Sensing, Probing, and Transmitting Strategy for Energy Harvesting Cognitive Radio

Keyu Wu, Hai Jiang, and Chintha Tellambura
Department of Electrical and Computer Engineering
University of Alberta, Edmonton, Alberta T6G 1H9, Canada

***Abstract*—We consider a single channel energy harvesting cognitive radio system, where the joint optimization of spectrum sensing, channel probing and transmission power control is considered with the goal to maximize the throughput. We model this control problem as a two-stage continuous-state Markov decision process with one stage for sensing and probing control, and the other for transmission power control. By utilizing the stochastic structure of the two-stage Markov decision process, we simplify the model via the notion of after-state, which reduces the state space and facilitates decision makings. Finally, the performance of the generated strategy is investigated via simulation.**

***Index Terms*—Energy harvesting, cognitive radio, spectrum sensing, power control, Markov decision process.**

I. Introduction

In an energy harvesting (EH) cognitive radio (CR) network, secondary users (SUs) can opportunistically share spectrum with primary users (PUs), leading to efficient use of spectrum and relieving the spectrum scarcity problem. With EH, SUs can harvest free energy, such as solar radiation, indoor illumination, etc., from surrounding environments, which helps to achieve environmentally friendly wireless networks.

The premise of CR is that SUs must give priority to reduce interference on PUs and can only access frequency bands that are temporally unused by PUs. Thus, SUs need to periodically sense the channel, and should not transmit whenever the channel is sensed to be busy. Spectrum sensing by SUs with constant power supply has been widely studied, and the sensing solution and the corresponding performance heavily depend on channel fading statistics [1]. In contrast, when SUs are powered by an EH process, the stochastic nature of harvested energy makes spectrum sensing significantly more challenging than the case with constant power supply.

It was shown that in EH CR networks, a SU's optimal sensing strategy [2] and ultimate achievable throughput [3] are not only affected by its obligation to protect PU's transmission but also constrained by its EH ability when the energy supply rate is below a certain level. Therefore, spectrum sensing in EH CR system should take into consideration characteristics of the harvested energy. Based on this observation, the problem of joint optimization of binary sensing decision, sensing energy and transmitting energy is investigated [4] for a single-channel EH CR system.

This work is jointly supported by China Scholarship Council/University of Alberta Scholarship, Alberta Innovates Graduate Student Scholarship and Natural Sciences and Engineering Research Council of Canada.

However, in previous works [2]–[4], the channel is modeled as static, and therefore, the sensing strategy and/or transmission power control are performed based on channel availability (not based on the channel quality). In conventional CR systems with constant power supply, it was shown that, taking channels' quality into SU's sensing decision can improve the throughput [5], [6], since the SU can make more efficient use of channel access opportunity by selecting channels with better quality.

The spectrum sensing and accessing problem under fading channel is considered in [7], where the wireless fading is represented as the amount of energy needed to accomplish one transmission. At each time slot, the SU needs to decide whether to sense or not, and which channel to sense. After a channel is sensed to be free and the channel state information (CSI) is obtained from the receiver's feedback, the SU needs to decide to transmit or not. However, in the work of [7], the battery is not rechargeable, and is with a finite energy budget, and the sensing is assumed to be perfect. An EH setting is considered in [8], as follows. The SU first probes a subset of channels. Based on the probed CSI, remaining energy and beliefs on channels' availability, the SU decides which channels to sense. After getting the sensing results, the SU further decides which channels to access. However, in [8], the channel probing is conducted before the spectrum sensing, which may suffer from high failure probability due to PU activity. Further, probing before sensing may cause severe interference to PU transmissions.

In this paper, we consider a slotted single channel EH CR system, and at each time slot the SU decides whether to sense or not, and if the channel is sensed to be free, the SU can probe the channel via sending a channel estimation sequence (CES). And given the probed CSI, the SU needs to decide on the transmission power to use. The problem is formulated as a continuous-state two-stage Markov decision process (MDP), in which one stage is responsible for the decision of sensing and probing, while the other controls the transmission power level. To the best of our knowledge, this is the first paper that uses a two-stage MDP (which can better represent a practical EH CR network than one-stage MDP) for modeling the sensing, probing and transmitting in an optimal control problem. Furthermore, we simplify the model based on the notion of after-state, which reduces the state space and facilitates decision makings. Finally, the performance of generated strategy is investigated via simulation.

The rest of paper is organized as follows. Section II describes the system model. The optimal control problem is formulated as a two-stage MDP in Section III, and it is further simplified via after-state transforming in Section IV. The performance of the generated strategy is investigated in Section V. Section VI concludes the paper.

II. System Model

Primary user model: We consider a single-channel system, where the PU transmits in a time-slotted manner, and across time slots, the PU's channel occupancy is modeled as an on-off Markov process (Fig. 1).

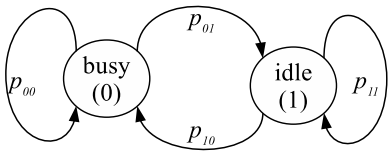

Fig. 1. The primary user's channel occupancy model

Channel sensing model: The SU's sensing functionality is implemented via an energy detector, which works under a fixed sensing duration τ_S and a predefined threshold, and each sensing process requires a fixed amount of energy e_S. The sensing result, denoted as Θ, is used by the SU to estimate the channel's true status C. And the imperfection of energy detector can be characterized by false alarm probability $p_{FA} = \text{prob}\{\Theta = 0|C = 1\}$ and miss detection probability $p_M = \text{prob}\{\Theta = 1|C = 0\}$. And $p_D = 1 - p_M$ and $p_O = 1 - p_{FA}$ represent the probability of correctly detecting the presence of the PU and correctly claiming an opportunity to access the channel, respectively. It is assumed that p_M is low enough to protect the PU system.

Sufficient statistic of channel's status: Due to the discontinuous monitoring of the channel, and imperfect sensing, the channel's true state is unknown in general. The best that the SU can do is to make decisions based on all observation information (including but not limited to sensing outcomes). The information can be summarized as a scale sufficient statistic, known as belief variable $p \in [0, 1]$, which represents the SU's belief on the channel's availability.

Energy harvesting model: We assume the SU also works under a time-slotted scheme, which is synchronized with the PU. The SU can harvest energy from its surroundings, which is not affected by either the PU's transmission or the SU's own actions. The harvested energy is assumed to arrive as an energy package at the beginning of each time slot. The amount of harvested energy at each time is modeled as an independent and identically distributed (i.i.d.) random variable (r.v.), denoted as E_H, with probability density function (pdf) $\mathrm{f}_E(\cdot)$. The SU equips a finite battery, with capacity E_{max}. The amount of remaining energy in battery is denoted as b.

Data transmission model: We assume the SU always has data to send. The channel gain between the SU and its receiver is assumed to be under block fading, and modeled as an i.i.d. r.v., denoted as H, with pdf $\mathrm{f}_H(\cdot)$. We assume the SU can adapt its transmission rate to different channel states via changing its transmission power, which can only be set to a finite number of levels. The CSI is available by probing channel. In order to get CSI, the SU can send a CES, if it senses a free channel. And if the channel is indeed free from the PU, the SU's receiver is assumed to always be able to receive the CES; otherwise, a collision will occur, and the decoding of CES is assumed to always fail. Upon successfully receiving the CES, the SU's receiver can estimate CSI and send it back to the SU. And the feedback is assumed to be always successful. We assume this whole channel probing process costs a fixed amount energy, denoted as e_P, and a fixed time duration, denoted as τ_P, no matter whether the feedback is received or not.

MAC protocol: As shown in Fig. 2, the SU's working time slot is further divided into four sub-slots. At the beginning of sensing sub-slot, the SU gets a harvested energy package. Based on the harvested energy e_H, current belief p, and battery level b, the SU needs to decide whether to sense the channel or not. If "to sense" is decided, and the channel is sensed to be free, i.e., $\Theta = 1$, it needs to decide whether to probe the channel or not. If the SU decides to probe the channel, it will send a CES to its receiver. And if the feedback from the receiver is received, the SU knows the CSI, and in the rest of the time slot, it needs to decide the transmission energy to use, which is denoted as e_T and can be taken from set $\mathbf{E_T}$ (the set $\mathbf{E_T}$ includes a finite number of energy levels). And if any of conditions is not satisfied, the SU will remain idle during the remaining time slot, repeat the procedure at the next time slot, and continue forever.

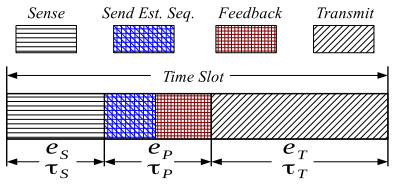

Fig. 2. Time slot structure

III. Two-stage MDP Formulation

A. FSM for MAC protocol

In this part, we will use a finite step machine[1] (FSM), as shown in Fig. 3, to elaborate on the MAC protocol introduced in Section II.

(1.1) At the sensing step of time slot t, the SU, initially with battery level b_t^S, belief p_t^S,[2] and harvested energy e_{Ht}, needs to decide whether to sense or not. If the SU chooses not to sense, it remains idle until the sensing step of time slot $t + 1$. And it will have energy $b_{t+1}^S = \phi(b_t^S + e_{Ht})$, where $\phi(b)$ is defined as:

$$\phi(b) \triangleq \max\{\min\{b, E_{max}\}, 0\}, \tag{1}$$

[1]Instead of the conventional terminology, finite state machine, we use "step" in order to distinguish the "state" that will be introduced in the MDP model in the next subsection.

[2]Superscript S represents sensing, and subscript t means slot index.

which is used to describe battery's transition after energy replenishment or consumption. And due to channel occupancy's transition, the SU's belief will change to $p_{t+1}^S = \psi(p_t^S)$, where $\psi(p)$ is defined as:

$$\psi(p) \triangleq \text{prob}\{C_{t+1} = 1 | p_t = p\} = p \cdot p_{11} + (1-p) \cdot p_{01}, \quad (2)$$

which is used to describe the SU's belief transition across time slot. The harvested energy of next slot e_{Ht+1} is drawn from pdf $\text{f}_E(\cdot)$.

(1.2) If the SU chooses to sense, with probability $1 - \text{p}_\Theta(p_t^S)$, it will get a negative sensing result, i.e., $\Theta = 0$, where $\text{p}_\Theta(p)$ is defined as:

$$\text{p}_\Theta(p) \triangleq \text{prob}\{\Theta = 1 | p\} = p \cdot p_O + (1-p) \cdot p_M. \quad (3)$$

Then it will remain idle until the sensing step of slot $t+1$, and we have $b_{t+1}^S = \phi(\phi(b_t^S + e_{Ht}) - e_S)$, and $p_{t+1}^S = \psi(\text{p}_N(p_t^S))$, where $\text{p}_N(p)$ means the probability that the channel is idle given belief p and negative sensing result, i.e.,

$$\text{p}_N(p) \triangleq \text{prob}\{C = 1 | p, \Theta = 0\} = \frac{p \cdot p_{FA}}{p \cdot p_{FA} + (1-p) \cdot p_D}. \quad (4)$$

(1.3) If the SU chooses to sense, with probability $\text{p}_\Theta(p_t^S)$, it will get a positive sensing result, i.e., $\Theta = 1$. Then it reaches the probing step, and at this moment, the battery level becomes $b_t^P = \phi(\phi(b_t^S + e_{Ht}) - e_S)$,[3] and the belief transits to $p_t^P = \text{p}_P(p_t^S)$, where $\text{p}_P(p)$ is the probability that channel is idle, given belief p and positive sensing result, i.e.,

$$\text{p}_P(p) = \text{prob}\{C = 1 | p, \Theta = 1\} = \frac{p \cdot p_O}{p \cdot p_O + (1-p) \cdot p_M}. \quad (5)$$

(2.1) At the probing step of time t, if the SU with (p_t^P, b_t^P) chooses not to probe, it will keep idle until the sensing step of slot $t+1$, and the battery remains the same $b_{t+1}^S = b_t^P$, and the belief becomes $p_{t+1}^S = \psi(p_t^P)$.

(2.2) If the SU chooses to probe, and after sending CES, there is probability $1 - p_t^P$ that channel is busy, and therefore, the receiver's feedback cannot be obtained. And then it keeps idle until the sensing step of slot $t+1$ with battery $b_{t+1}^S = \phi(b_t^P - e_P)$ and belief $p_{t+1}^S = p_{01}$.

(2.3) After sending CES, with probability p_t^P, it can get the feedback, and observe the channel gain information, h_t, which is draw from $\text{f}_H(\cdot)$. And the SU reaches the transmitting step. At this moment, it can make sure that the channel is free from the PU, i.e., $p_t^T = 1$,[4] and the remaining energy is $b_t^T = \phi(b_t^P - e_P)$.

(3) At the transmitting step of time t, the SU decides the amount of energy $e_T \in \mathbf{E_T}$ to use for transmission. After data transmission, it transits to the sensing step of slot $t+1$ with battery $b_{t+1}^S = \phi(b_t^T - e_T)$ and belief $p_{t+1}^S = p_{11}$.

B. Two-stage MDP based on FSM

Based on the FSM, in this part, we will use a MDP model to mathematically formulate the control problem. A

[3]Superscript P represents probing.

[4]Superscript T represents transmitting.

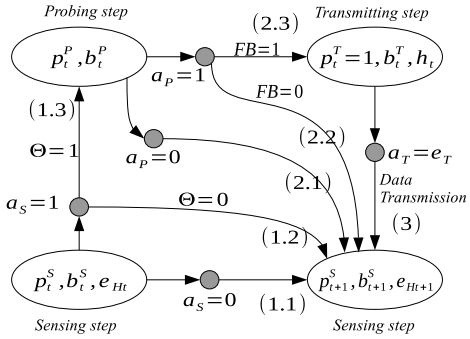

Fig. 3. FSM for MAC protocol

MDP can be fully characterized by specifying the 4-tuple ($\mathbb{S}$, $\{\mathbb{A}(s)\}_s$, $\text{f}(\cdot|s,a)$, $r(s,a)$), namely state space, allowed actions at different states, state transition kernel, and reward associated with each state-action pair, which are described as follows.

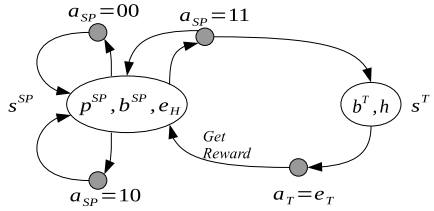

Fig. 4. Two-stage MDP

(1) In order to reduce the state space, the sensing step and probing step are merged into one stage (represented by superscript SP) via jointly deciding the actions of sensing and probing at the beginning of sensing step. And also observing that at transmitting step, the belief is always equal to 1, and there is no need to represent it. Therefore, the state space $\mathbb{S}$ is divided into two classes: 1) sensing-probing state $s^{SP} = [b^{SP}, p^{SP}, e_H]$, with $b^{SP} \in [0, E_{max}]$, $p^{SP} \in [0,1]$ and $e_H \in [0, \infty)$; and 2) transmitting state $s^T = [b^T, h]$, with $b^T \in [0, E_{max}]$ and $h \in [0, \infty)$.

(2) At sensing-probing states s^{SP}, the full set of available actions are "not to sense", "to sense but not to probe", and "to sense and to probe if possible", i.e., we have $a_{SP} \in \mathbb{A}(s^{SP}) = \{00, 10, 11\}$. If available energy $\phi(b^{SP} + e_H)$ is less than $e_S + e_P$, the available actions $\mathbb{A}(s^{SP})$ is limited to $\{00, 10\}$; and if it is less than e_S, we have $\mathbb{A}(s^{SP}) = \{00\}$. And at transmitting state s^T, the available actions are "transmission energy to use", i.e., $a_T \in \mathbb{A}(s^T) = \mathbf{E}_T$.

(3) $\text{f}(\cdot|s,a)$ is a pdf of next state $s' \in \mathbb{S}$ given initial state s and the taken action a. Denote $\delta(\cdot)$ as the Dirac delta function, which is used to generalize $\text{f}(\cdot|s,a)$ to include discrete transition components. And we can read out the state transition kernel following the description of the FSM. Starting from $s_t^{SP} = [p_t^{SP}, b_t^{SP}, e_{Ht}]$, it may transit to $s_{t+1}^{SP} = [p_{t+1}^{SP}, b_{t+1}^{SP}, e_{Ht+1}]$ or $s_t^T = [b_t^T, h_t]$ depending on chosen actions, with $\text{f}(\cdot|s_t^{SP}, a_{SP})$ shown in (6), (7), (8) and (9) on top of next page. From transmitting state $s_t^T = [b_t^T, h_t]$, it can only transit to $s_{t+1}^{SP} = [p_{t+1}^{SP}, b_{t+1}^{SP}, e_{Ht+1}]$, with $\text{f}(\cdot|s_t^T, a_T)$ shown

$$\mathrm{f}(s_{t+1}^{SP}|s_t^{SP}, a_{SP}=00) = \delta(p_{t+1}^{SP} - \psi(p_t^{SP}))\delta(b_{t+1}^{SP} - \phi(b_t^{SP} + e_{Ht}))\,\mathrm{f}_E(e_{Ht+1}) \tag{6}$$

$$\mathrm{f}(s_{t+1}^{SP}|s_t^{SP}, a_{SP}=10) = [(1-\mathrm{p}_\Theta(p_t^{SP}))\delta(p_{t+1}^{SP} - \psi(\mathrm{p}_N(p_t^{SP}))) + \mathrm{p}_\Theta(p_t^{SP})\delta(p_{t+1}^{SP} - \psi(\mathrm{p}_P(p_t^{SP})))]\times \delta(b_{t+1}^{SP} - \phi(\phi(b_t^{SP} + e_{Ht}) - e_S))\,\mathrm{f}_E(e_{Ht+1}) \tag{7}$$

$$\mathrm{f}(s_{t+1}^{SP}|s_t^{SP}, a_{SP}=11) = \mathrm{p}_\Theta(p_t^{SP})(1-\mathrm{p}_P(p_t^{SP}))\delta(p_{t+1}^{SP} - p_{01})\delta(b_{t+1}^{SP} - \phi(\phi(b_t^{SP} + e_{Ht}) - e_S - e_P))\times \mathrm{f}_E(e_{Ht+1}) + (1-\mathrm{p}_\Theta(p_t^{SP}))\delta(p_{t+1}^{SP} - \psi(\mathrm{p}_N(p_t^{SP})))\delta(b_{t+1}^{SP} - \phi(\phi(b_t^{SP} + e_{Ht}) - e_S))\,\mathrm{f}_E(e_{Ht+1}) \tag{8}$$

$$\mathrm{f}(s_t^{T}|s_t^{SP}, a_{SP}=11) = \mathrm{p}_\Theta(p_t^{SP})\,\mathrm{p}_P(p_t^{SP})\delta(b_t^{T} - \phi(\phi(b_t^{SP} + e_{Ht}) - e_S - e_P))\,\mathrm{f}_H(h_t) \tag{9}$$

$$\mathrm{f}(s_{t+1}^{SP}|s_t^{T}, a_T = e_T) = \delta(p_{t+1}^{SP} - p_{11})\delta(b_{t+1}^{SP} - \phi(b_t^{T} - e_T))\,\mathrm{f}_E(e_{Ht+1}) \tag{10}$$

in (10). Note that we treat $\mathrm{f}_H(\cdot)$ and $\mathrm{f}_E(\cdot)$ as generalized pdf's, which allows the development to equally cover the discrete or mixed r.v.s model for E_H and H.

(4) At sensing-probing states, because there is no data transmission happened yet, the reward is set to 0, i.e.,

$$r(s_t^{SP}, a^{SP}) = 0. \tag{11}$$

At transmitting states, the Shannon's formula is used to bridge the relationship between the transmission energy and data sent for simplicity, and our method can be extended to other formulations. Therefore, we have

$$r(s_t^T, a_T = e_T) = \tau_T W \log_2(1 + \frac{e_T h_t}{\tau_T N_0 W})\mathbf{1}(b_t^T \geq e_T), \tag{12}$$

where W is the spectrum bandwidth of the channel, N_0 is the thermal noise spectrum density, and $\mathbf{1}(\cdot)$ is the indicator function. Here we put one technical restriction on the r.v. H.

Assumption 1. *Given any battery level b^T and any transmission energy e_T, $\mathbb{E}[r(s^T, e_T)]$ exists and is bounded.*

C. Optimal control via classical MDP formulation

Denote set Π as all stationary deterministic strategies, which are maps from $s \in \mathbb{S}$ to $\mathbb{A}(s)$. We limit the control within Π. The optimization goal is to find a strategy which achieves the highest expected discounted accumulated reward for any starting state. To be specific, given any $\pi \in \Pi$, we define a so-called value function $V^\pi : \mathbb{S} \to \mathbb{R}$ for π as follows,

$$V^\pi(s) \triangleq \mathbb{E}[\sum_{\tau=t}^{\infty} \gamma^{\tau-t} r(S_\tau, a_\tau)|S_t = s, a_\tau = \pi(s_\tau)], \tag{13}$$

where $\gamma \in [0, 1)$ is a constant, called discounting factor, and the expectation is defined by the state transition kernel (6), (7), (8), (9) and (10). For the optimal strategy $\pi^* \in \Pi$, we have $V^{\pi^*}(s) = \sup_{\pi\in\Pi}\{V^\pi(s)\}$, $\forall s$. It is well-known that π^* can be identified by the so-called optimal Bellman equation [9, p. 154], which is defined as follows,

$$V(s) = \max_{a\in\mathbb{A}(s)}\{r(s,a) + \gamma\mathbb{E}[V(S')|s,a]\}, \tag{14}$$

where S' denotes the random next state given current state s and taken action a. The solution to (14), denoted as V^*, can be used to generate π^* as follows,

$$\pi^*(s) = \arg\max_{a\in\mathbb{A}(s)}\{r(s,a) + \gamma\mathbb{E}[V^*(S')|s,a]\}. \tag{15}$$

Note that, given V^*, decision generating with (15) needs to compute expectation, which can be time consuming. One solution is to pre-calculate the optimal action at each state and create a strategy lookup table, which is, however, space consuming. In next section, we will show via utilizing the stochastic structure of the two-stage MDP model, the expectation for action generating is eliminated, and furthermore, the required space to represent a value function can also be reduced.

IV. Simplification Via After-state MDP

A. Stochastic structure of the two-stage MDP

Notice that each state consists of an endogenous component and an exogenous component. Specifically, for $s^{SP} \in \mathbb{S}^{SP}$, the endogenous component, d^{SP}, is $[p^{SP}, b^{SP}]$, and denote the set of all possible d^{SP} as $\mathbb{D}^{SP}$, and the exogenous component, x^{SP}, is e_H, and denote the set of all possible x^{SP} as $\mathbb{X}^{SP}$. Similarly, for $s^T \in \mathbb{S}^T$, the endogenous component, d^T, is b^T, and denote the set of all possible d^T as $\mathbb{D}^T$, and the exogenous component, x^T, is h, and denote the set of all possible x^T as $\mathbb{X}^T$. We denote $\mathbb{D} = \mathbb{D}^{SP} \cup \mathbb{D}^T$, and $\mathbb{X} = \mathbb{X}^{SP} \cup \mathbb{X}^T$. Therefore, we have $\mathbb{S}^{SP} = \mathbb{D}^{SP} \times \mathbb{X}^{SP}$, $\mathbb{S}^T = \mathbb{D}^T \times \mathbb{X}^T$, and $\mathbb{S} = \mathbb{D} \times \mathbb{X}$.

Checking the state transition kernel defined via (6), (7), (8), (9) and (10), we can see that, given state $s = [d, x]$, and action $a \in \mathbb{A}(s)$, transitions to next state $s' = [d', x']$ (noting that $()'$ means state of next slot) have following two properties. 1) There are $N(d, a)$ possible transitions to d'. And for the i-th transition, the resulted endogenous component d' can be deterministically expressed as a function $\varrho_i(d, x, a)$, and the associated probability can be expressed as a function $p_i(d, a)$. 2) The transition to x' depends on (s, a) through $\varrho_i(d, x, a)$. Specifically, x' obeys $\mathrm{f}_E(\cdot)$, if $\varrho_i(d, x, a) \in \mathbb{D}^{SP}$; x' obeys $\mathrm{f}_H(\cdot)$, if $\varrho_i(d, x, a) \in \mathbb{D}^T$. And this defines a conditional pdf, and denote it as $\mathrm{f}_X(x'|\varrho_i(s, a))$. The values of N, p_i, ϱ_i and f_X for different d, x and a are listed in Table I. Therefore, the state transition kernel $\mathrm{f}(s'|s, a)$ can be rewritten as:

$$\begin{aligned}\mathrm{f}(s'|s,a) &= \mathrm{f}((d',x')|(d,x),a) \\ &= \sum_{i=1}^{N(d,a)} p_i(d,a)\delta(d' - \varrho_i(s,a))\,\mathrm{f}_X(x'|\varrho_i(s,a)).\end{aligned} \tag{16}$$

TABLE I
TRANSITION MODEL DECOUPLED VIA AFTER-STATE

d	x	$a \in \mathbb{A}(d,x)$	$N(d,a)$	$p_i(d,a)$	$d' = \varrho_i(d,x,a)$	$\mathrm{f}_X(x'\|\varrho_i)$
$[b,p]$	e_H	00	1	1	$[\psi(p), \phi(b+e_H)]$	$\mathrm{f}_E(\cdot)$
		10	2	$\mathrm{p}_\Theta(p)$	$[\psi(\mathrm{p}_P(p)), \phi(\phi(b+e_H)-e_S)]$	$\mathrm{f}_E(\cdot)$
				$1-\mathrm{p}_\Theta(p)$	$[\psi(\mathrm{p}_N(p)), \phi(\phi(b+e_H)-e_S)]$	$\mathrm{f}_E(\cdot)$
		11	3	$\mathrm{p}_\Theta(p)\,\mathrm{p}_P(p)$	$\phi(\phi(b+e_H)-e_S-e_P)$	$\mathrm{f}_H()$
				$\mathrm{p}_\Theta(p)(1-\mathrm{p}_P(p))$	$[p_{01}, \phi(\phi(b+e_H)-e_S-e_P)]$	$\mathrm{f}_E(\cdot)$
				$1-\mathrm{p}_\Theta(p)$	$[\psi(\mathrm{p}_N(p)), \phi(\phi(b+e_H)-e_S)]$	$\mathrm{f}_E(\cdot)$
b	h	e_T	1	1	$[p_{11}, \phi(b-e_T)]$	$\mathrm{f}_E(\cdot)$

It can be seen that given certain realization of endogenous component $\varrho_i(d,x,a)$, the state of next time slot $s' = [d', x']$ is independent of (s,a). Therefore, the value of $\varrho_i(d,x,a)$ is sufficient to determine the standing of current situation, which is similar to the "state" in the classical MDP model. And we name all possible $\varrho_i(d,x,a)$ as after-states, which is formally defined in the next part.

B. Optimal Bellman equation for after-states

Define the after-state space as $\Xi \triangleq \{\beta \in \mathbb{D} | \beta = \varrho_i(s,a), \forall (s,a,i)\}$, i.e., Ξ is the maximum subset of $\mathbb{D}$ such that for every element $\beta \in \Xi$, we can find a state s and action $a \in \mathbb{A}(s)$ such that $\beta = \varrho_i(s,a)$ for some i.

We define the optimal Bellman equation over after-state space as

$$J(\beta) = \gamma \mathop{\mathbb{E}}_{X'|\beta}[\max_{a' \in \mathbb{A}([\beta, X'])} \{r(\beta, X', a') + \sum_{i=1}^{N(\beta,a')} p_i(\beta, a') J(\varrho_i(\beta, X', a'))\}], \quad (17)$$

where X' is a r.v. presenting x' and $\mathop{\mathbb{E}}_{X'|\beta}[\cdot]$ means taking expectation over $\mathrm{f}_X(x'|\beta)$. The following theorem states the existence of solution to (17) and also gives a method to obtain the solution. The proof is omitted due to space limitation.

Theorem 1. *Given Assumption 1, there is a unique J^* that satisfies (17). And J^* can be calculated via value iteration algorithm, i.e., with J_0 being arbitrary bounded function, the sequence of functions $\{J_l\}_{l=0}^{L}$ defined by following iteration equation*

$$J_{l+1}(\beta) \leftarrow \gamma \mathop{\mathbb{E}}_{X'|\beta}[\max_{a' \in \mathbb{A}([\beta, X'])} \{r(\beta, X', a') + \sum_{i=1}^{N(\beta,a)} p_i(\beta, a') J_l(\varrho_i(\beta, X', a'))\}], \quad (18)$$

converges to J^* when $L \to \infty$.

C. Optimal control via J^*

We will now establish the relationship between J^* and V^*. Assuming the existence of V^*, define a function G over the after-state space as $G(\beta) \triangleq \gamma \mathop{\mathbb{E}}_{X'|\beta}[V^*(\beta, X')]$. Expanding V^* with (14), and also using (16), we have $G(\beta) = \gamma \mathop{\mathbb{E}}_{X'|\beta}[\max_{a' \in \mathbb{A}([\beta,X'])} \{r(\beta, X', a') + \sum_{i=1}^{N(\beta,a')} p_i(\beta, a') G(\varrho_i(\beta, X', a'))\}]$, which is just the optimal Bellman equation defined in (17). According to Theorem 1, this implies $G = J^*$. Therefore, from (14) and the definition of G, V^* can be expressed via J^* as

$$V^*([d,x]) = \max_{a \in \mathbb{A}([d,x])} \{r(d,x,a) + \sum_{i=1}^{N(d,a)} p_i(d,a) J^*(\varrho_i(d,x,a))\}, \quad (19)$$

which means the existence of J^* implies the existence of V^*. Similarly, the optimal strategy π^* defined by (15) can be expressed via J^* as

$$\pi^*([d,x]) = \arg\max_{a \in \mathbb{A}([d,x])} \{r(d,x,a) + \sum_{i=1}^{N(d,a)} p_i(d,a) J^*(\varrho_i(d,x,a))\}. \quad (20)$$

With discretization over Ξ, $\mathrm{f}_E(\cdot)$ and $\mathrm{f}_H(\cdot)$, which converts the problem into finite state case, the value iteration algorithm (18) can be used to compute the approximated value function of J^*. And the approximated value function can further be used to generate a near optimal strategy with (20).

We have shown that V^* and J^* are theoretically equivalent in achieving the optimal control. However, compared with V^*, J^* does not need to explicitly represent the exogenous space $\mathbb{X}$. From the implementation point of view, this can reduce the computation complicity in solving the MDP. Furthermore, unlike (15), if J^* is known, (20) can be used to generate actions without the need of taking expectation. This is favorable, since it eliminates the requirement of either time consuming expectation computation or space consuming strategy lookup table.

V. NUMERICAL SIMULATION

The performance of the generated strategy from the after-state MDP, which is named as Near Opt, is investigated. For comparison, we further construct two partially greedy strategies, namely Greedy-Sensing-Probing (GSP) and Greedy-Transmitting (GT). GSP always senses and probes the channel whenever it is able to do so, but carefully chooses the transmission energy. It is constructed via constraining $\mathbb{A}(s)$ in (18) and

(20) to have only the greedy action at sensing-probing states. Similarly, we construct GT, which only uses the maximum power level for transmission, but carefully chooses the sensing and probing action. Finally, we construct a purely greedy strategy, GSPT, which only uses the maximum power level for transmission, and always senses and probes the channel whenever the resulted energy at transmitting stage is enough for transmission.

We set $p_{00} = 0.2$, $p_{11} = 0.1$, $p_{FA} = 0.1$, $p_M = 0.01$, $W = 1$ MHz, $\tau_T = 10$ ms and $N_0B = -107$ dBm. The static power attenuation is set to be -100 dB and fast fading is modeled as Rayleigh fading with mean equal to 3. We set $e_S = 10^{-6}$ Joule, $e_P = 10^{-6}$ Joule, $e_T \in \{0, 1, 2, 3\} \times 10^{-6}$ Joule and $E_{max} = 10^{-5}$ Joule. The harvested energy process E_H is modeled as the gamma distribution with variance 10^{-9}, and with mean ranging from 10^{-7} to 2.5×10^{-5}.

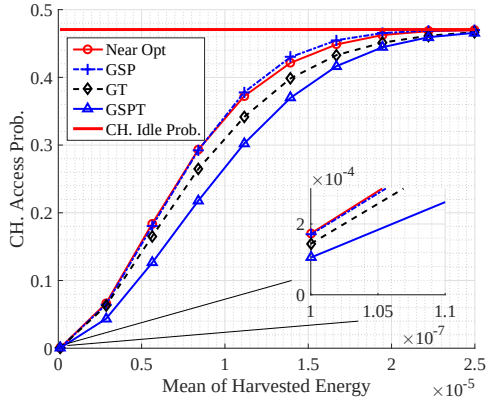

Fig. 5. Access probability under different mean of harvested energy

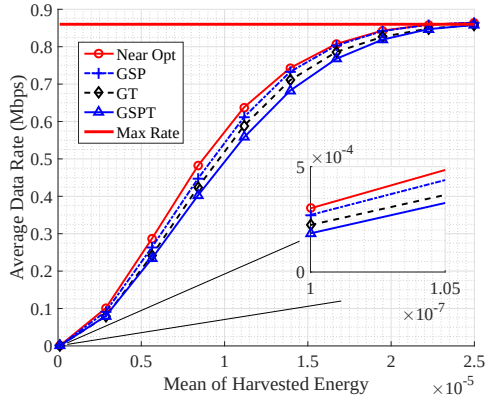

Fig. 6. Data rate under different mean of harvested energy

Fig. 5 shows the probabilities of accessing the channel for different strategies under different mean of harvested energy, which are upper bounded by the channel's idle probability $p_{01}/(p_{01} + p_{10})$. And Fig. 6 shows the ultimate achieved data rates for all strategies, which is upper bounded by $p_{01}/(p_{01} + p_{10}) \cdot p_O \cdot \mathbb{E}[W \log_2(1 + \frac{e_T^M H}{\tau_T N_0 W})] \approx 0.86$ Mbps, where $e_T^M = \max\{\mathbf{E_T}\}$. It can be seen that all strategies can exploit the increasing harvested energy to obtain more channel access opportunity and also more data throughput, and at highest energy supply rate, the purely greedy use of energy also trends to be optimal. Comparing Near Opt with GSPT, it can be observed that the lower harvested energy rate is, the higher the performance improves by intelligent use of energy. To be specific, under the smallest mean of harvested energy, Near Opt can achieve around 65% more data throughput than GSPT does; and when harvested energy supply increases, the relative increase achieved via Near Opt in the throughput, although less significant, is still not negligible. It is interesting to note that GSP captures almost the same channel access opportunity as Near Opt does at low energy rate region, and captures even more when energy rate is high. However, from the ultimate data throughput point of view, Near Opt always outperforms GSP under all energy rates. The reason is, although GSP's aggressive sensing-probing and intelligent transmitting strategy can achieve reasonable access probability, it wastes too much energy by blindly sensing and probing, which leaves less energy to transmitting stage. It can also be observed that GSP achieves more performance gain than GT. This is because the greedy transmission strategy needs more energy than greedy sensing and probing strategy, and without adapting transmission power to fading channel, GT cannot fully utilize the channel access opportunity obtained at sensing-probing stage. This observation also confirms the importance of the joint optimization of sensing, probing and transmitting under fading channel.

VI. Conclusion

In this paper, we have studied the optimal sensing, probing and power control problem under fading channel in EH CR systems. The problem is modeled as a two-stage continuous state MDP, which is further simplified via after-state formulation. Finally the performance of the generated strategy is investigated via simulation.

References

[1] S. Atapattu, C. Tellambura, and H. Jiang, *Energy detection for spectrum sensing in cognitive radio*. Springer, 2014.

[2] S. Park, H. Kim, and D. Hong, "Cognitive Radio Networks with Energy Harvesting," *IEEE Trans. Wirel. Commun.*, vol. 12, no. 3, pp. 1386–1397, 2013.

[3] S. Park and D. Hong, "Achievable throughput of energy harvesting cognitive radio networks," *IEEE Trans. Wirel. Commun.*, vol. 13, no. 2, pp. 1010–1022, 2014.

[4] A. Sultan, "Sensing and Transmit Energy Optimization for an Energy Harvesting Cognitive Radio," *IEEE Wirel. Commun. Lett.*, vol. 1, no. 5, pp. 500–503, 2012.

[5] T. Shu and M. Krunz, "Throughput-efficient sequential channel sensing and probing in cognitive radio networks under sensing errors," in *Proc. ACM Int. Conf. Mob. Comput. Netw.* ACM, 2009, pp. 37–48.

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

[7] 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.

[8] J. J. Pradha, S. S. Kalamkar, and A. Banerjee, "Energy Harvesting Cognitive Radio with Channel-Aware Sensing Strategy," *IEEE Commun. Lett.*, pp. 1–4, 2014.

[9] M. L. Puterman, *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. New York: John Wiley & Sons, 1994.