Tracking and Control of Gauss–Markov Processes over Packet-Drop Channels with Acknowledgments

Anatoly Khina, *Member, IEEE*, Victoria Kostina, *Member, IEEE*, Ashish Khisti, *Member, IEEE*, and Babak Hassibi, *Member, IEEE*

***Abstract*—We consider the problem of tracking the state of Gauss–Markov processes over rate-limited erasure-prone links. We concentrate first on the scenario in which several independent processes are seen by a single observer. The observer maps the processes into finite-rate packets that are sent over the erasure-prone links to a state estimator, and are acknowledged upon packet arrivals. The aim of the state estimator is to track the processes with zero delay and with minimum mean square error (MMSE). We show that, in the limit of many processes, greedy quantization with respect to the squared error distortion is optimal. That is, there is no tension between optimizing the MMSE of the process in the current time instant and that of future times. For the case of packet erasures with delayed acknowledgments, we connect the problem to that of compression with side information that is known at the observer and may be known at the state estimator—where the most recent packets serve as side information that may have been erased, and demonstrate that the loss due to a delay by one time unit is rather small. For the scenario where only one process is tracked by the observer–state estimator system, we further show that variable-length coding techniques are within a small gap of the many-process outer bound. We demonstrate the usefulness of the proposed approach for the simple setting of discrete-time scalar linear quadratic Gaussian control with a limited data-rate feedback that is susceptible to packet erasures.**

***Index Terms*—Networked control systems, packet loss, sequential coding of correlated sources, source coding with side information, state estimation, successive refinement.**

I. Introduction

TRACKING the state of a system from noisy and possibly partially observable measurements is of prime importance in many estimation scenarios, and serves as an important building block in many control setups.

The recent rapid growth in wireless connectivity and its ad hoc distributed nature, while offering a plethora of new and exciting possibilities, introduces new design challenges for control over such media. These challenges include, among others, the need to track processes with minimal error over digital links of limited data rate, which could be prone to (packet) erasures, and joint processing and reconstruction of distributed processes.

An important scenario, often encountered in practice, depicted in Fig. 1, is that of a *multi-track* system that tracks several processes over a single shared communication link. In this scenario, at each time instant, several processes are observed by a single observer. The observer, in turn, collects the measured states of these processes into a single vector state or *frame*, and maps them into finite-rate packets. These packets, in turn, are sent to the state-estimator over a channel which is prone to packet erasures. The state estimator tracks the latest states of the different processes, by constructing minimum mean square error (MMSE) estimates thereof using the available packets received thus far.

Since these settings incorporate communication components, we appeal to relevant tools and results from information theory. The information-theoretic framework for the multi-track setting with a large number of independent processes (large frames) and without packet erasures, was provided by Viswanathan and Berger [1] via the notion of *sequential coding* for the case of two time steps and for more steps in [2]–[5]. In these works, the optimal tradeoff between given (per-process) rates and MMSEs (referred to as *distortions*) were determined when the number of processes is large, in the form of an optimization problem.

A similar framework in the context of control was studied by Tatikonda [6]–[8] and Borkar *et al.* [7], who noticed the

Manuscript received October 27, 2017; revised May 17, 2018; accepted June 1, 2018. Date of publication June 25, 2018; date of current version May 28, 2019. This work has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie Grant 708932. The work of V. Kostina was supported in part by the National Science Foundation under Grant CCF-1566567 and Grant CCF-1751356. The work of A. Khisti was supported by the Canada Research Chairs Program. The work of B. Hassibi was supported in part by the National Science Foundation under Grant CNS-0932428, Grant CCF-1018927, Grant CCF-1423663, and Grant CCF-1409204; in part by a Grant from Qualcomm, Inc.; in part by NASA's Jet Propulsion Laboratory through the President and Director's Fund; and in part by King Abdullah University of Science and Technology. Recommended by Associate Editor M. Rabbat. This work was done in part while A. Khina and V. Kostina were visiting the Simons Institute for the Theory of Computing. This paper was presented in part at the IEEE Information Theory Workshop, Kaohsiung, Taiwan, November 2017. *(Corresponding author: Anatoly Khina.)*

A. Khina was with the Department of Electrical Engineering, California Institute of Technology, Pasadena, CA 91125 USA. He is now with the Department of Electrical Engineering-Systems, Tel Aviv University, Tel Aviv 69978, Israel (e-mail: anatolyk@eng.tau.ac.il)

V. Kostina and B. Hassibi are with the Department of Electrical Engineering, California Institute of Technology, Pasadena, CA 91125 USA (e-mail: vkostina@caltech.edu; hassibi@caltech.edu)

A. Khisti is with the Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON M5S 3G4, Canada (e-mail: akhisti@comm.utoronto.ca)

Digital Object Identifier 10.1109/TCNS.2018.2850225

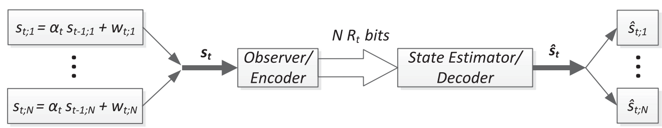

Fig. 1. Multi-track of Gauss–Markov processes over a finite-rate channel.

intimate connection to the early works of Gorbunov and Pinsker [9], [10]. Subsequent noteworthy efforts in the context of tracking include [11], [12] and references therein.

For the special case of Gauss–Markov processes, an explicit expression for the achievable sum-rate for given distortions was derived in [2] and [3] via the paradigms of predictive coding and differential pulse-code modulation (DPCM) [13]–[17] (see also [18, Ch. 6] and the references therein), and extended for the case of three time-steps of jointly Gaussian (not necessarily Markov) processes in [19].

In practice, packet-based protocols are prone to erasures and possible delays. The multi-track scenario in the presence of packet erasures was treated under various erasure models. The case where only the first packet is prone to an erasure was considered in [20]. A more general approach that trades between the performance given all previously sent packets and the performance given only the last packet was proposed in [21]. For random independent identically distributed (i.i.d.) packet erasures, a hybrid between pulse-code modulation (PCM) and DPCM, termed leaky DPCM was proposed in [22] and analyzed for the case of very low erasure probability in [23]. The scenario in which the erasures occur in bursts was considered in [24], [25].

These works correspond to UDP-based networks [26], in which no acknowledgment (ACK) upon packet arrival is available. That is, the observer does not know whether transmitted packets successfully arrived to the state estimator or not.

In contrast, in TCP-based networks, packet arrivals are acknowledged via a communication feedback link, in order to robustify the transmission of the overlying data [26]. Stabilizing control systems under this scenario has been studied in various works, [27]–[29], to name a few.

In this paper, we first consider the multi-track scenario of Gauss–Markov processes, which is defined formally in Section II. We determine the optimal tradeoff between rates and distortions when the number of processes (frame length) is large, in Section III. Specifically, we show, in Section III, that greedy quantization that optimizes the distortion at each time is also optimal for minimizing the distortion of future time instants. This insight allows us to extend the result to the case where the rate r_t available for the transmission of the packet at time t is determined just prior to its transmission, in Section IV.

The packet-erasure channel with instantaneous ACKs can be viewed as a special case of the above noiseless channel with random rate allocation, with $r_t = 0$ corresponding to a packet-erasure event [30]. The optimal tradeoff between rates and distortions for the multi-track scenario of Gauss–Markov processes in the presence of packet erasures and instantaneous ACKs thereby follows as a consequence, as is shown in Section V for both single- and multi-packet per state frame scenarios.

We further tackle, in Section VI, the more challenging delayed ACK setting, in which the observer does not know whether the most recently transmitted packets have arrived or not. By viewing these recent packets as side information (SI) that is available at the observer, and possibly at the state estimator, and leveraging the results of Kaspi [31] along with their specialization for the Gaussian case by Perron *et al.* [32],[1] we adapt our transmission scheme of Section III to the case of delayed ACKs. We provide a detailed description of the proposed scheme for the case where ACKs are delayed by one time unit and demonstrate that the loss compared to the case of instantaneous ACKs is small.

In Section VII, we go on and consider the case of tracking a single process—*single-track*, and a variable-length coding (VLC) scenario [35], [36, Ch. 5], in which the packet size is not fixed and is instead constrained to be below a desired rate *on average*. We consider a scheme that sequentially applies entropy-coded dithered quantization (ECDQ) [37]–[39], [40, Ch. 5], redolent of the scheme in [41], and show that it attains an MMSE–rate tradeoff that is close to the large-frame outer bound of Section III.

By supplementing the state-tracking task with appropriate control actions in Section VIII, we demonstrate the applicability of the derived results in Sections III and V to the scenario of linear quadratic Gaussian (LQG) networked control, where a scalar linear plant driven by an i.i.d. Gaussian process is stabilized by a controller that is not colocated with the observer and is separated from it, instead, by a packet-erasure (and more generally, a random-rate budget) channel. We derive inner and outer bounds on the optimal LQG cost that extend those in [42] and [43] to packet-erasure channels. We conclude the paper with Section IX, by discussing the cases of large delays, other types of VLC compression, and single-track with fixed-length coding (FLC) compression.

A. Notation

Throughout this paper, $\|\cdot\|$ denotes the Euclidean norm. $\mathbb{N}$ is the set of natural numbers. Random variables are denoted by lower-case possibly accented letters with temporal subscripts ($a_t, \hat{\hat{a}}_t$), and random vectors (frames) of length $N \in \mathbb{N}$ by boldface lower-case letters ($\boldsymbol{a}, \hat{\hat{\boldsymbol{a}}}_t$). We denote temporal sequences by $\boldsymbol{a}^t \triangleq (\boldsymbol{a}_1, \ldots, \boldsymbol{a}_t)$, where $\boldsymbol{a}_t \triangleq \mathrm{Transpose}\{(\, a_{t;1}\ a_{t;2}\ \cdots\ a_{t;N}\,)\}$, and $[T] \triangleq \{1, \ldots, T\}$ is the interval from 1 to $T \in \mathbb{N}$. All other notations represent deterministic scalars.

II. Problem Statement

The transmission spans the time interval $[T]$ of horizon T.

We next describe the state dynamics, and the operations carried by the observer and the state estimator, which communicate over a finite-rate channel (see Fig. 1).

State dynamics. Consider $N \in \mathbb{N}$ independent Gauss–Markov processes $\{s_{t;1}\}$, $\{s_{t;2}\}$, ..., $\{s_{t;N}\}$ with identical statistics. This can be compactly represented in a vector form as

[1]The scenario considered in [31] and [32] can also be viewed as special case of the results of Heegard and Berger [33], where the SI is not available at the observer, by adjusting the distortion measure and "augmenting" the state [34]. Interestingly, knowing the SI at the observer allows one to improve the optimal performance of this scenario in the Gaussian case; see Remark 11.

(we assume $\boldsymbol{s}_0 = 0$ for convenience)[2]

$$\boldsymbol{s}_t = \alpha_t \boldsymbol{s}_{t-1} + \boldsymbol{w}_t, \qquad t \in [T] \tag{1}$$

where $\boldsymbol{s}_t$ is the *vector state* or *frame* at time t, $\{\alpha_t\}$ are known process coefficients, the entries of $\boldsymbol{w}_t$ comprise N jointly independent driving noises, the temporal entries of which are i.i.d. Gaussian of zero mean and variance W_t.

Denote the average power of each state at time t by $S_t \triangleq \mathbb{E}\left[s_{t;n}^2\right]$, $n \in [N]$. Then, (1) implies the following recursive relation (with $S_0 = 0$):

$$S_t = \alpha_t^2 S_{t-1} + W_t, \qquad t \in [T]. \tag{2}$$

Observer: Sees the states $\{s_{t;1}, \ldots, s_{t;N}\}$ of all the N process at time t, collects them into the frame $\boldsymbol{s}_t$ and applies a causal function $\mathcal{E}_t$ to the observed frame sequence $\boldsymbol{s}^t$, to generate the packet $f_t \in \left[2^{NR_t}\right]$:

$$f_t = \mathcal{E}_t\left(\boldsymbol{s}^t\right), \tag{3}$$

where R_t is the per-process *rate* available for transmission over the channel at time t.

Channel: At time t, a packet $f_t \in \left[2^{NR_t}\right]$ is sent over a noiseless channel of (per-process) finite rate R_t.

State estimator: Applies a causal function $\mathcal{D}_t$ to the sequence of received packets f^t, to construct an estimate $\hat{\boldsymbol{s}}_t$ of $\boldsymbol{s}_t$, at time t:

$$\hat{\boldsymbol{s}}_t = \mathcal{D}_t\left(f^t\right). \tag{4}$$

Distortion: The average quadratic distortion (or MMSE) at time t is defined as

$$D_t \triangleq \frac{1}{N}\mathbb{E}\left[\|\boldsymbol{s}_t - \hat{\boldsymbol{s}}_t\|^2\right]. \tag{5}$$

In the important special case of fixed parameters

$$\begin{aligned} \alpha_t &\equiv \alpha, \\ W_t &\equiv W, \end{aligned} \qquad t \in [T], \tag{6}$$

the average process power, assuming $|\alpha| < 1$, converges to

$$S_\infty = \frac{W}{1-\alpha^2}.$$

In that case, by taking the rate-budget to be fixed too

$$R_t \equiv R, \qquad t \in [T], \tag{7}$$

define the steady-state distortion (assuming the limit exists)

$$D_\infty \triangleq \lim_{T\to\infty} D_t. \tag{8}$$

Definition 1 (Distortion–rate region): The *distortion–rate region* is the closure of all achievable distortion tuples $D^T \triangleq (D_1, \ldots, D_T)$ for a rate tuple $R^T \triangleq (R_1, \ldots, R_T)$, for any N, however large; its inverse is the *rate–distortion region*.

Definition 2 (Average-stage rate and distortion): The average-stage rate and distortion are defined as

$$\bar{R}_T \triangleq \frac{1}{T}\sum_{t=1}^{T} R_t, \qquad \bar{D}_T \triangleq \frac{1}{T}\sum_{t=1}^{T} D_t, \tag{9}$$

respectively. We further denote the steady-state average-stage rate and distortion by

$$\bar{R}_\infty = \limsup_{T\to\infty} \bar{R}_T, \qquad \bar{D}_\infty = \limsup_{T\to\infty} \bar{D}_T. \tag{10}$$

III. Distortion–Rate Region of Gauss–Markov Process Multi-Tracking

The optimal achievable distortions for given rates, under the model of Section II, are provided in the following theorem.

Theorem 1 (Distortion–rate region): The distortion–rate region of Gauss–Markov process multi-track for a rate tuple R^T is given by all distortion tuples D^T that satisfy $D_t \geq D_t^*$ with

$$D_t^* = \left(\alpha_t^2 D_{t-1}^* + W_t\right) 2^{-2R_t}, \qquad t \in [T] \tag{11a}$$

$$D_0^* = 0. \tag{11b}$$

Remark 1: The impossibility of Theorem 1 has been established in [8, Lemma 4.3]. We provide an alternative simple proof in Section III-B that allows us to treat random rates in the sequel.

Remark 2: The setting of Theorem 1 is referred to as "causal encoder–causal decoder" in [2]. We note that [2] provides an explicit result only for the sum-rate of the Gauss–Markov model [3]. Torbatian and Yang [19] extend the sum-rate result to the case of three-step general jointly Gaussian processes (which do not necessarily constitute a Markov chain). Our work, on the other hand, fully characterizes the rate–distortion region for the case of Gauss–Markov processes.

Remark 3: The results and proof (provided in the sequel) of Theorem 1 imply that optimal greedy quantization at every step—which is achieved via Gaussian backward [36, Ch. 10.3] or forward [36, pp. 338–339] channels—becomes optimal when N is large. Moreover, it achieves the optimum for all $t \in [T]$ simultaneously, meaning that there is no tension between minimizing the current distortion and future distortions.

To prove Theorem 1, we first construct the optimal greedy scheme and determine its performance in Section III-A. We then show that it is in fact optimal when N goes to infinity in Section III-B.

A. Achievability

We construct an inner bound using the optimal greedy scheme, which amounts to the classical causal DPCM scheme. In this scheme, all the quantizers are assumed to be MMSE quantizers, whose quantized values are well known to be uncorrelated with the resulting quantization errors.

Scheme 1 (DPCM).

Observer: At time t:

1) Generates the prediction error

$$\tilde{\boldsymbol{s}}_t \triangleq \boldsymbol{s}_t - \alpha_t \hat{\boldsymbol{s}}_{t-1} \tag{12}$$

where $\hat{\boldsymbol{s}}_{t-1}$, defined in (4), is the previous frame reconstruction at the state estimator, and $\hat{\boldsymbol{s}}_0 = 0$; a linear recursive relation for $\hat{\boldsymbol{s}}_t$ is provided in the sequel in (13).[3]

2) Generates $\hat{\tilde{\boldsymbol{s}}}_t$, the quantized reconstruction of the prediction error $\tilde{\boldsymbol{s}}_t$, by quantizing $\tilde{\boldsymbol{s}}_t$ using the MMSE quantizer of rate R_t and frame length N.

[2] The proposed treatment can be generalized to a matrix α_t, but is much more involved and, therefore, remains outside the scope of this paper.

[3] $\hat{\boldsymbol{s}}_{t-1} = \mathbb{E}\left[\boldsymbol{s}_{t-1}|f^{t-1}\right]$ and $\alpha_t \hat{\boldsymbol{s}}_{t-1} = \mathbb{E}\left[\boldsymbol{s}_t|f^{t-1}\right]$ are the MMSE estimators of $\boldsymbol{s}_{t-1}$ and $\boldsymbol{s}_t$, respectively, given all outputs until time $t-1$.

3) Sends $f_t \in [2^{NR_t}]$, the corresponding packet to $\hat{\tilde{\boldsymbol{s}}}_t$ over the channel.

State estimator: At time t:

1) Receives f_t.
2) Recovers the reconstruction $\hat{\tilde{\boldsymbol{s}}}_t$ of the prediction error $\tilde{\boldsymbol{s}}_t$.
3) Generates an estimate $\hat{\boldsymbol{s}}_t$ of $\boldsymbol{s}_t$:

$$\hat{\boldsymbol{s}}_t = \alpha_t \hat{\boldsymbol{s}}_{t-1} + \hat{\tilde{\boldsymbol{s}}}_t \,. \tag{13}$$

Performance analysis: First note that the error between $\boldsymbol{s}_t$ and $\hat{\boldsymbol{s}}_t$, $\boldsymbol{e}_t \triangleq \boldsymbol{s}_t - \hat{\boldsymbol{s}}_t$, is equal to $\boldsymbol{e}_t = \tilde{\boldsymbol{s}}_t - \hat{\tilde{\boldsymbol{s}}}_t$ by (12), (13). Thus, the distortion (5) is also the distortion in reconstructing $\tilde{\boldsymbol{s}}_t$.

This, along with (1) and (12) means that $\tilde{\boldsymbol{s}}_t = \alpha_t \boldsymbol{e}_{t-1} + \boldsymbol{w}_t$. Since $\boldsymbol{w}_t$ is independent of $\boldsymbol{e}_{t-1}$, the average power of the entries of $\tilde{\boldsymbol{s}}_t$ is equal to

$$\tilde{S}_t = \alpha_t^2 D_{t-1} + W_t \,.$$

Using the property that the rate–distortion function under mean square error distortion of a process with a given average variance is upper bounded by that of an i.i.d. Gaussian process with the same variance (see, e.g., [36, pp. 338–339]), we obtain $D_t \leq \left(\alpha_t^2 D_{t-1} + W_t\right) 2^{-2R_t}$, and hence, (11) is achievable within an arbitrarily small $\epsilon > 0$, for a sufficiently large N. ∎

B. Impossibility (Converse)

We now prove that, for any frame length $N \in \mathbb{N}$

$$D_t \geq 2^{-2R_t} \mathbb{E}_{\check{f}^{t-1}} \left[\mathcal{N}\left(\boldsymbol{s}_t | f^{t-1} = \check{f}^{t-1}\right) \right] \tag{14a}$$

$$\geq D_t^*, \qquad t \in [T] \tag{14b}$$

by induction, where the sequence $\{D_t^*\}$ is defined in (11)

$$\mathcal{N}(\boldsymbol{s}_t) \triangleq \frac{1}{2\pi e} 2^{\frac{2}{N} h \boldsymbol{s}_t}, \quad \mathcal{N}\left(\boldsymbol{s}_t | f^k = \check{f}^k\right) \triangleq \frac{1}{2\pi e} 2^{\frac{2}{N} h \left(\boldsymbol{s}_t | f^k = \check{f}^k\right)}$$

denote the entropy power (EP) and conditional EP of $\boldsymbol{s}_t$ given $f^k = \check{f}^k$, the expectation $\mathbb{E}_{\check{f}^{t-1}}[\cdot]$ is w.r.t. $\check{f}^{t-1}$, and the random vector $\check{f}^t$ is distributed the same as f^t.

Basic step ($t = 1$): Since $\boldsymbol{s}_0 = 0$, and the vector $\boldsymbol{w}_1$ consists of i.i.d. Gaussian entries of variance W_1, (14b) is satisfied with equality. To prove (14a), we use the fact that the optimal achievable distortion D_1 for a Gaussian process ($\boldsymbol{s}_1 = \boldsymbol{w}_1$) with i.i.d. entries of power W_1 and rate R_1 is dictated by its rate–distortion function [36, Ch. 10.3.2]: $D_1 \geq W_1 2^{-2R_1}$.

Inductive step: Let $k \geq 2$ and suppose (14) is true for $t = k - 1$. We shall now prove that it holds also for $t = k$.

$$D_k = \frac{1}{N} \mathbb{E}\left[\mathbb{E}\left[\left\|\boldsymbol{s}_k - \hat{\boldsymbol{s}}_k\right\|^2 | f^{k-1}\right]\right] \tag{15a}$$

$$= \frac{1}{N} \mathbb{E}_{\check{f}^{k-1}}\left[\mathbb{E}\left[\left\|\boldsymbol{s}_k - \hat{\boldsymbol{s}}_k\right\|^2 | f^{k-1} = \check{f}^{k-1}\right]\right] \tag{15b}$$

$$\geq \mathbb{E}_{\check{f}^{k-1}}\left[\mathcal{N}\left(\boldsymbol{s}_k | f^{k-1} = \check{f}^{k-1}\right) 2^{-2R_k}\right] \tag{15c}$$

$$= \mathbb{E}_{\check{f}^{k-1}}\left[\mathcal{N}\left(\alpha_k \boldsymbol{s}_{k-1} + \boldsymbol{w}_k | f^{k-1} = \check{f}^{k-1}\right)\right] 2^{-2R_k} \tag{15d}$$

$$\geq \left\{\mathbb{E}_{\check{f}^{k-2}}\left[\mathbb{E}_{\check{f}_{k-1}}\left[\mathcal{N}\left(\alpha_k \boldsymbol{s}_{k-1} | f^{k-1} = \check{f}^{k-1}\right) | \check{f}^{k-2}\right]\right] + \mathcal{N}(\boldsymbol{w}_k)\right\} 2^{-2R_k} \tag{15e}$$

$$\geq \left\{\alpha_k^2 \mathbb{E}_{\check{f}^{k-2}}\left[\mathcal{N}\left(\boldsymbol{s}_{k-1} | f^{k-2} = \check{f}^{k-2}, f_{k-1}\right)\right] + W_k\right\} 2^{-2R_k} \tag{15f}$$

$$\geq \left\{\alpha_k^2 \mathbb{E}_{\check{f}^{k-2}}\left[\mathcal{N}\left(\boldsymbol{s}_{k-1} | f^{k-2} = \check{f}^{k-2}\right)\right] 2^{-2R_{k-1}} + W_k\right\} 2^{-2R_k} \tag{15g}$$

$$\geq 2^{-2R_k}\left(\alpha_k^2 D_{k-1}^* + W_k\right) \tag{15h}$$

$$= D_k^* \tag{15i}$$

where (15a) follows from (5) and the law of total expectation; (15b) holds since f^{k-1} and $\check{f}^{k-1}$ have the same distribution; (15c) follows by bounding from below the inner expectation (conditional distortion) by the rate–distortion function and the Shannon lower bound [36, Ch. 10]—this also proves (14a); (15d) is due to (1); (15e) follows from the EP inequality [36, Ch. 17]; (15f) holds since $\boldsymbol{w}_k$ is Gaussian, the scaling property of differential entropies and Jensen's inequality:

$$\mathbb{E}_{\check{f}_{k-1}}\left[2^{\frac{2}{N} h\left(\boldsymbol{s}_{k-1} | f^{k-1} = \check{f}^{k-1}\right)} | \check{f}^{k-2}\right] \geq 2^{\frac{2}{N} h\left(\boldsymbol{s}_{k-1} | f^{k-2} = \check{f}^{k-2},\, f_{k-1}\right)};$$

(15g) follows from the following standard set of inequalities:

$$\begin{aligned} NR_{k-1} &\geq H\left(f_{k-1} | f^{k-2} = \check{f}^{k-2}\right) \\ &\geq I\left(\boldsymbol{s}_{k-1}; f_{k-1} | f^{k-2} = \check{f}^{k-2}\right) \\ &= h\left(\boldsymbol{s}_{k-1} | f^{k-2} = \check{f}^{k-2}\right) - h\left(\boldsymbol{s}_{k-1} | f^{k-2} = \check{f}^{k-2}, f_{k-1}\right); \end{aligned}$$

(15h) is by the induction hypothesis; and (15i) holds by the definition of $\{D_t^*\}$ (11)—which also proves (14b). ∎

Assertion 1 (Outer bound for non-Gaussian noise): Consider the setting of Section II with independent non-Gaussian noise entries $\{w_{t;n} | t \in [T], n \in [N]\}$. Then, the average achievable distortion D_t at time $t \in [T]$ is bounded from below by $D_t \geq D_t^*$, with $D_0^* = 0$ and D_t^* given by the recursion

$$D_t^* = \left(\alpha^2 D_{t-1}^* + \mathcal{N}(\boldsymbol{w}_t)\right) 2^{-2R_t}.$$

Proof: The proof is identical to that of the lower bound for the Gaussian case with W_t replaced by $\mathcal{N}(\boldsymbol{w}_t)$.[4] ∎

C. Fixed-Parameter Gauss–Markov Processes

For the case of fixed parameters (6) and fixed rate (7), the steady-state average distortion is given as follows.

Corollary 1 (Steady-state distortion with fixed rate): Assume a fixed-parameter (6) fixed-rate budget (7) setting. If $\alpha^2 2^{-2R} < 1$,[5] then the steady-state distortion is given by

$$D_\infty^* \triangleq \lim_{t \to \infty} D_t^* = \frac{W 2^{-2R}}{1 - \alpha^2 2^{-2R}}$$

and is otherwise unbounded.

Proof: The proof is immediate by noting that (11) constitutes a linear time-invariant (LTI) system and, therefore, is globally exponentially stable if the (only) pole of its transfer function lies strictly inside the unit circle, i.e., $\alpha^2 2^{-2R} < 1$, and is unstable otherwise [44, Ch. 6].

[4] Recall that in the Gaussian setting $\mathcal{N}\boldsymbol{w}_t \equiv W_t$.
[5] This is trivial for $|\alpha| < 1$.

Remark 4: As is evident from the proof, the result of Corollary 1 remains true for any initial value D_0^*.

Remark 5: The impossibility part of Corollary 1 can be traced back to the work of Gorbunov and Pinsker [10].

Interestingly, the optimal steady-state distortion achievable with a fixed-rate budget (7) is in fact optimal even if we loosen this restriction to a total rate-budget constraint as was previously observed, e.g., in [41]. This is a simple corollary of Theorem 1 and is formally proved next. The same conclusion holds if the frame entries are correlated Gaussians, as was recently proved by Tanaka [45].

Corollary 2 (Steady-state distortion with total rate): The average-stage steady-state distortion (10) $\bar{D}_\infty$, under a total rate-budget constraint (10) $\bar{R}_\infty \leq R$, is bounded from below by $\bar{D}_\infty \geq D_\infty^*$. Consequently, the fixed (a.k.a. uniform) rate allocation $R_t \equiv R$ is optimal in the limit of $T \to \infty$.

Proof: Without loss of generality, for a given tuple R^T, it suffices to consider distortion tuples D^T that belong to the boundary of the rate–distortion region, namely, distortion tuples satisfying (11) with equality

$$R_t = \frac{1}{2}\log\left(\alpha^2 D_{t-1} + W\right) - \frac{1}{2}\log D_t\,. \tag{16}$$

For the equivalent problem of minimizing the total rate budget (9) under an average-stage distortion constraint $\bar{D}_T \leq D$, the total rate budget can be bounded from below as

$$\bar{R}_T \equiv \frac{1}{T}\sum_{t=1}^{T} R_t \tag{17a}$$

$$= \frac{1}{T}\sum_{t=1}^{T}\left[\frac{1}{2}\log\left(\alpha^2 D_{t-1} + W\right) - \frac{1}{2}\log D_t\right] \tag{17b}$$

$$= \sum_{t=1}^{T}\frac{1}{2T}\log\left(\alpha^2 + \frac{W}{D_t}\right) - \frac{1}{2T}\log\left(1 + \frac{\alpha^2 D_T}{W}\right) \tag{17c}$$

$$\geq \frac{1}{2}\log\left(\alpha^2 + \frac{W}{\bar{D}_T}\right) - \frac{1}{2T}\log\left(1 + \frac{\alpha^2 T\bar{D}_T}{W}\right) \tag{17d}$$

$$\geq \frac{1}{2}\log\left(\alpha^2 + \frac{W}{D}\right) - \frac{1}{2T}\log\left(1 + \frac{\alpha^2 TD}{W}\right) \tag{17e}$$

where we use (9) in (17a), (17b) holds by substituting (16), (17d) follows from Jensen's inequality and $D_1 \leq T\bar{D}_T$, and (17e) holds due to the constraint $\bar{D}_T \leq D$.

Evaluating (17) in the limit $T \to \infty$ concludes the proof. ■

IV. Random-Rate Budgets

In practice, the available transmission rate may vary across time depending on the quality of service offered by the infrastructure, as well as, due to other applications sharing the same infrastructure. We, therefore, generalize next the results of Section III to random rates $\{r_t\}$ that are independent of each other and of $\{\boldsymbol{w}_t\}$. The rate r_t is revealed to the observer just before the transmission at time t.

Theorem 2 (Distortion–rate region): The distortion–rate region of Gauss–Markov multi-track with independent rates r^T is given by all distortion tuples D^T that satisfy $D_t \geq D_t^*$ with

$$D_t^* = \left(\alpha_t^2 D_{t-1}^* + W_t\right)\mathbb{E}\left[2^{-2r_t}\right], \qquad t \in [T], \tag{18a}$$

$$D_0^* = 0. \tag{18b}$$

Proof: *Achievability.* Since the achievability scheme in Theorem 1 does not use the knowledge of future transmission rates to encode or decode the packet at time t, we have

$$d_t \triangleq \frac{1}{N}\mathbb{E}\left[\|\boldsymbol{s}_t - \hat{\boldsymbol{s}}_t\|^2 \,|r^T\right] \tag{19a}$$

$$= \frac{1}{N}\mathbb{E}\left[\|\boldsymbol{s}_t - \hat{\boldsymbol{s}}_t\|^2 \,|r^t\right] \tag{19b}$$

$$\leq \left(\alpha_t^2 d_{t-1} + W_t\right)2^{-2r_t} + \epsilon \tag{19c}$$

for any $\epsilon > 0$, however small, and large enough N.

By taking an expectation of (19c) with respect to r^t and using the independence of r^{t-1} and r_t, we obtain (18).

Impossibility: Revealing the rates to the observer and the state estimator prior to the start of transmission can only improve the distortion. Thus, the distortions $\{d_t\}$ conditioned on $\{r_t\}$ (19a) are bounded from below as in Theorem 1; by taking the expectation w.r.t. $\{r_t\}$, we attain the desired result. ■

Remark 6: By applying Jensen's inequality to (18a): $\mathbb{E}\left[2^{-2r_t}\right] \geq 2^{-2\mathbb{E}[r_t]}$, we see that using packets of a fixed rate of $\mathbb{E}[r_t]$ performs better than using random rates r_t.

For the special case of fixed-parameters (6) and i.i.d. rates $\{r_t\}$, the steady-state distortion is given as follows.

Corollary 3 (Steady state): Assume a fixed-parameter setting (6) with i.i.d. rates $\{r_t\}$. If $\alpha^2 B < 1$,[6] where $B \triangleq \mathbb{E}\left[2^{-2r_1}\right]$, then the steady-state distortion is given by

$$D_\infty^* \triangleq \lim_{t\to\infty} D_t^* = \frac{BW}{1-\alpha^2 B} \tag{20}$$

and is otherwise unbounded.

Proof: The proof is identical to that of Corollary 1 with 2^{-R} replaced by B. ■

V. Packet Erasures With Instantaneous ACKs

A. One Packet Per Frame

An important scenario encompassed by the model of Section IV is that of packet erasures [30]. Since a packet erasure at time t can be viewed as $r_t = 0$, and assuming that the observer sends packets of fixed rate R and is cognizant of any packet erasures instantaneously, the packet erasure channel can be cast as the random rate channel of Section IV with

$$r_t = b_t R \tag{21a}$$

$$= \begin{cases} R, & b_t = 1 \\ 0, & b_t = 0 \end{cases} \tag{21b}$$

where $\{b_t\}$ are the packet-erasure events, such that $b_t = 1$ corresponds to a successful arrival of the packet f_t at time t, and $b_t = 0$ means it was erased. We further denote by

$$g_t \triangleq b_t f_t \tag{22}$$

the received output where $g_t = 0$ corresponds to an erasure, and otherwise $g_t = f_t$. We assume that $\{b_t\}$ are i.i.d. according to a Bernoulli distribution $\mathscr{B}er(\beta)$ with $\beta \in [0, 1]$.

Remark 7: We shall concentrate on the case of packets of fixed rate R to simplify the subsequent discussion. This way, the only randomness in rate comes from the packet-erasure effect. Nevertheless, all the results that follow can be easily extended

[6] Again, this is trivial for $|\alpha| < 1$.

to random/varying rate allocations to which the effect of packet erasures $\{b_t\}$ is added in the same manner as in (21).

Corollary 4 (Distortion–rate region): The distortion–rate region of Gauss–Markov multi-track with i.i.d. $\mathcal{B}\rceil\nabla(\beta)$ packet erasures and instantaneous ACKs is given as in Theorem 2 with

$$B \triangleq \mathbb{E}\left[2^{-2r_1}\right] = 1 - \beta\left(1 - 2^{-2R}\right). \tag{23}$$

Corollary 5 (Steady state): The steady-state distortion is given as in Corollary 3 with B as in (23).

Remark 8: In contrast to the scenario without packet erasures, the uniform rate allocation can be improved by allowing a *dynamic* rate allocation that *depends on the pattern of packet erasures* b^{t-1}. This setup can be thought of as the source-coding dual of the fast fading channel coding problem where the fading coefficient is known at both the transmitter and the receiver prior to transmission, and the transmitter optimizes the transmission rate via waterfilling across time [46, Ch. 5.4].

B. Multiple Packets Per Frame

In Section V-A, we assumed that one packet (f_t) was sent per each frame ($\boldsymbol{s}_t$). Instead, one may choose to transmit multiple packets of lower rate per one frame. If we assume that each packet arrival is instantly acknowledged, then the resulting scenario falls again in the random-rate budget framework of Section IV. Interestingly, it turns out that the optimal number of packets per frame depends on the PDF of the rate, i.e., increasing the number of packets can either improve or deteriorate the performance.

Specifically, assume that the observer uses K packets of equal rate R/K (and hence, a total rate of R) to successively refine [47, Ch. 13.5] a single state frame $\boldsymbol{s}_t$. Then, the rate probability distribution amounts to

$$r_t = \frac{b_t}{K}R$$

with b_t denoting the number of successful packet arrivals at time t, corresponding to state frame $\boldsymbol{s}_t$. Assuming that the erasure events of all packets are i.i.d. with probability $1-\beta$ implies that $\{b_t\}$ are i.i.d. according to a Binomial distribution $\mathscr{Bin}(K,\beta)$.

Interestingly, the optimal number of packets K depends on the (total) rate R and successful packet-arrival probability β, since by allocating more lower-rate packets, one trades a lower probability of receiving the maximal available rate at the state estimator with a higher probability of receiving intermediate rates. The optimal K is determined by the number that minimizes $\mathbb{E}\left[2^{-r_t}\right]$, as is demonstrated in Fig. 2.

We note that in the absence of ACKs of intermediate packets, the successive refinement encoding considered here cannot be used. One could use repetition coding to trade multiplexing gain with diversity [46] or multiple description coding [48], when ACKs are sent only after all the intermediate packets are transmitted. We do not discuss such extensions in this paper due to a lack of space.

Remark 9: We only considered uniform rate allocations for all the packets. Clearly, one can generalize the same approach to nonuniform packet rates.

Remark 10: In practice, one might expect longer packets to be prone to higher erasure probability. This can be taken into account when deciding on the K that minimizes $\mathbb{E}\left[2^{-2r_t}\right]$.

VI. Packet Erasures With Delayed ACKs

We now tackle the case of i.i.d. packet erasures with ACKs that are delayed by one time unit, i.e., the case where at time t, the observer does not know whether the last packet arrived or not (namely, it does not know b_{t-1}), but knows the erasure pattern of all preceding packets (knows b^{t-2}). The observer (3) and state estimator (4) mappings can be written as [recall the definition of $g_t \triangleq b_t f_t$ in (22)]

$$f_t = \mathcal{E}_t\left(\boldsymbol{s}^t, g^{t-2}\right), \qquad \hat{\boldsymbol{s}}_t = \mathcal{D}_t\left(g^t\right).$$

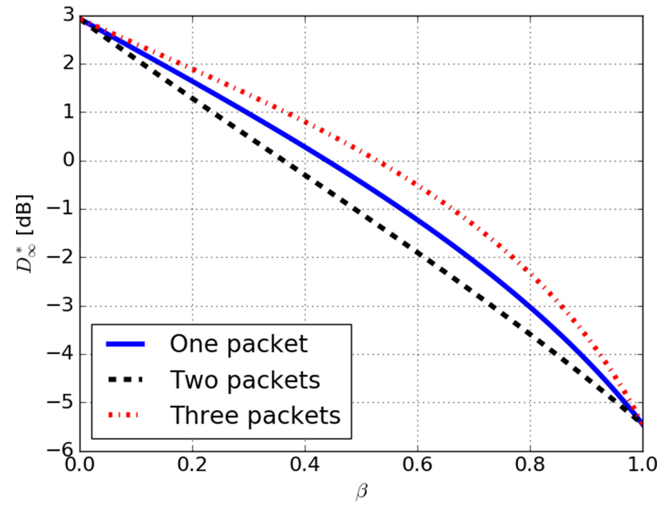

Fig. 2. Evaluation of D^*_∞ for $K = 1, 2,$ and 3 packets, all possible values of $\beta \in [0,1]$, $R = 1$, $\alpha = 0.7$ and $W = 1$.

To construct a transmission scheme for this case, we recall the following result by Perron *et al.* [32, Th. 2], which is a specialization to the jointly Gaussian case of the result by Kaspi [31, Th. 1], who established the rate–distortion region of lossy compression with two-sided SI where the SI may or may not be available at the state estimator.[7,8]

Theorem 3 ([32, Th. 2]): Let $\boldsymbol{s}$ be an i.i.d. zero-mean Gaussian process of power S, which is jointly Gaussian with SI $\boldsymbol{y}$ that is available at the observer and satisfies $\boldsymbol{s} = \boldsymbol{y} + \boldsymbol{z}$, where $\boldsymbol{z}$ is an i.i.d. Gaussian noise of power Z that is independent of $\boldsymbol{y}$. Denote by $\hat{\boldsymbol{s}}^+$ and $\hat{\boldsymbol{s}}^-$ the reconstructions of $\boldsymbol{s}$ with and without the SI $\boldsymbol{y}$, and by D^+ and D^-—their mean squared error distortion requirements, respectively. Then, the smallest rate required to achieve these distortions is given by

$$R^{\text{Kaspi}}(S,Z,D^-,D^+)$$
$$= \begin{cases} 0, & D^- \geq S \text{ and } D^+ \geq Z \\ \frac{1}{2}\log\left(\frac{S}{D^-}\right), & D^- < S \text{ and } D^+ \| S \geq D^- \| Z \\ \frac{1}{2}\log\left(\frac{Z}{D^+}\right), & D^+ < Z \text{ and } D^- \geq D^+ + S - Z \\ \frac{1}{2}\log\left(\frac{S}{D^- - \Delta^2}\right), & \begin{cases} D^- < S \text{ and } D^+ \| S < D^- \| Z \\ \text{and } D^- < D^+ + S - Z \end{cases} \end{cases}$$

where $a \| b \triangleq \frac{ab}{a+b}$ denotes the harmonic mean of a and b, and

$$\Delta \triangleq \frac{\sqrt{(S-Z)(S-D^-)}D^+ - \sqrt{(Z-D^+)(D^- - D^+)}S}{\sqrt{Z}\,(S-D^+)}.$$

[7] We use a backward channel to represent the SI $\boldsymbol{s} = \boldsymbol{y} + \boldsymbol{z}$, as opposed to the forward channel $\boldsymbol{y} = \boldsymbol{s} + \boldsymbol{z}$ used in [32] and [33].

[8] Kaspi's result [31, Th. 1] can also be viewed as a special case of [33] with some adjustments; see [34].

Remark 11: Surprisingly, as observed by Perron *et al.* [32], if the SI signal $\boldsymbol{y}$ is not available at the observer—a setting considered in [31, Th. 2], [33]—the required rate can be strictly higher than that in Theorem 3. This is in stark contrast to the case where the SI is not available at the observer, and the case where the SI is always available at the state estimator studied by Wyner and Ziv [49], [50]. Knowing the SI at the observer allows to (anti)correlate the noise $\boldsymbol{z}$ with the quantization error—an operation that is not possible when the SI is not available at the observer, as the two noises must be independent in that case. This leads to some improvement, though a modest one, as implied by the dual channel-coding results [51, Prop. 1], [52].

In our case, at time t, the previous packet f_{t-1} serves as the SI. Note that this SI is always available to the observer; the state estimator may or may not have access to it, depending on whether the previous packet arrived or not. Since the ACK is delayed, during the transmission of the current packet f_t, the observer does not know whether the previous packet was lost.

The tradeoff between D^+ and D^- for a given rate R will be determined by the probability of a successful packet arrival β.

Scheme 2 (Kaspi based).

Observer. At time t:

1) Generates the prediction error $\tilde{\boldsymbol{s}}_t \triangleq \boldsymbol{s}_t - \alpha_t \alpha_{t-1} \hat{\boldsymbol{s}}_{t-2}$.
2) Generates f_t by quantizing the prediction error $\tilde{\boldsymbol{s}}_t$ as in Theorem 3, where f_{t-1} is available as SI at the observer and possibly at the state estimator (depending on b_{t-1}) using the optimal quantizer of rate R and frame length N that minimizes the distortion averaged over b_{t-1}:
$$D_t^{\text{Weighted}} = \beta D_t^+ + (1-\beta) D_t^- ; \tag{24}$$
more precisely, since the observer does not know (b_{t-1}, b_t) at time t:
   a) Denote the reconstruction of $\tilde{\boldsymbol{s}}_t$ at the state estimator from f_t and g^{t-2}—namely given that $b_t = 1$ and $b_{t-1} = 0$ — by $Q_t^-(\tilde{\boldsymbol{s}}_t)$, and the corresponding distortion by D_t^-.
   b) Denote the reconstruction from (f_{t-1}, f_t) and g^{t-2}—namely given that $b_t = 1$ and $b_{t-1} = 1$ — by $Q_t^+(\tilde{\boldsymbol{s}}_t)$, and the corresponding distortion by D_t^+.
   c) Denote the reconstruction from f_t and g^{t-1}—namely given that $b_t = 1$ — by $Q_t(\tilde{\boldsymbol{s}}_t)$, and the corresponding distortion, averaged over b_{t-1}, by D_t^{Weighted}.

   Then, the observer sees $\alpha_t Q_{t-1}(\tilde{\boldsymbol{s}}_{t-1})$ as possible SI available at the state estimator to minimize D_t^{Weighted} as in (24).
(3) Sends f_t over the channel.

State estimator. At time t:

a) Receives g_t.
b) Generates a reconstruction $\hat{\tilde{\boldsymbol{s}}}_t$ of the prediction error $\tilde{\boldsymbol{s}}_t$
$$\hat{\tilde{\boldsymbol{s}}}_t = \begin{cases} Q_t^+(\tilde{\boldsymbol{s}}_t), & b_t = 1, b_{t-1} = 1 \\ Q_t^-(\tilde{\boldsymbol{s}}_t), & b_t = 1, b_{t-1} = 0 \\ 0, & b_t = 0 \end{cases} \tag{25}$$
c) Generates an estimate $\hat{\boldsymbol{s}}_t$ of $\boldsymbol{s}_t$: $\hat{\boldsymbol{s}}_t = \alpha_t \hat{\boldsymbol{s}}_{t-1} + \hat{\tilde{\boldsymbol{s}}}_t$.

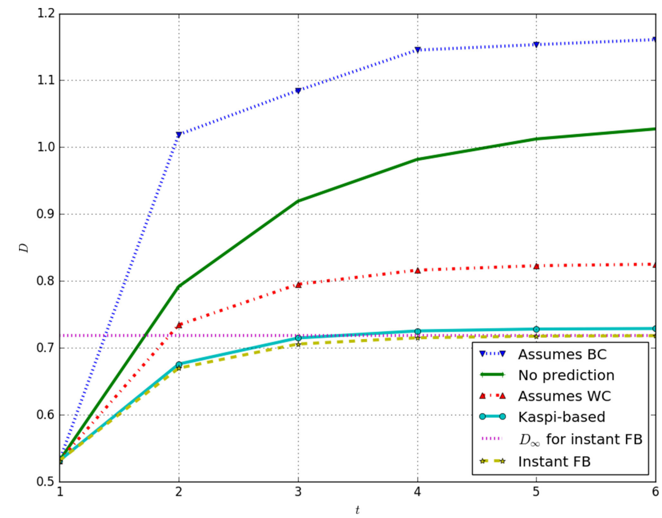

Fig. 3. Distortions D_t as a function of the time t of the various schemes presented in this section, along with that of the instantaneous-ACK scheme of Section V, for $\alpha = 0.7$, $W = 1$, $\beta = 0.5$, and $R = 2$.

This scheme is the optimal greedy scheme whose performance is stated next, in the limit of large N.

Theorem 4: The following distortions D^T can be approached arbitrarily closely in the limit $N \to \infty$ for $t \in [2, T]$:
$$D_t = \begin{cases} D_t^+, & b_t = 1, b_{t-1} = 1 \\ D_t^-, & b_t = 1, b_{t-1} = 0 \\ \alpha_t^2 D_{t-1} + W, & b_t = 0 \end{cases}$$
$$D_1 = D_1^+ = D_1^- = W_t 2^{-b_1 2R} + \epsilon$$
where D_t^+ and D_t^- are the distortions that minimize (24), such that the rate of Theorem 3 satisfies
$$R^{\text{Kaspi}}(\alpha_t D_{t-1}^- + W, \alpha_t D_{t-1}^+ + W, D_t^-, D_t^+) = R.$$

Proof: The proof is again the same as that of Theorems 1 and 2, with $\hat{\tilde{\boldsymbol{s}}}_t$ generated as in (25). ■

Remark 12: Here, in contrast to the case of instantaneous ACKs, evaluating the distortions $\{D_t\}$ in explicit form (recall Corollary 4) is more challenging. We do it numerically, instead.

Somewhat surprisingly, the loss in performance of the Kaspi-based scheme due to the ACK delay is rather small compared to the scenario in Section V where the ACKs are available instantaneously, for all values of β.[9] This is demonstrated in Fig. 3, where the perfomances of these schemes are compared along with the performances of the following three simple schemes for $\alpha_t \equiv 0.7, W \equiv 1, \beta = 0.5, R = 2$:

1) *No prediction*: A scheme that uses no prediction at all, as if the state frames were independent. This scheme achieves a distortion of $D_t = \beta S_t 2^{-2R} + (1-\beta) S_t$, where S_t is the power of the entries of $\boldsymbol{s}_t$ as given in (2).
2) *Assumes worst case*: Since at time t the observer does not know b_{t-1}, a "safe" way would be to work as if $b_{t-1} = 0$.

[9] For β values close to 0 or 1, the loss becomes even smaller as in these cases using the scheme of Section V that assumes that the previous packet arrived or was erased, respectively, becomes optimal.

This achieves a distortion of

$$\begin{aligned} D_t &= \left[\alpha^4 D_{t-2} + (1+\alpha^2)W\right]\left[\beta 2^{-2R} + (1-\beta)^2\right]; \\ &\quad + \beta(1-\beta)(\alpha^2 D_{t-1} + W), \qquad t = 2, \ldots, T; \\ D_0 &= 0, \qquad D_1 = W2^{-2R}. \end{aligned}$$

3) *Assumes best case*: The optimistic counterpart of the previous scheme is that, which always works as if $b_{t-1} = 1$. This scheme achieves a distortion of

$$\begin{aligned} D_t &= \beta\left[\alpha^2 D_{t-1|t-2} 2^{-2R} + W\right)\left[\beta 2^{-2R} + (1-\beta)\right]; \\ &\quad + (1-\beta)\left[\alpha^2 D_{t-1|t-2} + W\right], \quad t = 2, \ldots, T; \\ D_{t-1|t-2} &\triangleq \alpha^2 D_{t-2} + W, \qquad t = 2, \ldots, T; \\ D_0 &= 0, \quad D_1 = W2^{-2R}. \end{aligned}$$

VII. Variable-Length Coding

In contrast to previous sections where at time instant t *exactly* NR_t bits were available for the compression of the N-length vector $\boldsymbol{s}_t$, in this section, we consider the less restrictive case, commonly referred to as VLC, where the (transmit) rate is constrained to R only *on average* across time [35], [36, Ch. 5]. We assume again a packet-erasure case, where, as in Section V-A, the packet at time t is erased with probability $1-\beta$, and successfully arrives with probability β. The packet-erasure events $\{b_t\}$ take values in $\{0, 1\}$ where 0 corresponds to an erasure and 1—to a successful arrival; we assume that these events are i.i.d. We further concentrate on the scalar case, $N = 1$. The rate constraint can be, therefore, written as

$$\begin{aligned} \mathbb{E}\left[r_t | b_t = 1\right] &\le R, \\ \mathbb{E}\left[r_t | b_t = 0\right] &= 0, \end{aligned} \qquad t \in [T] \tag{26}$$

where, in contrast to previous sections, in this section, r_t *can depend on the exact value of* $\boldsymbol{s}^t$.

Remark 13: Similarly to the treatment in Section V-B, the treatment in this section can be extended to the case of multiple packets per state frame.

We first note that the lower bound of Theorem 2 remains valid for the VLC case, since Shannon's classical rate–distortion theorem [53]–[55] extends to the case of VLC (see, e.g., [56]). We next prove that this lower bound can be closely met by incorporating ECDQ [37]–[39], [40, Ch. 5], which is described as follows.

Scheme 3 (ECDQ).

Offline: The observer and the state estimator generate a common random dither z that is uniformly distributed over $[-\Delta/2, \Delta/2)$.

Observer.

1) Uses a uniform-grid (one-dimensional lattice) quantizer with quantization step Δ to quantize $\gamma s + z$: $Q_\Delta(\gamma s + z)$, where γ is a predetermined scalar.
2) Applies entropy coding to the output of the quantizer.
3) Sends the output of the entropy coder.

State estimator.

1) Receives the coded bits.
2) Reconstructs the output of the quantizer: $Q_\Delta(\gamma s + z)$.
3) Generates the state estimate by subtracting z from the quantizer's output and multiplies the result by γ

$$\hat{s} = \gamma\left[Q_\Delta(\gamma s + z) - z\right].$$

Theorem 5 (ECDQ performance [39], [40, Ch. 5]): The average rate R needed by the ECDQ scheme (for $N = 1$) to achieve a distortion D for a state s with variance S and γ set to $\gamma = \sqrt{1 - D/S}$ is bounded from above by

$$R \le \frac{1}{2}\log\frac{S}{D} + \frac{1}{2}\log\frac{2\pi \mathrm{e}}{12} \tag{27}$$

where the first element in (27) is the Gaussian rate–distortion function and the second element is the "shaping loss."

Equivalently, the average distortion D of ECDQ under an average rate constraint R (26) is bounded from above by

$$D \le \frac{2\pi\mathrm{e}}{12} S 2^{-2R}. \tag{28}$$

Remark 14 (One-to-one source coding): The entropy coding employed here is assumed to be one-to-one, that is, we do not require the resulting code to be prefix free. For a more thorough discussion of one-to-one versus prefix-free coding and the rationale behind using each, see Section IX-C.

Remark 15 (ECDQ for $N > 1$): For $N > 1$, one may replace the uniform scalar quantizer with a lattice-based one; the resulting distortion in this case is bounded from above by

$$R \le \frac{1}{2}\log\frac{S}{D} + \frac{1}{2}\log\left(2\pi\mathrm{e}G_N\right)$$

where G_N is the normalized second moment of the lattice [40, Ch. 3.2]. For the special case of a scalar lattice, $G_1 = 1/12$. It is known, by the isoperimetric inequality [40, Ch. 7], that $G_N > \frac{1}{2\pi\mathrm{e}}$ for any lattice of any dimensions N. Moreover, it is known that a sequence of lattices of growing dimensions N can be devised that attains this lower-bound in the limit of $N \to \infty$; see [40] for a thorough account of lattices and their application to ECDQ.

We next incorporate ECDQ in the DPCM scheme of Section III-A: we apply ECDQ (with i.i.d. dither z_t across time) to $\tilde{s}_t$, to generate $\hat{\tilde{s}}_t$ at the observer and recover it at the state estimator; the rest of the scheme remains exactly the same. We note that a similar scheme in the context of networked control (albeit without packet erasures) was previously proposed and analyzed in [41]. The performance of Scheme 3 is stated next.

Theorem 6 (ECDQ-based DPCM scheme performance): The ECDQ-based DPCM scheme (for $N = 1$) under an average rate constraint R (26) achieves a distortion D_t at time t that satisfies the recursion

$$D_t \le \frac{2\pi\mathrm{e}}{12} B\left(\alpha_t^2 D_{t-1} + W_t\right) \tag{29}$$

with $D_0 = 0$ and B as in (23).

Theorem 6 suggests that the gap in performance of scalar systems compared to their N-dimensional counterparts is bounded by a multiplicative factor of $2\pi\mathrm{e}/12$ in each recursive step (29).

Proof: The proof is identical to that in Section III-A and of Theorem 2, with $D_t \le (\alpha_t^2 D_{t-1} + W)B$ replaced with $D_t \le \frac{2\pi\mathrm{e}}{12}(\alpha_t^2 D_{t-1} + W)B$, due to the shaping loss of ECDQ. ■

Remark 16 (ECDQ-based DPCM scheme for $N > 1$): Following Remark 15, for the case of $N > 1$ the resulting distortion when applying ECDQ for $N > 1$ with an N-dimensional lattice is bounded from above by

$$D_t \leq 2\pi \mathrm{e} G_N B \left(\alpha_t^2 D_{t-1} + W_t\right)$$

where again $D_0 = 0$, G_N is the normalized second moment of the lattice and B is given in (23).

In the limit of large T, we attain the following steady-state distortion.

Corollary 6 (ECDQ-based DPCM scheme in steady-state): If $\frac{2\pi \mathrm{e}}{12}\alpha^2 B < 1$, then the steady-state distortion of the ECDQ-based DPCM scheme (for $N = 1$) under an average rate constraint R (26) is bounded from above by

$$D_\infty \leq \frac{\frac{2\pi \mathrm{e}}{12} W B}{1 - \frac{2\pi \mathrm{e}}{12}\alpha^2 B} \tag{30}$$

where B is given in (23).

Remark 17 (Stabilizability): The stabilizability condition $\frac{2\pi \mathrm{e}}{12}\alpha^2 B < 1$ is distant from that of the case of large frames by the shaping loss $\frac{2\pi \mathrm{e}}{12}$. This can be alliviated by applying downsampling, i.e., sending κR bits (on average) every $\kappa \in \mathbb{N}$ samples and remaining silent during the rest; the resulting stabilizability condtion in this case becomes $\sqrt[k]{\frac{2\pi \mathrm{e}}{12}}\alpha^2 B < 1$.

VIII. Application to Networked Control

An important application of state tracking is to networked control, namely, to the scenario where, in contrast to traditional control, the observer is not colocated with the controller, and communicates with it instead via a noiseless (packeted) channel. Hence, the controller assumes the additional role of a state estimator.

We concentrate on the following simple setting, also depicted in Fig. 4. The channel is the noiseless random-rate budget channel of Section IV.

We consider a stochastic system with linear scalar plant evolution, which is the same as in (1) (with $s_0 = 0$):

$$s_t = \alpha s_{t-1} + w_t + u_{t-1}$$

where the coefficient α (which is usually assumed to be fixed across time in control applications) can be greater than 1 in its absolute value, corresponding to an unstable open-loop process, with the additional term u_{t-1} serving as the control action that is generated by the controller from all past packets f^{t-1}, and is used to stabilize the system.

We consider the random-rate budget scenario of Section IV.

The goal of the system is to minimize the average-stage LQG cost upon reaching the horizon T:

$$\bar{\mathrm{J}}_T \triangleq \frac{1}{T}\mathbb{E}\left[\sum_{t=1}^{T-1}\left(\mathsf{Q}_t s_t^2 + \mathsf{R}_t u_t^2\right) + \mathsf{Q}_T s_T^2\right] \tag{31}$$

where $\{\mathsf{Q}_t\}$ and $\{\mathsf{R}_t\}$ are known nonnegative scalars, respectively, that penalize the cost for state deviations and control actuations, respectively.

In order to derive bounds on the LQG cost for this setting, we use a result by Fischer [57] and by Tatikonda *et al.* [8] that extends the celebrated control-theoretic separation principle to networked control systems.

Lemma 1 ([8], [57]): The optimal controller is given by

$$u_t = -K_t \hat{s}_t$$

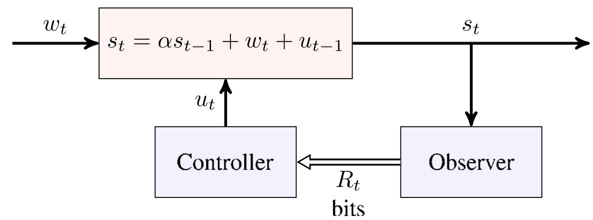

Fig. 4. Linear control system with a finite-rate feedback.

where $\hat{s}_t \triangleq \mathbb{E}\left[s_t | f^t\right]$, K_t is the optimal linear quadratic regulator (LQR) control gain

$$K_t = \frac{L_{t+1}}{\mathsf{R}_t + L_{t+1}}\alpha$$

and L_t satisfies the dynamic backward Riccati recursion [58]

$$L_t = \mathsf{Q}_t + \alpha \mathsf{R}_t K_t$$

with $L_{T+1} = 0$.[10] Moreover, this controller achieves a cost of

$$\bar{\mathrm{J}}_T = \frac{1}{T}\sum_{t=1}^{T}\left\{W L_t + \alpha K_t L_{t+1}\mathbb{E}\left[\left(s_t - \hat{s}_t\right)^2\right]\right\}$$

where we use the convention $R_T = 0$ and $f_T = 0$ for the definition of $\hat{s}_T$, as no transmission or control action are performed at time T.

A. Lower Bound

By substituting the result of Theorem 2 into Lemma 1, we attain the following lower bound for the achievable LQG cost, which extends the result of [43] to the case of random-rate budgets (packet-erasure scenario included).

Theorem 7 (LQG cost lower bound): The optimal LQG cost (31) with rate tuple R^T is bounded from below by

$$\bar{\mathrm{J}}_T \geq \frac{1}{T}\sum_{t=1}^{T}\left\{W L_t + \alpha K_t L_{t+1} D_t^*\right\}$$

where K_t and L_t are given in Lemma 1, and D_t^* is given in (18).

Proof: The proof is immediate by noting that, similarly to the Performance Analysis of Section III-A, at time t, given f^t, all the past control actions u^{t-1}—being a deterministic function of f^{t-1}—can be absorbed into $\hat{s}_t$. ■

B. Variable-Length Coding

Similarly to the proof of Theorem 7, by combining the results of Theorem 6 and Lemma 1 we attain the following upper bound for the achievable LQG cost, in the VLC scenario; following the exposition in Section VII, we concentrate here on the packet-erasure channel.

Theorem 8 (VLC LQG cost upper bound): The LQG cost (31) for the VLC scenario under an average-rate constraint R (26), is bounded from above by

$$\bar{\mathrm{J}}_T \leq \frac{1}{T}\sum_{t=1}^{T}\left\{W L_t + \alpha K_t L_{t+1} D_t\right\}$$

where K_t and L_t are given in Lemma 1, and D_t is bounded from above as in (29).

[10] In case $\mathsf{R}_T = 0$, define $K_T = 0$.

Proof: Again, the proof is immediate by noting that, similarly to the impossibility proof of Section III-B, at time t, given f^t, all the past control actions u^{t-1}—being a deterministic function of f^{t-1}—are fully determined. ∎

C. Steady State

We consider here the fixed-parameter fixed-rate case

$$\mathsf{Q}_t \equiv \mathsf{Q} \tag{32a}$$

$$\mathsf{R}_t \equiv \mathsf{R} \tag{32b}$$

$$R_t \equiv R \tag{32c}$$

and similarly to the steady-state distortion (8) and average-stage steady-state distortion (10), we wish to determine the optimal steady-state average-stage cost

$$\bar{\mathrm{J}}_\infty \triangleq \limsup_{T\to\infty} \bar{\mathrm{J}}_T \, .$$

Corollary 7 (LQG cost lower bound): The steady-state LQG cost for the fixed-parameter fixed-rate case (32) is bounded from below by

$$\bar{\mathrm{J}}_\infty \geq W L_\infty + \alpha K_\infty L_\infty D_\infty^* \tag{33}$$

where D_∞^* is given in (20):

$$K_\infty = \frac{L_\infty}{\mathsf{R} + L_\infty}\alpha, \tag{34}$$

and L_∞ is the positive solution of

$$L_\infty^2 - \left[\left(\alpha^2 - 1\right)\mathsf{R} + \mathsf{Q}\right] L_\infty - \mathsf{QR} = 0. \tag{35}$$

Remark 18: As noted in Section VII, the result of Corollary 7 holds true for VLC and, hence, also for the more restrictive FLC.

Remark 19 (Comparison to separation-based bounds): In [43], it is shown that the optimal steady-state LQG cost must satisfy (33) with the distortion D_∞^* dictated by the source–channel separation between the *causal rate–distortion* $R_{\mathrm{C}}(D_\infty)$ [6], [9] and the directed capacity (maximal directed information) [59]. Since in our case the directed capacity is upper bounded by the regular capacity of the channel, $C = \mathbb{E}\left[r_1\right]$, and the causal rate–distortion function (which is in itself a lower bound) is given by [6], [10]

$$R_{\mathrm{C}}(D_\infty^*) = \frac{1}{2}\log\left(\alpha^2 + \frac{W}{D_\infty^*}\right),$$

the source–channel separation-based bound $R_{\mathrm{C}}(D_\infty^*) \leq C$ reduces to the expression in (20) with $B \triangleq \mathbb{E}\left[2^{-2r_1}\right]$ replaced with $B_{\mathrm{Sep}} \triangleq 2^{-2\mathbb{E}[r_1]}$. By applying Jensen's inequality we see that $B < B_{\mathrm{Sep}}$ for any nondeterministic rate budget distirubtion. Thus, the joint source and channel treatment offered in this work strengthens the separation-based adaptation of the results in [43]. The difference becomes especially pronounced in the packet-erasure and instantaneous ACKs scenario of Section V-A with an infinite transmission rate R [recall (21)]—A setting extensively studied in the past two decades [26], [60], [61]. In this case, B_{Sep}, and consequently also the lower bound on D_∞^*, reduces to the trivial zero bound, whereas $B = 1 - \beta > 0$ unless $\beta = 1$.

Corollary 8 (VLC LQG cost upper bound): The steady-state LQG cost for the packet-erasure fixed-parameter case (32a), (32b) under an average rate constraint R (26) is bounded from above by

$$\bar{\mathrm{J}}_\infty \leq W L_\infty + \alpha K_\infty L_\infty D_\infty$$

where D_∞, K_∞, L_∞ are given in (30), (35) and (34), respectively.

IX. Discussion

A. ACKs With Larger Delays

To extend the scheme of Section VI for the case of delayed ACKs by one time instant to larger delays, a generalization of Theorem 3 is needed. Unfortunately, the optimal rate–distortion region for more than two SI options (e.g., with or without correlated SI $\boldsymbol{y}$) remains an open problem and is only known for the (degraded) case when the state and the possible SIs form a Markov chain. Nonetheless, achievable regions for multiple SI options have been proposed in [33], which can be used for the construction of schemes that accommodate larger delays.

B. Scalar FLC

In this paper, we derived lower bounds and proved that they are tight in the limit of large values of N. In the case of scalar FLC quantization, both design and analysis of good schemes are more involved and remain beyond the scope of this paper. For a treatment of the case of logarithmically-concave noise distributions (Gaussian included), see [62].

C. Prefix-Free Versus One-Shot Lossless Compression

The VLC ECDQ-based schemes throughout this paper employed one-to-one lossless coding. This is a reasonable assumption since, in packeted communications, the descriptions of subsequent symbols may be assumed to be parsed by the underlying protocol, which allows, in turn, to part with the prefix-free constraint and attain better performance [63]. Specifically, the one-bit loss with respect to the entropy of the process of prefix-free coding is circumvented by one-to-one coding [64]. Nonetheless, the results of this paper can be easily adjusted to the prefix-free coding case by adding an extra bit on the right-hand side of (27)—the maximal loss of prefix-free entropy coding above the entropy, and replacing the factor $2\pi\mathrm{e}/12$ in (28)–(30) by $2\pi\mathrm{e}/3$.

D. Packet-Erasure Modeling

In this paper, we modeled the packet erasures by an i.i.d. process. Nonetheless, the derived results can be extended far beyond this setting, as is evident from the proof of Theorem 2.

In the VLC setting, the erasure probability is likely to be higher for longer packets, and calls for further investigation.

E. Non-Gaussian

Following Assertion 1, the lower bounds in this paper can be extended to the case of a non-Gaussian driving process $\boldsymbol{w}_t$ in a straightforward fashion, with the variance of the elements of $\boldsymbol{w}_t$ in (18) replaced by its EP (recall that the two are equal in the Gaussian case), resulting in lower bounds reminiscent of those in [43].

REFERENCES

[1] H. Viswanathan and T. Berger, "Sequential coding of correlated sources," *IEEE Trans. Inf. Theory*, vol. 46, no. 1, pp. 236–246, Jan. 2000.

[2] N. Ma and P. Ishwar, "On delayed sequential coding of correlated sources," *IEEE Trans. Inf. Theory*, vol. 57, no. 6, pp. 3763–3782, Jun. 2011.

[3] N. Ma and P. Ishwar, "Erratum to "on delayed sequential coding of correlated sources"," *IEEE Trans. Inf. Theory*, vol. 58, no. 6, pp. 4074–4074, Jun. 2012.

[4] E.-H. Yang, L. Zheng, and D.-K. He, "Rate distortion theory for causal video coding: Characterization, computation algorithm, and comparison," *IEEE Trans. Inf. Theory*, vol. 57, no. 8, pp. 5258–5280, Aug. 2011.

[5] E.-H. Yang, L. Zheng, and D.-K. He, "On the information theoretic performance comparison of causal video coding and predictive video coding," *IEEE Trans. Inf. Theory*, vol. 60, no. 3, pp. 1428–1446, Mar. 2014.

[6] S. C. Tatikonda, "Control under communication constraints," Ph.D. dissertation, Dept. Elect. Eng. Comut. Sci., Massachusetts Inst. Technol., Cambridge, CA, USA, Sep. 2000.

[7] V. S. Borkar, S. K. Mitter, and S. Tatikonda, "Optimal sequential vector quantization of Markov sources," *SIAM J. Control Opt.*, vol. 40, no. 1, pp. 135–148, May 2001.

[8] S. Tatikonda, A. Sahai, and S. K. Mitter, "Stochastic linear control over a communication channel," *IEEE Trans. Autom. Control*, vol. 49, no. 9, pp. 1549–1561, Sep. 2004.

[9] A. K. Gorbunov and M. S. Pinsker, "Nonanticipatory and prognostic epsilon entropies and message generation rates," *(in Russian) Problemy Pered. Info. (Problems Info. Trans.)*, vol. 9, no. 3, pp. 12–21, 1973.

[10] A. K. Gorbunov and M. S. Pinsker, "Prognostic epsilon entropy of a Gaussian message and a Gaussian source," *(in Russian) Problemy Pered. Info. (Problems Info. Trans.)*, vol. 10, no. 2, pp. 5–25, 1974.

[11] T. Tanaka, K.-K. K. Kim, P. A. Parrilo, and S. K. Mitter, "Semidefinite programming approach to Gaussian sequential rate–distortion trade-offs," *IEEE Trans. Autom. Control*, vol. 62, no. 4, pp. 1896–1910, Apr. 2017.

[12] C. D. Charalambous, P. A. Stavrou, and N. U. Ahmed, "Nonanticipative rate distortion function and relations to filtering theory," *IEEE Trans. Autom. Control*, vol. 59, no. 4, pp. 937–952, Apr. 2014.

[13] P. Elias, "Predictive coding," Ph.D. dissertation, Faculty Arts Sci., Harvard Univ., Cambridge, CA, USA, May 1950.

[14] P. Elias, "Predictive coding—Parts I and II," *IRE Trans. Inf. Theory*, vol. 1, no. 1, pp. 16–33, Mar. 1955.

[15] B. M. Oliver, "Efficient coding," *Bell Syst. Tech. J.*, vol. 31, no. 4, pp. 724–763, Jul. 1952.

[16] C. W. Harrison, "Experiments with linear prediction in television," *Bell Sys. Tech. J.*, vol. 31, no. 4, pp. 764–783, Jul. 1952.

[17] C. C. Cutler, "Differential quantization of communication signals," Patent US 2/605 361, Jul. 29, 1952.

[18] N. S. Jayant and P. Noll, *Digital Coding of Waveform*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1984.

[19] M. Torbatian and E.-H. Yang, "Causal coding of multiple jointly Gaussian sources," in *Proc. Annu. Allerton Conf. Commun., Control, Comput.*, Monticello, IL, USA, Oct. 2012, pp. 2060–2067.

[20] M. Eslamifar, "On causal video coding with possible loss of the first encoded frame," Master's thesis, Dept. Electron. Commun. Eng., Univ. Waterloo, Waterloo, ON, Canada, 2013.

[21] L. Song, J. Chen, J. Wang, and T. Liu, "Gaussian robust sequential and predictive coding," *IEEE Trans. Inf. Theory*, vol. 59, no. 6, pp. 3635–3652, Jun. 2013.

[22] H. C. Huang, W. H. Peng, and T. Chiang, "Advances in the scalable amendment of H.264/AVC," *IEEE Commun. Mag.*, vol. 45, no. 1, pp. 68–76, Jan. 2007.

[23] Y.-Z. Huang, Y. Kochman, and G. W. Wornell, "Causal transmission of colored source frames over packet erasure channel," in *Proc. Data Comput. Conf.*, Snowbird, UT, USA, Mar. 2010, pp. 129–138.

[24] F. Etezadi, A. Khisti, and M. Trott, "Zero-delay sequential transmission of Markov sources over burst erasure channels," *IEEE Trans. Inf. Theory*, vol. 60, no. 8, pp. 4584–4613, Aug. 2014.

[25] F. Etezadi, A. Khisti, and J. Chen, "A truncated prediction framework for streaming over erasure channels," *IEEE Trans. Inf. Theory*, vol. 63, no. 11, pp. 7322–7351, Nov. 2017.

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

[27] K. You and L. Xie, "Minimum data rate for mean square stabilization of discrete LTI systems over lossy channels," *IEEE Trans. Autom. Control*, vol. 55, no. 10, pp. 2373–2378, Oct. 2010.

[28] S. Yüksel and S. P. Meyn, "Random-time, state-dependent stochastic drift for Markov chains and application to stochastic stabilization over erasure channels," *IEEE Trans. Autom. Control*, vol. 58, no. 1, pp. 47–59, Jan. 2013.

[29] B. R. Andrievsky, A. S. Matveev, and A. L. Fradkov, "Control and estimation under information constraints: Toward a unified theory of control, computation and communications," *Autom. Remote Control*, vol. 71, no. 4, pp. 572–633, Apr. 2010.

[30] P. Minero, M. Franceschetti, S. Dey, and G. N. Nair, "Data rate theorem for stabilization over time-varying feedback channels," *IEEE Trans. Autom. Control*, vol. 54, no. 2, pp. 243–255, Feb. 2009.

[31] A. H. Kaspi, "Rate–distortion when side-information may be present at the decoder," *IEEE Trans. Inf. Theory*, vol. 40, no. 6, pp. 2031–2034, Nov. 1994.

[32] E. Perron, S. Diggavi, and I. E. Telatar, "On the role of encoder side-information in source coding for multiple decoders," in *Proc. IEEE Int. Symp. Inf. Theory*, Seattle, WA, USA, Jul. 2006, pp. 331–335.

[33] C. Heegard and T. Berger, "Rate–distortion when side information may be absent," *IEEE Trans. Inf. Theory*, vol. IT-31, no. 6, pp. 727–734, Nov. 1985.

[34] A. Khina and U. Erez, "Source coding with composite side information at the decoder," in *Proc. IEEE Conv. Elect. Electron. Eng. Israel*, Eilat, Israel, Nov. 2012.

[35] D. Huffman, "A method for the construction of minimum redundancy codes," *Proc. IRE*, vol. 40, no. 9, pp. 1098–1101, Sep. 1952.

[36] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. New York, NY, USA: Wiley, 2006.

[37] J. Ziv, "On universal quantization," *IEEE Trans. Inf. Theory*, vol. IT-31, no. 3, pp. 344–347, May 1985.

[38] R. Zamir and M. Feder, "On universal quantization by randomized uniform/lattice quantizer," *IEEE Trans. Inf. Theory*, vol. 38, no. 2, pp. 428–436, Mar. 1992.

[39] R. Zamir and M. Feder, "Information rates of pre/post filtered dithered quantizers," *IEEE Trans. Inf. Theory*, vol. 42, no. 5, pp. 1340–1353, Sep. 1996.

[40] R. Zamir, *Lattice Coding for Signals and Networks*. Cambridge, U.K.: Cambridge Univ. Press, 2014.

[41] M. S. Derpich and J. Østergaard, "Improved upper bounds to the causal quadratic rate–distortion function for Gaussian stationary sources," *IEEE Trans. Inf. Theory*, vol. 58, no. 5, pp. 3131–3152, May 2012.

[42] E. I. Silva, M. S. Derpich, and J. Østergaard, "A framework for control system design subject to average data-rate constraints," *IEEE Trans. Autom. Control*, vol. 56, no. 8, pp. 1886–1899, Aug. 2011.

[43] V. Kostina and B. Hassibi, "Rate–cost tradeoffs in control," in *Proc. Annu. Allerton Conf. Commun., Control, Comput.*, Monticello, IL, USA, Sep. 2016, pp. 1157–1164.

[44] P. J. Antsaklis and A. N. Michel, *Linear Systems* (Electrical and Computer Engineering). New York, NY, USA: McGraw-Hill, 1997.

[45] T. Tanaka, "Semidefinite representation of sequential rate–distortion function for stationary Gauss–Markov processes," in *Proc. IEEE Conf. Control Appl.*, Sydney, NSW, Australia, Sep. 2015, pp. 1217–1222.

[46] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge, U.K: Cambridge Univ. Press, 2005.

[47] A. El Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge, U.K.: Cambridge Univ. Press, 2011.

[48] J. Østergaard and D. Quevedo, "Multiple descriptions for packetized predictive control," *EURASIP J. Adv. Signal Proc.*, vol. 2016, no. 45, Apr. 2016.

[49] A. D. Wyner and J. Ziv, "The rate–distortion function for source coding with side information at the decoder," *IEEE Trans. Inf. Theory*, vol. IT-22, no. 1, pp. 1–10, Jan. 1976.

[50] A. D. Wyner, "The rate–distortion function for source coding with side information at the decoder—II: General sources," *Inf. Control*, vol. 38, pp. 60–80, 1978.

[51] R. Zamir and U. Erez, "A Gaussian input is not too bad," *IEEE Trans. Inf. Theory*, vol. 50, no. 6, pp. 1362–1367, Jun. 2004.

[52] T. Philosof and R. Zamir, "The cost of uncorrelation and noncooperation in MIMO channels," *IEEE Trans. Inf. Theory*, vol. 53, no. 11, pp. 3904–3920, Nov. 2007.

[53] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423, Jul. 1948.

[54] C. E. Shannon, "Coding theorems for a discrete source with a fidelity criterion," in *Proc. Inst. Radio Eng., Int. Convention Record*, vol. 7, 1959, pp. 142–163.

[55] T. Berger, *Rate Distortion Theory: A Mathematical Basis for Data Compression*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1971.
[56] T. Hashimoto, "Source coding for average rate and average distortion: New variable-length coding theorems," *IEEE Trans. Inf. Theory*, vol. IT-29, no. 6, pp. 785–792, Nov. 1983.
[57] T. Fischer, "Optimal quantized control," *IEEE Trans. Autom. Control*, vol. AC-27, no. 4, pp. 996–998, Aug. 1982.
[58] D. P. Bertsekas, *Dynamic Programming and Optimal Control*, vol. I, 2nd ed. Belmont, MA, USA: Athena Scientific, 2000.
[59] J. Massey, "Causality, feedback and directed information," in *Proc. IEEE Int. Symp. Info. Theory Apps.*, Waikiki, HI, USA, Nov. 1990, pp. 303–305.
[60] B. Sinopoli *et al.*, "Kalman filtering with intermittent observations," *IEEE Trans. Autom. Control*, vol. 49, no. 9, pp. 1453–1464, Sep. 2004.
[61] V. Gupta, A. F. Dana, J. P. Hespanha, R. M. Murray, and B. Hassibi, "Data transmission over networks for estimation and control," *IEEE Trans. Autom. Control*, vol. 54, no. 8, pp. 1807–1819, Aug. 2009.
[62] A. Khina, Y. Nakahira, Y. Su, and B. Hassibi, "Algorithms for optimal control with fixed-rate feedback," in *Proc. IEEE Conf. Decis. Control*, Dec. 2017, pp. 6015–6020.
[63] N. Alon and A. Orlitsky, "A lower bound on the expected length of one-to-one codes," *IEEE Trans. Inf. Theory*, vol. 40, no. 5, pp. 1670–1672, Sep. 1994.
[64] A. D. Wyner, "An upper bound on the entropy series," *Inf. Control*, vol. 20, pp. 176–181, Mar. 1972.

Anatoly Khina (S'08–M'17) was born in Moscow, USSR, in 1984, and moved to Israel in 1991. He received the B.Sc. (*summa cum laude*), M.Sc. (*summa cum laude*), and Ph.D. degrees from Tel Aviv University, Tel Aviv, Israel, in 2006, 2010, and 2016, respectively, all in electrical engineering.

He is currently a Senior Lecturer with the Department of Electrical Engineering-Systems, Tel Aviv University. He was a Postdoctoral Scholar with the Department of Electrical Engineering, California Institute of Technology, Pasadena, CA, USA, from 2015 to 2018, and a Research Fellow with the Simons Institute for the Theory of Computing, University of California, Berkeley, Berkeley, CA, USA, during the Spring of 2018. His research interests include information theory, control theory, signal processing, and matrix analysis. In parallel to his studies, he was an Engineer in various roles focused on algorithms, software, and hardware R&D.

Dr. Khina is a recipient of the Simons-Berkeley and Qualcomm Research Fellowships; Fulbright, Rothschild and Marie Skłodowska-Curie Postdoctoral Fellowships; Clore Scholarship; Trotsky Award; Weinstein Prize in signal processing; Intel award for Ph.D. research; and the first prize for outstanding research work in the field of communication technologies from the Advanced Communication Center's Feder Family Award program.

Victoria Kostina (S'12–M'14) received the Bachelor's degree in applied mathematics and physics from Moscow institute of Physics and Technology, Dolgoprudny, Russia, in 2004, the Master's degree in electrical engineering from the University of Ottawa, Ottawa, ON, Canada, in 2006, and the Ph.D. degree in electrical engineering from Princeton University, Princeton, NJ, USA, in 2013.

In the fall of 2014, she joined Caltech as an Assistant Professor of Electrical Engineering. Her research interests include information theory, coding, control, and communications.

Dr. Kostina is a recipient of the Natural Sciences and Engineering Research Council of Canada postgraduate scholarship (2009–2012), the Princeton Electrical Engineering Best Dissertation Award (2013), the Simons-Berkeley research fellowship (2015), and the NSF CAREER award (2017).

Ashish Khisti (S'02–M'08) received the B.ASc. degree in engineering sciences (electrical option) from the University of Toronto, Toronto, ON, Canada, in 2002, and the S.M. and Ph.D. degrees in electrical engineering from the Massachusetts Institute of Technology, Cambridge, MA, USA, in 2004 and 2009, respectively.

During 2009–2015, he was an Assistant Professor with the Electrical and Computer Engineering Department, University of Toronto where he is currently an Associate Professor and a Canada Research Chair.

Dr. Khisti is a recipient of an Ontario Early Researcher Award, an Hewlett–Packard Innovation Research Award, a Cisco Research Center Award and the Harold H. Hazen Teaching Assistant Award from MIT. He is currently an Associate Editor for IEEE TRANSACTIONS ON INFORMATION THEORY and is also a Guest Editor for the PROCEEDINGS OF THE IEEE (Special Issue on Secure Communications via Physical-Layer and Information-Theoretic Techniques).

Babak Hassibi (M'08) was born in Tehran, Iran, in 1967. He received the B.S. degree from the University of Tehran, Tehran, Iran, in 1989, and the M.S. and Ph.D. degrees from Stanford University, Stanford, CA, USA, in 1993 and 1996, respectively, all in electrical engineering.

Since January 2001, he has been with the California Institute of Technology, Pasadena, CA, USA, where he is currently the Mose and Lilian S. Bohn Professor of Electrical Engineering. From 2013 to 2016, he was the Gordon M. Binder/Amgen Professor of Electrical Engineering and from 2008 to 2015, he was Executive Officer of Electrical Engineering, as well as Associate Director of Information Science and Technology. From October 1996 to October 1998, he was a Research Associate with the Information Systems Laboratory, Stanford University, and from November 1998 to December 2000, he was a Member of the Technical Staff in the Mathematical Sciences Research Center, Bell Laboratories, Murray Hill, NJ, USA. He has also held short-term appointments, Ricoh California Research Center, the Indian Institute of Science, and Linköping University, Sweden.

He is the coauthor of the books (both with A. H. Sayed and T. Kailath) *Indefinite Quadratic Estimation and Control: A Unified Approach to H^2 and H^∞ Theories* (New York: SIAM, 1999) and *Linear Estimation* (Englewood Cliffs, NJ, USA: Prentice Hall, 2000). His research interests include communications and information theory, control and network science, and signal processing and machine learning.

Dr. Hassibi was a recipient of an Alborz Foundation Fellowship, the 1999 O. Hugo Schuck best paper award of the American Automatic Control Council (with H. Hindi and S. P. Boyd), the 2002 NSF Career Award, the 2002 Okawa Foundation Research Grant for Information and Telecommunications, the 2003 David and Lucille Packard Fellowship for Science and Engineering, the 2003 Presidential Early Career Award for Scientists and Engineers (PECASE), and the 2009 Al-Marai Award for Innovative Research in Communications, and was a participant in the 2004 National Academy of Engineering "Frontiers in Engineering" program. He has been a Guest Editor for the IEEE TRANSACTIONS ON INFORMATION THEORY Special Issue on "space-time transmission, reception, coding and signal processing," was an Associate Editor for IEEE TRANSACTIONS ON INFORMATION THEORY during 2004–2006, and is currently an Editor for the *Journal Foundations and Trends in Information and Communication* and for the IEEE TRANSACTIONS ON NETWORK SCIENCE AND ENGINEERING.