When Simplicity Meets Optimality: Efficient Transmission Power Control with Stochastic Energy Harvesting

Qingsi Wang and Mingyan Liu
EECS Department
University of Michigan, Ann Arbor
(qingsi, mingyan)@umich.edu

***Abstract*—We consider the optimal transmission power control of a single wireless node with stochastic energy harvesting and an infinite/saturated queue with the objective of maximizing a certain reward function, e.g., the total data rate. We develop simple control policies that achieve near optimal performance in the finite-horizon case with finite energy storage. The same policies are shown to be asymptotically optimal in the infinite-horizon case for sufficiently large energy storage. Such policies are typically difficult to directly obtain using a Markov Decision Process (MDP) formulation or through a dynamic programming framework due to the computational complexity. We relate our results to those obtained in the unsaturated regime, and highlight a type of threshold-based policies that is universally optimal.**

I. Introduction

Harvesting renewable energy has been one of the most exciting technological developments in recent years. In the context of wireless communication such technologies promise to significantly alleviate the energy constraint that low cost (mobile) wireless devices operate under and prolong their operational lifetime, through harvesting ambient light, heat, vibration, etc; see e.g. [1], [2], [3] for more examples and surveys. While such technologies will undoubtedly improve the continuous operability of future wireless networks [4], it is also important to note that the amount of renewable energy available to a device usually fluctuates due to environmental factors (e.g., in the case of wind and solar energy) with typically low replenishing rates. Consequently, for wireless networks with energy harvesting nodes, efficient energy management both in terms of charging and discharging with quality of service (QoS) provisioning remains a critical issue and challenge that has motivated extensive research in various application contexts.

Literature Overview. [5], [6], [7], [8] have investigated optimal transmission policies with a non-causal formulation where future energy arrivals are assumed known *a priori*, which results in deterministic optimization. In [5], two problems are addressed through transmission power control, namely the maximization of reward (throughput, utility) given a deadline and the minimization of transmission completion time given a total reward objective; the optimal control is characterized for both problems. [6] revisits the completion time minimization problem, incorporating a deterministic data session arrival process. In [7], the throughput maximization problem is addressed using a geometric framework under a similar setting. They have also solved the minimization problem with the interesting observation of its connection to the throughput maximization counterpart. The above works all assume a packetized energy arrival model with the exception of [8], where similar problems are solved assuming continuous energy arrival as well as time-varying battery capacity due to degradation and energy leakage. For throughput maximization it is generally assumed that the data queue is *saturated*, that is, the transmitter always has data to transmit.

Stochastic energy harvesting has also been considered in the literature, see e.g., [9], [10], [11] with similar objectives like reward maximization as in the non-causal formulation, along with efficient online algorithms. [9] presented first a shortest path characterization of the optimal policy for the finite-horizon non-causal problem, and then an online algorithm that guarantees a fraction of the optimal performance when the estimate of the energy replenishment rate has bounded deviations. In [10], the throughput optimal policy, i.e., the data queue stabilizing policy that achieves the maximum throughput, is studied for a single node using results from $G/G/1$ queues. In [11], the authors considered a rate maximization problem using an infinite-horizon Markov Decision Process (MDP) framework with discount; main results include certain monotonicity and threshold properties of the optimal policy.

Contribution. In this paper, we consider the optimal transmission power control of a single node with stochastic energy harvesting, assuming a saturated data queue, and we present simple yet near-optimal online algorithms.

- Compared to previous work on stochastic throughput maximization in the saturated regime, we develop simple and explicit control policies that achieve near-optimal performance in the finite-horizon case with a finite battery capacity, and that are asymptotically optimal in the infinite-horizon scenario with sufficiently large battery capacity; such policies are typically difficult to directly obtain from an MDP or dynamic programming framework

This work was partially supported by the NASA under grants NNX06AD47G and NNX09AE91G.

due to the curse of dimensionality.

- We relate our results to those that have been obtained previously in the unsaturated regime, and highlight a type of threshold policies that is universally optimal.

Organization. We proceed as follows. In Section II, we formulate the finite-horizon reward maximization problem with preliminary structural results, motivating a policy named $\boldsymbol{g}_\theta$, of which performance bounds are derived in Section III. We also consider the infinite-horizon reward maximization in Section III, and show that a limiting form of $\boldsymbol{g}_\theta$ is asymptotically optimal as the battery capacity grows. We show numerical results for the proposed heuristics in Section IV and conclude in Section V. For the remainder of this paper, most proofs are omitted due to the space limit.

II. Problem Formulation and Preliminaries

Consider a discrete-time energy packet arrival process $E_t, t = 0, 1, \dots$ that models the energy harvesting process of the node, and suppose that each packet assumes values in a compact space $\mathcal{S} \subset \mathbb{R}_+$ with an independent and identical distribution (i.i.d.) with a mean $\overline{E}$ over the sample space Ω. The arriving energy is stored in an on-board battery up to a capacity B. Assume that the data queue of the node is saturated, that is, the node always has data to transmit. At each time step t, given any sample path of the energy arrivals $e_s, s = 0, 1, \dots, t$, the node determines the transmission power p_t to use so as to collect an amount of reward $r_t = r(p_t)$, where r is a concave and increasing continuous function. Our objective is to maximize the total amount of reward while satisfying the energy constraint imposed by the energy harvesting process and the battery capacity. Formally, let X_t denote the energy available in the battery after the energy arrival at time t, which evolves as follows,

$$X_t = \min\{X_{t-1} - P_{t-1} + E_t, B\}, t = 1, 2, \dots, \tag{1}$$

$$X_0 = \min\{B_{\text{init}} + E_0, B\}, \tag{2}$$

where B_{init} is the initial level of the battery. The node determines the transmission power at time t as a function of its history of battery states and past actions. Formally, $P_t = g_t(X^t, P^{t-1})$ with $0 \le P_t \le X_t$ for each t, and the function g_t is called the power control rule of the node at time t. Denote by T a finite time horizon, and denote the control policy, the collection of functions before the horizon by $\boldsymbol{g} = (g_0, g_1, \dots, g_T)$. We then consider the following stochastic optimization:

$$\text{maximize} \quad J^{\boldsymbol{g}} = \mathbb{E}^{\boldsymbol{g}}\left\{\sum_{t=0}^{T} r(P_t)\right\}, \tag{3}$$

where we use the superscript to emphasize the dependence on the control policy. Before proceeding, we introduce an upper bound on the optimum of (3), which we will then use as a benchmark to evaluate the performance of a given policy in the rest of this paper. Suppose that a sample path of energy arrivals from $t = 0$ to $t = T$ is given *a priori*, namely $\{E_t(\omega)\}_{t=0}^T$ for some $\omega \in \Omega$. Since $\sum_{t=0}^T P_t(\omega) \le B_{\text{init}} + \sum_{t=0}^T E_t(\omega)$, the maximum achievable average reward over this sample path $\overline{r}(T, \omega)$ is upper bounded as follows,

$$\overline{r}(T, \omega) \le r\left(\frac{B_{\text{init}} + \sum_{t=0}^T E_t(\omega)}{T+1}\right),$$

due to the concavity of the reward function. We will call this the *deterministic non-causal bound* on the average reward, as the energy causality and capacity limit may be violated at some time instants in achieving this bound. Using the strong law of large numbers, we have $\limsup_{T\to\infty} \overline{r}(T) \le r(\overline{E})$ almost surely. As we shall see, $r(\overline{E})$ is an achievable upper bound on the average reward for the infinite-horizon reward maximization problem.

In principle, (3) can be solved through the following dynamic program [12],

$$\begin{aligned} V_T(x) &= \max_{0\le p\le x} r(p) \\ V_t(x) &= \max_{0\le p\le x} \{r(p) + \mathbb{E}[V_{t+1}(X_{t+1}) | X_t = x, P_t = p]\} \\ &\quad t = 0, 1, \dots, T-1, \end{aligned}$$

where $V_t(x)$ is the value function denoting the maximum reward achievable at time t under state x. Note first that $V_T(x) = r(x)$. Furthermore, X_t is a controlled Markov process, thus an optimal policy exists in the space of Markov policies, i.e., we can limit our attention to the set of policies $P_t = g_t(X_t)$ [12]. We proceed with the following preliminary results.

Lemma 1: The value function V_t is a concave and nondecreasing function. Moreover, V_t satisfies the following inequality:

$$V_t(x) \le \begin{cases} (T-t+1) r(\frac{x+(T-t)\overline{E}}{T-t+1}), & \text{if } x \ge \overline{E} \\ r(x) + (T-t) r(\overline{E}), & \text{if } x < \overline{E} \end{cases} \tag{4}$$

for all t.

Remark 1: As a byproduct, the proof of Lemma 1 would yield a feasible Markov policy $\boldsymbol{g}$, namely, $p_t = g_t(x_t) = \min\{\frac{x_t+(T-t)\overline{E}}{T-t+1}, x_t\}$ for all t. This policy is intuitively appealing with the following interpretation. At time t, this policy essentially selects the optimal transmission power subject to the current battery level, if future energy arrivals are constant given by its statistical mean. More generally, note first that similar results as Lemma 1 can be obtained for the case where the energy arrivals are independent but with non-identical distributions. In particular, let $\overline{E}_t$ be the mean value of E_t, and a similar argument would give rise to the policy $p_t = \min\{\frac{x_t + \sum_{s=t+1}^T \overline{E}_s}{T-t+1}, x_t\}$. This policy in fact coincides with the so called “directional water-filling algorithm” which has been shown to be optimal for the deterministic optimization given non-causal knowledge of all energy arrivals [5] (also found by [13] under the name “staircase water-filling algorithm”).

This observation suggests that this policy may perform quite well if the energy arrival process has a small variation. However, if the variation is large, due to the concavity of the objective function, instead of depleting the battery when specified by the above policy, it may be beneficial to reserve a portion of storage in case the energy arrivals drop significantly. Accordingly we can modify $\boldsymbol{g}$ as follows:

$$p_t = \min\left\{\frac{x_t + (T-t)\overline{E}}{T-t+1}, \theta x_t\right\} \text{ when } t < T, \text{ and } p_T = x_T,$$

where $0 < \theta \leq 1$. We will denote this policy by $\boldsymbol{g}_\theta$, and argue in the next section that $\boldsymbol{g}_\theta$ can achieve a near-optimal average reward.

III. Optimality Analysis

In this section, we first examine the policy $\boldsymbol{g}_\theta$ for the finite-horizon reward maximization problem. Through a sequence of analysis combined with heuristic arguments, we characterize its average reward, justifying the claim of its near-optimal performance. We proceed in two steps.

1) We consider first a simple policy called the θ-policy, $p_t = \theta x_t$ for $t = 0, \ldots, T-1$ and $p_T = x_T$, and show that the gap between the average reward of this rudimentary policy and the non-causal bound can be significantly closed by properly tuning θ.
2) We given reasons why $\boldsymbol{g}_\theta$ should outperform the θ-policy in many cases.

Though the measure of suboptimality is not explicit in the finite-horizon case, we show in the second part that the limiting form of $\boldsymbol{g}_\theta$ is in fact asymptotically optimal in the infinite-horizon case with increasing battery capacity.

A. Performance characterization of the θ-policy

Assume for the time being that the battery capacity is infinite, or alternatively, it is sufficiently large: $B > (T+1)E_{\max}$, where $E_{\max} = \sup \mathcal{S}$ is the largest possible value of a single energy arrival. The state X_t then evolves without the minimization operator. Assume zero initial battery level for simplicity, $B_{\text{init}} = 0$. Consider a sample path of energy arrivals, $\{e_t\}_{t=0}^T$. By the θ-policy we have

$$p_t = \sum_{s=0}^{t} \theta(1-\theta)^{t-s} e_s \text{ if } t < T; \quad p_T = \sum_{s=0}^{T} (1-\theta)^{T-s} e_s.$$

To make our argument concrete, we consider in the following the log reward function given by $r_t = r(p_t) = \log(1+p_t)$. Note that $\frac{1}{2}r(\cdot)$ is the discrete-time Gaussian channel capacity with unit-variance noise [14]. Let $\rho_{\min} = \min_t \rho_t$ where $\rho_t = e_t / \sum_{s=0}^T e_s$ and let $\hat{\rho}_t = \sum_{s=0}^t \frac{\theta(1-\theta)^{t-s}}{1-(1-\theta)^{t+1}} \cdot \rho_s$, *i.e.*, $\hat{\rho}_t$ is a weighted average of ρ_s's where $s \leq t$, and weights are concentrated on the last few terms if θ is large while more evenly distributed when θ is small. Also, let

$$\mathcal{T} = \left\{ t : \sum_{s=0}^{t} (\theta + (1-\theta)\mathbb{1}_{\{t=T\}})(1-\theta)^{t-s}\rho_s < \frac{1}{T+1} \right\},$$

where $\mathbb{1}_{\{\cdot\}}$ is the indicator function of an event. Denote by Δ the per unit time difference between the deterministic non-causal bound and the average reward achieved by the θ-policy. It can be then shown that

$$\Delta \leq \frac{1}{T+1} \sum_{t \in \mathcal{T}} \left[\log\left(\frac{1 + (\theta-1)\mathbb{1}_{\{t=T\}}}{(T+1)\hat{\rho}_t} \right) + \frac{1}{\theta}(1-\theta)^{t+1} \right].$$

The above bound reveals how the parameter θ may be adjusted according to the property of the energy arrival process. If the arrival process has a small deviation, then $\hat{\rho}_t$ can be close to its statistical mean $1/(T+1)$ regardless of the choice of θ. In this case we can choose a large θ so that both terms would be small; this corresponds to the intuition of using most of the stored energy when one is fairly certain it would be replenished the next time instant. If on the other hand the arrival process has a large deviation, then we have a tradeoff: a small θ would keep $\hat{\rho}_t$ close to $1/(T+1)$ thereby keeping the first term small, while causing the second term to increase. This corresponds to the intuition that when there is greater uncertainty in energy arrival, we may need to "save for a rainy day".

We obtained the above result assuming that the battery has infinite capacity or $B > (T+1)E_{\max}$. These assumptions are however unrealistic, but using the proposition below we show that this assumption can be relaxed since the θ-policy in fact regularizes the battery level. Consider alternatively the scenario when the horizon is extended to the infinity when the θ-policy is applied and assume for the moment that $B > (T+1)E_{\max}$. Let $\tau_H := \min\{t : X_t < H\}$ for some constant $H > \frac{E_{\max}}{\theta}$. We then have the following result.

Proposition 1: Governed by the θ-policy, τ_H is finite. In particular, if $B_{\text{init}} < (\frac{1}{\theta} - 1)E_{\max}$, then $X_t < H$ for all t.

Remark 2: If the battery capacity satisfies $B > \frac{E_{\max}}{\theta}$, and the initial level is as in the lemma, we would never lose energy packets due to the battery capacity constraint. Moreover, the required capacity B under this setting is much smaller than $(T+1)E_{\max}$ that we supposed.

B. $\boldsymbol{g}_\theta$ versus θ-policy

The above bound suggests that the θ-policy could be practically efficient despite its seemingly naiveness. We have previously discussed why it may make sense to use the $\boldsymbol{g}_\theta$ policy rather than the θ-policy. In the rest of this section, we provide a heuristic argument for its performance advantage when $\theta = 1$. We first establish a property of deterministically bounded battery level under $\boldsymbol{g}_\theta$, which parallels Proposition 1. To this end, let $H' > (T+1)E_{\max} - T\overline{E}$ be some fixed number.

Proposition 2: Governed by $\boldsymbol{g}_\theta$ when $\theta = 1$, if $B_{\text{init}} < T(E_{\max} - \overline{E})$, then $X_t < H'$ for all t.

Remark 3: If $x_t \geq H'$, we have

$$x_{t+1} - x_t \leq -\frac{1}{T-t+1}H' - \frac{T-t}{T-t+1}\overline{E} + E_{\max} < 0.$$

Hence, if $B > H'$ and the initial state is greater than H', the battery state can be stabilized. Note also that the required battery capacity B under this setting is much smaller than $(T+1)E_{\max}$, if $\overline{E}$ is roughly of the same order as $E_{\max}$.

We present in the following a heuristic argument that $\boldsymbol{g}_\theta$ outperforms the θ-policy in many cases when $\theta = 1$. Assume that the battery capacity is sufficiently large as required in Propositions 1 and 2, i.e., $B > \max\{\frac{E_{\max}}{\theta}, (T+1)E_{\max} - T\overline{E}\}$, and assume that $B_{\text{init}} = 0$ for simplicity. Given the same sample path of energy arrivals, we compare the transmission power profiles over time under the θ-policy and $\boldsymbol{g}_\theta$. We first consider the trajectories of battery states under these two policies, which are denoted by x_t^θ and $x_t^{\boldsymbol{g}_\theta}$, respectively. We observe that whenever $x_{t_1}^{\boldsymbol{g}_\theta} \le \overline{E}$ at some t_1, $\boldsymbol{g}_\theta$ depletes the battery and x_t^θ and $x_t^{\boldsymbol{g}_\theta}$ start to coincide from t_1+1, resulting in the same transmission power and the same reward, until $x_{t_2}^\theta = x_{t_2}^{\boldsymbol{g}_\theta} > \overline{E}$ at some t_2. From t_2 on, x_t^θ and $x_t^{\boldsymbol{g}_\theta}$ diverge until they coincide again next time, which might not occur up to the end of the horizon T when both policies nevertheless deplete the battery. As a result, the comparison of the two power profiles reduces to a comparison over these divergent segments between their respective trajectories; an example is illustrated in Figure 1.

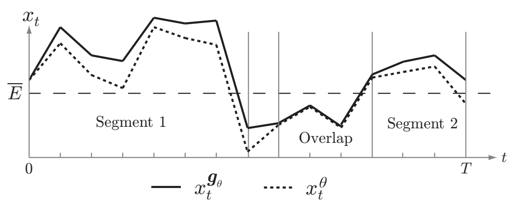

Fig. 1. The trajectories of $x_t^{\boldsymbol{g}_\theta}$ and x_t^θ.

All such segments including the last one ending at the horizon are defined as a period $[t_0, t_0+L]$ for some $L \ge 1$ such that the states under the two policies are only the same at t_0, i.e., $x_{t_0}^\theta = x_{t_0}^{\boldsymbol{g}_\theta}$, and moreover $x_t^{\boldsymbol{g}_\theta} > \overline{E}$ for all $t_0 \le t < t_0 + L$ and $p_{t_0+L} = x_{t_0+L}$ for both policies. Since the power profiles within such a segment are independent of the starting time t_0 for both policies ($\boldsymbol{g}_\theta$ only depends on the steps *to go* and the θ-policy is stationary), we will assume this segment of length $L+1$ starts at time 0 with an initial state e_0, and with a time to go T remaining in the horizon. We also note that $P_L^{\boldsymbol{g}_\theta} = X_L^{\boldsymbol{g}_\theta} \ge E_L = P_L^\theta$, that is, $\boldsymbol{g}_\theta$ always achieves a higher amount of reward than the θ-policy does at the final step of a segment. Therefore, we consider the first L steps of a segment. To simplify notation, we suppress the policy superscript whenever there is no ambiguity. Using $\boldsymbol{g}_\theta$, it can be show that

$$\boldsymbol{p} = (\Phi - \Phi\Psi + \Psi)\boldsymbol{e} + (\Phi\Psi - \Psi)\boldsymbol{\epsilon} =: Q\boldsymbol{e} + M\boldsymbol{\epsilon},$$

where

$$\Phi = \begin{bmatrix} \frac{1}{T+1} & 0 & \cdots & 0 \\ \frac{1}{T+1} & \frac{1}{T} & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ \frac{1}{T+1} & \frac{1}{T} & \cdots & \frac{1}{T-L+2} \end{bmatrix}, \Psi = \begin{bmatrix} \frac{1}{L} & \cdots & \frac{1}{L} \\ \vdots & \ddots & \\ \frac{1}{L} & \cdots & \frac{1}{L} \end{bmatrix},$$

$\boldsymbol{e} = (e_0, \ldots, e_{L-1})^{\mathrm{T}}$ and $\boldsymbol{p}$ similarly defined, $\epsilon = \mathbf{1}^{\mathrm{T}}\boldsymbol{e} - L\overline{E}$, and $\boldsymbol{\epsilon} = \begin{pmatrix} \mathbf{0} & \epsilon \end{pmatrix}^{\mathrm{T}}$. Assume that ϵ is small. Recall that $\boldsymbol{p}$ is the transmission power profile under $\boldsymbol{g}_\theta$ and $\boldsymbol{e}$ is the power profile under the θ-policy. We note that the sum reward function $\Gamma(\boldsymbol{p}) := \sum_{t=0}^{L-1} r(p_t)$ is a Schur concave function on $\mathbb{R}_+^L$, and hence $\Gamma(\boldsymbol{x}) \ge \Gamma(\boldsymbol{y})$ if $\boldsymbol{y}$ majorizes $\boldsymbol{x}$, which is equivalent to the existence of a doubly stochastic matrix Π such that $\boldsymbol{x} = \Pi\boldsymbol{y}$ (see [15] for more information). In our problem, however, Q is not doubly stochastic but nearly so. The sum of each row of Q is one, but the sum of the jth column is $1 + \frac{L+1-j}{T+2-j} - \frac{1}{L}(\frac{L}{T+1} + \frac{L-1}{T} + \cdots + \frac{1}{T-L+2})$, and the (i,j)th entry of Q is $\frac{1}{T+2-j} \cdot \mathbb{1}(i \ge j) + \frac{1}{L} - \frac{1}{L}(\frac{1}{T+1} + \frac{1}{T} + \cdots + \frac{1}{T+2-i})$. Hence, when T is much greater than L, i.e., this segment is far away from the end of the horizon, all elements are nonnegative, and the column sum approaches one as T increases. This suggests that when we are far away from the end, the $\boldsymbol{g}_\theta$ policy is at least as good as the θ policy. (Indeed in the next subsection we prove that the $\boldsymbol{g}_\theta$ policy is asymptotically optimal over an infinite horizon.) On the other hand, when L is close to T, while the column sum can be close to one, a small portion of entries in Q are negative with small values. This is a reflection of the boundary effect in a finite horizon problem when we get close to the end, and the advantage of the $\boldsymbol{g}_\theta$ policy cannot be as clearly ascertained.

C. Infinite-horizon case

In this part, we examine the infinite-horizon counterpart of the reward maximization problem with energy harvesting. Formally, for any power control policy $\boldsymbol{g} = (g_0, g_1, \ldots)$, where each g_t is similarly defined as in the finite-horizon case, we consider the following optimization problem

$$\text{maximize} \quad J^{\boldsymbol{g}} = \limsup_{T\to\infty} \mathbb{E}^{\boldsymbol{g}} \left\{ \frac{1}{T+1} \sum_{t=0}^{T} r(P_t) \right\}. \qquad (5)$$

We first note that as $T \to \infty$, the $\boldsymbol{g}_\theta$ policy that we studied in the finite-horizon case reduces to the stationary policy $p_t = g_\theta^\infty(x_t) = \min\{\overline{E}, \theta x_t\}$. We later show that when $\theta = 1$ the policy $\boldsymbol{g}_\theta^\infty = (g_\theta^\infty, g_\theta^\infty, \ldots)$ is asymptotically optimal as the battery capacity B increases.

In the rest of this part, we assume zero initial state of the battery for the simplicity of presentation, i.e., $B_{\text{init}} = 0$. Using Jensen's inequality, we have $\mathbb{E}^{\boldsymbol{g}}[r(P_t)] \le r(\mathbb{E}^{\boldsymbol{g}}[P_t])$ for any policy $\boldsymbol{g}$. Since $\sum_{t=0}^T P_t \le \sum_{t=0}^T E_t$ for any feasible policy $\boldsymbol{g}$ for any T, we have $\sum_{t=0}^T \mathbb{E}^{\boldsymbol{g}}[P_t] \le (T+1)\overline{E}$. Hence, combining the above results and further using the concavity of r, we have $\sum_{t=0}^T \mathbb{E}^{\boldsymbol{g}}[r(P_t)] \le (T+1)r(\overline{E})$, and therefore $J^{\boldsymbol{g}} \le r(\overline{E})$ for any policy $\boldsymbol{g}$. Fix $\epsilon > 0$ and let $\boldsymbol{g}_\epsilon = (g_\epsilon, g_\epsilon, \ldots)$ where $g_\epsilon(x_t) = \min\{\overline{E} - \epsilon, x_t\}$. We show in the following proposition that $\boldsymbol{g}_\epsilon$ approaches the neighborhood of $r(\overline{E})$.

Proposition 3: $J^{\boldsymbol{g}_\epsilon} \to r(\overline{E} - \epsilon)$ as capacity $B \to \infty$. In addition, $J^{\boldsymbol{g}_\epsilon}$ is continuous at infinity as a function of B.

Remark 4: It is worth noting that $\boldsymbol{g}_\epsilon$ coincides with the throughput optimal policy found in [10] under an unsaturated regime. Combining these results, it shows that this simple threshold policy is optimal in both the unsaturated and the saturated regimes. It remains an interesting open question whether similar results can be established with a more sophisticated energy harvesting and battery model.

IV. Numerical Results

In this section, we present numerical results for the algorithms we proposed for the finite-horizon reward maximization problem, and we consider three different distributions of the energy arrival E:

1) uniform distribution over $[0, 2\overline{E}]$;
2) triangle distribution, that is, $E = U_1 + U_2$ where U_1 and U_2 are i.i.d. and uniformly distributed over $[0, \overline{E}]$;
3) truncated Gaussian distribution, *i.e.*, $E = \max\{\min\{N, 2\overline{E}\}, 0\}$ where $N \sim \mathcal{N}(0, 1)$.

It is clearly that all distributions have the same mean value $\overline{E}$. Moreover, for our choice of $\overline{E}$ below, their variances follows a descending order. For each distribution, we generate 1000 sample paths of energy arrivals. For each sample path, we record the corresponding performance metrics of different policies and produce their sample means, respectively for the cases with an infinite and finite battery capacity. We compare the average total reward of both $\boldsymbol{g}_\theta$ and the θ-policy with two values of θ, against the non-causal bound (NCB) and a random policy (RP) defined as: $p_t = u_t x_t$ for $t < T$ and $p_T = x_T$ where u_t's are i.i.d samples generated uniformly over $[0, 1]$. Note that the random policy is included as an additional benchmark. As to the form of the reward function, we limit our attention to the rate function described in Section III. We set $\overline{E} = 10$, $B_{\text{init}} = 0$, $T = 100$ and $\epsilon = 0.1$, and the results are reported in Table I.

TABLE I
EMPIRICAL RESULTS.

			NCB	$\boldsymbol{g}_\theta$	θ-policy	RP
Total Reward $B = \infty$	Uniform	$\theta = 1$	242.21	**238.80**	222.13	211.24
		$\theta = .8$		**239.35**	230.58	
	Triangle	$\theta = 1$	242.04	**240.53**	233.44	214.68
		$\theta = .8$		**240.70**	236.75	
	Truncated Gaussian	$\theta = 1$	242.19	**242.12**	241.77	217.16
		$\theta = .8$		**242.07**	241.87	
$B = 20$	Uniform	$\theta = 1$	–	**227.95**	222.13	194.86
		$\theta = .8$		227.84	**229.41**	
	Triangle	$\theta = 1$	–	**235.85**	233.44	201.06
		$\theta = .8$		235.36	**236.54**	
	Truncated Gaussian	$\theta = 1$	–	**242.07**	241.77	206.19
		$\theta = .8$		**241.98**	241.87	

As can be seen, the proposed online algorithms provide efficient solutions to the respective problems, and the performance improves as the energy arrivals are more concentrated around their mean for both the infinite and the finite capacity cases. In the worst case of this numerical illustration, the best achieved total reward is no more than 5.9% below the NCB; similarly, the achieved completion time is no more than 6.5% above the NCB minimum. For sufficiently large battery capacities, $\boldsymbol{g}_\theta$ outperforms the θ-policy as we have argued. Interestingly, when the battery capacity is very limited, the θ-policy may be better in scenarios with larger deviations of energy arrivals. In such cases the choice of θ has non-negligible impact on the performance metric, especially for the θ-policy. Another remark is that when $\theta = 1$, in our setting we have $B \geq H = \frac{E_{\max}}{\theta}$ in both cases of the battery capacity for all testing distributions, and the θ-policy does result in the same performance regardless of the battery capacity as Proposition 1 suggests.

V. Concluding Remarks

In this paper we studied the transmission power control of a single node with stochastic energy harvesting, with simple yet efficient online algorithms presented. This work can be further pursued in the following two directions. An important assumption we have made in this work is the independence among energy arrivals, which is however unrealistic for most application scenarios. It is an interesting problem to investigate the optimal control in explicit forms with more realistic energy arrivals, e.g., that given by a Markovian model. We have also assumed that the battery has an infinite lifetime (or at least beyond the finite horizon). In practice a battery has limited lifetime and the efficient use of battery goes well beyond merely considering a capacity constraint: a battery typically has an optimal operating point depending on whether one wants to maximize the total number of cycles or the total energy output over its lifetime. These need to be taken into account in an optimal transmission control scheme when battery lifetime becomes a relevant issue.

References

[1] J. Paradiso and T. Starner, "Energy Scavenging for Mobile and Wireless Electronics," *Pervasive Computing, IEEE*, vol. 4, no. 1, pp. 18–27, 2005.
[2] S. Chalasani and J. Conrad, "A Survey of Energy Harvesting Sources for Embedded Systems," in *Southeastcon, 2008. IEEE*, pp. 442–447, Apr. 2008.
[3] S. Sudevalayam and P. Kulkarni, "Energy Harvesting Sensor Nodes: Survey and Implications," *Communications Surveys Tutorials, IEEE*, vol. 13, no. 3, pp. 443–461, 2011.
[4] X. Jiang, J. Polastre, and D. Culler, "Perpetual Environmentally Powered Sensor Networks," in *Information Processing in Sensor Networks, 2005. IPSN 2005. Fourth International Symposium on*, pp. 463–468, Apr. 2005.
[5] O. Ozel, K. Tutuncuoglu, J. Yang, S. Ulukus, and A. Yener, "Transmission with Energy Harvesting Nodes in Fading Wireless Channels: Optimal Policies," *Selected Areas in Communications, IEEE Journal on*, vol. 29, pp. 1732–1743, Sep. 2011.
[6] J. Yang and S. Ulukus, "Optimal Packet Scheduling in an Energy Harvesting Communication System," *Communications, IEEE Transactions on*, vol. 60, pp. 220–230, january 2012.
[7] K. Tutuncuoglu and A. Yener, "Optimum Transmission Policies for Battery Limited Energy Harvesting Nodes," *Wireless Communications, IEEE Transactions on*, vol. 11, pp. 1180–1189, Mar. 2012.
[8] B. Devillers and D. Gündüz, "A General Framework for the Optimization of Energy Harvesting Communication Systems with Battery Imperfections," *Arxiv preprint arXiv:1109.5490v2*, Mar. 2012.
[9] S. Chen, P. Sinha, N. Shroff, and C. Joo, "Finite-Horizon Energy Allocation and Routing Scheme in Rechargeable Sensor Networks," in *INFOCOM, 2011 Proceedings IEEE*, pp. 2273–2281, Apr. 2011.
[10] V. Sharma, U. Mukherji, V. Joseph, and S. Gupta, "Optimal Energy Management Policies for Energy Harvesting Sensor Nodes," *Wireless Communications, IEEE Transactions on*, vol. 9, pp. 1326–1336, Apr. 2010.
[11] A. Sinha and P. Chaporkar, "Optimal Power Allocation for a Renewable Energy Source," in *Communications (NCC), 2012 National Conference on*, pp. 1–5, Feb. 2012.
[12] P. R. Kumar and P. Varaiya, *Stochastic Systems: Estimation, Identification and Adaptive Control*. Prentice Hall, 1986.
[13] C. K. Ho and R. Zhang, "Optimal Energy Allocation for Wireless Communications with Energy Harvesting Constraints," *Arxiv preprint arXiv:1103.5290v2*, May 2012.
[14] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. John Wiley & Sons, Inc., 2001.
[15] A. W. Marshall, I. Olkin, and B. C. Arnold, *Inequalities: Theory of Majorization and Its Applications, Second Edition*. Springer Series in Statistics, 2011.