shows the result of the proposed wideband spectrum sensing algorithm. An '1' on the y-axis means that the frequency bin is found to be in the H_1 hypothesis, otherwise it is in the H_0 hypothesis.

As one can see, most of the frequency bins where a modulated signal is present are tagged as occupied (H_1 hypothesis). The number of false alarms, 8 in total, is in accordance with the pre-setted value of P_{fa} . The OFDM signal is not detected over it's complete band. This is due to the probability of detection of the GoF based spectrum sensing. An OFDM modulated signal is known to have a quite flat power spectral density. It can therefore be assumed that the SNR in each frequency bin within the bandwidth of the OFDM signal will have a SNR = -5dB and hence a $P_d = 0.2$, meaning that approximately 1 out of 5 bins will be tagged as occupied.

The latter simulation shows the strength of the wideband GoF spectrum sensing. Only on a few values (K = 40), an accurate sensing result can be obtained. The total number of samples for the wideband spectrum sensing method equals K.N. For a given sensing time, because K can be kept small in this method, N can be increased, resulting in a method with a good frequency domain resolution.



Fig. 4. Wideband sensing result on the 2 low SNR signals: N = 1024, K = 40, $\lambda = 3.89$.

VI. RESULT ON REAL DATA

In the previous section, the wideband AD detector was applied on synthetically generated data. In this section we will show the result of the wideband spectrum sensing technique on real data. For this, a 100 ms data segment was captured in the region of Brussels using a real-time spectrum analyser from Agilent (the N9020A MXA Signal Analyser). The frequency range was set from 50MHz to 108MHz, covering the majority of the military VHF band, including the FM broadcast band. The result is shown in figure 5. The upper subplot of the figure represents the power spectral density of the complete captured signal. The number of DFT-bins N is set to 8192 and a Hamming window is used. The middle subplot shows the waterfall representation of the power spectrum. The vertical lines in the waterfall represent signals present in the spectrum. One can see very clear the FM broadcast band, but also other weaker signals in the first part of the spectrum band. The lower subplot of figure 5 represents the result of the wideband AD detector. For the AD detector, the following parameters are used. The number of DFT-bins N is also set to 8192, leading to a frequency resolution of 9KHz, and a Hamming window is used to limit the leakage between frequency bins. K equals 40 and the P_{fa} is set to 0.01. Given the sample time of 0.1347 ns, the total sensing time for one wideband AD detection equals $4.4 \, ms$. The estimation of the noise power is done based on the Long-Term Gaussian noise power estimation method as described in [13]. The advantage of this method is that it can estimate the noise power in the presence of signals. On the complete captured dataset of 100 ms, a total of 22 nonoverlapping wideband AD detections could be performed. The result of these 22 detections are presented under a waterfall representation in the lower subplot of figure 5. The red color corresponds to the H1 hypothesis. A red vertical line in the waterfall means that the signal is all the time detected, vertical blue lines represent spectrum holes.

VII. CONCLUSION

In this paper, we have introduced a new wideband spectrum sensing scheme. The method is based on a conventional wideband spectrum sensing approach, using a DFT to divide the frequency band into small frequency bins. On each frequency bin a narrowband GoF based spectrum sensing algorithm is applied. It was shown that, if in a frequency bin only noise is present, the power coefficients will follow a chi-square distribution, which is tested by a GoF test using the Anderson-Darling test statistic. If the test fails, it is considered that also a signal is present in the frequency bin.



Fig. 5. Wideband sensing result on real data: (a) Power spectral density of the captured signal, (b) Waterfall of the power spectrum, (c) Waterfall plot of the wideband AD detector result, N = 8192, K = 40, $\lambda = 3.89$. The red color corresponds to the H1 hypothesis.

The narrowband GoF based spectrum sensing has the characterised by a short sensing time as it can take accurate decisions on only a few samples, typically less than 50. This feature is emphasised in a classical wideband spectrum sensing approach. An accurate decisions per frequency bin can be made after only a few DFTs. Typically, compared to a classical ED based wideband method, the gain in sensing time for the same performances is in the order of 2.5. This also implies that, compared to ED based wideband sensing methods, an increase of points for the DFT, resulting in a better frequency domain resolution, is less penalising in terms of sensing time.

These features, i.e. good sensitivity, together with a relatively small sensing time, while still maintaining a reasonable resolution in the frequency domain, makes this method interesting for cognitive radio applications.

REFERENCES

- Hongjian Sun and Arumugam Nallanathan: 'Wideband Spectrum Sensing for Cognitive Radio Networks: A Survey', IEEE Wireless Communications, Vol. 20, no. 2, pp. 74-81, April 2013
- [2] E. Axell, G. Leus, E. G. Larsson and H. V. Poor: 'Spectrum sensing for cognitive Radio : State-of-the-art and recent advances', IEEE Signal Processing Mag., vol. 29, no. 3, pp. 101-116, May. 2012.

- [3] D.D. Ariananda, M.K. Lakshmanan and H. Nikoo: 'A survey on spectrum sensing techniques for Cognitive Radio', Second International Workshop on Cognitive Radio and Advanced Spectrum Management CogART 2009, May 2009.
- [4] D. Teguig, V. Le Nir and B. Scheers : 'Spectrum sensing method based on goodness of fit test using chi-square distribution', Electronics Letters, Volume 50, Issue 9, 24 April 2014, p. 713-715.
- [5] D. Teguig, V. Le Nir, B. Scheers and F. Horlin: 'Spectrum Sensing Method Based on the Likelihood Ratio Goodness of Fit Test under Noise Uncertainty', International Journal of Engineering Research & Technology (IJERT), Volume 3(9), September 2014, p.488-494.
- [6] H. Wang, E. H. Yang, Z. Zhao and W. Zhang: 'Spectrum sensing in cognitive radio using goodness of fit testing', IEEE Trans. Wireless Commun., vol. 8, no. 11, pp. 5427-5430, Nov. 2009.
- [7] G. Zhang, X. Wang, Y. C. Liang and J. Liu: 'Fast and robust spectrum sensing via kolmogorov-smirnov test', IEEE Trans. Commun. vol. 58, no. 12, pp. 3410-3416, Dec. 2010.
- [8] S. Rostami, K. Arshad and K. Moessner: 'Order-Statistic Based Spectrum Sensing for Cognitive Radio', IEEE Commun. Lett.,vol. 16, no. 5, pp. 592-595, May. 2012.
- [9] T. Kieu-Xuan, I. Koo: 'Cramer-von Mises test spectrum sensing for cognitive radio systems', Wireless Telecommunication Symposium, 2011.
- [10] T. W. Anderson, D. A. Darling: 'Asymptotic Theory of Certain "Goodness of Fit" Criteria Based on Stochastic Processes', Annals of Mathematical Statistics, Vol. 23, pp.193212 ,1952.
- [11] T. W. Anderson, D. A. Darling: 'A Test of Goodness of Fit', Journal of the American Statistical Association, Vol. 49, No. 268., pp. 765-769, Dec. 1954.
- [12] M. A. Stephens: 'EDF Statistics for Goodness of Fit and Some Comparisons', Journal of the American Statistic Association, vol.69, No.347, pp.730-737, Sep. 1974.
- [13] S. Couturier, D. Rauschen: 'Energy Detection Based on Long-Term Estimation of Gaussian Noise Distribution', 8th Karlsruhe Workshop on Software Radios, Mar. 2014.