Research Article

Rate-Optimal and Reduced-Complexity Sequential Sensing Algorithms for Cognitive OFDM Radios

Seung-Jun Kim and Georgios B. Giannakis

Department of Electrical and Computer Engineering, University of Minnesota, 200 Union Street SE, Minneapolis, MN 55455, USA

Correspondence should be addressed to Georgios B. Giannakis, georgios@umn.edu

Sequential sensing algorithms are developed for OFDM-based hierarchical cognitive radio (CR) systems. Secondary users sense multiple subbands simultaneously for possible spectrum availabilities under hard misdetection constraints to prevent interference to the primary users. Accounting for the fact that the sensing time overhead can often be significant, a novel performance metric is introduced based on the effective achievable data rate. An optimization problem is formulated in the framework of optimal stopping problems to maximize the average effective data rate by determining the best time to stop taking samples and proceed to data transmission. A basis expansion-based suboptimal algorithm is developed to reduce the prohibitive complexity of the optimal solution. The numerical results presented verify the efficacy of the proposed approach.

1. Introduction

The cognitive radio (CR) paradigm aims to design intelligent radios that can sense the environment and adapt the transceiver parameters as well as the resource allocation decisions in order to exploit the spectrum availability aggressively [1]. The motivation stems from the observation that the current usage pattern of the licensed spectrum is not very efficient; there are plenty of "white spaces" in spectral and/or spatiotemporal domains that allow overlaying secondary transmissions on top of the existing licensed (a.k.a. primary) users without degrading the communication quality of the latter [2].

To exploit the spectrum holes opportunistically, orthogonal frequency-division multiplexing (OFDM) transceivers are often employed at the physical layer of the CR due to their flexibility in communicating over a wide range of spectrum bands efficiently [3]. OFDM radios can easily implement transmission filters that suppress the signal in the undesired subbands to prevent interference to the primary users (PUs). CR transmissions target only the bands in which the PUs are not present. In the time domain, the "quiet" periods of the PUs are identified so as to interleave the CR data in-between the PU transmissions [4].

Obviously, a key component of CR transceivers is the sensing module that monitors spectrum occupancy of the PUs in real time. Since the presence of CR links must be oblivious to the PUs, hard misdetection constraints need be imposed to the design of the detector in the sensing module. However, this inevitably leads to increased sensing time, which, in turn, leaves less time for the actual data transmission before the PUs may kick back in. Thus, it is important to factor in the sensing overhead in the design of the sensing module, especially for scenarios where the PU occupancy changes dynamically in time as well as in frequency.

However, the sensing task is often challenging. First of all, as the licensees may have invested heavily on the spectrum to run commercialized services in their bands, it is likely that very strict interference-preventing masks are placed to the CRs. In IEEE 802.22, for example, the sensing threshold for the digital TV (DTV) signal is set as low as -116 dBm over 6 MHz bandwidth [5]. Secondly, in some situations, coherent detectors such as matched filters cannot be used because the required a priori knowledge of the PU signal characteristics is not available or simply because hardware complexity is to be kept minimal. Thus, the sensing module often needs to employ noncoherent receivers with reduced complexity but

also performance such as energy detectors [6, 7] or feature-based detectors [8].

The wideband sensing problem for OFDM-based CRs was considered in [9], where the detection thresholds for a bank of energy detectors were optimized jointly to maximize throughput performance. To improve the sensing performance, cooperative sensing scenarios were also considered in [10] by collecting measurements from multiple CRs at the fusion center (FC) to improve detection performance. A cooperative sensing strategy was also pursued in [11], where a linear-quadratic fusion rule was developed based on the deflection criterion to process correlated observations. The tradeoff between sensing duration and throughput was studied in [12] using the energy detectors. These developments, however, assume batch (or fixed sample size (FSS)) detection strategies, where the number of samples collected for detection is a predetermined design parameter, which does not depend on the actual values of the received samples.

Sequential detection schemes on the other hand exploit the fact that the number of samples required to achieve a given reliability level may well be dependent on the actual realization of the observed samples. For example, in a simple binary hypothesis testing context, Wald's sequential probability ratio test (SPRT) compares the likelihood ratio with two thresholds, and the decision is made as soon as the test statistic exceeds either one of the thresholds. It is known that SPRT minimizes the average sample number (ASN) among all tests with the same false alarm and misdetection probabilities [13, page 21]. However, it is not clear how to apply the SPRT approach to the wideband sensing problem, where a bank of detectors must be run simultaneously. Moreover, the relevant optimization criterion in this case might not be as simple as the ASN. In [14], two layers of SPRTs were employed at individual CRs and the FC to reduce the overall detection delay for a single-channel sensing problem. However, no claims on optimality were provided.

In this work, rate-optimal wideband sequential sensing algorithms are developed in the framework of optimal stopping time problems. Such problems amount to determining the time to stop taking sequential observations so that an expected value of payoff based on the accumulated observations is maximized [15, 16]. The payoff in our CR sensing setup will be chosen as the total rate achieved by using all the available subchannels, where the availability is determined under hard interference constraints. The sensing overhead is captured by explicitly accounting for the sensing time, which consumes portion of the frame duration.

In a companion paper [17], generalizations to sequential *cooperative* sensing are also discussed, where a central processing unit sequentially collects either raw (analog) or quantized observations from multiple cooperating CRs. However, the underlying problem formulation is different in [17] in that the detector structure is optimally determined in closed form as a likelihood ratio test, whereas the conventional energy detector structure is adopted here. Also, a *recursive* (on-line) counterpart of the *batch* training algorithm developed here is presented in [17].

The rest of the paper is organized as follows. In Section 2, the signal model and the problem formulation are presented for the single-CR case. In Section 3, the optimal solution is derived and a reduced-dimension sufficient statistic is identified. Section 4 develops a basis expansion-based reduced-complexity algorithm to obtain a suboptimal yet tractable solution. Numerical results are presented in Section 5, and conclusions are provided in Section 6.

2. Problem Statement

2.1. Signal Model. Consider a CR that shares M orthogonal bands opportunistically with PUs in its network. In order not to interfere with on-going PU transmissions, the CR must identify the bands that are not occupied by the PUs before transmitting its own data.

The nth received signal sample at the CR on band $m \in \{1, 2, \ldots, M\}$, when a PU is transmitting on that band, can be modeled by

$$r_n^{(m)} = h_n^{(m)} s_n^{(m)} + z_n^{(m)}, \quad n \in \{1, 2, \ldots, N\}, \qquad (1)$$

where $\{h_n^{(m)}\}$ are the channel coefficients, $\{s_n^{(m)}\}$ are the PU signal samples, $\{z_n^{(m)}\}$ are independent and identically distributed (i.i.d.) complex Gaussian additive noise samples with mean 0 and variance σ^2, and N denotes the maximum number of samples that can be collected in the sensing phase. Under the assumption that PU occupancy is independent across the M bands, the CR must perform a binary hypothesis test per band m to discriminate the following two hypotheses:

$$\begin{aligned} H_0^{(m)} &: r_n^{(m)} = z_n^{(m)}, \quad n \in \{1, \ldots, N\}, \\ H_1^{(m)} &: r_n^{(m)} = h_n^{(m)} s_n^{(m)} + z_n^{(m)}, \quad n \in \{1, \ldots, N\}. \end{aligned} \qquad (2)$$

The prior probabilities $\Pr\{H_0^{(m)}\}$ and $\Pr\{H_1^{(m)}\}$ are denoted by $p_0^{(m)}$ and $1 - p_0^{(m)}$, respectively.

As in, for example, [6], each CR receiver relies on energy detection to decide the occupancy of each band. The test statistic $t_n^{(m)}$ is calculated at each time step n as

$$t_n^{(m)} = \sum_{k=1}^{n} y_k^{(m)}, \qquad y_n^{(m)} \triangleq \left| r_n^{(m)} \right|^2. \qquad (3)$$

For simplicity, assume that the channel coefficients $\{h_n^{(m)}\}$ do not vary over the detection interval such that $|h_n^{(m)}|^2$ can be denoted as $G^{(m)}$, $\forall n \in \{1, \ldots, N\}$, and the transmitted signal has unit power; that is, $E\{|s_n^{(m)}|^2\} = 1$, $\forall n, m$. The CR receiver in the scenarios considered can acquire its channel with the PU blindly, or, by overhearing the pilot signal transmitted by the PU system. This is clearly feasible when the transceivers are stationary so that the channel is quasistatic; see also [9] for a justifying argument in the context of DTV systems. In addition, it is possible to extend the ensuing formulation and algorithms to the case where only the distribution of $G^{(m)}$ is known by considering appropriate averages over those channel gains; however, this

case goes beyond the scope of the present paper and will not be detailed here.

Then, the observations $\{y_1^{(m)}, y_2^{(m)}, \ldots, y_N^{(m)}\}$ are i.i.d. under $H_0^{(m)}$ and $H_1^{(m)}$ with the conditional univariate densities under the two hypotheses for all m given by the (non)central χ^2 densities [18, Section 2.2]:

$$p\left(y \mid H_0^{(m)}\right) = \frac{1}{\sigma^2} e^{-y/\sigma^2} \mathbb{1}_{\{y \geq 0\}}, \qquad (4)$$

$$p\left(y \mid H_1^{(m)}\right) = \frac{1}{\sigma^2} e^{-(y+G^{(m)})/\sigma^2} I_0\left(\frac{\sqrt{G^{(m)} y}}{\sigma^2/2}\right), \qquad (5)$$

where $I_0(\cdot)$ denotes the zeroth-order-modified Bessel function of the first kind, and $\mathbb{1}_{\{\cdot\}}$ the indicator function that equals 1 if the condition $\{\cdot\}$ is true and 0 otherwise.

In the batch Neyman-Pearson framework, the decision with maximum detection probability under a given false-alarm rate is based on the threshold rule

$$\text{Decide } H_1^{(m)} \text{ if } t_n^{(m)} > \gamma_n^{(m)},$$
$$\text{Decide } H_0^{(m)} \text{ if } t_n^{(m)} < \gamma_n^{(m)} \qquad (6)$$

due to the monotonicity of the log-likelihood ratio, where $\gamma_n^{(m)}$ is the threshold at time n for band m, and the decision can be either way when $t_n^{(m)} = \gamma_n^{(m)}$. Invoking the central limit theorem, one can show that for large enough n the probability of false alarms $\alpha_n^{(m)}$ and the probability of misdetection $\beta_n^{(m)}$ for band m with sample size n are given by [19]

$$\alpha_n^{(m)} = \Pr\left\{t_n^{(m)} > \gamma_n^{(m)} \mid H_0^{(m)}\right\} = Q\left(\frac{\gamma_n^{(m)} - n\sigma^2}{\sigma^2 \sqrt{n}}\right), \qquad (7)$$

$$\beta_n^{(m)} = \Pr\left\{t_n^{(m)} < \gamma_n^{(m)} \mid H_1^{(m)}\right\}$$
$$= 1 - Q\left(\frac{\gamma_n^{(m)} - n\left(\sigma^2 + G^{(m)}\right)}{\sigma\sqrt{n(\sigma^2 + 2G^{(m)})}}\right), \qquad (8)$$

where $Q(\cdot)$ denotes the Gaussian tail function.

In the context of spectrum sensing for CRs, the misdetection probabilities signify the probabilities of failing to detect the presence of the PUs, which could lead to causing interference to the PU transmission. For this reason, the sensing algorithms must be designed to guarantee very small misdetection probabilities. On the other hand, small false alarm probabilities are desired to increase the usage of the available channels. These dual goals can be accomplished by increasing the sample size n. However, increasing the sample size leads to larger sensing overhead, which effectively reduces the time left for actual data transmission. In the next subsection, a sequential sensing problem is formulated to optimize the overall transmission rate by taking into account the overhead due to sensing.

2.2. Sequential CR Sensing as an Optimal Stopping Problem.

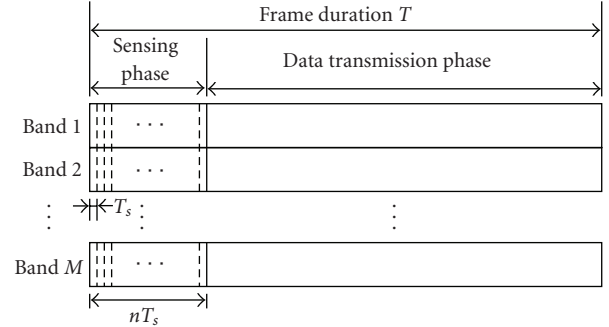The sequential sensing problem can be formulated in the

FIGURE 1: Frame structure of the CR.

framework of optimal stopping problems [20, Section 4.4], [15, 16, 21]. Based on the observations collected up to a certain point, the optimal stopping problems seek to find the "best" time to stop taking observations, "best" in the sense of maximizing the expected value of a chosen reward. An optimal stopping problem can thus be defined by specifying (i) a sequence of observed random variables with their joint distribution and (ii) a sequence of random variables, whose joint distributions with the observations are known, each representing the per-step reward that could be obtained should one decide to stop at the corresponding time step. In our setup, the sequence of i.i.d. observations is $\{\mathbf{y}_n\}_{n=1}^N$, whose joint p.d.f. is expressible in terms of the univariate p.d.f. in (4) or (5). In the ensuing subsections, the sequential CR sensing problem is posed as an optimal stopping problem by defining the reward sequence, which will be finally given in (27). But first it is prudent to specify the optimality criterion.

2.2.1. Average Throughput Criterion.

To capture the effective throughput of the CR while accounting for the sensing overhead, consider the frame structure shown in Figure 1. The deterministic frame duration T is divided into the sensing and the data transmission phases, where the durations of both phases are random variables. Denote the sampling interval of the detector by T_s, where $T \geq NT_s$. Let $\mathbf{y}_n \triangleq [y_n^{(1)} \quad y_n^{(2)} \quad \cdots \quad y_n^{(M)}]^T$ and likewise for the corresponding vector $\mathbf{t}_n \triangleq [t_n^{(1)} \quad \cdots \quad t_n^{(M)}]^T$, collecting the test statistics defined in (3). The PU occupancy over the M bands is denoted by an $M \times 1$ random vector $\mathcal{H}$ with the mth entry $\mathcal{H}^{(m)}$ taking values from $\{H_0^{(m)}, H_1^{(m)}\}$ and the rate that can be achieved in band m by $R^{(m)}$.

If the CR stops sensing after the nth sampling interval, the duration of the sensing phase is nT_s. The CR then proceeds to data transmission on the bands that are sensed idle for the remaining part of the frame of duration $T - nT_s$. Therefore, the overall throughput is given by

$$f_n'(\mathcal{H}, \mathbf{t}_n)$$
$$= \frac{T - nT_s}{T} \sum_{m=1}^M R^{(m)} \mathbb{1}_{\{H_0^{(m)}\}} \mathbb{1}_{\{t_n^{(m)} < \gamma_n^{(m)}\}}, \quad n = 1, 2, \ldots, N, \qquad (9)$$

where the first indicator function in (9) allows rate to be transmitted over band m only if $\mathcal{H}^{(m)}$ is indeed $H_0^{(m)}$; while the second indicator function allows usage of band m if also the energy detector correctly declares $\mathcal{H}^{(m)} = H_0^{(m)}$ (cf. (6)).

Note that the per-step reward $f_n'(\cdot)$ is a function of the underlying true spectrum occupancy $\mathcal{H}$, which is not directly observable. It can be shown that the optimal stopping problem based on the per-step reward $f_n'(\cdot)$ is equivalent to the one based on the conditional average per-step reward $E\{f_n'(\mathcal{H}, \mathbf{t}_n) \mid \mathbf{Y}_n\}$, $n = 1, 2, \ldots, N$, where $\mathbf{Y}_n \triangleq [\mathbf{y}_1 \cdots \mathbf{y}_n]$ and the expectation is taken over $\mathcal{H}$ [21], [20, Section 5.1]. Expected values of binary random variables corresponding to the indicator functions can be replaced by probabilities; for example, $E\{\mathbb{1}_{\{H_0^{(m)}\}} \mid \mathbf{Y}_n\} = \Pr\{H_0^{(m)} \mid y_1^{(m)}, \ldots, y_n^{(m)}\}$. Hence, the conditional average per-step reward can be expressed as (cf. (9))

$$f_n(\boldsymbol{\pi}_n, \mathbf{t}_n) \triangleq E\{f_n'(\mathcal{H}, \mathbf{t}_n) \mid \mathbf{Y}_n\}$$
$$= \frac{T - nT_s}{T} \sum_{m=1}^{M} R^{(m)} \pi_n^{(m)} \mathbb{1}_{\{t_n^{(m)} < \gamma_n^{(m)}\}}, \quad (10)$$

where $\boldsymbol{\pi}_n \triangleq [\pi_n^{(1)} \cdots \pi_n^{(M)}]^T$ denotes the belief vector with entries

$$\pi_n^{(m)} \triangleq \Pr\{H_0^{(m)} \mid y_1^{(m)}, \ldots, y_n^{(m)}\}. \quad (11)$$

Clearly, the factor $(T - nT_s)/T$ in the per-step rewards in both (9) and (10) diminishes as n approaches N, which encourages stopping as soon as possible, while collecting more samples in $\mathbf{t}_n$ can potentially identify more available bands, as captured by the indicator functions $\mathbb{1}_{\{t_n^{(m)} < \gamma_n^{(m)}\}}$, leading to a tradeoff.

The goal is to obtain the average (over $\mathcal{H}$ and $\mathbf{Y}_N$) throughput-optimal stopping policy that determines whether to stop sensing or not at each time $n \in \{1, 2, \ldots, N - 1\}$ given the observations $\mathbf{Y}_n$. To be precise, let $u_n \in \{\text{stop}, \text{continue}\}$ denote a control variable at time n. The stopping policy $\mathcal{U}_n(\cdot)$ per time n is the rule specifying the mapping $\mathbf{Y}_n \in (\mathbb{R}^+)^{nM} \rightsquigarrow \mathcal{U}_n(\mathbf{Y}_n) \in \{\text{stop}, \text{continue}\}$. With the introduction of an auxiliary state variable $x_n \in \{S, \overline{S}\}$, where $x_n = S$ indicates the "stop" state and $x_n = \overline{S}$ the "nonstop" state, the system evolution is characterized by

$$x_{n+1} = \begin{cases} S, & \text{if } x_n = S, \text{ or } x_n \neq S, \ u_n = \text{stop}, \\ \overline{S}, & \text{otherwise}, \end{cases} \quad (12)$$
$$n = 1, 2, \ldots, N - 1,$$

where $x_1 = \overline{S}$. The reward function at the final stage N is given by

$$\widehat{f}_N(\mathbf{Y}_N, x_N) = f_N(\boldsymbol{\pi}_N, \mathbf{t}_N) \mathbb{1}_{\{x_N \neq S\}} \quad (13)$$

and at stage n by

$$\widehat{f}_n(\mathbf{Y}_n, x_n, u_n)$$
$$= f_n(\boldsymbol{\pi}_n, \mathbf{t}_n) \mathbb{1}_{\{x_n \neq S, u_n = \text{stop}\}}, \quad n = 1, 2, \ldots, N - 1. \quad (14)$$

Note that the reward at stage N does not include the control variable because $\mathcal{U}_N \equiv \text{stop}$ since sensing must end a fortiori after N sampling intervals. The "bookkeeping" variable x_n ensures that the reward $\widehat{f}_n(\cdot)$ may be nonzero only upon the *first* instance of u_n being equal to "stop"; for the rest of the time steps, $\widehat{f}_n(\cdot)$ evaluates to zero due to the indicator functions in (13) and (14). (As a mnemonic for the notation used, functionals with hat (e.g., $\widehat{f}_n$) depend on the control (u_n) and the "bookkeeping" variables (x_n) as well as the rewards (e.g., f_n), which are functions of the data $\mathbf{Y}_n$ through $\boldsymbol{\pi}_n$ and $\mathbf{t}_n$.) Expressions (13) and (14) allow the average throughput to be expressed as the *sum* of the per-step rewards

$$\overline{\mathcal{R}}(\mathcal{U}_1, \ldots, \mathcal{U}_{N-1})$$
$$\triangleq E\left\{ \widehat{f}_N(\mathbf{Y}_N, x_N) + \sum_{n=1}^{N-1} \widehat{f}_n(\mathbf{Y}_n, x_n, \mathcal{U}_n(\mathbf{Y}_n)) \right\}, \quad (15)$$

where the expectation is with respect to $\mathbf{Y}_N$.

The average throughput in (15) is to be maximized over the finite horizon comprising N sampling intervals, with respect to the stopping policy $\{\mathcal{U}_1, \ldots, \mathcal{U}_{N-1}\}$ with $\mathcal{U}_n(\mathbf{Y}_n) = u_n \in \{\text{stop}, \text{continue}\}$. Certainly, to perform this functional (i.e., variational) optimization, it is necessary to know the data p.d.f. required to evaluate the expectation in (15). The reason behind expressing $\overline{\mathcal{R}}$ in (15) as a cumulative sum of the per-step rewards in (13) and (14) is our intention of casting the sequential CR sensing task as a dynamic programming (DP) problem, the subject discussed next.

2.2.2. Constrained Dynamic Programming Formulation.

Given the observations $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n$ per time n, the CR performs hypothesis testing in each band to determine whether the band is occupied or not. When the sensing is stopped, the CR transmits on those channels that the detector determines to be unoccupied. However, it is critical to constrain the "collision" probability $P_c^{(m)}$ for each channel m, which is the probability that the CR interferes with the PU transmission due to misdetection. It is thus desired for the DP formulation to be introduced later to enforce the constraints

$$P_c^{(m)} \leq \overline{\beta}, \quad m = 1, 2, \ldots, M, \quad (16)$$

where $\overline{\beta}$ is a small positive constant.

For *nonsequential* FSS tests with sample size n, the "collision" probability coincides with the misdetection probability $\beta_n^{(m)}$ in (8). Thus, (16) is equivalent to $\beta_n^{(m)} \leq \overline{\beta}$ for all m, and the threshold $\gamma_n^{(m)}$ that maximizes the probability of false alarm can be calculated from (8) as

$$\gamma_n^{(m)} = Q^{-1}\left(1 - \overline{\beta}\right) \sigma \sqrt{n(\sigma^2 + 2G^{(m)})} + n\left(G^{(m)} + \sigma^2\right). \quad (17)$$

Interestingly, in the case of *sequential* sensing, constraining $\beta_n^{(m)}$ is not equivalent to constraining the "collision" probabilities. To establish this, we recall that sensing can

terminate at any time $n \in \{1, \ldots, N\}$. Hence, $P_c^{(m)}$ can be expressed as

$$P_c^{(m)} = \sum_{n=1}^{N} \Pr\left\{ x_n \neq S, \, \mathcal{U}_n = \text{stop}, \, t_n^{(m)} < \gamma_n^{(m)} \mid H_1^{(m)} \right\}, \quad (18)$$

where $\mathcal{U}_N \equiv \text{stop}$ for the fixed horizon N considered here. Using Bayes' rule, (18) can be rewritten in terms of $\beta_n^{(m)}$ as

$$P_c^{(m)} = \sum_{n=1}^{N} \beta_n^{(m)} \Pr\left\{ x_n \neq S, \, \mathcal{U}_n = \text{stop} \mid t_n^{(m)} < \gamma_n^{(m)}, H_1^{(m)} \right\}. \quad (19)$$

Thus, if the events $\{x_n \neq S, \mathcal{U}_n = \text{stop}\}$ and $\{t_n^{(m)} < \gamma_n^{(m)}\}$ were independent conditioned on $H_1^{(m)}$ for all n, then imposing the condition $\beta_n^{(m)} \leq \bar{\beta}$ would be sufficient to guarantee $P_c^{(m)} \leq \bar{\beta}$. Since this is not the case in general, capping the misdetection probability under $\bar{\beta}$ does not necessarily ensure $P_c^{(m)} \leq \bar{\beta}$.

Ideally, the optimal $\{\gamma_n^{(m)}\}$ should be obtained by formulating a constrained DP problem that maximizes (15) under the constraints (16), after incorporating $\{\gamma_n^{(m)}\}$ as optimization variables in addition to $\{\mathcal{U}_n\}$. The resulting DP is significantly more complex to solve because the feasible space for $\{\gamma_n^{(m)}\}$ is infinite and discretization needs to be employed for tractability. For this reason, we adopt a less ideal yet intuitive rule for determining the thresholds $\gamma_n^{(m)}$, namely using the thresholds in (17) obtained from the FSS tests. The "collision" probability will still be guaranteed by explicitly imposing the conditions (16) in the constrained DP formulation described next.

First, note that the collision probability on band m given in (18) can be rewritten as

$$P_c^{(m)} = \sum_{n=1}^{N} E\left\{ \mathbb{1}_{\{x_n \neq S, \mathcal{U}_n = \text{stop}, t_n^{(m)} < \gamma_n^{(m)}\}} \mid H_1^{(m)} \right\}$$
$$= E\left\{ \sum_{n=1}^{N} \mathbb{1}_{\{x_n \neq S, \mathcal{U}_n = \text{stop}, t_n^{(m)} < \gamma_n^{(m)}\}} \frac{1 - \pi_n^{(m)}}{1 - p_0^{(m)}} \right\}, \quad (20)$$

where in deriving the second equality we used Bayes' rule and the facts that $\Pr\{H_1^{(m)} \mid y_1^{(m)}, \ldots, y_n^{(m)}\} = 1 - \pi_n^{(m)}$ and $\Pr\{H_1^{(m)}\} = 1 - p_0^{(m)}$. To render (20) compatible with the constrained DP formalism, define per-step costs for channel $m \in \{1, 2, \ldots, M\}$ as

$$\hat{c}_N^{(m)}(\mathbf{Y}_N, x_N) \triangleq \mathbb{1}_{\{x_N \neq S\}} c_N^{(m)}\left(\pi_N^{(m)}, t_N^{(m)} \right),$$

$$\hat{c}_n^{(m)}(\mathbf{Y}_n, x_n, u_n) \triangleq \mathbb{1}_{\{x_n \neq S, u_n = \text{stop}\}} c_n^{(m)}\left(\pi_n^{(m)}, t_n^{(m)} \right), \quad (21)$$
$$n = 1, 2, \ldots, N - 1,$$

where

$$c_n^{(m)}\left(\pi_n^{(m)}, t_n^{(m)} \right)$$
$$\triangleq \mathbb{1}_{\{t_n^{(m)} < \gamma_n^{(m)}\}} \frac{1 - \pi_n^{(m)}}{1 - p_0^{(m)}}, \quad n = 1, 2, \ldots, N. \quad (22)$$

Then, the desired optimization problem can be formulated as a constrained DP problem given by (cf. (15) and (20)–(22))

$$\max_{\{\mathcal{U}_n(\mathbf{Y}_n)\}_{n=1}^{N-1}} E\left\{ \hat{f}_N(\mathbf{Y}_N, x_N) + \sum_{n=1}^{N-1} \hat{f}_n(\mathbf{Y}_N, x_n, \mathcal{U}_n(\mathbf{Y}_n)) \right\} \quad (23)$$

$$\text{subject to } E\left\{ \hat{c}_N^{(m)}(\mathbf{Y}_N, x_N) + \sum_{n=1}^{N-1} \hat{c}_n^{(m)}(\mathbf{Y}_n, x_n, \mathcal{U}_n(\mathbf{Y}_n)) \right\}$$
$$\leq \bar{\beta}, \quad m = 1, 2, \ldots, M. \quad (24)$$

One approach to solving *constrained* DP problems is through Lagrange relaxation [22]. Given a set of Lagrange multipliers $\boldsymbol{\lambda} \triangleq [\lambda^{(1)} \quad \lambda^{(2)} \quad \cdots \quad \lambda^{(M)}]^T$ with $\lambda^{(m)} \geq 0$, $m \in \{1, 2, \ldots, M\}$, the relaxed *unconstrained* DP problem to solve is

$$\max_{\{\mathcal{U}_n(\mathbf{Y}_n)\}_{n=1}^{N-1}} E\left\{ \hat{g}_N(\mathbf{Y}_N, x_N; \boldsymbol{\lambda}) + \sum_{n=1}^{N-1} \hat{g}_n(\mathbf{Y}_n, x_n, \mathcal{U}_n(\mathbf{Y}_n); \boldsymbol{\lambda}) \right\}, \quad (25)$$

where

$$\hat{g}_N(\mathbf{Y}_N, x_N; \boldsymbol{\lambda}) \triangleq \hat{f}_N(\mathbf{Y}_N, x_N) - \sum_{m=1}^{M} \lambda^{(m)} \hat{c}_N^{(m)}(\mathbf{Y}_N, x_N),$$

$$\hat{g}_n(\mathbf{Y}_n, x_n, u_n; \boldsymbol{\lambda}) \triangleq \hat{f}_n(\mathbf{Y}_n, x_n, u_n) - \sum_{m=1}^{M} \lambda^{(m)} \hat{c}_n^{(m)}(\mathbf{Y}_n, x_n, u_n),$$
$$n = 1, 2, \ldots, N - 1. \quad (26)$$

Note that this is an optimal stopping problem with observation sequence $\{\mathbf{y}_n\}_{n=1}^{N}$ and reward sequence

$$g_n(\boldsymbol{\pi}_n, \mathbf{t}_n; \boldsymbol{\lambda})$$
$$\triangleq f_n(\boldsymbol{\pi}_n, \mathbf{t}_n) - \sum_{m=1}^{M} \lambda^{(m)} c_n^{(m)}\left(\pi_n^{(m)}, t_n^{(m)} \right), \quad n = 1, 2, \ldots, N. \quad (27)$$

The approaches to solve the problems (25) and (26) for a given $\boldsymbol{\lambda}$ are described in Sections 3 and 4.

The Lagrange multipliers $\boldsymbol{\lambda}$ need to be updated to satisfy the constraints in (24). A simple approach is to use the subgradient method (see, e.g., [23, Chapter 2])

$$\lambda_{\ell+1}^{(m)} = \max\left\{ 0, \lambda_\ell^{(m)} + \mu_\ell \left(P_c^{(m)} - \bar{\beta} \right) \right\}, \quad m = 1, 2, \ldots, M, \quad (28)$$

where ℓ denotes the iteration index, and μ_ℓ the step size. The subgradient update increases the Lagrange multipliers when the constraints are violated to effect larger penalty in the objective function of (25), so that the constraints are enforced. A remark on the signal model is in order.

Remark 1. Recall that the PU occupancy was assumed independent across the M bands. In practice, the resource allocation decision of the PU system may render the spectrum occupancy of the PU transmissions *correlated* across bands. In principle, it is possible to cope with correlated PU spectrum occupancy. Consider, for example, $M = 2$ channels, where the possible occupancy states are

$$\left\{ \left(H_0^{(1)}, H_0^{(2)}\right), \left(H_1^{(1)}, H_0^{(2)}\right), \left(H_0^{(1)}, H_1^{(2)}\right), \left(H_1^{(1)}, H_1^{(2)}\right) \right\} \tag{29}$$

with an associated prior for each state. By exploiting the correlation across subchannels, the sensing performance can be improved. However, the number of states and consequently the dimension of the belief vector increase exponentially in M, rendering the complexity prohibitive. This explains why we adopted the simple assumption on the spectrum occupancy being independent across the subbands.

3. Optimal Solution

In principle, the DP problem with finite horizon given in (25) and (26) can be solved optimally via the method of backward induction [20]. It can be easily verified that the backward induction for this dynamic program leads to the backward induction of the optimal stopping problems detailed next; see also [15]. To this end, define the value function $V_n(\cdot)$ recursively as follows:

$$V_N(\mathbf{Y}_N; \boldsymbol{\lambda}) = g_N(\boldsymbol{\pi}_N, \mathbf{t}_N; \boldsymbol{\lambda}), \tag{30}$$

$$V_n(\mathbf{Y}_n; \boldsymbol{\lambda}) = \max\{g_n(\boldsymbol{\pi}_n, \mathbf{t}_n; \boldsymbol{\lambda}), E\{V_{n+1}(\mathbf{Y}_n, \mathbf{y}_{n+1}; \boldsymbol{\lambda}) \mid \mathbf{Y}_n\}\}$$
$$\text{for } n = N - 1, \ldots, 1, 0, \tag{31}$$

where $f_0 \equiv 0$ is used for g_0 in (27). This recursive definition specifies offline functions $\{V_n(\cdot)\}_{n=0}^{N-1}$, which will be used in the optimal stopping policy that is implemented on-line.

Equation (31) proceeds backward from $n = N - 1$ to $n = 0$ by computing the value functions $V_n(\cdot)$ inductively. For each step n, this involves taking the expectation of the value function of the next step $V_{n+1}(\cdot)$ over the observation vector $\mathbf{y}_{n+1}$ conditioned on the observation history. Supposing for now that this expectation can be obtained, the optimal value of the objective in (25) for a given Lagrange multipliers $\boldsymbol{\lambda}$ is provided by $V_0(\boldsymbol{\lambda})$. In the following, it is understood that the exposition is for a given $\boldsymbol{\lambda}$, and $\boldsymbol{\lambda}$ is suppressed from the notation.

Once the value functions $\{V_n(\mathbf{Y}_n)\}_{n=1}^{N-1}$ are available, the optimal stopping rule can be applied forward in time and is given by [15]

$$\mathcal{U}_n(\mathbf{Y}_n) = \begin{cases} \text{stop}, & \text{if } V_n(\mathbf{Y}_n) = g_n(\boldsymbol{\pi}_n, \mathbf{t}_n), \\ \text{continue}, & \text{otherwise}, \end{cases} \tag{32}$$
$$n = 1, 2, \ldots, N - 1.$$

In practice, carrying out the backward induction is not computationally tractable. An immediate obstacle is that one has to deal with the state vector $\mathbf{Y}_n$, whose dimension increases over time. Thus, it is desirable to find a statistic with lower dimensionality that is still sufficient for finding the optimal control policy. The following proposition identifies such a sufficient statistic for the problem at hand.

Proposition 1. *The vector $\boldsymbol{\psi}_n \triangleq [\boldsymbol{\pi}_n \mathbf{t}_n]$ is a sufficient statistic for obtaining the optimal stopping policy.*

Proof. Note first that $\boldsymbol{\pi}_{n+1}$ and $\mathbf{t}_{n+1}$ can be updated via

$$\boldsymbol{\pi}_{n+1} = \boldsymbol{\Phi}_{\boldsymbol{\pi}}(\boldsymbol{\pi}_n, \mathbf{y}_{n+1})$$
$$\triangleq \left[\Phi^{(1)}\left(\pi_n^{(1)}, y_{n+1}^{(1)}\right) \cdots \Phi^{(M)}\left(\pi_n^{(M)}, y_{n+1}^{(M)}\right) \right]^T, \tag{33}$$
$$n = 0, 1, \ldots, N - 1,$$

where

$$\Phi^{(m)}\left(\pi_n^{(m)}, y_{n+1}^{(m)}\right)$$
$$= \frac{p\left(y_{n+1}^{(m)} \mid H_0^{(m)}\right)\pi_n^{(m)}}{p\left(y_{n+1}^{(m)} \mid H_0^{(m)}\right)\pi_n^{(m)} + p\left(y_{n+1}^{(m)} \mid H_1^{(m)}\right)\left(1 - \pi_n^{(m)}\right)}, \tag{34}$$

and from (3)

$$\mathbf{t}_{n+1} = \boldsymbol{\Phi}_{\mathbf{t}}(\mathbf{t}_n, \mathbf{y}_{n+1}) \triangleq \mathbf{t}_n + \mathbf{y}_{n+1}, \quad n = 0, \ldots, N - 1. \tag{35}$$

From (30) and (10), it is clear that $V_N(\mathbf{Y}_N)$ can be written as $\overline{V}_N(\boldsymbol{\psi}_N)$ for an appropriate function $\overline{V}_N(\cdot)$. To prove the proposition by induction, suppose that $V_{n+1}(\mathbf{Y}_{n+1})$ can be expressed as $\overline{V}_{n+1}(\boldsymbol{\psi}_{n+1})$. Then, by defining $\boldsymbol{\Phi}(\boldsymbol{\psi}_n, \mathbf{y}_{n+1}) \triangleq [\boldsymbol{\Phi}_{\boldsymbol{\pi}}(\boldsymbol{\pi}_n, \mathbf{y}_{n+1}) \quad \boldsymbol{\Phi}_{\mathbf{t}}(\mathbf{t}_n, \mathbf{y}_{n+1})]$, it follows that

$$V_n(\mathbf{Y}_n)$$
$$= \max\left\{g_n(\boldsymbol{\pi}_n, \mathbf{t}_n), E\left\{\overline{V}_{n+1}\left(\boldsymbol{\Phi}\left(\boldsymbol{\psi}_n, \mathbf{y}_{n+1}\right)\right) \mid \mathbf{y}_1, \ldots, \mathbf{y}_n\right\}\right\}, \tag{36}$$

which can be evaluated provided that $\boldsymbol{\psi}_n$ and the conditional p.d.f. $p(\mathbf{y}_{n+1} \mid \mathbf{y}_1, \ldots, \mathbf{y}_n)$ are known. However, the latter quantity is expressible in terms of $\boldsymbol{\pi}_n$ as

$$p(\mathbf{y}_{n+1} \mid \mathbf{y}_1, \ldots, \mathbf{y}_n)$$
$$= \prod_{m=1}^{M} p\left(y_{n+1}^{(m)} \mid y_1^{(m)}, \ldots, y_n^{(m)}\right)$$
$$= \prod_{m=1}^{M} \left[p\left(y_{n+1}^{(m)} \mid H_0^{(m)}\right)\pi_n^{(m)} + p\left(y_{n+1}^{(m)} \mid H_1^{(m)}\right)\left(1 - \pi_n^{(m)}\right) \right]. \tag{37}$$

Thus, the value function V_n is solely dependent on $\mathbf{t}_n$ and $\boldsymbol{\pi}_n$, and can thus be written as $\overline{V}_n(\boldsymbol{\psi}_n)$. □

Based on the preceding proposition, an alternative backward induction formula can be written by treating $\{\psi_n\}$ as the state variables:

$$\overline{V}_N\left(\psi_N\right) = g_N(\pi_N, \mathbf{t}_N), \tag{38}$$

$$\overline{V}_n\left(\psi_n\right) = \max\left\{g_n(\pi_n, \mathbf{t}_n), E\left\{\overline{V}_{n+1}\left(\Phi\left(\psi_n, \mathbf{y}_{n+1}\right)\right) \mid \mathbf{Y}_n\right\}\right\}$$
$$\text{for } n = 0, 1, \ldots, N-1. \tag{39}$$

However, since the observation space $(\mathbb{R}^+)^M$ and the state space $[0,1]^M \times (\mathbb{R}^+)^M$ are infinite spaces, the optimal backward induction must be implemented approximately via discretization per step [24]. That is, the state space at each step is partitioned into a finite number of sets and a grid point is chosen in each set. Then, the value functions are approximated to be piecewise-constant and the backward induction is used only for the grid points. The expectations are evaluated via finite sums by quantizing the observations to a finite number of quantization levels. Thus, if one uses a given number of grid points for each channel m, it is immediate that the complexity of the discretized algorithm grows exponentially in the number of subchannels M. Therefore, even with the reduced-dimension sufficient statistic, the implementation of the optimal backward induction can be prohibitively complex even for moderate number of OFDM subchannels.

4. A Basis Expansion-Based Reduced-Complexity Solution

To alleviate the "curse of dimensionality" associated with the optimal solution, suboptimal policies that approximate the optimal policy closely with reduced complexity are desired. For example, the k-step look-ahead policy decides to stop or continue based on the optimal rule truncated at k steps ahead of the current time [16], [20, Section 6.3]. In the case of $k = 1$, the one-step look-ahead rule decides to stop if

$$g_n(\pi_n, \mathbf{t}_n) > E\left\{g_{n+1}\left(\Phi(\pi_n, \mathbf{t}_n, \mathbf{y}_{n+1})\right) \mid \mathbf{Y}_n\right\} \tag{40}$$

and to continue otherwise at each step $n < N$.

Here, a regression-based method that has been applied to problems in quantitative finance [25–27] is adopted to the novel sequential CR sensing scenario. Let us define

$$\mathcal{V}_n\left(\psi_n\right)$$
$$\triangleq E\left\{\overline{V}_{n+1}\left(\Phi\left(\psi_n, \mathbf{y}_{n+1}\right)\right) \mid \mathbf{Y}_n\right\}, \quad n = 1, 2, \ldots, N-1. \tag{41}$$

The idea is to approximate $\mathcal{V}_n(\psi_n)$ by $\widehat{\mathcal{V}}_n(\psi_n) \triangleq \sum_{k=1}^{K} a_{n,k}\phi_{n,k}(\psi_n)$, where $\phi_{n,k}(\psi_n)$, $n = 1, \ldots, N-1$, $k = 1, \ldots, K$, are a set of basis functions and $\mathbf{a}_n \triangleq [a_{n,1} \quad a_{n,2} \quad \cdots \quad a_{n,K}]^T$ is the coefficient vector. To facilitate numerical computation, a finite set of sample trajectories of the state vector are simulated, and the coefficients $\mathbf{a}_n$ are

obtained via least-squares regression of the resulting sample paths of the $\mathcal{V}$-values [26].

Specifically, one first generates J independent sample paths $\{\mathbf{y}_n[j], n = 1, 2, \ldots, N\}$, $j = 1, 2, \ldots, J$, by sampling from the joint p.d.f.

$$\prod_{m=1}^{M}\left[\prod_{n=1}^{N} p\left(y_n^{(m)} \mid H_0^{(m)}\right) p_0^{(m)} + \prod_{n=1}^{N} p\left(y_n^{(m)} \mid H_1^{(m)}\right)\left(1 - p_0^{(m)}\right)\right]. \tag{42}$$

From the sample paths of the observations, the sample paths of $\{\psi_n[j], n = 1, 2, \ldots, N\}$ can be computed for $j = 1, 2, \ldots, J$ by applying (33) and (35), where $\mathbf{t}_0 = \mathbf{0}$ and $\pi_0 = [p_0^{(1)} \cdots p_0^{(M)}]^T$. Given the initial $\mathbf{a}_{N-1}$, the regression coefficients $\mathbf{a}_n$ for $n = N-2, N-3, \ldots, 1$ can be obtained recursively by

$$\mathbf{a}_n = \arg\min_{\bar{\mathbf{a}}_n}$$

$$\times \sum_{j=1}^{J}\left(\max\left\{g_{n+1}\left(\psi_{n+1}[j]\right), \sum_{k=1}^{K} a_{n+1,k}\phi_{n+1,k}\left(\psi_{n+1}[j]\right)\right\} - \sum_{k=1}^{K} \bar{a}_{n,k}\phi_{n,k}\left(\psi_n[j]\right)\right)^2, \tag{43}$$

and the initial $\mathbf{a}_{N-1}$ is found as

$$\mathbf{a}_{N-1}$$
$$= \arg\min_{\bar{\mathbf{a}}_{N-1}} \sum_{j=1}^{J}\left(g_N\left(\psi_N[j]\right) - \sum_{k=1}^{K} \bar{a}_{N-1,k}\phi_{N-1,k}(\psi_{N-1}[j])\right)^2. \tag{44}$$

Note that instead of specifying the functions $\{V_n\}_{n=0}^{N-1}$ in (31) or $\{\overline{V}_n\}_{n=0}^{N-1}$ in (39), the backward induction here is to specify the corresponding expansion coefficients. Once the regression coefficients are available, the optimal stopping rule at step $n \in \{1, \ldots, N-1\}$ is to stop if $g_n(\psi_n) > \widehat{\mathcal{V}}_n(\psi_n)$, and to continue otherwise.

Remark 2. Since the regression coefficients are obtained in a batch fashion using the simulated paths, the associated computational burden may be considerable. On the other hand, it is emphasized that the regression coefficients need to be updated only when the channel gains $\{G^{(m)}\}$ vary significantly. Thus, when the PU transmitters and the SU receivers are stationary such that $\{G^{(m)}\}$ do not change too often, the computational complexity of implementing the batch regression-based stopping policy is affordable. Moreover, it is also possible to pursue a recursive approach for computing the stopping policy, which allows online training using the actual measurement data. This approach is beyond the scope of this paper and will be presented in [17].

An important issue in applying the regression method is to choose the set of basis functions $\{\phi_{n,k}(\cdot)\}$. The a fortiori

chosen basis functions should be able to extract the salient features of the generally nonlinear value functions so that the projected $\mathcal{V}$-function onto the lower-dimensional space remains close to the exact one. Here, we choose the following $K = 18$ basis functions:

$$\phi_{n,1}\left(\boldsymbol{\psi}_n\right) = 1,$$

$$\phi_{n,2}\left(\boldsymbol{\psi}_n\right) = E\left\{g_{n+1}\left(\boldsymbol{\Phi}\left(\boldsymbol{\psi}_n, \mathbf{t}_n, \mathbf{y}_{n+1}\right)\right) \mid \mathbf{Y}_n\right\},$$

$$\phi_{n,3}\left(\boldsymbol{\psi}_n\right) = \frac{1}{M}\sum_{m=1}^{M}\pi_n^{(m)},$$

$$\phi_{n,4}\left(\boldsymbol{\psi}_n\right) = \frac{1}{M}\sum_{m=1}^{M}\pi_n^{(m)}\hat{m},$$

$$\phi_{n,5}\left(\boldsymbol{\psi}_n\right) = \frac{1}{M}\sum_{m=1}^{M}\pi_n^{(m)}\left(\frac{3\hat{m}^2}{2} - \frac{1}{2}\right),$$

$$\phi_{n,7}\left(\boldsymbol{\psi}_n\right) = \frac{1}{M}\sum_{m=1}^{M}\hat{t}_n^{(m)},$$

$$\phi_{n,8}\left(\boldsymbol{\psi}_n\right) = \frac{1}{M}\sum_{m=1}^{M}\hat{t}_n^{(m)}\hat{m},$$

$$\phi_{n,9}\left(\boldsymbol{\psi}_n\right) = \frac{1}{M}\sum_{m=1}^{M}\hat{t}_n^{(m)}\left(\frac{3\hat{m}^2}{2} - \frac{1}{2}\right),$$

$$\phi_{n,10}\left(\boldsymbol{\psi}_n\right) = \frac{1}{M}\sum_{m=1}^{M}\hat{t}_n^{(m)}\left(\frac{5\hat{m}^3}{2} - \frac{3\hat{m}}{2}\right),$$

$$\phi_{n,k}\left(\boldsymbol{\psi}_n\right) = \phi_{n,2}\left(\boldsymbol{\psi}_n\right)\phi_{n,k-8}\left(\boldsymbol{\psi}_n\right), \quad k = 11,\ldots,18,$$

(45)

where $n \in \{1, 2, \ldots, N\}$, $\hat{m} \triangleq 2(m-1)/(M-1) - 1$, and $\hat{t}_n^{(m)} \triangleq t_n^{(m)}/\gamma_n^{(m)}$.

The basis functions $\{\phi_{n,k}, k = 3,\ldots,6\}$ and $\{\phi_{n,k}, k = 7,\ldots,10\}$ are defined as the inner products of Legendre polynomials of order 0, 1, 2, and 3 with $\{\pi_n^{(m)}\}$ and $\{\hat{t}_n^{(m)}\}$, respectively; see also [25] for additional issues on selecting $\phi_{n,k}$ and the expansion order K. It is noted that albeit linear in the coefficients $\mathbf{a}_n$ and hence computationally tractable, the basis expansion approximates well the nonlinear dependency of $\mathcal{V}_n(\cdot)$ on $\boldsymbol{\pi}_n$ and $\mathbf{t}_n$ through the appropriately chosen basis functions.

5. Numerical Results

Scenarios with $M = 10$, $N = T = 100$, and $T_s = 1$ were tested. Rates $R^{(m)} = m$ for $m = 1,\ldots,M$ were used, and the channel gains $\{G^{(m)}\}$ were generated from the χ^2-distribution. The observation noise variance σ^2 was set to 10^{-6}, which corresponds to, for example, a subband bandwidth of $100\,\text{kHz}$, with double-sided power spectral density of the noise $N_0/2 = 5 \times 10^{-12}\,\text{W/Hz}$. The prior probabilities $p_0^{(m)} = 0.7$ for all m, and the false alarm rate $\bar{\beta} = 10^{-2}$ was used. To estimate the $\mathbf{a}_n$ coefficients involved
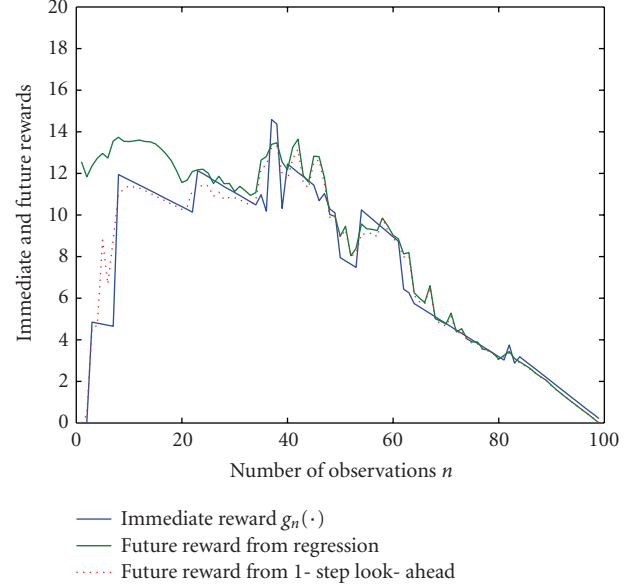

FIGURE 2: A sample path of reward values.

in the regression-based suboptimal policy, $J = 1000$ sample paths were generated.

Figure 2 depicts a sample path of the immediate reward $g_n(\boldsymbol{\psi}_n)$ for $n = 1, 2, \ldots, N$ and a sample path of the future reward $\hat{\mathcal{V}}_n(\boldsymbol{\psi}_n)$ of the regression-based method as well as that of the one-step look-ahead policy, given by the r.h.s. of (40), for mean SNR $= -3\,\text{dB}$ and $\boldsymbol{\lambda} = \mathbf{0}$. In this particular example, the one-step look-ahead policy declares "stop" at $n = 3$ (the smallest n such that (40) is satisfied) with a reward of 4.8, while the regression-based policy stops at $n = 26$ (the smallest n such that $g_n(\boldsymbol{\psi}_n) > \hat{\mathcal{V}}_n(\boldsymbol{\psi}_n)$) to obtain a total rate reward of 11.7. If it were possible to make a noncausal decision after seeing all N observations, a "genie-aided" rule would be able to maximize the objective in (25) over the $\mathcal{U}_n$'s that potentially depend on $\mathbf{Y}_N$. Thus, the "genie-aided" rule asserts that $\mathcal{U}_n = \text{stop}$ for a given $\boldsymbol{\lambda}$ if n is the maximizer of

$$\max_{n \in \{1,2,\ldots,N\}} g_n\left(\boldsymbol{\psi}_n; \boldsymbol{\lambda}\right). \tag{46}$$

Since the feasible set of policies for the "genie-aided" case is a superset of that of the original problem (25), the optimal objective corresponding to the "genie-aided" rule represents an upper-bound to the throughput that can be obtained from the optimal sensing policy. In the particular example under discussion, the "genie-aided" rule would be able to achieve a reward of 14.6 by stopping at time $n = 37$, at which the maximum of $g_n(\boldsymbol{\psi}_n)$ is achieved.

Figure 3 shows the convergence of the collision probabilities $\{P_c^{(m)}\}$ obtained from Monte Carlo simulations of 2000 independent runs, versus the iteration index ℓ. The corresponding Lagrange multipliers as updated per (28) are also shown, for the regression-based method at a mean SNR of $-3\,\text{dB}$. The different curves correspond to the $M = 10$ bands. A constant step size $\mu_\ell = \bar{\beta}^{-1}$ was used. The dual
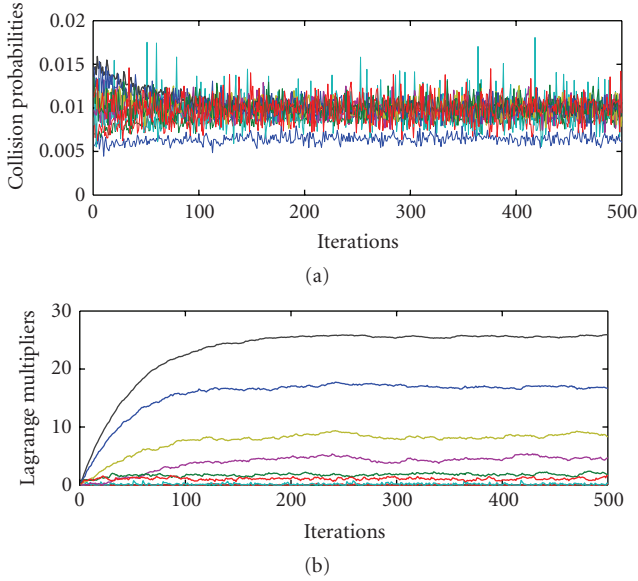
FIGURE 3: Evolution of the collision probabilities (a) and the Lagrange multipliers (b).
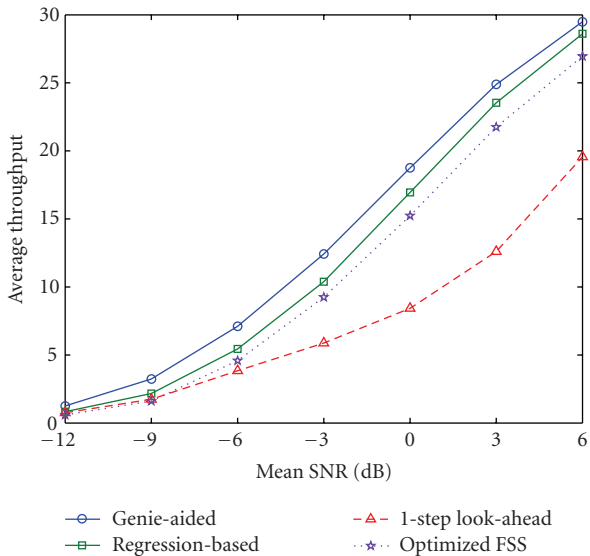

FIGURE 4: Average throughput versus mean SNR.

FIGURE 5: The ratio of the average throughput achieved to the genie-aided throughput.

update algorithm is seen to converge and the collision probability constraints are met in each band.

Figure 4 plots the achieved average total rates of the proposed regression-based scheme when the mean SNR of the PU-to-CR channels is varied. For comparison, the average throughput of the genie-aided and the one-step look-ahead schemes are also shown. Averaging was performed over 2,000 independent Monte Carlo realizations. It can be seen that the regression-based scheme attains throughput close to the genie-aided upper-bound over a wide range of SNR values. Also, the one-step look-ahead scheme is clearly suboptimal especially in the moderate-to-high SNR range.
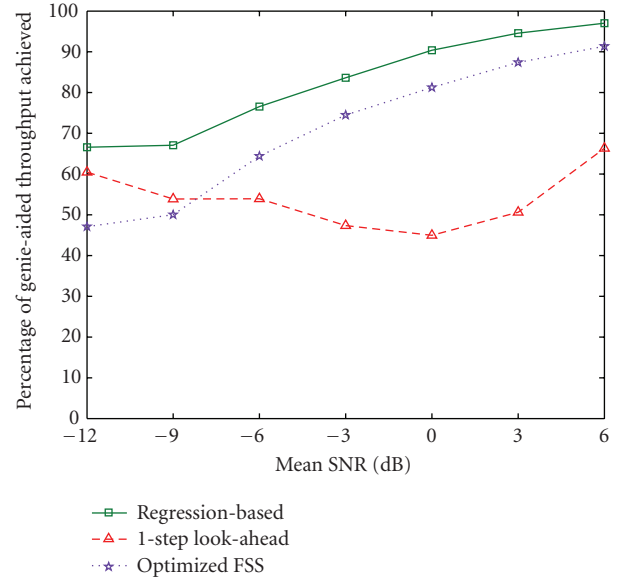
To compare the performance of the sequential regression-based sensing policy to that of the batch scheme, the FSS test is designed and optimized in the following way. For the given set of channel gain values $\{G^{(m)}\}$ and noise variance σ^2, the detection threshold to satisfy the misdetection probability constraint at the sample size n is given by (8), and the corresponding false alarm probability by (7). Then, the average throughput due to the FSS test with sample size n can be computed as

$$E\{f_n'(\boldsymbol{\mathcal{H}}, \mathbf{t}_n)\} = \frac{T - nT_s}{T} \sum_{m=1}^{M} R^{(m)} p_0^{(m)} \left(1 - \alpha_n^{(m)}\right). \quad (47)$$

The average throughput of the optimized FSS test is then defined as the value of

$$\max_{1 \le n \le N, n \in \mathbb{N}} E\{f_n'(\boldsymbol{\mathcal{H}}, \mathbf{t}_n)\}, \quad (48)$$

where the expectation is over $\boldsymbol{\mathcal{H}}$ and $\{\mathbf{y}_n\}$. The sample size n for the optimized FSS test is computed given the channel gains before receiving any samples. The computational complexity associated with the one-dimensional search in (48) is rather trivial.

Figure 4 depicts the average throughput of the optimized FSS scheme versus the mean SNR level. It can be seen that the proposed regression-based sequential sensing scheme outperforms the optimized FSS over all SNR levels tested. Figure 5 plots the ratio of the average throughput of the various schemes to that of the genie-aided scheme. It is seen that the regression-based policy achieves significant portion of the genie-aided throughput, but the optimized FSS test degrades as the SNR decreases. In fact, at SNR as low as -12 dB, even the one-step look-ahead policy outperforms the optimized FSS test, which corroborates the value of sequential CR sensing.
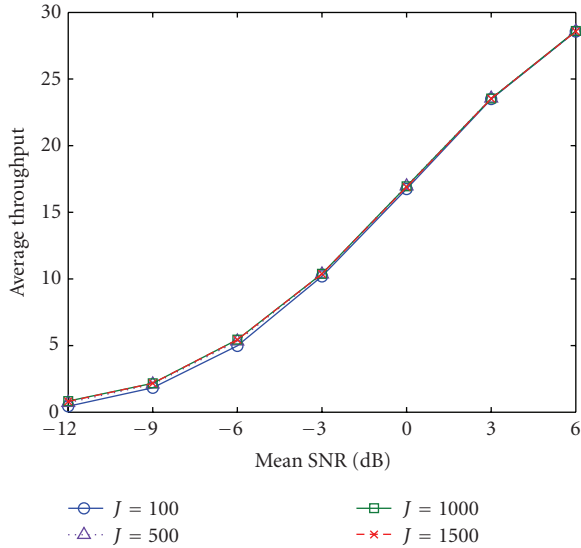
FIGURE 6: The average throughput of the regression-based policy for different numbers J of simulated paths.

To test the sensitivity of the proposed algorithm to the number J of the simulated paths used for computing the regression coefficients $\mathbf{a}_n$, Figure 6 plots the average throughput of the regression-based policy for different values of J. It can be seen that when J is as small as 100, there is some degradation in performance in the low SNR regime. However, the performance improvement as J is increased saturates very quickly when J is larger than 500.

6. Conclusions

Sequential sensing algorithms for OFDM-based wideband CR systems have been developed. The tradeoff between the sensing time and the chance of identifying more unoccupied subchannels were captured in the effective rate achieved by the CR system. Optimal stopping problems were formulated, which maximized the effective rate given the past and current observations. Although a sufficient statistic with dimension lower than that of the accumulated samples was identified, the computational complexity of the optimal solution of the associated dynamic programming problem was still prohibitive. A basis expansion-based reduced-complexity solution was derived, whose performance was shown to be close to the genie-aided upper-bounds and hence close to that of the optimal solution.

In a companion paper [17], an extension to the *cooperative* sequential sensing will be considered. A recursive (on-line) version of the batch training procedure developed here will also be presented, which further reduces the computational burden and the memory requirements associated with the batch counterpart developed here.

Acknowledgments

References

[1] Q. Zhao and A. Swami, "A survey of dynamic spectrum access: signal processing and networking perspectives," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '07)*, vol. 4, pp. 1349–1352, Honolulu, Hawaii, USA, April 2007.

[2] A. Sahai, N. Hoven, M. Mishra, and R. Tandra, "Fundamental tradeoffs in robust spectrum sensing for opportunistic frequency reuse," Tech. Rep., University of California, Berkeley, Calif, USA, 2006.

[3] T. A. Weiss and F. K. Jondral, "Spectrum pooling: an innovative strategy for the enhancement of spectrum efficiency," *IEEE Communications Magazine*, vol. 42, no. 3, pp. S8–S14, 2004.

[4] S. Geirhofer, L. Tong, and B. M. Sadler, "Dynamic spectrum access in the time domain: modeling and exploiting white space," *IEEE Communications Magazine*, vol. 45, no. 5, pp. 66–72, 2007.

[5] C. Cordeiro, K. Challapali, D. Birru, and N. Sai Shankar, "IEEE 802.22: the first worldwide wireless standard based on cognitive radios," in *Proceedings of the 1st IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN '05)*, pp. 328–337, Baltimore, Md, USA, November 2005.

[6] D. Cabric, A. Tkachenko, and R. W. Brodersen, "Spectrum sensing measurements of pilot, energy, and collaborative detection," in *Proceedings of the Military Communications Conference (MILCOM '06)*, pp. 1–7, Washington, DC, USA, October 2006.

[7] H. Urkowitz, "Energy detection of unknown deterministic signals," *Proceedings of the IEEE*, vol. 55, no. 4, pp. 523–531, 1967.

[8] M. Öner and F. Jondral, "On the extraction of the channel allocation information in spectrum pooling systems," *IEEE Journal on Selected Areas in Communications*, vol. 25, no. 3, pp. 558–565, 2007.

[9] Z. Quan, S. Cui, A. H. Sayed, and H. V. Poor, "Optimal multiband joint detection for spectrum sensing in cognitive radio networks," *IEEE Transactions on Signal Processing*, vol. 57, no. 3, pp. 1128–1140, 2009.

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

[11] J. Unnikrishnan and V. V. Veeravalli, "Cooperative sensing for primary detection in cognitive radio," *IEEE Journal on Selected Topics in Signal Processing*, vol. 2, no. 1, pp. 18–27, 2008.

[12] Y.-C. Liang, Y. Zeng, E. C. Y. Peh, and A. T. Hoang, "Sensing-throughput tradeoff for cognitive radio networks," *IEEE Transactions on Wireless Communications*, vol. 7, no. 4, pp. 1326–1337, 2008.

[13] D. Siegmund, *Sequential Analysis*, Springer, New York, NY, USA, 1985.

[14] N. Kundargi and A. Tewfik, "Hierarchical sequential detection in the context of dynamic spectrum access for cognitive radios," in *Proceedings of the IEEE International Conference*

on Electronics, Circuits, and Systems, pp. 514–517, Marrakech, Morocco, December 2007.

[15] Y. S. Chow, H. Robbins, and D. Siegmund, *Great Expectations: The Theory of Optimal Stopping*, Houghton Mifflin Company, Boston, Mass, USA, 1971.

[16] J. Jia, Q. Zhang, and X. Shen, "HC-MAC: a hardware-constrained cognitive MAC for efficient spectrum management," *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 1, pp. 106–117, 2008.

[17] S.-J. Kim and G. B. Giannakis, "Sequential sensing for multi-channel cognitive radios," submitted to *IEEE Transactions on Signal Processing*.

[18] S. M. Kay, *Fundamentals of Statistical Signal Processing: Detection Theory*, vol. 2, Prentice-Hall, Upper Saddle River, NJ, USA, 1993.

[19] Z. Quan, S. Cui, A. H. Sayed, and H. V. Poor, "Wideband spectrum sensing in cognitive radio networks," in *Proceedings of the IEEE International Conference on Communications*, pp. 901–906, Beijing, China, May 2008.

[20] D. P. Bertsekas, *Dynamic Programming and Optimal Control*, vol. 1, Athena Scientific, Belmont, Mass, USA, 2nd edition, 2000.

[21] T. S. Ferguson, *Optimal Stopping and Applications*, Mathematics Department, UCLA, Los Angeles, Calif, USA.

[22] D. A. Castañon, "Approximate dynamic programming for sensor management," in *Proceedings of the 36th IEEE Conference on Decision and Control (CDC '97)*, vol. 2, pp. 1202–1207, San Diego, Calif, USA, December 1997.

[23] N. Z. Shor, *Minimization Methods for Non-Differentiable Functions*, Springer, New York, NY, USA, 1985.

[24] D. P. Bertsekas, "Convergence of discretization procedures in dynamic programming," *IEEE Transactions on Automatic Control*, vol. 20, no. 3, pp. 415–419, 1975.

[25] J. N. Tsitsiklis and B. Van Roy, "Optimal stopping of Markov processes: hilbert space theory, approximation algorithms, and an application to pricing high-dimensional financial derivatives," *IEEE Transactions on Automatic Control*, vol. 44, no. 10, pp. 1840–1851, 1999.

[26] J. N. Tsitsiklis and B. Van Roy, "Regression methods for pricing complex American-style options," *IEEE Transactions on Neural Networks*, vol. 12, no. 4, pp. 694–703, 2001.

[27] F. A. Longstaff and E. S. Schwartz, "Valuing American options by simulation: a simple least-squares approach," *The Review of Financial Studies*, vol. 14, no. 1, pp. 113–147, 2001.