Optimal Energy Allocation in Multisensor Estimation Over Wireless Channels Using Energy Harvesting and Sharing

Steffi Knorn, Subhrakanti Dey, Anders Ahlén, and Daniel E. Quevedo

***Abstract*—We investigate the optimal power control for multisensor estimation of correlated random Gaussian sources. A group of wireless sensors obtains local measurements and transmits them to a remote fusion center (FC), which reconstructs the measurements using the minimum mean-square error estimator. All the sensors are equipped with an energy harvesting module and a transceiver unit for wireless, directed energy sharing between neighboring sensors. The sensor batteries are of finite storage capacity and prone to energy leakage. Our aim is to find optimal power control strategies, which determine the energies used to transmit data to the FC and shared between sensors, so as to minimize the long-term average distortion over an infinite horizon. We assume centralized causal information of the harvested energies and channel gains, which are generated by independent finite-state stationary Markov chains. The optimal power control policy is derived using a stochastic predictive control formulation. We also investigate the structure of the optimal solution, a Q-learning based suboptimal power control scheme and two computationally simple and easy-to-implement heuristic policies. Extensive numerical simulations illustrate the performance of the considered policies.**

***Index Terms*—Energy harvesting, energy sharing, fading, multisensor estimation, networks, power control, Q-learning.**

I. Introduction

Wireless sensors have become more powerful and affordable in recent years and are used in a growing number of areas, [1]–[4]. Often, several sensors are used to construct a wireless sensor network (WSN). Each sensor transmits its measurements wirelessly over a network to a remote fusion center (FC), which further processes the data, e.g., by reconstructing or analyzing the measured sources or computing an actuation signal. When using battery powered sensors, a significant challenge is to spend the available power in an optimal fashion, i.e., "power control" or "power management," [5]–[8].

Another promising alternative might be to harvest energy from the sensors' environment using, e.g., solar panels, windmills, thermoelectric elements, radio frequency harvesters, or vibration harvesters. However, since harvesting is an often unpredictable and unreliable power source, and rechargeable batteries have limited capacity, spending the available energy in an optimal fashion is a challenging task. Several optimal power control policies for different system settings with energy harvesting and optimizing a variety of performance criteria have been proposed. For example, power control policies were presented by [9] and [10] to maximize the throughput or minimize the mean delay or transmission completion time, respectively, and power control algorithms were derived to maximize the mutual information of a wireless link in [11]. An optimal packet scheduling problem for a single-user system with infinite battery and energy harvesting was investigated in [12]. The method to jointly control data queue and battery buffer to maximize the long-term average sensing rate of a WSN with energy harvesting was studied in [13]. The problem of designing optimal sensor transmission power control schemes under energy harvesting constraints was also investigated in [14].

An energy harvesting sensor that sends its measurements toward a remote estimator was considered in [15]. Also, a communication scheduling strategy for the sensor and an estimation strategy for the estimator, both of which that jointly minimize the expected sum of communication and distortion costs over a finite time horizon, was developed in [15]. A setting where sensor measurements are wirelessly sent over an unsure channel from an energy harvesting sensor to a remote estimator was investigated in [16]. This was extended to a closed control loop in [17].

Apart from energy harvesting, wireless energy transfer is another promising option to overcome the limitations of finite power resources since it allows harvested energy to be transferred and used in larger sensor networks, where all sensors might not be able to harvest sufficient amounts of energy at all times. It was experimentally demonstrated in [18] that energy can be transferred between two resonant objects with efficiencies more than 50% for distances up to 2 m. Similar energy transfer techniques were also discussed in [19]. Building on these results, the benefits of wireless energy transfer in wireless sensor systems were investigated in [20]–[25].

A significant hurdle while using batteries or capacitors in power wireless sensors, is the fact that these devices are not perfect. To address such issues, capacitor leakage aware algorithms for energy harvesting wireless devices were developed in [26]. The approach in [27] considered a single communication link with a hybrid power source including a constant energy supply and energy harvesting prone to energy leakage. A slightly different approach in [28] considered losses while saving harvested energy in the battery but lossless energy storing and energy retrieval from the battery.

A different line of research was conducted in [29] and [30], investigating a multisensor estimation problem. Wireless sensors report their measurements via a star network, over fading channels to a central FC, which reconstructs the random source observed by the sensors. All sensors are equipped with individual energy harvesting modules able to transfer energy via directed wireless links to neighboring sensors. Considering a finite time horizon, optimal power control policies for information transmission and energy sharing were derived to minimize the overall distortion at the FC. These results showed that energy shar-

Manuscript received September 16, 2018; accepted January 21, 2019. Date of publication January 29, 2019; date of current version September 25, 2019. This work was partially supported by Swedish Research Council Project under Grants 2017-04053 and 2017-04186. Recommended by Associate Editor G. Gu. *(Corresponding author: Steffi Knorn.)*

S. Knorn, S. Dey, and A. Ahlén are with the Department of Engineering Sciences, Uppsala University, Uppsala SE-75121, Sweden (e-mail: steffi.knorn@angstrom.uu.se; subhra.dey@angstrom.uu.se; anders.ahlen@angstrom.uu.se).

D. E. Quevedo is with the Department of Electrical Engineering (EIM-E), Paderborn University, Paderborn D-33098, Germany (e-mail: dquevedo@ieee.org).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TAC.2019.2896048

ing can be particularly beneficial (and potentially worth investing in) in case the harvesting and channel gain characteristics differ significantly between the sensors. However, implementing such finite time solutions is difficult. Optimal policies are time varying and require advance knowledge of the length of the time horizon of the application. Also, Knorn *et al.* [30] assumed perfect batteries and/or super-capacitors, independent and identically distributed (i.i.d.) channel gains and harvested energies and only a single point source. This paper considers more realistic scenarios by studying the more practically relevant case of infinite-time horizon power management. This leads to a *stationary* power control scheme, which can be implemented without knowing the run-time of the application *a priori*. Recalculating or adapting the policy is, hence, only necessary if the underlying statistics of the random processes change, which one assumes to be infrequent. The main contributions of this paper are as follows.

1) We investigate optimal power control schemes for information transmission and energy sharing in multisensor estimation of a spatially correlated random source vector, and minimize a long-term average distortion cost over an infinite horizon, with centralized causal information at the fusion center. We also consider Markovian fading channels and harvested energies and allow the sensor batteries/energy storage devices to be imperfect and subject to energy leakage.
2) The optimal stationary power control scheme is obtained by a stochastic control approach using a Markov decision process (MDP) formulation, where the optimal energy values for information transmission and sharing are found by solving a Bellman dynamic programming (DP) equation using *relative value iteration*; see [31]. Furthermore, some important structural properties of the optimal solution are established.
3) Motivated by practical limitations and on the basis of the structural properties, we show that the optimal choice of transmission energies is a simple threshold policy on the sensor battery level, provided that all other variables are fixed. We also consider a practical scenario where the exact statistical information of the underlying random processes may not be available, and present a Q-learning algorithm that yields a suboptimal solution to the power control problem at hand.

Section II presents the model, and Section III studies the infinite-horizon optimal power control problem. Section IV studies the structure of the optimal solution. Three suboptimal policies are proposed in Sections V and VI. The performances of the power control policies are compared using numerical examples presented in Section VII, followed by conclusion in Section VIII.

II. System Model

We consider a star-network with M sensors and an FC. Each sensor m individually measures a signal of interest, $\theta_m(k)$, at discrete-time instants, $k \in \{1, 2, 3, \dots\}$, subject to measurement noise. The measurements are spatially correlated between the sensors. The remote sensors transmit their information to the FC, which estimates the vector $\theta(k) = (\theta_1(k), \theta_2(k), \dots, \theta_M(k))^{\mathrm{T}}$, given the measurements received. We consider an analog amplify-and-forward uncoded transmission strategy subject to additive noise, [32]. Each sensor is equipped with a local battery/energy storage device, an energy harvester, and a unit to transmit and receive energy from other sensors, along with a transceiver for information transmission and reception, subject to transmission losses. A simple system is shown in Fig. 1.

A. Source Model and Sensor Measurements

We consider $\theta(k)$ to be an i.i.d., band-limited Gaussian process with zero mean. The measurements of the sensors are spatially correlated such that its covariance matrix (possibly nondiagonal) is $R_\theta = \mathbb{E}\{\theta(k)\theta^{\mathrm{T}}(k)\}$. We assume that $R_\theta > 0$ (positive definite). The measurements of sensor m, denoted by $x_m(k)$, are subject to measurement noise, $n_m(k)$, such that we have the following:

$$x_m(k) = \theta_m(k) + n_m(k) \tag{1}$$

where $1 \le m \le M$ and $k \ge 1$. The measurement noises, $n_m(k)$, are assumed to be i.i.d. Gaussian, mutually independent, and independent of $\theta(k)$ with zero mean and variances σ_m^2.

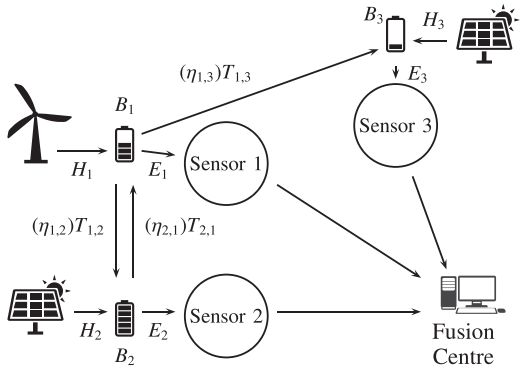

Fig. 1. Simple system with three sensors.

B. Energy Harvester, Energy Sharing, and Battery Dynamics

Each sensor is equipped with an energy harvester to gather energy from the environment. The harvested energy at sensor m at time k is denoted by $H_m(k)$ and is independent of the process, $\theta(k)$, and the measurement noise but may depend on $H_n(k)$ for $n \neq m$. The vector of harvested energies, $\mathbf{H}(k) = (H_1(k), \dots, H_M(k))$, is described as a first-order homogeneous finite-state irreducible and aperiodic Markov chain, motivated by empirical measurements reported in [33].[1] We further assume that the Markov chain is unichain, i.e., it has a single recurrent class and a possibly empty set of transient states; see [31]. We consider a slotted time model. For simplicity, each time-slot is assumed to be equal to the sampling period between two discrete sampling instants. The energy harvested at time slot k is stored in the battery, and can be used for data transmission to the FC or for energy sharing in time slot $k+1$. The energy used to transmit data from sensor m to the FC at time k is denoted by $E_m(k)$.

Each sensor can transmit energy to neighboring sensors and also receive energy from neighboring sensors via directed wireless energy transfer. This can be realized, for example, by energy transfer between two resonant objects such as discussed in [18] and [19], laser beams, or beamforming radio waves. The set of neighboring sensors from which sensor m can receive energy is denoted by $\mathcal{N}_{\mathrm{R},m}$, and the set of neighboring sensors to which sensor m can transmit energy is denoted by $\mathcal{N}_{\mathrm{T},m}$. The energy transferred from sensor m to sensor n at time k is denoted by $T_{m,n}(k)$. The efficiency of the energy transfer link from sensor m to sensor n, which accounts for losses in the wireless energy transfer process, is given by $\eta_{m,n} < 1$. In general, the efficiencies $\eta_{m,n}$ can be functions of time, i.e., $\eta_{m,n}(k)$. Here, we will assume time-invariant efficiencies.

Furthermore, we assume that during each time interval, some stored energy in the battery is lost because of leakage; see [26]. Thus, if no energy is added or used at time k, at time step $k+1$ only a fraction $\mu \in [0, 1]$ of the energy stored in the battery at time k is available for

[1] In case the harvested energies are mutually independent, each individual $H_m(k)$ would be described by an independent finite-state Markov chain.

use. Hence, using the notation mentioned above, the following are the dynamics of the battery level of sensor m at time $k+1$:

$$B_m(k+1) = \min\left\{\left(B_m(k) + H_m(k) - E_m(k) - \sum_{n\in\mathcal{N}_{\mathrm{T},m}} T_{m,n}(k) + \sum_{n\in\mathcal{N}_{\mathrm{R},m}} \eta_{n,m}\, T_{n,m}(k)\right)\mu;\quad B_m^{\max}\right\} \tag{2}$$

where $B_m^{\max}$ denotes the maximal battery capacity of sensor m.

C. Transmission Model

Each sensor has a transmitter using an analog amplify-and-forward uncoded strategy.[2] Hence, at each time-slot k, sensor m transmits its measurement, $x_m(k)$, amplified by a factor of $\sqrt{\alpha_m(k)}$. The energy needed for transmission is then given by $E_m(k) = \alpha_m(k)\left((R_\theta)_{m,m} + \sigma_m^2\right)$, where $(R_\theta)_{m,n}$ denotes element m,n of matrix R_θ. The channel power gain of the mth channel between sensor m and the FC is denoted by $g_m(k)$, and the vector of channel gains, $\mathbf{g}(k) = (g_1(k), \ldots, g_M(k))$, is assumed to be a first-order stationary and homogeneous finite-state Markov block-fading process. We assume that the Markov chain is unichain and that the channel gains are independent of the harvested energies, the process $\theta(k)$, and the measurement noises. We further assume that, within each block, the channel gains remain constant. For simplicity, the duration of each fading block is assumed to be the same as the duration of each transmission slot. We consider an orthogonal multiple access scheme between the sensors and the FC. The received signal at the FC from sensor m at time k is $z_m(k) = \sqrt{\alpha_m(k) g_m(k)}\, x_m(k) + \zeta_m(k)$ where $\zeta_m(k)$ is assumed to be an i.i.d. additive white Gaussian noise with variance ξ_m^2.

D. Distortion Measure at the Fusion Center

At the FC, the minimum mean-square error estimator, [34], provides the vector of estimates $\hat{\theta}(k) = \left(\hat{\theta}_1(k), \ldots, \hat{\theta}_M(k)\right)^{\mathrm{T}}$ given the vector of received signals $\mathbf{z}(k) = (z_1(k), \ldots, z_M(k))^{\mathrm{T}} = \mathbb{H}\theta(k) + \mathbf{v}(k)$ with $\mathbf{v} = \left(\sqrt{\alpha_1 g_1} n_1 + \zeta_1, \ldots, \sqrt{\alpha_M g_M} n_M + \zeta_M\right)^{\mathrm{T}}$ and $\mathbb{H} = \mathrm{diag}\left(\sqrt{\alpha_1 g_1}, \ldots, \sqrt{\alpha_M g_M}\right)$. So, the distortion is given by the following:

$$\begin{aligned} D(\mathbf{E}(k), \mathbf{g}(k)) &:= \mathrm{trace}\left(\mathbb{E}\left\{\left(\theta(k) - \hat{\theta}(k)\right)\left(\theta(k) - \hat{\theta}(k)\right)^{\mathrm{T}}\right\}\right) \\ &= \mathrm{trace}\left(\left(\mathbb{H}^{\mathrm{T}} R_{\mathrm{v}}^{-1} \mathbb{H} + R_\theta^{-1}\right)^{-1}\right) \end{aligned} \tag{3}$$

where $\mathbf{E}(k) = (E_1(k), \ldots, E_M(k))$ is the vector of transmission energies and $R_{\mathrm{v}} = \mathrm{diag}\left(\alpha_1 g_1 \sigma_1^2 + \xi_1^2, \ldots, \alpha_M g_M \sigma_M^2 + \xi_M^2\right)^{\mathrm{T}}$. The distortion is a random process as $\theta(k)$ is a random variable.

E. Information Patterns

We consider a *causal* information pattern using only information of current and past channel gains and harvested energies. Furthermore, we consider centralized information, where the FC has causal information of the channel gains, harvested energies, and battery levels of all sensors. This can be achieved in practice by the FC transmitting periodic pilot signals to the sensors at the beginning of each transmission slot, from which the sensors estimate their channels and report back their channel gains and previously harvested energies or current battery levels to the FC via orthogonal control channels. We assume the channels between the sensors and the FC are reciprocal, such as in a time-division-duplex framework. The FC computes the optimal power control schemes and informs the sensors at each slot.[3]

III. Infinite-Time Horizon Optimal Energy Allocation

In this section, we formulate an infinite-time horizon predictive control problem subject to the energy constraints in (2) to minimize the overall long-term average distortion (3) at the FC. It is considered that only causal information is available such that the information available at time $k \geq 1$ is $\mathcal{I}_k = \{\mathbf{g}(k), \mathbf{H}(k), \mathbf{B}(k), \mathcal{I}_{k-1}\}$, where $\mathbf{B}(k) = (B_1(k), \ldots, B_M(k))$ is the vector of battery levels, and $\mathcal{I}_1 = \{\mathbf{g}(1), \mathbf{H}(1), \mathbf{B}(1)\}$. The information $\mathcal{I}_k$ is used at each time k at the FC to decide the amount of the energy used for data transmission from the sensors to the FC, i.e., $E_m(k)$ for all $m = 1, \ldots, M$, and the amount of energy transferred between sensors, i.e., $T_{n,m}(k)$ for all $m = 1, \ldots, M$ and $n \in \mathcal{N}_{\mathrm{T},m}$. A power control policy is a set of functions to determine $(\{E_m(k)\}, \{T_{m,n}(k)\}) : m \in \{1, 2, \ldots, M\}$, and $n \in \mathcal{N}_{\mathrm{T},m}\}$. A policy is feasible if the following energy constraints:

$$E_m(k) \geq 0, \quad T_{m,n}(k) \geq 0, \quad E_m(k) + \sum_{n\in\mathcal{N}_{\mathrm{T},m}} T_{m,n}(k) \leq B_m(k) \tag{4}$$

are almost surely (a.s.) satisfied for all $1 \leq m, n \leq M$ and $k \geq 1$. The admissible control set is the set of all the possible power control policies, which are based only on $\mathcal{I}_k$ and do not violate the energy constraints in (4). For future reference, we define $\mathbf{T}(k)$ as the matrix with entries $(\mathbf{T}(k))_{m,n} = T_{m,n}(k)$ for $n \in \mathcal{N}_{\mathrm{T},m}$ and $(\mathbf{T}(k))_{m,n} = 0$ otherwise.

We aim to find the optimal power control scheme that minimizes the expected average distortion measure over an infinite-time horizon. The optimization problem is described as the following stochastic control problem: Find a power control policy, which determines $\mathbf{E}(k)$ and $\mathbf{T}(k)$, such that the following cost:

$$\limsup_{K\to\infty} \frac{1}{K} \sum_{k=1}^{K} \mathbb{E}\left\{D(\mathbf{E}(k), \mathbf{g}(k))\right\} \tag{5}$$

is minimized subject to (4) being satisfied a.s. for $1 \leq m, n \leq M$ and $1 \leq k \leq K$, and $B_m(k)$ satisfying (2).

The stochastic control problem in (5) with centralized information $\mathcal{I}_k$ can be regarded as an MDP formulation, $\{\mathcal{S}, \mathcal{A}, \mathcal{P}\}$, with state space, $\mathcal{S} = \{\mathbf{B}, \mathbf{g}, \mathbf{H}\}$, and action space, $\mathcal{A} = \{\mathbf{E}, \mathbf{T}\}$. The transition probability from state $\mathcal{S}$ to $\mathcal{S}'$ under action $\mathcal{A}$, i.e., $\mathcal{P}(\mathcal{S}'|\mathcal{S}, \mathcal{A})$, can be derived from the battery dynamics in (2) while considering the Markov chains describing the channel gains and harvested energies. See [31] and [35].

To simplify the notation, denote $\mathbf{g} = \mathbf{g}(k)$, $\mathbf{H} = \mathbf{H}(k)$, $\mathbf{B} = \mathbf{B}(k)$, $\mathbf{E} = \mathbf{E}(k)$, and $\mathbf{T} = \mathbf{T}(k)$, as well as $\tilde{\mathbf{g}} = \mathbf{g}(k+1)$, $\tilde{\mathbf{H}} = \mathbf{H}(k+1)$, and $\tilde{\mathbf{B}} = \mathbf{B}(k+1)$. Under the given assumptions, the existence of a stationary optimal power control policy computed offline from a Bellman DP equation follows.

[2]Optimality of analog transmission for multisensor estimation of a memoryless Gaussian source over a coherent multiaccess channel was shown in [32]. Furthermore, this scheme is very simple to implement since it does not require complex coding/decoding and incurs no delay other than propagation delay.

[3]The communication overhead between the sensors and the FC for reporting channel gains and battery levels also consumes energy at the sensors. This is not explicitly taken into account in this paper. However, if this energy consumption is constant for each transmission slot, it can be taken into account by subtracting it from the maximum battery level and defining a modified maximum battery level for each sensor.

Theorem 1: Suppose a unichain power control policy[4] exists and consider the average-cost optimality Bellman equation, as follows:

$$\rho+V(\mathbf{g},\mathbf{H},\mathbf{B})=\min_{\mathbf{E},\mathbf{T}}\left\{D(\mathbf{E},\mathbf{g})+\mathbb{E}\left\{V\left(\tilde{\mathbf{g}},\tilde{\mathbf{H}},\tilde{\mathbf{B}}\middle|\,\mathbf{g},\mathbf{H},\mathbf{E},\mathbf{T}\right)\right\}\right\} \tag{6}$$

where $\mathbf{E}$ and $\mathbf{T}$ satisfy (4), and V is the relative value function. Then, the infinite-time horizon stochastic control problem in (5) has a unique solution. Furthermore, if the set of possible policies includes at least one policy under which energy is used for data transmission or transferred to neighboring nodes, such that the associated Markov chain of battery levels is unichain, then the value of the infinite-time horizon stochastic control problem in (5) is given by ρ, which is the unique solution of (6). The optimal average cost, ρ, is independent of the initial conditions $\mathbf{g}(0)$, $\mathbf{H}(0)$, and $\mathbf{B}(0)$.

Proof: Since it is assumed that the Markov chains of the harvested energies and the channel gains are unichain and that a stationary unichain policy exists, it can be shown that (6) has a unique solution by following similar steps as in [36, Ch. 4.2, Proposition 2.5]. Then, by [36, Ch. 4.2, Proposition 2.6], the solution of (6) is independent of the initial state. ■

Remark 1: The stationary optimal solution to (5) is given by the following:

$$\begin{aligned}&\{\mathbf{E}^o(\mathbf{g},\mathbf{H},\mathbf{B}),\mathbf{T}^o(\mathbf{g},\mathbf{H},\mathbf{B})\}\\&=\arg\min_{\mathbf{E},\mathbf{T}}\left\{D(\mathbf{E},\mathbf{g})+\mathbb{E}\left[V(\tilde{\mathbf{g}},\tilde{\mathbf{H}},\tilde{\mathbf{B}})|\mathbf{g},\mathbf{H},\mathbf{E},\mathbf{T}\right]\right\}\end{aligned} \tag{7}$$

such that $\mathbf{E}$ and $\mathbf{T}$ satisfy the energy constraints in (4) with the battery dynamics in (2) for all m, and V constitutes the solution to the average cost Bellman equation in (6).

The Bellman equation in (6) can be solved using the relative value iteration algorithm; see [31]. In order to facilitate the numerical computation, the state and action spaces are discretized, in particular the battery levels and the power level space. It is expected that the solution of the discretized Bellman equation approaches the solution of the continuous-valued Bellman equation as the number of discretization levels grows [37].

IV. Structural Results of the Optimal Energy Allocation Policy

In this section, we investigate the structure of the optimal energy allocation solution. Given that V is convex in B (see [29], [34]), it will be shown that, if all other decision variables such as the shared energies $T_{m,n}$ for all m and n and the transmission energies E_n for all $n\neq m$ have been set, then the optimal transmission energy E_m is nondecreasing in B_m.

Theorem 2: Given $\mathbf{g},\mathbf{H},\mathbf{T}$ as well as B_n and E_n for all $n\neq m$, the optimal transmission energy allocation policy, E_m^o, is nondecreasing in B_m.

Proof: First, we define the right-hand side (RHS) of (7) for fixed $\mathbf{g},\mathbf{H},\mathbf{T}$ as well as B_n and E_n for all $n\neq m$, as follows:

$$L(E_m,B_m):=\arg\min_{0\leq E_m\leq B_m}\left\{D(\mathbf{E},\mathbf{g})+\mathbb{E}\left[V(\tilde{\mathbf{g}},\tilde{\mathbf{H}},\tilde{\mathbf{B}})|\mathbf{g},\mathbf{H},\mathbf{E},\mathbf{T}\right]\right\}. \tag{8}$$

Hence, the transmission energy of the single sensor m, i.e., E_m, is only constrained by the local battery level, B_m, and L in (8) is only a function of E_m and B_m.

[4]A unichain policy is a stationary policy under which the associated Markov chain has a single recurrent class, i.e., all states are visited an infinite number of times with probability 1.

To show that E_m^o is nondecreasing in B_m, it is sufficient to show that (8) is submodular in (E_m,B_m), as defined in [38], as $L(E_m',B_m')+L(E_m,B_m)\leq L(E_m,B_m')+L(E_m',B_m)$ for all $B_m'\geq B_m$ and $E_m'\geq E_m$. The first term of the RHS in (8) is independent of the battery levels, B, and thus submodular in (E_m,B_m). Define $\mathbf{x}=(x_1,x_2,\ldots,x_M)^{\mathrm{T}}$ with $x_m=B_m-E_m$ for all m, $\chi=(\chi_1,\chi_2,\ldots,\chi_M)^{\mathrm{T}}$ with $\chi_m=\min\{(B_m-E_m+H_m-\sum_{n\in\mathcal{N}_{\mathrm{T},m}}T_{m,n}+\sum_{n\in\mathcal{N}_{\mathrm{R},m}}\eta_{n,m}T_{n,m})\mu;B_m^{\max}\}$ for all m, and denote the last term in (8) as $Z(x_m):=\mathbb{E}\big[V(\tilde{\mathbf{g}},\tilde{\mathbf{H}},\chi|\mathbf{g},\mathbf{H},\mathbf{E},\mathbf{T}\big]$. Since V is convex in B_m as shown in [29, Lemma 5.1]; hence, $Z(x)$ is convex; this is equivalent to $Z(x+\epsilon)-Z(x)\leq Z(y+\epsilon)-Z(y)$ for all $x\leq y$ and $\epsilon\geq 0$. Setting $x=B_m-E_m'$, $y=B_m-E_m$, and $\epsilon=B_m'-B_m$, yields $Z(B_m'-E_m')-Z(B_m-E_m')\leq Z(B_m'-E_m)-Z(B_m-E_m)$, which shows the submodularity of L in (E_m,B_m). Submodularity implies that given $\mathbf{g},\mathbf{H},\mathbf{T}$, B_n, and E_n for all $n\neq m$, the optimal transmission energy, E_m, is, hence, a nondecreasing function of the battery level, B_m. ■

The abovementioned result is particularly useful for practical cases where the transmission energies, $\mathbf{E}$, at the power amplifier may only take values from a small finite set. In fact, a wireless sensor is often only able to transmit at a low power, E_{low} (or not at all, i.e., $E_{\mathrm{low}}=0$), or a high power, $E_{\mathrm{high}}>E_{\mathrm{low}}$. Then, it can be shown that for such cases, a threshold policy exists.

Corollary 1: Given $\mathbf{g},\mathbf{H},\mathbf{T}$ as well as B_n and E_n for all $n\neq m$, if E_m is restricted to take values from the finite set $E_m\in\{E_{\mathrm{low}},E_{\mathrm{high}}\}$, there exists a threshold $B_m^*>0$, such that the following holds true:

$$E_m^o(\mathbf{g},\mathbf{H},\mathbf{B})=\begin{cases}E_{\mathrm{low}} & \text{if } B<B_m^*\\ E_{\mathrm{high}} & \text{if } B\geq B_m^*\end{cases}. \tag{9}$$

Proof: The result follows directly from the fact that E_m^o is nondecreasing in B_m; see Theorem 2. ■

It should be noted that the above mentioned result is not only just limited to $E_m(k)$ taking binary values only. Indeed, in the case when $E_m(k)$ takes values from a larger discrete set, e.g., $E_m(k)\in\{0,E_{\mathrm{low}},E_{\mathrm{high}}\}$, two or more thresholds would exist instead of one to separate the energy levels.

Threshold-based policies for $E_m(k)$ taking values from a discrete set greatly reduce the search space to find the optimal policy. No analytical expression for the optimal thresholds exists, but several grid search techniques combined with stochastic-optimization based iterative algorithms can be used; see, e.g., [16] and especially the simultaneous perturbation stochastic optimization gradient algorithm in [29, Section V.A, Algorithm 1]. In particular, this algorithm can be applied to find locally optimal thresholds, B_m^*, with minor adaptations to the appropriate distortion cost function. Similarly, it can be also shown that there exists a threshold policy for the transferred energy, $T_{m,n}$, if all other variables are fixed.

V. Q-Learning

Solving the Bellman equation in (6) requires full knowledge of the underlying transition probability matrix, $\mathcal{P}$. In practice, the transition probabilities of the Markov process generating the channel gains and the harvested energies may not be perfectly known. In this case, the optimal power control cannot be determined by solving the Bellman DP equation presented in the previous section. Hence, finding suboptimal algorithms, which do not rely on complete knowledge of the underlying system, is an important task. In the case when state, $\mathcal{S}$, and action space, $\mathcal{A}$, are discrete or discretized (i.e., the channel gains, the harvested energies, the battery levels, and the allocated energy usage and energy transfer values belong to finite-discrete sets) and the fading channels

and harvested energies are independent finite-state Markov chains, the average-cost optimality Bellman equation (6) can be simplified to the Q-Bellman equation [39] as follows:

$$Q^*(\mathbf{g},\mathbf{H},\mathbf{B},\mathbf{E},\mathbf{T}) = D(\mathbf{E},\mathbf{g}) + \sum_{\tilde{\mathbf{g}},\tilde{\mathbf{H}},\tilde{\mathbf{B}}} \mathbb{P}(\tilde{\mathbf{g}}|\mathbf{g})\mathbb{P}(\tilde{\mathbf{H}}|\mathbf{H})\mathbb{P}(\tilde{\mathbf{B}}|\mathbf{B},\mathbf{H},\mathbf{E},\mathbf{T}) \min_{\tilde{\mathbf{E}},\tilde{\mathbf{T}}\in A(\tilde{\mathbf{B}})} Q^*(\tilde{\mathbf{g}},\tilde{\mathbf{H}},\tilde{\mathbf{B}},\tilde{\mathbf{E}},\tilde{\mathbf{T}}) \quad (10)$$

where $\tilde{\mathbf{E}}$ or $\tilde{\mathbf{T}}$ are the chosen values for $\mathbf{E}$ or $\mathbf{T}$ at the next time step, respectively, and $A(\tilde{\mathbf{B}})$ is the set of all feasible choices of $\tilde{\mathbf{E}}$ or $\tilde{\mathbf{T}}$ given $\tilde{\mathbf{B}}$. The iterative learning algorithm, which is referred to as Q-learning, approximates the average cost for a given set of states and actions, i.e., Q, by adjusting its value according to the recent observed cost, which is here the distortion denoted by D. The readers are referred to [39] and [40] for more details on the stochastic approximation Q-learning algorithm. Assuming that the probabilities $\mathbb{P}(\tilde{\mathbf{g}}|\mathbf{g})$, $\mathbb{P}(\tilde{\mathbf{H}}|\mathbf{H})$, and $\mathbb{P}(\tilde{\mathbf{B}}|\mathbf{B},\mathbf{H},\mathbf{E},\mathbf{T})$ are unknown, we obtain the following:

$$Q_1(\mathbf{g},\mathbf{H},\mathbf{B},\mathbf{E},\mathbf{T}) = 0 \quad \text{for all } \mathbf{g},\mathbf{H},\mathbf{B} \text{ and } \mathbf{E},\mathbf{T} \in A(\mathbf{B}) \quad (11)$$

and for all $k \geq 1$, we have the following:

$$Q_{k+1}(\mathbf{g},\mathbf{H},\mathbf{B},\mathbf{E},\mathbf{T}) = Q_k(\mathbf{g},\mathbf{H},\mathbf{B},\mathbf{E},\mathbf{T}) + \gamma(k)\left(D(\mathbf{E},\mathbf{g}) + \min_{\tilde{\mathbf{E}},\tilde{\mathbf{T}}\in A(\tilde{\mathbf{B}})} Q_k(\tilde{\mathbf{g}},\tilde{\mathbf{H}},\tilde{\mathbf{B}},\tilde{\mathbf{E}},\tilde{\mathbf{T}}) - Q_k(\mathbf{g},\mathbf{H},\mathbf{B},\mathbf{E},\mathbf{T})\right)$$

where $\{\tilde{\mathbf{g}},\tilde{\mathbf{H}},\tilde{\mathbf{B}},\tilde{\mathbf{E}},\tilde{\mathbf{T}}\}$ is the next state after $\{\mathbf{g},\mathbf{H},\mathbf{B},\mathbf{E},\mathbf{T}\}$ when $\mathbf{E},\mathbf{T} \in A(\mathbf{B})$ is selected according to the ϵ-greedy method as follows:

$$\{\mathbf{E},\mathbf{T}\} = \begin{cases} \arg\min_{\mathbf{E},\mathbf{T}\in A(\mathbf{B})} Q_k(\mathbf{g},\mathbf{H},\mathbf{B},\mathbf{E},\mathbf{T}) & \text{with prob. } 1-\epsilon \\ \text{chosen randomly } \in A(\mathbf{B}) & \text{with prob. } \epsilon \end{cases}.$$

The algorithm converges to the optimal Q values if the step sizes, $\gamma(k)$, for all $k \geq 1$ satisfy $\gamma(k) > 0$, $\sum_k \gamma(k) = \infty$, and $\sum_k \gamma^2(k) < \infty$, [39], [40]. Note that convergence is guaranteed for all $\epsilon > 0$. A small value of ϵ is usually preferred as it allows us to better exploit the knowledge regarding which choices of $\mathbf{E}$ and $\mathbf{T}$ lead to the minimal expected cost.

VI. Heuristic Policies

The proposed solutions to find power control policies by finding the optimal solution using (6) or solving the iterative learning algorithm (10), require a considerable computational effort and time. Even if the FC has sufficient energy resources, it may in practice be beneficial to find simple, suboptimal policies, which require less computational effort and time.

A. Heuristic 1: Modified Greedy Policy

A very simple policy is the greedy policy, where each sensor just uses all the available energy to transmit its data to the FC. Hence, $E_m(k) = B_m(k)$ for all m independent of the channel gain or any other states. While implementing this policy, there is a considerable risk of not having any energy available to transmit data from some sensor m to the FC at some time k if no energy has been harvested in the previous step. Thus, the greedy policy is slightly modified such that $E_m(k) = \frac{B_m(k)}{2}$, which ensures that at each time step, some energy is available to transmit data from every sensor to the FC, if $B_m(0) > 0$ $\forall m$.

B. Heuristic 2: Ad Hoc Policy

Inspired by our previous contribution [30], assume a simple system with two sensors, where the agents can share energy and have access to full causal information: the maximum battery level, mean channel gains and harvested energies, energy transfer efficiencies, and current channel gains and battery levels.[5] Aiming to minimize the overall distortion at the FC leads to the problem described in [30], for which necessary optimality conditions are derived. Those have to be simplified in order to reduce the computational complexity and to require only causal information. The simplified necessary conditions for using energy for data transmission to the FC ($E_1(k) \geq 0$), for storing energy in the battery for future use ($F_1(k) \geq 0$) and for transferring energy to sensor 2 ($T_{1,2}(k) \geq 0$), are as follows:

$$E_1(k) \geq 0 \quad \text{if } g_1(k) \geq \bar{g}_1 \text{ and } g_1(k) \geq \eta_{1,2}\bar{g}_2 \quad (12)$$

$$F_1(k) \geq 0 \quad \text{if } \bar{g}_1 \geq g_1(k) \text{ and } \bar{g}_1 \geq \eta_{1,2}\bar{g}_2 \quad (13)$$

$$T_{1,2}(k) \geq 0 \quad \text{if } \eta_{1,2}\bar{g}_2 \geq g_1(k) \text{ and } \eta_{1,2}\bar{g}_2 \geq \bar{g}_1. \quad (14)$$

In the case of unlimited battery capacity, these simplified necessary conditions could be used to allocate the energy at time step k. However, since both batteries have limited capacities, storing all energy at time k or transferring all energy from sensor 1 to sensor 2 at time k might be undesirable despite the necessary conditions (13) or (14) being satisfied, because it could lead to preventable battery overflow. Instead of determining the power control policy solely based on the necessary conditions, all three options (data transmission, storage, energy sharing) are prioritized, and energy is then allocated accordingly with the aim to minimize battery overflow as follows.

1) Denote the available power that is available at sensor 1 at time k by $\bar{B}_1 = B_1(k)$. Then, prioritize the three possible energy usage alternatives, i.e., data transmission $E_1(k)$, storage $F_1(k)$, and energy sharing $T_{1,2}(k)$, by sorting $g_1(k)$, $\bar{g}_1$, and $\eta_{1,2}\bar{g}_2$ from highest to lowest.[6] In the case when $g_1(k) = \bar{g}_1$ or $g_1(k) = \eta_{1,2}\bar{g}_2$, using energy for data transmission has higher priority than storing energy or transferring it to sensor 2, respectively. In the case when $\bar{g}_1 = \eta_{1,2}\bar{g}_2$, storing energy has higher priority than transferring it to sensor 2. Then, allocate $\bar{B}_1$ accordingly.

2) If transmitting data to the FC is the next highest priority, use all remaining energy to transmit data to the FC. (Thus, no energy is allocated to a task with a lower priority.)

3) If storing energy has the next highest priority, energy should be stored. To avoid battery overflow (i.e., energy waste), one should never store more energy than necessary to fill the battery to its maximal capacity minus the mean harvested energy: $F_1(k) = \min\left\{\max\left\{B_1^{\max}(k) - \bar{H}_1; 0\right\}; \bar{B}_1\right\}$. In the case there is more energy available in the battery than should be stored, the remaining energy should be used according to the next following priority, i.e., following the instructions in 2) or 4) and setting $\bar{B}_1 \to \bar{B}_1 - F_1(k)$.

4) If transferring energy to sensor 2 has the next highest priority, transfer as much energy to sensor 2 to have its battery full for the next time step. To avoid battery overflow, no more energy should be transferred than the battery capacity minus the mean harvested energy of sensor 2. Therefore, $T_{1,2}(k)$ for $\eta_{1,2} > 0$ is equal to $\min\left\{\max\left\{\left(B_2^{\max} - B_2(k) + E_2(k) - \bar{H}_2\right)/\eta_{1,2}; 0\right\}; \bar{B}_1\right\}$. If

[5]Note that in the case of Markovian channel gains or harvested energies, the mean channel gains, $\bar{g}_1$ and $\bar{g}_2$, and the mean harvested energies, $\bar{H}_1$ and $\bar{H}_2$, are calculated as the dot product of the channel gain levels or harvested energy levels, respectively, and the corresponding stationary distribution.

[6]For example, if $\bar{g}_1 > g_1(k) > \eta_{1,2}\bar{g}_2$, storing energy has the highest priority followed by data transmission to the FC, and transferring energy to the second sensor has the lowest priority.

$\eta_{1,2} = 0$, then $T_{1,2}(k) = 0$. In the case there is more energy in the battery than should be transferred, the remaining energy should be used according to the next following priority, i.e., following 2) or 3) and setting $\bar{B}_1 \to \bar{B}_1 - T_{1,2}(k)$.

Remark 2: This heuristic policy favors transmitting data to the FC if the current channel gain is higher than the mean, because then it is beneficial to minimize the overall distortion by transmitting data whenever the channel gain is better than the mean. In contrast, if a lot of energy is available because of higher mean harvested energy, then increasing the energy for data transmission further in the case of high channel gains leads to only a small reduction of the distortion. It would be better to store energy to be able to transmit data at time steps with poorer channel gains. This simple policy cannot distinguish between these two fundamentally different scenarios. It is designed to work well for scenarios with overall little energy availability but maybe not be as good for scenarios with higher amounts of energy.

VII. Simulation Examples

Example 1 (Effect of cross correlation): A system with two sensors is simulated with $\eta_{1,2} = \eta_{2,1} = 0.8$, $\mu = 0$ (no leakage), $B_1^{\max} = B_2^{\max} = 4$ mWh, and $R_\theta = (1, \varphi; \varphi, 1)$, where φ is the cross correlation between measurements θ_1 and θ_2 and is varied between 0 and 0.9. The channel gains and harvested energies are modeled as 3-level discrete Markov chains with the common transition matrix $\mathbb{T} = [0.2, 0.3, 0.5; 0.3, 0.4, 0.3; 0.1, 0.2, 0.7]$. We consider the "balanced case," where the state space for g_1 and g_2 is $\{0, 0.5, 1\}$ and for H_1 and H_2 is $\{0, 1, 2\}$, and the "unbalanced case," where g_2 and H_1 are four times lower than g_1 and H_2, respectively.

To facilitate the implementation of the DP and the Q-learning algorithm, the space for the battery levels and the power levels for data transmission or energy transfer were quantized uniformly with 16 levels. The discretization of the decision variables leads to numerical inaccuracies, which can be addressed by averaging the results over a sufficiently long time span. The Q-learning algorithm was evaluated by using two different training time horizons, i.e., 10^4 and 10^6, respectively, and with $\epsilon = 0.1$. After calculating the corresponding Q-values for both the training horizons, the performance of the algorithms were evaluated for a given simulation time span by using the Q-values as a lookup table to determine the best choice of $\mathbf{E}$ and $\mathbf{T}$ without adapting Q-values further. Third, the heuristics described in Section VI were implemented.

The average distortion and the energy usages for a simulation time span of 10^4 time steps for the optimal solution ("DP"), the Q-learning algorithm with the training time horizons 10^4 and 10^6 ("Q1" and "Q2," respectively), and the heuristics ("h1" and "h2") are illustrated in Fig. 2 . Increasing the cross correlation term, φ, leads to an overall reduced distortion. As expected, the average distortion is the smallest for the optimal algorithm ("DP"). The performance of the Q-learning algorithm is quite poor if a short training time horizon of 10^4 time steps is used ("Q1") but improves for the training horizon 10^6 ("Q2"). Also, the modified greedy policy ("h1") performs almost as good as the optimal solution ("DP") for the balanced case, but the ad hoc heuristic ("h2") outperforms the modified greedy policy in the unbalanced case.

Example 2 (Effect of energy transfer efficiency): The system settings from Example 1 were modified as follows: $\varphi = 0.2$ and $\eta = \eta_{1,2} = \eta_{2,1}$ varies between 0 and 1. See Fig. 3. In the balanced case, the average distortion hardly decreases when increasing the energy transfer efficiency despite the increase in the average energy transferred between the sensors. In the unbalanced case, the average distortions obtained for the optimal solution ("DP") and the Q-learning ("Q2") decrease for higher η. Again, the modified greedy policy ("h1") is more suitable for the balanced case while the ad hoc heuristic ("h2") achieves better results in the unbalanced case.

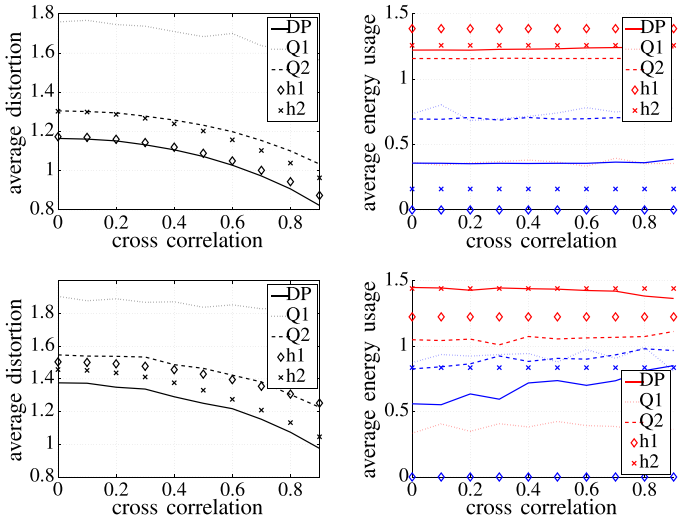

Fig. 2. Example 1: Average distortion (left) and average energy usage (right, $(E_1 + E_2)/2$ in red, $(T_{1,2} + T_{2,1})/2$ in blue), versus cross correlation term, φ, for the "balanced case" (top) and the "unbalanced case" (bottom).

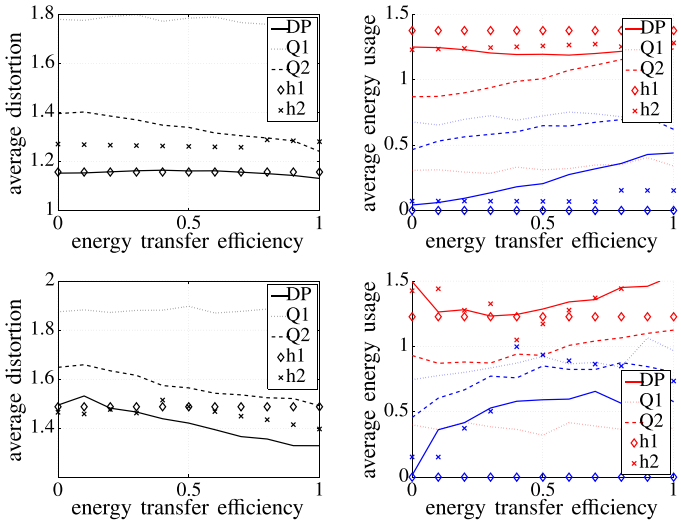

Fig. 3. Example 2: Distortion (left) and average energy usage (right, $(E_1 + E_2)/2$ in red, $(T_{1,2} + T_{2,1})/2$ in blue), versus energy transfer efficiency, η, for the "balanced case" (top) and the "unbalanced case" (bottom) for low cross correlation.

Example 3 (Effect of battery leakage): The system settings are similar to the examples mentioned above with $\varphi = 0.8$ and $\eta = 0.8$. The battery leakage parameter, μ, is varied between 0 (no leakage) and 0.5. The simulations in Fig. 4 show that a higher battery leakage parameter, μ, leads to an increase in the average distortion. It is also evident that energy sharing offers more benefits in the unbalanced case as compared to those in the balanced scenario. If the energy loss due to battery leakage increases, then the energy shared among the sensors approaches the average amount of energy used for data transmission. As in the examples mentioned above, the modified greedy policy ("h1") is outperformed by the ad hoc policy ("h2") in the case of unbalanced networks. In the case of balanced networks, the ad hoc heuristic ("h2") outperforms the modified greedy policy ("h1") for sufficiently high

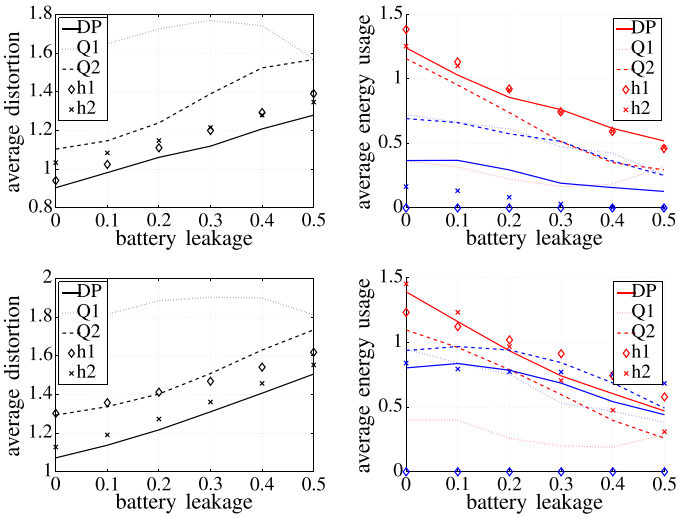

Fig. 4. Example 3: Distortion (left) and average energy usage (right, $(E_1 + E_2)/2$ in red, $(T_{1,2} + T_{2,1})/2$ in blue), versus battery leakage factor, μ, for the "balanced case" (top) and the "unbalanced case" (bottom).

battery leakage despite the ad hoc policy being developed for systems without battery leakage.

Using these simulations, it becomes clear that the optimal predictive power control scheme outperforms all suboptimal power control algorithms. Also, when considering the energy sharing between neighboring sensors by increasing the energy transfer efficiency, the overall distortion decreases, which indicates the usefulness of energy sharing. However, the extent of the reduction in the overall distortion while implementing the optimal power control solution compared to suboptimal schemes or while enabling wireless energy transfer, significantly depends on the system settings. If the system is balanced, little can be gained from applying the optimal power control or enabling wireless energy transfer. Implementing the simple modified greedy policy yields almost the same distortions as those yielded by the optimal solution. In unbalanced systems, the ad hoc heuristic outperforms the modified greedy policy.

VIII. Conclusion

This paper studied the distortion minimization problem of a multi-sensor system, where each sensor transmits its measurement to an FC over a fading channel for remote estimation at the FC. On the basis of causal information, the FC computes the optimal predictive power control policy to minimize a long-term average distortion cost, provided the following.

1) The batteries at the sensors have a limited capacity and are prone to energy leakage.
2) The sensors can harvest energy from their environment.
3) The sensors are able to wirelessly share energy with their neighbors subject to losses.

Harvested energies and channel gains are modeled as finite-state Markov chains.

The optimal solution is obtained using a stochastic predictive control approach, resulting in a Bellman DP equation. A suboptimal Q-learning algorithm, which does not require *a priori* knowledge of system parameters, is studied; two heuristic power control policies are also presented. Simulations reveal that the average distortion decreases as the cross correlation and the energy transfer efficiency increase. In most scenarios, the optimal solution clearly outperforms the suboptimal policies. It can be seen that an increase both in energy transfer efficiency and cross correlation have a significantly higher impact on the average distortion if the system is unbalanced, i.e., if one sensor has a substantially higher average harvested energy and a poorer channel compared to its neighbor.

The results in this paper reveal important insights into WSNs with energy harvesting and energy sharing. Even for simplistic network settings, the optimal energy allocation policy is far from trivial. Indeed, the findings presented here provide a benchmark for more complicated network topologies.

References

[1] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "A survey on sensor networks," *IEEE Commun. Mag.*, vol. 40, no. 8, pp. 102–114, Aug. 2002.
[2] X. Ren, J. Wu, K. H. Johansson, G. Shi, and L. Shi, "Infinite horizon optimal transmission power control for remote state estimation over fading channels," *IEEE Trans. Autom. Control*, vol. 63, no. 1, pp. 85–100, Jan. 2018.
[3] C.-Y. Chong and S. P. Kumar, "Sensor networks: Evolution, opportunities and challenges," *Proc. IEEE*, vol. 91, no. 8, pp. 1247–1256, Aug. 2003.
[4] J. Huang, D. Shi, and T. Chen, "Event-triggered state estimation with an energy harvesting sensor," *IEEE Trans. Autom. Control*, vol. 62, no. 9, pp. 4768–4775, Sep. 2017.
[5] D. E. Quevedo, A. Ahlén, and J. Ostergaard, "Energy efficient state estimation with wireless sensors through the use of predictive power control and coding," *IEEE Trans. Signal Process.*, vol. 58, no. 9, pp. 4811–4823, Sep. 2010.
[6] N. Pantazis and D. D. Vergados, "A survey on power control issues in wireless sensor networks," *IEEE Comm. Surv. Tuts.*, vol. 9, no. 4, pp. 86–107, Oct.–Dec. 2007.
[7] D. Han, P. Cheng, J. Chen, and L. Shi, "An online sensor power schedule for remote state estimation with communication energy constraint," *IEEE Trans. Autom. Control*, vol. 59, no. 7, pp. 1942–1947, Jul. 2014.
[8] K. Gatsis, A. Ribeiro, and G. J. Pappas, "Optimal power management in wireless control systems," *IEEE Trans. Autom. Control*, vol. 59, no. 6, pp. 1495–1510, Jun. 2014.
[9] V. Sharma, U. Mukherji, V. Joseph, and S. Gupta, "Optimal energy management policies for energy harvesting sensor nodes," *IEEE Trans. Wireless Commun.*, vol. 9, no. 4, pp. 1326–1336, Apr. 2010.
[10] K. Tutuncuoglu and A. Yener, "Optimum transmission policies for battery limited energy harvesting nodes," *IEEE Trans. Wireless Commun.*, vol. 11, no. 3, pp. 1180–1189, Mar. 2012.
[11] C. K. Ho and R. Zhang, "Optimal energy allocation for wireless communications with energy harvesting constraints," *IEEE Trans. Signal Process.*, vol. 60, no. 9, pp. 4808–4818, Sep. 2012.
[12] J. Yang, O. Ozel, and S. Ulukus, "Broadcasting with an energy harvesting rechargeable transmitter," *IEEE Trans. Wireless Commun.*, vol. 11, no. 2, pp. 571–583, Feb. 2012.
[13] Z. Mao, C. E. Koksal, and N. B. Schroff, "Near optimal power and rate control of multi-hop sensor networks with energy replenishment: Basic limitations with finite energy and data storage," *IEEE Trans. Autom. Control*, vol. 57, no. 4, pp. 815–829, Apr. 2012.
[14] Y. Li, F. Zhang, D. E. Quevedo, V. Lau, S. Dey, and L. Shi, "Power control of an energy harvesting sensor for remote state estimation," *IEEE Trans. Autom. Control*, vol. 62, no. 1, pp. 277–290, Jan. 2017.
[15] A. Nayyar, T. Başar, D. Teneketzis, and V. V. Veervalli, "Optimal strategies for communication and remote estimation with an energy harvesting sensor," *IEEE Trans. Autom. Control*, vol. 58, no. 9, pp. 2246–2260, Sep. 2013.
[16] M. Nourian, A. S. Leong, and S. Dey, "Optimal energy allocation for Kalman filtering over packet dropping links with imperfect acknowledgments and energy harvesting constraints," *IEEE Trans. Autom. Control*, vol. 59, no. 8, pp. 2128–2143, Aug. 2014.
[17] S. Knorn and S. Dey, "Optimal energy allocation for linear control with packet loss under energy harvesting constraints," *Automatica*, vol. 77, pp. 259–267, 2017.
[18] A. Kurs, A. Karalis, R. Moffatt, J. D. Joannopoulos, P. Fisher, and M. Soljačić, "Wireless power transfer via strongly coupled magnetic resonances," *Science*, vol. 317, no. 5834, pp. 83–86, 2007.
[19] A. Karalis, J. D. Joannopoulos, and M. Soljačić, "Efficient wireless non-radiative mid-range energy transfer," *Ann. Phys.*, vol. 323, no. 1, pp. 34–48, 2008.

[20] A. M. Fouladgar and O. Simeone, "On the transfer of information and energy in multi-user systems," *IEEE Commun. Lett.*, vol. 16, no. 11, pp. 1733–1736, Nov. 2012.
[21] R. Zhang and C. K. Ho, "MIMO broadcasting for simultaneous wireless information and power transfer," *IEEE Trans. Wireless Commun.*, vol. 12, no. 5, pp. 1989–2001, May 2013.
[22] L. Liu, R. Zhang, and K.-C. Chua, "Wireless information and power transfer: A dynamic power splitting approach," *IEEE Trans. Commun.*, vol. 61, no. 9, pp. 3990–4001, Sep. 2013.
[23] K. Huang and E. Larsson, "Simultaneous information and power transfer for broadband wireless systems," *IEEE Trans. Signal Process.*, vol. 61, no. 23, pp. 5972–5986, Dec. 2013.
[24] K. Huang and V. K. N. Lau, "Enabling wireless power transfer in cellular networks: Architecture, modeling and deployment," *IEEE Trans. Wireless Commun.*, vol. 13, no. 2, pp. 902–912, Feb. 2014.
[25] B. Gurakan, O. Ozel, J. Yang, and S. Ulukus, "Energy cooperation in energy harvesting communications," *IEEE Trans. Commun.*, vol. 61, no. 12, pp. 4884–4898, Dec. 2013.
[26] R. Shigeta, Y. Kawahara, and T. Asami, "Demo: Capacitor leakage aware duty cycle control for energy harvesting wireless sensor networks," in *Proc. 9th ACM Conf. Embedded Sen. Syst.*, 2011, pp. 387–388.
[27] I. Ahmed, A. Ikhlef, D. W. K. Ng, and R. Schober, "Power allocation for an energy harvesting transmitter with hybrid energy sources," *IEEE Trans. Wireless Commun.*, vol. 12, no. 12, pp. 6255–6267, Dec. 2013.
[28] K. Tutuncuoglu, A. Yener, and S. Ulukus, "Optimum policies for an energy harvesting transmitter under energy storage losses," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 3, pp. 467–481, Mar. 2015.
[29] M. Nourian, S. Dey, and A. Ahlén, "Distortion minimization in multi-sensor estimation with energy harvesting," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 3, pp. 524–539, Mar. 2015.
[30] S. Knorn, S. Dey, A. Ahlén, and D. E. Quevedo, "Distortion minimization in multi-sensor estimation using energy harvesting and energy sharing," *IEEE Trans. Signal Process.*, vol. 63, no. 11, pp. 2848–2863, Jun. 2015.
[31] D. P. Bertsekas, *Dynamic Programming and Optimal Control*, 3rd ed. Belmont, MA, USA: Athena Scientific, 1995, vol. 1.
[32] M. Gastpar, "Uncoded transmission is exactly optimal for a simple Gaussian "sensor" network," *IEEE Trans. Inf. Theory*, vol. 54, no. 11, pp. 5247–5251, Nov. 2008.
[33] C. K. Ho, P. D. Khoa, and P. C. Ming, "Markovian models for harvested energy in wireless communications," in *Proc. IEEE Int. Conf. Commun. Syst.*, 2010, pp. 311–315.
[34] I. Bahceci and A. K. Khandani, "Linear estimation of correlated data in wireless sensor networks with optimum power allocation and analog modulation," *IEEE Trans. Commun.*, vol. 56, no. 7, pp. 1146–1156, Jul. 2008.
[35] E. Altman, *Constrained Markov Decision Processes*. Boca Raton, FL, USA: CRC Press, 1999.
[36] D. P. Bertsekas, *Dynamic Programming and Optimal Control*, 3rd ed. Belmont, MA, USA: Athena Scientific, 1995, vol. 2.
[37] H. Yu and D. P. Bertsekas, "Discretized approximations for POMDP with average cost," in *Proc. 20th Conf. Uncertainty Artif. Intell.*, 2004, pp. 619–627.
[38] D. M. Topkis, *Supermodularity and Complementarity*. Princeton, NJ, USA: Princeton Univ. Press, 2001.
[39] R. Sutton and A. Barto, *Reinforcement Learning: An Introduction*. Cambridge, U.K.: Cambridge Univ. Press, 1998, vol. 1.
[40] K. J. Prabuchandran, S. K. Meena, and S. Bhatnagar, "Q-learning based energy management policies for a single sensor node with finite buffer," *IEEE Wireless Commun. Lett.*, vol. 2, no. 1, pp. 82–85, Feb. 2013.