Transmit Power Control Policies for Energy Harvesting Sensors with Retransmissions

Anup Aprem, Chandra R. Murthy, *Senior Member, IEEE*, and Neelesh B. Mehta, *Senior Member, IEEE*

***Abstract*—This paper addresses the problem of finding outage-optimal power control policies for wireless energy harvesting sensor (EHS) nodes with automatic repeat request (ARQ)-based packet transmissions. The power control policy of the EHS specifies the transmission power for each packet transmission attempt, based on all the information available at the EHS. In particular, the acknowledgement (ACK) or negative acknowledgement (NACK) messages received provide the EHS with partial information about the channel state. We solve the problem of finding an optimal power control policy by casting it as a partially observable Markov decision process (POMDP). We study the structure of the optimal power policy in two ways. First, for the special case of binary power levels at the EHS, we show that the optimal policy for the underlying Markov decision process (MDP) when the channel state is observable is a simple threshold policy in the battery state. Second, we benchmark the performance of the EHS by rigorously analyzing the outage probability of a general fixed-power transmission scheme, where the EHS uses a predetermined power level at each slot within the frame. Monte Carlo simulation results illustrate the performance of the POMDP approach and verify the accuracy of the analysis. They also show that the POMDP solutions can significantly outperform conventional ad hoc approaches.**

***Index Terms*— Energy harvesting sensors, power control, ARQ, retransmission, POMDP.**

I. Introduction

Wireless energy harvesting sensors (EHS) operate using energy harvested from environmental sources such as the sun, wind, vibrations, etc. Due to their promise of a potentially infinite lifetime, they are fast emerging as viable options for sensing-related applications ranging from inventory management and surveillance to structural health monitoring in buildings, bridges, and vehicles. However, due to the sporadic and random nature of the harvesting process, energy management becomes critical to ensure continuous and reliable operation of these nodes. The energy replenishment process of the natural phenomena, the time-varying nature of the wireless channel and the energy storage constraints of the node all need to be taken into consideration when designing efficient transmission strategies. Further, transmission policies for EHS need to satisfy the constraint of *energy neutrality*, which mandates that at every point in time, the cumulative energy consumed by a node cannot exceed the cumulative energy harvested by it. These requirements call for a cross-layer, energy-aware protocol design that optimizes energy consumption for reliable packet transmission. This, in turn, motivates the use of an automatic repeat request (ARQ) protocol, as it is resilient to channel variations and is known to be energy efficient [1]. Moreover, packet retransmissions and power control are already enabled in present-day low power communication standards such as the IEEE 802.15.4 [2], which makes schemes based on them readily suitable for implementation.

Transmit power management in EHS nodes has been studied in [3], with deterministic energy harvesting models. A Bernoulli injection model, in which the node harvests a fixed amount of energy with some probability or does not harvest at all, was proposed in [4]. Several different performance metrics for EHS have been considered in the literature, including minimizing the transmission time [5], maximizing the short-term throughput [6] or the quality of coverage [7], and throughput-optimal and delay-optimal policies [8]. The problem of power management was formulated as a Markov decision process (MDP) in [7], [9], but these works either ignored the channel variability [7] or considered it as perfectly known [9]. Further, none of the aforementioned works considered the outage probability as the performance metric, or exploited the implicit channel state information (CSI) at the transmitter available through the link-layer ARQ feedback messages from the destination, in designing the power control policies.

In this work, we consider an EHS node that transmits packets to a destination using an ARQ-based packet transmission scheme. In each *frame*, the EHS can make up to K attempts to transmit a packet. After every attempt, the node receives an acknowledgement (ACK) or negative acknowledgement (NACK) message, depending on whether the packet was successfully received or not by the destination. The ACK/NACK messages are assumed to be received without error at the EHS. In case a NACK is received, it retransmits the packet, provided it has enough energy to do so. If the packet is not successfully delivered by the end of the frame, it is declared to be in outage. The outage probability, which is defined as the average fraction of packets that suffer an outage, is thus an important performance metric in ARQ-based transmission systems. The ARQ-based transmission is pertinent in systems where the node makes periodic measurements, and where, once a new measurement is taken by the node, the old measurement is no longer relevant. This is the case, for example, in air temperature sensors [10] and environmental monitoring sensors [11].

The goal here is to select the power level for each packet

A. Aprem is currently working at Analog Devices, Inc. He was at the Dept. of ECE, Indian Institute of Science (IISc), Bangalore, during the course of this work. C. R. Murthy and N. B. Mehta are with the Dept. of ECE, IISc, Bangalore, India. Emails: anupaprem@gmail.com, {cmurthy, nbmehta}@ece.iisc.ernet.in.

This work was supported in part by a research grant from the Aerospace Network Research Consortium and in part by research grant no. 4900-IT-B funded by the Indo-French Centre for the Promotion of Advanced Research.

transmission attempt, based on the current and past information available at the EHS. For example, one option is to transmit at low power at the start of the frame, and successively increase the transmit power each time a NACK is received. In case the channel happens to be in a good state, the initial attempt would succeed, saving power for future transmissions. On the other hand, since failed transmission attempts are a waste of energy, one could transmit at a high power in the first attempt, and, in case a NACK is received, choose to incur an outage by not transmitting and thereby saving the power harvested over the rest of the frame for future transmissions. At first glance, it is not intuitively obvious as to which of these options would offer a better outage probability performance. Therefore, a systematic approach is essential.

In this paper, we address the problem of transmit power control with retransmissions by formulating it as one of making optimal sequential decisions. The ACK/NACK feedback messages implicitly provide the EHS with partial CSI, which can be exploited in deciding the transmit power level for subsequent transmission attempts. Since the CSI is only partially available at the EHS, we cast the problem as a partially observable Markov decision process (POMDP) [12]. Our focus on ARQ-based packet transmission and on exploiting the resulting partial observability of the channel using the POMDP framework makes our study fundamentally different from the past work that employs decision theory to design transmission policies for EHS [7]–[9], [13], [14].

The rigorous formulation shows how to optimally handle the following key trade-off that arises in the design of ARQ protocols for EHS nodes. Increasing the transmission power improves the odds of successful packet reception, but drains energy from the battery and decreases the probability that there will be sufficient energy to deliver future packets. On the other hand, a conservative approach of transmitting at a low or minimal power could lead to packet outages and wastage of energy if energy arrivals continue to occur after the battery gets full. Further, the history of transmit powers and corresponding ACK/NACK messages, the time correlation of the channel, the statistics of the energy arrival process, and the battery capacity. can be utilized in choosing the transmit power. The POMDP framework uses all of the available past information to optimally manage the power available at the EHS, subject to the constraint of energy neutrality. To the best of our knowledge, this is the first time in the literature that power management in EHS with ARQ-based packet retransmission has been studied in the POMDP framework.

The main advantage of the POMDP formulation is that it allows one to choose from a gamut of available techniques for finding an optimal and also near-optimal energy management policies. Due to the large size of the state space, the exact solution to the POMDP at hand turns out to be computationally infeasible. Further, our problem differs from the classical POMDP due to the mixed observability of the state process. The CSI is partially observable through ACK/NACK messages, whereas the battery energy level and transmission index which also form part of the state description, are fully observable. Hence, we adapt two popular and computationally efficient suboptimal solutions to the POMDP, namely, the voting policy and the ML-estimation policy, to design the power control policy for the EHS.

To gain further insights about the proposed solution, we study the structure of the underlying MDP that results when the channel is fully observable. When the EHS is restricted to employ a binary power control policy, we show that the optimal policy is a threshold policy in the battery energy state, i.e., the EHS transmits if and only if the battery energy level exceeds a threshold. This not only reduces the search complexity for finding the optimum power policy, but also provides easily implementable policies.

Finally, we benchmark the proposed solutions against other solutions proposed in the literature [15]. To this end, we rigorously analyze the outage probability performance of these approaches under a quasi-static, block fading channel. The analysis generalizes that in [15] by allowing the EHS to employ arbitrary power levels for each retransmission of a given packet. The theoretical expressions are useful in understanding the critical dependence of the outage performance on the power control policy. Through simulations, we illustrate the superior performance of the POMDP over fixed-power policies, as they optimally tune the power control policy as a function of the current state and the past information available at the EHS. For example, to achieve the same outage probability, the POMDP solution requires only about 60-80% of the energy harvesting rate required by the scheme in [15].

The organization of the paper is as follows. Section II introduces the system model and the problem definition. Section III presents the POMDP formulation of the power management problem. Section IV discusses the approximate solutions of the POMDP. Section V analyzes the outage probability performance of the EHS with ARQ transmissions and a given power control policy. Section VI presents simulation results, and concluding remarks are offered in Section VII.

II. System Model

Consider an EHS node that wishes to send a measurement packet of size ℓ bits periodically, in a frame of duration T_m s, to a destination, over a time-varying wireless channel. Each packet transmission attempt happens during a *slot* of duration T_p s, which includes the time for sending the packet and receiving an ACK from the destination. Hence, the node can make at most $K \triangleq \lfloor T_m/T_p \rfloor$ attempts to transmit the packet within the frame, where $\lfloor \cdot \rfloor$ denotes the floor function. If the EHS is unable to deliver the packet within the frame duration, a measurement *outage* occurs.

In a slot, the receiver may fail to decode the packet if the EHS node does not have sufficient energy to transmit, or if the transmitted packet is corrupted by the channel or noise at the receiver. The following ARQ protocol is assumed at the link layer. If the EHS receives a NACK from the receiver, it retransmits its packet until it receives an ACK, or it runs out of energy, or it is time to transmit the next packet. If it receives an ACK, the node stops transmitting and just accumulates the harvested energy during the rest of the frame. A finite energy buffer (e.g., a battery) is used to store the harvested energy, and it is assumed that there are no storage inefficiencies in

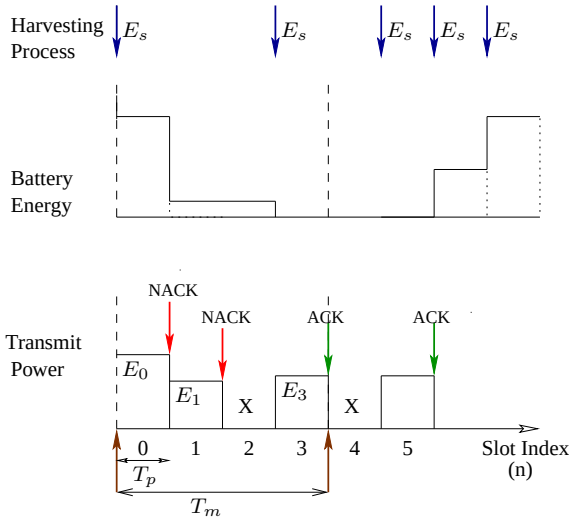


Fig. 1. Transmission timeline of the EHS node for $K = 4$, showing the random energy harvesting process $(\downarrow)$ and periodic data arrivals $(\uparrow)$. The marker "X" denotes slots where the EHS does not transmit. By convention, in a given slot, both the energy harvested and the energy already present in the battery are assumed to be available for data transmission by the EHS. For example, in slot 3, the EHS harvests energy E_s; this energy, along with the energy in the battery at the end of slot 2, equals the energy E_3 used by the node for transmitting the data packet. Hence, the EHS is left with no energy in the battery at the end of slot 3.

the buffer. Apart from the ACK/NACK, no CSI is assumed to be available at the EHS. Figure 1 illustrates the timeline of events, showing the sporadic energy injections, battery energy evolution, and the packet transmission powers.

For the energy harvesting process, an independent and identically distributed (i.i.d.) Bernoulli model is considered, in which an energy E_s is injected into the EHS node at the beginning of every slot with probability ρ, and with probability $1-\rho$, no energy is harvested. This model is motivated by switch-based harvesting mechanisms [4]; a similar model for energy recharging was also considered in [16]. Other models for the energy harvesting process include the leaky-bucket model [3] and the Markov model [17], etc. While the Bernoulli model is simple, it does capture the sporadic and random nature of energy availability at the EHS and facilitates analysis.

Let B_n denote the battery energy level at the beginning of the n^{th} slot, and let $E_n \leq B_n$ denote the energy used for packet transmission. The battery energy itself gets replenished whenever the node harvests energy, and, consequently, obeys the following Markovian evolution:

$$B_{n+1} = \begin{cases} \min\left(B_n + E_s - E_n, B_{\max}\right), & \text{with probability } \rho \\ B_n - E_n, & \text{with probability } 1-\rho \end{cases}$$

where $B_{\max}$ denotes the battery capacity. The next section presents the outage analysis of the EHS, for a power policy specified as a function of transmission index. In the remainder of this paper, we normalize all energies with respect to a minimum possible transmit energy E, which is typically imposed by the lower end of the linearity range of RF amplifier on the EHS node. Thus, the battery energy level is considered

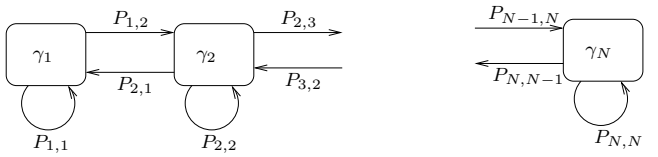


Fig. 2. Finite state Markov chain model for the Rayleigh fading channel.

to be an integer multiple of E. Further, we let $L \triangleq E_s/E$ be an integer. That is, each time an energy injection occurs, the EHS obtains sufficient energy to make L packet transmission attempts at the lowest possible transmit energy level.

III. POMDP Formulation

Our goal in this section is to sequentially decide on the optimum packet transmit power levels, $\{E_0, E_1, \dots\}$, based on the transmission index, battery energy level, and the history of transmission energies and ACK/NACK messages received, to minimize the long term expected outage probability.

The POMDP formulation naturally equips the EHS to exploit the time correlation in the wireless channel, and, hence, we consider two channel models: the correlated channel model and the block fading channel model. In both cases, we need the state space to be finite. To facilitate this, we discretize the channel into N levels, $\gamma_1, \dots, \gamma_N$. In the *correlated channel model*, the channel is modeled as the finite state Markov chain (FSMC) shown in Fig. 2, with known channel transition probabilities[1] $P_{i,i+1}, P_{i,i-1}$ and $P_{i,i}$. Such a first-order model is known to be accurate for packet-level studies [18] and in cross-layer optimization with slowly-varying channels [19], [20]. The channel levels $\gamma_1, \dots, \gamma_N$ and the transition probabilities can be computed based on the underlying fading distribution and Doppler frequency, following the procedure in [21], [22]. In the *block fading channel model*, the channel is assumed to remain fixed for the duration of a frame, and changes independently and identically from one frame to the next. This assumption is valid when the channel coherence time equals the frame duration, due to which, the initial packet transmission and all of the retransmission attempts see the same channel state [15]. To facilitate comparisons between the two channel models, we assume that the stationary distribution of the quantized channel is the same for both the correlated and block fading models.

Recall that the channel state in a given slot is only partially available at the EHS transmitter through the ACK/NACK messages. Hence, we cast the sequential decision problem in a POMDP framework. The POMDP formulation consists of the following components:

a) State Space: The finite set of states denoted by $\mathcal{S} \triangleq \mathcal{B} \times \mathcal{G} \times \mathcal{K} \times \mathcal{U}$, where

- $\mathcal{B} \triangleq \{0, 1, \dots, B_{\max}\}$ is the set of battery states, normalized with respect to the minimum transmit energy E. Recall that $B_{\max}$ is the battery capacity.

[1]In this paper, for convenience, we use the notations $P_{i,j}$ and P_{γ_i,γ_j} interchangeably, to represent the probability of going from channel state γ_i to γ_j.

- $\mathcal{G}$ is the set of discrete channel states. Under the channel model explained above, $\mathcal{G} \triangleq \{\gamma_1, \gamma_2, \ldots, \gamma_N\}$.
- $\mathcal{K}$ is the set of possible packet transmission attempt indices within a frame. Since the EHS can make at most K attempts in a frame, $\mathcal{K} \triangleq \{0, 1, \ldots, K-1\}$. In the sequel, we use $k \in \mathcal{K}$ to index the slot count within a frame, and $n \in \{1, 2, \ldots\}$ as an absolute index for the slot count, which increments indefinitely with time.
- $\mathcal{U} \triangleq \{0, 1\}$ is the set of *packet reception states*. The packet reception state takes the value 1 when an ACK is received by the EHS, and is 0 otherwise. At the beginning of the frame, i.e., when the packet transmission attempt index k is 0, the EHS node is always in packet reception state 0, since an ACK has not yet been received. If the receiver successfully decodes the current packet and an ACK is received by the EHS, the packet reception state changes to 1 for the rest of the frame. Irrespective of the system state at $k = K-1$, the packet reception state is reset to 0 at the next slot, as it corresponds to the beginning of a new frame.

b) Observation Space: The observations are the ACK/NACK messages received by the EHS node after each packet transmission attempt. The observation space is the finite set $\mathcal{O} \triangleq \{\text{ACK}, \text{NACK}\}$. Since the ACK/NACK messages are assumed to be received error-free, the observations always match the packet reception state of the system.

c) Action Space: An *action* a by the EHS node corresponds to sending a packet at power level aE. The action space is the set of possible actions, and is denoted by $\mathcal{A} \triangleq \{0, 1, \ldots, B\}$, with $B \in \mathcal{B}$ representing the battery level in the current slot.

d) State Transition Function: Let two arbitrary states in $\mathcal{S}$ be $s \triangleq (b, \gamma, k, u)$ and $s' \triangleq (b', \gamma', k', u')$. The state transition function is the probability that the system starts in state s, takes an action a, and lands in state s'. Since the channel state evolution is independent of the packet reception errors at the receiver, the state transition function depends on the product of the probability of the channel state transiting from γ to γ' and the probability that the battery and packet reception state transits from (b, u) to (b', u') when the EHS node takes action a and the channel state is γ. Under the correlated fading channel model, it is given by

$$\mathcal{T}(s, a, s') = \delta(k', k_+)\, P_{\gamma,\gamma'}\, \psi((b, u), a, (b', u'), k, \gamma) \quad (1)$$

where $k_+ \triangleq (k+1) \bmod K$, $\delta(k', k)$ is the Kronecker delta function, and $P_{\gamma,\gamma'}$ is the channel transition probability, as defined earlier in this section. Here, the term $\delta(k', k_+)$ captures the fact that the packet transmission index always increases one at a time until the end of the frame, where it resets to 0. Also, $\psi((b, u), a, (b', u'), k, \gamma)$ represents the probability that the EHS node starts from battery state b and packet reception state u, takes an action a, and lands in the state (b', u') when the current channel state and packet transmission index are γ and k, respectively. It is given as follows. Let

$$\eta(b, a, b') \triangleq \rho\delta(b', b + L - a) + (1 - \rho)\delta(b', b - a).$$

When $k = K-1$, it can be shown that

$$\psi((b, u), a, (b', u'), K-1, \gamma) = \begin{cases} \eta(b, a, b'), & u' = 0 \\ 0, & \text{else.} \end{cases} \quad (2)$$

Also, when $k \neq K-1$, it can be shown that

$$\psi((b, u), a, (b', u'), k, \gamma) = \begin{cases} \eta(b, a, b'), & u' = 1, u = 1 \\ \eta(b, a, b')(1 - P_e(\gamma, aE)), & u' = 1, u = 0 \\ \eta(b, a, b')P_e(\gamma, aE), & u' = 0, u = 0 \\ 0, & \text{else.} \end{cases} \quad (3)$$

Here, $P_e(\gamma, aE)$ is the probability that a packet transmitted at power aE will be received in error when the channel state is γ. This probability depends on the modulation and coding scheme used. For example, with uncoded binary phase shift keying (BPSK) transmission, the packet error probability is given by

$$P_e(\gamma, aE) = 1 - \left(1 - Q\left(\sqrt{\frac{2\gamma aE}{N_0}}\right)\right)^{\ell}, \quad (4)$$

where $Q(\cdot)$ is the Gaussian tail function, ℓ is the packet size in bits, and N_0 is the noise power spectral density.

The expression in (3) is obtained by tracking the probabilities of the following events: i) Whether energy has been harvested in the current slot or not; ii) The packet reception state of the system; and iii) The probability of successful packet reception given the channel state and action. Also, the expression in the $k = K-1$ case arises because the packet reception state always resets to zero at the end of the frame. For example, to get the first term in (3), note that, when the EHS has already received an ACK, it transits from battery state b to $b + L - a$ upon taking an action[2] a if it harvests energy (which occurs with probability ρ). If it does not harvest energy (which occurs with probability $1 - \rho$), it transits to the state $b - a$.

Under the block fading model, $\mathcal{T}(s, a, s')$ is given by

$$\mathcal{T}(s, a, s') = \delta(k', k_+)\, \zeta(\gamma, \gamma'; k)\, \psi((b, u), a, (b', u'), k, \gamma), \quad (5)$$

$$\text{with } \zeta(\gamma, \gamma'; k) = \begin{cases} \delta(\gamma, \gamma'), & k \neq K-1 \\ \pi_{\gamma'}, & \text{else.} \end{cases} \quad (6)$$

Here, $\pi_{\gamma'}$ represents the stationary probability of the channel state γ'. In contrast with the correlated fading channel, the $\zeta(\gamma, \gamma'; k)$ term in the above equation captures the fact that the channel remains constant for the duration of a frame and transitions in an i.i.d. fashion from one frame to the next.

e) Observation Function: The observation function is the probability of observing an ACK or a NACK given the current state and action. Since this probability depends only on the current channel state and action, it is given by

$$\begin{aligned} P(\text{NACK}|a, \gamma) &= P_e(\gamma, aE) \\ P(\text{ACK}|a, \gamma) &= 1 - P_e(\gamma, aE) \end{aligned} \quad (7)$$

[2]Note that, in this example, a obviously equals 0 since the EHS has already received an ACK. An arbitrary a is incorporated here for the sake of generality in the expression.

where P_e is packet error probability as defined in (4).

f) Cost: Let $s \triangleq (b, \gamma, k, u)$ be the state of the system. The expected immediate cost is defined as

$$c(s,a) = \begin{cases} P_e(\gamma, aE), & a \leq b, u = 0 \\ 1, & (a > b, u = 0) \text{ or } (a \neq 0, u = 1) \\ 0, & \text{else.} \end{cases} \tag{8}$$

The immediate cost of 1 is used to preempt the EHS from using a nonzero energy to transmit a packet when it has already received an ACK, or from attempting to use more than the energy available in the battery.

g) Objective: A power control *policy* π describes a decision rule that determines the action taken by the EHS. It maps the history of actions and observations to the action to be taken in the current time slot. The goal here is to find a policy that minimizes the expected cost incurred by the EHS node over an infinite time horizon, which is given by

$$J_\pi(s_0) = \lim_{m\to\infty} \frac{1}{m} \mathbb{E}_\pi \left[\sum_{n=1}^{m} c(S_n, a_n) | S_0 = s_0 \right] \tag{9}$$

where $n \in \{1, 2, \dots\}$ denotes the slot index, S_n is the state sequence, a_n is the action sequence, and s_0 is the initial state. The expectation in (9) is over the distribution of the state sequence S_n. The next section discusses the techniques for solving the POMDP considered in this work.

IV. Solution Techniques

Recall that in the POMDP formulation above, the system state s is not fully observable. In particular, the channel state component of s is unknown to the EHS, while the battery state, slot count within the frame, and the packet reception state are observable. However, given the history of actions and observations, a so-called *belief state* $\beta(\gamma)$ can be computed, that represents the probability that the channel is in state γ. It is known that the belief state is a sufficient statistic for finding the optimal policy [23]. The belief state can be updated at the end of each slot, based on the previous belief state, the current observation, and the state transition function. Then, the solution to the POMDP can be found as the solution of a fully observable Markov decision process (MDP) on the belief states [12]. Although several exact algorithms [12] for solving the belief MDP exist, these algorithms are computationally feasible only when the cardinality of the state space is of the order of ten [24], [25]. Even approximate solution methods can only handle a state space with cardinality of about a hundred [26]. In our case, the state space is much larger, as it is indexed by the number of battery energy levels, the number of channel states, the packet transmission index, and the ACK/NACK state. As a result, finding an exact or even an approximate solution to the POMDP is computationally infeasible. Hence, in this paper, we explore two computationally efficient suboptimal solutions for the POMDP. For this, we first describe the solution to the MDP that is obtained when the system state is fully observable. Both the solution techniques we investigate rely on solving this underlying MDP.

The solution to the MDP when the system state is fully observable yields an optimal policy μ^*_{MDP} that maximizes the expected long-term reward defined in (9). The optimal policy is the solution to the following Bellman equation [27]:

$$\lambda^* + h^*(s) = \min_{a\in\mathcal{A}, a\leq B(s)} \left[c(s,a) + \sum_{s'\in\mathcal{S}} \mathcal{T}(s,a,s') h^*(s') \right] \tag{10}$$

for all $s \in \mathcal{S}$, where λ^* is the optimal average cost and $h^*(s)$ is the optimal differential cost when starting at state s. Here, with a slight abuse of notation, the battery energy level is written as $B(s)$ to indicate that it is one of the components of the system state s.

The value iteration method [12] can be used to solve the Bellman equation (10). This involves iteratively solving

$$J_{k+1}(s) = \min_{a\in\mathcal{A}, a\leq B(s)} \left[c(s,a) + \sum_{s'\in\mathcal{S}} \mathcal{T}(s,a,s') J_k(s') \right] \tag{11}$$

for all $s \in \mathcal{S}$, where J_k is the value function at the k^{th} iteration, $k = 0, 1, \dots$. It can be shown that [27]

$$\lim_{k\to\infty} \frac{J_k(s)}{k} = \lambda^*, \quad \forall s \in \mathcal{S}. \tag{12}$$

In practice, it is standard to use relative value iteration to solve the MDP, which is a numerically stable version of the above procedure. We refer the interested reader to [27] for a comprehensive treatment of relative value iteration and its convergence properties. The convergence is guaranteed, provided that one of the states is visited with positive probability at least once within the first m slots, for some integer m, for all initial states and all possible policies. In our problem, this requirement is trivially satisfied, and all states are reachable from any given state in a finite number of steps, for all possible policies. We have found that this algorithm converges reliably and quickly for the problem at hand. Upon convergence, we obtain the optimal action as the argument that minimizes the right hand side in (11). We denote the solution to the MDP obtained using relative value iteration as $\mu^*_{\text{MDP}}(s)$.

Next, we present a structural property of the solution $\mu^*_{\text{MDP}}(s)$. Establishing structural properties not only provides useful insights into the form of the solution, but more importantly, helps in reducing the computational complexity of finding the optimal solution.

A. Structure of the MDP Solution

In order to study the structural properties of the MDP, we lean on the theory of discounted cost MDPs. We exploit the fact that the average cost MDP under study in this paper is the limit of a sequence of discounted cost MDPs, with the discount factor $\nu \to 1$ [28], [29]. The discounted long term cost associated with policy π and discount factor $\nu \in (0,1)$ is given by

$$J^\nu_\pi(s_0) = \limsup_{m\to\infty} \mathbb{E}_\pi \left[\sum_{n=1}^{m} \nu^{n-1} c(S_n, a_n) \middle| S_0 = s_0 \right]. \tag{13}$$

The optimal discounted cost function, $V^\nu(s_0)$, is given by

$$V^\nu(s_0) = \inf_{\pi\in\mathbf{\Pi}_{\text{sd}}} J^\nu_\pi(s_0), \tag{14}$$

where $\mathbf{\Pi}_{\text{sd}}$ is the set of all stationary deterministic policies. It is known that for all discrete state MDPs with bounded cost, an optimal stationary deterministic policy exists [30]. The optimal discounted cost $V^{\nu}(s)$ satisfies the Bellman equation for optimality, which is given by

$$V^{\nu}(s) = \min_{a\in\mathcal{A}, a\leq B(s)} \left\{ c(s,a) + \nu \sum_{s'\in\mathcal{S}} \mathcal{T}(s,a,s') V^{\nu}(s') \right\}. \tag{15}$$

The optimal action a^*_{ν} is the action that satisfies (15). The discounted cost MDP can be solved using the value iteration algorithm given by

$$V^{\nu}_{n+1}(s) = \min_{a\in\mathcal{A}, a\leq B(s)} \left\{ c(s,a) + \nu \sum_{s'\in\mathcal{S}} \mathcal{T}(s,a,s') V^{\nu}_{n}(s') \right\}. \tag{16}$$

Let

$$Q^{\nu}_{n+1}(s,a) \triangleq c(s,a) + \nu \sum_{s'\in\mathcal{S}} \mathcal{T}(s,a,s') V^{\nu}_{n}(s'). \tag{17}$$

The threshold structure of discounted cost optimal policies can now be established for the important special case of binary actions, i.e., $\mathcal{A} = \{0,1\}$. Binary actions correspond to on-off power control at the EHS, where the node decides whether or not to transmit a packet based on the history of packet attempts and corresponding ACK/NACK observations. This is summarized in the following theorem. The proof, which is shown in Appendix A, shows that $Q^{\nu}_{\infty}([b,\gamma,u,k],a)$ is submodular in (b,a), for some initial condition. This is sufficient to establish that the discounted cost optimal policy is a threshold policy in the energy buffer state b, due to the convergence of (16) for all initial conditions [31].

Theorem 1. *When the action set is binary, i.e., $\mathcal{A} = \{0,1\}$, for any discount factor $\nu \in (0,1)$, the optimal policy is a threshold policy in the energy buffer state. That is, given the channel gain γ, transmission index k, and packet reception state u, there is a threshold $B_{thresh}(\gamma,k,u)$ that defines the optimal policy: the node transmits the packet if and only if $B > B_{thresh}(\gamma,k,u)$, where B is the current battery level.*

The simplicity of the threshold policies makes them attractive for implementation on power and hardware resource starved EHS nodes. Determining the optimal threshold requires solving the underlying MDP, but this is an offline search that does not impose a computational burden on the EHS. Moreover, the fact that the optimal policy is a threshold policy in the battery state can be used to reduce the computational complexity in finding the optimal policy.

We next discuss two computationally efficient suboptimal solutions to the original POMDP.

B. Solution of the POMDP

Recall that, in our system model, the battery state, packet reception state, and the transmission attempt index are fully observable, while the channel state component of the system state is only partially observable through the ACK/NACK messages. Therefore, it is sufficient to maintain the belief over only the channel state component of the system state, denoted by $\beta(\gamma)$, and use it to approximately solve the POMDP. When new observations are obtained, $\beta(\gamma)$ is updated as follows.

Let $o_n \in \mathcal{O}$ be the observation at time slot n, and let[3] $\beta_n(\gamma_j)$ denote the belief that the channel state at time slot n equals γ_j, given the history $\mathcal{F}_{n-1}$ of actions and observations up to time $n-1$. That is, $\beta_n(\gamma_j) \triangleq P\{\text{channel state at time slot } n = \gamma_j | \mathcal{F}_{n-1}\}$. In the correlated fading model, the belief state $\beta_n(\gamma_j)$ can be updated using Bayes' rule as

$$\beta_n(\gamma_j) = \frac{\sum_i P_{\gamma_i,\gamma_j} P(o_{n-1}|a_{n-1},\gamma_i)\beta_{n-1}(\gamma_i)}{\sum_l \sum_i P_{\gamma_i,\gamma_l} P(o_{n-1}|a_{n-1},\gamma_i)\beta_{n-1}(\gamma_i)} \tag{18}$$

for $j = 1,2,\ldots,N$, where $P(o|a,\gamma)$ is given by (7). In the block fading model, P_{γ_i,γ_j} in the above is replaced by $\zeta(\gamma_i,\gamma_j;k)$ as defined in (6), where k is the transmission index of the packet within the current frame.

The final task is to use the belief state of the channel $\beta_n(\gamma_j)$ obtained above to convert the POMDP to an MDP, and use the solution of the MDP as an approximate solution for the POMDP. To this end, we consider two popular computationally efficient approaches.

1) Maximum Likelihood (ML) Heuristic [32]: Here, at each slot n, we find the most probable channel state, $\gamma_{\text{ML}} \triangleq \arg\max_{\gamma\in\mathcal{G}} \beta_n(\gamma)$, of the system. Then, the ML state of the system is defined as $s_{\text{ML}} \triangleq (b,\gamma_{\text{ML}},k,u)$, where b, k and u are the current battery, packet transmission attempt index and packet reception state, respectively. The ML heuristic method adopts the action corresponding to the solution of the MDP with the ML state as the solution of the POMDP. Thus,

$$\mu_{\text{ML}} = \mu^*_{\text{MDP}}(s_{\text{ML}}), \tag{19}$$

where, as mentioned earlier, $\mu^*_{\text{MDP}}(s)$ denotes the solution to the MDP obtained by solving (10).

2) Voting Heuristic [33]: Here, for a given b, u, and k, we consider the set of states (b,γ,k,u), for $\gamma\in\mathcal{G}$. A given state in this set, say state s, votes for an action a, as determined by the optimal policy $\mu^*_{\text{MDP}}(s)$ of the underlying MDP corresponding to that particular state. In any given time slot n, these votes are weighed by the component of the belief state corresponding to each $\gamma\in\mathcal{G}$, and the sum of the weighted votes for each action is determined. The action with the largest sum, denoted by μ_{voting}, is selected as the optimal action:

$$\mu_{\text{voting}} = \arg\max_{a\in\mathcal{A}} \sum_{\substack{s=(b,\gamma,k,u)\\ \gamma\in\mathcal{G}}} \beta_n(\gamma)\delta(\mu^*_{\text{MDP}}(s),a), \tag{20}$$

where, as before, $\delta(x,a)$ is the Kronecker delta function.

Note that, in order to implement the above policies, the main computational burden on the EHS is in updating the belief state using (18). Further, based on the belief state, the node needs to find the ML state (eq. (19)) or the action with the highest number of votes (eq. (20)). The MDP policy itself can be implemented as a simple look-up table, and hence

[3]Note that we use $\beta(\gamma)$ to denote the current belief state of the channel and $\beta_n(\gamma)$ to indicate the belief state of the channel at time slot n.

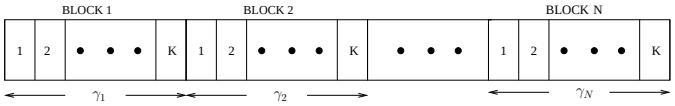


Fig. 3. The block fading channel model.

does not impose a significant computational burden on the EHS. Precisely accounting for the energy cost involved in the computations and storage is beyond the scope of our work.

V. Outage Probability Analysis

In this section, we consider the outage analysis of the EHS for the block fading channel model. Here, the channel remains constant throughout the frame, but changes in an i.i.d. fashion at the beginning of the next frame, as shown in Figure 3. In this case, the *power control policy* specifies the packet transmission power as a function of the transmission index and the current battery state. In the foregoing analysis, we derive the outage probability of a given fixed-power control policy, which not only generalizes the result in [15], but also serves as an important performance benchmark.

We consider the following general class of power control policies. In the first slot of a given frame, the EHS sends the packet at a power L_1E if the current battery state B_1 exceeds L_1, otherwise, is sends the packet at a power B_1. In general, at the k^{th} attempt to transmit a given packet, the EHS sends the packet at a power L_kE if the current battery state B_k exceeds L_k; otherwise, it sends it at a power B_k. This is a generalization of the transmit power policy considered in [15], as setting $L_1 = L_2 = \cdots = L_K$ results in the single-power transmission scheme considered there.

With the above power control scheme, the evolution of the system is a Discrete Time Markov Chain (DTMC), as shown in Figure 4. The state at time n is denoted by (U_n, B_n), where B_n is the battery state, as before, and U_n represents the *ACK state*. The ACK state of the EHS node is defined as

$$U_n \triangleq \begin{cases} -1, & \text{ACK received,} \\ 0, & \text{Start of transmission,} \\ i, & i \text{ NACKs received, } i \in \{1, \cdots, K\}. \end{cases} \quad (21)$$

In this section, we assume that the battery capacity is infinite, so $B_n \in \{0, 1, \cdots, \infty\}$. However, the extension to the case of a finite battery capacity is straightforward. As shown in Figure 4, the EHS does not transmit in the frame when $U_n = -1$, as an ACK has been received. Under mild conditions on the transmit power policy, it is easy to show that the DTMC is irreducible and positive recurrent. Now, in state (i, B_n), the packet transmission power is given by

$$P_{(i,B_n)} = \min(L_iE, B_n). \quad (22)$$

The m^{th} frame consists of slots $mK,\ mK+1, \ldots,\ (m+1)K-1$. The ACK state is $U = 0$ at the start of the frame. At the end of each slot, U is incremented by 1 if a NACK is received, otherwise it transits to $U = -1$ for the rest of the frame (i.e., when an ACK is received). Irrespective of the value of U at the end of the frame, it always resets to 0 at the start of the $(m+1)^{\text{th}}$ frame. An outage occurs in the m^{th} frame if the packet has not yet been successfully received even at the end of the frame. Since B_{mK} is independent of the channel state in $[mK, (m+1)K]$, the outage probability as a function of $K \geq 1$ can be written as

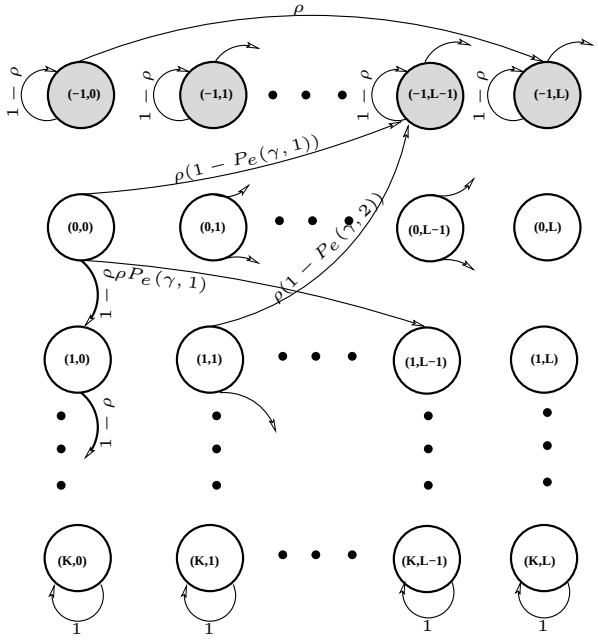


Fig. 4. DTMC for the power control policy $L_1 = 1, L_2 = 2, L_3 = 3, L_4 = 4$, and $K = 4$. The energy states are normalized with respect to E.

$$P_{\text{out}}(K) = \sum_i \pi(i) \mathbb{E}_\gamma \left\{ P_{\text{out}}(K \,|i, \gamma, r = 0) \right\}, \quad (23)$$

where $\pi(i)$ is the stationary probability that the node has energy iE at the beginning of the frame and $P_{\text{out}}(K|i, \gamma, r)$ is the outage probability conditioned on the channel gain γ, the battery state iE, and the ACK state r, at the beginning of the m^{th} frame. Also, $\mathbb{E}_\gamma\{\cdot\}$ represents the expectation over the distribution of the channel gain.

Now, $P_{\text{out}}(K|i, \gamma, r)$ is given by the recursive relation over K in (24) at the bottom of the next page. In (24), $\Psi^{(1)} \triangleq \min(L_r, i + L), \Psi^{(2)} \triangleq \min(L_r, i)$, and $P_e(\gamma, E)$ is the packet error probability, as defined in (4). The terms in (24) are obtained by tracking the battery transitions that happen under different possible energy harvesting and packet transmission events, and accounting for the outage probability of the packet over the remaining $K-1$ attempts. For example, if battery energy level i at the start of the frame exceeds $\sum_{r=1}^{K} L_r$, regardless of the energy harvested, the EHS node can make all K attempts to transmit the packet, at power levels $L_1, L_2, \ldots, L_K$. Hence, a packet outage occurs only if all K transmission attempts fail, and is given the product of their probabilities. Note that we have omitted the dependence of $\Psi^{(1)}$ and $\Psi^{(2)}$ on r and i for notational convenience.

The stationary probabilities π can be obtained by solving the balance equation

$$\pi(j) = \sum_i \mathbb{E}\left[\Pr\left(B_{(m+1)K} = j \,|B_{mK} = i, \gamma\right)\right] \pi(i), \quad (25)$$

where the expectation is taken with respect to the distribution of γ. To obtain $\Pr(B_{(m+1)K} = j | B_{mK} = i, \gamma)$, we use the state transition probability matrix $G(\gamma)$, whose elements represent the probability of a transition from (i, j) to (r, s). It is defined as

$$G_{ij}^{rs} = \Pr(B_{n+1} = j, U_{n+1} = s | B_n = i, U_n = r, \gamma). \quad (26)$$

It is easy to show that, for $r = 0, 1, \ldots, K-1$ and $i \geq 1$,

$$G_{ij}^{rs} = \begin{cases} \rho P_e(\gamma, \Psi^{(1)}), & j = i + L - \Psi^{(1)}, s = r + 1, \\ \rho(1 - P_e(\gamma, \Psi^{(1)})), & j = i + L - \Psi^{(1)}, s = -1, \\ (1-\rho) P_e(\gamma, \Psi^{(2)}), & j = i - \Psi^{(2)}, s = r + 1, \\ (1-\rho)(1 - P_e(\gamma, \Psi^{(2)})), & j = i - \Psi^{(2)}, s = -1, \\ 0, & \text{else.} \end{cases} \quad (27)$$

For $r = 0, 1, \ldots, K-1$ and $i = 0$,

$$G_{ij}^{rs} = \begin{cases} \rho P_e(\gamma, \Psi^{(1)}), & j = i + L - \Psi^{(1)}, s = r + 1, \\ \rho(1 - P_e(\gamma, \Psi^{(1)})), & j = i + L - \Psi^{(1)}, s = -1, \\ (1-\rho), & j = i, s = r + 1, \\ 0, & \text{else.} \end{cases} \quad (28)$$

For $r = -1$ and $i \geq 0$,

$$G_{ij}^{rs} = \begin{cases} \rho, & j = i + L, s = -1, \\ (1-\rho), & j = i, s = -1, \\ 0, & \text{else.} \end{cases} \quad (29)$$

The states $(i, K), \forall i$, are absorbing states. The above transition probabilities follow from the given power control policy, and track the following independent events: i) Whether energy was harvested in the slot, which happens with probability ρ. ii) Whether the transmitted packet was correctly decoded by the receiver, which happens with probability $1 - P_e(\gamma, r)$.

We can now summarize the procedure for computing the outage probability:

1) Compute $G^K(\gamma)$, where $G(\gamma)$ is the state transition probability matrix in (26).
2) Using the entries of $G^K(\gamma)$, compute the probability $\Pr(B_{(m+1)K} = j | B_{mK} = i, \gamma, r = 0)$ as:
$$\sum_{u=-1}^{K} \Pr(B_{(m+1)K} = j, U_{(m+1)K} = u | B_{mK} = i, U_{mK} = 0, \gamma). \quad (30)$$
3) Obtain the stationary probabilities $\pi(j)$ by solving (25).
4) Obtain $P_{\text{out}}(K | i, \gamma, r = 0)$ from (24).
5) Compute $P_{\text{out}}(K)$ using (23).

The simulation results in the next section illustrate that the above analysis brings out the importance of tuning the transmit power control parameters in optimizing the EHS link performance. Moreover, by comparing with the POMDP approach presented in the previous section, we show how the POMDP solution improves performance over fixed-power transmission policies by exploiting all the information available at the EHS.

VI. Simulation Results

In this section, we present simulation results that demonstrate the performance improvement obtainable from the POMDP approach compared to ad hoc policies that only admit a fixed transmit power level [15]. We also validate the outage probability analysis in Sec. V. The channel from the EHS node to its destination is assumed to be Rayleigh faded in the block-fading case, and with the time-variation following the popular Jakes' spectrum in the correlated fading case. The noise power spectral density is taken as $N_0 = 1$, and the packets are assumed to be uncoded BPSK modulated with $\ell = 10$ bits. The value of E_s follows from the energy budget requirement of a typical wireless sensor node. For example, $E_s = 15$ dB corresponds to $L = 2$, i.e., twice the energy required to transmit a packet, with a transmit power of 1.38 mW, carrier frequency of 2 GHz, slot duration of $T_p = 10$ ms, distance $d = 10 d_0$, where $d_0 = 10$ m is a reference distance, a path loss exponent of 3.9, an additive noise corresponding to a temperature of 300 K, and a bandwidth of 1 MHz [34]. We consider $K = 4$ attempts per packet, and all the simulations are run for over 10^7 slots.

a) Linear Power Algorithm: Recall the general power policy considered in Sec. V. During the i^{th} slot within a frame, the packet is transmitted at power $L_i E$, provided there is enough energy in the battery and an ACK has not yet been received. Let us represent the power policy as $[L_1\ L_2\ L_3\ L_4]$. For example, $L_i = i$, for $i = 1, 2, \ldots, K$, represents a *linear* power increase policy. Also, $[2\ 2\ 2\ 2]$ corresponds to the fixed-power retransmission scheme considered in [15], where all four transmission attempts are made at a power level $2E$.

The outage probability performance of the general power policy is shown as a function of the energy harvesting probability in Fig. 5. In the legend, `Sim` and `Analy` correspond to the results obtained from simulations and from the analytical expression in (23), respectively. In the region marked `Better performance`, the linear policy outperforms the scheme considered in [15], showing that fixed-power retransmissions schemes, are, in general, suboptimal. Thus, it is possible to tune the L_is to improve the outage performance. Also note

$$P_{\text{out}}(K | i, \gamma, r) = \begin{cases} \rho P_e\left(\gamma, \Psi^{(1)} E\right) P_{\text{out}}\left(K - 1 \left| i + L - \Psi^{(1)}, \gamma, r + 1\right.\right) & \\ \quad + (1-\rho) P_{\text{out}}\left(K - 1 \left| i, \gamma, r + 1\right.\right), & \text{if } i = 0 \\ \rho P_e\left(\gamma, \Psi^{(1)} E\right) P_{\text{out}}\left(K - 1 \left| i + L - \Psi^{(1)}, \gamma, r + 1\right.\right) & \\ \quad + (1-\rho) P_e\left(\gamma, \Psi^{(2)} E\right) P_{\text{out}}\left(K - 1 \left| i - \Psi^{(2)}, \gamma, r + 1\right.\right), & \text{if } 0 < i \leq \sum_{r=1}^{K} L_r \\ \Pi_{r=1}^{K} P_e\left(\gamma, L_r E\right), & \text{else.} \end{cases} \quad (24)$$

the close agreement between the simulation and the analytical results, which validates the analysis.

b) POMDP Solution: In the following, we compare the packet outage performance of the algorithms in Sec. IV against the conventional fixed-power retransmission scheme. Let the fixed-power be denoted by E_w and define $W \triangleq E_w/E_s$. For comparison with past work [15], we assume that W is an integer multiple or integer fraction of E_s. Figure 7 shows the simulation results for the block fading channel case. The voting policy and the ML estimation policy perform almost equally well, and they significantly outperform the fixed-power retransmission scheme.

In the correlated fading case, the channel correlation is assumed to follow the FSMC shown in Fig. 2. The parameters for the FSMC model are taken as $T_p = 10$ ms and $f_d T_p = 0.03$, where f_d denotes the Doppler frequency. Figure 6 shows the threshold nature of the optimal policy of the fully observable MDP for the correlated fading channel, when the action space is restricted to be binary-valued. Recall that the optimal policy is characterized by a single threshold B_{thresh} on the battery state, one for each combination of the channel, transmission index and ACK states. From the simulations, we observe that the threshold decreases with transmission index. This is intuitive, since towards the end of frame, the EHS must transmit even if it has a lower battery level. The policy also tries to ensure that each packet is attempted at least once towards the end of frame. This not only improves the packet outage, but also improves the belief state of the channel by collecting observations.

Figure 8 compares the performance of the fixed and linear power policies with that of the POMDP policies for the correlated fading channel. As before, the ML heuristic and the voting heuristic outperform the other schemes. To achieve the same outage probability, the POMDP solution typically requires only about 60-80% of the energy harvesting rate compared to the best fixed-power scheme. Also, we see that the ML heuristic and the voting heuristic policies perform nearly equally well. Hence, the ML heuristic policy is a better choice for implementation on EHS platforms, since it is computationally simpler.

Finally, Figure 9 shows how the outage probability varies as a function of ρ with various battery capacities. In particular, when $B_{\max} = 0$, the optimal policy is to transmit the packet in slots where energy is harvested, and until an ACK is received. An outage does not occur if the EHS is able to harvest energy and successfully deliver the packet in any of the K slots within a given frame. The plot thus highlights the role of the battery capacity in improving the outage probability.

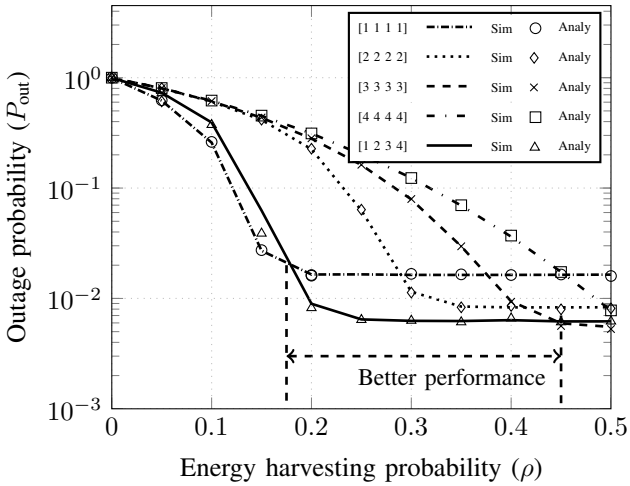


Fig. 5. Comparison of the linear power policy ($L_i = i, i = 1, 2, \ldots, K$) and the fixed-power algorithm [15] with various fixed-power levels, for $K = 4$, $L = 2$, and $E_s = 18$ dB. In the legend, the numbers in the square brackets represent the packet transmit power levels for each of the 4 transmission attempts. `Sim` and `Analy` correspond to the results obtained from simulations and from the analytical expression in (23), respectively. In the region marked `Better performance`, the linear policy, [1 2 3 4], outperforms the fixed-power retransmission schemes.

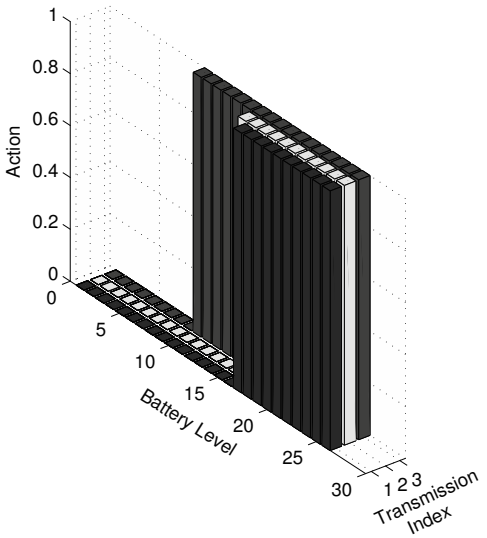


Fig. 6. Correlated channel case: Optimal policy for the binary action case. The optimal policy is a threshold policy in the battery state. Here, $K = 4$, $L = 4$, $N = 7$, $E_s = 15$ dB, and $B_{\max} = 10E_s$.

VII. Conclusions

In this paper, we considered the problem of power management for EHS nodes with packet retransmissions. A sequential decision-theoretic framework was used to obtain outage optimal policies for the correlated channel. Since the channel state is only partially observable through the ACK/NACK messages, we formulated the problem as a POMDP. Exact solutions were found to be computationally infeasible, motivating us to explore two computationally efficient suboptimal solutions: the ML policy and the voting heuristic policy. The solutions were based on the belief state of the channel and the solution of the underlying MDP. We also derived structural results for the fully observable MDP in the binary action case. Thus, the decision-theoretic approach adopted in this paper is a promising technique for the design of power management policies for EHS nodes that need to operate under stringent energy constraints and yet achieve high reliability or throughput. We also benchmarked the performance of the POMDP solution by deriving analytical expressions for the packet outage probability for the case of a block fading channel with a fixed-power retransmission scheme. Simulation results showed

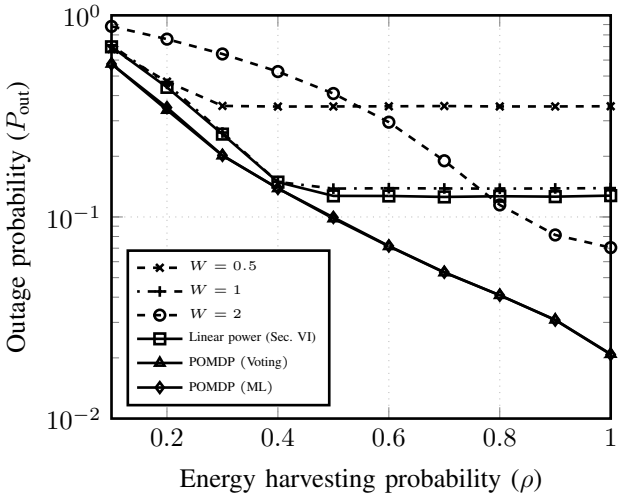


Fig. 7. Block fading channel case: Comparison of ML and voting heuristic solutions of the EHS POMDP against fixed-power transmission, for $K = 4$, $L = 4$, $N = 7$, $E_s = 12$ dB, and $B_{\max} = 20E_s$.

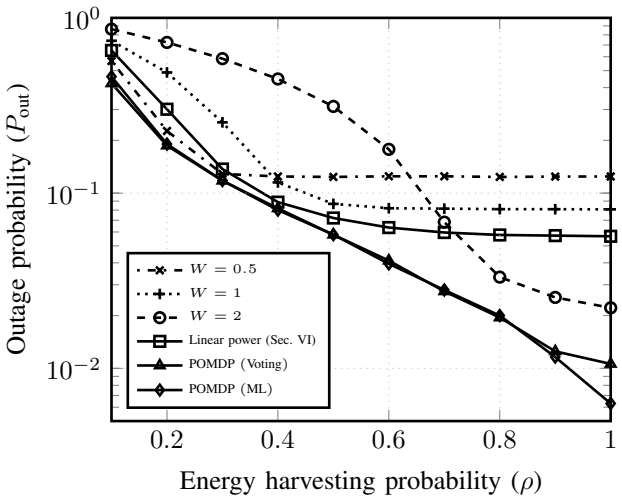


Fig. 8. Correlated channel case: Comparison of ML and voting heuristic solutions of the EHS POMDP against fixed-power transmission for $K = 4$, $L = 4$, $N = 7$, $E_s = 15$ dB, and $B_{\max} = 10E_s$.

that the proposed POMDP solutions significantly outperform the existing fixed-power retransmission schemes. Extensions of this work could involve considering other objective functions, such as maximizing the average rate by adapting the modulation and coding scheme based on the ACK/NACK messages, considering battery inefficiencies, and including other energy harvesting models. These different issues promise to offer interesting avenues for future work.

APPENDIX

A. Proof of Theorem 1

As mentioned earlier, due to the convergence of (16) for all initial conditions, the discounted cost optimal policy would be a threshold policy in the energy buffer state b, provided $Q^{\nu}_{\infty}([b,\gamma,u,k],a)$ is submodular in (b,a), for some initial condition [31]. Hence, we need to show that, with suitable initialization, $Q^{\nu}_{n}([b,\gamma,u,k],a)$ is submodular in (b,a). That

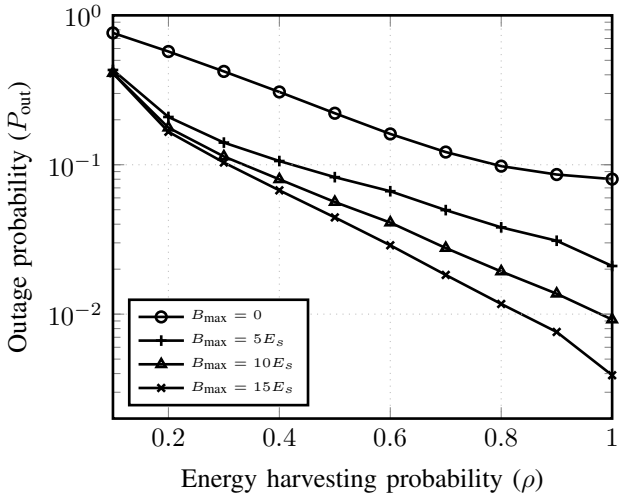


Fig. 9. Correlated channel case: Increasing the battery capacity decreases the outage probability. Here, $K = 4$, $L = 4$, $N = 7$, and $E_s = 15$ dB.

is, since $\mathcal{A} = \{0,1\}$, $Q^{\nu}_{n}([b,\gamma,u,k],1) - Q^{\nu}_{n}([b,\gamma,u,k],0)$ should monotonically decrease with b. In the following, we consider only the case of $u = 0$. For $u = 1$, the possible action set is restricted to $\{0\}$ and hence monotonicity is trivial.

Notation: Let $V_{n,u}(b) \triangleq V^{\nu}_{n}([b,\gamma,u,k])$ and $V'_{n,u'}(b') \triangleq V^{\nu}_{n}([b',\gamma',u',k'])$. Similarly, let $Q_{n,u}(b,a) \triangleq Q^{\nu}_{n}([b,\gamma,u,k],a)$ and $Q'_{n,u'}(b',a) \triangleq Q^{\nu}_{n}([b',\gamma',u',k'],a)$. Finally, let $c(a) \triangleq c([b,\gamma,u,k],a)$. The simplified notation above is obtained by dropping the dependence on the parameters that remain unchanged through the rest of the proof.

In the following, we consider the case corresponding to transmission at the end of the frame ($k = K-1$) and transmission at the middle of the frame ($k \neq K-1$) separately, as their state transition functions are different.

End of Frame ($k = K-1$): We have

$$\begin{aligned} &Q_{n+1,0}(b,1) - Q_{n+1,0}(b,0) = c(1) - c(0) \\ &+\nu \sum_{\gamma' \in \mathcal{G}} P_{\gamma,\gamma'} \left\{ \rho \left\{ V'_{n,0}(b+L-1) - V'_{n,0}(b+L) \right\} \right. \\ &\left. +(1-\rho) \left\{ V'_{n,0}(b-1) - V'_{n,0}(b) \right\} \right\}. \end{aligned} \tag{31}$$

For $Q_{n+1,0}(s,a)$ to be submodular in (b,a), we need the terms in (31) to monotonically decrease with b. Since the two terms in (31) are identical expect for a shift in b by L, the value function must satisfy the following sufficient condition:

$$V_{n,0}(b-1) - V_{n,0}(b) \geq V_{n,0}(b) - V_{n,0}(b+1). \tag{32}$$

Rearranging terms, we need to show that

$$V_{n,0}(b+1) - V_{n,0}(b) \geq V_{n,0}(b) - V_{n,0}(b-1), \tag{33}$$

i.e., that $V_{n,0}(b)$ has an increasing difference in b. We show this using an inductive argument. First, note that one can always choose $V_{0,0}(b) = V^{\nu}_{0}([b,\gamma,u,k])$ to have an increasing difference in b; an example is the exponential function. Assume that $V^{\nu}_{n}([b,\gamma,u,k])$ has increasing difference in b. We now prove that $V^{\nu}_{n+1}([b,\gamma,u,k])$ has increasing difference in b. In (17), let $a_0,\ a_1,\ a_2 \in \mathcal{A}$ correspond to the optimal

action in the states $[b-1,\gamma,0,k]$, $[b,\gamma,0,k]$ and $[b+1,\gamma,0,k]$, respectively. Then,

$$\begin{aligned} V_{n+1,0}\left(b+1\right) &= Q_{n+1,0}\left(b+1,a_2\right), &(34)\\ V_{n+1,0}\left(b\right) &= Q_{n+1,0}\left(b,a_1\right), &(35)\\ V_{n+1,0}\left(b-1\right) &= Q_{n+1,0}\left(b-1,a_0\right). &(36)\end{aligned}$$

Substituting (34) through (36) in (33) with n replaced with $n+1$, we get the difference between the left- and right-hand-side as

$$\begin{aligned} &\left(Q_{n+1,0}\left(b+1,a_2\right)-Q_{n+1,0}\left(b,a_1\right)\right)\\ &-\left(Q_{n+1,0}\left(b,a_1\right)-Q_{n+1,0}\left(b-1,a_0\right)\right)\\ &=\left(Q_{n+1,0}\left(b+1,a_2\right)-Q_{n+1,0}\left(b,a_2\right)\right)\\ &\quad+\left(Q_{n+1,0}\left(b,a_2\right)-Q_{n+1,0}\left(b,a_1\right)\right)\\ &\quad+\left(Q_{n+1,0}\left(b,a_0\right)-Q_{n+1,0}\left(b,a_1\right)\right)\\ &\quad-\left(Q_{n+1,0}\left(b,a_0\right)-Q_{n+1,0}\left(b-1,a_0\right)\right). \qquad (37)\end{aligned}$$

The terms $Q_{n+1,0}\left(b,a_2\right)-Q_{n+1,0}\left(b,a_1\right)$ and $Q_{n+1,0}\left(b,a_0\right)-Q_{n+1,0}\left(b,a_1\right)$ are non-negative due to optimality of action a_1 for state $[b,\gamma,0,k]$. Define

$$\begin{aligned} A &\triangleq Q_{n+1,0}\left(b+1,a_2\right)-Q_{n+1,0}\left(b,a_2\right), &(38)\\ B &\triangleq Q_{n+1,0}\left(b,a_0\right)-Q_{n+1,0}\left(b-1,a_0\right). &(39)\end{aligned}$$

In order to show that (37) is non-negative, we need to prove that $A \geq B$. We lower bound A as follows:

$$\begin{aligned} A=&\nu\sum_{\gamma'\in\mathcal{G}}P_{\gamma\gamma'}\left\{\mathbb{1}_{\{a_2=1\}}\left\{\rho\left(V'_{n,0}\left(b+L\right)\right.\right.\right.\\ &\left.-V'_{n,0}\left(b+L-1\right)\right)+\left(1-\rho\right)\left(V'_{n,0}\left(b\right)-V'_{n,0}\left(b-1\right)\right)\Big\}\\ &+\mathbb{1}_{\{a_2=0\}}\left\{\rho\left(V'_{n,0}\left(b+L+1\right)-V'_{n,0}\left(b+L\right)\right)\right.\\ &\left.\left.+\left(1-\rho\right)\left(V'_{n,0}\left(b+1\right)-V'_{n,0}\left(b\right)\right)\right\}\right\},\\ \geq&\nu\sum_{\gamma'\in\mathcal{G}}P_{\gamma\gamma'}\left\{\rho\left(V'_{n,0}\left(b+L\right)-V'_{n,0}\left(b+L-1\right)\right)\right.\\ &\left.+\left(1-\rho\right)\left(V'_{n,0}\left(b\right)-V'_{n,0}\left(b-1\right)\right)\right\} \qquad (40)\end{aligned}$$

where the inequality in (40) is obtained by applying (33).

Similarly, we upper bound B to obtain

$$\begin{aligned} B=&\nu\sum_{\gamma'\in\mathcal{G}}P_{\gamma\gamma'}\left\{\mathbb{1}_{\{a_0=1\}}\left\{\rho\left(V'_{n,0}\left(b+L-1\right)\right.\right.\right.\\ &\left.-V'_{n,0}\left(b+L-2\right)\right)+\bar{\rho}\left(V'_{n,0}\left(b-1\right)-V'_{n,0}\left(b-2\right)\right)\Big\}\\ &+\mathbb{1}_{\{a_2=0\}}\left\{\rho\left(V'_{n,0}\left(b+L\right)-V'_{n,0}\left(b+L-1\right)\right)\right.\\ &\left.\left.+\bar{\rho}\left(V'_{n,0}\left(b\right)-V'_{n,0}\left(b-1\right)\right)\right\}\right\},\\ \leq&\nu\sum_{\gamma'\in\mathcal{G}}P_{\gamma\gamma'}\left\{\rho\left(V'_{n,0}\left(b+L\right)-V'_{n,0}\left(b+L-1\right)\right)\right.\\ &\left.+\bar{\rho}\left(V'_{n,0}\left(b\right)-V'_{n,0}\left(b-1\right)\right)\right\}, \qquad (41)\end{aligned}$$

where $\bar{\rho}\triangleq 1-\rho$, and the inequality in (41) is again obtained by applying (33). Thus, from (40) and (41), $A \geq B$, and hence $Q_{n+1,0}\left(b,a\right)$ is submodular in (b,a).

Within a frame ($k \neq K-1$): As before, we start with the goal of showing that $Q^{\nu}_{n+1}([b,\gamma,u,k],1) - Q^{\nu}_{n+1}([b,\gamma,u,k],0)$ is monotonically decreasing with b. With a little manipulation,

$$\begin{aligned} &Q_{n+1,0}\left(b,1\right)-Q_{n+1,0}\left(b,0\right)\\ &=c(1)-c(0)\\ &+\nu\sum_{\gamma'\in\mathcal{G}}\left\{P_{\gamma\gamma'}\left\{\rho P_e\left\{V'_{n,0}\left(b+L-1\right)-V'_{n,0}\left(b+L\right)\right\}\right.\right.\\ &+\rho(1-P_e)\left\{V'_{n,1}\left(b+L-1\right)-V'_{n,0}\left(b+L\right)\right\}\\ &+(1-\rho)P_e\left\{V'_{n,0}\left(b-1\right)-V'_{n,0}\left(b\right)\right\}\\ &\left.\left.+(1-\rho)(1-P_e)\left\{V'_{n,1}\left(b-1\right)-V'_{n,0}\left(b\right)\right\}\right\}\right. \qquad (42)\end{aligned}$$

Following a similar procedure as for the end of frame ($k = K-1$) case, we obtain the following three sufficient conditions for $Q_{n,0}\left(b,a\right)$ to be monotonically decreasing in b:

$$\begin{aligned} &1)\ V_{n,0}\left(b+1\right)-V_{n,0}\left(b\right)\geq V_{n,0}\left(b\right)-V_{n,0}\left(b-1\right). &(43)\\ &2)\ V_{n,0}\left(b+1\right)-V_{n,0}\left(b\right)\geq V_{n,1}\left(b\right)-V_{n,1}\left(b-1\right). &(44)\\ &3)\ V_{n,1}\left(b\right)-V_{n,1}\left(b-1\right)\geq V_{n,0}\left(b\right)-V_{n,0}\left(b-1\right). &(45)\end{aligned}$$

Using a technique similar to that employed for showing (32), it can be shown that (43) and (44) hold. Hence, we focus our attention on showing (45). As before, one can always choose the initialization $V_{0,u}\left(b\right) = V^{\nu}_{0}([b,\gamma,u,k])$ such that (43) to (45) is satisfied. Assume that $V_{n,u}\left(b\right) = V^{\nu}_{n}([b,\gamma,u,k])$ satisfies (45).

We now use induction to show that $V_{n+1,u}\left(b\right) = V^{\nu}_{n+1}([b,\gamma,u,k])$ also satisfies (45), i.e., that

$$\begin{aligned} &\left(V_{n+1,1}\left(b\right)-V_{n+1,1}\left(b-1\right)\right)\\ &\qquad-\left(V_{n+1,0}\left(b\right)\right)-V_{n+1,0}\left(b-1\right)\right)\geq 0. \qquad (46)\end{aligned}$$

Since $\{0\}$ is the only allowed action when $u=1$, we obtain

$$\begin{aligned} &\left(V_{n+1,1}\left(b\right)-V_{n+1,1}\left(b-1\right)\right)-\left(V_{n+1,0}\left(b\right)-V_{n+1,0}\left(b-1\right)\right)\\ =&\left(Q_{n+1,1}\left(b,0\right)-Q_{n+1,1}\left(b-1,0\right)\right)\\ &-\left(Q_{n+1,0}\left(b,a_1\right)-Q_{n+1,0}\left(b-1,a_0\right)\right), \qquad (47)\end{aligned}$$

for some $a_1,\ a_0\ \in\mathcal{A}$. In (47), a_0 and a_1 denote the optimal actions in states $[b-1,\gamma,0,k]$ and $[b,\gamma,0,k]$, respectively. Define P and Q as follows:

$$\begin{aligned} P\triangleq&Q_{n+1,1}\left(b,0\right)-Q_{n+1,1}\left(b-1,0\right), &(48)\\ =&\nu\sum_{\gamma'\in\mathcal{G}}P_{\gamma\gamma'}\left\{\rho\left\{V'_{n,1}\left(b+L\right)\right)-V'_{n,1}\left(b+L-1\right)\right\}\\ &+\left(1-\rho\right)\left\{V'_{n,1}\left(b\right)-V'_{n,1}\left(b-1\right)\right\}\Big\}. &(49)\\ Q\triangleq&Q_{n+1,0}\left(b,a_1\right)-Q_{n+1,0}\left(b-1,a_0\right), &(50)\\ =&\left\{Q_{n+1,0}\left(b,a_1\right)-Q_{n+1,0}\left(b,a_0\right)\right\}\\ &+\left\{Q_{n+1,0}\left(b,a_0\right)-Q_{n+1,0}\left(b-1,a_0\right)\right\}, &(51)\\ \leq&Q_{n+1,0}\left(b,a_0\right)-Q_{n+1,0}\left(b-1,a_0\right), &(52)\end{aligned}$$

where the inequality follows because action a_1 is optimal in state $[b,\gamma,0,k]$. The right hand side in (52) can be written as

$$\begin{aligned} =&\nu\sum_{\gamma'\in\mathcal{G}}P_{\gamma\gamma'}\left\{\mathbb{1}_{\{a_0=0\}}\left\{\rho\left\{V'_{n,0}\left(b+L\right)-V'_{n,0}\left(b+L-1\right)\right\}\right.\right.\\ &\left.+\left(1-\rho\right)\left\{V'_{n,0}\left(b\right)-V'_{n,0}\left(b-1\right)\right\}\right\}\\ &+\mathbb{1}_{\{a_0=1\}}\left\{\rho P_e\left\{V'_{n,0}\left(b+L-1\right)-V'_{n,0}\left(b+L-2\right)\right\}\right.\\ &+\rho(1-P_e)\left\{V'_{n,1}\left(b+L-1\right)-V'_{n,1}\left(b+L-2\right)\right\}\\ &\left.+\left\{(1-\rho)P_e\left\{V'_{n,0}\left(b-1\right)-V'_{n,0}\left(b-2\right)\right\}\right\}\right.\end{aligned}$$

$$+\left\{(1-\rho)(1-P_e)\left\{V'_{n,1}(b-1)-V'_{n,1}(b-2)\right\}\right\}\Big\},$$
$$\leq \nu \sum_{\gamma'\in\mathcal{G}} P_{\gamma\gamma'}\left\{\rho\left\{V'_{n,0}(b+L)-V'_{n,0}(b+L-1)\right\}\right.$$
$$\left.+(1-\rho)\left\{V'_{n,0}(b)-V'_{n,0}(b-1)\right\}\right\}. \quad (53)$$

The inequality above is obtained by using (43) and (44). Finally, the right hand side above is

$$\leq \nu \sum_{\gamma'\in\mathcal{G}} P_{\gamma\gamma'}\left\{\rho\left\{V'_{n,1}(b+L)-V'_{n,1}(b+L-1)\right\}\right.$$
$$\left.+(1-\rho)\left\{V'_{n,1}(b)-V'_{n,1}(b-1)\right\}\right\}, \quad (54)$$
$$= P, \quad (55)$$

where the inequality is obtained using (45). Thus, by induction, (43), (44), (45) are satisfied, and hence, $Q^{\nu}_{n+1}(b,a)$ is submodular in (b,a). Therefore, the optimal policy is monotone in b, and consequently is a threshold policy in the battery state. □

REFERENCES

[1] I. Stanojev, O. Simeone, Y. Bar-Ness, and D. Kim, "On the energy efficiency of hybrid-ARQ protocols in fading channels," in *Proc. ICC*, 2007, pp. 3173–3177.

[2] "*IEEE standard 802, part 15.4: wireless medium access control (MAC) and physical layer (PHY) specifications for low rate wireless personal area networks (WPANs)*," Tech. Rep., 2003.

[3] A. Kansal, J. Hsu, S. Zahedi, and M. Srivastava, "Power management in energy harvesting sensor networks," *ACM Trans. Embedded Computing Sys.*, vol. 7, pp. 1–38, 2007.

[4] J. Lei, R. Yates, and L. Greenstein, "A generic model for optimizing single-hop transmission policy of replenishable sensors," *IEEE Trans. Wireless Commun.*, vol. 8, no. 2, pp. 547–551, Feb. 2009.

[5] J. Yang and S. Ulukus, "Transmission completion time minimization in an energy harvesting system," in *Proc. Conf. on Inform. Sci. and Syst. (CISS)*, 2010, pp. 1–6.

[6] K. Tutuncuoglu and A. Yener, "Short-term throughput maximization for battery limited energy harvesting nodes," in *Proc. ICC*, 2011, pp. 1–5.

[7] A. Seyedi and B. Sikdar, "Energy efficient transmission strategies for body sensor networks with energy harvesting," *IEEE Trans. Commun.*, vol. 58, no. 7, pp. 2116–2126, 2010.

[8] V. Sharma, U. Mukherji, V. Joseph, and S. Gupta, "Optimal energy management policies for energy harvesting sensor nodes," *IEEE Trans. Wireless Commun.*, vol. 9, no. 4, pp. 1326–1336, 2010.

[9] A. Sinha and P. Chaporkar, "Optimal power allocation for a renewable energy source," in *Proc. NCC*, 2012.

[10] E. Sardini and M. Serpelloni, "Self-powered wireless sensor for air temperature and velocity measurements with energy harvesting capability," *IEEE Trans. Instrum. Meas.*, vol. 60, no. 5, pp. 1838–1844, May 2011.

[11] V. Kyriatzis, N. Samaras, P. Stavroulakis, H. Takruri-Rizk, and S. Tzortzios, "Enviromote: A new solar-harvesting platform prototype for wireless sensor networks/work-in-progress report," in *Proc. PIMRC*, Sep. 2007, pp. 1–5.

[12] L. Kaelbling, M. Littman, and A. Cassandra, "Planning and acting in partially observable stochastic domains," *Artificial Intelligence*, vol. 101, no. 1-2, pp. 99–134, 1998.

[13] H. Li, N. Jaggi, and B. Sikdar, "Cooperative relay scheduling under partial state information in energy harvesting sensor networks," in *Proc. Globecom*, 2010, pp. 1–5.

[14] N. Jaggi, K. Kar, and A. Krishnamurthy, "Rechargeable sensor activation under temporally correlated events," *Wireless Networks*, vol. 15, no. 5, pp. 619–635, 2009.

[15] B. Medepally, N. B. Mehta, and C. R. Murthy, "Implications of energy profile and storage on energy harvesting sensor link performance," in *Proc. Globecom*, 2009, pp. 1–6.

[16] J. A. Paradiso and M. Feldmeier, "A compact, wireless, self powered pushbutton controller," in *Proc. Int. Conf. Ubiquitous Comput.*, 2001, pp. 299–304.

[17] D. Niyato, E. Hossain, and A. Fallahi, "Sleep and wakeup strategies in solar-powered wireless sensor/mesh networks: performance analysis and optimization," *IEEE Trans. Mobile Comput.*, vol. 6, no. 2, pp. 221–236, Feb. 2007.

[18] C. Tan and N. Beaulieu, "On first-order Markov modeling for the Rayleigh fading channel," *IEEE Trans. Commun.*, vol. 48, no. 12, pp. 2032–2040, 2000.

[19] A. Farrokh, V. Krishnamurthy, and R. Schober, "Optimal adaptive modulation and coding with switching costs," *IEEE Trans. Commun.*, vol. 57, no. 3, pp. 697 –706, Mar. 2009.

[20] A. Karmokar, D. Djonin, and V. Bhargava, "POMDP-based coding rate adaptation for type-I hybrid ARQ systems over fading channels with memory," *IEEE Trans. Wireless Commun.*, vol. 5, no. 12, pp. 3512–3523, 2006.

[21] H. Wang and N. Moayeri, "Finite-state Markov channel- A useful model for radio communication channels," *IEEE Trans. Veh. Technol.*, vol. 44, no. 1, pp. 163–171, 1995.

[22] Q. Zhang and S. Kassam, "Finite-state Markov model for Rayleigh fading channels," *IEEE Trans. Commun.*, vol. 47, no. 11, pp. 1688–1692, 1999.

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

[24] C. H. Papadimitriou and J. N. Tsitsiklis, "The complexity of optimal queuing network control," *Mathematics of Operations Research*, vol. 24, no. 2, pp. pp. 293–305, 1999.

[25] M. Littman, A. Cassandra, and L. Kaelbling, "Learning policies for partially observable environments: Scaling up," in *Int. Conf. Machine Learning*, 1995, pp. 362–370.

[26] J. Pineau, G. Gordon, and S. Thrun, "Point-based value iteration: An anytime algorithm for POMDPs," in *Int. Joint Conf. on Artificial Intelligence*, vol. 18, 2003, pp. 1025–1032.

[27] D. P. Bertsekas, *Dynamic programming and optimal control*, 3rd ed. Athena Scientific Belmont, MA, 2005, vol. 1.

[28] S. Ross, *Introduction to stochastic dynamic programming: Probability and mathematical*. Academic Press, Inc., 1983.

[29] L. Sennott, "Average cost optimal stationary policies in infinite state markov decision processes with unbounded costs," *Operations Research*, vol. 37, pp. 626–633, 1989.

[30] M. Puterman, *Markov decision processes: Discrete stochastic dynamic programming*. John Wiley & Sons, Inc., 1994.

[31] D. Topkis, *Supermodularity and complementarity*. Princeton University Press, 1998.

[32] A. R. Cassandra, "Exact and approximate algorithms for partially observable Markov decision processes," Ph.D. dissertation, 1998.

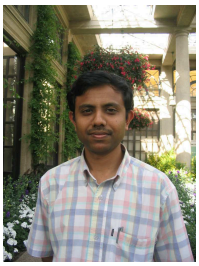
[33] R. Simmons and S. Koenig, "Probabilistic robot navigation in partially observable environments," in *Proc. Int. Joint Conf. on Artificial Intelligence*, vol. 14, 1995, pp. 1080–1087.

[34] A. Goldsmith, *Wireless communications*. Cambridge University Press, 2005.

Anup Aprem received the B. Tech. Degree in Electronics and Communication Engineering from the College of Engineering, Trivandrum, India, in 2007 and the M. E. degree in Signal Processing from the Electrical Communication Engineering department at Indian Institute of Science, Bangalore, India, in 2012. Currently, he is working as a design engineer at Analog Devices, Inc. His research interest includes power management in energy harvesting sensors, cross layer design of communication systems and MIMO systems.

Chandra R. Murthy (S'03-M'06-SM'11) received the B.Tech. degree in Electrical Engineering from the Indian Institute of Technology, Madras in 1998, the M.S. and Ph.D. degrees in Electrical and Computer Engineering from Purdue University and the University of California, San Diego, in 2000 and 2006, respectively.

From 2000 to 2002, he worked as an engineer for Qualcomm Inc., where he worked on WCDMA baseband transceiver design and 802.11b baseband receivers. From Aug. 2006 to Aug. 2007, he worked as a staff engineer at Beceem Communications Inc. on advanced receiver architectures for the 802.16e Mobile WiMAX standard. In Sept. 2007, he joined as an assistant professor at the Department of Electrical Communication Engineering at the Indian Institute of Science, where he is currently working. His research interests are in the areas of Cognitive Radio, Energy Harvesting Wireless Sensors and MIMO systems with channel-state feedback. He is currently serving as an associate editor for the IEEE Signal Processing Letters.



Neelesh B. Mehta (S'98-M'01-SM'06) received his Bachelor of Technology degree in Electronics and Communications Eng. from the Indian Institute of Technology (IIT), Madras in 1996, and his M.S. and Ph.D. degrees in Electrical Eng. from the California Institute of Technology, Pasadena, CA, USA in 1997 and 2001, respectively. He is now an Associate Professor in the Dept. of Electrical Communication Eng., Indian Institute of Science (IISc), Bangalore, India. Prior to joining IISc in 2007, he was a research scientist in AT&T Laboratories, NJ, USA, Broadcom Corp., NJ, USA, and Mitsubishi Electric Research Laboratories (MERL), MA, USA from 2001 to 2007.

His research includes work on link adaptation, multiple access protocols, cellular system design, MIMO and antenna selection, cooperative communications, energy harvesting networks, and cognitive radio. He was also actively involved in the Radio Access Network (RAN1) standardization activities in 3GPP from 2003 to 2007. He has served as a TPC co-chair for tracks/symposia in COMSNETS 2014, Globecom 2013, ICC 2013, SPCOM 2012, WISARD 2010 & 2011, NCC 2011, VTC 2009 (Fall), and Chinacom 2008. He has co-authored 40+ IEEE transactions papers, 65+ conference papers, and three book chapters, and is a co-inventor in 20 issued US patents. He is an Editor of IEEE Wireless Communications Letters and the Journal for Communications and Networks, and currently serves as the Director of Conference Publications on the Board of Governors of the IEEE Communications Society.