

# Optimal energy allocation for linear control with packet loss under energy harvesting constraints

Steffi Knorn <sup>a</sup> and Subhrakanti Dey <sup>a</sup>

<sup>a</sup>*Uppsala University, Uppsala, Sweden*

---

## Abstract

This paper studies a closed loop linear control system over a lossy communication link. A sensor computes a state estimate of the observed discrete-time linear dynamical system and sends it (in the form of packetized transmission) to the controller in the receiver block over a randomly time-varying (fading) packet dropping link. The receiver sends an ACK/NACK packet to the transmitter over an acknowledgement channel which might also be prone to packet loss. It is assumed that the energy used in packet transmission depletes a battery of limited capacity at the sensor, but is also replenished by an energy harvester which has access to a source of everlasting but random harvested energy. Under an assumption of finite-state Markov chain models of the energy harvesting and the fading channel gain processes, the objective is to design an optimal energy allocation policy at the transmitter and an optimal control policy at the receiver so that an average infinite horizon linear quadratic Gaussian (LQG) control cost is minimized. It is shown that in the case of perfect channel feedback a separation principle holds, the optimal LQG controller is linear and the optimal energy allocation policy at the transmitter can be obtained via solving the Bellman dynamic programming equation. A Q-learning algorithm is used to approximate the optimal energy allocation policy in case the system parameters (such as the transition probabilities of the underlying Markov chains) are unknown. Numerical simulation examples are used to illustrate the relative performance of the proposed algorithms and various other heuristic algorithms. It is seen that the dynamic programming based policies outperform the simple heuristic policies, especially at higher battery capacities.

*Key words:* Kalman filtering (KF), energy harvesting, imperfect acknowledgments

---

## 1 Introduction

Wireless sensors are used in many areas such as environmental data gathering Akyildiz et al. (2002), industrial process monitoring Gungor & Hancke (2009), mobile robots and autonomous vehicles Chong & Kumar (2003), and for monitoring of smart electricity grids Gungor et al. (2010). Sensors are often located in remote places and cannot be connected to reliable power sources. Thus, they are often powered by batteries and can only use a limited amount of energy for sensing, processing and communicating information. When using wireless communication and battery powered devices, the communication links are unreliable and information might be lost in a random manner. It is therefore an important task to study the effects of such unreliable communication channels on filtering and control. A line of research in this area started with Sinopoli, Schenato, Franceschetti, Poolla, Jordan & Sastry (2004), studying a Kalman filter relying on measurements, that are received from the sensor via a packet dropping channel. It is shown

that the resulting Kalman filter and its error covariance matrix are time-varying and stochastic. The mean state covariance can be guaranteed to be bounded if the probability of receiving a packet is above a certain lower bound. These results were further extended to derive conditions on the packet arrival rate to guarantee stability in Liu & Goldsmith (2004), Xu & Hespanha (2005), Schenato et al. (2007), Huang & Dey (2007), Epstein et al. (2008), Schenato (2008), Mo & Sinopoli (2008), Quevedo et al. (2012).

Other researchers studied the performance of the Kalman filter, that is minimizing the expected estimation error, instead of only ensuring its stability, see Quevedo et al. (2010), Shi et al. (2011).

The impact of packet dropping links was also studied for the closed loop control problem in a large number of works. Sinopoli, Schenato, Franceschetti, Poolla & Sastry (2004) studied a closed loop control system with a linear Gaussian quadratic optimal controller and showed that the separation principle holds in the presence of data losses between the sensor and the receiver Kalman filter when the sensor receives perfect channel feedback and there exists a critical arrival probability below which the resulting optimal controller fails to stabilize the system. Sinopoli et al. (2005c) extended these results by assuming that the control

---

\* Corresponding author: Steffi Knorn. Tel +46-18-4717389.

*Email addresses:* steffi.knorn@angstrom.uu.se (Steffi Knorn), subhra.dey@angstrom.uu.se (Subhrakanti Dey).

signal also is transmitted via an unreliable communication channel. If the arrival of the control packet is acknowledged without error at the receiving actuator, the separation principle holds and the optimal LQG control is linear. However, if no such channel feedback exists, the separation principle does not hold and the resulting optimal controller is in general nonlinear, Sinopoli et al. (2005a), apart from some special cases, Sinopoli et al. (2005b).

Energy harvesting using solar panels, wind mills or other devices might help to reduce the limitations of limited battery capacities. The harvested energy can be used for data transmission or be stored for future use. Due to the unreliable nature of most renewable energy sources, allocating the available energy in an optimal fashion is a challenging task.

Sharma et al. (2010) studied throughput optimal and mean delay optimal energy allocation policies in a single sensor node. The optimal energy allocation policies to maximize the mutual information of a wireless link were derived in Ho & Zhang (2012) under either causal or non-causal side information at the transmitters. Yang et al. (2012) investigated an optimal packet scheduling problem for a single-user energy harvesting wireless communication system and developed optimal off-line scheduling policies given non-causal information. Optimal off-line transmission policies with batteries with limited capacities are investigated in Tutuncuoglu & Yener (2012), where a finite horizon throughput maximization and the related problem of minimization of the transmission completion time are studied. These results are further generalized in Ozel et al. (2011), considering fading channels and optimal online policies. Nourian, Leong & Dey (2014) studied estimation of a dynamical system with a packet dropping link under energy harvesting constraints and derived transmission energy allocation policies, that minimize the expected error covariance in the presence of perfect or imperfect channel feedback. A smart sensor, which can decide at each time step to either send a quantized version of its local state estimate or its local innovation via a packet dropping link, was considered in Nourian, Leong, Dey & Quevedo (2014).

This paper extends Nourian, Leong & Dey (2014) to the case of a closed loop control system with a packet dropping link between the sensor and the controller at the receiver, and a packet dropping acknowledgement channel. We study the optimal energy allocation policy at the transmitter and the optimal control design at the receiver such that an average infinite-time horizon LQG control cost is minimized. The “smart” sensor estimates the state of the linear dynamical system and transmits the current estimate (as opposed to the measurements, Nourian, Leong & Dey (2014)) to the receiver unit via a packet dropping link. This transmission strategy is chosen based on Gupta et al. (2007), where it was shown that it is optimal to send estimates, in contrast to sending measurements, over packet dropping links. The receiver sends an acknowledgement to the transmitter. In contrast to Sinopoli, Schenato, Franceschetti, Poolla & Sastry (2004), Sinopoli et al. (2005c,a,b) where no restriction is placed on the sensor transmission energy, the transmitter at the sensor is equipped with a battery with finite capacity and

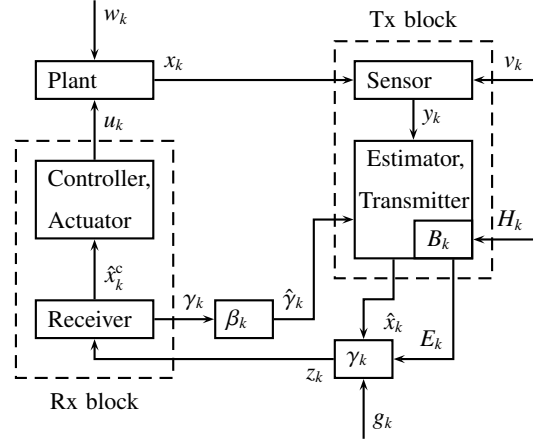


Figure 1. Scheme of system model when transmitter sends estimates

an energy harvester. Hence, the transmitter can choose how much energy should be used to transmit the current state estimate, which is limited by the available energy at the battery, which fluctuates randomly due to the stochastic nature of harvested energy. It is assumed that the time varying fading channel gain and the harvested energy amounts are described by finite-state Markov chains. Hence, the probability of dropping the packet depends on the transmission energy and the current channel gain and is, therefore, time varying (in contrast to Sinopoli, Schenato, Franceschetti, Poolla & Sastry (2004), Sinopoli et al. (2005c,a,b)). The transmitter is required to find a tradeoff amongst spending energy to transmit the current state estimate, storing energy for future transmissions, and reducing energy overflow due to a finite battery capacity. If the acknowledgement channel is perfect, the separation principle holds and the optimal controller is linear. The optimal energy allocation policy is obtained solving a average cost optimality Bellman equation. In case the acknowledgement channel is erroneous, the separation principle does not hold, leading to a coupling between the controller and energy allocation policy in general. For this case, a simple suboptimal policy with a linear controller and a suboptimal transmission energy allocation policy minimizing an average expected sum of the trace of an *estimated* receiver state estimation error covariance matrix is proposed. Both cases are also studied numerically and compared to other strategies.

Section 2 describes the system model and problem formulation. The cases of perfect or imperfect channel feedback are considered in Section 3 or Section 4, respectively. Section 5 describes the Q learning algorithm and two suboptimal heuristic energy allocation policies and all policies are compared via numerical studies in Section 6, followed by concluding remarks in Section 7.

## 2 System Model

This section describes the general structure of the system. Sections 3 and 4 describe the two different cases of perfect and imperfect channel feedback in more detail. A scheme of the system model can be found in Figure 1.

## 2.1 Plant Model and Sensor

The plant is modeled as a simple linear system with state  $x_k \in \mathbb{R}^n$  at time slot  $k$ , process noise  $w_k \in \mathbb{R}^n$ , and a control input  $u_k \in \mathbb{R}^p$ :

$$x_{k+1} = Ax_k + Bu_k + w_k. \quad (1)$$

The process noise is assumed to be i.i.d. Gaussian noise with zero mean and covariance  $M = \mathbb{E}\{w_k w_k^T\} \geq 0$ , where  $\mathbb{E}\{x\}$  denotes the expected value of  $x$ . The initial state  $x_0$  is also Gaussian with mean  $\bar{x}_0$ , and covariance  $\bar{P}_0$ . We assume that  $(A, B)$  and  $(A, M^{\frac{1}{2}})$  are controllable.

The sensor produces a noisy measurement of the state  $y_k = Cx_k + v_k$  where  $(A, C)$  is assumed to be observable,  $y_k \in \mathbb{R}^q$ , and  $v_k \in \mathbb{R}^q$  is i.i.d. Gaussian noise (independent of  $x_0$  and  $w_k$ ) with zero mean and covariance  $N = \mathbb{E}\{v_k v_k^T\} > 0$ .

## 2.2 State Estimator at the Transmitter

The sensor is assumed to be smart with computational capability, and forwards a state estimate to the remote estimator/controller. The sensor measurements are used at the transmitter to estimate the current state based on the information set  $\mathcal{I}_k := \{\hat{x}_0, y_l, \hat{y}_{l-1} : 1 \leq l \leq k\}$ , where  $\hat{y}_l$  is the channel feedback acknowledgment, which will be discussed in detail in Section 2.5. The estimate is given by

$$\begin{aligned} \hat{x}_k &:= \hat{x}_{k|k} = \mathbb{E}\{x_k | \mathcal{I}_k\} \\ &= \hat{x}_{k|k-1} + K_k(y_k - C\hat{x}_{k|k-1}) \\ \hat{x}_{k+1|k} &= \mathbb{E}\{x_{k+1} | \mathcal{I}_k\} = A\hat{x}_{k|k} + B\hat{u}_k \end{aligned} \quad (2)$$

where  $\hat{u}_k$  is the estimated control input which depends on whether the latest channel feedback signal (explained in detail below) has been received or not. The matrix  $K_k$  should be chosen such that it minimizes the error covariance matrix of the state estimation error. Details of the case of perfect and imperfect channel feedback are discussed in Sections 3 and 4.

## 2.3 Energy Harvester and Battery Dynamics

The transmitter has a rechargeable battery and an energy harvester, that can gather energy from the environment. The amount of energy available to be harvested, denoted by  $H_k$ , is unpredictable and is described as a stationary first-order homogeneous Markov chain, Ho et al. (2010). The energy used for computational purposes at the transmitter is assumed to be negligible compared to the transmission energy, such as if the receiver is a long distance away. The amount of energy in the battery  $B_k$  evolves according to

$$B_{k+1} = \min\{B_k - E_k + H_{k+1}; \bar{B}\} \quad (4)$$

with  $0 \leq B_0 \leq \bar{B}$  and where  $\bar{B}$  is the battery capacity, and  $E_k$  is the energy used for transmission during the  $k$ -th slot.

## 2.4 Forward Communication Channel

A wireless communication channel is used to transmit the estimate  $\hat{x}_k$  to the controller/actuator unit, referred to as Rx block. The channel is a packet dropping link such that the

estimate is either exactly received (for  $\gamma_k = 1$ ) or completely lost due to corrupted data or substantial delay (for  $\gamma_k = 0$ ), where  $\gamma_k$  is the Bernoulli random variable modeling the packet loss process. The received signal is  $z_k = \gamma_k \hat{x}_k$ . The probability of successfully transmitting the packet is

$$\mathbb{P}\{\gamma_k = 1 | g_k, E_k\} := h(g_k E_k) \quad (5)$$

where  $\mathbb{P}\{x|y\}$  is the probability of  $x$  conditioned on  $y$  and  $g_k$  is the time-varying wireless fading channel gain.  $h : [0, \infty] \rightarrow [0, 1]$  is monotonically increasing and continuous. We assume that the fading channel gain  $\{g_k\}$  is a first-order stationary homogeneous finite-state Markov chain where the channel gains remain constant over each fading block and vary in a Markovian fashion from block to block. Finite-state Markov models for fading channels are quite common in both wireless communications and information theoretic literature - see Sadeghi et al. (2008) for a survey in this regard. We also assume that the fading channel is independent of the energy harvesting process  $H_k$ , and known to the transmitter. The knowledge of the channel at the transmitter can be achieved by the receiver sending a pilot signal at the beginning of each slot for the transmitter to estimate the channel. The estimated channel is used as an estimate for the sensor to receiver channel also based on channel reciprocity assumption, which is valid for schemes such as time-division-duplex (TDD). Based on the current channel gain  $g_k$ , and the battery level  $B_k$ , the transmitter finds an optimal energy allocation policy  $\{E_k\}^*$  to minimize a suitable infinite horizon control cost.

## 2.5 Erroneous Channel Feedback

After receiving  $z_k$  from the transmitter, the receiver sends an acknowledgment (ACK/NACK) to the transmitter. The acknowledgement is sent via a packet dropping ACK/NACK channel such that the received acknowledgment is

$$\hat{\gamma}_k = \begin{cases} \gamma_k & \text{if } \beta_k = 1 \\ 2 & \text{otherwise} \end{cases} \quad (6)$$

where  $\beta_k$  is another Bernoulli random variable indicating if the ACK/NACK packet has been received and the probability of receiving the ACK/NACK packet is

$$\mathbb{P}\{\beta_k = 1\} = \eta \in [0, 1]. \quad (7)$$

In case no ACK/NACK is received, the transmitter receives  $\hat{\gamma}_k = 2$  indicating that the ACK/NACK packet was dropped.

## 2.6 Estimator/Controller and Actuator in the Rx block

The controller in the receiver block has access to the information set  $\mathcal{I}_k^c := \{\hat{x}_0^c, z_l, \gamma_l : 1 \leq l \leq k\}$ . As the estimates from the transmitter Kalman filter are sometimes dropped, the state estimate at the Rx block,  $\hat{x}_k^c = \mathbb{E}\{x_k | \mathcal{I}_k^c\}$ , is

$$\hat{x}_k^c := \gamma_k \hat{x}_k + (1 - \gamma_k)(A\hat{x}_{k-1}^c + Bu_{k-1}). \quad (8)$$

It can easily be shown that this state estimate at the receiver is optimal given the information set  $\mathcal{I}_k^c$ . The task of the controller is to design an optimal control sequence  $\{u_k\}$  based on  $\mathcal{I}_k^c$  to minimize the average control cost. It is assumed that the link between the Rx block and the plant is lossless.

### 2.7 Optimal Transmission Energy and Controller Design

The aim is to find the stationary optimal transmission energy allocation policy  $\{E\}^*$  and the optimal control policy  $\{u\}^*$ , that jointly minimize the average LQG control cost

$$\begin{aligned} J(\{u\}, \{E\}, \bar{x}_0, \bar{P}_0) &= \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{k=0}^{T-1} \mathbb{E} \{x_k^T W x_k + u_k^T U u_k\} \\ &= \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{k=0}^{T-1} \mathbb{E} \{ \mathbb{E} \{x_k^T W x_k + u_k^T U u_k \mid \mathcal{I}_k^c\} \} \end{aligned} \quad (9)$$

where  $(A, W^{\frac{1}{2}})$  is assumed to be observable. Section 3 shows, that if the ACK/NACK channel is perfect, the separation principle holds and the design of the optimal control input  $u_k$  and the optimal transmission energy  $E_k$  can be separated. However, in the case of imperfect channel feedback, the optimal choices of  $u_k$  and  $E_k$  depend on each other, as shown in Section 4, due to the absence of a separation principle.

**Remark 1** *The joint optimization of the transmission energy allocation and the control policy can be done at the Rx block if the battery level is known, and consequently the optimal energy allocation policy can be fed back to the transmitter. However, this is difficult as this information needs to be included in the transmitted packet, which may be lost. Therefore, the Tx block designs the optimal energy allocation policy whereas the Rx block designs the optimal control policy. It will be seen below that in the case of perfect channel feedback, the transmitter can design the optimal energy allocation policy based on the combined state information  $(g_k, B_k)$ . This is a consequence of the separation principle. In the case of imperfect channel feedback, the design of the energy allocation policy at the Tx block is suboptimal.*

## 3 Perfect Acknowledgements

In case the channel feedback link is perfect, the control of the closed loop system follows well known principles. After clarifying the dynamics of the error covariance matrices at the estimator and the controller, it will be shown that the separation principle holds. This is followed by a stability proof given a sufficiently low property of dropping packets in the forward communication channel.

### 3.1 Error Covariance Matrices

The Tx block calculates an estimate of the system state via a Kalman filter based on the information set  $\mathcal{I}_k := \{\hat{x}_0, y_l, \gamma_{l-1} : 1 \leq l \leq k\}$ . The estimate, (3), is transmitted via the packet dropping link to the receiver. Since it is assumed that the acknowledgements are received without faults, the estimator has perfect knowledge of the used state estimate at the controller, i.e.,  $\hat{x}_k^c$  since in case of  $\gamma_k = 1$ ,  $\hat{x}_k^c = \hat{x}_k$ , while in case  $\gamma_k = 0$ ,  $\hat{x}_k^c$  is known to be updated by the prediction

step  $\hat{x}_k^c = A\hat{x}_{k-1}^c + Bu_{k-1}$ . Since the control law  $u_k$  is a function of  $\hat{x}_k^c$  (see (15) below) and assumed to be known, the transmitter has perfect knowledge of the receiver estimate  $\hat{x}_k^c$ , and consequently the applied control input,  $u_k, \forall k \geq 0$ , such that  $\hat{u}_k = u_k$  by virtue of an inductive logic. The corresponding Kalman filter error covariance matrices at the transmitter are

$$P_{k|k} = \mathbb{E} \{(x_k - \hat{x}_{k|k})(x_k - \hat{x}_{k|k})^T \mid \mathcal{I}_k\}, \quad (10)$$

$$P_{k+1} := P_{k+1|k} = \mathbb{E} \{(x_{k+1} - \hat{x}_{k+1|k})(x_{k+1} - \hat{x}_{k+1|k})^T \mid \mathcal{I}_k\}. \quad (11)$$

With  $e_{k|k} = x_k - \hat{x}_{k|k}$  and  $e_{k+1|k} = x_{k+1} - \hat{x}_{k+1|k} = Ae_{k|k} + w_k$ , this yields  $P_{k+1} = \mathbb{E} \{(Ae_{k|k} + w_k)(Ae_{k|k} + w_k)^T\} = AP_{k|k}A^T + M$ .

Further, choosing  $K_k = P_{k|k-1}C^T(CP_{k|k-1}C^T + N)^{-1}$  leads to the minimal error covariance matrix after updating the estimate in the standard form  $P_{k|k} = P_{k-1} - P_{k-1}C^T(CP_{k-1}C^T + N)^{-1}CP_{k-1}$ . The initial covariance matrix is  $P_0 = \bar{P}_0$ . We assume that the Kalman filter at the transmitter has been running long enough to reach a steady state<sup>1</sup> such that the error covariance matrix at the transmitter is given by  $\lim_{k \rightarrow \infty} P_{k|k} = P_\infty$ . Since the current state estimate is dropped with a probability of  $1 - h(g_k E_k)$ , it is intermittently unavailable at the Rx block and is replaced by (8). Thus, the corresponding estimation error covariance matrix at the Rx block,  $P_k^c := \mathbb{E} \{(x_k - \hat{x}_k^c)(x_k - \hat{x}_k^c)^T \mid \mathcal{I}_k^c\}$ , is

$$P_k^c = \gamma_k P_\infty + (1 - \gamma_k)(AP_{k-1}^c A^T + M) \text{ with } P_0^c := \bar{P}_0. \quad (12)$$

### 3.2 Separation Principle and Optimal Controller

It will be shown in this section that the traditional separation principle, i.e., the separation of estimation and control design holds under the assumption of perfect ACK/NACK feedback. Note that in the context of networked control systems with packet losses on both the sensor-estimator and controller-actuator links but with transmission of measurements, it was shown that the separation principle holds when the transmitter has full knowledge of the control signals applied to the plant Schenato et al. (2007), and the optimal control law is linear. In our setting, transmission of state estimates is considered and packet loss is only present in the forward link. However, as the estimator needs to know the control applied to the plant, the assumption of perfect ACK/NACK feedback is necessary and it is shown below that under this assumption, the same separation principle holds. What follows as a consequence is also the separation of the design of the stationary optimal transmission energy allocation policy at the transmitter and the optimal control policy at the receiver, that jointly minimize the control cost

<sup>1</sup> Note that in this paper we investigate the optimization of the infinite-time horizon case of a long term average of an LQG control cost. Even if the system is set up when the Kalman filter is not in the steady state, transients towards the equilibrium state do not contribute to the average over the infinite time horizon, see Bertsekas (1995).

(9). Essentially, the LQG control cost can be minimized by first choosing  $u_k^*$ , the optimal controller, which is a linear function of  $\hat{x}_k^c$ . It can be shown that minimizing the control cost is then equivalent to minimizing the long term average of the estimation error covariance matrix with respect to the energy allocation policy at the transmitter. The analysis largely follows a similar result in Schenato et al. (2007), and we only present a sketch of the most important details for the sake of completeness.

First, define a finite horizon version of the control cost (9) as follows:  $J_T = \mathbb{E} \left\{ x_T^T W x_T + \sum_{k=0}^{T-1} (x_k^T W x_k + u_k^T U u_k) \right\}$ . Then a fixed sequence of energy allocation  $\{E\}_0^{T-1} = \{E_0, E_1, \dots, E_{T-1}\}$  and consider minimization of the above cost only with respect to the control sequence  $\{u\}_0^{T-1} = \{u_0, u_1, \dots, u_{T-1}\}$ . One can then present the following lemma:

**Lemma 2** *The value function  $V_T(\mathcal{I}_T^c) := \mathbb{E} \{ x_T^T W x_T | \mathcal{I}_T^c \}$ ,  $V_k(\mathcal{I}_k^c) := \min_{u_k} \mathbb{E} \{ x_k^T W x_k + u_k^T U u_k + V_{k+1}(\mathcal{I}_{k+1}^c) | \mathcal{I}_k^c \}$  for the system (1) with the information set  $\mathcal{I}_k^c$ ,  $k = T-1, \dots, 1$ , can be written as  $V_k(\mathcal{I}_k^c) = \mathbb{E} \{ x_k^T S_k x_k | \mathcal{I}_k^c \} + c_k$  for  $k = T, \dots, 0$  where  $S_k$  and  $c_k$  can be calculated recursively according to*

$$S_k = A^T S_{k+1} A + W - A^T S_{k+1} B (B^T S_{k+1} B + U)^{-1} B^T S_{k+1} A \quad (13)$$

$$c_k = \text{trace} \left( (A^T S_{k+1} A + W - S_k) P_k^c \right) + \text{trace} (S_{k+1} M) + \mathbb{E} \{ c_{k+1} | \mathcal{I}_k^c \} \quad (14)$$

with  $S_T = W$  and  $c_T = 0$ . Further,

$$u_k = - (B^T S_{k+1} B + U)^{-1} B^T S_{k+1} A \hat{x}_k^c = L_k \hat{x}_k^c. \quad (15)$$

**PROOF.** The proof follows along similar lines as in (Schenato et al. 2007, Lem 5.1): The claim is true for  $k = N$ . When evaluating  $V_k(\mathcal{I}_k^c)$  for  $k \leq N-1$  and using results in Schenato et al. (2007) it can be shown that  $V_k(\mathcal{I}_k^c)$  does not depend on  $E_k$ . Hence, it suffices to minimize with respect to  $u_k$  while keeping  $E_k$  fixed. Solving  $\frac{\partial V_k(\mathcal{I}_k^c)}{\partial u_k} = 0$  yields the optimal control input (15), which together with (Schenato et al. 2007, Lem 4.1(b)), (13) and (14) proves the results.  $\square$

The above result can be applied repeatedly to obtain from dynamic programming theory that the optimal cost  $J_T^*$  is given by  $\min_{\{E\}_0^{T-1}} J_T = V_0(\mathcal{I}_0^c)$  which is (similar to Schenato et al. (2007))

$$J_T^* = \min_{\{E\}_0^{T-1}} \left[ \bar{x}_0^T \bar{P}_0 \bar{x}_0 + \text{tr}(S_0 \bar{P}_0) + \sum_{k=0}^{T-1} \text{tr}(S_{k+1} M) + \sum_{k=0}^{T-1} \text{tr} \left( (A^T S_{k+1} A + W - A_k) \mathbb{E}(P_k^c) \right) \right] \quad (16)$$

Hence, calculating the optimal control input at the Rx block, and the optimal energy allocation at the Tx block can be done separately. The infinite horizon cost (9) can now be considered by taking the limit of the above equations as  $T \rightarrow \infty$ , provided the cost is bounded. A sufficient condition for boundedness is provided in Section 3.3 below. Since  $(A, B)$  is controllable and  $(A, W^{\frac{1}{2}})$  is observable, the optimal controller for the infinite horizon case has gain  $L = L_\infty = - (B^T S_\infty B + U)^{-1} B^T S_\infty A$  where  $S_\infty$  is the solution of the standard ARE  $S_\infty = A^T S_\infty A + W - A^T S_\infty B (B^T S_\infty B + U)^{-1} B^T S_\infty A$ , see Sinopoli, Schenato, Franceschetti, Poolla & Sastry (2004) for further details.

### 3.3 Stability

Recall that we aim to find the stationary optimal transmission energy allocation policy  $\{E\}^*$  (if it exists) and the optimal control policy  $\{u\}^*$ , that jointly minimise the infinite-horizon average LQG control cost (9). Since the separation principle holds as shown above, it follows from (16) that the stochastic control problem to determine  $\{E_k\}^*$  is given by

$$\min_{E_k: k \geq 1} \limsup_{T \rightarrow \infty} \frac{1}{T} \sum_{k=0}^{T-1} \mathbb{E} \{ \text{tr}(P_k^c) \}. \quad (17)$$

This is a stochastic control problem based on a Markov Decision Process (MDP) Bertsekas (1995), with state space  $S = \{P_k^c, g_k, H_k, B_k\}$  and action space  $\mathcal{A} = \{E_k\}$ . Stability of the process with an unstable open loop can be guaranteed if the packet dropping probability is sufficiently low:

**Theorem 3** *Assume the error covariance matrix at the controller  $P_k^c$  in (12). If there exists a  $\xi \in [0, 1)$  such that*

$$\sup_{g, H} \int_{g_k} \int_{H_k} (1 - h(g_k \min \{H_k, \bar{B}\})) \times \mathbb{P}(g_k | g_{k-1} = g) \times \mathbb{P}(H_k | H_{k-1} = H) dg_k dH_k \leq \xi / \|A\|^2 \quad (18)$$

for all  $k \geq 0$ , then there exists a policy  $\{E_k\}$  such that for some scalars  $\alpha, \beta > 0$  the norm of  $P_k^c$  is satisfies for all  $k \geq 0$

$$\mathbb{E} \{ \|P_k^c\| \} \leq \alpha \xi^k + \beta. \quad (19)$$

**PROOF.** The proof is based on (Quevedo et al. 2013, Thm 1) showing that a sufficient condition for exponential stability in the sense of (19) is  $\sup_{g, H} \mathbb{P}(\gamma_k = 0 | g_{k-1} = g, H_{k-1} = H) \leq \frac{\xi}{\|A\|^2}$  for some  $\xi \in [0, 1)$ . Since it is assumed that both the channel gains  $\{g_k\}$  and the harvested energies  $\{H_k\}$  are described by mutually independent stationary first-order homogeneous Markov chains, the exponential stability condition yields (18) with  $\min \{H_k, \bar{B}\} = E_k$  for some  $\xi \in [0, 1)$ . Assume all the harvested energy at each time step is used. Hence,  $E_0 = B_0$  and  $E_k = \min \{H_k, \bar{B}\}$  for  $k \geq 1$ . Then (18) is a sufficient condition to guarantee (19).  $\square$

**Remark 4** *Condition (18) describes, that if one uses all*

the energy in the battery for data packet transmission at each step, i.e.,  $E_k = \min\{H_k, \bar{B}\}$ , a sufficiently low value of the resulting average probability of the packet being lost, i.e.,  $1 - h(g_k E_k)$  averaged over the conditional probability distributions of the Markov channel gain  $g_k$  and the Markov harvested energy process  $H_k$ , can guarantee stabilizability of the possibly unstable system (1). If this condition is satisfied for the suboptimal energy allocation policy  $E_k = \min\{H_k, \bar{B}\}$ , then Theorem 3 shows that an optimal energy allocation exists, such that the norm of the expected error covariance matrix is bounded.

### 3.4 Optimal Energy Allocation Policy

Note that even though the channel and harvested energy state spaces are finite-discrete by assumption, the error covariance matrix  $P_k^c$  belongs to a general Borel space, and so does the energy allocation  $E_k$ . Existence of solutions to average cost MDP problems with general (Borel) state and action spaces is not straightforward and requires various assumptions on the cost function, state and action spaces, and the transition probabilities of the MDP to be satisfied. Below, we first state the main result in Theorem 5, followed by the assumptions required for this result to hold and how they are satisfied in the short proof sketch. Assume the channel gain, harvested energy, battery level and energy consumption at time  $k$  are denoted  $g = g_k$ ,  $H = H_k$ ,  $B = B_k$ , and  $E = E_k$ , respectively, and the corresponding channel gain, harvested energy and battery level at time  $k + 1$  are denoted  $\tilde{g} = g_{k+1}$ ,  $\tilde{H} = H_{k+1}$  and  $\tilde{B} = B_{k+1}$ , respectively. The following theorem states that the stationary optimal energy allocation policy is given by:

**Theorem 5** *The optimal cost of the infinite-time horizon stochastic control problem (17) is given by  $\rho$ , which is the unique solution of the average-cost Bellman optimality equation*

$$\rho + V(P^c, g, H, B) = \min_{E \in [0, \bar{B}]} \mathbb{E} \left\{ \text{tr}(P^c) + V(\tilde{P}^c, \tilde{g}, \tilde{H}, \tilde{B} | P^c, g, H, B, E) \right\} \quad (20)$$

where  $V$  is the relative value function, and  $\rho$  is independent of the initial conditions  $P_0$ ,  $g_0$ ,  $H_0$ ,  $B_0$ .

**PROOF.** The proof follows similar steps as the proof of (Nourian, Leong & Dey 2014, Thm 4.2). Here we only provide a brief sketch due to space constraints. Essentially, one needs to first show the existence to the solution of average cost optimality *inequality* (see Equation (A.8) in Nourian, Leong & Dey (2014)), by verifying two conditions [W] and [B] from Schäl (1993). Condition [W] essentially requires that the state space is *locally compact*, the action space is compact and the state to action function is upper semi-continuous, and transition probabilities of the MDP are weakly continuous, and finally, the cost function  $\text{tr}(P^c)$  is lower semi-continuous. Due to the finite state Markov chain modelling of the harvested energy and the fading channel processes, and the energy allocation being limited by the maximum battery capacity, all of these conditions can be

verified for our problem. Condition [B] of Schäl (1993) is trickier and essentially it requires the finiteness of a relative value function of a corresponding *discounted* cost MDP problem. It can be also shown that this condition holds in our case, similar to the arguments used in Nourian, Leong & Dey (2014). This guarantees the existence of a solution to the average cost optimality inequality. To show equality, as demanded by (20), one requires a further equicontinuity property of the optimal cost for the related discounted cost MDP with respect to the state, as shown in the proof of Proposition 3.2 in Huang & Dey (2006). Finally, the assumptions on Sections 5.4 and 5.5 of Hernández-Lerma & Lasserre (1996) can be then verified to conclude the existence of a solution to the average cost optimality equation (20).  $\square$

The stationary optimal solution to the stochastic control problem (17) is hence

$$E^*(P^c, g, H, B) = \underset{E \in [0, \bar{B}]}{\text{argmin}} \mathbb{E} \left\{ \text{tr}(P^c) + V(\tilde{P}^c, \tilde{g}, \tilde{H}, \tilde{B} | P^c, g, H, B, E) \right\} \quad (21)$$

with (4) and  $V$  is the solution to the Bellman equation (20).

**Remark 6** *In the setup described in the current paper, the packet loss process  $\gamma_k$  is described by a Bernoulli process, independent of any other random processes such as the process and measurement noise and the initial state, fading channel or harvested energy processes. The choice of the optimal energy allocation  $E_k^*$  does not also depend on the measurements  $\{y_k\}$ , or the receiver/transmitter estimates. In particular, optimal values of  $E_k$  are found by minimizing the long-term average of the trace of the estimation error covariance at the receiver. Thus, the optimal energy allocation  $E_k^*$  is adapted to the current receiver error covariance  $P_k^c$  and the current fading channel gain and battery state at the sensor, which are known at the transmitter. Due to the perfect ACK/NACK feedback scenario, the receiver error covariance is also known at the transmitter. Thus separation principle holds. Note however that this is in contrast with many event-triggering schemes Astrom & Bernhardsson (2002), Sijs & Lazar (2012), Molin & Hirche (2014), where the decision to transmit data may depend on the state, or the estimate (or estimation error) or the measurement at the transmitter, and this makes the received signal a nonlinear function of the state/measurement, leading to non-optimality of the Kalman filter and non-existence of the separation principle.*

## 4 Imperfect Acknowledgements

### 4.1 Assumed State Estimate

Similar to the case of perfect channel feedback in Section 3, the state estimate at the Rx block is given by (8). The estimate at the Tx block depends on the estimated input signal  $\hat{u}_k$ , which substitutes the unknown input signal  $u_k = L\hat{x}_k^c$ . The control input is estimated by the transmitter using the estimate of the state estimate available at the con-

troller  $\hat{u}_k = L\hat{x}_k^{\text{ce}}$  with

$$\begin{aligned}\hat{x}_k^{\text{ce}} = & (1 - \beta_k) \left( h(g_k E_k) \hat{x}_k + (1 - h(g_k E_k)) (A\hat{x}_{k-1}^{\text{ce}} + B\hat{u}_{k-1}) \right) \\ & + \beta_k \left( \gamma_k \hat{x}_k + (1 - \gamma_k) (A\hat{x}_{k-1}^{\text{ce}} + B\hat{u}_{k-1}) \right).\end{aligned}\quad (22)$$

In case an acknowledgment was received, the information of the acknowledgment is used. Otherwise the estimate used at the controller is estimated using the packet dropping probability of the forward channel. Hence, the estimation error  $e_{k+1|k} = Ae_{k|k} + BL e_k^{\text{e}} + w_k$  now depends on the error  $e_k^{\text{e}} := \hat{x}_k^{\text{c}} - \hat{x}_k^{\text{ce}}$ . Due to this,  $e_{k+1|k}$  and the remaining error signals  $e_k^{\text{c}} := x_k - \hat{x}_k^{\text{c}}$  and  $\hat{e}_k^{\text{ce}} := x_k - \hat{x}_k^{\text{ce}}$  are coupled.

**Remark 7** In order to calculate the assumed estimate  $\hat{x}_k^{\text{ce}}$ , other suboptimal policies based on additional information such as the current measurement can be used. The approach here (based solely on the drop-out probability) is simple to implement but suboptimal. Alternative suboptimal schemes in a slightly different context can be found in Dey et al. (2015). Further, the choice of  $u_k = L\hat{x}_k^{\text{c}}$  is also suboptimal (see below), but allows a simple and straight forward implementation.

#### 4.2 Error Covariance Matrices and Kalman Filter

Defining  $P_k^{\text{e}} := \mathbb{E} \left\{ e_k^{\text{e}} (e_k^{\text{e}})^{\text{T}} \right\}$  it follows

$$\begin{aligned}P_{k+1} = & AP_{k|k}A^{\text{T}} + M + BL P_k^{\text{e}} L^{\text{T}} B^{\text{T}} \\ & + BL \mathbb{E} \left\{ e_k^{\text{e}} (e_{k|k})^{\text{T}} \right\} A^{\text{T}} + A \mathbb{E} \left\{ e_{k|k} (e_k^{\text{e}})^{\text{T}} \right\} L^{\text{T}} B^{\text{T}}.\end{aligned}\quad (23)$$

Note that  $P_{k+1}$  depends on the controller matrix  $L$  and  $P_k^{\text{e}}$  and the cross-correlation matrix  $\mathbb{E} \left\{ e_k^{\text{e}} (e_{k|k})^{\text{T}} \right\}$ . Below we derive some approximate expressions for  $P_k^{\text{e}}$  and the cross-correlation matrix  $\tilde{P}_k^{\text{corr}} = \mathbb{E} \left\{ e_k^{\text{e}} (e_{k|k})^{\text{T}} \right\}$ , and show that they can be asymptotically neglected if the packet loss probability for the ACK/NACK channel,  $\mathbb{P}\{\beta_k = 0\} = (1 - \eta) \approx 0$ .

To this end, we use (8) and (22), and  $\hat{u}_{k-1} = L\hat{x}_{k-1}^{\text{ce}}$  to first derive that when  $\beta_k = 1$ ,  $e_k^{\text{e}} = (1 - \gamma_k)(A + BL)e_{k-1}^{\text{e}}$ . Similarly, when  $\beta_k = 0$ , we have

$$e_k^{\text{e}} = (\gamma_k - \lambda_k) \hat{x}_k + (A + BL) \left[ (1 - \gamma_k) \hat{x}_{k-1}^{\text{c}} - (1 - \lambda_k) \hat{x}_{k-1}^{\text{ce}} \right]$$

where we have used for simplicity,  $\lambda_k = \mathbb{P}\{\gamma_k = 1 | g_k, E_k\} := h(g_k E_k)$ . This allows us to write (after some algebra)

$$\begin{aligned}P_k^{\text{e}} = & \eta(1 - \lambda_k)(A + BL)P_{k-1}^{\text{e}}(A + BL)^{\text{T}} + (1 - \eta) \\ & \times \left[ \lambda_k(1 - \lambda_k)^2 \mathbb{E} \left( \hat{x}_k + (A + BL)\hat{x}_{k-1}^{\text{ce}} \right) \left( \hat{x}_k + (A + BL)\hat{x}_{k-1}^{\text{ce}} \right)^{\text{T}} \right. \\ & + (1 - \lambda_k) \mathbb{E} \left( -\lambda_k \hat{x}_k + (A + BL)(\hat{x}_k^{\text{c}} - (1 - \lambda_k)\hat{x}_{k-1}^{\text{ce}}) \right. \\ & \times \left. \left. \left( -\lambda_k \hat{x}_k + (A + BL)(\hat{x}_k^{\text{c}} - (1 - \lambda_k)\hat{x}_{k-1}^{\text{ce}}) \right)^{\text{T}} \right) \right]\end{aligned}\quad (24)$$

It is obvious that the second term on the right hand side following the multiplication factor  $(1 - \eta)$  is difficult to compute explicitly. Under the assumption that  $(1 - \eta) \approx 0$ , we can approximate the above equation as  $P_k^{\text{e}} = \eta(1 -$

$\lambda_k)(A + BL)P_{k-1}^{\text{e}}(A + BL)^{\text{T}}$ , and since  $(A, B)$  is controllable with  $(A + BL)$  stable, it automatically follows that  $P_k^{\text{e}} \rightarrow 0$  as  $k \rightarrow \infty$ .

Now, using the state-space model (1), (2) and (3), we can also write

$$e_{k|k} = (I - K_{\infty}C)(Ae_{k-1|k-1} + BL e_{k-1}^{\text{e}}) - K_{\infty}v_k + w_k$$

where  $K_{\infty}$  is the steady state Kalman prediction gain at the transmitter. Again, using the approximation  $e_k^{\text{e}} \approx (1 - \gamma_k)(A + BL)e_{k-1}^{\text{e}}$ , we can obtain the following approximation for the cross-correlation matrix  $\tilde{P}_k^{\text{corr}} = \mathbb{E} \left\{ e_k^{\text{e}} (e_{k|k})^{\text{T}} \right\}$ :

$$\begin{aligned}\tilde{P}_k^{\text{corr}} = & (1 - \lambda_k)(A + BL) \left[ P_{k-1}^{\text{e}} L^{\text{T}} B^{\text{T}} (I - K_{\infty}C)^{\text{T}} \right. \\ & \left. + \tilde{P}_{k-1}^{\text{corr}} A^{\text{T}} (I - K_{\infty}C)^{\text{T}} \right] \\ \approx & (1 - \lambda_k)(A + BL) \tilde{P}_{k-1}^{\text{corr}} [(I - K_{\infty}C)A]^{\text{T}}, k \rightarrow \infty\end{aligned}$$

where the last line follows since  $P_k^{\text{e}} \rightarrow 0$  as  $k \rightarrow \infty$ . Since  $(A, C)$  is also observable, we know from Kalman filtering theory that the standard Kalman prediction error and hence the standard Kalman filtering error for the state space system observed at the transmitter becomes zero-mean asymptotically. This implies that  $(I - K_{\infty}C)A$  must also be stable. Therefore, it follows directly from the last line of the above equation that  $\tilde{P}_k^{\text{corr}} \rightarrow 0$  as  $k \rightarrow \infty$ , since both  $(A + BL)$  and  $(I - K_{\infty}C)A$  are stable matrices. Thus, asymptotically, from (23),  $P_{k+1}$  is approximated by  $\tilde{P}_{k+1} = A\tilde{P}_{k|k}A^{\text{T}} + M$ .

Following similar steps as in the perfect channel feedback case leads to  $\tilde{P}_{k|k} = \tilde{P}_{k-1} - \tilde{P}_{k-1}C^{\text{T}}(C\tilde{P}_{k-1}C^{\text{T}} + N)^{-1}C\tilde{P}_{k-1}$ . Hence, due to these simplifications, it follows that  $\tilde{P}_{k|k}$  also approaches  $P_{\infty}$  as in the perfect channel feedback case. However, whereas  $P_{k|k}$  describes the true error covariance matrix at the transmitter in the perfect channel feedback case,  $\tilde{P}_{k|k}$  in the imperfect channel feedback case is merely an approximation of the true, unknown  $P_{k|k}$ . It can further be shown that the error covariance matrix  $P_k^{\text{e}}$  in the case of imperfect channel feedback can be approximated by  $\tilde{P}_k^{\text{e}} = \gamma_k P_{\infty} + (1 - \gamma_k)(AP_{k-1}^{\text{c}}A^{\text{T}} + M)$  and  $P_k^{\text{ce}}$  is approximated by

$$\begin{aligned}\tilde{P}_k^{\text{ce}} = & (\beta_k \gamma_k + (1 - \beta_k)h(g_k E_k)) P_{\infty} \\ & + (\beta_k(1 - \gamma_k) + (1 - \beta_k)(1 - h(g_k E_k))) (A\tilde{P}_{k-1}^{\text{ce}}A^{\text{T}} + M)\end{aligned}\quad (25)$$

Note that this simplification is achieved by ignoring the term  $BLP_{k-1}^{\text{e}}L^{\text{T}}B^{\text{T}}$  and a variety of mixed terms containing  $e_{k-1}^{\text{e}}$ ,  $e_{k-1}^{\text{ce}}$  and  $e_{k|k}$ . Hence, the approximation might differ significantly from the true  $P_k^{\text{ce}}$ , but allows simple calculations.

#### 4.3 Suboptimal Controller Design and Energy Allocation

Note that in case of imperfect acknowledgements the separation principle does not hold since the estimate and the estimation error covariance matrix in (23) and the error covariance matrix of the assumed controller error  $P_k^{\text{ce}}$  depend on the controller matrix  $L$ . Hence, it is not optimal to design the estimator, the LQR controller and the energy allo-

cation policy separately as done in Section 3, and the optimal control is in general nonlinear. However, assuming that the probability of dropping the acknowledgement is small, a suboptimal linear controller  $u_k = L\hat{x}_k^c$  (using  $L = L_\infty$  for the perfect channel feedback case) can still be used. A simple, suboptimal energy allocation policy at the transmitter can be realized using the look-up table created for the perfect channel feedback case. Since the error covariance matrix at the receiver is not perfectly known at the transmitter, the optimal energy allocation from the look-up table for setting  $P_k^c = P_k^{ce}$  is used. Numerical results in Section 6 show that the suboptimal policy performs well compared to the perfect channel feedback case.

## 5 Suboptimal Energy Allocation Policies

### 5.1 Q-Learning

The Bellman equation (20) cannot be used if some of the system parameters are not known, such as, for instance, the transition probabilities of the Markov chains generating the channel gains or the energy harvesting process. Hence, finding algorithms, which do not rely on the complete knowledge of the system, is an important task. Assume that the state space is discrete or can be discretized. Since the fading channel gains and harvested energy are independent Markov chains, the Q-Bellman equation yields

$$Q^*(P^c, g, H, B, E) = \mathbb{E}\{tr(P^c)\} + \sum_{\tilde{g}, \tilde{H}, \tilde{B}} \mathbb{P}(\tilde{g}|g)\mathbb{P}(\tilde{H}|H)\mathbb{P}(\tilde{B}|B, H, E) \min_{\tilde{E} \in \mathcal{A}(\tilde{B})} Q^*(\tilde{P}^c, \tilde{g}, \tilde{H}, \tilde{B}, \tilde{E}) \quad (26)$$

where  $\mathcal{A}(\tilde{B})$  is the set of all feasible choices of  $\tilde{E}$  given  $\tilde{B}$ . (26) can be solved using the stochastic approximation based Q-learning algorithm, Sutton & Barto (1998), Prabuchandran et al. (2013). Assuming that the probabilities  $\mathbb{P}(\tilde{g}|g)$ ,  $\mathbb{P}(\tilde{H}|H)$  and  $\mathbb{P}(\tilde{B}|B, H, E)$  are unknown, it is

$$Q_0(P^c, g, H, B, E) = 0 \quad \forall P^c, g, H, B \text{ and } E \in \mathcal{A}(B), \quad (27)$$

$$Q_{k+1}(P^c, g, H, B, E) = Q_k(P^c, g, H, B, E) + \gamma(k) \cdot \left( \mathbb{E}\{tr(P^c)\} + \min_{\tilde{E} \in \mathcal{A}(\tilde{B})} Q_k(\tilde{P}^c, \tilde{g}, \tilde{H}, \tilde{B}, \tilde{E}) - Q_k(P^c, g, H, B, E) \right) \quad (28)$$

for all  $k \geq 0$ , where  $\{\tilde{P}^c, \tilde{g}, \tilde{H}, \tilde{B}, \tilde{E}\}$  is the next state, and  $\{P^c, g, H, B, E\}$  is the previous state at which  $E \in \mathcal{A}(B)$  is

$$E = \begin{cases} \text{argmin}_{E \in \mathcal{A}(B)} Q_k(P^c, g, H, B, E) & \text{w/ prob. } 1 - \epsilon \\ \text{chosen randomly } \in \mathcal{A}(B) & \text{w/ prob. } \epsilon \end{cases} \quad (29)$$

The algorithm in (28) converges to the optimal Q values if the step sizes  $\gamma(k)$  satisfies  $\sum_k \gamma(k) = \infty$  and  $\sum_k \gamma^2(k) < \infty$ , and the convergence is guaranteed for all  $\epsilon > 0$ , Sutton & Barto (1998), Prabuchandran et al. (2013).

### 5.2 Heuristic policies

It is desirable to find suboptimal policies, that require less computational effort than solving (20). One very simple sub-

optimal policy is a “greedy policy” which sets  $E_k = B_k, \forall k$ . A second simple heuristic policy is the “inverted channel policy”. Assume the required transmission energy such that the expected drop-out probability of the communication channel with channel gain  $g_k$  is equal to a desired probability  $\bar{\gamma}$ , is denoted by  $E_{\bar{\gamma}}(\bar{\gamma}, g_k)$ . Then, the inverted channel energy allocation policy follows the simple rule  $E_k = \min\{B_k, E_{\bar{\gamma}}(\bar{\gamma}, g_k)\}$ .

## 6 Numerical Examples

**Example 1:** A scalar system with parameters  $A = 1.1$ ,  $B = 1$ ,  $C = 1$ ,  $M = 1$ ,  $N = 1$  and  $P_{x_0} = 1$  is considered. It is assumed that the sensor uses a binary phase shift keying (BPSK) transmission scheme, Proakis (2001), with  $b = 4$  bits per packet. Hence,  $\mathbb{P}\{\gamma_k = 1|g_k, E_k\} = h(g_k E_k) = \left( \int_{-\infty}^{\sqrt{g_k E_k}} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt \right)^b$ . The battery capacity is varied between 1mW and 5mW. The fading channel gain and harvested energy are given by independent 3-level discrete Markov chains with values  $\{0, 0.5, 1\}$  and  $\{0, 1, 2\}$ , respectively, and the transition probability matrix for both processes is taken to be the same  $\mathbb{T} = [0.2, 0.3, 0.5; 0.3, 0.4, 0.3; 0.1, 0.2, 0.7]$ .

Eight scenarios have been simulated. In the first three scenarios, the optimal solution is obtained using dynamic programming to solve the Bellman optimality equation. Sending measurements vs. sending estimates and perfect vs. imperfect channel feedback with  $\eta = 0.8$ , is considered. In the scenarios 4,5 and 6 the Q-learning algorithm is used and the current state estimate is sent via the communication channel. The learning horizon is increased from  $10^4$  (QL1) to  $10^6$  (QL2) and to  $10^8$  (QL3). Scenarios 7 and 8 consider the two heuristic policies described in Section 5.2. In scenario 7, the greedy policy is used whereas in scenario 8 the inverted channel policy with  $\bar{\gamma} = 0.7$  is used. The average control cost  $J$  for all scenarios (for  $10^6$  time steps) is shown in Fig. 2. It can be observed that for dynamic programming (DP) and Q-learning  $J$  decreases from 0.7 – 0.77 to 0.4 – 0.42 for DP and from 1.08 to 0.49 for QL3 if the battery level increases from 1mWh to 5mWh. As expected, sending measurements leads to a slightly worse performance (that is,  $J \approx 0.42 - 0.77$  for different battery limits) compared to sending the state estimate ( $J \approx 0.4 - 0.7$ ). When sending state estimates,  $J$  for the scenario with  $\eta = 0.8$  differs at most 10% from the scenario with perfect channel feedback. When the learning horizon of the Q-learning algorithm increases,  $J$  also decreases by up to 30% comparing QL1 and QL3 but lies at best still 15% above  $J$  obtained by using dynamic programming. The two heuristic policies perform much better than the Q-learning algorithm for  $\bar{B} = 1mWh$ , that is  $J = 0.71$  and  $J = 0.77$  vs.  $J = 1.08$  but for the heuristics  $J$  stays constant at 0.5 when  $\bar{B}$  lies between 2 and 5mWh. A small increase in the average control cost with increasing  $\bar{B}$  for the case of QL1 can be attributed to numerical inaccuracies due to insufficient number of discretization levels for the receiver error covariance values. This can be rectified by using larger number of discretization levels and averaging over increased number of independent simulation runs, but at the cost of increased computational time.

**Example 2:** The system dynamics are chosen as in Ex-



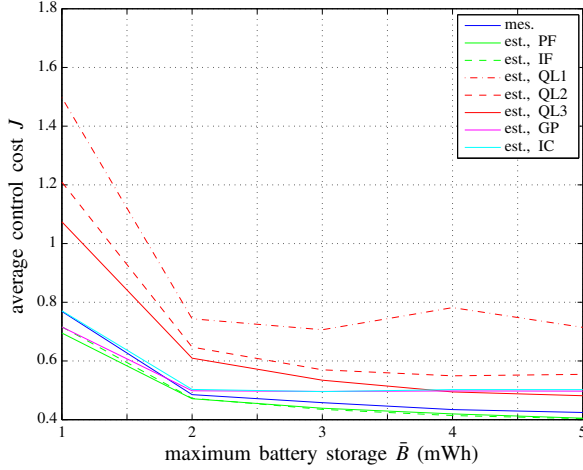


Figure 2. Example 1:  $J$  vs.  $\bar{B}$ ; ‘mes’ = send measurements, ‘est’ = send estimates, ‘PF’ = perfect channel feedback, ‘IF’ = imperfect channel feedback, ‘QL’ = Q-learning horizon, ‘GP’ = greedy policy, ‘IC’ = inverted channel policy

ample 1 but setting  $A = 1.25$  such that the critical packet arrival rate to guarantee stability for perfect channel feedback is 0.36. The values of the channel gains are chosen as multiples of  $\{0, 1, 2\}$  to vary the average forward packet arrival rate. Fig. 3 shows the simulation results for using the lookup table obtained from dynamic programming (‘DP’) and for the constant energy allocation policy (‘CE’)  $E_k = \min\{B_k, \bar{H}\}$ , where  $\bar{H}$  is the average harvested energy. In both cases, it can be observed that when the average channel gain is decreased such that the average forward packet arrival rate decreases below the critical rate  $\bar{h} = 0.36$ , the average control cost increases to values of  $J$  above 4 and grows without bound as  $\bar{h}$  approaches 0. Further, the average control cost slightly increases when lowering  $\eta$  from 0.9 to 0.5 for both energy allocation policies. Using the energy allocation lookup table calculated for the perfect channel feedback case in case of imperfect channel feedback case, yields acceptable simulation results. Despite the solution being suboptimal, it outperforms simple heuristic policies.

## 7 Conclusions

We studied a linear control system with a packet dropping link between the smart sensor and the controller. Since the link is a time-varying fading channel, the probability of receiving the current state estimate is time-varying. The receiver at the controller sends an ACK/NACK packet to the transmitter to acknowledge the arrival of the packet. The acknowledgement channel is assumed to be prone to losses. The transmitter at the sensor is equipped with a finite battery and an energy harvester. The objective is to design a jointly optimal sensor transmission energy allocation and optimal control policy to minimize an infinite-horizon LQG control cost.

When assuming perfect channel feedback, the separation principle holds. Hence, the Kalman filters at the sen-

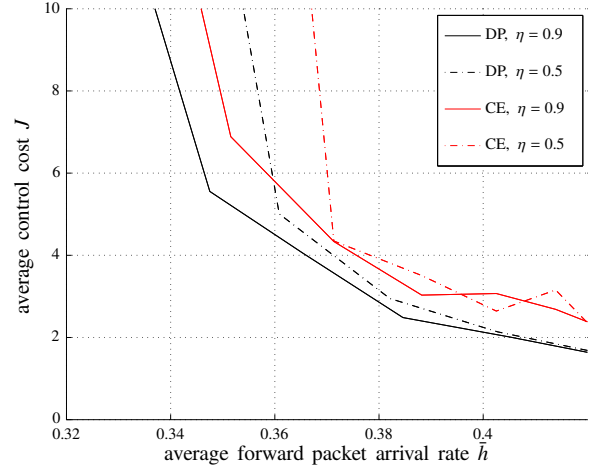


Figure 3. Example 2:  $J$  vs.  $\bar{h}$ ; ‘DP’ = energy allocation according to solution of dynamic programming, ‘CE’ = constant energy allocation policy

sor and a linear controller are optimal, and the transmission energy allocation policy can be obtained by standard dynamic programming techniques for solving the average-cost optimality equation. In case the acknowledgement channel is imperfect, the separation principle no longer holds. This implies that the optimal estimator and controller might no longer be linear. Yet, numerical studies revealed that when adjusting the estimator, controller and energy allocation policy obtained for the case of perfect channel feedback, the average control cost only increase slightly in case of imperfect channel feedback.

In case certain underlying system parameters such as the transition probabilities of the associated Markov chains are unknown, we employ Q-learning based suboptimal energy allocation algorithms. We also proposed two simple heuristic energy allocation algorithms. All algorithms are compared by numerical studies illustrating that the optimal energy allocation policy obtained by dynamic programming outperforms several suboptimal policies.

## References

- Akyildiz, I., Su, W., Sankarasubramaniam, Y. & Cayirci, E. (2002), ‘A survey on sensor networks’, *IEEE Commun. Mag.* **40**(8), 102–114.
- Astrom, K. & Bernhardsson, B. (2002), Comparison of riemann and lebesgue sampling for first order stochastic systems, in ‘Proc. 41st IEEE Conf. on Decision and Control’, pp. 2011–2016.
- Bertsekas, D. (1995), *Dynamic Programming and Optimal Control*, Vol. 1, Athena Scientific.
- Chong, C.-Y. & Kumar, S. (2003), ‘Sensor networks: Evolution, opportunities and challenges’, *Proc. IEEE* **91**(8), 1247–1256.
- Dey, S., Chiuso, A. & Schenato, L. (2015), Linear encoder-decoder-controller design over channels with packet loss and quantization noise, in ‘ECC’.
- Epstein, M., Shi, L., Tiwari, A. & Murray, R. (2008), ‘Prob-

- abilistic performance of state estimation across a lossy network', *Automatica* **44**(12), 3046–3053.
- Gungor, V. & Hancke, G. (2009), 'Industrial wireless sensor networks: Challenges in design, principles and technical approaches', *IEEE Trans. Ind. Electron.* **56**(10), 4258–4265.
- Gungor, V., Lu, B. & Hancke, G. (2010), 'Opportunities and challenges of wireless sensor networks in smart grid', *IEEE Trans. Ind. Electron.* **57**(10), 3557–3564.
- Gupta, V., Hassibi, B. & Murray, R. (2007), 'Optimal LQG control across packet-dropping links', *System and Control Letters* **56**(6), 439–446.
- Hernández-Lerma, O. & Lasserre, J. B. (1996), *Discrete-Time Markov Control Processes: Basic Optimality Criteria*, Springer-Verlag, New York.
- Ho, C., Khoa, P. & Ming, P. (2010), Markovian models for harvested energy in wireless communications, in 'IEEE Intern. Conf. Comm. Systems', pp. 311–315.
- Ho, C. & Zhang, R. (2012), 'Optimal energy allocation for wireless communications with energy harvesting constraints', *IEEE Trans. Signal Process.* **60**(9), 4808–4818.
- Huang, M. & Dey, S. (2006), 'Dynamic quantizer design for hidden Markov state estimation via multiple sensors with fusion center feedback', *IEEE Trans. Signal Process.* **54**(8), 2887–2896.
- Huang, M. & Dey, S. (2007), 'Stability of Kalman filtering with Markovian packet losses', *Automatica* **43**(4), 698–707.
- Liu, X. & Goldsmith, A. (2004), Kalman filtering with partial observation losses, in 'CDC', pp. 4180–4186.
- Mo, Y. & Sinopoli, B. (2008), A characterization of the critical value for Kalman filtering with intermittent observations, in 'CDC', pp. 2692–2216.
- Molin, A. & Hirche, S. (2014), 'On the optimality of certainty equivalence for event-triggered control systems', **58**(2), 470–474.
- Nourian, M., Leong, A. & Dey, S. (2014), 'Optimal energy allocation for Kalman filtering over packet dropping links with imperfect acknowledgments and energy harvesting constraints', *IEEE Trans. Autom. Control* **59**(8), 2128–2143.
- Nourian, M., Leong, A., Dey, S. & Quevedo, D. (2014), 'An optimal transmission strategy for kalman filtering over packet dropping links with imperfect acknowledgements', *IEEE Trans. Control Netw. Syst.* **1**(3), 259–271.
- Ozel, O., Tutuncuoglu, K., Yang, J., Ulukus, S. & Yener, A. (2011), 'Transmission with energy harvesting nodes in fading wireless channels: Optimal policies', *IEEE J. Sel. Areas Commun.* **29**(8), 1732–1743.
- Prabuchandran, K. J., Meena, S. & Bhatnagar, S. (2013), 'Q-learning based energy management policies for a single sensor node with finite buffer', *IEEE Wireless Commun. Lett.* **2**(1), 82–85.
- Proakis, J. (2001), *Digital Communications*, 4th edn, New York: McGraw-Hill.
- Quevedo, D., Ahlén, A. & Johansson, K. (2013), 'State estimation over sensor networks with correlated wireless fading channels', *IEEE Trans. Autom. Control* **58**(3), 581–593.
- Quevedo, D., Ahlén, A., Leong, A. & Dey, S. (2012), 'On Kalman filtering over fading wireless channels with controlled transmission powers', *Automatica* **48**(7), 1306–1316.
- Quevedo, D., Ahlén, A. & Østergaard, J. (2010), 'Energy efficient state estimation with wireless sensors through the use of predictive power control and coding', *IEEE Trans. Signal Processing* **58**(9), 4811–4823.
- Sadeghi, P., Kennedy, R., Rapajic, P. & Shams, R. (2008), 'Finite-state markov modeling of fading channels - a survey of principles and applications', *IEEE Signal Processing Magazine* **25**(5), 57–80.
- Schäl, M. (1993), 'Average optimality in dynamic programming with general state space', *Mathematics of Operational Research* **18**(1), 163–172.
- Schenato, L. (2008), 'Optimal estimation in networked control systems subject to random delay and packet drop', *IEEE Trans. Autom. Control* **53**(5), 1311–1317.
- Schenato, L., Sinopoli, B., Franceschetti, M., Poolla, K. & Sastry, S. (2007), 'Foundations of control and estimation over lossy networks', *Proc. IEEE* **95**(1), 163–187.
- Sharma, V., Mukherji, U., Joseph, V. & Gupta, S. (2010), 'Optimal energy management policies for energy harvesting sensor nodes', *IEEE Trans. Wireless Commun.* **9**(4).
- Shi, L., Cheng, P. & Chen, J. (2011), 'Sensor data scheduling for optimal state estimation with communication energy constraints', *Automatica* **47**(8), 1693–1698.
- Sijs, J. & Lazar, M. (2012), 'Event based state estimation with time synchronous updates', *IEEE Transactions on Automatic Control* **57**(10), 2650–2655.
- Sinopoli, B., Schenato, L., Franceschetti, M., Poolla, K., Jordan, M. & Sastry, S. (2004), 'Kalman filtering with intermittent observations', *IEEE Trans. Autom. Control* **49**(9), 1453–1464.
- Sinopoli, B., Schenato, L., Franceschetti, M., Poolla, K. & Sastry, S. (2004), Time varying optimal control with packet losses, in 'CDC', pp. 1938–1943.
- Sinopoli, B., Schenato, L., Franceschetti, M., Poolla, K. & Sastry, S. (2005a), LQG control with missing observation and control packets, in 'IFAC World Congress'.
- Sinopoli, B., Schenato, L., Franceschetti, M., Poolla, K. & Sastry, S. (2005b), An LQG optimal linear controller for control systems with packet losses, in 'CDC', pp. 458–463.
- Sinopoli, B., Schenato, L., Franceschetti, M., Poolla, K. & Sastry, S. (2005c), Optimal control with unreliable communication: the TCP case, in 'ACC', pp. 3354–3359.
- Sutton, R. & Barto, A. (1998), *Reinforcement learning: An introduction*, Cambridge Univ Press.
- Tutuncuoglu, K. & Yener, A. (2012), 'Optimum transmission policies for battery limited energy harvesting nodes', *IEEE Trans. Wireless Commun.* **11**(3), 1180–1189.
- Xu, Y. & Hespanha, J. (2005), Estimation under uncontrolled and controlled communications in networked control systems, in 'CDC', pp. 842–847.
- Yang, J., Ozel, O. & Ulukus, S. (2012), 'Broadcasting with an energy harvesting rechargeable transmitter', *IEEE Trans. Wireless Commun.* **11**(2), 571–583.