Jointly Optimal Sensing and Resource Allocation for Multiuser Interweave Cognitive Radios

Luis M. Lopez-Ramos, *Student Member, IEEE,* Antonio G. Marques, *Member, IEEE,* and Javier Ramos

Abstract—Successful deployment of cognitive radios requires efficient sensing of the spectrum and dynamic adaptation of the available resources according to the sensed (imperfect) information. While most works design these two tasks separately, in this paper the sensing and resource allocation schemes are designed jointly. We investigate an interweave CR with multiple secondary users that access orthogonally a set frequency bands originally devoted to primary users. The schemes are designed to optimize the performance of the secondary users while limiting the "probability of interfering" the primary users. The joint design is addressed using dynamic programming and nonlinear optimization techniques. A two-step strategy that first finds the optimal resource allocation for any sensing scheme and then uses that solution as input to solve for the optimal sensing policy is implemented. The two-step strategy is optimal and entails a computational complexity much lower than that required to solve the original formulation.

Index Terms—Cognitive radios, dynamic programming, imperfect channel state information, resource management.

I. INTRODUCTION

Cognitive radios (CRs) are viewed as the next-generation solution to alleviate the perceived spectrum scarcity. When CRs are deployed, the secondary users (SUs) have to sense their radio environment to optimize their communication performance while avoiding (limiting) the interference to the primary users (PUs). As a result, efficient operation of CR requires the implementation of two critical tasks: i) sensing the spectrum and ii) dynamic adaptation of the available resources according to the sensed information [1]. To carry out the sensing task two important challenges are: C1) the presence of errors in the measurements that lead to errors on the channel occupancy detection; and C2) the inability to sense the totality of the time-frequency lattice due to scarcity of resources (time, energy or sensing devices). Two additional challenges that arise to carry out the resource allocation (RA) task are: C3) the ability of the RA algorithms to deal with channel imperfections; and C4) the selection of metrics that properly quantify the reward for the SUs and the damage for the PUs.

Many alternatives have been proposed in the CR literature to deal with these challenges. Regarding C1, some works deal with noisy CSI [3] or quantized CSI [5]. However, in the context of CR only a few works have considered the fact that the CSI may be not only noisy but also outdated, or have incorporated those imperfections into the design of

resource allocation algorithms [6]. Alternatives to deal with C2 based on convex optimization [7] and dynamic programming (DP) have also been explored [6]. Regarding C3, many works consider that the CSI is imperfect, but only a few exploit the statistical model of these imperfections (especially for the time correlation) to mitigate them; see, e.g., [6], [14]. Finally, different alternatives have been considered to limit the harm that the SUs cause to the PUs [13]. The most widely used is to set limits on the peak (instantaneous) and average interfering power. Some works also have tried to limit the rate loss that PU experience [5], while others look at limiting the instantaneous or average probability of interfering the PU (bounds on the short-term or long-term outage probability) [2], [14].

The objective of this work is to design the sensing and the RA policies *jointly* while accounting for the challenges C1-C4. Only a few works have addressed the joint design of the sensing and RA policies [7], [8]. Those works consider operating conditions different than those in this paper, which are described next. An interweave CR with multiple SUs and PUs is considered. SUs are able to adapt their power and rate loadings and access orthogonally a set frequency bands. Those bands are originally devoted to primary transmissions. Orthogonal here means that if a secondary user is transmitting, no other secondary user can be active in the same band. The schemes are designed to maximize the sum-average rate of the SUs while adhering to constraints that *limit* the maximum "average power" that SUs transmit and the average "probability of interfering" the PUs. It is assumed that the CSI of the SU links is instantaneous and free of errors, while the CSI of the PUs activity is outdated and noisy. A simple first-order hidden Markov model is used to characterize such imperfections. Sensing a channel band entails a given cost, and at each instant the system has to decide which channels (if any) are sensed.

The jointly optimal sensing and RA schemes will be designed using dynamic programming (DP) and nonlinear optimization techniques. DP techniques are required because the activity of PUs is assumed to be correlated across time, so that sensing a channel has an impact not only for the current instant, but also for future time instants. See, e.g., [15], [8] for relevant examples that have applied DP techniques to design CR schemes. To solve the joint design, a two-step strategy is implemented. First, the sensing is considered given and the optimal RA is found for *any* fixed sensing scheme. This first problem was recently solved in [9], [14]. Then, the results of the first step are used as input to obtain the optimal sensing policy. The motivation for using this strategy is twofold. First, while the joint design is non convex and has to be solved using DP techniques, the problem in the first step (optimum

RA for a fixed sensing scheme) is convex. Second, when the optimal RA is substituted back into the original joint design, the resulting problem (which has to be solved over the sensing policy) has a more favorable structure. Specifically, while the original design problem was a constrained DP, the updated one is an unconstrained DP problem which can be solved separately for each of the channels.[1]

II. SYSTEM MODEL

Consider a CR scenario with M secondary users (indexed by m) transmitting opportunistically over a frequency band divided into K orthogonal sub-channels (indexed by k). For simplicity, we consider that each band has the same bandwidth and is occupied by a different PU. We also assume that there is a secondary base station (SBS) which is the destination of all SU transmissions, acts as a central scheduler, collects the CSI, and also performs the task of sensing the medium for primary presence. Although this work focuses on an uplink setup, the results can be easily extended to downlink and any-to-any setups.

In this section we: a) describe the model for the CSI of the secondary (S-CSI) and primary (P-CSI) networks, paying special attention to the sensing errors; b) describe the variables to be designed; and c) formulate the constraints that such variables need to satisfy. The constraints will account for the operating conditions of the SUs, the quality of service requirements and the rules the SUs need to implement to limit the damage (interference) caused to the PUs.

Starting with a), we first present the model for the S-CSI. Specifically, the power gain of the channel between the mth SU and the SBS in the kth channel at time n is denoted as $h_k^m[n]$; it represents the noise-normalized square magnitude of the fading coefficient. Channels are assumed to be ergodic and independent across bands and time, and the exact value of their instantaneous gain is assumed to be available at every time instant. Now, we address the problem of modeling the P-CSI. Since we consider an interweave scenario, it suffices to know whether a given channel is occupied or not. For that purpose, let $a_k[n]$ denote a binary variable which is one if the kth primary link is active at time n and zero otherwise. The process $a_k[n]$ is modeled as a two-state, time invariant Markov chain. With $P_{xy} = \Pr(a_k[n] = x | a_k[n-1] = y)$, the time dynamics of a channel are fully described by the transition matrix $\mathbf{P} = [P_{00}, P_{01}; P_{10}, P_{11}]$. The Markovian property will be useful to keep the DP modeling simple, and will also allow to recursively keep track of the P-CSI [15]. The results can be extended to account for more realistic occupation models. The main price in such a case is an increase on the computational load required to solve the DP.

In our model, we suppose that the SBS is equipped with sensors to measure $a_k[n]$. However, presence of imperfections render deterministic knowledge of $a_k[n]$ impossible. Two sources of imperfections are considered here: i) errors in the sensing process and ii) outdated information (because the

channels are not always sensed). For that purpose let $s_k[n]$ denote a binary variable which is one if the kth channel is sensed at time n and zero otherwise. Moreover, let $z_k[n]$ denote the output of the sensor if indeed $s_k[n] = 1$; i.e., if the kth channel is sensed. We will assume that the output of the sensor is binary and may contain errors. To model the sensing errors, consider the probabilities of false alarm $P_k^{FA} = \Pr(z_k[n] = 1 | a_k = 0)$ and miss detection $P_k^{MD} = \Pr(z_k[n] = 0 | a_k = 1)$. Such probabilities are assumed to be known and constant across time.

Due to the aforementioned sources of imperfections, at time n the SBS does not know the actual value of $a_k[n]$. Rather, only a probabilistic description of $a_k[n]$ is available. The knowledge about $a_k[n]$ will be referred to as (instantaneous) belief $\mathbf{b}_k[n]$. Basically, $\mathbf{b}_k[n]$ contains the probabilities of channel k being idle and busy conditioned to *all past measurements*. Mathematically, $\mathbf{b}_k[n] := [\Pr(a_k[n] = 0 | n-1), \Pr(a_k[n] = 1 | n-1)]$. To account for the instants where $s_k[n] = 0$, we will refer to $\mathbf{b}_k[n]$ as the *pre-decision* belief, and introduce $\mathbf{b}_k^S[n]$ as the *post-decision* belief. Intuitively, $\mathbf{b}_k[n]$ contains the information about $a_k[n]$ at instant $n-1$, while $\mathbf{b}_k^S[n]$ contains the information about $a_k[n]$ once $s_k[n]$ and $z_k[n]$ (if $s_k[n] = 1$) are known. Clearly, using the time-correlation model, the expression to get the pre-decision belief is $\mathbf{b}_k[n] = \mathbf{P}_k \mathbf{b}_k^S[n-1]$. Differently, the expression to get $\mathbf{b}_k^S[n]$, depends on the sensing decision $s_k[n]$. If $s_k[n] = 0$, no additional information is available, so that $\mathbf{b}_k^S[n] = \mathbf{b}_k[n]$. If $s_k[n] = 1$, the sensor output $z_k[n]$ can be either 0 (idle) or 1 (busy), and the belief is corrected as follows:

$$\mathbf{b}_k^S[n] = \mathbf{b}_k^S[n]\Big(\mathbf{b}_k[n], z\Big) = \frac{\mathbf{D}_z \mathbf{b}_k[n]}{\Pr(z_k[n] = z | \mathbf{b}_k[n])} \quad (1)$$

where $\mathbf{D}_z$ with $z \in \{0, 1\}$ is a 2×2 diagonal matrix with entries $[\mathbf{D}_z]_{1,1} := \Pr(z_k[n] = z | a_k = 0)$ and $[\mathbf{D}_z]_{2,2} := \Pr(z_k[n] = z | a_k = 1)$. The denominator is the likelihood of the sensor outcome, and it can be calculated as $\Pr(z_k[n] = z | \mathbf{b}_k[n]) = \mathbf{1}^T \mathbf{D}_z \mathbf{b}_k[n]$. Finally, if no information about the initial state is available, $\mathbf{b}_k[0]$ is initialized as the stationary distribution of the Markov chain associated to the channel.

Vectors $\mathbf{b}_k[n]$ and $\mathbf{b}_k^S[n]$ represent the P-CSI at instant n and will be used as input to design our optimal schemes. For mathematical convenience, vectors $\mathbf{b}_k[n]$ and $\mathbf{b}_k^S[n]$ (for all k) and variables $h_k^m[n]$ (for all k, m) are gathered into vector $\mathbf{i}[n]$. Clearly, $\mathbf{i}[n]$ represents the overall CSI of the system.

Now, we move to the description of the variables to be designed [cf. b)]. The first set of variables to be optimized is the sensing decision variables $s_k[n]$. For such a purpose, we define ξ_k as the (fixed) positive cost which is paid every time that $s_k[n] = 1$. Clearly, if $\xi_k = 0$, then the optimal solution is to set $s_k[n] = 1$ for all n. However, if $\xi_k > 0$, then the system has to decide whether to sense or not. As it will be shown in the upcoming sections, such a decision depends on the level of CSI uncertainty and the potential benefits that using the channel would bring to the system. The second set of variables to be optimized is the access (scheduling) decision variables $w_k^m[n]$. Specifically, $w_k^m[n]$ is one if the mth SU is scheduled to transmit into the kth band at time n and zero otherwise. The third and last set of variables is the

---

[1]*Notation:* x^* denotes the optimal value of variable x; $\mathbb{E}[\cdot]$ expectation; $\wedge$ the boolean "and" operator; $\mathbb{1}_{\{\cdot\}}$ the indicator function ($\mathbb{1}_{\{x\}} = 1$ if x is true and zero otherwise); and $[x]_+$ the projection of x onto the non-negative orthant, i.e., $[x]_+ := \max\{x, 0\}$.

power loadings $p_k^m[n]$. Specifically, provided that $w_k^m[n] = 1$, variable $p_k^m[n]$ denotes the instantaneous power transmitted over the kth band by the mth SU. Under bit error rate or capacity constraints, instantaneous rate and power variables are coupled. This rate-power coupling will be represented by the function $C_k^m(h_k^m[n], p_k^m[n])$. Throughout this paper it is assumed that the rate-power function $C_k^m(h_k^m[n], \cdot)$ is given by Shannon's capacity formula $\log(1 + h_k^m[n]p_k^m[n]/\Gamma_k^m)$, where Γ_k^m represents the SNR-gap which depends on the coding scheme implemented [10]. At every time instant n, the $\mathbf{i}[n]$ will be used to find the (optimum) value of $w_k^m[n]$, $p_k^m[n]$ and $s_k[n]$. Throughout the manuscript, we will write $\mathbf{i}$, $s_k(\mathbf{i})$, $w_k^m(\mathbf{i})$ and $p_k^m(\mathbf{i})$, or $\mathbf{i}[n]$, $s_k[n]$, $w_k^m[n]$, and $p_k^m[n]$, wherever is convenient to emphasize the corresponding dependence.

Once the design variables have been introduced, we identify the different constraints that the optimal schemes need to satisfy [cf. c)]. The variable $s_k^m[n]$ is constrained to be boolean, so that $s_k[n] \in \{0,1\} \; \forall k$. The variable $w_k^m[n]$ needs to obey two constraints: i) being a boolean variable, so that $w_k^m[n] \in \{0,1\} \; \forall m,k$; and ii) the fact that the access is orthogonal, so that only one user can access the channel. Mathematically, the latter requires

$$\sum_m w_k^m[n] \leq 1 \;\; \forall k, n. \qquad (2)$$

The constraints that the power variables need to satisfy are: i) $p_k^m[n] \geq 0$; ii) each user is allowed to transmit a maximum long-term power, denoted by $\check{p}^m$. To formulate the second constraint, we consider the *normalized* exponentially decreasing window $(1-\gamma)\gamma^n$ with $0 < \gamma < 1$, and require the satisfaction of the following M constraints

$$\mathbb{E}\left[\lim_{N\to\infty}\sum_{n=0}^{N-1}(1-\gamma)\gamma^n\sum_k w_k^m[n]p_k^m[n]\right] \leq \check{p}^m, \;\; \forall m, \quad (3)$$

which also involves the scheduling variables $w_k^m[n]$. Note that (3) is an average constraint (it needs to hold on the long-term), while (2) is an instantaneous (short-term) constraint.

The harm (interference) that SUs cause to PU must be kept under control. Interference occurs when $a_k[n] = 1$ and $\sum_m w_k^m[n] = 1$ (a secondary user is transmitting into the kth band). In this work, we are interested in limiting the probability of SUs interfering each of the PUs. Note that since errors in sensing are inevitable, a zero probability of interference can not be guaranteed. Specifically, if $\check{o}_k$ represents the maximum probability of interference that the kth PU can tolerate, then the constraint $\Pr\{\sum_m w_k^m[n] = 1 | a_k = 1\} \leq \check{o}_k$ needs to be satisfied. Using Bayes' theorem, the constraint can be re-written as:

$$\mathbb{E}\left[\lim_{N\to\infty}\sum_{n=0}^{N-1}(1-\gamma)\gamma^n a_k[n]\sum_m w_k^m[n]\right]/A_k \leq \check{o}_k, \;\; \forall k, \quad (4)$$

where A_k denotes the long-term probability of the kth band being occupied by the kth PU. Clearly, (4) is a long-term constraint, so that $\check{o}_k$ can be readily interpreted as the maximum fraction of time that SUs can interfere the kth PU.

III. PROBLEM FORMULATION AND OPTIMAL RA

The previous section described the variables to be optimized, the constraints that those variables needed to satisfy, and the CSI that will be used as input for the design. The only remaining step to formulate our optimization problem is to identify the metric to be maximized; different utility (reward) functions can be used for such a purpose. In this work, we use a linear combination between the sum-rate transmitted by the users minus the sensing cost. Specifically, at time n we define the instantaneous utility as $U[n] := \sum_k \left(\sum_m w_k^m[n]C_k^m(h_k^m[n], p_k^m[n])\right) - \xi_k s_k[n]$. Then, we are interested in maximizing the average long-term utility

$$\bar{U} := \mathbb{E}\left[\lim_{N\to\infty}\sum_{n=0}^{N-1}(1-\gamma)\gamma^n U[n]\right]. \qquad (5)$$

Before addressing the optimal design, it is convenient to clarify the operation of our system. At each slot n the CR will run five sequential steps: s1) the Markov transition matrix and the post-decision beliefs $\mathbf{b}_k^S[n-1]$ of the previous instant are used to obtain pre-decision beliefs $\mathbf{b}_k[n]$ at instant n; s2) the perfect (deterministic) values of $h_k^m[n]$ are acquired; s3) $h_k^m[n]$ and $\mathbf{b}_k[n]$ are used to find $s_k^*[n]$; s4) the values of $s_k^*[n]$, and $z_k[n]$ for the channels for which $s_k^*[n] = 1$, are used to get the post-decision beliefs $\mathbf{b}_k^S[n]$; s5) $h_k^m[n]$ and $\mathbf{b}_k^S[n]$ are used to find the optimal value of $w_k^{m*}[n]$ and $p_k^{m*}[n]$, and the SUs transmit accordingly. Several of these steps require exchange of signalling. Except for that related to the sensing of the activity of the PUs, we assume that the acquisition and exchange of information are ideal (instantaneous and error free) and do not entail any cost.

As already mentioned, most works design the sensing policy [cf. s3)] and the RA policy [cf. s5)] separately. The approach in this paper is to design them jointly. To do so, we start by finding the optimal RA scheme for a *fixed* sensing policy. This is useful because: i) it corresponds to the task that has to be carried out in s5); and ii) the expression for the optimal $w_k^{m*}[n]$ and $p_k^{m*}[n]$ as a function of $h_k^m[n]$ and $b_k^S[n]$ will be used to find the optimal sensing in s3).

A. Optimal RA

To formulate the problem that gives rise to the optimal RA we have to take into account two important facts. The first one is that the expression for the optimal RA needs to hold for any sensing policy. Since modifying the value of $s_k[n]$ will modify the value of $\mathbf{b}_k^S[n]$, the latter implies that the expression of the optimal RA needs to hold for any value of $\mathbf{b}_k^S[n]$; i.e., needs to be a function of $\mathbf{b}_k^S[n]$. The second one is that the sensing decisions $s_k(\mathbf{i}[n])$ are assumed to be fixed. This implies that the terms and constraints which only depend on $s_k[n]$ can be dropped. Hence, i) the constraint $s_k[n] \in \{0,1\}$ does not need to be considered here, and ii) the sensing cost term in (5) can be dropped. To account for ii) we define the simplified utility as $\bar{U}_{RA} := \sum_{k,m}\mathbb{E}[\lim_{N\to\infty}\sum_{n=0}^{N-1}(1-\gamma)$ $\gamma^n w_k^m[n]C_k^m(h_k^m[n], p_k^m[n])]$. Under all previous considerations, the optimal RA schemes are obtained as the solution

of the following problem:

$$\max_{w_k^m[n],p_k^m[n]} \bar{U}_{RA} \tag{6a}$$

$$\text{s. to:} \quad (2), (3), (4), \ w_k^m[n] \in \{0,1\}, p_k^m[n] \geq 0. \tag{6b}$$

This problem was solved in [9], [14]. The approach to find the optimal RA is to dualize the long-term constraints (3) and (4). For this purpose, let π^m and θ_k denote the Lagrange multipliers respectively associated to these constraints. Then, it is shown in [9], [14] that the solution to the problem is

$$p_k^{m*}[n] := \left[(\dot{C}_k^m)^{-1} \left(h_k^m[n], \pi^m \right) \right]_+; \tag{7}$$

$$w_k^{m*}[n] := \mathbb{1}_{\{(\phi_k^m[n]=\max_l \phi_k^l[n]) \ \wedge \ (\phi_k^m[n]>0)\}}, \ \text{with} \tag{8}$$

$$\phi_k^m[n] := \varphi_k^m[n] - \theta_k \left[\mathbf{b}_k^S[n] \right]_2, \ \text{and} \tag{9}$$

$$\varphi_k^m[n] := C_k^m(h_k^m[n], p_k^{m*}[n]) - \pi^m p_k^{m*}[n], \tag{10}$$

where $(\dot{C}_k^m)^{-1}(h_k^m[n], \cdot)$ denotes the inverse of the derivative of $C_k^m(h_k^m[n], \cdot)$. The auxiliary variables $\varphi_k^m[n]$ and $\phi_k^m[n]$ can be viewed as link quality indicators (LQIs). The indicator $\varphi_k^m[n]$ represents the best achievable trade-off between power and bit rate. Similarly, The indicator $\phi_k^m[n]$ represents a trade off among the rate, power, and short-term probability of interfering. The multipliers π^m and θ_k play the role of power and probability of interfering price (cost). Lastly, note that if the LQI for all users in a given channel is negative, then that channel should not contain secondary transmissions.

There are several methods to set the value of the dual variables π^m and θ_k. The classical one, consists in using the values that minimize the dual function associated with (6): π^{m*} and θ_k^*. Since (6) has zero duality gap, the optimal RA is found by substituting $\pi^m = \pi^{m*}$ and $\theta_k = \theta_k^*$ into (7) and (9). Alternative methods that guarantee tracking capabilities and give rise to RA schemes that are asymptotically optimal entail replacing π^m and θ_k with time-varying stochastic estimates; see, e.g., [14] for details.

The optimal RA in (7)-(10) will be used in the next section as an input for the algorithms that design the optimal sensing policy. Before addressing that design, we introduce some auxiliary variables that will be helpful to simplify the mathematical derivations in the next section. Such variables are based on the expressions in (7)-(10). Specifically, let $J_k[n]$ be the kth channel's *resource score* variable, defined as $J_k[n] := \sum_m w_k^{m*}[n]\phi_k^m[n]$. Using (8), this resource score can be written as $J_k[n] = \left[\max_l \phi_k^{l*}[n] \right]_+$. Defining the variable $\varphi_k^{win}[n] := \max_l \varphi_k^l[n]$, the vector $\mathbf{F}_k[n] := [\varphi_k^{win}[n], \ \varphi_k^{win}[n] - \theta_k[n]]^T$, and using (9), the previous equality can be written as

$$J_k[n] = \left[\mathbf{F}_k^T[n] \mathbf{b}_k^S[n] \right]_+. \tag{11}$$

This expression will be useful because $\mathbf{F}_k[n]$ encapsulates all the information which is relevant for the RA and does not depend on the sensing policy. Note that maximizing $\sum_k J_k[n]$ is equivalent to maximizing the lagrangian of (6). In fact, the optimal value that can be achieved at the original problem is the same than that of an equivalent problem, in which there is a single virtual user (the *winner*), whose LQI in each channel k and instant n is $\varphi_k^{win}[n]$. For this reason, $\varphi_k^{win}[n]$ will be referred to as the LQI of channel k.

IV. Optimal Sensing decision

In section III-A, the RA scheme was obtained as the solution of a convex problem, which was solved using a dual approach. The design optimization was carried out supposing that the sensing decisions were fixed. However, it became apparent that the reward obtained by the system depends also on $\mathbf{b}_k^S[n]$ [cf. (9)], which itself depends on the sensing policy.

The main objective of this section is to develop an scheme that optimizes the sensing policy taking as inputs: $h_k^m[n]$, $\mathbf{b}_k[n]$, and the expressions for the optimal RA in (7)-(10). Before formulating the optimization problem that will give rise to the optimal sensing policy, some additional considerations are in order. First, note that each sensing decision will have an impact on future sensing decisions. This is because, due to the time correlation of the primary occupancy process, the action of sensing increases the available information in the current and future time slots; and the decision of sensing a channel depends itself on the information available at that moment. In other words, all sensing decisions form a string of events, and the benefit associated to an individual decision is spread along time. Second, if there were no cost on sensing a channel, the solution to this problem would be to simply sense all the time, because the sensing task generally entails a positive effect on the objective. However, since sensing a channel entails a cost, we need to know the corresponding benefit in order to make a decision. Having all this in mind, it becomes clear that this problem requires the use of DP to be solved optimally.

A. Finding the optimal sensing policy

Adopting the classic DP formulation [12], the optimization problem can be written in the form of a constrained DP as:

$$\max_{s_k[n],w_k^m[n],p_k^m[n]} \bar{U} \tag{12a}$$

$$\text{s. to:} \quad (6b), \ s_k[n] \in \{0,1\}. \tag{12b}$$

Here, we also optimize over the sensing variables $s_k[n]$ [cf. (6)]. Using the same approach than that for solving (6), the long-term constraints in (6b) are dualized. Moreover, if we substitute the optimal RA (7)-(10) and the auxiliary variables (11) into the Lagrangian of (12), then the problem to be optimized can be reformulated as the following *unconstrained* DP [11]:

$$\max_{s_k[n]\in\{0,1\}} \sum_k \mathbb{E}\left[\lim_{N \to \infty} \sum_{n=0}^{N-1} (1-\gamma)\gamma^n \left(\left[\mathbf{F}_k^T[n] \mathbf{b}_k^S[n] \right]_+ - \xi_k s_k[n] \right) \right] \tag{13}$$

where the only remaining variables to be optimized are $s_k[n]$. Note that the two terms in (13) depend on $s_k[n]$. The first term is the dot product of vectors $\mathbf{F}_k[n]$ (which does not depend on $s_k[n]$) and $\mathbf{b}_k^S[n]$ (which depends on $s_k[n]$) [cf. (11)]. The second term is just the product of constant ξ_k and the sensing variable $s_k[n]$. The expression in (13) also reveals that $\mathbf{F}_k[n]$ encapsulates all the information pertaining SUs which is relevant to find $s_k^*[n]$. In other words, instead of knowing $h_k^m[n]$, $w_k^{m*}[n]$ and $p_k^{m*}[n]$, it suffices to know $\mathbf{F}_k[n]$. Last but not least, (13) manifests that the sensing decision will be made comparing the sensing cost with the increase in the expected LQI after sensing the channel.

To facilitate the solution of (13), the following facts will be considered. The problem in (13) can be modeled as a partially observable Markov decision process (POMDP) which depends on random variables that are discrete (occupancy variables $a_k[n]$) and continuous (channel gains $h_k^m[n]$). However, since $h_k^m[n]$ are perfectly known and independent across time, the problem can be posed as a POMDP whose state space is only $a_k[n]$. The second fact is that problem (13) can be separated (decoupled) across channels. This fact is true because the RA problem in the dual domain was separable across channels and because $a_k[n]$ are assumed to be independent across channels. As a result, the optimal sensing policy can be found by solving K discrete-state-space POMDPs . The reward function [12] corresponding to kth POMDP is $R_k[n] := -\xi_k s_k[n] + \left[\mathbf{F}_k^T[n]\mathbf{b}_k^S[n]\right]_+$.

B. Formulation of Bellman's Equation

In order to obtain the optimal sensing decision $s_k^*[n]$, we just have to compare the expected sum reward for $s_k[n] = 0$ and for $s_k[n] = 1$. The value of $s_k[n]$ will affect: i) the expected reward at time slot n, and ii) the expected rewards for the time slots after n. The latter is true because the beliefs for future instants depend on the current sensing decision. To account for this effect in the formulation, we introduce the *(expected) value function* $\bar{V}_k^n(\mathbf{b}_k[n])$, which represents the expected sum reward if we reach time slot n with a specified $\mathbf{b}_k[n]$. The standard Bellman's equations that drive the optimal sensing decision and the value function are:

$$s_k^*[n] = \arg\max_{s\in\{0,1\}} \mathbb{E}\left[R_k[n]\big|s_k[n] = s\right] + \\ \gamma\mathbb{E}\left[\bar{V}_k^{n+1}(\mathbf{b}_k[n+1])\big|s_k[n] = s\right] \quad (14)$$

$$\bar{V}_k^n(\mathbf{b}_k[n]) = \max_{s\in\{0,1\}} \mathbb{E}\left[R_k[n]\big|s_k[n] = s\right] + \\ \gamma\mathbb{E}\left[\bar{V}_k^{n+1}(\mathbf{b}_k[n+1])\big|s_k[n] = s\right] \quad (15)$$

where the term $\mathbb{E}[R_k[n]|s_k[n] = s]$ is the *short-term* expected reward conditioned to $s_k[n] = s$; and the term $\mathbb{E}[\bar{V}_k^{n+1}(\mathbf{b}_k[n+1])|s_k[n] = s]$ is the expected sum-reward to be obtained in all future time instants, conditioned to $s_k[n] = s$. Next, we describe how to obtain the expressions for these two terms.

The expected short-term reward is calculated as follows. If $s_k^*[n] = 0$, then $\mathbf{b}_k^S[n] = \mathbf{b}_k[n]$ and $\mathbb{E}\left[R_k[n]\big|s_k[n] = 0\right] = [\mathbf{F}_k[n]^T\mathbf{b}_k[n]]_+$. If $s_k^*[n] = 1$, then the expected short-term reward is found by averaging over the sensor outcome $z_k[n]$ and subtracting the cost of sensing: $\mathbb{E}[R_k[n]|s_k[n] = 1] = -\xi_k + \sum_{z=0}^1 \Pr(z_k[n] = z|\mathbf{b}_k[n])[\mathbf{F}_k[n]^T\mathbf{b}_k^S(\mathbf{b}_k[n], z)]_+$. Substituting (1) into the latter yields

$$\mathbb{E}[R_k[n]\big|s_k[n] = 1] = -\xi_k + \sum_{z=0}^1 \left[\mathbf{F}_k[n]^T\mathbf{D}_z\mathbf{b}_k[n]\right]_+. \quad (16)$$

The expression $\mathbb{E}\left[\bar{V}_k^{n+1}(\mathbf{b}_k[n+1])\big|s_k[n] = s\right]$ is the expectation of the value function in the following time instant $n + 1$, which equals the expected, discounted, sum-reward in all future time instants. If $s_k[n] = 0$, then there is no correction step, only the update step is performed, and

$$\mathbb{E}\left[\bar{V}_k^{n+1}(\mathbf{b}_k[n+1])\big|s_k[n] = 0\right] = \bar{V}_k^{n+1}(\mathbf{P}_k\mathbf{b}_k[n]). \quad (17)$$

If, for the contrary, $s_k[n] = 1$, then the belief is corrected according to (1), and updated in the transition to slot $n + 1$:

$$\mathbb{E}\left[\bar{V}_k^{n+1}(\mathbf{b}_k[n+1])\big|s_k[n] = 1\right] = \\ \sum_{z=0}^1 \Pr(z|\mathbf{b}_k[n])\bar{V}_k^{n+1}(\mathbf{P}_k\mathbf{b}_k^S(\mathbf{b}_k[n], z)). \quad (18)$$

Substituting the previous equalities into (15):

$$\bar{V}_k^n(\mathbf{b}_k[n]) = \mathbb{E}\Big[\max\big\{[\mathbf{F}_k[n]^T\mathbf{b}_k[n]]_+ + \gamma\bar{V}_k^{n+1}(\mathbf{P}_k\mathbf{b}_k[n]),$$
$$-\xi_k + \sum_{z=0}^1 [\mathbf{F}_k[n]^T\mathbf{D}_z\mathbf{b}_k[n]]_+ + \gamma\Pr(z|\mathbf{b}_k[n])\bar{V}_k^{n+1}\big(\tfrac{\mathbf{P}_k\mathbf{D}_z\mathbf{b}_k[n]}{\mathbf{1}^T\mathbf{D}_z\mathbf{b}_k[n]}\big)\big\}\Big] \quad (19)$$

where the expectation operator is taken over $\mathbf{F}_k[n]$. The sensing decision can be written as:

$$s_k^*[n] = \underset{s\in\{0,1\}}{\arg\max} \; s\cdot\left([\mathbf{F}_k^T[n]\mathbf{b}_k[n]]_+ + \gamma\bar{V}_k^{n+1}(\mathbf{P}_k\mathbf{b}_k[n])\right) + (1-s)\cdot$$
$$\left(-\xi_k + \sum_{z=0}^1 [\mathbf{F}_k^T[n]\mathbf{D}_z\mathbf{b}_k[n]]_+ + \gamma\Pr(z|\mathbf{b}_k[n])\bar{V}_k^{n+1}\big(\tfrac{\mathbf{P}_k\mathbf{D}_z\mathbf{b}_k[n]}{\mathbf{1}^T\mathbf{D}_z\mathbf{b}_k[n]}\big)\right) \quad (20)$$

The only step left is to find the expression for function $\bar{V}_k^{n+1}(\cdot)$. Due to the formulation chosen for the DP, $\bar{V}_k^{n+1}(\cdot)$ is a stationary function whose input is the instantaneous belief. Since the gains of the secondary channels are continuous, the value function is a piecewise linear convex (PWLC) function with infinite linear components. Hence, the value function must be computed numerically. A standard approach is to sample the belief space and use a Monte Carlo method to estimate the mean over $\mathbf{F}_k[n]$ [12].

V. NUMERICAL RESULTS

This section analyzes the feasibility and optimality of the developed joint sensing and RA schemes. Since the optimal RA developed in section III-A were analyzed in [9], [14], this section focuses on the optimal joint sensing and RA scheme. The default setup of the simulation is the following: $M = 4$; $K = 4$; the secondary normalized channel gains are Rayleigh distributed with $\mathbb{E}[h_k^m[n]] = 3.16 \forall k, m$. The average power limits are $[\breve{p}^1, \breve{p}^2, \breve{p}^3, \breve{p}^4] = [20, 16, 14, 10]$, normalized to the average noise power. The maximum interference probabilities allowed in these channels are $[\breve{o}_1, \breve{o}_2, \breve{o}_3, \breve{o}_4] = [0.3, 0.05, 0.1, 0.1]$. The primary occupancy dynamics and the sensor probability errors follow a set-up similar to that used in the simulations of [6]. The chosen parameters are $\mathbf{P}_k = [0.95, 0.02; 0.05, 0.98]\forall k$; $[P_1^{FA}, P_2^{FA}, P_3^{FA}, P_4^{FA}] = [0.09, 0.09, 0.05, 0.05]$, and $[P_1^{MD}, P_2^{MD}, P_3^{MD}, P_4^{MD}] = [0.08, 0.08, 0.03, 0.03]$. The sensing costs have been chosen to be approximately the average utility gain in a single time slot in a single channel: $[\xi_1, \xi_2, \xi_3, \xi_4] = [1, 1.8, 1, 1.8]$.

The sensing and RA scheme developed in this paper is compared with three alternative sensing schemes: SS1) a simple sensing scheme that senses the channel uniformly at random with probability p_s; SS2) a sequential (round-robin) sensing scheme used in the simulations of [9] which senses the channel one out of each N_s time slots; SS3) a myopic policy that uses the scheme developed in this paper but set the value function to zero (i.e., $\bar{V}_k^{n+1} = 0$); and SS4) optimal sensing scheme

developed in this paper: the value function is found using the Value Iteration procedure [12], approximating the value function with polynomial regression. The hyperparameters p_s and N_s in SS1 and SS3 are chosen such that the proportion of sensed/not sensed slots is the same as in SS4.

Three test cases are considered: TC1) the default setup described at the beginning of this section; TC2) the same as TC1 with different values for the sensing error probabilities: $[P_1^{FA}, P_2^{FA}, P_3^{FA}, P_4^{FA}] = [0.05, 0.04, 0.07, 0.06]$, and $[P_1^{MD}, P_2^{MD}, P_3^{MD}, P_4^{MD}] = [0.10, 0.05, 0.03, 0.03]$; TC3) the same as TC2 with $\mathbf{P}_k = [0.921, 0.032; 0.079, 0.968] \forall k$ (the idle/busy ratio is the same as TC1 and TC2 but there is weaker time correlation). Results are listed in Table I.

TABLE I
AVERAGE UTILITY FOR DIFFERENT SENSING SCHEMES

	SS1	SS2	SS3	**SS4**
TC1	5,875	4,829	4,872	**6,592**
TC2	5,729	4,736	4,884	**6,729**
TC3	5,508	4,723	4,928	**6,271**

The results confirm that the optimal policy SS4 outperforms the suboptimal alternatives. In addition, we observe that if the time correlation of $a_k[n]$ is stronger (TC2 with respect to TC3), then the utility gap of the optimal scheme relative to the suboptimal ones is higher.

Finally, to gain insights on the behavior of the optimal sensing schemes, Figure 1 plots the decision regions for the four channels in TC1 (one plot per channel). Since the sensing decision function is two-dimensional (it depends on $\mathbf{b}_k[n]$ and $\varphi_k^{win}[n]$), the decision function can be represented as an image. In all the plots, three regions are represented. Each of these regions is associated with a different (output) value of $s_k^*[n]$ and $\sum_m w_k^{m*}[n]$ (one if one user access the channel and zero otherwise). Note that the size and shape of the sensing region for channel k depend on the values of $\{\mathbf{P}_k, P_k^{FA}, P_k^{MD}, \xi_k\}$.
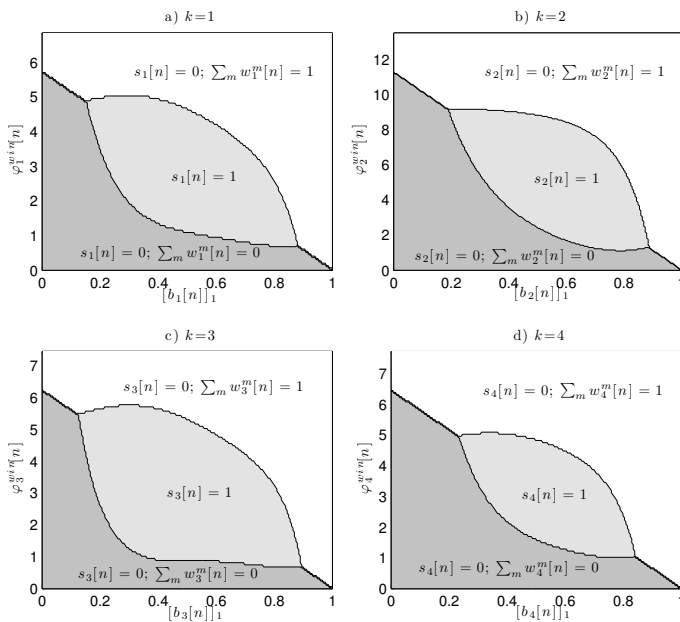

Fig. 1. Decision maps (regions) for the four channels in TC1. The light gray area in the center corresponds to the sensing decision.

VI. CONCLUSIONS

Jointly optimal sensing and RA schemes for an interweave CR with multiple SUs and PUs have been designed. The schemes were designed to maximize the sum-rate of the SUs while adhering to constraints that limit the long-term power transmitted by the SUs and the long-term probability of interfering the PUs. The schemes were designed assuming perfect CSI of the SU links, and outdated/noisy CSI of the PUs . Sensing a channel entails a given cost and at each instant the system has to decide which channels (if any) are sensed. The optimal design was addressed using a two-step strategy. First, the optimal RA for any sensing scheme was found. This problem was convex and thus efficiently solvable. The second step used the solution found in the first step as input to solve for the optimal sensing policy. The problem in the second step was non-convex and was solved using DP techniques. Moreover, it was shown that due to the operating conditions considered in the paper and the strategy followed to solve the problem, the DP that was solved in the second step entailed much less complexity than that of the original joint design.

REFERENCES

[1] S. Haykin, "Cognitive radio: brain-empowered wireless communications," *IEEE. J. Sel. Areas Commun.*, vol. 23, no. 2, pp. 201 – 220, feb. 2005.

[2] R. Urgaonkar and M. Neely, "Opportunistic scheduling with reliability guarantees in cognitive radio networks," *IEEE Trans. Mobile Comp.*, vol. 8, no. 6, pp.766-777, Jun. 2009.

[3] L. Musavian and S. Aissa, "Fundamental capacity limits of cognitive radio in fading environments with imperfect channel information," *IEEE Trans. Commun.*, vol. 57, no. 11, pp. 3472-3480, Nov. 2009.

[4] X. Wang, "Jointly Optimal Sensing Selection and Power Allocation for Cognitive Communications," *IEEE Proc. of Globecom Conf.*, Miami, FL, Dec. 6-10, 2010.

[5] A. G. Marques, X. Wang, and G. B. Giannakis, "Optimal stochastic dual resource allocation for cognitive radios based on quantized CSI," *IEEE Proc. of Intl. Conf. on Acoustics, Speech and Signal Process.*, Las Vegas, NV, Mar. 30 - Apr. 4, 2008.

[6] Y. Chen, Q. Zhao, and A. Swami, "Joint design and separation principle for opportunistic spectrum access in the presence of sensing errors," *IEEE Trans. Info. Theory*, vol. 54, no. 5, pp. 2053 –2071, may 2008.

[7] X. Wang, "Joint sensing-channel selection and power control for cognitive radios," *IEEE Trans. Wireless Commun.*, vol. 10, no. 3, pp. 958 –967, march 2011.

[8] S.-J. Kim and G. Giannakis, "Sequential and cooperative sensing for multi-channel cognitive radios," *IEEE Trans. Signal Process.*, vol. 58, no. 8, pp. 4239 –4253, Aug. 2010.

[9] A. G. Marques, G. B. Giannakis, L. M. Lopez-Ramos, and J. Ramos, "Stochastic resource allocation for cognitive radio networks based on imperfect state information," *IEEE Proc. of Intl. Conf. on Acoustics, Speech and Signal Process.*, Prague, Czech Rep., May. 22- 27, 2011.

[10] A. Goldsmith, *Wireless Communications*, Cambridge Univ. Press, 2005.

[11] D. A. Castañon, "Approximate Dynamic Programming for sensor management," in *Proc. of the 36th Conf. on Decision and Contr.*, vol. 2, San Diego, CA, Dec. 1997, pp. 1202-1207.

[12] W. B. Powell, *Approximate Dynamic Programming: Solving the Curses of Dimensionality (Wiley Series in Probability and Statistics)*. Wiley-Interscience, 2007.

[13] X. Gong, S. Vorobyov, C. Tellambura, "Joint bandwidth and power allocation in cognitive radio networks under fading channels", *IEEE Proc. of Intl. Conf. on Acoustics, Speech and Signal Process.*, Prague, Czech Rep., May. 22- 27, 2011.

[14] A. G. Marques, G. B. Giannakis, L. M. Lopez-Ramos, J. Ramos, "Resource Allocation for Interweave and Underlay CRs under Probability-of-Interference Constraints", *IEEE J. Sel. Areas Commun.* (submitted).

[15] Q. Zhao, L. Tong, A. Swami, and Y. Chen, "Decentralized cognitive mac for opportunistic spectrum access in ad hoc networks: A pomdp framework," *Selected Areas in Communications, IEEE Journal on*, vol. 25, no. 3, pp. 589 –600, april 2007.