

Communication over a Time Correlated Channel with an Energy Harvesting Transmitter

Mehdi Salehi Heydar Abad

Faculty of Engineering
and Natural Sciences

Sabanci University, Istanbul, Turkey
mehdis@sabanciuniv.edu

Deniz Gündüz

Department of Electrical
and Electronic Engineering,

Imperial College London, U.K.
d.gunduz@imperial.ac.uk

Ozgur Ercetin

Faculty of Engineering
and Natural Sciences

Sabanci University, Istanbul, Turkey
oercetin@sabanciuniv.edu

Abstract—In this work, communication over a time-correlated point-to-point wireless channel is studied for an energy harvesting (EH) transmitter. In this model, we take into account the time and energy cost of acquiring channel state information. At the beginning of the time slot, the EH transmitter, has to choose among three possible actions: i) deferring the transmission to save its energy for future use, ii) transmitting without sensing, and iii) sensing the channel before transmission. At each time slot, the transmitter chooses one of the three possible actions to maximize the total expected discounted number of bits transmitted over an infinite time horizon. This problem can be formulated as a partially observable Markov decision process (POMDP) which is then converted to an ordinary MDP by introducing a belief on the channel state, and the optimal policy is shown to exhibit a threshold behavior on the belief state, with battery-dependent threshold values. Optimal threshold values and corresponding optimal performance are characterized through numerical simulations, and it is shown that having the sensing action and intelligently using it to track the channel state improves the achievable long-term throughput significantly.

I. INTRODUCTION

Due to the tremendous increase in the number of battery-powered wireless communication devices over the past decade, replenishing the batteries of these devices by harvesting energy from natural resources has become an important research area [1]. Transmitters may harvest energy via wind turbines, photovoltaic cells, thermoelectric generators, or from mechanical vibrations through piezoelectric or electromagnetic technology [2]. Regardless of which type of energy harvesting (EH) device and natural energy source is employed, a main concern is the stochastic nature of the EH process driving the wireless communications. The associated battery recharging process can be modeled either as a continuous, or a discrete, [3], [4] stochastic process.

We consider a wireless point-to-point link with a transmitter equipped with a finite-capacity battery fed by an EH device. At each time slot, a unit of energy is harvested by the transmitter

according to a binary random process independent over time¹. We assume that the transmitter can accurately observe the current energy level of the battery, and it has the knowledge of the statistics of the EH process. The wireless channel is time-varying and has memory across time. The channel memory is modeled with a finite state Markov chain [5], where the next channel state depends only on the current state. A convenient and often-employed simplification of the Markov model is a two state Markov chain, known as the Gilbert-Elliot channel [6]. This model assumes that the channel can be either in a *good* or a *bad* state. We assume that in the bad state, transmitter cannot transmit any information reliably, while in the good state it may transmit R bits per time slot by spending exactly one unit of energy from its battery.

In this work, differently from most of the literature on EH systems, we take into account the energy cost of acquiring channel state information (CSI). At the beginning of each time slot, without knowing the current CSI, EH transmitter has three possible actions: i) deferring the transmission to save its energy for future use, ii) transmitting at a rate of R bits per time slot, and iii) sensing the channel to reveal the channel state by consuming a portion of its energy and transmission time, followed by transmission at a reduced rate consuming the remainder of the energy unit, if the channel is in the good state. If the channel is in a bad state, the transmitter remains silent in the rest of the time slot, saving its energy for future. If the level of the battery is less than a unit of energy at the beginning of a time slot, no transmission is possible. Our objective is to maximize the total expected discounted number of bits transmitted over an infinite time horizon.

Markov decision process (MDP) tools have been extensively utilized in the recent literature in solving communication problems involving EH devices. In [7] authors propose a simple single-threshold policy for a solar-powered sensor operating over a fading wireless channel. Optimality of a single-threshold policy is proven [8] when transmitting packets with importance values on EH transmitter. Problem of

This work was in part supported by EC H2020-MSCA-RISE-2015 programme under grant number 690893, by Tubitak under grant number 114E955 and by British Council Institutional Links Program under grant number 173605884.

¹Typically, the EH process is neither memoryless nor discrete, and the energy is accumulated continuously over time. However, in order to develop the analytical model underlying this paper, we follow the common assumption in the literature [3], and assume that the continuous energy arrival is accumulated in an intermediate energy storage device to form quantas.

energy allocation for gathering and transmitting data in an EH communication system is studied in [9] and [10]. The scheduling of EH transmitters with time correlated energy arrivals to optimize the long term sum throughput is investigated in [11]. The allocation of energy over a finite horizon to optimize the throughput is considered in [12], where it is assumed that either the current or the future energy and channel states are provided to the transmitter. In [13], for a Markov EH process, and a static channel, a discrete power allocation problem is studied to maximize the throughput. In [14] throughput is optimized over a multiple access channel with collisions, considering spatially correlated energy arrivals at the transmitters.

In a closely related work [15], scheduling of an EH transmitter over a Gilbert-Elliot channel is considered. However, unlike our work, the transmitter [15] always has perfect CSI, obtained by sensing at every time slot, and makes a decision to defer or to transmit, based on the current CSI and battery state. Similarly, without considering the channel sensing capability, [16] addresses the problem of optimal power management for an EH sensor over a multi-state wireless channel with memory, using the ACK/NACK channel feedback to track the channel state. In our work, instead, we take into account the energy cost of channel sensing which can be significant for EH transmitters. Therefore, the EH transmitter does not necessarily have perfect CSI, but keeps an updated belief of the channel state according to its past observations. Hence, the transmitter may occasionally take a third decision (in addition to defer and to transmit) of sensing the current channel state to improve its belief. Channel sensing is an essential part of opportunistic and cognitive spectrum access. In [17], the authors investigate the problem of optimal access to a Gilbert-Elliot channel, wherein an energy-unlimited transmitter senses the channel at every time slot. In [18] channel sensing is done only occasionally. The transmitter can decide to transmit at a high or a low rate without sensing the channel; or can first sense the channel and transmit at a reduced rate due to the time spent for sensing. The energy cost of sensing is ignored in [18].

In Section II we explain the channel and EH process models under consideration, and elaborate on the transmission protocol. In Section III, we formulate the problem as a two state partially observable MDP (POMDP) which is then converted to a continuous-state MDP by introducing a belief state. In Section IV we show that the optimal policy is of threshold type, for which the optimal threshold values depend on the state of the battery. In Section V we present simulation results that numerically obtain the optimal threshold values and the optimal performance. In Section VI we conclude the paper and present future research directions.

II. SYSTEM MODEL

A. Channel and energy harvesting models

Consider the communication system illustrated in Fig. 1, in which an EH transmitter communicates over a slotted Gilbert-Elliot channel. Let G_t denote the state of the channel at

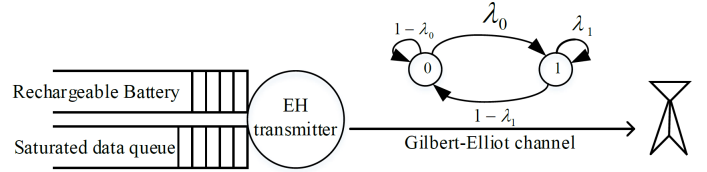


Fig. 1: System model.

time slot t which is modeled as a one-dimensional Markov chain with two states: a good state denoted by 1, and a bad state denoted by 0. Channel transitions occur at the beginning of each time slot. The transition probabilities are given by $\mathbb{P}[G_t = 1 | G_{t-1} = 1] = \lambda_1$ and $\mathbb{P}[G_t = 1 | G_{t-1} = 0] = \lambda_0$. The transmitter can transmit R bits per time slot if $G_t = 1$, and zero bits if $G_t = 0$.

A unit of energy arrives at the end of time slot t according to an independent and identically distributed (i.i.d.) Bernoulli process, denoted by E_t , with probability q , i.e., $\mathbb{P}[E_t = 1] = q$ for all t . The transmitter stores the energy packets in a battery with a storage capacity of B_{max} units of energy. We denote the state of the battery, i.e., the energy available in the battery at the beginning of time slot t , by B_t . An energy unit is consumed at each slot if the transmitter decides to transmit in that slot. A unit of energy consumed per slot includes the energy cost of sensing (if the transmitter decides to sense the channel), transmission of the message, and the reception of ACK or NACK from the receiver. We assume that the transmitter has an infinitely backlogged data queue, and thus, it always has a packet to transmit.

B. Transmission protocol

At the beginning of each time slot, the transmitter may choose among three possible actions: *i*) deferring the transmission, *ii*) channel sensing and transmitting opportunistically, and *iii*) transmitting without sensing.

Deferring the transmission: This action (denoted by D) corresponds to the case in which the transmitter either believes that the channel is in a bad state, or observes that its battery has low energy. If this action is chosen, there is no message exchange between the transmitter and the receiver. Hence, the receiver does not send any feedback, and therefore the transmitter cannot obtain any knowledge about the current channel state. The scenario in which the transmitter is informed about the current channel state even when it does not transmit any data packet is equivalent to the system model investigated in [15].

Channel sensing and transmitting opportunistically: This action (denoted by O) corresponds to the case in which the transmitter decides to sense the channel at the beginning of the time slot. We assume that sensing consumes a fraction $0 < \tau < 1$ of an energy unit. Sensing is carried out by the transmitter first sending a control/probing packet, to which, the receiver responds with a packet indicating the channel state. We assume that the time it takes to sense the channel is τ seconds and the transmitter consumes on average the same

power as data transmission over the sensing period. Therefore, we equivalently assume for simplicity that $\tau = 1/k$ for some $k \in \mathbb{Z}^+$. In the remaining $1 - \tau$ seconds, the transmitter may choose to transmit data at the same rate it would without channel sensing, which means that by the end of the time slot it transmits $(1 - \tau)R$ bits per time slot.

If the channel is revealed to be in the bad state, transmitter defers its transmission and saves the rest of the energy unit (i.e., $1 - \tau$). Note that thanks to the channel sensing capability, in the case of a bad state, the transmitter wastes only τ portion of a unit energy packet, and saves the remaining energy by deferring its transmission, which as we will show later in this paper, is an important advantage in EH networks with scarce energy sources.

Transmitting without sensing: This action (denoted by T) corresponds to the case when transmitter attempts to transmit R bits in the current time slot without sensing the channel. If the channel is in a good state, the transmission is successful and the receiver sends an ACK. Otherwise, the transmission fails, and the receiver sends a NACK. Note that, at the end of the slot the transmitter has the perfect knowledge of the current channel state.

III. PARTIALLY OBSERVABLE MARKOV DECISION PROCESS (POMDP) FORMULATION

At the beginning of each time slot, the transmitter chooses among the three possible actions based on the state of its battery, and its belief about the channel state to maximize a long-term discounted reward to be defined shortly. Although the transmitter is perfectly aware of its battery state, it cannot directly observe the current channel state. Hence, the problem in hand becomes a partially observable Markov decision process (POMDP).

Let the state of the system at time t be denoted by $S_t = (B_t, X_t)$. We define the *belief* of the transmitter at time slot t , denoted by X_t , as the conditional probability that the channel is in the good state at the beginning of the current slot, i.e., $X_t = \mathbb{P}[G_t = 1 | \mathcal{H}_t]$, given the history \mathcal{H}_t , where \mathcal{H}_t represents all the past actions and observations of the transmitter up to slot t . The transmitter's belief constitutes a sufficient statistic to characterize its optimal actions [19]. Note that with this definition of the state, the POMDP problem is converted into a MDP with an uncountable state space $[0, \tau, 2\tau, \dots, B_{\max}] \times [0, 1]^2$.

A transmission policy π describes a set of rules that dictates which action to take depending on the history. Let $V^\pi(b, p)$ be the expected infinite-horizon discounted reward with initial state $S_0 = (b, \mathbb{P}[G_0 = 1 | \mathcal{H}_0] = p)$ under policy π with discount factor $\beta \in [0, 1)$. The use of the expected discounted reward allows us to obtain a tractable solution, and one can gain insights into the optimal policy for the average reward when β is close to 1. It is also discussed in [5] that β can be interpreted as the probability that a particular user is allowed

to use the channel, or as the probability of the transmitter to remain active at each time slot as in [20]. For an initial belief p the expected discounted reward has the following expression

$$V^\pi(b, p) = \mathbb{E} \left[\sum_{t=0}^{\infty} \beta^t R(S_t, A_t) | S_0 = (b, p) \right], \quad (1)$$

where t is the time index, $A_t \in \{D, O, T\}$ is the action chosen at time t , and $R(S_t, A_t)$ is the expected reward acquired when action A_t is taken at state S_t . The expectation in (1) is over state sequence distribution induced by the given transmission policy π . The expected reward when action A_t is chosen at state S_t is given as follows:

$$R(S_t, A_t) = \begin{cases} X_t R & \text{if } A_t = T \text{ and } B_t \geq 1, \\ (1 - \tau)X_t R & \text{if } A_t = O \text{ and } B_t \geq 1, \\ 0 & \text{otherwise.} \end{cases} \quad (2)$$

Since at least one energy unit is required for transmission, if the battery state is less than one unit, the reward becomes zero. Hence, in explaining the expected reward function in (2), we consider actions when the battery state is greater than or equal to one. If the action of transmitting without sensing is chosen, R bits per time slot are transmitted successfully if the channel is in a good state, and 0 bits if the channel is in a bad state. Since the belief, X_t , represents the probability of the channel being in a good state, the expected reward is given by $X_t R$. If the action of transmitting opportunistically is chosen, τ fraction of energy unit is spent sensing the channel with the remaining energy being used for transmission if the channel is sensed to be in a good state. In this case, $(1 - \tau)R$ bits per time slot are transmitted successfully. If the channel is sensed to be in a bad state, the transmitter remains silent in the rest of the time slot. The expected reward in this case is $(1 - \tau)X_t R$. Finally, if the action of deferring the transmission is taken the transmitter neither senses the channel nor transmits, so the reward is zero.

Define the value function $V(b, p)$ as

$$V(b, p) = \max_{\pi} V^\pi(b, p) \quad \text{for all } b \in [0, \tau, 2\tau, \dots, B_{\max}] \text{ and } p \in [0, 1]. \quad (3)$$

It is well known that the optimal value of the infinite-horizon expected reward can be achieved by a stationary policy, i.e., there exists a stationary policy π^* such that $V(b, p) = V^{\pi^*}(b, p)$ [21]. The value function $V(b, p)$ satisfies the Bellman equation

$$V(b, p) = \max_{A \in \{D, O, T\}} \{V_A(b, p)\}, \quad (4)$$

where $V_A(b, p)$ is the action-value function, defined as the expected infinite-horizon discounted reward acquired by taking action A when the state is (b, p) , and is given by

$$V_A(b, p) = R((b, p), A) + \beta \mathbb{E}_{(\hat{b}, \hat{p})} [V(\hat{b}, \hat{p}) | S_0 = (b, p), A_0 = A], \quad (5)$$

where (\hat{b}, \hat{p}) denotes the next state when action A is chosen at

²Note that since sensing without transmission is possible, i.e., consuming only τ fraction of the energy unit, the battery can take fraction of units as states.

state $S_0 = (b, p)$. The expectation in (5) is over the distribution of possible next states. In the following, we define and explain the value function $V_A(b, p)$, and how the system state evolves for each action.

Deferring the transmission: If this action is taken, since there is no transmission, there is no ACK or NAK from the receiver, and thus, the transmitter does not learn the state of the channel. Therefore the next belief is obtained as the probability of finding the channel in a good state given the current belief state. If the transmitter had a belief $X_t = p$ at time slot t , after taking action D, its belief at the beginning of the next slot is updated as

$$J(p) = \lambda_0(1 - p) + \lambda_1 p. \quad (6)$$

In every time slot, a unit of energy is harvested with probability q . Thus, after taking action D, the value function evolves as follows:

$$\begin{aligned} V_D(b, p) &= \beta [qV(\min\{b+1, B_{\max}\}, J(p)) + (1-q)V(b, J(p))]. \end{aligned} \quad (7)$$

Note that the term $\min\{b+1, B_{\max}\}$ is used to ensure that the battery state does not exceed the battery capacity, B_{\max} .

Channel sensing and transmitting opportunistically: For this action, two scenarios are possible. If $b \geq 1$ and EH decides to transmit opportunistically, then it consumes τ fraction of energy to first sense the channel and obtain the current channel state. Based on the outcome of the channel sensing, if the channel is found to be in a good state, $(1-\tau)$ units of energy is used to transmit $(1-\tau)R$ bits per time slot. Also, the belief state is updated as λ_1 for the next time slot.

On the other hand, if the outcome of the channel sensing reveals the channel to be in a bad state, then the transmitter defers its transmission, and saves $(1-\tau)$ units of energy for possible future transmissions. Also, the channel belief is updated as λ_0 for the next time slot. Based on the aforementioned discussion, for $b \geq 1$ the evolution of the value function can be written as:

$$\begin{aligned} V_O(b, p) &= p[(1-\tau)R + \beta(qV(b, \lambda_1) + (1-q)V(b-1, \lambda_1))] \\ &\quad + (1-p)\beta[qV(\min\{b-\tau+1, B_{\max}\}, \lambda_0) \\ &\quad + (1-q)V(b-\tau, \lambda_0)]. \end{aligned} \quad (8)$$

If $\tau \leq b < 1$, then transmission is not possible since transmission requires at least one unit of energy. However, it is still possible to sense the channel, since it only requires τ fraction of energy. This may happen when transmitter believes that learning the channel state will help its decision in the future. Thus for $\tau \leq b < 1$, the value function evolves as:

$$\begin{aligned} V_O(b, p) &= \beta[qpV(b-\tau+1, \lambda_1) \\ &\quad + q(1-p)V(b-\tau+1, \lambda_0) + (1-q)pV(b-\tau, \lambda_1) \\ &\quad + (1-q)(1-p)V(b-\tau, \lambda_0)]. \end{aligned} \quad (9)$$

Transmitting without sensing: This action can only be chosen if the battery state is greater than or equal to one, i.e.,

$b \geq 1$ ³. Under this action, the transmitter transmits regardless of the actual state of the channel, costing one unit of energy. If the channel is in the good state, R bits per time slot are successfully delivered to the receiver, and the receiver sends back an ACK. Otherwise, the channel is in the bad state, so the transmission fails, and the receiver sends back a NAK. Meanwhile, the channel is in a good state with probability p , i.e., the current belief state, and the belief in the next time slot will be λ_1 . Also the channel is in a bad state with probability $1-p$ and the belief in the next time slot will be λ_0 . Hence, the value function evolves as:

$$\begin{aligned} V_T(b, p) &= p[R + \beta(qV(b, \lambda_1) + (1-q)V(b-1, \lambda_1))] \\ &\quad + (1-p)\beta[qV(b, \lambda_0) + (1-q)V(b-1, \lambda_0)] \end{aligned} \quad (10)$$

IV. STRUCTURE OF THE OPTIMAL POLICY

In this section, we prove that the optimal policy has a threshold type structure on the belief state. First, we need to prove some of the properties of the value function. We begin with establishing the convexity of the optimal value function with respect to the belief state.

Lemma 1. *For any given $b \geq 0$, $V(b, p)$ is convex in p .*

Proof. The proof is given in [22]. \square

In the following lemma, we show that the value function is a non decreasing function of battery state, b . This lemma provides the intuition why deferring or sensing actions are advantageous in some states. The incentive of taking these actions is that the value function transitions into higher values without consuming any energy. Moreover, it states that the value function is also non-decreasing with respect to the belief state, p .

Lemma 2. *Given any belief p , $V(b_1, p) \geq V(b_0, p)$ when $b_1 > b_0$. Moreover, for a fixed battery state b , if $p_1 > p_0$ then $V(b, p_1) \geq V(b, p_0)$.*

Proof. The proof is given in [22]. \square

Finally, Theorem 1 below shows that the optimal solution of the problem is a threshold policy with two or three thresholds depending on the system parameters. The threshold values depend on the state of the battery.

Theorem 1. *For any $p \in [0, 1]$ and $b \geq 0$, there exists thresholds $0 \leq \rho_1(b) \leq \rho_2(b) \leq \rho_3(b) \leq 1$, all of which are functions of the battery state b , such that, for $b \geq 1$*

$$\pi^*(b, p) = \begin{cases} D, & \text{if } 0 \leq p \leq \rho_1(b) \text{ or } \rho_2(b) \leq p \leq \rho_3(b) \\ O, & \text{if } \rho_1(b) \leq p \leq \rho_2(b), \\ T, & \text{if } \rho_3(b) \leq p \leq 1, \end{cases} \quad (11)$$

³Note that we are aware that in the generic MDP formulation, in every state, we should have the same set of actions. We can re-define the reward function by assigning $-\infty$ reward for those actions that are not possible to be taken in specific states to account for this issue. For the ease of comprehension, we chose to present the formulation in this manner.

and for $\tau \leq b < 1$,

$$\pi^*(b, p) = \begin{cases} D, & \text{if } 0 \leq p \leq \rho_1(b) \text{ or } \rho_2(b) \leq p \leq 1, \\ O, & \text{if } \rho_1(b) \leq p \leq \rho_2(b). \end{cases} \quad (12)$$

Proof. The detailed proof of the theorem is given in [22]. \square

Theorem 1 proves that at any battery state $b \geq 1$, at most three threshold values are sufficient to characterize the optimal policy; whereas two thresholds suffice for $0 \leq b < 1$. However the optimal policy can even be simpler for some battery states and some instances of the problem as it is possible to have $\rho_2(b) = \rho_3(b)$, or even $\rho_1(b) = \rho_2(b) = \rho_3(b)$.

V. NUMERICAL RESULTS

In this section we use numerical techniques to characterize the optimal policy, and evaluate its performance. We utilize the value iteration algorithm to calculate the optimal value function. We numerically identify the thresholds for the optimal policy for different scenarios. We also evaluate the performance of the optimal policy, and compare it with some alternative policies in terms of throughput.

A. Optimal policy evaluation

In the following, we assume that $B_{max} = 5$, $\tau = 0.2$, $\beta = 0.98$, $\lambda_1 = 0.9$, $\lambda_0 = 0.6$, $R = 3$ and $q = 0.1$. The optimal policy is evaluated using the value iteration algorithm. In Fig. 2 each state (b, p) is illustrated with a different color corresponding to the optimal policy at that state. In the figure, the areas highlighted with blue color correspond to those states at which deferring the transmission is optimal, green areas correspond to the states at which transmitting opportunistically is optimal, and finally yellow areas correspond to the states for which transmitting without sensing is optimal. As seen in Fig. 2 any of the three policies (one, two, or three threshold policies) may be optimal depending on the level of the battery state. For example, when the battery state is $b = 2$, one-threshold policy is optimal. The transmitter defers transmission up to a belief of state of $p = 0.8$ and starts transmitting without sensing beyond this value. For no value of the belief state it opts for sensing the channel. On the other hand, when the battery state is 3.8, two-threshold policy is optimal, and when the battery state is 2.8, three-threshold policy is optimal. Considering the low probability of energy arrivals ($q = 0.1$) and the relative high cost of sensing ($\tau = 0.2$), it is interesting to notice that the transmitter senses the channel even when its battery state is below the transmission threshold, i.e., $b < 1$.

Next, we investigate the effect of the sensing cost, τ , on the optimal policy. To illustrate this effect, we choose the system parameters as before, but increase the sensing cost from $\tau = 0.2$ to $\tau = 0.5$. Optimal action regions for this setup are shown in Fig. 3. By comparing Fig. 2 and Fig. 3, it is evident that a higher cost of sensing results in less incentive for sensing the channel. We observe in Fig. 3 that the green area has shrunk almost to nothing, i.e., the transmitter is more likely to take a risk and transmit without sensing, or defer its transmission, when sensing consumes more energy.

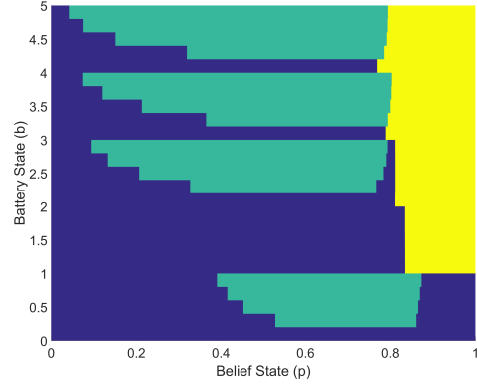


Fig. 2: Optimal thresholds for taking the actions D (blue), O (green), T (yellow) for $B_{max} = 5$, $\tau = 0.2$, $\beta = 0.98$, $\lambda_1 = 0.9$, $\lambda_0 = 0.6$, $R = 3$ and $q = 0.1$.

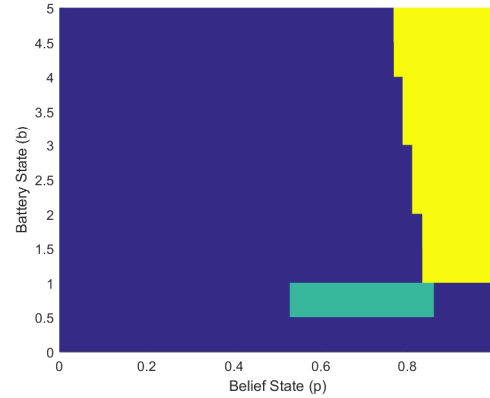


Fig. 3: Optimal thresholds for taking the actions D (blue), O (green), T (yellow) for $B_{max} = 5$, $\tau = 0.5$, $\beta = 0.98$, $\lambda_1 = 0.9$, $\lambda_0 = 0.6$, $R = 3$ and $q = 0.1$.

B. Throughput performance

In this section, we compare the performance of the optimal policy with two alternative policies, namely a greedy policy and a single-threshold policy. In the greedy policy, the transmitter transmits whenever it has energy in its battery. In the single-threshold policy there are only two actions: defer (D) or transmit (T). We optimize the threshold corresponding to each battery state for the single-threshold policy using the value iteration algorithm. By choosing the parameters $B_{max} = 5$, $\tau = 0.1$, $\beta = 0.999$, $\lambda_1 = 0.7$, $\lambda_0 = 0.2$, $R = 2$, the throughput achieved by these three policies are plotted in Fig. 4 with respect to the EH rate q .

As expected we observe that the greedy policy performs the worst as it does not exploit the transmitter's knowledge about the state of the channel. We can see that by simply exploiting the ACK/NACK feedback from the receiver, it is possible to achieve a higher throughput than the greedy policy for all values of the EH rate. On the other hand, by further introducing the channel sensing action the throughput

of the system is substantially increased. The improvement is particularly higher for the mid-range of q values, for which the transmitter benefits more from the flexibility offered by three actions.

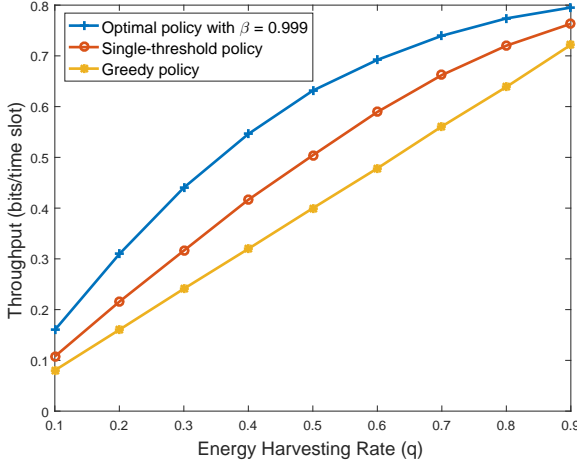


Fig. 4: Throughput comparison among the optimal, greedy, and single-threshold policies as a function of the EH rate, q .

VI. CONCLUSIONS AND FUTURE WORK

In this work we considered an EH transmitter equipped with a battery, operating over a time varying finite-capacity wireless channel with memory, modeled as a Gilbert-Elliot channel. The transmitter receives ACK/NACK feedback after each transmission, which can be used to track the channel. We further consider channel sensing, which the transmitter can use to learn the current channel state at a certain energy and time cost. Therefore, at the beginning of each time slot, the transmitter has three possible actions to maximize the total expected discounted number of bits transmitted over an infinite time horizon: i) deferring the transmission to save its energy for future use, ii) transmitting at a rate of R bits per time slot, and iii) sensing the channel to reveal the current channel state by consuming a portion of its energy and time, followed by transmission at a reduced rate consuming the remainder of the energy unit, only if the channel is in the good state. We formulated the problem as a POMDP, which is then converted into a MDP with continuous state space by introducing a belief parameter for the channel state. Then we proved that the optimal policy is a threshold policy, where the threshold values on the belief parameter depends on the battery state. We find the optimal threshold values numerically using the value iteration algorithm. In terms of throughput, we compared the optimal policy to the alternative policies, the greedy policy and a single-threshold policy which does not have channel sensing capability. We have shown through simulations that the channel sensing capability improves the performance significantly, thanks to the increased adaptability to the channel conditions it provides. For future studies, we will consider the case where the sensing is not perfect. Another interesting problem is to consider the case in which the EH

transmitter has the option to choose the duration of the sensing which determines its accuracy.

REFERENCES

- [1] J. A. Paradiso and T. Starner. Energy scavenging for mobile and wireless electronics. *IEEE Pervasive Computing*, 4(1):18–27, Jan. 2005.
- [2] G. Park, T. Rosing, M. D. Todd, C. R. Farrar, and W. Hodgkiss. Energy harvesting for structural health monitoring sensor networks. *Journal of Infrastructure Systems*, 14(1):64–79, Mar. 2008.
- [3] D. Gunduz, K. Stamatiou, N. Michelusi, and M. Zorzi. Designing intelligent energy harvesting communication systems. *IEEE Communications Magazine*, 52(1):210–216, Jan. 2014.
- [4] H. Li, C. Huang, P. Zhang, S. Cui, and J. Zhang. Distributed opportunistic scheduling for energy harvesting based wireless networks: A two-stage probing approach. *IEEE/ACM Transactions on Networking*, 24(3):1618–1631, Jun. 2016.
- [5] Q. Zhang and S. A. Kassam. Finite-state Markov model for Rayleigh fading channels. *IEEE Trans. on Commun.*, 47(11):1688–1692, Nov. 1999.
- [6] E. N. Gilbert. Capacity of a burst-noise channel. *The Bell System Technical Journal*, 39(5):1253–1265, Sep. 1960.
- [7] M. L. Ku, Y. Chen, and K. J. R. Liu. Data-driven stochastic models and policies for energy harvesting sensor communications. *IEEE Journal on Selected Areas in Communications*, 33(8):1505–1520, Aug. 2015.
- [8] N. Michelusi, K. Stamatiou, and M. Zorzi. Transmission policies for energy harvesting sensors with time-correlated energy supply. *IEEE Transactions on Communications*, 61(7):2988–3001, Jul. 2013.
- [9] A. Hentati, F. Abdelkefi, and W. Ajib. Energy allocation for sensing and transmission in WSNs with energy harvesting Tx/Rx. In *IEEE Vehicular Technology Conf. (VTC Fall)*, pages 1–5, Sep. 2015.
- [10] S. Mao, M. H. Cheung, and V. W. S. Wong. Joint energy allocation for sensing and transmission in rechargeable wireless sensor networks. *IEEE Transactions on Vehicular Technology*, 63(6):2862–2875, Jul. 2014.
- [11] P. Blasco and D. Gunduz. Multi-access communications with energy harvesting: A multi-armed bandit model and the optimality of the myopic policy. *IEEE Journal on Selected Areas in Communications*, 33(3):585–597, Mar. 2015.
- [12] C. K. Ho and R. Zhang. Optimal energy allocation for wireless communications with energy harvesting constraints. *IEEE Transactions on Signal Processing*, 60(9):4808–4818, Sep. 2012.
- [13] B. T. Bacinoglu and E. Uysal-Biyikoglu. Finite-horizon online transmission scheduling on an energy harvesting communication link with a discrete set of rates. *Journal of Communications and Networks*, 16(3):393–300, Jun. 2014.
- [14] M. S. H. Abad, D. Gunduz, and O. Ercetin. Energy harvesting wireless networks with correlated energy sources. In *2016 IEEE Wireless Communications and Networking Conference*, pages 1–6, Apr. 2016.
- [15] M. Kashef and A. Ephremides. Optimal packet scheduling for energy harvesting sources on time varying wireless channels. *Journal of Communications and Networks*, 14(2):121–129, Apr. 2012.
- [16] A. Aprem, C. R. Murthy, and N. B. Mehta. Transmit power control policies for energy harvesting sensors with retransmissions. *IEEE Journal of Selected Topics in Signal Processing*, 7(5):895–906, Oct. 2013.
- [17] Q. Zhao, L. Tong, A. Swami, and Y. Chen. Decentralized cognitive MAC for opportunistic spectrum access in ad hoc networks: A POMDP framework. *IEEE Journal on Selected Areas in Communications*, 25(3):589–600, Apr. 2007.
- [18] A. Laourine and L. Tong. Betting on Gilbert-Elliot channels. *IEEE Transactions on Wireless Communications*, 9(2):723–733, Feb. 2010.
- [19] William S. Lovejoy. A survey of algorithmic methods for partially observed Markov decision processes. *Annals of Operations Research*, 28(1):47–65, Dec. 1991.
- [20] P. Blasco, D. Gunduz, and M. Dohler. A learning theoretic approach to energy harvesting communication system optimization. *IEEE Transactions on Wireless Communications*, 12(4):1872–1882, Apr. 2013.
- [21] M. L. Puterman. *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. John Wiley & Sons, Inc., New York, NY, USA, 1st edition, 1994.
- [22] M.S.H. Abad, D. Gunduz, and O. Ercetin. Channel sensing and communication over a time-correlated channel with an energy harvesting transmitter. *arXiv:1703.10519 [cs.IT]*, Mar. 2017.