Multi-Access Communications with Energy Harvesting: A Multi-Armed Bandit Model and the Optimality of the Myopic Policy

Pol Blasco and Deniz Gündüz
Imperial College London, UK
Emails: {p.blasco-moreno12, d.gunduz}@imperial.ac.uk

***Abstract*—A time-slotted multi-access wireless network with N transmitting nodes, each equipped with an energy harvesting (EH) device and a rechargeable battery of finite capacity, is studied. At each time slot (TS) each node is *operative* with a certain probability, which may depend on the availability of data at the node or on the state of the channel. The energy arrival process at each node is modeled as an independent two-state Markov process, such that a node either harvests one unit of energy, or none, at each TS. The access point (AP) schedules a subset of the nodes at each TS. The scheduling policy that maximises the total throughput is studied for a system in which the AP does not know the EH processes and nodes' battery states. The problem is identified as a restless multi-armed bandit (RMAB) problem, and an upper bound on the optimal scheduling policy is found. Under certain assumptions regarding the EH processes and the battery sizes, the optimal scheduling policy is characterized explicitly, and is shown to be myopic. For the general case, the performance of the myopic policy is compared numerically to that of the upper bound.**

***Index Terms*—Energy harvesting, Myopic policy, multi-acess, online scheduing, partiably obserbable Markov decision process, restless multi-armed problem.**

I. Introduction

Low-power wireless networks, such as machine-to-machine and wireless sensor networks, can be complemented with energy harvesting (EH) technology to extend the network lifetime. Tipically, a low-power wireless node is powered by a battery and has a limited lifetime, constrained by the battery size; but when complemented with an energy harvester, that scavenges available energy from the environment, and a rechargeable battery, its lifetime can be prolonged significantly. However, energy availability at the EH nodes is scarce, and, due to the random nature of the energy sources, it arrives at random times and in arbitrary amounts. Hence, in order to take the most out of the scarce energy in the system, it is critical to optimise the scheduling policy of the wireless network using the available information regarding the energy and data arrival processes as well as the channel and battery states of the nodes [1].

Previous research on EH wireless networks can be grouped into three, based on the information available regarding the random processes governing the system [2]. In the offline optimization framework, availability of non-causal information on the exact realizations of the random processes governing the system is assumed at the transmitter [3], [4]. In the online optimization framework [5]–[12], the statistics governing the random processes are assumed to be available at the transmitter, and their realizations are known only causally. The EH communication system is modeled as a Markov decision process (MDP) [5], or as a partially observable MDP (POMDP) [6], and dynamic programming (DP) [13] can be used to optimise the EH communication system numerically. In many practical applications, the state space of the corresponding MDPs and POMDPs is large, and DP becomes computationally prohibitive [14], and the numerical results of DP do not provide much intuition about the structure of the optimal scheduling policy. In order to avoid complex numerical optimisations it is important to characterize the behaviour of the optimal scheduling policy and identify properties about its structure; however, this is possible only in some special cases [7], [9], [10]. In the learning optimization framework, the knowledge about the system behaviour is further relaxed, and even the statistical knowledge about the random processes governing the system is not assumed, and the optimal policy scheduling is learnt over time [12].optimally solve MDPs and POMDPs, however the convergence rate of those techniques decrease as the system state space gets larger.

In this paper, we study the optimal uplink scheduling of low-power wireless nodes by an access point (AP) in the online optimization framework. The low-power wireless nodes are equipped with EH devices, and powered by rechargeable batteries. At each time slot (TS) a node is operative with a certain probability, which may depend on the channel conditions or the availability of data at the node. The EH process at each node is modelled as an independent Markov process, and at each TS, a node either harvests one unit of energy or does not harvest any. The AP is in charge of scheduling, at each TS, the EH nodes to the available orthogonal channels. The nodes transmit data and control information to the AP only if they are scheduled and operative at the same time. Hence, at each TS the AP readily knows the EH process states and battery levels of the operative nodes that are scheduled, but does not receive new information about the other nodes. The AP is interested in maximising the expected sum throughput within a given time horizon. This problem can be model as a POMDP and solved numerically using DP at the expense of a high-computational cost. Instead, we model it as a restless multi-armed bandit (RMAB) problem [15], and prove the optimality of a low-complexity policy in two special cases. Moreover, by relaxing the constraint on the number of nodes that the AP can schedule

at each TS, we obtain an upper bound on the performance of the optimal scheduling policy. Finally, the performance of the low complexity policy is compared to that of the upper bound numerically. The main technical contributions of the paper are summarised as follows:

- We study optimal scheduling policies in multi-access communication with EH, assuming that the AP does not know the battery levels of the nodes or the states of their EH processes.
- We show the optimality of a myopic policy if the nodes do not harvest energy and transmit data at the same time, and the EH process is affected by the scheduling policy.
- We show the optimality of a myopic policy if the nodes do not have batteries and can transmit only if they have harvested energy in the previous TS.
- We provide an upper bound on the performance of any scheduling policy for the general case by relaxing the constraint on the number of nodes that can be scheduled at each TS.
- We show numerically that the myopic policy performs close to the upper bound for the general case.

The rest of this paper is organized as follows. Section II is dedicated to a summary of the related literature. In Section III, we present the EH wireless multi-access network model. In Sections IV and V we characterize explicitly the structure of the optimal policy that maximises the sum throughput for two special cases. In Section VI, we provide an upper bound on the performance. Finally, in Section VII we compare the performance of the myopic policy with that of the upperbound through numerical analysis. Section VIII concludes the paper.

II. Related Work

There is a growing research interest in EH wireless communication systems, and in particular, in developing scheduling policies that exploit the scarce harvested energy in the most efficient manner. In large EH wireless networks, since numerical optimization is computationally prohibitive, it is important to characterise the optimal scheduling policy explicitly, or at least characterize certain properties of the optimal policy.

In [7], the authors assume that the data packets arrive at the EH transmitter as a Poisson process, and each packet has an intrinsic value assigned to it, which also is a random variable. The optimal transmission policy that maximizes the average value of the received packets at the destination is proven to be a threshold policy; that is, for each possible battery level there is a threshold value, and only packets with a higher value than this threshold are transmitted, and the rest are dropped. However, the values of the thresholds have to be computed using numerical techniques, such as DP or linear programming (LP). Reference [8] extends the problem in [7] to the multi-access scenario, in which several nodes access the channel without central coordination while the EH processes at different nodes are time-correlated.

Multi-access in EH wireless networks with a central scheduler, static channels and backlogged nodes has been studied in [9]–[11]. The network central scheduler in [9] does not know the battery levels or the states of the EH processes at the nodes. Assuming that the nodes have unit size batteries, the system is modeled as an RMAB, and the myopic policy, which has a round robin (RR) structure, is shown to maximise the sum throughput. Reference [10] considers nodes with batteries of arbitrary capacity, and MP is found to be optimal in two special cases. In contrast to the present paper, [10] considers static channels and backlogged nodes, and the optimality proof exploits the RR structure of MP. In [11] the UROP policy is proposed, and, under the assumption that the EH nodes have infinite-capacity battery, is shown to be asymptotically optimal.

The problem studied in this paper is modeled as an RMAB problem. In the classic RMAB problem there are several arms, each of which is modelled as a Markov chain [15]. The arms' states are unknown, and at each TS an arm is played. The played arm reveals its state and yields a reward, which is a function of the state. The objective is to find a policy that maximises the total reward over time. RMAB problems have been shown to be, in general, PSPACE hard [16], and our knowledge on the structure of the optimal policy for general RMAB problem is limited.

Recently, the RMAB model has been used to study channel access and cognitive radio problems, and new results on the optimality of MP have been obtained [17]–[21]. The structure and the optimality of MP is proven in [17] and [18] for single and multiple plays, respectively, under certain conditions on the Markov transition probabilities. In [19] the optimality of MP is shown for a general class of monotone affine reward functions, which include arms with arbitrary number of states. The optimality of MP is proven in [20] when the arms' states follow non-identical Markov chains. The case of imperfect channel detection is studied in [21], and MP is found to be optimal when the false alarm probability of the channel state detector is below a certain value.

III. System Model

We consider an EH wireless network with N EH nodes and one AP, as depicted in Figure 1. Time is divided into TSs of constant duration, and the AP is in charge of scheduling K of the N nodes to the K available orthogonal channels at each TS. A node is *operative* at each TS with a fixed probability p independent over TSs and among nodes, and *inoperative* otherwise. We consider that a node is in the operative state if it has a data packet to transmit in its buffer and its channel to the AP is in a good state, while it is inoperative otherwise even if it is scheduled to a channel. The EH process is modelled as a Markov chain, which can be either in the harvesting or in the non-harvesting state, denoted by states 1 and 0, respectively. We denote by p_{ij} the transition probability from state i to j, and assume that $p_{11} \geq p_{01}$, that is, the EH process is positively correlated in time, and hence, if the EH process is in state i, it is more likely to remain in state i than switching to the other state. We denote by $E_i^s(n)$ and $E_i^h(n)$ the state of the EH process and the amount of energy harvested by node i, respectively, in TS n. The energy harvested in TS n is available for transmission in TS $n+1$. We assume that one fundamental unit of energy is harvested when the Markov process makes a

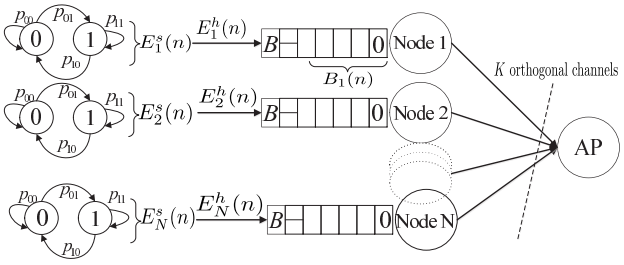


Figure 1. System model with N nodes and one AP. The EH process associated with each node is modeled by independent two-state Markov processes, the batteries have a capacity of B units, and $E_i^h(n)$ energy units are harvested by node i in TS n

transition to the harvesting state, that is, $E_i^h(n) = E_i^s(n+1)$[1]. Each node is equipped with a battery of capacity B, and we denote by $B_i(n) \in \{0, \dots, B\}$ the amount of energy stored in the battery of node i at the beginning of TS n. The state of node i in TS n, $S_i(n)$, is given by its battery level and EH process state, and we have that $S_i(n) = (E_i^s(n), B_i(n)) \in \{0,1\} \times \{0, \dots, B\}$. The system state is characterized by the joint states of all the nodes.

The system functions as follows. At the beginning of each TS, the AP schedules K out of N nodes, such that a single node is allocated to each channel. When a node is scheduled, if it is operative in that TS, i.e., it has data to transmit and its channel is in a good state, it transmits a data packet as well as the current state of its EH process to the AP over its scheduled channel. If it is not operative it transmits a status beacon to the AP, and backs off. We say that a node is *active* in a TS if it is scheduled by the AP and is operative; and hence, it transmits a data packet to the AP, otherwise we say that the node is *idle* in this TS, that is, the node is not scheduled or it is scheduled but is not operative. We denote by $\mathcal{K}(n)$ and $\mathcal{K}^a(n)$ the set of nodes scheduled by the AP, and the set of active nodes in TS n, respectively, where $\mathcal{K}^a(n) \subseteq \mathcal{K}(n)$.

We assume that the transmission rate of the nodes is a linear function of their transmit power. This is a typical assumption for low-power sensor nodes [22]. Note that, when the power-rate function is linear the total number of bits transmitted to the AP by any scheduling policy is maximised when an active node transmits at a constant power throughout the TS and use all the energy in its battery. Hence, we assume that when a node is active it transmits at the maximum rate throughout the TS, and uses up all the energy. To simplify the notation we normalise the power-rate function such that the number of bits transmitted by a node within a TS is equal to the total energy used for transmission. The expected throughput in TS n is then given by

$$R(\mathcal{K}(n)) = \mathbb{E}\left[\sum_{i \in \mathcal{K}^a(n)} B_i(n)\right] = p \sum_{i \in \mathcal{K}(n)} B_i(n). \tag{1}$$

The objective of the AP is to schedule the best set of nodes, $\mathcal{K}(n)$, at each TS in order to maximize the system throughput, without knowing which nodes are operative, the battery levels, or the EH states. The only information the AP receives is the EH state of the nodes that are active at each TS. Note that the AP also knows the battery state of the active nodes after transmission since they use all their energy.

A scheduling policy is an algorithm that schedules nodes at each TS n, based on the previous observations of the EH states and battery levels. The objective of the AP is to find the scheduling policy $\mathcal{K}(n)$, $\forall n \in [1, T]$, that maximizes the total discounted throughput, given by

$$\begin{aligned} \max_{\{\mathcal{K}(n)\}_{n=1}^{T}} \quad & \sum_{n=1}^{T} \beta^{n-1} R(\mathcal{K}(n)), \\ \text{s.t. } & B_i(n+1) = \min\{B_i(n) \\ & \quad + E_i^h(n), B\} \cdot \mathbb{1}_{i \notin \mathcal{K}^a(n)} + E_i^h(n) \cdot \mathbb{1}_{i \in \mathcal{K}^a(n)}, \end{aligned} \tag{2}$$

where $0 < \beta \leq 1$ is the discount factor, and $\mathbb{1}_a$ is the indicator function, defined as $\mathbb{1}_a = 1$ if a is true, and $\mathbb{1}_a = 0$, otherwise.

If the AP is somehow informed on the current state of all the nodes at each TS, this problem would be formulated as an MDP, and solved using LP or DP [13]. However, in practice transmitting all the nodes' states to the AP introduces further overhead and energy consumption; and hence, is not studied here. Accordingly, the appropriate model for our problem setting is a POMDP. It can be shown that a sufficient statistic for optimal decision making in a POMDP is given by the conditional probability of the system states given all the past actions and observations, which, in our problem, depends only on the number of TSs each node has been idle for, and on the realisation of each node's EH state last time it was active. Hence, we can reformulate the POMDP into an equivalent MDP with an extended state space. The belief states, that is, the states in the equivalent MDP, are characterized by all the past actions and observations. We denote by l_i and h_i the number of TSs that node i has been idle for, and the state of the EH process the last time it was active, respectively. The belief state of node i, $s_i(n)$, is given by $s_i(n) = (l_i, h_i)$, and the belief state of the whole system is the joint belief states of all the nodes. In TS n, the belief state of node i is updated as $s_i(n+1) = (0, E_i^s(n))$, if $i \in \mathcal{K}^a(n)$, and as $s_i(n+1) = (l_i + 1, h_i)$, otherwise. That is, at each TS, l_i is set to 0 if node i is active, and increased by one if it is idle. In principle, since the number of TSs a node can be idle is unbounded, the state space of the equivalent MDP is infinite, and hence, the POMDP in (2) is hard to solve, and numerical techniques such as DP turn out to be PSPACE-complete. In Sections IV and V, we focus on two particular settings of the problem in (2), and show that, under certain assumptions regarding the EH processes and the battery sizes, there exist optimal low-complexity scheduling policies.

[1]Note that since the objective is to maximise the expected throughput, the results of this paper can be generalised to a broader class of two-state Markovian EH processes in which the amount of energy harvested in each state of the EH process is an independent and identically distributed (iid) random variable, and the expected amount of energy harvested in their harvesting state is larger than that in the non-harvesting state. However, the EH model used in this paper captures the random nature of the energy arrivals, and has been used in [5], [9], [10], [12].

IV. Non Simultaneous Energy Harvesting and Data Transmission

In this section we assume that the nodes are not able to harvest energy and transmit data simultaneously, and that if node i is active in TS $n-1$, then its EH state in TS n, $E_i^s(n)$, is either 0 or 1 with probabilities e_0 and e_1, respectively, independent of the EH state in TS $n-1$, where $e_0 \leq \frac{p_{10}}{p_{01}+p_{10}}$. These assumptions may account for nodes equipped with electromagnetic energy harvesters in which the same antenna is used for harvesting as well as transmission; and hence, it is not possible to transmit data and harvest energy simultaneously, and the RF hardware has to be reset into the harvesting mode after each transmission.

Since the EH process is either 0 or 1 with fixed probabilities after a node transmits, the EH process states of active nodes are not relevant. As a consequence, the belief state of a node, $s_i(n)$, is characterized only by the number of TSs the node has been idle for, l_i. There is a one-to-one correspondence between l_i and the expected amount of energy in the battery of node i; therefore, we redefine the belief state, $s_i(n)$, as the expected battery level of node i in TS n, normalised by the battery capacity. The expected throughput in (1) can be rewritten as

$$R(\mathcal{K}(n)) = pB \sum_{i \in \mathcal{K}(n)} s_i(n). \tag{3}$$

Notice that $s_i(n)$ in (3) is normalised, i.e., $s_i(n) \in [0,1]$. If the belief states of all the nodes are 1, that is, the AP is certain that all the batteries are full, the expected throughput would be $p \cdot B \cdot K$; whereas if the belief states are 0 for all the nodes, the expected throughput would be 0.

Due to the Markovian nature of the EH processes, the future belief state is only a function of the current belief state and the scheduling policy. If a node is active in TS n, since it uses all the available energy in the battery and does not harvest energy, the belief state is set to 0 in TS $n+1$. If a node is not active in TS n, then the belief state evolves according to the belief state transition function $\tau(\cdot)$. The belief state of node i in TS $n+1$ is

$$s_i(n+1) = \begin{cases} \tau(s_i(n)) & \text{if } i \notin \mathcal{K}^a(n), \\ 0 & \text{if } i \in \mathcal{K}^a(n), \end{cases} \tag{4}$$

Property 1. *The belief state transition function, $\tau(\cdot)$, is a monotonically increasing contracting map, that is, $\tau(s_i(n)) > \tau(s_j(n))$ if $s_i(n) > s_j(n)$, and $\|\tau(s_i(n)) - \tau(s_j(n))\| \leq \|s_i(n) - s_j(n)\|$.*

Proof. The proof is in Appendix A. □

Note that the assumption $p_{11} \geq p_{01}$ is a necessary condition for Property 1. We denote by $\mathbf{s}(n) = (s_1(n), \ldots, s_N(n))$ the belief vector in TS n, which contains the belief states of all the nodes, and by $\mathbf{s}_{\mathcal{E}}(n)$ the belief vector of the nodes in set $\mathcal{E}$. For the sake of clarity we drop the n from $\mathbf{s}(n)$, and $\mathbf{s}_{\mathcal{E}}(n)$ when the time index is clear from the context. We denote the expected throughput by $R(\mathbf{s}_{\mathcal{E}})$ if the belief vector is $\mathbf{s}$ and nodes in $\mathcal{E}$ are scheduled.

The probability that a particular set of nodes, $\mathcal{K}^a(n) \subseteq \mathcal{K}(n)$, is active while the rest of the scheduled nodes remain idle in TS n is a function of the cardinality of $\mathcal{K}^a(n)$ and the probability that a node is operative, p. For $a \triangleq |\mathcal{K}^a(n)|$ we denote this probability by

$$q(a, K) \triangleq (1-p)^{K-a} p^a, \tag{5}$$

where a and $K-a$ are the number of active and idle nodes in $\mathcal{K}(n)$, respectively.

The AP is interested in finding the scheduling policy π, which, in each TS, schedules the nodes according to the belief vector, $\mathbf{s}(n)$, that is $\mathcal{K}(n) = \pi(\mathbf{s}(n))$, such that the expected throughput within the time horizon T is maximised. The associated optimization problem is expressed through the Bellman value functions,

$$\begin{aligned} V_n^{\pi}(\mathbf{s}) \quad = \quad & R(\mathbf{s}_{\pi(\mathbf{s})}) + \beta \sum_{\mathcal{E} \subseteq \pi(\mathbf{s})} q(|\mathcal{E}|, K) \\ & \times V_{n+1}^{\pi}((s_1(n+1), \ldots, s_j(n+1) = 0, \\ & \ldots, s_i(n+1) = \tau(s_i(n)), \ldots)), \end{aligned} \tag{6}$$

where the sum is over all possible sets of active nodes, $\mathcal{E}$, among the scheduled nodes, $\mathcal{K}(n) = \pi(\mathbf{s}(n))$, and nodes j and i are active and idle, respectively. The optimal policy, π^*, is the one that maximises (6).

A. Definitions

Definition 1. (Myopic policy) At TS n the myopic policy (MP) schedules the K nodes that maximise the expected instantaneous reward function, $R(\cdot)$. For the reward function in (3) the MP schedules the K nodes with the highest belief states.

MP schedules the nodes similarly to a round robin (RR) policy that orders the nodes according to the time they have been idle for, and at each TS schedules the nodes with the highest idle time values. If a node is active in this TS, it is sent to the bottom of this ordered list in the next TS. If a node is idle it moves forward in the order. Notice that due to the monotonicity of $\tau(\cdot)$ the order of the idle nodes is preserved.

We denote by $\mathbf{s}_{\Pi} = (s_{\Pi(1)}, \ldots, s_{\Pi(N)})$, the permutation of the vector $\mathbf{s}$, where $\Pi(\cdot)$ is a permutation function, by $\mathbf{s}_{\Pi}^K = (s_{\Pi(1)}, \ldots, s_{\Pi(K)})$ the vector containing the first K elements of $\mathbf{s}_{\mathbf{\Pi}}$. By $\mathcal{S}_{\Pi}^K = \{\Pi(1), \cdots, \Pi(K)\}$, the set of indices of the nodes in positions from 1 to K in vector $\mathbf{s}_{\Pi}$. We say that a vector is ordered if its elements are in decreasing order. We denote by $\mathring{\Pi}$ the permutation that orders a vector, that is, the vector $\mathbf{s}_{\mathring{\Pi}}$ is ordered, i.e., $s_{\mathring{\Pi}(1)} \geq s_{\mathring{\Pi}(2)} \geq \ldots \geq s_{\mathring{\Pi}(N)}$. We denote the vector operator that first orders the vector $\mathbf{s}_{\mathcal{E}}$ of $|\mathcal{E}|$ components, and then applies $\tau(\cdot)$ to each of the components of the resulting vector by $\mathrm{T}(\mathbf{s}_{\mathcal{E}}) = (\tau(s_{\mathring{\Pi}(1)}), \cdots, \tau(s_{\mathring{\Pi}(|\mathcal{E}|)}))$, with $\mathring{\Pi}(i) \in \mathcal{E}, 1 \leq i \leq |\mathcal{E}|$. Note that due to the monotonicity of $\tau(\cdot)$ the vector $\mathrm{T}(\mathbf{s}_{\mathcal{E}})$ is always ordered. Finally, we denote the zero vector of length k by $\mathbf{0}(k)$.

Definition 2. (Pseudo value function)

$$\begin{aligned} & W_n(\mathbf{s}_{\Pi}) \triangleq R(\mathbf{s}_{\Pi}^K) + \beta \sum_{\mathcal{E} \subseteq \mathcal{S}_{\Pi}^K} q(|\mathcal{E}|, K) W_{n+1}([\mathrm{T}(\mathbf{s}_{\overline{\mathcal{E}}}) \ , \mathbf{0}(|\mathcal{E}|)]), \\ & W_T(\mathbf{s}_{\Pi}) \triangleq R(\mathbf{s}_{\Pi}^K), \end{aligned} \tag{7}$$

where $[\cdot]$ is the vector concatenation operator.

The pseudo value function is characterized solely by the belief vector $\mathbf{s}$ and its initial permutation Π. In TS n, the first K nodes according to permutation Π are scheduled, and the nodes are scheduled according to MP thereafter. The belief vector in TS $n+1$ is $\mathbf{s}^{\circ}_{\Pi}(n+1) = [\mathrm{T}(\mathbf{s}_{\overline{\mathcal{E}}}), \mathbf{0}(|\mathcal{E}|)]$, where $\mathcal{E}$ is the set of active nodes in TS n, and, since $\mathrm{T}(\cdot)$ implicitly orders the output vector, $\mathbf{s}^{\circ}_{\Pi}(n+1)$ is ordered. Hence, the nodes that are active in TS n have belief state 0 in TS $n+1$, and are moved to the rightmost position in the belief vector. If the vector $\mathbf{s}_{\Pi}$ is ordered, (7) corresponds to the value function of MP, that is, corresponds to(6) where π is MP.

Definition 3. (Swap permutations)

- A permutation Π is an i,j-swap of permutation $\widehat{\Pi}$ if $\Pi(k) = \widehat{\Pi}(k)$, for $\forall k \neq \{i,j\}$, and $\Pi(j) = \widehat{\Pi}(i)$ and $\Pi(i) = \widehat{\Pi}(j)$. That is, all the nodes but those in positions i and j are in the same positions in $\mathbf{s}_{\Pi}$ and $\mathbf{s}_{\widehat{\Pi}}$, and the nodes in positions i and j are swapped.
- A permutation Π is an i,j-swap if $\Pi(k) = k$, for $\forall k \neq \{i,j\}$, and $\Pi(i) = j$ and $\Pi(j) = i$. That is, all the nodes but those in positions i and j are in the same position in $\mathbf{s}$ and $\mathbf{s}_{\Pi}$, and the nodes in positions i and j are swapped.

Definition 4. (Regularity [20]) A function $f(\mathbf{x})$, $f : \mathbb{R}^k \to \mathbb{R}$ and $\mathbf{x} = (x_1, \ldots, x_k)$, is said to be *regular* if it is symmetric, monotonically increasing, and decomposable.

- Symmetry: $f(\mathbf{x})$ is symmetric if $f(\ldots, x_i, \ldots, x_j, \ldots) = f(\ldots, x_j, \ldots, x_i, \ldots)$.
- Monotonicity: $f(\mathbf{x})$ is monotonically increasing in each of its components, that is, if $x_j \leq \tilde{x}_j$ then $f(\ldots, x_j, \ldots) \leq f(\ldots, \tilde{x}_j, \ldots)$.
- Decomposability: $f(\mathbf{x})$ is decomposable if $f(\ldots, x_j, \ldots) = x_j f(\ldots, 1, \ldots) + (1 - x_j) f(\ldots, 0, \ldots)$.

Definition 5. (Boundedness) A function $f(\mathbf{x})$, $f : \mathbb{R}^k \to \mathbb{R}$ and $\mathbf{x} = (x_1, \ldots, x_k)$, is said to be *bounded* if $\Delta_l \leq f(\ldots, 1, \ldots) - f(\ldots, 0, \ldots) \leq \Delta_u$.

We note that the expected throughput $R(\cdot)$ is a linear function of the belief vector, which has bounded elements, and that all the nodes that are scheduled have the same coefficient; hence, $R(\cdot)$ is a bounded regular function. The pseudo value function, $W_n(\cdot)$, is symmetric, that is,

$$W_n(\mathbf{s}_{\Pi}) = W_n(\mathbf{s}_{\widehat{\Pi}}), \tag{8}$$

where Π is a i,j-swap permutation of $\widehat{\Pi}$, and $j, i \leq K$ or $j, i > K$. To see this we can use the symmetry of $R(\cdot)$, and the fact that $\mathrm{T}(\cdot)$ orders the belief vector in decreasing order.

B. Proof of the optimality of MP

In this section we prove the optimality of MP under the assumptions that $\tau(\cdot)$ is a monotonically increasing contracting map[2], and that $R(\cdot)$ is a bounded regular function. Hence, the results in this section can be applied to a boarder class of EH processes and reward functions than those studied in this paper.

The proof of the optimality of MP is structured as follows: Lemma 1 gives sufficient conditions for the optimality of MP in TS n, given that MP is optimal from TS $n+1$ onwards. In Lemma 2 we show that the difference in the values of the pseudo value function of two different vectors is bounded. In particular, we bound the difference between the value functions of two belief vectors $\mathbf{s}^{\circ}_{\Pi}$ and $\tilde{\mathbf{s}}^{\circ}_{\Pi}$, which are both ordered and different only for the belief state of node i. In Lemma 3 we show that, under certain conditions, the sufficient conditions for the optimality of MP given in Lemma 1 hold. In Theorem 1 we prove the optimality of MP, and Theorem 2 proves the optimality of MP for the EH scheduling problem studied in this section.

Lemma 1. *Assume that MP is optimal from TS $n+1$ until TS T. A sufficient condition for the optimality of MP in TS n is*

$$W_n(\mathbf{s}) \geq W_n(\mathbf{s}_{\Pi}), \tag{9}$$

for any Π that is an i,j-swap, with $s_j \geq s_i$ and $j \leq i$.

Proof. To prove that a policy is optimal, we need to show that it maximizes (6). By assumption MP is optimal from TS $n+1$ onwards; and hence, it is only necessary to prove that scheduling any set of nodes and following MP thereafter is no better than following MP directly in TS n. The value function corresponding to the latter policy is $W_n([\mathbf{s}_{\mathcal{O}}, \mathbf{s}_{\overline{\mathcal{O}}}])$, where $\mathbf{s}_{\mathcal{O}}$ contains the K nodes with the highest belief states in $\mathbf{s}$, and $\mathbf{s}_{\overline{\mathcal{O}}}$ contains the rest of the nodes not necessarily ordered. The value function corresponding to the former policy is $W_n([\mathbf{s}_{\mathcal{U}}, \mathbf{s}_{\overline{\mathcal{U}}}])$, where $\mathbf{s}_{\mathcal{U}}$ contains the K nodes scheduled in TS n, and $\mathbf{s}_{\overline{\mathcal{U}}}$ is the set of the remaining nodes. There exist at least a pair of nodes s_i and s_j such that, $j \in \overline{\mathcal{U}}$ and $j \notin \overline{\mathcal{O}}$, $i \in \mathcal{U}$ and $i \notin \mathcal{O}$, and $s_j \geq s_i$. By swapping each pair of such nodes, that is, swapping $j \in \overline{\mathcal{U}}$ for $i \in \mathcal{U}$, we can obtain $W_n([\mathbf{s}_{\mathcal{O}}, \mathbf{s}_{\overline{\mathcal{O}}}])$ from $W_n([\mathbf{s}_{\mathcal{U}}, \mathbf{s}_{\overline{\mathcal{U}}}])$ through a cascade of inequalities using (9). Accordingly, $W_n([\mathbf{s}_{\mathcal{O}}, \mathbf{s}_{\overline{\mathcal{O}}}])$ is an upper bound for any $W_n([\mathbf{s}_{\mathcal{U}}, \mathbf{s}_{\overline{\mathcal{U}}}])$, and, hence, MP is optimal. □

In general, the optimality of a policy can be established by showing that it maximises the value function in (6). Lemma 1 shows that, under certain conditions, the optimality of MP can be established through the pseudo value function. In particular, under the conditions of Lemma 1, if swapping a node in the belief vector with another node with a lower position and a lower belief state does not decrease the pseudo value function, then MP is optimal.

Lemma 2. *Consider a pair of belief vectors $\mathbf{s}$ and $\tilde{\mathbf{s}}$, that differ only in one element, that is, $s_i = \tilde{s}_i$ for $\forall i \neq j$ and $s_j \geq \tilde{s}_j$. If $R(\cdot)$ is a bounded regular function, $\tau(\cdot)$ a monotonically increasing contracting map, and $\beta \leq 1$, then we have*

$$W_n(\mathbf{s}^{\circ}_{\Pi}) - W_n(\tilde{\mathbf{s}}^{\circ}_{\Pi}) \leq \Delta_u (s_j - \tilde{s}_j) u(n), \tag{10}$$

where $u(n) \triangleq \sum_{i=0}^{T-n} (\beta(1-p))^i$.

Proof. See Appendix B. □

[2]Our results can also be applied to the case in which the state transition function is a monotonically increasing contracting map with parameter α, that is, $\tau(s_i(n)) > \tau(s_j(n))$ if $s_i(n) > s_j(n)$, and $\|\tau(s_i(n)) - \tau(s_j(n))\| \leq \alpha \|s_i(n) - s_j(n)\|$, if $0 \leq \alpha \cdot \beta \leq 1$.

The result of Lemma 2 establishes that increasing the belief state of a node j from $\tilde{s}_j$ to s_j may increase the value of the pseudo value function, which is bounded by a linear function of the increase in the belief, $s_j - \tilde{s}_j$, and the function $u(n)$, which decreases with n and corresponds to the maximum accumulated loss from TS n to TS T.

Lemma 3. *Consider two belief vectors* $\mathbf{s}$ *and* $\mathbf{s}_\Pi$*, such that permutation* Π *is an* i,j*-swap, and* $s_j \geq s_i$ *for some* $j \leq i$*. If* $R(\cdot)$ *is a bounded regular function,* $\tau(\cdot)$ *a monotonically increasing contracting map, and* $\beta \leq 1$*, then*

$$W_n(\mathbf{s}) - W_n(\mathbf{s}_\Pi) \geq 0 \text{ if } \Delta_l \geq \Delta_u \beta p \frac{1-(\beta(1-p))^{T+1}}{1-\beta(1-p)}. \tag{11}$$

Proof. See Appendix C. □

Theorem 1. *If* $R(\cdot)$ *is a bounded regular function,* $\tau(\cdot)$ *a monotonically increasing contracting map,* $\beta \leq 1$*, and* $\Delta_l \geq \Delta_u \beta p \frac{1-(\beta(1-p))^{T+1}}{1-\beta(1-p)}$*, then MP is the optimal policy, that is, it maximises (6).*

Proof. The proof is done by backward induction. We have already shown that MP is optimal at TS T. Then we assume that MP is optimal from TS $n+1$ until TS T, and we need to show that MP is optimal at TS n. To show that MP is optimal at TS n, using Lemma 1, we only need to show that (9) holds. This is proven in Lemma 3, which completes the proof. □

The result of Theorem 1 holds for any $R(\cdot)$ that is a bounded regular function. The reward function of the scheduling problem studied in this paper, that is, the sum expected throughput in (3), is a bounded regular function, and we have $\Delta_u = \Delta_l = pB$. Finally, we can state the optimality of the MP for the EH problem studied in this section in the following theorem.

Theorem 2. *If the reward function* $R(\cdot)$ *is defined as the expected throughput in (3), and the transition probabilities among the EH states satisfy* $p_{11} \geq p_{01}$ *and* $e_0 \leq \frac{p_{10}}{p_{01}+p_{10}}$*, then MP is the optimal policy, that is, it maximises the sum discounted throughput in (6).*

V. Simultaneous Energy Harvesting and Data Transmission with Batteryless Nodes

Now we consider another special case of the general system model introduced in Section III. We assume that the nodes cannot store energy, and the harvested energy is lost if not used immediately. This is practically relevant for low-cost batteryless nodes with physical size constraints. Energy available for transmission in TS n is equal to the energy harvested in TS $n-1$, that is, $B_i(n) = E_i^h(n-1)$. We denote by $s_i(n)$ the belief state of node i at TS n, which is the expected energy available for transmission, that is, the probability that the node is in the harvesting state. The belief state transition probabilities are

$$s_i(n+1) = \begin{cases} \tau(s_i(n)) & \text{if } i \notin \mathcal{K}^a(n), \\ p_{11} & \text{if } i \in \mathcal{K}^a(n) \text{ w.p. } s_i(n), \\ p_{01} & \text{if } i \in \mathcal{K}^a(n) \text{ w.p. } 1-s_i(n), \end{cases} \tag{12}$$

where $\tau(s) = (p_{11} - p_{01})s + p_{01}$, and since $p_{11} \leq p_{01}$, it is a monotonically increasing affine function. This implies that if $s_i \geq s_j$ then $\tau(s_i) \geq \tau(s_j)$, that is, the order of the idle nodes is preserved. We note that $i \in \mathcal{K}^a(n)$ with probability p, if $i \in \mathcal{K}(n)$. The problem is to find a scheduling policy, $\mathcal{K}(n)$, such that the expected discounted sum throughput, $R(\mathcal{K}(n)) = p \sum_{i \in \mathcal{K}(n)} s_i(n)$, is maximised over a time horizon T under the battery constraint.

A. Definitions

We define the pseudo value function as follows

$$\begin{aligned} W_n(\mathbf{s}_\Pi) &\triangleq R(\mathbf{s}_\Pi^K) + \beta \sum_{\mathcal{E} \subseteq \mathcal{S}_\Pi^K} \sum_{l_\mathcal{E} \in \{0,1\}^{|\mathcal{E}|}} h(l_\mathcal{E}, K) \\ &\quad \times W_{n+1}\left(\mathbf{P}_{11}\left(\Sigma l_\mathcal{E}\right), \boldsymbol{\tau}(\mathbf{s}_{\overline{\mathcal{E}}}), \mathbf{P}_{01}\left(\overline{\Sigma} l_\mathcal{E}\right)\right), \\ W_T(\mathbf{s}_\Pi) &\triangleq R(\mathbf{s}_\Pi^K), \end{aligned} \tag{13}$$

where we denote the set of active nodes by $\mathcal{E}$ and the ith active node by $\mathcal{E}(i)$. We define $l_\mathcal{E} = (l_{\mathcal{E}(1)}, \ldots, l_{\mathcal{E}(|\mathcal{E}|)})$, such that $l_{\mathcal{E}(i)} = 1$ if the EH process of the corresponding node is in the harvesting state and $l_{\mathcal{E}(i)} = 0$ otherwise. We define the function $h(l_\mathcal{E}, K) \triangleq q(|\mathcal{E}|, K) \prod_{j \in \mathcal{E}} s_j^{l_j} (1-s_j)^{(1-l_j)}$, where $q(|\mathcal{E}|, K)$ is defined in (5). We denote by $\mathbf{P}_{01}(a)$ and $\mathbf{P}_{11}(a)$ the vectors $(p_{01}, \ldots, p_{01})$ and $(p_{11}, \ldots, p_{11})$, respectively, of length a, and we define $\Sigma l_\mathcal{E} \triangleq \sum_{i \in \mathcal{E}} l_i$, and $\overline{\Sigma} l_\mathcal{E} \triangleq |\mathcal{E}| - \sum_{i \in \mathcal{E}} l_i$. The operator $\boldsymbol{\tau}(\cdot)$ applies the mapping $\tau(\cdot)$ to all its components. The pseudo value function schedules the nodes according to permutation Π, and if $\mathbf{s}_\Pi$ is ordered (13) is the value function of MP.

We note that swapping the order of two nodes that are scheduled does not change the value of the pseudo value function, that is, the pseudo value function is symmetric. This property is similar to that in (8), but only for $i,j \leq K$. Similarly to [17] and [18], the mapping $\tau(\cdot)$ is linear, and hence, the pseudo value function is affine in each of its elements. This implies that, if Π is an i,j-swap of $\widehat{\Pi}$, then

$$\begin{aligned} &W_n\left(\mathbf{s}_\Pi\right) - W_n(\mathbf{s}_{\widehat{\Pi}}) \\ &\quad = (s_{\Pi(j)} - s_{\Pi(i)}) \Big(W_n(\ldots, s_{\Pi(j)} = 1, \ldots, s_{\Pi(i)} = 0, \ldots) \\ &\qquad - W_n(\ldots, s_{\Pi(j)} = 0, \ldots, s_{\Pi(i)} = 1, \ldots) \Big). \end{aligned} \tag{14}$$

MP has the following structure: it schedules the nodes whose EH processes are more likely to be in the harvesting state at each TS. Initially, nodes are ordered according to an initial belief. If a node is active, it is sent to the first position of the queue if it is in the harvesting state, and to the last position if it is in the non-harvesting state. The idle nodes are moved forward in the queue. Due to the monotonicity of $\tau(\cdot)$, MP continues scheduling a node until it is active and its EH process is in the non-harvesting state.

B. Proof of the optimality of MP

We note that the result of Lemma 1 is applicable in this case. If Lemma 4 holds, the same arguments as in Theorem 1 can be used to prove the optimality of MP.

Lemma 4. *Let Π be an i,j-swap, and consider a permutation $\widehat{\Pi}$, such that $\widehat{\Pi}(k) = k-1$, for $\forall k \neq 1$ and $\widehat{\Pi}(1) = N$. If $s_j \geq s_i$ for some $j \leq i$, then we have the inequalities*

$$1 + W_n(\mathbf{s}_{\widehat{\Pi}}) \geq W_n(\mathbf{s}), \tag{15a}$$

$$W_n(\mathbf{s}) \geq W_n(\mathbf{s}_{\Pi}). \tag{15b}$$

Proof. The proof follows from the similar arguments as in [18]. In particular, we use backward induction in (15a) and (15b), and a sample-path argument. A sketch of the proof is provided in Appendix D. □

Note that (15a) and (15b) are similar to (10) and (11), respectively.

Theorem 3. *If the reward function is $R(\mathcal{K}(n)) = p\sum_{i\in\mathcal{K}(n)} s_i(n)$, and the EH processes of the nodes satisfy that $p_{11} \geq p_{01}$, MP is the optimal policy, that is, it maximises the sum discounted throughput in (6).*

Proof. Theorem 3 can be proven by using the same arguments as in Theorem 1 and Lemmas 1 and 4. □

Remark 1. This problem is similar to the opportunistic multi-channel access problem studied in [17]–[20], with imperfect channel sensing, such that, at each attempt, a channel can not be sensed with probability $1-p$, independent of its channel state. While the MP has been proven to be optimal in the case of perfect channel sensing, i.e., $p = 1$, [18], the case with sensing errors, i.e., $p \neq 1$, has not been considered in the literature. We also note that this model of imperfect channel detection is different from that in [21].

Remark 2. Using similar techniques as in [17] the MP optimality results of Sections IV and V can be extended from the finite horizon discounted reward criteria to the infinite horizon with discounted reward, and to the infinite horizon with average reward criteria.

VI. Upper Bound on the Performance of the Optimal Scheduling Policy

In this section we derive an upper bound on the performance of the optimal policy for the general case described in Section III under the average reward criteria and infinite time horizon. The RMAB problem with an infinite horizon discounted reward criteria is studied in [23], and it is shown that an upper bound can be computed in polynomial time using LP.

The decision of scheduling a node in a TS affects the scheduling of the other nodes in the same TS, since exactly K nodes have to be scheduled at each TS. Whittle [15] proposed to relax the original problem constraint, and impose instead that the number of nodes that are scheduled at each TS is K *on average*. In the relaxed problem, since the nodes are symmetric, one can decouple the original RMAB problem into N RMAB problems, one for each node. As before, we denote by $s = (l,h) \in \mathcal{W}$ the belief state of a node, where l is the number of TSs the node has been idle for, and h the EH state last time the node was scheduled, and $\mathcal{W}$ the belief state space. We denote by $\pi(s)$ the probability that a node is scheduled if it is in state s, by $p(s)$ the steady state probability of state s, and by $p_{\tilde{s},s}(a)$ the state transition probability function from state $\tilde{s}$ to s if action $a \in \{0,1\}$ is taken, where if $a = 1$ the node is scheduled in this TS, and $a = 0$, otherwise. The optimization problem is

$$\begin{aligned} \max_{\pi(s),p(s)} & \sum_{s\in\mathcal{W}} R(s)\pi(s)p(s) \\ \text{s.t. } & p(s) = \sum_{\tilde{s}\in\mathcal{W}} p(\tilde{s})[(1-\pi(s))p_{\tilde{s},s}(0) + \pi(s)p_{\tilde{s},s}(1)], \\ & \sum_{s\in\mathcal{W}} \pi(s)p(s) = \frac{K}{N}, \text{ and } \sum_{s\in\mathcal{W}} p(s) = 1, \end{aligned} \tag{16}$$

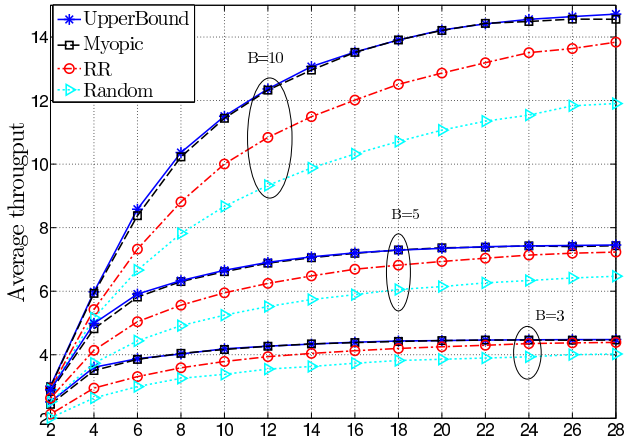
where $0 \leq \pi(s),\ p(s) \leq 1$, and $R(s)$ is the expected throughput of a node if it is in state s. Note that the constraint $\sum \pi(s)p(s) = \frac{K}{N}$ imposes that the node is scheduled on average every $\frac{N}{K}$ TSs. This implies that, for $p = 1$, the maximum time a node can be idle is finite, and hence, the state space $\mathcal{W}$ is finite. For the case in which $p \neq 1$, one can truncate the state space by bounding the maximum time a node can be idle, i.e., imposing that l is bounded. The problem (16) has linear objective functions and constrains, the state space is finite, therefore it can be solved in polynomial time with LP.

VII. Numerical Results

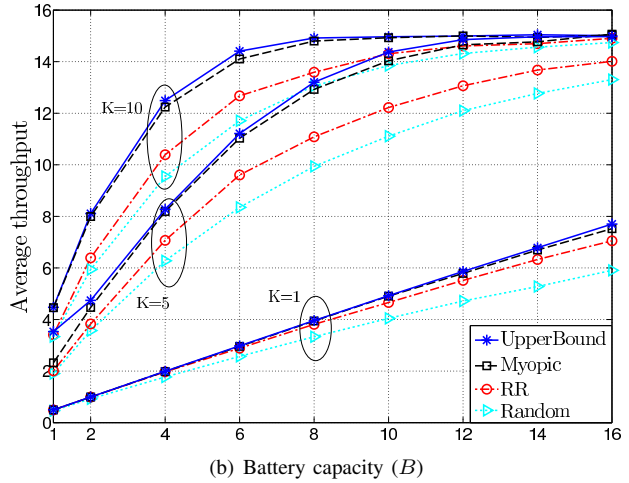
In this section we study the performances of different scheduling policies for the general case described in Section III numerically, in which nodes can harvest energy and transmit at the same TS, and the EH process state is not affected by the scheduling policy. In particular, we consider MP which is optimal for the cases studied in Sections IV and V, the RR policy, which schedules the nodes in a cyclic fashion according to an initial random order, and a random policy, which at each TS schedules K random nodes regardless of the history. We measure the performance of the scheduling policies as the average throughput per TS over a time horizon of $T = 1000$, that is, we consider $\beta = 1$ and normalise (2) by T. We perform 100 repetitions for each experiment and average the results. We assume, unless otherwise stated, a total of $N = 30$ EH nodes, $K = 5$ available channels, and a probability $p = 0.5$ for a node to be operative in each TS. We assume that all the nodes and EH processes are symmetric, the batteries have a capacity of $B = 5$ energy units, and the transition probabilities of the EH processes are $p_{11} = p_{00} = 0.9$. Notice that, on average, each node is scheduled every $\frac{N}{K}$ TSs. Hence, if $\frac{N}{K}$ is large the nodes remain idle for larger periods. This implies that when $\frac{N}{K}$ is large, since the nodes harvest over many TSs without being scheduled, there are more energy overflows in the system. In the numerical results we have included the infinite horizon upper bound of Section VI, which for large T is tight to the upper bound of the finite horizon case.

In Figure 2(a) we investigate the impact of the number of nodes on the throughput, when the number of available channels, K, is fixed. The throughput increases with the number of nodes, and, due to the battery overflows, saturates when the number of nodes is large. By increasing the battery capacity, hence reducing the battery overflows, the throughput saturates with a higher number of nodes and at higher value. We observe that MP has a performance close to that of the upper bound, the random policy has a lower performance than the others; and the gap between different curves wides wit the increase in the battery capacity.

In Figure 2(b) we investigate the effect of the battery capacity, B, on the system throughput when the number of nodes is fixed. Clearly, the larger the battery capacity the fewer battery overflows will occur. The throughput increases with the battery capacity, and due to the limited amount of energy that the nodes can harvest, it saturates at a certain value. By increasing the number of available channels, K, which also reduces the battery overflow, the throughput saturates more quickly as a function of the battery capacity, but at higher values. The performances of the scheduling policies are similar to those observed in Figure 2(a).

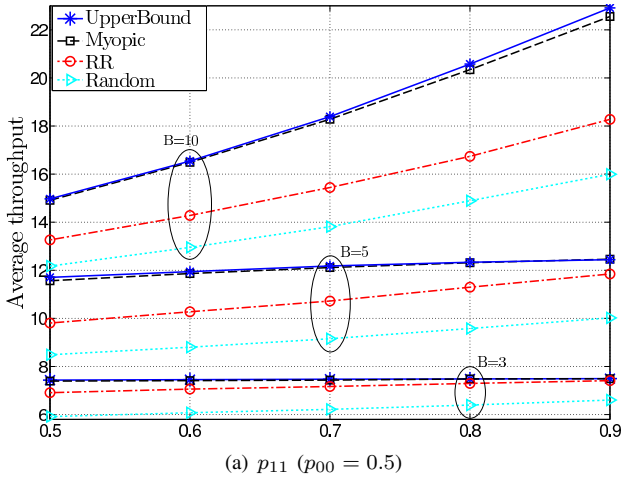


(a) N/K

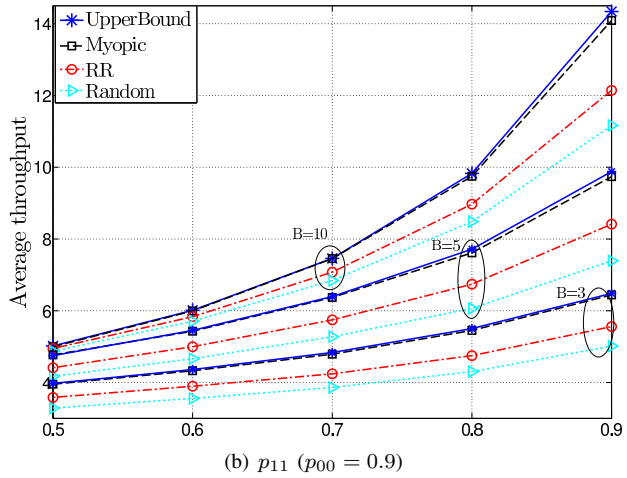


(b) Battery capacity (B)

Figure 2. (a) Average throughput vs. number of nodes, N, with $K = 5$ channels, and battery capacity $B = 3, 5, 10$, and (b) average throughput vs. battery capacity, B, for $N = 30$ nodes, and $K = 1, 5, 10$ channels.



(a) p_{11} ($p_{00} = 0.5$)



(b) p_{11} ($p_{00} = 0.9$)

Figure 3. Average throughput for different EH process transition probabilities, for $N = 30$ nodes, $K = 5$ available channels, and battery capacity $B = 3, 5, 10$.

Figure 3 shows the average throughput for different EH process transition probabilities. We note that the amount of energy arriving to the system increases with p_{11} and decreases with p_{00}. As expected, we observe in Figure 3 the throughput increases with p_{11}, and the throughput values in Figure 3(a) are notably higher than those in Figure 3(b). MP is a policy which maximises the immediate throughput at each TS, and does not take into account the future TSs. We observe in Figure 3(b) for $B = \{5, 10\}$ and in Figure 3(a) for $B = 10$ that, if the EH state has low correlation across TSs, that is, $p_{11} = \{0.5, 0.6\}$, the throughput obtained with MP is similar to that of the upper bound. On the contrary, if it has high correlation across TSs, that is $p_{11} = \{0.8, 0.9\}$, the throughput falls below the upper bound. This is due to the fact that when the state transitions have low correlation it is difficult to reliably predict the impact of the actions on the future rewards, and no transmission strategy can improve upon MP. Our numerical results indicate, that even in scenarios in which the MP cannot be shown to be theoretically optimal, it performs very close to the upper bound, obtained for an infinite horizon problem.

VIII. Conclusions

We have studied a scheduling problem in a multi-access communication system with EH nodes, in which the harvested energy at each node is modeled as a Markov process. At each TS a node is operative with a certain probability, which may model the random data availability at the nodes, or the state of their channels to the AP. We have modeled the system as an RMAB problem, and shown the optimality of MP in two settings: i) when the nodes cannot harvest energy and transmit simultaneously and the EH process state is independent of the past states after a node is active; ii) when the nodes have no battery. We have also provided an upper bound for the general setting on the average throughput for the infinite horizon problem, and compared the performance of MP to that of the upper bound numerically. The results of this paper suggest that although the optimal scheduling in large EH networks requires high computational complexity, in some cases there exist simple and practical scheduling policies that have almost optimal performance. This can have an impact on the design of scheduling policies for large low-power wireless sensor networks equipped with energy harvesting devices and limited storage. In particular, these results are interesting for network deployments which have an inherent symmetry, such as vibration-based EH nodes deployed in industrial machines of a factory in a star topology with the AP located at the centre, e.g., in the factory roof.

Appendix A

We denote the probability that the battery of a node is not full if the node has been idle for the last n TSs by $p_{nf}(n)$. It is easy to note that $p_{nf}(n)$ is a decreasing function of n. If the node has been idle n TSs, we denote the probability of the EH process being in state 0 and 1, by $p_0(n) \triangleq p_{10} + p_0(n-1)(p_{11}-p_{01})$ and $p_1(n) \triangleq 1-p_0(n)$, respectively. We set $p_0(0)=e_0$. Since $p_{11} \geq p_{01}$ and $e_0 \leq \frac{p_{10}}{p_{01}+p_{10}}$, $p_0(n)$ monotonically increases to the steady state distribution ([24, Appendix B]).

We denote the belief state of a node that has been idle for n TSs by z_n. If the node has been idle for $n+1$ TSs, the expected battery level is $z_{n+1}=\tau(z_n)=z_n+\frac{p_{nf}(n)}{B}(p_{01}p_0(n)+p_{11}p_1(n))$, which is a monotonically increasing function. If $n \geq m$, then $z_n \geq z_m$ and $\tau(z_n) \geq \tau(z_m)$. By applying the definition of $p_1(n)$, we get $z_{n+1}=z_n+\frac{p_{nf}(n)}{B}(p_{11}-p_0(n)(p_{11}-p_{00}))$. If we assume that $n \geq m$, we have

$$\begin{aligned}\|\tau(z_n)-\tau(z_m)\| &= z_n - z_m + \frac{p_{nf}(n)}{B}(p_{11}-p_0(n)(p_{11}-p_{01})) \\ &\quad - \frac{p_{nf}(m)}{B}(p_{11}-p_0(m)(p_{11}-p_{01})) \\ &\leq z_n - z_m - \frac{p_{nf}(n)}{B}(p_{11}-p_{01})(p_0(n)-p_0(m)) \\ &\leq z_n - z_m,\end{aligned}$$

where the first inequality follows since $p_{nf}(n) \leq p_{nf}(m)$, and the second inequality follows since $p_0(n)$ is monotonically increasing and $p_{11} \geq p_{01}$.

Appendix B

The proof uses backward induction. We denote by $\mathcal{S}^K_{\circ_\Pi}$ and $\tilde{\mathcal{S}}^K_{\circ_\Pi}$ the nodes scheduled from $\mathbf{s}_{\circ_\Pi}$ and $\tilde{\mathbf{s}}_{\circ_\Pi}$, respectively. We first observe that (10) holds for $n=T$. This follows from the bounded regularity of $R(\cdot)$, noting that $u(T)=1$, and distinguishing four possible cases.

- Case 1: $j \in \mathcal{S}^K_{\circ_\Pi}$ and $j \in \tilde{\mathcal{S}}^K_{\circ_\Pi}$, i.e., node j is scheduled in both cases.

$$\begin{aligned}&W_T(\mathbf{s}_{\circ_\Pi}) - W_T(\tilde{\mathbf{s}}_{\circ_\Pi}) \\ &=R(s_{\circ_{\Pi(1)}},\ldots,s_j,\ldots,s_{\circ_{\Pi(K)}}) - R(\tilde{s}_{\circ_{\Pi(1)}},\ldots,\tilde{s}_j,\ldots,\tilde{s}_{\circ_{\Pi(K)}}) \\ &=s_j R(s_{\circ_{\Pi(1)}},\ldots,1,\ldots,s_{\circ_{\Pi(K)}}) + (1-s_j)R(s_{\circ_{\Pi(1)}},\ldots,0, \\ &\quad \ldots,s_{\circ_{\Pi(K)}}) - \tilde{s}_j R(\tilde{s}_{\circ_{\Pi(1)}},\ldots,1,\ldots,\tilde{s}_{\circ_{\Pi(K)}}) \\ &\quad -(1-\tilde{s}_j)R(\tilde{s}_{\circ_{\Pi(1)}},\ldots,0,\ldots,\tilde{s}_{\circ_{\Pi(K)}}) \\ &=(s_j-\tilde{s}_j)\big(R(s_{\circ_{\Pi(1)}},\ldots,1,\ldots,s_{\circ_{\Pi(K)}}) \\ &\quad -R(s_{\circ_{\Pi(1)}},\ldots,0,\ldots,s_{\circ_{\Pi(K)}})\big) \\ &\leq (s_j-\tilde{s}_j)\Delta_u u(T),\end{aligned}$$

  where the second equality follows from the decomposability of $R(\cdot)$. Since $R(\cdot)$ is symmetric and the belief vectors are equal but for node j, we have $R(s_{\circ_{\Pi(1)}},\ldots,\tilde{s}_j=k,\ldots,s_{\circ_{\Pi(K)}}) = R(\tilde{s}_{\circ_{\Pi(1)}},\ldots,\tilde{s}_j=k,\ldots,\tilde{s}_{\circ_{\Pi(N)}})$, which we use in the third equality. Finally, the inequality follows from the boundedness of $R(\cdot)$.
- Case 2: $j \notin \mathcal{S}^K_{\circ_\Pi}$ and $j \notin \tilde{\mathcal{S}}^K_{\circ_\Pi}$, i.e., node j is not scheduled in either case. The same nodes with the same beliefs are scheduled in both cases, hence, $\mathbf{s}^K_{\circ_\Pi} = \tilde{\mathbf{s}}^K_{\circ_\Pi}$, and $W_T(\mathbf{s}_{\circ_\Pi}) - W_T(\tilde{\mathbf{s}}_{\circ_\Pi}) = 0$.
- Case 3: $j \in \mathcal{S}^K_{\circ_\Pi}$ and $j \notin \tilde{\mathcal{S}}^K_{\circ_\Pi}$. In this case there exists a node $m \in \tilde{\mathcal{S}}^K_{\circ_\Pi}$ such that $s_j \geq s_m \geq \tilde{s}_j$, and $m \notin \mathcal{S}^K_{\circ_\Pi}$

$$\begin{aligned}&W_T(\mathbf{s}_{\circ_\Pi}) - W_T(\tilde{\mathbf{s}}_{\circ_\Pi}) \\ &=(s_j-s_m)\big(R(s_{\circ_{\Pi(1)}},\ldots,1,\ldots,s_{\circ_{\Pi(K)}}) \\ &\quad -R(s_{\circ_{\Pi(1)}},\ldots,0,\ldots,s_{\circ_{\Pi(K)}})\big) \\ &\leq (s_j-\tilde{s}_j)\big(R(s_{\circ_{\Pi(1)}},\ldots,1,\ldots,s_{\circ_{\Pi(K)}}) \\ &\quad -R(s_{\circ_{\Pi(1)}},\ldots,0,\ldots,s_{\circ_{\Pi(K)}})\big) \\ &\leq (s_j-\tilde{s}_j)\Delta_u u(T),\end{aligned}$$

  where the first equality follows similar to Case 1, the second equality from the fact that $s_m \geq \tilde{s}_j$, and the last inequality from the boundedness of $R(\cdot)$. Note that node m is the node with the highest belief state that is not scheduled in $W_T(\mathbf{s}_{\circ_\Pi})$, and the node with the lowest belief state scheduled in $W_T(\tilde{\mathbf{s}}_{\circ_\Pi})$.
- Case 4: $j \notin \mathcal{S}^K_{\circ_\Pi}$ and $j \in \tilde{\mathcal{S}}^K_{\circ_\Pi}$. This case is not possible since the vectors $\mathbf{s}_{\circ_\Pi}$ and $\tilde{\mathbf{s}}_{\circ_\Pi}$ are ordered and $s_j \geq \tilde{s}_j$, hence, if $\tilde{s}_j$ is scheduled then s_j must be scheduled too.

Now, we assume that (10) holds from TS $n+1$ up to TS T, and show that it holds for TS n as well. We distinguish three cases:

- Case 1: $j \in \mathcal{S}^K_{\circ_\Pi}$ and $j \in \tilde{\mathcal{S}}^K_{\circ_\Pi}$ in (18), i.e., node j is scheduled in both cases. The first and second summations in the first line of (18a) correspond to the cases in which node $j \in \mathcal{S}^K_{\circ_\Pi}$ is idle and active, respectively, in TS n. Similarly, first and second summations in the second line of (18a) correspond to the cases in which node $j \in \tilde{\mathcal{S}}^K_{\circ_\Pi}$ is idle and active, respectively, in TS n.

$$
\begin{aligned}
&W_n(\mathbf{s}_{\Pi^\circ}) - W_n(\tilde{\mathbf{s}}_{\Pi^\circ}) \\
&= R(\mathbf{s}^K_{\Pi^\circ}) + (1-p)\beta \sum_{\mathcal{E}\subseteq \mathcal{S}^K_{\Pi^\circ}\setminus\{j\}} q(|\mathcal{E}|,K-1) W_{n+1}([T(\mathbf{s}_{\overline{\mathcal{E}}}), \mathbf{0}(|\mathcal{E}|)]) + p\beta \sum_{\mathcal{E}\subseteq \mathcal{S}^K_{\Pi^\circ}\setminus\{j\}} q(|\mathcal{E}|,K-1) W_{n+1}([T(\mathbf{s}_{\overline{\mathcal{E}\cup j}}), \mathbf{0}(|\mathcal{E}|+1)]) \\
&- R(\tilde{\mathbf{s}}^K_{\Pi^\circ}) - (1-p)\beta \sum_{\mathcal{E}\subseteq \tilde{\mathcal{S}}^K_{\Pi^\circ}\setminus\{j\}} q(|\mathcal{E}|,K-1) W_{n+1}([T(\tilde{\mathbf{s}}_{\overline{\mathcal{E}}}), \mathbf{0}(|\mathcal{E}|)]) - p\beta \sum_{\mathcal{E}\subseteq \tilde{\mathcal{S}}^K_{\Pi^\circ}\setminus\{j\}} q(|\mathcal{E}|,K-1) W_{n+1}([T(\tilde{\mathbf{s}}_{\overline{\mathcal{E}\cup j}}), \mathbf{0}(|\mathcal{E}|+1)]) && \text{(18a)} \\
&= R(\mathbf{s}^K_{\Pi^\circ}) - R(\tilde{\mathbf{s}}^K_{\Pi^\circ}) + (1-p)\beta \sum_{\mathcal{E}\subseteq \mathcal{S}^K_{\Pi^\circ}\setminus\{j\}} q(|\mathcal{E}|,K-1) \Big(W_{n+1}([\mathrm{T}(\mathbf{s}_{\overline{\mathcal{E}}}), \mathbf{0}(|\mathcal{E}|)]) - W_{n+1}([\mathrm{T}(\tilde{\mathbf{s}}_{\overline{\mathcal{E}}}), \mathbf{0}(|\mathcal{E}|)]) \Big) && \text{(18b)} \\
&\leq R(\mathbf{s}^K_{\Pi^\circ}) - R(\tilde{\mathbf{s}}^K_{\Pi^\circ}) + (1-p)\beta \sum_{\mathcal{E}\subseteq \mathcal{S}^K_{\Pi^\circ}\setminus\{j\}} q(|\mathcal{E}|,K-1) \Big(\Delta_u(\tau(s_j) - \tau(\tilde{s}_j)) u(n+1) \Big) && \text{(18c)} \\
&\leq \Delta_u(s_j - \tilde{s}_j) + (1-p)\beta \Delta_u(\tau(s_j) - \tau(\tilde{s}_j)) u(n+1) && \text{(18d)} \\
&\leq \Delta_u(s_j - \tilde{s}_j) + (1-p)\beta \Delta_u(s_j - \tilde{s}_j) u(n+1) && \text{(18e)} \\
&\leq \Delta_u(s_j - \tilde{s}_j) \Big(1 + \beta(1-p) \sum_{i=0}^{T-n-1} (\beta(1-p))^i \Big) && \text{(18f)} \\
&= \Delta_u(s_j - \tilde{s}_j) u(n), && \text{(18g)}
\end{aligned}
$$

Note that the belief state vector $\tilde{\mathbf{s}}_{\overline{\mathcal{E}\cup j}}$ includes the belief states of all the nodes in $\tilde{\mathbf{s}}_{\Pi^\circ}$, but those in $\mathcal{E}$ and $\tilde{s}_j$, hence, it is equivalent to the belief state vector $\mathbf{s}_{\overline{\mathcal{E}\cup j}}$. We use this fact to get (18b). Note that the belief state vectors in (18b) differ only in the belief states of node j, namely, $\tau(s_j)$ and $\tau(\tilde{s}_j)$ are the beliefs of node j in vectors $[\mathrm{T}(\mathbf{s}_{\overline{\mathcal{E}}}), \mathbf{0}(|\mathcal{E}|)]$ and $[\mathrm{T}(\tilde{\mathbf{s}}_{\overline{\mathcal{E}}}), \mathbf{0}(|\mathcal{E}|)]$, respectively; and hence, we use the induction hypothesis in the summation of (18b) to obtain (18c). The summation in (18c) is over all possible operative/inoperative combinations of the nodes in $\mathcal{S}^K_{\Pi^\circ}\setminus\{j\}$, and it is equal to one. This fact together with the boundedness and the decomposability of $R(\cdot)$ are used in (18c) to get (18d). The contracting property of $\tau(\cdot)$, and the definition of $u(n)$ are used in (18e) and (18f), respectively.

- Case 2:$j \notin \mathcal{S}^K_{\Pi^\circ}$ and $j \notin \tilde{\mathcal{S}}^K_{\Pi^\circ}$, i.e., the same nodes are scheduled from $\mathbf{s}_{\Pi^\circ}$ and $\tilde{\mathbf{s}}_{\Pi^\circ}$, and node j is not scheduled in either case. Then

$$
\begin{aligned}
&W_n(\mathbf{s}_{\Pi^\circ}) - W_n(\tilde{\mathbf{s}}_{\Pi^\circ}) \\
&\quad = \beta \sum_{\mathcal{E}\subseteq \mathcal{S}^K_{\Pi^\circ}} q(|\mathcal{E}|, K) \Big(W_{n+1}([\mathrm{T}(\mathbf{s}_{\overline{\mathcal{E}}}), \mathbf{0}(|\mathcal{E}|)]) \\
&\qquad - W_{n+1}([\mathrm{T}(\tilde{\mathbf{s}}_{\overline{\mathcal{E}}}) \mathbf{0}(|\mathcal{E}|)]) \Big) && \text{(19a)} \\
&\quad \leq \Delta_u(s_j - \tilde{s}_j) \beta u(n+1) && \text{(19b)} \\
&\quad \leq \Delta_u(s_j - \tilde{s}_j) \beta \sum_{i=0}^{T-n-1} (\beta(1-p))^i && \text{(19c)} \\
&\quad \leq \Delta_u(s_j - \tilde{s}_j) u(n), && \text{(19d)}
\end{aligned}
$$

where (19a) follows since the value of the expected immediate rewards in TS n are the same. The belief state vectors at TS $n+1$ are equal but for the belief state of node j, that is, $\tau(s_j)$ and $\tau(\tilde{s}_j)$ are the beliefs of node j in $\mathrm{T}(\mathbf{s}_{\overline{\mathcal{E}}})$ and $\mathrm{T}(\tilde{\mathbf{s}}_{\overline{\mathcal{E}}})$, respectively. In (19a), similarly to (18c), (18d), and (18e), we apply the induction hypothesis, the contracting map property, and the fact that the summation is equal to one, to obtain (19b). We use $\beta \leq 1$ and the definition of $u(n)$ to obtain (19c) and (19d), respectively.

- Case 3: $j \in \mathcal{S}^K_{\Pi^\circ}$ and $j \notin \tilde{\mathcal{S}}^K_{\Pi^\circ}$ in (20), i.e., there exists $m \in \tilde{\mathcal{S}}^K_{\Pi^\circ}$ such that $s_j \geq s_m = \tilde{s}_m \geq \tilde{s}_j$ and that $m \notin \mathcal{S}^K_{\Pi^\circ}$. Hence, $\mathcal{S}^K_{\Pi^\circ}$ and $\tilde{\mathcal{S}}^K_{\Pi^\circ}$ differ only in one element. To obtain (20a) we use the symmetry property of the pseudo value function and the fact that the belief vectors are equal but for node j; in (20b) we add and subtract a pseudo value function, which has two nodes with the same belief state s_m, and one is scheduled while the other is not. We can group the pseudo value functions, and apply the results of Case 1 and Case 2 above. In particular, for the pseudo value functions in the first line of (20b), the belief vectors are equal but for s_j and s_m, moreover $j \in \mathcal{S}^K_{\Pi^\circ}$ and $m \in \tilde{\mathcal{S}}^K_{\Pi^\circ}$, and $s_j \geq s_m$, so we can apply the results of Case 1. Similarly, for the two pseudo value functions in the second line of (20b) we can use the results of Case 2.

APPENDIX C

We note that set $\mathcal{S} = \{1, \ldots, K\}$ is the set of K nodes scheduled from $\mathbf{s}$, and that the set $\mathcal{S}^K_\Pi$ is the set of nodes scheduled from $\mathbf{s}_\Pi$, that is, the first K nodes as ordered according to permutation Π. We only need to study the cases in which $\mathcal{S}$ and $\mathcal{S}^K_\Pi$ are different, since the claim holds for the others due to the symmetric property of the pseudo value function, (8). We study the case $j \in \mathcal{S}$, $i \in \mathcal{S}^K_\Pi$, $i \notin \mathcal{S}$, and $j \notin \mathcal{S}^K_\Pi$ in (21). The summation in (21a) is over all operative/inoperative combinations of the nodes in $\mathcal{S}\setminus\{j\}$. We denote the belief state of all nodes but those in $\mathcal{E}$ and s_j by

$$
\begin{aligned}
&W_n(s_{\overset{\circ}{\Pi}(1)},\dots,s_j,\dots,s_{\overset{\circ}{\Pi}(K)},s_m,\dots,s_{\overset{\circ}{\Pi}(N)}) - W_n(\tilde{s}_{\overset{\circ}{\Pi}(1)},\dots,\tilde{s}_m,\tilde{s}_{\overset{\circ}{\Pi}(K+1)},\dots,\tilde{s}_j,\dots,\tilde{s}_{\overset{\circ}{\Pi}(N)}) \\
&\quad = W_n(s_{\overset{\circ}{\Pi}(1)},\dots,s_j,\dots,s_{\overset{\circ}{\Pi}(K)},\dots,s_m,\dots,s_{\overset{\circ}{\Pi}(N)}) - W_n(s_{\overset{\circ}{\Pi}(1)},\dots,s_m,\dots,s_{\overset{\circ}{\Pi}(K)},\dots,\tilde{s}_j,\dots,s_{\overset{\circ}{\Pi}(N)}) && \text{(20a)} \\
&\quad = W_n(s_{\overset{\circ}{\Pi}(1)},\dots,s_j,\dots,s_{\overset{\circ}{\Pi}(K)},\dots,s_m,\dots,s_{\overset{\circ}{\Pi}(N)}) - W_n(s_{\overset{\circ}{\Pi}(1)},\dots,s_m,\dots,s_{\overset{\circ}{\Pi}(K)},\dots,s_m,\dots,s_{\overset{\circ}{\Pi}(N)}) \\
&\quad + W_n(s_{\overset{\circ}{\Pi}(1)},\dots,s_m,\dots,s_{\overset{\circ}{\Pi}(K)},\dots,s_m,\dots,s_{\overset{\circ}{\Pi}(N)}) - W_n(s_{\overset{\circ}{\Pi}(1)},\dots,s_m,\dots,s_{\overset{\circ}{\Pi}(K)},\dots,\tilde{s}_j,\dots,s_{\overset{\circ}{\Pi}(N)}) && \text{(20b)} \\
&\quad \le \Delta_u(s_j - s_m)u(n) + \Delta_u(s_m - \tilde{s}_j)u(n) && \text{(20c)} \\
&\quad = \Delta_u(s_j - \tilde{s}_j)u(n). && \text{(20d)}
\end{aligned}
$$

$$
\begin{aligned}
&W_n(\mathbf{s}) - W_n(\mathbf{s}_\Pi) \\
&= R(\mathbf{s}^K) - R(\mathbf{s}_\Pi^K) + \beta \sum_{\mathcal{E}\subseteq\mathcal{S}\setminus\{j\}} q(|\mathcal{E}|,K-1)\Big(pW_{n+1}\big([\mathrm{T}(\mathbf{s}_{\overline{\mathcal{E}\cup j}}),\mathbf{0}(|\mathcal{E}|+1)]\big) + (1-p)W_{n+1}\big([\mathrm{T}(\mathbf{s}_{\overline{\mathcal{E}}}),\mathbf{0}(|\mathcal{E}|)]\big) \\
&\qquad\qquad - pW_{n+1}\big([\mathrm{T}(\mathbf{s}_{\overline{\mathcal{E}\cup i}}),\mathbf{0}(|\mathcal{E}|+1)]\big) - (1-p)W_{n+1}\big([\mathrm{T}(\mathbf{s}_{\overline{\mathcal{E}}}),\mathbf{0}(|\mathcal{E}|)]\big)\Big) && \text{(21a)} \\
&= R(\mathbf{s}^K) - R(\mathbf{s}_\Pi^K) - p\beta \sum_{\mathcal{E}\subseteq\mathcal{S}\setminus\{j\}} q(|\mathcal{E}|,K-1)\Big(W_{n+1}\big([\mathrm{T}(\mathbf{s}_{\overline{\mathcal{E}\cup i}}),\mathbf{0}(|\mathcal{E}|+1)]\big) - W_{n+1}\big([\mathrm{T}(\mathbf{s}_{\overline{\mathcal{E}\cup j}}),\mathbf{0}(|\mathcal{E}|+1)]\big)\Big) && \text{(21b)} \\
&\ge \Delta_l(s_j - s_i) - p\beta \sum_{\mathcal{E}\subseteq\mathcal{S}\setminus\{j\}} q(|\mathcal{E}|,K-1)\Big(W_{n+1}\big([\mathrm{T}(\mathbf{s}_{\overline{\mathcal{E}\cup i}}),\mathbf{0}(|\mathcal{E}|+1)]\big) - W_{n+1}\big([\mathrm{T}(\mathbf{s}_{\overline{\mathcal{E}\cup j}}),\mathbf{0}(|\mathcal{E}|+1)]\big)\Big) && \text{(21c)} \\
&\ge \Delta_l(s_j - s_i) - p\beta \sum_{\mathcal{E}\subseteq\mathcal{S}\setminus\{j\}} \Big(q(|\mathcal{E}|,K-1)\Delta_u(\tau(s_j) - \tau(s_i))u(n+1)\Big) && \text{(21d)} \\
&\ge \Delta_l(s_j - s_i) - p\beta\Delta_u(s_j - s_i)u(n+1) && \text{(21e)} \\
&\ge \Delta_l(s_j - s_i) - p\beta\Delta_u(s_j - s_i)u(0) && \text{(21f)} \\
&= (s_j - s_i)\left(\Delta_l - p\beta\frac{1-\beta(1-p)^{T+1}}{1-\beta(1-p)}\Delta_u\right) \ge 0 && \text{(21g)}
\end{aligned}
$$

$\mathbf{s}_{\overline{\mathcal{E}\cup j}}$. The belief state of node i in TS $n+1$, $\tau(s_i)$, is in vector $\mathrm{T}(\mathbf{s}_{\overline{\mathcal{E}\cup j}})$. Similarly, the belief state of node j in TS $n+1$, $\tau(s_j)$, is in vector $\mathrm{T}(\mathbf{s}_{\overline{\mathcal{E}\cup i}})$. The second pseudo value functions in the first and second lines in (21a) cancel out, and (21b) is obtained. We have applied the decomposability and boundedness of $R(\cdot)$ to obtain (21c). Belief vectors $\mathrm{T}(\mathbf{s}_{\overline{\mathcal{E}\cup j}})$ and $\mathrm{T}(\mathbf{s}_{\overline{\mathcal{E}\cup i}})$ in (21c) are ordered and only differ in one element, $\tau(s_i)$ and $\tau(s_j)$, respectively, where $\tau(s_i) \le \tau(s_j)$, and hence, we use Lemma 2 to get (21d); (21e) follows since $\tau(\cdot)$ is a monotonically increasing contracting map, (21f) since $u(n)$ is decreasing in n; finally (21g) follows since $u(0)$ is the sum of a geometric series.

APPENDIX D

We again use backward induction. Lemma 4 holds trivially for $n = T$. Note that in (15a) the set of nodes scheduled in the pseudo value functions $W_n(\mathbf{s}_{\widehat{\Pi}})$ and $W_n(\mathbf{s})$ are $\{1,\dots,K-1,N\}$ and $\{1,\dots,K\}$, respectively. That is, node K is scheduled in $W_n(\mathbf{s})$, but not in $W_n(\mathbf{s}_{\widehat{\Pi}})$; and node N is scheduled in $W_n(\mathbf{s}_{\widehat{\Pi}})$, but not in $W_n(\mathbf{s})$. To prove that (15a) holds at TS n we use a sample path argument similarly to [18], and assume that the realizations of the EH processes of nodes K and N are either 0 or 1. There are four different cases, but here we only consider one, since the others follow similarly.

We consider the case in which the EH processes have realizations $E_K^s(n) = 1$ and $E_N^s(n) = 0$. We denote by $\mathcal{K} = \{1,\dots,K-1\}$ the set of nodes scheduled in both sides of (15a). If $\mathcal{E}$ is the set of active nodes, we denote the set of nodes in $\mathcal{K}$ that remain idle by $\mathcal{K}^i = \mathcal{K}\backslash\mathcal{E}$. We denote the nodes that are not scheduled in either side of (15a) by $\mathcal{K}^s = \overline{\mathcal{K}\cup\{K,N\}}$. We denote the set $\{0,1\}^{|\mathcal{E}|}$ by $\mathcal{B}^{|\mathcal{E}|}$. From the left hand side of (15a) we obtain (22), where in (22c) we have applied the induction hypothesis of (15a), the symmetry of the pseudo value function, the inequality $p_{11} \ge p_{00}$, and the definition of $R(\cdot)$. This concludes the proof of (15a).

Now we prove the second part of Lemma 4, (15b). There are three cases:

- Case 1: $j,i \le K$, i.e., nodes j and i are scheduled on both sides of (15b). The inequality holds since the pseudo value function is symmetric.
- Case 2: $j \le K$ and $i > K$ in (23), i.e., nodes i and j are scheduled on the left and right hand sides of (15b), respectively. To prove the inequality we use the linearity of the pseudo value function (14). Since $s_j \ge s_i$, using (14), we only need to prove that $W_n(s_1,\dots,1,\dots,0,\dots,s_N) - W_n(s_1,\dots,0,\dots,1,\dots,s_N) \ge 0$. We denote the scheduled nodes in both sides of (15b) by $\mathcal{K} = \{1,\dots,K\}\backslash\{j\}$, the set of nodes in $\mathcal{K}$ that remain idle

$$
\begin{aligned}
&1+W_n(s_N,s_1,\ldots,s_{N-1})\\
&=1+R(\mathbf{s}^K_{\widehat{\Pi}})+\beta\sum_{\mathcal{E}\subseteq\mathcal{K}}\sum_{l_\mathcal{E}\in\mathcal{B}^{|\mathcal{E}|}}h(l_\mathcal{E},K-1)\Big[pW_{n+1}\left(\mathbf{P}_{11}\left(\Sigma l_\mathcal{E}\right),\boldsymbol{\tau}(\mathbf{s}_{\mathcal{K}^i}),s_K=p_{11},\boldsymbol{\tau}(\mathbf{s}_{\mathcal{K}^s}),s_N=p_{01},\mathbf{P}_{01}\left(\overline{\Sigma}l_\mathcal{E}\right)\right)\\
&\qquad+(1-p)W_{n+1}\left(\mathbf{P}_{11}\left(\Sigma l_\mathcal{E}\right),s_N=p_{01},\boldsymbol{\tau}(\mathbf{s}_{\mathcal{K}^i}),s_K=p_{11},\boldsymbol{\tau}(\mathbf{s}_{\mathcal{K}^s}),\mathbf{P}_{01}\left(\overline{\Sigma}l_\mathcal{E}\right)\right)\Big] &\text{(22a)}\\
&\geq p+R(\mathbf{s}^K_{\widehat{\Pi}})+\beta\sum_{\mathcal{E}\subseteq\mathcal{K}}\sum_{l_\mathcal{E}\in\mathcal{B}^{|\mathcal{E}|}}h(l_\mathcal{E},K-1)\Big[pW_{n+1}\left(\mathbf{P}_{11}\left(\Sigma l_\mathcal{E}\right),\boldsymbol{\tau}(\mathbf{s}_{\mathcal{K}^i}),s_K=p_{11},\boldsymbol{\tau}(\mathbf{s}_{\mathcal{K}^s}),s_N=p_{01},\mathbf{P}_{01}\left(\overline{\Sigma}l_\mathcal{E}\right)\right)\\
&\qquad+(1-p)\Big(1+W_{n+1}\left(\mathbf{P}_{11}\left(\Sigma l_\mathcal{E}\right),s_N=p_{01},\boldsymbol{\tau}(\mathbf{s}_{\mathcal{K}^i}),s_K=p_{11},\boldsymbol{\tau}(\mathbf{s}_{\mathcal{K}^s}),\mathbf{P}_{01}\left(\overline{\Sigma}l_\mathcal{E}\right)\right)\Big)\Big] &\text{(22b)}\\
&\geq p+R(\mathbf{s}^K_{\widehat{\Pi}})+\beta\sum_{\mathcal{E}\subseteq\mathcal{K}}\sum_{l_\mathcal{E}\in\mathcal{B}^{|\mathcal{E}|}}h(l_\mathcal{E},K-1)\Big[pW_{n+1}\left(\mathbf{P}_{11}\left(\Sigma l_\mathcal{E}\right),\boldsymbol{\tau}(\mathbf{s}_{\mathcal{K}^i}),s_K=p_{11},\boldsymbol{\tau}(\mathbf{s}_{\mathcal{K}^s}),s_N=p_{01},\mathbf{P}_{01}\left(\overline{\Sigma}l_\mathcal{E}\right)\right)\\
&\qquad+(1-p)W_{n+1}\left(\mathbf{P}_{11}\left(\Sigma l_\mathcal{E}\right),\boldsymbol{\tau}(\mathbf{s}_{\mathcal{K}^i}),s_K=p_{11},\boldsymbol{\tau}(\mathbf{s}_{\mathcal{K}^s}),\mathbf{P}_{01}\left(\overline{\Sigma}l_\mathcal{E}\right),s_N=p_{01}\right)\Big] &\text{(22c)}\\
&=R(\mathbf{s}^K)+\beta\sum_{\mathcal{E}\subseteq\mathcal{K}}\sum_{l_\mathcal{E}\in\mathcal{B}^{|\mathcal{E}|}}h(l_\mathcal{E},K-1)\Big[pW_{n+1}\left(\mathbf{P}_{11}\left(\Sigma l_\mathcal{E}\right),s_K=p_{11},\boldsymbol{\tau}(\mathbf{s}_{\mathcal{K}^i}),\boldsymbol{\tau}(\mathbf{s}_{\mathcal{K}^s}),s_N=p_{01},\mathbf{P}_{01}\left(\overline{\Sigma}l_\mathcal{E}\right)\right)\\
&\qquad+(1-p)W_{n+1}\left(\mathbf{P}_{11}\left(\Sigma l_\mathcal{E}\right),\boldsymbol{\tau}(\mathbf{s}_{\mathcal{K}^i}),s_K=p_{11},\boldsymbol{\tau}(\mathbf{s}_{\mathcal{K}^s}),s_N=p_{01},\mathbf{P}_{01}\left(\overline{\Sigma}l_\mathcal{E}\right)\right)\Big] &\text{(22d)}\\
&=W_n(\mathbf{s}) &\text{(22e)}
\end{aligned}
$$

$$
\begin{aligned}
W_n(\tilde{\mathbf{s}})&=R(\tilde{\mathbf{s}}^K)+\beta\sum_{\mathcal{E}\subseteq\mathcal{K}}\sum_{l_\mathcal{E}\in\mathcal{B}^{|\mathcal{E}|}}h(l_\mathcal{E},K-1)\Big[pW_{n+1}\left(\mathbf{P}_{11}\left(\Sigma l_\mathcal{E}\right),s_j=p_{11},\boldsymbol{\tau}(\mathbf{s}_{\mathcal{K}^i}),\boldsymbol{\tau}(\mathbf{s}_{\mathcal{K}^s\cup i}),\mathbf{P}_{01}\left(\overline{\Sigma}l_\mathcal{E}\right)\right)\\
&\qquad+(1-p)W_{n+1}\left(\mathbf{P}_{11}\left(\Sigma l_\mathcal{E}\right),\boldsymbol{\tau}(\mathbf{s}_{\mathcal{K}^i\cup j}),\boldsymbol{\tau}(\mathbf{s}_{\mathcal{K}^s\cup i}),\mathbf{P}_{01}\left(\overline{\Sigma}l_\mathcal{E}\right)\right)\Big] &\text{(23a)}\\
&\geq R(\tilde{\mathbf{s}}^K)-p+\beta\sum_{\mathcal{E}\subseteq\mathcal{K}}\sum_{l_\mathcal{E}\in\mathcal{B}^{|\mathcal{E}|}}h(l_\mathcal{E},K-1)\Big[p\Big(1+W_{n+1}\left(\mathbf{P}_{11}\left(\Sigma l_\mathcal{E}\right),s_i=p_{01},\boldsymbol{\tau}(\mathbf{s}_{\mathcal{K}^i}),\boldsymbol{\tau}(\mathbf{s}_{\mathcal{K}^s\cup j}),\mathbf{P}_{01}\left(\overline{\Sigma}l_\mathcal{E}\right)\right)\Big)\\
&\qquad+(1-p)W_{n+1}\left(\mathbf{P}_{11}\left(\Sigma l_\mathcal{E}\right),\boldsymbol{\tau}(\mathbf{s}_{\mathcal{K}^i\cup i}),\boldsymbol{\tau}(\mathbf{s}_{\mathcal{K}^s\cup j}),\mathbf{P}_{01}\left(\overline{\Sigma}l_\mathcal{E}\right)\right)\Big] &\text{(23b)}\\
&\geq R(\tilde{\mathbf{s}}^K)-p+\beta\sum_{\mathcal{E}\subseteq\mathcal{K}}\sum_{l_\mathcal{E}\in\mathcal{B}^{|\mathcal{E}|}}h(l_\mathcal{E},K-1)\Big[pW_{n+1}\left(\mathbf{P}_{11}\left(\Sigma l_\mathcal{E}\right),\boldsymbol{\tau}(\mathbf{s}_{\mathcal{K}^i}),\boldsymbol{\tau}(\mathbf{s}_{\mathcal{K}^s\cup j}),\mathbf{P}_{01}\left(\overline{\Sigma}l_\mathcal{E}\right),s_i=p_{01}\right)\\
&\qquad+(1-p)W_{n+1}\left(\mathbf{P}_{11}\left(\Sigma l_\mathcal{E}\right),\boldsymbol{\tau}(\mathbf{s}_{\mathcal{K}^i\cup i}),\boldsymbol{\tau}(\mathbf{s}_{\mathcal{K}^s\cup j}),\mathbf{P}_{01}\left(\overline{\Sigma}l_\mathcal{E}\right)\right)\Big] &\text{(23c)}\\
&=W_n(\tilde{\mathbf{s}}_\Pi) &\text{(23d)}
\end{aligned}
$$

by $\mathcal{K}^i=\mathcal{K}\backslash\mathcal{E}$, and the nodes that are not scheduled in either side of (15b) by $\mathcal{K}^s=\overline{\mathcal{K}\cup\{j,i\}}$. We denote the belief vector $(s_1,\ldots,s_j=1,\ldots,s_i=0,\ldots,s_N)$ by $\tilde{\mathbf{s}}$, its i,j-swap by $\tilde{\mathbf{s}}_\Pi$, and define $\tilde{\mathbf{s}}^K\triangleq(\tilde{s}_1,\ldots,\tilde{s}_K)$. In (23) have used the induction hypothesis of (15b) and (15a) in (23b) and (23c), respectively, and the fact that $\beta\leq 1$.

- Case 3: nodes s_j and s_i are not scheduled. Inequality holds in this case, by applying the definition of (13) and the induction hypothesis of (15b).

REFERENCES

[1] M. Gorlatova, P. Kinget, I. Kymissis, D. Rubenstein, X. Wang, and G. Zussman, "Energy harvesting active networked tags (EnHANTs) for ubiquitous object networking," *Wireless Communications, IEEE*, vol. 17, no. 6, pp. 18–25, December 2010.

[2] D. Gunduz, K. Stamatiou, N. Michelusi, and M. Zorzi, "Designing intelligent energy harvesting communication systems," *IEEE Commun. Mag.*, vol. 52, no. 1, pp. 210–216, Jan. 2014.

[3] J. Yang and S. Ulukus, "Optimal packet scheduling in an energy harvesting communication system," *IEEE Trans. Commun.*, vol. 60, no. 1, pp. 220–230, Jan. 2012.

[4] B. Devillers and D. Gündüz, "A general framework for the optimization of energy harvesting communication systems," *J. of Commun. and Nerworks., Special Issue on Energy Harvesting in Wireless Networks*, vol. 14, no. 2, pp. 130–139, Apr. 2012.

[5] Z. Wang, A. Tajer, and X. Wang, "Communication of energy harvesting tags," *IEEE Trans. Commun.*, vol. 60, no. 4, pp. 1159–1166, Apr. 2012.

[6] A. Aprem, C. Murthy, and N. Mehta, "Transmit power control policies for energy harvesting sensors with retransmissions," *IEEE J. of Selected Topics in Signal Processing*, vol. 7, no. 5, pp. 895–906, Oct. 2013.

[7] 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. 4, pp. 547–551, Apr. 2009.

[8] N. Michelusi and M. Zorzi, "Optimal random multiaccess in energy harvesting wireless sensor networks," in *IEEE International Conference in Communications (ICC)-E2Nets workshop*, Budapest, Hungary, Jun. 2013.

[9] F. Iannello, O. Simeone, and U. Spagnolini, "Optimality of myopic scheduling and Whittle indexability for energy harvesting sensors," in *Conf. on Information Sciences and Systems (CISS)*, Princeton, NJ, Mar. 2012, pp. 1–6.

[10] P. Blasco, D. Gündüz, and M. Dohler, "Low-complexity scheduling policies for energy harvesting communication networks," in *IEEE International Symposium on Information Theory (ISIT)*, Istanbul, Turkey, Jul. 2013.
[11] O. M. Gul and E. Uysal-Biyikoglu, "A randomized scheduling algorithm for energy harvesting wireless sensor networks achieving nearly 100% throughput," in *IEEE Wireless Communications and Networking Conference (WCNC)*, Istanbul, Turkey, Apr. 2014.
[12] P. Blasco, D. Gündüz, and M. Dohler, "A learning theoretic approach to energy harvesting communication system optimization," *IEEE Trans. Wireless Commun.*, vol. 12, no. 4, pp. 1872–1882, Apr. 2013.
[13] D. P. Bertsekas, *Dynamic Programming and Optimal Control*. Athenea Scientific, 2005.
[14] M. L. Littman, T. L. Dean, and L. P. Kaelbling, "On the complexity of solving Markov decision problems," in *Conference on Uncertainty in Artificial Intelligence (UAI)*, Montreal, QU, Aug. 1995, pp. 394–402.
[15] P. Whittle, "Restless bandits: Activity allocation in a changing world," *Journal of Applied Probability, a celabration of Applied Probability*, vol. 25, pp. 287–298, 1988.
[16] C. H. Papadimitriou and J. Tsitsiklis, "The complexity of optimal queueing network control," in *Structure in Complexity Theory Conference*, Amsterdam, The Netherlands, Jun. 1994, pp. 318–322.
[17] S. H. A. Ahmad, M. Liu, T. Javidi, Q. Zhao, and B. Krishnamachari, "Optimality of myopic sensing in multichannel opportunistic access," *IEEE Trans. Inf. Theory*, vol. 55, no. 9, pp. 4040–4050, Sep. 2009.
[18] S. H. A. Ahmad and M. Liu, "Multi-channel opportunistic access: A case of restless bandits with multiple plays," in *Allerton conference on Communication, Control, and Computing*, Monticello, IL, Sep. 2009, pp. 1361–1368.
[19] P. Mansourifard, T. Javidi, and B. Krishnamachari, "Optimality of myopic policy for a class of monotone affine restless multi-armed bandits," in *IEEE Conference on Decision and Control (CDC)*, Grand Wailea Maui, HI, USA, Dec. 2012, pp. 877–882.
[20] K. Wang and L. Chen, "On optimality of myopic policy for restless multi-armed bandit problem: An axiomatic approach," *IEEE Trans. Signal Process.*, vol. 60, no. 1, pp. 300–309, Jan 2012.
[21] K. Liu, Q. Zhao, and B. Krishnamachari, "Dynamic multichannel access with imperfect channel state detection," *IEEE Trans. Signal Process.*, vol. 58, no. 5, pp. 2795–2808, May 2010.
[22] I. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "A survey on sensor networks," *IEEE Commun. Mag.*, vol. 40, no. 8, pp. 102–114, Aug. 2002.
[23] D. Bertsimas and J. Niño-Mora, "Restless bandits, linear programming relaxations, and primal-dual index heuristic," *Operational Research*, vol. 48, no. 1, pp. 80–90, Jan.-Feb. 2000.
[24] Q. Zhao, Krishnamachari, and K. Liu, "On myopic sensing for multi-channel opportunistic access: Structure, optimality and performance," *IEEE Trans. Wireless Commun.*, vol. 7, no. 12, pp. 5431–5440, Dec. 2008.

Pol Blasco received the B.Eng. degree from Technische Universität Darmstadt, Germany, and BarcelonaTech (formally UPC), Spain, in 2008 and 2009, respectively, the M. S. degree from BarcelonaTech, in 2011, and the Ph.D. degree in electrical engineering from Imperial College London, in 2014. He was research assistant at CTTC in Barcelona, Spain, from November 2009 until August 2013. He was a visiting scholar in the Centre for Wireless Communication in University of Oulu, Finland, during the last semester of 2011, and to Imperial College during the first half of 2013. In 2008 he pursued the B.Eng. thesis in European Space Operation Center in the OPS-GSS section, Darmstadt, Germany. He also has carried on research in neuroscience in IDIBAPS, Barcelona, Spain, and in the Technische Universität Darmstadt in collaboration with the Max-Planck-Institut, Frankfurt, Germany, in 2009 and 2008, respectively. His current research interest cover communication of energy harvesting devices, cognitive radio, machine learning, control theory, decision making, and neuroscience.

Deniz Gündüz received the B.S. degree in electrical and electronics engineering from METU, Turkey in 2002, and the M.S. and Ph.D. degrees in electrical engineering from NYU Polytechnic School of Engineering in 2004 and 2007, respectively. After his PhD he served as a postdoctoral research associate at the Department of Electrical Engineering, Princeton University, and as a consulting assistant professor at the Department of Electrical Engineering, Stanford University. Since September 2012 he is a Lecturer in the Electrical and Electronic Engineering Department of Imperial College London, UK. Previously he was a research associate at CTTC in Barcelona, Spain. He also held a visiting researcher position at Princeton University from November 2009 until November 2011.

Dr. Gunduz is an Associate Editor of the IEEE TRANSACTIONS ON COMMUNICATIONS, and served as a guest editor of the EURASIP Journal on Wireless Communications and Networking, Special Issue on Recent Advances in Optimization Techniques in Wireless Communication Networks. He is the recipient of a Marie Curie Fellowship awarded by the European Commission, and a recipient of the Best Student Paper Award at the 2007 IEEE International Symposium on Information Theory (ISIT).

He is serving as a co-chair of the IEEE Information Theory Society Student Committee. He is also the co-director of the Imperial College Probability Center. He has served as the co-chair of the Network Theory Symposium at the 2013 and 2014 IEEE Global Conference on Signal and Information Processing (GlobalSIP), and he was a co-chair of the 2012 IEEE European School of Information Theory (ESIT). His research interests lie in the areas of communication theory and information theory with special emphasis on joint source-channel coding, multi-user networks, energy efficient communications and privacy.