Stabilizing a linear system using phone calls: when time is information

Mohammad Javad Khojasteh, Massimo Franceschetti, Gireeja Ranade

Abstract—**We consider the problem of stabilizing an undisturbed, scalar, linear system over a "timing" channel, namely a channel where information is communicated through the timestamps of the transmitted symbols. Each symbol transmitted from a sensor to a controller in a closed-loop system is received subject to some to random delay. The sensor can encode messages in the waiting times between successive transmissions and the controller must decode them from the inter-reception times of successive symbols. This set-up is analogous to a telephone system where a transmitter signals a phone call to a receiver through a "ring" and, after the random delay required to establish the connection; the receiver is aware of the "ring" being received. Since there is no data payload exchange between the sensor and the controller, this set-up provides an abstraction for performing event-triggering control with zero-payload rate. We show the following requirement for stabilization: for the state of the system to converge to zero in probability, the *timing capacity* of the channel should be, essentially, at least as large as the *entropy rate* of the system. Conversely, in the case the symbol delays are exponentially distributed, we show an "almost" tight sufficient condition using a coding strategy that refines the estimate of the decoded message every time a new symbol is received. Our results generalize previous zero-payload event-triggering control strategies, revealing a fundamental limit in using timing information for stabilization, independent of any transmission strategy.**

Index Terms—**Timing channel, control with communication constraints, event-triggered control, linear systems.**

I. Introduction

A networked control system with a feedback loop over a communication channel provides a first-order approximation of a cyber-physical system (CPS), where the interplay between the communication and control aspects of the system leads to new and unexpected analysis and design challenges [3], [4]. In this setting, data-rate theorems quantify the impact of the communication channel on the ability to stabilize the system. Roughly speaking, these theorems state that stabilization requires a communication rate in the feedback loop at least as large as the intrinsic *entropy rate* of the system, expressed by the sum of the logarithms of its unstable eigenvalues [5]–[12].

We consider a specific communication channel in the loop — a *timing channel*. Here, information is communicated through the timestamps of the symbols transmitted over the channel; the time is carrying the message. This formulation is motivated by recent works in event-triggering control, showing that the timing of the triggering events carries information that can be used for stabilization [13]–[20]. By encoding information in timing, stabilization can be achieved by transmitting additional data at a rate arbitrarily close to zero. However, in these works, the timing information was not explicitly quantified, and the analysis was limited to specific event-triggering strategies. In this paper, our goal is to determine the value of a timestamp from an information-theoretic perspective, when this timestamp is used for control. We are further motivated by the results on the impact of multiplicative noise in control [21], [22], since timing uncertainty can lead to multiplicative noise in systems and thus can serve as an information bottleneck.

To illustrate the proof of concept that timing carries information useful for control, we consider the simple case of stabilization of a scalar, undisturbed, continuous-time, unstable, linear system over a timing channel and rely on the information-theoretic notion of *timing capacity* of the channel, namely the amount of information that can be encoded using time stamps [23]–[39]. In this setting, the sensor can communicate with the controller by choosing the timestamps at which symbols from a unitary alphabet are transmitted. The controller receives each transmitted symbol after a random delay is added to the timestamp. We show the following data-rate theorem. For the state to converge to zero in probability, the timing capacity of the channel should be, essentially, at least as large as the entropy rate of the system. Conversely, in the case the random delays are exponentially distributed, we show that when the timing capacity is strictly greater than the entropy rate of the system, we can drive the state to zero in probability by using a decoder that refines its estimate of the transmitted message every time a new symbol is received [40]. We also derive analogous necessary and sufficient conditions for the problem of estimating the state of the system with an error that tends to zero in probability.

The books [5], [6], [50], [51] and the surveys [7], [8], [52] provide detailed discussions of data-rate theorems and related results that heavily inspire this work. A portion of the literature studied stabilization over "bit-pipe channels," where a rate-limited, possibly time-varying and erasure-prone communication channel is present in the feedback loop [41], [46]–[48], [53]. For more general noisy channels, Tatikonda and Mitter [42] and Matveev and Savkin [43] showed that the state of undisturbed linear systems can be forced to converge to zero almost surely (a.s.) if and only if the Shannon capacity of the channel is larger than the entropy rate of the system. In the presence of disturbances, in order to keep the state bounded a.s., a more stringent condition is required, namely the zero-error capacity of the channel must be larger than

The material in this paper was presented in part at the 18th European Control Conference (ECC), 2019 [1], and at 57th IEEE Conference on on Decision and Control, 2018 [2]. This research was partially supported by NSF awards CNS-1446891 and ECCS-1917177.

M. J. Khojasteh is with the Wireless Communication and Network Sciences Laboratory (WINS Lab), Massachusetts Institute of Technology. Some of this work was performed while at University of California, San Diego, and also visiting Microsoft Research. (mkhojast@mit.edu)

M. Franceschetti is with the Department of Electrical and Computer Engineering, University of California, San Diego. (massimo@ucsd.edu)

G. Ranade is with EECS department at the University of California, Berkeley. Some of this work was performed while at Microsoft Research, and also at the Simons Institute for the Theory of Computing. (gireeja@eecs.berkeley.edu)

TABLE I: Capacity notions used to derive data-rate theorems in the literature under different notions of stability, channel types, and system disturbances.

Work	Disturbance	Channel	Stability condition	Capacity
[41]	NO	Bit-pipe	$\|X(t)\| \to 0$ a.s.	Shannon
[42], [43]	NO	DMC	$\|X(t)\| \to 0$ a.s.	Shannon
[44]	bounded	DMC	$\mathbb{P}(\sup_t \|X(t)\| < \infty) = 1$	Zero-Error
[6, Ch. 8]	bounded	DMC	$\mathbb{P}(\sup_t \|X(t)\| < K_\epsilon) > 1 - \epsilon$	Shannon
[45]	bounded	DMC	$\sup_t \mathbb{E}(\|X(t)\|^m) < \infty$	Anytime
[46]	unbounded	Bit-Pipe	$\sup_t \mathbb{E}(\|X(t)\|^2) < \infty$	Shannon
[47]–[49]	unbounded	Var. Bit-pipe	$\sup_t \mathbb{E}(\|X(t)\|^m) < \infty$	Anytime
This paper	NO	Timing	$\|X(t)\| \xrightarrow{P} 0$	Timing

the entropy rate of the system [44]. Nair derived a similar information-theoretic result in a non-stochastic setting [54]. Sahai and Mitter [45] considered moment-stabilization over noisy channels and in the presence of system disturbances of bounded support, and provided a data-rate theorem in terms of the anytime capacity of the channel. They showed that to keep the mth moment of the state bounded, the anytime capacity of order m should be larger than the entropy rate of the system. The anytime capacity has been further investigated in [49], [55]–[57]. Matveev and Savkin [6, Chapter 8] have also introduced a weaker notion of stability in probability, requiring the state to be bounded with probability $(1-\epsilon)$ by a constant that diverges as $\epsilon \to 0$, and showed that in this case it is possible to stabilize linear systems with bounded disturbances over noisy channels provided that the Shannon capacity of the channel is larger than the entropy rate of the system. The various results, along with our contribution, are summarized in Table I. The main point that can be drawn from all of these results is that the relevant capacity notion for stabilization over a communication channel critically depends on the notion of stability and on the system's model.

From the system's perspective, our set-up is closest to the one in [41]–[43], as there are no disturbances and the objective is to drive the state to zero. Our convergence in probability provides a stronger necessary condition for stabilization, but a weaker sufficient condition than the one in these works. We also point out that our notion of stability is considerably stronger than the notion of probabilistic stability proposed in [6, Chapter 8]. Some additional works considered nonlinear plants without disturbances [58]–[60], and switched linear systems [61], [62] where communication between the sensor and the controller occurs over a bit-pipe communication channel. The recent work in [63] studies estimation of nonlinear systems over noisy communication channels and the work in [64] investigates the trade-offs between the communication channel rate and the cost of the linear quadratic regulator for linear plants.

Parallel work in control theory has investigated the possibility of stabilizing linear systems using timing information. One primary focus of the emerging paradigm of event-triggered control [65]–[77] has been on minimizing the number of transmissions while simultaneously ensuring the control objective [16], [78], [79]. Rather than performing periodic communication between the system and the controller, in event-triggered control communication occurs only as needed, in an opportunistic manner. In this setting, the timing of the triggering events can carry useful information about the state of the system, that can be used for stabilization [13]–[20]. In this context, it has been shown that the amount of timing information is sensitive to the delay in the communication channel. While for small delay stabilization can be achieved using only timing information and transmitting data payload (i.e. physical data) at a rate arbitrarily close to zero, for large values of the delay this is not the case, and the data payload rate must be increased [15], [19]. In this paper, we extend these results from an information-theoretic perspective, as we explicitly quantify the value of the timing information, independent of any transmission strategy. To quantify the amount of timing information alone, we restrict to transmitting symbols from a unitary alphabet, i.e. at zero data payload rate. Research directions left open for future investigation include the study of "mixed" strategies, using both timing information and physical data transmitted over a larger alphabet, as well as generalizations to vector systems and the study of systems with disturbances. In the latter case, it is likely that the usage of stronger notions of capacity, or weaker notions of stability, will be necessary.

The rest of the paper is organized as follows. Section II introduces the system and channels models. The main results are presented in Section III. Section IV considers the estimation problem, and Section V considers the stabilization problem. Section VI provides a comparison with related work, and Section VII presents a numerical example. Conclusions are drawn in Section VIII.

A. Notation

Let $X^n = (X_1, \cdots, X_n)$ denote a vector of random variables and let $x^n = (x_1, \cdots, x_n)$ denote its realization. If $X_1, \cdots, X_n$ are independent and identically distributed (i.i.d) random variables, then we refer to a generic $X_i \in X^n$ by X and skip the subscript i. We use $\log$ and $\ln$ to denote the logarithms base 2 and base e respectively. We use $H(X)$ to denote the Shannon entropy of a discrete random variable X and $h(X)$ to denote the differential entropy of a continuous random variable X. Further, we use $I(X;Y)$ to indicate the mutual information between random variables X and Y. We

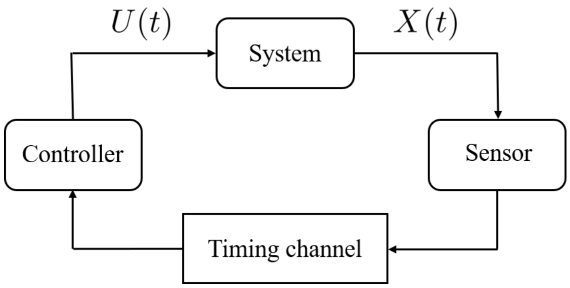

Fig. 1: Model of a networked control system where the feedback loop is closed over a timing channel.

write $X_n \xrightarrow{P} X$ if X_n converges in probability to X. Similarly, we write $X_n \xrightarrow{a.s.} X$ if X_n converges almost surely to X. For any set $\mathscr{X}$ and any $n \in \mathbb{N}$ we let

$$\pi_n : \mathscr{X}^{\mathbb{N}} \to \mathscr{X}^n \tag{1}$$

be the truncation operator, namely the projection of a sequence in $\mathscr{X}^{\mathbb{N}}$ into its first n symbols.

II. System and channel model

We consider the networked control system depicted in Fig. 1. The system dynamics are described by a scalar, continuous-time, noiseless, linear time-invariant (LTI) system

$$\dot{X}(t) = aX(t) + bU(t), \tag{2}$$

where $X(t) \in \mathbb{R}$ and $U(t) \in \mathbb{R}$ are the system state and the control input respectively. The constants $a, b \in \mathbb{R}$ are such that $a > 0$ and $b \neq 0$. The initial state $X(0)$ is random and is drawn from a distribution of bounded differential entropy and bounded support, namely $h(X(0)) < \infty$ and $|X(0)| < L$, where L is known to both the sensor and the controller. Conditioned on the realization of $X(0)$, the system evolution is deterministic. Both controller and sensor have knowledge of the system dynamics in (2). We assume the sensor can measure the state of the system with infinite precision, and the controller can apply the control input to the system with infinite precision and with zero delay.

The sensor is connected to the controller through a *timing channel* (the telephone signaling channel defined in [23]). The operation of this channel is analogous to that of a telephone system where a transmitter signals a phone call to the receiver through a "ring" and, after a random time required to establish the connection, is aware of the "ring" being received. Communication between transmitter and receiver can then occur without any vocal exchange, but by encoding messages in the "waiting times" between consecutive calls.

A. The channel

We model the channel as carrying symbols ♠ from a unitary alphabet, and each transmission is received after a random delay. Every time a symbol is received, the sender is notified of the reception by an instantaneous acknowledgment. The channel is initialized with a ♠ received at time $t = 0$. After receiving the acknowledgment for the ith ♠, the sender waits for W_{i+1} seconds and then transmits the next ♠. Transmitted symbols are subject to i.i.d. random delays $\{S_i\}$. Letting D_i be the inter-reception time between two consecutive symbols, we have

$$D_i = W_i + S_i. \tag{3}$$

It follows that the reception time of the nth symbol is

$$\mathcal{T}_n = \sum_{i=1}^{n} D_i. \tag{4}$$

Fig. 2 provides an example of the timing channel in action.

B. Capacity of the channel

We start by reviewing some definitions from [23].

Definition 1: A (n, M, T, δ)-timing code for the telephone signaling channel consists of a codebook of M codewords $\{(w_{i,m}, i = 1, \ldots, n),\ m = 1 \ldots M\}$, as well as a decoder, which upon observation of $(D_1, \ldots, D_n)$ selects the correct transmitted codeword with probability at least $1-\delta$. Moreover, the codebook is such that the expected random arrival time of the nth symbol is at most T, namely

$$\mathbb{E}(\mathcal{T}_n) \leq T. \tag{5}$$

Definition 2: The rate of an (n, M, T, δ)-timing code is

$$R = (\log M)/T. \tag{6}$$

Definition 3: The timing capacity C of the telephone signaling channel is the supremum of the achievable rates, namely the largest R such that for every $\gamma > 0$ there exists a sequence of $(n, M_n, T_n, \delta_{T_n})$-timing codes that satisfy

$$\frac{\log M_n}{T_n} > R - \gamma, \tag{7}$$

and $\delta_{T_n} \to 0$ as $n \to \infty$.

The following result [23, Theorem 8] characterizes the capacity of the telephone signaling channel.

Theorem 1 (Anantharam and Verdú): The timing capacity of the telephone signaling channel is given by

$$C = \sup_{\chi > 0} \sup_{\substack{W \geq 0 \\ \mathbb{E}(W) \leq \chi}} \frac{I(W; W+S)}{\mathbb{E}(S) + \chi}, \tag{8}$$

and if S is exponentially distributed then

$$C = \frac{1}{e\mathbb{E}(S)} \quad \text{[nats/sec]}. \tag{9}$$

In this paper, we assume that the waiting times $\{W_i\}$ used to encode any given message are generated at random in an i.i.d. fashion, and are also independent of the random delays $\{S_i\}$. Assuming the symbols in each codeword are picked i.i.d. from a common distribution restricts the encoder to using a fixed random telephoning policy. This assumption comes at no loss of generality since: (i) the capacity in (8) is achieved by i.i.d. random codes [23], and (ii) in our system model there are no disturbances and therefore the control problem reduces to the communication of a fixed real-valued variable representing the initial condition with exponential reliability over a digital channel, which can be performed optimally using a fixed random coding strategy [40].

C. The sensor

The sensor in Fig. 1 can act as a source and channel encoder. Based on its source knowledge, namely the knowledge of the initial condition $X(0)$, system dynamics (2), and L, it selects the waiting times $\{W_i\}$ between the reception and the transmission of consecutive ♠ symbols.

As in [23], [26] we assume that the causal acknowledgments received by the sensor every time a ♠ is delivered to the controller are not used to choose the waiting times, but only to avoid queuing, ensuring that every symbol is sent after the previous one has been received. This applies to TCP-based networks, where packet deliveries are acknowledged via a feedback link [80]–[84]. For networked control systems, this causal acknowledgment can be obtained without assuming an additional communication channel in the feedback loop. The controller can signal the acknowledgment to the sensor by applying a control input to the system that excites a specific frequency of the state each time a symbol has been received. This strategy is known in the literature as "acknowledgment through the control input" [6], [13], [42], [45].

D. The controller

The controller in Fig. 1 can act as a source and channel decoder. It uses the reception times of all the symbols received up to time t, along with the knowledge of L and of the system dynamics (2) to decode the source message, compute the control input $U(t)$, and apply it to the system. The control input can be refined over time, as the estimate of the source can be decoded with increasing accuracy when more and more symbols are received. The objective is to design an encoding and decoding strategy to stabilize the system by driving the state to zero in probability, i.e. we want $|X(t)| \xrightarrow{P} 0$ as $t \to \infty$.

Although the computational complexity of different encoding-decoding schemes is a key practical issue, in this paper we are concerned with the existence of schemes satisfying our objective, rather than with their practical implementation.

III. Main results

A. Necessary condition

To derive a necessary condition for the stabilization of the feedback loop system depicted in Fig. 1, we first consider the problem of estimating the state in open-loop over the timing channel along a specific sequence of estimation times. We show that if the estimation error tends to zero in probability along this sequence, then for all $\nu > 0$ the timing capacity must be at least as large as $(1-\nu)$ times the entry rate of the system. This result holds for any source and random channel coding strategies adopted by the sensor, and for any strategy adopted by the controller to generate the control input. Our proof employs a rate-distortion argument to compute a lower bound on the minimum number of bits required to represent the state up to any given accuracy, and this leads to a corresponding lower bound on the required timing capacity of the channel. We then show that the same bound on the timing capacity holds for stabilization, since in order to have $|X(t)| \xrightarrow{P} 0$ as $t \to \infty$ in closed-loop, the estimation error in open-loop must tend to zero in probability as $t \to \infty$, and therefore, in particular, along the designed sequence of estimation times.

B. Sufficient condition

To derive a sufficient condition for stabilization, we first consider the problem of estimating the state in open-loop over the timing channel. We focus on a specific sequence of estimation times. We provide an explicit source-channel coding scheme which guarantees that if for all $\nu > 0$ the timing capacity is larger than $(1+\nu)$ times the entropy entropy rate of the system, then the estimation error tends to zero in probability. We then show that this condition is also sufficient to construct a control scheme such that $|X(t)| \xrightarrow{P} 0$ as $t \to \infty$. The main idea behind our strategy is based on the realization that in the absence of disturbances all that is needed to drive the state to zero is communicating the initial condition $X(0)$ to the controller with accuracy that increases exponentially over time. Once this is achieved, the controller can estimate the state $X(t)$ with increasing accuracy over time, and continuously apply an input that drives the state to zero. This idea has been exploited before in the literature [41], [42], and the problem is related to the anytime reliable transmission of a real-valued variable over a digital channel [40]. Here, we cast this problem in the framework of the timing channel. A main difficulty in our case is to ensure that we can drive the system's state to zero in probability despite the unbounded random delays occurring in the timing channel.

In the source coding process, we quantize the interval $[-L, L]$ uniformly using a tree-structured quantizer [85]. We then map the obtained source code into a channel code suitable for transmission over the timing channel, using the capacity-achieving random codebook of [23]. Given $X(0)$, the encoder picks a codeword from an arbitrarily large codebook and starts transmitting the real numbers of the codeword one by one, where each real number corresponds to a holding time, and proceeds in this way forever. Every time a sufficiently large number of symbols are received, we use a maximum likelihood decoder to successively refine the controller's estimate of $X(0)$. Namely, the controller re-estimates $X(0)$ based on the new inter-reception times and all previous inter-reception times, and uses it to compute the new state estimate of $X(t)$ and control input $U(t)$. We show that when the sensor quantizes $X(0)$ at sufficiently high resolution, and when the timing capacity is larger than the entropy rate of the system, the controller can construct a sufficiently accurate estimate of $X(t)$ and compute $U(t)$ such that $|X(t)| \xrightarrow{P} 0$ as $t \to \infty$.

IV. The estimation problem

We start considering the estimation problem depicted in Fig. 3. By letting $b = 0$ in (2) we obtain the open-loop equation

$$\dot{X}_e(t) = aX_e(t). \tag{10}$$

We assume that the encoder has causal knowledge of the reception times via acknowledgments through the system as depicted in Fig. 3. Our first objective is to obtain a necessary condition on the capacity of the timing channel required to

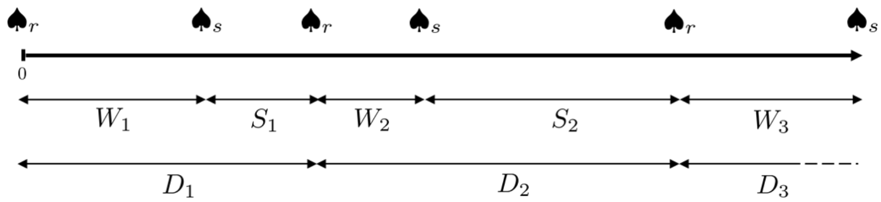

Fig. 2: The timing channel. Subscripts s and r are used to denote sent and received symbols, respectively.

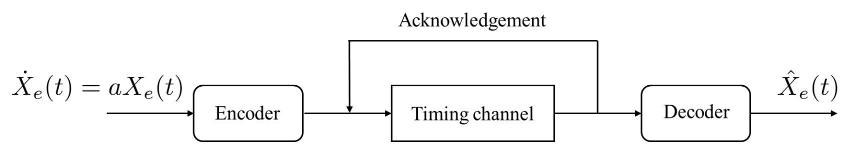

Fig. 3: The estimation problem.

construct an estimate $\hat{X}_e(t_n)$ such that for all $\nu > 0$ and any sequence of estimation times t_n that satisfies

$$1 - \nu \leq \lim_{n\to\infty} \frac{t_n}{\mathbb{E}(\mathcal{T}_n)} < 1, \tag{11}$$

we have $|X_e(t_n) - \hat{X}_e(t_n)| \xrightarrow{P} 0$ as $n \to \infty$.

Our second objective is to obtain a sufficient condition on the capacity of the timing channel that ensures the construction of an estimate $\hat{X}_e(t'_n)$ such that for all $\nu > 0$ and any sequence of estimation times t'_n that satisfies

$$1 < \lim_{n\to\infty} \frac{t'_n}{\mathbb{E}(\mathcal{T}_n)} \leq 1 + \nu, \tag{12}$$

we have $|X_e(t'_n) - \hat{X}_e(t'_n)| \xrightarrow{P} 0$ as $n \to \infty$.

The sequences of estimation times that satisfy (11) and (12) are "close" in the sense that

$$\lim_{n\to\infty} \frac{t'_n - t_n}{\mathbb{E}(\mathcal{T}_n)} \leq 2\nu. \tag{13}$$

Given the conditions in (11) and (12), the next lemma provides probabilistic bounds on the number of symbols that are received up to time t_n and t'_n, respectively.

Lemma 1: Given the conditions in (11), the probability $\mathbb{P}(\mathcal{T}_{n+1} \leq t_n)$ tends to zero. Moreover, given the condition in (12), as $n \to \infty$, the probability $\mathbb{P}(\mathcal{T}_n > t'_n)$ tends to zero.

Proof: We start by proving that $\mathbb{P}(\mathcal{T}_{n+1} \leq t_n)$ tends to zero as $n \to \infty$. For large enough n, using (11), we have that the probability of receiving the $n+1$ symbols before t_n is

$$\begin{aligned}\mathbb{P}(\mathcal{T}_{n+1} \leq t_n) &\leq \mathbb{P}(\mathcal{T}_{n+1}/(n+1) < \mathbb{E}(\mathcal{T}_n)/(n+1)) \\ &\leq \mathbb{P}(\mathcal{T}_{n+1}/(n+1) < \mathbb{E}(D)). \end{aligned} \tag{14}$$

Since the waiting times $\{W_i\}$ and the random delays $\{S_i\}$ are i.i.d. sequences and independent of each other, it follows by the strong law of large numbers that (14) tends to zero as $n \to \infty$.

We continue by bounding the probability of the event that the n-th symbol does not arrive by the estimation deadline t'_n. For large enough n, using (12), we have that the probability of missing the deadline is

$$\begin{aligned}\mathbb{P}(\mathcal{T}_n > t'_n) &\leq \mathbb{P}(\mathcal{T}_n/n > \mathbb{E}(\mathcal{T}_n)/n) \\ &= \mathbb{P}(\mathcal{T}_n/n > \mathbb{E}(D)). \end{aligned} \tag{15}$$

Since the waiting times $\{W_i\}$ and the random delays $\{S_i\}$ are i.i.d. sequences and independent of each other, it follows by the strong law of large numbers that (15) tends to zero as $n \to \infty$. ■

Lemma 1 leads to the following conclusions. First, with high probability as $n \to \infty$, by time t_n at most n symbols have been received. Second, with high probability as $n \to \infty$ the estimation at time t'_n is evaluated, after at least n symbols have been received.

A. Necessary condition

The next theorem provides a necessary condition on the timing capacity for the state estimation error to tend to zero in probability.

Theorem 2: Consider the estimation problem depicted in Fig. 3 with system dynamics (10). Consider transmitting n symbols over the telephone signaling channel (3), and the sequence of estimation times satisfying (11). If $|X_e(t_n) - \hat{X}_e(t_n)| \xrightarrow{P} 0$, then

$$I(W; W+S) \geq \ a\ (1-\nu)\ \mathbb{E}(W+S) \quad \text{[nats]}, \tag{16}$$

and consequently

$$C \geq a\,(1-\nu) \quad \text{[nats/sec]}. \tag{17}$$

The proof of Theorem 2 is given in the appendix.

B. Sufficient condition

The next theorem provides a sufficient condition for convergence of the state estimation error to zero in probability along any sequence of estimation times t'_n satisfying (12), in the case of exponentially distributed delays.

Theorem 3: Consider the estimation problem depicted in Fig. 3 with system dynamics (10). Consider transmitting n symbols over the telephone signaling channel (3). Assume $\{S_i\}$ are drawn i.i.d. from exponential distribution with mean $\mathbb{E}(S)$. If the capacity of the timing channel is at least

$$C \geq a\,(1+\nu) \quad \text{[nats/sec]}, \tag{18}$$

then for any sequence of times $\{t'_n\}$ that satisfies (12), we can compute an estimate $\hat{X}_e(t'_n)$ such that as $n \to \infty$, we have

$$|X_e(t'_n) - \hat{X}_e(t'_n)| \xrightarrow{P} 0. \tag{19}$$

The proof of Theorem 3 is given in the appendix. The result is strengthened in the next section (see Corollary 1), showing that $C > a\,(1+\nu)$ is also sufficient to drive the state estimation error to zero in probability for all $t \to \infty$.

Remark 1: Since $\nu > 0$ can be chosen to be sufficiently small Theorems 2 and 3 provide an "almost" tight necessary and sufficient condition for the estimation problem. The entropy-rate of our system is a nats/time [58], [86]–[89]. This represents the amount of uncertainty per unit time generated by the system in open loop. In fact, (16), can be seen as a typical scenario in data-rate theorems: to drive the error to zero the mutual information between an encoding symbol W and its received noisy version $W + S$ should be larger than the average "information growth" of the state during the inter-reception interval D, which is given by

$$\mathbb{E}(aD) = a\ \mathbb{E}(W + S). \tag{20}$$

•

V. The stabilization problem

A. Necessary condition

We now turn to consider the stabilization problem. Our first lemma states that if in closed-loop we are able to drive the state to zero in probability, then in open-loop we are also able to estimate the state with vanishing error in probability.

Lemma 2: Consider stabilization of the closed-loop system (2) and estimation of the open-loop system (10) over the timing channel (3). If there exists a controller such that $|X(t)| \xrightarrow{P} 0$ as $t \to \infty$, in closed-loop, then there exists an estimator such that $|X_e(t) - \hat{X}_e(t)| \xrightarrow{P} 0$ as $t \to \infty$, in open-loop.

Proof: From (2), we have in closed loop

$$X(t) = e^{at}X(0) + \zeta(t), \tag{21}$$

$$\zeta(t) = e^{at}\int_0^t e^{-a\varrho}bU(\varrho)d\varrho. \tag{22}$$

It follows that if

$$\lim_{t\to\infty} \mathbb{P}\left(|X(t)| \le \epsilon\right) = 1, \tag{23}$$

then we also have

$$\lim_{t\to\infty} \mathbb{P}\left(\left|e^{at}X(0) + \zeta(t)\right| \le \epsilon\right) = 1. \tag{24}$$

On the other hand, from (10) we have in open loop

$$X_e(t) = e^{at}X(0), \tag{25}$$

and we can choose $\hat{X}_e(t) = -\zeta(t)$ so that

$$|X_e(t) - \hat{X}_e(t)| = |e^{at}X(0) + \zeta(t)| \xrightarrow{P} 0, \tag{26}$$

where the last step follows from (24). ■

The next theorem provides a necessary rate for the stabilization problem.

Theorem 4: Consider the stabilization of the closed-loop system (2). If $|X(t)| \xrightarrow{P} 0$ as $t \to \infty$, then

$$I(W; W + S)\ \ge\ a\ (1-\nu)\ \mathbb{E}(W + S)\quad \text{[nats]}, \tag{27}$$

and consequently

$$C\ \ge a\ (1-\nu)\quad \text{[nats/sec]}. \tag{28}$$

Proof: By Lemma 2 we have that if $|X(t)| \xrightarrow{P} 0$, then $|X_e(t) - \hat{X}_e(t)| \xrightarrow{P} 0$ for all $t \to \infty$, and in particular along a sequence $\{t_n\}$ satisfying (11). The result now follows from Theorem 2. ■

B. Sufficient condition

Our next lemma strengthens our estimation results, stating that it is enough for the state estimation error to converge to zero in probability as $n \to \infty$ along a sequence of estimation times $\{t'_n\}$ satisfying (12), to ensure it converges to zero for all $t \to \infty$.

Lemma 3: Consider estimation of the system (10) over the timing channel (3). If there exists $\Gamma_0 > 1$ such that along the sequence of estimation times $t'_n = \Gamma_0\mathbb{E}(\mathcal{T}_n)$ we have $|X_e(t'_n) - \hat{X}_e(t'_n)| \xrightarrow{P} 0$ as $n \to \infty$, then for all $t \to \infty$ we also have $|X_e(t) - \hat{X}_e(t)| \xrightarrow{P} 0$.

Proof: We have that for $t'_n = \Gamma_0\mathbb{E}(\mathcal{T}_n)$ and for all $\epsilon' > 0$, and $\phi > 0$, there exist n_ϕ such that for all $n \ge n_\phi$

$$\mathbb{P}\left(|X_e(t'_n) - \hat{X}_e(t'_n)| > \epsilon'\right) \le \phi. \tag{29}$$

Let $t_{n_\phi} = \Gamma_0\mathbb{E}(\mathcal{T}_{n_\phi})$ be the time at which we estimate the state for the n_ϕth time. We want to show that for all $t \in [t_{n_\phi}, t_{n_\phi+1}]$ and $\epsilon > 0$, we also have

$$\mathbb{P}\left(|X_e(t) - \hat{X}_e(t)| > \epsilon\right) \le \phi. \tag{30}$$

Consider the random time $\mathcal{T}_{n_\phi}$ at which ♠ is received for the n_ϕth time. We have

$$\begin{aligned} t_{n_\phi+1} - t_{n_\phi} &= \Gamma_0\,\mathbb{E}(\mathcal{T}_{n_\phi+1}) - \Gamma_0\,\mathbb{E}(\mathcal{T}_{n_\phi}) \\ &= (n_\phi+1)\Gamma_0\,\mathbb{E}(D) - n_\phi\Gamma_0\,\mathbb{E}(D) \\ &= \Gamma_0\,\mathbb{E}(D). \end{aligned} \tag{31}$$

For all $t \in [t_{n_\phi}, t_{n_\phi+1}]$, from the open-loop equation (10) we have

$$X_e(t) = e^{a(t-t_{n_\phi})}X_e(t_{n_\phi}). \tag{32}$$

We then let

$$\hat{X}_e(t) = e^{a(t-t_{n_\phi})}\hat{X}_e(t_{n_\phi}). \tag{33}$$

Combining (32) and (33) and using (31), we obtain that for all $t \in [t_{n_\phi}, t_{n_\phi+1}]$

$$|X_e(t) - \hat{X}_e(t)| \le e^{a\Gamma_0\,\mathbb{E}(D)}|X_e(t_{n_\phi}) - \hat{X}_e(t_{n_\phi})|. \tag{34}$$

From which it follows that

$$\begin{aligned} &\mathbb{P}\left(|X_e(t) - \hat{X}_e(t)| > \epsilon' e^{a\Gamma_0\,\mathbb{E}(D)}\right) \\ &\le \mathbb{P}\left(|X_e(t_{n_\phi}) - \hat{X}_e(t_{n_\phi})| > \epsilon'\right). \end{aligned} \tag{35}$$

Since (29) holds for all $n \geq n_\phi$, we also have

$$\mathbb{P}\left(|X_e(t_{n_\phi}) - \hat{X}_e(t_{n_\phi})| \geq \epsilon'\right) \leq \phi. \tag{36}$$

We can now let $\epsilon' < \epsilon e^{-a\Gamma_0 \mathbb{E}(D)}$ and the result follows. ■

Lemma 3 yields the following corollary, which is an immediate extension of Theorem 3.

Corollary 1: Consider the estimation problem depicted in Fig. 3 with system dynamics (10). Consider transmitting n symbols over the telephone signaling channel (3). Assume $\{S_i\}$ are drawn i.i.d. from exponential distribution with mean $\mathbb{E}(S)$. If the capacity of the timing channel is at least $C \geq a\,(1+\nu)$, then we have $|X_e(t) - \hat{X}_e(t)| \xrightarrow{P} 0$ as $t \to \infty$.

Proof: We start by considering the sequence of estimation times $t'_n = (1+\nu)\,\mathbb{E}(\mathcal{T}_n)$. Since $C \geq a\,(1+\nu)$, by Theorem 3 we have $|X_e(t'_n) - \hat{X}_e(t'_n)| \xrightarrow{P} 0$ as $n \to \infty$. Then, by Lemma 3 we also have $|X_e(t) - \hat{X}_e(t)| \xrightarrow{P} 0$ as $t \to \infty$. ■

The next key lemma states that if we are able to estimate the state with vanishing error in probability, then we are also able to drive the state to zero in probability.

Lemma 4: Consider stabilization of the closed-loop system (2) and estimation of the open-loop system (10) over the timing channel (3). If there exists an estimator such that $|X_e(t) - \hat{X}_e(t)| \xrightarrow{P} 0$ as $t \to \infty$, in open-loop, then there exists a controller such that $|X(t)| \xrightarrow{P} 0$ as $t \to \infty$, in closed-loop.

Proof: We start by showing that if there exists an open-loop estimator such that $|X_e(t) - \hat{X}_e(t)| \xrightarrow{P} 0$ as $t \to \infty$, then there also exists a closed-loop estimator such that $|X(t) - \hat{X}(t)| \xrightarrow{P} 0$ as $t \to \infty$. We construct the closed-loop estimator based on the open-loop estimator as follows. The sensor in closed-loop runs a copy of the open-loop system by constructing the virtual open-loop dynamic

$$X_e(t) = X(0)e^{at}. \tag{37}$$

Using the open-loop estimator, for all $t > 0$ the controller acquires the open-loop estimate $\hat{X}_e(t)$ such that $|X_e(t) - \hat{X}_e(t)| \xrightarrow{P} 0$. It then uses this estimate to construct the closed-loop estimate

$$\hat{X}(t) = \hat{X}_e(t) + e^{at}\int_0^t e^{-a\varrho} bU(\varrho)d\varrho. \tag{38}$$

Since from (2) the true state in closed loop is

$$X(t) = X(0)e^{at} + e^{at}\int_0^t e^{-a\varrho} bU(\varrho)d\varrho, \tag{39}$$

it follows by combining (37), (38) and (39) that

$$|X(t) - \hat{X}(t)| = |X_e(t) - \hat{X}_e(t)| \xrightarrow{P} 0. \tag{40}$$

What remains to be proven is that if $|X(t) - \hat{X}(t)| \xrightarrow{P} 0$, then there exists a controller such that $|X(t)| \xrightarrow{P} 0$.

Let $b > 0$ and choose k so large that $a - bk < 0$. Let $U(t) = -k\hat{X}(t)$. From (2), we have

$$\dot{X}(t) = (a - bk)X(t) + bk[X(t) - \hat{X}(t)]. \tag{41}$$

By solving (41) and using the triangle inequality, we get

$$|X(t)| \leq |e^{(a-bk)t}X(0)| + \left|\int_0^t e^{(t-\varrho)(a-bk)} bk(X(\varrho) - \hat{X}(\varrho))d\varrho\right|. \tag{42}$$

Since $|X(0)| < L$ and $a - bk < 0$, the first term in (42) tends to zero as $t \to \infty$. Namely, for any $\epsilon > 0$ there exists a number N_ϵ such that for all $t \geq N_\epsilon$, we have

$$|e^{(a-bk)t}X(0)| \leq \epsilon. \tag{43}$$

Since by (40) we have that $|X(t) - \hat{X}(t)| \xrightarrow{P} 0$, we also have that for any $\epsilon, \delta > 0$ there exist a number N'_ϵ such that for all $t \geq N'_\epsilon$, we have

$$\mathbb{P}\left(|X(t) - \hat{X}(t)| \leq \epsilon\right) \geq 1 - \delta. \tag{44}$$

It now follows from (42) that for all $t \geq \max\{N_\epsilon, N'_\epsilon\}$ the following inequality holds with probability at least $(1-\delta)$

$$|X(t)| \leq \epsilon + bke^{t(a-bk)}\int_0^{N'_\epsilon} e^{-\varrho(a-bk)}|X(\varrho) - \hat{X}(\varrho)|d\varrho + \epsilon bke^{t(a-bk)}\int_{N'_\epsilon}^t e^{-\varrho(a-bk)}d\varrho. \tag{45}$$

Since both sensor and controller are aware that $|X(0)| < L$, by (37) we have that for all $t \geq 0$ the open-loop estimate acquired by the controller satisfies $\hat{X}_e(t) \in [-Le^{at}, Le^{at}]$. By (40) the closed-loop estimation error is the same as the open-loop estimation error, and we then have that for all $\varrho \in [0, N'_\epsilon]$

$$|X(\varrho) - \hat{X}(\varrho)| = |X_e(\varrho) - \hat{X}_e(\varrho)| \leq 2Le^{aN'_\epsilon}. \tag{46}$$

Substituting (46) into (45), we obtain that with probability at least $(1-\delta)$

$$|X(t)| \leq \epsilon + 2Lbke^{[t(a-bk)+aN'_\epsilon]}\frac{e^{-N'_\epsilon(a-bk)} - 1}{-(a-bk)} + \epsilon bke^{t(a-bk)}\,\frac{e^{-t(a-bk)} - e^{-N'_\epsilon(a-bk)}}{-(a-bk)}. \tag{47}$$

By first letting ϵ be sufficiently close to zero, and then letting t be sufficiently large, we can make the right-hand side of (47) arbitrarily small, and the result follows. ■

The next theorem combines the results above, providing a sufficient condition for convergence of the state to zero in probability in the case of exponentially distributed delays.

Theorem 5: Consider the stabilization of the system (2). Assume $\{S_i\}$ are drawn i.i.d. from an exponential distribution with mean $\mathbb{E}(S)$. If the capacity of the timing channel is at least

$$C \geq a\,(1+\nu) \quad \text{[nats/sec]}, \tag{48}$$

then $|X(t)| \xrightarrow{P} 0$ as $t \to \infty$.

VI. Comparison with previous work

A. Comparison with stabilization over an erasure channel

In [42] the problem of stabilization of the discrete-time version of the system in (2) over an erasure channel has been considered. In this discrete model, at each time step of the system's evolution the sensor transmits I bits to the controller and these bits are successfully delivered with probability $1-\mu$, or they are dropped with probability μ, in an independent fashion. It is shown that a necessary condition for $X(k) \xrightarrow{a.s} 0$ is that the capacity of this I-bit erasure channel is

$$(1-\mu)I \geq \log a \quad \text{[bits/sec]}. \tag{49}$$

Since almost sure convergence implies convergence in probability, by Theorem 4 we have that the following necessary condition holds in our setting for $X(t) \xrightarrow{a.s.} 0$:

$$\frac{I(W;W+S)}{\mathbb{E}(W+S)} \geq a\,(1-\nu) \qquad \text{[nats/sec]}, \tag{50}$$

where $\nu > 0$ can be arbitrarily small.

We now compare (49) and (50). The rate of expansion of the state space of the continuous system in open loop is a nats per unit time, while for the discrete system is $\log a$ bits per unit time. Accordingly, (49) and (50) are parallel to each other: in the case of (50) the controller must receive at least $a\,\mathbb{E}(W+S)$ nats representing the initial state during a time interval of average length $\mathbb{E}(W+S)$. In the case of (49) the controller must receive at least $\log a/(1-\mu)$ bits representing the initial state over a time interval whose average length corresponds to the average number of trials before the first successful reception

$$(1-\mu)\sum_{k=0}^{\infty}(k+1)\mu^k = \frac{1}{1-\mu}. \tag{51}$$

B. Comparison with event triggering strategies

The works [13]–[20] use event-triggering strategies that exploit timing information for stabilization over a digital communication channel. These strategies encode information over time in a specific state-dependent fashion and use a combination of timing information and data payload to convey information used for stabilization.

Our framework, by considering the transmission of symbols from a unitary alphabet, uses only timing information for stabilization. In Theorem 4 we provide a fundamental limit on the rate at which information can be encoded in time, independent of any transmission strategy. Theorem 5 then shows that this limit can be almost achieved, in the case of exponentially distributed delays.

The work [14] shows that using event triggering it is possible to achieve stabilization with any positive transmission rate over a zero-delay digital communication channel. Indeed, for channels without delay achieving stabilization at zero rate is easy. One could for example transmit a single symbol at a time equal to any bijective mapping of $x(0)$ into a point of the non-negative reals. For example, we could transmit ♠ at time $t = \tan^{-1}(x(0))$ for $t \in [0, \pi]$. The reception of the symbol would reveal the initial state exactly, and the system could be stabilized.

The work in [15] shows that when the delay is positive, but sufficiently small, a triggering policy can still achieve stabilization with any positive transmission rate. However, as the delay increases past a critical threshold, the timing information becomes so much out-of-date that the transmission rate must begin to increase. In our case, since the capacity of our timing channel depends on the distribution of the delay, we may also expect that a large value of the capacity, corresponding to a small average delay, would allow for stabilization to occur using only timing information. Indeed, when delays are distributed exponentially, from (9) and Theorem 5 it follows that as longs as the expected value of delay is

$$\mathbb{E}(S) < \frac{1}{e\,a}, \tag{52}$$

it is possible to stabilize the system by using only timing information. On the other hand, the system is not stabilizable using only timing information if the expected value of the delay becomes larger than $(e\,a)^{-1}$.

VII. Numerical Example

We now present a numerical simulation of stabilization over the telephone signaling channel. While our analysis is for continuous-time systems, the simulation is performed in discrete time, considering the system

$$X[m] = aX[m] + U[m], \quad \text{for} \quad m \in \mathbb{N}, \tag{53}$$

where $a > 1$ so that the system is unstable.

In this case, assuming i.i.d. geometrically distributed delays $\{S_i\}$, the sufficient condition for stabilization becomes

$$C \geq \log a\,(1+\nu) \quad \text{[nats/sec]}, \tag{54}$$

where C is the timing capacity of the discrete telephone signaling channel [24]. The timing capacity is achieved in this case using i.i.d. waiting times $\{W_i\}$ that are distributed according to a mixture of a geometric and a delta distribution. This results in $\{D_i\}$ also being i.i.d. geometric [24], [26].

Assuming that a decoding operation occurs at time m using all k_m symbols received up to this time, and following the source-channel coding scheme described in the proof of Theorem 3, the controller decodes an estimate $\hat{X}_m[0]$ of the initial state and estimates the current state as

$$\hat{X}[m] = a^m \hat{X}_m[0] + \sum_{j=0}^{m-1} a^{m-1-j} U[j]. \tag{55}$$

The estimate $\hat{X}_m[0]$ corresponds to the binary representation of $X(0)$ using $\lceil k_m \mathbb{E}(D)C \rceil$ bits, provided that there is no decoding error in the tranmsission. Accordingly, in our simulation, we let $\eta > 0$ and $P_e = e^{-\eta k_m}$, and we assume that at every decoding time, with probability $(1-P_e)$ we construct a correct quantized estimate of the initial state $\hat{X}_m[0]$ using $\lceil k_m \mathbb{E}(D)C \rceil$ bits. Alternatively, with probability P_e we construct an incorrect quantized estimate. In the case of a correct estimate, we apply the asymptotically optimal control input $U[m] = -K\hat{X}[m]$, where $K > 0$ is the control gain and $\hat{X}[m]$ is obtained from (55). In the case of an incorrect estimate, the state estimate used to construct the control input can be arbitrary. We consider three cases: (i) we do not apply any control input and let the system evolve in open loop, (ii) we apply the control input using the previous estimate, (iii) we apply the opposite of the asymptotically optimal control input: $U[m] = K\hat{X}[m]$. In all cases, the control input remains fixed to its most recent value during the time required for a new estimate to be performed.

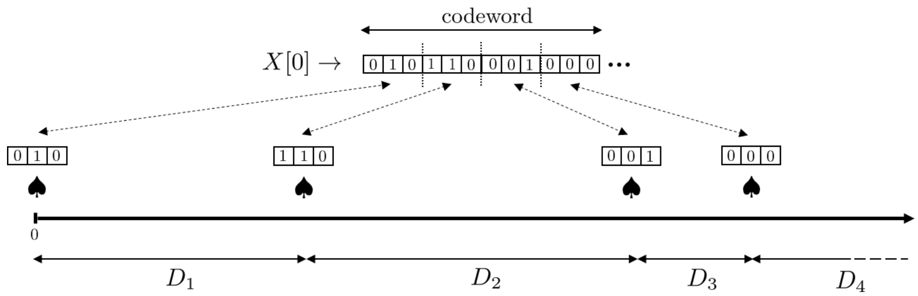

Fig. 4: Evolution of the channel used in the simulation in an error-free case. Each time ♠ is received, a new codeword is decoded using all the symbols received up to that time. The decoded codeword represents the initial state $X[0]$ with a precision that increases by $\mathbb{E}(D)C$ bits at each symbol reception. In the figure, for illustration purposes we have assumed $\mathbb{E}(D)C = 3$ bits.

Fig. 4 pictorially illustrates the evolution of our simulation in an error-free case in which the binary representation of $X[0]$ is refined by $\mathbb{E}(D)C = 3$ bits at each symbol reception.

Numerical results are depicted in Fig. 5, showing convergence of the state to zero in all cases, provided that the timing capacity is above the entropy rate of the system. In contrast, when the timing capacity is below the entropy rate, the state diverges. The plots also show the absolute value of the control input used for stabilization in the various cases.

Fig. 6 illustrates the percentage of times at which the controller successfully stabilized the plant versus the capacity of the channel in a run of 500 Monte Carlo simulations. The phase transition behavior at the critical value $C = \log a$ is clearly evident.

VIII. Conclusions

In the framework of control of dynamical systems over communication channels, it has recently been observed that event-triggering policies encoding information over time in a state-dependent fashion can exploit timing information for stabilization in addition to the information traditionally carried by data packets [13]–[20]. In a more general framework, this paper studied from an information-theoretic perspective the fundamental limitation of using *only* timing information for stabilization, independent of any transmission strategy. We showed that for stabilization of an undisturbed scalar linear system over a channel with a unitary alphabet, the timing capacity [23] should be, essentially, at least as large as the entropy rate of the system. In addition, in the case of exponentially distributed delays, we provided an almost tight sufficient condition using a coding strategy that refines the estimate of the decoded message as more and more symbols are received. Important open problems for future research include the effect of system disturbances, understanding the combination of timing information and packets with data payload, and extensions to vector systems.

Our derivation ensures that when the timing capacity is larger than the entropy rate, the estimation error does not grow unbounded, in probability, even in the presence of the random delays occurring in the timing channel. This is made possible by communicating a real-valued variable (the initial state) at an increasingly higher resolution and with vanishing probability of error. This strategy has been previously studied in [40] in the context of estimation over the binary erasure channel, rather than over the timing channel. It is also related to communication at increasing resolution over channels with feedback via posterior matching [90], [91]. The classic Horstein [92] and Schalkwijk-Kailath [93] schemes are special cases of posterior matching for the binary symmetric channel and the additive Gaussian channel respectively. The main idea in our setting is to employ a tree-structured quantizer in conjunction to a capacity-achieving timing channel codebook that grows exponentially with the tree depth, and re-compute the estimate of the real-valued variable as more and more channel symbols are received. The estimate is re-computed for a number of received symbols that depends on the channel rate and on the average delay. In contrast to posterior matching, we are not concerned with the complexity of the encoding-decoding strategy, but only with its existence. We also do not assume a specific distribution for the real value we need to communicate, and we do not use the feedback signal to perform encoding, but only to avoid queuing [23], [26]. We point out that our control strategy does not work in the presence of disturbances: in this case, one needs to track a state that depends not only on the initial condition, but also on the evolution of the disturbance. This requires to update the entire history of the system's states at each symbol reception [45], leading to a different, i.e. non-classical, coding model. Alternatively, remaining in a classical setting one could aim for less, and attempt to obtain results using weaker probabilistic notions of stability, such as the one in [6, Chapter 8].

Finally, by showing that in the case of no disturbances and exponentially distributed delay it is possible to achieve stabilization at zero data-rate only for sufficiently small average delay $\mathbb{E}(S) < (e\,a)^{-1}$, we confirmed from an information-theoretic perspective the observation made in [15] regarding the existence of a critical delay value for stabilization at zero data-rate.

Appendix

A. Proof of Theorem 2

We start by introducing a few definitions and proving some useful lemmas.

Definition 4: For any $\epsilon > 0$ and $\phi > 0$, we define the rate-distortion function of the source $\dot{X}_e = aX_e(t)$ at times $\{t_n\}$ as

$$R^{\epsilon}_{t_n}(\phi) = \inf_{\mathbb{P}\left(\hat{X}_e(t_n)|X_e(t_n)\right)} \left\{ I\left(X_e(t_n); \hat{X}_e(t_n)\right) : \right. \quad (56)$$

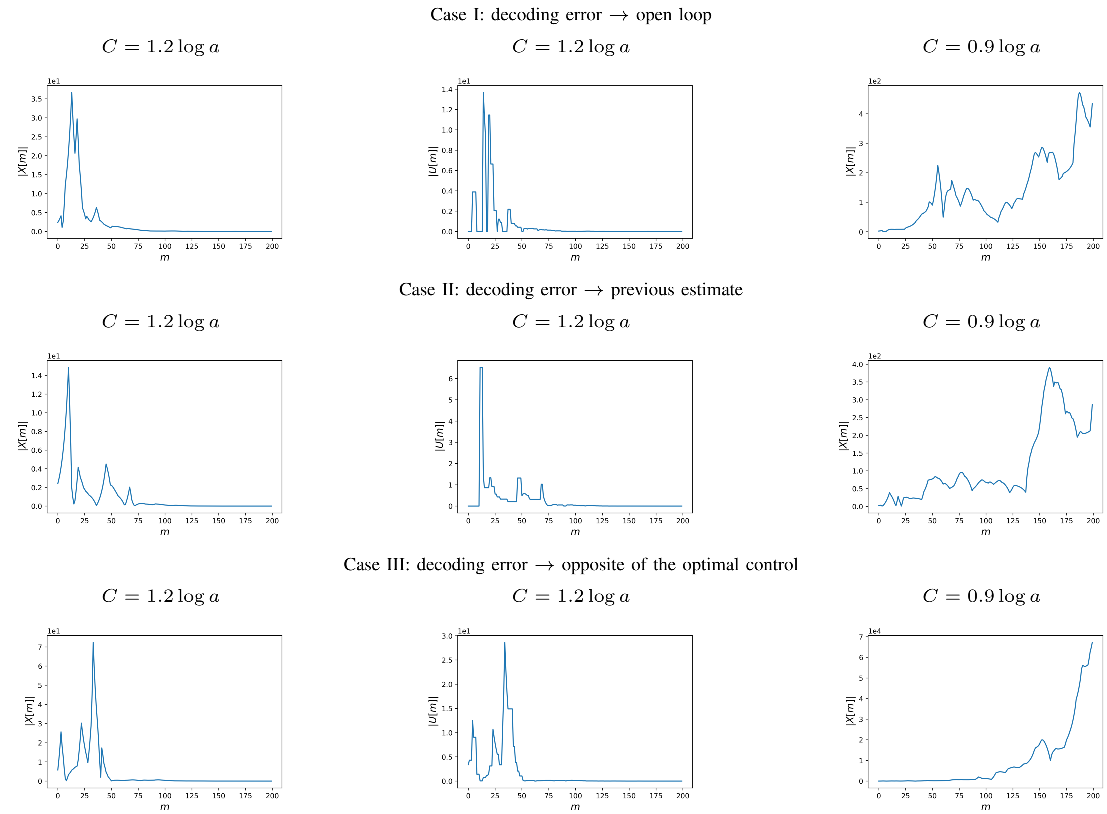

Fig. 5: Here we show the evolution of a single run of a system with different capacities for the timing channel. The first and second columns represent the absolute value of the state and control input, respectively, when the timing capacity is larger than the entropy rate of the system ($C > \log a$). The third column represents the absolute value of the state when the timing capacity is smaller than the entropy rate of the system ($C < \log a$). In the first row, in the presence of a decoding error, we do not apply any control input and let the system evolve in open-loop; in the second row, we apply the control using the previous estimate; the third row, we apply the opposite of the optimal control. The simulation parameters were chosen as follows: $a = 1.2$, $\mathbb{E}(D) = 2$, and $P_e = e^{-\eta k_m}$, where $\eta = 0.09$. For the optimal control gain we have chosen $K = 0.4$, which is optimal with respect to the (time-averaged) linear quadratic regulator (LQR) control cost $(1/200)\mathbb{E}[\sum_{m=0}^{199}(0.01X_k^2 + 0.5U_k^2) + 0.01X_{200}^2]$.

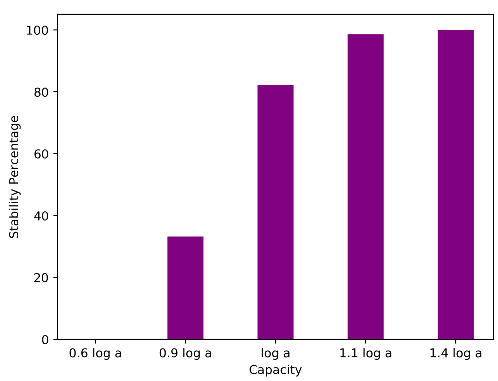

Fig. 6: Here we show the fraction of times stabilization was achieved versus the capacity of the channel across a run of 500 simulations for each value of the capacity. Successful stabilization is defined in these simulations as $|X[250]| \leq 0.05$. In the case of a decoding error, no control input is applied and we let the system evolve in open loop. The simulation parameters were chosen as follows: $a = 1.2$, $\mathbb{E}(D) = 2$, $P_e = e^{-\eta k_m}$, where $\eta = 0.09$, and the control gain is $K = 0.4$.

$$\left.\mathbb{P}\left(|X_e(t_n) - \hat{X}_e(t_n)| > \epsilon\right) \leq \phi\right\}.$$

The proof of the following lemma adapts an argument of [42] to our continuous-time setting.

Lemma 5: We have

$$R_{t_n}^{\epsilon}(\phi) \geq (1-\phi)\left[a\, t_n + h(X(0))\right] - \ln 2\epsilon - \frac{\ln 2}{2} \quad [\text{nats}]. \tag{57}$$

Proof: Let

$$\xi = \begin{cases} 0 & \text{if } |X_e(t_n) - \hat{X}_e(t_n)| \leq \epsilon \\ 1 & \text{if } |X_e(t_n) - \hat{X}_e(t_n)| > \epsilon. \end{cases} \tag{58}$$

Using the chain rule, we have

$$\begin{aligned} & I(X_e(t_n); \hat{X}_e(t_n)) \\ & = I(X_e(t_n); \xi, \hat{X}_e(t_n)) - I(X_e(t_n); \xi | \hat{X}_e(t_n)) \\ & = I(X_e(t_n); \xi, \hat{X}_e(t_n)) - H(\xi | \hat{X}_e(t_n)) \\ & \quad + H(\xi | X_e(t_n), \hat{X}_e(t_n)). \end{aligned} \tag{59}$$

Given $X(t_n)$ and $\hat{X}(t_n)$, there is no uncertainty in ξ, hence we deduce

$$\begin{aligned} & I(X_e(t_n); \hat{X}_e(t_n)) \\ & = I(X_e(t_n); \xi, \hat{X}_e(t_n)) - H(\xi | \hat{X}_e(t_n)) \\ & = h(X_e(t_n)) - h(X_e(t_n) | \xi, \hat{X}_e(t_n)) - H(\xi | \hat{X}_e(t_n)) \end{aligned}$$

$$= h(X_e(t_n)) - h(X_e(t_n)|\xi = 0, \hat{X}_e(t_n))\mathbb{P}(\xi = 0) \quad (60)$$
$$- h(X_e(t_n)|\xi = 1, \hat{X}_e(t_n))\mathbb{P}(\xi = 1) - H(\xi|\hat{X}_e(t_n)).$$

Since $H(\xi|\hat{X}_e(t_n)) \leq H(\xi) \leq \ln 2/2$ [nats], $\mathbb{P}(\xi = 0) \leq 1$, and $\mathbb{P}(\xi = 1) \leq \phi$, it then follows that

$$I(X_e(t_n); \hat{X}_e(t_n)) \geq$$
$$h\left(X_e(t_n)\right) - h\left(X_e(t_n) - \hat{X}_e(t_n)|\xi = 0, \hat{X}_e(t_n)\right)$$
$$- h\left(X_e(t_n)|\xi = 1, \hat{X}_e(t_n)\right)\phi - \frac{\ln 2}{2}. \quad (61)$$

Since conditioning reduces the entropy, we have

$$I(X_e(t_n); \hat{X}_e(t_n)) \geq h(X_e(t_n)) \quad (62)$$
$$- h(X_e(t_n) - \hat{X}_e(t_n)|\xi = 0) - h(X_e(t_n))\phi - \frac{\ln 2}{2}$$
$$= (1-\phi)h(X_e(t_n)) - h(X_e(t_n) - \hat{X}_e(t_n)|\xi = 0) - \frac{\ln 2}{2}.$$

By (58) and since the uniform distribution maximizes the differential entropy among all distributions with bounded support, we have

$$I(X_e(t_n); \hat{X}_e(t_n)) \geq (1-\phi)h(X_e(t_n)) - \ln 2\epsilon - \frac{\ln 2}{2}. \quad (63)$$

Since $X_e(t_n) = X(0)\ e^{at_n}$, we have

$$h(X_e(t_n)) = \ln e^{at_n} + h(X(0)) = at_n + h(X(0)). \quad (64)$$

Combining (63), and (64) we obtain

$$I(X_e(t_n); \hat{X}_e(t_n)) \geq (1-\phi)\left(at_n + h(X(0))\right) - \ln 2\epsilon - \frac{\ln 2}{2}. \quad (65)$$

Finally, noting that this inequality is independent of $\mathbb{P}(\hat{X}_e(t_n)|X_e(t_n))$ the result follows. ■

Remark 2: By letting $\phi = \epsilon$ in (57), we have

$$R^{\epsilon}_{t_n}(\epsilon) \geq (1-\epsilon)at_n + \epsilon', \quad (66)$$

where

$$\epsilon' = (1-\epsilon)h\left(X(0)\right) - \ln 2\epsilon - \frac{\ln 2}{2}. \quad (67)$$

For sufficiently small ϵ we have that $\epsilon' \geq 0$, and hence

$$\frac{R^{\epsilon}_{t_n}(\epsilon)}{t_n} \geq (1-\epsilon)a. \quad (68)$$

It follows that for sufficiently small ϵ the rate-distortion per unit time of the source must be at least as large as the entropy rate of the system. Since the rate-distortion represents the number of bits required to represent the state of the process up to a given fidelity, this provides an operational characterization of the entropy rate of the system. •

The proof of the following lemma follows a converse argument of [23] with some modifications due to our different setting.

Lemma 6: Under the same assumptions as in Theorem 2, if by time t_n, κ_n symbol is received by the controller, we have

$$I\left(X_e(t_n); \hat{X}_e(t_n)\right) \leq \kappa_n I(W; W+S). \quad (69)$$

Proof: We denote the transmitted message by $V \in \{1, \ldots, M\}$ and the decoded message by $U \in \{1, \ldots, M\}$. Then

$$X_e(t_n) \to V \to (D_1, \ldots, D_{\kappa_n}) \to U \to \hat{X}_e(t_n), \quad (70)$$

is a Markov chain. Therefore, using the data-processing inequality [94], we have

$$I\left(X_e(t_n); \hat{X}_e(t_n)\right) \leq I(V; U) \leq I(V; D_1, \ldots, D_{\kappa_n}). \quad (71)$$

By the chain rule for the mutual information, we have

$$I(V; D_1, \ldots, D_{\kappa_n}) = \sum_{i=1}^{\kappa_n} I(V; D_i|D^{i-1}). \quad (72)$$

Since W_i is uniquely determined by the encoder from V, using the chain rule we deduce

$$\sum_{i=1}^{\kappa_n} I(V; D_i|D^{i-1}) = \sum_{i=1}^{\kappa_n} I(V, W_i; D_i|D^{i-1}). \quad (73)$$

In addition, again using the chain rule, we have

$$\sum_{i=1}^{\kappa_n} I(V, W_i; D_i|D^{i-1}) = \sum_{i=1}^{\kappa_n} I(W_i; D_i|D^{i-1}) \quad (74)$$
$$+ \sum_{i=1}^{\kappa_n} I(V; D_i|D^{i-1}, W_i).$$

D_i is conditionally independent of V when given W_i, hence,

$$\sum_{i=1}^{\kappa_n} I(V; D_i|D^{i-1}, W_i) = 0. \quad (75)$$

Combining (73), (74), and (75) it follows that

$$\sum_{i=1}^{\kappa_n} I(V; D_i|D^{i-1}) = \sum_{i=1}^{\kappa_n} I(W_i; D_i|D^{i-1}). \quad (76)$$

Since the sequences $\{S_i\}$ and $\{W_i\}$ are i.i.d. and independent of each other, it follows that the sequence $\{D_i\}$ is also i.i.d., and we have

$$\sum_{i=1}^{\kappa_n} I(W_i; D_i|D^{i-1}) = \kappa_n I(W; D). \quad (77)$$

By combining (71), (72), (76) and (77) the result follows. ■

We are now ready to finish the proof of Theorem 2.

Proof: If $\mathbb{E}(W+S) = 0$, (16) is straightforward. Thus, for the rest of the proof, we assume $\mathbb{E}(W+S) > 0$. Using Lemma 1, as $n \to \infty$, by time t_n, given in (11), with a probability that tends to one, at most n symbols are received by the controller. In this case, using Lemma (6), it follows that

$$n\, I(W; W+S) \geq I\left(X_e(t_n); \hat{X}_e(t_n)\right). \quad (78)$$

By the assumption of the theorem, for any $\epsilon > 0$ we have

$$\lim_{n\to\infty} \mathbb{P}\left(|X_e(t_n) - \hat{X}_e(t_n)| \leq \epsilon\right) = 1. \quad (79)$$

Hence, for any $\epsilon > 0$ and any $\phi > 0$ there exist n_ϕ such that for $n \geq n_\phi$

$$\mathbb{P}\left(|X_e(t_n) - \hat{X}_e(t_n)| > \epsilon\right) \leq \phi. \tag{80}$$

Using (80), (56), and Lemma 5 it follows that for $n \geq n_\phi$

$$R^{\epsilon}_{t_n}(\phi) \geq (1-\phi)\left[at_n + h(X(0))\right] - \ln 2\epsilon - \frac{\ln 2}{2}. \tag{81}$$

By (56), we have

$$I(X_e(t_n); \hat{X}_e(t_n)) \geq R^{\epsilon}_{t_n}(\phi), \tag{82}$$

and combining (81), and (82) we obtain that for $n \geq n_\phi$

$$\frac{I\left(X_e(t_n); \hat{X}_e(t_n)\right)}{n} \geq \tag{83}$$
$$\frac{(1-\phi)at_n}{n} + \frac{(1-\phi)h(X(0)) - \ln 2\epsilon - \frac{\ln 2}{2}}{n}.$$

We now let $\phi \to 0$, so that $n \to \infty$. Using (78) we have

$$I(W; W+S) \geq a \lim_{n\to\infty} \frac{t_n}{n}. \tag{84}$$

Since, $\mathbb{E}(\mathcal{T}_n) = n\mathbb{E}(D_n)$ from (11) it follows that

$$\lim_{n\to\infty} \frac{t_n}{n} \geq (1-\nu)\,\mathbb{E}(D). \tag{85}$$

Combining (85) and (84), (16) follows. Finally, using (8) and noticing

$$\sup_{\substack{W\geq 0\\ \mathbb{E}(W)\leq\chi}} \frac{I(W;W+S)}{\mathbb{E}(S)+\chi} \geq \sup_{\substack{W\geq 0\\ \mathbb{E}(W)=\chi}} \frac{I(W;W+S)}{\mathbb{E}(S)+\chi}, \tag{86}$$

we deduce that if (16) holds then (17) holds as well. ■

B. Proof of Theorem 3

Proof: If $\mathbb{E}(S) = 0$ the timing capacity is infinite, and the result is trivial. Hence, for the rest of the proof, we assume that

$$\mathbb{E}(S+W) \geq \mathbb{E}(S) > 0, \tag{87}$$

which by (4) implies that $\mathbb{E}(\mathcal{T}_n) \to \infty$ as $n \to \infty$. As a consequence, by (12) we also have that $t'_n \to \infty$ as $n \to \infty$.

The objective is to design an encoding and decoding strategy, such that for all $\epsilon, \delta > 0$ and sufficiently large n, we have

$$\mathbb{P}(|X_e(t'_n) - \hat{X}_e(t'_n)| > \epsilon) < \delta. \tag{88}$$

We have

$$\begin{aligned}
&\mathbb{P}(|X_e(t'_n) - \hat{X}_e(t'_n)| > \epsilon) = \\
&\mathbb{P}(|X_e(t'_n) - \hat{X}_e(t'_n)| > \epsilon \mid t'_n \geq \mathcal{T}_n)\mathbb{P}(t'_n \geq \mathcal{T}_n) \\
&+ \mathbb{P}(|X_e(t'_n) - \hat{X}_e(t'_n)| > \epsilon \mid t'_n < \mathcal{T}_n)\mathbb{P}(t'_n < \mathcal{T}_n) \\
&\leq \mathbb{P}(|X_e(t'_n) - \hat{X}_e(t'_n)| > \epsilon \mid t'_n \geq \mathcal{T}_n) + \mathbb{P}(t'_n < \mathcal{T}_n), \quad (89)
\end{aligned}$$

where, using Lemma 1, the second term in the sum (89), tends to zero as $n \to \infty$. It follows that to ensure (88) it suffices to design an encoding and decoding scheme, such that for all $\epsilon, \delta > 0$ and sufficiently large n, we have that the conditional probability

$$\mathbb{P}(|X_e(t'_n) - \hat{X}_e(t'_n)| > \epsilon \mid t'_n \geq \mathcal{T}_n) < \delta. \tag{90}$$

From the open-loop equation (10), we have

$$X_e(t'_n) = e^{at'_n} X(0), \tag{91}$$

from which it follows that the decoder can construct the estimate

$$\hat{X}_e(t'_n) = e^{at'_n} \hat{X}_{t'_n}(0), \tag{92}$$

where $\hat{X}_{t'_n}(0)$ is an estimate of $X(0)$ constructed at time t'_n using all the symbols received by this time.

By (91) and (92), we now have that (90) is equivalent to

$$\mathbb{P}(|X(0) - \hat{X}_{t'_n}(0)| > \epsilon e^{-at'_n} \mid t'_n \geq \mathcal{T}_n) < \delta, \tag{93}$$

namely it suffices to design an encoding and decoding scheme to communicate the initial condition with exponentially increasing reliability in probability. Our coding procedure that achieves this objective is described next.

Source coding: We let the source coding map

$$\mathcal{Q} : [-L, L] \to \{0,1\}^{\mathbb{N}} \tag{94}$$

be an infinite tree-structured quantizer [85]. This map constructs the infinite binary sequence $\mathcal{Q}(X(0)) = \{Q_1, Q_2, \ldots\}$ as follows. $Q_1 = 0$ if $X(0)$ falls into the left-half of the interval $[-L, L]$, otherwise $Q_1 = 1$. The sub-interval where $X(0)$ falls is then divided into half and we let $Q_2 = 0$ if $X(0)$ falls into the left-half of this sub-interval, otherwise $Q_2 = 1$. The process then continues in the natural way, and Q_i is determined accordingly for all $i \geq 3$.

Using the definition of truncation operator (1), for any $n' \geq 1$ we can define

$$\mathcal{Q}_{n'} = \pi_{n'} \circ \mathcal{Q}. \tag{95}$$

It follows that $\mathcal{Q}_{n'}(X(0))$ is a binary sequence of length n' that identifies an interval of length $L/2^{n'-1}$ that contains $X(0)$. We also let

$$\mathcal{Q}^{-1}_{n'} : \{0,1\}^{n'} \to [-L, L] \tag{96}$$

be the right-inverse map of $\mathcal{Q}_{n'}$, which assigns the middle point of the last interval identified by the sequence that contains $X(0)$. It follows that for any $n' \geq 1$, this procedure achieves a quantization error

$$|X(0) - \mathcal{Q}^{-1}_{n'} \circ \mathcal{Q}_{n'}(X(0))| \leq \frac{L}{2^{n'}}. \tag{97}$$

•

Channel coding: In order to communicate the quantized initial condition over the timing channel, the truncated binary sequence $Q_{n'}(X(0))$ needs to be mapped into a channel codeword of length n.

We consider a channel codebook of n columns and M_n rows. The codeword symbols $\{w_{i,m}, i = 1, \cdots, n; m = 1 \cdots M_n\}$ are drawn i.i.d. from a distribution which is mixture of a delta function and an exponential and such that $\mathbb{P}(W_i = 0) = e^{-1}$, and $\mathbb{P}(W_i > w \mid W_i > 0) = \exp\{\frac{-w}{e\mathbb{E}(S)}\}$. By Theorem 3 of [23], if the delays $\{S_i\}$ are exponentially distributed, using a maximum likelihood decoder this construc-

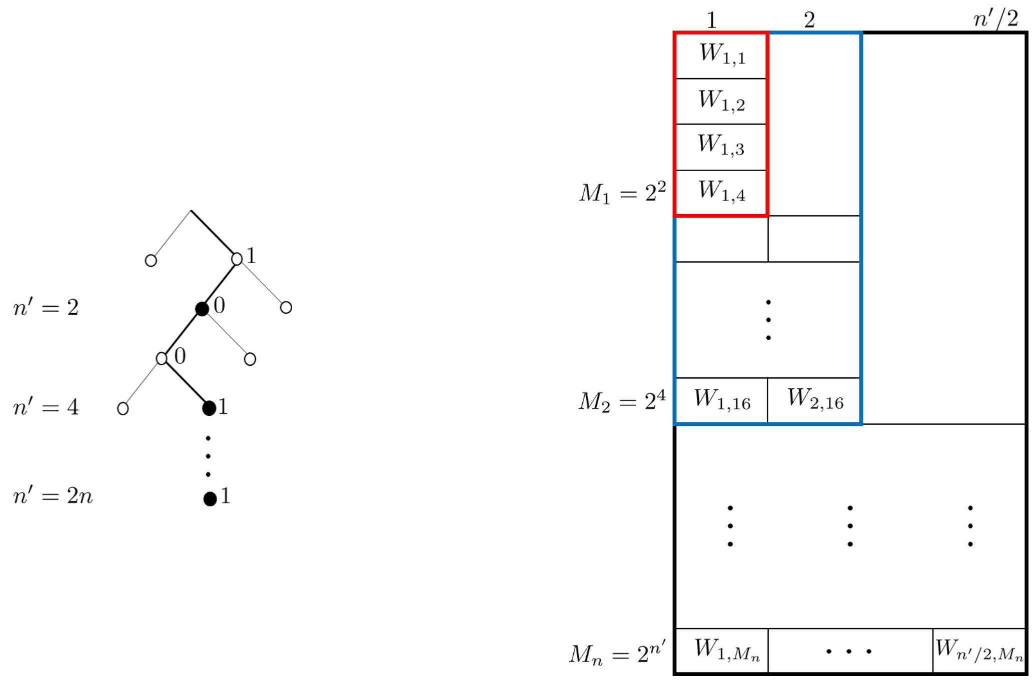

Fig. 7: Tree-structured quantizer and the corresponding codebook for $R\mathbb{E}(D) = 2$. In this case, every received channel symbol refines the source coding representation by two bits. Here the black nodes in the quantization tree at level $n' = \lceil iR\mathbb{E}(D) \rceil = 2, 4, 6, \ldots,$ are mapped into the rows of the codebook.

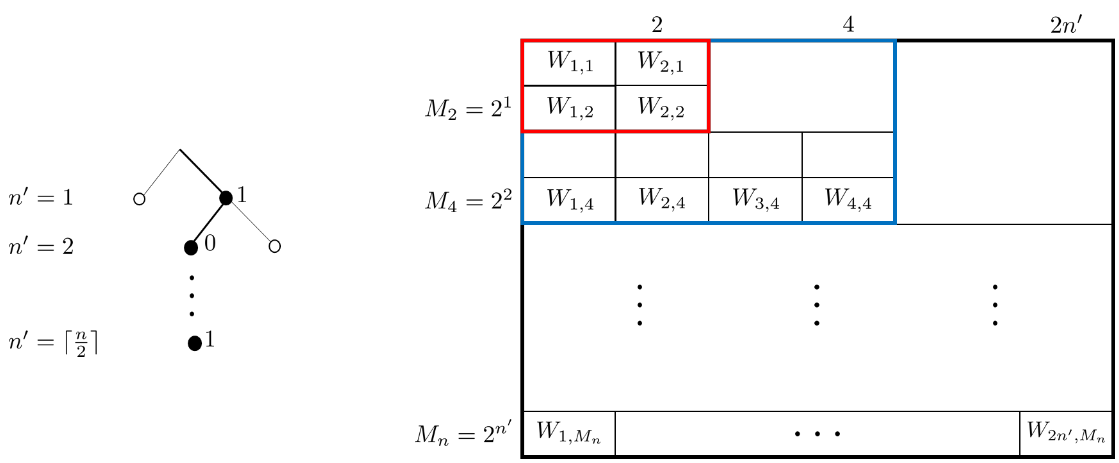

Fig. 8: Tree-structured quantizer and the corresponding codebook for $R\mathbb{E}(D) = 1/2$. In this case, every two received channel symbols refine the source coding representation by one bit.

tion achieves the timing capacity. Namely, letting

$$T_n = \mathbb{E}(\mathcal{T}_n) = n\mathbb{E}(D), \tag{98}$$

using this codebook we can achieve any rate

$$R = \lim_{n\to\infty} \frac{\log M_n}{T_n} \leq C \tag{99}$$

over the timing channel.

Next, we describe the mapping between the source coding and the channel coding constructions. •

Source-channel mapping: We first consider the direct mapping. For all $i \geq 1$, we let $n' = \lceil iR\mathbb{E}(D) \rceil$ and consider the $2^{n'}$ possible outcomes of the source coding map $\mathcal{Q}_{n'}(X(0))$. We associate them, in a one-to-one fashion, to the rows of a codebook $\Psi_{n'}$ of size $2^{n'} \times \lceil n'/R\mathbb{E}(D) \rceil$. This mapping is defined as

$$\mathcal{E}_{n'} : \{0,1\}^{n'} \to \Psi_{n'}. \tag{100}$$

By letting $i \to \infty$, the codebook becomes a double-infinite matrix Ψ_∞, and the map becomes

$$\mathcal{E} : \{0,1\}^{\mathbb{N}} \to \Psi_\infty. \tag{101}$$

Thus, as $i \to \infty$, $X(0)$ is encoded as

$$X(0) \xrightarrow{\mathcal{Q}} \{0,1\}^{\mathbb{N}} \xrightarrow{\mathcal{E}} \Psi_\infty. \tag{102}$$

We now consider the inverse mapping. Since the elements of $\Psi_{n'}$ are drawn independently from a continuous distribution, with probability one, no two rows of the codebook are equal to each other, so for any $i \geq 1$ and number of received symbols $n = \lceil i/R\mathbb{E}(D) \rceil$ we define

$$\mathcal{E}_{n'}^{-1} : \Psi_{n'} \to \{0,1\}^{n'}, \tag{103}$$

where $n' = \lceil nR\mathbb{E}(D) \rceil$. This map associates to every row in the codebook a corresponding node in the quantization tree at level n'.

Figures 7 and 8 show the constructions described above for the cases $R\mathbb{E}(D) = 2$ and $R\mathbb{E}(D) = 0.5$, respectively. In Fig. 7, the nodes in the quantization tree at level $n' = \lceil iR\mathbb{E}(D) \rceil = 2, 4, 6, \ldots,$ are mapped into the rows of a table of $M_n = 2^2, 2^4, 2^6, \ldots$ rows and $n = 1, 2, 3 \ldots$ columns. Conversely, the rows in each table are mapped into the corresponding nodes in the tree. In Fig. 8, the nodes in the quantization tree at level $n' = \lceil iR\mathbb{E}(D) \rceil = 1, 2, 3, \ldots,$ are mapped into the rows of a table of $M_n = 2, 2^2, 2^3, \ldots$ rows and $n = 2, 4, 6, \ldots$ columns. Conversely, the rows in each table are mapped into the corresponding nodes in the tree.

Next, we describe how the encoding and decoding operations are performed using these maps and how transmission occurs over the channel. •

One-time encoding: The encoding of the initial state $X(0)$ occurs at the sensor in one-shot and then the corresponding symbols are transmitted over the channel, one by one. Given $X(0)$, the source encoder computes $Q(X(0))$ according to the source coding map (94) and the channel encoder picks the corresponding codeword $\mathcal{E}(Q(X(0)))$ from the doubly-infinite codebook according to the map (101). This codeword is an infinite sequence of real numbers, which also corresponds to a leaf at infinite depth in the quantization tree. Then, the encoder starts transmitting the real numbers of the codeword one by one, where each real number corresponds to a holding time, and proceeds in this way forever. According to the source-channel mapping described above, transmitting $n = \lceil n'/R\mathbb{E}(D) \rceil$ symbols using this scheme corresponds to transmitting, for all $i \geq 1$, $n' = \lceil iR\mathbb{E}(D) \rceil$ source bits, encoded into a codeword $\mathcal{E}_{n'}(Q_{n'}(X(0)))$, picked from a truncated codebook of $2^{n'}$ rows and n columns. •

Anytime Decoding: The decoding of the initial state $X(0)$ occurs at the controller in an anytime fashion, refining the estimate of $X(0)$ as more and more symbols are received.

For all $i \geq 1$ the decoder updates its guess for the value of $X(0)$ any time the number of symbols received equals $n = \lceil i/R\mathbb{E}(D) \rceil$. Assuming a decoding operation occurs after n symbols have been received, the decoder picks the maximum likelihood codeword from a truncated codebook of size $M_n \times n$ and by inverse mapping, it finds the corresponding node in the tree. It follows that at the nth random reception time $\mathcal{T}_n$, the decoder utilizes the inter-reception times of all n symbols received up to this time to construct the estimate $\hat{X}_{\mathcal{T}_n}(0)$. First, a maximum likelihood decoder $\mathcal{D}_n$ is employed to map the inter-reception times $(D_1, \ldots, D_n)$ to an element of $\Psi_{n'}$. This element is then mapped to a binary sequence of length n' using $\mathcal{E}_{n'}^{-1}$. Finally, $\mathcal{Q}_{n'}^{-1}$ is used to construct $\hat{X}_{\mathcal{T}_n}(0)$. It follows that at the nth reception time where decoding occurs, we have

$$(D_1, \ldots, D_n) \xrightarrow{\mathcal{D}_n} \Psi_{n'} \xrightarrow{\mathcal{E}_{n'}^{-1}} \{0,1\}^{n'} \xrightarrow{\mathcal{Q}_{n'}^{-1}} [-L, L], \tag{104}$$

and we let

$$\hat{X}_{\mathcal{T}_n}(0) = \mathcal{Q}_{n'}^{-1}\left(\mathcal{E}_{n'}^{-1}\left(\mathcal{D}_n(D_1, \ldots, D_n)\right)\right). \tag{105}$$

Thus, as $n \to \infty$ the final decoding process becomes

$$(D_1, D_n, \ldots) \xrightarrow{\mathcal{D}} \Psi_\infty \xrightarrow{\mathcal{E}^{-1}} \{0,1\}^{\mathbb{N}} \xrightarrow{\mathcal{Q}^{-1}} [-L, L]. \tag{106}$$

•

To conclude the proof, we now show that if $C \geq (1+\nu)a$, then it is possible to perform the above encoding and decoding operations with an arbitrarily small probability of error while using a codebook so large that it can accommodate a quantization error at most $L/2^{n'} < \epsilon e^{-at'_n}$.

Since the channel coding scheme achieves the timing capacity, we have that for any $R \leq C$, as $n \to \infty$ the maximum likelihood decoder selects the correct transmitted codeword with arbitrarily high probability. It follows that for any $\delta > 0$ and n sufficiently large, we have with probability at least $(1-\delta)$ that

$$\mathcal{Q}_{n'}(X(0)) = \mathcal{E}_{n'}^{-1}\left(\mathcal{D}_n(D_1, \ldots, D_n)\right), \tag{107}$$

and then by (97) we have

$$|X(0) - \hat{X}_{\mathcal{T}_n}(0)| \leq \frac{L}{2^{n'}}. \tag{108}$$

We now consider a sequence of estimation times $\{t'_n\}$ satisfying (12) and let the estimate at time $t'_n \geq \mathcal{T}_n$ in (93) be $\hat{X}_{t'_n}(0) = \hat{X}_{\mathcal{T}_n}(0)$. By (108) we have that the sufficient condition for estimation reduces to

$$\frac{L}{2^{n'}} \leq \epsilon e^{-at'_n}, \tag{109}$$

which means having the size of the codebook M_n be such that

$$\frac{L}{M_n} \leq \epsilon e^{-at'_n}, \tag{110}$$

or equivalently

$$\frac{\log M_n - \log L + \log \epsilon}{t'_n} \geq a. \tag{111}$$

Using (98), we have

$$\begin{aligned}\frac{\log M_n - \log L + \log \epsilon}{t'_n} &= \frac{\log M_n - \log L + \log \epsilon}{T_n} \cdot \frac{T_n}{t'_n} \\ &= \frac{\log M_n - \log L + \log \epsilon}{T_n} \\ &\quad \cdot \frac{\mathbb{E}(\mathcal{T}_n)}{t'_n}.\end{aligned} \tag{112}$$

Taking the limit for $n \to \infty$, we have

$$\lim_{n\to\infty} \frac{\log M_n - \log L + \log \epsilon}{T_n} \cdot \frac{\mathbb{E}(\mathcal{T}_n)}{t'_n} \geq R \cdot \frac{1}{1+\nu}. \tag{113}$$

It follows that as $n \to \infty$ the sufficient condition (111) can be expressed in terms of the rate as

$$R \geq (1+\nu)a. \tag{114}$$

It follows that the rate must satisfy

$$C \geq R \geq (1+\nu)a \tag{115}$$

and since $C \geq (1+\nu)a$, the proof is complete. ■

REFERENCES

[1] M. J. Khojasteh, M. Franceschetti, and G. Ranade, "Stabilizing a linear system using phone calls," in *18th Eur. Cont. Conf. (ECC)*. IEEE, 2019, pp. 2856–2861.

[2] ——, "Estimating a linear process using phone calls," in *IEEE Conf. Decis. and Cont. (CDC)*, 2018, pp. 127–131.

[3] K.-D. Kim and P. R. Kumar, "Cyber–physical systems: A perspective at the centennial," *Proc. IEEE*, vol. 100 (Special Centennial Issue), pp. 1287–1308, 2012.

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

[5] S. Yüksel and T. Başar, *Stochastic Networked Control Systems: Stabilization and Optimization under Information Constraints*. Springer Science & Business Media, 2013.

[6] A. S. Matveev and A. V. Savkin, *Estimation and control over communication networks*. Springer Science & Business Media, 2009.

[7] M. Franceschetti and P. Minero, "Elements of information theory for networked control systems," in *Information and Control in Networks*. Springer, 2014, pp. 3–37.

[8] 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.

[9] N. C. Martins, M. A. Dahleh, and N. Elia, "Feedback stabilization of uncertain systems in the presence of a direct link," *IEEE Trans. Autom. Control*, vol. 51, no. 3, pp. 438–447, 2006.

[10] D. F. Delchamps, "Stabilizing a linear system with quantized state feedback," *IEEE Trans. Autom. Control*, vol. 35, no. 8, pp. 916–924, 1990.

[11] W. S. Wong and R. W. Brockett, "Systems with finite communication bandwidth constraints. II. stabilization with limited information feedback," *IEEE Trans. Autom. Control*, vol. 44, no. 5, pp. 1049–1053, 1999.

[12] J. Baillieul, "Feedback designs for controlling device arrays with communication channel bandwidth constraints," in *ARO Workshop on Smart Structures, Pennsylvania State Univ*, 1999, pp. 16–18.

[13] M. J. Khojasteh, M. Hedayatpour, J. Cortés, and M. Franceschetti, "Exploiting timing information in event-triggered stabilization of linear systems with disturbances," *IEEE Transactions on Control of Network Systems*, vol. 8, no. 1, pp. 15–27, 2020.

[14] E. Kofman and J. H. Braslavsky, "Level crossing sampling in feedback stabilization under data-rate constraints," in *45st IEEE Conf. Decision and Cont. (CDC)*, 2006, pp. 4423–4428.

[15] M. J. Khojasteh, P. Tallapragada, J. Cortés, and M. Franceschetti, "The value of timing information in event-triggered control," *IEEE Transactions on Automatic Control*, vol. 65, no. 3, pp. 925–940, 2019.

[16] M. J. Khojasteh, P. Tallapragada, J. Cortés, and M. Franceschetti, "Time-triggering versus event-triggering control over communication channels," in *56th IEEE Conf. Decision and Cont. (CDC)*, Dec 2017, pp. 5432–5437.

[17] S. Linsenmayer, R. Blind, and F. Allgöwer, "Delay-dependent data rate bounds for containability of scalar systems," *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 7875–7880, 2017.

[18] H. Yildiz, Y. Su, A. Khina, and B. Hassibi, "Event-triggered stochastic control via constrained quantization," in *Data Comp. Conf.* IEEE, 2019, pp. 612–612.

[19] M. J. Khojasteh, M. Hedayatpour, and M. Franceschetti, "Theory and implementation of event-triggered stabilization over digital channels," in *IEEE Conf. Decis. and Cont. (CDC)*, 2019, pp. 4183–4188.

[20] N. Guo and V. Kostina, "Optimal causal rate-constrained sampling of the wiener process," in *57th Ann. Allerton Conf. on Comm., Cont., and Comp. (Allerton)*. IEEE, 2019, pp. 1090–1097.

[21] G. Ranade and A. Sahai, "Non-coherence in estimation and control," in *51st Communication, Annu. Allerton Conf. Commun., Control, Comput.* IEEE, 2013, pp. 189–196.

[22] ——, "Control capacity," *IEEE Transactions on Information Theory*, vol. 65, no. 1, pp. 235–254, 2018.

[23] V. Anantharam and S. Verdú, "Bits through queues," *IEEE Trans. Inf. Theory*, vol. 42, no. 1, pp. 4–18, 1996.

[24] A. S. Bedekar and M. Azizoglu, "The information-theoretic capacity of discrete-time queues," *IEEE Trans. Inf. Theory*, vol. 44, no. 2, pp. 446–461, 1998.

[25] E. Arikan, "On the reliability exponent of the exponential timing channel," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1681–1689, 2002.

[26] R. Sundaresan and S. Verdú, "Robust decoding for timing channels," *IEEE Trans. Inf. Theory*, vol. 46, no. 2, pp. 405–419, 2000.

[27] C. Rose and I. S. Mian, "Inscribed matter communication: Part I," *IEEE Transactions on Molecular, Biological and Multi-Scale Communications*, vol. 2, no. 2, pp. 209–227, 2016.

[28] A. Gohari, M. Mirmohseni, and M. Nasiri-Kenari, "Information theory of molecular communication: Directions and challenges," *IEEE Transactions on Molecular, Biological and Multi-Scale Communications*, vol. 2, no. 2, pp. 120–142, 2016.

[29] T. J. Riedl, T. P. Coleman, and A. C. Singer, "Finite block-length achievable rates for queuing timing channels," in *Information Theory Workshop (ITW), 2011 IEEE*. IEEE, 2011, pp. 200–204.

[30] A. B. Wagner and V. Anantharam, "Zero-rate reliability of the exponential-server timing channel," *IEEE Trans. Inf. Theory*, vol. 51, no. 2, pp. 447–465, 2005.

[31] M. Tavan, R. D. Yates, and W. U. Bajwa, "Bits through bufferless queues," in *51st Annu. Allerton Conf. Commun., Control, Comput.*, Oct 2013, pp. 755–762.

[32] B. Prabhakar and R. Gallager, "Entropy and the timing capacity of discrete queues," *IEEE Trans. Inf. Theory*, vol. 49, no. 2, pp. 357–370, 2003.

[33] L. Aptel and A. Tchamkerten, "Feedback increases the capacity of queues," in *2018 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2018, pp. 1116–1120.

[34] J. Giles and B. Hajek, "An information-theoretic and game-theoretic study of timing channels," *IEEE Trans. Inf. Theory*, vol. 48, no. 9, pp. 2455–2477, 2002.

[35] T. P. Coleman and M. Raginsky, "Mutual information saddle points in channels of exponential family type," in *Information Theory Proceedings (ISIT), 2010 IEEE International Symposium on*. IEEE, 2010, pp. 1355–1359.

[36] X. Liu and R. Srikant, "The timing capacity of single-server queues with multiple flows," *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, 2004.

[37] S. H. Sellke, C.-C. Wang, N. Shroff, and S. Bagchi, "Capacity bounds on timing channels with bounded service times," in *2007 IEEE International Symposium on Information Theory*. IEEE, 2007, pp. 981–985.

[38] G. C. Ferrante, T. Q. Quek, and M. Z. Win, "An achievable rate region for superposed timing channels," in *2016 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2016, pp. 365–369.

[39] ——, "Timing capacity of queues with random arrival and modified service times," in *2016 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2016, pp. 370–374.

[40] G. Como, F. Fagnani, and S. Zampieri, "Anytime reliable transmission of real-valued information through digital noisy channels," *SIAM J. Control Optimiz.*, vol. 48, no. 6, pp. 3903–3924, 2010.

[41] S. Tatikonda and S. Mitter, "Control under communication constraints," *IEEE Trans. Autom. Control*, vol. 49, no. 7, pp. 1056–1068, 2004.

[42] ——, "Control over noisy channels," *IEEE Trans. Autom. Control*, vol. 49, no. 7, pp. 1196–1201, 2004.

[43] A. S. Matveev and A. V. Savkin, "An analogue of shannon information theory for detection and stabilization via noisy discrete communication channels," *SIAM J. Control Optimiz.*, vol. 46, no. 4, pp. 1323–1367, 2007.

[44] ——, "Shannon zero error capacity in the problems of state estimation and stabilization via noisy communication channels," *International Journal of Control*, vol. 80, no. 2, pp. 241–255, 2007.

[45] 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, 2006.

[46] G. N. Nair and R. J. Evans, "Stabilizability of stochastic linear systems with finite feedback data rates," *SIAM J. Control Optimiz.*, vol. 43, no. 2, pp. 413–436, 2004.

[47] 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, p. 243, 2009.

[48] P. Minero, L. Coviello, and M. Franceschetti, "Stabilization over Markov feedback channels: the general case," *IEEE Trans. Autom. Control*, vol. 58, no. 2, pp. 349–362, 2013.

[49] P. Minero and M. Franceschetti, "Anytime capacity of a class of markov channels," *IEEE Trans. Autom. Control*, vol. 62, no. 3, pp. 1356–1367, 2017.

[50] S. Fang, J. Chen, and I. Hideaki, *Towards integrating control and information theories*. Springer, 2017.

[51] C. Kawan, "Invariance entropy for deterministic control systems," *Lecture notes in mathematics*, vol. 2089, 2013.

[52] F. Colonius, U. Helmke, J. Jordan, C. Kawan, R. Sailer, and F. Wirth, "Analysis of networked systems," in *Control Theory of Digitally Networked Dynamic Systems*. Springer, 2014, pp. 31–79.

[53] J. Hespanha, A. Ortega, and L. Vasudevan, "Towards the control of linear systems with minimum bit-rate," in *Proc. 15th Int. Symp. on Mathematical Theory of Networks and Systems (MTNS)*, 2002.

[54] G. Nair, "A non-stochastic information theory for communication and state estimation," *IEEE Trans. Autom. Control*, vol. 58, pp. 1497–1510, 2013.

[55] 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, 2009.

[56] 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, 2016.

[57] A. Khina, W. Halbawi, and B. Hassibi, "(Almost) practical tree codes," in *Information Theory (ISIT), 2016 IEEE International Symposium on*. IEEE, 2016, pp. 2404–2408.

[58] G. N. Nair, R. J. Evans, I. M. Mareels, and W. Moran, "Topological feedback entropy and nonlinear stabilization," *IEEE Trans. Autom. Control*, vol. 49, no. 9, pp. 1585–1597, 2004.

[59] D. Liberzon and J. P. Hespanha, "Stabilization of nonlinear systems with limited information feedback," *IEEE Trans. Autom. Control*, vol. 50, no. 6, pp. 910–915, 2005.

[60] C. De Persis, "n-bit stabilization of n-dimensional nonlinear systems in feedforward form," *IEEE Trans. Autom. Control*, vol. 50, no. 3, pp. 299–311, 2005.

[61] D. Liberzon, "Finite data-rate feedback stabilization of switched and hybrid linear systems," *Automatica*, vol. 50, no. 2, pp. 409–420, 2014.

[62] G. Yang and D. Liberzon, "Feedback stabilization of switched linear systems with unknown disturbances under data-rate constraints," *IEEE Trans. Autom. Control*, vol. 63, no. 7, pp. 2107–2122, 2018.

[63] V. Sanjaroon, A. Farhadi, A. S. Motahari, and B. H. Khalaj, "Estimation of nonlinear dynamic systems over communication channels," *IEEE Trans. Autom. Control*, vol. 63, no. 9, pp. 3024–3031, 2018.

[64] V. Kostina and B. Hassibi, "Rate-cost tradeoffs in control," *IEEE Trans. Autom. Control*, vol. 64, no. 11, pp. 4525–4540, 2019.

[65] P. Tabuada, "Event-triggered real-time scheduling of stabilizing control tasks," *IEEE Trans. Autom. Control*, vol. 52, no. 9, pp. 1680–1685, 2007.

[66] W. P. M. H. Heemels, K. H. Johansson, and P. Tabuada, "An introduction to event-triggered and self-triggered control," in *51st IEEE Conf. Decision and Cont. (CDC)*, 2012, pp. 3270–3285.

[67] K. J. Astrom and B. M. Bernhardsson, "Comparison of riemann and lebesgue sampling for first order stochastic systems," in *41st IEEE Conf. Decision and Cont. (CDC)*, vol. 2, 2002, pp. 2011–2016.

[68] X. Wang and M. D. Lemmon, "Event-triggering in distributed networked control systems," *IEEE Trans. Autom. Control*, vol. 56, no. 3, p. 586, 2011.

[69] D. V. Dimarogonas, E. Frazzoli, and K. H. Johansson, "Distributed event-triggered control for multi-agent systems," *IEEE Trans. Autom. Control*, vol. 57, no. 5, pp. 1291–1297, 2012.

[70] B. A. Khashooei, D. J. Antunes, and W. Heemels, "A consistent threshold-based policy for event-triggered control," *IEEE Control Systems Letters*, vol. 2, no. 3, pp. 447–452, 2018.

[71] W. H. Heemels, M. Donkers, and A. R. Teel, "Periodic event-triggered control for linear systems," *IEEE Trans. Autom. Control*, vol. 58, no. 4, pp. 847–861, 2013.

[72] L. Li, X. Wang, and M. Lemmon, "Stabilizing bit-rates in quantized event triggered control systems," in *Proceedings of the 15th ACM international conference on Hybrid Systems: Computation and Control*. ACM, 2012, pp. 245–254.

[73] B. Demirel, V. Gupta, D. E. Quevedo, and M. Johansson, "On the trade-off between communication and control cost in event-triggered dead-beat control," *IEEE Trans. Autom. Control*, vol. 62, no. 6, pp. 2973–2980, 2017.

[74] D. E. Quevedo, V. Gupta, W.-J. Ma, and S. Yüksel, "Stochastic stability of event-triggered anytime control," *IEEE Trans. Autom. Control*, vol. 59, no. 12, pp. 3373–3379, 2014.

[75] L. Lindemann, D. Maity, J. S. Baras, and D. V. Dimarogonas, "Event-triggered feedback control for signal temporal logic tasks," in *IEEE Conf. Decision and Cont. (CDC)*, 2018, pp. 146–151.

[76] A. Girard, "Dynamic triggering mechanisms for event-triggered control," *IEEE Trans. Autom. Control*, vol. 60, no. 7, pp. 1992–1997, 2015.

[77] A. Seuret, C. Prieur, S. Tarbouriech, and L. Zaccarian, "LQ-based event-triggered controller co-design for saturated linear systems," *Automatica*, vol. 74, pp. 47–54, 2016.

[78] 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, 2016.

[79] J. Pearson, J. P. Hespanha, and D. Liberzon, "Control with minimal cost-per-symbol encoding and quasi-optimality of event-based encoders," *IEEE Trans. Autom. Control*, vol. 62, no. 5, pp. 2286–2301, 2017.

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

[81] K. You and L. Xie, "Minimum data rate for mean square stabilization of discrete LTI systems over lossy channels," *IEEE Tran. Auto. Cont.*, vol. 55, no. 10, pp. 2373–2378, 2010.

[82] S. Yuksel and S. P. Meyn, "Random-time, state-dependent stochastic drift for Markov chains and application to stochastic stabilization over erasure channels," *IEEE Tran. on Auto. Cont.*, vol. 58, no. 1, pp. 47–59, 2012.

[83] V. Gupta, A. F. Dana, J. P. Hespanha, R. M. Murray, and B. Hassibi, "Data transmission over networks for estimation and control," *IEEE Tran. on Auto. Cont.*, vol. 54, no. 8, pp. 1807–1819, 2009.

[84] A. Khina, V. Kostina, A. Khisti, and B. Hassibi, "Tracking and control of Gauss–Markov processes over packet-drop channels with acknowledgments," *IEEE Tran. on Cont. of Net. Sys.*, vol. 6, no. 2, pp. 549–560, 2018.

[85] A. Gersho and R. M. Gray, *Vector quantization and signal compression*. Springer Science & Business Media, 2012, vol. 159.

[86] D. Liberzon and S. Mitra, "Entropy and minimal bit rates for state estimation and model detection," *IEEE Trans. Autom. Control*, 2017.

[87] F. Colonius and C. Kawan, "Invariance entropy for control systems," *SIAM J. Control Optimiz.*, vol. 48, no. 3, pp. 1701–1721, 2009.

[88] F. Colonius, "Minimal bit rates and entropy for exponential stabilization," *SIAM J. Control Optimiz.*, vol. 50, no. 5, pp. 2988–3010, 2012.

[89] M. Rungger and M. Zamani, "On the invariance feedback entropy of linear perturbed control systems," in *56th IEEE Conf. Decision and Cont. (CDC)*, 2017, pp. 3998–4003.

[90] O. Shayevitz and M. Feder, "Optimal feedback communication via posterior matching," *IEEE Trans. Inf. Theory*, vol. 57, no. 3, pp. 1186–1222, 2011.

[91] M. Naghshvar, T. Javidi, and M. Wigger, "Extrinsic jensen–shannon divergence: Applications to variable-length coding," *IEEE Trans. Inf. Theory*, vol. 61, no. 4, pp. 2148–2164, 2015.

[92] M. Horstein, "Sequential transmission using noiseless feedback," *IEEE Trans. Inf. Theory*, vol. 9, no. 3, pp. 136–143, 1963.

[93] J. Schalkwijk and T. Kailath, "A coding scheme for additive noise channels with feedback–i: No bandwidth constraint," *IEEE Trans. Inf. Theory*, vol. 12, no. 2, pp. 172–182, 1966.

[94] T. M. Cover and J. A. Thomas, *Elements of information theory*. John Wiley & Sons, 2012.