Optimal transmission policies for Energy Harvesting Devices with Limited State-of-Charge Knowledge

Nicolò Michelusi, Leonardo Badia, and Michele Zorzi

Abstract

Wireless sensors can be integrated with rechargeable batteries and energy harvesting (EH) devices to allow for long-term, autonomous operation. To this end, efficient energy management policies are needed. Existing research relies on the assumption that the energy available to the sensor is known; however, the accurate estimation of the battery state-of-charge (SOC) in real-world devices is typically costly or impractical. This paper investigates the impact of imperfect SOC knowledge, and tackles the design of optimal operation policies to cope with such imperfect knowledge. The performance degradation with respect to the idealized scenario where the SOC is perfectly known is quantified, and it is shown that it decreases with increasing storage capacity and decreasing uncertainty in the EH source. It is shown that the performance degradation is within 5% for most cases of practical interest, and that near-optimal performance is achieved by only a loose knowledge of the SOC, which distinguishes between high and low SOC levels. Moreover, the impact of time correlation in the EH source is investigated, and it is shown that knowledge of the EH state is more critical than accurate SOC knowledge, so that a precise knowledge of the former can obviate the need for accurate information about the latter.

Index Terms

Wireless sensor networks; battery management; green design; renewable energy sources.

N. Michelusi is with the Dept. of Electrical Engineering, University of Southern California. email: michelus@usc.edu.

L. Badia and M. Zorzi are with the Dept. of Information Engineering, University of Padova, Padova, Italy. email: {badia, zorzi}@dei.unipd.it.

Preliminary versions of this paper have appeared at the IEEE International Workshop for Energy Harvesting for Communications – Co-located with IEEE ICC 2012 [1], and at the 8th International Wireless Communications and Mobile Computing Conference (IWCMC) [2].

I. Introduction

Energy harvesting devices (EHDs) can operate autonomously over long periods of time, as they are capable of collecting energy from the surrounding environment, *e.g.*, solar, motion, heat, and aeolian EH [3], [4], to sustain tasks such as data acquisition, processing and transmission. EHDs may find a very important application in wireless sensor networks (WSNs), which are distributed systems of autonomous devices, deployed over an area of interest with the purpose of monitoring (*e.g.*, sensing) environmental conditions and other relevant data, processing this information and transmitting or relaying it to a fusion center, with applications such as health care, disaster prevention, and industrial, agricultural, or home monitoring [5]. The wireless sensor devices in a WSN typically integrate multiple capabilities concerning the actual environmental sensing (data collection) but also transmission via radio techniques, as well as autonomous control of the overall network operation. In this regard, the exploitation of energy collected by the sensor devices from the surrounding environment makes it possible to overcome a severe limitation of battery powered WSN deployments, namely, the finite lifetime of wireless devices, which results in WSN failure and connectivity loss in those scenarios where battery replacement is costly or prohibitive. Since wireless sensors already involve an integration process between communication and sensing elements, it may be sensible to think of a further integration with EHDs, with the aim to achieve virtually continuous and indefinitely long network operation [6], [7], at least within the limits of physical degradation of the rechargeable batteries [8].

Combining EHDs with wireless sensors in a single terminal poses a new challenge, *i.e.*, how to optimally manage the harvested energy with the goal of optimizing the long-term performance related to sensing and data-communication tasks [7], [9], [10]. The EHD component of a sensor can be modeled as an *energy buffer*, where energy is stored according to a given statistical process, and from where it is drawn to feed sensor microprocessors and transceiver equipments, whenever needed. To find whether and how much energy can be used, a *battery management* algorithm is employed, which makes its decision based on internal state information, such as the amount of energy stored in the buffer or "state-of-charge" (SOC) [11], [12], the importance of the data packets to be transmitted [13], or the health state of the battery [8], as well as on external state information, such as the state of the external energy source [13] or the channel state [14]. In this regard, exact knowledge of the SOC is especially useful since, when the SOC

is low, the wireless sensor can remain idle to preserve energy, thus avoiding running out of energy during the execution of an important task.

However, practical EHDs store energy in electrochemical rechargeable batteries and/or supercapacitors. Thus, it is questionable to assume that the SOC of these devices can be characterized with infinite precision and immediate availability at any time. For example, [4] argues that the actual value of a supercapacitor capacitance may fluctuate by approximately 30% with respect to the value reported on the data sheet. The SOC level can be estimated online, although this comes at the price of additional energy and processing costs. Other algorithms [15] have been proposed to estimate the open circuit voltage, which is closely related to the SOC, but have non-negligible complexity, which may be an issue for resource-limited small devices. We thus conclude that SOC estimation is a complex and resource-expensive task and a precise knowledge of the SOC may be difficult to acquire.

Motivated by the aforementioned real-world concern, we claim that it is of interest to design battery management policies for scenarios where SOC knowledge is, if not entirely unavailable, at least imperfect. Thus, in this paper we consider an EHD where the controller knows the SOC only to the extent of a rough quantization, *i.e.*, a range of values it falls within, but not the exact value. We investigate policies regulating how energy should be drawn from the battery. These can be seen as the result of an optimization problem where the goal is to maximize a long-term reward metric, such as throughput. Differently from traditional investigations of sensor networks, where a constraint is set on the average long-term power used to perform a specific task, so as to attain a target lifetime of the device, here the constraints on the operation of the device are induced by the random fluctuations in the EH source and by the finite battery *storage capacity*. Specifically, we determine how partial knowledge of the SOC influences the goodness of the resulting solution. For example, in the special case of a linear reward function, it is shown that, under loose assumptions, the optimal policy when SOC knowledge is limited to two intervals, which can be denoted as LOW and HIGH, incurs no performance loss with respect to the idealized scenario where the SOC is perfectly known. The numerical evaluations, based on a logarithmic reward function, which models, for instance, the achievable capacity of a Gaussian channel, demonstrate that in typical scenarios the performance penalty due to imperfect SOC knowledge is at most 5%. The intuition behind all these results is that the optimal policy should aim at avoiding *energy outage* (*i.e.*, depleting the battery) when the SOC is LOW, while at the

same time being aggressive when the SOC is HIGH, in order to stay away from *energy overflow* (that is, receiving energy that cannot be stored because the buffer is already full). Beyond these operational criteria, a precise knowledge of the SOC yields only a marginal additional benefit.

Further, we address the impact of correlation in the EH source. We model the EH process as a hidden Markov chain, reflecting the underlying scenario of the energy generation process [16]. Intuitively, an efficient battery operation policy should also exploit knowledge of the state of this generation process, to better reduce the occurrences of energy outage and overflow events. In this regard, we show that adaptation to the underlying scenario process is extremely beneficial to achieve near-optimal performance. As a result, we show that knowledge of the state of the EH process and adaptation to it are more critical than perfect knowledge of the SOC. In particular, in the time-correlated setting, near-optimal performance (within 2% of the globally optimal policy) can be achieved by only knowing whether the SOC is high or low, but perfectly knowing the state of the EH source. On the other hand, a more significant performance degradation is incurred by neglecting the state of the EH source while knowing the SOC perfectly. This confirms our previous findings [13], where a simple balanced policy, which only adapts to the scenario process but not to the exact energy level in the battery, achieves performance within 3% of the globally optimal solution.

The rest of this paper is organized as follows. Sec. II introduces the mathematical representation of the EHD and the variables involved. Sec. III discusses the optimization methods and also derives some notable performance bounds and discusses some special scenarios. In Sec. IV, we present the main numerical results, In Sec. V, we discuss some extensions of the model, and Sec. VI concludes the paper.

II. System model

The main parameters of the model are listed in Table I. We consider an EHD, which scavenges energy from the surrounding environment (*e.g.*, solar, kinetic, wind, radio frequency). We assume a slotted-time system, where time slot $k \in \mathbb{N}_0$ corresponds to the time interval $[k, k+1)$. The harvested energy is stored in an energy buffer, in the form of energy quanta of value Δe [J], with capacity $e_{\max}$ (quanta). In the following sections, we further describe the model in terms of battery dynamics, scenario process, imperfect SOC, exogenous process, and reward function.

Table I

PARAMETERS OF THE MODEL.

Δe	Energy quantum (in J)	$e_{\max}$	Battery storage capacity (quanta)
$E_k \in \mathcal{E}$	SOC state (quanta)	$N_k \in \mathcal{N}$	Interval index of SOC state E_k
$Q_k \in \mathcal{Q}$	Action (quanta)	$B_k \sim p_B(B_k\|S_k)$	Energy harvested & distribution (quanta)
$S_k \sim p_S(S_k\|S_{k-1})$	EH scenario & transition prob.	$q_{\min},\ q_{\max}$	Minimum & maximum load requirements (quanta)
$C_k \sim p_{C\|W}(C_k\|W_k)$	Exogenous process & distribution	$W_k \sim p_W(W_k\|W_{k-1})$	Underlying exogenous state & transition prob.
$g(Q_k, C_k, E_k)$	Reward function	$G(\mu)$	Long-term average reward per time slot
μ	Policy		

A. Battery dynamics

The SOC available in the buffer at time instant k is denoted by $E_k \in \mathcal{E}$, taking values in the set $\mathcal{E} \equiv \{0, \ldots, e_{\max}\}$. At the beginning of the kth time slot, the EHD controller requests a number of energy quanta Q_k to be drawn from the buffer to perform a certain task, drawn from the *action space* $\mathcal{Q} = \{0\} \cup \{q_{\min}, q_{\min}+1, \ldots, q_{\max}\}$, where $0 < q_{\min} \leq q_{\max} \leq e_{\max}$. $q_{\min} \in \mathbb{N}$ and $q_{\max} \in \mathbb{N}$ represent the minimum and maximum load requirements, respectively. In particular, $q_{\min}$ may capture fixed energy costs, circuit power, as well as the energy cost of performing data acquisition and transmission, *e.g.*, see [17], [18]. Action $Q_k = 0$ accounts for the possibility to remain idle in a given time slot, due to either a controller's decision or energy outage. During the time slot duration, the EHD harvests $B_k \in \mathcal{B}$ energy quanta from the environment, which are stored in the buffer, where $\mathcal{B} = \{0, 1, \ldots, b_{\max}\}$ is the set of arrival values, with $b_{\max} \leq e_{\max}$. Accordingly, starting from the initial SOC level $E_0 \in \mathcal{E}$ available at time instant 0, the temporal evolution of the random variable E_k follows the equation

$$E_{k+1} = \min\left\{[E_k - Q_k]^+ + B_k, e_{\max}\right\}, \quad k \geq 0, \tag{1}$$

where $[\cdot]^+ \triangleq \max\{\cdot, 0\}$.

Remark 1 *Note that E_k, B_k and Q_k are discrete variables expressed in number of energy quanta. Since the energy quantum is Δe[J], the corresponding physical quantities are given by $\Delta e E_k$[J], $\Delta e B_k$[J] and $\Delta e Q_k$[J], respectively. However, in practice, these are continuous variables taking values in $\mathbb{R}$. This assumption thus represents a quantization of the corresponding real-valued variables. The error introduced by this quantization decreases with a finer granularity of the energy quantum Δe. However, the smaller the energy quantum, the larger the quantized battery capacity $e_{\max}$, maximum harvested energy $b_{\max}$ and maximum action $q_{\max}$ needed to represent such quantities, hence the higher the optimization complexity. Therefore, the value of*

the energy quantum reflects a trade-off between optimization complexity and modeling accuracy. Optimization of the energy quantum according to this trade-off is not considered in this paper and is left for future work.

The EH/consumption mechanism given by equation (1) entails two important phenomena. The former, denoted as *energy outage*, corresponds to the EHD running out of energy before the completion of the requested task, which happens when $Q_k > E_k$. This determines the failure of the requested task, and the energy available in the battery is depleted. Alternatively, *energy overflow* may occur if $B_k > e_{\max} - [E_k - Q_k]^+$, *i.e.*, the energy buffer is unable to store all of the harvested energy B_k, resulting in the loss of $B_k - e_{\max} + [E_k - Q_k]^+$ energy quanta. This is a consequence of the limited energy buffer capacity.

B. Scenario process

We model the energy arrival process $\{B_k\}$ as a homogeneous hidden Markov process, taking values in the set $\mathcal{B}$. We define an underlying *scenario process* $\{S_k\}$, taking values in the finite set $\mathcal{S}$, which evolves according to a stationary irreducible Markov chain with transition probability $p_S(s_{k+1}|s_k) \triangleq \mathbb{P}(S_{k+1}{=}s_{k+1}|S_k{=}s_k)$. Given the scenario $S_k = s$, the energy harvest B_k is drawn with probability mass function $p_B(b|s) \triangleq \mathbb{P}(B_k = b|S_k = s)$, for all $b \in \mathcal{B}, s \in \mathcal{S}$. We define $\pi_S(s),\ s \in \mathcal{S}$, as the steady state distribution of the scenario process, and we refer to $\bar{b} = \mathbb{E}[B_k] = \sum_{s\in\mathcal{S}} \pi_S(s) \sum_{b\in\mathcal{B}} b p_B(b|s)$ as the *average EH rate*. Note that this model is a special instance of the *generalized Markov model* presented in [16]. Therein, the scenario process is modeled as a first-order Markov chain, whereas B_k statistically depends on B_{k-L}^{k-1} and on S_k, for some order $L \geq 0$, where we have defined $X_i^j = (X_i, X_{i+1}, \ldots, X_j)$. In particular, in [16] it is shown that, by quantizing B_k with 20 states, $L = 0$ models well a piezoelectric energy source, whereas $L = 1$ models well a solar energy source. For simplicity, in this paper we assume $L = 0$. The analysis can be extended to the case $L = 1$, similarly to [13], but this extension is beyond the scope of this paper. The scenario S_{k-1} can be estimated from measurements of the past energy arrivals B_0^{k-1}. The posterior distribution of state S_{k-1} can be inferred recursively as

$$\mathbb{P}(S_{k-1} = s|B_0^{k-1}) \propto p_B(B_{k-1}|s) \sum_{\sigma\in\mathcal{S}} p_S(s|\sigma)\mathbb{P}(S_{k-2} = \sigma|B_0^{k-2}),$$

where $\propto$ denotes proportionality up to a normalization factor, independent of $s \in \mathcal{S}$, and $\mathbb{P}(S_{k-2} = \sigma|B_0^{k-2})$ is the posterior distribution inferred in the previous time slot. S_{k-1} can

then be estimated using, *e.g.*, a maximum a-posteriori (MAP) probability criterion $\hat{S}_{k-1} = \arg\max_{s\in\mathcal{S}} \mathbb{P}(S_{k-1} = s|B_0^{k-1})$. Typically, the scenario process $\{S_k\}$ is slowly varying over time, hence it can be estimated accurately from the harvested sequence $\{B_k\}$. In this paper, we assume that perfect knowledge of S_{k-1} is available at the EHD controller at time instant k. The case where S_{k-1} is only partially known (*e.g.*, via the posterior distribution) can be treated using the framework of POMDP, but its analysis is beyond the scope of this paper. On the other hand, only statistical knowledge of S_k is available at time k, since the energy arrival B_k, drawn from $p_B(B_k|S_k)$, has not yet been observed.

C. Imperfect SOC

By keeping track of B_k, Q_k, it is possible, to some extent, to gain some aggregate knowledge on the SOC E_k, via (1). However, this *open loop* approach is prone to error propagation, so that the estimate of E_k may become unreliable; yet, this information can serve to identify whether the SOC state is generally "high" or "low." Instead, accurate measurement of the current value of the charging state (scenario process S_{k-1}) may be much easier to acquire, as discussed above. Thus, we assume that only partial knowledge of the SOC E_k is available, *e.g.*, due to uncertainty in its estimation. We model this uncertainty by defining a partition of the SOC space, $\{\mathcal{I}(n), n \in \mathcal{N}\}$, where $\mathcal{I}(n) = \{\tilde{e}_n, \ldots, \tilde{e}_{n+1} - 1\}$, $\mathcal{N} \equiv \{0, \ldots, \tilde{n} - 1\}$, n is the nth SOC interval, and $0 = \tilde{e}_0 < \tilde{e}_1 < \cdots < \tilde{e}_{\tilde{n}} = e_{\max} + 1$ define the interval boundaries. Suppose that, at time k, $E_k \in \mathcal{I}(N_k)$, for some $N_k \in \{0, \ldots, \tilde{n} - 1\}$. Then, we assume that the EHD controller knows only the interval index N_k, *i.e.*, it knows that $E_k \in \mathcal{I}(N_k)$, rather than the exact SOC E_k. We define the *interval index* process $\{N_k, k \geq 0\}$, taking values in $\{0, \ldots, \tilde{n} - 1\}$. The special case with perfect SOC knowledge is obtained by letting $\tilde{n} = e_{\max} + 1$, hence $E_k = N_k$. In particular, we are interested in the case $\tilde{n} = 2$ with $\tilde{e}_1 = \lceil e_{\max}/2 \rceil$. In this case, we denote the two intervals $\mathcal{I}(0)$ and $\mathcal{I}(1)$ as LOW and HIGH SOC, respectively.

D. Exogenous process

We define an exogenous process $\{C_k, k \geq 0\}$, where C_k takes value in the set $\mathcal{C} \subset \mathbb{R}$, and an underlying Markov chain $\{W_k\}$, taking value in the finite set $\mathcal{W}$, with transition probabilities $p_W(w_2|w_1),\ w_1, w_2 \in \mathcal{W}$ and steady state distribution $\pi_W(w),\ w \in \mathcal{W}$. Given $W_k = w$, $C_k = c$ is distributed according to $p_{C|W}(c|w),\ c \in \mathcal{C}$, and is conditionally independent of $\{(W_j, C_j), j < k\}$. C_k models, *e.g.*, the channel gain in slot k, or the priority of the current data packet [13]. The

Markov assumption on W_k models memory effects, *e.g.*, time correlation in the channel gains. Similarly to the EH scenario S_{k-1}, W_k can be estimated by measuring the past values of the exogenous process $\{C_j, j \leq k\}$ (we assume that C_k is known to the EHD controller at time instant k). The posterior distribution of state W_k can then be inferred recursively as

$$\mathbb{P}(W_k = w|C_0^k) \propto p_B(C_k|w) \sum_{\omega \in \mathcal{W}} p_W(w|\omega)\mathbb{P}(W_{k-1} = \omega|C_0^{k-1}).$$

where $\mathbb{P}(W_{k-1} = \omega|C_0^{k-1})$ is the posterior distribution inferred in the previous time slot. W_k can then be estimated using, *e.g.*, the MAP criterion $\hat{W}_k = \arg\max_w \mathbb{P}(W_k = w|C_0^k)$. Typically, $\{W_k\}$ varies slowly over time, hence it can be estimated accurately from the exogenous sequence $\{C_k\}$. For instance, if W_k represents the average channel gain associated to the large-scale fading, whereas C_k is the actual small-scale fading channel gain, W_k can be estimated accurately by a moving time average of appropriate window length, which depends on the coherence time of the small- and large-scale fading. For simplicity, in this paper we assume that the state of the exogenous process at time k, denoted as (W_k, C_k), is known to the EHD, which can schedule the amount of energy Q_k accordingly. The case where (W_k, C_k) is only partially known (*e.g.*, via the posterior distribution) can be treated using the framework of POMDP, but its analysis is beyond the scope of this paper.

E. Reward function

We define the reward function $g(q, c, e)$ when the SOC level is $E_k = e \in \mathcal{E}$, action $Q_k = q$ is chosen, and the exogenous process takes value $C_k = c$, as

$$g(q, c, e) = \begin{cases} 0 & q > e, \\ \tilde{g}(q, c) & q \leq e, \end{cases} \tag{2}$$

where $\tilde{g} : \mathcal{Q} \times \mathcal{C} \mapsto \mathbb{R}^+$ is an increasing function of q with $\tilde{g}(0, c) = 0$, for any $c \in \mathcal{C}$, and increasing in c, for each q. Notice that, if $q > e$, then $g(q, c, e) = 0$, which models an energy outage event, *i.e.*, the incapability of the wireless sensor to complete the task assigned. More in general, the reward function $g(q, c, e)$ may be the expected reward with respect to a "nuisance" process U_k, independent and identically distributed (i.i.d.) over time, as seen in Example 1.

Example 1 *Let C_k be the channel gain, known to the EHD controller. Letting U_k be the power of the interference from nearby terminals, modeled as an exponential random variable with mean m_U ($U_k \sim Exp(m_U)$), unknown to the EHD controller, the signal-to-noise ratio (SNR) at the*

receiver is given by $\frac{Q_k C_k}{1+U_k}$.[1] *In this case,* $\tilde{g}(q,c)$ *may model the success probability to transmit* R *bits,* i.e.*, assuming that the transmission is successful if and only if* $R < \frac{1}{2}\log_2\left(1+\frac{qc}{1+U_k}\right)$ *[19], or equivalently,* $U_k > \frac{qc}{2^{2R}-1} - 1$*, and using the fact that* $U_k \sim Exp(m_U)$*, we obtain*

$$\tilde{g}(q,c) = \mathbb{P}\left(U_k \leq \frac{qc}{2^{2R}-1} - 1\right) = \begin{cases} 0, & q \leq \frac{2^{2R}-1}{c}, \\ 1 - \exp\left\{-\frac{\frac{qc}{2^{2R}-1}-1}{m_U}\right\}, & q > \frac{2^{2R}-1}{c}, \end{cases} \tag{3}$$

Knowledge of C_k *can be exploited to perform power adaptation* (Q_k) *to the channel gain.*

Remark 2 *Note that, in the special case* $q_{\min} = q_{\max} = 1$*,* $B_k \in \{0,1\}$ *with binary scenario process* $S_k \in \{G,B\}$*, where* $S_k = G$ *and* $S_k = B$ *denote the "good" and "bad" EH scenarios, respectively,* $p_B(0|B) = 1$ *(*$B_k = 0$ *if* $S_k = B$*),* $p_B(1|G) = \lambda_G$ *(*$B_k = 1$ *with probability* λ_G *and* $B_k = 0$ *otherwise, if* $S_k = G$*), and* $W_k = 1,\ \forall k$*, we obtain the model considered in [13] as a special case, where* C_k *represents the i.i.d. "importance" of the current data packet. Therein,* balanced *policies are developed which adapt only to the scenario state* S_{k-1} *but not to the exact SOC* E_k*, thus not requiring knowledge of the SOC, and it is shown that they achieve near-optimal performance with respect to the globally optimal policy. In this paper, we investigate the impact of imperfect SOC knowledge for the more general setting. Moreover, when* $C_k = 1,\ \forall k$*, we obtain the model analyzed in [2] as a special case and, by further letting* $S_k = 1,\ \forall k$*, we obtain the case with i.i.d. EH process analyzed in [1].*

III. Optimization Problem

A. Policy definition and problem statement

A policy μ is a function that decides on the amount of energy Q_k to be requested from the buffer, based on the interval index N_k at time instant k, the previous scenario state S_{k-1}, the previous energy outage event $O_{k-1} = \chi(Q_{k-1} > E_{k-1})$, where $\chi(\cdot)$ denotes the indicator function, the exogenous state (C_k, W_k), and, possibly, the history $\mathcal{H}_k$, which includes

[1]In the SNR expression, it is assumed that the interference is treated as a Gaussian random variable and the noise is assumed to be zero-mean Gaussian with unit mean. Note that the exact expression of the SNR should include the noise power spectral density at the denominator. For notational simplicity, we ignore this issue, which amounts to assuming the presence of a normalization factor, *e.g.*, in the definition of the transmit power.

all past values of these quantities, *i.e.*, $(N_0^{k-1}, O_0^{k-2}, C_0^{k-1}, W_0^{k-1}, S_{-1}^{k-2}, Q_0^{k-1})$. In particular, $\mu(q|N_k, O_{k-1}, \mathcal{H}_k, C_k, W_k, S_{k-1})$ denotes the probability that $Q_k = q \in \mathcal{Q}$ in slot k.[2]

We define the *long-term average reward per time slot* under policy μ, starting from state $E_0{=}e_0$, $S_{-1}{=}s_{-1}$ and $W_0 = w_0$, as

$$G(\mu, e_0, s_{-1}, w_0) \triangleq \lim_{K\to\infty} \inf \frac{1}{K} \mathbb{E}_\mu \left[\sum_{k=0}^{K-1} g(Q_k, C_k, E_k) \,\middle|\, E_0 = e_0, S_{-1} = s_{-1}, W_0 = w_0 \right], \quad (4)$$

where the expectation is computed with respect to the realization of the random variables $\{B_k, S_k, Q_k, C_k, W_k, O_k\}$ induced by policy μ, for $k = 0, \ldots, K{-}1$. In all cases of practical interest, the Markov chain induced by a policy μ has a unique communicating class, hence the long-term reward is independent of the initial state (E_0, S_{-1}, W_0) [21], therefore we denote it as $G(\mu)$ in the following treatment. The problem is to determine a policy μ^* such that

$$\mu^* = \arg\max_{\mu} G(\mu). \quad (5)$$

Due to the partial knowledge of the SOC, problem (5) can be recast in the context of partially observable Markov decision processes (POMDPs) [20], and can be solved by using numerical optimization tools available in the literature [22]. For this case, due to properties of MDPs and POMDPs [23], the optimal policy μ is deterministic and is a function of the *belief state on the energy level* $\Pi_k(e)$, $e \in \mathcal{E}$ (since E_k is not perfectly observed) and the other state variables (C_k, W_k, S_{k-1}) (perfectly observed), *i.e.*, $Q^*(\Pi, c, w, s)$ is the optimal action in state $(\Pi_k, C_k, W_k, S_{k-1}) = (\Pi, c, w, s)$, and $\mu^*(q|\Pi, c, w, s) = \chi(q = Q^*(\Pi, c, w, s))$[3]. However, the optimal POMDP formulation suffers from the curse of dimensionality for the following reasons:

- *Policy optimization complexity:* the optimization via, *e.g.*, value iteration, has high complexity and poor convergence properties, since the optimal action needs to be determined for every possible value of (Π, c, w, s), where the belief Π is a probability distribution over a $|\mathcal{I}(n)|$-dimensional space, $n \in \mathcal{N}$ and $(c, w, s) \in \mathcal{C} \times \mathcal{W} \times \mathcal{S}$;
- *Operational complexity:* once the optimal policy μ^* has been determined, it needs to be stored in a look-up table; in particular, the optimal action needs to be stored for every

[2]We remark that, for the sake of generality, a randomized policy is defined. However, the optimal POMDP policy is deterministic for this case [20]. See the discussion following (5).

[3]Note that μ^* is independent of $N_k, O_{k-1}, \mathcal{H}_k$, since this information is used to compute the belief state Π_k, which is sufficient for decision making [20].

possible value of (Π, c, w, s), resulting in very demanding storage requirements, which are not available in practical deployments; moreover, the belief $\Pi_k(\cdot)$ needs to be updated in each slot, as observations are collected, resulting in additional complexity overhead.

Therefore, due to the limited processing capability of EHDs, in this paper we focus on suboptimal policies, which neglect the history $\mathcal{H}_k$ available up to time slot k, and only map the current interval index, exogenous state and scenario state (N_k, C_k, W_k, S_{k-1}) to the probability $\mu(q|N_k, C_k, W_k, S_{k-1})$ of drawing q energy quanta from the buffer. Note that, for this case, the optimal policy may not be deterministic, unlike the optimal POMDP formulation. In fact, by using randomization, the risk of incurring energy outage, resulting from the uncertainty on the true SOC E_k, can be optimally balanced. The advantages resulting from such formulation and other approximations will be discussed in Remark 5.

Remark 3 *Note that, in the case of perfect SOC knowledge $E_k = N_k$, neglecting the history does not incur any loss of optimality, since the sequence $\{(E_k, C_k, W_k, S_{k-1}, Q_k), k \geq 0\}$ constitutes a Markov decision process [23]. In contrast, in the case of imperfect SOC knowledge, the history could bring additional information about the current energy level E_k, hence neglecting it incurs a loss of optimality in general. For instance, knowing only that $N_k = n$ implies that $E_k \in \{\tilde{e}_n, \ldots, \tilde{e}_{n+1} - 1\}$. If, additionally, it is known that $N_{k-1} = n + 1$ (*i.e.*, $E_{k-1} \in \{\tilde{e}_{n+1}, \ldots, \tilde{e}_{n+2} - 1\}$), $B_{k-1} = 0$ and $1 \leq Q_{k-1} \leq \min\{\tilde{e}_{n+2} - \tilde{e}_{n+1}, \tilde{e}_{n+1} - \tilde{e}_n\}$, and that no outage nor overflow occurred in slot $k-1$ (resulting in $E_k = E_{k-1} - Q_{k-1}$ from (1)), by combining this information with $N_k = n$ we obtain $\tilde{e}_n < \tilde{e}_{n+1} - Q_{k-1} \leq E_k \leq \tilde{e}_{n+1} - 1$, hence $E_k \notin \{\tilde{e}_n, \tilde{e}_n + 1, \ldots, \tilde{e}_{n+1} - Q_{k-1} - 1\}$, so that more information can be inferred on the value of E_k. In the numerical results, we will evaluate the validity of such approximation by comparing it with the idealized scenario where the SOC is known perfectly.*

B. Upper bound and Balanced policy

In this section, we derive an upper bound to the long-term reward $G(\mu)$. To this end, notice that the EH mechanism induces the constraint $\sum_{k=0}^{K-2} B_k + E_0 \geq \sum_{k=0}^{K-1} \min\{Q_k, E_k\}$, *i.e.*, the amount of energy consumed up to slot $K-1$ cannot exceed the amount of energy harvested up to slot $K-2$. Dividing by K and taking the expectation and the limit for $K \to \infty$, we obtain

$$\lim_{K\to\infty} \frac{1}{K} \mathbb{E}_\mu \left[\sum_{k=0}^{K-1} \min\{Q_k, E_k\} \right] \leq \bar{b}. \tag{6}$$

We thus obtain the bound

$$G(\mu) = \lim_{K\to\infty} \inf \frac{1}{K}\mathbb{E}_\mu \left[\sum_{k=0}^{K-1} g(Q_k, C_k, E_k)\right]$$
$$\leq \lim_{K\to\infty} \inf \frac{1}{K}\mathbb{E}_\mu \left[\sum_{k=0}^{K-1} \tilde{g}(\min\{Q_k, E_k\}, C_k)\right] \leq g^*(\min\{q_{\max}, \bar{b}\}), \quad (7)$$

where we have defined $g^*(x)$, $x \in [0, q_{\max}]$ as the solution of the linear program (LP)

$$g^*(x) = \max_{y(\cdot|\cdot)} \sum_{q\in\mathcal{Q}, c\in\mathcal{C}} y(q|c)\pi_C(c)\tilde{g}(q,c) \quad (8)$$

$$\text{s.t.} \sum_{q\in\mathcal{Q}, c\in\mathcal{C}} y(q|c)\pi_C(c)q \leq x, \quad (9)$$

$$\sum_{q\in\mathcal{Q}} y(q|c) = 1, \ \forall c \in \mathcal{C}, \qquad y(q|c) \geq 0, \ \forall q \in \mathcal{Q}, c \in \mathcal{C},$$

where $\pi_C(c) = \sum_w p_{C|W}(c|w)\pi_W(w), c \in \mathcal{C}$ is the steady-state distribution of C_k, $y(q|c)$ represents the conditional probability of using action $Q_k = q$ when $C_k = c$, and the EH mechanism is replaced by a (looser) average energy per slot constraint with intensity x. The LP (8) can be solved efficiently using, *e.g.*, interior point methods [24]. In particular, since $\tilde{g}(q, c)$ is an increasing function of q, the inequality constraint (9) is attained with equality under the optimal $y^*(\cdot|\cdot)$, and therefore it can be replaced with the equality constraint $\sum_{q\in\mathcal{Q}, c\in\mathcal{C}} y(q|c)\pi_C(c)q = x$.

We denote the policy solving the optimization problem (8) for $x = \min\{q_{\max}, \bar{b}\}$ as *balanced policy*, $\hat{\mu}_{BP}$, which draws $Q_k = q$ energy quanta with probability $\hat{\mu}_{BP}(q|c)$, when the exogenous state is $C_k = c$, *independently* of the SOC E_k, N_k and exogenous state W_k. In particular, if $\bar{b} \leq q_{\max}$, this policy draws, on average, $\bar{b}$ energy quanta from the battery in each slot, equal to the average EH rate, hence the name *balanced policy* [13]. This policy achieves the upper bound (7), for asymptotically large battery capacity $e_{\max} \to \infty$. In fact, for infinite battery capacity, no energy overflow can occur, hence all the harvested energy is used for reward accrual.

C. Perfect and Imperfect SOC knowledge

When perfect SOC knowledge is available at the EHD controller, policy μ maps the state of the system (E_k, C_k, W_k, S_{k-1}) to the probability of drawing q energy quanta from the buffer. The sequence $\{(E_k, C_k, W_k, S_{k-1}, Q_k), k \geq 0\}$ constitutes a Markov decision process, and the long-term reward is maximized by a stationary, deterministic policy [25]. In this case, the optimal policy is found by using standard tools, such as policy or value iteration [25]. Thus,

the long-term reward under perfect SOC knowledge represents an upper bound to the performance of any policy under SOC uncertainty. Conversely, under SOC uncertainty, the sequence $\{(N_k, C_k, W_k, S_{k-1}, Q_k), k \geq 0\}$ does not constitute a Markov process, hence the optimal policy μ^* cannot be found via the policy/value iteration algorithm. Moreover, problem (5) under policies of the form $\mu(q|n,c,w,s)$ cannot in general be recast as a convex optimization problem, so that multiple *local* maxima may exist, and convergence to the *global* maximum may not be achieved under any finite complexity algorithm. Therefore, we recur to sub-optimal methods, that aim at achieving a local maximum rather than a global one. Furthermore, due to the high dimensionality of the optimization, we restrict it to policies of the form

$$\mu(q|n,c,w,s) = \hat{\mu}(q|n,c,w,\bar{Q}(n,w,s)), \tag{10}$$

where $\bar{Q}(n,w,s)$ is the *expected action* in state $(N_k, W_k, S_{k-1}) = (n,w,s)$, after marginalization with respect to C_k, and $\hat{\mu}(q|n,c,w,x)$, $x \in [0, \min\{q_{\max}, \tilde{e}_{n+1} - 1\}]$ is the solution of the LP

$$\hat{\mu}(\cdot|n,\cdot,w,x) = \arg\max_{y(\cdot|\cdot)} \sum_{q\in\mathcal{Q},c\in\mathcal{C}} y(q|c) p_{C|W}(c|w) \tilde{g}(q,c) \tag{11}$$

$$\text{s.t.} \sum_{q\in\mathcal{Q},c\in\mathcal{C}} y(q|c) p_{C|W}(c|w) q \leq x, \quad \sum_{q\in\mathcal{Q}} y(q|c) = 1, \ \forall c \in \mathcal{C}, \tag{12}$$

$$y(q|c) = 0, \ \forall q \geq \tilde{e}_{n+1}, \ \forall c \in \mathcal{C}, \ y(q|c) \geq 0, \ \forall q \in \mathcal{Q}, c \in \mathcal{C}.$$

Similarly to (8), this problem can be solved efficiently [24]. Moreover, since $\tilde{g}(q,c)$ is an increasing function of q, the inequality constraint (12) is attained with equality under the optimal $y^*(\cdot|\cdot)$, and therefore it can be replaced with the equality constraint $\sum_{q\in\mathcal{Q},c\in\mathcal{C}} y(q|c) p_{C|W}(c|w) q = x$. Moreover, the constraint $y(q|c) = 0, \ \forall q \geq \tilde{e}_{n+1}, \ \forall c \in \mathcal{C}$ is due to the fact that, when $E_k \in \mathcal{I}(n)$, then $E_k \leq \tilde{e}_{n+1} - 1$, hence any $Q_k \geq \tilde{e}_{n+1}$ would lead to outage and is thus sub-optimal. Note that (11) and (8) are different optimization problems. In fact, in (11), we are conditioning to $W_k = w$, whereas in (11) we are marginalizing with respect to its steady-state distribution $\pi_W(w), \ w \in \mathcal{W}$. The maximization in (11) is independent of the EH process (S_{k-1}, B_k), and in fact the EH mechanism is replaced by a (looser) constraint on the expected amount of energy drawn in each slot (12).

Policy (10) can be interpreted as follows. When the EHD is in state (N_k, W_k, S_{k-1}), it draws from the battery an expected amount of energy equal to $\bar{Q}(N_k, W_k, S_{k-1})$. The actual action Q_k is random, depending on the realization of the exogenous random variable C_k. The dependence

of (10) on C_k is such that, neglecting the impact of energy outage, the *expected instantaneous* reward with respect to C_k is maximized, over all policies that draw an expected amount of $\bar{Q}(N_k, W_k, S_{k-1})$ quanta from the battery, as expressed in the optimization problem (11). Note that policy (10) is fully described by the expected action $\bar{Q}(\cdot)$, which does not depend on the exogenous process C_k, thus yielding a reduction in the optimization complexity. Using the functional constraint (10), the optimization of the expected action policy $\bar{Q} : \mathcal{N} \times \mathcal{W} \times \mathcal{S} \mapsto [0, q_{\max}]$ is expressed as the solution of the non-convex optimization problem

$$\bar{Q}^* = \arg\max_{\bar{Q}} G(\bar{Q}), \tag{13}$$

where we have defined

$$G(\bar{Q}) = \sum_{n\in\mathcal{N}} \sum_{e\in\mathcal{I}(n)} \sum_{w\in\mathcal{W}} \sum_{s\in\mathcal{S}} \pi_{\bar{Q}}(e, w, s) \sum_{q,c} \hat{\mu}(q|n, c, w, \bar{Q}(n, w, s)) p_{C|W}(c|w) g(q, c, e), \tag{14}$$

and $\pi_{\bar{Q}}(e, w, s)$ is the steady state distribution of state $(E_k, W_k, S_{k-1}) = (e, w, s)$ induced by the expected action policy $\bar{Q}$, defined as the solution of

$$\begin{cases} \sum_{n\in\mathcal{N}} \sum_{e\in\mathcal{I}(n)} \sum_{w\in\mathcal{W}} \sum_{s\in\mathcal{S}} \pi_{\bar{Q}}(e, w, s) = 1, \text{ (normalization)}, \\ \pi_{\bar{Q}}(e, w, s) \geq 0, \ \forall (e, w, s) \in \mathcal{E} \times \mathcal{W} \times \mathcal{S}, \text{ (non-negativity)}, \\ \sum_{n_k\in\mathcal{N}} \sum_{e_k\in\mathcal{I}(n_k)} \sum_{w_k\in\mathcal{W}} \sum_{s_{k-1}\in\mathcal{S}} \pi_{\bar{Q}}(e_k, w_k, s_{k-1}) \mathbb{P}_{\bar{Q}}\left(e_{k+1}, s_k \,|e_k, w_k, s_{k-1}\right) p_W(w_{k+1}|w_k) \\ \quad = \pi_{\bar{Q}}(e_{k+1}, w_{k+1}, s_k), \ \forall (e_{k+1}, w_{k+1}, s_k) \in \mathcal{E} \times \mathcal{W} \times \mathcal{S}, \text{ (steady-state)}, \end{cases} \tag{15}$$

where $\mathbb{P}_{\bar{Q}}\left(e_{k+1}, s_k \,|e_k, w_k, s_{k-1}\right) p_W(w_{k+1}|w_k)$ is the transition probability of the Markov chain $\{E_k, W_k, S_{k-1}\}$ under the expected action policy $\bar{Q}$, and is given by

$$\mathbb{P}_{\bar{Q}}\left(e_{k+1}, s_k|e_k, w_k, s_{k-1}\right) = p_S(s_k|s_{k-1}) \sum_{q,b,c} p_B(b|s_k) \hat{\mu}(q|\eta(e_k), c, w_k, \bar{Q}(\eta(e_k), w_k, s_{k-1}))$$

$$\times\, p_{C|W}(c|w_k) \chi\left(\min\left\{[e_k - q]^+ + b, e_{\max}\right\} = e_{k+1}\right), \tag{16}$$

where we have used (1) and we have defined $\eta(e_k)$ as the index of the SOC interval which e_k belongs to, *i.e.*, $e_k \in \mathcal{I}(\eta(e_k))$. Note that the randomness induced by C_k is captured in the expected instantaneous reward in (14) and is marginalized in the steady-state probabilities (15).

Remark 4 *The same policy (11) can be defined for the case of perfect SOC knowledge ($N_k = E_k$), in order to reduce the dimensionality of the optimization (*i.e.*, remove the dependence on the*

exogenous random variable C_k). In this case, the optimal expected action $\bar{Q}^*(e,w,s),\ (e,w,s) \in \mathcal{E} \times \mathcal{W} \times \mathcal{S}$, can be found efficiently using the policy or value iteration algorithms.

In general, policy (10) may be sub-optimal for all expected action policies $\bar{Q}$, since it neglects the impact of the distribution of Q_k, induced by the random C_k, on the evolution of E_k. However, we argue that, if $e_{\max} \gg q_{\max}$, *i.e.*, the battery storage capacity is sufficiently large, then (10) yields a good approximation, since the actual distribution of Q_k impacts the performance only at the battery boundaries (*i.e.*, when the battery is almost depleted or almost fully charged), whereas, for intermediate values of the SOC, the performance is primarily affected by the average amount of energy drawn from the battery in each slot, $\bar{Q}(N_k, W_k, S_{k-1})$. In fact, assuming W_k varies slowly over time, *e.g.*, it is constant over T slots, then C_k is i.i.d. over this interval of T slots, and the fluctuations in the energy drawn from the buffer Q_k and in the reward $\tilde{g}(Q_k, C_k)$, induced by the realization of C_k, are averaged out over relatively short time scales, thus not impacting the average long-term performance. This is indeed true for asymptotically large battery capacity for the balanced policy $\hat{\mu}_{BP}$, as discussed in Sec. III-B.

D. Special case

We now consider the special scenario of a linear reward function with a two-interval quantization of the SOC and constant exogenous process, for which the optimal policy can be derived in closed form. Proposition 1 shows that always transmitting when the SOC is in the HIGH state, and refraining from transmitting when it is LOW, is a sufficient condition for optimality.

Proposition 1 (Linear Reward) *Under a linear reward function $\tilde{g}(q,c) = \alpha q$, constant exogenous process $C_k = 1$, a general energy arrival process, and the following assumptions:*
(a) Two-interval SOC uncertainty, i.e., $\tilde{n} = 2$, $\mathcal{I}(0) = \{0, \dots, \tilde{e}_1 - 1\}$ and $\mathcal{I}(1) = \{\tilde{e}_1, \dots, e_{\max}\}$,
(b) $b_{\max} \leq \min\{\tilde{e}_1, e_{\max} + 1 - \tilde{e}_1, q_{\max}\},\ \tilde{e}_1 \geq q_{\min}$,
the optimal reward is $G^* = \alpha \bar{b}$, and one optimal policy is

$$Q_k = N_k \max\{b_{\max}, q_{\min}\}. \tag{17}$$

Proof: As discussed in Sec. III-B, we have the upper bound $G(\bar{Q}) \leq g^*(\bar{b}) = \alpha \bar{b}$, since $\bar{b} \leq b_{\max} \leq q_{\max}$ by hypothesis *(b)*. We now prove that policy (17) achieves this upper bound. Since the bound holds for any policy μ, (17) is also optimal under perfect SOC knowledge.

If $E_k \in \mathcal{I}(0)$, then $Q_k = 0$ from (17). From (1), we have $E_{k+1} = \min\{E_k + B_k, e_{\max}\}$. Since $B_k \leq b_{\max}$ and $E_k \leq \tilde{e}_1 - 1$ (from $E_k \in \mathcal{I}(0)$), from hypothesis *(b)* we have that $E_k + B_k \leq \tilde{e}_1 - 1 + b_{\max} \leq e_{\max}$. This implies that neither overflow nor outage occurs when $E_k \in \mathcal{I}(0)$, hence $g(Q_k, C_k, E_k) = \tilde{g}(Q_k, C_k) = \alpha Q_k$.

If $E_k \in \mathcal{I}(1)$, then $Q_k = \max\{b_{\max}, q_{\min}\}$ from (17). Since $E_k \geq \tilde{e}_1 \geq Q_k$, outage does not occur, hence $g(Q_k, C_k, E_k) = \tilde{g}(Q_k, C_k) = \alpha Q_k$. Moreover, since $B_k \leq b_{\max} \leq Q_k$, at any time slot enough energy quanta are drawn from the buffer to make room for the new arrivals, hence overflow does not occur.

Since neither overflow nor outage occurs at any time, we have $E_{k+1} = E_k - Q_k + B_k$ and $g(Q_k, C_k, E_k) = \tilde{g}(Q_k, C_k) = \alpha Q_k$. All harvested energy contributes to reward accrual, hence

$$G^* = \lim_{K\to\infty} \inf \frac{1}{K} \mathbb{E}\left[\sum_{k=0}^{K-1} \alpha Q_k \middle| E_0 = e_0, S_{k-1} = s_{k-1}\right] = \alpha \bar{b},$$

which proves the achievability of the upper bound. ■

If the length of the intervals $\mathcal{I}(0), \mathcal{I}(1)$ differ by at most one unit, *i.e.*, $\tilde{e}_1 = \lceil e_{\max}/2 \rceil$, then assumption *(b)* simplifies to $b_{\max} \leq \lceil e_{\max}/2 \rceil$, *i.e.*, the buffer capacity is at least twice the maximum energy that can be harvested in a time slot. Note that any policy avoiding energy outages and overflows is optimal in the linear reward case, so that, in general, (17) may not be the only optimal solution.

Harvested energy is wasted, thus incurring a performance degradation, when there is energy outage due to uncertain SOC knowledge, or energy overflow due to limited energy buffer capacity. When the energy arrival or exogenous processes are random, the controller has limited knowledge about the future arrivals of energy and packet importance. In this case, overflow can be avoided by an *aggressive* policy, which draws $Q_k \geq b_{\max}$ energy quanta when the battery SOC approaches its capacity. This choice guarantees that enough energy quanta are drawn from the buffer, thus making room for the new energy arrival. Moreover, outage can be avoided by a *conservative* policy $Q_k = 0$, which stays idle when the battery SOC approaches depletion. According to Proposition 1, this approach is optimal under a deterministic exogenous process and linear reward function. In this case, the EHD only needs to know whether the energy available is either LOW or HIGH, so that the controller can remain idle or transmit at high power in order to avoid outage and overflow, respectively. Optimal performance is thus achieved by only adapting to the (quantized) SOC, but not to the scenario state S_k. Due to the time-sharing nature of this solution,

sub-optimal performance may be achieved when the reward function is not linear, and/or when C_k is random.

E. Exhaustive and Local Search Algorithms

For a more general exogenous process $\{(C_k, W_k)\}$, EH process $\{(B_k, S_k)\}$ and reward function $\tilde{g}(q, c)$, the optimal policy is difficult to characterize in closed-form, due to the random fluctuations in these quantities. In order to further reduce the dimensionality of the optimization problem (13), the expected action $\bar{Q}(n, w, s)$ can be restricted to take value in the discrete set $\mathcal{J} \equiv \{jq_{\max}/M, j = 0, 1, \ldots, M\}$, for some $M > 0$, rather than the continuous interval $[0, q_{\max}]$. In order to solve (13) optimally under this restriction, an exhaustive search can be carried out. Note that there are $(M+1)^{|\mathcal{W}|\cdot\tilde{n}\cdot|\mathcal{S}|}$ possible expected action policies. Therefore, an exhaustive search algorithm requires to compute (14) and (15) for every such policy, thus determining the policy with maximum reward. An exhaustive search algorithm is thus feasible only for some special cases, *e.g.*, if the exogenous (C_k) and EH (B_k) processes are i.i.d. ($|\mathcal{W}| = 1$ and $|\mathcal{S}| = 1$), and the battery is quantized to two intervals ($\tilde{n} = 2$), with M not too large. For the other cases of practical interest, the complexity of the exhaustive search algorithm may be too large, due to the number of policies that need to be evaluated. For these cases, we resort to a *local search method* to optimize the policy, as described below, which guarantees convergence to a local maximum of (13), rather than the global one.

Algorithm 1 (Local Search) *1) Let $\bar{Q}^{(0)} : \mathcal{W} \times \mathcal{N} \times \mathcal{S} \mapsto \mathcal{J}$ be an initial expected action policy, and $i = 0$.*

2) In stage i:

- *Initialize $\bar{Q}^{(i+1)} = \bar{Q}^{(i)}$. For $w \in \mathcal{W}$ $n \in \mathcal{N}$, $s \in \mathcal{S}$, sequentially update $\bar{Q}^{(i+1)}$ as*

$$\bar{Q}^{(i+1)}(n, w, s) := \underset{\bar{Q} \in \mathcal{Q}^{(i+1)}(n,w,s)}{\arg\max} G(\bar{Q}), \tag{18}$$

where we have defined

$$\mathcal{Q}^{(i+1)}(n, w, s) \equiv \big\{\bar{Q} : \mathcal{N} \times \mathcal{W} \times \mathcal{S} \mapsto \mathcal{J} : \tag{19}$$

$$\bar{Q}(\hat{n}, \hat{w}, \hat{s}) = \bar{Q}^{(i+1)}(\hat{n}, \hat{w}, \hat{s}), \forall (\hat{n}, \hat{w}, \hat{s}) \neq (n, w, s), \bar{Q}(n, w, s) \in \mathcal{J}\big\}.$$

- *If $\bar{Q}^{(i+1)} = \bar{Q}^{(i)}$, return $\bar{Q}^{(i+1)}$. Else, update the counter as $i := i + 1$ and repeat 2).*

This algorithm sequentially determines a local optimum of (13) by unilaterally optimizing the action performed on each tuple (w, n, s), until convergence to a local stable point (note that the

optimization is performed on a discrete set, so this is not exactly a "local maximum" in a strict sense). This happens when $\bar{Q}^{(i+1)} = \bar{Q}^{(i)}$ for some $i \geq 0$, *i.e.*, any unilateral change in policy $\bar{Q}^{(i)}$ does not lead to an improved reward. Since the set of policies has size $(M+1)^{|\mathcal{W}|\tilde{n}|\mathcal{S}|}$, convergence of the algorithm is guaranteed within at most $(M+1)^{|\mathcal{W}|\tilde{n}|\mathcal{S}|}$ evaluations of the reward $G(\bar{Q})$ (typically, much fewer iterations are needed). A generally good initialization is the *balanced policy*

$$\bar{Q}^{(0)}(n,w,s) = \sum_{q\in\mathcal{Q}}\sum_{c\in\mathcal{C}} q\hat{\mu}_{BP}(q|c)p_{C|W}(c|w), \qquad \forall (w,n,s) \in \mathcal{W}\times\mathcal{N}\times\mathcal{S}, \tag{20}$$

which is asymptotically optimal for large battery capacity (Sec. III-B).

Remark 5 *The local search algorithm brings the following benefits with respect to the "optimal" POMDP formulation (5):*

- Policy optimization complexity: *we have verified that, typically, only few iterations of the local search algorithm are sufficient, when initialized with the balanced policy, thus requiring only few evaluations of (14) and (15) in stage 2). This represents a significant complexity saving with respect to the POMDP formulation, where the optimal action needs to be determined,* e.g., *using the iterative value iteration algorithm, for each value of the belief state* (Π, c, w, s)*;*
- Operational complexity: *once the expected action* $\bar{Q}(n,w,s)$ *has been determined, it can be stored in a look-up table, for each possible value of* (n,w,s)*. Hence,* $\log_2(M+1)|\mathcal{W}|\tilde{n}|\mathcal{S}|$ *bits are required to store such policy. This represents a significant storage saving with respect to the POMDP formulation, where the optimal action needs to be stored for every possible value of* (Π, c, w, s)*. Moreover, the belief state* Π_k *need not be tracked.*

IV. Numerical Results

We present quantitative evaluations for both the case of independent energy arrivals, and a more realistic scenario with correlated EH. For simplicity, we assume that the exogenous process is i.i.d. over time, *i.e.*, $W_k = 1,\ \forall k$ and $\{C_k\}$ is i.i.d. with distribution $\pi_C(c)$, so that any dependence on the exogenous state W_k can be neglected. In fact, the main focus of this paper is on the impact of time correlation of the EH process and imperfect knowledge of the SOC on the performance of EHDs. However, the framework we have proposed in this paper allows also to model time-correlated exogenous processes, whose study is left as an item for future research.

We consider the reward function

$$\tilde{g}(q,c) = \frac{1}{2}\log_2(1+qc), \tag{21}$$

which represents the achievable capacity under Gaussian signaling over an AWGN channel with gain $C_k = c$ [19], so that qc represents the signal-to-noise ratio (SNR) at the receiver. Therefore, the average long-term metric $G(\bar{Q})$ represents the throughput of scheme $\bar{Q}$. The channel is Rayleigh fading, and its gain C_k has uniform distribution in the discrete set $\mathcal{C} \equiv \{-\alpha\ln(u), u \in \{j/(N_C+1), j = 1,2,\ldots,N_C\}\}$, which represents a quantized exponential random variable. In fact, when $N_C \to \infty$, we have $C_k = -\alpha\ln(U_k)$, where $U_k \sim \mathcal{U}((0,1))$, so that C_k approaches the exponential distribution $\pi_C(c) = \frac{1}{\alpha}e^{-c/\alpha}$. Herein, we use $N_C = 10$. The parameter α is set so as to achieve a target average SNR $\Lambda = \bar{b}\sum_{c\in\mathcal{C}} c\pi_C(c)$ at the receiver, where Λ is computed assuming transmissions occur with expected energy $\bar{b}$ in each slot, due to the EH constraint. We then obtain

$$\alpha = -\frac{\Lambda N_C}{\bar{b}\ln\left(\frac{N_C!}{(N_C+1)^{N_C}}\right)}. \tag{22}$$

For this case, if the action Q_k were allowed to take values in the compact set $[0, q_{\max}]$, the solution of the optimization problem (11) would be given by the well-known water-filling solution [19] (with maximum power constraint $\min\{q_{\max}, \tilde{e}_{n+1}-1\}$ when $E_k \in \mathcal{I}(n)$), given by $\hat{\mu}(q|n,c,x) = \chi(q = q_{WF}(n,c,x))$, where

$$q_{WF}(n,c,x) \triangleq \min\left\{\left[\lambda(x) - \frac{1}{c}\right]^+, q_{\max}, \tilde{e}_{n+1}-1\right\}, \tag{23}$$

where $\lambda(x)$ uniquely solves

$$\frac{1}{N_C}\sum_{j=1}^{N_C} q_{WF}(n, \alpha\ln((N_C+1)/j), x) = x. \tag{24}$$

The water-filling solution is asymptotically achieved by a sufficiently small energy quantum, *i.e.*, large $e_{\max}$ [energy quanta] and $q_{\max}$ [energy quanta]. As in Sec. III-B, assuming $\bar{b} \le q_{\max}$, an upper bound $G_{\rm UP}$ to $G(\mu)$, achieved for asymptotically large battery capacity $e_{\max}$ and fine grained quantization (small energy quantum) is then given by

$$G_{\rm UP} = \frac{1}{2}\mathbb{E}\left[\min\left\{[\log_2(\lambda(\bar{b})C_k)]^+, \log_2(1+q_{\max}C_k)\right\}\right], \tag{25}$$

where the expectation is computed with respect to C_k.

In the next two sections, we consider the i.i.d. and time-correlated EH scenarios

A. I.i.d. Energy Arrivals

We consider a scenario with $\bar{b} = 10$ and (unless otherwise stated) a geometric energy arrival distribution truncated at $b_{\max} = 4\bar{b}$, with probability mass function

$$p_B(b) = e^{-\beta b} \frac{1 - e^{-\beta}}{1 - e^{-\beta(b_{\max}+1)}}, \tag{26}$$

where β uniquely solves $\bar{b} = \sum_{b=0}^{b_{\max}} b p_B(b)$, which can be determined using a bisection method [24]. We let $q_{\min} = 1$ and $q_{\max} = b_{\max} = M$ (recall that M is the quantization of the average action space $[0, q_{\max}]$). This choice represents a good trade-off between a sufficiently fine-grained quantization of the physical quantities of interest and, at the same time, manageable computation time. Note that, due to the discrete action space $\mathcal{Q}$, the water-filling solution (23) cannot be achieved exactly. Therefore, we approximate $\hat{\mu}(q|n,c,x)$, *i.e.*, the solution of the optimization problem (11), as

$$\hat{\mu}\left(q|n,c,x\right) = \begin{cases} q_{WF}(n,c,x) + 1 - \lceil q_{WF}(n,c,x) \rceil, & q = \lceil q_{WF}(n,c,x) \rceil, \\ \lceil q_{WF}(n,c,x) \rceil - q_{WF}(n,c,x), & q = \lceil q_{WF}(n,c,x) \rceil - 1, \\ 0 & \text{otherwise.} \end{cases} \tag{27}$$

The smaller the energy quantum Δe employed to discretize the action space $\mathcal{Q}$, the better this approximation, *i.e.*, the water-filling solution (23) is asymptotically achieved by $\hat{\mu}\left(q|n,c,x\right)$ as the energy quantum Δe approaches 0. We have verified that, for the chosen value $q_{\max} = 40$, this approximation is indeed very good. This solution is such that $\sum_{q\in\mathcal{Q}} q\hat{\mu}\left(q|n,c,x\right) = q_{WF}(n,c,x)$, *i.e.*, for each value of the exogenous variable $C_k = c$ and SOC interval $N_k = n$, the average energy expenditure of policy $\hat{\mu}\left(q|n,c,x\right)$ is the same as that of the water-filling solution. Unless otherwise stated, we use the average SNR $\Lambda = 10$.

We consider the following policies: balanced policy $\hat{\mu}_{BP}$ (BP), defined in Sec. III-B, policy with perfect SOC knowledge (IID), optimized via the policy iteration algorithm [25], policy with no SOC knowledge (P1iid, *i.e.*, one-interval uncertainty $\mathcal{I}(0) = \mathcal{E}$), and policy with two-equal-interval uncertainty (P2iid, *i.e.*, $\mathcal{I}(0) = \{0, \ldots, \tilde{e}_1 - 1\}$ and $\mathcal{I}(1) = \{\tilde{e}_1, \ldots, e_{\max}\}$ with $\tilde{e}_1 = \lceil \frac{e_{\max}}{2} \rceil$). Moreover, we plot the upper bound (UB), given by (25). The label "iid" is used for policies that neglect the correlation in the EH process and treat it as i.i.d. While this is optimal in the case considered in this section, where the EH process is indeed i.i.d., these policies are sub-optimal in the time-correlated case considered in Sec. IV-B. Note that BP can be seen as a specific, non-optimal, instance of a policy with one-interval uncertainty.

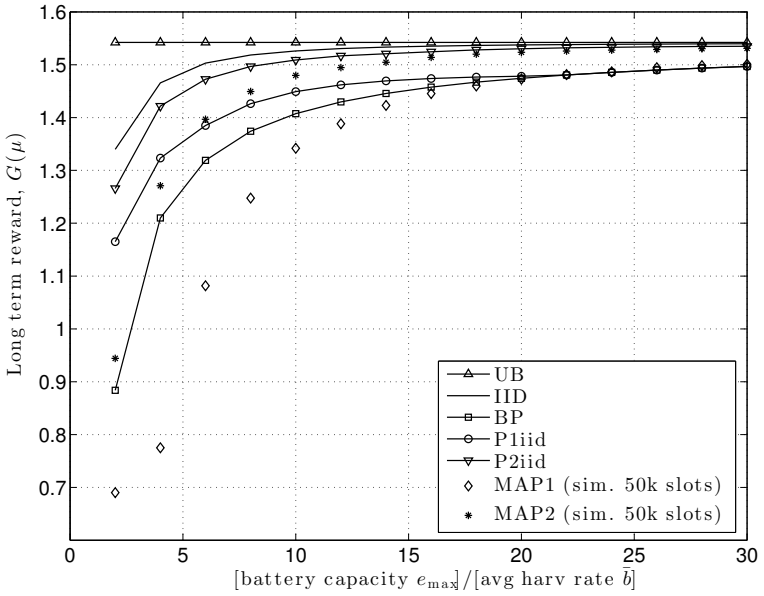

Figure 1. Throughput as a function of $e_{\max}/\bar{b}$, for the i.i.d. EH scenario, for different policies. ($\Lambda = 10$, $\bar{b} = 10$)

In Fig. 1, we plot the long-term throughput $G(\mu)$ vs. the ratio of the buffer capacity over the average EH rate $e_{\max}/\bar{b}$. For this case, we plot also the performance of the globally optimal policy computed via policy iteration under the assumption of perfect SOC knowledge, but operated based on a MAP estimate $\hat{E}_k$ of the SOC E_k, rather than the true SOC E_k. The MAP estimate is obtained from the posterior belief $\Pi_k(e)$, *i.e.*, $\hat{E}_k = \arg\max_{e \in \mathcal{I}(N_k)} \Pi_k(e)$, which is updated over time as a function of the action Q_{k-1}, the outage event $O_{k-1} = \chi(Q_{k-1} > E_{k-1})$, the SOC measurement N_k, and the previous belief $\Pi_{k-1}(E_{k-1})$. We denote these policies as MAP1 and MAP2, for the cases of one-interval and two-equal-interval uncertainty, respectively. As expected, the best performance is achieved by IID, followed by P2iid, P1iid and BP. At a buffer capacity $e_{\max} = 2\bar{b}$, the performance degradation of P2iid with respect to IID is about 5%, and that of P1iid with respect to IID is about 13%. As $e_{\max}$ increases, the degradation becomes smaller (*e.g.*, 0.5% for P2iid and 4% for P1iid, at $e_{\max} = 20\bar{b}$), since the impact of outage and overflow, which occur when the SOC approaches 0 and $e_{\max}$, respectively, becomes smaller. Also note that P1iid performs better than BP, since BP is a special instance of a policy that does not exploit SOC knowledge, and P1iid is optimized among such policies. We have verified that P1iid is more conservative than BP for small values of $e_{\max}$. On the other hand, for large $e_{\max}$, P1iid draws energy with rate $\bar{b}$, so that it performs the same as BP. Finally, we note that MAP1

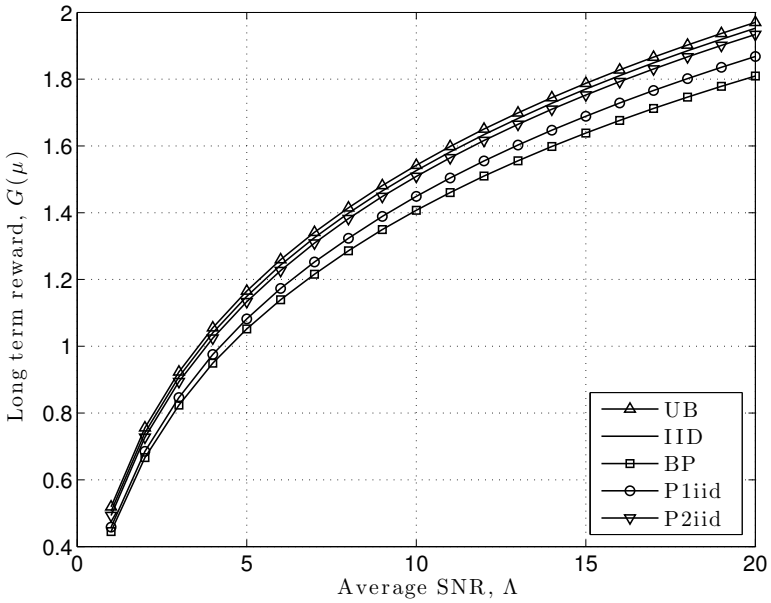

Figure 2. Throughput as a function of the average SNR Λ. ($e_{\max} = 100$, $\bar{b} = 10$)

and MAP2 perform poorly, compared to P1iid and P2iid, respectively. This is due to the fact that MAP1 and MAP2 operate based on the optimistic assumption of perfect SOC knowledge, and thus incur frequent outages, whereas P1iid and P2iid are designed to cope with SOC uncertainty, and thus keep into account the risk of outage events resulting from such uncertainty.

In Fig. 2, we plot $G(\mu)$ versus Λ, for $e_{\max} = 10\bar{b}$. It is seen that, as the average SNR increases, the performance also improves, as expected. In this case, the performance degradations of P2iid and P1iid with respect to IID are within 2% and 10%, respectively, for all SNR values, and decrease for increasing SNR, approaching 1% and 4%, respectively, for $\Lambda = 20$.

Fig. 3 examines the dependence of $G(\mu)$ on the energy arrival statistics. All the EH processes considered have the same average rate $\bar{b} = 10$. Other than the geometric EH process, we consider a Bernoulli process taking values 0 and B with probability $1-\bar{b}/B$ and $\bar{b}/B$, respectively, where $B \geq \bar{b}$ is varied. In particular, for $B = \bar{b}$, we obtain a deterministic process $B_k = B, \ \forall k$. It is seen that, the larger the variance of the EH process $\mathrm{var}(B_k)$, the larger the performance degradation, under both perfect and imperfect SOC knowledge. Interestingly, the geometric and Bernoulli energy arrival distributions with the same mean $\bar{b} = 10$ and variance $\mathrm{var}(B_k) = 84$ yield approximately the same performance. This suggests that the performance depends on the distribution of the energy arrivals mostly through its second order statistics. For the case of a

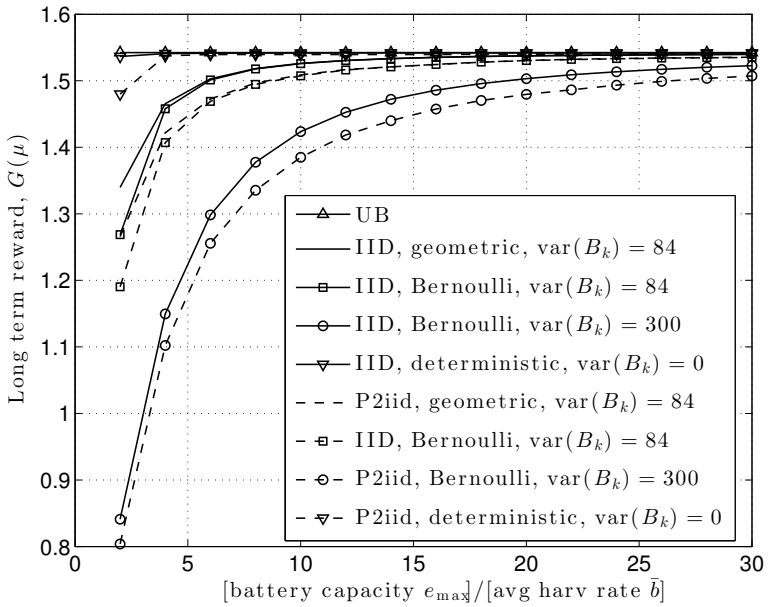

Figure 3. Throughput as a function of $e_{\max}/\bar{b}$, for different statistics of the EH process with the same rate $\bar{b}$. ($\Lambda = 10$, $\bar{b} = 10$)

constant energy arrival process $B_k = \bar{b}$, $\forall k$, we notice that the upper bound is attained (except for low battery capacity). As the capacity increases, the impact of an erratic energy source, in terms of outage and overflow, becomes smaller, hence UB is approached for large $e_{\max}$.

Note that the model developed in this paper does not account for quantization errors. However, in practice, the quantization region N_k may be measured with errors. In Fig. 4, we evaluate the impact of such errors on the performance of P2iid. In particular, letting Y_k be the observation of the quantization region N_k available to the EHD controller, we assume that $Y_k \neq N_k$ with *quantization error probability* p_e, and $Y_k = N_k$ otherwise. For instance, if $N_k = 0$ (LOW), then $Y_k = 1$ (HIGH) with probability p_e. The *expected action* $\bar{Q}(y, w, s)$ (note that this is a function of the observation $Y_k = y$, rather than the true region N_k) can then be optimized using Algorithm 1, where from (16), the transition probability used to compute $G(\bar{Q})$ is given by

$$\mathbb{P}_{\bar{Q}}(e_{k+1}, s_k | e_k, w_k, s_{k-1}) = p_S(s_k|s_{k-1}) \sum_{q,b,c} p_B(b|s_k) p_{C|W}(c|w_k) \chi\big(\min\{[e_k - q]^+ + b, e_{\max}\} = e_{k+1}\big)$$

$$\times [(1-p_e)\hat{\mu}(q|\eta(e_k), c, w_k, \bar{Q}(\eta(e_k), w_k, s_{k-1})) + p_e \hat{\mu}(q|1-\eta(e_k), c, w_k, \bar{Q}(1-\eta(e_k), w_k, s_{k-1}))],$$

where we have used the fact that, if $N_k = \eta(E_k) \in \{0, 1\}$, then a quantization error results in $Y_k = 1 - \eta(E_k)$. From Fig. 4, we note that the performance of P2iid degrades as the quantization error probability increases. For $p_e \leq 0.12$, the performance degradation of P2iid due

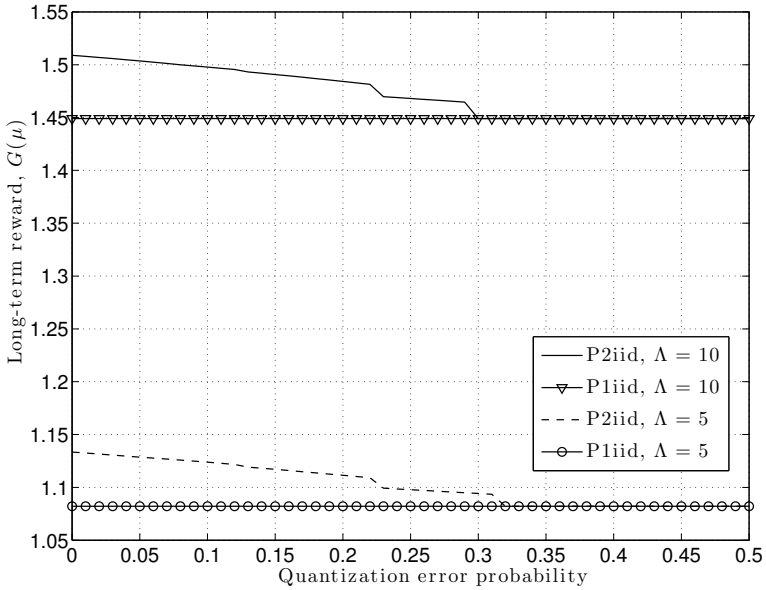

Figure 4. Throughput as a function of the quantization error probability. ($e_{\max} = 100$, $\bar{b} = 10$)

to quantization errors is within 1%. Moreover, for $p_e \geq 0.32$, P2iid attains the same performance as P1iid. In fact, P1iid is unaffected by quantization errors, since it completely neglects the SOC information N_k. As p_e increases, SOC readings become more and more unreliable and, when $p_e = 0.5$, they don't carry any information as to the correct value of N_k, so that the performance of P1iid, which does not have access to such information, is attained.

B. Correlated Energy Arrivals

In this section, we present numerical results for the case of a time-correlated EH process. The analysis of Sec. III can be applied to any distribution of the energy arrivals that depends on an underlying Markov "scenario" process. As a specific example, we consider an energy arrival process $\{B_k\}$ with average EH rate $\bar{b} = 10$, and three possible generation scenarios $S_k \in \{G, B, R\}$, representing a "good," "bad," and "random" state of the charging process, respectively. In the "random" scenario $S_k = R$, the arrival process follows a geometric distribution with mean $\bar{b}$, truncated at $b_{\max} = 4\bar{b}$, as in the previous section. In the "good" scenario $S_k = G$, the arrival B_k takes the value $B_k = 2\bar{b}$ deterministically, *i.e.*, $p_B(2\bar{b}|G) = 1$, $p_B(b|G) = 0,\ \forall\ b \neq 2\bar{b}$. Finally, in the "bad" scenario B, the arrival B_k takes the value $B_k = 0$ deterministically, *i.e.*, $p_B(0|B) = 1$, $p_B(b|B) = 0,\ \forall\ b \neq 0$.

The transition probabilities of S_k are defined via the transition matrix $[\mathbf{P}_S]_{s0,s1} = p_S(s1|s0)$ as

$$\mathbf{P}_S = \begin{bmatrix} p_S(R|R) & \frac{1-p_S(R|R)}{2} & \frac{1-p_S(R|R)}{2} \\ 0.05 & 0.95 & 0 \\ 0.05 & 0 & 0.95 \end{bmatrix}, \tag{28}$$

where $p_S(R|R) \in [0.5, 1]$. This process represents a situation where, under given conditions, *i.e.*, in the good scenario $S_k = G$, arrivals of energy are guaranteed. On the other hand, in the bad scenario $S_k = B$, no energy is harvested at all. Finally, the random scenario $S_k = R$ is a transient scenario between the good and bad scenarios, where the arrival process exhibits a random behavior. Notice that the transition probabilities $p_S(G|B) = p_S(B|G) = 0$, so that scenario G (respectively, B) cannot be directly reached by scenario B (G), but only via the transient scenario R. The steady state distribution of the energy arrival states is

$$\pi_S(R) = \frac{0.05}{1.05 - p_S(R|R)}, \ \pi_S(G) = \pi_S(B) = \frac{1}{2}\frac{1 - p_S(R|R)}{1.05 - p_S(R|R)}. \tag{29}$$

Note that, for any value of $p_S(R|R)$, the average EH rate is constant and equals $\bar{b}$. Moreover, when $p_S(R|R) = 1$, we obtain the i.i.d. energy arrival scenario of the previous section. In this case, the "random" scenario $S_k = R$ absorbs the Markov chain $\{S_k\}$, hence the energy arrival process $\{B_k\}$ exhibits an i.i.d. behavior with probability mass function $p_B(b|R), \ b \in \mathcal{B}$. We use the same reward function $\tilde{g}(q, c)$ as in the previous subsection, defined in (21).

In the numerical results, we compare two classes of policies. Namely, we consider a set of policies with perfect knowledge of the scenario S_{k-1}, in particular: policy with perfect SOC knowledge (OPT), policy with no SOC knowledge (P1corr, *i.e.*, one-interval uncertainty $\mathcal{I}(0)$=$\mathcal{E}$), and policy with two intervals uncertainty of equal size (P2corr, *i.e.*, $\mathcal{I}(0)$=$\{0, \dots, \tilde{e}_1 - 1\}$ and $\mathcal{I}(1)$=$\{\tilde{e}_1, \dots, e_{\max}\}$ with $\tilde{e}_1 = \lceil \frac{e_{\max}}{2} \rceil$). Additionally, we consider a set of policies which, on the other hand, neglect the underlying Markov structure of the energy arrivals and the time correlation, and treat them as i.i.d. with distribution given by the marginal $p_B(b) = \sum_{s \in \mathcal{S}} \pi_S(s) p_B(b|s)$. This set of policies is optimized assuming such marginal i.i.d. setting, but their performance is computed for the actual setting where B_k is time-correlated. For this case, we consider: balanced policy (BP), policy with perfect SOC knowledge (IID), policy with no SOC knowledge (P1iid, *i.e.*, one-interval uncertainty $\mathcal{I}(0) = \mathcal{E}$), and policy with two intervals uncertainty of equal size (P2iid, *i.e.*, $\mathcal{I}(0) = \{0, \dots, \tilde{e}_1 - 1\}$ and $\mathcal{I}(1) = \{\tilde{e}_1, \dots, e_{\max}\}$ with $\tilde{e}_1 = \lceil \frac{e_{\max}}{2} \rceil$). Policies OPT and IID are obtained via policy iteration as discussed in Sec. III-C (assuming Markovian and

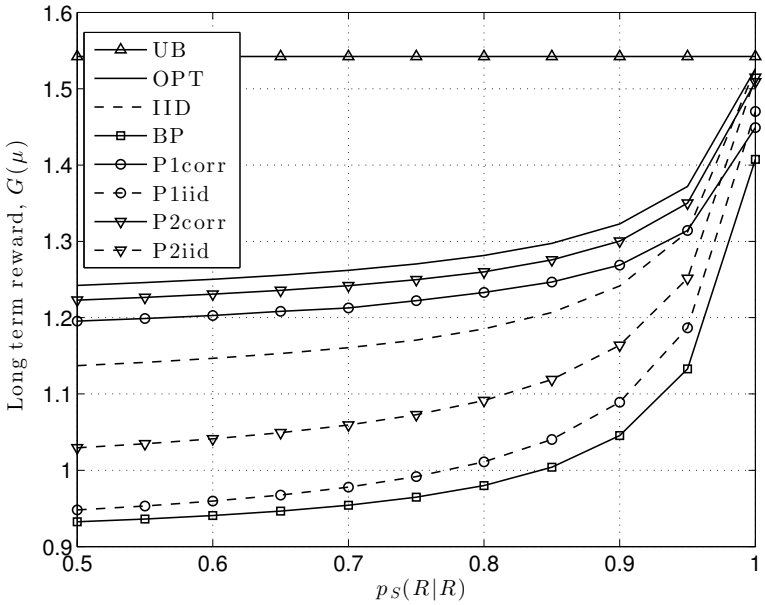

Figure 5. Throughput as a function of $p_S(R|R)$, for different policies. ($\Lambda = 10$, $\bar{b} = 10$)

i.i.d. energy arrivals, respectively). Policies P1corr, P2corr, P1iid and P2iid, on the other hand, are obtained using the local search Algorithm 1.

In Fig. 5, we plot the long-term throughput as a function of the transition probability $p_S(R|R)$. As discussed above, when $p_S(R|R) \to 1$, we approach the i.i.d. setting considered in the previous section, hence the policies which neglect the Markov structure of the EH process B_k and instead treat it as i.i.d. become optimal as $p_S(R|R) \to 1$. In general, we notice that the long-term throughput improves for increasing values of $p_S(R|R)$, *i.e.*, as the time correlation decreases and the process B_k approaches the i.i.d. distribution. This result is in line with the analysis done in [13] for a special case of the model considered in this paper, where $S_k \in \{G, B\}$ and $B_k \in \{0, 1\}$ with $p_B(1|G) = \lambda_G \in [0, 1]$, $p_B(1|B) = 0$, $Q_k \in \{0, 1\}$ and C_k represents the "importance" of the current data packet. Therein, it is shown that, for fixed battery capacity, the performance degrades as the EH process becomes more time-correlated. In fact, the device experiences intervals in which no energy is harvested ("bad" scenario), during which the battery is discharged until it becomes depleted, thus incurring energy outage, and intervals in which energy is harvested abundantly ("good" scenario), during which the battery is recharged until it becomes fully charged, thus incurring energy overflow. Conversely, when the EH process is i.i.d., a small battery suffices to filter out the randomness of (C_k, B_k) over short time scales.

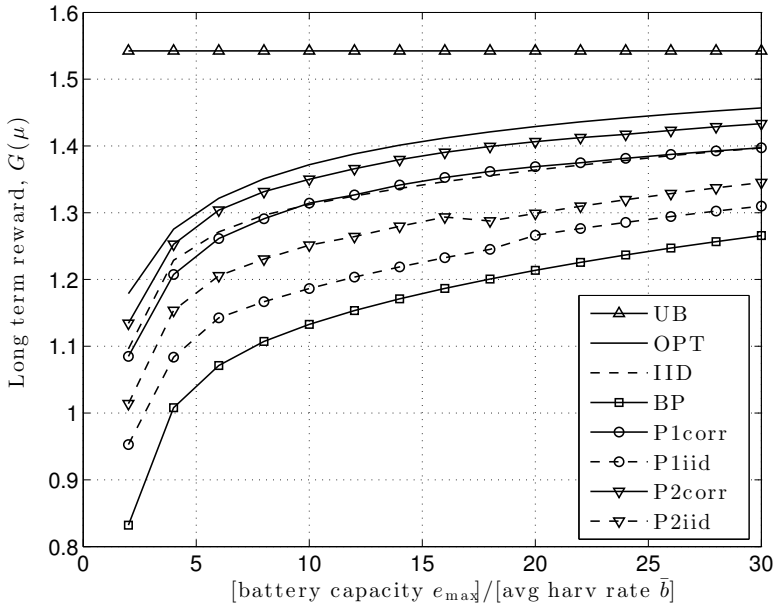

Figure 6. Throughput as a function of $e_{\max}/\bar{b}$ ($p_S(R|R) = 0.95$, $\Lambda = 10$, $\bar{b} = 10$)

In general, we observe that the policies designed under the assumption of an i.i.d. EH process perform significantly worse than those policies that, instead, exploit the actual energy arrival distribution and knowledge of S_{k-1}. This result suggests that perfect knowledge of the scenario process S_{k-1}, but only a loose knowledge of the SOC, suffices to achieve near-optimal performance (in particular, the performance degradation of P2corr and P1corr is within 2% and 5% of the globally optimal policy OPT, respectively). On the other hand, if perfect knowledge of the SOC is available, but the time correlation in the EH process is neglected (policy IID), a much more severe performance degradation may be incurred with respect to OPT (17% for $p_S(R|R) = 0.5$). Clearly, as $p_S(R|R)$ approaches 1, knowledge of S_{k-1} becomes less and less critical, and the performance of IID approaches that of OPT. We conclude that, in the time-correlated scenario, perfect knowledge of the scenario process S_{k-1} is more critical than perfect knowledge of the SOC, in order to achieve near-optimal performance. This result is in line with [13], where low-complexity policies achieving near-optimal performance are designed that only adapt to the scenario process S_{k-1}, but not to the SOC.

In Fig. 6, we plot $G(\mu)$ as a function of the ratio $e_{\max}/\bar{b}$ for $p_S(R|R) = 0.95$. The best performance is achieved by OPT, followed by P2corr. P1corr and IID obtain lower long-term reward but, interestingly, perform very similarly to each other. Even lower rewards are obtained

by P2iid, P1iid, and BP (the worst policy). Unlike Fig. 5, where P1corr outperforms IID (*i.e.*, knowledge of S_{k-1} is more critical than knowledge of the SOC), in this case P1corr and IID perform the same. This is because $p_S(R|R) = 0.95$, so that the EH process is almost i.i.d. As $p_S(R|R)$ approaches 1, knowledge of S_{k-1} becomes less and less critical and, in the limit as $p_S(R|R) \to 1$, the EH process is i.i.d. and IID becomes optimal. The performance degradation of P2corr is within 2% of OPT, and that of P1corr and IID is within 5%.

As a general conclusion, we can observe that the performance in the case of incomplete information on the SOC, but perfect knowledge of the scenario S_{k-1}, is affected only by a limited loss. Especially, knowing only whether the SOC is HIGH or LOW (2 quantization intervals) but perfectly knowing S_{k-1} (P2corr), incurs a small degradation with respect to OPT (typically within 2%). Similarly, not knowing at all the SOC, but perfectly knowing S_{k-1} (P1corr), incurs a degradation within 5% of OPT. On the other hand, the effect of perfectly knowing the current SOC but not adapting to the scenario S_{k-1} (policy IID) incurs a more significant performance degradation in the time-correlated case, which vanishes as the EH process becomes closer to i.i.d. When accurate knowledge of neither the SOC nor the energy arrival process is available, the performance significantly degrades, yet, if the battery capacity $e_{\max}$ is sufficiently large, the degradation decreases, and the performance approaches that of OPT for $e_{\max} \to \infty$. In this asymptotic case, the balanced policy BP, which does not adapt to the SOC nor to the scenario S_{k-1}, becomes optimal, as discussed in Sec. III-B.

V. Extensions

In this section, as in [8], we show how the model (1) can be extended to include non-idealities, such as battery leakage, sensing, processing and activation costs. The impact of some of these phenomena has been analyzed from an information theoretic perspective in [26], [27] (battery leakage) and [12] (sensing and processing costs). In particular, (1) can be extended to

$$E_{k+1} = \min\left\{[E_k - Q_k - L_k]^+ + B_k, e_{\max}\right\}, \tag{30}$$

where L_k is the overall energy cost in slot k, not including the control Q_k, resulting from battery leakage, sensing, processing and activation of the circuitry after the node goes to sleep (if $Q_{k-1} = 0$). We model L_k as a random variable with probability distribution $p_L(L_k|Q_k, A_k)$ taking values in the set $\mathcal{L} \triangleq \{0, 1, \ldots, L_{\max}\}$, possibly dependent on the action Q_k, and on the *activity* state A_k. The activity state $A_k = \chi(Q_{k-1} > 0)$ tracks the idle/active mode of the

sensor node, so that, if $A_k = 1$, then the node was active in the previous slot $k-1$ ($Q_{k-1} > 0$); otherwise, the node was idle ($Q_{k-1} = 0$). The dependence of p_L on A_k may be used to model activation costs of the sensor circuitry, *i.e.*, $\mathbb{P}(L_k \geq l|Q_k, A_k = 0) \geq \mathbb{P}(L_k \geq l|Q_k, A_k = 1)$, $\forall l, \forall Q_k > 0$, so that a higher energy cost is incurred when switching from idle to active mode ($A_k = 0$ and $Q_k > 0$) than when staying active ($A_k = 1$ and $Q_k > 0$).

For this more general model, policy μ decides on the amount of energy Q_k to be requested from the buffer, given $(N_k, C_k, W_k, S_{k-1}, Q_k, A_k)$. Note that, in this case, the policy is also a function of the activity state $A_k \in \{0, 1\}$. The reward function (2) when $(Q_k, C_k, E_k, L_k) = (q, c, e, l)$ can then be extended to accommodate such non-idealities as

$$g(q, c, e, l) = \begin{cases} 0 & q > [e - l]^+ \\ \tilde{g}(q, c) & q \leq [e - l]^+. \end{cases} \tag{31}$$

As in Sec. III-C, we can reduce the dimensionality of the optimization problem by restricting it to policies of the form $\mu(q|n, c, w, s, a) = \hat{\mu}(q|n, c, w, a, \bar{T}(n, w, s, a))$, where $\bar{T}(n, w, s, a)$ is the *total expected energy cost* in state $(N_k, W_k, S_{k-1}, A_k) = (n, w, s, a)$ (comprising both the non-ideality cost L_k and the action Q_k), after marginalization with respect to the realization of the exogenous state C_k, and $\hat{\mu}(q|n, c, w, a, x)$ is defined as the solution of the LP

$$\hat{\mu}(\cdot|n, \cdot, w, a, x) = \arg\max_{y(\cdot|\cdot)} \sum_{q \in \mathcal{Q}, c \in \mathcal{C}} y(q|c) p_{C|W}(c|w) \tilde{g}(q, c) \tag{32}$$

$$\text{s.t.} \sum_{q \in \mathcal{Q}, c \in \mathcal{C}} y(q|c) p_{C|W}(c|w) \left(q + \bar{L}(q, a)\right) \leq x, \tag{33}$$

$$\sum_{q \in \mathcal{Q}} y(q|c) = 1, \ \forall c \in \mathcal{C}, \qquad y(q|c) \geq 0, \ \forall q \in \mathcal{Q}, c \in \mathcal{C},$$

and $\bar{L}(q, a) \triangleq \sum_{l \in \mathcal{L}} p_L(l|q, a) l$. Policy $\bar{T}(n, w, s, a)$ can then be optimized by Algorithm 1, by replacing the *expected action* $\bar{Q}^{(i)}$ with the *total expected energy cost* $\bar{T}^{(i)}$, and the local optimization is done with respect to the extended argument $w \in \mathcal{W}$ $n \in \mathcal{N}$, $s \in \mathcal{S}$, $a \in \{0, 1\}$, so as to account for the additional activity state $A_k \in \{0, 1\}$.

VI. Conclusions and Future Developments

Motivated by the characteristics of real-world implementations, we have investigated energy management policies for EHDs, under the assumption of imperfect knowledge of the SOC of the battery and time-correlated energy arrivals. In both cases, having partial information on

either of them improves the performance with respect to having no information at all. Yet, our numerical evaluations suggest that there is little gain in having costly procedures to determine the SOC with high accuracy, while accurate estimation of the state of the EH source appears to be more critical to achieve near-optimal performance. In particular, the degradation due to SOC uncertainty increases with decreasing battery capacity and increasing variance of the energy arrival process, and knowing only if the SOC is HIGH or LOW performs within 2% of the globally optimal policy for typical parameter values, so that close-to-optimal performance may be achieved by having only a loose knowledge of the SOC and an accurate knowledge of the state of the EH source.

The model can be extended to consider the impact of battery degradation phenomena induced by the frequent charge and discharge cycles of the battery [28], [29], using the battery degradation model developed in [8]. More in general, several challenges can be expected in the future to determine efficient and sustainable usage of wireless terminals, and the contributions made in the present paper represent a step forward in this direction. The integration of realistic modeling considerations and advanced optimization techniques can therefore have important consequences on the joint design of batteries, network elements, and control and actuation policies.

REFERENCES

[1] N. Michelusi, K. Stamatiou, L. Badia, and M. Zorzi, "Operation policies for energy harvesting devices with imperfect state-of-charge knowledge," in *Proc. IEEE ICC*, 2012, pp. 5782–5787.

[2] N. Michelusi, L. Badia, R. Carli, K. Stamatiou, and M. Zorzi, "Correlated energy generation and imperfect state-of-charge knowledge in energy harvesting devices," in *Proc. IEEE International Wireless Communications and Mobile Computing Conference (IWCMC)*, 2012, pp. 401–406.

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

[4] C. Renner and V. Turau, "CapLibrate: self-calibration of an energy harvesting power supply with supercapacitors," in *Proc. Int. conf. on Architecture of Comput. Syst. (ARCS)*, Hannover, Germany, Feb. 2010, pp. 1–10.

[5] P. Casari, A. P. Castellani, A. Cenedese, C. Lora, M. Rossi, L. Schenato, and M. Zorzi, "The wireless sensor networks for city-wide ambient intelligence (WISE-WAI) project," *Sensors*, vol. 9, no. 6, pp. 4056–4082, May 2009.

[6] K. Lin, J. Yu, J. Hsu, S. Zahedi, D. Lee, J. Friedman, A. Kansal, V. Raghunathan, and M. Srivastava, "Heliomote: enabling long-lived sensor networks through solar energy harvesting," in *Proc. ACM SenSys*, 2005.

[7] K. Kar, A. Krishnamurthy, and N. Jaggi, "Dynamic node activation in networks of rechargeable sensors," *IEEE/ACM Transactions on Networking*, vol. 14, pp. 15–26, Feb. 2006.

[8] N. Michelusi, L. Badia, R. Carli, L. Corradini, and M. Zorzi, "Energy Management Policies for Harvesting-Based Wireless Sensor Devices with Battery Degradation," *IEEE Transactions on Communications*, vol. 61, no. 12, pp. 4934–4947, Dec. 2013.

[9] D. Niyato, E. Hossain, and A. Fallahi, "Sleep and wakeup strategies in solar-powered wireless sensor/mesh networks: performance analysis and optimization," *IEEE Trans. Mobile Comput.*, vol. 6, pp. 221–236, Feb. 2007.

[10] A. Kansal, J. Hsu, S. Zahedi, and M. B. Srivastava, "Power management in energy harvesting sensor networks," *ACM Transactions on Embedded Computing Systems*, vol. 6, no. 4, Sep. 2007. [Online]. Available: http://doi.acm.org/10.1145/1274858.1274870

[11] M. Gatzianas, L. Georgiadis, and L. Tassiulas, "Control of wireless networks with rechargeable batteries," *IEEE Transactions on Wireless Communications*, vol. 9, no. 2, pp. 581–593, Feb. 2010.

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

[13] N. Michelusi, K. Stamatiou, and M. Zorzi, "Transmission policies for energy harvesting sensors with time-correlated energy supply," *IEEE Transactions on Communications*, vol. 61, no. 7, pp. 2988–3001, July 2013.

[14] O. Ozel, K. Tutuncuoglu, J. Yang, S. Ulukus, and A. Yener, "Transmission with Energy Harvesting Nodes in Fading Wireless Channels: Optimal Policies," *IEEE Journal on Selected Areas in Communications*, vol. 29, no. 8, pp. 1732–1743, Aug. 2011.

[15] B. Xiao, Y. Shi, and L. He, "A universal state-of-charge algorithm for batteries," in *ACM/IEEE Design Automation Conference (DAC)*, Anaheim, CA, Jun. 2010, pp. 687–692.

[16] C. K. Ho, P. D. Khoa, and P. C. Ming, "Markovian models for harvested energy in wireless communications," in *Proc. IEEE Int. Conf. on Communication Systems (ICCS)*, 2010, pp. 311–315.

[17] J. Xu and R. Zhang, "Throughput Optimal Policies for Energy Harvesting Wireless Transmitters with Non-Ideal Circuit Power," *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 2, pp. 322–332, February 2014.

[18] O. Orhan, D. Gunduz, and E. Erkip, "Throughput maximization for an energy harvesting communication system with processing cost," in *IEEE Information Theory Workshop (ITW)*, Sept 2012, pp. 84–88.

[19] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. John Wiley & Sons, Inc., New York, 2006.

[20] E. Sondik, "The optimal control of partially observable Markov processes," Stanford University, Tech. Rep. AD0730503, May 1971.

[21] D. A. Levin, Y. Peres, and E. L. Wilmer, *Markov chains and mixing times*. American Mathematical Society, 2006.

[22] J. M. Porta, N. Vlassis, M. T. Spaan, and P. Poupart, "Point-based value iteration for continuous POMDPs," *Journal of Machine Learning Research*, vol. 7, 2006.

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

[24] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

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

[26] B. Devillers and D. Gunduz, "A general framework for the optimization of energy harvesting communication systems with battery imperfections," *Journal of Communications and Networks*, vol. 14, no. 2, pp. 130–139, 2012.

[27] R. Rajesh, V. Sharma, and P. Viswanath, "Information capacity of energy harvesting sensor nodes," in *Proceedings of the IEEE International Symposium on Information Theory (ISIT)*, 2011, pp. 2363–2367.

[28] S. T. Hung, D. C. Hopkins, and C. R. Mosling, "Extension of battery life via charge equalization control," *IEEE Trans. Ind. Electron.*, vol. 40, no. 1, pp. 96–104, Feb. 1993.

[29] K. Lahiri, S. Dey, D. Panigrahi, and A. Raghunathan, "Battery-driven system design: A new frontier in low power design," in *Proc. Asia and South Pacific Design Automation Conference*, 2002.