An EM Approach for Cooperative Spectrum Sensing in Multiantenna CR Networks

Ayman Assra, *Member, IEEE*, Jiaxin Yang, *Student Member, IEEE*, and Benoit Champagne, *Senior Member, IEEE*

***Abstract*—In this paper, a cooperative wideband spectrum sensing scheme based on the expectation–maximization (EM) algorithm is proposed for the detection of a primary user (PU) system in multiantenna cognitive radio (CR) networks. Given noisy signal observations from N secondary users (SUs) over multiple subbands at a fusion center (FC), prior works on cooperative spectrum sensing often use the set of received subband energy as decision statistics over the sensing interval. However, to achieve satisfactory performance, knowledge of the channel state information (CSI) and the noise variances at all the SUs is required by these algorithms. To overcome this limitation, our proposed method, which is referred to as joint detection and estimation (JDE), adopts the EM algorithm to jointly detect the PU signal and estimate the unknown channel frequency responses and noise variances over multiple subbands in an iterative manner. Various aspects of this proposed EM-JDE scheme are investigated, including a reliable initialization strategy to ensure convergence under practical conditions and a distributed implementation to reduce communication overhead. Under the assumption of perfect estimation for the channel frequency responses and noise variances, we further show that the proposed EM-JDE converges to the maximum-likelihood (ML) solution, which serves as an upper bound on its performance. Monte Carlo simulations over Rayleigh fading channels show that the proposed scheme significantly improves the performance of spectrum detection by exploiting the diversity of the spatially distributed SUs with multiple antennas.**

***Index Terms*—Channel state information, cognitive radio, expectation-maximization, joint detection and estimation, multi-antenna, spectrum sensing.**

I. Introduction

In recent years, cognitive radio (CR) has emerged as a key technology paradigm to alleviate the frequency spectrum scarcity [1]. The basic idea behind CR is that unlicensed or secondary users (SUs) share the frequency spectrum opportunistically with licensed or primary users (PUs) without causing harmful interference. This can be achieved by enabling the SUs to monitor the presence of PUs over a particular band of frequencies. In the literature, several spectrum sensing techniques have been proposed, which include energy detection (ED), matched filter detection, and cyclostationary feature detection [2], [3]. The appropriate spectrum sensing technique is chosen based on *a priori* knowledge about the PU's signal and the receiver complexity. For example, the matched filter is the optimal detection technique when the SU has complete information about the PU signal, which is rarely the case in practice [4]. The cyclostationary feature detector exploits the periodicity of the modulated signal to distinguish it from the stationary noise; however, it suffers from high computational complexity [5], [6]. The ED is the optimal detection scheme if the PU's signal is unknown; it also offers the advantage of lower complexity [7].

In many CR applications, detailed *a priori* knowledge about the PU's signal structure and modulation is not readily available, and in this context, ED is the most common choice for spectrum sensing. However, ED suffers from a poor performance in wireless environments characterized by low signal-to-noise ratio (SNR), multipath fading, or shadowing [8]. Multiantenna techniques have been employed along with ED to combat the fading effects by exploiting the spatial diversity of the observations at the SU terminal in [9]; they also help to reduce the sensing time compared with single-antenna ED. Another drawback of ED is its inherent susceptibility to uncertainties about the noise variance at the SU side. In [10], it was noticed that there is a minimum value of SNR, which is referred to as the *SNR wall*, and below which the spectrum detection fails even with infinite sensing intervals. In [11], necessary conditions for the existence of an SNR wall in ED techniques coupled with noise power estimation were introduced. In [12], the performance of multiantenna-based cooperative spectrum sensing was investigated under Rayleigh fading channels when an improved form of ED is employed, where the decision statistic is an arbitrary positive power of the amplitudes of the PU's signal samples.

Most works presented for the spectrum sensing problem assume the perfect knowledge of the channel conditions and noise variance by the SU, and few researchers have investigated the effect of estimation errors in these parameters on the PU detection process or possible estimation techniques that can be used jointly [13]. In [14] and [15], spectrum detectors based on the generalized likelihood ratio test (LRT) were proposed for multiantenna CR, and their performance was examined

Manuscript received February 6, 2014; revised September 29, 2014 and January 2, 2015; accepted January 9, 2015. Date of publication February 27, 2015; date of current version March 10, 2016. This work was supported by a research grant from the Natural Sciences and Engineering Research Council of Canada. The review of this paper was coordinated by Prof. X. Wang.

A. Assra is with TechTalent Consulting and TeamBuilding Inc., Calabasass, CA 91302 USA (e-mail: ayman.assra@mail.mcgill.ca).

J. Yang and B. Champagne are with the Department of Electrical and Computer Engineering, McGill University, Montreal, QC H3A 0E9, Canada (e-mail: jiaxin.yang@mail.mcgill.ca; benoit.champagne@mcgill.ca).

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/TVT.2015.2408369

under flat-fading channel conditions, assuming unknown channel gains and noise variance. An eigenvalue-based signal detection scheme was developed in [16] under noise variance or signal correlation uncertainty. In [17], spectrum sensing techniques for a finite-rank PU signal with unknown spatial covariance matrix were studied. In [18], a multiantenna spectrum sensing technique based on discrete Fourier transform (DFT) analysis of the received signals over flat-fading channels was presented, which does not require the knowledge of the noise variances at the different receive antennas.

In the aforementioned spectrum sensing schemes, the test statistics, e.g., energy measures, are derived as a function of the *received* signals at different SU's antennas, making the decision process on spectral occupancy vulnerable to common channel impairments, such as time variations and multipath fading, particularly for mobile networks. In this paper, to overcome this limitation, we propose a spectrum sensing scheme where a binary hypothesis test is applied on estimates of the average power *transmitted* by the PU over the frequency spectrum of interest during the sensing interval. Consequently, it is possible to make decision on the spectral occupancies of the current state of the channel, which in turn improves the sensing performance. We first develop our proposed approach for a multiantenna CR network operating over a wide frequency bandwidth, where a fusion center (FC) collects observations from N spatially distributed SUs over multiple subbands through different reporting channels. The estimation of the PU's average transmitted power requires the knowledge of the channel gains and noise variances at each SU over different frequency subbands; therefore, reliable channel and noise variance estimation becomes crucial for the proposed spectrum sensing scheme. As opposed to a more conventional approach, in which the channel state information (CSI) and noise variances are estimated first followed by spectrum detection, we *jointly* detect the PU's signal and estimate the unknown channel and noise parameters in each subband, thereby forming a joint detection and estimation (JDE) scheme. In the literature, iterative JDE techniques have been applied in wireless communication systems, particularly for multiuser detection, because of their ability to achieve accurate estimation without wasting the system resources [19]. In particular, the expectation–maximization (EM) algorithm has been proposed in iterative receivers due to attractive features such as iteratively attaining the maximum-likelihood (ML) solution with reduced complexity [20].

The application of the EM algorithm in spectrum sensing has been considered in our earlier works [21], [22], which focus on a specific noncooperative scenario, i.e., a single SU with time-invariant channels. In this paper, we propose a new and more general EM-JDE scheme for *cooperative* spectrum sensing in multiuser multiantenna CR networks, where multiple spatially distributed SUs cooperate to detect the state of occupancy of a wideband frequency spectrum. In addition to the extended modeling of spatial dimension, which involves more elaborate mathematics in the derivations, the new contributions in this paper include the following essential aspects.

1) *Cooperation Mechanisms*: Different cooperation mechanisms among SUs are considered, i.e., centralized versus distributed implementations, as a means to tradeoff communication overhead for local processing complexity at the SUs.
2) *Initialization for EM Algorithm*: A practical and reliable initialization strategy for the parameters to be estimated is proposed, which provides a reasonably "good" starting point for fast convergence of the iterative algorithm.
3) *Performance Evaluation*: The asymptotic ML-based spectrum sensing for the multiuser scenario is investigated, to provide an upper bound on the performance of the proposed algorithm. New closed-form expressions of the corresponding probability of false alarm and missed detection are also derived.
4) *Numerical Experiments*: New simulation results for the multiuser CR network are reported. In addition to the convergence behavior of the iterative algorithm with the proposed initialization, the effects of time-varying fading channels, the number of cooperating SUs, and the distributed implementation are thoroughly studied.

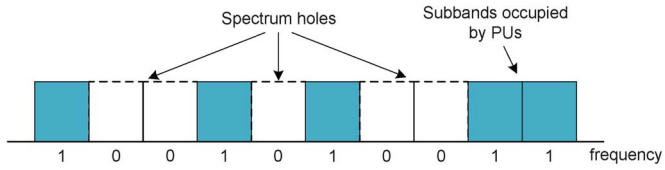

Fig. 1. Occupancy of a wideband spectrum.

Our theoretical findings and simulation results demonstrate the advantages of the proposed cooperative EM-JDE scheme in improving the performance of spectrum detection by exploiting the diversity offered by spatially distributed SUs with multiple antennas.

The remainder of this paper is organized as follows. The system model and problem formulation are presented in Section II. In Section III, we derive the EM-JDE scheme for cooperative spectrum sensing in multiantenna CR networks. An ML-based upper bound on the performance of the proposed scheme is developed in Section IV. In Section V, the simulation results and discussions are presented. Finally, conclusions are drawn in Section VI.

II. System Model and Problem Formulation

In our work, we assume a CR network comprised of N SUs, which are indexed by $n \in \{0, \ldots, N-1\}$, where each SU terminal is equipped with P receiving antennas. We consider a wideband frequency spectrum, which is divided into K subbands, as shown in Fig. 1. Here, the concept of a subband is identical to that used in a multicarrier modulation system, where each subband represents a narrow band of frequency centered on a single subcarrier, upon which the corresponding subband information is modulated.

The time-domain measurements, after sampling using Nyquist rate at the pth antenna of the nth SU, can be represented as[1]

$$r_{n,p}(l) = \sum_{q=0}^{L-1} h_{n,p}(q)s(l-q) + v_{n,p}(l) \tag{1}$$

where $p \in \{0, \ldots, P-1\}$; $s(l)$ is the lth sample of the time-domain signal transmitted by the PU; $h_{n,p}(q)$ for $q \in \{0, \ldots, L-1\}$ denotes the qth tap of the channel impulse response between the PU and the pth antenna of the nth SU, which is assumed of finite length L; and $v_{n,p}(l)$ is an additive noise process. In this paper, we employ a frequency-domain detector, where a K-point DFT operation is applied on successive frames of $r_{n,p}(l)$ to obtain the following narrowband discrete frequency components:

$$R_{k,n,p}(m) = \sum_{l=0}^{K-1} w(l) r_{n,p}(mK+l) e^{-j2\pi lk/K} \tag{2}$$

where frequency index $k \in \{0, \ldots, K-1\}$, frame index $m \in \{1, \ldots, M-1\}$ with M being the number of frames available for detection, and $w(l)$ is a normalized frequency analysis window [25]. Under the assumption that the frequency subbands are sufficiently narrow, in comparison to the interval of variations in the channel frequency responses (or equivalently, $K \gg L$), the linear convolution in (1) can be approximated as the circular convolution, such that application of DFT in the time domain is equivalent to the pointwise multiplication of the corresponding K points in the discrete frequency domain [26], i.e.,

$$R_{k,n,p}(m) = H_{k,n,p} S_k(m) + V_{k,n,p}(m) \tag{3}$$

where $S_k(m)$ represents the DFT coefficient of the PU signal $s(t)$ over the mth time frame in the kth subband, $H_{k,n,p}$ is the DFT coefficient of the channel $h_{n,p}(q)$ between the PU and the pth receive antenna of the nth SU in the kth subband, and $V_{k,n,p}(m)$ is the DFT coefficient of the noise $v_{n,p}(l)$ at the pth receive antenna of the nth user, as obtained over the mth frame in the kth subband.

The linear model (3), with multiplicative channel effect on the PU signal in each subband, is common in the wideband spectrum sensing literature (see, e.g., in [27]–[29]). Nevertheless, in practice, the DFT operation will suffer from spectral leakage, which may cause interference between neighboring subbands. Traditionally, a properly chosen window function $w(l)$ can be applied in (2) to allow a design tradeoff between frequency resolution and leakage [25]. Specific solutions to the suppression of spectral leakage in the context of spectrum sensing have been studied in [30]–[32]. In this paper, we assume that such a suppression technique has been employed to eliminate or, at least, reduce the spectral leakage among successive frequency bands to a level that is comparable to that of the additive background noise. Considering the high levels of radio noise and interference often encountered in CR applications, this does not represent a very stringent requirement. Thus, the effect of the spectral leakage can be minimized or neglected, which facilitates the derivation and analysis of the EM algorithm.[2]

In this paper, we make use of the statistical model described in the following for the characterization of the received signal samples $\{R_{k,n,p}(m)\}$, which is widely adopted in the literature (see again [27]–[29] and the references therein). To begin with, we define the vectors $\mathbf{S} = [\mathbf{S}_0^T, \ldots, \mathbf{S}_{K-1}^T]^T$ and $\mathbf{S}_k = [S_k(0), \ldots, S_k(M-1)]^T$, where superscript ^{T} denotes the transpose operation. Since we have no prior knowledge about the PU signal, $\mathbf{S}_k$ is assumed to follow a complex circular Gaussian distribution with zero mean and covariance matrix $B_k \mathbf{I}_M$, which is denoted $\mathcal{CN}(\mathbf{0}_M, B_k \mathbf{I}_M)$, where $\mathbf{0}_M$ is an $M \times 1$ zero vector, $\mathbf{I}_M$ is an identity matrix of order M, and B_k is an occupancy parameter as explained in the following. The complex circular Gaussian assumption of the subband PU signal is widely employed in the spectrum sensing literature as it can be naturally justified in many applications. For instance, when the PU system employs a broadband form of modulation, such as multicarrier modulation [33], [34] or spread spectrum [35], the received signal in each subband in (2) is the sum of K nearly independent contributions. In this case, one can invoke the central limit theorem [36] to motivate the Gaussian assumption since, in practice, the number of subbands K used for spectrum sensing can be fairly large, e.g., $K = 2^l$, where $l \geq 6$. An exception to this would be when the PU system uses OFDM modulation with the same frequency plan as the one used by the wideband SU detector and with perfect synchronization in time and frequency, but in practice, this is unlikely to be the case. Notwithstanding the above, the Gaussian model corresponds to a worst-case assumption, according to the principle of maximum entropy [37].

We model the subband occupancy B_k as a binary random variable, which indicates the status of the PU activity in the kth subband: $B_k = 0$ when the kth subband is vacant, whereas $B_k = 1$ when the PU signal is present.[3] We assume independent subband occupancy, i.e., the joint probability mass

[1]Recently, allowing sub-Nyquist sampling rates while still maintaining a certain level of detection quality has received increasing attention in the physical implementation of wideband spectrum sensing where a high sampling rate is needed [23], [24]. In effect, the proposed EM-JDE scheme in this paper could be combined with an existing compressive sensing technique in order to operate at a sub-Nyquist sampling rate. Specifically, under the assumption of sparse spectrum, the first step is to compress and collect a set of time-domain samples with reduced dimensionality by means of a linear transformation. This can be expressed as $\mathbf{r}'_{n,p} = \mathbf{S}_n \mathbf{r}_{n,p}$, where $\mathbf{r}_{n,p}$ is the K-dimensional vector of observation from (1), $\mathbf{S}_n$ is an $K' \times K$ compressive sensing matrix, and $\mathbf{r}'_{n,p}$ is the compressed vector of observations with dimension $K' < K$. From this point on, the proposed EM-JDE can be applied with appropriate modifications, which are needed to take into account the effects of the known transformation matrix $\mathbf{S}_n$ on the unknown channel coefficients and noise variances in (1). However, the suitable choice of the compression matrices $\mathbf{S}_n$ and the dimensionality parameter K' in the CR network applications remains an open issue, which, we believe, falls outside the scope of this paper.

[2]At moderate SNR, and in the absence of correlation between adjacent subband occupancy by the PU, the consideration of leakage is conceptually equivalent to a slight increase in the additive background noise variance. We have been able to confirm this point for a standard DFT-based analysis (rectangular window) by independent simulations not reported in this paper.

[3]Without loss of generality, to simplify the presentation, we assume that, when the PU is present, the signal power in the kth subband is normalized to unity, i.e., $E[|S_k|^2 | B_k = 1] = 1$.

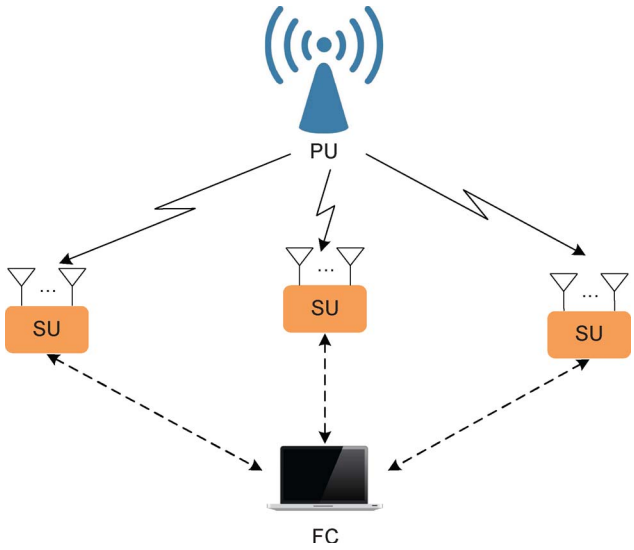

Fig. 2. Cooperative spectrum sensing in a CR network.

function (pmf) of $\mathbf{B} = [B_0, \ldots, B_{K-1}]^T$ is given by $P(\mathbf{B}) = \prod_{k=0}^{K-1} P(B_k)$, where $P(B_k)$ denotes the marginal pmf of B_k. Moreover, given $\mathbf{B}$, signal samples from different subbands are independent, i.e., the conditional probability density function (pdf) $f(\mathbf{S}|\mathbf{B}) = \prod_{k=0}^{K-1} f(\mathbf{S}_k|B_k)$, where $f(\mathbf{S}_k|B_k)$ denotes the conditional pdf of $\mathbf{S}_k$ given B_k.

The channel coefficients $H_{k,n,p}$ in (3) are assumed to remain constant during the sensing interval and hence are modeled as deterministic but unknown quantities. For later reference, we define the channel coefficient vectors $\mathbf{H} = [\mathbf{H}_0^T, \ldots, \mathbf{H}_{K-1}^T]^T$, $\mathbf{H}_k = [\mathbf{H}_{k,0}^T, \ldots, \mathbf{H}_{k,N-1}^T]^T$, and $\mathbf{H}_{k,n} = [H_{k,n,0}, \ldots, H_{k,n,P-1}]^T$. The additive noise samples $V_{k,n,p}(m)$ for $m \in \{0, \ldots, M-1\}$ in (3) are represented by the vector $\mathbf{V}_{k,n,p} = [V_{k,n,p}(0), \ldots, V_{k,n,p}(M-1)]^T$, which is modeled as $\mathcal{CN}(\mathbf{0}_M, \varsigma_{k,n}\mathbf{I}_M)$, where the noise variance $\varsigma_{k,n} \triangleq E[|V_{k,n,p}(m)|^2]$ has the same value over the different receive antennas of the nth SU. For future reference, we also introduce the vectors $\boldsymbol{\varsigma} = [\boldsymbol{\varsigma}_0^T, \ldots, \boldsymbol{\varsigma}_{K-1}^T]^T$ and $\boldsymbol{\varsigma}_k = [\varsigma_{k,0}, \ldots, \varsigma_{k,N-1}]^T$, which are also treated as deterministic but unknown quantities. The noise vector $\mathbf{V}_{k,n,p}$ has independent distribution across the frequency, user, and antenna indexes and is assumed independent of the signal generation mechanism at the PU, as represented by random vectors $\mathbf{S}$ and $\mathbf{B}$.

In our work, as shown in Fig. 2, we assume a centralized spectrum sensing scheme, where local observations from the N SUs are reported assuming error free transmission to the FC, which can be a cognitive base station or a selected CR node. Then, a final decision is taken at the FC based on these observations to determine the presence or absence of the PU signal. On this basis, the problem at hand can be formulated as follows: Our goal is to detect the value of the occupancy random variable in each subband, i.e., $B_k \in \{0, 1\}$, based on the set of observations $\{R_{k,n,p}(m) : \forall k, n, p \text{ and } m\}$ collected from the N SUs and considering that the information about the channel vector $\mathbf{H}_k$ and the noise variance vector $\boldsymbol{\varsigma}_k$ in each subband is missing. Therefore, the unknown parameter vector of this wideband spectrum sensing problem can be represented by the concatenated vector $\mathbf{U} = [\mathbf{B}^T, \mathbf{H}^T, \boldsymbol{\varsigma}^T]^T$. Considering the given statistical model of the received signal samples $R_{k,n,p}(m)$ in (3), ML estimation of this parameter vector results in a multidimensional joint optimization problem, whose solution can be achieved in theory by applying an exhaustive search. However, due to the very large number of unknown parameters subject to estimation, the ML solution is not feasible in practice.

III. Expectation–Maximization-Based Joint Detection and Estimation

In the literature, the EM algorithm is proposed to achieve the ML solution iteratively with low computational complexity [38]. Therefore, here, we use the EM algorithm to estimate the unknown parameter vector of each subband while performing the spectrum detection process concurrently. This technique, which is referred to as iterative JDE, has been used to solve many problems of practical interest in the wireless communications, but to the best of our knowledge, its application to spectrum sensing has not been extensively researched. In the following, we first apply the EM formalism to derive the proposed JDE algorithm for cooperative spectrum sensing in multiantenna CR networks. This is followed by a discussion of related implementation aspects, including computational complexity, distributed implementation in cooperative frameworks, and the proposed initialization scheme.

A. Algorithm Derivation

While the EM algorithm is often developed and studied for the case of continuous parameters, here, we are faced with a mixed situation in which the unknown channel and noise parameters $\mathbf{H}$ and $\boldsymbol{\varsigma}$ are continuous, whereas the spectral occupancy vector $\mathbf{B}$ takes on discrete (binary) values. Therefore, to simplify the application of the EM formalism and the convergence analysis of the resulting algorithm, we initially adopt a purely continuous approach in which the occupancy variables $\{B_k\}$ are first treated as continuous within the interval of [0, 1]. In this way, the intermediate estimates of each B_k obtained through the sequence of EM iterations, which are denoted $\hat{B}_k^{(i)}$, where the iteration index $i \in \mathbb{N}$, may be viewed as soft estimates of the occupancy in subband k. This makes it possible to find closed-form expressions for the maximum of the expected conditional likelihood during the maximization step of the EM procedure. Once the sequence of soft estimates $\hat{B}_k^{(i)}$ has been judged to converge to an adequate level (as will be explained in the following), a hard estimate of B_k is finally obtained by applying a binary test, i.e., comparing $\hat{B}_k^{(\infty)}$ to a properly selected threshold. The specific details of our derivation follows.

According to the EM terminology, the *incomplete* data $\mathbf{R} = [\mathbf{R}_0^T, \ldots, \mathbf{R}_{K-1}^T]^T$ consists of the observations from all the receive antennas of the N SUs over the K subbands, where we define $\mathbf{R}_k = [\mathbf{R}_k(0)^T, \ldots, \mathbf{R}_k(M-1)^T]^T$, $\mathbf{R}_k(m) = [\mathbf{R}_{k,0}(m)^T, \ldots, \mathbf{R}_{k,N-1}(m)^T]^T$, and $\mathbf{R}_{k,n}(m) = [R_{k,n,0}(m), \ldots, R_{k,n,P-1}(m)]^T$. The so-called *complete* data $\mathbf{Y}$ is defined as a combination of K independent pairs of $\mathbf{R}_k$ and $\mathbf{S}_k$, i.e., $\mathbf{Y} = [\mathbf{Y}_0^T, \ldots, \mathbf{Y}_{K-1}^T]^T$, where $\mathbf{Y}_k = [\mathbf{R}_k^T, \mathbf{S}_k^T]^T$.

Conditioned on $\mathbf{U} = \left[\mathbf{B}^T, \mathbf{H}^T, \varsigma^T\right]^T$, the K component vectors of $\mathbf{Y}$ are statistically independent, and consequently

$$f(\mathbf{Y}|\mathbf{U}) = \prod_{k=1}^{K} f(\mathbf{R}_k, \mathbf{S}_k|\mathbf{U}_k) \tag{4}$$

where

$$f(\mathbf{R}_k, \mathbf{S}_k|\mathbf{U}_k) = f(\mathbf{R}_k|\mathbf{S}_k, \mathbf{U}_k) f(\mathbf{S}_k|\mathbf{U}_k). \tag{5}$$

Invoking the Gaussian assumptions on the signal and noise made in Section II, it follows that the complete data log-likelihood function $L(\mathbf{Y}|\mathbf{U})$ is given by

$$L(\mathbf{Y}|\mathbf{U}) = \sum_{k=0}^{K-1} L(\mathbf{Y}_k|\mathbf{U}_k) \tag{6}$$

where

$$\begin{aligned} &L(\mathbf{Y}_k|\mathbf{U}_k) \\ &\quad = L(\mathbf{R}_k|\mathbf{S}_k, \mathbf{U}_k) + L(\mathbf{S}_k|\mathbf{U}_k) \\ &\quad = -MP\sum_{n=0}^{N-1} \ln(\pi\varsigma_{k,n}) - M\ln(\pi B_k) - \frac{1}{B_k}\sum_{m=0}^{M-1} |S_k(m)|^2 \\ &\qquad - \sum_{n=0}^{N-1} \frac{1}{\varsigma_{k,n}} \sum_{p=0}^{P-1}\sum_{m=0}^{M-1} |R_{k,n,p}(m) - H_{k,n,p}S_k(m)|^2 \end{aligned} \tag{7}$$

and the unknown parameter vector of the kth subband $\mathbf{U}_k = [B_k, \mathbf{H}_k^T, \varsigma_k^T]^T$.

In the expectation step (E-step) of the EM algorithm, we compute the conditional expectation of (6) given $\mathbf{R}$ and the estimation of $\mathbf{U}$ at the ith iteration, i.e., $\mathbf{U} = \hat{\mathbf{U}}^{(i)}$, where $\hat{\mathbf{U}}^{(i)} = [(\hat{\mathbf{B}}^{(i)})^T, (\hat{\mathbf{H}}^{(i)})^T, (\hat{\varsigma}^{(i)})^T]^T$, and $\hat{\mathbf{B}}^{(i)}$, $\hat{\mathbf{H}}^{(i)}$, and $\hat{\varsigma}^{(i)}$ are the EM estimates of $\mathbf{B}$, $\mathbf{H}$, and ς at the ith iteration, respectively. The result of this conditional expectation is a function of the unknown parameter vector $\mathbf{U}$, which we denote by $\Delta(\mathbf{U}|\hat{\mathbf{U}}^{(i)})$. Hence, we have

$$\Delta(\mathbf{U}|\hat{\mathbf{U}}^{(i)}) = \sum_{k=0}^{K-1} \Delta\left(\mathbf{U}_k|\hat{\mathbf{U}}^{(i)}\right) \tag{8}$$

where $\Delta(\mathbf{U}_k|\hat{\mathbf{U}}^{(i)})$ is expressed as (9), shown at the bottom of the page, and $E[\cdot]$ denotes the expectation operator.

In the maximization step (M-step) of the algorithm, the updated parameter estimate $\hat{\mathbf{U}}^{(i+1)}$ is obtained by maximizing $\Delta(\mathbf{U}|\hat{\mathbf{U}}^{(i)})$ in (8) with respect to $\mathbf{U}$. Here, since the samples of each process are statistically independent across the frequency index, $\mathbf{U}_k$ can be estimated individually by maximizing its corresponding term, i.e., $\Delta(\mathbf{U}_k|\hat{\mathbf{U}}^{(i)})$ in (9). Notice that the occupancy parameter B_k in the conditional expectation in (9) is actually decoupled from the channel and variance parameters $H_{k,n,p}$ and $\varsigma_{k,n}$; therefore, we can first estimate B_k, followed by the estimation of $\mathbf{H}_k$ and ς_k.

To begin with, $\hat{B}_k^{(i)}$ can be updated by maximizing (9) with respect to B_k. By neglecting the terms in (9) that are independent of B_k, we obtain

$$\hat{B}_k^{(i+1)} = \arg\max_{B_k\in[0,1]} g\left(B_k, \sum_{m=0}^{M-1} E\left[|S_k(m)|^2\|\mathbf{R}, \hat{\mathbf{U}}^{(i)}\right]\right) \tag{10}$$

where

$$g(B_k, \phi) \triangleq -M\ln(\pi B_k) - \frac{1}{B_k}\phi. \tag{11}$$

As explained earlier, the parameter space for B_k is discrete, and the convergence of the EM algorithm in this case is not always guaranteed. To overcome this problem, we artificially extend the search range of the M-step maximization from the discrete space $\{0, 1\}$ to the continuous space [0, 1]; therefore, the solution to the M-step update (10) can be given by

$$\hat{B}_k^{(i+1)} = \min\left\{\frac{1}{M}\sum_{m=0}^{M-1} E\left[|S_k(m)|^2|\mathbf{R}, \hat{\mathbf{U}}^{(i)}\right], 1\right\}. \tag{12}$$

Under this continuous parameter space assumption, the convergence of the resulting EM estimator to a stationary point follows from well-known results on the analysis of the conventional EM algorithm [39], [40]. To compute $\hat{B}_k^{(i+1)}$ in (12), we note that $E[|S_k(m)|^2|\mathbf{R}, \hat{\mathbf{U}}^{(i)}] = |E[S_k(m)|\mathbf{R}, \hat{\mathbf{U}}^{(i)}]|^2 + \mathrm{Var}[S_k(m)|\mathbf{R}, \hat{\mathbf{U}}^{(i)}]$, and $\mathrm{Var}[\cdot]$ is the variance operator. Since $\mathbf{R}$ and $\mathbf{S}$ are jointly Gaussian, the conditional mean and variance of $S_k(m)$ given $\mathbf{R}$ and $\hat{\mathbf{U}}^{(i)}$ can be expressed in the following forms [41]:

$$E\left[S_k(m)|\mathbf{R}, \hat{\mathbf{U}}^{(i)}\right] = \hat{B}_k^{(i)}\hat{\mathbf{H}}_k^{(i)H}\hat{\mathbf{\Gamma}}_k^{(i)^{-1}}\mathbf{R}_k(m) \tag{13}$$

$$\mathrm{Var}\left[S_k(m)|\mathbf{R}, \hat{\mathbf{U}}^{(i)}\right] = \hat{B}_k^{(i)} - \hat{B}_k^{(i)}\hat{\mathbf{H}}_k^{(i)H}\hat{\mathbf{\Gamma}}_k^{(i)^{-1}}\hat{\mathbf{H}}_k^{(i)}\hat{B}_k^{(i)} \tag{14}$$

where the superscript ^{H} denotes the conjugate transpose, and $\hat{\mathbf{\Gamma}}_k^{(i)} = \hat{B}_k^{(i)}\hat{\mathbf{H}}_k^{(i)}\hat{\mathbf{H}}_k^{(i)H} + \hat{\mathbf{\Sigma}}_k^{(i)}$. The matrix $\hat{\mathbf{\Sigma}}_k^{(i)}$ that represents

$$\begin{aligned} \Delta(\mathbf{U}_k|\hat{\mathbf{U}}^{(i)}) = &-MP\sum_{n=0}^{N-1}\ln(\pi\varsigma_{k,n}) - M\ln(\pi B_k) - \frac{1}{B_k}\sum_{m=0}^{M-1} E\left[|S_k(m)|^2|\mathbf{R}, \hat{\mathbf{U}}^{(i)}\right] \\ &- \sum_{n=0}^{N-1}\frac{1}{\varsigma_{k,n}}\sum_{p=0}^{P-1}\sum_{m=0}^{M-1} \times E\left[|R_{k,n,p}(m) - H_{k,n,p}S_k(m)|^2|\mathbf{R}, \hat{\mathbf{U}}^{(i)}\right] \end{aligned} \tag{9}$$

the ith iterative estimate of the noise covariance matrix in the kth subband for the N SUs is given by

$$\hat{\mathbf{\Sigma}}_k^{(i)} = \begin{pmatrix} \hat{\mathbf{\Sigma}}_{k,0}^{(i)} & \cdots & \cdots & \mathbf{0}_P \\ \mathbf{0}_P & \hat{\mathbf{\Sigma}}_{k,1}^{(i)} & \cdots & \mathbf{0}_P \\ \vdots & \cdots & \cdots & \vdots \\ \mathbf{0}_P & \cdots & \cdots & \hat{\mathbf{\Sigma}}_{k,N-1}^{(i)} \end{pmatrix} \tag{15}$$

where $\hat{\mathbf{\Sigma}}_{k,n}^{(i)} = \hat{\varsigma}_{k,n}^{(i)} \mathbf{I}_P$. From (12)–(14), it follows that the EM algorithm provides an iterative estimate of the average transmission power of the PU system in each subband, i.e., an iterative energy detector. In turn, this can be used to provide binary-valued subband occupancy decisions through the application of a final hard limiter, as will be explained in the following.

Next, we obtain the estimate of $\mathbf{H}_k$ at the $(i+1)$th iteration as follows. Each element of $\hat{\mathbf{H}}_k^{(i+1)}$ is obtained by maximizing its corresponding summand in the right-hand side of (9), which yields

$$\hat{H}_{k,n,p}^{(i+1)} = \arg \min_{H_{k,n,p}} \sum_{m=0}^{M-1} \times E\left[|R_{k,n,p}(m) - H_{k,n,p} S_k(m)|^2 | \mathbf{R}, \hat{\mathbf{U}}^{(i)}\right] \tag{16}$$

and subsequently

$$\hat{H}_{k,n,p}^{(i+1)} = \frac{\sum_{m=0}^{M-1} R_{k,n,p}(m) E\left[S_k(m)^* | \mathbf{R}, \hat{\mathbf{U}}^{(i)}\right]}{\sum_{m=0}^{M-1} E\left[|S_k(m)|^2 | \mathbf{R}, \hat{\mathbf{U}}^{(i)}\right]} \tag{17}$$

where the superscript * denotes the complex conjugate. Introducing $\mathbf{R}_{k,n,p} = [R_{k,n,p}(0), \ldots, R_{k,n,p}(M-1)]^T$, $\hat{H}_{k,n,p}^{(i+1)}$ in (17) can be represented in a more compact form as

$$\hat{H}_{k,n,p}^{(i+1)} = E\left[\mathbf{S}_k^H \mathbf{S}_k | \mathbf{R}, \hat{\mathbf{U}}^{(i)}\right]^{-1} E\left[\mathbf{S}_k | \mathbf{R}, \hat{\mathbf{U}}^{(i)}\right]^H \mathbf{R}_{k,n,p} \tag{18}$$

which can be seen as an iterative least-squares channel estimation of $H_{k,n,p}$ at the $(i+1)$th iteration. During the sensing periods where the kth subband is vacant, the SUs only receive noise samples. Therefore, it is not necessary to estimate the channels, and the channel estimation step has to be dropped from the EM iterative loop. We use the following approach to make a decision on the EM channel estimation step. We compare the value of $\hat{B}_k^{(i)}$ after running the EM algorithm for a few iterations with a preset threshold $\delta_B > 0$. The value of δ_B is chosen to be less than the mid-value between the minimum and maximum values of B_k, i.e., in our case, $\delta_B < 0.5$. The selection tends to be conservative in the sense that it prevents stopping the channel estimation unnecessarily due to an erroneous missed detection. Consequently, the values of $\hat{B}_k^{(i)}$ after these few iterations can be classified within two regions: $\mathcal{R}_0$ with $0 \le \hat{B}_k^{(i)} < \delta_B$ and $\mathcal{R}_1$ with $\delta_B \le \hat{B}_k^{(i)} \le 1$. The channel estimation step is excluded from the EM iterative loop if $\hat{B}_k^{(i)}$ is located within $\mathcal{R}_0$. To realize this condition without hindering the update of $\hat{B}_k^{(i)}$ in the EM algorithm, $\hat{H}_{k,n,p}^{(i+1)}$ in (17) can be redefined as follows:

$$\hat{H}_{k,n,p}^{(i+1)} = \frac{\sum_{m=0}^{M-1} R_{k,n,p}(m) E\left[S_k(m)^* | \mathbf{R}, \hat{\mathbf{U}}^{(i)}\right]}{\vartheta + \sum_{m=0}^{M-1} E\left[|S_k(m)|^2 | \mathbf{R}, \hat{\mathbf{U}}^{(i)}\right]} \tag{19}$$

where

$$\vartheta = \begin{cases} \varrho, & \hat{B}_k^{(i)} \in \mathcal{R}_0 \\ 0, & \hat{B}_k^{(i)} \in \mathcal{R}_1 \end{cases} \tag{20}$$

and ϱ is a large positive number. Note that the classification in (20) is based on the soft estimate $\hat{B}_k^{(i)}$ at the ith iteration and may therefore involve error. To minimize the effects of this error, variations in the values of $\hat{B}_k^{(i)}$ over successive stages, i.e., groups of T iterations as explained in Section III-B4, can be exploited to improve the exactness of the decision in comparing $\hat{B}_k^{(i)}$ with threshold δ_B. This stems from the fact that, when the kth subband is vacant, the soft estimate of B_k tends to decline toward zero, but not monotonically, whereas in the case when the kth subbed is occupied by the PU system, $\hat{B}_k^{(i)}$ will tend to increase toward the true value of 1. Therefore, in our implementation, the decision in (20) is performed only at the end of a stage, such that the descent in the values of $\hat{B}_k^{(i)}$ can be more accurately detected. We remark that this approach is empirical in nature but works well in simulation experiments. However, a formal proof of its advantages is not currently available, which remains an interesting topic of future research.

Finally, we update $\hat{\varsigma}_k^{(i)}$ as follows. In (9), we substitute $\hat{\mathbf{H}}_{k,n}^{(i+1)}$ for $\mathbf{H}_{k,n}$ and maximize the objective in (9) with respect to $\varsigma_{k,n}$. That is

$$\hat{\varsigma}_{k,n}^{(i+1)} = \arg \min_{\varsigma_{k,n}} \left(MP \ln(\varsigma_{k,n}) + \frac{1}{\varsigma_{k,n}} \sum_{p=0}^{P-1} \sum_{m=0}^{M-1} \mathcal{V}_{k,n,p}(m) \right) \tag{21}$$

where

$$\begin{aligned} &\mathcal{V}_{k,n,p}(m) \\ &\quad = E\left[|R_{k,n,p}(m) - \hat{H}_{k,n,p}^{(i+1)} S_k(m)|^2 | \mathbf{R}, \hat{\mathbf{U}}^{(i)}\right] \\ &\quad = |R_{k,n,p}(m)|^2 - R_{k,n,p}(m)^* \hat{H}_{k,n,p}^{(i+1)} E\left[S_k(m) | \mathbf{R}, \hat{\mathbf{U}}^{(i)}\right] \\ &\qquad - R_{k,n,p}(m) \hat{H}_{k,n,p}^{(i+1)*} E\left[S_k(m)^* | \mathbf{R}, \hat{\mathbf{U}}^{(i)}\right] \\ &\qquad + \left|\hat{H}_{k,n,p}^{(i+1)}\right|^2 E\left[|S_k(m)|^2 | \mathbf{R}, \hat{\mathbf{U}}^{(i)}\right]. \end{aligned} \tag{22}$$

This yields

$$\hat{\varsigma}_{k,n}^{(i+1)} = \frac{1}{MP} \sum_{p=0}^{P-1} \sum_{m=0}^{M-1} \mathcal{V}_{k,n,p}(m). \tag{23}$$

Up to this point, as explained earlier, the proposed EM-JDE produces a soft estimate of B_k at each iteration. Finally, a hard (i.e., binary) estimation of B_k is performed once the sequence of EM iterative steps for the soft estimates is judged to have

reached an adequate level of convergence. In our case, the convergence condition is defined by comparing the difference between the estimates of $\hat{B}_k^{(i)}$ in two successive iterations with a small positive threshold ϵ. That is, the EM iterations stop if $|\hat{B}_k^{(i+1)} - \hat{B}_k^{(i)}| \leq \epsilon$, where $\epsilon \to 0^+$. Then, hard limiting with threshold γ_k is applied on the EM estimate of B_k after convergence, which we simply denote $\hat{B}_k^{(\infty)} = \lim_{i\to\infty} \hat{B}_k^{(i)}$. The corresponding test can be expressed as

$$\hat{B}_k^{(\infty)} \underset{\hat{B}_k^{\mathrm{EM}}=0}{\overset{\hat{B}_k^{\mathrm{EM}}=1}{\gtrless}} \gamma_k \tag{24}$$

where $\hat{B}_k^{\mathrm{EM}} \in \{0, 1\}$ is the final binary estimate of the kth subband occupancy. The choice of γ_k, as in other binary detection problems, gives rise to a practical tradeoff between the probability of false alarm $P_{f,k}(\gamma_k)$ and the probability of missed detection $P_{m,k}(\gamma_k)$. Specifically, as γ_k increases from 0 to 1, $P_{f,k}(\gamma_k)$ decreases from 1 to 0, whereas $P_{m,k}(\gamma_k)$ increases from 0 to 1. The various receive operating characteristic (ROC) curves shown in Section V are indeed obtained by varying γ_k in this way. In particular, if a desired level of $P_{f,k}(\gamma_k)$ is needed in a given application, a *rough* estimate of γ_k can be obtained from the asymptotic performance analysis given in Section IV-B, as per (47); this value can be then further refined through simulations or experimental measurements.

Remark 1 (Assumption on Subband Occupancy): In many applications, there will be correlation in the occupancy of adjacent subbands, as evidenced in recent experimental studies [42]. In fact, the modeling of the subband occupancy highly depends on the operation of the primary system. If the PU represents a multiuser multicarrier modulation-based network, e.g., OFDMA, where each user operates independently on a subset of dedicated subcarriers, the subband occupancy could be assumed independent, particularly if adaptive loading is employed. However, if the PU is a broadcast television or wireless local area network system, the PU signal usually occupies a range of contiguous subbands, and therefore, the subband occupancy becomes correlated. For the latter case, investigating cooperative wideband spectrum sensing with correlated subband occupancy represents an interesting, yet challenging extension of this work. To be specific, since the random variables $\{B_k\}$ are no longer independent, to solve the M step in (10)–(12), a joint optimization problem over multiple discrete variables needs to be considered (see, e.g., [27] and [43]), which is quite computationally intensive. Therefore, in this paper, we focus on the former case, which facilitates the derivation of a low-complexity algorithm. However, investigating cooperative wideband spectrum sensing with correlated subband occupancy, possibly by exploiting a simpler but suboptimal mathematical framework, remains an interesting research direction for our future work.

B. Implementation Aspects

1) Computational Complexity: Using the EM algorithm, the N_d-dimensional optimization problem for each subband k is decomposed into N_d independent 1-D optimization problems,

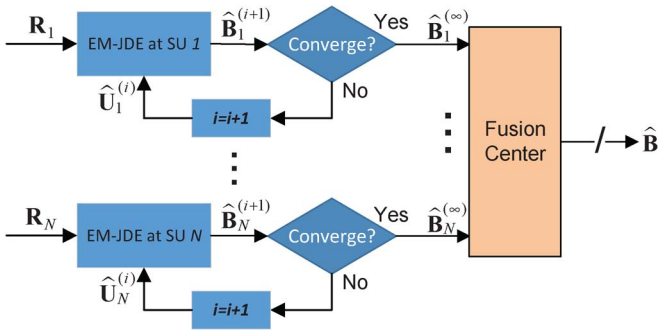

Fig. 3. Block diagram of the proposed EM-JDE algorithm with distributed implementation at each SU.

where $N_d = 1 + N(P + 1)$ is the number of unknown parameters of subband k, leading to a computationally feasible scheme. Furthermore, at each iteration of the EM algorithm, the solution of these 1-D optimization problems is obtained in closed form where the computation of the conditional moments in (13) and (14) dominates the computational complexity of the proposed EM-JDE scheme. By exploiting the specific structure of the matrix $\hat{\mathbf{\Gamma}}_k^{(i)}$ as a sum of a diagonal matrix plus a rank-1 modification term and invoking the Sherman–Morrison formula [44], it can be shown that this step requires $\mathcal{O}(NP)$ mathematical operations.

2) Distributed EM-JDE Implementation: The proposed EM-JDE scheme adopts centralized spectrum sensing, where the observations from multiple SUs need to be reported to the FC. The advantage of centralized processing is that the CR units benefit from a much reduced hardware complexity since the decision-making process is performed at the FC. However, the communication overhead between the SUs and the FC is increased since each SU must transmit their KP complex-valued observed frequency samples per sensing frame for M consecutive frames before the EM algorithm can be run by the FC. This overhead can be reduced by adopting a distributed or localized implementation for the EM-JDE, as explained in the following.

The block diagram of the distributed implementation of the proposed EM-JDE is presented in Fig. 3. In the proposed distributed implementation, each SU generates an iterative soft estimate of B_k locally by running a simplified version of the EM algorithm (the derivation is similar to the one in Section III-A and therefore omitted for brevity), thereby producing the sequence $\hat{B}_{k,n}^{(i)}$. Then, each SU reports its estimate of B_k after convergence, i.e., $\hat{B}_{k,n}^{(\infty)}$, to the FC to make a decision on the spectral occupancy. Using these reported estimates from the N SUs, where error-free transmission is assumed, the decision statistic on the kth subband is defined as

$$\hat{B}_k = \frac{1}{N} \sum_{n=0}^{N-1} \hat{B}_{k,n}^{(\infty)}. \tag{25}$$

Compared with the centralized implementation described earlier, the distributed implementation significantly reduces the communication overhead between the FC and the SUs to K complex values per SU per sensing period of M frames. However,

this is achieved at the expense of increasing the hardware complexity of the CR units used by the SUs. The performance of the given low-overhead distributed approach is compared with that of the centralized spectrum sensing in Section V.

3) Initialization: In theory, the EM algorithm monotonically increases the log-likelihood function of the observed data at each iteration. Therefore, it is guaranteed to converge to a stationary point of the likelihood function, which can be a local maxima or a saddle point[4] (see, e.g., [39] and [40]). In practice, the convergence of the EM algorithm to a global maxima (i.e., the ML solution) can be achieved by using reliable initialization of the unknown parameters. This can also guarantee fast convergence to the ML solution within a reasonable number of iterations. In the following, we present an initialization strategy that ensures good convergence under practical conditions of operation in the CR network with multiple SUs operating over time-varying channels.

Starting with $\hat{\varsigma}_{k,n}^{(0)}$, it is assumed that the FC has *a priori* knowledge about the recent history of the PU's activity, i.e., the FC can record the intervals of presence and absence of the PU signal in the time–frequency plane (e.g., based on some basic form of ED or other available data) to determine the most likely intervals over which the kth subband is not occupied, and then report this information to the SUs. By targeting such an idle period of the PU activity, the initial estimate of $\varsigma_{k,n}$ at the nth SU is determined by computing the sample variance of the observations from the P receive antennas, as follows[5]:

$$\hat{\varsigma}_{k,n}^{(0)} = \frac{1}{J_n P} \sum_{j=0}^{J_n-1} \sum_{p=0}^{P-1} |V_{k,n,p}(j)|^2 \tag{26}$$

where J_n is the number of available frames during an idle period. In the centralized implementation, the values of $\hat{\varsigma}_{k,n}^{(0)}$ from the N SUs are reported along with the frequency samples $R_{k,n,p}(m)$ to the FC, where the complete EM-JDE algorithm can be run; whereas in the distributed implementation, the value of $\hat{\varsigma}_{k,n}^{(0)}$ computed at the nth SU is used locally to run the distributed version of the EM-JDE, as shown in Fig. 3. Following initialization for the first block of M frames based on *a priori* knowledge of the PU's activity as given earlier, initialization for subsequent blocks can be based on the noise variance estimates obtained via application of the EM-JDE algorithm in the previous frame.

The estimate of B_k is initialized by the FC as follows. Let $\hat{\kappa}_{k,n}^{(0)}$ represent the sample variance of the received signals at the P antennas of the nth SU normalized by $\hat{\varsigma}_{k,n}^{(0)}$, as given in the following:

$$\hat{\kappa}_{k,n}^{(0)} = \frac{1}{MP\hat{\varsigma}_{k,n}^{(0)}} \sum_{m=0}^{M-1} \sum_{p=0}^{P-1} |R_{k,n,p}(m)|^2 . \tag{27}$$

On the one hand, when the kth subband is occupied by the PU (i.e., $B_k = 1$), $\hat{\kappa}_{k,n}^{(0)}$ represents an initial estimate of the received SNR at the nth SU in that subband. On the other hand, when the kth subband is idle (i.e., $B_k = 0$), we have $R_{k,n,p}(m) = V_{k,n,p}(m)$, and then $\hat{\kappa}_{k,n}^{(0)} \approx 1$. Each SU transmits its variance estimate $\hat{\kappa}_{k,n}^{(0)}$ to the FC, which subsequently computes the initial estimate

$$\hat{B}_k^{(0)} = \min \left\{ \frac{1}{N} \sum_{n=0}^{N-1} \left| \hat{\kappa}_{k,n}^{(0)} - 1 \right|, 1 \right\}. \tag{28}$$

As a result, $\hat{B}_k^{(0)}$ is more likely to take values close to zero when the kth subband is vacant, as well as values larger than zero when the PU signal is present.

Finally, we discuss the initialization of the channel estimates. In a traditional (i.e., nonopportunistic) communication network, the source can transmit a short sequence of known training symbols to help the receiver initiate the channel estimation [46]. In CR networks, this approach is not feasible since the SUs have no *a priori* information about the PU signal. One possible alternative is simply to initialize the unknown channel coefficients to zero, i.e., $\hat{\mathbf{H}}_{k,n}^{(0)} = \mathbf{0}_P$, where $\mathbf{0}_P$ denotes a $P \times 1$ zero vector. However, the zero initialization might increase the probability of missed detection, which in turn leads to higher interfering rate with the PU. In this paper, we consider a more practical approach where each SU has only very limited knowledge of the PU-to-SU channels. The true but unknown channel between the PU and the pth receive antenna of the nth SU in the kth subband can be represented by the following additive model [47]:

$$H_{k,n,p} = \bar{H}_{k,n,p} + \Delta H_{k,n,p} \tag{29}$$

where $\bar{H}_{k,n,p}$ is the available channel estimate at the SU, which can be inaccurate, and $\Delta H_{k,n,p}$ captures the underlying channel uncertainty. Specifically, the uncertainty $\Delta H_{k,n,p}$ is assumed to take values from the following bounded set:

$$\mathcal{H}_{k,n,p} \triangleq \{\Delta H : |\Delta H| \leq \epsilon_{k,n,p}\} \tag{30}$$

where $\epsilon_{k,n,p} > 0$ specifies the radius of $\mathcal{H}_{k,n,p}$ and therefore reflects the degree of uncertainty associated with the available channel estimate $\bar{H}_{k,n,p}$. Such a model has been extensively used in transceiver design for CR networks [48], [49], where

[4]While the convergence rate of the EM algorithm near a stationary point can be analyzed by means of Taylor series expansion and Jacobian computation of the EM iterative mapping function [45], this analysis in the current setup would be extremely complicated and of limited value as it would provide information about convergence in a small neighborhood of a stationary point. In a practical setup as considered here, the convergence rate of the EM algorithm highly depends on the initialization of the unknown parameters and can vary on a case-by-case basis. In our work, as an alternative to such analysis and in order to guarantee fast convergence within a reasonable number of iterations, we propose some novel initialization techniques that are specifically designed for the use of the EM-JDE algorithm in wideband spectrum sensing applications.

[5]The effect of the SNR wall on ED with ML noise variance estimation has been thoroughly investigated in [10]. In [10, Th. 1], it has been proven that ED with noise power estimation can avoid the SNR wall phenomenon if the variance of the noise estimator decreases in $o(1)$ with the number of the sensing frames. In the special case of interest here, where the narrow-band noise power is estimated according to the sample average in (26), we have that $\lim_{J_n \to \infty} \mathrm{var}\{\hat{\varsigma}_{k,n}^{(0)}\} = 0$ under a standard Gaussian assumption for the noise samples. The condition of the theorem is therefore satisfied: In theory, there is no SNR wall, and any arbitrary pair $(P_{f,k}, P_{d,k})$ can be achieved for the ED by increasing the observation interval.

$\bar{H}_{k,n,p}$ can be obtained by calculating the deterministic path-loss coefficients between PU and the different SU antennas, whereas the size of the uncertainty region can be derived from the fading channel statistics. Therefore, for each triplet (k, n, p), an initial guess of the channel frequency response can be obtained as

$$\hat{H}_{k,n,p}^{(0)} = \bar{H}_{k,n,p} + \Delta H_{k,n,p} \tag{31}$$

where $\Delta H_{k,n,p}$ is randomly generated from the set $\mathcal{H}_{k,n,p}$. To further reduce the probability of missed detection of the proposed EM-JDE algorithm, multiple initializations are generated according to (31) in parallel, and the one with the largest value of the corresponding complete data log-likelihood function, i.e., $L(\mathbf{Y}|\mathbf{U})$ in (6), is selected at the expense of using additional computing resources.

4) Operation of EM-JDE Algorithm: In our implementation of the proposed EM-JDE scheme, we propose the following strategy to operate the iterative algorithm with the goal of enhancing its convergence and stability. In this strategy, the sequence of EM iterations is divided into S consecutive stages, which are indexed by $s \in \{1, \ldots, S\}$, where each stage comprises T iterations, which are indexed by $i = (s-1)T + j$, where $j \in \{1, \ldots, T\}$. In each stage, we fix the value of $\hat{B}_k^{(i+1)} = \hat{B}_k^{(i)}$ for the first $T-1$ iterations (i.e., for $j = 1, \ldots, T-1$), whereas the values of $\hat{H}_k^{(i+1)}$ and $\hat{\varsigma}_k^{(i+1)}$ are updated using the EM formulas derived in Section III-A with every iteration i. At the last iteration of each stage (i.e., when we reach $j = T$), $\hat{B}_k^{(i+1)}$ is updated along with $\hat{H}_k^{(i+1)}$ and $\hat{\varsigma}_k^{(i+1)}$ using the most recent estimates $\hat{H}_k^{(i)}$ and $\hat{\varsigma}_k^{(i)}$. This means that the values of $\hat{H}_k^{(i)}$ and $\hat{\varsigma}_k^{(i)}$ always change with the iteration index i, whereas the value of $\hat{B}_k^{(i)}$ only changes with the stage order s, i.e., when $i = sT$. Indeed, with imperfect knowledge of these quantities, the iterative EM estimate $\hat{B}_k^{(i)}$ might not converge sufficiently rapidly to the true value B_k due to the uncertainty in the channel and noise parameters. We have found experimentally that by devoting more iterations for improving the channel and noise estimation before updating $\hat{B}_k^{(i+1)}$, we can ascertain a unique slope of the EM estimate $\hat{B}_k^{(i)}$ toward one of the two limiting values, i.e., 0 or 1. Specifically, this increases the likelihood that at the end of a given iteration, $\hat{B}_k^{sT} > \hat{B}_k^{(s-1)T}$ when $B_k = 1$ and $\hat{B}_k^{sT} < \hat{B}_k^{(s-1)T}$ when $B_k = 0$. Hence, the stability of the EM algorithm is improved, even with imperfect channel and noise variance estimation.

IV. Upper Bound on the Asymptotic Performance of Expectation–Maximaization–Joint Detection and Estimatiox

The performance of spectrum sensing schemes is evaluated using the Neyman–Pearson criterion, where for a given probability of false alarm in the kth subband, e.g., $P_{f,k}(\gamma_k) = \alpha_k$ with $\alpha_k \in [0, 1]$, the optimum threshold in the decision-making process γ_k^{opt} and, subsequently, the optimum probability of detection $P_{d,k}^{\text{opt}}(\gamma_k^{\text{opt}})$ are provided [41]. This analytical evaluation cannot be applied in iterative JDE schemes since the derivation of a closed-form expression for the final decision statistics, i.e., $\hat{B}_k^{(\infty)}$, is not apparently feasible. Therefore, we present an analytical evaluation of the EM-JDE assuming a perfect knowledge of the CSI and noise variances by the SUs. In this case, the ROC curve represents an upper bound on the performance of the EM-JDE scheme. The derivations can be summarized as follows. First, we present the EM estimation of B_k assuming that ς_k and $\mathbf{H}_k$ are perfectly known by the SU, which enables us to express $\hat{B}_k^{(i+1)}$ in terms of $\hat{B}_k^{(i)}$. Then, we obtain an explicit expression of $\hat{B}_k^{(\infty)}$ by deriving the ML solution of the same problem, which is denoted $\hat{B}_k^{\text{ML}}$, and by proving that $\lim_{i\to\infty} \hat{B}_k^{(i)} = \hat{B}_k^{\text{ML}}$. Finally, using $\hat{B}_k^{\text{ML}}$, closed-form expressions of γ_k^{opt} and $P_{d,k}^{\text{opt}}(\gamma_k^{\text{opt}})$ are derived.

The analysis presented in the following is important for several reasons. First, it sheds light on the convergence properties of the proposed EM-JDE scheme under idealized conditions. Second, the closed-form expressions of $P_{f,k}(\gamma_k)$ and $P_{d,k}(\gamma_k)$ obtained for the ideal ML solution can be used as an upper bound on the performance of the proposed scheme, allowing setting of the preliminary values of the detection thresholds γ_k to achieve a specified false-alarm rate.

A. EM-Based Spectrum Sensing

Here, we determine the spectrum occupancy assuming the perfect knowledge of channel frequency responses $\mathbf{H}$ and noise variances ς. Therefore, the only unknown parameter in the kth subband is the occupancy parameter B_k, and we denote the EM solution in this case as the ideal EM-based spectrum sensing (IdEM-SS). Following the same procedure as in Section III-A, $\hat{B}_k^{(i+1)}$ is obtained as in (10), i.e.,

$$\hat{B}_k^{(i+1)} = \frac{1}{M} \sum_{m=0}^{M-1} E\left[|S_k(m)|^2 \,|\mathbf{R}, \hat{\mathbf{B}}^{(i)}\right] \tag{32}$$

where $E[|S_k(m)|^2|\mathbf{R}, \hat{\mathbf{B}}^{(i)}]$ can be derived using (13) and (14) with $\hat{\mathbf{H}}_k^{(i)}$ and $\hat{\varsigma}_k^{(i)}$ replaced by their true values. To justify the convergence of $\hat{B}_k^{(i+1)}$ to $\hat{B}_k^{\text{ML}}$ as $i \to \infty$, we first derive a closed-form expression for the ML estimator of $\mathbf{B}$, which is referred to as the ideal ML-based spectrum sensing (IdML-SS), as follows.

The log-likelihood function of $\mathbf{R}$ given $\mathbf{B}$ (assuming that $\mathbf{H}$ and ς are known) is

$$L(\mathbf{R}|\mathbf{B}) = \sum_{k=0}^{K-1} L(\mathbf{R}_k|B_k) \tag{33}$$

where

$$L(\mathbf{R}_k|B_k) = -MNP\ln(\pi) - M\ln\left(\det(\mathbf{\Gamma}_k)\right) - \sum_{m=0}^{M-1} \mathbf{R}_k(m)^H \mathbf{\Gamma}_k^{-1} \mathbf{R}_k(m). \tag{34}$$

In (34), $\mathbf{\Gamma}_k = B_k \mathbf{H}_k \mathbf{H}_k^H + \mathbf{\Sigma}_k$ with $\mathbf{\Sigma}_k$ defined as in (15) but using the true values of $\varsigma_{k,n}$, i.e., $\hat{\varsigma}_{k,n}^{(i)} = \varsigma_{k,n}$, and $\det(\cdot)$

denotes the matrix determinant. Using the matrix determinant lemma [50], $\det(\mathbf{\Gamma}_k)$ is reduced to

$$\begin{aligned}\det(\mathbf{\Gamma}_k) &= \left(1 + B_k \mathbf{H}_k^H \mathbf{\Sigma}_k^{-1} \mathbf{H}_k\right) \det(\mathbf{\Sigma}_k) \\ &= (1 + B_k \mathbf{H}_k^H \mathbf{\Sigma}_k^{-1} \mathbf{H}_k) \prod_{n=0}^{N-1} \varsigma_{k,n}^P. \end{aligned} \tag{35}$$

Moreover, by using the Sherman–Morrison formula [44], $\mathbf{\Gamma}_k^{-1}$ can be expressed as

$$\mathbf{\Gamma}_k^{-1} = \mathbf{\Sigma}_k^{-1} - \frac{B_k \mathbf{\Sigma}_k^{-1} \mathbf{H}_k \mathbf{H}_k^H \mathbf{\Sigma}_k^{-1}}{1 + B_k \mathbf{H}_k^H \mathbf{\Sigma}_k^{-1} \mathbf{H}_k}. \tag{36}$$

Substituting (35) and (36) in (34) and neglecting the terms independent of B_k, we obtain

$$\begin{aligned} L(\mathbf{R}_k|B_k) &= -M \ln(1 + B_k \mathbf{H}_k^H \mathbf{\Sigma}_k^{-1} \mathbf{H}_k) + \frac{B_k}{1 + B_k \mathbf{H}_k^H \mathbf{\Sigma}_k^{-1} \mathbf{H}_k} \\ &\quad \times \sum_{m=0}^{M-1} \mathbf{R}_k(m)^H \mathbf{\Sigma}_k^{-1} \mathbf{H}_k \mathbf{H}_k^H \mathbf{\Sigma}_k^{-1} \mathbf{R}_k(m). \end{aligned} \tag{37}$$

Since the subband occupancies $\{B_k\}$ are assumed statistically independent, the maximization process of (33) with respect to $\mathbf{B}$ is done separately for each subband. The ML estimate of the kth subband occupancy is obtained by maximizing the log-likelihood function (37) with respect to B_k, i.e., $\hat{B}_k^{\mathrm{ML}} = \arg\max_{B_k} L(\mathbf{R}_k|B_k)$. To simplify its derivation, let us introduce random variable $\Upsilon_k(m) = \mathbf{R}_k(m)^H \mathbf{\Sigma}_k^{-1} \mathbf{H}_k$ and define $\Psi_k = \mathbf{H}_k^H \mathbf{\Sigma}_k^{-1} \mathbf{H}_k$. Then, $\hat{B}_k^{\mathrm{ML}}$ is obtained by taking the derivative of (37) with respect to B_k and equating the resultant equation to 0 as follows:

$$\begin{aligned} \frac{-M\Psi_k}{1 + B_k\Psi_k} + \left(\frac{1}{1 + B_k\Psi_k} - \frac{B_k\Psi_k}{(1 + B_k\Psi_k)^2}\right) \\ \times \sum_{m=0}^{M-1} |\Upsilon_k(m)|^2 = 0 \end{aligned} \tag{38}$$

which yield

$$\hat{B}_k^{\mathrm{ML}} = \frac{1}{M\Psi_k^2} \sum_{m=0}^{M-1} \left(|\Upsilon_k(m)|^2 - \Psi_k\right). \tag{39}$$

Using (36) with $B_k = \hat{B}_k^{(i)}$, (13) is reduced to

$$\begin{aligned} E[S_k(m)|\mathbf{R}, \hat{\mathbf{B}}^{(i)}] &= \hat{B}_k^{(i)} \mathbf{H}_k^H \left(\mathbf{\Sigma}_k^{-1} - \frac{\hat{B}_k^{(i)} \mathbf{\Sigma}_k^{-1} \mathbf{H}_k \mathbf{H}_k^H \mathbf{\Sigma}_k^{-1}}{1 + \hat{B}_k^{(i)} \Psi_k}\right) \mathbf{R}_k(m) \\ &= C_k A_{k,m} \end{aligned} \tag{40}$$

where we define $C_k = (1 + \hat{B}_k^{(i)} \Psi_k)^{-1}$ and $A_{k,m} = \hat{B}_k^{(i)} \Upsilon_k(m)^H$. Similarly, (14) is expressed as

$$\mathrm{Var}\left[S_k(m)|\mathbf{R}, \hat{\mathbf{B}}^{(i)}\right] = \frac{\hat{B}_k^{(i)}}{1 + \hat{B}_k^{(i)} \Psi_k}. \tag{41}$$

Substituting (40) and (41) in (32), we have

$$\hat{B}_k^{(i+1)} = \frac{1}{M} \sum_{m=0}^{M-1} \frac{\hat{B}_k^{(i)2} \, |\Upsilon_k(m)|^2}{\left(1 + \hat{B}_k^{(i)} \Psi_k\right)^2} + \frac{\hat{B}_k^{(i)}}{1 + \hat{B}_k^{(i)} \Psi_k}. \tag{42}$$

Substituting $\hat{B}_k^{(i)} = \hat{B}_k^{(i+1)} = \hat{B}_k^{(\infty)}$ in (42) and solving the resultant equation, we obtain $\hat{B}_k^{(\infty)} = \hat{B}_k^{\mathrm{ML}}$ as given by (39).

B. Performance Evaluation

Here, we derive closed-form expressions for the probability of false alarm and missed detection, i.e., $P_{f,k}(\gamma_k)$ and $P_{d,k}(\gamma_k)$, respectively, for the IdML-SS. Based on the analysis in Part A, it follows that these expressions will also be applicable to the IdEM-SS approach in the limit of large i, assuming that $\mathbf{H}_k$ and ς_k are known. The desired performance metrics are derived by considering the binary test in (24), with $\hat{B}_k^{\mathrm{EM}}$ now replaced by $\hat{B}_k^{\mathrm{ML}}$, and making use of the expression of $\hat{B}_k^{\mathrm{ML}}$ in (39).

Conditioned on $\mathbf{H}_k$ and ς_k, the random variable $\Upsilon_k(m) = \mathbf{R}_k^H(m) \mathbf{\Sigma}_k^{-1} \mathbf{H}_k$ has a complex Gaussian distribution with zero mean and variance $E[|\Upsilon_k(m)|^2] = E[|\mathbf{R}_k(m)^H \mathbf{\Sigma}_k^{-1} \mathbf{H}_k \mathbf{H}_k^H \mathbf{\Sigma}_k^{-1} \mathbf{R}_k(m)|]$, whose expression can be derived as follows. We first expand $|\Upsilon_k(m)|^2$ as the following sum:

$$|\Upsilon_k(m)|^2 = \sum_{n=0}^{N-1} \sum_{n'=0}^{N-1} \zeta_{k,n}(m) \zeta_{k,n'}(m)^* \tag{43}$$

where $\zeta_{k,n}(m) = \varsigma_{k,n}^{-1} \mathbf{R}_{k,n}(m)^H \mathbf{H}_{k,n}$. Then, we derive the expected value of the cross term $\zeta_{k,n}(m)\zeta_{k,n'}(m)^*$, which is given as follows (see the Appendix):

$$\begin{aligned} E\left[\zeta_{k,n}(m)\zeta_{k,n'}(m)^*\right] \\ = \varsigma_{k,n}^{-1} \varsigma_{k,n'}^{-1} \|\mathbf{H}_{k,n}\|^2 \left(\|\mathbf{H}_{k,n'}\|^2 B_k + \varsigma_{k,n} \delta_{n,n'}\right) \end{aligned} \tag{44}$$

where $\|\cdot\|$ returns the 2-norm of a vector. Using (44), $E[|\Upsilon_k(m)|^2]$ is obtained as

$$\begin{aligned} E\left[|\Upsilon_k(m)|^2\right] &= \sum_{n=0}^{N-1} \varsigma_{k,n}^{-2} \|\mathbf{H}_{k,n}\|^2 \left(\|\mathbf{H}_{k,n}\|^2 B_k + \varsigma_{k,n}\right) \\ &\quad + B_k \sum_{n=0}^{N-1} \sum_{n'=0, n' \neq n}^{N-1} \varsigma_{k,n}^{-1} \varsigma_{k,n'}^{-1} \|\mathbf{H}_{k,n}\|^2 \|\mathbf{H}_{k,n'}\|^2. \end{aligned} \tag{45}$$

Let $|\Upsilon_k|^2 = \sum_{m=0}^{M-1} |\Upsilon_k(m)|^2$, which is a sum of M statistically independent terms, where each term $|\Upsilon_k(m)|^2$, once normalized by half its variance, follows a chi-square distribution with two degrees of freedom, which is denoted χ_2^2. Therefore, $(|\Upsilon_k|^2 / (E[|\Upsilon_k(m)|^2]/2))$ has a chi-square distribution with

$2M$ degrees of freedom, i.e., χ^2_{2M}. Under the hypothesis $B_k = 0$, $E[|\Upsilon_k(m)|^2]_{B_k=0} \triangleq \sum_{n=0}^{N-1} \varsigma_{k,n}^{-1} \|\mathbf{H}_{k,n}\|^2$. Using the closed-form expression of the complementary cumulative distribution function of the chi-square distribution, $P_{f,k}(\gamma_k)$ is given by [14]

$$\begin{aligned} P_{f,k}(\gamma_k) &= P\left(\hat{B}_k^{\mathrm{ML}} > \gamma_k | B_k = 0\right) \\ &= P\left(\frac{|\Upsilon_k|^2}{E\left[|\Upsilon_k(m)|^2\right]_{B_k=0}} > \frac{\gamma_k M \Psi_k^2 + M\Psi_k}{\sum_{n=0}^{N-1} \varsigma_{k,n}^{-1} \|\mathbf{H}_{k,n}\|^2}\right) \\ &= \Gamma\left(M, \frac{\gamma_k M \Psi_k^2 + M\Psi_k}{\sum_{n=0}^{N-1} \varsigma_{k,n}^{-1} \|\mathbf{H}_{k,n}\|^2}\right) \end{aligned} \quad (46)$$

where $\Gamma(c,y) = (1/\Gamma(c)) \int_y^\infty x^{c-1} e^{-x} dx = w$ is the normalized upper incomplete gamma function, and $\Gamma(c) = \int_0^\infty x^{c-1} e^{-x} dx$ represents the complete gamma function. Under the constraint $P_{f,k}(\gamma_k) = \alpha_k$, the optimum threshold is given as follows:

$$\gamma_k^{\mathrm{opt}} = \frac{1}{M\Psi_k^2}\left(\Gamma^{\mathrm{inv}}(M, \alpha_k) \sum_{n=0}^{N-1} \left(\varsigma_{k,n}^{-1} \|\mathbf{H}_{k,n}\|^2\right) - M\Psi_k\right) \quad (47)$$

where $\Gamma^{\mathrm{inv}}(c,w)$ is the inverse incomplete gamma function of $\Gamma(c,y)$ [51]. Similarly, under the hypothesis $B_k = 1$, the optimum probability of detection under the constraint that $P_{f,k}(\gamma_k) = \alpha_k$, i.e., according to the Neyman–Pearson formulation of the binary hypothesis testing problem, is obtained as follows [52]:

$$\begin{aligned} &P_{d,k}^{\mathrm{opt}}\left(\gamma_k^{\mathrm{opt}}\right) \\ &= P\left(\hat{B}_k^{\mathrm{ML}} > \gamma_k^{\mathrm{opt}} | B_k = 1\right) \\ &= P\left(\frac{|\Upsilon_k|^2}{E\left[|\Upsilon_k(m)|^2\right]_{B_k=1}} > \frac{\gamma_k^{\mathrm{opt}} M \Psi_k^2 + M\Psi_k}{E\left[|\Upsilon_k(m)|^2\right]_{B_k=1}}\right) \\ &= \Gamma\left(M, \frac{\gamma_k^{\mathrm{opt}} M \Psi_k^2 + M\Psi_k}{E\left[|\Upsilon_k(m)|^2\right]_{B_k=1}}\right) \end{aligned} \quad (48)$$

where

$$\begin{aligned} &E\left[|\Upsilon_k(m)|^2\right]_{B_k=1} \\ &= \sum_{n=0}^{N-1} \varsigma_{k,n}^{-2} \|\mathbf{H}_{k,n}\|^2 \left(\|\mathbf{H}_{k,n}\|^2 + \varsigma_{k,n}\right) \\ &\quad + \sum_{n=0}^{N-1} \sum_{n'=0, n' \neq n}^{N-1} \varsigma_{k,n}^{-1} \varsigma_{k,n'}^{-1} \|\mathbf{H}_{k,n}\|^2 \|\mathbf{H}_{k,n'}\|^2. \end{aligned} \quad (49)$$

We remark that, by using the threshold derived in (47), the theoretical ROC obtained from (46) and (48) acts as an upper bound for the proposed EM-based JDE with unknown CSI and noise variance. Therefore, the "ideal" threshold (47) in itself can actually provide some insights on how to determine a practical threshold, although it is not necessarily optimal for the EM-JDE. From (47), we observe that the threshold can be sensitive to the sensing interval M, the probability of false alarm via α_k, the entries of the instantaneous fading channel vector $\mathbf{H}_k$, and the noise variance $\varsigma_{k,n}$.

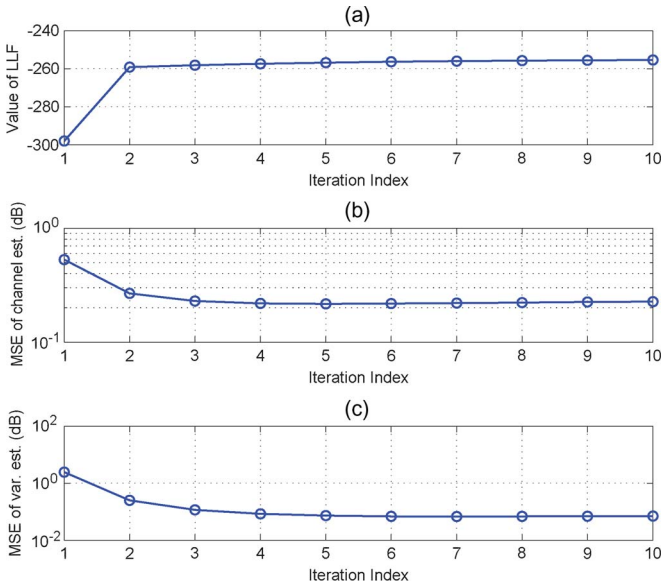

Fig. 4. Convergence behavior of the proposed EM-JDE scheme for a CR scenario with $N = 1$ SU equipped with $P = 2$ antennas. (a) Log-likelihood function of the complete data in (6). (b) MSE of channel estimation in (50). (c) MSE of noise variance estimation in (51) (SNR $= -3$ dB).

V. Simulation Experiments

The performance of the proposed JDE scheme based on the EM algorithm is evaluated through its ROC curves. Throughout our simulations, we assume a CR network of N SUs, where each SU is equipped with P receive antennas and operates in a wideband frequency spectrum with K subbands. Since the estimation of the unknown parameters is performed independently for each subband, our results are presented for a selected subband, e.g., $k = 0$. Moreover, B_k is estimated on a block-by-block basis in which the channel vector $\mathbf{H}_k$ has a constant value within a block of M samples[6] and changes independently from a block to the next. Our results are produced for $M = 50$ samples, and the channel vector $\mathbf{H}_k$ is modeled as complex Gaussian vector with zero mean and covariance $\sigma_h^2 \mathbf{I}$. We assume that the noise variance $\varsigma_{k,n}$ is constant over the P receive antennas of each SU. In this case, the average SNR per receive antenna of each SU in the kth subband is given by SNR $= (B_k \sigma_h^2 / \varsigma_{k,n})$. For each choice of the detection threshold γ_k in (24), 10^5 trials are used to estimate the performance metrics of the proposed EM-JDE scheme. Except for the results in Fig. 4, where $T = 1$, we run the EM algorithm for $S = 4$ stages, where each stage comprises $T = 5$ iterations; initialization is performed as explained in Section III-B3.

[6]In practice, the value of M can be determined according to the coherence time of the channel.

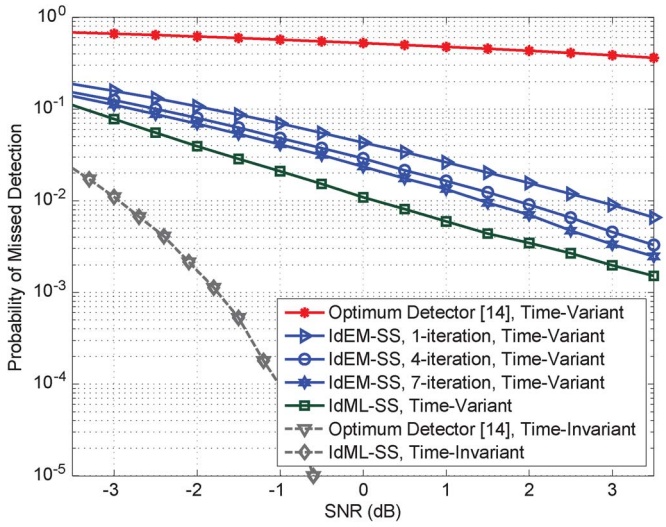

Fig. 5. Probability of missed detection versus SNR of the proposed EM-based spectrum sensing schemes for $N = 1$ SU with $P = 2$ antennas over time-varying Rayleigh fading channels ($P_{f,k}(\gamma_k) = 0.05$).

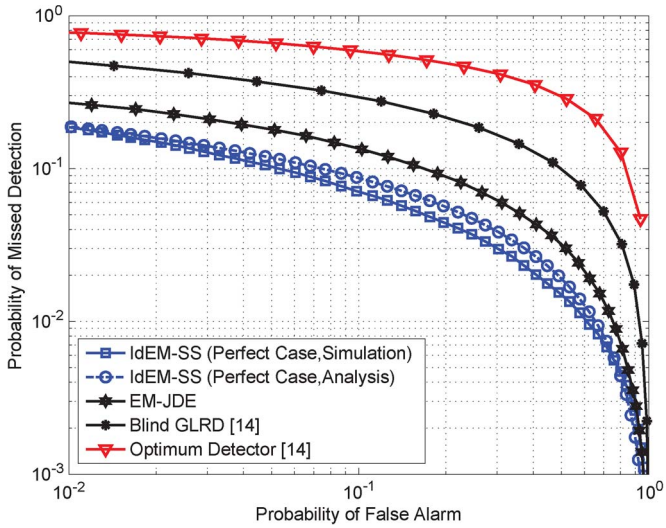

Fig. 6. ROCs of different spectrum sensing schemes for the CR scenario with $N = 1$ SU equipped with $P = 2$ antennas (SNR $= -3$ dB).

In Figs. 4–6, we first study the performance of the proposed spectrum sensing techniques for the basic CR configuration with $N = 1$ SU equipped with $P = 2$ antennas. Subsequently, some of these parameters will be varied. In Fig. 4, we show the convergence of the proposed EM-JDE scheme for the basic CR configuration; the value of the log-likelihood function of the complete data (9) is plotted in Fig. 4(a), whereas the MSE of channel estimation and noise power estimation, which are defined as

$$\frac{1}{P}\sum_{p=1}^{P} E\left[\left|H_{k_0,0,p} - \hat{H}_{k_0,0,p}\right|^2\right] \tag{50}$$

$$\frac{1}{P}\sum_{p=1}^{P} E\left[\left|\varsigma_{k_0,0,p} - \hat{\varsigma}_{k_0,0,p}\right|^2\right] \tag{51}$$

respectively, are plotted in Fig. 4(b) and (c). The results show that the value of the log-likelihood function increases monotonically and reaches a local maximum within a few iterations, whereas simultaneously, the channel and noise parameter estimates converge to stationary points.

For the same configuration as earlier, Fig. 5 examines the performance of spectrum sensing techniques over time-variant Rayleigh fading channels with a perfect estimation of $\mathbf{H}_k$ and ς_k for a fixed probability of false alarm $P_{f,k}(\gamma_k) = 0.05$. We also include the optimum detector proposed in [14] as a performance benchmark.[7] We first note that the IdML-SS outperforms the optimum detector [14] with significant gains in PU's signal detection. The better performance of the IdML-SS is due to the fact that it uses estimates of the average transmit power as decision statistics [see (39)], which in turn is independent of the time-varying channel gains. The proposed IdEM-SS also outperforms the optimum detector [14] by a wide margin in the time-variant case, with the probability of missed detection converging to that of the IdML-SS as the number of iterations i increases. On this figure, as a lower bound on the missed detection performance of the spectrum sensing detectors, we also present results for the case of time-invariant channels, where $\mathbf{H}_k$ is constant over all sensing intervals. Both the IdML-SS and the optimum detector achieve the same performance in this ideal situation. Comparing the results for time-varying channels to those for time-invariant channels, we conclude that user mobility can have a major impact on, i.e., significantly degrade the achievable detection performance, spectrum sensing schemes.

Still for the case of $N = 1$ SU with $P = 2$ antennas, Fig. 6 evaluates the performance of the proposed EM-JDE scheme over time-variant Rayleigh fading channels in the practical case where $\mathbf{H}_k$ and ς_k are unknown by the SU. For the purpose of comparison, we also include the ROC curve of the blind general likelihood ration detector (GLRD) proposed in [14]. The results show that the EM-JDE scheme enhances the spectrum detection process compared with the blind GLRD. As a benchmark on the performance of the proposed scheme, we add the ROC curve of the IdEM-SS with the same simulation parameters assuming perfect channel and noise variance estimations. We note that, in this case, the performance of the proposed JDE scheme with unknown $\mathbf{H}_k$ and ς_k comes very close to the results obtained for this ideal case, whereas the validity of the theoretical results derived in Section IV-B is demonstrated.

Fig. 7 presents the performance of the proposed JDE scheme for multiantenna CR networks with $N \in \{1, 2, 3, 4\}$ spatially distributed SUs, where each SU has $P = 2$ antennas. The simulation results show that the proposed JDE scheme can significantly enhance the spectrum detection by exploiting the spatial diversity of the CR network. In Fig. 8, the performance of the EM-JDE scheme for a CR network with $N = 3$ SU, each

[7]The optimum detector in [14] refers to the Neyman–Pearson detector under the assumption of known and time-invariant channel gain and noise variance parameters. This detector implements an LRT based on a Gaussian signal model, where the detection threshold γ_k is adjusted to minimize the probability of missed detection $P_{m,k}(\gamma_k)$, which is subject to a constraint on the probability of false alarm, i.e., $P_{f,k}(\gamma_k) < \alpha$.

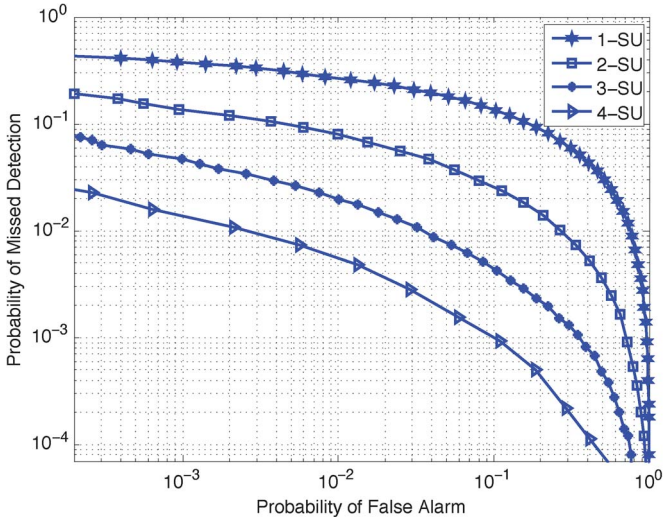

Fig. 7. ROC of the proposed EM-JDE scheme for a multiuser CR network with $N \in \{1, 2, 3, 4\}$ SUs, each equipped with $P = 2$ antennas (SNR $= -3$ dB).

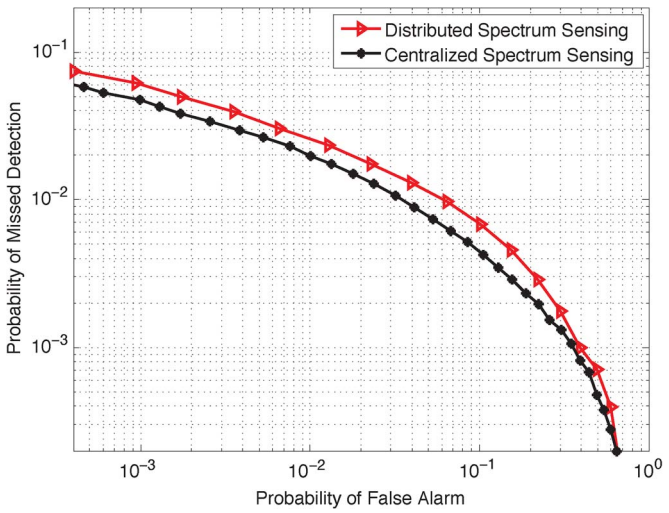

Fig. 8. ROC of the proposed EM-JDE scheme with centralized and distributed implementations for a CR network with $N = 3$ SUs, each with $P = 2$ antennas (SNR $= -3$ dB).

with $P = 2$ antennas, is presented, considering two different implementation approaches, that is centralized versus distributed spectrum sensing. The results show that the performance of the two schemes almost match with each other. Considering the challenges of physical implementations, one needs to consider the tradeoff between the communications overhead and the hardware complexity in choosing between these two different options.

Fig. 9 studies the effect of increasing the number of samples available for sensing, as represented here by the parameter M, on the performance of the proposed EM-JDE scheme for a CR network with $N = 2$ SUs, each equipped with $P = 2$ antennas. The plotted curves represent the relationship between the probability of missed detection in subband k $P_{m,k}(\gamma_k) = 1 - P_{d,k}(\gamma_k)$ against M for a fixed value of the probability of false alarm $P_{f,k}(\gamma_k) = 0.05$. The results show that the values of $P_{m,k}(\gamma_k)$ are reduced significantly by increasing M.

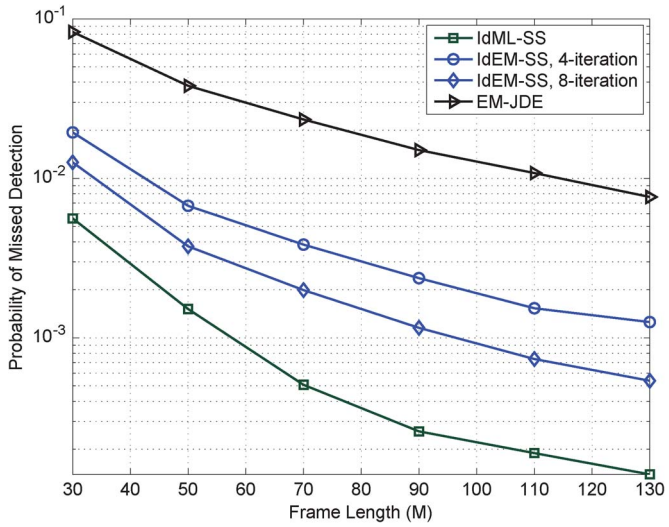

Fig. 9. Effect of M on the performance of the proposed EM-JDE scheme for a CR network with $N = 2$ SUs, each with $P = 2$ antennas ($P_{f,k}(\gamma_k) = 0.05$, and SNR $= -3$ dB).

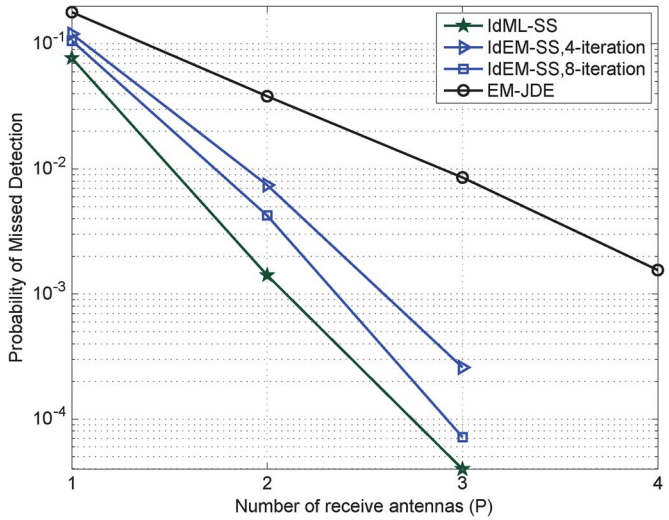

Fig. 10. Effect of P on the performance of the proposed EM-JDE scheme for a CR network with $N = 2$ SUs, each with $P = 2$ antennas ($P_{f,k}(\gamma_k) = 0.05$, and SNR $= -3$ dB).

Finally, Fig. 10 shows the effect of increasing the number of antennas P on the performance of the proposed EM-JDE scheme for a CR network with $N = 2$ SUs. Given $P_{f,k}(\gamma_k) = 0.05$, $P_{m,k}(\gamma_k)$ is plotted versus different values of P in the case of time-variant Rayleigh fading channels. From the results, we notice that increasing the number of antennas at the SUs can significantly enhance the ability of the proposed scheme to efficiently sense the available spectrum in cooperative CR networks.

VI. Conclusion

In this paper, a cooperative spectrum sensing scheme based on the EM algorithm for multiantenna CR networks has been proposed. In this scheme, a binary hypothesis test is applied on estimates of the average power transmitted by the PU over

a wideband frequency spectrum during the sensing interval, making the decision on the spectral occupancies independent of the channel states. However, knowledge of CSI and noise variance at each SU is crucial to obtain reliable estimates of the PU's transmitted power over different frequency subbands. Therefore, the FC employs the EM algorithm in a nontrivial way to jointly estimate the unknown continuous parameters in each subband along with the PU detection, thereby forming an EM-JDE scheme for *multiuser* multiantenna CR networks. Various aspects of this proposed EM-JDE scheme were investigated, including a reliable initialization strategy to ensure convergence under practical conditions and a distributed implementation to reduce communication overhead. We also introduced an analytical evaluation of the IdML-SS based on the Neyman–Pearson criterion as a lower bound on the missed detection performance of cooperative spectrum sensing. The results show that the proposed EM-JDE scheme can provide a significant improvement in the PU spectrum detection, particularly for time-variant channels. This paper has extended the application of advanced JDE schemes, which are widely employed in the design of modern wireless communication systems, to the topic of spectrum sensing in wideband CR networks.

APPENDIX

Here, we derive a closed-form expression for $E[\zeta_{k,n}(m)\zeta_{k,n'}(m)^*]$ in (44). First, we expand the product $\zeta_{k,n}(m)\zeta_{k,n'}(m)^*$ as

$$\begin{aligned}\zeta_{k,n}(m)&\zeta_{k,n'}(m)^* \\ &= \varsigma_{k,n}^{-1}\varsigma_{k,n'}^{-1}\mathbf{R}_{k,n}(m)^H\mathbf{H}_{k,n}\mathbf{H}_{k,n'}^H\mathbf{R}_{k,n'}(m) \\ &= \varsigma_{k,n}^{-1}\varsigma_{k,n'}^{-1}\sum_{p=0}^{P-1}\sum_{p'=0}^{P-1}H_{k,n,p}H_{k,n',p'}^*R_{k,n,p}(m)^*R_{k,n',p'(m)}.\end{aligned} \tag{52}$$

The expectation of the cross term $R_{k,n,p}(m)^*R_{k,n',p'}(m)$ is given by

$$E\left[R_{k,n,p}(m)^*R_{k,n',p'}(m)\right] = H_{k,n,p}^*H_{k,n',p'}B_k + \varsigma_{k,n}\delta_{n,n'}\delta_{p,p'}. \tag{53}$$

Using (53), we can obtain $E[\zeta_{k,n}(m)\zeta_{k,n'}(m)^*]$ in

$$\begin{aligned}E&\left[\zeta_{k,n}(m)\zeta_{k,n'}(m)^*\right] \\ &= \varsigma_{k,n}^{-1}\varsigma_{k,n'}^{-1}\left(B_k\sum_{p=0}^{P-1}\sum_{p'=0}^{P-1}|H_{k,n,p}|^2|H_{k,n',p'}|^2\right. \\ &\qquad\left. + \varsigma_{k,n}\sum_{p=0}^{P-1}|H_{k,n,p}|^2\delta_{n,n'}\right) \\ &= \varsigma_{k,n}^{-1}\varsigma_{k,n'}^{-1}\|\mathbf{H}_{k,n}\|^2\left(\|\mathbf{H}_{k,n'}\|^2B_k + \varsigma_{k,n}\delta_{n,n'}\right).\end{aligned} \tag{54}$$

REFERENCES

[1] Y.-C. Liang, K.-C. Chen, G. Y. Li, and P. Mahonen, "Cognitive radio networking and communications: An overview," *IEEE Trans. Veh. Technol.*, vol. 60, no. 7, pp. 3386–3407, Sep. 2011.

[2] Z. Quan, S. Cui, H. V. Poor, and A. Sayed, "Collaborative wideband sensing for cognitive radios," *IEEE Signal Process. Mag.*, vol. 25, no. 6, pp. 60–73, Nov. 2008.

[3] H. Sun, A. Nallanathan, C.-X. Wang, and Y. Chen, "Wideband spectrum sensing for cognitive radio networks: A survey," *IEEE Wireless Commun.*, vol. 20, no. 2, pp. 74–81, Apr. 2013.

[4] Z. Zhang, Q. Yang, L. Wang, and X. Zhou, "A novel hybrid matched filter structure for IEEE 802.22 standard," in *Proc. IEEE Asia Pacific Conf. Circuits Syst.*, Kuala Lumpur, Malaysia, Dec. 2010, pp. 652–655.

[5] K. W. Choi, W. S. Jeon, and D. G. Jeong, "Sequential detection of cyclostationary signal for cognitive radio systems," *IEEE Trans. Wireless Commun.*, vol. 8, no. 9, pp. 4480–4485, Sep. 2009.

[6] E. Rebeiz, P. Urriza, and D. Cabric, "Optimizing wideband cyclostationary spectrum sensing under receiver impairments," *IEEE Trans. Signal Process.*, vol. 61, no. 15, pp. 3931–3943, Aug. 2013.

[7] S. Atapattu, C. Tellambura, and H. Jiang, "Energy detection based cooperative spectrum sensing in cognitive radio networks," *IEEE Trans. Wireless Commun.*, vol. 10, no. 4, pp. 1232–1241, Apr. 2011.

[8] A. Pandharipande and J.-P. Linnartz, "Performance analysis of primary user detection in a multiple antenna cognitive radio," in *Proc. IEEE Int. Conf. Commun.*, Glasgow, U.K., Jun. 2007, pp. 6482–6486.

[9] V. R. S. Banjade, N. Rajatheva, and C. Tellambura, "Performance analysis of energy detection with multiple correlated antenna cognitive radio in Nakagami-m fading," *IEEE Commun. Lett.*, vol. 16, no. 4, pp. 502–505, Apr. 2012.

[10] R. Tandra and A. Sahai, "SNR walls for signal detection," *IEEE J. Sel. Topics Signal Process.*, vol. 2, no. 1, pp. 4–17, Feb. 2008.

[11] A. Mariani, A. Giorgetti, and M. Chiani, "Effects of noise power estimation on energy detection for cognitive radio applications," *IEEE Trans. Commun.*, vol. 59, no. 12, pp. 3410–3420, Dec. 2011.

[12] A. Singh, M. R. Bhatnagar, and R. K. Mallik, "Cooperative spectrum sensing in multiple antenna based cognitive radio network using an improved energy detector," *IEEE Commun. Lett.*, vol. 16, no. 1, pp. 64–67, Jan. 2012.

[13] L. Shen, H. Wang, W. Zhang, and Z. Zhao, "Multiple antennas assisted blind spectrum sensing in cognitive radio channels," *IEEE Commun. Lett.*, vol. 16, no. 1, pp. 92–94, Jan. 2012.

[14] A. Taherpour, M. Nasiri-Kenari, and S. Gazor, "Multiple antenna spectrum sensing in cognitive radios," *IEEE Trans. Wireless Commun.*, vol. 9, no. 2, pp. 814–823, Feb. 2010.

[15] P. Wang, J. Fang, N. Han, and H. Li, "Multiantenna-assisted spectrum sensing for cognitive radio," *IEEE Trans. Veh. Technol.*, vol. 59, no. 4, pp. 1791–1800, May 2010.

[16] R. Zhang, T. J. Lim, Y. C. Liang, and Y. Zeng, "Multi-antenna based spectrum sensing for cognitive radios: A GLRT approach," *IEEE Trans. Commun.*, vol. 58, no. 1, pp. 84–88, Jan. 2010.

[17] D. Ramirez, G. Vazquez-Vilar, R. Lopez-Valcarce, J. Via, and I. Santamaria, "Detection of rank-P signals in cognitive radio networks with uncalibrated multiple antennas," *IEEE Trans. Signal Process.*, vol. 59, no. 8, pp. 3764–3774, Aug. 2011.

[18] J. K. Tugnait, "On multiple antenna spectrum sensing under noise variance uncertainty and flat fading," *IEEE Trans. Signal Process.*, vol. 60, no. 4, pp. 1823–1832, Apr. 2012.

[19] K. Takeuchi, "Large-system analysis of joint channel and data estimation for MIMO DS-CDMA systems," *IEEE Trans. Inf. Theory*, vol. 58, no. 3, pp. 1385–1412, Mar. 2012.

[20] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via EM algorithm," *J. Royal Statist. Soc.*, vol. 39, no. 1, pp. 1–38, Jan. 1977.

[21] A. Assra and B. Champagne, "EM-based joint estimation and detection for multiple antenna cognitive radios," in *Proc. IEEE Int. Conf. Commun.*, Ottawa, ON, Canada, Jun. 2012, pp. 1502–1506.

[22] A. Assra, A. Vakili, and B. Champagne, "Iterative joint channel and noise variance estimation and primary user signal detection for cognitive radios," in *Proc. IEEE ISSPIT*, Dec. 2011, pp. 305–309.

[23] D. Romero and G. Leus, "Wideband spectrum sensing from compressed measurements using spectral prior information," *IEEE Trans. Signal Process.*, vol. 61, no. 24, pp. 6232–6246, Dec. 2013.

[24] F. Zeng, C. Li, and Z. Tian, "Distributed compressive spectrum sensing in cooperative multihop cognitive networks," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 24, pp. 37–48, Feb. 2011.

[25] P. Stoica and R. Moses, *Spectral Analysis of Signals*. Englewood Cliffs, NJ, USA: Prentice-Hall, 2005.
[26] J. G. Proakis and D. K. Mnolakis, *Digital Signal Processing*, 4th ed. Upper Saddle River, NJ, USA: Prentice-Hall, 2007.
[27] Z. Quan, S. Cui, A. Sayed, and H. V. Poor, "Optimal multiband joint detection for spectrum sensing in cognitive radio networks," *IEEE Trans. Signal Process.*, vol. 57, no. 3, pp. 1128–1140, Mar. 2009.
[28] P. P.-Hoseini and N. C. Beaulieu, "Optimal wideband spectrum sensing framework for cognitive radio systems," *IEEE Trans. Signal Process.*, vol. 59, no. 3, pp. 1170–1182, Mar. 2011.
[29] K. Hossain and B. Champagne, "Wideband spectrum sensing for cognitive radios with correlated subband occupancy," *IEEE Signal Process. Lett.*, vol. 18, no. 1, pp. 35–38, Jan. 2011.
[30] R. Xu and M. Chen, "Spectral leakage suppression for DFT-based OFDM via adjacent subcarriers correlative coding," in *Proc. IEEE GLOBECOM Conf.*, Nov. 2008, pp. 1–5.
[31] T. Hunziker, U. U. Rehman, and D. Dahlhaus, "Spectrum sensing in cognitive radios: Design of DFT filter banks achieving maximal timefrequency resolution," in *Proc. ICICS*, Dec. 2011, pp. 1–5.
[32] B. Farhang-Boroujeny, "Filter bank spectrum sensing for cognitive radios," *IEEE Trans. Signal Process.*, vol. 56, no. 5, pp. 1801–18011, May 2008.
[33] T. Hwang, C. Yang, G. Wu, S. Li, and G. Y. Li, "OFDM and its wireless applications: A survey," *IEEE Trans. Veh. Technol.*, vol. 58, no. 4, pp. 1673–1694, May 2009.
[34] B. Farhang-Boroujeny, "OFDM versus filter bank multicarrier," *IEEE Signal Process. Mag.*, vol. 28, no. 3, pp. 92–112, May 2011.
[35] V. P. Ipatov, *Spread Specturm and CDMA: Principles and Applications*. New York, NY, USA: Wiley, 2005.
[36] A. Papoulis and S. U. Pillai, *Probability, Random Variables and Stochastic Processes*, 4th ed. New York, NY, USA: McGraw-Hill, 2002.
[37] R. M. Gray, *Entropy and Information Theory*, 2nd ed. New York, NY, USA: Springer-Verlag, 2011.
[38] Q. Guo and D. Huang, "EM-based joint channel estimation and detection for frequency selective channels using Gaussian message passing," *IEEE Trans. Signal Process.*, vol. 59, no. 8, pp. 4030–4035, Aug. 2011.
[39] T. K. Moon, "The expectation–maximization algorithm," *IEEE Signal Process. Mag.*, vol. 13, no. 6, pp. 45–59, Nov. 1996.
[40] K. P. Murphy, *Machine Learning*. Cambridge, MA, USA: MIT Press, 2012.
[41] H. V. Poor, *An Introduction to Signal Detection and Estimation*, 2nd ed. New York, NY, USA: Springer-Verlag, 1994.
[42] S. Arkoulis, E. Anifantis, V. Karyotis, S. Papavassiliou, and N. Mitrou, "Discovering and exploiting spectrum power correlations in cognitive radio networks: An experimentally driven approach," *EURASIP J. Wireless Commun. Netw.*, vol. 2014, no. 17, pp. 1–14, Jan. 2014.
[43] K. Hossain, B. Champagne, and A. Assra, "Cooperative multiband joint detection with correlated spectral occupancy in cognitive radio networks," *IEEE Trans. Signal Process.*, vol. 60, no. 5, pp. 2682–2687, May 2012.
[44] J. Sherman and W. J. Morrison, "Adjustment of an inverse matrix corresponding to a change in one element of a given matrix," *Ann. Math. Statist.*, vol. 21, no. 1, pp. 124–127, Mar. 1950.
[45] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Royal Statist. Soc., Series B (Methodol.)*, vol. 39, no. 1, pp. 1–38, Mar. 1977.
[46] A. Kocian and B. H. Fleury, "EM-based joint data detection and channel estimation of DS-CDMA signals," *IEEE Trans. Commun.*, vol. 51, no. 10, pp. 1709–1720, Oct. 2003.
[47] A. Pascual-Iserte, D. P. Palomar, A. I. Perez-Neira, and M. A. Lagunas, "A robust maximin approach for MIMO communications with imperfect channel state information based on convex optimization," *IEEE Trans. Signal Process.*, vol. 54, no. 1, pp. 346–360, Jan. 2006.
[48] E. A. Gharavol, Y.-C. Liang, and K. Mouthaan, "Robust downlink beamforming in multiuser MISO cognitive radio networks with imperfect channel-state information," *IEEE Trans. Veh. Technol.*, vol. 59, no. 6, pp. 2852–2860, May 2010.
[49] Y. Zhang, E. Dall'Anese, and G. B. Giannakis, "Distributed optimal beamformer for cognitive radios robust to channel uncertainties," *IEEE Trans. Signal Process.*, vol. 60, no. 12, pp. 6495–6508, Dec. 2012.
[50] D. A. Harville, *Matrix Algebra From a Statistician's Perspective*. New York, NY, USA: Springer-Verlag, 1997.
[51] N. M. Temme, "Asymptotic inversion of incomplete gamma functions," *Math. Comput.*, vol. 58, no. 198, pp. 755–764, Apr. 1992.
[52] B. C. Levy, *Principles of Signal Detection and Parameter Estimation*. New York, NY, USA: Springer-Verlag, 2008.

Ayman Assra (M'15) received the B.Sc. degree in electrical engineering from Alexandria University, Alexandria, Egypt, in 2000; the M.Sc. degree from the Arab Academy for Science and Technology and Maritime Transport, Alexandria, in 2004; and the Ph.D. degree from Concordia University, Montreal, QC, Canada, in 2010.

He was a Postdoctoral Research Fellow with Institut National de Recherche sur les Transports et leurs Sécurité (INRETS), Lille, France; McGill University, Montreal, QC, Canada; Communications Research Centre, Ottawa, ON, Canada; and Alexander von Humboldt Foundation, Bonn, Germany. Since 2014, he has been a Technical Engineering Systems Architect for Broadband Networks with TechTalent Consulting and Team-Building Inc., Calabasass, CA, USA. His current research interests include space–time processing, multiple-input–multiple-output (MIMO) communications, cognitive radios, massive MIMO, and Long-Term Evolution unlicensed.

Dr. Assra received the Best Paper Award at the 2012 International Symposium on Wireless Communication Systems and at the 2013 International Conference on Wireless and Mobile Communications.

Jiaxin Yang (S'11) received the B.Eng. degree in information engineering from Shanghai Jiao Tong University, Shanghai, China, in 2009 and the M.E.Sc. degree in electrical and computer engineering from The University of Western Ontario, London, ON, Canada, in 2011. He is currently working toward the Ph.D. degree with the Department of Electrical and Computer Engineering, McGill University, Montreal, QC, Canada.

His research interests include optimization theory; statistical signal processing, detection, and estimation; and applications thereof in wireless communications such as multiple-input–multiple-output systems, cooperative communications, and physical-layer security.

Mr. Yang has received several awards and scholarships, including the McGill Engineering Doctoral Award, the Graduate Excellence Fellowship, the FRQNT International Internship Scholarship, and the PERSWADE Ph.D. Scholarship.

Benoit Champagne (SM'03) received the B.Ing. degree in engineering physics from the École Polytechnique de Montréal, Montréal, QC, Canada, in 1983; the M.Sc. degree in physics from the Université de Montréal in 1985; and the Ph.D. degree in electrical engineering from the University of Toronto, Toronto, ON, Canada, in 1990.

From 1990 to 1999, he was an Assistant Professor and then an Associate Professor with INRS-Telecommunications, Université du Quebec, Montréal. Since 1999, he has been with McGill University, Montréal, where he is currently a Full Professor with the Department of Electrical and Computer Engineering. From 2004 to 2007, he served as an Associate Chair of Graduate Studies with the Department of Electrical and Computer Engineering, McGill University. He is the author or coauthor of more than 200 referred publications. His research has been funded by the Natural Sciences and Engineering Research Council of Canada, by the "Fonds de Recherche sur la Nature et les Technologies" from the Government of Quebec, and by some major industrial sponsors, including Nortel Networks, Bell Canada, InterDigital, and Microsemi. His research interests include advanced algorithms for the processing of information bearing signals by digital means. His interests span many areas of statistical signal processing, including detection and estimation, sensor array processing, adaptive filtering, and applications thereof to broadband communications and audio processing.

Dr. Champagne was the Cochair, Wide Area Cellular Communications Track, for the IEEE International Symposium on Personal, Indoor, and Mobile Radio Communications in September 2011, held in Toronto; the Cochair, Antenna and Propagation Track, for the IEEE Vehicular Technology Conference (VTC Fall) in September 2004, held in Los Angeles, CA, USA; and the Registration Chair for the IEEE International Conference on Application-Specific Systems, Architecture, and Processors in May 2004, held in Montréal. He has also served on the Technical Committees of several international conferences in the fields of communications and signal processing. He has served as an Associate Editor for the IEEE SIGNAL PROCESSING LETTERS, the IEEE TRANSACTIONS ON SIGNAL PROCESSING, and the *EURASIP Journal on Applied Signal Processing*.