

# Chapter 1

## Elements of Information Theory for Networked Control Systems

Massimo Franceschetti and Paolo Minero

### 1.1 Introduction

Next generation cyber-physical systems [35] will integrate computing, communication, and control technologies, to respond to the increased societal need to build large-scale systems using digital technology and interacting with the physical world. These include energy systems where the generation, transmission, and distribution of energy is made more efficient through the integration of information technologies; transportation systems that integrate intelligent vehicles and intelligent infrastructures; and health care systems where medical devices have high degree of intelligence and interoperability, integrating wireless networking and sensing capabilities.

One of the fundamental characteristics of cyber-physical systems is that communication among computing and physical entities occurs over communication channels of limited bandwidth and is subject to interference and noise. This challenges the standard assumption of classical control theory that communication can be performed instantaneously, reliably, and with infinite precision, and leads to the development of a new theory of networked control systems (NCS) [7, 8, 24, 30].

This chapter complements the surveys [2, 50] that focus on the communication constraints imposed by the network on the ability to estimate and control dynamical systems. We describe in a tutorial style the main ideas and techniques that contributed shaping the field, with particular attention to the connections with Shannon's information theory. A compendium of additional related results can be found in the recent monograph [44], relating results to Kolmogorov's approach to information theory via the concept of topological entropy [1].

---

M. Franceschetti (✉)

Department of Electrical and Computer Engineering, University of California, San Diego,  
CA 92093, USA

e-mail: [massimo@ece.ucsd.edu](mailto:massimo@ece.ucsd.edu)

P. Minero

Department of Electrical Engineering, University of Notre Dame, Notre Dame, IN 46556, USA

e-mail: [pminero@nd.edu](mailto:pminero@nd.edu)

We shall not repeat proofs here that are readily available in the literature, but rather concentrate on providing specific illustrative examples and on bridging between different results, with the objective of outlining the leitmotiv and the central theoretical issues underlying this research area. We also present some new results that were not mentioned in the above works, and draw attention to a recent approach, based on the theory of Markov jump linear systems (MJLS) [15], that can be used to derive in a unified way many earlier results obtained using different techniques. Finally, we give a perspective on the open problems that are the natural candidates for future research in the field.

The rest of the chapter is organized as follows. In the next section, we describe a standard model of NCS. In Sect. 1.3, we present a basic result on the data-rate required in the feedback loop to guarantee system's stabilization. This is an important point of contact between communication and control theories and can be written in various forms. These are illustrated in Sect. 1.4, along with their connections with different notions of information capacity and their associated reliability constraints. Section 1.5 focuses on challenges in the design of suitable error correcting codes to satisfy these constraints. Section 1.6 looks more closely at a specific communication channel, illustrating how the theory of MJLS can be used to recover in a unified way many of the results on system stabilization that appeared in the literature. Finally, Sect. 1.7 discusses some of the problems and challenges that lay ahead.

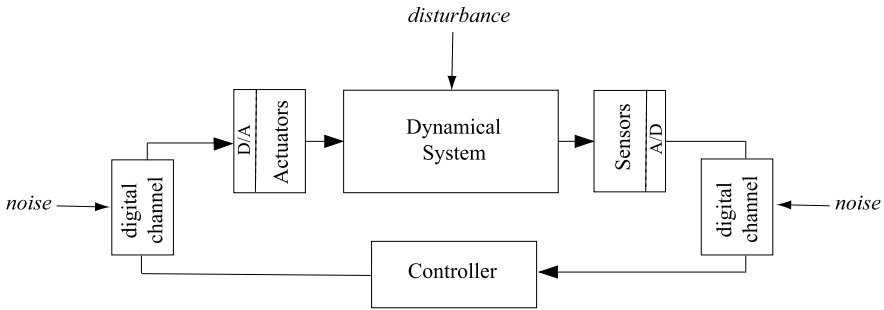
## 1.2 Networked Control Systems

The block diagram of a typical NCS is depicted in Fig. 1.1. The state of a dynamical system evolves over time according to deterministic plant dynamics, possibly affected by stochastic disturbances. Sensors feed back the plant's output to a controller over a digital communication channel. The control action is then sent back to the plant over another digital communication channel for actuation. Communication is affected by noise, and the channel has limited bandwidth as it may be shared among different components in a network setting. This limits the amount of information that can be transferred in the feedback loop at each time step of the evolution of the system.

A natural mathematical abstraction of the above scenario considers the plant to be a discrete-time, linear, dynamical system, affected by additive disturbances

$$\begin{cases} x_{k+1} = Ax_k + Bu_k + v_k, & (1.1a) \\ y_k = Cx_k + w_k, & (1.1b) \end{cases}$$

where  $k = 0, 1, \dots$  is time,  $x_k \in \mathbb{R}^d$  represents the state variable of the system,  $u_k \in \mathbb{R}^m$  is the control input,  $v_k \in \mathbb{R}^d$  is an additive disturbance,  $y_k \in \mathbb{R}^p$  is the sensor measurement,  $w_k \in \mathbb{R}^p$  is the measurement disturbance, and  $A, B, C$  are constant real matrices of matching dimensions. Standard conditions on  $(A, B)$  to



**Fig. 1.1** Feedback loop model of a networked control system

be reachable,  $(C, A)$  observable, are added to make the problems considered well-posed.

In a first approximation, noise and bandwidth limitations in the communication channels can be captured by modeling the channels as “bit pipes” capable of transmitting only a fixed number  $r$  of bits in each time slot of the system’s evolution. In this way, each channel can represent a network connection with a limited available bit-rate. This approach was originally proposed in [10] in the context of linear quadratic Gaussian (LQG) control of *stable* dynamical systems. In this case, by sending to the controller a quantized version of the innovation step of the minimum variance estimator, it was shown that the separation principle between estimation and control holds, and the optimal controller is a linear function of the state. Hence, the estimation problem is formally equivalent to the control one. Extensions of this result to LQG control of unstable systems and to other kind of channel models are highly dependent on the information pattern available to the sender and receiver and are explored in [26, 27, 56, 66]. In particular, when channel errors make the encoder uncertain of what the decoder received, the optimal controller is in general nonlinear [56], a result reminiscent of Witsenhausen’s famous counter example [67].

### 1.3 The Data-Rate Theorem

For unstable systems under the bit-pipe communication model, when the control objective is to keep the state of the system bounded, or asymptotically drive it to zero, the control law is always a linear function of the state, and the central issue is to characterize the ability to perform a reliable estimate of the state at the receiving end of the communication channel. The central result in this case is the *data-rate theorem*. Loosely speaking, this states that the information rate  $r$  supported by the channel to keep the system stable 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,

$$r > \sum_{i \in \mathcal{U}} \log |\lambda_i|, \quad (1.2)$$

where the  $\mathcal{U}$  is the set of indexes of the unstable eigenvalues of the open loop system and the logarithm is base 2. In the simple setting considered, the result is oblivious to the presence of two communication channels between the sensor and the controller and between the controller and the actuator. From the perspective of the system, the location of the controller is purely nominal. Since the key issue is communication of a reliable state estimate, the “bottleneck” link determines the effective data rate of the feedback loop. This intuitive reasoning can easily be made rigorous [49, Proposition 2.2]. The situation is, of course, different in the presence of channel uncertainties that, as already mentioned, make the problem highly dependent on the available information pattern at different points in the feedback loop. In this case, (1.2) should be modified using an appropriate notion of information capacity available in the feedback loop that depends, as we shall see, on the particular notion of stability employed, and on the characteristics of the disturbance.

The intuition behind the data-rate theorem is evident by considering the scalar case of (1.2)

$$2^r > |\lambda|, \quad (1.3)$$

and noticing that while the squared volume of the state of the open loop system increases by  $|\lambda|^2$  at each time step, in closed loop this expansion is compensated by a factor  $2^{-2r}$  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, the result follows. Another interpretation arises if one identifies the logarithm of the right-hand side of (1.2) 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 transport information as fast as it is produced.

Early incarnations of this fundamental result appeared in [5, 6, 68, 69] where it was shown that the state of an *undisturbed*, scalar, unstable plant with mode  $\lambda$  can be kept bounded if and only if the data rate in the feedback loop is at least  $\log |\lambda|$  bits per unit time. While an improvement of the result from maintaining a bounded state to obtaining a state that asymptotically approaches zero cannot be achieved using a fixed quantizer [18], the works [12, 22, 37] showed that this can be obtained letting the encoder to have memory and using of 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 system disturbances, asymptotic stability can only be guaranteed within the range of the disturbances. Disturbances of unbounded support can drive the state arbitrarily far from zero. In this case, it is possible to guarantee stability only in a weaker, probabilistic sense. The work [65] proved the data-rate theorem for vector systems affected by unknown, but bounded disturbances, while the work [49] proved the data-rate theorem under the weaker condition of stochastic disturbances having unbounded support but a uniformly bounded higher moment, and using the probabilistic notion of mean-square stability. The work in [72] provides a related result by characterizing the limit for the second moment of the state in the infinite time horizon.

Since  $\eta$ -moment stability requires

$$\sup_{k \in \mathbb{N}} \mathbb{E}(\|X_k\|^\eta) < \infty, \quad (1.4)$$

the index  $\eta$  gives an estimate of the quality of the stability attainable: large stabilization errors occur more rarely as  $\eta$  increases and in this sense the system is better stabilized. One interpretation of the results in [49, 65] is that in order to achieve stability in a strong, almost deterministic sense ( $\eta \rightarrow \infty$ ), one needs to assume almost surely bounded disturbances and bounded initial condition; on the other hand, relaxing the condition on stability to the weaker mean-square sense ( $\eta = 2$ ), one can use the weaker assumption of bounded higher moments

$$\exists \varepsilon > 0: \quad \mathbb{E}(\|X_0\|^{2+\varepsilon}) < \infty, \quad \sup_{k \in \mathbb{N}} \mathbb{E}(\|V_k\|^{2+\varepsilon}) < \infty, \quad \sup_{k \in \mathbb{N}} \mathbb{E}(\|W_k\|^{2+\varepsilon}) < \infty. \quad (1.5)$$

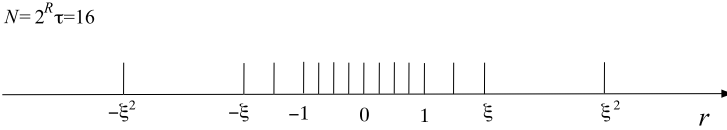
In short, better stability is guaranteed with better behaved disturbances, while “wild disturbances” can only guarantee second moment stability.

The strict necessity of the data-rate theorem is proven in the deterministic setting of bounded disturbances by a recursive argument using the Brunn–Minkowski inequality, which states that the effective radius of the union of two sets is greater than the sum of their effective radii. In the stochastic setting, it is proven using the stochastic counterpart of the inequality, namely the entropy power inequality of information theory which states that the effective variance (entropy power) of the sum of two independent random variables is greater than the sum of their effective variances. The similarity between these two tools is well documented in [14]. In the stochastic case, it is required that the disturbances and the initial state have finite differential entropy.

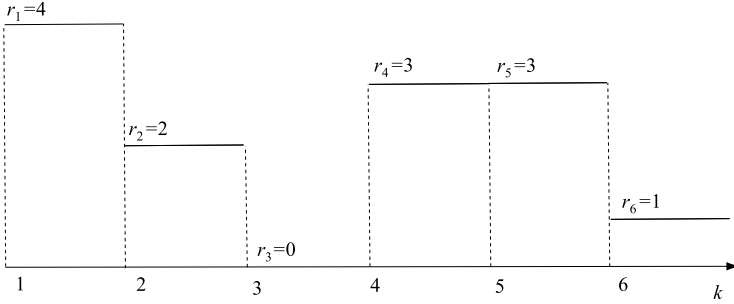
The difficulty in proving the sufficiency of the data-rate theorem in the unbounded support case is due to the uncertainty about the state that cannot be confined in any bounded interval. This is overcome by using an adaptive quantizer depicted in Fig. 1.2 whose number of levels  $N$  depends on the rate process and whose resolution exponentially increases near the origin and diverges far from it, so that it can avoid saturation. The constant  $\xi$  depends on the statistics of the disturbance and it is used to recursively split the open semi-infinite intervals on the real axis into two, while every other finite interval is recursively divided in half. The main idea is then to divide time into cycles of length  $\tau$  and at the beginning of each cycle quantize the estimated state using  $N = 2^R \tau$  levels. Using this strategy, it can be shown that the estimated state satisfies a recurrence of the type

$$\mathbb{E}(\|X_{k\tau}\|^2) \leq c_1 \left( \frac{|\lambda|^2}{2^{2R}} \right)^\tau \mathbb{E}(\|X_{(k-1)\tau}\|^2) + c_2, \quad (1.6)$$

where  $c_1$  and  $c_2$  are constants. This converges in virtue of (1.2) and by choosing  $\tau$  large enough. In practice, the strategy allows the system to evolve in open loop for  $\tau$  time steps and then applies a sufficiently refined control input that makes the state decrease at an exponential rate higher than the exponential divergence rate of the system.



**Fig. 1.2** Adaptive quantizer used to avoid saturation due to unbounded disturbances



**Fig. 1.3** Example of a realization of a stochastic rate channel  $R_k$

## 1.4 Stochastic Time-Varying Channels

### 1.4.1 Stochastic Rate Channel

A different set of extensions concern the stochastic variability of the channel depicted in Fig. 1.3. This can be a first-order approximation of a wireless communication channel where the rate varies randomly in a slotted fashion. When the channel rate varies randomly with time in an independent, identically distributed (i.i.d.) fashion  $R_k \sim R$  and there is causal knowledge of the rate process at both ends of the communication channel, the data-rate theorem for second moment stability in the scalar case becomes

$$|\lambda|^2 \mathbb{E}(2^{-2R}) < 1. \quad (1.7)$$

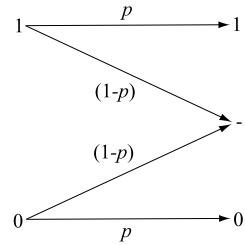
The work [39] proves the result for scalar systems with bounded disturbances and also provides the extension to  $\eta$ -moment stability

$$|\lambda|^\eta \mathbb{E}(2^{-\eta R}) < 1. \quad (1.8)$$

The intuition that to keep the state bounded it is required to balance the expansion of the state variable of the unstable system with the contraction provided by the received information bits still holds. The contraction rate is now a random variable, whose  $\eta$ -moment trades off the  $\eta$ -power of the unstable mode.

The work [46] proves the result for unbounded disturbances and second moment stability, and also provides necessary and sufficient conditions for vector systems that are tight in some special cases. The tools required to prove these results are the

**Fig. 1.4** The binary erasure channel



same as the ones described in the previous section. The additional complication due to the time-varying nature of the channel in the unbounded support case is solved using the idea of *successive refinements*. Namely, at the beginning of each cycle of duration  $\tau$  the quantizer sends an initial estimate of the state using the quantizer depicted in Fig. 1.2, with a resolution dictated by the current value of the rate. In the remaining part of the cycle, the initial estimate is refined using the appropriate quantizer resolution allowed by the channel at each step. The refined state is then used for control at the end of the cycle. Notice that in this case the number of bits per cycle is a random variable dependent on the rate process and the mean square of the state is with respect to both the channel variations and the system disturbances.

The difficulties associated with the vector extension amount to the design of a bit allocation algorithm that dynamically allocates the available rate to the different unstable modes of the plant. The work [46] solves the problem using time-sharing techniques reminiscent of the ones developed in the context of network information theory for the multiple access channel [19]. Some extensions showing the tightness of the construction for some specific class of vector systems are provided in [70].

The stochastic rate channel includes the erasure channel as a special case that corresponds to the rate distribution

$$\begin{cases} \mathbb{P}(R = r) = p, \\ \mathbb{P}(R = 0) = 1 - p. \end{cases} \quad (1.9)$$

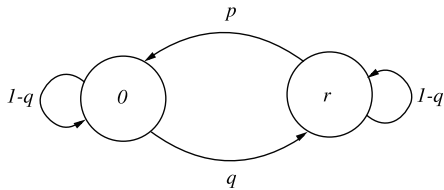
This reduces, for  $r = 1$ , to the binary erasure channel depicted in Fig. 1.4 and, for  $r \rightarrow \infty$ , to the continuous intermittent channel. We explore these reductions in more detail in the next section.

In real networks, many channels exhibit correlations over time. When the rate process follows a two-state Markov chain that corresponds to an erasure channel with two-state memory called the Gilbert–Elliott channel and depicted in Fig. 1.5, the data-rate theorem for mean-square stability in the scalar case with unbounded disturbances becomes [71]

$$r > \frac{1}{2} \log \mathbb{E}(|\lambda|^{2T}), \quad (1.10)$$

where  $T$  is the excursion time of state  $r$ . A more general result is provided in [17] that models the time-varying rate of the channel as an arbitrary time-invariant, positive-recurrent Markov chain of  $n$  states. This allows arbitrary temporal correlations of the channel variations and includes all previous models mentioned above,

**Fig. 1.5** The  $r$ -bit erasure channel with two-state memory (Gilbert–Elliott channel)



including extensions to the vector case. The technique used to provide this extension is based on the theory of MJLS.

In the scalar case, it is shown that stabilizing the system is equivalent to stabilizing

$$z_{k+1} = \frac{\lambda}{2^{r_k}} z_k + c, \quad (1.11)$$

where  $z_k \in \mathbb{R}$  with  $z_0 < \infty$ ,  $c > 0$ ,  $\{R_k\}_{k \geq 0}$  is the Markov rate process whose evolution through one time step is described by the transition probabilities

$$p_{ij} = \mathbb{P}\{R_{k+1} = r_j | R_k = r_i\}, \quad (1.12)$$

for all  $k \in \mathbb{N}$  and  $i, j \in \{1, \dots, n\}$ . This equivalent MJLS describes the stochastic evolution of the estimation error  $\|x_k - \hat{x}_k\|$  at the decoder, which at every time step  $k$  increases by  $\lambda$  because of the system dynamics, and is reduced by  $2^{R_k}$  because of the information sent across the channel. A tight condition for second-moment stability is then expressed in terms of the spectral radius of an augmented matrix describing the dynamics of the second moment of this MJLS.

Letting  $H$  be the  $n \times n$  matrix with elements

$$h_{ij} = \frac{p_{ij}}{2^{2r_j}}, \quad (1.13)$$

with spectral radius  $\rho(H)$ , the data-rate theorem becomes

$$|\lambda|^2 < \frac{1}{\rho(H)}. \quad (1.14)$$

A similar approach provides stability conditions for the case of vector systems. Necessary conditions use the idea of a “genie”-aided proof. First, it is assumed that a genie helps the channel decoder by stabilizing a subset of the unstable states. Then, the stability of the reduced vector system is related to the one of a scalar MJLS whose evolution depends on the remaining unstable modes. By considering all possible subsets of unstable modes, a family of conditions is obtained that relate the degree of instability of the system to the parameters governing the rate process. On the other hand, a sufficient condition for mean-square stability is given using a control scheme in which each unstable component of the system is quantized using a separate scalar quantizer. A bit allocation function determines how the bits available for communication over the Markov feedback channel are distributed among



the various unstable sub-systems. Given a bit allocation function, the sufficient condition is then given as the intersection of the stability conditions for the scalar jump linear systems that describe the evolution of the estimation error for each unstable mode.

The data-rate theorem for general Markovian rates presented in [17] recovers all results in [39, 46, 49, 65, 71] for constant, i.i.d., or two-state Markov data rates, with bounded or unbounded disturbances, in the scalar or vector cases. In addition, it also recovers results for the intermittent continuous channel and for the erasure channel, as discussed next. We discuss the techniques used to derive the results using the theory of MJLS in more detail in Sect. 1.6.

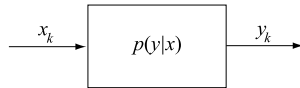
### 1.4.2 Intermittent Channel

The study of the intermittent continuous channel for estimation of the state of a dynamical system first initiated in [48]. The study of this channel was boosted in more recent times by the paper [61] in the context of Kalman filtering with intermittent observations. This work was inspired by computer networks in which packets can be dropped randomly and are sufficiently long that can be thought as representing real, continuous values. The analysis does not involve quantization, but only erasures occurring at each time step of the evolution of the system. Hence, the system in Fig. 1.1 is observed “intermittently”, through an analog, rather than digital channel, and  $y_k$  in (1.1a), (1.1b) can be lost, with some probability, at each time step  $k$ . Similar to the data-rate theorem, it is of interest to characterize the *critical packet loss probability*, defined in [61], above which the mean-square estimation error remains bounded and below which it grows unbounded. This threshold value depends, once again, on the unstable modes of the system. Extensions providing large deviation bounds on the error covariance and conditions on its weak convergence to a stationary distribution are given in [47, 59, 62].

The model is easily extended to stabilization and control by considering an intermittent continuous channel also between the controller and the actuator. The work [56] considers LQG control over i.i.d. packet dropping links and shows that in the presence of acknowledgement of received packets the separation between estimation and control holds and the optimal controller is a linear function of the state. On the other hand, when there is uncertainty regarding the delivery of the packet, the optimal control is in general nonlinear. Similar results in the slightly more restrictive setting of the system being fully observable and the disturbance affecting only the system and not the observation, also appear in [32]. The critical role of the available information pattern on the optimal control is well known [67] and is further explored for stochastic rate channel models in [66].

The critical packet loss probability for mean-square stabilization is characterized in [26], under the assumption of i.i.d. erasures, and in [28] in the case of Markov erasures. The work [21] shows that such critical packet loss probabilities can be obtained as a solution of a robust control synthesis problem. These results can also

**Fig. 1.6** The discrete memoryless channel



be obtained from the stochastic rate channel model, considering the erasure channel in (1.9) and letting  $r \rightarrow \infty$ . An easy derivation of the critical packet loss probability for stabilization is obtained in the scalar case by evaluating the expectation in (1.7), immediately yielding the result in [26]

$$p < \frac{1}{\lambda^2}. \quad (1.15)$$

Similarly, evaluating the condition in [71] for the Gilbert–Elliott channel as  $r \rightarrow \infty$ , one recovers the critical probability for the two-state Markov model of [28]. The works [17, 46] give matching reductions for the vector case as well. The latter of these works considers the most general channel model described so far, being an arbitrary Markov chain of  $n$  states, where  $r$  can be as low as zero (erasure) and as high as  $\infty$  (continuous channel).

### 1.4.3 Discrete Memoryless Channels

Information theory treats the communication channel as a stochastic system described by the conditional probability distribution of the channel output under the given input. Figure 1.6 gives a visual representation of this information-theoretic model for the discrete memoryless channel (DMC).

In this context, the Shannon capacity of the channel is the supremum of the achievable rates of transmissions with an arbitrarily small error probability. It follows that the erasure channel of bit-rate  $r$  described previously is a special case of the DMC and has Shannon capacity [16]

$$C = (1 - p)r. \quad (1.16)$$

In the presence of system disturbances, for the erasure channel it follows from (1.7) that to ensure second moment stability a necessary and sufficient condition is

$$|\lambda|^2(2^{-2r}(1 - p) + p) < 1. \quad (1.17)$$

Comparing (1.16) with (1.17), it is evident that the Shannon capacity does not capture the ability to stabilize the system: not only the left-hand side of (1.17) is different from (1.16), but as  $r \rightarrow \infty$  the Shannon capacity of the channel grows unboundedly, while the data-rate condition for stabilization reduces to (1.15) and critically depends on the erasure probability. Despite the infinite channel capacity, the system may be unstable when the erasure probability is high.

The reason for the insufficiency of the of Shannon capacity to characterize the trade-off between communication and information rate production of a dynamical

system lies in its operational definition. Roughly speaking, the notion of Shannon capacity implies that the message is encoded into a finite length codeword that is then transmitted over the channel. The message is communicated reliably only *asymptotically*, as the length of the codeword transmitted over the channel increases. The probability of decoding the wrong codeword is never zero, but it approaches zero as the length of the code increases. This asymptotic notion clashes with the dynamic nature of the system. A very large Shannon capacity can be useless from the system's perspective if it cannot be used in time for control. As argued at the end of Sect. 1.3, the system requires to receive without error a *sufficiently refined* control signal *every time*  $\tau$  that makes the state decrease by a factor exponential in  $\tau$ . The ability to receive a control input without error in a given time interval can be characterized in a classical information-theoretic setting using the notion of error exponent. However, for the control signal to be effective it must also be appropriate to the current state of the system. The state depends on the history of whether previous codewords were decoded correctly or not, since decoding the wrong codeword implies applying a wrong signal and driving the system away from the stability. In essence, this problem is an example of *interactive communication*, where two-way communication occurs through the feedback loop between the plant and the controller to stabilize the system. Error correcting codes developed in this context have a natural tree structure representing past history [51, 57] and are natural candidates to be used for control over channels with errors. They satisfy more stringent reliability constraints than the ones required to achieve Shannon capacity and can be used, as we shall see in Sect. 1.5, to obtain moment stabilization over the DMC.

Alternative notions of capacity have been proposed to capture the hard reliability constraints dictated by the control problem. The zero-error capacity  $C_0$  was also introduced by Shannon [58] and considers the maximum data rate that can be communicated over the channel with no error. Assuming that the encoder knows the channel output perfectly, this notion of capacity can be used to obtain a data-rate theorem for systems with bounded disturbances with probability one in the form [43]

$$C_0 \gtrsim \sum_{i \in \mathcal{U}} \log |\lambda_i|, \quad (1.18)$$

where we have used the symbol  $\gtrsim$  to indicate that the inequality is strict for the sufficient but not for the necessary condition. It was noted in [43] that even if a feedback channel from decoder to encoder is not available, in the absence of bounded external disturbances “virtual feedback” from decoder to encoder can always be established because the controller affects the plant's motion in a deterministic way and the sensor observes such motion. The controller can then encode its message depending on the observed state motion. For this reason, it is customary in the literature to assume the presence of communication feedback. This assumption is particularly important in the case of (1.18) because, unlike in the classical Shannon capacity, the zero-error capacity of the DMC increases in the presence of feedback.

The insufficiency of classical Shannon capacity to describe stabilization with probability one in the presence of disturbances over erasure channels was first

pointed out in [41], which led to the zero-error capacity framework of [43]. Unfortunately, the zero-error capacity (with or without feedback) of most practical channels (including the erasure channel) is zero [36], which implies that unstable systems cannot keep a bounded state with probability one when controlled over such channels. In practice, a long sequence of decoding errors always arises with probability one, and the small unknown disturbances that accumulate in this long time interval can always drive the system state without bound.

The situation drastically changes for *undisturbed* systems. In this case, the classical Shannon capacity  $C$  can be used to derive a data-rate theorem with probability one in the form [42]

$$C \gtrsim \sum_{i \in \mathcal{U}} \log |\lambda_i|. \quad (1.19)$$

This result was proven for the special case of the erasure channel in [64] and in the more general form for the DMC in [42].

Zero-error capacity and Shannon capacity provide data-rate theorems for plants with and without disturbances, respectively, over the DMC. They both require the strong notion of keeping the state bounded with probability one. Another notion of capacity arises by relaxing the constraint on stabilization with probability one to the weaker constraint of moment stability (1.4) that we used to describe stabilization over stochastic rate channels with unbounded system disturbances. In this case, the data-rate theorem can be written in terms of a parametric notion of channel capacity called *anytime capacity* [52]. Consider a system for information transmission that allows the time for processing the received codeword at the decoder to be infinite, and improves the reliability as time progresses. More precisely, at each step  $k$  in the evolution of the plant a new message  $m_k$  of  $r$  bits is generated that must be sent over the channel. The coder sends a bit over the channel at each  $k$  and the decoder upon reception of the new bit updates the estimates for all messages up to time  $k$ . It follows that at time  $k$  messages

$$m_0, m_1, \dots, m_k$$

are considered for estimation, while estimates

$$\hat{m}_{0|k}, \hat{m}_{1|k}, \dots, \hat{m}_{k|k}$$

are constructed, given all the bits received up to time  $k$ . Hence, the processing operation for any message  $m_i$  continues indefinitely for all  $k \geq i$ . A reliability level  $\alpha$  is achieved in the given transmission system if for all  $k$  the probability that there exists at least one message in the past whose estimate is incorrect decreases  $\alpha$ -exponentially with the number of bits received, namely

$$\mathbb{P}((\hat{M}_{0|k}, \dots, \hat{M}_{d|k}) \neq (M_0, \dots, M_d)) = O(2^{-\alpha d}) \quad \text{for all } d \leq k. \quad (1.20)$$

The described communication system is then characterized by a rate–reliability pair  $(r, \alpha)$ . It turns out that the ability to stabilize a dynamical system depends on the

ability to construct such a communication system, in terms of achievable coding and decoding schemes, with a given rate–reliability constraints.

Let the supremum of the rate  $r$  that can be achieved with reliability  $\alpha$  be the  $\alpha$ -anytime capacity  $C_A(\alpha)$  of a given DMC with channel feedback. The necessary and sufficient condition of the data-rate theorem for  $\eta$ -moment stabilization of a scalar system with bounded disturbances and in the presence of channel output feedback is [53]

$$C_A(\eta \log |\lambda| + \varepsilon) \gtrsim \log |\lambda|. \quad (1.21)$$

Extensions to vector systems appear in preprint form in [54].

The anytime capacity has been introduced as an intermediate quantity between the hard notion of zero-error capacity and the soft notion of Shannon capacity. Not surprisingly, we have

$$C_0 \leq C_A(\alpha) \leq C, \quad (1.22)$$

and in the limiting cases

$$C_A(0^+) = C, \quad C_A(\infty) = C_0. \quad (1.23)$$

Zero-error capacity requires transmission without error. Shannon capacity requires the decoding error go to zero with the length of the code. In the presence of disturbances, only the zero-error capacity can guarantee the almost sure stability of the system. The anytime capacity requires transmission with codeword reliability increasing exponentially in the delay of the single received bit. For scalar systems in presence of bounded disturbances, it is able to characterize the ability to stabilize the system in the weaker  $\eta$ -moment sense [53].

Unfortunately, the anytime capacity can be computed only for the special cases of the erasure channel and the additive white Gaussian noise channel with input power constraint, and in both of these cases it provides data-rate theorems that can also be derived directly in a more classical setting. For the  $r$ -bit erasure channel with feedback, we have

$$C_A(\alpha) = \frac{r\alpha}{\alpha + \log[(1-p)(1-2^\alpha p)^{-1}]}. \quad (1.24)$$

Substituting (1.24) into (1.21), we obtain after some algebra

$$|\lambda|^\eta (2^{-\eta r} (1-p) + p) \lesssim 1. \quad (1.25)$$

Comparing (1.25) with (1.17), it follows that (1.25) is consistent with the result for the stochastic rate channel in [17], which, in fact, gives a stronger version of the anytime capacity data-rate theorem for the case of the erasure channel with feedback, providing a single (necessary and sufficient) strict inequality condition for second moment stability. Furthermore, it also extends the result for this particular channel to disturbances with unbounded support.

For the additive white Gaussian noise channel with input power constraint, the anytime capacity is independent of the reliability level  $\alpha$  and it coincides with the

Shannon capacity. In this case, the data-rate theorem can be given in terms of signal-to-noise ratio and available bandwidth [11, 25, 66].

The anytime capacity of more general channel models remains unknown. In addition, there may be cases in which the output of the noisy channel may not be available at the encoder and is impracticable to use the plant to signal from the decoder to the encoder. In this case, it is only known that the anytime capacity of a DMC without feedback is lower bounded by the exponent  $\beta$  of the error probability of block codes; namely, for any rate  $r < C$  we have

$$C_A(\beta(r) \log_2 e) \geq r \log_2 e. \quad (1.26)$$

The work [53] proposes an ingenious control scheme to achieve (1.26) based on the idea of random binning: the observer maps to state using a time-varying randomly labeled lattice quantizer and outputs a random label for the bin index; the controller, on the other hand, makes use of the common randomness used to select the random bin labels to decode the quantized state value. This proof technique, however, only applies to plants with bounded disturbances.

Despite these shortcomings, the anytime capacity has been influential in the definition of the reliability constraints for the coding–decoding schemes that can achieve moment stabilization of linear systems in the presence of bounded disturbances, thus providing inspiration for further research in coding [13, 51, 60, 63].

#### 1.4.4 Additive Gaussian channels

The additive white Gaussian noise communication channel with power constraint  $P$  is defined as the system

$$y_k = x_k + z_k, \quad (1.27)$$

where  $z_k$  is the realization of an i.i.d. Gaussian process with zero mean and variance  $\sigma^2$ , and the input is constrained by

$$\mathbb{E}(X_k^2) \leq P, \quad \forall k. \quad (1.28)$$

The Shannon capacity of this channel is perhaps the most notorious formula in information theory

$$C = \frac{1}{2} \log(1 + P/\sigma^2). \quad (1.29)$$

In this case, the data-rate theorem for second moment stabilization becomes [11, 25]

$$\frac{P}{\sigma^2} > \prod_{i \in \mathcal{U}} |\lambda_i|^2 - 1, \quad (1.30)$$

that is equivalent to

$$C > \sum_{i \in \mathcal{U}} \log |\lambda_i|. \quad (1.31)$$

The work in [11] also shows that stabilization can be achieved, provided (1.31) holds, using a linear controller with constant gain, if the system's output sent to the controller consists of the entire state vector. If the output consists only of a linear combination of state elements, then the required signal-to-noise ratio for stabilization using linear constant feedback exceeds the bound in (1.30), unless the plant is minimum phase. The work in [25] also shows that (1.31) is also required for second moment stability using nonlinear, time-varying control and provides an explicit lower bound on the second moment of the state that diverges as one approaches the data-rate capacity threshold. Earlier incarnation of these results go back to [66], with slightly stronger assumptions on the available information pattern, and to [20] that connected the recursive capacity-achieving scheme in [55] for the AWGN with feedback to the stabilization problem of scalar systems over AWGN channels.

Extensions to additive *colored* Gaussian channels (ACGC) provide additional connections between the ability to stabilize dynamical systems and the *feedback capacity*  $C_F$  of the channel. This is defined as the capacity, in Shannon's sense, in the presence of an additional noiseless feedback link between the output and the input of the channel. While for the AWGN channel feedback does not improve capacity, for ACGC it does improve it. The feedback capacity of the first order moving average (MA1) additive Gaussian channel has been determined in [33] and for the general case of stationary ACGC in [34]. The work in [45] exploits the result in [33] to show that mean-square stabilization of an undisturbed minimum phase plant with a single unstable pole over a MA1 additive Gaussian channel is possible if and only if

$$C_F > \log |\lambda|. \quad (1.32)$$

The work in [3] exploits the result in [34] to show that the feedback capacity of the general stationary ACGC with power constraint  $P$  is

$$C_F = \sup_{\mathcal{L}} U, \quad (1.33)$$

where

$$U = \sum_{i \in \mathcal{U}} \log |\lambda_i| \quad (1.34)$$

and  $\mathcal{L}$  is the set of all undisturbed (vector) linear systems that can be stabilized using a linear controller over the same additive Gaussian channel, with power constraint

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} |T(e^{j\omega})|^2 S_Z(\omega) d\omega \leq P, \quad (1.35)$$

**Table 1.1** Summary of data-rate theorems for stabilization over noisy channels

Condition	Channel	Stabilization	Disturbance
$C \gtrsim U$	DMC	a.s.	0
$C_0 \gtrsim U$	DMC	a.s.	bounded
$C_A(\eta \log  \lambda ) \gtrsim \eta \log  \lambda $	DMC	$\eta$ -moment	bounded
$ \lambda ^2(2^{-2r}(1-p) + p) < 1$	Erasure	2nd moment	unbounded
$C > U$	AWGN	$\eta$ -moment	unbounded
$C_F = \sup U$	ACGN	2nd moment	0

where  $S_z(\omega)$  is the power spectral density of the noise, and  $T$  is the complementary sensitivity function of the system. This result shows that the maximum “tolerable instability”  $U$  of an LTI system with a given power constraint  $P$ , controlled by a linear controller over a general stationary Gaussian channel, corresponds to the feedback capacity of that channel subject to the same power constraint  $P$ . Hence, there is a natural duality between feedback stabilization and communication over the Gaussian channel. This duality can also be exploited to construct efficient communication schemes over the Gaussian channel with feedback in the context of network information theory, using control tools. This theme was first explored in [20] and later expanded in [4].

We provide a summary of the results for different noisy channels Table 1.1.

## 1.5 Error Correcting Codes for Control

Independent of research in stabilization and control, error correcting codes with exponential reliability constraints in the form of (1.20) were introduced in the context of interactive communication [57]. These codes possess a natural tree structure that can be used to maintain synchronization between the controller and system when communication occurs over noisy channels. Although it is not known whether tree codes are anytime capacity achieving, they can be used for stabilization of networked control systems when their rate-reliability parameters fall within a region needed for stabilization of the given system. We motivate them with the following example.

Consider the problem of tracking a scalar unstable process with dynamics

$$x_{k+1} = \lambda x_k + v_k, \quad (1.36)$$

with  $\lambda > 1$ . The initial condition and the additive disturbance are supposed to be random but bounded, i.e.,  $|X_0| \leq \alpha$  and  $|V_k| \leq \beta$  for some  $\alpha, \beta < \infty$ . We consider the setup where a coder having access to the state communicates over a binary noisy channel to a decoder that wishes to track the state of the system. The objective is to



design a coder–decoder pair such that

$$\sup_k \mathbb{E}(|X_k - \hat{X}_k|^2) < \infty. \quad (1.37)$$

If the communication channel is noiseless and allows transmission without errors of  $r$  bits per unit of time, then we obtain the usual data-rate theorem in the form (1.3). The strategy used for estimation follows the one described in [65]. Let  $\mathcal{U}_0 = [-\alpha, +\alpha]$  denote the set containing the initial condition  $x_0$ . At time  $k = 0$ , the coder and the decoder partition  $\mathcal{U}_0$  into  $2^r$  intervals  $\mathcal{U}_0(1), \dots, \mathcal{U}_0(2^r)$  of equal size. The coder communicates to the decoder the index  $m_0$  of the interval  $\mathcal{U}_0(m_0)$  containing the state, so the decoder can form a state estimate  $\bar{x}_0$  as the midpoint of  $\mathcal{U}_0(m_0)$ . This construction implies

$$|x_0 - \bar{x}_0| \leq \alpha 2^{-r}$$

for any  $x_0 \in \mathcal{U}_0$ . Also, notice that  $x_1$  is contained inside the set  $\mathcal{U}_1 := \lambda \mathcal{U}_0(m_0) + [-\beta, +\beta]$ , where the sum denotes the Minkowski sum of sets. This means that the same scheme can be used at time  $k = 1$  to estimate the state  $x_1$ . Specifically, the coder and the decoder partition the set  $\mathcal{U}_1$  into  $2^r$  intervals  $\mathcal{U}_1(1), \dots, \mathcal{U}_1(2^r)$  of equal size, the coder transmits the index  $m_1$  of the subinterval containing the state, and the decoder sets  $\bar{x}_1$  equal to the midpoint of  $\mathcal{U}_1(m_1)$ , so that

$$|x_1 - \bar{x}_1| \leq \alpha \lambda 2^{-2r} + \beta 2^{-r}.$$

By iterating the same procedure  $k$  times, at time  $k$  the coder and the decoder agree that  $x_k$  belongs to a set  $\mathcal{U}_k := \lambda \mathcal{U}_{k-1}(m_{k-1}) + [-\beta, +\beta]$ . Then, the coder sends over the channel the index  $m_k$  of the subinterval  $\mathcal{U}_k(m_k) \subseteq \mathcal{U}_k$  containing  $x_k$  and the decoder forms an estimate  $\bar{x}_k$  as the midpoint of the uncertainty interval  $\mathcal{U}_k(m_k)$ . It can be shown by induction that

$$|x_k - \bar{x}_k| \leq (\lambda 2^{-r})^k \alpha 2^{-r} + \beta 2^{-r} \sum_{j=0}^{k-1} (\lambda 2^{-r})^{k-1-j}.$$

It follows that a sufficient condition for the estimation error at the decoder to remain bounded for all  $k$  coincides with (1.3).

Consider now the case of a noisy channel in which synchronism between coder and decoder can be lost in the event that the sequence  $m_0, \dots, m_k$  is not correctly decoded at the estimator. To prevent this, at every time  $k$  a channel encoder maps the sequence  $m_0, \dots, m_k$  into an  $r$ -bit channel input sequence  $f_k(m_0, \dots, m_k)$  that is transmitted over the channel. A channel decoder maps the received channel bits up to time  $k$  into an estimate  $\hat{m}_{0|k}, \dots, \hat{m}_{k|k}$  for the input sequence, which, in turn, is used to form the state estimate  $\hat{x}_k$  as the midpoint of the interval  $\mathcal{U}_k(\hat{m}_{k|k})$  which is formed by recursively partitioning  $\lambda \mathcal{U}_j(\hat{m}_{j|k}) + [-\beta, \beta]$ ,  $j = 0, \dots, k-1$ , into  $2^r$  intervals.

If the index of the first wrong estimate at the decoder is  $k-d$ , that is, if  $\hat{m}_{0|k} = m_0, \dots, \hat{m}_{k-d-1|k} = m_{k-d-1}$  and  $\hat{m}_{m-d|k} \neq m_{m-d}$ , then the error between

the estimators at coder and decoder is

$$|\bar{x}_k - \hat{x}_k| = O(\lambda^d), \quad (1.38)$$

because the difference between the two estimates at time  $k - d$  is amplified by  $\lambda$  at each iteration due to the expansion of the state process. It follows that the mean-square estimation error can be upper bounded as

$$\begin{aligned} \mathbb{E}(|X_k - \hat{X}_k|^2) &\leq 2\mathbb{E}(|X_k - \bar{X}_k|^2) + 2\mathbb{E}(|\bar{X}_k - \hat{X}_k|^2) \\ &= O\left(\frac{\lambda^{2k}}{2^{2kr}} + \sum_{d=0}^{k-1} P_{d,k} \lambda^{2d}\right), \end{aligned} \quad (1.39)$$

where

$$P_{d,k} = P\{\hat{M}_{0|k} = M_0, \dots, \hat{M}_{k-d-1|k} = M_{k-d-1}, \hat{M}_{k-d|k} \neq M_{k-d}\},$$

denotes the probability that the index of the first wrong estimate at time  $k$  is  $k - d$ ,  $d = 0, 1, \dots, k$ . Observe that (1.39) is obtained by separately bounding two terms, the first of which represents the mean-square estimation error under the assumption that the channel is noise free, that goes to zero if (1.3) is satisfied, while the second denotes the mean-square error between the estimator  $\bar{x}_k$  available at the encoder and the estimator  $\hat{x}_k$  available at the decoder, and is bounded provided  $P_{d,k}$  decays fast enough as  $d$  grows. It follows that a sufficient condition for second moment stabilization is given by

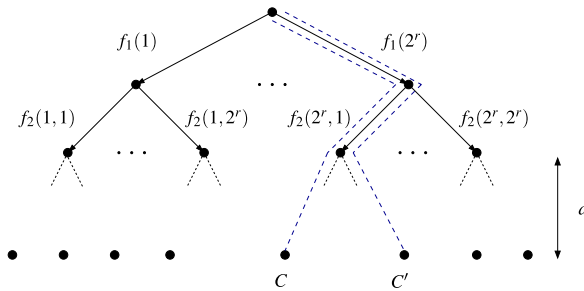
$$\begin{cases} r \geq \log |\lambda|, & (1.40a) \end{cases}$$

$$\begin{cases} P_{d,k} = O(2^{-2(\log |\lambda| + \varepsilon)d}) \quad \text{for all } d \leq k, & (1.40b) \end{cases}$$

that corresponds to the sufficient condition given in (1.21) in terms of anytime capacity.

### 1.5.1 Tree Codes

The reliability condition imposed by (1.40a), (1.40b) is amenable to the following visual interpretation. First, notice that the coding–decoding scheme can be visualized on a tree of depth  $k$ , as depicted in Fig. 1.7, where the nodes at level  $i$  denote the uncertainty intervals  $\mathcal{U}_j(1), \dots, \mathcal{U}_j(2^r)$ , while the label on each branch denotes the  $r$ -bit sequence transmitted over the channel at each time instant. The codeword associated to a given path in the tree is given by the concatenation of the branch symbols along that path. The sequence  $m_0, \dots, m_k$  determines the path in the tree followed up to time  $k$  by the encoder, while  $\hat{m}_0, \dots, \hat{m}_k$  determines the path followed by the decoder. Then, (1.40a), (1.40b) implies that the uncertainty at the controller about



**Fig. 1.7** Binary tree visualizing the evolution of the uncertainty set containing the initial condition. The coding–decoding scheme described in Sect. 1.5 can be visualized on this tree by labeling each branch with the symbols sent over the channel. The codeword associated to a given path is given by the concatenation of the branch symbols along that path

the path followed in the binary tree must decrease exponentially at rate  $2(\log |\lambda| + \varepsilon)$  with the distance  $d$  from the bottom of the tree.

Tree codes and their maximum likelihood analysis were first introduced in [23], but finding explicit deterministic constructions of codes achieving a given rate–reliability pair  $(r, \alpha)$  is still an important open problem. The work [57] applied the random coding argument in [23] to prove the existence of codes within a specific  $(r, \alpha)$  region. The codes introduced in [57] are defined by the property that the Hamming distance between any two codewords associated with distinct paths of equal depth in the binary tree is proportional to the height from the bottom of the tree of the least common ancestor between the two paths. For example, the Hamming distance between the codewords  $C$  and  $C'$  illustrated in Fig. 1.7 should be proportional to  $h$ . This property on the minimum distance translates into different guarantees on the reliability of the code depending on the communication channel. The preprint [63] proves the existence with high probability of *linear*  $(r, \alpha)$  tree codes, i.e., codes where the channel input sequence  $f_k(m_0, \dots, m_k)$  transmitted over the channel at time  $k$  is a linear function of  $m_0, \dots, m_k$ . The  $(r, \alpha)$  region of existence obtained in [63] is currently the largest known region of existence. An important open problem is to show the existence of (possibly nonlinear)  $(2 \log |\lambda|)$ -reliable codes for any rate  $r$  greater than  $\log |\lambda|$ . This result would show that tree codes are anytime-capacity achieving and therefore they are both necessary and sufficient for moment stabilization of unstable scalar systems over noisy channels.

The argument in [57] relies on the probabilistic method and only ensures the existence of tree codes, not their explicit construction. A new class of codes with explicit constructions that are computationally efficient have been presented in [51], but they exhibit weaker reliability constraints that are only useful for stabilization of plants whose state space grows polynomially with time. The preprint [63] offers an explicit construction for the binary erasure channel that does not require causal knowledge of the erasure process, as it was assumed to derive the data-rate theorem in [17].

It is important to emphasize that explicit constructions require coding and decoding operations to be computationally efficient. One could, in principle, consider

using traditional convolutional codes developed in the context of wireless communication to stabilize dynamical systems [38]. These codes perform “on-line” encoding and decoding in which the estimate of the received message is refined as more bits are received within the *constraint length* window of the code. The constraint length is analogous to the block length of traditional block codes, but it allows incremental, on-line refinement of the received message estimate at the decoder. The error probability decreases exponentially with the constraint length of the code, thus providing the required reliability constraint. Unfortunately, the complexity of the construction increases with the constraint length and computationally efficient convolutional codes only exist for small constraint lengths. Convolutional codes are heavily used in mobile phones, where occasional errors translate in call drops or audio disturbances. In control applications, however, the accumulation of errors over long time periods resulting from finite constraint lengths would make them unsuitable for practical implementations as they would drive the system to instability.

## 1.6 Stochastic Time-Varying Rate: An In-Depth Look

We now provide a more rigorous treatment of the data-rate theorem for stochastic time-varying rate channels, with the objective of illustrating recently developed techniques based on the theory of MJLS that can be used to derive many of the results available in the literature. We follow the approach developed in [17]; however, we consider here the special case of a scalar system in which there are only system disturbances and no observation disturbances. This allows presenting simplified proofs that are considerably shorter, more easily accessible, and better suited to grasp the main ideas behind them.

Consider the special case of a scalar system with state feedback

$$\begin{cases} x_{k+1} = \lambda x_k + u_k + v_k, & (1.41a) \\ y_k = x_k, & (1.41b) \end{cases}$$

where  $k = 0, 1, \dots$  and  $|\lambda| \geq 1$ , and suppose that the following assumptions hold:

**Assumption 1.1** The initial condition  $X_0$  and the plant disturbance  $V_k$ ,  $k \geq 0$ , are zero mean and have continuous probability density functions of finite differential entropy, so there exists a constant  $\beta > 0$  such that  $e^{2h(V_k)} \geq \beta$  for all  $k$ .

**Assumption 1.2** The initial condition  $X_0$  and the plant disturbance  $V_k$ ,  $k \geq 0$ , have uniformly bounded  $(2 + \varepsilon)$ th moments so there exists a constant  $\alpha < \infty$  such that  $\mathbb{E}(|V_k|^{2+\varepsilon}) \leq \alpha$  for all  $k$ .

We also assume that the sensor measurements  $y_k$  are transmitted from the state observer to the actuator over a noiseless digital communication link that at each time  $k$  allows transmission without errors of  $r_k$  bits. The rate sequence  $r_0, r_1, \dots$  is the

realization of a stochastic process  $R_1, R_2, \dots$ , that is modeled as a homogeneous positive-recurrent Markov chain taking values in a finite subset of the nonnegative integers

$$\mathcal{R} = \{\bar{r}_1, \dots, \bar{r}_n\},$$

and whose evolution through one time step is described by the transition probabilities (1.12), i.e.,

$$p_{ij} = \mathbb{P}\{R_{k+1} = \bar{r}_j | R_k = \bar{r}_i\}$$

for all  $k \in \mathbb{N}$  and  $i, j \in \{1, \dots, n\}$ . The rate process is independent of the other quantities describing the system and is causally known at observer and controller.

At each time  $k$ , a coding function (coder)  $s_k = s_k(y_0, \dots, y_k)$  maps all past and present measurements into the set  $\{1, \dots, 2^{r_k}\}$ . The digital link is mathematically modeled as the identity function on the set  $\{1, \dots, 2^{r_k}\}$ , so the symbols  $s_k$  are reliably transmitted without distortion. The received channel outputs are transformed by a decoding function (decoder)  $u_k = \hat{x}_k(s_0, \dots, s_k)$  that maps all past and present symbols sent over the digital link into a control input  $u_k$  that is sent to the plant.

The problem is to find conditions on the rate process and the system parameters to ensure stability of the closed loop system. We adopt the probabilistic notion of mean-square stability and require that

$$\sup_k \mathbb{E}[|X_k|^2] < \infty, \quad (1.42)$$

where the expectation is taken with respect to the rate process, the initial condition, and the plant disturbance.

We now proceed to establish necessary and sufficient conditions for mean-square stability of the scalar linear system (1.41a), (1.41a).

**Theorem 1.1** *Let  $H$  be the  $n \times n$  matrix with nonnegative real elements*

$$h_{ij} = \frac{1}{2^{\bar{r}_j}} p_{ji} \quad (1.43)$$

for all  $1 \leq i, j \leq n$ . If Assumption 1.1 holds, then (1.41a), (1.41b) is mean-square stable only if

$$|\lambda|^2 < \frac{1}{\rho(H)}. \quad (1.44)$$

Conversely, if Assumption 1.2 holds, then there exists a coder–decoder pair that stabilizes (1.41a), (1.41b) in mean-square sense if (1.44) is satisfied.

If both Assumptions 1.1 and 1.2 hold, then Theorem (1.1) asserts that condition (1.44) is both necessary and sufficient to ensure mean-square stability. Application of Theorem 1.1 yields the following results as special cases.

- (a) *Constant rate.* When the channel supports a constant rate, i.e., the rate process is identically equal to  $\bar{r}$  at all times, the matrix  $H$  is equal to  $1/2^{2\bar{r}}$  and thus (1.44) reduces to

$$\bar{r} > \log|\lambda|, \quad (1.45)$$

which is the condition given by the data-rate theorem in its basic formulation. It should be remarked that here  $\bar{r}$  is restricted to be an integer, but this assumption can be relaxed by taking the approach followed in [49, 65], where the rate process is allowed to vary deterministically and  $\bar{r}$  is defined as the infinite horizon time-average of the process.

- (b) *Independent rate process.* Consider the special case of an independent rate process where each random variable  $R_k$  in the rate process is identically distributed as a random variable  $R$  with probability mass function  $p_i = \mathbb{P}\{R = \bar{r}_i\}$ ,  $\bar{r}_i \in \mathcal{R}$ . It can be easily seen that in this case  $H$  reduces to a rank-one matrix with only one nonzero eigenvalue equal to  $\sum_{i=1}^n p_i |\lambda|^2 2^{-2\bar{r}_i}$ . Therefore, (1.44) specializes to

$$\begin{aligned} |\lambda|^2 \rho(H) &= \sum_{i=1}^n p_i |\lambda|^2 2^{-2\bar{r}_i} \\ &= \mathbb{E}(|\lambda|^2 2^{-2R}) < 1. \end{aligned} \quad (1.46)$$

The necessity and sufficiency of (1.46) for mean-square stability in this setting was established in [46]. This condition is also a special case of a result in [39], where it is established under the assumption of bounded disturbances that necessary and sufficient condition for  $\eta$ th moment stability, i.e., boundedness of the  $\eta$ th moment of the plant, is  $\mathbb{E}(|\lambda|^\eta 2^{-\eta R}) < 1$ .

- (c) *Two-state Markov process.* Consider the special case of a rate process that randomly switches between two different states, state  $\bar{r}_1$  and  $\bar{r}_2$ , and where the transition probabilities from  $\bar{r}_1$  to  $\bar{r}_2$  and from  $\bar{r}_2$  to  $\bar{r}_1$  are denoted by  $p$  and  $q$ , respectively. In this case, it is possible to relate the spectral radius of  $H$  to its determinant  $\det(H)$  and its trace  $\text{tr}(H)$ . Specifically, the condition in Theorem 1.1 reduces to

$$\frac{|\lambda|^2}{2} (\text{tr}(H) + \sqrt{\text{tr}(H)^2 - 4 \det(H)}) < 1. \quad (1.47)$$

- (d) *Erasure Channel.* Another special case that has been studied in the literature is the case of an erasure channel, which is further specialization of the two-state Markov process described above in the case where  $\bar{r}_1 = 0$ ,  $\bar{r}_2 = \bar{r}$ . Necessary and sufficient conditions for mean-square stability under this channel model were established in [71], for the Markovian case, and in [46, 52] in the special case of independent rate process. If we further specialize to the case where  $\bar{r} \rightarrow \infty$ , then (1.47) recovers a result that was first established in [26].

### 1.6.1 Necessity

The following lemma states that if Assumption 1.1 is satisfied, then the second moment of the state in (1.41a), (1.41b) is lower bounded by the first moment of a MJLS whose dynamics depends on the Markov rate process  $\{R_k\}$  and on the constant  $\beta$  defined in Assumption 1.1.

**Lemma 1.1** *Let Assumption 1.1 hold. Then, for every  $k = 0, 1, \dots$  the second moment of  $X_k$  satisfies*

$$\mathbb{E}(|X_k|^2) > \frac{1}{2\pi e} \mathbb{E}(Z_k),$$

where  $\{Z_k\}$  is a non-homogeneous MJLS with dynamics  $z_0 = e^{2h(X_0)}$  and

$$z_{k+1} = \frac{|\lambda|^2}{2^{2R_k}} z_k + \beta, \quad k = 0, 1, \dots \quad (1.48)$$

*Proof* Let  $S^k = \{S_0, \dots, S_k\}$  denote the symbols transmitted over the digital link up to time  $k$ . By the law of total expectation and the maximum entropy theorem [19], we have

$$\begin{aligned} \mathbb{E}(|X_{k+1}|^2) &= \sum_{s^k} P\{S^k = s^k\} \mathbb{E}(|X_{k+1}|^2 | S^k = s^k) \\ &= \frac{1}{2\pi e} \sum_{s^k} P\{S^k = s^k\} e^{\ln 2\pi e \mathbb{E}(|X_{k+1}|^2 | S^k = s^k)} \\ &\geq \frac{1}{2\pi e} \sum_{s^k} P\{S^k = s^k\} e^{\ln 2\pi e h(X_{k+1} | S^k = s^k)} \\ &=: \frac{1}{2\pi e} \mathbb{E}_{S^k} (e^{2h(X_{k+1} | S^k = s^k)}), \end{aligned} \quad (1.49)$$

where the summation is over  $s_i \in \mathcal{S} := \cup_{r \in \mathcal{R}} \{1, \dots, 2^{2r}\}$ ,  $0 \leq i \leq k$ . It follows that the second moment of the state is lower bounded by the average entropy power of  $X_k$  conditional on  $S^k$ . From the translation invariance property of the differential entropy, the conditional version of entropy power inequality [19], and Assumption 1.1, it follows that

$$\begin{aligned} \mathbb{E}_{S^k} (e^{2h(X_{k+1} | S^k = s^k)}) &= \mathbb{E}_{S^k} (e^{2h(\lambda X_k + \hat{x}(s^k) + V_k | S^k = s^k)}) \\ &\geq \mathbb{E}_{S^k} (e^{2h(\lambda X_k | S^k = s^k)}) + e^{2h(v_k)} \\ &\geq |\lambda|^2 \mathbb{E}_{S^k} (e^{2h(X_k | S^k = s^k)}) + \beta. \end{aligned} \quad (1.50)$$

We can further lower bound (1.50) making use of a result proved in [46, 49], which states that for every time  $k \geq 0$ ,  $s^{k-1} \in \mathcal{S}^{k-1}$ , and  $r \in \mathcal{R}$

$$\sum_{s_k} P\{S^k = s_k \mid S^{k-1} = s^{k-1}, R_k = r\} e^{2h(X_k | S^k = s^k)} \geq \frac{1}{22r} e^{2h(X_k | S^{k-1} = s^{k-1})}, \quad (1.51)$$

where  $S_{-1} := \emptyset$ . By the tower rule of conditional expectation, it then follows that

$$\mathbb{E}_{S^k} (e^{2h(X_k | S^k = s^k)}) \geq \mathbb{E}_{S^{k-1}, R_k} \left( \frac{1}{22R_k} e^{2h(X_k | S^{k-1} = s^{k-1})} \right). \quad (1.52)$$

Combining (1.52) and (1.50) gives

$$\begin{aligned} & \mathbb{E}_{S^k} (e^{2h(X_{k+1} | S^k = s^k)}) \\ & \geq \mathbb{E}_{R_k} \left( \frac{|\lambda|^2}{22R_k} \mathbb{E}_{S^{k-1} | R_k} (e^{2h(X_k | S^{k-1} = s^{k-1})}) \right) + \beta. \end{aligned} \quad (1.53)$$

Following similar steps and using the Markov chain  $S^{k-1} \rightarrow (S^{k-2}, R_{k-1}) \rightarrow R_k$ , we obtain

$$\begin{aligned} & \mathbb{E}_{S^{k-1} | R_k} (e^{2h(X_k | S^{k-1} = s^{k-1})}) \\ & \geq |\lambda|^2 \mathbb{E}_{S^{k-1} | R_k} (e^{2h(X_{k-1} | S^{k-1} = s^{k-1})}) + \beta \\ & \geq \mathbb{E}_{S^{k-2}, R_{k-1} | R_k} \left( \frac{|\lambda|^2}{22R_{k-1}} e^{2h(X_{k-1} | S^{k-2} = s^{k-2})} \right) + \beta \\ & = \mathbb{E}_{R_{k-1} | R_k} \left( \frac{|\lambda|^2}{22R_{k-1}} \mathbb{E}_{S^{k-2} | R_{k-1}, R_k} (e^{2h(X_{k-1} | S^{k-2} = s^{k-2})}) \right) + \beta. \end{aligned} \quad (1.54)$$

Substituting (1.54) into (1.53) and re-iterating  $k$  times, it follows that

$$\begin{aligned} & \mathbb{E}_{S^k} (e^{2h(X_{k+1} | S^k = s^k)}) \\ & \geq \mathbb{E}_{R_{k-1}, R_k} \left( \frac{|\lambda|^4}{22(R_{k-1} + R_k)} \mathbb{E}_{S^{k-2} | R_{k-1}, R_k} (e^{2h(X_{k-1} | S^{k-2} = s^{k-2})}) \right) \\ & \quad + \beta \left( 1 + \mathbb{E}_{R_k} \left( \frac{|\lambda|^4}{22R_k} \right) \right) \\ & \geq \mathbb{E}_{R_1, \dots, R_k} \left( \frac{|\lambda|^{2k}}{22(R_1 + \dots + R_k)} \mathbb{E}_{S_1 | R_1, \dots, R_k} (e^{2h(X_1 | S_0 = s_0)}) \right) \\ & \quad + \beta \left( 1 + \sum_{j=2}^k \mathbb{E}_{R_1, \dots, R_k} \left( \frac{|\lambda|^{2(k-j+1)}}{22(R_j + \dots + R_k)} \right) \right) \end{aligned} \quad (1.55)$$



$$= \mathbb{E} \left( \frac{|\lambda|^{2(k+1)}}{2^{2(R_1+\dots+R_k)}} \right) e^{2h(X_0)} + \beta \left( 1 + \sum_{j=1}^k \mathbb{E} \left( \frac{|\lambda|^{2(k-j+1)}}{2^{2(R_j+\dots+R_k)}} \right) \right), \quad (1.56)$$

where (1.55) uses the fact that the initial condition of the state  $X_0$  is independent of the rate process  $R_k$ . By taking the expectation on both sides of (1.48) and iterating  $k$  times, it is easy to see that the right hand side of (1.56) is the first moment of the non-homogeneous MJLS  $z_{k+1}$  with dynamics given in (1.48). Hence, combining (1.53)–(1.56), we conclude that  $\mathbb{E}(|X_k|^2) > \frac{1}{2\pi e} \mathbb{E}(Z_k)$ , which is the claim.  $\square$

Lemma 1.1 shows that the state cannot be mean-square stable if the average of the  $\{Z_k\}$  process is unbounded. Next, we establish that (1.44) is a necessary condition for the first-moment stability of  $\{Z_k\}$ . For every  $k \geq 0$ , let  $\mu_{k,i} = \mathbb{E}[Z_k 1_{\{R_k = \bar{r}_i\}}]$  denote the expectation of  $Z_k$  in the event that the rate at time  $k$  is  $\bar{r}_i$ . Since  $Z_{k+1} \rightarrow R_k \rightarrow R_{k+1}$  form a Markov chain, the following recursion holds for every  $1 \leq i, j \leq n$ :

$$\mu_{k+1,j} = \sum_{i=1}^n \frac{|\lambda|^2}{2^{2\bar{r}_i}} p_{ij} \mu_{k,i} + \beta \sum_{i=1}^n p_{ij} P\{R_k = \bar{r}_i\}, \quad k = 0, 1, \dots$$

It follows that the vector  $\mu_k = (\mu_{k,1}, \dots, \mu_{k,n})^T \in \mathbb{R}^n$  evolves over time according to the linear system

$$\mu_{k+1} = |\lambda|^2 H \mu_k + b_k, \quad k = 0, 1, \dots, \quad (1.57)$$

where  $H$  is the transition probability matrix defined in (1.43) and  $b_k \in \mathbb{R}^n$  is a vector with  $j$ th element equal to  $\beta \sum_{i=1}^n p_{ij} P\{R_k = \bar{r}_i\}$ . Notice that  $\rho(|\lambda|^2 H) < 1$  is a necessary condition to ensure that the linear system (1.57) is stable, i.e.,  $\sup_k \|\mu_k\|_1 < \infty$ . On the other hand, by the law of total probability,  $\mathbb{E}(Z_k) = \sum_{i=1}^n \mu_{k,i} = \|\mu_k\|_1$  and so the plant is mean-square stable only if  $\sup_k \|\mu_k\|_1 < \infty$ . This establishes that (1.44) is a necessary condition for the second moment stability of the plant.

## 1.6.2 Sufficiency

Consider now the system (1.41a), (1.41b) and suppose that Assumption 1.2 is satisfied. In this section, we build a coder–decoder pair that stabilizes the system under the assumption that (1.44) holds. We first describe the adaptive quantizer that is at the base of the constructive scheme. This is based on the construction given in [49].

**Adaptive Quantizer** For any  $r \geq 2$ , the quantizer  $q_r$  proposed in [49] induces the following partition of the real line:

- The set  $[-1, 1]$  is divided into  $2^{r-1}$  intervals of the same length;
- The sets  $(\xi^{i-2}, \xi^{i-1}]$  and  $(-\xi^{i-1}, -\xi^{i-2}]$  are divided into  $2^{r-1-i}$  intervals of the same length, for each  $i \in \{2, \dots, r-1\}$ ;

- The leftmost and rightmost intervals are the semi-open sets  $(-\infty, -\xi^{r-2}]$  and  $(\xi^{r-2}, \infty)$ .

A sketch of the quantizer for  $r = 4$  is depicted in Fig. 1.2. Here  $\xi > 1$  is a parameter that determines the concentration of intervals around the origin. We can see that the width of the quantization regions increases with  $\xi$ , so the partition becomes more spread out as  $\xi$  increases. Given a real number  $x$ , the output value of the quantizer  $q_r(x)$  is the midpoint of the interval in the partition containing  $x$ . In the sequel, we will also make use of the function  $\kappa_r(x)$ , which instead returns the half-length of such interval, such that the quantization error is bounded by  $\kappa_r(x)$ . If  $x$  is in one of the two semi-open sets at the two extremes of the partition, then we set  $q_r(x) = \text{sign}(x)\xi^r$  and  $\kappa_r(x) = \xi^r - \xi^{r-1}$ .

A fundamental property of this construction is that, loosely speaking, the estimation error produced by the mapping  $q_r$  decays exponentially fast  $r$ . The precise statement of this property involves a functional that was first introduced in [49]. For any pair of random variables  $(X, L)$ , where  $L \geq 0$ , let

$$\|X, L\| := \sqrt{\mathbb{E}[L^2 + |X|^{2+\varepsilon}L^{-\varepsilon}]}. \quad (1.58)$$

In [29], it is shown that the non-negative functional  $\|X, L\|$  is a *pseudo-norm* in the space of random vectors  $(X, L) \in \mathbb{R} \times \mathbb{R}_+$  and satisfies the following properties:

- (i) Second moment bound:

$$\mathbb{E}(|X|^2) \leq \|dX, dL\|^2. \quad (1.59)$$

- (ii) Positive homogeneity: For any  $d \geq 0$

$$\|dX, dL\| = d\|X, L\|. \quad (1.60)$$

- (iii) Triangle inequality: For any  $X_1, X_2 \in \mathbb{R}$  and  $L_1, L_2 \geq 0$ ,

$$\|X_1 + X_2, L_1 + L_2\| \leq \|X_1, L_1\| + \|X_2, L_2\|. \quad (1.61)$$

Lemma 5.2 in [49] proves that if  $\xi > 2^{2/\varepsilon}$ , then the average quantization error produced by  $q_r$  satisfies

$$\left\| X - Lq_r\left(\frac{X}{L}\right), L\kappa_r\left(\frac{X}{L}\right) \right\|^2 \leq \frac{\zeta}{2^{2r}} \|X, L\|^2, \quad (1.62)$$

for some constant  $\zeta > 0$  only determined by  $\varepsilon$  and  $\xi$ .

Another important property of this quantizer is that it is *successively refinable*. Observe in fact that the partition of the  $r$ -bit quantizer can be obtained recursively from the one of the  $(r - 1)$ -bit quantizer by dividing each bounded interval into two intervals of the same length and the two semi-open intervals into two intervals each. In particular, the interval  $(\xi^{r-2}, \infty)$  is divided into the bounded interval  $(\xi^{r-2}, \xi^{r-1}]$  and the semi-open interval  $(\xi^{r-1}, \infty)$ , and similarly for the interval

$(-\infty, -\xi^{r-2}]$ . Thus,  $q_{r+r'}(x)$  can be computed recursively starting from  $q_r(x)$  by repeating the above procedure  $r'$  times. We will make use of this property in our control scheme, where we use the fact that if coder and decoder know  $q_{r_k}(x)$  at time  $k$ , then the coder can communicate to the decoder  $q_{r_k+r_{k+1}}(x)$  by sending  $r_{k+1}$  bits at time  $k+1$ .

The stabilizing scheme can be described as follows. Coder and decoder share at each time  $k$  a state estimator  $\hat{x}_k$  that is recursively updated using the symbols sent over the digital link. Time is divided into cycles of fixed duration  $\tau$ . At the beginning of each cycle, the coder sends a scaled version of the estimation error that is quantized at a resolution dictated by the current value of the rate. In the remaining part of the cycle, the coder sends refinements of the original transmission at a resolution determined by the rate process at each step. At the end of each cycle, the decoder updates the state estimator and sends a control signal to the plant. The scaling factor that is applied to the error prior to quantization is updated at the end of each cycle. The basic idea is to adjust the range of the quantizer as in the zoom-in zoom-out strategy proposed in [37, 69]: the range is increased (zoom-out phase) when atypically large disturbances affect the system, and decreased as the state reduces its size (zoom-in phase). Next, the coder and decoder are described in detail.

**Coder** At the beginning of the  $j$ th cycle, i.e., at time  $j\tau$ , the coder computes

$$q_{r_{j\tau}}((x_{j\tau} - \hat{x}_{j\tau})/l_j), \quad (1.63)$$

where  $l_j$  is the scaling factor updated at the beginning of each cycle, and communicates to the decoder the index  $s_{j\tau} \in \{1, \dots, 2^{r_{j\tau}}\}$  of the quantization interval containing the scaled estimation error. At time  $j\tau + 1$ , coder and decoder divide the quantization interval into  $2^{r_{j\tau+1}}$  subintervals according to the recursive procedure described above. The coder sets  $s_{j\tau+1} \in \{1, \dots, 2^{r_{j\tau+1}}\}$  equal to the subinterval containing  $(x_{j\tau} - \hat{x}_{j\tau})/l_j$ , so the decoder can compute

$$q_{r_{j\tau+r_{j\tau+1}}}((x_{j\tau} - \hat{x}_{j\tau})/l_j).$$

By repeating the same procedure for the rest of the cycle, at time  $(j+1)\tau - 1$  the decoder knows  $(x_{j\tau} - \hat{x}_{j\tau})/l_j$  at the resolution provided by a quantizer with

$$r(j) = r_{j\tau} + \dots + r_{(j+1)\tau-1}$$

bits. Before the beginning of the next cycle, coder and decoder compute

$$\hat{x}_{(j+1)\tau} = \lambda^\tau \left( \hat{x}_{j\tau} + l_j q_{r(j)} \left( \frac{x_{j\tau} - \hat{x}_{j\tau}}{l_j} \right) \right), \quad (1.64)$$

and

$$l_{j+1} = \max \left\{ \varphi, |\lambda|^\tau l_j \kappa_{r(j)} \left( \frac{x_{j\tau} - \hat{x}_{j\tau}}{l_j} \right) \right\}, \quad (1.65)$$

with  $\hat{x}_0 = 0$ ,  $l_0 = \varphi$ , where  $\varphi$  is any constant that only depends on  $\varepsilon$ .

**Decoder** At every time  $k$  the decoder sends to the plant the control signal

$$u_k = \begin{cases} -\lambda \hat{x}_k & \text{if } k = \tau, 2\tau, \dots, \\ 0 & \text{otherwise,} \end{cases} \quad (1.66)$$

where  $\hat{x}_{j\tau}$  is updated as in (1.64) at the beginning of each cycle.

**Analysis** First, we prove that if (1.44) holds, then the second moment of the mean-squared estimation error at the beginning of each cycle is bounded. The following lemma shows that  $\mathbb{E}(|X_{j\tau} - \hat{X}_{j\tau}|^2)$  is lower bounded by the second moment of a MJLS whose dynamics depends on the Markov rate process  $\{R_k\}$  and on the constants  $\alpha$  and  $\varepsilon$  defined in Assumption 1.2.

**Lemma 1.2** *Let Assumption 1.2 hold. Then, for every  $k = 0, 1, \dots$ , the estimation error  $X_{j\tau} - \hat{X}_{j\tau}$  satisfies*

$$\mathbb{E}(|X_{j\tau} - \hat{X}_{j\tau}|^2) \leq \mathbb{E}(Z_{j\tau}^2),$$

where  $\{Z_{j\tau}\}$  is a non-homogeneous MJLS with dynamics

$$z_{(j+1)\tau} = \phi \frac{|\lambda|^\tau}{2^{R_{j\tau} + \dots + R_{(j+1)\tau-1}}} z_{j\tau} + \varsigma, \quad j = 0, 1, \dots, \quad (1.67)$$

for some constants  $z_0 > 0$ ,  $\phi > 1$ , and  $\varsigma > 0$  that are only determined by  $\varepsilon$ ,  $\tau$ , and  $\alpha$ .

*Proof* Let  $e_{j\tau} = x_{j\tau} - \hat{x}_{j\tau}$  denote the estimation error at the beginning each cycle. By (1.59) and the fact that scaling factor  $L_j$  updated by coder and controller at the end of each cycle is nonnegative,

$$\mathbb{E}(|E_{(j+1)\tau}|^2) \leq \|E_{(j+1)\tau}, L_{j+1}\|^2. \quad (1.68)$$

Notice from (1.65) that

$$l_{j+1} \leq |\lambda|^\tau l_j \kappa_{R(j)} \left( \frac{x_{j\tau} - \hat{x}_{j\tau}}{l_j} \right) + \varphi,$$

and that by iteration of (1.41a), (1.41b) and (1.64) for  $\tau$  time steps

$$e_{(j+1)\tau} = |\lambda|^\tau \left( e_{j\tau} - l_j q_{R(j)} \left( \frac{e_{j\tau}}{l_j} \right) \right) + \eta_j,$$

where  $\eta_j := \sum_{i=0}^{\tau-1} \lambda^{\tau-1-i} v_{j\tau+i}$ . Thus, properties (1.60) and (1.61) yield

$$\begin{aligned} \|E_{(j+1)\tau}, L_{j+1}\|^2 &\leq 2|\lambda|^{2\tau} \left\| E_{j\tau} - L_j q_{R(j)} \left( \frac{E_{j\tau}}{L_j} \right), L_j \kappa_{R(j)} \left( \frac{X_{j\tau} - \hat{X}_{j\tau}}{L_j} \right) \right\|^2 \\ &\quad + 2\|H_j, \varphi\|^2. \end{aligned} \quad (1.69)$$

Notice that  $\|H_j, \varphi\|^2$  is upper bounded by a constant  $\varsigma^2$  that only depends on  $\varepsilon$ ,  $\tau$ , and  $\alpha$ .

Let

$$\theta_{j,i} = \mathbb{E}((L_j^2 + |E_{j\tau}|^{2+\varepsilon} L_j^{-\varepsilon}) 1_{\{R_{j\tau}=r_i\}}), \quad i \in \mathcal{R}.$$

Combining (1.62) and (1.69) and making use of the law of total probability,

$$\begin{aligned} \theta_{j+1,i_\tau} &\leq 2\zeta \sum_{i_0} \left( \sum_{i_1, \dots, i_{\tau-1}} \frac{|\lambda|^{2\tau}}{2^{2(r_{i_0} + \dots + r_{i_{\tau-1}})}} p_{i_0, i_1} \cdots p_{i_{\tau-1}, i_\tau} \right) \theta_{j, i_0} \\ &\quad + \varsigma^2 P\{R_{(j+1)\tau} = r_{i_\tau}\}, \end{aligned} \quad (1.70)$$

which provides a recursive formula for the  $\theta_{j,i}$  subsequences.

Next, we claim that, for every  $j \geq 0$ ,

$$\theta_{j+1,i} \leq \mathbb{E}[Z_{(j+1)\tau}^2 1_{\{R_{(j+1)\tau}=r_i\}}], \quad r_i \in \mathcal{R}, \quad (1.71)$$

where the process  $\{Z_{j\tau}\}$  is formed recursively from  $z_0 = \theta_0$  as

$$z_{(j+1)\tau} = \phi \frac{|\lambda|^\tau}{2^{r_{j\tau} + \dots + r_{(j+1)\tau-1}}} z_{j\tau} + \varsigma, \quad j \geq 1, \quad (1.72)$$

where  $\phi = \sqrt{2\zeta} > 1$ . To see this, consider the following inductive argument. By construction  $z_0 = \theta_0$ , hence the claim holds for  $k = 0$ . Now, suppose that the claim is true up to time  $j$ . Then, for any  $r_{i_\tau} \in \mathcal{R}$ ,

$$\begin{aligned} &\mathbb{E}[Z_{(j+1)\tau}^2 1_{\{R_{(j+1)\tau}=r_{i_\tau}\}}] \\ &= \mathbb{E}\left(\left(\sqrt{2\zeta} \frac{|\lambda|^\tau}{2^{R_{j\tau} + \dots + R_{(j+1)\tau-1}}} Z_{j\tau} + \varsigma\right)^2 1_{\{R_{(j+1)\tau}=r_{i_\tau}\}}\right) \\ &\geq \mathbb{E}\left(\left(\sqrt{2\zeta} \frac{|\lambda|^\tau}{2^{R_{j\tau} + \dots + R_{(j+1)\tau-1}}} Z_{j\tau}\right)^2 1_{\{R_{(j+1)\tau}=r_{i_\tau}\}}\right) + \varsigma^2 P\{R_{(j+1)\tau} = r_{i_\tau}\} \\ &= 2\zeta \sum_{i_0, \dots, i_{\tau-1}} \frac{|\lambda|^{2\tau} p_{i_0, i_1} \cdots p_{i_{\tau-1}, i_\tau}}{2^{2(r_{i_0} + \dots + r_{i_{\tau-1}})}} \mathbb{E}[Z_{j\tau}^2 1_{\{R_{j\tau}=r_{i_0}\}}] + \varsigma^2 P\{R_{(j+1)\tau} = r_{i_\tau}\} \\ &\geq 2\zeta \sum_{i_0, \dots, i_{\tau-1}} \frac{|\lambda|^{2\tau}}{2^{2(r_{i_0} + \dots + r_{i_{\tau-1}})}} p_{i_0, i_1} \cdots p_{i_{\tau-1}, i_\tau} \theta_{j, i_0} + \varsigma^2 P\{R_{(j+1)\tau} = r_{i_\tau}\} \\ &\geq \theta_{j+1, i_\tau} \end{aligned}$$

where the first inequality follows from the fact that  $(a + b)^2 \geq a^2 + b^2$  for all nonnegative numbers  $a$  and  $b$ , the second inequality uses the induction hypothesis, while the last inequality uses (1.70). Hence, the claim holds at time  $k + 1$  as well.

Summing both sides of (1.71) over  $r_i \in \mathcal{R}$  and making use of (1.68), it follows that  $\mathbb{E}(E_{j\tau}^2) \leq \mathbb{E}(Z_{j\tau}^2)$ , as claimed.  $\square$

Lemma 1.2 shows that the mean-squared estimation error at the beginning of each cycle is finite if the process  $\{Z_k\}$  is mean-square stable. Next, we establish that (1.44) is a sufficient condition for the second-moment stability  $\{Z_k\}$ . Let  $\sigma_{k,i}^2 = \mathbb{E}[Z_k^2 1_{\{R_k = \bar{r}_i\}}]$  denote the second moment of  $Z_k$  in the event that the rate at time  $k$  takes value  $\bar{r}_i$ . Making use of the fact that  $(a + b)^2 \leq 2(a^2 + b^2)$ , it can be verified that the vector  $\sigma_k^2 = (\sigma_{k,1}^2, \dots, \sigma_{k,n}^2)^T \in \mathbb{R}^n$  satisfies

$$\sigma_{k+1}^2 \leq 2\phi^2 |\lambda|^{2\tau} H^\tau \sigma_k^2 + 2\varsigma_k^2, \quad k = 0, 1, \dots, \quad (1.73)$$

where  $H$  is the transition probability matrix defined in (1.43) and  $\varsigma_k \in \mathbb{R}^n$  is a vector with the  $i$ th component equal to  $\varsigma P\{R_k = \bar{r}_i\}$ . A sufficient condition for the recursion in (1.73) to be bounded is

$$2\phi^2 (|\lambda|^2 \rho(H))^\tau < 1. \quad (1.74)$$

Since by the law of total probability  $\mathbb{E}(|Z_k|^2) \leq \sum_{i=1}^n \sigma_{k,i}^2 = \|\sigma_k^2\|_1$ , it follows that (1.74) is a sufficient condition for  $Z_k$  to be mean-square stable. On the other hand, if the condition of Theorem 1.1 is satisfied, that is, if  $|\lambda|^2 \rho(H) < 1$ , then we can choose the duration of a cycle  $\tau$  large enough to ensure that (1.74) holds and, as a consequence, the second moment of the estimation error at the beginning of each cycle is bounded. Notice that the choice of a larger  $\tau$  translates into larger oscillations of the system state because, according to our quantization scheme, the system evolves in open loop during a cycle.

Finally, for any  $i = 1, \dots, \tau - 1$ , the triangle inequality implies that  $|x_{j\tau+i}| \leq |\lambda|^i |x_{j\tau} - \hat{x}_{j\tau}| + \sum_{k=0}^{i-1} |\lambda|^{i-1-k} |v_{j\tau+i}|$ , so the state remain bounded at all times. This establishes that (1.44) is a sufficient condition for the second moment stability of the plant.

## 1.7 Conclusion

Understanding the operational mechanism of feedback loops over limited data-rate communication channels will be of utmost importance in the near future, as cyber-physical systems (CPS) continue to impact our society more broadly. This requires the development of a rigorous theory of information transmission for control systems. This theory must identify the trade-offs between the amount of information that can be communicated through the control loop and the ability of achieving the required control objectives.

In the past decade, a number of results appeared in the literature, but much remains to be done. Obtained results show that the control objective is fundamentally limited by *both* the channel noise and the intrinsic system noise that affects the plant

in the form of external disturbances. For channels that allow transmission of a given number of bits without error, the “quality” of the achievable stabilization in terms of moment constraints depends on the corresponding constraints on the noise process disturbances. Loosely speaking, better stability can only be guaranteed with better behaved disturbances, while “wild disturbances” can only guarantee lower moment stability. In all cases, the region where the system can be stabilized is clearly demarcated by a data-rate theorem relating the amount of instability of the system to the available communication rate.

For noisy channels, the quality of the stabilization depends on the notion of channel capacity employed. Zero-error capacity, guaranteeing reliable transmission without error, allows for almost sure stabilization. Shannon capacity, guaranteeing reliable transmission with error that decays to zero asymptotically, allows for almost sure stabilization only for systems without disturbances. The parametric notion of anytime capacity, with communication reliability stronger than Shannon’s capacity, but weaker than zero-error capacity, can be used to characterize stabilization of disturbed systems in a moment sense. Again, the region where the system can be stabilized is determined by a data-rate theorem written using the appropriate notion of capacity.

For limited rate channels, the theory of MJLS provides a general framework that can be used to develop data-rate theorems characterizing necessary and sufficient conditions for stabilization that hold in a variety of cases, including for the erasure channel, and for the continuous intermittent channel, with or without memory. On the other hand, the study of the DMC with memory in the context of control remains an important open problem.

Beside the formulation of data-rate theorems for different channels and noise models, a field open for further research is error correcting codes for automatic control over noisy channels. For the Gaussian channel, uncoded transmission is sufficient to achieve stabilization when the Shannon capacity is above the threshold dictated by the data-rate theorem, but for the DMC stabilization requires development of error correcting codes with specific rate-reliability constraints dictated by the corresponding data-rate theorem. These constructions are, at present, largely unknown, although recent advancements in tree codes for the erasure channel appear promising.

We conclude this chapter by mentioning some open problems. As remarked in Sect. 1.4, tight conditions for moment stability of a vector system over a time-varying bit pipe link are not known, in general. Even in the simple setting where the process on the feedback link is an i.i.d. process, only partial results are available. All existing works on stability of linear systems under stochastic disturbance of unbounded support focus on the restrictive notion of second-moment stability [17, 46, 49, 70, 71]. The generalization to  $\eta$ -moment stability, which is currently known only in the case where the disturbance is bounded [39, 53], is an open problem. Similarly, most of the existing works assume a perfect channel from the controller to the actuator. The case where both the sensor–controller and the controller–actuator channels are noisy was studied in [73], which provides conditions for second moment stability using Markov stability theory. In general, however, it is not known when the criteria summarized in this chapter continue to hold

after replacing the relevant notion of capacity with the capacity of the bottleneck channel. Our previous work [46] has revealed a connection between stabilization over the intermittent continuous channel and the rate-limited channel. It would be of interest to establish a similar connection in the case of optimal control over finite-capacity channels. Previous works [26, 56] have considered the LQG problem under the network-theoretic approach where packets can be lost, while [9, 31, 40] studied the same problem under the assumption that the feedback channel is a bit pipe with constant rate  $R$ . In order to create a connection between these two lines of work, one would have to formulate an LQG problem over a time-varying bit pipe channel whose rate oscillates independently over time between 0 and  $R$ . As a final remark, notice that the proof techniques used in [53] only apply to plants with bounded disturbances. A question that requires further investigation is to extend the result in [53] to the case of noise with infinite support. A possible approach based on variable rate coding is outlined in [52, 73].

As control systems gradually evolve towards usage of wireless platforms, the developed theory will have a direct applicability in a practical setting. The move towards wireless is dictated by both technological advancements and economic factors, as the cost of “wiring” large CPS can easily dominate development costs. The theory developed so far has shown that existing error correcting codes for wireless communication are not immediately applicable in the context of control, due to their soft reliability constraints that are not sufficient to ensure even low-moment stability for safety critical applications. In the next decades, we will witness a refinement of the theory to gain additional understanding of fundamental limitations, as well as the development of new communication schemes needed to address the growing industrial need for control over noisy channels.

**Acknowledgement** This research was supported by LCCC—Linnaeus Grant VR 2007-8646, Swedish Research Council.

## References

1. Adler, R.L., Konheim, A.G., McAndrew, M.H.: Topological entropy. *Trans. Am. Math. Soc.* **114**, 309–319 (1965)
2. Andrievsky, B., Matveev, A., Fradkov, A.: Control and estimation under information constraints: toward a unified theory of control, computation and communications. *Autom. Remote Control* **71**, 572–633 (2010). Original Russian text in *Autom. Telemekh.* **4**, 34–99 (2010)
3. Ardestanizadeh, E., Franceschetti, M.: Control-theoretic approach to communication with feedback. *IEEE Trans. Autom. Control* (2012)
4. Ardestanizadeh, E., Minero, P., Franceschetti, M.: LQG control approach to Gaussian broadcast channels with feedback. *IEEE Trans. Inf. Theory* **58**(8), 5267–5278 (2012)
5. Baillieul, J.: Feedback designs for controlling device arrays with communication channel bandwidth constraints. In: *ARO Workshop on Smart Structures*, Penn. State U., USA (1999)
6. Baillieul, J.: Feedback designs in information-based control. In: Pasik-Duncan, B. (ed.) *Proceedings of the Workshop on Stochastic Theory and Control*. Springer, Lawrence (2001)
7. Baillieul, J., Antsaklis, P.J. (eds.): Special issue on Networked Control Systems. *IEEE Trans. Autom. Control* **49**(9) (2004)



8. Baillieul, J., Antsaklis, P.J.: Control and communication challenges in networked real-time systems. *Proc. IEEE* **95**(1), 9–28 (2007)
9. Borkar, V., Mitter, S.: LQG Control with Communication Constraints. LIDS-P. Massachusetts Institute of Technology, Laboratory for Information and Decision Systems (1995)
10. Borkar, V., Mitter, S.: LQG Control with Communication Constraints. *Communications, Computation, Control and Signal Processing: A Tribute to Thomas Kailath*. Kluwer Academic, Dordrecht (1997)
11. Braslavsky, J., Middleton, R., Freudenberg, J.: Feedback stabilization over signal-to-noise ratio constrained channels. *IEEE Trans. Autom. Control* **52**(8), 1391–1403 (2007)
12. Brockett, R., Liberzon, D.: Quantized feedback stabilization of linear systems. *IEEE Trans. Autom. Control* **45**(7), 1279–1289 (2000)
13. Como, G., Fagnani, F., Zampieri, S.: Anytime reliable transmission of real-valued information through digital noisy channels. *SIAM J. Control Optim.* **48**(6), 3903–3924 (2010)
14. Costa, M., Cover, T.: On the similarity of the entropy power inequality and the Brunn-Minkowski inequality. *IEEE Trans. Inf. Theory* **30**(6), 837–839 (1984)
15. Costa, O., Fragoso, D., Marques, R.: *Discrete-Time Markov Jump Linear Systems*. Probability and Its Applications. Springer, Berlin (2004)
16. Cover, T., Thomas, J.: *Elements of Information Theory*. Wiley, New York (2006)
17. Coviello, L., Minero, P., Franceschetti, M.: Stabilization over Markov feedback channels: the general case. *IEEE Trans. Autom. Control* **58**(2), 349–362 (2013)
18. Delchamps, D.: Stabilizing a linear system with quantized state feedback. *IEEE Trans. Autom. Control* **35**(8), 916–924 (1990)
19. El Gamal, A., Kim, Y.H.: *Network Information Theory*. Cambridge University Press, Cambridge (2011)
20. Elia, N.: When Bode meets Shannon: control-oriented feedback communication schemes. *IEEE Trans. Autom. Control* **49**(9), 1477–1488 (2004)
21. Elia, N.: Remote stabilization over fading channels. *Syst. Control Lett.* **54**(3), 237–249 (2005)
22. Elia, N., Mitter, S.K.: Stabilization of linear systems with limited information. *IEEE Trans. Autom. Control* **46**(9), 1384–1400 (2001)
23. Forney, G.D.: Convolutional codes II. Maximum-likelihood decoding. *Inf. Control* **25**(3), 222–266 (1974)
24. Franceschetti, M., Javidi, T., Kumar, P.R., Mitter, S.K., Teneketzis, D. (eds.): Special issue on Control and Communications. *IEEE J. Sel. Areas Commun.* **26**(4) (2008)
25. Freudenberg, J., Middleton, R.H., Solo, V.: Stabilization and disturbance attenuation over a Gaussian communication channel. *IEEE Trans. Autom. Control* **55**(3), 795–799 (2010)
26. Gupta, V., Spanos, D., Hassibi, B., Murray, R.M.: Optimal LQG control across packet-dropping links. *Syst. Control Lett.* **56**(6), 439–446 (2007)
27. Gupta, V., Dana, A., Hespanha, J., Murray, R., Hassibi, B.: Data transmission over networks for estimation and control. *IEEE Trans. Autom. Control* **54**(8), 1807–1819 (2009)
28. Gupta, V., Martins, N., Baras, J.: Optimal output feedback control using two remote sensors over erasure channels. *IEEE Trans. Autom. Control* **54**(7), 1463–1476 (2009)
29. Gurt, A., Nair, G.N.: Internal stability of dynamic quantised control for stochastic linear plants. *Automatica* **45**(6), 1387–1396 (2009)
30. Hespanha, J., Naghshtabrizi, P., Xu, Y.: A survey of recent results in networked control systems. *Proc. IEEE* **95**(1), 138–162 (2007)
31. Huang, M., Nair, G., Evans, R.: Finite horizon LQ optimal control and computation with data rate constraints. In: 44th IEEE Conference on Decision and Control, 2005 and 2005 European Control Conference. CDC-ECC'05, pp. 179–184 (2005)
32. Imer, O.C., Yüksel, S., Başar, T.: Optimal control of LTI systems over unreliable communication links. *Automatica* **42**(9), 1429–1439 (2006)
33. Kim, Y.H.: Feedback capacity of the first-order moving average Gaussian channel. *IEEE Trans. Inf. Theory* **52**(7), 3063–3079 (2006)
34. Kim, Y.H.: Feedback capacity of stationary Gaussian channels. *IEEE Trans. Inf. Theory* **56**(1), 57–85 (2010)

35. Kim, K.D., Kumar, P.R.: Cyber-physical systems: a perspective at the centennial. *Proc. IEEE* **100**(13), 1287–1308 (2012)
36. Korner, J., Orlitsky, A.: Zero-error information theory. *IEEE Trans. Inf. Theory* **44**(6), 2207–2229 (1998)
37. Liberzon, D.: On stabilization of linear systems with limited information. *IEEE Trans. Autom. Control* **48**(2), 304–307 (2003)
38. Lin, S., Costello, D.J. Jr.: *Error Control Coding: Fundamentals and Applications*. Prentice Hall, New York (1983). TUB-HH 2413-469 3
39. Martins, N., Dahleh, M., Elia, N.: Feedback stabilization of uncertain systems in the presence of a direct link. *IEEE Trans. Autom. Control* **51**(3), 438–447 (2006)
40. Matveev, A.S., Savkin, A.V.: The problem of LQG optimal control via a limited capacity communication channel. *Syst. Control Lett.* **53**(1), 51–64 (2004)
41. Matveev, A., Savkin, A.: Comments on “control over noisy channels” and relevant negative results. *IEEE Trans. Autom. Control* **50**(12), 2105–2110 (2005)
42. Matveev, A.S., Savkin, A.V.: An analogue of Shannon information theory for detection and stabilization via noisy discrete communication channels. *SIAM J. Control Optim.* **46**(4), 1323–1367 (2007)
43. Matveev, A.S., Savkin, A.V.: Shannon zero error capacity in the problems of state estimation and stabilization via noisy communication channels. *Int. J. Control* **80**(2), 241–255 (2007)
44. Matveev, A., Savkin, A.: *Estimation and Control over Communication Networks*. Birkhäuser, Basel (2009). *Control Engineering*
45. Middleton, R., Rojas, A., Freudenberg, J., Braslavsky, J.: Feedback stabilization over a first order moving average Gaussian noise channel. *IEEE Trans. Autom. Control* **54**(1), 163–167 (2009)
46. Minero, P., Franceschetti, M., Dey, S., Nair, G.: Data rate theorem for stabilization over time-varying feedback channels. *IEEE Trans. Autom. Control* **54**(2), 243–255 (2009)
47. Mo, Y., Sinopoli, B.: Kalman filtering with intermittent observations: tail distribution and critical value. *IEEE Trans. Autom. Control* **57**(3), 677–689 (2012)
48. Nahi, N.: Optimal recursive estimation with uncertain observation. *Automatica* **15**(4), 457–462 (1969)
49. Nair, G.N., Evans, R.J.: Stabilizability of stochastic linear systems with finite feedback data rates. *SIAM J. Control Optim.* **43**(2), 413–436 (2004)
50. Nair, G., Fagnani, F., Zampieri, S., Evans, R.: Feedback control under data rate constraints: an overview. *Proc. IEEE* **95**(1), 108–137 (2007)
51. Ostrovsky, R., Rabani, Y., Schulman, L.: Error-correcting codes for automatic control. *IEEE Trans. Inf. Theory* **55**(7), 2931–2941 (2009)
52. Sahai, A.: Anytime information theory. Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, MA (2001)
53. Sahai, A., Mitter, S.K.: 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 (2006)
54. Sahai, A., Mitter, S.K.: The necessity and sufficiency of anytime capacity for stabilization of a linear system over a noisy communication link—Part II: vector systems (2006). Available on-line at [arXiv:cs/0610146v2](https://arxiv.org/abs/cs/0610146v2) [cs.IT]
55. Schalkwijk, J.P.M., Kailath, T.: A coding scheme for additive noise channels with feedback—I: no bandwidth constraint. *IEEE Trans. Inf. Theory* **12**, 172–182 (1966)
56. Schenato, L., Sinopoli, B., Franceschetti, M., Poolla, K., Sastry, S.: Foundations of control and estimation over lossy networks. *Proc. IEEE* **95**(1), 163–187 (2007)
57. Schulman, L.: Coding for interactive communication. *IEEE Trans. Inf. Theory* **42**(6), 1745–1756 (1996)
58. Shannon, C.E.: The zero-error capacity of a noisy channel. *IRE Trans. Inf. Theory* **2**, 8–19 (1956)
59. Shi, L., Epstein, M., Murray, R.: Kalman filtering over a packet-dropping network: a probabilistic perspective. *IEEE Trans. Autom. Control* **55**(3), 594–604 (2010)

60. Simşek, T., Jain, R., Varaiya, P.: Scalar estimation and control with noisy binary observations. *IEEE Trans. Autom. Control* **49**(9), 1598–1603 (2004)
61. Sinopoli, B., Schenato, L., Franceschetti, M., Poolla, K., Jordan, M., Sastry, S.: Kalman filtering with intermittent observations. *IEEE Trans. Autom. Control* **49**(9), 1453–1464 (2004)
62. Soumya, K., Sinopoli, B., Moura, J.M.F.: Kalman filtering with intermittent observations: weak convergence to a stationary distribution. *IEEE Trans. Autom. Control* **57**(2), 405–420 (2012)
63. Sukhavasi, R., Hassibi, B.: Error correcting codes for distributed control (2011). Available on-line at [arXiv:1112.4236v2](https://arxiv.org/abs/1112.4236v2) [cs.IT]
64. Tatikonda, S., Mitter, S.K.: Control over noisy channels. *IEEE Trans. Autom. Control* **49**(7), 1196–1201 (2004)
65. Tatikonda, S., Mitter, S.K.: Control under communication constraints. *IEEE Trans. Autom. Control* **49**(7), 1056–1068 (2004)
66. Tatikonda, S., Sahai, A., Mitter, S.K.: Stochastic linear control over a communication channel. *IEEE Trans. Autom. Control* **49**(9), 1549–1561 (2004)
67. Witsenhausen, H.S.: A counterexample in stochastic optimum control. *SIAM J. Control* **6**(1), 131–147 (1968)
68. Wong, W.S., Brockett, R.: Systems with finite communication bandwidth constraints. I. State estimation problems. *IEEE Trans. Autom. Control* **42**(9), 1294–1299 (1997)
69. Wong, W.S., Brockett, R.: Systems with finite communication bandwidth constraints. II. Stabilization with limited information feedback. *IEEE Trans. Autom. Control* **44**(5), 1049–1053 (1999)
70. You, K., Xie, L.: Minimum data rate for mean square stabilization of discrete LTI systems over lossy channels. *IEEE Trans. Autom. Control* **55**(10), 2373–2378 (2010)
71. You, K., Xie, L.: Minimum data rate for mean square stabilizability of linear systems with Markovian packet losses. *IEEE Trans. Autom. Control* **56**(4), 772–785 (2011)
72. Yüksel, S.: Stochastic stabilization of noisy linear systems with fixed-rate limited feedback. *IEEE Trans. Autom. Control* **55**(12), 2847–2853 (2010)
73. Yüksel, S., Başar, T.: Control over noisy forward and reverse channels. *IEEE Trans. Autom. Control* **56**(5), 1014–1029 (2011)