Infinite Horizon Optimal Transmission Power Control for Remote State Estimation Over Fading Channels

Xiaoqiang Ren, Junfeng Wu, Karl Henrik Johansson, *Fellow, IEEE*, Guodong Shi, *Member, IEEE*, and Ling Shi

***Abstract*—This paper studies the joint design over an infinite horizon of the transmission power controller and remote estimator for state estimation over fading channels. A sensor observes a dynamic process and sends its observations to a remote estimator over a wireless fading channel characterized by a time-homogeneous Markov chain. The successful transmission probability depends on both the channel gains and the transmission power used by the sensor. The transmission power control rule and the remote estimator should be jointly designed, aiming to minimize an infinite-horizon cost consisting of the power usage and the remote estimation error. We formulate the joint optimization problem as an average cost belief-state Markov decision process and prove that there exists an optimal deterministic and stationary policy. We then show that when the monitored dynamic process is scalar or the system matrix is orthogonal, the optimal remote estimates depend only on the most recently received sensor observation, and the optimal transmission power is symmetric and monotonically increasing with respect to the norm of the innovation error.**

***Index Terms*—Estimation, fading channel, Kalman filtering, Markov decision process, power control.**

I. Introduction

IN NETWORKED control systems, control loops are often closed over a shared wireless communication network. This motivates research on remote state estimation problems, where a sensor measures the state of a linear system and transmits its observations to a remote estimator over a wireless fading channel. Such monitoring problems appear in a wide range of applications in environmental monitoring, space exploration, smart grids, intelligent buildings, among others. The challenges introduced by the networked setting lie in the fact that nonideal communication environment and constrained power supplies at sensing nodes may result in overall system performance degradation. The past decade has witnessed tremendous research efforts devoted to communication-constrained estimation problems, with the purpose of establishing a balance between estimation performance and communication cost.

A. Related Work

Wireless communications are being widely used nowadays in sensor networks and networked control systems. The interface of control and wireless communication has been a central theme in the study of networked sensing and control systems in the past decade. Early works assumed finite-capacity digital channels and focused on the minimum channel capacity or data rate needed for feedback stabilization, and on constructing encoder–decoder pairs to improve performance, e.g., [1]–[3]. Motivated by the fact that packets are the fundamental information carrier in most modern data networks [4], networked control and estimation subject to packet delays [5] and packet losses [6], [7] has been extensively studied.

State estimation is embedded in many networked control applications, playing a fundamental role therein. For networked state estimation subject to limited communication resource, the research on controlled communication has been extensive, see the survey [4]. Controlled communication, in general referring to reducing the communication rate intentionally to obtain a desirable tradeoff between the estimation performance and the communication rate, is motivated from at least two facts: 1) wireless sensors are usually battery-powered and sparsely deployed, and replacement of battery is difficult or even impossible, so the amount of communication needs to be kept at a minimum as communication is often the dominating on-board energy consumer [8]; and 2) traffic congestion in a sensor network may lead to packet losses and other network performance degradation. To minimize the inevitable enlarged estimation error due to reduced communication rate, a communication scheduling strategy for the sensor is needed. Two lines of research

Manuscript received February 21, 2017; accepted May 17, 2017. Date of publication May 30, 2017; date of current version December 27, 2017. The work of X. Ren and L. Shi was supported by a HK RGC theme-based project T23-701/14N. The work of J. Wu and K. H. Johansson was supported in part by the Knut and Alice Wallenberg Foundation, in part by the Swedish Foundation for Strategic Research, in part by the Swedish Research Council, VINNOVA, and in part by the NNSF of China under Grant 61120106011. Recommended by Associate Editor M. Verhaegen. (*Corresponding author: Junfeng Wu.*)

X. Ren and L. Shi are with the Department of Electronic and Computer Engineering, The Hong Kong University of Science and Technology, Hong Kong (e-mail: xren@connect.ust.hk; eesling@ust.hk).

J. Wu and K. H. Johansson are with the ACCESS Linnaeus Center, School of Electrical Engineering, Royal Institute of Technology, 114 28 Stockholm, Sweden (e-mail: junfengw@kth.se; kallej@kth.se).

G. Shi is with the College of Engineering and Computer Science, The Australian National University, Canberra ACT 0200, Australia (e-mail: guodong.shi@anu.edu.au).

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.2017.2709914

directions are present in the literature. The first line is known as time-based (offline) scheduling, whereby the communication decisions are simply specified only according to the time. Informally, a purely time-based strategy is likely to lead to a periodic communication schedule [9], [10]. The second line is known as event-based scheduling, whereby the communication decisions are specified according to the system state. The idea of event-based scheduling was popularized by Lebesgue sampling [11]. Deterministic event-based transmission schedules have been proposed in [12]–[18] for different application scenarios, and randomized event-based transmission schedules can be found in [19] and [20]. Essentially, event-based scheduling is a sequential decision problem with a team of two agents (a sensor and an estimator). Due to the nonclassical information structure of the two agents, joint optimization of the communication controller and the estimator is hard [21], and the interested readers are referred to [22] and references therein to see more on the team decision theory. Most works [12], [13], [15]–[18] bypassed the challenge by imposing restricted information structures or by approximations, while some authors have obtained structural descriptions of the agents under the joint optimization framework, using a majorization argument [14], [16] or an iterative procedure [18]. In all these works, communication models were highly simplified, restricted to a binary switching model.

Fading is nonignorable impairment to wireless communication [23]. The effects of fading have been taken into account in networked control systems [24], [25]. There are works that are concerned with transmission power management for state estimation [26]–[28]. The power allocated to transmission affects the probability of successful reception of the measurement, thus affecting the estimation performance. In [28], imperfect acknowledgments of communication links and energy harvesting were taken into account. In [26], power allocation for the estimation outage minimization problem was investigated in estimation of a scalar Gauss–Markov source. In all of the aforementioned works, the estimation error covariances are a Markov chain controlled by the transmission power, so the Markov decision process (MDP) theory is ready for solving this kind of problems. Gatsis *et al.* [27] considered the case when plant state is transmitted from a sensor to the controller over a wireless fading channel. The transmission power is adapted to the channel gain and the plant states. Due to nonclassical information structure, joint optimization of plant input and transmit power policies, although desired, is difficult. A restricted information structure was, therefore, imposed, i.e., only a subset of the full information history available at the sensor is utilized when determining the transmission power, to allow separate design at expense of loss of optimality. It seems that such a challenge involved in these joint optimization problems always exists.

B. Contributions

In this paper, we consider a remote state estimation scheme, where a sensor measures the state of a linear time-invariant discrete-time process and transmits its observations to a remote estimator over a wireless fading channel characterized by a time-homogeneous Markov chain. The successful transmission probability depends on both the channel gain and the transmission power used by the sensor. The objective is to minimize an infinite horizon cost consisting of the power consumption and the remote estimation error. In contrast to [27], no approximations are made to prevent loss of optimality, which however renders the analysis challenging. We formulate our problem as an infinite horizon belief-state MDP with an average cost criterion. Contrary to the ideal “send or not” communication scheduling model considered in [14] and [16], for which the majorization argument applies for randomized policies, a first question facing our fading channel model with an infinite horizon is whether or not the formulated MDP has an optimal stationary and deterministic policy. The answer is yes provided certain conditions given in this paper. On top of this, we present structural results on the optimal transmission power controller and the remote estimator for some special systems, which can be seen as the extension of the results in [14], [16], and [18] for the power management scenario. The analysis tools used in this paper (i.e., the partially observable Markov decision process (POMDP) formulation and the majorization interpretation) is inspired by [16] (the majorization technique of which is a variation of [14] and [29]). Nevertheless, the contributions of the two works are distinct. In [16], the authors mainly studied the threshold structure of the optimal communication strategy within a finite horizon, while this paper focuses on the asymptotic analysis of the joint optimization problem over an infinite horizon. A slightly more general model than [16] is studied in [30] under infinite time horizon, where the focus was on explicit characterization of the threshold policy with a Markov chain source and symmetric noises assumed *a priori*. The existence establishment of the solution (stationary and deterministic) relied heavily on the threshold structure. The general modeling of the monitored process and the fading channel, however, makes our analysis much more challenging.

In summary, the main contributions of this paper are listed as follows. We prove that a deterministic and stationary policy is an optimal solution to the formulated average cost belief-state MDP. We should remark that the abstractness of the considered state and action spaces (the state space is a probability measure space and the action space a function space) renders the analysis rather challenging. Then, we prove that both the optimal estimator and the optimal power control have simple structures when the dynamic process monitored is scalar or the system matrix is orthogonal. To be precise, the remote estimator synchronizes its estimates with the data received in the presence of successful transmissions, and linearly projects its estimates a step forward otherwise. For a certain belief, the optimal transmission power is a symmetric and monotonically increasing function of the norm of the innovation error. Thanks to these properties, both the offline computation and the online implementation of the optimal transmission power rule are greatly simplified, especially when the available power levels are discrete, for which only thresholds of switchings between power levels are to be determined.

This paper provides a theory in support of the study of infinite horizon communication-constrained estimation problems. Deterministic and stationary policies are relatively easy to compute and implement, thus, it is important to know that an optimal solution that such a policy exists. The structural characteristic of the jointly optimal transmission power and estimation policies provides insights into the design of energy-efficient state estimation algorithms.

C. Paper Organization

In Section II, we provide the mathematical formulation of the system model adopted, including the monitored dynamic process, the wireless fading channel, the transmission power controller, and the remote estimator. We then present the considered problem and formulate it as an average cost MDP in Section III. In Section IV, we prove that there exists a deterministic and stationary policy that is optimal to the formulated MDP. Some structural results about the optimal remote estimator and the optimal transmission power control strategy are presented in Section V. In Section VI, we discuss about the practical implementation of the whole system. Concluding remarks are given in Section VII. All the proofs and some auxiliary background results are provided in the appendixes.

C. Notation

$\mathbb{N}$ and $\mathbb{R}_+$ ($\mathbb{R}_{++}$) are the sets of nonnegative integers and nonnegative (positive) real numbers, respectively. $\mathbb{S}^n_+$ (and $\mathbb{S}^n_{++}$) is the set of n by n positive semidefinite matrices (and positive definite matrices). When $X \in \mathbb{S}^n_+$ (and $\mathbb{S}^n_{++}$), we write $X \succeq 0$ (and $X \succ 0$). $X \succeq Y$ if $X - Y \in \mathbb{S}^n_+$. $\mathrm{Tr}(\cdot)$ and $\det(\cdot)$ are the trace and the determinant of a matrix, respectively. $\lambda_{\max}(\cdot)$ represents the eigenvalue, having the largest magnitude, of a matrix. The superscripts $^\top$ and $^{-1}$ stand for matrix transposition and matrix inversion, respectively. The indicator function of a set $\mathcal{A}$ is defined as

$$\mathbb{1}_{\mathcal{A}}(\omega) = \begin{cases} 1, & \omega \in \mathcal{A} \\ 0, & \omega \notin \mathcal{A}. \end{cases}$$

The notation $p(\mathbf{x}; x)$ represents the probability density function (pdf) of a random variable $\mathbf{x}$ with x as the input variable. If being clear in the context, $\mathbf{x}$ is omitted. For a random variable $\mathbf{x}$ and a pdf θ, the notation $\mathbf{x} \sim \theta$ means that $\mathbf{x}$ follows the distribution defined by θ. For measurable functions $f, g : \mathbb{R}^n \to \mathbb{R}$, we use $f * g$ to denote the convolution of f and g. For a Lebesgue measurable set $A \subset \mathbb{R}^n$, $\mathfrak{L}(A)$ denotes the Lebesgue measure of A. Let $\|x\|$ denote the L^2 norm of a vector $x \in \mathbb{R}^n$. δ_{ij} is the Dirac delta function, i.e., δ_{ij} equals to 1 when $i = j$ and 0 otherwise. In addition, $\mathbb{P}(\cdot)$ (or $\mathbb{P}(\cdot|\cdot)$) refers to (conditional) probability.

II. System Model

In this paper, we focus on dynamic power control for remote state estimation. We consider a remote state estimation scheme as depicted in Fig. 1. In this scheme, a sensor measures a linear time-invariant discrete-time process and sends its measurement in the form of data packets, to a remote estimator over a wireless link. The remote estimator produces an estimate of the process state based on the received data. When sending packets through the wireless channel, transmissions may fail due to interference and weak channel gains. Packet losses lead to distortion of the remote estimation and packet loss probabilities depend on transmission power levels used by the transmitter and on the channel gains. Lower loss probabilities require higher transmission power usage; on the other hand, energy saving is critical to expand the lifetime of the sensor. The wireless communication overhead dominates the total power consumption, therefore, we introduce a transmission power controller, which aims to balance the transmission energy cost and distortion penalty as the channel gain varies over time.

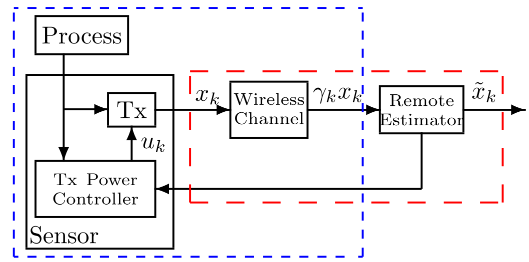


Fig. 1. Remote state estimation scheme.

In what follows, the attention is devoted to laying out the main components in Fig. 1.

A. State Process

We consider the following linear time-invariant discrete-time process:

$$x_{k+1} = Ax_k + w_k \tag{1}$$

where $k \in \mathbb{N}$, $x_k \in \mathbb{R}^n$ is the process state vector at time k, $w_k \in \mathbb{R}^n$ is zero-mean independent and identically distributed (i.i.d.) noises, described by the probability density function (pdf) $\mathfrak{f}_w$, with $\mathbb{E}[w_k w_k^\top] = W$ ($W \succeq 0$). We further assume that the support of the noise distribution is unbounded, i.e., for any $C > 0$, there holds $\int_{\|w\| \geq C} \mathfrak{f}_w(w)\mathrm{d}w > 0$. The initial state x_0, independent of $w_k, k \in \mathbb{N}$, is described by the pdf $\mathfrak{f}_{x_0}$, with mean $\mathbb{E}[x_0]$ and covariance Σ_0. Without loss of generality, we assume $\mathbb{E}[x_0] = 0$, as nonzero-mean cases can be translated into zero-mean one by coordinate change $x'_k = x_k - \mathbb{E}[x_0]$. The system parameters are all known to the sensor as well as the remote estimator. Notice that we do not impose any constraint on the stability of the process in (1), i.e., $|\lambda_{\max}(A)|$ may take any value in $\mathbb{R}_+$.

B. Wireless Communication Model

The sensor measures and sends the process state x_k to the remote estimator over an additive white Gaussian noise (AWGN) channel that suffers from channel fading (see Fig. 2)

$$\mathbf{y} = g_k \mathbf{x} + v_k$$

where g_k is a random complex number, and v_k is additive white Gaussian noise; $\mathbf{x}$ represents the signal (e.g., x_k) sent by the transmitter and $\mathbf{y}$ the signal received by the receiver. Let the channel gain $h_k = |g_k|^2$ take values in a finite set $\mathbb{h} \subseteq \mathbb{R}_{++}$, where l is the size of $\mathbb{h}$, and $\{h_k\}_{k\in\mathbb{N}}$ possess temporal correlation modeled by a time-homogenous Markov chain. The

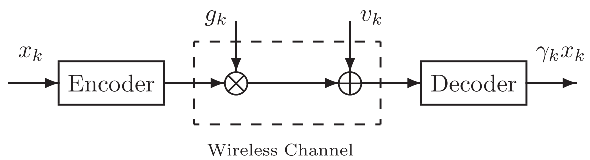


Fig. 2. Wireless communication model, where g_k is a random complex number, and v_k is additive white Gaussian noise.

one-step transition probability for this chain is denotedby

$$\Xi(\cdot|\cdot) : \mathbb{h} \times \mathbb{h} \longmapsto [0,1].$$

The function $\Xi(\cdot|\cdot)$ is known *a priori*. We assume the remote estimator or the sensor can access the channel state information (CSI), so the channel gain h_k is available at each time before transmission. This might be achieved by channel reciprocity techniques, which are typical in time-division-duplex-based transmissions [23]. The estimation errors of the channel gains are not taken into account in this paper.

To facilitate our analysis, the following assumption is made.

Assumption 1 (Communication model).

1) The channel gain h_k is independent of the system parameters.
2) The channel is block fading, i.e., the channel gain remains constant during each packet transmission and varies from block to block.
3) The quantization effect is negligible and does not effect the remote estimator.
4) The receiver can detect symbol errors.[1] Only the data reconstructed error free are regarded as successfully reception. The receiver perfectly realizes whether the instantaneous communication succeeds or not.
5) The Markov chain governing the channel gains, $\Xi(\cdot|\cdot)$, is aperiodic and irreducible.

Assumption 1-1)–4) are standard for fading channel model. Note that Assumption 1-1), 3), 4) were used in [6] and [27], and that Assumption 1-2) was used in [25]. From Assumption 1-4), whether or not the data sent by the sensor is successfully received by the remote estimator is indicated by a sequence $\{\gamma_k\}_{k\in\mathbb{N}}$ of random variables, where

$$\gamma_k = \begin{cases} 1, & \text{if } \ x_k \text{ is received error free at time } k \\ 0, & \text{otherwise (regarded as dropout)} \end{cases} \tag{2}$$

initialized with $\gamma_0 = 1$. Assumption 1-5) is a technical requirement for our analysis. One notes that both the i.i.d. channel gains model and the Gilbert–Elliott model with the good/bad state transition probability not equal to 1 satisfy Assumption 1-5).

C. Transmission Power Controller

Let $u_k \in \mathbb{R}_+$ be the transmission power at time k, the power supplied to the radio transmitter. Due to constraints with respect to radio power amplifiers, the admissible transmission power is restricted. Let u_k take values in $\mathcal{U} \subset \mathbb{R}_+$, which may be an infinite or a finite set depending on the radio implementation. It is further assumed that $\mathcal{U}$ is compact and contains zero. Under Assumption 1-3), the successful packet reception is statistically determined by the signal-to-noise ratio (SNR) $h_k u_k / N_0$ at the receiver, where N_0 is the power spectral density of v_k. The different modulation models may be characterized by the conditional packet reception probability

$$q(u_k, h_k) \triangleq \mathbb{P}(\gamma_k = 1 | u_k, h_k). \tag{3}$$

Assumption 2: The function $q(u,h) : \mathcal{U} \times \mathbb{h} \to [0,1]$ is nondecreasing in both u and h.

This assumption is consistent with the intuition that more transmission power or a better channel state will lead to a higher packet arrival rate, which is common for a fading channel model [25], [27].

Assumption 3: The function $q(u,h) : \mathcal{U} \times \mathbb{h} \to [0,1]$ is continuous almost everywhere with respect to u for any fixed h. Moreover, $q(0,h) = 0$ and $q(\bar{u}, h) > 0$ for all $h \in \mathbb{h}$, where $\bar{u}$ is the highest available power level: $\bar{u} \triangleq \max\{u : u \in \mathcal{U}\}$.

Remark 1: Notice that since $\mathcal{U}$ is compact, $\bar{u}$ always exist. Let $\mathcal{U} = \{0, 1\}$ with

$$q(u_k, h_k) = \begin{cases} 1, \text{if } u_k = 1 \\ 0, \text{if } u_k = 0. \end{cases}$$

Then the "ON–OFF" controlled communication problem considered in [12–20] and [31]–[33] becomes a special case of the transmission power control problem considered here.

We assume that packet reception probabilities are conditionally independent for given channel gains and transmission power levels, which is stated in the following assumption.

Assumption 4: The following equality holds for any $k \in \mathbb{N}$:

$$\mathbb{P}(\gamma_k = r_k, \ldots, \gamma_1 = r_1 | u_{1:k}, h_{1:k}) = \prod_{j=1}^{k} \mathbb{P}(\gamma_j = r_j | u_j, h_j).$$

Remark 2: Assumption 2 is standard for digital communication over fading channels. Assumption 3 is in accordance with the common sense that the symbol error rate statistically depends on the instantaneous SNR at the receiver. Many digital communication modulation methods are embraced by these assumptions [25].

Assumption 5: The following relation holds:

$$\min_{h \in \mathbb{h}} \mathbb{E}\left[q(\bar{u}, h_{k+1}) | h_k = h\right] > 1 - \frac{1}{|\lambda_{\max}(A)|^2} \tag{4}$$

where A is the system matrix in (1).

Remark 3: Assumption 5 provides a sufficient condition under which the expected estimation error covariance is bounded when the maximum power level is consistently used. Notice that when the channel gain $\{h_k\}$ is i.i.d., Assumption 5 coincides with [27, Assumption 1]. Notice also that when the system is stable, i.e., $|\lambda_{\max}(A)| < 1$, for any communication model (4) trivially holds.

D. Remote Estimator

At the base station side, each time a remote estimator generates an estimate based on what it has received from the sensor.

[1] In practice, symbol errors can be detected via a cyclic redundancy check code.

In many applications, the remote estimator is powered by an external source or is connected with an energy-abundant controller/actuator, thus having sufficient communication energy in contrast to the energy-constrained sensor. This energy asymmetry allows us to assume that the estimator can send messages back to the sensor. The content of feedback messages are separatively defined under different system implementations, the details of which will be discussed later in Section VI. Denote by $\mathcal{O}_k^-$ the observation obtained by the remote estimator up to *before* the communication at time k, i.e.

$$\mathcal{O}_k^- \triangleq \{\gamma_1 x_1, ..., \gamma_{k-1} x_{k-1}\} \cup \{\gamma_1, \ldots, \gamma_{k-1}\} \cup \{h_1, \ldots, h_k\}.$$

Similarly, denote by $\mathcal{O}_k^+$ the observation obtained by the remote estimator up to *after* the communication at time k, where

$$\mathcal{O}_k^+ \triangleq \mathcal{O}_k^- \cup \{\gamma_k, \gamma_k x_k\}.$$

III. Problem Definition

We take into account both the estimation quality at the remote estimator and the transmission energy consumed by the sensor. To this purpose, joint design of the transmission power controller and the remote estimator is desired. Measurement realizations, communication indicators, and channel gains are adopted to manage the usage of transmission power

$$u_k = f_k\big(x_{1:k}, h_{1:k}, \gamma_{1:k-1}\big). \tag{5}$$

Given the transmission power controller, the remote estimator generates an estimate as a function of what it has received from the sensor, i.e.

$$\hat{x}_k \triangleq g_k(\mathcal{O}_k^+). \tag{6}$$

We emphasize that since the transmission power controller $f_{1:k}$ affects the arrival of the data, the optimal estimate $\hat{x}_k$ should also depend on $f_{1:k}$. The average remote estimation quality over an infinite time horizon is quantified by

$$\mathcal{E}(\mathbf{f}, \mathbf{g}) \triangleq \mathbb{E}_{\mathbf{f},\mathbf{g}}\left[\limsup_{T\to\infty} \frac{1}{T}\sum_{k=1}^{T} \|x_k - \hat{x}_k\|^2\right] \tag{7}$$

correspondingly, the average transmission power cost, denoted as $\mathcal{W}(\mathbf{f})$, is given by

$$\mathcal{W}(\mathbf{f}) \triangleq \mathbb{E}_{\mathbf{f}}\left[\limsup_{T\to\infty} \frac{1}{T}\sum_{k=1}^{T} u_k\right] \tag{8}$$

where $\mathbf{f} \triangleq \{f_1, \ldots, f_k, \ldots\}$ and $\mathbf{g} \triangleq \{g_1, \ldots, g_t, \ldots\}$. It is clear from the common arguments in $\mathcal{E}(\cdot,\cdot)$ and $\mathcal{W}(\cdot)$ that the transmission power controller and the remote estimator must be designed jointly. Note that in (7) and (8), the expectations are taken with respect to the randomness of the system and the transmission outcomes for given $\mathbf{f}$ and $\mathbf{g}$. For the remote state estimation system, we naturally wonder how to find a jointly optimal transmission power controller f_k^* and remote state estimator g_k^* satisfying

$$\text{minimize}_{\mathbf{f},\mathbf{g}}\, [\mathcal{E}(\mathbf{f}, \mathbf{g}) + \alpha \mathcal{W}(\mathbf{f})] \tag{9}$$

where the constant α can be interpreted as a Lagrange multiplier. We should remark that (9) is difficult to solve due to the nonclassical information structure [21]. What is more, (9) has an average cost criterion that depends only on the limiting behavior of $\mathbf{f}$ and $\mathbf{g}$, adding additional analysis difficulty.

A. Belief-State Markov Decision Process

Before proceeding, we first give in the following lemma that the variables of the transmission power controller f_k defined in (5) can be changed without any loss of performance. The proof is similar to that of [16, Lemma 1].

Lemma 1: Without any loss of performance, the transmission power controller f_k defined in (5) can be restricted to the following form:

$$u_k = f_k\big(x_k, \mathcal{O}_k^-\big). \tag{10}$$

To find a solution to the optimization problem (9), we first observe from (8) that $\mathcal{W}(\mathbf{f})$ does not depend on $\mathbf{g}$, thus leading to an insight into the structure of g_k^*—Lemma 2, the proof of which follows from optimal filtering theory: the conditional mean is the minimum-variance estimate. Similar results can be seen in [14], [16], and [18].

Lemma 2: For any given transmission power controller f_k, the optimal remote estimator g_k^* is the MMSE estimator

$$\hat{x}_k \triangleq g_k^*(\mathcal{O}_k^+) = \mathbb{E}_{f_{1:k}}[x_k|\mathcal{O}_k^+]. \tag{11}$$

Problem (9) still remains hard since g_k^* depends on the choice of $f_{1:k}$. To address this issue, by adopting the common information approach [34], we formulate (9) as a POMDP at a fictional coordinator. The fictional coordinator, with the common information of the sensor and the estimator, will generate *prescriptions* that map from each side's private information to the optimal action. Notice that due to the feedback structure, there is no private information for the remote estimator. Also, the optimal action for the remote estimator (i.e., the optimal estimator) has been provided in Lemma 2. Thus, the goal of the POMDP is to find the optimal prescription for the sensor based on the common information. From (10), the private information at time k for the sensor is x_k. Hence, one may define the prescription $l_k : \mathbb{R}^n \to \mathcal{U}$ as

$$l_k(\cdot) = f_k(\mathcal{O}_k^-, \cdot).$$

Following the conventional treatment of the POMDP, we are allowed to equivalently study its belief-state MDP. For technical reasons, we pose two moderate constraints on the action space. We will present the formal belief-state MDP model and remark that the resulting gap between the formulated belief-state MDP and (9) is negligible (see Remark 6). Before doing so, a few definitions and notations are needed. Define innovation e_k as

$$e_k \triangleq x_k - A^{k-\tau(k)} x_{\tau(k)} \tag{12}$$

with e_k taking values in $\mathbb{R}^n$ and τ_k being the most recent time the remote estimator received data before time k as

$$\tau(k) \triangleq \max_{1\leqslant t\leqslant k-1}\{t : \gamma_t = 1\}. \tag{13}$$

Let $\hat{e}_k \triangleq \mathbb{E}_{f_{1:k}}[e_k|\mathcal{O}_k^+]$. Since $\tau(k), x_{\tau(k)} \in \mathcal{O}_{k-1}^+$, the equality

$$e_k - \hat{e}_k = x_k - \hat{x}_k \tag{14}$$

holds for all $k \in \mathbb{N}$. In other words, e_k can be treated as x_k offset by a variable that is measurable to $\mathcal{O}_{k-1}^+$. We define the belief state on e_k. From (14), the belief state on x_k can be equally defined. Here, we use e_k instead of x_k for ease of presentation.

Definition 1: Before the transmission at time k, the belief state $\theta_k(\cdot) : \mathbb{R}^n \to \mathbb{R}_+$ is defined as $\theta_k(e) \triangleq p(e_k; e|f_{1:k-1}, \mathcal{O}_k^-)$.

To define the action space accurately, we also need some definitions related to a partition of a set.

Definition 2: A collection Δ of sets is a partition of a set $\mathcal{X}$ if the following conditions are satisfied:

1) $\emptyset \notin \Delta$;
2) $\cup_{\mathcal{B} \in \Delta} \mathcal{B} = \mathcal{X}$;
3) if $\mathcal{B}_1, \mathcal{B}_2 \in \Delta$ and $\mathcal{B}_1 \neq \mathcal{B}_2$, then $\mathcal{B}_1 \cap \mathcal{B}_2 = \emptyset$.

An element of Δ is also called a cell of Δ. If $\mathcal{X} \subset \mathbb{R}^n$, we define the size of Δ as

$$|\Delta| \triangleq \sup_{\mathcal{B}, x, y} \{\|x - y\| : x, y \in \mathcal{B}, \mathcal{B} \in \Delta\}.$$

Definition 3: For two partitions, denoted as Δ_1 and Δ_2, of a set $\mathcal{X}$, Δ_1 is called a refinement of Δ_2 if every cell of Δ_1 is a subset of some cell of Δ_2. Formally it is written as $\Delta_1 \preceq \Delta_2$.

One can verify that the relation $\preceq$ is a partial order, and the set of partitions together with this relation form a lattice. We denote the meet [35, Definition 1.3][2] of partitions Δ_1 and Δ_2 as $\Delta_1 \wedge \Delta_2$.

Now, we are able to mathematically describe the belief-state MDP by a quintuplet $(\mathbb{N}, \mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{C})$. Each item in the tuple is elaborated as follows.

1) The set of decision epochs is $\mathbb{N}$.
2) State space $\mathcal{S} = \Theta \times \mathbb{h}$: Θ is the set of beliefs over $\mathbb{R}^n$, i.e., the space of probability measures on $\mathbb{R}^n$. The set Θ is further constrained as follows. Let μ be a generic element of Θ. Then, μ is absolutely continuous with respect to the Lebesgue measure,[3] and μ has the finite second moment, i.e., $\int_{\mathbb{R}^n} \|e\|^2 \mathrm{d}\mu(e) < \infty$. Let $\theta(e) = \frac{\mathrm{d}\mu(e)}{\mathrm{d}\mathfrak{L}(e)}$ be the Radon–Nikodym derivative. Note that $\theta(e)$ is uniquely defined up to a $\mathfrak{L}$-null set (i.e., a set having Lebesgue measure zero). We thus use μ and $\theta(e)$ interchangeably to represent a probability measure on $\mathbb{R}^n$, and we do not distinguish between any two functions $\theta(e)$ and $\theta'(e)$ with $\mathfrak{L}(\{e : \theta(e) - \theta(e)' \neq 0\}) = 0$ by convention. We assume that Θ is endowed with the topology of weak convergence [36]. Denote by $s \triangleq (\mu, h)$ a generic element of $\mathcal{S}$. Let $\mathbb{d}_P(\cdot, \cdot)$ denote the Prohorov metric [36] on Θ. We define the metric on $\mathcal{S}$ as $\mathbb{d}_s((\mu_1, h_1), (\mu_2, h_2)) = \max\{\mathbb{d}_P(\mu_1, \mu_2), |h_1 - h_2|\}$.
3) Action space $\mathcal{A}$ is the set of all functions that have the following structure:

$$a(e) = \begin{cases} \bar{u}, & \text{if } \|e\| > L \\ a'(e), & \text{otherwise} \end{cases} \tag{15}$$

where $a' \in \mathcal{A}' : \mathcal{E} \to \mathcal{U}$ with $\mathcal{E} \triangleq \{e \in \mathbb{R}^n : \|e\| \leq L\}$. The space $\mathcal{A}'$ is further defined as follows. Let $a' \in \mathcal{A}'$ be a generic element, then there exists a finite partition $\Delta_{a'}$ of $\mathcal{E}$ such that each cell of $\Delta_{a'}$ is a $\mathfrak{L}$-continuity set[4] and on each cell $a'(e)$ is Lipschitz continuous with Lipschitz constant uniformly bounded by M. It is further assumed that $\overline{\Delta} = \wedge_{a' \in \mathcal{A}'} \Delta_{a'}$ is a finite partition of $\mathcal{E}$. We adopt the Skorohod distance defined in Appendix A, for which $\mathcal{X} = \mathcal{E}$. By convention, we do not distinguish two functions in $\mathcal{A}$ that have zero distance and we consider the space of the resulting equivalence classes. Note that the argument of the function $a(\cdot)$ is the innovation e_k defined in (12), and by the definition of e_k, one obtains that $a_k(e) = l_k(e + A^{k-\tau(k)} x_{\tau(k)})$.
4) The function $\mathcal{P}(\theta', h'|\theta, h, a) : \mathcal{S} \times \mathcal{A} \times \mathcal{S}$ defines the conditional state transition probability. To be precise

$$\begin{aligned}
&\mathcal{P}(\theta', h'|\theta, h, a) \\
&\triangleq p(\theta_{k+1}, h_{k+1}; \theta', h'|\theta_k = \theta, h_k = h, a_k = a) \\
&= \begin{cases} \Xi(h'|h)\left(1 - \varphi(\theta, h, a)\right), & \text{if } \theta' = \phi(\theta, h, a, 0) \\ \Xi(h'|h)\varphi(\theta, h, a), & \text{if } \theta' = \phi(\theta, h, a, 1) \\ 0, & \text{otherwise} \end{cases}
\end{aligned}$$

where $\varphi(\theta, h, a) \triangleq \int_{\mathbb{R}^n} q(a(e), h)\theta(e)\mathrm{d}e$, and

$$\begin{aligned}
&\phi(\theta, h, a, \gamma) \\
&\triangleq \begin{cases} \frac{1}{|\det(A)|}\theta^+_{\theta,h,a}(A^{-1}e) * \mathfrak{f}_w(e), & \text{if } \gamma = 0 \\ \mathfrak{f}_w(e), & \text{if } \gamma = 1 \end{cases}
\end{aligned} \tag{16}$$

where $\theta^+_{\theta,h,a}(e) \triangleq \frac{(1 - q(a(e),h))\theta(e)}{1 - \varphi(\theta,h,a)}$ is interpreted as the posttransmission belief when the transmission fails, and $\mathfrak{f}_w(\cdot)$, recall, is the pdf of the system noises in (1). One obtains (16) by noticing that $e_{k+1} = Ae_k + w_k$ if $\gamma_k = 0$ and $e_{k+1} = w_k$ otherwise.
5) The function $\mathcal{C}(\theta, h, a) : \mathcal{S} \times \mathcal{A} \to \mathbb{R}_+$ is the cost function when performing $a \in \mathcal{A}$ for $\theta \in \Theta$ and $h \in \mathbb{h}$ at time k, which is given by

$$\mathcal{C}(\theta, h, a) = \int_{\mathbb{R}^n} \theta(e)c(e, h, a)\mathrm{d}e. \tag{17}$$

In the aforementioned equation, the function $c(\cdot, \cdot, \cdot) : \mathbb{R}^n \times \mathbb{h} \times \mathcal{A} \to \mathbb{R}_+$ is defined as $c(e, h, a) = \alpha a(e) + (1 - q(a(e), h))\|e - \hat{e}_+\|^2$ with $\hat{e}_+ = \mathbb{E}_{\theta^+_{\theta,h,a}}[e] \triangleq \mathbb{E}$

[2] For $z, x, y \in \mathcal{A}$ with $\mathcal{A}$ being a partially ordered set, z is the meet of x and y, if the following two conditions are satisfied:

1) $z \preceq x$ and $z \preceq y$;
2) for any $w \in \mathcal{A}$ such that $w \preceq x$ and $w \preceq y$, there holds $w \preceq z$.

[3] Let μ_1 and μ_2 be measures on the same measurable space. Then, μ_1 is said to be absolutely continuous with respect to μ_2 if for any Borel subset $\mathcal{B}$, $\mu_2(\mathcal{B}) = 0 \Rightarrow \mu_1(\mathcal{B}) = 0$.

[4] A Borel subset $\mathcal{B}$ is said to be a μ-continuity set if $\mu(\partial\mathcal{B}) = 0$, where $\partial\mathcal{B}$ is the boundary set of $\mathcal{B}$.

$[e|e \sim \theta^{+}_{\theta,h,a}]$, where the communication cost is counted by the first term and the distortion $\|e - \hat{e}_{+}\|^2$ with probability $1 - q(a(e), h)$ is counted by the second term.

Remark 4: The initial belief $\theta_1(e) = 1/\det(A)\mathfrak{f}_{x_0}(A^{-1}e) * \mathfrak{f}_w(e)$ is absolutely continuous with respect to the Lebesgue measure. The belief evolution in (16) gives that, whatever policy is used, θ_k is absolutely continuous with respect to the Lebesgue measure for $k \geq 2$. Also, notice that if there exists a channel gain $h \in \mathbb{h}$ such that $q(\bar{u}, h) < 1$ and if θ has infinite second moment, then $\mathcal{C}(\theta, h, a) = \infty$ for any action a. Thus, to solve (9), without any performance loss, we can restrict beliefs into the state space Θ.

Remark 5: The action $a(e) \in \mathcal{A}$ is allowed to have a $\mathfrak{L}$-null set of discontinuity points. The assumption that on each cell of a partition, $a(e)$ is a Lipschitz function is a technical requirement for Theorem 1. The intuition is that given θ_k, except for $\mathfrak{L}$-null set of points, the difference between the power used for e_k and e'_k is at most proportional to the distance between e_k and e'_k. The saturation structure in (15), i.e., $a(e) = \bar{u}$ if $\|e\| > L$ is also a technical requirement for Theorem 1. Intuitively, this ensures that, when the transmission fails, the second moment of the post-transmission belief $\theta^{+}_{\theta,h,a}(e)$ is bounded by a function of the second moment of $\theta(e)$. The saturation assumption can also be found in [27].

An admissible k-history for this MDP is defined as $\mathbf{h}_k \triangleq \{\theta_1, h_1, a_1, \ldots, \theta_{k-1}, h_{k-1}, a_{k-1}, \theta_k, h_k\}$. Let $\mathcal{H}_k$ denote the class of all the admissible k-history $\mathbf{h}_k$. A generic policy $\mathbf{d}$ for $(\mathbb{N}, \mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{C})$ is a sequence of decision rules $\{d_k\}_{k\in\mathbb{N}}$, with each $d_k : \mathcal{H}_k \to \mathcal{A}$. In general, d_k may be a stochastic mapping. Let $\mathcal{D}$ denote the whole class of $\mathbf{d}$. In some cases, we may write $\mathbf{d}$ as $\mathbf{d}(d_k)$ to explicitly point out the decision rules used at each stage. We focus on the following problem. For any initial state (θ, h)

$$\begin{aligned} \min \quad & \mathcal{J}(\mathbf{d}, \theta, h) \triangleq \limsup_{T\to\infty} \frac{1}{T} \sum_{k=1}^{T} \mathbb{E}^{\theta,h}_{\mathbf{d}} \left[\mathcal{C}(\theta_k, h_k, a_k)\right] \\ \text{s.t.} \quad & \mathbf{d} \in \mathcal{D}. \end{aligned} \tag{18}$$

Remark 6: The gap between (9) and (18) arises from the structure assumptions for the action space. These structure constraints, however, are moderate, since the saturation level L and the uniform Lipschitz constant M can be arbitrarily large and the size of $|\overline{\Delta}|$ can be arbitrarily small.

IV. Optimal Deterministic Stationary Policy: Existence

The definition of the policy $\mathbf{d}$ in the aforementioned section allows the dependence of d_k on the full k-history, $\mathbf{h}_k$. Fortunately, with the aid of the results of average cost MDPs [37]–[39], we prove that there exists a deterministic stationary policy that is optimal to (18). Before showing the main theorem, we introduce some notations.

We define the class of deterministic and stationary policies $\mathcal{D}_{\rm ds}$ as follows: $\mathbf{d}(d_k) \in \mathcal{D}_{\rm ds}$ if and only if there exists a Borel measurable function $d : \mathcal{S} \to \mathcal{A}$ such that $\forall i$

$$d_k(\mathcal{H}_{k-1}, a_{k-1}, \theta_k = \theta, h_k = h) = d(\theta, h).$$

Since the decision rules d_k's are identical (equal d) along the time horizon for a stationary policy $\mathbf{d}(\{d_k\}_{k\in\mathbb{N}}) \in \mathcal{D}_{\rm ds}$, we write it as $\mathbf{d}(d)$ for the ease of notation.

Theorem 1: There exists a deterministic and stationary policy $\mathbf{d}^*(d) \in \mathcal{D}_{\rm ds}$ such that for any $(\theta, h) \in \mathcal{S}$, there holds

$$\mathcal{J}(\mathbf{d}^*(d), \theta, h) \leq \mathcal{J}(\mathbf{d}, \theta, h) \quad \forall\, \mathbf{d} \in \mathcal{D}$$

Moreover, the optimal policy is given by

$$\mathbf{d}^*(d) = \arg\min_{\mathbf{d}\in\mathcal{D}_{\rm ds}} \{\mathcal{C}_d(\theta, h) + \mathbb{E}_{\mathbf{d}}[\mathcal{Q}^*(\theta', h')|\theta, h]\} \tag{19}$$

and the optimal cost is

$$\mathcal{J}(\mathbf{d}^*(d), \theta, h) = \rho^* \quad \forall\, (\theta, h) \in \mathcal{S}$$

where the functions $\mathcal{Q}^* : \mathcal{S} \to \mathbb{R}$ and $\rho^* \in \mathbb{R}$ satisfy

$$\mathcal{Q}^*(\theta, h) = \min_{\mathbf{d}\in\mathcal{D}_{\rm ds}} \{\mathcal{C}_d(\theta, h) - \rho^* + \mathbb{E}_{\mathbf{d}}[\mathcal{Q}^*(\theta', h')|\theta, h]\}$$

with $\mathcal{C}_d(\theta, h) \triangleq \mathcal{C}(\theta, h, d(\theta, h))$ and $\mathbb{E}_{\mathbf{d}}[\mathcal{Q}^*(\theta', h')|\theta, h] \triangleq \int_{\mathcal{S}} \mathcal{Q}^*(\theta', h')\mathcal{P}(\theta', h'|\theta, h, d(\theta, h))\mathrm{d}(\theta', h')$.

The proof is given in Appendix B. The aforementioned theorem says that the optimal power transmission policy exists and is deterministic and stationary, i.e., the power used at the sensor node u_k only depends on (θ_k, h_k) and e_k. Since the belief state θ_k can be updated recursively as in (16), this property facilitates the related performance analysis. The optimal deterministic and stationary policy to an average cost MDP with finite state and action spaces can be obtained by the well-established algorithms, such as value iteration, policy iteration, and linear programming approach; see, e.g.,[40, ch. 4] and [41, ch. 6]. However, it is not computationally tractable to solve (19), since neither the state space nor the action space is finite. One might apply the algorithm proposed in [42], which involves discretization of the state and action spaces. While the algorithm involving discretization may not work well when the dimension of the system (1) is large, developing efficient numerical algorithms is out of the scope of this paper and we refer the readers to [43] for numerical algorithms for POMDPs with average cost criteria. Nevertheless, Theorem 1 provides a qualitative characteristic of the optimal transmission power control rule.

V. Structural Description: Majorization Interpretation

In this section, based on the results obtained in Section IV, we borrow the technical reasoning from [14], [16], and [29], to show that the optimal transmission power allocation strategy has a symmetric and monotonic structure and the optimal estimator has a simple form for cases where the system is scalar or the system matrix is orthogonal.

Before presenting the main theorem, we introduce a notation as follows. For a policy $\mathbf{d}(d) \in \mathcal{D}_{\rm ds}$ with $d(\theta, h) = a(e)$, with a little abuse of notations, we write $a(e)$ as $a_{\theta,h}(e)$ to emphasize its dependence on the state (θ, h). We also use $a_{\theta,h}(e)$

to represent the deterministic and stationary policy $\mathbf{d}(d)$ with $d(\theta, h) = a(e)$.

We further introduce Assumption 6, to present that, we need the following definitions.

Definition 4 (Symmetry). A function $f : \mathbb{R}^n \to \mathbb{R}$ is said to be symmetric (point symmetric) about a point $o \in \mathbb{R}^n$, if, for any two points $x, y \in \mathbb{R}^n$, $\|y - o\| = \|x - o\|$ $(y - o = -x + o)$ implies $f(x) = f(y)$.

Definition 5 (Unimodality). A function $f : \mathbb{R}^n \to \mathbb{R}$ is said to be unimodal about $o \in \mathbb{R}^n$ if $f(o) \geq f(o + \alpha_0 v) \geq f(o + \alpha_1 v)$ holds for any $v \in \mathbb{R}^n$ and any $\alpha_1 \geq \alpha_0 \geq 0$.

For the symmetry and unimodality defined previously, if the point o is not specified, it is assumed to be the origin $\mathbf{0}$ by default.

Assumption 6: The pdf of the system noises $\mathfrak{f}_w$ is symmetric and unimodal.

According to Theorem 1, to solve (18), we can restrict the optimal policy to be deterministic and stationary without any performance loss. The following theorem suggests that the optimal policy can be further restricted to be a specific class of functions.

Theorem 2: Suppose Assumption 6 holds. Let A in (1) be either a scalar or an orthogonal matrix. Then, there exists an optimal deterministic and stationary policy $a^*_{\theta,h}(e)$ such that $a^*_{\theta,h}(e)$ is symmetric and monotonically increasing with respect to $\|e\|$, i.e., for any given $(\theta, h) \in \mathcal{S}$, there holds

1) $a^*_{\theta,h}(e) = a^*_{\theta,h}(-e)$ for all $e \in \mathbb{R}^n$;
2) $a^*_{\theta,h}(e_1) \geq a^*_{\theta,h}(e_2)$ when $\|e_1\| \geq \|e_2\|$ with equality for $\|e_1\| = \|e_2\|$.

The proof is given in Appendix C. Note that Theorem 2 does not require a symmetric initial distribution $\mathfrak{f}_{x_0}$. Intuitively, this is because 1) whatever the initial distribution is, the belief state will reach the very special state $\mathfrak{f}_w$ sooner or later; and 2) we focus on the long-term average cost and the cost incurred by finite transient states can be omitted.

Remark 7: When there exists only a finite number of power levels, only the norms of the thresholds used to switch the power levels are to be determined for computation of the optimal transmission power control strategy. This significantly simplifies both the offline computational complexity and the online implementation. While the online implementation simplification is straightforward, we shall discuss more about the offline computational complexity reduction. In general, structure of feasible policies will make the search space much smaller and some specialized algorithms utilizing the structure may be developed. When it comes to our case, to apply the algorithm in [42], the discretization of the action space is not necessary. Instead, gradient-based optimization algorithms, such as simultaneous perturbation stochastic approximation algorithm [44, ch. 7], can be used to find the optimal policy.

In the following theorem, we give the optimal estimator (11) when the transmission power controller has certain symmetric structure, which includes the structure results stated in Theorem 2 as special cases. Recall that $\tau(k)$ is defined in (13) and $\mathfrak{f}_{x_0}$ is the pdf of the initial state x_0.

Theorem 3: Assume both $\mathfrak{f}_{x_0}$ and $\mathfrak{f}_w$ are point symmetric. Consider the transmission power controller $f^\sharp_k$ as

$$u_k = f^\sharp_k(x_k, \mathcal{O}^-_k) \triangleq a^\sharp_{\theta_k, h_k}(e_k)$$

where $a^\sharp_{\theta,h}(e)$ is point symmetric. Then, the optimal remote state estimator g^*_k is given by

$$\hat{x}_k = g^*_k(\mathcal{O}^+_k) = \begin{cases} x_k, & \text{if } \gamma_k = 1 \quad (20a) \\ A^{k-\tau(k)}\hat{x}_{\tau(k)}, & \text{if } \gamma_k = 0. \quad (20b) \end{cases}$$

Notice that we do not impose any constraint on the system matrix in the aforementioned theorem. Here, for the sake of space, we only present the main idea of the proof. Equation (20a) holds trivially. Moreover, if θ is point symmetric and a point symmetric power action $a(e)$ is used, given $\gamma_k = 0$, both the posttransmission belief $\theta^+_{\theta,h,a}(e)$ and the next time belief $\phi(\theta, h, a, 0)$ defined in (16) are point symmetric as well. By mathematical induction, the point symmetric structure remains if consecutive packet dropouts occur. Then by (12), the posttransmission belief of x_k is point symmetric about $A^{k-\tau(k)}\hat{x}_{\tau(k)}$, which yields (20b).

Remark 8: Let us consider related structural problems when our problem is formulated over a finite-time horizon. Using the techniques in the proof of Theorem 1, one easily verifies that an optimal deterministic policy exists (see, e.g., [39, ch. 3.3]). Then, the same structural results of the action policy as in Theorem 2 (except that the action is time dependent) can be concluded by the same arguments as in the proof of Theorem 2. Since Lemma 5 is correct regardless of the time horizon, structural results of the optimal remote estimator in Theorem 3 continue to hold.

VI. Practical Implementation

Here, we discuss about the implementation of the system, which is illustrated in Fig. 3. The optimal policy of the MDP is computed offline, and the state and its optimal action are stored as a lookup table in advance of online implementation. Depending on the storage capacity of the senor node, the system we study can work either as in (a) or in (b). The main difference between the systems in (a) and (b) is where the MDP algorithm is implemented. The content of feedback messages are correspondingly different. In (a), the MDP algorithm is implemented at the remote estimator and the action l_k is fed back to the sensor. In practice, for a generic l_k, only an approximate version (e.g., lookup tables) can be transmitted due to bandwidth limitation. An accurate feedback of l_k is possible if l_k has a special structure. For example, if $\mathcal{U}$ is a finite set, by Theorem 2, $a_k(e)$ (recall that that $a_k(e) = l_k(e + A^{k-\tau(k)}x_{\tau(k)})$) is a monotonic step function. Then, only those points, where a_k jumps, are needed to represent l_k (note that $A^{k-\tau(k)}x_{\tau(k)}$ is available at the sensor node). Since the function l_k is directly fed back to the sensor, the only task carried out by the sensor is computing $l_k(x_k)$. When the sensor node is capable of storing the MDP algorithm locally, the system can be implemented as illustrated in (b). In this case, only γ_k (a binary variable) is fed back. Note that when γ_k is fed back, the sensor knows exactly the information

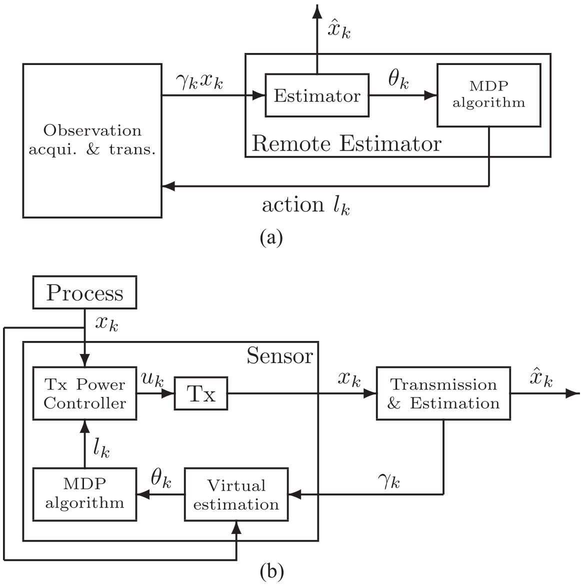


Fig. 3. Implementation of the system. The block "Observation acqui. & trans." in (a) corresponds to the blue-dashed rectangle in Fig. 1 and the block "Transmission & estimation" the red-dashed rectangle. In (a), the MDP algorithm is implemented at the remote estimator and the action l_k is fed back to the sensor. While in (b), the MDP algorithm is implemented at the sensor node and γ_k is fed back by the remote estimator.

available at the remote estimator. It can run a virtual estimator locally that has the same behavior as the remote estimator.

VII. Conclusion and Future Work

In this paper, we studied the remote estimation problem where the sensor communicates with the remote estimator over a fading channel. The transmission power control strategy, which affects the behavior of communications, as well as the remote estimator were optimally co-designed to minimize an infinite horizon cost consisting of power consumption and estimation error. We showed that when determining the optimal transmission power, the full information history available at the sensor is equivalent to its belief state. Since no constraints on the information structure are imposed and the belief state is updated recursively, the results we obtained provide some insights into the qualitative characterization of the optimal power allocation strategy and facilitate the related performance analyses. In particular, we provided some structural results on the optimal power allocation strategy and the optimal estimator, which simplifies the practical implementation of the algorithm significantly. One direction of future work is to explore the structural description of the optimal remote estimator and the optimal transmission power control rule when the system matrix is a general one. We also note that developing an efficient numerical algorithms for POMDPs with average cost is still in an early stage.

Appendix A
Generalized Skorohod Space [45]

Let $(\mathcal{X}, \mathbb{d}_{\mathcal{X}}(\cdot,\cdot))$ be a compact metric space and Λ be a set of homeomorphisms from $\mathcal{X}$ onto itself. Let π be a generic element of Λ, then on Λ, define the following three norms:

$$\|\pi\|_{\mathrm{s}} = \sup_{x\in\mathcal{X}} \mathbb{d}_{\mathcal{X}}(\pi x, x)$$

$$\|\pi\|_{\mathrm{t}} = \sup_{x,y\in\mathcal{X}:x\neq y} \left|\log\frac{\mathbb{d}_{\mathcal{X}}(\pi x,\pi y)}{\mathbb{d}_{\mathcal{X}}(x,y)}\right|$$

$$\|\pi\|_{\mathrm{m}} = \|\pi\|_{\mathrm{s}} + \|\pi\|_{\mathrm{t}}.$$

Note that $\|\pi\|_{\mathrm{t}} = \|\pi^{-1}\|_{\mathrm{t}}$. Let $\Lambda_{\mathrm{t}} \subseteq \Lambda$ be the group of homeomorphisms with finite $\|\cdot\|_{\mathrm{t}}$, i.e.

$$\Lambda_{\mathrm{t}} = \{\pi\in\Lambda : \|\pi\|_{\mathrm{t}} < \infty\}.$$

Note that since $\mathcal{X}$ is compact, each element in Λ_{t} also has finite $\|\cdot\|_{\mathrm{m}}$. Let $\mathcal{B}_r(\mathcal{X})$ be the set of bounded real-valued functions defined on $\mathcal{X}$, then the Skorohod distance $\mathbb{d}(\cdot,\cdot)$ for $f,g\in\mathcal{B}_r(\mathcal{X})$ is defined by

$$\mathbb{d}(f,g) = \inf_{\epsilon}\{\epsilon>0 : \exists\,\pi\in\Lambda_{\mathrm{t}} \text{ such that } \|\pi\|_{\mathrm{m}} < \epsilon \text{ and } \sup_{x\in\mathcal{X}}|f(x)-g(\pi x)|<\epsilon\}. \tag{21}$$

Let $\mathcal{W}$ be the set of all finite partitions of $\mathcal{X}$ that are invariant under Λ. Let I_Δ be the collection of functions that are constant on each cell of a partition $\Delta\in\mathcal{W}$. Then, the generalized Skorohod space on $\mathcal{X}$ are defined by

$$\mathscr{D}(\mathcal{X}) = \{f\in\mathcal{B}_r(\mathcal{X}) :\ \exists\,\Delta\in\mathcal{W},\ g\in I_\Delta \text{ such that } \mathbb{d}(f,g)=0\}. \tag{22}$$

By convention, two functions f and g with $\mathbb{d}(a,b)=0$ are not distinguished. Then, by [45, Lemma 3.4, Ths. 3.7 and 3.8], the space $\mathscr{D}(\mathcal{X})$ of the resulting equivalence classes with metric $\mathbb{d}(\cdot,\cdot)$ defined in (21) is a complete metric space. For $f\in\mathcal{B}_r(X)$ and $\Delta=\{\delta_j\}\in\mathcal{W}$, define

$$\mathbb{w}(f,\Delta) = \max_{\delta_j}\sup_{x,y}\{|f(x)-f(y)| : x,y\in\delta_j\}. \tag{23}$$

For $f\in\mathcal{B}_r(X)$, $f\in\mathscr{D}(\mathcal{X})$ if and only if $\lim_\Delta \mathbb{w}(f,\Delta)\to 0$, with the limits taken along the direction of refinements.

Appendix B
Proof of Theorem 1

Before proceeding, we give two supporting lemmas. In Lemma 3, a condition on the probability measures is provided, under which the weak convergence implies set-wise convergence. Lemma 4 shows that the packet arrival rate at each time can be uniformly lower bounded.

Lemma 3: Let μ and $\{\mu_{i,i\in\mathbb{N}}\}$ be probability measures defined on $(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n))$, where $\mathscr{B}(\mathbb{R}^n)$ denotes the Borel $\sigma-$ algebra of $\mathbb{R}^n$. Suppose they are absolutely continuous with respect to the Lebesgue measure. Then, the following holds:

$$\mu_i \xrightarrow{w} \mu \Leftrightarrow \mu_i \xrightarrow{\mathrm{sw}} \mu \tag{24}$$

where $\mu_i \xrightarrow{w} \mu$ stands for weak convergence [36], $\mu_i \xrightarrow{\mathrm{sw}} \mu$ represents set-wise convergence, i.e., for any $\mathcal{A}\in\mathscr{B}(\mathbb{R}^n)$, $\mu_i(\mathcal{A})\to\mu(\mathcal{A})$.

Proof: Notice that $\mu_i \xrightarrow{\mathrm{sw}} \mu \Rightarrow \mu_i \xrightarrow{w} \mu$ holds trivially [39, Appendix E] and in the following, we focus on the proof of $\mu_i \xrightarrow{w}$

$\mu \Rightarrow \mu_i \xrightarrow{\text{sw}} \mu$. The Borel $\sigma-$ algebra $\mathscr{B}(\mathbb{R}^n)$ can be generated by n-dimensional rectangles, i.e.

$$\mathscr{B}(\mathbb{R}^n) = \sigma(\{(x_1, y_1] \times \cdots \times (x_n, y_n] : x_j, y_j \in \mathbb{R}\}). \quad (25)$$

Since μ is absolutely continuous with respect to Lebesgue measure, all the rectangles are μ-continuity sets. By the Portmanteau Theorem [36], for any $x_j, y_j \in \mathbb{R}$

$$\mu_i((x_1, y_1] \times \cdots \times (x_n, y_n]) \to \mu((x_1, y_1] \times \cdots \times (x_n, y_n]).$$

Then, statement (24) follows from (25), which completes the proof.

Lemma 4: For any initial state $(\theta_1, h_1) \in \mathcal{S}$ and any policy $\mathbf{d} \in \mathcal{D}$, there exists a *uniform* lower bound $\varepsilon > 0$ such that

$$\mathbb{P}(\gamma_k = 1) \geq \varepsilon$$

holds for every $k \geq 1$.

Proof: By the saturation structure assumed for the actions space in (15), one concludes that for any $k \geq 1$

$$\mathbb{P}(\gamma_k = 1) \geq q(\bar{u}, \underline{h}) \int_{\|e\| \geq L} \theta_k(e) \mathrm{d}e.$$

Since $q(\bar{u}, \underline{h}) > 0$ by Assumption 3, we then focus on $\int_{\|e\| \geq L} \theta_k(e) \mathrm{d}e$. By the evolution of $\{\theta_k\}$ showed in (16), we prove $\int_{\|e\| \geq L} \theta_k(e) \mathrm{d}e > 0$ by cases.

If $\gamma_{k-1} = 1$, one has $\theta_k = \mathfrak{f}_w$ and $\int_{\|e\| \geq L} \mathfrak{f}_w(e) \mathrm{d}e > 0$ follows from the assumption that the support of $\mathfrak{f}_w$ is unbounded.

When $\gamma_{k-1} = 0$, we prove that for any belief θ, $\int_{\|e\| \geq L} \theta'(e) \mathrm{d}e > 0$ with $\theta' \triangleq \theta * \mathfrak{f}_w$. This is done as follows: for any $L' > 0$

$$\begin{aligned} &\int_{\|e\| \geq L} \theta'(e) \mathrm{d}e \\ &\geq \int_{\|e\| \geq L + L'} \theta(e) \mathrm{d}e \int_{\|e\| \leq L'} \mathfrak{f}_w(e) \mathrm{d}e \\ &+ \int_{\|e\| \leq L + L'} \theta(e) \mathrm{d}e \int_{\|e\| \geq 2L + L'} \mathfrak{f}_w(e) \mathrm{d}e \\ &\geq \min\left\{ \int_{\|e\| \leq L'} \mathfrak{f}_w(e) \mathrm{d}e, \int_{\|e\| \geq 2L + L'} \mathfrak{f}_w(e) \mathrm{d}e \right\}. \end{aligned} \quad (26)$$

Notice that (26) holds for any $L' > 0$, and thus, one can always find a $L' > 0$ such that $\int_{\|e\| \leq L'} \mathfrak{f}_w(e) \mathrm{d}e > 0$. Furthermore, since $\int_{\|e\| \geq 2L + L'} \mathfrak{f}_w(e) \mathrm{d}e > 0$ holds for any $L' > 0$ by the assumption that the support of $\mathfrak{f}_w$ is unbounded, one concludes that $\int_{\|e\| \geq L} \theta_k(e) \mathrm{d}e > 0$ if $\gamma_{k-1} = 0$. Notice that the aforementioned arguments do not rely on any specific policy or initial state, the uniform lower bound in Lemma 4, thus, is obtained. The proof thus is complete.

We now turn to the main body of this proof. Define

$$\mathcal{J}_\beta(\mathbf{d}, \theta, h) \triangleq \mathbb{E}^{\mathbf{d}}_{\theta, h} \left[\sum_{k=1}^{\infty} \beta^{k-1} \mathcal{C}(\theta_k, h_k, a_k) \right] \quad (27)$$

as the expected total discounted cost with the discount factor $0 < \beta < 1$. Let $v_\beta(\theta, h) \triangleq \inf_{\mathbf{d} \in \mathcal{D}} \mathcal{J}_\beta(\mathbf{d}, \theta, h)$ be the least cost associated with the initial state (θ, h), and let $m_\beta = \inf_{(\theta, h) \in \mathcal{S}} v_\beta(\theta, h)$.

By [37, Th. 3.8], in order to prove Theorem 1, it is sufficient to verify the following conditions.

C1 *(State Space)*: The state space $\mathcal{S}$ is locally compact with countable base.

C2 *(Regularity)*: Let $\mathcal{M}$ be a mapping assigning to each $s \in \mathcal{S}$ the nonempty available action space $\mathcal{A}(s)$. Then, for each $s \in \mathcal{S}$, $\mathcal{A}(s)$ is compact, and $\mathcal{M}$ is upper semicontinuous.

C3 *(Transition Kernel)*: The state transition kernel $\mathcal{P}(\cdot | s, a)$ is weakly continuous.[5]

C4 *(Cost Function)*: The one stage cost function $\mathcal{C}(s, a)$ is lower semicontinuous.

C5 *(Relative Discounted Value Function)*: There holds

$$\sup_{0 < \beta < 1} [v_\beta(\theta, h) - m_\beta] < \infty \; \forall (\theta, h) \in \mathcal{S}. \quad (28)$$

We now verify each of the aforementioned conditions for the considered problem, by which we establish the proof of Theorem 1.

A. State-Space Condition C1

We prove that both $\mathcal{S}$ and $\mathcal{A}$ are Borel subsets of Polish spaces (i.e., separable completely metrizable topological spaces) instead. Then, as pointed out in [38], by the Arsenin–Kunugui Thoerem, the condition C1 holds.

To show that $\mathcal{S}$ is a Borel subset of a Polish space, by the well-known results about the product topology [46], it suffices to prove that Θ and $\mathbb{h}$ are Borel subsets of Polish spaces. Since $\mathbb{h}$ is a compact subset of $\mathbb{R}$, we only need to prove Θ is a Borel subset of a Polish space. Let $\mathcal{M}(\mathbb{R}^n)$ be the space of probability measures on $\mathbb{R}^n$ endowed with the topology of weak convergence. It is well known that $\mathcal{M}(\mathbb{R}^n)$ is a Polish space [36]. Let $\mathcal{M}_2(\mathbb{R}^n) \subseteq \mathcal{M}(\mathbb{R}^n)$ be the set of probability measures with finite second moment, and $\mathcal{M}_{\mathrm{a}}(\mathbb{R}^n) \subseteq \mathcal{M}(\mathbb{R}^n)$ be the set of probability measures absolutely continuous with respect to $\mathfrak{L}$. By [47, Th. 3.5], $\mathcal{M}_{\mathrm{a}}(\mathbb{R}^n)$ is a Borel set. We then show that $\mathcal{M}_2(\mathbb{R}^n)$ is closed. Suppose $\{\mu_{i, i \in \mathbb{N}}\} \in \mathcal{M}_2(\mathbb{R}^n)$ and $\mu_i \xrightarrow{w} \mu$. Since $\mathcal{M}(\mathbb{R}^n)$ is complete, $\mu \in \mathcal{M}(\mathbb{R}^n)$, and using the fact that norms are continuous, by [48, Th. 1.1]

$$\int_{\mathbb{R}^n} \|e\|^2 \mu(\mathrm{d}e) \leq \liminf_{i \to \infty} \int_{\mathbb{R}^n} \|e\|^2 \mu_i(\mathrm{d}e) < \infty.$$

Then, $\mu \in \mathcal{M}_2(\mathbb{R}^n)$, implying that $\mathcal{M}_2(\mathbb{R}^n)$ is closed. Since $\Theta = \mathcal{M}_2(\mathbb{R}^n) \cap \mathcal{M}_{\mathrm{a}}(\mathbb{R}^n)$, Θ is a Borel subset of $\mathcal{M}(\mathbb{R}^n)$. The state space $\mathcal{S}$ thus is a Borel subset of a Polish space.

Now we shall show that $\mathcal{A}$ is a Borel subset of a Polish space. Considering the structure relation between $\mathcal{A}$ and $\mathcal{A}'$ in (15), we do this by proving $\mathcal{A}'$ is a Polish space. First, as *Step 1*, we

[5]We say $\mathcal{P}(\cdot | s, a)$ is weakly continuous if as $s_i \to s$ and $a_i \to a$

$$\int_{\mathcal{S}} b(s') \mathcal{P}(\mathrm{d}s' | s_i, a_i) \to \int_{\mathcal{S}} b(s') \mathcal{P}(\mathrm{d}s' | s, a)$$

for any sequence $\{(s_i, a_i), i \geq 1\}$ converging to (s, a) with $s_i, s \in \mathcal{S}$ and $a_i, a \in \mathcal{A}$, and for any bounded and continuous function $b : \mathcal{S} \to \mathbb{R}$.

show that the closure of $\mathcal{A}'$, denoted as $\mathrm{cl}(\mathcal{A}')$, is a Polish space. Then, as *Step 2*, we prove that $\mathcal{A}'$ is closed, i.e., $\mathrm{cl}(\mathcal{A}') = \mathcal{A}'$.

Step 1: Since a bounded function can be approximated by simple functions uniformly, the space $\mathcal{A}'$ is a subset of the general Skorohod space defined on $\mathcal{E}$ (see Appendix A), i.e., $\mathcal{A}' \subseteq \mathscr{D}(\mathcal{E})$. Then by [45, Th. 3.11], if (3.37) and (3.38) thereof hold, $\mathrm{cl}(\mathcal{A}')$ is compact. Since a generic $a \in \mathcal{A}'$ maps from $\mathcal{E}$ to $[0, \bar{u}]$, (3.37) thereof obviously holds. Notice that (3.38) thereof is equivalent to

$$\lim_{\Delta} \sup_{a\in\mathcal{A}'} \mathbb{w}(a, \Delta) \to 0. \tag{29}$$

By the definition of $\overline{\Delta}$, *all* the functions in $\mathcal{A}'$ are Lipschitz continuous with Lipschitz constant uniformly bounded by M on each cell of $\overline{\Delta}$. Thus, for $\Delta \preceq \overline{\Delta}$

$$\sup_{a\in\mathcal{A}'} \mathbb{w}(a, \Delta) \le M|\Delta|$$

which yields (29). Using the fact that every compact metric space is complete and separable, one obtains that $\mathrm{cl}(\mathcal{A}')$ is a Polish space.

Step 2: Suppose that $a_i \in \mathcal{A}'$ converges to a limit a in the Skorohod topology (we write as $a_i \xrightarrow{s} a$), we then show that $a \in \mathcal{A}'$. By the definition of the Skorohod distance $\mathbb{d}(\cdot,\cdot)$ in (21), $a_i \xrightarrow{s} a$ if and only if, there exist mappings $\pi_i \in \Lambda_t$ such that

$$\lim_i a_i(\pi_i x) = a(x) \text{ uniformly in } \mathcal{E}$$

$$\text{and} \quad \lim_i \pi_i x = x \text{ uniformly in } \mathcal{E}. \tag{30}$$

Since $\lim_i \pi_i x = x$ uniformly in $\mathcal{E}$, for any $\epsilon > 0$, there exists i_0 such that $\|\pi_i\|_t < \epsilon$ with $i \ge i_0$. Note that if $\|\pi_i\|_t < \epsilon$, π_i is a bi-Lipschitz homeomorphism. By the definition of $\mathcal{A}'$, any $a_i \in \mathcal{A}'$ has $\mathfrak{L}$-null set of discontinuity points. Since measure-null sets are preserved by a Lipschitz homeomorphism, by (30), one obtains that

$$\mathfrak{L}(\text{the set of discontinuity points of } a) = 0. \tag{31}$$

Following the same reasoning for one dimensional Skorohod space $\mathscr{D}[0,1]$ (see, e.g., [36, P124]), one obtains that $a_i \xrightarrow{s} a$ implies that $a_i(x) \to a(x)$ uniformly for all continuity points x of a. Since on each cell of $\overline{\Delta} = \{\delta_j\}$, *all* the functions in $\mathcal{A}'$ are Lipschitz continuous, the interior points of δ_j (write the set as δ_j^o) must be continuity points of a. By the fact that if a sequence of Lipschitz functions with Lipschitz constant uniformly bounded by M converge to a limit function, then this limit function is also a Lipschitz function with Lipschitz constant bounded by the same M, a is Lipschitz continuous with Lipschitz constant uniformly bounded by M on the interior set of each cell of $\overline{\Delta}$. For a boundary point x of the cells of $\overline{\Delta}$, denote the collection of cells whose boundary contains x as $\delta_x \triangleq \{\delta_j : x \in \partial\delta_j\}$. Then, one obtains that $a(x)$ must be a limit of a from one cell in δ_x, i.e., there exists $\delta_j \in \delta_x$ such that $\lim_{y\to x, y\in\delta_j^o} a(y) = a(x)$. Now we define a function a^* such that for each $\delta_j \in \overline{\Delta}$, $a^*(x) = a(x)$ if $x \in \delta_j^o$ and $a^*(x)$ are continuous on δ_j. Then, one obtains that $\mathbb{d}(a, a^*) = 0$, which implies that $a = a^*$ since $\mathscr{D}(\mathcal{E})$ is a metric space. Combining (31), one obtains that $a \in \mathcal{A}'$. Thus, $\mathcal{A}'$ is closed.

B. Regularity Condition C2

Since $\mathcal{A}$ is compact and $\mathcal{A}(s) = \mathcal{A}$ for every $s \in \mathcal{S}$, C2 is readily verified.

C. Transition Kernel Condition C3

Since $\mathcal{S}$ is separable and given (θ_k, h_k, a), h_{k+1} and θ_{k+1} are independent, then by [36, Th. 2.8], it suffices to prove that for any $h \in \mathbb{h}$, as $\theta_i \xrightarrow{w} \theta$ ($\mu_i \xrightarrow{w} \mu$) and $a_i \xrightarrow{s} a$, the followings hold:

$$\varphi(\theta_i, h, a_i) \to \varphi(\theta, h, a) \tag{32}$$

$$\text{and} \quad \phi(\theta_i, h, a_i, 0) \xrightarrow{w} \phi(\theta, h, a, 0). \tag{33}$$

Notice that since the set of discontinuity points of a has Lebesgue measure zero, $a_i \xrightarrow{s} a$ implies $a_i \to a$ $\mathfrak{L}-$ a.e. Furthermore, the fact that μ is absolutely continuous with respect to $\mathfrak{L}$ yields $a_i \to a$ $\mu-$ a.e. Then, it follows that $q(a_i, h) \to q(a, h)$ $\mu-$ a.e., since q is continuous $\mathfrak{L}-$ a.e. by Assumption 3. Also, by Lemma 3, $\mu_i \xrightarrow{\text{sw}} \mu$. Then by [49, Th. 2.2], one obtains that

$$\liminf_{i\to\infty} \int_{\mathbb{R}^n} q(a_i(e), h)\mu_i(\mathrm{d}e) \ge \int_{\mathbb{R}^n} q(a(e), h)\mu(\mathrm{d}e)$$

$$\text{and} \liminf_{i\to\infty} \int_{\mathbb{R}^n} -q(a_i(e), h)\mu_i(\mathrm{d}e) \ge -\int_{\mathbb{R}^n} q(a(e), h)\mu(\mathrm{d}e).$$

Combing the aforementioned two equations, one obtains that $\lim_{i\to\infty} \int_{\mathbb{R}^n} q(a_i(e), h)\mu_i(\mathrm{d}e) \to \int_{\mathbb{R}^n} q(a(e), h)\mu(\mathrm{d}e)$, i.e., $\varphi(\theta_i, h, a_i) \to \varphi(\theta, h, a)$.

We now prove that (33) holds. Noting that $\theta_i \xrightarrow{\text{sw}} \theta$ implies that $\theta_i(e) \to \theta(e)$ $\mathfrak{L}-$ a.e., it thus follows that

$$\theta^+_{\theta_i, h, a_i}(e) \to \theta^+_{\theta, h, a}(e) \tag{34}$$

$\mathfrak{L}-$ a.e. Note that $\theta^+_{\theta_i, h_i, a_i}(e)$ and $\theta^+_{\theta, h, a}(e)$ can be viewed as probability density functions of e, and for simplicity, we write the corresponding probability measures as μ_i^+ and μ^+, respectively. Then, it follows from (34) that

$$\mu_i^+ \xrightarrow{\text{sw}} \mu^+. \tag{35}$$

Let $b(e)$ be any bounded and continuous function defined on $\mathbb{R}^n$, then

$$\begin{aligned}
&\int_{\mathbb{R}^n} b(e)\phi(\theta, h, a, 0)(e)\mathrm{d}e \\
&= \int_{\mathbb{R}^n} b(e) \int_{\mathbb{R}^n} \theta^+_{\theta, h, a}(e')\mathfrak{f}_w(e - Ae')\mathrm{d}e'\mathrm{d}e \\
&= \int_{\mathbb{R}^n} \theta^+_{\theta, h, a}(e') \int_{\mathbb{R}^n} b(e)\mathfrak{f}_w(e - Ae')\mathrm{d}e\mathrm{d}e' \\
&\triangleq \int_{\mathbb{R}^n} \tilde{b}(e')\mu^+(\mathrm{d}e')
\end{aligned}$$

where $\tilde{b}(e') \triangleq \int_{\mathbb{R}^n} b(e)\mathfrak{f}_w(e - Ae')\mathrm{d}e$. Noting that $\tilde{b}(e')$ is a bounded function, then by [39, Appendix E] and (35),

$$\int_{\mathbb{R}^n} \tilde{b}(e')\mu_i^+(\mathrm{d}e') \to \int_{\mathbb{R}^n} \tilde{b}(e')\mu^+(\mathrm{d}e').$$

Equation (33) thus follows by the Portmanteau Theorem [36].

D. Cost Function Condition C4

Let $\hat{e}_+^i = \mathbb{E}_{\theta^+_{\theta_i,h,a_i}}[e]$, we first prove as $\theta_i \xrightarrow{w} \theta$ (i.e., $\mu_i \xrightarrow{w} \mu$) and $a_i \xrightarrow{s} a$

$$\hat{e}_+^i \to \hat{e}_+.$$

By (34) and [36, Th. 3.5], it remains to prove $e \sim \theta^+_{\theta_i,h,a_i}$ is uniformly integrable. We do this by showing that μ_i^+ $(\theta^+_{\theta_i,h,a_i})$ also has finite second moment given that $\mu(\theta_i)$ has finite second moment.

$$\begin{aligned}&\int_{\mathbb{R}^n} \|e\|^2 \mathrm{d}\mu_i^+(e)\\ &= \int_{\mathcal{E}} \|e\|^2 \mathrm{d}\mu_i^+(e) + \int_{\mathbb{R}^n\setminus\mathcal{E}} \|e\|^2 \mathrm{d}\mu_i^+(e)\\ &\le L^2 + \int_{\mathbb{R}^n\setminus\mathcal{E}} \|e\|^2 \mathrm{d}\mu_i(e), \text{ for any } h, a_i\\ &< \infty\end{aligned}$$

where the first inequality follows from the structure of $a_i(e)$ in (15). Since $\theta^+_{\theta_i,h,a_i}$ has finite second moment, $e \sim \theta^+_{\theta_i,h,a_i}$ is uniformly integrable.

Note that

$$\begin{aligned}C(\theta,h,a) &= \int_{\mathbb{R}^n} \theta(e) c(e,h,a) \mathrm{d}e.\\ &= \int_{\mathbb{R}^n} \alpha a(e) + (1-q(a(e),h))\|e-\hat{e}_+\|^2 \mathrm{d}\mu(e).\end{aligned}$$

Since $a_i(e) + (1-q(a_i(e),h_i))\|e_i - \hat{e}_+^i\|^2 \ge 0$, then by [49, Th. 2.2], one obtains that

$$\begin{aligned}&\int_{\mathbb{R}^n} \alpha a(e) + (1-q(a(e),h))\|e-\hat{e}_+\|^2 \mathrm{d}\mu(e)\\ &\le \liminf_{i\to\infty} \int_{\mathbb{R}^n} \alpha a_i(e) + (1-q(a_i(e),h))\|e-\hat{e}_+^i\|^2 \mathrm{d}\mu_i(e)\end{aligned}$$

which means that $C(\theta,h,a)$ is lower semicontinuous.

E. Relative Discounted Value Function Condition C5

Note that by [38, Lemma 5], if

$$\inf_{\mathbf{d},\theta,h} \mathcal{J}(\mathbf{d},\theta,h) < \infty \tag{36}$$

then (28) can be equivalently written as

$$\limsup_{\beta\uparrow 1} [\upsilon_\beta(\theta,h) - m_\beta] < \infty \; \forall (\theta,h) \in \mathcal{S}. \tag{37}$$

Step 1: Verification of (36). Consider a suboptimal policy, denoted by $\mathbf{d}^\diamond$, where at each time instant the maximal transmission power $\bar{u}$ is used. Given a belief θ, denote by $\mathtt{Var}(\theta)$ the second central moment, i.e.

$$\mathtt{Var}(\theta) = \int_{\mathbb{R}^n} \theta(e)(e-\hat{e})(e-\hat{e})^\top \mathrm{d}e \tag{38}$$

where $\hat{e} = \mathbb{E}[e|e \sim \theta]$ is the mean. Then, for any initial state $(\theta,h) \in \mathcal{S}$, if the policy $\mathbf{d}^\diamond$ is used, one can rewrite (17) as

$$C(\theta_k,h_k,a_k) = \alpha\bar{u} + (1-q(\bar{u},h_k))\mathrm{Tr}(\mathtt{Var}(\theta_k))$$

and for any $k \ge 1$

$$\mathtt{Var}(\theta_{k+1}) = \begin{cases} A\mathtt{Var}(\theta_k)A^\top + W, & \text{if } \gamma_k = 0\\ W, & \text{otherwise}\end{cases}$$

with $\mathbb{P}(\gamma_k = 0) = 1 - q(\bar{u},h_k)$ and $\mathtt{Var}(\theta_1) = A\Sigma_0 A^\top + W$. Then, for any initial state $(\theta,h) \in \mathcal{S}$, with Assumption 5, there exists a finite upper bound $\kappa(\theta)$, which depends on the initial state θ, such that for any $k \ge 1$

$$\mathbb{E}^{\mathbf{d}^\diamond}_{\theta,h}[\mathrm{Tr}(\mathtt{Var}(\theta_k))] \le \kappa(\theta). \tag{39}$$

This relation can be shown by describing the evolution of $\mathtt{Var}(\theta_k)$) using a Markov jump linear system and Assumption 5 implies the system's stability.

Then, one obtains that

$$\begin{aligned}\inf_{\mathbf{d},\theta,h} \mathcal{J}(\mathbf{d},\theta,h) &\le \inf_{\theta,h} \mathcal{J}(\mathbf{d}^\diamond,\theta,h)\\ &< \inf_\theta \kappa(\theta) + \alpha\bar{u}\\ &< \infty.\end{aligned}$$

Step 2: Verification of (37). Define the stopping time

$$\mathbb{T}_\beta \triangleq \inf\{k \ge 1 : \upsilon_\beta(\theta_k,h_k) \le \underline{\upsilon}_\beta(\mathfrak{f}_w)\}$$

where $\underline{\upsilon}_\beta(\mathfrak{f}_w) = \min_{h\in\mathbb{h}} \upsilon_\beta(\mathfrak{f}_w,h)$. Then, by [37, Lemma 4.1], one has for any $\beta < 1$ and $(\theta,h) \in \mathcal{S}$,

$$\begin{aligned}\upsilon_\beta(\theta,h) - m_\beta &\le \underline{\upsilon}_\beta(\mathfrak{f}_w) - m_\beta\\ &+ \inf_{\mathbf{d}\in\mathcal{D}} \mathbb{E}^{\mathbf{d}}_{\theta,h}\left[\sum_{k=1}^{\mathbb{T}_\beta - 1} C(\theta_k,h_k,a_k)\right].\end{aligned} \tag{40}$$

Then, by proving the finiteness of the right-hand side of (40) as β approaches 1, we show (37) holds. First, we focus on the term $\inf_{\mathbf{d}\in\mathcal{D}} \mathbb{E}^{\mathbf{d}}_{\theta,h}\left[\sum_{k=1}^{\mathbb{T}_\beta-1} C(\theta_k,h_k,a_k)\right]$. We now prove the uniform finiteness (with respect to β) of $\mathbb{E}^{\mathbf{d}^\diamond}_{\theta,h}[\mathbb{T}_\beta]$ for any initial state (θ,h). To this end, let $h^* = \arg\min_{h\in\mathbb{h}} \upsilon_\beta(\mathfrak{f}_w,h)$ and

$$\mathbb{T}^*_\beta \triangleq \inf\{k \ge 1 : (\theta_k,h_k) = (\mathfrak{f}_w,h^*)\}.$$

Note that the dependence of $\mathbb{T}^*_\beta$ on β is due to h^*. Then, one can see that for any realization of $\{\theta_k\}$ and $\{h_k\}$

$$\mathbb{T}^*_\beta \ge \mathbb{T}_\beta$$

always holds. Note that $\{h_k\}$ evolves independently. Though $\{\theta_k\}$ depends on the realization of $\{h_k\}$, under the policy $\mathbf{d}^\diamond$, $\mathbb{P}(\theta_k = \mathfrak{f}_w) \ge q(\bar{u},\underline{h})$ with $\underline{h} \triangleq \min\{h : h \in \mathbb{h}\}$ for all $k > 1$ with any initial state θ. Based on the aforementioned two observations, we construct a *uniform* (for any $0 < \beta < 1$) upper bound of $\mathbb{E}^{\mathbf{d}^\diamond}_{\theta,h}[\mathbb{T}^*_\beta]$ as follows. Define

$$\mathcal{K}(h,h') = \min\{k > 1 : h_k = h', h_1 = h\}$$

as the first time h_k reaches h' when starting at h. Then, given the initial state h, let $\{T_i\}_{i\ge 1}$ be a sequence of independent random variables such that $\mathbb{E}[T_1] = \mathbb{E}[\mathcal{K}(h,h^*)]$ and $\mathbb{E}[T_i] = \mathbb{E}[\mathcal{K}(h^*,h^*)]$, $i > 1$. Let χ be a geometrically distributed random variable with success probability $q(\bar{u}, \underline{h})$. Then, one obtains that

$$\begin{aligned}
&\mathbb{E}^{\mathbf{d}^\diamond}_{\theta,h}[\mathbb{T}^*_\beta] \\
&\leq \mathbb{E}\left[\sum_{i=1}^{\chi} T_i\right] \\
&\leq \frac{1}{q(\bar{u},\underline{h})}\max\{\mathbb{E}[\mathcal{K}(h,h^*)],\mathbb{E}[\mathcal{K}(h^*,h^*)]\} \\
&\leq \frac{1}{q(\bar{u},\underline{h})}\max_{h,h'\in\mathbb{h}}\{\max\{\mathbb{E}[\mathcal{K}(h,h')],\mathbb{E}[\mathcal{K}(h',h')]\}\} \quad (41)\\
&< \infty \quad (42)
\end{aligned}$$

where the second inequality follows from the Wald's identity and Assumption 3 that $q(\bar{u}, \underline{h}) > 0$, and the last inequality follows from the assumption that $\mathbb{h}$ is a finite set and Assumption 1-5). Note that since (41) is independent of β, $\mathbb{E}^{\mathbf{d}^\diamond}_{\theta,h}[\mathbb{T}^*_\beta]$ is uniformly bounded. One thus obtains that for any $(\theta, h) \in \mathcal{S}$

$$\begin{aligned}
&\limsup_{\beta\uparrow 1}\inf_{\mathbf{d}\in\mathcal{D}}\mathbb{E}^{\mathbf{d}}_{\theta,h}\left[\sum_{k=1}^{\mathbb{T}_\beta-1} C(\theta_k,h_k,a_k)\right] \\
&\leq \limsup_{\beta\uparrow 1}\mathbb{E}^{\mathbf{d}^\diamond}_{\theta,h}\left[\sum_{k=1}^{\mathbb{T}_\beta-1} C(\theta_k,h_k,a_k)\right] \\
&\leq \limsup_{\beta\uparrow 1}\mathbb{E}^{\mathbf{d}^\diamond}_{\theta,h}\left[\sum_{k=1}^{\mathbb{T}^*_\beta-1} C(\theta_k,h_k,a_k)\right] \\
&\leq \limsup_{\beta\uparrow 1}\mathbb{E}^{\mathbf{d}^\diamond}_{\theta,h}[\mathbb{T}^*_\beta-1](\kappa(\theta)+\alpha\bar{u}) \\
&< \infty \quad (43)
\end{aligned}$$

where the last second inequality follows from the Wald's identity and the last inequality follows from (42).

We now turn to the term $\underline{v}_\beta(\mathfrak{f}_w) - m_\beta$ and we shall show that $\limsup_{\beta\uparrow 1}(\underline{v}_\beta(\mathfrak{f}_w) - m_\beta) < \infty$. Notice that if $\arg\min_{(\theta,h)\in\mathcal{S}} v_\beta(\theta, h) = (\mathfrak{f}_w, h)$ for some $h \in \mathbb{h}$, then by the definition of $\underline{v}_\beta(\mathfrak{f}_w)$, there holds $\underline{v}_\beta(\mathfrak{f}_w) - m_\beta = 0$. In the following, we then focus on the cases when it is possible that $\arg\min_{(\theta,h)\in\mathcal{S}} v_\beta(\theta, h) \neq (\mathfrak{f}_w, h)$ for any $h \in \mathbb{h}$.

To proceed, with a little abuse of notation, define

$$\mathcal{K}(\theta,\theta') = \min\{k > 0 : \theta_{k+1} = \theta', \theta_1 = \theta\}. \quad (44)$$

By Lemma 4, one sees that for any $k > 1$, $\mathbb{P}(\mathcal{K}(\theta, \mathfrak{f}_w) > k) \leq (1-\varepsilon)^k$, which means that $\sum_{k=1}^{\infty}\mathbb{P}(\mathcal{K}(\theta,\mathfrak{f}_w) = k) = 1$ holds, for any initial state (θ, h) with $\theta \neq \mathfrak{f}_w$ and policy. Then, together with the definition of $\underline{v}_\beta(\mathfrak{f}_w)$, nonnegativity of the cost at each stage and the principle of optimality for dynamic programming, it yields that for any initial state (θ, h) and any $0 < \beta < 1$

$$\begin{aligned}
v_\beta(\theta,h) &\geq \underline{v}_\beta(\mathfrak{f}_w)\sum_{k=1}^{\infty}\mathbb{P}(\mathcal{K}(\theta,\mathfrak{f}_w)=k)\beta^k \\
&\geq \underline{v}_\beta(\mathfrak{f}_w)\sum_{k=1}^{\infty}\varepsilon(1-\varepsilon)^{k-1}\beta^k
\end{aligned}$$

where the second inequality follows from Lemma 4: for any $k_0 \geq 1$, there holds $\sum_{k=1}^{k_0}\mathbb{P}(\mathcal{K}(\theta,\mathfrak{f}_w) = k) \geq \sum_{k=1}^{k_0}\varepsilon(1-\varepsilon)^{k-1}$. Furthermore, the arguments in *Step 1* [specifically, (39)] give that $\underline{v}_\beta(\mathfrak{f}_w) \leq 1/(1-\beta)\,(\kappa(\mathfrak{f}_w) + \alpha\bar{u})$. Then, one obtains that

$$\begin{aligned}
&\limsup_{\beta\uparrow 1}(\underline{v}_\beta(\mathfrak{f}_w) - m_\beta) \\
&\leq \limsup_{\beta\uparrow 1}\underline{v}_\beta(\mathfrak{f}_w)\left(1-\sum_{k=1}^{\infty}\varepsilon(1-\varepsilon)^{k-1}\beta^k\right) \\
&\leq \limsup_{\beta\uparrow 1}\frac{1-\sum_{k=1}^{\infty}\varepsilon(1-\varepsilon)^{k-1}\beta^k}{1-\beta}(\kappa(\mathfrak{f}_w)+\alpha\bar{u}) \\
&= \lim_{\beta\to 1}(\kappa(\mathfrak{f}_w)+\alpha\bar{u})\sum_{k=1}^{\infty}k\varepsilon(1-\varepsilon)^{k-1}\beta^{k-1} \quad (45)\\
&< \infty \quad (46)
\end{aligned}$$

where the equality applies the L'Hospital's Rule and holds because the power series is differentiable (and continuous of course, which implies that $\limsup_{\beta\uparrow 1}\sum_{k=1}^{\infty}\varepsilon(1-\varepsilon)^{k-1}\beta^k = \sum_{k=1}^{\infty}\varepsilon(1-\varepsilon)^{k-1} = 1$) at the interior points of the convergence domain $(-1/(1-\varepsilon), 1/(1-\varepsilon))$; the last inequality follows from that the convergence domain for the power series in (45) is also $(-1/(1-\varepsilon), 1/(1-\varepsilon))$.

Therefore, one obtains that for any $(\theta, h) \in \mathcal{S}$

$$\limsup_{\beta\uparrow 1}[v_\beta(\theta,h) - m_\beta] < \infty$$

by (43) and (46), and the relation that $\limsup_{\beta\uparrow 1}[v_\beta(\theta,h) - m_\beta] \leq \limsup_{\beta\uparrow 1}(\underline{v}_\beta(\mathfrak{f}_w) - m_\beta) + \limsup_{\beta\uparrow 1}\inf_{\mathbf{d}\in\mathcal{D}}\mathbb{E}^{\mathbf{d}}_{\theta,h}\left[\sum_{k=1}^{\mathbb{T}_\beta-1}C(\theta_k,h_k,a_k)\right]$. The condition (relative discounted value function) thus is verified.

The proof of Theorem 1 now is complete.

Appendix C
Proof of Theorem 2

We first give some supporting definitions and lemmas as follows.

Definition 6: For any given Borel measurable set $\mathcal{B} \subset \mathbb{R}^n$, where $\mathfrak{L}(\mathcal{B}) < \infty$, we denote the symmetric rearrangement of $\mathcal{B}$ by $\mathcal{B}^\sigma$, i.e., $\mathcal{B}^\sigma$ is a ball centered at 0 with the Lebesgue measure $\mathfrak{L}(\mathcal{B})$. For a given integrable, nonnegative function $f : \mathbb{R}^n \to \mathbb{R}$, we denote the symmetric nonincreasing rearrangement of f by f^σ, where f^σ is defined as

$$f^\sigma(x) \triangleq \int_0^\infty \mathbb{1}_{\{o\in\mathbb{R}^n : f(o)>t\}^\sigma}(x)\mathrm{d}t.$$

Definition 7: For any given two integrable, nonnegative functions $f, g : \mathbb{R}^n \to \mathbb{R}$, we say that f majorizes g, which is denoted as $g \prec f$, if the following conditions hold:

$$\int_{\|x\|\leq t} f^\sigma(x)\mathrm{d}x \geq \int_{\|x\|\leq t} g^\sigma(x)\mathrm{d}x \quad \forall t \geq 0 \quad (47)$$

and

$$\int_{\mathbb{R}^n} f(x)\mathrm{d}x = \int_{\mathbb{R}^n} g(x)\mathrm{d}x.$$

Equivalently, (47) can be altered by the following condition: for any Borel set $\mathcal{B} \subset \mathbb{R}^n$, there always exists another Borel set $\mathcal{B}'$ with $\mathfrak{L}(\mathcal{B}') = \mathfrak{L}(\mathcal{B})$ such that $\int_{\mathcal{B}} g(x)\mathrm{d}x \le \int_{\mathcal{B}'} f(x)\mathrm{d}x$.

Recall that L, which is introduced in (15), is the saturation threshold for actions. As in [16], we define a binary relation on the belief state space as follows.

Definition 8 (Binary Relation $\mathcal{R}$ on Belief States). For any two belief states $\theta, \theta_* \in \Theta$, we say that $\theta\mathcal{R}\theta_*$ if the following conditions hold:

1) there holds $\theta \prec \theta_*$;
2) θ_* is symmetric and unimodal about the origin point 0;
3) $\theta(e) = \theta_*(e)$ for any $e \in \mathbb{R}^n \backslash \mathcal{E}$, where $\mathcal{E} \triangleq \{e \in \mathbb{R}^n : \|e\| \le L\}$ is defined later (15).

In the following, we define a symmetric increasing rearrangement of an action $a \in \mathcal{A}$, which preserves the average power consumption and successful transmission probability.

Definition 9: For any given Borel measurable $\mathcal{B} \subset \mathbb{R}^n$, where $\mathfrak{L}(\mathcal{B}) < \infty$, we define

$$\mathcal{B}^{\sigma}_{\theta,\hat{\theta}} \triangleq \{e \in \mathbb{R}^n : \|e\| \ge r\}$$

where $\theta, \hat{\theta} \in \Theta$, and r is determined such that $\int_{\mathcal{B}} \theta(e)\mathrm{d}e = \int_{\mathcal{B}^{\sigma}_{\theta,\hat{\theta}}} \hat{\theta}(e)\mathrm{d}e$. Given an action $a \in \mathcal{A}$, define

$$a^{\sigma}_{\theta,\hat{\theta}}(e) \triangleq \int_0^{\infty} \mathbb{1}_{\{o\in\mathbb{R}^n : a(o)>t\}^{\sigma}_{\theta,\hat{\theta}}}(e)\mathrm{d}t. \tag{48}$$

It can be verified that if $\int_{\mathbb{R}^n\backslash\mathcal{E}} \theta(e)\mathrm{d}e = \int_{\mathbb{R}^n\backslash\mathcal{E}} \hat{\theta}(e)\mathrm{d}e$, $a^{\sigma}_{\theta,\hat{\theta}}(e) \in \mathcal{A}$. One also obtains that

$$\int_{\mathbb{R}^n} a(e)\theta(e)\mathrm{d}e = \int_{\mathbb{R}^n} a^{\sigma}_{\theta,\hat{\theta}}(e)\hat{\theta}(e)\mathrm{d}e \tag{49}$$

and for any h

$$\int_{\mathbb{R}^n} q(a(e),h)\theta(e)\mathrm{d}e = \int_{\mathbb{R}^n} q\left(a^{\sigma}_{\theta,\hat{\theta}}(e),h\right)\hat{\theta}(e)\mathrm{d}e.$$

Then, the following lemma follows straightforwardly.

Lemma 5: If A is a scalar or orthogonal, then $\theta\mathcal{R}\theta_*$ implies $\phi(\theta,h,a,0)\mathcal{R}\phi(\theta_*,h,a^{\sigma}_{\theta,\theta_*},0)$, where $\phi(\cdot,\cdot,\cdot,\cdot)$ is the belief update equation defined in (16).

Note that if $\theta\mathcal{R}\hat{\theta}$, then $q(a(e),h)\theta(e)\mathcal{R}q\left(a^{\sigma}_{\theta,\hat{\theta}}(e),h\right)\hat{\theta}(e)$. Then, based on (49), following the same reasoning as in [16, Lemma 15], one obtains the following lemma.

Lemma 6: If $\theta\mathcal{R}\hat{\theta}$, then the following inequality about the one stage cost holds: $\mathcal{C}(\theta,h,a) \ge \mathcal{C}(\hat{\theta},h,a^{\sigma}_{\theta,\hat{\theta}})$.

We then proceed to prove Theorem 2 in a constructive way. To be specific, we show that for any initial state (θ,h), and any deterministic and stationary policy $\mathbf{d}(d) \in \mathcal{D}_{\mathrm{ds}}$[6], there exists another policy $\hat{\mathbf{d}}(\hat{d}) \in \mathcal{D}_{\mathrm{ds}}$ with a symmetric and monotonic structure defined in Theorem 2 such that $\mathcal{J}(\hat{\mathbf{d}}(\hat{d}),\theta,h) \le \mathcal{J}(\mathbf{d}(d),\theta,h)$. Notice that by Lemma 4, for any initial state (θ,h) and policy, there holds $\mathbb{P}\left(\mathcal{K}(\theta,\mathfrak{f}_w) < \infty\right) = 1$, where $\mathcal{K}(\cdot,\cdot)$ is defined in (44). Hence, without loss of generality, we assume that the initial state $\theta = \mathfrak{f}_w$. Let $d(\theta,h) = a_{\theta,h}(e)$, then under the policy $\mathbf{d}(d)$, the evolution of belief states is illustrated in Fig. 4. Notice that the evolution of channel gains is independent of action a, we thus assume the channel gain to be a constant h in Fig. 4 for simplicity of presentation. Notice also that the notation θ^i is different from θ_k: θ^i denotes an element in Θ, while θ_k is the belief state of the MDP at time instant k. Let $\hat{d}(\theta,h) \triangleq \hat{a}_{\theta,h}(e)$, and $\hat{p}_i$ and $\hat{\theta}^i$ be the counterparts of p_i and θ^i in Fig. 4, respectively. To facilitate presentation, let $a^i \triangleq a_{\theta^i,h}$ and $\hat{a}^i \triangleq \hat{a}_{\hat{\theta}^i,h}$. Then, $\{\hat{a}^i\}_{i\in\mathbb{N}}$ are constructed as in (48) as

$$\hat{a}^i = \left(a^i\right)^{\sigma}_{\theta^i,\hat{\theta}^i}.$$

Then, by Lemmas 5 and 6, one obtains that

$$\hat{p}_i = p_i \forall i \ge 0$$
$$\hat{\theta}^0 = \theta^0 = \mathfrak{f}_w, \quad \theta^i\mathcal{R}\hat{\theta}^i \forall i \ge 1$$
$$\mathcal{C}(\theta^i,h,a^i) \ge \mathcal{C}(\hat{\theta}^i,h,\hat{a}^i) \forall i \ge 0, h \in \mathbb{h}.$$

It then follows that $\mathcal{J}(\hat{\mathbf{d}}(\hat{d}),\theta,h) \le \mathcal{J}(\mathbf{d}(d),\theta,h)$. Since $\{\hat{a}^i\}_{i\in\mathbb{N}}$ is symmetric and increasing, and $\hat{\theta}^i$ is symmetric, one concludes the results of the theorem.

[6]By Theorem 1, without any performance loss, we just focus on the class of deterministic and stationary policies $\mathcal{D}_{\mathrm{ds}}$.

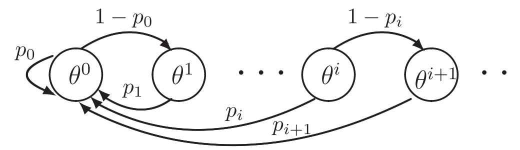


Fig. 4. Evolution of belief states under the policy $\mathbf{d}(d)$ with $d(\theta,h) \triangleq a_{\theta,h}(e)$. The special state $\theta^0 = \mathfrak{f}_w$, $p_i = \varphi(\theta^i,h,a_{\theta^i,h}) \forall i \ge 0$ is the successful transmission probability defined just above (16), and $\theta^{i+1} = \phi(\theta^i,h,a_{\theta^i,h},0) \forall i \ge 0$, where ϕ is the belief state update rule defined in (16). When the belief state is θ^i, it incurs cost $\mathcal{C}(\theta^i,h,a_{\theta^i,h})$.

References

[1] W. S. Wong and R. W. Brockett, "Systems with finite communication bandwidth-Part II: Stabilization with limited information feedback," *IEEE Trans. Autom. Control*, vol. 44, no. 5, pp. 1049–1053, May 1999.
[2] H. Ishii and B. A. Francis, "Quadratic stabilization of sampled-data systems with quantization," *Automatica*, vol. 39, pp. 1793–1800, 2003.
[3] M. Fu and L. Xie, "The sector bound approach to quantized feedback control," *IEEE Trans. Autom. Control*, vol. 50, no. 11, pp. 1698–1711, Nov. 2005.
[4] J. Hespanha, P. Naghshtabrizi, and Y. Xu, "A survey of recent results in networked control systems," *Proc. IEEE*, vol. 95, no. 1, pp. 138–162, Jan. 2007.
[5] K. You, M. Fu, and L. Xie, "Mean square stability for Kalman filtering with Markovian packet losses," *Automatica*, vol. 47, no. 12, pp. 2647–2657, 2011.
[6] B. Sinopoli, L. Schenato, M. Franceschetti, K. Poolla, M. I. Jordan, and S. S. Sastry, "Kalman filtering with intermittent observations," *IEEE Trans. Autom. Control*, vol. 49, no. 9, pp. 1453–1464, Sep. 2004.
[7] M. Huang and S. Dey, "Stability of Kalman filtering with Markovian packet losses," *Automatica*, vol. 43, pp. 598–607, 2007.
[8] A. Mainwaring, D. Culler, J. Polastre, R. Szewczyk, and J. Anderson, "Wireless sensor networks for habitat monitoring," in *Proc. Int. Workshop Wireless Sensor Netw. Appl.*, 2002, pp. 88–97.

[9] C. Yang and L. Shi, "Deterministic sensor data scheduling under limited communication resource," *IEEE Trans. Signal Process.*, vol. 59, no. 10, pp. 5050–5056, Oct. 2011.

[10] L. Zhao, W. Zhang, J. Hu, A. Abate, and C. J. Tomlin, "On the optimal solutions of the infinite-horizon linear sensor scheduling problem," *IEEE Trans. Autom. Control*, vol. 59, no. 10, pp. 2825–2830, Oct. 2014.

[11] K. J. Åström and B. Bernhardsson, "Comparison of Riemann and Lebesque sampling for first order stochastic systems," in *Proc. 41st IEEE Conf. Decision Control*, 2002, vol. 2, pp. 2011–2016.

[12] Y. Xu and J. Hespanha, "Estimation under uncontrolled and controlled communications in networked control systems," in *Proc. IEEE Conf. Decision Control Eur. Control Conf.*, Dec. 2005, pp. 842–847.

[13] J. Sijs and M. Lazar, "On event based state estimation," in *Hybrid Systems: Computation and Control*. New York, NY, USA: Springer, 2009, pp. 336–350.

[14] G. Lipsa and N. Martins, "Remote state estimation with communication costs for first-order LTI systems," *IEEE Trans. Autom. Control*, vol. 56, no. 9, pp. 2013–2025, Sep. 2011.

[15] J. Wu, Q.-S. Jia, K. H. Johansson, and L. Shi, "Event-based sensor data scheduling: Trade-off between communication rate and estimation quality," *IEEE Trans. Autom. Control*, vol. 58, no. 4, pp. 1041–1046, Apr. 2013.

[16] A. Nayyar, T. Başar, D. Teneketzis, and V. V. Veeravalli, "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.

[17] C. Ramesh, H. Sandberg, and K. H. Johansson, "Design of state-based schedulers for a network of control loops," *IEEE Trans. Autom. Control*, vol. 58, no. 8, pp. 1962–1975, Aug. 2013.

[18] A. Molin, "Optimal event-triggered control with communication constraints," Ph.D. dissertation, Technische Universität München, , München, Germany, 2014.

[19] V. Gupta, T. Chung, B. Hassibi, and R. M. Murray, "On a stochastic sensor selection algorithm with applications in sensor scheduling and dynamic sensor coverage," *Automatica*, vol. 42, no. 2, pp. 251–260, 2006.

[20] D. Han, Y. Mo, J. Wu, S. Weerakkody, B. Sinopoli, and L. Shi, "Stochastic event-triggered sensor schedule for remote state estimation," *IEEE Trans. Autom. Control*, vol. 60, no. 10, pp. 2661–2675, Oct. 2015.

[21] S. Yüksel and T. Başar, *Stochastic Networked Control Systems: Stabilization and Optimization Under Information Constraints*. New York, NY, USA: Springer, 2013.

[22] A. Gupta, S. Yuksel, T. Basar, and C. Langbort, "On the existence of optimal policies for a class of static and sequential dynamic teams," *SIAM J. Control Optimization*, vol. 53, no. 3, pp. 1681–1712, 2015.

[23] A. Goldsmith, *Wireless Communications*. Cambridge, U.K.: Cambridge Univ. Press, 2005.

[24] N. Elia, "Remote stabilization over fading channels," *Syst. Control Lett.*, vol. 54, no. 3, pp. 237–249, 2005.

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

[26] A. S. Leong, S. Dey, G. N. Nair, and P. Sharma, "Power allocation for outage minimization in state estimation over fading channels," *IEEE Trans. Signal Process.*, vol. 59, no. 7, pp. 3382–3397, Jul. 2011.

[27] 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.

[28] 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.

[29] B. Hajek, K. Mitzel, and S. Yang, "Paging and registration in cellular networks: Jointly optimal policies and an iterative algorithm," *IEEE Trans. Inf. Theory*, vol. 54, no. 2, pp. 608–622, Feb. 2008.

[30] J. Chakravorty and A. Mahajan, "Distortion-transmission trade-off in real-time transmission of Markov sources," *CoRR*, vol. abs/1412.3199, 2014. [Online]. Available: http://arxiv.org/abs/1412.3199

[31] M. F. Huber, "Optimal pruning for multi-step sensor scheduling," *IEEE Trans. Autom. Control*, vol. 57, no. 5, pp. 1338–1343, May 2012.

[32] D. Shi and T. Chen, "Optimal periodic scheduling of sensor networks: A branch and bound approach," *Syst. Control Lett.*, vol. 62, no. 9, pp. 732–738, 2013.

[33] S. Liu, M. Fardad, P. K. Varshney, and E. Masazade, "Optimal periodic sensor scheduling in networks of dynamical systems," *IEEE Trans. Signal Process.*, vol. 62, no. 12, pp. 3055–3068, Jun. 2014.

[34] A. Nayyar, "Sequential decision making in decentralized systems," Ph.D. dissertation, University of California, Berkeley, CA, USA, 2011.

[35] M. Darnel, *Theory of Lattice-Ordered Groups*. New York, NY, USA: Marcel Dekker, 1995, vol. 187.

[36] P. Billingsley, *Convergence of Probability Measures*. New York, NY, USA: Wiley, 1999.

[37] M. Schäl, "Average optimality in dynamic programming with general state space," *Math. Oper. Res.*, vol. 18, no. 1, pp. 163–172, 1993.

[38] E. A. Feinberg, P. O. Kasyanov, and N. V. Zadoianchuk, "Average cost Markov decision processes with weakly continuous transition probabilities," *Math. Oper. Res.*, vol. 37, no. 4, pp. 591–607, 2012.

[39] O. Hernández-Lerma and J. B. Lasserre, *Discrete-time Markov Control Processes: Basic Optimality Criteria*, vol. 30. New York, NY, USA: Springer, 1996.

[40] D. P. Bertsekas, *Dynamic Programming and Optimal Control, Vol II*. Belmont, MA, USA: Athena Scientific, 2007.

[41] L. I. Sennott, *Stochastic Dynamic Programming and the Control of Queueing Systems*. New York, NY, USA: Wiley, 1999, vol. 504.

[42] H. Yu and D. P. Bertsekas, "Discretized approximations for POMDP with average cost," in *Proc. 20th Conf. Uncertainty Artif. Intell.*, 2004, pp. 619–627.

[43] H. Yu, "Approximate solution methods for partially observable Markov and semi-Markov decision processes," Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, MA, USA, 2006.

[44] J. C. Spall, *Introduction to Stochastic Search and Optimization: Estimation, Simulation, and Control*. New York, NY, USA: Wiley, 2003, vol. 65.

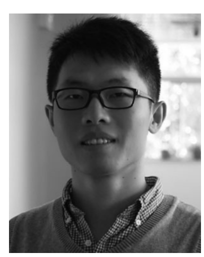
[45] M. L. Straf, "Weak convergence of stochastic processes with several parameters," in *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, Volume 2: Probability Theory*. Berkeley, CA, USA: Univ. California Press, 1972, pp. 187–221.

[46] M. Yan, *Introduction to Topology: Theory and Applications*. Beijing, China: Higher Education Press, 2010.

[47] K. Lange, "Borel sets of probability measures," *Pacific J. Math.*, vol. 48, pp. 141–161, 1973.

[48] E. A. Feinberg, P. O. Kasyanov, and N. V. Zadoianchuk, "Fatou's lemma for weakly converging probabilities," *Theory Probability Appl.*, vol. 58, no. 4, pp. 683–689, 2014.

[49] O. Hernández-Lerma and J. B. Lasserre, "Fatou's lemma and Lebesgue's convergence theorem for measures," *Int. J. Stochastic Anal.*, vol. 13, no. 2, pp. 137–146, 2000.



Xiaoqiang Ren received the B.E. degree from the Department of Control Science and Engineering, Zhejiang University, Hangzhou, China, in 2012, and the Ph.D. degree from the Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology, Kowloon, Hong Kong, in 2016.

From September to November 2016, he was a Research Associate with the Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology. He is currently a Research Fellow with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore. His research interests include sequential detection, security of cyber-physical systems, and networked control and estimation.

Junfeng Wu received the B.Eng. degree from the Department of Automatic Control, Zhejiang University, Hangzhou, China, in 2009 and the Ph.D. degree in electrical and computer engineering from the Hong Kong University of Science and Technology, Kowloon, Hong Kong, in 2013.

From September to December 2013, he was a Research Associate with the Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology. He is currently a Postdoctoral Researcher with Autonomic Complex Communication Networks, Signals and Systems Linnaeus Center, School of Electrical Engineering, KTH Royal Institute of Technology, Stockholm, Sweden. His research interests include networked control systems, state estimation, and wireless sensor networks, multiagent systems.

Dr. Wu received the Guan Zhao-Zhi Best Paper Award at the 34th Chinese Control Conference in 2015. He was selected to the national "1000-Youth Talent Program" of China in 2016.

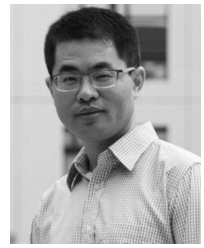
Karl Henrik Johansson (F'xx) received the M.Sc. and Ph.D. degrees in electrical engineering from Lund University, Lund, Sweden.

He is the Director of the Stockholm Strategic Research Area ICT The Next Generation and a Professor with the School of Electrical Engineering, KTH Royal Institute of Technology, Stockholm, Sweden. He has held visiting positions at University of California, Berkeley, California Institute of Technology, Nanyang Technological University, Hong Kong University of Science and Technology Institute of Advanced Studies, and Norwegian University of Science and Technology. His research interests include networked control systems, cyber-physical systems, and applications in transportation, energy, and automation.

Dr. Johansson is a member of the IEEE Control Systems Society Board of Governors and the European Control Association Council. He has received several best paper awards and other distinctions, including a ten-year Wallenberg Scholar Grant, a Senior Researcher Position with the Swedish Research Council, the Future Research Leader Award from the Swedish Foundation for Strategic Research, and the triennial Young Author Prize from the International Federation of Automatic Control. He is the IEEE Distinguished Lecturer.

Guodong Shi (M'15) received the Ph.D. degree in systems theory from the Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, China, in July 2010.

From August 2010 to April 2014, he was a Postdoctoral Researcher with the Autonomic Complex Communication Networks, Signals and Systems Linnaeus Centre, KTH Royal Institute of Technology, Stockholm, Sweden. Since May 2014, he has been with the Research School of Engineering, The Australian National University, Canberra, Australia, where he is currently a Senior Lecturer and Future Engineering Research Leadership Fellow. His current research interests include distributed control systems, quantum networking and decisions, and social opinion dynamics.



Ling Shi received the B.S. degree in electrical and electronic engineering from Hong Kong University of Science and Technology, Kowloon, Hong Kong, in 2002, and the Ph.D. degree in control and dynamical systems from the California Institute of Technology, Pasadena, CA, USA, in 2008.

He is currently an Associate Professor with the Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology. His research interests include cyber-physical systems security, networked control systems, sensor scheduling, and event-based state estimation.

Dr. Shi has been serving as a Subject Editor for the *International Journal of Robust and Nonlinear Control* from March 2015, an Associate Editor for the IEEE Transactions on Control of Network Systems from July 2016, and an Associate Editor for the IEEE Control Systems Letters from February 2017. He also served as an Associate Editor for a special issue on Secure Control of Cyber Physical Systems in the IEEE Transactions on Control of Network Systems in 2015–2017.