Decision-Theoretic Distributed Channel Selection for Opportunistic Spectrum Access: Strategies, Challenges and Solutions

Yuhua Xu, *Student Member, IEEE,* Alagan Anpalagan, *Senior Member, IEEE,* Qihui Wu, *Senior Member, IEEE,* Liang Shen, Zhan Gao, and Jinglong Wang, *Senior Member, IEEE*

***Abstract*—Opportunistic spectrum access (OSA) has been regarded as the most promising approach to solve the paradox between spectrum *scarcity* and *waste*. Intelligent decision making is key to OSA and differentiates it from previous wireless technologies. In this article, a survey of decision-theoretic solutions for channel selection and access strategies for OSA system is presented. We analyze the challenges facing OSA systems globally, which mainly include interactions among multiple users, dynamic spectrum opportunity, tradeoff between sequential sensing cost and expected reward, and tradeoff between exploitation and exploration in the absence of prior statistical information. We provide comprehensive review and comparison of each kind of existing decision-theoretic solution, i.e., game models, Markovian decision process, optimal stopping problem and multi-armed bandit problem. We analyze their strengths and limitations and outline further research for both technical contents and methodologies. In particular, these solutions are critically analyzed in terms of information, cost and convergence speed, which are key concerns for practical implementation. Moreover, it is noted that each kind of existing decision-theoretic solution mainly addresses one aspect of the challenges, which implies that two or more kinds of decision-theoretic solutions should be incorporated to address more challenges simultaneously.**

***Index Terms*—Opportunistic spectrum access, cognitive radio, distributed channel selection, game theory, Markovian decision process, optimal stopping problem, multi-armed bandit problem.**

I. Introduction

THE EXPLOSIVE increase in wireless service demand has made *spectrum scarcity* a serious problem facing today's wireless communication systems. Historically, most of the spectrum less than 6 GHz has been almost completely assigned to different services, which were called the licensed spectrum owners. However, the reports released by FCC and the practical measurements carried out by researchers show that the allocated spectrum is largely under-utilized in time and space, which is referred to as *spectrum waste*. To address the paradox between spectrum scarcity and waste, lots of attentions have been given to the following two aspects of effort: (i) explore and utilize unknown spectrum, e.g., non-line-of-sight ultraviolet communications [1], visible light wireless communications [2] and Terahertz communications [3], and (ii) seek and opportunistically utilize the "spectrum hole" that is not utilized by the licensed owners in the current spectrum, which is referred to as opportunistic spectrum access (OSA) technology. In comparison, OSA technology operates in the current spectrum and hence is compatible with existing technologies, which has made it a hot research topic in the last decade. In this article, we focus on opportunistic spectrum access technologies.

Technically, the successful implementation of OSA technology is mainly relying on cognitive radio (CR), which was firstly coined by J. Mitola [4]. More importantly, some developed hardware and software platforms, e.g., GNU Radio [5], Universal Software Radio Peripheral (USRP) [6], Shared Spectrums XG Radio [7], Wireless open-Access Research Platform (WARP) [8]–[10] and Open-Source SCA Implementation Embedded (OSSIE) [11]–[13], make it realizable. Generally, the operational procedure of OSA technology can be described by the following steps [14]:

1) Observation: sample and collect multi-domain information about the environment, which may include information about spectrum occupancy, location, user preference, traffic and network state.
2) Decision: make intelligent decisions based on the observation results, e.g., identify spectrum holes and find the optimal spectrum bands to use, learn the behaviors of other users, predict the interactions among multiple users and infer useful knowledge from the collected data.
3) Reconfiguration: change the parameters of the radio accordingly to implement the decisions.

In summary, observation belongs to the field of traditional signal detection and processing while reconfiguration is related to hardware operation. On the other hand, decision making is key to technologies and differentiates them from previous wireless technologies.

There are always multiple channels in opportunistic spectrum access systems and the key concern is the joint optimization for channel sensing and access, which can be abstracted as channel selection [15]. For the problem of channel selection, there are two basic decision issues: (i) in the parallel sensing strategies, which channels are chosen to sense and access in

Manuscript received September 21, 2012; revised January 22, 2013. This work was supported by the National Basic Research Program of China under Grant No. 2009CB320400, the National Science Foundation of China under Grant No. 60932002 and No. 61172062, and in part by Jiangsu Province Natural Science Foundation of China under Grant No. BK2011116.

Y. Xu, Q.Wu, L. Shen, Z. Gao and J. Wang are with the Institute of Communications Engineering, PLA University of Science and Technology, Nanjing 21007, China (e-mail: yuhuaenator@gmail.com; wqhtxdk@yahoo.cn; ShenLiang671104@sina.com; gzck111@sina.com; wjl543@sina.com).

A. Anpalagan is with the Department of Electrical and Computer Engineering, Ryerson University, Toronto, Canada (e-mail: alagan@ee.ryerson.ca).

Digital Object Identifier 10.1109/SURV.2013.030713.00189

each slot? (ii) in the sequential sensing strategies, how to determine the sensing order and control the sensing process to balance the tradeoff between sensing cost and expected reward? To address the two issues, different decision-theoretic solutions were proposed.

In addition, various challenges facing the problem of channel selection in opportunistic spectrum access systems arise as the results of flexible speculum usage. Specifically, the challenges mainly include: interactions among multiple users, dynamic spectrum opportunity, tradeoff between sequential sensing cost and expected reward, and tradeoff of between exploitation and exploration in the absence of prior statistical information. To address these challenges, some powerful decision-theoretic solutions have been extensively studied, e.g., game models, Markovian decision process, optimal stopping problem and multi-armed bandit problem. It is seen that each kind of these solution mainly addresses one aspect of challenges.

Thus, the focus of this article is to provide a comprehensive review of the state-of-the-art on decision-theoretic solutions for opportunistic spectrum access systems. Based on the in-depth review and comparison of results, strengths and limitations of each kind of existing decision-theoretic solutions are discussed. More importantly, several future research problems for both technical content and methodology are suggested.

A few surveys that review existing solutions for opportunistic spectrum access systems have already been out. General review related to opportunistic spectrum access technologies and applications were provided in [16]–[20]. A multitude of articles for reviewing medium access control (MAC) protocols in opportunistic spectrum access systems can be found in [21]–[25]. A Markovian decision process framework for OSA technologies was proposed in [26] and game-theoretic solutions for opportunistic spectrum access systems were surveyed in [27]–[29]. Furthermore, a survey on spectrum decision in cognitive radio networks was presented in [30], machine-learning techniques in cognitive radios was surveyed in [31], techniques for improving reliability of wireless networks using cognitive radios were surveyed in [32] and a survey of artificial intelligence for cognitive radios was presented in [33].

With respect to previous surveys, the contributions of this work are threefold. First, we present a global and integrated analysis of the challenges facing the problem of channel selection in OSA systems. Second, we provide comprehensive review and comparison for each kind of existing decision-theoretic solution and analyze their strengths and limitations. Third, we provide global and critical analysis for the four important decision-theoretic solutions and outline future research for both technical contents and methodologies. In particular, these solutions are contrastively analyzed in terms of information, cost and convergence speed, which are concerns for practical implementation. Moreover, it is noted that each kind of existing decision-theoretic solution mainly addresses one aspect of the challenges, which implies that two or more kinds of decision-theoretic solutions should be incorporated to address more challenges simultaneously.

The rest of this article is organized as follows. In Section II, background and challenges of opportunistic spectrum access systems are presented. We review and compare four kinds of decision-theoretic solutions for opportunistic spectrum access systems in Sections III–VI respectively; specifically, game models in Section III, Markovian decision process in Section IV, optimal stopping problems in Section V and multi-armed bandit problem in Section VI. Finally, contrastive analysis for the four kinds of solutions and future research directions are presented in Section VII, and summary is provided in Section VIII.

II. Background and Challenges of Opportunistic Spectrum Access

A. Background

We mainly consider decision-theoretic solutions for slotted OSA systems in this article, i.e., the secondary user (SU) employs a sensing-then-access structure with equal slot length. There are two traffic models for the primary users (PUs): (i) slotted traffic [15], i.e., the state of PU (idle or active) remains unchanged in each slot and changes over slots independently or correlatively, and (ii) continuous traffic [35], i.e., the PU does not have slotted transmission structure and it may become idle or active at any time. For decision-theoretic solutions, continuous traffic models can be converted to slotted traffic models by imposing collision constraints on the SUs [36]. We focus on slotted opportunistic spectrum access systems since the used slotted transmission structures coincide with the nature of periodic procedure of decision theories[1].

There are two basic components in opportunistic spectrum access systems: (i) a spectrum sensing strategy for detecting the activities of PUs to determine whether to perform sensing, the number of sensing channels and which channels to sense, and (ii) a channel access rule for determining whether access the channels and which channels to access based on the sensing results. Axiomatically, the access channels are exactly those sensed idle or a subset of them. Basically, channel sensing and channel access strategies should be jointly optimized, which is a distinct feature differentiating OSA from traditional wireless technologies.

There are always multiple channels in opportunistic spectrum access systems and the key concern of the task for joint channel sensing and access can be abstracted as channel selection [15]. Specifically, the task of channel selection in opportunistic spectrum access systems mainly includes determining channels to sense and selecting the channels to access. When more than one SUs choose to access the same channel, only one or none of them will receive positive payoff depending on the channel collision models, e.g., using perfect contention resolution mechanisms or not. Surely, the task of optimizing the parameters in the contention resolution mechanisms is also an important issue, but it is key to traditional multiple access control problem but not to slotted opportunistic spectrum access systems. Therefore, we mainly discuss the problem of channel selection for spectrum sensing strategy and channel access rule in this article.

[1]It should be pointed out that there are also a large number of studies for unslotted OSA systems, where the SUs do not have slotted transmission structure and can sense and access the licensed channels in arbitrary time [37]. However, we limited our work to slotted opportunistic spectrum access systems as the decision theories used in opportunistic spectrum access systems are more classical and typical.

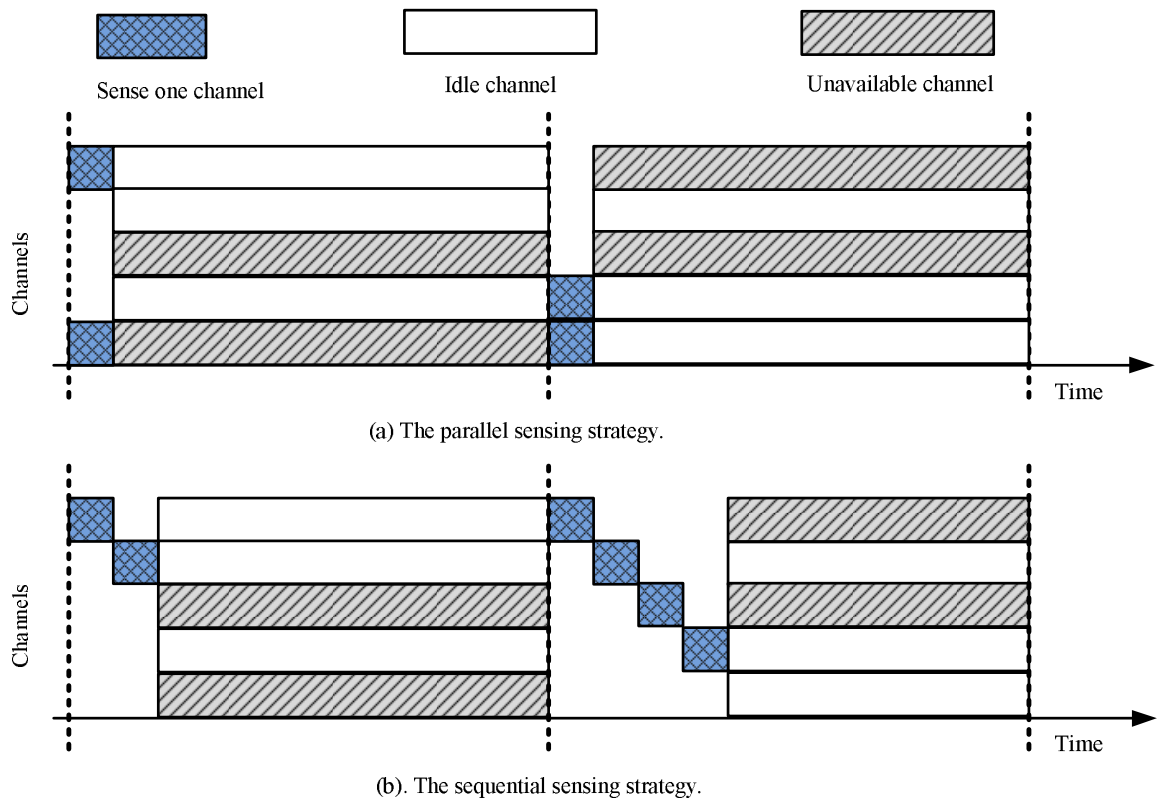

Fig. 1. An illustrative and general diagram of two basic sensing strategies for the slotted opportunistic spectrum access. In the parallel sensing strategy, the SU simultaneously senses a *fixed* number set of channels at the beginning of the slot and updates the channel selections in the next slot. In the sequential sensing strategy, the SU sequentially senses a *variable* number set of channels according to a control policy. In the parallel sensing strategy, an extensively studied model is that only one channel is sensed in a slot.

From the perspective of control model, OSA technologies can be implemented in centralized [38], [39] or distributed [15] manner. In the centralized OSA model, there is a central controller which schedules the sensing and access of SUs. On the contrary, the SUs in the distributed OSA models behave distributively and autonomously. In comparison, the centralized OSA models involve heavy computational complexity and communication overhead, while the distributed models can be implemented with low computational complexity and communication overhead. Furthermore, distributed control model is easy to implement and robust to observation error and link failure. Based on this consideration, we discuss distributed and slotted OSA models in this article.

An illustrative and general diagram for sensing strategies in opportunistic spectrum access systems are shown in Fig. 1. Due to hardware limitation, the SUs can not sense all the channels simultaneously; instead, they can only sense a small part of channels in a slot [40]. As a result, there are two basic sensing strategies in the literature: *parallel sensing* [15], i.e., the SU simultaneously senses a fixed set of channels in a slot, and *sequential sensing* [41], i.e., the SU sequentially senses the channels according to a pre-defined order and stops to sense when some criterion is met. From a systematic perspective, the decision-theoretic research problems are: (i) in the parallel sensing strategies, which channels are chosen to sense and access in each slot? (ii) in the sequential sensing strategies, how to determine the sensing order and control the sensing process to balance the tradeoff between sensing cost and achieved performance? In different scenarios and under different conditions, variants of the above two basic problems have been investigated using different decision-theoretic solutions. Our goal in this article is to provide a comprehensive review and comparison of existing decision-theoretic studies and seek for new solutions to cope with the challenges facing opportunistic spectrum access systems, which will be discussed in the following subsection.

Remark 1: Besides the above discussed share-use OSA systems, there is also another form of OSA, i.e., the free-use model [43]–[45]. The free-use model is differentiated from the share-use one in that there is no PU and all the SUs can access the spectrum equally and freely. Although spectrum sensing may not be needed anymore, channel selection is also a key concern. Therefore, the decision-theoretic methodologies are suitable not only for share-use OSA systems but also free-use OSA systems. We will also discuss existing studies for free-use OSA systems when the decision-theoretic models are commonly used in the two kinds of opportunistic spectrum access systems.

It should be pointed out that the channels considered in this article are also referred to as abstract channels. For orthogonal frequency division multiplex (OFDM) systems, spectrum sensing and access exhibit distinct attributes, e.g., exploiting the features of OFDM signals to improve the sensing performance [46] and the task of channel selection involves a constraint that no more than two users can choose the same sub-carrier simultaneously [47]. Generally, these

attributes involve optimization techniques rather than decision-theoretic solutions. Thus, spectrum sensing and access in OFDM systems is beyond the scope this article and we do not discuss them.

B. Challenges

In this subsection, we summarize the main challenges facing the distributed opportunistic spectrum access systems and briefly present existing decision-theoretic solutions. Basically, opportunistic spectrum access systems suffer from challenges in traditional wireless communications; more importantly, they suffer from challenges caused by the manner of opportunistic spectrum access. To summarize, the main challenges facing the distributed opportunistic spectrum access systems are:

1) ***Interactions among multiple SUs***. There are generally multiple SUs competing for the limited spectrum resources and their decisions are mutually affected. Such interactions are not easy to analyze by traditional optimization methods, but can be well analyzed by game theory. Depending on the system models and the degree of information availability, different game models can be formulated to address the interactions. Furthermore, when encountering other challenges which will be illustrated below, game based solutions become complicated and challenging.

It is emphasized here that the basic assumption in game theory is that the users are rational and utilitarian. In other words, the objective of each user in a game is to maximize its individual utility function. When the utility function only considers the individual payoff of each player, it is referred to as a non-cooperative game [48]. Such selfish behaviors may cause inefficiency. In some other game models, users may cooperate to complete tasks to achieve increased aggregate payoffs, and the key concern therein is how to allocate the increased payoffs among the participants. These models are referred to as cooperative games [49].

Based on the above analysis, it should be pointed out that some existing research, e.g., [15], [43], which explicitly consider multiple SUs, do not belong to game based solutions. The reason is that SUs in these work are not utilitarian, despite the impact of interactions among SUs being considered.

2) ***Dynamic spectrum opportunity***. In OSA models, the PU may become idle and busy randomly in each slot on each channel, which thus makes the available spectrum opportunities for the SUs change dynamic. On one hand, the spectrum opportunity dynamics may be correlated or independent from slot to slot, and the commonly used correlated dynamics over successive slots is Markovian process [50]. On the other hand, the spectrum opportunity dynamics is generally assumed to be independent from channel to channel.

For the decision-theoretic solutions, the spectrum opportunity dynamics cause uncertainty. Specifically, such uncertainty leads to random payoff in each decision episode. If the dynamics is Markovian from slot to slot, Markovian decision process (MDP) can be used to address this uncertainty for single user opportunistic spectrum access systems and Markovian game models [51] can be applied to multiuser opportunistic spectrum access systems. If the dynamic is independent from slot to slot, the SU would choose the channel with the highest availability to access, which naturally achieves the highest expected throughput. However, all the users choosing the channel with the best channel seems unreasonable for multiuser opportunistic spectrum access systems, since the best channel will be overcrowded while others are not utilized by any SU. To deal with this problem, a game with random payoff [52] can be formulated.

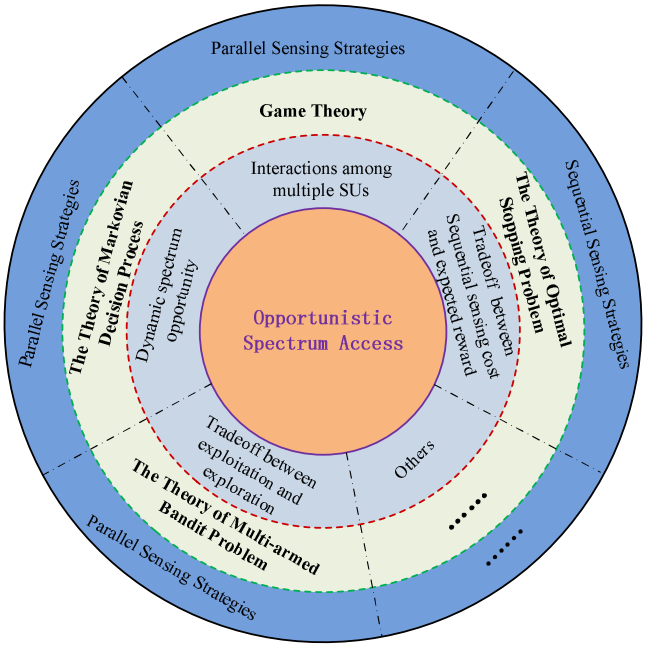

Fig. 2. The challenges and solutions in distributed and slotted OSA systems.

Some previous research, e.g., [55], [56], have assumed quasi-static spectrum opportunities for SUs, i.e., the spectrum opportunities remain unchanged during the convergence of the decision solutions. Thus, these research do not consider dynamic spectrum opportunities, since the algorithms proposed therein were designed for static spectrum opportunities.

3) ***Tradeoff between sequential sensing cost and expected reward***. The channel status, which mainly includes the occupation of PU and channel quality, are random and uncovered to the SUs until they are sensed. In the sequential sensing strategies, there is a fundamental tradeoff between sensing cost and the expected reward. Specifically, as the number of sensing channels increases, the expected received reward increases and so does the sensing cost. In wireless communication systems, sensing cost mainly includes time, bandwidth and energy; in addition, it is proportional to the number of sensing channels. Thus, it needs random optimization approaches to control the sequential sensing procedure, which is adaptively determined and is based on the previously sensed channel results.

For illustration, we present an example below. In a share-use opportunistic spectrum access system, the SUs would select an idle channel with the strongest quality to access, which is referred to as achieving multichannel diversity. It is known that identifying the channel status leads to sensing cost. In such systems, as the number of sensed channels increases, the achieved multichannel diversity increases and so does the sensing cost. Thus, there is a tradeoff between the achieved multiple diversity and sensing cost. To be more specific, there

are two issues in such a random control problem. First, the SUs need to determine the sensing order. Second, based on the observed results, they need to decide whether to stop to explore the residual channels or not, after each channel sensing. This problem can be solved by the theory of optimal stopping [57].

4) ***Tradeoff between exploitation and exploration in the absence of priori statistical information***. In practical opportunistic spectrum access systems, the SUs may not have the statistical information about the spectrum environment a priori, e.g., the idle probabilities of licensed channels. This generally occurs when the SUs move into a new region covered by a PU system. Intuitively, a candidate method is that the SUs firstly collect information and estimate the channel idle probabilities, and then access the channels based on the estimated results. Due to hardware limitation, the SUs can only sense a small part of channels in a slot [40], which implies that the SU needs to use a round-robin scheduler to collect information of all the channels. Such an offline approach is inefficient and costly, since the transmissions of SUs can not be optimized during the estimation periods.

To overcome the problem of lack of prior statistical information, an online learning approach which estimates the statistical information from the decision-payoff history information of the SUs and meanwhile optimizes their selections, is desirable. For online learning approaches, there is a fundamental tradeoff between *exploitation*, i.e., maximizing the reward based on the current estimated statistics, and *exploration*, i.e., spending time on sampling the resources to increase the accuracy of estimated statistic with the prospect of better future rewards.

For single user opportunistic spectrum access systems, the theory of multi-armed bandit problem [58] provides efficient online learning solutions. For multiuser opportunistic spectrum access systems, the task of dealing with the interactions among multiple users in the absence of prior statistical information about the spectrum environment is challenging and is beginning to draw attention recently.

The main challenges facing opportunistic spectrum access systems and the corresponding decision theories are summarized in Fig. 2. Other decision-theoretic solutions for the problem of channel selection in opportunistic spectrum access systems include graph coloring [59], rule-regulated strategies [60] and evolutionary algorithms [61]. We will mainly discuss the above four decision theories, i.e., *game theory*, *Markovian decision process*, *optimal stopping theory* and *multi-armed bandit problem*. It is noted that these decision-theoretic solutions only address a single challenge facing OSA systems respectively, which implies that the problem of channel selection would be more complicated and challenging if two or more aspects of the challenges are jointly considered. We will clearly classify the applied scenarios of each decision theory, review and compare existing studies, and discuss how to incorporate two or more of them to study the problem of channel selection for opportunistic spectrum access systems more practically.

Remark 2: Besides the above four decision-theoretic solutions, there are also some other important research issues in opportunistic spectrum access systems, e.g., cooperative spectrum sensing [62], sensing and access duration optimization [63]–[66], queuing analysis [38], [67]–[69] and capacity analysis [70]–[72]. Technically, these issues can be solved by other optimization techniques, e.g., convex optimization and discrete optimization. However, we do not discuss these studies in this article since they are centralized optimization technologies rather than distributed decision-theoretic solutions.

III. The Application of Game Theory in Opportunistic Spectrum Access

Game theory is an applied mathematical tool that models and analyzes the mutual interactions in multiuser systems [48]. It was originally developed to study the competition/cooperation behavior in economics. Nowadays, it has been widely used in other scenarios, e.g., biological systems [73], human society [74] and engineering [75]. Recently, game theory has been successfully applied to distributed wireless communication systems [76]–[79]. Furthermore, due to the distributed and autonomous decision process, game theory has been regarded as one of the most important decision-theoretic solutions for future communication systems, e.g., 4G [80] and femtocells [81].

In this section, we firstly present the basic models of game theory, review and compare the state-of-the-art game-theoretic solutions for distributed and slotted OSA systems, and finally conclude their strengths and limitations.

A. Basic models of game theory

There are two major branches in game theory: cooperative game and non-cooperative game. In a cooperative game, the users make rational decisions to maximize their individual utility function. In a cooperative game, the users are grouped together according to an enforceable agreement for payoff allocation.

1) Non-cooperative game: A non-cooperative game is generally denoted by $\mathcal{G} = \{\mathcal{N}, A_n, u_n\}$, where $\mathcal{N} = \{1, \ldots, N\}$ is the player (user) set, A_n is the strategy set of player n and u_n is the utility function of player n. A player selects a pure strategy if a single action is chosen from its action set. Denote $a_n \in A_n$ as a chosen strategy of player n and $a = \{a_1, \ldots, a_N\}$ as a strategy profile of all the players. The space of pure strategy profiles is defined as the Cartesian product of the individual strategy spaces: $A = \times_{n\in\mathcal{N}} A_n$. It is conventional to denote a_{-n} as the pure strategy profile of all the users except n. Similarly, $A_{-n} = \times_{i\in\mathcal{N}, i\neq n} A_i$. The utility function in pure strategies can be expressed as $u_n(a_n, a_{-n})$.

Besides pure strategies, a player can also select strategies in a stochastic manner, which is referred to as a mixed strategy. A mixed strategy of player n is denoted as σ_n. Specifically, we denote by $\sigma_n(a_n)$ the probability that player n selects strategy a_n. Then, the mixed strategy space of player n is determined by $\Sigma_n = \{\sigma_n : \sum_{a_n\in A_n} \sigma_n(a_n) = 1 \text{ and } 0 \leq \sigma_n(a_n) \leq 1\}$. A mixed strategy profile of all the players is denoted as $\sigma = \{\sigma_1, \ldots, \sigma_N\}$ and the space of mixed strategy profiles is defined as the Cartesian product of the individual strategy spaces: $\Sigma = \times_{n\in\mathcal{N}} \Sigma_n$. Similarly, σ_{-n} denotes a mixed strategy profile and $\Sigma_{-n} = \times_{i\in\mathcal{N}, i\neq n} \Sigma_i$ denotes the space of mixed strategy profiles of all the users

except player n respectively. The utility function in mixed strategies can be expressed as the expected utility under the mixed strategy profile $\sigma = (\sigma_n, \sigma_{-n})$, i.e., $u_n(\sigma_n, \sigma_{-n}) = \sum_{a \in A} \Big(\prod_{n \in \mathcal{N}} \sigma_n(a_n) \Big) u_n(a)$.

In a non-cooperative game, each user maximizes its individual utility function. To address interactions among multiple players, Nash equilibrium (NE) [48] is the most well-known solution concept. Specifically, a strategy profile $a^* = \{a_1^*, \ldots, a_N^*\}$ is a pure strategy NE if and only if no player can improve its utility function by deviating unilaterally, i.e.,

$$u_n(a_n^*, a_{-n}^*) \geq u_n(a_n, a_{-n}^*), \forall n \in \mathcal{N}, \forall a_n \in A_n, a_n \neq a_n^* \quad (1)$$

Similarly, a mixed strategy profile $\sigma^* = \{\sigma_1^*, \ldots, \sigma_N^*\}$ is a mixed strategy NE if and only if the following equation holds:

$$u_n(\sigma_n^*, \sigma_{-n}^*) \geq u_n(\sigma_n, \sigma_{-n}^*), \forall n \in \mathcal{N}, \forall \sigma_n \in \Sigma_n, \sigma_n \neq \sigma_n^* \quad (2)$$

Another important solution concept of non-cooperative games is correlated equilibrium (CE) [82]. The key difference is that the decisions of the players in NE are independent while those in CE are correlated. Suppose that there is a third party sending a recommendation signal to all the players[2]. These recommendation signals allow players to coordinate their actions and to perform joint randomization over their strategies according to a certain probability distribution. Formally, a joint probability distribution π over A is a CE if and only if for all $n \in \mathcal{N}$, for all actions $a_n \in A_n$ and all alternative actions $a_n' \in A_n$,

$$\sum_{a_{-n} \in A_{-n}} \pi(a_n, a_{-n}) \big(u_n(a_n, a_{-n}) - u_n(a_n', a_{-n}) \big) \geq 0, \quad (3)$$

where $\pi(a_n, a_{-n}) \geq 0$ represents the probability that player n takes strategy a_n while all other players take strategy profile a_{-n}, and $u_n(a_n, a_{-n})$ is the utility function of player n. The inequality (3) indicates that when the recommendation to player n is to choose action a_n, then choosing any other action a_n' rather than a_n can not yield a higher expected utility for player n. In fact, every NE is also a CE and Nash equilibria correspond to the special case where $\pi(a_n, a_{-n})$ is a product of each individual user's probability for different actions. Moreover, CE may include the distribution that is not in the convex hull of the NE distributions.

Compared with NE, CE admits the following advantages: (i) the set of correlated equilibria is structurally simpler than that of Nash equilibria. Because it is a convex set, whereas the Nash equilibria are isolated points at the extrema of this set, and (ii) since the set of correlated equilibria is convex, fairness between players can be well addressed in this domain.

To formulate non-cooperative game-theoretic models, whatever NE or CE solutions are used, the following two aspects should be considered carefully [56]:

- Design a utility function carefully to prevent the users from utilizing the resources selfishly and greedily, i.e., the well-known *tragedy of common* [83]. A current and efficient way is to define the utility function as the received payoff minus the cost [84]–[86]. Using this method, the cost needs to be carefully designed.
- Develop efficient behavior update rules for the users that can achieve the desirable solutions. This issue, however, was underestimated in the previous literature, since achieving NE/CE solutions in the presence of perfect information about other users is not challenging. For example, best response [87] converges to pure strategy NE points for a potential game and regret matching [88] converges to the set of correlated equilibria of a general game. However, having information about others is not always possible in practical applications, especially in wireless communication systems. Thus, achieving desirable solutions without information about others is an interesting but challenging task.

There are several non-cooperative game models that have been widely used in wireless communication engineering, e.g., static game, graphical game, dynamic game, repeated game and evolution game. We will illustrate them in detail in the next subsection.

2) Cooperative game: While noncooperative games study competitive behavior, cooperative games focus on cooperative behavior among rational players. Specifically, cooperative games are concerned primarily with coalitions, which are groups of players, and coordinate their actions to achieve increased payoffs based on an enforceable agreement for payoff allocation. Consequently, the primary problem in cooperative games is how to fairly divide the extra benefit among the members of the formed coalition. For mathematical formulation of models in cooperative game theory, refer to [49]. In the next subsection, we will discuss some important cooperative games that were applied in OSA systems.

B. The application of game models in opportunistic spectrum access systems

In a multiuser opportunistic spectrum access system, channel selections of the SUs are highly interactive. Specifically, SUs choosing the same channel cause interference to each other when they transmit simultaneously, or share the channel using some multiple access control mechanisms. Such interactions can be well modeled and analyzed by game models. In this subsection, we review and compare existing game models in opportunistic spectrum access systems in Table I. There, the behavior update rule is highlighted since it is a key step for practical implementation of game models. Moreover, it is noted that a common characteristic is that all existing game based solutions employ the parallel sensing strategies.

1) Static game: The basic model of non-cooperative games is static game, in which the game is played only once. In this sense, a static game is also referred to as an one-shot game. In quasi static or slowly time-varying environment, e.g., the available channels remain unchanged for a long duration, and in this case OSA problems can be well modeled and analyzed as static games. Other interpretation of static games in opportunistic spectrum access systems is that it is designed only for one slot.

In [89], N. Nie et al. formulated the problem of channel selection in a free-use opportunistic spectrum access system

[2]In fact, the third party can fictitiously exist and the CE can be achieved in a distributed and autonomous manner.

TABLE I
SUMMARY OF GAME MODELS IN OSA SYSTEMS.

Game type	Game model	Objective	Spectrum sharing model	Spectrum dynamics	Behavior update rule	Solution	Ref.
Non-cooperative	Static game	**Interference avoidance–** choose a channel to minimize the aggregate interference	Free-use	Static	Φ-no-regret learning	NE	[89]
		Price of Anarchy Characterization	Free-use	Static	—	NE	[90]
		Interference avoidance– choose a channel to minimize the aggregate interference	Free-use	Static	Best response & Spatial adaptive play	NE	[91]
		Throughput maximization– choose transmission rates over different channels to maximize the achievable throughput	Free-use	Static	Regret matching	CE	[95]
		Global utility maximization– choose idle channels to maximize the satisfaction level of the worst-off user	Share-use	Static	Regret tracking	CE	[55]
	Repeated game	**Throughput maximization–** choose a channel to maximize the expected throughput with contention overhead consideration	Share-use	Time-varying	Stochastic learning automata	NE	[52]
		Throughput maximization– choose a channel to maximize the expected throughput	Free-use	Static	Reinforcement learning	NE	[96] [97]
	Graphical game	**Throughput maximization–** choose a channel to maximize the achievable throughput	Free-use	Static	Regret minimization	NE	[100]
		Throughput maximization– choose a channel to maximize the achievable throughput	Share-use	Static	Exponential learning	ESS	[101]
		Collision level maximization– choose a channel to minimize the collision level	Share-use	Static	Best response	NE	[102]
		Throughput maximization & Collision level Minimization	Share-use	Static	Spatial adaptive play	NE	[56]
		Congestion minimization– choose a resource to minimize the congestion level	Free-use	Static	Best response	NE	[99] [103] [104]
	Evolutionary game	**Throughput maximization–** choose a channel to maximize the achievable throughput	Share-use	Time-varying	Replicator dynamic & A distributed algorithm	ESS	[111]
Cooperative	Coalition game	**Jointly improve sensing and access performance**	Share-use	Static	Coalition formation algorithms	–	[114]

as a static game. They proposed two utility functions: the first is the aggregate interference experienced by a SU and the second is the aggregate interference experienced by a SU plus the interference it caused to other SUs. The first utility function is for selfish users while the second is for cooperative users. In particular, with the second proposed utility function, the channel selection game admits a potential function. The proposed behavior update rule therein is Φ-no-regret learning, which converges to mixed strategy NE for the first utility function and pure strategy NE for the second utility function.

In [90], L. Law et al. studied the price of anarchy (PoA) of channel selection game for a free-use OSA system. PoA, which is an important metric of non-cooperative games, is defined as the ratio between the aggregate payoffs of all the players in the worst NE point and the social optimum. They derived closed-form expressions of PoA for both symmetric and asymmetric games, and presented several insights to improve the PoA of channel selection games for OSA systems. It is noted that PoA is an inherent feature of non-cooperative games, regardless of the used behavior update rule.

In our earlier work [91], we re-studied the greedy asynchronous distributed interference avoidance (GADIA) algorithms, which were originally proposed in [92] and investigated from a non-cooperative game theoretic perspective therein. The problem of channel selection for a free-use OSA system was formulated as a potential game in [91] and some incremental results in favor of [92] were obtained. Specifically, it is shown that the basic and soft GADIA algorithms proposed in [92] correspond to the best response [87] and spatial adaptive play [93], [94] respectively.

The solution concept of correlated equilibrium for opportunistic spectrum access systems was firstly introduced by Z. Han et al. in [95], where they formulated the problem of rate adaptation over different channels for a free-use OSA system as a static game. A distributed channel algorithm named regret matching [88] was proposed to converge to the set of correlated equilibria.

Another static game model with CE solution for distributed channel selection in a share-use OSA system was proposed by M. Maskery et al. in [55]. The payoff of a user is defined as the proportion of received resources divided by its demand, which is also interpreted as satisfaction level. Interestingly, although the formulated game is non-cooperative, the proposed behavior update rule, i.e., the regret tracking, achieves cooperative design solution. Furthermore, it should be pointed out that although some simulation results for dynamic spectrum opportunities were presented therein, the considered spectrum opportunities are actually static, or varying slowly in time.

2) Repeated game: Unlike static games, a repeated game is repeatedly played in finite or infinite horizon. The players

update their strategies based on their action-payoff history in previous plays. Recalling the illustrative diagrams of opportunistic spectrum access systems, as shown in Fig. 1, it is seen that the SUs always make decisions repeatedly from slot to slot. This characteristic makes repeated games suitable for modeling and analyzing opportunistic spectrum access systems in long-run time.

H. Li formulated the problems of distributed channel selection in free-use OSA systems as repeated games. The authors proposed reinforcement learning based channel selection algorithms for two-user two-channel systems in [96] and multiuser multichannel systems in [97]. It is shown the reinforcement learning based algorithms converge to Nash equilibria of the game. However, they only considered static spectrum, which is not always true in opportunistic spectrum access systems.

Recently, the problem of distributed channel selection for a share-use OSA system with time-varying spectrum environment was considered in our earlier work [52]. We formulated the problem as a repeated with random payoffs. An interesting result presented in this work is that although the spectrum is varying from slot to slot independently with unknown statistics, a learning algorithm called stochastic learning automata is proposed to converge to NE of the formulated repeated game.

3) Graphical game: A graphical game, which is also called as a local interaction game [98] or spatial game [99], is characterized by: the action of a player only affects its neighboring players rather than all other players. In a wide area opportunistic spectrum access system, the transmission from a SU only causes interference to the nearby SUs and hardly interferes with distant players. In fact, it leads to the so-called spatial reuse in wireless communication systems. Recently, graphical games have been applied to opportunistic spectrum access systems as it captures the limited range of mutual impact of the SUs.

The graphical game was firstly introduced into opportunistic spectrum access systems by H. Li et al. in [100], where they considered a free-use system and proposed regret-minimization algorithms to converge to NE. In a subsequent study [101], M. Azarafrooz et al. considered graphical game for a share-use OSA system and proposed an exponential learning algorithm to converge to evolutionary stable strategy (ESS), which is the solution concept of evolutionary games [106].

Motivated by the game formulation in [100], we further studied graphical games for share-use OSA systems in our earlier work [102] and [56] respectively. The objective considered in [102] is minimizing the collision level, while those considered in [56] include both maximizing the throughput and minimizing the collision level. The formulated graphical games in [56], [102] are potential games. The behavior update rule in [102] is best response, which is averagely suboptimal, and that in [56] is spatial adaptive play [93], [94], which was shown to be asymptotically optimal with local information exchange.

In another study series [99], [103], [104], the authors also formulated graphical games for free-use OSA systems. The games are called spatial congestion games therein, since the payoff of a player is a function of the number of players who interact with it and use the same resource (channel), which is similar to that in traditional congestion game [105]. Compared with [56], [102], the focus of [99], [103], [104] is to investigate conditions under which the spatial games possess a pure strategy NE.

4) Evolutionary game: Evolutionary game theory was first introduced by biologists studying population dynamics [106]. The solution concept of evolutionary games is the evolutionary stable strategy (ESS), which was first defined in [107]. ESS is characterized by robustness against invaders (mutations): i) the proportions of each population remains unchanged, as far as an ESS is reached, and ii) at ESS, the populations are robust to perturbations by a small fraction of players. Evolutionary games are being applied to wireless communication systems and several related work can be found in the literature, e.g., an evolutionary game framework for power control and multiple-access control [108], cooperative spectrum sensing [109], and network selection in heterogeneous networks [110].

Very recently, X. Chen et al. formulated the problem of channel selection in a share-use OSA system as an evolutionary game in [111]. The spectrum opportunities are time-varying with Bernoulli distribution in each slot. With complete network information, the replicator dynamics was applied to converge to an ESS. More importantly, a distributed learning mechanism with incomplete network information was also proposed to converge to an ESS. This work provides several insights in terms of the interactions among SUs and the dynamics of channel selections, and hence would draw great attention in the near future.

5) Coalition game: In essence, coalitional games involve a set of players who seek to form cooperative groups, i.e., coalitions, to achieve increased payoffs. A good tutorial of coalition game theory for communication networks can be found in [112]. The formation of coalitions is ubiquitous from human society to wireless communications. For example, countries can form coalitions for improving their human potential while SUs can form coalitions for improving the spectrum sensing performance [113].

W. Saad et al. proposed a coalition game in partition form to study the problem of spectrum sensing and access in a share-use OSA systems in [114]. The proposed coalition approach is promising in three aspects: (i) the SUs in a coalition sense different channels and share their sensing results to reduce the sensing times, which eventually leads to increased throughput, (ii) the SUs in a coalition jointly coordinate the channel access order to reduce the mutual interference, and (iii) the SUs in a coalition share their instantaneous sensing results to improve capacity by distributing their total power over multiple channels.

Remark 3: In addition to the above presented game models, a large numbers of papers using game theory to study OSA technologies from economic perspectives can be found in the literature [115]–[121]. Specifically, game models for spectrum trading between SUs and PUs were formulated in these studies. We do not analyze these studies, since they are not in the scope of channel selection.

C. Discussion of hierarchical game models

It is noted that the players in the above reviewed game models are treated equally with no hierarchy, which is caused

TABLE II
INFORMATION REQUIREMENT OF BASIC LEARNING ALGORITHMS IN GAME MODELS.

Information	Best response	Better response	Fictitious play	Regret matching	Spatial adaptive play	Reinforcement learning	Learning automata
a_n action of user n				✓		✓	✓
a_{-n} actions of other users	✓	✓	✓	✓	✓		
$u_n(a_n, a_{-n})$ received payoff of user n		✓	✓	✓	✓	✓	✓
$u_n(a'_n, a_{-n}), a'_n \neq a_n$ payoff for unchosen actions of user n	✓	✓	✓	✓	✓		

by the completely distributed structure of the considered system. However, the system architecture of the opportunistic spectrum access systems is hierarchic in essence, i.e., the secondary users use the idle spectrum of the primary users. Thus, it is interesting to include the primary users into the game by considering the utilities of both secondary users and primary users. To achieve this, it needs to introduce hierarchy into the games. In game theory, some approaches utilize the hierarchical game models. The most important one is Stackelberg game [122]–[124], which consists of a leader and several followers competing with each other on certain resources. The leader takes an action first and the followers take actions subsequently. The solution concept is the Stackelberg equilibria from which neither the leader nor the followers have incentives to deviate.

In Stackelberg game, the leader maximizes its utility function and so do the followers. However, in some scenarios, the leaders has no individual utility and its goal is just to maximize the aggregate utility of the followers. To cope with this problem, J. Park et al. [125]–[129] recently proposed a new hierarchal game in which the leader first chooses an intervention rule and then the followers choose their actions according to the intervention rule. They are called intervention games and more suitable than Stackelberg games when the leader is not a resource user but a manager who regulates resource shared by followers. Essentially, intervention game is a variant of Stackelberg game.

It is seen that Stackelberg games have been widely applied to opportunistic spectrum access systems for several research issues, e.g., power control [130], [131], bandwidth allocation [132], [133], and joint power control and bandwidth allocation [134], [135]. The investigations in these references are informative since it has been shown that the Stackelberg equilibria can improve the efficiency of Nash equilibria significantly. Although the application of hierarchical games for distributed channel selection has not been reported yet, we believe that the hierarchical control models are promising and would be applied to solve the distributed channel selection problem for opportunistic spectrum access systems.

D. Information requirement of learning algorithms in game models

As stated before, the task of achieving stable solutions, e.g. NE or CE, is important for practical implementation; in particular, different learning algorithms require different information. In this subsection, we present information requirement of basic learning algorithms in game models. The comparison results are shown in Table II. We divide these algorithms into two groups: *coupled* and *uncoupled*. Specifically, the former needs information about other players in terms of chosen actions and/or payoffs, while the latter only needs local information of a player. For the presented algorithms in the table, coupled algorithms include best (better) response [87], fictitious play, regret matching [88] and spatial adaptive play [56], while uncoupled algorithms include reinforcement learning [96] and learning automata [52]–[54].

For wireless systems, uncoupled algorithms are more preferable than coupled algorithms, since obtaining information about other players cause heavy communication overhead or is not even feasible in some scenarios. However, it is noted from Table II that most existing game learning algorithms in the literature belong to coupled algorithms. The reasons are twofold: (i) coupled algorithms have been well investigated in pure game theory and (ii) it is generally hard to develop uncoupled algorithms that converge to some stable solutions for game models. Clearly, uncoupled game learning algorithms are desirable in OSA systems and should be studied in future.

E. Strengths and limitations of game models

Based on the review and comparison results of game models in OSA systems, its strengths can be summarized as follows:

1) Game models provide an efficient framework for capturing the interactions among multiple SUs. With game formulations, the system steady states can be well predicted and achieved by learning algorithms. Also, the performance of steady states can be analytically characterized.
2) Game models yield flexible design. As has been shown before, utility function is key to game models, as it determines the structure and the steady state of the game. Thus, some new utility functions have been proposed to guarantee the existence of game solutions (e.g., NE) and their optimality. For example, an approach commonly used for utility design is minus cost or price in the received payoff. Such an idea has been extensively and successfully applied to many wireless research issues, e.g., rate adaptation [84], distributed power control [85] and multiple access control [86].

However, game models also have some inherent limitations as summarized below:

1) Game models fail to find more speculum opportunities in OSA systems. It can be seen that almost all the referenced game models employ the parallel sensing strategies, i.e., sense a fixed number set of channels in

a slot. Such strategies enjoy convenience in analysis but also lead to conservative throughput. In a scenario where the chosen channels are busy in a slot, the SU has to suspend its transmission until the next slot. However, there is a probability that channels which are not chosen to sense may be idle in the slot. Thus, an alternative and desirable approach is that the SU does not limit itself to a fixed number of channels; instead, it proceeds to sense the residual channels to find more spectrum opportunities in the current slot.

2) Most dynamic game models need to know the statistical information about the spectrum environment, e.g., the idle probabilities of licensed channels with independent and identical distributions or the transition probabilities of Markovian channels. Relying on them, repeated games or stochastic games can be formulated to cope with the spectrum dynamics. However, the statistical information of the spectrum is always not known a priori, which makes traditional game models not workable in the absence of statistics information.

IV. The Application of Markovian Decision Process in Opportunistic Spectrum Access

Markov decision process (MDP) models, which were introduced in 1960 [136], are mainly used to analyze and solve a sequential decision making problem with multi-periods in Markovian environment. MDP models have been studied extensively and successfully applied into several engineering fields [137], e.g., telecommunication, signal processing, artificial intelligence and economics. In OSA systems where the activities of PUs evolve in a Markovian stochastic manner, the spectrum sensing and channel selection strategies can be naturally formulated as a MDP problem. In this section, we introduce the basic models of MDP, review and compare existing MDP models in OSA systems, and finally discuss their strengths and limitations.

A. Basic models of Markovian decision process (MDP)

There are three basic branches of MDPs which have been commonly used in communication systems: the basic discrete time MDP (DTMDP), partially observable MDP and constrained MDP. A DTMDP model is formally defined by the following elements:

- Discrete time $k = 0, 1, 2, \ldots$
- A discrete set of countable states $s \in S$.
- A discrete set of countable actions $a \in \mathcal{A}$.
- A reward function $R : S \times A \longmapsto \mathbb{R}$ indicating the received reward $R(s, a)$ when it takes action a at state s.
- A stochastic transition model $p(s'|s, a)$ indicating the probability that the system will transfer to state s' in the next period when the player takes action a at state s.

The goal of the player is to find an optimal policy $\pi(s)$ mapping states to actions so as to achieve the optimal value function of state s, which is defined as the maximum expected aggregate discounted reward starting from state s, i.e.,

$$V^*(s) = \max_{\pi} \mathbf{E}\Big[\sum_{k=0}^{\infty} \gamma^k R(s_k, a_k) | s_0 = s, a_k = \pi(s_k)\Big], \tag{4}$$

where $\mathbf{E}[\cdot]$ is the expectation operation, $\gamma \in [0, 1)$ is the discount factor. Similarly, the optimal Q-value function of state-action pair (s, a) is the maximum discounted future reward the player can receive after taking action a in state s:

$$Q^*(s, a) = \max_{\pi} \mathbf{E}\Big[\sum_{k=0}^{\infty} \gamma^k R(s_k, a_k) | s_0 = s, a_0 = a, a_{k>0} = \pi(s_k)\Big]. \tag{5}$$

It has been proved that a DTMDP has an optimal policy which is deterministic and stationary. Deterministic means that $\pi^*(s)$ specifies a single action per state, while stationary means that every time the user observes a state s, the optimal action is always $\pi^*(s)$. The optimal policy is given by:

$$\pi^*(s) = \arg\max_{a \in \mathcal{A}} Q^*(s, a), \tag{6}$$

where the optimal Q-value $Q^*(s, a)$ of each state-action pair (s, a) can be calculated by the value iteration method [136] or Q-learning [138]. For specific solutions for DTMDP, refer to [136]. For comparison, it should be pointed out that in DTMDP models, the system state is completely observed by the player in each decision period.

Different from the basic DTMDP model, the state information in a partially observable MDP (POMDP) model is only partially observed by the player in each decision period. Thus, the internal state of the underlying Markov process is unknown in POMDP problems. To address this, the player needs to construct a belief vector which is the conditional probability (given the decision and observation history) that the system is in each state and update it after each decision. Based on the maintained belief vector, the user would find an optimal policy. For detailed discussion on the solutions of POMDP, refer to [139].

In addition, unlike DTMDP and POMDP where no constraint is imposed, a constrained MDP (CMDP) is to model and analyze sequential decision problems with constraints. This model is very useful since there are always constraints that must be met in practice, e.g., the interference requirement imposed by the PUs. For detailed discussion on solutions for CMDP, refer to [140].

B. The application of MDP in opportunistic spectrum access systems

Experimental measurement results [50] have validated that the activities of PUs can be approximately characterized by discrete-time or continuous Markovian processes. This feature makes MDP models efficient and promising solutions in OSA systems. Table III provides an overview of existing MDP models for different OSA scenarios, in terms of action, optimization objective, sensing strategy and sensing reliability. In all the referred MDP models, the system state is defined as the speculum occupancy of PUs.

1) DTMDP: In [141], S. Yin et al. exploited spectrum correlation between different channels [142] and proposed a new metric, i.e., channel availability vector, to characterize the state information by spectrum prediction. Due to prediction error, collision with PU is not inevitable. Based on the real-time prediction results, the user decides to sense which

TABLE III
SUMMARY OF MDP MODELS IN OSA SYSTEMS

Models	Action	Objective	Traffic of PU	Sensing Strategy	Sensing Reliability	Ref.
DTMDP	Sense which channels or not to sense	Maximize throughput subject to collision constraints	Slotted	Sequential sensing	Perfect	[141]
CMDP	Transmit or not after each channel sensing	Maximize throughput subject to collision constraints	Continuous	Sense all channels periodically	Perfect	[143]
	Determine which channel to sense	Maximize throughput subject to collision constraints	Continuous	Sense one channel in a slot	Perfect	[144]
	Determine which channel to sense	Obtain maximum throughput region of multiuser OSA systems subject to collision constraints	Continuous	Sense one channel in a slot	Perfect	[36]
	Determine which channel to sense	Maximize throughput subject to collision constraints	Slotted	Access one channel in a slot	Perfect	[145]
POMDP	Determine which channel to sense	Maximize throughput subject to collision constraints	Slotted	Sense one channel in a slot	Perfect and Imperfect	[15]
	Determine which channel to sense	Maximize throughput subject to collision constraints	Slotted	Sense one channel & Sense multiple channels in a slot	Imperfect	[146]
	Determine which channel to sense	Maximize throughput subject to collision constraints	Slotted	Sense one channel in a slot	Imperfect	[147]
	Determine the sensing duration in a slot	Maximize net reward with energy consideration	Slotted	Sense one channel in a slot (there is only one channel)	Imperfect	[148]
	Sense which channel or to sleep	Maximize throughput during the battery lifetime	Slotted	Sense one channel in a slot	Perfect	[149]

channels and when to stop sensing, aiming to maximize the throughput while satisfying the collision probability with the primary user. Interestingly, although the activities of the primary users are independent from slot to slot in each channel, the authors formulated a MDP model by using the channel availability vector. In this work, the sensing overhead (time) is explicitly considered and the SU may sense multiple channels sequentially in a slot, which is determined by the channel availability vector and the collision probability requirement.

2) CMDP: Q. Zhao et al. [143] proposed a novel periodic channel sensing strategy. The traffic of the PU is modeled as a continuous-time Markovian process and the SU senses only one channel in a slot. In order to get full observation of the channels, the SU is designed to sense all the channels periodically. The action of the SU is determine whether to transmit after each sensing or not. Since the activities of the PUs are continuous, collision with the PU is inevitable. As a result, the authors formulated a CMDP model whose objective is to maximize the throughput while satisfying the collision probability requirement. Also, two simple heuristic algorithms, i.e., memoryless access and greedy access, were proposed to achieve suboptimal solutions.

Another CMDP model for opportunistic spectrum access systems was proposed by X. Li et al. [144], where the traffic of PU is also modeled as a continuous Markovian process. The SU senses one channel in a slot and the objective is to maximize the throughput subject to collision constraints imposed by the PU. The heuristic memoryless access (MA) algorithm proposed in [143] was also applied to solve the formulated CMDP problem. It is shown that the MA algorithm is optimal when the collision constraints are tight. In another recent study [36], where the considered scenario is similar to that in [144], the maximum throughput region of a multiuser OSA system was obtained. Specifically, the authors established inner and outer bounds for the maximum throughput region respectively. An interesting and promising result in [36] is that when collision constraints are tight, the outer and inner bounds match.

In the above references, the PU traffic is modeled as a continuous-time Markovian process and the interference fraction with PU were naturally formulated as the constraints. A drawback is that the QoS of the SU was not considered. Recently, D. Niyato et al. [145] also formulated the channel selection problem as a CMDP model in cognitive vehicular networks with QoS support, where the PU traffic is slotted and the considered constraints include not only the maximum probability of collision with PU, but also the maximum packet loss probability and the maximum packet delay for the vehicular nodes. Since both share-use and exclusive-use channels were considered, a CMDP for opportunistic spectrum access and another CMDP for exclusive-use channel reservation and clustering control form a hierarchical MDP model. The former was summarized in Table III.

3) POMDP: The pioneer work related to the application of POMDP into opportunistic spectrum access systems was in [15], which kindled great attention later. The traffic of PU is modeled as a slotted Markovian process therein. The SU can sense one channel in a slot, which makes the system state partially observable. Both perfect and imperfect spectrum sensing were considered. For perfect sensing scenario, the objective is to maximize the throughput, while that for imperfect sensing scenario is to maximize the throughput subject to collision constraints. The main contribution of [15] is that a POMDP framework was established and a simple but efficient algorithm was proposed. The methodology proposed there is constructive.

Another important existing work related to POMDP model is [146], which is a significant improvement in favor of [15]. The objective therein is also to maximize the throughput subject to collision constraints. In particular, the joint design of an sensing and access strategy is formulated as a constrained POMDP. The authors established a separation principle which reveals that the optimality of myopic policies for the design of the sensing strategy and the access strategy lead to closed-form optimal solutions. The spectrum sensing is imperfect and the SU can either sense one channel or multiple channels simultaneously in a slot.

J. Unnikrishnan et al. [147] considered a similar but more practical scenario also using a POMDP model. In particular, they investigated two scenarios where the channel availability

statistics are known or unknown. When the channel availability statistics is unknown, a greedy channel selection algorithm was proposed and an upper bound on the performance of the optimal policy was derived. When unknown, an algorithm is developed to learn the true statistics while guaranteeing the collision constraints.

Different from the above studies, A. Hoang et al. [148] considered energy consumption for spectrum sensing and also formulated a POMDP model to solve the sensing and access control problem. Only one channel is considered in this work. The achievable net reward in a slot is jointly determined by the channel idle probability and the sensing duration, which determines the mis-detection probability, false alarm probability and energy consumption. As a result, the objective is to decide the sensing duration in each slot (zero duration corresponds to the event of not sensing) to maximize the net reward.

Similarly, Y. Chen et al. [149] considered energy consumption for both spectrum sensing and data transmission. Two operating modes of the SU are considered therein: one is sleeping, i.e., the SU does nothing, and the other is sensing, i.e., the SU chooses one channel to sense. Both continuous (saturated) and bursty traffic models of the SU are considered. The objective is to maximize the throughput during its battery lifetime using a POMDP model. Note that the formulation for energy consumption in this work is interesting and should be considered in future research.

Remark 4: The above reviewed solutions are for single user systems. For multi-user systems, Markovian games, which is also referred to as stochastic games [28], [51], [150], can be applied. This will be discussed in Section VII.

C. Strengths and limitations of MDP

The most attractive strength of MDP models is that the Markovian dynamics of spectrum environment can be well modeled and analyzed, as all the above referred work focused on them. In addition, it should be pointed out that the considered spectrum environment in the opportunistic spectrum access systems are *non-reactive*, i.e., the system evolution is not affected by the actions taken by the SU. More specifically, the state transition probability $p(s'|s,a)$ is degraded to $p(s'|s)$, where a in the taken action, s and s' are the system states in the current slot and the next slot respectively. This feature eventually simplifies the analysis of the MDP models in OSA systems.

On the other hand, the limitations of MDP models in OSA systems are summarized in the following:

1) MDP models are more suitable for single SU systems rather than multiple SU systems. The reason is that in order to obtain the optimal policies of a MDP model, it requires the environment to be stationary [136]. However, in a multiple SU system, other SUs naturally serve as parts of the environment facing one SU, which implies that the environment is non-stationary. Although some existing work, e.g., [15], [146], [147], consider multiple SUs explicitly, their solutions are originally designed for single SU and the interactions among multiple SUs on the spectrum access problem are not well considered yet.
2) Most MDP models only seek for idle channels but not consider the channel quality. However, only finding an idle channel is not enough, since an idle channel may have poor transmission quality. Thus, one may want to jointly consider the occupancy state of PU and the channel quality of the data channel to formulate new MDP models with channel quality consideration.
3) MDP models need to know the transition probabilities of the Markovian process. In practical OSA systems, however, such statistical information are always unknown a priori. In the absence of these statistical information, the algorithms originally proposed for solving MDP models will not function.
4) Most MDP models fail to find more speculum opportunities in OSA systems as game models, since they also employ the parallel sensing strategies.

V. The Application of Optimal Stopping Theory in Opportunistic Spectrum Access

The theory of optimal stopping is concerned with the problem of sequentially taking actions to maximize the expected reward based on a sequence of observed random variables [57]. Specifically, the observed variables determine the current reward. The achieved reward of proceeding to observe the residual variables is random and may be greater or less than the current received reward. Therefore, a rational action, i.e., stopping or proceeding to observe, should be taken after each observation. When an action of stopping is taken, the decision process is terminated. We refer to this model as optimal stopping problem (OSP) in this article. OSP models have been extensively applied to many fields [151], e.g., management science, economics and wireless communications. In this section, we firstly introduce the basic models of OSP, review and compare existing OSP models in OSA systems, and finally discuss their strengths and limitations.

A. Basic models for optimal stopping problem (OSP)

Generally, an OSP model has the following two elements:

- A sequence of random variables, $X_1, X_2, \ldots, X_N$, whose joint distribution is known and their realizations are denoted by $x_1, x_2, \ldots, x_N$.
- A sequence of functions mapping the observed random variables to real-valued rewards, i.e.,

$$y_1(x_1), y_2(x_1, x_2), \ldots, y_N(x_1, \ldots, x_N) \tag{7}$$

For presentation, suppose that you are now involving in the sequential decision problem. If you choose to stop after observing the nth variable, x_n, you will receive a known reward $y_n(x_1, \ldots, x_n)$. If you choose to proceed to observe and stop in the mth variable, $m > n$, the achieved reward in the future, i.e., $y_m(x_1, \ldots, x_n, \ldots, x_m)$, is random and unknown. This is naturally due to the fact that the unobserved variables $x_{n+1}, \ldots, x_m$ are random and unknown. In this circumstance, deterministic optimization is not feasible and your objective would be to choose the right time to stop to maximize the expected reward, based on the sequence of observed variables. The decision horizon N may be infinite

and finite. We only consider OSP models with finite horizon in this article, since those with infinite horizon are rare in practice.

OSP models with finite horizon can be solved by the method of backward induction [57]. Since we must stop at stage N, we first find the optimal rule at stage $N-1$. Then, we can find the optimal rule at stage $N-2$ with the known optimal rule at stage $N-1$, Inductively, the optimal rule backward to the initial stage (stage 1) can be found with the known optimal rules in future stages. Mathematically, we can define the stage value as follows:

$$V_n = \begin{cases} y_N(x_1, ..., x_N), & n = N \\ \max\left\{y_n(x_1, ..., x_n), \right. & \\ \quad \left. \mathrm{E}[V_{n+1}|\{x_i\}_{i=1}^n]\right\}, & n < N, \end{cases} \tag{8}$$

where $\mathrm{E}[V_{n+1}|\{x_i\}_{i=1}^n]$ represents the expected reward of proceeding to observe when $X_1 = x_1, X_2 = x_2, \ldots, X_i = x_i$ have been observed, and V_n represents the maximum achieved reward one can obtain starting from stage i. At stage n, we compare the current reward of stopping, i.e., $y_n(x_1, \ldots, x_n)$, and the expected reward of proceeding to observe the residual stages using the optimal rules, which at stage n is $\mathrm{E}[V_{n+1}|\{x_i\}_{i=1}^n]$. In order to maximize the expected reward, it is optimal to stop at stage n if $y_n(x_1, \ldots, x_n) \geq \mathrm{E}[V_{n+1}|\{x_i\}_{i=1}^n]$, and to continue otherwise.

There are two main branches: OSP with no recall (NR-OSP) and OSP with recall (R-OSP). In NR-OSP models, recalling a previously observed variable is prohibited and the decision maker only can access the currently observed variable, which implies that the current reward can be simplified as $y_n(x_n)$. In R-OSP models, recalling previously observed variables is allowed, i.e., at stage n the decision maker can recall x_k, $k \leq n$. For NR-OSP models, the backward induction method can be easily constructed. In contrast, the backward induction method is not feasible for R-OSP models, since it involves huge computation complexity which increases exponentially as the number of decision horizon increases. To address this challenge, it is possible to consider a truncated version of the original problem, e.g., the commonly used k-stage look-ahead rule (k-SLA). In the k-SLA rule, the expected reward of continuing to observe, i.e., $\mathrm{E}[V_{n+1}|\{x_i\}_{i=1}^n]$, is replaced by that of continuing to observe the following consecutive k stages and then stop. The simplest truncated rules, i.e., 1-SLA, are quite good in general; moreover, they are optimal for monotone R-OSP models [57].

OSP models have been extensively applied to wireless communication systems, which always encounter sequential decision problems. For example, in a system exploiting multi-channel diversity, the user would explore state information of the channels and then choose one with the strongest quality for transmission. However, increasing the number of explored channels definitely increases both achievable diversity and exploration cost; in addition, the state information of unexplored channels are random and unknown until being explored. Clearly, such a problem can be formulated and analyzed by OSP models. There are several existing work related to the application of OSP models in wireless communications, e.g., distributed opportunistic scheduling for random access control [152], opportunistic relaying control [153], traffic scheduling for vehicular delay tolerant networks [154] and opportunistic scheduling for multiuser diversity [155].

B. The application of OSP in opportunistic spectrum access systems

As stated before, the sequential sensing strategy is more efficient than the parallel one, since it determines the sensing channels adaptively based on the sequential observation results. Mathematically, sequential sensing strategies can be well analyzed by OSP models and the ultimate objectives are to achieve good balance between the achieved performance and sensing cost. In order to comply with OSP formulations, one needs to first clarify the random variables in opportunistic spectrum access systems. Specifically, the random variables in the share-use OSA systems include the occupancy state of PUs, a pair consisting of the occupancy state of PUs and the instantaneous channel quality, while those in the free-use OSA systems mainly include the instantaneous channel quality. In this subsection, we review and compare existing OSP models in opportunistic spectrum access systems as given in Table IV. The review and comparison results are in terms of actions, objectives, spectrum sharing models, and observed random variables.

1) NR-OSP: A. Sabharwal et al. [45] investigated the problem of balancing the tradeoff of finding a good quality channel (band) and time overhead of exploring multiple channels in a free-use OSA system. The considered random variable therein is the instantaneous channel quality. Specifically, the channels are assumed to undergo Rayleigh block-fading identically and independently. The fading statistics, i.e., the mean value of the Rayleigh fading, is known. It is not allowed to use channels which have already been visited, which naturally leads to the formulated NR-OSP model therein. The taken actions are to use the current channel or to explore the residual channels, and the objective is to find the optimal stopping rule to maximize the expected throughput. Since the channels experience homogeneous fading, exploration order can be arbitrarily picked out and hence was not considered.

In [156], H. Jiang et al. studied the sensing order with optimal stopping rule for a share-use OSA system. The considered random variables therein are the occupancy state of PU and the instantaneous channel quality. Their formulated model also belongs to NR-OSP. An interesting phenomenon found by the authors is that the intuitive sensing order, i.e., descending order of the channel availability probabilities, is not optimal even for scenarios with homogeneous channel fading. Thus, they proposed a dynamic programming method to find an optimal sensing order. In addition, the scenario of not knowing the channel availability probabilities was considered. However, it should be pointed out that the proposed dynamic programming method involves heavy computational complexity.

The sensing order with optimal stopping for a share-use OSA system was also investigated in another work [41]. The channel transmission rates are assumed to be fixed and the observed random variable therein is the occupancy state of PU. Their formulated model also belongs to NR-OSP. An interesting result is that it is optimal to sense the channels according to the descending order of their achievable rates with optimal

TABLE IV
SUMMARY OF OSP MODELS IN OSA SYSTEMS

Models	Actions	Objective	Spectrum Sharing Model	Observed Random Variables	Ref.
NR-OSP	**A1.** Use the current channel **A2.** Proceed to sense the residual channels	Find the optimal stopping rule to maximize expected throughput	Free-use	Channel quality	[45]
	A1. Use the current channel **A2.** Proceed to sense the residual channels	Find the optimal sensing order with optimal stopping rule to maximize expected throughput	Share-use	Channel quality and occupancy state of PU	[156]
	A1. Use the current channel **A2.** Proceed to sesne the residual channels	Find the optimal sensing order with optimal stopping rule to maximize expected throughput	Share-use	Occupancy state of PU	[41]
	A1. Use the current channel **A2.** Proceed to sesne the residual channels	Find the optimal stopping rule, power allocation strategy and sensing order to maximize throughput normalized by the aggregate energy consumption	Share-use	Channel quality and occupancy state of PU	[157]
	A1. Use the current channel **A2.** Proceed to sense the residual channels. **A3.** Release the current accessed channel.	Find the optimal rule for the formulated two-dimensional OSP model.	Free-use	Channel quality	[158]
R-OSP	**A1.** Use observed idle channels **A2.** Proceed to sense the residual channels	Find the optimal stopping rule to maximize expected throughput	Share-use	Occupancy state of PU	[40]
	A1. Use an sensed channel **A2.** Use an unsensed channel **A3.** Proceed to sense the residual channels.	Find the optimal control rule to maximize expected throughput	Free-use	Channel quality	[164]
	Stop sensing or not after each sampling	Find the optimal control rule to maximize expected throughput	Share-use	The received cumulative energy	[42]
	A1. Use an sensed channel **A2.** Proceed to sense the residual channels	Find the optimal stopping rule to maximize throughput normalized by the aggregate energy consumption	Free-use	Channel quality	[44]
	A1. Use an sensed channel **A2.** Proceed to sense the residual channels	Find the optimal control rule to maximize expected throughput	Share-use	Channel quality and occupancy state of PU	[165]

stopping, regardless of the channel availability probabilities. In comparison, [41] is differentiated from [156] in two aspects: (i) instantaneous channel fading is not considered, and (ii) the channel availability probabilities do not affect the sensing order and the optimal stopping rule.

Recently, besides time overhead of exploring multiple channels, energy overhead also begins to draw attention. Y. Pei et al. [157] jointly considered the optimal stopping rule, power allocation and sensing order for a share-use OSA system with energy consumption consideration. The spectrum sensing is imperfect. The observed random variables are the channel quality and the occupancy state of PU. The optimal stopping rule and sensing order are obtained by a common dynamic program method.

In all the above referenced OSP models, the SU releases the accessed channel for a pre-defined duration and then re-starts the sequential sensing procedure in the next decision epoch. These models can be regarded as opportunistic spectrum access in time domain. Very recently, B. Li et al. [158] formulated the sequential sensing and access problem for a free-use OSA system as a two-dimensional NR-OSP model, by investigating this problem in time-frequency domain. Specifically, they considered finite-state Markovian channels and formulated a novel NR-OSP model for determining when and which channel to access and when to release it. There are three actions: (i) use the current channel, (ii) proceed to sense the residual channels, and (iii) release the current accessed channel. This study is a significant improvement of other NR-OSP models in OSA systems. Also, the authors have investigated this problem extensively and some interesting results were reported in [159]–[163].

2) R-OSP: J. Jia et al. [40] investigated the problem of balancing the tradeoff of finding idle channels and time overhead of sensing multiple channels in a share-use OSA system. The observed random variable is the occupancy state of PU. In this work, the SU can use the previously and currently observed idle channels, which naturally leads to an optimal stopping problem with recall (R-OSP). As stated before, finding an optimal rule for an R-OSP model is generally challenging, which motivated the authors to apply the k-SLA rules, where $k = 1, 2$ are considered. It is shown that the used 1-SLA rule is quite good since its performance is close to that of the optimal solution. In addition to the theoretical analysis, the authors also proposed a MAC protocol to implement the proposed channel sensing algorithm.

Another important existing work addressing the application of R-OSP model in free-use OSA systems is [164]. Different from other existing work, there are three actions: (i) use an observed channel, (ii) use an unobserved channel, and (iii) sense unobserved channels. With the formulated R-OSP model, the authors proposed a dynamic program to compute the optimal strategy within a finite number of steps, and also applied the 2-SLA rule to achieve a suboptimal but more efficient solution.

S. Kim et al. [42] formulated a novel R-OSP model in a share-use OSA system. Interestingly, although an R-OSP model is formulated, the SU in this work adopts a parallel sensing strategy rather than a sequential one. The spectrum sensing is imperfect and the observed variable in this work is the received accumulative energy. As the sampling duration increases, the sensing performance improves, which implies more spectrum opportunities for the SU. However, increasing sampling duration definitely results in a decrease in the transmission time. Thus, there is a fundamental tradeoff between sampling duration and achievable throughput. To adhere to the collision constraints imposed by the PU, a constrained dynamic programming problem is proposed to obtain the optimal rule that chooses the best time to stop sensing and

the best set of channels to access. Suboptimal algorithms have also been proposed to avoid computational complexity.

It should be pointed that the above R-OSP models only considered time consumption and ignored other kinds of cost, e.g., energy consumption. Recently, study on both time and energy overhead of exploring multiple channels can be found in our earlier work [44], where an energy-efficient channel exploration problem using an R-OSP model is formulated. The involved random variable is the instantaneous channel quality. The objective is to maximize the throughput normalized by the aggregate energy consumption for channel exploration and data transmission. It is proved that the 1-SLA rule is optimal for this problem. We also formulated an R-OSP models for a share-use OSA system in [165]. Besides the 1-SLA rule for single SU systems, a stochastic recalling algorithm for multiple SU systems was also proposed to alleviate collisions among SUs. In addition, we also applied the 1-SLA rule proposed in [44] to study the tradeoff between channel exploration and exploitation in spectrum sharing systems in [166] and [167] respectively, where the interference temperature constraints imposed by the primary users are explicitly considered.

C. Strengths and limitations of OSP

Based on the review and comparison results of existing work, the strengths of OSP models in OSA systems can be summarized as follows:

1) Adaptive and efficient spectrum opportunity discovery. Using OSP models, the SU would discard the parallel sensing strategy, i.e., sensing a fixed number of licensed channels in a slot; instead, it adaptively senses the channels with overhead consideration. Such a design leads to more efficient and adaptive spectrum opportunity discovery mechanisms.
2) Exploiting instantaneous channel fading to achieve multichannel diversity. Traditionally, the unreliability caused by the time-varying fading needs to be mitigated. With the OSP models, however, such channel fluctuations can assist to find a strong channel for transmission. In fact, it can be regarded as multichannel diversity with overhead consideration, which is a more practical version of those without overhead consideration.

However, OSP models also have some inherent limitations as summarized below:

1) All the OSP models need to know statistical information about the channels. In order to calculate the expected payoff of proceeding to sense the residual channels, the SUs need to know the statistics information of the channels, e.g., channel idle probabilities. However, such statistical information may not be available a priori in some scenarios.
2) The OSP models, including both NR-OSP and R-OSP, are more suitable for single SU systems rather than multiple SU systems. In multiple SU systems, an unexplored channel may have been occupied by other SUs. As a result, the achievable reward of continuing to explore the residual channels can not be calculated and would be less than that of assuming only one SU in the system, which brings about new challenges for the OSP models with multiple users. It should be pointed out that although some existing studies, e.g., [41], [165], have considered multiple SUs with numeric simulation, theoretical analysis for optimal stopping rules for multiple SU systems have not been studied yet.
3) The NR-OSP models admit mathematical tractability but lead to relatively conservative design. With the constraint of no recalling, i.e., it is not allowed to use a previously explored channel, optimal stopping rules can be easily obtained by backward induction methods. However, it may miss some transmission opportunities in the recently explored channels, even though they are idle and good.
4) The R-OSP models may suffer from outdated channel state information, especially in a system with large number of channels. Outdated channel state information [168] means that the channel state at the moment of accessing is generally correlated with but different from that at the moment of observation. The larger the interval between observation and accessing, the lower the correlation coefficient between the channel status. Thus, recalling an observed channel which is far away from the current channel may not achieve the expected reward since the qualities of these channels may have became outdated.

VI. The Application of Multi-armed Bandit Problem in Opportunistic Spectrum Access

Multi-armed bandit (MAB) problem [169] is a powerful tool in online learning theory. Specifically, it can be described as: a player chooses one or more resources among several alternative candidates whose statistical information are unknown. Basically, in order to achieve desirable performance, it needs to sample the resources and collect the statistics during the decision process. Therefore, there is a fundamental tradeoff between *exploitation*, maximizing the current reward based on the current estimated statistics, and *exploration*, i.e., spending time on sampling the resources to increase the accuracy of estimated statistic with the prospect of better future rewards. It is noted that the attribute of not knowing a priori statistical information and partially monitoring in MAB models is very common in several scenarios. As a result, the MAB models have been extensively studied and successfully applied in several technological and scientific fields such as economics, manufacturing systems, control theory, search theory, communication networks, etc.

A. Basic models of multi-armed bandits (MAB) problem

The classical MAB problem can be described as a player playing with K arms. Time is divided into slots with equal length. At each slot, the player can choose any one of the arms to play and get a real-valued random reward. The received reward of playing arm $k \in \{1, 2, \ldots, K\}$ in the tth slot, which is denoted as r_{tk}, is an independent identical distributed (i.i.d.) variable following some deterministic but unknown distribution Φ_k. Denote the expected reward of playing arm k

as μ_k. The objective of MAB is to develop an learning policy to maximize the cumulative reward.

A learning policy π is defined as a function that maps previous plays and observed rewards into current decision. To evaluate the performance of a learning policy, the mostly widely used metric is *regret*, which is defined as the reward loss of the policy compared with the optimal policy under an ideal assumption that all the statistical information are known. Formally, the cumulative regret of a learning policy π until decision epoch T is give by:

$$R_\pi(T) = T\mu^* - \sum_{t=1}^{T} r_{t\pi(t)}, \tag{9}$$

where $\mu^* = \max\{\mu_k\}$ is the maximum expected reward, $\pi(t)$ is the chosen arm in epoch t, and $r_{n\pi(t)}$ is the received reward in slot n. The objective is to keep $R_\pi(T)$ as small as possible. Specifically, if it is sublinear with respective to time, the time-averaged regret will tend to zero, i.e., $\frac{R_\pi(T)}{T} \to 0$, as $T \to \infty$. Accordingly, the maximum time-averaged reward can be achieved.

This problem was originally formulated around 1940 [169]. However, no substantial progress in finding its optimal solution was made until Lai and Robbins [170] presented a general policy that provides expected regret which is in order of $O(K \log T)$, i.e., linear in the number of arms and asymptotically logarithmic in time. More importantly, they showed that this policy provides a lower bound on the expected regret, which implies that it is order-optimal and no policy can do better than the logarithmic order. Lai and Robbins's results are for the scenario of playing exactly one arm at a time, which are extended by Anantharam et al. [171] to the case of playing multiple arms simultaneously. R. Agrawal in [172] proposed sample mean based index policies for that also achieves logarithmic regret. P. Auer et al. [58] also proposed upper confidence bound (UCB) based policies, which are simpler and more general than those proposed in [172].

We briefly describe the UCB1 algorithm in [58] since a large number of its variant have been proposed for different problems in the literature. Using UCB1 policy, the arm with the highest index $\hat{\mu}_k(T) + \sqrt{\frac{2\log T}{m_k}}$ is selected at each decision period T, where $\hat{\mu}_k(T)$ is the measured expected reward of arm k until T and m_k is the number of times that arm k has been played. The first term in the index corresponds to exploitation while the second term corresponds to exploration since more chances are given to arms that are not played often. For detail procedure of solutions for the classical MAB problem, refer to these references.

Besides the i.i.d. MAB problems, in which the rewards are independently identically distributed over time, there is also a class of MAB problems with Markovian rewards [173]. Basically, Markovian MAB can be categorized into two branches: (i) rested MAB, in which the state of an arm evolves only when it is played and frozen when it is not played, and (ii) restless MAB, in which the arm state evolution is independent from taken actions and is potentially determined by some underlying mechanisms. The optimal policy for the rested MAB is an index policy of playing the arm with the highest Gittins'index [174] at each time, while that for the restless MAB is also an index policy of playing the arm with the highest Whittle' index [175].

B. The application of MAB in OSA systems

In practical opportunistic spectrum access systems, the channel availability statistics are initially unknown, e.g., in the initialization phase or when the SUs move into a new region. Moreover, due to hardware limitation, the SUs can only sense and access a part of channels (one or more) rather than all the channels at a time, which hence exhibits an attribute of partial monitoring. Due to the above two features, parallels between MAB models and opportunistic spectrum access systems can be constructed. In the following, we review and compare the applications of MAB models in opportunistic spectrum access systems, and summarize in Table V. As has been shown before, the dynamics of channel availability may be independent over time or follow a Markovian process. Accordingly, we classified existing studies into two branches: I.I.D. and restless.

1) I.I.D. MAB: It is the first time to consider a share-use OSA system with unknown channel availability statistics that was formulated as an i.i.d. MAB model by L. Lai et al. in [176]. Specifically, they considered scenarios for accessing a single channel or multiple channels at a time; also, both scenarios for single SU and multiple SUs were considered. Each channel is modeled as an arm. For the scenario with a single player, the UCB1 algorithm [58] was directly applied. For the scenario with multiple players, a naive stochastic algorithm in which the probability of choosing channels is proportional to the measured expected rewards was proposed. An extended version of this work can be found in [177].

Y. Gai et al. in [178] considered a user-channel matching problem in a free-use OSA system with a more general model of one channel offering different transmission rates for different users, and formulated a combinatorial multi-armed bandit problem. Interestingly, each user-channel matching is formulated as an arm, which is different from other work in which each channel is formulated as an arm. Since the UCB1 algorithm scales with the number of all matching profiles, which is exponential with respect to the number of channels and users, the authors proposed a modified UCB1 algorithm by utilizing the inherent correlation between different arms. The proposed algorithm is order-optimal, since it achieves regret which grows logarithmic in time and polynomially in the number of channels. Although this formulation involves multiple SUs, the algorithm is implemented in centralized manner with information exchange and cooperation. An extended version of this work can be found in [179].

A. Anandkumar et al. [43] also formulated a distributed I.I.D. MAB model for a free-use multiuser OSA system with unknown statistics. The objective is to achieve a collision-free channel selection profile over the best channels. For the scenario with single SU, the ε-greedy algorithm which was originally proposed in [58] was directly applied. It is noted that any single player MAB algorithm will lead to collisions, since all the users are prone to choose the same channel. Thus, for scenario with multiple SUs, an adaptive random UCB1 algorithm was used, where each SU randomly chooses

TABLE V
SUMMARY OF MAB MODELS IN OSA SYSTEMS

Model	Spectrum Sharing Model	Action	Arm	Learning Algorithm	Ref.
I.I.D MAB	Share-use	**A1.** Access one channel in a slot **A2.** Access multiple channels simultaneously in a slot	Channel	i. UCB1 for the scenario with a single SU ii. A naive distributed stochastic algorithm for multiuser setting	[176] [177]
	Free-use	Access one channel in a slot	User-channel matching profile	Modified UCB1 algorithm by utilizing the inherent correlation between different arms	[178] [179]
	Free-use	Access one channel in a slot	Channel	i. ε-greedy algorithm for the scenario with a single SU ii. Modified UCB1 algorithm with adaptive randomization for the scenario with multiple SUs	[43] [180]
	—	Access multiple channels simultaneously in a slot	Channel	N-parallel UCB1 algorithm based on time-division access	[181]
	Share-use	Access one channel in a slot	Channel	A decentralized policy called the synchronized learning under corrupted data	[182]
	Free-use	Access one channel in a slot	Channel	Two distributed algorithms for the scenarios with prioritized users and fair access respectively, which were based on the general UCB1 for k-th largest selection	[183]
	Free-use	Access one channel in a slot	Channel	A block-based UCB1 algorithm	[184]
Restless MAB	Share-use	Access one channel in a slot	Channel	The tiling algorithm	[186]
	Free-use	Access one channel in a slot	Channel	The regenerative cycle algorithm based on UCB1	[187]
	—	Access multiple channels simultaneously in a slot	Channel	A myopic but optimal learning algorithm	[188]

a channel only if a collision occurs in the previous slot; otherwise, it follows UCB1. This policy captures well the effect of collision among multiple SUs choosing the same channel and also has logarithmic order. This work is further extended in [180].

K. Liu and Q. Zhao [181] considered a more general problem than that in [43] and proposed another distributed I.I.D. MAB solutions. In their formulation, the players can access multiple channels simultaneously in a slot. They proposed a N-parallel UCB algorithm, where N is the number of channels. Their solutions are based on time-division access. Specifically, a pre-agreement mechanism is essentially needed to orthogonalize the players via settling them at different offsets in the time sharing schedule, which eventually diminish collisions among users. As the users are orthogonalized in time, they can follow the UCB1 algorithm directly in parallel. However, it is noted that although multiple SUs are considered in this work, the proposed learning policies are not fully distributed.

As an extension of their prior work [181], K. Liu and Q. Zhao considered minimizing the system regret for an OSA system with multiple SUs in [182]. Imperfect sensing is also considered. The proposed decentralized learning policy, called the synchronized learning under corrupted data, is shown to be order-optimal since the system regret is increased in logarithmic order. In fact, the algorithm in [182] can be regarded as a general version of that proposed in [181].

Recently, Y. Gai et al. re-considered the same scenario for their priori work [178] but formulated it as a distributed I.I.D MAB problem in [183]. They proposed a selective learning policy of the K-th largest expected reward, which is a general version of UCB1. Based on the selective learning policy, two distributed algorithms were proposed for the scenarios with prioritized users and fair access respectively.

It is perhaps the first time to consider the channel switching cost for opportunistic spectrum access systems in the I.I.D MAB model by L. Chen et al. [184]. They explicitly included the channel switching cost in the problem formulation. Also, they proposed a block-based UCB1 algorithm to avoid frequent channel switching.

2) Restless MAB: As has been validated by some experiment results [50], the activities of PUs can be approximately modeled as Markovian process; also, the channel quality exhibits Markovian properties [185]. If the transition properties are known, Markovian decision process (MDP) models can be applied for the OSA problems. However, such statistics are always not known a priori in practice, which makes restless MAB models desirable solutions.

S. Filippi et al. [186] formulated a restless MAB model for Markovian OSA systems with unknown statistical information. They considered a share-use OSA system with single SU and proposed a tiling algorithm which achieves expected regret in logarithmic order. C. Tekin et al. [187] also formulated a single-user restless MAB model for free-use OSA systems. Based on the well-known UCB1 algorithm, they proposed a sample mean index based policy, the regenerative cycle algorithm, to achieve expected regret in logarithmic order. In addition, K. Wang et al. [188] considered a similar restless MAB problem and proposed a myopic policy which is optimal. There is also a restless MAB formulation [189] which is naturally extended from its I.I.D counterpart [178].

However, a common limitation of existing restless MAB models is that only single user is considered and the proposed learning policies can not be applied to multiuser OSA systems.

C. Strengths and Limitations of MAB

Based on the review and comparison results of existing MAB models in opportunistic spectrum access systems, the strengths of MAB models are summarized as follows:

1) It provides a framework for online learning approaches in opportunistic spectrum access systems with unknown statistical information. It addresses the fundamental tradeoff between exploitation and exploration. Since the statistical information about the spectrum environment in opportunistic spectrum access systems, e.g., the channel availability, is always not known a priori, MAB models are very promising and useful.
2) It also provides flexible and efficient design. Generally, the arms in MAB models correspond to resources (e.g.,

channels), as has been considered in most applications. In a broader perspective, however, any action of the player can also be regarded as an arm, e.g., the user-channel matching profile in [178]. In this sense, one can also formulate other actions in opportunistic spectrum access systems as arms, e.g., to sense or not, to switch or not.

However, it also admits some inherent limitations as summarized below:

1) The MAB models essentially care more about the interaction with the environment rather than those with other players. In fact, it is noted that all existing studies involving multiple SUs, e.g., [43], [178]–[181], have a common assumption that the number of users is less than that of channels. Under this assumption, the users are finally spread over distinct channels. In these scenarios, the interactions with the environment are dominant while those among users are relatively unimportant. However, in scenarios where the number of users is larger than that of channels, which is very common in opportunistic spectrum access systems, the interactions among multiple users turn dominant and traditional MAB learning policies can not be applied.
2) Most existing policies in MAB models are deterministic, i.e., the players asymptotically choose certain arms. However, such deterministic strategies are only efficient for a single player but not in multiuser setting.
3) All MAB policies employ the parallel sensing models, which means that they also fail to find more speculum opportunities in opportunistic spectrum access systems, as game theory and Markovian decision process.

VII. Contrastive Analysis and Future Research

In this section, we discuss some concerns of decision-theoretic solutions for practical implementation. Based on this, we re-consider the above-discussed decision-theoretic solutions from a global and contrastive perspective. Moreover, we outline some future research directions.

A. Contrastive analysis of concerns for practical implementation

In this subsection, we list some important concerns of decision-theoretic solutions in practical opportunistic spectrum access systems. Specifically, three concerns including *information*, *cost* and *convergence speed* are discussed below.

1) ***Information****:* As the basic element, information plays a vital role in decision-theoretic solutions. Information in opportunistic spectrum access systems mainly includes spectrum occupancy status, channel quality and traffic demand of users. Generally, these information may be local or global, static or dynamic, deterministic or uncertain, and known or unknown.

In fact, the presented four decision-theoretic solutions in this survey address different scenarios with different information considerations. Specifically, the theory of multi-armed bandit problem considers scenarios where the statistical information are unknown a priori. The theory of optimal stopping problem considers scenarios where the information of realizations of unexplored channels are uncertain. The theory of Markovian decision process consideration scenarios where the system state (information) is dynamic and correlated. Furthermore, game theory considers interactions among multiple users and studies how the behavior of a user is affected by other users; in particular, coupled game learning algorithms that need information about other users or uncoupled algorithms that only rely on local information are designed to achieve some stable solutions.

2) ***Cost****:* Generally, there are two kinds of costs for decision-theoretic solutions with respect to practical implementation. The first kind is due to information exchange in the network, which consumes resources and causes extra overhead, e.g., time, power or bandwidth. The second kind is due to action switching, which involves hardware reconfiguration and signalling transmission to re-synchronize the transmitter and the receiver to a new chosen action.

To reduce the first kind cost, uncoupled algorithms that do not need information exchange and are only relying on local information have begun to draw great attention. Specifically, the algorithms for Markovian decision process, multi-armed bandit problem and optimal stopping problem, are originally developed for single user systems and of course are uncoupled. However, it is urgent and important to develop uncoupled algorithms for these models, when they are applied to multiuser systems. Furthermore, traditional learning algorithms in game models are coupled while uncoupled learning algorithms in game models are current active research topics. Table II provides detailed discussion on information requirement of different learning algorithms in game models.

To reduce the second kind cost, an efficient approach is to include the action cost into the optimization objectives. For the optimal stopping problems, action switching cost is obvious and explicitly included in the problem formulation. In addition, it is seen that the learning algorithms in other solutions, i.e., game models, Markovian decision process and multi-armed bandit problem, converge via learning from the trial-payoff history of SUs. As a result, the SUs frequently change their selections before convergence, which also leads to a large amount of switching cost. In current research, however, only the theory of optimal stopping problems considers this cost while the other three decision-theoretic solutions do not[3].

3) ***Convergence speed****:* Fast convergence speed of learning algorithms is desirable for two reasons: it can better adapt to dynamic environment and result in less overhead and cost.

However, this concern is not well studied in the presented decision-theoretic solutions. Specifically, game models and Markovian decision process mainly investigate the problem of whether the learning algorithms converge or not, and do not consider the convergence speed. In fact, most learning algorithms in the two models admit a feature of asymptotical convergence, i.e., they converge as the iteration number goes sufficiently large. Also, the convergence of learning approaches in the multi-armed bandit problem is generally studied as the iteration number goes sufficiently large, i.e., asymptotical convergence is studied. Finally, it is worthy to mention that the optimal stopping problems only involve one-

[3]In MAB models, the channel switching cost was only preliminarily considered in [184].

TABLE VI
COMPARISON RESULTS OF THE FOUR DECISION-THEORETIC SOLUTIONS

		Game Models	Markovian Decision Process	Optimal Stopping Problem	Multi-armed Bandit Problem
Focus		Interactions among multiple SUs	Dynamic spectrum opportunity	Tradeoff between sequential sensing cost and expected reward	Tradeoff between exploitation and exploration
Concerns	**Information**	A user's action is influenced by the actions of other users.	System state information is dynamic and correlated.	The realizations of unexplored channels are uncertain.	The statistical information of channels is unknown a priori.
	Cost	i. Type-I cost: cost for information exchange among users. ii. Type-II cost: cost for switching action before convergence. iii. Two types of cost are not considered in existing game models.	Type-II cost exists but has not been considered yet.	Type-II cost exists and is explicitly included in the problem formulation.	Type-II cost exists but is not well studied.
	Convergence Speed	i. Convergence speed for general algorithms is not studied. ii. Asymptotical convergence for a few algorithms is studied.	Asymptotical convergence is studied.	—	Asymptotical convergence is studied.

shot decision, which implies that the concern of convergence speed does not exist.

Based on the above global and contrastive analysis, we present the comparison results for the four presented decision-theoretic solutions in Table VI. It is seen that each kind of decision-theoretic solutions mainly addresses one challenge facing opportunistic spectrum access systems and admits its inherent strengths and limitations. In comparison, they are summarized below:

- Game theory captures the interactions among multiple SUs well, while the theories of Markovian decision process, optimal stopping problem, and multi-armed bandit problem require the environment to be stationary. In this sense, the latter three solutions are more suitable for single SU systems rather than multiple SU systems.
- The theory of optimal stopping problem provides more spectrum opportunities as the sequential sensing strategy is applied, while all other three solutions fail to find more speculum opportunities as they employ the parallel sensing strategies.
- The theory of multi-armed bandit problem provides an efficient online learning framework for unknown statistical information, while all other three solutions generally need to know the statistical information a priori.
- Cost in the context of decision-theoretic solutions is very important. However, only the optimal stopping problem involves this issue whereas all other three solutions have not yet considered it.

B. Future research for technical contents

Based on the above comparison results, we list some future research problems. The presented items mainly consist of two parts: the first are future research problems for technical contents and the second are those for methodologies. First, we list some research issues for technical content in opportunistic spectrum access systems, which would be investigated by any kind of reviewed decision theories.

1) ***Joint optimization of channel sensing and selection****:* It is seen that most decision-theoretic solutions are for the problem of channel selection based on the assumption of perfect sensing. In addition, although there are some existing work that considered the impact of imperfect sensing on the decision results, channel sensing and selection are still studied separately. Thus, it would be worthy to use decision theories to jointly optimize channel sensing and selection.

2) ***Incorporating user demand in metric formulation****:* It is noted that most decision-theoretic solutions for opportunistic spectrum access systems only considered optimization of allocated resources of SUs. For example, the optimization objectives in most existing studies are throughput maximization with or without energy constraint. These optimization metrics are indeed important but not enough for practical applications, for which user demand must be considered. The quality of experience (QoE) [190], which is jointly determined by the user demand and the allocated resources, should serve as new optimization metric in future research.

3) ***Including channel switching cost into the decision-theoretic solutions****:* In the decision-theoretic learning algorithms, the SUs frequently switch their selections before convergence. Furthermore, channel switching always consumes resources since it needs additional signaling to re-synchronize the transmitter and the receiver in the switched channels, as analyzed before. However, as shown in Table VI, except for the optimal stopping problem, the cost of learning algorithms are not studied in the other three decision-theoretic solutions. Thus, a desirable approach should be that the algorithms still converge with decreased switching frequency. To achieve this, the switching action cost should be explicitly considered in the three decision theories, which is interesting and challenging. This issue has begun to draw attention in wireless communications, e.g., cost of learning in heterogeneous 4G networks was reported in a recent work [80].

C. Future research for methodologies

The previously presented four decision theories, i.e., game theory, Markovian decision process, optimal stopping problem and multi-armed bandit problem, mainly address one challenge facing opportunistic spectrum access systems. Thus, in order to address two or more challenges simultaneously, it is natural to incorporate two or more decision theories. Following this idea, the methodology roadmap for opportunistic spectrum access systems is shown in Fig. 3. In the following, we discuss some possible future research issues.

1) ***Game models with uncoupled algorithms****:* As discussed in Section III. C, traditional learning algorithms in game models are coupled since a user's action update needs information about other users. Moreover, information exchange

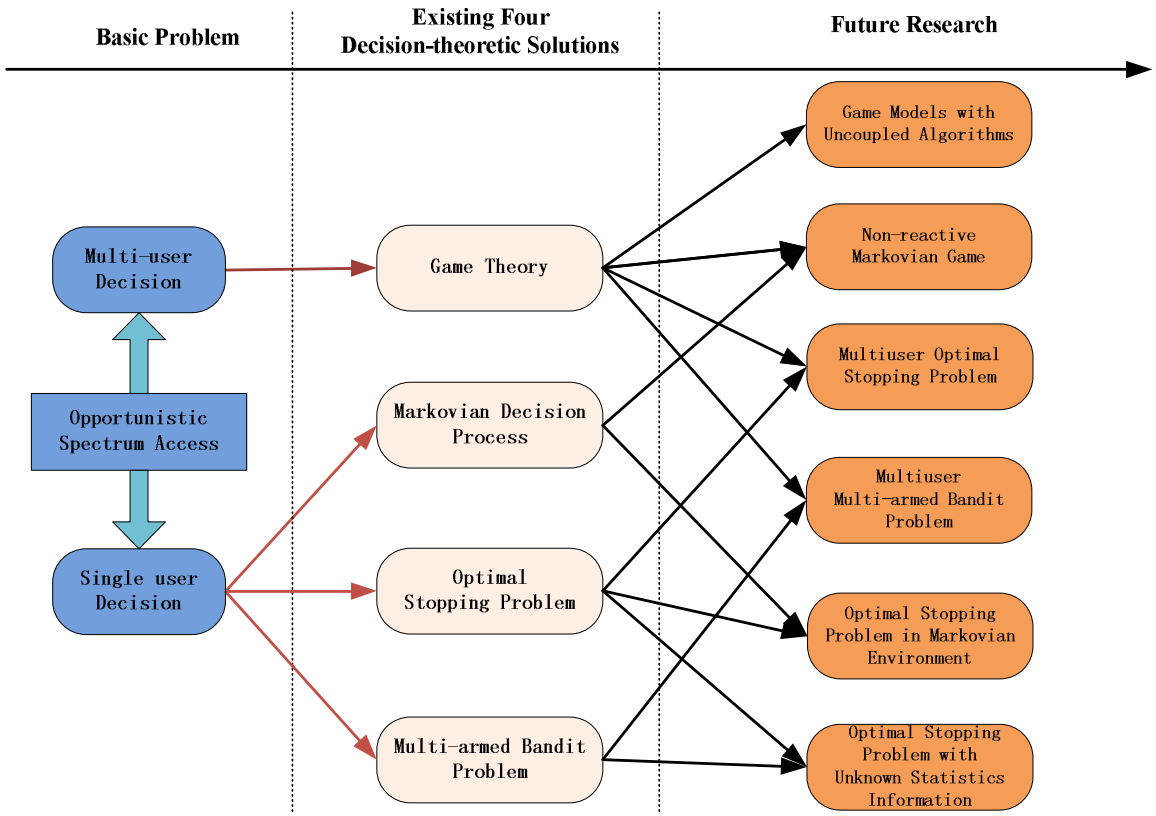

Fig. 3. The methodology roadmap for opportunistic spectrum access systems.

leads to Type-I cost, as shown in Table VI. Thus, it is desirable to develop game models with uncoupled learning algorithms for opportunistic spectrum access systems. Since the action of a user is influenced by the actions of other users, the convergence and optimality of uncoupled algorithms for game models are generally hard to guarantee and need to be carefully designed.

*2) **Non-reactive markovian game**:* This naturally emerges as an incorporation of game theory and Markovian decision process. In such a game, there are multiple users competing for resources in Markovian environment. In a classical Markovian game [51], the system state evolution is jointly determined by the current system state and the actions chosen by the players. We call this kind of game *reactive* Markovian game, as the system state evolution is affected by current actions. Different from the classical reactive Markovian games, the games used in opportunistic spectrum access systems with Markovian environment exhibit an interesting property of the system state evolution being *non-reactive*. To be more specific, the system state in opportunistic spectrum access systems, which is generally defined as the spectrum occupancy state, evolves according to the traffic model of PUs and regardless of the actions of the SUs. Thus, although there are some solutions for the classical reactive Markovian games, e.g., value iteration and stochastic approximation algorithms proposed in [51], the solutions for non-reactive Markovian games would be more concise and elegant. In particular, distributed solutions that do not need information about other users are desirable.

*3) **Multiuser optimal stopping problem**:* As discussed before, the theory of optimal stopping problem is more suitable for single user OSA systems rather than multiuser systems. The fact that there are always multiple SUs in an opportunistic spectrum access system naturally yields the multiuser optimal stopping problem. For solving the multiuser optimal stopping problem, the difficulties primarily lie on the interactions among multiple users. One can incorporate game theory into traditional optimal stopping problem to develop desirable solutions. It should be pointed out that although few existing work, e.g., [41], [165], have considered multiple SUs with numeric simulation, theoretical analysis for multiuser optimal stopping problems has not yet been reported.

*4) **Multiuser multi-armed bandit problem**:* Also, traditional multi-armed bandit problem is more suitable for single user OSA systems rather than multiuser systems, which naturally yields the multiuser multi-armed bandit problem. Although some existing studies have begun to consider the effect of multiple users, e.g., [180], [181], the presented solutions are only for some specific applications and can not be extended to general scenarios. Specifically, some methods were proposed to create orthogonality among users (time orthogonality [180] and channel orthogonality in [181]), which eventually leads to stationary environment for the learning policies proposed therein. In a broad perspective, however, the environment for multiuser multi-armed bandit problems is non-stationary, which means that traditional learning policies can not be applied directly. To achieve a general solution of multiuser multi-armed bandit problem, one can incorporate game theory into traditional multi-armed bandit problem. Some preliminary studies can be found in [182] and further investigations are needed.

*5) **Optimal stopping problem in markovian environment**:* In traditional optimal stopping problem, the observed random variables are distributed independently, from decision epoch to decision epoch, which means that the decisions in each epoch are independent from those in previous epoches. For opportunistic spectrum access systems, however, the dynamics of the spectrum opportunities between successive epoches are always formulated as Markovian process. Considering such dynamics will bring about fundamental challenges; for example, the sensing order in each epoch will be adaptively optimized based on the observations in the previous epoches.

*6) **Optimal stopping problem with unknown statistical information**:* As pointed out before, traditional optimal stopping problem needs to know the statistical information of the channels to calculate the expected reward of proceeding to sense residual channels. However, the statistical information of the channel availability and quality are always unknown a priori, which means that there is a fundamental tradeoff between exploration and exploitation. Incorporating the approaches of multi-armed bandit problems into traditional optimal stopping problems would yield desirable solutions for this problem.

Remark 5: Surely, one can develop more complicated solutions to address more than two challenges of opportunistic spectrum access systems. This might be achieved by incorporating more than two of the four decision theories, which would be definitely extremely challenging. Although we can not discuss this issue in detail at present, we believe that it will eventually be achieved in the near future.

VIII. Summary

In this article, we surveyed decision-theoretic solutions for channel sensing and access strategies for opportunistic spectrum access (OSA) systems. The main contributions of this work are threefold. First, we globally analyzed the challenges facing OSA systems, which mainly include interactions among multiple users, dynamic spectrum opportunity, tradeoff between sequential sensing cost and expected reward, and tradeoff between exploitation and exploration in the absence of prior statistical information. Second, we provided comprehensive review and comparison of each kind of existing decision-theoretic solutions, i.e., game models, Markovian decision process, optimal stopping problem and multi-armed bandit problem. Third, we analyzed their strengths and limitations and outlined further research issues for both technical content and methodology. In particular, contrastive analysis for the four kinds of solutions are provided in terms of information, cost and convergence speed, which are key concerns for practical implementation. Moreover, each kind of existing decision-theoretic solution mainly addresses a single aspect of the challenges facing OSA systems, which implies that two or more kinds of decision-theoretic solutions should be incorporated to address more challenges simultaneously.

References

[1] H. Zhang, H. Yin, H. Jia, et al., "Study of effects of obstacle on non-line-of-sight ultraviolet communication links," *Opt. Express,* vol. 19, pp. 21216-21226, 2011.

[2] T. Komine, J. Lee, S. Haruyama, et al., "Adaptive equalization system for visible light wireless communication utilizing multiple white LED lighting equipment," *IEEE Trans. Wireless Commun.*, vol. 8, no. 6, pp. pp. 2892-2900, 2009.

[3] R. Piesiewicz, C. Jansen, D. Mittleman, et al, "Scattering analysis for the modeling of THz communication systems," *IEEE Transactions on Antennas and Propagation*, vol. 55, no. 11, pp. 3002–3009, 2007.

[4] J. Mitola, and G. Q. Maguire, "Cognitive radio: making software radios more personal," *IEEE Pers. Commun.*, vol. 6, no. 4, pp. 13 –18, Aug. 1999.

[5] E. Blossom, "GNU radio: tools for exploring the radio frequency spectrum," *Linux Journal*, vol. 2004, no. 122, June 2004.

[6] M. Ettus, "Universal software radio peripheral," [Online]. Available: http://www.ettus.com

[7] M. McHenry, E. Livsics, T. Nguyen, and N. Majumdar, "XG dynamic spectrum sharing field test results," *in Proc. IEEE Int. Symposium on New Frontiers in Dynamic Spectrum Access Networks,* Dublin, Ireland, Apr. 2007, pp. 676-684.

[8] http://warp.rice.edu/.

[9] P. Murphy, A. Sabharwal, and B. Aazhang, "Design of WARP: Wireless open-access research platform," In *European Signal Processing Conference*, June 2006.

[10] A. Khattab, J. Camp, C. Hunter, et al., "WARP: A flexible platform for clean-slate wireless medium access protocol design," *Mobile Computing and Communications Review,* vo. 12, no. 1, pp. 56-58, Jan., 2008.

[11] "Open-Source SCA Implementation::Embedded (OSSIE)"; http://ossie.wireless.vt.edu

[12] C. R. Aguayo González, C. B. Dietrich, and J. H. Reed, "Understanding the software communications architecture," *IEEE Commun. Mag.,* vol. 47, no. 9, pp. 50-57, Sept., 2009.

[13] C. R. Aguayo González, C. B. Dietrich, S. Sayed, et al., "Open-source SCA-based core framework and rapid development tools enable software-defined radio education and research," *IEEE Commun. Mag.,* vol. 47, no. 10, pp. 48-55, Oct., 2009.

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

[15] Q. Zhao, L. Tong, A. Swami; Y. Chen, "Decentralized cognitive MAC for opportunistic spectrum access in ad hoc networks: A POMDP framework," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 3, pp.589-600, 2007

[16] T. Yücek and H. Arslan, "A survey of spectrum sensing algorithms for cognitive radio applications," *IEEE Commun. Surveys & Tutorials,* vol. 11, no. 1, first quarter 2009.

[17] IF. Akyildiz, W. Lee, M. Vuran, et al., "NeXt generation/dynamic spectrum access/cognitive radio wireless networks: A survey," *Computer Networks*, vol. 50, vol. 13, pp. 2127-2159, 2006.

[18] IF. Akyildiz, W. Lee, M. Vuran, et al., "A survey on spectrum management in cognitive radio networks," *IEEE Commun. Mag.,* vol. 46, no. 4, pp. 40-48, 2008.

[19] K. Shin, H. Kim, A. Min, A. Kumar, "Cognitive radios for dynamic spectrum access: From concept to reality," *IEEE Wireless Commun.,* vol. 17, no. 6, pp. 64-74, 2010.

[20] Y. Zhao, S. Mao, J. Neel, and J. Reed, "Performance evaluation of cognitive radios: Metrics, utility functions, and methodology," *Proc. IEEE*, vol. 97, no. 4, pp. 642 –659, 2009.

[21] A. De Domenico, E. Strinati and M. Di Benedetto, "A survey on MAC strategies for cognitive radio networks," *IEEE Commun. Surveys & Tutorials,* vol. 14, no. 1, pp. 21-44, 2012.

[22] C. Cormio, K. Chowdhury, "A survey on MAC protocols for cognitive radio networks," *Ad Hoc Networks,* vol. 7, no. 7, pp. 1315-1329, 2009.

[23] T. Krishna and A. Das, "A survey on MAC protocols in OSA networks," *Computer Networks,* vol. 53, vol. 9, pp. 1377-1394, 2009.

[24] P. Ren, Y. Wang, Q. Du, et al., "A survey on dynamic spectrum access protocols for distributed cognitive wireless networks," *EURASIP J. Wireless Communications and Networking,* pp. 1-21, 2012.

[25] J. Xiang, Y. Zhang and T. Skeie, "Medium access control protocols in cognitive radio networks," *Wireless Communications & Mobile Computing,* vol. 10, no. 1, pp. 31-49, 2010.

[26] Q. Zhao, A. Swami, "A Decision-theoretic framework for opportunistic spectrum access," *IEEE Wireless Commun.,* vol. 14, no. 4, pp.14–20, 2007

[27] A. MacKenzie, J. Reed, P. Athanas, et al., "Cognitive radio and networking research at Virginia Tech," *Proc. IEEE*, vol. 97, no. 4, pp. 660 –688, 2009.

[28] M. van der Schaar and F. Fu, "Spectrum access games and strategic learning in cognitive radio networks for delay-critical applications," *Proc. IEEE*, vol. 97, no. 4, pp. 720 –740, 2009.

[29] B. Wang, Y. Wu and K. Liu, "Game theory for cognitive radio networks: An overview," *Computer Networks,* vol. 54, no. 14, pp. 2537-2561, 2010.

[30] M. T. Masonta, M. Mzyece, and N. Ntlatlapa, "Spectrum decision in cognitive radio networks: A survey," *IEEE Commun. Surveys & Tutorials,* DOI: 10.1109./SURV.2012.111412.00160.

[31] M. Bkassiny, Y. Li, and S. K. Jayaweera, "A survey on machine-learning techniques in cognitive radios," *IEEE Commun. Surveys & Tutorials,* DOI: 10.1109./SURV.2012.100412.00017.

[32] A. Azarfar, J. F. Frigon, and B. Sanso, "Improving the reliability of wireless networks using cognitive radios," *IEEE Commun. Surveys & Tutorials,* vol. 14, no. 2, pp. 338-354, 2012.

[33] A. He, K. K. Bae, T. R. Newman, et al., "A survey of artificial intelligence for cognitive radios," *IEEE Trans. Veh. Technol.,* vol. 59, no. 4, pp. 1578-1592, May 2010.

[34] A. Attar, M. Nakhai, and A. Aghvami, "Cognitive radio game for secondary spectrum access problem," *IEEE Trans. Wireless Commun.*, vol. 8, no. 4, pp. 2121 –2131, 2009.

[35] W-Y. Lee and I. F. Akyildiz, "Optimal spectrum sensing framework for cognitive radio networks," *IEEE Trans. Wireless Commun.*, vol. 7, no. 10, pp. 3845-3857, October 2008.

[36] S. Chen and L. Tong, "Maximum throughput region of multiuser cognitive access of continuous time markovian channels," *IEEE J. Sel. Areas Commun.*, vol. 29, no.12, pp. 1959-1969, Dec 2011.

[37] H. Su and X. Zhang, "Interference-confined adaptive transmission scheme for cognitive radio networks," in *Pro. IEEE ICC*, pp. 1-5, 2010.

[38] M. Rashid, M. Hossain, E. Hossain, et al., "Opportunistic spectrum scheduling for multiuser cognitive radio: A queueing analysis," *IEEE Trans. Wireless Commun.*, vol. 8, no. 10, pp. 5259-5269, 2010.

[39] R. Fan and H. Jiang, "Channel sensing-order setting in cognitive radio networks: A two-user case," *IEEE Trans. Veh. Technol.*, vol. 58, no. 9, pp. 4997-5008, 2009.

[40] J. Jia, Q. Zhang, and X. Shen, "HC-MAC: A hardware-constrained cognitive MAC for efficient spectrum management," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 1, pp. 106-117, 2008.

[41] H. T. Cheng and W. Zhuang, "Simple channel sensing order in cognitive radio networks," *IEEE J. Sel. Areas Commun.*, vol. 29, no. 4, pp. 676-688, 2011.

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

[43] A. Anandkumar, N. Michael, and A. Tang, "Opportunistic spectrum access with multiple users: Learning under competition," in *Proc. IEEE INFOCOM,* San Deigo, USA, March 2010.

[44] Y. Xu, J. Wang, Q. Wu, et al, "Optimal energy-efficient channel exploration for opportunistic spectrum usage," *IEEE Wireless Commun. Lett.,* vol. 1, no. 2, pp. 77-80, 2012.

[45] A. Sabharwal, A. Khoshnevis, and E. Knightly, "Opportunistic spectral usage: bounds and a multi-band CSMA/CA protocol," *IEEE/ACM Trans. Netw.,* vol. 15, no. 3, pp. 533-545, 2007.

[46] E. Axell and E. G. Larsson, "Optimal and sub-optimal spectrum sensing of ofdm signals in known and unknown noise variance," *IEEE J. Sel. Areas Commun.*, vol. 29, no. 2, pp. 290-304, 2011.

[47] M. Ge and S. Wang, "Fast optimal resource allocation is possible for multiuser OFDM-based cognitive radio networks with heterogeneous services," *IEEE Trans. Wireless Commun.*, vol. 11, no. 4, pp. 1500-1509, 2012.

[48] R. Myerson, *Game Theory: Analysis of Conflict*. Cambridge, MA: Harvard Univ. Press, 1991.

[49] R. Branzei, D. Dimitrov, and S. Tijs, "Models in cooperative game theory," Second Edition, Springer.

[50] S. Geirhofer, L. Tong, and B. M. Sadler, "Dynamic spectrum access in WLAN channels: Empirical model and its stochastic analysis," in *Proc. First International Workshop on Technology and Policy in Accessing Spectrum (TAPAS)*, Boston, MA, August, 2006.

[51] J. Huang and V. Krishnamurthy, "Transmission control in cognitive radio as a Markovian dynamic game: Structural result on randomized threshold policies," *IEEE Trans. Commun.,* vol. 58, no. 1, pp. 301-310, 2010.

[52] Y. Xu, J. Wang, and Q. Wu, et al, "Opportunistic spectrum access in unknown dynamic environment: A game-theoretic stochastic learning solution", *IEEE Trans. Wireless Commun.*, vol. 11, no. 4, pp.1380-1391, 2012.

[53] Y. Xu, Q. Wu and J. Wang,"Game theoretic channel selection for opportunistic spectrum access with unknown prior information," in *Proc. IEEE ICC,* 2011.

[54] Y. Xu, A. Anpalagan, Q. Wu, et al., "Game-theoretic channel selection for interference mitigation in cognitive networks with block-fading channels," in *Proc. IEEE WCNC*, 2013.

[55] M. Maskery, V. Krishnamurthy, and Q. Zhao, "Decentralized dynamic spectrum access for cognitive radios: Cooperative design of a non-cooperative game," *IEEE Trans. Commun.*, vol. 57, no. 2, pp. 459 -469, 2009.

[56] Y. Xu, J. Wang, Q. Wu, A. Anpalagan, and Y.-D. Yao, "Opportunistic spectrum access in cognitive radio networks: Global optimization using local interaction games," *IEEE J. Sel. Topics Signal Process.*, vol. 6, no. 2, pp. 180-194, 2012.

[57] T. S. Ferguson, *Optimal stopping and applications*. [Online]. Available: http://www.math.ucla.edu/ tom/Stopping/Contents.html.

[58] P. Auer, N. Cesa-Bianchi, and P. Fischer, "Finite-time analysis of the multiarmed bandit problem," *Machine Learning,* vol. 47, pp. 235-256, 2002.

[59] C. Peng, H. Zheng, and B. Zhao, "Utilization and fairness in spectrum assignment for opportunistic spectrum access," *Mobile Networks & Applications*, vol. 11, no. 4, pp. 555-576, 2006.

[60] L. Cao and H. Zheng, "Distributed rule-regulated spectrum sharing," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 1, pp. 130–145, 2008.

[61] Z. Zhao, Z. Peng, S. Zheng, et al, "Cognitive radio spectrum allocation using evolutionary algorithms," *IEEE Trans. Wireless Commun.*, vol. 8, no. 9, pp. 4421-4425, 2009.

[62] K. M. Thilina, K. W. Choi, N. Saquib, and E. Hossain, "Pattern Classification Techniques for Cooperative Spectrum Sensing in Cognitive Radio Networks: SVM and W-KNN Approaches, in *Proc. Globecom 2012*.

[63] Y-C. Liang, Y. Zeng, E. Peh, and A. Hoang, "Sensing-throughput tradeoff for cognitive radio networks," *IEEE Trans. Wireless Commun.,* vol. 7, no. 4, pp. 1326-1336, 2008.

[64] J. Zhang, L. Qi, and H. Zhu, "Optimization of MAC frame structure for opportunistic spectrum access," *IEEE Trans. Wireless Commun.,* vol. 11, no. 6, pp. 2036-2045, 2012.

[65] A. Hoang, Y-C. Liang, and Y. Zeng, "Adaptive joint scheduling of spectrum sensing and data transmission in cognitive radio networks," *IEEE Trans. Commun.,* vol. 58, no. 1, pp. 235-246, 2010.

[66] Y. Xu, J. Wang, Q. Wu, "Interference-throughput tradeoff in dynamic spectrum access: Analysis based on discrete-time queuing subjected to bursty preemption," in *Proc. 4th International Conference on Cognitive Radio Oriented Wireless Networks (CROWNCOM)*, 2009.

[67] S. Li and Z. Han, 'Socially optimal queuing control in cognitive radio networks subject to service interruptions: To queue or not to queue?" *IEEE Trans. Wireless Commun.,* vol. 10, no. 5, pp. 1656-1666 , 2011.

[68] J. Wang, Y. Xu, Z. Gao, and Q. Wu, "Discrete-time queuing analysis of opportunistic spectrum access: Single user case" *Frequenz*, vol. 65, no. 11-12, pp. 335-341.

[69] I. Krikidis, N. Devroye, and J. Thompson, "Stability analysis for cognitive radio with multi-access primary transmission," *IEEE Trans. Wireless Commun.,* vol. 9, no. 1, pp. 72-77, 2010.

[70] C. Chou, N. S. Shankar, H. Kim, et al., "What and how much to gain by spectrum agility?" *IEEE J. Sel. Areas Commun.*, vol. 25, no. 3, pp. 576 –588, 2007.

[71] S. Srinivasa and S. Jafar, "How much spectrum sharing is optimal in cognitive radio networks?" *IEEE Trans. Wireless Commun.*, vol. 7, no. 10, pp. 4010 –4018, 2008.

[72] Y. Xu, J. Wang and Q. Wu, "Effective capacity region of two-user opportunistic spectrum access," *Sci China Inf Sci,* vol. 54, no. 9, pp. 1928-1937, 2011.

[73] R. Axelrod and W. Hamilton, "The evolution of cooperation," *Science,* vol. 211, no. 4489, pp. 1390-1396, 1981.

[74] E. Fehr and U. Fischbacher, "The nature of human altruism," *Nature,* vol. 425, no. 6960, pp. 785-791, 2003.

[75] C. Yeung, A. Poon and F. Wu, "Game theoretical multi-agent modelling of coalition formation for multilateral trades," *IEEE Transactions on Power Systems,* vol. 14, no. 3, pp. 929-934, 1999.

[76] K. Akkarajitsakul, E. Hossain, D. Niyato, et al., "Game theoretic approaches for multiple access in wireless networks: A survey," *IEEE Commun. Surveys & Tutorials,* vol. 13, no. 3, pp. 372-395, 2011.

[77] R. Trestian, O. Ormond and G. Muntean, "Game theory-based network selection: Solutions and challenges," *IEEE Commun. Surveys & Tutorials,* vol. 14, no. 4, pp. 1212–1231, 2012.

[78] V. Srivastava, J. Neel, A. B. Mackenzie, et al., "Using game theory to analyze wireless ad hoc networks," *IEEE Commun. Surveys & Tutorials,* vol. 7, no. 4, pp. 46-56, 2005.

[79] H. Yaiche, R. Mazumdar, and C. Rosenberg, "A game theoretic framework for bandwidth allocation and pricing in broadband networks," *IEEE J. Sel. Areas Commun.*, vol. 8, no. 5, pp. 667-678, 2000.

[80] M. Khan, H. Tembine and A. Vasilakos, "Game dynamics and cost of learning in heterogeneous 4G networks," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 1, pp. 198-213, 2012.

[81] S. Lien, Y. Lin, K.-C. Chen, "Cognitive and game-theoretical radio resource management for autonomous femtocells with QoS guarantees," *IEEE Trans. Wireless Commun.*, vol. 10, no. 7, pp. 2196-2206, 2011.

[82] R. J. Aumann, "Subjectivity and correlation in randomized strategy", *Journal of Mathematical Economics,* vol. 1, no. 1, pp. 67-96, 1974.

[83] H. Kameda and E. Altman, "Inefficient noncooperation in networking games of common-pool resources," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 7, pp. 1260 –1268, 2008.

[84] L. Chen and J. Leneutre, "A game theoretic framework of distributed power and rate control in IEEE 802.11 WLANs," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 7, pp. 1128 –1137, 2008.

[85] Y. Song, C. Zhang, and Y. Fang, "Joint channel and power allocation in wireless mesh networks: A game theoretical perspective," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 7, pp. 1149 –1159, 2008.

[86] T. Cui, L. Chen, and S. H. Low, "A game-theoretic framework for medium access control," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 7, pp. 1116-1127, 2008.

[87] D. Monderer. and L. S. Shapley, "Potential games," *Games and Economic Behavior*, vol. 14, pp. 124–143, 1996.

[88] S. Hart and A. Mas-Colell, "A simple adaptive procedure leading to correlated equilibrium," *Econometrica,* vol. 68, no. 5, pp. 1127-1150, 2000.

[89] N. Nie and C. Comaniciu, "Adaptive channel allocation spectrum etiquette for cognitive radio networks," *Mobile Networks & Applications*, vol. 11, no. 6, pp. 779–797, 2006.

[90] L. Law, J. Huang, M. Liu, and S. R. Li, "Price of anarchy for cognitive MAC games," in *in Proc. IEEE GLOBECOM 2009*, pp. 1-6.

[91] Q. Wu, Y. Xu, L. Shen, and J. Wang, "Investigation on GADIA algorithms for interference avoidance: A game-theoretic perspective," *IEEE Commun. Lett.,* vol. 16, no. 7, pp. 1041-1043, 2012.

[92] B. Babadi, and V. Tarokh, "GADIA: A greedy asynchronous distributed interference avoidance algorithm," *IEEE Trans. Signal Process.*, vol. 56, no. 12, pp.6228–6252, 2010.

[93] H. P. Young, Individual Strategy and Social Structure. Princeton University Press, 1998.

[94] J. Marden, G. Arslan, and J. Shamma, "Cooperative control and potential games," *IEEE Trans. Syst., Man, Cybern. B*, vol. 39, no. 6, pp. 1393 –1407, 2009.

[95] Z. Han, C. Pandana, and K. Liu, "Distributive opportunistic spectrum access for cognitive radio using correlated equilibrium and no-regret learning," in *in Proc. IEEE WCNC*, pp. 11 –15, 2007.

[96] 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 2009.*, pp. 1893–1898, 2009

[97] H. Li, "Multi-agent q-learning for competitive spectrum access in cognitive radio systems." in *IEEE Fifth Workshop on Networking Technologies for Software Defined Radio Networks*, 2010.

[98] A. Montanari and A. Saberi, "Convergence to equilibrium in local interaction games," in *50th Annual IEEE Symposium on Foundations of Computer Science*, pp. 303 –312, 2009.

[99] S. Ahmad, C. Tekin, M. Liu, R. Southwell, J. Huang, "Spectrum sharing as spatial congestion games," 2010 [Online]. Available: http://arxiv.org/abs/1011.5384.

[100] H. Li and Z. Han, "Competitive spectrum access in cognitive radio networks: Graphical game and learning," in *Proc. IEEE WCNC*, 2010, pp. 1-6.

[101] M. Azarafrooz and R. Chandramouli, "Distributed learning in secondary spectrum sharing graphical game," in *Pro. IEEE GLOBECOM,* pp.1-6, 2011.

[102] Y. Xu, Q. Wu, J. Wang, N. Ming, and A. Anpalagan, "Distributed channel selection in CRAHNs with heterogeneous spectrum opportunities: A local congestion game approach," *IEICE Trans. Commun.,* vol. E95-B, no. 3, pp. 991-994, 2012.

[103] M. Liu, S. Ahmad, Y. Wu, "Congestion games with resource reuse and applications in spectrum sharing," *GameNets* , pp. 171–179, 2009.

[104] C. Tekin, M. Liu, R. Southwell, J. Huang, and S. Ahmad, "Atomic congestion games on graphs and their applications in networking," *IEEE/ACM Trans. Netw.,* vol. 20, no. 5, pp. 1541-1552, 2012.

[105] B. Vcking and R. Aachen, "Congestion games: Optimization in competition," in *Proc. 2nd Algorithms Complexity Durham Workshop,* 2006, pp. 9-20, Kings College Publications.

[106] J. M. Smith, Evolution and the theory of games. Cambridge University Press, 1982.

[107] J. M. Smith and G. R. Price, "The logic of animal conflict," *Nature,* vol. 246, no. 5427, pp. 15-18, 1973.

[108] H. Tembine, E. Altman, R. El-Azouzi, and Y. Hayel, "Evolutionary games in wireless networks," *IEEE Trans. Syst., Man, Cybern. B, Cybern.,* vol. 40, no. 6, pp. 634-646, Jun. 2010.

[109] B. Wang, K. J. R. Liu, and T. C. Clancy, "Evolutionary cooperative spectrum sensing game: How to collaborate?" *IEEE Trans. Commun.,* vol. 58, no. 3, 2010.

[110] D. Niyato and E. Hossain, "Dynamics of network selection in heterogeneous wireless networks: An evolutionary game approach," *IEEE Trans. Veh. Technol.,* vol. 58, no. 4, pp. 2008-2017, 2009.

[111] X. Chen and J. Huang, "Evolutionarily stable spectrum access," *IEEE Trans. Mobile Computing,* DOI: 10.1109/TMC.2012.94.

[112] W. Saad, Z. Han, M. Debbah, A. Hjøungnes, and T. Başr, "Coalitional game theory for communication networks: A tutorial," *IEEE Signal Process. Mag.,* vol. 26, no. 5, pp. 77-97, 2009.

[113] W. Saad, Z. Han, T. Basar, M. Debbah, and A. Hjørungnes, "Coalition formation games for collaborative spectrum sensing," *IEEE Trans. Veh. Technol.*, vol. 60, no. 1, pp. 276-297, 2011.

[114] W. Saad, Z. Han, R. Zheng, A. Hjøungnes, T. Başr, and H. V. Poor, "Coalitional games in partition form for joint spectrum sensing and access in cognitive radio networks," *IEEE J. Sel. Topics Signal Process.*, vol. 6, no. 2, pp. 195-209, 2012.

[115] D. Niyato and E. Hossain, "Competitive spectrum sharing in cognitive radio networks: A dynamic game approach," *IEEE Trans. Wireless Commun.*, vol. 7, no. 7, pp. 2651–2660, 2008.

[116] D. Niyato and E. Hossain, "Competitive pricing for spectrum sharing in cognitive radio networks: Dynamic game, inefficiency of nash equilibrium, and collusion," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 1, pp. 192 –202, 2008.

[117] D. Niyato and E. Hossain, "Market-Equilibrium, Competitive, and Cooperative Pricing for Spectrum Sharing in Cognitive Radio Networks: Analysis and Comparison," *IEEE Trans. Wireless Commun.*, vol. 7, no. 11, pp. 4273–4283, 2008.

[118] D. Niyato and E. Hossain, "Spectrum trading in cognitive radio networks: A market-equilibrium-based approach," *IEEE Wireless Commun.*, vol. 15, no. 6, pp. 71-80, 2008.

[119] L. Gao, X. Wang, Y. Xu, et al., "Spectrum trading in cognitive radio networks: A contract-theoretic modeling approach," *IEEE J. Sel. Areas Commun.*, vol. 29, no. 4, pp. 843–855, 2011.

[120] L. Qian, F. Ye, L. Gao, et al., "Spectrum trading in cognitive radio networks: An agent-based model under demand uncertainty," *IEEE Trans. Commun.,* vol. 59, no. 11, pp. 3192-3203, 2011.

[121] D. Niyato, E. Hossain and Z. Han, "Dynamics of multiple-seller and multiple-buyer spectrum trading in cognitive radio networks: A game-theoretic modeling approach," *IEEE Trans. Mobile Computing,* vol. 8, no. 8, pp. 1009-1022, 2009.

[122] S. Lasaulce, Y. Hayel, R. E. Azouzi and M. Debbah, "Introducing hierarchy in energy games," *IEEE Trans. Wireless Commun.*, vol. 8, no. 7, pp. 3833-3843, 2009.

[123] M. Haddad, S. E. Elayoubi, E. Altman and Z. Altman, "A Hybrid Approach for Radio Resource Management in Heterogeneous Cognitive Networks," *IEEE J. Sel. Areas Commun.,* vol. 29, no. 4, pp. 831–842, 2012.

[124] X. Kang, R. Zhang and M. Motani, "Price-based resource allocation for spectrum-sharing femtocell networks: a stackelberg game approach," *IEEE J. Sel. Areas Commun.,* vol. 30, no. 3, pp. 538–549, 2012.

[125] J. Park and M. van der Schaar, "Stackelberg contention games in multiuser networks," *EURASIP J. Advances in Signal Process.,* pp. 1-15, 2009.

[126] J. Park and M. van der Schaar, "Designing incentive schemes based on intervention: the case of imperfect monitoring," in *Proc. GameNets 2011.*

[127] J. Park and M. van der Schaar, "The theory of intervention games for resource sharing in wireless communications," *IEEE J. Sel. Areas Commun.,* vol. 30, no. 1, pp. 165C175, Jan. 2012.

[128] Y. Xiao, J. Park, and M. van der Schaar, "Intervention in power control games with selfish users," *IEEE J. Sel. Topics Signal Process.,* vol. 6, no. 2, pp. 165C179, Apr. 2012.

[129] Y. Xiao, J. Park, and M. van der Schaar, "Repeated games with intervention: Theory and applications in communications," *IEEE Trans. Commun.,* vol. 60, no. 10, pp. 3123-3132, 2012.

[130] A. Daoud, T Alpcan, S. Agarwal, and M. Alanyali, "A Stackelberg game for pricing uplink power in wide-band cognitive radio networks," in *47th IEEE Conference on Decision and Control (CDC)*, Cancun, Mexico, 2008.

[131] Y. Wu, T. Zhang, and D. Tsang, "Joint pricing and power allocation for dynamic spectrum access networks with Stackelberg game model," *IEEE Trans. Wireless Commun.,* vol. 10, no. 1, pp. 12-19, 2011.

[132] M. Razaviyayn, M. Yao, and L. Zhi-Quan, "A Stackelberg game approach to distributed spectrum management," in *IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)*, Dallas, TX, 2010.

[133] L. Duan, J. Huang and B. Shou, "Investment and pricing with spectrum uncertainty: A cognitive operators perspective," *IEEE Trans. Mobile Computing,* vol. 10, no. 11, pp. 1590-1604, 2011.

[134] M. Bloem, T. Alpcan, and T. Başar,"A Stackelberg game for power control and channel allocation in cognitive radio networks," in *Proceedings of the 2nd international conference on performance evaluation methodologies and tools*, 2007.

[135] Y. Xiao, G. Bi, D. Niyato, L. A. DaSilva, "A hierarchical game theoretic framework for cognitive radio networks," *IEEE J. Sel. Areas Commun.,* vol. 30, no. 10, pp. 2053-2069, 2012.

[136] Martin L. Puterman, Markov Decision Processes: Discrete Stochastic Dynamic Programming, John Wiley & Sons, Inc., New York, NY, 1994.

[137] Q. Hu and W. Yue, Markov Decision Processes with Their Applications, Springer, 2008.

[138] L. Kaelbling, M. Littman and A. Moore, "Reinforcement learning: A survey," *Journal of Artificial Intelligence Research*, vol. 4, pp. 237-285, 1996.

[139] S. Ross, J. Pineau, S. Paquet, et al, "Online planning algorithms for POMDPs," *J. Artificial Intelligence Research*, vol. 32, pp. 663-704, 2008.

[140] A. Piunovskiy and X. Mao, "Constrained Markovian decision processes: the dynamic programming approach," *Operations Research Letters,* vol. 27, no. 3, pp. 119-126, 2000.

[141] S. Yin, D. Chen, Q. Zhang, et al, "Prediction-based throughput optimization for dynamic spectrum access," *IEEE Trans. Veh. Technol.*, vol. 60, no. 3, pp. 1284-1289, 2011.

[142] D. Chen, S. Yin, M. Liu, et al, "Mining spectrum usage data: A large-scale spectrum measurement study," in *Proc. ACM MOBICOM*, pp. 13-24, Sep. 2009.

[143] Q. Zhao, S. Geirhofer, L. Tong, et al, "Opportunistic spectrum access via periodic channel sensing," *IEEE Trans. Signal Process.*, vol. 56, pp.785-796, 2008.

[144] X. Li, Q. Zhao, X. Guan et la, "Optimal cognitive access of markovian channels under tight collision constraints," *IEEE J. Sel. Areas Commun.*, vol. 29, no. 4, pp. 746-756, Apr 2011.

[145] D. Niyato, E. Hossain, and P. Wang, "Optimal channel access management with QoS support for cognitive vehicular networks," *IEEE Trans. Mobile Computing,* vol. 10, no. 4, pp. 573-591, Feb., 2011.

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

[147] J. Unnikrishnan and V. V. Veeravalli, "Algorithms for dynamic spectrum access with learning for cognitive radio," *IEEE Trans. Signal Process.*, vol. 58, no. 2, pp. 750-760, Feb 2010.

[148] A. T. Hoang, Y.-C. Liang, D. Wong, et al, "Opportunistic spectrum access for energy-constrained cognitive radios," *IEEE Trans. Wireless Commun.*, vol. 8, no. 3, pp. 1206-1211, Mar 2009.

[149] Y. Chen, Q. Zhao, and A. Swami, "Distributed spectrum sensing and access in cognitive radio networks with energy constraint," *IEEE Trans. Signal Process.,* vol. 57, no. 2, pp.-783-797, 2009.

[150] F. Fu and M. van der Schaar, "Learning to compete for resources in wireless stochastic games," *IEEE Trans. Veh. Technol.*, vol. 58, no. 4, pp. 1904-1919, 2009.

[151] Y. Ohtsubo, "Risk minimization in optimal stopping problem and applications," *Journal Of The Operations Research Society Of Japan,* vol. 46, no. 3, pp. 342-352, 2003.

[152] D. Zheng, W. Ge, and J. Zhang, "Distributed opportunistic scheduling for ad-hoc communications: An optimal stopping approach," *IEEE. Trans. Inf. Theory,* vol. 55, no. 1, pp. 205-222, Jan. 2009.

[153] X. Gong, T. P. S., Chandrashekhar, J. Zhang, H. V. Poor, "Opportunistic cooperative networking: To relay or not to relay?," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 2, pp. 307-314, 2012.

[154] Z. Yan, Z. Zhand, H. Jiang, et al, "Optimal traffic scheduling in vehicular delay tolerant networks," *IEEE Commun. Letters,* vol. 16, no. 1, pp. 50-53, 2012.

[155] P. Chaporkar and A. Proutiere, "Optimal joint probing and transmission strategy for maximizing throughput in wireless systems," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 8, pp. 1546–1555, 2008.

[156] H. Jiang, L. Lai, R. Fan, and H. V. Poor, "Optimal selection of channel sensing order in cognitive radio," *IEEE Trans. Wireless Commun.*, vol. 8, no. 1, pp. 297-307, 2009.

[157] Y. Pei, Y.-C. Liang, K. C. Teh, and K. H. Li, "Energy-efficient design of sequential channel sensing in cognitive radio networks: Optimal sensing strategy, power allocation, and sensing order," *IEEE J. Sel. Areas Commun.*, vol. 29, no. 8, pp. 1648-1659, 2011.

[158] B. Li, P. Yang, J. Wang, et al. "Optimal frequency-temporal opportunity exploitation for multichannel ad hoc networks," *IEEE Trans. Parallel Distrib. Syst.,* vol. 23, no. 12, pp. 2289–2302, 2012.

[159] B. Li, P. Yang, J. Wang, et al., "Optimal action point for dynamic spectrum utilization under rayleigh fading," *Ad Hoc & Sensor Wireless Networks,* vol. 17, no. 1-2, pp. 1-32, 2012.

[160] B. Li, P. Yang, X. Y. Li, S. Tang, et al., "Almost optimal dynamically-ordered multi-channel accessing for cognitive networks," in *Proc. IEEE INFOCOM*, 2012.

[161] B. Li, P. Yang, J. Wang, et al., "Statistics exploration vs. diversity exploitation: Online sequential channel access in CRN," in *Proc. IEEE MASS,* 2012.

[162] B. Li, P. Yang, J. Wang, et al., "Optimal time-frequency diversity exploitation for multichannel system under rayleigh fading," in *IEEE Proc. MASS,* 2011.

[163] B. Li, P. Yang, X. -Y. Li, et al., "Finding optimal action point for multi-stage spectrum access in cognitive radio networks," in *Proc. IEEE ICC*, 2011.

[164] N. B. Chang and M. Liu, "Optimal channel probing and transmission scheduling for opportunistic spectrum access," *IEEE/ACM Trans. Netw.,*, vol. 17, no. 6, pp. 1805-1818, 2009.

[165] Y. Xu, Z. Gao, J. Wang, and Q. Wu, "Multichannel opportunistic spectrum access in fading environment using optimal stopping rule," in *Proc. First International Conference of Wireless Communications and Applications, (ICWCA 2011)*, pp. 275-286, 2011.

[166] Y. Xu, J. Wang, Q. Wu, et al., "Exploiting multichannel diversity in spectrum sharing systems using optimal stopping rule," *ETRI Journal*, vol. 34, no. 2, pp. 272-275, 2012.

[167] Y. Xu, L. Shen, A. Anpalagan, et al., "Energy-efficient exploration and exploitation of multichannel diversity in spectrum sharing systems," *Trans. Emerging Telecommunications Technologies,* vol. 23, no. 8, pp. 701-706, 2012.

[168] J. L. Vicario, A. Bel, J. A. Lopez-Salcedo, and G. Seco, "Opportunistic relay selection with outdated CSI: Outage probability and diversity analysis," *IEEE Trans. Wireless Commun.*, vol. 8, no. 6, pp. 2872-2876, 2009.

[169] A. Mahajan and D. Teneketzis, "Multi-Armed Bandit Problems," Foundations and Applications of Sensor Management, Springer US, 2008.

[170] T. L. Lai and H. Robbins, "Asymptotically efficient adaptive allocation rules," *Advances in Applied Mathematics,* vol. 6, no. 1, pp.4-22, 1985.

[171] V. Anantharam, P. Varaiya, and J. Walrand, "Asymptotically efficient allocation rules for the multiarmed bandit problem with multiple playspart i: IID rewards," *IEEE Trans. Automat. Contr.,* vol. AC-32, no. 11, pp. 968-975, November 1987.

[172] R. Agrawal, "Sample mean based index policies with $O(\log n)$ regret for the multi-armed bandit problem," *Advances in Applied Probability,* vol. 27, no. 4, pp. 1054-1078, December 1995.

[173] V. Anantharam, P. Varaiya, and J. Walrand, "Asymptotically efficient allocation rules for the multiarmed bandit problem with multiple playspart i: Markovian rewards," *IEEE Trans. Automat. Contr.,* vol. AC-32, no. 11, pp. 977-982, 1987.

[174] J. Gittins, "Bandit processes and dynamic allocation indices," *J. Roy. Statist. Soc.,* vol. 41, no. 2, pp. 148-177, 1979.

[175] P. Whitlle, "Restless bandits," *J. Appl. Prob.,* pp. 301-313, 1988.

[176] L. Lai, H. E. Gamal, H. Jiang, and H. V. Poor, "Cognitive medium access: Exploration, exploitation and competition," CoRR, vol. abs/0710.1385, 2007.

[177] L. Lai, H. E. Gamal, H. Jiang, and H. V. Poor, "Cognitive medium access: Exploration, exploitation and competition," *IEEE Trans. Mobile Computing,* vol. 10, no. 2, pp. 239-253, 2011.

[178] Y. Gai, B. Krishnamachari, and R. Jain, "Learning multiuser channel allocations in cognitive radio networks: A combinatorial multi-armed bandit formulation," in *IEEE Symp. Dynamic Spectrum Access Networks (DySPAN),* Singapore, April 2010.

[179] Y. Gai, B. Krishnamachari, and R. Jain, "Combinatorial network optimization with unknown variables: Multi-armed banditswith linear rewards and individual observations," *IEEE/ACM Trans. Netw.,* to appear.

[180] A. Anandkumar, N. Michael, A. Tang, and A. Swami, "Distributed algorithms for learning and cognitive medium access with logarithmic regret," *IEEE J. Sel. Areas Commun.*, vol. 29, no. 4, pp. 731-745, 2011.

[181] K. Liu and Q. Zhao, "Distributed learning in multi-armed bandit with multiple players," *IEEE Trans. Signal Process.,* vol. 58, no. 11, pp. 5667-5681, 2010.
[182] K. Liu and Q. Zhao, "Cooperative game in dynamic spectrum access with unknown model and imperfect sensing," *IEEE Trans. Wireless Commun.,* vol. 11, no. 4, 2012
[183] Y. Gai, B. Krishnamachari, "Decentralized online learning algorithms for opportunistic spectrum access,", in *Pro. IEEE GLOBECOM,* 2011.
[184] L. Chen, S. Iellamo and M. Coupechoux, "Opportunistic spectrum access with channel switching cost for cognitive radio networks," in *Proc. IEEE ICC, 2011.*
[185] T. S. Rappaport, "Wireless Communications: Principles and Practice (2nd Edition)," 2nd ed. Prentice Hall, Jan. 2002.
[186] S. Filippi, O. Cappé, and A. Garivier, "Optimally sensing a single channel without prior information: The tiling algorithm and regret bounds," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 1, pp.68-76, Feb 2011.
[187] C. Tekin and M. Liu, "Online learning in opportunistic spectrum access: A restless bandit approach," in *Proc. IEEE INFOCOM 2011.*
[188] K. Wang and L. Chen, "On optimality of myopic policy for restless multi-armed bandit problem: An axiomatic approach," *IEEE Trans. Signal Process.,* vol. 60, no. 1, pp. 300-309, 2012.
[189] Y. Gai, B. Krishnamachari, and M. Liu, "On the combinatorial multi-armed bandit problem with markovian rewards," in *Proc. IEEE GLOBECOM 2011.*
[190] J. Zhang and N. Ansari, "On assuring end-to-end QoE in next generation networks: Challenges and a possible solution," *IEEE Commun. Mag.,* vol. 49, no. 7, pp. 185-191, 2011.

Yuhua Xu received his B.S. degree in communications engineering, M.S. degree in communications and information system from Institute of Communications Engineering, Nanjing, China, in 2006 and 2008, respectively. He is currently pursuing the Ph.D. degree in communications and information system in Institute of Communications Engineering, PLA University of Science and Technology. His research interests focus on opportunistic spectrum access, learning theory, game theory, and optimization techniques. He was an Exemplary Reviewer for the IEEE Communications Letters in 2011 and 2012.

Alagan Anpalagan received the B.A.Sc.,M.A.Sc., and Ph.D. degrees in Electrical Engineering from the University of Toronto, Canada in 1995,1997 and 2001 respectively. Since August 2001, he has been with the Ryerson University, Toronto, Canada, where he co-founded WINCORE laboratory in 2002 and leads the Radio Resource Management (RRM) and Wireless Access and Networking (WAN) R&D groups. Currently, he is an Associate Professor and Program Director for Graduate Studies in the Department of Electrical and Computer Engineering at Ryerson University.

His research interests are in general, wireless communication, mobile networks and system performance analysis; and in particular, QoS-aware radio resource management, joint study of wireless physical/link layer characteristics, cooperative communications, cognitive radios, cross-layer resource optimization, and wireless sensor networking. He has published extensively in international conferences and journals in his research area.

Dr. Anpalagans editorial duties include Guest Editor, Special Issues on Radio Resource Management in 3G+ Wireless Systems (2005-06), Fairness in Radio Resource Management for Wireless Networks; and Associate Editor, EURASIP Journal of Wireless Communications and Networking. He previously served as IEEE Toronto Section Chair (2006-07) and Communications Chapter Chair (2004-05) and Technical Program Co-Chair, IEEE Canadian Conference on Electrical and Computer Engineering (2008, 2004). He is an IEEE Senior Member and a Registered Professional Engineer in the province of Ontario, Canada.

Qihui Wu received his B.S. degree in communications engineering, M.S. degree and Ph.D. degree in communications and information system from Institute of Communications Engineering, Nanjing, China, in 1994, 1997 and 2000, respectively. He is currently a professor at the PLA University of Science and Technology, China. His current research interests are algorithms and optimization for cognitive wireless networks, soft-defined radio and wireless communication systems. He is an IEEE Senior Member.

Liang Shen received his BS degree in Communications Engineering and MS degree in Communications and Information System from the Institute of Communications Engineering, Nanjing, China, in 1988 and 1991 respectively. He is currently a professor at the PLA University of Science and Technology, China. His current research interests are information theory, digital signal processing, and wireless networking.

Zhan Gao received his BS degree in Communications Engineering, MS and PhD degrees in Communications and Information System from the Institute of Communications Engineering, Nanjing, China, in 1999, 2001 and 2004, respectively. He is currently an associate professor at the PLA University of Science and Technology, China. His current research interests are cognitive radio networks, distributed optimization algorithms and digital signal processing.

Jinlong Wang received the B.S. degree in mobile communications, M.S. degree and Ph.D. degree in communications engineering and information systems from Institute of Communications Engineering, Nanjing, China, in 1983, 1986 and 1992, respectively. Since 1979, Dr. Wang has been with the Institute of Communications Engineering, PLA University of Science and Technology, where he is currently a Full Professor and the Head of Institute of Communications Engineering. He has published over 100 papers in refereed mainstream journals and reputed international conferences and has been granted over 20 patents in his research areas. His current research interests are the broad area of digital communications systems with emphasis on cooperative communication, adaptive modulation, multiple-input-multiple-output systems, soft defined radio, cognitive radio, green wireless communications, and game theory.

Dr. Wang also has served as the Founding Chair and Publication Chair of WCSP 2009, a member of the Steering Committees of WCSP2010-2012, a TPC member for several international conferences and a reviewer for many famous journals. He currently is the vice-chair of the IEEE Communications Society Nanjing Chapter and is an IEEE Senior Member.