# Sequential and Cooperative Sensing for Multi-Channel Cognitive Radios

Seung-Jun Kim and Georgios B. Giannakis, *Fellow, IEEE*

***Abstract*—Effective spectrum sensing is a critical prerequisite for multi-channel cognitive radio (CR) networks, where multiple spectrum bands are sensed to identify transmission opportunities, while preventing interference to the primary users. The present paper develops sequential spectrum sensing algorithms which explicitly take into account the sensing time overhead, and optimize a performance metric capturing the effective average data rate of CR transmitters. A constrained dynamic programming problem is formulated to obtain the policy that chooses the best time to stop taking measurements and the best set of channels to access for data transmission, while adhering to hard "collision" constraints imposed to protect primary links. Given the associated Lagrange multipliers, the optimal access policy is obtained in closed form, and the subsequent problem reduces to an optimal stopping problem. A basis expansion-based sub-optimal strategy is employed to mitigate the prohibitive computational complexity of the optimal stopping policy. A novel on-line implementation based on the recursive least-squares (RLS) algorithm along with a stochastic dual update procedure is then developed to obviate the lengthy training phase of the batch scheme. Cooperative sequential sensing generalizations are also provided with either raw or quantized measurements collected at a central processing unit. The numerical results presented verify the efficacy of the proposed algorithms.**

***Index Terms*—Cognitive radio, optimal stopping, sequential detection, spectrum sensing.**

## I. Introduction

It was recognized recently that the cognitive radio (CR) paradigm has the potential to mitigate the apparent scarcity of spectral resources caused by static licensing of the spectrum. A CR is capable of sensing the environment to detect the underutilized "white spaces" in the temporal and/or spectral domains and dynamically adapt its transmission parameters, thus exploiting the emergent communication opportunities without disturbing the licensed primary users (PUs) [27].

A key enabler of CR systems is the sensing module, which is needed to monitor spectrum usage of the primary system in real time. However, there are a number of challenges to the sensing task. First, to protect the PU links, stringent interference requirements must be imposed on the CR operation [8]. Moreover, coherent receivers based on matched filtering are often infeasible because PU signal characteristics are unknown. In such cases, non-coherent receivers based on energy [4], [35] or feature (e.g., cyclostationarity) detection are employed [22]. These less efficient detectors and the strict miss detection probability may require excessive sensing time. This can lead to having the CR left with little time for actual data transmission in between PU transmissions. Hence, it is critical to take into account the sensing time overhead explicitly in designing the sensing module, especially when the PU occupancy changes dynamically in time and frequency [11], [37].

Having a means to sense efficiently, multi-channel CR architectures, such as orthogonal frequency-division multiplexing (OFDM), allow sensing to be performed in parallel over multiple channels. The multi-channel sensing problem was considered in [26]. A cooperative sensing strategy was studied in [33], where a linear-quadratic combiner was considered for correlated measurements based on the deflection criterion. All of these works, however, rely on *batch* (or fixed sample size (FSS)) detectors that do not account for the sensing time overhead in conjunction with the achievable throughput.

The present paper deals with *sequential* schemes, for which the number of data used for detection is a random variable dependent on the realization of the observed samples. As a result, sequential detection schemes can opportunistically reduce the sample size required to meet the specified reliability target. In Wald's sequential probability ratio test (SPRT), for instance, the likelihood ratio of the underlying simple binary hypothesis testing problem is compared with two thresholds per sampling interval, and the decision is made as soon as either of the two thresholds is exceeded [29, p. 21]. An SPRT-like sequential energy detection scheme was proposed in [36], where dual thresholds were applied to the accumulated sample energy to reduce implementation complexity. A hierarchical architecture was developed in [16] for suboptimal cooperative sensing, where SPRTs were employed at the individual CRs and also at the fusion center (FC), with the objective of minimizing the overall detection delay. However, none of these works considered multi-channel CR systems. In [6], an SPRT was employed at the FC for detection of OFDM signals, but it was assumed that the PU signal occupied the entire multi-channels when present. In fact, it is not straightforward to apply the SPRT framework to the multi-channel case, where a bank of detectors operate in parallel with the same stopping time, and the contributions of the individual channels to the performance objective might not be of equal weight.

In this work, sequential sensing algorithms are developed to maximize the average total achievable rate of a multi-channel

Manuscript received August 20, 2009; accepted April 16, 2010. Date of publication April 26, 2010; date of current version July 14, 2010. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Philippe Ciblat. This work was supported by NSF grants CCF 0830480 and CON 0824007; and also through collaborative participation in the Communications and Networks Consortium sponsored by the U.S. Army Research Laboratory under the Collaborative Technology Alliance Program, Cooperative Agreement DAAD19-01-2-0011. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation thereon.

The authors are with the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55455 USA (e-mail: seungjun@umn.edu; georgios@umn.edu).

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/TSP.2010.2049106

CR system under prescribed miss detection probability constraints on each channel. The time spent on sensing is accounted for in the effective rate criterion, which captures the trade-off between sensing accuracy and sensing time. The formulation entails a constrained dynamic program (DP). Its solution yields the optimal stopping policy, which dictates when to stop sensing, as well as the optimal channel access policy, which identifies the set of channels to be used for data transmission. To mitigate the prohibitive computational complexity for obtaining the optimal stopping policy, a basis expansion-based sub-optimal approach is pursued, where both a batch least-squares (LS)-based algorithm as well as a recursive least-squares (RLS)-based on-line implementation are developed. Useful extensions to the *cooperative* sensing scenarios are also studied, where it is shown that the solution approach for the single-CR case can be applied to obtain the optimal policies.

A throughput-sensing trade-off related to the one studied here was examined in [18] and [12]. However, [18] considered non-sequential FSS energy detection schemes for a single-channel CR. Multi-channel sensing with serial channel scanning was pursued in [12], [13], [38] as optimal stopping problems. Channel selection strategies under Markov PU occupancy models were studied in [37], [34] using DP formulations.

In our companion paper [15], a sequential multi-channel CR sensing strategy also related to the one considered here was reported. However, a sub-optimal access policy based on the accumulated sample energy statistic was enforced in [15]; whereas, the optimal channel access policy here is identified through Lagrange relaxation of a constrained DP formulation. Moreover, only the batch training scheme was considered for the DP solver in [15], while a novel on-line implementation is developed here. Finally, the extension to cooperative sequential sensing is unique to the present work.

The rest of the paper is organized as follows. In Section II, the signal model and the problem formulation are presented. The optimal solution is developed in Section III. Section IV introduces a basis expansion-based reduced-complexity algorithm to obtain sub-optimal yet computationally tractable batch and on-line solutions. Section V deals with the extension to the cooperative sequential sensing. Alternative batch detection strategies, useful for assessing the performance benefit of the proposed sequential schemes, are provided in Section VI. Numerical tests are presented in Section VII, followed by the conclusions in Section VIII.

## II. Modeling and Problem Formulation

### A. Signal Model

Consider a CR system operating on a set of $M$ orthogonal bands shared opportunistically with PUs. The focus of the present work is on the CR sensing task, with the goal of identifying the set of available bands that are not occupied by the PUs.

The $n$th 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

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

where $\{h_n^{(m)}\}$ are the channel coefficients, $\{s_n^{(m)}\}$ are the PU signal samples, and $\{z_n^{(m)}\}$ are independent and identically distributed (*i.i.d.*) complex Gaussian additive noise samples with mean 0 and variance $\sigma^2$. Assuming that PU occupancy is independent across the $M$ bands, the CR performs a binary hypothesis test per band $m$ to discriminate between the following two hypotheses:

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

Adopting the Bayesian hypothesis testing framework, the prior probabilities $\Pr\{H_0^{(m)}\}$ and $\Pr\{H_1^{(m)}\}$ are denoted by $\pi_0^{(m)}$ and $1-\pi_0^{(m)}$, respectively.

For reasons well documented, e.g., in [4], each CR receiver relies on energy detection to decide the occupancy of each band. Thus, the observation at each time step $n$ is defined as $y_n^{(m)} \triangleq |r_n^{(m)}|^2$. Suppose that the channel coefficients $\{h_n^{(m)}\}$ do not vary over the detection interval so that $|h_n^{(m)}|^2$ can be denoted as $G^{(m)}$, $\forall n$, and the transmitted signal is PSK-modulated and has unit power; i.e., $E\{|s_n^{(m)}|^2\} = 1$, $\forall n,m$. Clearly, these assumptions are ensured if the CR receiver can learn its channel with the PU blindly, or, by overhearing the pilot signal transmitted by the PU system. This is feasible when the transceivers are not moving so that the channel is quasi-static; see e.g., [26] for a justifying argument in the context of digital television systems.

The PU-to-SU channels are obtained during the *acquisition stage*. Since the delay incurred by channel detection-estimation schemes is not of primary concern during the acquisition stage, a sufficient number of samples can be used to ensure accuracy of the acquired channel. Relying on these acquired PU-to-SU channels, the algorithms of the present paper are intended for the *operational stage* to sense PU transmissions even based on energy detection, while optimally taking the detection delay into account, and attaining a desirable tradeoff with the achievable throughput.

The observations $\{y_1^{(m)}, y_2^{(m)}, \ldots, y_N^{(m)}\}$ are i.i.d. (non-) central $\chi^2$-distributed under $H_0^{(m)}$ and $H_1^{(m)}$ with the conditional densities for each $m$ given by [14, Sec. 2.2]

$$p(y^{(m)} \,|\, H_0^{(m)}) = \frac{1}{\sigma^2} e^{-\frac{y^{(m)}}{\sigma^2}} \mathbb{1}_{\{y^{(m)} \geq 0\}} \tag{3}$$

$$p(y^{(m)} | H_1^{(m)}) = \frac{1}{\sigma^2} e^{-\frac{y^{(m)}+G^{(m)}}{\sigma^2}} I_0\left(\frac{\sqrt{G^{(m)} y^{(m)}}}{\sigma^2/2}\right) \tag{4}$$

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.[1] Let us define $\mathbf{y}_n \triangleq [y_n^{(1)} \ y_n^{(2)} \cdots y_n^{(M)}]^T$.

Miss detection in the CR sensing context represents failing to detect PU presence, which could lead to causing interference to the PU transmission. For this reason, the sensing algorithms must guarantee very low miss-detection probability. On the other hand, low false-alarm probability is desirable as it increases the usage of the available channels. These conflicting objectives can be accommodated by increasing the sample size $n$. However, increasing the $n$ increases the sensing overhead,

[1]More generally, it is possible to apply the results of the present paper even when the channel gains are unknown, provided that one has available the conditional data *p.d.f.*s in (3) and (4) capturing the randomness due to channel uncertainty; the rest of the derivations can proceed without major modification.

which effectively reduces the time remaining for transmission of useful data. In the next subsection, a sequential sensing problem is set up to optimize the effective throughput by taking into account the sensing overhead.

### B. Problem Formulation

*1) Average Throughput Criterion:* During the operational stage, consider a CR frame of duration $T$ divided into a sensing phase and a data transmission phase. In the sensing phase, one sample per channel is taken at every sampling interval of duration $\tau$. Let $\delta_n^{(m)}$ denote the permit-to-access decision for channel $m$ after the $n$th sampling interval, which is equal to 1 if the channel is detected idle (that is, $H_0^{(m)}$ is deemed to be in effect), and 0 otherwise. Define the $M \times 1$ vector $\boldsymbol{\delta}_n \triangleq [\delta_n^{(1)}\ \delta_n^{(2)} \ldots \delta_n^{(M)}]^T$ collecting all these binary decision variables. The PU occupancy over the $M$ channels is denoted by an $M \times 1$ random vector $\boldsymbol{\mathcal{H}}$, whose $m$th entry $\mathcal{H}^{(m)}$ takes values from $\{H_0^{(m)}, H_1^{(m)}\}$. Let also $R^{(m)}$ denote the known rate that can be achieved when transmitting over channel $m$. Then, if the CR stops sensing after $n$ sampling intervals and proceeds to data transmission based on the channel access decisions $\boldsymbol{\delta}_n$, the throughput over all channels will be given by

$$f_n'(\boldsymbol{\mathcal{H}}, \boldsymbol{\delta}_n) = \frac{T - n\tau}{T} \sum_{m=1}^{M} R^{(m)} \mathbb{1}_{\{H_0^{(m)}\}} \delta_n^{(m)}, \quad n = 1, 2, \ldots, N \tag{5}$$

where $N\tau \leq T$. The presence of the indicator functions and the decision variables in (5) accounts for the fact that the CR will attain throughput $f_n'(\cdot)$ provided that certain channels are truly idle, and are also detected correctly as being idle.

From (5), the throughput-sensing trade-off is apparent: as the number of observed samples increases, the factor $(T - n\tau/T)$ diminishes, but more available bands may be correctly identified to yield a higher value for the sum rate in (5). Thus, the goal is to obtain the average throughput-optimal *stopping policy* that determines whether to stop or not at each time $n$, and the *access policy* indicating which bands may be used for data transmission if the sensing stops at time $n$, given the accumulated observations $\mathbf{Y}_n \triangleq [\mathbf{y}_1\ \mathbf{y}_2 \ldots \mathbf{y}_n]$.

Let $\Delta_n(\mathbf{Y}_n)$ per time $n$ denote the stopping rule that maps $\mathbf{Y}_n$ to the stopping decision variable $\Delta_n \in \{S, \bar{S}\}$, where $S$ signifies the "stop" decision and $\bar{S}$ "continue"; and $\boldsymbol{\delta}_n(\mathbf{Y}_n) \triangleq [\delta_n^{(1)}(\mathbf{Y}_n) \ldots \delta_n^{(M)}(\mathbf{Y}_n)]^T$ the channel access rule that maps $\mathbf{Y}_n$ to $\boldsymbol{\delta}_n \in \{1, 0\}^M$. The stopping policy refers to $\{\Delta_n(\cdot)\}_{n=1}^{N-1}$, and the channel access policy to $\{\boldsymbol{\delta}_n(\cdot)\}_{n=1}^{N}$. Note that $\Delta_N \equiv S$ by design, since we deal with a finite-horizon sequential detection problem. Our goal is to maximize the average throughput

$$E_{\mathbf{Y}_N, \boldsymbol{\mathcal{H}}} \left\{ \sum_{n=1}^{N} \mathbb{1}_{\{\boldsymbol{\mathcal{D}}_{n-1} = \bar{S}, \Delta_n = S\}} f_n'(\boldsymbol{\mathcal{H}}, \boldsymbol{\delta}_n) \right\} \tag{6}$$

over the control policies $\{\Delta_n(\cdot)\}_{n=1}^{N-1}$ and $\{\boldsymbol{\delta}_n(\cdot)\}_{n=1}^{N}$, where $\boldsymbol{\mathcal{D}}_{n-1} \triangleq [\Delta_0\ \Delta_1 \ldots \Delta_{n-1}]$ with $\Delta_0 \equiv \bar{S}$ and $\boldsymbol{\mathcal{D}}_{n-1} = \bar{S}$ is a shorthand for $\Delta_0 = \Delta_1 = \cdots = \Delta_{n-1} = \bar{S}$. The indicator function in (6) ensures that the reward $f_n'(\cdot)$ is evaluated only at the smallest time slot $n^*$ satisfying $\Delta_{n^*} = S$; for the rest of the time steps, $n < n^*$ and $n > n^*$, the summands are zero. A remark is now in order.

*Remark 1:* Practical reasons dictate sensing to terminate simultaneously across all channels. Indeed, hardware considerations discourage sensing on one channel while transmitting concurrently on another one. This explains why the stopping time was assumed common to all channels, and justifies the choice of optimizing the sum throughput rather than each channel's throughput independently. A naive alternative would be to employ an SPRT per channel and continue sensing until all detectors have finished sensing. However, this would be suboptimal because the detectors that stop earlier than others could have lowered their false alarm rates by not stopping earlier, and thus contribute to increasing the throughput.

Clearly, the per-step reward $f_n'(\cdot)$ in (5) is a function of the underlying spectrum occupancy $\boldsymbol{\mathcal{H}}$, which is not directly observable, and also of the data $\{y_n^{(m)}\}$ through $\{\delta_n^{(m)}\}$. However, note that (6) can be re-written as

$$E_{\mathbf{Y}_N} \left\{ E_{\boldsymbol{\mathcal{H}}|\mathbf{Y}_N} \left\{ \sum_{n=1}^{N} \mathbb{1}_{\{\boldsymbol{\mathcal{D}}_{n-1} = \bar{S}, \Delta_n = S\}} f_n'(\boldsymbol{\mathcal{H}}, \boldsymbol{\delta}_n) \middle| \mathbf{Y}_N \right\} \right\} \tag{7}$$

$$= E_{\mathbf{Y}_N} \left\{ \sum_{n=1}^{N} \mathbb{1}_{\{\boldsymbol{\mathcal{D}}_{n-1} = \bar{S}, \Delta_n = S\}} E_{\boldsymbol{\mathcal{H}}|\mathbf{Y}_N} \{ f_n'(\boldsymbol{\mathcal{H}}, \boldsymbol{\delta}_n) | \mathbf{Y}_N \} \right\}. \tag{8}$$

Consider now the data-conditioned average per-step reward given by [cf. (5)]

$$f_n(\boldsymbol{\pi}_n, \boldsymbol{\delta}_n) \triangleq E_{\boldsymbol{\mathcal{H}}|\mathbf{Y}_n} \{ f_n'(\boldsymbol{\mathcal{H}}, \boldsymbol{\delta}_n) | \mathbf{Y}_n \} = \frac{T - n\tau}{T} \sum_{m=1}^{M} R^{(m)} \pi_n^{(m)} \delta_n^{(m)} \tag{9}$$

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

$$\pi_n^{(m)} \triangleq \Pr \left\{ H_0^{(m)} \middle| y_1^{(m)}, \ldots, y_n^{(m)} \right\}. \tag{10}$$

Based on (9), the average throughput (6) can be expressed as

$$E_{\mathbf{Y}_N} \left\{ \sum_{n=1}^{N} \mathbb{1}_{\{\boldsymbol{\mathcal{D}}_{n-1} = \bar{S}, \Delta_n = S\}} f_n(\boldsymbol{\pi}_n, \boldsymbol{\delta}_n) \right\}. \tag{11}$$

Using Bayes' rule, the belief vector can be expressed recursively as

$$\boldsymbol{\pi}_{n+1} = \boldsymbol{\Phi}(\boldsymbol{\pi}_n, \mathbf{y}_{n+1}) \triangleq \left[ \Phi^{(1)} \left( \pi_n^{(1)}, y_{n+1}^{(1)} \right) \ldots \Phi^{(M)} \left( \pi_n^{(M)}, y_{n+1}^{(M)} \right) \right], \quad n = 0, 1, \ldots, N-1 \tag{12}$$

where

$$\Phi^{(m)} \left( \pi_n^{(m)}, y_{n+1}^{(m)} \right) \triangleq \frac{p \left( y_{n+1}^{(m)} \middle| H_0^{(m)} \right) \pi_n^{(m)}}{p \left( y_{n+1}^{(m)} \middle| H_0^{(m)} \right) \pi_n^{(m)} + p \left( y_{n+1}^{(m)} \middle| H_1^{(m)} \right) \left( 1 - \pi_n^{(m)} \right)}. \tag{13}$$

For future reference, it is important to remember that substitution of (3)–(4) into (13) allows computation of $\boldsymbol{\pi}_{n+1}$ from $\boldsymbol{\pi}_n$ and $\mathbf{y}_{n+1}$ using the first-order recursion (12) initialized with the *a priori* known $\boldsymbol{\pi}_0$.

It can be verified that (11) is a valid dynamic program (DP) objective function by considering the augmented state $\boldsymbol{\mathcal{S}}_{n+1} \triangleq [\mathbf{Y}_{n+1},\ \boldsymbol{\mathcal{D}}_n]$, which depends on the state $\boldsymbol{\mathcal{S}}_n$ of the present step as well as the value of $[\mathbf{y}_{n+1},\ \Delta_n]$. It is clear that the $n$th summand of (11) is a function of the state $\boldsymbol{\mathcal{S}}_n$ and the control $[\Delta_n,\ \boldsymbol{\delta}_n]$ of the $n$th step, as required by the DP formalism [3, Ch. 1].

In addition to maximizing the average throughput in (11), the CR sensing must adhere to prescribed constraints preventing interference to PU communications. To this end, a constrained formulation is developed in the sequel to find the optimal control policies $\{\Delta_n(\mathbf{Y}_n)\}$ and $\{\boldsymbol{\delta}_n(\mathbf{Y}_n)\}$ such that the average throughput in (11) is maximized over the finite horizon comprising $N$ sampling intervals.

*2) Constrained Dynamic Programming Formulation:* The CR access policy has to ensure low "collision" probability, which is the probability that the CR interferes with the PU transmission due to miss detection of spectrum occupancy. The "collision" probability $P_c^{(m)}$ on channel $m$ can be expressed as

$$P_c^{(m)} = \sum_{n=1}^{N} \Pr\left\{\boldsymbol{\mathcal{D}}_{n-1} = \bar{S}, \Delta_n = S, \delta_n^{(m)} = 1 \,\middle|\, H_1^{(m)}\right\}. \quad (14)$$

It is desired that for each band the "collision" probability is kept under a small positive threshold $\bar{\beta}$; i.e.,

$$P_c^{(m)} \leq \bar{\beta}, \quad m \in \{1, 2, \ldots, M\}. \quad (15)$$

Now, recall that the expected value of a $\{0,1\}$-binary random variable is equal to the probability of the random variable being equal to 1. Hence, $P_c^{(m)}$ in (14) can be re-written as

$$P_c^{(m)} = \sum_{n=1}^{N} E_{\mathbf{Y}_n|H_1^{(m)}}\left\{\mathbb{1}_{\left\{\boldsymbol{\mathcal{D}}_{n-1}=\bar{S}, \Delta_n = S, \delta_n^{(m)}=1\right\}} \,\middle|\, H_1^{(m)}\right\} \quad (16)$$

$$= E_{\mathbf{Y}_N}\left\{\sum_{n=1}^{N} \mathbb{1}_{\left\{\boldsymbol{\mathcal{D}}_{n-1}=\bar{S}, \Delta_n = S, \delta_n^{(m)}=1\right\}} \frac{1-\pi_n^{(m)}}{\Pr\left\{H_1^{(m)}\right\}}\right\} \quad (17)$$

$$= E_{\mathbf{Y}_N}\left\{\sum_{n=1}^{N} \mathbb{1}_{\{\boldsymbol{\mathcal{D}}_{n-1}=\bar{S}, \Delta_n = S\}} \delta_n^{(m)} \frac{1-\pi_n^{(m)}}{1-\pi_0^{(m)}}\right\}. \quad (18)$$

Define

$$c_n^{(m)}\left(\pi_n^{(m)}, \delta_n^{(m)}\right) \triangleq \frac{1-\pi_n^{(m)}}{1-\pi_0^{(m)}} \delta_n^{(m)}, \quad n = 1, \ldots, N. \quad (19)$$

Then, the desired optimization problem can be formulated as a constrained DP problem given by

$$\max_{\substack{\{\Delta_n(\mathbf{Y}_n)\}_{n=1}^{N-1} \\ \{\boldsymbol{\delta}_n(\mathbf{Y}_n)\}_{n=1}^{N}}} E_{\mathbf{Y}_N}\left\{\sum_{n=1}^{N} \mathbb{1}_{\{\boldsymbol{\mathcal{D}}_{n-1}=\bar{S}, \Delta_n = S\}} f_n(\boldsymbol{\pi}_n, \boldsymbol{\delta}_n)\right\} \quad (20)$$

subject to

$$E_{\mathbf{Y}_N}\left\{\sum_{n=1}^{N} \mathbb{1}_{\{\boldsymbol{\mathcal{D}}_{n-1}=\bar{S}, \Delta_n = S\}} c_n^{(m)}\left(\pi_n^{(m)}, \delta_n^{(m)}\right)\right\} \leq \bar{\beta}, \quad m = 1, 2, \ldots, M. \quad (21)$$

Note that the functional optimization problem defined in (20)–(21) is feasible since never accessing any channel (i.e., $\delta_n^{(m)}(\mathbf{Y}_n) \equiv 0,\ \forall n, m$) is a feasible solution.

*3) Lagrange Relaxation Approach:* One approach to solve constrained DP problems is to use Lagrange relaxation; see e.g., [5]. Let $\boldsymbol{\lambda} \triangleq [\lambda^{(1)}\lambda^{(2)}\ldots\lambda^{(M)}]^T$ be a set of Lagrange multipliers with $\lambda^{(m)} \geq 0, m \in \{1, 2, \ldots, M\}$. Upon defining

$$g_n(\boldsymbol{\pi}_n, \boldsymbol{\delta}_n; \boldsymbol{\lambda}) \triangleq f_n(\boldsymbol{\pi}_n, \boldsymbol{\delta}_n) - \sum_{m=1}^{M} \lambda^{(m)} c_n^{(m)}\left(\pi_n^{(m)}, \delta_n^{(m)}\right), \quad n = 1, 2, \ldots, N \quad (22)$$

for any given $\boldsymbol{\lambda}$, the relaxed unconstrained DP problem to solve is

$$\max_{\substack{\{\Delta_n(\mathbf{Y}_n)\}_{n=1}^{N-1} \\ \{\boldsymbol{\delta}_n(\mathbf{Y}_n)\}_{n=1}^{N}}} E_{\mathbf{Y}_N}\left\{\sum_{n=1}^{N} \mathbb{1}_{\{\boldsymbol{\mathcal{D}}_{n-1}=\bar{S}, \Delta_n = S\}} g_n(\boldsymbol{\pi}_n, \boldsymbol{\delta}_n; \boldsymbol{\lambda})\right\}. \quad (23)$$

The approaches to solve the problem (23) for a given $\boldsymbol{\lambda}$ will be developed in Sections III and IV.

The optimal Lagrange multiplier vector can be obtained iteratively so that the constraints in (15) are satisfied. A simple approach to this end, is to use the subgradient method offline [28]

$$\lambda_{\ell+1}^{(m)} = \max\left\{0, \lambda_\ell^{(m)} + \mu_\ell\left(P_c^{(m)} - \bar{\beta}\right)\right\}, \quad m = 1, 2, \ldots, M \quad (24)$$

where $\ell$ is the iteration index, and $\mu_\ell$ is a preselected sequence of step sizes.

Carrying through the iteration (24) presumes ability to obtain the expected value in (18) needed to compute $P_c^{(m)}$. Section IV will detail how this computation is possible offline using Monte Carlo runs. But for now, it is instructive to focus on the functional maximization in (23) for a fixed $\boldsymbol{\lambda}$ that will be assumed known in the next section.

## III. Optimal Solution

The relaxed unconstrained DP problem in (23) can be solved via backward induction.[2] Moreover, it is proven that the belief states $\{\boldsymbol{\pi}_n\}$ constitute sufficient statistics for computing the optimal control policies [3, Sec. 5.4]. Consider the so-called value functions $\{V_n(\cdot)\}$ defined recursively as follows:

$$V_N(\boldsymbol{\pi}_N; \boldsymbol{\lambda}) \triangleq \max_{\boldsymbol{\delta}_N} g_N(\boldsymbol{\pi}_N, \boldsymbol{\delta}_N; \boldsymbol{\lambda}) \quad (25)$$

$$V_n(\boldsymbol{\pi}_n; \boldsymbol{\lambda}) \triangleq \max\{\max_{\boldsymbol{\delta}_n} g_n(\boldsymbol{\pi}_n, \boldsymbol{\delta}_n; \boldsymbol{\lambda}), E_{\mathbf{y}_{n+1}|\mathbf{Y}_n}\{V_{n+1}(\boldsymbol{\Phi}(\boldsymbol{\pi}_n, \mathbf{y}_{n+1}); \boldsymbol{\lambda})|\mathbf{Y}_n\}\}, \quad n = N-1, N-2, \ldots, 1. \quad (26)$$

[2]It should be noted that the backward induction is not part of the actual sensing procedure. In principle, the mappings $\{\Delta_n(\mathbf{Y}_n)\}$ and $\{\boldsymbol{\delta}_n(\mathbf{Y}_n)\}$ need to be first computed and stored, and sequential sensing takes place by applying these mappings to the observation sequence. However, we will also develop an on-line version in Section IV-A-2), where the computation (and refinement) of the policy itself and its application for sensing occur at the same time.

Then, as shown in Appendix A, the optimal stopping rule at any slot $n \in \{1,2,\ldots,N-1\}$ is given by

$$\Delta_n^*(\boldsymbol{\pi}_n;\boldsymbol{\lambda}) = \begin{cases} S, & \text{if} \quad V_n(\boldsymbol{\pi}_n;\boldsymbol{\lambda}) = \max_{\boldsymbol{\delta}_n} g_n(\boldsymbol{\pi}_n,\boldsymbol{\delta}_n;\boldsymbol{\lambda}) \\ \bar{S}, & \text{otherwise} \end{cases} \tag{27}$$

and upon stopping, the optimal access policy at slot $n \in \{1,2,\ldots,N\}$ is

$$\boldsymbol{\delta}_n^*(\boldsymbol{\pi}_n;\boldsymbol{\lambda}) = \arg \max_{\boldsymbol{\delta}_n \in \{1,0\}^M} g_n(\boldsymbol{\pi}_n,\boldsymbol{\delta}_n;\boldsymbol{\lambda}) \tag{28}$$

$$= \arg \max_{\boldsymbol{\delta}_n \in \{1,0\}^M} \sum_{m=1}^{M} \left( \frac{T-n\tau}{T} R^{(m)} \pi_n^{(m)} - \lambda^{(m)} \frac{1-\pi_n^{(m)}}{1-\pi_0^{(m)}} \right) \delta_n^{(m)}. \tag{29}$$

Backward induction proceeds from the last stage $n = N$ backward to $n = 1$, computing the optimal stopping and access decisions $\Delta_n$ and $\boldsymbol{\delta}_n$ for each value of $\boldsymbol{\pi}_n$ according to (27) and (28). The principle of optimality implies that the control policies $\{\Delta_n^*(\boldsymbol{\pi}_n;\boldsymbol{\lambda})\}$ and $\{\boldsymbol{\delta}_n^*(\boldsymbol{\pi}_n;\boldsymbol{\lambda})\}$ so obtained offer the optimal solution of (23) [3, Ch. 1].

The term $E_{\mathbf{y}_{n+1}|\mathbf{Y}_n}\{V_{n+1}(\boldsymbol{\Phi}(\boldsymbol{\pi}_n,\mathbf{y}_{n+1});\boldsymbol{\lambda})|\mathbf{Y}_n\}$ in (26) represents the expected throughput that can be achieved if sensing is stopped some time *after* the $n$th slot. Thus, the optimal stopping rule (27) dictates that sensing must cease as soon as the throughput gained when stopping at the present time slot $n$ exceeds the expected throughput due to stopping in the future.

It is also worth noting that the optimal solution in (29) dictates setting $\delta_n^{(m)}$ to 1 when the coefficient for $\delta_n^{(m)}$ in the objective of (29) is non-negative, and to 0 otherwise, for each $m$. Thus, after some manipulation, the optimal access rule in (29) is given in closed form as

$$\delta_n^{*(m)}\left(\pi_n^{(m)}\right) \triangleq \mathbb{1}_{\left\{\Gamma_n^{(m)} \le \gamma_n^{(m)}\right\}}, \quad m = 1,2,\ldots,M \tag{30}$$

where

$$\Gamma_n^{(m)}(\pi_n^{(m)}) \triangleq \frac{p\left(y_1^{(m)},\ldots,y_n^{(m)} \middle| H_1^{(m)}\right)}{p\left(y_1^{(m)},\ldots,y_n^{(m)} \middle| H_0^{(m)}\right)} = \frac{1-\pi_n^{(m)}}{\pi_n^{(m)}} \frac{\pi_0^{(m)}}{1-\pi_0^{(m)}} \tag{31}$$

$$\gamma_n^{(m)} \triangleq \frac{T-n\tau}{T} R^{(m)} \cdot \frac{\pi_0^{(m)}}{\lambda^{(m)}}. \tag{32}$$

Therefore, it is clear that the optimal access decision of each band is based on a likelihood ratio test whose threshold depends on the Lagrange multiplier associated with the band. It is seen from (32) that the optimum access policy becomes less aggressive as $n$ grows, since the effective rate $(T-n\tau)/(T)R^{(m)}$ diminishes and hence so does the incentive to access the channel. In terms of implementation, the likelihood ratio $\Gamma_n^{(m)}$ can be easily computed on-line using (12) in the last equality of (31). The threshold $\gamma_n^{(m)}$ is obtained from the known parameters once the Lagrange multiplier vector $\boldsymbol{\lambda}$ is fixed.

*Remark 2:* In our companion paper [15], the energy detector output was intuitively (yet suboptimally) employed for the channel access decision. The analysis here reveals that the *optimal* access policy can be obtained naturally from the DP formulation as a likelihood ratio test.

By substituting the optimal access policy to $g_n(\boldsymbol{\pi}_n,\boldsymbol{\delta}_n;\boldsymbol{\lambda})$, the per-step reward becomes

$$g_n^*(\boldsymbol{\pi}_n;\boldsymbol{\lambda}) \triangleq g_n(\boldsymbol{\pi}_n,\boldsymbol{\delta}_n^*(\boldsymbol{\pi}_n);\boldsymbol{\lambda}) = \max_{\boldsymbol{\delta}_n} g_n(\boldsymbol{\pi}_n,\boldsymbol{\delta}_n;\boldsymbol{\lambda}) \tag{33}$$

$$= \sum_{m=1}^{M} \left( \frac{T-n\tau}{T} R^{(m)} \pi_n^{(m)} - \lambda^{(m)} \frac{1-\pi_n^{(m)}}{1-\pi_0^{(m)}} \right)^+ \tag{34}$$

where $(\cdot)^+ \triangleq \max\{0,\cdot\}$. Thus, the problem in (23) reduces to an *optimal stopping problem* defined precisely through the sequence of observations $\{\mathbf{y}_n\}$ and the sequence of rewards $\{g_n^*(\boldsymbol{\pi}_n;\boldsymbol{\lambda})\}$ [7], [10], [3, Sec. 4.4]. The optimal stopping problem seeks the "best" time to stop taking observations in the sense of maximizing the expected value of the chosen reward. The optimal stopping rule for this problem is given by [cf. (26) and (27)]

$$\Delta_n^*(\boldsymbol{\pi}_n;\boldsymbol{\lambda}) = \begin{cases} S, & \text{if } g_n^*(\boldsymbol{\pi}_n;\boldsymbol{\lambda}) > E_{\mathbf{y}_{n+1}|\mathbf{Y}_n} \\ & \{ \quad V_{n+1}(\boldsymbol{\Phi}(\boldsymbol{\pi}_n,\mathbf{y}_{n+1});\boldsymbol{\lambda})|\mathbf{Y}_n\} \\ \bar{S}, & \text{otherwise.} \end{cases} \tag{35}$$

Since the observation space $(\mathbb{R}^+)^M$ and the state space $[0,1]^M$ for the state variables $\{\boldsymbol{\pi}_n\}$ are infinite sets, the conditional expectation in (26) is hard to evaluate. Therefore, the optimal backward induction must be implemented approximately via discretization per step [2]. That is, the state space is partitioned into a finite number of sets and a grid point is chosen in each set. Similarly, the observation space is discretized to finite quantization levels. Then, the value functions are approximated to be piecewise-constant, and backward induction is carried out only at the grid points by approximating the expectations via finite sums. 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 with the number of channels $M$. Therefore, implementation of the optimal backward induction can be prohibitively complex even for a moderate number of channels.

## IV. Reduced-Complexity Suboptimal Solutions

### A. Basis Expansion-Based Approach

To alleviate the "curse of dimensionality" associated with carrying out the backward induction in (25)–(26), sub-optimal policies that approximate the optimal policy closely with reduced complexity are desired. Here, a regression-based method that has been applied to problems in quantitative finance [31], [32], [20] is adopted to the novel sequential CR sensing scenario. Let us define

$$\mathcal{V}_n(\boldsymbol{\pi}_n;\boldsymbol{\lambda}) \triangleq E_{\mathbf{y}_{n+1}|\mathbf{Y}_n}\{V_{n+1}(\boldsymbol{\Phi}(\boldsymbol{\pi}_n,\mathbf{y}_{n+1});\boldsymbol{\lambda})|\mathbf{Y}_n\}, \quad n = 1,\ldots,N-1. \tag{36}$$

The idea is to approximate $\mathcal{V}_n(\boldsymbol{\pi}_n;\boldsymbol{\lambda})$ by

$$\hat{\mathcal{V}}_n(\boldsymbol{\pi}_n;\boldsymbol{\lambda}) \triangleq \sum_{k=1}^{K} a_{n,k} \boldsymbol{\mathcal{b}}_{n,k}(\boldsymbol{\pi}_n) = \mathbf{a}_n^T \boldsymbol{\mathcal{b}}_n(\boldsymbol{\pi}_n) \tag{37}$$

where $\boldsymbol{\mathcal{b}}_n(\cdot) \triangleq [\mathcal{b}_{n,1}(\cdot)\ \mathcal{b}_{n,2}(\cdot)\ldots\mathcal{b}_{n,K}(\cdot)]^T$ per $n \in \{1,\ldots,N-1\}$ is a vector of $K$ basis functions and $\mathbf{a}_n \triangleq [a_{n,1}\ a_{n,2}\ldots a_{n,K}]^T$ is the coefficient vector.

An important issue in applying the regression method is to choose the set of basis functions $\{\mathcal{b}_{n,k}(\cdot)\}$. The chosen basis functions should be able to extract the salient features of the value functions so that the projected $\mathcal{V}$-function onto the lower-dimensional space remains close to the exact one. An example set of basis functions used for the numerical experiments of Section VII are given in Appendix B. It is noted that albeit linear in the coefficients $\mathbf{a}_n$, and hence computationally tractable, the basis expansion approximates well the non-linear dependency of $\mathcal{V}_n(\cdot)$ on $\boldsymbol{\pi}_n$ through the appropriately chosen basis functions.

The coefficient vectors $\{\mathbf{a}_n\}$ are obtained via regression. Two alternatives are considered in the following. In the batch training scheme, the batch least-squares (LS) method [21, Ch. 3] is employed to the data samples generated via Monte Carlo simulations. In the on-line alternative, recursive least-squares (RLS) is used on the actual observations to learn the coefficient vectors on-line.

*1) Batch Algorithm:* In the case of batch estimation of the vectors $\{\mathbf{a}_n\}$, a finite set of sample trajectories of the observation vector are simulated from their joint distribution, and the coefficients $\{\mathbf{a}_n\}$ are obtained via LS regression of the resulting sample paths of the $\mathcal{V}$-values [32].

Specifically, one first generates $J$ independent sample paths $\{\mathbf{y}_n[j], n = 1,2,\ldots,N\}, j = 1,2,\ldots,J$, according to the signal model set forth in Section II-A; that is, by sampling from the probability density function

$$\prod_{m=1}^{M} \sum_{q\in\{1,0\}} \left[\prod_{n=1}^{N} p\left(y_n^{(m)} \middle| H_q^{(m)}\right) \Pr\left\{H_q^{(m)}\right\}\right]. \tag{38}$$

From the sample paths of the observations, the sample paths of $\{\boldsymbol{\pi}_n[j], n = 1,2,\ldots,N\}$ can be computed for $j = 1,2,\ldots,J$ by applying (12), where $\boldsymbol{\pi}_0 \triangleq [\pi_0^{(1)}\ldots\pi_0^{(M)}]^T$. Vector $\mathbf{a}_{N-1}$ is computed first using

$$\mathbf{a}_{N-1} = \arg\min_{\bar{\mathbf{a}}_{N-1}} \sum_{j=1}^{J} \left( g_N^*(\boldsymbol{\pi}_N[j]) - \sum_{k=1}^{K} \bar{a}_{N-1,k} \mathcal{b}_{N-1,k}(\boldsymbol{\pi}_{N-1}[j]) \right)^2 \tag{39}$$

where the dependence of $g_n^*(\cdot)$ on $\boldsymbol{\lambda}$ has been suppressed for simplicity of notation. Given $\mathbf{a}_{N-1}$, the regression coefficients $\mathbf{a}_n$ for $n = N-2, N-3,\ldots,1$ can be obtained recursively by [cf. (25), (26) and (37)]

$$\mathbf{a}_n = \arg\min_{\bar{\mathbf{a}}_n} \sum_{j=1}^{J} \left( \max\left\{ g_{n+1}^*(\boldsymbol{\pi}_{n+1}[j]), \sum_{k=1}^{K} a_{n+1,k} \mathcal{b}_{n+1,k}(\boldsymbol{\pi}_{n+1}[j]) \right\} - \sum_{k=1}^{K} \bar{a}_{n,k} \mathcal{b}_{n,k}(\boldsymbol{\pi}_n[j]) \right)^2. \tag{40}$$

Once the regression coefficients are computed, the optimal stopping rule at step $n \in \{1,2,\ldots,N-1\}$, is [cf. (35)]

$$\hat{\Delta}_n(\boldsymbol{\pi}_n;\boldsymbol{\lambda}) = \begin{cases} S, & \text{if } g_n^*(\boldsymbol{\pi}_n;\boldsymbol{\lambda}) > \mathbf{a}_n^T \boldsymbol{\mathcal{b}}_n(\boldsymbol{\pi}_n) \\ \bar{S}, & \text{otherwise.} \end{cases} \tag{41}$$

The overall procedure to approximate the solution to the constrained DP in (20)–(21) is as follows. First, generate $J$ independent sample paths and use them in a batch form to obtain the estimates of the parameterized value functions $\hat{\mathcal{V}}_n(\cdot)$ as described earlier. Then, the sequential sensing algorithm must be run many times, again using simulated paths, to obtain estimates of the "collision" probabilities $P_c^{(m)}, m \in \{1,2,\ldots,M\}$, which are necessary to update the Lagrange multipliers $\boldsymbol{\lambda}$ per (24). Specifically, a sample average-version of (18) may be used to obtain

$$\hat{P}_c^{(m)} = \frac{1}{J} \sum_{j=1}^{J} \sum_{n=1}^{N} \delta_n^{*(m)} \left(\pi_n^{(m)}[j]\right) \frac{1-\pi_n^{(m)}[j]}{1-\pi_0^{(m)}} \cdot \mathbb{1}_{\{\hat{\Delta}_1(\boldsymbol{\pi}_1[j])=\ldots=\hat{\Delta}_{n-1}(\boldsymbol{\pi}_{n-1}[j])=\bar{S},\hat{\Delta}_n(\boldsymbol{\pi}_n[j])=S\}}. \tag{42}$$

These two steps must be repeated until the Lagrange multipliers converge. Therefore, the overall algorithm is computationally demanding, which can be justified only when the channel variation is slow. To remedy this, an on-line algorithm based on recursive estimation of the value functions as well as on stochastic optimization of the Lagrange multipliers is developed next.

*2) On-Line Algorithm:* In solving approximately DP problems on-line, one naturally faces two conflicting objectives: obtaining better estimates of the value functions, and achieving higher overall reward based on the current estimates of the value functions [24, Ch. 10]. In our problem setup, this trade-off between so-called *exploration versus exploitation* manifests itself as a dilemma of choosing to stop at time $n$ whenever the current estimate $\hat{\mathcal{V}}_n(\cdot)$ drops below the immediate reward $g_n^*(\cdot)$, versus to continue nonetheless as the current estimate $\hat{\mathcal{V}}_n(\cdot)$ might not be very accurate and needs to be refined by visiting additional states.

Our approach is to use the RLS updates in covariance form for estimating the value functions on-line, while using an index policy based on the confidence level of the value function estimate. Use of an index policy in the approximate DP solution herein is motivated by the similarity of the challenge with the one encountered in the class of multi-armed bandit problems, where an index policy is known to be optimal [1], [24, Ch. 10].[3]

A detailed description of the proposed on-line algorithm is given in Table I. In Steps 0 and 1, the regression coefficients $\mathbf{a}_n$, the Lagrange multiplier vector $\boldsymbol{\lambda}$, the collision probability estimates $P_c^{(m)}$ of time slot $n$, the covariance matrices $\mathbf{P}_n$ of the RLS algorithm, and the prior probability vector $\boldsymbol{\pi}_0$ are first initialized. In Step 2, the observations $\mathbf{y}_n$ are taken at time $n$ by sensing the channels $m = 1,2,\ldots,M$, and the belief vector $\boldsymbol{\pi}_n$ is updated based on (12) in Step 3. The sensing process is

[3]The multi-armed bandit problem refers to the challenge of choosing one from $K$ arms to receive a random reward dictated by the state of the chosen arm. The state of the chosen arm evolves according to a Markov process. The objective is to maximize the expected total discounted reward over an infinite horizon. An index policy refers to the one of choosing the arm with the highest associated index, where each index is computed as a function of the state solely of the corresponding arm.

TABLE I
ON-LINE ALGORITHM FOR SEQUENTIAL SENSING

Step 0: Initialization
Initialize $\mathbf{a}_n = \mathbf{0}$ for $n \in \{1, 2, \ldots, N\}$
Initialize $\mathbf{P}_n = \epsilon^{-1}\mathbf{I}$ for $n \in \{1, 2, \ldots, N\}$,
where $\epsilon$ is a small positive constant.
Initialize $\lambda^{(m)},\ m \in \{1, 2, \ldots, M\}$, to positive values.
Initialize $P_c^{(m)} = 0$ for all $m \in \{1, 2, \ldots, M\}$.
Set $\ell = 1$.
Step 1: Initialize $\pi_0^{(m)} = \Pr(H_0^{(m)})$ for $m \in \{1, 2, \ldots, M\}$
and set $n = 1$.
Step 2: Take observation $\mathbf{y}_n$.
Step 3: Compute updated belief vector $\boldsymbol{\pi}_n = \boldsymbol{\Phi}(\boldsymbol{\pi}_{n-1}, \mathbf{y}_n)$
using (12) and (13).
Step 4: If $n = N$ or $g_n^*(\boldsymbol{\pi}_n; \boldsymbol{\lambda}) > \hat{\mathcal{V}}_n(\boldsymbol{\pi}_n; \boldsymbol{\lambda}) + \zeta\sqrt{\boldsymbol{\phi}_n^T(\boldsymbol{\pi}_n)\mathbf{P}_n\boldsymbol{\phi}_n(\boldsymbol{\pi}_n)}$,
stop sensing and go to Step 5;
otherwise, set $n \leftarrow n + 1$ and go to Step 2.
Step 5: For $k = n-1, n-2, \ldots, 1$, perform RLS update
$\mathbf{k}_k = \frac{\xi^{-1}\mathbf{P}_k\boldsymbol{\phi}_k(\boldsymbol{\pi}_k)}{1+\xi^{-1}\boldsymbol{\phi}_k^T(\boldsymbol{\pi}_k)\mathbf{P}_k\boldsymbol{\phi}_k(\boldsymbol{\pi}_k)}$
$\epsilon_k = \max\left\{g_{k+1}^*(\boldsymbol{\pi}_{k+1}; \boldsymbol{\lambda}), \hat{\mathcal{V}}_{k+1}(\boldsymbol{\pi}_{k+1}; \boldsymbol{\lambda})\right\} - \mathbf{a}_k^T\boldsymbol{\phi}_k(\boldsymbol{\pi}_k)$
Set $\mathbf{a}_k \leftarrow \mathbf{a}_k + \mathbf{k}_k\epsilon_k$
Set $\mathbf{P}_k \leftarrow \xi^{-1}\mathbf{P}_k - \xi^{-1}\mathbf{k}_k\boldsymbol{\phi}_k^T(\boldsymbol{\pi}_k)\mathbf{P}_k$
Step 6: Proceed to data transmission on channels $\mathcal{M} \triangleq \{m : \delta_n^{*(m)} = 1\}$
selected via (30)–(32), and collect per-channel ACKs.
Set $\mathcal{M}_A = \{m : \text{ACK received on channel } m\}$.
Step 7: Update collision probability estimates
$P_c^{(m)} \leftarrow \xi P_c^{(m)} + (1-\xi)\frac{\mathbb{1}_{\{m \in \mathcal{M}\setminus\mathcal{M}_A\}}}{1-\pi_0^{(m)}},\ m \in \{1, 2, \ldots, M\}$
Step 8: Update Lagrange multipliers
$\lambda^{(m)} \leftarrow [\lambda^{(m)} + \mu_\ell(P_c^{(m)} - \bar{\beta})]^+,\ m \in \{1, 2, \ldots, M\}$
Step 9: Set $\ell \leftarrow \ell + 1$ and go to Step 1.

carried on until the end of the $N$-step horizon is reached, or, for a prescribed margin constant $\zeta$,

$$g_n^*(\boldsymbol{\pi}_n; \boldsymbol{\lambda}) > \hat{\mathcal{V}}_n(\boldsymbol{\pi}_n; \boldsymbol{\lambda}) + \zeta\sqrt{\boldsymbol{\phi}_n^T(\boldsymbol{\pi}_n)\mathbf{P}_n\boldsymbol{\phi}_n(\boldsymbol{\pi}_n)} \tag{43}$$

is satisfied at a certain sampling time $n = n^*$. Upon stopping, Step 5 of the algorithm performs the RLS update on the regression coefficients $\mathbf{a}_k$ starting from $k = n^* - 1$ down to $k = 1$ in the backward direction. Notice that the trajectory of the belief vector $\boldsymbol{\pi}_k$ for $k = 1, 2, \ldots, n^*$ must be stored in the memory. This strategy is referred to as a double-pass procedure for approximately solving a DP, in contrast to a possible single-pass procedure, where the value function update occurs forward in time simultaneously with the state evolution (that is, as the sensing operation) takes place. The double-pass procedure is often observed to have faster convergence [24, Ch. 8].

The weighted least-squares (WLS) estimate obtained by the RLS algorithm at iteration index $\ell$ with the forgetting factor $\xi \leq 1$ corresponds to the best linear unbiased estimate (BLUE) when the measurements are corrupted by a zero-mean noise sequence with a diagonal covariance matrix whose diagonal entries are given by $\{1, \xi^{-1}, \xi^{-2}, \ldots, \xi^{1-\ell}\}$ [21, Ch. 9]. Note that this choice of noise covariance regards "older" measurements to be more corrupted, thus giving larger weight to the recent measurements in the estimation. With this interpretation, matrices $\mathbf{P}_n$ in the RLS update can be viewed as the covariance of the BLUE [21, p. 126]. Therefore, the variance of the value function estimate under this noise model is given by

$$\text{var}\{\hat{\mathcal{V}}_n(\boldsymbol{\pi}_n; \boldsymbol{\lambda})\} = \boldsymbol{\phi}_n^T(\boldsymbol{\pi}_n)\mathbf{P}_n\boldsymbol{\phi}_n(\boldsymbol{\pi}_n). \tag{44}$$

According to (43), the stop decision is made in Step 4 only when the immediate reward $g_n^*(\cdot)$ exceeds the estimated future reward $\hat{\mathcal{V}}_n(\cdot)$ at least by a margin $\zeta(\geq 0)$, which is a design parameter, times the standard deviation of $\hat{\mathcal{V}}_n(\cdot)$ obtained from (44). In other words, when the confidence level for the value function estimates is low (i.e., the variance is large), the algorithm tends to opt for "exploring" future stages to refine the estimates of the future reward, even though current estimates indicate that such a choice could sacrifice performance. As confidence on the estimates improves, the algorithm will tend to "exploit" the current knowledge of the value function to maximize the overall expected reward.

In Step 6, the CR can proceed to the data transmission phase by transmitting on the set of channels $\mathcal{M} \triangleq \{m : \delta_n^{*(m)} = 1\}$, as dictated by the optimal access rule described in (30)–(32). If the CR protocol has a feature for the CR transmitter to receive an acknowledgment (ACK) reliably from the receiver for each channel $m \in \mathcal{M}$ provided that the transmission was successful on that channel, the CR transmitter can update the collision probability estimates based on the per-channel ACKs. By defining the set of channels for which the ACKs were received as $\mathcal{M}_A$, the update is done as (cf. Step 7)

$$P_c^{(m)} \leftarrow \xi P_c^{(m)} + (1-\xi)\frac{\mathbb{1}_{\{m \in \mathcal{M}\setminus\mathcal{M}_A\}}}{\Pr\{H_1^{(m)}\}}, \quad m \in \{1, 2, \ldots, M\} \tag{45}$$

which approximates the expectation in (18) on-line via exponentially weighted sample averaging.

The Lagrange multiplier vector is updated in Step 8 based on the collision probability estimates following (24), and the next sensing phase is resumed in Step 9 when there are data to transmit.

### B. One-Step Look-Ahead Heuristic

For performance comparison, the one-step look-ahead heuristic policy is also derived. The $k$-step look-ahead policy decides to stop or continue based on the optimal rules truncated $k$ steps ahead of the current time [3, Sec. 6.3], [12]. In the case of $k = 1$, the one-step look-ahead rule decides to stop if [cf. (35)]

$$g_n^*(\boldsymbol{\pi}_n; \boldsymbol{\lambda}) > E_{\mathbf{y}_{n+1}|\mathbf{Y}_n}\{g_{n+1}^*(\boldsymbol{\Phi}(\boldsymbol{\pi}_n, \mathbf{y}_{n+1}); \boldsymbol{\lambda})|\mathbf{Y}_n\} \tag{46}$$

or to continue sensing otherwise at stage $n \in \{1, 2, \ldots, N-1\}$.

The expected value on the r.h.s. of (46) can be obtained explicitly. Define the likelihood ratio $\tilde{\Gamma}^{(m)}(y) \triangleq (p(y|H_1^{(m)}))/(p(y|H_0^{(m)}))$, and let $F(\cdot|H_0^{(m)})$ and $F(\cdot|H_1^{(m)})$ denote the conditional cumulative distribution functions (*c.d.f.*'s) of $y_n^{(m)}$. With the derivation relegated to Appendix C, the expected value in (46) turns out to be

$$\begin{aligned} &E_{\mathbf{y}_{n+1}|\mathbf{Y}_n}\{g_{n+1}^*(\boldsymbol{\Phi}(\boldsymbol{\pi}_n, \mathbf{y}_{n+1}); \boldsymbol{\lambda})|\mathbf{Y}_n\} \\ &\quad = \sum_{m=1}^{M}\left(\frac{T-(n+1)\tau}{T}R^{(m)}\pi_n^{(m)}F\left(\tilde{\theta}_{n+1}^{(m)}\middle|H_0^{(m)}\right)\right. \\ &\qquad \left. - \frac{\lambda^{(m)}\left(1-\pi_n^{(m)}\right)}{1-\pi_0^{(m)}}F\left(\tilde{\theta}_{n+1}^{(m)}\middle|H_1^{(m)}\right)\right) \end{aligned} \tag{47}$$

where

$$\tilde{\theta}_{n+1}^{(m)} \triangleq \left(\tilde{\Gamma}^{(m)}\right)^{-1}\left(\frac{\gamma_{n+1}^{(m)}}{\Gamma_n^{(m)}\left(\pi_n^{(m)}\right)}\right). \tag{48}$$

The sequential sensing based on the one-step look-ahead rule proceeds as follows. At step $n$, based on the observation vector $\mathbf{y}_n$, the belief vector $\boldsymbol{\pi}_n$ is computed via (12). Then, the immediate reward $g_n^*(\boldsymbol{\pi}_n;\boldsymbol{\lambda})$ as well as the expected next-step reward $E_{\mathbf{y}_{n+1}|\mathbf{Y}_n}\{g_{n+1}^*(\boldsymbol{\Phi}(\boldsymbol{\pi}_n,\mathbf{y}_{n+1});\boldsymbol{\lambda})|\mathbf{Y}_n\}$ are computed. If the immediate reward is greater, the sensing stops, and the channels are accessed according to the access policy $\boldsymbol{\delta}_n^*(\boldsymbol{\pi}_n;\boldsymbol{\lambda})$. Otherwise, the sensing continues by taking another observation vector. After the sensing is stopped, the collision probability estimate is updated just like Steps 6 and 7 in Table I, and subsequently, the Lagrange multiplier vector $\boldsymbol{\lambda}$ is updated as in Step 8 in Table I.

Since the (next-step) future reward can be computed explicitly, there is no training involved. Therefore, the described sensing procedure can be executed either off-line with simulated samples (to compute the optimal Lagrange multipliers $\boldsymbol{\lambda}$), or in an on-line fashion relying on the actual observations. Although the one-step look-ahead sequential sensing is simpler and thus less complex than the optimal one, the performance is degraded. In fact, the simulated tests of Section VII will demonstrate that the basis-expansion-based approximation of the optimal procedure outperforms the simple one-step look-ahead heuristic in all operating conditions experimented.

## V. Sequential Cooperative Sensing

The sequential sensing algorithm developed so far for an individual CR can be extended to the cooperative scenario, where multiple CRs cooperate to detect the presence of PUs. By leveraging diversity of the (conditionally) independent observations made by multiple CRs, the cooperative sensing improves detection performance and robustness against fading. Suppose that a CR cluster head node (henceforth referred to as FC) collects the measurements from the individual CRs to perform joint decision.

Consider $I$ CRs that are close enough to each other so that they observe the same PU occupancy over the set of $M$ orthogonal channels. Let $y_n^{(m)}(i)$ denote the measurements of the $i$th CR on channel $m$ at time $n$. It is assumed that the observations are conditionally independent over time and across CRs, conditioned on the PU occupancy. The conditional densities of $y_n^{(m)}(i)$ under $H_0^{(m)}$ and $H_1^{(m)}$ are given by (3) and (4), respectively, with $G^{(m)}$ and $\sigma^2$ substituted with $G^{(m)}(i)$ and $\sigma^2(i)$, respectively. Here, $G^{(m)}(i)$ denotes the channel gain from the PU transmitter to the receiver of CR $i$ on channel $m$, and $\sigma^2(i)$ the receiver noise variance per channel at CR $i$.

Two different strategies for transmitting the observed samples to the FC are considered in the sequel. One entails the raw (analog) observations transmitted to the FC over a noisy channel. The other quantizes first the observations and transmits the quantized bits to the FC error free. The latter is ensured provided that a sufficiently powerful error control coding scheme is employed.

### A. Sequential Cooperative Sensing With Raw Observations

*1) Signal Model:* A simple way to relay the individual observations $y_n^{(m)}(i)$ from the CRs to the FC is to transmit the raw measurements using analog modulation. Following the model described in [25], consider that the CRs transmit their observations over orthogonal channels in the presence of additive white Gaussian noise (AWGN). To facilitate analysis, it is further assumed that each CR transmits a summary of observations every $\eta \in \mathbb{N}$ sampling intervals per frame.[4] That is, CR $i$ transmits at time $n = k\eta$, where $k \in \{1,2,\ldots,\lfloor(N/\eta)\rfloor\}$, the sum of $\eta$ samples generated during $(k-1)\eta < n \leq k\eta$. Stacking the received samples from the $I$ CRs in a vector form, the FC thus obtains at time $n = k\eta$

$$\tilde{\mathbf{y}}_k^{(m)} = \sum_{n=(k-1)\eta+1}^{k\eta} \mathbf{y}_n^{(m)} + \tilde{\mathbf{z}}_k^{(m)} \tag{49}$$

where $\mathbf{y}_n^{(m)} \triangleq [y_n^{(m)}(1) \ \ldots \ y_n^{(m)}(I)]^T$, $\tilde{\mathbf{y}}_k^{(m)} \triangleq [\tilde{y}_k^{(m)}(1)\ldots\tilde{y}_k^{(m)}(I)]^T$, $\tilde{\mathbf{z}}_k^{(m)} \triangleq [\tilde{z}_k^{(m)}(1)\ldots\tilde{z}_k^{(m)}(I)]^T$, and $\tilde{z}_k^{(m)}(i)$ is a real-valued Gaussian random variable with mean 0 and (cross-) correlation

$$E\left\{\tilde{z}_k^{(m)}(i)\tilde{z}_{k'}^{(m')}(i')\right\} = \left(\tilde{\sigma}^{(m)}(i)\right)^2 \mathbb{1}_{\{k=k',m=m',i=i'\}}.$$

Even for modest values of $\eta$, one can appeal to the central limit theorem to approximate $\tilde{y}_k^{(m)}(i)$ as Gaussian, when conditioned on $H_0^{(m)}$ or $H_1^{(m)}$. Thus, the conditional densities of $\tilde{y}_k^{(m)}(i)$ are given by

$$p\left(y|H_0^{(m)}\right) = \mathcal{N}\left(y;\eta\sigma^2(i),\eta\sigma^4(i)+\left(\tilde{\sigma}^{(m)}(i)\right)^2\right) \tag{50}$$

$$p\left(y|H_1^{(m)}\right) = \mathcal{N}\Big(y;\eta\left(\sigma^2(i)+G^{(m)}(i)\right), \eta\left[\sigma^4(i)+2G^{(m)}(i)\sigma^2(i)\right]+\left(\tilde{\sigma}^{(m)}(i)\right)^2\Big) \tag{51}$$

where $\mathcal{N}(y;a,v^2) \triangleq (1)/(\sqrt{2\pi v^2})e^{-((y-a)^2)/(2v^2)}$. Due to independence, the conditional joint *p.d.f.*s of $\tilde{\mathbf{y}}_k^{(m)}$ are simply the products

$$p\left(\tilde{\mathbf{y}}_k^{(m)}\middle|\, H_q^{(m)}\right) = \prod_{i=1}^{I} p\left(\tilde{y}_k^{(m)}(i)\middle|\, H_q^{(m)}\right), \quad q=0,1. \tag{52}$$

Using (52), one can compute recursively the belief vector $\tilde{\boldsymbol{\pi}}_k \triangleq [\tilde{\pi}_k^{(1)}\ldots\tilde{\pi}_k^{(M)}]^T$ via Bayes rule as in (12), where $\tilde{\pi}_k^{(m)} \triangleq \Pr\{H_0^{(m)}|\tilde{\mathbf{Y}}_k^{(m)}\}$ and $\tilde{\mathbf{Y}}_k^{(m)} \triangleq [\tilde{\mathbf{y}}_1^{(m)}\ldots\tilde{\mathbf{y}}_k^{(m)}]$.

*2) Problem Formulation:* Similar to the single-CR case, it is desired to maximize the average network-wide throughput by optimizing the stopping and access policies. To formulate such a problem, the following rate-based reward function is adopted [cf. (9)]

$$\tilde{f}_k(\tilde{pi}_k,\tilde{\boldsymbol{\delta}}_k) = \frac{T-k\eta\tau}{T}\sum_{m=1}^{M}\hat{R}^{(m)}\tilde{\pi}_k^{(m)}\tilde{\delta}_k^{(m)}, \quad k=1,2,\ldots,\left\lfloor\frac{N}{\eta}\right\rfloor \tag{53}$$

[4]An interesting future research direction is to find the value of $\eta$ attaining a desirable tradeoff between feedback overhead and throughput, but this goes beyond the scope of this work.

where $\hat{R}^{(m)} \triangleq \max_i R^{(m)}(i)$ with $R^{(m)}(i)$ denoting the achievable rate on channel $m$ if CR $i$ is allowed to access that channel, and $\tilde{\boldsymbol{\delta}}_k \triangleq [\tilde{\delta}_k^{(1)} \dots \tilde{\delta}_k^{(M)}]^T$ is the vector of access decisions. Modulo these modifications, the optimal stopping and access policies are obtained in an analogous fashion to the single-CR case treated in Section III.

### B. Quantized Observations Case

*1) Signal Model:* Here we suppose that the CRs quantize their observations $y_n^{(m)}(i)$ and relay the quantized bits to the FC. Further, we assume that each CR transmits only a single bit at each time step $n$ per channel. The quantized bit at CR $i$ for channel $m$ at time $n$ is obtained as

$$b_n^{(m)}(i) \triangleq \mathbb{1}_{\left\{y_n^{(m)}(i) \geq \theta^{(m)}(i)\right\}} \tag{54}$$

where $\theta^{(m)}(i)$ denotes the quantization threshold.

The reception of quantized observations at the FC is assumed to be error free. Let $F(y|H_q^{(m)})$ denote the *c.d.f.* of $y_n^{(m)}(i)$ on channel $m$ at CR $i$ under hypothesis $H_q^{(m)}$ for $q = 0, 1$. Define also

$$\begin{aligned} F_q^{(m)}(i) &\triangleq \Pr\left\{ b_n^{(m)}(i) = 0 \middle| H_q^{(m)} \right\} \\ &= F\left( \theta^{(m)}(i) \middle| H_q^{(m)} \right), \quad q = 0, 1. \end{aligned} \tag{55}$$

Then, the conditional probability for $\mathbf{b}_n^{(m)} \triangleq [b_n^{(m)}(1) \dots b_n^{(m)}(I)]^T$ under $H_q^{(m)}$ is given by

$$\begin{aligned} \Pr\left\{ \mathbf{b}_n^{(m)} \middle| H_q^{(m)} \right\} = \prod_{i=1}^{I} &\left[ F_q^{(m)}(i) \right]^{\left(1 - b_n^{(m)}(i)\right)} \\ &\times \left[ 1 - F_q^{(m)} \right]^{b_n^{(m)}(i)}, \quad q = 0, 1. \end{aligned} \tag{56}$$

Thus, the belief vector, defined as $\bar{\boldsymbol{\pi}}_n \triangleq [\bar{\pi}_n^{(1)} \dots \bar{\pi}_n^{(M)}]^T$, where $\bar{\pi}_n^{(m)} \triangleq \Pr\{H_0^{(m)} | \mathbf{b}_1^{(m)}, \dots \mathbf{b}_n^{(m)}\}$, can be updated as [cf. (13)]

$$\bar{\pi}_{n+1}^{(m)} = \frac{\Pr\left\{ \mathbf{b}_{n+1}^{(m)} \middle| H_0^{(m)} \right\} \bar{\pi}_n^{(m)}}{\Pr\left\{ \mathbf{b}_{n+1}^{(m)} \middle| H_0^{(m)} \right\} \bar{\pi}_n^{(m)} + \Pr\left\{ \mathbf{b}_{n+1}^{(m)} \middle| H_1^{(m)} \right\} \left( 1 - \bar{\pi}_n^{(m)} \right)}. \tag{57}$$

Recall now that the data $\{\mathbf{b}_n^{(m)}\}$ and their conditional *p.d.f.*s under $H_q^{(m)}$ along with the belief update equation is all that is needed to run the (sub-)optimal sequential sensing schemes of Section IV. Hence, analogous algorithms apply to the cooperating sensing scenarios with raw or quantized observations.

*2) Quantizer Design:* It remains to select the quantization thresholds $\{\theta^{(m)}(i)\}$ appropriately. For mathematical tractability, maximization of an Ali-Silvey distance metric is pursued; see also [23], [19], [17]. Specifically, the Kullback-Leibler divergence of the quantizer output, given by

$$\sum_{b \in \{0,1\}} \Pr\left\{ b_n^{(m)}(i) = b \middle| H_0^{(m)} \right\} \log \left( \frac{\Pr\left\{ b_n^{(m)}(i) = b \middle| H_0^{(m)} \right\}}{\Pr\left\{ b_n^{(m)}(i) = b \middle| H_1^{(m)} \right\}} \right) \tag{58}$$

is maximized. The Kullback-Leibler divergence is related to the asymptotic decay rate of the false alarm probability of Neyman-Pearson hypothesis tests [9, p. 309]. The necessary conditions for optimality are given by [23]

$$\begin{aligned} \frac{p\left( \theta^{(m)}(i) \middle| H_1^{(m)} \right)}{p\left( \theta^{(m)}(i) \middle| H_0^{(m)} \right)} &= \frac{F_1^{(m)}(i)\left[ 1 - F_1^{(m)}(i) \right]}{F_0^{(m)}(i) - F_1^{(m)}(i)} \\ &\quad \cdot \log \left( \frac{F_0^{(m)}(i)\left[ 1 - F_1^{(m)}(i) \right]}{F_1^{(m)}(i)\left[ 1 - F_0^{(m)}(i) \right]} \right), \quad \forall m, i \end{aligned} \tag{59}$$

which can be solved numerically for $\theta^{(m)}(i)$ to find the desired quantization breakpoints.

*3) Problem Formulation:* As in Section V-A, the immediate reward function at time $n$ is given by the effective total achievable rate:

$$\bar{f}_n(\bar{\boldsymbol{\pi}}_n, \bar{\boldsymbol{\delta}}_n) = \frac{T - n\tau}{T} \sum_{m=1}^{M} \hat{R}^{(m)} \bar{\pi}_n^{(m)} \bar{\delta}_n^{(m)}, \quad n = 1, 2, \dots, N \tag{60}$$

where $\bar{\boldsymbol{\delta}}_n \triangleq [\bar{\delta}_n^{(1)} \ \dots \ \bar{\delta}_n^{(M)}]^T$, and $\bar{\delta}_n^{(m)}$ is the access decision for channel $m$ when stopped at time $n$.

*Remark 3:* Note that with the quantized observations $\{\mathbf{b}_n^{(m)}\}$, one can alternatively employ efficient solvers for partially observable Markov decision processes (POMDPs) to obtain the *exact* solution to the sequential cooperative sensing problem; see, e.g., [30]. However, these approaches are considerably more complex than the basis expansion approach, and do not easily lend themselves to an on-line implementation. Moreover, it is of interest to compare the performance of the algorithms developed in this work on an equal footing. For this reason, the basis expansion-based sub-optimal approach is pursued even for the quantized case.

## VI. Optimized FSS Tests for Cooperative Sensing

To compare the performance of the sequential sensing policy with that of a batch detection scheme, FSS tests based on energy detection are designed and optimized in this section with respect to their sample size $n$. A related optimization approach was pursued in [18] for a single-channel CR. The multi-channel throughput-optimal FSS sensing scheme for the single CR case was presented in [15, Sec. 5]. Here, extensions of the throughput-optimal FSS test for multiple cooperating CRs are developed.

### A. Raw Observations Case

The signal model for this case is provided in Section V-A-1) as multi-variate Gaussian. It is well known that the optimal fusion rule for the binary hypothesis testing in the Neyman-Pearson framework is linear-quadratic [14, Sec. 5.6]. However, in order to obtain an analytic formula for the detection thresholds, which facilitates the subsequent optimization with respect to the sample size, a linear fusion rule is adopted. Thus, the collected observations are linearly combined and tested against a threshold. The test statistic for channel $m$ is given by

$$\tilde{t}_k^{(m)} \triangleq \mathbf{w}^{(m)T} \sum_{k'=1}^{k} \tilde{\mathbf{y}}_{k'}^{(m)} \tag{61}$$

where $\mathbf{w}^{(m)} \triangleq [w^{(m)}(1)\dots w^{(m)}(I)]^T$ is the combining weight vector for channel $m$ with $||\mathbf{w}^{(m)}||_2 = 1$.

The conditional expectations and variances of the test statistics $\{\tilde{t}_k^{(m)}\}$ for $m \in \{1,2,\dots,M\}$ under the two hypotheses can be readily found to be

$$E\left\{\tilde{t}_k^{(m)}\middle| H_0^{(m)}\right\} = k\eta\sum_{i=1}^{I} w^{(m)}(i)\sigma^2(i) \tag{62}$$

$$E\left\{\tilde{t}_k^{(m)}\middle| H_1^{(m)}\right\} = k\eta\sum_{i=1}^{I} w^{(m)}(i)\left(\sigma^2(i)+G^{(m)}(i)\right) \tag{63}$$

$$\mathrm{var}\left\{\tilde{t}_k^{(m)}\middle| H_0^{(m)}\right\} = k\sum_{i=1}^{I}\left(w^{(m)}(i)\right)^2\left[\eta\sigma^4(i)+(\tilde{\sigma}^{(m)}(i))^2\right] \tag{64}$$

$$\mathrm{var}\left\{\tilde{t}_k^{(m)}\middle| H_1^{(m)}\right\} = k\sum_{i=1}^{I}\left(w^{(m)}(i)\right)^2 \cdot\left\{\eta\left[\sigma^4(i)+2G^{(m)}(i)\sigma^2(i)\right]+(\tilde{\sigma}^{(m)}(i))^2\right\}. \tag{65}$$

The weight vectors $\{\mathbf{w}^{(m)}\}$ will be chosen to minimize the false alarm probabilities. In [25], it is shown that the combiner weights that maximize the modified deflection coefficients defined by

$$\left(d^{(m)}\right)^2 \triangleq \frac{\left[E\left\{\tilde{t}_k^{(m)}\middle| H_0^{(m)}\right\} - E\left\{\tilde{t}_k^{(m)}\middle| H_1^{(m)}\right\}\right]^2}{\mathrm{var}\left\{\tilde{t}_k^{(m)}\middle| H_1^{(m)}\right\}} \tag{66}$$

achieve near-optimum performance in the sense of minimizing false alarms. Substituting (62)–(65) into (66) and differentiating with respect to $w^{(m)}(i)$, the combiner weight vector based on this criterion can be readily obtained as $\mathbf{w}^{(m)} = \mathbf{w}^{(m)\prime}/||\mathbf{w}^{(m)\prime}||$, where

$$w^{(m)\prime}(i) = \frac{\eta G^{(m)}(i)}{\sqrt{\eta\left[\sigma^4(i)+2G^{(m)}(i)\sigma^2(i)\right]+\left(\tilde{\sigma}^{(m)}(i)\right)^2}}. \tag{67}$$

The detection rule is thus given by $\tilde{\boldsymbol{\delta}}_k^{\mathrm{FSS}} \triangleq [\tilde{\delta}_k^{\mathrm{FSS},(1)}\dots\tilde{\delta}_k^{\mathrm{FSS},(M)}]^T$ with $\tilde{\delta}_k^{\mathrm{FSS},(m)} \triangleq \mathbb{1}_{\{\tilde{t}_k^{(m)} < \tilde{\gamma}_k^{\mathrm{FSS},(m)}\}}$, where the threshold $\tilde{\gamma}_k^{\mathrm{FSS},(m)}$ for satisfying the miss detection probability target is given by

$$\tilde{\gamma}_k^{\mathrm{FSS},(m)} = Q^{-1}(1-\bar{\beta})\sqrt{\mathrm{var}\left\{\tilde{t}_k^{(m)}\middle| H_1^{(m)}\right\}} + E\left\{\tilde{t}_k^{(m)}\middle| H_1^{(m)}\right\}. \tag{68}$$

Then, the optimized average throughput is given by

$$E_{\mathbf{G}}\left\{\max_{1\le k\le\lfloor N/\eta\rfloor, k\in\mathbb{N}} \frac{T-\eta k\tau}{T}\sum_{m=1}^{M}\hat{R}^{(m)}\pi_0^{(m)}\left(1-\tilde{\alpha}_k^{(m)}\right)\right\} \tag{69}$$

where the false alarm probabilities are obtained as

$$\tilde{\alpha}_k^{(m)} = Q\left(\frac{\tilde{\gamma}_k^{\mathrm{FSS},(m)} - E\left\{\tilde{t}_k^{(m)}\middle| H_0^{(m)}\right\}}{\sqrt{\mathrm{var}\left\{\tilde{t}_k^{(m)}\middle| H_0^{(m)}\right\}}}\right). \tag{70}$$

The optimization in (69) is carried out using an exhaustive search over $k$.

### B. Quantized Observations Case

The FC forms the decision statistic $\bar{t}_n^{(m)}$ by linearly combining the quantized observations from the $I$ CRs; i.e.,

$$\bar{t}_n^{(m)} = \bar{\mathbf{w}}^{(m)T}\sum_{k=1}^{n}\mathbf{b}_k^{(m)} \tag{71}$$

where $\bar{\mathbf{w}}^{(m)} \triangleq [\bar{w}^{(m)}(1)\dots\bar{w}^{(m)}(I)]^T$ is the normalized combiner weight vector with $||\bar{\mathbf{w}}^{(m)}||_2 = 1$. Note that this is essentially equivalent to forming a linear combination of the log-likelihood ratios of the quantized observations, since

$$\sum_{k=1}^{n}\log\left(\frac{\Pr\left\{b_k^{(m)}(i)\middle| H_1^{(m)}\right\}}{\Pr\left\{b_k^{(m)}(i)\middle| H_0^{(m)}\right\}}\right) = C^{(m)}(i)\sum_{k=1}^{n}b_k^{(m)}(i) + nD^{(m)}(i) \tag{72}$$

where the constants $C^{(m)}(i)$ and $D^{(m)}(i)$ are expressed in terms of the *c.d.f.*s in (55) as

$$C^{(m)}(i) \triangleq \log\frac{F_0^{(m)}(i)\left(1-F_1^{(m)}(i)\right)}{F_1^{(m)}(i)\left(1-F_0^{(m)}(i)\right)} \tag{73}$$

$$D^{(m)}(i) \triangleq \log\frac{F_1^{(m)}(i)}{F_0^{(m)}(i)}. \tag{74}$$

Hence, the detection strategy of thresholding the test statistic $\bar{t}_n^{(m)}$ is equivalent to a threshold test based on the linear combination of log-likelihood ratios of the quantized observations.

To optimize the combiner weights, we again resort to maximizing the modified deflection coefficients, which can be defined as in (66) with $\tilde{t}_k^{(m)}$'s replaced by $\bar{t}_n^{(m)}$'s. The conditional means and variances for $\bar{t}_n^{(m)}$ are given by

$$E\left\{\bar{t}_n^{(m)}\middle| H_q(m)\right\} = n\sum_{i=1}^{I}\bar{w}^{(m)}(i)\left[1-F_q^{(m)}(i)\right], \quad q=0,1 \tag{75}$$

$$\mathrm{var}\left\{\bar{t}_n^{(m)}\middle| H_q^{(m)}\right\} = n\sum_{i=1}^{I}\left(\bar{w}^{(m)}(i)\right)^2 \times\left[1-F_q^{(m)}(i)\right]F_q^{(m)}(i), \quad q=0,1. \tag{76}$$

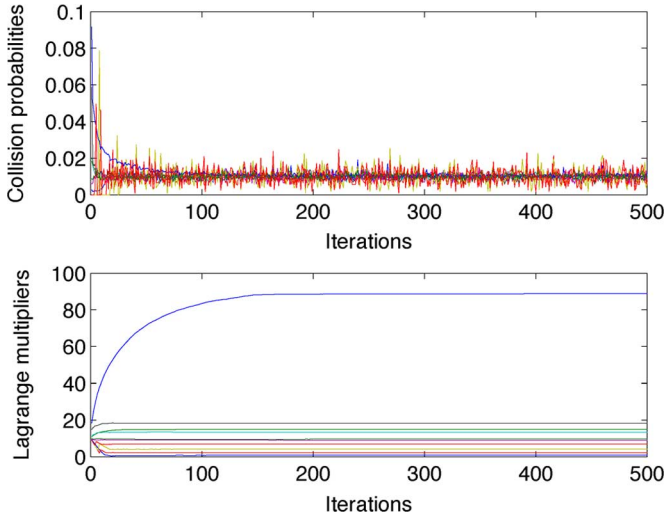


Fig. 1. Evolution of the collision probabilities (top) and the Lagrange multipliers (bottom). Different curves in each panel correspond to different channels $m$.

The deflection-optimal combiner weight vector $\bar{\mathbf{w}}^{(m)}$ is determined by $\bar{\mathbf{w}}^{(m)} = \bar{\mathbf{w}}^{(m)\prime}/||\bar{\mathbf{w}}^{(m)\prime}||$, where

$$\bar{w}^{(m)\prime}(i) = \frac{F_0^{(m)}(i) - F_1^{(m)}(i)}{\sqrt{\left[1 - F_1^{(m)}(i)\right] F_1^{(m)}(i)}}. \tag{77}$$

Using (75)–(76) and invoking the central limit theorem once more, the detection rule is $\bar{\boldsymbol{\delta}}_n^{\mathrm{FSS},(m)} \triangleq [\bar{\delta}_n^{\mathrm{FSS},(1)} \ \ldots \ \bar{\delta}_n^{\mathrm{FSS},(M)}]^T$, where $\bar{\delta}_n^{\mathrm{FSS},(m)} \triangleq \mathbb{1}_{\{\bar{t}_n^{(m)} < \bar{\gamma}_n^{(m)}\}}$. The decision thresholds $\bar{\gamma}_n^{(m)}$ and the false alarm probabilities $\bar{\alpha}_n^{(m)}$ are given by (68) and (70), respectively, with $k$ replaced by $n$ and all the tilded symbols by their barred counterparts. The associated throughput is given by [cf. (69)]

$$E_{\mathbf{G}}\left\{\max_{1\le n\le N, n\in\mathbb{N}} \frac{Th\tau}{T} \sum_{m=1}^{M} \hat{R}^{(m)} \pi_0^{(m)} \left(1 - \bar{\alpha}_k^{(m)}\right)\right\}. \tag{78}$$

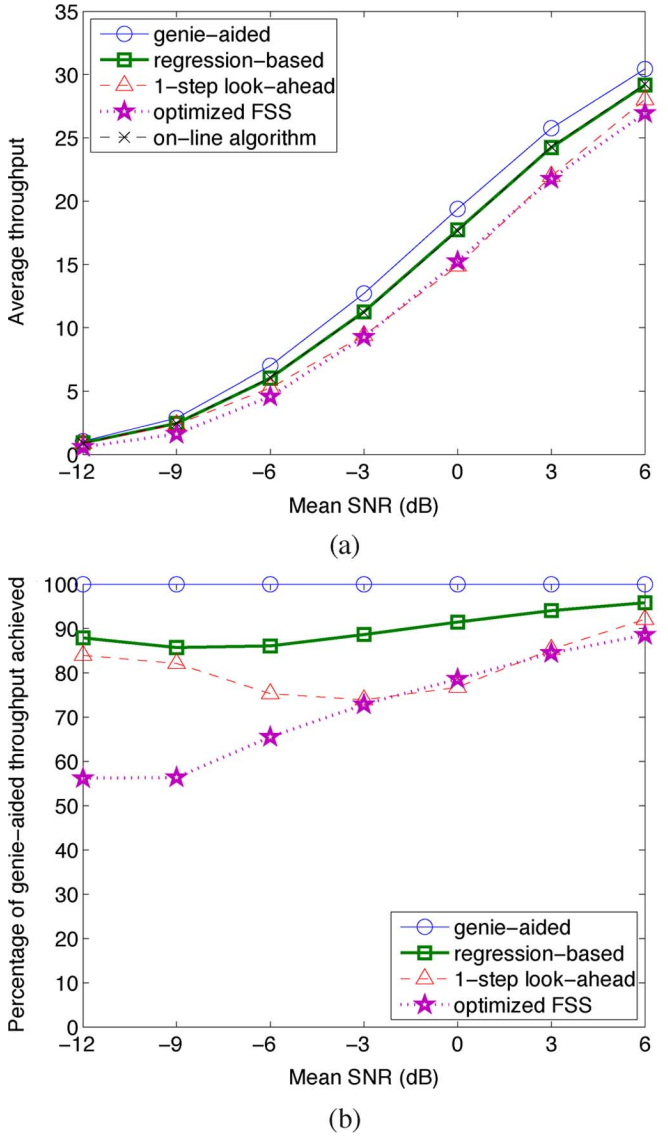


Fig. 2. Throughput performance of the single CR case. (a) Average throughput versus mean SNR. (b) Ratio of the average throughput achieved relative to the genie-aided throughput.

## VII. Numerical Tests

To test the proposed sequential sensing algorithms, scenarios with $M = 10, N = T = 100$, and $\tau = 1$ were considered. Rates $R^{(m)} = m$ for $m = 1, \ldots, M$ were used, and the channel gains $\{G^{(m)}\}$ were generated from the $\chi^2$-distribution. The observation noise variance $\sigma^2$ was set to $10^{-6}$. The prior probabilities were set to $\Pr\{H_0^{(m)}\} = 0.7$ for all $m$, and the collision probability target to $\bar{\beta} = 10^{-2}$.

### A. Single-CR Case

*1) Batch LS-Based Algorithm:* To estimate the $\mathbf{a}_n$ coefficient vector for the regression-based sub-optimal policy using the batch LS method, $J = 1\,000$ sample paths were generated.

Fig. 1 shows the collision probabilities $\{P_c^{(m)}\}$ in the $M = 10$ bands as a function of the iteration index $\ell$, along with the corresponding Lagrange multipliers, for the regression-based method at a mean SNR of $-3$ dB for all $M$ channels. The dual update iteration (24) is seen to converge relatively fast and the collision probability constraints are met in each band.

Fig. 2(a) plots the achieved average total rates of the proposed regression-based scheme when the mean SNRs of the PU-to-CR channels are varied. For comparison, the average throughput of the one-step look-ahead scheme of Section IV-B is shown. Also depicted is a genie-aided scheme, which chooses the stopping time $n$ that maximizes $g_n^*(\boldsymbol{\pi}_n; \boldsymbol{\lambda})$ over all $n \in \{1, 2, \ldots, N\}$, assuming non-causal knowledge of $\{\boldsymbol{\pi}_n\}$, and thus can be used to benchmark the achievable performance. Averaging was performed over 20 000 Monte Carlo realizations. It can be seen that the regression-based scheme attains throughput close to the genie-aided upperbound over a wide range of SNR values. The one-step look-ahead scheme is clearly sub-optimal especially in the moderate-to-high SNR range. Also, it can be seen that the proposed regression-based sequential sensing scheme outperforms the optimized FSS over all SNR levels tested.

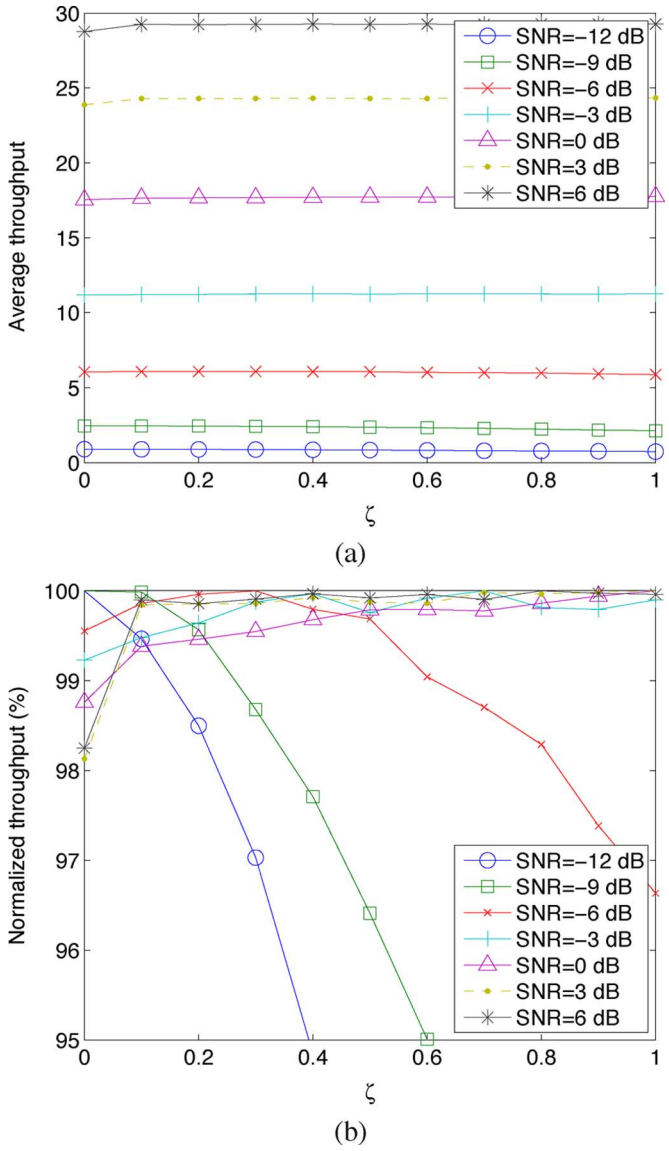


Fig. 3. Average throughput performance versus the parameter $\zeta$. (a) Average throughput. (b) Throughput normalized by the maximum throughput value at each SNR in percent.

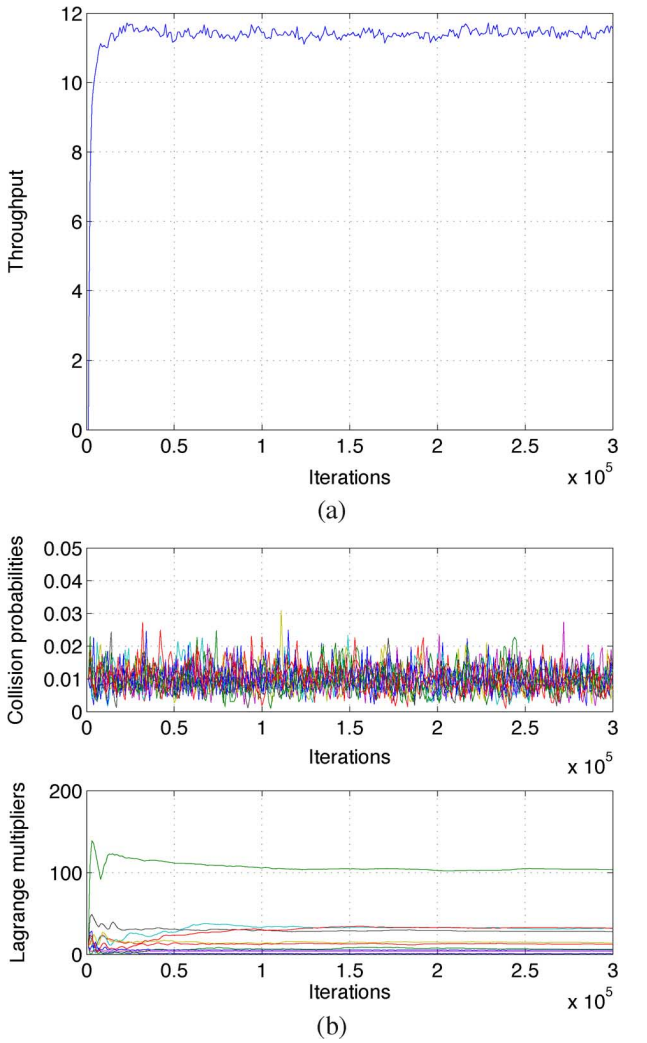


Fig. 4. Evolution of the throughput, the collision probabilities and the Lagrange multipliers for the on-line algorithm. (a) Throughput evolution. (b) Collision probabilities (top) and Lagrange multipliers (bottom).

Fig. 2(b) plots the ratio of the average throughput of the various schemes over 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.

*2) RLS-Based Algorithm:* Performance of the RLS-based on-line algorithm in Table I was tested with the forgetting factor $\xi$ set to 0.999.

In order to first determine an appropriate value of the margin $\zeta$, the algorithm in Table I was run for different values of $\zeta$ at a number of SNR conditions. It turns out that, as Fig. 3(a) shows, the throughput performance is quite insensitive to the choice of $\zeta$, although a slight increase in throughput is perceived especially at high SNR values when non-zero $\zeta$ values are used. An intuitive explanation is the following. As we initialized the value functions to zero (i.e., $\mathbf{a}_n = \mathbf{0}$ for all $n$), the algorithm naturally explores all the way to the end of the horizon in the early stages of iteration, even if $\zeta = 0$. In fact, when the observations are more noisy, which encourages exploratory behavior, smaller values of $\zeta$ tend to perform better. Fig. 3(b) plots the average throughput, normalized by the maximum throughput value corresponding to the optimal $\zeta$. Again, it is seen that as the SNR grows high, the throughput can be increased slightly by picking non-zero values of $\zeta$. However, since the performance is quite flat with respect to $\zeta$, the value $\zeta = 0.1$ will be used for the rest of the numerical results.

Fig. 4 shows a sample path of the on-line algorithm at $\mathrm{SNR} = -3$ dB. Fig. 4(a) depicts the evolution of the achieved throughput of the CR, while Fig. 4(b) the evolution of the collision probabilities and the Lagrange multipliers. The algorithm is seen to converge in about $10^5$ iterations with the collision probability constraints enforced on the average.

To assess the average throughput performance, the readers are again referred to Fig. 2(a), where it can be noticed that the performance of the on-line algorithm, represented by the dashed line with the 'x' marks, is virtually indistinguishable

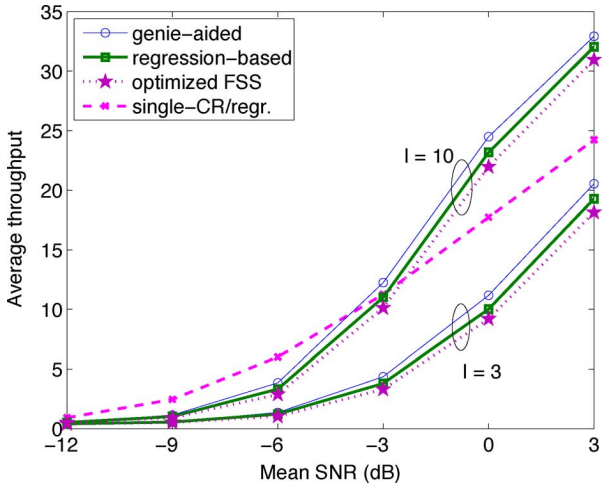


Fig. 5. Average throughput of the cooperative sensing scheme with raw observations.

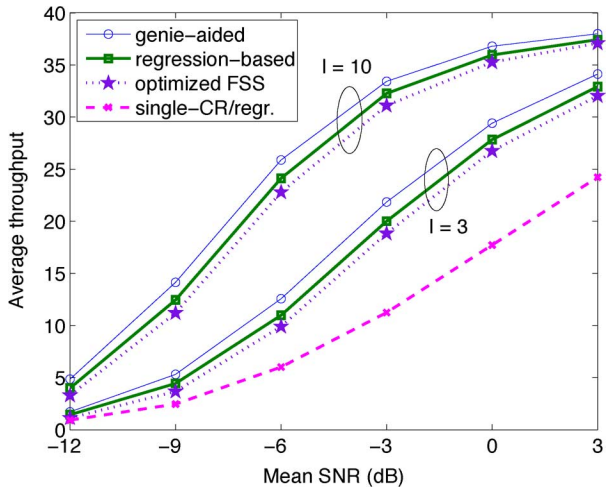


Fig. 6. Average throughput of the cooperative sensing scheme with quantized observations.

from that of the batch version. This is because the on-line algorithm learned the optimal Lagrange multipliers during the transient state. Thus, the more practical RLS-based algorithm incurs virtually no performance penalty compared to its batch counterpart in the steady state.

### B. Cooperative Sensing Case

Fig. 5 depicts the average throughput performance of the sequential cooperative sensing schemes when the FC collects the raw observations corrupted by AWGN with variance $(\check{\sigma}^{(m)}(i))^2 = 10^{-10}$ for all $m \in \{1,2,\ldots,M\}$ and $i \in \{1,2,\ldots,I\}$. Networks with for $I = 3$ and $I = 10$ cooperating CRs are considered. The value of $\eta$ was set to 3. It can be seen again that the proposed regression-based algorithm achieves a significant portion of the upper-bound set by the genie-aided scheme. However, due to the low receiver SNR at the FC, performance of the cooperative sensing with $I = 3$ is actually seen to be inferior to that of the single-CR case that uses the regression-based policy. The cooperative sensing performance can be improved by increasing the number of cooperating radios, or, by increasing the power for transmitting the observations to the FC.

Fig. 6 shows the throughput performance for the quantized observations case. Even though only 1 bit is transmitted per measurement taken at the individual CRs, the cooperative sensing policy with $I = 3$ already outperforms the single-CR benchmark. For $I = 10$, significant performance improvement is observed over the single-CR case.

Figs. 5 and 6 also verify that the sequential sensing algorithms outperform the optimized FSS test over all SNRs. Interestingly, however, it can be noted that the gain over the FSS test alternative is less than that of the single-CR case shown in Fig. 2. This is intuitively reasonable because the sequential detection algorithms yield performance advantage by exploiting the opportune situations where the received samples happen to strongly favor one of the two hypotheses. As the number of cooperating CRs increases, one can expect that the diversity effect actually saps the benefit of such opportunistic mechanisms.

## VIII. Conclusion

Sequential sensing algorithms for multi-channel CR systems have been developed. The trade-off between the sensing time and the chance of identifying more unoccupied channels was captured in the effective rate achieved by the CR system. A constrained DP problem was formulated to maximize the effective rate given past and current observations, under the constraints of specified probability of collision with on-going PU transmissions. The optimal access policy was derived in closed form given the Lagrange multipliers that were used to relax the constrained DP to an unconstrained DP.

The optimal stopping policy, which determines the number of samples required for sensing, was obtained from the solution of an optimal stopping problem. A basis expansion-based reduced-complexity solution was derived both in the batch and in the on-line forms, whose performance was shown to be close to the genie-aided upper-bounds, and hence close to that of the optimal solution. Extensions to cooperative sensing cases were also studied using the framework similar to the single-CR case. It was illustrated through numerical simulations that using only 1-bit quantized samples collected from individual CRs, yields significant performance gains in sequential CR sensing.

## Appendix A
## Derivation of Optimal Policies (27) and (28)

By applying the standard backward induction to the DP formulation in (23), the optimal control policies, call them $\{\Delta'_n, \boldsymbol{\delta}'_n\}$, can be obtained by solving the following maximization recursively:

$$V'_N(\boldsymbol{\mathcal{S}}_N;\boldsymbol{\lambda}) \triangleq \max_{\boldsymbol{\delta}_N} \mathbb{1}_{\{\boldsymbol{\mathcal{D}}_{N-1}=\bar{S}\}} g_N(\boldsymbol{\pi}_N,\boldsymbol{\delta}_N;\boldsymbol{\lambda}) \tag{79}$$

$$V'_n(\boldsymbol{\mathcal{S}}_n;\boldsymbol{\lambda}) \triangleq \max_{\Delta_n,\boldsymbol{\delta}_n} \Big\{ \mathbb{1}_{\{\boldsymbol{\mathcal{D}}_{n-1}=\bar{S},\Delta_n=S\}} g_n(\boldsymbol{\pi}_n,\boldsymbol{\delta}_n;\boldsymbol{\lambda}) + E_{\mathbf{y}_{n+1}|\mathbf{Y}_n}\{V'_{n+1}(\boldsymbol{\mathcal{S}}_{n+1};\boldsymbol{\lambda})|\mathbf{Y}_n\}\Big\}, \quad n = N-1, N-2, \ldots, 1 \tag{80}$$

where the state definition $\boldsymbol{\mathcal{S}}_n \triangleq [\mathbf{Y}_n,\ \boldsymbol{\mathcal{D}}_{n-1}]$ has been used.

The optimal controls $\Delta'_n(\boldsymbol{\mathcal{S}}_n)$ and $\boldsymbol{\delta}'_n(\boldsymbol{\mathcal{S}}_n)$ after sensing is stopped, i.e., when $\boldsymbol{\mathcal{D}}_{n-1} \neq \bar{S}$, do not affect the throughput since they contribute zero summands in the reward; in fact, they are never used for actual sensing. Thus, one only needs to verify that when $\boldsymbol{\mathcal{D}}_{n-1} = \bar{S}$, the optimal policies $\{\Delta'_n(\boldsymbol{\mathcal{S}}_n), \boldsymbol{\delta}'_n(\boldsymbol{\mathcal{S}}_n)\}$

$$V_n'(\boldsymbol{\mathcal{S}}_n;\boldsymbol{\lambda}) = \max_{\Delta_n}\left\{\max_{\boldsymbol{\delta}_n} \mathbb{1}_{\{\boldsymbol{\mathcal{D}}_{n-1}=\bar{S},\Delta_n=S\}} g_n(\boldsymbol{\pi}_n,\boldsymbol{\delta}_n;\boldsymbol{\lambda}) + E_{\mathbf{y}_{n+1}|\mathbf{Y}_n}\{V_{n+1}'(\boldsymbol{\mathcal{S}}_{n+1};\boldsymbol{\lambda})|\mathbf{Y}_n\}\right\} \quad (81)$$

$$= \begin{cases} \max\left\{\max_{\boldsymbol{\delta}_n} g_n(\boldsymbol{\pi}_n,\boldsymbol{\delta}_n;\boldsymbol{\lambda}), E_{\mathbf{y}_{n+1}|\mathbf{Y}_n}\{V_{n+1}'([\mathbf{Y}_{n+1},[\boldsymbol{\mathcal{D}}_{n-1}\ \bar{S}]];\boldsymbol{\lambda})|\mathbf{Y}_n\}\right\} & \text{if } \boldsymbol{\mathcal{D}}_{n-1}=\bar{S} \\ 0 & \text{if } \boldsymbol{\mathcal{D}}_{n-1}\neq\bar{S} \end{cases}$$

$$n = N-1, N-2, \ldots, 1. \quad (82)$$

maximizing (79)–(80) are identical to $\{\Delta_n^*(\boldsymbol{\pi}_n), \boldsymbol{\delta}_n^*(\boldsymbol{\pi}_n)\}$, which maximize (25)–(26).

It is clear that the statement holds for $n = N$ by comparing (25) and (79). To verify the same for $n < N$, consider re-writing (80) as (81)–(82), shown at the top of the page. By comparing (26) with (82), one can inductively deduce that $V_n'(\boldsymbol{\mathcal{S}}_n;\boldsymbol{\lambda}) = V_n(\boldsymbol{\pi}_n;\boldsymbol{\lambda})$ for $n = N-1, N-2, \ldots, 1$, provided that $\boldsymbol{\mathcal{D}}_{n-1} = \bar{S}$. Hence, the optimal control policies are also identical.

## Appendix B
## Example Basis Functions

For the numerical results in Section VII, we used the following $K = 10$ basis functions: $\phi_{n,1}(\boldsymbol{\pi}_n) = 1, \phi_{n,2}(\boldsymbol{\pi}_n) = g_n^*(\boldsymbol{\pi}_n), \phi_{n,3}(\boldsymbol{\pi}_n) = (1/M)\sum_{m=1}^M \pi_n^{(m)}, \phi_{n,4}(\boldsymbol{\pi}_n) = (1/M)\sum_{m=1}^M \pi_n^{(m)}\hat{m}, \phi_{n,5}(\boldsymbol{\pi}_n) = (1/M)\sum_{m=1}^M \pi_n^{(m)}((3\hat{m}^2)/(2) - (1/2)), \phi_{n,6}(\boldsymbol{\pi}_n) = (1/M)\sum_{m=1}^M \pi_n^{(m)}((5\hat{m}^3)/(2) - (3\hat{m})/(2))$ and $\phi_{n,k}(\boldsymbol{\pi}_n) = \phi_{n,2}(\boldsymbol{\pi}_n)\phi_{n,k-4}(\boldsymbol{\pi}_n), k = 7, \ldots, 10$, where $n \in \{1,2,\ldots,N\}$ and $\hat{m} \triangleq (2(m-1))/(M-1) - 1$. The basis functions $\{\phi_{n,k}, k = 3,\ldots,6\}$ are defined as the inner products of Legendre polynomials of order 0, 1, 2 and 3 with $\{\pi_n^{(m)}\}$; see also [31] for additional issues on selecting $\phi_{n,k}$ and the expansion order $K$.

## Appendix C
## Derivation of (47)

First, note that

$$E_{\mathbf{y}_{n+1}|\mathbf{Y}_n}\{g_{n+1}^*(\boldsymbol{\Phi}(\boldsymbol{\pi}_n,\mathbf{y}_{n+1});\boldsymbol{\lambda})|\mathbf{Y}_n\}$$

$$= \int \sum_{m=1}^M \left(\frac{T-(n+1)\tau}{T} R^{(m)}\pi_{n+1}^{(m)} - \lambda^{(m)}\frac{1-\pi_{n+1}^{(m)}}{1-\pi_0^{(m)}}\right)^+ \cdot p(\mathbf{y}_{n+1}|\mathbf{Y}_n)d\mathbf{y}_{n+1} \quad (83)$$

$$= \sum_{m=1}^M \int \left(\frac{T-(n+1)\tau}{T} R^{(m)}\pi_{n+1}^{(m)} - \lambda^{(m)}\frac{1-\pi_{n+1}^{(m)}}{1-\pi_0^{(m)}}\right)^+ \cdot p\left(y_{n+1}^{(m)}|y_n^{(m)},\ldots,y_1^{(m)}\right) dy_{n+1}^{(m)}. \quad (84)$$

For the generic $m$th term of (84), consider the case where $\lambda^{(m)} > 0$. Recalling the definition in (31) and (32), the $m$th summand in (84) can be re-written as

$$\int \left(\gamma_{n+1}^{(m)} - \Gamma_{n+1}^{(m)}\right)^+ \frac{\lambda^{(m)}}{\pi_0^{(m)}}\Phi^{(m)}\left(\pi_n^{(m)}, y_{n+1}^{(m)}\right) \cdot p\left(y_{n+1}^{(m)}\middle| y_n^{(m)},\ldots,y_1^{(m)}\right) dy_{n+1}^{(m)} \quad (85)$$

$$= \int \left(\gamma_{n+1}^{(m)} - \tilde{\Gamma}^{(m)}(y_{n+1}^{(m)})\cdot\Gamma_n^{(m)}\right)^+ \frac{\lambda^{(m)}\pi_n^{(m)}}{\pi_0^{(m)}} \cdot p\left(y_{n+1}^{(m)}\middle| H_0^{(m)}\right) dy_{n+1}^{(m)} \quad (86)$$

where $\tilde{\Gamma}^{(m)}(y_{n+1}^{(m)}) \triangleq (p(y_{n+1}^{(m)}|H_1^{(m)}))/(p(y_{n+1}^{(m)}|H_0^{(m)})) = e^{-(G^{(m)})/(\sigma^2)} I_0((\sqrt{G^{(m)}y_{n+1}^{(m)}})/(\sigma^2/2))$ is monotone in $y_{n+1}^{(m)}$. Thus, with $\tilde{\theta}_{n+1}^{(m)}$ defined as in (48), the integral in (86) can be expressed as

$$\int_0^{\tilde{\theta}_{n+1}^{(m)}} \left(\gamma_{n+1}^{(m)} - \tilde{\Gamma}^{(m)}\left(y_{n+1}^{(m)}\right)\cdot\Gamma_n^{(m)}\right)\frac{\lambda^{(m)}\pi_n^{(m)}}{\pi_0^{(m)}} \cdot p\left(y_{n+1}^{(m)}\middle| H_0^{(m)}\right) dy_{n+1}^{(m)}$$

$$= \frac{T-(n+1)\tau}{T} R^{(m)}\pi_n^{(m)} F\left(\tilde{\theta}_{n+1}^{(m)}\middle| H_0^{(m)}\right) - \frac{\lambda^{(m)}\left(1-\pi_n^{(m)}\right)}{1-\pi_0^{(m)}} F\left(\tilde{\theta}_{n+1}^{(m)}\middle| H_1^{(m)}\right) \quad (87)$$

which verifies the $m$th term of (87). It is also easily verified that when $\lambda^{(m)} \downarrow 0$ the $m$th term of (84) is equal to $(T-(n+1)\tau)/(T)R^{(m)}\pi_n^{(m)}$, which is also equal to (87) with $\lambda^{(m)} \downarrow 0$ and $\tilde{\theta}_{n+1}^{(m)} = +\infty$.

## References


[1] D. A. Berry and B. Fristedt, *Bandit Problems: Sequential Allocation of Experiments*. New York: Chapman & Hall, 1985.

[2] D. P. Bertsekas, "Convergence of discretization procedures in dynamic programming," *IEEE Trans. Autom. Control*, vol. 20, no. 3, pp. 415–419, Jun. 1975.

[3] D. P. Bertsekas, *Dynamic Programming and Optimal Control*, 2nd ed. Belmont, MA: Athena Scientific, 2000, vol. 1.

[4] D. Cabric, A. Tkachenko, and R. W. Brodersen, "Spectrum sensing measurements of pilot, energy, and collaborative detection," in *Proc. MILCOM Conf.*, Washington, DC, Oct. 2006, pp. 1–7.

[5] D. A. Castañon, "Approximate dynamic programming for sensor management," in *Proc. 36th Conf. Decision Control*, San Diego, CA, Dec. 1997, vol. 2, pp. 1202–1207.

[6] S. Chaudhari, V. Koivunen, and H. V. Poor, "Autocorrelation-based decentralized sequential detection of OFDM signals in cognitive radios," *IEEE Trans. Signal Process.*, vol. 57, no. 7, pp. 2690–2700, Jul. 2009.

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

[8] C. Cordeiro, K. Challapali, D. Birru, and S. Shankar, "IEEE 802.22: the first worldwide wireless standard based on cognitive radios," in *Proc. DySPAN Conf.*, Baltimore, MD, Nov. 2005, pp. 328–337.

[9] T. M. Cover and . A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.

[10] T. S. Ferguson, Optimal Stopping and Applications. [Online]. Available: http://www.math.ucla.edu/tom/Stopping/Contents.html

[11] S. Geirhofer, L. Tong, and B. Sadler, "Dynamic spectrum access in the time domain: Modeling and exploiting white space," *IEEE Commun. Mag.*, vol. 45, no. 5, pp. 66–72, May 2007.

[12] J. Jia, Q. Zhang, and X. Shen, "HC-MAC: A hardware-constrained cognitive MAC for efficient spectrum management," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 1, pp. 106–117, Jan. 2008.

[13] H. Jiang, L. Lai, R. Fan, and H. V. Poor, "Optimal selection of channel sensing order in cognitive radio," *IEEE Trans. Wireless Commun.*, vol. 8, no. 1, pp. 297–307, Jan. 2009.
[14] S. M. Kay, *Fundamentals of Statistical Signal Processing: Detection Theory*. Upper Saddle River, NJ: Prentice-Hall, 1993, vol. 2.
[15] S.-J. Kim and G. B. Giannakis, "Rate-optimal and reduced-complexity sequential sensing algorithms for cognitive OFDM radios," *EURASIP J. Adv. Sig. Proc.*, 2009, article ID 421540, 11 pages.
[16] N. Kundargi and A. Tewfik, "Hierarchical sequential detection in the context of dynamic spectrum access for cognitive radios," in *Proc. of the Intl. Conf. on Electr., Circuits and Syst.*, Marrakech, Morocco, Dec. 2007, pp. 514–517.
[17] M. A. Lexa and D. H. Johnson, "Distributed structures, sequential optimization, and quantization for detection," *IEEE Trans. Signal Process.*, vol. 56, no. 4, pp. 1740–1745, Apr. 2008.
[18] Y.-C. Liang, Y. Zeng, E. C. Y. Peh, and T. Hoang, "Sensing-throughput trade off for cognitive radio networks," *IEEE Trans. Wireless Commun.*, vol. 7, no. 4, pp. 1326–1337, Apr. 2008.
[19] M. Longo, T. D. Lookabaugh, and R. M. Gray, "Quantization for decentralized hypothesis testing under communication constraints," *IEEE Trans. Inf. Theory*, vol. 36, no. 2, pp. 241–255, Mar. 1990.
[20] F. A. Longstaff and E. S. Schwartz, "Valuing American options by simulation: A simple least-squares approach," *Rev. Financ. Stud.*, vol. 14, no. 1, pp. 113–147, Spring 2001.
[21] J. M. Mendel, *Lessons in Estimation Theory for Signal Processing, Communications, and Control*. Upper Saddle River, NJ: Prentice-Hall, 1995.
[22] M. Öner and F. Jondral, "On the extraction of the channel allocation information in spectrum pooling systems," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 3, pp. 558–565, Apr. 2007.
[23] H. V. Poor and J. B. Thomas, "Applications of Ali-Silvey distance measures in the design of generalized quantizers for binary decision systems," *IEEE Trans. Commun.*, vol. COM-25, no. 9, pp. 893–900, Sep. 1977.
[24] W. B. Powell, *Approximate Dynamic Programming: Solving the Curses of Dimensionality*. Hoboken, NJ: Wiley, 2007.
[25] Z. Quan, S. Cui, and A. H. Sayed, "Optimal linear cooperation for spectrum sensing in cognitive radio networks," *IEEE J. Sel. Topics Signal Process.*, vol. 2, no. 1, pp. 28–39, Feb. 2008.
[26] Z. Quan, S. Cui, A. H. 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.
[27] A. Sahai, N. Hoven, M. Mishra, and R. Tandra, "Fundamental tradeoffs in robust spectrum sensing for opportunistic frequency reuse," Univ. Calif., Berkeley, CA, Tech. Rep., 2006.
[28] N. Z. Shor, *Minimization Methods for Non-Differentiable Functions*. New York: Springer-Verlag, 1985.
[29] D. Siegmund, *Sequential Analysis*. New York: Springer-Verlag, 1985.
[30] R. D. Smallwood and E. J. Sondik, "The optimal control of partially observable Markov processes over a finite horizon," *Oper. Res.*, vol. 21, no. 5, pp. 1071–1088, Sept.–Oct. 1973.
[31] 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 Trans. Autom. Control*, vol. 44, no. 10, pp. 1840–1851, Oct. 1999.
[32] J. N. Tsitsiklis and B. Van Roy, "Regression methods for pricing complex American-style options," *IEEE Trans. Neural Netw.*, vol. 12, no. 4, pp. 694–703, Jul. 2001.
[33] J. Unnikrishnan and V. V. Veeravalli, "Cooperative sensing for primary detection in cognitive radio," *IEEE J. Sel. Topics Signal Process.*, vol. 2, no. 1, pp. 18–27, Feb. 2008.
[34] J. Unnikrishnan and V. V. Veeravalli, "Algorithms for dynamic spectrum access with learning for cognitive radio," *IEEE Trans. Signal Process.*, vol. 58, no. 2, pp. 750–760, Feb. 2010.
[35] H. Urkowitz, "Energy detection of unknown deterministic signals," *Proc. IEEE*, vol. 55, no. 4, pp. 523–531, Apr. 1967.
[36] Y. Xin and H. Zhang, "A simple sequential spectrum sensing scheme for cognitive radio," *IEEE Trans. Signal Process.*, May 2009, submitted for publication.
[37] Q. Zhao, L. Tong, A. Swami, and Y. Chen, "Decentralized cognitive MAC for opportunistic spectrum access in ad hoc networks: A POMDP framework," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 3, pp. 589–600, Apr. 2007.
[38] Q. Zhao and J. Ye, "When to quit for a new job: Quickest detection of spectrum opportunities in multiple channels," presented at the MILCOM Conf., San Diego, CA, Nov. 2008.

**Seung-Jun Kim** received the B.S. and M.S. degrees from Seoul National University, Seoul, Korea in 1996 and 1998, respectively, and the Ph.D. degree from the University of California at Santa Barbara in 2005, all in electrical engineering.

From 2005 to 2008, he worked for NEC Laboratories America, Princeton, NJ, initially as a Visiting Researcher, and then as a Research Staff Member. Since 2008, he has been with the Department of Electrical and Computer Engineering at the University of Minnesota, where he is currently a Research Assistant Professor. His research interests lie in wireless communications and networking, statistical signal processing, and optimization.

**Georgios B. Giannakis** (F'97) received the Diploma degree in electrical engineering from the National Technical University of Athens, Greece, in 1981 and the M.Sc. degree in electrical engineering, the M.Sc. degree in mathematics, and the Ph.D. degree in electrical engineering from the University of Southern California (USC) in 1983, 1986 and 1986, respectively.

Since 1999, he has been a Professor with the University of Minnesota, where he now holds an ADC Chair in Wireless Telecommunications in the Electric and Computer Engineering Department and serves as Director of the Digital Technology Center. His general interests span the areas of communications, networking and statistical signal processingsubjects on which he has published more than 285 journal papers, 485 conference papers, two edited books, and two research monographs. Current research focuses on compressive sensing, cognitive radios, network coding, cross-layer designs, mobile ad hoc networks, wireless sensor, and social networks.

Dr. Giannakis is the (co)recipient of seven paper awards from the IEEE Signal Processing (SP) and Communications Societies, including the G. Marconi Prize Paper Award inWireless Communications. He also received Technical Achievement Awards from the SP Society (2000), from EURASIP (2005), a Young Faculty Teaching Award, and the G. W. Taylor Award for Distinguished Research from the University of Minnesota. He is a Fellowof EURASIP, has served the IEEE in a number of posts, and is also as a Distinguished Lecturer for the IEEE-SP Society.