Transmission policies for energy harvesting sensors with time-correlated energy supply

Nicolò Michelusi, Member, IEEE, Kostas Stamatiou, Member, IEEE, and Michele Zorzi, Fellow, IEEE

Abstract—This paper considers a wireless sensor powered by an energy harvesting device, which reports data of varying importance to its receiver. Modeling the ambient energy supply by a two-state Markov chain (“GOOD” and “BAD”), assuming a finite battery capacity constraint, and associating data transmission with a given energy cost, we propose low-complexity transmission policies, that achieve near-optimal performance in terms of the average long-term importance of the reported data. In particular, we derive the performance of the Balanced Policy (BP), which adapts the transmission probability to the harvesting state, such that energy harvesting and consumption are balanced. Our analysis demonstrates that the performance of the BP largely depends on the power-to-depletion, defined as the power that a fully charged battery can supply on average over a BAD period. Numerical results show that the optimal BP achieves near-optimal performance and that a BP which avoids energy overflow further reduces the gap with respect to the globally optimal policy. A heuristic BP, based on the analysis of a system with a deterministic and periodic energy supply, is also proposed, and the parallels between the deterministic system and its stochastic counterpart are discussed.

Index Terms—Energy harvesting, Wireless Sensor Networks, Markov Decision Processes

I. INTRODUCTION

Energy Harvesting (EH) devices collect, or harvest, energy from the environment in order to power electronics in a wide range of applications [2], [3]. EH technology is particularly appealing in the deployment of wireless sensor networks, where autonomous operation is required and factors such as the sheer number of nodes or their inaccessibility render battery replacement unrealistic and cost-prohibitive [4]. In contrast to a battery-operated sensor, where energy efficiency and conservation are crucial to prolong the device life-time, in an EH Sensor (EHS) the energy supply is potentially unlimited, but its availability is random and intermittent over time; the focus thus shifts to the management of the harvested energy, so as to ensure a stable operation of the EHS and to minimize the deleterious impact of energy depletion.

In this paper, we are concerned with a fundamental question: how should statistical information on the ambient energy supply be exploited in order to optimize EHS operation, in the presence of a finite battery capacity constraint? We consider a general system model consisting of an EHS, which judiciously reports data of varying importance to its receiver (RX). Practical examples of this setting include: a temperature sensor, where higher temperature readings, being indicators of overheating or fire, are more important; a sensor which acts as a relay of different priority packets in a wireless network [5]; data transmission over a fading channel, where the number of bits which can be reliably transmitted depends on the instantaneous channel realization. Energy is harvested from an ambient source modeled by a two-state Markov chain, where “GOOD” and “BAD” correspond to an abundance and scarcity of ambient energy, respectively, and is stored in the sensor battery. Given that data transmission incurs an energy cost, our objective is to characterize low-complexity transmission policies, which achieve near-optimal performance in terms of the average long-term importance of the reported data.

We derive analytically the performance of a Balanced Policy (BP), which adapts the transmission probability based only on the harvesting state, such that, on average, energy harvesting and consumption are balanced. Numerical results demonstrate that the optimal BP performs very well with respect to the globally optimal policy, and the gap between the two is reduced even further if the sensor is always forced to transmit when the battery is fully charged. The main implication of these results is that near-optimal performance can be achieved with simple adaptation to the ambient energy supply, without precise knowledge of the energy stored in the sensor battery at any given time.

In contrast to prior related work, e.g., [6]–[9], we aim at explicitly capturing the effect of the finite battery capacity on the system performance, given the prolonged charge and discharge cycles induced by a time-correlated EH process, which may lead to energy overflow and depletion, respectively, if the battery is not sufficiently large. In this regard, a key result of this paper is that the EHS performance is heavily dependent on the power-to-depletion \( \rho \), defined as the power that a fully charged battery can supply over a BAD period, such that, on average, it is depleted at the end of the period. Intuitively, a large \( \rho \) indicates that the battery capacity is sufficiently large to “absorb” the fluctuations in the EH process, and to provide a stable energy supply to the sensor circuitry. In the spirit of the “offline” optimization framework of [10]–[12], we also study an EHS with a deterministic and periodic energy supply and show that the optimal policy depends exclusively on \( \rho \). Based on this analysis, we propose a heuristic BP which, in certain regimes, is shown to perform well under the original system model.

A problem similar to the one studied in this paper has
previously been considered in [13]. However, in [13], the battery recharge process is modeled as memoryless, whereas we introduce time correlation in the energy supply. The issue of energy management for solar-powered EHSs and RFID has previously been addressed in [14] and [15], respectively, primarily from a numerical standpoint. [6] derived the policy which maximizes the long-term detection probability of a random event and [8] considered a similar problem in the context of body sensor networks. In [9], data transmission for a two-state time-varying fading channel is studied, and properties of the policy that maximizes the log-term discounted throughput are derived, under the assumptions of infinite battery capacity and i.i.d. Bernoulli EH process. In contrast, our work considers both a finite battery capacity and correlation in the EH process, as well as a more general, but temporally i.i.d., data importance model. Moreover, we consider an average long-term, rather than discounted, reward metric. In [7], [16], policies which stabilize the data queue of an EHS with random data arrivals were proposed and analyzed. [10]–[12] derived policies that maximize the data throughput of the EHS to RX link by a deadline, relying, however, on the assumption that energy arrivals (and the channel fading profile in [10]) are known beforehand. Other related work includes [17], which explored activation policies in a network of EHSs; [18], which derived power management algorithms for EHSs with battery inefficiencies; and, recently, [19], which proposed a framework to model battery degradation phenomena, and derived power management policies to maximize battery lifetime under a quality of service constraint at the EHS. The model considered in the present paper is a generalization of the one studied in [20], where the optimal policy was characterized for a temporally independent energy supply.

The remainder of this paper is organized as follows. In Section II, we describe the system model. The optimization problem and the formal policy definitions are presented in Section III, followed by the analysis of the BP in Section IV. Section V is devoted to the analysis of a continuous-time, deterministic model and to the discussion of its connection to the stochastic model. Numerical results are presented in Section VI, followed by our concluding remarks in Section VII.

We close this section with a note on the notation employed throughout the paper: \( x = 1 - \bar{x} \) is the complement of \( x \in [0, 1] \) and \( \chi(\cdot) \) is the indicator function; random variables and their values are denoted by uppercase and lowercase letters, respectively. A list of symbols is provided in Table I.

### II. System Model

The block diagram of a wireless EHS is shown in Fig. 1. The EH device (EHD) collects ambient energy, which is stored in a battery (or super-capacitor), which powers the sensing apparatus and the RF circuitry. A processing unit, e.g., a microcontroller, manages the energy consumption of the EHS. 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 EHS has a new data packet to send to the RX. We assume that a strict delay constraint is enforced at the EHS: the packet is either sent to the RX or it is dropped.

The EHS battery is modeled by a buffer. As in previous work [6], [8], [20], we assume 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, i.e., the battery capacity, is \( e_{\text{max}} \) and the set of possible energy levels is denoted by \( \mathcal{E} = \{0, 1, \ldots, e_{\text{max}}\} \). At time \( k + 1, k \in \mathbb{Z}^+ \), the amount of energy in the buffer is

\[
E_{k+1} = \min \{ E_k - Q_k + B_k, e_{\text{max}} \},
\]

where \( \{B_k\} \) is the energy arrival process and \( \{Q_k\} \) is the action process. \( Q_k = 1 \) if the current data packet is transmitted, which results in the expenditure of one energy quantum, and \( Q_k = 0 \) otherwise. \( B_k \) models the randomness in the energy harvested in slot \( k \). We assume that \( B_k \in \{0, 1\} \), i.e., either one energy quantum is harvested, or no energy is harvested at all. Moreover, the energy harvested in time-slot \( k \) can be used only in a later time-slot. As a consequence, if the battery is depleted, i.e., \( E_k = 0 \), then \( Q_k = 0 \). We model the underlying EH process \( \{A_k\} \) as a two-state Markov chain, with state space \( \{G, B\} \), where \( G \) and \( B \) denote the GOOD and BAD harvesting states, respectively. If \( A_k = G \) (GOOD state), then

\[1^{\delta} \in \{0, 1\}\] models a typical characteristic of EHS systems (e.g., see [21]): the energy to perform a given task (transmit a packet) is spent much faster than it is collected. Note that the value of \( \delta \) has no impact on the subsequent analysis.

\[2^\text{We only consider the energy expenditure associated with RF transmission.}\]
\( B_k = 1 \) with probability \( \lambda_G \), where \( \lambda_G \in (0, 1] \), and \( B_k = 0 \) with probability \( 1 - \lambda_G \); if \( A_k = B \) (BAD state), then \( B_k = 0 \). When \( \lambda_G < 1 \), in the GOOD state \( 1/\lambda_G > 1 \) time-slots are required on average to harvest one energy quantum, i.e., the amount of energy required for the transmission of one data packet. We denote the transition probabilities of \{\( A_k \)\} from G to G and from B to B as \( p_G = \Pr(A_k = G | A_{k-1} = G) \) and \( p_B = \Pr(A_k = B | A_{k-1} = B) \), respectively. The steady-state distribution of \{\( A_k \)\} is thus

\[
\pi_A(G) = \frac{\bar{p}_B}{\bar{p}_B + \bar{p}_G}, \quad \pi_A(B) = \frac{\bar{p}_G}{\bar{p}_B + \bar{p}_G}. \tag{2}
\]

The average durations of the GOOD and BAD EH periods are denoted by \( D_G \) and \( D_B \), respectively, and their ratio by \( \gamma = D_G/D_B \). Simple calculations yield that \( D_G = 1/\bar{p}_G \), \( D_B = 1/\bar{p}_B \) and \( \gamma = \pi_A(G)/\pi_A(B) \). Finally, since one energy quantum is harvested with probability \( \lambda_G \) in every GOOD time-slot, the average EH rate is

\[
\beta = \lim_{K \to \infty} \frac{1}{K} E \left[ \sum_{k=0}^{K-1} B_k \right] = \lambda_G \pi_A(G), \tag{3}
\]

where \( \beta \in (0, 1) \). Note that \( \beta, \gamma \) and \( \lambda_G \) are related as

\[
\beta = \frac{\lambda_G}{\gamma + 1}. \tag{4}
\]

We now formally define the events of energy outage and overflow.

**Definition 1 (Outage)** In slot \( k \), energy outage occurs if \( E_k = 0 \).

**Definition 2 (Overflow)** In slot \( k \), energy overflow occurs if \( (E_k = e_{\text{max}}) \cap (B_k = 1) \cap (Q_k = 0) \).

Under energy outage, no transmissions can be performed, i.e., \( Q_k = 0 \). Energy overflow occurs when a harvested energy quantum (\( B_k = 1 \)) cannot be stored due to a fully charged battery (\( E_k = e_{\text{max}} \)) in an idle time-slot (\( Q_k = 0 \)), and is thus lost.

The state of the EHS at time \( k \) is given by \( (S_k, V_k) \), where \( S_k = (E_k, A_{k-1}) \in S \) is the joint energy level and EH state, with \( S = \mathcal{E} \times \{G, B\} \), and \( V_k \in \mathbb{R}^+ \) is the importance value of the current data packet. We model \( V_k \) as a continuous random variable with probability density function (pdf) \( f_v(v), v \geq 0 \), and assume that \{\( V_k \)\} are i.i.d. Note that, at time \( k \), the EHS controller can infer the posterior distribution of \( A_{k-1}, \Pr(A_{k-1} = a | B_0, \ldots, B_{k-1}) \) for \( a \in \{G, B\} \), from the observation of the EH process \{\( B_0, \ldots, B_{k-1} \)\}. Herein, we assume that perfect knowledge of \( A_{k-1} \) is available at the EHS controller, and leave the problem of estimating \( A_{k-1} \) as future work.

### III. Optimization Problem and Policy Definitions

#### A. Optimization Problem

Given \( s_k = (e, a) \in S \) and \( V_k = \nu \in \mathbb{R}^+ \), the policy \( \mu \) implemented by the controller in Fig. 1 is defined by the probability \( \mu(e, a, \nu) \) of transmitting the data packet in slot \( k \). The respective probability of discarding the data packet is \( 1 - \mu(e, a, \nu) \). Given an initial state \( S_0 \in S \), the average long-term importance of the reported data (from now on referred to as average reward for brevity) under policy \( \mu \) is

\[
G_\mu(S_0) = \lim_{K \to \infty} \inf \frac{1}{K} \mathbb{E} \left[ \sum_{k=0}^{K-1} Q_k V_k \bigg| S_0 \right]. \tag{5}
\]

The expectation in (5) is taken with respect to \{\( B_k, A_k, Q_k, V_k \)\}, where, at each instant \( k \), \( Q_k \) is drawn according to policy \( \mu \) and depends on the state \( (E_k, A_{k-1}, V_k) \), and \( E_k \) is given by (1).

The optimization problem at hand is to determine the optimal policy \( \mu^* \) such that

\[
\mu^* = \arg \max_\mu G_\mu(S_0). \tag{6}
\]

We now establish that \( \mu^* \) has a threshold structure with respect to the data importance.

**Lemma 1** For each state \((e, a) \in S\), there exists a threshold \( v_{\text{th}}^*(e, a) \) such that

\[
\mu^*(e, a, \nu) = \begin{cases} 1, & \nu \geq v_{\text{th}}^*(e, a), \\ 0, & \nu < v_{\text{th}}^*(e, a). \end{cases} \tag{7}
\]

*Proof:* See Appendix A.

Intuitively, Lemma 1 states that the optimal policy transmits data only if it is of sufficient importance. As a consequence, we henceforth only consider policies with the structure defined in (7). For a threshold policy \( \mu \), the transmission probability in state \((e, a)\) is

\[
\eta(e, a) = \mathbb{E}_V [\mu(e, a, V)] = F_V(v_{\text{th}}^*(e, a)), \tag{8}
\]

where \( F_V(v), v \geq 0 \), is the complementary cumulative distribution function (ccdf) of the importance value process. The expected reported data importance in state \((e, a)\) is \( g(\eta(e, a)) \), where \( g(x), x \in [0, 1] \), is a function defined as

\[
g(x) = \mathbb{E}_V \left[ \chi \left( V \geq F_V^{-1}(x) \right) \right] = \int_0^\infty v f_V(v) dv, \tag{9}
\]

and \( F_V^{-1}(x) \) denotes the inverse of \( F_V(v) \). In words, \( g(x) \) is the expected accrued reward when only the data with importance above the threshold \( v = F_V^{-1}(x) \) is reported. The function \( g(x) \) has the following properties, which are stated without proof.

**Lemma 2** The function \( g(x) \) is strictly increasing, strictly concave in \( x \), with \( g(0) = 0 \) and \( g'(x) = F_V^{-1}(x) \).

From (7) and (8), it is seen that the mapping between a threshold policy \( \mu \) and its respective \( v_{\text{th}}(\cdot) \) and \( \eta(\cdot) \) is one-to-one. Moreover, due to the independence between \((A_k, B_k)\) and \( V_k \), the transition probabilities of the time-homogeneous Markov chain \{\( S_k \)\} are governed by \( \eta \). Therefore, in the remainder of the paper, we refer to a threshold policy \( \mu \) in terms of its corresponding transmission probability function \( \eta(e, a), (e, a) \in S \).

\(^3\) For the sake of maximizing an average long-term reward function of the state and action processes, it is sufficient to consider only stationary policies depending on the present state [22].
B. Policy Definitions

For the sake of mathematical tractability and without loss of optimality in (6), we only consider the set of policies that result in an average reward independent of the initial state $S_0$.

Definition 3 The set $\mathcal{U}$ of admissible policies is defined as

$$\mathcal{U} = \{ \eta : \eta(0,a) = 0, \eta(e,a) \in (0,1], \eta(e,a) \in (0,1),
\quad e = 1, \ldots, e_{\max} - 1, \forall a \in \{G, B\} \}.$$  

It can be shown that the Markov chain $(\{E_k, A_{k-1}\})$ under a policy $\eta \in \mathcal{U}$ has a unique recurrent class. Hence, there exists a unique steady-state distribution, $\pi_\eta(e,a), (e,a) \in S$, independent of $S_0$ [23]. From (5), for any $\eta \in \mathcal{U}$, we have

$$G_\eta = \lim_{K \to \infty} \frac{1}{K} \mathbb{E} \left[ \sum_{k=0}^{K-1} \chi \left( V_k \geq \bar{F}_V^{-1}(\eta(E_k, A_{k-1})) \right) \mid S_0 \right]$$

$$= \sum_{e=1}^{e_{\max}} \sum_{a \in \{G, B\}} \pi_\eta(e,a)g(\eta(e,a)).$$  

(10)

The optimization problem (6) over the class of admissible policies is stated as

$$\eta^* = \arg \max_{\eta \in \mathcal{U}} G_\eta. \tag{11}$$

The optimal policy $\eta^*$ can be found numerically using the Policy Iteration Algorithm (PIA) for infinite horizon, average cost-per-stage problems [24], [25]. In general, $\eta^*$ is a function of the EH state $A_{k-1}$ and the energy available in the battery, $E_k$. This implies a high implementation complexity for three reasons: the controller must make decisions based on the energy level, which may be too computationally intensive for the ultra-low power electronics typically found in practical EHSSs (for example, PIA requires to update iteratively the transmission probability $\eta(e,a)$ for each value of the energy level $e \in \mathcal{E}$ and of the EH state $a \in \{G, B\}$; the transmission probability for each state needs to be stored in a $2 \times e_{\max}$ look-up table, which takes up an amount of memory proportional to the size of the battery; and knowledge of $E_k$ might be hard to obtain or imprecise at best [26], [27]. Motivated by these observations, we focus on the low-complexity Balanced Policy (BP), defined below.

Definition 4 A BP is any policy $\eta \in \mathcal{U}$ such that, for $a \in \{G, B\}$,

$$\eta(e,a) = \begin{cases} \eta_a, & e \in \{1,2,\ldots,e_{\max} - 1\}, \\ \theta + \bar{\theta} \eta_a, & e = e_{\max}, \end{cases}$$

(12)

where $\theta \in \{0,1\}$ is the Overflow Avoidance (OA) parameter and $\eta_G$ and $\eta_B$ are such that

$$\pi_A(G)\eta_G + \pi_A(B)\eta_B = \beta.$$  

(13)

If $\theta = 0$, the transmission probability of the BP depends only on the EH state, i.e., it is $\eta_G$ in the GOOD state and $\eta_B$ in the BAD state. If $\theta = 1$, the sensor always transmits when the battery is fully charged, thus avoiding energy overflow (Def. 2). OA introduces a mild dependence of the BP on the energy level, since the controller is required to know when the battery is fully charged.

According to (13), the BP “balances” the average energy consumption rate (left hand side of (13)) with the average EH rate (right hand side of (13)), if the impact of energy outage and overflow due to the finite battery capacity is neglected. Alternatively, since $\gamma = D_G/D_B = \pi_A(G)/\pi_A(B)$ and $\beta = \lambda_G\pi_A(G)$, (13) is equivalent to $D_G(\lambda_G - \eta_G) = D_B\eta_B$, i.e., under the BP, an equilibrium amongst the recharge/discharge phases is achieved, in the sense that the expected energy recharge over the GOOD EH period, $D_G(\lambda_G - \eta_G)$, equals the expected energy discharge over the BAD EH period, $D_B\eta_B$.

From (12) and (13), it is seen that a BP is uniquely defined by the parameters $(\eta_G, \theta)$, where $\eta_G \in (\max\{\lambda_G - \gamma^{-1}, 0\}, \lambda_G)$ and $\theta \in \{0,1\}$. In the remainder of the paper, we thus refer to a BP $\eta$ in terms of its corresponding pair $(\eta_G, \theta)$. The next section is devoted to the derivation of the average reward under the BP and the characterization of the optimal BP.

IV. PERFORMANCE ANALYSIS OF THE BP

The main theoretical result of this section is a closed-form expression of the average reward of the BP and is presented in Theorem 1. The proof involves a crafty manipulation of the steady-state equations of the Markov chain $(E_k, A_{k-1})$ and is found in Appendix B. The complicated general expression does not lend itself to a straightforward interpretation. We thus consider an asymptotic regime where energy arrivals are highly correlated and the battery capacity is very large. In this regime, we derive the average reward and its main properties (Theorem 2), and characterize the optimal BP (Lemma 4).

Theorem 1 The average reward of the BP $(\eta_G, \theta)$ is

$$G_\eta = (\pi_A(G) - \eta_G(0,G))g(\eta_G) + (\pi_A(B) - \eta_G(0,B))g(\eta_B)$$

$$+ \theta\eta_G(e_{\max}, G)(g(1) - g(\eta_G)),$$  

(14)

where

$$\begin{bmatrix} \pi_G(0, G) \\ \pi_G(0, B) \end{bmatrix} = \mathbf{ZJ}^{e_{\max} - 1} \mathbf{t}(\theta) \pi_\eta(e_{\max}, G),$$

(15)

$$\pi_G(e_{\max}, G) = \begin{bmatrix} 1 & 1 \end{bmatrix} \mathbf{ZJ}^{e_{\max} - 1} \mathbf{t}(\theta)$$

$$+ \sum_{e=0}^{e_{\max} - 1} \begin{bmatrix} 1 & 1 \end{bmatrix} \mathbf{J}^e \mathbf{t}(\theta) - \theta \frac{\eta_G}{\gamma},$$  

(16)

and

$$\mathbf{t}(\theta) = \begin{bmatrix} \theta & 0 \\ \frac{\eta_G}{\theta} & \frac{1}{\theta} \end{bmatrix},$$

(17)

$$\mathbf{Z} = \begin{bmatrix} \eta_G \lambda_G & \eta_B \lambda_G \\ \eta_B^{\gamma - 1} \eta_G & \eta_G^{\gamma - 1} \lambda_G \end{bmatrix},$$

(18)

$$\mathbf{J} = \begin{bmatrix} \frac{\eta_G}{\eta_B} & \frac{\eta_G}{\eta_B} \lambda_G \\ \frac{\eta_G}{\eta_B} (\frac{\eta_G}{\eta_B} - 1) & \frac{\eta_G}{\eta_B} \lambda_G \end{bmatrix},$$

(19)

Proof: See Appendix B.
The interpretation of (14) is as follows. The terms $(\pi_A(G) - \pi_n(0, G))g(\eta_G)$ and $(\pi_A(B) - \pi_n(0, B))g(\eta_B)$ are the average rewards accrued in the GOOD and BAD states, respectively, where the terms $\pi_n(0, G)$ and $\pi_n(0, B)$ account for the performance loss due to energy outage events. The last term $\theta_n(\epsilon_{\max}, G) (g(1) - g(\eta_G))$ accounts for the impact of OA, i.e., the fact that, in state $(\epsilon_{\max}, G)$, a data packet is always transmitted irrespective of its importance.

In general, $G_n$ does not admit a simple expression, due to the presence of the matrix exponential $J^c$ in (15) and (16). However, a simple expression can be obtained when $\lambda_G = 1$, i.e., when one energy quantum is always harvested in the GOOD state [1].

**Lemma 3** If $\lambda_G = 1$, the average reward of a BP $n$ with parameters $(\eta_G, \theta)$ is

$$G_n = \beta g(\eta_G) + \frac{\epsilon_{\max} - \eta_G}{\epsilon_{\max} + \eta_B(\theta)} g(\eta_B) \tag{20}$$

$$+ \theta \frac{1 + \eta_B B - 1 - \eta_G}{\epsilon_{\max} + \eta_B(\theta)} (g(1) - g(\eta_G)).$$

**Proof:** Letting $\lambda_G = 1$ in (19), the second term becomes zero and we can verify that $J^n = J, \forall n \geq 1$. Therefore, (15)-(16) can be readily computed and Eq. (20) follows from (14).

In order to understand (20), let us focus on a simple non-adaptive BP (NABP) which always transmits with probability $\beta$ and performs no OA, i.e., $(\eta_G, \theta) = (\beta, 0)$. In this case,

$$G_n = g(\beta) \left( \beta + \frac{\epsilon_{\max} - \beta}{\epsilon_{\max} + \beta(\theta)} \right). \tag{21}$$

Moreover, let $\epsilon_{\max} \gg 1$. This is typical of real EHS deployments, e.g., in [21] the capacitance is much larger than the energy to transmit a packet. In slow-changing environments, it is also reasonable to assume that $D_B \gg 1$. In this setting, we can approximate $G_n$ in (21) as

$$G_n \approx g(\beta) \left( \beta + \frac{\rho}{\rho + \beta} \right), \tag{22}$$

where we have defined $\rho \equiv \frac{\epsilon_{\max}}{D_B}$. Note that $\rho$ is the normalized power\(^4\) that can be continuously supplied from a fully charged battery over a BAD period, such that, on average, the battery is empty at the end of that period; we thus name $\rho$ the power-to-depletion. Its effect on the performance of NABP can be explained as follows. In an ideal scenario with infinite battery capacity, NABP is optimal, owing to the concavity of the function $g(x)$ (Lemma 2). $D_B \beta$ is the expected amount of energy drawn from the battery during the BAD period, and $D_G(\lambda_G - \beta) = D_B \beta$ is the expected amount of energy by which the battery is recharged over the GOOD period. If a system with finite battery capacity is operated under this policy, $\rho$ captures the extent to which the battery can absorb the fluctuations in the EH process. If $\rho \gg \beta$, i.e., the power-to-depletion is much larger than the transmission probability, the battery has a “large” capacity and can sustain a constant energy consumption rate $\beta$, while being rarely subject to outage and overflow events. In contrast, if $\rho \ll \beta$, the battery has a “small” capacity, hence it is deeply discharged over the BAD EH period, and fully recharged over the GOOD EH period. The performance is thus severely affected by energy outage and overflow, as can be seen from (22): letting $\rho \to 0$, $G_n \approx \beta g(\beta)$, i.e., transmitting with constant probability $\beta$ achieves only a fraction $\beta$ of the theoretical upper bound $g(\beta)$. This indicates that, for $\rho \ll \beta$, adaptation to the EH state is critical to achieve good performance.

In order to study the impact of $\rho$ on the performance of a general BP, we focus on the asymptotic regime $D_B \to \infty$, $D_G \to \infty$ and $\epsilon_{\max} \to \infty$, where the ratios $\rho = \epsilon_{\max}/D_B$ and $\gamma = D_G/D_B$ (hence $\beta = \lambda_G(1/(1 + \gamma))$) are kept fixed. This regime corresponds to a scenario of extreme correlation in the EH process, where the GOOD and BAD periods are much longer than a time slot, and the battery capacity is much larger than an energy quantum. For the BP $(\eta_G, \theta)$, we denote the asymptotic average reward as $G^{(\infty)}(\eta_G, \theta; \rho) = \lim_{D_B \to \infty} G_n$. In Theorem 2, we derive $G^{(\infty)}(\eta_G, \theta; \rho)$ and characterize its main properties.

**Theorem 2** The asymptotic average reward for the BP $(\eta_G, \theta)$ is

$$G^{(\infty)}(\eta_G, \theta; \rho) = \pi_A(A) g(\eta_G) + \pi_A(B) \frac{\rho}{\rho + \eta_B} g(\eta_B) \tag{23}$$

$$+ \theta \pi_A(A) \frac{\lambda_G - \eta_G}{1 - \eta_G} \frac{\eta_B}{\rho + \eta_B} (g(1) - g(\eta_G)).$$

**Proof:** See Appendix C.

Eq. (23) is a generalization of (22) to any BP $(\eta_G, \theta)$ and any $\lambda_G \leq 1$. Property 1 shows that OA increases the (asymptotic) average reward for any $\eta_G$ and $\rho$. Intuitively, without OA, part of the energy is lost due to overflow, whereas, with OA, all the harvested energy is used towards data transmission. Property 2 generalizes our previous observations on the performance of NABP: for any BP $(\eta_G, \theta)$, the (asymptotic) average reward increases with $\rho$, i.e., as the battery capacity becomes larger with respect to $D_B$. For $\rho \to \infty$, there is no outage nor overflow, which explains the limit in (24). In contrast, for $\rho \to 0$, the battery is almost surely led to outage in the BAD state and, in the long term, reward is only accrued in the GOOD state; hence the limit in (25).

Having derived the asymptotic average reward for any BP, we now characterize the optimal BP in the asymptotic regime. Let $\eta_{G^*}(\theta; \rho) = \arg \max_{\eta_G} G^{(\infty)}(\eta_G, \theta; \rho)$.\(^4\)Note that $\rho$ has units of [energy quanta]/[time-slots], hence it represents a normalized power.
**Lemma 4** The optimal BP in the asymptotic regime, \( \eta_G^*(\theta; \rho) \), is the unique solution of

\[
L(\eta_G, \theta; \rho) = 0
\]

in \( \eta_G \in (\max\{\lambda_G - \gamma^{-1}, 0\}, \lambda_G) \), where

\[
L(\eta_G, \theta; \rho) = \left(1 + \frac{\eta_B}{\rho}\right)^2 g'(\eta_G) - \left(1 + \frac{\eta_B}{\rho}\right) g'(\eta_B) + \frac{g(\eta_B)}{\rho} - \frac{\eta_B}{\rho} \left(\eta_G + \lambda_G + \frac{\eta_B}{\rho} \lambda_G\right) (g(1) - g(\eta_G)) - \frac{\theta \lambda_G - \eta_B}{1 - \eta_G} \left(1 + \frac{\eta_B}{\rho}\right) g'(\eta_G).
\]

Moreover:

1. \( \eta_G^*(\theta; \rho) \in (\beta, \lambda_G) \);
2. \( \eta_G^*(1; \rho) < \eta_G^*(0; \rho) \);
3. \( \eta_G^*(\theta; \rho) \) is a decreasing function of \( \rho \), for \( \rho \geq \beta \), and

\[
\lim_{\rho \to \infty} \eta_G^*(\theta; \rho) = \beta,
\]

\[
\lim_{\rho \to 0} \eta_G^*(\theta; \rho) = \lambda_G.
\]

**Proof:** See Appendix D.

The main implication of Lemma 4 is that the optimal BP in the asymptotic regime can be easily found numerically: from property 1, we know that \( \eta_G^* \) lies in \( (\beta, \lambda_G) \). Moreover, \( L(\eta_G, \theta; \rho) \) is a decreasing function of \( \eta_G \), with \( L(\beta, \theta; \rho) > 0 \) and \( L(\lambda_G, \theta; \rho) < 0 \) (see Appendix D). Hence, \( L(\eta_G, \theta; \rho) = 0 \) can be solved using the bisection method [28].

Property 2) reveals that the optimal BP without OA is more “aggressive” in the GOOD state than the optimal BP with OA. In other words, since there is no protection from overflow, the policy itself tries to minimize energy spillover by forcing a higher consumption rate when energy is available.

Finally, property 3) provides yet further insight as to the characteristics of \( \eta_G^*(\theta; \rho) \). In the limit \( \rho \to \infty \), the battery capacity is large enough, so that transmitting with constant probability \( \beta \) is optimal. From (24), it is seen that \( G^{(\infty)} \to g(\beta) \), i.e., the upper bound is achieved. In contrast, when \( \rho \to 0 \), the battery capacity is so small relative to the time scale of the EH process that it is optimal to use all the energy as it is being harvested during the GOOD state, i.e., \( \eta_G^* \to \lambda_G \). In this case, (24) yields \( G^{(\infty)} \to \pi_A(G)g(\lambda_G) \). As \( \rho \) takes values from 0 to \( \infty \), \( \eta_G^* \) decreases from \( \lambda_G \) to \( \beta \), and the optimal \( G^{(\infty)} \) increases from \( \pi_A(G)g(\lambda_G) \) to \( g(\beta) \).

### V. CONTINUOUS-TIME MODEL WITH DETERMINISTIC EH PROCESS

In this section, we investigate a continuous-time model with a deterministic and periodic EH process. We refer to this model as CDM, to differentiate it from the discrete-time stochastic model of Section II, which, in this section, we denote as DSM. The motivation for considering CDM is twofold. In the asymptotic regime \( D_B, D_G, e_{\max} \to \infty \) with \( \rho = e_{\max}/D_B \) and \( \gamma = D_G/D_B \) fixed, DSM resembles a continuous-time-energy model, in that the time-slot duration is much smaller than the time scale of the EH process and the battery capacity is much larger than an energy quantum. Moreover, in CDM, the controller has perfect knowledge of the EH profile, similar to the offline optimization problems addressed in [11], [12]. It is thus of interest to derive further insight on DSM using CDM and also to draw parallels with [11], [12].

Adhering to the notation in Section II, in CDM, the battery capacity is denoted by \( e_{\max} \). The EH process is periodic with period \( D = D_G + D_B \), where \( D_G \) and \( D_B \) denote the (deterministic) durations of the GOOD and BAD periods. Mathematically, letting \( t \) be a time counter, when \( t \in T_G \), the EH state is GOOD and, when \( t \in T_B \), it is BAD, where \( T_G = \bigcup_{k \in \mathbb{Z}^+} T_G(k) \) and \( T_B = \bigcup_{k \in \mathbb{Z}^+} T_B(k) \), \( T_G(k) = [kD, kD + D_G) \) and \( T_B(k) = [kD + D_G, (k+1)D) \) denote the sets of GOOD and BAD time intervals, respectively. During the GOOD period, energy is harvested at rate \( \lambda_G \), and, during the BAD period, no energy is harvested; therefore, as in DSM, the average EH rate is \( \beta = \lambda_G D_G/D \), and we denote the average long-term fraction of time spent in the GOOD and BAD EH periods as \( \pi_A(G) = D_G/D \) and \( \pi_A(B) = D_B/D \), respectively, with \( \gamma = D_G/D_B = \pi_A(G)/\pi_A(B) \). A policy for CDM is defined by the energy drawing rate \( \eta^{(CDM)}(t) \in [0, 1] \), which specifies the rate according to which energy is drawn from the battery. In particular, if the battery is empty, then \( \eta^{(CDM)}(t) = 0 \); if it is full for \( t \in T_G \), then \( \eta^{(CDM)}(t) \geq \lambda_G \), so that no energy is lost due to overflow. The energy level at time \( t \), \( \eta(t) \), is thus given by

\[
E(t) = \min \left\{ E(kD) + \int_{kD}^{(k+1)D} \left[ \lambda_G - \eta^{(CDM)}(\tau) \right] d\tau, e_{\max} \right\},
\]

for \( t \in T_G(k) \) and, for \( t \in T_B(k) \),

\[
E(t) = E(kD + D_G) - \int_{kD + D_G}^{(k+1)D} \eta^{(CDM)}(\tau) d\tau.
\]

Since the EH process is periodic, it is sufficient to consider a periodic policy

\[
\eta^{(CDM)}(t + kD) = \eta^{(CDM)}(t), \quad \forall t \in [0, D), \quad \forall k \in \mathbb{Z}^+.
\]

Therefore, the following boundary conditions must hold for \( k \in \mathbb{Z}^+ \) (after, possibly, an initial transient phase, whose impact on the average long-term reward is negligible)

\[
E(kD + e_L) = e_L, \quad E(kD + D_G + e_H) = e_H,
\]

i.e., during the GOOD period, the battery is recharged from \( e_L \) to \( e_H \), and, during the BAD period, it is discharged from \( e_H \) to \( e_L \). By definition, \( e_L, e_H \in [0, e_{\max}] \) and \( e_L \leq e_H \).

We define the instantaneous reward rate in CDM as \( g(\eta^{(CDM)}(t)) \). Note that we employ the same mathematical reward function as DSM for the purpose of comparing the two models later in Lemma 6. However, it is emphasized that the physical meaning of \( g(x) \) is different for each model: in CDM, the argument \( x \) is the rate according to which energy is drawn from the battery, and \( g(x) \) is the corresponding instantaneous reward rate; in DSM, \( x \) is the transmission probability and \( g(x) \), defined in (9), is the corresponding expected data importance. With these remarks in place, the average long-term reward per unit time in CDM is

\[
G^{(CDM)}(\eta^{(CDM)}) = \frac{1}{D} \int_0^D g(\eta^{(CDM)}(\tau)) d\tau.
\]
The optimal policy in CDM is the solution of
\[ \eta^{(CDM)*} = \arg \max_{\eta^{(CDM)}} G_{CDM}(\eta^{(CDM)}) . \] (34)

The following lemma determines \( \eta^{(CDM)*} \) and the respective optimal average reward per unit time. As in DSM, let \( \rho = \epsilon_{\infty}/D_B \) be the power-to-depletion in CDM.

**Lemma 5** The optimal policy for CDM is
\[ \eta^{(CDM)*}(t) = \begin{cases} \lambda_G - \frac{1}{\gamma} \min\{\beta, \rho\}, & t \in T_G, \\ \min\{\beta, \rho\}, & t \notin T_B, \end{cases} \] (35)
and the optimal average reward per unit time is
\[ G_{CDM}(\eta^{(CDM)*}) = \pi_A(G)g\left(\lambda_G - \frac{1}{\gamma} \min\{\beta, \rho\}\right) \]
\[ + \pi_A(B)g\left(\min\{\beta, \rho\}\right). \] (36)

**Proof:** See Appendix E.

Lemma 5 distinguishes two regimes of operation in CDM which are determined exclusively by the relation between \( \rho \) and \( \beta \). As illustrated in Fig. 2, if \( \rho \geq \beta \), energy can be drawn with constant rate \( \beta \) and the optimal average reward per unit time is thus \( g(\beta) \). If \( \rho < \beta \), energy is drawn with rate \( \lambda_G - \rho/\gamma = \lambda_G - \epsilon_{\infty}/D_B > \beta \) during the GOOD phase, and with rate \( \epsilon_{\infty}/D_B < \beta \) during the BAD phase, i.e., the battery is completely recharged and discharged over each cycle \( (e_H = \epsilon_{\infty}, e_L = 0) \). Under the prism of [11], [12], the optimal energy expenditure curve in the interval \([kD, kD + D]\) is the unique minimum-length curve that lies in the feasible energy “tunnel” defined by the energy arrival curve, and its downward-shifted version by \( \epsilon_{\infty} \). If \( \rho \geq \beta \), the slope of the expenditure curve is constant and equal to \( \beta \), whereas, if \( \rho < \beta \), it is \( \lambda_G - \epsilon_{\infty}/D_G \) during the GOOD phase and \( \epsilon_{\infty}/D_B < \beta \) during the BAD phase.

Note that, both in Lemma 4 and in Lemma 5, the value of \( \rho \) essentially determines the optimal policy. The main difference is that in CDM the EH profile is completely known, thus the energy consumption rate can be optimally adjusted to completely avoid outage and overflow. In contrast, in DSM energy outage may occur, and energy may also be wasted (if OA is not employed) due to the randomness in the energy arrivals. In the following lemma, we formalize these intuitive remarks by comparing the asymptotic average reward for DSM, \( G^{(\infty)}(\eta_G; \theta, \rho) \), with \( G_{CDM}(\eta^{(CDM)*}) \).

**Lemma 6** For any BP \( (\eta_G, \theta) \) in DSM,
\[ G^{(\infty)}(\eta_G; \theta, \rho) \leq G_{CDM}(\eta^{(CDM)*}). \] (37)

**Proof:** See Appendix F.

We close this section by proposing the following suboptimal policy for DSM, based on the optimal policy found for CDM in Lemma 5: \( \forall e \in \mathcal{E} \setminus \{0\} \),
\[ \eta(e, G) = \lambda_G - \min\{\beta, \rho\}/\gamma, \quad \eta(e, B) = \min\{\beta, \rho\}. \] (38)
This is a BP for DSM, as can be verified from Def. 4, and its performance can thus be evaluated analytically from Theorem 1.

**VI. NUMERICAL RESULTS**

In this section, we present numerical performance results for the policies that have been considered so far in this paper:

- **Optimal policy (OP)**, obtained numerically via the PIA [25];
- **Optimal BP with OA (OBP-OA) and without OA (OBP)**, determined by solving (26) for \( \theta = 1 \) and \( \theta = 0 \), respectively, using the bisection method;
- **Heuristic BP (HBP)**, defined in (38);
- **Non-Adaptive BP (NABP)**, the BP with \( (\eta_G, \theta) = (\beta, 0) \);
- **Greedy Policy (GP)**, which always transmits when there is energy in the buffer.

The average reward of OP is computed numerically via the PIA [25]; those of OBP-OA, OBP, HBP and NABP can be computed analytically from Theorem 1. By definition, the average reward of GP is \( \beta g(1) \).

For the purposes of this section, we let \( V_k = \log_2(1 + \Lambda H_k) \), where \( H_k \) is exponentially distributed with unit mean and \( \Lambda > 0 \) is a scaling parameter. This choice of \( V_k \) corresponds to the information rate achievable on a Rayleigh fading channel with gain \( H_k \), where \( \Lambda \) is the average receive SNR, and the transmitter and receiver have full channel state information, so that the former can perform rate adaptation, whereas the latter can employ coherent detection [29]. The cdf of the data importance is \( \tilde{F}_V(v) = \Pr(H_k \geq 2^{\frac{v-1}{\Lambda}}) = \exp\left\{-\frac{2^{v-1}}{\Lambda}\right\} \).

From (9),
\[ g(x) = \int_{-\ln x}^{\infty} \log_2(1 + \Lambda h)e^{-h}dh \]
\[ = x \log_2(1 - \Lambda \ln x) + \log_2(x)e^{-x}E_1(\Lambda^{-1} - \ln x), \] (39)
where \( E_1(x) \) is the exponential integral function [30], defined as \( E_1(t) = \int_t^{\infty} e^{-\tau}/\tau d\tau \), and computed in Matlab using \( \text{expint}(t) \). Note that (39) is a generalization of [31, Eq. (17)] and [32, Eq. (5)] for \( x \geq 1 \). Unless otherwise stated, we let \( \beta = 0.25 \) and \( \lambda_G = 0.5 \); hence, from (4), \( \gamma = 1 \). Moreover, we set \( \Lambda = 6.31 \), which corresponds to an average SNR of 8 dB.

The numerical results provided in this section are derived for representative values of the system parameters. However, we have verified that the following observations hold for broader parameter ranges.

In Fig. 3, we plot the average reward as a function of \( \epsilon_{\infty} \), for \( \rho \in \{0.2\beta, \beta, 5\beta\} \). For each value of \( \rho \), \( D_B \) is
Figure 3. Average reward as a function of $e_{\text{max}}$ for the considered policies. The respective asymptotic average rewards, obtained from (23), are plotted with bold markers in the right side of each subplot. The performance is heavily dependent on $\rho$ and only mildly affected by the absolute value of $e_{\text{max}}$. ($\beta = 0.25$, $\lambda_G = 0.5$ and SNR = 8 dB)

Figure 4. Asymptotic average reward (23) vs. $\rho/\beta$. ($\beta = 0.25$, $\lambda_G = 0.25$ and SNR = 8 dB)

Figure 5. Transmission probability in the GOOD EH state vs. $\rho/\beta$ corresponding to Fig. 4. HBP resembles OBP for small and large $\rho/\beta$; this explains the behavior of the respective reward curves in Fig. 4. ($\beta = 0.25$, $\lambda_G = 0.5$, SNR = 8 dB)

In Fig. 4, we plot the asymptotic average reward (23) for the considered policies as a function of $\rho/\beta$, and, in Fig. 5, the corresponding $\eta_G$ for the BPs (in Fig. 5, we do not plot OP and GP, since in the former the transmit probability is also a function of the energy level in the battery, whereas the latter transmits with probability one whenever energy is available). For OP in particular, an approximation of the asymptotic average reward is obtained by using the PIA for $e_{\text{max}} = 100$. In Fig. 4, we also plot the curve for CDM, which is an upper bound for the asymptotic average reward achieved by any BP, as proved in Lemma 6. We note that OBP and OBP-OA are within 5% and 2.5% of OP, respectively, for all values considered. HBP, proposed in Section V, attains close to optimal performance for very large values of $\rho/\beta$ and for $\rho/\beta \rightarrow 0$. This behavior is explained in Fig. 5, where it is seen that the transmission probability of HBP approaches that of OBP for these ranges of $\rho$. However, for $\rho$ in the vicinity of $\beta$, HBP incurs a performance loss, which shows that offline policies are not always suitable in a random setting. NABP performs poorly for small values of $\rho$ (60% loss compared to OP) and approaches OBP (and the upper bound $g(\beta)$) for large values of $\rho$. The properties of the optimal transmission probability for OBP and OBP-OA, derived in Lemma 4, are confirmed in Fig. 5.

Overall, OBP performs so well with respect to OP because it adjusts the transmission probability in the BAD state to avoid outage and in the GOOD state to avoid overflow, which are the main factors that cause a performance degradation in a finite-capacity system. If they are avoided, then close-to-optimal performance can be achieved, even without exact knowledge of the energy level in the battery at any given time. As the power-to-depletion $\rho$ increases, the battery becomes more and more resilient to the randomness of the ambient energy, and adaptation is less crucial. As shown in Figs. 4 and 5, in the limit of large $\rho$, it becomes optimal to transmit with constant probability $\beta$ irrespective of the state of the EH process.

determined as $D_B = e_{\text{max}}/\rho$ and $D_G = \gamma D_B$. The asymptotic average reward (23) is shown with a bold marker in the right side of each subplot. Note that, for all policies, $G$ quickly approaches the asymptotic value, i.e., for $e_{\text{max}} \gtrsim 20$, and displays a constant behavior as a function of $e_{\text{max}}$. This suggests that the absolute value of $e_{\text{max}}$ only mildly affects the system performance in the range $e_{\text{max}} \gtrsim 20$. In general, the performance of all policies except GP improves with increasing $\rho$, and approaches more closely the upper bound $g(\beta)$. It is seen that OBP incurs only a small performance degradation with respect to OP: within 6%, for all values of $\rho$ and $e_{\text{max}} \geq 12$. OA reduces the gap even further: within 3% of OP, for $e_{\text{max}} \geq 4$. As discussed in Section IV, NABP, which does not adapt the transmission probability to the EH state, approaches OBP for large values of $\rho$, but incurs a significant performance loss for small values of $\rho$ ($\sim 35\%$ compared to OP).
In agreement with all previous results, the performance loss both (EHS) in the presence of a stochastic ambient energy source.

For the broad range of parameter values considered in the paper, the optimal BP consumption and harvesting are balanced. For the broad range solely adapt to the EH state, such that, on average, energy overflow due to the fact that the power-to-depletion parameter values considered in the paper, the optimal BP consumption and harvesting are balanced. For the broad range solely adapt to the EH state, such that, on average, energy overflow due to the fact that the power-to-depletion parameter values considered in the paper, the optimal BP consumption and harvesting are balanced. For the broad range solely adapt to the EH state, such that, on average, energy overflow due to the fact that the power-to-depletion parameter values considered in the paper, the optimal BP consumption and harvesting are balanced. For the broad range solely adapt to the EH state, such that, on average, energy overflow due to the fact that the power-to-depletion parameter values considered in the paper, the optimal BP consumption and harvesting are balanced. For the broad range solely adapt to the EH state, such that, on average, energy overflow due to the fact that the power-to-depletion parameter values considered in the paper, the optimal BP consumption and harvesting are balanced. For the broad range solely adapt to the EH state, such that, on average, energy overflow due to the fact that the power-to-depletion parameter values considered in the paper, the optimal BP consumption and harvesting are balanced. For the broad range solely adapt to the EH state, such that, on average, energy overflow due to the fact that the power-to-depletion parameter values considered in the paper, the optimal BP consumption and harvesting are balanced. For the broad range solely adapt to the EH state, such that, on average, energy overflow due to the fact that the power-to-depletion parameter values considered in the paper, the optimal BP consumption and harvesting are balanced. For the broad range solely adapt to the EH state, such that, on average, energy overflow due to the fact that the power-to-depletion parameter values considered in the paper, the optimal BP consumption and harvesting are balanced. For the broad range solely adapt to the EH state, such that, on average, energy overflow due to the fact that the power-to-depletion parameter values considered in the paper, the optimal BP consumption and harvesting are balanced. For the broad range solely adapt to the EH state, such that, on average, energy overflow due to the fact that the power-to-depletion parameter values considered in the paper, the optimal BP consumption and harvesting are balanced. For the broad range solely adapt to the EH state, such that, on average, energy overflow due to the fact that the power-to-depletion parameter values considered in the paper, the optimal BP consumption and harvesting are balanced. For the broad range solely adapt to the EH state, such that, on average, energy overflow due to the fact that the power-to-depletion parameter values considered in the paper, the optimal BP consumption and harvesting are balanced. For the broad range solely adapt to the EH state, such that, on average, energy overflow due to the fact that the power-to-depletion parameter values considered in the paper, the optimal BP consumption and harvesting are balanced. For the broad range solely adapt to the EH state, such that, on average, energy overflow due to the fact that the power-to-depletion parameter values considered in the paper, the optimal BP consumption and harvesting are balanced. For the broad range solely adapt to the EH state, such that, on average, energy overflow due to the fact that the power-to-depletion parameter values considered in the paper, the optimal BP consumption and harvesting are balanced. For the broad range solely adapt to the EH state, such that, on average, energy overflow due to the fact that the power-to-depletion parameter values considered in the paper, the optimal BP consumption and harvesting are balanced. For the broad range solely adapt to the EH state, such that, on average, energy overflow due to the fact that the power-to-depletion parameter values considered in the paper, the optimal BP consumption and harvesting are balanced. For the broad range solely adapt to the EH state, such that, on average, energy overflow due to the fact that the power-to-depletion parameter values considered in the paper, the optimal BP consumption and harvesting are balanced. For the broad range solely adapt to the EH state, such that, on average, energy overflow due to the fact that the power-to-depletion parameter values considered in the paper, the optimal BP consumption and harvesting are balanced. For the broad range solely adapt to the EH state, such that, on average, energy overflow due to the fact that the power-to-depletion parameter values considered in the paper, the optimal BP consumption and harvesting are balanced. For the broad range solely adapt to the EH state, such that, on average, energy overflow due to the fact that the power-to-depletion parameter values considered in the paper, the optimal BP consumption and harvesting are balanced. For the broad range solely adapt to the EH state, such that, on average, energy overflow due to the fact that the power-to-depletion parameter values considered in the paper, the optimal BP consumption and harvesting are balanced. For the broad range solely adapt to the EH state, such that, on average, energy overflow due to the fact that the power-to-depletion parameter values considered in the paper, the optimal BP consumption and harvesting are balanced. For the broad range solely adapt to the EH state, such that, on average, energy overflow due to the fact that the power-to-depletion parameter values considered in the paper, the optimal BP consumption and harvesting are balanced. For the broad range solely adapt to the EH state, such that, on average, energy overflow due to the fact that the power-to-depletion parameter values considered in the paper, the optimal BP consumption and harvesting are balanced. For the broad range solely adapt to the EH state, such that, on average, energy overflow due to the fact that the power-to-depletion parameter values considered in the paper, the optimal BP consumption and harvesting are balanced. For the broad range solely adapt to the EH state, such that, on average, energy overflow due to the fact that the power-to-depletion parameter values considered in the paper, the optimal BP consumption and harvesting are balanced. For the broad range solely adapt to the EH state, such that, on average, energy overflow due to the fact that the power-to-depletion parameter values considered in the paper, the optimal BP consumption and harvesting are balanced. For the broad range solely adapt to the EH state, such that, on average, energy overflow due to the fact that the power-to-depletion parameter values considered in the paper, the optimal BP consumption and harvesting are balanced. For the broad range solely adapt to the EH state, such that, on average, energy overflow due to the fact that the power-to-depletion parameter values considered in the paper, the optimal BP consumption and harvesting are balanced. For the broad range solely adapt to the EH state, such that, on average, energy overflow due to the fact that the power-to-depletion parameter values considered in the paper, the optimal BP consumption and harvesting are balanced. For the broad range solely adapt to the EH state, such that, on average, energy overflow due to the fact that the power-to-depletion parameter values considered in the paper, the optimal BP consumption and harvesting are balanced. For the broad range solely adapt to the EH state, such that, on average, energy overflow due to the fact that the power-to-depletion parameter values considered in the paper, the optimal BP consumption and harvesting are balanced. For the broad range solely adapt to the EH state, such that, on average, energy overflow due to the fact that the power-to-depletion parameter values considered in the paper, the optimal BP consumption and harvesting are balanced. For the broad range solely adapt to the EH state, such that, on average, energy overflow due to the fact that the power-to-depletion parameter values considered in the paper, the optimal BP consumption and harvesting are balanced. For the broad range solely adapt to the EH state, such that, on average, energy overflow due to the fact that the power-to-depletion parameter values considered in the paper, the optimal BP consumption and harvesting are balanced. For the broad range solely adapt to the EH state, such that, on average, energy overflow due to the fact that the power-to-depletion parameter values considered in the paper, the optimal BP consumption and harvesting are balanced. For the broad range solely adapt to the EH state, such that, on average, energy overflow due to the fact that the power-to-depletion parameter values considered in the paper, the optimal BP consumption and harvesting are balanced. For the broad range solely adapt to the EH state, such that, on average, energy overflow due to the fact that the power-to-depletion parameter values considered in the paper, the optimal BP consumption and harvesting are balanced. For the broad range solely adapt to the EH state, such that, on average, energy overflow due to the fact that the power-to-depletion parameter values considered in the paper, the optimal BP consumption and harvesting are balanced. For the broad range solely adapt to the EH state, such that, on average, energy overflow due to the fact that the power-to-depletion parameter values considered in the paper, the optimal BP consumption and harvesting are balanced. For the broad range solely adapt to the EH state, such that, on average, energy overflow due to the fact that the power-to-depletion parameter values considered in the paper, the optimal BP consumption and harvesting are balanced. For the broad range solely adapt to the EH state, such that, on average, energy overflow due to the fact that the power-to-depletion parameter values considered in the paper, the optimal BP consumption and harvesting are balanced. For the broad range solely adapt to the EH state, such that, on average, energy overflow due to the fact that the power-to-depletion parameter values considered in the paper, the optimal BP consumption and harvesting are balanced. For the broad range solely adapt to the EH state, such that, on average, energy overflow due to the fact that the power-to-depletion parameter values considered in the paper, the optimal BP consumption and harvesting are balanced. For the broad range solely adapt to the EH state, such that, on average, energy overflow due to the fact that the power-to-depletion parameter values considered in the paper, the optimal BP consumption and harvesting are balanced. For the broad range solely adapt to the EH state, such that, on average, energy overflow due to the fact that the power-to-depletion parameter values considered in the paper, the optimal BP consumption and harvesting are balanced. For the broad range solely adapt to the EH state, such that, on average, energy overflow due to the fact that the power-to-depletion parameter values considered in the paper, the optimal BP consumption and harvesting are balanced. For the broad range solely adapt to the EH state, such that, on average, energy overflow due to the fact that the power-to-depletion parameter values considered in the paper, the optimal BP consumption and harvesting are balanced. For the broad range solely adapt to the EH state, such that, on average, energy overflow due to the fact that the power-to-depletion parameter values considered in the paper, the optimal BP consumption and harvesting are balanced. For the broad range solely adapt to the EH state, such that, on average, energy overflow due to the fact that the power-to-depletion parameter values considered in the paper, the optimal BP consumption and harvesting are balanced. For the broad range solely adapt to the EH state, such that, on average, energy overflow due to the fact that the power-to-depletion parameter values considered in the paper, the optimal BP consumption and harvesting are balanced. For the broad range solely adapt to the EH state, such that, on average, energy overflow due to the fact that the power-to-depletion parameter values considered in the paper, the optimal BP consumption and harvesting are balanced. For the broad range solely adapt to the EH state, such that, on average, energy overflow due to the fact that the power-to-depletion parameter values considered in the paper, the optimal BP consumption and harvesting are balanced. For the broad range solely adapt to the EH state, such that, on average, energy overflow due to the fact that the power-to-depletion parameter values considered in the paper, the optimal BP consumption and harvesting are balanced. For the broad range solely adapt to the EH state, such that, on average, energy overflow due to the fact that the power-to-depletion parameter values considered in the paper, the optimal BP consumption and harvesting are balanced. For the broad range solely adapt to the EH state, such that, on average, energy overflow due to the fact that the power-to-depletion parameter values considered in the paper, the optimal BP consumption and harvesting are balanced. For the broad range solely adapt to the EH state, such that, on average, energy overflow due to the fact that the power-to-depletion parameter values considered in the paper, the optimal BP consumption and harvesting are balanced. For the broad range solely adapt to the EH state, such that, on average, energy overflow due to the fact that the power-to-depletion parameter values considered in the paper, the optimal BP consumption and harvesting are balanced. For the broad range solely adapt to the EH state, such that, on average, energy overflow due to the fact that the power-to-depletion parameter values considered in the paper, the optimal BP consumption and harvesting are balanced. For the broad range solely adapt to the EH state, such that, on average, energy overflow due to the fact that the power-to-depletion parameter values considered in the paper, the optimal BP consumption and harvesting are balanced. For the broad range solely adapt to the EH state, such that, on average, energy overflow due to the fact that the power-to-depletion parameter values consid...
\[ \Pr_{\tilde{\mu}}(E_k = e, A_{k-1} = a|S_0) \] depends on \( \tilde{\mu} \) only through the expectation \( \eta(e, a) \triangleq \mathbb{E}_V[\tilde{\mu}(e, a, V)] \), which is common to all \( \tilde{\mu} \in \mathcal{R}_\mu \), and therefore the steady state distribution of \( \{E_k, A_{k-1}\} \) is the same for all \( \tilde{\mu} \in \mathcal{R}_\mu \), i.e., \( \pi_{\tilde{\mu}}(e, a; S_0) = \pi_{\mu}(e, a; S_0) \). Therefore, from (40) and (41) we obtain

\[
G_{\mu}(S_0) \leq \max_{\tilde{\mu} \in \mathcal{R}_\mu} G_{\tilde{\mu}}(S_0) = \sum_{(e,a) \in S} \pi_{\tilde{\mu}}(e, a; S_0)\mathbb{E}_V[\mu^*(e, a, V)],
\]

where, for each \((e, a) \in S\), \(\mu^*(e, a, \cdot)\) is defined as

\[
\mu^*(e, a, \cdot) = \arg \max_{\tilde{\mu}(e, a, \cdot) : \mathbb{R}^+ \rightarrow [0, 1]} \mathbb{E}_V[\tilde{\mu}(e, a, V)],
\]

s.t. \(\mathbb{E}_V[\tilde{\mu}(e, a, V)] = \eta(e, a)\). (43)

Since (43) is a convex optimization problem, using the Lagrangian method [33], with Lagrangian multiplier \(\psi_{\tilde{\mu}}(e, a)\) associated to the constraint \(\mathbb{E}_V[\tilde{\mu}(e, a, V)] = \eta(e, a)\), we obtain

\[
\mu^*(e, a, \cdot) = \arg \max_{\tilde{\mu}(e, a, \cdot) : \mathbb{R}^+ \rightarrow [0, 1]} \mathbb{E}_V[\tilde{\mu}(e, a, V) - v_{th}(e, a)],
\]

yielding (7), where \(v_{th}(e, a)\) is such that \(\mathbb{E}_V[\mu^*(e, a, V)] = \eta(e, a)\). The threshold structure in (7) is thus proved. \(\blacksquare\)

**APPENDIX B**

**PROOF OF THEOREM 1**

Let us consider a BP \(\eta\) with parameters \((\eta_G, \theta)\). We have the following steady-state equation for state \((e_{\text{max}}, B)\) under the BP:

\[
\pi_{\eta}(e_{\text{max}}, B) = \pi_{\eta}(e_{\text{max}}, B) p_{\text{BP}} \eta(e_{\text{max}}, B) + \pi_{\eta}(e_{\text{max}}, G) \tilde{\beta} \eta \eta_G e_{\text{max}, G} + \pi_{\eta}(e_{\text{max}}, G) \tilde{\beta} \eta \eta_G e_{\text{max}, G},
\]

where the first equality follows from the fact that, if the system is in state \((E_k, A_{k-1}) = (e_{\text{max}}, B)\) at time \(k\), no energy can be harvested in time-slot \(k - 1\), hence state \((e_{\text{max}}, B)\) can be reached only from states \((e_{\text{max}}, B)\) or \((e_{\text{max}}, G)\), provided that the EHS does not transmit and the proper transition occurs in the harvesting process. The second equality follows from the definition of BP in (12). Then, using the fact that \(p_{\text{BP}} = 1 - D_B^{-1}\) and \(p_G = 1 - \gamma^{-1} D_B^{-1}\), after algebraic manipulation we obtain

\[
\begin{bmatrix}
\pi_{\eta}(e_{\text{max}}, G) \\
\pi_{\eta}(e_{\text{max}}, B)
\end{bmatrix} = (\theta \eta_G + \tilde{\theta}) t(\theta) \pi_{\eta}(e_{\text{max}}, G),
\]

where \(t(\theta)\) is given by (17).\(^5\)

In the long-term, the frequency of transitions from energy level \(e - 1\) to \(e\) and from \(e\) to \(e - 1\) must be the same. Therefore, for \(e \in \{1, \ldots, e_{\text{max}}\}\) we have

\[
\pi_{\eta}(e - 1, G) p_G \lambda_G \tilde{\eta}(e - 1, G) + \pi_{\eta}(e - 1, B) \tilde{\beta} \lambda_B \tilde{\eta}(e - 1, B) = \pi_{\eta}(e, G) \left( p_G \lambda_G + \tilde{\beta} \lambda_B \right) \eta(e, G) + \pi_{\eta}(e, B) \left( p_B \lambda_B + \tilde{\beta} \lambda_B \right) \eta(e, B),
\]

where the left hand expression is the probability of moving from \(e - 1\) to \(e\), and is computed by noting that the energy level can increase by one unit only if the EH state moves to state \(G\), one energy quantum is harvested and the EHS does not transmit. Similarly, the right hand expression is the probability of moving from \(e\) to \(e - 1\), and is computed by noting that the energy level decreases by one unit when no energy is harvested (because of either a transition to EH state \(B\), or a transition to state \(G\) and no energy harvested) and the EHS transmits.

Moreover, we have the following steady-state equation for state \((e - 1, B)\), for \(e \in \{1, \ldots, e_{\text{max}}\}\):

\[
\pi_{\eta}(e - 1, B) = \pi_{\eta}(e - 1, B) p_{\text{BP}} \eta(e - 1, B) + \pi_{\eta}(e, B) p_{\text{BP}} \eta(e, B) + \pi_{\eta}(e - 1, G) \tilde{\beta} \eta(e - 1, G) + \pi_{\eta}(e, G) \tilde{\beta} \eta(e, G),
\]

where the first two terms account for transitions from EH state \(B\) to \(B\), i.e., moving to state \((e - 1, B)\) from either \((e, B)\) (when the EHS transmits) or \((e - 1, B)\) (when the EHS remains idle), and the latter two terms have a similar interpretation, and correspond to transitions from EH state \(G\) to \(B\). (Note that on all transitions into state \((e - 1, B)\) no harvesting can occur since \(A_{k-1} = B\).)

Solving the system of equations (46), (47) with respect to \(\pi_{\eta}(e - 1, G), \pi_{\eta}(e - 1, B)\), for \(e \in \{1, \ldots, e_{\text{max}}\}\) we obtain

\[
\begin{bmatrix}
\pi_{\eta}(e - 1, G) \\
\pi_{\eta}(e - 1, B)
\end{bmatrix} = K(e - 1) \begin{bmatrix}
\pi_{\eta}(e, G) \\
\pi_{\eta}(e, B)
\end{bmatrix},
\]

where we have defined the \(2 \times 2\) matrix

\[
K(e - 1) = \begin{bmatrix}
K_{GG}(e - 1) & K_{GB}(e - 1) \\
K_{BG}(e - 1) & K_{BB}(e - 1)
\end{bmatrix},
\]

with components

\[
K_{GG}(e - 1) = \frac{(D_B - 1) \lambda_G + \gamma^{-1} \lambda_G \eta(e - 1, B) + \lambda_B}{\eta(e - 1, G) W/\eta(e, G)},
\]

\[
K_{GB}(e - 1) = \frac{\gamma^{-1} \eta(e, G)}{W},
\]

\[
K_{BG}(e - 1) = \frac{\eta(e, G) [\gamma^{-1} \lambda_B + (D_B - 1) \lambda_G]}{W},
\]

and

\[
W = \lambda_G[D_B \eta(e - 1, B) + \tilde{\eta}(e - 1, B) - \eta(e - 1, B) \gamma^{-1}].
\]

In particular, substituting the expression of the BP in Def. 4, we have \(K(0) = Z\), where \(Z\) is defined in (18), \(K(e) = J, \forall e \in \{1, 2, \ldots, e_{\text{max}} - 2\}\), where \(J\) is defined in (19), and \(K(e_{\text{max}} - 1) = \theta J\) \(\begin{bmatrix}
\frac{1}{\eta_G} & 0 \\
0 & \frac{1}{\eta_B}
\end{bmatrix} + \tilde{\theta} J\). Using (48), (45), and the fact that \(K(e_{\text{max}} - 1) t(\theta) = J t(\theta) (\tilde{\theta} + \eta_G)\), we then obtain, for \(e \in \{1, \ldots, e_{\text{max}} - 1\}\)

\[
\begin{bmatrix}
\pi_{\eta}(e, G) \\
\pi_{\eta}(e, B)
\end{bmatrix} = J^{e_{\text{max}} - e} t(\theta) \pi_{\eta}(e_{\text{max}}, G),
\]

and \(\pi_{\eta}(0, G); \pi_{\eta}(0, B))^T = Z [\pi_{\eta}(1, G); \pi_{\eta}(1, B))^T\), yielding (15). \(\pi_{\eta}(e_{\text{max}}, G)\) is finally obtained by imposing the normalization \(\sum_{e = 0}^{e_{\text{max}}} [\pi_{\eta}(e, G) + \pi_{\eta}(e, B)] = 1\), yielding (16).
The average reward under the BP directly follows by substituting the expressions of the BP and of the steady state distribution in (10), and using the fact that, by marginalization over the battery state, \( \sum_{e=0}^{e_{\text{max}}} \pi_{\eta}(e, a) = \pi_{A}(a), \forall a \in \{G, B\} \), and, when overflow avoidance is employed (\( \theta = 1 \)), \( \pi_{\eta}(e_{\text{max}}, B) = 0 \) from (44).

**APPENDIX C**

**PROOF OF THEOREM 2**

*Proof of (23):* In this proof, the notation \( f(D_B) = h(D_B) + \sigma(D_B^2) \) is equivalent to \( \limsup_{D_B \to \infty} \frac{f(D_B) - h(D_B)}{D_B^2} < \infty \). If \( f(D_B) \) and \( h(D_B) \) are matrices of the same size, this definition applies to each component. Moreover, we vary \( D_B \), \( D_G \) and \( e_{\text{max}} \), while keeping \( \gamma = D_G/D_B \), \( \rho = e_{\text{max}}/D_B \) and all the other system parameters fixed.

It can be shown that the eigenvalue decomposition of matrix \( J \) in (19) is given by \( J = V \Sigma B^{-1} \), where, letting \( \nu = \frac{\lambda_0 \lambda_G}{\lambda_0 \gamma G} \left( 1 - \frac{1}{\nu + \lambda_0 (1 - \gamma)} \right) \),

\[
V = \begin{bmatrix} \gamma & 1 \\ 1 & 0 \end{bmatrix} + D_B^{-1} \begin{bmatrix} 0 & -\frac{\lambda_0 \lambda_G}{\nu + \lambda_0 (1 - \gamma)} \\ 0 & \frac{\lambda_G}{\nu + \lambda_0 (1 - \gamma)} \end{bmatrix}, \quad D = \begin{bmatrix} 1 & 0 \\ 0 & \nu \end{bmatrix}. \tag{52}
\]

Since \( V \Sigma V^{-1} \) is the \( 2 \times 2 \) identity matrix, it can be shown that

\[
V^{-1} = D_B^{-1} \begin{bmatrix} \gamma & 1 \\ 1 & 0 \end{bmatrix} + \frac{\lambda_G \lambda_0}{\nu + \lambda_0 (1 - \gamma)} \begin{bmatrix} 0 & 1 \\ 1 & -\gamma \end{bmatrix} + \sigma(D_B^2). \tag{53}
\]

Moreover, from (17) and (18) we have

\[
t(\theta) = \begin{bmatrix} \theta & 1 \\ 1 & 0 \end{bmatrix} + D_B^{-1} \theta \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} + \sigma(D_B^2), \tag{54}
\]

\[
Z = D_B \eta_B \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix} + \sigma(1). \tag{55}
\]

Now, letting \( e_{\text{max}} = \rho D_B \), where \( \rho \) is fixed, we have

\[
Z \Sigma V^{-1} t(\theta) \overset{(a)}{=} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} V^{-1} t(\theta) \overset{(b)}{=} Z V \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} V^{-1} t(\theta) + \sigma(D_B^{-2}), \tag{56}
\]

where (a) follows from the eigenvalue decomposition of \( J \), and (b) follows from the fact that \( \liminf_{D_B \to \infty} \nu \in (0, 1) \), hence \( \nu^{e_{\text{max}}-1} = \nu^{e_{\text{max}}-1} \) decays exponentially fast to zero for \( D_B \to \infty \) so that \( \nu^{e_{\text{max}}-1} = \sigma(D_B^2) \). From (53) and (54), and since \( \eta_G + 1 - \gamma \lambda_B = 1 \), we have that

\[
V^{-1} t(\theta) = D_B^{-1} \frac{\eta_G}{\eta_B} (\theta + \theta \lambda_G) + \sigma(D_B^{-2}). \tag{57}
\]

Moreover, from (52) and (55), we obtain

\[
Z V \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} = D_B \eta_B \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix} + \sigma(1). \tag{58}
\]

Then, substituting (57) and (58) in (56), we obtain

\[
Z \Sigma V^{-1} t(\theta) = \frac{\eta_G}{\eta_B} (\theta + \theta \lambda_G) \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} + \sigma(D_B^{-1}). \tag{59}
\]

Using a similar approach, it can be proved that

\[
e_{\text{max}}^{-1} \begin{bmatrix} 1 & 1 \end{bmatrix} J \Sigma V^{-1} t(\theta) = \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 - e_{\text{max}}^{-1} \nu \end{bmatrix} V^{-1} t(\theta)
\]

\[
= \rho(\gamma + 1) \frac{\eta_G}{\eta_B} (\theta + \theta \lambda_G) + \gamma \frac{\eta_G}{\eta_B} \frac{\lambda_G}{\eta_B} (\theta + \theta \lambda_G) + \sigma(D_B^{-1}). \tag{60}
\]

Then, substituting (59) and (60) in (16), we obtain

\[
\pi_{\eta}(e_{\text{max}}, G) = \pi_{A}(G) \eta_B \frac{\lambda_G - \eta_G \theta + \theta \lambda_G}{\lambda_G \gamma G} + \sigma(D_B^{-1}), \tag{61}
\]

where we have used the fact that \( (1 + \sigma(D_B^{-1}))^{-1} = 1 + \sigma(D_B^{-1}) \). Then, from (15) and (59), we obtain

\[
\pi_{\eta}(0, G) = \sigma(D_B^{-1}), \tag{62}
\]

\[
\pi_{\eta}(0, B) = \frac{\eta_G}{\eta_B} (\theta + \theta \lambda_G) \pi_{\eta}(e_{\text{max}}, G) + \sigma(D_B^{-1}) = \pi_{A}(B) \frac{\eta_B}{\rho + \eta_B} + \sigma(D_B^{-1}). \tag{63}
\]

The asymptotic expression (23) is finally obtained by substituting (61), (62) and (63) in (14), and letting \( D_B \to \infty \), with \( \lim_{D_B \to \infty} \sigma(D_B^{-1}) = 0 \). (Note that (61) is only used in (14) when \( \theta = 1 \).)

*Proof of property 1):* The inequality \( G^{(\infty)}(\eta_G, 1; \rho) > G^{(\infty)}(\eta_G, 0; \rho) \) is proved by direct inspection of (23), since \( g(1) > g(0) \) and \( \lambda_G > \eta_G \).

*Proof of property 2):* We want to prove that \( G^{(\infty)}(\eta_G, \theta; \rho) \) is an increasing function of \( \rho \). Equivalently, \( \frac{d G^{(\infty)}(\eta_G, \theta, \rho)}{d \rho} > 0 \).

The derivative of (23) with respect to \( \rho \) can be written as

\[
d G^{(\infty)}(\eta_G, \theta; \rho) \frac{d \rho}{d \rho} = \pi_{A}(B) \frac{\eta_B}{\rho + \eta_B} \left[ g(\eta_B) - \eta_B g(1) \right]
\]

\[
+ \theta \pi_{A}(B) \frac{\eta_B^2}{(\rho + \eta_B)^2} g(1)
\]

\[
+ \theta \pi_{A}(A) \eta_B \frac{\lambda_G - \eta_G}{1 - \eta_G} \frac{1}{\rho + \eta_B} \left[ g(\eta_G) - \eta_G g(1) \right] > 0,
\]

where we have used the fact that, since \( g(x) \) is a concave function of \( x \) and \( g(0) = 0 \) (Lemma 2), \( g(x) > x g(1) \). The limits for \( \rho \to \infty \) and \( \rho \to 0 \) are finally obtained by computing the corresponding limits in (23).

**APPENDIX D**

**PROOF OF LEMMA 4**

*Proof of (26) and property 1)* We first prove that the optimal BP \( (\eta_G(\theta; \rho), \theta) \) uniquely solves (26). After algebraic manipulations, we find that the derivative of \( G^{(\infty)}(\eta_G, \theta; \rho) \) in (23) with respect to \( \eta_G \) is positive if and only if \( L(\eta_G, \theta; \rho) > 0 \), where \( L(\eta_G, \theta; \rho) \) is given in (27). Moreover,

\[
d L(\eta_G, \theta; \rho) \frac{d \eta_G}{d \eta_G} \rho g''(\eta_G) + g''(\eta_G) \frac{\eta_G}{\eta_G} (\theta \lambda_G + \theta \eta) + \gamma \rho g''(\eta_B) - 2 \theta \lambda_G \gamma g''(\eta_G) - g(1) < 0, \tag{64}
\]

where \( g^{(\infty)}(x) \) denotes proportionality up to a positive multiplicative factor, and the inequality holds since \( g''(\eta_G) > 0, g''(x) < 0 \) and \( g(\eta_G) + \eta_G g'(\eta_G) - g(1) > 0 \), from the concavity of \( g(x) \). Therefore, \( L(\eta_G, \theta; \rho) \) is a decreasing function of \( \eta_G \).
Moreover, direct substitution in (27) and rearrangement of terms yields
\[ L(\lambda_G, \theta; \rho) = g'(\lambda_G) - g'(0) < 0, \]
\[ L(\beta, \theta; \rho) = \frac{\beta g(\beta)}{\rho} + \theta \frac{\beta}{\rho} \left( 1 + \frac{\beta}{\rho} \right) g'(\beta) + \theta g(\beta) - \beta g'(\beta) \]
\[ + \theta \frac{\beta}{\rho^2} \left( \frac{\beta}{\rho} + 1 + \frac{\beta}{\rho} + \frac{\beta}{\rho^2} \right) g(\beta) > 0, \]
for \( \eta_G = \lambda_G, \eta_B = 0 \) and for \( \eta_G = \eta_B = \beta, \) respectively.

We conclude that there exists a unique \( \eta_G \in (\beta, \lambda_G) \) that maximizes \( G^{(\infty)}(\eta_G, \theta; \rho), \) as obtained as the unique solution of \( L(\eta_G, \theta; \rho) = 0. \)

**Proof of property 2)** Since \( L(\eta_G; 1; \rho) < L(\eta_G; 0; \rho), \) \( \forall \eta_G, \) it follows that, under the optimal BP,
\[ L(\eta_G^*(0; \rho); 0; \rho) = 0 = L(\eta_G^*(1; \rho); 1; \rho) < L(\eta_G^*(1; \rho); 0; \rho). \]

Since \( L(\eta_G; 0; \rho) \) is a decreasing function of \( \rho, \) it follows that \( \eta_G^*(1; \rho) < \eta_G^*(0; \rho). \)

**Proof of property 3)**: We now prove that \( \eta_G^*(\theta; \rho) \) is a decreasing function of \( \rho, \) for \( \rho \geq \beta, \) i.e., \( \frac{d\eta_G^*(\theta; \rho)}{d\rho} < 0. \)

For the optimal BP \( \eta_G^*(\theta; \rho), \) we have \( L(\eta_G^*(\theta; \rho); \theta; \rho) = 0, \)
\[ \frac{dL(\eta_G^*(\theta; \rho); \theta; \rho)}{d\rho} = 0, \]
where we have defined \( f(\eta_G)' = f(\eta_G^*(\theta; \rho)). \) Then, since \( \frac{d\eta_G^*(\theta; \rho)}{d\rho} < 0 \) from (64), \( \frac{dL(\eta_G^*(\theta; \rho); \theta; \rho)}{d\rho} 

Finally, from (27) and since \( \eta_G^* = \eta_G^*(\theta; \rho) \in (\beta, \lambda_G) \) and \( \eta_B = \gamma(\lambda_G - \eta_G^*) = 0, \beta, \lambda_G, \) we have
\[ (\rho + \beta)^3 \frac{d}{d\rho} \left[ L(\eta_G, \theta; \rho) \cdot \frac{\rho^2}{\rho + \eta_B g(\eta_B)} \right] = -\frac{\beta}{\rho} \cdot \left( g(\eta_B) - \eta_B g'(\eta_B) \right) (\rho - \eta_B) \]
\[ - \frac{2\theta}{\rho} \left( g'(\eta_B) - g'(\eta_B) \right) g(\eta_B) + \eta_B g'(\eta_B) \]
\[ \times \left[ \frac{\eta_B}{\eta_G} (\rho - \eta_B) + \frac{\lambda_G}{\rho} (\rho - \eta_B) \right] \leq 0, \]
where the inequality holds for \( \rho \geq \beta \) (which implies \( \rho \geq \eta_B \)), since \( \eta_B \leq \beta, \) and we have used the fact that \( g(x) \) is a concave increasing function of \( x \) with \( g(0) = 0 (Lemma 2), \) hence \( g(\eta_B) - \eta_B g'(\eta_B) > 0, g'(\eta_B) > g'(\eta_B) \) (since \( \eta_B < \eta_G^* \)) and \( g(\eta_B) + \eta_G^* g'(\eta_B) - g(1) > 0. \)

Equivalently, \( \frac{dL(\eta_G^*(\theta; \rho); \theta; \rho)}{d\rho} < 0 \) and \( \frac{d\eta_G^*(\theta; \rho)}{d\rho} < 0. \)

In the limit \( \rho \to \infty, \) we have \( \lim_{\rho \to \infty} L(\eta_G, \theta; \rho) = g'(\eta_G) - g'(\eta_B), \) which is equal to zero if and only if \( \eta_G^*(\theta; \infty) = \beta, \) proving (28). For \( \rho \to 0, \) we have
\[ \lim_{\rho \to 0} g'(\eta_G) = \frac{2}{\rho^2} \frac{dL(\eta_G, \theta; \rho)}{d\rho} \]
\[ = \frac{\beta}{\rho} g'(\eta_G) + \frac{\beta}{\rho} \lambda_G (g(\eta_G) + \eta_G^* g'(\eta_G) - g(1)) > 0. \]

Hence, for \( \rho \to 0, \) the asymptotic reward is a strictly increasing function of \( \eta_G, \) for \( \eta_G \in (\beta, \lambda_G), \) and is therefore maximized by \( \eta_G = \lambda_G, \) proving (29).

**APPENDIX E**

**PROOF OF Lemma 5**

From the concavity of \( g(x) \) and Jensen’s inequality [33], (33) implies that
\[ G_{CDM}(\eta; \theta; \rho) = D_{G} g \left( \frac{1}{D_{G}} \int_{0}^{D_{G}} \eta(\theta; \tau) d\tau \right) + D_{B} g \left( \frac{1}{D_{B}} \int_{0}^{D_{G}+D_{B}} \eta(\theta; \tau) d\tau \right). \]

Therefore, letting \( \eta_G = \frac{1}{D_{G}} \int_{0}^{D_{G}} \eta(\theta; \tau) d\tau \) and \( \eta_B = \frac{1}{D_{B}} \int_{0}^{D_{G}+D_{B}} \eta(\theta; \tau) d\tau \), the upper bound above is attained with equality if the energy drawing rates are constant over the GOOD and BAD periods, i.e., \( \eta(\theta; \tau) = \eta_G, \forall \tau \in T_{G} \) and \( \eta(\theta; \tau) = \eta_B, \forall \tau \in T_{B} \) (note that \( \eta_G, \eta_B \in [0, 1], \) since \( \eta(\theta; \tau) \in [0, 1], \)) so that it is sufficient to consider policies with such structure in order to achieve optimality. Substituting in (33), we obtain
\[ G_{CDM}(\eta; \theta; \rho) = \pi_{A}(G) g(\eta_B) + \pi_{A}(B) g(\eta_B). \]

Moreover, from (30), (31) and (32), \( \eta_B \) and \( \eta_G \) are related to \( e_L \) and \( e_H \) by
\[ e_H = \min \left\{ e_L + D_{G}(\lambda_G - \eta_G), e_{\text{max}} \right\}, \]
\[ e_L = \max \left\{ e_H - D_{B} \eta_B, 0 \right\}. \]

Note that a policy such that \( e_L + D_{G}(\lambda_G - \eta_G) > e_{\text{max}} \) incurs energy overflow, hence it is strictly sub-optimal. This can be shown by defining an improved policy \( \tilde{\eta}(\theta; \rho) \), with \( \tilde{\eta}(\theta; \rho) = \eta_G, \forall \theta \in T_{G} \), where \( \eta_G \) is the unique solution of \( e_L + D_{G}(\lambda_G - \eta_G) = e_{\text{max}}. \) Under the new policy, we have
\[ G_{CDM}(\eta(\theta; \rho)) = \pi_{A}(G) g(\eta_B) + \pi_{A}(B) g(\eta_B) > \pi_{A}(G) g(\eta_G) + \pi_{A}(B) g(\eta_B) = G_{CDM}(\eta(\theta; \rho)). \]

We thus only consider \( \eta_G \) such that \( e_L + D_{G}(\lambda_G - \eta_G) \leq e_{\text{max}}. \)

From (67), letting \( \Delta = e_H - e_L, \eta_G \) and \( \eta_B \) are then given by
\[ \eta_G = \lambda_G - \frac{\Delta}{D_{G}}, \eta_B = \frac{\Delta}{D_{B}}. \]

Note that \( \Delta \in [0, \min \left\{ \lambda_G D_{G}, D_{B}, e_{\text{max}} \right\}], \) since, during the GOOD EH period, the battery cannot be recharged by more than \( \lambda_G D_{G} \), and, during the BAD EH period, it cannot be discharged by more than \( D_{B}. \)

Substituting (68) in (66), we obtain
\[ G_{CDM}(\eta(\theta; \rho)) = \pi_{A}(G) g(\lambda_G - \frac{\Delta}{D_{G}}) + \pi_{A}(B) g \left( \frac{\Delta}{D_{B}} \right) \]

We now maximize the right hand side with respect to \( \Delta \in [0, \min \left\{ \lambda_G D_{G}, D_{B}, e_{\text{max}} \right\}] \). We have
\[ \frac{d}{d\Delta} G_{CDM}(\eta(\theta; \rho)) = -\pi_{A}(G) \frac{1}{D_{G}} g'(\lambda_G - \frac{\Delta}{D_{G}}) \]
\[ + \pi_{A}(B) \frac{1}{D_{B}} g' \left( \frac{\Delta}{D_{B}} \right). \]
so that $\frac{\partial}{\partial \Delta} G_{\text{CDM}}(\eta^{(\text{CDM})}) > 0$ if and only if $\Delta < \beta D_B$ (we have used the fact that $\beta = \lambda G/(1 + \gamma)$). Therefore, $G_{\text{CDM}}(\eta^{(\text{CDM})})$ is maximized by

$$
\Delta^* = \min \{ \beta D_B, \lambda G D_B, D_B, \epsilon_{\max} \} = D_B \min \{ \beta, \lambda G; 1, \rho \} = D_B \min \{ \beta, \rho \},
$$

(71)

where the last equality follows from the fact that $\beta < \min \{ \lambda G; 1 \}$. The optimal energy drawing rates $\eta G$ and $\eta B$ are obtained by substituting (71) in (68) and (69), thus proving (35) and (36).

**APPENDIX F**

**PROOF OF LEMMA 6**

We consider a BP for DSM with parameters $(\eta G, \theta)$. Without loss of generality, we assume that $\eta G \in (\beta, \lambda G)$, i.e., $\eta B \in (0, \beta)$, since the optimal BP, which maximizes the asymptotic average reward (23), must satisfy this condition (Property 1 of Lemma 4). If $\rho \geq \beta$, then $G_{\text{CDM}}(\eta^{(\text{CDM})}) = \pi A(G)g(\lambda G - \frac{\rho}{\gamma}) + \pi A(B)g(\beta) = g(\beta)$, since $\lambda G - \frac{\rho}{\gamma} = \beta$.

From the concavity of $g(x)$ and from the fact that for a BP $\beta$ is the average energy expenditure per time-slot, application of Jensen’s inequality shows that the reward is upper bounded by $g(\beta)$, thereby proving (37) for $\rho \geq \beta$.

If $\rho < \beta$, we have that $G_{\text{CDM}}(\eta^{(\text{CDM})}) = \pi A(G)g(\lambda G - \frac{\rho}{\gamma}) + \pi A(B)g(\rho)$. Define

$$
Z(\eta G) = \begin{cases} 
\pi A(G)g(\eta G) + \pi A(B)g(\eta B) & \text{if } \eta B \leq \rho, \\
\pi A(G) \frac{\rho}{\eta G} g(\eta G) + \pi A(G) \frac{\eta G - \rho}{\eta G} g(\lambda G) + \pi A(B) \frac{\rho}{\eta B} g(\eta B) & \text{if } \eta B > \rho.
\end{cases}
$$

(72)

We have that $Z(\eta G) \leq G_{\text{CDM}}(\eta^{(\text{CDM})}) = Z(\lambda G - \gamma^{-1} \rho)$. This can be proved, for $\eta G \geq \lambda G - \gamma^{-1} \rho$ (i.e., $\eta B \leq \rho$), by observing that in this regime $Z(\eta G)$ is a decreasing function of $\eta G$. On the other hand, for $\eta G < \lambda G - \gamma^{-1} \rho$ (i.e., $\eta B > \rho$), by applying Jensen’s inequality to the first two terms in the expression of $Z(\eta G)$, we obtain

$$
Z(\eta G) = \pi A(G) \frac{\rho}{\eta B} g(\eta G) + \pi A(B) \frac{\rho}{\eta B} g(\lambda G) + \pi A(B) \frac{\rho}{\eta B} g(\eta B)
\leq \pi A(G) g(\lambda G - \gamma^{-1} \rho) + \pi A(B) \frac{\rho}{\eta B} g(\eta B)
\leq G_{\text{CDM}}(\eta^{(\text{CDM})}),
$$

(73)

where in the first step we have used the fact that $\lambda G - \gamma^{-1} \eta G = \gamma^{-1} \eta B$, and in the last step we have used the fact that $g(\eta B)/\eta B$ is a decreasing function of $\eta B$, so that $\rho g(\eta B)/\eta B \leq g(\rho)$ for $\eta B > \rho$. From property 1 of Theorem 2, and from the fact that $\theta \in \{0, 1\}$, we have that $G^{(\infty)}(\eta G, \theta; \rho) \leq G^{(\infty)}(\eta G, 1; \rho)$ which, combined with (73), yields

$$
G^{(\infty)}(\eta G, \theta; \rho) - G_{\text{CDM}}(\eta^{(\text{CDM})}) \leq G^{(\infty)}(\eta G, 1; \rho) - Z(\eta G).
$$

(74)

We finally prove that the right hand side of (74) is negative. In fact, if $\eta G \geq \lambda G - \frac{\rho}{\gamma}$, from (72) and Lemma 2 we obtain

$$
G^{(\infty)}(\eta G, 1; \rho) - Z(\eta G) = -\frac{\eta B}{\rho + \eta B} g(\eta B) g(\eta G) - \frac{\eta B}{\eta G} (g(\eta G) - g(\eta G))
\times \left[ (g(\eta B) - \eta B g(1)) + \frac{\eta B}{\eta G} (g(\eta G) - \eta B g(1)) \right] < 0.
$$

On the other hand, if $\eta G < \lambda G - \frac{\rho}{\gamma}$, we obtain

$$
G^{(\infty)}(\eta G, 1; \rho) - Z(\eta G) = -\frac{\rho^2}{\rho + \eta B} g(\eta B) - \frac{\eta B}{\eta G} g(\eta G)
- \frac{\rho^2}{\rho + \eta B} \frac{\lambda G}{\eta B} + \frac{\eta B}{\eta G} g(\eta G) - \frac{\rho}{\rho + \eta B} + \frac{\eta B}{\eta G} g(\eta G)
\times \left[ (g(\eta B) - \eta B g(1)) + \frac{\eta B}{\eta G} (g(\eta G) - \eta B g(1)) \right] < 0.
$$

(75)

where the inequality holds from the concavity of $g(x)$ (Lemma 2), which implies $(g(y) - g(x))/(y - x) > g(z) - g(x)/(z - x)$, for any $x < y < z$, and from the fact that $0 < \eta B < \eta G < \lambda G \leq 1$. The lemma is thus proved.

**REFERENCES**


Nicolò Michelusi (S’09-M’13) received the B.Sc. (with honors), M.Sc. (with honors) and Ph.D. degrees from the University of Padova, Italy, in 2006, 2009 and 2013, respectively, and the M.Sc. degree in Telecommunications Engineering from the Technical University of Denmark in 2009, as part of the T.I.M.E. double degree program. In 2011, he was on leave at the University of Southern California, Los Angeles, USA, as a visiting PhD student. He is currently a post-doctoral research fellow at the Ming Hsieh Department of Electrical Engineering, University of Southern California, USA. His research interests lie in the areas of communication theory, wireless networks, cognitive radio networks, energy harvesting systems, stochastic optimization and optimal control.


Kostas Stamatiou received his Diploma in Electrical and Computer Engineering from the National Technical University of Athens in 2000 and the M.Sc. and Ph.D. degrees in Electrical Engineering in 2004 and 2009, respectively, from the University of California San Diego. From 2009 to 2010 he was a post-doctoral scholar in the Department of Electrical Engineering at the University of Notre Dame, South Bend, Indiana, and from 2010 to 2012 he held a research appointment at the Department of Information Engineering at the University of Padova, Italy. He currently holds a Research Associate position at the Centre Tecnologic de Telecomunicacions de Catalunya (CTTC), Barcelona, Spain. His research interests lie in the areas of communication theory, stochastic geometry and random networks, energy harvesting systems and future smart grids. Dr. Stamatiou received the California Institute of Telecommunications and Information Technology Fellowship (CalIT2) in 2002 and the Marie Curie Fellowship in 2010. He has served as TPC member for the Communication Theory Symposium of the IEEE International Conference on Communications (ICC), and as reviewer for the IEEE Transactions on Communications, IEEE Transactions on Wireless Communications, IEEE Journal on Selected Areas in Communications and the IEEE Communications Letters.
Michele Zorzi (S’89, M’95, SM’98, F’07) was born in Venice, Italy, on December 6th, 1966. He received the Laurea and the PhD degrees in Electrical Engineering from the University of Padova, Italy, in 1990 and 1994, respectively. During the Academic Year 1992/93, he was on leave at the University of California, San Diego (UCSD) as a visiting PhD student, working on multiple access in mobile radio networks. In 1993, he joined the faculty of the Dipartimento di Elettronica e Informazione, Politecnico di Milano, Italy. After spending three years with the Center for Wireless Communications at UCSD, in 1998 he joined the School of Engineering of the University of Ferrara, Italy, where he became a Professor in 2000. Since November 2003, he has been on the faculty at the Information Engineering Department of the University of Padova. His present research interests include performance evaluation in mobile communications systems, random access in mobile radio networks, ad hoc and sensor networks, energy constrained communications protocols, broadband wireless access and underwater acoustic communications and networking.

Dr. Zorzi was the Editor-In-Chief of the IEEE WIRELESS COMMUNICATIONS MAGAZINE from 2003 to 2005 and the Editor-In-Chief of the IEEE TRANSACTIONS ON COMMUNICATIONS from 2008 to 2011, and currently serves on the Editorial Board of the WILEY JOURNAL OF WIRELESS COMMUNICATIONS AND MOBILE COMPUTING. He was also guest editor for special issues in the IEEE PERSONAL COMMUNICATIONS MAGAZINE (Energy Management in Personal Communications Systems) IEEE WIRELESS COMMUNICATIONS MAGAZINE (Cognitive Wireless Networks) and the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS (Multi-media Network Radios, and Underwater Wireless Communications Networks). He served as a Member-at-large of the Board of Governors of the IEEE Communications Society from 2009 to 2011.