

# On Linear Encoder-Decoder Design for Multi-sensor State Estimation Subject to Quantization Noise and Channel Erasure

Amirpasha Shirazinia and Subhrakanti Dey

**Abstract**—We consider remote state estimation of a scalar stationary linear Gauss-Markov process observed via noisy measurements obtained by two sensors. The sensors can construct a causal linear function of their measurements, which are quantized and transmitted to a decoder (or fusion centre (FC)) over channels which are prone to packet erasures. We design linear encoding and decoding strategies for estimating the state of the linear system that allow improved estimation performance in the presence of packet erasures and quantization errors. To this end, we construct and compare various distributed encoding and decoding methods without any feedback from the FC regarding the channel erasures. We also design various decentralized benchmark methods that either assume perfect feedback from the FC or in addition co-location of the two sensors resulting in a centralized scheme with diversity. These benchmark methods provide various lower bounds for the distributed encoding-decoding schemes designed without feedback. Numerical results indicate i) that optimal decentralized design of the encoders and the decoder in the absence of feedback can provide a remote state estimation performance that is comparable to those achieved by the lower bounds (with feedback) particularly when the sensors are identical and their channels are symmetric, and (ii) a little feedback from the decoder can improve the performance considerably when the channels are asymmetric (i.e. the packet erasure probabilities are unequal).

## I. INTRODUCTION

Rapid advances in sensing and wireless technologies have led to a significant interest in understanding and development of wireless sensor networks (WSNs). Though easier to realize, wireless communication systems come at the price of lower channel capacity which results in higher quantization noise, packet loss and delay. Therefore, such communication constraints need to be considered in the design of WSNs. In the last decade or so, remote sensing and estimation in WSNs have witnessed significant progress in dynamical systems which evolve over time (see e.g., [1]–[3] and references therein).

In this work, we are interested in exploring the problem of remotely estimating the state of a stable stochastic scalar dynamic linear time-invariant (LTI) system (equivalently, a stationary Gauss-Markov process) that is being observed in the presence of noise by two sensors. From each terminal, the noisy measurement is encoded (pre-processed), quantized and then transmitted over a communication channel subject to packet erasures, to a remote node (or FC) for decoding and estimation purposes. In such system, we aim to design strategies for encoding the measurements of the sensors and decoding the state of the LTI system that allow improved estimation performance in the presence of packet erasures, quantization errors, process and measurement noises.

Assuming Gaussian communication channels, linear encoding/decoding strategies are shown to be optimal with respect to minimizing a quadratic cost function in a single-sensor LTI system with Gaussian noise [4]. The optimality, however, cannot be generalized to multi-sensor LTI systems. This is shown, for example, in [5] via a counter example, and in [6], where the authors proposed a non-linear encoding strategy in a multi-sensor LTI system, equipped with a controller, which outperforms the best linear scheme in mean-square sense. Linear schemes, although not optimal in multi-sensor setups, are easier to analyze. Moreover, finding the optimal non-linear policy is difficult in general. Hence, in this work, we focus on exploring linear encoding/decoding strategies that can be implemented with reasonable complexity.

Under the model of channel erasures in a multi-sensor system (but without quantization noise), the authors in [7] obtain necessary and sufficient conditions for stabilization of an LTI system in mean-square sense by identifying an encoding algorithm, where each sensor has access to channel erasure information of all sensors. In [8], the authors proposed a linear encoding/decoding scheme for a single-sensor system setup when channel erasures are considered along with quantization noise. An important result of [8] is that the optimal strategy among all possible linear encoders corresponds to the transmission of the Kalman filter innovations under the assumption that the encoder has perfect knowledge of the channel. In this work, we generalize the result of [8] by considering two sensors and correspondingly two transmission links to the receiver. The results out of this generalization, as will be shown later, are non-trivial. Note that although we consider only two sensors, the results derived here can be easily generalized to any number of sensors.

In particular, we propose two benchmark methods where the first (Benchmark I) is based on the fact that a perfect feedback link is available from the decoder to the encoders, and in the second benchmark method (Benchmark II), besides the perfect receiver feedback, we assume that sensors are co-located such that encoders can be designed in a centralized manner. Due to resource constraints, channel feedback may not be always available, and sensors are generally not co-located in distributed remote estimation applications. Therefore we also propose distributed encoding/decoding schemes, where the encoders are not aware of the receiver state estimate. We propose a soft-innovation forwarding (SIF) scheme where each encoder transmits a convex combination of transmitting the state estimate at the encoder (state-forwarding (SF)) and the

difference between encoder state estimate and the predicted state estimate at the receiver (innovation forwarding (IF)). Although the SIF strategy has a higher design complexity compared to SF or IF, its performance is superior to the latter methods for low erasure probabilities. Our numerical simulations also show that the encoding-decoding designs with receiver feedback considerably improve the performance compared to the case where the feedback is not available.

The rest of the paper is organized as follows. Section II presents the system description and problem formulation. Section III presents the encoding/decoding schemes with perfect receiver feedback, while Section IV presents the soft innovation forwarding scheme with no receiver feedback. Section V presents numerical results, followed by concluding remarks in Section VI.

## II. SYSTEM DESCRIPTION AND PRELIMINARIES

We study the system model depicted in Figure 1. We

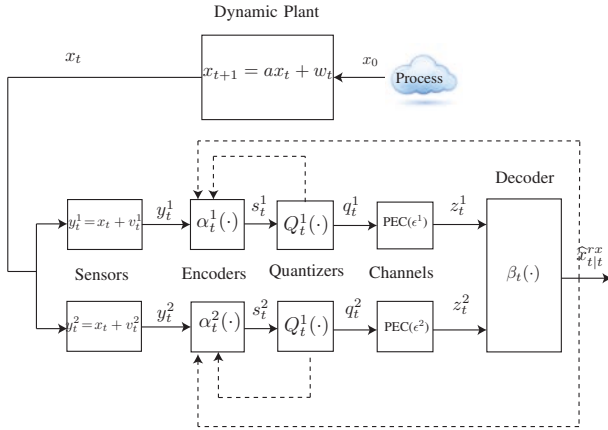


Fig. 1: Dynamical system model for decentralized transmission and estimation with two terminals. Potential feedback links are specified by dashed arrow lines.

consider the problem of remotely estimating the state of a scalar linear stochastic dynamical system in a two-terminal setup (comprised of two sensors, two encoders, two quantizers and two communication channels). System and Measurement equations are expressed as follows<sup>1</sup>

$$x_{t+1} = ax_t + w_t, \quad y_t^i = c^i x_t + v_t^i, \quad i \in \{1, 2\}, \quad (1)$$

where the process noise  $w_t \sim \mathcal{N}(0, (\sigma_w)^2)$  and the measurement noise  $v_t^i \sim \mathcal{N}(0, (\sigma_v^i)^2)$ ,  $i \in \{1, 2\}$  are white, mutually uncorrelated and uncorrelated with the initial state  $x_0 \sim \mathcal{N}(\bar{x}_0, (\sigma_0)^2)$ . As shown in Figure 1, the measurements  $y_t^i$  are pre-processed by causal linear encoders  $\alpha_t^i(\cdot)$  into an analog signal  $s_t^i$  before transmission. The signals are then quantized, using quantizers  $Q_t^i(\cdot)$ , into code-points  $q_t^i$  from finite alphabets, which are then transmitted over digital noisy channels. The channels are modeled as packet erasure channels (PECs), where their outputs are either exactly received or erased. Therefore, channel outputs are  $z_t^i = \gamma_t^i q_t^i$ , where  $\gamma_t^i \in \{0, 1\} \equiv \{\text{erased, received}\}$ , and  $\Pr(\gamma_t^i = 0) = \epsilon^i$ .

<sup>1</sup>We denote time indexes by subscripts, and indexes associated with a terminal are denoted by superscripts. In order to remove ambiguity, squared values are demonstrated by parentheses; for example,  $(g_t^2)^2$  means the squared value of  $g_t$  associated with terminal 2, while  $g_t^1$  means the value of  $g_t$  associated with terminal 1.

At the receiver side, the linear decoder, characterized by the function  $\beta_t^i(\cdot)$ , uses the noisy received signals  $z_t^i$  and provides the estimate of the state at time instant  $t$ .

Throughout the paper, we consider that the dynamical plant is stable, i.e.,  $|a| < 1$ , and is operating in open loop. This assumption is reasonable in many applications, such as remote sensing and estimation, where the dynamical system to be controlled is already stable, but the existing communication and feedback performance can be substantially improved. Without loss of generality, we can assume that the scaling coefficient  $c^i = 1$ . Further, we assume that the quantizer outputs  $q_t^i$  are available to their corresponding sensors. Motivated by the fact that the quantization noise incurred by using a uniform scalar quantizer can be modeled as an additive Gaussian random variable with zero mean and known variance under a high rate quantization assumption (see [8]–[10]), we have  $q_t^i = s_t^i + n_t^i$ , where  $n_t^i \sim \mathcal{N}(0, \frac{1}{\Lambda^i} \mathbb{E}[(s_t^i)^2])$ . Here,  $\Lambda^i$  is the signal-to-quantization noise ratio (SQNR) of the quantizer. Later, in Section V, we show how to calculate  $\Lambda^i$  based on a given quantization rate.

We also assume that there is potentially a delay-free error-free channel feedback from the receiver to the encoders. Note that this assumption is an idealized and may not always be practical. However, the performance provided by schemes based on such feedback, which is the focus of Section III, serves as a benchmark against the methods (focus of Section IV) in which the feedback is not available.

Define the measurement sets  $\mathcal{Y}_t^i$ , quantizer output sets  $\mathcal{Q}_t^i$ , and received signal set  $\mathcal{R}_t$ , respectively, as

$$\begin{aligned} \mathcal{Y}_t^i &\triangleq \{y_0^i, y_1^i, \dots, y_t^i\}, \quad \mathcal{Q}_t^i \triangleq \{q_0^i, q_1^i, \dots, q_t^i\}, \quad i \in \{1, 2\}, \\ \mathcal{R}_t &\triangleq \{z_0^1, \dots, z_t^1, z_0^2, \dots, z_t^2, \gamma_0^1, \dots, \gamma_t^1, \gamma_0^2, \dots, \gamma_t^2\}. \end{aligned} \quad (2)$$

We aim to design the linear encoders  $\alpha_t^i(\cdot)$  and the linear decoder  $\beta_t(\cdot)$  such that the estimation prediction error variance at the decoder, i.e.,

$$p_{t+1|t}^{rx} = \mathbb{E}[(x_{t+1} - \hat{x}_{t+1|t}^{rx})^2 | \mathcal{R}_t] \quad (3)$$

is minimized. In (3), the expectation is taken over all sources of randomness in the system including process, measurement and quantization noises as well as the erasure process  $\{\gamma_t^i\}$ .  $\hat{x}_{t+1|t}^{rx}$  denotes the decoder's predicted state estimate given  $\mathcal{R}_t$ .

We also define the state estimator at the transmitter side as  $\hat{x}_{t|t}^{tx} \triangleq \mathbb{E}[x_t | \mathcal{Y}_t^i]$ . Using Kalman filter analysis [11, Chapter 13], it follows that

$$\begin{aligned} \hat{x}_{t|t}^{tx} &= a\hat{x}_{t-1|t-1}^{tx} + \hat{k}_t^i(y_t^i - a\hat{x}_{t-1|t-1}^{tx}), \\ \hat{x}_{t+1|t}^{tx} &= a\hat{x}_{t|t}^{tx} \end{aligned} \quad (4)$$

where  $\hat{k}_t^i$  is the Kalman filter gain to be optimized. Now, let us define the estimation error at encoder  $i$  as  $\tilde{x}_{t|t}^{tx} = x_t - \hat{x}_{t|t}^{tx}$  and the error variance as  $p_{t|t}^{tx} \triangleq \mathbb{E}[(\tilde{x}_{t|t}^{tx})^2]$ , then based on the fact that the system is asymptotically stable, the estimation error variance has the property that  $\lim_{t \rightarrow \infty} p_{t|t}^{tx} = p^{tx}$  and  $\lim_{t \rightarrow \infty} \hat{k}_t^i = \hat{k}^i$  where  $p^{tx}$  is the solution of the following quadratic equation [8]

$$p^{tx} = (\sigma_v^i)^2 \frac{(a)^2 p^{tx} + (\sigma_w)^2}{(a)^2 p^{tx} + (\sigma_w)^2 + (\sigma_v^i)^2}, \quad (5)$$

and the (asymptotic) optimal Kalman filter gain is given by

$$\hat{k}^i = \frac{(a)^2 p^{tx^i} + (\sigma_w)^2}{(a)^2 p^{tx^i} + (\sigma_w)^2 + (\sigma_v^i)^2}. \quad (6)$$

The above equations (5) and (6) are later used in the design of encoding schemes for the case with no receiver feedback.

### III. IDEALIZED DESIGN METHODS: PERFECT CHANNEL FEEDBACK & A CENTRALIZED SCHEME

In this section, we investigate two benchmark methods, where in the first one (Benchmark I) perfect channel feedback is available to the encoders where the decoder is able to send its predicted estimate to both encoders, and in the second one (Benchmark II), along with the channel feedback, the encoders have access to each other's measurements (centralized scenario). Note that with a slight abuse of notation, we denote the transmitter Kalman gains by  $\hat{k}^i$  although we actually optimize over them when there is perfect feedback, instead of using (6). This allows for greater flexibility in this scenario.

#### A. Benchmark Method I

Inspired by the results of [8, Theorem 1] for a single-sensor scenario, we propose an encoding scheme with perfect channel feedback where the decoder sends its predicted state estimate to both encoders. The linear encoding scheme at the sensors is shown in (7), which provides a lower-bound on the estimation error variance for the linear system (1) achievable under any linear encoding scheme when channel feedback is not available. The proposed linear encoding scheme is based on transmitting the innovation signal

$$s_t^i = \alpha_t^i (\mathcal{Y}_t^i, \mathcal{R}_{t-1}) = \hat{x}_{t|t}^{tx^i} - \hat{x}_{t|t-1}^{rx} = \mathbb{E}[x_t | \mathcal{Y}_t^i] - \mathbb{E}[x_t | \mathcal{R}_{t-1}], \quad (7)$$

where, based on the perfect knowledge of the receiver predicted estimate  $\hat{x}_{t|t-1}^{rx}$ , we can obtain the state estimates at the encoder as

$$\hat{x}_{t|t}^{tx^i} = \hat{x}_{t|t-1}^{rx} + \hat{k}^i (y_t^i - \hat{x}_{t|t-1}^{rx}) = \hat{x}_{t|t-1}^{rx} + \hat{k}^i (\hat{x}_{t|t-1}^{rx} + v_t^i), \quad (8)$$

where  $\hat{k}^i$  is the Kalman filter gain to be optimized, and  $\hat{x}_{t|t-1}^{rx} \triangleq x_t - \hat{x}_{t|t-1}^{rx}$  is the receiver prediction error. Note that (8) is obtained by replacing the sensors' individual predicted estimates in (4) by the receiver predicted estimate (which contains information (albeit lossy) from both sensors). Later in Remark 2, we explain a variation of (8) where the sensors' own predicted estimates can be used in updating their filtered estimates.

The Kalman filter at the decoder is given by

$$\hat{x}_{t+1|t}^{rx} = a \hat{x}_{t|t-1}^{rx} + a [k^1 \ k^2] \begin{bmatrix} \gamma_t^1 (\hat{x}_{t|t}^{tx^1} - \hat{x}_{t|t-1}^{rx} + n_t^1) \\ \gamma_t^2 (\hat{x}_{t|t}^{tx^2} - \hat{x}_{t|t-1}^{rx} + n_t^2) \end{bmatrix} \quad (9)$$

where  $k^i$  are the decoder's Kalman filter gains to be optimized. Subtracting from  $x_{t+1} = ax_t + w_t$ , it follows that

$$\tilde{x}_{t+1|t}^{rx} = a \left( 1 - \sum_{i=1}^2 k^i \hat{k}^i \gamma_t^i \right) \tilde{x}_{t|t-1}^{rx} + w_t - a \sum_{i=1}^2 k^i \gamma_t^i (\hat{k}^i v_t^i + n_t^i). \quad (10)$$

**Remark 1.** It is well known that the optimal Kalman gain for a filter using intermittent signals transmitted from the sensors in general is time-varying [12] but imposing a constant gain that minimizes the expected state estimation error (when the

expectation is taken over the packet loss sequence) can achieve a performance close to that with the optimal time-varying gain, and at the same time, makes the analysis more tractable [13]. In the single-sensor case, the optimal receiver Kalman gain was also shown to be constant when the sensor transmits innovations [8].

Now, the error variance given the sequences  $\gamma_t^i$ ,  $i \in \{1, 2\}$ , is defined as  $p_{t+1|t}^{rx} \triangleq \mathbb{E}[(\tilde{x}_{t+1|t})^2 | \gamma_t^1 \gamma_t^2]$ . Then, as the transmitted signals over the channel are  $s_t^i = \hat{x}_{t|t}^{tx^i} - \hat{x}_{t|t-1}^{rx} = \hat{k}^i (\hat{x}_{t|t-1}^{rx} + v_t^i)$ , the variance of quantization noise becomes

$$(\sigma_n^i)^2 = \frac{\mathbb{E}[(\hat{x}_{t|t}^{tx^i} - \hat{x}_{t|t-1}^{rx})^2]}{\Lambda^i} = \frac{(\hat{k}^i)^2 (p_{t|t-1}^{rx} + (\sigma_v^i)^2)}{\Lambda^i}. \quad (11)$$

Taking the mean of squared terms in right- and left-hand sides of (10) (conditioned on the sequences  $\gamma_t^1$  and  $\gamma_t^2$ ) yields  $p_{t+1|t}^{rx} = A_t p_{t|t-1}^{rx} + N_t$ , where  $A_t \triangleq a^2 \left( 1 - \sum_{i=1}^2 k^i \hat{k}^i \gamma_t^i \right)^2$ , and  $N_t \triangleq \sigma_w^2 + a^2 \sum_{i=1}^2 (k^i)^2 \gamma_t^i \left( (\hat{k}^i \sigma_v^i)^2 + (\sigma_n^i)^2 \right)$ .

Next, let us define  $A_{ij} \triangleq A_t |_{\gamma_t^i=i, \gamma_t^j=j}$  and  $N_{ij} \triangleq N_t |_{\gamma_t^i=i, \gamma_t^j=j}$ ,  $i, j \in \{0, 1\}$  and  $\bar{p}_{t|t-1}^{rx} = \mathbb{E}_{\gamma_t^1 \gamma_t^2} [p_{t|t-1}^{rx}]$ . Then by averaging  $p_{t+1|t}^{rx}$  over the sequences  $\gamma_t^1, \gamma_t^2$ , and using the Bayes' rule, we get

$$\bar{p}_{t+1|t}^{rx} = (A_{11} \bar{p}_{t|t-1}^{rx} + N_{11})(1 - \epsilon^1)(1 - \epsilon^2) + (A_{00} \bar{p}_{t|t-1}^{rx} + N_{00})\epsilon^1 \epsilon^2 + (A_{01} \bar{p}_{t|t-1}^{rx} + N_{01})\epsilon^1(1 - \epsilon^2) + (A_{10} \bar{p}_{t|t-1}^{rx} + N_{10})(1 - \epsilon^1)\epsilon^2. \quad (12)$$

Since  $|a| < 1$ ,  $A_{ij} < 1$  and  $N_{ij} < 1$ ,  $\forall i, j \in \{0, 1\}$ , it can be shown that  $\bar{p}_{t|t-1}^{rx}$  reaches its steady-state value, denoted by  $\bar{p}^{rx}$ , as  $t \rightarrow \infty$ . We note that our goal is to optimize the Kalman filter gains at the encoders  $\hat{k}^i$  and at the decoder  $k^i$  in the sense of minimizing the estimation prediction error variance (12). To do so, we define  $\beta^i \triangleq k^i \hat{k}^i$ , and solve  $\frac{\partial \bar{p}^{rx}}{\partial \beta^1} = 0$ ,  $\frac{\partial \bar{p}^{rx}}{\partial \beta^2} = 0$  for  $\beta^1$  and  $\beta^2$ , which yields

$$\beta^1 = \frac{\bar{p}^{rx} [(1 + 1/\Lambda^2)(\bar{p}^{rx} + (\sigma_v^2)^2) - (1 - \epsilon^2)\bar{p}^{rx}]}{\prod_{i=1}^2 (1 + 1/\Lambda^i)(\bar{p}^{rx} + (\sigma_v^i)^2) - (1 - \epsilon^1)(1 - \epsilon^2)(\bar{p}^{rx})^2} \quad (13a)$$

$$\beta^2 = \frac{\bar{p}^{rx} [(1 + 1/\Lambda^1)(\bar{p}^{rx} + (\sigma_v^1)^2) - (1 - \epsilon^1)\bar{p}^{rx}]}{\prod_{i=1}^2 (1 + 1/\Lambda^i)(\bar{p}^{rx} + (\sigma_v^i)^2) - (1 - \epsilon^1)(1 - \epsilon^2)(\bar{p}^{rx})^2} \quad (13b)$$

It is clear that the optimal  $k^i$  and  $\hat{k}^i$  can be chosen as arbitrary positive values whose product gives the optimal  $\beta^i$ . This is beneficial because one can choose the transmitter gain  $\hat{k}^i$  to satisfy a transmit power constraint for example, and then choose  $k^i$  as  $\frac{\beta^i}{\hat{k}^i}$ . Finally, the resulting steady-state estimation prediction error  $\bar{p}^{rx}$  is obtained by plugging (13a) and (13b) back into (12) (assuming steady-state parameters), which yields a quartic equation in terms of  $\bar{p}^{rx}$  that can be solved numerically using standard algorithms.

**Remark 2.** Note that in (8), we consider the receiver predicted estimate in order to design the encoding scheme. However, another variation of this method can be considered by only using the sensors' individual predicted estimates. The derivations of this new scheme (which can be regarded as another benchmark) are omitted due to space limitations. However, we have observed via numerical studies that the performance of



this scheme closely follows that of benchmark I unless there is a substantial asymmetry in terminals' parameters, e.g., erasure probabilities, quantization rates, etc.

### B. Benchmark Method II

Now, we investigate a centralized scheme, where the encoders at the transmitter side can be combined to give a single state estimation. As we exploit the information from both sensors, it is expected that the state estimation is more accurate using this centralized scheme. We still assume that a perfect channel feedback is available so that we can exploit the decoder's estimate at the centralized encoder to obtain the innovation signal, which is transmitted by both sensors to achieve a diversity scheme over the erasure channels. Note that encoding and decoding processes are the same as those of Benchmark I, introduced in Section III-A. Therefore, this method is expected to give the best performance among the benchmark methods based on perfect feedback. Once again, we use constant Kalman filter gains at the encoder and decoder.

The centralized state estimate at the transmitter side can be written as

$$\hat{x}_{t|t}^{tx} = \hat{x}_{t|t-1}^{rx} + \hat{k}^1(\hat{x}_{t|t-1}^{rx} + v_t^1) + \hat{k}^2(\hat{x}_{t|t-1}^{rx} + v_t^2) \quad (14)$$

Hence,  $\hat{x}_{t|t}^{tx} - \hat{x}_{t|t-1}^{rx} = \sum_{i=1}^2 \hat{k}^i \hat{x}_{t|t-1}^{rx} + \sum_{i=1}^2 \hat{k}^i v_t^i$ . Also,

$$\hat{x}_{t+1|t} = a\hat{x}_{t|t-1}^{rx} + a[k_1 \ k_2] \begin{bmatrix} \gamma_t^1(\hat{x}_{t|t}^{tx} - \hat{x}_{t|t-1}^{rx} + n_t^1) \\ \gamma_t^2(\hat{x}_{t|t}^{tx} - \hat{x}_{t|t-1}^{rx} + n_t^2) \end{bmatrix}, \quad (15)$$

Subtracting (15) from  $x_{t+1} = ax_t + w_t$ , and defining  $\hat{x}_{t|t-1}^{rx} = x_t - \hat{x}_{t|t-1}^{tx}$ , we obtain

$$\begin{aligned} \hat{x}_{t+1|t}^{rx} &= a \left( 1 - (\hat{k}^1 + \hat{k}^2) \sum_{i=1}^2 k^i \gamma_t^i \right) \hat{x}_{t|t-1}^{rx} \\ &+ w_t - a(k^1 \gamma_t^1 + k^2 \gamma_t^2)(\hat{k}^1 v_t^1 + \hat{k}^2 v_t^2) - a \sum_{i=1}^2 k^i \gamma_t^i n_t^i. \end{aligned} \quad (16)$$

Similar to the previous benchmark method, we define the variance of receiver prediction error given the sequences  $\gamma_t^i$ ,  $i \in \{1, 2\}$ , as  $p_{t+1|t}^{rx} \triangleq \mathbb{E}[(\hat{x}_{t+1|t} - \hat{x}_{t|t-1}^{rx})^2 | \gamma_t^1, \gamma_t^2]$ . Then, as the transmitted signals over the channel are  $s_t^i = \hat{x}_{t|t}^{tx} - \hat{x}_{t|t-1}^{rx} = \hat{x}_{t|t-1}^{rx} - \hat{x}_{t|t-1}^{tx}$ , the variance of quantization noise becomes

$$(\sigma_n^i)^2 = \frac{(\hat{k}^1 + \hat{k}^2)^2 p_{t|t-1}^{rx} + \sum_{i=1}^2 (\hat{k}^i \sigma_v^i)^2}{\Lambda^i}. \quad (17)$$

Taking the mean of squared terms in right- and left-hand sides of (16) (conditioned on the sequences  $\gamma_t^1$  and  $\gamma_t^2$ ) yields  $p_{t+1|t}^{rx} = A_t p_{t|t-1}^{rx} + N_t$ , where  $A_t \triangleq a^2 \left( 1 - (\hat{k}^1 + \hat{k}^2) \sum_{i=1}^2 k^i \gamma_t^i \right)^2$ , and

$$N_t \triangleq \sigma_w^2 + a^2(k^1 \gamma_t^1 + k^2 \gamma_t^2)^2 ((\hat{k}^1 \sigma_v^1)^2 + (\hat{k}^2 \sigma_v^2)^2) + a^2 \sum_{i=1}^2 \gamma_t^i (k^i \sigma_n^i)^2.$$

Next, let us define  $A_{ij} \triangleq A_t |_{\gamma_t^1=i, \gamma_t^2=j}$  and  $N_{ij} \triangleq N_t |_{\gamma_t^1=i, \gamma_t^2=j}$ ,  $i, j \in \{0, 1\}$  and  $\bar{p}_{t|t-1}^{rx} = \mathbb{E}_{\gamma_t^1, \gamma_t^2} [p_{t|t-1}^{rx}]$ . Then by averaging  $p_{t+1|t}^{rx}$  over the sequence  $\{\gamma_t^1, \gamma_t^2\}$ , and using Bayes' rule, we get

$$\begin{aligned} \bar{p}_{t+1|t}^{rx} &= (A_{11} \bar{p}_{t|t-1}^{rx} + N_{11})(1-\epsilon^1)(1-\epsilon^2) + (A_{00} \bar{p}_{t|t-1}^{rx} + N_{00})\epsilon^1 \epsilon^2 \\ &+ (A_{01} \bar{p}_{t|t-1}^{rx} + N_{01})\epsilon^1(1-\epsilon^2) + (A_{10} \bar{p}_{t|t-1}^{rx} + N_{10})(1-\epsilon^1)\epsilon^2. \end{aligned} \quad (18)$$

Finally, we can obtain the optimal gains  $\{k^1, k^2, \hat{k}^1, \hat{k}^2\}^*$  by minimizing  $\lim_{t \rightarrow \infty} \bar{p}_{t|t-1}^{rx}$ .

### IV. LINEAR ENCODING/DECODING WITH NO FEEDBACK: SOFT-INNOVATION FORWARDING (SIF)

Here we investigate a scenario where there is no channel feedback available from the decoder to the encoders. For this purpose, we can instead exploit the information from quantizers' outputs. That is to say, instead of using  $\hat{x}_{t|t-1}^{rx}$  we employ  $\bar{x}_{t|t-1}^{tx} \triangleq \mathbb{E}[x_t | \mathcal{Q}_{t-1}^i]$  by combining it with the state estimate at the sensors, i.e.,  $\hat{x}_{t|t-1}^{tx} = \mathbb{E}[x_t | \mathcal{Y}^i]$ . This technique is motivated by the *soft innovation forwarding strategy (SIF)* pioneered in [8] for the single sensor case, where a convex combination of the innovations and the state estimate at the encoder is transmitted. Using this method, the transmitted signal is

$$\begin{aligned} s_t^i &\triangleq \alpha_t^i(\mathcal{Y}_t^i, \mathcal{Q}_{t-1}^i, \nu^i) = \nu^i \hat{x}_{t|t}^{tx} + (1-\nu^i)(\hat{x}_{t|t}^{tx} - \bar{x}_{t|t-1}^{tx}) \\ &= \hat{x}_{t|t}^{tx} - (1-\nu^i)\bar{x}_{t|t-1}^{tx} \end{aligned} \quad (19)$$

where  $0 < \nu^i < 1$  is a mixing coefficient to be optimized. Note that state-forwarding (SF) method (sending the state estimate at the sensor, i.e.,  $\nu^i = 1$ ) and innovation-forwarding (IF) method (sending the innovation, which is optimal for the perfect feedback case i.e.,  $\nu^i = 0$ ) are special cases of SIF.

We first write augmented state equations as in the form

$$\boldsymbol{\mu}_{t+1} = \mathbf{A}_t \boldsymbol{\mu}_t + \boldsymbol{\eta}_t, \quad (20)$$

where

$$\boldsymbol{\mu}_t = \begin{bmatrix} x_{t+1} & \hat{x}_{t+1|t+1}^{tx1} & \bar{x}_{t+1|t}^{tx1} & \hat{x}_{t+1|t+1}^{tx2} & \bar{x}_{t+1|t}^{tx2} \end{bmatrix}^\top, \quad (21a)$$

$$\mathbf{A}_t = \begin{bmatrix} a & 0 & 0 & 0 & 0 \\ a\hat{k}^1 & a(1-\hat{k}^1) & 0 & 0 & 0 \\ 0 & \bar{k}^1 & a-\bar{k}^1 & 0 & 0 \\ a\hat{k}^2 & 0 & 0 & a(1-\hat{k}^2) & 0 \\ 0 & 0 & 0 & \bar{k}^2 & a-\bar{k}^2 \end{bmatrix}, \quad (21b)$$

$$\boldsymbol{\eta}_t = \begin{bmatrix} w_t & \hat{k}^1(w_t + v_{t+1}^1) & \bar{k}^1 n_t^1 & \hat{k}^2(w_t + v_{t+1}^2) & \bar{k}^2 n_t^2 \end{bmatrix}^\top. \quad (21c)$$

Here,  $\bar{k}^i$ ,  $i \in \{1, 2\}$ , is the Kalman filter gain (to be optimized) associated with the quantizer predicted estimate. Also, the received measurement equation can be stated as  $\mathbf{z}_t = \mathbf{C}_t \boldsymbol{\mu}_t + \mathbf{n}_t$ , where  $\mathbf{z}_t = [z_t^1 \ z_t^2]^\top$ ,  $\mathbf{n}_t = [\gamma_t^1 n_t^1 \ \gamma_t^2 n_t^2]^\top$ , and

$$\mathbf{C}_t = \begin{bmatrix} 0 & \gamma_t^1 & -(1-\nu_1)\gamma_t^1 & 0 & 0 \\ 0 & 0 & 0 & \gamma_t^2 & -(1-\nu_2)\gamma_t^2 \end{bmatrix}.$$

Now, defining  $\tilde{\boldsymbol{\mu}}_t \triangleq \boldsymbol{\mu}_t - \boldsymbol{\mu}_{t|t-1}$ , we have

$$\begin{aligned} \tilde{\boldsymbol{\mu}}_{t+1} &= \mathbf{A}_t \boldsymbol{\mu}_t + \boldsymbol{\eta}_t - (\mathbf{A}_t - \mathbf{K}_t \mathbf{C}_t) \tilde{\boldsymbol{\mu}}_{t|t-1} - \mathbf{K}_t (\mathbf{C}_t \boldsymbol{\mu}_t + \mathbf{n}_t) \\ &= (\mathbf{A}_t - \mathbf{K}_t \mathbf{C}_t) \tilde{\boldsymbol{\mu}}_t + \boldsymbol{\eta}_t - \mathbf{K}_t \mathbf{n}_t, \end{aligned} \quad (22)$$

where  $\mathbf{K}_t \in \mathbb{R}^{5 \times 2}$  is the Kalman filter gain matrix to be optimized. Note that we are interested in the entries of its first row that minimize  $p_{t+1|t}^{rx}$ .

Next, we define the following covariance matrices

$$\begin{aligned} \mathbf{P}_t &\triangleq \mathbb{E}[\tilde{\boldsymbol{\mu}}_t \tilde{\boldsymbol{\mu}}_t^\top | \gamma_t^1, \gamma_t^2], & \mathbf{Q} &\triangleq \mathbb{E}[\boldsymbol{\eta}_t \boldsymbol{\eta}_t^\top] \\ \mathbf{N}_t &\triangleq \mathbb{E}[\mathbf{n}_t \mathbf{n}_t^\top | \gamma_t^1, \gamma_t^2], & \mathbf{S}_t &\triangleq \mathbb{E}[\boldsymbol{\eta}_t \mathbf{n}_t^\top | \gamma_t^1, \gamma_t^2]. \end{aligned} \quad (23)$$

More explicitly, we can derive  $\mathbf{N}_t$ ,  $\mathbf{S}_t$  and  $\mathbf{Q}$ , which are stated by (24), on top of next page. Note that  $(\sigma_n^i)^2$  can be computed from [8, Equation (38)].

Then, from (22) and (23), we obtain

$$\mathbf{P}_{t+1} = \mathbb{E}[(\mathbf{A}_t - \mathbf{K}_t \mathbf{C}_t) \mathbf{P}_t (\mathbf{A}_t - \mathbf{K}_t \mathbf{C}_t)^\top | \gamma_t^1 \gamma_t^2] + \mathbf{Q} + \mathbf{K}_t \mathbf{N}_t \mathbf{K}_t^\top - \mathbf{K}_t \mathbf{S}_t^\top - \mathbf{S}_t \mathbf{K}_t^\top. \quad (25)$$

Now, let us define  $\bar{\mathbf{P}}_t \triangleq \mathbb{E}_{\gamma_t^1 \gamma_t^2}[\mathbf{P}_t]$ ,  $\mathbf{N}_{ij} \triangleq \mathbf{N}_t | \gamma_t^1 = i, \gamma_t^2 = j$  and  $\mathbf{S}_{ij} \triangleq \mathbf{S}_t | \gamma_t^1 = i, \gamma_t^2 = j$ ,  $i, j \in \{0, 1\}$ . Using the Bayes' rule on (25), we get

$$\bar{\mathbf{P}}_{t+1} = \mathbf{K}_t \mathbf{X}_t \mathbf{K}_t^\top + \mathbf{K}_t \mathbf{Y}_t + \mathbf{Y}_t^\top \mathbf{K}_t^\top + \mathbf{Z}_t = (\mathbf{K}_t + \mathbf{Y}_t \mathbf{X}_t^{-1}) \mathbf{X}_t (\mathbf{K}_t + \mathbf{Y}_t \mathbf{X}_t^{-1})^\top + \mathbf{Z}_t - \mathbf{Y}_t \mathbf{X}_t^{-1} \mathbf{Y}_t^\top \quad (26)$$

where

$$\mathbf{X}_t = (1 - \epsilon^1)(1 - \epsilon^2)(\mathbf{N}_{11} + \mathbf{C}_{11} \bar{\mathbf{P}}_t \mathbf{C}_{11}^\top) - \epsilon^1(1 - \epsilon^2)(\mathbf{N}_{01} + \mathbf{C}_{01} \bar{\mathbf{P}}_t \mathbf{C}_{01}^\top) - \epsilon^2(1 - \epsilon^1)(\mathbf{N}_{10} + \mathbf{C}_{10} \bar{\mathbf{P}}_t \mathbf{C}_{10}^\top) \quad (27a)$$

$$\mathbf{Y}_t = -(1 - \epsilon^1)(1 - \epsilon^2)(\mathbf{S}_{11}^\top + \mathbf{C}_{11} \bar{\mathbf{P}}_t \mathbf{A}_t^\top) - \epsilon^1(1 - \epsilon^2)(\mathbf{S}_{01}^\top + \mathbf{C}_{01} \bar{\mathbf{P}}_t \mathbf{A}_t^\top) - \epsilon^2(1 - \epsilon^1)(\mathbf{S}_{10}^\top + \mathbf{C}_{10} \bar{\mathbf{P}}_t \mathbf{A}_t^\top) \quad (27b)$$

$$\mathbf{Z}_t = \mathbf{A}_t \bar{\mathbf{P}}_t \mathbf{A}_t^\top + \mathbf{Q} \quad (27c)$$

Next, minimizing (26) with respect to  $\mathbf{K}_t$  yields

$$\mathbf{K}_t = -\mathbf{Y}_t \mathbf{X}_t^{-1}, \quad \bar{\mathbf{P}}_t = \mathbf{Z}_t - \mathbf{Y}_t \mathbf{X}_t^{-1} \mathbf{Y}_t^\top. \quad (28)$$

Note that (28) represents the optimal filter gain matrix and corresponding error covariance matrix for a fixed  $\nu^i$ . The optimal  $\mathbf{K}_t$  can be substituted (26) to obtain  $\lim_{t \rightarrow \infty} \bar{\mathbf{P}}_t = \bar{\mathbf{P}}_\infty$ , which can be minimized with respect to  $0 \leq \nu^i \leq 1$ , to obtain

$$\{\nu^1, \nu^2\}^* = \underset{0 \leq \nu^i \leq 1}{\operatorname{argmin}}_{i=1} [\bar{\mathbf{P}}_\infty]_{1,1}. \quad (29)$$

Hence,  $[\bar{\mathbf{P}}_\infty]_{1,1}$  is the optimal receiver prediction error variance, given the optimized  $\{\nu^1, \nu^2\}^*$  and the optimal Kalman filter gains  $k^1$  and  $k^2$  are the ones which are associated with the first row of  $\mathbf{K}_\infty = \lim_{t \rightarrow \infty} \mathbf{K}_t$ . Note that this minimization can only be done numerically as  $[\bar{\mathbf{P}}_\infty]_{1,1}$  is not convex in  $\nu^1, \nu^2$ .

## V. NUMERICAL RESULTS

In this section, we demonstrate the performance of the proposed design methods characterized in Section III and Section IV. We use a uniform quantizer to quantize  $s_t^i$  ( $i \in \{1, 2\}$ ) with a suitable number of quantization levels and saturation thresholds to ensure SQNR equals  $\Lambda^i$ . Following a similar method to [8], the quantization step  $\Delta_q^i$  is chosen so that  $(\sigma_n^i)^2 = \frac{(\Delta_q^i)^2}{12}$ , where  $(\sigma_n^i)^2 = \frac{\operatorname{Var}\{s_t^i\}}{\Lambda^i}$ , which yields  $\Delta_q^i = \sqrt{\frac{12 \operatorname{Var}\{s_t^i\}}{\Lambda^i}}$ . By setting the saturation thresholds  $\pm T_q^i$  according to  $T_q^i = 4\sqrt{\operatorname{Var}\{s_t^i\}}$ , the number of quantization levels is given by  $N_b^i = \lceil \frac{2T_q^i}{\Delta_q^i} \rceil = \lceil 4\sqrt{\frac{\Lambda^i}{3}} \rceil$ , which corresponds to  $N_b^i = \lceil \log_2(4\sqrt{\frac{\Lambda^i}{3}}) \rceil$  number of quantization bits per sample. We also set  $a = 0.95$ ,  $(\sigma_w)^2 = 0.1$  and  $(\sigma_v^i)^2 = 0.05$ .

In our simulation studies, we compare the following strategies: Benchmarks I and II, and designs with no feedback, i.e., SIF, SF (when  $\nu^i = 1$ ) and IF (when  $\nu^i = 0$ ).

We illustrate, in Figure 2, the performance of the ideal methods and non-ideal methods in a symmetric scenario,

where  $\epsilon_1 = \epsilon_2$ . For the SIF method, we optimize  $\nu^i$  using a grid search by varying  $\nu^i$  from 0 to 1 in a step size 0.02.

As expected, the centralized design (benchmark II) outperforms other design methods. Among the non-ideal design methods, SIF reaches the lowest error since it finds the optimal mixing coefficients  $\nu^i$ . It should be noted that the gap between the performance of Benchmark II and the SIF strategy is, of course, achieved at the price of perfect channel feedback. Comparing the methods based on no feedback where both terminals use SF or IF, we see that the former improves the performance for high packet loss probabilities, while the latter gives a better performance at a low packet loss regime, which is consistent with the results in [8]. The *critical* erasure probability where these two curves intersect is approximately  $\epsilon^{cri^1} = \epsilon^{cri^2} = 0.18$ . Now, if a terminal uses the SF method, and the other uses IF, we observe that at an intermediate low packet loss regime, this scheme performs than the schemes where both terminals use SF or IF. As the erasure probability increases, the scheme where both terminals use SF becomes better, and approaches the SIF scheme for large packet loss probabilities.

In the next study, in Figure 3, we choose an asymmetric scenario by fixing  $\epsilon^1 = 0.05$  and varying  $\epsilon^2$ . Note that we have chosen  $\epsilon^1$  such that  $\epsilon^1 < \epsilon^{cri^1} = 0.18$  (obtained by previous simulations). We omit the graphs for the case where  $\epsilon^1 > \epsilon^{cri^1}$  due to space limitations. In Figure 3, the benchmark methods I and II outperform the other schemes, and the SIF method, by optimizing the mixing coefficients  $\nu^i$ , provides a better performance among all the schemes based on no feedback. At low packet loss regime, we observe that both terminals using IF gives a lower error, due to the fact that terminals 1 and 2 benefit from the IF method at low erasure probabilities. When  $\epsilon^2$  increases to 1, it can be also seen that the curve corresponding to terminal 1 using IF and terminal 2 using SF (or, vice versa, terminal 1 using SF and terminal 2 using IF) tends to the curve associated with both terminals using IF (or, vice versa, SF). The reason is that for large  $\epsilon^2$ , terminal 2 becomes practically immaterial, and terminal 1 becomes the only functional link. Thus, the ultimate performance corresponds to that of the forwarding technique used at terminal 1. One can also see that for  $\epsilon_2 < \epsilon_1 = 0.05$ , the strategy with terminal 1 using SF and terminal 2 using IF does better than vice versa, while the behavior reverses when  $\epsilon_2 > \epsilon_1 = 0.05$ .

In the last study, we plot the behavior of optimal mixing coefficient  $\nu^i$  by varying  $\epsilon^2$  and fixing  $\epsilon^1 = 0.2$  in the SIF method. We observe as  $\epsilon^2$  increases,  $\nu^2$  also increases. Interestingly, as  $\epsilon^2$  increases,  $\nu^1$  also increases even though  $\epsilon^1$  is fixed at 0.2. This is an indication that both encoders need to adapt in order to improve the overall performance. Further, the two curves intersect at  $\epsilon^2 = 0.2$ , the point where the system becomes symmetric, i.e.,  $\epsilon^1 = \epsilon^2$  such that  $\nu^1 = \nu^2$ .

## VI. CONCLUSIONS AND EXTENSIONS

In this paper, we proposed distributed linear encoding schemes and the associated decoder design methods for remote estimation of the state of a scalar stable linear dy-

$$\mathbf{N}_t = \begin{bmatrix} \gamma_t^1 (\sigma_n^1)^2 & 0 \\ 0 & \gamma_t^2 (\sigma_n^2)^2 \end{bmatrix}, \quad \mathbf{S}_t = \begin{bmatrix} 0 & 0 & \bar{k}^1 \gamma_t^1 (\sigma_n^1)^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & \bar{k}^2 \gamma_t^2 (\sigma_n^2)^2 \end{bmatrix}^\top,$$

$$\mathbf{Q} = \begin{bmatrix} (\sigma_w)^2 & \hat{k}^1 (\sigma_w)^2 & 0 & \hat{k}^2 (\sigma_w)^2 & 0 \\ \hat{k}^1 (\sigma_w)^2 & (\hat{k}^1)^2 ((\sigma_w)^2 + (\sigma_v^1)^2) & 0 & \hat{k}^1 \hat{k}^2 (\sigma_w)^2 & 0 \\ 0 & 0 & (\bar{k}^1 \sigma_n^1)^2 & 0 & 0 \\ \hat{k}^2 (\sigma_w)^2 & \hat{k}^1 \hat{k}^2 (\sigma_w)^2 & 0 & (\hat{k}^2)^2 ((\sigma_w)^2 + (\sigma_v^2)^2) & 0 \\ 0 & 0 & 0 & 0 & (\bar{k}^2 \sigma_n^2)^2 \end{bmatrix}. \quad (24)$$

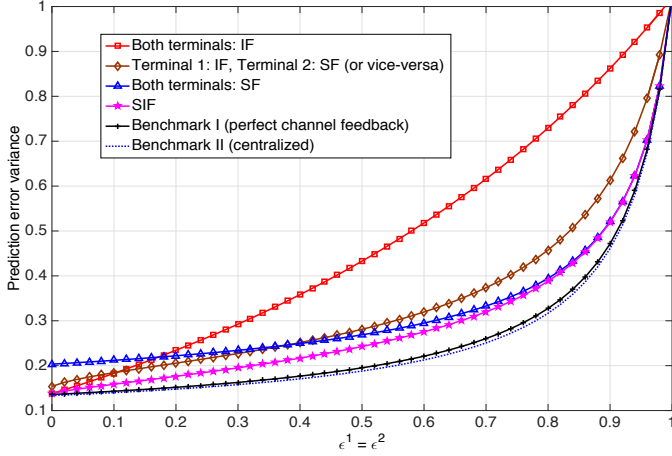


Fig. 2: Prediction error covariance of proposed strategies as a function of erasure probability in a symmetric setup.

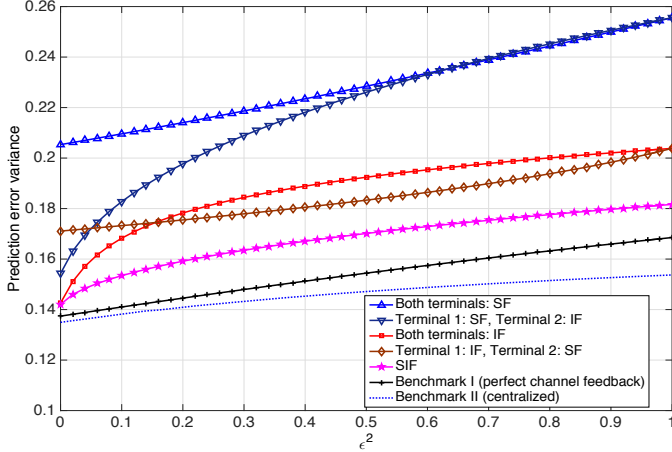


Fig. 3: Prediction error covariance of proposed strategies as a function of  $\epsilon^2$  in an asymmetric setup for a fixed  $\epsilon^1 = 0.05$ .

namical system (or equivalently a stationary Gauss-Markov source) when multiple sensors take noisy measurements and send their linearly encoded signals over channels subject to quantization noise and packet erasures. We propose various benchmark methods when there is perfect feedback from the receiver/decoder, and an extension of the soft innovation forwarding method for the single sensor case in [8] to multiple sensors. Although we present our results for a 2-sensor case, the proposed encoding/decoding methods can be easily extended to an arbitrary number of sensors. How the number of sensors affect the estimation error performance for the various encoding schemes is however requires further analysis, and will be investigated in future work.

#### REFERENCES

- [1] S. Tatikonda and S. Mitter, "Control over noisy channels," *IEEE Transactions on Automatic Control*, vol. 49, no. 7, pp. 1196–1201, 2004.

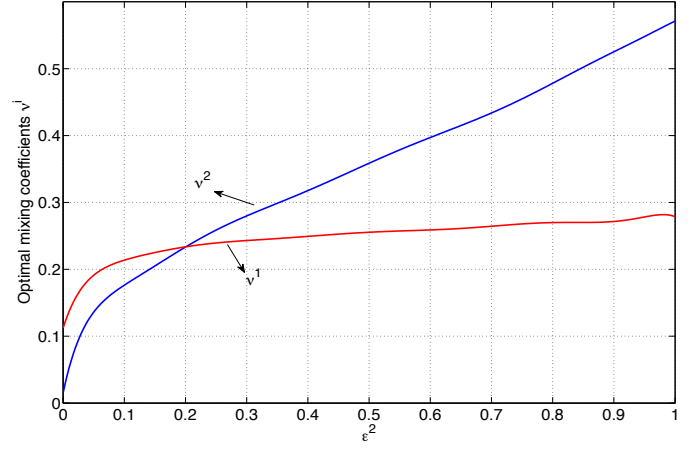


Fig. 4: Optimal mixing coefficients  $\nu^i$  as a function of  $\epsilon^2$  by fixing  $\epsilon^1$  for the SIF method.

- [2] A. Sahai and S. Mitter, "The necessity and sufficiency of anytime capacity for stabilization of a linear system over a noisy communication link; Part I: Scalar systems," *IEEE Transactions on Information Theory*, vol. 52, no. 8, pp. 3369–3395, 2006.
- [3] B. G. N. Nair, F. Fagnani, S. Zampieri, and R. J. Evans, "Feedback control under data rate constraints: An overview," *Proceedings of the IEEE*, vol. 95, no. 1, pp. 108–137, 2007.
- [4] R. Bansal and T. Başar, "Simultaneous design of measurement and control strategies for stochastic systems with feedback," *Automatica*, vol. 25, no. 5, pp. 679–694, 1989.
- [5] S. Yüksel and S. Tatikonda, "A counterexample in distributed optimal sensing and control," *IEEE Transactions on Automatic Control*, vol. 54, no. 4, p. 841, 2009.
- [6] M. Andersson, A. A. Zaidi, N. Wernersson, and M. Skoglund, "Nonlinear distributed sensing for closed-loop control over gaussian channels," in *IEEE Swedish Communication Technologies Workshop (Swe-CTW)*. IEEE, 2011, pp. 19–23.
- [7] V. Gupta, N. C. Martins, and J. S. Baras, "Optimal output feedback control using two remote sensors over erasure channels," *IEEE Transactions on Automatic Control*, vol. 54, no. 7, pp. 1463–1476, 2009.
- [8] S. Dey, A. Chiuso, and L. Schenato, "Remote estimation with noisy measurements subject to packet loss and quantization noise," *IEEE Transactions on Control of Network Systems*, vol. 1, no. 3, pp. 204–217, Sep. 2014.
- [9] D. Marco and D. Neuhoff, "The validity of the additive noise model for uniform scalar quantizers," *IEEE Trans. Info. Theory*, vol. 51, no. 5, pp. 1739–1755, 2005.
- [10] A. Leong, S. Dey, and G. Nair, "Quantized filtering schemes for multi-sensor linear state estimation: Stability and performance under high rate quantization," *IEEE Trans. Sig. Proc.*, vol. 61, no. 15, pp. 3852–3865, 2013.
- [11] S. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ: Prentice Hall, 1993.
- [12] B. Sinopoli, L. Schenato, M. Franceschetti, K. Poolla, M. Jordan, and S. Sastry, "Kalman filtering with intermittent observations," *IEEE Transactions on Automatic Control*, vol. 49, no. 9, pp. 1453–1464, September 2004.
- [13] L. Schenato, "Optimal estimation in networked control systems subject to random delay and packet drop," *IEEE Transactions on Automatic Control*, vol. 53, no. 5, pp. 1311–1317, May 2008.