Jointly Optimal Sensing and Resource Allocation for Multiuser Interweave Cognitive Radios

Luis M. Lopez-Ramos *Student Member, IEEE*, Antonio G. Marques *Senior 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 we address them jointly. In particular, we investigate an interweave cognitive radio with multiple secondary users that access orthogonally a set of frequency bands originally devoted to primary users. The schemes are designed to minimize the cost of sensing, maximize the performance of the secondary users (weighted sum rate), and limit the probability of interfering with the primary users. The joint design is addressed using nonlinear optimization and dynamic programming, which is able to leverage the time correlation in the activity of the primary network. A two-step strategy is implemented: it first finds the optimal resource allocation for any sensing scheme and then uses that solution as input to solve for the optimal sensing policy. The two-step strategy is optimal, gives rise to intuitive optimal policies, and entails a computational complexity much lower than that required to solve the original formulation.**

***Index Terms*—Cognitive radios, sequential decision making, dual decomposition, partially observable Markov decision processes**

I. Introduction

Cognitive radios (CRs) are viewed as the next-generation solution to alleviate the perceived spectrum scarcity. To do so, CRs must be aware of their radio environment and adapt their transmission parameters accordingly. When CRs are deployed, the secondary users (SUs) (users that look for spectrum holes to transmit opportunistically) have to sense their radio environment to optimize their communication performance while avoiding (limiting) the interference to the primary users (PUs) (pre-existent or licensed users). More precisely, interweave[1] CR is the setup where interference is limited by allowing the SUs to be active only a small proportion of the time the PUs are active. As a result, effective operation of CRs requires the implementation of two critical tasks: i) sensing the spectrum seeking for transmit opportunities and ii) dynamic adaptation of the available resources according to the sensed information [2]. This work aims to optimize these two tasks jointly.

TABLE I
List of most important symbols

Symbol	Meaning
n	Time slot index
M , m	Number of users / User index
K , k	Number of channels / Channel index
$h_k^m[n]$	Fading gain for user m on channel k
$a_k[n]$	Presence of PU in channel k
$\mathbf{P}_k$	Transition probability matrix of $a_k[n]$
$s_k[n]$, $z_k[n]$	Sensing decision / Sensing output
P_k^{FA}, P_k^{FA}	False alarm / missed detection probabilities
$B_k[n]$, $B_k^S[n]$	Pre-decision / post decision belief on $a_k[n]$
$\mathbf{b}_k[n]$, $\mathbf{b}_k^S[n]$	Vector form of pre-decision/post-decision belief
$w_k^m[n]$	Scheduling variable (1 if user m occupies channel k)
$p_k^m[n]$	Nominal power of user m in channel k
$\check{p}^m[n]$	Maximum average power consumed by user m
$\check{o}_k[n]$	Maximum prob. of interference on channel k
π^m , θ_k	Lagrange multipliers associated with (5), (6)
γ	Discount factor $0 < \gamma < 1$
ξ_k	Sensing price
$\mathbf{l}_k^m[n]$	Instantaneous reward indicator (IRI) vector
$\bar{V}_k(\cdot)$	Value function of the POMDP associated to channel k

To carry out the *sensing task* two important challenges are: C1) the presence of errors in the measurements that lead to errors in the channel occupancy detection and thus render harmless SU transmissions impossible; 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 need of the RA algorithms to deal with channel imperfections such as noise or quantization; and C4) the selection of metrics that properly quantify the reward for the SUs and the harm for the PUs in case of interference.

A. Related Work

Many alternatives have been proposed in the CR literature to deal with these challenges. Different forms of imperfect channel state information (CSI), such as quantized or noisy CSI, have been used to deal with C1 [3]. However, in the context of CR, fewer 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 RA algorithms [4]. The inherent tradeoff between sensing cost and throughput gains in C2 has been investigated [5]; designs that account for such a tradeoff based on convex optimization [6] and dynamic programming (DP) [4] for specific system setups have been

This work is supported by the Spanish Ministry of Science, under FPU Grant AP2010-1050 and EU contract FP7-ICT-2011-9-601102; and the Spanish Ministry of Economy, under grant No TEC2013-41604-R. Parts of this paper were presented at CROWNCOM 2012.

All authors are with the Dept. of Signal Theory and Comms., King Juan Carlos Univ., Camino del Molino s/n, Fuenlabrada, Madrid 28943, Spain. Contact info available at: http://www.tsc.urjc.es/

[1] Initially, most works referred to interweave CRs as overlay CRs (in contrast to underlay CRs). Nowadays, the difference between the three paradigms is clear and the term interweave CRs to refer to the scenario where the SUs exploit spectrum holes to transmit opportunistically is widely accepted; see e.g., [1].

proposed. 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 [4], [7]. Finally, different alternatives have been considered to deal with C4 and limit the harm that the SUs cause to the PUs [8]. The most widely used is to set limits on the peak (instantaneous) and average interfering power. Some works also have imposed limits on the rate loss that PUs experience [9], [10], while others look at limiting the instantaneous or average probability of interfering the PU (bounds on the short-term or long-term outage probability) [11], [7].

Regardless of the challenges addressed and the formulation chosen, the sensing and RA policies have been traditionally designed separately. Each of the tasks has been investigated thoroughly and relevant results are available in the literature. However, a globally optimum solution requires a joint design capable of leveraging the interactions between those two tasks. Clearly, more accurate sensing enables more efficient RA, but at the expense of higher time and/or energy consumption [5]. Early works dealing with joint design of sensing and RA are [12] and [4]. In such works, imperfections in the sensors, and also time correlation of the state of the primary channel, are considered. As a result, the sensing design is modeled as a partially observable Markov decision process (POMDP) [13], [14, Ch. 12], which is a particular instance of DP [15]. The design of the RA in these works amounts simply to select the user transmitting on each channel (also known as user scheduling). Under mild conditions, the authors establish that a separation principle holds in the design of the optimal access and sensing policies. Additional works addressing the joint design of sensing and RA, and considering more complex operating conditions, were published recently [6], [16]. For a single SU operating multiple fading channels, [6] relies on convex optimization to optimally design both the RA and the indices of the channels to be sensed at every time instant. Assuming that the number of channels that can be sensed at every instant is fixed and that the PU activity is independent across time, the author establishes that the channels to sense are those that can potentially yield a higher reward for the secondary user. Joint optimal design is also pursued in [16], although for a very different setup. Specifically, [16] postulates that at each slot, the CR must compute the fraction of time devoted to sense the channel and the fraction devoted to transmit in the bands which are found to be unoccupied. Clearly, a tradeoff between sensing accuracy and transmission rate emerges. The design is formulated as an optimal stopping problem, and solved by means of Lagrange relaxation of DP [17]. However, none of these two works takes into account the temporal correlation of the state information of the primary network (SIPN).

B. Objective and Contributions

The objective of this work is to design the sensing and the RA policies *jointly* while accounting for the challenges C1-C4. The specific operating conditions considered in the paper are described next. We analyze an *interweave* CR with multiple SUs and PUs. SUs are able to adapt their transmit power and rate, and access orthogonally a set of frequency bands originally devoted to PU transmissions. *Orthogonally* here means that if a SU is transmitting, no other SU 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 main novelty (and contribution) of this work is the combined use of DP and dual nonlinear optimization techniques to design the jointly optimal sensing and RA schemes. 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 [12]. To solve the joint design, a two-step strategy is implemented. In the first step, we obtain an analytical expression for the performance achieved by the *optimal* RA as a function of the state of the system after the sensing task (this expression is valid for *any* fixed sensing scheme). This (sub-) problem was recently solved in [18], [7]. In the second step, the *analytical expression* obtained in the first step is used as input to obtain the optimal sensing policy. This two-step strategy has a double motivation. First, while the joint design is non convex and has to be solved using DP techniques, the problem to be solved in the first step (optimal RA for a fixed sensing scheme) can be recast as a convex one. Second, when the *expression* for the optimal RA performance is substituted back into the original joint design, the resulting sensing optimization problem (which does need to be solved using DP techniques) has a more favorable structure. More specifically, while the original design problem was a constrained DP, the problem to be solved in the second step is an unconstrained DP which can be solved separately for each of the channels. These facts will make our problem computationally affordable without entailing a loss of optimality [cf. Sec. III].

The rest of the paper is organized as follows. Sec. II describes the system setup and introduces notation. The optimization problem that gives rise to the optimal sensing and RA schemes is formulated in Sec. III. The solution for the optimal RA given the sensing scheme is presented in Sec. IV. The optimization of the sensing scheme is addressed in Sec. V, formulating the problem in the DP framework and developing its solution. Numerical simulations validating the theoretical claims and providing insights on our optimal schemes are presented in Sec. VI. Sec. VII summarizes the main properties of our jointly optimal RA and sensing policies and points out future lines of work[2].

[2] *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); $[\mathbf{x}]_i$ the ith entry of vector $\mathbf{x}$, and $[x]_+$ the projection of x onto the non-negative orthant, i.e., $[x]_+ := \max\{x, 0\}$.

II. System setup and state information

The section begins by briefly describing the system setup and the main operation steps (tasks that the system runs at every time slot). Then, the model for the CSI, which will play a critical role in the problem formulation, is explained in detail. The resources that SUs will adapt as a function of the CSI are described in the last part of the section.

We consider a CR scenario with several PUs and SUs. The frequency band of interest (portion of spectrum that is licensed to PUs, or the subset of this shared with the SUs) is divided into K frequency-flat orthogonal subchannels (indexed by k). Each of the M secondary users (indexed by m) opportunistically accesses any number of these channels during a time slot (indexed by n). Opportunistic here means that the user accessing each channel will vary with time as a function of the current CSI, with the objective of optimally utilizing the available channel resources. For simplicity, we assume that there exists a network controller (NC) which acts as a central scheduler and will also perform the task of sensing the medium for primary presence. The scheduling information will be forwarded to the mobile stations through a parallel feedback channel. The results hold for one-hop (either cellular or any-to-any) setups.

Next, we briefly describe the operation of the system. A more detailed description will be given in Sec. III, which will rely on the notation and problem formulation introduced in the following sections. Before starting, it is important to clarify that we focus on systems where the SIPN is more difficult to acquire than the state information of the secondary network (SISN). As a result, we will assume that SISN is error-free and acquired at every slot n, while SIPN is not. With these considerations in mind, the CR operates as follows. At every slot n the following tasks are run sequentially: T1) the NC acquires the SISN; T2) the NC relies on the output of T1 (and on previous measurements) to decide which channels to sense (if any), then the output of the sensing is used to update the SIPN; and T3) the NC uses the outputs of T1 and T2 to find the optimal RA for instant n. Overheads associated with acquisition of the SISN and notification of the optimal RA to the SUs are considered negligible. Such an assumption facilitates the analysis, and it is reasonable for scenarios where the SUs are deployed in a relatively small area which allows for low-cost signaling transmissions.

A. State information and sensing scheme

We begin by introducing the model for the SISN. Let $\tilde{h}_k^m[n]$ be the square magnitude of the fading coefficient of the channel between the mth SU and its intended receiver on frequency k during slot n. With $\sigma_k^m[n]$ denoting the corresponding noise-plus-interference power, $h_k^m[n] := \tilde{h}_k^m[n]/\sigma_k^m[n]$ is defined as the noise-normalized power gain for the mth SU on frequency k. The stochastic process $h_k^m[n]$ will be assumed to be independent and identically distributed (i.i.d.) across time. The values of $h_k^m[n]$ for all m and k constitute the SISN at slot n. The SISN is assumed perfect, so that the values of $h_k^m[n]$ at every time slot n are known with no errors. This assumption may be unrealistic, but it is made to focus on the challenges due to SIPN imperfections, which are always more severe. Nonetheless, comments on how to modify the RA when this assumption does not hold true will be provided in Sec. IV-A.

The SIPN accounts for the channel occupancy. We will assume that the primary system contains one user per channel. This assumption keeps the modeling simple and it is accurate for some practical scenarios, e.g. a primary system of mobile telephony where a single narrow-band channel is assigned to a single user during the course of a call. Since we consider an interweave scenario (i.e., if an SU accesses the channel when the PU is active, interference takes place regardless of the transmit power), it suffices to know whether a given channel is occupied or not [1]. This way, when a PU is not active, opportunities for SUs to transmit in the corresponding channel arise. The primary system is not assumed to collaborate with the secondary system. Hence, from the point of view of the SUs, the behavior of PUs is a stochastic process independent of $h_k^m[n]$. With these considerations in mind, the presence of the primary user in channel k at time n is represented by the binary state variable $a_k[n]$ (0/idle, 1/busy). Each primary user's behavior will be modeled as a simple, discrete-time, Gilbert-Elliot channel model, so that $a_k[n]$ is assumed to remain constant during the whole time slot, and then change according to a two-state, time invariant Markov chain. The Markovian property will be useful to keep the DP modeling simple and will also be exploited to recursively keep track of the SIPN. Nonetheless, more refined models can be considered without paying a big computational price [7], [19]. With $P_k^{xy} := \Pr(a_k[n] = x|a_k[n-1] = y)$, the dynamics for the Gilbert-Elliot model are fully described by the 2×2 Markov transition matrix $\mathbf{P}_k := [P_k^{00}, P_k^{01}; P_k^{10}, P_k^{11}]$. Sec. VII discusses the implications of relaxing some of these assumptions.

While knowledge of $h_k^m[n]$ at instant n was assumed to be perfect (deterministic), knowledge of $a_k[n]$ at instant n is assumed to be imperfect (probabilistic). Two important sources of imperfections are: i) errors in the sensing process and ii) outdated information (because the channels are not always sensed). To model the sensing task, let $s_k[n]$ denote a binary design variable which is 1 if the kth channel is sensed at time n, and 0 otherwise. Moreover, let $z_k[n]$ denote the output of the sensor if indeed $s_k[n] = 1$; i.e., if the kth channel has been sensed. We will assume that the output of the sensor is binary and may contain errors. To account for asymmetric errors, the probabilities of false alarm $P_k^{FA} = \Pr(z_k[n] = 1|a_k[n] = 0)$ and missed detection $P_k^{MD} = \Pr(z_k[n] = 0|a_k[n] = 1)$ are considered. Clearly, the specific values of P_k^{FA} and P_k^{MD} will depend on the detection technique the sensors implement and the sensor parameters (operating point). For simplicity, P_k^{FA} and P_k^{MD} are considered known and time invariant[3]. As already mentioned,

[3] This is reasonable if: i) the primary-secondary fading conditions are stationary and ii) complete information about their statistics (but not about their instantaneous values) is available. Under such conditions, the optimal operating point of the sensor is constant during the CR operation [20]. Nonetheless, our results can be adapted to handle time-variant $P_k^{FA}[n]$ and $P_k^{MD}[n]$. Specifically, results in section IV are not affected, and results in Section V can be adapted by accounting for the distribution of $P_k^{FA}[n]$ and $P_k^{MD}[n]$.

the sensing imperfections render the knowledge of $a_k[n]$ at instant n probabilistic; in other words, $a_k[n]$ is a partially observable state variable. The knowledge about the value of $a_k[n]$ at instant n will be referred to as the instantaneous belief. For a given instant n, two different types of belief are considered: the *pre-decision* belief $B_k[n]$ and the *post-decision* belief $B_k^S[n]$. Intuitively, $B_k[n]$ contains the information about $a_k[n]$ before the sensing decision has been made (i.e., at the beginning of task T2), while $B_k^S[n]$ contains the information about $a_k[n]$ once $s_k[n]$ and $z_k[n]$ (if $s_k[n]=1$) are known (i.e., at the end of task T2). Mathematically, if $\mathcal{H}_n$ represents the history of all past sensing decisions and measurements, i.e., $\mathcal{H}_n := \{s_k[0], z_k[0], \ldots, s_k[n], z_k[n]\}$; then $B_k[n] := \Pr(a_k[n]=1|\mathcal{H}_{n-1})$ and $B_k^S[n] := \Pr(a_k[n]=1|\mathcal{H}_n)$. For notational convenience, the beliefs will also be expressed as vectors, with $\mathbf{b}_k[n] := \left[1-B_k[n], B_k[n]\right]^T$ and $\mathbf{b}_k^S[n] = \left[1-B_k^S[n], B_k^S[n]\right]^T$. Provided that the Markov matrix $\mathbf{P}_k$ is known, the expression to get the pre-decision belief at time slot n is

$$\mathbf{b}_k[n] = \mathbf{P}_k \mathbf{b}_k^S[n-1]. \tag{1}$$

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

$$\mathbf{b}_k^S[n] = \mathbf{b}_k[n]. \tag{2}$$

If $s_k[n]=1$, the belief is updated as $\mathbf{b}_k^S[n] = \mathbf{b}_k^S\Big(\mathbf{b}_k[n], z_k[n]\Big)$, with

$$\mathbf{b}_k^S\Big(\mathbf{b}_k[n], z\Big) := \frac{\mathbf{D}_z \mathbf{b}_k[n]}{\Pr(z_k[n]=z\big|\mathbf{b}_k[n])}, \tag{3}$$

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 $[\mathbf{D}_z]_{2,2} := \Pr(z_k[n]=z|a_k=1)$. Using this same notation, the denominator can be rewritten as $\Pr(z_k[n]=z\big|\mathbf{b}_k[n]) = \mathbf{1}^T\mathbf{D}_z\mathbf{b}_k[n]$. Note that (1) and (2, 3) correspond to the prediction and update steps of a Bayesian recursive estimator, respectively. If no information about the initial state of the PU is available, the best choice is to initialize $\mathbf{b}_k[0]$ to the stationary distribution of the Markov chain associated with channel k.

In a nutshell, the actual state of the primary and secondary networks is given by the random processes $a_k[n]$ and $h_k^m[n]$, which are assumed to be mutually independent. The operating conditions of our CR are such that at instant n, the value of $h_k^m[n]$ is perfectly known, while the SIPN is formed by $\mathbf{b}_k[n]$ and $\mathbf{b}_k^S[n]$, which are a *probabilistic* description of $a_k[n]$. The system will perform the sensing and RA tasks based on the available SISN and SIPN. In particular, the sensing decision will be made based on $h_k^m[n]$ and $\mathbf{b}_k[n]$, while the RA will be implemented based on $h_k^m[n]$ and $\mathbf{b}_k^S[n]$.

B. Resources at the secondary network

We consider a secondary network where users implement adaptive modulation and power control, and share orthogonally the available channels. To describe the channel access scheme (scheduling) rigorously, let $w_k^m[n]$ be a Boolean variable so that $w_k^m[n]=1$ if SU m accesses channel k and zero otherwise. Moreover, let $p_k^m[n]$ be a nonnegative variable denoting the *nominal* power assigned for SU m to transmit in channel k, and let $C_k^m[n]$ be its corresponding rate. We say that the $p_k^m[n]$ is a nominal power in the sense that power is consumed only if the user is actually accessing the channel. Otherwise the power is zero, so that the actual (effective) power user m loads in channel k can be written as $w_k^m[n]p_k^m[n]$.

The transmission bit rate is obtained through Shannon's capacity formula [21]: $C_k^m[n] := C_k^m(h_k^m[n], p_k^m[n]) := \log_2(1+h_k^m[n]p_k^m[n]/\Gamma)$ where Γ is a signal-to-noise ratio (SNR) gap that accounts for the difference between the theoretical capacity and the actual rate achieved by the modulation and coding scheme the SU implements. This is a bijective, nondecreasing, concave function with $p_k^m[n]$ and it establishes a relationship between power and rate in the sense that controlling $p_k^m[n]$ implies also controlling $C_k^m[n]$.

The fact of the access being orthogonal implies that, at any time instant, at most one SU can access the channel. Mathematically,

$$\sum_m w_k^m[n] \leq 1 \ \ \forall k, n. \tag{4}$$

Note that (4) allows for the event of all $w_k^m[n]$ being zero for a given channel k. That would happen if, for example, the system believes that, at instant n, it is very likely that channel k is occupied by a PU.

III. Problem statement

The approach in this paper is to design the sensing and RA schemes as the solution of a judiciously formulated optimization problem. Consequently, it is critical to identify: i) the design (optimization) variables, ii) the state variables, iii) the constraints that design and state variables must obey, and iv) the objective of the optimization problem.

The first two steps were accomplished in Sec. II, stating that the design variables are $s_k[n]$, $w_k^m[n]$ and $p_k^m[n]$ (recall that there is no need to optimize over $C_k^m[n]$); and that the state variables are $h_k^m[n]$ (SISN), and $\mathbf{b}_k[n]$ and $\mathbf{b}_k^S[n]$ (SIPN).

Moving to step iii), the constraints that the variables need to satisfy can be grouped into two classes. The first class is formed by constraints that account for the system setup. This class includes constraint (4) as well as the following constraints that were implicitly introduced in the previous section: $s_k[n] \in \{0,1\}$, $w_k^m[n] \in \{0,1\}$ and $p_k^m[n] \geq 0$. The second class is formed by constraints that account for quality of service (QoS). In particular, we consider the following two constraints. The first one is a limit on the maximum average (long-term) power an SU can transmit. By enforcing an average consumption constraint, opportunistic strategies are favored because energy can be saved during deep fadings (or when the channel is known to be occupied) and used during transmission opportunities. Transmission opportunities are time slots where the channel is certainly known to be idle and the fading conditions are favorable. Mathematically, with $\check{p}^m$ denoting such maximum value, the average power constraint is written as:

$$\mathbb{E}\left[\lim_{N\to\infty}(1-\gamma)\sum_{n=0}^{N-1}\gamma^n\sum\nolimits_k w_k^m[n]p_k^m[n]\right]\leq \check{p}^m, \quad \forall m, \tag{5}$$

where $0<\gamma<1$ is a discount factor such that more emphasis is placed in near future instants. The factor $(1-\gamma)$ ensures that the averaging operator is normalized; i.e., that $\lim_{N\to\infty}\sum_{n=0}^{N-1}(1-\gamma)\gamma^n = 1$. As explained in more detail in Sec. V, using an exponentially decaying average is also useful from a mathematical perspective (convergence and existence of stationary policies are guaranteed).

While the previous constraint guarantees QoS for the SUs, we also need to guarantee a level of QoS for the PUs. As explained in the introduction, there are different strategies to limit the interference that SUs cause to PUs; e.g., by imposing limits on the interfering power at the PUs, or on the rate loss that such interference generates [7]. In this paper, we will guarantee that the *long-term* probability of a PU being interfered by SUs is below a certain prespecified threshold $\breve{o}_k$. Mathematically, we require $\Pr\{\sum_m w_k^m = 1|a_k = 1\}\leq \breve{o}_k$ for each band $k=1,\ldots,K$. Using the definition of conditional probability, the constraint can be rewritten as $\Pr\{\sum_m w_k^m = 1, a_k = 1\}/\Pr\{a_k=1\}\leq \breve{o}_k$ and, capitalizing on the fact that both a_k and $\sum_m w_k^m$ are Boolean variables:

$$\mathbb{E}\left[\lim_{N\to\infty}\sum_{n=0}^{N-1}(1-\gamma)\gamma^n a_k[n]\sum\nolimits_m w_k^m[n]\right]/A_k\leq \breve{o}_k, \quad \forall k, \tag{6}$$

where A_k, which is assumed known, denotes the stationary probability of the kth band being occupied by the corresponding primary user. Writing the constraint in this form reveals its underlying convexity. Before moving to the next step, two clarifications are in order. The first one is on the practicality of (6). Constraints that allow for a certain level of interference are reasonable because error-free sensing is unrealistic. Indeed, our model assumes that even if channel k is sensed as idle, there is a probability P_k^{MD} of being occupied. Moreover, when the interference limit is formulated as a long-term constraint (as it is in our case), there is an additional motivation for the constraint. The system is able to exploit the so-called interference diversity [22]. Such diversity allows SUs to take advantage of very favorable channel realizations even if they are likely to interfere PUs. To balance the outcome, SUs will be conservative when channel realizations are not that good and may remain silent even if it is likely that the PU is not present. The second clarification is that we implicitly assumed that SU transmissions are possible even if the PU is present. The reason is twofold. First, the fact that a SU transmitter is interfering a PU receiver, does not necessarily imply that the reciprocal is true. Second, since the NC does not have any control over the power that primary transmitters use, the interfering power at the secondary receiver could be incorporated into $h_k^m[n]$ as an additional source of noise.

The fourth (and last) step to formulate the optimization problem is to design the metric (objective) to be maximized. Different utility (reward) and cost functions can be used to such purpose. As mentioned in the introduction, in this work we are interested in schemes that maximize the weighted sum rate of the SUs and minimize the cost associated with sensing. Specifically, we consider that every time that channel k is sensed, the system has to pay a price $\xi_k > 0$. We assume that such a price is fixed and known beforehand, but time-varying prices can be accommodated into our formulation too (see Sec. V-C for additional details). This way, the sensing cost at time n is $U_S[n] := \sum_k \xi_k s_k[n]$. Similarly, we define the utility for the SUs at time n as $U_{SU}[n] := \sum_k\left(\sum_m \beta^m w_k^m[n]C_k^m(h_k^m[n], p_k^m[n])\right)$, where $\beta^m > 0$ is a user-priority coefficient. Based on these definitions, the utility for our CR at time n is $U_T[n] := U_{SU}[n] - U_S[n]$. Finally, we aim to maximize the long-term utility of the system denoted by $\bar{U}_T$ and defined as

$$\bar{U}_T := \mathbb{E}\left[\lim_{N\to\infty}\sum_{n=0}^{N-1}(1-\gamma)\gamma^n U_T[n]\right]. \tag{7}$$

With these notational conventions, the optimal $s_k^*[n]$, $w_k^{m*}[n]$ and $p_k^{m*}[n]$ will be obtained as the solution of the following constrained optimization problem.

$$\max_{\{s_k[n], w_k^m[n], p_k^m[n]\}} \quad \bar{U}_T \tag{8a}$$

$$\text{s. to}: \quad (4),\ w_k^m[n]\in\{0,1\},\ p_k^m[n]\geq 0, \tag{8b}$$

$$(5),\ (6) \tag{8c}$$

$$s_k[n]\in\{0,1\}. \tag{8d}$$

The constraints in (8) have been gathered into three groups: (8b) refers to constraints that affect the RA (i.e., $w_k^m[n]$ and $p_k^m[n]$) and need to hold at every time instant; (8c) refers to constraints that affect the RA and need to hold in the long term; and (8d) affects the design variables involved in the sensing task ($s_k[n]$).

The main difficulty in solving (8) is that the solution for all time instants has to be found jointly. The key reason is that sensing decisions at instant n have an impact not only at that instant, but at future instants as well. As a result, a separate per-slot optimization approach is not optimal in the long term. More specifically, (8) belongs to the class of sequential decision problems that needs to be solved by means of DP. This techniques usually give rise to algorithms of high complexity, so that a careful analysis must be performed to keep the problem tractable. In this work, the algorithmic strategy exploits some details of the problem structure that will help to reduce the complexity of the solution.

Two common practices to render DP problems tractable are: i) analyzing the problem within a well-studied framework; and ii) looking for approximation strategies that allow to reduce the computational load in exchange for a small loss of optimality. The problem at hand will be analyzed within the framework of POMDPs. Markov Decision Processes (MDPs)s are a class of DPs where state transitions and average rewards only depend on the current state-action pair. POMDPs can be viewed as a generalization of MDPs where the system state is not known perfectly; only an observation (affected by errors, missing data, or ambiguity) is available. By using a belief variable [cf. Sec. II-A] a POMDP can be recast as a MDP.

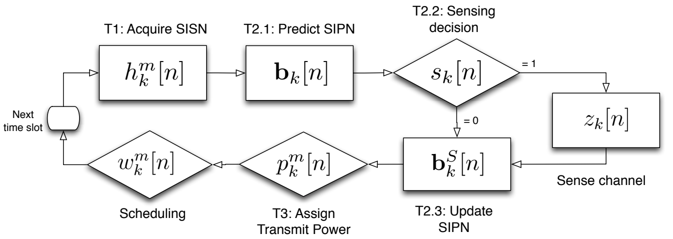


Fig. 1. Sequential operation of the CR system.

In addition, a two-step strategy will considerably reduce the computational burden to solve (8) without sacrificing optimality. To explain this strategy, let us revisit (and further clarify) the operation of the system. In Sec. II we explained that, at each slot n, our CR had to implement three main tasks: T1) acquisition of the SISN, T2) sensing and update of the SIPN, and T3) allocation of resources. In what follows, task T2 is split into 3 subtasks, so that the CR runs five sequential steps for each channel k:

- T1) $h_k^m[n]$ is acquired;
- T2.1) $\mathbf{b}_k[n]$ is computed using $\mathbf{P}_k$ and $\mathbf{b}_k^S[n-1]$ via (1);
- T2.2) $h_k^m[n]$ and $\mathbf{b}_k[n]$ are used to find $s_k^*[n]$;
- T2.3) $\mathbf{b}_k[n]$ and $z_k[n]$ (for the channels where $s_k^*[n] = 1$) are used to get $\mathbf{b}_k^S[n]$ via (2) and (3);
- T3) $h_k^m[n]$ 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.

The two-step strategy to solve (8) proceeds as follows. In step I, we obtain the expression for the optimal $w_k^m[n]$ and $p_k^m[n]$ for any sensing scheme. The resultant optimization problem is simpler than the original one in (8) because the dimensionality of the optimization space is smaller and the terms in (8) that depend only on $s_k[n]$ can be dropped. More importantly, if the sensing is not optimized, a per-slot optimization can be rendered optimal (see discussion on the per-slot separability of the Lagrangian in Sec. IV.A). In step II, we substitute the output of step I into (8) and solve for the optimal $s_k[n]$. Note that this does not entail a loss of optimality because the solution of step I is a *function* of the sensing scheme, and the latter is optimized in step II. Mathematically, for a generic function $f(x, y)$, the approach amounts to find $(x^*, y^*) = \arg\min_{x,y} f(x, y)$ as follows: i) $x^*(y) = \arg\min_x f(x, y)$ and ii) $y^* = \arg\min_y f(x^*(y), y)$. The last (trivial) step is to find x^* as $x^* = x^*(y^*)$. Here, the RA variables correspond to x and the sensing variables correspond to y. Clearly, the output of step I will be used in T2.2 (to find the optimal sensing) and in T3 (to find the optimal RA once the optimal sensing is known). The output of step II will be used in T2.2.[4] The optimization in step I (RA) is addressed next, while the one in step II (sensing) is addressed in Sec. V.

IV. Optimal RA for the secondary network

According to what we just explained, the objective of this section is to design the optimal RA (scheduling and powers) for a fixed sensing policy. Solving this problem is convenient because: i) it corresponds to one of the tasks our CR has to implement; ii) it is a much simpler problem than the original problem in (8), indeed the problem in this section has a smaller dimensionality and, more importantly, can be recast as a convex optimization problem; and iii) it will serve as an input for the design of the optimal sensing, simplifying the task of finding the global solution of (8).

Because in this section the sensing policy is considered given (fixed), $s_k[n]$ is not a design variable, and all the terms that depend only on $s_k[n]$ can be ignored. Specifically, the constraint in (8d) and the contribution of the sensing cost $U_S[n]$ to the total utility $U_T[n]$ in (8a) can be dropped. As a result, we define the new objective to optimize as $\bar{U}_{SU} := \sum_{k,m} \mathbb{E}\big[\lim_{N\to\infty} \sum_{n=0}^{N-1} (1-\gamma)\gamma^n \beta^m w_k^m[n] C_k^m(h_k^m[n], p_k^m[n])\big]$ and aim at solving the following problem [cf. (8)]

$$\max_{\{w_k^m[n], p_k^m[n]\}} \quad \bar{U}_{SU} \tag{9a}$$

$$\text{s. to}: \ (8b),\ (8c). \tag{9b}$$

A slightly modified version of this problem was recently solved in [7]. For this reason, we organize the remaining of this section into two parts. The first one summarizes (and adapts) the results in [7], presenting the optimal RA. The second part is devoted to introduce new variables that will serve as input for the design of the optimal sensing in Sec. V.

A. Solving for the RA

It can be shown that the problem in (9) can be reformulated as a convex one; see [7] as well as [23] for details. Specifically, if the constraint $w_k^m[n] \in \{0, 1\}$ is relaxed to yield $w_k^m[n] \in [0, 1]$ and an auxiliary (dummy) variable $x_k^m[n] := p_k^m[n] w_k^m[n]$ is introduced, then the problem is convex in $x_k^m[n]$ and $w_k^m[n]$. Moreover, with probability one (w.p.1) the solution to the relaxed problem is feasible (hence, optimal) for the original problem [7], [23]. The approach to solve the reformulated version of (9) is to dualize the long-term constraints in (8c). For such a purpose, let π^m and θ_k be the Lagrange multipliers associated with the (now convex) constraints (5) and (6), respectively, and define the following auxiliary variables:

$$\tilde{p}_k^m[n] \ := \ \Big[(\dot{C}_k^m)^{-1}\left(h_k^m[n], \pi^m/\beta^m\right)\Big]_+, \tag{10}$$

$$L_{SU,k}^m[n] := \ \beta^m C_k^m(h_k^m[n], \tilde{p}_k^m[n]) - \pi^m \tilde{p}_k^m[n], \text{ and} \tag{11}$$

$$L_k^m[n] \ := \ L_{SU,k}^m[n] - \theta_k B_k^S[n]. \tag{12}$$

Using (10)–(12), it can be shown that the optimal $w_k^{m*}[n]$ and $x_k^{m*}[n]$ that solve the convex version of (9) are

$$w_k^{m*}[n] := \mathbb{1}_{\{(L_k^m[n] = \max_q L_k^q[n]) \wedge (L_k^m[n] > 0)\}}. \tag{13}$$

and $x_k^{m*}[n] := \tilde{p}_k^m[n] w_k^{m*}[n]$. The latter readily implies that the optimal power allocation can be written as $p_k^{m*}[n] := \tilde{p}_k^m[n]$.

The auxiliary variables $L_{SU,k}^m[n]$ and $L_k^m[n]$ are useful to express the optimal RA, but also to gain insights on how the optimal RA operates. Both variables can be viewed as

[4] Note that the steps to obtain the optimal solution and those to implement the optimal solution during the CR operation do not follow the same order.

instantaneous reward indicators (IRIs) representing the reward that can be obtained if $w_k^m[n]$ is set to one. The indicator $L_{SU,k}^m[n]$ considers only SISN and represents the best achievable tradeoff between the rate and power transmitted by the SU. The risk of interfering the PU is considered in $L_k^m[n]$, which is obtained by adding an interference-related term to $L_{SU,k}^m[n]$. According to (13), only the SU with highest IRI can access the channel; moreover, if all users obtain a negative IRI, then none of them should access the channel (meaning that the utility for the SUs does not compensate for the risk of interfering the PU). The expressions in (10)–(13) also reveal the favorable structure of the optimal RA. The only parameters coupling users, channels and time slots are the multipliers (prices) π^m and θ_k. Once they are known, the Lagrangian of the convex version of (9) is separable, and the optimal RA at time n can be found using only information at time n.

Before moving to the next section, two comments are in order. The first one is related to the computation of the Lagrange multipliers. Several methods to set the value of the dual variables π^m and θ_k are available. Since, after relaxation, the problem has zero duality gap, there exists a *constant* (stationary) optimal value for each multiplier, denoted as π^{m*} and θ_k^*, such that substituting $\pi^m = \pi^{m*}$ and $\theta_k = \theta_k^*$ into (10) and (12) yields the optimal solution to the RA problem. Optimal Lagrange multipliers are rarely available in closed form and they have to be found through numerical search, for example, by using a dual subgradient method aimed at maximizing the dual function associated with (9) [24]. The second comment addresses the assumption of perfect SISN. Key to dealing with this issue is to acquire the instantaneous belief of the SISN, which is denoted as $H_k^m[n]$ and basically corresponds to the instantaneous distribution of the actual $h_k^m[n]$ at time n. Once $H_k^m[n]$ is available, the presence of imperfections has to be incorporated into the SU rate-power function $C(h_k^m[n], p_k^m[n])$, now $C(H_k^m[n], p_k^m[n])$, and substitute it into (10) and (11). Several alternatives to design $C(H_k^m[n], p_k^m[n])$ arise. Under a robust (worst-case) approach, the updated rate-power function would be $C(H_k^m[n], p_k^m[n]) = \min_{h_k^m[n] \in H_k^m[n]} C(h_k^m[n], p_k^m[n])$. Under a ergodic approach, it would be $C(H_k^m[n], p_k^m[n]) = \mathbb{E}_{h_k^m[n] \in H_k^m[n]}[C(h_k^m[n], p_k^m[n])]$. In any case, the basic structure (separability) of the RA remains the same.

B. RA as input for the design of the optimal sensing

The optimal solution in (10)-(11) will serve as input for the algorithms that design the optimal sensing scheme. For this reason, we introduce some auxiliary notation that will simplify the mathematical derivations in the next section. Specifically, let $L[n]$ be an auxiliary variable referred to as global IRI, which is defined as $L[n] := \sum_k L_k[n]$, with

$$L_k[n] := \sum\nolimits_m w_k^{m*}[n] L_k^m[n] \tag{14}$$

Due to the structure of the optimal RA, the IRI for channel k can be rewritten as [cf. (13), (12)]:

$$L_k[n] := \left[\max_q L_k^q[n]\right]_+ \tag{15}$$

Mathematically, $L[n]$ represents the contribution to the Lagrangian of (9) at instant n when $p_k^m[n] = p_k^{m*}[n]$ and $w_k^m[n] = w_k^{m*}[n]$ for all k and m. Intuitively, one can view $L[n]$ as the *instantaneous* functional that the optimal RA maximizes at instant n.

Key for the design of the optimal sensing is to understand the effect of the belief on the performance of the secondary network, thus, on $L[n]$. For such a purpose, we first define the IRI for the SUs in channel k as $L_{SU,k}[n] := \max_q L_{SU,k}^q[n]$. Then, we use $L_{SU,k}[n]$ to define the nominal IRI vector as

$$\boldsymbol{l}_k[n] := \begin{pmatrix} L_{SU,k}[n] \\ L_{SU,k}[n] - \theta_k[n] \end{pmatrix}. \tag{16}$$

Such a vector can be used to write $L_k[n]$ as a function of the belief $\mathbf{b}_k^S[n]$. Specifically,

$$L_k[n] = \left[\boldsymbol{l}_k^T[n]\mathbf{b}_k^S[n]\right]_+. \tag{17}$$

This suggests that the optimization of the sensing (which affects the value of $\mathbf{b}_k^S[n]$) can be performed separately for each of the channels. Moreover, (17) also reveals that $L_k[n]$ can be viewed as the expected IRI: $\left[\boldsymbol{l}_k[n]\right]_2$ is the IRI if the PU is present, $\left[\boldsymbol{l}_k[n]\right]_1$ is the IRI if it is not, and the entries of $\mathbf{b}_k^S[n]$ account for the corresponding probabilities, so that the expectation is carried over the SIPN uncertainties. Equally important, while the value of $\mathbf{b}_k^S[n]$ is only available after making the sensing decision, the value of $\boldsymbol{l}_k^T[n]$ is available before making such a decision. In other words, sensing decisions do not have an impact on $\boldsymbol{l}_k^T[n]$, but only on $\mathbf{b}_k^S[n]$. These properties will be exploited in the next section.

V. Optimal sensing

The aim of this section is to leverage the results of Secs. III and IV to design the optimal sensing scheme. Recall that current sensing decisions have an impact not only on the current reward (cost) of the system, but also on future rewards. This in turn implies that future sensing decisions are affected by the current decision, so that the sensing decisions across time form a string of events that has to be optimized jointly. Consequently, the optimization problem has to be posed as a DP. The section is organized as follows. First, we substitute the optimal RA policy obtained in Sec. IV into the original optimization problem presented in Sec. III and show that the design of the optimal sensing amounts to solving a set of separate unconstrained DP problems (Sec. V-A). Then, we obtain the solution to each of the DP problems formulated (Sec. V-B). It turns out that the optimal sensing leverages: ξ_k, the sensing cost at time n; the expected channel IRI at time n, which basically depends on $\boldsymbol{l}_k[n]$ (SISN) and the pre-decision belief (SIPN); and the future reward for time slots $n' > n$. The future reward is quantified by the value function associated with each channel's DP, which plays a fundamental role in the design of our sensing policies. Intuitively, a channel is sensed if there is uncertainty on the actual channel occupancy (SIPN) and the potential reward for the secondary network is high enough (SISN). The expression for the optimal sensing provided at the end of this section will corroborate this intuition.

A. Formulating the optimal sensing problem

The aim of this section is to formulate the optimal decision problem as a standard (unconstrained) DP. To do so, we substitute the optimal RA into the original optimization problem in (8). Recall that optimization in (8) involved variables $s_k[n]$, $w_k^m[n]$ and $p_k^m[n]$, and the sets of constraints in (8b), (8c) and (8d), the latter requiring $s_k[n] \in \{0,1\}$. When the optimal solution for $w_k^{m*}[n], p_k^{m*}[n]$ presented in Sec. IV is substituted into (8), the resulting optimization problem is

$$\max_{\{s_k[n]\}} \quad \bar{U}_{T|RA^*} \tag{18a}$$

$$\text{s. to :} \qquad s_k[n] \in \{0,1\}, \tag{18b}$$

where $\bar{U}_{T|RA^*}$ stands for the total utility given the optimal RA and is defined as

$$\bar{U}_{T|RA^*} := \mathbb{E}\Big[\lim_{N\to\infty} \sum_{n=0}^{N-1} (1-\gamma)\gamma^n \times \sum_k \Big(-\xi_k s_k[n] + \sum_m w_k^{m*}[n] L_k^m[n] \Big)\Big]. \tag{19}$$

Using definitions (14) and (17), we have that $\sum_m w_k^{m*}[n]L_k^m[n] = L_k[n] = \big[\mathbf{l}_k^T[n]\mathbf{b}_k^S[n]\big]_+$. Therefore, (19) can be rewritten as

$$\bar{U}_{T|RA^*} := \mathbb{E}\Big[\lim_{N\to\infty} \sum_{n=0}^{N-1} (1-\gamma)\gamma^n \times \sum_k \Big(-\xi_k s_k[n] + \Big[\boldsymbol{l}_k^T[n]\mathbf{b}_k^S[n]\Big]_+ \Big)\Big]. \tag{20}$$

The three main differences between (18) and the original formulation in (8) are that now: i) the only optimization variables are $s_k[n]$; ii) because the optimal RA fulfills the constraints in (8b) and (8c), the only constraints that need to be enforced are (8d), which simply require $s_k[n] \in \{0,1\}$ [cf. (18b)]; and iii) as a result of the Lagrangian relaxation of the DP, the objective has been augmented with the terms accounting for the dualized constraints.

Key to find the solution of (18) will be the facts that: i) $a_k[n]$ is independent of $h_k^m[n]$, ii) $a_k[n]$ is independent of $a_{k'}[n]$ for $k \neq k'$; and iii) $h_k^m[n]$ is independent of $h_{k'}^m[n]$ for $k \neq k'$. The former implies that the state transition functions for $a_k[n]$ do not depend on $h_k^m[n]$, while the two last allow to solve for each of the channels separately. These properties, together with the separability of (9) in the dual domain, allow us to obtain the optimal sensing policy by solving separate DPs (POMDPs), which will rely only on state information of the corresponding channel[5]. Specifically, the optimal sensing can be found as the solution of the following DP:

$$\max_{\{s_k[n]\in\{0,1\}\}} \sum_k \mathbb{E}\Big[\lim_{N\to\infty} \sum_{n=0}^{N-1} (1-\gamma)\gamma^n \times \Big(-\xi_k s_k[n] + \Big[\boldsymbol{l}_k^T[n]\mathbf{b}_k^S[n]\Big]_+ \Big)\Big], \tag{21}$$

which can be separated channel-wise into K simpler POMDPs. As a result, rather than exponential, the computational complexity to solve the optimal sensing is linear on K. Clearly, the reward function for the kth DP is

$$R_k[n] = -\xi_k s_k[n] + \Big[\boldsymbol{l}_k^T[n]\mathbf{b}_k^S[n]\Big]_+. \tag{22}$$

The structure of (22) manifests clearly that this is a joint design because $s_k[n]$ affects the two terms in (22). The first term (which accounts for the cost of the sensing decision) is just the product of the constant ξ_k and the sensing variable $s_k[n]$. The second term (which accounts for the reward of the RA) is the dot product of vectors $\boldsymbol{l}_k[n]$ (which does not depend on $s_k[n]$) and $\mathbf{b}_k^S[n]$ (which does depend on $s_k[n]$). The expression in (22) also reveals that $\boldsymbol{l}_k[n]$ encapsulates all the information pertaining to the SUs which is relevant to find $s_k^*[n]$. In other words, in lieu of knowing $h_k^m[n]$, $w_k^{m*}[n]$ and $p_k^{m*}[n]$, it suffices to know $\boldsymbol{l}_k[n]$.

Relying on (21) and (22), and taking into account that the problem can be separated across channels, *at each time slot* n the optimal sensing for channel k can be obtained as

$$s_k^*[n] = \arg\max_{s\in\{0,1\}} \Big\{ \lim_{N\to\infty} \sum_{t=n}^{N-1} (1-\gamma)\gamma^t \mathbb{E}\Big[R_k[t]|s_k[n]=s\Big]\Big\}. \tag{23}$$

B. Bellman equations and optimal solution

To find $s_k^*[n]$, we will derive the Bellman equations [15], [25] associated with (23). For such a purpose, we split the objective in (23) into the present and future rewards and drop the constant factor $(1-\gamma)\gamma^n$. Then, (23) can be rewritten as

$$s_k^*[n] = \arg\max_{s\in\{0,1\}} \Big\{ \mathbb{E}\Big[R_k[n]|s_k[n]=s\Big] + \gamma \lim_{N\to\infty} \sum_{t=n+1}^{N-1} \gamma^{t-n-1} \mathbb{E}\Big[R_k[t]|s_k[n]=s\Big]\Big\}. \tag{24}$$

It is clear that the expected reward at time slot $t = n$ depends on $s_k[n]$ –recall that both terms in (22) depend on $s_k[n]$. Moreover, the expected reward at time slots $t > n$ also depend on the current $s_k[n]$. The reason is that $\mathbf{b}_k^S[t]$ for $t > n$ depend on the $s_k[n]$ [cf. (3)]. This is testament to the fact that our problem is indeed a POMDP: current actions that improve the information about the current state have also an impact on the information about the state in future instants and thus, on future rewards.

To account for that effect in the formulation, we need to introduce the value function $V_k(\cdot)$ that quantifies the expected sum reward on channel k for all future instants. Since (23) is an infinite horizon problem with $\gamma < 1$, the value function is stationary and its existence is guaranteed [25]. Stationarity implies that the expression for $V_k(\cdot)$ is time-invariant. Since in our problem the state information is formed by the SISN and the SIPN, $V_k(\cdot)$ should be written as $V_k(B_k[n], \mathbf{h}_k[n])$. Since $\mathbf{h}_k[n]$ is i.i.d. across time and independent of $s_k[n]$, it can be rigorously shown that the optimal solution can be expressed in terms of the alternative value function $\bar{V}_k(B_k[n]) := \mathbb{E}_{\mathbf{h}}[V_k(B_k[n], \mathbf{h}_k[n])]$, where $\mathbb{E}_{\mathbf{h}}$

[5] Coupling among channels due to the constraint in (5) is implicitly accounted for via the optimum Lagrange multipliers and scheduling.

denotes that the expectation is taken over all possible values of $\mathbf{h}_k[n]$. This is useful not only because it emphasizes the fact that the impact of the sensing decisions on the future reward is encapsulated into $B_k[n]$, but also because $\bar{V}_k(\cdot)$ is a one-dimensional function, so that the numerical methods to compute it require lower computational burden that those to compute the original value function.

Based on the previous notation, the standard Bellman equations that drive the optimal sensing decision and the value function are (25) and (26) (shown at the bottom of the page) where $\mathbb{E}_{\mathbf{z}}$ denotes taking the expectation over the sensor outcome distribution. Equation (25) exploits the fact of the value function being stationary, manifests the dynamic nature of our problem, and provides further intuition about how sensing decisions have to be designed. The first term in (25), $\mathbb{E}_{\mathbf{z}}\left[R_k[n]\middle|s_k[n]=s\right]$, is the expected short-term reward conditioned to $s_k[n]=s$, while the second term, $\mathbb{E}_{\mathbf{z}}\left[\bar{V}_k(B_k[n+1])\middle|s_k[n]=s\right]$, is the expected long-term sum reward to be obtained in all future time instants, conditioned to $s_k[n]=s$ and that every future decision is optimal. Equation (26) expresses the condition that the value function $\bar{V}_k(\cdot)$ must satisfy in order to be optimal (and stationary) and provides a way to compute it iteratively.

Since obtaining the optimal sensing decision $s_k^*[n]$ at time slot n (and also evaluating the stationarity condition for the value function) boils down to evaluate the objective in (25) for $s_k[n]=0$ and $s_k[n]=1$, in the following we obtain the expressions for each of the two terms in (25) for both $s_k[n]=0$ and $s_k[n]=1$. Key for this purpose will be the expressions to update the belief presented in Sec. II-A. Specifically, expressions in (1)-(3) describe how the future beliefs depend on the current belief, on the set of possible actions (sensing decision), and on the random variables associated with those actions (outcome of the sensing process if the channel is indeed sensed).

The expressions for the expected *short-term reward* [cf. first summand in (25)] are the following. If $s_k[n]=0$, the channel is not sensed, there is no correction step, and the post-decision belief coincides with the pre-decision belief [cf. (2)]. The expected short-term reward in this case is:

$$\mathbb{E}_{\mathbf{z}}\left[R_k[n]\middle|s_k[n]=0\right]=\left[\boldsymbol{l}_k[n]^T\mathbf{b}_k[n]\right]_+. \tag{27}$$

On the other hand, if $s_k[n]=1$, the expected short-term reward is found by averaging over the probability mass of the sensor outcome $z_k[n]$ and subtracting the cost of sensing

$$\begin{aligned}\mathbb{E}_{\mathbf{z}}[R_k[n]\big|s_k[n]{=}1]=-\xi_k+\sum\nolimits_{z\in\{0,1\}}\\ \Pr(z_k[n]=z\big|\mathbf{b}_k[n])\left[\boldsymbol{l}_k[n]^T\mathbf{b}_k^S(\mathbf{b}_k[n],z)\right]_+,\end{aligned} \tag{28}$$

which, by substituting (3) into (28), yields

$$\mathbb{E}_{\mathbf{z}}[R_k[n]\big|s_k[n]{=}1]=-\xi_k+\sum_{z\in\{0,1\}}\left[\boldsymbol{l}_k[n]^T\mathbf{D}_z\mathbf{b}_k[n]\right]_+. \tag{29}$$

Once the expressions for the expected short-term reward are known, we find the expressions for the expected *long-term sum reward* [cf. second summand in (25)] for both $s_k[n]=0$ and $s_k[n]=1$. If $s_k[n]=0$, then there is no correction step [cf. (2)], and using (1)

$$\mathbb{E}_{\mathbf{z}}\left[\bar{V}_k(B_k[n+1])\big|s_k[n]{=}0\right]=\bar{V}_k\Big(\left[\mathbf{P}_k\mathbf{b}_k[n]\right]_2\Big). \tag{30}$$

On the other hand, if $s_k[n]=1$, the belief for instant n is corrected according to (3), and updated for instant $n+1$ using the prediction step in (1) as:

$$\begin{aligned}\mathbb{E}_{\mathbf{z}}\left[\bar{V}_k(B_k[n+1])\big|s_k[n]{=}1\right]=\\ \sum\nolimits_{z\in\{0,1\}}\Pr(z\big|\mathbf{b}_k[n])\bar{V}_k\Big(\left[\mathbf{P}_k\mathbf{b}_k^S(\mathbf{b}_k[n],z)\right]_2\Big)\end{aligned} \tag{31}$$

Clearly, the expressions for the expected *long-term reward* in (30) and (31) account for the expected value of $\bar{V}_k$ at time $n+1$. Substituting (27), (29), (30) and (31) into (26) yields (32) (shown at the bottom of the page) where for the last term we have used the expression for $\mathbf{b}_k^S(\mathbf{b}_k[n],z)$ in (3). The importance of (32) is that it allows to compute the value function numerically by means of well-studied algorithms such as value iteration and policy iteration [25, Ch. 3], which guarantee convergence to the optimal value function and, hence, to the optimal policy. Moreover, upon defining the auxiliary function $Q(s):=\mathbb{E}_{\mathbf{z}}\left[R_k[n]\middle|s_k[n]=s\right]+\gamma\mathbb{E}_{\mathbf{z}}\left[\bar{V}_k(B_k[n+1])\middle|s_k[n]=s\right]$, we can rewrite (25) as $s_k^*[n]=\arg\max_{s\in\{0,1\}}\{(1-s)Q(0)+sQ(1)\}$. Substituting (27)-(31) into the new expression, we get the optimal sensing at time n as (33) (shown at the bottom of the page).

Let us summarize the most relevant properties (several of them were already pointed out) of the optimal sensing policy in (33): i) it can be found separately for each of the channels; ii) since it amounts to a decision problem, we only have to evaluate the long-term aggregate reward if $s_k[n]=1$ (the channel is sensed at time n) and that if $s_k[n]=0$ (the channel is not sensed at time n), and make the decision which gives rise to a higher reward; iii) the reward takes into account not only the sensing cost but also the utility and QoS for the SUs (joint design); iv) the sensing at instant n is found as a function of both the instantaneous and the future reward (the problem is a DP); vi) the instantaneous reward depends on both the current SISN and the current SIPN, while the future reward depends on the current SIPN and not on the current SISN; and vii) to quantify the future reward, the value function $\bar{V}_k(\cdot)$ is required. The input of this function is the SIPN. Additional insights on the optimal sensing will be given in Sec. VII.

$$s_k^*[n]=\arg\max_{s\in\{0,1\}}\left\{\mathbb{E}_{\mathbf{z}}\left[R_k[n]\middle|s_k[n]=s\right]+\gamma\mathbb{E}_{\mathbf{z}}\left[\bar{V}_k(B_k[n+1])\middle|s_k[n]=s\right]\right\} \tag{25}$$

$$\bar{V}_k(B_k[n])=\mathbb{E}_{\mathbf{h}}\Big[\max_{s\in\{0,1\}}\left\{\mathbb{E}_{\mathbf{z}}\left[R_k[n]\middle|s_k[n]=s\right]+\gamma\mathbb{E}_{\mathbf{z}}\left[\bar{V}_k(B_k[n+1])\middle|s_k[n]=s\right]\right\}\Big], \tag{26}$$

To conclude this section, some comments regarding the computational complexity associated with the optimal sensing policy are in order. Note that, once the optimal value functions $\bar{V}_k(\cdot)$ are known for all k, the per-slot computational complexity boils down to evaluating (33) once per channel. Since there are K channels and computing $\boldsymbol{l}_k[n]$ entails a maximization over M terms [cf. (16)], the associated complexity is $O(KM)$. Similarly, once the optimal Lagrange multipliers are known, the per-slot complexity associated with the RA task is $O(KM)$ [cf. (13)]. Clearly, it is necessary to find the optimal multipliers and value functions; this can be done off-line by means of iterative methods whose complexity has been well studied in the literature [24], [14].

C. Sensing cost

To account for the cost of sensing a given channel, the additive and constant cost ξ_k was introduced. So far, we considered that the value of ξ_k was pre-specified. However, the value of ξ_k can be tuned to represent physical properties of the CR. Space limitations prevent us from elaborating formally on this issue, so that only three illustrative examples are given. Example 1: If the NC spends a power P_k^{NC} to sense channel k, then ξ_k can be set to $\xi_k = \pi^{NC} P_k^{NC}$, where π^{NC} stands for the Lagrange multiplier associated with a long-term power constraint on the NC. Example 2: Consider a setup for which the long-term rate of sensing is limited; mathematically, this can be accomplished by imposing that $\mathbb{E}[\lim_{N\to\infty}\sum_{n=0}^{N-1}(1-\gamma)\gamma^n s_k[n]] \leq \eta$, where η represents the maximum sensing rate (say 10%). Let ρ_k be the Lagrange multiplier associated with such a constraint, in this scenario ξ_k should be set to $\xi_k = \rho_k$. Example 3: Suppose now that to sense the channel, the NC needs a fraction ψ_k of total duration of the slot (that, otherwise, would have been used for SU transmissions). In this scenario, $\xi_k[n] = \psi_k L_k[n]$ (time-varying opportunity cost). Linear combinations and stochastic versions of any of those costs are possible too. Similarly, if a collaborative sensing scheme is assumed, aggregation of costs across users can also be considered.

VI. Numerical results

Numerical experiments to corroborate the theoretical findings and gain insights on the optimal policies are implemented in this section. Since an *RA scheme* similar to the one presented in this paper was analyzed in [7], the focus is on analyzing the properties of the optimal *sensing scheme*. The readers interested can find additional simulations as well as the Matlab codes used to run them in http://www.tsc.urjc.es/∼amarques/simulations/NumSimulations_lramjr12.html.

The experiments are grouped into two test cases. In the first one, we compare the performance of our algorithms with that of other existing (suboptimal) alternatives. Moreover, we analyze the behavior of the sensing schemes and assess the impact of variation of different parameters (correlation of the PUs activity, sensing cost, sensor quality, and average SNR). In the second test case, we provide a graphical representation of the sensing functions in the form of two-dimensional decision maps. Such representation will help us to understand the behavior of the optimal schemes.

The parameters for the default test case are listed in Table II. We consider $K = 4$ channels, each of them with different values for the sensor quality, the sensing cost and the QoS requirements. In most cases, the value of $\check{o}_k$ has been chosen to be larger than the value of P_k^{MD} (so that the cognitive diversity can be effectively exploited), while the values of the remaining parameters have been chosen so that the test-case yields illustrative results. The secondary link gains are Rayleigh distributed, their average SNR is 5dB, and the frequency selectivity is such that the gains are uncorrelated across channels. We consider $M = 4$ SUs, and their average power limits are set to $[\check{p}_1, \check{p}_2, \check{p}_3, \check{p}_4] = [20, 16, 18, 10]$. The discount factor γ is set to 0.95, so that the autocorrelation function of $a_k[n]$ and the implemented window have comparable length. The multipliers associated with the system parameters have been calculated by gradient descent, using a Monte Carlo approach to average over the channel processes. Unless otherwise stated, the remaining parameters are set to one.

TABLE II
Parameters of the system under test.

k	P_k^{FA}	P_k^{MD}	$\mathbf{P}_k$	ξ_k	$\check{o}_k$
1	0.09	0.08	[0.95, 0.05; 0.02, 0.98]	1.00	0.30
2	0.09	0.08	[0.95, 0.05; 0.02, 0.98]	1.80	0.05
3	0.05	0.03	[0.95, 0.05; 0.02, 0.98]	1.00	0.10
4	0.05	0.03	[0.95, 0.05; 0.02, 0.98]	1.80	0.10

Test case 1: Optimality and performance analysis. The objective here is twofold. First, we want to numerically demonstrate that our schemes are indeed optimal. Second,

$$\begin{aligned}\bar{V}_k(B_k[n]) \quad = \quad & \mathbb{E}_{\mathbf{h}}\Bigg[\max\Bigg\{\Big[\boldsymbol{l}_k[n]^T\mathbf{b}_k[n]\Big]_+ + \gamma\bar{V}_k\left(\big[\mathbf{P}_k\mathbf{b}_k[n]\big]_2\right); \\ - \quad & \xi_k + \sum\nolimits_{z\in\{0,1\}}\left(\Big[\boldsymbol{l}_k[n]^T\mathbf{D}_z\mathbf{b}_k[n]\Big]_+ + \gamma\Pr(z_k[n]\big|\mathbf{b}_k[n])\bar{V}_k\left(\frac{\big[\mathbf{P}_k\mathbf{D}_z\mathbf{b}_k[n]\big]_2}{\mathbf{1}^T\mathbf{D}_z\mathbf{b}_k[n]}\right)\right)\Bigg\}\Bigg].\end{aligned} \tag{32}$$

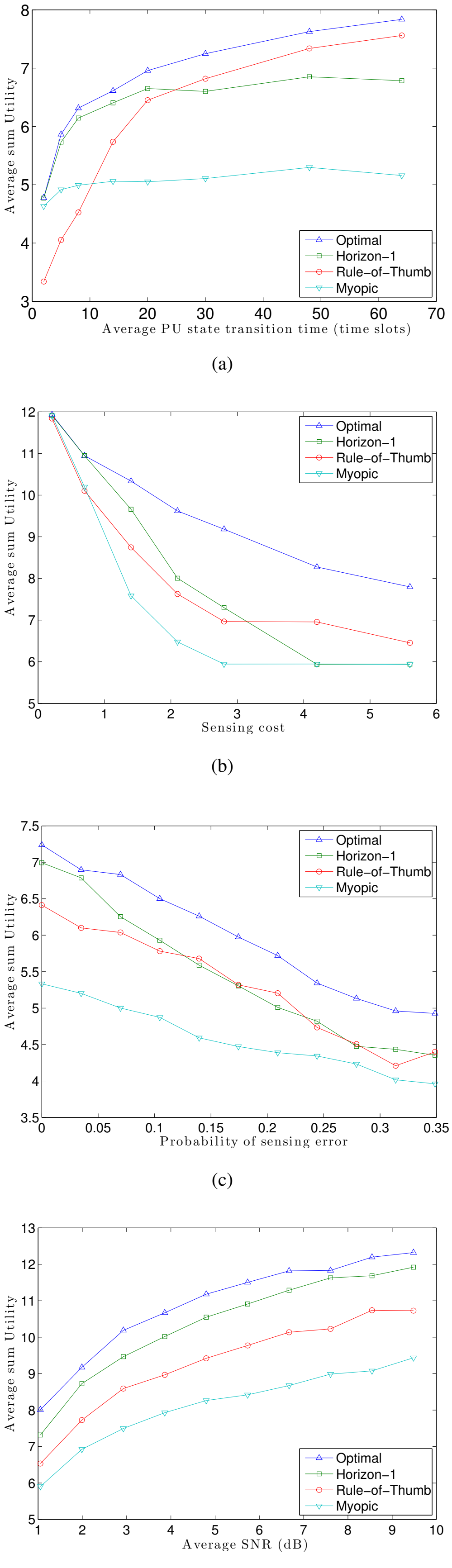
$$\begin{aligned}s_k^*[n] = \arg\max_{s\in\{0,1\}}&\Bigg\{(1-s)\Bigg[\Big[\boldsymbol{l}_k^T[n]\mathbf{b}_k[n]\Big]_+ + \gamma\bar{V}_k\Big(\big[\mathbf{P}_k\mathbf{b}_k[n]\big]_2\Big)\Bigg] \\ &+ s\Bigg[-\xi_k + \sum\nolimits_{z\in\{0,1\}}\left(\Big[\boldsymbol{l}_k^T[n]\mathbf{D}_z\mathbf{b}_k[n]\Big]_+ + \gamma\Pr(z_k[n]\big|\mathbf{b}_k[n])\bar{V}_k\left(\frac{\big[\mathbf{P}_k\mathbf{D}_z\mathbf{b}_k[n]\big]_2}{\mathbf{1}^T\mathbf{D}_z\mathbf{b}_k[n]}\right)\right)\Bigg]\Bigg\}.\end{aligned} \tag{33}$$

we are also interested in assessing the loss of optimality incurred by suboptimal schemes with low computational burden. Specifically, the optimal sensing scheme is compared with the three suboptimal alternatives described next. A) A myopic policy, which is implemented by setting $\bar{V}(B) = 0 \ \forall B$ and is equivalent to the greedy sensing and RA technique proposed in [6]. B) A policy that replaces the infinite horizon value function with a ”horizon-1” value function. (i.e., it makes the sensing decision at time n considering the expected reward for instants n and $n+1$). C) A rule-of-thumb sensing scheme implementing the simple (separable) decision function: $s_k[n] = \mathbb{1}_{\{L_k[n]\in[\xi_k,\theta_k-\xi_k]\}}\mathbb{1}_{\{B_k[n]\in[\mathbf{b}_k^S(A_k,0),\mathbf{b}_k^S(A_k,1)]\}}$. This simple policy senses the channel only if two conditions hold: a) the IRI $L_k[n]$ is neither too low nor too high (so that the scheduling decision is uncertain), and b) the instantaneous belief $B_k[n]$ is close to the long-term belief A_k (so that the information given by the observation will modify the value of $B_k[n]$ substantially).

To evaluate the effect of the variation of different parameters, we run 4 experiments. In each experiment, one of the parameters is swept and the rest remain constant (with the values listed in Table II): a) average PU state transition time (by modifying $\mathbf{P}_k$); b) sensing cost; c) probability of sensing error (by modifying P_k^{FA} and P_k^{MD}) and d) average SNR. Results are plotted in Fig. 2. The slight lack of monotonicity observed in the curves is due to the fact that simulations have been run using a Monte-Carlo approach. As expected, the optimal sensing scheme achieves the best performance for all test cases. Moreover, Figs. 2(a) and 2(b) reveal that the ”horizon-1” value function approximation constitutes a good approximation to the optimal value function in two cases: i) when the expected PU transition time is short (low time correlation) and ii) when the sensing cost is relatively small. The performance of the myopic policy is shown to be far from the optimal. This finding is in disagreement with the results obtained for simpler models in the opportunistic spectrum access literature [12] where it was suggested that the myopic policy could be a good approximation to solve the associated POMDP efficiently. The reason can be that the CR models considered were substantially different (the RA schemes in this paper are more complex and the interference constraints are formulated differently). In fact, the only cases where the myopic policy seems to approximate the optimal performance are: i) if $\xi_k \to 0$, this is unsurprising because then the optimal policy is to sense at every time instant; and ii) if the PUs activity is not correlated across time (which was the assumption in [6]).

Fig. 2(c) suggests that the benefits of implementing the optimal sensing policies are stronger when sensors are inaccurate. In other words, the proposed schemes can help to soften the negative impact of deploying low-quality (low-cost) sensing devices. Finally, results in Fig. 2(d) also suggest that changes in the average SNR between SU and NC, have similar effects on the performance of all analyzed schemes.

Test case 2: Sensing decision maps. To gain insights on the behavior of the optimal sensing schemes, Fig. 3 plots the sensing decisions as a function of $B_k[n]$ and $L_{SU,k}[n]$. Simulations are run using the parameters for the default test



(a)

(b)

(c)

(d)

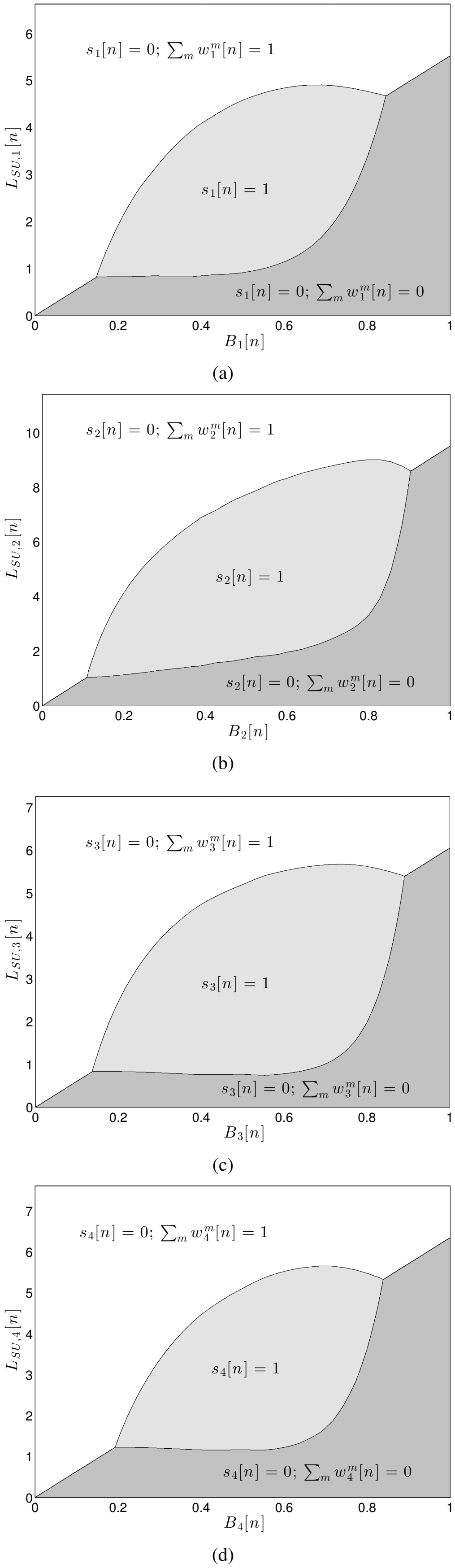
Fig. 2. Performance of the optimal scheme vs. some suboptimal schemes for variations in (a) expected primary transition time; (b) sensing cost; (c) probability of error; (d) average SNR.

case (see Table II) and each subplot corresponds to a different channel k. Since the domain of the sensing decision function is two dimensional, the function itself can be visually displayed as a decision map. Two main regions are identified, one corresponding to the pairs $(B_k[n], L_k[n])$ which give rise to $s_k[n] = 1$, and one corresponding to the pairs giving rise to $s_k[n] = 0$. Moreover, the region where $s_k[n] = 0$ is split into two subregions, the first one corresponding to $\sum_m w_k^m[n] = 1$ (i.e., when there is a user accessing the channel) and the second one when $\sum_m w_k^m[n] = 0$ (i.e., when the system decides that no user will access the channel). Note that for the region where $s_k[n] = 1$, the access decision basically depends on the outcome of the sensing process $z_k[n]$ (if fact, it can be rigorously shown that $\sum_m w_k^m[n] = 1$ if and only if $z_k[n] = 1$).

Upon comparing the different subplots, one can easily conclude that the size and shape of the $s_k[n] = 1$ region depend on $\mathbf{P}_k$, P_k^{FA}, P_k^{MD}, ξ_k, and $\check{o}_k$. For example, the simulations reveal that channels with stricter interference constraint need to be more frequently sensed and thus the sensing region is larger: Fig. 3(a) vs. Fig. 3(b). They also reveal that if the sensing cost ξ_k increases, then the sensing region becomes smaller: Fig. 3(c) vs. Fig. 3(d). This is not surprising: more expensive sensing implies more conservative sensing decisions.

VII. Summarizing conclusions and future work

The aim of this paper was to design jointly optimal RA and sensing schemes for an interweave CR with multiple primary and secondary users. Since, due to time correlation in the SIPN, sensing decisions were coupled across time, the problem fell into the class of DP (more precisely, POMDP). To address the complexity typically associated with DP, both the objective and the QoS constraints were formulated as long-term time averages (which are less restrictive than their short-term counterparts and, hence, give rise to a better global objective). Additionally, a discounted, infinite time-horizon formulation was chosen for the DP (giving rise to stationary value functions). Duality was used to handle the QoS constraints and dual decomposition methods were leveraged to facilitate the optimization. A two-step approach that solved first for the optimal RA for *any* sensing scheme and, then, solved for the optimal sensing was implemented. As a result, the DP finally solved had a state space much smaller than that of the original formulation. In particular, the optimization was separable across channels, partially separable across SUs, and separable across time. More specifically, to find the optimal solution at time n, the required inputs were: the SISN and SIPN at n; the stationary Lagrange *multipliers* associated with the dualized constraints; and the stationary *value function* associated with the long-term objective. Both the value functions and the multipliers accounted for the effect of sensing and RA in time instants other than n. The expressions for the optimal RA and the optimal sensing policies were intuitive and relatively simple. Provided that the stationary value function and multipliers were available (they are found off-line via numerical search during the initialization phase), the online implementation of the optimal schemes entailed



(a)

(b)

(c)

(d)

Fig. 3. Decision maps (regions) for the four channels in the default test case (see Table II). The light gray area in the center corresponds to the sensing decision.

very low computational complexity. Numerical results showed that both optimal and near-optimal policies corresponding to the formulated DP perform significantly better than myopic policies, conversely to the model studied in [12].

There are multiple meaningful ways to extend our results. One of them is to develop low-complexity stochastic estimations for the value functions and the multipliers. Those are useful to decrease complexity and to cope with cases where the statistics of the random processes are not known or they are not stationary. Dual stochastic subgradient methods [6], [7] can be used to estimate online the Lagrange multipliers; to estimate the value function, alternatives such as Q-learning [25, Ch.8] can be explored. Considering more complex models for the CSI is also a line of work. For example, non-Markovian models for the PU activity can be used (this requires looking for efficient ways to represent and update the belief, e.g., by using recursive Bayesian estimation; see [7] and references therein). Additional sources of correlation (correlation across time for the SISN and correlation across channels for the SIPN) can be considered too, rendering the POMDP more challenging to solve. Another line of work is to address the optimal design for underlay CR networks. In such a case, information about the channel gains between the SUs and PUs would be required. In this paper we limited the interference to the PU by bounding the average probability of interference. Formulations limiting the average interfering power or the average rate loss due to the interfering power are other reasonable options. Last but not least, developing distributed implementations for our novel schemes (addressing cooperative sensing and distributed RA) is also a relevant line of work. For some of these extensions, designs based on suboptimal but low-complexity solutions may be an alternative worth exploring.

References

[1] A. Goldsmith, S. A. Jafar, I. Mari, and S. Srinivasa, "Breaking spectrum gridlock with cognitive radios: An information theoretic perspective," *Proc. IEEE*, vol. 97, no. 5, pp. 894–914, May 2009.

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

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

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

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

[7] A. G. Marques, L. M. Lopez-Ramos, G. B. Giannakis, and J. Ramos, "Resource allocation for interweave and underlay CRs under probability-of-interference constraints", *IEEE J. Sel. Areas Commun.*, vol. 30, no. 10, pp. 1922–1933, Nov. 2012.

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

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

[10] A. G. Marques, L. M. Lopez-Ramos, and J. Ramos, "Cognitive Radios with Ergodic Capacity Guarantees for Primary Users," in *Proc. Intl. Conf. on Cognitive Radio Oriented Wireless Networks*, Stockholm, Sweden, Jun. 18-20, 2012.

[11] 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.

[12] 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.

[13] L. P. Kaelbling, M. L. Littman, and A.R. Cassandra, "Planning and acting in partially observable stochastic domains", *Artificial Intelligence*, vol. 101, pp. 99–139, Jan. 1998.

[14] M. Wiering, and van Otterlo, M. (Eds.), *Reinforcement Learning: State-of-the-art (Vol.12)*, Springer, 2012.

[15] D. Bertsekas, *Dynamic Programming and Optimal Control*, Athena Scientific, 1995.

[16] 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.

[17] D. A. Castañon, "Approximate Dynamic Programming for sensor management," in *Proc. IEEE Conf. on Decision and Contr.*, San Diego, CA, Dec. 10–12, 1997.

[18] 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," in *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Process.*, Prague, Czech Rep., May. 22–27, 2011.

[19] X. Zhang and H. Su, "Opportunistic spectrum sharing schemes for CDMA-based uplink MAC in cognitive radio networks," *IEEE J. Sel. Areas Commun.*, vol. 29, no. 4, pp. 716–730, Apr. 2011.

[20] A. Ghasemi and E. Sousa, "Collaborative spectrum sensing for opportunistic access in fading environments," in *Proc. IEEE Int. Symposium on New Frontiers in Dynamic Spectrum Access Networks*, Baltimore, Maryland, USA, Nov. 8–11, 2005.

[21] L. Li and A. J. Goldsmith, "Capacity and optimal resource allocation for fading broadcast channels–Part I: Ergodic capacity," *IEEE Trans. Inf. Theory*, vol. 47, no. 3, pp. 1083–1102, Mar. 2001.

[22] R. Zhang, YC Liang, and S. Cui, "Dynamic resource allocation in cognitive radio networks," *IEEE Signal Process. Mag.*, vol. 27, no. 3, pp.102–114, May 2010.

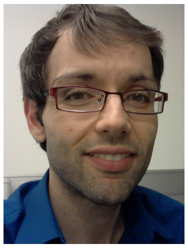
[23] A. G. Marques, G. B. Giannakis, and J. Ramos, "Optimizing orthogonal multiple access based on quantized channel state information," *IEEE Trans. Signal Process.*, vol. 59, no. 10, pp. 5023–5038, Oct. 2011.

[24] D. Bertsekas, *Nonlinear Programming*, Athena Scientific, 1999.

[25] W. B. Powell, *Approximate Dynamic Programming: Solving the Curses of Dimensionality*, Wiley-Interscience, 2007.

Luis M. Lopez-Ramos (S10) received the B.Sc. degree (with highest honors) in telecommunications engineering from King Juan Carlos University (URJC), Madrid, Spain, in 2010 and the M.Sc. degree in multimedia and communications from Carlos III University of Madrid, Madrid, Spain, in 2012. He is currently pursuing the Ph.D. degree in telecommunication engineering at URJC. In 2013, he was a visiting scholar at the Dept. of Electrical Engineering, University of Minnesota, Minneapolis.

Since 2010, he has been a Research Assistant with the Dept. of Signal Theory and Communications at URJC, where he has developed research and teaching activities under the Spanish Ministry of Education's FPU program. His research interests include nonlinear optimization, dynamic programming and reinforcement learning, and their applications on wireless networks, cognitive radios, and power networks.



Antonio G. Marques (SM13) received the Telecommunications Engineering degree and the Doctorate degree (together equivalent to the B.Sc., M.Sc., and Ph.D. degrees in electrical engineering), both with highest honors, from the Carlos III University of Madrid, Spain, in 2002 and 2007, respectively. In 2003, he joined the Dept. of Signal Theory and Communications, King Juan Carlos University, Madrid, Spain, where he currently develops his research and teaching activities as an Associate Professor. Since 2005, he has held different visiting positions at the Dept. of Electrical Engineering, University of Minnesota, Minneapolis.

His research interests lie in the areas of communication theory, signal processing, and networking. His current research focuses on stochastic resource allocation for wireless systems, cognitive radios, nonlinear network optimization, and signal processing for graphs. Dr. Marques work has been awarded in several conferences and workshops.

Javier Ramos received the B.Sc and M.Sc. degrees from the Polytechnic University of Madrid, Spain. Between 1992 and 1995 he participated in several research projects at Purdue University, Indiana, USA, working on the field of Signal Processing for Communications. He received the Ph.D. degree on 1995. During 1996 he was Post-Doctoral Research Associate at Purdue University. Dr. Ramos received the Ericsson award to the best Ph.D. dissertation on Mobile Communications in 1996. From 1997 to 2003 he was an Associate Professor at Carlos III University of Madrid. Since 2003 Dr. Ramos is the Dean of the Telecommunications Engineering School at King Juan Carlos University, Madrid, Spain.

His current research interests are broadband wireless services technologies, distributed sensing, and eHealth.