# Latency-Reliability Tradeoffs for State Estimation

Konstantinos Gatsis, *Member, IEEE*, Hamed Hassani, *Member, IEEE*, and George J. Pappas, *Fellow, IEEE*

***Abstract*—The emerging interest in low-latency high-reliability applications, such as connected vehicles, necessitates a new abstraction between communication and control. Thanks to advances in cyber-physical systems over the past decades, we understand this interface for classical bit-rate models of channels as well as packet-loss-type channels. This article proposes a new abstraction characterized as a tradeoff curve between latency, reliability, and rate. Our aim is to understand—do we (control engineers) prefer faster but less reliable communications (with shorter codes), or slower but more reliable communications (with longer codes)? In this article, we examine the tradeoffs between latency and reliability for the problem of estimating dynamical systems over communication channels. Employing different latency-reliability curves derived from practical coding schemes, we develop a cross-layer design methodology, i.e., select the code length depending on the system dynamics to optimize system performance.**

***Index Terms*—Channel coding, high-reliability, low-latency, networked control systems, state estimation.**

## I. Introduction

RECENT interest in the Internet-of-Things (IoT) and the next generation wireless communication standards (5G) is targeting applications such as connected vehicles, collaborative swarm planning, smart cities, and industrial control [1]. These are challenging applications due to their low-latency high-reliability closed-loop control requirements. This necessitates *rethinking* the communication stack, practical codes, networking architecture, and control design that can provide ultralow latency ($< 1$ms) and very high reliability ($>99.999\%$), which is not possible in today's wireless systems. Even more fundamentally, we lack an understanding of the limits for stability, estimation, and control over low-latency, high-reliability communications.

Manuscript received August 31, 2019; revised January 31, 2020; accepted April 15, 2020. Date of publication May 4, 2020; date of current version February 26, 2021. This work was supported in part by NSF under Grant CPS-1837253, and Grant NSF-CCF 1755707, in part by the Intel Science and Technology Center for Wireless Autonomous Systems (ISTC-WAS) and in part by ARL DCIST CRA under Grant W911NF-17-2-0181. Recommended by Associate Editor W. X. Zheng. *(Corresponding author: Konstantinos Gatsis.)*

Konstantinos Gatsis is with the Department of Engineering Science, University of Oxford, OX1 3PJ Oxford, U.K. (e-mail: konstantinos.gatsis@eng.ox.ac.uk).

Hamed Hassani and George J. Pappas are with the Department of Electrical and Systems Engineering, University of Pennsylvania, Philadelphia, PA 19104 USA (e-mail: hassani@seas.upenn.edu; pappasg@seas.upenn.edu).

Color versions of one or more of the figures in this article are available online at https://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TAC.2020.2992563

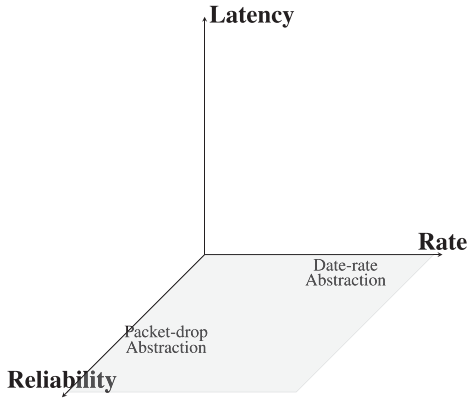



Fig. 1. Rate-latency-reliability design space for estimation and control. Existing works explore the design space spanned by rate and reliability axes.

In this article, we take the first step in understanding the fundamental tradeoffs between latency, reliability, stability, and control performance. More simply, we ask the following.

*Does a dynamical system need faster but less reliable information or slower but more reliable information?*

To answer this question, we propose a new communication abstraction (see Fig. 1) based on the latency-rate-reliability curves obtained from recent developments in information/coding theory for finite blocklengths. Furthermore, we argue fundamentally about the opportunity of designing the communication block (channel coding) adapted to a dynamical (control) system and examine the role the key parameters (i.e. rate, reliability, error, system dynamics) play in such a design. One of the earliest abstractions between communication and control is based on bit rate characterizations. This has permitted a fundamental understanding of the minimum bit rate required for stability over data-rate limited channels [2]–[6] as well as for the case where besides data rate constraints channels also introduce noise [7], [8] or delays [3], [9]. Beyond fundamental characterizations, this has led to an extensive literature on appropriate quantizer, encoder, and decoder designs for control [2], [10]–[20]. A different widely adopted abstraction is packet-based communication where quantization and data rate effects are often ignored. This has permitted the analysis of the maximum packet drop rate above which controlling a plant becomes impossible [21]–[24], as well as observer and controller design to counteract random packet drops [22]–[27].

Finally, the recent interest in low-power IoT devices has increased the development of abstractions in terms of resources, e.g., the number of transmissions, and has resulted in the

0018-9286 © 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: University of Pennsylvania. Downloaded on October 29,2026 at 15:10:15 UTC from IEEE Xplore. Restrictions apply.

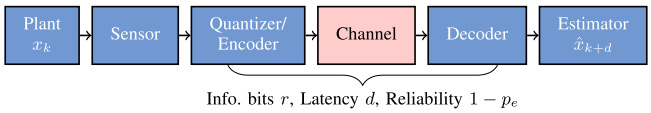



Fig. 2. System for low-latency high-reliability remote estimation. The communication block consisting of encoder, channel, and decoder is abstracted as a latency-reliability curve.

development of frameworks such as event-triggered control [28]–[35] and transmit power allocation [36], [37]. Resource allocation in more complex networks with multiple plants, sensors, or actuators has also attracted attention but primarily without latency considerations. Examples include sharing a communication medium between different sensor and actuators [38]–[42], decentralized mechanisms subject to interferences [43], [44], and control over shared wireless channels [45], [46]. Recent approaches also consider control system operation over multiple links where each link has a different given delay parameter [47], [48].

On the other hand, low-latency communication introduces a novel design parameter that is absent in the above literature. In particular, delay has been treated as a given disturbance to a control system either fixed or random. Longer delays are undesirable because, as expected, they degrade system performance, but they have not been explicitly part of the design. In that sense the existing literature explores the design space spanned by the reliability and data rate axes of Fig. 1 from the perspective of control systems. Our novelty is on analyzing the effect of latency in performance and appropriately guiding the selection of code blocklength.

We note that interest in low-latency communications has also started to appear within the networking community, using the concept of age or freshness of information [49]–[51], or analyzing latency and reliability using network coding and cooperative communication [52]–[55].

In information theory, fundamental limits for low-latency communication have been characterized in [56]–[58] (also known as "nonasymptotic channel coding," see Section III and 28 for details). Such laws demonstrate how the minimum communication (coding) latency scales with rate and reliability. Such laws are typically used as a benchmark for comparing the latency of a coding scheme with the optimal latency. Even though it has been shown that codes with optimal latency exist, so far none of the practical coding schemes (such as polar or iterative codes) have been capable of achieving such optimal latencies [59], [60]. Hence, designing practical codes with optimal latency is a key research frontier in information theory. In this article instead we use such fundamental scaling laws from information theory to abstract the communication block (encoding/channel/decoding) of the system shown in Fig. 2 that includes the physical plant dynamics. In other words, we can find out what the optimal (shortest) range of the transmission length (coding latency) should be from the perspective of dynamical system performance. Such an abstraction of the communication block can then help us to provide fundamental tradeoffs between reliability and latency for the whole system shown in Fig. 2. Even though we do not derive any new blocklength-reliability curves, to the best of our knowledge we are the first to use them to optimize estimation performance for dynamical systems.

In Section II, we describe our setup which involves the remote estimation of a scalar dynamical system over a communication channel. We introduce our abstraction capturing the available information bits, latency, and reliability of the communication block. Using a sequential quantization scheme from the control literature [2], we analyze estimation performance over such a communication block. In particular, we show that reliability plays a crucial role in keeping the estimation stable (cf., Theorem 1) while both latency and reliability need to be taken into account to optimize the steady state estimation performance (cf., Theorem 2).

We proceed in Section III to characterize the role of coding in the proposed abstraction. We utilize known blocklength-reliability curves of both theoretically optimal and practical coding schemes and we consider their effect on the system performance, i.e., the estimation error, exploiting the control-theoretic results of Section II. Longer codes, even though they improve reliability, should be avoided due to the impact their latency has on performance. On the other hand, too short codes do not have the necessary reliability to even stabilize the system. Our contribution is a methodology that facilitates the choice of optimal code length and reveals its relationship with system dynamics and channel conditions. To the best of our knowledge, this is the first time such a fundamental relationship is derived[1]. This analysis is extended to higher-dimensional systems in Section IV.

In summary, the contributions of our article are the following.

1) A new communication abstraction that includes rate, latency, and reliability as design parameters based on finite blocklength codes.
2) An analysis of state estimation performance over this abstraction (see Theorem 2).
3) A methodology for selection of the optimal coding blocklength for the first time for the problem of state estimation.

## II. Problem Setup

Consider the setup indicated in Fig. 2. A sensor is measuring the state of a dynamical process and a remote estimator is interested in maintaining a state estimate of the dynamical process. This is achieved by *communication* between the two entities over a *noisy channel*. Specifically, we consider the discrete time scalar dynamical system—the general case of higher-dimensional systems is treated in Section IV

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

where $k = 0, 1, \ldots$ are the discrete time steps, $x_k \in \mathbb{R}$ is the state of the dynamical system where the initial state $x_0$ is known to lie

[1]We would like to point out that some works have used the term blocklength as a synonym to sampling period over noiseless channels and answer, for example, should one send one message every one time step, or send a message of double the size every two time steps [17], [61]. This is a different problem than our work where sampling period is fixed and communication is over an unreliable channel, so longer blocklengths per sample help counteract the unreliable channel at the expense of transmission latency.

Authorized licensed use limited to: University of Pennsylvania. Downloaded on October 29,2026 at 15:10:15 UTC from IEEE Xplore. Restrictions apply.

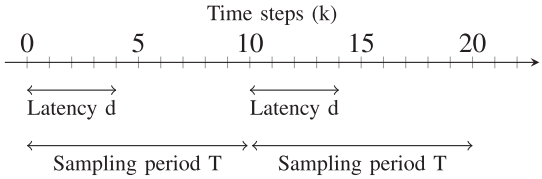



Fig. 3. Time steps indexed by $k = 0, 1, \ldots$ correspond to time intervals required for the transmission of single bits over the channel. The sensor samples and transmits the state of the plant periodically and this period corresponds to $T$ time steps, e.g., $T = 10$ above. The transmission of each sample which consists of multiple bits requires $d$ time steps, e.g., $d = 4$ above.

in some set $[\underline{X}_0, \bar{X}_0] \subset \mathbb{R}$ , and $w_k$ is an unknown disturbance of magnitude $w_k \in [-W/2, W/2]$. This assumption of *bounded* initial condition and disturbance is common in the literature (see, e.g., [2], [7], [13]—see also available approaches for the case of unbounded stochastic disturbances [3], [19]). Equation (1) is derived from discretizing a continuous time dynamical system over small units of time. Without loss of generality, as we will also make clear in Section III we may take each discrete time step normalized to correspond to the time interval required for the transmission of a single bit over the channel. Then we may define all other time intervals as multiples of these discrete units of time.

We let $A \geq 0$ and note that similar analysis can be given for the symmetric case $A \leq 0$. In general, this model captures the important case $A > 1$, i.e., where the system is unstable and the remote estimator is interested in tracking the state with a bounded error even though the state can grow unbounded.

The sensor samples the dynamical system 1 every $T$ discrete time steps, i.e., at times $k = 0, T, 2T, \ldots$. Thus, from the perspective of the sensor, this is equivalent to sampling a dynamical system of the form

$$x_{(\ell+1)T} = A^T x_{\ell T} + \sum_{j=0}^{T-1} A^{T-1-j} w_{\ell T+j} \tag{2}$$

with the index $\ell = 0, 1, \ldots$ counting the number of generated samples. Throughout the article, this sampling period $T$ is fixed and not a design parameter. For notational convenience, we represent the system dynamics in the form 1 with respect to the (shorter) discrete time steps corresponding to transmission intervals of single bits, while the (longer) sampling period is a multiple of these discrete time steps (see also Fig. 3).

At the sensor each sample is converted (quantized) to an $r$-bit message. The message to be communicated is transformed into a generally longer $n$-bit message by using some form of channel coding procedure (see details in Section III). The transmission of the coded message to the remote estimator introduces a *delay*[2]. In particular, in this article we account for the time to transmit each bit over the channel. This is motivated for example by interest in new wireless networks that can provide transmission latencies at the order of 1 ms while industrial applications have sampling periods at the order of 10 s of ms [62], [63]. We note that in practice we can extend our model to include other message overheads that introduce delays, as well as the process of encoding and decoding which introduces computational delay, or there might be uncertainty in the delay. Additionally, communication introduces noise. We make the following assumptions for the delay, noise, and availability of channel feedback. As we will see, these assumptions are fairly general depending on how the parameters are set and cover a variety of practical scenarios.

*Assumption 1:* Each $r$-bit message requires a delay of $d$ time steps, with $d \leq T$. Each message is either successfully decoded with probability $1 - p_e$, or with probability $p_e$ it is corrupted during the communication and discarded.

*Assumption 2:* The receiver/remote estimator sends perfect acknowledgment signals to the transmitter about whether each message is successfully received or not.

Assumption 1 describes our proposed *abstraction for low-latency high-reliability dynamical systems*. As shown in Fig. 2, this abstraction by the length of the message $r$, the latency $d$, and the reliability $1 - p_e$ corresponds to the whole *communication block* including the encoding, channel, and decoding. In this section, we analyze the estimation performance treating these as given parameters. In practice, as we discuss in Section III, these are interrelated parameters as derived from practical coding schemes and cross-layer optimization. For example, keeping the channel reliability $1 - p_e$ fixed, we can increase the quantization level $r$ to get more precise information, but this will increase the delay $d$ of the communication which may adversely affect the estimation performance. Similar tradeoffs arise by tuning any of the parameters. We also note that the assumption that latency is less than the sampling period ($d \leq T$) is added to practically ensure that messages are received before new messages are being generated to be sent over the communication block.

The fact that the receiver knows whether the message is successfully decoded or not (Assumption 1) is an important assumption and is often made in the control literature [2], [8] (see also Remark 2). Assumption 2 is also typical.

To sum up, the state is sampled at time steps $0, T, 2T, \ldots$ and transmission takes $d$ time steps (see also Fig. 3). Similar models are common in the literature [40]. At time steps $d, T + d, 2T + d, \ldots$ the remote estimator builds some estimates $\hat{x}_d, \hat{x}_{T+d}, \hat{x}_{2T+d}, \ldots$ about the corresponding state and we measure performance by the magnitudes of the estimation error $|x_d - \hat{x}_d|, |x_{T+d} - \hat{x}_{T+d}|, |x_{2T+d} - \hat{x}_{2T+d}|, \ldots$. Other performance metrics are also possible.

### A. Quantization Scheme

The way each state sample is quantized to an $r$-bit message is an important choice. Here we employ the scheme of Tatikonda and Mitter [2] which is a sequential uniform quantization scheme. We also point out that other quantization schemes are also available in the literature, e.g., logarithmic quantizers [3], or zoom-in/zoom-out quantizers [19]. These are not further

[2]We note that a different notion of delay is discussed in [7] and [13]. In particular, this has to do with the case where communication is over a noisy channel and it takes a random number of channel uses for the decoder to correctly decode past transmitted messages. This is different than transmission latency in our article.

Authorized licensed use limited to: University of Pennsylvania. Downloaded on October 29,2026 at 15:10:15 UTC from IEEE Xplore. Restrictions apply.

explored in this article but we believe a similar latency-reliability analysis can be performed in these cases.

Our quantization scheme is as follows. At time $k = 0$ the receiver/estimator knows that the state belongs in the initial set $[\underline{X}_0, \bar{X}_0]$. The transmitter uniformly quantizes this range of values into $2^r$ bins and transmits the $r$-bit message indicating the bin in which the measured state $x_0$ belongs. After the transmission duration of $d$ time steps we have two cases.

*Case I:* If the message corresponding to the sample $x_0$ is not successfully received at time $k = d$, the remote estimator only knows the initial information, i.e., that $x_0 \in [\underline{X}_0, \bar{X}_0]$. Then it can form an estimate about $x_0$ as the center value of this interval

$$\hat{x}_0 = \frac{\bar{X}_0 + \underline{X}_0}{2}. \tag{3}$$

However, the estimator is interested in the *current* value of the state $x_d$ at this reception time $k = d$. To obtain an estimate of the current state $x_d$ and counteract the effect of the delay, the estimator can propagate the obtained estimate $\hat{x}_0$ by the system dynamics 1 assuming zero noise to form

$$\hat{x}_d = A^d \hat{x}_0. \tag{4}$$

Moreover, by Assumption 2, the receiver sends an acknowledgment signal to the transmitter to notify that the message was not successfully received. At the next sampling time $k = T$, the receiver only knows that the state $x_T$ due to the model 1 has changed to some value

$$x_T = A^T x_0 + \sum_{m=0}^{T-1} A^{T-1-m} w_m \tag{5}$$

which lies in a set given by

$$\left[ A^T \underline{X}_0 - \frac{A^T - 1}{A - 1} \frac{W}{2}, A^T \bar{X}_0 + \frac{A^T - 1}{A - 1} \frac{W}{2} \right] \tag{6}$$

summing up the magnitude of the unknown values of the system noise $w_0, \ldots, w_{T-1}$. Then the process repeats, i.e., the sensor quantizes uniformly this new set, it sends the bin in which the state $x_T$ belongs, etc.

*Case II:* Alternatively, if the message corresponding to the sample $x_0$ is successfully decoded at time $d$, the remote estimator learns in which of the $2^r$ bins the state $x_0$ belongs, e.g., $x_0 \in [\underline{X}'_0, \bar{X}'_0]$ with length $\bar{X}'_0 - \underline{X}'_0 = (\bar{X}_0 - \underline{X}_0)/2^r$. It can construct an estimate as the center of that interval $\hat{x}_0 = (\underline{X}'_0 + \bar{X}'_0)/2$. As in the previous case the estimator can propagate the initial estimate $\hat{x}_0$ via the system dynamics to obtain a current estimate of the form $\hat{x}_d = A^d \hat{x}_0$. Moreover the receiver sends an acknowledgment signal to the transmitter to notify that the message was successfully received. At the next sampling time $T$ the receiver knows that the state $x_T$ given by (5) lies in a new set

$$\left[ A^T \underline{X}'_0 - \frac{A^T - 1}{A - 1} \frac{W}{2}, A^T \bar{X}'_0 + \frac{A^T - 1}{A - 1} \frac{W}{2} \right] \tag{7}$$

because it knows that $x_0 \in [\underline{X}'_0, \bar{X}'_0]$ – note also that the set 7 in this case is smaller than the set in the opposite case in 6. Then the process repeats. An illustration of this sequential coding scheme is shown in Fig. 4.

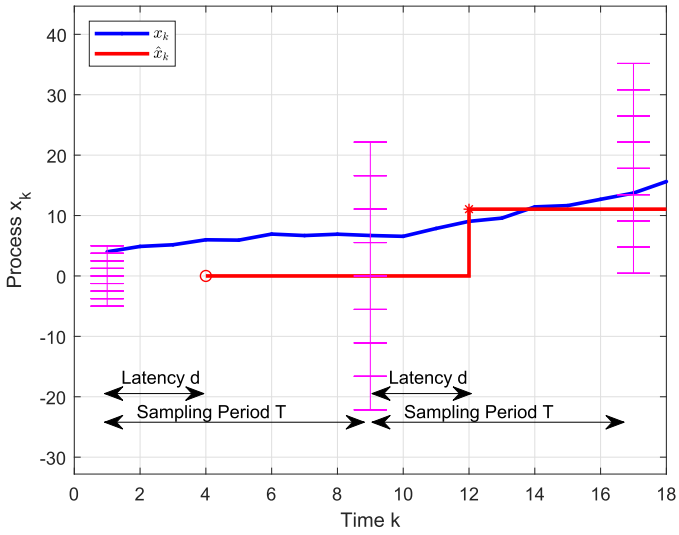



Fig. 4. Process $x_k$ is sampled periodically at time steps $k = 1, 9, 17, \ldots$ and quantized to the 8 ($r = 3$) bins shown. After a latency of $d = 3$ time steps, the receiver either decodes or discards the message. The first message at time $k = 4$ is discarded (shown with an circle), resulting in a larger uncertainty about the state. The second message at time $k = 12$ is correctly decoded (shown with a star) and the estimator obtains a good estimate.

Given the communication uncertainty described in the two cases above, we can derive the performance of the estimator at time $d$ as

$$\mathbb{E}|x_d - \hat{x_d}| \leq \frac{1}{2} \left[ A^d \left( p_e + \frac{1 - p_e}{2^r} \right) (\bar{X}_0 - \underline{X}_0) + \frac{A^d - 1}{A - 1} W \right]. \tag{8}$$

Here, the expectation accounts for the uncertainty in the success of decoding. We observe in particular that the estimation error increases linearly as the decoding error probability increases. On the contrary the estimation error increases exponentially as the latency (delay) increases, hence latency has a detrimental effect on the performance motivating our analysis. We also point out that this bound depends on both the initial uncertainty about the state as well as the level of disturbance.

## B. Steady State Estimation Performance

Since the state changes over time by 1, we are interested in whether the remote estimator can track this process, and furthermore how well it can do so in the long run. Our first result derives conditions on the communication block parameters under which estimation is feasible.

*Theorem 1 (Scalar Stability Condition):* Consider the remote estimation of the scalar dynamical system in 1 over the communication channel and let Assumptions 1 and 2 hold. Suppose we employ the sequential quantization scheme described in 3–7. Then the expected value of the estimation error

Authorized licensed use limited to: University of Pennsylvania. Downloaded on October 29,2026 at 15:10:15 UTC from IEEE Xplore. Restrictions apply.

$|x_{\ell T+d} - \hat{x}_{\ell T+d}|$, $\ell = 0, 1, \ldots$ is bounded if and only if[3]

$$\left(p_e + \frac{1-p_e}{2^r}\right) A^T < 1. \tag{9}$$

*Proof:* As mentioned above for the first message there are two cases depending on the success of decoding the message. In the first case where the transmission is unsuccessful the remote estimator learns that the state $x_0$ belongs in the set $[\underline{X}_0, \bar{X}_0]$ with width $[\bar{X}_0 - \underline{X}_0]$ and constructs the estimate $\hat{x}_0 = (\bar{X}_0 + \underline{X}_0)/2$. In this case at time $k = d$, the transmitter has an estimation error about $x_0$ bounded by

$$|x_0 - \hat{x}_0| \leq \frac{1}{2}(\bar{X}_0 - \underline{X}_0). \tag{10}$$

In the opposite case where the remote estimator receives the message correctly, it learns that the state $x_0$ belongs in a set $x_0 \in [\underline{X}'_0, \bar{X}'_0]$ with length $\bar{X}'_0 - \underline{X}'_0 = (\bar{X}_0 - \underline{X}_0)/2^r$. It constructs an estimate as the center of that interval $\hat{x}_0 = (\underline{X}'_0 + \bar{X}'_0)/2$ with error bounded by

$$|x_0 - \hat{x}_0| \leq \frac{1}{2}(\bar{X}'_0 - \underline{X}'_0) = \frac{1}{2}\frac{\bar{X}_0 - \underline{X}_0}{2^r}. \tag{11}$$

Let us define a random variable $\Delta_0$ as the length of the set concerning $x_0$ at the estimator at time $k = d$. That is

$$\Delta_0 := \begin{cases} \bar{X}_0 - \underline{X}_0 & \text{w.p. } p_e \\ \frac{1}{2^r}(\bar{X}_0 - \underline{X}_0) & \text{w.p. } 1 - p_e. \end{cases} \tag{12}$$

From the inequalities (10) and (11) we also have that the estimation error is in both cases bounded by

$$|x_0 - \hat{x}_0| \leq \frac{1}{2}\Delta_0 \tag{13}$$

where there is an initial condition $x_0$ that makes the inequality tight.

Moreover using this notation, given the success or failure of the first message, at the next sampling time $T$ the state is known to lie in a set of magnitude $A^T\Delta_0 + \frac{A^T-1}{A-1}W$. This is quantized in $2^r$ bins and sent again.

Similarly, let us define $\Delta_T$ as the width of the set where the remote estimator knows that the state $x_T$ belongs depending on the success or failure of the message at time $k = T + d$. This is a new random variable defined as

$$\Delta_T := \begin{cases} A^T\Delta_0 + \frac{A^T-1}{A-1}W & \text{w. prob. } p_e \\ \frac{1}{2^r}(A^T\Delta_0 + \frac{A^T-1}{A-1}W) & \text{w. prob. } 1 - p_e. \end{cases} \tag{14}$$

Again we have that the new estimate satisfies $|x_T - \hat{x}_T| \leq 1/2\Delta_T$, where there is an initial condition $x_0$ and disturbance $w_0$ that makes the inequality tight.

The process repeats so that we have the recursion at the $(\ell + 1)$th transmission

$$\Delta_{(\ell+1)T} := \begin{cases} A^T\Delta_{\ell T} + \frac{A^T-1}{A-1}W & \text{w. prob. } p_e \\ \frac{1}{2^r}(A^T\Delta_{\ell T} + \frac{A^T-1}{A-1}W) & \text{w. prob. } 1 - p_e \end{cases} \tag{15}$$

and moreover we have that

$$|x_{(\ell+1)T} - \hat{x}_{(\ell+1)T}| \leq \frac{1}{2}\Delta_{(\ell+1)T}. \tag{16}$$

Taking expectation with respect to the success or failure of all transmissions at both sides we can bound the expected estimation error by

$$\mathbb{E}|x_{(\ell+1)T} - \hat{x}_{(\ell+1)T}| \leq \frac{1}{2}\mathbb{E}\Delta_{(\ell+1)T} \tag{17}$$

where there is an initial condition $x_0$ and disturbance $w_0, w_1, \ldots$ that makes the inequality tight.

Also taking the expectation in the above recursion we have that

$$\begin{aligned}\mathbb{E}\Delta_{(\ell+1)T} = &\left(p_e + (1-p_e)\frac{1}{2^r}\right) A^T \mathbb{E}\Delta_{\ell T} \\ &+ \left(p_e + (1-p_e)\frac{1}{2^r}\right)\frac{A^T - 1}{A-1}W.\end{aligned} \tag{18}$$

This is a linear system of equations that converges to a finite value if and only if (9) holds. Hence the expected estimation error is bounded if and only if (9) holds. ■

From this theorem we observe that to maintain bounded estimation there is a critical dependence between reliability, quantization level, and system dynamics. Assuming dynamics $A$ and quantization level $r$ fixed, there is a critical threshold on the communication error rate $p_e$ above which communication becomes too noisy and it is impossible to maintain a finite estimation error. Assuming dynamics $A$ and communication error rate $p_e$ fixed, there is a critical threshold on the quantization level $r$ below which information becomes too coarse and it is impossible to maintain a finite estimation error. Related stability results are also known in the literature (see Remark 1).

On the other hand, latency $d$ does not play a role for stability as it does not appear in 9—apart from the practical condition that latency should be smaller than sampling period ($d \leq T$ in Assumption 1). The fact that latency does not affect stability is also mentioned in [7, Sec. IV.F] and [3] based on an input–output channel model where information from the input is always delayed by a constant amount of time at the output. We will show now however that it does play a significant role for the steady state estimation performance.

*Theorem 2 (Scalar Estimation Performance):* Consider the remote estimation of the scalar dynamical system in 1 over the communication channel and let Assumptions 1 and 2 hold. Suppose we employ the sequential quantization scheme described in (3)–(7). If (9) holds, then the bound on the expected steady state error converges to the value[4]

$$\begin{aligned}\limsup_{\ell\to\infty}\ &\mathbb{E}|x_{\ell T+d} - \hat{x}_{\ell T+d}| \\ &\leq \frac{1}{2}\frac{A^d + \left(p_e + \frac{1-p_e}{2^r}\right)(A^T - A^d) - 1}{1 - \left(p_e + \frac{1-p_e}{2^r}\right)A^T}\frac{W}{A-1}.\end{aligned} \tag{19}$$

[3]More precisely, if condition (9) fails, then there exists an initial state and disturbance sequence such that the expected value of the estimation error is unbounded.

[4]For the marginally stable case $A = 1$, the bound becomes $\frac{d+(p_e+(1-p_e)2^{-r})(T-d)}{(1-p_e)(1-2^{-r})}\frac{W}{2}$.

Authorized licensed use limited to: University of Pennsylvania. Downloaded on October 29,2026 at 15:10:15 UTC from IEEE Xplore. Restrictions apply.

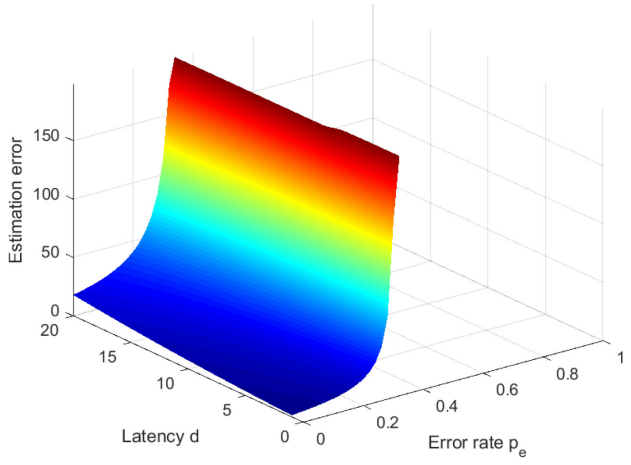



Fig. 5. Steady state estimation performance 19 as a function of reliability and latency. Estimation error becomes unbounded (unstable) when the reliability exceed some threshold, while larger latency adversely affects performance.

*Proof.* In our scheme, at each reception time $\ell T + d$ for $\ell = 0, 1, \ldots$ the estimator constructs an estimate $\hat{x}_{\ell T}$ about the state value $x_{\ell T}$ and then, to counteract the effect of delay, this estimate is propagated through the system dynamics 1 to obtain an estimate about the current state $x_{\ell T+d}$ of the form

$$\hat{x}_{\ell T+d} = A^d \hat{x}_{\ell T}. \tag{20}$$

The current estimation error then is bounded by

$$|x_{\ell T+d} - \hat{x}_{\ell T+d}| \leq A^d |x_{\ell T} - \hat{x}_{\ell T}| + \frac{A^d - 1}{A - 1} \frac{W}{2} \tag{21}$$

accounting for the noise of the system 1 during this delay.

Similar to the proof of Theorem 1, we define define $\Delta_{\ell T}$ as the width of the set where the remote estimator knows that the state $x_{\ell T}$ belongs depending on the success or failure of the message at time $k = \ell T + d$. Moreover, we have shown in (16) that $|x_{\ell T} - \hat{x}_{\ell T}| \leq \Delta_{\ell T}/2$ holds always. Using this fact and taking expectation of 21 we get that

$$\mathbb{E}|x_{\ell T+d} - \hat{x}_{\ell T+d}| \leq A^d \frac{\mathbb{E}\Delta_{\ell T}}{2} + \frac{A^d - 1}{A - 1} \frac{W}{2}. \tag{22}$$

Recall also that $\mathbb{E}\Delta_{\ell T}$ satisfies the recursion in (18). Hence, if condition (9) holds then the value $\mathbb{E}\Delta_{\ell T}$ converges to

$$\lim_{\ell \to \infty} \mathbb{E}\Delta_{\ell T} = \frac{\left(p_e + \frac{1-p_e}{2^r}\right) \frac{A^T - 1}{A - 1} W}{1 - \left(p_e + \frac{1-p_e}{2^r}\right) A^T}. \tag{23}$$

Plugging this limit expression in (22) we obtain the result (19). ■

From (19) it can be seen that, unlike the stability analysis, latency $d$ has a crucial effect on performance (due to the term $A^d$). We plot the bound (19) in Fig. 5 for different values of the latency parameter $d$ and communication reliability $p_e$ and for a fixed quantization rate $r$. As mentioned above in (9) there is a minimum reliability below which the remote estimator cannot track the system and the estimation error grows to infinity. Otherwise the estimation performance improves as the error rate becomes smaller or the latency decreases. We also note that a related tradeoff between data rate and delay and with perfect reliability ($p_e = 0$) appears in [3, Sec. II]. Also [20, Sec. V-A] contains a performance analysis similar to Theorem 2 for the case of unreliable links ($p_e \neq 0$) and finite rate but without latency considerations. Our Theorem 2 provides a performance metric that accounts for rate, latency, and reliability, hence extending the aforementioned results. More importantly this enables us to consider novel tradeoffs in channel coding design as explained in the next section.

Specifically the parameters $r, d, p_e$ of our abstraction are interdependent variables of the employed channel coding scheme. In the following section, we discuss how these dependencies arise from practical coding schemes, and how we can tune over these parameters to obtain the latency-reliability tradeoff that optimizes the steady state system performance.

*Remark 1:* There is an extensive literature on related stability results over communication links, including [8, Prop. 4.2] for the case without disturbances which is a special case of our setting, including [4, Sec. 1.4.3] and [5, Th. 4.1] for the case with stochastic disturbances of potentially unbounded support (such as Gaussians) which is more general than our setting, and including [21] for the packet-drop channel without quantization which can be thought of as a special case of our setting when $r \to \infty$.

We further point out that our necessary and sufficient stability condition is derived for the specific quantization scheme employed. One may wonder if using other quantization (or channel coding) schemes may help relax this condition, e.g, support systems with even larger eigenvalues (faster systems) than in (9). If we ignore latency and consider our communication block abstraction as an input–output channel with input $r$ bits and output that is erased with probability $p_e$, and assuming acknowledgments, then [4, Sec. 1.4.3] uses the notion of anytime capacity introduced in [7] and argues that (9) is in fact necessary for stability under any quantization (or channel coding) scheme. Hence that is a more general result than ours. We emphasize that the main goal of our article is to not to derive new stability results, but instead examine the effects that latency and reliability have in estimation performance as provided in the following section. □

## III. Code Length Selection for Estimation Performance

In this section, we investigate how channel coding can play a role in improving the system performance. That is, we use error correcting codes to alleviate the effect of channel noise on the transmitted messages. While coding leads to a dramatic increase in reliability, it causes an extra penalty in latency as reliability is obtained at the cost of sending longer messages (codewords). We will study latency-reliability tradeoffs obtained from codes and their consequences regarding the system performance. We also discuss cross-layer coding design, i.e., the selection of code parameters that yield the optimal performance taking into account the system dynamics.

We consider error-correction coding systems which encode the $r$-bit information messages into longer $n$-bit codewords.

Authorized licensed use limited to: University of Pennsylvania. Downloaded on October 29,2026 at 15:10:15 UTC from IEEE Xplore. Restrictions apply.

Here we are assuming that the transmission takes place over a channel with binary input [e.g. the binary erasure channel (BEC)]. Clearly, the rate of the code needs to be lower than the channel capacity $\frac{r}{n} < C$.

We attribute the latency $d$ of communicating the $n$-bit message in transmission delay of sending these $n$ bits—in practice there will also be other overheads contributing to delay that we omit from the present analysis. As mentioned in Section II, we specifically model that the transmission of each bit requires a single normalized unit time interval, hence

$$d = n. \tag{24}$$

The reliability (error probability) of the code is denoted by $p_e$, i.e. with probability $1 - p_e$ we can decode the information message correctly (from the noisy outcome of the channel) and with probability $p_e$ the decoding procedure is unsuccessful.

For a given channel and a given error-correcting coding scheme the error probability is a function of the message length $r$ and the code length $n$, that is

$$p_e = p_e(r, n). \tag{25}$$

As mentioned in Assumptions 1 and 2, we further assume that the decoder can *detect* whether or not the decoding procedure has been successful, and will notify the transmitter about it by using a one bit ACK/NACK feedback (see also Remark 2). As a result, in case of decoding failure, the transmitted packet will be discarded. We can thus attribute the channel reliability $1 - p_e$ to the probability of successful decoding of the coded messages.

As a result, using Theorem 2 and substituting our model for the latency $d$ by (24) and the reliability $p_e$ by (25), the expected steady state error of the overall system can be expressed as

$$\lim_{\ell\to\infty} \mathbb{E}|x_{\ell T+d} - \hat{x}_{\ell T+d}| \leq \frac{A^n + \left(p_e(r,n) + \frac{1-p_e(r,n)}{2^r}\right)(A^T - A^n) - 1}{1 - \left(p_e(r,n) + \frac{1-p_e(r,n)}{2^r}\right) A^T} \frac{W}{A-1}. \tag{26}$$

This expression reveals a relation between code parameters $r, n$ and the steady-state error, and this relationship can be understood through the the lens of latency-reliability tradeoffs derived in the previous section.

Next we illustrate these tradeoffs through different choice of coding schemes.

1) *Uncoded transmission:* As a warm-up, we first consider uncoded transmission, i.e., each time we send $r$ (information) bits through the channel uncoded and without using any error-correction mechanism, i.e., $n = r$. In this case, code length/latency becomes trivially equivalent to the message length $d = n = r$.

   As a concrete illustration, we consider the simplest type of channel, the so-called *BEC* with erasure probability $\zeta$ (BEC($\zeta$)). On this channel, a bit is either passed through perfectly with probability $1 - \zeta$ or completely erased with probability $\zeta$. Hence, each $r$-bit message is either successfully received with probability $(1 - \zeta)^r$, or otherwise with the complementary probability at least one bit is erased and the message is discarded. This corresponds to a model of the form (25) with

$$p_e(r, n) = 1 - (1 - \zeta)^r. \tag{27}$$

So error probability increases exponentially in the message length.

With these expressions in place our communication scheme and the corresponding estimation error performance in (26) is parameterized by the single parameter $r$, the quantization level, taking values in $[1, T]$, as there are at most $T$ available time slots to transmit bits until the next sampling time.

We plot in Fig. 6–8 the relationship between the code length/latency $d = n = r$ and the steady state estimation error in (26) for different values of the channel quality $\zeta$ and dynamics $A$. From these figures we can see that longer messages eventually lead to unbounded costs (instability) as the probability of error increases dramatically in these cases. The performance degradation is less dramatic when the channel quality is good (lower erasure probability $\zeta$). For the special case of stable dynamics in Fig. 8, unbounded estimation error cannot occur because even if the probability of error is $p_e = 1$ the estimation error is finite. But the performance quickly saturates to its peak value. Overall for uncoded communication very short messages achieve the best performance.

2) *Coded transmission using optimal finite blocklength codes:* In reality, uncoded communication is never used as it is inefficient, hence we consider error-correction coding systems which encode the $r$-bit information messages into longer $n$-bit codewords contributing to longer latencies through (24). Clearly, the rate of the code $R = r/n$

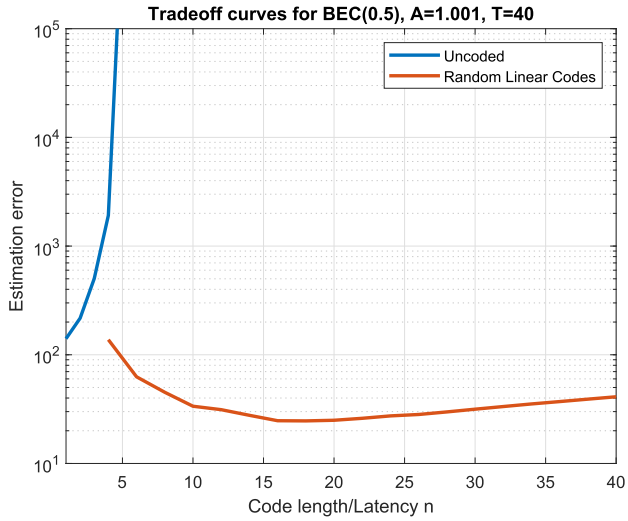



Fig. 6. Estimation performance of different coding schemes over the BEC as a function of codelegth/latency. We observe that there is an optimal codelength that achieves the best latency-reliability tradeoff. We also observe that the random linear codes (which are the optimal in the sense of shortest blocklength) achieve the best performance. Uncoded communication is very inefficient and leads to instability for longer message lengths.

Authorized licensed use limited to: University of Pennsylvania. Downloaded on October 29,2026 at 15:10:15 UTC from IEEE Xplore. Restrictions apply.

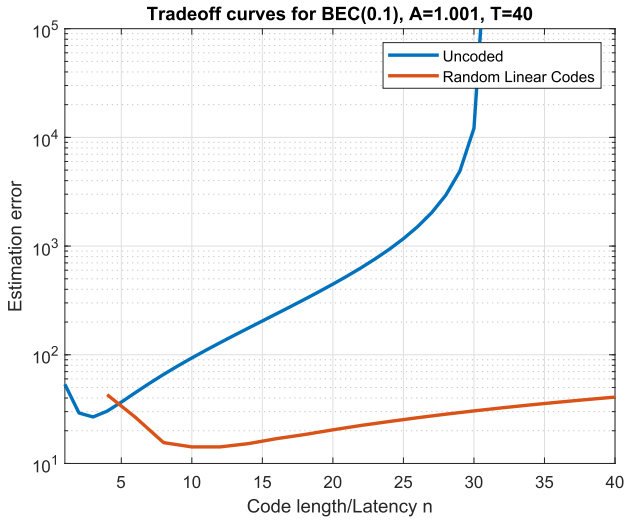



Fig. 7. Estimation performance of different coding schemes over the BEC as a function of codelegth/latency. We observe that there is an optimal codelength that achieves the best latency-reliability tradeoff. We also observe that the random linear codes (which are the optimal in the sense of shortest blocklength) achieve the best performance.

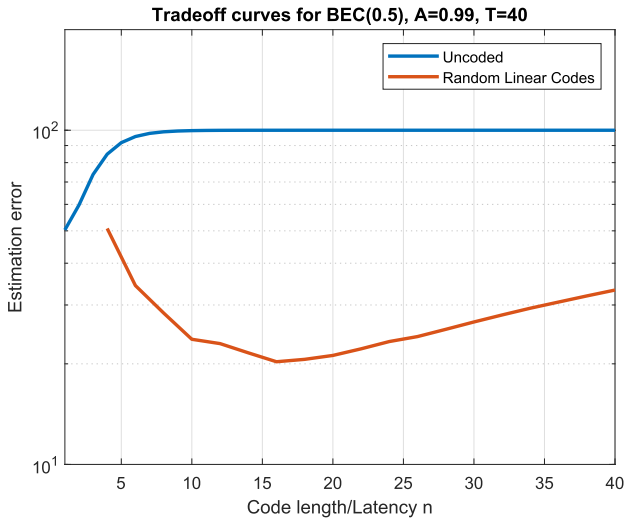



Fig. 8. Estimation performance for a stable system using different coding schemes over the BEC as a function of codelegth/latency. In contrast to previous figures, here performance always remains bounded, but still optimal codes obtain a best performance.

needs to be lower than the channel capacity $\frac{r}{n} < C$. Fundamental information–theoretic laws [56]–[58] state that to reliably communicate over a channel with capacity $C$, the *optimal* (in the sense of shortest possible) blocklength $n$ scales as

$$n \approx \frac{VQ^{-1}(p_e)}{(C-R)^2}, \quad \text{or equivalently:}$$

$$p_e(r,n) = Q\left(\sqrt{\frac{n}{V}}\left(C - \frac{r}{n}\right) + O(\log n)\right) \qquad (28)$$

where $Q(\cdot)$ is the tail probability of the standard normal distribution, and $V$ is a characteristic of the channel referred to as channel dispersion. Unfortunately the only optimal codes [in the sense of 28] we currently know have an exponential decoding complexity in general. Instead, in practice we quest for codes with low-complexity encoder/decoder design (ideally, linear in blocklength). For the error-correcting codes used in practice, one can compute or simulate the *tradeoff* curves showing how the codelength $n$ varies with the rate $R = r/n$ at different values of error probability $p_e$. By plugging such tradeoff curves into (25) and (26) we can obtain tradeoff curves for reliability versus latency.

For the special case of the *BEC* as the transmission medium, we may consider *random linear codes* for error correction. Such codes are constructed by selecting uniformly at random a linear mapping from the space $\{0,1\}^r$ to $\{0,1\}^n$. Such a mapping is specified by a $r \times n$ binary matrix whose entries are chosen i.i.d according to Bernoulli($\frac{1}{2}$). Every such random mapping (matrix) is a code that can be use to encode an $r$-bit information vector to a codeword of size $n$. For the specific case of BEC, random linear codes are known to have the optimal rate-reliability-length as in (28) (see e.g., [64]). In other words, such codes minimize the latency of the communication block of the system in Fig. 2. These codes are decoded in cubic time in terms of the length and thus their complexity is rather high and they are seldom used in practice. Other codes with linear complexity but higher latency are preferred (see for example the following section).

Despite their practical shortcomings, we illustrate the effect of using random linear codes in our estimation setup over a BEC channel. For given system dynamics $A$, sampling period $T$, and channel erasure probability $\zeta$ we consider different code choices given by the quantization level $r$ and codelength $n$. We obtain the error rate of such coding by simulation, and then the resulting estimation error by (26) as a function of the two parameters $r$ and $n$. To obtain a clear relation between estimation error and latency/code length $n$, we optimize over the quantization level $r$ by performing a search for each code length $n$, and the resulting relationship is plotted.

From the figures, we can search for an optimal codelength that minimizes estimation error. Practically an optimal blocklength that minimizes the estimation error should exist, because it is just a choice from a finite set of integer numbers, e.g., $n = 1, 2, \ldots, n_{\max}$. From our numerical plots, we observe the following intuitive behavior. Shorter codes increase estimation error because they have lower reliability, while longer codes have better reliability but estimation performance is worse due to the required longer latency. The gap in performance can be significant, for example in the figures there is more than 50% improvement in estimation quality between the optimal length and the longest length. Finally comparing the optimal (i.e., shortest) codes with the case of uncoded communication we point out that for the considered

Authorized licensed use limited to: University of Pennsylvania. Downloaded on October 29,2026 at 15:10:15 UTC from IEEE Xplore. Restrictions apply.

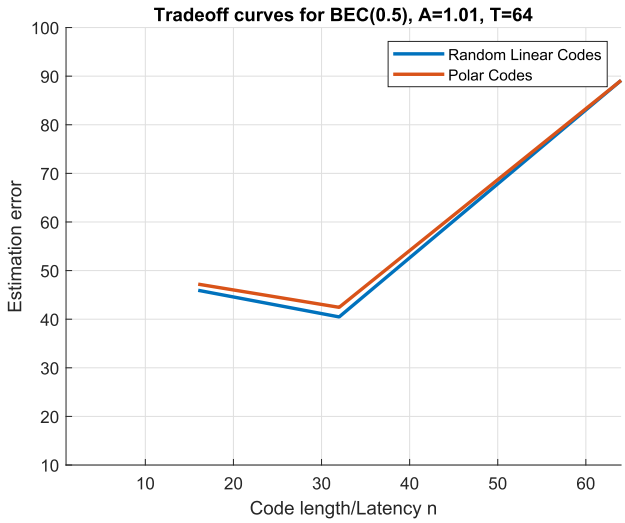



Fig. 9. Estimation performance using polar codes (which are practical in terms of decoding complexity) and random linear codes (which are the optimal in the sense of shortest blocklength). We observe very similar performance among the two.

channel and dynamics values uncoded communication is very inefficient, as expected.

From the plots we also obtain the following practical insights: for faster dynamics (larger $A^T$) and noisy channel conditions (larger probability of erasure $\zeta$) longer codelengths are preferred. This is expected as faster systems require higher reliability and noisy channels require longer codes with more redundancy. On the other hand, for slower dynamics and less noisy channels shorter codelengths are optimal, i.e., there is no performance gain from longer codes. In other cases the codelengths should be neither too short nor too long.

3) *Coded transmission using polar codes:* Beyond the theoretically optimal coding schemes we may also consider practical codes. In particular, we consider polar codes [65]. Polar codes achieve the capacity of BEC in the asymptotic limit of length $n$. They can be efficiently implemented, i.e., the encoding and decoding complexity of polar codes is $O(n \log n)$ for any channel. In comparison, random codes also achieve the capacity but their decoding complexity is cubic in $n$ for the BEC (and for other channels that there is no polynomial time algorithm to decode random codes). Both polar and random codes are capable of erasure detection when used over BEC. For error detection, we should add cyclic redundancy check (CRC) bits as overhead to the codeword.

We compare random linear codes (which are the optimal in the sense of shortest blocklength) with polar codes in Fig. 9. As in the previous figures we plot the estimation performance as a function of the codelength $n$ after optimizing over the message quantization length $r$. Even though polar codes are not optimal we observe a very similar performance among the two. That is for the codelength values considered the reliability of both coding schemes is within the same order of magnitude.

*Remark 2:* We note that when transmission is over the BEC channel, as in the numerical analysis of this section, almost all code designs in practice are capable of detecting decoding failures [66]. For other channels, even though our assumption that decoding errors can be detected is idealized, it is justified because the task of detecting decoding failures is typically facilitated by the use of CRC codes. Adding a CRC code on top of e.g., the error-correcting channel codes discussed in Section III, can drive the probability of incorrect decoding to very small values. Alternatively the case where the communication block may corrupt the transmitted message so that the receiver may incorrectly decode a message is a challenging problem in estimation and control (see for example recent work [13] and is not further explored in this article). □

### A. Theoretically Optimal Blocklength Analysis

In principle, we can plug in expression (28) in (26) to find the optimal blocklength $n$ that minimizes the steady state error for the optimal (shortest) codes. However, the resulting equation does not attain a simple closed form solution. Nevertheless, the equations can be solved numerically by search and the optimal blocklength $n$ can be found in terms of $r$ and the parameters of the dynamical system and the channel. Alternatively, we need to resort to approximate solutions.

Let us now provide a heuristic argument to approximate the optimal blocklength $n$ in the regime where the value of $A$ is close to 1, e.g., $A = 1.001$ or smaller. Let us also assume the code rate $R = r/n$ is fixed to some value and independent of $n$, so that the error probability in (28) is just a function of blocklength $n$ as $p_e(n)$. Denote $\theta(n) := p_e(n) + \frac{1-p_e(n)}{2^r}$. Then taking the derivative of (26) with respect to $n$ and equaling to zero, we obtain that the optimal codelength satisfies

$$\log(A)(1-\theta(n))(1-\theta(n)A^T) + \frac{d\theta}{dn}(A^T - 1) = 0 \quad (29)$$

where $\log$ is the natural logarithm. Suppose $\theta(n)$ is small enough so that we may approximate $(1-\theta(n))(1-\theta(n)A^T) \approx 1 - \theta(n)(A^T+1)$. Further suppose $A \approx 1 + a$ for some small $a > 0$ so that $\log(A) \approx a$ and $A^T \approx 1 + aT$. Using these approximations in (29) we get

$$1 - \theta(n)(2 + aT) + \frac{d\theta}{dn}T \approx 0. \quad (30)$$

Moreover, let us approximate $\theta(n) \approx p_e(n)$ which is reasonable for a large number of information bits $r$. From (28) let us approximate the normal tail as $Q(x) \approx \phi(x)/x$ where $\phi(x)$ is the standard normal density function. We also have that $\frac{dQ(x)}{dx} = -\phi(x)$. Finally, since we fixed the code rate $R = r/n$ to some value, we obtain from (30) that the optimal code length approximately satisfies

$$1 - \frac{\phi(\sqrt{\frac{n}{V}}(C-R))}{\sqrt{\frac{n}{V}}(C-R)}(2+aT) - \phi\left(\sqrt{\frac{n}{V}}(C-R)\right)\frac{C-R}{2\sqrt{Vn}}T \approx 0 \quad (31)$$

Authorized licensed use limited to: University of Pennsylvania. Downloaded on October 29,2026 at 15:10:15 UTC from IEEE Xplore. Restrictions apply.

or equivalently

$$\frac{\phi(\sqrt{\frac{n}{V}}(C-R))}{\sqrt{\frac{n}{V}}(C-R)} \approx \left[2 + aT + \frac{(C-R)^2}{2V}T\right]^{-1}. \quad (32)$$

This equation does not have a closed form solution with respect to $n$, but is straightforward to solve numerically, instead of searching over the set of all possible codelength choices. From this equation, we see that both increasing the system eigenvalue $A = 1 + a$ (faster systems) has the effect of increasing the optimal codelength $n$. That is because the left-hand side of (32) is an increasing function of $n$. Moreover, decreasing the sampling period $T$ for fixed system dynamics (sampling faster) has the effect of decreasing the codelength. However the effect of the sampling period is more significant at code rates further from the channel capacity ($R << C$).

## IV. Latency-Reliability Tradeoffs for Higher Dimensional Systems

In this section, we consider higher dimensional systems. Our goal is to fix a quantization scheme and develop a methodology to analyze the rate-latency-reliability tradeoff in state estimation. In particular, this also includes the case where a system may have both unstable and stable modes.

Consider the system

$$x_{k+1} = Ax_k + w_k \quad (33)$$

where $x_k \in \mathbb{R}^n$, and we denote each element as $x_k(i), i = 1, \ldots, n$, and $A \in \mathbb{R}^{n\times n}$ and $w_k$ is bounded noise with $w_k(i) \in [-W(i)/2, W(i)/2]$.

Tatikonda and Mitter [2] propose a scheme that goes as follows. We can perform a similarity transform to obtain a system state evolution in the normal Jordan form. In the new coordinate system, the receiver/estimator keeps a hypercube where the state belongs. The sensor/transmitter also keeps track of this hypercube and quantizes uniformly each element of this hypercube with a different number of bits per dimension. At the next sampling time, the sensor updates each coordinate of the hypercube to obtain a new hypercube where the new state sample is guaranteed to belong. The estimation error eventually accounts for the quantization performed along each dimension and along all Jordan blocks.

We explain this scheme in more detail. For simplicity of exposition, we first illustrate the case for systems with Jordan blocks corresponding to only real eigenvalues. Then we illustrate the case for Jordan blocks with complex eigenvalues. To obtain the estimation error of a system with Jordan blocks of both real and complex eigenvalues, it just requires adding up the estimation error terms from separate blocks.

### A. Systems With Only Real Eigenvalues

Suppose that the system matrix $A$ has only real eigenvalues. As the sensor in our setup is sampling the system every $T$ time steps this corresponds to system dynamics

$$x_T = A^T x_0 + \sum_{\ell=0}^{T-1} A^{T-1-\ell} w_\ell. \quad (34)$$

Specifically, let $A^T = \Phi^{-1} J \Phi$ be a Jordan decomposition. At sampling time $T$, the sensor observes the current state $x_T$ and applies the transform $z_T = \Phi x_T$. The sensor needs to quantize the transformed state. The receiver/estimator knows that the state belongs in some $n$-dimensional cube. The transmitter uniformly quantizes each dimension $i$ of this set into $2^{r_i}$ bins and transmits the $r_i$-bit message indicating the bin in which the measured state $x_T(i)$ belongs for all dimensions $i$. In our article, we take these values $r_i$ fixed and they satisfy $\sum_{i=1}^n r_i = r$, but it is possible to also optimize over how we allocate them. After the latency of $d$ time steps, given the success or fail of the transmission, the receiver builds a state estimate $\hat{z}_T$ of the state $z_T$, applies the inverse transform $\hat{x}_T = \Phi^{-1}\hat{z}_T$ to obtain an estimate about the past state $x_T$, and propagates the dynamics to obtain an estimate $\hat{x}_{T+d} = A^d \hat{x}_T = A^d \Phi^{-1} \hat{z}_T$ to counteract the effect of latency as in Section II. Next, we present the quantization method and the evolution of the estimation error with respect to the transformed state $z_T$, and at the end we discuss the performance with respect to the original system state $x_T$.

Consider the situation at time step $k = T$. The sensor measures $z_T$. At the same time, the remote estimator knows that the *previous* sample $z_0$ lies in a hypercube, described by $z_0(i) \in [\underline{Z}_0(i), \bar{Z}_0(i)]$ for each dimension $i = 1, \ldots, n$. Similar to the proof of Theorem 1, let us define the random variable $\Delta_0(i) = \bar{Z}_0(i) - \underline{Z}_0(i)$ which denotes the width along each dimension.

At the sampling time $k = T$ and before the transmission takes place the receiver only knows that the state $z_T$ has changed due to the model (34) to some value

$$z_T = Jz_0 + \sum_{m=0}^{T-1} \Phi A^{T-1-m} w_m \quad (35)$$

where we employed the Jordan decomposition. Consider the $i$th element of this vector and suppose that it corresponds to some (real by assumption) eigenvalue $\lambda$. It takes the value

$$z_T(i) = \lambda z_0(i) + J_{i,i+1} z_0(i+1) + \sum_{m=0}^{T-1} \sum_{j=1}^{n} (\Phi A^{T-1-m})_{ij} w_m(j) \quad (36)$$

where we used the fact that $J$ is in the Jordan normal form and the element $J_{i,i+1}$ of this matrix is nonnegative because it takes the value either 0 or 1. Given the previous information about $z_0(i) \in [\underline{Z}_0(i), \bar{Z}_0(i)]$ then, the remote estimator, before transmission, knows that the state $z_T$ lies in a hypercube where the $i$th element of this vector is upper bounded by

$$z_T(i) \leq \max\{\lambda \bar{Z}_0(i), \lambda \underline{Z}_0(i)\} + J_{i,i+1} \bar{Z}_0(i+1) \quad (37)$$

$$+ \sum_{m=0}^{T-1} \sum_{j=1}^{n} |(\Phi A^m)_{ij}| W(j)/2 \quad (38)$$

Authorized licensed use limited to: University of Pennsylvania. Downloaded on October 29,2026 at 15:10:15 UTC from IEEE Xplore. Restrictions apply.

where we employed the two extreme cases for $z_0(i)$ and we sum up the magnitude of the unknown values of the system noise $w_0, \ldots, w_{T-1}$. Similarly, it is bounded as follows by:

$$z_T(i) \geq \min\{\lambda \bar{Z}_0(i), \lambda \underline{Z}_0(i)\} + J_{i,i+1}\underline{Z}_0(i+1) - \sum_{m=0}^{T-1}\sum_{j=1}^{n} |(\Phi A^m)_{ij}| W(j)/2. \quad (39)$$

Combining the upper and lower bounds we have that $z_T(i)$ lies in an interval of length

$$|\lambda|(\bar{Z}_0(i) - \underline{Z}_0(i)) + J_{i,i+1}(\bar{Z}_0(i+1) - \underline{Z}_0(i+1)) + \sum_{m=0}^{T-1}\sum_{j=1}^{n} |(\Phi A^m)_{ij}| W(j). \quad (40)$$

Note that the absolute value of the eigenvalue has appeared. Using the above introduced notation this length is equal to

$$|\lambda|\Delta_0(i) + J_{i,i+1}\Delta_0(i+1) + \sum_{m=0}^{T-1}\sum_{j=1}^{n} |(\Phi A^m)_{ij}| W(j). \quad (41)$$

We can further express these interval lengths along all dimensions $i = 1, \ldots, n$ in a vector form

$$|J|\Delta_0 + \sum_{m=0}^{T-1} |\Phi A^m| W \quad (42)$$

where by $|M|$ we denote a matrix whose elements are the absolute values of the elements of the matrix $M$.

The sensor quantizes uniformly each dimension $i$ of this new hypercube with a selected number of bits $r_i$, and it sends the bin in which the state belongs.

*Case I:* With probability $p_e$, the message corresponding to the sample $z_T$ is not successfully received at time $k = T + d$, the remote estimator only knows the original information before transmission, i.e., that $z_T$ lies in the above hypercube with widths given by $|J|\Delta_0 + \sum_{m=0}^{T-1} |\Phi A^m| W$. The estimator selects the center of this hypercube as the estimate $\hat{z}_T$.

*Case II:* Alternatively, if the message corresponding to the sample $z_T$ is successfully decoded at time $T + d$, the remote estimator learns in which of the $2^{r_1+\ldots+r_n}$ bins the state $z_T$ belongs, it can construct an estimate as the center of that interval. In that case the estimator knows that the state lies in a hypercube with smaller widths (as compared to Case I) given by $\text{diag}\{2^{-r_i}\}(|J|\Delta_0 + \sum_{m=0}^{T-1} |\Phi A^m| W)$.

Let us define $\Delta_T$ as the width of the set where the remote estimator knows that the state $z_T$ belongs depending on the success or failure of the message at time $k = T + d$. This is a new random variable defined as

$$\Delta_T := \begin{cases} |J|\Delta_0 + \sum_{m=0}^{T-1} |\Phi A^m| W & \text{w. p. } p_e \\ \text{diag}\{2^{-r_i}\}(|J|\Delta_0 + \sum_{m=0}^{T-1} |\Phi A^m| W) & \text{w. p. } 1 - p_e \end{cases} \quad (43)$$

depending on the Cases I or II above. By Assumption 2, the receiver sends an acknowledgment signal to the transmitter to notify whether the message was successfully received or not. Then the process repeats.

At a general sampling time $\ell T$, we have that the width of the expected hypercube follows the general recursion

$$\mathbb{E}\Delta_{(\ell+1)T} = \left(p_e I + (1 - p_e)\text{diag}\left\{2^{-r_i}\right\}\right) \left(|J|\mathbb{E}\Delta_{\ell T} + \sum_{\ell=0}^{T-1} |\Phi A^\ell| W\right) \quad (44)$$

which follows by taking the expectation in (43).

Since $|J|$ is an upper triangular matrix (due to the Jordan form), its eigenvalues are the diagonal elements, which correspond to the absolute values $|\lambda_i(A^T)|$ of the eigenvalues of matrix $A^T$ which are the same as $|\lambda_i(A)|^T$. It follows that the above recursion (44) converges to a finite value if and only if the number of bits per dimension and the reliability satisfy $(p_e + (1 - p_e)\frac{1}{2^{r_i}})|\lambda_i(A)|^T < 1$ for all $i$. This is a generalization of Theorem 1. Similar results are shown in [2, Prop. 5.1]. Let $\Delta^*$ denote the vector that is the unique steady state solution of the above recursion (44), i.e.,

$$\Delta^* = \left(p_e I + (1 - p_e)\text{diag}\{2^{-r_i}\}\right) \left(|J|\Delta^* + \sum_{\ell=0}^{T-1} |\Phi A^\ell| W\right). \quad (45)$$

Let us now return to the estimation error. We have by construction that for all sampling times and for all dimensions $i = 1, \ldots, n$ that

$$|z_{\ell T}(i) - \hat{z}_{\ell T}(i)| \leq \frac{1}{2}\Delta_{\ell T}(i). \quad (46)$$

We can then measure the estimation error with respect to the original system dynamics as well as with respect to the current state[5] by

$$\|x_{\ell T+d} - \hat{x}_{\ell T+d}\| = \| A^d\Phi^{-1}(z_{\ell T} - \hat{z}_{\ell T}) + \sum_{m=0}^{d-1} A^{d-1-m} w_{\ell T+m} \| \quad (47)$$

where we used the fact that $\hat{x}_{\ell T+d} = A^d\Phi^{-1}\hat{z}_{\ell T}$ as explained in the beginning of this section. We can bound this estimation error as

$$\begin{aligned} &\|x_{\ell T+d} - \hat{x}_{\ell T+d}\| \\ &\leq \|A^d\Phi^{-1}\|\|z_{\ell T+d} - \hat{z}_{\ell T+d}\| + \sum_{m=0}^{d-1} \|A^{d-1-m}\|\|w_{\ell T+m}\| \\ &\leq \frac{1}{2}\|A^d\Phi^{-1}\| \sum_{i=1}^{n} \Delta_{\ell T+d}(i) + \sum_{m=0}^{d-1} \|A^m\| \sum_{i=1}^{n} \frac{1}{2} W(i) \end{aligned} \quad (48)$$

where in the last inequality we used (46). Taking the expectation and the limit in the above expression we obtain the following result.

[5] In this section, we consider the $\ell_1$ norm: $\|x - \hat{x}\|_1 = \sum_{i=1}^{n} |x(i) - \hat{x}(i)|$.

Authorized licensed use limited to: University of Pennsylvania. Downloaded on October 29,2026 at 15:10:15 UTC from IEEE Xplore. Restrictions apply.

*Theorem 3 (Vector Estimation Performance):* Consider the remote estimation of the dynamical system in (33) over the communication channel and let Assumptions 1 and 2 hold. Suppose we employ the sequential quantization scheme described in this section. Let $A^T = \Phi^{-1} J \Phi$ be a Jordan decomposition and suppose the system has only real eigenvalues. Moreover suppose

$$(p_e + (1-p_e)2^{-r_i})|\lambda_i(A)|^T < 1 \tag{49}$$

holds for all $i = 1, \dots, n$. Then the expected steady state estimation error is bounded by

$$\limsup_{\ell\to\infty} \mathbb{E}\|x_{\ell T+d} - \hat{x}_{\ell T+d}\| \leq \frac{1}{2}\|A^d\Phi^{-1}\| \sum_{i=1}^{n} \Delta^*(i) + \sum_{m=0}^{d-1} \|A^m\| \sum_{i=1}^{n} \frac{1}{2} W(i) \tag{50}$$

where $\Delta^*$ is the solution to the linear equation (45).

To summarize, we can solve for the recursion steady state solution (45), plug in (50), and obtain an expression on the estimation cost. This is a generalization of Theorem 2 which is in fact recovered in the scalar dynamics case. Again we see that rate $r$, reliability $p_e$ and latency $d$ have a significant effect on performance.

In Section IV-C, we analyze numerically the above rate-latency-reliability tradeoff.

### B. Systems With Only Complex Eigenvalues

We employ the quantization scheme of [2] for systems with complex eigenvalues. This setup ends up in a slightly different coordinate state transformation and a slightly different recursion than (44) and hence a different estimation error bound in (50). The stability condition (49) remains the same. Again we can decompose $A^T$ into Jordan normal form $A^T = \Phi^{-1} J \Phi$ to obtain the matrices $\Phi, J$. For clarity of exposition, we present the exact derivation for a single $4x4$ Jordan block matrix $J$, while extension to larger blocks follows in a similar fashion, with just larger matrices of similar structure. Suppose

$$J = \begin{bmatrix} \rho\cos(\theta) & \rho\sin(\theta) & 1 & 0 \\ -\rho\sin(\theta) & \rho\cos(\theta) & 0 & 1 \\ 0 & 0 & \rho\cos(\theta) & \rho\sin(\theta) \\ 0 & 0 & -\rho\sin(\theta) & \rho\cos(\theta) \end{bmatrix}. \tag{51}$$

Define the matrix associated with the above system

$$H = \begin{bmatrix} \cos(\theta) & -\sin(\theta) & 0 & 0 \\ \sin(\theta) & \cos(\theta) & 0 & 0 \\ 0 & 0 & \cos(\theta) & -\sin(\theta) \\ 0 & 0 & \sin(\theta) & \cos(\theta) \end{bmatrix}. \tag{52}$$

This matrix has the property that $HJH^{-1} = J$. This matrix will be used as follows. At the $\ell$th sample, i.e., when the sensor measures the state $x_{\ell T}$, the sensor applies the transform $z_{\ell T} = H^\ell \Phi x_{\ell T}$ on the measurement. Note that this is a time-varying but still invertible transformation, which is the main difference to the time invariant transformation explained for systems with real eigenvalues. The sensor then quantizes uniformly each element $i$ of the transformed vector into $r_i$ bits according to each dimension of a hypercube, and transmits all the indices. If a transmission is successful the receiver builds a state estimate $\hat{z}_{\ell T}$ of the state $z_{\ell T}$, applies the inverse time-varying transform $\hat{x}_{\ell T} = \Phi^{-1} H^{-\ell} \hat{z}_{\ell T}$ to obtain an estimate about the past state $x_{\ell T}$, and propagates the dynamics to obtain an estimate $\hat{x}_{\ell T+d} = A^d \hat{x}_{\ell T} = A^d \Phi^{-1} H^{-\ell} \hat{z}_{\ell T}$ to counteract the effect of latency as in Section II.

Using an analysis similar to the one for the case with real eigenvalues in 44, it can be shown that the width along all dimensions of the hypercube where the transformed states $z_{\ell T}$ take values evolves according to

$$\mathbb{E}\Delta_{(l+1)T} = \left(p_e I + (1-p_e)\text{diag}\{2^{-r_i}\}\right) \left(\tilde{J}\,\mathbb{E}\Delta_{\ell T} + \sum_{m=0}^{T-1} |H^{\ell+1}\Phi A^m| W\right) \tag{53}$$

where $\tilde{J}$ is a matrix that "upper bounds" the growth of the product $HJ$ and is defined by

$$\tilde{J} = \begin{bmatrix} \rho & 0 & 1 & 1 \\ 0 & \rho & 1 & 1 \\ 0 & 0 & \rho & 0 \\ 0 & 0 & 0 & \rho \end{bmatrix}. \tag{54}$$

This last fact is also shown in [2, Lemma 4.2]. This recursion remains stable if $(p_e + (1-p_e)2^{-r_i})\rho < 1$. Note that the last term in the right-hand side of this recursion is time-varying. Nevertheless, each element of the matrix $H^{\ell+1}$ takes values either $\cos((\ell+1)\theta)$ or $\sin((\ell+1)\theta)$ or 0 and hence can be upper bounded by 1 or 0. Hence, we can upper bound the steady state expected width by the solution to the linear equation

$$\Delta^* = \left(p_e I + (1-p_e)\text{diag}\{2^{-r_i}\}\right) \left(\tilde{J}\Delta^* + \sum\nolimits_{m=0}^{T-1} \tilde{H}|\Phi A^m| W\right) \tag{55}$$

where the matrix $\tilde{H}$ is defined as

$$\tilde{H} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix}. \tag{56}$$

Finally, we can bound the expected estimation error similar to (48) as

$$\begin{aligned} &\limsup_{\ell\to\infty} \mathbb{E}\|x_{\ell T+d} - \hat{x}_{\ell T+d}\| \\ &\leq \limsup_{\ell\to\infty} \|A^d\Phi^{-1}\|\,\|H^{-\ell}\|\,\mathbb{E}\|z_{\ell T+d} - \hat{z}_{\ell T+d}\| \\ &\quad + \sum_{m=0}^{d-1} \|A^m\| \sum_{i=1}^{n} \frac{1}{2} W(i) \\ &\leq \frac{\sqrt{2}}{2}\|A^d\Phi^{-1}\| \sum_{i=1}^{n} \Delta^*(i) + \sum_{m=0}^{d-1} \|A^m\| \sum_{i=1}^{n} \frac{1}{2} W(i) \end{aligned} \tag{57}$$

where in the last inequality we used the fact that $\|H^{-\ell}\| = |\cos(-\ell\theta)| + |\sin(-\ell\theta)| \leq \sqrt{2}$ for all $\ell$, and $\Delta^*$ is given by the solution to 55. The derivation for Jordan blocks of larger

Authorized licensed use limited to: University of Pennsylvania. Downloaded on October 29,2026 at 15:10:15 UTC from IEEE Xplore. Restrictions apply.

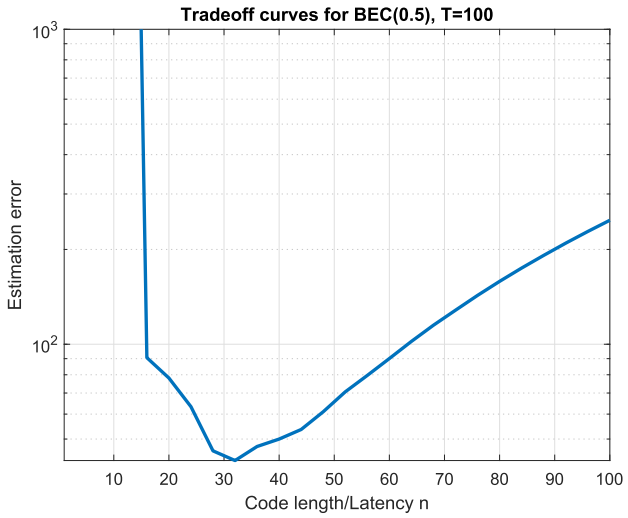



Fig. 10. Estimation performance using random linear codes for a higher dimensional system over the BEC with erasure probability 0.5. We observe that there is an optimal codelength that achieves the best latency-reliability tradeoff.

size follows similarly, i.e, by defining matrices $H$, $\tilde{J}$, $\tilde{H}$ of similar structure, solving for $\Delta^*$ in (55), and upper bounding the estimation error as in (57).

### C. Finite Blocklength Selection for Higher Dimensional Systems

Consider the system with double integrator dynamics $A=\begin{bmatrix}1 & 0.1\\ 0 & 1\end{bmatrix}$ and driven by noise in the second state with magnitude $W=\begin{bmatrix}0\\ 1\end{bmatrix}$. This corresponds for example to tracking the position of an object subject to random acceleration. Suppose we use different quantization levels $r$ and for simplicity we allocate an equal number of bits to quantize each dimension of this system. We consider again the BEC with parameter $\zeta$ and random linear codes for this channel as explained in Section III. In Fig. 10, we plot the estimation performance for the first state as a function of the code length. This is computed using the expression as in (50). Note that in this example even though the first state is not subject to noise there is still the effect of noise carried over from the second state. We observe quantitatively similar tradeoffs between shorter or longer codes as in the scalar system case of Section III. Longer codes have better reliability but they increase latency which in turn affects estimation, hence they should be avoided.

## V. Conclusion

We consider a new communication abstraction motivated by the recent interest for low-latency high-reliability applications in the IoT. More specifically we examine the tradeoffs between latency and reliability for the problem of estimating dynamical systems over communication channels. We couple this approach with different latency-reliability curves derived from practical coding schemes. Our methodology enables a novel cross-layer design, i.e., select the appropriate codelength depending on the system dynamics to optimize estimation performance.

A number of open research questions arise in characterizing latency-reliability tradeoffs for estimation of dynamical systems perturbed by stochastic potentially unbounded noise, as well as more general classes of channels. These problems may also require the consideration of alternative quantization and coding schemes. Future work is also focused on extending the design methodology from estimation to low-latency high-reliability control of fast dynamical systems.

## References

[1] M. Agiwal, A. Roy, and N. Saxena, "Next generation 5G wireless networks: A comprehensive survey," *IEEE Commun. Surv. Tut.*, vol. 18, no. 3, pp. 1617–1655, thirdquarter 2016.
[2] S. Tatikonda and S. Mitter, "Control under communication constraints," *IEEE Trans. Autom. Control*, vol. 49, no. 7, pp. 1056–1068, Jul. 2004.
[3] G. Nair, F. Fagnani, S. Zampieri, and R. Evans, "Feedback control under data rate constraints: An overview," *Proc. IEEE*, vol. 95, no. 1, pp. 108–137, Jan. 2007.
[4] M. Franceschetti and P. Minero, "Elements of information theory for networked control systems," in *Information and Control in Networks*. Cham, Switzerland: Springer, 2014, pp. 3–37.
[5] P. Minero, M. Franceschetti, S. Dey, and G. N. Nair, "Data rate theorem for stabilization over time-varying feedback channels," *IEEE Trans. Autom. Control*, vol. 54, no. 2, pp. 243–255, Feb. 2009.
[6] D. Liberzon and S. Mitra, "Entropy and minimal bit rates for state estimation and model detection," *IEEE Trans. Autom. Control*, vol. 63, no. 10, pp. 3330–3344, Oct. 2018.
[7] 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 Trans. Inf. Theory*, vol. 52, no. 8, pp. 3369–3395, Aug. 2006.
[8] S. Tatikonda and S. Mitter, "Control over noisy channels," *IEEE Trans. Autom. Control*, vol. 49, no. 7, pp. 1196–1201, Jul. 2004.
[9] M. Barforooshan, J. Østergaard, and M. S. Derpich, "Interplay between transmission delay, average data rate, and performance in output feedback control over digital communication channels," in *Proc. Amer. Control Conf.*, 2017, pp. 1691–1696.
[10] R. W. Brockett and D. Liberzon, "Quantized feedback stabilization of linear systems," *IEEE Trans. Autom. Control*, vol. 45, no. 7, pp. 1279–1289, Jul. 2000.
[11] S. Yuksel and T. Başar, "Minimum rate coding for LTI systems over noiseless channels," *IEEE Trans. Autom. Control*, vol. 51, no. 12, pp. 1878–1887, Dec. 2006.
[12] R. Ostrovsky, Y. Rabani, and L. J. Schulman, "Error-correcting codes for automatic control," *IEEE Trans. Inf. Theory*, vol. 55, no. 7, pp. 2931–2941, Jul. 2009.
[13] R. T. Sukhavasi and B. Hassibi, "Linear time-invariant anytime codes for control over noisy channels," *IEEE Trans. Autom. Control*, vol. 61, no. 12, pp. 3826–3841, Dec. 2016.
[14] V. Kostina and B. Hassibi, "Rate-cost tradeoffs in control," in *Proc. 54th Annu. Allerton Conf. Commun., Control, and Comput.*, 2019, pp. 1157–1164.
[15] A. Khina, Y. Nakahira, Y. Su, and B. Hassibi, "Algorithms for optimal control with fixed-rate feedback," in *Proc. IEEE Conf. Decis. Control*, 2017, pp. 6015–6020.
[16] P. Varaiya and J. Walrand, "Causal coding and control for Markov chains," *Syst. Control Lett.*, vol. 3, no. 4, pp. 189–192, 1983.
[17] V. S. Borkar and S. K. Mitter, *LQG Control With Communication Constraints*. Berlin, Germany: Springer, 1997.
[18] A. Khina, E. R. Garding, G. M. Pettersson, V. Kostina, and B. Hassibi, "Control over Gaussian channels with and without source-channel separation," *IEEE Trans. Autom. Control*, vol. 64, no. 9, pp. 3690–3705, Sep. 2019.
[19] V. Kostina, Y. Peres, G. Ranade, and M. Sellke, "Exact minimum number of bits to stabilize a linear system," in *Proc. IEEE Conf. Decis. Control*, 2018, pp. 453–458.

Authorized licensed use limited to: University of Pennsylvania. Downloaded on October 29,2026 at 15:10:15 UTC from IEEE Xplore. Restrictions apply.

[20] A. Khina, V. Kostina, A. Khisti, and B. Hassibi, "Tracking and control of Gauss-Markov processes over packet-drop channels with acknowledgments," *IEEE Trans. Control Netw. Syst.*, vol. 6, no. 2, pp. 549–560, Jun. 2019.

[21] B. Sinopoli, L. Schenato, M. Franceschetti, K. Poolla, M. Jordan, and S. Sastry, "Kalman filtering with intermittent observations," *IEEE Trans. Autom. Control*, vol. 49, no. 9, pp. 1453–1464, Sep. 2004.

[22] O. C. Imer, S. Yüksel, and T. Başar, "Optimal control of LTI systems over unreliable communication links," *Automatica*, vol. 42, no. 9, pp. 1429–1439, 2006.

[23] J. Hespanha, P. Naghshtabrizi, and Y. Xu, "A survey of recent results in networked control systems," *Proc. IEEE*, vol. 95, no. 1, pp. 138–162, Jan. 2007.

[24] V. Gupta, B. Hassibi, and R. Murray, "Optimal LQG control across packet-dropping links," *Syst. Control Lett.*, vol. 56, no. 6, pp. 439–446, 2007.

[25] G. C. Walsh, H. Ye, and L. G. Bushnell, "Stability analysis of networked control systems," *IEEE Trans. Control Syst. Technol.*, vol. 10, no. 3, pp. 438–446, May 2002.

[26] S. C. Smith and P. Seiler, "Estimation with lossy measurements: Jump estimators for jump systems," *IEEE Trans. Autom. Control*, vol. 48, no. 12, pp. 2163–2171, Dec. 2003.

[27] L. Schenato, B. Sinopoli, M. Franceschetti, K. Poolla, and S. Sastry, "Foundations of control and estimation over lossy networks," *Proc. IEEE*, vol. 95, no. 1, pp. 163–187, Jan. 2007.

[28] W. Heemels, K. H. Johansson, and P. Tabuada, "An introduction to event-triggered and self-triggered control," in *Proc. 51st Annu. Conf. Decis. Control*, 2012, pp. 3270–3285.

[29] Y. Xu and J. Hespanha, "Optimal communication logics in networked control systems," in *Proc. 43rd IEEE Conf. Decis. Control*, 2004, vol. 4, pp. 3527–3532.

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

[31] O. C. Imer and T. Başar, "Optimal estimation with limited measurements," *Int. J. Syst., Control Commun.*, vol. 2, no. 1/3, pp. 5–29, 2010.

[32] A. Molin and S. Hirche, "On LQG joint optimal scheduling and control under communication constraints," in *Proc. 48th IEEE Conf. Decis. Control*, 2009, pp. 5832–5838.

[33] A. Nayyar, T. Başar, D. Teneketzis, and V. V. Veeravalli, "Optimal strategies for communication and remote estimation with an energy harvesting sensor," *IEEE Trans. Autom. Control*, vol. 58, no. 9, pp. 2246–2260, Sep. 2013.

[34] P. Tallapragada and J. Cortés, "Event-triggered stabilization of linear systems under bounded bit rates," *IEEE Trans. Autom. Control*, vol. 61, no. 6, pp. 1575–1589, Jun. 2016.

[35] M. J. Khojasteh, P. Tallapragada, J. Cortés, and M. Franceschetti, "Time-triggering versus event-triggering control over communication channels," in *Proc. IEEE 56th Annu. Conf. Decis. Control*, 2017, pp. 5432–5437.

[36] K. Gatsis, A. Ribeiro, and G. J. Pappas, "Optimal power management in wireless control systems," *IEEE Trans. Autom. Control*, vol. 59, no. 6, pp. 1495–1510, Jun. 2014.

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

[38] W. Zhang, M. Branicky, and S. Phillips, "Stability of networked control systems," *IEEE Control Syst. Mag.*, vol. 21, no. 1, pp. 84–99, Feb. 2001.

[39] W. M. H. Heemels, A. R. Teel, N. van de Wouw, and D. Nesic, "Networked control systems with communication constraints: Tradeoffs between transmission intervals, delays and performance," *IEEE Trans. Autom. Control*, vol. 55, no. 8, pp. 1781–1796, Aug. 2010.

[40] M. Donkers, W. Heemels, N. Van De Wouw, and L. Hetel, "Stability analysis of networked control systems using a switched linear systems approach," *IEEE Trans. Autom. Control*, vol. 56, no. 9, pp. 2101–2115, Sep. 2011.

[41] J. Le Ny, E. Feron, and G. J. Pappas, "Resource constrained LQR control under fast sampling," in *Proc. 14th Int. Conf. Hybrid Syst.: Comput. Control*, 2011, pp. 271–280.

[42] A. Molin and S. Hirche, "Price-based adaptive scheduling in multi-loop control systems with resource constraints," *IEEE Trans. Autom. Control*, vol. 59, no. 12, pp. 3282–3295, Dec. 2014.

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

[44] M. Vasconcelos and N. Martins, "Remote estimation games over shared networks," in *Proc. 52nd Annu. Allerton Conf. Commun., Control, Comput.*, 2014, pp. 12–18.

[45] K. Gatsis, M. Pajic, A. Ribeiro, and G. J. Pappas, "Opportunistic control over shared wireless channels," *IEEE Trans. Autom. Control*, vol. 60, no. 12, pp. 3140–3155, Dec. 2015.

[46] K. Gatsis, A. Ribeiro, and G. J. Pappas, "Random access design for wireless control systems," *Automatica*, vol. 91, pp. 1–9, May 2018.

[47] F. Smarra, M. D. Di Benedetto, and A. D'Innocenzo, "Efficient routing redundancy design over lossy networks," *Int. J. Robust Nonlinear Control*, vol. 28, no. 6, pp. 2574–2597, 2018.

[48] D. Maity, M. H. Mamduhi, S. Hirche, K. H. Johansson, and J. S. Baras, "Optimal LQG control under delay-dependent costly information," *IEEE Control Syst. Lett.*, vol. 3, no. 1, pp. 102–107, Jan. 2019.

[49] Y. Sun, E. Uysal-Biyikoglu, R. D. Yates, C. E. Koksal, and N. B. Shroff, "Update or wait: How to keep your data fresh," *IEEE Trans. Inf. Theory*, vol. 63, no. 11, pp. 7492–7508, Nov. 2017.

[50] Q. He, D. Yuan, and A. Ephremides, "Optimal link scheduling for age minimization in wireless systems," *IEEE Trans. Inf. Theory*, vol. 64, no. 7, pp. 5381–5394, Jul. 2018.

[51] I. Kadota, A. Sinha, and E. Modiano, "Optimizing age of information in wireless networks with throughput constraints," in *Proc. IEEE Conf. Comput. Commun.*, 2018, pp. 1844–1852.

[52] V. N. Swamy *et al.*"Real-time cooperative communication for automation over wireless," *IEEE Trans. Wireless Commun.*, vol. 16, no. 11, pp. 7168–7183, Nov. 2017.

[53] V. N. Swamy, P. Rigge, G. Ranade, A. Sahai, and B. Nikolić, "Network coding for high-reliability low-latency wireless control," in *Proc. IEEE Wireless Commun. Netw. Conf.*, 2016, pp. 1–7.

[54] L. Dickstein, V. N. Swamy, G. Ranade, and A. Sahai, "Finite block length coding for low-latency high-reliability wireless communication," in *Proc. 54th Annu. Allerton Conf. Commun., Control, Comput.*, 2016, pp. 908–915.

[55] Y. Hu, M. Serror, K. Wehrle, and J. Gross, "Finite blocklength performance of cooperative multi-terminal wireless industrial networks," *IEEE Trans. Veh. Technol.*, vol. 67, no. 7, pp. 5778–5792, Jul. 2018.

[56] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2307–2359, May 2010.

[57] R. Dobrushin, "Mathematical problems in the Shannon theory of optimal coding of information," in *Proc. 4th Berkeley Symp. Math., Statist., Probability*, 1961, vol. 1, pp. 211–252.

[58] V. Strassen, "Asymptotische abschätzungen in Shannons informationstheorie," in *Proc. Trans. 3rd Prague Conf. Inf. Theory*, 1967, pp. 689–723.

[59] H. Hassani, K. Alishahi, and R. Urbanke, "Finite-length scaling for polar codes," *IEEE Trans. Inf. Theory*, vol. 60, no. 10, pp. 5875–5898, Oct. 2014.

[60] H. Hassani and R. Urbanke, "Scaling exponent of spatially coupled codes," *IEEE Trans. Inf. Theory*, vol. 61, no. 9, pp. 4838–4851, 2015.

[61] T. Tanaka, K. H. Johansson, and M. Skoglund, "Optimal block length for data-rate minimization in networked LQG control," *IFAC-PapersOnLine*, vol. 49, no. 22, pp. 133–138, 2016.

[62] N. Brahmi, O. N. Yilmaz, K. W. Helmersson, S. A. Ashraf, and J. Torsner, "Deployment strategies for ultra-reliable and low-latency communication in factory automation," in *Proc. IEEE Globecom Workshops*, 2015, pp. 1–6.

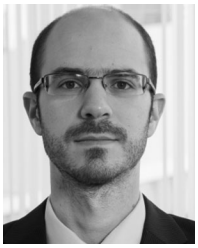
[63] Z. Pang, M. Luvisotto, and D. Dzung, "Wireless high-performance communications: The challenges and opportunities of a new target," *IEEE Ind. Electron. Mag.*, vol. 11, no. 3, pp. 20–25, Sep. 2017.

[64] A. Amraoui, A. Montanari, T. Richardson, and R. Urbanke, "Finite-length scaling for iteratively decoded LDPC ensembles," *IEEE Trans. Inf. Theory*, vol. 55, no. 2, pp. 473–498, Feb. 2009.

[65] E. Arikan, "Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels," *IEEE Trans. Inf. Theory*, vol. 55, no. 7, pp. 3051–3073, Jul. 2009.

[66] T. Richardson and R. Urbanke, *Modern Coding Theory*. Cambridge, U.K.: Cambridge Univ. Press, 2008.


Authorized licensed use limited to: University of Pennsylvania. Downloaded on October 29,2026 at 15:10:15 UTC from IEEE Xplore. Restrictions apply.



**Konstantinos Gatsis** (Member, IEEE) received the Ph.D. degree in electrical and systems engineering from the University of Pennsylvania, Philadelphia, PA, USA, in 2016.

Currently, he is a Departmental Lecturer with the Department of Engineering Science, University of Oxford, Oxford, U.K. His research interests include control and optimization applied to cyber-physical systems.

Dr. Gatsis was the recipient of the 2014 O. Hugo Schuck Best Paper Award, the Student Best Paper Award at the 2013 American Control Conference, and was a Best Paper Award Finalist at the 2014 ACM/IEEE International Conference on Cyber-Physical Systems.



**Hamed Hassani** (Member, IEEE) received the B.Sc. degree in electrical engineering and mathematics from the Sharif University of Technology, Tehran, Iran, in 2007, and the Ph.D. degree in computer and communication sciences from EPFL, Lausanne, Switzerland, in 2013.

He is an Assistant Professor with the Department of Electrical and Systems Engineering (ESE), University of Pennsylvania, Philadelphia, PA, USA. Prior to that, he was a Research Fellow with the Simons Institute at UC Berkeley, Berkeley, CA, USA, and a Postdoctoral Scholar with the Institute for Machine Learning, ETH Zurich, Zrich, Switzerland.

Dr. Hassani was the recipient of 2014 IEEE Information Theory Society Dissertation Award, as well as 2015 Jack K. Wolf ISIT Student Paper Award.



**George J. Pappas** (Fellow, IEEE) received the Ph.D. degree in electrical engineering and computer sciences from the University of California, Berkeley, CA, USA, in 1998.

He is currently the Joseph Moore Professor and Chair of the Department of Electrical and Systems Engineering, University of Pennsylvania, Philadelphia, PA, USA. He also holds a secondary appointment with the Department of Computer and Information Sciences and the Department of Mechanical Engineering and Applied Mechanics. He is a member of the GRASP Lab and the PRECISE Center. He had previously served as the Deputy Dean for Research with the School of Engineering and Applied Science. His research interests include control theory and, in particular, hybrid systems, embedded systems, cyber-physical systems, and hierarchical and distributed control systems, with applications to unmanned aerial vehicles, distributed robotics, green buildings, and biomolecular networks.

Dr. Pappas was the recipient of various awards, such as the Antonio Ruberti Young Researcher Prize, the George S. Axelby Award, the Hugo Schuck Best Paper Award, the George H. Heilmeier Award, the National Science Foundation PECASE Award, and numerous best student papers awards at American Control Conference (ACC), Conference on Decision and Control (CDC), and International Conference on Cyber-physical systems (ICCPS).

Authorized licensed use limited to: University of Pennsylvania. Downloaded on October 29,2026 at 15:10:15 UTC from IEEE Xplore. Restrictions apply.