On Optimal Transmission Policies for Energy Harvesting Devices

Nicolò Michelusi*, Kostas Stamatiou† and Michele Zorzi*†

{nicolo.michelusi,kstamat,zorzi}@dei.unipd.it

*Department of Information Engineering, University of Padova - via Gradenigo, 6b, 35131 Padova, Italy

†Consorzio Ferrara Ricerche (CFR) - via Saragat, 1, 44122 Ferrara, Italy

Abstract—We consider an energy harvesting device whose state at a given time is determined by its energy level and an "importance" value, associated to the transmission of a data packet to the receiver at that particular time. We consider policies that, at each time, elect whether to transmit the data packet or not, based on the current energy level and data importance, so as to maximize the long-term average transmitted data importance. Under the assumption of i.i.d. Bernoulli energy arrivals, we show that the sensor should report only data with an importance value above a given threshold, which is a strictly decreasing function of the energy level, and we derive upper and lower bounds on the thresholds for any energy level. Leveraging on these findings, we construct a suboptimal policy that performs very close to the optimal one, at a fraction of the complexity. Finally, we demonstrate that a threshold policy, which on the average transmits with probability equal to the average energy arrival rate is asymptotically optimal as the energy storage capacity grows large.

I. INTRODUCTION

Energy harvesting devices (EHD) collect, or "harvest", energy from the environment in order to perform sensing and data-communication tasks, thus having the ability to operate autonomously for extended periods of time [1]. In contrast to traditional sensors, where the objective is to minimize energy consumption under a performance constraint, *e.g.*, delay [2], in the case of an EHD, the objective is the "management" of the harvested energy. Intuitively, an EHD should judiciously perform its assigned task based on its available energy, becoming more "conservative" as its energy supply runs low to ensure uninterrupted operation, and more "aggressive" when energy is abundant.

In this paper, we study the case of an EHD which reports incoming data of different "importance" levels to a receiver (RX), with the overall goal to maximize the long-term average importance of the reported data. A number of practical examples fall under this general model: a temperature-sensing EHD, where the importance is an increasing function of the temperature value, higher temperatures being the indicator of overheating or fire; an EHD which relays different priority packets in a sensor network [3]; an EHD which adjusts the packet information rate based on the channel condition to the RX, in which case the importance level corresponds to the instantaneous rate and the objective is to maximize the long-term average throughput [4].

We consider a slotted-time system with i.i.d. Bernoulli energy arrivals, and i.i.d. importance values which follow an arbitrary continuous distribution. We show that, for the class of binary (transmit/no transmit) policies, the optimal policy dictates the transmission of data with importance above a given threshold, which is decreasing in the energy level. In other words, the EHD becomes more frugal as its energy level decreases, reserving its energy only for the transmission of important data. This is an intuitive result, yet its proof is quite involved, requiring a crafty manipulation of the energy level steady-state distribution. Using the structure of the optimal policy, we derive upper and lower bounds on the transmission thresholds for any energy level and construct a suboptimal policy which is shown to perform close to the optimal one. Moreover, we show that a "balanced" policy, *i.e.*, a threshold policy which on the average (over the importance value distribution) transmits with probability equal to the average energy harvesting rate becomes asymptotically optimal, as the energy storage capacity goes to infinity.

The problem of maximizing the average value of the reported data from an energy-aware replenishable sensor was formulated in [5]. However, [5] considered a continuous-time system and employed policy iteration to determine the optimal thresholds. [3] investigated the relaying of packets of different priorities in a network of energy-limited sensors, but did not account for energy harvesting capability. In [4], an EHD with a data queue was considered and sufficient stability conditions, as well as heuristic delay-minimizing policies, were derived. Other related work on EHDs includes [6], [7], which consider variants of the system model employed in this work, but are concerned with a different performance metric, namely the probability of detection of a randomly occurring event, and [8], which proposes the use of RF-energy harvesting to enhance the performance of passive RFID systems.

This paper is organized as follows. In Section II, the system model is described in detail. Our theoretical results are derived in Section III. In Section IV, we present numerical results related to the performance of the optimal policy and various suboptimal policies. Finally, Section V concludes the paper.

II. SYSTEM MODEL

A. General

We consider a slotted-time system, where slot k is the time interval $[k, k+1)$, $k \in \mathbb{Z}^+$. At each time instant k, the EHD has a new data packet to send to the RX, and the packet duration is one slot. If the packet is not sent, then it is discarded.

The EHD energy storage capability is modeled by a buffer. As in previous work [6], [7], we assume for simplicity that each position in the buffer can hold one energy quantum and that the transmission of one data packet requires the

expenditure of one energy quantum. The maximum number of quanta that can be stored is $e_{\max}$ and the set of possible energy levels is denoted by $\mathcal{E} = \{0, 1, \ldots, e_{\max}\}$.

Denote the amount of energy quanta at time k as E_k. The evolution of E_k is determined by the following equation

$$E_{k+1} = \min\left\{[E_k - Q_k]^+ + B_k, e_{\max}\right\}, \quad (1)$$

where:

- $\{B_k\}$ is the energy arrival process, which models the randomness in the energy harvesting mechanism, *e.g.*, due to an erratic energy supply. We assume that $\{B_k\}$ are i.i.d. Bernoulli random variables with mean $\bar{b} \in (0, 1)$.[1] Moreover, we assume that a new energy quantum harvested in slot k can only be used at a later time instant $> k$.
- $\{Q_k\}$ is the action process, which is one if the current data packet is transmitted and one energy quantum is drawn from the buffer, and zero otherwise.

We now formally define energy *outage* and *overflow*.

Definition 1 *In slot k, energy outage occurs if $E_k = 0$ and energy overflow occurs if $(E_k = e_{\max}) \cap (B_k = 1) \cap (Q_k = 0)$.*

Under energy outage, no transmissions can be performed, hence $Q_k = 0$ regardless of the importance of the current data packet. When energy overflow occurs, an incoming energy quantum cannot be stored due to the finite storage capacity. Since energy is lost, an energy overflow potentially represents a lost transmission opportunity in the future.

At time k, the EHD state $\boldsymbol{S}_k$ is defined as $\boldsymbol{S}_k = (E_k, V_k)$, where V_k is the importance value of the current data packet. We assume that $V_k \in \mathbb{R}^+$ is a continuous random variable with probability density function (pdf) $f_V(v)$, $v \geq 0$, and that $\{V_k\}$ are i.i.d.

B. Policy definition and general optimization problem

Given $\boldsymbol{S}_k$, a *policy* determines $Q_k \in \{0, 1\}$ at time k. Formally, a policy μ is a probability measure on the action space $\{0, 1\}$, parameterized by the state $\boldsymbol{S}_k$, *i.e.*, given that $\boldsymbol{S}_k = (e, v) \in \mathcal{E} \times \mathbb{R}^+$, $\mu(1; e, v)$ and $\mu(0; e, v) = 1 - \mu(1; e, v)$ are the probabilities of drawing one energy quantum (*i.e.*, transmitting the data packet) and not drawing an energy quantum (*i.e.*, discarding the data packet), respectively.[2]

Given an initial state $\boldsymbol{S}_0$, the *long-term average reward* under policy μ is defined as

$$G(\mu, \boldsymbol{S}_0) = \lim_{K \to \infty} \inf \frac{1}{K} \mathbb{E}\left[\sum_{k=0}^{K-1} Q_k V_k \,\middle|\, \boldsymbol{S}_0\right], \quad (2)$$

where the expectation is taken with respect to $\{B_k, Q_k, V_k\}$ and Q_k is drawn according to μ. The optimization problem at hand is to determine the optimal μ^* such that

$$\mu^* = \arg\max_\mu G(\mu, \boldsymbol{S}_0). \quad (3)$$

[1] $\bar{b} = 0$ corresponds to the case of no energy harvesting and $\bar{b} = 1$ to the case where energy is always available, hence no energy management issues arise.

[2] For the sake of maximizing a long-term average reward function of the state and action processes, it is sufficient to consider only state-dependent stationary policies [9].

It can be proved that μ^* must have a threshold structure [10], *i.e.*, for each energy level $e \in \mathcal{E}$, there exists a threshold $v_{\mathrm{th}}(e)$ such that

$$\begin{cases} \mu(1; e, v) = 1 & v \geq v_{\mathrm{th}}(e) \\ \mu(1; e, v) = 0 & v < v_{\mathrm{th}}(e). \end{cases} \quad (4)$$

As a result, we henceforth consider only the subset of policies with threshold structure.

For ease of notation, we introduce the function $g(x), x \in [0, 1]$, defined as

$$g(x) = E_V\left[\chi\left(V \geq \bar{F}_V^{-1}(x)\right) V\right] = \int_{\bar{F}_V^{-1}(x)}^{+\infty} v\lambda(v)\mathrm{d}v, \quad (5)$$

where $\chi(\cdot)$ is the indicator function and $\bar{F}_V^{-1}(x)$ is the inverse of the complementary cumulative distribution function of the importance value process $\bar{F}_V(v)$, *i.e.*,

$$x = \bar{F}_V(v) = \int_v^{+\infty} f_V(\nu)\mathrm{d}\nu, \quad x \in [0, 1]. \quad (6)$$

The function $g(x)$ is the average reward accrued by transmitting only the data packets with importance value above the threshold $v = \bar{F}_V^{-1}(x)$, which corresponds to an average transmission probability $x = \bar{F}_V(v)$.

With these definitions in place, let $\eta(e) = \mathbb{E}_V[\mu(1; e, V)]$ denote the average transmission probability when the energy level is e. From (6), we have that $v_{\mathrm{th}}(e) = \bar{F}_V^{-1}((\eta(e))$. Moreover, from (5), the average reward accrued, when the energy level is e, is $g(\eta(e))$.

Due to the threshold structure, the mapping between μ, $v_{\mathrm{th}}(\cdot)$ and $\eta(\cdot)$ is one-to-one, and the transition probabilities of the time-homogeneous Markov chain $\{E_k\}$ are governed by η. Henceforth, we refer to policy μ in terms of the corresponding average transmission probability η.

We close this section by stating without proof some properties of the function $g(x)$, which are used in Section III.

Lemma 1 *The function $g(x)$ defined in (5) is strictly increasing, strictly concave in $(0, 1)$ and $g'(x) = \bar{F}_V^{-1}(x)$.*

III. CHARACTERIZATION OF THE OPTIMAL POLICY

A. Definitions and preliminary results

Before addressing (3), we provide some definitions and preliminary results.

Definition 2 *A policy η is said to be admissible if the Markov chain $\{E_k\}$ is irreducible.*

Under an admissible policy η, there exists a unique steady-state distribution of the energy level states, denoted by $\pi_\eta(e)$, $e \in \mathcal{E}$. Moreover, the long-term reward does not depend on the initial state. With a slight abuse of notation, (2) becomes

$$G(\eta) = \lim_{K \to \infty} \frac{1}{K} \mathbb{E}\left[\sum_{k=0}^{K-1} \chi\left(V_k \geq \bar{F}_V^{-1}(\eta(E_k))\right) V_k \,\middle|\, \boldsymbol{S}_0\right]$$

$$= \sum_{e=0}^{e_{\max}} \pi_\eta(e) g(\eta(e)). \quad (7)$$

In the following lemma, we determine the set of admissible policies.

Lemma 2 *The set of admissible policies is*

$$\mathcal{U} = \{\eta : \eta(0) = 0, \ \eta(e_{\max}) \in (0,1],$$
$$\eta(e) \in (0,1), e = 1, \dots, e_{\max} - 1\}.$$

Lemma 2 is proved in [10] by showing that E_k is in a unique communicating class, if and only if $\eta \in \mathcal{U}$. If $\eta \notin \mathcal{U}$, after a transient phase which depends on the initial condition E_0, E_k is absorbed by a subset of energy levels $\tilde{\mathcal{E}} \subset \mathcal{E}$. As a result, the EHD emulates the behavior of an equivalent EHD with a smaller energy-level set $\tilde{\mathcal{E}}$. Since the system resources are under-utilized, a policy $\eta \notin \mathcal{U}$ is suboptimal.

Since it is sufficient to consider only $\eta \in \mathcal{U}$, the optimization problem in (3) becomes

$$\eta^* = \arg\max_{\eta \in \mathcal{U}} G(\eta). \tag{8}$$

B. Example: $e_{\max} = 1$

Before proceeding to the analysis for general $e_{\max}$, we give a simple example for the case $e_{\max} = 1$. Solving for the steady state distribution, we obtain

$$\pi_\eta(0) = \frac{(1-\bar{b})\eta(1)}{\bar{b} + (1-\bar{b})\eta(1)}, \tag{9}$$

$$\pi_\eta(1) = \frac{\bar{b}}{\bar{b} + (1-\bar{b})\eta(1)}. \tag{10}$$

From (7), the long-term reward is given by

$$G(\eta) = \pi_\eta(1)g(\eta(1)) = \frac{\bar{b}}{\bar{b} + (1-\bar{b})\eta(1)}g(\eta(1)). \tag{11}$$

Note that a large $\eta(1)$ gives a large accrued reward $g(\eta(1))$, but induces frequent energy outage with long-term probability $\Pr(\text{outage}) = 1 - \pi_\eta(1)$. Conversely, a small $\eta(1)$ results in infrequent outage, but less accrued reward $g(\eta(1))$. Therefore, the optimal $\eta^*(1)$ reflects the trade-off between maximizing $g(\eta(1))$ and minimizing $\Pr(\text{outage})$. Taking the derivative of $G(\eta)$ over $\eta(1)$, we obtain

$$\frac{dG(\eta)}{d\eta(1)} \propto g'(\eta(1))\big[\bar{b} + (1-\bar{b})\eta(1)\big] - (1-\bar{b})g(\eta(1)) \triangleq q(\eta(1)),$$

where $\propto$ denotes equality up to a positive factor (which does not affect the sign of $q(\eta(1))$). Using the concavity of the reward function $g(x)$ (Lemma 1), it can be shown that $q(\eta(1))$ is a decreasing function of $\eta(1)$, with $q(0) > 0$ and $q(1) < 0$. Therefore, $G(\eta)$ is maximized at $\eta^*(1)$, which is the unique solution of $q(\eta^*(1)) = 0$.

In the next section, we consider the case $e_{\max} > 1$ and determine structural properties of the optimal policy.

C. Structure of the optimal policy for $e_{\max} > 1$

The sequence $\{(\mathbf{S}_k, Q_k), k \geq 0\}$ is a Markov Decision Process [11], hence the optimal policy can be evaluated numerically as the solution of a Linear Program (LP) [9]. The objective of this section is to characterize the structure of the optimal policy. Our main result is stated in the following theorem.

Theorem 1 *The optimal policy η^* has the following properties:*

P1) $\eta^*(e)$ *is a strictly increasing function of $e \in \mathcal{E}$.*
P2) $\eta^*(e) \in (\eta_L, \eta_U), \forall e \in \mathcal{E} \setminus \{0\}$, *where $\eta_L \in (0, \bar{b})$, $\eta_U \in (\bar{b}, 1)$ uniquely solve*

$$g(\eta_L) + (1 - \eta_L)g'(\eta_L) = \frac{g(\bar{b})}{\bar{b}}, \tag{12}$$

$$g(\eta_U) - \eta_U g'(\eta_U) = g(\bar{b}). \tag{13}$$

Proof: We give an outline of the proof. A detailed proof is provided in [10].

We prove P1 by contradiction. Let $\eta_0 \in \mathcal{U}$ be a generic transmission policy, which violates P1. Then, we can show that there exists $\epsilon \in \{1, \dots, e_{\max} - 1\}$ such that

$$\eta_0(\epsilon - 1) < \eta_0(\epsilon) \geq \eta_0(\epsilon + 1). \tag{14}$$

The violation of P1 is due to the fact that $\eta_0(\epsilon) \geq \eta_0(\epsilon + 1)$, *i.e.*, η_0 is not strictly increasing from ϵ to $\epsilon + 1$. We now define a new transmission policy, η_δ, parameterized by $\delta > 0$, as:

$$\eta_\delta(e) = \begin{cases} \eta_0(e), & e \in \mathcal{E} \setminus \{\epsilon - 1, \epsilon, \epsilon + 1\} \\ \eta_0(\epsilon - 1) + h(\delta), & e = \epsilon - 1 \\ \eta_0(\epsilon) - \delta, & e = \epsilon \\ \eta_0(\epsilon + 1) + r(\delta), & e = \epsilon + 1. \end{cases}$$

Intuitively, policy η_δ is constructed from the original policy η_0 by *transferring* some transmissions from energy state ϵ to states $(\epsilon + 1)$ and $(\epsilon - 1)$, whereas transmissions in all other states are left unchanged. The functions $r(\delta) > 0$ and $h(\delta) \geq 0$ are uniquely defined as follows. If $\epsilon > 1$, the transmission transfer is done by preserving the steady state distribution of visiting the *low* energy states $\{0, \dots, \epsilon - 2\}$ and the *high* energy states $\{\epsilon + 2, \dots, e_{\max}\}$ (a similar consideration holds for $\epsilon = 1$, with the difference that in this case $h(\delta) = 0$, since $\eta_\delta(0) = 0$). Therefore, on the one hand, the new policy η_δ recovers from the structure that violates P1, by diminishing the gap $\eta(\epsilon) - \eta(\epsilon + 1)$ by a quantity $\delta + r(\delta) > 0$; on the other hand, it confines the perturbations on the steady state distribution only to states $\{\epsilon - 1, \epsilon, \epsilon + 1\}$, thus simplifying the analysis. Formally,

1) if $\epsilon = 1$, let $h(\delta) = 0$ and let $r(\delta)$ be such that $\pi_{\eta_\delta}(e_{\max}) = \pi_{\eta_0}(e_{\max}), \forall \delta < \kappa$,
2) if $\epsilon > 1$, let $h(\delta)$ and $r(\delta)$ be such that

$$\begin{cases} \pi_{\eta_\delta}(e_{\max}) = \pi_{\eta_0}(e_{\max}) \\ \pi_{\eta_\delta}(0) = \pi_{\eta_0}(0) \end{cases}, \forall \delta < \kappa, \tag{15}$$

where $0 < \kappa \ll 1$ is an arbitrarily small constant, which guarantees an admissible policy $\eta_\delta \in \mathcal{U}$, *i.e.*, $\eta_\delta(\epsilon - 1) \in (0,1)$, $\eta_\delta(\epsilon) \in (0,1)$, $\eta_\delta(\epsilon + 1) \in (0,1)$ for $\epsilon + 1 < e_{\max}$, $\eta_\delta(\epsilon + 1) \in (0,1]$ for $\epsilon + 1 = e_{\max}$.

It can be shown that this choice also gives $\pi_{\eta_\delta}(e) = \pi_{\eta_0}(e)$ for $\{e < \epsilon - 1\} \cup \{e > \epsilon + 1\}$. In fact, states $\{e < \epsilon - 1\}$ and $\{e > \epsilon + 1\}$ communicate with states $\{\epsilon - 1, \epsilon, \epsilon + 1\}$ only through states $\epsilon - 2$ and $\epsilon + 2$, respectively. Therefore, since the policy is left unchanged in states $\{e < \epsilon - 1\} \cup \{e > \epsilon + 1\}$,

by imposing no perturbation in the steady state distribution of states 0 and $e_{\max}$, no perturbation is induced in states $\{e < \epsilon - 1\}$ and $\{e > \epsilon + 1\}$, respectively. Then, by comparing the long-term rewards under policies η_0 and η_δ, we have

$$D_{\eta_0}(\delta) \triangleq \frac{G(\eta_\delta) - G(\eta_0)}{\delta} = \tag{16}$$

$$\frac{\pi_{\eta_\delta}(\epsilon - 1)g(\eta_0(\epsilon - 1) + h(\delta)) - \pi_{\eta_0}(\epsilon - 1)g(\eta_0(\epsilon - 1))}{\delta}$$

$$+ \frac{\pi_{\eta_\delta}(\epsilon)g(\eta_0(\epsilon) - \delta) - \pi_{\eta_0}(\epsilon)g(\eta_0(\epsilon))}{\delta}$$

$$+ \frac{\pi_{\eta_\delta}(\epsilon + 1)g(\eta_0(\epsilon + 1) + r(\delta)) - \pi_{\eta_0}(\epsilon + 1)g(\eta_0(\epsilon + 1))}{\delta}.$$

Finally, it can be shown that $D_{\eta_0}(0^+) \triangleq \lim_{\delta \to 0^+} D_{\eta_0}(\delta) > 0$, therefore, for some $0 < \kappa \ll 1$ we have $D_{\eta_0}(\delta) > 0, \forall \delta \in (0, \kappa)$. Equivalently, $G(\eta_\delta) > G(\eta_0)$, hence η_0 and any policy violating P1 are *strictly* suboptimal.

For the proof of P2, we proceed as follows. Computing the derivative of $G(\eta)$ with respect to $\eta(1)$, we have, after some algebraic manipulation,

$$\frac{\mathrm{d}G(\eta)}{\mathrm{d}\eta(1)} \propto g(\eta(1)) + (1 - \eta(1))g'(\eta(1)) - \frac{G(\eta)}{\bar{b}}$$

$$> g(\eta(1)) + (1 - \eta(1))g'(\eta(1)) - \frac{g(\bar{b})}{\bar{b}} \triangleq L(\eta(1), \bar{b}),$$

where the second step follows from $G(\eta) < g(\bar{b})$ (see Appendix). Using the concavity of $g(x)$ (Lemma 1), it can be shown that $L(\eta(1), \bar{b})$ is a decreasing function of $\eta(1)$, with $\lim_{x \to 0^+} L(x, \bar{b}) > 0$ and $L(\bar{b}, \bar{b}) < 0$. Therefore, there exists a unique $\eta_L \in (0, \bar{b})$ that solves $L(\eta_L, \bar{b}) = 0$. Then, for all $\eta(1) \le \eta_L$ we have $L(\eta(1), \bar{b}) \ge 0$, hence $\frac{\mathrm{d}G(\eta)}{\mathrm{d}\eta(1)} > 0$, which proves that $\eta(1) \le \eta_L$ is strictly suboptimal.

By computing the derivative of $G(\eta)$ with respect to $\eta(e_{\max})$, we have

$$\frac{\mathrm{d}G(\eta)}{\mathrm{d}\eta(e_{\max})} \propto -g(\eta(e_{\max})) + \eta(e_{\max})g'(\eta(e_{\max})) + G(\eta)$$

$$< -g(\eta(e_{\max})) + \eta(e_{\max})g'(\eta(e_{\max})) + g(\bar{b})$$

$$\triangleq U(\eta(e_{\max}), \bar{b}).$$

Since $g(x)$ is concave, $U(\eta(e_{\max}), \bar{b})$ is a decreasing function of $\eta(e_{\max})$, with $U(\bar{b}, \bar{b}) > 0$ and $\lim_{x \to 1^-} U(x, \bar{b}) < 0$. Therefore, there exists a unique $\eta_U \in (\bar{b}, 1)$ that solves $U(\eta_U, \bar{b}) = 0$. Then, for all $\eta(e_{\max}) \ge \eta_U$ we have $U(\eta(e_{\max}), \bar{b}) \le 0$, hence $\frac{\mathrm{d}G(\eta)}{\mathrm{d}\eta(e_{\max})} < 0$, which proves that $\eta(e_{\max}) \ge \eta_U$ is strictly suboptimal. Finally, by combining these results with P1, we obtain

$$\eta_L < \eta(1) < \eta(2) < \cdots < \eta(e_{\max}) < \eta_U. \tag{17}$$

∎

Remarks on Theorem 1: P1 of Theorem 1 states the intuitive fact that the more energy available in the buffer, the higher the incentive to transmit. Moreover, if the energy level in time-slot k is $E_k = e_{\max}$, energy overflow will occur with probability $\bar{b}(1 - \eta(e_{\max}))$. Therefore, the large transmission probability in the high energy states reflects the incentive to avoid visiting the full energy state $E_k = e_{\max}$, thus
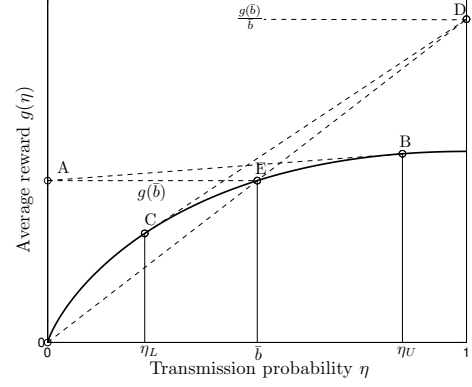

Figure 1. Reward function $g(\eta)$, and geometric interpretation of the upper bound η_U and lower bound η_L to the transmission probability $\eta(e)$.

minimizing the impact of overflow to the system performance. In contrast, the small transmission probability in the low energy states aims to minimize the impact of energy outage, by conservatively prioritizing the battery recharge process over data transmission.

The bounds η_U and η_L of P2 can be interpreted geometrically with the help of Fig. 1. The tangent line drawn from point $\left(1, \frac{g(\bar{b})}{\bar{b}}\right)$ (D) to the curve $\{(\eta, g(\eta)), \eta \in (0, 1)\}$ touches it at $(\eta_L, g(\eta_L))$ (C). Similarly, the tangent line drawn from the point $(0, g(\bar{b}))$ (A) to the curve $\{(\eta, g(\eta)), \eta \in (0, 1)\}$ touches it at $(\eta_U, g(\eta_U))$ (B).

IV. NUMERICAL RESULTS

In this section, we employ the framework developed in Sections II and III to maximize the long-term average throughput from the EHD to the RX, for i.i.d. channel gains H_k, which are exponentially distributed with unit mean, *i.e.*, with pdf $f_H(h) = \mathrm{e}^{-h}$, $h > 0$. The achievable rate in slot k is

$$V_k = \ln(1 + \mathrm{SNR}H_k), \tag{18}$$

where SNR is the average signal-to-noise-ratio at RX. From (5), we have

$$g(\eta(e)) = \int_{h_{\mathrm{th}}(e)}^{+\infty} \ln(1 + \mathrm{SNR}h)\mathrm{e}^{-h}\mathrm{d}h, \tag{19}$$

where $\eta(e)$ is the transmission probability induced by using a channel threshold $h_{\mathrm{th}}(e)$ (corresponding to the rate threshold $v_{\mathrm{th}}(e) = \ln(1 + \mathrm{SNR}h_{\mathrm{th}}(e))$), and is given by

$$\eta(e) = \int_{h_{\mathrm{th}}(e)}^{+\infty} \mathrm{e}^{-h}\mathrm{d}h = \mathrm{e}^{-h_{\mathrm{th}}(e)}. \tag{20}$$

We consider the following policies.
- *Optimal policy (OP)*: numerically evaluated by solving (8) as a linear program [9].
- *Balanced policy (BP)*: a threshold policy which, on average, transmits in all non-zero energy levels with probability $\bar{b}$, *i.e.*, $\eta_{\mathrm{BP}}(e) = \bar{b}$, $e \in \mathcal{E} \setminus \{0\}$. In each non-zero energy state, BP matches the energy consumption
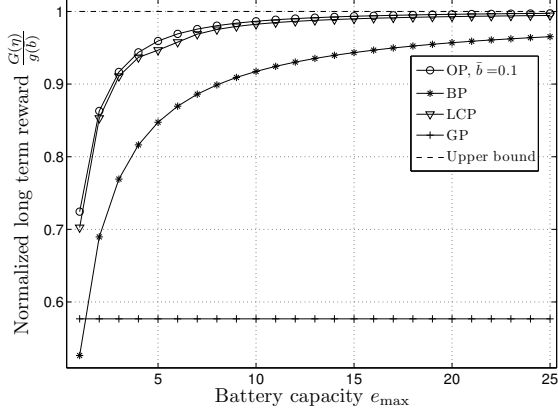
Figure 2. Long-term reward for different policies, normalized to $g(\bar{b})$, versus battery capacity. Energy harvesting rate $\bar{b} = 0.1$. SNR $= 10\,\mathrm{dB}$.

Figure 3. Percentage gain of OP and LCP over BP, versus battery capacity. Energy harvesting rate $\bar{b} \in \{0.1, 0.01\}$, SNR $= 10\,\mathrm{dB}$.

rate to the energy harvesting rate, thus "balancing" the EHD operation. For BP, we have that

$$G(\eta_{\mathrm{BP}}) = \frac{e_{\max}}{e_{\max} + 1 - \bar{b}} g(\bar{b}). \qquad (21)$$

From (21), it is seen that, when the energy storage capacity grows large, BP approaches the upper bound $g(\bar{b})$, *i.e.*, it is asymptotically optimal.

- *Greedy policy (GP)*: always transmits as long as there is energy in the storage unit, independently of the channel realization, *i.e.*, $\eta_{\mathrm{GP}}(e) = 1$, $e \in \mathcal{E} \setminus \{0\}$. For this policy, we have

$$G(\eta_{\mathrm{GP}}) = \bar{b} g(1). \qquad (22)$$

- *Low-complexity policy (LCP)*: based on Theorem 1, and the fact that the balanced policy is asymptotically optimal, we construct a heuristic policy which (a) is conservative when energy is low, (b) emulates the balanced policy in the middle-energy levels, (c) is aggressive when the energy storage capacity is approached. Mathematically, LCP is defined as

$$\eta_{\mathrm{LCP}}(e) = \begin{cases} \eta_{\mathrm{lo}}(e), & 1 \le e \le 3 \\ \bar{b}, & 4 \le e \le e_{\max} - 3 \\ \eta_{\mathrm{hi}}(e), & e_{\max} - 2 \le e \le e_{\max} \end{cases}$$

for $e_{\max} \ge 6$, and

$$\eta_{\mathrm{LCP}}(e) = \begin{cases} \eta_{\mathrm{lo}}(e) & 1 \le e < e_{\max} - 2 \\ \frac{\eta_{\mathrm{lo}}(e) + \eta_{\mathrm{hi}}(e)}{2} & \max\{e_{\max} - 2, 1\} \le e \\ & \le \min\{e_{\max}, 3\} \\ \eta_{\mathrm{hi}}(e) & 3 < e \le e_{\max}, \end{cases}$$

for $e_{\max} < 6$, where

$$\eta_{\mathrm{lo}}(e) = \frac{e - 1}{3} \bar{b} + \frac{4 - e}{3} \eta_L,$$

$$\eta_{\mathrm{hi}}(e) = \frac{e_{\max} - e}{3} \bar{b} + \frac{e + 3 - e_{\max}}{3} \eta_U.$$

and η_L, η_U are defined in (12)-(13). LCP is constructed such that the average transmission probability increases linearly from the lower bound η_L to $\bar{b}$ in the low energy states
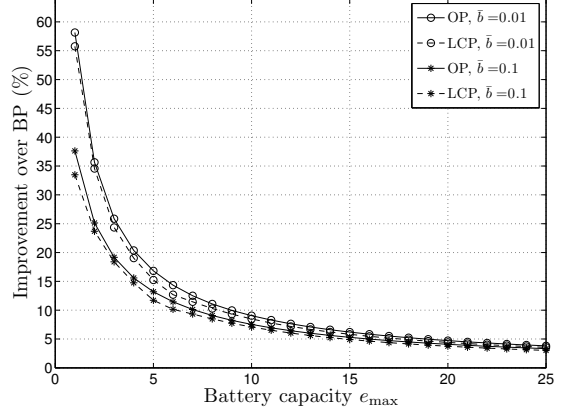
$(1 \le e \le 3)$, it is constant or "balanced" in the middle energy states $(3 < e < e_{\max} - 2)$, and it increases linearly to the upper bound η_U in the high energy levels $(e \ge e_{\max} - 2)$. When $e_{\max} < 6$, the two regions where the transmission probability increases linearly overlap in the interval $\max\{e_{\max} - 2, 1\}, \ldots, \min\{e_{\max}, 3\}$, in which case the transmission probability is taken as the average between $\eta_{\mathrm{lo}}(\cdot)$ and $\eta_{\mathrm{hi}}(\cdot)$. $\eta_{\mathrm{LCP}}(e)$ is plotted versus e for $e_{\max} = 25$ in Fig. 5. Note that LCP can be easily implemented in an environment of changing statistics, *e.g.*, variable $\bar{b}$. In contrast, OP requires the solution of a new LP every time $\bar{b}$ changes, which could be challenging for an EHD with limited computational resources.

In Fig. 2, we plot the long-term reward (throughput) for $\bar{b} = 0.1$, normalized to the upper bound $g(\bar{b})$. The gain of OP with respect to BP, plotted in Fig. 3, is as much as 58% for small storage capacity, and slowly decays to zero as the storage capacity grows large. Figs. 2-3 also demonstrate that the performance of LCP is very close to that of OP for the whole range of $e_{\max}$ considered. Regarding GP, it is seen in Fig. 2 that it performs very poorly, due to the fact that it transmits indiscriminately, wasting energy on channels that can only support a low rate. Moreover, its performance does not vary with $e_{\max}$, as each energy quantum is immediately spent when it becomes available, so that the energy level keeps bouncing between $E_k = 0$ and $E_k = 1$.

To give further insight on the operation mechanisms of the various policies, in Fig. 4 we plot the steady-state distribution of the energy levels for OP, LCP and BP. For BP, it is spread equally likely over all states, so that the system, in the long-term, spends the same amount of time in each state. Therefore, BP cannot prevent the system from visiting the low/high energy states and suffers from frequent energy outage and overflow events. In contrast, for OP and LCP, the states characterized by small/high energy levels are hit less frequently in comparison to BP, since these policies are more conservative in the low energy states (by transmitting with smaller probability than $\bar{b}$), and more aggressive in the high energy states (by transmitting with higher probability than $\bar{b}$).
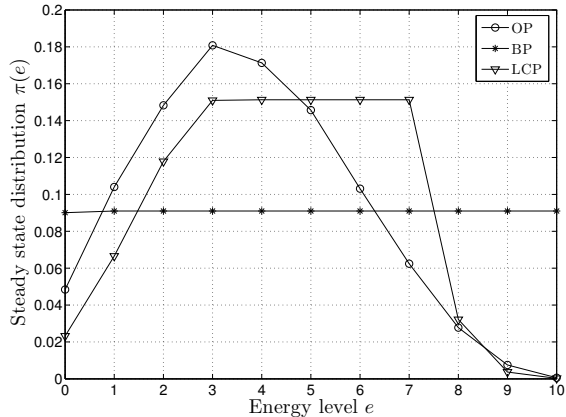
Figure 4. Steady state distributions of the energy levels. Energy harvesting rate $\bar{b} = 0.1$, SNR $= 10\,\mathrm{dB}$, battery capacity $e_{\max} = 10$.

Figure 5. Transmission probability as a function of the energy level. Energy harvesting rate $\bar{b} = 0.1$, SNR $= 10\,\mathrm{dB}$, battery capacity $e_{\max} = 25$.

The result is that the steady-state distribution is concentrated in the intermediate energy levels, where neither outage nor overflow occurs. Hence, in the long-term, the impact of these events to the system performance is minimized.

Finally, in Fig. 5 we plot the transmission probability as a function of the energy level, for OP, LCP and BP. It is seen that LCP follows closely the structure of OP and that the upper and lower bounds to the transmission probability derived in Theorem 1 are approached at $e = e_{\max}$ and $e = 1$, respectively.

V. Conclusions

We derived properties of the optimal binary policy for an EHD which transmits data of different importance, under i.i.d. energy arrival and data importance processes. A three-level suboptimal policy was constructed: conservative in the low energy states, so as to prevent energy outages, aggressive in the high energy states, so as to prevent energy overflows, and balanced in the middle energy states, so as to match the energy consumption rate to the energy rate harvested from the environment. It was demonstrated through numerical examples that this policy attains close to optimal performance.

As part of our future work, we will generalize the assumptions of Bernoulli i.i.d. energy arrival process and of binary policies, and include more realistic energy harvesting scenarios and battery models.

Acknowledgments

Appendix

We derive an upper bound to $G(\eta)$ for any $\eta \in \mathcal{U}$. Since $g(x)$ is strictly concave (Lemma 1), applying Jensen's inequality on (7) we obtain

$$G(\eta) = \sum_{e=0}^{e_{\max}} \pi_\eta(e) g(\eta(e)) < g\left(\sum_{e=0}^{e_{\max}} \pi_\eta(e)\eta(e)\right), \quad (23)$$
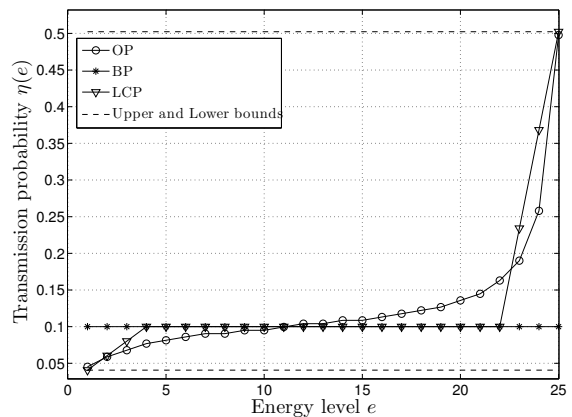
where the strict inequality comes from the fact that $\eta(e) > \eta(0) = 0, \forall e \in \{1, \ldots, e_{\max}\}$. Now, note that (1) implies an average long-term transmission constraint dictated by the average harvesting rate $\bar{b}$, *i.e.*,

$$\sum_{e=0}^{e_{\max}} \pi_\eta(e)\eta(e) \le \bar{b}. \quad (24)$$

Since $g(x)$ is increasing in x (Lemma 1),

$$G(\eta) < g\left(\sum_{e=0}^{e_{\max}} \pi_\eta(e)\eta(e)\right) \le g(\bar{b}). \quad (25)$$

References

[1] J. A. Paradiso and T. Starner, "Energy scavenging for mobile and wireless electronics," *IEEE Pervasive Computing*, vol. 4, pp. 18–27, Jan. 2005.

[2] R. Berry and R. Gallager, "Communication over fading channels with delay constraints," *IEEE Trans. Inform. Theory*, vol. 48, pp. 1135–1149, May 2002.

[3] R. A. Valles, A. G. Marques, and J. G. Sueiro, "Optimal selective forwarding for energy saving in wireless sensor networks," *IEEE Trans. Wireless Commun.*, vol. 10, pp. 164–175, Jan. 2011.

[4] V. Sharma, U. Mukherji, V. Joseph, and S. Gupta, "Optimal energy management policies for energy harvesting sensor nodes," *IEEE Trans. Wireless Commun.*, vol. 9, pp. 1326–1336, Apr. 2010.

[5] 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, pp. 547–551, Feb. 2009.

[6] N. Jaggi, K. Kar, and A. Krishnamurthy, "Rechargeable sensor activation under temporally correlated events," *Springer Wireless Networks (WINET)*, vol. 15, pp. 619–635, Jul. 2009.

[7] A. Seyedi and B. Sikdar, "Energy efficient transmission strategies for body sensor networks with energy harvesting," *IEEE Trans. Commun.*, vol. 58, pp. 2116–2126, Jul. 2010.

[8] F. Iannello, O. Simeone, and U. Spagnolini, "Energy Management Policies for Passive RFID Sensors with RF-Energy Harvesting," in *IEEE International Conference on Communications (ICC)*, May 2010, pp. 1 –6.

[9] K. W. Ross, "Randomized and past-dependent policies for Markov decision processes with multiple constraints," *Operations Research*, vol. 37, pp. 474–477, June 1989.

[10] N. Michelusi, K. Stamatiou, and M. Zorzi, "Optimal transmission policies for energy harvesting devices," Tech. Rep., under preparation.

[11] D. Bertsekas, *Dynamic programming and optimal control*. Athena Scientific, Belmont, Massachusetts, 2005.