

# Spectrum Sensing in Cognitive Radio Networks via Approximate POMDP

Bharath Keshavamurthy, Nicolò Michelusi

**Abstract**—Cognitive radio technologies will be critical to the wireless communication infrastructure due to their potential to alleviate the spectrum crunch, both in the commercial and the military spheres. In this paper, we propose a novel spectrum sensing and access strategy based on Partially Observable Markov Decision Processes (POMDPs) wherein a cognitive radio node learns the correlation model defining the occupancy behavior of the incumbents and devises an optimal strategy to perform spectrum sensing and access that exploits the learned correlation model. To ameliorate the complexity of the POMDP optimization, we employ the PERSEUS algorithm, an approximate value iteration method. We compare numerically the performance of the proposed framework with state-of-the-art MAP-based state estimators and correlation-coefficient based clustering algorithms, and demonstrate that the proposed system achieves the performance attained by MAP-based state estimators which possess prior knowledge about the correlation model; outperforms the correlation-coefficient based state-of-the-art by an average of 50%; and surpasses a Neyman-Pearson Detector that assumes independence among channels, by an average of 33%.

**Index Terms**—Hidden Markov Model, Cognitive Radio, Spectrum Sensing, POMDP

## I. INTRODUCTION

With the advent of fifth-generation wireless communication networks, the problem of spectrum scarcity has been exacerbated [1]. For some time now, cognitive radio technologies have been in the spotlight as a potential solution to this problem in commercial and military applications [2]. Cognitive radio networks facilitate efficient spectrum utilization by intelligently accessing *white spaces* left unused by the sparse and infrequent transmissions of licensed users, while ensuring rigorous incumbent non-interference compliance [3]. A crucial aspect underlying the design of cognitive radio networks is the channel access protocol in the MAC layer of the stack. In this regard, the current state-of-the-art involves channel access strategies dictated by multi-armed bandits [4], reinforcement learning agents [5], and other custom heuristics [6]–[8]. However, almost all these works, such as [4], [5], [9], [10], assume independence among channels in the discretized spectrum which is imprudent because licensed users exhibit correlation across both frequency and time in their channel occupancy behavior: the primary users frequently occupy a set of adjacent channels (frequency correlation), repeating similar motifs in behavior over an extended period of time (temporal correlation) [11]–[13]. This pattern in occupancy behavior of the incumbents imputes very high levels of correlation among channels which need to be leveraged for more accurate predictions of spectrum holes. In this paper, we propose a

parameter estimation algorithm to learn the aforementioned correlation model, and an approach to solve for the optimal channel sensing and access policy that exploits this learned correlation structure.

The works [4], [14] develop spectrum sensing and access algorithms under the assumption that the occupancy behavior of incumbents is independent across both time and frequencies. In our work, we exploit both frequency and temporal correlations. In [15], a compressed spectrum sensing scheme is devised that exploits sparse temporal dynamics in the occupancy of licensed users, and in [16], an efficient spectrum sensing strategy is proposed for dense multi-cell cognitive networks, that also exploits the spatial structure of interference; however, both works assume independence across frequencies. Spectrum sensing and access strategies in a distributed multi-agent cognitive radio setting have been considered in [5] and solved using SARSA with linear value function approximation. However, frequency correlation is precluded, and errors in state estimation are neglected in the decision process. Unlike [5], we consider a model with correlation across frequencies, and we account for uncertainty in the occupancy state via a POMDP formulation. Although the spectrum sensing algorithms detailed in [6] consider the correlation in incumbent occupancy behavior across frequencies, the authors assume a perfect, noise-free observation model, unlike which we construct a more reliable and realistic AWGN model; and like other works modeling correlation across both time and frequencies, such as [7], [8], the algorithms operate based on a data-driven strategy wherein pre-loaded databases are employed offline to estimate the correlation models, while we, in our work, present a fully online framework that estimates the correlation models and simultaneously solves for the optimal channel sensing and access policy.

The rest of the paper is organized as follows: in Sec. II, we define the signal model, followed by the formulations, approaches, and algorithms in Sec. III; in Sec. IV, we present numerical evaluations, followed by our conclusions in Sec. V.

## II. SYSTEM MODEL

### A. Signal Model

We consider a network consisting of  $P$  licensed users termed the Primary Users (PUs) and one cognitive radio node termed the Secondary User (SU) equipped with a spectrum sensor. The objective of the SU is to opportunistically access portions of the spectrum left unused by the PUs in order to maximize its own throughput. To this end, the SU should learn how to intelligently access spectrum holes (white spaces) intending to maximize its throughput while maintaining strict non-interference compliance with incumbent transmissions.

This research has been funded in part by NSF under grant CNS-1642982.

The authors are with the School of Electrical and Computer Engineering, Purdue University. email: {bkeshava,michelusi}@purdue.edu.

The wideband signal received at the SU receiver at time  $n$  is denoted as  $y(n)$  and is given by

$$y(n) = \sum_{p=1}^P \sum_{l=0}^{L_p-1} h_p(l) x_p(n-l) + v(n), \quad (1)$$

where  $y(n)$  is expressed as a convolution of the signal  $x_p(n)$  of the  $p$ th PU with the channel impulse response  $h_p(n)$ , and  $v(n)$  denotes additive white Gaussian noise (AWGN) with variances  $\sigma_v^2$ . Eq. (1) can be written in the frequency domain by taking a  $K$ -point DFT which decomposes the observed wideband signal into  $K$  discrete narrow-band components as

$$Y_k(i) = \sum_{p=1}^P H_{p,k}(i) X_{p,k}(i) + V_k(i), \quad (2)$$

where  $i \in \{1, 2, 3, \dots, T\}$  represents the time index;  $k \in \{1, 2, 3, \dots, K\}$  represents the index of the components in the frequency domain;  $V_k(i) \sim \mathcal{CN}(0, \sigma_v^2)$  represents circularly symmetric additive complex Gaussian noise, i.i.d across frequency and across time, and independent of  $H$  and  $X$ ;  $X_{p,k}(i)$  is the signal of the  $p$ th PU in the frequency domain, and  $H_{p,k}(i)$  is its frequency domain channel. We further assume that the  $P$  PUs employ an orthogonal access to the spectrum (e.g., OFDMA) so that  $X_{p,k}(i) X_{q,k}(i) = 0, \forall p \neq q$ . Thus, letting  $p_k$  be the index of the PU that contributes to the signal in the  $k$ th spectrum band (possibly,  $p_k = 0$  if no PU is transmitting in the  $k$ th spectrum band), and letting  $H_k(i) = H_{p_k,k}(i)$  and  $X_k(i) = X_{p_k,k}(i)$ , we can rewrite (2) as

$$Y_k(i) = H_k(i) X_k(i) + V_k(i). \quad (3)$$

Thus,  $H_k(i)$  represents the  $k$ th DFT coefficient of the impulse response  $h_{p_k,k}(n)$  of the channel in between the PU operating on the  $k$ th spectrum band and the SU, at time  $i$ ; we model it as a zero-mean circularly symmetric complex Gaussian random variable with variance  $\sigma_H^2$ ,  $H_k \sim \mathcal{CN}(0, \sigma_H^2)$ , i.i.d. across frequency bands, over time, and independent of the occupancy state of the channels.

### B. PU Spectrum Occupancy Model

We now introduce the model of PU occupancy over time and across the frequency domain. We model each  $X_k(i)$  as

$$X_k(i) = \sqrt{P_{tx}} B_k(i) S_k(i), \quad (4)$$

where  $P_{tx}$  is the transmission power of the PUs,  $S_k(i)$  is the transmitted symbol modelled as a constant amplitude signal,  $|S_k(i)|=1$ , i.i.d. over time and across frequency bands;<sup>1</sup>  $B_k(i) \in \{0, 1\}$  is the binary spectrum occupancy variable, with  $B_k(i)=1$  if the  $k$ th spectrum band is occupied by a PU at time  $i$ , and  $B_k(i)=0$  otherwise. Therefore, the PU occupancy behavior in the entire wideband spectrum of interest at time  $i$ , discretized into narrow-band frequency components can be modeled as the vector

$$\vec{B}(i) = [B_1(i), B_2(i), B_3(i), \dots, B_K(i)]^T \in \{0, 1\}^K. \quad (5)$$

<sup>1</sup>In the case where  $S_k(i)$  does not have constant amplitude, we may approximate  $H_k(i) S_k(i)$  as complex Gaussian with zero mean and variance  $\sigma_H^2 \mathbb{E}[|S_k(i)|^2]$ , without any modification to the subsequent analysis.

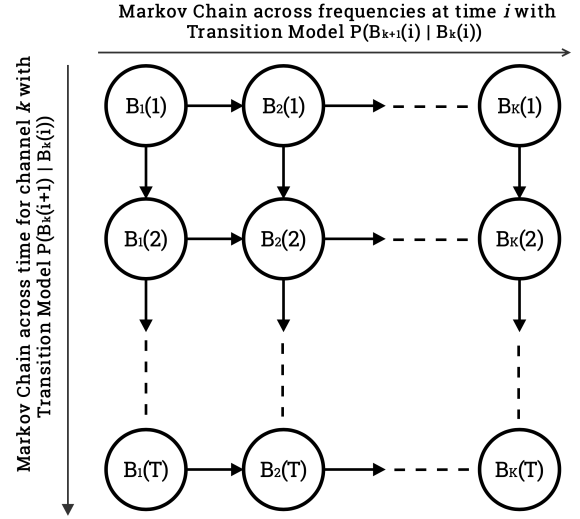


Fig. 1. The correlation model across time and frequencies underlying the occupancy behavior of incumbents in the network

PUs join and leave the spectrum at random times. To capture this temporal correlation in the spectrum occupancy dynamics of PUs, we model  $\vec{B}(i)$  as a Markov process,

$$\mathbb{P}(\vec{B}(i+1) | \vec{B}(j), \forall j \leq i) = \mathbb{P}(\vec{B}(i+1) | \vec{B}(i)). \quad (6)$$

Additionally, when joining the spectrum pool, PUs occupy a number of adjacent spectrum bands, and may vary their spectrum needs depending on traffic demands, channel conditions, etc. To capture this behavior, we model  $\vec{B}(i)$  as having Markovian correlation across the bands as,

$$\begin{aligned} \mathbb{P}(\vec{B}(i+1) | \vec{B}(i)) \\ = \mathbb{P}(B_1(i+1) | B_1(i)) \prod_{k=2}^K \mathbb{P}(B_k(i+1) | B_k(i), B_{k-1}(i+1)). \end{aligned} \quad (7)$$

That is, the spectrum occupancy at time  $i+1$  in frequency band  $k$ ,  $B_k(i+1)$ , depends on the occupancy state of the adjacent spectrum band at the same time,  $B_{k-1}(i+1)$ , and that of the same spectrum band  $k$  in the previous time index  $i$ ,  $B_k(i)$  as shown in Fig. 1.

### C. Spectrum Sensing Model

In order to detect the available spectrum holes, the SU performs spectrum sensing. However, owing to physical design limitations at the SU's spectrum sensor [17], not all channels in the discretized spectrum can be sensed at once: the SU can sense only  $\kappa$  out of  $K$  spectrum bands at any given time, with  $1 \leq \kappa \leq K$ . Let  $\mathcal{K}_i \subseteq \{1, 2, \dots, K\}$  with  $|\mathcal{K}_i| \leq \kappa$  be the set of indices of spectrum bands sensed by the SU at time  $i$ , which is part of our design. Then, we define the observation vector

$$\vec{Y}(i) = [Y_k(i)]_{k \in \mathcal{K}_i}, \quad (8)$$

where  $Y_k(i)$  is given by (3). The true states  $\vec{B}(i)$  encapsulate the actual occupancy behavior of the PU and the measurements at the SU are noisy observations of these true states which are

modeled to be the observed states of a Hidden Markov Model (HMM). Conditional on  $\vec{B}(i)$  and  $\mathcal{K}_i$ , the probability density function of  $\vec{Y}(i)$  is expressed as

$$f(\vec{Y}(i)|\vec{B}(i), \mathcal{K}_i) = \prod_{k \in \mathcal{K}_i} f(Y_k(i)|B_k(i)), \quad (9)$$

owing to the independence of channels, noise, and transmitted symbols across frequency bands. Moreover, from (3),

$$Y_k(i)|B_k(i) \sim \mathcal{CN}(0, \sigma_H^2 P_{tx} B_k(i) + \sigma_V^2). \quad (10)$$

#### D. POMDP Agent Model

In this section, we model the spectrum access scheme of the SU as a POMDP, whose goal is to devise an optimal sensing and access policy in order to maximize its throughput while maintaining strict non-interference compliance with incumbent transmissions. In fact, the agent's limited sensing capabilities coupled with its noisy observations result in an increased level of uncertainty at the agent's end about the occupancy state of the spectrum under consideration and the exact effect of executing an action on the radio environment. The transition model of the underlying MDP as described by (7), is denoted by  $\mathbf{A}$  and is learned by the agent by interacting with the radio environment (see Sec. III). The emission model is denoted by  $\mathbf{M}$  and is given by (9), with  $f(Y_k(i)|B_k(i))$  given by (10).

We model the POMDP as a tuple  $(\mathcal{B}, \mathcal{A}, \mathcal{Y}, \mathbf{P}, \mathbf{M})$  where  $\mathcal{B} \equiv \{0, 1\}^K$  represents the state space of the underlying MDP with states  $\vec{B}$  given by all possible realizations of the spectrum occupancy vector as described by (3),  $\mathcal{A}$  represents the action space of the agent, given by all possible combinations in which the  $\kappa$  spectrum bands are chosen to be sensed out of  $K$  at any given time; and  $\mathcal{Y}$  represents the observation space of the agent based on the signal model outlined in the previous subsection. The state of the POMDP at time  $i$  is given by the *prior belief*  $\beta_i$ , which represents the probability distribution of the underlying MDP state  $\vec{B}(i)$ , given the information collected by the agent up to time  $i$ , but before collecting the new information in slot  $i$ . At the beginning of each time index  $i$ , given  $\beta_i$ , the agent selects  $\kappa$  spectrum bands out of  $K$  according to a policy  $\pi(\beta_i)$ , thus defining the sensing set  $\mathcal{K}_i$ , performs spectrum sensing on these spectrum bands, observes  $\vec{Y}(i) \in \mathcal{Y}$ , and updates its *posterior belief*  $\hat{\beta}_i$  of the current spectrum occupancy  $\vec{B}(i)$  as

$$\begin{aligned} \hat{\beta}_i(\vec{B}') &= \mathbb{P}(\vec{B}(i) = \vec{B}' | \vec{Y}(i), \mathcal{K}_i, \beta_i) \\ &= \frac{\mathbb{P}(\vec{Y}(i) | \vec{B}', \mathcal{K}_i) \beta_i(\vec{B}')}{\sum_{\vec{B}'' \in \{0,1\}^K} \mathbb{P}(\vec{Y}(i) | \vec{B}'', \mathcal{K}_i) \beta_i(\vec{B}'')}. \end{aligned} \quad (11)$$

We denote the function that maps the prior belief  $\beta_i$  to the posterior belief  $\hat{\beta}_i$  through the spectrum sensing action  $\mathcal{K}_i$  and the observation signal  $\vec{Y}(i)$  as  $\hat{\beta}_i = \mathbb{B}(\beta_i, \mathcal{K}_i, \vec{Y}(i))$ .

Given the posterior belief  $\hat{\beta}_i$ , we estimate the occupancy state of the discretized spectrum under consideration as

$$\vec{B}(i)^* = \arg \max_{\vec{B} \in \mathcal{B}} \hat{\beta}_i(\vec{B}). \quad (12)$$

Let  $B_k(i)^* = \phi_k(\hat{\beta}_i) \in \{0, 1\}$  be the state of channel  $k$  at time  $i$ . If the channel is deemed to be idle as a result of

this MAP estimation procedure, i.e.,  $\phi_k(\hat{\beta}_i) = 0$ , the SU accesses the channel for delivering its network flows. Else, it leaves it untouched. Given the PU occupancy state  $\vec{B}(i)$  and posterior belief  $\hat{\beta}_i$ , the reward metric of the POMDP is given by the number of *truly idle* bands detected by the SU accounting for the throughput maximization aspect of the agent's objective and a penalty for *missed detections* accounting for the incumbent non-interference constraint, i.e.,

$$R(\vec{B}(i), \hat{\beta}_i) = \sum_{k=1}^K (1 - B_k(i))(1 - \phi_k(\hat{\beta}_i)) - \lambda B_k(i)(1 - \phi_k(\hat{\beta}_i)), \quad (13)$$

where  $\lambda > 0$  represents the cost term penalizing the agent for missed detections, i.e. interference with the incumbent. After performing data transmission, the SU computes the prior belief for the next slot based on the dynamics of the Markov chain as

$$\beta_{i+1}(\vec{B}') = \mathbb{P}(\vec{B}(i+1) = \vec{B}' | \hat{\beta}_i). \quad (14)$$

We denote the function that maps the posterior belief  $\hat{\beta}_i$  to the prior belief  $\beta_{i+1}$  as  $\beta_{i+1} = \mathbb{B}(\hat{\beta}_i)$ . The goal of the problem at hand is to determine an optimal spectrum sensing policy to maximize the infinite-horizon discounted reward,

$$\pi^* = \arg \max_{\pi} V^{\pi}(\beta) \triangleq \mathbb{E}_{\pi} \left[ \sum_{i=0}^{\infty} \gamma^i R(\vec{B}(i), \hat{\beta}_i) | \beta_0 = \beta \right], \quad (15)$$

where  $0 < \gamma < 1$  is the discount factor,  $\beta_0$  is the initial belief, and  $\hat{\beta}_i$  is the posterior belief induced by policy  $\pi$ , which maps the beliefs  $\beta_i$  to actions  $\mathcal{K}_i$  at time  $i$ , and we have defined the value function  $V^{\pi}(\beta)$  under policy  $\pi$  starting from belief  $\beta$ . The optimal policy  $\pi^*$  and the corresponding optimal reward  $V^*(\beta)$  are the solutions of Bellman's optimality equation  $V^* = H[V^*]$ , where we have defined the operator  $V_{n+1} = H[V_n]$  as,  $\forall \beta$ ,

$$\begin{aligned} V_{n+1}(\beta) &= \max_{\mathcal{K} \in \mathcal{A}} \sum_{\vec{B} \in \mathcal{B}} \beta(\vec{B}) \mathbb{E}_{\vec{Y} | \vec{B}, \mathcal{K}} \left[ R(\vec{B}, \mathbb{B}(\beta, \mathcal{K}, \vec{Y})) + \right. \\ &\quad \left. \gamma V_n(\mathbb{B}(\mathbb{B}(\beta, \mathcal{K}, \vec{Y}))) \right]. \end{aligned} \quad (16)$$

This problem can be solved using the value iteration algorithm, i.e., by solving (16) iteratively until convergence to a fixed point. However, given the high dimensionality of the spectrum sensing and access problem, i.e. the number of states of the underlying MDP scales exponentially with the number of bands in the spectrum, solving equation (16) using Exact Value Iteration and Policy Iteration algorithms is computationally infeasible. Additionally, solving for the optimal policy from equation (16) requires prior knowledge about the underlying MDP's transition model. Therefore, in this paper we present a framework to estimate the transition model of the underlying MDP online, while utilizing this learned model to solve for the optimal policy by employing Randomized Point-Based Value Iteration techniques, namely, the PERSEUS algorithm [18].

### III. APPROACHES AND ALGORITHMS

#### A. Occupancy Behavior Transition Model Estimation

In real-world implementations of cognitive radio systems, the transition model of the occupancy behavior of the PUs is not known to the SUs in the network and needs to be learned over time. The learned model then needs to be fed back to the POMDP agent which is solving for the optimal spectrum sensing and access policy simultaneously. Inherently, the approach constitutes solving a parameter estimation problem formulated as

$$\mathbf{A}^* = \arg \max_{\mathbf{A}} \mathbb{P}([\vec{Y}(i)]_{i=1}^{\tau} | \mathbf{A}), \quad (17)$$

which is a Maximum Likelihood Estimation (MLE) problem, where  $\mathbf{A}$  is defined as  $\mathbb{P}(\vec{B}(i+1) | \vec{B}(i))$  and  $\tau$  refers to the learning period of the parameter estimator: this can be equal to the entire duration of the POMDP agent's interaction with the radio environment implying simultaneous model learning or can be a predefined parameter learning period before triggering the POMDP agent. In order to facilitate better readability, for the description of this parameter estimator, we denote  $[\vec{Y}(i)]_{i=1}^{\tau}$  as  $\mathbf{Y}$  and  $[\vec{B}(i)]_{i=0}^{\tau}$  as  $\mathbf{B}$ . Re-framing (17) as an optimization of the log-likelihood, we get,

$$\mathbf{A} = \arg \max_{\mathbf{A}} \log \left( \sum_{\mathbf{B}} \mathbb{P}(\mathbf{B}, \mathbf{Y} | \mathbf{A}) \right). \quad (18)$$

This problem can be solved using the Expectation-Maximization algorithm, where the E-step constitutes

$$Q(\mathbf{A} | \hat{\mathbf{A}}^{(t)}) = \mathbb{E}_{\mathbf{B} | \mathbf{Y}, \hat{\mathbf{A}}^{(t)}} \left[ \log \left( \sum_{\mathbf{B}} \mathbb{P}(\mathbf{B}, \mathbf{Y} | \hat{\mathbf{A}}^{(t)}) \right) \right], \text{ and} \quad (19)$$

the M-step constitutes

$$\hat{\mathbf{A}}^{(t+1)} = \arg \max_{\mathbf{A}} Q(\mathbf{A} | \hat{\mathbf{A}}^{(t)}). \quad (20)$$

In order to solve for  $a_{LR} \in \mathbf{A}$  where,  $L, R \in \{0, 1\}^K$  and bring the solution into a simplified iterative form, we use the forward and backward probability members defined as

$$\begin{aligned} F(i, R) &= \mathbb{P}([\vec{Y}(t)]_{t=1}^i, \vec{B}(i) = R) \\ &= \sum_L \mathbb{P}(\vec{B}(i) = R, \vec{Y}(i) | \vec{B}(i-1) = L) F(i-1, L), \end{aligned} \quad (21)$$

$$\begin{aligned} D(i, L) &= \mathbb{P}([\vec{Y}(t)]_{t=i}^{\tau} | \vec{B}(i-1) = L) \\ &= \sum_R \mathbb{P}(\vec{B}(i) = R, \vec{Y}(i) | \vec{B}(i-1) = L) D(i+1, R). \end{aligned} \quad (22)$$

Using these definitions, the iterative procedure turns out to be,

$$a_{LR}^{(t+1)} = \frac{\sum_{i=1}^{\tau} F(i-1, L) \mathbf{M}_R(\vec{Y}(i)) a_{LR}^{(t)} D(i+1, R)}{\sum_R \sum_{i=1}^{\tau} F(i-1, L) \mathbf{M}_R(\vec{Y}(i)) a_{LR}^{(t)} D(i+1, R)}, \quad (23)$$

until the convergence condition  $|a_{LR}^{(t+1)} - a_{LR}^{(t)}| \leq \epsilon$  is satisfied, where  $\epsilon > 0$ , a very small quantity.

#### B. The PERSEUS Algorithm

As discussed in Sec. II of this article, solving the Bellman equation (16) for POMDPs with large state and action spaces using exact value iteration and policy iteration techniques [19] is computationally infeasible [18], [20]. Hence, we resort to approximate value iteration techniques to ensure that the system scales well to a large number of bands in the spectrum of interest. For infinite-horizon POMDPs,  $V^*$  in (16) can be approximated by a Piece-Wise Linear and Convex function (PWLC) [19]. The core idea behind the PERSEUS algorithm is that the value function in time index  $i$  can be parameterized by a set of hyperplanes  $\{\vec{\alpha}_i^u\}$ ,  $u = \{0, 1, 2, \dots, |V_i|\}$ , each of which represents a region of the belief space for which it is the maximizing element. The backup step is defined as the process of determining the optimal hyperplane out of the set of available hyperplanes in time index  $i$  as

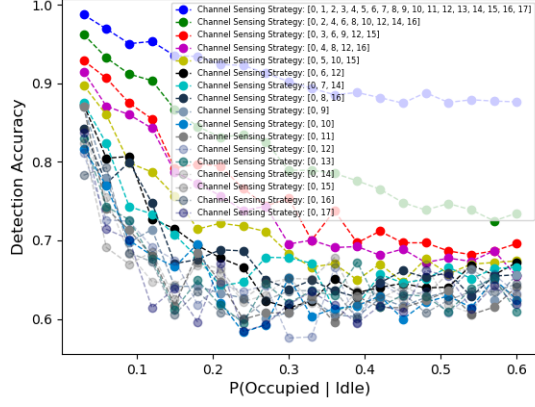
$$\vec{\alpha}_i(\beta) = \arg \max_{\vec{\alpha}_i^u, u \in \{0, 1, 2, \dots, |V_i(\beta)|\}} \beta \cdot \vec{\alpha}_i^u, \quad (24)$$

which implies that,

$$\begin{aligned} V_i(\beta) &= \max_{\vec{\alpha}_i^u, u \in \{0, 1, 2, \dots, |V_i(\beta)|\}} \beta \cdot \vec{\alpha}_i^u, \\ \pi_i(\beta) &= a(\vec{\alpha}_i^{u*}), \end{aligned} \quad (25)$$

where  $a(\vec{\alpha}_i^u)$  is the action corresponding to hyperplane  $\vec{\alpha}_i^u$ . PERSEUS is a randomized point-based value iteration algorithm whose goal is to improve the value of all belief points by updating the value and gradient of only a subset of these belief points, chosen at random. The procedure constitutes an initial phase wherein the POMDP agent randomly interacts with the radio environment to collect a set of so-called *reachable beliefs* which are to be improved over numerous iterative backup stages. In each backup stage, the agent samples a belief  $\beta$  uniformly at random from the set of unimproved points and performs a backup on this sampled belief point according to (24). **[NM: how do you compute the new hyperplanes? You are not showing that] [BK: Isn't that given by (24)? What other details should I add?] [NM: ok, how do you determine the new hyperplanes then? (24) is just telling you how to approximate the V function given hyperplanes, but it does not tell you how hyperplanes are constantly improved upon in the algorithm..]** to determine the optimal hyperplane  $\vec{\alpha}$ . Considering an arbitrary time index  $i$ , if  $V_{i+1}(\beta) = \beta \cdot \vec{\alpha} \geq V_i(\beta)$ , then the belief point  $\beta$  is said to be improved along with any other belief points  $\beta'$  in the unimproved set for which  $V_{i+1}(\beta') = \beta' \cdot \vec{\alpha} \geq V_i(\beta')$ . If  $V_{i+1}(\beta) = \beta \cdot \vec{\alpha} < V_i(\beta)$ , then a copy of the maximizing hyperplane for  $V_i(\beta)$  is used for  $V_{i+1}(\beta)$  and the belief point  $\beta$  is then removed from the set of unimproved points. The backup stage continues until the set of unimproved points is empty and the agent performs a series of backup stages until there the number of policy changes between consecutive iterations is below a specified threshold  $\eta$ .

The belief update procedure outlined in (11) is an essential aspect of the PERSEUS algorithm which can turn into a performance bottleneck for large state spaces due to the inherent iteration over all possible states. In order to circumvent this problem, we fragment the spectrum into much smaller,

Detection Accuracy v/s  $P(\text{Occupied} | \text{Idle})$  for 18 channels at  $P(X_i = 1) = 0.6$  with varying uniform channel sensing strategiesFig. 2. The detection accuracies of the constrained Viterbi algorithm for different sensing strategies over varying values of  $P(\text{Occupied} | \text{Idle})$ 

independent sets of correlated channels and then run the PERSEUS algorithm on these fragments by leveraging multi-processing and multi-threading tools available at our disposal in software frameworks. Furthermore, we avoid iterating over all possible states and allow only those state transitions we deem to be the most probable - for example, we allow only those state transitions that involve a Hamming distance of up to 3 between the previous state vector and the current state vector in an 18 channel radio environment. [NM: you need to provide more details here, including math.] [BK: Do you mean more details about PERSEUS in general or more details about my modifications to PERSEUS (fragmentation and belief simplification)?] [NM: I mean, you are hastily presenting your modifications of PERSEUS but you should describe them in more details.. anyway, believe that this is a lower priority thing, given the space constraint, so fix all the rest first and then we can decide about this one.]

#### IV. NUMERICAL EVALUATION

[NM: I thought you were working on the state of the art comparison, where is that?] The given framework is simulated in Python for a system with 18 channels and a channel model constituting an SNR of 19dB when an incumbent occupies a specific channel. The same Markovian transition model, emission model, and steady-state model is employed across both the chain across time and the chain across frequencies. The first line trace (in blue) depicted in Fig. 2 illustrates the detection accuracies of the Viterbi algorithm wherein the SU makes observations of all the channels in the radio environment and estimates the occupancy states of these channels over varying values of  $P(\text{Occupied} | \text{Idle})$ , i.e., as the channels transition toward independence. We note that the detection accuracy of this MAP-based state estimator degrades as the channels transition toward independence, which is as surmised, because the Viterbi algorithm's structure begins to crumble if the Markovian correlation ceases to exist among cells in the grid, where a cell corresponds to the state of a certain channel at a given time index. This plot is an important illustration to prove that leveraging correlation in incumbent

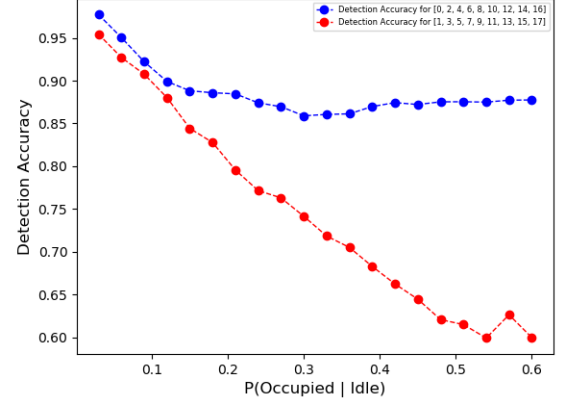
Detection Accuracy v/s  $P(\text{Occupied} | \text{Idle})$  for 18 channels at  $P(X_i = 1) = 0.6$  with uniform channel sensing strategy [0, 2, 4, 6, 8, 10, 12, 14, 16]

Fig. 3. The detection accuracies of the constrained Viterbi algorithm for sensed and un-sensed channels under a given channel sensing strategy

Mean Square Error Convergence of the Markov Correlated Parameter Estimation Algorithm for a Static PU with Complete Information

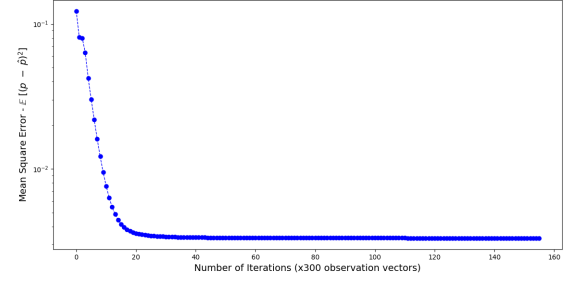


Fig. 4. The mean square error convergence plot of the parameter estimation algorithm[NM: Show x axis up to 40]

occupancy behavior across both time and frequencies gives us a boost in detection accuracy, and thereby, a boost in secondary network throughput while complying with the required non-interference constraints. Fig. 2 also illustrates the detection accuracies of the Viterbi algorithm (refer to all the other line traces) wherein the SU makes observations of only the channels in the given channel sensing strategy and estimates the occupancy states of all the channels in the discretized spectrum of interest over varying values of  $P(\text{Occupied} | \text{Idle})$ . In Fig. 3, the plot differentiates the trend in detection accuracies for the sensed and the un-sensed channels in relation to a constrained Viterbi algorithm in which the SU only senses channels whose indices correspond to the multiples of 2 and uses these channels to estimate the occupancy behavior of the incumbents across all 18 channels over varying values of  $p = P(\text{Occupied} | \text{Idle})$ . In both these plots i.e., Fig. 2 and 3, we observe that, as anticipated, the detection accuracy of this MAP-based state estimator deteriorates as the amount of information available for estimation decreases. These plots illustrate that the imposition of sensing limitations on the SU's spectrum sensor in view of quick turnaround times and improved energy efficiency, will put a dent in the detection accuracy of the cognitive radio node, and this needs to be addressed by leveraging the correlation in incumbent occupancy behavior across both time and frequencies. The plot

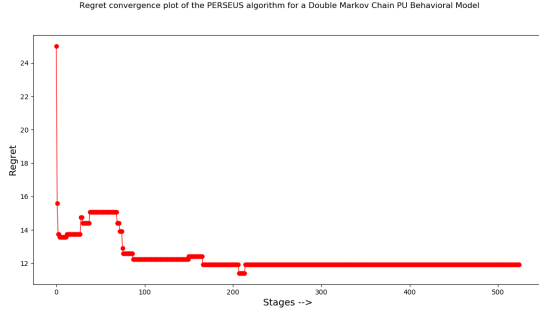


Fig. 5. The regret convergence plot of the PERSEUS algorithm over several backup and wrapper stages

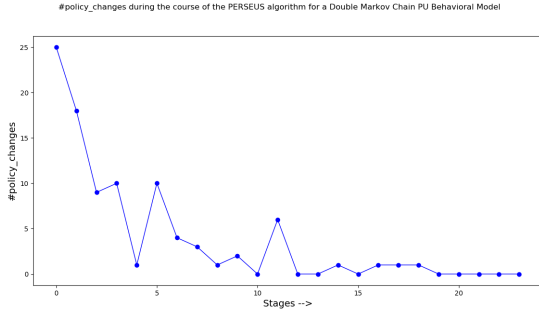


Fig. 6. The number of policy changes involved in the PERSEUS algorithm as it transitions toward convergence over numerous backup and wrapper stages

depicted in Fig. 4 shows the Mean Square Error convergence of the Parameter Estimation algorithm while determining the transition model of the MDP underlying the problem under consideration. Starting with an initial estimate of 0.0, the EM algorithm detailed in Sec. III converges to the true transition model with an error of  $\epsilon \leq 10^{-8}$  over numerous iterations, each iteration corresponding to an averaging operation of 300 observation vectors. Here, we observe the mean square error given by  $\mathbb{E}[(p - \hat{p})^2]$  iteratively reduces as it goes through the E-step which finds the distribution for the parameters in the current time-step, given the previous estimate and the M-step which determines the maximum likelihood estimate of the parameters, given the distribution obtained in the E-step. It has been theoretically shown to converge, i.e., each iteration either improves the true likelihood or leaves it unchanged [21]. Since the EM algorithm is susceptible to premature convergence to local optima and saddle points, we mitigate this by averaging the procedure over several cycles. Fig. 5 illustrates the Regret Convergence plot of the PERSEUS algorithm over several backup and wrapper stages wherein the regret metric corresponds to the difference in utility between the PERSEUS algorithm at a certain stage and an Oracle which has complete information with respect to the occupancy behavior of the incumbents in the radio environment. Furthermore, the algorithm involves an online estimation of the transition model of the underlying MDP and a random exploration strategy to gather the initial set of reachable beliefs. Fig. 6 depicts the number of policy changes involved in each individual stage of the PERSEUS algorithm as it moves toward

convergence. The termination condition for the PERSEUS algorithm is that the number of policy changes over several consecutive backup stages should be zero, i.e.,  $\eta = 0$ . These plots, similar to the *Reward v/s Time* and  $\Delta\pi$  v/s *time* plots in [18], serve as a measure of convergence for our fragmented PERSEUS algorithm with simplified belief updates, and an online transition model estimation.

## V. CONCLUSION

In this paper, we formulate the optimal spectrum sensing and access problem in an AWGN observation model with multiple licensed users and a cognitive radio node restricted in terms of its sensing capabilities, as a POMDP. In a radio environment wherein the occupancy behavior of the incumbents is correlated across time and frequencies, we present a consolidated framework that employs the Expectation-Maximization algorithm to estimate the transition model of this occupancy behavior and leverage a fragmented PERSEUS algorithm with belief update heuristics to simultaneously solve for the optimal spectrum sensing and access policy. Through system simulations, we show that our framework outperforms the existing correlation-coefficient based state-of-the-art; surpasses Neyman-Pearson based occupancy detection schemes that fail to exploit the correlation across time and frequencies; and matches the performance of standard MAP-estimators which possess the transition model statistics as an apriori. **[NM: you need to cut down the reference list to 15-16 max]**

## REFERENCES

- [1] C. Pradhan, K. Sankhe, S. Kumar, and G. R. Murthy, "Revamp of enodeb for 5g networks: Detracting spectrum scarcity," in *2015 12th Annual IEEE Consumer Communications and Networking Conference (CCNC)*, Jan 2015, pp. 862–868.
- [2] M. Danneberg, R. Datta, A. Festag, and G. Fettweis, "Experimental testbed for 5g cognitive radio access in 4g lte cellular systems," in *2014 IEEE 8th Sensor Array and Multichannel Signal Processing Workshop (SAM)*, June 2014, pp. 321–324.
- [3] F. Xu, L. Zhang, Z. Zhou, and Y. Ye, "Architecture for next-generation reconfigurable wireless networks using cognitive radio," in *2008 3rd International Conference on Cognitive Radio Oriented Wireless Networks and Communications (CrownCom 2008)*, May 2008, pp. 1–5.
- [4] K. Cohen, Q. Zhao, and A. Scaglione, "Restless multi-armed bandits under time-varying activation constraints for dynamic spectrum access," in *2014 48th Asilomar Conference on Signals, Systems and Computers*, Nov 2014, pp. 1575–1578.
- [5] J. Lundén, S. R. Kulkarni, V. Koivunen, and H. V. Poor, "Multiagent reinforcement learning based spectrum sensing policies for cognitive radio networks," *IEEE Journal of Selected Topics in Signal Processing*, vol. 7, no. 5, pp. 858–868, Oct 2013.
- [6] M. Gao, X. Yan, Y. Zhang, C. Liu, Y. Zhang, and Z. Feng, "Fast spectrum sensing: A combination of channel correlation and markov model," in *2014 IEEE Military Communications Conference*, Oct 2014, pp. 405–410.
- [7] C. Park, S. Kim, S. Lim, and M. Song, "Hmm based channel status predictor for cognitive radio," in *2007 Asia-Pacific Microwave Conference*, Dec 2007, pp. 1–4.
- [8] G. Ding, J. Wang, Q. Wu, L. Yu, Y. Jiao, and X. Gao, "Joint spectral-temporal spectrum prediction from incomplete historical observations," in *2014 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, Dec 2014, pp. 1325–1329.
- [9] J. Oksanen, V. Koivunen, J. Lundén, and A. Huttunen, "Diversity-based spectrum sensing policy for detecting primary signals over multiple frequency bands," in *2010 IEEE International Conference on Acoustics, Speech and Signal Processing*, March 2010, pp. 3130–3133.



- [10] S. Chaudhari, V. Koivunen, and H. V. Poor, "Autocorrelation-based decentralized sequential detection of ofdm signals in cognitive radios," *IEEE Transactions on Signal Processing*, vol. 57, no. 7, pp. 2690–2700, July 2009.
- [11] S. Yin, D. Chen, Q. Zhang, M. Liu, and S. Li, "Mining spectrum usage data: A large-scale spectrum measurement study," *IEEE Transactions on Mobile Computing*, vol. 11, no. 6, pp. 1033–1046, June 2012.
- [12] R. I. C. Chiang, G. B. Rowe, and K. W. Sowerby, "A quantitative analysis of spectral occupancy measurements for cognitive radio," in *2007 IEEE 65th Vehicular Technology Conference - VTC2007-Spring*, April 2007, pp. 3016–3020.
- [13] M. A. McHenry, P. A. Tenhula, D. McCloskey, D. A. Roberson, and C. S. Hood, "Chicago spectrum occupancy measurements & analysis and a long-term studies proposal," in *Proceedings of the First International Workshop on Technology and Policy for Accessing Spectrum*, ser. TAPAS '06. New York, NY, USA: ACM, 2006. [Online]. Available: <http://doi.acm.org/10.1145/1234388.1234389>
- [14] L. Ferrari, Q. Zhao, and A. Scaglione, "Utility maximizing sequential sensing over a finite horizon," *IEEE Transactions on Signal Processing*, vol. 65, no. 13, pp. 3430–3445, July 2017.
- [15] N. Michelusi and U. Mitra, "Cross-layer estimation and control for cognitive radio: Exploiting sparse network dynamics," *IEEE Transactions on Cognitive Communications and Networking*, vol. 1, no. 1, pp. 128–145, March 2015.
- [16] N. Michelusi, M. Nokleby, U. Mitra, and R. Calderbank, "Multi-Scale Spectrum Sensing in Dense Multi-Cell Cognitive Networks," *IEEE Transactions on Communications*, vol. 67, no. 4, pp. 2673–2688, April 2019.
- [17] S. Maleki, S. P. Chepuri, and G. Leus, "Energy and throughput efficient strategies for cooperative spectrum sensing in cognitive radios," in *2011 IEEE 12th International Workshop on Signal Processing Advances in Wireless Communications*, June 2011, pp. 71–75.
- [18] M. T. J. Spaan and N. A. Vlassis, "Perseus: Randomized point-based value iteration for pomdps," *CoRR*, vol. abs/1109.2145, 2011. [Online]. Available: <http://arxiv.org/abs/1109.2145>
- [19] L. P. Kaelbling, M. L. Littman, and A. R. Cassandra, "Planning and acting in partially observable stochastic domains," *Artif. Intell.*, vol. 101, no. 1-2, pp. 99–134, May 1998. [Online]. Available: [http://dx.doi.org/10.1016/S0004-3702\(98\)00023-X](http://dx.doi.org/10.1016/S0004-3702(98)00023-X)
- [20] J. Pineau, G. Gordon, and S. Thrun, "Point-based value iteration: An anytime algorithm for pomdps," in *Proceedings of the 18th International Joint Conference on Artificial Intelligence*, ser. IJCAI'03. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2003, pp. 1025–1030. [Online]. Available: <http://dl.acm.org/citation.cfm?id=1630659.1630806>
- [21] R. M. Neal and G. E. Hinton, *A View of the Em Algorithm that Justifies Incremental, Sparse, and other Variants*. Dordrecht: Springer Netherlands, 1998, pp. 355–368. [Online]. Available: [https://doi.org/10.1007/978-94-011-5014-9\\_12](https://doi.org/10.1007/978-94-011-5014-9_12)