Abstract—We introduce zero-delay source coding for scalar-valued Gaussian processes and mean-squared error (MSE) distortion constraint with causal side information (CSI) at the encoder-decoder (ED) pair. We introduce a lower bound by minimizing a variant of causally conditioned directed information subject to a MSE distortion constraint. For the lower bound, we first solve the optimization problem assuming general processes and then we restrict our attention to Gaussian processes. For the latter, we derive the optimal realization scheme and the corresponding lower bound in both finite-time and in the asymptotic limit. Then, we restrict the result to Gaussian processes for the asymptotic limit, we draw direct connections to existing results in the literature that include among others the classical rate distortion function (RDF) with jointly Gaussian random variables (RVs) with ED side information [6], [7] and nonanticipatory $\epsilon$-entropy of [8]. Finally, using the optimal realization scheme of the optimal test channel obtained for the lower bound we explain how by invoking standard techniques of ECDQ [9] we can get an achievable bound on the zero-delay Gaussian RDF with CSI at the ED pair. Our framework is enhanced with a simulation experiment.

Notation: We let $\mathbb{R} = (-\infty, \infty)$, $\mathbb{Z} = \{\ldots, -1, 0, 1, \ldots\}$, $N_0 = \{0, 1, \ldots, n\}$, $n \in N_0$. A RV defined on some probability space $(\Omega, \mathcal{F}, P)$ is a map $x: \Omega \rightarrow \mathcal{X}$, where $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ is a measurable space. We denote a sequence of RVs by $x^{T}_t \triangleq (x_r, x_{r+1}, \ldots, x_t), (r,t) \in \mathbb{Z} \times \mathbb{Z}, t \geq r$, and their values by $x^{T}_r \equiv x_{k=r}^t \equiv x^T_k$, with $x_{k} = \mathcal{X}$, for simplicity. If $r = -\infty$ and $t = -1$, we use the notation $x^{-1}_\infty = x^{-1}$, and if $r = 0$, we use the notation $x^{0}_t = x^t$. The distribution of the RV $x$ on $\mathcal{X}$ is denoted by $P_x(dx) \equiv P(dx)$. The conditional distribution of RV $y$ given $x = x$ is denoted by $P_{y|x}(dy|x) \equiv P(dy|x)$. The transpose of a vector $S$ is denoted by $S^t$ and the the expectation operator by $E\{\cdot\}$.

I. INTRODUCTION

Zero-delay source coding is relevant in understanding and identifying the fundamental performance limitations of networked control systems because it requires by design instantaneous encoding and decoding, thus, mitigating the processing latency in a communication link. This coding approach is adopted in many works that consider closed-loop control systems with communication constraints, see, e.g., [1]–[3].

In this paper, we introduce and analyze a generalization of the zero-delay source coding framework where a scalar-valued Gaussian source modeled as a Gauss-Markov process is conveyed via an encoder through a noiseless link to a decoder subject to a MSE distortion constraint. Compared to existing works on this field (see, e.g., [2], [4]) here we allow the ED pair to have access to some CSI which is an noisy measurement of the source. For this setting, we first obtain a characterization of a generic lower bound formulated as an optimization problem where we minimize a variant of causally conditioned directed information [5] assuming correlated general sources and CSI processes that are unaffected by the previous estimates of the system subject to a MSE distortion constraint. For this optimization problem, we give the structure of the optimal minimizer obtained sequentially backward in time. Then, we restrict the result to Gaussian processes for which we find the optimal realization and characterize the lower bound in both finite-time and in the asymptotic limit. For the asymptotic limit, we draw direct connections to existing results in the literature that include among others the classical rate distortion function (RDF) with jointly Gaussian RVs with ED side information [6], [7] and nonanticipatory $\epsilon$-entropy of [8]. Finally, using the optimal realization scheme of the optimal test channel obtained for the lower bound we explain how by invoking standard techniques of ECDQ [9] we can get an achievable bound on the zero-delay Gaussian RDF with CSI at the ED pair. Our framework is enhanced with a simulation experiment.

II. PROBLEM STATEMENT

We consider the zero-delay source coding setting illustrated in Fig. 1 where the source and the CSI processes are generated by the joint process $\{(x_t, z_t): t \in N_0\}$ that is assumed to be jointly Gaussian.

In particular, the Gaussian source is modeled by the following discrete-time time-invariant Gauss-Markov process

$$x_{t+1} = \alpha x_t + w_t, \quad x_0 \in \mathbb{R}, \quad t \in N_0,$$

where $\alpha \in \mathbb{R}$ is a non-random coefficient, $x_0 \in \mathcal{X}$ is the initial state, $w_t \in \mathbb{R} \sim \mathcal{N}(0; \sigma_w^2)$ is an independent Gaussian process, independent of $(x_0, \{w_t: t \in N_0\})$. The CSI is modeled by the following discrete-time time-invariant Gaussian process

$$z_t = c x_t + n_t, \quad t \in N_0,$$

where $c \in \mathbb{R}$ are non-random coefficients, and $n_t \in \mathcal{X}$ is an independent Gaussian process, independent of $(x_0, \{w_t: t \in N_0\})$.

The system operates as follows. At every time step $t \in N_0$, the encoder observes $(x_t, z_t)$ and has access to $x^{t-1}$, $z^{t-1}$. Then, it produces a single binary codeword $I_t \in \{0, 1\}^*$ from a
predefined set of codewords $L_t$ of at most countable number of codewords. Since the source and the CSI are random, $I_t$ is a RV. Upon receiving $I_t$, the MMSE decoder produces an estimate $y_t$ of the source/CSI sample $(x_t, z_t)$, under the assumption that $Y^{t-1}, x^t, z^t$ have been received. We assume that both the encoder and decoder process information without delay.

![Fig. 1. Zero-delay source coding with CSI at the encoder/decoder.](image)

The analysis of the noiseless digital channel is restricted to the class of all codewords (codebook) $L_t$, which is a RV. The countable set of codewords (codebook) $L_t$ is time-varying to allow the binary sequence $I_t$ to be arbitrarily long.

**Zero-delay source coding with CSI.** Formally, the zero-delay source coding problem with CSI of Fig. 1 can be explained as follows. Define the input and output alphabet of the noiseless digital channel by $\mathcal{M} = \{1, 2, \ldots, M_t\}$ where $M_t = \max_{\{x_t, z_t\}} |L_t|$ that is allowed to be infinite. The elements in $\mathcal{M}$ enumerate the codewords of $L_t$. The encoder and the decoder are specified by the sequence of functions $\{(f_t, g_t) : t \in \mathbb{N}_0\}$ as follows:

$$f_t : \mathcal{M}^{t-1} \times \mathcal{X}^t \times \mathcal{Z}^t \rightarrow \mathcal{M}_t, \quad t \in \mathbb{N}_0,$$

$$g_t : \mathcal{M}^t \times \mathcal{Z}^t \rightarrow \mathcal{Y}_t.$$  

(3)

For time $t \in \mathbb{N}_0$, the output of the encoder is a message $l_t = f_t(l^{t-1}, x^t, z^t)$ with $f_0 = f_0(x_0, z_0)$ which is transmitted through a noiseless channel to the MMSE decoder. The latter generates $y_t = g_t(l^t, z^t)$ with $y_0 = g_0(l_0, z_0)$ assuming $l^{t-1}, z^{t-1}$ is already generated. Moreover, the expected rate (in bits) at each time instant, is defined as $R_t \triangleq E[I_t]$, where $|\cdot|$ denotes the cardinality of the binary sequence $I_t$.

**Performance.** The objective is to minimize the average coding rate, i.e., the average expected number of bits received by the decoder at the time it reproduces $\{y_t : t \in \mathbb{N}_0\}$, denoted by $\limsup_{n \rightarrow \infty} \frac{1}{n+1} \sum_{t=1}^{n} R_t$, over all encoding and decoding policies $\{(f_t, g_t) : t \in \mathbb{N}_0\}$. This can be cast to the following optimization problem:

$$R_{\text{ZD}}^{\text{CS}} \triangleq \inf_{(f_t, g_t), t \in \mathbb{N}_0} \limsup_{n \rightarrow \infty} \frac{1}{n+1} \sum_{t=1}^{n} R_t, \quad \text{s.t.} \quad \exists y_t \triangleq E[I_t]$$  

(4)

where $R_{\text{ZD}}^{\text{CS}}(D)$ is the operational zero-delay Gaussian RDF with CSI at the ED and $D \in [0, D_{\text{max}}]$.

Our goal is to derive lower and upper bounds on the optimal rates given by (4) as the problem as such is very difficult to compute.

III. LOWER BOUND

In this section, we first derive a lower bound on (4). To do so, we first introduce a lower bound for general sources and CSI at the ED pair with the information structure assumed in (1), (2), respectively, subject to a MSE distortion. This bound is cast by minimizing a variant of causally conditioned directed information [5, Eq. (3.8)]. For the resulting optimization problem we compute the optimal solution via recursions obtained backward in time and we derive information structures of the optimal test channel. Then, we turn our attention and solve the Gaussian case.

### A. General sources and CSI with MSE distortion

We consider a source that randomly generates sequences $x_t = x_t \in \mathcal{X}_t, t \in \mathbb{N}_0$ and a correlated CSI that randomly generates sequences $z_t = z_t \in \mathcal{Z}_t, t \in \mathbb{N}_0$ and we wish to reproduce or reconstruct by $y_t = y_t \in \mathcal{Y}_t, t \in \mathbb{N}_0$, subject to a MSE distortion constraint defined by $d_{0,n}(x^n, y^n) \triangleq \sum_{t=0}^{n} \|x_t - y_t\|^2$.

**Source and CSI.** The source distribution satisfies conditional independence

$$P_{x_t | x^{t-1}, y^{t-1}} \triangleq P(dx_t | x_{t-1}), \quad t \in \mathbb{N}_0,$$  

(5)

whereas the CSI distribution satisfies conditional independence

$$P_{z_t | x^{t-1}, y^{t-1}} \triangleq P(dz_t | x_{t}), \quad t \in \mathbb{N}_0.$$  

(6)

We assume no initial information in the system, thus, at $t = 0$ we have $P(dz_0 | x_0) = P(dx_0)$. Also, by Bayes’ rule we obtain $P_{x^n} \equiv P(dx^n) \triangleq \otimes_{t=0}^{n} P(dx_t | x_{t-1})$ and $P_{z^n | x^n} \equiv P(dz^n | x^n) \triangleq \otimes_{t=0}^{n} P(dz_t | x_t)$.

**Reproduction or “test-channel”,** Suppose the reproduction $y_t = y_t, t \in \mathbb{N}_0$ of $x^t, z^t$ is randomly generated, according to the collection of conditional distributions by

$$P_{y_t | x^{t-1}, z^t} \triangleq P(dy_t | x^{t-1}, z^t), \quad t \in \mathbb{N}_0.$$  

(7)

Again, at $t = 0$, we assume that $P(dy_0 | x^{-1}, z_0, x_0) = P(dy_0 | z_0, x_0)$. Similar to [10], we can uniquely define the family of conditional distributions on $y^n$ parametrized by $(x^n, z^n) \in \mathcal{X}^n \times \mathcal{Z}^n$ by $P(dy^n | x^n, z^n) \triangleq \otimes_{t=0}^{n} P(dy_t | x^n, z^n)$, and vice-versa. Further, from (5)-(7), we can uniquely define the joint distribution of $\{(x_t, z_t, y_t) : t \in \mathbb{N}_0\}$ by

$$P_{x^n, y^n, z^n} = P(dx^n, dz^n, dy^n) = P(dx^n) \otimes \mathcal{P}(dz^n | x^n) \otimes \mathcal{Q}(dy^n | z^n, x^n)$$  

(8)

$$= P(dx^n, dz^n) \otimes \mathcal{Q}(dy^n | z^n, x^n).$$  

Further, from (8), we can uniquely define the $Y_t$-marginal distribution parametrized by $z^t \in \mathcal{Z}^t$ by $P_{y_t | y^{-1}, z^t} \triangleq P(dy_t | y^{t-1}, z^t)$, and by Bayes’ rule the conditional distribution $P_{y_t | x^n} \triangleq \otimes_{t=0}^{n} P(dy_t | x^n, y^{t-1}, z^t)$. At $t = 0$, $P(dy_0 | y^{-1}, z_0) = P(dy_0 | z_0)$. 


Given the above construction of distributions, we introduce the corresponding information measure as follows:

\[
I(x^n; y^n|z^n) = \sum_{t=0}^{n} \mathbb{E} \left\{ \log \left( \frac{\mathbb{P}(y_t|y_{t-1}, z^t)}{\mathbb{P}(y_t|y_{t-1}, z^t)} \right) \right\} = \sum_{t=0}^{n} I(x_t; y_t|y_{t-1}, z^t).
\]  

(9)

Clearly, (9) is a special case of directed information from \( x^n \) to \( y^n \) causally conditioned on \( z^n \) (see [5, Eq. (3.8)]).

Next, we formally introduce the generic lower bound which we refer to as NRDF with CSI at the ED subject to total MSE distortion constraint.

**Definition 1:** (NRDF with CSI at the ED)

For the finite-time NRDF with CSI at the ED is defined by

\[
R_{\text{na}}^{\text{CSI-ED}}(D) \triangleq \inf_{\mathbb{P}(dy^n|y_{n-1}, z^n), \sum_{t=0}^{n} E|x_t-y_t|^2 \leq D} I(x^n; y^n|z^n).
\]  

(10)

The asymptotic limit of (10) is defined by

\[
\lim_{n \to \infty} \frac{1}{n+1} R_{\text{na}}^{\text{CSI-ED}}(D) = \lim_{n \to \infty} \frac{1}{n+1} I(x^n; y^n|z^n),
\]  

assuming the limit exists and it is finite.

By interchanging \( \inf \) to \( \lim \) in (11), an upper bound on \( R_{\text{na}}^{\text{CSI-ED}}(D) \) is obtained, as defined as follows:

\[
\tilde{R}_{\text{na}}^{\text{CSI-ED}}(D) \triangleq \inf_{\mathbb{P}(dy^n|z^n), \sum_{t=0}^{n} E|x_t-y_t|^2 \leq D} \lim_{n \to \infty} \frac{1}{n+1} I(x^n; y^n|z^n).
\]  

(12)

where \( \tilde{Q}(dy^n|z^n, x^n) \) denotes the sequence of conditional probability distributions \( \{\mathbb{P}(dy_t|y_{t-1}, z^t, x^t) : t \in \mathbb{N}_0\} \).

In the next remark, we stress some connections of (10) to existing results in the literature and discuss some of its properties.

**Remark 1:** (Connections to existing results and properties of \( R_{\text{na}}^{\text{CSI-ED}}(D) \))

(a) The asymptotic limit in (11) serves as an upper bound to the classical RDF with non-causal side information at the ED pair that was originally introduced in [6, 11], i.e.,

\[
R_{\text{na}}^{\text{SI-ED}}(D) = \lim_{n \to \infty} \frac{1}{n+1} R_{\text{na}}^{\text{CSI-ED}}(D),
\]  

(13)

assuming the limit exists and it is finite, where

\[
R_{\text{na}}^{\text{SI-ED}}(D) \triangleq \inf_{\mathbb{P}(dy^n|y_{n-1}, x^n), \sum_{t=0}^{n} E|x_t-y_t|^2 \leq D} I(x^n; y^n|x^n),
\]  

(14)

and \( I(x^n; y^n|x^n) \) denotes the mutual information between \( x^n \) and \( y^n \) non-causally conditioned on \( x^n \) (see [12]). This is because the constraint set in (10) is smaller than the one in (14).

(b) The optimization problem in (10), in contrast to the one in (4), is convex with respect to the test channel (see [10]), for \( D \in [D_{\text{min}}, D_{\text{max}}] \subseteq [0, \infty) \). Moreover, under certain conditions (see for instance [10, Thm 15]), the infimum is achieved and \( R_{\text{na}}^{\text{CSI-ED}}(D) < \infty \).

Next, we give a theorem where we compute recursively backward in time the elements \( \{\mathbb{P}^*(dy_t|y_{t-1}, z^t, x^t) : t \in \mathbb{N}_0\} \) the optimization problem in (10).

**Theorem 1:** (Time-varying solution of (10))

Suppose there exists \( \tilde{Q}^*(\cdot|z^n, x^n) \), which solves (10) for \( D \in [D_{\text{min}}, D_{\text{max}}] \), the distortion constraint is non-empty and the objective function in (10) is Gâteaux differentiable in every direction of \( \{\mathbb{P}(\cdot|y_{t-1}, x^t, z^t) : t \in \mathbb{N}_0\} \) for fixed \( \mathbb{P}(\cdot) \in \mathcal{M}(x^t) \), \( P(\cdot) \in \mathcal{M}(z^t) \), where \( \mathcal{M}(\cdot) \) denotes the set of the distributions on \( x^t \) and \( z^t \), respectively. Then, the following hold:

(i) The optimal time-varying reproduction distributions are given by the following recursive equations backwards in time.

\[
P^*(dy_t|y_{t-1}, z^t, x^t) = e^{s|x_t-y_{t-1}|^2} \mathbb{P}^*(dy_{t+1}|y_{t+1}, z^t),
\]  

(15)

(ii) For \( t = n-1, n-2, \ldots, 1, 0 \):

\[
P^*(dy_t|y_{t-1}, z^t, x^t) = e^{s|x_t-y_{t-1}|^2-V_t(\cdot)} \mathbb{P}^*(dy_{t+1}|y_{t+1}, z^t),
\]  

(16)

where \( s < 0 \) is a Lagrange multiplier, and \( V_t(\cdot) \) is given by

\[
V_t(x_t, z^t, y^t) \triangleq \int_{\mathbb{R}^t} \mathbb{P}(dx_{t+1}|x_t) \mathbb{P}(dz_{t+1}|x_{t+1}) \log \left( \int_{\mathbb{R}^t} e^{s|x_{t+1-1}-y_{t+1}|^2-V_{t+1}(\cdot)} \mathbb{P}^*(dy_{t+1}|y_{t+1}, z^t) \right).
\]

(17)

with \( V_0(x_0, z^n, y^n) = 0 \).

**Proof:** The derivation uses similar steps to [13, Thm 4.1].

**Remark 2:** (Gaussianity of \( \mathbb{P}^*(dy_t|y_{t-1}, z^t, x^t) \))

If the joint process \( \{(x_t, z_t) : t \in \mathbb{N}_0\} \) is jointly Gaussian, then, one can show using the recursions in Theorem 1 and Kalman filter equations that \( \mathbb{P}^*(dy_t|y_{t-1}, z^t, x^t) \) is conditionally Gaussian.

B. A lower bound on (4)

In this subsection, we derive the analytical solution of \( R_{\text{na}}^{\text{CSI-ED}}(D) \) for Gaussian processes.

First, we derive the following lemma.

**Lemma 1:** (Realization of optimal reproduction distribution)

Since the joint process \( \{(x_t, z_t, y_t) : t \in \mathbb{N}_0\} \) is jointly Gaussian, then, the following statements hold.

(a) Any candidate of the optimal reproduction distribution \( \{\mathbb{P}^*(dy_t|y_{t-1}, z^t, x_t) : t \in \mathbb{N}_0\} \) is realized by the recursion

\[
y_t = h_t (x_t - \tilde{x}_t|t-1) + g_t (z_t - \tilde{z}_t|t-1) + \tilde{z}_t + v_t,
\]

(17)
and the fact that

\( \hat{\epsilon}_t = y_t - x_t \).

Moreover, the innovations process, denoted by \( \{ I_t : t \in \mathbb{N}_0 \} \), is given by

\[
I_t \triangleq y_t - E[y_t | z_{t-1}, y_{t-1}^*] = y_t - x_t - \hat{\epsilon}_{t-1} = (h_t + g_c)(x_t - \hat{x}_{t-1}) + g_t n_t + v_t,
\]

where \( I_t \sim \mathcal{N}(0; \sigma_{I_t}^2) \), \( \sigma_{I_t}^2 = (h_t + g_c)^2 \Sigma_{t-1} + g_t^2 \sigma_n^2 + \sigma_r^2 \), and \( \Sigma_{t-1} \triangleq E \{ (x_t - \hat{x}_{t-1})^2 | z_{t-1}, y_{t-1}^* \} \).

(b) Let \( \hat{x}_{t|t-1} \triangleq E \{ x_t | z_{t-1}, y_{t-1}^* \} \) and \( \Sigma_{t|t} \triangleq E \{ (x_t - \hat{x}_{t|t})^2 | z_{t-1}, y_{t-1}^* \} \). Then, \( \{ \hat{x}_{t|t-1}, \Sigma_{t|t} : t \in \mathbb{N}_0 \} \) satisfy the following scalar-valued filtering recursions:

\[
\hat{x}_{t|t-1} = a \hat{x}_{t-1} + b u_t, \\
\Sigma_{t|t} = \alpha^2 \Sigma_{t-1} + \sigma_n^2,
\]

(18)

\( k_t = \Sigma_{t-1} (h_t + g_c)(\sigma^2_{I_t})^{-1}, \) (Kalman Gain)

\( \Sigma_{t|t} = \Sigma_{t-1} + (h_t + g_c)^2 (\sigma^2_{I_t})^{-1}. \)

(c) \( R_{\text{CSI},[n,0]}(D) \) is defined by

\[
R_{\text{CSI},[n,0]}(D) = \inf_{h_t \in \mathbb{R}, g_t \in \mathbb{R}} \frac{1}{2} \sum_{t=0}^{n} \left[ \log \left( \frac{\text{var} \{ x_t | z_{t-1}, y_{t-1}^* \}}{\hat{\epsilon}_{t|t}} \right) \right]^+,
\]

(19a)

\[
\text{s.t. } \frac{1}{n+1} \sum_{t=0}^{n} \left[ (1 - (h_t + g_c))^2 \Sigma_{t|t-1} + g_t^2 \sigma_n^2 + \sigma_r^2 \right] \leq D,
\]

(19b)

where \( \text{var} \{ x_t | z_{t-1}, y_{t-1}^* \} = \frac{\sigma_n^2}{\sigma_{I_t}^2 + \sigma_n^2} \), for \( D \in [D_{\text{min}}, D_{\text{max}}] \subseteq [0, \infty] \), and \( \{ x_t \} \) are the innovations.

Proof: We sketch the proof due to space limitations.

(a) From Remark 2 we have the following orthogonal realization

\( y_t = h_t x_t + g_t z_t + r_t (z_t^{*-1}, y_t^{*-1}) + v_t, \)

where \( r_t (z_t^{*-1}, y_t^{*-1}) \) are vectors of appropriate dimensions and \( \{ v_t : t \in \mathbb{N}_0 \} \) is an independent Gaussian process. For such a realization, \( I(x^2, y^2) = \sum_{t=0}^{n} I(x_t \mid y_t^{*-1}, z^*_t) \) does not depend on \( r_t (z_t^{*-1}, y_t^{*-1}) \). Since \( E \{ \sum_{t=0}^{n} | x_t - y_t |^2 \} = \sum_{t=0}^{n} \left[ (1 - (h_t + g_c))^2 \sigma_n^2 + \sigma_r^2 \right] + \sum_{t=0}^{n} (g_t^2 \sigma_n^2 + \sigma_r^2) \), then, by mean-squared estimation theory, a smaller average distortion occurs when

\( r_t (z_t^{*-1}, y_t^{*-1}) = (1 - (h_t + g_c))^2 \sigma_n^2 + \sigma_r^2 \), \quad \forall t \in \mathbb{N}_0.

The second part of (a) follows from well-known properties of innovations process. (b) This follows from the discrete-time Kalman filtering equations [14, (c)]. (c) This follows from (a), (b) and the fact that \( \text{dist}(x^2, y^2) \) is conditionally Gaussian.

The next theorem gives the complete characterization of (10) for scalar-valued time-varying Gaussian processes and total MSE distortion constraint.

**Theorem 2:** (Characterization of (10))

Define \( \lambda_t \triangleq \Sigma_{t|t-1} \) and \( D_t \triangleq \Sigma_{t|t} \). Then, the following holds.

(a) The optimal reproduction of (19) is realized by

\[
y_t = h_t x_t + g_t z_t + (1 - (h_t + g_c)) \lambda_t y_{t-1} + v_t, \quad y_{t-1} = 0,
\]

(20)

where

\[
h_{t-1} = 1 - \frac{D_t}{\text{var} \{ x_t, z_t, y_{t-1}^* \}}, \quad g_t = \frac{D_t c}{\sigma_n^2} + \sigma_r^2, \quad \lambda_t = h_t D_t \geq 0,
\]

(21)

where \( \text{var} \{ x_t | z_t, y_{t-1}^* \} = \frac{\sigma_n^2}{\sigma_{I_t}^2 + \sigma_n^2} \), and \( \lambda_t = \alpha^2 D_t + \sigma_r^2 \), \( \lambda_0 = \sigma_n^2 \), \( t \in \mathbb{N}_0 \). Moreover, for the above realization it holds that \( \hat{x}_{t|t} = y_t, \quad \hat{x}_{t|t-1} = \alpha y_{t-1}, \quad \text{P}^*(d_{y_t} | y_{t-1}, z_t, x_t), \)

\( I(x^2; y^2 | z^2) = \sum_{t=0}^{n} I(x_t; y_t | y_{t-1}, z_t). \)

(b) The characterization of \( R_{\text{CSI},[n,0]}(D) \) is

\[
R_{\text{CSI},[n,0]}(D) = \min_{D_{t|t} \geq 0, t \in \mathbb{N}_0} \frac{1}{2} \sum_{t=0}^{n} \left[ \log \left( \frac{\text{var} \{ x_t | z_{t-1}, y_{t-1}^* \}}{D_t} \right) \right]^+,
\]

(22)

where \( D_{t|t} = \min \{ D_{t|t}^1, D_{t|t}^2 \}, D \subseteq [D_{\text{min}}, D_{\text{max}}] \subseteq [0, \infty] \).

Proof: We sketch the proof due to space limitations.

(a) From mean-squared estimation theory we know that the inequality

\[
E \left[ \sum_{t=0}^{n} | x_t - y_t |^2 \right] \geq E \left[ \sum_{t=0}^{n} | \hat{x}_{t|t} - \text{E} \{ x_t | z_{t-1}, y_{t-1}^* \} |^2 \right],
\]

holds \( \forall y_t, t \in \mathbb{N}_0 \), i.e., for all \( (h_t, g_t, \sigma^2_{I_t}), t \in \mathbb{N}_0 \), and it is achieved if and only if \( \text{E} \{ x_t | z_{t-1}, y_{t-1}^* \} = y_t \). The choice of (21) achieves this lower bound, i.e., a smaller distortion is achieved for a given rate. (b) This is immediate from (a).
with approximately \( D \) or the analytical expression of (26) or the analytical solution of (24) (without the optimal realization of (26) is replaced by a subtractively dithered uniform scalar quantizer [9] followed by memoryless entropy coding [12], (26) is approximated in (25) continue to hold even if \( |\alpha| > 1 \), i.e., if the source process in (1) is unstable.

In the next remark, we draw connections to existing results in the literature.

**Remark 3:** (Connections to existing results)

(a) The analytical solution of (24) (without the optimal realization of (26)) or the analytical expression of (25) recently derived in [16, Eq. (88)] using a different approach.

(b) If we assume that \( |\alpha| = 0 \) in the source model (1), then, \( R_{\text{na}}^{\text{CSI-ED}}(D) \equiv R_{\text{SI-ED}}(D) \), where

\[
R_{\text{SI-ED}}(D) = \frac{1}{2} \log \left( \frac{\sigma_n^2 \sigma_w^2}{\sigma_n^2 + \sigma_w^2 D} \right), \quad D \leq D_{\text{max}},
\]

with \( D_{\text{max}} = \frac{\sigma_w^2}{\sigma_n^2} \) and the optimal realization (26) simplifies to

\[
y_t = h x_t + g z_t + v_t,
\]

with \( h = 1 - \frac{D}{\sigma_w^2 + \sigma_n^2} \), \( g = \frac{D}{\sigma_n^2} \), and \( \sigma_v = h D \). This special case corresponds to the well-known result of [7, §3] for jointly Gaussian RVs \((x, z)\) (or i.i.d. sequences \( (x_t, z_{t-1}) : t \in \mathbb{N}_0 \)) when \( z_t \) has the structure of (2).

(c) If we assume \( c = 0 \) in the CSI model of (2), then, the analytical solution of (24) simplifies to

\[
R_{\text{na}}^{\text{CSI-ED}}(D) = \frac{1}{2} \log \left( \frac{\sigma_n^2}{\sigma_n^2 + D} \right), \quad D \leq D_{\text{max}},
\]

with \( D_{\text{max}} = \frac{\sigma_n^2}{\sigma_n^2} \) and the optimal realization of (26) simplifies to

\[
y_t = h x_t + (1 - h) \alpha y_{t-1} + v_t,
\]

where \( h = 1 - \frac{D}{\sigma_n^2} \), \( g = 0 \), and \( \sigma_v = h D \). This special case corresponds precisely to the NRDF or sequential RDF without side information, see, e.g., [15, Eq. (1.43)].

**IV. ACHIEVABLE BOUND ON (4)**

The optimal realization of the lower bound on (4) obtained in §III-B, gives the possibility to construct an achievable scheme on (4). In particular, when the AWGN channel in (26) is replaced by a subtractively dithered uniform scalar quantizer [9] followed by memoryless entropy coding [12], then, following similar steps to the approach put forward in [2, §V], we can obtain an upper bound on (4) given by

\[
R_{\text{ZD}}^{\text{CSI-ED}}(D) \leq R_{\text{na}}^{\text{CSI-ED}}(D) + \frac{1}{2} \log \left( \frac{\pi e}{6} \right) + 1,
\]

i.e., the gap between the lower and upper bound is approximately 1.254 bits/sample. A simulation experiment with random values for \((\alpha, c, \sigma_w^2, \sigma_n^2)\) is illustrated in Fig. 2. In this experiment we demonstrate the tightness of the achievable bound compared to the optimal rates of (4) and the lower bound obtained in (11). In addition, the plot illustrates the comparison between \( R_{\text{SI-ED}}^{\text{CSI}} \) and \( R_{\text{na}}^{\text{CSI-ED}}(D) \).

**V. ACKNOWLEDGEMENTS**

This work has received funding by the KAW Foundation and the Swedish Foundation for Strategic Research.

**REFERENCES**


