sensors

MDPI

Article

Outage Probability Minimization for Energy Harvesting Cognitive Radio Sensor Networks

Fan Zhang *, Tao Jing, Yan Huo and Kaiwei Jiang

School of Electronic and Information Engineering, Beijing Jiaotong University, Beijing 100044, China; tjing@bjtu.edu.cn (T.J.); yhuo@bjtu.edu.cn (Y.H.); kwjiang@bjtu.edu.cn (K.J.)
* Correspondence: fanzhang2@bjtu.edu.cn; Tel.: +86-10-5168-4060

Academic Editor: Leonhard M. Reindl
Received: 29 November 2016; Accepted: 17 January 2017; Published: 24 January 2017

Abstract: The incorporation of cognitive radio (CR) capability in wireless sensor networks yields a promising network paradigm known as CR sensor networks (CRSNs), which is able to provide spectrum efficient data communication. However, due to the high energy consumption results from spectrum sensing, as well as subsequent data transmission, the energy supply for the conventional sensor nodes powered by batteries is regarded as a severe bottleneck for sustainable operation. The energy harvesting technique, which gathers energy from the ambient environment, is regarded as a promising solution to perpetually power-up energy-limited devices with a continual source of energy. Therefore, applying the energy harvesting (EH) technique in CRSNs is able to facilitate the self-sustainability of the energy-limited sensors. The primary concern of this study is to design sensing-transmission policies to minimize the long-term outage probability of EH-powered CR sensor nodes. We formulate this problem as an infinite-horizon discounted Markov decision process and propose an ϵ-optimal sensing-transmission (ST) policy through using the value iteration algorithm. ϵ is the error bound between the ST policy and the optimal policy, which can be pre-defined according to the actual need. Moreover, for a special case that the signal-to-noise (SNR) power ratio is sufficiently high, we present an efficient transmission (ET) policy and prove that the ET policy achieves the same performance with the ST policy. Finally, extensive simulations are conducted to evaluate the performance of the proposed policies and the impaction of various network parameters.

Keywords: cognitive radio; energy harvesting; sensor networks; Markov decision process

1. Introduction

During the last decade, bandwidth demand for the limited spectrum has been greatly increasing due to the explosive growth of wireless services. The current static frequency allocation schemes, with a severe underutilization of the licensed spectrum over vast temporal and geographic expanses [1], cannot support numerous emerging wireless services. This motivates the concept of cognitive radio (CR) [2–4], which has been envisioned as an intelligent and promising approach to alleviate the problem of spectrum utilization inefficiency. In CR networks (CRNs), unlicensed secondary users (SUs) opportunistically access the spectrum dedicated to some licensed primary users (PUs) without interfering with the PU operation [5]. Through enabling the CR users to dynamically access the available bands in the licensed spectrum, spectrum efficiency can be improved significantly.

The wireless sensor network (WSN), which is capable of performing event monitoring and data gathering, has been applied to various fields, including environment monitoring, military surveillance, smart homes and other industrial applications [6,7]. Currently, most WSNs work in the license-free band and are expected to suffer from heavy interference caused by other applications sharing the same spectrum. It is therefore imperative to employ CR in WSNs to exploit the dynamic spectrum access techniques, hence giving birth to the CR sensor networks (CRSNs) [8,9]. In CRSNs, in order

to guarantee the quality-of-service (QoS) of primary users, it is indispensable for CR sensor nodes to sense the licensed spectrum to ensure that the spectrum is free of primary activities before data transmission. The exclusive operation of spectrum sensing along with subsequent data transmission results in high energy consumption in CRSNs, which traditionally operate powered by batteries. Consequently, one of the looming challenges that threatens the successful deployment of CRSNs is the energy efficiency [6,10].

Energy harvesting (EH) technology, which is used to replenish energy from various energy sources, such as solar, wind and thermal, has been flagged as one of the effective approaches for improving the energy efficiency with more eco-friendliness [11]. Compared with traditional communication devices powered by batteries, EH-enabled devices could scavenge unlimited energy from the ambient environment energy sources, which enable them to operate continuously without battery replacement [6]. This self-sustainable feature is very important because in many situations, periodically replacing or recharging batteries may be inconvenient or even impossible due to various physical restrictions [12]. Besides, powering wireless networks with renewable energy source could also significantly reduce the harmful effects to the environment caused by fossil-based energy. Furthermore, energy harvesting systems can be built inexpensively in small dimensions, which could be a significant advantage in the manufacturing of small communication devices, such as sensor nodes [13]. Recently, apart from traditional energy sources (e.g., solar, wind, thermal), the ambient radio signal is also regarded as a helpful optional source, which can be consistently available regardless of the time and location in urban areas [14]. In light of the above advanced features, applying EH in CRSNs to improve energy efficiency has become increasingly eye-catching recently [10,15–17].

In this paper, we consider a time-slotted EH CR sensor network, where the secondary sensor node (also called SU) with a finite-capacity battery has no fixed energy supply and is powered exclusively by energy harvested from the ambient environment. There are multiple tradeoffs involved in the design of the parameters to achieve the optimal system performance of the SU. First, due to the existence of sensing errors, with a longer time allocated for channel sensing, the SU can acquire the status of a licensed spectrum with higher accuracy, such that the performance of the SU may be improved. However, in the slotted operating mode, with more time allocated for channel sensing, less time remains for data transmission, leading to possible performance reduction. Besides, as the amount of transmitting power used upon transmission will affect both the performance and energy consumption, a crucial challenge lies in adaptively tuning the transmission power levels according to the energy replenishment process, as well as channel variation. An overly conservative power allocation may limit the performance by failing to take full advantage of harvested energy, while an overly aggressive allocation of power may cause the energy in the battery to run out and affect the performance of the future time slots. Additionally, different from traditional CR systems with a fixed power supply, the energy consumption on the channel sensing is non-negligible in EH CRSNs; therefore, the problem of designing parameters, which should jointly consider the energy consumption of channel sensing and data transmission, as well as the dynamic battery replenishment process, becomes even more complicated than the traditional CR systems.

The objective of this paper is to minimize the long-term outage probability of the secondary sensor node by adapting the sensing time and the transmit power to the system states, including the battery energy, channel fading and the arrival energy by harvesting. The main contributions of this work are summarized as follows:

1. Considering the status of primary channels, the diversity of channel conditions, the energy replenishment process, as well as the imperfection of spectrum sensing, we investigate the joint optimization of channel sensing and adaptive transmit power allocation to minimize the SU's long-term outage probability. The above design problem is formulated as a discounted Markov decision process (MDP).
2. We theoretically prove the existence of an optimal stationary deterministic policy and obtain the ϵ-optimal sensing-transmission (ST) policy, which specifies the allocation of sensing time

and transmission power through using the value iteration in the MDP. Moreover, an interesting structural property regarding the optimal transmission policy is obtained. It is proven that the optimal long-term outage probability is non-increasing with the amount of the available energy in the battery.

3. For a special case where the signal-to-noise (SNR) power ratio is sufficiently high, we propose an efficient transmission (ET) policy with reduced computational complexity. It is theoretically proven that the efficient transmission policy achieves the same performance as the proposed sensing-transmission policy when the SNR is sufficiently high, which has also been validated through computer simulations.
4. We provide extensive simulation results to compare the performance of the sensing-transmission policy and the efficient transmission policy with that of a benchmark policy. It is shown that the proposed sensing-transmission policy achieves significant gains with respect to the benchmark policy, and both the sensing-transmission and the efficient transmission policies converge to the same value in high SNR regions. In addition, the impacts of various system parameters on the performance of proposed policies are also investigated.

The rest of the paper is organized as follows. The related work is reviewed in Section 2. The network model and the related assumptions are presented in Section 3. We formulate the outage probability minimization problem as an MDP in Section 4. The proposed policies and the related theorems are illustrated in Section 5. The performance and characteristics of the proposed policies are evaluated through numerical results in Section 6. Finally, we conclude this paper in Section 7.

2. Related Work

In the literature, the topic of energy harvesting and cognitive radio receive increasing attention. Three groups of existing works are most related. First, the CR technique has received significant attention during the past few years [18–22]. In [18], the authors focus on designing a database access strategy that allows the SUs to jointly consider the requirements of the existing rules, as well as the maximization of the expected communication opportunities through on-demand database access. The optimal strategy introduced in [18], which is computationally unfeasible with the brute-force approach, can be solved by the efficient algorithm proposed in [19]. In [19], by proving that the optimal strategy has a threshold structure, an efficient algorithm is introduced by exploiting the threshold property. In [20], the authors investigate the achievable throughput of an unlicensed sensor network operating over the TV white space spectrum. The achievable throughput is analytically derived as a function of the channel ordering. Additionally, the closed-form expression of the maximum expected throughput is illustrated. The work in [21] studies the problem of coexistence interference among multiple secondary networks without the secondary cooperation. Under a reasonable assumption, a computationally-efficient algorithm for finding the optimal strategy is presented. The work in [22] develops robust power control strategies for cognitive radios in the case of sensing delay and model parameter uncertainty. A robust power control framework that optimizes the worst-case system performance is proposed. All of the problems considered in the above works are formulated as Markov decision process problems, and the technical contributions are very important and valuable. However, due to the unique features of the EH CRSNs, such as the dynamic energy replenishment process, which stipulates a new design constraint on energy usage in the time axis, there is a need to revisit resource allocation policies so that the energy expenditure can efficiently adapt to the dynamics of energy arrivals.

Second, the energy harvesting technique has been widely studied in wireless communication systems [23–30]. The works in [23–26] consider the point-to-point wireless communications. In [23], through optimizing the time sequence of transmit powers, the authors focus on maximizing the throughput by a deadline and minimizing the transmission completion time. For the offline policy, a directional water-filling algorithm is introduced to find the optimal power allocation. For the online policy, dynamic programming is applied to solve the optimal power allocation. In [24],

the authors consider the problem of energy allocation over a finite horizon to maximize the throughput. A water-filling energy allocation where the water-level follows a staircase function is introduced. The work in [25] studies the problem of energy allocation for sensing and transmission to maximize the throughput in an energy harvesting wireless sensor network. The problem studied in [25] considers the finite horizon case, which is extended in [26] to an infinite-horizon case. In [26], the authors study the energy allocation for sensing and transmission for an energy harvesting sensor node. An optimal energy allocation algorithm and an optimal transmission energy allocation algorithm are introduced. The works in [27,28] consider the problem of hybrid energy supply. In [27], the authors investigate the minimization of the power consumption stemming from the constant energy source for transmitting a given number of data packets. In [28], for a hybrid energy supply system employing a save-then-transmit protocol, the authors explore the transmission scheduling problem. In [29], the authors study the transmission power allocation strategy to achieve the energy-efficient transmission. The harvest-use technique is adopted, which means that the harvested energy cannot be stored and must be used immediately. In [30], for a solar-powered wireless sensor network, the authors present an optimal transmission policy based on a data-driven approach. However, due to the distinctive operation of cognitive radio, such as spectrum sensing, spectrum management, etc., directly applying the strategies mentioned above to EH CRSNs can be ineffective or inefficient.

Third, much recent research has been tightly focused on CR systems powered by energy harvesting. The work in [11] focuses on an energy harvesting cognitive radio network with the save-then-transmit protocol; the authors mainly investigate the joint optimization of saving factor, sensing duration, sensing threshold and fusion rules to maximize the achievable throughput. In [31], for a single-user multi-channel setting, jointly considering probabilistic arrival energy, channel conditions and the probability of PU's occupation, the authors propose a channel selection criterion. In [32], jointly considering the battery replenish process and the secondary belief regarding the primary activities, the authors introduce an energy allocation for sensing and transmission to maximize the long-term throughput. Different from [32], a suboptimal energy allocation algorithm that allocates energy in an online approach is introduced in [33]. In [34], in order to maximize the throughput, the authors derive an optimal sensing strategy through optimizing the access probabilities of idle channels and busy channels. In [35], a joint design of the spectrum sensing and detection threshold to maximize the long-term throughput is studied. Furthermore, the upper bound of the achievable throughput is derived as a function of the energy arrival rate, the statistical behavior of the primary network traffic and the detection threshold in [36]. In [10], the authors propose a spectrum and energy-efficient heterogeneous cognitive radio sensor network (HCRSNs), where EH-enabled spectrum sensors cooperatively detect the status of the licensed channels, while the data sensors transmit data to the sink. Compared with these works, the salient feature of this paper is that, according to the current knowledge of the battery state, channel fading, as well as the arrival energy based on EH, we jointly optimize the action of channel sensing and transmission power allocation for an energy harvesting cognitive sensor node, to minimize the long-term outage probability.

3. Network Model

3.1. Primary Network Model

We consider a primary network where a primary user (PU) owns the usage right of a channel with bandwidth B. The PU is assumed to employ synchronous slotted communication with a time slot duration T. The primary traffic is modeled as a two-state time-homogeneous random process, in which the channel randomly switches its state between idle and occupied, as assumed in [34]. Let θ_t represent the status of the channel in time slot t, with $\theta_t = 0$ or 1 indicating that the channel is occupied or idle, respectively. The probability that the channel is occupied by the PU is denoted as $p_o \triangleq Pr(\theta_t = 0)$. Correspondingly, the idle probability of the channel is defined as $p_i \triangleq Pr(\theta_t = 1) = 1 - p_o$. It is

assumed that p_o and p_i are available for the secondary users based on the long-term spectrum measurements [35].

3.2. Secondary Network Model

3.2.1. Energy Model and Opportunistic Spectrum Access

We consider a point-to-point communication link between two secondary sensor nodes, which are also referred to as secondary users (SUs). An EH-enabled SU opportunistically accesses the primary channel to convey data to its receiver. The EH SU is powered exclusively by energy harvested from the ambient environment (e.g., solar, wind, thermal, ambient radio power) and stores the energy in a rechargeable battery with finite energy storage capacity. A correlated time process following a first-order discrete-time Markovian model is adopted for modeling the energy arrivals [26,37]. According to the harvest-store-use model, the harvested energy in the current time slot can only be used in the next time slot.

Since the PU has priority in utilizing the channel, in order to opportunistically use the channel, the SU has to perform real-time monitoring of the channel to avoid collisions with the PU. Thus, for each time slot, the overall transmission process consists of two phases, namely the channel acquisition phase and the transmission phase. The allocation of time durations for the two phases is illustrated in Figure 1, where the channel acquisition phase and the transmission phase consume α_t and $1-\alpha_t$ fractions of one time slot, respectively, and α_t is referred to as spectrum sensing overhead for time slot t, which can be altered to optimize the performance of the system. For the channel acquisition phase, the SU senses the status of the spectrum with $\alpha_t T$ time through the energy detection technique [38]. As the complexity is roughly linear in sensing duration, we can assume that the energy consumption e_s for sensing is proportional to $\alpha_t T$ with a constant sensing power p_s [32], namely:

$$e_s(\alpha_t) = \alpha_t T p_s. \tag{1}$$

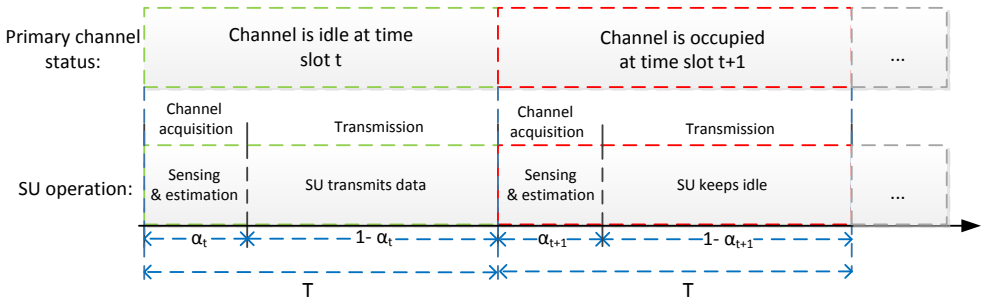

Figure 1. The allocation of time durations: channel acquisition phase versus transmission phase.

If the channel is sensed to be idle, the SU starts data transmission using energy stemming from the battery. Let P_t be the transmit power of SU, then the energy consumption for the transmission phase can be expressed as:

$$e_d^1(\alpha_t, P_t) = (1-\alpha_t) T P_t. \tag{2}$$

If the channel is sensed to be occupied, the SU stays in the idle state with a constant idle power p_c, which is considerably less than the transmit power [39]; therefore, the energy consumption for the data transmission phase is:

$$e_d^0(\alpha_t) = (1-\alpha_t) T p_c. \tag{3}$$

3.2.2. Spectrum Sensing and Transmission Data Rate

During the channel acquisition phase mentioned above, the SU acquires the status of the channel by performing a binary hypothesis test to determine between idle $\mathcal{H}_0$ (i.e., $\theta_t = 1$) and occupied $\mathcal{H}_1$

(i.e., $\theta_t = 0$). Due to the existence of sensing errors, the reliability of spectrum sensing is evaluated by two indicators, namely the false alarm probability P_f and the detection probability P_d, which are defined as follows:

$$P_f = Pr\{\hat{\theta}_t = 0 \mid \theta_t = 1\}, \tag{4}$$

$$P_d = Pr\{\hat{\theta}_t = 0 \mid \theta_t = 0\}, \tag{5}$$

where $\hat{\theta}_t$ is the binary decision on the primary channel, with $\hat{\theta}_t = 0$ or 1 representing that the primary channel is determined to be occupied or idle, respectively. Considering ensuring sufficient protection to the PU, the SU should satisfy a target detection probability $\bar{P}_d$ on the primary channel. Regarding the complex-valued primary signal and circularly symmetric complex Gaussian (CSCG) noise case, the probability of a false alarm is given by [40]:

$$P_f(\alpha_t) = Q(\sqrt{2\beta + 1}Q^{-1}(\bar{P}_d) + \sqrt{\alpha_t T f_s}\beta), \tag{6}$$

where β is the received signal-to-noise ratio (SNR) of the primary signal at SU and f_s is the sampling frequency. The function $Q(\cdot)$ is $Q(x) = (1/\sqrt{2\pi})\int_x^\infty \exp(-t^2/2)dt$.

For time slot t, after acquiring the status of the channel, the SU performs channel estimation to obtain the channel condition. Specifically, the SU will send pilot signals to the receiver and acquires the channel power gain, denoted as γ_t, through an error-free and dedicated feedback channel [31]. Since the above channel estimation takes a very short time and limited power, we assume the time and energy consumed in the channel estimation are negligible compared to the sensing time, and hence, we ignore it in our analysis (For example, if PUs are TV bands where each channel occupies 6 MHz in the case of the IEEE802.22 wireless regional area network (WRAN), the typical sensing time is about a few milliseconds, which will result in thousands of samples [40]. However, for channel estimation, a few pilot symbols would be enough. For example, in IEEE802.11*a*, only four pilot symbols are used for channel estimation [41]), similar to [42,43]. Then, the transmission data rate of the SU is:

$$r(\alpha_t, P_t, \gamma_t) = (1 - \alpha_t)\log(1 + \frac{P_t\gamma_t}{N_0}), \tag{7}$$

where N_0 is the destination noise power. The coefficient $1 - \alpha_t$ is due to the fact that only a $1 - \alpha_t$ fraction of a time slot is utilized for the SU's data transmission phase. If, on the other hand, the channel is sensed to be occupied, the sensor node abstains from transmission and stays idle for the rest of the time slot; thus, the transmission data rate $r(\alpha_t, P_t, \gamma_t)$ is zero.

The overall objective of this paper is to design optimal policies by jointly considering the sensing overhead and the transmit power allocation, to minimize the long-term outage probability of the EH-enabled cognitive sensor nodes. In the following section, we will exhibit the procedure of formulating the problem of outage probability minimization as an Markov decision process in detail.

4. Problem Formulation

In this section, we formulate the problem of long-term outage probability minimization as an MDP. The MDP model is mainly composed of decision epochs, states, actions, state transition probabilities and rewards. The decision epoch is time slot $t \in \mathcal{T} = \{0, 1, 2, \cdots\}$. The state of the system is denoted as $s = (b, g, h)$, where b indicates the battery energy state, g indicates the channel state and h indicates the state of arrival energy based on EH. We assume that b, g and h take discrete values from discrete finite set $\mathcal{B} = \{0, 1, 2, \cdots, N_B - 1\}$, $\mathcal{G} = \{0, 1, 2, \cdots N_G - 1\}$ and $\mathcal{H} = \{0, 1, 2, \cdots N_H - 1\}$, respectively. Thus, the state space can be expressed as $\mathcal{S} = \mathcal{B} \times \mathcal{G} \times \mathcal{H}$, where $\times$ denotes the Cartesian product. We assume the battery is quantized in units of e_u, which can be referred to as one unit of energy quantum . Additionally, we denote the battery energy State 0 corresponds to the energy $B_0 \triangleq \lceil \frac{p_c T}{e_u} \rceil e_u$, which is the energy consumption when the SU stays in the idle state within the entire time slot, and for battery state $b \in \mathcal{B}\backslash\{0\}$, the total energy in the battery is $B_0 + be_u$. As for the arrival energy, if the

arrival energy state is $h \in \mathcal{H}$, then the actually arrival energy is $Q_h e_u$, where $Q_h \in N$. It should be noted that as the channel state and arrival energy state can only be acquired casually, at the beginning of time slot t, the SU only attains the exact channel state and the arrival energy state of the previous time slot. Therefore, the system state for time slot t can be represented as $s_t = (b_t, g_{t-1}, h_{t-1})$, where $b_t \in \mathcal{B}$ is the energy state for the current time slot, whereas $g_{t-1} \in \mathcal{G}$ and $h_{t-1} \in \mathcal{H}$ are the states of channel and arrival energy of the previous time slot. The evolvement of the arrival energy h_t is assumed to be a first-order discrete-time Markovian model introduced in Section 3.2; hence, in the following, we will first introduce the update process of the battery energy state b_t along with the SU's channel capacity. Then, the evolvement of the channel state g_t is presented.

First, as to the battery energy state update process, a combination of sensing overhead α_t and transmit power P_t leads to one of the following four possible consequences:

1. Idle detection with probability $p_i(1 - P_f(\alpha_t))$: the primary channel is idle while the sensing result is correct. Then, channel capacity:
$$R = r(\alpha_t, P_t, \gamma_t) \tag{8}$$
is gained, and the battery energy state updates as:
$$b_{t+1} = \min\{\lfloor b_t - e_s(\alpha_t)/e_u - e_d^1(\alpha_t, P_t)/e_u + Q_{h_t} \rfloor, N_B - 1\}. \tag{9}$$
2. False alarm with probability $p_i P_f(\alpha_t)$: the primary channel is idle while the sensing result is wrong. The SU abstains from the transmission, and the channel capacity R is zero. The battery energy state is:
$$b_{t+1} = \min\left\{\lfloor b_t - e_s(\alpha_t)/e_u - e_d^0(\alpha_t)/e_u + Q_{h_t} \rfloor, N_B - 1\right\}. \tag{10}$$
3. Occupied detection with probability $p_o \bar{P}_d$: the primary channel is occupied while the sensing result is correct. SU abstains from the transmission, and channel capacity R is zero; the battery energy state is the same as (10).
4. Misdetection with probability $p_o(1 - \bar{P}_d)$: the primary channel is occupied while the sensing result is wrong. Channel capacity R is zero due to the collision with PU and the battery energy state updates the same as (9).

Second, we formulate the evolvement of channel states. The channel fading process can be modeled as a time-homogeneous finite-state Markov chain (FSMC), which has been widely used to model the block fading channel [44–47]. Specifically, the channel power is quantized using a finite number of thresholds $\mathbb{G} = \{G_0 = 0, G_1, G_2, \cdots, G_{N_G} = \infty\}$, where $G_i < G_j$ when $0 \le i < j \le N_G - 1$. The channel is considered to be in state i, $0 \le i \le N_G - 1$, if the instantaneous channel power gain belongs to the interval $[G_i, G_{i+1})$. We consider that the wireless channel fluctuates slowly over time slots and remains constant within a time slot, as assumed in [48,49]. Hence, the channel state transition occurs only from the current state to its neighboring states at the beginning of each time slot [30]. Considering the Rayleigh fading channel, the channel state transition probability is determined by [50]:

$$P(g_{t+1} = j | g_t = i) = \begin{cases} \dfrac{h(G_{i+1})}{P(g = i)}, & j = i + 1, i = 0, ..., N_G - 2; \\ \dfrac{h(G_i)}{P(g = i)}, & j = i - 1, i = 1, ..., N_G - 1; \\ 1 - \dfrac{h(G_i)}{P(g = i)} - \dfrac{h(G_{i+1})}{P(g = i)}, & j = i, i = 1, ..., N_G - 2, \end{cases} \tag{11}$$

where $P(g = i)$ is the stationary probability that the channel state is i, and $P(g = i) = \exp(-\frac{G_i}{G_a}) - \exp(-\frac{G_{i+1}}{G_a})$; G_a is the average channel power gain. $h(\beta) = \sqrt{2\pi\beta/G_a} f_D \exp(-\beta/G_a)$ is the level

crossing rate, where f_D is the maximum Doppler frequency, normalized by $1/T$. The boundary transition probabilities for channel states are:

$$P(g_{t+1} = 0|g_t = 0) = 1 - P(g_{t+1} = 1|g_t = 0), \tag{12}$$

$$P(g_{t+1}{=}N_G{-}1|g_t{=}N_G{-}1){=}1{-}P(g_{t+1}{=}N_G{-}2|g_t{=}N_G{-}1). \tag{13}$$

According to the current system state $s_t = (b_t, g_{t-1}, h_{t-1})$, we introduce the action set of the SU. The sensing overhead α_t is quantized in units of $\alpha_u = \frac{e_u}{p_s T}$, and the the action set of sensing overhead can be expressed as follows:

$$A_\alpha^{s_t} = \begin{cases} \{0\} \text{ if } b_t = 0, \\ \{1, 2, \cdots, \min\{\left\lfloor \dfrac{p_s T}{e_u} \right\rfloor, b_t\}\} \text{ otherwise,} \end{cases} \tag{14}$$

where $\lfloor \cdot \rfloor$ is the floor function. $b_t = 0$ indicates that the energy level in the battery is so low (the energy stored in the battery is B_0) that the available energy is merely enough to compensate the energy expenditure when the SU stays in the idle state within the entire time slot. In this case, the SU stops the sensing, as well as transmission and keeps on harvesting energy. Respecting the constraint $\min\{\left\lfloor \frac{p_s T}{e_u} \right\rfloor, b_t\}$, the first constraint indicates that the sensing duration should be less than the time slot T; the second constraint indicates that the energy consumption for sensing should be less than the available energy $b_t e_u$. When an action $a_\alpha \in A_\alpha^{s_t}$ is taken, the sensing overhead is $a_\alpha \cdot \alpha_u$, the sensing time is $a_\alpha \cdot \alpha_u \cdot T = a_\alpha \cdot \frac{e_u}{P_s}$ and the energy consumption for sensing is $e_s(a_\alpha \cdot \alpha_u) = a_\alpha \cdot \alpha_u \cdot T \cdot P_s = a_\alpha e_u$. According to the action of sensing overhead, the action set of transmission power is quantized in units of $P_u = \frac{e_u}{(T - a_\alpha \alpha_u T)}$, and the action set can be expressed as:

$$A_p^{(s_t, a_\alpha)} = \{0, 1, 2, \cdots, b_t - a_\alpha\}. \tag{15}$$

For an action $a_p \in A_p^{(s_t, a_\alpha)}$, SU will consume $a_p e_u$ energy for data transmission.

Therefore, given a system state $s_t = (b_t, g_{t-1}, h_{t-1})$, the action set can be represented as:

$$\mathcal{A}_{s_t} = \left\{ (a_\alpha, a_p) | a_\alpha \in A_\alpha^{s_t}, a_p \in A_p^{(s_t, a_\alpha)} \right\}. \tag{16}$$

We use $P(s_{t+1}|s_t, a)$ to denote the system state transition probability, which indicates the probability that the system will go into state $s_{t+1} = (b_{t+1} = b', g_t = g', h_t = h')$ in the case that the current system state is $s_t = (b_t = b, g_t = g, h_t = h)$ and SU takes an action $a = (a_\alpha, a_p) \in \mathcal{A}_{s_t}$. The state transition probability can be derived as follows:

$$\begin{aligned} P(s_{t+1}|s_t, a) &= P(b', g', h'|b, g, h, a_\alpha, a_P) \\ &= P(b'|b, g', h', a_\alpha, a_P) P(g'|g) P(h'|h) \end{aligned} \tag{17}$$

where:

$$P(b'|b, g', h', a_\alpha, a_P) = \begin{cases} 1 \text{ if } b' = \min\{\lfloor b - a_\alpha - I_{a_P > 0} a_P - I_{a_P = 0} \dfrac{e_d^0(a_\alpha \alpha_u)}{e_u} + Q_{h'} \rfloor, N_B - 1\}, \\ 0 \text{ otherwise,} \end{cases} \tag{18}$$

since b' is a certain value, which is determined by b, h' and action a_α, a_p. I_x denotes the indicator function which takes the value of one if x is true, otherwise zero.

The reward function is defined as the outage probability regarding the system state $s_t = (b_t, g_{t-1}, h_{t-1})$ and the corresponding action $a = (a_\alpha, a_p)$, which is given by [51]:

$$\begin{aligned} R(s_t, a) &\triangleq P_{out}(R < R_{th}) \\ &= p_i(1 - P_f(a_\alpha \alpha_u))Pr(r(a_\alpha \alpha_u, a_p P_u, \gamma_t) < R_{th}) + p_i P_f(a_\alpha \alpha_u) + p_o \bar{P}_d + p_o(1 - \bar{P}_d) \\ &= p_i(1 - P_f(a_\alpha \alpha_u)) \sum_{g_t \in \mathcal{G}} P(g_t | g_{t-1}) Pr(\gamma_t < \gamma_{th} | G_{g_t} \le \gamma_t < G_{g_t+1}) \\ &+ p_i P_f(a_\alpha \alpha_u) + (1 - p_i), \end{aligned} \tag{19}$$

where $\gamma_{th} = \frac{N_0}{a_p P_u}(2^{\frac{R_{th}}{1-a_\alpha \alpha_u}} - 1)$. If $\gamma_{th} \ge G_{g_t+1}$, then $Pr(\gamma_t < \gamma_{th} | G_{g_t} \le \gamma_t < G_{g_t+1}) = 1$; if $\gamma_{th} < G_{g_t}$, then $Pr(\gamma_t < \gamma_{th} | G_{g_t} \le \gamma_t < G_{g_t+1}) = 0$; otherwise, $Pr(\gamma_t < \gamma_{th} | G_{g_t} \le \gamma_t < G_{g_t+1}) = \frac{Pr\{G_{g_t} \le \gamma_t < \gamma_{th}\}}{Pr\{G_{g_t} \le \gamma_t < G_{g_t+1}\}} = \frac{\exp(-G_{g_t}/G_a) - \exp(-\gamma_{th}/G_a)}{\exp(-G_{g_t}/G_a) - \exp(-G_{g_t+1}/G_a)}$.

In the following section, we first mainly study the existence of the optimal transmission policy. Then, the ϵ-optimal sensing-transmission policy that specifies the actions concerning the sensing overhead and the transmit power to minimize the long-term outage probability is introduced. Last, for a special case where the signal-to-noise power ratio is sufficiently high, we introduce an efficient transmission policy, which achieves the same performance as the ϵ-optimal sensing-transmission policy.

5. Proposed Transmission Policies

In this section, we focus on deriving policies that specify the actions regarding the sensing overhead and transmit power, with the goal of minimizing the long-term outage probability. First, we introduce the concept of the stationary deterministic policy. Second, we prove the convergence and the existence of the stationary deterministic policy. Then, based on the Bellman equation, we propose an ϵ-optimal stationary deterministic policy named the sensing-transmission policy through the value iteration approach. Last, for the special case where the signal-to-noise (SNR) is sufficiently high, we introduce an efficient transmission policy.

Denote $\pi(s) = \{d_0(s_0), d_1(s_1), d_2(s_2), \cdots\}$ as the decision policy that specifies the decision rules to be used at each time slot, and d_t is the decision rule that prescribes a procedure for action selection in time slot t. A policy is stationary deterministic if d_t is deterministic Markovian and $d_t = d$ for all $t \in \mathcal{T}$ [26]; therefore, the stationary deterministic policy can be represented as $\pi(s) = \{d(s_0), d(s_1), d(s_2), \cdots\}$. For an infinite-horizon MDP, our primary focus will be on the stationary deterministic policy because the decision rules do not change over time, and they are easiest to implement and evaluate [52]. We denote the feasible set of stationary deterministic policies as Π^{SD}. Given the initial state $s_0 = (b_0, g_{-1}, h_{-1})$ and the policy $\pi \in \Pi^{SD}$, the expected discounted infinite-horizon reward that represents the long-term outage probability is defined to be [52]:

$$V_\pi(s_0) = \mathbb{E}\Big\{ \sum_{t=0}^{\infty} \lambda^t R(s_t, a) | s_0, \pi \Big\}, s_t \in \mathcal{S}, a \in \mathcal{A}_{s_t}, \tag{20}$$

where $V_\pi(s_0)$ is the long-term expected reward with respect to the initial state s_0, $0 \le \lambda < 1$ is the discount factor, $R(s_t, a)$ is the reward function defined by (19) and a is the action determined by the policy π. The alteration of λ brings a wide range of performance characteristics, which can be altered according to the actual needs.

The objective of the SU is to find the optimal stationary deterministic policy π^* that minimize the long-term expected reward defined in (20), that is:

$$\pi^* = \min_{\pi \in \Pi^{SD}} V_\pi(s_0). \tag{21}$$

First, we prove that the long-term expected reward $V_\pi(s_0)$, where $\pi \in \Pi^{SD}$, is finite.

Lemma 1. *$V_\pi(s_0)$ is finite, namely $|V_\pi(s_0)| < \infty$, where $\pi \in \Pi^{SD}$ and $s_0 \in \mathcal{S}$.*

Proof of Lemma 1. In order to prove that the value of $|V_\pi(s_0)|$ is limited, according to [52], we only need to prove that $\sup_{a\in\mathcal{A}_s, s\in S} |R(s,a)| < \infty$. As $Pr(\gamma_t < \gamma_{th} | G_{g_t} \leq \gamma_t < G_{g_t+1}) \leq 1$, $P(g_t|g_{t-1}) \leq 1$ and $\mathcal{G}$ is discrete and finite, we can deduce that $\sum_{g_t\in\mathcal{G}} P(g_t|g_{t-1}) Pr(\gamma_t < \gamma_{th} | G_{g_t} \leq \gamma_t < G_{g_t+1})$ is finite. Since $p_i \leq 1$, $P_f \leq 1$, it can be derived that $|R(s,a)|$ is limited. Thus, we can conclude that $\sup_{a\in\mathcal{A}_s, s\in S} |R(s,a)| < \infty$, and therefore, $V_\pi(s_0)$ is finite. □

Lemma 1 indicates that for any initial system state, the value of $V_\pi(s_0)$ converges to a certain value. Next, we explain the existence of the optimal stationary deterministic policy π^*.

Theorem 1. *There exists an optimal stationary deterministic policy π^* to minimize the long-term expected reward displayed in Equation (20).*

Proof of Theorem 1. Since the system state $\mathcal{S} = \mathcal{B} \times \mathcal{G} \times \mathcal{H}$ is discrete and finite and for an arbitrary $s \in \mathcal{S}$, the corresponding action space $\mathcal{A}_s$ is also discrete and finite, thus there exists an optimal stationary deterministic policy [52]. □

Given an arbitrary system system s, the optimal long-term expected reward $V_{\pi^*}(s)$ should satisfy the following Bellman optimality equation:

$$V_{\pi^*}(s) = \min_{a\in\mathcal{A}_s} \left\{ R(s,a) + \lambda \sum_{s'\in\mathcal{S}} P(s'|s,a) V_{\pi^*}(s') \right\}, s \in \mathcal{S}. \tag{22}$$

The first term on the right-hand side of Equation (22) is the immediate reward for the current time slot, and the second term is the expected total discount future reward if SU chooses action a. The well-known value iteration approach is then applied to find the ϵ-optimal stationary deterministic policy, as shown in Algorithm 1.

Algorithm 1 Sensing-transmission (ST) policy.

1: Set $V_0(s) = 0$ for all $s \in \mathcal{S}$, set $i = 0$, specify $\epsilon > 0$.
2: For each $s \in \mathcal{S}$, calculate the $V_{i+1}(s)$ according to
$V^a_{i+1}(s) = \left\{ R(s,a) + \lambda \sum_{s'\in\mathcal{S}} P(s'|s,a) V_i(s') \right\}, a \in \mathcal{A}_s,$
$V_{i+1}(s) = \min_{a\in\mathcal{A}_s} \left\{ V^a_{i+1}(s) \right\}.$
3: If $\|V_{i+1} - V_i\| < \epsilon(1-\lambda)/2\lambda$, go to Step 4. Otherwise, increase i by 1 and go back to Step 2.
4: For each $s \in \mathcal{S}$, choose $d(s) = \arg\min_{a\in\mathcal{A}_s} \left\{ R(s,a) + \lambda \sum_{s'\in\mathcal{S}} P(s'|s,a) V_{i+1}(s') \right\}$
5: Obtain the ϵ-optimal transmission policy $\pi^*_\epsilon = \{d, d, \cdots\}$

In Algorithm 1, the SU iteratively finds the optimal policy. Specifically, in Step 1, $V_0(s)$ is initialized to zero for all $s \in \mathcal{S}$; the error bound ϵ is specified; and set the iteration sequence i to be zero. In Step 2, we compute the $V_{i+1}(s)$ for each $s \in \mathcal{S}$ according to the knowledge of $V_i(s)$. Then, in Step 3, the SU first estimates whether $\|V_{i+1} - V_i\| < \epsilon(1-\lambda)/2\lambda$ holds, where $V_{i+1} = \{V_{i+1}(s), \forall s \in \mathcal{S}\}$, $V_i = \{V_i(s), \forall s \in \mathcal{S}\}$ and $\|V_{i+1} - V_i\| = \max_{s\in\mathcal{S}} |V_{i+1}(s) - V_i(s)|$. If the inequality holds, which means that the value iteration algorithm has converged, then we proceed to Step 4 to obtain the decision rule and then formulate the sensing-transmission policy. Otherwise, we need to go back to Step 2 and continue to perform the iteration. According to Algorithm 1, the SU can pre-compute the policy and records it in a look-up table. Then, based on the specific system state, the SU can check the look-up table to find out the corresponding action.

As to the convergence, $V_i(s)$ computed by Step 2 converges to $V_{\pi^*}(s)$ for all $s \in \mathcal{S}$. Once the inequality condition in Step 3 is satisfied, then the obtained optimal policy ensures that $\left\| V_{\pi_\epsilon^*} - V_{\pi^*} \right\| < \epsilon$, where $V_{\pi_\epsilon^*} = \{V_{\pi_\epsilon^*}(s), \forall s \in \mathcal{S}\}$ is the long-term expected reward achieved by the ϵ-optimal policy obtained in Step 5 of the Algorithm 1. In practice, according to the actual needs, SU can predefine the value of ϵ to control the accuracy of convergence. Choosing ϵ small enough ensures that the algorithm stops with a policy that is very close to optimal. Next, we introduce the complexity of Algorithm 1. The complexity of each iteration in the value iteration algorithm is $O(N_{state}N'_{state}N_{action})$ [53], where N_{state} represents the total number of states in the state space, N'_{state} indicates the total number of states that the system can possibly transmit to and N_{action} represents the total number of actions in the action space. For our MDP problem, the total number of states in state space $\mathcal{S}$ is $N_B \cdot N_G \cdot N_H$. As the battery state of the next time slot is deterministic and the channel can only transmit to the neighbor state or remains in its current state, therefore the total possible states the current system state can transmit to is $3N_H$. The maximum number of actions regarding the sensing overhead, as well as the transmit power is $(N_B + 1)N_B/2$. Hence, the complexity of each iteration in Algorithm 1 is $O(N_B^3 N_H^2 N_G)$.

Next, we study the structural property of the proposed sensing-transmission policy. Regarding the reward function, we have the following lemma:

Lemma 2. *Given a system s, for an arbitrary certain action of a_α, the immediate reward $R(s, a_\alpha, a_p)$ is non-increasing with a_p, namely $R(s, a_\alpha, a_p + 1) \leq R(s, a_\alpha, a_p)$, where $a_\alpha \in A_\alpha^s$, a_p and $a_p + 1 \in A_p^{(s,a_\alpha)}$.*

Proof of Lemma 2. First, we prove that for a certain action of a_α, $Pr(\gamma_t < \gamma_{th}(a_p) | G_{g'} \leq \gamma_t < G_{g'+1})$ defined in Equation (19) is non-increasing with transmit action a_p, namely $Pr(\gamma_t < \gamma_{th}(a_p + 1) | G_{g'} \leq \gamma_t < G_{g'+1}) \leq Pr(\gamma_t < \gamma_{th}(a_p) | G_{g'} \leq \gamma_t < G_{g'+1})$, where $g' \in \mathcal{G}$. As $\gamma_{th}(a_p)$ is decreasing with a_p, we have $\gamma_{th}(a_p + 1) < \gamma_{th}(a_p)$. If $\gamma_{th}(a_p) > \gamma_{th}(a_p + 1) \geq G_{g'+1}$, we can derive that $Pr(\gamma_t < \gamma_{th}(a_p + 1) | G_{g'} \leq \gamma_t < G_{g'+1}) = Pr(\gamma_t < \gamma_{th}(a_p) | G_{g'} \leq \gamma_t < G_{g'+1}) = 1$. If $G_{g'} > \gamma_{th}(a_p) > \gamma_{th}(a_p + 1)$, we can derive that $Pr(\gamma_t < \gamma_{th}(a_p + 1) | G_{g'} \leq \gamma_t < G_{g'+1}) \leq Pr(\gamma_t < \gamma_{th}(a_p) | G_{g'} \leq \gamma_t < G_{g'+1}) = 0$. Otherwise, it can be derived that $Pr(\gamma_t < \gamma_{th}(a_p + 1) | G_{g'} \leq \gamma_t < G_{g'+1}) < Pr(\gamma_t < \gamma_{th}(a_p) | G_{g'} \leq \gamma_t < G_{g'+1})$. Therefore, we can conclude that $Pr(\gamma_t < \gamma_{th}(a_p) | G_{g'} \leq \gamma_t < G_{g'+1})$ is non-increasing with transmit action a_p.

Next, we calculate the difference between $R(s, a_\alpha, a_p)$ and $R(s, a_\alpha, a_p + 1)$:

$$\begin{aligned} & R(s, a_\alpha, a_p) - R(s, a_\alpha, a_p + 1) \\ & = p_i(1 - P_f(a_\alpha \alpha_u)) \sum_{g' \in \mathcal{G}} P(g'|g) \Big[Pr(\gamma_t < \gamma_{th}(a_p) | G_{g'} \leq \gamma_t < G_{g'+1}) \\ & - Pr(\gamma_t < \gamma_{th}(a_p + 1) | G_{g'} \leq \gamma_t < G_{g'+1}) \Big], \end{aligned} \tag{23}$$

since $Pr(\gamma_t < \gamma_{th}(a_p) | G_{g'} \leq \gamma_t < G_{g'+1})$ is non-increasing with a_p, we can derive that $R(s, a_\alpha, a_p) - R(s, a_\alpha, a_p + 1) \geq 0$, that is $R(s, a_\alpha, a_p + 1) \leq R(s, a_\alpha, a_p)$. □

Lemma 3. *For any given channel state $g \in \mathcal{G}$ and arrival energy state $h \in \mathcal{H}$, the minimum immediate reward $R(s, a)$ is non-increasing in battery state $b \in \mathcal{B}$. That is, $\min_{a^+ \in \mathcal{A}_{s^+}} \{R(s^+, a^+)\} \leq \min_{a \in \mathcal{A}_s} \{R(s, a)\}$, where $s^+ = \{b + 1, g, h\}$, $s = \{b, g, h\}$, $\forall b \in \mathcal{B} \backslash \{N_B - 1\}$, $g \in \mathcal{G}, h \in \mathcal{H}$.*

Proof of Lemma 3. The action set for s^+ can be expressed as $\mathcal{A}_{s^+} = \left\{ (a_\alpha^+, a_p^+) | a_\alpha^+ \in A_\alpha^{s^+}, a_p^+ \in A_p^{(s^+, a_\alpha^+)} \right\}$, and the action set for s can be expressed as $\mathcal{A}_s = \left\{ (a_\alpha, a_p) | a_\alpha \in A_\alpha^s, a_p \in A_p^{(s,a_\alpha)} \right\}$. When $a_\alpha^+ = a_\alpha = w$, we can derive that the unit of transmit power $\frac{e_u}{T - a_\alpha^+ \alpha_u T} = \frac{e_u}{T - a_\alpha \alpha_u T}$, and $A_p^{(s^+, a_\alpha^+)} =$

$\{0,1,2,\cdots,\max\{b+1-w,0\}\} \supseteq A_p^{(s,a_\alpha)} = \{0,1,2,\cdots,\max\{b-w,0\}\}$; according to Lemma 2, we have $\min_{a_p^+ \in A_p^{(s^+,a_\alpha^+)}} R(s^+,w,a_p^+) = R(s^+,w,b+1-w)$, and $\min_{a_p \in A_p^{(s,a_\alpha)}} R(s,w,a_p) = R(s,w,b-w)$. Since:

$$\begin{aligned} & R(s^+,w,b+1-w) - R(s,w,b-w) \\ & = p_i(1-P_f(a_\alpha\alpha_u)) \sum_{g' \in \mathcal{G}} P(g'|g) \Big[Pr(\gamma_t < \gamma_{th}(b+1-w) | G_{g'} \le \gamma_t < G_{g'+1}) \\ & - Pr(\gamma_t < \gamma_{th}(b-w) | G_{g'} \le \gamma_t < G_{g'+1}) \Big] \\ & \le 0, \end{aligned} \tag{24}$$

therefore, we have $\min_{a_p^+ \in A_p^{(s^+,a_\alpha^+)}} R(s^+,w,a_p^+) \le \min_{a_p \in A_p^{(s,a_\alpha)}} R(s,w,a_p)$.

As $\min\{\left\lfloor \frac{P_s T}{e_u} \right\rfloor, b+1\} \ge \min\{\left\lfloor \frac{P_s T}{e_u} \right\rfloor, b\}$, thus $A_\alpha^{s^+} \supseteq A_\alpha^s$; therefore, we have:

$$\min_{a_\alpha^+ \in A_\alpha^{s^+}} \min_{a_p^+ \in A_p^{(s^+,a_\alpha^+)}} R(s^+,a_\alpha^+,a_p^+) \le \min_{a_\alpha \in A_\alpha^s} \min_{a_p \in A_p^{(s,a_\alpha)}} R(s,a_\alpha,a_p), \tag{25}$$

namely $\min_{a^+ \in \mathcal{A}_{s^+}} \{R(s^+,a^+)\} \le \min_{a \in \mathcal{A}_s} \{R(s,a)\}$. □

Based on Lemma 3, we have following lemma:

Lemma 4. *For any given channel state $g \in \mathcal{G}$ and arrival energy state $h \in \mathcal{H}$, we have that $V_i(b,g,h)$ is non-increasing in the battery state $b \in \mathcal{B}$, that is $V_i(b+1,g,h) \le V_i(b,g,h)$ $\forall b \in \mathcal{B}\backslash\{N_B-1\}$.*

Proof of Lemma 4. We prove this lemma by the induction. When $i=1$, as the initial condition $V_0(s)=0$ for all $s \in \mathcal{S}$, thus $V_1(s) = \min_{a \in \mathcal{A}_s} \Big\{ R(s,a) \Big\}$. According to Lemma 3, we have $V_1(b+1,g,h) \le V_1(b,g,h)$. Assume when $i=k$, for any given $g \in \mathcal{G}$, $h \in \mathcal{H}$ and $\forall b \in \mathcal{B}\backslash\{N_B-1\}$, $V_k(b+1,g,h) \le V_k(b,g,h)$ holds. When $i=k+1$, we use s^+ to indicate system state $(b+1,g,h)$ and use s to indicate system state (b,g,h). The action sets for s^+ and s are $\mathcal{A}_{s^+} = \Big\{ (a_\alpha^+,a_p^+) | a_\alpha^+ \in A_\alpha^{s^+}, a_p^+ \in A_p^{(s^+,a_\alpha^+)} \Big\}$ and $\mathcal{A}_s = \Big\{ (a_\alpha,a_p) | a_\alpha \in A_\alpha^s, a_p \in A_p^{(s,a_\alpha)} \Big\}$, respectively. When $a_\alpha^+ = a_\alpha = w$, for arbitrary $a_p^+ = a_p = m$, we have $R(b+1,g,h,w,m) = R(b,g,h,w,m)$. Since $\min\{\lfloor b+1-w-I_{m>0}m - I_{m=0}\frac{e_d^0(w\alpha_u)}{e_u} + Q_h \rfloor, N_B-1\} \ge \min\{\lfloor b-w-I_{m>0}m - I_{m=0}\frac{e_d^0(w\alpha_u)}{e_u} + Q_h \rfloor, N_B-1\}$, for any $g \in \mathcal{G}, h \in \mathcal{H}$, we have that:

$$\begin{aligned} & V_k(\min\{\lfloor b+1-w-I_{m>0}m - I_{m=0}\frac{e_d^0(w\alpha_u)}{e_u} + Q_h \rfloor, N_B-1\}, g, h) \le \\ & V_k(\min\{\lfloor b-w-I_{m>0}m - I_{m=0}\frac{e_d^0(w\alpha_u)}{e_u} + Q_h \rfloor, N_B-1\}, g, h). \end{aligned} \tag{26}$$

Since $A_p^{(s^+,w)} \supseteq A_p^{(s,w)}$, we can deduce that:

$$\begin{aligned}
&\min_{a_p^+ \in A_p^{s^+}} \Big\{ R(b+1,g,h,w,a_p^+) + \\
&\lambda \sum_{h' \in \mathcal{H}} \sum_{g' \in \mathcal{G}} P(h'|h) P(g'|g) V_k(\min\{\lfloor b+1-w-I_{a_p^+>0} a_p^+ - I_{a_p^+=0} \frac{e_d^0(w\alpha_u)}{e_u} + Q_{h'} \rfloor, N_B - 1\}, g', h') \Big\} \\
&\leq \min_{a_p \in A_p^s} \Big\{ R(b,g,h,w,a_p) + \\
&\lambda \sum_{h' \in \mathcal{H}} \sum_{g' \in \mathcal{G}} P(h'|h) P(g'|g) V_k(\min\{\lfloor b-w-I_{a_p>0} a_p - I_{a_p=0} \frac{e_d^0(w\alpha_u)}{e_u} + Q_{h'} \rfloor, N_B - 1\}, g', h') \Big\}.
\end{aligned} \tag{27}$$

As $A_\alpha^{s^+} \supseteq A_\alpha^s$, we have:

$$\begin{aligned}
&V_{k+1}(b+1,g,h) = \\
&\min_{a_\alpha^+ \in A_\alpha^{s^+}} \min_{a_p^+ \in A_p^{s^+}} \Big\{ R(b+1,g,h,a_\alpha^+,a_p^+) + \\
&\lambda \sum_{h' \in \mathcal{H}} \sum_{g' \in \mathcal{G}} P(h'|h) P(g'|g) V_k(\min\{b+1-a_\alpha^+ - I_{a_p^+>0} a_p^+ - I_{a_p^+=0} \frac{e_d^0(w\alpha_u)}{e_u} + Q_{h'}, N_B - 1\}, g', h') \Big\} \\
&\leq \min_{a_\alpha \in A_\alpha^s} \min_{a_p \in A_p^s} \Big\{ R(b,g,h,a_\alpha,a_p) + \\
&\lambda \sum_{h' \in \mathcal{H}} \sum_{g' \in \mathcal{G}} P(h'|h) P(g'|g) V_k(\min\{b-a_\alpha - I_{a_p>0} a_p - I_{a_p=0} \frac{e_d^0(w\alpha_u)}{e_u} + Q_{h'}, N_B - 1\}, g', h') \Big\} \\
&= V_{k+1}(b,g,h).
\end{aligned} \tag{28}$$

□

According to Lemma 4, we have the following theorem:

Theorem 2. *For any given channel state $g \in \mathcal{G}$ and arrival energy state $h \in \mathcal{H}$, the long-term expected reward achieved by the proposed sensing-transmission policy is non-increasing in the battery state b, that is $V_{\pi_\epsilon^*}(b+1,g,h) \leq V_{\pi_\epsilon^*}(b,g,h)$, $\forall b \in \mathcal{B} \backslash \{N_B - 1\}$.*

Proof of Theorem 2. Assume when $i = k$, the inequality $\|V_{k+1} - V_k\| < \epsilon(1-\lambda)/2\lambda$ holds. According to Step 4 in Algorithm 1, $V_{\pi_\epsilon^*}(s)$ is actually $V_{k+2}(s)$. Based on Lemma 3, we can conclude that $V_{k+2}(b+1,g,h) \leq V_{k+2}(b,g,h)$, namely $V_{\pi_\epsilon^*}(b+1,g,h) \leq V_{\pi_\epsilon^*}(b,g,h)$, $\forall b \in \mathcal{B} \backslash \{N_B - 1\}$. □

From Theorem 2, we perceive that the long-term reward $V_{\pi_\epsilon^*}(s)$ is non-increasing in the battery state b. By taking the parameters in Section 6 except as otherwise stated, the reward of the proposed ϵ-optimal sensing-access policy is depicted in Figure 2. From Figure 2, we can see that $V_{\pi_\epsilon^*}(s)$ is non-increasing in the direction along the battery state, which validates Theorem 2.

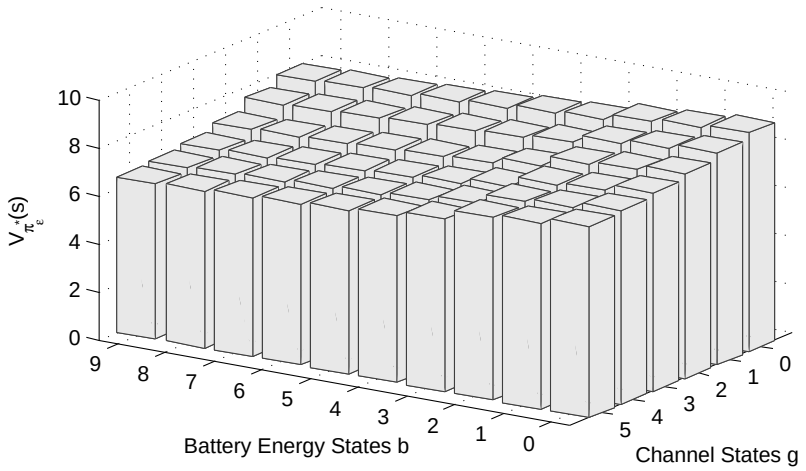

Figure 2. Long-term expected reward $V_{\pi^*_\epsilon}(s)$ with battery energy states and channel states. The arrival energy state is $h = 1$, and the number of battery energy states is $N_B = 10$.

Theorem 3. *For any given channel state $g \in \mathcal{G}$ and arrival energy state $h \in \mathcal{H}$, the optimal long-term expected reward achieved by optimal policy π^* is non-increasing in battery state b, that is $V_{\pi^*}(b+1,g,h) \leq V_{\pi^*}(b,g,h)$, $\forall b \in \mathcal{B}\backslash\{N_B-1\}$.*

Proof of Theorem 3. According to Theorem 2, we acquire that the ϵ-optimal policy is non-increasing in battery state b; therefore, the optimal long-term expected reward $V_{\pi^*}(b,g,h) = \lim_{\epsilon\to 0} V_{\pi^*_\epsilon}(b,g,h)$ is non-increasing in b for any given g and h. □

In the following, we consider a special case where the signal-to-noise ratio (SNR) is sufficiently high. When SNR is sufficiently high, namely $N_0 \to 0$, the reward function for the system state $s = (b,g,h)$ and the corresponding action $a = (a_\alpha, a_p)$ are degenerated to:

$$\lim_{N_0\to 0} R(s,a) = \begin{cases} 1, & a_p = 0, \\ p_i P_f(a_\alpha \alpha_u) + 1 - p_i, & a_p \geq 1, a_\alpha \geq 1. \end{cases} \tag{29}$$

For the i-th iteration, denote the long-term expected reward function with respect to action $a = (a_\alpha, a_p)$ as $V_i^{(a_\alpha,a_p)}$. Then, we have the following theorem.

Theorem 4. *When the SNR is sufficiently high, for any iteration i, the expected reward with action $a = (a_\alpha, 1)$ is no greater than the expected reward with action $a = (a_\alpha, a_p)$, where $a_p \geq 1$. That is, $V_i^{(a_\alpha,1)}(s) \leq V_i^{(a_\alpha,a_p)}(s)$, where $1 \leq a_p \in A_p^{(s,a_\alpha)}$.*

Proof of Theorem 4. The value difference of the two long-term expected rewards with actions $a = (a_\alpha, 1)$ and $a = (a_\alpha, a_p)$ can be calculated as:

$$\begin{aligned} &V_{i+1}^{(a_\alpha,1)}(s) - V_{i+1}^{(a_\alpha,a_p)}(s) \\ &= p_i P_f(a_\alpha\alpha_u) + 1 - p_i + \lambda \sum_{h'\in\mathcal{H}} \sum_{g'\in\mathcal{G}} P(h'|h)P(g'|g)V_i(b'_{a_\alpha+1}, g', h') - \\ &p_i P_f(a_\alpha\alpha_u) + 1 - p_i + \lambda \sum_{h'\in\mathcal{H}} \sum_{g'\in\mathcal{G}} P(h'|h)P(g'|g)V_i(b'_{a_\alpha+a_p}, g', h') \\ &= \lambda \sum_{h'\in\mathcal{H}} \sum_{g'\in\mathcal{G}} P(h'|h)P(g'|g)[V_i(b'_{a_\alpha+1}, g', h') - V_i(b'_{a_\alpha+a_p}, g', h')], \end{aligned} \tag{30}$$

where $b'_x = \min\{b - x + Q_{h'}, N_B - 1\}$. As $b'_{a_\alpha+1} = \min\{b - a_\alpha - 1 + Q_{h'}, N_B - 1\} \geq \min\{b - a_\alpha - a_p + Q_{h'}, N_B - 1\} = b'_{a_\alpha+a_p}$, according to Lemma 4, we have $V_i(b'_{a_\alpha+1}, g', h') - V_i(b'_{a_\alpha+a_p}, g', h') \leq 0$; thus, we can derive $V_{i+1}^{(a_\alpha,1)}(s) - V_{i+1}^{(a_\alpha,a_p)}(s) \leq 0$. □

Based on Theorem 4, we can deduce the following theorem:

Theorem 5. *When the SNR is sufficiently high, for any iteration i with a certain action of sensing overhead a_α, the action set of transmit power to minimize the expected reward is $A_{p_{new}}^{(s,a_\alpha)} = \{0, \min\{1, b - a_\alpha\}\}$, where $a_\alpha \in A_\alpha^s$, $s \in \mathcal{S}$.*

Proof of Theorem 5. When $b \leq 1$, the available transmit power set is $\{0\} \in A_{p_{new}}^{(s,a_\alpha)} = \{0\}$. When $b = 2$, if $a_\alpha = 1$, the available transmit power set is $\{1\} \in A_{p_{new}}^{(s,a_\alpha)} = \{0, \min\{1,1\}\} = \{0,1\}$; otherwise $a_\alpha = 2$; the available transmit power set is $\{0\} \in A_{p_{new}}^{(s,a_\alpha)} = \{0\}$. When $b \geq 3$, we have two cases:

- Case 1: $\min\{\left\lfloor \frac{p_s T}{e_u} \right\rfloor, b\} = b$, then if $a_\alpha = b$, the action set is $\{0\} \in A_{p_{new}}^{(s,a_\alpha)} = \{0\}$; otherwise, for arbitrary $a_\alpha \in A_\alpha^s \backslash \{b\} \leq b - 1$, according to Theorem 4, we have $V_i^{(a_\alpha,1)}(s) \leq V_i^{(a_\alpha,a_p)}(s)$ where $a_p \geq 1$; therefore, the transmit power set to minimize the long-term value $V_i^{(a_\alpha,1)}(s)$ is $\{0,1\} = A_{p_{new}}^{(s,a_\alpha)} = \{0,1\}$.
- Case 2: $\min\{\left\lfloor \frac{p_s T}{e_u} \right\rfloor, b\} < b$, for arbitrary $a_\alpha \in A_\alpha^s \leq b - 1$, according to Theorem 4, we have $V_i^{(a_\alpha,1)}(s) \leq V_i^{(a_\alpha,a_p)}(s)$ where $a_p \geq 1$; therefore, the action set to minimize the long-term value $V_i^{(a_\alpha,1)}(s)$ is $\{0,1\} = A_{p_{new}}^{(s,a_\alpha)} = \{0,1\}$.

Thus, we can derive that the action set to minimize the long-term reward is $A_{p_{new}}^{(s,a_\alpha)} = \{0, \min\{1, b - a_\alpha\}\}$. □

Based on Theorem (5), we present an efficient transmission policy with reduced computational complexity, which is suitable for the case that the *SNR* is sufficiently high, as shown in Algorithm 2.

Algorithm 2 Efficient transmission (ET) policy.

1: Set $V_0(s) = 0$ for all $s \in \mathcal{S}$, set $i = 0$, specify $\epsilon > 0$.
2: For each $s = (b, g, h) \in \mathcal{S}$, formulate the new action space:
$$A_\alpha^s = \begin{cases} \{0\} \ \text{if} \ b_t = 0, \\ \{1, 2, \cdots, \min\{\left\lfloor \frac{p_s T}{e_u} \right\rfloor, b_t\}\} \ \text{otherwise}, \end{cases}$$
$A_{p_{new}}^{(s,a_\alpha)} = \{0, \min\{1, b - a_\alpha\}\}$,
$\mathcal{A}_{s_{new}} = \left\{(a_\alpha, a_p) | a_\alpha \in A_\alpha^s, a_p \in A_{p_{new}}^{(s,a_\alpha)}\right\}$,
Calculate the $V_{i+1}(s)$ according to
$V_{i+1}^a(s) = \left\{R(s,a) + \lambda \sum_{s' \in \mathcal{S}} P(s'|s,a) V_i(s')\right\}$, $a \in \mathcal{A}_{s_{new}}$,
$V_{i+1}(s) = \min\limits_{a \in \mathcal{A}_{s_{new}}} \left\{V_{i+1}^a(s)\right\}$.
3: If $\|V_{i+1} - V_i\| < \epsilon(1 - \lambda)/2\lambda$, go to Step 4. Otherwise, increase i by 1 and go back to step 2.
4: For each $s \in \mathcal{S}$, choose $d(s) = \arg\min\limits_{a \in \mathcal{A}_{s_{new}}} \left\{R(s,a) + \lambda \sum_{s' \in \mathcal{S}} P(s'|s,a) V_{i+1}(s')\right\}$
5: Obtain the efficient transmission policy $\pi_\epsilon^* = \{d, d, \cdots\}$

In Algorithm 2, since $V_i^{(a_\alpha,1)}(s) \leq V_i^{(a_\alpha,a_p)}(s)$ illustrated in Theorem 4, we can ignore the actions that $a_p > 1$ and formulate the new action space $\mathcal{A}_{s_{new}}$ with a lesser number of candidate actions, which reduces the computational complexity significantly. The total number of states in the state space is $N_B \cdot N_G \cdot N_H$. Similar to the analysis of Algorithm 1, the total possible states the current system state

can transmit to is $3N_H$. The maximum number of actions regarding the sensing overhead is N_B, and the maximum number of actions regarding the transmit power is two. Therefore, the complexity of each iteration in Algorithm 2 is $O(N_B^2 N_H^2 N_G)$.

6. Numerical Results and Discussion

In this section, we evaluate the performance and characteristics of the proposed policies by extensive simulations on MatlabR2012a. Unless otherwise stated, the system parameters employed in the simulation are summarized in Table 1, which draws mainly from [26,30,31,42]. The unit of the energy quantum is $e_u = 0.5$ mJ, and $N_B = 20$. The quantization levels of the channel power are $\mathbb{G} = \{0, 0.3, 0.6, 1.0, 2.0, 3.0\}$. The arrival energy takes values from the finite set $\{0, 4e_u, 6e_u, 8e_u\}$ mJ per time slot, namely $Q_0 = 0$, $Q_1 = 4$, $Q_2 = 6$, $Q_3 = 8$, and evolves according to the four-state Markov chain with the state transition probability given by:

$$P_h = \begin{bmatrix} P_{0,0} & P_{0,1} & P_{0,2} & P_{0,3} \\ P_{1,0} & P_{1,1} & P_{1,2} & P_{1,3} \\ P_{2,0} & P_{2,1} & P_{2,2} & P_{2,3} \\ P_{3,0} & P_{3,1} & P_{3,2} & P_{3,3} \end{bmatrix} = \begin{bmatrix} 0.5 & 0.5 & 0 & 0 \\ 0.25 & 0.5 & 0.25 & 0 \\ 0 & 0.25 & 0.5 & 0.25 \\ 0 & 0 & 0.5 & 0.5 \end{bmatrix}. \quad (31)$$

A normalized SNR γ_c (i.e., $\gamma_c = 1/N_0$) is defined with respect to the transmit power of 1 mW throughout the simulation. We choose ϵ to be 10^{-2}. The initial energy state is $b_0 = 6$; the initial channel state is $g_{-1} = 1$; and the initial arrival energy state is $h_{-1} = 1$. The total simulation duration is 500 time slots. All of the numerical results are averaged over 500 independent runs.

Table 1. Simulation parameters.

Parameter	Notation	Value
Duration of a time slot	T	100 ms
Sampling frequency	f_s	1 MHz
Channel idle probability	p_i	0.8
Sensing power	p_s	100 mw
Target detection probability	P_d	0.99
Primary signal's SNR	β	-15 dB
Average channel gain	G_a	2
Normalized Doppler frequency	f_D	0.05
Discount factor	λ	0.99
Normalized SNR	γ_c	10 dB
Data rate threshold	R_{th}	4 bits/time slot/Hz
Idle state power	P_c	3 mw

We compare the proposed sensing-transmission (ST) and efficient transmission (ET) policies with a benchmark named shortsighted policy [32,54] in terms of the performance in Figures 3–6. The primary concern of the shortsighted policy is to minimize the immediate reward of the current time slot, without considering the impact of the current action on the future reward, i.e., $\lambda = 0$. However, the policies proposed in this paper take into account not only the current immediate reward, but also the future expected reward. Therefore, by comparing with the shortsighted policy, we can evaluate the benefit and advantage of proposed policies. Figure 3 depicts the outage probability of ST, ET and the shortsighted policies under different normalized SNRs and channel idle probabilities. First, it can be seen that the ST policy outperforms the shortsighted policy for all settings of normalized SNR. This can be explained by the fact that the ST policy considers a tradeoff between the current immediate reward and the future achievable reward; while the shortsighted policy only focuses on maximizing the current immediate reward, ignoring the impact of the current action on the future reward. It should be noted that despite the better performance of the ST policy, it is much more computationally extensive than the shortsighted policy. Second, we can see that for ST and ET policies,

when γ_c is sufficiently high, the curves of ST and ET policies almost overlap, and a saturation effect is observed, namely the outage probability gradually converges to the same value. This phenomenon coincides with Theorem 5, that is when γ_c is sufficiently high, the transmit action set that SU needs to consider is $A_{p_{new}}^{s,a_\alpha}$, and that ET policy is equivalent to the ST policy in high γ_c regions. Third, we also observe that the saturation outage probability of the three policies in high SNR regions becomes smaller when p_i gets larger. This is because larger p_i indicates more probability of employing the licensed channel for data transmission, resulting in lower outage probability.

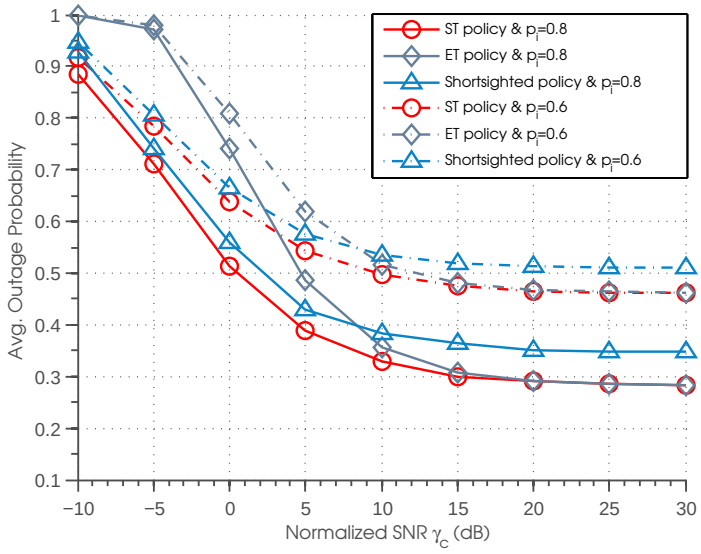

Figure 3. Average outage probability vs. normalized SNR.

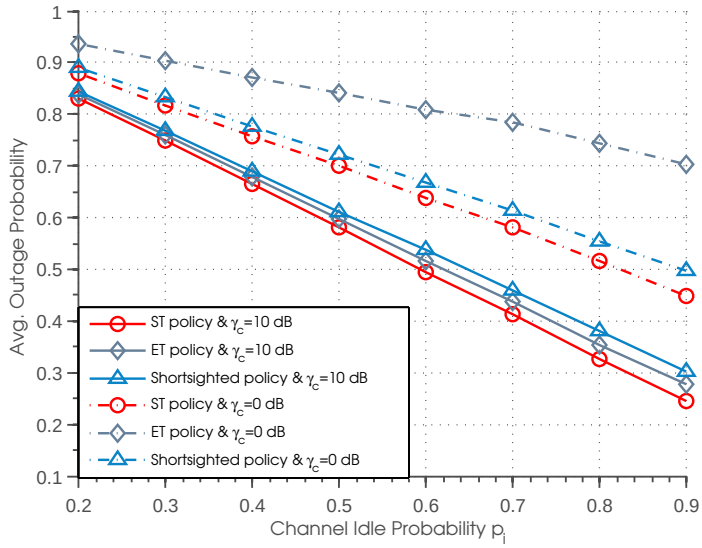

Figure 4. Average outage probability vs. channel idle probability.

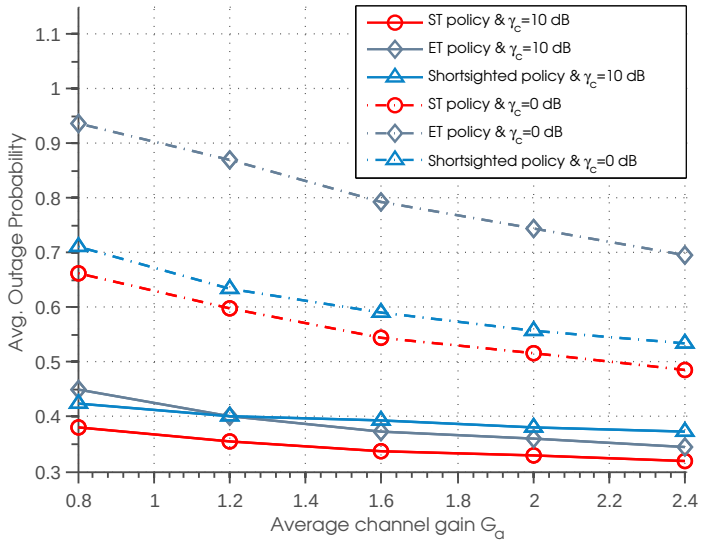

Figure 5. Average outage probability vs. average channel gain.

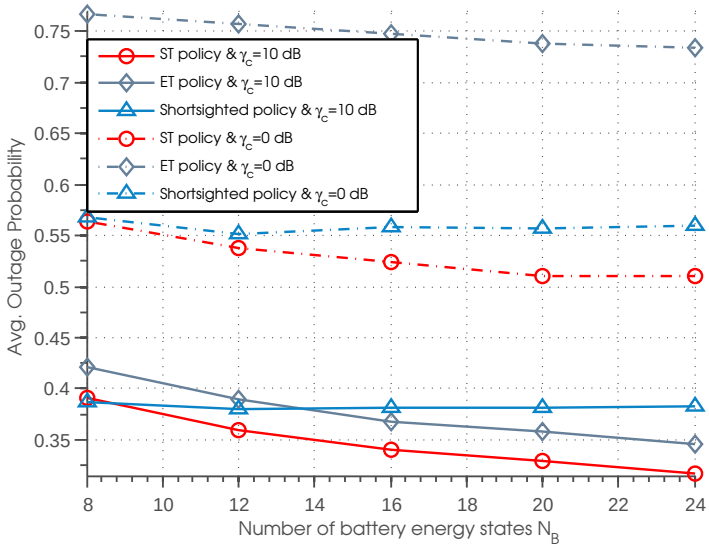

Figure 6. Average outage probability vs. the number of battery energy states.

Figure 4 plots the outage probability of three policies versus the channel idle probability for different values of normalized SNR, where the performance curves plotted correspond to $\gamma_c = 0$ dB and $\gamma_c = 10$ dB, respectively. It can be seen that ST policy outperforms the other two policies for all settings of p_i. Besides, we can observe that the outage probability of all three policies decreases with the increase of channel idle probability, which can be easily understood since a higher value of p_i results in a higher possibility of successful data transmission and therefore reduces the outage probability. We can also observe that when γ_c is small ($\gamma_c = 0$ dB), the gap between the ST and ET policies becomes larger as p_i increases, and the shortsighted policy achieves better performance than the ET policy. While when γ_c is large ($\gamma_c = 10$ dB), there is only a tiny difference between the ST and ET policies, and the ET policy achieves better performance than the shortsighted policy.

Figure 5 illustrates the outage probability of three policies as a function of average channel gain G_a for different γ_c. It can be observed that the outage probability goes down with the increase of G_a. This is due to the fact that as G_a increases, the data transmission is more efficient when the primary channel is idle, resulting in lower outage probability. Besides, we can see that the ST policy outperforms the other two policies for all of the settings of G_a. It is also shown that when γ_c is small ($\gamma_c = 0$ dB),

the shortsighted policy outperforms the ET policy, while when γ_c is large ($\gamma_c = 10$ dB), the ET policy achieve better performance than shortsighted performance in the case that $G_a \geq 1.2$. Thus, we can conclude that in the case that the γ_c is small or the channel quality is poor, the shortsighted policy outperforms the ET policy.

Figure 6 plots the outage probability of three policies with different settings of battery energy state and normalized SNR. It can be seen that the outage probability with respect to ST and ET policies decreases as N_B increases; while the outage probability regarding the shortsighted policy almost remains unchanged under different values of N_B. This phenomenon indicates that by increasing the capacity of the battery, we can efficiently decrease the outage probability, but the performance of the shortsighted policy is almost independent of the battery capacity. Besides, we can also observe that for lower γ_c, the performance of shortsighted policy outperforms the ET policy. For higher γ_c, the ET policy achieves better performance than the shortsighted policy when $N_B \geq 14$.

Figure 7 shows the outage probability of ST and ET policies as a function of γ_c for different data rate threshold R_{th}. We can see that for lower γ_c, a lower data rate threshold leads to a lower outage probability, and the curves with $R_{th} = 2$ outperform the curves with $R_{th} = 4$ and $R_{th} = 6$. However, when γ_c is sufficiently high, we observe that the curves correspond to $R_{th} = 2$, $R_{th} = 4$ and $R_{th} = 6$ all converge to the same value. This is because when γ_c is sufficiently high, according to Equation (29), the reward functions have no relation to R_{th}; the curves with different R_{th} achieve the same outage probability.

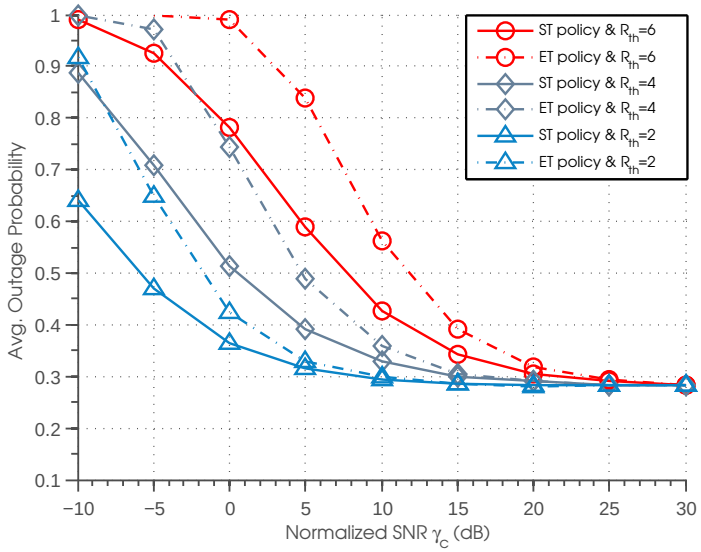

Figure 7. Average outage probability vs. normalized SNR.

The outage probability of ST and ET policies with different settings of battery energy state N_B and idle probability p_i is shown in Figure 8. It can be seen that outage probability of the ST and ET policies decreases as the battery storage capacity N_B increases. This is because with a higher N_B, SU can allocate the energy more efficiently: if the expected channel condition of the next time slot is good and the channel occupancy is estimated to be idle with high probability, the SU can allocate more energy for data transmission; otherwise, the SU can allocate less energy for data transmission and save more energy for future utilization.

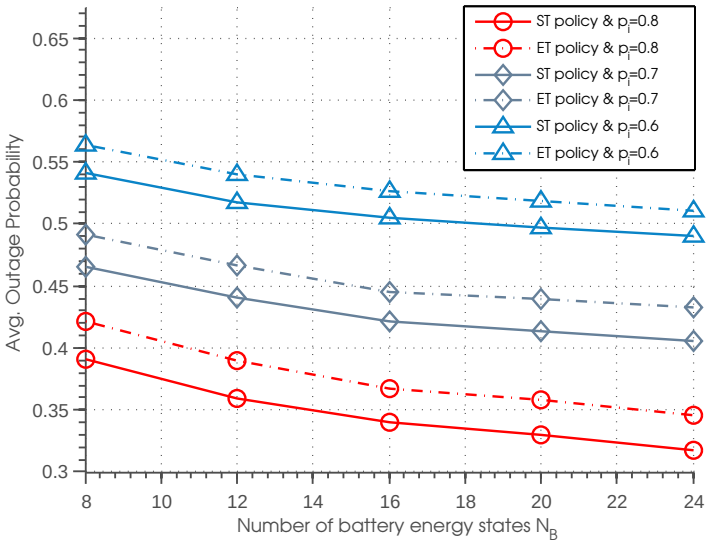

Figure 8. Average outage probability vs. the number of battery energy states.

7. Conclusions

In this paper, we have considered a time-slotted energy harvesting cognitive radio sensor network, where the cognitive sensor nodes solely rely on harvested energy for spectrum sensing and data transmission. Our goal is to minimize the long-term outage probability of the sensor node by adapting the sensing time and transmission power to the current sensor node's knowledge of battery energy, channel fading and harvested energy. This problem has been formulated as an infinite-horizon discounted MDP . The existence of the optimal stationary deterministic policy has been proven, and an ϵ-optimal sensing-transmission policy has been presented through using value iterations. ϵ is the error bound between the ST policy and the optimal policy, which can be pre-defined according to the actual need. Moreover, for a special case where the signal-to-noise (SNR) power ratio is sufficiently high, we have introduced an efficient optimal transmission policy with reduced computational complexity. It has been illustrated that the efficient transmission policy is equivalent to the sensing-transmission policy for high regions of SNR. Finally, we have conducted extensive simulations to verify the performance of the proposed policies, and the impacts of system parameters have also been investigated.

Acknowledgments: The authors would like to thank the support from the Science Foundation of Beijing Jiaotong University (No. 2016YJS027).

Author Contributions: Fan Zhang conceived of and designed the transmission policies. Yan Huo and Kaiwei Jiang designed the simulations and analyzed the data. Tao Jing supported and supervised the research. All of the authors participated in the project, and they read and approved the final manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Staple, G.; Werbach, K. The end of spectrum scarcity [spectrum allocation and utilization]. *IEEE Spectr.* **2004**, *41*, 48–52.
2. Li, S.; Zheng, Z.; Ekici, E.; Shroff, N. Maximizing System Throughput by Cooperative Sensing in Cognitive Radio Networks. *IEEE/ACM Trans. Netw.* **2014**, *22*, 1245–1256.
3. Almasaeid, H.M.; Kamal, A.E. Exploiting Multichannel Diversity for Cooperative Multicast in Cognitive Radio Mesh Networks. *IEEE/ACM Trans. Netw.* **2014**, *22*, 770–783.
4. Joshi, G.P.; Kim, S.W. A Survey on Node Clustering in Cognitive Radio Wireless Sensor Networks. *Sensors* **2016**, *16*, 1465.

5. Mitola, J.; Maguire, G.Q. Cognitive radio: Making software radios more personal. *IEEE Pers. Commun.* **1999**, *6*, 13–18.
6. Ren, J.; Zhang, Y.; Zhang, N.; Zhang, D.; Shen, X. Dynamic Channel Access to Improve Energy Efficiency in Cognitive Radio Sensor Networks. *IEEE Trans. Wirel. Commun.* **2016**, *15*, 3143–3156.
7. Borges, L.M.; Velez, F.J.; Lebres, A.S. Survey on the Characterization and Classification of Wireless Sensor Network Applications. *IEEE Commun. Surv. Tutor.* **2014**, *16*, 1860–1890.
8. Shah, G.A.; Akan, O.B. Cognitive Adaptive Medium Access Control in Cognitive Radio Sensor Networks. *IEEE Trans. Veh. Technol.* **2015**, *64*, 757–767.
9. Salim, S.; Moh, S. An Energy-Efficient Game-Theory-Based Spectrum Decision Scheme for Cognitive Radio Sensor Networks. *Sensors* **2016**, *16*, 1009.
10. Zhang, D.; Chen, Z.; Ren, J.; Zhang, N.; Awad, M.; Zhou, H.; Shen, X. Energy Harvesting-Aided Spectrum Sensing and Data Transmission in Heterogeneous Cognitive Radio Sensor Network. *arXiv* **2016**, arXiv:1604.01519.
11. Yin, S.; Qu, Z.; Li, S. Achievable Throughput Optimization in Energy Harvesting Cognitive Radio Systems. *IEEE J. Sel. Areas Commun.* **2015**, *33*, 407–422.
12. Zhao, Y.; Chen, B.; Zhang, R. Optimal Power Management for Remote Estimation with an Energy Harvesting Sensor. *IEEE Trans. Wirel. Commun.* **2015**, *14*, 6471–6480.
13. Mohjazi, L.; Dianati, M.; Karagiannidis, G.K.; Muhaidat, S.; Al-Qutayri, M. RF-powered cognitive radio networks: Technical challenges and limitations. *IEEE Commun. Mag.* **2015**, *53*, 94–100.
14. Valenta, C.R.; Durgin, G.D. Harvesting Wireless Power: Survey of Energy-Harvester Conversion Efficiency in Far-Field, Wireless Power Transfer Systems. *IEEE Microw. Mag.* **2014**, *15*, 108–120.
15. Zhang, D.; Chen, Z.; Awad, M.K.; Zhang, N.; Zhou, H.; Shen, X.S. Utility-optimal Resource Management and Allocation Algorithm for Energy Harvesting Cognitive Radio Sensor Networks. *IEEE J. Sel. Areas Commun.* **2016**, *34*, 3552–3565.
16. Ren, J.; Zhang, Y.; Deng, R.; Zhang, N.; Zhang, D.; Shen, X. Joint Channel Access and Sampling Rate Control in Energy Harvesting Cognitive Radio Sensor Networks. *IEEE Trans. Emerg. Top. Comput.* **2016**, *99*, 1.
17. Park, S.; Heo, J.; Kim, B.; Chung, W.; Wang, H.; Hong, D. Optimal mode selection for cognitive radio sensor networks with RF energy harvesting. In Proceedings of the IEEE International Symposium on Personal, Indoor and Mobile Radio Communications—(PIMRC), Sydney, Australia, 9–12 September 2012; pp. 2155–2159.
18. Caleffi, M.; Cacciapuoti, A.S. Database access strategy for TV White Space cognitive radio networks. In Proceedings of the IEEE International Conference on Sensing, Communication, and Networking Workshops (SECON Workshops), Singapore, 30 June–3 July 2014; pp. 34–38.
19. Caleffi, M.; Cacciapuoti, A.S. Optimal Database Access for TV White Space. *IEEE Trans. Commun.* **2016**, *64*, 83–93.
20. Caleffi, M.; Cacciapuoti, A.S. On the Achievable Throughput over TVWS Sensor Networks. *Sensors* **2016**, *16*, 457.
21. Cacciapuoti, A.S.; Caleffi, M.; Paura, L. Optimal Strategy Design for Enabling the Coexistence of Heterogeneous Networks in TV White Space. *IEEE Trans. Veh. Technol.* **2016**, *65*, 7361–7373.
22. Xiao, H.; Yang, K.; Wang, X. Robust Power Control under Channel Uncertainty for Cognitive Radios with Sensing Delays. *IEEE Trans. Wirel. Commun.* **2013**, *12*, 646–655.
23. Ozel, O.; Tutuncuoglu, K.; Yang, J.; Ulukus, S.; Yener, A. Transmission with Energy Harvesting Nodes in Fading Wireless Channels: Optimal Policies. *IEEE J. Sel. Areas Commun.* **2011**, *29*, 1732–1743.
24. Ho, C.K.; Zhang, R. Optimal Energy Allocation for Wireless Communications With Energy Harvesting Constraints. *IEEE Trans. Signal Process.* **2012**, *60*, 4808–4818.
25. Mao, S.; Cheung, M.H.; Wong, V.W.S. An optimal energy allocation algorithm for energy harvesting wireless sensor networks. In Proceedings of the IEEE International Conference on Communications (ICC), Ottawa, ON, Canada, 10–15 June 2012; pp. 265–270.
26. Mao, S.; Cheung, M.H.; Wong, V.W.S. Optimal Joint Energy Allocation for Sensing and Transmission in Rechargeable Wireless Sensor Networks. *IEEE Trans. Veh. Technol.* **2014**, *63*, 2862–2875 .
27. Ahmed, I.; Ikhlef, A.; Ng, D.W.K.; Schober, R. Power Allocation for an Energy Harvesting Transmitter with Hybrid Energy Sources. *IEEE Trans. Wirel. Commun.* **2013**, *12*, 6255–6267.

28. Mao, Y.; Yu, G.; Zhang, Z. On the Optimal Transmission Policy in Hybrid Energy Supply Wireless Communication Systems. *IEEE Trans. Wirel. Commun.* **2014**, *13*, 6422–6430.
29. Siddiqui, A.M.; Musavian, L.; Ni, Q. Energy efficiency optimization with energy harvesting using harvest-use approach. In Proceedings of the IEEE International Conference on Communication Workshop (ICCW), London, UK, 8–12 June 2015; pp. 1982–1987.
30. Ku, M.L.; Chen, Y.; Liu, K.J.R. Data-Driven Stochastic Models and Policies for Energy Harvesting Sensor Communications. *IEEE J. Sel. Areas Commun.* **2015**, *33*, 1505–1520.
31. Jeya, J.P.; Kalamkar, S.S.; Banerjee, A. Energy Harvesting Cognitive Radio With Channel-Aware Sensing Strategy. *IEEE Commun. Lett.* **2014**, *18*, 1171–1174.
32. Sultan, A. Sensing and Transmit Energy Optimization for an Energy Harvesting Cognitive Radio. *IEEE Wirel. Commun. Lett.* **2012**, *1*, 500–503.
33. Gao, X.; Xu, W.; Li, S.; Lin, J. An online energy allocation strategy for energy harvesting cognitive radio systems. In Proceedings of the International Conference on Wireless Communications Signal Processing (WCSP), Hangzhou, China, 24–26 October 2013; pp. 1–5.
34. Park, S.; Kim, H.; Hong, D. Cognitive Radio Networks with Energy Harvesting. *IEEE Trans. Wirel. Commun.* **2013**, *12*, 1386–1397.
35. Park, S.; Hong, D. Optimal Spectrum Access for Energy Harvesting Cognitive Radio Networks. *IEEE Trans. Wirel. Commun.* **2013**, *12*, 6166–6179.
36. Park, S.; Hong, D. Achievable Throughput of Energy Harvesting Cognitive Radio Networks. *IEEE Trans. Wirel. Commun.* **2014**, *13*, 1010–1022.
37. Blasco, P.; Gunduz, D.; Dohler, M. A Learning Theoretic Approach to Energy Harvesting Communication System Optimization. *IEEE Trans. Wirel. Commun.* **2013**, *12*, 1872–1882.
38. Razavi, A.; Valkama, M.; Cabric, D. Compressive Detection of Random Subspace Signals. *IEEE Trans. Signal Process.* **2016**, doi:10.1109/TSP.2016.2560132.
39. Sinha, K.; Sinha, B.P.; Datta, D. An Energy-Efficient Communication Scheme for Wireless Networks: A Redundant Radix-Based Approach. *IEEE Trans. Wirel. Commun.* **2011**, *10*, 550–559.
40. Liang, Y.C.; Zeng, Y.; Peh, E.C.Y.; Hoang, A.T. Sensing-Throughput Tradeoff for Cognitive Radio Networks. *IEEE Trans. Wirel. Commun.* **2008**, *7*, 1326–1337.
41. Goldsmith, A. *Wireless Communications*; Cambridge University Press: New York, NY, USA, 2005.
42. Pei, Y.; Liang, Y.C.; Teh, K.C.; Li, K.H. Energy-Efficient Design of Sequential Channel Sensing in Cognitive Radio Networks: Optimal Sensing Strategy, Power Allocation, and Sensing Order. *IEEE J. Sel. Areas Commun.* **2011**, *29*, 1648–1659.
43. Cheng, H.T.; Zhuang, W. Simple Channel Sensing Order in Cognitive Radio Networks. *IEEE J. Sel. Areas Commun.* **2011**, *29*, 676–688.
44. Wu, Y.; Lau, V.K.N.; Tsang, D.H.K.; Qian, L.P. Energy-Efficient Delay-Constrained Transmission and Sensing for Cognitive Radio Systems. *IEEE Trans. Veh. Technol.* **2012**, *61*, 3100–3113.
45. Chen, Y.; Zhao, Q.; Swami, A. Distributed Spectrum Sensing and Access in Cognitive Radio Networks With Energy Constraint. *IEEE Trans. Signal Process.* **2009**, *57*, 783–797.
46. Ngo, M.H.; Krishnamurthy, V. Monotonicity of Constrained Optimal Transmission Policies in Correlated Fading Channels With ARQ. *IEEE Trans. Signal Process.* **2010**, *58*, 438–451.
47. Wang, Y.; Xu, Y.; Shen, L.; Xu, C.; Cheng, Y. Two-dimensional POMDP-based opportunistic spectrum access in time-varying environment with fading channels. *J. Commun. Netw.* **2014**, *16*, 217–226.
48. Amirnavaei, F.; Dong, M. Online Power Control Optimization for Wireless Transmission With Energy Harvesting and Storage. *IEEE Trans. Wirel. Commun.* **2016**, *15*, 4888–4901.
49. Khairnar, P.S.; Mehta, N.B. Power and Discrete Rate Adaptation for Energy Harvesting Wireless Nodes. In Proceedings of the IEEE International Conference on Communications (ICC), Kyoto, Japan, 5–9 June 2011; pp. 1–5.
50. Wang, H.S.; Moayeriu, N. Finite-state Markov channel-a useful model for radio communication channels. *IEEE Trans. Veh. Technol.* **1995**, *44*, 163–171.
51. Zou, Y.; Yao, Y.; Zheng, B. Outage probability analysis of cognitive transmissions: The impact of spectrum sensing overhead. *IEEE Trans. Wirel. Commun.* **2010**, *33*, 2676–2688.
52. Puterman, M.L. *Markov Decision Processes: Discrete Stochastic Dynamic Programming*; John Wiley & Sons: New York, NY, USA, 2005.

53. Littman, M.L.; Dean, T.L.; Kaelbling, L.P. On the Complexity of Solving Markov Decision Problems. In Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence, Montreal, QC, USA, 18–20 August 1995; pp. 394–402.
54. Zhao, Q.; Krishnamachari, B.; Liu, K. On myopic sensing for multi-channel opportunistic access: Structure, optimality, and performance. *IEEE Trans. Wirel. Commun.* **2008**, *7*, 5431–5440.