Channel Energy Statistics Learning in Compressive Spectrum Sensing

Haoran Qi, *Student Member, IEEE*, Xingjian Zhang, *Student Member, IEEE*, and Yue Gao, *Senior Member, IEEE*

***Abstract*— Spectrum sensing is a proactive way in cognitive radio systems to achieve dynamic spectrum access; and compressive spectrum sensing (CSS) techniques alleviate the demand for high-speed sampling in wideband spectrum sensing. Most existing literature discusses Neyman–Pearson channel energy detection and threshold adaption schemes to achieve an optimal performance of detection in a conventional non-compressive spectrum sensing scenario. However, in the CSS, it is found that the channel energy statistics and optimal threshold depend not only on noise energy but also on compression ratio, sparsity of spectrum, and nature of recovery algorithms. To investigate the channel energy statistics of recovered spectrum, we postulate a statistical model of channel energy for CSS and propose a learning algorithm based on a mixture model and expectation–maximization techniques. In addition, having verified the validity of the postulated model, we propose a practical threshold adaption scheme for CSS aiming to maintain constant false alarm rates in channel energy detection. In simulations, it is shown that the postulated channel energy statistic models with parameters learned by the proposed learning algorithm fit well with empirical distributions under circumstances of various channel models and recovery algorithms. Moreover, it is presented that the proposed threshold adaption scheme maintains the false alarm rate near the predefined constant, which in turn validates the postulated model.**

***Index Terms*— Compressive spectrum sensing, energy detection, mixture model, threshold adaption, constant false alarm rate.**

I. Introduction

With the rapidly increasing demands for data rates and service coverage, spectrum scarcity is one of the significant challenges faced by today's wireless communications. The fact that the spectrum resource is underutilized in certain bands [1] has motivated the dynamic spectrum access (DSA) which enables unlicensed secondary users (SUs) to access the spectrum without causing significant interference to primary users (PUs). Spectrum sensing techniques in cognitive radio are proactive ways of sensing nodes to acquire the surrounding spectrum availability information. The traditional way of spectrum sensing requires analog-to-digital converters (ADCs) sampling at Nyquist rate. However, for spectrum sensing in wideband scenarios, Nyquist-rate processing tends to be unrealistic due to high power consumption and hardware complexity of ADCs. Compressive sensing (CS) techniques have been proposed to be applied to spectrum sensing to achieve reliable spectrum reconstruction on condition that the spectrum occupancy is sparse [2]–[7].

A real-world example of DSA application is in TV white-space (TVWS) where opportunistic access to the UHF band for digital terrestrial TV (DTTV) is allowed to made by low-power local radio applications, for example, machine-to-machine communications and wireless personal area networks. In UK's practice, it is stated that only up to 6 out of 32 channels over the 320MHz band are used by PUs at arbitrary location [1]. This low spectrum occupancy over DTTV bands not only enables potentially large throughput of SUs but also suits the application of wideband compressive spectrum sensing (CSS).

Detection of occupied channels using spectrum reconstructed by CS recovery algorithms is a crucial procedure of CSS. Cyclostationary feature detection (CFD) and channel energy detection are two major types of detection techniques applied in CSS appeared in literature. CFD exploits the cyclic stationary property in modulated radio signals and it conducts detection in the spectrum of cyclic frequency and spectral frequency [8], [9]. Compared with energy detection, CFD has been shown outperforming in lower signal-to-noise ratio (SNR) region [10], [11]. However, it adds considerable computational complexity to receivers and CFD-based detection algorithms only work with specific and known types of modulation of signals. The simpler yet more commonly used method is Neyman-Pearson (NP) channel energy detection of which the objective is to directly differentiate the present signal from noise in certain channel's power spectrum by setting a proper threshold [12]. A common problem of energy detectors is that statistics of the noise, for example the variance, are a priori unknown in most cases, because the noise in the received signal depends on receiver's noise figure and gain control, temperature, ambient radio interference, etc. Thus, the noise statistics need to be estimated to achieve optimal detection performance. As in most literature of conventional spectrum sensing, it is assumed noise level is Gaussian distributed in NP energy detection method and detec-

Manuscript received December 27, 2017; revised July 31, 2018; accepted September 18, 2018. Date of publication October 16, 2018; date of current version December 10, 2018. This work was supported by the Engineering and Physical Sciences Research Council, U.K., under Grant EP/R00711X/1. This paper was presented at the IEEE/CIC International Conference on Communications in China, Beijing, China, August 2018. The associate editor coordinating the review of this paper and approving it for publication was X. Zhou. *(Corresponding author: Haoran Qi.)*

The authors are with the Department of Electronic Engineering and Computer Science, Queen Mary University of London, London E1 4NS, U.K. (e-mail: h.qi@qmul.ac.uk; xingjian.zhang@qmul.ac.uk; yue.gao@ieee.org).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TWC.2018.2872712

tion performance can be evaluated in various channel models [13], [14]. Under the assumption of Gaussian noise, it has been proposed in [13]–[16] that noise variance estimation and threshold adaption can be achieved in an online fashion. Other discussions on threshold adaption schemes without prior assumptions on signal statistics are seen in literature where supervision in the adaption process is necessary [8], [17]. However, the supervised learning process requires training radio pilots of which the radio activity information is readily known to sensing nodes, which is unrealistic in real-world spectrum sensing applications.

In CSS, the desired spectrum for energy detection is not directly available from time-domain samples of analog-to-information converter (AIC) [18], [19]. The spectrum we interest in needs to be recovered from these sub-Nyquist-rate samples by sparse recovery algorithms. From both restricted isotropic property theory of CS [20] and practice of wideband CSS [3]–[5], [21], [22], the sparse recovery of a noisy signal results in an inconsistent recovered spectrum compared to the true spectrum, and such inconsistency depends on spectrum sparsity, compression ratio and SNR of the sensed signal. In our past experiments [3], [23], we discovered that in CSS the energy statistics of recovered spectrum differ from that of the true spectrum. In order to achieve optimal energy detection performance, a proper statistical model and unknown statistics in recovered spectrum need to be learned.

The main contribution of this paper is that, to the best knowledge of the authors, it is the first work to address and model the statistics of the recovered signal in the energy detection problem of CSS. We discovered that the channel energy statistics in CSS is fundamentally different from that in conventional non-compressive spectrum sensing. Specifically, to set NP energy detection hypotheses for recovered signals, we postulate that the channel energy statistics model of recovered spectrum still conforms to that of the original spectrum, however, parameters of the model for recovered signals are treated as unknown. Mixture Model (MM) and Expectation-Maximization (EM) techniques [24] have been commonly used to obtain maximal-likelihood estimates the parameters given analytic distributions of statistics, and the specific use of Rayleigh-Gaussian MM has been seen in [25] and [26] to learn the signal statistics in non-compressive spectrum sensing. In this paper, we focus on the channel energy in CSS and a customized EM-based algorithm for a chi-square-MM is proposed to learn the channel energy statistics of recovered signal. Simulations have shown that the postulated statistical model for recovered spectrum is a reasonably good fit with parameters learned from sample data set by the proposed algorithm.

An additional contribution is that, furthermore, we propose a novel and practical threshold adaption scheme based on the newly-addressed statistic model to achieve the detection performance of constant false alarm rate (CFAR) for energy detection in CSS. In simulations, it is showed how differently the thresholds should be set in various settings of the CSS. Moreover, the results that the probability of false alarm can be kept near the predefined constant also validate the good fitness of our postulated model of channel energy statistics and learning algorithm.

The rest of this paper is organized as follows. In Section II, the signal model and the NP energy detection problem in CSS are illustrated. In Section III, the postulated statistical model for recovered signal in CSS is presented and an algorithm for learning the parameters in the postulated statistical model is proposed. In Section IV, based on the results in Section III, the threshold adaption scheme based on noise statistics estimation is proposed. Simulation results to prove the effectiveness of the postulated channel energy statistical model and the proposed learning algorithm, as well as the performance of the proposed threshold adaption scheme are shown in Section V.

II. Signal Model and Problem Statement

A. Compressive Spectrum Sensing

Consider CSS by single sensing node in non-cooperative scenario. Denote the Nyquist time-domain signal as $\boldsymbol{s_t} = [s_t^{(1)}\ s_t^{(2)} \dots s_t^{(N)}]^T$ and its frequency-domain representation $\boldsymbol{s_f} = \boldsymbol{\mathcal{F}}\boldsymbol{s_t} = [s_f^{(1)}\ s_f^{(2)} \dots s_f^{(N)}]^T$ where $\boldsymbol{\mathcal{F}}_{N\times N}$ stands for N-point Discrete Fourier Transform (DFT) matrix. A signal sparse in frequency domain can be recovered based on M sub-Nyquist-rate measurements $\boldsymbol{y} = [y^{(1)}\ y^{(2)} \dots y^{(M)}]^T$ where $M < N$. The sub-Nyquist sampling can be expressed as a linear system

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{s_t} + \boldsymbol{b} = \boldsymbol{A}\boldsymbol{\mathcal{F}}^{-1}\boldsymbol{s_f} + \boldsymbol{b}, \tag{1}$$

where $\boldsymbol{A}_{M\times N}$ is the sampling matrix with structured random entries corresponding to AIC sampler settings, and $\boldsymbol{b} = [b^{(1)}\ b^{(2)} \dots b^{(M)}]$ is additive noise on measurements.

The recovery of sparse signal $\boldsymbol{s_f}$ can be achieved by solving the optimization problem, writing

$$\boldsymbol{s_f}^o = \arg\min_{\boldsymbol{s_f}} ||\boldsymbol{s_f}||_l + \frac{\mu}{2}||A\mathcal{F}^{-1}\boldsymbol{s_f} - \boldsymbol{y}||_2^2, \tag{2}$$

where the norm $0 < l \leq 1$ and μ accounts for penalization. Specifically when $l = 1$, the optimization problem is convex and can be solved by Basis Pursuit Denoising algorithm. Furthermore, when $\mu = \infty$, which is usually adopted as penalization term is often a priori unknown, the recovery algorithm is called Basis Pursuit (BP). Besides, greedy algorithms are also an efficient category of sparse recovery solvers, among which the most commonly used is the Orthogonal Matching Pursuit (OMP) [27]. Spare Bayesian Learning (SBL) has been recently brought to attention for solving sparse recovery problem in a probabilistic setting. By introducing Gaussian assumption on noise, specifically $p(\boldsymbol{y}|\boldsymbol{s_f})$ being Gaussian and assuming parameterized Gaussian priors $\boldsymbol{s_f} \sim \mathcal{CN}(\boldsymbol{0}, \mathrm{diag}(\boldsymbol{\gamma}))$ that induces sparsity in the recovered signal, it aims to find maximum a posteriori probability of the hyperparameters $\boldsymbol{\gamma}$ [28], [29]:

$$\boldsymbol{s_f}^o = \arg\max_{\boldsymbol{\gamma}} p(\boldsymbol{\gamma}|\boldsymbol{y}) = \arg\max_{\boldsymbol{\gamma}} \int p(\boldsymbol{y}|\boldsymbol{s_f})p(\boldsymbol{s_f};\boldsymbol{\gamma})\mathrm{d}\boldsymbol{s_f}. \tag{3}$$

B. Channel Energy Detection in Spectrum Sensing

Consider wideband spectrum sensing in a frequency-division band where a channel is the unit of spectrum resource. To detect the channel occupancy, we calculate the average of

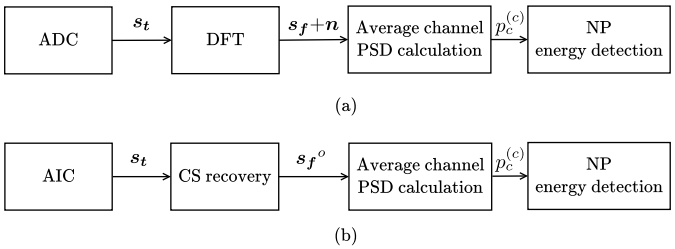


Fig. 1. System architecture of **(a)** conventional non-compressive compressive sensing and **(b)** CSS.

power spectrum density (PSD) bins within each channel for NP energy detection as

$$p_c^{(c)} = \sum_{(c-1)R \le i < cR} \frac{1}{R} |x_f^{(i+1)}|^2, \tag{4}$$

for $c = 1, 2, \ldots, C$, where C is the total number of channels and R is the number of PSD bins in each channel. The system architecture diagrams of non-compressive spectrum sensing and CSS with channel energy detection are illustrated in Fig. 1 (a) and (b) respectively.

Next, we will formulate the statistical model of channel energy for conventional non-compressive spectrum sensing, and elaborate the differences in the channel energy statistics for compressive spectrum sensing cases.

For non-compressive spectrum sensing, express the sensed spectrum as $\boldsymbol{x_f} = \boldsymbol{s_f} + \boldsymbol{n_f} = \boldsymbol{\mathcal{F}} \boldsymbol{x_t} = \boldsymbol{\mathcal{F}}(\boldsymbol{s_t} + \boldsymbol{n_t})$, where $\boldsymbol{n_t}_{N\times 1} = [\omega_t^{(1)} \ \omega_t^{(2)} \cdots \omega_t^{(N)}]^T$ denotes complex independent-and-identically-distributed (i.i.d.) additive white Gaussian noise (AWGN) as is widely adopted in literature [3], [12], [14]. Denote $\boldsymbol{n_t} \sim \mathcal{CN}(\boldsymbol{0}, \sigma^2 \boldsymbol{I})$. Note that the dimension of $\boldsymbol{x_f}$ here is $N = C \cdot R$. Consider a general multipath scenario, the expression of $x_t^{(n)}$ of null hypotheses $\mathcal{H}_0$ corresponding to absent radio activity and $\mathcal{H}_1$ corresponding to active radio activity, is modelled by

$$x_t^{(n)} = \begin{cases} w_t^{(n)} & \mathcal{H}_0 \\ \sqrt{K} e_t^{(n)} + \sqrt{1-K} e_t^{(n)} h_t^{(n)} + w_t^{(n)} & \mathcal{H}_1 \end{cases}$$

for $n = 1, 2, \cdots, N$, where we consider a commonly-adopted model $\boldsymbol{h_t} = [h_t^{(1)} \ h_t^{(2)} \cdots h_t^{(N)}]^T \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{I})$ that characterizes the multipath effect of channel. K is the power ratio of line-of-sight against multipath component. $\boldsymbol{e_t} = [e_t^{(1)} \ e_t^{(2)} \cdots e_t^{(N)}]^T$ represents the deterministic samples in a time frame of the PUs' transmitted signal attenuated by channel gain, and we define $\boldsymbol{e_f} = [e_f^{(1)} \ e_f^{(2)} \cdots e_f^{(N)}] := \boldsymbol{\mathcal{F}} \boldsymbol{e_t}$. After performing a linear transform (i.e. the DFT) on $\boldsymbol{s_t}$ and some rearrangements, we reach the statistical model of $\boldsymbol{x_f}$

$$x_f^{(n)} = \begin{cases} w_f^{(n)} \sim \mathcal{CN}(0, N\sigma^2) & \mathcal{H}_0 \\ \sqrt{K} e_f^{(n)} + \sqrt{1-K} e_f^{(n)} h_f^{(n)} + w_f^{(n)} \\ \sim \mathcal{CN}(\sqrt{K} e_f^{(n)}, (1-K)|e_f^{(n)}|^2 + N\sigma^2) & \mathcal{H}_1 \end{cases} \tag{5}$$

where we can find $\boldsymbol{h_f} = [h_f^{(1)} \ h_f^{(2)} \cdots h_f^{(N)}]^T \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{I})$ and denote $\boldsymbol{n_f} = [\omega_f^{(1)} \ \omega_f^{(2)} \cdots \omega_f^{(N)}]^T := \boldsymbol{\mathcal{F}} \boldsymbol{n_t} \sim \mathcal{CN}(\boldsymbol{0}, N\sigma^2 \boldsymbol{I})$.

Moreover, from (4), the statistical model of average channel PSD level $p_c^{(c)}$ is found characterized by a central and a non-central chi-square distribution, as a direct result of summing the squared Gaussian distributions of $\boldsymbol{x_f}$ in (5) with zero means and non-zero means respectively. Specifically, we write

$$p_c^{(c)} = \begin{cases} \displaystyle\sum_{(c-1)R \le i < cR} \frac{|w_f^{(i+1)}|^2}{R} = \frac{N\sigma^2}{2R} r_0, \ r_0 \sim \chi^2(2R) & \mathcal{H}_0 \\ \displaystyle\sum_{(c-1)R \le i < cR} \frac{1}{R} |\sqrt{K} e_f^{(i+1)} \\ +\sqrt{1-K} e_f^{(i+1)} h^{(i+1)} + w_f^{(i+1)}|^2 \\ = \beta^{(c)} r_1, \ r_1 \sim \chi'^2(2R, \alpha^{(c)}) & \mathcal{H}_1 \end{cases} \tag{6}$$

for $c = 1, 2, \cdots, C$, where $\chi^2(k)$ and $\chi'^2(k, \lambda)$ represents central chi-square distribution of degree of freedom (DoF) k and non-central chi-square distribution of DoF k and non-centrality parameter λ respectively. In (6), $r_0 \sim \chi^2(2R)$ and $r_1 \sim \chi'^2(2R, \alpha^{(c)})$ are random variables. Parameters $\alpha^{(c)}$ and $\beta^{(c)}$ are expressed as

$$\alpha^{(c)} = \sum_{(c-1)R \le i < cR} \frac{K|e_f^{(i+1)}|^2}{(1-K)|e_f^{(i+1)}|^2 + N\sigma^2}, \tag{7}$$

and

$$\beta^{(c)} = \frac{(1-K) \displaystyle\sum_{(c-1)R \le i < cR} |e_f^{(i+1)}|^2 / R + N\sigma^2}{2R}. \tag{8}$$

As special cases, expressions for Rayleigh channel and AWGN channel can be obtained by setting $K = 0$ and $K = 1$, respectively. In Rayleigh channel case, noting that $\alpha^{(c)} = 0$, the general noncentral chi-square distribution of $\mathcal{H}_1$ degenerates to a central chi-square distribution.

In conventional non-compressive spectrum sensing, it is generally valid to assume Gaussianity of noise added on the signal as a prior information. However, in CSS, when we assume Gaussian noise and certain channel model of original signal, it is spotted that the two hypotheses' distributions in recovered signal do not preserve distributions in original signal. Specifically, let the sensed spectrum to be $\boldsymbol{x_f} = \boldsymbol{s_f}{}^o$ in CSS scenario, and it is found that the distributions in (6) parameterized by R, (7) and (8) no longer hold. This finding should not be surprising, as it is a common conclusion that the recovery performance goes worse with decreasing compression ratio and spectrum sparsity, as a result of more occurrences of miss detected falsely detected spectrum supports. A direct result of this discrepancy is that attempts in CSS to achieve CFAR in detection using threshold setting methods [12] for non-compressive spectrum sensing will lead to varying false alarm probability, which have been appeared in [3, Figs. 7 and 8], [23, Figs. 6 and 8], and [30, Fig. 4]. To give a direct example of such changes on signal statistics, comparisons of original and recovered on spectrum, average channel PSD, and histogram of average channel PSD are given in Fig. 2. To simplify our problem, from this point, we assume the hypothesis $\mathcal{H}_1$ to have same parameters across all channels,

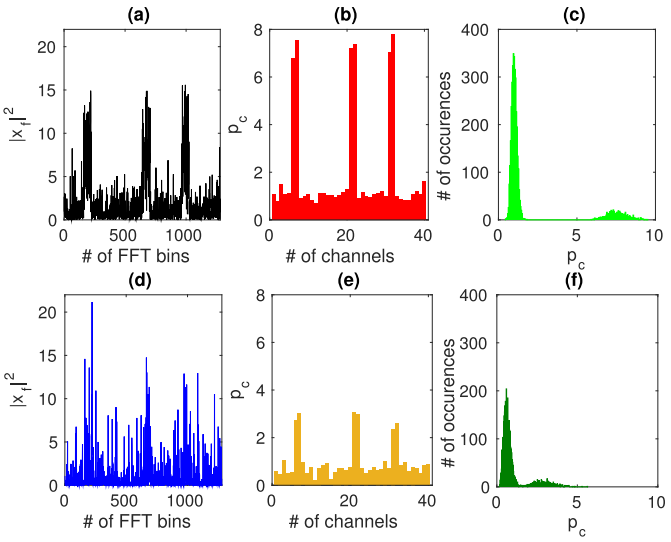


Fig. 2. Comparisons of original and recovered signal and their statistics. **(a)(d)** PSDs of original and recovered signal, respectively; **(b)(e)** average channel PSDs of original and recovered signal, respectively; **(c)(f)** average channel PSD distributions of original and recovered signal, respectively.

i.e. $\alpha = \alpha^{(1)} = \cdots = \alpha^{(C)}$ and $\beta = \beta^{(1)} = \cdots = \beta^{(C)}$. Hence, we do not differentiate the average channel PSDs $p_c^{(c)}$ from different channels. Instead, they are treated as multiple samples of p_c as they are drawn from the same statistic model. The scales of axis in each subfigure pairs for comparison are fixed to give a clear view of the differences. It is particularly noted that the difference in the statistics of p_c between original and recovered signal is obvious in Fig. 2 (c) and (f).

III. Modeling and Learning of Channel Energy Statistics for Compressive Spectrum Sensing

In this section, we formulate the statistic model of channel energy in the CSS, and present an EM-based algorithm to learn the unknown parameters in the formulated model.

A. Model and Problem Formulation

To examine the statistics of the recovered signal of compressive spectrum sensing, the concise and direct way is through mathematical analysis of the probability density function of the output signal step by step following certain compressive sensing algorithm. However, commonplace compressive sensing algorithms include optimization sub-routines that have complex and highly-nonlinear forms [27]–[29], which makes the PDF derivation rather challenging. Instead, we tentatively postulate a general statistical model where the hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$ of recovered signal still conform to central and non-central chi-square distribution respectively with parameters relaxed as unknown. To express this postulated statistic model on the two hypotheses, we have

$$p_c = \begin{cases} \dfrac{\sigma^2}{2R} r_0, & r_0 \sim \chi^2(2R) \quad \mathcal{H}_0 \\ \beta r_1, & r_1 \sim \chi'^2(2R, \alpha) \quad \mathcal{H}_1 \end{cases} \tag{9}$$

where σ^2, R, α and β are no longer defined as in (4), (7) and (8), and are treated as unknowns which are to be estimated from the learning dataset. This effectively formulates an MM learning problem of two components in each channel. Specifically, after performing multiple runs of CS recovery of the spectrum, suppose we have a series of T recovered spectrum as the learning dataset. Thus, we can obtain CT samples of $p_c[t]$ $(t = 1, 2, \cdots, CT)$ according to (4). The expression of likelihood function is formulated as

$$\begin{aligned} &f(p_c; \sigma^2, R, \alpha, \beta, \pi_{\mathcal{H}_0}, \pi_{\mathcal{H}_1}) \\ &\quad = \pi_{\mathcal{H}_0} f_{\mathcal{H}_0}(p_c; \sigma^2, R) + \pi_{\mathcal{H}_1} f_{\mathcal{H}_1}(p_c; R, \alpha, \beta), \end{aligned} \tag{10}$$

where prior probability is represented as $\pi_{\mathcal{H}_0} = \Pr(Z = \mathcal{H}_0)$ and $\pi_{\mathcal{H}_1} = \Pr(Z = \mathcal{H}_1)$. The MM learning problem mentioned above is to find the optimal $\boldsymbol{\theta}^o$ of parameters

$$\boldsymbol{\theta} = [\sigma^2, R, \alpha, \beta, \pi_{\mathcal{H}_0}, \pi_{\mathcal{H}_1}] \tag{11}$$

to obtain maximal likelihood (ML) estimation of parameters, expressed by

$$\begin{aligned} \boldsymbol{\theta}^o &= \arg\max_{\boldsymbol{\theta}} \prod_{t=1}^{CT} f(p_c[t]; \boldsymbol{\theta}) \\ &= \arg\max_{\boldsymbol{\theta}} \sum_{t=1}^{CT} \log\left[f(p_c[t]; \boldsymbol{\theta})\right]. \end{aligned} \tag{12}$$

Problem (12) can be solved by EM algorithm [31] which sets a surrogate majorization function

$$\begin{aligned} &Q(\boldsymbol{\theta}; \boldsymbol{\theta}^{(j)}) \\ &= \mathrm{E}_{Z|p_c;\boldsymbol{\theta}^{(j)}} \left\{ \sum_{t=1}^{CT} \log\left[f(p_c[t], z[t]; \boldsymbol{\theta})\right] \right\} \\ &= \sum_{z=\mathcal{H}_0}^{\mathcal{H}_1} \sum_{t=1}^{CT} \Pr(Z = z | p_c = p_c[t]; \boldsymbol{\theta}^{(j)}) \log\left\{ \pi_z f_z(p_c[t]; \boldsymbol{\theta}) \right\} \end{aligned} \tag{13}$$

parameterized by $\boldsymbol{\theta}^{(j)}$ and maximize the surrogate function over $\boldsymbol{\theta}$ iteratively. The joint probability density in (13) of p_c and latent variable of hypothesis $Z = z$ in (13) is expressed as

$$\begin{aligned} f(p_c, z; \boldsymbol{\theta}) &= \pi_{\mathcal{H}_0} \delta(z = \mathcal{H}_0) f_{\mathcal{H}_0}(p_c; \sigma^2, R) \\ &\quad + \pi_{\mathcal{H}_1} \delta(z = \mathcal{H}_1) f_{\mathcal{H}_1}(p_c; R, \alpha, \beta), \end{aligned} \tag{14}$$

where $\delta(a = A)$ is indicator function which equals to 1 only if $a = A$ and elsewhere 0. It is proved [31] that by iteratively decreasing the surrogate function, each EM procedure will not decrease the objective until convergence to at least a local optimum.

B. EM Algorithm: Expectation

In expectation step, the probability term conditioned on $\boldsymbol{\theta}^{(j)}$ in (13) is called "membership probability" and obtained by the definition

$$\begin{aligned} M_z^{(j)}[t] &:= \Pr(Z = z | p_c = p_c[t]; \boldsymbol{\theta}^{(j)}) \\ &= \frac{\pi_z^{(j)} f_z(p_c[t]; \boldsymbol{\theta}^{(j)})}{\pi_{\mathcal{H}_0}^{(j)} f_{\mathcal{H}_0}(p_c[t]; \boldsymbol{\theta}^{(j)}) + \pi_{\mathcal{H}_1}^{(j)} f_{\mathcal{H}_1}(p_c[t]; \boldsymbol{\theta}^{(j)})}, \end{aligned} \tag{15}$$

where $z = \mathcal{H}_0, \mathcal{H}_1$.

C. EM Algorithm: Maximization

The maximization step of EM procedure finds the next update of parameters $\boldsymbol{\theta}^{(j+1)}$ by setting partial derivatives of $Q(\boldsymbol{\theta};\boldsymbol{\theta}^{(j)})$ to zero if closed-form partial derivatives are available. For prior probabilities, considering the normalization relationship $M_{\mathcal{H}_0}^{(j)} + M_{\mathcal{H}_1}^{(j)} = 1$, the update, which is irrelevant to PDF of chi-square distribution, simply follow

$$\begin{aligned}\frac{\partial Q(\boldsymbol{\theta};\boldsymbol{\theta}^{(j)})}{\partial \pi_{\mathcal{H}_0}} &= \sum_{t=1}^{CT} M_{\mathcal{H}_0}^{(j)}[t](\pi_{\mathcal{H}_0})^{-1} - M_{\mathcal{H}_1}^{(j)}[t](1-\pi_{\mathcal{H}_0})^{-1} \\ &= 0 \Rightarrow \pi_{\mathcal{H}_0}^{(j+1)} = \sum_{t=1}^{CT} \frac{M_{\mathcal{H}_0}^{(j)}}{CT} \end{aligned} \tag{16}$$

and

$$\pi_{\mathcal{H}_1}^{(j+1)} = 1 - \pi_{\mathcal{H}_0}^{(j+1)}. \tag{17}$$

Regarding updating σ^2, it only relates to central chi-square-like PDF $f_{\mathcal{H}_0}$, which has closed-form expression

$$f_{\mathcal{H}_0}(p_c;\sigma^2,R) = \begin{cases} \left(\frac{R}{\sigma^2}\right)^R \cdot \frac{(p_c)^{R-1} e^{-\frac{Rp_c}{\sigma^2}}}{\Gamma(R)}, & p_c > 0 \\ 0, & p_c \le 0. \end{cases} \tag{18}$$

Thus, the partial derivatives of logarithm PDF over σ^2 is not relevant to Gamma function $\Gamma(\cdot)$, and the update can be presented by,

$$\begin{aligned}\frac{\partial Q(\boldsymbol{\theta};\boldsymbol{\theta}^{(j)})}{\partial \sigma^2} &= \sum_{t=1}^{CT} M_{\mathcal{H}_0}^{(j)}[t]\Big(-\frac{R}{\sigma^2} + \frac{p_c[t]R}{\sigma^4}\Big) \\ &= 0 \Rightarrow (\sigma^2)^{(j+1)} = \frac{\sum_{t=1}^{T} M_{\mathcal{H}_0}^{(j)}[t]p_c[t]}{\sum_{t=1}^{T} M_{\mathcal{H}_0}^{(j)}[t]}. \end{aligned} \tag{19}$$

Exact optimization of the surrogate function over parameters R, α and β requires differentiating non-central chi-square PDF $f_{\mathcal{H}_1}$ which is known to have a modified Bessel function term including infinite series of Gamma function [32], so it is difficult to derive a handy expression of derivatives. The moment-matching method can be adopted as an alternative approach of ML to estimate parameters of MMs by directly matching the moment of mixture's PDF and moment estimations of samples [24]. In our case, however, the dimension of parameter vector $\boldsymbol{\theta}$ to be estimated is so large that we need to match high-order moments and solve high-order equations which is impractical. Although it is unfeasible to directly apply optimization or moment-matching estimation, we propose to use simpler moment-matching updates on the above mentioned parameters in EM's maximization step to induce increase of the surrogate function. Relating to the theory of EM algorithm [31], the non-decreasing property of objective function in (12) can be preserved as long as $Q(\boldsymbol{\theta}^{(j+1)};\boldsymbol{\theta}^{(j)}) \ge Q(\boldsymbol{\theta}^{(j)};\boldsymbol{\theta}^{(j)})$, thus the maximization step can be relaxed to an increasing step at a price of possibly slower convergence rate. To illustrate the moment-matching method in maximization step, we start by the update of K. The second-order central moment of hypothesis $\mathcal{H}_0$ is expressed as

$$\mathrm{Var}_{\mathcal{H}_0;\boldsymbol{\theta}} = \mathrm{E}\{(p_c - \sigma^2)^2 | Z = \mathcal{H}_0;\boldsymbol{\theta}\} = \sigma^4/R. \tag{20}$$

The second-order central moment estimation of samples $p_c[t]$ on condition of $\mathcal{H}_0$ and parameters $\boldsymbol{\theta}^j$ is expressed as (21). Note that we treat a prior occurrence probability of samples $\Pr(p_c = p_c[t])$ as uniform across all t.

$$\begin{aligned}&\sum_{t=1}^{CT}(p_c[t]-\sigma^2)^2 \cdot \Pr(p_c = p_c[t] | Z = \mathcal{H}_0;\boldsymbol{\theta}^{(j)}) \\ &= \frac{\sum_{t=1}^{CT} \Pr(Z = \mathcal{H}_0 | p_c = p_c[t];\boldsymbol{\theta}^{(j)})[p_c[t] - (\sigma^2)^{(j)}]^2}{\sum_{t=1}^{T} \Pr(Z = \mathcal{H}_0 | p_c = p_c[t];\boldsymbol{\theta}^{(j)})} \\ &= \frac{\sum_{t=1}^{CT} M_{\mathcal{H}_0}^{(j)}[t][p_c[t] - (\sigma^2)^{(j)}]^2}{\sum_{t=1}^{CT} M_{\mathcal{H}_0}^{(j)}[t]}. \end{aligned} \tag{21}$$

Then we use this second-order central moment estimation (21) over $\boldsymbol{\theta}^{(j)}$ to match the parameters given in (20) for next update $\boldsymbol{\theta}^{(j+1)}$,

$$\frac{[(\sigma^2)^{(j+1)}]^2}{R^{(j+1)}} = \frac{\sum_{t=1}^{CT} M_{\mathcal{H}_0}^{(j)}[t](p_c[t] - (\sigma^2)^{(j)})^2}{\sum_{t=1}^{CT} M_{\mathcal{H}_0}^{(j)}[t]}, \tag{22}$$

where $R^{(j+1)}$ is solved with $(\sigma^2)^{(j+1)}$ given in (19). Similarly, given $R^{(j+1)}$, the moment-matching updates for α and β is obtained by the first-order moment and second-order central moment of $\mathcal{H}_1$,

$$\begin{aligned}\mu_{\mathcal{H}_1;\boldsymbol{\theta}^{(j+1)}} &= \beta^{(j+1)}[2R^{(j+1)} + \alpha^{(j+1)}] \\ &= \frac{\sum_{t=1}^{CT} M_{\mathcal{H}_1}^{(j)}[t]p_c[t]}{\sum_{t=1}^{CT} M_{\mathcal{H}_1}^{(j)}[t]}, \end{aligned} \tag{23}$$

and

$$\begin{aligned}\mathrm{Var}_{\mathcal{H}_1;\boldsymbol{\theta}^{(j+1)}} &= 4[\beta^{(j+1)}]^2[R^{(j+1)} + \alpha^{(j+1)}] \\ &= \frac{\sum_{t=1}^{CT} M_{\mathcal{H}_1}^{(j)}[t]\{p_c[t] - \beta^{(j)}[2R^{(j)} + \alpha^{(j)}]\}^2}{\sum_{t=1}^{CT} M_{\mathcal{H}_1}^{(j)}[t]}. \end{aligned} \tag{24}$$

It should be noted that the updates using moment-matching method do not necessarily decrease $Q(\boldsymbol{\theta};\boldsymbol{\theta}^{(j)})$. In order to guarantee an explicit non-decreasing step, we propose an additional step where the moment-matching method only updates the corresponding parameter if the value of the surrogate function is not decreased by moment-maching method. This is accomplished by calculating the surrogate function $Q(\boldsymbol{\theta}^{(j+1)};\boldsymbol{\theta}^{(j)})$ and $Q(\boldsymbol{\theta}^{(j)};\boldsymbol{\theta}^{(j)})$ where the value of central and non-central chi-square PDFs can be well approximated by numerical methods. Moreover, as (23) and (24) consist of a quadratic equation set, it is noted that there will exist two sets of solutions $[\alpha_1^{(j+1)}, \beta_1^{(j+1)}]$ and $[\alpha_2^{(j+1)}, \beta_2^{(j+1)}]$. For each update, we choose the set of solution leading to the greater increase, if there is any, of the surrogate function.

To finalize this section, the proposed EM-based learning algorithm above is summarized as in Algorithm 1.

IV. Threshold Adaption via Noise Statistics Estimation in Compressive Spectrum Sensing

Although in Section III we have presented that our postulated statistical model (6), it is noted that Algorithm 1

Algorithm 1 EM-Based Learning of Channel Energy Statistics in Compressive Spectrum Sensing

Input: CT average channel PSD samples of $p_c[t]$, $t = 1, 2, \cdots, CT$.

Output: Parameters $\boldsymbol{\theta}^{(j)}$ of two hypotheses' PDF in (6) optimized by (12).

initialize $(\sigma^2)^{(0)} > 0$, $R^{(0)} > 0$, $\alpha^{(0)} > 0$, $\beta^{(0)} => 0$, $1 > \pi_{\mathcal{H}_0}^{(0)} = 1 - \pi_{\mathcal{H}_1}^{(0)} > 0$, $j = 0$.
repeat
  for each hypothesis $z = \mathcal{H}_0$ and $\mathcal{H}_1$ and each sample $t = 0$ to CT **do**
    update membership probability $M_z^{(j+1)}[t]$ as in (15)
  end for
  update $(\sigma^2)^{(j+1)}$ as in (19)
  update $R^{(j+1)}$ as in (22)
  $\boldsymbol{\theta}_{\text{temp0}} \leftarrow \boldsymbol{\theta}^{(j)}$ with element $R^{(j)}$ replaced by $R^{(j+1)}$
  if $Q(\boldsymbol{\theta}_{\text{temp0}}; \boldsymbol{\theta}^{(j)}) < Q(\boldsymbol{\theta}^{(j)}; \boldsymbol{\theta}^{(j)})$ **then**
    $R^{(j+1)} \leftarrow R^{(j)}$
  end if
  update $\pi_{\mathcal{H}_0}^{(j+1)}$ and $\pi_{\mathcal{H}_1}^{(j+1)}$ as in (16) and (17)
  solve (23) (24) to get two sets of solutions $[\alpha_1^{(j+1)}, \beta_1^{(j+1)}]$ and $[\alpha_2^{(j+1)}, \beta_2^{(j+1)}]$
  $\boldsymbol{\theta}_{\text{temp1}} \leftarrow \boldsymbol{\theta}^{(j)}$ with elements $\alpha^{(j)}$ and $\beta^{(j)}$ replaced by $\alpha_1^{(j+1)}$ and $\beta_1^{(j+1)}$
  $\boldsymbol{\theta}_{\text{temp2}} \leftarrow \boldsymbol{\theta}^{(j)}$ with elements $\alpha^{(j)}$ and $\beta^{(j)}$ replaced by $\alpha_2^{(j+1)}$ and $\beta_2^{(j+1)}$
  if $Q(\boldsymbol{\theta}_{\text{temp1}}; \boldsymbol{\theta}^{(j)}) > Q(\boldsymbol{\theta}_{\text{temp2}}; \boldsymbol{\theta}^{(j)})$ **then**
    $\alpha^{(j+1)} \leftarrow \alpha_1^{(j+1)}$, $\beta^{(j+1)} \leftarrow \beta_1^{(j+1)}$
  else
    $\alpha^{(j+1)} \leftarrow \alpha_2^{(j+1)}$, $\beta^{(j+1)} \leftarrow \beta_2^{(j+1)}$
  end if
  if $Q(\boldsymbol{\theta}_{\text{temp1}}; \boldsymbol{\theta}^{(j)}) < Q(\boldsymbol{\theta}^{(j)}; \boldsymbol{\theta}^{(j)})$ and $Q(\boldsymbol{\theta}_{\text{temp2}}; \boldsymbol{\theta}^{(j)}) < Q(\boldsymbol{\theta}^{(j)}; \boldsymbol{\theta}^{(j)})$ **then**
    $\alpha^{(j+1)} \leftarrow \alpha^{(j)}$, $\beta^{(j+1)} \leftarrow \beta^{(j)}$
  end if
  $j \leftarrow j + 1$
until $||Q(\boldsymbol{\theta}^{(j)}; \boldsymbol{\theta}^{(j)}) - Q(\boldsymbol{\theta}^{(j-1)}; \boldsymbol{\theta}^{(j)}||_2 < \epsilon$

is impractical in CSS applications due to the following reasons:

1. Algorithm 1 involves multiple calculations of the surrogate function in each iteration, which requires considerable computational effort;

2. As a common drawback of EM algorithms, Algorithm 1 can converge to one of many local maxima. Setting initial values $\boldsymbol{\theta}^{(0)}$ close to the real maxima helps the algorithm converge to the global maximum [33]. In order to find the global maximum, a common practice is that the algorithm should run multiple times with $\boldsymbol{\theta}^{(0)}$ randomly valued, which adds more computational complexity. Alternatively, human involvement to choose the proper $\boldsymbol{\theta}^{(0)}$ or other initial value selection scheme should be applied;

3. In real-world applications, the sensed channel energy or channel model of each channel is hardly likely to be the same. Hence the samples from active channels would be drawn from differently parameterized hypothesis $\mathcal{H}_1$, which may lead to failure to fit with the postulated statistical model (6).

Due to these impracticalities, in this section, we propose a robust and practical threshold adaption scheme via noise statistics estimation based on the verified statistical model of hypothesis $\mathcal{H}_0$ in (6). However, these drawbacks do not decrease the necessity of Algorithm 1. The purpose of Algorithm 1 is not for practice after all - it is proposed to verify the postulated statistical model (9) and for the general interest of the parameters $\boldsymbol{\theta}$ with various CSS settings.

A. The Proposed Threshold Adaption Scheme

In this subsection, a threshold adaption scheme which specifically aims to achieve CFAR in detection phase is presented. According to NP detection theory, CFAR only relates to hypothesis $\mathcal{H}_0$, which means that only parameter learning of $\mathcal{H}_0$ is required. In the following, procedures of the proposed threshold adaption scheme are detailed.

1) Identification of Vacant Channels in the Learning Dataset: Given learning dataset being T' samples of spectrum recovered by CS algorithm, prior to estimating the parameters relating to $\mathcal{H}_0$, the first step of the proposed threshold adaption is to initially identify these samples of hypothesis $\mathcal{H}_0$. Specifically, given $t = 0, 1, \cdots, T'$ observations of recovered spectrum and consequently $p_c^{(c)}[t]$'s, we try to identify these channels that are free throughout these observations. Due to sparse spectrum usage and rapid observation acquisition, it is a reasonable assumption that the channel occupancy is static with at least one occupied and multiple vacant channels during the acquisition process. This identification problem falls into the category of clustering [34]. The average channel PSD level $p_c^{(c)}$ of these vacant channels conforms to the same distribution; however, other channels have different statistics (usually larger mean and standard deviation) from these of vacant channels. We exploit this feature and k-means clustering algorithm to identify these vacant channels. In this specific problem, k-means clustering are conducted in two dimensions - sample means and sample standard deviation - to produce two clusters. K-means is a basic model of clustering which aims to find the clustering solution with minimum intra-cluster distance. In this particular case, it aims to solve the following optimization problem to obtain the vacant channel set $\mathcal{C}$, which writes

$$\mathcal{C} = \arg\min_{\mathcal{X}} \sum_{\mathcal{X}=\mathcal{S},\mathcal{S}\setminus\mathcal{X}} \sum_{k\in\mathcal{X}} \left\| \boldsymbol{\omega}^{(k)} - \sum_{v\in\mathcal{X}} \boldsymbol{\omega}^{(v)}/|\mathcal{X}| \right\|_2, \quad (25)$$

where $\mathcal{S} = 1, 2, \cdots, C$ is the set of all channels and

$$\boldsymbol{\omega}^{(c)} = \left(\frac{\sum_{t=1}^{T'} p_c^{(c)}[t]}{T'}, \sqrt{\frac{\sum_{t=1}^{T'} \left[p_c^{(c)}[t] - \frac{\sum_{\tau=1}^{T'} p_c^{(c)}[\tau]}{T'} \right]^2}{T' - 1}} \right). \quad (26)$$

Details of k-means clustering algorithm are not to be elaborated in this paper and readers are referred to [34] and [35].

2) ML Estimation of Parameters in $\mathcal{H}_0$: After these vacant channels being identified, the samples of these channels are used to estimate the parameters R and σ^2 of chi-square-like distribution in (18). Denote the number of identified vacant channels as $C' = |\mathcal{C}|$. ML estimator which maximizes the joint probability of each sample in the dataset for estimation [36] is adopted, expressed by

$$\begin{aligned}(R^o, (\sigma^2)^o) &= \arg\max_{R,\sigma^2} L(\sigma^2, R) \\ &= \arg\max_{R,\sigma^2} \prod_{t=1}^{T'} \prod_{c\in\mathcal{C}} f_{\mathcal{H}_0}(p_c^{(c)}[t]; \sigma^2, R). \end{aligned} \quad (27)$$

The maximization[1] is achieved by solving the stationary point of the log likelihood function, where

$$\begin{aligned}\frac{\partial \log[L(\sigma^2, R)]}{\partial \sigma^2} &= \frac{C'T'R}{\sigma^2} - \frac{R\sum_{t=1}^{T'}\sum_{c\in\mathcal{C}} p_c^{(c)}[t]}{\sigma^4} = 0 \\ &\Rightarrow (\sigma^2)^o = \frac{\sum_{t=1}^{T'}\sum_{c\in\mathcal{C}} p_c^{(c)}[t]}{C'T'} \end{aligned} \quad (28)$$

and

$$\begin{aligned}&\frac{\partial \log[L(\sigma^2, R)]}{\partial R} \\ &= C'T'[\log R + 1 - \log\sigma^2 - \frac{\Gamma'(R)}{\Gamma(R)}] \\ &\quad + \sum_{t=1}^{T'}\sum_{c\in\mathcal{C}} \log p_c^{(c)}[t] - \frac{\sum_{t=1}^{T'}\sum_{c\in\mathcal{C}} p_c^{(c)}[t]}{\sigma^2} = 0. \end{aligned} \quad (29)$$

$(\sigma^2)^o$ can be easily solved in (28). However, R^o is not easy to solve analytically but can be solved numerically. Use Stirling's expansion [37] of digamma function when R is relatively large

$$\begin{aligned}\frac{\Gamma'(R)}{\Gamma(R)} = \log(R) - \frac{1}{2R} - \frac{1}{12R^2} + \frac{1}{120R^4} \\ - \frac{1}{252R^6} + O\left(\frac{1}{R^8}\right), \end{aligned} \quad (30)$$

and adopt a partial cut-off sum of the series to insert into (29). Thus the equation (29) can be solved by Newton-Ralphson method which implements the iteration

$$R^{(j+1)} \leftarrow R^{(j)} - \left.\frac{\partial \log[L(\sigma^2, R)]/\partial R}{\partial^2 \log[L(\sigma^2, R)]/\partial R^2}\right|_{R=R^{(j)}}, \quad (31)$$

where the first and second partial derivatives are easy to obtain as it only relates to polynomial of R. The initial value of iteration should be near the root to ensure fast convergence and it can be obtained by moment-matching. Recall (20), and the initial value is obtained by

$$R^{(0)} = \frac{C'T'[(\sigma^2)^o]^2}{\sum_{t=1}^{T'}\sum_{c\in\mathcal{C}}[p_c^{(c)}[t] - (\sigma^2)^o]^2}. \quad (32)$$

[1]Examine the equations (28) and (29). It is noted that partial derivatives of $\log\left[L(\sigma^2, R)\right]$ with regards to $R > 0$ and σ^2 both have only one zero point. Hence global optimality is guaranteed.

3) Threshold Adaption: NP tests on channel energy is deployed based on the threshold δ

$$p_c^{(c)} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \delta, \quad (33)$$

where the threshold can be determined by CFAR strategy. Probability of false alarm are found to satisfy the constant complementary cumulative probability density (CCPD) of $\mathcal{H}_0$,

$$P_f = \Pr(p_c^{(c)} > \delta|\mathcal{H}_0) = \frac{\Gamma(R, \frac{R\delta}{\sigma^2})}{\Gamma(R)}, \quad (34)$$

where $\Gamma(a, b)$ is the incomplete Gamma function. For given P_f in CFAR, the corresponding threshold δ can be easily found numerically, for example, by binary search method, as CCPD values can be computed[2] and is monotonically decreasing.

B. Asymptotic Performance of the Proposed Threshold Adaption Scheme

Here we analyze the asymptotic performance of the noise energy statistics estimation and threshold adaption. As a direct conclusion from central limit theorem (CLT), the ML estimator (27) has the property of asymptotic minimum-variance and unbiased estimator [36], specifically,

$$\left(R^o, (\sigma^2)^o\right) \overset{C'T'}{\sim} \mathcal{N}\left((R, \sigma^2), \mathbf{I}^{-1}(R, \sigma^2)\right), \quad (35)$$

where Fisher information matrix is

$$\begin{aligned}&\mathbf{I}(R, \sigma^2) \\ &= \begin{bmatrix} -\mathrm{E}\left[\frac{\partial^2 \log[L(\sigma^2, R)]}{\partial R^2}\right] & -\mathrm{E}\left[\frac{\partial^2 \log[L(\sigma^2, R)]}{\partial R \partial \sigma^2}\right] \\ -\mathrm{E}\left[\frac{\partial^2 \log[L(\sigma^2, R)]}{\partial \sigma^2 \partial R}\right] & -\mathrm{E}\left[\frac{\partial^2 \log[L(\sigma^2, R)]}{\partial (\sigma^2)^2}\right] \end{bmatrix} \\ &= \begin{bmatrix} C'T'\left(\frac{1}{R} - \frac{\Gamma''(R)}{\Gamma'(R)}\right) & 0 \\ 0 & \frac{C'T'R}{\sigma^4} \end{bmatrix}. \end{aligned} \quad (36)$$

The variance of estimated parameters $\left(R^o, (\sigma^2)^o\right)$ asymptotically approaches Cramer-Rao lower bound,

$$\mathrm{Var}(R^o) \overset{C'T'}{\sim} \frac{1}{C'T'}\left(\frac{1}{R} - \frac{\Gamma''(R)}{\Gamma'(R)}\right)^{-1}, \quad (37)$$

and

$$\mathrm{Var}\left((\sigma^2)^o\right) \overset{C'T'}{\sim} \frac{\sigma^4}{C'T'R}. \quad (38)$$

With the increasing and sufficiently large number of samples $C'T'$, these lower bounds (37) and (38) can be approached asymptotically and well approximates the real variance, which represents the analytic performance of the ML estimation (27).

Moreover, the variance of the real P_f is another interested parameter, as it indicates how well the proposed noise statistics estimation and threshold adaption can determine a threshold to keep a constant P_f. To quantify the variance of P_f, the asymptotic variance bounds (37) and (38) are used. To obtain the closed-form expression of the variance of P_f, the variance of P_f with regards to random variables R^o and $(\sigma^2)^o$ should

[2]Note R in (34), as an output of the proposed ML estimation, can be arbitrary positive value. In simulations we use Matlab functions `chi2cdf` to calculate CCPD.

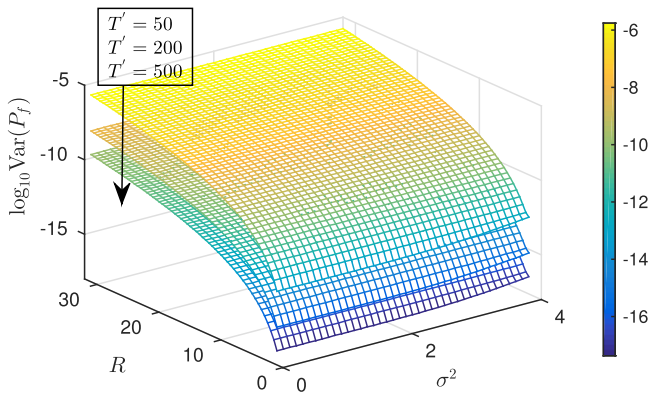


Fig. 3. Asymptotic variance lower bound (logarithm scale) of P_f which is determined by proposed ML estimation and threshold adaption against the two parameters.

be derived using the relationship (34). As the expression (34) has complex formulation which contains incomplete Gamma function, the derivation of the variance of P_f is not a trivial task. Alternatively, Monte-Carlo simulations are used to inspect this variance, where $T' = 34$ (the same value used in Section V) is adopted and sample variance is used as the approximation to the real variance. The legitimacy behind this Monte-Carlo simulation is the law of large numbers - for the simulation times large enough, the sample variance of multiple independent observations of P_f converges to the variance of the random variable P_f. Fig. 3 gives the numerical results of variance lower bound of P_f against R and σ^2 given T' value of 50, 200 and 500 respectively. It is noted that the variance of P_f is irrelevant of parameter σ^2, as a result of the fact that σ^2 is a factor only accounting for the dilation of the channel energy statistics in (9). However, P_f has shown increasing variance with R evidently. With the interested range of $R \in (0, 32]$ in Fig. 3, the variance of P_f compared with the target value can be kept small, by selecting a proper value of T'.

V. Numerical Analysis

In this section, simulation results of the proposed signal energy statistics algorithm and threshold adaption via noise statistics estimation are presented. We generate time-domain signals as in (39) to emulate signal with continuous support in frequency domain,

$$s_t{}^{(n)} = \sum_{i=1}^{\mathcal{K}} \sqrt{E_i B}\mathrm{sinc}[B(n - n_i)]e^{j2\pi f_i t}, \qquad (39)$$

where $\mathcal{K}$ is the number of active channels; B is channel bandwidth; E_i, n_i, f_i stands for total energy, time offset and central frequency of the ith active channel. To simulate the band of TVWS in UK (470-790MHz), we choose the interested spectrum bandwidth to be 320MHz and set $N = 1280$, $B = 8$MHz, $\mathcal{K} = 6$, and $E_i = 1280/6$. For each combination of recovery algorithm, SNR and compression ratio, learning dataset is generated by the channel occupancy patterns as $[f_1, \cdots, f_6] = [36, 44, 164, 172, 244, 252]$(MHz). SNR is defined as SNR $= ||\boldsymbol{s_t}||_2^2/(N\sigma^2)$ where σ^2 is the power of complex zero-mean additive noise $w_t^{(n)} \sim \mathcal{CN}(0, \sigma^2)$.

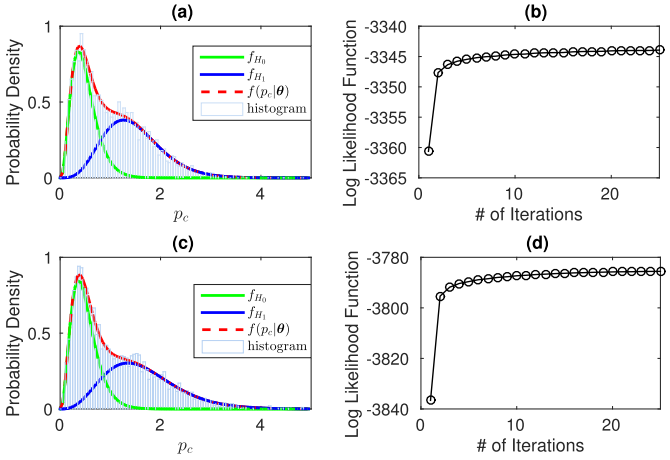


Fig. 4. **(a)(c)** Histogram and learned distributions of BP recovery, SNR $=$ 0dB, compression ratio of 0.2, in AWGN and Rayleigh channel respectively; **(b)(d)** Likelihood function over first 25 iterations of learning process in **(a)(c)**, respectively.

A. The Effectiveness of Channel Energy Statistics Learning for Compressive Spectrum Sensing

In this subsection, we focus on and implement the proposed channel energy statistics learning algorithm using the samples $p_c[t]$ from channel set whose central frequencies are $[f_1, \cdots, f_6]$ and $[f_7, \cdots, f_{12}] = [60, 68, 188, 196, 268, 276]$(MHz) so that this set of $C = 12$ channels has channel occupancy rate of 0.5. The dimension of the dataset is $T = 300$.

We experiment the proposed learning algorithm over recovered signals by three major sparse recovery algorithms - BP, OMP and SBL, and two propagation scenarios - AWGN and Rayleigh channel. Fig. 4 (a) and (c) exemplify the distributions of average channel PSD level $p_c^{(c)}$ conditioned on hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$ and unconditional distribution whose parameters are learned from the proposed algorithm, using BP, compression ratio of 0.2 and SNR of 0dB, over AWGN and Rayleigh channel respectively. It can be seen that the learned distributions align well with the histogram of learning dataset in both cases. Fig. 4 (b) and (d) correspond to (a) and (c) respectively and show the increasing and convergent results of likelihood functions which the proposed learning algorithm aims to maximize. Fig. 5 (a) and (b) show the learning results of signals recovered by OMP and SBL algorithm respectively. Again, we set SNR to 0dB in AWGN channel and compression ratio of 0.5 and 0.25 in (a) and (b) respectively. The examples in Fig. 5 indicate that the learning results may also describe the empirical distribution of learning dataset with OMP and SBL recovery algorithms.

To further characterize how well the empirical distribution from simulation matches our postulated model in (9), we adopt the Kullback-Leibler divergence (KLD) as a measure of the similarity of the two distribution [38]. KLD may be any non-negative value. KLD of 0 reveals two identical distributions, and a smaller KLD near 0 indicates more similarity. From simulations, the learning dataset forms a histogram of empirical distribution, instead of a continuous probability

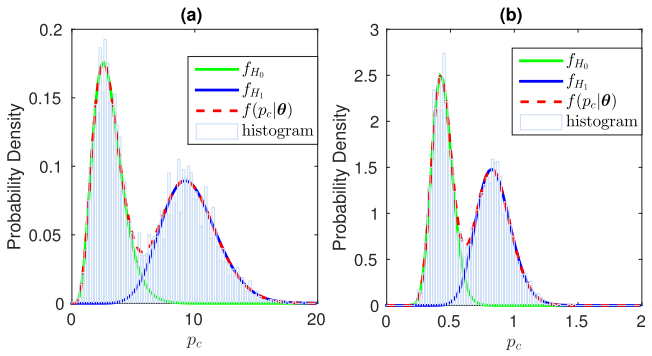


Fig. 5. Histogram and learned distributions, SNR = 0dB in AWGN channel, using **(a)** OMP with compression ratio of 0.5 and **(b)** SBL with compression ratio of 0.25.

density function. Therefore, we adopt the discrete probability form of KLD definition,

$$D_{\mathrm{KL}}(P_{\mathrm{empirical}} \| P_{\mathrm{model}}) = \sum_{i=1}^{I} P_{\mathrm{empirical}}(i) \log \left(\frac{P_{\mathrm{empirical}}(i)}{P_{\mathrm{model}}(i)} \right), \tag{40}$$

where $P_{\mathrm{empirical}}(i)$ is the histogram value of empirical distribution, normalized by

$$\sum_{i=1}^{I} P_{\mathrm{empirical}}(i) = 1; \tag{41}$$

$P_{\mathrm{model}}(i)$ is the normalized probability in the ith interval of the postulated model, expressed by

$$P_{\mathrm{model}}(i) = \frac{\int_{i\text{th interval}} f_z(p_c; \boldsymbol{\theta}^o) \mathrm{d}p_c}{\int_{\text{all intervals in } \mathcal{I}} f_z(p_c; \boldsymbol{\theta}^o) \mathrm{d}p_c}, \quad z = \mathcal{H}_0, \mathcal{H}_1. \tag{42}$$

The range of interested intervals is determined as

$$[a, b] = \Big[\min \{p_c[t] \,|t = 1, 2, \cdots, T'\}, \max \{p_c[t] \,|t = 1, 2, \cdots, T'\} \Big]. \tag{43}$$

Then the interval is evenly divided into I (in simulations $I = 250$ is used) smaller ones such that the ith interval corresponds to $\left[a + (i-1)\frac{b-a}{I}, a + i\frac{b-a}{I}\right]$. Fig. 6 plots the KLD values from empirical data to our postulated model, for three kinds of interested algorithms, against compression ratio and SNR respectively. In all cases of our simulations, KLD values turn out to be low, within the range from 10^{-2} to 5×10^{-2}. This result verified that the proposed learning algorithm fits well with our postulated model.

Next, we change the compression ratio and focus on the parameter changes learned by the proposed algorithm. Fig. 7 gives these results of three recovery algorithms in AWGN channel under the condition SNR = 0dB. Moreover, each parameter value of the average channel PSD level distribution of the original signal is drawn as a reference. As a general and intuitive trend, most parameters from signals recovered by three algorithms approach to the orignal signal's reference with increasing compression ratio. An interesting exception is R of SBL which fluctuates slightly near 32. This can be explained by SBL's assumption of Gaussian prior distribution

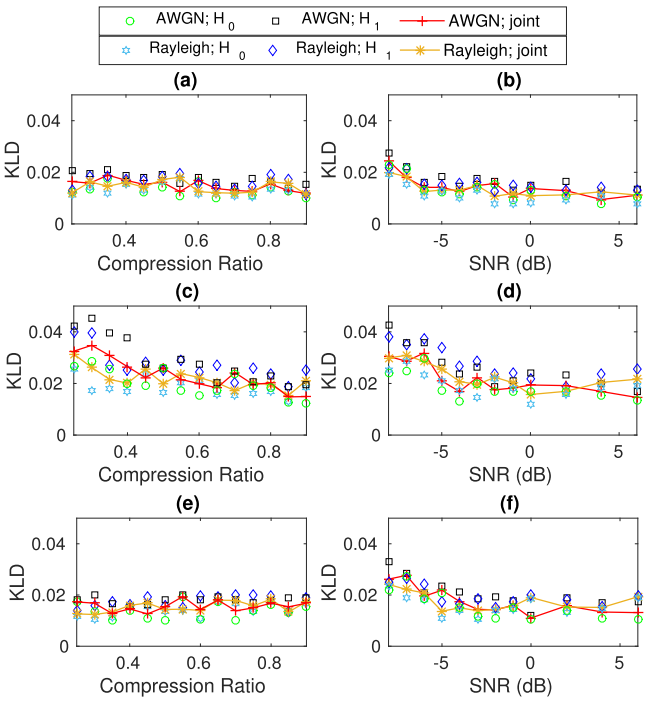


Fig. 6. KLD from empirical data to postulated models versus compression ratio and SNR. **(a)(b)** BP; **(c)(d)** OMP; **(e)(f)** SBL.

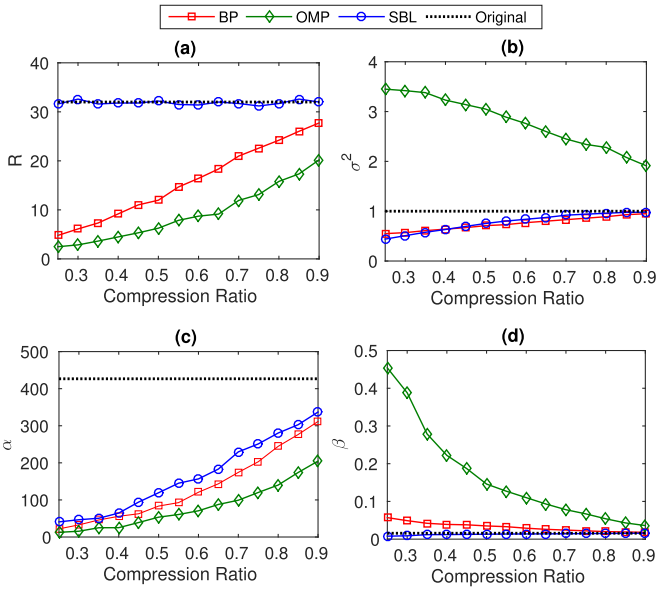


Fig. 7. Learned parameters against compression ratio using three major recovery algorithms, compared with original signal parameters. SNR = 0dB, AWGN channel.

of the signal to be recovered [28]. Given this assumption and by the definition of chi-square distribution, the distribution of average channel PSD level has a definite DoF which equals twice of the number of PSD bins in each channel, considering complex Gaussian noise over each PSD bin. Fig. 8 gives similar results in Rayleigh channel situation with the same settings as in Fig. 7. It is noted that the SBL's parameter α is negligible and approximates the original signal's reference of 0. In Rayleigh channel scenario, not only the elements in $\boldsymbol{s_f}$ corresponding to the hypothesis $\mathcal{H}_0$ but also $\mathcal{H}_1$ have Gaussian

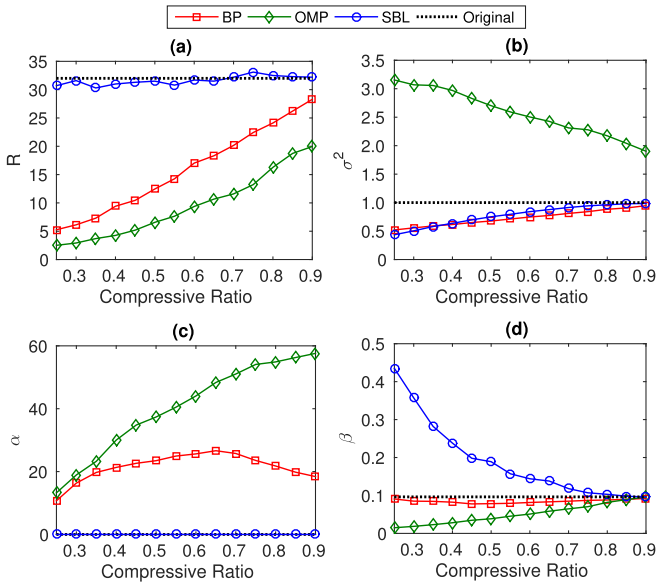


Fig. 8. Learned parameters against compression ratio using three major recovery algorithms, compared with original signal parameters. SNR = 0dB, Rayleigh channel.

distribution. This prior distribution of original signals matches the SBL's Gaussian assumption. As a result, SBL induces the Gaussainity of the original signal set, which leads to central chi-square statistics in $\mathcal{H}_1$ in the recovered signal.

B. Performance of Noise Energy Estimation and Threshold Adaption

In this section, we evaluated the performance of the proposed noise energy estimation and threshold adaption in Section III. As the first step, k-means clustering is performed for several learning datasets with different SNRs and dataset dimensions. Qualitatively, it is obvious from CLT that the larger dataset dimension is, the more concentrated the statistics $\boldsymbol{\omega}^{(c)}$ within one cluster (for either vacant or incumbent channels) tend to be. Thus, it means that the global maximum of clustering objective (25) is more likely to correspond to the true channel occupancy pattern. On the other hand, in lower SNR cases, the statistics $\boldsymbol{\omega}^{(c)}$ from vacant and incumbent channel clusters are less seperated. To investigate the effectiveness of the proposed k-means clustering step, we illustrate the clustering results with a small dataset ($T' = 20$) and low SNR (-10dB), accompanied by other three clustering results with larger dataset $T' = 200$ and higher SNR of -8dB. Although the clustering performance tends to be deteriorated by decreasing SNR and sample size T', the results in Fig. 9 show that the proposed k-means clustering suffices to identify the vacant channels in a worst case with $SNR = -10$dB and small dataset size $T' = 20$. Here, the SNR of -10dB is a level where compressive spectrum recovery becomes unreliable so that the detection performance is not usable.

Next, we extract the samples of $p_c^{(c)}$ from identified vacant channels, perform ML estimation based on $T' = 200$ spectrum samples and obtain threshold δ given $P_f = 0.01$. In Fig. 10 and Fig. 11, the adapted thresholds are plotted

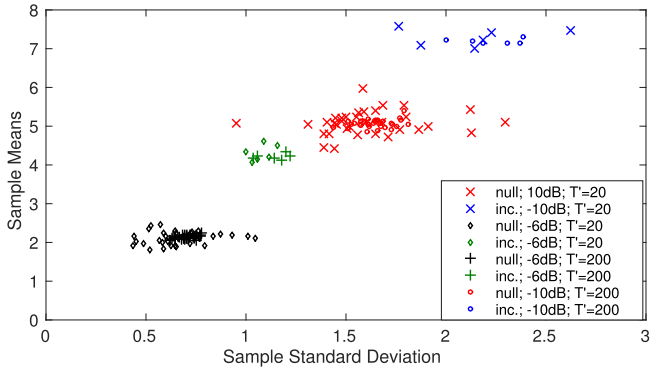


Fig. 9. K-means clustering for identifying the vacant channels from incumbent ones. It is shown that the clustering algorithm successfully identifies the vacant channels with low SNR (-10dB) and small dataset ($T' = 20$).

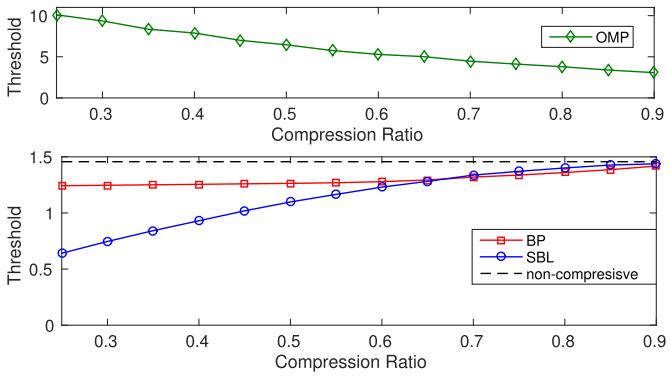


Fig. 10. Threshold adapted from our proposed scheme against compression ratio of three major recovery algorithms, given target $P_f = 0.01$. SNR = 0dB.

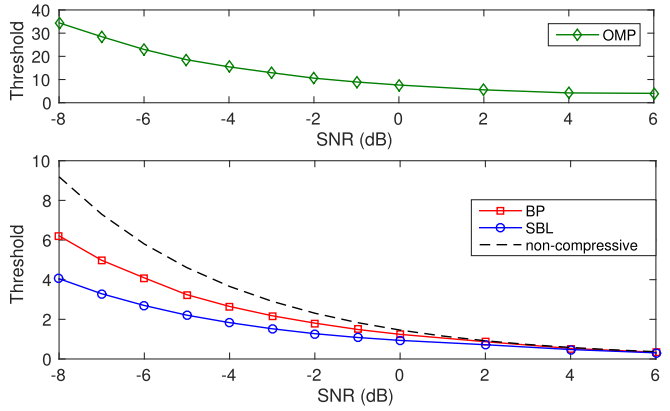


Fig. 11. Threshold adapted from our proposed scheme against SNR of three major recovery algorithms, given target $P_f = 0.01$. Compression ratio is 0.4.

for three major recovery algorithms, against compression ratio and SNR respectively. To stress the necessity of the proposed threshold adaption scheme, again, thresholds for non-compressive cases are calculated based on original signals $\boldsymbol{s_t}$ and drawn for reference. These two figures are direct illustrations of how evidently discrepant the thresholds should be valued among varying SNRs, compression ratios and different choices of recovery algorithms, and especially between non-compressive and compressive cases. Moreover, as an intuitive rule, thresholds of all three CS algorithms

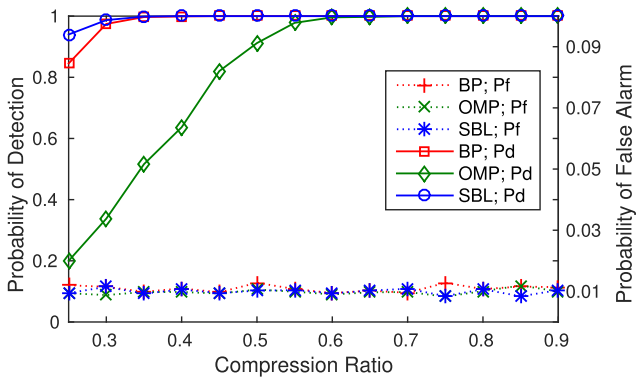


Fig. 12. CFR detection performance of three major recovery algorithms against compression ratio using proposed noise energy statistics estimation and threshold adaption. SNR = 0dB.

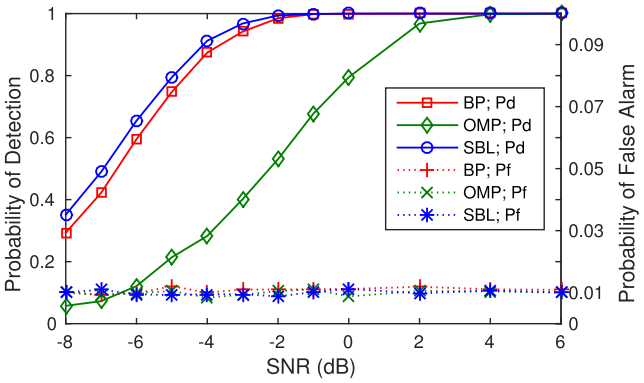


Fig. 13. CFR detection performance of three major recovery algorithms against SNR using proposed noise energy statistics estimation and threshold adaption. Compression ratio is 0.4.

approach these of non-compressive cases with compression ratio and SNR increasing.

Using the thresholds adapted, we perform Monte-Carlo test of the detection performance by using these thresholds for detection over a new sample set of recovered spectrum. Note that the dataset we used here for detection performance simulation is other than the learning dataset we used for threshold adaption. Fig. 12 illustrates the probability of detection and false alarm against compression ratio using the proposed threshold adaption scheme for the three major recovery algorithms with SNR = 0dB. It is discovered that our proposed threshold adaption scheme can maintain a close approximate of predefined P_f of 0.01. Similarly, Fig. 13 gives CFAR detection performance against SNR for the three selected recovery algorithms while keeping compression ratio 0.4. Again, the actual P_f tends to maintain constant with minor variance. Moreover, the results of P_f which are kept near constant from Fig. 12 and Fig. 13 in turn verify the reasonableness of our postulated model (9) and the effectiveness of our proposed threshold adaption scheme via noise energy statistics estimation.

VI. Conclusion

In this paper, we for the first time addressed the inconsistency of channel energy statistics and hence optimal threshold between CSS and conventional non-compressive spectrum sensing. Then a channel energy statistics model was postulated for CSS and an algorithm used for learning the parameters of the postulated model was proposed. We postulated that the two hypotheses of channel energy of absent and active radio activity conform to central and noncentral chi-square distributions, respectively, both with unknown parameters. Given the proposed statistics model being validated, a robust and practical scheme of threshold adaption achieved by channel noise statistics estimation was proposed. The asymptotic performance of the scheme was analyzed by presenting the variance lower bound of actual false alarm probability. By numerical simulations, it was shown that our postulated model provided good fit to the empirical distributions, in the sense of data-to-model KLD. Next, out of interest of the trend of these parameters in different CS settings, we experimented our learning algorithm in various compression ratio settings and presented the trend of each parameter versus compression ratio. Furthermore, the adapted thresholds given target false alarm probability were plotted against SNR and compression ratio, where the discrepancies in thresholds between compressive and non-compressive spectrum sensing are directly illustrated. Moreover, the actual false alarm probability using the thresholds adapted by our proposed scheme was shown to be near the target value, which inversely verified the validity of our postulated model.

References

[1] Office of Communications, U.K. (Jul. 2009). *Statement on Licence-Exempting Cognitive Devices Using Interleaved Spectrum*. [Online]. Available: https://www.ofcom.org.uk/__data/assets/pdf_file/0023/40838/statement.pdf

[2] J. J. Meng, W. Yin, H. Li, E. Hossain, and Z. Han, "Collaborative spectrum sensing from sparse observations in cognitive radio networks," *IEEE J. Sel. Areas Commun.*, vol. 29, no. 2, pp. 327–337, Feb. 2011.

[3] Z. Qin, Y. Gao, M. D. Plumbley, and C. G. Parini, "Wideband spectrum sensing on real-time signals at sub-Nyquist sampling rates in single and cooperative multiple nodes," *IEEE Trans. Signal Process.*, vol. 64, no. 12, pp. 3106–3117, Jun. 2016.

[4] Z. Qin, Y. Gao, and C. G. Parini, "Data-assisted low complexity compressive spectrum sensing on real-time signals under sub-Nyquist rate," *IEEE Trans. Wireless Commun.*, vol. 15, no. 2, pp. 1174–1185, Feb. 2016.

[5] Y. Ma, Y. Gao, Y.-C. Liang, and S. Cui, "Reliable and efficient sub-Nyquist wideband spectrum sensing in cooperative cognitive radio networks," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 10, pp. 2750–2762, Oct. 2016.

[6] X. Zhang, Y. Ma, Y. Gao, and S. Cui, "Real-time adaptively regularized compressive sensing in cognitive radio networks," *IEEE Trans. Veh. Technol.*, vol. 67, no. 2, pp. 95–103, Feb. 2017.

[7] F. Salahdine, N. Kaabouch, and H. El Ghazi, "A survey on compressive sensing techniques for cognitive radio networks," *Phys. Commun.*, vol. 20, pp. 61–73, Sep. 2016.

[8] M. Bkassiny, S. K. Jayaweera, Y. Li, and K. A. Avery, "Wideband spectrum sensing and non-parametric signal classification for autonomous self-learning cognitive radios," *IEEE Trans. Wireless Commun.*, vol. 11, no. 7, pp. 2596–2605, Jul. 2012.

[9] M. Yang, Y. Li, X. Liu, and W. Tang, "Cyclostationary feature detection based spectrum sensing algorithm under complicated electromagnetic environment in cognitive radio networks," *China Commun.*, vol. 12, no. 9, pp. 35–44, Sep. 2015.

[10] Z. Tian, Y. Tafesse, and B. M. Sadler, "Cyclic feature detection with sub-Nyquist sampling for wideband spectrum sensing," *IEEE J. Sel. Topics Signal Process.*, vol. 6, no. 1, pp. 58–69, Feb. 2012.

[11] M. O. Mughal, T. Nawaz, L. Marcenaro, and C. S. Regazzoni, "Cyclostationary-based jammer detection algorithm for wide-band radios using compressed sensing," in *Proc. IEEE Global Conf. Signal Inf. Process. (GlobalSIP)*, Orlando, FL, USA, Dec. 2015, pp. 280–284.

[12] W. Zhang, R. K. Mallik, and K. B. Letaief, "Optimization of cooperative spectrum sensing with energy detection in cognitive radio networks," *IEEE Trans. Wireless Commun.*, vol. 8, no. 12, pp. 5761–5766, Dec. 2009.

[13] Z. Ye, G. Memik, and J. Grosspietsch, "Energy detection using estimated noise variance for spectrum sensing in cognitive radio networks," in *Proc. IEEE Wireless Commun. Netw. Conf.*, Las Vegas, NV, USA, Mar./Apr. 2008, pp. 711–716.

[14] F. F. Digham, M.-S. Alouini, and M. K. Simon, "On the energy detection of unknown signals over fading channels," *IEEE Trans. Commun.*, vol. 55, no. 1, pp. 21–24, Jan. 2007.

[15] D. A. Pados, P. Papantoni-Kazakos, D. Kazakos, and A. G. Koyiantis, "On-line threshold learning for Neyman–Pearson distributed detection," *IEEE Trans. Syst., Man, Cybern.*, vol. 24, no. 10, pp. 1519–1531, Oct. 1994.

[16] D. R. Joshi, D. C. Popescu, and O. A. Dobre, "Adaptive spectrum sensing with noise variance estimation for dynamic cognitive radio systems," in *Proc. Annu. Conf. Inf. Sci. Syst.*, Mar. 2010, pp. 1–5.

[17] S. Gong, W. Liu, W. Yuan, W. Cheng, and S. Wang, "Threshold-learning in local spectrum sensing of cognitive radio," in *Proc. IEEE Veh. Technol. Conf. Spring*, Apr. 2009, pp. 1–6.

[18] M. Mishali and Y. C. Eldar, "From theory to practice: Sub-Nyquist sampling of sparse wideband analog signals," *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 2, pp. 375–391, Apr. 2010.

[19] J. A. Tropp, J. N. Laska, M. F. Duarte, J. K. Romberg, and R. G. Baraniuk, "Beyond Nyquist: Efficient sampling of sparse bandlimited signals," *IEEE Trans. Inf. Theory*, vol. 56, no. 1, pp. 520–544, Jan. 2010.

[20] T. T. Cai, L. Wang, and G. Xu, "New bounds for restricted isometry constants," *IEEE Trans. Inf. Theory*, vol. 56, no. 9, pp. 4388–4394, Aug. 2010.

[21] X. Zhang, Z. Qin, and Y. Gao, "Dynamic adjustment of sparsity upper bound in wideband compressive spectrum sensing," in *Proc. IEEE Global Conf. Signal Inf. Process.*, Atlanta, GA, USA, Dec. 2014, pp. 1214–1218.

[22] H. Qi, X. Zhang, and Y. Gao, "Mixture model-based channel energy statistics modeling in compressive spectrum sensing," in *Proc. IEEE/CIC Int. Conf. Commun. China*, Beijing, China, Aug. 2018.

[23] H. Qi and Y. Gao, "Two-dimensional compressive spectrum sensing in collaborative cognitive radio networks," in *Proc. IEEE Globe Conf. Commun.*, Singapore, Dec. 2017, pp. 1–6.

[24] G. J. McLachlan and K. E. Basford, "Mixture models. Inference and applications to clustering," in *Statistics: Textbooks and Monographs*. New York, NY, USA: Dekker, 1988.

[25] M. Z. Zheleva *et al.*, "Enabling a nationwide radio frequency inventory using the spectrum observatory," *IEEE Trans. Mobile Comput.*, vol. 17, no. 2, pp. 362–375, Feb. 2018.

[26] M. Zheleva, R. Chandra, A. Chowdhery, A. Kapoor, and P. Garnett, "TxMiner: Identifying transmitters in real-world spectrum measurements," in *Proc. IEEE Int. Symp. Dyn. Spectr. Access Netw.*, Sep./Oct. 2015, pp. 94–105.

[27] Z. Han, H. Li, and W. Yin, "Sparse optimization algorithms," in *Compressive Sensing for Wireless Networks*. Cambridge, U.K.: Cambridge Univ. Press, 2013, ch. 4, pp. 69–117.

[28] D. P. Wipf and B. D. Rao, "Sparse Bayesian learning for basis selection," *IEEE Trans. Signal Process.*, vol. 52, no. 8, pp. 2153–2164, Aug. 2004.

[29] M. E. Tipping, "Sparse Bayesian learning and the relevance vector machine," *J. Mach. Learn. Res.*, vol. 1, pp. 211–244, Sep. 2001.

[30] Y. Wang, A. Pandharipande, Y. L. Polo, and G. Leus, "Distributed compressive wide-band spectrum sensing," in *Proc. Inf. Theory Appl. Workshop*, Feb. 2009, pp. 178–183.

[31] T. Hastie, R. Tibshirani, and J. Friedman, "Model inference and averaging," in *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. New York, NY, USA: Springer, 2013, ch. 8, pp. 236–242.

[32] N. Johnson, S. Kotz, and N. Balakrishnan, "Noncentral chi-square distributions," in *Continuous Univariate Distributions*, vol. 2. New York, NY, USA: Wiley, 1994.

[33] C. F. J. Wu, "On the convergence properties of the EM algorithm," *Ann. Statist.*, vol. 11, no. 1, pp. 95–103, 1983.

[34] S. Lloyd, "Least squares quantization in PCM," *IEEE Trans. Inf. Theory*, vol. IT-28, no. 2, pp. 129–137, Mar. 1982.

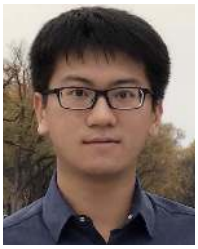
[35] D. MacKay, "An example inference task: Clustering," in *Information Theory, Inference and Learning Algorithms*. Cambridge, U.K.: Cambridge Univ. Press, 2003, ch. 20.

[36] S. M. Kay, "Maximum likelihood estimation," in *Fundamentals of Statistical Signal Processing: Estimation Theory*. Upper Saddle River, NJ, USA: Prentice-Hall, 1993, ch. 7.

[37] J. M. Bernardo, "Algorithm AS 103: Psi (Digamma) function," *J. Roy. Stat. Soc. Ser. C, Appl. Statist.*, vol. 25, no. 3, pp. 315–317, 1976.

[38] K. P. Burnham and D. Anderson, "Information and likelihood theory: A basis for model selection and inference," in *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. New York, NY, USA: Springer, 2007, ch. 2.

Haoran Qi (S'16) received the B.Eng. degree from Beihang University, Beijing, China, and the M.Eng. degree (Hons.) from the University of Nottingham, Nottingham, U.K., in 2016. He is currently pursuing the Ph.D. degree with the Department of Electronic Engineering and Computer Science, Queen Mary University of London, London, U.K. His current research interests include spectrum sensing, cooperative cognitive radio networks, and sub-Nyquist signal processing.



Xingjian Zhang (S'16) received the B.Sc. degree (Hons.) in telecommunications engineering from the Beijing University of Posts and Telecommunications, Beijing, China, in 2014, and the Ph.D. degree in electronic engineering from the Queen Mary University of London, London, U.K., in 2018. His research interests include cooperative wireless sensor networks, compressive sensing, real-time spectrum monitoring and analysis, and Internet of Things applications.

Yue Gao (S'03–M'07–SM'13) received the Ph.D. degree from the Queen Mary University of London (QMUL), U.K., in 2007. He was a Research Assistant, a Lecturer (Assistant Professor), and a Senior Lecturer (Associate Professor) with QMUL. He is currently a Reader in antennas and signal processing, and the Director of the Whitespace Machine Communication Lab, School of Electronic Engineering and Computer Science, QMUL. He is also leading a team developing theoretical research into practice in the interdisciplinary area among smart antennas, signal processing, spectrum sharing, and Internet of Things applications. He has published over 130 peer-reviewed journal and conference papers, two patents, and two book chapters. He is an Engineering and Physical Sciences Research Council Fellow from 2018 to 2023. He was a co-recipient of the EU Horizon Prize Award on Collaborative Spectrum Sharing in 2016, and the Research Performance Award from the Faculty of Science and Engineering at QMUL in 2017.

He served as the Signal Processing for Communications Symposium Co-Chair for IEEE ICCC 2016, the Publicity Co-Chair for the IEEE GLOBECOM 2016, the Cognitive Radio Symposium Co-Chair for the IEEE GLOBECOM 2017, and the General Chair of the the IEEE WoWMoM and iWEM 2017. He is a Secretary of the IEEE Technical Committee on Cognitive Networks and the IEEE Distinguished Lecturer of the Vehicular Technology Society. He is an Editor for the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, the IEEE WIRELESS COMMUNICATIONS LETTERS, and *China Communications*.