Computationally Efficient Wideband Spectrum Sensing through Cumulative Distribution Function and Machine Learning

Jakub Nikonowicz, *Member, IEEE,* Mieczysław Jessa, *Member, IEEE,* and Łukasz Matuszewski, *Member, IEEE*

***Abstract*—Blind spectrum sensing (BSS) is crucial for identifying unknown signals in scenarios with limited prior knowledge. Traditional methods face challenges with unknown and time-varying signals, especially in the presence of noise interference. This paper addresses these issues by introducing a statistical signal processing framework that extends the use of machine learning (ML) features. Our approach improves BSS by incorporating cumulative distribution functions (CDFs) into unsupervised ML, enabling effective clustering of diverse transmission states without assumptions about specific noise distributions. Additionally, we introduce a temporal decomposition technique using shorter Fast Fourier Transforms (FFTs), enhancing the learning process, reducing system inertia, and minimizing data requirements for retraining under dynamic conditions. We evaluate our method, focusing on various features/approaches for incorporating CDFs into ML, including centroid, linear approximation, and low-order statistics. Simulation results demonstrate robust detection in a standard transmission scenario with a Gaussian pulse amidst additive white Gaussian noise, maintaining a consistently low false alarm rate. These findings highlight our BSS approach's effectiveness and practical potential in handling unknown signals in challenging environments. This research provides valuable insights, laying the groundwork for practical implementation in real-world scenarios.**

***Index Terms*—Blind detection, cumulative distribution function, machine learning, spectrum sensing, unknown signals.**

I. Introduction

BLIND spectrum sensing (BSS) is a well-known approach for detecting unknown signals in scenarios where prior knowledge about the signal is minimal, inaccessible, or missing. Conventional detection methods often encounter operational deficiencies when faced with unknown, dynamic signals immersed in significant noise and interference levels. A primary challenge of blind detection lies in the demand for extensive datasets, high computational costs, and less precision compared to methods equipped with prior signal knowledge. Nonetheless, the capability to detect unknown signals holds significant value across diverse applications.

In domains such as wireless communications and cognitive radio networks, detecting and tracking signals from multiple sources is crucial to avert disruptive interference [1]. In radio astronomy, detecting signals from distant, unknown sources is an everyday necessity [2]. Passive radar systems, for instance, rely on naturally occurring signals like TV or FM broadcasts as sources of reflection rather than active transmitters [3]. In medical imaging, the task often involves detecting and locating subtle signals within vast datasets. These challenges are not confined to a single field but have broader relevance.

Various advanced and well-established techniques have found application across this interdisciplinary domain. Among them, subspace methods, notably Principal Component Analysis (PCA) and Independent Component Analysis (ICA) are popular choices for blind spectrum sensing. These techniques seek to identify a lower-dimensional subspace that captures the most critical signal features [4]. Although subspace methods excel at disentangling mixed or correlated signals and are relatively straightforward to implement, they demand substantial data and can be sensitive to noise and outliers. Cyclostationary feature detection, on the other hand, leverages the cyclostationarity property of signals, which denotes statistical dependencies between a signal and its time-shifted versions. This approach is robust to noise and interference but applies primarily to narrowband signals [5]. In recent times, sparse representations, including compressive sensing, have emerged as a BSS technique that aims to represent signals as sparse linear combinations of basis functions. This method excels at extracting meaningful signal information, even in the presence of noise and interference; however, its effectiveness is conditional upon the availability of substantial data and the selection of appropriate basis functions [6].

In recent years, BSS has witnessed a surge in interest, resulting in the developing of novel algorithms and techniques to enhance signal detection performance and robustness. Notably, machine learning (ML) approaches have garnered substantial attention [7], [8], [9], [10], [11]. These methods deploy ML algorithms capable of adapting to the characteristics of signals and noise. However, two significant challenges persist in ML-based BSS research:

- The development of versatile ML features that can handle non-stationary and non-linear signals, independent of their characteristics, is of pivotal importance. These features should facilitate the detection of unknown signals without prior knowledge of their frequency, modulation, or structure, and they should adapt to signals with varying

Manuscript received December 1, 2023; revised January 18, 2023. Date of publication January 31, 2024. Date of current version January 31, 2024.

All authors are with the Faculty of Computing and Telecommunications, Poznań University of Technology, Poznań, Poland (e-mails: {jakub.nikonowicz, mieczyslaw.jessa, lukasz.matuszewski}@put.poznan.pl).

This paper was presented in part at the International conference on Software, Telecommunications and Computer Networks (SoftCOM) 2023.
Digital Object Identifier (DOI): 10.24138/jcomss-2023-0175

characteristics over time.

- The substantial data requirements for achieving high accuracy using machine learning-based methods can pose significant limitations, especially in scenarios with limited data or computationally intensive detection tasks.

Addressing these challenges is vital for advancing BSS with ML techniques.

This paper presents a significant advancement in the field of blind spectrum sensing. Building upon the research presented in [12], our work investigates a novel statistical signal processing method that incorporates cumulative distribution functions (CDFs) into the ML process and employs unsupervised ML. The model effectively discriminates between statistically distinct states without making specific assumptions about noise distribution. The temporal decomposition technique enhances learning by utilizing multiple shorter Fast Fourier Transforms (FFTs) within a single time frame, reducing system inertia and minimizing data requirements for model retraining under changing propagation conditions. This study extends prior work by streamlining the procedure's most intricate step/node—simplifying the complex centroid determination in CDF assessment and offering considerations for alternative approaches.

The remainder of this paper is structured as follows: Section II introduces the ML-based Blind Spectrum Sensing (BSS) system, offering an overview of the commonly used measurements and features. Section III introduces the system model and the statistical basis for CDF-based detection. Section IV presents the CDF-based approaches to ML-based detection, including simulation results on CDF measurements in BSS and a discussion of the improvements achieved through various CDF-processing techniques. Finally, in Section VI, we summarize the findings and outline potential trajectories for future research.

II. Machine Learning in BSS

Machine learning (ML) has emerged as a potent tool for signal detection in recent years [7], [8], [9], [10], [11]. Its ability to extract meaningful information from signals, even in noisy environments, without prior knowledge of the signal's characteristics, has become a focal element. ML employs various features to classify the presence of unknown signals in the noise. Common features for spectrum sensing in ML encompass power spectral density, autocorrelation, and cyclostationary features like cyclic correlation and cyclic spectral density. Additional features derive from signal statistics, such as mean, variance, skewness, and kurtosis. Time-domain features, such as energy, entropy, and correlation coefficients between subcarriers, are also applied, contingent on the specific application and signal characteristics [8]. In some instances, a combination of multiple features enhances the algorithm's accuracy, especially when dealing with non-stationary and non-linear signals.

ML algorithms enhance detectors' decision-making capabilities through efficient information inference. Zhang et al. in [9] combined unsupervised and supervised learning for spectrum sensing under varying transmit powers. The method learns transmission patterns and statistics through a modified K-means algorithm during a learning phase, subsequently distinguishing energy feature vectors with support vector machines (SVM). Xiao et al. in [11] introduced a random forest spectrum sensing algorithm for signal recognition in low signal-to-noise ratio (SNR) settings. Awe et al. in [10] developed supervised and semi-supervised learning algorithms using the eigenvalues of the received signal covariance matrix as features. Vyas et al. in [13] adopted a binary classification-based artificial neural network (ANN) with received signal energy and likelihood ratio test statistics at different SNRs. In the work of Tian et al. [14], received signal power and cyclic prefix-induced correlation are used as features. Khalfi et al. in [15] proposed a wideband detection scheme that leverages regression and compressive sampling techniques for improved detection performance.

To further enhance ML-based detection, three approaches are considered:

- In the study by Mikaeil et al. [16], a cooperative sensing algorithm utilizes ML. The classifier undergoes initial training with a dataset containing energy test statistics and their corresponding decisions on signal presence or absence. It is then used to predict new decisions based on new energy test statistics. Li et al. in [17] introduced an SVM-based model for cooperative sensing, optimizing signal grouping to reduce cooperation overhead and improve spectrum sensing performance.
- Hybrid models combine various modeling approaches, such as physical models, data-driven models, or rule-based models, to leverage the strengths of each. This leads to a more profound understanding of the system and more precise decision-making. An example of a hybrid solution is the energy/entropy-based model in [18], which combines physical-driven detection and data-driven entropy. These feature as vectors for the classifier.
- Decision functions play a vital role in hybrid modeling. Nikonowicz et al. in [19] underscore the importance of integrating decision-making into the modeling process. This involves identifying optimal solutions based on model outputs and considering factors like cost, feasibility, and environmental impact. In [20], various ML techniques, including SVM, random forest, decision tree, K-NN, logistic regression, and NBC, are trained, validated, and tested.

It is crucial to note that while integrating ML into hybrid solutions with traditional spectrum sensing features can significantly enhance spectrum sensing performance, it may lead to increased processing time and implementation complexity. The trade-off between model accuracy and processing efficiency becomes a key consideration. Furthermore, dimensionality reduction techniques, as discussed in [21], can offer substantial benefits by reducing data complexity without compromising accuracy, which is an important aspect to consider in optimizing data utilization.

III. System Model

Incorporating domain-specific knowledge into detection models is a challenging task, particularly in the face of complex spectral signals influenced by various factors like ever-changing environmental conditions and frequency-dependent propagation and attenuation. The presence of non-uniform signal sources further adds complexity to the detection process. Statistical feature analysis has emerged as a promising research direction to tackle these intricacies. This approach avoids the need for pre-determined signal parameterization, opting to identify external signals as anomalies or rare occurrences within what is initially assumed to be a homogenous noise environment. The efficacy of this methodology has been affirmed in prior investigations [22], [12].

Furthermore, Scheers et al. in [23] illustrate integrating a Goodness-Of-Fit-based spectrum sensing approach into a conventional wideband spectrum sensing framework. Their work highlights a precise technique with a short sensing time, providing empirical evidence for the credibility of the statistical approach in the context of wideband signals.

Additive white Gaussian noise (AWGN) is a prevalent choice for modeling channel distortion in radio environments, and numerous motivations can justify this preference. First and foremost, practical radio channels contend with a wide range of noise sources, ranging from thermal to atmospheric and man-made distortions. The central limit theorem, a cornerstone of statistics, posits that the collective sum of numerous independent random variables converges toward a Gaussian distribution. In radio environments, the independence of multiple noise sources, treated as random variables, collectively contributes to the overall noise on a channel. This, in turn, establishes AWGN as a fitting model for encapsulating the combined noise in a radio channel. Secondly, AWGN aligns with well-defined statistical parameters that closely approximate the characteristics of various real-world noise sources [24]. With its characteristic bell-shaped curve, the Gaussian distribution lends itself to easy mathematical analysis. Its property of closure under convolution simplifies the assessment of noise's impact on a signal. Furthermore, it is worth noting that Gaussian noise represents a worst-case scenario for many radio systems. If a communication system can effectively operate in the presence of AWGN, it will likely demonstrate robust performance when confronted with other, potentially more complex noise sources. Therefore, investigating the effects of AWGN provides valuable insights for designing communication systems that exhibit resilience against a diverse array of noise sources [25], [26].

In the system model under consideration, we introduce the following key variables:

- L: the number of sampling points.
- s_n: the signal under investigation.
- x_n: the deterministic signal targeted for detection.
- z_n: complex Gaussian noise, comprising independent and identically distributed (iid) samples with zero mean and a variance of σ^2.

Furthermore, we establish two critical hypotheses:

- H_0: Corresponding to the scenario where no signal is transmitted.
- H_1: Corresponding to the scenario where a signal is indeed transmitted.

The signal, which is a composite of the actual signal and Gaussian noise, can be expressed as follows:

$$s_n = \begin{cases} z_n, & \text{if } H_0 \\ x_n + z_n, & \text{if } H_1 \end{cases} \tag{1}$$

In order to ascertain the single-sided spectrum $S(k)$ of the signal $s(n)$, the initial step entails applying a Fourier transform to the incoming signal

$$S(k) = \sum_{n=0}^{L-1} s(n) e^{-j2\pi \frac{k}{L}}, k = 0, 1, 2 \ldots \frac{L}{2} - 1, \tag{2}$$

subsequently, leveraging the inherent nature of $s(n)$ as a composite of $x(n)$ and $z(n)$, in conjunction with the properties of the Fourier transform, allows us to compose the following

$$S(k) = X(k) + Z(k), \tag{3}$$

where $X(k)$ and $Z(k)$ denote the Fourier transforms of $x(n)$ and $z(n)$, respectively. Consequently, the expression for the kth bin in the power spectrum manifests as

$$P(k) = \frac{1}{L} \left((X_R(k) + Z_R(k))^2 + (X_I(k) + Z_I(k))^2 \right). \tag{4}$$

The notations $X_R(k)$ and $X_I(k)$ refer to the real and imaginary components of the signal, respectively. Similarly, $Z_R(k)$ and $Z_I(k)$ represent the real and imaginary parts of the noise. The probability distribution of spectrum bins within the states H_0 and H_1 can be ascertained by investigating the statistical characteristics of the received signal $s(n)$ under each hypothesis. This analysis allows for determining how spectral bins are distributed under different signal and noise conditions.

Drawing from the insights provided by [23], we delve into analyzing Fourier coefficients' distribution in the presence of AWGN. As a weighted summation of Gaussian random variables, the Fourier coefficient $Z(k)$ for a specific frequency bin k follows a Gaussian distribution. Consequently, a complex Gaussian noise process, denoted as $z(n)$ and characterized by $z(n) \sim \mathcal{N}(0, \sigma^2)$, results in complex Gaussian Fourier coefficients $Z(k)$. Given that $z(n)$ has a zero mean, it naturally follows that the mean of $Z(k)$ is also zero. In probability theory, it is well-established that when two independent random variables, $Z(i)$ and $Z(j)$, each possess variances of $\sigma^2 i$ and $\sigma^2 j$ respectively, the random variable $Z(i) + Z(j)$ exhibits a variance equal to the sum of their individual variances, i.e., $\sigma^2 i + \sigma^2 j$. Consequently, the variance of $Z(k)$ can be calculated as expounded in [23].

$$\text{var}(Z(k)) = \sum_{n=0}^{L-1} |e^{-j2\pi \frac{kn}{L}}| \sigma^2 = L\sigma^2, \tag{5}$$

where L represents the length of the DFT. Consequently, we can deduce that the Fourier coefficients follow a Gaussian distribution $\sim \mathcal{N}(0, L\sigma^2)$. Furthermore, this information leads us to an important observation regarding the k-th power spectrum coefficient $P(k)$. It follows a Chi-squared distribution with two degrees of freedom, as supported by [27], [28], [23], given by

$$\frac{2|Z(k)|^2}{L\sigma^2} \sim \chi_2^2. \tag{6}$$

This distribution is defined as the summation of the squares of independent standard normal variables, and the factor $2/L\sigma^2$ arises from the normalization of the real and imaginary components of the coefficients $Z(k) \sim \mathcal{N}(0,1)$, following [23].

Under the hypothesis H_1, the received signal comprises a composite of the signal $x(n)$ and the noise $z(n)$. Since $x(n)$ and $z(n)$ are mutually independent, their respective Fourier transforms also remain independent. Consequently, under H_1, the power spectrum bin adheres to a non-central Chi-squared distribution, denoted as $P(k) \sim \chi_2^2(\lambda_k)$. This distribution possesses two degrees of freedom and a non-centrality parameter, as elaborated in [27], [28], [23].

$$\lambda_k = \frac{1}{\sigma^2}\left|X(k) + Z(k)\right|^2. \tag{7}$$

The central chi-squared distribution arises from the sum of squared independent standard normal distributions, while the noncentral chi-squared distribution extends this concept to include normal distributions with any mean and variance. This property enables the incorporation of deterministic narrow or wideband signals in noise, resulting in non-centrality. The presence of non-centrality is a crucial foundation for the detection mechanism, as outlined in the subsequent section.

IV. Proposed Solution

Integrating new features into the array of machine learning parameters may not pose a substantial challenge. However, ensuring that each processing step remains as straightforward as possible and effectively handling the surging data volumes demands thoughtful consideration.

Our solution tackles the former by examining the optimal conversion of CDFs into data that is manageable for learning purposes. Simultaneously, we address the latter by departing from the previous approaches, which involved aggregating data from multiple FFT measurement campaigns. Instead, we minimize the data requirement for empirical distribution analysis by constructing an FFT matrix based on decimated signals. This approach streamlines data management and enhances the efficiency of our analysis.

The detection is based on a single capture of the signal time frame and three assumptions:

1) Adjacent noise samples, as iid, can be arbitrarily combined without changing the statistical parameters.
2) The temporal form of signals can be subsampled without loss in average power.
3) The imperfect matching of the sampling frequency and the frequency of the analyzed signal causes energy leakage into adjacent FFT bins.

Up to this point, the singly captured sample vector $\mathbf{s}$ of length L is transformed according to the following procedure:

1) We determine the decimation coefficient d.
2) We create a measurement matrix $\mathbf{m}$ with dimensions of $d \times \lfloor L/d \rfloor$.
3) The rows of $\mathbf{m}$ are filled with samples of $\mathbf{s}$ in a non-overlapping, interleaved pattern:

$$m(i,j) = S\left((j-1)d + i\right). \tag{8}$$

4) Proceed the $\lfloor L/d \rfloor$-point FFT for each row of the matrix $\mathbf{m}$.
5) For row-oriented FFTs matrix, determine the column-oriented empirical CDFs.
6) Resolve manageable representation for each CDF curve:
   a) centroid of a polyfigure,
   b) amplitude of 0.5 probability,
   c) length and deviation,
   d) slope coefficient and length.
7) Evaluate the 2D points representing the CDFs with respect to the demarcation of the points clustered in the ML procedure.

The method expounded in this discourse holds several significant advantages. Primarily, it integrates a direct incorporation of CDF into the ML process. Unlike the previous solution [23] that evaluated goodness-of-fit against expected distribution patterns, our approach introduces ML without assuming specific distributions. Instead, it rests on the premise that the two states are statistically discernible, leveraging unsupervised clustering performed by ML. Moreover, temporal decomposition facilitates the handling of multiple notably shorter FFTs, thereby enabling the acquisition of a comprehensive CDF representation for both H_0 and H_1 states from a single measurement capture. This methodology reduces system inertia and diminishes the data required for model relearning in response to changing propagation conditions that necessitate algorithmic adjustments.

V. Simulation

To evaluate the performance of our proposed algorithm, we established a simulation framework designed to generate random instances of a deterministic signal within a complex AWGN environment. Our analysis is centered on non-overlapping time intervals, each comprising 4096 samples of complex noise. In this simulation, there is a 50% probability of introducing a complex Gaussian-modulated sinusoidal Radio Frequency (RF) pulse into the frame, as illustrated in Figure 1a. The transmitted signal, denoted as Tx, remains clear, while the received signal in the presence of noise is represented as Rx, with its real and imaginary parts representing the in-phase and quadrature-phase components, respectively. Key parameters include a receiver (Rx) bandwidth (BW) of 120 MHz, RF pulse fractional bandwidth following a normal distribution $\sim \mathcal{N}(0.5, 0.1)$, and center frequency $\sim \mathcal{N}(0.15, 0.1)$ of the bandwidth. Additionally, the phase position of the pulse within the frame is governed by a uniform distribution.

Considering SNR below 0 dB, as depicted in the full-scale FFT (Figure 1b), there is no clear indication of signal presence. However, our proposed solution employs a method where a frame is subdivided into 32 subframes. As a consequence of this approach, rather than computing a 4096-point FFT, we calculate 128 instances of a row-oriented 32-point FFT for a row-oriented time-decomposition matrix. Following this, a column-oriented cumulative distribution function (CDF) is generated, reflecting the empirical distributions for individual spectral bins (Figure 2). The non-centrality (shift) of the CDFs corresponding to spectral bands carrying the signal becomes

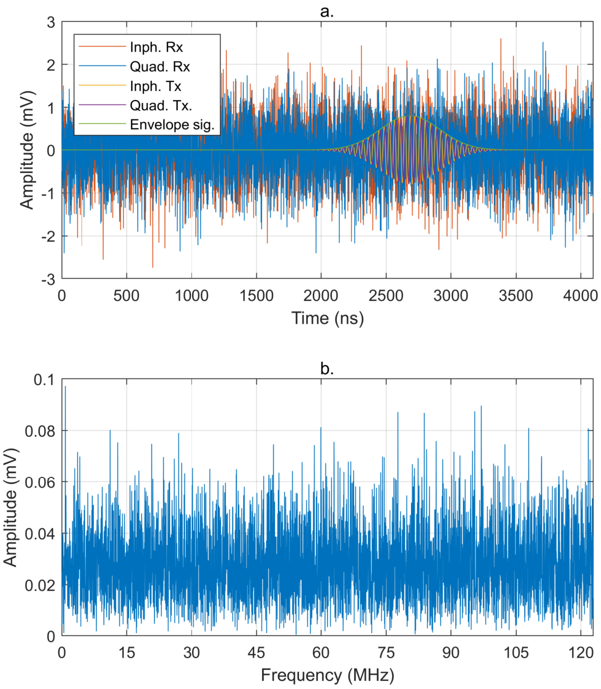

Fig. 1. Single capture of a time frame (a) at the transmitter (Tx) and the receiver (Rx); FFT of the received signal (b).

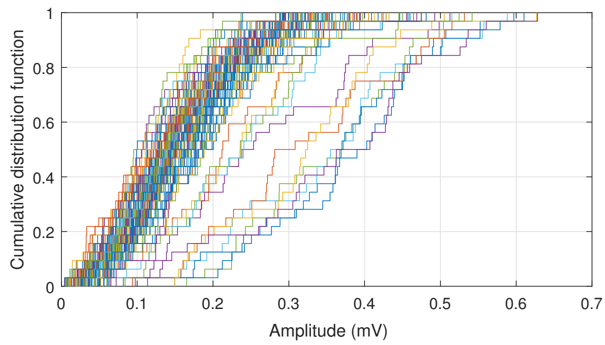

Fig. 2. Column-oriented CDFs of 128 row-oriented 32-point FFTs.

distinctly observable. Subsequently, converting each curve into a more manageable representation becomes imperative.

A. *Centroid of a Polyfigure*

The fundamental approach supported in [12] involves representing each curve by redefining it as a polyfigure and then finding the centroid (Fig. 3). Thus, each curve is represented by a single point on a 2D amplitude-probability plane. The resulting cloud of points, with an intensely concentrated group representing noise, is readily amenable to clustering using the k-means algorithm and a cosine-based distance criterion (Fig. 4). Both groups can be well separated by a single demarcation line, which adjustment through the tuning of a constant coefficient allows for regulating the probability of false alarms.

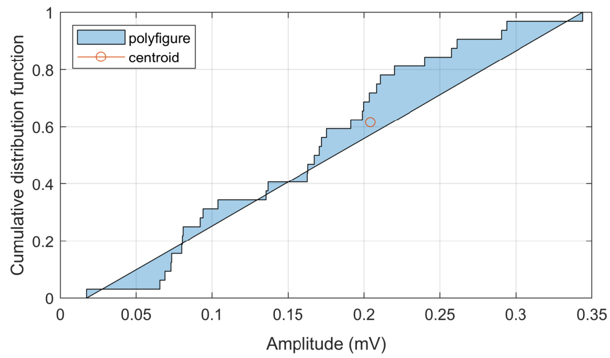

Fig. 3. Transformation of a single CDF into the centroid of a polyfigure.

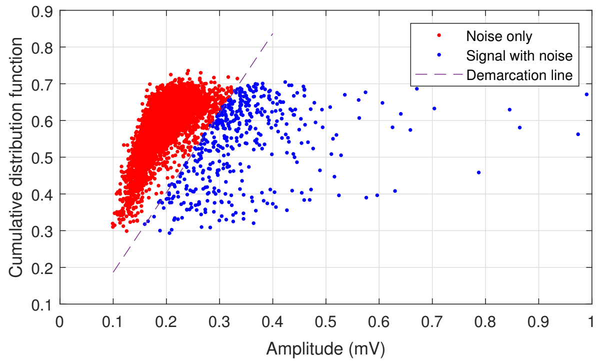

Fig. 4. Allocation of clusters using the K-means method.

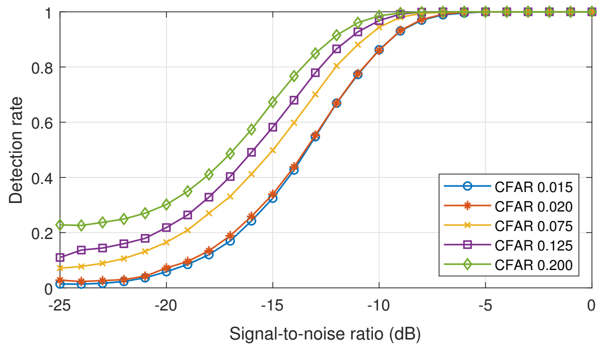

Fig. 5. Detection rate in a CFAR scenario utilizing centroid classification.

The results of the exemplary detection scenario show that the presented method can be effectively applied even for very weak signals. The analysis for the constant false alarm rate (CFAR) set to 0.075 shows that the detection rate (DR) keeps above 90% to over -11 dB. However, the technique has its challenges. We place a strong emphasis on the low complexity of the proposed solution. Conversely, polyshape determination and centroid calculation are complex operations that, while yielding good detection results, conflict with the adopted assumption. Therefore, in the following sections, we present simplified approaches to representing CDF, their impact on learning, and detection effectiveness.

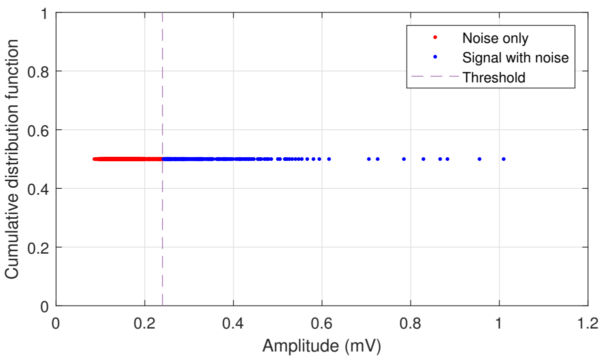

Fig. 6. Assignment of clusters using the K-means method.

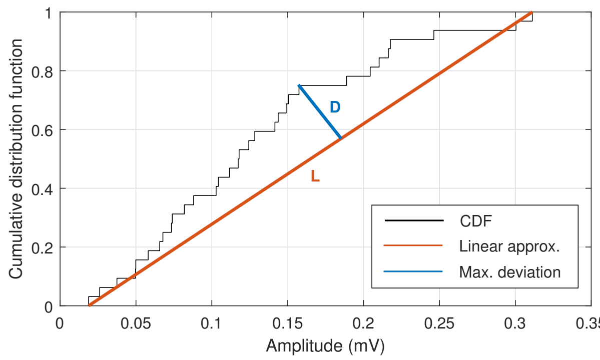

Fig. 8. Conversion of a single CDF into a length-difference (L/D) pair.

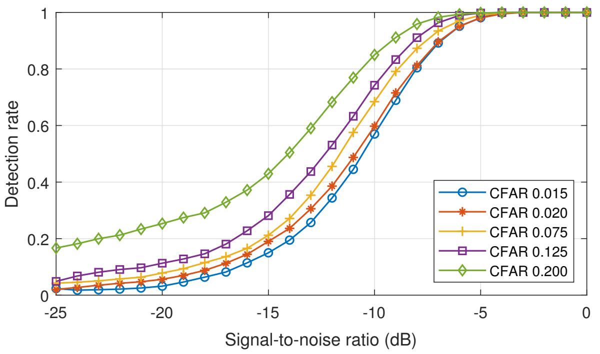

Fig. 7. Detection rate in a CFAR scenario using mean classification.

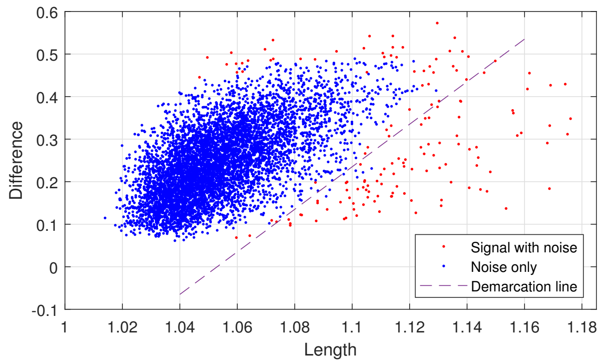

Fig. 9. Cluster assignment using the DBSCAN method.

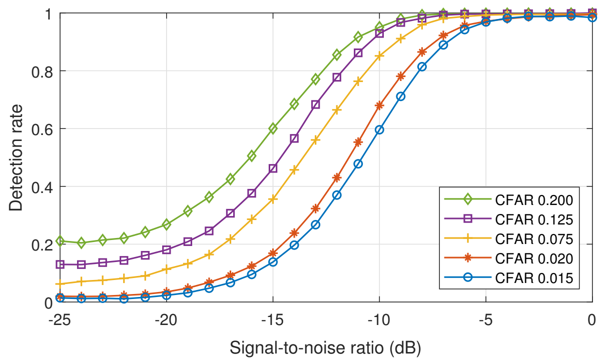

Fig. 10. Detection rate in a CFAR scenario with L/D classification.

B. Mean

One of the most straightforward and intuitive approaches to evaluating the centrality/non-centrality of a distribution is through the assessment of the mean value. This reduces the earlier 2D planeto to a one-dimensional amplitude criterion for a probability of 0.5. The value is easily distinguishable using k-means clustering with Euclidean distance criterion (Fig. 6). In this case, the false alarm probability is adjusted by adopting a threshold value for the observed amplitude. For such a tailored method, the detection efficiency at a 90% level in a scenario analogous to the previous one is -7 dB (Fig. 7).

C. Linear Approximation: Length and Deviation

A more intricate, yet still accessible, evaluation of CDFs focuses on measuring curvature rather than mean-shift. In this regard, the curve undergoes approximation through a linear function, resulting in the length L as the first acquired dimension. Subsequently, the evaluation involves assessing the maximum deviation of the empirical CDF from the approximation, yielding second dimension D (Fig.8). The cloud of points in the D/L plane serves as the basis for clustering in the ML process (Fig.9). The clustering utilizes the DBSCAN algorithm, created on a density-based spatial clustering (control parameters: a minimum of 10 points within a distance of 0.02). The simplified 2D approach produces more promising results than the 1D analysis of the mean, ensuring a 90% detection accuracy at approximately -9 dB (Fig. 10).

D. Linear Approximation: Slope Coefficient and Length

The third considerable factor that captures the distinctiveness of CDFs between 'noise-only' and 'signal-with-noise' states is the slope coefficient of the curve in its linear approximation. In this context, we assess the length of the curve, denoted as L, approximated based on the points where the curve reaches probabilities of 0.1 and 0.9. For the linear approximation, we determine the slope coefficient α (Fig.11). These measurements enable the construction of a two-dimensional space, denoted as α/L, where, once again, a cloud of points undergoes clustering into two groups using the DBSCAN strategy (control parameters: a minimum of 10 points within a distance of 0.01) (Fig.12). It's important to

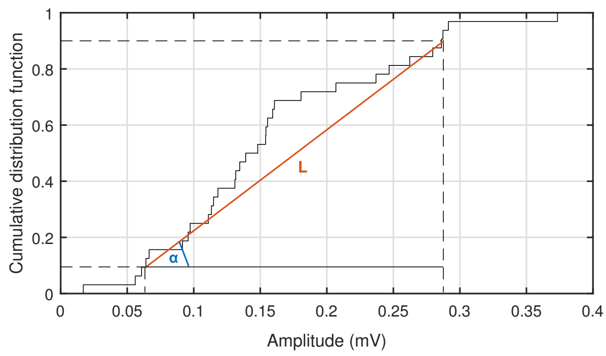

Fig. 11. Conversion of a single CDF into a slope-length (α/L) pair.

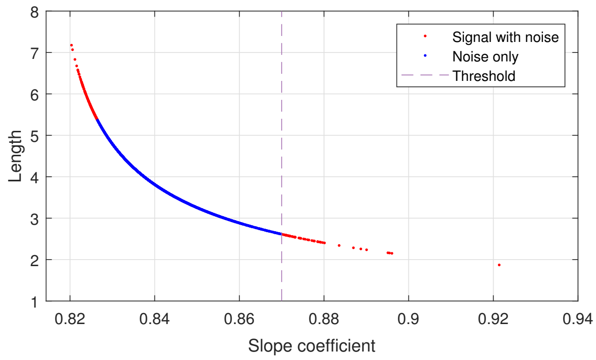

Fig. 12. Assignment of clusters using the DBSCAN method.

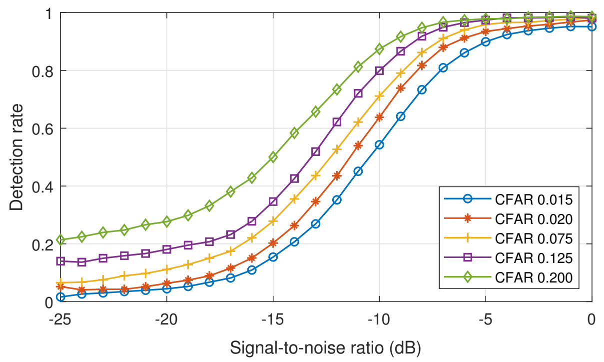

Fig. 13. Detection rate in a CFAR scenario with α/L classification.

highlight that the pivotal factor in partitioning the plane is not the length but rather the determined slope coefficient. This 2D approach yields a detection accuracy of approximately -7 dB for a 90% detection rate when considering a 7.5% false alarm rate (Fig. 13).

E. Complexity

The investigation into time complexity was conducted by averaging the duration of 25,000 detection processes executed serially on a 3 GHz CPU. According to previous considerations, relinquishing the determination of the polyshape in favor of other methods results in a significant decrease in the detection time by order of magnitude (Fig. 14). In the case of

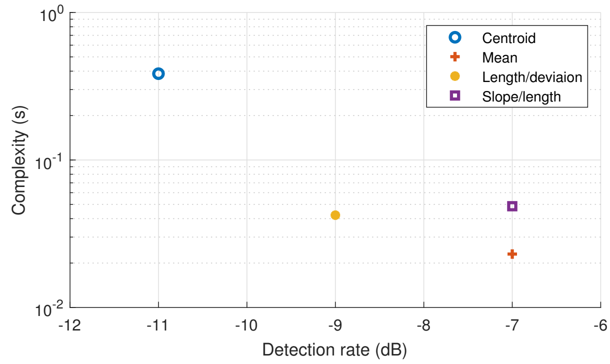

Fig. 14. Time complexity comparison.

the centroid-based method, the total process of single detection took 385 ms, with the shape determination itself accounting for 304 ms. In comparison, for methods based on D/L and α/L, taking 42 ms and 48.5 ms, respectively, the two most complex operations were linear approximation at 16 ms and building CDF matrices at 14.2 ms (common to all methods). The most straightforward approach, i.e., assessing the mean value, took 23 ms.

F. Comparison to other Methods

Referring to the study [29] is valuable to relate the results obtained to other alternative methods. In a similar impulse transmission scenario, the authors present the effectiveness of detection based on statistical approaches and relate them to the fundamental techniques of energy detection and its modification, utilizing short windows. Table I provides a comparison of methods in two scenarios: S_1) the minimum SNR at which the method achieves 90% detection rate with a CFAR of 10% and 1024 samples per frame (for CDF-based methods decimation coefficient d set to 16); S_2) the detection rate achieved for an SNR cutoff at -6 dB with a CFAR of 10% and 4096 samples per frame.

TABLE I
PERFORMANCE COMPARISON OF DETECTION METHODS IN TWO SIMULATION SCENARIOS

Detection method	SNR in S_1 (DR over 90%)	DR in S_2 (SNR at -6dB)
Conventional energy det.	0 dB	78%
Short windows energy det.	-1 dB	80%
Jarque-Bera normality test	3 dB	81%
Higher-order-statistics test	3 dB	83%
Gaussianity testing	-2 dB	88%
CDF mean-based det.	-2 dB	92%
CDF α/L-based det.	-2 dB	95%
CDF L/D-based det.	-3 dB	99%
CDF centroid-based det.	-5 dB	99%

Comparison with previously studied approaches shows that the proposed solution, even with a short frame of 1024 samples, ensures better detectability than energy detection or direct sample distribution analysis. Reference to results obtained for 4096 samples in the frame indicates that statistical methods significantly benefit from increased frame length used to build the empirical distribution (for details, see [29]).

Nonetheless, the approach based not on predefined patterns but on a learned reference obtained through autonomous clustering occurs superior in both cases, which seems intuitively justified due to the better adaptation to the case achieved for learning methods.

Regarding time complexity, we can draw insights from [30]. The results outlined in subsection V-E surpass the 1 ms threshold observed in energy detection but bring the basic mean-based approach close to 19.2 ms, as seen in normality testing (adjusted for equal input size). It is noteworthy that, from the detectors' standpoint, the entire learning and establishment of the appropriate demarcation function, crucial for detection, occur independently of standalone detection. This process occurs in the background or between detection campaigns.

VI. Conclusions

The presented research builds upon prior investigations into the effective integration of time decomposition and distribution analysis with machine learning. The conducted survey of methods showcases the diversity of solutions achievable through these tools and underscores their practical implementation in an effective and straightforward manner.

The undertaken studies reveal that, despite a relatively modest increment in the complexity of a singular detection process, the utility of detection increases significantly. The adaptability of these methods, enabled by effective learning through autonomous clustering, positions them as efficient solutions for blind signal detection applications, thereby outperforming classical statistical approaches.

Moreover, it is worth noting that even a simple, two-dimensional assessment of CDF curvature proves to enable effective detection. However, exploring additional refinement and optimization of CDF-based detection methods could enhance their efficiency and adaptability in more diverse signal detection scenarios. This avenue of research would contribute to ensuring the robustness and versatility of the proposed methods, making them more applicable across various real-world situations.

Acknowledgment

This work was supported by the Polish Ministry of Science and Higher Education under the status activity task 0314/SBAD/0231 in 2023.

References

[1] 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, p. 126, 2019.

[2] S. Rezaei, J. P. McKean, M. Biehl, and A. Javadpour, "DECORAS: detection and characterization of radio-astronomical sources using deep learning," *Monthly Notices of the Royal Astronomical Society*, vol. 510, no. 4, pp. 5891–5907, 2021.

[3] Z. Geng, H. Yan, J. Zhang, and D. Zhu, "Deep-learning for radar: A survey," *IEEE Access*, vol. 9, pp. 141 800–141 818, 2021.

[4] H. Ma, X. Yuan, J. Wang, and B. Li, "Revisiting model order selection: A sub-nyquist sampling blind spectrum sensing scheme," *IEEE Transactions on Wireless Communications*, pp. 1–13, 2022.

[5] G. George and C. Prema, "Cyclostationary feature detection based blind approach for spectrum sensing and classification," *Radioengineering*, vol. 27, pp. 298–303, 04 2019.

[6] S. Benazzouza, M. Ridouani, F. Salahdine, and A. Hayar, "A survey on compressive spectrum sensing for cognitive radio networks," in *2019 IEEE International Smart Cities Conference (ISC2)*, 2019, pp. 535–541.

[7] R. Ahmed, Y. Chen, B. Hassan, and L. Du, "Cr-iotnet: Machine learning based joint spectrum sensing and allocation for cognitive radio enabled iot cellular networks," *Ad Hoc Networks*, vol. 112, p. 102390, 2021.

[8] A. Mohammad, F. Awin, and E. Abdel-Raheem, "Case study of tv spectrum sensing model based on machine learning techniques," *Ain Shams Engineering Journal*, vol. 13, no. 2, p. 101540, 2022.

[9] K. Zhang, J. Li, and F. Gao, "Machine learning techniques for spectrum sensing when primary user has multiple transmit powers," in *2014 IEEE International Conference on Communication Systems, ICCS*. Macau, China: IEEE, 2014, pp. 137–141.

[10] O. P. Awe, "Machine learning algorithms for cognitive radio wireless networks," Ph.D. dissertation, Loughborough University, Loughborough, U.K., November 2015.

[11] H. Xiao, X. Zhou, and Y. Tian, "Research on wireless spectrum sensing technology based on machine learning," in *International Conference on Security, Privacy and Anonymity in Computation, Communication and Storage*. Cham, Melbourne, Australia: Springer, 2018, pp. 472–479.

[12] J. Nikonowicz and M. Jessa, "Wideband spectrum sensing utilizing cumulative distribution function and machine learning," in *2023 International Conference on Software, Telecommunications and Computer Networks (SoftCOM)*. Split, Croatia: IEEE, 2023, pp. 1–6.

[13] M. R. Vyas, D. K. Patel, and M. Lopez-Benitez, "Artificial neural network based hybrid spectrum sensing scheme for cognitive radio," in *Proc. IEEE Int. Symp. Personal, Indoor, Mobile Radio Commun.* Montreal, QC, Canada: IEEE, Oct. 2017, pp. 1–7.

[14] J. Tian, P. Cheng, Z. Chen, M. Li, H. Hu, Y. Li, and B. Vucetic, "A machine learning-enabled spectrum sensing method for OFDM systems," *IEEE Trans. Veh. Technol.*, vol. 68, no. 11, pp. 11 374–11 378, 2019.

[15] B. Khalfi, A. Zaid, and B. Hamdaoui, "When machine learning meets compressive sampling for wideband spectrum sensing," in *2017 13th International Wireless Communications and Mobile Computing Conference (IWCMC)*. Valencia, Spain: IEEE, 2017, pp. 1120–1125.

[16] A. M. Mikaeil, B. Guo, and Z. Wang, "Machine learning to data fusion approach for cooperative spectrum sensing," in *2014 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC)*. Shanghai, China: IEEE, 2014, pp. 429–434.

[17] Z. Li, W. Wu, X. Liu, and P. Qi, "Improved cooperative spectrum sensing model based on machine learning for cognitive radio networks," *IET Communications*, vol. 12, no. 19, pp. 2485–2492, 2018.

[18] J. Nikonowicz, P. Kubczak, and L. Matuszewski, "Hybrid detection based on energy and entropy analysis as a novel approach for spectrum sensing," in *2016 International Conference on Signals and Electronic Systems (ICSES)*. Krakow, Poland: IEEE, 2016.

[19] ——, "Impact of the decision function on the overall performance of an energy-and entropy-based hybrid detector," in *2018 International Conference on Signals and Electronic Systems (ICSES)*. Krakow, Poland: IEEE, 2018, pp. 114–117.

[20] Y. Arjoune and N. Kaabouch, "On spectrum sensing, a machine learning method for cognitive radio systems," in *2019 IEEE International Conference on Electro Information Technology (EIT)*. Brookings, SD, USA: IEEE, 2019, pp. 333–338.

[21] G. T. Reddy, M. P. K. Reddy, K. Lakshmanna, R. Kaluri, D. S. Rajput, G. Srivastava, and T. Baker, "Analysis of dimensionality reduction techniques on big data," *IEEE Access*, vol. 8, pp. 54 776–54 788, 2020.

[22] J. Nikonowicz and M. Jessa, "A novel method of blind signal detection using the distribution of the bin values of the power spectrum density and the moving average," *Digital Signal Processing*, vol. 66, pp. 18–28, 2017.

[23] B. Scheers, D. Teguig, and V. Le Nir, "Wideband spectrum sensing technique based on goodness-of-fit testing," in *2015 International Conference on Military Communications and Information Systems (ICMCIS)*, 2015, pp. 1–6.

[24] A. Goldsmith, *Wireless Communications*, 1st ed. Cambridge University Press, 2005.

[25] R. W. Heath Jr and A. Goldsmith, "Why gaussian? an introduction to the mathematics of randomness in wireless communications," *IEEE Signal Processing Magazine*, vol. 16, no. 4, pp. 65–70, 1999.

[26] B. Sklar, *Digital Communications: Fundamentals and Applications*, 2nd ed. Prentice Hall, 2001, ch. 3, pp. 53–87.

[27] W. A. Gardner, *Statistical spectral analysis: A nonprobabilistic theory*. Prentice Hall, 1995.

[28] S. M. Kay, *Fundamentals of Statistical Signal Processing, Volume II: Detection Theory*. Prentice Hall, 1998.

[29] J. Nikonowicz and M. Jessa, “Gaussianity testing as an effective technique for detecting discontinuous transmission in 5g networks,” *IEEE Access*, vol. 9, pp. 22 186–22 194, 2021.

[30] ——, “A novel method of blind signal detection using the distribution of the bin values of the power spectrum density and the moving average,” *Digital Signal Processing*, vol. 66, pp. 18–28, 2017.

Jakub Nikonowicz received an M.Sc. degree in Electronics and Telecommunication and a Ph.D. degree (with honors) in Telecommunication from Poznan University of Technology, Poznan, Poland, in 2014 and 2019, respectively. He has authored or co-authored over a dozen scientific publications in refereed journals and international conference proceedings. He also served as a project manager for grants supporting young scientists and as a co-principal investigator in an international research consortium, contributing significantly to collaborative research efforts. His current research interests encompass innovations in statistical signal processing for next-generation communication and security technologies.

Mieczyslaw Jessa received the M.Sc. degree with honors from Poznan University of Technology in 1985 and the Ph.D. degree in 1992 from the same University. Initially, his research interest included phase-locked loops and PDH/SDH network synchronization. In the years 1995-1997 he was an expert of the Polish Ministry of Communications in the field of digital network synchronization. His current research concerns the mathematical models of randomness and pseudo-randomness. He is the author or co-author of over one hundred journal and conference papers and fifteen patents.

Łukasz Matuszewski graduated from the Faculty of Electronics and Telecommunications of the Poznań University of Technology in 2010. In 2011, he was employed at the Department of Telecommunications Systems and Optoelectronics at the Faculty of Electronics and Telecommunications of the Poznań University of Technology. In 2019 he defended his Ph.D. thesis and since 2020 he has been an assistant professor at the Institute of Multimedia Telecommunications at the Faculty of Computing and Telecommunications at the Poznań University of Technology. He was a project manager for grants for young scientists, as well as a contractor in projects for a Polish telecommunications operator. He is the author and co-author of 25 scientific publications in peer-reviewed journals and materials from national and international conferences. His research interests include designing devices with the use of reprogrammable circuits, in particular cryptographic circuits, random number generators, and synchronization circuits.