The Many Facets of Information in Networked Estimation and Control

Massimo Franceschetti,[1] Mohammad Javad Khojasteh,[2] and Moe Z. Win[3]

[1]Department of Electrical and Computer Engineering, University of California San Diego, La Jolla, CA, 92093, email: massimo@ece.ucsd.edu
[2]Wireless Information and Network Sciences Laboratory, Massachusetts Institute of Technology, Cambridge, MA, 02139, email: mkhojast@mit.edu
[3]Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, MA, 02139, email: moewin@mit.edu

Published in the Annual Review of Control, Robotics, and Autonomous Systems.

This unofficial version compiled on: November 4, 2022

Keywords

Networked Control Systems, Cyber-Physical Systems, Control Under Communication Constraints, Decentralized Inference, Timing Information, Event Triggering

Abstract

Networked control systems, where feedback loops are closed over communication networks, arise in several domains including smart energy grids, autonomous driving, unmanned aerial vehicles, and many industrial and robotic systems active in service, production, agriculture, and smart homes and cities. In these settings, the two main layers of the system, control and communication, strongly affect each other's performance, and they also reveal the interaction between a cyber-system component, represented by information-based computing and communication technologies, and a physical-system component, represented by the environment that needs to be controlled. The information access and distribution constraints required to achieve reliable state estimation and stabilization in networked control systems have been intensively studied over the course of roughly two decades. This article reviews some of the cornerstone results in this area, draws a map for what we have learned over these years, and describes the new challenges that we will face in the future. Rather than simply listing different results, we present them in a coherent fashion using a uniform notation, and we also put them in context, highlighting both their theoretical insights and their practical significance. Particular attention given to recent developments related to decentralized estimation in distributed sensing and communication systems and the information-theoretic value of event timing in the context of networked control.

Contents

1. Introduction

In this paper we are concerned with *Networked Control Systems* (NCS) composed by a network of interacting elements, including sensors, actuators, computing and communication devices connected in closed loop, with the objective of performing tasks that require interaction with the physical world. A schematic representation of such systems is depicted in Figure 1. Examples include autonomous and remotely controlled robots, unmanned aerial vehicles (UAVs), autonomous vehicles (AV), and several industrial and consumer control systems. One of their key features is the interaction between a *Cyber* component, namely a networked computing and communication infrastructure composed by controllers, encoders and decoders, and a *Physical* component, namely a physical plant —which occurs through distributed sensors and actuators. For this reason, they are also referred to as *Cyber-Physical Systems* (CPS) (1). In this framework, two fundamental questions that we wish to address are: 1. What is the minimum amount of information transfer among the different components of the system that is needed to keep the overall system stable? 2. How can we design encoding, decoding, and control policies that best exploit the available information flow to reach stability? As we shall see, these questions are closely related to the ability of performing decentralized estimation through distributed sensing and communication, since achieving a reliable estimate of the state is key to determine the correct control action, and many of the results on control and stabilization also have counterparts in this setting. Deriving them requires the development of a new information-theoretic paradigm in which the dynamical system aspects of the problem pose strong constraints on the communication aspects. These constraints are typically ignored in classical information theory, but must be taken into account in the context of control. Namely, in our setting the "utility" of the received information at any given time, rather than merely the rate at which the information flows, determines the ability to perform control. This utility is associated to system-dependent parameters, as information quickly becomes outdated and thus unusable for control. This new paradigm also leads to the realization that in NCS information can have many facets. For example, the event corresponding to the availability of new data, along with the data itself, can encode information that can be useful for control. In some cases, this allows to perform control with a remarkably low data-rate, which can be counter-

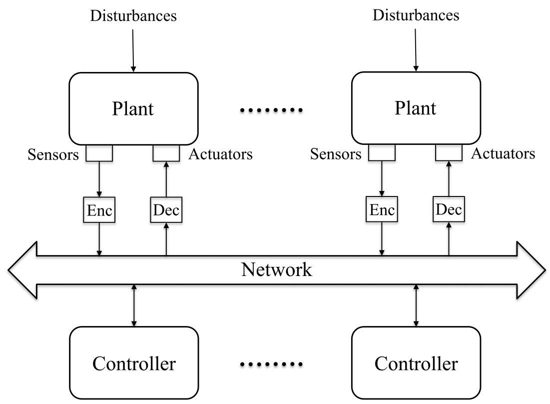

Figure 1

Distributed communication and control system.

intuitive from a more classical information-theoretic perspective. The main theme of this paper is to show how these considerations can be cast into a rigorous theory that lays the foundations for the developments of next-generation NCS. Related works investigating information decay over time include the "age of information" paradigm (2, 3, 4, 5). However, as pointed out in (6, 7, 8), while these works are relevant for some specific applications, e.g. news feeds, they do not consider that when information is used in the context of control, a relevant age metric should be related to system parameters. We do not wish to review here all the alternative measures of information utility that have been proposed in the literature, but we focus on the different aspects in which information can be encoded to be useful for control and identify the communication constraint needed for stability in NCS.

In the last two decades, the research community has studied information constraints in NCS by developing several mathematical abstractions of system components and interconnections. The results that have been obtained shed light on the behavior of real systems and provide guidelines to develop effective control policies. Surveys of this literature appear in (9, 10, 11, 12, 13) and in the books (14, 15, 16, 17). We extend these reviews, focusing on data-rate requirements for stabilization, and in particular on recent advancements and insights obtained through the study of event-triggered control policies and distributed inference systems. One key point that we wish to put forward is that is that the information flow in feedback systems is not only associated to data flowing through the links connecting the different devices, but it is more generally encoded in "events" that occur over time. This new point of view leads to several extensions of classic results and to a broader perspective on the information constraints associated to control systems. Another point is that decentralized inference is an important building block for performing control in NCS and information constraints can be derived in this case for both single plant and multi-node networks. While in this paper we focus primarily on system stabilization, we point out that related studies of optimal control under communication constraints have also been performed (18, 19, 20, 6, 7). Recently, the tradeoffs between rate and linear–quadratic regulator (LQR) cost for periodic control schemes have been evaluated in (21).

Notation: Random variables are displayed in sans serif, upright fonts; their realizations in serif, italic fonts. Vectors and matrices are denoted by bold lowercase and uppercase letters, respectively. For example, a random variable and its realization are denoted by x

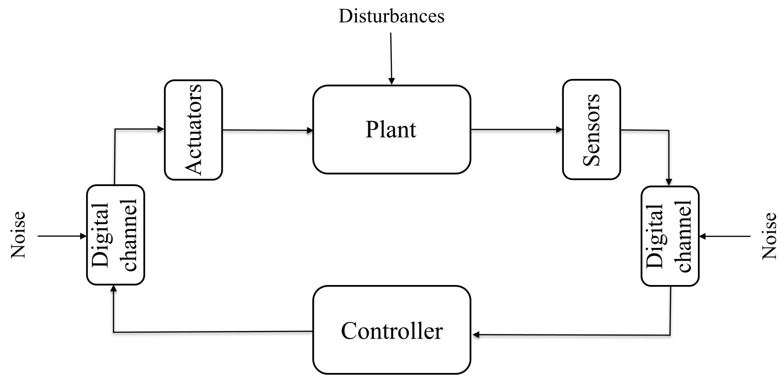

Figure 2

High level abstraction of a CPS from a communication perspective.

and x; a random vector and its realization are denoted by x and $\boldsymbol{x}$; a random matrix and its realization are denoted by X and $\boldsymbol{X}$, respectively. The expectation of x is denoted by $\mathbb{E}\{\mathsf{x}\}$. Given a discrete-time stochastic process $\{\mathsf{x}_t\}_{t\geqslant 0}$, the notation $\mathsf{x}_{s:t}$ represents the vertical concatenation of x_τ for integers $s \leqslant \tau \leqslant t$. Logarithms of a positive number x with base 2 is denoted by $\log x$. The Euclidean norm of vector $\boldsymbol{x}$ is denoted by $\|\boldsymbol{x}\|$. $\mathcal{N}(a, b)$ denotes a Gaussian distribution with mean a and variance b.

2. Data-rate theorem

We start considering a simple high-level abstraction of a single-plant network as the block diagram as depicted in Figure 2, and later extend the treatment to the multi-node case. A dynamical system evolves over time according to deterministic state equations, affected by stochastic disturbances. Sensors monitor the system's output and their readings are encoded and sent through a digital communication channel to a controller, whose action is fed back to the actuators through another digital communication channel. We can further simplify this model by assuming that the controller is co-located with the actuators and the only communication channel is between the sensors and the controller. This comes at no loss of generality so long as the information available to perform encoding and decoding is the same at the sensor, controller, and actuators. In this case, performing decoding and re-encoding at the controller is redundant, and the bottleneck link determines the effective data-rate. The information flow through the feedback loop can then be viewed as occurring over a single channel, which can also represent a multi-hop network connection, and in this case the effective data-rate refers to the rate available at the endpoints of the connection. On the other hand, we point out that in practice the information available for encoding and decoding may be different at different points in the network and solutions in this case are highly dependent on the assumed information pattern. Nevertheless, a global view of the network can be achieved by running a distributed consensus protocol (22) before attempting to perform control.

The first basic result on the information flow requirements for stabilization that we wish to describe is the so-called *data-rate theorem* (23, 24), which has also been the starting point for much of the research in the area of information constraints in NCS. This quantifies the

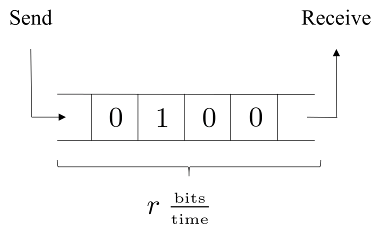

Figure 3

Bit-pipe channel

effect that communication has on closed-loop stabilization of unstable systems by stating that the communication rate available in the feedback loop should be at least as large as the *intrinsic entropy rate* of the system. For continuous linear systems the intrinsic entropy rate corresponds to the sum of the unstable modes of the system and for discrete systems it corresponds to the sum of the logarithms of the unstable modes. When this condition is satisfied, the controller can compensate for the growth of the state space occurring during the communication process and is able to keep the system stable. To illustrate this result for linear systems, consider the set of equations

$$\begin{cases} \mathbf{x}_{k+1} = \boldsymbol{A}\mathbf{x}_k + \boldsymbol{B}\mathbf{u}_k + \mathbf{v}_k & \text{1a.} \\ \mathbf{y}_k = \boldsymbol{C}\mathbf{x}_k + \mathbf{w}_k, & \text{1b.} \end{cases}$$

where $k = 0, 1, \dots$ is time, $\mathbf{x}_k \in \mathbb{R}^d$ represents the state variable of the system, $\mathbf{u}_k \in \mathbb{R}^m$ is the control input, $\mathbf{v}_k \in \mathbb{R}^d$ is an additive disturbance, $\mathbf{y}_k \in \mathbb{R}^p$ is the sensor measurement, $\mathbf{w}_k \in \mathbb{R}^p$ is the measurement disturbance, and $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$ are constant real matrixes of matching dimensions. Standard conditions on $(\boldsymbol{A}, \boldsymbol{B})$ to be reachable, $(\boldsymbol{C}, \boldsymbol{A})$ observable, are added to make the problems considered well posed. The equivalent continuous formulation is

$$\begin{cases} \dot{\mathbf{x}}(t) = \boldsymbol{A}\mathbf{x}(t) + \boldsymbol{B}\mathbf{u}(t) + \mathbf{v}(t) & \text{2a.} \\ \mathbf{y}(t) = \boldsymbol{C}\mathbf{x}(t) + \mathbf{w}(t). & \text{2b.} \end{cases}$$

In a first approximation, noise and bandwidth limitations in the communication channel can be captured by modeling the channel as a rate-limited "bit pipe" capable of transmitting only a fixed number r of bits in each time slot of the system's evolution, see Figure 3. In this way, the channel can represent a network connection with a limited available bit-rate, when transmitting below this rate communication errors are assumed to be negligible, and only quantization of the transmitted messages is accounted for.

When the control objective is to keep its state bounded, or asymptotically drive it to zero, the control law can be a linear function of the state estimate. Hence, for unstable linear systems under this rate-limited bit-pipe communication model, the central issue is to characterize the ability to perform a reliable estimate of the state at the receiving end of the communication channel. In order to keep the system stable, the data-rate theorem states that the information rate r supported by the channel must be large enough compared to the unstable modes of the system, so that it can compensate for the expansion of the state during the communication process. Namely,

$$r > \sum_{|\lambda_i| \geqslant 1} \log_2 |\lambda_i| \ \text{[bits/sec]}, \qquad \text{3.}$$

for discrete systems, where $\{\lambda_i\}$ are the open-loop eigenvalues raised to their corresponding algebraic multiplicities, and

$$r > \sum_{\mathrm{Re}\{\lambda_i\}>0} \lambda_i \log_2 e \ \ \text{[bits/sec]}, \tag{4.}$$

for continuous systems. If the real parts of all the eigenvalues of A are positive (unstable), this can be written as

$$r > \mathrm{tr}(A) \log_2 e \ \ \text{[bits/sec]}. \tag{5.}$$

The intuition behind the data-rate theorem is evident by considering a scalar system and noticing that while the volume of the state of the open loop system increases by $|\lambda|$ in a unit time step in the discrete setting —or by $|e^{\lambda}|$ in the continuous setting— in closed loop this expansion is compensated by a factor 2^{-r} due to the partitioning induced by the coder providing r bits of information through the communication channel. By imposing the product to be less than one and taking the logarithm base two, the results follow. Another interpretation arises if one identifies the right-hand side of (3) and (5) as a measure of the rate at which information is generated by the unstable plant, then the theorem essentially states that to achieve stability the channel must be able to transport information as fast as it is produced.

Early incarnations of this fundamental result appeared in (25, 26, 27, 28, 29) for undisturbed, scalar, unstable plants, when the objective is to keep the state bounded at all times. Improvement of the result from maintaining a bounded state to obtaining a state that asymptotically approaches zero are shown in (30, 31, 32) and require an adaptive "zoom-in, zoom-out" strategy that adjusts the range of the quantizer so that it increases as the plant's state approaches the target and decreases if the state diverges from the target. This follows the intuition that in order to drive the state to zero, the quantizer's resolution should become higher close to the target.

In the presence of stochastic disturbances, asymptotic stability can only be guaranteed within the range of the disturbances. The work (24) showed that for almost surely (a.s.) bounded disturbances and initial condition, the data-rate theorem holds and we can have

$$\sup_{k\in\mathbb{N}} \|\mathbf{x}_k\|^2 < \infty, \quad \text{a.s.} \tag{6.}$$

On the other hand, unbounded disturbances can drive the state arbitrarily far from zero, and one can only guarantee stability in a weaker, probabilistic sense. The typical approach is to consider mean-square (m.s.) stability, namely

$$\sup_{k\in\mathbb{N}} \mathbb{E}\{\|\mathbf{x}_k\|^2\} < \infty. \tag{7.}$$

The work (23) proved the data rate theorem using mean-square stability for systems with unbounded stochastic disturbances provided that higher moments are bounded, namely

$$\exists \epsilon > 0 : \mathbb{E}\{\|\mathbf{x}_0\|^{2+\epsilon}\} < \infty, \ \sup_{k\in\mathbb{N}} \mathbb{E}\{\|\mathbf{v}_k\|^{2+\epsilon}\} < \infty, \ \sup_{k\in\mathbb{N}} \mathbb{E}\{\|\mathbf{w}_k\|^{2+\epsilon}\} < \infty. \tag{8.}$$

A similar data-rate theorem formulation also holds for nonlinear systems. In this case one may consider a partially observed, time-invariant, dynamical system

$$\begin{cases} \dot{\mathbf{x}}(t) = f(\mathbf{x}(t), \mathbf{u}(t), \mathbf{v}(t)), & \text{(9a.)} \\ \mathbf{y}(t) = h(\mathbf{x}(t), \mathbf{w}(t)), & \text{(9b.)} \end{cases}$$

where f denotes the state transition function, h denotes the observation measurement function, $\mathbf{y}, \mathbf{x}, \mathbf{u}, \mathbf{v}, \mathbf{w}$ the observation, state, control input, state disturbance, and observation disturbance, respectively, as before. In order to express the data-rate theorem, one needs to quantify the rate at which the dynamical system generates information, which for the linear case corresponds to the right-hand side of 3. and 5. Obviously, this rate should be intrinsic to the nonlinear dynamical system and thus independent of encoding and decoding processes, controllers, and feedback communication constraints. One way to obtain such quantification is to refer to the *topological entropy* of the system (33), a construction spin-off from Kolmogorov's entropy definition for completely deterministic nonlinear maps (34). The idea behind this definition is to first fix an open cover for the space, through which each iteration of the map is observed. As the number of iterations increases, the family of all possible intersections of initial state open sets forms an increasingly fine open cover for the space. The topological entropy of the map is then obtained by taking the supremum of the asymptotic rate of increase of the cardinality of this open cover over all observation open covers. This measures the fastest rate at which uncertainty about the initial state can be reduced, or equivalently the fastest rate at which initial state information can be generated. There is an analogy here with source coding in classical information theory (35), which Kolmogorov credits as the inspiration for his own work (36). Source coding is concerned with determining the smallest data rate at which a stochastic source can be encoded, transmitted and reliably decoded over a noiseless digital channel. Shannon's source coding theorem states that the smallest possible data rate is equal to the *Shannon entropy* of the source, independent of external constructs. Parallels between Kolmogorov's deterministic theory and Shannon's stochastic theory are further explored in (37, 38). The work (33) considers fully observable, undisturbed systems closed over a a bit-pipe communication channel and uses the notion of topological entropy to determine necessary and sufficient bit rates for local uniform asymptotic stability. The work (39) provides the extension for partially observable systems. The works (40, 41) follow a different approach, expressing sufficient conditions for stabilization in terms of the Lipshitz constant. The work (42) considers noiseless and fully observed nonlinear systems with a special upper triangular structure, i.e. feedforward systems, providing a tight condition for global stability that matches the topological entropy formulation of (33). An extensive compendium of related results connecting different variants of topological entropy definitions to data rate requirements for different stability notions can be found in the monograph (17) and in the recent work (43) and references therein.

3. Extensions to noisy channels

Several generalizations of the simple bit-pipe communication model have been considered in the literature. An important body of work regards extensions to *stochastic channels*, namely channels whose behavior varies randomly due to noise. In this case, the rate available for transmission through the noisy channel must be defined in terms of information capacity. In this setting, a key result is that for undisturbed systems one can derive a data-rate theorem expressing the rate available for transmission in terms of the *Shannon capacity* of the channel (44, 45), which must be larger than the entropy rate of the system in order to guarantee stability. In contrast, when systems are subject to disturbances the standard notion of Shannon capacity turns out to be insufficient to express the ability to stabilize the system in both the a.s. and m.s. sense. In this case, alternative notions of capacity that have stronger reliability constraints must be used to formulate data-rate theorems, namely the *anytime capacity* must be used to express m.s. stability (46), or more generally α-moment stability, and the *zero-error* capacity must be used to express a.s. stability (47). The main difference between these notions of capacity is that while the Shannon capacity is defined as the supremum of the rates that can achieve an arbitrarily small probability of error, the anytime capacity has more stringent conditions on the probability of error, requiring "anytime" decoding of all codewords every time a new symbol is received, and imposing that the probability of having an error in any of the decoded codewords tends to zero exponentially as more and more symbols are received. On the other hand, the zero-error capacity requires that the probability of error is exactly zero for every transmitted codeword. In short, the Shannon capacity offers only "weak" reliability constraints and it is generally insufficient to characterize the ability to stabilize the system in the presence of external disturbances; while the zero-error and anytime capacity offer stronger reliability constraints and can be used to characterize the ability to stabilize the system in the presence of external disturbances in an a.s. and moment setting, respectively. For a more extensive discussion of the relationship between the different capacity definitions, we refer the reader to (9, 48).

To illustrate the main results for noisy channels, first consider stabilization of a scalar system over a simple stochastic erasure channel where the rate varies randomly between the two values $\{r, 0\}$ in an i.i.d. fashion. Namely, for all k we have the stochastic rate process

$$\mathbf{r}_k = \begin{cases} 0 & \text{w.p.} \quad p \\ r & \text{w.p.} \quad 1-p, \end{cases} \tag{10}$$

and we assume that both encoder and decoder have causal knowledge of the channel realization. In information-theoretic terms, this is as an r-bit packet erasure channel with acknowledgement of packet reception and erasure probability p. This channel is visually illustrated in Figure 4(a).

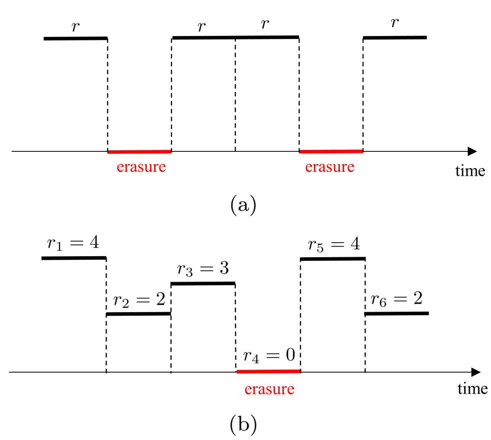

Figure 4

Example of stochastic channels. (a) r-bit packet erasure channel. (b) stochastic-rate channel, including erasures.

The condition to achieve m.s. stabilization 7. over this channel is expressed as

$$\mathbb{E}\left\{\frac{|\lambda|^2}{2^{2\mathbf{r}_k}}\right\} = p\frac{|\lambda|^2}{2^0} + (1-p)\frac{|\lambda|^2}{2^{2r}} < 1. \qquad 11.$$

Using the same interpretation of production and consumption of information mentioned above, this condition states that the average of the product of the open loop state expansion and compensation through r-bit quantization should be kept less than one to ensure stability. By rearranging terms, we obtain an expression where the anytime capacity of the channel appears on the left-hand side of the inequality (46)

$$\breve{C}(2) \equiv \frac{2^{2r}}{2^{2r}p + 1 - p} = \frac{2^{2r}}{p(2^{2r}-1)+1} > |\lambda|^2, \qquad 12.$$

here the anytime capacity $\breve{C}(\alpha)$ is parametrized by the mean-square stability exponent $\alpha = 2$. The expression in 12. shows a clear a trade-off between the reliability of the channel and the quantization rate. Namely, when the quantization rate $r \to \infty$, we obtain

$$\frac{1}{p} > |\lambda|^2, \qquad 13.$$

indicating that the erasure probability p must be small enough to guarantee stability. In contrast, the Shannon capacity of the r-bit erasure channel is (49)

$$C = (1-p)r, \qquad 14.$$

which diverges as $r \to \infty$, independent of the value of p. It follows that the Shannon capacity does not give in this case any indication on the ability to achieve stabilization. In general, for any finite value of r both the quantization rate and the reliability of the channel play a role in determining the ability to stabilize the system. We also note that when $r \to 0$ we obtain $|\lambda| < 1$, namely the system cannot be stabilized regardless how small the erasure probability p is.

The anytime capacity is the correct figure of merit to express data-rate theorems describing the ability to achieve α-moment stabilization (46) for more general noisy channels, beside the simple erasure one. The price to pay to have a complete characterization, however, is the computation of the anytime capacity that becomes increasingly difficult. Only for a few channels anytime capacity stabilization conditions similar to 12. have been obtained. These include time-varying rate channels (50, 51) where the rate process $\mathbf{r}_k$ varies randomly over time in an i.i.d. fashion, taking values in a subset of the non-negative integers, see Figure 4(b). The erasure channel considered above is a special case, when the rate process takes values in $\{0, r\}$. Results have also been obtained for stochastic channels where the rate varies according to a Markov process (52, 53, 48). This allows arbitrary temporal correlations of the channel variations over time, and results rely on the theory of Markov Jump Linear Systems (9).

In the special case of additive Gaussian channels, it turns out that the Shannon capacity is indeed sufficient to characterize m.s. stability and we refer to (9, Sec 1.4.4) and references therein for a description of these results. Another way to use the weaker notion of Shannon capacity to characterize the ability to stabilize the system is to relax the notion of stability. The work in (15, Chapter 8) considers the weaker notion of stability in probability, requiring the state to be bounded with probability at least $(1-\epsilon)$ by a constant K_ϵ that diverges as $\epsilon \to 0$, namely

$$\mathbb{P}\Big\{ \sup_t |\mathsf{x}(t)| < K_\epsilon \Big\} > 1-\epsilon, \tag{15.}$$

and shows 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. At the opposite extreme, for general stochastic channels, and systems with bounded disturbances, if instead of m.s. stability 7. one wants to achieve the more stringent a.s. stabilization condition 6., a basic result in (47) shows that the capacity notion to use is the zero-error one, and the work (54) shows that the zero-error capacity can be written in terms of an information functional describing the flow of information through the feedback loop.

Some additional extensions regard stabilization over channels with multiplicative noise that can be used to model fast-fading wireless communication channels or synchronization errors in system sampling. One example of this case is the work (55), which considers the following scalar system

$$\begin{cases} \mathsf{x}_{k+1} = A\,\mathsf{x}_k - \mathsf{u}_k & \text{16a.} \\ \mathsf{x}'_k = \mathsf{z}_k\,\mathsf{x}_k, & \text{16b.} \end{cases}$$

where A is constant, $\mathsf{x}_0 \sim \mathcal{N}(0,1)$, u_k can be any function of the current and previous observations and z_k are random variables representing the multiplicative noise. In this setting, the work (55) shows a result that is reminiscent of a data-rate theorem. Namely, if z_k are i.i.d. with a known bounded density with unit mean and variance equal to σ^2, letting $A^* = \sqrt{1+(1/\sigma^2)}$, we have that a memoryless linear controller can stabilize the system in a second-moment sense if $A \leqslant A^*$. Moreover, if $A > A^*$ the system cannot be second-moment stabilized using a linear control strategy.

4. Event-triggered control

Event-triggering (56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71) is a recent control paradigm that seeks to prescribe information exchange between the controller and the plant in an opportunistic manner. Rather than communicating periodically the control action, communication in event-triggering control occurs only when triggered by some events indicating the need to send fresh information to guarantee the correct execution of the task at hand (e.g., stabilization, tracking). The primary focus then is on minimizing the number of transmissions while guaranteeing the control objectives.

At a high level, one can view event triggering as *sampling in time* with the objective to identify the minimum sampling rate at which information may be transmitted through the feedback loop. Similarly, a bit-pipe communication model can be viewed as *sampling in space*, namely as quantization of the signal, and the data-rate theorem corresponds to the identification of the minimum quantization rate that can still guarantee stabilization. In the case of stochastic channels, we have both sampling in space, since we are transmitting through a digital channel messages of finite precision, and sampling in time, through errors and erasures. This view suggests that there should be a close connection between data-rate theorems and event-triggered control (69).

A first connection is revealed in (72), which presents a data-rate theorem for event-triggering strategies for systems subject to bounded disturbances and controlled over a bit-pipe communication channel. Consider the system's equations

$$\begin{cases} \dot{\mathbf{x}}(t) = \boldsymbol{A}\mathbf{x}(t) + \boldsymbol{B}\mathbf{u}(t) + \mathbf{v}(t) & \text{17a.} \\ \mathbf{y}(t) = \mathbf{x}(t), & \text{17b.} \end{cases}$$

where the initial condition $\mathbf{x}(0)$ and the system disturbance $\mathbf{v}(t)$ are a.s. bounded. At each triggering event, the sensor transmits to the controller a packet of a fixed number of bits. Letting $b_{\mathrm{s}}(t)$ be the total number of bits transmitted by the sensor to the controller up to time t, the data-rate theorem is expressed in terms of the asymptotic average transmission bit-rate of the sensor

$$r \equiv \limsup_{t\to\infty} \frac{b_{\mathrm{s}}(t)}{t} \quad \text{[bits/sec]}. \qquad 18.$$

Letting $\|x_\infty\|^2$ be a deterministic bound on the the steady state and κ be a sufficiently large constant (both depending on the range of the disturbance and the initial condition), it turns out that to obtain exponential stability at rate $\sigma \geqslant 0$, namely requiring that for all $t \geqslant 0$

$$\|\mathbf{x}(t)\|^2 \leqslant (\kappa - \|x_\infty\|^2)e^{-2\sigma t} + \|x_\infty\|^2 \quad \text{a.s.}, \qquad 19.$$

we need

$$r \geqslant (\mathrm{tr}(\boldsymbol{A}) + \sigma d)\log_2 e \quad \text{[bits/sec]}, \qquad 20.$$

where d indicates the dimension of the system. The expressions 19. and 20. are consistent with 5. and 6., where σd represent the extra bits required for exponential convergence to the steady state. It follows that the result in (72) can be viewed as being analogous to the one in (24), but obtained here in the context of event triggering for continuous systems and with exponential convergence guarantees. However, while the result in (24) is a data-rate theorem that is both necessary and sufficient for stabilization, the event-triggering controller design proposed in (72) uses an asymptotic data rate that is within a constant factor from the necessary condition 20. This sufficient rate clearly depends on the triggering strategy. In particular, the event-triggering strategy utilized in (72) is based on a Lyapunov function that ensures the desired convergence rate of the state. Nevertheless, the proposed design adjusts the communication rate in accordance with state information in an opportunistic fashion and it guarantees a uniform positive lower bound on the times between successive triggering events, so that degenerate cases where triggering occurs infinitely often in a finite interval are avoided.

5. Timing information in event-triggering

The results in (72) seem to indicate that the data-rate theorem is in complete harmony with event-triggering, in the sense that whether the transmission rate is limited by channel conditions, or it is limited opportunistically through event-triggering, the same fundamental limitation applies, which is dictated by the unstable modes of the system and by the desired convergence rate and expressed by 20.

It turns out, however, that event-triggering can also exploit an additional resource that is not accounted for in the current formulation and which allows to achieve stabilization with a dramatically lower data-rate. The work in (73) reveals that if the channel does not introduce any delay and the controller is aware of the triggering strategy used by the sensor, then one can achieve stabilization by transmitting at a rate that is arbitrarily close to zero. To illustrate this point, consider the following undisturbed system

$$\begin{cases} \dot{\mathsf{x}}(t) = \boldsymbol{A}\mathsf{x}(t) + \boldsymbol{B}\mathsf{u}(t) & \text{21a.} \\ \mathsf{y}(t) = \boldsymbol{C}\mathsf{x}(t), & \text{21b.} \end{cases}$$

where $\mathsf{x} \in \mathbb{R}^d$ and $\mathsf{y} \in \mathbb{R}$ (that is $\boldsymbol{C} \in \mathbb{R}^{1\times d}$) and the only uncertainty is due to the random initial condition. The channel connecting the sensor to the controller is assumed to be capable of transmitting one bit in an arbitrarily small time unit, so that communication of this binary symbol can be considered instantaneous when compared to the system dynamics. Let $\{t_{\mathrm{s}}^k\}_{k\in\mathbb{N}}$ be the sequence of times at which the sensor transmits a bit to the controller. These times are set by event triggering accordingly to a level crossing strategy. Letting h be a given threshold, a transmission occurs every time absolute value of the difference between two successive output samples crosses the threshold, namely the triggering condition is given by

$$|\mathsf{y}(t_{\mathrm{s}}^k) - \mathsf{y}(t_{\mathrm{s}}^{k+1})| = h. \qquad 22.$$

The system initially evolves in open-loop by letting $\boldsymbol{u} = 0$. Then, at each triggering time the sensor transmits a single bit to the controller that encodes the sign of the h step change in the y value. By receiving at least $d+1$ bits, and solving the following system of equations

$$\begin{bmatrix} \boldsymbol{C}(e^{At_{\mathrm{s}}^1} - e^{At_{\mathrm{s}}^0}) \\ \boldsymbol{C}(e^{At_{\mathrm{s}}^2} - e^{At_{\mathrm{s}}^1}) \\ \vdots \\ \boldsymbol{C}(e^{At_{\mathrm{s}}^d} - e^{At_{\mathrm{s}}^{d-1}}) \end{bmatrix} \mathsf{x}_0 = \begin{bmatrix} \mathsf{y}(t_{\mathrm{s}}^1) - \mathsf{y}(t_{\mathrm{s}}^0) \\ \mathsf{y}(t_{\mathrm{s}}^2) - \mathsf{y}(t_{\mathrm{s}}^1) \\ \vdots \\ \mathsf{y}(t_{\mathrm{s}}^d) - \mathsf{y}(t_{\mathrm{s}}^{d-1}) \end{bmatrix},$$

where the right-hand side is a column vector of $\pm h$ values, the controller can infer the initial condition provided that the matrix on the left-hand side is nonsingular. Once the initial condition is known, it can then stabilize the system in a closed-loop fashion. Since we have

$$\limsup_{t\to\infty} \frac{b_{\mathrm{s}}(t)}{t} = \lim_{t\to\infty} \frac{(d+1)}{t} = 0, \qquad 23.$$

it follows that we can stabilize the system with an arbitrarily small transmission rate.

The intuition behind the result follows by noting that, like pauses are used in spoken language to convey information, in the context of event-triggering control it is possible transmit information in the feedback loop not only by message content, but also with its timing. Specifically, in the absence of delay in the communication channel, the mere act of sending one bit at a given time can reveal the state of the system with arbitrary precision, and transmitting a single data payload bit at every triggering event is enough to compute the appropriate control action. In fact, we may take this intuition one step further and also notice that under the same assumptions of (73) we do not even need to transmit a bit at each triggering time. To reveal any component x of $\mathsf{x}(0)$, we could transmit a single arbitrary symbol ♠ at a time equal to any bijective mapping of x into a point of the non-negative reals. For example, we could transmit a symbol ♠ at time $t = \tan^{-1}(\mathsf{x})$, where $t \in [0, \pi]$. Since there is no choice associated to the symbol selection, in principle the reception of the symbol should not carry any information. However, its arrival time carries information and can reveal x with arbitrary precision. To communicate the whole vector $\mathsf{x}(0)$, we could then send $d+1$ identical ♠ symbols at different times and encode all the components of $\mathbf{x}(0)$ in their inter-transmission times. In principle, one could even send a single ♠ symbol to encode the whole $\mathbf{x}(0)$ vector by using a d-dimensional space-filling curve and selecting a time of transmission for which a point on the curve is mapped onto $\mathbf{x}(0)$.

The important message to be taken from (73) is that using event triggering information can be transmitted in the feedback loop not only by sending data but also by carefully selecting the times of transmission. A similar observation is also made in (74). This work considers the system to be fully observable, namely $\boldsymbol{C}$ to be the identity, and the sensor to transmit, at a *fixed* sequence of transmission times, $\{t_k\}$ symbols from a finite alphabet over a delay-free and error-free communication channel. It is further assumed that a special symbol in the alphabet can be transmitted without consuming any communication resources, effectively representing the absence of an explicit transmission, while the other symbols require one unit of communication resource per transmission. From an information-theoretic perspective, this set up is related to the silence-based communication paradigm of (75). Letting $s(t_k)$ be the total number of non-free symbols transmitted by the sensor to the controller up to time t_k, the asymptotic average cost per unit time is given by

$$c \equiv \limsup_{k\to\infty} \frac{s(t_k)}{t_k} \quad \text{[symbols/sec]}, \tag{24.}$$

where $t_k \to \infty$ as k tends to infinity. This can also be interpreted as an effective data rate, since it represents the rate accounting for only the non-free transmissions, and should be compared with 18. In both cases, the rate depends on the transmission strategy. It is shown in (74) that stabilization can be achieved with arbitrarily small values of 24. by letting the transmission times $t_k = kT$, decreasing the sampling period T, and transmitting non-free symbols rarely. In this regime, the transmission policy resembles an event-triggering strategy where the transmission of a non-free symbol may occur at any given time, which depends on the encoding strategy, and can be chosen with arbitrary precision as $T \to 0$. At all other times only free symbols are sent —which is analogous to sending nothing. As in the case of (73), in this setup the act of transmitting a non-free symbol now carries an amount of information that can be made arbitrarily large by decreasing the sampling time T. This allows to decrease the number of transmitted non-free symbols and drives the effective rate 24. arbitrarily close to zero.

6. Timing information in the presence of delay

The works we have described suggest that a more general formulation of data-rate theorems should account for two distinct information flows: one is through data payload (possibly corrupted by noise) and another is through timing (possibly corrupted by delay). Traditionally, only the data payload case has been considered, but the timing information can be very relevant especially in the context of event triggering. The work in (72) only considers communication through data payload and does not attempt to exploit timing information. As a result, it recovers the traditional data-rate theorem formulation. In contrast, the works in (73) and (74) show that stabilization can be achieved with arbitrarily small data payload rate, by exploiting the timing information implicit in event-triggering schemes, when sender and receiver are perfectly synchronized.

At this point, one may suspect that the ability to stabilize the system with zero payload rate is an artifact of the assumed perfect synchronization between the sensor and the controller achieved through a zero-delay channel. As we shall see next, this is not the case. In the presence of unknown delay, the value of the timing information in the triggering events decreases, because in this case the sensor may only reveal the state of the system with a finite precision —which depends on the range of the unknown delay. However, as long as the amount of information supplied by timing is above what prescribed by the data-rate theorem for stabilization, it is still possible to stabilize the system with an arbitrarily small data payload rate. Next, we illustrate this point in more detail.

6.1. Information Access Rate vs Information Transmission Rate

The works (76, 77) make a key distinction between the *information access rate*, that is the rate at which the controller needs to receive information, conveyed by both data payload and timing information, and that is subject to the requirement expressed by the classic data-rate theorem; and the *information transmission rate*, that is the rate at which the sensor needs to send data in the form of payload bits, that depends on the triggering scheme and that can become arbitrarily small without affecting the ability to stabilize the system.

First, let us take the viewpoint of the sensor and examine the amount of information in the data payload transmissions to the controller. Let $b_{\mathrm{s}}(t)$ be the number of bits in the data payload transmitted by the sensor up to time t, and define the information transmission rate as

$$r_{\mathrm{s}} \equiv \limsup_{t \to \infty} \frac{b_{\mathrm{s}}(t)}{t}.$$

Let us now consider the viewpoint of the controller and examine the amount of information that it needs to receive in order to be able to select its stabilizing policy. This includes both payload and timing information. It is also the same as the number of bits needed to construct a reliable state estimate (78, Theorem 1). We let $b_{\mathrm{c}}(t)$ be the number of bits required at the controller to perform its selection at time t and define the *information access rate* as

$$r_{\mathrm{c}} \equiv \limsup_{t \to \infty} \frac{b_{\mathrm{c}}(t)}{t}.$$

In classic data-rate theorems r_{c} coincides with r_{s} because the controller uses only data payload bits to select its control law. On the other hand, as discussed above, by exploiting timing information r_{s} and r_{c} can be substantially different and the classic data-rate limitation applies to r_{c} only, while we can achieve stabilization with r_{s} arbitrarily close to zero.

To view the limitation on r_{c}, we consider the same system's equations as in 17. In this case, a necessary and sufficient condition to achieve exponential stabilization at rate σ is given by the usual data-rate theorem formula expressed in terms of r_{c}

$$r_{\mathrm{c}} \geqslant (\mathrm{tr}(\boldsymbol{A}) + \sigma d) \log_2 e \quad [\text{bits/sec}]. \qquad 25.$$

This result should be compared with 20. It is important to stress that the limitation in 25. describes what is required by the controller, and it does not depend on the feedback structure — including aspects such as communication delays, information pattern at the sensor and the controller, and whether the times at which transmissions occur are state-dependent, as in event-triggered control, or periodic, as in time-triggered control. In order to obtain 25., one considers for any control input trajectory $u(t)$ the subset of initial conditions for which the plant is stabilized by such input. Then, one constructs a cover of the set of all initial conditions by stabilizing control policies. This leads to a discrete set of choices for selecting the stabilization policy for any given realization of the initial condition. It follows that the logarithm of the covering number is the number of bits needed by the controller by time t to select a stabilizing control policy. A usual balance of information argument between the rate of expansion of the uncertainty in the state due to the random initial condition, and the quantization due to the covering finally leads to the lower bound in 25.

Having established the classic data-rate theorem result for the information access rate, we can now ask what is the data-rate requirement on the information transmission rate r_{s}, assuming that the sensor has access to causal feedback regarding what has been received by the controller and for different ranges of the possible delay. While we have already established that r_{s} can be arbitrarily close to zero in the absence of delay, the presence of unknown delay decreases the amount of information that can be communicated by timing, and this may require r_{s} to become positive.

To illustrate the results, we denote by $\{t_{\mathrm{s}}^k\}_{k\in\mathbb{N}}$ the sequence of times when the sensor transmits a packet of a certain number $g(t_{\mathrm{s}}^k)$ of bits to the controller. We assume the packet is delivered to the controller without error and entirely but with an unknown delay. Letting $\{t_{\mathrm{c}}^k\}_{k\in\mathbb{N}}$ be the sequence of times when the controller receives the packets transmitted at times $\{t_{\mathrm{s}}^k\}_{k\in\mathbb{N}}$, we assume that for all $k \in \mathbb{N}$ the communication delay $\Delta_k = t_{\mathrm{c}}^k - t_{\mathrm{s}}^k$ satisfies

$$\Delta_k \leqslant \gamma, \tag{26.}$$

where $\gamma \in \mathbb{R}_{\geqslant 0}$, and that both t_{s}^k and t_{c}^k tend to infinity as $k \to \infty$. We can then study how the rate r_{s} required for stabilization using an event-triggering varies as a function of γ. The work (76) considers the case of systems without disturbances, where the objective is to drive the state to zero at an exponential rate σ. The work (77) considers the case of systems with disturbances using a notion of input to state stability, which guarantees that the state is bounded at all times and this bound, as usual, depends on the range of the disturbance. While results hold for both scalar and vector systems, for illustrative purposes in the following we review the results in (77) for scalar system.

A plot of the rate required to keep the state bounded when using any threshold-based event-triggering policy based on the value of the state estimation error is depicted in Figure 5. This shows that the required rate for stabilization undergoes a *phase transition*: for small values of the delay upper bound γ the system can be stabilized with an arbitrarily small information transmission rate. However, when γ reaches the critical threshold

$$\gamma_{\mathrm{c}} = \frac{\ln 2}{A}, \tag{27.}$$

the required rate begins to increase, eventually surpassing the data-rate theorem requirement $A/\ln 2$, where A here is a positive scalar. This indicates that for $\gamma < \gamma_{\mathrm{c}}$ the amount of information contained in the timing of the triggering events is large enough that the rate that must be supplied by data-payload to guarantee stability is zero. On the other hand, when $\gamma > \gamma_{\mathrm{c}}$ the information contained in the timing of the triggering events is not enough to guarantee stability and the rate must begin to increase. One way to interpret this result is that in the presence of delay the value of the timing information supplied by event triggering "deteriorates" and eventually becomes insufficient to be used alone for stabilization. On the other hand, increasing the the delay also affects the rate at which the transmitted payload bits are received, which results in a higher transmission rate requirement that can surpass the data-rate theorem requirement.

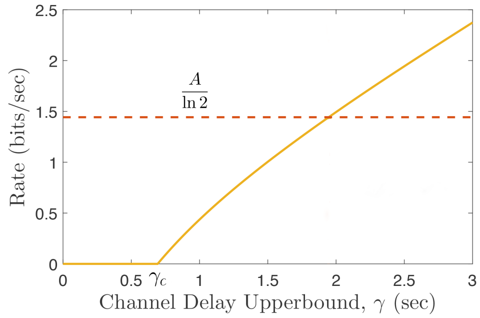

Figure 5

Phase transition of the necessary information transmission rate for stabilization. The graph is valid for any generic system. In this example we have a scalar system with no disturbance, $A = 1$, $\gamma_c = \ln 2/A = 0.6931$ and the rate dictated by the data-rate theorem is $r_c \geqslant A/\ln 2 = 1.4427$.

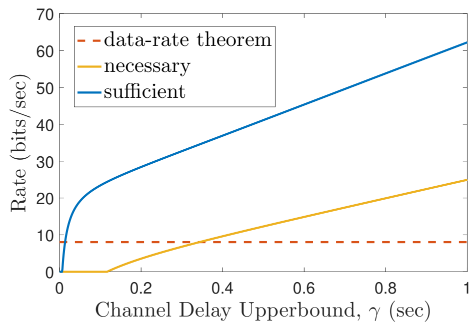

Figure 6

Sufficient and necessary transmission rates for stabilization. Here we have a scalar system with disturbances bounded by 0.4, with $A = 5.5651$. The rate dictated by the data-rate theorem is $r_c \geqslant A/\ln 2 = 8.02874$.

We also note that the critical value γ_c at which the information transmission rate becomes positive equals the inverse of the entropy rate of the system, namely 27. is the inverse of the critical rate in the data-rate theorem formula 5.. Recalling the production and consumption of information analogy that we discussed in Section 2, we have that for $\gamma = \gamma_c$ the entropy of the system can expand by one new bit at every delay occurrence and this amount of information cannot be counter-balanced by the information carried by the event triggering times. In other words, the information supplied by the triggering events can always be "one bit short" due to the uncertainty introduced by the delay, and this bit must be supplied by the data payload to ensure stabilization. For this reason, the rate r_s begins to increase once γ reaches the critical value γ_c. Figure 6 also shows a sufficient condition for stabilization obtained using a given triggering strategy described in (77) that employs a fixed threshold policy and compares it with the necessary condition and with the data-rate theorem requirement. Results have been validated in a real system configuration in (79).

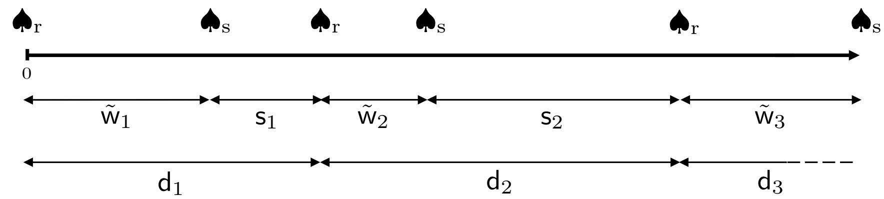

Figure 7

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

6.2. The information value of event timing

We have shown that information useful for control can be carried through the feedback loop by both data packets and event timing. Event timing may allow to achieve stabilization by sending packets at a bit-rate that is lower than what the classic data-rate theorem prescribes. We now ask whether it is possible to provide an information-theoretic characterization of event timing, and whether it is possible to recover the classic data-rate theorem formulation relating the amount of information carried by event timing alone, to the intrinsic entropy rate of the system.

To quantify the amount of timing information the work in (80) considers a channel carrying symbols ♠ from a unitary alphabet, where each transmission is received after a random delay. Since the alphabet is composed by a single symbol, there is no information conveyed in data packets and communication can only occur by selecting the times of transmission of the unitary symbols. 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 $\tilde{\mathsf{w}}_{i+1}$ seconds and then transmits the next ♠. Transmitted symbols are subject to i.i.d. random delays $\{\mathsf{s}_i\}$. Letting d_i be the inter-reception time between two consecutive symbols, we have

$$\mathsf{d}_i = \tilde{\mathsf{w}}_i + \mathsf{s}_i. \qquad 28.$$

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. Fig. 7 provides an example of the timing channel in action. The work (81) defines the *timing capacity* for this telephoning signaling system in terms of mutual information between transmitter and receiver. This notion is the analogous of the Shannon capacity for the timing channel. The work in (80) considers stabilization of the following scalar continuous-time system with no disturbance over the timing channel described above

$$\dot{\mathsf{x}} = A\mathsf{x}(t) + B\mathsf{u}(t). \qquad 29.$$

The constants $A, B \in \mathbb{R}$ are such that $A > 0$ and $B \neq 0$. Since the system in 29. is not subject to disturbances, we expect that a notion of capacity analogous to the Shannon one is sufficient to characterize the ability to stabilize the system, as discussed in Section 2. Indeed, by using the notion of timing capacity of (81), the work (80) shows that for the state to converge to zero in probability, the timing capacity of the channel should be at least as large as the entropy rate of the system. Conversely, in the case the random delays are exponentially distributed, 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. Finally, since the timing capacity depends on the distribution of the delay, it is also shown that in the case of exponentially distributed delay it is possible to achieve stabilization at zero data-rate only for sufficiently small average delay, namely when

$$\mathbb{E}\{\mathsf{s}\} < (e\,A)^{-1}, \qquad 30.$$

which confirms the intuition from the event-triggering results that to achieve stabilization at zero data-rate the delay should be sufficiently small.

7. Estimation under communication constraints

In closed-loop systems, the ability to select the correct control action to keep the system stable boils down to that of constructing a reliable state estimate that can be used for stabilization. As we discussed in Section 2, in order to keep the system stable the amount of information that must flow through the feedback loop must compensate for the expansion in the uncertainty of the state, and this dictates the communication constraints expressed by the various data-rate theorem formulations.

In the absence of the controller, the problem of estimating the state of an open-loop dynamical system observed over a communication channel is also of interest, and is further motivated by additional applications such as situation awareness (82, 83, 84), asset tracking (85, 86, 87), smart cities (88, 89, 90), Internet of Things (91, 92, 93) and network localization and navigation (NLN) (94, 95, 96, 97), where nodes in a network aim to infer their positions and possibly other position-related quantities using observations obtained via different types of sensors. In this case, results analogous to the data-rate theorem for stabilization have been obtained, and in what follows we wish to compare and put them in the context of those that we have already described.

The works (98, 99, 100) exploit dynamical system entropy notions for estimation that are inspired by the topological entropy approach of (17, 33, 43) that was used to determine stabilization conditions for nonlinear systems over bit-pipe communication channels. In particular, the work (100) introduces the notion of *estimation entropy* in terms of the number of system trajectories that approximate all other trajectories up to an exponentially decaying error. In the case of linear systems of the form $\dot{\boldsymbol{x}} = \boldsymbol{A}\,\boldsymbol{x}(t)$, and exponential error decay rate $\sigma \geqslant 0$, the estimation entropy reduces to $(\mathrm{tr}(\boldsymbol{A}) + \sigma d)\log_2 e$ [bits/sec], which is analogous to the stabilization result 25. Furthermore, for more general nonlinear systems of the form $\dot{\boldsymbol{x}} = f(\boldsymbol{x})$, where $\mathbf{x}(t) \in \mathbb{R}^d$, the work (100) shows supper and lower bounds on the estimation entropy.

In the context of exploiting timing information for estimation in event-based transmission, the work (101) considers estimation over a finite-size packet communication channel with delay analogous to the one in (76, 77). Here the aim is to remotely estimate a discrete-event process

$$\mathsf{x}(t) = \sum_{k=0}^{\infty} \eta_k\, \mathbb{1}(\tau_k \leqslant t < \tau_{k+1}), \qquad 31.$$

where the system states η_k are discrete i.i.d. random variables that belongs to a finite set, $\tau_k \in \mathbb{N}$ denotes the random time when a state transition to η_k occurs, $0 < \tau_0 < \tau_1 < \ldots$, and $\mathbb{1}(\cdot)$ is the indicator function. It follows that $\mathsf{x}(t)$ describes the state evolution in continuous time and the random duration time of each state is $\mathsf{t}_k = \tau_{k+1} - \tau_k$. Since the process $\mathsf{x}(t)$ remains constant during inter-event times, it is sufficient to describe $\mathsf{x}(t)$ with $\{\mathsf{x}(k) \mid k \in \mathbb{N}\}$.

To perform remote estimation, when a state transition to η_k occurs, a packet with a finite number of bits is transmitted over a channel. Like in the communication setting of (76, 77), the packet is delivered to the receiver without error but with an unknown delay, denoted by $\tilde{\Delta}_k$. The transmission delays $\{\tilde{\Delta}_k\}$ are assumed to be random, i.i.d, and independent of the states $\{\eta_k\}$. The amount of information that the system produces can be expressed in terms of the of the *Shannon entropy* of two stochastic sources, namely the unknown state value η and the unknown inter-event time t. Letting $H\{\cdot\}$ be the joint Shannon entropy of an ensemble of random variables (49), we have that the entropy rate of the information produced is

$$r_0 = \lim_{k\to\infty} \frac{H\{\mathsf{x}(0), \mathsf{x}(1), \ldots, \mathsf{x}(k)\}}{k} = \frac{H\{\eta\} + H\{\mathsf{t}\}}{\mathbb{E}\{\mathsf{t}\}} \quad \text{[bits/sec]}. \qquad 32.$$

This represents the average rate at which the system generates information and the receiver needs to have access to at least this amount of information to construct a correct estimate of the state. It can also be interpreted as the stochastic analogue of the information access rate defined in Section 6.1 for closed-loop stabilization. The information access rate was defined in a deterministic setting under worst-case delay conditions, while here we have a stochastic setting and an average transmission rate expressed in terms of entropy.

On the other hand, by exploiting knowledge of the reception times $\mathsf{t}_k + \tilde{\Delta}_\mathsf{k}$, the receiver may be able to perform estimation with a rate lower than 32. It turns out that the information rate required by the receiver for real time estimation of the process 31. is given by

$$r_0' = \frac{H\{\eta\} + H\{\mathsf{t} | \mathsf{t} + \tilde{\Delta}\}}{\mathbb{E}\{\mathsf{t}\}} \quad \text{[bits/sec]}. \qquad 33.$$

Since conditioning reduces the entropy, we immediately deduce that $r_0' \leqslant r_0$. There is an analogy in this case with the information transmission rate defined in Section 6.1, which is the deterministic counterpart of this reduced entropy rate for the case of stabilization. In the case of small delay, we have $H\{\mathsf{t}|\mathsf{t} + \tilde{\Delta}\} \approx H\{\mathsf{t}|\mathsf{t}\} = 0$, and the rate required for estimation reduces to

$$r_0' = \frac{H\{\eta\}}{\mathbb{E}\{\mathsf{t}\}} \quad \text{[bits/sec]}, \qquad 34.$$

namely we need to transmit only the number of bits required on average to describe the value of the process $\{\eta_k\}$ and we do not need to encode any information regarding the inter-transmission times $\{\mathsf{t}_k\}$. We also point out that in the event-triggering results discussed in Section 6.1 the transmission time was a function of the system state. Here the transmission time and the symbol are assumed to be independent. Consequently, the timing information cannot further reduce the uncertainty in the state and 34. remains a nonzero lower bound on the required rate. In contrast, in the event-triggering results described in Section 6.1, since timing information can also encode information about the state, the minimum required transmission rate can become arbitrarily small for small delay values.

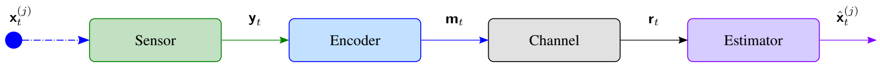

Figure 8

Block diagram for decentralized inference in a two-node system. The state $\mathbf{x}_t^{(j)}$ is measured by a sensor to generate observation $\mathbf{y}_t$ at time step t. Observations $\mathbf{y}_{0:t}$ are used by the encoder to generate a message $\mathbf{m}_t$, which is then transmitted via the channel. Using the received messages $\mathbf{m}_{0:t}$, a decentralized estimator $\hat{\mathbf{x}}_t^{(j)}$ of $\mathbf{x}_t^{(j)}$ is evaluated by node j.

8. Estimation over noisy channels

In a decentralized estimation problem a node can exchange messages that contain information of the states of interest with other nodes via noisy communication channels. In Section 3 we have seen that to guarantee moment stability over noisy channels a suitable metric for characterizing the quality of a communication channel is anytime capacity, a notion introduced by Sahai and Mitter in (46) that is is parameterized by a positive scalar α, which specifies the requirements on the communication reliability. In particular, this notion is used to establish tight necessary and sufficient conditions for stabilizing a system over a noisy channel in the presence of bounded disturbances (46).

In the following, we present the results in (102, 103, 104, 105), showing that the anytime capacity is also a relevant measure of information transmission for estimation of open-loop systems over noisy channels in both a single-node and multiple-node scenarios. Specifically, while a necessary condition for bounded moment error is expressed in terms of the Shannon capacity, a sufficient condition is obtained in terms of the anytime capacity. Since the anytime capacity expresses communication with stronger reliability constraints, for any $\alpha > 0$ the α-anytime capacity of a channel is greater than or equal to its Shannon capacity. It follows that these conditions are not tight in general. Nevertheless, their form resemble analogous results for closed loop stabilization and can also be extended to multiple-node networks settings.

8.1. Two-node system

Consider a system consisting of node i and node j in discrete-time scenarios. In particular, node j is associated with a time-varying state that this node aims to infer. The state of node j at time step t is denoted by a d-dimensional random vector $\mathbf{x}_t^{(j)}$ (see Fig. 8), which satisfies

$$\mathbf{x}_t^{(j)} = \boldsymbol{A}^{(j)}\mathbf{x}_{t-1}^{(j)} + \mathbf{v}_t^{(j)}, \qquad t = 1, 2, \dots \tag{35}$$

where $\boldsymbol{A}^{(j)} \in \mathbb{R}^{d\times d}$ is a deterministic matrix know to both nodes, and $\mathbf{v}_t^{(j)} \in \mathbb{R}^d$ is a zero-mean random vector representing the disturbance to the state.

The other node in the system, node i, obtains an observation $\mathbf{y}_t$ of $\mathbf{x}_t^{(j)}$ at each time step t given by

$$\mathbf{y}_t = \boldsymbol{C}\mathbf{x}_t^{(j)} + \mathbf{w}_t, \qquad t = 0, 1, \dots$$

where $\boldsymbol{C}$ is the sensor gain matrix known to both nodes, and $\mathbf{w}_t$ is a zero-mean random vector representing the observation noise at time step t. Moreover, node i generates an encoded message represented by a random vector $\mathbf{m}_t$ at time step t based on its observations

$\mathbf{y}_{0:t}$, i.e., $\mathbf{m}_t$ is a function of $\mathbf{y}_{0:t}$. The concatenation of such functions from time step 0 to the last time step of interest is referred to as an encoding strategy, which is designed by node i. Message $\mathbf{m}_t$ is transmitted via a memoryless channel to node j and the received message is represented by a random vector $\mathbf{r}_t$ that may be different from $\mathbf{m}_t$ due to noise, fading, and interference in the channel.

The following assumptions are made on the state disturbance and the observations.

1. $\mathbf{v}_t^{(j)}$ are independent over time steps t for all $j \in \mathcal{V}$. That is to say, random vectors $\mathbf{v}_{t_0}^{(j)}, \mathbf{v}_{t_1}^{(j)}, \ldots, \mathbf{v}_{t_n}^{(j)}$ are independent for any positive integer n and $0 \leqslant t_0 < t_1 < \cdots < t_n$. Similarly, $\mathbf{w}_t$ are independent over t. In addition, the process $\{\mathbf{v}_t^{(j)}\}_{t\geqslant 0}$ and $\{\mathbf{w}_t\}_{t\geqslant 0}$ are independent.
2. There exists a scalar $a > 2$ such that sequences $\{\mathbb{E}\{\|\mathbf{v}_t^{(j)}\|^a\}\}_{t\geqslant 0}$ and $\{\mathbb{E}\{\|\mathbf{w}_t\|^a\}\}_{t\geqslant 0}$ are bounded over time, namely

$$\sup_{t\geqslant 0} \mathbb{E}\Big\{\big\|\mathbf{v}_t^{(j)}\big\|^a\Big\} < \infty\,, \qquad \sup_{t\geqslant 0} \mathbb{E}\Big\{\big\|\mathbf{w}_t\big\|^a\Big\} < \infty\,.$$

3. There exists a constant $\underline{h} > -\infty$ such that the differential entropy $h(\mathbf{v}_t^{(j)})$ of $\mathbf{v}_t^{(j)}$ satisfies $h(\mathbf{v}_t^{(j)}) > \underline{h}$ for all $t \geqslant 0$.

The above assumptions are mild and the adopted models for the states, observations, as well as message transmission are general. In particular, Assumption 1 is widely adopted in literature on inference and filtering. Assumption 2 holds if the tail of the distribution for each entry of $\mathbf{v}_t^{(j)}$ and $\mathbf{w}_t$ is not heavy. As an example, if both $\mathbf{v}_t^{(j)}$ and $\mathbf{w}_t$ have an identical distribution for different time steps t, and entries of $\mathbf{v}_t^{(j)}$ and $\mathbf{w}_t$ are sub-exponential random variables,[1] then Assumption 2 holds for all a. Assumption 2 indicates that the uncertainty in the state disturbance does not vanish as time approaches infinity.

Node j evaluates a decentralized estimator $\hat{\mathbf{x}}_t^{(j)}$ of $\mathbf{x}_t^{(j)}$ at time step t using its received messages $\mathbf{r}_{0:t}$, namely $\hat{\mathbf{x}}_t^{(j)}$ is a function of $\mathbf{r}_{0:t}$. The optimal design of the estimator depends on the metric for inference error. The metric we consider for the inference error at time step t is the bth moment of $\|\hat{\mathbf{x}}_t^{(j)} - \mathbf{x}_t^{(j)}\|$, namely

$$\mathbb{E}\Big\{\big\|\hat{\mathbf{x}}_t^{(j)} - \mathbf{x}_t^{(j)}\big\|^b\Big\} =: e_t^{(j)} \tag{36}$$

where $b \geqslant 2$. In the special case where $b = 2$, this metric becomes the mean-square error (MSE). We establish conditions under which there exist an encoder at node i and a decentralized estimator at node j such that $e_t^{(j)}$ is bounded over time, i.e.,

$$\sup_{t\geqslant 0} e_t^{(j)} < \infty\,. \tag{37}$$

Intuitively, whether 37. holds or not depends on the quality of the channel from node i to node j: a better channel allows a higher rate of messages to be transmitted reliably, thus increasing the amount of information of the unknown state obtained by node j.

The work in (102) shows that

[1] A random variable x is sub-exponential if there exists a constant $c > 0$ such that $\mathbb{P}\{|\mathsf{x}| > x\} \leqslant 2\exp\{-cx\}$ for any $x \geqslant 0$ (106, Chapter 2). For example, a random variable is sub-exponential if it has a Gaussian distribution or if it has a bounded support.

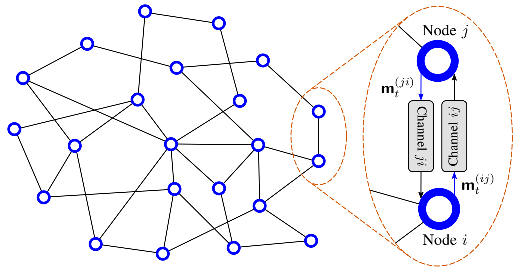

Figure 9

Decentralized inference via sensing and communication in a multi-node network. A pair of nodes in the network is connected by an edge if they are neighbors to each other (figure from (105)).

1. If there exist an encoder at node i and a decentralized estimator at node j such that 37. holds, then the Shannon capacity C of the channel satisfies

$$C > \sum_{|\lambda_i| \geqslant 1} \log_2 |\lambda_i^{(j)}| \,. \tag{38}$$

2. Conversely, if there exists a parameter $\alpha > \frac{ab}{a-b} \log\left|\rho(\boldsymbol{A}^{(j)})\right|$ such that the α-anytime capacity $\breve{C}(\alpha)$ of the channel satisfies

$$\breve{C}(\alpha) > \sum_{|\lambda_i| \geqslant 1} \log_2 |\lambda_i^{(j)}| \tag{39}$$

   then there exist an encoder at node i and a decentralized estimator at node j such that 37. holds. Here, $\rho(\boldsymbol{A}^{(j)})$ represents the spectral radius, i.e., the largest of the magnitudes of all the eigenvalues of $\boldsymbol{A}^{(j)}$.

This result specifies a necessary condition and a sufficient condition in 38. and 39., respectively, for the inference error to be bounded over time and is is parallel to the data rate theorem for stabilization over noisy channels as both 38. and 39. compare the capacity of the channel with a threshold determined by eigenvalues of $\boldsymbol{A}^{(j)}$.

Since the right-hand sides of 38. and 39. are equal the necessary and sufficient conditions become tight for channels whose Shannon capacity and anytime capacity coincide. Such channels include the noiseless bit-pipe channels and the Gaussian channel with feedback (46). As a final remark, we also point out that Assumption 2 is required only for establishing the sufficient condition 39. and not required for the necessary condition 38., whereas Assumption 3 is required only for 38. and not required for 39..

8.2. Multi-node network

Conditions for the boundedness of inference error for decentralized inference can also be established in a general network with multiple nodes. Specifically, consider a network comprising a set $\mathcal{V}$ of nodes where each node is associated with a time-varying unknown state (see Fig. 9). In particular, the state of node j at time step t is represented by $\mathbf{x}_t^{(j)}$ and it satisfies 35.. Each node in the network can perform observations and exchange messages with other nodes within its sensing and communication range, which are referred to as neighbors of the node. Specifically, node i obtains an intra-node observation $\mathbf{y}_t^{(ii)}$ as well as an inter-node observation $\mathbf{y}_t^{(ij)}$ for each neighbor $j \in \mathcal{N}^{(i)}$ at time step t, where $\mathcal{N}^{(i)}$ represents the set of neighbors of node i. Observations $\mathbf{y}_t^{(ii)}$ and $\mathbf{y}_t^{(ij)}$ are given by

$$\mathbf{y}_t^{(ii)} = \boldsymbol{C}^{(ii)}\mathbf{x}_t^{(i)} + \mathbf{w}_t^{(ii)}\,,$$
$$\mathbf{y}_t^{(ij)} = \boldsymbol{C}_1^{(ij)}\mathbf{x}_t^{(i)} + \boldsymbol{C}_2^{(ij)}\mathbf{x}_t^{(j)} + \mathbf{w}_t^{(ij)}\,, \qquad t = 0, 1, \ldots$$

where $\boldsymbol{C}^{(ii)}$, $\boldsymbol{C}_1^{(ij)}$, and $\boldsymbol{C}_2^{(ij)}$ are sensor gain matrices, whereas random vectors $\mathbf{w}_t^{(ij)}$ and $\mathbf{w}_t^{(ij)}$ represent observation noise. Moreover, node i transmits an encoded message $\mathbf{m}_t^{(ij)}$ to each neighbor $j \in \mathcal{N}^{(i)}$ at time step t. In particular, $\mathbf{m}_t^{(ij)}$ is generated based on the observations obtained by node i up to time step t and messages received by node i up to time step $t-1$. In other words, $\mathbf{m}_t^{(ij)}$ is a function of $\mathbf{y}_{0:t}^{(ii)}$, $\{\mathbf{y}_{0:t}^{(ij)} : j \in \mathcal{N}^{(i)}\}$, and $\{\mathbf{r}_{0:t-1}^{(ki)} : k \in \mathcal{N}^{(i)}\}$. Here $\mathbf{r}_\tau^{(ki)}$ is the message received by node i from node k at any time step $\tau \geqslant 0$.

A subset $\mathcal{V}_\mathrm{a} \subseteq \mathcal{V}$ of nodes in the network referred to as agents aim to infer their states in real time.[2] In particular, each agent j evaluates an estimator $\hat{\mathbf{x}}_t^{(j)}$ of $\mathbf{x}_t^{(j)}$ at time step t using the observations and received messages obtained up to time t. In other words, $\hat{\mathbf{x}}_t^{(j)}$ is a function of $\mathbf{y}_{0:t}^{(jj)}$, $\{\mathbf{y}_{0:t}^{(ji)} : i \in \mathcal{N}^{(j)}\}$, and $\{\mathbf{r}_{0:t}^{(ij)} : i \in \mathcal{N}^{(j)}\}$. The metric $e_t^{(j)}$ defined in 36. is adopted for the inference error of agent j at time step t.

A necessary condition and a sufficient condition for the inference error of all the agents to be bounded over time, i.e., 37. holds for all $j \in \mathcal{V}_\mathrm{a}$, are presented in (105). Both the necessary condition and the sufficient condition consist of a sensing sub-condition and a communication sub-condition. The sensing sub-condition in the necessary condition is the same as that in the sufficient condition. In particular, this sub-condition is stated in terms of the sensor gain matrices of nodes in the network. On the other hand, the communication sub-condition is stated in terms of the Shannon capacities and anytime capacities of channels in the network, respectively, in the necessary condition and in the sufficient condition. The gap between the established necessary condition and sufficient condition is discussed in (105) and it is shown that such a gap is small in certain scenarios.

8.2.1. Decentralized estimation in continuous-time. Results for decentralized inference are also established for continuous-time scenarios. Specifically, consider a system consisting of nodes i and j. The unknown state of node j at time t is represented by a random variable $\mathrm{x}_t^{(j)}$, which satisfies the following stochastic differential equation (SDE):

$$\mathrm{d}\mathrm{x}_t^{(j)} = A^{(j)}\mathrm{x}_t^{(j)}\,\mathrm{d}t + \boldsymbol{B}^{(j)}\,\mathrm{d}\mathbf{v}_t^{(j)}, \qquad t \in [0, \infty)$$

where $A^{(j)} \in \mathbb{R}$ satisfies $|A^{(j)}| \geqslant 1$ and is known to both nodes. Quantity $\boldsymbol{B}^{(j)}$ is a row vector and is also known to the two nodes. Process $\{\mathbf{v}_t^{(j)}\}_{t\geqslant 0}$ is a Brownian motion corresponding to the disturbance to the state of node j.

Node i obtains an observation of the node j's state at each time. The observation obtained by node i at time t is represented by a random vector $\mathbf{y}_t$, which satisfies

$$\mathrm{d}\mathbf{y}_t = \boldsymbol{C}\mathrm{x}_t^{(j)}\,\mathrm{d}t + \boldsymbol{\Xi}\,\mathrm{d}\mathbf{w}_t, \qquad t \in [0, \infty)$$

where the sensor gain vector $\boldsymbol{C}$ and matrix $\boldsymbol{\Xi}$ are deterministic and are known to both nodes. Process $\{\mathbf{w}_t\}_{t\geqslant 0}$ is a Brownian motion corresponding to the observation noise. Moreover,

[2]The subset $\mathcal{V}_\mathrm{a}$ can be chosen arbitrarily, from a singleton $\{j\}$ to the entire set $\mathcal{V}$.

node i generates an encoded message m_t at each time t and transmits the message via a scalar Gaussian channel with noiseless feedback. The message m_t is a function of $\{\mathbf{y}_\tau\}_{\tau\in[0,t]}$ and $\{\mathsf{r}_\tau\}_{\tau\in[0,t]}$, where r_τ represents the message received by node j at time τ. Specifically, r_t satisfies

$$\mathrm{d}\mathsf{r}_t = \mathsf{m}_t\,\mathrm{d}t + \kappa\,\mathrm{d}\mathsf{w}_t\,, \qquad t\in[0,\infty)$$

where κ is a known scalar, and $\{\mathsf{w}_t\}_{t\geqslant 0}$ is a one-dimensional Brownian motion corresponding to the additive Gaussian noise in the channel.

Node j evaluates a decentralized estimator $\hat{\mathsf{x}}_t^{(j)}$ of its state $\mathsf{x}_t^{(j)}$ at each time t. Consider the MSE $e_t^{(j)} := \mathbb{E}\Big\{\big\|\hat{\mathsf{x}}_t^{(j)} - \mathsf{x}_t^{(j)}\big\|^2\Big\}$ as the metric for the inference error at time t. The work in (104) establishes a necessary and sufficient condition under which $e_t^{(j)}$ is bounded over time, namely there exist an encoder at node i and a decentralized estimator at node j that achieve 37. if and only if the Shannon capacity C of the channel satisfies $C > \log(|A|)$.

Decentralized inference has also been studied from an information-theoretical perspective in (104, 105). Specifically, building on the pioneering work of Mitter and Newton (107, 108), a relationship between mutual information and Fisher information is established for decentralized inference.

9. Discussion and outlook on the field

Research at the intersection of information theory and control theory has now entered its third decade and we can reflect on the body of knowledge that has been developed so far, as well as on the new challenges that we see appearing at the horizon. Examining the large body of research conducted, we draw the following basic conclusions. First, there is the realization that in order to describe the ability to stabilize a NCS, a notion of information capacity must be related to system parameters expressing the "value" or "utility" of the information available for control. Second, we have that the information produced by the dynamical system can be quantified in terms of intrinsic entropy of the system. Since the intrinsic entropy can grow exponentially over time whenever there is no information available at the decoder, or in the presence of decoding errors, we compare the intrinsic entropy of the system to a notion of information capacity available through the system loop that takes into account these dynamics. It follows that, depending on the notion of stability employed, we can use different capacity notions that range from the Shannon one, to the anytime and to the zero-error ones. Third, there is the realization that in NCS information useful for control can be transmitted not only through data packets but also through events that occur over time, like in the case of event-triggering strategies, and in this case capacity notions should include timing information. Finally, there is a certain duality between results obtained in the context of stabilization and the ones obtained in the context of estimation. In this latter case, recent advancements have also shown results for multiple-node networks.

Moving forward, we expect that a complete theory of communication over feedback loops can be constructed by considering encoding and decoding strategies accounting for both communication by timing and data payload, as well as accounting for the distributed nature of many system implementations. This theory will make impact on practical developments that will take into account the information constraints that need to be satisfied to achieve different objectives. Some experimental platforms have already demonstrated the applicability of the theoretical results obtained so far and we expect more and more impact to emerge in the future as the theory will be used to develop industrial systems.

A grand challenge will be to extend the treatment to distributed networks where partial state information is available at the different nodes. In this case, since network information theory as well as a distributed control theory are not yet fully developed, studying the intersection between the two will present additional challenges. Nevertheless, while recognizing that much needs to be done, given the amount of progress we have witnessed in the last two decades, we can look at the future with a positive outlook and dare to say: much *will* be done.

ACKNOWLEDGMENTS

The research described in this paper was supported in part by NSF awards ECCS-2127946 CNS-1446891 and ECCS-1917177 and in part by the Army Research Office through the MIT Institute for Soldier Nanotechnologies under Contract W911NF-13-D-0001.

LITERATURE CITED

1. Kim KD, Kumar PR. 2012. Cyber-physical systems: A perspective at the centennial. *Proc. IEEE* 100 (Special Centennial Issue):1287–1308
2. Sun Y, Kadota I, Talak R, Modiano E. 2019. Age of information: A new metric for information freshness. *Synthesis Lectures on Communication Networks* 12(2):1–224
3. Yates RD, Sun Y, Brown DR, Kaul SK, Modiano E, Ulukus S. 2021. Age of information: An introduction and survey. *IEEE J. Sel. Areas Commun.* 39(5):1183–1210
4. Yates RD. 2020. The age of information in networks: Moments, distributions, and sampling. *IEEE Trans. Inf. Theory* 66(9):5712–5728
5. Talak R, Karaman S, Modiano E. 2020. Improving age of information in wireless networks with perfect channel state information. *IEEE/ACM Trans. Netw.* 28(4):1765–1778
6. Soleymani T, Baras JS, Hirche S. 2022. Value of information in feedback control: Quantification. *IEEE Trans. Autom. Control* 67(7):3730–3737
7. Soleymani T, Baras JS, Hirche S, Johansson KH. 2022. Value of information in feedback control: Global optimality. *IEEE Trans. Autom. Control* To appear
8. Uysal E, Kaya O, Ephremides A, Gross J, Codreanu M, et al. 2022. Semantic communications in networked systems: A data significance perspective. *IEEE Network* To appear
9. Franceschetti M, Minero P. 2014. Elements of information theory for networked control systems. In *Information and Control in Networks*. Springer
10. Nair BGN, Fagnani F, Zampieri S, Evans RJ. 2007. Feedback control under data rate constraints: An overview. *Proc. IEEE* 95(1):108–137
11. Liberzon D. 2009. Nonlinear control with limited information. *Communications in Information & Systems* 9(1):41–58
12. Colonius F, Helmke U, Jordan J, Kawan C, Sailer R, Wirth F. 2014. Analysis of networked systems. In *Control Theory of Digitally Networked Dynamic Systems*. Springer
13. Hespanha JP, Naghshtabrizi P, Xu Y. 2007. A survey of recent results in networked control systems. *Proc. IEEE* 95(1):138–162
14. Yüksel S, Başar T. 2013. *Stochastic Networked Control Systems: Stabilization and Optimization under Information Constraints*. Springer Science & Business Media
15. Matveev AS, Savkin AV. 2009. *Estimation and control over communication networks*. Springer Science & Business Media
16. Fang S, Chen J, Ishii H. 2017. *Towards integrating control and information theories*. Springer
17. Kawan C. 2013. Invariance entropy for deterministic control systems. *Lecture Notes in Mathematics* 2089
18. Fischer T. 1982. Optimal quantized control. *IEEE Trans. Autom. Control* 27(4):996–998

19. Tatikonda S, Sahai A, Mitter S. 2004. Stochastic linear control over a communication channel. *IEEE Trans. Autom. Control* 49(9):1549–1561
20. Khina A, Garding ER, Pettersson GM, Kostina V, Hassibi B. 2019. Control over Gaussian channels with and without source-channel separation. *IEEE Trans. Autom. Control* 64(9):3690–3705
21. Kostina V, Hassibi B. 2019. Rate-cost tradeoffs in control. *IEEE Trans. Autom. Control* 64(11):4525–4540
22. Olfati-Saber R, Fax JA, Murray RM. 2007. Consensus and cooperation in networked multi-agent systems. *Proc. IEEE* 95(1):215–233
23. Nair GN, Evans RJ. 2004. Stabilizability of stochastic linear systems with finite feedback data rates. *SIAM J Control Optim* 43(2):413–436
24. Tatikonda S, Mitter SK. 2004. Control under communication constraints. *IEEE Trans. Autom. Control* 49(7):1056–1068
25. Hespanha J, Ortega A, Vasudevan L. 2002. *Towards the control of linear systems with minimum bit-rate*. In *Proc. 15th Int. Symp. on Mathematical Theory of Networks and Systems*, pp. 1–15
26. Baillieul J. 1999. *Feedback designs for controlling device arrays with communication channel bandwidth constraints*. In *ARO Workshop on Smart Structures*, pp. 16–18. Penn. State U.
27. Baillieul J. Springer, 2001. *Feedback designs in information-based control*. In *B. Pasik-Duncan (ed.), Proceedings of the Workshop on Stochastic Theory and Control*. Lawrence, Kansas
28. Wong WS, Brockett R. 1997. Systems with finite communication bandwidth constraints. I. State estimation problems. *IEEE Trans. Autom. Control* 42(9):1294–1299
29. Wong WS, Brockett RW. 1999. Systems with finite communication bandwidth constraints. II. stabilization with limited information feedback. *IEEE Trans. Autom. Control* 44(5):1049–1053
30. Brockett R, Liberzon D. 2000. Quantized feedback stabilization of linear systems. *IEEE Trans. Autom. Control* 45(7):1279–1289
31. Elia N, Mitter SK. 2001. Stabilization of linear systems with limited information. *IEEE Trans. Autom. Control* 46(9):1384–1400
32. Liberzon D. 2003. On stabilization of linear systems with limited information. *IEEE Trans. Autom. Control* 48(2):304–307
33. Nair GN, Evans RJ, Mareels IM, Moran W. 2004. Topological feedback entropy and nonlinear stabilization. *IEEE Trans. Autom. Control* 49(9):1585–1597
34. Adler RL, Konheim AG, McAndrew MH. 1965. Topological entropy. *Trans. Amer. Math. Soc.* 114(2):309–319
35. Shannon CE. 1948. A mathematical theory of communication. *Bell System Technical Journal* 27(7):379–423
36. Kolmogorov AN, Tikhomirov VM. 1959. ε-entropy and ε-capacity of sets in function spaces. *Uspekhi Matematicheskikh Nauk* 14(2):3–86
37. Lim TJ, Franceschetti M. 2017. Information without rolling dice. *IEEE Trans. Inf. Theory* 63(3):1349–1363
38. Donoho D. 2000. Wald lecture i: counting bits with shannon and kolmogorov. *Dept. Statist., Stanford Univ., Stanford, CA, USA, Tech. Rep*
39. Hagihara R, Nair GN. 2013. Two extensions of topological feedback entropy. *Math. Cont. Sign. Syst.* 25(4):473–490
40. Liberzon D, Hespanha JP. 2005. Stabilization of nonlinear systems with limited information feedback. *IEEE Trans. Autom. Control* 50(6):910–915
41. Sharon Y, Liberzon D. 2012. Input to state stabilizing controller for systems with coarse quantization. *IEEE Trans. Autom. Control* 57(4):830–844
42. De Persis C. 2005. n-bit stabilization of n-dimensional nonlinear systems in feedforward form. *IEEE Trans. Autom. Control* 50(3):299–311
43. Colonius F, Hamzi B. 2021. Entropy for practical stabilization. *SIAM J. Control Optim.*

59(3):2195–2222
44. Tatikonda S, Mitter SK. 2004. Control over noisy channels. *IEEE Trans. Autom. Control* 49(7):1196–1201
45. Matveev AS, Savkin AV. 2007. An analogue of shannon information theory for detection and stabilization via noisy discrete communication channels. *SIAM J Control Optim* 46(4):1323–1367
46. Sahai A, Mitter SK. 2006. 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* 52(8):3369–3395
47. Matveev AS, Savkin AV. 2007. Shannon zero error capacity in the problems of state estimation and stabilization via noisy communication channels. *Int. J. Control* 80(2):241–255
48. Minero P, Franceschetti M. 2017. Anytime capacity of a class of Markov channels. *IEEE Trans. Autom. Control* 62(3):1356–1367
49. Cover TM, Thomas JA. 2006. *Elements of Information Theory.* Hoboken, NJ: John Wiley & Sons, Inc., 2nd ed.
50. Martins NC, Dahleh MA, Elia N. 2006. Feedback stabilization of uncertain systems in the presence of a direct link. *IEEE Trans. Autom. Control* 51(3):438–447
51. Minero P, Franceschetti M, Dey S, Nair GN. 2009. Data rate theorem for stabilization over time-varying feedback channels. *IEEE Trans. Autom. Control* 54(2):243–255
52. You K, Xie L. 2010. Minimum data rate for mean square stabilizability of linear systems with markovian packet losses. *IEEE Trans. Autom. Control* 56(4):772–785
53. Minero P, Coviello L, Franceschetti M. 2013. Stabilization over Markov feedback channels: the general case. *IEEE Trans. Autom. Control* 58(2):349–362
54. Nair G. 2013. A non-stochastic information theory for communication and state estimation. *IEEE Trans. Autom. Control* 58:1497–1510
55. Ding J, Peres Y, Ranade G, Zhai A. 2019. When multiplicative noise stymies control. *Ann Appl Probab* 29(4):1963–1992
56. Astrom KJ, Bernhardsson BM. 2002. *Comparison of Riemann and Lebesgue sampling for first order stochastic systems.* In *Proc. IEEE Conf. Decision and Control*, vol. 2, pp. 2011–2016
57. Tabuada P. 2007. Event-triggered real-time scheduling of stabilizing control tasks. *IEEE Trans. Autom. Control* 52(9):1680–1685
58. Khashooei BA, Antunes DJ, Heemels W. 2018. A consistent threshold-based policy for event-triggered control. *IEEE Control Syst. Lett* 2(3):447–452
59. Girard A. 2014. Dynamic triggering mechanisms for event-triggered control. *IEEE Trans. Autom. Control* 60(7):1992–1997
60. Wang X, Lemmon MD. 2011. Event-triggering in distributed networked control systems. *IEEE Trans. Autom. Control* 56(3):586
61. Dimarogonas DV, Frazzoli E, Johansson KH. 2012. Distributed event-triggered control for multi-agent systems. *IEEE Trans. Autom. Control* 57(5):1291–1297
62. Demirel B, Gupta V, Quevedo DE, Johansson M. 2017. On the trade-off between communication and control cost in event-triggered dead-beat control. *IEEE Trans. Autom. Control* 62(6):2973–2980
63. Trimpe S, D'Andrea R. 2014. Event-based state estimation with variance-based triggering. *IEEE Trans. Autom. Control* 59(12):3266–3281
64. Seuret A, Prieur C, Tarbouriech S, Zaccarian L. 2016. Lq-based event-triggered controller co-design for saturated linear systems. *Automatica* 74:47–54
65. Umlauft J, Hirche S. 2019. Feedback linearization based on gaussian processes with event-triggered online learning. *IEEE Trans. Autom. Control* 65(10):4154–4169
66. Maity D, Baras JS. 2019. Optimal event-triggered control of nondeterministic linear systems. *IEEE Trans. Autom. Control* 65(2):604–619
67. Heemels W, Johansson KH, Tabuada P. 2012. *An introduction to event-triggered and self-*

triggered control. In *Proc. IEEE Conf. Decision and Control*, pp. 3270–3285
68. Tanwani A, Teel A. 2017. *Stabilization with event-driven controllers over a digital communication channel with random transmissions*. In *Proc. IEEE Conf. Decision and Control*, pp. 6063–6068
69. Lemmon M. 2010. Event-triggered feedback in control, estimation, and optimization. London: Springer London, 293–358
70. Miskowicz M. 2018. *Event-based control and signal processing*. CRC press
71. Tolić D, Hirche S. 2017. *Networked control systems with intermittent feedback*. CRC Press
72. Tallapragada P, Cortés J. 2015. Event-triggered stabilization of linear systems under bounded bit rates. *IEEE Trans. Autom. Control* 61(6):1575–1589
73. Kofman E, Braslavsky JH. 2006. *Level crossing sampling in feedback stabilization under data-rate constraints*. In *Proc. IEEE Conf. Decision and Control*, pp. 4423–4428. IEEE
74. Pearson J, Hespanha JP, Liberzon D. 2017. Control with minimal cost-per-symbol encoding and quasi-optimality of event-based encoders. *IEEE Trans. Autom. Control* 62(5):2286–2301
75. Dhulipala AK, Fragouli C, Orlitsky A. 2009. Silence-based communication. *IEEE Trans. Inf. Theory* 56(1):350–366
76. Khojasteh MJ, Tallapragada P, Cortés J, Franceschetti M. 2019. The value of timing information in event-triggered control. *IEEE Trans. Autom. Control* 65(3):925–940
77. Khojasteh MJ, Hedayatpour M, Cortés J, Franceschetti M. 2021. Exploiting timing information in event-triggered stabilization of linear systems with disturbances. *IEEE Trans. Control Netw. Syst.* 8(1):15–27
78. Khojasteh MJ, Tallapragada P, Cortés J, Franceschetti M. 2017. *Time-triggering versus event-triggering control over communication channels*. In *Proc. IEEE Conf. Decision and Control*, pp. 5432–5437
79. Khojasteh MJ, Hedayatpour M, Franceschetti M. 2019. *Theory and implementation of event-triggered stabilization over digital channels*. In *Proc. IEEE Conf. Decision and Control*, pp. 4183–4188
80. Khojasteh MJ, Franceschetti M, Ranade G. 2018. Stabilizing a linear system using phone calls: when time is information. *arXiv preprint arXiv:1804.00351*
81. Anantharam V, Verdú S. 1996. Bits through queues. *IEEE Trans. Inf. Theory* 42(1):4–18
82. Win MZ, Conti A, Mazuelas S, Shen Y, Gifford WM, et al. 2011. Network localization and navigation via cooperation. *IEEE Commun. Mag.* 49(5):56–62
83. Conti A, Morselli F, Liu Z, Bartoletti S, Mazuelas S, et al. 2021. Location awareness in beyond 5G networks. *IEEE Commun. Mag.* To appear
84. Patwari N, Ash JN, Kyperountas S, Hero AO, Moses RL, Correal NS. 2005. Locating the nodes: Cooperative localization in wireless sensor networks. *IEEE Signal Process. Mag.* 22(4):54–69
85. Chiani M, Giorgetti A, Paolini E. 2018. Sensor radar for object tracking. *Proc. IEEE* 106(6):1022–1041
86. Bartoletti S, Giorgetti A, Win MZ, Conti A. 2015. Blind selection of representative observations for sensor radar networks. *IEEE Trans. Veh. Technol.* 64(4):1388–1400
87. Bartoletti S, Conti A, Giorgetti A, Win MZ. 2014. Sensor radar networks for indoor tracking. *IEEE Wireless Commun. Lett.* 3(2):157–160
88. Cardone G, Foschini L, Bellavista P, Corradi A, Borcea C, et al. 2013. Fostering participaction in smart cities: a geo-social crowdsensing platform. *IEEE Commun. Mag.* 51(6):112–119
89. Bartoletti S, Conti A, Win MZ. 2017. Device-free counting via wideband signals. *IEEE J. Sel. Areas Commun.* 35(5):1163–1174
90. Moreno V, Zamora MA, Skarmeta AF. 2016. A low-cost indoor localization system for energy sustainability in smart buildings. *IEEE Sensors J.* 16(9):3246–3262
91. Atzori L, Iera A, Morabito G. 2010. The Internet of Things: A survey. *Computer Networks* 54(15):2787–2805
92. Win MZ, Meyer F, Liu Z, Dai W, Bartoletti S, Conti A. 2018. Efficient multi-sensor localization

for the Internet of Things. *IEEE Signal Process. Mag.* 35(5):153–167
93. Amadeo M, Campolo C, Quevedo J, Corujo D, Molinaro A, et al. 2016. Information-centric networking for the Internet of Things: Challenges and opportunities. *IEEE Netw.* 30(2):92–100
94. Win MZ, Shen Y, Dai W. 2018. A theoretical foundation of network localization and navigation. *Proc. IEEE* 106(7):1136–1165Special issue on *Foundations and Trends in Localization Technologies*
95. Win MZ, Dai W, Shen Y, Chrisikos G, Poor HV. 2018. Network operation strategies for efficient localization and navigation. *Proc. IEEE* 106(7):1224–1254Special issue on *Foundations and Trends in Localization Technologies*
96. Conti A, Mazuelas S, Bartoletti S, Lindsey WC, Win MZ. 2019. Soft information for localization-of-things. *Proc. IEEE* 107(11):2240–2264
97. Conti A, Guerra M, Dardari D, Decarli N, Win MZ. 2012. Network experimentation for cooperative localization. *IEEE J. Sel. Areas Commun.* 30(2):467–475
98. Savkin AV. 2006. Analysis and synthesis of networked control systems: Topological entropy, observability, robustness and optimal control. *Automatica* 42(1):51–62
99. Matveev A, Pogromsky A. 2016. Observation of nonlinear systems via finite capacity channels: Constructive data rate limits. *Automatica* 70:217–229
100. Liberzon D, Mitra S. 2017. Entropy and minimal bit rates for state estimation and model detection. *IEEE Trans. Autom. Control* 63(10):3330–3344
101. Yu S, Chen W, Poor HV. 2021. *Timing Side Information Aided Real-Time Monitoring of Discrete-Event Systems*. In *Proc. IEEE Global Comm. Conf.*, pp. 1–6
102. Liu Z, Conti A, Mitter SK, Win MZ. 2022. Filtering over non-Gaussian channels: The role of anytime capacity :1–6In revision
103. Liu Z, Conti A, Mitter SK, Win MZ. 2021. *Networked Filtering with Feedback for Discrete-Time Observations*. In *Proc. IEEE Conf. Decision and Control*, pp. 5882–5889. Austin, TX
104. Liu Z, Conti A, Mitter SK, Win MZ. 2022. *Networked Filtering with Feedback for Continuous-Time Observations*. In *American Control Conference*, pp. 1–8. Atlanta, GA
105. Liu Z. 2022. *Decentralized Inference and its Application to Network Localization and Navigation*. Ph.D. thesis, Department of Aeronautics and Astronautics, Massachusetts Institute of Technology, Cambridge, MA. Thesis advisor: Professor Moe Z. Win
106. Vershynin R. 2018. *High-dimensional probability: An introduction with applications in data science*. New York, NY, USA: Cambridge University Press
107. Mitter SK, Newton NJ. 2005. Information and entropy flow in the Kalman-Bucy filter. *J. Stat. Phys.* 118:145–176
108. Newton NJ. 2008. Interactive statistical mechanics and nonlinear filtering. *J. Stat. Phys.* 133(4):711–737