

Learning-aided Sub-band Selection Algorithms for Spectrum Sensing in Wide-band Cognitive Radios

Yang Li[†], Sudharman K. Jayaweera[†], Mario Bkassiny[†] and Chittabrata Ghosh[‡]

[†]Department of Electrical and Computer Engineering, University of New Mexico, Albuquerque, NM, USA

Email: {yangli, jayaweera, bkassiny}@ece.unm.edu

[‡]Nokia Research Center, Berkeley, CA, USA

Email: chittabrata.ghosh@nokia.com

Abstract—We propose wide-band spectrum sensing scheduling solutions for cognitive radios that are equipped with reconfigurable RF front-ends. The wide frequency spectrum of interest is segmented into frequency sub-bands due to software and hardware limitations. These sub-bands can be non-contiguous, and each may contain an arbitrary number of channels from an arbitrary number of systems. It is assumed that the CR can only sense one sub-band at a time. Three sub-band selection policies are proposed to find spectrum opportunities taking into account realistic hardware reconfiguration energy consumptions and time delays. Two of the proposed policies rely on the individual channel Markov properties and the sub-band Markov properties, respectively. Although these two policies may achieve good performance, they rely on complete knowledge of RF environment dynamics and thus may become computationally demanding. The third sub-band selection policy based on Q-learning is proposed to circumvent this. Performance of the three policies are compared and discussed against a performance upper-bound of the optimal solution to the corresponding partially observable Markov decision process formulation. The suitability of the Q-learning technique is validated by showing that it achieves good performance through numerical results in both simulated and real measured RF environments.

Index Terms—Bandwidth aggregation, cognitive radios, Markov decision processes, Partially observable Markov decision processes, Q-learning, sub-band selection, wide-band cognitive radios, wide-band spectrum sensing.

I. INTRODUCTION

The radio frequency (RF) spectrum is a limited resource regulated by government agencies. Conventional radios are designed to communicate within a specified RF spectrum range. Nowadays, the increasing demand for mobile wireless services, such as web browsing, video telephony, and video streaming, with various constraints on delay and bandwidth requirements, poses new challenges to be met by future generation wireless communication networks. On the other hand, it has been reported that the static RF spectrum allocation scheme has caused low efficiency of the spectrum utilizations. Unlike conventional radios, cognitive radios (CRs) [1]–[5] are proposed to achieve dynamic utilization of the limited RF spectrum resource and to settle the spectral under-utilization problem.

The National Broadband Plan (NBP) [6] recommends to free up 500 MHz of spectrum for broadband use in the next 10 years with 300 MHz being made available for mobile use in the next five years. The plan proposes to achieve this goal in a number of ways: incentive auctions, repacking spectrum, and

enabling innovative spectrum access models that take advantage of opportunistic spectrum access and cognitive techniques to better utilize the spectrum. The plan urges the FCC to initiate further developments on opportunistic spectrum access beyond the already completed TV white spaces proceedings. The Radiobot architecture proposed in [3], [7] is in-line with above vision and proposes CRs that are wide-band, multi-mode and multi-band. A Radiobot is a wide-band CR that would be able to optimally respond to its RF environment in order to achieve its performance objectives. However, these kind of wide-band CR capabilities do rely on both state-of-the-art RF hardware front-end (such as wide-band antennas, real-time reconfigurable antennas, etc.) and sophisticated signal processing techniques¹.

Spectrum sensing has been identified as a fundamental task for CRs to detect spectrum opportunities and achieve awareness of the surrounding RF environment [1], [4], [5], [8], [9]. Several sensing techniques have been proposed for sensing primary signals² in either narrow or wide frequency bands [8]–[12]. In narrowband applications, a CR senses a particular channel (or a particular set of channels) to identify the existence of primary signals. In this case, the decision-making reduces to a binary hypothesis testing problem to determine whether a particular channel is idle or busy [13]–[16]. In a wide-band CR application, however, in order to maximize its communication throughput, a CR not only has to determine the existence of primary signals, but it also has to determine the spectrum range to sense in the first place. This is due to the limitations of the RF hardware and the signal processing capabilities, which often prohibit a wide-band CR from sensing the whole spectrum range of interest at the same time and the spectrum usage patterns are in general non-homogeneous across the wide spectrum range of interest [17].

In this paper, we propose a dynamic spectrum sensing scheduling framework for wide-band CRs. The considered wide-band CR is assumed to be equipped with a reconfigurable RF front-end (reconfigurable antennas and reconfigurable RF circuitry) that may operate over several wide frequency bands, with each configuration corresponding to one of the wide frequency bands. Each of the wide frequency bands is assumed

¹The details on the software and hardware requirements for a Radiobot architecture are discussed in [3].

²A primary signal refers to a signal that is licensed to a certain frequency range by the regulations of static RF spectrum allocation.

to be further segmented into several non-overlapping sub-bands and each of the segmented sub-bands is assumed to contain multiple communication channels. Without loss of generality, we assume that the CR can only operate in one of the sub-bands at a time due to hardware and signal processing limitations. We also assume that the wide-band CR is capable of simultaneous transmissions of multiple signals on multiple channels within a single sub-band. Note that there may exist multiple distinguishable radio interfaces or communication protocols within any particular sub-band. The simultaneous transmission over multiple radio interfaces by a single mobile terminal has been previously discussed in the literature under the term of bandwidth aggregation (BAG) [18]–[20], which aims at performing simultaneous use of multiple interfaces to improve transmission quality or throughput depending on designs. In this work, however, the focus is on the sub-band selection problem that arise in wide-band spectrum sensing instead of the optimization of the BAG problem.

Note that many schemes presented in CR literature, such as in [13], [21], [22], have previously proposed and derived the channel sensing algorithms for narrow-band scenarios. As opposed to the wide-band spectrum sensing, in narrow-band spectrum sensing problems, hardware reconfigurations are generally not considered. For example, the authors in [13] developed an optimal myopic³ sensing scheduling policy in a centralized multi-agent setup for a group of traditional narrow-band CRs with a given set of channels. In [21], assuming that the channel state transition probabilities are partially known, the authors developed a myopic channel sensing strategy for the narrow-band CRs and showed that this myopic policy is the optimal Partially Observable Markov Decision Process (POMDP) solution under the assumption of a certain ordering of the state transition probabilities of individual channels. In [22], the authors developed stationary optimal spectrum sensing and access policies under the framework of POMDP to maximize the CR’s throughput on a given set of channels in a narrow-band setup with battery life constraints. However, these spectrum sensing policies cannot be easily applied in a wide-band spectrum sensing scenario since the reconfigurable RF front-end is not considered and the reconfiguration costs are not taken into account to jointly optimize the performance. As a result, in this paper, we propose the wide-band spectrum sensing scheduling policies with realistic reconfigurable RF front-end considerations. In [25], the authors investigated optimal sensing time and power allocation strategies in order to maximize the transmission throughput in a wide-band sensing setup. However, there is a fundamental difference between our system setup and the one in [25]. In particular, what is meant by ‘wide-band’ in our system is different from that of [25], and all similar previous work. In [25], wide-band sensing refers to simultaneous sensing of a frequency band containing multiple narrowband channels. The term wide-band is justified because the spectrum spanned by these channels can be larger compared to a single narrowband channel. However, the wide-

band system assumed in this work is conceivably much wider than that of [25]. In fact, the wide spectrum band considered in [25] is somewhat equivalent to a single sub-band assumed in our setup. In [25] and other similar previous work, the wide-band operation is limited by the RF front-end and the A/D circuits, whereas our wide-band CRs are presumed to be equipped with real-time reconfigurable RF front-ends covering a set of wide spectrum ranges in each mode of operation, and each of these spectrum ranges are divided into a set of sub-bands that are still wide and may contain multiple (narrowband) channels [3], [7], [26]. Clearly, given the state-of-the-art wide-band antenna/RF front-end designs [27]–[30], and the signal processing burdens, the wide-band assumption in those previous proposals can only imply something akin to one of the sub-bands assumed in our work. As a result, while spectrum sensing decisions in many of the previous proposals are concerned with channel selection, our focus is on the problem of which subset of channels (i.e. the sub-band) to sense.

Note that although the wide-band spectrum sensing scheduling problem may be formulated as a POMDP problem⁴ when the RF environment exhibit Markov properties, the optimal solution to the POMDP is *computationally prohibitive* because of the continuum of the state space, as also noted in [13], [21], [31]. As a result, three myopic sub-band selection policies are proposed in this paper to myopically maximize the probability of finding spectrum opportunities and communication throughput. The proposed policies take into account realistic reconfiguration energy consumptions and time delays. The first sub-band selection policy rely on the knowledge of the channel Markov properties. The second sub-band selection policy is proposed to rely on the Markov properties of the sub-bands to reduce the complexity. Note that, although both of these two policies may achieve good results, they rely on the knowledge of the Markov properties of the RF environment and thus may become computationally infeasible when the knowledge of the Markov models are unavailable. As a result, the third sensing policy based on the Q-learning [32] technique is proposed to avoid the necessity of any knowledge of the Markov properties.

The Q-learning algorithm is one of the most important *temporal difference* (TD) reinforcement learning (RL) methods and it has been shown to converge to the optimal policy when applied to single agent Markov decision process (MDP) models [32], [33]. The Q-learning has also been recently applied to CRs [34], [35]. Although the sub-band selection problem is a POMDP problem, we may still use the Q-learning technique to achieve reasonable performance results since it has been shown that the application of Q-learning in POMDP problems may achieve near-optimal solutions [14], [36], [37]. Performance of the three policies are compared and discussed against a performance upper-bound of the optimal solution to the POMDP formulation. We validate the suitability of the Q-learning technique for this type of wide-band spectrum sensing

³Myopic policies aims at maximizing an instantaneous reward at each time step, as opposed to a long-term reward as considered in a Partially Observable Markov Decision Process (POMDP) setup [23], [24]. The optimal myopic solution refers to the optimal solution within the class of myopic policies.

⁴The wide-band spectrum sensing scheduling problem can be formulated as a POMDP problem since at each time step, only the state of the sensed sub-band is revealed and the complete state of the RF environment is not fully observable.

problems by showing that it achieves good performance in both simulated and real measured RF environments.

The remainder of the paper is organized as follows: In Section II we introduce the system model and problem formulation. In Section III, the sub-band selection policies for spectrum sensing are developed. In Section IV, the alternative Q-learning based solution is proposed. In Section V we show the simulation results. In Section VI we conclude by summarizing our results.

II. SYSTEM MODEL AND PROBLEM FORMULATION

A. Spectrum Segmentation Model for Wide-band Sensing

The proposed CR architecture consists of a tunable RF front-end with wide-band capabilities and a cognitive engine (CE), as shown in Fig. 1. The CE is equipped with signal processing, autonomous learning and decision-making capabilities, as proposed in the Radiobot architecture in [3]. The CE controls the RF front-end to perform spectrum sensing and communication functionalities.

We assume that a reconfigurable antenna is adopted to cover R number of different frequency bands $\mathcal{W}_1, \dots, \mathcal{W}_R$ spanning a wide range of frequency spectrum. Note that the frequency bands $\mathcal{W}_1, \dots, \mathcal{W}_R$ are determined by the capabilities of each configuration of the antenna. We denote by $W_l = |\mathcal{W}_l| > 0$ the bandwidth of the frequency band \mathcal{W}_l , for $l \in \{1, \dots, R\}$. Since the bandwidths W_1, \dots, W_R are considered to be still wide, which may require further segmenting those frequency bands into smaller sub-bands prior to processing. Therefore, the sensing reconfigurable antenna will be connected to a reconfigurable band-pass filter or a filter bank of reconfigurable band-pass filters allowing proper segmentation of each of the frequency bands. We also assume that spectrum sensing can only be performed on a single sub-band at a time due to software and hardware limitations. There are several characteristics that need to be specified in order to determine the optimal number of sub-bands in each frequency band, such as the sampling rate of the ADC, the required quantization accuracy, and the power consumptions to name a few. However, we omit the problem of finding the optimal number of sub-bands in each frequency band due to the focus of this work. Without loss of generality, we may assume that there are N_l number of sub-bands in the l -th frequency band \mathcal{W}_l and denote by \mathcal{N}_l the set of sub-bands contained in the l -th RF configuration mode, with $|\mathcal{N}_l| = N_l$. An illustration of the frequency bands and the further segmented sub-bands is shown in Fig. 2. Note that the collection of the operable wide frequency bands may not perfectly cover the whole spectrum range due to antenna imperfections. The operable wide frequency bands may also overlap and/or be non-contiguous. Such reconfigurable antenna designs can be found in [27]–[30].

A spectrum sensing scheduling policy can be designed to dynamically change the RF front-end configurations to aim at suitable sub-bands to perform spectrum sensing. This sensing scheduling policy chooses a sub-band according to the real-time variations of the RF environment in order to maximize potential communication opportunities. We propose such a

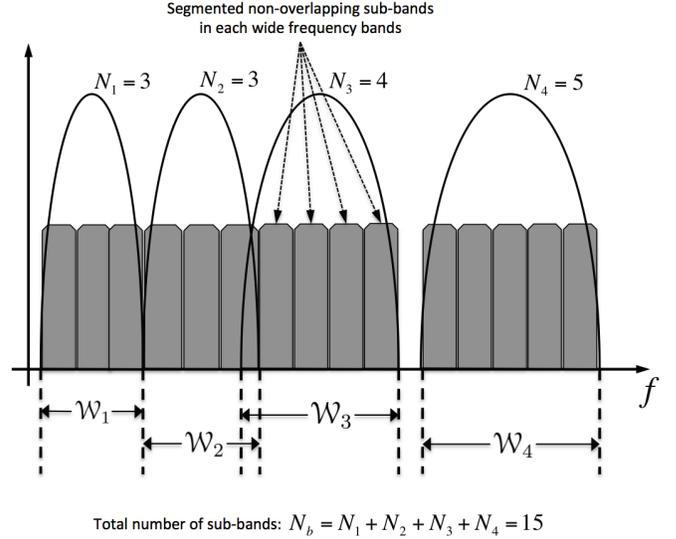


Fig. 2. An illustration of wide frequency bands and further segmented sub-bands in each wide frequency band.

sensing selection policy for the CR to perform spectrum sensing. We assume that the total bandwidth of interest is divided into $N_b = \sum_{l=1}^R N_l$ sub-bands and there are M_1, \dots, M_{N_b} number of identified communication channels in each of the N_b sub-bands respectively. In order to develop the proposed sub-band selection policies in Section III, we introduce the channel and sub-band Markov models in the rest of this section.

B. Channel Markov Model

We assume a semi-infinite slotted time horizon with each time slot having an equal time length of T sec. We denote by $k = \{0, 1, 2, \dots\}$ the time indices of the time slots. For simplicity, we assume that the state of a communication channel does not change within a single time slot, so that the CR may spend a short period of time at the beginning of each time slot to determine the corresponding state. We denote by $S_{i,j}(k) \in \{0, 1\}$ the true state of the (i, j) -th channel (the j -th channel in the i -th sub-band) at time k , for $j \in \{1, \dots, D_i\}$ and $i \in \{1, \dots, N_b\}$. As shown in Fig. 3, for a single channel, we may assume that the state *busy* (state 0) indicates the channel is occupied by other radio activities, and the state *idle* (state 1) indicates no radio activities over that channel and it is available for a CR to access. As a result, the state dynamics of each communication channel may be modeled as a two-state Markov chain. This Markov model, also known as the Gilbert-Elliot model [38], has been commonly used to abstract physical primary channels with memory (see, for example [13], [39]). Note that it is worth mentioning that the choice of the value T may play a critical role in terms of the validity of the channel Markov models. In particular, the channel Markov property may not hold for some choices of T , or the channel dynamics may be better represented by higher-order Markov models as opposed to the first order Markov model considered in this paper. However, due to the focus of this work, the problem of finding the appropriate value of T

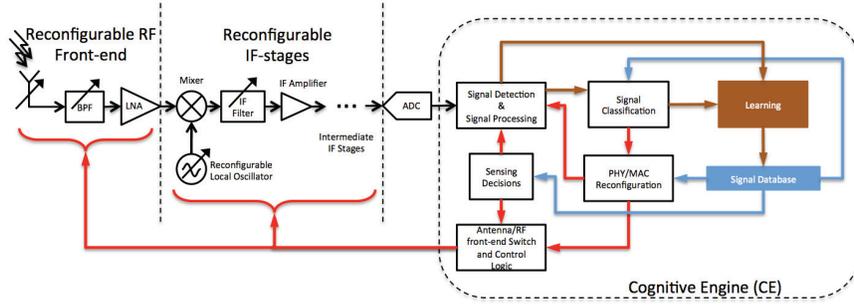


Fig. 1. System architecture for the proposed wide-band CR.

is not investigated. Detailed discussions on this topic can be found in [40] and the references therein.

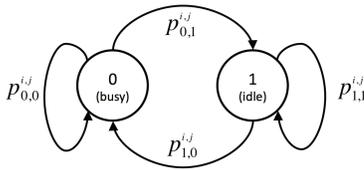


Fig. 3. The Markov chain model for a single communication channel (Gilbert-Elliot model).

Since different channels may exhibit non-identical statistical behaviors [17], the assigned Markov chain models are, in general, non-identical, i.e. the state transition probabilities and the stationary distributions are different. The time-invariant transition probability of the (i, j) -th channel Markov model from state x to state y is defined as $p_{x,y}^{i,j} = \Pr\{S_{i,j}(k+1) = y \mid S_{i,j}(k) = x\}$, $\forall x, y \in \{0, 1\}$. The *transition probability matrix* of the (i, j) -th channel Markov model is denoted by $\mathbf{P}^{i,j} = \begin{pmatrix} p_{0,0}^{i,j} & p_{0,1}^{i,j} \\ p_{1,0}^{i,j} & p_{1,1}^{i,j} \end{pmatrix}$. We denote by vector $\boldsymbol{\pi}^{i,j} = [\pi_0^{i,j}, \pi_1^{i,j}]$ the stationary distribution vector, such that $\boldsymbol{\pi}^{i,j} = \boldsymbol{\pi}^{i,j} \mathbf{P}^{i,j}$, with $\pi_0^{i,j}$ and $\pi_1^{i,j}$ being the stationary probabilities of busy and idle, respectively.

C. Sub-band Markov Models

We may further define the random variable $N_i^{idle}(k)$ as the number of idle channels in the i -th sub-band at time k . Note that, due to the Markov property of the communication channels, the dynamic of $N_i^{idle}(k)$ also forms a Markov chain as shown in Fig. 4. Since there are D_i number of channels in

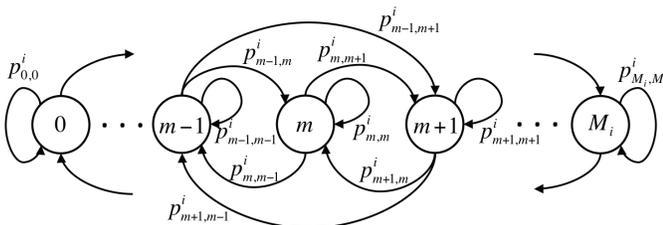


Fig. 4. The Markov model of the i -th sub-band. The state of the Markov model is defined as the number of idle channels in the i -th sub-band.

the i -th sub-band, we obtain a $(D_i + 1)$ -state Markov chain for the i -th sub-band, with a state space of $\{0, 1, \dots, D_i\}$. As shown in Fig. 4, the time-invariant transition probability of the Markov model from state m to state n is defined as

$$p_{m,n}^i = \Pr\{N_i^{idle}(k+1) = n \mid N_i^{idle}(k) = m\}, \quad \forall m, n \in \{0, 1, \dots, D_i\}. \quad (1)$$

The $(D_i + 1) \times (D_i + 1)$ *transition probability matrix* of the i -th sub-band is then denoted by $\mathbf{P}_i = \begin{pmatrix} p_{0,0}^i & p_{0,1}^i & \dots & p_{0,D_i}^i \\ p_{1,0}^i & p_{1,1}^i & \dots & p_{1,D_i}^i \\ \vdots & \vdots & \ddots & \vdots \\ p_{D_i,0}^i & p_{D_i,1}^i & \dots & p_{D_i,D_i}^i \end{pmatrix}$. We denote by vector $\boldsymbol{\pi}_i = [\pi_0^i, \dots, \pi_{D_i}^i]$ the stationary distribution vector, such that $\boldsymbol{\pi}_i = \boldsymbol{\pi}_i \mathbf{P}_i$ with $\pi_0^i, \dots, \pi_{D_i}^i$ being the stationary probabilities of the states $0, 1, \dots, D_i$, respectively.

III. SUB-BAND SELECTION IN WIDE-BAND SPECTRUM SENSING

A. Spectrum Sensing Detector Characteristics

Under the assumption that the CR has no knowledge of the signaling on the communication channels, we adopt an energy detection based detector for spectrum sensing [13]. Since the optimality criterion is to maximize the probability of detection of idle channel under the constraint of the collision probability (claiming a channel idle when it is actually busy leads to a collision), we develop an energy-based Neyman-Pearson (NP) detector [13]. Note that although Matched-filter based or cyclostationarity-based NP detectors may be adopted under different assumptions on the knowledge of the channel signaling, in this paper we adopt the energy-based NP detector for illustration purpose.

Within the k -th time step, for the i -th sub-band, we consider a sampled data sequence $\{y(t, k, i)\}_{t=0}^{U_i-1}$, with data length of U_i , and T_s as the sampling period. As a result, the sensing duration for the i -th sub-band is $T_0^i = U_i T_s$. We denote by $\{Y(n, k, i)\}_{n=0}^{U_i-1}$ its discrete Fourier transform (DFT):

$$Y(n, k, i) = \sum_{t=0}^{U_i-1} y(t, k, i) e^{-j2\pi n t / U_i}, \quad \text{for } n = 0, \dots, U_i - 1. \quad (2)$$

In order to detect the state of each and every channel, we find the average power in a spectral window of odd length $L_{i,j}$, centered at $f_c^{i,j}$, which can be approximated by

$T(f_c^{i,j}, k, i) = \sum_{l=-L_{i,j-1}/2}^{(L_{i,j-1})/2} |Y(n^{i,j} + l, k, i)|^2$. Note that $n^{i,j}$ is the discrete frequency point corresponding to $f_c^{i,j}$. The collection of non-overlapping spectral windows then represent the channels within the target sub-band. In order to derive an NP detector, we determine the distribution of $T(f_c^{i,j}, k, i)$ under the two hypotheses:

$$\mathcal{H}_0 : y_{i,j}(t, k) = s_{i,j}(t, k) + w_{i,j}(t, k), \quad (3)$$

$$\mathcal{H}_1 : y_{i,j}(t, k) = w_{i,j}(t, k), \quad (4)$$

where we denote by $y_{i,j}(t, k)$, $s_{i,j}(t, k)$, and $w_{i,j}(t, k)$ the assumed receiver samples, the received signal samples, and the noise samples, corresponding to the (i, j) -th channel, respectively. Note that $\{w_{i,j}(t, k)\}_{t=0}^{U_i-1}$ are modeled as i.i.d. Gaussian random variables, s.t. $w_{i,j}(t, k) \sim \mathcal{N}(0, P_w)$. The signal $\{s_{i,j}(t, k)\}_{t=0}^{U_i-1}$ in (3) can be modeled as i.i.d. Gaussian random variables, s.t. $s_{i,j}(t, k) \sim \mathcal{N}(0, P_s)$. This is a reasonable assumption for signals that are perturbed by propagation through turbulent media and multipath fading [41]. In the following, we drop the time step and sub-band/channel indices for notational simplicity: we let $y(t) = y_{i,j}(t, k)$, $Y(n) = Y(n, k, i)$, $T(f_c) = T(f_c^{i,j}, k, i)$, $n = n^{i,j}$, $L = L_{i,j}$, and $U = U_i$. We denote by $\mathbf{y} = [y(0), \dots, y(U-1)]^T$, $\mathbf{Y} = [Y(n - \frac{L-1}{2}), \dots, Y(n + \frac{L-1}{2})]$, $\mathbf{Y}^R = \Re\{\mathbf{Y}\}$ and $\mathbf{Y}^I = \Im\{\mathbf{Y}\}$, where $\Re\{\}$ and $\Im\{\}$ denote the real and imaginary parts, respectively. The DFT operation can be equivalently expressed as: $\mathbf{Y}^C \triangleq [\mathbf{Y}^R \ \mathbf{Y}^I]^T = \mathbf{A}\mathbf{y}$, where \mathbf{A} is a $2L$ -by- U matrix of DFT coefficients. Since the time domain samples $\{y(t)\}_{t=0}^{U-1}$ are zero-mean i.i.d. Gaussian random variables, then \mathbf{Y}^C is also a jointly Gaussian random vector. It can be shown that, under \mathcal{H}_0 , $\mathbb{E}\{\mathbf{Y}^C (\mathbf{Y}^C)^T\} = L(P_s + P_w)\mathbf{I}_{2L}$ (where \mathbf{I}_{2L} is an $2L$ -by- $2L$ identity matrix) and under \mathcal{H}_1 , $\mathbb{E}\{\mathbf{Y}^C (\mathbf{Y}^C)^T\} = LP_w\mathbf{I}_{2L}$. Therefore, elements of \mathbf{Y}^C are uncorrelated. Since \mathbf{Y}^C is jointly Gaussian with uncorrelated elements, the elements of \mathbf{Y}^C are then independent. Also, since all the elements have the same variance under each of the hypotheses, elements of \mathbf{Y}^C are assumed to be i.i.d. zero-mean Gaussian random variables with variance $L(P_w + P_s)$ under \mathcal{H}_0 , and LP_w under \mathcal{H}_1 . Under the above assumptions, $T'(f_c) \triangleq \frac{1}{L(P_w + P_s)}T(f_c)$ is a sufficient statistic for the hypothesis testing and follows a χ_{2L}^2 distribution. The threshold η for idle channel detection is defined s.t. $\Pr\{T'(f_c) < \eta | \mathcal{H}_0\} \leq \alpha_F$, where α_F is the acceptable false alarm probability, or in our case, the collision probability with the undetected signal activities on the channel. Note that the signal and noise power can be estimated, for example, by using the method proposed in [42]. The resulting threshold can then be found as $\eta = 2\gamma^{-1}(L; \alpha_F \Gamma(L))$ from the cumulative distribution function (cdf) of the χ_{2L}^2 distribution, where γ^{-1} is the inverse lower incomplete gamma function (where $\gamma(k; x) = \int_0^x t^{k-1} e^{-t} dt$ and the inverse is w.r.t. the second argument) and $\Gamma(k) = \int_0^\infty t^{k-1} e^{-t} dt$ is the gamma function. The NP decision rule δ for idle state detection of channel centered at f_c is then defined as:

$$\delta(T'(f_c)) = \begin{cases} 1 & \text{if } T'(f_c) < \eta \\ 0 & \text{otherwise} \end{cases}, \quad (5)$$

where the decision 1 stands for claiming a channel as

in idle state (state '1'), and the decision 0 stands for claiming a channel as in busy state (state '0'). The detection probability (detecting idle channel) is $P_D = \Pr\{T'(f_c) < \eta | \mathcal{H}_1\}$, which can also be computed as $P_D = \Pr\{(1 + \text{SNR})T'(f_c) < (1 + \text{SNR})\eta | \mathcal{H}_1\}$, where we denote by $\text{SNR} = P_s/P_w$ the signal-to-noise ratio. Since $(1 + \text{SNR})T'(f_c) = \frac{1}{LP_w}T(f_c)$ is χ_{2L}^2 distributed under \mathcal{H}_1 , the detection probability can be found as $P_D = \frac{1}{\Gamma(L)}\gamma\left(L; \frac{(1 + \text{SNR})\eta}{2}\right)$. Appending back the time step and sub-band indices, we obtain the probability of detection of idle channel, at the k -th time step for the i -th sub-band, as $P_D(k, i) = \frac{1}{\Gamma(L)}\gamma\left(L; \frac{(1 + \text{SNR}(k, i))\eta(i)}{2}\right)$, where the threshold $\eta(i) = 2\gamma^{-1}(L; \alpha_F(i)\Gamma(L))$, with the acceptable false alarm probability of $\alpha_F(i)$ in the i -th sub-band.

Note that since different frequency bands may have different spectrum sensing requirements. For instance, in some licensed frequency bands, there can be a more stringent regulation of collisions with licensed users such that the upper bound on the probability of collision is low. This requires a CR to spend more time on spectrum sensing in order to achieve the required level of probability of detecting idle channels. On the other hand, in an unlicensed frequency band, such as the Industrial, Scientific and Medical (ISM) band, the collision is not often strictly controlled. Hence, a CR may spend less time to detect a transmission opportunity at the expense of a possible higher collision probability. This can be easily understood by examining the expression of $P_D(k, i)$. In particular, since $P_D(k, i) = \frac{1}{\Gamma(L)}\gamma\left(L; (1 + \text{SNR}(k, i))\gamma^{-1}(L; \alpha_F(i)\Gamma(L))\right)$, it is straightforward to confirm that $P_D(k, i)$ decreases as $\alpha_F(i)$ decreases, and $P_D(k, i)$ increases as L increases. As a result, for a lower value of $\alpha_F(i)$, $P_D(k, i)$ is decreased. However, one may increase $P_D(k, i)$ back to the desired level by increasing L . One effective way to increase L is to increase the sensing duration (or increasing the number of samples D_i under the same sampling rate) in order to obtain a higher resolution of the DFT in the frequency domain, since it would need a larger L to cover the bandwidth of a channel with higher frequency resolution.

B. Channel State Belief Update

We may denote by $\hat{S}_{i,j}(k)$ the outcome of the NP detector at time k , as the estimate of the state of the (i, j) -th channel. In particular, we have $\hat{S}_{i,j}(k) \triangleq \delta(T'(f_c^{i,j}))$. We may then define the channel state belief $b_s^{i,j}(k)$ for the (i, j) -th channel as the probability of the channel in state $s \in \{0, 1\}$ at time k , given the observation history on that particular channel. In particular, we define $b_s^{i,j}(k) \triangleq \Pr\{S_{i,j}(k) = s | \hat{\mathbf{S}}_{i,j}^{0:k-1}\}$, for $s \in \{0, 1\}$, where we denote by $\hat{\mathbf{S}}_{i,j}^{0:k-1} = [\hat{S}_{i,j}(0), \dots, \hat{S}_{i,j}(k-1)]^T$ the channel state detection history from time step 0 to time step $k-1$. When the channel state detection result $\hat{S}_{i,j}(k-1)$ is obtained, the channel state belief $b_s^{i,j}(k)$ can be found iteratively as in (6), where we have the following from the NP detector

$$\begin{aligned}
b_s^{i,j}(k) &= \Pr\{S_{i,j}(k) = s \mid \hat{S}_{i,j}^{0:k-1}\} = \frac{\sum_{s' \in \{0,1\}} p_{s',s}^{i,j} \Pr\{\hat{S}_{i,j}(k-1) | S_{i,j}(k-1) = s'\} \Pr\{S_{i,j}(k-1) = s' | \hat{S}_{i,j}^{0:k-2}\}}{\sum_{s' \in \{0,1\}} \Pr\{\hat{S}_{i,j}(k-1) | S_{i,j}(k-1) = s'\} \Pr\{S_{i,j}(k-1) = s' | \hat{S}_{i,j}^{0:k-2}\}} \\
&= \frac{\sum_{s' \in \{0,1\}} p_{s',s}^{i,j} \Pr\{\hat{S}_{i,j}(k-1) | S_{i,j}(k-1) = s'\} b_s^{i,j}(k-1)}{\sum_{s' \in \{0,1\}} \Pr\{\hat{S}_{i,j}(k-1) | S_{i,j}(k-1) = s'\} b_s^{i,j}(k-1)}, \text{ for } s \in \{0,1\}.
\end{aligned} \tag{6}$$

characteristics:

$$\begin{cases} \Pr\{\hat{S}_{i,j}(k-1) = 1 | S_{i,j}(k-1) = 1\} = P_D(k-1, i) \\ \Pr\{\hat{S}_{i,j}(k-1) = 0 | S_{i,j}(k-1) = 1\} = 1 - P_D(k-1, i) \\ \Pr\{\hat{S}_{i,j}(k-1) = 1 | S_{i,j}(k-1) = 0\} = \alpha_F(i) \\ \Pr\{\hat{S}_{i,j}(k-1) = 0 | S_{i,j}(k-1) = 0\} = 1 - \alpha_F(i) \end{cases} . \tag{7}$$

Note that, however, to obtain $b_s^{i,j}(k)$, for $k \in \{1, 2, \dots\}$ using (6), it requires that the (i, j) -th channel is sensed at time $k-1$. When this is not the case, we use the Markovian property to update the channel belief. In particular, we have $[b_0^{i,j}(k) \ b_1^{i,j}(k)] = [b_0^{i,j}(k-1) \ b_1^{i,j}(k-1)] \mathbf{P}^{i,j}$, where $\mathbf{P}^{i,j}$ is the transition probability matrix of the (i, j) -th channel.

We denote by $T_{i,j}(k) \in \{0, T, 2T, \dots\}$ the discrete-valued random variable of the idle sojourn time of the (i, j) -th channel starting from time k . The idle sojourn time refers to the time duration of the channel being consecutively idle. Since we assumed that the state of any communication channel does not change within a single time slot, the sojourn time of a channel is discrete-valued. The probability mass function (pmf) of $T_{i,j}(k)$ can be found as in (8) by using the Markov properties and the channel belief, where we denote by $(p_{1,1}^{i,j})^{n-1}$ the $(n-1)$ -th power of $p_{1,1}^{i,j}$. The expected value of $T_{i,j}(k)$ can then be found as

$$\mathbb{E}\{T_{i,j}(k)\} = \sum_{n=0}^{\infty} f_{T_{i,j},k}(nT) \cdot nT. \tag{9}$$

C. Sub-band Selection Policy Based On The Channel Markov Models

In order to derive the sensing sub-band selection policy, let us first denote by $BW_{i,j}$ the identified channel bandwidth of the j -th channel in the i -th sub-band, for $i \in \{1, \dots, N_b\}$ and $j \in \{1, \dots, D_i\}$. Note that the instantaneous transmission rate of a channel with a bandwidth of B is $r = B \log_2 \left(1 + \frac{h^2 P}{BN_0}\right)$ bits/sec, where we denote by h , P , and N_0 the channel coefficient between the receiver and the transmitter, the transmission power, and the single-sided noise power spectrum density (PSD) level, respectively. We assume that the distributions of the channel coefficients are either known or can be obtained through pilot signal learning within the CR devices. We denote by $f_{H_{i,j}}$ the corresponding distribution function of the channel coefficient of the (i, j) -th channel.

In order to take into account the practical RF front-end reconfigurable energy consumptions in the sub-band selection decision-making, we may denote by $c_s(i', i)$ the switching energy cost from the i' -th sub-band to the i -th sub-band, such that

$$c_s(i', i) = \begin{cases} c_1 + c(T_0^i), & \text{if } i' \in \mathcal{N}_{l'}, i \in \mathcal{N}_l, \text{ and } l' \neq l \\ c_2 + c(T_0^i), & \text{if } i' \in \mathcal{N}_{l'}, i \in \mathcal{N}_{l'}, \text{ and } i' \neq i \\ c(T_0^i), & \text{if } i' = i \end{cases}, \tag{10}$$

where c_1 denotes the energy cost when switching between different RF configuration modes, and c_2 denotes the energy cost when switching between different sub-bands within the same RF configuration mode. The quantity $c(T_0^i)$ denotes the energy cost required for spectrum sensing in the i -th sub-band, as a function of the required sensing time T_0^i . Since hardware reconfiguration may require more energy consumption, we assume that $c_1 > c_2$. Note that in practice c_1 and c_2 may not necessarily be constant. In such cases, we may easily re-adjust them depending on the specific adopted RF front-end.

We may also define $t_s(i', i)$ the switching time delay incurred when the CR switches from the i -th sub-band to the i' -th sub-band, such that

$$t_s(i', i) = \begin{cases} t_1, & \text{if } i' \in \mathcal{N}_{l'}, i \in \mathcal{N}_l, \text{ and } l' \neq l \\ t_2, & \text{if } i' \in \mathcal{N}_{l'}, i \in \mathcal{N}_{l'}, \text{ and } i' \neq i \\ t_3, & \text{if } i' = i \end{cases}, \tag{11}$$

where t_1, t_2 and t_3 include the computation time of decision-making at each time step, the circuit switching time, software reconfiguration time, and settling time for the RF front-end (especially the settling time for the phase-locked loop (PLL) in the frequency synthesizer [43]).

In order to consider the bandwidth aggregation, we may assume that the CR is capable of utilizing up to a maximum of G idle channels simultaneously, all from a single sub-band. When the CR has the knowledge of channel Markov models but not the Markov models of the sub-bands, we may define the total expected communication throughput by switching from the i' -th sub-band to the i -th sub-band in time slot k as

$$\begin{aligned}
R_{i'}(i, k) &= \sum_{j \in \mathcal{M}_{i,G}^*} \mathbb{E}_{H_{i,j}} \{r_{i,j}\} \times \\
&\times \min \left\{ \left[\mathbb{E}\{T_{i,j}(k)\} \left(1 - \frac{T_0^i}{T}\right) - t_s(i', i) \right], T_{max} \right\}, \tag{12}
\end{aligned}$$

where function $\min\{x, y\} = x$, if $x \leq y$ and $\min(x, y) = y$ otherwise. Note that the expectation of transmission rate $\mathbb{E}_{H_{i,j}} \{r_{i,j}\}$ on the (i, j) -th channel in (12) is with respect to the channel coefficient and is defined as $\mathbb{E}_{H_{i,j}} \{r_{i,j}\} = \int BW_{i,j} \log_2 \left(1 + \frac{h^2 P}{BW_{i,j} N_0}\right) \times f_{H_{i,j}}(h) dh$. The expression $\left[\mathbb{E}\{T_{i,j}(k)\} \left(1 - \frac{T_0^i}{T}\right) - t_s(i', i) \right]$ in (12) gives the expected transmission time on the (i, j) -th channel. We denote by T_{max} the maximum considered staying time for any sub-band. The T_{max} is introduced to prevent the CR from selecting a sub-band when the achievable transmission rate in a sub-band is extremely low, but the expected channel idle sojourn time is extremely large. In this case, although the expected throughput may be large, the extremely low transmission rate may not be desirable. We denote by $\mathcal{M}_{i,G}^*$ in (12) the set of G channels in the i -th sub-band that have top G highest expected transmission throughput.

$$\begin{aligned}
f_{T_{i,j},k}(nT) &= \Pr\{T_{i,j}(k) = nT \mid \hat{\mathbf{S}}_{i,j}^{0:k-1}\} = \frac{\sum_{s \in \{0,1\}} \Pr\{T_{i,j}(k) = nT, \hat{\mathbf{S}}_{i,j}^{0:k-1} | S_{i,j}(k) = s\} \Pr\{S_{i,j}(k) = s\}}{\Pr\{\hat{\mathbf{S}}_{i,j}^{0:k-1}\}} \\
&= \frac{\sum_{s \in \{0,1\}} \Pr\{T_{i,j}(k) = nT | S_{i,j}(k) = s\} \Pr\{\hat{\mathbf{S}}_{i,j}^{0:k-1}, S_{i,j}(k) = s\}}{\Pr\{\hat{\mathbf{S}}_{i,j}^{0:k-1}\}} \\
&= \sum_{s \in \{0,1\}} \Pr\{T_{i,j}(k) = nT | S_{i,j}(k) = s\} \Pr\{S_{i,j}(k) = s | \hat{\mathbf{S}}_{i,j}^{0:k-1}\} = \begin{cases} b_0^{i,j}(k), & \text{if } n = 0 \\ b_1^{i,j}(k) \cdot (p_{1,1}^{i,j})^{n-1} \cdot p_{1,0}^{i,j}, & \text{if } n \in \{1, 2, 3, \dots\} \end{cases} \quad (8)
\end{aligned}$$

When the CR has only the knowledge of channel Markov models, by taking the switching energy and time delays into account, we may then define the quality of the i -th sub-band (switching from the i' -th sub-band) at time k as

$$Q_{i'}(i, k) = R_{i'}(i, k) - \beta c_s(i', i), \quad (13)$$

where the coefficient β (bits/Joule) is used to convert the units and to help weighting the energy consumption priority. The sub-band selection policy $a(i', k)$ (in the i' -th sub-band and time slot k) may then be defined as

$$a(i', k) = \arg \max_{i \in \{1, \dots, N_b\}} Q_{i'}(i, k). \quad (14)$$

D. Sub-band Selection Policy based on the Sub-band Markov Models

In the case when the knowledge of both the sub-band Markov models and the channel Models are available, we may define the total expected communication throughput by switching from the i' -th sub-band to the i -th sub-band in time slot k as in (15) where we denote by $\hat{N}_i^{idle}(k)$ the estimate of the number of idle channels in the i -th sub-band at time k . The term $\min\{\hat{N}_i^{idle}(k), G\}$ in (15) is the estimated number of accessible and usable channels at time k . The estimate of $\hat{N}_i^{idle}(k)$ may be obtained, for example, using the following two criteria:

- 1) The maximum a posteriori (MAP) criterion: $\hat{N}_i^{idle}(k) = \arg \max_{n \in \{0, \dots, D_i\}} \Pr\{N_i^{idle}(k) = n \mid \mathbf{b}_0^i(k)\}$, where we denote by $\mathbf{b}_0^i(k) = [b_0^{i,1}(k), \dots, b_0^{i,D_i}(k)]^T$ the belief vector. The probability $\Pr\{N_i^{idle}(k) = n \mid \mathbf{b}_0^i(k)\}$ is found as

$$\begin{aligned}
&\Pr\{N_i^{idle}(k) = n \mid \mathbf{b}_0^i(k)\} \\
&= \Pr\left\{\sum_{j=1}^{D_i} S_{i,j}(k) = n \mid \mathbf{b}_0^i(k)\right\} \\
&= \sum_{\mathcal{A}_{i,n}} \left[\left(\prod_{j \in \mathcal{A}_{i,n}} b_0^{i,j}(k) \right) \left(\prod_{j \in \mathcal{A}'_{i,n}} (1 - b_0^{i,j}(k)) \right) \right],
\end{aligned}$$

where we denote by $\mathcal{A}_{i,n}$ a subset of channels in the i -th sub-band, with cardinality n and we denote by $\mathcal{A}'_{i,n}$ the relative complement of $\mathcal{A}_{i,n}$, with respect to the set of all channels in the i -th sub-band. Note that the summation is over all possible $\mathcal{A}_{i,n}$'s.

- 2) The minimum mean square error (MMSE) criterion: $\hat{N}_i^{idle}(k) = \mathbb{E}\{N_i^{idle}(k) \mid \mathbf{b}_0^i(k)\} = \sum_{n=1}^{D_i} n \Pr\{N_i^{idle}(k) = n \mid \mathbf{b}_0^i(k)\}$.

Note that although the MMSE estimator may give a non-integer result for $\hat{N}_i^{idle}(k)$, it would still make sense when we use $\hat{N}_i^{idle}(k)$ to obtain the expected sub-band communication throughput. We verify in simulations that both methods achieve close results and thus we choose to use the MAP criterion since its computation is straightforward.

Note that, the average expected channel throughput per channel within the i -th sub-band in (15) requires the knowledge of the individual channel Markov models in order to obtain $\mathbb{E}\{T_{i,j}(k)\}$. This, of course, is not possible when the channel Markov parameters are unavailable. However, when only the sub-band Markov model is assumed to be known, we may replace the average expected channel throughput term by $\bar{r}_i \min\left\{\left[\bar{T}_i \left(1 - \frac{T_i}{T}\right) - t_s(i', i)\right], T_{max}\right\}$, where we denote by \bar{r}_i and \bar{T}_i the average achievable individual channel throughput and the average idle sojourn time of the channels in the i -th sub-band. Note that the average individual channel throughput and the average channel idle sojourn time may be easily summarized from past channel access history. However, due to the space limitation, we do not go into details of estimation methods for \bar{r}_i and \bar{T}_i . Note that the function $\min\{x, y\} = x$, if $x \leq y$, and $\min\{x, y\} = y$ otherwise.

The quality of the i -th sub-band may then be defined as:

$$Q_{i'}(i, k, \hat{N}_i^{idle}(k)) = R_{i'}(i, k, \hat{N}_i^{idle}(k)) - \beta c_s(i', i). \quad (16)$$

The sub-band selection policy $a(i', k)$ (in i' -th sub-band and time slot k) is defined as

$$a(i', k) = \arg \max_{i \in \{1, \dots, N_b\}} Q_{i'}(i, k, \hat{N}_i^{idle}(k)). \quad (17)$$

When the knowledge of the sub-band Markov models is not directly available, but the knowledge of the channel Markov models is available, one may obtain the knowledge of the sub-band Markov models from the knowledge of the channel Markov models, at least in theory (However, note that this is extremely unlikely when channels are non-i.i.d.). Note that the time-invariant transition probability $p_{m,n}^i$ of the i -th sub-band may be expressed as in (18), for all $m \in \{0, \dots, D_i\}$ and $n \in \{0, \dots, D_i\}$, where we denote by $\mathcal{A}_{i,m}$ a subset of channels in the i -th sub-band, with cardinality m and we denote by $\mathcal{A}'_{i,m}$ the relative complement of $\mathcal{A}_{i,m}$ with respect to the set of all channels in the i -th sub-band. The summation in (18) is taken over all possible $\mathcal{A}_{i,m}$'s and all possible combination of states $s_{i,1}, \dots, s_{i,D_i}$, where $s_{i,j} \in \{0, 1\}$ for all $j \in \{1, \dots, D_i\}$, such that $\sum_{j=1}^{D_i} s_{i,j} = n$.

$$R_{i'}(i, k, \hat{N}_i^{idle}(k)) = \frac{1}{D_i} \left(\sum_{j=1}^{D_i} \mathbb{E}_{H_{i,j}} \{r_{i,j}\} \min \left\{ \left[\mathbb{E}\{T_{i,j}(k)\} \left(1 - \frac{T_0^i}{T}\right) - t_s(i', i) \right], T_{max} \right\} \right) \min \{ \hat{N}_i^{idle}(k), G \}, \quad (15)$$

$$p_{m,n}^i = \Pr \left\{ \sum_{j=1}^{D_i} S_{i,j}(k+1) = n \mid \sum_{j=1}^{D_i} S_{i,j}(k) = m \right\} = \sum_{\mathcal{A}_{i,m}} \sum_{\{s_{i,j}\} \text{ s.t. } \sum_{j=1}^{D_i} s_{i,j} = n} \left[\left(\prod_{j \in \mathcal{A}_{i,m}} P_{1,s_{i,j}}^{i,j} \right) \left(\prod_{j \in \mathcal{A}'_{i,m}} P_{0,s_{i,j}}^{i,j} \right) \right]. \quad (18)$$

On the other hand, the stationary probability π_m^i , for $m \in \{0, \dots, D_i\}$ can be expressed as

$$\pi_m^i = \Pr \left\{ \sum_{j=1}^{D_i} S_{i,j}(k) = m \right\}. \quad (19)$$

In case the channels are independent, the stationary distribution can be further expressed as

$$\pi_m^i = \sum_{\mathcal{A}_{i,m}} \left(\prod_{j \in \mathcal{A}_{i,m}} \pi_1^{i,j} \right) \left(\prod_{j \in \mathcal{A}'_{i,m}} \pi_0^{i,j} \right). \quad (20)$$

The summation in (20) is taken over all possible $\mathcal{A}_{i,m}$'s. The closed-form expression of (20) requires a computational complexity of $\left[\binom{D_i}{m} \times D_i \right] - 1$ [17] when channels are assumed non-identical, but independent. In case the channels in a sub-band are non-identical, but statistically independent, we may also approximate the stationary distributions using the Poisson-Normal approximation method that is proposed in [17]. In this paper, however, since we assume that channels may be correlated (i.e. non-independent) in general, to obtain the closed-form expression of π_m^i requires the knowledge of the joint distribution of $S_{i,j}(k)$'s, which is even harder to be obtained. We can see that the computational complexity to obtain the transition probabilities is at least $\left[\binom{D_i}{m} \binom{D_i}{n} \times D_i \right] - 1$, which is even higher than that of obtaining the stationary distributions. This observation suggests that to obtain the knowledge of the sub-band Markov models from the channel Markov models may not be advisable.

As an alternative, we may adopt the hidden Markov model (HMM)-based parameter estimation algorithm proposed in [13] to perform on-line estimation of the transition probabilities of the Markov chain model, without the computation of (18). The estimation algorithm has been shown to have a computation complexity linear in the number of the states of the Markov chain, or in this case, D_i , the number of channels in the i -th sub-band. However, when the number of sub-bands and the number of channels in each sub-band are both large, the overall computational complexity is still high. Moreover, to obtain accurate estimates of the transition probability matrices, it may require a long period of time. As a result, in the case when the sub-band transition probabilities are unknown but the channel Markov models are known, we suggest to use the channel Markov models based sub-band selection policy defined in (14). When both the knowledge of the channel Markov models and the sub-band Markov models are available, one may choose either (12) or (15)

to express the expected sub-band throughput. We compare the resulting performances between these two strategies in simulations later in Section V. In the case when both channel and sub-band Markov models are unknown, we propose a Q-learning based Machine learning technique in Section IV to bypass the computation complexity.

IV. MACHINE LEARNING AIDED SUB-BAND SELECTION

In the case when neither the channels' nor the sub-bands' Markov models are known, we may rely on Reinforcement Learning (RL) techniques [32]. A Q-table $Q(s, a)$ is maintained that is used to summarize the value (benefit) of each action a in each and every state s . In our case, the action a refers to the selection of a sub-band, with $a \in \mathcal{N}_1 \cup \mathcal{N}_2 \dots \cup \mathcal{N}_R$. Each time an action is chosen in a certain state, the Q-table may be updated using the following rule:

$$Q(s_{k-1}, a_{k-1}) \leftarrow (1 - \alpha)Q(s_{k-1}, a_{k-1}) + \alpha \left[r_k(s_{k-1}, a_{k-1}) + \gamma \max_a Q(s_k, a) \right], \quad (21)$$

where we denote by s_{k-1} and a_{k-1} the observed state and the action in time interval $k-1$, respectively. Note that the state s_k does not refer to the state of the whole RF environment. This is explained in the following. The action a_{k-1} denotes the index of the sub-band selected that is to be sensed at time k . We denote by $\alpha \in (0, 1)$ the learning rate. The function $r_k(s_{k-1}, a_{k-1})$ denotes the reward obtained at time k , as a result of the action a_{k-1} in state s_{k-1} , which can be defined as the actual achieved performance. In the simulation, the reward is calculated as

$$r_k(s_{k-1}, a_{k-1}) = \tilde{r}(s_{k-1}, a_{k-1}) - \beta c_s(a_{k-2}, a_{k-1}) \quad (22)$$

where we denote by $\tilde{r}(s_{k-1}, a_{k-1})$ the actual achieved communication throughput by taking action a_{k-1} in state s_{k-1} . The term $c_s(a_{k-2}, a_{k-1})$ in (22) is the switching energy cost from the a_{k-2} -th sub-band to a_{k-1} -th sub-band as defined in (10), and β is the same coefficient as in (13). We denote by γ the discount factor, with $\gamma \in [0, 1)$. Note that the state at time $k-1$ may be defined as $s_{k-1} = [a(k-2), \hat{N}_{a(k-2)}^{idle}(k-1)]$, where $a(k-2)$ denotes the index of the sensed sub-band in time interval $k-1$. Also note that, the state s_k in (21) is the result of taking action a_{k-1} in state s_{k-1} and the term $\gamma \max_a Q(s_k, a)$ represents the discounted delayed reward by taking action a_{k-1} in state s_{k-1} . The value of $\max_a Q(s_k, a)$ is obtained by finding the maximum value in the row of the Q-table corresponding to the state s_k . The decision-making rule for choosing an action a^* in the state s may be defined

as $a^* = \arg \max_a Q(s, a)$. Since the state of the whole RF environment is not obtained at each time due to the RF hardware limitation (sensing can be done only in one sub-band at a time), the Q-learning application is for the POMDP case as discussed in the introduction section.

Note that the Q-learning is usually implemented as a balance between exploration and exploitation. Exploration refers to the effort of searching new opportunities, whereas exploitation refers to taking actions for immediate reward. Maintaining a certain level of exploration may help the agent avoid being trapped in local maxima. An exploration rate $\epsilon \in (0, 1)$ is often defined, such that the agent each time takes an action using a^* with probability $1 - \epsilon$ and uniformly choose an action out of all the possible actions with probability ϵ . Choosing a high exploration rate may help the agent to quickly *understand* the environment. However, it may also reduce the overall performance due to excessively exploring. On the other hand, a low exploration rate may increase the required time for the algorithm to converge to the optimal solution. In the simulation section, we investigate the performance of the Q-learning technique using different parameter options. The variable parameters include the exploration rate ϵ , the learning rate α , and the discount factor γ . Since the Q-learning technique is simple to implement and it does not require any prior knowledge of the environment, we also compare its performance to the previously proposed sub-band selection policies to validate the application of Q-learning techniques in this type of problems. A temporal illustration of the Q-learning procedure on the slotted time horizon is shown in Fig. 5.

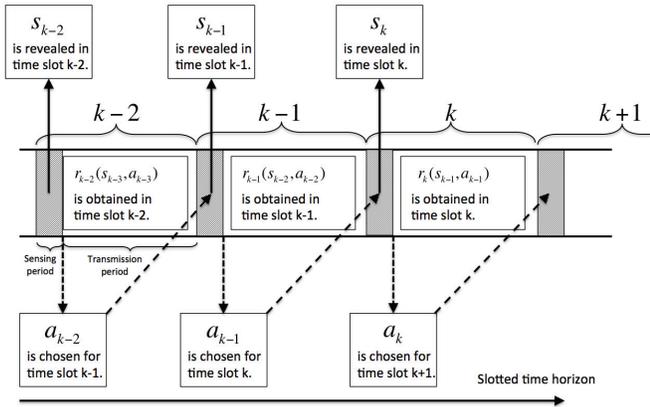


Fig. 5. An illustration of the Q-learning procedure on the slotted time horizon.

V. SIMULATION RESULTS AND DISCUSSIONS

In order to evaluate the performance of the proposed sub-band selection policies, we have conducted simulations for 3 test cases. For all the test cases, we assume that the spectrum sensing is errorless for illustration. In other words, the channel/sub-band states are revealed exactly each time the sub-band is sensed. Note that the errorless spectrum sensing is a special case of the formulation presented in this paper, when assumed, the whole formulation remains unchanged except that we have the state belief $b_0^{i,j}(k) = 0$, $b_1^{i,j}(k) = 1$ when

the (i, j) -th channel state is revealed as $S_{i,j}(k) = 1$, and $b_0^{i,j}(k) = 1$, $b_1^{i,j}(k) = 0$ when $S_{i,j}(k) = 0$. The simulation settings for the 3 test cases are summarized in Table. I. Note that the test cases 1, and 2 are based on simulated RF environments, whereas the test case 3 is based on real RF measurements for the 20 – 1500MHz band, with center frequency at 770 MHz inside a modern office building at Aachen, Germany [44].

For test cases 1 and 3, we assume that all channels have the same bandwidth, but the channel coefficients are independently Rayleigh-distributed. On the other hand, in test case 2, the individual channel throughputs are specifically assigned with non-random values for comparison purposes: in each configuration mode, one of the sub-bands is assumed to have channels with the same individual channel throughputs, whereas the other sub-band is assumed to have 2 channels with very high channel throughput and the other 8 channels with very low individual throughputs, such that all the sub-bands have the same sum of channel throughputs.

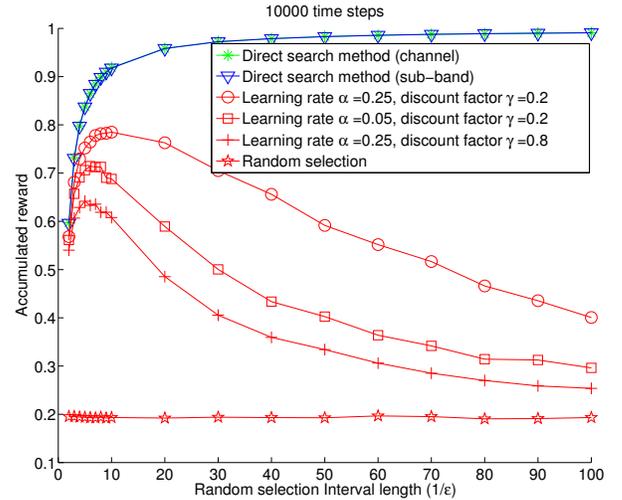


Fig. 6. Comparison of normalized accumulated reward of sub-band selection policies in 10,000 time steps for the first test case. The considered random selection interval length is set from 2 to 100.

In Fig. 6, we show the performance of the sub-band selection policies in the first test case. The simulated policies are: 1) the channel Markov models based policy using (14), 2) the sub-band Markov models based policy using (17), and 3) the Q-learning policy without any knowledge of the channel and sub-band Markov models. A trivial random policy is also included for comparison. The reward for all policies is defined as the actual obtained throughput less the energy consumption due to hardware reconfigurations (the energy consumption is weighted by the coefficient β), similar to the way the sub-band quality is defined in (13) and (16). The accumulated reward is then normalized with respect to a performance upper-bound. The performance upper-bound is obtained by assuming that each time after a sub-band selection decision is made, not only the state of the selected sub-band is revealed, but the states of all other sub-bands are also revealed. Since each time the sub-band selection maximize the

TABLE I
SIMULATION SETTINGS FOR THE CONSIDERED 4 TEST CASES.

Settings	Test case 1	Test case 2	Test case 3 with real measurement data
# of configuration modes	2	2	2
# of sub-bands in each mode	[3 3]	[2 2]	[2 3]
Total # of sub-bands	6	4	5
# of channels in each sub-band	10 each	10 each	10 each
Total # of channels	60	40	50
Max # of channels can be used for each time step: G	2	2	2
Time slot duration: T (seconds)	1	1	1
# of simulation time steps	10,000	10,000	12,000
Channel Markov models	Randomly generated.		Estimated from real world measurement data.
Required sensing time duration	The required sensing time duration in each sub-band is chosen uniformly between 0.1 sec and 1.0 sec.		
Sub-band Markov models	Obtained from channel Markov models.		
Reconfiguration coefficients	$c_1 = 1, c_2 = 0.8; t_1 = 0.1, t_2 = 0.05, t_3 = 0.01; \beta = 1.$		

immediate reward without affecting information update for the next step, the policy achieves the performance upper-bound for the POMDP solution. Note that this performance upper-bound is commonly used for the optimal POMDP solutions [13], [31]. The normalized accumulated reward is plotted against the random selection interval length. The random selection interval length refers to the average number of steps for which the CR makes a random selection. For instance, when the random selection interval is 100, the CR makes a random selection for every 100 steps on average. In all other time steps, the sub-band selection decisions are made accordingly to the selected policy. Note that the random selection interval length is equivalent to the inverse of the exploration rate ϵ in Q-learning. The trivial random selection policy selects a sub-band randomly and stays in that sub-band until the next time step in which another sub-band is randomly selected.

As shown in Fig. 6, the trivial random selection policy can only achieve a 20% of the performance whereas the two direct search methods (using (14) and (17)), achieve almost 100% of performance when the random selection interval is long (low exploration rate). In this case, the channel Markov model based policy and the sub-band Markov model based policy achieve almost the same performance. This may be explained by the structure of the simulated RF environment: all channels are statistically identical such that the product of the expected average individual channel throughput and the expected number of accessible channels is rather close to the sum of the expected highest throughputs from the expected accessible channels. As a result, the two different approaches of defining the sub-band qualities does not make a difference.

In the case of the Q-learning, the performance achieves the highest value of 78% when the random selection interval is roughly between 5 and 10, corresponding to an exploration rate in the range from $1/10$ to $1/5$. The highest performance of the Q-learning technique is achieved when the learning rate $\alpha = 0.25$ and the discount factor $\gamma = 0.8$. Since there is a total of 60 channels, without sufficient exploration (long random selection intervals), the performance of the Q-learning technique degrades. On the other hand, when the exploration rate is too high (very short random selection intervals), the performance degrades as well. Note that this

delicate balance between the exploration and exploitation is a well-known aspect of all RL algorithms [14], [32], [33]. A detailed performance of the Q-learning based policy for the first test case is shown in Fig. 7 for various combinations of the exploration rate ϵ , the learning rate α and the discount factor γ . For all the selected parameter combinations, the highest achieved performance is observed to be 78.03%, which is achieved when when $\epsilon = 1/7$, $\alpha = 0.05$ and $\gamma = 0.2$.

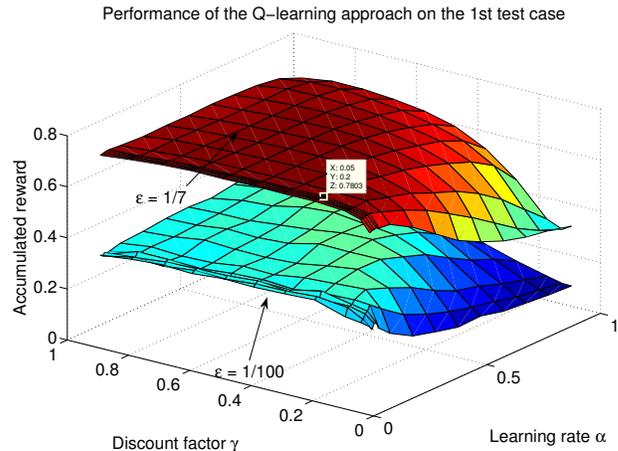


Fig. 7. Comparison of normalized accumulated reward of the Q-learning-based sub-band selection policy in 10,000 time steps for the first test case with different Q-learning parameter settings.

In Fig. 8, we show the performance of the sub-band selection policies in the second test case. The performance is normalized with respect to the performance upper-bound as introduced in the first test case. We can see that the trivial random selection method may achieve roughly 65% of the performance whereas the sub-band selection policy using the channel Markov models achieves almost 100% performance at low exploration rate (long random selection interval). The sub-band selection policy using the sub-band Markov models can only achieve roughly 50% with a high exploration rate. The performance difference between the channel Markov model based policy and the sub-band Markov model based policy can be explained as follows. Note that the expected individual

channel throughputs are specifically assigned such that in each configuration mode, one of the sub-bands is assumed to have channels with the same individual channel throughputs, whereas the other sub-band is assumed to have 2 channels with very high channel throughput and the other 8 channels with very low individual throughputs but the resulting sum throughputs of all individual sub-bands are the same. Also note that the sub-band quality defined in the sub-band Markov model based policy computes the expected sub-band throughput by finding the product of the expected average individual channel throughput and the expected number of accessible channels. On the other hand, the channel Markov model based policy computes the expected sub-band throughput by finding the sum of the expected highest individual channel throughputs of the expected accessible channels. The latter gives a better estimate of the expected sub-band throughputs with the setting of $G = 2$, since the sub-band Markov model based policy sees all the sub-bands having the same expected sub-band throughput. However, the channels are distinct and the actual communication throughput is much lower in those sub-bands with channels of the same channel throughput, compared to those sub-bands with 2 channels with very high channel throughput. As a result, the channel Markov model based policy gives much better performance compared to the sub-band Markov model based policy. Note that although the channel Markov models based sub-band selection policy may achieve better results, the unavailability of the required knowledge in practical scenarios may prohibit the application of the policy. In this case, using the Q-learning based policy may be a better choice. As shown in Fig. 8, the Q-learning based policy is capable of achieving the performance at 90%, when $\alpha = 0.25$, $\gamma = 0.2$, and the exploration rate ϵ between $1/8$ and $1/6$.

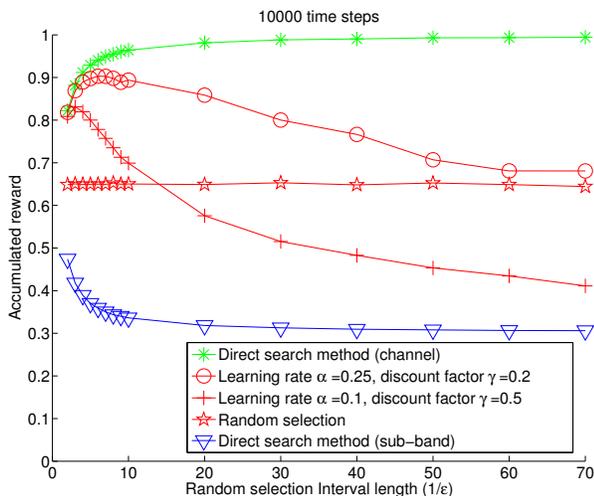


Fig. 8. Comparison of normalized accumulated reward of sub-band selection policies in 10,000 time steps for the second test case. The considered random selection interval length is set from 2 to 70.

A detailed performance of the Q-learning policy in the second test case is shown in Fig. 9 for various combinations of the exploration rate ϵ , the learning rate α and the discount

factor γ . For all the selected parameter combinations, the highest achieved performance is observed to be 92.24%, which is achieved when $\epsilon = 1/6$, $\alpha = 0.01$ and $\gamma = 0.5$.

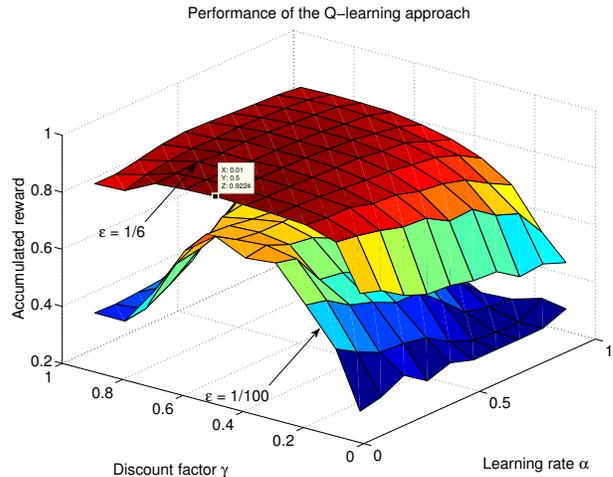


Fig. 9. Comparison of normalized accumulated reward of the Q-learning-based sub-band selection policy in 10,000 time steps for the second test case with different Q-learning parameter settings.

In Fig. 10, we show the Q-learning policy for the third test case with real RF measurement data for the 20 – 1500MHz band, with center frequency at 770 MHz inside a modern office building at Aachen, Germany [44]. The data is the measured values of the power spectrum density (PSD) with a resolution bandwidth of 200kHz taken each second. For simplicity, the communication channels are also considered as spaced at 200kHz and each data point corresponds to a channel [44]. We randomly selected 50 channels over a time duration of 12,000 seconds for the simulation. We assume that the wide-band CR has two reconfiguration modes with the first mode contains two sub-bands and the other contains three sub-bands and that each sub-band contains 10 channels as shown in Table. I. The channel occupancies (idle and busy states) are then determined by a thresholding test of the measurement data of each channel, similar to [44]. In this test case, we found that the channel and sub-band state transitions do not exhibit stationary Markov properties. This is found out by performing the built-in Matlab function *hmmestimate* on the data such that different portions of the data (with each portion corresponds to 2,000 seconds of data) give significantly different estimated state transition probabilities. Note that this is similar to the observation in [45] that a simple discrete-time Markov chain model is not able to accurately capture the channel load variations⁵. In this case, in order to obtain the performance upper-bound as used in previous two test cases, we obtained the Markov model parameters for the entire data. However, we observed that the Q-learning base policy outperforms the ‘upper-bound’. This is due to the non-stationarity of the state dynamics of the measured RF environment and the

⁵When the channel is sparsely used (low load), the length of idle periods is significantly higher than that of busy periods. On the other hand, when the channel is subject to an intensive usage (high load), the length of busy periods increases, whereas idle periods become notably shorter.

assumptions of the time-invariant transition probabilities of the channels and sub-bands do not *capture* the non-stationary scenario, so that the obtained performance ‘upper-bound’ is not a performance upper-bound. As a result, we obtained a loose performance upper-bound by assuming that before a sub-band selection decision is about made, all sub-band and channels states are exactly revealed for the next time step. As shown in Fig. 10, the obtained Q-learning policy performance is normalized to the loose upper-bound. A performance of 78.9% is achieved when the exploration rate $\epsilon = 1/6$, the learning rate $\alpha = 0.01$, and the discount factor $\gamma = 0.7$. For comparison, the trivial random selection policy as introduced in the first test case can only achieve a 52% of performance. Due to the space limitation, we do not show the performance of the random selection policy.

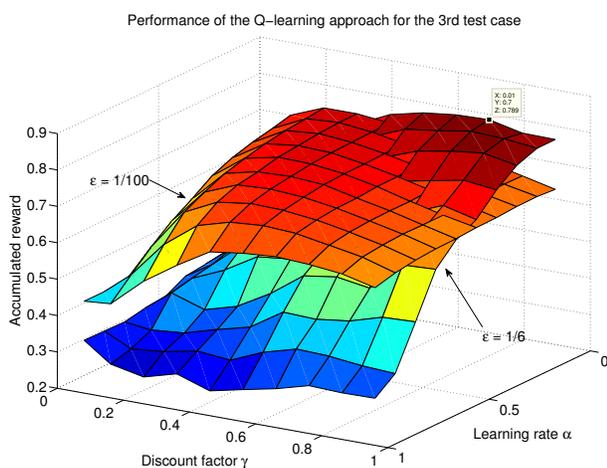


Fig. 10. Comparison of normalized accumulated reward of the Q-learning-based sub-band selection policy in 12,000 time steps for the third test case with different Q-learning parameter settings.

In summary, the two Markov-based sub-band selection policies may achieve good results. However, the performance may vary depending on the RF environment. The required Markov knowledge may not be easy to obtain in some cases. On the other hand, the Q-learning policy achieves reasonable results (around 80 – 90% performance) in all test cases, with a much lower computational effort without any knowledge of the channel/sub-band Markov models. As a result, we validate the application of the Q-learning technique in the wide-band spectrum sensing problem. In order to achieve the autonomous operation of the CR in practical RF environments, the CR may adopt a certain Machine-learning technique to fine tune the parameters of the Q-learning method. However, due to the focus of this paper, the higher level autonomous behavior is out of the scope of this work.

VI. CONCLUSION

In this paper, we investigate a frequency spectrum sensing scheduling problem in a realistic wide-band spectrum sensing setup for a CR equipped with a reconfigurable RF front-end with several operation modes to cover a wide frequency range

of interest. We assume that within each operation mode, the frequency range is further divided into several frequency sub-bands and that the CR can only perform spectrum sensing in one sub-band at a time. We propose three different sub-band selection policies for the spectrum sensing scheduling problem: 1) a myopic sub-band selection policy based on the channel Markov models; 2) a myopic sub-band selection policy based on the sub-band Markov models; 3) a Q-learning policy without the knowledge of the channel and sub-band Markov models. Realistic RF front-end reconfiguration costs such as energy consumption and time delays are considered. We show that the proposed sub-band selection policies achieve good results comparing to a commonly used performance upper-bound for the POMDP solution. We also show that in both simulated and real measured RF environments, the Q-learning technique may achieve around 80 – 90% of the performance upper-bound without any knowledge of the RF environment, which validates the Q-learning application in the wide-band spectrum sensing problems.

REFERENCES

- [1] J. Mitola, III and G. Maguire, Jr., “Cognitive radio: making software radios more personal,” *IEEE Personal Communications*, vol. 6, no. 4, pp. 13 –18, Aug. 1999.
- [2] B. Wang and K. J. R. Liu, “Advances in cognitive radio networks: A survey,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 1, pp. 5 –23, Feb. 2011.
- [3] S. K. Jayaweera and C. G. Christodoulou, “Radiobots: Architecture, algorithms and realtime reconfigurable antenna designs for autonomous, self-learning future cognitive radios,” University of New Mexico, Technical Report EECE-TR-11-0001, Mar. 2011. [Online]. Available: <http://repository.unm.edu/handle/1928/12306>
- [4] S. Haykin, “Cognitive radio: brain-empowered wireless communications,” *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 2, pp. 201 – 220, Feb. 2005.
- [5] J. Mitola, “Cognitive radio architecture evolution,” *Proceedings of the IEEE*, vol. 97, no. 4, pp. 626 –641, Apr. 2009.
- [6] “IEEE-USA President Commends FCC for National Broadband Plan [IEEE;USA],” *Antennas and Propagation Magazine, IEEE*, vol. 52, no. 2, p. 179, April 2010.
- [7] S. Jayaweera, Y. Li, M. Bkassiny, C. Christodoulou, and K. Avery, “Radiobots: The autonomous, self-learning future cognitive radios,” in *International Symposium on Intelligent Signal Processing and Communications Systems (ISPACS '11)*, Chiangmai, Thailand, Dec. 2011, pp. 1 –5.
- [8] T. Yucek and H. Arslan, “A survey of spectrum sensing algorithms for cognitive radio applications,” *IEEE Communications Surveys Tutorials*, vol. 11, no. 1, pp. 116–130, Mar. 2009.
- [9] E. Biglieri, “Effect of uncertainties in modeling interferences in coherent and energy detectors for spectrum sensing,” in *IEEE International Symposium on Information Theory Proceedings (ISIT '11)*, Aug. 2011, pp. 2418 –2421.
- [10] Z. Tian and G. B. Giannakis, “A wavelet approach to wideband spectrum sensing for cognitive radios,” in *1st International Conference on Cognitive Radio Oriented Wireless Networks and Communications*, Mykonos Island, Greece, June 2006, pp. 1–5.
- [11] Z. Tian and G. Giannakis, “Compressed sensing for wideband cognitive radios,” in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '07)*, vol. 4, Honolulu, HI, Apr. 2007, pp. IV-1357 –IV-1360.
- [12] S. Maleki, A. Pandharipande, and G. Leus, “Two-stage spectrum sensing for cognitive radios,” in *IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP '10)*, Dallas, TX, Mar. 2010, pp. 2946 –2949.
- [13] Y. Li, S. Jayaweera, M. Bkassiny, and K. Avery, “Optimal myopic sensing and dynamic spectrum access in cognitive radio networks with low-complexity implementations,” *IEEE Transactions on Wireless Communications*, vol. 11, no. 7, pp. 2412 –2423, July 2012.

- [14] M. Bkassiny, S. K. Jayaweera, and K. A. Avery, "Distributed Reinforcement Learning based MAC protocols for autonomous cognitive secondary users," in *20th Annual Wireless and Optical Communications Conference (WOCC '11)*, Newark, NJ, Apr. 2011, pp. 1–6.
- [15] Y. Li, S. K. Jayaweera, M. Bkassiny, and K. A. Avery, "Optimal myopic sensing and dynamic spectrum access with low-complexity implementations," in *IEEE Vehicular Technology Conference (VTC-spring '11)*, Budapest, Hungary, May 2011.
- [16] M. Bkassiny, S. K. Jayaweera, Y. Li, and K. A. Avery, "Optimal and low-complexity algorithms for dynamic spectrum access in centralized cognitive radio networks with fading channels," in *IEEE Vehicular Technology Conference (VTC-spring '11)*, Budapest, Hungary, May 2011.
- [17] C. Ghosh, S. Roy, and M. B. Rao, "Modeling and validation of channel idleness and spectrum availability for cognitive networks," *IEEE Journal on Selected Areas in Communications*, Nov. 2012.
- [18] Y. Li, S. Jayaweera, and C. Christodoulou, "Wideband PHY/MAC bandwidth aggregation optimization for cognitive radios," in *Cognitive Information Processing (CIP), 2012 3rd International Workshop on*, May 2012.
- [19] J. Lee and J. So, "Analysis of cognitive radio networks with channel aggregation," in *Wireless Communications and Networking Conference (WCNC), 2010 IEEE*, April 2010, pp. 1–6.
- [20] F. Huang, W. Wang, H. Luo, G. Yu, and Z. Zhang, "Prediction-based spectrum aggregation with hardware limitation in cognitive radio networks," in *Vehicular Technology Conference (VTC 2010-Spring), 2010 IEEE 71st*, May 2010, pp. 1–5.
- [21] K. Liu, Q. Zhao, and B. Krishnamachari, "Dynamic multichannel access with imperfect channel state detection," *Signal Processing, IEEE Transactions on*, vol. 58, no. 5, pp. 2795–2808, May 2010.
- [22] Y. Chen, Q. Zhao, and A. Swami, "Distributed spectrum sensing and access in cognitive radio networks with energy constraint," *IEEE Transactions on Signal Processing*, vol. 57, no. 2, pp. 783–797, 2009.
- [23] R. D. Smallwood and E. J. Sondik, "The optimal control of partially observable Markov processes over a finite horizon," *Operations Research*, vol. 21, no. 5, pp. 1071–1088, Sept.-Oct. 1973.
- [24] E. J. Sondik, "The optimal control of partially observable markov processes over the infinite horizon: Discounted costs," *Operations Research*, vol. 26, no. 2, pp. pp. 282–304, Mar.-Apr. 1978.
- [25] Y. Pei, Y.-C. Liang, K. Teh, and K. H. Li, "How much time is needed for wideband spectrum sensing?" *IEEE Transactions on Wireless Communications*, vol. 8, no. 11, pp. 5466–5471, 2009.
- [26] M. Bkassiny, S. Jayaweera, Y. Li, and K. Avery, "Wideband Spectrum Sensing and Non-Parametric Signal Classification for Autonomous Self-Learning Cognitive Radios," *IEEE Transactions on Wireless Communications*, vol. 11, no. 7, pp. 2596–2605, 2012.
- [27] Y. Tawk, J. Costantine, and C. Christodoulou, "A rotatable reconfigurable antenna for cognitive radio applications," in *2011 IEEE Radio and Wireless Symposium (RWS)*, Jan. 2011.
- [28] Y. Tawk, M. Bkassiny, G. El-Howayek, S. Jayaweera, K. Avery, and C. Christodoulou, "Reconfigurable front-end antennas for cognitive radio applications," *IET Microwaves, Antennas Propagation*, Jan. 2011.
- [29] J. W. Kim, M. Chu, P. Jacob, A. Zia, R. Kraft, and J. McDonald, "Reconfigurable 40 ghz bimos uniform delay crossbar switch for broadband and wide tuning range narrowband applications," *IET Circuits, Devices Systems*, vol. 5, no. 3, pp. 159–169, 2011.
- [30] J. Papapolymerou, K. Lange, C. Goldsmith, A. Malczewski, and J. Kleber, "Reconfigurable double-stub tuners using mems switches for intelligent rf front-ends," *IEEE Transactions on Microwave Theory and Techniques*, vol. 51, no. 1, pp. 271–278, 2003.
- [31] J. Unnikrishnan and V. Veeravalli, "Algorithms for dynamic spectrum access with learning for cognitive radio," *IEEE Transactions on Signal Processing*, vol. 58, no. 2, pp. 750–760, Feb. 2010.
- [32] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. Cambridge, MA: MIT Press, 1998.
- [33] C. Watkins, "Learning from delayed rewards," Ph.D. dissertation, University of Cambridge, United Kingdom, 1989.
- [34] Y. Reddy, "Detecting Primary Signals for Efficient Utilization of Spectrum Using Q-Learning," in *Fifth International Conference on Information Technology: New Generations (ITNG '08)*, Las Vegas, NV, Apr. 2008, pp. 360–365.
- [35] H. Li, "Multi-agent Q-learning of channel selection in multi-user cognitive radio systems: A two by two case," in *IEEE International Conference on Systems, Man and Cybernetics (SMC '09)*, San Antonio, TX, Oct. 2009, pp. 1893–1898.
- [36] A. Galindo-Serrano and L. Giupponi, "Distributed Q-Learning for Aggregated Interference Control in Cognitive Radio Networks," *IEEE Transactions on Vehicular Technology*, vol. 59, no. 4, pp. 1823–1834, May 2010.
- [37] M. L. Littman, A. R. Cassandra, and L. P. Kaelbling, "Learning Policies for Partially Observable Environments: Scaling Up," *Readings in agents*, pp. 495–503, 1998.
- [38] E. N. Gilbert, "Capacity of a burst-noise channel," *Bell System Technical Journal*, vol. 39, pp. 1253–1265, Sept. 1960.
- [39] S. Ahmad, M. Liu, T. Javidi, Q. Zhao, and B. Krishnamachari, "Optimality of Myopic Sensing in Multichannel Opportunistic Access," *IEEE Transactions on Information Theory*, vol. 55, no. 9, pp. 4040–4050, Sept. 2009.
- [40] J. Aruz, "Discrete rayleigh fading channel modeling," *Wireless Communications and Mobile Computing 2004*, vol. 4, pp. 413–425, 2004.
- [41] H. V. Poor, *An Introduction to Signal Detection and Estimation*, 2nd ed. New York: Springer, 1998, ch. 3, pp. 72–76.
- [42] F. Millioz and N. Martin, "Estimation of a white gaussian noise in the short time fourier transform based on the spectral kurtosis of the minimal statistics: Application to underwater noise," in *IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP '10)*, Dallas, TX, Mar. 2010, pp. 5638–5641.
- [43] L. Luo, N. Neihart, S. Roy, and D. Allstot, "A two-stage sensing technique for dynamic spectrum access," *IEEE Transactions on Wireless Communications*, vol. 8, no. 6, pp. 3028–3037, 2009.
- [44] M. Wellens and P. Mahonen, "Lessons learned from an extensive spectrum occupancy measurement campaign and a stochastic duty cycle model," in *5th International Conference on Testbeds and Research Infrastructures for the Development of Networks Communities and Workshops (TridentCom), 2009, Washington DC, USA*, 2009, pp. 1–9.
- [45] M. Lopez-Benitez and F. Casadevall, "Empirical Time-Dimension Model of Spectrum Use Based on a Discrete-Time Markov Chain With Deterministic and Stochastic Duty Cycle Models," *IEEE Transactions on Vehicular Technology*, vol. 60, no. 6, pp. 2519–2533, 2011.



Yang Li received the B.E. degree in Electrical Engineering from the Beijing University of Aeronautics and Astronautics, Beijing, China, in 2005 and the M.S. degree in Electrical Engineering from New Mexico Institute of Mining and Technology, Socorro, New Mexico, USA in 2009. He received his PhD degree in Electrical Engineering with Distinction from the University of New Mexico, Albuquerque, NM, USA in 2013. His current research interests are in cognitive radios, machine learning, wireless sensor networks, digital signal processing, signal detection and estimation, embedded systems, wearable technologies, and home automations, etc.



Sudharman K. Jayaweera (S00, M04, SM09) was born in Matara, Sri Lanka. He completed his High school education at the Rahula College, Matara, and worked as a science journalist at the Associated Newspapers Ceylon Limited (ANCL) till 1993. Later, he received the B.E. degree in Electrical and Electronic Engineering with First Class Honors from the University of Melbourne, Australia, in 1997 and M.A. and PhD degrees in Electrical Engineering from Princeton University, USA in 2001 and 2003, respectively. He is currently an Associate Professor

in Electrical Engineering at the Department of Electrical and Computer Engineering at University of New Mexico, Albuquerque, NM. Dr. Jayaweera held an Air Force Summer Faculty Fellowship at the Air Force Research Laboratory, Space Vehicles Directorate (AFRL/RVSV) from 2009-2011 and was a National Research Council (NRC) Senior Fellow at the Naval Postgraduate School in Monterey, CA in 2013.

Dr. Jayaweera is currently an associate editor of IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY. He has also served as a member of the Technical Program Committees of numerous IEEE conferences and was the Tutorial and Workshop Chair of the 2013 Fall IEEE Vehicular Technology Conference. His current research interests include cooperative and cognitive communications, machine learning, smart-grid control and optimization, statistical signal processing and information theory of networked-control systems.



Mario Bkassiny (S'06, M'14) received the B.E. degree in Electrical Engineering with High Distinction and the M.S. degree in Computer Engineering from the Lebanese American University, Lebanon, in 2008 and 2009, respectively. He received his PhD degree in Electrical Engineering from the University of New Mexico, Albuquerque, NM, USA in 2013. He is currently an Assistant Professor in Electrical Engineering at the Department of Electrical and Computer Engineering at State University of New York (SUNY) at Oswego, Oswego, NY, USA. He

worked as a research assistant at the Communication and Information Sciences Laboratory (CISL), Department of Electrical and Computer Engineering at the University of New Mexico from 2009 to 2013.

Dr. Bkassiny served as a technical co-chair of the Workshop on Wideband Cognitive Radio Communication and Networks (WCRCN) at the IEEE Vehicular Technology Conference (VTC-Fall 2013). His current research interests are in cognitive radios, distributed learning and reasoning, signal classification, wideband spectrum sensing, cyclostationary detection and wireless communications.



Chittabrata Ghosh is currently a Principal Researcher at Nokia Research Center, Berkeley. Before joining Nokia in July 2011 as a Senior Researcher, he pursued Postdoctoral Research in the Electrical Engineering Department at University of Washington, Seattle. He is an active contributor to IEEE 802.11ah, a long-range low power amendment to IEEE 802.11 Wireless Local Area Network standard and was a member of the Executive Review Committee towards the draft development. His research interests lie in developing spectrum sharing concepts

for next generation commercial communications systems and solve coexistence challenges among cooperating heterogeneous wireless networks. He is an author of more than 30 conference and journal publications and an author of over 30 filed and published patents. He has received the IEEE PACRIM Gold Award for Best Communications paper in 2011 and is also a recipient of the Best Paper Award in IEEE PIMRC 2012. He is serving as TPC member of various esteemed conferences and reviews peer-reviewed journal papers. He is currently serving as the Vice Chair of the IEEE Technical Committee on Simulations (TCSIM).