This is an electronic reprint of the original article.
This reprint may differ from the original in pagination and typographic detail.

Farjam, Tahmoores; Fardno, Fatemeh; Charalambous, Themistoklis

Power Allocation of Sensor Transmission for Remote Estimation Over an Unknown Gilbert-Elliott Channel

Published in:
European Control Conference 2020, ECC 2020

Published: 01/05/2020

Document Version
Peer reviewed version

Please cite the original version:

Farjam, T., Fardno, F., & Charalambous, T. (2020). Power Allocation of Sensor Transmission for Remote Estimation Over an Unknown Gilbert-Elliott Channel. In *European Control Conference 2020, ECC 2020* (pp. 1461-1467). [9143632] IEEE.

This material is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the repository collections is not permitted, except that material may be duplicated by you for your research use or educational purposes in electronic or print form. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone who is not an authorised user.

Power Allocation of Sensor Transmission for Remote Estimation Over an Unknown Gilbert-Elliott Channel

Tahmoore Farjam, Fatemeh Fardno, and Themistoklis Charalambous

Abstract—In this paper, we consider the problem of scheduling the power of a sensor when transmitting over an unknown Gilbert-Elliott (GE) channel for remote state estimation. The sensor supports two power modes, namely low power and high power, which are to be selected for transmission over the channel in order to minimize a cost on the error covariance, while satisfying the energy constraints. The remote estimator provides error-free acknowledgement/negative-acknowledgement (ACK/NACK) messages to the sensor only when low power is utilized. We first consider the Partially Observable Markov Decision Process (POMDP) problem for the case of known GE channels and derive conditions for optimality of a stationary schedule. Next, a Bayesian inference approach is used through which the channels statistics are approximately learned when they are initially unknown. An algorithm is proposed in which the sensor adjusts its scheduling policy based on the energy constraint.

Index Terms—Power allocation, remote estimation, Gilbert-Elliott channel, partially observable Markov decision process, Bayesian inference.

I. INTRODUCTION

In modern control environments wireless technologies are indispensable for communication to accomplish the control tasks. Such systems are known as Wireless Networked Control Systems (WNCSs); see, e.g., [1]. This is mainly due to advancements in sensor technology, which facilitate the use of small and low-cost sensors with sensing, actuating, communication and computing capabilities. Along with these technological advancements come a plethora of new and foreseeable applications with potentials to improve our quality of life, such as, environmental monitoring, industrial monitoring, health care, and intelligent transportations.

Recent research in WNCSs often concentrates on designing estimators and controllers of process states without taking into account the characteristics of the deployed battery-operated sensors. On the one hand, low-power transmission is desirable for sensors due to their energy limitations. On the other hand, using high-power for transmission over a fading channel increases the probability of successful reception of the data packet, thus leading to better performance. This trade off has lead to extensive research on energy-aware sensor transmission scheduling policies; see [2] for a thorough discussion on the topic. Several works considered Kalman

filtering over packet dropping wireless links focusing on the quality of estimation and control performance without taking into consideration any energy constraints; see, for example, [3]–[10] and references therein. Although fading plays a key role in networked estimation and control, the implications of energy limitations in scheduling and performance of such systems have often been ignored, apart from a limited number of works [11]–[14]. From, all these works, Qi *et al.* [14] were the first to consider the power allocation problem of sensor transmission for remote estimation, taking into consideration both the energy constraint and the temporal correlated variation of the wireless channel.

In this work, we study the setup in which a sensor is sensing a dynamical system and transmits its state estimate to a remote estimator. We assume that the sensor has two power modes: low power and high power. Additionally, the wireless channel is a GE channel with two states: good state in which packets are successfully transmitted to the remote estimator at any power level, and bad state in which packets are successfully transmitted only if the sensor transmits at high power level. The remote estimator provides feedback to the sensor by transmitting error-free ACK/NACK messages to the sensor over a separate error-free narrow-band channel. For this setup, [14] proposed the optimal dynamic scheduling policy, provided that the channels state can be determined even when using the high power mode in addition to requiring *a priori* knowledge of the channel statistics. To address more realistic scenarios, we lift these assumptions which leads to a POMDP problem and make the following contributions:

- First, we lift the assumption that the channel state can be always inferred from the ACK/NACK messages from the remote estimator. This is reasonable, since when low power is selected for transmission the sensor can infer the channel state from the ACK/NACK signal. However, when high power is used the transmission is assumed to be always successful and thus the underlying channel state cannot be determined. For this reason, we model this problem as a POMDP and solve it by exploiting the structure of the optimal policy derived in [14] and establish the conditions that guarantee its optimality in this new setting.
- Next, we consider the scenario where no *a priori* knowledge of the transition probabilities of the GE channel exists and apply Bayesian inference for learning them online and adopt a heuristic posterior sampling approach

T. Farjam and T. Charalambous are with the Department of Electrical Engineering and Automation, School of Electrical Engineering, Aalto University, Espoo, Finland. E-mails: {name.surname@aalto.fi}

F. Fardno is with the Department of Electrical Engineering, Sharif University of Technology, Azadi Street, Tehran, Iran. E-mail: fardno_fatemeh@ee.sharif.edu

to ensure computational tractability. In addition, we propose a novel method for iteratively readjusting the policy to meet the constraints despite possible inaccuracies of the learned parameters.

The remainder of the paper is organized as follows. In Section II, we provide the system model and preliminaries necessary for the development of our results. In Section III, we describe the scheduling policy and derive optimality conditions for the case of known channel parameters. In Section IV, a learning method based on Bayesian inference is proposed for near-optimal scheduling with respect to energy constraints. In Section V we demonstrate its performance and discuss its limitations and finally we draw conclusions and discuss future directions in Section VI.

Notation: Vectors and matrices are denoted by lowercase and uppercase letters, respectively. Real and nonnegative real numbers sets are denoted by $\mathbb{R}$ and $\mathbb{R}_+$, respectively. $\mathbb{S}_+^n$ is the set of n by n positive semi-definite matrices. The transpose matrix of matrix A is denoted with A^T and its inverse with A^{-1} . The notation $A \succeq 0$ ($A \succ 0$) means that matrix A is semi-positive (positive) definite. The n by n identity matrix is represented by I_n . $\mathbb{E}\{\cdot\}$ represents the expectation of its argument. For functions $f_1, f_2 : \mathbb{S}_+^n \rightarrow \mathbb{S}_+^n$, $f_1 \circ f_2$ is defined as $f_1 \circ f_2(X) \triangleq f_1(f_2(X))$.

II. PROBLEM FORMULATION

A. System model

The dynamics of the system is modeled by a linear time-invariant stochastic process governed by the following discrete-time state-space representation:

$$x_{k+1} = Ax_k + w_k, \quad (1a)$$

$$y_k = Cx_k + v_k, \quad (1b)$$

where $x_k \in \mathbb{R}^n$ is the system state, $y_k \in \mathbb{R}^m$ is the measurement taken by the sensor, A and C are known time-invariant real matrices of appropriate dimensions, and $w_k \in \mathbb{R}^n$ and $v_k \in \mathbb{R}^m$ are Gaussian random noise modeled as $w_k \sim \mathcal{N}(0, W)$ and $v_k \sim \mathcal{N}(0, V)$, where W and V denote the respective covariances. Similarly, the initial state x_0 is a zero-mean Gaussian random vector which is uncorrelated with w_k and v_k and has a positive semi-definite covariance, $P_0 \succeq 0$. We assume the pairs $(A, W^{1/2})$ and (A, C) are controllable and observable, respectively.

The sensor is capable of running a Kalman filter which provides the minimum mean squared error (MMSE) estimate of the state of the system as $\hat{x}_k^s \triangleq \mathbb{E}\{x_k | \mathcal{Y}_k^s\}$, where $\mathcal{Y}_k^s$ is a set of all measurement data collected by the sensor up to and including time k . The corresponding estimation error covariance is defined as $P_k^s \triangleq \mathbb{E}\{(x_k - \hat{x}_k^s)(x_k - \hat{x}_k^s)^T | \mathcal{Y}_k^s\}$.

Define functions $h, g : \mathbb{S}_+^n \rightarrow \mathbb{S}_+^n$ as

$$h(X) \triangleq AXA^T + W,$$

$$g(X) \triangleq X - XC^T[XCX^T + V]^{-1}CX.$$

The sensor has access to the entire measurement history and thus a unique positive semi-definite solution of $g \circ h(\bar{P}) = \bar{P}$ exists, which is the steady-state error covariance, denoted

by $\bar{P}$. By initiating from $P_0 \succeq 0$, P_k^s converges to $\bar{P}$ exponentially fast. Consequently, without loss of generality, we assume that the Kalman filter has already entered steady-state at the sensor side.

We assume that the sensor supports two transmission power levels, namely δ and Δ , where $\delta < \Delta$. At every time step k , it needs to decide the power level for transmission of the latest state estimate, $\hat{x}_k^s$, to the estimator over a time-varying channel.

B. Communication Channel

We model the communication channel connecting the sensor to the estimator as a two-state Markov chain depicted in Fig 1. This model is in accordance with most industrial applications, since it captures the temporal correlation between the variation of channel states due to presence of shadow fading and burst error [15]. The channel can be either in the good (G) or bad state (B) where the failure rate (transition from G to B) and recovery rate (transition from B to G) are denoted by p and q , respectively, where $0 \leq p, q \leq 1$. We consider the case of positively correlated channels in this work, i.e., $q \leq 1 - p$.

Fig. 1. The GE channel model where the communication channel is modeled as a two-state Markov chain.

When the channel is in G, data transmission is successful regardless of the selected power level. However, when the channel is in B, in case the data packet is transmitted with high energy level, Δ , it is successfully received by the remote estimator while the packet is dropped low energy, δ , is used for transmission. This scenario is in accordance with the *reliable data flow* offered by commercial sensors when using their highest energy level [16]. Since a sufficiently high transmission power would result in a high Signal to Noise Ratio (SNR) regardless of the channel condition, the energy level Δ can be determined such that this assumption holds.

C. Remote State Estimation

Let $\mathbf{A} \triangleq \{0, 1\}$ denote the action space and $a_k \in \mathbf{A}$ denote the energy level selected by the sensor for transmission at time k , which is defined as

$$a_k = \begin{cases} 1, & \hat{x}_k^s \text{ is transmitted with energy } \Delta, \\ 0, & \text{otherwise.} \end{cases} \quad (2)$$

We assume perfect feedback channel is available from the estimator to the sensor and define a binary variable γ_k as the indicator of packet reception at k such that $\gamma_k = 1$ if $\hat{x}_k^s$ is received and $\gamma_k = 0$, otherwise. Let θ denote the scheduling scheme that determines a_k at each k . Then, the information available at the estimator side can then be described by

$$\mathcal{Y}_k(\theta) = \{\hat{x}_1^s \gamma_1, \dots, \hat{x}_k^s \gamma_k\},$$

The estimator computes the MMSE state estimate, $\hat{x}_k(\theta) \triangleq \mathbb{E}\{x_k | \mathcal{Y}_k(\theta)\}$, and the corresponding error covariance, $P_k(\theta) \triangleq \mathbb{E}\{(x_k - \hat{x}_k)(x_k - \hat{x}_k)^T | \mathcal{Y}_k(\theta)\}$, recursively:

$$\hat{x}_k(\theta) = A^{t_k} \hat{x}_{k-t_k}^s, \quad P_k(\theta) = h^{t_k}(\bar{P}),$$

where $t_k \triangleq k - \max_{1 \leq t \leq k} \{t : \gamma_t = 1\}$ denotes the time elapsed since the last successful packet reception.

D. Problem of interest

In this work, we use the trace of the error covariance matrix as the performance metric. By considering the performance over the infinite horizon, the average expected value of this metric for a given schedule θ is considered as the objective

$$P_a(\theta) = \limsup_{T \rightarrow \infty} \frac{1}{T} \mathbb{E} \left\{ \sum_{k=1}^T \text{tr}(P_k(\theta)) \right\}. \quad (3)$$

Furthermore, the average energy cost incurred by θ is given by

$$J_a(\theta) = \limsup_{T \rightarrow \infty} \frac{1}{T} \mathbb{E} \left\{ \sum_{k=1}^T (1 - a_k) \delta + a_k \Delta \right\}.$$

We intend to find θ which minimizes (3) such that $J_a(\theta) \leq L$, where $\delta \leq L \leq \Delta$ is a given energy budget. This problem is formally formulated as

$$\min_{\theta} P_a(\theta), \quad \text{s.t. } J_a(\theta) \leq L. \quad (4)$$

III. SCHEDULING OVER KNOWN GE CHANNELS

Assuming that the transition probabilities of the GE model are known and the state of the channel is observed at every time step, problem (4) can be solved by considering the equivalent constrained MDP formulation [14]. It is reasonable to assume that the channel state is observed when low power is selected for transmission since the sensor can infer this from the ACK/NACK signal. However, using high power will always result in successful transmission and thus the underlying channel state cannot be determined, unless Channel State Information (CSI) acquisition is available. Nonetheless, due to the dynamic nature of the considered process and the changing environment, the coherence time of the channel is relatively small and thus fast fading occurs, rendering it impossible to have instantaneous CSI acquisition. In this section, we first present the structure of the optimal policy derived for the constrained MDP problem. Then, we consider the realistic scenario of partial observations and exploit the existing results for solving the POMDP problem and derive conditions that guarantee optimality of this policy.

A. Optimal schedule for fully observable channels

By considering the case where state of the channel is observable regardless of the selected energy level by the sensor, problem (4) can be transformed into a constrained MDP. The state space of this MDP is given by $\mathcal{S} = \{G, B\} \times \{\bar{P}, h(\bar{P}), h^2(\bar{P}), \dots\}$. The data packet is guaranteed to be successfully transmitted in case the channel

is in G and thus the states $\{G, h^i(\bar{P})\}$ are non-existent for $i \geq 1$. As a result, the state space can be defined as $\mathcal{S} = \{h^{-1}(\bar{P}), h^0(\bar{P}), h(\bar{P}), \dots\}$, where $h^{-1}(\bar{P})$ corresponds to the case where the channel is in state G and $h^{-1}(\bar{P}) = \bar{P}$, whereas $h^i(\bar{P})$ for $i \geq 0$ refers to i consecutive packet dropouts while the channel is in state B.

It has been proven in [14] that a stationary schedule exists which solves (4) and by investigating the stationary distribution of the induced Markov chain, the optimal schedule is derived. By adopting this schedule, denoted by θ^* , the sensor continuously transmits with low power until the Markov chain reaches state i^* . Then, the sensor switches to high power with probability μ_{i^*} . If the Markov chain hits state $i^* + 1$, however, Δ is used for transmission with probability one, which ensures successful packet reception. Under this schedule, the decision variable (2), i.e., action, is given by

$$\mathbb{P}\{a_i(\theta^*(i^*)) = 1\} = \begin{cases} 0, & \text{if } i < i^*, \\ \mu_{i^*}, & \text{if } i = i^*, \\ 1, & \text{otherwise,} \end{cases} \quad (5)$$

where μ_{i^*} is computed as

$$\mu_{i^*} = \begin{cases} \frac{E - [1 + E(1 - q)^2] (1 - q)^{i^*}}{[1 + E(1 - q)^2] q (1 - q)^{i^* - 1}}, & \text{if } i^* > 0, \\ \frac{E [1 - (1 - q)^2] - 1}{Eq(1 - q)}, & \text{if } i^* = 0, \\ p(E - 1/q), & \text{if } i^* = -1, \end{cases} \quad (6)$$

where $E = [(p + q)(L - \delta)]/[pq(\Delta - \delta)]$.

B. Scheduling with partial observations

In the scenario considered here, if $a_k = 1$, i.e., the sensor transmits the data packet with high energy Δ at time k , the channel could either be in G or B. By exploiting the knowledge of the exact values of p and q , however, the probability of the channel being in each state can be determined. We define *belief* as the probability of the channel being in G and at a given time k denote it by b_k . If $a_k = 0$, the state of the channel is observed and the belief is updated as

$$b_{k+1} = \begin{cases} 1 - p, & \text{if } s_k = G, \\ q, & \text{if } s_k = B, \end{cases}$$

where $s_k \in \{G, B\}$ denotes the state of channel at time k . In case high energy is used for transmission, however, the belief at the next step is a function of belief at the current step as well,

$$b_{k+1} = b_k(1 - p) + (1 - b_k)q.$$

It has been shown that POMDP can be converted to a corresponding MDP with the belief as a state, since b is a sufficient statistic for decision making given the past action history and observation history [17]. In this work, however, due to the specific structure of our problem, the optimal policy for MDP can be applied in this new setting and is optimal under stricter energy constraints.

Adopting any stationary scheduling policy of the form

$$\mathbb{P}\{a_i(\theta) = 1\} = \begin{cases} 0, & \text{if } i < 1, \\ \mu_i, & \text{if } 1 \leq i \leq j, \\ 1, & \text{if } i > j, \end{cases} \quad (7)$$

where $\mu_i \in [0, 1]$ and $j \in \mathbb{N}$, for this POMDP will lead to the same results as the fully observable scenario. The sensor will eventually transmit with high power at some time k when $i \geq 1$. As a result, at $k+1$, it can be deduced that $i \in \{-1, 0\}$, since successful packet transmission is guaranteed and the error covariance is reset to $\bar{P}$. Although the exact state of the Markov chain is unknown due to transmission with high energy, the sensor will definitely use low energy for transmission at $k+1$ according to (7). Consequently, the missing observations when $i \in \{-1, 0\}$ have no effect on the performance of this policy and it leads to the same result as the MDP counterpart.

Since for $i^* \geq 1$ the optimal policy in (5) has the same structure as (7), $\theta^*(i^*)$ will be optimal for the POMDP considered here as well. To derive the optimality condition, we use the result of the following lemma [14, Lemma 4.1]:

Lemma 1. Define a schedule $\bar{\theta}(\bar{i})$ such that

$$a_i(\bar{\theta}(\bar{i})) = \begin{cases} 0, & \text{if } i < \bar{i}, \\ 1, & \text{if } i \geq \bar{i}. \end{cases} \quad (8)$$

Under this schedule, $J_a(\bar{\theta}(\bar{i}))$ is a monotonically decreasing function of $\bar{i}$ and for $\bar{i} \geq 1$

$$J_a(\bar{\theta}(\bar{i})) = \delta + \frac{pq(\Delta - \delta)}{(p+q)[(1-q)^{1-m} - (1-q)^2]} \quad (9)$$

By comparing $\bar{\theta}(\bar{i})$ with the schedule in (5), it is easy to verify that $J_a(\bar{\theta}(\bar{i})) \geq J_a(\theta^*(i^*))$ for all $\bar{i} = i^*$. Hence, if the energy budget is such that

$$\delta \leq L \leq \delta + \frac{pq(\Delta - \delta)}{(p+q)[1 - (1-q)^2]}, \quad (10)$$

then $i^* \geq 1$ and thus $\theta^*(i^*)$ is the optimal for the POMDP.

Remark 1. Despite the fact that the proposed optimal schedule is only applicable when the energy budget satisfies (10), for larger L , the policy $\bar{\theta}(1)$ can be adopted. Although this results in suboptimal performance due to lower energy consumption than the given budget, it will guarantee satisfaction of the energy constraint.

IV. SCHEDULING OVER UNKNOWN GE CHANNELS

The policy given in (5) is optimal for the POMDP so long as the energy budget satisfies (10). Calculation of (6) for implementing this policy requires *a priori* knowledge of the exact values of recovery and failure rate of the channel. This is, however, a very strong assumption which restricts implementation in practical scenarios. In this section, we consider the case where the transition probabilities are unknown and adopt a Bayesian framework for learning them. We then adopt a heuristic posterior sampling algorithm to ensure computational tractability, and extract the optimal policy

from the learning outcome. Since the system is partially observable, the learned transition probabilities are (possibly) inaccurate and the policy is no longer optimal, leading either to violation of the energy constraint or to underuse of the available energy resources. Hence, we propose a novel method, with which we adapt the policy accordingly to meet the energy constraints and achieve near-optimal performance.

A. A Bayesian framework

The underlying channel parameters, p and q , are on the interval $[0, 1]$. As a result, since Beta distribution is the conjugate prior for Bernoulli distributions, we assume the prior distribution of the channel parameters follow the Beta distribution parameterized by $\Phi \triangleq [\phi_1 \ \phi_2 \ \phi_3 \ \phi_4] \in \mathbb{Z}_+^4$ and due to their independence we derive

$$\mathbb{P}(p, q; \Phi) = \mathbb{P}(p; \phi_1, \phi_2) \mathbb{P}(q; \phi_3, \phi_4),$$

where

$$\mathbb{P}(p; \phi_1, \phi_2) = \frac{p^{\phi_1-1}(1-p)^{\phi_2-1}}{B(\phi_1, \phi_2)}, \quad (11)$$

$$\mathbb{P}(q; \phi_3, \phi_4) = \frac{q^{\phi_3-1}(1-q)^{\phi_4-1}}{B(\phi_3, \phi_4)}, \quad (12)$$

where B denotes the Beta function. As a result of this particular choice of prior distribution which is parametrized by posterior counts (ϕ_1, ϕ_2) for p and similarly, (ϕ_3, ϕ_4) for the unknown recovery rate q , the posterior update is easily computed after obtaining new observations; see [18].

Without loss of generality, we assume that each posterior count is initially set to 1, i.e., $\Phi = [1 \ 1 \ 1 \ 1]$. The channel parameters are estimated as the mean of the respective distribution, i.e., $p = \phi_1/(\phi_1 + \phi_2)$ and $q = \phi_3/(\phi_3 + \phi_4)$. Let $z_k \in \{G, B, V\}$ denote the observation at time k and denote the observation history as $z^k = \{z_1, \dots, z_k\}$. Similarly, the channel state history is denoted as $s^k = \{s_1, \dots, s_k\}$. No observation of the channel state is represented by $z_k = V$ which corresponds to $a_k = 1$. Moreover, $z_k \in \{G, B\}$ corresponds to observing the channel state due to transmission with low energy, i.e., $a_k = 0$. Hence, the action history can be inferred from the observation history.

When $z_k = V$ the channel state can be either G or B, i.e., $s_k \in \{G, B\}$. We denote the set of all state histories that can lead to the same observation history by $S(z^{k-1})$ which is defined as

$$S(z^{k-1}) \triangleq \{s^{k-1} | s_t = z_t, \forall t \in \{\tau | z_\tau \neq V\}\}.$$

Multiple state histories can lead to the same posterior count. For instance, assume at k a state history consists of c_1, c_2, c_3 , and c_4 number of transitions from G to B , G to G , B to G , and B to B , respectively, which yields

$$\mathbb{P}(s^k | p, q) \mathbb{P}(p, q) = \mathbb{P}(s^k | p, q) = p^{c_1}(1-p)^{c_2} q^{c_3}(1-q)^{c_4}.$$

Consequently, since the posterior count is independent of the order of occurrence of the state transitions, multiple state histories can induce the same probability distribution on p and q . In this regard, we define *appearance count*, denoted

Fig. 2. Graphical representation of Bayesian inference where the contents of each rectangle denote the channel state, posterior count and appearance count, respectively. Due to transmission with high energy at k , the channel state is not observed which leads to an increase in the number of possible posteriors as captured in (14).

by $\Psi(\Phi, S(z^{k-1}), s_k)$, as the total number of state histories up to k which lead to the same posterior count Φ .

We consider the joint probability distribution of the state and transition probabilities given the observation history, i.e., $\mathbb{P}(s_k, p, q|z^{t-1})$, which can be fully described by the posterior count and appearance count as [19]

$$\mathbb{P}(s_k, p, q|z^{k-1})\mathbb{P}(z^{k-1}) = \sum_{\Phi} \Psi(\Phi, S(z^{k-1}), s_k) p^{\phi_1-1} (1-p)^{\phi_2-1} q^{\phi_3-1} (1-q)^{\phi_4-1}, \quad (13)$$

where $\mathbb{P}(z^{k-1})$ is the normalization term. Similarly, as a new observation is obtained, the posterior is updated recursively as

$$\mathbb{P}(s_{k+1}, p, q|z^k) = \sum_{s_k} \mathbb{P}(s_k, p, q|z^{k-1}) \mathbb{P}(s_{k+1}, z_k | s_k, p, q) / \mathbb{P}(z_k | z^{k-1}). \quad (14)$$

When $a_k = 0$, the sensor transmits with low energy δ and the state of the channel is observed, i.e., $s_k = z_k$. As a result, the number of posteriors remains constant according to (14). However, using Δ will lead to $z_k = V$ and thus the summation in (14) is taken over both possibilities, i.e., $s_k \in \{G, B\}$. Fig. 2 provides a graphical representation of the update procedure as well as the effect of $a_k = 1$ on the growth of the number of possible posteriors.

B. Learning the channel parameters for scheduling

The aforementioned method enables incorporating the uncertainty in the transition probabilities in the decision making process in a recursive manner. At any given time, the underlying channel parameters are learned with respect to the available state and observation history. These parameters are then adopted in (6) to determine the corresponding optimal schedule. By enforcing the determined policy for power management in the current step, a new observation is obtained which is used for learning the parameters more accurately and readjusting the policy accordingly. However, since with each transmission with high energy the number of posterior counts grows, the number of possible posteriors can grow to infinity over time rendering the problem intractable. To address this issue, we adopt an approach based on the idea of approximate belief monitoring [20, Section 5.1]. In essence, this method ensures feasibility of computations by keeping only K number of posteriors during the Bayesian update by randomly drawing them with respect to the appearance count

Algorithm 1: Joint learning and scheduling design

Input: posterior sampling parameter K , initial state of the Markov chain $i = -1$, energy states δ and Δ , average energy constraint L , $\xi^G = \{[1, 1, 1, 1], 1, 0.5\}$ and $\xi^B = \{[1, 1, 1, 1], 1, 0.5\}$

```

1 for  $k \leftarrow 1$  to  $\infty$  do
2   determine the average consumed energy so far
    $J = \mathbb{E}\{(1 - a_k)\delta + a_k\Delta\}$ 
3   determine the average energy of the new policy
    $\hat{L} \leftarrow L - (J - L)$ 
4   estimate  $p, q$  w.r.t. probability of each posterior  $\Pi$ 
5   determine  $i^*$  and  $\mu_{i^*}$  using  $\hat{L}$  and estimated  $p, q$ 
6   apply  $a_i$  determined by the policy
7   if  $a_i = 0$  and  $z_k = G$  then
8     | G Update () and  $i \leftarrow -1$ 
9   else if  $a_i = 0$  and  $z_k = B$  then
10    | B Update () and  $i \leftarrow i + 1$ 
11   else
12     | G Update () and B Update () and
      |  $i \leftarrow -1$ 
13   end
14    $\xi^G \leftarrow \xi^{G2G} \cup \xi^{B2G}$  and  $\xi^B \leftarrow \xi^{G2B} \cup \xi^{B2B}$ 
15   keep  $K$  number of posteriors of  $\xi^G$  and  $\xi^B$  by
      randomly drawing w.r.t.
16 end
17 Procedure G Update ()
18    $\xi^{G2B} \leftarrow \xi^G$  and  $\xi^{G2G} \leftarrow \xi^G$ 
19    $\xi^{G2B}(\phi_1) \leftarrow \xi^G(\phi_1) + 1$  and
      $\xi^{G2B}(\Pi) \leftarrow p\xi^B(\Pi)$ 
20    $\xi^{G2G}(\phi_2) \leftarrow \xi^B(\phi_2) + 1$  and
      $\xi^{G2G}(\Pi) \leftarrow (1-p)\xi^B(\Pi)$ 
21 Procedure B Update ()
22    $\xi^{B2G} \leftarrow \xi^B$  and  $\xi^{B2B} \leftarrow \xi^B$ 
23    $\xi^{B2G}(\phi_3) \leftarrow \xi^B(\phi_3) + 1$  and
      $\xi^{B2G}(\Pi) \leftarrow q\xi^B(\Pi)$ 
24    $\xi^{B2B}(\phi_4) \leftarrow \xi^B(\phi_4) + 1$  and
      $\xi^{B2B}(\Pi) \leftarrow (1-q)\xi^B(\Pi)$ 

```

ψ . The implementation of this heuristic approach in parallel with scheduling design and power allocation is presented in Algorithm 1.

We define $\xi^G \triangleq \{[\phi_1, \phi_2, \phi_3, \phi_4], \psi, \Pi\}$ as the set of information on the posteriors of the channel being in G and let Π denote the joint probability distribution (13), which is set to 0.5 initially. Each posterior in ξ^G can transition into G or B with probability of $1 - p$ and p , respectively, where p is the mean of the corresponding Beta distribution. Similarly, we define ξ^B for the posteriors of being in the bad state. These operations are shown by the two procedures on Line 17 and Line 21. After the update, these sets are merged, shown by the operator $\cup$, such that ψ and Π of the items with identical posterior counts are summed and then only K posteriors are kept as shown in Line 14 and Line 15, respectively.

Despite the possible inaccuracy of the learned parameters due to partial observations, we can exploit the structure of optimal policy (5) by ensuring that the energy constraint is satisfied. In this regard,

- First, the average energy consumed in previous steps is computed in Line 2.
- The average energy constraint for the new policy, denoted by $\hat{J}$ is then determined according to Line 3.
- New policy is determined based on $\hat{J}$ and the estimated values for p and q which are the outcome of the posterior sampling.

This ensures that the energy consumed satisfies the actual energy constraint regardless of the accuracy of the learned parameters. Adopting the novel policy for the readjusted energy constraint results in near-optimal performance as shown in Section V.

V. NUMERICAL RESULTS

In this section, we evaluate the performance of our proposed scheduling methods by numerical simulations for a system described by

$$A = \begin{bmatrix} 1.4 & 0 \\ 0 & 1 \end{bmatrix}, C = I_2, Q = 0.5I_2, R = 0.3I_2.$$

The energy consumed by the sensor when transmitting with low and high power are assumed to be $\delta = 1$ and $\Delta = 10$, respectively. Moreover, the GE channel parameters are assumed to be $p = 0.2$ and $q = 0.4$. For these values, it follows from (10) that the schedule given in (5) will be optimal for both MDP and POMDP scenarios if the $1 \leq L \leq 2.875$.

Fig. 3 and 4 depict the average energy consumption and average trace of the error covariance, respectively, for various scenarios when the energy constraint is chosen as $L = 2$. Since this constraint is within the bounds derived in (10), when channel parameters are known, the optimal schedule proposed for MDP is expected to lead to optimal performance for the POMDP as well. This is in agreement with the obtained numerical results since the average cost for all scenarios converge to the one obtained for $\theta^*(i^*)$ without violating the constraint. For the scenario considered here the parameters of the optimal schedule are given by $i^* = 1$ and $\mu_{i^*} = 0.1026$.

It should be noted the scheduling policy is time-varying when the channel parameters are unknown since the learned parameters, denoted by $\hat{p}$ and $\hat{q}$, are time-varying. Nonetheless, the average of μ_{i^*} for the last 10000 steps is $\mu_{i^*} = 0.1065$ with $i^* = 1$. Although the deviation of the probability of switching to high energy is 4% higher for the POMDP case with unknown parameters, its performance is similar to the optimal case over time. Due to the partial observations of the channel states, the parameters are not learned accurately over time as depicted in Fig. 5. However, by readjusting the average energy consumption of the policy through the feedback mechanism proposed in Algorithm 1, the cost converges to the one obtained by the optimal setup while ensuring that the average energy constraint is not violated.

Fig. 3. Average energy consumption of the sensor for both fully observable and partially observable scenarios when $L = 2$.

Fig. 4. Performance of the proposed policies in terms of cost (3) for $L = 2$.

For the considered setup, if the energy budget violates constraint (10), POMDP cannot be treated as MDP since i^* will be smaller than 1 in the optimal policy. This scenario is investigated for $L = 3$ and Fig. 6 and 7 illustrate the results. The optimal solution of the MDP problem is given by a scheduling policy of the form (5) with $i^* = 0$ and $\mu_{i^*} = 0.1667$. To be able to treat the POMDP as the MDP counterpart, we use the policy defined in (8) with $\hat{i} = 1$ which guarantees satisfying the average energy constraint despite suboptimality in terms of average cost. This is in agreement with the numerical results. The average energy consumption for the POMDP converges to 2.875 which leads to higher cost in terms of the estimation quality. This is expected since due to lower average transmission power, the probability of packet loss is higher.

Fig. 5. Value of the learned channel parameters over time in comparison with the actual values $p = 0.2$ and $q = 0.4$.

Fig. 6. Average transmission energy used by the sensor when $L = 3$ for both fully observable and partially observable scenarios.

Fig. 7. Average of the trace of the error covariance matrix when the energy budget is $L = 3$.

VI. CONCLUSIONS AND FUTURE DIRECTIONS

A. Conclusions

We considered the problem of power allocation of a sensor with two power modes and with an energy constraint when transmitting over a GE channel for remote state estimation. We considered the scenario where the channel state is not observed when using high transmission power and derived conditions which preserve optimality of a given stationary schedule for the resulting POMDP problem. We then extended the results to the case of unknown GE channels and proposed a novel method for learning the channel statistics online while ensuring the resulting scheduling policy satisfies the energy constraint.

B. Future Directions

Part of ongoing work is to investigate how the learning approach should be modified when the ACK/NACK mechanism is not error-free, thus resulting in more unobservable states. In addition, we intend to find the optimal scheduling policy for POMDP when the optimality of the results obtained for MDP is not preserved under less strict energy constraints.

In [21] the problem of orchestrating the scheduling of a group of wireless sensors subject to energy constraints was considered, but without considering fading channels. We plan to study scheduling policies for networks of wireless sensors under known and unknown GE channels.

REFERENCES

- [1] P. Park, S. C. Ergen, C. Fischione, C. Lu, and K. H. Johansson, "Wireless network design for control systems: A survey," *IEEE Communications Surveys Tutorials*, vol. 20, no. 2, pp. 978–1013, Secondquarter 2018.
- [2] A. S. Leong, D. E. Quevedo, and S. Dey, *Optimal Control of Energy Resources for State Estimation Over Wireless Channels*, 1st ed. Springer Publishing Company, Incorporated, 2018.
- [3] B. Sinopoli, L. Schenato, M. Franceschetti, K. Poolla, M. I. Jordan, and S. S. Sastry, "Kalman filtering with intermittent observations," *IEEE Transactions on Automatic Control*, vol. 49, no. 9, pp. 1453–1464, Sept 2004.
- [4] L. Schenato, B. Sinopoli, M. Franceschetti, K. Poolla, and S. S. Sastry, "Foundations of control and estimation over lossy networks," *Proceedings of the IEEE*, vol. 95, no. 1, pp. 163–187, Jan. 2007.
- [5] M. Huang and S. Dey, "Stability of Kalman filtering with Markovian packet losses," *Automatica*, vol. 43, no. 4, pp. 598–607, 2007.
- [6] T. Farjam, T. Charalambous, and H. Wymeersch, "Timer-based distributed channel access for control over unknown unreliable time-varying communication channels," in *European Control Conference (ECC)*, June 2019, pp. 2975–2982.
- [7] A. Censi, "Kalman filtering with intermittent observations: Convergence for semi-Markov chains and an intrinsic performance measure," *IEEE Transactions on Automatic Control*, vol. 56, no. 2, pp. 376–381, Feb. 2011.
- [8] T. Charalambous, A. Ozcelikkale, M. Zanon, P. Falcone, and H. Wymeersch, "On the resource allocation problem in wireless networked control systems," in *Conference on Decision and Control (CDC)*, Dec 2017, pp. 4147–4154.
- [9] N. Xiao, L. Xie, and L. Qiu, "Feedback stabilization of discrete-time networked systems over fading channels," *IEEE Transactions on Automatic Control*, vol. 57, no. 9, pp. 2176–2189, Sep. 2012.
- [10] M. Zanon, T. Charalambous, H. Wymeersch, and P. Falcone, "Optimal scheduling of downlink communication for a multi-agent system with a central observation post," *IEEE Control Systems Letters*, vol. 2, no. 1, pp. 37–42, Jan 2018.
- [11] L. Shi, P. Cheng, and J. Chen, "Sensor data scheduling for optimal state estimation with communication energy constraint," *Automatica*, vol. 47, no. 8, pp. 1693–1698, 2011.
- [12] K. Gatsis, A. Ribeiro, and G. J. Pappas, "Optimal power management in wireless control systems," *IEEE Transactions on Automatic Control*, vol. 59, no. 6, pp. 1495–1510, June 2014.
- [13] Z. Ren, P. Cheng, J. Chen, L. Shi, and H. Zhang, "Dynamic sensor transmission power scheduling for remote state estimation," *Automatica*, vol. 50, no. 4, pp. 1235–1242, 2014.
- [14] Y. Qi, P. Cheng, and J. Chen, "Optimal sensor data scheduling for remote estimation over a time-varying channel," *IEEE Transactions on Automatic Control*, vol. 62, no. 9, pp. 4611–4617, Sep. 2017.
- [15] D. E. Quevedo, A. Ahlen, and K. H. Johansson, "State estimation over sensor networks with correlated wireless fading channels," *IEEE Trans. on Automatic Control*, vol. 58, no. 3, pp. 581–593, March 2013.
- [16] J.-J. Xiao, S. Cui, Z.-Q. Luo, and A. J. Goldsmith, "Power scheduling of universal decentralized estimation in sensor networks," *IEEE Trans. on Signal Processing*, vol. 54, no. 2, pp. 413–422, Feb. 2006.
- [17] L. P. Kaelbling, M. L. Littman, and A. R. Cassandra, "Planning and acting in partially observable stochastic domains," *Artificial intelligence*, vol. 101, no. 1-2, pp. 99–134, 1998.
- [18] T. Farjam, T. Charalambous, and H. Wymeersch, "Timer-based distributed channel access in networked control systems over known and unknown Gilbert-Elliott channels," in *European Control Conference (ECC)*, June 2019, pp. 2983–2989.
- [19] Z. Zou, A. Gidmark, T. Charalambous, and M. Johansson, "Optimal radio frequency energy harvesting with limited energy arrival knowledge," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 12, pp. 3528–3539, Dec. 2016.
- [20] S. Ross, J. Pineau, B. Chaib-draa, and P. Kreitmann, "A Bayesian approach for learning and planning in partially observable Markov decision processes," *Journal of Machine Learning Research*, vol. 12, no. May, pp. 1729–1770, 2011.
- [21] T. Iwaki, Y. Wu, J. Wu, H. Sandberg, and K. H. Johansson, "Wireless sensor network scheduling for remote estimation under energy constraints," in *IEEE Conference on Decision and Control (CDC)*, Dec. 2017, pp. 3362–3367.