Sequential Source Coding for Stochastic Systems Subject to Finite Rate Constraints

Photios A. Stavrou, *Member, IEEE*, Mikael Skoglund, *Fellow, IEEE*, and Takashi Tanaka, *Member, IEEE*

***Abstract*—In this article, we revisit the sequential source-coding framework to analyze fundamental performance limitations of discrete-time stochastic control systems subject to feedback data-rate constraints in finite-time horizon. The basis of our results is a new characterization of the lower bound on the minimum total-rate achieved by sequential codes subject to a total (across time) distortion constraint and a computational algorithm that allocates optimally the rate-distortion, for a given distortion level, at each instant of time and any fixed finite-time horizon. The idea behind this characterization facilitates the derivation of *analytical*, *nonasymptotic*, and *finite-dimensional* lower and upper bounds in two control-related scenarios: a) A parallel time-varying Gauss–Markov process with identically distributed spatial components that are quantized and transmitted through a noiseless channel to a minimum mean-squared error decoder; and b) a time-varying quantized linear quadratic Gaussian (LQG) closed-loop control system, with identically distributed spatial components and with a random data-rate allocation. Our nonasymptotic lower bound on the quantized LQG control problem reveals the absolute minimum data-rates for (mean square) stability of our time-varying plant for any fixed finite-time horizon. We supplement our framework with illustrative simulation experiments.**

***Index Terms*—Finite-time horizon, quantization, reverse-waterfilling, sequential causal coding, stochastic systems.**

I. Introduction

ONE of the fundamental characteristics of networked control systems (NCSs) [1] is the existence of an imperfect communication network between computational and physical entities. In such setups, an analytical framework to assess impacts of communication and data-rate limitations on the control performance is strongly required. In this article, we adopt information-theoretic tools to analyze such requirements. Specifically, we consider *sequential coding of correlated sources* initially introduced by [2] (see also [3]), which is a generalization of the successive refinement source-coding problem [4], [5]. In successive refinement, source coding is performed in (time) stages and the goal is to "refine" the description of the source with every new stage when more information is available. Sequential coding differs from successive refinement in that at the second stage, encoding involves describing a correlated (in time) source as opposed to improving the description of the same source. To accomplish this task, sequential coding encompasses a spatio-temporal coding method. In addition, sequential coding is a temporally zero-delay coding paradigm since both *encoding* and *decoding* must occur in real time. The resulting zero-delay coding approach should not be confused with other existing works on zero-delay coding, see, e.g., [6]–[11], because it relies on the use of a spatio-temporal coding approach whereas the aforementioned papers rely solely on temporal coding approaches.

A. Literature Review on Sequential Source Coding

In what follows, we provide a detailed literature review on sequential source coding. To shed more light on the historical route of this coding paradigm, we distinguish the work of [2] (see also [12], [13]) with the work of [3] because although their results complement each other, their underlining motivation has been different. Indeed, Viswanathan and Berger [2] initiated this coding approach targeting video-coding applications, whereas Tatikonda [3] aimed to develop a framework for delay-constrained systems and to study the communication theory in classical closed-loop control setups.

Sequential Coding Via [2]: The authors of [2] characterized the minimum achievable rate-distortion region for two temporally correlated random variables (RVs) with each being a vector of spatially independent and identically distributed (IID) processes (also called "frames"), subject to a coupled average distortion criterion. In the last decade, sequential coding approach of [2] was further studied in [12]–[14]. Ma and Ishwar [12] used an extension of [2] to three time instants subject to a per-time distortion constraint to investigate the effect of sequential coding when possible coding delays occur within a multi-input–multi-output system. Around the same time, Yang *et al.* [13] generalized [2] to a finite number of time instants.

Manuscript received 19 April 2021; accepted 17 August 2021. Date of publication 8 September 2021; date of current version 29 July 2022. The work of Photios A. Stavrou and Mikael Skoglund was supported by the KAW Foundation, and the Swedish Research Council (VR) under Grant 2019-03606. The work of Takashi Tanaka was supported by the National Science Foundation (NSF) CAREER Award under Grant 1944318. Recommended by Associate Editor W. X. Zheng. *(Corresponding author: Photios A. Stavrou.)*

Photios A. Stavrou and Mikael Skoglund are with the Division of Information Science and Engineering, KTH Royal Institute of Technology, 114 28 Stockholm, Sweden (e-mail: fstavrou@kth.se; skoglund@s3.kth.se).

Takashi Tanaka is with the Department of Aerospace Engineering and Engineering Mechanics, University of Austin, Austin, TX 78712 USA (e-mail: ttanaka@utexas.edu).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TAC.2021.3110430.

Digital Object Identifier 10.1109/TAC.2021.3110430

Compared to [2], [12], their spatio-temporal source process is correlated over time, whereas each frame is spatially jointly stationary and totally ergodic subject to a per-time average distortion criterion. More recently, Yang *et al.* [14] drew connections between sequential causal coding and *predictive sequential causal coding*, that is, for (first-order) Markov sources subject to a single-letter fidelity constraint, sequential causal coding and sequential predictive coding coincide. For three time instants of an IID vector source containing jointly Gaussian correlated processes (not necessarily Markov), an explicit expression of the minimum achievable sum-rate for a per-time mean-squared error (MSE) distortion is obtained in [15]. Inspired by the framework of [2], [12], Khina *et al.* [16] derived fundamental performance limitations in control-related applications. In particular, they considered a multitrack system that tracks several parallel time-varying Gauss–Markov processes with IID spatial components conveyed over a single shared wireless communication link (possibly prone to packet drops) to a minimum mean-squared error (MMSE) decoder. In their Gauss–Markov multitracking scenario, they provided lower and upper bounds in finite-time and in the per-unit time asymptotic limit for the distortion-rate region of time-varying Gauss–Markov sources subject to a mean-squared error (MSE) distortion constraint. Their lower bound is characterized by a forward-in time-distortion allocation algorithm operating with *fixed data-rates at each time instant for a finite time horizon* whereas their upper bound is obtained by means of a differential pulse-code modulation (DPCM) scheme using entropy-coded dithered quantization (ECDQ) with one-dimensional lattice constrained by *uniform data rates across time*. Subsequently, they used these bounds in a scalar-valued quantized linear quadratic Gaussian (LQG) closed-loop control problem to find similar bounds on the minimum cost of control.

Sequential Coding Via [3]: Tatikonda in his Ph.D thesis [3, Ch. 5] (see also [17]) studied sequential source coding in the context of delay-constrained and control-related applications. Therein an information theoretic quantity called *sequential rate distortion function (*RDF*)* was introduced that is attributed to the prior works of Gorbunov and Pinsker [18], [19]. Using the sequential RDF, Tatikonda *et al.* [20] studied the performance analysis and synthesis of a multidimensional fully observable time-invariant Gaussian closed-loop control system when a memoryless communication link exists between a stochastic linear plant and a controller and the performance criterion is the classical linear quadratic cost. The use of sequential RDF (also encountered as nonanticipative or causal RDF in the literature) in filtering applications is stressed in Refs. [21]–[23]. Analytical suboptimal expressions of lower and upper bounds for the setup of [20], including the cases where a linear fully observable time-invariant plant is driven by IID non-Gaussian noise processes or when the system is modeled by time-invariant partially observable Gaussian processes, are derived in [24]. Tanaka *et al.* [25], [26] studied the performance analysis and synthesis of a linear fully observable and partially observable Gaussian closed-loop control problem when the performance criterion is the linear quadratic cost. Moreover, they showed that one can derive lower bounds in finite time and in the per-unit time asymptotic limit by casting the problems as semidefinite representable and thus numerically computable by known solvers. An achievability bound on the asymptotic limit using a DPCM-based ECDQ scheme that uses one-dimensional quantizer at each dimension was also proposed. Lower and upper bounds for a general closed-loop control system subject to *asymptotically average total data-rate constraints* across the time are also investigated in [27], [28]. The lower bounds are obtained using sequential coding and directed information [29], whereas the upper bounds are obtained via a sequential ECDQ scheme using scalar quantizers.

B. Contributions

In this article, we first revisit the sequential coding framework developed by [2], [3], [12], [13] to obtain the following new results.

1) Analytical, nonasymptotic, and finite-dimensional lower and upper bounds on the minimum achievable total-rates (per-dimension) for a multitrack communication scenario similar to the one considered in [16]. However, compared to [16], which derived distortion-rate bounds via forward recursions with *given data rates across a finite time horizon*, here we derive a lower bound subject to a dynamic reverse-waterfilling solution in which we only require a given distortion threshold $D > 0$ (Theorem 1). We also implement the solution in Algorithm 1. The idea to obtain our lower bound is subsequently used to derive an upper bound on the minimum achievable total-rates (per dimension) using a sequential DPCM-based ECDQ scheme that is constrained by total-rates for a fixed finite-time horizon. For the specific rate constraint, we use the dynamic reverse-waterfilling algorithm obtained from our lower bound to allocate the rate and the MSE distortion at each time instant for the whole finite-time horizon. *This rate constraint is the fundamental difference compared to similar upper bounds derived in [16, Th. 6] and [27, Corollary 5.2] (see also [11], [28]) that restrict their transmit rates to have either fixed rates that are averaged across the time horizon or that are asymptotically averaged across the time*.

2) We obtain analogous bounds to 1) on the minimum achievable total (across time) cost-rate function of control (per-dimension) for a networked control system (NCS) with time-varying quantized LQG closed-loops operating with data-rate obtained subject to a solution of a reverse-waterfilling algorithm (Theorems 3 and 4).

Discussion of the Contributions and Additional Results: The nonasymptotic lower bound in 1) is obtained because for parallel processes, all involved matrices in the characterization of the corresponding optimization problem *commute by pairs* [30, p. 5]; thus, they are *simultaneously diagonalizable* by an orthogonal matrix [30, Th. 21.13.1] and the resulting optimization problem simplifies to one that resembles scalar-valued processes. The upper bound in 1) is obtained because we are able to employ a lattice quantizer [31] using a quantization scheme with existing performance guarantees such as the DPCM-based ECDQ scheme and using existing approximations from quantization theory for high-dimensional but possibly finite-dimensional quantizers with an MSE performance criterion (see, e.g., [32]). The nonasymptotic bounds derived in 2) are obtained using the so-called "weak separation principle" of quantized LQG control (for details, see Section IV) and well-known information theoretic inequalities. Interestingly, our lower bound in 2) also reveals the minimum allowable data rates at each time instant to ensure (mean square) stability of the plant (see e.g., [33] for the definition) using nonuniform rate-distortion allocation for the specific NCS (Remark 6). Finally, for every bound in this article, we explain how to recover known results for the steady-state solution (see Corollaries 1–4).

This article is organized as follows. In Section II, we give an overview of known results on sequential coding. In Section III, we derive nonasymptotic bounds and their corresponding per-unit time asymptotic limits for a quantized state estimation problem. In Section IV, we use the results of Section III and the weak separation principle to derive nonasymptotic bounds and their corresponding per-unit time asymptotic limits for a quantized LQG closed-loop control problem. We draw conclusions and discuss open questions in Section V.

Notation: $\mathbb{R} \triangleq (-\infty, \infty)$, $\mathbb{N}_1 \triangleq \{1, 2, \ldots\}$, and $\mathbb{N}_1^n \triangleq \{1, \ldots, n\}$, $n \in \mathbb{N}_1$, respectively. Let $\mathbb{X}$ be a finite-dimensional Euclidean space and $\mathcal{B}(\mathbb{X})$ be the Borel σ-algebra on $\mathbb{X}$. An RV X defined on some probability space $(\Omega, \mathcal{F}, \mathbf{P})$ is a map $X : \Omega \mapsto \mathbb{X}$. The probability distribution of an RV X with realization $X = x$ on $\mathbb{X}$ is denoted by $\mathbf{P}_X \equiv p(x)$. The conditional distribution of an RV Y with realization $Y = y$, given $X = x$, is denoted by $\mathbf{Q}_{Y|X} \equiv q(y|x)$. We denote the sequence of one-sided RVs by $X_{t,j} \triangleq (X_t, X_{t+1}, \ldots, X_j), t \leq j, (t, j) \in \mathbb{N}_1 \times \mathbb{N}_1$, and their values by $x_{t,j} \in \mathbb{X}_{t,j} \triangleq \times_{k=t}^{j} \mathbb{X}_k$. We denote the sequence of ordered RVs with ith spatial components by $X_{t,j}^i$, so that $X_{t,j}^i$ is a vector of dimension "i," and their values by $x_{t,j}^i \in \mathbb{X}_{t,j}^i \triangleq \times_{k=t}^{j} \mathbb{X}_k^i$, where $\mathbb{X}_k^i \triangleq (\mathbb{X}_k(1), \ldots \mathbb{X}_k(i))$. The notation $X \leftrightarrow Y \leftrightarrow Z$ denotes a Markov Chain (MC), which means that $p(x|y, z) = p(x|y)$. We denote the diagonal of a square matrix by $\mathrm{diag}(\cdot)$ and the $p \times p$ identity matrix by I_p. If $A \in \mathbb{R}^{p \times p}$, we denote by $A \succeq 0$ (respectively, $A \succ 0$) a positive semidefinite matrix (respectively positive definite matrix). We denote the determinant and trace of some square matrix $A \in \mathbb{R}^{p \times p}$ by $|A|$ and $\mathrm{trace}(A)$, respectively. We denote by $h(x)$ (resp. $h(x|y)$) the differential entropy of a distribution $p(x)$ (resp. $p(x|y)$). We denote $\mathcal{D}(P||Q)$ the relative entropy of probability distributions P and Q. We denote by $\mathbf{E}\{\cdot\}$ the expectation operator and $||\cdot||_2$ the Euclidean norm. When we say "total" distortion, "total-rate," or "total-cost," we mean with respect to time. Similarly, by referring to "average total," we mean normalized over the total time horizon.

II. Known Results on Sequential Coding

In this section, we give an overview of the sequential causal coding [3, Ch. 5], [2], [12], [13].

In the following analysis, we consider processes for a fixed time-span $t \in \mathbb{N}_1^n$, i.e., $(X_1, \ldots, X_n)$. Following [12], [13], we assume that the sequences of RVs are defined on alphabet spaces with finite cardinality. Nevertheless, these can be extended following, for instance, the techniques employed in [34] to continuous alphabet spaces as well (i.e., Gaussian processes) with MSE distortion constraints.

First, we use some definitions (with slight modifications to ease the readability of the article) from [12, §II] and [13, §I].

Definition 1: (Sequential Causal Coding) A spatial order p sequential causal code $\mathcal{C}_p$ for the (joint) vector source $(X_1^p, X_2^p, \ldots, X_n^p)$ is formally defined by a sequence of encoder and decoder pairs $(f_1^{(p)}, g_1^{(p)}), \ldots, (f_n^{(p)}, g_n^{(p)})$ such that

$$f_t^{(p)} : \mathbb{X}_{1,t}^p \times \underbrace{\{0,1\}^* \times \ldots \times \{0,1\}^*}_{t-1 \ times} \longrightarrow \{0,1\}^*$$
$$g_t^{(p)} : \underbrace{\{0,1\}^* \times \ldots \times \{0,1\}^*}_{t \ times} \longrightarrow \mathbb{Y}_t^p, \ t \in \mathbb{N}_1^n \tag{1}$$

where $\{0,1\}^*$ denotes the set of all binary sequences of finite length satisfying the property that at each time instant t, the range of $\{f_t : t \in \mathbb{N}_1^n\}$ given any $t-1$ binary sequences is an instantaneous code. The encoded and reconstructed sequences of $\{X_t^p : t \in \mathbb{N}_1^n\}$ are given by $S_t = f_t(X_{1,t}^p, S_{1,t-1})$, with $S_t \in \mathbb{S}_t \subset \{0,1\}^*$, and $Y_t^p = g_t(S_{1,t})$, respectively, with cardinality $\mathrm{card}(\mathbb{Y}_t) < \infty$ for any t. The expected rate in bits per symbol at each time instant (normalized over the spatial components) is defined as

$$r_t \triangleq \frac{\mathbf{E}|S_t|}{p}, \ t \in \mathbb{N}_1^n \tag{2}$$

where $|S_t|$ denotes the length of the binary sequence S_t.

Distortion Criterion: For each $t \in \mathbb{N}_1^n$, we consider a total (in dimension) single-letter distortion criterion. This means that the distortion between X_t^p and Y_t^p is measured by a function $d_t : \mathbb{X}_t^p \times \mathbb{Y}_t^p \longrightarrow [0, \infty)$ with maximum distortion $d_t^{\max} = \max_{x_t^p, y_t^p} d_t(x_t^p, y_t^p) < \infty$ such that

$$d_t(x_t^p, y_t^p) \triangleq \frac{1}{p} \sum_{i=1}^{p} d_t(x_t(i), y_t(i)). \tag{3}$$

The average distortion is defined as

$$\mathbf{E}\left\{d_t(X_t^p, Y_t^p)\right\} \triangleq \frac{1}{p} \sum_{i=1}^{p} \mathbf{E}\left\{d_t(X_t(i), Y_t(i))\right\}. \tag{4}$$

Definition 2: (Achievability) A rate-distortion tuple $(R_{1,n}, D_{1,n}) \triangleq (R_1, \ldots, R_n, D_1, \ldots, D_n)$ for any "n" is said to be *achievable* for a given sequential causal coding system if for all $\epsilon > 0$, there exists a sequential code $\{(f_t^{(p)}, g_t^{(p)}) : t \in \mathbb{N}_1^n\}$ such that there exists $\mathcal{P}$ for which

$$r_t \leq R_t + \epsilon$$
$$\mathbf{E}\left\{d_t(X_t^p, Y_t^p)\right\} \leq D_t + \epsilon, \ D_t \geq 0, \ \forall t \in \mathbb{N}_1^n \tag{5}$$

holds $\forall p \geq \mathcal{P}$. Moreover, let the set of all achievable rate-distortion tuples $(R_{1,n}, D_{1,n})$ be denoted by $\mathcal{R}^*$. Then, the minimum total-rate required to achieve the distortion tuple $(D_1, D_2, \ldots, D_n)$ is defined by

$$\mathcal{R}_{\text{sum}}^{\text{op}}(D_{1,n}) \triangleq \inf_{(R_{1,n}, D_{1,n}) \in \mathcal{R}^*} \sum_{t=1}^{n} R_t. \tag{6}$$

Source Model: The finite alphabet source randomly generates symbols $X^p_{1,n} = x^p_{1,n} \in \mathbb{X}^p_{1,n}$ according to the following temporally correlated joint probability mass function (PMF):

$$p(x^p_{1,n}) \triangleq \otimes_{i=1}^p p(x_1(i), \ldots, x_n(i)) \tag{7}$$

where the joint process $\{(X_1(i), \ldots, X_n(i))\}_{i=1}^p$ is identically distributed. This means that for each $i = 1, \ldots, p$, the temporally correlated joint process $(X_1(i), \ldots, X_n(i))$ is independent of every other temporally correlated joint process $(X_1(j), \ldots, X_n(j))$, such that $i \neq j$. Furthermore, each temporally correlated joint process $(X_1(i), \ldots, X_n(i))$ is spatially identically distributed.

Achievable Rate-Distortion Regions and Minimum Achievable Total Rate: Next, we characterize the achievable rate-distortion regions and the minimum achievable total rate for the source model (7) with distortion constraint (4).

The following lemma is given in [13, Theorem 5].

Lemma 1: (Achievable Rate-Distortion Region) Consider the source model (7) with the average distortion of (4). Then, the "spatially" single-letter characterization of the rate-distortion region $(R_{1,n}, D_{1,n})$ is given by

$$\mathcal{R}^{\text{IID}} = \left\{ (R_{1,n}, D_{1,n}) \middle| \exists S_{1,n-1}, Y_{1,n}, \ \{g_t(\cdot)\}_{t=1}^n \right.$$
$$\text{s.t.} \quad R_1 \geq I(X_1; S_1) \ \textit{(initial time)}$$
$$R_t \geq I(X_{1,t}; S_t | S_{1,t-1}), \ t = 2, \ldots, n-1,$$
$$R_n \geq I(X_{1,n}; Y_n | S_{1,n-1}) \ \textit{(terminal time)}$$
$$D_t \geq \mathbf{E}\left\{d_t(X_t, Y_t)\right\}, \ t \in \mathbb{N}_1^n$$
$$Y_1 = g_1(S_1), \ Y_t = g_t(S_{1,t}), \ t = 2, \ldots, n-1$$
$$S_1 \leftrightarrow (X_1) \leftrightarrow X_{2,n}$$
$$\left. S_t \leftrightarrow (X_{1,t}, S_{1,t-1}) \leftrightarrow X_{t+1,n}, \ t = 2, \ldots, n-1 \right\} \tag{8}$$

where $\{S_{1,n-1}, Y_{1,n}\}$ are the auxiliary (encoded) and reproduction RVs, respectively, taking values in some finite alphabet spaces $\{\mathbb{S}_{1,n-1}, \mathbb{Y}_{1,n}\}$, and $\{g_t(\cdot) : t \in \mathbb{N}_1^n\}$ are deterministic functions.

Remark 1: (On Lemma 1) In the characterization of Lemma 1, we exclude the spatial index because the rate and distortion regions are normalized with the total number of spatial components. Following [12], [13], Lemma 1 gives a set $\mathcal{R}^{\text{IID}}$ that is convex and closed (this can be shown by trivially generalizing the time-sharing and continuity arguments of [12, Appendix C2] to n time-steps). This in turn means that $\mathcal{R}^* = \mathcal{R}^{\text{IID}}$ (see, e.g., [13, Theorem 5]). Thus, (6) can be reformulated to the following problem:

$$\mathcal{R}^{\text{IID,op}}_{\text{sum}}(D_{1,n}) \triangleq \min_{(R_{1,n}, D_{1,n}) \in \mathcal{R}^{\text{IID}}} \sum_{t=1}^n R_t. \tag{9}$$

Next, we state a lemma (without a proof) that gives a lower bound on $\mathcal{R}^{\text{IID,op}}_{\text{sum}}(D_{1,n})$. The derivation of the proof can be found, for instance, in [3, Theorem 5.3.1, Lemma 5.4.1], [27, Theorem 4.1], [12, Corollary 1.1].

Lemma 2: (Lower Bound on (9)) For any p sufficiently large, the following lower bound holds:

$$\mathcal{R}^{\text{IID,op}}_{\text{sum}}(D_{1,n}) \geq \mathcal{R}^{\text{IID}}_{\text{sum}}(D_{1,n})$$
$$\triangleq \min_{\substack{\mathbf{E}\{d_t(X_t, Y_t)\} \leq D_t, \ t \in \mathbb{N}_1^n \\ Y_1 \leftrightarrow X_1 \leftrightarrow X_{2,n}, \\ Y_t \leftrightarrow (X_{1,t}, Y_{1,t-1}) \leftrightarrow X_{t+1,n}, \ t=2,\ldots,n-1}} I(X_{1,n}; Y_{1,n}) \tag{10}$$

where $I(X_{1,n}; Y_{1,n}) = \sum_{t=1}^n I(X_{1,t}; Y_t | Y_{1,t-1})$ is a variant of directed information [29], [35] obtained by the conditional independence constraints imposed in the constraint set of (10).

The lower bound in Lemma 2 can be found in the literature by the name nonanticipatory ϵ-entropy and sequential or nonanticipative RDF.

Remark 2: (When do we achieve the lower bound in (10)?) In [13, Theorem 4], they showed via an algorithmic approach (see also [13, Theorem 5] for an equivalent proof via a direct and converse coding theorem) that Lemma 2 is achieved with equality if the number of IID spatial components tends to infinity, i.e., $p \longrightarrow \infty$, which also means that the optimal minimizer or "test-channel" at each time instant in (10) corresponds precisely to the distribution generated by a sequential encoder, i.e., $S_t = Y_t$, for any $t \in \mathbb{N}_1^n$ (see also [12, Corollary 1.1]). In other words, the equality holds if the encoder (or quantizer for continuous alphabet sources) simulates exactly the corresponding "test-channel" distribution of (10). This claim was demonstrated via an application example for jointly Gaussian RVs and per-time MSE distortion in [12, Corollary 1.2] and also stated as a corollary referring to an "ideal" DPCM-based MSE quantizer in [12, Corollary 1.3]. In general, however, for any $p < \infty$, the equality in (10) is not achievable.

Next, we state the generalization of Lemma 2 when the constrained set is subject to an average total distortion constraint defined as $\frac{1}{n}\sum_{t=1}^n \mathbf{E}\{d_t(X_t, Y_t)\} \leq D$ with $\mathbf{E}\{d_t(X_t, Y_t)\}$ given in (4). This lemma was derived in [3, Theorem 5.3.1 and Lemma 5.4.1].

Lemma 3: (Generalization of Lemma 2) For any p sufficiently large, the following lower bound holds:

$$\mathcal{R}^{\text{IID,op}}_{\text{sum}}(D) \geq \mathcal{R}^{\text{IID}}_{\text{sum}}(D)$$
$$= \min_{\substack{\frac{1}{n}\sum_{t=1}^n \mathbf{E}\{d_t(X_t, Y_t)\} \leq D, \ t \in \mathbb{N}_1^n \\ Y_1 \leftrightarrow (X_1) \leftrightarrow X_{2,n}, \\ Y_t \leftrightarrow (X_{1,t}, Y_{1,t-1}) \leftrightarrow X_{t+1,n}, \ t \in \mathbb{N}_2^{n-1}}} I(X_{1,n}; Y_{1,n}). \tag{11}$$

Clearly, one can use the same methodology applied in [13, Ths. 4 and 5] to demonstrate that the lower bound in (11) is achieved once $p \longrightarrow \infty$ (see the discussion in Remark 2). However, we once again point out that in general, (11) is a lower bound on the minimum achievable rates achieved by causal sequential codes.

Information Structures: Next, we state a few well-known structural results related to the bounds in Lemmas 2 and 3. Specifically, if the temporally correlated joint PMF in (7) follows a finite-order Markov process, then, the description in Lemma 1 and the corresponding bounds in Lemmas 2 and 3 can be simplified following, for instance, the framework of [6], [17], [23]. For the important special case of first-order Markov

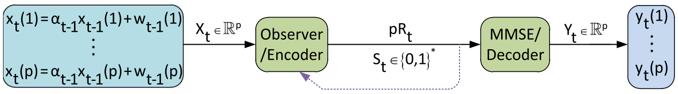


Fig. 1. Multitrack state estimation system model.

process, (8) simplifies to

$$\mathcal{R}^{\text{IID},1} = \left\{ (R_{1,n}, D_{1,n}) \middle| \exists S_{1,n-1}, Y_{1,n},\ \{g_t(\cdot)\}_{t=1}^{n} \right.$$
$$\begin{aligned} s.t.\ & R_1 \geq I(X_1; S_1) \text{ (initial time)} \\ & R_t \geq I(X_t; S_t | S_{1,t-1}),\ t = 2, \ldots, n-1 \\ & R_n \geq I(X_n; Y_n | S_{1,n-1}) \text{ (terminal time)}, \\ & D_t \geq \mathbf{E}\left\{ d_t(X_t, Y_t) \right\},\ t \in \mathbb{N}_1^n \\ & Y_1 = g_1(S_1),\ Y_t = g_t(S_{1,t}),\ t = 2, \ldots, n-1 \\ & S_1 \leftrightarrow (X_1) \leftrightarrow X_{2,n} \\ & \left. S_t \leftrightarrow (X_t, S_{1,t-1}) \leftrightarrow (X_{1,t-1}, X_{t+1,n}) \right\}. \end{aligned} \tag{12}$$

Using (12), the minimum achievable total-rate in turn can be simplified as follows:

$$\mathcal{R}_{\text{sum}}^{\text{IID,op},1}(D_{1,n}) \triangleq \min_{(R_{1,n}, D_{1,n}) \in \mathcal{R}^{\text{IID},1}} \sum_{t=1}^{n} R_t. \tag{13}$$

In addition, from (13), we can simplify (10) and (11), accordingly. In the sequel of the article, we only consider average total MSE distortion constraint to obtain our results, and for this reason, we give its simplification below

$$\begin{aligned} \mathcal{R}_{\text{sum}}^{\text{IID,op},1}(D) &\geq \mathcal{R}_{\text{sum}}^{\text{IID},1}(D) \\ &\triangleq \min_{\substack{\frac{1}{n}\sum_{t=1}^{n} \mathbf{E}\{d_t(X_t, Y_t)\} \leq D,\ t \in \mathbb{N}_1^n \\ Y_1 \leftrightarrow X_1 \leftrightarrow X_{2,n}, \\ Y_t \leftrightarrow (X_t, Y_{1,t-1}) \leftrightarrow (X_{1,t-1},\ X_{t+1,n}),\ t=2,\ldots,n-1}} I(X_{1,n}; Y_{1,n}) \end{aligned} \tag{14}$$

where $I(X_{1,n}; Y_{1,n}) = \sum_{t=1}^{n} I(X_t; Y_t | Y_{1,t-1})$.

III. Application in Quantized State Estimation

In this section, we apply the framework of Section II to a state estimation problem and obtain new results in such applications. The setup is similar to [16, Section II] where a multitrack system estimates several "parallel" Gaussian processes over a single shared communication link as illustrated in Fig. 1. However, in contrary to the result of [16, Theorem 1] which derives a dynamic forward in time recursion of a distortion-rate allocation algorithm when the rate is given at each time instant, here we derive a dynamic rate-distortion reverse-waterfilling algorithm operating forward in time for which we only consider a given distortion threshold $D > 0$.

We start with the description of the problem.

State process: Consider p-parallel time-varying Gauss–Markov processes with IID spatial components as follows:

$$x_t(i) = \alpha_{t-1} x_{t-1}(i) + w_{t-1}(i),\ i \in \mathbb{N}_1^p,\ t \in \mathbb{N}_1^n \tag{15}$$

where $x_1(i) \equiv x_1$ is given, with $x_1 \sim \mathcal{N}(0; \sigma_{x_1}^2)$; the nonrandom coefficient $\alpha_t \in \mathbb{R}$ is known at each time step t, and $\{w_t(i) \equiv w_t :\ i \in \mathbb{N}_1^p\}$, $w_t \sim \mathcal{N}(0; \sigma_{w_t}^2)$, is an independent Gaussian noise process at each t, independent of $x_1, \forall i \in \mathbb{N}_1^p$. Clearly, (15) can be compactly written as a vector or frame as follows:

$$X_t = A_{t-1} X_{t-1} + W_{t-1},\ X_1 = \text{given},\ t \in \mathbb{N}_2^n \tag{16}$$

where $A_{t-1} = \text{diag}(\alpha_{t-1}, \ldots, \alpha_{t-1}) \in \mathbb{R}^{p \times p}$, $X_t \in \mathbb{R}^p$, and the independent Gaussian noise process $W_t \in \mathbb{R}^p \sim \mathcal{N}(0; \Sigma_{W_t})$, where $\Sigma_{W_t} = \text{diag}(\sigma_{w_t}^2, \ldots, \sigma_{w_t}^2) \succ 0 \in \mathbb{R}^{p \times p}$ independent of the initial state X_1.

Observer/Encoder: At the observer, the spatially IID time-varying $\mathbb{R}^p$-valued Gauss–Markov processes are collected into a frame $X_t \in \mathbb{R}^p$ and mapped using sequential coding with encoded sequence

$$S_t = f_t(X_{1,t}, S_{1,t-1}) \tag{17}$$

where at $t = 1$, we assume $S_1 = f_1(X_1)$, and $R_t = \frac{\mathbf{E}|S_t|}{p}$ is the expected rate (per dimension) at each time instant t transmitted through the noiseless link.

MMSE Decoder: The data packet S_t is received using the following reconstructed sequence:

$$Y_t = g_t(S_{1,t}) \tag{18}$$

where at $t = 1$, we have $Y_1 = g_1(S_1)$.

Distortion: We consider the average total MSE distortion normalized over all spatial components as follows:

$$\frac{1}{n} \sum_{t=1}^{n} D_t \text{ with } D_t \triangleq \frac{1}{p} \mathbf{E}\left\{ ||X_t - Y_t||_2^2 \right\}. \tag{19}$$

Performance: The performance of the system (per dimension) for a given $D > 0$ can be cast as follows:

$$\mathcal{R}_{\text{sum}}^{\text{IID,op},1}(D) = \min_{\substack{(f_t,\ g_t):\ t=1,\ldots,n \\ \frac{1}{n}\sum_{t=1}^{n} D_t \leq D}} \sum_{t=1}^{n} R_t. \tag{20}$$

The next theorem is our first main result in this article. It derives a lower bound on the performance of Fig. 1 by means of a dynamic reverse-waterfilling algorithm.

Theorem 1: (Lower Bound on (20)) For the multitrack system in Fig. 1, the minimum achievable total-rate for any "n" and any p, however large, is $\mathcal{R}_{\text{sum}}^{\text{IID,op},1}(D) = \sum_{t=1}^{n} R_t^{\text{op}}$ with the minimum achievable rate-distortion at each time instant (per dimension) given by some $R_t^{\text{op}} \geq R_t^*$ such that

$$R_t^* = \frac{1}{2} \log_2 \left(\frac{\lambda_t}{D_t} \right) \tag{21}$$

where $\lambda_t \triangleq \alpha_{t-1}^2 D_{t-1} + \sigma_{w_{t-1}}^2$ and D_t is the distortion at each time instant evaluated based on a dynamic reverse-waterfilling algorithm operating forward in time as follows:

$$D_t \triangleq \begin{cases} \xi_t & \text{if} \quad \xi_t \leq \lambda_t \\ \lambda_t & \text{if} \quad \xi_t > \lambda_t \end{cases}, \ \forall t \tag{22}$$

with $\sum_{t=1}^n D_t = nD$, and

$$\xi_t = \begin{cases} \frac{1}{2b_t^2}\left(\sqrt{1+\frac{2b_t^2}{\theta}}-1\right), \ \forall t \in \mathbb{N}_1^{n-1} \\ \frac{1}{2\theta}, \ t=n \end{cases} \tag{23}$$

where $\theta > 0$ is the Lagrangian multiplier tuned to obtain equality $\sum_{t=1}^n D_t = nD$, $b_t^2 \triangleq \frac{\alpha_t^2}{\sigma_{w_t}^2}$, and $D \in (0,\infty)$.

Proof: See Appendix A. ■

Next we remark some technical observations on Theorem 1 and draw connections with [12, Corollary 1.2].

Remark 3: (On Theorem 1) 1) The optimization problem in the derivation of Theorem 1 suggests that $(A_t, \Sigma_{W_t}, \Delta_t, \Lambda_t)$ commute by pairs [30, p. 5] because they are all scalar matrices. This means that they are simultaneously diagonalizable by an orthogonal matrix [30, Theorem 21.13.1] (in this case, the orthogonal matrix is the identity matrix and thus can be omitted from the optimization problem). 2) Theorem 1 extends the result of [12, Corollary 1.2] which found an explicit expression of the minimum total rate $\sum_{t=1}^n R_t^*$ for $n=3$ subject to a per-time MSE distortion, to a similar problem constrained by an average total distortion that we solve using a dynamic reverse-waterfilling algorithm for any fixed finite-time horizon.

Implementation of the Dynamic Reverse-Waterfilling: A way to implement the dynamic reverse-waterfilling algorithm of Theorem 1 is proposed in [36, Algorithm 1]. A different algorithm using the *bisection method* (for details, see, e.g., [37, Chapter 2.1]) is proposed in Algorithm 1. This method guarantees linear convergence. On the other hand, [36, Algorithm 1] requires a specific proportionality gain factor $\gamma \in (0,1]$ chosen appropriately at each time instant that affects the rate of convergence and it does not guarantee global convergence of the algorithm.

Algorithm 1: Dynamic Reverse-Waterfilling Algorithm.

Initialize: number of time-steps n; distortion level D; error tolerance ϵ; nominal minimum and maximum value of θ, i.e., $\theta^{\min}=0$ and $\theta^{\max}=\frac{1}{2D}$; initial variance $\lambda_1=\sigma_{x_1}^2$, values a_t and $\sigma_{w_t}^2$ of (15).
Set $\theta = 1/2D$; flag $=0$.
while flag $=0$ **do**
  for $t=1:n$ **do**
    Compute ξ_t according to (23).
    Compute D_t according to (22).
    if $t<n$ **then**
      Compute λ_{t+1} according to $\lambda_{t+1} \triangleq \alpha_t^2 D_t + \sigma_{w_t}^2$.
    end if
  end for
  if $\frac{1}{n}\sum D_t - D \geq \epsilon$ **then**
    Set $\theta^{\min}=\theta$.
  else
    Set $\theta^{\max}=\theta$.
  end if
  if $\theta^{\max}-\theta^{\min} \geq \frac{\epsilon}{n}$ **then**
    Compute $\theta = \frac{(\theta^{\min}+\theta^{\max})}{2}$.
  else
    flag $\leftarrow 1$
  end if
end while
Output: $\{D_t : t \in \mathbb{N}_1^n\}$, $\{\lambda_t : t \in \mathbb{N}_1^n\}$, for a given distortion level D.

A. Steady-State Solution of Theorem 1

In this subsection, we briefly discuss the steady-state case of the lower bound obtained in Theorem 1. To do it, first, we restrict the state process of our setup to be time-invariant, which means that in (15), the coefficients $\alpha_{t-1} \equiv \alpha, \forall t$ and $w_t \sim \mathcal{N}(0;\sigma_w^2), \forall t$, or similarly, in (16), the matrix $A_{t-1} \equiv A = \text{diag}(\alpha,\ldots,\alpha), \forall t$ and $W_t \sim \mathcal{N}(0;\Sigma_W), \forall t$, where $\Sigma_W = \text{diag}(\sigma_w^2,\ldots,\sigma_w^2) \succ 0$. We also define the steady-state average total rate and distortion as follows:

$$R_\infty \triangleq \limsup_{n\longrightarrow\infty} \frac{1}{n}\sum_{t=1}^n R_t, \quad D_\infty \triangleq \limsup_{n\longrightarrow\infty}\frac{1}{n}\sum_{t=1}^n D_t. \tag{24}$$

Steady-State Performance: For p-parallel time-invariant Gauss–Markov processes (per dimension), the minimum achievable steady-state performance of the multitrack system of Fig. 1 can be cast as follows:

$$\mathcal{R}_{\text{sum},ss}^{\text{IID,op},1}(D) = \min_{\substack{(f_t,\ g_t):\ t=1,\ldots,\infty \\ D_\infty \leq D}} R_\infty. \tag{25}$$

Under the previous assumptions, one can obtain a lower bound on the minimum achievable steady-state total rate subject to steady-state total distortion constraint. This result is equivalent to having the minimum achievable steady-state total rate subject to a fixed (uniform) distortion budget, i.e., $D_t = D, \forall t$. In what follows, we state this result without a proof because it follows using similar steps to [16, Corollary 2] or [24, Theorem 9].

Corollary 1: (Lower Bound on (25)) The minimum achievable steady-state performance of (25), under a steady-state total distortion constraint $D_\infty \leq D$ for any p, however larger, is bounded from below by $\mathcal{R}_{\text{sum},ss}^{\text{IID,op},1}(D) \geq R_\infty^*$, such that

$$R_\infty^* = \frac{1}{2}\log_2\left(\alpha^2 + \frac{\sigma_w^2}{D}\right) \tag{26}$$

where $R_\infty^* \triangleq \lim_{n\longrightarrow\infty}\frac{1}{n}\sum_{t=1}^n R_t^*$. Consequently, assuming $D_t = D, \forall t$, achieves (26) as $n \longrightarrow \infty$.

Remark 4: (On Corollary 1) 1) The steady-state lower bound of Corollary 1 corresponds precisely to the solution of the time-invariant scalar-valued Gauss–Markov processes with per-time MSE distortion constraint derived in [20, (14)] and to the solution of stationary Gauss-Markov processes with MSE distortion constraint derived in [38, Theorem 3], [18, (1.43)]. 2) In Fig. 2 , we illustrate the behavior of the average total rate obtained in Theorem 1 versus (vs.) the average total rate obtained using uniform distortion allocation vs. the steady-state lower bound of (26), as a function of t. We observe that although the three lines seem to meet really fast, they do not coincide. In fact, for

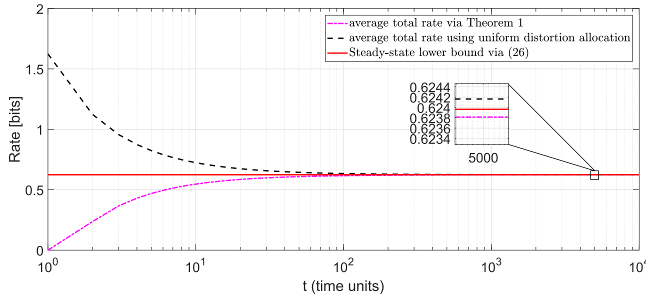


Fig. 2. Comparison of the average total rate of Theorem 1 vs. average total rate with uniform distortion vs. the lower bound of Corollary 1, as a function of t. For this simulation, we considered $D = 2$ and $(\alpha, \sigma_w^2) = (1.5, 0.5)$.

a reasonable precision error (i.e., of order 10^{-9}), the lines will converge in the asymptotic limit as Corollary 1 suggests. The starting point of the plot obtained using Theorem 1 depends on the initial value of λ_1.

B. Upper Bounds on the Minimum Achievable Total-Rate

In this section, we employ a sequential causal DPCM-based scheme using pre-/post-filtered ECDQ (for details on this scheme, see, e.g., [31, Chapter 5]) that ensures standard performance guarantees (achievable upper bounds) on the minimum achievable sum-rate $\mathcal{R}_{\text{sum}}^{\text{IID,op},1}(D) = \sum_{t=1}^{n} R_t^{\text{op}}$ of the multitrack setup of Fig. 1. The reason for the choice of this quantization scheme is twofold. First, it can be implemented in practice and, second, it allows to find analytical achievable bounds and approximations on finite-dimensional quantizers which generate near-Gaussian quantization noise and Gaussian quantization noise for infinite dimensional quantizers [39].

We first describe the sequential causal DPCM scheme using an MMSE quantizer for parallel time-varying Gauss–Markov processes. Then, we bound the rate performance of such scheme using ECDQ and vector quantization followed by memoryless entropy coding. This can be seen as a generalization of [12, Corollary 1.2] to any finite time when the rate is nonuniformly allocated at each time instant.

DPCM Scheme: At each time instant t, the *encoder or innovations' encoder* performs the linear operation

$$\widehat{X}_t = X_t - A_{t-1}Y_{t-1} \tag{27}$$

where at $t=1$, we have $\widehat{X}_1 = X_1$ and also $Y_{t-1} \triangleq \mathbf{E}\{X_{t-1}|S_{1,t-1}\}$, i.e., an estimate of X_{t-1} given the previous quantized symbols $S_{1,t-1}$. Then, by means of a $\mathbb{R}^p$-valued MMSE quantizer that operates at a rate (per dimension) R_t, we generate the quantized reconstruction $\widehat{Y}_t$ of the residual source $\widehat{X}_t$ denoted by $\widehat{Y}_t = Y_t - A_{t-1}Y_{t-1}$. Afterwards, we send S_t over the channel (the corresponding data packet to $\widehat{Y}_t$). At the *decoder*, we receive S_t and recover the quantized symbol $\widehat{Y}_t$ of $\widehat{X}_t$. Then, we generate the estimate Y_t using the linear operation

$$Y_t = \widehat{Y}_t + A_{t-1}Y_{t-1}. \tag{28}$$

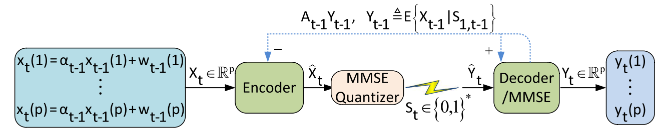


Fig. 3. DPCM of parallel processes.

Combining (27) and (28), we obtain

$$X_t - Y_t = \widehat{X}_t - \widehat{Y}_t. \tag{29}$$

MSE Performance: From (29), we see that the error between X_t and Y_t is equal to the quantization error introduced by $\widehat{X}_t$ and $\widehat{Y}_t$. This also means that the MSE distortion (per dimension) at each instant of time satisfies

$$D_t = \frac{1}{p}\mathbf{E}\{||X_t - Y_t||_2^2\} = \frac{1}{p}\mathbf{E}\{||\widehat{X}_t - \widehat{Y}_t||_2^2\}. \tag{30}$$

A pictorial view of the DPCM scheme is given in Fig. 3.

The following theorem is another main result.

Theorem 2: (Upper Bound to $R_{\text{sum}}^{\text{IID,op},1}(D)$) Suppose that in (20), we apply a sequential causal DPCM-based ECDQ with a lattice quantizer. Then, the minimum achievable total rate $\mathcal{R}_{\text{sum}}^{\text{IID,op},1}(D) = \sum_{t=1}^{n} R_t^{\text{op}}$, where at each time instant R_t^{op} is upper bounded as follows:

$$R_t^{\text{op}} \leq R_t^* + \frac{1}{2}\log_2(2\pi e G_p) + \frac{1}{p}, \ \forall t, (bits/dimension) \tag{31}$$

where R_t^* is obtained from Theorem 1, $\frac{1}{2}\log_2(2\pi e G_p)$ is the divergence of the quantization noise from Gaussianity; G_p is the dimensionless normalized second moment of the lattice [31, Definition 3.2.2]; and $\frac{1}{p}$ is the additional cost due to having prefix-free (instantaneous) coding.

Proof: See Appendix B.

Next, we remark some technical comments on Theorem 2. ■

Remark 5: (On Theorem 2) 1) Theorem 2 allows a nonuniform rate at each time instant for a finite-time horizon while it achieves the MMSE distortion at each time step t. This is because our DPCM-based ECDQ scheme makes use of the dynamic reverse-waterfilling algorithm of Theorem 1. This general rate-constraint is the new input of our bound compared to similar existing bounds in the literature (see, e.g., [16, Theorem 6, Remark 16], [27, Corollary 5.2], [11, Theorem 5]) that assume *fixed (uniform) rates averaged across the time*. 2) Recently, in [16], [24], it is pointed out that for discrete-time processes, one can assume in the ECDQ coding scheme the clocks of the entropy encoder and the entropy decoder to be synchronized, thus, eliminating the additional rate-loss due to prefix-free coding, i.e., $\frac{1}{p}$ in (31) can be removed.

Steady-State Performance: If we restrict the system model to be time-invariant (per dimension) similar to §III-A, we can obtain the following upper bound on (25).

Corollary 2: (Upper Bound on (25)) Suppose that in (20), we apply a sequential causal DPCM-based ECDQ with a lattice quantizer assuming the system is time-invariant and $D_t = D$,

$\forall t$. Then, $\mathcal{R}^{\text{IID,op},1}_{\text{sum},ss}(D) = R^{\text{op}}_\infty$ is upper bounded as follows:

$$R^{\text{op}}_\infty \leq R^*_\infty + \frac{1}{2}\log_2(2\pi e G_p) + \frac{1}{p} \text{ (bits/dimension)} \tag{32}$$

where R^*_∞ is given by (26). ∎

Proof: This follows from Theorem 2 and Corollary 1. ∎

We note that Corollary 2 is a known result derived in several papers in the past, such as those discussed in Remark 5, 1).

Computational Aspects of Theorem 2 for High-Dimensional Systems: Unfortunately, finding G_p in (31) for good high-dimensional quantizers of possibly finite dimension is currently an open problem (although it can be approximated for any dimension using, for example, product lattices [32]). For this reason, next we bring to spotlight some existing computable bounds to the achievable upper bound of Theorem 2 for any high-dimensional lattice quantizer. These bounds were derived as a consequence of a main result by Zador (see, e.g., [32]), namely, it is possible to reduce the MSE distortion normalized per dimension using higher dimensional quantizers. Toward this end, Zador introduced a lower bound on G_p using the dimensionless normalized second moment of a p-dimensional sphere, hereinafter denoted by $G(S_p)$, for which it holds that

$$G(S_p) = \frac{1}{(p+2)\pi}\Gamma\left(\frac{p}{2}+1\right)^{\frac{2}{p}} \tag{33}$$

where $\Gamma(\cdot)$ is the gamma function. Moreover, G_p and $G(S_p)$ are connected via the following inequalities:

$$\frac{1}{2\pi e} \overset{(a)}{\leq} G(S_p) \overset{(b)}{\leq} G_p \overset{(c)}{\leq} \frac{1}{12} \tag{34}$$

where $(a), (b)$ holds with equality for $p \longrightarrow \infty$; (c) holds with equality if $p = 1$. Note that in [32, (82)], there is also an upper bound on G_p due to Zador, i.e.,

$$G_p \leq \frac{1}{p\pi}\Gamma\left(\frac{p}{2}+1\right)^{\frac{2}{p}}\Gamma\left(1+\frac{2}{p}\right). \tag{35}$$

In Fig. 4, we illustrate two plots using the bounds derived in Theorems 1 and 2 for two different scenarios. In Fig. 4(a), we choose $t = \{1, \ldots, 20\}$, $a_t \in (0, 1.5)$, $\sigma^2_{w_t} = 1$, and $D = 1$, to illustrate the gap between the nonuniform rate-distortion allocation obtained via the lower bound (21) and the upper bound (31) when the latter is approximated with the best known quantizer up to 24 dimensions, i.e., Leech lattice quantizer (for details, see, e.g., [32, Table 2.3]). For this experiment, the gap between the two bounds is approximately 0.126 bits/dimension. In Fig. 4(b), we assume the same values for $(a_t, \sigma^2_{w_t}, D)$, whereas the quantization is performed for 500 dimensions. We observe that the achievable bounds obtained via (33) and (35) are quite tight (they have a gap of approximately 0.0014 bits/dimension) whereas the gap between the lower bound (21) with the achievable upper bound (31) approximated by (33) is 0.0097 bits/dimension, and the one approximated by (35) is approximately 0.011 bits/dimension. Compared to the first experiment where $p = 24$, the gap between the bounds on the minimum achievable rate R^{op}_t is considerably decreased because we increase the number of dimensions in the system. Moreover, when the number of dimensions in the system increases, the gap between (21) and the

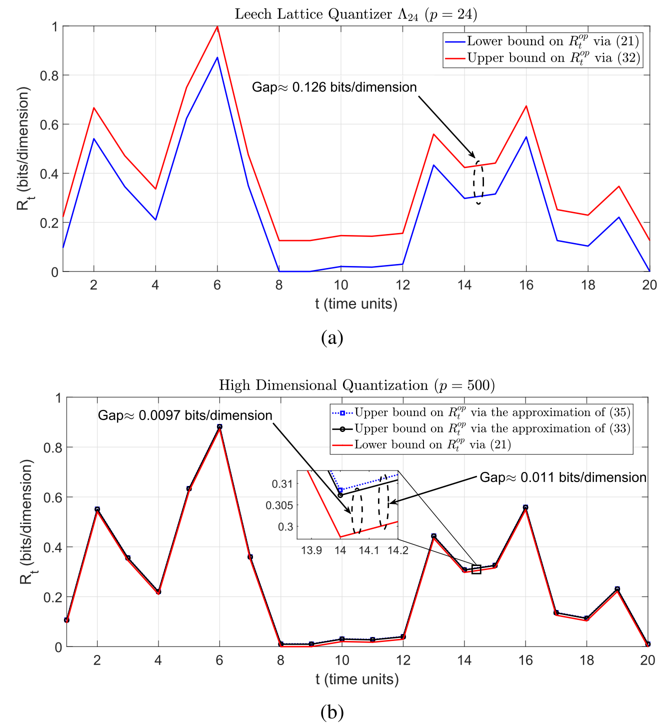


Fig. 4. Bounds on the minimum achievable total rate.

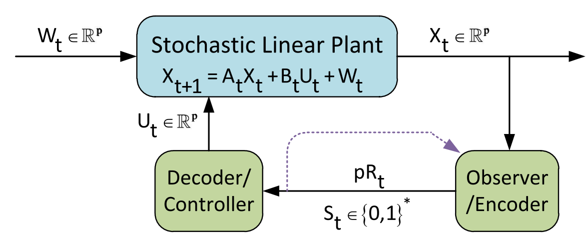


Fig. 5. Closed-loop control system model.

high-dimensional approximations of (31) will become arbitrary small. The two bounds will coincide as $p \longrightarrow \infty$, because then, the gap of coding noise from Gaussianity goes to zero (see, e.g., [40], [39, Lemma 1]), which implies that (33) is equal to (35) (see, e.g.,[32, (83)]).

IV. Application in NCSs

In this section, we use sequential coding in the NCS setup of Fig. 5 by applying the results obtained in Section III. We first describe each component of Fig. 5.

Plant: Consider p parallel time-varying controlled Gauss–Markov processes as follows:

$$x_{t+1}(i) = \alpha_t x_t(i) + \beta_t u_t(i) + w_t(i),\ i \in \mathbb{N}^p_1,\ t \in \mathbb{N}^n_1 \tag{36}$$

where $x_1(i) \equiv x_1$ is given with $x_1 \sim \mathcal{N}(0; \sigma^2_{x_1})$, $\forall i$; the nonrandom coefficients $(\alpha_t, \beta_t) \in \mathbb{R}$ are known to the system with $(\alpha_t, \beta_t) \neq 0, \forall t$; $\{u_t(i) : i \in \mathbb{N}^p_1\}$ is the controlled process with $u_t(i) \neq u_t(\ell)$, for any $(i, \ell) \in \mathbb{N}^p_1$; $\{w_t(i) \equiv w_t : i \in \mathbb{N}^p_1\}$ is an independent Gaussian noise process such that $w_t \sim \mathcal{N}(0; \sigma^2_{w_t})$,

$\sigma^2_{w_t} > 0$, independent of x_1, $\forall i$. Again, similar to Section III, (36) can be compactly written as

$$X_{t+1} = A_t X_t + B_t U_t + W_t, X_1 = \text{given}, t \in \mathbb{N}_1^n \tag{37}$$

where $A_t = \text{diag}(\alpha_t, \ldots, \alpha_t) \in \mathbb{R}^{p\times p}$, $B_t = \text{diag}(\beta_t, \ldots, \beta_t) \in \mathbb{R}^{p\times p}$, $U_t \in \mathbb{R}^p$, $W_t \in \mathbb{R}^p \sim \mathcal{N}(0; \Sigma_{W_t})$, $\Sigma_{W_t} = \text{diag}(\sigma^2_{w_t}, \ldots, \sigma^2_{w_t}) \succ 0$ is an independent Gaussian noise process independent of X_1. In this setup, the plant is fully observable for the observer that acts as an encoder but not for the controller due to the quantization noise (coding noise).

Observer/Encoder: At the encoder, the controlled process is collected into a frame $X_t \in \mathbb{R}^p$ from the plant and encoded as follows:

$$S_t = f_t(X_{1,t}, S_{1,t-1}) \tag{38}$$

where at $t = 1$, we have $S_1 = f_1(X_1)$, and $R_t = \frac{\mathbf{E}|S_t|}{p}$ is the rate at each time instant t available for transmission via the noiseless channel. In the design of Fig. 5, the channel is noiseless and the controller/decoder are deterministic mappings; thus, the observer/encoder implicitly has access to past control signals $U_{1,t-1} \in \mathbb{U}_{1,t-1}$.

Decoder/Controller: The data packet S_t is received by the controller using the following reconstructed sequence:

$$U_t = g_t(S_{1,t}). \tag{39}$$

Following (39), when the sequence $S_{1,t}$ is available at the decoder/controller, all past control signals $U_{1,t-1}$ are completely specified.

Quadratic Cost: The cost of control (per dimension) is defined as

$$\text{LQG}_{1,n} \triangleq \frac{1}{p}\mathbf{E}\left\{\sum_{t=1}^{n-1}\left(X_t^{\mathrm{T}}\widetilde{Q}_t X_t + U_t^{\mathrm{T}}\widetilde{N}_t U_t\right) + X_n^{\mathrm{T}}\widetilde{Q}_n X_n\right\} \tag{40}$$

where $\widetilde{Q}_t = \text{diag}(Q_t \ldots, Q_t) \succeq 0$, $\widetilde{Q}_t \in \mathbb{R}^{p\times p}$ and $\widetilde{N}_t = \text{diag}(N_t, \ldots, N_t) \succ 0$, $\widetilde{N}_t \in \mathbb{R}^{p\times p}$, are designing parameters that penalize the state variables or the control signals.

Performance: The performance of Fig. 5 (per dimension) can be cast to a finite-time horizon-quantized LQG control problem subject to the joint design of quantizer/controller as follows:

$$\Gamma_{\text{sum}}^{\text{IID,op}}(R) = \min_{\substack{(f_t,\, g_t):\, t=1,\ldots,n\\ \frac{1}{n}\sum_{t=1}^n R_t \le R}} \text{LQG}_{1,n}. \tag{41}$$

Iterative Encoder/Controller Design: In general, as (41) suggests, the optimal performance of the system in Fig. 5 is achieved when the encoder/controller pair is designed jointly. This is a quite challenging task, especially when the channel is noisy because information structure is non-nested in such cases (for details, see, e.g., [41]). There are examples, however, where the separation principle applies and the task comes much easier. More precisely, the so-called certainty equivalent controller remains optimal if the estimation errors are independent of previous control commands (i.e., dual effect is absent) [42]. In our case, the optimal control strategy will be a certainty equivalence controller *if we assume a fixed sequence of encoders* $\{f_t^*:\ t \in \mathbb{N}_1^n\}$ *and the corresponding quantizer follows a predictive quantizer policy (similar to the* DPCM*-based* ECDQ *scheme proposed in Section III-B), i.e., at each time instant, it subtracts the effect of the previous control signals at the encoder and adds them at the decoder* (see, e.g., [43, Proposition 3], [44], [45, Section III]). Moreover, the separation principle will also be optimal if we consider an MMSE estimate of the state (similar to what we have established in Section III), and an encoder that minimizes a distortion for state estimation at the controller. The resulting separation principle is termed "weak separation principle" [44] as it relies on the fixed (given) quantization policies. This is different from the well-known full separation principle in the classical LQG stochastic control problem [46] where the problem separates naturally into a state estimator and a state feedback controller without any loss of optimality. The previous analysis is described by a modified version of (41) as follows:

$$\Gamma_{\text{sum}}^{\text{IID,op}}(R) \le \Gamma_{\text{sum}}^{\text{IID,op},ws} = \min_{\substack{(f_t^*,\, g_t):\, t=1,\ldots,n\\ \frac{1}{n}\sum_{t=1}^n R_t \le R}} \text{LQG}_{1,n}. \tag{42}$$

Next, we state the solution of (42) in the form of a lemma. The derivation of the proof can be found in [20], [44], [45].

Lemma 4: (Weak Separation Principle for Fig. 5) The optimal controller that minimizes (41) is given by

$$U_t = -L_t \mathbf{E}\left\{X_t | S_{1,t}\right\} \tag{43}$$

where $\mathbf{E}\{X_t|S_{1,t}\}$ are the fixed quantized state estimates obtained from the estimation problem in Section III; $\widetilde{L}_t = \text{diag}(L_t, \ldots, L_t) \in \mathbb{R}^p$ is the optimal LQG control (feedback) gain obtained as follows:

$$\tilde{L}_t = \left(B_t^2 \widetilde{K}_{t+1} + \widetilde{N}_t\right)^{-1} B_t \widetilde{K}_{t+1} A_t \tag{44}$$

and $\widetilde{K}_t = \text{diag}(K_t, \ldots, K_t) \succeq 0$ is obtained using the backward recursions

$$\widetilde{K}_t = A_t^2\left(\widetilde{K}_{t+1} - \widetilde{K}_{t+1}B_t^2(B_t^2\widetilde{K}_{t+1} + \widetilde{N}_t)^{-1}\widetilde{K}_{t+1}\right) + \widetilde{Q}_t \tag{45}$$

with $\widetilde{K}_{n+1} = 0$. Moreover, this controller achieves a minimum linear quadratic cost of

$$\Gamma_{\text{sum}}^{\text{IID,op},ws} = \frac{1}{p}\sum_{t=1}^{n}\Big\{\text{trace}(\Sigma_{W_t}\widetilde{K}_t) + \text{trace}(A_t B_t \widetilde{L}_t \widetilde{K}_{t+1}\mathbf{E}\{||X_t - Y_t||_2^2\})\Big\} \tag{46}$$

where $\mathbf{E}\{||X_t - Y_t||_2^2\}$ is the MMSE distortion obtained using any quantization (coding) in the control/estimation system.

Before we prove one main result of the article, we define the instantaneous cost of control as follows:

$$\text{LQG}_t^{\text{op}} \triangleq \frac{1}{p}\Big\{\text{trace}(\Sigma_{W_t}\widetilde{K}_t) + \text{trace}(A_t B_t \widetilde{L}_t \widetilde{K}_{t+1}\mathbf{E}\{||X_t - Y_t||_2^2\})\Big\},\ t \in \mathbb{N}_1^n. \tag{47}$$

Next, we use Lemma 4 to derive a lower bound on (42).

Theorem 3: (Lower bound on (42)) For fixed coding policies, the minimum total-cost of control (per dimension) of (42), for any "n" and any p, however large, is $\Gamma_{\text{sum}}^{\text{IID,op},ws} = \sum_{t=1}^n \text{LQG}_t^{\text{op}}$, with $\text{LQG}_t^{\text{op}} \geq \text{LQG}_t^*$ such that

$$\text{LQG}_t^* = \sigma_{w_t}^2 K_t + \alpha_t\beta_t L_t K_{t+1} D(R_t^*) \tag{48}$$

where $D(R_t^*)$ is given by

$$D(R_t^*) \triangleq \begin{cases} \frac{\sigma_{w_t}^2}{2^{2R_t^*}-\alpha_t^2}, & \forall t \in \mathbb{N}_1^{n-1} \\ 2^{-2R_n^*}, & for\ t = n \end{cases} \tag{49}$$

with the pair $(D(R_t^*), R_t^*)$ given by (21)–(23).

Proof: See Appendix C. ■

The following remark on Theorem 3 reveals a new major result related to the absolute minimum rates at each instant of time for a mean square stabilizable system.

Remark 6: (On Theorem 3) The expression of the lower bound in Theorem 3 can be reformulated for any n, and any p, to the equivalent expression of the total rate-cost function, denoted hereinafter by $\sum_{t=1}^n R(\text{LQG}_t^*)$, as follows:

$$R(\text{LQG}_t^*) = \frac{1}{2}\log_2\left(\alpha_t^2 + \frac{\alpha_t\beta_t L_t K_{t+1}\sigma_{w_t}^2}{\text{LQG}_t^* - \sigma_{w_t}^2 K_t}\right),\ t \in \mathbb{N}_1^{n-1} \tag{50}$$

with $R(\text{LQG}_n^*) \equiv R_n^*$ as it is independent of LQG_n^*. Interestingly, one can observe that by substituting in (50) the per-dimension version of (44), we obtain

$$R(\text{LQG}_t^*) = \frac{1}{2}\log_2\left(\alpha_t^2\left(1 + \frac{\frac{\beta_t^2\ K_{t+1}^2\sigma_{w_t}^2}{\beta_t^2\ K_{t+1}^2 + N_t}}{\text{LQG}_t^* - \sigma_{w_t}^2 K_t}\right)\right) \tag{51}$$

$$= \frac{1}{2}\left[\log_2(\alpha_t^2) + \log_2\left(1 + \frac{\frac{\beta_t^2\ K_{t+1}^2\sigma_{w_t}^2}{\beta_t^2\ K_{t+1}^2 + N_t}}{\text{LQG}_t^* - \sigma_{w_t}^2 K_t}\right)\right]. \tag{52}$$

The bound in (52) extends the result of [24, (16)] from uniform (fixed) data rates to nonuniform rates at each instant of time because the rate-cost function is obtained using an allocation of LQG_t^* due to the nonuniform rate allocation of the quantized state estimation problem of Theorem 1. Additionally, the expression in (52) reveals an interesting observation regarding the absolute minimum data rates for mean square stability of the plant (per dimension), i.e., $\sup_t \mathbf{E}\{(x_t)^2\} < \infty$ (see, e.g., [33, (25)] for the definition) for any fixed finite-time horizon. In particular, (52) suggests that for unstable time-varying plants with arbitrary disturbances modeled as in (37), and provided that at each time instant the cost of control (per dimension) is with communication constraints, i.e., $\text{LQG}_t^* > \sigma_{w_t}^2 K_t$ (the derivation without communication constraints is well known as the separation principle holds without a loss and $\text{LQG}_t^* = \sigma_{w_t}^2 K_t,\ \forall t$ [46]), the minimum possible rates at each time instant t, namely, $R(\text{LQG}_t^*)$, cannot be lower than $\log_2|\alpha_t|$, when $|\alpha_t| > 1$. This result extends known observations obtained for time-invariant plants (see, e.g., [24, Remark 1]) to parallel and (possibly unbounded) time-varying plants for any fixed finite-time horizon.

Next, we use Theorem 2 to find an upper bound on $\Gamma_{\text{sum}}^{\text{IID,op},ws}$.

Theorem 4: (Upper Bound on (42)) Suppose that in the system of Fig. 5, the fixed coding policies are obtained using the predictive coding scheme via sequential causal DPCM-based ECDQ coding scheme with an $\mathbb{R}^p$-valued lattice quantizer described in Theorem 2. Then, $\Gamma_{\text{sum}}^{\text{IID,op},ws} = \sum_{t=1}^n \text{LQG}_t^{\text{op}}$ for any n, and any p, with the instantaneous cost of control $\{\text{LQG}_t : t \in \mathbb{N}_1^{n-1}\}$ (per dimension) to be upper bounded as follows:

$$\text{LQG}_t^{\text{op}} \leq \sigma_{w_t}^2 K_t + \alpha_t\beta_t L_t K_{t+1}\frac{4^{\frac{1}{p}}(2\pi e G_p)\sigma_{w_t}^2}{2^{2R_t^{\text{op}}} - 4^{\frac{1}{p}}(2\pi e G_p)\alpha_t^2} \tag{53}$$

whereas, at $t = n$, $\text{LQG}_n^{\text{op}} = \sigma_{w_n}^2 K_n$ and R_t^{op} is bounded above as in (31).

Proof: See Appendix D. ■

Remark 7: (On Theorem 4) For infinitely large spatial components, i.e., $p \longrightarrow \infty$, the upper bound in (53) approaches the lower bound in Theorem 3 because $G_\infty \longrightarrow \frac{1}{2\pi e}$ (see, e.g, [39, Lemma 1]). Moreover, one can easily obtain the equivalent inverse problem of the total rate-cost function for the upper bound in (53) similar to Remark 6.

Next, we note the main technical difference of the new results obtained in Theorems 3 and 4 compared to existing results in the literature.

Remark 8: (Connections to Existing Works) 1) Our bounds on LQG cost extend similar bounds derived in [16, Ths. 7 and 8] to nonuniform rate constraints for any fixed finite-time horizon. Such constraints require the use of the dynamic reverse-waterfilling optimization algorithm derived in Theorem 1. In contrast, the uniform rate constraint assumed in [16, Ths. 7 and 8] does not require a similar optimization technique because at each instant of time, the transmit rate is the same. Another structural difference compared to [16, Ths. 7 and 8] is that in our bound, we decouple the dependency of D_{t-1} at each time instant and that is why we are able to obtain the major result of Remark 6. 2) Clearly, our results extend the steady-state bounds on LQG cost obtained in [25], [27], [28] to nonasymptotic bounds constrained by nonuniform rates for any fixed finite-time horizon.

Steady-State Solution of Theorems 3 and 4: Next, we discuss the steady-state case of the bounds derived in Theorems 3 and 4 using the following assumptions: (A1) restricts the controlled process (37) to be time-invariant, which means that $A_t \equiv A = \text{diag}(\alpha, \ldots, \alpha) \in \mathbb{R}^{p\times p},\ B_t \equiv B = \text{diag}(\beta, \ldots, \beta) \in \mathbb{R}^{p\times p},\ W_t \in \mathbb{R}^p \sim \mathcal{N}(0; \Sigma_W),\ \Sigma_W = \text{diag}(\sigma_w^2, \ldots, \sigma_w^2) \succ 0,\ \forall t$; (A2) restricts the design parameters that penalize the control cost (40) to also be time-invariant, i.e., $\widetilde{Q}_t \equiv \text{diag}(Q, \ldots, Q),\ \widetilde{N}_t \equiv \text{diag}(N, \ldots, N)$; (A3) fixes $D_t \equiv D,\ \forall t$. We denote the steady-state value of the total cost of control (per dimension) as follows:

$$\text{LQG}_\infty = \limsup_{n\longrightarrow\infty}\frac{1}{n}\sum_{t=1}^n \text{LQG}_t. \tag{54}$$

Steady-State Performance: The minimum achievable steady-state performance (per dimension) of the quantized LQG control problem of Fig. 5 under the weak separation principle can be cast as follows:

$$\Gamma^{\mathrm{IID,op},ws}_{\mathrm{sum},ss} = \min_{\substack{(f^*_t,\ g_t):\ t=1,\ldots,\infty\\ R_\infty \le R}} \mathrm{LQG}_\infty. \tag{55}$$

In the next two corollaries, we state the lower and upper bounds on (55). These bounds stem from the Assumptions (A1)–(A3) and Corollaries 1 and 2.

Corollary 3: (Lower Bound on (55)) The minimum achievable steady-state performance of (55), under Assumptions (A1)–(A3), for any p, is such that $\Gamma^{\mathrm{IID,op},ws}_{\mathrm{sum},ss} \ge \mathrm{LQG}^*_\infty$, where

$$\mathrm{LQG}^*_\infty = \sigma^2_w K_\infty + \alpha\beta L_\infty K_\infty \frac{\sigma^2_w}{2^{2R^*_\infty} - \alpha^2} \tag{56}$$

where $\mathrm{LQG}^*_\infty \triangleq \lim_{n\longrightarrow\infty} \frac{1}{n}\sum_{t=1}^n \mathrm{LQG}^*_t$

$$L_\infty = \frac{\alpha\beta K_\infty}{\beta^2 K_\infty + N} \tag{57}$$

and K_∞ is the positive solution of the quadratic equation

$$\beta^2 K_\infty + \left((1-\alpha^2)N - \beta^2 Q\right) K_\infty - QN = 0 \tag{58}$$

given by the formula

$$K_\infty = \frac{1}{2\beta^2}\left(\sqrt{\bar{f}^2 + 4\beta^2 QN} - \bar{f}\right) \tag{59}$$

with $\bar{f} = (1-\alpha^2)N - \beta^2 Q$.

Corollary 4: (Upper Bound on (55)) The minimum achievable steady-state performance of (55), under Assumptions (A1)–(A3), for any p, is upper bounded as follows:

$$\Gamma^{\mathrm{IID,op},ws}_{\mathrm{sum},ss} \le \sigma^2_w K_\infty + \alpha\beta L_\infty K_\infty \frac{4^{\frac{1}{p}}(2\pi e G_p)\sigma^2_w}{2^{2R^{\mathrm{op}}_\infty} - 4^{\frac{1}{p}}(2\pi e G_p)\alpha^2}$$

where R^{op}_∞ is upper bounded by (32) and K_∞, L_∞ are given by (59) and (57), respectively.

We omit the derivation of the proofs for Corollaries 3 and 4 because they follow using similar arguments to previous papers (see, e.g., [20, Section V], [47, Section 6] for the lower bound and [16], [27], [28] for the upper bound). The upper bound is also similar to the one obtained in [25] albeit their space-filling term is obtained differently.

V. Epilogue and Open Questions

We revisited the sequential coding of correlated sources with independent spatial components to use it in the derivation of nonasymptotic, finite-dimensional lower and upper bounds for two application examples in stochastic systems. Our application examples included a parallel time-varying quantized state-estimation problem subject to a total MSE distortion constraint and a parallel time-varying quantized LQG closed-loop control system with linear quadratic cost. For the latter, its lower bound revealed the absolute minimum rates for mean square stability of the plant at each time instant under nonuniform rates for any fixed finite-time horizon.

There are various open questions that stem from the results of this article. For instance, it would be interesting to explore strong structural properties on matrices $(A_t, \Sigma_{W_t}, \Delta_t, \Lambda_t)$ in the derivation of Theorem 1, which will allow its extension to more general cases of time-varying multivariate Gauss–Markov processes. Another question not addressed in this article is whether the nonasymptotic lower bounds derived in Theorems 1 and 3 can be extended to linear Markov models driven by additive independent non-Gaussian noise processes. Finally, of interest is the extension of our setup to take into account noiseless communication links prone to packet drops.

Appendix A
Proof of Theorem 1

Using (14), we obtain

$$\mathcal{R}^{\mathrm{IID,op},1}_{\mathrm{sum}}(D) \ge \mathcal{R}^{\mathrm{IID},1}_{\mathrm{sum}}(D)$$

$$= \min_{\substack{\frac{1}{n}\frac{1}{p}\sum_{t=1}^n \mathbf{E}\{\|X_t - Y_t\|^2_2\} \le D,\\ Y_1 \leftrightarrow X_1 \leftrightarrow X_{2,n},\\ Y_t \leftrightarrow (X_t, Y_{1,t-1}) \leftrightarrow (X_{1,t-1}, X_{t+1,n})}} \frac{1}{p}\sum_{t=1}^n I(X_t; Y_t | Y_{1,t-1}). \tag{60}$$

It is easy to see that the RHS term in (60) corresponds precisely to the sequential or NRDF obtained for parallel Gauss–Markov processes with a total MSE distortion constraint which is a simple generalization of the scalar-valued problem that has already been studied in [36]. Therefore, using the analysis of [36], we can obtain

$$\mathcal{R}^{\mathrm{IID},1}_{\mathrm{sum}}(D)$$

$$\stackrel{(a)}{=} \min_{\text{contraint in (60)}} \frac{1}{p}\sum_{t=1}^n \{h(X_t|Y_{1,t-1}) - h(X_t|Y_{1,t})\}$$

$$\stackrel{(b)}{=} \frac{1}{p} \min_{\substack{\Delta_t \succeq 0,\ t\in\mathbb{N}_1^n\\ \frac{1}{n}\frac{1}{p}\sum_{t=1}^n \mathrm{trace}\,(\Delta_t) \le D}} \sum_{t=1}^n \max\left[0, \frac{1}{2}\log_2\left(\frac{|\Lambda_t|}{|\Delta_t|}\right)\right]$$

$$= \min_{\substack{D_t \ge 0,\ t\in\mathbb{N}_1^n\\ \frac{1}{n}\sum_{t=1}^n D_t \le D}} \sum_{t=1}^n \max\left[0, \frac{1}{2}\log_2\left(\frac{\lambda_t}{D_t}\right)\right] \tag{61}$$

where (a) follows by definition; (b) follows from the fact that $h(X_t|Y_{1,t-1}) = \frac{1}{2}\log_2(2\pi e)^p|\Lambda_t|$ where $\Lambda_t = \mathrm{diag}\,(\lambda_t, \ldots \lambda_t) \in \mathbb{R}^{p\times p}$ with $\lambda_t = \alpha^2_{t-1} D_{t-1} + \sigma^2_{w_{t-1}}$, and that $h(X_t|Y_{1,t}) = \frac{1}{2}$ $\log_2(2\pi e)^p|\Delta_t|$ where $\Delta_t = \mathrm{diag}(D_t, \ldots, D_t) \in \mathbb{R}^{p\times p}$ for $D \in [0,\infty)$. The optimization problem of (61) is already solved in [36, Theorem 2] and is given by (21)–(23).

Appendix B
Proof of Theorem 2

In this article, we bound the rate performance of the DPCM scheme described in Section III-B at each time instant, for any fixed finite time n, using an ECDQ scheme that utilizes the forward Gaussian test-channel realization that achieves the lower bound of Theorem 1. The scheme relies on the replacement of the quantization noise with an additive Gaussian noise with the same second moments (see e.g., [48] or [31, Chapter 5] and the references therein). First, note that the Gaussian test-channel linear realization of the lower bound in Theorem 1 is known to be [36]

$$Y_t = H_t X_t + (I_p - H_t)A_{t-1}Y_{t-1} + H^{\frac{1}{2}}V_t,\ V_t \sim \mathcal{N}(0;\Delta_t) \tag{62}$$

where $H_t \triangleq I_p - \Delta_t\Lambda_t^{-1} \succeq 0$, $\Delta_t \triangleq \text{diag}(D_t,\ldots,D_t) \succ 0$, $\Lambda_t = \text{diag}(\lambda_t,\ldots\lambda_t) \succ 0$.

Pre-/Post-Filtered ECDQ With Multiplicative Factors for Parallel Sources: [48] First, consider a p-dimensional lattice quantizer Q_p [32] such that $\mathbf{E}\{Z_tZ_t^{\rm T}\} = \Sigma_{V_t^c}, \Sigma_{V_t^c} \succ 0$, where $Z_t \in \mathbb{R}^p$ is a random dither vector (shared randomness) generated both at the encoder/decoder independent of the input signals $\widehat{X}_t$ and the previous realizations of the dither, uniformly distributed over the basic Voronoi cell of the p-dimensional lattice quantizer Q_p such that $V_t^c \sim Unif(0;\Sigma_{V_t^c})$. At the *encoder*, the lattice quantizer quantizes $H_t^{\frac{1}{2}}\widehat{X}_t + Z_t$, that is, $Q_p(H_t^{\frac{1}{2}}\widehat{X}_t + Z_t)$, where $\widehat{X}_t$ is given by (27). Then, the encoder applies entropy coding to the output of the quantizer and transmits the output of the entropy coder. At the *decoder*, the coded bits are received and the output of the quantizer is reconstructed, i.e., $Q_p(H_t^{\frac{1}{2}}\widehat{X}_t + Z_t)$. Then, it generates an estimate by subtracting Z_t from the quantizer's output and multiplies the result by Φ_t as follows:

$$Y_t = \Phi_t(Q_p(H_t^{\frac{1}{2}}\widehat{X}_t + Z_t) - Z_t) \tag{63}$$

where $\Phi_t = H_t^{\frac{1}{2}}$. The *coding rate at each time instant* of the conditional entropy of the MSE quantizer is given by

$$\begin{aligned} H(Q_p|Z_t) &= I(H^{\frac{1}{2}}\widehat{X}_t; H^{\frac{1}{2}}\widehat{X}_t + V_t^c)\\ &\overset{(a)}{=} I(H^{\frac{1}{2}}\widehat{X}_t; H^{\frac{1}{2}}\widehat{X}_t + V_t) + \mathcal{D}(V_t^c||V_t)\\ &\quad - \mathcal{D}(H^{\frac{1}{2}}\widehat{X}_t + V_t^c||H^{\frac{1}{2}}\widehat{X}_t + V_t)\\ &\overset{(b)}{\leq} I(H^{\frac{1}{2}}\widehat{X}_t; H^{\frac{1}{2}}\widehat{X}_t + V_t) + \mathcal{D}(V_t^c||V_t)\\ &\overset{(c)}{\leq} I(H^{\frac{1}{2}}\widehat{X}_t; H^{\frac{1}{2}}\widehat{X}_t + V_t) + \frac{p}{2}\log(2\pi e G_p)\\ &\overset{(d)}{=} I(X_t; Y_t|Y_{1,t-1}) + \frac{p}{2}\log(2\pi e G_p) \end{aligned} \tag{64}$$

where $V_t^c \in \mathbb{R}^p$ is the (uniform) coding noise in the ECDQ scheme and V_t is the corresponding Gaussian counterpart; (a) follows because the two random vectors V_t^c, V_t have the same second moments, and hence we can use the identity $\mathcal{D}(x||x') = h(x') - h(x)$; (b) follows because $\mathcal{D}(H\widehat{X}_t + V_t^c||H\widehat{X}_t + V_t) \geq 0$; (c) follows because the divergence of the coding noise from Gaussianity is less than or equal to $\frac{p}{2}\log(2\pi e G_p)$ [39], where G_p is the dimensionless normalized second moment of the lattice [31, Definition 3.2.2]; (d) follows from data processing properties, i.e., $I(X_t;Y_t|Y_{1,t-1}) \overset{(*)}{=} I(X_t;Y_t|Y_{t-1}) \overset{(**)}{=} I(\widehat{X}_t;\widehat{Y}_t) \overset{(***)}{=} I(H^{\frac{1}{2}}\widehat{X}_t; H^{\frac{1}{2}}\widehat{X}_t + V_t)$, where $(*)$ follows from the realization of (62), $(**)$ follows from the fact that $\widehat{X}_t$ and $\widehat{Y}_t$ [obtained by (28)] are independent of Y_{t-1}, and $(***)$ follows from (27), (62), and the fact that $H^{\frac{1}{2}}$ is an invertible operation. Since we assume joint (memoryless) entropy coding with lattice quantizers, then, the total coding rate per dimension is obtained as follows[49, Chapter 5.4]:

$$\begin{aligned} \sum_{t=1}^n \frac{\mathbf{E}|S_t|}{p} &\leq \frac{1}{p}\sum_{t=1}^n (H(Q_p|Z_t) + 1)\\ &\overset{(e)}{\leq} \frac{1}{p}\sum_{t=1}^n I(X_t;Y_t|Y_{1,t-1}) + \frac{n}{2}\log(2\pi e G_p) + \frac{n}{p}\\ &\overset{(f)}{=} \frac{1}{2p}\sum_{t=1}^n \log_2\frac{|\Lambda_t|}{|\Delta_t|} + \frac{n}{2}\log(2\pi e G_p) + \frac{n}{p} \end{aligned} \tag{65}$$

where (e) follows from (64); (f) follows from the derivation of Theorem 1. The derivation is complete once we minimize both sides of the inequality in (65) with the appropriate constraint sets.

Appendix C
Proof of Theorem 3

Note that from (46), we obtain

$$\begin{aligned} &\Gamma^{\text{IID,op},ws} = \sum_{t=1}^n \text{LQG}_t^{\text{op}}\\ &= \frac{1}{p}\sum_{t=1}^n \Big\{\text{trace}(\Sigma_{W_t}\widetilde{K}_t)\\ &\quad + \text{trace}(A_tB_t\widetilde{L}_t\widetilde{K}_{t+1}\mathbf{E}\{||X_t - Y_t||_2^2\})\Big\}\\ &\overset{(a)}{\geq} \frac{1}{p}\sum_{t=1}^n \Big\{\text{trace}(\Sigma_{W_t}\widetilde{K}_t)\\ &\quad + \text{trace}(A_tB_t\widetilde{L}_t\widetilde{K}_{t+1}\mathbf{E}\{||X_t - \mathbf{E}\{X_t|S_{1,t}\}||_2^2\})\Big\}\\ &\overset{(b)}{\geq} \frac{1}{p}\sum_{t=1}^n \left\{\text{trace}(\Sigma_{W_t}\widetilde{K}_t) + \text{trace}\left(A_tB_t\widetilde{L}_t\widetilde{K}_{t+1}\right.\right.\\ &\left.\left.\mathbf{E}_{\bar{S}_{1,t-1}}\left\{\frac{1}{2\pi e}2^{\frac{2}{p}h(X_t|S_{1,t-1}=\bar{S}_{1,t-1})}\right\}2^{-2R_t^*}\right)\right\}\\ &\overset{(c)}{\geq} \frac{1}{p}\sum_{t=1}^n \left\{\text{trace}(\Sigma_{W_t}\widetilde{K}_t) + \text{trace}\left(A_tB_t\widetilde{L}_t\widetilde{K}_{t+1}\right.\right.\\ &\left.\left.\left\{\frac{1}{2\pi e}2^{\frac{2}{p}h(X_t|S_{1,t-1})}2^{-2R_t^*}\right\}\right)\right\}\\ &\overset{(d)}{\geq} \sum_{t=1}^n \{\sigma_{w_t}^2K_t + \alpha_t\beta_tL_tK_{t+1}D(R_t^*)\} \triangleq \sum_{t=1}^n \text{LQG}_t^* \end{aligned} \tag{66}$$

where (a) follows from the fact that Y_t is $\mathbb{S}_{1,t}$-measurable and the MMSE is obtained for $Y_t = \mathbf{E}\{X_t|S_{1,t}\}$; (b) follows from the fact that $\mathbf{E}\{||X_t - \mathbf{E}\{X_t|S_{1,t}\}||_2^2\}) = \mathbf{E}_{\bar{S}_{1,t-1}}\{\mathbf{E}\{||X_t - \mathbf{E}\{X_t|S_{1,t}\}||_2^2|S_{1,t-1} = \bar{S}_{1,t-1}\}\}$, where $\mathbf{E}_{\bar{S}_{1,t}}\{\cdot\}$ is the expectation with respect to some vector $\bar{S}_{1,t-1}$ that is distributed similarly to $S_{1,t-1}$, also from the MSE inequality in [49, Theorem 17.3.2] and, finally, from the fact that $R_t^* \geq 0$, where $R_t^* = \frac{1}{p}\{h^*(X_t|Y_{1,t-1}) - h^*(X_t|Y_{1,t})\}$ (see the derivation of Theorem 1, 1)) with $h^*(X_t|Y_{1,t-1})$, $h^*(X_t|Y_{1,t})$ being the minimized values in (61); (c) follows from Jensen's inequality [49, Theorem 2.6.2], i.e., $\mathbf{E}_{\bar{S}_{1,t-1}}\{2^{\frac{2}{p}h(X_t|S_{1,t-1}=\bar{S}_{1,t-1})}\} \geq 2^{\frac{2}{p}h(X_t|S_{1,t-1})}$; (d) follows from the fact that $\{h(X_t|S_{1,t-1}) = h(A_{t-1}X_{t-1} + B_{t-1}U_{t-1} + W_{t-1}|S_{1,t-1}) : t \in \mathbb{N}_2^n\}$ is completely specified from the independent Gaussian noise process $\{W_{t-1} : t \in \mathbb{N}_2^n\}$ because $\{U_{t-1} = g_t(S_{1,t-1}) : t \in \mathbb{N}_2^n\}$ [see (39)] are constants conditioned on $S_{1,t-1}$. Therefore, $h(X_t|S_{1,t-1})$ is conditionally Gaussian, thus equivalent to $h(X_t|Y_{1,t-1})$. This further means that $\frac{1}{2\pi e}2^{\frac{2}{p}\ h(X_t|Y_{1,t-1})}2^{-2R_t^*} \geq \frac{1}{2\pi e}2^{\frac{2}{p}h^*(X_t|Y_{1,t-1})}2^{-2R_t^*} \overset{(\star)}{=} \frac{1}{2\pi e}2^{\frac{1}{p}\log_2(2\pi e)^p|\Delta_t^*|} \overset{(\star\star)}{=} \min\{D_t\} \equiv D(R_t^*)$, where $(\star)$ follows because $h^*(X_t|Y_{1,t}) = \frac{1}{2}\log_2(2\pi e)^p|\Delta_t^*|$ and $(\star\star)$ follows because $\Delta_t^* = \mathrm{diag}(\min\{D_t\}, \ldots, \min\{D_t\})$.

It remains to find $D(R_t^*)$ at each time instant in (66). To do so, we reformulate the solution of the dynamic reverse-waterfilling solution in (21) as follows:

$$nR \equiv \mathcal{R}_{\text{sum}}^{\text{IID},1} = \sum_{t=1}^{n} R_t^* \equiv \frac{1}{2}\sum_{t=1}^{n}\log_2\left(\frac{\lambda_t}{D_t}\right)$$
$$= \frac{1}{2}\Bigg\{\underbrace{\cancelto{0}{\log_2(\lambda_1)}}_{\text{initial step}}$$
$$+ \sum_{t=1}^{n-1}\log_2\left(\alpha_t^2 + \frac{\sigma_{w_t}^2}{D_t}\right) - \underbrace{\log_2 D_n}_{\text{final step}}\Bigg\}. \quad (67)$$

From (67), we observe that at each time instant, the rate R_t^* is a function of only one distortion D_t since we have now decoupled the correlation with D_{t-1}. Moreover, we can assume without loss of generality, the initial step is zero because it is independent of D_0. Thus, from (67), we can find at each time instant, a $D_t \in (0, \infty)$ such that the rate is $R_t^* \in [0, \infty)$. Since the rate distortion problem is equivalent to the distortion rate problem (see, e.g., [49, Chapter 10]), we can immediately compute the total distortion rate function, denoted by $D_{\text{sum}}^{\text{IID},1}(R)$, as follows:

$$D_{\text{sum}}^{\text{IID},1}(R) \triangleq \sum_{t=1}^{n} D(R_t^*) = \sum_{t=1}^{n-1}\frac{\sigma_{w_t}^2}{2^{2R_t^*} - \alpha_t^2} + 2^{-2R_n^*}. \quad (68)$$

Substituting $D(R_t^*)$ at each time instant in (66), the result follows. This completes the proof.

Appendix D
Proof of Theorem 4

Note that from Lemma 4, (46), we obtain

$$\Gamma_{\text{sum}}^{\text{IID,op},ws} = \frac{1}{p}\sum_{t=1}^{n}\Big\{\mathrm{trace}(\Sigma_{W_t}\widetilde{K}_t)$$
$$+ \mathrm{trace}(A_t B_t \widetilde{L}_t \widetilde{K}_{t+1}\mathbf{E}\{||X_t - Y_t||_2^2\})\Big\}$$
$$= \frac{1}{p}\sum_{t=1}^{n}\Big\{\mathrm{trace}(\Sigma_{W_t}\widetilde{K}_t) + \mathrm{trace}(A_t B_t \widetilde{L}_t \widetilde{K}_{t+1} D(R_t^{\text{op}}))\Big\}$$
$$\overset{(a)}{\leq} \sum_{t=1}^{n-1}\left\{\sigma_{w_t}^2 K_t + \alpha_t\beta_t L_t K_{t+1}\frac{4^{\frac{1}{p}}(2\pi e G_p)\sigma_{w_t}^2}{2^{2R_t^{\text{op}}} - 4^{\frac{1}{p}}(2\pi e G_p)\alpha_t^2}\right\}$$
$$+ \sigma_{w_n}^2 K_n \quad (69)$$

where (a) follows because we can use Theorem 2, (31), to reformulate $\{R_t^* : t \in \mathbb{N}_1^n\}$ similar to (67) (in the derivation of Theorem 3) so that we decouple the dependence on D_{t-1} at each time step. Finally, for each $R_t^{\text{op}}, t = 1, 2 \ldots, n-1$, in (31), we solve with respect to the equivalent inverse problem of the distortion rate function, i.e., $D(R_t^{\text{op}})$, which gives

$$D(R_t^{\text{op}}) \leq \frac{4^{\frac{1}{p}}(2\pi e G_p)\sigma_{w_t}^2}{2^{2R_t^{\text{op}}} - 4^{\frac{1}{p}}(2\pi e G_p)\alpha_t^2},\ t \in \mathbb{N}_1^{n-1}. \quad (70)$$

Observe that the last step $t = n$ is not needed because in (46), we have $K_{n+1} = 0$. This completes the article.

Acknowledgment

The authors wish to thank the Associate Editor and the anonymous reviewers for their valuable comments and suggestions. They are also indebted to Prof. T. Charalambous for reading the paper and proposing the idea of bisection method for Algorithm 1. They are also grateful to Prof. J. Østergaard for fruitful discussions on technical issues of the article.

References

[1] X. Zhang, Q. Han, and X. Yu, "Survey on recent advances in networked control systems," *IEEE Trans. Ind. Inform.*, vol. 12, no. 5, pp. 1740–1752, Oct. 2016.
[2] H. Viswanathan and T. Berger, "Sequential coding of correlated sources," *IEEE Trans. Inf. Theory*, vol. 46, no. 1, pp. 236–246, Jan. 2000.
[3] S. C. Tatikonda, "Control under communication constraints," Ph.D. dissertation, Mass. Inst. Tech. (M. I. T.), Cambridge, MA, USA, 2000.
[4] V. N. Koshelev, "Hierarchical coding of discrete sources," *Problems Inf. Transmiss.*, vol. 3, no. 16, pp. 31–49, 1980.
[5] W. H. R. Equitz and T. M. Cover, "Successive refinement of information," *IEEE Trans. Inf. Theory*, vol. 37, no. 2, pp. 269–275, Mar. 1991.
[6] H. S. Witsenhausen, "On the structure of real-time source coders," *Bell Syst. Tech. J.*, vol. 58, no. 6, pp. 1437–1451, Jul. 1979.
[7] N. Gaarder and D. Slepian, "On optimal finite-state digital transmission systems," *IEEE Trans. Inf. Theory*, vol. 28, no. 2, pp. 167–186, Mar. 1982.
[8] D. Teneketzis, "On the structure of optimal real-time encoders and decoders in noisy communication," *IEEE Trans. Inf. Theory*, vol. 52, no. 9, pp. 4017–4035, Sep. 2006.
[9] T. Linder and S. Yüksel, "On optimal zero-delay coding of vector Markov sources," *IEEE Trans. Inf. Theory*, vol. 60, no. 10, pp. 5975–5991, Oct. 2014.
[10] R. G. Wood, T. Linder, and S. Yüksel, "Optimal zero delay coding of Markov sources: Stationary and finite memory codes," *IEEE Trans. Inf. Theory*, vol. 63, no. 9, pp. 5968–5980, Sep. 2017.
[11] P. A. Stavrou, J. østergaard, and C. D. Charalambous, "Zero-delay rate distortion via filtering for vector-valued Gaussian sources," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 5, pp. 841–856, Oct. 2018.
[12] 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.
[13] E. H. Yang, L. Zheng, D. K. He, and Z. Zhang, "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.
[14] 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.
[15] M. Torbatian and E. h. Yang, "Causal coding of multiple jointly Gaussian sources," in *Proc. 50th Annu. Allerton Conf. Commun., Control, Comput.*, 2012, pp. 2060–2067.
[16] A. Khina, V. Kostina, A. Khisti, and B. Hassibi, "Tracking and control of Gauss–Markov processes over packet-drop channels with acknowledgments," *IEEE Control Netw. Syst.*, vol. 6, no. 2, pp. 549–560, Jun. 2019.
[17] V. S. Borkar, S. K. Mitter, and S. Tatikonda, "Optimal sequential vector quantization of Markov sources," *SIAM J. Control Optim.*, vol. 40, no. 1, pp. 135–148, 2001.

[18] A. K. Gorbunov and M. S. Pinsker, "Nonanticipatory and prognostic epsilon entropies and message generation rates," *Problems Inf. Transmiss.*, vol. 9, no. 3, pp. 184–191, 1973.

[19] A. K. Gorbunov and M. S. Pinsker, "Prognostic epsilon entropy of a Gaussian message and a Gaussian source," *Problems Inf. Transmiss.*, vol. 10, no. 2, pp. 93–109, 1974.

[20] S. Tatikonda, A. Sahai, and S. Mitter, "Stochastic linear control over a communication channel," *IEEE Trans. Autom. Control*, vol. 49, pp. 1549–1561, Sep. 2004.

[21] 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.

[22] T. Tanaka, 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.

[23] P. A. Stavrou, T. Charalambous, C. D. Charalambous, and S. Loyka, "Optimal estimation via nonanticipative rate distortion function and applications to time-varying Gauss-Markov processes," *SIAM J. Control Optim.*, vol. 56, no. 5, pp. 3731–3765, Oct. 2018.

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

[25] T. Tanaka, K. H. Johansson, T. Oechtering, H. Sandberg, and M. Skoglund, "Rate of prefix-free codes in LQG control systems," in *Proc. IEEE Int. Symp. Inf. Theory*, 2016, pp. 2399–2403.

[26] T. Tanaka, P. M. Esfahani, and S. K. Mitter, "LQG control with minimum directed information: Semidefinite programming approach," *IEEE Trans. Autom. Control*, vol. 63, no. 1, pp. 37–52, Jan. 2018.

[27] 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.

[28] E. I. Silva, M. S. Derpich, J. Østergaard, and M. A. Encina, "A characterization of the minimal average data rate that guarantees a given closed-loop performance level," *IEEE Trans. Inf. Theory*, vol. 61, no. 8, pp. 2171–2186, Aug. 2016.

[29] J. L. Massey, "Causality, feedback and directed information," in *Proc. Int. Symp. Inf. Theory Appl.*, Nov. 1990, pp. 303–305.

[30] D. A. Harville, *Matrix Algebra From a Statistician's Perspective*. New York, NY, USA: Springer, 1997.

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

[32] J. H. Conway and N. J. A. Sloane, *Sphere-Packings, Lattices, and Groups*, 3rd ed., New York, NY, USA: Springer, 1999.

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

[34] Y. Oohama, "The rate-distortion function for the quadratic Gaussian CEO problem," *IEEE Trans. Inf. Theory*, vol. 44, no. 3, pp. 1057–1070, May 1998.

[35] H. Marko, "The bidirectional communication theory–A generalization of information theory," *IEEE Trans. Commun.*, vol. 21, no. 12, pp. 1345–1351, Dec. 1973.

[36] P. A. Stavrou, T. Charalambous, and C. D. Charalambous, "Finite-time nonanticipative rate distortion function for time-varying scalar-valued Gauss–Markov sources," *IEEE Control Syst. Lett.*, vol. 2, no. 1, pp. 175–180, Jan. 2018.

[37] K. Atkinson, *An Introduction to Numerical Analysis*, 2nd ed., Hoboken, NJ, USA: Wiley, 1991.

[38] 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.

[39] R. Zamir and M. Feder, "On lattice quantization noise," *IEEE Trans. Inf. Theory*, vol. 42, no. 4, pp. 1152–1159, Jul. 1996.

[40] P. Zador, "Asymptotic quantization error of continuous signals and the quantization dimension," *IEEE Trans. Inf. Theory*, vol. 28, no. 2, pp. 139–149, Mar. 1982.

[41] S. Yüksel and T. Basar, *Stochastic Networked Control Systems: Stabilization and Optimization Under Information Constraints*. New York, NY, USA: Springer, 2013.

[42] Y. Bar-Shalom and E. Tse, "Dual effect, certainty equivalence, and separation in stochastic control," *IEEE Trans. Autom. Control*, vol. 19, no. 5, pp. 494–500, Oct. 1974.

[43] L. Bao, M. Skoglund, and K. H. Johansson, "Iterative encoder-controller design for feedback control over noisy channels," *IEEE Trans. Autom. Control*, vol. 56, no. 2, pp. 265–278, Feb. 2011.

[44] M. Fu, "Lack of separation principle for quantized linear quadratic Gaussian control," *IEEE Trans. Autom. Control*, vol. 57, no. 9, pp. 2385–2390, Sep. 2012.

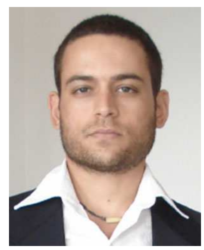
[45] S. Yüksel, "Jointly optimal LQG quantization and control policies for multi-dimensional systems," *IEEE Trans. Autom. Control*, vol. 59, no. 6, pp. 1612–1617, Jun. 2014.

[46] D. P. Bertsekas, *Dynamic Programming and Optimal Control*, Belmont, MA, USA: Athena Scientific, 2005.

[47] S. Tatikonda, A. Sahai, and S. Mitter, "Control of LQG systems under communication constraints," in *Proc. 37th IEEE Conf. Decis. Control*, 1998, pp. 1165–1170.

[48] 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.

[49] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed., Hoboken, NJ, USA: Wiley, 2006.



Photios A. Stavrou (Member, IEEE) received the D.Eng. degree from the Department of Electrical and Computer Engineering (ECE), Faculty of Engineering, Aristotle University of Thessaloniki, Thessaloniki, Greece, in 2008 and the Ph.D degree from the Department of ECE, Faculty of Engineering, University of Cyprus, Nicosia, Cyprus, in 2016.

From November 2016 to October 2017, he was a Postdoctoral Researcher with Aalborg University and from November 2017 to October 2019, he was a Postdoc with the KTH Royal Institute of Technology, Stockholm, Sweden, where he has been a Researcher since 2019. His research interests include information theory, coding in networked control systems, convex optimization, state estimation, and game theory.

Mikael Skoglund (Fellow, IEEE) received the M.Sc. degree in electrical engineering and the Ph.D. degree in information theory from Chalmers University of Technology, Gothenburg, Sweden in in 1992 and 1997, respectively.

In 1997, he joined the Royal Institute of Technology (KTH), Stockholm, Sweden, where he was appointed as the Chair in Communication Theory in 2003, and currently heads the Division of Information Science and Engineering. He has worked on problems in source-channel coding, coding, and transmission for wireless communications, Shannon theory, information and control, and statistical signal processing. He has authored and coauthored 150 journals and 350 conference papers.

Dr. Skoglund was an Associate Editor for the IEEE Transactions on Communications during 2003–08, and, during 2008–12, he was on the editorial board for the IEEE Transactions on Information Theory. He has served on numerous technical program committees for IEEE-sponsored conferences, General Co-Chair for IEEE ITW 2019, and he will serve as TPC Co-Chair for IEEE International Symposium on Information Theory 2022.

Takashi Tanaka (Memeber, IEEE) received the B.S. degree from the University of Tokyo, Tokyo, Japan, in 2006, and the M.S. and Ph.D. degrees in aerospace engineering (automatic control) from the University of Illinois at Urbana-Champaign (UIUC), Champaign, IL, USA, in 2009 and 2012, respectively.

He was a Postdoctoral Associate with the Laboratory for Information and Decision Systems (LIDS), Massachusetts Institute of Technology (MIT), Cambridge, MA, USA, from 2012 to 2015, and a Postdoctoral Researcher with the KTH Royal Institute of Technology, Stockholm, Sweden, from 2015 to 2017. He is currently an Assistant Professor with the Department of Aerospace Engineering and Engineering Mechanics, University of Texas at Austin, Austin, TX, USA. His research interests include control theory and its applications, most recently the information-theoretic perspectives of optimal control problems.

Dr. Tanaka was the recipient of the DARPA Young Faculty Award, the AFOSR Young Investigator Program Award, and the National Science Foundation Career Award.