Distributed Spectrum Sensing and Access in Cognitive Radio Networks with Energy Constraint

Yunxia Chen, Qing Zhao*, Ananthram Swami

Abstract

We design distributed spectrum sensing and access strategies for opportunistic spectrum access (OSA) under an energy constraint on secondary users. Both the continuous and the bursty traffic models are considered for different applications of the secondary network. In each slot, a secondary user sequentially decides whether to sense, where in the spectrum to sense, and whether to access. By casting this sequential decision-making problem into the framework of partially observable Markov decision processes, we obtain stationary optimal spectrum sensing and access polices that maximize the throughput of the secondary user during its battery lifetime. We also establish threshold structures of the optimal policies and study the fundamental tradeoffs involved in the energy-constrained OSA design. Numerical results are provided to investigate the impact of the secondary user's residual energy on the optimal spectrum sensing and access decisions.

Index Terms

Opportunistic spectrum access, spectrum sensing, cognitive radio, partially observable Markov decision process.

This work was supported in part by the Army Research Laboratory CTA on Communication and Networks under Grant DAAD19-01-2-0011 and by the National Science Foundation under Grants CNS-0627090 and ECS-0622200.

Part of this work was presented in the *IEEE MILCOM Conference*, Oct. - Nov. 2006 and the *IEEE GLOBECOM Conference*, Nov. 2007.

Yunxia Chen and Qing Zhao are with the Department of Electrical and Computer Engineering, University of California, Davis, CA 95616. Emails: {yxchen,qzhao}@ece.ucdavis.edu. Ananthram Swami is with the Army Research Laboratory, Adelphi, MD 20783. Email: aswami@arl.army.mil.

* Corresponding author. Phone: 1-530-752-7390. Fax: 1-530-752-8428.

I. Introduction

Opportunistic spectrum access (OSA), also referred to as spectrum overlay, is one of the approaches envisioned for dynamic spectrum management. It has received increasing attention due to its potential for improving spectrum efficiency and its compatible with the current spectrum management policy and legacy wireless systems. The basic idea of OSA is to allow secondary users to search for and exploit local and instantaneous spectrum opportunities with limited interference to primary users. The physical platform of OSA and other dynamic spectrum access strategies is cognitive radio, which is capable of agile sensing and communication through adaptive learning [1]. As such, cognitive radio is often used as a synonym for different dynamic spectrum access strategies (see [2] for a survey of different approaches envisioned for dynamic spectrum access).

In this paper, we focus on the design of distributed medium access control (MAC) protocols for OSA under an energy constraint on secondary users. We consider secondary users, each with a finite amount of initial energy, exploiting temporal spectrum opportunities in a slotted primary system. In each slot, a secondary user either turns off its transceiver to save energy or chooses a channel in the spectrum to sense and possibly access, resulting in different levels of reduction in its battery energy. A MAC protocol governing such a sequential decision-making process thus consists of two components: (i) a sensing strategy that specifies whether to sense and where in the spectrum to sense; (ii) an access strategy that determines whether to access based on the sensing outcomes regarding the occupancy state (idle or occupied by primary users) and the fading condition of the channel. The design objective is to maximize the throughput of a secondary user during its battery lifetime. We propose optimal MAC protocols for both the continuous and the bursty traffic models. For brevity, we adopt the continuous traffic model, where the secondary user always has packets to transmit, unless otherwise specified.

A. Energy-Constrained OSA Design

While optimal distributed MAC protocols for OSA have been proposed in [3], [4], the impact of energy constraint on optimal sensing and access protocols has not been studied. The incorporation of the energy constraint significantly complicates the problem. First consider the sensing strategy. Without the energy constraint, the secondary user should always sense, and its channel selection should exploit the spectrum occupancy statistics to achieve the best tradeoff between gaining immediate access and gaining statistical information about the spectrum occupancy [3], [4]. With the energy constraint, however, the secondary user, even with packets to transmit, may choose to sleep to conserve energy. Moreover, channel selection should also exploit channel fading statistics since a channel in deep fading requires more energy for transmission. The design tradeoff involved in sensing decisions thus lies among three often conflicting objectives: gaining immediate access, gaining spectrum occupancy information, and conserving energy.

It has been shown in [4] that without the energy constraint, the optimal access strategy is to access if and only if the channel is sensed as idle, provided that the operating characteristics (false alarm rate vs. miss detection rate) of the spectrum sensor is chosen optimally according to the interference constraint. With the energy constraint, channel fading statistics play an important role in access decision-making. For example, when the sensed channel is idle but has poor fading condition, should the secondary user with packets to send access this channel to gain immediate reward or wait for better channel realizations to save transmission energy but waste the energy already used in sensing? Clearly, such a decision depends on the secondary user's residual energy level and its energy consumption characteristics, as well as the channel fading statistics.

Bursty Traffic Bursty traffic of the secondary user further complicates the design. In this case, the design tradeoffs vary with the secondary user's buffer state. Specifically, when the buffer is empty, the secondary user does not need to gain immediate access in the current slot. Hence, sensing decisions are made for the sole purpose of gaining statistical information about spectrum occupancy. The question raised here is whether the secondary user should continue tracking the dynamics of the spectrum for future use or turn off its transceiver to save energy. Intuitively, the sensing strategy employed by the secondary user when its buffer is empty should be fundamentally different from the one used when it has packets to transmit.

B. Main Results

Within the framework of partially observable Markov decision process (POMDP), we tackle the optimal MAC design for energy-constrained OSA. By modeling the primary users' traffic as a Markov chain, we formulate the problem of dynamically choosing whether to sense, where in the spectrum to sense, and whether to access for maximum throughput as a POMDP with a finite but random time horizon. This formulation allows us to integrate the dynamics of spectrum occupancy and channel fading into the MAC design. The optimal MAC design is given by the stationary optimal policy of this POMDP, which can be solved using existing POMDP algorithms.

To gain insights into the energy-constrained OSA problem, we search for structures of the optimal sensing and access policies. We show that in the single-channel case, the optimal sensing decision (whether to sleep or sense) has a threshold structure: the secondary user should sense the channel if and only if the conditional probability that the channel is idle in the current slot (conditioned on the entire sensing and observation history) is above a certain threshold (referred to as the sensing threshold). We also show that the optimal access strategy is a threshold policy in terms of the channel fading condition. That is, the secondary user should access the channel if and only if the sensing outcome indicates that the channel is idle and its fading condition is better than a certain threshold (referred to as the access threshold). These structural results not only reveal the fundamental design tradeoffs but also reduce the computational complexity in searching for the optimal policies.

These structural results are complemented with numerical examples. We study different factors that affect the optimal sensing and access decisions. We find that the impact of the secondary user's residual energy on the optimal decisions diminishes as the residual energy increases. This observation indicates that energy conservation only plays a critical role in sensing and access decisions when the battery of the secondary user is close to depletion. We also find that when the sensing energy consumption is large, the secondary user should be more conservative in sensing, but more aggressive in access. Specifically, the secondary user should increase the sensing threshold and lower the access threshold.

Bursty Traffic We also extend our analysis to the case where the secondary user has bursty traffic. As explained in Section I-A, the optimal sensing decisions in this case should incorporate the secondary user's buffer state. We, however, note that due to random packet arrivals, the receiver does not know the secondary user's buffer state. This impedes optimal distributed design since in the absence of additional control channels, transceiver synchronization requires the secondary user and its receiver to have the same information for decision-making [3], [4]. We overcome this obstacle by treating the buffer state as a partially observable parameter. The secondary user and its receiver can thus make sensing decisions based on the conditional probability mass function (PMF) of the buffer state. We show that the secondary user with an empty buffer can benefit from sensing a channel if the time-correlation of the spectrum occupancy state is sufficiently large.

C. Related Work

Cognitive MAC design for OSA has been addressed under different network architectures (see [3]–[6] and references therein). In [5], the authors address the implementation of a MAC protocol for OSA in a GSM primary network. A dedicated control channel is required for the secondary transmitter and receiver to exchange information about channel selection. In [3], optimal distributed MAC protocols are proposed for OSA in slotted primary systems. The proposed protocols ensure synchronous hopping of the secondary transmitter and receiver in the spectrum without requiring central controllers or control channels. More recently, sensing errors have been taken into account in the MAC design [3], [4]. Significantly, a separation principle is established in [4] for the optimal joint design of the physical layer spectrum sensor and the MAC layer sensing and access strategies. In [6], access strategies for a slotted secondary user searching for opportunities in an un-slotted primary network are considered, where a round-robin single-channel sensing scheme is used and sensing is considered to be perfect. The joint design of the spectrum sensor and sensing and access strategies for OSA in un-slotted primary systems has been addressed in [7]. To our best knowledge, energy-constrained OSA design has not been considered in the literature.

Statistical models for spectrum usage of primary systems are important for OSA protocol design. Existing work along this line can be found in [8]–[10]. Measurements obtained from spectrum monitoring test-beds demonstrate the Makovian transition between busy and idle channel states in wireless LANs [8], a model similar to that used in this paper. With these active experimental research activities, we can perhaps foresee a public database of statistical models of spectrum usage in different bands and at different times and locations. Secondary users can then download the required model for the design of spectrum sensing and access strategies.

An overview of challenges and recent developments in OSA can be found in [11].

D. Organization and Notation

The rest of this paper is organized as follows. After describing the primary and the secondary network models in Section II, we formulate the optimal MAC design for energy-constrained OSA as a POMDP over a random horizon in Section III. In Section IV, we derive recursive formulas for solving this POMDP and establish structures of the solution. We also address the distributed implementation of the obtained optimal design. In Section V, we further establish the threshold structures of the optimal sensing and access policies and study different factors that affect the optimal decisions. Finally, Section VI focuses on the energy-constrained OSA design for secondary users with bursty traffic, and Section VII concludes the paper.

Random variables and their realizations are denoted by capital and small letters, respectively. Vectors are denoted by boldfaced letters. For two equal-length vectors $\mathbf{x} = [x_1, x_2, \ldots, x_N]$ and $\mathbf{y} = [y_1, y_2, \ldots, y_N]$, we say $\mathbf{x} \geq \mathbf{y}$ if $x_k \geq y_k$ for all k. Let $1_{[X]}$ denote the indicator function: $1_{[X]} = 1$ if event X occurs and zero otherwise.

II. Network Model

A. Primary Network Model

Consider a spectrum consisting of N channels, each with bandwidth B_n ($n = 1, \cdots, N$). These N channels are licensed to a primary network employing a synchronous slotted communication protocol.

The primary traffic is modeled as a time-homogeneous discrete Markov process. Specifically, let $S_n(t) \in \{0 \text{ (busy)}, 1 \text{ (idle)}\}$ denote the occupancy of channel n by the primary network in slot t. The spectrum occupancy state (SOS), denoted by $\mathbf{S}(t) \stackrel{\Delta}{=} [S_1(t), \ldots, S_N(t)]$, forms a Markov chain with state space $\mathcal{S} \stackrel{\Delta}{=} \{0,1\}^N$. The transition probabilities are denoted by

$$P_{\mathbf{S}}(\mathbf{s}'|\mathbf{s}) \stackrel{\Delta}{=} \Pr\{\mathbf{S}(t) = \mathbf{s}' \,|\, \mathbf{S}(t-1) = \mathbf{s}\}, \quad \mathbf{s}, \mathbf{s}' \in \mathcal{S}, \tag{1}$$

which are determined by the statistics of the primary traffic and assumed known to secondary users.

B. Secondary Network Model

Consider an overlay ad hoc secondary network whose users independently and selfishly search for, according to a MAC protocol, instantaneous spectrum opportunities in these N channels. We assume that each secondary user can only sense and access one channel in a slot. At the beginning of each slot, a secondary user first determines its operation mode: sleeping or sensing. If the former, the user turns off its transceiver until the next slot. If the latter, the user chooses one channel to sense and then decides whether to access this channel based on the sensing outcome. We assume that sensing errors are negligible.

The optimal sensing and access decisions are made based on the user's statistical knowledge of the SOS and its own residual energy. Our goal is to design the optimal sensing and access strategies that maximizes the throughput of an individual secondary user during its battery lifetime.

Channel Fading Model We adopt a block channel fading model[1]. Specifically, we assume that the channel gain between the secondary user and its receiver is a random variable independently and identically distributed (i.i.d.) across slots but not necessarily i.i.d. across channels.

Energy Model The secondary user is powered by a battery with finite initial energy $\mathring{e}$. Energy consumption in a slot may include the following: (i) the energy e_p consumed in the sleeping mode; (ii) the energy e_s consumed in sensing the channel occupancy and estimating the channel fading condition[2]; (iii) the energy $E_{tx}(n)$ consumed in successfully transmitting over channel n in a slot. In general, we have $e_p \ll e_s < E_{tx}(n)$. For ease of presentation, we assume that the sleeping energy e_p and the sensing energy e_s are constants, invariant to channel fading.

[1]Our analysis can be readily extended to a more general Markovian fading channel model. See details in Section III-B.

[2]An interesting variation is to separate the energy for sensing channel occupancy from that for estimating channel fades; the latter would be consumed only if the channel is sensed to be idle. This variation is easily incorporated into the framework developed here.

Due to hardware and power limitations, the secondary user only has a finite number L of transmission power levels. We assume that the user transmits at a fixed rate. Hence, to ensure successful transmission, the user has to adjust its transmission power according to the current channel fading condition. The transmission energy consumption $E_{tx}(n)$ is thus a random variable depending on the current channel fading condition. In general, the better the channel, the lower the transmission power level. Let ε_k denote the energy consumed in transmitting at the kth power level with $\varepsilon_1 < \ldots < \varepsilon_L$. The PMF of $E_{tx}(n)$ is determined by channel fading statistics, and is denoted by

$$p_n(k) \stackrel{\Delta}{=} \Pr\{E_{tx}(n) = \varepsilon_k\}, \quad k = 1, \ldots, L. \tag{2}$$

Let $E(t)$ denote the secondary user's residual energy at the beginning of slot t. Due to random transmission energy consumption, $E(t)$ is also a random variable taking values from a finite set $\mathcal{E}$:

$$\mathcal{E} \stackrel{\Delta}{=} \left\{ e : e = \mathring{e} - \sum_{k=1}^{L} c_k \varepsilon_k - c_s e_s - c_p e_p \geq 0; c_s \geq \sum_{k=1}^{L} c_k; c_k, c_s, c_p \in \{0\} \cup \mathbb{Z}^+; \right\}, \tag{3}$$

where c_p, c_s, c_k are, respectively, the numbers of slots when the secondary user turns into the sleeping mode, senses a channel, and transmits at the kth power level. Since the secondary user is required to sense a channel before accessing it in order to avoid collisions with primary users, we have $c_s \geq \sum_{k=1}^{L} c_k$.

Traffic Model In Sections III - V, we adopt a continuous traffic model, *i.e.,* the secondary user always has packets to transmit. The case where secondary users have bursty traffic is considered in Section VI.

III. A Decision-Theoretic Framework

In this section, we formulate the optimal energy-constrained OSA design as a POMDP. This formulation allows us to incorporate the secondary user's residual energy into sensing and access decisions at the MAC layer. We show that the optimal energy-constrained OSA strategy is given by the optimal policies of this POMDP.

A. Sequential Decision-Making

Sensing Decision A the beginning of slot t, based on its statistical knowledge of the SOS and its current residual energy, the secondary user first determines its operating mode in this slot: sleeping or sensing. If the sleeping mode is chosen, no more decisions need to be made in this slot. Otherwise, the user chooses a channel n to sense. Let 0 represent the sleeping mode. We define sensing action $a(t)$ as

$$a(t) \in \{0 \text{ (sleeping)}, 1, \ldots, N\}. \tag{4}$$

Sensing Observation Suppose that the user has decided to sense channel $a(t) \in \{1, \ldots, N\}$ in this slot. Then, the user observes the occupancy state and the fading condition of this channel (see Section IV-E for implementation details). Combining these two observations, we define sensing outcome $\Theta(t)$ as

$$\Theta(t) \in \{0 \text{ (busy)}, 1, \ldots, L\}, \tag{5}$$

where $\Theta(t) = 0$ indicates that the chosen channel is busy, and $\Theta(t) = k > 0$ indicates that the chosen channel is idle and the fading condition requires the user to transmit at the kth power level.

Given $\mathbf{S}(t) = \mathbf{s} \in \mathcal{S}$, the conditional PMF of sensing outcome $\Theta(t)$ for channel $a(t) > 0$ is given by

$$U_a(k|\mathbf{s}) \stackrel{\Delta}{=} \Pr\{\Theta(t) = k \,|\, \mathbf{S}(t) = \mathbf{s}\} = \begin{cases} p_a(k), & \text{if } S_a(t) = 1, k > 0, \\ 1, & \text{if } S_a(t) = 0, k = 0, \\ 0, & \text{otherwise.} \end{cases} \tag{6}$$

where $p_a(k)$ is determined by channel fading statistics, and is defined in (2).

Access Decision After observing $\Theta(t)$ from the chosen channel, the user determines whether to access. Let $\Phi_k(t)$ denote the access decision given $\Theta(t) = k$:

$$\Phi_k(t) \in \{0 \text{ (no access), } 1 \text{ (access)}\}. \tag{7}$$

Note that to avoid collisions with primary users, the user should refrain from transmission when the channel is sensed as busy: $\Phi_0(t) = 0$ (note that from eq. 6, $U_a(0|0) = 1$). Furthermore, the user should not access when its residual energy is insufficient for accessing the channel in the current fading condition. With the above in mind, we define a set $\varrho(e|a, k)$ of admissible access decisions when the user has residual energy $E(t) = e$ and obtains sensing outcome $\Theta(t) = k$ at channel $a(t) > 0$:

$$\varrho(e|a, k) \stackrel{\Delta}{=} \begin{cases} \{0\}, & \text{if } k = 0 \text{ or } e < e_s + \varepsilon_k, \\ \{0, 1\} & \text{otherwise,} \end{cases} \tag{8}$$

where ε_k is the energy required for a successful transmission under the current sensing outcome $\Theta(t) = k$. Hence, access decisions $\mathbf{\Phi}(t) \stackrel{\Delta}{=} [\Phi_0(t), \Phi_1(t), \ldots, \Phi_L(t)]$ for different sensing outcomes should be chosen from the composite set $\varrho(e|a)$:

$$\varrho(e|a) \stackrel{\Delta}{=} \varrho(e|a, 0) \times \ldots \times \varrho(e|a, L). \tag{9}$$

At the end of the slot, the user updates its statistical knowledge of the SOS by incorporating its decisions and observations in this slot (see Section III-B for details). Depending on its sensing and access decisions, the user's residual energy is reduced from $E(t) = e$ to

$$E(t+1) = \begin{cases} e - e_p, & \text{if } a(t) = 0, \\ e - e_s - \Phi_k(t)\varepsilon_k & \text{if } a(t) > 0 \text{ and } \Theta(t) = k. \end{cases} \tag{10}$$

Note that when $\Theta(t) = 0$, the user should not access ($\Phi_0(t) = 0$); its residual energy is reduced to $E(t+1) = e - e_s$. The updated SOS statistics, together with the reduced residual energy $E(t+1)$, are then used by the user to make optimal decisions in slot $t+1$. The above procedure repeats until the secondary user is incapable of successful transmission under any channel fading conditions, *i.e.,* $E(t) < e_s + \varepsilon_1$.

B. A POMDP Formulation

We show that the sequential decision-making process described above can be formulated as a POMDP. Specifically, the system state is characterized by the SOS of the primary network $\mathbf{S}(t)$ and the residual energy $E(t)$ of the secondary user[3]. While the residual energy is fully observable to the user, the current SOS of the primary network cannot be directly observed due to partial spectrum monitoring. We thus have a POMDP with a random horizon determined by the stopping time $T = \min\{t \geq 0 : E(t) < e_s + \varepsilon_1\}$.

Sufficient Statistics At the beginning of slot t, the user's statistical knowledge of the SOS is provided by its decision and observation history $H(t) \stackrel{\Delta}{=} \{a(\tau), \Theta(\tau)\}_{\tau=1}^{t-1}$. As shown in [14], a sufficient statistic for the SOS is given by a *belief vector* $\mathbf{\Lambda}(t) \stackrel{\Delta}{=} \{\lambda_{\mathbf{s}}(t)\}_{\mathbf{s} \in \mathcal{S}}$ of size 2^N, where each element $\lambda_{\mathbf{s}}(t)$ represents the conditional probability (given the decision and observation history $H(t)$) that the SOS is given by $\mathbf{S}(t) = \mathbf{s}$, *i.e.,*

$$\lambda_{\mathbf{s}}(t) \stackrel{\Delta}{=} \Pr\{\mathbf{S}(t) = \mathbf{s} | H(t)\}. \tag{11}$$

At the beginning of slot $t+1$, the belief vector $\mathbf{\Lambda}(t+1)$ can be obtained from $\mathbf{\Lambda}(t)$ by incorporating the sensing decision $a(t)$ and possibly the observation $\Theta(t)$ in slot t. Specifically, when the user chooses to operate in the sleeping mode ($a(t) = 0$), no observation is made, and the belief vector is updated based solely on the underlying Markovian model of the primary traffic:

$$\mathbf{\Lambda}(t+1) \stackrel{\Delta}{=} \{\lambda_{\mathbf{s}}(t+1)\}_{\mathbf{s} \in \mathcal{S}} = \mathcal{T}(\mathbf{\Lambda}(t)|0), \quad \text{where } \lambda_{\mathbf{s}}(t+1) = \sum_{\mathbf{s}' \in \mathcal{S}} \lambda_{\mathbf{s}'}(t) P_{\mathbf{S}}(\mathbf{s}|\mathbf{s}'). \tag{12}$$

[3]If a Markovian fading model is adopted, the system state should also include the fading conditions $\mathbf{C} \stackrel{\Delta}{=} [C_1(t), \ldots, C_N(t)]$, where $C_n(t)$ represents the current fading condition of channel n. Due to partial spectrum monitoring, fading conditions $\mathbf{C}$ are also partially observable.

When the user chooses a channel $a(t) > 0$ to sense, the belief vector can be updated using Bayes rule based on the sensing outcome $\Theta(t) = k$:

$$\mathbf{\Lambda}(t+1) = \mathcal{T}(\mathbf{\Lambda}(t)|a,k), \quad \text{where } \lambda_{\mathbf{s}}(t+1) = \frac{\lambda_{\mathbf{s}'}(t) P_{\mathbf{S}}(\mathbf{s}|\mathbf{s}') U_a(k|\mathbf{s}')}{\sum_{\mathbf{s}' \in \mathcal{S}} \lambda_{\mathbf{s}'}(t) U_a(k|\mathbf{s}')}. \tag{13}$$

The belief vector $\mathbf{\Lambda}(t)$ together with the residual energy $E(t) = e$ is a sufficient statistic[4] for the system state $(\mathbf{S}(t), E(t))$. That is $(\mathbf{\Lambda}(t), e)$, referred to as the *information state*, is sufficient for making optimal sensing and access decisions. A sensing policy π_s is thus given by a sequence of functions $\pi_s \stackrel{\Delta}{=} [\mu_1, \mu_2, \ldots]$, where μ_t maps an information state $(\mathbf{\Lambda}, e)$ to a sensing decision $a(t) \in \{0, 1, \ldots, N\}$ in slot t. Given sensing policy π_s, an access policy π_c is given by a sequence of functions $\pi_c \stackrel{\Delta}{=} [\nu_1, \nu_2, \ldots]$, where ν_t maps an information state $(\mathbf{\Lambda}, e)$ satisfying $a(t) = \mu_t(\mathbf{\Lambda}, e) > 0$ (*i.e.,* the user operates in the sensing mode) to an admissible access decision $\mathbf{\Phi}(t)$. If functions μ_t (ν_t) are identical for all t, π_s (π_c) is a *stationary policy*.

Reward and Objective A natural definition of the reward is the number of bits delivered by the user in a slot. The immediate reward $R(t)$ can thus be written as

$$R(t) = \begin{cases} 0, & a(t) = 0, \\ g_a(B_a)\Phi_k(t), & a(t) > 0, \Theta(t) = k, \end{cases} \tag{14}$$

where $g_a(\cdot)$ is a given function of the channel bandwidth B_a, determined by the modulation and coding scheme used by the user. For simplicity, we assume $g_a(B_a) = B_a$.

The expected total reward of this POMDP over a random time horizon represents the expected total number of bits delivered by the user during its battery lifetime. The optimal sensing and access policies are thus given by

$$\{\pi_s^*, \pi_c^*\} = \arg\max_{\pi_s, \pi_c} \mathbb{E}\left[\sum_{t=1}^{\infty} R(t) \;\middle|\; \mathbf{\Lambda}(1), E(1) = \overset{\circ}{e} \right], \tag{15}$$

where the initial belief vector $\mathbf{\Lambda}(1)$ can be set to the stationary distribution $\overset{\circ}{\mathbf{\Lambda}}$ of the SOS if no information about the initial state is available.

IV. Optimal Energy-Constrained OSA Design

In this section, we tackle the optimal MAC design for energy-constrained OSA defined in (15). We first show that the optimal sensing and access policies $\{\pi_s^*, \pi_c^*\}$ are stationary and then derive recursive

[4]If a Makovian fading model is adopted, the sufficient statistic consists of three parameters: the belief vector, the residual energy, and the conditional distribution (given the decision and observation history) of the fading conditions $\mathbf{C}$.

formulas for solving (15). We also show the structure of the optimal solution and describe an efficient algorithm for obtaining the optimal decisions. At the end of this section, we discuss the distributed implementation of the optimal MAC design.

A. Stationary Optimal Policy

Stationary policies are usually preferred due to their reduced memory requirements and low complexity in implementation. The fact that the user consumes nonzero energy in each slot and that its battery has finite initial energy implies that the system always reaches a terminating state (*i.e.,* $E(t) < e_s + \varepsilon_1$) in a finite but random time. The inevitable termination makes the energy-constrained OSA design an example of a stochastic shortest path problem, which always has a stationary optimal policy [12].

Proposition 1: For the energy-constrained OSA design given by (15)*, there exist stationary optimal sensing and access policies.*

Proof: See Appendix A. ■

B. Value Function

Proposition 1 allows us to focus on stationary policies without losing optimality. We can thus omit the time index t for notational convenience. The next step to solving (15) is to express the objective explicitly as a function of the information state $(\mathbf{\Lambda}, e)$ and the sensing and access actions $\{a, \mathbf{\Phi}\}$.

Let $Q_a(\mathbf{\Lambda}, e)$ denote the *action-value function* or the Q-function, which represents the maximum expected total reward that can be obtained by taking sensing action $a \in \{0, \ldots, N\}$ in the current slot when the information state is $(\mathbf{\Lambda}, e)$. The *value function*, denoted by $V(\mathbf{\Lambda}, e)$, is the maximum expected total reward that can be accumulated starting from information state $(\mathbf{\Lambda}, e)$. The value function $V(\mathbf{\Lambda}, e)$ and the corresponding optimal sensing action $a^*(\mathbf{\Lambda}, e)$ are given by

$$V(\mathbf{\Lambda}, e) = \max_{a \in \{0,1,\ldots,N\}} Q_a(\mathbf{\Lambda}, e), \qquad a^*(\mathbf{\Lambda}, e) = \arg \max_{a \in \{0,1,\ldots,N\}} Q_a(\mathbf{\Lambda}, e). \tag{16}$$

Since no reward will be earned after the user's residual energy $E(t)$ drops below the minimum energy requirement $e_s + \varepsilon_1$, we have $V(\mathbf{\Lambda}, e) = 0$ for all information states $(\mathbf{\Lambda}, e)$ with $e < e_s + \varepsilon_1$.

Next, we derive iterative formulas for calculating the value function $V(\mathbf{\Lambda}, e)$ and the action-value functions $Q_a(\mathbf{\Lambda}, e)$.

1) Sleeping Mode: In the sleeping mode ($a = 0$), the user consumes e_p energy and no reward will be earned in this slot. The action-value function $Q_0(\mathbf{\Lambda}, e)$ is thus given by the maximum expected remaining reward from the next slot:

$$Q_0(\mathbf{\Lambda}, e) = V(\mathcal{T}(\mathbf{\Lambda}|0), e - e_p), \tag{17}$$

where $\mathcal{T}(\mathbf{\Lambda}|0)$ is the updated belief vector given in (12).

2) Sensing Mode: If the user chooses channel $a > 0$ to sense, it will observe a sensing outcome $\Theta = k$ with probability

$$O_a(k) \stackrel{\Delta}{=} \Pr\{\Theta = k | \mathbf{\Lambda}\} = \sum_{\mathbf{s} \in \mathcal{S}} \lambda_{\mathbf{s}} U_a(k|\mathbf{s}), \tag{18}$$

where $U_a(k|\mathbf{s})$ is the conditional observation probability given in (6).

Given sensing outcome $\Theta = k$ at the chosen channel a, we can calculate the conditional maximum expected reward $Q_a(\mathbf{\Lambda}, e|k, \Phi_k)$ achieved by adopting an admissible access decision Φ_k. Specifically, $Q_a(\mathbf{\Lambda}, e|k, \Phi_k)$ consists of two parts: (i) the immediate reward obtained in this slot, which is given by (14); (ii) the maximum expected remaining reward $V(\mathbf{\Lambda}', e')$ starting from the updated information state $(\mathbf{\Lambda}', e')$, where $\mathbf{\Lambda}' = \mathcal{T}(\mathbf{\Lambda}|a, k)$ given in (13) represents the updated knowledge of the SOS after incorporating sensing action a and observation $\Theta = k$, and $e' = e - e_s - \Phi_k \varepsilon_k$ is the reduced residual energy. We arrive at

$$Q_a(\mathbf{\Lambda}, e|k, \Phi_k) = B_a \Phi_k + V(\mathcal{T}(\mathbf{\Lambda}|a, k), e - e_s - \Phi_k \varepsilon_k). \tag{19}$$

Optimizing over all admissible access decisions Φ_k and then averaging over sensing outcomes $\Theta = k$, we obtain the maximum expected reward achieved by choosing channel $a > 0$ and the corresponding optimal access decision $\Phi_k^*(\mathbf{\Lambda}, e|a)$ as

$$Q_a(\mathbf{\Lambda}, e) = \sum_{k=0}^{L} O_a(k) \max_{\Phi_k \in \varrho(e|a,k)} Q_a(\mathbf{\Lambda}, e|k, \Phi_k), \qquad \Phi_k^*(\mathbf{\Lambda}, e|a) = \arg \max_{\Phi_k \in \varrho(e|a,k)} Q_a(\mathbf{\Lambda}, e|k, \Phi_k). \tag{20}$$

C. Solution Structure

We note that obtaining the optimal sensing and access decisions hinges on the computation of the action-value and the value functions. We thus seek structures of the value function that lead to efficient computation of the optimal decisions.

1) Reduced Dimension: One of the difficulties in calculating the value function $V(\mathbf{\Lambda}, e)$ is that the dimension of the belief vector $\mathbf{\Lambda}$ grows exponentially with the number N of channels. It has been shown in [3] that for independently evolving channels, an alternative sufficient statistic for the SOS is given by a the marginal distribution $\mathbf{\Omega}(t) \stackrel{\Delta}{=} [\omega_1(t), \ldots, \omega_N(t)]$ of the SOS, where $\omega_n(t)$ denotes the probability (conditioned on the entire decision and observation history $H(t)$) that channel n is idle at the beginning of slot t:

$$\omega_n(t) \stackrel{\Delta}{=} \Pr\{S_n(t) = 1 | H(t)\}. \tag{21}$$

Let $\boldsymbol{\alpha}\triangleq[\alpha_1,\ldots,\alpha_N]$ and $\boldsymbol{\beta}\triangleq[\beta_1,\ldots,\beta_N]$ denote the transition probabilities of channel n, where $\alpha_n\triangleq\Pr\{S_n(t)=1|S_n(t-1)=0\}$ and $\beta_n\triangleq\Pr\{S_n(t)=1|S_n(t-1)=1\}$. We can then obtain the belief updates similar to (12) and (13). Specifically, when the user operates in the sleeping mode, we have

$$\boldsymbol{\Omega}(t+1)=\hat{\mathcal{T}}(\boldsymbol{\Omega}(t)|0), \quad \text{where } \omega_n(t+1)=\alpha_n+(\beta_n-\alpha_n)\omega_n(t). \tag{22}$$

When the user chooses channel $a(t)>0$, then the belief vector $\boldsymbol{\Omega}(t+1)$ is updated according to the sensing outcome $\Theta(t)=k$:

$$\boldsymbol{\Omega}(t+1)=\hat{\mathcal{T}}(\boldsymbol{\Omega}(t)|a,k), \quad \text{where } \omega_n(t+1)=\begin{cases}\alpha_n+(\beta_n-\alpha_n)\omega_n(t) & \text{if } n\neq a(t),\\ \beta, & \text{if } n=a(t),k>0,\\ \alpha, & \text{if } n=a(t),k=0.\end{cases} \tag{23}$$

Following Section IV-B, we can also develop a simpler recursion for the value function $V(\boldsymbol{\Omega},e)$:

$$V(\boldsymbol{\Omega},e)=\max_{a\in\{0,1,\ldots,N\}} Q_a(\boldsymbol{\Omega},e), \tag{24}$$

where

$$Q_0(\boldsymbol{\Omega},e)=V(\hat{\mathcal{T}}(\boldsymbol{\Omega}),e-e_p), \tag{25a}$$

$$Q_a(\boldsymbol{\Omega},e)=(1-\omega_a)V(\hat{\mathcal{T}}(\boldsymbol{\Omega}|a,0),e-e_s)+\omega_a\sum_{k=1}^{L}p_a(k)\max_{\Phi_k\in\varrho(e|a,k)}Q_a(\boldsymbol{\Omega},e|k,\Phi_k), \quad a>0 \tag{25b}$$

$$Q_a(\boldsymbol{\Omega},e|k,\Phi_k)=B_a\Phi_k+V(\hat{\mathcal{T}}(\boldsymbol{\Omega}|a,k),e-e_s-\Phi_k\varepsilon_k), \quad a>0. \tag{25c}$$

Compared with the original value function $V(\boldsymbol{\Lambda},e)$ developed in Section IV-B, the above value function $V(\boldsymbol{\Omega},e)$ not only has simpler belief updates $\hat{\mathcal{T}}$ but also avoids computation of the summation in (18).

2) Monotonicity: Monotonicity results for the value function are usually desired since they not only provides insights into the underlying problem but also serves as a stepping stone for establishing the structure of optimal policies (see [13] for an example). In Proposition 2, we study the monotonicity of the value function with respect to each of its parameters.

Proposition 2: **Monotonicity of Value Function**

P2.1 *The value function $V(\boldsymbol{\Lambda},e)$ is monotonically increasing with the residual energy $e\in\mathcal{E}$, i.e., $V(\boldsymbol{\Lambda},e)\geq V(\boldsymbol{\Lambda},e')$ for $e\geq e'$.*

P2.2 *Assume that the SOS evolves independently across channels. If $\boldsymbol{\beta}\geq\boldsymbol{\alpha}$, then the value function $V(\boldsymbol{\Omega},e)$ given in* (24) *is monotonically increasing with the belief vector $\boldsymbol{\Omega}$, i.e., $V(\boldsymbol{\Omega},e)\geq V(\boldsymbol{\Omega}',e)$ for $\boldsymbol{\Omega}\geq\boldsymbol{\Omega}'$.*

Proof: See Appendix B. ■

P2.1 is straightforward. P2.2 considers the case where the SOS evolves independently across channels. It provides a sufficient condition for the value function $V(\mathbf{\Omega}, e)$ to be monotonically increasing with the belief vector $\mathbf{\Omega}$ defined in (21). Note that $\boldsymbol{\beta} \geq \boldsymbol{\alpha}$ represents the case where the channel occupancy state is positively correlated across time. In this case, a larger current belief vector $\mathbf{\Omega}(t)$ indicates a larger probability that channels will be idle in all the future slots, leading to a higher chance of getting rewards. When $\boldsymbol{\beta} < \boldsymbol{\alpha}$, the channel occupancy state is negatively correlated across time. The value function is not necessarily monotonic. This is because when $\boldsymbol{\beta} < \boldsymbol{\alpha}$, a larger belief vector $\mathbf{\Omega}(t)$ indicates a smaller probability that channels are idle in the next slot. The probabilities of channels being idle oscillates over time.

3) Piecewise Linearity and Convexity: It has been shown in [14] that the value function for a POMDP over a finite and fixed time horizon is piecewise linear and convex with respect to the belief vector. In Proposition 3, we show that the value function $V(\mathbf{\Lambda}, e)$ for a POMDP over a finite but random time horizon also has this property.

Proposition 3: **Piecewise Linear and Convex Value Function**

The value function $V(\mathbf{\Lambda}, e)$ given in (16) is piecewise linear and convex with respect to the belief vector $\mathbf{\Lambda} \in \Pi$. That is, for a given residual energy $e \in \mathcal{E}$, the value function $V(\mathbf{\Lambda}, e)$ can be written as

$$V(\mathbf{\Lambda}, e) = \max_{\mathbf{\Upsilon} \in \mathbf{\Gamma}_e} < \mathbf{\Lambda}, \mathbf{\Upsilon} >, \tag{26}$$

where $< \cdot, \cdot >$ denotes inner product, $\mathbf{\Upsilon}$ is a vector of size $|\mathcal{S}| = 2^N$, and $\mathbf{\Gamma}_e$ is a finite set of such vectors $\mathbf{\Upsilon}$.

Proof: The proof proceeds by mathematical induction on the residual energy e. See Appendix C. ■

As illustrated in Fig. 1, Proposition 3 shows that the domain of the value function $V(\mathbf{\Lambda}, e)$ can be partitioned into a finite number of convex regions, each of which is associated with an Υ-vector $\mathbf{\Upsilon}_e^i \in \mathbf{\Gamma}_e$. The value function of a certain belief vector $\mathbf{\Lambda}$ is simply given by the inner product of this belief vector and the Υ-vector associated with the region where $\mathbf{\Lambda}$ lies. For the example in Fig. 1, the value function of $\mathbf{\Lambda}(t)$ is given by $V(\mathbf{\Lambda}(t), e) =< \mathbf{\Lambda}(t), \mathbf{\Upsilon}_e^2 >$. Hence, calculating the value function over the entire continuous belief space is equivalent to finding a finite set $\mathbf{\Gamma}_e$ of Υ-vectors. Readers are referred to [15]–[19] for different dynamic programming algorithms for constructing Υ-vectors.

D. A Solution Procedure

At the beginning, the secondary user may not have any information about the SOS other than its transmission probabilities $P_{\mathbf{S}}$. Hence, the initial belief vector $\mathbf{\Lambda}(1)$ is usually set to the stationary

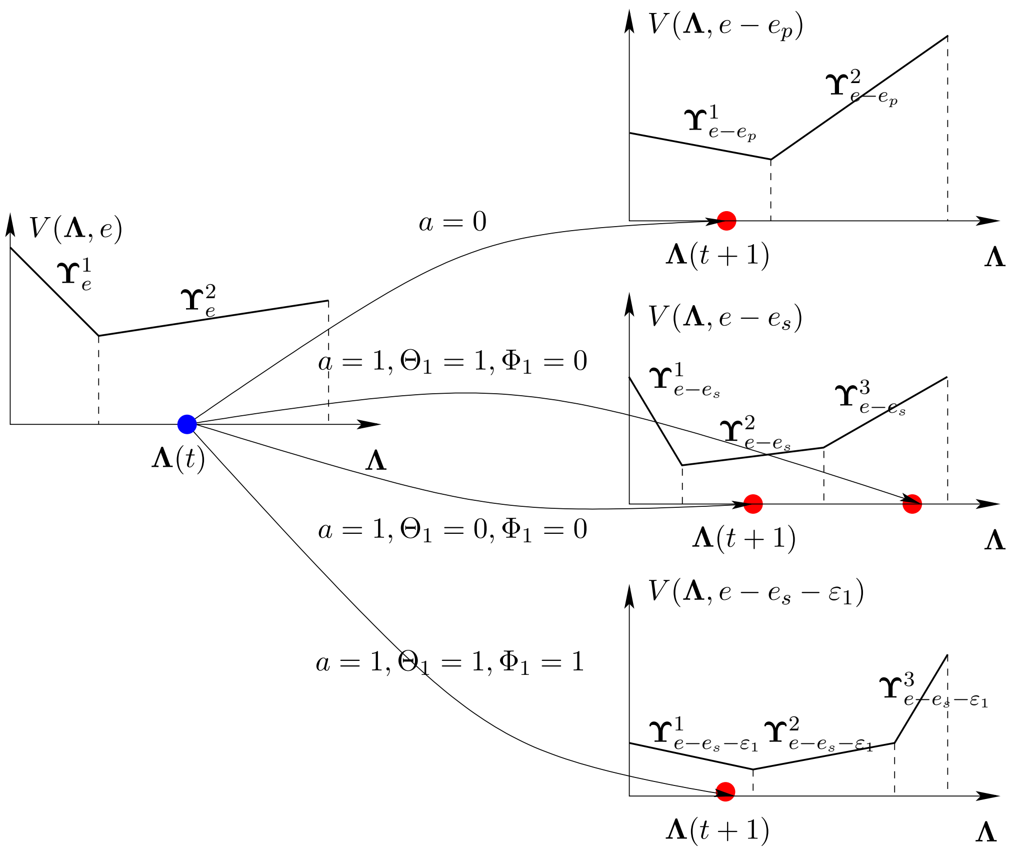

Fig. 1. An illustration of the structure of the value function $V(\mathbf{\Lambda}, e)$. Each point on the x-axis represents a possible belief vector $\mathbf{\Lambda}$. Each arrow represents the update of an information state $(\mathbf{\Lambda}, e)$ under certain decisions and observations.

distribution $\overset{\circ}{\mathbf{\Lambda}}$ of the SOS. We note that given an initial belief vector and an initial energy, the secondary user can only experience a finite number of possible information states $(\mathbf{\Lambda}, e)$ during its battery lifetime. This is due to the fact that a belief vector $\mathbf{\Lambda}(t)$ in a slot can only transit to a finite number of possible belief vectors $\mathbf{\Lambda}(t+1)$ in the next slot (see Fig. 1), and that the POMDP given in (15) terminates in a finite time (see Section IV-A). The above observation suggests that to obtain optimal sensing and access decisions for a given initial information state, we only need to calculate the value function for a finite number of possible information states.

Also note that due to energy consumption in sleeping and sensing, the user's residual energy decreases after each slot. Hence, the value function and the action-value function of an information state $(\mathbf{\Lambda}, e)$ only depend on those with less residual energies. We can thus compute the value function in an increasing order of the residual energy $e \in \mathcal{E}$, which leads to the following algorithm for computing the optimal sensing and access decisions.

Algorithm for computing optimal sensing and access decisions

S0: According to the initial belief vector $\mathbf{\Lambda}(1)$ and the initial battery energy $\overset{\circ}{e}$, enumerate all possible information states $(\mathbf{\Lambda}, e)$ that the user may experience during its battery lifetime. Let $\mathcal{U}$ include all such $(\mathbf{\Lambda}, e)$ with $e \geq e_s + \varepsilon_1$.

S1: Let $V(\mathbf{\Lambda}, e) = 0$ for all $(\mathbf{\Lambda}, e)$ with $e \in \mathcal{E}$ and $e < e_s + \varepsilon_1$.

S2: Use (16), (17), (19), and (20) to calculate the value function for the information state $(\mathbf{\Lambda}, e) \in \mathcal{U}$ satisfying $e \leq e'$ for all $(\mathbf{\Lambda}', e') \in \mathcal{U}$.

S3: Remove $(\mathbf{\Lambda}, e)$ from set $\mathcal{U}$: *i.e.,* $\mathcal{U} = \mathcal{U} \backslash (\mathbf{\Lambda}, e)$. If $\mathcal{U}$ is non-empty, then goto S2. Otherwise, stop the calculation.

We point out that the optimal sensing and access decisions for all possible information states can be pre-computed and stored by each user before it operates. At the beginning of each slot, the user simply looks up the optimal decisions using its current information state $(\mathbf{\Lambda}, e)$. Hence, the proposed optimal OSA design does not impose any computational burden on the user.

E. Distributed Implementation

Next, we show that the optimal energy-constrained OSA strategy obtained under the POMDP framework can be implemented in a distribution fashion.

1) Channel State Acquisition: Suppose that the transmitter and the receiver hop to the same channel at the beginning of a slot. If the channel is sensed as idle, the transmitter adopts carrier sensing (*i.e.,* wait for a random backoff time before transmission attempts) to avoid collisions among competing secondary users. If the channel remains idle when its backoff time expires, it transmits a short request-to-send (RTS) message at full power[5] to the receiver. Upon receiving the RTS, the receiver estimates the channel fading condition using the RTS, and then replies with a clear-to-send (CTS) message which contains the estimated channel fading condition. After a successful exchange of RTS-CTS, the transmitter can adjust its transmission power according to the channel fading condition and communicate with the receiver over this channel.

2) Synchronous Hopping: Suppose that the transmitter and the receiver have tuned to the same channel after the initial handshake (one scheme for initial handshake can be found in [3]). To ensure synchronous hopping in the spectrum afterwards without extra control message exchange, the receiver must be aware of the transmitter's sensing decisions at the beginning of each slot. For this purpose, the transmitter and the receiver must maintain the same information state $(\mathbf{\Lambda}, e)$ in each slot.

[5]Note that the energy consumed in channel state estimation is absorbed into the sensing energy consumption e_s.

We point out that when both the transmitter and the receiver can observe the true state of the sensed channel, they will have the same update of the belief vector and the residual energy, thus reaching the same information state. When the transmitter and the receiver are affected by different sets of primary users or when sensing errors cannot be ignored, the exchange of RTS-CTS for fading state acquisition can be exploited to ensure synchronous hopping between the transmitter and the receiver. In this case, the common observation used for updating the information state is whether there is a successful exchange of RTS-CTS. A similar discussion on using common observations to ensure synchronous hopping can be found in [4].

V. Threshold Structures of Optimal Policies

In this section, we study different factors that affect the optimal decisions obtained in Section IV-B. We focus on the operating decision (sleeping vs. sensing) and the access decision, which are unique to the energy-constrained OSA problem.

Careful inspection of (17) and (19) reveals that the user's decision affects the total expected reward in three ways: (i) it may acquire an immediate reward B_a; (ii) it transforms the current belief vector $\mathbf{\Lambda}(t)$ to $\mathbf{\Lambda}(t+1) = \mathcal{T}(\mathbf{\Lambda}|0)$ or $\mathcal{T}(\mathbf{\Lambda}|a,k)$ which summarizes the information of the SOS up to this slot; (iii) it causes a reduction in battery energy. Hence, to maximize the total expected reward during the battery lifetime, optimal decisions should be made to achieve a tradeoff among gaining instantaneous reward, gaining information for future use, and conserving energy.

A. To Sense or Not to Sense?

Without the energy constraint, the user should always operate in a sensing mode since sensing provides not only a chance to gain immediate access but also statistical information about the SOS. With the energy constraint, however, the user may choose to sleep since sensing costs energy. In this case, the optimal operating decision should strike a balance between gaining reward/information and conserving energy.

1) Analytical Study: We first provide a sufficient condition for the user to operate in the sensing mode.

Proposition 4: When the secondary user's belief vector is given by the stationary distribution $\mathring{\mathbf{\Lambda}}$ of the underlying SOS, its optimal operating mode is to sense, *i.e.,* $a^*(\mathbf{\Lambda}, e) > 0$ if $\mathbf{\Lambda} = \mathring{\mathbf{\Lambda}}$.

Proof: See Appendix D. ■

The intuition behind Proposition 4 is explained as follows. Suppose that the secondary user chooses to operate in the sleeping mode when its belief vector is given by the stationary distribution of the SOS. Then, it will have the same belief vector but reduced residual energy at the beginning of the next slot.

The energy consumed in sleeping is thus wasted without gaining any statistical information about the SOS. This suggests that the optimal operating mode is to sense.

Next, we consider the single-channel ($N = 1$) case, where the belief vector $\mathbf{\Lambda}$ can be characterized by a scalar ω as defined in (21), and the transition probabilities of this channel can be denoted by $\beta \stackrel{\Delta}{=} \Pr\{S(t+1) = 1 | S(t) = 1\}$ and $\alpha \stackrel{\Delta}{=} \Pr\{S(t+1) = 1 | S(t) = 0\}$.

Proposition 5: **Threshold Optimal Sensing Decision**

Consider the single-channel ($N = 1$) case. For any given residual energy e, the optimal sensing decision has a threshold structure:

$$a^*(\omega, e) = \begin{cases} 1 \textit{ (sense)} & \textit{if } \omega \geq r_{th}(e) \\ 0 \textit{ (sleep)} & \textit{otherwise}, \end{cases} \tag{27}$$

where $r_{th}(e) \in [\min\{\alpha, \beta\}, \frac{\alpha}{1+\alpha-\beta}]$ is the optimal sensing threshold.

Proof: See Appendix D. ■

Proposition 5 states that the user should sense when the belief ω of the channel is large and should sleep when the channel is less likely to be idle. This agrees with our intuition.

Corollary 1: Consider the single-channel ($N = 1$) case. When $\alpha = \beta$, the secondary user should always operate in the sensing mode, i.e., $a^*(\omega, e) = 1$.

Proof: See Appendix D. ■

2) Numerical Study: As indicated by Proposition 5, the user's residual energy e affects the optimal operating decision through sensing threshold $r_{th}(e)$. To study the role of the residual energy e in choosing operate modes, we plot the optimal sensing threshold $r_{th}(e)$ in Fig. 2 for different sensing energy consumption e_s and channel occupancy statistics $\{\alpha, \beta\}$.

We find that the optimal sensing threshold $r_{th}(e)$ is highly dependent on the user's residual energy e when e is small. As e increases, the impact of the residual energy e on the user's operating decision $r_{th}(e)$ diminishes. When e is sufficiently large, the optimal sensing threshold $r_{th}(e)$ becomes independent of the residual energy e. This observation implies that when the battery is depleting, the user should focus more on how to fully utilize its residual energy.

We also see that the optimal sensing threshold $r_{th}(e)$ fluctuates more dramatically when the channel occupancy state is negatively correlated (*i.e.,* $\beta < \alpha$). That is, in this case, the residual energy plays a more important role in decision-making. As explained below Proposition 2, the probability that the channel is idle fluctuates when $\beta < \alpha$. Hence, the user should focus more on its residual energy in this case to save energy for those slots when the channel is more likely to be idle.

Furthermore, we find that the optimal sensing threshold $r_{th}(e)$ increases with the sensing energy consumption e_s. That is, the user should be more conservative in making operating decisions when e_s is large. This observation agrees with our expectation because when e_s is large, the extra energy consumed in sensing can only be paid off when the chance of gaining immediate access is higher. On the other hand, when e_s is comparable to e_p, the user can afford sensing more often to gain statistical information.

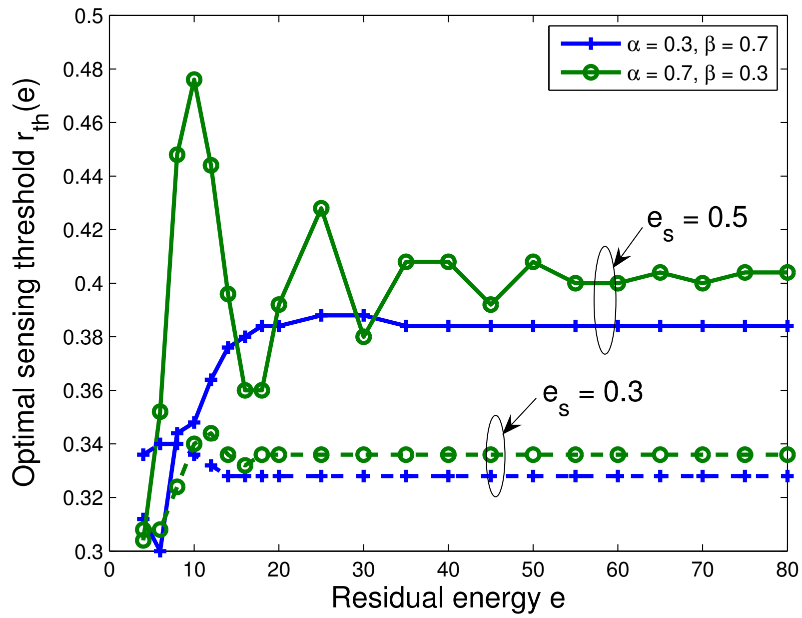

Fig. 2. Optimal sensing threshold $r_{th}(e)$. $B = 1$ (bandwidth), $e_s = 0.3, 0.5$ (sensing energy), $e_p = 0.1$ (sleeping energy), $E_{tx} \in \{1, 2, 3, 4\}$ (transmission energy), $N = 1$ (number of channels), $[p(1), p(2), p(3), p(4)] = [0.2, 0.3, 0.3, 0.2]$ (channel fading statistics).

B. To Access or Not to Access?

Without the energy constraint, the user should always access an idle channel. With the energy constraint, however, the access decision should take into account both the energy consumption characteristics and the channel fading statistics. For example, when the channel is idle but has poor fading condition, should the user access this channel to gain immediate reward or wait for better channel realizations for less transmission energy? We find that such a decision is a monotonic function of the channel fading condition.

1) Analytical Study:

Proposition 6: Given that a channel $a \in \{1, \ldots, N\}$ is sensed, the optimal access decision is monotonically increasing with the channel fading condition. Specifically, for any given residual energy $e \geq e_s + \varepsilon_1$, the optimal access decision is given by

$$\Phi_k^*(\mathbf{\Lambda}, e|a) = \begin{cases} 1 & \textit{if } 0 < k \leq k_{th}(\mathbf{\Lambda}, e) \\ 0 & \textit{otherwise}, \end{cases} \tag{28}$$

where $k_{th}(\mathbf{\Lambda}, e) \in \{1, \ldots, L\}$ is the optimal access threshold. Furthermore, when $N = 1$ (i.e., the single-channel case), the threshold $k_{th}(\mathbf{\Lambda}, e) = k_{th}(e)$ is independent of the belief vector $\mathbf{\Lambda}$.

Proof: See Appendix E. ■

Note that the better the channel fading condition, the lower the sensing outcome. Proposition 6 indicates that the user should access when the channel is in good condition and not access when the channel experiences deep fading. In particular, when the sensed channel is in the best fading condition (*i.e.,* $\Theta(t) = 1$), then the user should always access, *i.e.,* $\Phi_1^*(\mathbf{\Lambda}, e|a) = 1$, for any $e \geq e_s + \varepsilon_1$.

Proposition 6 also helps us reduce the size of the access decision space $\varrho(e|a)$ from exponential $\mathcal{O}(2^L)$ to linear $\mathcal{O}(L)$ with respect to the number L of power levels, leading to a more efficient search for the optimal access policy. Specifically, we can restrict our search for the optimal access decision to the following set:

$$\varrho'(e|a) = \{[\Phi_0, \Phi_1, \ldots, \Phi_L] : \Phi_k \in \varrho(e|a, k), \Phi_1 \geq \ldots \geq \Phi_L\}, \tag{29}$$

where the size of $\varrho'(e|a)$ is on the order of L.

2) Numerical Study: For simplicity, we consider the single-channel case ($N = 1$) in the numerical study. As shown in Proposition 6, the optimal access threshold $k_{th}(\mathbf{\Lambda}, e)$ in this case reduces to $k_{th}(e)$, which is independent of the belief vector $\mathbf{\Lambda}$. In Fig. 3, we plot the optimal access threshold $k_{th}(e)$ as a function of the residual energy e for different sensing energy consumption e_s.

Similar to the behavior of the optimal sensing threshold $r_{th}(e)$, the optimal access threshold $k_{th}(e)$ may differ from each other when e is small but a common steady value is reached when e is sufficiently large. That is, the impact of e on optimal access decisions diminishes.

We further see that the optimal access threshold $k_{th}(e)$ increases with the sensing energy consumption e_s. Hence, when e_s is small, the user should refrain from transmission under poor channel conditions and wait for better channel realization. On the other hand, when e_s is large, the user should be more aggressive in making access decisions: it should grab the instantaneous opportunity even when the channel is in a deep fade. This is because when e_s is large, the sensing energy consumed in waiting for the best channel realization may exceed the extra energy consumed in combating the poor channel fading.

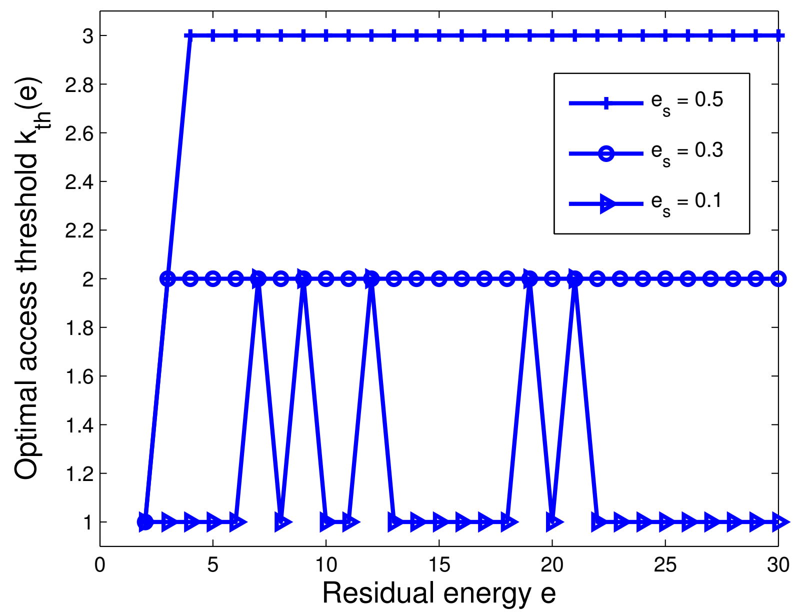

Fig. 3. Optimal access thresholds $k_{th}(\mathbf{\Lambda}, e)$. $B = 1$ (bandwidth), $\beta = 0.7$, $\alpha = 0.3$ (transition probabilities), $e_p = 0.1$ (sleeping energy), $E_{tx} \in \{1, 2, 3, 4\}$ (transmission energy), $[p(1), p(2), p(3), p(4)] = [0.2, 0.3, 0.3, 0.2]$ (channel fading statistics).

TABLE I

A SAMPLE PATH OF THE SOS EVOLUTION AND THE CORRESPONDING OPTIMAL SENSING AND ACCESS DECISIONS. ($e_p = 0.1$, $e_s = 0.5$, $E_{tx} \in \{1, 2, 3, 4\}$, $[p_1(1), p_1(2), p_1(3), p_1(4)] = [0.4, 0.3, 0.2, 0.1]$, $[p_2(1), p_2(2), p_2(3), p_2(4)] = [0.2, 0.3, 0.4, 0.1]$.)

Time t	1	2	3	4	5	6	7	8	9
SOS $\mathbf{S}(t)$	[1, 0]	[0, 1]	[1, 0]	[1, 1]	[0, 0]	[1, 0]	[0, 1]	[1, 0]	[0, 1]
Belief vector $\mathbf{\Omega}(t)$	[0.5,0.5]	[0.3,0.5]	[0.58,0.5]	[0.3,0.5]	[0.58,0.5]	[0.7,0.5]	[0.3,0.5]	[0.58,0.5]	[0.3,0.5]
Residual energy $E(t)$	8	6.5	6.4	3.9	3.8	3.3	2.8	2.7	0.2
Sensing decision $a(t)$	1	0	1	0	1	1	0	1	-
Sensing outcome $\Theta(t)$	1	-	2	-	0	3	-	2	-
Access decision $\Phi(t)$	1	-	1	-	0	0	-	1	-

C. A Sample Path

To further illustrate the behavior of the optimal sensing and access policies, we study an example of the SOS evolution and the corresponding optimal decisions. In Table I, we consider $N = 2$ independent channels with identical transition probabilities ($\alpha = 0.7$, $\beta = 0.3$) but different channel fading statistics. At the beginning of the first slot, the user operates in the sensing mode since its belief vector is given by the stationary distribution of the SOS. This agrees with Proposition 4. We find that to conserve energy, the user never chooses the channel (*i.e.,* Channel 2) in deep fading even if there is a higher probability that this channel is idle. This demonstrates the important role of channel fading statistics in deciding whether to sense. We also see that the exploitation of channel occupancy dynamics allows the user to efficiently track spectrum opportunities. Specifically, when the channel is less likely to be idle, the user operates in the sleeping mode to save energy (see $t = 2, 4, 7$). It wakes up when the probability that the channel is idle is large.

VI. Bursty Traffic in Energy-Constrained OSA

In this section, we address the optimal distributed MAC design for energy-constrained OSA when the secondary user has bursty traffic. We show that in this case, the optimal sensing and access decisions should also take into account the traffic dynamics of the secondary user. We illustrate the impact of the secondary user's buffer state on the optimal operating decision.

A. Bursty Traffic Model

We assume that the packet arrival process is i.i.d. across slots, for example, the Poisson packet arrival process. Let q_m, $m = 0, 1, \ldots$, denote the probability that m packets arrive in a slot. We assume that the user has a finite buffer with maximum size M. It receives packets in every slot even if it operates in the sleeping mode. Packets are dropped when its buffer overflows. Let $D(t)$ denote the number of packets in the user's buffer at the beginning of slot t. Depending on the packet arrivals and departures, the buffer state $D(t)$ follows a Markov chain with state space $\{0 \text{ (empty)}, 1, \ldots, M\}$ and transition probabilities given by

$$\begin{aligned} P_D(d'|d, i) &\stackrel{\Delta}{=} \Pr\{D(t+1) = d'|D(t) = d, i \text{ packets were sent in slot } t\} \\ &= \sum_{m=0}^{\infty} q_m \mathbb{1}_{[d'=\min\{d-i+m, M\}]}, \quad d, d' \in \{0, 1, \ldots, M\}. \end{aligned} \tag{30}$$

We assume that the transmission time of a packet over a channel with unit bandwidth is equal to the slot length. Hence, the number i of packets transmitted over channel a in a slot is either 0 or B_a.

B. POMDP Formulation

The POMDP framework developed in Section III-B for energy-constrained OSA design in the continuous traffic case can be extended to the bursty traffic case. Specifically, the new system is characterized by the following three components: (i) the primary network's SOS $\mathbf{S}(t)$; (ii) the secondary user's residual energy $E(t)$; (iii) the secondary user's buffer size $D(t)$.

Sufficient Statistics As explained in Section IV-E, to ensure synchronous hopping in the spectrum without extra control messages, the user (*i.e.,* transmitter) and its receiver must use a common knowledge of the system state for decision-making in each slot. We note that while the user and its receiver can maintain the same belief vector $\mathbf{\Lambda}(t)$ and residual energy $E(t) = e$, the receiver does not know the exact buffer state $D(t)$ until notified by the user during the exchange of RTS-CTS[6]. Hence, when making sensing decisions (which occurs before the exchange of RTS-CTS), both the user and its receiver should treat the buffer state $D(t)$ as a partially observable parameter and use statistical information about $D(t)$. On the other hand, since both the user and its receiver know the exact buffer state after a successful exchange of RTS-CTS, access decisions should be made by taking into account the exact buffer state $D(t)$. Let $\Phi_{k,d}(t) \in \varrho(e|a, k, d)$ denote an admissible access decision under sensing outcome $\Theta(t) = k$ and buffer state $D(t) = d$, where

$$\varrho(e|a,k,d) \stackrel{\Delta}{=} \begin{cases} \{0\}, & \text{if } k = 0 \text{ or } e < e_s + \varepsilon_k \text{ or } d = 0, \\ \{0,1\} & \text{otherwise.} \end{cases} \tag{31}$$

The statistical information about the buffer state $D(t)$ can be summarized by a conditional PMF $\mathbf{\Psi}(t) \stackrel{\Delta}{=} \{\psi_d(t)\}_{d=0}^M$, where $\psi_d(t) \in [0, 1]$ and $\sum_{d=0}^M \psi_d(t) = 1$. Each element $\psi_d(t)$ denotes the conditional probability (given the user's notifications of the buffer state) that the user's buffer state is $D(t) = d$ at the beginning of slot t. When the user operates in the sleeping mode (*i.e.,* $a(t) = 0$) or the chosen channel is sensed as busy ($a(t) > 0, \Theta(t) = 0$), the user is unable to inform the receiver of its buffer state. Hence, the statistical information $\mathbf{\Psi}(t)$ of $D(t)$ is updated at both the user and its receiver based solely on the packet arrival process:

$$\mathbf{\Psi}(t+1) \stackrel{\Delta}{=} \{\psi_d(t+1)\}_{d=0}^M = \mathcal{F}(\mathbf{\Psi}(t)|0), \quad \text{where } \psi_d(t+1) = \sum_{d'=0}^{M} \psi_{d'}(t) P_D(d|d', 0). \tag{32}$$

When a channel $a(t) > 0$ is sensed as idle, the receiver knows the buffer state $D(t) = d$ from the user's RTS message. The statistical information about the buffer state can thus be updated based on the user's sensing decision $a(t)$ and access decision $\Phi_{k,d}(t) \in \varrho(e|a, k, d)$:

$$\mathbf{\Psi}(t+1) = \mathcal{F}(\mathbf{\Psi}(t)|a, k, d, \Phi_{k,d}), \quad \text{where } \psi_{d'}(t+1) = P_D(d'|d, \Phi_{k,d} B_a). \tag{33}$$

Based on the above discussion, we see that in the bursty traffic case, the information state used for sensing and access decision-making consists of the belief vector Λ, the residual energy $E(t) = e$, and

[6]The secondary user can piggyback its buffer state $D(t)$ to the RTS message.

the statistical information $\mathbf{\Psi}(t)$ on the buffer state. The design objective is thus given by

$$\{\pi_s^*, \pi_c^*\} = \arg\max_{\pi_s, \pi_c} \mathbb{E}\left[\sum_{t=1}^{\infty} R(t) \,\middle|\, \mathbf{\Lambda}(1), E(1) = \overset{\circ}{e}, \mathbf{\Psi}(1)\right]. \tag{34}$$

C. Optimal Solution

We derive here the value function $V(\mathbf{\Lambda}, e, \mathbf{\Psi})$ and the action-value function $Q_a(\mathbf{\Lambda}, e, \mathbf{\Psi})$ for the POMDP given in (34). Following Section IV, we can readily obtain the maximum expected reward that can be achieved in the sleeping mode as

$$Q_0(\mathbf{\Lambda}, e, \mathbf{\Psi}) = V(\mathcal{T}(\mathbf{\Lambda}), e - e_p, \mathcal{F}(\mathbf{\Psi}|0)), \tag{35}$$

where $\mathcal{F}(\mathbf{\Psi}|0)$ is the updated knowledge of the buffer state given in (32).

Next, we derive the maximum expected reward $Q_a(\mathbf{\Lambda}, e, \mathbf{\Psi})$ that can be achieved in the sensing mode. Consider the scenario where the secondary user chooses access decision $\Phi_{k,d} \in \varrho(e|a, k, d)$ under sensing outcome $\Theta = k$ when its buffer state is $D = d$. In this case, the maximum expected reward can be calculated as

$$Q_a(\mathbf{\Lambda}, e, \mathbf{\Psi}|k, d, \Phi_{k,d}) = B_a \Phi_k + V(\mathcal{T}(\mathbf{\Lambda}|a, k), e - e_s - \Phi_{k,d}\varepsilon_k, \mathcal{F}(\mathbf{\Psi}|a, k, d, \Phi_{k,d})). \tag{36}$$

Optimizing over all admissible access decisions $\Phi_{k,d} \in \varrho(e|a, k, d)$ and then averaging over all sensing outcomes $\Theta = k$ with (18) and all buffer states $D = d$ with current statistical information $\mathbf{\Psi}$, we obtain that

$$Q_a(\mathbf{\Lambda}, e, \mathbf{\Psi}) = \sum_{d=0}^{M} \Psi_d \sum_{k=0}^{L} O_a(k) \max_{\Phi_{k,d} \in \varrho(e|a,k,d)} Q_a(\mathbf{\Lambda}, e, \mathbf{\Psi}|k, d, \Phi_{k,d}). \tag{37}$$

With (35), (37), (36), the value function can be obtained as

$$V(\mathbf{\Lambda}, e, \mathbf{\Psi}) = \max_{a \in \{0,\ldots,N\}} Q_a(\mathbf{\Lambda}, e, \mathbf{\Psi}). \tag{38}$$

We can readily generalize the solution procedure described in Section IV-D and calculate the above value function in an increasing order of the residual energy e starting from $e < e_s + \varepsilon_1$. After computing the value function, we can obtain the optimal sensing and access decisions as

$$a^*(\mathbf{\Lambda}, e, \mathbf{\Psi}) = \arg\max_{a \in \{0,\ldots,N\}} Q_a(\mathbf{\Lambda}, e, \mathbf{\Psi}), \qquad \Phi_{k,d}^*(\mathbf{\Lambda}, e, \mathbf{\Psi}|a) = \arg\max_{\Phi_{k,d} \in \varrho(e|a,k,d)} Q_a(\mathbf{\Lambda}, e, \mathbf{\Psi}|k, d, \Phi_{k,d}). \tag{39}$$

We point out that the structures of the value function developed in Section IV-C and the threshold structure of the optimal access policy developed in Proposition 6 hold for the bursty traffic case. The structural results for the optimal sensing policy (*i.e.,* Proposition 4-5 and Corollary 1), however, do not hold since the optimal operating decision in the bursty traffic case is highly dependent of the user's buffer state. For example, we find that in the bursty traffic case, the user may choose to sleep even if its belief vector is given by the stationary distribution of the underlying SOS (contrary to Proposition 4). This happens when the probability that the user has packets to transmit is small. To avoid wasting sensing energy, the user and its receiver should wait until the buffer is more likely to be non-empty.

D. Numerical Study: Optimal Operating Decision for Empty Buffer

It is interesting to note that even if the buffer is empty, the user may want to sense a channel in order to gain information about the SOS for future use, especially when the SOS is highly correlated in time. We study below the optimal operating decision (sleeping vs. sensing) in the empty buffer case[7].

Consider two coupled channels $N = 2$ where the SOS is either $\mathbf{S}(t) = [0, 1]$ (*i.e.,* only channel 2 is idle) or $\mathbf{S}(t) = [1, 0]$ (*i.e.,* only channel 1 is idle). We assume that $P_S([1, 0] \,|\, [0, 1]) = P_S([0, 1] \,|\, [1, 0]) = \varpi$, *i.e.,* the channel occupancy state changes with probability ϖ in each slot. In this case, the correlation between the SOS in two successive slots can be characterized by a single parameter $\eta = 1-2\varpi$. Extensive numerical results show that the optimal operating decision is a monotonically increasing with the SOS time correlation $|\eta|$. Specifically, given the user's residual energy e, there exists a threshold $\eta_{th} \in [0, 1]$ such that

$$a^*(\mathbf{\Lambda}, e, \mathbf{\Psi}) \begin{cases} > 0(\text{ sensing }), & \text{if } |\eta| > \eta_{th}, \\ = 0(\text{ sleeping }), & \text{otherwise.} \end{cases} \tag{40}$$

We adopt Poisson packet arrival process. In Fig. 4, we plot the threshold η_{th} on the SOS correlation as a function of the packet arrival rate ρ for different sensing energy consumption e_s. We see that the threshold η_{th} decreases with the packet arrival rate ρ. Intuitively, when ρ is large, there is a high probability that packets will arrive in this slot, and hence the user should be more active in collecting information about the SOS for better channel selection in the next slot. As the packet arrival rate ρ keeps increasing, the threshold η_{th} approaches zero, *i.e.,* the user should always sense a channel. This observation demonstrates Proposition 4 since we have the continuous traffic case when ρ is infinite. As expected, the threshold η_{th} also increases with the sensing energy consumption e_s. As sensing cost e_s increases, the user with an empty buffer tends to operate in the sleeping mode; it only senses a channel when the resulting sensing outcome can provide more information about the SOS, *i.e.,* the time correlation of the SOS is high.

[7]Similar observations are obtained for the case when the probability $\Psi_0(t) = \Pr\{D(t) = 0\}$ of empty buffer is close to 1.

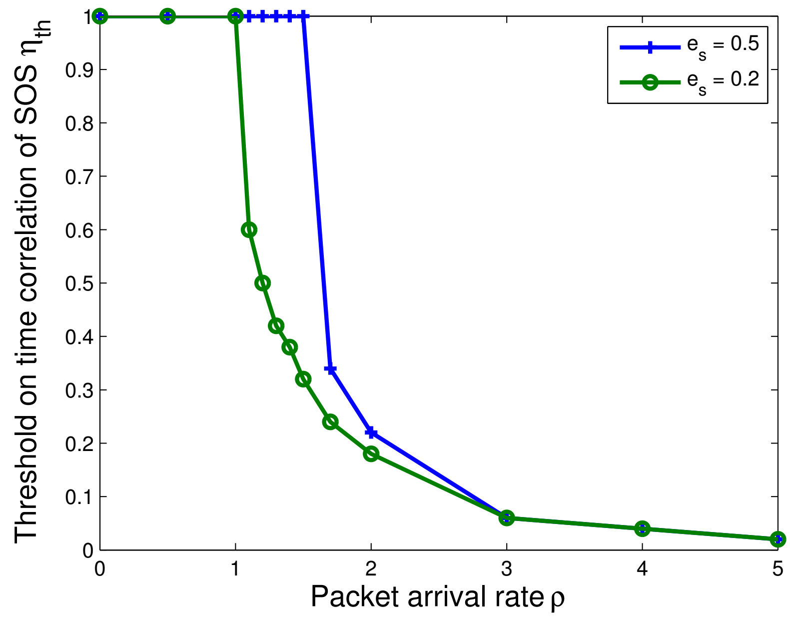

Fig. 4. Optimal thresholds η_{th} on the SOS correlation. $\overset{\circ}{e} = 30$ (initial energy), $B_1 = B_2 = 1$ (bandwidth), $\mathbf{\Omega}(1) = [0.5, 0.5]$ (initial belief vector), $e_p = 0.1$ (sleeping energy), $E_{tx} \in \{1, 2\}$ (transmission energy), $[p_i(1), p_i(2)] = [0.6, 0.4]$, $i = 1, 2$, (channel fading statistics).

VII. Conclusion and Discussion

Within the POMDP framework, we have developed optimal distributed MAC protocols for energy-constrained OSA under both the continuous and the bursty traffic models. To study the fundamental design tradeoff, we have established threshold structures of the optimal sensing and access policies. We have also provided numerical examples to study the impact of different factors that affect the optimal decisions. We find that the residual energy has more significant impact on the optimal sensing and access decisions when the battery is close to depletion or the channel occupancy state is negatively correlated in time. When the sensing cost is high, the secondary user should be more conservative in sensing but more aggressive in accessing the channel. Interestingly, we also find that even if a secondary user does not have any packet to send in the current slot, it should still choose to sense a channel when the time-correlation of the channel occupancy state is large. These results provide not only insights into the energy-constrained OSA design but also guidelines for suboptimal designs.

In this paper, we have assumed that secondary users have perfect knowledge of the statistical model of the spectrum usage. We take the viewpoint that such statistical models of a particular spectrum region should be obtained through measurements before the deployment of secondary networks in that spectrum region. This is for the purpose of evaluating the potential gain or profit of secondary market in that spectrum region. Such statistical models can then be made available to secondary users to facilitate the design. We are, however, aware that in some scenarios, secondary users may not have access to spectrum usage models. In this case, we have a POMDP with unknown model, and existing reinforcement learning algorithms may be borrowed [20].

We have not considered sensing errors in this paper. When a secondary user may mistake a busy channel as an idle one and *vice versa*, the joint design of the access strategy and the operating characteristics of the spectrum sensor is crucial in order to minimize overlooked spectrum opportunities without violating the interference constraint. This issue has been fully addressed in [4] in the absence of energy constraint. The impact of sensing errors on energy-constrained OSA design is one of the future directions. In particular, how to exploit the RTS-CTS exchange to combat sensing errors and to ensure synchronous hopping is worth investigating. Another interesting extension is to consider a scenario where batteries could be slowly recharged.

The interaction among secondary users has not been taken into account. The sensing and access protocols proposed in this paper can be applied to a network of secondary users. Their performance is, however, suboptimal in terms of *network* throughput. Preliminary results on spectrum sharing among distributed competing secondary users have been obtained in [21] without considering energy constraints. We hope that the proposed optimal single-user energy-constrained MAC protocols provide insights for the design of multi-user OSA with energy constraint.

APPENDIX A: PROOF OF PROPOSITION 1

As explained in Section IV-D, given any initial energy $\mathring{e}$ and any initial belief vector $\mathring{\mathbf{\Lambda}}$, the secondary user can only experience a finite number of information states $(\mathbf{\Lambda}, e)$ during its entire battery lifetime. Hence, an energy-constrained OSA problem can be viewed as a MDP with a finite state space consisting of all possible information states. Moreover, the immediate reward defined in (14) is non-negative. This, together with the inevitable termination, makes the energy-constrained OSA design an example of a stochastic shortest path problem.

The key to understanding the existence of stationary optimal polices is to note that the residual energy of the secondary user is part of the system state. Since the residual energy determines the remaining lifetime, the system state contains all the time-dependent information for decision-making. The optimal actions thus depend only on the system state and are stationary in time

APPENDIX B: PROOF OF PROPOSITION 2

Proof of P2.1: Note that the secondary user with residual energy can always act as if it has a lower residual energy. Hence, the secondary user with a larger initial energy earns no fewer rewards. □□□

Proof of P2.2: We prove P2.2 by induction over residual energies e. Specifically, for the lowest possible residual energy $e = \min \mathcal{E}$, the value function of any information state is $V(\mathbf{\Omega}, e) = 0$ and hence P2.2 holds. Suppose that it holds for all possible residual energies $e' \in \mathcal{E}$ lower than e. Since $\mathbf{\Omega} \geq \mathbf{\Omega}'$ implies $\hat{\mathcal{T}}(\mathbf{\Omega}|0) \geq \hat{\mathcal{T}}(\mathbf{\Omega}'|0)$ as seen from (22), we obtain from (25a) that $Q_0(\mathbf{\Omega}, e) \geq Q_0(\mathbf{\Omega}', e)$. Next, we show that $Q_a(\mathbf{\Omega}, e) \geq Q_a(\mathbf{\Omega}', e)$ for $\mathbf{\Omega} \geq \mathbf{\Omega}'$. We note that when $\mathbf{\Omega} \geq \mathbf{\Omega}'$, we have $\hat{\mathcal{T}}(\mathbf{\Omega}|a,k) \geq \hat{\mathcal{T}}(\mathbf{\Omega}'|a,k)$ from (23) and hence $Q_a(\mathbf{\Omega}, e|k, \Phi_k) \geq Q_a(\mathbf{\Omega}', e|k, \Phi_k)$ from (25c). Since $\hat{\mathcal{T}}(\mathbf{\Omega}'|a,k) \geq \hat{\mathcal{T}}(\mathbf{\Omega}'|a,0)$ as seen from (23), we have $Q_a(\mathbf{\Omega}', e|k, \Phi_k) \geq Q_a(\mathbf{\Omega}', e|0, 0) = V(\hat{\mathcal{T}}(\mathbf{\Omega}'|a,0), e - e_s)$ from (25c). Using (25b), we then obtain that

$$\begin{aligned}
Q_a(\mathbf{\Omega}, e) &\geq (1-\omega_a) V(\hat{\mathcal{T}}(\mathbf{\Omega}'|a,0), e-e_s) + \omega_a \sum_{k=1}^{L} p_a(k) \max_{\Phi_k \in \varrho(e|a,k)} Q_a(\mathbf{\Omega}', e|k, \Phi_k) \\
&\geq V(\hat{\mathcal{T}}(\mathbf{\Omega}'|a,0), e-e_s) + \omega_a' \left[\sum_{k=1}^{L} p_a(k) \max_{\Phi_k \in \varrho(e|a,k)} Q_a(\mathbf{\Omega}', e|k, \Phi_k) - V(\hat{\mathcal{T}}(\mathbf{\Omega}'|a,0), e-e_s) \right] \\
&= (1-\omega_a') V(\hat{\mathcal{T}}(\mathbf{\Omega}'|a,0), e-e_s) + \omega_a' \sum_{k=1}^{L} p_a(k) \max_{\Phi_k \in \varrho(e|a,k)} Q_a(\mathbf{\Omega}', e|k, \Phi_k) = Q_a(\mathbf{\Omega}', e).
\end{aligned}$$

Hence, by (24), we have $V(\mathbf{\Omega}, e) \geq V(\mathbf{\Omega}', e)$, which completes the proof. □□□

APPENDIX C: PROOF OF PROPOSITION 3

The proof of Proposition 3 is very similar to that provided in [14] for a POMDP with finite and fixed time horizon. Hence, we only briefly describe the procedure for this proof.

For any residual energy $e < e_s + \varepsilon_1$, we have $V(\mathbf{\Lambda}, e) = 0$, which can be written as an inner product of the belief vector $\mathbf{\Lambda}$ and an all-zero Υ-vector. Suppose that Proposition 3 holds for all residual energies $e' \in \mathcal{E}$ that are lower than e. After some algebra, we can re-write the action-value functions given in (17) and (20) in terms of the Υ-vectors:

$$Q_0(\mathbf{\Lambda}, e) = \max_{\mathbf{\Upsilon} \in \mathbf{\Gamma}_{e-e_p}} < \mathcal{T}(\mathbf{\Lambda}|0), \mathbf{\Upsilon} > = \sum_{\mathbf{s}' \in \mathcal{S}} \lambda_{\mathbf{s}'} \left[\sum_{\mathbf{s} \in \mathcal{S}} P_{\mathbf{S}}(\mathbf{s}|\mathbf{s}') \Upsilon_{e-e_p, \mathbf{s}}^{l(\mathbf{\Lambda}, 0)} \right], \tag{41}$$

$$\begin{aligned}
Q_a(\mathbf{\Lambda}, e) &= \max_{\mathbf{\Phi} \in \varrho(e|a)} \sum_{\mathbf{s} \in \mathcal{S}} \lambda_{\mathbf{s}} \sum_{k=0}^{L} U_a(k|\mathbf{s}) \left(B_a \Phi_k + \max_{\mathbf{\Upsilon} \in \mathbf{\Gamma}_{e-e_s-\Phi_k \varepsilon_k}} < \mathcal{T}(\mathbf{\Lambda}|a,k), \mathbf{\Upsilon} > \right) \\
&= \max_{\mathbf{\Phi} \in \varrho(e|a)} \sum_{\mathbf{s}' \in \mathcal{S}} \lambda_{\mathbf{s}'} \left[\sum_{k=0}^{L} U_a(k|\mathbf{s}') \left(B_a \Phi_k + \sum_{\mathbf{s} \in \mathcal{S}} P_{\mathbf{S}}(\mathbf{s}|\mathbf{s}') \Upsilon_{e-e_s-\Phi_k \varepsilon_k, \mathbf{s}}^{l(\mathbf{\Lambda}, a, k)} \right) \right],
\end{aligned} \tag{42}$$

where $\Upsilon_{e-e_p, \mathbf{s}}^{l(\mathbf{\Lambda}, 0)}$ and $\Upsilon_{e-e_s-\Phi_k \varepsilon_k, \mathbf{s}}^{l(\mathbf{\Lambda}, a, k)}$ are, respectively, the Υ-vectors associated with the regions containing belief vectors $\mathcal{T}(\mathbf{\Lambda}|0)$ and $\mathcal{T}(\mathbf{\Lambda}|a,k)$, respectively. Viewing each term in the square brackets of (41) and (42) as an element $\Upsilon_{e,\mathbf{s}}$ of a possible Υ-vector $\mathbf{\Upsilon}_e$, we find that the action-value functions can be written as an inner product of the belief vector and an Υ-vector $\mathbf{\Upsilon}_e$. Moreover, there are only a finite number of such Υ-vectors $\mathbf{\Upsilon}_e$ since we have assumed that sets $\mathbf{\Gamma}_{e'}$ are finite for all $e' < e$. Since the maximum of a finite set of piecewise linear and convex functions is also piecewise linear and convex, Proposition 3 holds.

APPENDIX D: PROOF OF PROPOSITIONS 4 - 5 AND COROLLARY 1

Proof of Proposition 4 We prove by induction that $a^*(\mathring{\mathbf{\Lambda}}, e) = 1$, *i.e.,* $V(\mathring{\mathbf{\Lambda}}, e) = \max_{a\in\{1,\ldots,N\}} Q_a(\mathring{\mathbf{\Lambda}}, e)$. Clearly, $V(\mathring{\mathbf{\Lambda}}, e) = Q_a(\mathring{\mathbf{\Lambda}}, e) = 0$ holds for any $a \in \{1, \ldots, N\}$ when $e = \min \mathcal{E}$. Suppose that this equality holds for all residual energies $e' \in \mathcal{E}$ lower than e. Since $\mathring{\mathbf{\Lambda}}$ is the stationary distribution of the underlying SOS, we have $\mathcal{T}(\mathbf{\Lambda}|0) = \mathbf{\Lambda}$. We thus obtain from (16) and (17) that

$$\begin{aligned} V(\mathring{\mathbf{\Lambda}}, e) &= \max\{V(\mathring{\mathbf{\Lambda}}, e - e_p), \max_{a\in\{1,\ldots,N\}} Q_a(\mathring{\mathbf{\Lambda}}, e)\} \\ &= \max\{\max_{a\in\{1,\ldots,N\}} Q_a(\mathring{\mathbf{\Lambda}}, e - e_p), \max_{a\in\{1,\ldots,N\}} Q_a(\mathring{\mathbf{\Lambda}}, e)\} = \max_{a\in\{1,\ldots,N\}} Q_a(\mathring{\mathbf{\Lambda}}, e), \end{aligned} \tag{43}$$

where the last equality is due to the monotonicity of the value function in terms of the residual energy. Proposition 4 thus follows. □□□

Proof of Proposition 5 We consider the single-channel case ($N = 1$) and adopt the value function defined in (24) and (25). We note that when $N = 1$, the belief vector $\mathbf{\Lambda}(t)$ reduces to a scalar $\omega(t)$ as defined in (21), and the corresponding belief update $\hat{T}(\omega(t)|a, k)$ under sensing outcome k from this channel reduces from (23) to $\omega(t + 1) = \beta$ if $k > 0$ and $\omega(t + 1) = \alpha$ if $k = 0$, which is independent of the current belief vector $\omega(t)$.

Lemma 1: Consider the single-channel case ($N = 1$). Given current residual energy e, we have that for any belief vector ω,

$$G(e) \stackrel{\Delta}{=} \sum_{k=1}^{L} p(k) \max_{\Phi_k\in\varrho(e|a,k)} Q_1(\omega, e|k, \Phi_k) \geq V(\omega, e - e_p). \tag{44}$$

Proof: We prove this lemma by induction. For any residual energy $e < e_s + \varepsilon_1$, the value function of any information state is $V(\Omega, e) = 0$ and hence (44) holds. Suppose that it holds for all possible residual energies $e' \in \mathcal{E}$ lower than e. Then, applying (25) to (24), we obtain that

$$\begin{aligned} V(\omega, e - e_p) &= \max\{V(\alpha + (\beta - \alpha)\omega, e - 2e_p); (1 - \omega)V(\alpha, e - e_p - e_s) + \omega G(e - e_p)\} \\ &\leq \max\{G(e - e_p); (1 - \omega)G(e - e_s) + \omega G(e - e_p)\} \leq G(e - e_p) \leq G(e), \end{aligned} \tag{45}$$

where the last two inequalities follow from the fact that $e_s \geq e_p$ and the value function is monotonically increasing with the residual energy. This completes the proof of (44). ■

Suppose that the optimal sensing action is $a^*(\omega_1, e) = 1$ (*i.e.,* sensing) when the current belief vector is ω_1. That is, $Q_1(\omega_1, e) \geq Q_0(\omega_1, e)$. Consider any belief vector ω_2 such that $\omega_2 \geq \omega_1$. We obtain from (25b) that

$$\begin{aligned} Q_1(\omega_2, e) &= (1 - \omega_2)V(\alpha, e - e_s) + \omega_2 G(e) = (1 - \omega_2)\frac{Q_1(\omega_1, e) - \omega_1 G(e)}{1 - \omega_1} + \omega_2 G(e) \\ &= \frac{1 - \omega_2}{1 - \omega_1} Q_1(\omega_1, e) + \frac{\omega_2 - \omega_1}{1 - \omega_1} G(e) \geq \frac{1 - \omega_2}{1 - \omega_1} Q_0(\omega_1, e) + \frac{\omega_2 - \omega_1}{1 - \omega_1} G(e). \end{aligned} \tag{46}$$

Applying (25a) and (44) to (46), we obtain that

$$Q_1(\omega_2, e) \geq \frac{1 - \omega_2}{1 - \omega_1} V(\alpha + (\beta - \alpha)\omega_1, e - e_p) + \frac{\omega_2 - \omega_1}{1 - \omega_1} V(\beta, e - e_p). \tag{47}$$

Since the value function $V(\omega, e)$ is convex in belief vector ω, we obtain from (47) that

$$Q_1(\omega_2, e) \geq V\left(\frac{1-\omega_2}{1-\omega_1}[\alpha + (\beta-\alpha)\omega_1] + \frac{\omega_2-\omega_1}{1-\omega_1}\beta, e-e_p\right) = V(\alpha + (\beta-\alpha)\omega_2, e-e_p) = Q_0(\omega_2, e). \quad (48)$$

Hence, the optimal sensing action for ω_2 is $a^*(\omega_2, e) = 1$. That is, $a^*(\omega, e)$ is monotonically increasing in $\hat{\omega}$.

We see from (22) and (23) that the belief vector $\omega \in [\min\{\alpha,\beta\}, \max\{\alpha,\beta\}]$. Furthermore, by Proposition 4, the optimal sensing action is given by $a^*(\overset{\circ}{\omega}, e) = 1$ where $\overset{\circ}{\omega} = \frac{\alpha}{1+\alpha-\beta}$ is the stationary distribution of the SOS. Hence, threshold $r_{th}(e)$ is upper bounded by $\frac{\alpha}{1+\alpha-\beta}$. □□□

Proof of Corollary 1 When $\alpha = \beta$, we have $\omega = \alpha$ as seen from (22) and (23). By Proposition 5, the sensing threshold is given by $r_{th}(e) = \alpha$, and hence $a^*(\omega, e) = 1$. □□□

APPENDIX E: PROOF OF PROPOSITION 6

Consider the case where the secondary user operates in the sensing mode and observes sensing outcome $\Theta(t) > 0$. Inspection of (6) and (13) reveals that the belief update $\mathcal{T}(\mathbf{\Lambda}|a,k)$ is independent of k when $k > 0$. Hence, $Q_a(\mathbf{\Lambda}, e|k, 0)$ is identical for all positive k. It thus suffices to show that $Q_a(\mathbf{\Lambda}, e|k, 1)$ is monotonically decreasing with k. This follows straightforwardly from (19) and the monotonicity of the value function with the residual energy.

Furthermore, when $N = 1$, the updated belief vector $\mathcal{T}(\mathbf{\Lambda}|a,k)$ is determined solely by the current observation. The action-value function $Q_a(\mathbf{\Lambda}, e|k, \phi_k)$ given in (19) and, hence, the optimal access decision are thus independent of the current belief vector.

REFERENCES

[1] J. Mitola, "Cognitive radio for flexible mobile multimedia communications," in *Proceedings of IEEE International Workshop on Mobile Multimedia Communications*, pp. 3 - 9, 1999.

[2] Q. Zhao and A. Swami, "A Survey of dynamic spectrum access: signal processing and networking perspectives," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 4, pp. 1349 - 1352, Honolulu, Hawaii, Apr. 2007.

[3] Q. Zhao, L. Tong, A. Swami, and Y. Chen "Decentralized cognitive MAC for opportunistic spectrum access in ad hoc networks: A POMDP framework," *IEEE Journal on Selected Areas in Communications*, vol. 25, no. 3, pp. 589 - 600, Apr. 2007 (also see *Proceedings of the first IEEE Symposium on New Frontiers in Dynamic Spectrum Access Networks*, pp. 224 - 232, Nov. 2005).

[4] Y. Chen, Q. Zhao, and A. Swami, " Joint design and separation principle for opportunistic spectrum access in the presence of sensing errors," *IEEE Transactions on Information Theory*, vol. 54, no. 5, pp. 2053-2071, May, 2008 (also see *Proc. of IEEE Asilomar Conference on Signals, Systems, and Computers*, Oct. 2006)

[5] P. Papadimitratos, S. Sankaranarayanan, and A. Mishra, "A bandwidth sharing approach to improve licensed spectrum utilization," *IEEE Communications Magazine*, vol. 43, pp. 10-14, Dec. 2005.

[6] Qianchun Zhao, S. Geirhofer, L. Tong, and B. M. Sadler, "Optimal dynamic spectrum access via periodic channel sensing," in *Proceedings of Wireless Communications and Networking Conference (WCNC)*, pp. 33 - 37, Hong Kong, Mar. 2007.

[7] Q. Zhao and K. Liu, "Detecting, tracking, and exploiting spectrum opportunities in unslotted primary systems," in *Proceedings of IEEE Radio and Wireless Symposium (RWS)*, San Diego, CA, Jan. 2008.

[8] S. Geirhofer, L. Tong, and B. M. Sadler, "Dynamic spectrum access in WLAN channels: empirical model and its stochastic analysis," in *Proceedings of the First International Workshop on Technology and Policy in Accessing Spectrum (TAPAS), Boston, MA, Aug. 2006.*

[9] D. Datla, A. M. Wyglinski, and G. J. Minden, "A spectrum surveying framework for dynamic spectrum access networks," submitted to *IEEE Journal on Selected Areas in Communications: Cognitive Radio: Theory and Applications*, March 2007.

[10] S. D. Jones, E. Jung, X. Liu, N. Merheb, and I. -J. Wang, "Characterization of spectrum activities in the U.S. public safety band for opportunistic spectrum access," in *Proceedings of the IEEE Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN)*, Dublin, Ireland, Apr. 2007.

[11] Q. Zhao and B. Sadler, "A survey of dynamic spectrum access," *IEEE Signal Processing Magazine*, vol. 24, no. 3, pp. 79 - 89, May 2007.

[12] D. P. Bertsekas, *Dynamic Programming and Optimal Control*, vol. 1. Belmont, Massachusetts: Athena Scientific, 1995.

[13] W. S. Lovejoy, "Some monotonicity results for partially observed Markov decision processes," *Operations Research*, vol. 35, no. 5, pp. 736 - 743, Sep. - Oct. 1987.

[14] R. D. Smallwood and E. J. Sondik, "The optimal control of partially observable Markov processes over a finite horizon," *Operations Research*, vol. 21, pp. 1071–1088, 1973.

[15] H. -T. Cheng, *Algorithms for partially observable Markov decision processes*, PhD thesis, University of British Columbia, British Columbia, Canada, 1988.

[16] M. Littman, "The witness algorithm: solving partially observable Markov decision processes," *Technical Report CS-94-40*, Brown University, Providence RI, 1994.

[17] A. Cassandra, M. Littman, and N. L. Zhang, "Incremental pruning: a simple, fast, exact algorithm for partially observable Markov decision processes," in *Proc. of the 13th Conference on Uncertainties in Artificial Intelligence*.

[18] A. Cassandra, "Optimal policies for partially observable Markov decision processes," *Technical Report CS-94-14*, Brown University, Providence RI, 1994.

[19] E. J. Sondik, *The optimal control of partially observable Markov processes*, PhD thesis, Stanford University, Stanford, California, 1971.

[20] D. Aberdeen, "A survey of approximate methods for solving partially observable Markov decision processes," tech. rep., National ICT Australia, December 2003. `http://users.rsise.anu.edu.au/~daa/papers.html`.

[21] K. Liu, Q. Zhao, and Y. Chen, "Distributed Sensing and Access in Cognitive Radio Networks," to appear in *Proceedings of 10th International Symposium on Spread Spectrum Techniques and Applications (ISSSTA)*, Bologna, Italy, August, 2008.