

# Fundamental limits of distributed tracking

Victoria Kostina, Babak Hassibi

**Abstract**—Consider the following communication scenario. An  $n$ -dimensional source with memory is observed by  $K$  isolated encoders via parallel channels, who causally compress their observations to transmit to the decoder via noiseless rate-constrained links. At each time instant, the decoder receives  $K$  new codewords from the observers, combines them with the past received codewords, and produces a minimum-distortion estimate of the latest block of  $n$  source symbols. This scenario extends the classical one-shot CEO problem to multiple rounds of communication with communicators maintaining memory of the past.

We prove a coding theorem showing that the minimum asymptotically (as  $n \rightarrow \infty$ ) achievable sum rate required to achieve a target distortion is equal to the directed mutual information from the observers to the decoder minimized subject to the distortion constraint and the separate encoding constraint. For the Gauss-Markov source observed via  $K$  parallel AWGN channels, we solve that minimal directed mutual information problem, thereby establishing the minimum asymptotically achievable sum rate. Finally, we explicitly bound the rate loss due to a lack of communication among the observers; that bound is attained with equality in the case of identical observation channels.

The general coding theorem is proved via a new nonasymptotic bound that uses stochastic likelihood coders and whose asymptotic analysis yields an extension of the Berger-Tung inner bound to the causal setting. The analysis of the Gaussian case is facilitated by reversing the channels of the observers.

**Index Terms**—CEO problem, Berger-Tung bound, distributed source coding, causal rate-distortion theory, Gauss-Markov source, LQG control.

## I. INTRODUCTION

We set up the causal CEO (chief executive or estimation officer) problem as follows. An information source  $\{X_i\}$  outputs  $X_i \in \mathcal{A}^n$  at time  $i$ ; it is observed by  $K$  encoders through  $K$  noisy channels; at time  $i$ ,  $k$ th encoder sees  $Y_i^k$  generated according to  $P_{Y_i^k|X_1, \dots, X_i, Y_1^k, \dots, Y_{i-1}^k}$ . See Fig. 1. The encoders (observers) communicate to the decoder (CEO) via their separate noiseless rate-constrained links. At each time  $i$ ,  $k$ th observer forms a codeword based on the observations it has seen so far, i.e.,  $Y_1^k, \dots, Y_i^k$ . The decoder at time  $i$  chooses  $\hat{X}_i \in \hat{\mathcal{A}}^n$  based on the codewords it received thus far. The goal is to minimize the average distortion

$$\frac{1}{t} \sum_{i=1}^t \mathbb{E} \left[ d(X_i, \hat{X}_i) \right], \quad (1)$$

where  $t$  is the *time horizon* over which the source is being tracked, and  $d: \mathcal{A}^n \times \hat{\mathcal{A}}^n \mapsto \mathbb{R}_+$  is the distortion measure.

The authors are with California Institute of Technology (e-mail: [vkostina@caltech.edu](mailto:vkostina@caltech.edu), [hassibi@caltech.edu](mailto:hassibi@caltech.edu)). This work was supported in part by the National Science Foundation (NSF) under grants CCF-1751356 and CCF-1817241. The work of Babak Hassibi was supported in part by the NSF under grants CNS-0932428, CCF-1018927, CCF-1423663 and CCF-1409204, by a grant from Qualcomm Inc., by NASA's Jet Propulsion Laboratory through the President and Director's Fund, and by King Abdullah University of Science and Technology. A part of this work was accepted for presentation at ISIT 2020 [1].

Encoding and decoding operations leverage memory of the past but cannot look in the future. In this causal setting no delay is allowed in producing  $\hat{X}_i$ .

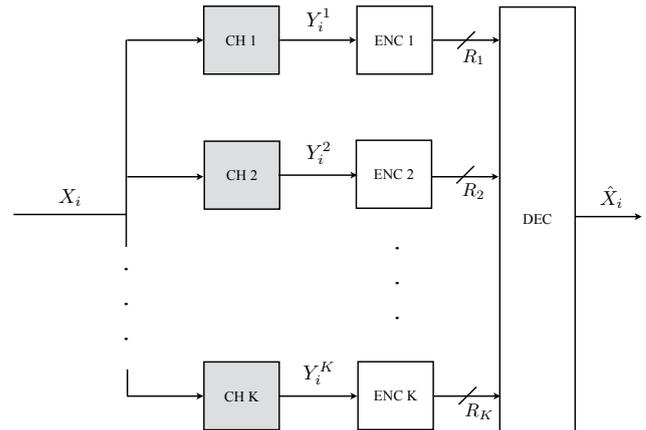


Fig. 1. The causal CEO problem.

In the classical setting with  $t = 1$ , the CEO problem was first introduced by Berger et al. [2] for a finite alphabet source. In the classical Gaussian CEO problem, a Gaussian source is observed via Gaussian channels and reproduced under mean-square error (MSE) distortion. The Gaussian CEO problem was studied by Viswanathan and Berger [3], who proved an achievability bound on the rate-distortion dimension for the case of  $K$  identical Gaussian channels, by Oohama [4], who derived the sum-rate rate-distortion region for that special case, by Prabhakaran et al. [5], who determined the full Gaussian CEO rate region, by Chen et al. [6], who proved that the minimum sum rate is achieved via waterfilling, by Behroozi and Soleymani [7] and by Chen and Berger [8], who showed rate-optimal successive coding schemes. Wagner et al. [9] found the rate region of the distributed Gaussian lossy compression problem by coupling it to the Gaussian CEO problem. Wagner and Anantharam [10] showed an outer bound to the rate region of the multiterminal source coding problem that is tighter than the Berger-Tung outer bound [11], [12]. Wang et al. [13] showed a simple converse on the sum rate of the vector Gaussian CEO problem. Concurrently, Ekrem and Ulukus [14] and Wang and Chen [15] showed an outer bound to the rate region of the vector Gaussian CEO problem that is tight in some cases and not tight in others and that particularizes the outer bound in [10] to the Gaussian case. Courtade and Weissman [16] determined the distortion region of the distributed source coding and the CEO problem under logarithmic loss.

None of the above results directly apply to the causal tracking problem in Fig. 1 because of the causality constraint in encoding the observations and in producing  $\hat{X}_i$  in (1).

The most basic scenario of source coding with causality constraints is that of a single observer directly seeing the information source [17]. The causal rate-distortion function for the Gauss-Markov source was computed by Gorbunov and Pinsker [18]. The link between the minimum attainable linear quadratic Gaussian (LQG) control cost and the causal rate-distortion function is elucidated in [19]–[21]. A semidefinite program to compute the causal rate-distortion function for vector Gauss-Markov sources is provided in [22]. The remote Gaussian causal rate-distortion function, which corresponds to setting  $K = 1$  in Fig. 1, is computed in [21]. The causal rate-distortion function of the Gauss-Markov source with Gaussian side observation available at decoder (the causal counterpart of the Wyner-Ziv setting) is computed in [23]. That causal Wyner-Ziv setting can be viewed a special case of our causal CEO problem (2), (3) with two observers, with the second observer enjoying an infinite rate. Stability of linear Gaussian systems with multiple isolated observers is investigated in [24].

The first contribution of this paper is a characterization the minimum asymptotically achievable (as  $n \rightarrow \infty$ ) sum rate  $R_1 + \dots + R_K$  required to achieve a given average distortion (1) in the causal distributed tracking setting of Fig. 1. Provided that the components of each  $X_i \in \mathcal{A}^n$  are i.i.d. ( $X_i$  can still depend on  $X_1, \dots, X_{i-1}$ ), the channels act on each of those components independently, and the distortion measure is separable, that minimum sum rate is equal to the directed mutual information from the observers to the decoder minimized subject to the distortion constraint and the separate encoding constraint. The converse to this coding theorem follows via standard data processing and single-letterization arguments. To prove the achievability, we show a nonasymptotic bound for causal distributed lossy source coding that can be viewed as an extension of the nonasymptotic Berger-Tung bound by Yassaee et al. [25], [26] to the setting with  $K > 2$  sources and  $t > 1$  time instances. We view the horizon- $t$  causal coding problem as a multiterminal coding problem in which at each step coded side information from past steps is available, and we use a stochastic likelihood coder (SLC) by Yassaee et al. [25], [26] to perform encoding operations. The SLC-based encoder mimics the operation of the joint typicality encoder while admitting sharp nonasymptotic bounds on its performance. Unfortunately, the SLC-based decoder of [25], [26] is ill-suited to the case  $K > 2$ . We propose a novel decoder that falls into the class of generalized likelihood decoders [27] and uses  $K$  different threshold tests depending on the point of the rate-distortion region the code is operating at. An asymptotic analysis of our nonasymptotic bound yields an extension of the Berger-Tung bound [11], [12] to the causal coding setting.

The second contribution of the paper is an explicit evaluation of the minimum sum rate for the causal Gaussian CEO problem. In that scenario, the source is an  $n$ -dimensional Gauss-Markov source,

$$X_{i+1} = aX_i + V_i, \quad (2)$$

$k$ -th observer sees

$$Y_i^k = X_i + W_i^k, \quad k = 1, \dots, K, \quad (3)$$

where  $X_1$  and  $\{V_i, W_i^1, W_i^2, \dots, W_i^K\}_{i=1}^T$  are independent Gaussian vectors of length  $n$ ;  $V_i \sim \mathcal{N}(0, \sigma_V^2 \mathbf{1})$ ;  $W_i^k \sim \mathcal{N}(0, \sigma_{W_k}^2 \mathbf{1})$ . The distortion measure is mean-square error (MSE)

$$d(X_i, \hat{X}_i) = \|X_i - \hat{X}_i\|^2. \quad (4)$$

We characterize the minimum sum rate as a convex optimization problem over  $K$  parameters; an explicit formula is given in the case of identical observation channels. To prove this result, we split up the directed mutual information problem whose operational meaning we established as the first contribution of this paper into a sum of easier-to-solve optimization problems. To tie the parameters of those optimization problems back to those of the original optimization problem, we extend the technique developed by Wang et al. [13] for the time horizon  $t = 1$ , to  $t > 1$ . That extension is nontrivial. A device that facilitates an understanding of how estimation errors behave over multiple time instances is the reversal of the channels from  $\{X_i\}$  to  $\{Y_i^k\}$ :

$$X_i = \bar{X}_i^k + W_i^{k'}, \quad (5)$$

where

$$\bar{X}_i^k \triangleq \mathbb{E}[X_i | Y_1^k, \dots, Y_i^k], \quad (6)$$

and  $W_i^{k'} \perp \bar{X}_i^k$  are Gaussian independent random vectors representing the errors in estimating  $X_i$  from  $\{Y_j^k\}_{j=1}^i$ . While for  $t = 1$ , it does not matter whether the encoders compress  $Y_1$  or  $\bar{X}_1$  since the latter is just a scaled version of the former, for  $t > 1$ , compressing  $Y_i$  instead of  $\bar{X}_i^k$  is only suboptimal.

The third contribution of the paper is a bound on the rate loss due to a lack of communication among the different encoders in the causal Gaussian CEO problem: as long as the target distortion is not too small, the rate loss is bounded above by  $K - 1$  times the difference between the remote and the direct rate-distortion functions. The bound is attained with equality if the observation channels are identical, indicating that among all possible observer channels with the same error in estimation  $\{X_i\}$  from  $\{Y_j^k\}_{j \leq i, k=1, \dots, K}$ , the identical channels case is the hardest to compress.

The rest of the paper is organized as follows. In Section II, we consider the general (non-Gaussian) causal CEO problem and prove direct and converse coding theorems to establish that the minimum sum rate is given by the directed mutual information minimized subject to a distortion constraint under separate encoding (Theorem 1). In Section III, we characterize the causal Gaussian CEO sum rate - distortion function (Theorem 4). In Section IV, we bound the rate loss due to isolated observers (Theorem 5).

*Notation:* Logarithms are natural base. For a natural number  $M$ ,  $[M] \triangleq \{1, \dots, M\}$ . Notation  $X \leftrightarrow Y$  reads “replace  $X$  by  $Y$ ”. We indicate the temporal index in the subscript and the spatial index in the superscript:  $Y_{[t]}^k$  is the temporal vector  $(Y_1^k, \dots, Y_t^k)$ ;  $Y_i^{[K]}$  is the spatial vector  $(Y_i^1, \dots, Y_i^K)$ ;  $Y_{[t]}^{[K]} \triangleq (Y_{[t]}^1, \dots, Y_{[t]}^K)$ .  $\mathcal{D}$  denotes delay by one, i.e.  $\mathcal{D}X_{[t]} = (0, X_1, \dots, X_{t-1})$ . We use the

following shorthand notation for causally conditional [28] probability kernels:

$$P_{Y_{[t]}|X_{[t]}} \triangleq \prod_{i=1}^t P_{Y_i|Y_{[i-1]}, X_{[i]}}. \quad (7)$$

For a random vector  $X$  with i.i.d. components,  $X$  denotes a random variable distributed the same as each component of  $X$ .

## II. SUM RATE DISTORTION FUNCTION VIA DIRECTED INFORMATION

### A. Overview

In Section II, we present and prove our main coding theorem that links the minimum achievable sum rate to a minimal directed mutual information problem. The theorem applies to an abstract source with abstract observations. The operational scenario and achievable rates are formally defined in Section II-B. The coding theorem is presented in Section II-C. Its converse is proven in Section II-D. A nonasymptotic achievability bound and its asymptotic analysis establishing the achievability direction of the coding theorem are presented in Section II-E.

### B. Operational problem setting

A causal CEO code is formally defined as follows.

**Definition 1** (Causal CEO code). *Consider a discrete-time random process  $\{X_i\}_{i=1}^t$  on  $\mathcal{X}$ , observed by  $K$  causal observers via the channels*

$$P_{Y_{[t]}^k|X_{[t]}}: \mathcal{X}^{\otimes t} \mapsto \mathcal{Y}^{\otimes t}, \quad k \in [K]. \quad (8)$$

Let  $d: \mathcal{X} \times \hat{\mathcal{X}} \mapsto \mathbb{R}_+$  be the distortion measure.

A causal CEO code consists of:

#### a) $K$ causal encoding policies

$$P_{B_{[t]}^k|Y_{[t]}^k}: \mathcal{Y}^{\otimes t} \mapsto \prod_{i=1}^t [M_i^k], \quad k \in [K], \quad (9)$$

#### b) a decoding policy

$$P_{\hat{X}_{[t]}^{[K]}|B_{[t]}^{[K]}}: \prod_{i=1}^t [M_i^k] \mapsto \hat{\mathcal{X}}^{\otimes t}. \quad (10)$$

If the encoding and decoding policies satisfy

$$\frac{1}{t} \sum_{i=1}^t \mathbb{E} \left[ d \left( X_i, \hat{X}_i \right) \right] \leq d, \quad (11)$$

we say that they form an  $(M_{[t]}^{[K]}, d)$  average distortion code.

If the encoding and decoding policies satisfy

$$\mathbb{P} \left[ \bigcup_{i=1}^t \left\{ d \left( X_i, \hat{X}_i \right) > d_i \right\} \right] \leq \epsilon, \quad (12)$$

we say that they form an  $(M_{[t]}^{[K]}, d_{[t]}, \epsilon)$  excess distortion code.

The probability measure in (11) and (12) is generated by the joint distribution  $P_{X_{[t]}} P_{Y_{[t]}^{[K]}|X_{[t]}} P_{\hat{X}_{[t]}^{[K]}|B_{[t]}^{[K]}} \prod_{k=1}^K P_{B_{[t]}^k|Y_{[t]}^k}$ .

A distortion measure  $d_n: \mathcal{A}^n \times \hat{\mathcal{A}}^n \mapsto \mathbb{R}_+$  is called *separable* if

$$d_n(x, \hat{x}) = \frac{1}{n} \sum_{i=1}^n d(x(i), \hat{x}(i)), \quad (13)$$

where  $d: \mathcal{A} \times \hat{\mathcal{A}} \mapsto \mathbb{R}_+$ , and  $x(i)$ ,  $\hat{x}(i)$  denote the  $i$ -th components of vectors  $x \in \mathcal{A}^n$  and  $\hat{x} \in \hat{\mathcal{A}}^n$ , respectively.

**Definition 2** (Operational sum rate - distortion function). *Consider a discrete-time random process  $\{X_i\}_{i=1}^t$  on  $\mathcal{X} = \mathcal{A}^n$  equipped with a separable distortion measure, observed by  $K$  causal observers via the channels (8).*

The rate-distortion tuple  $(R^{[K]}, d)$  is asymptotically achievable at time horizon  $t$  if for  $\forall \gamma > 0$ ,  $\exists n_0 \in \mathbb{N}$  such that  $\forall n \geq n_0$ , an  $(M_{[t]}^{[K]}, d + \gamma)$  average distortion causal CEO code exists, where

$$\frac{1}{nt} \sum_{i=1}^t \log M_i^k \leq R^k, \quad k \in [K]. \quad (14)$$

The sum rate - distortion pair  $(R, d)$  is asymptotically achievable if a rate-distortion tuple  $(R^{[K]}, d)$  with

$$\sum_{k=1}^K R^k \leq R \quad (15)$$

is asymptotically achievable.

The causal sum rate - distortion function at time horizon  $t$  is defined as follows:

$$R_{t \text{ CEO}}(d) \triangleq \inf \left\{ R: (R, d) \text{ is achievable at time horizon } t \text{ in the CEO problem.} \right\} \quad (16)$$

### C. Main coding theorem

Given a distribution  $P_{X_{[t]}}$  and a causal kernel  $P_{Y_{[t]}|X_{[t]}}$ , the directed mutual information is defined as [29]

$$I(X_{[t]} \rightarrow Y_{[t]}) \triangleq \sum_{i=1}^t I(X_{[i]}; Y_i | Y_{[i-1]}). \quad (17)$$

**Theorem 1** (Main coding theorem). *Consider a discrete-time random process  $\{X_i\}_{i=1}^t$  on  $\mathcal{X} = \mathcal{A}^n$  equipped with a separable distortion measure, observed by  $K$  causal observers via the channels (8) with  $\mathcal{Y} = \mathcal{B}^n$  and*

$$P_{X_i|X_{[i-1]}} = P_{X_i|X_{[i-1]}}^{\otimes n} \quad (18)$$

$$P_{Y_i^k|X_{[i]}, Y_{[i-1]}^k} = P_{Y_i^k|X_{[i]}, Y_{[i-1]}^k}^{\otimes n}. \quad (19)$$

Suppose further that for some  $p > 1$ , there exists a vector  $\hat{x}_{[t]}$  such that

$$\mathbb{E} \left[ \left( \frac{1}{t} \sum_{i=1}^t d(X_i, \hat{x}_i) \right)^p \right] \leq d_p < \infty. \quad (20)$$

The causal sum rate - distortion function is given by

$$R_{t \text{ CEO}}(d) = \inf_{\substack{P_{U_{[t]}^{[K]}|Y_{[t]}^{[K]}} = \prod_{k=1}^K P_{U_{[t]}^k|Y_{[t]}^k}, \\ P_{\hat{X}_{[t]}^{[K]}|U_{[t]}^{[K]}}: \\ \frac{1}{t} \sum_{i=1}^t \mathbb{E}[d(X_i, \hat{x}_i)] \leq d}} I(Y_{[t]}^{[K]} \rightarrow U_{[t]}^{[K]}). \quad (21)$$

Theorem 1 establishes the operational meaning of the minimal directed mutual information in (21). Note that  $R_{t\text{CEO}}(d)$  is a convex function of  $d$ . (Convexity can be proven in the standard way, using time sharing between kernels achieving different  $d_j$ 's and the convexity of mutual information in those kernels.)

#### D. Theorem 1: proof of converse

Converse proof uses standard techniques. We will use the following definition and lemma.

Causally conditioned directed information is defined as

$$I(X_{[t]} \rightarrow Y_{[t]} \| Z_{[t]}) \triangleq \sum_{i=1}^t I(X_{[i]}; Y_i | Y_{[i-1]}, Z_{[i]}). \quad (22)$$

**Lemma 1** ([28, (3.14)–(3.16)]). *Directed information chain rules:*

$$I((X_{[t]}, Y_{[t]}) \rightarrow Z_{[t]}) = I(X_{[t]} \rightarrow Z_{[t]}) + I(Y_{[t]} \rightarrow Z_{[t]} \| X_{[t]}), \quad (23)$$

$$I(X_{[t]} \rightarrow (Y_{[t]}, Z_{[t]})) = I(X_{[t]} \rightarrow Y_{[t]} \| \mathcal{D}Z_{[t]}) + I(X_{[t]} \rightarrow Z_{[t]} \| Y_{[t]}). \quad (24)$$

Fix an  $(M_{[t]}^{[K]}, d)$  code in Definition 1. Denote by  $B_i^k \in [M_i^k]$  the codeword sent by  $k$ -th encoder at time  $i$ . Since the codewords satisfy the sum rate constraint (15),

$$nR \geq \sum_{k=1}^K H(B_{[t]}^k) \quad (25)$$

$$\geq H(B_{[t]}^{[K]}) \quad (26)$$

$$\geq I(Y_{[t]}^{[K]} \rightarrow B_{[t]}^{[K]}) \quad (27)$$

$$\geq \inf_{\substack{P_{B_{[t]}^{[K]} \| Y_{[t]}^{[K]}} = \prod_{k=1}^K P_{B_{[t]}^k \| Y_{[t]}^k}, \\ P_{X_{[t]}^{[K]} \| B_{[t]}^{[K]}} : (11) \text{ holds}}} I(Y_{[t]}^{[K]} \rightarrow B_{[t]}^{[K]}), \quad (28)$$

where (26) holds because joint entropy is upper-bounded by the sum of individual entropies, and (27) holds because mutual information is upper-bounded by entropy.

We proceed to apply a standard single-letterization argument to (28). For an  $n$ -dimensional vector  $Y_i^k$ , we denote by  $Y_i^k(j)$  its  $j$ -th component; for sets  $\mathcal{K} \subseteq [K]$  and  $\mathcal{I} \subseteq [n]$ , we denote by  $Y_i^{\mathcal{K}}(\mathcal{I})$  the components of the vectors  $(Y_i^k : k \in \mathcal{K})$  indexed by  $\mathcal{I}$ .

Denote by  $R_{t\text{CEO}}(d)$  the right-hand side of (21). We introduce auxiliary random objects

$$U_i^k(j) = (B_i^k, Y_i^k([j-1])), \quad j \in [n] \quad (29)$$

The directed mutual information in the right side of (28) can be rewritten in terms of  $U_i^{[K]}$  and bounded as follows.

$$I(Y_{[t]}^{[K]} \rightarrow B_{[t]}^{[K]}) = \sum_{j=1}^n I(Y_{[t]}^{[K]}(j) \rightarrow B_{[t]}^{[K]} \| Y_{[t]}^{[K]}([j-1])) \quad (30)$$

$$= \sum_{j=1}^n I(Y_{[t]}^{[K]}(j) \rightarrow (B_{[t]}^{[K]}, Y_{[t]}^{[K]}([j-1]))) - I(Y_{[t]}^{[K]}(j) \rightarrow Y_{[t]}^{[K]}([j-1]) \| \mathcal{D}B_{[t]}^{[K]}) \quad (31)$$

$$= \sum_{j=1}^n I(Y_{[t]}^{[K]}(j) \rightarrow (B_{[t]}^{[K]}, Y_{[t]}^{[K]}([j-1]))) \quad (32)$$

$$= \sum_{j=1}^n I(Y_{[t]}^{[K]}(j) \rightarrow U_{[t]}^{[K]}(j)) \quad (33)$$

$$\geq \min_{\substack{d_j, j \in [n]: \\ \sum d_j \leq nd}} \sum_{j=1}^n R_{t\text{CEO}}(d_j) \quad (34)$$

$$\geq n R_{t\text{CEO}}(d) \quad (35)$$

where (30) is by the chain rule of mutual information; (31) is by the chain rule of directed information (24); (32) holds because  $P_{B_{[t]}^{[K]} | Y_{[t]}^{[K]}} = P_{B_{[t]}^{[K]} \| Y_{[t]}^{[K]}}$  is a causal kernel, which means that  $P_{Y_{[t]}^{[K]} \| \mathcal{D}B_{[t]}^{[K]}} = P_{Y_{[t]}^{[K]}}$ , hence conditioning on  $\mathcal{D}B_{[t]}^{[K]}$  in (31) can be eliminated, and the resulting directed information is zero because different components of the vector  $Y_i^k$  are independent due to (18), (19); (33) is by substituting (29); (34) holds because  $U_i^k(j)$  (29) satisfies the separate encoding constraint  $P_{U_{[t]}^{[K]}(j) | Y_{[t]}^{[K]}} = \prod_{k=1}^K P_{U_{[t]}^k(j) | Y_{[t]}^k}$ , the distortion measure is separable and (18), (19) hold; (35) is by the convexity of  $R_{t\text{CEO}}(d)$  as a function of  $d$ .  $\square$

#### E. Theorem 1: proof of achievability

To show that (21) is achievable in the asymptotics  $n \rightarrow \infty$ , we first show a nonasymptotic bound. Then, via an asymptotic analysis of the bound, we derive an extension of the Berger-Tung bound [11], [12] to the causal coding setting.

Before we present our nonasymptotic achievability bound in Theorem 2 below, we prepare some notation.

For a fixed conditional distribution  $P_{U_i^k Y_{[i]}^k | U_{[i-1]}^k}$ , denote the conditional information density

$$i(y_{[i]}^k; u_i^k | u_{[i-1]}^k) \triangleq \log \frac{dP_{U_i^k | Y_{[i]}^k, U_{[i-1]}^k}(u_i^k | y_{[i]}^k, u_{[i-1]}^k)}{dP_{U_i^k | U_{[i-1]}^k}(u_i^k | u_{[i-1]}^k)}. \quad (36)$$

For a fixed joint distribution  $P_{U_{[i]}^{[K]}}$ , denote the relative conditional information densities

$$j^k(u_{[i]}^{[K]}) \triangleq \log \frac{dP_{U_i^k | U_{[i]}^{[k-1]} U_{[i-1]}^{[K]}}(u_i^k | u_{[i]}^{[k-1]} u_{[i-1]}^{[K]})}{dP_{U_i^k | U_{[i-1]}^k}(u_i^k | u_{[i-1]}^k)}. \quad (37)$$

For a permutation  $\pi: [K] \mapsto [K]$ , we denote the ordered set

$$\pi(\mathcal{K}) \triangleq (\pi(k) : k \in \mathcal{K}). \quad (38)$$

**Theorem 2** (nonasymptotic causal Berger-Tung bound). *Fix  $P_{Y_{[t]}^{[K]}}$  and parameters  $M_{[t]}^{[K]}, d_{[t]}^{[K]}, \epsilon$ . For any scalars  $\alpha_i^k, \beta_i^k$ , any integers  $L_i^k \geq M_i^k$ ,  $i \in [t], k \in [K]$ , any causal kernels  $P_{U_{[t]}^{[K]} \| Y_{[t]}^{[K]}} = \prod_{k=1}^K P_{U_{[t]}^k \| Y_{[t]}^k}$  and  $P_{\hat{X}_{[t]}^{[K]} \| U_{[t]}^{[K]}}$ , and any permutation  $\pi: [K] \mapsto [K]$ , there exists an  $(M_{[t]}^{[K]}, d_{[t]}^{[K]}, \epsilon)$  excess distortion causal CEO code with*

$$\epsilon \leq \mathbb{P}[\mathcal{E}] + \gamma, \quad (39)$$

where event  $\mathcal{E}$  is given by

$$\begin{aligned} \mathcal{E} \triangleq & \bigcup_{i=1}^t \left\{ d(X_i, \hat{X}_i^k) > d_i \right\} \\ & \bigcup_{i=1}^t \bigcup_{k=1}^K \left\{ i(Y_{[i]}^k; U_i^k | U_{[i-1]}^k) > \log L_i^k - \alpha_i^k \right\} \\ & \bigcup_{i=1}^t \bigcup_{k=1}^K \left\{ j^{\pi(k)}(u_{[i]}^{\pi(k)}) < \log \frac{L_i^{\pi(k)}}{M_i^{\pi(k)}} + \beta_i^{\pi(k)} \right\}, \end{aligned} \quad (40)$$

and constant  $\gamma$  is given by

$$\gamma \triangleq 1 - \frac{1}{\prod_{i=1}^t \left[ \sum_{\mathcal{K} \subseteq K} \exp(-\sum_{k \in \mathcal{K}} \beta_i^k) \right] \prod_{k=1}^K [1 + \exp(-\alpha_i^k)]}. \quad (41)$$

*Proof.* Appendix A.  $\square$

**Theorem 3** (causal Berger-Tung inner bound). *Under the assumptions of Theorem 1, the rate-distortion tuple  $(R_{[K]}, d)$  is asymptotically achievable at time horizon  $t$  if for some single-letter causal kernels  $\prod_{k=1}^K P_{U_{[t]}^k \| Y_{[t]}^k}$ ,  $P_{\hat{X}_{[t]}^{[K]} \| U_{[t]}^{[K]}}$  and some permutation  $\pi: [K] \mapsto [K]$ , it holds that*

$$\frac{1}{t} \sum_{i=1}^t \mathbb{E} \left[ d(X_i, \hat{X}_i) \right] \leq d \quad (42)$$

and for all  $k \in [K]$

$$R_{\pi(k)} > I(Y_{[t]}^{\pi(k)} \rightarrow U_{[t]}^{\pi(k)} \| U_{[t]}^{\pi(k-1)}), \mathcal{D}U_{[t]}^{[K]}. \quad (43)$$

*Proof.* Appendix B.  $\square$

We conclude Section II-E with a set of remarks. In particular, Remark 1 below completes the proof of achievability of the sum rate in Theorem 1.

1. Theorem 3 implies that the sum rate

$$\sum_{k=1}^K R_k > I(Y_{[t]}^{[K]} \rightarrow U_{[t]}^{[K]}) \quad (44)$$

is achievable. Indeed, summing (43) over  $k$  and using  $U_i^k - (Y_{[i]}^k, U_{[i-1]}^k) - U_{[i]}^{[K] \setminus \{k\}}$  leads to (44).

2. Theorems 2 and 3 are easily extended to causal distributed source coding, where the goal is to separately

compress (and jointly decompress)  $K$  processes  $\{Y_i^k\}$  under the individual distortion constraints

$$\frac{1}{t} \sum_{i=1}^t \mathbb{E} \left[ d^k(Y_i^k, \hat{Y}_i^k) \right] \leq d^k, \quad k \in [K]. \quad (45)$$

Theorem 2 continues to hold with  $d(X_i, \hat{X}_i^k) > d_i$  in (40) replaced by  $d^k(Y_i^k, \hat{Y}_i^k) > d_i^k$ . Consequently, Theorem 3 also continues to hold, replacing the constraint in (42) by

$$\frac{1}{t} \sum_{i=1}^t \mathbb{E} \left[ d^k(Y_i^k, \hat{Y}_i^k) \right] \leq d^k, \quad k \in [K]. \quad (46)$$

3. While the sum rate bound in (44) is the same regardless of the choice of permutation  $\pi$ , different  $\pi$ 's correspond to different orders in which the chain rule of mutual information can be applied, and are needed to specify the full achievable region of rates and distortions.

4. Case  $t = 1$  corresponds to the classical CEO / distributed source coding problems. The region in (43) simplifies to

$$\begin{aligned} R^{\pi(k)} &> I(Y^{\pi(k)}; U^{\pi(k)} | U^{\pi(k-1)}), \\ &\forall k \in [K], \forall \text{permutation } \pi: [K] \mapsto [K]. \end{aligned} \quad (47)$$

The multiterminal Berger-Tung region is usually (e.g. [16, Def. 7], [5, eq. (2)]) specified as

$$\sum_{k \in \mathcal{A}} R_k > I(Y^{\mathcal{A}}; U^{\mathcal{A}} | U^{\mathcal{A}^c}), \quad \forall \mathcal{A} \subseteq [K]. \quad (48)$$

These characterizations are equivalent (Appendix C).

5. To prove Theorem 2, we employ the achievability proof technique developed by Yassaee et al. [25], [26] that uses a stochastic likelihood coder (SLC) to perform encoding operations. An SLC makes a randomized decision that coincides with high probability with the choice that a maximum likelihood (ML) coder would make (in fact, the error probability of the SLC exceeds by at most a factor of 2 the error probability of the ML coder [30, Th. 7]). We view the horizon- $t$  causal coding problem as a multiterminal coding problem in which at each step coded side information from past steps is available, and we define the SLC based on the auxiliary transition probability kernel  $P_{U_i^k | Y_{[i]}^k, U_{[i-1]}^k}$  (see (131) in Appendix A below) that is also used to generate random codebooks.

6. While [26, Th. 6] shows a sharp nonasymptotic bound for the classical distributed source coding problem with  $K = 2$  terminals, the decoder employed there does not extend to the case  $K > 2$ . In (135) in Appendix A below, we propose a novel decoder that falls into the class of generalized likelihood decoders (GLD) conceptualized by Merhav [27, eq. (4)] and uses an auxiliary indicator function  $g(u_{[t]}^{[K]})$  (136). With our GLD we are able to recover the full Berger-Tung region (48) for any  $K$ . One can view the set of outcomes  $u_{[t]}^{[K]}$  for which  $g(u_{[t]}^{[K]}) = 1$  as a jointly typical set. That set depends on the choice of  $\pi$  and thus on the particular rate point that the code is operating at. Checking for membership in that set involves  $K$  threshold tests. In contrast, the jointly typical

set defined by Oohama [4, eq. (46)] involves  $2^K - 1$  threshold tests, one for each nonempty subset of  $[K]$ .

### III. GAUSSIAN SUM RATE - DISTORTION FUNCTION

#### A. Problem setup

In Section III, we focus on the scenario of the Gauss-Markov source in (2) observed through the Gaussian channels in (3) under MSE distortion (4). Given an encoding policy in Definition 1, the optimal decoding policy  $P_{\hat{X}_{[t]}^{[K]} \| B_{[t]}^{[K]}}$  that achieves the minimum of  $\mathbb{E} [\|X_i - \hat{X}_i\|^2]$  is  $\hat{X}_i = \mathbb{E} [X_i | B_{[i]}^{[K]}]$ .

For simplicity we focus on the infinite time-horizon limit.

$$R_{\text{CEO}}(d) \triangleq \limsup_{t \rightarrow \infty} R_{t, \text{CEO}}(d). \quad (49)$$

In other words,  $R_{\text{CEO}}(d)$  is the infimum of  $R$ 's such that  $\forall \gamma > 0, \exists t_0 \geq 0$  such that  $\forall t \geq t_0, \exists n_0 \in \mathbb{N}$  such that  $\forall n \geq n_0$ , an  $(M_{[t]}^{[K]}, d + \gamma)$  average distortion causal CEO code exists with  $M_{[t]}^{[K]}$  satisfying (14) and (15).

Taking the limit  $t \rightarrow \infty$  simplifies the solution of many minimal directed mutual information problems ([21, Th. 9], [23, Th. 6, Th. 7], [31, Th. 1], [32, Th. 2]) by eliminating the transient effects due to the starting location  $X_1$  of the process  $\{X_i\}$  that is being transmitted. In this steady state regime, the optimal rate allocation across time is uniform (i.e.,  $\log M_1^k = \dots = \log M_t^k$  in (14)). Furthermore,  $R_{t, \text{CEO}}(d)$  approaches its steady-state value (49) as  $O(\frac{1}{t})$  (this is a consequence of [23, eq. (83)-(85), (92)] and (79), (85), (92) below).

In Section III-B, we present the Gaussian sum rate - distortion function as a convex optimization problem over  $K$  parameters (Theorem 4), which reduces to an explicit formula in the identical-channels case (Corollary 1). These results are obtained by evaluating the minimal directed mutual information in Theorem 1 in the Gaussian case. In Section III-C, we give auxiliary estimation lemmas that are useful in the proof of Theorem 4, and we give the proof of Theorem 4 in Section III-D.

*Notation:* For a random process  $\{X_i\}$  on  $\mathbb{R}$ , its stationary variance (can be  $+\infty$ ) is denoted by

$$\sigma_X^2 \triangleq \limsup_{i \rightarrow \infty} \frac{1}{n} \mathbb{E} [X_i^2]. \quad (50)$$

The minimum MSE (MMSE) in the estimation of  $X_i$  from  $Y_{[i]}^{[K]}$  is denoted by

$$\sigma_{X_i | Y_{[i]}^{[K]}}^2 \triangleq \mathbb{E} \left[ \left( X - \mathbb{E} [X_i | Y_{[i]}^{[K]}] \right)^2 \right], \quad (51)$$

and the steady-state causal MMSE by

$$\sigma_{X \| Y^{[K]}}^2 \triangleq \limsup_{i \rightarrow \infty} \sigma_{X_i | Y_{[i]}^{[K]}}^2. \quad (52)$$

#### B. Gaussian sum rate - distortion function

In Theorem 4, the causal sum rate - distortion function is expressed as a convex optimization problem over parameters  $\{d_k\}_{k=1}^K$  that determine the individual rates of the transmitters and that correspond to the MSE achievable at the decoder in estimation of  $\{X_i\}_{i=1}^t$  provided that it correctly decoded the codewords from  $k$ -th transmitter.

**Theorem 4.** For all  $\sigma_{X \| Y^{[K]}}^2 < d < \sigma_X^2$ , the causal sum rate - distortion function (49) for the Gauss-Markov source in (2) observed through the Gaussian channels in (3) is given by

$$R_{\text{CEO}}(d) = \frac{1}{2} \log \frac{\bar{d}}{d} + \min_{\{d_k\}_{k=1}^K} \sum_{k=1}^K \frac{1}{2} \log \frac{\bar{d}_k - \sigma_{X \| Y^k}^2 d_k}{d_k - \sigma_{X \| Y^k}^2 \bar{d}_k}, \quad (53)$$

where

$$\bar{d} \triangleq a^2 d + \sigma_V^2, \quad (54)$$

$$\bar{d}_k \triangleq a^2 d_k + \sigma_V^2, \quad (55)$$

and the minimum is over  $d_k, k \in [K]$ , that satisfy

$$\frac{1}{d} \leq \frac{1}{\sigma_{X \| Y^{[K]}}^2} - \sum_{k=1}^K \left( \frac{1}{\sigma_{X \| Y^k}^2} - \frac{1}{d_k} \right), \quad (56)$$

$$\sigma_{X \| Y^k}^2 \leq d_k \leq \sigma_X^2, \quad (57)$$

*Proof.* Section III-D.  $\square$

If the source is observed directly by one or more of the encoders, say if  $\sigma_{X \| Y^1}^2 = 0$ , then  $d_1 = d, d_2 = \dots = d_K = \sigma_X^2$  is optimal, and (53) reduces to the causal rate-distortion function [18, eq. (1.43)] (and e.g. [19], [33, Th. 3], [21, (64)], [23, Th. 6]):

$$R(d) = \frac{1}{2} \log \frac{\bar{d}}{d}. \quad (58)$$

The sum over  $k \in [K]$  in (53) is thus the penalty due to the encoders not observing the source directly and not communicating with each other.

If the observation channels satisfy

$$\sigma_{X \| Y^1}^2 = \dots = \sigma_{X \| Y^K}^2, \quad (59)$$

we can explicitly write the sum rate - distortion function  $R_{\text{CEO}}^{K\text{-sym}}(d)$  for this symmetrical scenario.

**Corollary 1.** If in the scenario of Theorem 4 the observation channels satisfy (59), the causal sum rate - distortion function (49) is given by

$$R_{\text{CEO}}^{K\text{-sym}}(d) = \frac{1}{2} \log \frac{\bar{d}}{d} + \frac{K}{2} \log \frac{\bar{d}_1 - \sigma_{X \| Y^1}^2 d_1}{d_1 - \sigma_{X \| Y^1}^2 \bar{d}_1}, \quad (60)$$

where  $d_1$  satisfies

$$\frac{1}{d} = \frac{1}{\sigma_{X \| Y^{[K]}}^2} - \frac{K}{\sigma_{X \| Y^1}^2} + \frac{K}{d_1}. \quad (61)$$

*Proof.* It suffices to show that the minimum in (53) is attained by  $d_1 = \dots = d_K$ . Since each of the terms in the sum in (53) is a convex function of  $d_k$ , applying Jensen's inequality concludes the proof.  $\square$

Think now of adding identical observers by letting  $K \rightarrow \infty$  in (59). Since  $\sigma_{X \| Y^{[K]}}^2 \rightarrow 0$ , had the observers communicated with each other, they could recover the source exactly, and they could operate at sum rate (58) in the limit. As the following result demonstrates,  $\lim_{K \rightarrow \infty} R_{\text{CEO}}^{K\text{-sym}}(d)$  is actually strictly greater than (58), thus a nonvanishing

penalty due to separate encoding is present in this regime. See Section IV for a more thorough discussion on the loss due to separate encoding.

**Corollary 2.**

$$\lim_{K \rightarrow \infty} R_{\text{CEO}}^{K\text{-sym}}(d) = \frac{1}{2} \log \frac{\bar{d}}{d} + \frac{1}{2} \frac{\frac{1}{\bar{d}} - \frac{1}{d}}{\frac{1}{\sigma_{X|Y^1}^2} - \frac{1}{\sigma_X^2}}. \quad (62)$$

*Proof.* By Lemma 3 in Section III-C below,

$$\frac{1}{\sigma_{X|Y^{[K]}}^2} = \frac{K}{\sigma_{X|Y^1}^2} - \frac{K-1}{\sigma_X^2}. \quad (63)$$

Eliminating  $d_1$  and  $\sigma_{X|Y^{[K]}}^2$  from (60) using (61) and (63), one readily verifies that

$$R_{\text{CEO}}^{K\text{-sym}}(d) - \frac{1}{2} \log \frac{\bar{d}}{d} = \frac{1}{2} \frac{\frac{1}{\bar{d}} - \frac{1}{d}}{\frac{1}{\sigma_{X|Y^1}^2} - \frac{1}{\sigma_X^2}} + O\left(\frac{1}{K}\right), \quad (64)$$

and (62) follows.  $\square$

Corollary 2 extends the result of Oohama [4, Cor. 1] to causal compression, and recovers it if  $a = 0$ .

Considering a scenario where the encoders and the decoder do not keep any memory of past observations and codewords, we may invoke the results on the classical Gaussian CEO problem in [5], [6] to express the minimum achievable sum rate as

$$R_{\text{CEO}}^{\text{no memory}}(d) = \frac{1}{2} \log \frac{\sigma_X^2}{d} + \min_{\{d_k\}_{k=1}^K} \sum_{k=1}^K \frac{1}{2} \log \frac{\sigma_X^2 - \sigma_{X|Y^k}^2}{d_k - \sigma_{X|Y^k}^2} \frac{d_k}{\sigma_X^2}, \quad (65)$$

where the minimum is over

$$\frac{1}{d} \leq \frac{1}{\sigma_{X|Y^{[K]}}^2} - \sum_{k=1}^K \left( \frac{1}{\sigma_{X|Y^k}^2} - \frac{1}{d_k} \right), \quad (66)$$

$$\sigma_{X|Y^k}^2 \leq d_k \leq \sigma_X^2. \quad (67)$$

Here  $\sigma_{X|Y^k}^2 \triangleq \lim_{i \rightarrow \infty} \sigma_{X_i|Y_i^k}$  and  $\sigma_{X|Y^{[K]}}^2 \triangleq \lim_{i \rightarrow \infty} \sigma_{X_i|Y_i^{[K]}}^2$  denote the stationary MMSE achievable in the estimation of  $X_i$  from  $Y_i^k$  and  $Y_i^{[K]}$  respectively, i.e., without memory of the past.

If  $a = 0$ , the observed process (2) becomes a stationary memoryless Gaussian process, the predictive MMSEs reduce to the variance of  $X_i$ :  $\bar{d} = \bar{d}_k = \sigma_X^2 = \sigma_Y^2$ ; similarly,  $\sigma_{X|Y^k}^2 = \sigma_{X||Y^k}^2$  and  $\sigma_{X|Y^{[K]}}^2 = \sigma_{X||Y^{[K]}}^2$ , and the result of Theorem 4 coincides with the classical Gaussian CEO sum rate - distortion function (65). This shows that if the source is memoryless, asymptotically there is no benefit in keeping the memory of previously encoded estimates as permitted by Definition 1. Classical codes that forget the past after encoding the current block of length  $n$  perform just as well.

If  $|a| > 1$ , the benefit due to memory is infinite: indeed, since the source is unstable,  $\sigma_X^2 = \infty$ , while  $\bar{d} < \infty$ . If  $|a| < 1$ , that benefit is finite and is characterized by the discrepancy between the stationary variance of the process  $\{X_i\}_{i=1}^{\infty}$   $\sigma_X^2 = \frac{\sigma_Y^2}{1-a^2}$  and the steady-state predictive MMSE  $\bar{d} < \sigma_X^2$ , as well as that between  $\sigma_{X|Y^k}^2$  and  $\sigma_{X||Y^k}^2$ .

### C. MMSE estimation lemmas

We record two elementary estimation lemmas that will be instrumental in the proof of Theorem 4.

**Lemma 2.** Let  $X \sim \mathcal{N}(0, \sigma_X^2)$ ,  $W \sim \mathcal{N}(0, \sigma_W^2)$ ,  $W \perp X$ , and let

$$Y = X + W. \quad (68)$$

Then,

$$\sigma_{X|Y}^2 = \sigma_X^2 \left( 1 - \frac{\sigma_X^2}{\sigma_Y^2} \right) \quad (69)$$

*Proof.* Appendix D.  $\square$

**Lemma 3.** Let  $\bar{X}_k$  and  $W'_k$  be Gaussian random variables,  $\{\bar{X}_k\}_{k=1}^K \perp \{W'_j\}_{j=1}^K$ , such that  $W'_k \perp W'_j$ ,  $j \neq k$ , and

$$X = \bar{X}_k + W'_k. \quad (70)$$

Then, the MMSE estimate and the estimation error  $\sigma_{W'}^2 \triangleq \sigma_{X|\bar{X}_{[K]}}^2$  of  $X$  given the vector  $\bar{X}_{[K]}$  are given by

$$\mathbb{E}[X|\bar{X}_{[K]}] = \sum_{k=1}^K \frac{\sigma_{W'}^2}{\sigma_{W'_k}^2} \bar{X}_k, \quad (71)$$

$$\frac{1}{\sigma_{W'}^2} = \sum_{k=1}^K \frac{1}{\sigma_{W'_k}^2} - \frac{K-1}{\sigma_X^2} \quad (72)$$

*Proof.* Appendix D.  $\square$

Lemma 3 converts the ‘‘forward channels’’ from  $X$  to observations  $Y_k$

$$Y_k = X + W_k, \quad k = 1, \dots, K, \quad (73)$$

where  $W_k \sim \mathcal{N}(0, \sigma_{W_k}^2)$ ,  $W_k \perp W_j$ ,  $j \neq k$ , into ‘‘backward channels’’ from estimates  $\bar{X}_k$  to  $X$  (70). While both representations are equivalent, (70) is more convenient to work with. Backward channel representations find widespread use in rate-distortion theory [34].

### D. Proof of Theorem 4

1) *Proof overview:* We break up the minimal directed mutual information problem in Theorem 1 into subproblems, and we use the tools we developed in [23] to evaluate the causal rate-distortion functions for each subproblem. As it turns out, the additive white Gaussian kernels in (94) below attain the minimal directed mutual information in Theorem 1. To link the parameters of the subproblems together to obtain the solution of the original problem, we extend the proof technique by Wang et al. [13], developed for the case  $t = 1$ , to  $t > 1$ . Converting the ‘‘forward channels’’ from  $X_{[t]}$  to observations  $Y_{[t]}^k$  into the ‘‘backward channels’’ from MMSE estimates  $\bar{X}_{[t]}^k$  to  $X_{[t]}$  and applying the lemmas in Section III-C above are key to that extension.

2) *Decoupling the problem into  $K$  subproblems:* Denote the MMSE estimate of  $X_i$  given  $Y_{[t]}^k$

$$\bar{X}_i^k \triangleq \mathbb{E} \left[ X_i | Y_{[t]}^k \right]. \quad (74)$$

We expand the right-hand side of (21) in Theorem 1 as

$$R_{t \text{ CEO}}(d) = \inf I \left( Y_{[t]}^{[K]} \rightarrow U_{[t]}^{[K]} \right) \quad (75)$$

$$\geq \inf I \left( \bar{X}_{[t]}^{[K]} \rightarrow U_{[t]}^{[K]} \right) \quad (76)$$

$$= \inf I \left( X_{[t]}, \bar{X}_{[t]}^{[K]} \rightarrow U_{[t]}^{[K]} \right) \quad (77)$$

$$= \inf \left\{ I \left( X_{[t]} \rightarrow U_{[t]}^{[K]} \right) + I \left( \bar{X}_{[t]}^{[K]} \rightarrow U_{[t]}^{[K]} \| X_{[t]} \right) \right\} \quad (78)$$

$$= \inf \left\{ I \left( X_{[t]} \rightarrow U_{[t]}^{[K]} \right) + \sum_{k=1}^K I \left( \bar{X}_{[t]}^k \rightarrow U_{[t]}^k \| X_{[t]} \right) \right\} \quad (79)$$

where

- (75) is by Theorem 1;
- (76) holds by the chain rule (23) using  $I \left( \bar{X}_{[t]}^{[K]} \rightarrow U_{[t]}^{[K]} \| Y_{[t]}^{[K]} \right) = 0$ .<sup>1</sup> The infimum is over single-letter causal kernels  $P_{U_{[t]}^{[K]} \| \bar{X}_{[t]}^{[K]}}$  satisfying both the separate encoding constraint

$$P_{U_{[t]}^{[K]} \| \bar{X}_{[t]}^{[K]}} = \prod_{k=1}^K P_{U_{[t]}^k \| \bar{X}_{[t]}^k}, \quad (80)$$

and the distortion constraint

$$\frac{1}{t} \sum_{i=1}^t \mathbb{E} \left[ (X_i - \hat{X}_i)^2 \right] \leq d, \quad (81)$$

where

$$\hat{X}_i = \mathbb{E} \left[ X_i | U_{[i]}^{[K]} \right] \quad (82)$$

is the MMSE estimate of  $X_i$  given  $U_{[i]}^{[K]}$ ;

- (77) is due to the chain rule of directed information (23), and  $I \left( X_{[t]} \rightarrow U_{[t]}^{[K]} \| \bar{X}_{[t]}^{[K]} \right) = 0$ ;
- (78) is by the chain rule of directed information (23);
- (79) is due to (80) and  $P_{U_{[t]}^{[K]} \| X_{[t]}, \mathcal{D}U_{[t]}^{[K]}} = \prod_{k=1}^K P_{U_{[t]}^k \| X_{[t]}, \mathcal{D}U_{[t]}^k}$ .

3) *Using causal rate-distortion functions to evaluate the terms in (79):* We lower-bound the first term in (79) using a classical result on the point-to-point causal Gaussian rate-

<sup>1</sup>In fact, equality holds in (76). For our setting of scalar  $Y_i^k$  one can see this immediately because  $\{Y_i^k\}_{i=1}^t$  and  $\{\bar{X}_i^k\}_{i=1}^t$  generate the same filtration. Our proof generalizes verbatim to the case of scalar  $X_i$  and vector  $Y_i^k$ . In that case, equality in (76) is assured by the form of the optimal kernels (94).

distortion function [18, eq. (1.43)] (see [23, Th. 6] for a modern exposition) as

$$\lim_{t \rightarrow \infty} \inf_{(80):} I \left( X_{[t]} \rightarrow U_{[t]}^{[K]} \right) \quad (83)$$

$$\geq \lim_{t \rightarrow \infty} \inf_{P_{\hat{X}_{[t]}^{[K]} \| X_{[t]}}:} I \left( X_{[t]} \rightarrow \hat{X}_{[t]}^{[K]} \right) \quad (84)$$

$$= \frac{1}{2} \log \frac{\bar{d}}{d}, \quad (85)$$

where  $\bar{d}$  is uniquely determined by  $d$  via (54). Furthermore, (85) is achieved by the Gaussian kernel  $P_{\hat{X}_{[t]}^{[K]} \| X_{[t]}}$  such that

$$X_i = \hat{X}_i^* + Z_i', \quad Z_i' \sim \mathcal{N}(0, d), \quad (86)$$

$\{Z_i'\}$  are i.i.d. and independent of  $\{\hat{X}_i^*\}$ , and

$$d = \sigma_{X \| \hat{X}^*}^2 \quad (87)$$

$$\bar{d} = \sigma_{X \| \mathcal{D}\hat{X}^*}^2. \quad (88)$$

For each of the remaining  $K$  terms in (79), note that  $\{\bar{X}_i^k\}$  is a Gauss-Markov process

$$\bar{X}_{i+1}^k = a\bar{X}_i^k + \bar{V}_i^k, \quad (89)$$

where  $\bar{V}_i^k \sim \mathcal{N} \left( 0, \sigma_{\bar{X}_i^k | Y_{[i]}^k}^2 - \sigma_{\bar{X}_i^k | Y_{[i-1]}^k}^2 \right)$ . The process  $\{X_i\}$  can be expressed through  $\{\bar{X}_i^k\}$  as

$$X_i = \bar{X}_i^k + W_i^{k'}, \quad (90)$$

where  $W_i^{k'}$  are independent,  $W_i^{k'} \sim \mathcal{N} \left( 0, \sigma_{X_i | Y_{[i]}^k}^2 \right)$ , and  $W_i^{k'} \perp X_i^k$ . Thus, we may apply the result [23, Th. 7] on the causal counterpart of Gaussian Wyner-Ziv rate-distortion function to the process  $\{\bar{X}_i^k\}$  (89) with side information  $\{X_i\}$  (90) to write

$$\lim_{t \rightarrow \infty} \inf_{P_{U_{[t]}^k \| \bar{X}_{[t]}^k}:} I \left( \bar{X}_{[t]}^k \rightarrow U_{[t]}^k \| X_{[t]} \right) \quad (91)$$

$$= \frac{1}{2} \log \frac{\bar{\rho}_k}{\rho_k}, \quad (92)$$

where  $\bar{\rho}_k$  is uniquely determined by  $\rho_k$  via

$$\frac{1}{\bar{\rho}_k} = \frac{1}{\sigma_{W^{k'}}^2} + \frac{1}{a^2 \rho_k + \sigma_V^2}. \quad (93)$$

Furthermore, (92) is attained by the Gaussian kernel  $P_{U_{[t]}^{k*} \| \bar{X}_{[t]}^k}$

$$U_{[t]}^{k*} = \bar{X}_{[t]}^k + Z_{[t]}^k, \quad Z_i \sim \mathcal{N}(0, \sigma_{Z_k}^2), \quad (94)$$

$\{Z_i\}$  are i.i.d. and independent of  $\{\bar{X}_i^k\}$ , and

$$\rho_k = \sigma_{X^k \| X, U^{k*}}^2, \quad (95)$$

$$\bar{\rho}_k = \sigma_{X^k \| X, \mathcal{D}U^{k*}}^2. \quad (96)$$

The variances  $\sigma_{Z_k}^2$  in (94) are set to satisfy (95).

4) *Achievability*: Using  $\mathbf{X}_{[i-1]} - (\mathbf{X}_i, \mathbf{U}_{[i-1]}^{[K]*}) - \mathbf{U}_i^{[K]*}$ , we note that

$$I(\mathbf{X}_{[i]}; \mathbf{U}_i^{[K]*} | \mathbf{U}_{[i-1]}^{[K]*}) = I(\mathbf{X}_i; \mathbf{U}_i^{[K]*} | \mathbf{U}_{[i-1]}^{[K]*}) \quad (97)$$

$$= \frac{1}{2} \log \frac{\sigma_{\mathbf{X}_i | \mathbf{U}_{[i-1]}^{[K]*}}^2}{\sigma_{\mathbf{X}_i | \mathbf{U}_{[i]}^{[K]*}}^2}, \quad (98)$$

and equality in (84) is attained by setting

$$\hat{\mathbf{X}}_i^* = \mathbb{E}[\mathbf{X}_i | \mathbf{U}_i^{[K]*}] \quad (99)$$

in (86). We conclude that the Gaussian kernels in (94) actually attain the  $\lim_{t \rightarrow \infty}$  of the infimum in (79).

5) *Linking*  $\{\rho_k\}_{k=1}^K$  to  $d$ : It remains to establish the connection between  $\{\rho_k\}_{k=1}^K$  (95) and  $d$  (87). Putting together (79), (85) and (92), we have

$$R_{\text{CEO}}(d) = \inf_{\left\{ \frac{\sigma_{\bar{\mathbf{X}}^k \| \mathbf{X}, \mathbf{U}^{k*}}^2}{\sigma_{\mathbf{X} \| \mathbf{U}^{[K]*}}^2} \right\}_{k=1}^K : \frac{\sigma_{\bar{\mathbf{X}}^k \| \mathbf{X}, \mathbf{U}^{k*}}^2}{\sigma_{\mathbf{X} \| \mathbf{U}^{[K]*}}^2} = d \right\}} \left\{ \frac{1}{2} \log \frac{\sigma_{\mathbf{X} \| \mathcal{D}\mathbf{U}^{[K]*}}^2}{\sigma_{\mathbf{X} \| \mathbf{U}^{[K]*}}^2} + \sum_{k=1}^K \frac{1}{2} \log \frac{\sigma_{\bar{\mathbf{X}}^k \| \mathbf{X}, \mathcal{D}\mathbf{U}^{k*}}^2}{\sigma_{\bar{\mathbf{X}}^k \| \mathbf{X}, \mathbf{U}^{k*}}^2} \right\} \quad (100)$$

Invoking Lemma 3 with  $X \leftrightarrow \mathbf{X}_i$ ,  $\bar{X}_k \leftrightarrow \bar{\mathbf{X}}_i^k$ ,  $W'_k \leftrightarrow W_i^k$ , we express

$$\bar{\mathbf{X}}_i \triangleq \mathbb{E}[\mathbf{X}_i | \mathbf{Y}_{[i]}^{[K]}] \quad (101)$$

$$= \sum_{k=1}^K \frac{\sigma_{\mathbf{X}_i | \mathbf{Y}_{[i]}^{[K]}}^2}{\sigma_{\mathbf{X}_i | \mathbf{Y}_{[i]}^k}^2} \bar{\mathbf{X}}_i^k, \quad (102)$$

which implies in particular

$$\mathbb{E}[\bar{\mathbf{X}}_i | \mathbf{X}_{[i]}, \mathbf{U}_{[i]}^{[K]*}] = \sum_{k=1}^K \frac{\sigma_{\mathbf{X}_i | \mathbf{Y}_{[i]}^{[K]}}^2}{\sigma_{\mathbf{X}_i | \mathbf{Y}_{[i]}^k}^2} \mathbb{E}[\bar{\mathbf{X}}_i^k | \mathbf{X}_{[i]}, \mathbf{U}_{[i]}^{[K]*}] \quad (103)$$

$$= \sum_{k=1}^K \frac{\sigma_{\mathbf{X}_i | \mathbf{Y}_{[i]}^{[K]}}^2}{\sigma_{\mathbf{X}_i | \mathbf{Y}_{[i]}^k}^2} \mathbb{E}[\bar{\mathbf{X}}_i^k | \mathbf{X}_{[i]}, \mathbf{U}_{[i]}^{k*}]. \quad (104)$$

It follows that steady-state causal MMSE in estimating  $\bar{\mathbf{X}}_i$  from  $\mathbf{X}_{[i]}$  and  $\mathbf{U}_{[i]}^{[K]*}$  satisfies

$$\sigma_{\bar{\mathbf{X}}_i \| \mathbf{X}_{[i]}, \mathbf{U}_{[i]}^{[K]*}}^2 = \sum_{k=1}^K \frac{\sigma_{\mathbf{X} \| \mathbf{Y}}^4}{\sigma_{\mathbf{X} \| \mathbf{Y}^k}^4} \rho_k. \quad (105)$$

Observe that

$$\sigma_{\bar{\mathbf{X}}_i | \mathbf{X}_{[i]}, \mathbf{U}_{[i]}^{[K]*}}^2 = \sigma_{\bar{\mathbf{X}}_i - \mathbb{E}[\bar{\mathbf{X}}_i | \mathbf{X}_{[i]}, \mathbf{U}_{[i]}^{[K]*}]}^2 \quad (106)$$

$$= \sigma_{\bar{\mathbf{X}}_i - \mathbf{X}_i - \mathbb{E}[\bar{\mathbf{X}}_i - \mathbf{X}_i | \mathbf{X}_{[i]}, \mathbf{U}_{[i]}^{[K]*}]}^2 \quad (107)$$

$$\leq \sigma_{\bar{\mathbf{X}}_i - \mathbf{X}_i - \mathbb{E}[\bar{\mathbf{X}}_i - \mathbf{X}_i | \mathbf{X}_i - \hat{\mathbf{X}}_i^*]}^2 \quad (108)$$

$$= \sigma_{\bar{\mathbf{X}}_i - \bar{\mathbf{X}}_i | \mathbf{X}_i - \hat{\mathbf{X}}_i^*}^2, \quad (109)$$

Now, we apply Lemma 2 with  $X \leftrightarrow \mathbf{X}_i - \bar{\mathbf{X}}_i$ ,  $Y \leftrightarrow \mathbf{X}_i - \hat{\mathbf{X}}_i^*$ ,  $W \leftrightarrow \bar{\mathbf{X}}_i - \bar{\mathbf{X}}_i$  to establish

$$\lim_{i \rightarrow \infty} \sigma_{\bar{\mathbf{X}}_i - \bar{\mathbf{X}}_i | \mathbf{X}_i - \hat{\mathbf{X}}_i^*}^2 = \sigma_{\bar{\mathbf{X}} \| \mathbf{Y}}^2 \left( 1 - \frac{\sigma_{\mathbf{X} \| \mathbf{Y}}^2}{d} \right), \quad (110)$$

which, together with (105) and (109), means

$$\frac{1}{d} \leq \frac{1}{\sigma_{\mathbf{X} \| \mathbf{Y}}^2} - \sum_{k=1}^K \frac{\rho_k}{\sigma_{\mathbf{X} \| \mathbf{Y}^k}^4}. \quad (111)$$

Also, note that

$$0 \leq \rho_k \leq \sigma_{\bar{\mathbf{X}}^k \| \mathbf{X}}^2. \quad (112)$$

We can now simplify the constraint set in the infimum in (100): the infimum is over  $\{\rho_k\}_{k=1}^K$  that satisfy (111) and (112).

It remains to clarify how the form in (53), (56), (57), parameterized in terms of

$$d_k \triangleq \sigma_{\mathbf{X} \| \mathbf{U}^{k*}}^2 \quad (113)$$

rather than  $\rho_k$ , is obtained. An application of Lemma 2 with  $X \leftrightarrow \mathbf{X}_i - \bar{\mathbf{X}}_i^k$ ,  $Y \leftrightarrow \mathbf{X}_i - \hat{\mathbf{X}}_i^k$ ,  $W \leftrightarrow \bar{\mathbf{X}}_i^k - \hat{\mathbf{X}}_i^k$  leads to

$$\rho_k = \sigma_{\mathbf{X} \| \mathbf{Y}^k}^2 \left( 1 - \frac{\sigma_{\mathbf{X} \| \mathbf{Y}^k}^2}{d_k} \right). \quad (114)$$

Plugging (114) into (111) leads to (56). Applying Lemma 2 with  $X \leftrightarrow \mathbf{X}_i - \bar{\mathbf{X}}_i^k$ ,  $Y \leftrightarrow \mathbf{X}_i$ ,  $W \leftrightarrow \bar{\mathbf{X}}_i^k$ , we express

$$\sigma_{\bar{\mathbf{X}}^k \| \mathbf{X}}^2 = \sigma_{\mathbf{X} \| \mathbf{Y}^k}^2 \left( 1 - \frac{\sigma_{\mathbf{X} \| \mathbf{Y}^k}^2}{\sigma_{\bar{\mathbf{X}}^k}^2} \right), \quad (115)$$

which, together with (114), implies the equivalence of (112) and (57). Finally, applying Lemma 2 with  $X \leftrightarrow \mathbf{X}_i - \bar{\mathbf{X}}_i^k$ ,  $Y \leftrightarrow \mathbf{X}_i - a\hat{\mathbf{X}}_{i-1}^k$ ,  $W \leftrightarrow \bar{\mathbf{X}}_i^k - a\hat{\mathbf{X}}_{i-1}^k$ , we express

$$\bar{\rho}_k = \sigma_{\mathbf{X} \| \mathbf{Y}^k}^2 \left( 1 - \frac{\sigma_{\mathbf{X} \| \mathbf{Y}^k}^2}{\bar{d}_k} \right). \quad (116)$$

Plugging (114) and (116) into (100), we conclude the equivalence of (100) and (53).  $\square$

## IV. LOSS DUE TO ISOLATED OBSERVERS

### A. Overview

In Section IV, we investigate how the sum rate - distortion function in Theorem 4 compares to what would be achievable had the encoders communicated with each other. A tight upper bound on the rate loss due to separate encoding is presented in Section IV-B (Theorem 5). Its proof relies on an upper bound on  $R_{\text{CEO}}(d)$  presented in Section IV-C (Proposition 1). The proof of Theorem 5 in Section IV-D concludes the section.

### B. Loss due to isolated observers

Unrestricted communication among the encoders is equivalent to having one encoder that sees all the observation processes  $\{Y_i^{[K]}\}$ . It is also equivalent to allowing joint encoding policies  $P_{B_{[i]}^{[K]} \| Y_{[i]}^{[K]}}$  in lieu of independent encoding policies  $\prod_{k=1}^K P_{B_{[i]}^k \| Y_{[i]}^k}$  in Definition 1.

The lossy compression setup in which the encoder has access only to a noise-corrupted version of the source has been referred to as ‘‘remote’’, ‘‘indirect’’, or ‘‘noisy’’ rate-distortion problem [34]–[37]. A causal setting was considered in [21, Th. 5–8, Cor. 1].

We denote the joint encoding counterpart of the operational fundamental limit  $R_{\text{CEO}}(d)$  (49) by  $R_{\text{rm}}(d)$  (remote).

The following result is a corollary to Theorem 4.

**Corollary 3** (Causal remote rate-distortion function). *For all  $\sigma_{\mathbf{X}||\mathbf{Y}^{[K]}}^2 < d < \sigma_{\mathbf{X}}^2$ , the sum rate - distortion function with joint encoding for the Gauss-Markov source in (2) observed through the Gaussian channels in (3) is given by*

$$R_{\text{rm}}(d) = \frac{1}{2} \log \frac{\bar{d} - \sigma_{\mathbf{X}||\mathbf{Y}^{[K]}}^2}{d - \sigma_{\mathbf{X}||\mathbf{Y}^{[K]}}^2}. \quad (117)$$

*Proof.* Examining its proof, it is easy to see that Theorem 4 continues to hold in the scenario with vector observations  $\mathbf{Y}_i^k$  (that are still required to be jointly Gaussian with  $\mathbf{X}_i$ ). In light of this fact, we view the joint encoding scenario as the CEO scenario with a single encoder that has access to all  $K$  observations, and we see that (53) indeed reduces to (117) in that case.

Previously, the minimal mutual information problem leading to  $R_{\text{rm}}(d)$  was solved in [21] in a different form using a different method; both forms are equivalent (Appendix E).  $\square$

The loss due to isolated encoders is bounded as follows.

**Theorem 5** (Loss due to isolated observers). *Consider the causal Gaussian CEO problem (2), (3). Assume that target distortion  $d$  satisfies*

$$\frac{1}{d} \geq \frac{1}{\sigma_{\mathbf{X}||\mathbf{Y}^{[K]}}^2} + \frac{K}{\sigma_{\mathbf{X}}^2} - \min_{k \in [K]} \frac{K}{\sigma_{\mathbf{X}||\mathbf{Y}^k}^2}. \quad (118)$$

*Then, the rate loss due to isolated observers is bounded as*

$$R_{\text{CEO}}(d) - R_{\text{rm}}(d) \leq (K - 1) (R_{\text{rm}}(d) - R(d)), \quad (119)$$

*with equality if and only if  $\sigma_{\mathbf{X}||\mathbf{Y}^k}^2$  are all the same.*

*Proof.* Section IV-D.  $\square$

Theorem 5 parallels the corresponding result for the classical Gaussian CEO problem [38, Cor. 1], and recovers it if  $a = 0$ . It's interesting that in both cases, the rate loss is bounded above by  $K - 1$  times the difference between the remote and the direct rate-distortion functions.

### C. A suboptimal waterfilling allocation

We present an upper bound to  $R_{\text{CEO}}(d)$  obtained by waterfilling over  $d_k$ 's. This parallels the corresponding result for the classical Gaussian CEO problem [38, Cor. 1]. Like [38], we use waterfilling to obtain this result, but unlike the case  $t = 1$  considered in [38] where waterfilling is optimal [6], it is only suboptimal if  $t > 1$  due to the memory of the past steps at the encoders and the decoder. This is unsurprising as for the same reason waterfilling cannot be applied to solve the vector Gaussian rate-distortion problem for  $t > 1$  [21, Remark 2].

**Proposition 1.** *For all  $\sigma_{\mathbf{X}||\mathbf{Y}^{[K]}}^2 < d < \sigma_{\mathbf{X}}^2$ , the causal sum rate - distortion function for the Gauss-Markov source in*

(2) observed through the Gaussian channels in (3) is upper-bounded as

$$R_{\text{CEO}}(d) \leq \frac{1}{2} \log \frac{\bar{d}}{d} + \sum_{k=1}^K \frac{1}{2} \log \frac{\bar{d}_k - \sigma_{\mathbf{X}||\mathbf{Y}^k}^2}{d_k - \sigma_{\mathbf{X}||\mathbf{Y}^k}^2} \frac{d_k}{\bar{d}_k}, \quad (120)$$

where  $d_k, k \in [K]$  satisfy

$$\frac{1}{\sigma_{\mathbf{X}||\mathbf{Y}^k}^2} - \frac{1}{d_k} = \min \left\{ \frac{1}{\lambda}, \frac{1}{\sigma_{\mathbf{X}||\mathbf{Y}^k}^2} - \frac{1}{\sigma_{\mathbf{X}}^2} \right\}, \quad (121)$$

$\lambda$  is the solution to

$$\sum_{k=1}^K \min \left\{ \frac{1}{\lambda}, \frac{1}{\sigma_{\mathbf{X}||\mathbf{Y}^k}^2} - \frac{1}{\sigma_{\mathbf{X}}^2} \right\} = \frac{1}{\sigma_{\mathbf{X}||\mathbf{Y}^{[K]}}^2} - \frac{1}{d}, \quad (122)$$

and  $\bar{d}, \bar{d}_k$  are defined in (54), (55) respectively. Inequality in (120) holds with equality if all  $\sigma_{\mathbf{X}||\mathbf{Y}^k}^2$  are equal.

*Proof.* We first check that the choice in (121) is feasible. Since the right side of (121) is lower-bounded by 0 and upper bounded by  $\frac{1}{\sigma_{\mathbf{X}||\mathbf{Y}^k}^2} - \frac{1}{\sigma_{\mathbf{X}}^2}$ , (57) is satisfied. Furthermore, substituting (122) ensures that (56) is satisfied with equality.

To claim equality in the symmetrical case, it suffices to recall that in that case, the minimum in (53) is attained by  $d_1 = \dots = d_K$  (Corollary 1).  $\square$

### D. Proof of Theorem 5

Under the assumption (118), the waterfilling allocation in Proposition 1 results in all active transmitters, and (121) reduces to

$$\frac{1}{\sigma_{\mathbf{X}||\mathbf{Y}^k}^2} - \frac{1}{d_k} = \frac{1}{\lambda}, \quad (123)$$

while (122) reduces to

$$\lambda = K \left( \frac{1}{\sigma_{\mathbf{X}||\mathbf{Y}^{[K]}}^2} - \frac{1}{d} \right)^{-1}. \quad (124)$$

Substituting (123) into (120) we conclude that under assumption (118),

$$R_{\text{CEO}}(d) = \frac{1}{2} \log \frac{\bar{d}}{d} + \frac{1}{2} \sum_{k=1}^K \log \left[ \left( \frac{1}{\sigma_{\mathbf{X}||\mathbf{Y}^k}^2} - \frac{1}{d_k} \right) \lambda \right] \quad (125)$$

$$\leq \frac{1}{2} \log \frac{\bar{d}}{d} + \frac{K}{2} \log \left[ \sum_{k=1}^K \left( \frac{1}{\sigma_{\mathbf{X}||\mathbf{Y}^k}^2} - \frac{1}{\bar{d}_k} \right) \frac{\lambda}{K} \right] \quad (126)$$

$$= \frac{1}{2} \log \frac{\bar{d}}{d} + \frac{K}{2} \log \left( \frac{1}{\sigma_{\mathbf{X}||\mathbf{Y}^{[K]}}^2} - \frac{1}{\bar{d}} \right) \frac{\lambda}{K} \quad (127)$$

$$= \frac{1}{2} \log \frac{\bar{d} - \sigma_{\mathbf{X}||\mathbf{Y}^{[K]}}^2}{d - \sigma_{\mathbf{X}||\mathbf{Y}^{[K]}}^2} + \frac{K - 1}{2} \log \frac{\bar{d} - \sigma_{\mathbf{X}||\mathbf{Y}^{[K]}}^2}{d - \sigma_{\mathbf{X}||\mathbf{Y}^{[K]}}^2} \frac{d}{\bar{d}} \quad (128)$$

where

- (126) is by Jensen's inequality, since log is concave;
- (127) is due to

$$\frac{1}{\sigma_{\mathbf{X}||\mathbf{Y}^{[K]}}^2} = \sum_{k=1}^K \frac{1}{\sigma_{\mathbf{X}||\mathbf{Y}^k}^2} - \frac{K - 1}{\sigma_{\mathbf{X}}^2}, \quad (129)$$

$$\frac{1}{\bar{d}} = \sum_{k=1}^K \frac{1}{\bar{d}_k} - \frac{K - 1}{\sigma_{\mathbf{X}}^2}, \quad (130)$$

which holds by Lemma 3 even if the source is nonstationary (that is,  $|a| \geq 1$  and  $\sigma_X^2 = \infty$ ), as a simple limiting argument taking  $\frac{K-1}{\sigma_X^2}$  to 0 confirms.

- (128) holds by substituting (123) into (127).

Notice that (119) is just another way to write (128), using (117) and (58). To verify the condition for equality, note that ‘=’ holds in (125) in the symmetrical case by Proposition 1, and that ‘=’ holds in (126) only in the symmetrical case due to strict concavity of the log function.  $\square$

## V. CONCLUSION

In this paper, we set up the causal CEO problem (Definition 1, Definition 2) and we prove that the sum rate - distortion function is given by the directed mutual information from the encoders to the decoder minimized subject to the distortion constraint and the separate encoding constraint (Theorem 1). The proof of the direct coding theorem hinges upon an SLC-based nonasymptotic bound (Theorem 2) that extends [26, Th. 6] to the case with  $K > 2$  observers and  $t > 1$  time steps. An asymptotic analysis of Theorem 2 leads to an extension of the Berger-Tung inner bound [11], [12] to  $t > 1$  time steps (Theorem 3).

By solving the minimal directed mutual information problem in Theorem 1, we characterize the Gaussian sum rate - distortion function as a convex optimization problem over  $K$  parameters (Theorem 4). We give an explicit formula in the identical-channels case (Corollary 1) and study its asymptotic behavior as  $K \rightarrow \infty$  (Corollary 2). We derive the causal Gaussian remote rate-distortion function as a corollary to Theorem 4 with  $K = 1$  (Corollary 3). Using a suboptimal waterfilling allocation over the  $K$  optimization parameters in Theorem 4 (Proposition 1), we upper-bound the rate loss due to separated observers (Theorem 5).

As future work, it will be interesting to determine the full rate-distortion region of the causal Gaussian CEO problem as opposed to the sum rate we found in this paper. While Theorem 3 already gives an inner bound to that region, developing a converse remains open. The techniques in [10], [14], [15] appear promising in that pursuit. Certain causal multiterminal source coding problems also appear within reach in view of the result in [9] and the applicability of Theorem 3 to multiterminal source coding. Finally, computing the rate-distortion region for Gaussian processes beyond the Gauss-Markov source with i.i.d. components would be an important advance.

## REFERENCES

- [1] V. Kostina and B. Hassibi, “Fundamental limits of distributed tracking,” in *Proceedings 2020 IEEE International Symposium on Information Theory*, June 2020, to appear.
- [2] T. Berger, Z. Zhang, and H. Viswanathan, “The CEO problem [multiterminal source coding],” *IEEE Transactions on Information Theory*, vol. 42, no. 3, pp. 887–902, 1996.
- [3] H. Viswanathan and T. Berger, “The quadratic Gaussian CEO problem,” *IEEE Transactions on Information Theory*, vol. 43, no. 5, pp. 1549–1559, 1997.
- [4] Y. Oohama, “The rate-distortion function for the quadratic Gaussian CEO problem,” *IEEE Transactions on Information Theory*, vol. 44, no. 3, pp. 1057–1070, May 1998.
- [5] V. Prabhakaran, D. Tse, and K. Ramachandran, “Rate region of the quadratic Gaussian CEO problem,” in *Proceedings 2004 International Symposium on Information Theory*, June 2004, p. 119.
- [6] J. Chen, X. Zhang, T. Berger, and S. B. Wicker, “An upper bound on the sum-rate distortion function and its corresponding rate allocation schemes for the CEO problem,” *IEEE Journal on Selected Areas in Communications*, vol. 22, no. 6, pp. 977–987, 2004.
- [7] H. Behroozi and M. R. Soleymani, “Optimal rate allocation in successively structured Gaussian CEO problem,” *IEEE Transactions on Wireless Communications*, vol. 8, no. 2, pp. 627–632, 2009.
- [8] J. Chen and T. Berger, “Successive Wyner–Ziv coding scheme and its application to the quadratic Gaussian CEO problem,” *IEEE Transactions on Information Theory*, vol. 54, no. 4, pp. 1586–1603, 2008.
- [9] A. B. Wagner, S. Tavildar, and P. Viswanath, “Rate region of the quadratic Gaussian two-encoder source-coding problem,” *IEEE Transactions on Information Theory*, vol. 54, no. 5, pp. 1938–1961, 2008.
- [10] A. B. Wagner and V. Anantharam, “An improved outer bound for multiterminal source coding,” *IEEE Transactions on Information Theory*, vol. 54, no. 5, pp. 1919–1937, 2008.
- [11] T. Berger, *Multi-terminal source coding. Chapter in The Information Theory Approach to Communications*. Springer-Verlag, 1978.
- [12] S.-Y. Tung, “Multiterminal source coding,” Ph.D. dissertation, School of Electrical Engineering, Cornell University, 1978.
- [13] J. Wang, J. Chen, and X. Wu, “On the sum rate of Gaussian multi-terminal source coding: New proofs and results,” *IEEE Transactions on Information Theory*, vol. 56, no. 8, pp. 3946–3960, 2010.
- [14] E. Ekrem and S. Ulukus, “An outer bound for the vector Gaussian CEO problem,” *IEEE Transactions on Information Theory*, vol. 60, no. 11, pp. 6870–6887, 2014.
- [15] J. Wang and J. Chen, “Vector Gaussian multiterminal source coding,” *IEEE Transactions on Information Theory*, vol. 60, no. 9, pp. 5533–5552, 2014.
- [16] T. A. Courtade and T. Weissman, “Multiterminal source coding under logarithmic loss,” *IEEE Transactions on Information Theory*, vol. 60, no. 1, pp. 740–761, Jan. 2014.
- [17] A. Gorbunov and M. S. Pinsker, “Nonanticipatory and prognostic  $\epsilon$ -entropies and message generation rates,” *Problemy Peredachi Informatsii*, vol. 9, no. 3, pp. 12–21, 1973.
- [18] —, “Prognostic epsilon entropy of a Gaussian message and a Gaussian source,” *Problemy Peredachi Informatsii*, vol. 10, no. 2, pp. 5–25, 1974.
- [19] S. Tatikonda, A. Sahai, and S. Mitter, “Stochastic linear control over a communication channel,” *IEEE Transactions on Automatic Control*, vol. 49, no. 9, pp. 1549–1561, 2004.
- [20] E. Silva, M. Derpich, J. Ostergaard, and M. Encina, “A characterization of the minimal average data rate that guarantees a given closed-loop performance level,” *IEEE Transactions on Automatic Control*, 2016.
- [21] V. Kostina and B. Hassibi, “Rate-cost tradeoffs in control,” *IEEE Transactions on Automatic Control*, Apr. 2019.
- [22] T. Tanaka, K.-K. Kim, P. A. Parrilo, and S. K. Mitter, “Semidefinite programming approach to Gaussian sequential rate-distortion tradeoffs,” *IEEE Transactions on Automatic Control*, vol. 62, no. 4, pp. 1896–1910, 2017.
- [23] V. Kostina and B. Hassibi, “Rate-cost tradeoffs in scalar LQG control and tracking with side information,” in *2018 56th Annual Allerton Conference on Communication, Control, and Computing*, Monticello, IL, Oct. 2018, pp. 421–428.
- [24] A. P. Johnston and S. Yüksel, “Stochastic stabilization of partially observed and multi-sensor systems driven by unbounded noise under fixed-rate information constraints,” *IEEE Transactions on Automatic Control*, vol. 59, no. 3, pp. 792–798, 2014.
- [25] M. H. Yassaee, M. R. Aref, and A. Gohari, “A technique for deriving one-shot achievability results in network information theory,” in *Proceedings 2013 IEEE International Symposium on Information Theory*, Istanbul, Turkey, July 2013.
- [26] —, “A technique for deriving one-shot achievability results in network information theory,” 2013.
- [27] N. Merhav, “The generalized stochastic likelihood decoder: Random coding and expurgated bounds,” *IEEE Transactions on Information Theory*, vol. 63, no. 8, pp. 5039–5051, 2017.
- [28] G. Kramer, “Directed information for channels with feedback,” Ph.D. dissertation, ETH Zurich, 1998.
- [29] J. Massey, “Causality, feedback and directed information,” in *Proc. Int. Symp. Inf. Theory Applic. (ISITA-90)*, 1990, pp. 303–305.
- [30] J. Liu, P. Cuff, and S. Verdú, “On  $\alpha$ -decodability and  $\alpha$ -likelihood decoder,” in *Proceedings 55th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, Monticello, IL, Oct. 2017, pp. 118–124.

- [31] T. Tanaka, "Semidefinite representation of sequential rate-distortion function for stationary Gauss-Markov processes," in *Proceedings 2015 IEEE Conference on Control Applications (CCA)*, Sep. 2015, pp. 1217–1222.
- [32] N. Guo and V. Kostina, "Optimal causal rate-constrained sampling of the Wiener process," in *Proceedings 57th Annual Allerton Conference on Communication, Control and Computing*, Monticello, IL, Sep. 2019.
- [33] M. S. Derpich and J. Ostergaard, "Improved upper bounds to the causal quadratic rate-distortion function for Gaussian stationary sources," *IEEE Transactions on Information Theory*, vol. 58, no. 5, pp. 3131–3152, May 2012.
- [34] T. Berger, *Rate distortion theory*. Prentice-Hall, Englewood Cliffs, NJ, 1971.
- [35] R. Dobrushin and B. Tsybakov, "Information transmission with additional noise," *IRE Transactions on Information Theory*, vol. 8, no. 5, pp. 293–304, Sep. 1962.
- [36] H. S. Witsenhausen, "Indirect rate distortion problems," *IEEE Transactions on Information Theory*, vol. 26, no. 5, pp. 518–521, Sep. 1980.
- [37] V. Kostina and S. Verdú, "Nonasymptotic noisy lossy source coding," *IEEE Transactions on Information Theory*, vol. 62, no. 11, pp. 6111–6123, Nov. 2016.
- [38] V. Kostina, "Rate loss in the Gaussian CEO problem," in *Proceedings 2019 IEEE Information Theory Workshop*, Visby, Gotland, Sweden, Aug. 2019.
- [39] Y. Polyanskiy, "Information theory lecture notes," *Dep. Electrical Engineering and Computer Science, M.I.T.*, 2012.

## APPENDIX A PROOF OF THEOREM 2

**Codebooks:** Encoder  $k$  maintains separate codebooks  $\underline{U}_1^k, \underline{U}_2^k, \dots, \underline{U}_t^k$  to use at the transmission instances  $1, 2, \dots, t$  respectively. Codebook  $\underline{U}_i^k$  is an  $n \times L_1^k \times \dots \times L_i^k$ -dimensional array: there is a separate codebook for each possible realization of past chosen codewords.

For vector of indices  $\ell_{[i]} \in \prod_{j=1}^i [L_j^k]$ , we denote by  $\underline{U}_i^k(\ell_{[i]})$  the codeword corresponding to index  $\ell_i$ , given the past indices  $\ell_{[i-1]}$ . For subsets  $\mathcal{K} \subseteq [K]$  and  $\mathcal{I} \subseteq [t]$ , we denote the collection of codebooks  $\underline{U}_{\mathcal{I}}^{\mathcal{K}} \triangleq (\underline{U}_i^k : k \in \mathcal{K}, i \in \mathcal{I})$ . For indices  $\ell_i^k \in [L_i^k]$ ,  $i \in [t]$ ,  $k \in [K]$ , we denote their collection  $\ell_{\mathcal{I}}^{\mathcal{K}} \triangleq (\ell_i^k : k \in \mathcal{K}, i \in \mathcal{I})$ . Finally,  $\underline{U}_{\mathcal{I}}^{\mathcal{K}}(\ell_{\mathcal{I}}^{\mathcal{K}}) \triangleq (\underline{U}_i^k(\ell_{[i]}^k) : k \in \mathcal{K}, i \in \mathcal{I})$  denotes the codewords corresponding to  $\ell_{\mathcal{I}}^{\mathcal{K}}$ ;  $1_{\mathcal{I}}^{\mathcal{K}}$  denotes the array of 1's of dimension  $|\mathcal{K}| \times |\mathcal{I}|$ .

Codebook 1 for encoder  $k$ ,  $\underline{U}_1^k$ , consists of  $L_1^k$  codewords drawn i.i.d. from  $P_{U_1^k}$ . For  $i = 2, \dots, t$ , codebook  $i$  for user  $k$ ,  $\underline{U}_i^k$ , consists of  $L_i^k$  codewords drawn i.i.d. from  $P_{U_i^k | U_{[i-1]}^k = \underline{U}_{[i-1]}^k}(\ell_{[i-1]}^k)$ , for each  $\ell_{[i-1]}^k \in \prod_{j=1}^{i-1} [L_j^k]$ .

**Random binning:** Let  $B_i^k : [L_i^k] \mapsto [M_i^k]$ ,  $i = 1, 2, \dots, t$ , be random mappings in which each element of  $[L_i^k]$  is mapped equiprobably and independently to the set  $[M_i^k]$ .

We will use the notation  $B_{\mathcal{I}}^{\mathcal{K}}(\ell_{\mathcal{I}}^{\mathcal{K}}) \triangleq (B_i^k(\ell_{[i]}^k) : k \in \mathcal{K}, i \in \mathcal{I})$  denotes the codewords corresponding to  $\ell_{\mathcal{I}}^{\mathcal{K}}$ .

In the description of coding operations that follows, we denote the instances of the random codebooks in operation by  $\underline{u}_i^k$  and those of the random binning functions by  $b_i^k$ .

**Encoders:** The encoders use the stochastic likelihood coder (SLC) [25], [26] followed by random binning. Each user  $k$  maintains a collection of encoders indexed by time  $i = 1, 2, \dots, t$ ; at time  $i$ , encoder  $i$  is invoked to form and transmit a codeword at that time.

**Encoder  $i$  for user  $k$ :** Given an observation  $y_i^k \in \mathcal{Y}_i^k$  and past codewords  $\ell_{[i-1]}^k \in \prod_{j=1}^{i-1} [L_j^k]$ , the SLC chooses an index  $\ell_i^k \in [L_i^k]$  with probability

$$Q_{U_i^k | Y_{[i]}^k = y_{[i]}^k, U_{[i-1]}^k = \underline{u}_{[i-1]}^k}(\ell_{[i]}^k) \left( \underline{u}_i^k(\ell_{[i]}^k) \right) \quad (131)$$

$$= \frac{\exp \left( \iota \left( y_{[i]}^k; \underline{u}_i^k(\ell_{[i]}^k) | \underline{u}_{[i-1]}^k(\ell_{[i-1]}^k) \right) \right)}{\sum_{\ell=1}^{L_i^k} \exp \left( \iota \left( y_{[i]}^k; \underline{u}_i^k \left( (\ell_{[i-1]}^k, \ell) \right) | \underline{u}_{[i-1]}^k(\ell_{[i-1]}^k) \right) \right)},$$

where the conditional information density is with respect to the given distribution  $P_{Y_i^k | U_{[i-1]}^k}$ . Encoder  $i$  transmits  $m_i^k = b_i^k(\ell_i^k)$  to the decoder, a realization of the random variable we denote by  $B_i^k$ .

The causal encoder  $k$  is the resulting causal probability kernel

$$Q_{B_{[t]}^k | Y_{[t]}^k} (m_{[t]}^k | y_{[t]}^k)$$

$$= \sum_{\ell_{[t]}^k} \mathbb{1} \left\{ b_{[t]}^k(\ell_{[t]}^k) = m_{[t]}^k \right\} Q_{U_{[t]}^k | Y_{[t]}^k}(\ell_{[t]}^k | y_{[t]}^k). \quad (132)$$

Since the encoders operate independently,

$$Q_{U_{[t]}^{[K]} | Y_{[t]}^{[K]}} = \prod_{k=1}^K Q_{U_{[t]}^k | Y_{[t]}^k}, \quad (133)$$

$$Q_{B_{[t]}^{[K]} | Y_{[t]}^{[K]}} = \prod_{k=1}^K Q_{B_{[t]}^k | Y_{[t]}^k}. \quad (134)$$

**Decoder:** Having received the collection of bin numbers  $m_i^{[K]} \in \prod_{k=1}^K [M_i^k]$  at time  $i$  and remembering the past, the decoder invokes a generalized likelihood decoder (GLD) [27, eq. (4)] to select among indices that fall into those bins a collection of indices  $\hat{\ell}_i^{[K]} \in \prod_{k=1}^K [L_i^k]$  with probability

$$Q_{\hat{U}_i^{[K]} | B_{[i]}^{[K]} = b_{[i]}^{[K]}, \hat{U}_{[i-1]}^{[K]} = \underline{u}_{[i-1]}^{[K]}(\hat{\ell}_{[i-1]}^{[K]})} \left( \underline{u}_i^{[K]}(\hat{\ell}_i^{[K]}) \right) =$$

$$\frac{\mathfrak{g} \left( \underline{u}_{[i]}^{[K]}(\hat{\ell}_{[i]}^{[K]}) \right) \mathbb{1} \left\{ b_{[i]}^{[K]}(\hat{\ell}_i^{[K]}) = m_i^{[K]} \right\}}{\sum_{\ell_i^{[K]}} \mathfrak{g} \left( \underline{u}_{[i]}^{[K]}(\ell_{[i]}^{[K]}, \ell_i^{[K]}) \right) \mathbb{1} \left\{ b_{[i]}^{[K]}(\ell_i^{[K]}) = m_i^{[K]} \right\}}, \quad (135)$$

where

$$\mathfrak{g} \left( \underline{u}_{[i]}^{[K]} \right) \triangleq \prod_{k=1}^K \mathbb{1} \left\{ j^{\pi(k)} \left( \underline{u}_{[i]}^{\pi(k)} \right) \geq \log \frac{L_i^{\pi(k)}}{M_i^{\pi(k)}} + \beta_i^{\pi(k)} \right\} \quad (136)$$

Having determined  $\hat{\ell}_i^{[K]}$ , the decoder applies the given transformation  $P_{\hat{X}_i | U_{[i]}^{[K]}, \hat{X}_{[i-1]}^{[K]}}$  to form the estimate of the source  $\hat{X}_i \left( \underline{u}_{[i]}^{[K]}(\hat{\ell}_{[i]}^{[K]}) \right)$ . The causal decoder is the resulting causal kernel  $Q_{\hat{X}_{[t]} | B_{[t]}^{[K]}}$ .

**Error analysis:** We consider two error events:

$$\mathcal{E}_{\text{dec}} : \hat{U}_{[t]}^{[K]} \neq U_{[t]}^{[K]} \quad (137)$$

$$\mathcal{E}_{\text{enc}} : \bigcup_{i=1}^t \left\{ d \left( X_i, \hat{X}_i \left( U_{[i]}^{[K]} \right) \right) > d_i \right\}, \quad (138)$$

where  $U_{[i]}^{[K]}$  are the codewords chosen by the encoders at encoding step (131), and  $\hat{U}_{[t]}^{[K]}$  is the decoder's estimate of those codewords after decoding step (135).

Note that  $\mathcal{E}_{\text{dec}}$  is the event that some codewords are not recovered (decoding error), and  $\mathcal{E}_{\text{enc}}$  is the event that some distortions exceed threshold even if all the codewords are recovered correctly (encoding error). We denote for brevity by  $\mathcal{F}$  the sigma-algebra generated by  $Y_{[t]}^{[K]}$ ,  $\underline{U}_{[t]}^{[K]}(1_{[t]}^{[K]})$ ,  $\mathbf{B}_{[t]}^{[K]}(1_{[t]}^{[K]})$ ,  $\hat{X}_{[t]}(\underline{U}_{[t]}^{[K]}(1_{[t]}^{[K]}))$ ; by  $\mathbb{Q}$  the probability measure generated by the code; and by  $F_i^k$ ,  $G_i$  the denominators in (131) and (135), respectively. Following Shannon's random coding argument and the Jensen inequality technique of Yassaee et al. [25], [26], we proceed to bound an expectation of the indicator of the correct decoding event with respect to both the actual source code and the random codebooks.

$$\begin{aligned} & \mathbb{E} \left[ \mathbb{Q} \left[ \prod_{i=1}^t 1 \left\{ d \left( X_i, \hat{X}_i \right) \leq d_i \mid \underline{U}_{[t]}^{[K]}, \mathbf{B}_{[t]}^{[K]} \right\} \right] \right] \\ & \geq \mathbb{E} \left[ \mathbb{Q} \left[ \mathcal{E}_{\text{enc}}^c \cap \mathcal{E}_{\text{dec}}^c \mid \underline{U}_{[t]}^{[K]}, \mathbf{B}_{[t]}^{[K]} \right] \right] \end{aligned} \quad (139)$$

$$\begin{aligned} & = \mathbb{E} \left[ \sum_{\ell_{[t]}^{[K]} \in \prod_{i=1}^t \prod_{k=1}^K [L_i^k]} Q_{U_{[t]}^{[K]} \parallel Y_{[t]}^{[K]}} \left( \underline{U}_{[t]}^{[K]} \left( \ell_{[t]}^{[K]} \right) \parallel Y_{[t]}^{[K]} \right) \right. \\ & \quad \cdot \sum_{m_{[t]}^{[K]} \in \prod_{i=1}^t \prod_{k=1}^K [M_i^k]} 1 \left\{ \mathbf{B}_{[t]}^{[K]} \left( \ell_{[t]}^{[K]} \right) = m_{[t]}^{[K]} \right\} \\ & \quad \cdot \left. Q_{\hat{U}_{[t]}^{[K]} \parallel \mathbf{B}_{[t]}^{[K]} = m_{[t]}^{[K]}} \left( \underline{U}_{[t]}^{[K]} \left( \ell_{[t]}^{[K]} \right) \right) 1 \left\{ \mathcal{E}_{\text{enc}}^c \right\} \right] \end{aligned} \quad (140)$$

$$\begin{aligned} & = \prod_{k=1}^K \prod_{i=1}^t M_i^k L_i^k \\ & \quad \cdot \mathbb{E} \left[ \mathbb{E} \left[ Q_{U_{[t]}^{[K]} \parallel Y_{[t]}^{[K]}} \left( \underline{U}_{[t]}^{[K]} \left( 1_{[t]}^{[K]} \right) \parallel Y_{[t]}^{[K]} \right) \right. \right. \\ & \quad \cdot 1 \left\{ \mathbf{B}_{[t]}^{[K]} \left( 1_{[t]}^{[K]} \right) = 1_{[t]}^{[K]} \right\} \\ & \quad \cdot \left. \left. Q_{\hat{U}_{[t]}^{[K]} \parallel \mathbf{B}_{[t]}^{[K]} = 1_{[t]}^{[K]}} \left( \underline{U}_{[t]}^{[K]} \left( 1_{[t]}^{[K]} \right) \right) 1 \left\{ \mathcal{E}_{\text{enc}}^c \mid \mathcal{F} \right\} \right] \right] \quad (141) \\ & \geq \prod_{k=1}^K \prod_{i=1}^t M_i^k L_i^k \\ & \quad \cdot \mathbb{E} \left[ \prod_{k=1}^K \prod_{i=1}^t \frac{\exp \left( \iota \left( Y_{[i]}^k; \underline{U}_i^k \left( 1_{[i]} \right) \mid \underline{U}_{[i-1]}^k \left( 1_{[i-1]} \right) \right) \right)}{\mathbb{E} \left[ F_i^k \mid \mathcal{F} \right]} \right. \\ & \quad \cdot \frac{\mathfrak{g} \left( \underline{U}_{[i]}^{[K]} \left( 1_{[i]}^{[K]} \right) \right) 1 \left\{ \mathbf{B}_i^{[K]} \left( 1_{[i]}^{[K]} \right) = 1_{[i]}^{[K]} \right\}}{\mathbb{E} \left[ G_i \mid \mathcal{F} \right]} \\ & \quad \cdot \left. 1 \left\{ d \left( X_i, \hat{X}_i \left( \underline{U}_{[i]}^{[K]} \left( 1_{[i]}^{[K]} \right) \right) \right) \leq d_i \right\} \right], \end{aligned} \quad (142)$$

where

- the expectation  $\mathbb{E}$  in (140) is with respect to the codebooks  $\underline{U}_{[t]}^{[K]}$ , the random binning functions  $\mathbf{B}_{[t]}^{[K]}$ , the decoder  $P_{\hat{X}_{[t]} \parallel U_{[t]}^{[K]}}$  and  $X_{[t]}$ ,  $Y_{[t]}^{[K]}$ ;
- (141) uses that both the codewords and the binning functions for the  $i$ -th time instant are independently and identically distributed, thus each choice of  $\ell_{[t]}^{[K]}$  and  $m_{[t]}^{[K]}$  results in the same probability as the choice  $\ell_{[t]}^{[K]} = 1_{[t]}^{[K]}$  and  $m_{[t]}^{[K]} = 1_{[t]}^{[K]}$ . Here we also conditioned on  $\mathcal{F}$  before taking an outer expectation

with respect to it, which will facilitate the next step of the calculation.

- the main step (142) is shown as follows. The product  $Q_{U_{[t]}^{[K]} \parallel Y_{[t]}^{[K]}} Q_{\hat{U}_{[t]}^{[K]} \parallel \mathbf{B}_{[t]}^{[K]}}$  is proportional to the product of  $(K+1)t$  factors  $\prod_{i=1}^t \frac{1}{G_i} \prod_{k=1}^K \frac{1}{F_i^k}$ . Applying Jensen's inequality to this jointly convex function of  $(K+1)t$  variables yields

$$\mathbb{E} \left[ \prod_{i=1}^t \frac{1}{G_i} \prod_{k=1}^K \frac{1}{F_i^k} \mid \mathcal{F} \right] \geq \prod_{i=1}^t \frac{1}{\mathbb{E} \left[ G_i \mid \mathcal{F} \right]} \prod_{k=1}^K \frac{1}{\mathbb{E} \left[ F_i^k \mid \mathcal{F} \right]} \quad (143)$$

We compute each factor in (143) as follows.

$$\begin{aligned} & \mathbb{E} \left[ F_i^k \mid \mathcal{F} \right] \quad (144) \\ & = \mathbb{E} \left[ \sum_{\ell=1}^{L_i^k} \exp \left( \iota \left( Y_{[i]}^k; \underline{U}_i^k \left( 1_{[i-1]} \right), \ell \right) \mid \underline{U}_{[i-1]}^k \left( 1_{[i-1]} \right) \right) \mid \mathcal{F} \right] \\ & = \exp \left( \iota \left( Y_{[i]}^k; \underline{U}_i^k \left( 1_{[i]} \right) \mid \underline{U}_{[i-1]}^k \left( 1_{[i-1]} \right) \right) \right) + (L_i^k - 1) \cdot \\ & \quad \mathbb{E} \left[ \exp \left( \iota \left( Y_{[i]}^k; \underline{U}_i^k \left( 1_{[i-1]} \right), 2 \right) \mid \underline{U}_{[i-1]}^k \left( 1_{[i-1]} \right) \right) \mid \mathcal{F} \right] \quad (145) \\ & = \exp \left( \iota \left( Y_{[i]}^k; \underline{U}_i^k \left( 1_{[i]} \right) \mid \underline{U}_{[i-1]}^k \left( 1_{[i-1]} \right) \right) \right) + (L_i^k - 1), \end{aligned} \quad (146)$$

where to write (145) we used that the codewords  $\{\underline{U}_i^k(1_{[i-1]}, \ell) : \ell \neq 1\}$  are identically distributed conditioned on  $\mathcal{F}$ .

To evaluate  $\mathbb{E} \left[ G_i \mid \mathcal{F} \right]$ , we partition the set of all  $\ell_i^{[K]} \in \prod_{i=1}^K [L_i^k]$  into index sets parameterized by  $\mathcal{K} \subseteq [K]$ :

$$\mathcal{L}_i(\mathcal{K}) \triangleq \left\{ \ell^{[K]} \in \prod_{i=1}^K [L_i^k] : \ell^{\pi(k)} = 1, k \in \mathcal{K}, \right. \\ \left. \ell^{\pi(k)} \neq 1, k \in \mathcal{K}^c \right\}, \quad (147)$$

and for each  $\ell_i^{[K]} \in \mathcal{L}_i(\mathcal{K})$ ,  $\mathcal{K} \subset K$ , we upper-bound  $\mathfrak{g}(\cdot)$  as

$$\begin{aligned} & \mathfrak{g} \left( \underline{u}_{[i]}^{[K]} \left( 1_{[i-1]}^{[K]}, \ell_i^{[K]} \right) \right) \quad (148) \\ & \leq \prod_{k \in \mathcal{K}^c} 1 \left\{ j^{\pi(k)} \left( \underline{u}_{[i]}^{\pi([K])} \right) \geq \log \frac{L_i^{\pi(k)}}{M_i^{\pi(k)}} + \beta_i^{\pi(k)} \right\} \end{aligned} \quad (149)$$

$$\leq \prod_{k \in \mathcal{K}^c} \frac{M_i^{\pi(k)}}{L_i^{\pi(k)}} \exp \left( j^{\pi(k)} \left( \underline{u}_{[i]}^{\pi([K])} \left( 1_{[i-1]}^{[K]}, \ell_i^{[K]} \right) \right) - \beta_i^{\pi(k)} \right), \quad (150)$$

while for  $\mathcal{K} = [K]$ , we upper-bound it as

$$\mathfrak{g} \left( \underline{u}_{[i]}^{[K]} \left( 1_{[i]}^{[K]} \right) \right) \leq 1. \quad (151)$$

Note that for each  $\ell_i^{[K]} \in \mathcal{L}_i(\mathcal{K})$ ,  $\mathcal{K} \subset K$

$$\mathbb{E} \left[ \prod_{k \in \mathcal{K}^c} \exp \left( j^{\pi(k)} \left( \underline{u}_{[i]}^{\pi([K])} \left( 1_{[i-1]}^{[K]}, \ell_i^{[K]} \right) \right) \right) \mid \mathcal{F} \right] = 1. \quad (152)$$

The upper-bound in (150) and the equality in (152) are key to the analysis of our GLD (135).

Now,  $\mathbb{E}[G_i|\mathcal{F}]$  is bounded as

$$\begin{aligned} & \mathbb{E}[G_i|\mathcal{F}] = \\ & \mathbb{E} \left[ \sum_{\ell_i^{[K]} \in \prod_{k=1}^K [L_i^k]} \mathbf{g} \left( \underline{U}_{[i]}^{[K]} \left( 1_{[i-1]}^{[K]}, \ell_i^{[K]} \right) \right) \right. \\ & \quad \left. \cdot 1 \left\{ \mathbf{B}_i^{[K]}(\ell_i^{[K]}) = 1^{[K]} \right\} \mid \mathcal{F} \right] \quad (153) \\ & = \mathbf{g} \left( \underline{U}_{[i]}^{[K]} \left( 1_{[i]}^{[K]} \right) \right) 1 \left\{ \mathbf{B}_i^{[K]}(1^{[K]}) = 1^{[K]} \right\} \\ & \quad + \sum_{\mathcal{K} \subset [K]} \mathbb{E} \left[ \sum_{\ell_i^{[K]} \in \mathcal{L}_i(\mathcal{K})} \mathbf{g} \left( \underline{U}_{[i]}^{[K]} \left( 1_{[i-1]}^{[K]}, \ell_i^{[K]} \right) \right) \mid \mathcal{F} \right] \\ & \quad \cdot \prod_{k \in \mathcal{K}^c} \frac{1}{M_i^{\pi(k)}} \cdot 1 \left\{ \mathbf{B}_i^{\pi(\mathcal{K})}(1^{\mathcal{K}}) = 1^{\mathcal{K}} \right\} \quad (154) \\ & \leq 1 \left\{ \mathbf{B}_i^{[K]}(1^{[K]}) = 1^{[K]} \right\} \quad (155) \\ & \quad + \sum_{\mathcal{K} \subset [K]} \exp \left( - \sum_{k \in \mathcal{K}^c} \beta_i^{\pi(k)} \right) 1 \left\{ \mathbf{B}_i^{\pi(\mathcal{K})}(1^{\mathcal{K}}) = 1^{\mathcal{K}} \right\}, \end{aligned}$$

where (155) follows from (150), (151) and (152).

Now, plugging (146) and (155) into (142) and computing the expectation in (142) with respect to the codebooks and the binning functions, we conclude that the probability of successful decoding is bounded below as

$$\begin{aligned} & 1 - \epsilon \geq \quad (156) \\ & \mathbb{E} \left[ \prod_{k=1}^K \prod_{i=1}^t \frac{1}{L_i^k \exp \left( \iota \left( Y_{[i]}^k; U_i^k | U_{[i-1]}^k \right) \right) + \left( 1 - \frac{1}{L_i^k} \right)} \right. \\ & \quad \left. \cdot \frac{\mathbf{g} \left( U_{[i]}^{[K]} \right) 1 \left\{ \mathbf{d} \left( X_i, \hat{X}_i \left( U_{[i]}^{[K]} \right) \right) \leq d_i \right\}}{1 + \sum_{\mathcal{K} \subset [K]} \exp \left( - \sum_{k \in \mathcal{K}^c} \beta_i^k \right)} \right]. \end{aligned}$$

*Loosening the bound (156):* Here we again follow the recipe of Yassaee et al. [25], [26].

$$\begin{aligned} & 1 - \epsilon \geq \\ & \mathbb{E} \left[ \prod_{k=1}^K \prod_{i=1}^t \frac{1}{(L_i^k)^{-1} \exp \left( \iota \left( Y_{[i]}^k; U_i^k | U_{[i-1]}^k \right) \right) + 1} \right. \\ & \quad \left. \cdot \frac{\mathbf{g} \left( U_{[i]}^{[K]} \right) 1 \left\{ \mathbf{d} \left( X_i, \hat{X}_i \left( U_{[i]}^{[K]} \right) \right) \leq d_i \right\}}{\sum_{\mathcal{K} \subset [K]} \exp \left( - \sum_{k \in \mathcal{K}} \beta_i^k \right)} \right] \quad (157) \\ & \geq \prod_{k=1}^K \prod_{i=1}^t \frac{\mathbb{P}[\mathcal{E}^c]}{[1 + \exp(-\alpha_i^k)] \left[ \sum_{\mathcal{K} \subset [K]} \exp \left( - \sum_{k \in \mathcal{K}} \beta_i^k \right) \right]} \quad (158) \end{aligned}$$

where (157) holds by weakening (156) using  $1 - (L_i^k)^{-1} \leq 1$  and rewriting for brevity

$$1 + \sum_{\mathcal{K} \subset [K]} \exp \left( - \sum_{k \in \mathcal{K}^c} \beta_i^k \right) = \sum_{\mathcal{K} \subset [K]} \exp \left( - \sum_{k \in \mathcal{K}} \beta_i^k \right); \quad (159)$$

(158) is obtained by weakening (157) by multiplying the random variable inside the expectation by  $1 \{\mathcal{E}^c\}$  and using

the conditions in  $\mathcal{E}$  (40) to upper-bound  $\iota \left( Y_{[i]}^k; U_i^k | U_{[i-1]}^k \right)$  in the denominator.

Rewriting (158), we obtain

$$\epsilon \leq 1 - \quad (160)$$

$$\begin{aligned} & \prod_{k=1}^K \prod_{i=1}^t \frac{\mathbb{P}[\mathcal{E}^c]}{[1 + \exp(-\alpha_i^k)] \left[ \sum_{\mathcal{K} \subset [K]} \exp \left( - \sum_{k \in \mathcal{K}} \beta_i^k \right) \right]} \\ & = \mathbb{P}[\mathcal{E}] + \gamma \mathbb{P}[\mathcal{E}^c] \quad (161) \\ & \leq \mathbb{P}[\mathcal{E}] + \gamma. \quad (162) \end{aligned}$$

## APPENDIX B

### PROOF OF THEOREM 3

We analyze the bound in Theorem 2 with

$$P_{U_{[t]}^k \| Y_{[t]}^k} = P_{U_{[t]}^k \| Y_{[t]}^k}, \quad (163)$$

$$P_{\hat{X}_{[t]}^{[K]} \| U_{[t]}^{[K]}} = P_{\hat{X}_{[t]}^{[K]} \| U_{[t]}^{[K]}}, \quad (164)$$

single-letter kernels chosen so that

$$\mathbb{E} \left[ \mathbf{d} \left( X_i, \hat{X}_i \left( U_{[i]}^{[K]} \right) \right) \right] = d_i + \delta, \quad (165)$$

for some  $\delta > 0$ . We also fix an arbitrary permutation  $\pi: [K] \mapsto [K]$ . Denote for brevity the divergences

$$D_i^{\pi(k)} \triangleq \mathbb{E} \left[ J^{\pi(k)} \left( U_{[i]}^{\pi([K])} \right) \right] \quad (166)$$

$$= D \left( P_{U_i^{\pi(k)} | U_i^{\pi([K])}} \| P_{U_i^{\pi(k)} | U_{[i-1]}^{\pi(k)}} | P_{U_i^{\pi(k-1)} | U_{[i-1]}^{\pi([K])}} \right)$$

For  $k \in [K]$ ,  $i \in [t]$ , let

$$\alpha_i^k = \beta_i^k = n\delta, \quad (167)$$

and choose  $L_i^k$ ,  $M_i^k$  to satisfy

$$\log L_i^k \geq n I \left( Y_{[i]}^k; U_i^k | U_{[i-1]}^k \right) + 2\alpha_i^k, \quad (168)$$

$$\log M_i^{\pi(k)} \geq \log L_i^{\pi(k)} - n D_i^{\pi(k)} + 2\beta_i^k. \quad (169)$$

Note that since  $U_i^k - \left( Y_{[i]}^k, U_{[i-1]}^k \right) - U_{[i]}^{[K] \setminus \{k\}}$ , it holds that

$$\begin{aligned} & I \left( Y_{[i]}^{\pi(k)}; U_i^{\pi(k)} | U_i^{\pi([k-1])}, U_{[i-1]}^{\pi([K])} \right) \\ & = I \left( Y_{[i]}^{\pi(k)}; U_i^{\pi(k)} | U_{[i-1]}^{\pi(k)} \right) - D_i^{\pi(k)}, \quad (170) \end{aligned}$$

and thus summing both sides of (169) over  $i \in [t]$  we obtain (cf. (43))

$$\begin{aligned} & \frac{1}{n} \sum_{i=1}^t \log M_i^k \geq I \left( Y_{[t]}^{\pi(k)} \rightarrow U_{[t]}^{\pi(k)} \| U_{[t]}^{\pi([k-1])}, \mathcal{D}U_{[t]}^{[K]} \right) \\ & \quad + 4t\delta. \quad (171) \end{aligned}$$

Applying the union bound to  $\mathbb{P}[\mathcal{E}]$  and the law of large numbers to each of the resultant  $(2K+1)t$  terms, we further conclude that  $\mathbb{P}[\mathcal{E}] \rightarrow 0$  as  $n \rightarrow \infty$ . Furthermore,  $\gamma \rightarrow 0$  as  $n \rightarrow \infty$ , and therefore by Theorem 2 there exists a sequence of codes with  $\log L_i^k$  and  $\log M_i^k$  satisfying (168), (169) with excess-distortion probability  $\epsilon \rightarrow 0$  as  $n \rightarrow \infty$ .

Under our assumption on the  $p$ -th moment of the distortion measure (20), the existence of an  $(M_{[t]}^{[K]}, d_{[t]}, \epsilon)$  excess-distortion code with  $\frac{1}{t} \sum_{i=1}^t d_i \leq d$  implies the existence of an  $(M_{[t]}^{[K]}, d(1-\epsilon) + d_p \epsilon^{1-1/p})$  average distortion code via a standard argument using Hölder's inequality [39, Th. 25.5].

## APPENDIX C

### TWO CHARACTERIZATIONS OF BERGER-TUNG BOUND

**Proposition 2.** *The region  $\mathcal{R}$  in (48) is equivalent to the region  $\mathcal{R}'$  in (47).*

*Proof of Proposition 2.* Observe that any subset  $\mathcal{A}$  of  $[K]$  with cardinality  $k$  is equal to  $\pi([k])$ , for some permutation  $\pi$  on  $[K]$ .

First, we show that  $\mathcal{R}' \subseteq \mathcal{R}$ . Fix  $\pi$  and consider  $\mathcal{K} = \pi([k])$ . Since given  $Y_k, U_k$  is independent of  $\cup^{[K] \setminus \{k\}}$ .

$$I(Y^{\mathcal{K}}; U^{\mathcal{K}}) = \sum_{j=1}^k I(Y_{\pi(j)}; U_{\pi(j)} | U^{\pi[j-1]}) \quad (172)$$

$$I(Y^{\mathcal{K}^c}; U^{\mathcal{K}^c} | U^{\mathcal{K}}) = \sum_{j=k+1}^K I(Y_{\pi(j)}; U_{\pi(j)} | U^{\pi[j-1]}) \quad (173)$$

From (173), we conclude that any set of rates that satisfies (47) for  $\pi$  must also satisfy (48) for  $\mathcal{A} = \mathcal{K}^c$ . Thus,  $\mathcal{R}' \subseteq \mathcal{R}$ .

To show that  $\mathcal{R} \subseteq \mathcal{R}'$ , note, using the operational Markov chain condition  $U^{\mathcal{B}} - Y^{\mathcal{B}} - Y^{\mathcal{A} \setminus \mathcal{B}} - U^{\mathcal{A} \setminus \mathcal{B}}$ , that for all  $\mathcal{B} \subseteq \mathcal{A}$ ,

$$I(Y^{\mathcal{A}}; U^{\mathcal{A}}) = I(Y^{\mathcal{A} \setminus \mathcal{B}}; U^{\mathcal{A} \setminus \mathcal{B}} | U^{\mathcal{B}}) + I(Y^{\mathcal{B}}; U^{\mathcal{B}}). \quad (174)$$

Since

$$\begin{cases} S_1 \geq I_1 \\ S_1 + S_2 \geq I_1 + I_2 \end{cases} \iff \begin{cases} S_1 \geq I_1 \\ S_2 \geq I_2 \end{cases}, \quad (175)$$

(174) implies that for any  $\mathcal{A} \subseteq [K]$ ,

$$\begin{cases} \sum_{k \in \mathcal{A}^c} R_k \geq I(Y^{\mathcal{A}^c}; U^{\mathcal{A}^c} | U^{\mathcal{A}}) \\ \sum_{k \in [K]} R_k \geq I(Y^{[K]}; U^{[K]}) \end{cases} \quad (176)$$

$$\iff \begin{cases} \sum_{k \in \mathcal{A}} R_k \geq I(Y^{\mathcal{A}}; U^{\mathcal{A}}) \\ \sum_{k \in \mathcal{A}^c} R_k \geq I(Y^{\mathcal{A}^c}; U^{\mathcal{A}^c} | U^{\mathcal{A}}) \end{cases} \quad (177)$$

and for any  $\mathcal{B} \subseteq \mathcal{A}$ ,

$$\begin{cases} \sum_{k \in \mathcal{B}} R_k \geq I(Y^{\mathcal{B}}; U^{\mathcal{B}}) \\ \sum_{k \in \mathcal{A}} R_k \geq I(Y^{\mathcal{A}}; U^{\mathcal{A}}) \end{cases} \quad (178)$$

$$\iff \begin{cases} \sum_{k \in \mathcal{B}} R_k \geq I(Y^{\mathcal{B}}; U^{\mathcal{B}}) \\ \sum_{k \in \mathcal{A} \setminus \mathcal{B}} R_k \geq I(Y^{\mathcal{A} \setminus \mathcal{B}}; U^{\mathcal{A} \setminus \mathcal{B}} | U^{\mathcal{B}}) \end{cases} \quad (179)$$

For  $\mathcal{B} = \pi([k-1])$  and  $\mathcal{A} = \pi([k])$ , the second inequality in (179) is exactly the inequality (47). Since any set of rates satisfying (48) must also satisfy (179) for all  $\mathcal{B} \subseteq \mathcal{A} \subseteq [K]$ , we conclude that  $\mathcal{R} \subseteq \mathcal{R}'$ .  $\square$

## APPENDIX D

### MMSE ESTIMATION LEMMAS

Lemmas 2 and 3 are corollaries to the following result.

**Lemma 4.** *Let  $X \sim \mathcal{N}(0, \sigma_X^2)$ , and let*

$$Y_k = X + W_k, \quad k = 1, \dots, K, \quad (180)$$

where  $W_k \sim \mathcal{N}(0, \sigma_{W_k}^2)$ ,  $W_k \perp W_j$ ,  $j \neq k$ . Then, the MMSE estimate and the normalized estimation error of  $X$  given  $Y_{[K]}$  are given by

$$\mathbb{E}[X|Y_{[K]}] = \sum_{k=1}^K \frac{\sigma_{X|Y_{[K]}}^2}{\sigma_{W_k}^2} Y_k, \quad (181)$$

$$\frac{1}{\sigma_{X|Y_{[K]}}^2} = \frac{1}{\sigma_X^2} + \sum_{k=1}^K \frac{1}{\sigma_{W_k}^2}. \quad (182)$$

*Proof of Lemma 4.* The result is well known; we provide a proof for completeness. For jointly Gaussian random vectors  $X, Y$ ,

$$\mathbb{E}[X|Y = y] = \mathbb{E}[X] + \Sigma_{XY} \Sigma_{YY}^{-1} (y - \mathbb{E}[Y]), \quad (183)$$

$$\text{Cov}[X|Y] = \Sigma_{XX} - \Sigma_{XY} \Sigma_{YY}^{-1} \Sigma_{YX}. \quad (184)$$

Denote for brevity

$$\Sigma_W \triangleq \begin{bmatrix} \sigma_{W_1}^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_{W_K}^2 \end{bmatrix} \quad (185)$$

In our case,  $X$  is a scalar and  $Y = Y_{[K]}$  is a vector, and

$$\Sigma_{XX} = \sigma_X^2 \quad (186)$$

$$\Sigma_{YY} = \Sigma_W + \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \sigma_X^2 [1 \quad \dots \quad 1] \quad (187)$$

$$\Sigma_{XY} = \sigma_X^2 [1 \quad \dots \quad 1] \quad (188)$$

Using the matrix inversion lemma, we compute readily

$$\text{Cov}[X|Y]^{-1} = \Sigma_{XX}^{-1} - \Sigma_{XX}^{-1} \Sigma_{XY} (\Sigma_{YX} \Sigma_{XX}^{-1} \Sigma_{XY} - \Sigma_{YY})^{-1} \Sigma_{YX} \Sigma_{XX}^{-1} \quad (189)$$

$$= \Sigma_{XX}^{-1} + \Sigma_{XX}^{-1} \Sigma_{XY} \Sigma_W^{-1} \Sigma_{YX} \Sigma_{XX}^{-1} \quad (190)$$

$$= \frac{1}{\sigma_X^2} + \frac{1}{\sigma_{W_1}^2} + \dots + \frac{1}{\sigma_{W_K}^2}, \quad (191)$$

which shows (182). To show (181), we apply the matrix inversion lemma to  $\Sigma_{YY}$  to write:

$$\Sigma_{YY}^{-1} = \Sigma_W^{-1} - \Sigma_W^{-1} \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \sigma_{X|Y_{[K]}}^2 [1 \quad \dots \quad 1] \Sigma_W^{-1}$$

It's easy to verify that

$$\sigma_X^2 [1 \quad \dots \quad 1] \left( \frac{1}{\sigma_{X|Y_{[K]}}^2} - \Sigma_W^{-1} \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} [1 \quad \dots \quad 1] \right) = [1 \quad \dots \quad 1], \quad (192)$$

so

$$\mathbb{E}[X|Y = y] = \Sigma_{XY} \Sigma_{YY}^{-1} y \quad (193)$$

$$= [1 \quad \dots \quad 1] \Sigma_W^{-1} \sigma_{X|Y_{[K]}}^2 y, \quad (194)$$

which is equivalent to (181).  $\square$

*Proof of Lemma 2.* Equality (69) follows from

$$\sigma_Y^2 = \sigma_X^2 + \sigma_W^2, \quad (195)$$

$$\frac{1}{\sigma_{X|Y}^2} = \frac{1}{\sigma_X^2} + \frac{1}{\sigma_W^2}, \quad (196)$$

where (196) is a particularization of (182).  $\square$

*Proof of Lemma 3.* Notice that (70) with  $\bar{X}_k = \mathbb{E}[X|Y_k]$  and  $W'_k \sim \mathcal{N}(0, \sigma_{X|Y_k}^2)$  is just another way to write (180). Reparameterizing (181) and (182) accordingly, one recovers (71) and (72).  $\square$

*Remark 1.* We may use Lemma 4 to derive the Kalman filter for the estimation of  $X_i$  (2) given the history of observations  $Y_{[i]}^{[K]}$  (3):

$$\bar{X}_i = a\bar{X}_{i-1} + \sum_{k=1}^K \frac{\sigma_{X_i|Y_{[i]}^{[K]}}^2}{\sigma_{W_k}^2} (Y_i^k - a\bar{X}_{i-1}), \quad (197)$$

$$\frac{1}{\sigma_{X_i|Y_{[i]}^{[K]}}^2} = \frac{1}{\sigma_{X_i|Y_{[i-1]}^{[K]}}^2} + \sum_{k=1}^K \frac{1}{\sigma_{W_k}^2}. \quad (198)$$

where  $\bar{X}_i$  is defined in (101). Equation (197) is the Kalman filter recursion with Kalman filter gain equal to the row vector  $\sigma_{X_i|Y_{[i]}^{[K]}}^2 \left( \frac{1}{\sigma_{W_1}^2}, \dots, \frac{1}{\sigma_{W_K}^2} \right)$ , and (198) is the corresponding Riccati recursion for the MSE.

#### APPENDIX E

##### TWO EQUIVALENT REPRESENTATIONS OF $R_{\text{rm}}(d)$

In this appendix, we verify that (117) coincides with the lower bound on the causal remote rate-distortion function derived in [21]. Indeed, [21, Cor. 1 and Th. 9] imply

$$R_{\text{rm}}(d) \geq \frac{1}{2} \log \left( a^2 + \frac{\sigma_{X_{\parallel}^2|\mathcal{D}Y^{[K]}} - \sigma_{X_{\parallel}^2|Y^{[K]}}}{d - \sigma_{X_{\parallel}^2|Y^{[K]}}} \right). \quad (199)$$

Here,  $\sigma_{X_{\parallel}^2|\mathcal{D}Y^{[K]}} - \sigma_{X_{\parallel}^2|Y^{[K]}}$  is the variance of the innovations of the Gauss-Markov process  $\{\bar{X}_i\}$ , i.e.

$$\bar{X}_{i+1} = a\bar{X}_i + \bar{V}_i, \quad (200)$$

$\bar{V}_i \sim \mathcal{N}(0, \sigma_{X_{\parallel}^2|\mathcal{D}Y^{[K]}} - \sigma_{X_{\parallel}^2|Y^{[K]}})$ . The form in (199) leads to that in (117) via (54) and

$$\sigma_{X_{\parallel}^2|\mathcal{D}Y^{[K]}} = a^2 \sigma_{X_{\parallel}^2|Y^{[K]}} + \sigma_{\bar{V}}^2. \quad (201)$$

□