Average Cost Distortion Minimization in Multi Sensor Estimation over Wireless Channels using Energy Harvesting and Energy Sharing

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

***Abstract*—In this paper, we investigate an optimal energy allocation problem for multi sensor estimation of correlated random Gaussian sources. A group of wireless sensors obtains a local measurement and transmits their measurements to a remote fusion centre (FC) via orthogonal fading wireless channels using uncoded analog transmissions. The vector of measurements is reconstructed at the FC using the minimum mean-square error (MMSE) estimator. All sensors are equipped with an energy harvesting module to gather energy from their environment to replenish their batteries or transmit data to the FC. Sensors are also fitted with a transceiver unit for sharing energy, which allows to transmit energy wirelessly between neighboring sensors in a directed fashion. The sensor batteries are of finite storage capacity and may be subject to energy leakage as well. Our aim is to find optimal energy allocation strategies, which determine the energies used to transmit data to the FC and shared between sensors, that 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. We derive these optimal policies using a stochastic control formulation, resulting in a Bellman dynamic programming equation. The requirement of full statistical information regarding the channel and harvested energy dynamics at the FC can be impractical, and we also investigate a Q-learning based sub-optimal energy allocation policy that does not need to know such statistical information a priori. In order to avoid the computational burden imposed by the curse of dimensionality associated with implementation of dynamic programming, we also investigate two computationally simple heuristic policies. All these energy allocation policies are explored and their performances compared via suitably chosen numerical examples.**

***Index Terms*—multi sensor estimation, energy harvesting, energy sharing, energy allocation, fading channels, Q-learning, networks**

I. Introduction

Wireless sensors have become much more powerful and affordable in recent years. Hence, they are used in a growing number of areas such as environmental data gathering [1], industrial process monitoring [2], mobile robots and autonomous vehicles [3], and for monitoring of smart electricity grids [4]. Often, several sensors are used to construct a wireless sensor network. Each sensor wirelessly transmits its measurements over a network to a remote fusion center (FC), which further processes the received data, e.g. by reconstructing the measured sources or computing an actuation signal.

Steffi Knorn, Subhrakanti Dey and Anders Ahlén are with the Department of Engineering Sciences, Uppsala University, Sweden; Email: {steffi.knorn, subhra.dey, anders.ahlen}@angstrom.uu.se

Daniel E. Quevedo is with the Department of Electrical Engineering (EIM-E), The University of Paderborn, Germany; Email: dquevedo@ieee.org

A significant challenge in this area is the powering of wireless sensors. Since some sensors cannot be connected to a reliable energy source or the connection to the power grid is not desirable, alternative power sources have to be considered. One such alternative are battery powered sensors. To maximize the lifetime and minimize the costs due to battery exchange it is crucial to spend the limited available energy stored in the battery in an optimal fashion. See for instance [5]–[7] and the references therein.

Another promising alternative might be to harvest energy. With appropriate hardware, such as e.g. solar panels, windmills, thermoelectric elements, radio frequency energy harvesting units or vibration harvesters, sensors can gather energy from their direct environment and hence potentially make battery changes obsolete. However, since harvesting is an often unpredictable and unreliable energy source and the rechargeable batteries have limited capacity, spending the available energy in an optimal fashion is a challenging task. Several optimal energy allocation policies for different system settings with energy harvesting and optimizing a variety of performance criteria have been proposed in the recent literature.

For instance, energy allocation policies to maximize the throughput or minimize the mean delay for a single transmitting device were presented in [8]. The authors of [9] derived optimal energy allocation policies that maximize the mutual information of a wireless link considering either causal or non-causal side information. An optimal packet scheduling problem for a single-user communication system with an infinite battery and energy harvesting considering, that data packets and energy packets arrive at the transmitter in a random manner, was investigated in [10]. The authors of [10] develop optimal off-line scheduling policies for minimizing the delivery time for all packets to the destination in a deterministic setting, where the energy harvesting times and the amounts of energy harvested are all known before transmission starts. In contrast to assuming a battery with infinite capacity as in [10], the work [11] studied the optimal off-line transmission policies with batteries with limited storage. Finite horizon throughput maximization and the related problem of minimizing the transmission completion time for a given amount of data were studied, [11]. These results were further generalized in [12] to include transmission over fading channels and optimal online policies. The effect of energy harvesting in optimal energy allocation for source acquisition/compression and transmission was studied in [13], [14]. Recently, optimal and suitable suboptimal energy allocation policies for a network of multiple sensors, which share a single energy harvester, have been studied in [15].

Apart from energy harvesting, wireless energy transfer is another promising option to overcome the limitations of finite energy resources due to finite battery capacities. Since wireless energy transfer is becoming more efficient and less costly, it can potentially be used to recharge batteries of wireless sensors. The authors of [16] showed through experiments that energy can be efficiently transferred between two resonant objects of the same resonant frequency; efficiencies of over 50% were achieved for distances up to 2 meters. By choosing different resonant frequencies between each pair, which are coupled by an energy transfer link, it is hence possible to allow for highly efficient energy transfer. Similar energy transfer techniques were also discussed in [17]. Another promising experiment conducted by Mitsubishi Heavy Industries demonstrated effective wireless energy transfer of 10kW over 500m, see [18].

Indeed, there is an increasing commercial interest in developing wireless energy transfer products, [19], [20], ranging from charging small devices such as cell phones in coffee shops [21] to charging electric vehicles [22]. It seems that it is merely a question of time when the application of wireless energy transfer becomes feasible in a wider range of technical areas [23]. Optimal design of energy and information transmission through wireless communication channels has also been of interest. Important works along this direction include [24]–[28], where the energy is assumed to be broadcasted in all directions, in contrast to the techniques discussed in [16], [17].

The important question of what benefits wireless energy transfer could bring to wireless sensor systems has already attracted some attention. The conference contributions [29] and [30] considered a wireless sensor network with a fixed base station and a wireless charging vehicle driving from sensor to sensor assuming wireless energy transfer as discussed in [16].

An optimal power allocation policy was derived and multiple necessary conditions for optimality are given for throughput maximization for a two-hop relay channel with one-way energy transfer from the source to the relay in [31]. The authors of [31] also investigated throughput maximization for a Gaussian two-way channel with one-way energy transfer. The optimal energy allocation policy is shown to be a directional two-way water filling algorithm, where one dimension relates to time while the second dimension describes the relationship between users.

A significant hurdle when using batteries, or other energy storage options such as capacitors, to 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 and successfully evaluated in [32], [33]. The work [34] considered a single communication link with a hybrid power source including a constant energy supply and energy harvesting prone to energy leakage. The authors in [35] studied throughput maximization of a single communication link, where the transmitter has full noncausal information of the fading channel gains and harvested energy but harvested energy is randomly lost. A slightly different approach was studied in [36], where saving harvested energy in the battery is assumed to be prone to losses whereas storing and retrieving the energy from the battery is considered lossless. In this situation, the policy that maximizes the communication rate, is found to be a double-threshold policy.

Following a different line of research, our recent work documented in [37] investigated a multi sensor estimation problem via a star network of wireless sensors that report their measurements over temporally independent block fading channels to a central FC, which reconstructs the random source observed by the sensors. All sensors are equipped with individual energy harvesting modules and can in addition transfer energy via directed wireless links to neighboring sensors. Optimal energy allocation policies for information transmission and energy sharing[1] were derived to minimize the overall distortion at the FC over a finite horizon using non-causal, causal centralized and causal local information at the sensors.

The present manuscript extends the results reported in [37] in several important ways: First, instead of assuming that all sensors measure the same source as in [37], the sensors in this paper are assumed to measure a field of correlated sources. Further, here, the fading channels and harvested energies are described by finite state Markov chains instead of assuming independent and identically distributed channel gains and harvested energies. In contrast to [37] and [38], we here consider sensor batteries/energy storage devices that may be prone to energy leakage. Also, this paper studies the infinite-time horizon case instead of the finite-time horizon considered in [37] and our conference contribution [38]. Choosing an infinite-time horizon approach has a significant advantage compared to implementing finite-time horizon solutions. Most networked sensor systems deployed for remote monitoring and estimation tasks are expected to operate over a long period of time. Thus, finding a stationary optimal energy allocation policy as a solution to an infinite horizon average distortion minimization problem is more practical. The solution is independent of the time horizon of application, can be implemented based on causal information only, and does not require recalculations as long as the statistics of the underlying random processes remain unchanged.

In particular, the current work contributes in the following ways:

1) We investigate optimal energy allocation policies for information transmission and energy sharing in a multi sensor estimation problem with a correlated field of data and minimizing a long-term average distortion cost over an infinite horizon, with centralized causal information and Markovian fading channels and harvested energies.
2) We allow the sensor batteries / energy storage devices to be imperfect and subject to energy leakage.
3) The optimal stationary energy allocation policy is found by a stochastic control approach using a Markov de-

[1]Note that the term ‘energy sharing’ in [37] refers to wireless energy transfer between neighboring sensor nodes. This is in contract to ‘energy sharing’ in [15], where multiple sensors have to share a single energy harvester.

cision process (MDP) formulation, where the optimal energy values for information transmission and sharing are found by solving a Bellman dynamic programming equation using *relative value iteration*, see [39].

4) Motivated by practical scenarios, where full statistical information about the harvested energy and fading channel dynamics may not be available, we present a Q-learning algorithm, that yields a suboptimal solution for the energy allocation policies without requiring exact knowledge of all system parameters.
5) We conduct a comparative performance investigation of the optimal solution obtained by using the relative value iteration algorithm with the suboptimal Q-learning algorithm and two simple heuristic policies via suitably chosen numerical examples, illustrating the advantages and disadvantages of each scheme. The benefits of energy sharing, and how the average distortion depends on various parameters such as cross correlation terms, energy transfer efficiency and energy leakage is discussed in detail.

The rest of the paper is organized as follows: The system model is presented in Section II. Section III studies the infinite-horizon optimal energy allocation problem. Three suboptimal energy allocation policies, namely Q-learning and two heuristic policies, are discussed in Sections IV and V, respectively. The performances of all considered energy allocation policies are compared by means of numerical examples presented in Section VI, followed by concluding remarks in Section VII.

II. System Model

We consider a star-network with M sensors and a FC. Each sensor m individually measures a signal of interest $\theta_m(k)$, at discrete-time instants $k \in \{1,2,3,\ldots\}$ 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 $\theta(k) = (\theta_1(k),\theta_2(k),\ldots,\theta_M(k))^{\mathrm{T}}$ given the measurements received. We consider an analog amplify and forward uncoded transmission strategy subject to additive noise. 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 scheme showing a simple system with three sensors is depicted in Fig. 1. The description of the individual parts is given below.

A. Source Model and Sensor Measurements

We consider $\theta_m(k)$ to be an independent and identically distributed (i.i.d.) (with respect to time) band-limited Gaussian process with zero mean. The measurements of the sensors are spatially correlated such that its covariance matrix (possibly non-diagonal) is $R_\theta = \mathbb{E}\left\{\theta\theta^{\mathrm{T}}\right\}$. We assume that $R_\theta > 0$ (positive definite). The measurements of sensor m, denoted $x_m(k)$, are subject to measurement noise, $n_m(k)$, such that

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

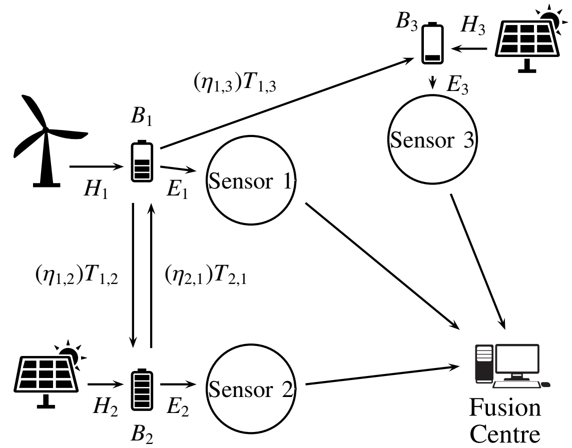

Figure 1: System setting (icons taken from [40])

for $1 \leq m \leq M$ and $k \geq 1$. The measurement noises $n_m(k)$ are assumed to be i.i.d. Gaussian, mutually independent and also independent of $\theta(k)$. Further, it is assumed that they have zero mean and variances σ_m^2.

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, denoted by $H_m(k)$, is described as a first-order homogeneous finite-state irreducible and aperiodic Markov chain, motivated by empirical measurements reported in [41]. We further assume that the Markov chain is unichain, that is, it has a single recurrent class and a possibly empty set of transient states. It is assumed that the harvested energies are mutually independent and independent from the process $\theta(k)$ and the measurement noise. 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 with neighboring sensors in time slot $k+1$. The energy used to transmit data from sensor m to the FC at time k is denoted $E_m(k)$. The transmission model will be described in detail in the next subsection.

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 instance, by energy transfer between two resonant objects such as discussed in [16], [17], the use of laser beams, or by the use of beamforming radiowaves. 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., $\eta_{m,n}(k)$. Unless explicitly mentioned, we will assume time-invariant efficiencies throughout this work.

Further, we assume that during each time interval, some stored energy in the battery is lost due to leakage, [32], [33]. 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 use. Hence, using the notation above, the dynamics of the battery level of sensor m at time $k+1$ is similar to the model in [34] and is given by

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

where $\hat{B}_m$ denotes the maximal battery capacity of sensor m. However, the energy storage model (2) differs from the model in [34] in two ways: (i) the model in [34] does not consider energy transfer, and (ii) the energy harvesting term is not affected by energy leakage. However, by simply rescaling the harvested energy levels accordingly, both models can be converted into each other when the energy transfer terms are ignored.

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

$$E_m(k) = \alpha_m(k)\left((R_\theta)_{m,m} + \sigma_m^2\right) \tag{3}$$

where $(R_\theta)_{m,n}$ denotes element m,n of matrix R_θ. The channel power gain of the m-th channel between sensor m and the FC, $g_m(k)$, is assumed to be a first-order stationary and homogeneous finite-state Markov block-fading process [43]. We assume that the channel gains are mutually independent and independent of the harvested energies. Similar to the harvested energies, we assume that the Markov chain is unichain. We further assume that within each block, the channel remains 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, which can be implemented for instance via orthogonal frequency division multiple access (OFDMA). 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 i.i.d. additive white Gaussian noise with variance ξ_m^2.

D. Distortion Measure at the Fusion Centre

At the FC, the minimum mean-square error (MMSE) estimator (see [44]) provides the vector of estimates $\hat{\theta}(k) = \left(\hat{\theta}_1(k),\hat{\theta}_2(k),\ldots,\hat{\theta}_M(k)\right)^{\mathrm{T}}$ 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 $\mathbb{H} = \mathrm{diag}\left(\sqrt{\alpha_1 g_1}, \sqrt{\alpha_2 g_2}, \ldots, \sqrt{\alpha_M g_M}\right)$ and $\mathbf{v} = \left(\sqrt{\alpha_1 g_1}n_1 + \zeta_1, \sqrt{\alpha_2 g_2}n_2 + \zeta_2, \ldots, \sqrt{\alpha_M g_M}n_M + \zeta_M\right)^{\mathrm{T}}$ (where we dropped the dependence on k for brevity).[3] Then, the distortion measure at the FC is

$$\begin{aligned} D(k) :=& \mathrm{trace}\left(\mathbb{E}\left\{\left(\theta(k) - \hat{\theta}(k)\right)\left(\theta(k) - \hat{\theta}(k)\right)^{\mathrm{T}}\right\}\right) \\ =& \mathrm{trace}\left(\left(\mathbb{H}^{\mathrm{T}} R_{\mathrm{v}}^{-1}\mathbb{H} + R_\theta^{-1}\right)^{-1}\right) \end{aligned} \tag{4}$$

where $R_{\mathrm{v}} = \mathrm{diag}\left(\alpha_1 g_1 \sigma_1^2 + \xi_1^2, \ldots, \alpha_M g_M \sigma_M^2 + \xi_M^2\right)^{\mathrm{T}}$.

E. Information Patterns

In this paper, we will consider a *causal* information pattern where only information of current and past channel gains and harvested energies is assumed. In particular, we consider centralized information, where the FC has causal information of all 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 (TDD) framework. The FC computes the optimal energy allocation policies and informs the sensors at each slot.[4]

III. Infinite-Time Horizon Optimal Energy Allocation

In this section, we formulate an infinite-time horizon optimal energy allocation problem subject to energy constraints (2) to minimize the overall long-term average distortion (4) at the FC. It is considered that only causal information is available. Hence, the unpredictable future wireless fading channel gains and harvested energies are not known a priori and the information available at time $k \geq 1$ is

$$\mathcal{I}_k = \{\mathbf{g}(k), \mathbf{H}(k), \mathbf{B}(k), \mathcal{I}_{k-1}\} \tag{5}$$

where $\mathbf{g}(k) = (g_1(k), g_2(k), \ldots, g_M(k))$ is the complete vectors of all channel gains, $\mathbf{H}(k) = (H_1(k), H_2(k), \ldots, H_M(k))$ is the vector of harvested energies and $\mathbf{B}(k) = (B_1(k), B_2(k), \ldots, B_M(k))$ is the vector of battery levels at time k, and $\mathcal{I}_1 = \{\mathbf{g}(1), \mathbf{H}(1), \mathbf{B}(1)\}$. The information $\mathcal{I}_k$ is used at each time slot k at the FC to decide the amount of energy used for data transmission from the sensors to the FC, i.e., $E_m(k)$ for all $m = 1,2,\ldots,M$, and the amount of energy transferred between sensors, i.e., $T_{n,m}(k)$ for all $m = 1,2,\ldots,M$ and $n \in \mathcal{N}_{\mathrm{T},m}$. An energy allocation 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 energy constraints

$$E_m(k) \geq 0, \quad T_{m,n}(k) \geq 0, \quad E_m(k) + \sum_{n\in\mathcal{N}_{\mathrm{T},m}} T_{m,n}(k) \leq B_m(k) \tag{6}$$

[2]Optimality of analog transmission for multi sensor estimation of a memoryless Gaussian source over a coherent multiaccess channel was shown in [42]. Further, this scheme is very simple to implement since it does not require complex coding/decoding, and incurs no other delay than propagation delay.

[3]It is also assumed that the sensor noise parameters σ_m, and the channel noise variances ξ_m are known at the FC.

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

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 possible energy allocation policies, which are based only on the causal information set $\mathcal{I}_k$ and do not violate the energy constraints (6). 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.

A. Infinite-Time Horizon Stochastic Control Problem

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

$$\limsup_{K\to\infty} \frac{1}{K} \sum_{k=1}^{K} \mathbb{E}\left\{D(k)\right\}, \tag{7}$$

(6) is satisfied a.s. for $1 \leq m,n \leq M$ and $1 \leq k \leq K$, and $B_m(k)$ satisfies (2).

B. Stationary Optimal Energy Allocation Policy

The stochastic control problem (7) with centralized information (5) can be regarded as a Markov Decision Process (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 (2) while considering the Markov chains describing the channel gains and harvested energies. See [39], [45] for further details.

To simplify notation, the vector of channel gains, harvested energies, battery levels and energy consumption and the matrix of energy shared at time k are denoted $\mathbf{g} = \mathbf{g}(k)$, $\mathbf{H} = \mathbf{H}(k)$, $\mathbf{B} = \mathbf{B}(k)$, $\mathbf{E} = \mathbf{E}(k)$ and $\mathbf{T} = \mathbf{T}(k)$, respectively, and the corresponding vectors of channel gains, harvested energies and battery levels at time $k+1$ are denoted $\tilde{\mathbf{g}} = \mathbf{g}(k+1)$, $\tilde{\mathbf{H}} = \mathbf{H}(k+1)$ and $\tilde{\mathbf{B}} = \mathbf{B}(k+1)$, respectively.

Under the given assumptions, one can show the existence of a stationary optimal energy allocation policy computed offline from a Bellman dynamic programming equation given in Theorem 1 below.

Theorem 1. *Suppose that a unichain energy allocation policy*[5] *exists. Then the infinite-time horizon stochastic control problem* (7) *has a unique solution.*

Further, if the set of possible policies includes at least one policy under which energy is used for data transmission or energy transfer 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 (7) *is given by* ρ*, which is the unique solution of the average-cost optimality Bellman equation*

$$\rho + V(\mathbf{g},\mathbf{H},\mathbf{B}) = \min_{\mathbf{E},\mathbf{T}} \left\{D + \mathbb{E}\left\{V\left(\tilde{\mathbf{g}},\tilde{\mathbf{H}},\tilde{\mathbf{B}}\middle|\, \mathbf{g},\mathbf{H},\mathbf{E},\mathbf{T}\right)\right\}\right\} \tag{8}$$

where $\mathbf{E}$ *and* $\mathbf{T}$ *satisfy the energy constraints given in* (6) *and* V *is the relative value function. The optimal average cost* ρ *is independent of the initial conditions* $\mathbf{g}(0)$, $\mathbf{H}(0)$ *and* $\mathbf{B}(0)$.

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

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 (8) has a unique solution by following similar steps as in [46, Chap. 4.2, Prop. 2.5]. Then, by [46, Chap. 4.2, Prop. 2.6], the solution of (8) is independent of the initial state. □

Remark 1. The stationary optimal solution to the stochastic control problem (7) is given by

$$\begin{aligned}&\{\mathbf{E}^o(\mathbf{g},\mathbf{H},\mathbf{B}),\mathbf{T}^o(\mathbf{g},\mathbf{H},\mathbf{B})\}\\ &= \operatorname*{argmin}_{\mathbf{E},\mathbf{T}} \left\{D + \mathbb{E}\left[V(\tilde{\mathbf{g}},\tilde{\mathbf{H}},\tilde{\mathbf{B}})|\mathbf{g},\mathbf{H},\mathbf{E},\mathbf{T}\right]\right\}\end{aligned} \tag{9}$$

such that $\mathbf{E}$ and $\mathbf{T}$, which satisfy the energy constraints (6) with battery dynamics (2) for all m, and V constitute the solution to the average cost Bellman equation (8).

Remark 2. If a control policy $\{\mathbf{E}^o,\mathbf{T}^o\}$, a measurable function V, and a constant ρ exist, which solve equations (8) and (9), then the control $\{\mathbf{E}^o,\mathbf{T}^o\}$ is optimal and ρ is the optimal cost

$$\rho = \limsup_{K\to\infty} \frac{1}{K} \sum_{k=1}^{K} \mathbb{E}\left\{D(k)\right\} \tag{10}$$

and, for any other feasible and causal control policy $\{\mathbf{E},\mathbf{T}\}$, we have

$$\rho \leq \limsup_{K\to\infty} \frac{1}{K} \sum_{k=1}^{K} \mathbb{E}\left\{D(k)\right\}. \tag{11}$$

More details can be found in [39].

Remark 3. Since the processes $\mathbf{g}$ and $\mathbf{H}$ are mutually independent (also across sensors) finite state Markov chains, the second right hand term involving the expectation in (8) becomes

$$\int_{\tilde{\mathbf{g}},\tilde{\mathbf{H}}} V\left(\tilde{\mathbf{g}},\tilde{\mathbf{H}},\tilde{\mathbf{B}}\right) \prod_{m=1}^{M} \left(\mathbb{P}(\tilde{g}_m|g_m)\mathbb{P}(\tilde{H}_m|H_m)\right) d\tilde{\mathbf{g}}d\tilde{\mathbf{H}} \tag{12}$$

where $\mathbb{P}(\mathbf{x}|\mathbf{y})$ is the probability of $\mathbf{x}$ given $\mathbf{y}$.

If the processes $\mathbf{g}$ and $\mathbf{H}$ are i.i.d. over time and across the sensors, then the same term in (8) simplifies to

$$\int_{\tilde{\mathbf{g}},\tilde{\mathbf{H}}} V\left(\tilde{\mathbf{g}},\tilde{\mathbf{H}},\tilde{\mathbf{B}}\right) \prod_{m=1}^{M} \left(\mathbb{P}(\tilde{g}_m)\mathbb{P}(\tilde{H}_m)\right) d\tilde{\mathbf{g}}d\tilde{\mathbf{H}}. \tag{13}$$

The Bellman equation (8) can be solved using the relative value iteration algorithm. Details can be found in [39]. In order to facilitate the numerical computation, the Bellman equation (8) is solved by discretizing the state and action space, in particular the battery levels and the energy allocation space. Recall that the state components involving the fading channels and the harvested energy levels are already assumed to be discrete due to the finite-state Markov chain assumption. 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 [47].

IV. Q-Learning

Solving the average-cost optimality Bellman equation (8) 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 energy allocation solution cannot be determined by solving the Bellman dynamic programming equation presented in the previous section. Hence, finding suboptimal algorithms, which do not rely on the complete knowledge of the underlying system, is an important task. In case the state, $\mathcal{S}$, and action space, $\mathcal{A}$, are discrete or discretized (that is, 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 (8) can be simplified to the Q-Bellman equation

$$Q^*(\mathbf{g},\mathbf{H},\mathbf{B},\mathbf{E},\mathbf{T}) = D + \sum_{\tilde{\mathbf{g}},\tilde{\mathbf{H}},\tilde{\mathbf{B}}} \mathbb{P}(\tilde{\mathbf{g}}|\mathbf{g})\mathbb{P}(\tilde{\mathbf{H}}|\mathbf{H})\mathbb{P}(\tilde{\mathbf{B}}|\mathbf{B},\mathbf{H},\mathbf{E},\mathbf{T}) \min_{\tilde{\mathbf{E}},\tilde{\mathbf{T}}\in A(\tilde{\mathbf{B}})} Q^*(\tilde{\mathbf{g}},\tilde{\mathbf{H}},\tilde{\mathbf{B}},\tilde{\mathbf{E}},\tilde{\mathbf{T}}) \tag{14}$$

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 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 D. See also [48] and [49], 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

$$Q_1(\mathbf{g},\mathbf{H},\mathbf{B},\mathbf{E},\mathbf{T}) = 0 \quad \forall \mathbf{g},\mathbf{H},\mathbf{B} \text{ and } \mathbf{E},\mathbf{T} \in A(\mathbf{B}) \tag{15}$$

and for all $k \geq 1$

$$Q_{k+1}(\mathbf{g},\mathbf{H},\mathbf{B},\mathbf{E},\mathbf{T}) = Q_k(\mathbf{g},\mathbf{H},\mathbf{B},\mathbf{E},\mathbf{T}) + \gamma(k)\left(D + \min_{\tilde{\mathbf{E}},\tilde{\mathbf{T}}\in A(\tilde{\mathbf{B}})} Q_k(\tilde{\mathbf{g}},\tilde{\mathbf{H}},\tilde{\mathbf{B}},\tilde{\mathbf{E}},\tilde{\mathbf{T}}) - Q_k(\mathbf{g},\mathbf{H},\mathbf{B},\mathbf{E},\mathbf{T})\right) \tag{16}$$

where now $\{\tilde{\mathbf{g}},\tilde{\mathbf{H}},\tilde{\mathbf{B}},\tilde{\mathbf{E}},\tilde{\mathbf{T}}\}$ 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:

$$\{\mathbf{E},\mathbf{T}\} = \begin{cases} \operatorname{argmin}_{\mathbf{E},\mathbf{T}\in A(\mathbf{B})} Q_k(\mathbf{g},\mathbf{H},\mathbf{B},\mathbf{E},\mathbf{T}) & \text{with prob. } 1-\epsilon \\ \text{chosen randomly } \in A(\mathbf{B}) & \text{with prob. } \epsilon \end{cases} \tag{17}$$

The algorithm in (16) 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 $\sum_k \gamma^2(k) < \infty$, [48], [49]. Note that convergence is guaranteed for all $\epsilon > 0$, [48], [49]. If ϵ is large, then the algorithm spends more computational effort in exploring the effect of possible choices of $\mathbf{E}$ and $\mathbf{T}$. However, a small value of ϵ is usually preferred as it often allows to better exploit the knowledge of which choice of $\mathbf{E}$ and $\mathbf{T}$ leads to the minimal expected cost based on the current Q_k.

V. Heuristic policies

The proposed solutions to find energy allocation policies in the two previous sections, i.e., finding the optimal solution via (8) or solving the iterative learning algorithm (14), require a considerable computational effort. In practice, it is often beneficial to investigate simple policies, that are providing suboptimal solutions, but require very little computational effort.

A. Heuristic 1: Modified greedy policy

A very simple policy is the greedy policy, where each sensor just uses all available energy to transmit its data to the FC. Hence, $E_m(k) = B_m(k)$ for all m independently of the channel gain or any other states. When implementing this policy, there is 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 the initial battery levels are not zero.

B. Heuristic 2: Ad hoc policy

The second heuristic policy was derived for a related but slightly different problem in [37], where instead of a correlated field, all sensors measure the same scalar signal of interest $\theta(k)$. We recapitulate the basic principles next.

Assume a simple system with two sensors, where both agents can share energy between each other and have access to full causal information, such as the maximal battery level, mean channel gains and harvested energies, energy transfer efficiencies as well as current channel gains and battery levels.[6] Aiming to minimise the overall distortion at the FC, leads to the problem described in [37], for which necessary optimality conditions is 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 ($F_1(k) \geq 0$)[7] and for transferring energy to sensor 2 ($T_{1,2}(k) \geq 0$) are as follows:

$$E_1(k) \geq 0 \quad \text{if } g_1(k) \geq \bar{g}_1 \text{ and } g_1(k) \geq \eta_{1,2}\bar{g}_2 \tag{18}$$

$$F_1(k) \geq 0 \quad \text{if } \bar{g}_1 \geq g_1(k) \text{ and } \bar{g}_1 \geq \eta_{1,2}\bar{g}_2 \tag{19}$$

$$T_{1,2}(k) \geq 0 \quad \text{if } \eta_{1,2}\bar{g}_2 \geq g_1(k) \text{ and } \eta_{1,2}\bar{g}_2 \geq \bar{g}_1 \tag{20}$$

In 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 (19) or (20) being satisfied because it could lead to preventable battery overflow. Instead of

[6]Note that in case of assuming 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.

[7]That is, $F_1(k)$ is the total amount of energy left in battery 1 for future use.

determining the energy allocation policy solely on the necessary condition, all three options (data transmission, storage, energy sharing) are prioritized and energy is then allocated accordingly with the aim to minimize battery overflow.

This suggests the following basic rules:

(i) 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.[8] In case $g_1(k) = \bar{g}_1$ 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 case $\bar{g}_1 = \eta_{1,2}\bar{g}_2$ storing energy has higher priority than transferring it to sensor 2. Then allocate the available energy according to these priorities.

(ii) 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.)

(iii) 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. That is

$$F_1(k) = \min\left\{\max\left\{\hat{B}_1(k) - \bar{H}_1; 0\right\}; B_1(k)\right\}.$$

In 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, that is, following the instructions in (ii) or (iv).

(iv) 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, again 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 given by

$$\min\left\{\max\left\{\left(\hat{B}_2 - B_2(k) + E_2(k) - \bar{H}_2\right)/\eta_{1,2}; 0\right\}; B_1(k)\right\}.$$

If $\eta_{1,2} = 0$, then $T_{1,2}(k) = 0$. In case there is more energy in the battery than should be transferred, the remaining energy should be used according to the next following priority, that is, following the instructions in (ii) or (iii).

Remark 4. The necessary conditions reported in [37], which lead to the heuristic algorithm given above, have been derived for a system without battery leakage, that is, with $\mu = 1$. However, when assuming little battery leakage, that is, μ close to 1, it can be expected that the heuristic policy can still be applied.

Remark 5. It should be noted that this heuristic policy favors transmitting data to the FC if the current channel gain is higher than the mean. This policy works well for cases where the overall amount of energy available is low. If only little energy is available, 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 already available due to higher mean harvested energy or higher battery capacity, increasing the energy for data transmission further in case of high channel gains leads to a small reduction of the distortion. In these cases it would be better to store energy to be able to transmit data at time steps with poorer channel gains. However, 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 its performance may not be as good when higher amounts of energy are available.

[8]For instance, 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.

Despite this heuristic policy being derived from optimality conditions in our earlier study for the different problem on decentralized estimation of a point source $\theta(k)$, the same simple rules for energy allocation will be evaluated for the case of individual measurements $\theta_m(k)$ and compared to other energy allocation policies below.

VI. Numerical Results

In this section, we provide a collection of numerical results that illustrate the performance of the optimal dynamic programming based algorithm, the Q-learning based algorithm and the two heuristic policies against various important parameters such as cross correlation, energy transfer efficiency and battery leakage.

Example 1 (Effect of Cross Correlation). A system with two sensors is simulated where $\eta_{1,2} = \eta_{2,1} = 0.8$, $\mu = 0$ (no battery leakage), $\hat{B}_1 = \hat{B}_2 = 4mWh$ and $R_\theta = (1,\varphi;\varphi,1)$, where φ describes the cross correlation between the two measurements θ_1 and θ_2 and is varied between 0 and 0.9.

The fading channel gains and harvested energies are modeled as 3-level discrete Markov chains with the common transition matrix

$$\mathbb{T} = \begin{bmatrix} 0.2 & 0.3 & 0.5 \\ 0.3 & 0.4 & 0.3 \\ 0.1 & 0.2 & 0.7 \end{bmatrix}. \tag{21}$$

Two cases have been simulated: In the 'balanced scenario', the state space for g_1, g_2 is {0,0.5,1} and for H_1 and H_2 is {0,1,2}. In the 'unbalanced scenario' g_2 and H_1 are 4 times lower than g_1 and H_2, respectively. That is, the state space for g_1 and g_2 are {0,0.5,1} and {0,0.125,0.25}, respectively, while the state spaces of H_1 and H_2 are {0,0.5,1} and {0,2,4}, respectively.

To facilitate the implementation of the dynamic programming algorithm and the Q-learning algorithm, the space for the battery levels and the space for energy allocation for data transmission or energy transfer to the neighboring sensor were quantized uniformly. Despite these discretizations, the dynamic programming based algorithm can be time-consuming for calculating the optimal energy allocation look-up tables, due to the well known *curse of dimensionality*. In addition, 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 the use of 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 training horizons, the performance of the algorithms were evaluated for a given simulation time span by using

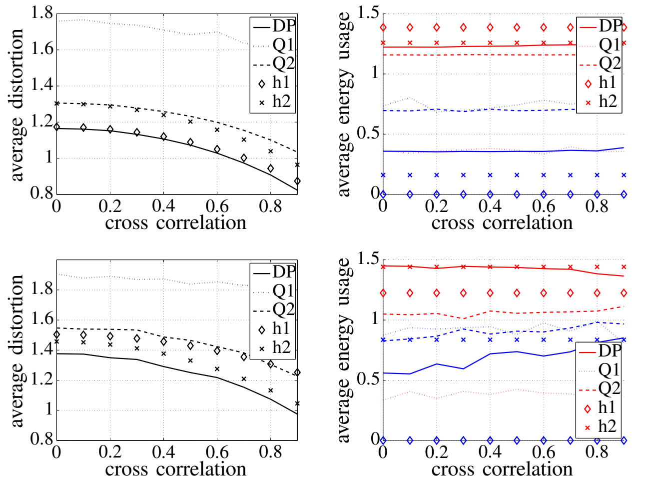

Figure 2: Example 1: average distortion (left) and average energy usage (right, $(E_1 + E_2)/2$ in red, $(T_{1,2} + T_{2,1})/2$ in blue), vs. cross correlation term φ for the 'balanced case' (top) and the 'unbalanced case' (bottom)

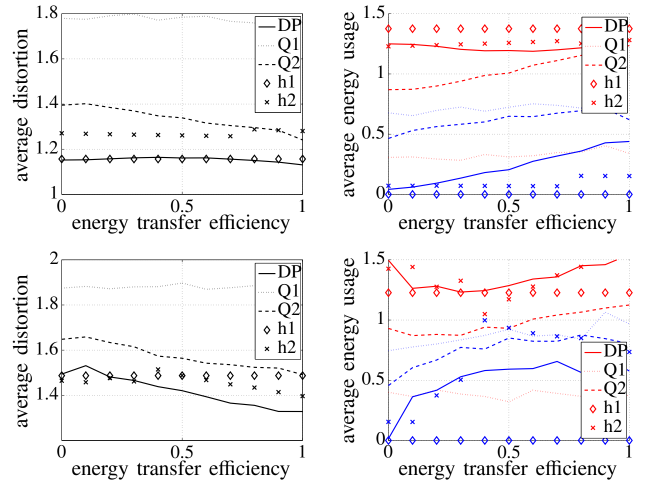

Figure 3: Example 2: distortion (left) and average energy usage (right, $(E_1 + E_2)/2$ in red, $(T_{1,2} + T_{2,1})/2$ in blue), vs. energy transfer efficiency η for the 'balanced case' (top) and the 'unbalanced case' (bottom) for low cross correlation

the Q-values as a look-up table to determine the best choice of $\mathbf{E}$ and $\mathbf{T}$ without adapting Q-values further. Third, the heuristics described in Section V were implemented.

The average distortion and the average energy usages for a simulation time span of 10^4 time steps for the optimal solution based on dynamic programming ('DP'), the Q-learning algorithm with the two different training time horizons 10^4 and 10^6 ('Q1' and 'Q2', respectively), and the two heuristics ('h1' and 'h2') are illustrated in the plots in Fig. 2.

It is evident that increasing the cross correlation term φ leads to an overall reduced distortion. As expected, the average distortion is the smallest for the optimal algorithm based on dynamic programming. The performance of the Q-learning algorithm is quite poor if a short training time horizon of 10^4 time steps is used ('Q1'). However, when increasing the training horizon to 10^6 ('Q2') the average distortion is significantly reduced since the optimal policy is better approximated. It is expected that the performance can be further improved using even longer training time horizons. Observe also, that the modified greedy policy ('h1') performs almost as good as the optimal solution ('DP') for the balanced case. In contrast, the ad hoc heuristic ('h2') derived for the related setting in [37] (every sensor measures the same θ) clearly outperforms the modified greedy policy in the unbalanced case.

Example 2 (Effect of Energy Transfer Efficiency for Low Cross Correlation). The system settings from Example 1 were modified in the following way: Instead of varying the cross correlation term, it is set to $\varphi = 0.2$ while $\eta = \eta_{1,2} = \eta_{2,1}$ is varied between 0 and 1.

The simulations are shown in Figure 3. In the balanced case, the average distortion hardly decreases when increasing the energy transfer efficiency despite the increase of 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. As in the previous example, 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. Note that the average distortion for the optimal solution slightly increases between $\eta = 0$ and $\eta = 0.1$. This can be explained by the loss of optimality due to discretization, which is necessary to implement the solution of the Bellman equation on a digital computer resulting in small deviations from the true optimal solution.

In the unbalanced case, it should also be noted that the optimal shared energy increases when the energy transfer efficiency increases from 0 to approximately 0.3. If the energy transfer efficiency is increased further, the optimal amount of energy shared among the sensors remains roughly the same. Since the measurements from the two sensors carry information about two different sources (although correlated) the FC needs to receive data from both sensors in order to estimate both sources. Hence, in the unbalanced case, one sensor needs to share some energy to allow the other sensor to transmit data that can be received at the FC with an acceptable quality. In case wireless energy transfer is possible with a sufficiently high efficiency (such that at least 30% of the transmitted energy is actually received at the receiving sensor), sharing more energy is not beneficial since the other sensor has enough energy already for information transmission with an acceptable distortion level at the FC.

It can also be observed that the curve of the average energy used for data transmission has a "bowl shaped" behavior: Due to the increase in average shared energy when increasing the energy transfer efficiency from 0 to 0.3, on average, less energy is available for data transmission to the FC. Hence, the average energy usage decreases for low energy transfer efficiencies. However, for higher energy transfer efficiencies, the amount of shared energy remains almost the same, leading to an increase in average available energy to be used for data transmission.

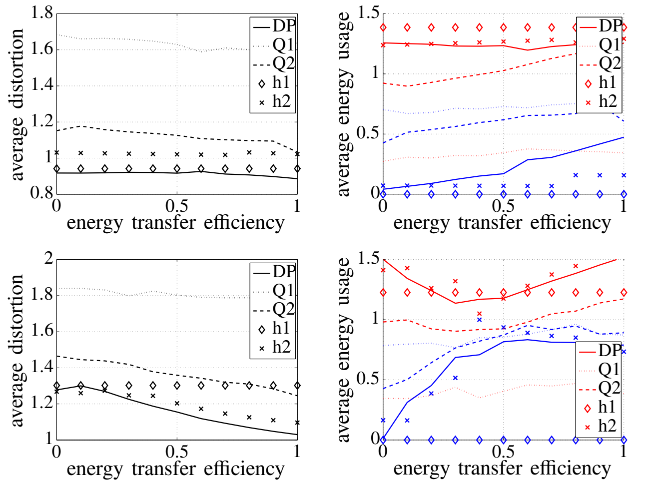

Figure 4: Example 3: distortion (left) and average energy usage (right, $(E_1 + E_2)/2$ in red, $(T_{1,2} + T_{2,1})/2$ in blue), vs. energy transfer efficiency η for the 'balanced case' (top) and the 'unbalanced case' (bottom) for high cross correlation

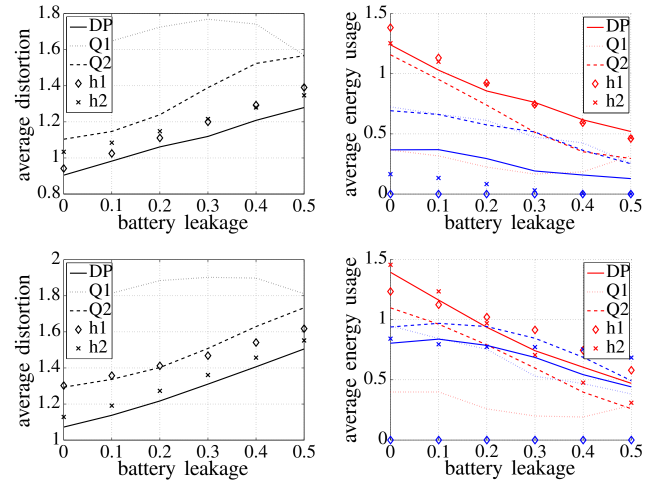

Figure 5: Example 4: distortion (left) and average energy usage (right, $(E_1 + E_2)/2$ in red, $(T_{1,2} + T_{2,1})/2$ in blue), vs. battery leakage factor μ for the 'balanced case' (top) and the 'unbalanced case' (bottom)

Example 3 (Effect of Energy Transfer Efficiency for High Cross Correlation). The system settings from Example 2 were solely modified by setting $\varphi = 0.8$.

The simulations in Figure 4 show similar results as in Example 2 for the case of low cross correlation ($\varphi = 0.2$). Due to the higher cross correlation, the average distortion is generally lower in Figure 4 compared to the results in Figure 3. Further, it seems that in case of higher cross correlation, energy transfer offers a higher benefit, since the average energy transferred between the sensors increases more for $\varphi = 0.8$ than for $\varphi = 0.2$.

Example 4 (Effect of Battery Leakage). Here, the system settings are similar to the examples above with setting $\varphi = 0.8$ and $\eta = 0.8$. In contrast to above, the battery leakage parameter μ is varied between 0 (no leakage) to 0.5.

The simulations in Figure 5 show that for all energy allocation policies, in both cases (balanced and unbalanced case), 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 despite energy leakage compared to the balanced scenario. For increasing energy loss due to battery leakage energy shared among the sensors approaches the average amount of energy used for data transmission. As in the examples above, the modified greedy policy ('h1') is outperformed by the ad hoc policy ('h2') in case of unbalanced networks. In case of balanced networks, the ad hoc heuristic ('h2') outperforms the modified greedy policy ('h1') for sufficiently high battery leakage. This is despite the ad hoc policy being developed for systems without battery leakage.

VII. Conclusions

This paper studied the distortion minimization problem of a multi sensor system, where each sensor transmits its measurement to a FC over a fading channel using uncoded analog forwarding for remote estimation at the FC. The FC computes the optimal energy allocation policy to minimize a long term average distortion cost when using the minimum mean-square error (MMSE) estimator under the following energy constraints: (i) the batteries at the sensors have a limited capacity and are prone to energy leakage, (ii) the sensors can harvest energy from their environment but only causal information about the harvested energies is available, and (iii) the sensors are fitted with transceiver units, that allow them to share energy with their neighbors subject to some loss. Random harvested energies and channel gains are modeled as independent finite-state Markov chains. The FC has causal information about the sensors' channel gains and harvested energy levels.

The optimal solution is obtained via a stochastic control approach resulting in a Bellman dynamic programming equation. A suboptimal Q-learning algorithm, which does not require a priori knowledge of all system parameters, is also studied. Further, to avoid the computational burden of the optimal solution based on dynamic programming techniques, two heuristic ad hoc energy allocation policies are presented and the performances of all policies are compared via numerical examples. The simulations reveal that the average distortion decreases as the cross correlation and the energy transfer efficiency increase. Further, in most scenarios, the optimal solution (obtained by dynamic programming) clearly outperforms the two sub-optimal policies. It can also be seen that an increase in energy transfer efficiency (for energy sharing) and an increase in the cross correlation term have a significantly higher impact on the average distortion if the system is unbalanced, that is, if a 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 wireless sensor networks with energy harvesting and energy sharing. Despite this paper focusing on relatively simple star networks, the results show that even for those simplistic network settings, the optimal energy allocation policy is far from trivial. Indeed, the findings presented here, form an important base for further investigation in this area as they provide a benchmark for more complicated network topologies. As a next step, more advanced sensor networks should be considered.

References

[1] I. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "A survey on sensor networks," *IEEE Communications Magazine*, vol. 40, no. 8, pp. 102–114, August 2002.

[2] V. Gungor and G. Hancke, "Industrial wireless sensor networks: Challenges, design, principles and technical approaches," *IEEE Transactions on Industrial Electronics*, vol. 56, no. 10, pp. 4258–4265, 2009.

[3] C.-Y. Chong and S. Kumar, "Sensor networks: Evolution, opportunities and challenges," *Proceedings of the IEEE*, vol. 91, no. 8, pp. 1247–1256, April 2003.

[4] V. Gungor, B. Lu, and G. Hancke, "Opportunities and challenges of wireless sensor networks in smart grid," *IEEE Transactions on Industrial Electronics*, vol. 57, no. 10, pp. 3557–3564, October 2010.

[5] C. E. Jones, K. M. Sivalingam, P. Agrawal, and J. C. Chen, "A survey of energy efficient network protocols for wireless networks," *Wireless Networks*, vol. 7, no. 4, pp. 343–358, 2001.

[6] A. J. Goldsmith and S. B. Wicker, "Design challenges for energy-constrained ad hoc wireless networks," *IEEE Wireless Communications*, vol. 9, no. 4, pp. 8–27, 2002.

[7] D. E. Quevedo, A. Ahlén, and J. Østergaard, "Energy efficient state estimation with wireless sensors through the use of predictive power control and coding," *IEEE Transactions on Signal Processing*, vol. 58, no. 9, pp. 4811–4823, September 2010.

[8] V. Sharma, U. Mukherji, V. Joseph, and S. Gupta, "Optimal energy managment policies for energy harvesting sensor nodes," *IEEE Transactions on Wireless Communications*, vol. 9, no. 4, 2010.

[9] C. K. Ho and R. Zhang, "Optimal energy allocation for wireless communications with energy harvesting constraints," *IEEE Transactions on Signal Processing*, vol. 60, no. 9, pp. 4808–4818, September 2012.

[10] J. Yang, O. Ozel, and S. Ulukus, "Broadcasting with an energy harvesting rechargeable transmitter," *IEEE Transactions on Wireless Communications*, vol. 11, no. 2, pp. 571–583, February 2012.

[11] K. Tutuncuoglu and A. Yener, "Optimum transmission policies for battery limited energy harvesting nodes," *IEEE Transactions on Wireless Communications*, vol. 11, no. 3, pp. 1180–1189, March 2012.

[12] O. Ozel, K. Tutuncuoglu, J. Yang, S. Ulukus, and A. Yener, "Transmission with energy harvesting nodes in fading wireless channels: Optimal policies," *IEEE Journal of Selected Areas in Commnunication*, vol. 29, no. 8, pp. 1732–1743, September 2011.

[13] P. Castiglione, O. Simeone, E. Erkip, and T. Zemen, "Energy management policies for energy-neutral source-channel coding," *IEEE Transactions on Communications*, vol. 60, no. 9, pp. 2668–2678, Sep. 2012.

[14] P. Castiglione and G. Matz, "Energy-neutral source-channel coding with battery and memory size constraints," *IEEE Transactions on Communications*, vol. 62, no. 4, pp. 1373–1381, April 2014.

[15] S. Padakandla, K. J. Prabuchandran, and S. Bhatnagar, "Energy sharing for multiple sensor nodes with finite buffers," *IEEE Transactions on Communications*, vol. 63, no. 5, pp. 1811–1823, May 2015.

[16] 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. 83, pp. 83–86, July 2007.

[17] A. Karalis, J. D. Joannopoulos, and M. Soljačić, "Efficient wireless non-radiative mid-range energy transfer," *Annals of Physics*, vol. 323, no. 1, pp. 34–48, January 2008.

[18] [Online]. Available: http://news.softpedia.com/news/The-Wonders-of-Wireless-Electricity-475837.shtml

[19] [Online]. Available: http://www.wirelesspowerconsortium.com

[20] [Online]. Available: http://www.powercastco.com

[21] [Online]. Available: http://money.cnn.com/2014/06/12/technology/mobile/starbucks-recharging/

[22] [Online]. Available: http://www.pluglesspower.com

[23] W. Lumpkins, "Nikola Tesla's dream realized: Wireless power energy harvesting," *IEEE Consumer Electronics Magazine*, vol. 3, no. 1, pp. 39–42, January 2014.

[24] A. M. Fouladgar and O. Simeone, "On the transfer of information and energy in multi-user systems," *IEEE Communications Letters*, vol. 16, no. 11, pp. 1733–1736, November 2012.

[25] R. Zhang and C. K. Ho, "MIMO broadcasting for simultaneous wireless information and power transfer," *IEEE Transactions on Wireless Communications*, vol. 12, no. 5, pp. 1989–2001, May 2013.

[26] L. Liu, R. Zhang, and K.-C. Chua, "Wireless information and power transfer: A dynamic power splitting approach," *IEEE Transactions on Communications*, vol. 61, no. 9, pp. 3990–4001, September 2013.

[27] K. Huang and E. G. Larsson, "Simultaneous information and power transfer for broadband wireless systems," *IEEE Transactions on Signal Processing*, vol. 61, no. 23, pp. 5972–5986, December 2013.

[28] K. Huang and V. K. N. Lau, "Enabling wireless power transfer in cellular networks: Architecture, modeling and deployment," *IEEE Transactions on Wireless Communications*, vol. 13, no. 2, 2014.

[29] Y. Shi, L. Xie, Y. T. Hou, and H. D. Sherali, "On renewable sensor networks with wireless energy transfer," in *Proceedings IEEE INFOCOM*, April 2011, pp. 1350–1358.

[30] L. Xie, Y. Shi, Y. T. Hou, W. Lou, H. D. Sherali, and S. F. Midkiff, "Renewable sensor networks with wireless energy transfer: The multi-node case," in *9th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks*, 2012.

[31] B. Gurakan, O. Ozel, J. Yang, and S. Ulukus, "Energy cooperation in energy harvesting communications," *IEEE Transactions on Communications*, vol. 61, no. 12, pp. 4884–4898, December 2013.

[32] T. Zhu, Z. Zhong, Y. Gu, T. He, and Z.-L. Zhang, "Leakage-aware energy synchronization for wireless sensor networks," in *Proceedings of the 7th international conference on Mobile systems, applications, and services*, June 2009, pp. 319–332.

[33] R. Shigeta, Y. Kawahara, and T. Asami, "Demo: Capacitor leakage aware duty cycle control for energy harvesting wireless sensor networks," in *Proceedings of the 9th ACM Conference on Embedded Networked Sensor Systems*, November 2011, pp. 387–388.

[34] I. Ahmed, A. Ikhlef, D. W. K. Ng, and R. Schober, "Power allocation for an energy harvesting transmitter with hybrid energy sources," *IEEE Transactions on Wireless Communications*, vol. 12, no. 12, pp. 6255–6267, December 2013.

[35] N. Su and M. Koca, "Stochastic transmission policies for energy harvesting nodes with random energy leakage," in *Proceedings of European Wireless*, May 2014, pp. 896–901.

[36] K. Tutuncuoglu, A. Yener, and S. Ulukus, "Optimum policies for an energy harvesting transmitter under energy storage losses," *IEEE Journal of Selected Areas in Commnunication*, vol. 33, no. 3, pp. 467–481, March 2015.

[37] S. Knorn, S. Dey, A. Ahlén, and D. E. Quevedo, "Distortion minimization in multi-sensor estimation using energy harvesting and energy sharing," *IEEE Transactions on Signal Processing*, vol. 63, no. 11, pp. 2848–2863, June 2015.

[38] ——, "Multi-sensor estimation using energy harvesting and energy sharing," in *Proceedings of the IEEE International Conference on Communications*, June 2015.

[39] D. P. Bertsekas, *Dynamic Programming and Optimal Control*, 3rd ed. Athena Scientific, 1995, vol. 1.

[40] [Online]. Available: www.freepik.com

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

[42] M. Gastpar, "Uncoded transmission is exactly optimal for a simple Gaussian "sensor" network," *IEEE Transactions on Information Theory*, vol. 54, no. 11, pp. 5247–5251, November 2008.

[43] D. E. Quevedo, A. Ahlén, and K. H. Johansson, "State estimation over sensor networks with correlated wireless fading channels," *IEEE Trans. Automatic Control*, vol. 58, no. 3, pp. 581–593, March 2013.

[44] I. Bahceci and A. K. Khandani, "Linear estimation of correlated data in wireless sensor networks with optimum power allocation and analog modulation," *IEEE Transactions on Communications*, vol. 56, no. 7, pp. 1146–1156, July 2008.

[45] E. Altman, *Constrained Markov Decision Processes*. CRC Press, 1999, vol. 7.

[46] D. P. Bertsekas, *Dynamic Programming and Optimal Control*, 3rd ed. Athena Scientific, 1995, vol. 2.

[47] H. Yu and D. P. Bertsekas, "Discretized approximations for POMDP with average cost," in *Proc. 20th Conference on Uncertainty in Artifical Intelligence*, Banff, Canada, 2004, pp. 619–627.

[48] R. Sutton and A. Barto, *Reinforcement learning: An introduction*. Cambridge Univ Press, 1998, vol. 1.

[49] 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 Communications Letters*, vol. 2, no. 1, pp. 82–85, February 2013.