The CEO problem with inter-block memory

Victoria Kostina, Babak Hassibi

***Abstract*—An n-dimensional source with memory is observed by K isolated encoders via parallel channels, who compress their observations to transmit to the decoder via noiseless rate-constrained links while leveraging their memory of the past. 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 the memory of the past.**

We extend the Berger-Tung inner and outer bounds to the scenario with inter-block memory, showing that the minimum asymptotically (as $n \to \infty$) achievable sum rate required to achieve a target distortion is bounded by minimal directed mutual information problems. For the Gauss-Markov source observed via K parallel AWGN channels, we show that the inner bound is tight and solve the corresponding 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, directed information.**

I. Introduction

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

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

where t is the *time horizon* over which the source is being tracked, and $\mathsf{d}\colon \mathcal{A}^n \times \hat{\mathcal{A}}^n \mapsto \mathbb{R}_+$ is the distortion measure. Encoding and decoding operations leverage the memory of the past but cannot look in the future. In this causal setting no delay is allowed neither at the encoders in producing codewords to encode X_i nor at the decoder in producing $\hat{X}_i$.

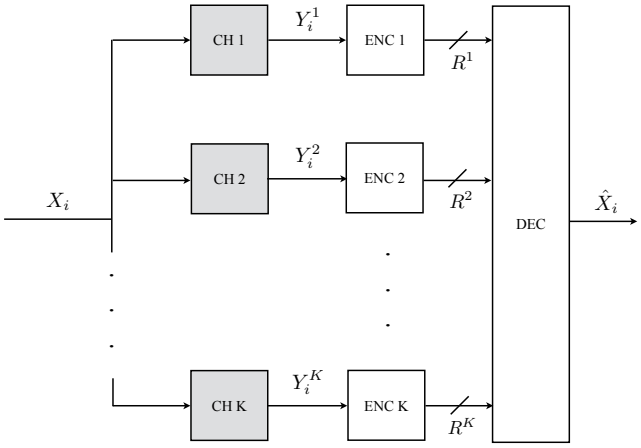

Fig. 1. The CEO problem with inter-block memory: the encoders and the decoder keep the memory of their past observations.

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, an i.i.d. Gaussian source is observed via AWGN channels and reproduced under mean squared 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 Prabharan et al. [5] and Oohama [6], who determined the full Gaussian CEO rate region, by Chen et al. [7], who proved that the minimum sum rate is achieved via waterfilling, by Behroozi and Soleymani [8] and by Chen and Berger [9], who showed rate-optimal successive coding schemes. Wagner et al. [10] found the rate region of the distributed Gaussian lossy compression problem by coupling it to the Gaussian CEO problem. Wagner and Anantharam [11] showed an outer bound to the rate region of the multiterminal source coding problem that is tighter than the Berger-Tung outer bound [12], [13]. Wang et al. [14] showed a simple converse on the sum rate of the vector Gaussian CEO problem. Concurrently, Ekrem and Ulukus [15] and Wang and Chen [16] 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 [11] to the Gaussian case. Courtade and Weissman [17] determined

The authors are with California Institute of Technology (e-mail: vkostina@caltech.edu, 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 presented at ISIT 2020 [1].

the distortion region of the distributed source coding and the CEO problem under logarithmic loss.

None of the above results directly apply to the tracking problem in Fig. 1 because of the past memory in encoding the n-blocks of observations and in producing $\hat{X}_i$ in (1), which imposes blockwise causality constraints onto the coding process. The most basic scenario of source coding with causality constraints is that of a single observer directly seeing the information source [18]. The causal rate-distortion function for the Gauss-Markov source was computed by Gorbunov and Pinsker [19]. The link between the minimum attainable linear quadratic Gaussian (LQG) control cost and the causal rate-distortion function is elucidated in [20]–[22]. A semidefinite program to compute the causal rate-distortion function for vector Gauss-Markov sources is provided in [23]. The remote Gaussian causal rate-distortion function, which corresponds to setting $K = 1$ in Fig. 1, is computed in [22]. The causal rate-distortion function of the Gauss-Markov source with a Gaussian side observation available at the decoder (the causal counterpart of the Wyner-Ziv setting) is computed in [24] for the scalar source and in [25] for the vector source. That causal Wyner-Ziv setting can be viewed a special case of our causal CEO problem (2), (3) with two observers, with one of the observers enjoying an infinite rate. Stability of linear Gaussian systems with multiple isolated observers is investigated in [26].

The first contribution of this paper is an extension of the Berger-Tung inner and outer bounds [12], [13] to the distributed tracking setting of Fig. 1 that sandwich the minimum asymptotically achievable (as $n \to \infty$) sum rate $R^1 + \ldots + R^K$ required to achieve a given average distortion (1). Provided that the components of each $X_i \in \mathcal{A}^n$ are i.i.d. (X_i can still depend on $X_1, \ldots, X_{i-1}$), the channels act on each of those components independently, and the distortion measure is separable, that minimum sum rate is bounded in terms of the directed mutual information from the encoders to the decoder. The converse (outer bound) follows via standard data processing and single-letterization arguments. To prove the achievability, we show a nonasymptotic bound for blockwise-causal distributed lossy source coding that can be viewed as an extension of the nonasymptotic Berger-Tung inner bound by Yassaee et al. [27], [28], applicable to the setting with $K = 2$ sources and $t = 1$ rounds of communication, to the setting with an arbitrary number of sources and communication rounds. 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. [27], [28] to perform encoding operations. The SLC-based encoder mimics the operation of the joint typicality encoder while admitting sharp nonasymptotic bounds on its performance. While the SLC-based decoder of [27], [28] is ill-suited to the case $K > 2$, we propose a novel decoder that falls into the class of generalized likelihood decoders [29] 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 inner bound [12], [13] to the setting with inter-block memory.

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,

$$X_{i+1} = aX_i + V_i, \tag{2}$$

and the k-th observer sees

$$Y_i^k = X_i + W_i^k, \quad k = 1, \ldots, K, \tag{3}$$

where X_1 and $\{V_i, W_i^1, W_i^2, \ldots, W_i^K\}_{i=1}^T$ are independent Gaussian vectors of length n with i.i.d. components; each component of V_i is distributed as $\mathcal{N}(0, \sigma_{\mathsf{V}}^2)$, and each component of W_i^k as $\mathcal{N}(0, \sigma_{\mathsf{W}_k}^2)$. Note that different observation channels can have different noise powers. The distortion measure is the normalized squared error

$$\mathsf{d}\left(X_i, \hat{X}_i\right) = \frac{1}{n}\|X_i - \hat{X}_i\|^2. \tag{4}$$

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. Similar to the corresponding result for $t = 1$ [5], [6], [30, Th. 12.3], our extension of the Berger-Tung inner bound is tight in this case. To compute the bound, we split up the directed minimal mutual information problem 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. [14] for the time horizon $t = 1$, to $t > 1$. A device that helps us track the behavior of optimal estimation errors over multiple time instances is the reversal of the channels from $\{X_i\}$ to $\{Y_i^k\}$:

$$X_i = \bar{X}_i^k + W_i^{k\prime}, \tag{5}$$

where

$$\bar{X}_i^k \triangleq \mathbb{E}\left[X_i | Y_1^k, \ldots, Y_i^k\right], \tag{6}$$

and $W_i^{k\prime} \perp \bar{X}_i^k$ 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^k or $\bar{X}_1$ since the latter is just a scaled version of the former, for $t > 1$, compressing Y_i^k 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 minimum MSE in the estimation of $\{X_i\}$ from $\{Y_j^k\}_{j \le i, k=1,\ldots,K}$, the identical channels case is the hardest to compress. This result contributes to the discussions of the rate loss in the classical CEO [31, Cor. 1] and multiple descriptions [32, Lemma 3] problems.

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 bounds to the minimum sum rate in terms of minimal directed mutual information problems (Theorem 1). In Section III, we characterize the causal Gaussian CEO 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, $[M] \triangleq \{1, \ldots, M\}$. Notation $X \leftarrow Y$ reads "replace X by Y"; notation $X \perp Y$ reads "X is independent of Y"; notation $\triangleq$ reads "by definition". The temporal index is indicated in the subscript and the spatial index in the superscript: $Y_{[t]}^k$ is the temporal vector $(Y_1^k, \ldots, Y_t^k)$; $Y_i^{[K]}$ is the spatial vector $(Y_i^1, \ldots, Y_i^K)^{\mathsf{T}}$; $Y_{[t]}^{[K]} \triangleq (Y_{[t]}^1, \ldots, Y_{[t]}^K)$. Delay operator $\mathcal{D}$ acts as $\mathcal{D}X_{[t]} \triangleq (0, X_1, \ldots, X_{t-1})$. For a random vector X with i.i.d. components, X denotes a random variable distributed the same as each component of X. We adopt the following shorthand notation for causally conditional [33] probability kernels:

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

Given a distribution $P_{X_{[t]}}$ and a causal kernel $P_{Y_{[t]}\|X_{[t]}}$, the directed mutual information is defined as [34]

$$I\left(X_{[t]} \to Y_{[t]}\right) \triangleq \sum_{i=1}^{t} I\left(X_{[i]}; Y_i | Y_{[i-1]}\right). \tag{8}$$

II. Sum rate via directed information

A. Overview

In this section, we present and prove our extension of the Berger-Tung bounds to the setting inter-block memory that sandwich the minimum achievable sum rate in terms of minimal directed mutual information problems. The bounds apply to an abstract source with abstract observations. The operational scenario and achievable rates are formally defined in Section II-B. The directed mutual information bounds are presented in Section II-C. The converse is proven in Section II-D. The nonasymptotic achievability bound and its asymptotic analysis are presented in Section II-E. A set of remarks in Section II-F completes Section II.

B. Operational problem setting

A CEO code with inter-block memory, or a causal CEO code, is formally defined as follows.

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

$$P_{Y_{[t]}^k\|X_{[t]}} \colon \mathcal{X}^{\otimes t} \mapsto \mathcal{Y}^{\otimes t}, \quad k \in [K]. \tag{9}$$

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

A CEO code with inter-block memory consists of:

a) K encoding policies

$$P_{B_{[t]}^k\|Y_{[t]}^k} : \mathcal{Y}^{\otimes t} \mapsto \prod_{i=1}^{t} [M_i^k], \quad k \in [K], \tag{10}$$

b) a decoding policy

$$P_{\hat{X}_{[t]}^{[K]}\|B_{[t]}^{[K]}} \colon \prod_{i=1}^{t} [M_i^k] \mapsto \hat{\mathcal{X}}^{\otimes t}. \tag{11}$$

If the encoding and decoding policies satisfy

$$\frac{1}{t}\sum_{i=1}^{t} \mathbb{E}\left[\mathsf{d}\left(X_i, \hat{X}_i\right)\right] \le d, \tag{12}$$

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

If the encoding and decoding policies satisfy

$$\mathbb{P}\left[\bigcup_{i=1}^{t}\left\{\mathsf{d}\left(X_i, \hat{X}_i\right) > d_i\right\}\right] \le \epsilon, \tag{13}$$

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

The probability measure in (12) and (13) 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 $\mathsf{d}_n \colon \mathcal{A}^n \times \hat{\mathcal{A}}^n \mapsto \mathbb{R}_+$ is called *separable* if

$$\mathsf{d}_n(x, \hat{x}) = \frac{1}{n}\sum_{i=1}^{n} \mathsf{d}(x(i), \hat{x}(i)), \tag{14}$$

where $\mathsf{d} \colon \mathcal{A} \times \hat{\mathcal{A}} \mapsto \mathbb{R}_+$, 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 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 (9).*

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

$$\frac{1}{nt}\sum_{i=1}^{t} \log M_i^k \le R^k, \quad k \in [K]. \tag{15}$$

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

$$\sum_{k=1}^{K} R^k \le R \tag{16}$$

is asymptotically achievable.

The causal CEO rate-distortion function *at time horizon t is defined as follows:*

$$R_{t\,\mathrm{CEO}}(d) \triangleq \inf\Big\{R\colon \ (R, d) \ \textit{is achievable} \ \textit{at time horizon } t \textit{ in the CEO problem.}\Big\} \tag{17}$$

C. Berger-Tung bounds with inter-block memory

Consider a discrete-time random process $\{X_i\}_{i=1}^t$ on $\mathcal{X} = \mathcal{A}^n$ equipped with separable distortion measure d, observed by K causal observers via the channels (9) with $\mathcal{Y} = \mathcal{B}^n$ and

$$P_{X_i|X_{[i-1]}} = P_{\mathsf{X}_i|\mathsf{X}_{[i-1]}}^{\otimes n} \tag{18}$$

$$P_{Y_i^k|X_{[i]},Y_{[i-1]}^k} = P_{\mathsf{Y}_i^k|\mathsf{X}_{[i]},\mathsf{Y}_{[i-1]}^k}^{\otimes n}. \tag{19}$$

Denote the minimal directed mutual information problems

$$\overline{\mathsf{R}}_{t\,\mathrm{CEO}}(d) \triangleq \inf_{\substack{P_{\mathsf{U}_{[t]}^{[K]}\|\mathsf{Y}_{[t]}^{[K]}} : (22) \\ P_{\hat{\mathsf{X}}_{[t]}\|\mathsf{U}_{[t]}^{[K]}} : (24)}} I\left(\mathsf{Y}_{[t]}^{[K]} \to \mathsf{U}_{[t]}^{[K]}\right) \tag{20}$$

$$\underline{\mathsf{R}}_{t\,\mathrm{CEO}}(d) \triangleq \inf_{\substack{P_{\mathsf{U}_{[t]}^{[K]}\|\mathsf{Y}_{[t]}^{[K]}} : (23) \\ P_{\hat{\mathsf{X}}_{[t]}\|\mathsf{U}_{[t]}^{[K]}} : (24)}} I\left(\mathsf{Y}_{[t]}^{[K]} \to \mathsf{U}_{[t]}^{[K]}\right) \tag{21}$$

where the constraints are as follows:

$$P_{\mathsf{U}_{[t]}^{[K]}\|\mathsf{Y}_{[t]}^{[K]}} = \prod_{k=1}^K P_{\mathsf{U}_{[t]}^k\|\mathsf{Y}_{[t]}^k} \tag{22}$$

$$P_{\mathsf{U}_{[t]}^k\|\mathsf{Y}_{[t]}^{[K]}} = P_{\mathsf{U}_{[t]}^k\|\mathsf{Y}_{[t]}^k} \quad \forall k \in [K] \tag{23}$$

$$\frac{1}{t}\sum_{i=1}^t \mathbb{E}\left[\mathsf{d}\left(\mathsf{X}_i, \hat{\mathsf{X}}_i\right)\right] \le d. \tag{24}$$

Fixing a $k \in [K]$ and marginalizing $\{U_{[t]}^{k'},\ k' \neq k\}$ out of both sides of (22), one can see that any joint distribution that satisfies the separate encoding constraint (22) also satisfies (23). Thus, the optimization problems (20) and (21) differ in that the constraint (22) is more stringent than (23). They represent extensions of the Berger-Tung inner ((20)) and outer ((21)) bounds [30, Th. 12.1, 12.2] to the causal setting.

One can convexify $\overline{\mathsf{R}}_{t\,\mathrm{CEO}}(d)$ by adding to the optimization parameters a scalar $\alpha \in (0, 1]$ and a distribution $P_{\tilde{\mathsf{U}}_{[t]}^{[K]}\|\tilde{\mathsf{Y}}_{[t]}^{[K]}}$ satisfying the separate encoding constraint analogous to (22), and replacing the directed information in (20) by $\alpha I\left(\mathsf{Y}_{[t]}^{[K]} \to \mathsf{U}_{[t]}^{[K]}\right) + (1-\alpha) I\left(\tilde{\mathsf{Y}}_{[t]}^{[K]} \to \tilde{\mathsf{U}}_{[t]}^{[K]}\right)$. This is equivalent to introducing into (20) a binary time sharing random variable. Given the achievability of (20), the achievability of the convexification follows by the standard time sharing argument [30, Ch. 4.4].

Since a mixture of distributions $P_{\mathsf{U}_{[t]}^{[K]}\|\mathsf{Y}_{[t]}^{[K]}}$ satisfying (23) also satisfies (23), the convexity of $\underline{\mathsf{R}}_{t\,\mathrm{CEO}}(d)$ follows from the convexity of directed mutual information in $P_{\mathsf{U}_{[t]}^{[K]}\|\mathsf{Y}_{[t]}^{[K]}}$, with no need for an explicit auxiliary time sharing random variable.

Theorem 1 (Berger-Tung bounds with inter-block memory)**.** *Consider a discrete-time random process $\{X_i\}_{i=1}^t$ on $\mathcal{X} = \mathcal{A}^n$ equipped with a separable distortion measure d, observed by K causal observers via the channels (9) with $\mathcal{Y} = \mathcal{B}^n$ and (18), (19) satisfied. Suppose further that for some $p > 1$, there exists a vector $\hat{\mathsf{x}}_{[t]}$ such that*

$$\left(\mathbb{E}\left[\left(\frac{1}{t}\sum_{i=1}^t \mathsf{d}(\mathsf{X}_i, \hat{\mathsf{x}}_i)\right)^p\right]\right)^{\frac{1}{p}} \le d_p < \infty. \tag{25}$$

The causal rate-distortion function is bounded as

$$\underline{\mathsf{R}}_{t\,\mathrm{CEO}}(d) \le R_{t\,\mathrm{CEO}}(d) \le \overline{\mathsf{R}}_{t\,\mathrm{CEO}}(d). \tag{26}$$

Condition (25) is a technical condition needed to apply a standard argument using Hölder's inequality to pass from an excess to average distortion in the proof of the achievability bound (Appendix B).

To prove the upper bound on the sum rate in (26), we actually show a more accurate characterization of the entire rate tuple $R^{[K]}$ (Theorem 3, below).

We will see in Section III below that the inner (upper) bound in (26) is tight in the quadratic Gaussian setting. This is in line with the corresponding result in the setting of block coding without inter-block memory [30, Th. 12.3].

While in general the t-step optimization problems (20) and (21) are challenging to compute, we illustrate in this paper that the normalized limit as $t \to \infty$ is possible to compute in the Gaussian setting. Similar limit results in other communication scenarios were shown in [19], [22], [24], [25], [35]–[38].

D. Theorem 1: proof of converse

The proof of the converse uses standard techniques. We will use the following definition and lemma.

Causally conditioned directed information is defined as

$$I(X_{[t]} \to Y_{[t]} \| Z_{[t]}) \triangleq \sum_{i=1}^t I(X_{[i]}; Y_i | Y_{[i-1]}, Z_{[i]}). \tag{27}$$

Lemma 1 ([33, (3.14)–(3.16)])**.** *Directed information chain rules:*

$$\begin{aligned} I((X_{[t]}, Y_{[t]}) \to Z_{[t]}) &= I(X_{[t]} \to Z_{[t]}) \\ &\quad + I(Y_{[t]} \to Z_{[t]} \| X_{[t]}), \end{aligned} \tag{28}$$

$$\begin{aligned} I(X_{[t]} \to (Y_{[t]}, Z_{[t]})) &= I(X_{[t]} \to Y_{[t]} \| \mathcal{D} Z_{[t]}) \\ &\quad + I(X_{[t]} \to Z_{[t]} \| Y_{[t]}). \end{aligned} \tag{29}$$

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 (16),

$$nR \ge \sum_{k=1}^K H(B_{[t]}^k) \tag{30}$$

$$\ge H\left(B_{[t]}^{[K]}\right) \tag{31}$$

$$\ge I\left(Y_{[t]}^{[K]} \to B_{[t]}^{[K]}\right) \tag{32}$$

$$\ge \inf_{\substack{P_{B_{[t]}^{[K]}\|Y_{[t]}^{[K]}} = \prod_{k=1}^K P_{B_{[t]}^k\|Y_{[t]}^k}, \\ P_{\hat{X}_{[t]}^{[K]}\|B_{[t]}^{[K]}} : (12) \text{ holds}}} I\left(Y_{[t]}^{[K]} \to B_{[t]}^{[K]}\right), \tag{33}$$

where (31) holds because the joint entropy is upper-bounded by the sum of individual entropies, and (32) holds because the mutual information is upper-bounded by the entropy. Note that (33) is the n-letter version of (20).

We proceed to apply a standard single-letterization argument to (33). 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 $\left(Y_i^k \colon k \in \mathcal{K}\right)$ indexed by $\mathcal{I}$.

Denote by $\mathsf{R}_{t\,\mathrm{CEO}}(d)$ the right-hand side of (26). We introduce auxiliary random objects

$$U_i^k(j) = \left(B_i^k, Y_i^{[K]}([j-1])\right), \quad j \in [n] \tag{34}$$

The directed mutual information in the right side of (33) can be rewritten in terms of $U_i^{[K]}$ and bounded as follows.

$$\begin{aligned}
& I\left(Y_{[t]}^{[K]} \to B_{[t]}^{[K]}\right) \\
&= \sum_{j=1}^{n} I\left(Y_{[t]}^{[K]}(j) \to B_{[t]}^{[K]} \| Y_{[t]}^{[K]}([j-1])\right) && (35) \\
&= \sum_{j=1}^{n} I\left(Y_{[t]}^{[K]}(j) \to \left(B_{[t]}^{[K]}, Y_{[t]}^{[K]}([j-1])\right)\right) \\
&\quad - I\left(Y_{[t]}^{[K]}(j) \to Y_{[t]}^{[K]}([j-1]) \| \mathcal{D} B_{[t]}^{[K]}\right) && (36) \\
&= \sum_{j=1}^{n} I\left(Y_{[t]}^{[K]}(j) \to \left(B_{[t]}^{[K]}, Y_{[t]}^{[K]}([j-1])\right)\right) && (37) \\
&= \sum_{j=1}^{n} I\left(Y_{[t]}^{[K]}(j) \to U_{[t]}^{[K]}(j)\right) && (38) \\
&\geq \min_{\substack{d_j, j \in [n]: \\ \sum d_j \leq nd}} \sum_{j=1}^{n} \underline{\mathsf{R}}_{t\,\mathrm{CEO}}(d_j) && (39) \\
&\geq n\, \underline{\mathsf{R}}_{t\,\mathrm{CEO}}(d) && (40)
\end{aligned}$$

where (35) is by the chain rule of mutual information; (36) is by the chain rule of directed information (29); (37) 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 (36) 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); (38) is by substituting (34); (39) holds because $U_i^k(j)$ (34) satisfies $P_{U_{[t]}^k(j)\|Y_{[t]}^{[K]}(j)} = P_{U_{[t]}^k(j)\|Y_{[t]}^k(j)}$, the distortion measure is separable and (18), (19) hold; and (40) is by the convexity of $\underline{\mathsf{R}}_{t\,\mathrm{CEO}}(d)$ as a function of d. □

E. Theorem 1: proof of achievability

To show that (26) is achievable in the asymptotics $n \to \infty$, we first show a nonasymptotic bound. Then, via an asymptotic analysis of the bound, we derive an extension of the Berger-Tung inner bound [12], [13] to the setting with inter-block memory.

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

$$\imath\left(y_{[i]}^k; u_i^k | u_{[i-1]}^k\right) \triangleq \log \frac{dP_{U_i^k|Y_{[i]}^k, U_{[i-1]}^k}\left(u_i^k | y_{[i]}^k, u_{[i-1]}^k\right)}{dP_{U_i^k|U_{[i-1]}^k}\left(u_i^k | u_{[i-1]}^k\right)}. \tag{41}$$

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

$$\jmath^k\left(u_{[i]}^{[K]}\right) \triangleq \log \frac{dP_{U_i^k|U_i^{[k-1]}U_{[i-1]}^{[K]}}\left(u_i^k \mid u_i^{[k-1]} u_{[i-1]}^{[K]}\right)}{dP_{U_i^k|U_{[i-1]}^k}\left(u_i^k \mid u_{[i-1]}^k\right)}. \tag{42}$$

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

$$\pi(\mathcal{K}) \triangleq \{\pi(k) \colon k \in \mathcal{K}\}. \tag{43}$$

Theorem 2 (nonasymptotic Berger-Tung inner bound with inter-block memory)**.** *Fix $P_{Y_{[t]}^{[K]}}$ and parameters $M_{[t]}^{[K]}, d_{[t]}^{[K]}, \epsilon$. For any scalars α_i^k, β_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 $\pi \colon [K] \mapsto [K]$, there exists an $(M_{[t]}^{[K]}, d_{[t]}, \epsilon)$ excess distortion CEO code with inter-block memory such that*

$$\epsilon \leq \mathbb{P}[\mathcal{E}] + \gamma, \tag{44}$$

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

$$\begin{aligned}
\mathcal{E} \triangleq & \bigcup_{i=1}^{t} \left\{\mathsf{d}\left(X_i, \hat{X}_i^k\right) > d_i\right\} && (45) \\
& \bigcup_{i=1}^{t} \bigcup_{k=1}^{K} \left\{\imath\left(Y_{[i]}^k; U_i^k | U_{[i-1]}^k\right) > \log L_i^k - \alpha_i^k\right\} \\
& \bigcup_{i=1}^{t} \bigcup_{k=1}^{K} \left\{\jmath^{\pi(k)}\left(u_{[i]}^{\pi([K])}\right) < \log \frac{L_i^{\pi(k)}}{M_i^{\pi(k)}} + \beta_i^{\pi(k)}\right\},
\end{aligned}$$

and constant γ is given by

$$\gamma \triangleq 1 - \frac{1}{\prod_{i=1}^{t}\left[\sum_{\mathcal{K} \subseteq K} \exp\left(-\sum_{k \in \mathcal{K}} \beta_i^k\right)\right] \prod_{k=1}^{K}\left[1 + \exp(-\alpha_i^k)\right]}. \tag{46}$$

Proof sketch. We employ the achievability proof technique developed by Yassaee et al. [27], [28] 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 [39, 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 (132) in Appendix A) that is also used to generate random codebooks.

While [28, 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 (136) in Appendix A, we propose a novel decoder that falls into the class of generalized likelihood decoders (GLD) conceptualized by Merhav [29, eq. (4)] and that uses an auxiliary indicator function $\mathsf{g}\left(u_{[i]}^{[K]}\right)$ (137). With our GLD we are able to recover the full Berger-Tung

region ((52), below) for any K. One can view the set of outcomes $u_{[i]}^{[K]}$ for which $\mathsf{g}\left(u_{[i]}^{[K]}\right)=1$ as a jointly typical set. That set depends on the choice of π 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 $[K]$.

Full details are given in Appendix A. □

Theorem 3 (Berger-Tung inner bound with inter-block memory)**.** *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 $P_{\mathsf{U}_{[t]}^{[K]}\|\mathsf{Y}_{[t]}^{[K]}}$, $P_{\hat{\mathsf{X}}_{[t]}^{[K]}\|\mathsf{U}_{[t]}^{[K]}}$ satisfying (22), (24) and some permutation $\pi\colon [K]\mapsto[K]$, it holds for all $k\in[K]$*

$$R^{\pi(k)} > I\left(\mathsf{Y}_{[t]}^{\pi(k)} \to \mathsf{U}_{[t]}^{\pi(k)} \| \mathsf{U}_{[t]}^{\pi([k-1])}, \mathcal{D}\mathsf{U}_{[t]}^{[K]}\right). \quad (47)$$

Proof. Appendix B. □

Theorem 3 implies that the sum rate

$$\sum_{k=1}^{K} R^k > I\left(\mathsf{Y}_{[t]}^{[K]} \to \mathsf{U}_{[t]}^{[K]}\right) \quad (48)$$

is achievable. Indeed, summing (47) over k and using $\mathsf{U}_i^k - \left(\mathsf{Y}_{[i]}^k, \mathsf{U}_{[i-1]}^k\right) - \mathsf{U}_{[i]}^{[K]\setminus\{k\}}$ leads to (48). Therefore, the sum rate in (21) is achievable. □

F. Remarks

We conclude Section II with a set of remarks.

1. Theorems 2 and 3 are easily extended to distributed source coding with inter-block memory, where the goal is to separately compress (and jointly decompress) K processes $\{Y_i^k\}$ under the individual distortion constraints

$$\frac{1}{t}\sum_{i=1}^{t}\mathbb{E}\left[\mathsf{d}^k(Y_i^k,\hat{Y}_i^k)\right] \le d^k, \quad k\in[K]. \quad (49)$$

Theorem 2 continues to hold with $\mathsf{d}\left(X_i,\hat{X}_i^k\right) > d_i$ in (45) replaced by $\mathsf{d}^k\left(Y_i^k,\hat{Y}_i^k\right) > d_i^k$. Consequently, Theorem 3 also continues to hold, replacing the constraint in (24) by

$$\frac{1}{t}\sum_{i=1}^{t}\mathbb{E}\left[\mathsf{d}^k(\mathsf{Y}_i^k,\hat{\mathsf{Y}}_i^k)\right] \le d^k, \quad k\in[K]. \quad (50)$$

2. Case $t=1$ corresponds to the classical CEO / distributed source coding problems. The region in (47) simplifies to

$$R^{\pi(k)} > I(\mathsf{Y}^{\pi(k)};\mathsf{U}^{\pi(k)}|\mathsf{U}^{\pi([k-1])}), \quad \forall k\in[K], \forall \text{ permutation } \pi\colon[K]\mapsto[K]. \quad (51)$$

The multiterminal Berger-Tung inner region is usually (e.g. [17, Def. 7], [5, eq. (2)]) specified as

$$\sum_{k\in\mathcal{A}} R^k > I(\mathsf{Y}^{\mathcal{A}};\mathsf{U}^{\mathcal{A}}|\mathsf{U}^{\mathcal{A}^c}), \quad \forall\mathcal{A}\subseteq[K]. \quad (52)$$

These characterizations are equivalent (Appendix C).

3. While the sum rate bound in (48) is the same regardless of the choice of permutation π, different π's in (47) 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. We chose to omit the time-sharing random variable in Theorem 3 for simplicity of presentation. It can be introduced in (47) using the standard time sharing argument [30, Ch. 4.4].

III. Gaussian rate-distortion function

A. Problem setup

This section focuses on the scenario of the Gauss-Markov source in (2) observed through the Gaussian channels in (3) under squared error 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 expected squared error is

$$\hat{X}_i \triangleq \mathbb{E}\left[X_i|B_{[i]}^{[K]}\right]. \quad (53)$$

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

$$R_{\mathrm{CEO}}(d) \triangleq \limsup_{t\to\infty} R_{t\,\mathrm{CEO}}(d). \quad (54)$$

In other words, the causal CEO rate-distortion function $R_{\mathrm{CEO}}(d)$ is the infimum of R's such that $\forall\gamma>0$, $\exists t_0\ge 0$ such that $\forall t\ge t_0$, $\exists n_0\in\mathbb{N}$ such that $\forall n\ge n_0$, an $\left(M_{[t]}^{[K]}, d+\gamma\right)$ average distortion causal CEO code exists with $M_{[t]}^{[K]}$ satisfying (15) and (16).

Taking the limit $t\to\infty$ simplifies the solution of many minimal directed mutual information problems ([22, Th. 9], [24, Th. 6, Th. 7], [35, Th. 1], [36, Th. 2], [37, Th. 1]) 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., $\log M_1^k = \ldots = \log M_t^k$ in (15)). Furthermore, $R_{t\,\mathrm{CEO}}(d)$ approaches its steady-state value (54) as $O\left(\frac{1}{t}\right)$ (this is a consequence of [24, eq. (83)-(85), (92)] and (82), (86), (93) below).

In Section III-B, we present the Gaussian 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 showing that the inner bound in Theorem 1 is tight in the Gaussian case and by evaluating the corresponding minimal directed mutual information. In Section III-C, we give auxiliary estimation lemmas that are useful in the proof of Theorem 4. We give the proof of Theorem 4 in Section III-D.

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

$$\sigma_{\mathsf{X}}^2 \triangleq \limsup_{i\to\infty}\mathbb{E}\left[\mathsf{X}_i^2\right]. \quad (55)$$

The minimum mean squared error (MMSE) in the estimation of X_i from $\mathsf{Y}_{[i]}^{[K]}$ is denoted by

$$\sigma^2_{\mathsf{X}_i|\mathsf{Y}_{[i]}^{[K]}} \triangleq \mathbb{E}\left[\left(\mathsf{X}-\mathbb{E}\left[\mathsf{X}_i|\mathsf{Y}_{[i]}^{[K]}\right]\right)^2\right], \quad (56)$$

and the steady-state causal MMSE by

$$\sigma^2_{\mathsf{X}\|\mathsf{Y}^{[K]}} \triangleq \limsup_{i\to\infty}\sigma^2_{\mathsf{X}_i|\mathsf{Y}_{[i]}^{[K]}}. \quad (57)$$

B. Gaussian rate-distortion function

In Theorem 4, the Gaussian 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 the estimation of $\{X_i\}_{i=1}^t$ provided that the codewords from k-th transmitter are decoded correctly.

Theorem 4 (Gaussian rate-distortion function with inter-block memory)**.** *For all $\sigma^2_{\mathsf{X}\|\mathsf{Y}^{[K]}} < d < \sigma^2_{\mathsf{X}}$, the causal CEO rate-distortion function (54) for the Gauss-Markov source in (2) observed through the Gaussian channels in (3) is given by*

$$R_{\mathrm{CEO}}(d) = \frac{1}{2}\log\frac{\bar{d}}{d} + \min_{\{d_k\}_{k=1}^K} \sum_{k=1}^K \frac{1}{2}\log\frac{\bar{d}_k - \sigma^2_{\mathsf{X}\|\mathsf{Y}^k}}{d_k - \sigma^2_{\mathsf{X}\|\mathsf{Y}^k}}\frac{d_k}{\bar{d}_k}, \tag{58}$$

where

$$\bar{d} \triangleq a^2 d + \sigma_V^2, \tag{59}$$

$$\bar{d}_k \triangleq a^2 d_k + \sigma_V^2, \tag{60}$$

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

$$\frac{1}{d} \le \frac{1}{\sigma^2_{\mathsf{X}\|\mathsf{Y}^{[K]}}} - \sum_{k=1}^K \left(\frac{1}{\sigma^2_{\mathsf{X}\|\mathsf{Y}^k}} - \frac{1}{d_k}\right), \tag{61}$$

$$\sigma^2_{\mathsf{X}\|\mathsf{Y}^k} \le d_k \le \sigma^2_{\mathsf{X}}. \tag{62}$$

Proof. Section III-D. □

If the source is observed directly by one or more of the encoders, say if $\sigma^2_{\mathsf{X}\|\mathsf{Y}^1} = 0$, then $d_1 = d$, $d_2 = \ldots = d_K = \sigma^2_{\mathsf{X}}$ is optimal, and (58) reduces to the causal rate-distortion function [19, eq. (1.43)] (and e.g. [20], [40, Th. 3], [22, (64)]), [24, Th. 6]):

$$R(d) = \frac{1}{2}\log\frac{\bar{d}}{d}. \tag{63}$$

The sum over $k \in [K]$ in (58) is thus the penalty due to the encoders not observing the source directly and not communicating with each other.

If the observation channels satisfy

$$\sigma^2_{\mathsf{X}\|\mathsf{Y}^1} = \ldots = \sigma^2_{\mathsf{X}\|\mathsf{Y}^K}, \tag{64}$$

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

Corollary 1 (Observation channels with the same SNR)**.** *If, in the scenario of Theorem 4, the observation channels satisfy (64), the causal CEO rate-distortion function (54) is given by*

$$R_{\mathrm{CEO}}^{K-sym}(d) = \frac{1}{2}\log\frac{\bar{d}}{d} + \frac{K}{2}\log\frac{\bar{d}_1 - \sigma^2_{\mathsf{X}\|\mathsf{Y}^1}}{d_1 - \sigma^2_{\mathsf{X}\|\mathsf{Y}^1}}\frac{d_1}{\bar{d}_1}, \tag{65}$$

where d_1 satisfies

$$\frac{1}{d} = \frac{1}{\sigma^2_{\mathsf{X}\|\mathsf{Y}^{[K]}}} - \frac{K}{\sigma^2_{\mathsf{X}\|\mathsf{Y}^1}} + \frac{K}{d_1}. \tag{66}$$

Proof. It suffices to show that the minimum in (58) is attained by $d_1 = \ldots = d_K$. Since each of the terms in the sum in (58) is a convex function of d_k, applying Jensen's inequality concludes the proof. □

Let us think now of adding identical observers by letting $K \to \infty$ in (64). Since $\sigma^2_{\mathsf{X}\|\mathsf{Y}^{[K]}} \to 0$, had the observers communicated with each other, they could have recovered the source exactly, and they could have operated at the sum rate (63) in the limit. As the following result demonstrates, $\lim_{K\to\infty} R_{\mathrm{CEO}}^{K-\mathrm{sym}}(d)$ is actually strictly greater than (63), 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 (Many channels asymptotics)**.** *In the scenario of Corollary 1,*

$$\lim_{K\to\infty} R_{\mathrm{CEO}}^{K-\mathrm{sym}}(d) = \frac{1}{2}\log\frac{\bar{d}}{d} + \frac{1}{2}\frac{\frac{1}{d} - \frac{1}{\bar{d}}}{\frac{1}{\sigma^2_{\mathsf{X}\|\mathsf{Y}^1}} - \frac{1}{\sigma^2_{\mathsf{X}}}}. \tag{67}$$

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

$$\frac{1}{\sigma^2_{\mathsf{X}\|\mathsf{Y}^{[K]}}} = \frac{K}{\sigma^2_{\mathsf{X}\|\mathsf{Y}^1}} - \frac{K-1}{\sigma^2_{\mathsf{X}}}. \tag{68}$$

Eliminating d_1 and $\sigma^2_{\mathsf{X}\|\mathsf{Y}^{[K]}}$ from (65) using (66) and (68), one readily verifies that

$$R_{\mathrm{CEO}}^{K-\mathrm{sym}}(d) - \frac{1}{2}\log\frac{\bar{d}}{d} = \frac{1}{2}\frac{\frac{1}{d} - \frac{1}{\bar{d}}}{\frac{1}{\sigma^2_{\mathsf{X}\|\mathsf{Y}^1}} - \frac{1}{\sigma^2_{\mathsf{X}}}} + O\left(\frac{1}{K}\right), \tag{69}$$

and (67) follows. □

Corollary 2 extends the result of Oohama [4, Cor. 1] to the compression with inter-block memory, and coincides with it if $a = 0$.

Considering the scenario where the encoders and the decoder do not memorize past observations or codewords, we may invoke the results on the classical Gaussian CEO problem in [5], [7] to express the minimum achievable sum rate as

$$R_{\mathrm{CEO}}^{\text{no memory}}(d) = \frac{1}{2}\log\frac{\sigma^2_{\mathsf{X}}}{d} + \min_{\{d_k\}_{k=1}^K} \sum_{k=1}^K \frac{1}{2}\log\frac{\sigma^2_{\mathsf{X}} - \sigma^2_{\mathsf{X}|\mathsf{Y}^k}}{d_k - \sigma^2_{\mathsf{X}|\mathsf{Y}^k}}\frac{d_k}{\sigma^2_{\mathsf{X}}}, \tag{70}$$

where the minimum is over

$$\frac{1}{d} \le \frac{1}{\sigma^2_{\mathsf{X}|\mathsf{Y}^{[K]}}} - \sum_{k=1}^K \left(\frac{1}{\sigma^2_{\mathsf{X}|\mathsf{Y}^k}} - \frac{1}{d_k}\right), \tag{71}$$

$$\sigma^2_{\mathsf{X}|\mathsf{Y}^k} \le d_k \le \sigma^2_{\mathsf{X}}. \tag{72}$$

Here $\sigma^2_{\mathsf{X}|\mathsf{Y}^k} \triangleq \lim_{i\to\infty} \sigma_{\mathsf{X}_i|\mathsf{Y}_i^k}$ and $\sigma^2_{\mathsf{X}|\mathsf{Y}^{[K]}} \triangleq \lim_{i\to\infty} \sigma^2_{\mathsf{X}_i|\mathsf{Y}_i^{[K]}}$ denote the stationary MMSE achievable in the estimation of X_i from Y_i^k and $\mathsf{Y}_i^{[K]}$ 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: $\bar{d} = \bar{d}_k = \sigma^2_{\mathsf{X}} = \sigma^2_{\mathsf{V}}$; similarly, $\sigma^2_{\mathsf{X}|\mathsf{Y}^k} = \sigma^2_{\mathsf{X}\|\mathsf{Y}^k}$ and $\sigma^2_{\mathsf{X}|\mathsf{Y}^{[K]}} = \sigma^2_{\mathsf{X}\|\mathsf{Y}^{[K]}}$, and the result of

Theorem 4 coincides with the classical Gaussian CEO rate-distortion function (70). 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, $\sigma^2_{\mathsf{X}} = \infty$, while $\bar{d} < \infty$. If $|a| < 1$, that benefit is finite and is characterized by the discrepancy between the stationary variance $\sigma^2_{\mathsf{X}} = \frac{\sigma^2_{\mathsf{V}}}{1-a^2}$ of the process $\{\mathsf{X}_i\}_{i=1}^{\infty}$ and the steady-state predictive MMSE $\bar{d} < \sigma^2_{\mathsf{X}}$, as well as that between $\sigma^2_{\mathsf{X}|\mathsf{Y}^k}$ and $\sigma^2_{\mathsf{X}\|\mathsf{Y}^k}$.

C. MMSE estimation lemmas

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

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

$$Y = X + W. \tag{73}$$

Then,

$$\sigma^2_{X|Y} = \sigma^2_X \left(1 - \frac{\sigma^2_X}{\sigma^2_Y}\right). \tag{74}$$

Proof. Appendix D. □

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

$$X = \bar{X}_k + W'_k. \tag{75}$$

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

$$\mathbb{E}\left[X|\bar{X}_{[K]}\right] = \sum_{k=1}^K \frac{\sigma^2_{W'}}{\sigma^2_{W'_k}} \bar{X}_k, \tag{76}$$

$$\frac{1}{\sigma^2_{W'}} = \sum_{k=1}^K \frac{1}{\sigma^2_{W'_k}} - \frac{K-1}{\sigma^2_X}. \tag{77}$$

Proof. Appendix D. □

Lemma 3 converts the "forward channels" from X to observations Y_k

$$Y_k = X + W_k, \ k = 1, \ldots, K, \tag{78}$$

where $W_k \sim \mathcal{N}\left(0, \sigma^2_{W_k}\mathbf{I}\right)$ ($\mathbf{I}$ denotes the identity matrix), $W_k \perp W_j$, $j \neq k$, into "backward channels" from estimates $\bar{X}_k$ to X (75). While both representations are equivalent, (75) is more convenient to work with. Backward channel representations find a widespread use in rate-distortion theory [41].

D. Proof of Theorem 4: converse

1) Proof overview: We evaluate the n-letter converse bound (33). We break up the minimal directed mutual information problem in (33) into subproblems, and we use the tools we developed in [24] to evaluate the causal rate-distortion functions for each subproblem. 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. [14], developed for the case $t = 1$, to $t > 1$. Converting the "forward channels" from $X_{[t]}$ to observations $Y^k_{[t]}$ into the "backward channels" from MMSE estimates $\bar{X}^k_{[t]}$ to $X_{[t]}$ and applying the lemmas in Section III-C above are key to that extension.

2) Decoupling the problem into K subproblems: Recall the notation in (6). We expand the right-hand side of (33):

$$\begin{aligned}
&\inf I\left(Y^{[K]}_{[t]} \to B^{[K]}_{[t]}\right) \\
&\geq \inf I\left(\bar{X}^{[K]}_{[t]} \to B^{[K]}_{[t]}\right) && (79)\\
&= \inf I\left(\left(X_{[t]}, \bar{X}^{[K]}_{[t]}\right) \to B^{[K]}_{[t]}\right) && (80)\\
&= \inf\left\{I\left(X_{[t]} \to B^{[K]}_{[t]}\right) + I\left(\bar{X}^{[K]}_{[t]} \to B^{[K]}_{[t]} \| X_{[t]}\right)\right\} && (81)\\
&= \inf\left\{I\left(X_{[t]} \to B^{[K]}_{[t]}\right) + \sum_{k=1}^K I\left(\bar{X}^k_{[t]} \to B^k_{[t]} \| X_{[t]}\right)\right\} && (82)
\end{aligned}$$

where

- (79) holds by the chain rule (28) using $I\left(\bar{X}^{[K]}_{[t]} \to B^{[K]}_{[t]} \| Y^{[K]}_{[t]}\right) = 0$. The infimum is over kernels $P_{B^{[K]}_{[t]}\|\bar{X}^{[K]}_{[t]}}$ satisfying both the separate encoding constraint

$$P_{B^{[K]}_{[t]}\|\bar{X}^{[K]}_{[t]}} = \prod_{k=1}^K P_{B^k_{[t]}\|\bar{X}^k_{[t]}} \tag{83}$$

  and the distortion constraint

$$\frac{1}{nt}\sum_{i=1}^t \mathbb{E}\left[\|X_i - \hat{X}_i\|^2\right] \leq d, \tag{84}$$

  where $\hat{X}_i$ (53) is the MMSE estimate of X_i given $B^{[K]}_{[i]}$;
- (80) is due to the chain rule of directed information (28), and $I\left(X_{[t]} \to B^{[K]}_{[t]} \| \bar{X}^{[K]}_{[t]}\right) = 0$;
- (81) is by the chain rule of directed information (28);
- (82) is due to (83).

3) Using causal rate-distortion functions to evaluate the terms in (82): We lower-bound the first term in (82) using a classical result on the point-to-point causal Gaussian rate-

distortion function [19, eq. (1.43)][1] as

$$\lim_{t\to\infty} \inf_{\substack{(83):\\ (84)\text{ holds}}} I\left(X_{[t]} \to B_{[t]}^{[K]}\right)$$
$$\geq \lim_{t\to\infty} \inf_{\substack{P_{\hat{X}_{[t]}^{[K]}\|X_{[t]}}:\\ (84)\text{ holds}}} I\left(X_{[t]} \to \hat{X}_{[t]}^{[K]}\right) \tag{85}$$
$$= \frac{n}{2}\log\frac{\bar{d}}{d}, \tag{86}$$

where $\bar{d}$ is uniquely determined by d via (59). Furthermore, (86) is achieved by the Gaussian kernel $P_{\hat{X}_{[t]}^{\star}\|X_{[t]}}$ such that

$$X_i = \hat{X}_i^{\star} + Z_i', \quad Z_i' \sim \mathcal{N}(0, d\,\mathbf{I}), \tag{87}$$

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

$$d = \sigma^2_{\mathsf{X}\|\hat{\mathsf{X}}^{\star}} \tag{88}$$
$$\bar{d} = \sigma^2_{\mathsf{X}\|\mathcal{D}\hat{\mathsf{X}}^{\star}}. \tag{89}$$

For each of the remaining K terms in (82), note that $\{\bar{X}_i^k\}$ is a Gauss-Markov process

$$\bar{X}_{i+1}^k = a\bar{X}_i^k + \bar{V}_i^k, \tag{90}$$

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

$$X_i = \bar{X}_i^k + W_i^{k\prime}, \tag{91}$$

where $W_i^{k\prime}$ are independent, $W_i^{k\prime} \sim \mathcal{N}\left(0, \sigma^2_{\mathsf{X}_i|\mathsf{Y}_{[i]}^k}\mathbf{I}\right)$, and $W_i^{k\prime} \perp X_i^k$. Thus, we may apply the result [24, Th. 7] on the causal counterpart of Gaussian Wyner-Ziv rate-distortion function to the process $\{\bar{X}_i^k\}$ (90) with side information $\{X_i\}$ (91) to write (while stated for the scalar Gaussian source, the same argument applies to n parallel Gaussian sources of the same power, as is the case here; see [25] for the general vector case)

$$\lim_{t\to\infty} \inf_{\substack{P_{B_{[t]}^k\|\bar{X}_{[t]}^k}:\\ \frac{1}{t}\sum_{i=1}^t \sigma^2_{\bar{X}_i^k|X_{[i]},B_{[i]}^k} \leq \rho_k}} I\left(\bar{X}_{[t]}^k \to B_{[t]}^k \| X_{[t]}\right) \tag{92}$$
$$= \frac{n}{2}\log\frac{\bar{\rho}_k}{\rho_k}, \tag{93}$$

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

$$\frac{1}{\bar{\rho}_k} = \frac{1}{\sigma^2_{\bar{\mathsf{W}}^{k\prime}}} + \frac{1}{a^2\rho_k + \sigma^2_{\bar{\mathsf{V}}}}. \tag{94}$$

Furthermore, (93) is attained by the Gaussian kernel $P_{B^{k\star}\|\bar{X}^k}$

$$B_i^{k\star} = \bar{X}_i^k + Z_i^k, \quad Z_i \sim \mathcal{N}(0, \sigma^2_{Z^k}\mathbf{I}), \tag{95}$$

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

$$\rho_k = \sigma^2_{\bar{\mathsf{X}}^k\|\mathsf{X},\mathsf{B}^{k\star}}, \tag{96}$$
$$\bar{\rho}_k = \sigma^2_{\bar{\mathsf{X}}^k\|\mathsf{X},\mathcal{D}\mathsf{B}^{k\star}}. \tag{97}$$

The variances $\sigma^2_{Z^k}$ in (95) are set to satisfy (96).

[1] See also [24, Th. 6]; while stated for the scalar Gaussian source, the same argument applies to n parallel Gaussian sources of the same power, as is the case here; see [23] for the general vector case.

4) Linking $\{\rho_k\}_{k=1}^K$ to d: It remains to establish the connection between $\{\rho_k\}_{k=1}^K$ (96) and d (88).

Setting $\hat{X}_i^{\star}$ in (87) to

$$\hat{X}_i^{\star} \triangleq \mathbb{E}\left[X_i | B_{[i]}^{[K]\star}\right] \tag{98}$$

attains equality in (85), implying that the same Gaussian kernel (95) simultaneously attains the infima of both terms in (82). Thus, putting together (82), (86) and (93), we have

$$R_{\mathrm{CEO}}(d) \geq \tag{99}$$
$$\inf_{\substack{\left\{\sigma^2_{\bar{\mathsf{X}}^k\|\mathsf{X},\mathsf{U}^{k\star}}\right\}_{k=1}^K:\\ \sigma^2_{\mathsf{X}\|\mathsf{U}^{[K]\star}}=d}} \left\{\frac{1}{2}\log\frac{\sigma^2_{\mathsf{X}\|\mathcal{D}\mathsf{B}^{[K]\star}}}{\sigma^2_{\mathsf{X}\|\mathsf{B}^{[K]\star}}} + \sum_{k=1}^K \frac{1}{2}\log\frac{\sigma^2_{\bar{\mathsf{X}}^k\|\mathsf{X},\mathcal{D}\mathsf{B}^{\star k}}}{\sigma^2_{\bar{\mathsf{X}}^k\|\mathsf{X},\mathsf{B}^{k\star}}}\right\}$$

Invoking Lemma 3 with $X \leftarrow \mathsf{X}_i$, $\bar{X}_k \leftarrow \bar{\mathsf{X}}_i^k$, $W_k' \leftarrow \mathsf{W}_i^{k\prime}$, we express

$$\bar{\mathsf{X}}_i \triangleq \mathbb{E}\left[\mathsf{X}_i | \mathsf{Y}_{[i]}^{[K]}\right] \tag{100}$$
$$= \sum_{k=1}^K \frac{\sigma^2_{\mathsf{X}_i|\mathsf{Y}_{[i]}^{[K]}}}{\sigma^2_{\mathsf{X}_i|\mathsf{Y}_{[i]}^k}} \bar{\mathsf{X}}_i^k, \tag{101}$$

which implies in particular

$$\mathbb{E}\left[\bar{\mathsf{X}}_i | \mathsf{X}_{[i]}, \mathsf{B}_{[i]}^{[K]\star}\right] = \sum_{k=1}^K \frac{\sigma^2_{\mathsf{X}_i|\mathsf{Y}_{[i]}^{[K]}}}{\sigma^2_{\mathsf{X}_i|\mathsf{Y}_{[i]}^k}} \mathbb{E}\left[\bar{\mathsf{X}}_i^k | \mathsf{X}_{[i]}, \mathsf{B}_{[i]}^{[K]\star}\right] \tag{102}$$
$$= \sum_{k=1}^K \frac{\sigma^2_{\mathsf{X}_i|\mathsf{Y}_{[i]}^{[K]}}}{\sigma^2_{\mathsf{X}_i|\mathsf{Y}_{[i]}^k}} \mathbb{E}\left[\bar{\mathsf{X}}_i^k | \mathsf{X}_{[i]}, \mathsf{B}_{[i]}^{k\star}\right]. \tag{103}$$

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

$$\sigma^2_{\bar{\mathsf{X}}\|\mathsf{X},\mathsf{B}^{[K]\star}} = \sum_{k=1}^K \frac{\sigma^4_{\mathsf{X}\|\mathsf{Y}}}{\sigma^4_{\mathsf{X}\|\mathsf{Y}^k}}\rho_k. \tag{104}$$

Observe that

$$\sigma^2_{\bar{\mathsf{X}}_i|\mathsf{X}_{[i]},\mathsf{B}_{[i]}^{[K]\star}} = \sigma^2_{\bar{\mathsf{X}}_i - \mathbb{E}\left[\bar{\mathsf{X}}_i|\mathsf{X}_{[i]},\mathsf{B}_{[i]}^{[K]\star}\right]} \tag{105}$$
$$= \sigma^2_{\bar{\mathsf{X}}_i - \mathsf{X}_i - \mathbb{E}\left[\bar{\mathsf{X}}_i - \mathsf{X}_i|\mathsf{X}_{[i]},\mathsf{B}_{[i]}^{[K]\star}\right]} \tag{106}$$
$$= \sigma^2_{\bar{\mathsf{X}}_i - \mathsf{X}_i - \mathbb{E}\left[\bar{\mathsf{X}}_i - \mathsf{X}_i|\mathsf{X}_i - \hat{\mathsf{X}}_i^{\star}\right]} \tag{107}$$
$$= \sigma^2_{\mathsf{X}_i - \bar{\mathsf{X}}_i|\mathsf{X}_i - \hat{\mathsf{X}}_i^{\star}}, \tag{108}$$

Now, we apply Lemma 2 with $X \leftarrow \mathsf{X}_i - \bar{\mathsf{X}}_i$, $Y \leftarrow \mathsf{X}_i - \hat{\mathsf{X}}_i$, $W \leftarrow \bar{\mathsf{X}}_i - \hat{\mathsf{X}}_i$ to establish

$$\lim_{i\to\infty} \sigma^2_{\mathsf{X}_i - \bar{\mathsf{X}}_i|\mathsf{X}_i - \hat{\mathsf{X}}_i} = \sigma^2_{\mathsf{X}\|\mathsf{Y}}\left(1 - \frac{\sigma^2_{\mathsf{X}\|\mathsf{Y}}}{d}\right), \tag{109}$$

which, together with (104) and (108), means

$$\frac{1}{d} \leq \frac{1}{\sigma^2_{\mathsf{X}\|\mathsf{Y}}} - \sum_{k=1}^K \frac{\rho_k}{\sigma^4_{\mathsf{X}\|\mathsf{Y}_k}}. \tag{110}$$

Also, note that

$$0 \leq \rho_k \leq \sigma^2_{\bar{\mathsf{X}}^k\|\mathsf{X}}. \tag{111}$$

We can now simplify the constraint set in the infimum in (99): the infimum is over $\{\rho_k\}_{k=1}^K$ that satisfy (110) and (111).

It remains to clarify how the form in (58), (61), (62), parameterized in terms of

$$d_k \triangleq \sigma^2_{\mathsf{X}\|\mathsf{B}^{k\star}} \tag{112}$$

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

$$\rho_k = \sigma^2_{\mathsf{X}\|\mathsf{Y}^k}\left(1 - \frac{\sigma^2_{\mathsf{X}\|\mathsf{Y}^k}}{d_k}\right). \tag{113}$$

Plugging (113) into (110) leads to (61). Applying Lemma 2 with $X \leftarrow \mathsf{X}_i - \bar{\mathsf{X}}_i^k$, $Y \leftarrow \mathsf{X}_i$, $W \leftarrow \bar{\mathsf{X}}_i^k$, we express

$$\sigma^2_{\bar{\mathsf{X}}^k\|\mathsf{X}} = \sigma^2_{\mathsf{X}\|\mathsf{Y}^k}\left(1 - \frac{\sigma^2_{\mathsf{X}\|\mathsf{Y}^k}}{\sigma^2_{\mathsf{X}}}\right), \tag{114}$$

which, together with (113), implies the equivalence of (111) and (62). Finally, applying Lemma 2 with $X \leftarrow \mathsf{X}_i - \bar{\mathsf{X}}_i^k$, $Y \leftarrow \mathsf{X}_i - a\hat{\mathsf{X}}_{i-1}^k$, $W \leftarrow \bar{\mathsf{X}}_i^k - a\hat{\mathsf{X}}_{i-1}^k$, we express

$$\bar{\rho}_k = \sigma^2_{\mathsf{X}\|\mathsf{Y}^k}\left(1 - \frac{\sigma^2_{\mathsf{X}\|\mathsf{Y}^k}}{\bar{d}_k}\right). \tag{115}$$

Plugging (113) and (115) into (99), we conclude the equivalence of (99) and (58). □

E. Proof of Theorem 4: achievability

We evaluate the Berger-Tung inner bound with inter-block memory (20). In the proof of the converse, we lower-bounded the n-letter version of that bound, i.e., (33), by computing the right-hand side of (82). Thus, it suffices to show that equality holds in (79). But this is easily verified by substituting the optimal kernel (95) into the left side of (79). □

IV. Loss due to isolated observers

A. Overview

In Section IV, we investigate how the rate-distortion function in Theorem 4 compares to what would have been 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 $R_{\mathrm{CEO}}(d)$ 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 $\left\{Y_i^{[K]}\right\}$. It is also equivalent to allowing joint encoding policies $P_{B_{[t]}^{[K]}\|Y_{[t]}^{[K]}}$ in lieu of independent encoding policies $\prod_{k=1}^K P_{B_{[t]}^k\|Y_{[t]}^k}$ 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 in the literature [41]–[44]. The setting with causal coding was considered in [22, Th. 5–8, Cor. 1].

We denote the joint encoding counterpart of the operational fundamental limit $R_{\mathrm{CEO}}(d)$ (54) by $R_{\mathrm{rm}}(d)$ (remote).

The following result is a corollary to Theorem 4.

Corollary 3 (Remote rate-distortion function with inter-block memory)**.** *For all $\sigma^2_{\mathsf{X}\|\mathsf{Y}^{[K]}} < d < \sigma^2_{\mathsf{X}}$, the rate-distortion function with joint encoding for the Gauss-Markov source in (2) observed through the Gaussian channels in (3) is given by*

$$R_{\mathrm{rm}}(d) = \frac{1}{2}\log\frac{\bar{d} - \sigma^2_{\mathsf{X}\|\mathsf{Y}^{[K]}}}{d - \sigma^2_{\mathsf{X}\|\mathsf{Y}^{[K]}}}, \tag{116}$$

where $\bar{d}$ is defined in (59).

Proof. Examining its proof, it is easy to see that Theorem 4 continues to hold in the scenario with vector observations Y_i^k (that are still required to be jointly Gaussian with 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 (58) indeed reduces to (116) in that case.

Previously, the minimal mutual information problem leading to $R_{\mathrm{rm}}(d)$ was solved in [22] in a different form using a different method; both forms are equivalent (Appendix E). □

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 $\sigma^2_{\mathsf{X}\|\mathsf{Y}^{[K]}} < d$ and*

$$\frac{1}{d} \geq \frac{1}{\sigma^2_{\mathsf{X}\|\mathsf{Y}^{[K]}}} + \frac{K}{\sigma^2_{\mathsf{X}}} - \min_{k\in[K]}\frac{K}{\sigma^2_{\mathsf{X}\|\mathsf{Y}^k}}. \tag{117}$$

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

$$R_{\mathrm{CEO}}(d) - R_{\mathrm{rm}}(d) \leq (K-1)\left(R_{\mathrm{rm}}(d) - R(d)\right), \tag{118}$$

with equality if and only if $\sigma^2_{\mathsf{X}\|\mathsf{Y}^k}$ are all the same, where $R(d)$ is given in (63) and $R_{\mathrm{rm}}(d)$ is given in (116).

Proof. Section IV-D. □

Theorem 5 parallels the corresponding result for the classical Gaussian CEO problem [31, Cor. 1], and recovers it if $a = 0$. It is 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. In the case of identical observation channels, condition (117) reduces to $d \leq \sigma^2_{\mathsf{X}}$. The rate loss (118) grows without bound in the high resolution regime $d \downarrow \sigma^2_{\mathsf{X}\|\mathsf{Y}^{[K]}}$ and vanishes in the low resolution regime $d \uparrow \sigma^2_{\mathsf{X}}$.

C. A suboptimal waterfilling allocation

We present an upper bound to $R_{\mathrm{CEO}}(d)$, which is obtained by waterfilling over d_k's. This parallels the corresponding result for the classical Gaussian CEO problem [31, Cor. 1]. Like [31], we use waterfilling to obtain this result, but unlike the case $t = 1$ considered in [31] where

waterfilling is optimal [7], 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 $t > 1$ [22, Remark 2].

Proposition 1 (Suboptimal waterfilling rate allocation)**.** *For all $\sigma^2_{\mathsf{X}\|\mathsf{Y}^{[K]}} < d < \sigma^2_{\mathsf{X}}$, the causal CEO rate-distortion function for the Gauss-Markov source in (2) observed through the Gaussian channels in (3) is upper-bounded as*

$$R_{\mathrm{CEO}}(d) \le \frac{1}{2}\log\frac{\bar{d}}{d} + \sum_{k=1}^{K}\frac{1}{2}\log\frac{\bar{d}_k - \sigma^2_{\mathsf{X}\|\mathsf{Y}^k}}{d_k - \sigma^2_{\mathsf{X}\|\mathsf{Y}^k}}\frac{d_k}{\bar{d}_k}, \quad (119)$$

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

$$\frac{1}{\sigma^2_{\mathsf{X}\|\mathsf{Y}^k}} - \frac{1}{d_k} = \min\left\{\frac{1}{\lambda}, \frac{1}{\sigma^2_{\mathsf{X}\|\mathsf{Y}^k}} - \frac{1}{\sigma^2_{\mathsf{X}}}\right\}, \quad (120)$$

λ is the solution to

$$\sum_{k=1}^{K}\min\left\{\frac{1}{\lambda}, \frac{1}{\sigma^2_{\mathsf{X}\|\mathsf{Y}^k}} - \frac{1}{\sigma^2_{\mathsf{X}}}\right\} = \frac{1}{\sigma^2_{\mathsf{X}\|\mathsf{Y}^{[K]}}} - \frac{1}{d}, \quad (121)$$

and $\bar{d}$, $\bar{d}_k$ are defined in (59), (60) respectively. Inequality in (119) holds with equality if all $\sigma^2_{\mathsf{X}\|\mathsf{Y}^k}$ are equal.

Proof. We first check that the choice in (120) is feasible. Since the right side of (120) is lower-bounded by 0 and upper bounded by $\frac{1}{\sigma^2_{\mathsf{X}\|\mathsf{Y}^k}} - \frac{1}{\sigma^2_{\mathsf{X}}}$, (62) is satisfied. Furthermore, substituting (121) ensures that (61) is satisfied with equality.

To claim equality in the symmetrical case, it suffices to recall that in that case, the minimum in (58) is attained by $d_1 = \ldots = d_K$ (Corollary 1). □

D. Proof of Theorem 5

Under the assumption (117), the waterfilling allocation in Proposition 1 results in all active transmitters, and (120) reduces to

$$\frac{1}{\sigma^2_{\mathsf{X}\|\mathsf{Y}^k}} - \frac{1}{d_k} = \frac{1}{\lambda}, \quad (122)$$

while (121) reduces to

$$\lambda = K\left(\frac{1}{\sigma^2_{\mathsf{X}\|\mathsf{Y}^{[K]}}} - \frac{1}{d}\right)^{-1}. \quad (123)$$

Substituting (122) into (119) we conclude that under assumption (117),

$$\begin{aligned}
&R_{\mathrm{CEO}}(d)\\
&= \frac{1}{2}\log\frac{\bar{d}}{d} + \frac{1}{2}\sum_{k=1}^{K}\log\left[\left(\frac{1}{\sigma^2_{\mathsf{X}\|\mathsf{Y}^k}} - \frac{1}{\bar{d}_k}\right)\lambda\right] && (124)\\
&\le \frac{1}{2}\log\frac{\bar{d}}{d} + \frac{K}{2}\log\left[\sum_{k=1}^{K}\left(\frac{1}{\sigma^2_{\mathsf{X}\|\mathsf{Y}^k}} - \frac{1}{\bar{d}_k}\right)\frac{\lambda}{K}\right] && (125)\\
&= \frac{1}{2}\log\frac{\bar{d}}{d} + \frac{K}{2}\log\left(\frac{1}{\sigma^2_{\mathsf{X}\|\mathsf{Y}^{[K]}}} - \frac{1}{\bar{d}}\right)\frac{\lambda}{K} && (126)\\
&= \frac{1}{2}\log\frac{\bar{d} - \sigma^2_{\mathsf{X}\|\mathsf{Y}^{[K]}}}{d - \sigma^2_{\mathsf{X}\|\mathsf{Y}^{[K]}}} + \frac{K-1}{2}\log\frac{\bar{d} - \sigma^2_{\mathsf{X}\|\mathsf{Y}^{[K]}}}{d - \sigma^2_{\mathsf{X}\|\mathsf{Y}^{[K]}}}\frac{d}{\bar{d}} && (127)
\end{aligned}$$

where

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

$$\frac{1}{\sigma^2_{\mathsf{X}\|\mathsf{Y}^{[K]}}} = \sum_{k=1}^{K}\frac{1}{\sigma^2_{\mathsf{X}\|\mathsf{Y}^k}} - \frac{K-1}{\sigma^2_{\mathsf{X}}}, \quad (128)$$

$$\frac{1}{\bar{d}} = \sum_{k=1}^{K}\frac{1}{\bar{d}_k} - \frac{K-1}{\sigma^2_{\mathsf{X}}}, \quad (129)$$

  which holds by Lemma 3 even if the source is nonstationary (that is, $|a| \ge 1$ and $\sigma^2_{\mathsf{X}} = \infty$), as a simple limiting argument taking $\frac{K-1}{\sigma^2_{\mathsf{X}}}$ to 0 confirms.
- (127) holds by substituting (122) into (126).

Notice that (118) is just another way to write (127), using (116) and (63). To verify the condition for equality, note that '=' holds in (124) in the symmetrical case by Proposition 1, and that '=' holds in (125) only in the symmetrical case due to strict concavity of the log function. □

V. Conclusion

In this paper, we set up the causal CEO problem (Definition 1, Definition 2) and we prove that the rate-distortion function is upper bounded by the directed mutual information from the encoders to the decoder minimized subject to the distortion constraint and the separate encoding constraint, and lower bounded by the minimal directed mutual information subject to a weaker constraint (Theorem 1). The proof of the direct coding theorem hinges upon an SLC-based nonasymptotic bound (Theorem 2) that extends [28, 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 [12], [13] to $t > 1$ time steps (Theorem 3).

By showing that the achievability bound in Theorem 1 is tight in the Gaussian case and by solving the correspoding minimal directed mutual information problem, we characterize the causal Gaussian CEO 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 we study its asymptotic behavior as $K \to \infty$ (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).

We chose not to treat correlation between n components of X_i and W_i^k in this paper merely to keep things simple. We expect our results to generalize to the scenario in which the components of the source and the noise are not i.i.d. A further interesting generalization would be to consider the general vector state-space model

$$X_{i+1} = AX_i + V_i \quad (130)$$

$$Y_i^k = CX_i + W_i^k, \quad (131)$$

where A is an $n \times n$ matrix and C is an $m \times n$ matrix. It will also 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 [11], [15], [16] appear promising in that pursuit. Certain causal multiterminal source coding problems also appear within reach in view of the result in [10] and the applicability of Theorem 3 to multiterminal source coding.

VI. Acknowledgement

We thank both anonymous reviewers for their insightful and careful reviews, which are reflected in the final version.

References

[1] V. Kostina and B. Hassibi, "Fundamental limits of distributed tracking," in *Proceedings 2020 IEEE International Symposium on Information Theory*, June 2020, pp. 2438–2443.

[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] Y. Oohama, "Rate-distortion theory for Gaussian multiterminal source coding systems with several side informations at the decoder," *IEEE Transactions on Information Theory*, vol. 51, no. 7, pp. 2577–2593, 2005.

[7] 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.

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] T. Berger, *Multi-terminal source coding. Chapter in The Information Theory Approach to Communications*. Springer-Verlag, 1978.

[13] S.-Y. Tung, "Multiterminal source coding," Ph.D. dissertation, School of Electrical Engineering, Cornell University, 1978.

[14] J. Wang, J. Chen, and X. Wu, "On the sum rate of Gaussian multiterminal source coding: New proofs and results," *IEEE Transactions on Information Theory*, vol. 56, no. 8, pp. 3946–3960, 2010.

[15] 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.

[16] J. Wang and J. Chen, "Vector Gaussian multiterminal source coding," *IEEE Transactions on Information Theory*, vol. 60, no. 9, pp. 5533–5552, 2014.

[17] 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.

[18] A. Gorbunov and M. S. Pinsker, "Nonanticipatory and prognostic ϵ-entropies and message generation rates," *Problemy Peredachi Informatsii*, vol. 9, no. 3, pp. 12–21, 1973.

[19] ——, "Prognostic epsilon entropy of a Gaussian message and a Gaussian source," *Problemy Peredachi Informatsii*, vol. 10, no. 2, pp. 5–25, 1974.

[20] 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, Sep. 2004.

[21] 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*, vol. 61, no. 8, pp. 2171–2186, Nov. 2016.

[22] V. Kostina and B. Hassibi, "Rate-cost tradeoffs in control," *IEEE Transactions on Automatic Control*, vol. 64, no. 11, pp. 4525–4540, Apr. 2019.

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

[24] 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.

[25] O. Sabag, P. Tian, V. Kostina, and B. Hassibi, "The minimal directed information needed to improve the LQG cost," in *2020 59th IEEE Conference on Decision and Control (CDC)*, 2020, pp. 1842–1847.

[26] 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.

[27] 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.

[28] ——, "A technique for deriving one-shot achievability results in network information theory," 2013.

[29] 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.

[30] A. El Gamal and Y.-H. Kim, *Network information theory*. Cambridge university press, 2011.

[31] V. Kostina, "Rate loss in the Gaussian CEO problem," in *Proceedings 2019 IEEE Information Theory Workshop*, Visby, Gotland, Sweden, Aug. 2019.

[32] J. Østergaard and R. Zamir, "Incremental refinement using a gaussian test channel," in *2011 IEEE International Symposium on Information Theory Proceedings*, 2011, pp. 2233–2237.

[33] G. Kramer, "Directed information for channels with feedback," Ph.D. dissertation, ETH Zurich, 1998.

[34] J. Massey, "Causality, feedback and directed information," in *Proc. Int. Symp. Inf. Theory Applic.(ISITA-90)*, Nov. 1990, pp. 303–305.

[35] 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.

[36] 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.

[37] O. Sabag, V. Kostina, and B. Hassibi, "Feedback capacity of MIMO gaussian channels," in *Proceedings 2021 IEEE International Symposium on Information Theory*, July 2021.

[38] ——, "Feedback capacity of MIMO Gaussian channels," *arXiv preprint arXiv:2106.01994*, June 2021.

[39] J. Liu, P. Cuff, and S. Verdú, "On α-decodability and α-likelihood decoder," in *Proceedings 55th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, Monticello, IL, Oct. 2017, pp. 118–124.

[40] 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.

[41] T. Berger, *Rate distortion theory*. Prentice-Hall, Englewood Cliffs, NJ, 1971.

[42] R. Dobrushin and B. Tsybakov, "Information transmission with additional noise," *IRE Transactions on Information Theory*, vol. 8, no. 5, pp. 293 –304, Sep. 1962.

[43] H. S. Witsenhausen, "Indirect rate distortion problems," *IEEE Transactions on Information Theory*, vol. 26, no. 5, pp. 518–521, Sep. 1980.

[44] V. Kostina and S. Verdú, "Nonasymptotic noisy lossy source coding," *IEEE Transactions on Information Theory*, vol. 62, no. 11, pp. 6111–6123, Nov. 2016.

[45] 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 $\underline{U}_1^k, \underline{U}_2^k, \ldots, \underline{U}_t^k$ to use at the transmission instances $1, 2, \ldots, t$ respectively. Codebook $\underline{U}_i^k$ is an $n \times L_1^k \times \ldots \times L_i^k$-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 ℓ_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 $\underline{U}_{\mathcal{I}}^{\mathcal{K}} \triangleq (\underline{U}_i^k \colon k \in \mathcal{K}, i \in \mathcal{I})$. For indices $\ell_i^k \in [L_i^k]$, $i \in [t]$, $k \in [K]$, we denote their collection $\ell_{\mathcal{I}}^{\mathcal{K}} \triangleq (\ell_i^k \colon k \in \mathcal{K}, i \in \mathcal{I})$. Finally, $\underline{U}_{\mathcal{I}}^{\mathcal{K}}(\ell_{\mathcal{I}}^{\mathcal{K}}) \triangleq (\underline{U}_i^k(\ell_{[i]}^k) \colon k \in \mathcal{K}, i \in \mathcal{I})$ 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 $i = 2, \ldots, t$, codebook i for user k, $\underline{U}_i^k$, consists of L_i^k codewords drawn i.i.d. from $P_{U_i^k|U_{[i-1]}^k = \underline{U}_{[i-1]}^k\left(\ell_{[i-1]}^k\right)}$, for each $\ell_{[i-1]}^k \in \prod_{j=1}^{i-1}[L_j^k]$.

Random binning: Let $\mathsf{B}_i^k \colon [L_i^k] \mapsto [M_i^k]$, $i = 1, 2, \ldots, t$, 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 $\mathsf{B}_{\mathcal{I}}^{\mathcal{K}}(\ell_{\mathcal{I}}^{\mathcal{K}}) \triangleq (\mathsf{B}_i^k(\ell_{[i]}^k) \colon k \in \mathcal{K}, i \in \mathcal{I})$ 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) [27], [28] followed by random binning. Each user k maintains a collection of encoders indexed by time $i = 1, 2, \ldots, t$; 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

$$Q_{U_i^k|Y_{[i]}^k = y_{[i]}^k, U_{[i-1]}^k = \underline{u}_{[i-1]}^k\left(\ell_{[i-1]}^k\right)}\left(\underline{u}_i^k\left(\ell_{[i]}^k\right)\right) = \frac{\exp\left(\imath\left(y_{[i]}^k; \underline{u}_i^k\left(\ell_{[i]}^k\right) | \underline{u}_{[i-1]}^k\left(\ell_{[i-1]}^k\right)\right)\right)}{\sum_{\ell=1}^{L_i^k} \exp\left(\imath\left(y_{[i]}^k; \underline{u}_i^k\left(\left(\ell_{[i-1]}^k, \ell\right)\right) | \underline{u}_{[i-1]}^k\left(\ell_{[i-1]}^k\right)\right)\right)}, \tag{132}$$

where the conditional information density is with respect to the given distribution $P_{Y_i^k U_i^k | U_{[i-1]}^k}$. Encoder i transmits $m_i^k = \mathsf{b}_i^k(\ell_i^k)$ 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

$$Q_{B_{[t]}^k \| Y_{[t]}^k}(m_{[t]}^k \| y_{[t]}^k) = \sum_{\ell_{[t]}^k} 1\left\{\mathsf{b}_{[t]}^k\left(\ell_{[t]}^k\right) = m_{[t]}^k\right\} Q_{U_{[t]}^k \| Y_{[t]}^k}(\ell_{[t]}^k \| y_{[t]}^k). \tag{133}$$

Since the encoders operate independently,

$$Q_{U_{[t]}^{[K]} \| Y_{[t]}^{[K]}} = \prod_{k=1}^{K} Q_{U_{[t]}^k \| Y_{[t]}^k}, \tag{134}$$

$$Q_{B_{[t]}^{[K]} \| Y_{[t]}^{[K]}} = \prod_{k=1}^{K} Q_{B_{[t]}^k \| Y_{[t]}^k}. \tag{135}$$

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) [29, 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

$$Q_{\hat{U}_i^{[K]} | B_{[i]}^{[K]} = b_{[i]}^{[K]}, \hat{U}_{[i-1]}^{[K]} = \underline{u}_{[i-1]}^{[K]}\left(\hat{\ell}_{[i-1]}^{[K]}\right)}\left(\underline{u}_i^{[K]}\left(\hat{\ell}_i^{[K]}\right)\right) = \frac{\mathsf{g}\left(\underline{u}_{[i]}^{[K]}\left(\hat{\ell}_{[i]}^{K}\right)\right) 1\left\{\mathsf{b}_i^{[K]}\left(\hat{\ell}_i^{[K]}\right) = m_i^{[K]}\right\}}{\sum_{\ell_i^{[K]}} \mathsf{g}\left(\underline{u}_{[i]}^{[K]}\left(\hat{\ell}_{[i-1]}^{[K]}, \ell_i^{[K]}\right)\right) 1\left\{\mathsf{b}_i^{[K]}(\ell_i^{[K]}) = m_i^{[K]}\right\}}, \tag{136}$$

where

$$\mathsf{g}\left(u_{[i]}^{[K]}\right) \triangleq \prod_{k=1}^{K} 1\left\{\jmath^{\pi(k)}\left(u_{[i]}^{\pi([K])}\right) \geq \log \frac{L_i^{\pi(k)}}{M_i^{\pi(k)}} + \beta_i^{\pi(k)}\right\}. \tag{137}$$

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

Error analysis: We consider two error events:

$$\mathcal{E}_{\text{dec}} \colon \hat{U}_{[t]}^{[K]} \neq U_{[t]}^{[K]} \tag{138}$$

$$\mathcal{E}_{\text{enc}} \colon \bigcup_{i=1}^{t} \left\{\mathsf{d}\left(X_i, \hat{X}_i\left(U_{[i]}^{[K]}\right)\right) > d_i\right\}, \tag{139}$$

where $U_{[i]}^{[K]}$ are the codewords chosen by the encoders at encoding step (132), and $\hat{U}_{[t]}^{[K]}$ is the decoder's estimate of those codewords after decoding step (136). 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 $Y_{[t]}^{[K]}$, $\underline{U}_{[t]}^{[K]}\left(1_{[t]}^{[K]}\right)$, $\mathsf{B}_{[t]}^{[K]}\left(1_{[t]}^{[K]}\right)$, $\hat{X}_{[i]}\left(\underline{U}_{[i]}^{[K]}\left(1_{[i]}^{[K]}\right)\right)$; by $\mathbb{Q}$ the probability measure generated by the code; and by F_i^k, G_i the denominators in (132) and (136), respectively. Following Shannon's random coding argument and the Jensen inequality technique of Yassaee et al. [27], [28], 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\{\mathsf{d}\left(X_i,\hat{X}_i\right)\le d_i \mid \underline{U}_{[t]}^{[K]},\mathsf{B}_{[t]}^{[K]}\right\}\right]\right] \\
&\ge \mathbb{E}\left[\mathbb{Q}\left[\mathcal{E}_{\text{enc}}^{c}\cap\mathcal{E}_{\text{dec}}^{c}\mid \underline{U}_{[t]}^{[K]},\mathsf{B}_{[t]}^{[K]}\right]\right] \qquad (140)\\
&= \mathbb{E}\Bigg[\sum_{\ell_{[t]}^{[K]}\in\prod_{i=1}^{t}\prod_{k=1}^{K}[L_i^k]} Q_{U_{[t]}\|Y_{[t]}^{[K]}}\left(\underline{U}_{[t]}^{[K]}\left(\ell_{[t]}^{[K]}\right)\|Y_{[t]}^{[K]}\right)\\
&\quad\cdot\sum_{m_{[t]}^{[K]}\in\prod_{i=1}^{t}\prod_{k=1}^{K}[M_i^k]} 1\left\{\mathsf{B}_{[t]}^{[K]}\left(\ell_{[t]}^{[K]}\right)=m_{[t]}^{[K]}\right\}\\
&\quad\cdot Q_{\hat{U}_{[t]}^{[K]}\|B_{[t]}^{[K]}=m_{[t]}^{[K]}}\left(\underline{U}_{[t]}^{[K]}\left(\ell_{[t]}^{[K]}\right)\right)1\{\mathcal{E}_{\text{enc}}^{c}\}\Bigg] \qquad (141)\\
&= \prod_{k=1}^{K}\prod_{i=1}^{t} M_i^k L_i^k\\
&\quad\cdot\mathbb{E}\Bigg[\mathbb{E}\Bigg[Q_{U_{[t]}^{[K]}\|Y_{[t]}^{[K]}}\left(\underline{U}_{[t]}^{[K]}\left(1_{[t]}^{[K]}\right)\|Y_{[t]}^{[K]}\right)\\
&\quad\cdot 1\left\{\mathsf{B}_{[t]}^{[K]}\left(1_{[t]}^{[K]}\right)=1_{[t]}^{[K]}\right\}\\
&\quad\cdot Q_{\hat{U}_{[t]}^{[K]}\|B_{[t]}^{[K]}=1_{[t]}^{[K]}}\left(\underline{U}_{[t]}^{[K]}\left(1_{[t]}^{[K]}\right)\right)1\{\mathcal{E}_{\text{enc}}^{c}\}\mid\mathcal{F}\Bigg]\Bigg] \qquad (142)\\
&\ge \prod_{k=1}^{K}\prod_{i=1}^{t} M_i^k L_i^k\\
&\quad\cdot\mathbb{E}\Bigg[\prod_{k=1}^{K}\prod_{i=1}^{t}\frac{\exp\left(\imath\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]}\\
&\quad\cdot\frac{\mathsf{g}\left(\underline{U}_{[i]}^{[K]}\left(1_{[i]}^{[K]}\right)\right)1\left\{\mathsf{B}_i^{[K]}\left(1^{[K]}\right)=1^{[K]}\right\}}{\mathbb{E}\left[G_i\mid\mathcal{F}\right]}\\
&\quad\cdot 1\left\{\mathsf{d}\left(X_i,\hat{X}_i\left(\underline{U}_{[i]}^{[K]}\left(1_{[i]}^{[K]}\right)\right)\right)\le d_i\right\}\Bigg], \qquad (143)
\end{aligned}$$

where

- the expectation $\mathbb{E}$ in (141) is with respect to the codebooks $\underline{U}_{[t]}^{[K]}$, the random binning functions $\mathsf{B}_{[t]}^{[K]}$, the decoder $P_{\hat{X}_{[t]}\|U_{[t]}^{[K]}}$ and $X_{[t]}$, $Y_{[t]}^{[K]}$;
- (142) 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 (143) is shown as follows. The product $Q_{U_{[t]}^{[K]}\|Y_{[t]}^{[K]}}Q_{\hat{U}_{[t]}^{[K]}\|B_{[t]}^{[K]}}$ 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

$$\mathbb{E}\left[\prod_{i=1}^{t}\frac{1}{G_i}\prod_{k=1}^{K}\frac{1}{F_i^k}\mid\mathcal{F}\right]\ge\prod_{i=1}^{t}\frac{1}{\mathbb{E}\left[G_i\mid\mathcal{F}\right]}\prod_{k=1}^{K}\frac{1}{\mathbb{E}\left[F_i^k\mid\mathcal{F}\right]} \qquad (144)$$

We compute each factor in (144) as follows.

$$\begin{aligned}
&\mathbb{E}\left[F_i^k|\mathcal{F}\right] \qquad (145)\\
&=\mathbb{E}\left[\sum_{\ell=1}^{L_i^k}\exp\left(\imath\left(Y_{[i]}^{k};\underline{U}_i^k\left(1_{[i-1]},\ell\right)|\underline{U}_{[i-1]}^{k}\left(1_{[i-1]}\right)\right)\right)|\mathcal{F}\right]\\
&=\exp\left(\imath\left(Y_{[i]}^{k};\underline{U}_i^k\left(1_{[i]}\right)|\underline{U}_{[i-1]}^{k}\left(1_{[i-1]}\right)\right)\right)+(L_i^k-1)\cdot\\
&\quad\mathbb{E}\left[\exp\left(\imath\left(Y_{[i]}^{k};\underline{U}_i^k(1_{[i-1]},2)|\underline{U}_{[i-1]}^{k}\left(1_{[i-1]}\right)\right)\right)\mid\mathcal{F}\right] \qquad (146)\\
&=\exp\left(\imath\left(Y_{[i]}^{k};\underline{U}_i^k(1_{[i]})|\underline{U}_{[i-1]}^{k}\left(1_{[i-1]}\right)\right)\right)+(L_i^k-1), \qquad (147)
\end{aligned}$$

where to write (146) we used that the codewords $\{\underline{U}_i^k(1_{[i-1]},\ell)\colon \ell\neq 1\}$ are identically distributed conditioned on $\mathcal{F}$.

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

$$\begin{aligned}
\mathcal{L}_i(\mathcal{K})\triangleq\Big\{\ell^{[K]}\in\prod_{i=1}^{K}[L_i^k]\colon\ &\ell^{\pi(k)}=1,k\in\mathcal{K},\\
&\ell^{\pi(k)}\neq 1,k\in\mathcal{K}^c\Big\}, \qquad (148)
\end{aligned}$$

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

$$\begin{aligned}
&\mathsf{g}\left(\underline{u}_{[i]}^{[K]}\left(1_{[i-1]}^{[K]},\ell_i^{[K]}\right)\right) \qquad (149)\\
&\le\prod_{k\in\mathcal{K}^c}1\left\{\jmath^{\pi(k)}\left(\underline{u}_{[i]}^{\pi([K])}\right)\ge\log\frac{L_i^{\pi(k)}}{M_i^{\pi(k)}}+\beta_i^{\pi(k)}\right\} \qquad (150)\\
&\le\prod_{k\in\mathcal{K}^c}\frac{M_i^{\pi(k)}}{L_i^{\pi(k)}}\exp\left(\jmath^{\pi(k)}\left(\underline{u}_{[i]}^{\pi([K])}\left(1_{[i-1]}^{[K]},\ell_i^{[K]}\right)\right)-\beta_i^{\pi(k)}\right), \qquad (151)
\end{aligned}$$

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

$$\mathsf{g}\left(\underline{u}_{[i]}^{[K]}\left(1_{[i]}^{[K]}\right)\right)\le 1. \qquad (152)$$

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

$$\mathbb{E}\left[\prod_{k\in\mathcal{K}^c}\exp\left(\jmath^{\pi(k)}\left(\underline{U}_{[i]}^{\pi([K])}\left(1_{[i-1]}^{[K]},\ell_i^{[K]}\right)\right)\right)|\mathcal{F}\right]=1. \qquad (153)$$

The upper-bound in (151) and the equality in (153) are key to the analysis of our GLD (136).

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

$$\begin{aligned}
&\mathbb{E}\left[G_i|\mathcal{F}\right] = \\
&\mathbb{E}\Bigg[\sum_{\ell_i^{[K]}\in\prod_{k=1}^K[L_i^k]} \mathsf{g}\left(\underline{U}_{[i]}^{[K]}\left(1_{[i-1]}^{[K]},\ell_i^{[K]}\right)\right) \\
&\qquad\qquad \cdot 1\left\{\mathsf{B}_i^{[K]}(\ell_i^{[K]}) = 1^{[K]}\right\} \mid \mathcal{F}\Bigg] \qquad (154)\\
&= \mathsf{g}\left(\underline{U}_{[i]}^{[K]}\left(1_{[i]}^{[K]}\right)\right) 1\left\{\mathsf{B}_i^{[K]}(1^{[K]}) = 1^{[K]}\right\} \\
&\quad + \sum_{\mathcal{K}\subset[K]} \mathbb{E}\left[\sum_{\ell_i^{[K]}\in\mathcal{L}_i(\mathcal{K})} \mathsf{g}\left(\underline{U}_{[i]}^{[K]}\left(1_{[i-1]}^{[K]},\ell_i^{[K]}\right)\right) \mid \mathcal{F}\right] \\
&\quad \cdot \prod_{k\in\mathcal{K}^c}\frac{1}{M_i^{\pi(k)}} \cdot 1\left\{\mathsf{B}_i^{\pi(\mathcal{K})}(1^{\mathcal{K}}) = 1^{\mathcal{K}}\right\} \qquad (155)\\
&\le 1\left\{\mathsf{B}_i^{[K]}(1^{[K]}) = 1^{[K]}\right\} \qquad (156)\\
&\quad + \sum_{\mathcal{K}\subset[K]} \exp\left(-\sum_{k\in\mathcal{K}^c}\beta_i^{\pi(k)}\right) 1\left\{\mathsf{B}_i^{\pi(\mathcal{K})}(1^{\mathcal{K}}) = 1^{\mathcal{K}}\right\},
\end{aligned}$$

where (156) follows from (151), (152) and (153).

Now, plugging (147) and (156) into (143) and computing the expectation in (143) with respect to the codebooks and the binning functions, we conclude that the probability of successful decoding is bounded below as

$$\begin{aligned}
&1-\epsilon \ge \qquad (157)\\
&\mathbb{E}\Bigg[\prod_{k=1}^K\prod_{i=1}^t \frac{1}{\frac{1}{L_i^k}\exp\left(\imath\left(Y_{[i]}^k;U_i^k|U_{[i-1]}^k\right)\right)+\left(1-\frac{1}{L_i^k}\right)} \\
&\cdot \frac{\mathsf{g}\left(U_{[i]}^{[K]}\right) 1\left\{\mathsf{d}\left(X_i,\hat{X}_i\left(U_{[i]}^{[K]}\right)\right)\le d_i\right\}}{1+\sum_{\mathcal{K}\subset[K]}\exp\left(-\sum_{k\in\mathcal{K}^c}\beta_i^k\right)}\Bigg].
\end{aligned}$$

Loosening the bound (157)*:* Here we again follow the recipe of Yassaee et al. [27], [28].

$$\begin{aligned}
&1-\epsilon \ge \\
&\mathbb{E}\Bigg[\prod_{k=1}^K\prod_{i=1}^t \frac{1}{(L_i^k)^{-1}\exp\left(\imath\left(Y_{[i]}^k;U_i^k|U_{[i-1]}^k\right)\right)+1} \\
&\cdot \frac{\mathsf{g}\left(U_{[i]}^{[K]}\right) 1\left\{\mathsf{d}\left(X_i,\hat{X}_i\left(U_{[i]}^{[K]}\right)\right)\le d_i\right\}}{\sum_{\mathcal{K}\subseteq[K]}\exp\left(-\sum_{k\in\mathcal{K}}\beta_i^k\right)}\Bigg] \qquad (158)\\
&\ge \prod_{k=1}^K\prod_{i=1}^t \frac{\mathbb{P}\left[\mathcal{E}^c\right]}{\left[1+\exp(-\alpha_i^k)\right]\left[\sum_{\mathcal{K}\subseteq[K]}\exp\left(-\sum_{k\in\mathcal{K}}\beta_i^k\right)\right]} \qquad (159)
\end{aligned}$$

where (158) holds by weakening (157) using $1-\left(L_t^k\right)^{-1}\le 1$ and rewriting for brevity

$$1+\sum_{\mathcal{K}\subset[K]}\exp\left(-\sum_{k\in\mathcal{K}^c}\beta_i^k\right) = \sum_{\mathcal{K}\subseteq[K]}\exp\left(-\sum_{k\in\mathcal{K}}\beta_i^k\right); \qquad (160)$$

(159) is obtained by weakening (158) by multiplying the random variable inside the expectation by $1\left\{\mathcal{E}^c\right\}$ and using the conditions in $\mathcal{E}$ (45) to upper-bound $\imath\left(Y_{[i]}^k;U_i^k|U_{[i-1]}^k\right)$ in the denominator.

Rewriting (159), we obtain

$$\begin{aligned}
\epsilon &\le 1- \qquad (161)\\
&\prod_{k=1}^K\prod_{i=1}^t \frac{\mathbb{P}\left[\mathcal{E}^c\right]}{\left[1+\exp(-\alpha_i^k)\right]\left[\sum_{\mathcal{K}\subseteq K}\exp(-\sum_{k\in\mathcal{K}}\beta_i^k)\right]} \\
&= \mathbb{P}\left[\mathcal{E}\right] + \gamma\,\mathbb{P}\left[\mathcal{E}^c\right] \qquad (162)\\
&\le \mathbb{P}\left[\mathcal{E}\right] + \gamma. \qquad (163)
\end{aligned}$$

Appendix B
Proof of Theorem 3

We analyze the bound in Theorem 2 with

$$P_{U_{[t]}^k\|Y_{[t]}^k} = P_{\mathsf{U}_{[t]}^k\|\mathsf{Y}_{[t]}^k}^{\otimes n}, \qquad (164)$$

$$P_{\hat{X}_{[t]}^{[K]}\|U_{[t]}^{[K]}} = P_{\hat{\mathsf{X}}_{[t]}^{[K]}\|\mathsf{U}_{[t]}^{[K]}}^{\otimes n}, \qquad (165)$$

single-letter kernels chosen so that

$$\mathbb{E}\left[\mathsf{d}\left(\mathsf{X}_i,\hat{\mathsf{X}}_i\left(\mathsf{U}_{[i]}^{[K]}\right)\right)\right] = d_i+\delta, \qquad (166)$$

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

$$\begin{aligned}
D_i^{\pi(k)} &\triangleq \mathbb{E}\left[\jmath^{\pi(k)}\left(\mathsf{U}_{[i]}^{\pi([K])}\right)\right] \qquad (167)\\
&= D\left(P_{\mathsf{U}_i^{\pi(k)}|\mathsf{U}_i^{\pi([k-1])}\mathsf{U}_{[i-1]}^{\pi([K])}}\|P_{\mathsf{U}_i^{\pi(k)}|\mathsf{U}_{[i-1]}^{\pi(k)}}|P_{\mathsf{U}_i^{\pi([k-1])}\mathsf{U}_{[i-1]}^{\pi([K])}}\right)
\end{aligned}$$

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

$$\alpha_i^k = \beta_i^k = n\delta, \qquad (168)$$

and choose L_i^k, M_i^k to satisfy

$$\log L_i^k \ge n\,I\left(\mathsf{Y}_{[i]}^k;\mathsf{U}_i^k|\mathsf{U}_{[i-1]}^k\right)+2\alpha_i^k, \qquad (169)$$

$$\log M_i^{\pi(k)} \ge \log L_i^{\pi(k)} - n\,D_i^{\pi(k)} + 2\beta_i^k. \qquad (170)$$

Note that since $\mathsf{U}_i^k - \left(\mathsf{Y}_{[i]}^k,\mathsf{U}_{[i-1]}^k\right) - \mathsf{U}_{[i]}^{[K]\setminus\{k\}}$, it holds that

$$\begin{aligned}
&I\left(\mathsf{Y}_{[i]}^{\pi(k)};\mathsf{U}_i^{\pi(k)}|\mathsf{U}_i^{\pi([k-1])},\mathsf{U}_{[i-1]}^{\pi([K])}\right) \\
&= I\left(\mathsf{Y}_{[i]}^{\pi(k)};\mathsf{U}_i^{\pi(k)}|\mathsf{U}_{[i-1]}^{\pi(k)}\right) - D_i^{\pi(k)}, \qquad (171)
\end{aligned}$$

and thus summing both sides of (170) over $i\in[t]$ we obtain (cf. (47))

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^t \log M_i^k \ge\ & I\left(\mathsf{Y}_{[t]}^{\pi(k)}\to\mathsf{U}_{[t]}^{\pi(k)}\|\mathsf{U}_{[t]}^{\pi([k-1])},\mathcal{D}\mathsf{U}_{[t]}^{[K]}\right) \\
&+ 4t\delta. \qquad (172)
\end{aligned}$$

Applying the union bound to $\mathbb{P}\left[\mathcal{E}\right]$ and the law of large numbers to each of the resultant $(2K+1)t$ terms, we further conclude that $\mathbb{P}\left[\mathcal{E}\right]\to 0$ as $n\to\infty$. Furthermore, $\gamma\to 0$ as $n\to\infty$, and therefore by Theorem 2 there exists a sequence of codes with $\log L_i^k$ and $\log M_i^k$ satisfying (169), (170) with excess-distortion probability $\epsilon\to 0$ as $n\to\infty$.

Under our assumption on the p-th moment of the distortion measure (25), the existence of an $(M_{[t]}^{[K]},d_{[t]},\epsilon)$ excess-distortion code with $\frac{1}{t}\sum_{i=1}^t d_i\le d$ 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 [45, Th. 25.5].

APPENDIX C
TWO CHARACTERIZATIONS OF BERGER-TUNG BOUND

Proposition 2. *The region $\mathcal{R}$ in (52) is equivalent to the region $\mathcal{R}'$ in (51).*

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

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

$$I(\mathsf{Y}^{\mathcal{K}};\mathsf{U}^{\mathcal{K}}) = \sum_{j=1}^{k} I(\mathsf{Y}_{\pi(j)};\mathsf{U}_{\pi(j)}|\mathsf{U}^{\pi[j-1]}), \tag{173}$$

$$I(\mathsf{Y}^{\mathcal{K}^c};\mathsf{U}^{\mathcal{K}^c}|\mathsf{U}^{\mathcal{K}}) = \sum_{j=k+1}^{K} I(\mathsf{Y}_{\pi(j)};\mathsf{U}_{\pi(j)}|\mathsf{U}^{\pi[j-1]}). \tag{174}$$

From (174), we conclude that any set of rates that satisfies (51) for π must also satisfy (52) 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 $\mathsf{U}^{\mathcal{B}} - \mathsf{Y}^{\mathcal{B}} - \mathsf{Y}^{\mathcal{A}\setminus\mathcal{B}} - \mathsf{U}^{\mathcal{A}\setminus\mathcal{B}}$, that for all $\mathcal{B} \subseteq \mathcal{A}$,

$$I(\mathsf{Y}^{\mathcal{A}};\mathsf{U}^{\mathcal{A}}) = I(\mathsf{Y}^{\mathcal{A}\setminus\mathcal{B}};\mathsf{U}^{\mathcal{A}\setminus\mathcal{B}}|\mathsf{U}^{\mathcal{B}}) + I(\mathsf{Y}^{\mathcal{B}};\mathsf{U}^{\mathcal{B}}). \tag{175}$$

Since

$$\begin{cases} S_1 \geq I_1 \\ S_1 + S_2 \geq I_1 + I_2 \end{cases} \iff \begin{cases} S_1 \geq I_1 \\ S_2 \geq I_2 \end{cases}, \tag{176}$$

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

$$\begin{cases} \sum_{k\in\mathcal{A}^c} R^k \geq I(\mathsf{Y}^{\mathcal{A}^c};\mathsf{U}^{\mathcal{A}^c}|\mathsf{U}^{\mathcal{A}}) \\ \sum_{k\in[K]} R^k \geq I(\mathsf{Y}^{[K]};\mathsf{U}^{[K]}) \end{cases} \tag{177}$$

$$\iff \begin{cases} \sum_{k\in\mathcal{A}} R^k \geq I(\mathsf{Y}^{\mathcal{A}};\mathsf{U}^{\mathcal{A}}) \\ \sum_{k\in\mathcal{A}^c} R^k \geq I(\mathsf{Y}^{\mathcal{A}^c};\mathsf{U}^{\mathcal{A}^c}|\mathsf{U}^{\mathcal{A}}) \end{cases} \tag{178}$$

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

$$\begin{cases} \sum_{k\in\mathcal{B}} R^k \geq I(\mathsf{Y}^{\mathcal{B}};\mathsf{U}^{\mathcal{B}}) \\ \sum_{k\in\mathcal{A}} R^k \geq I(\mathsf{Y}^{\mathcal{A}};\mathsf{U}^{\mathcal{A}}) \end{cases} \tag{179}$$

$$\iff \begin{cases} \sum_{k\in\mathcal{B}} R^k \geq I(\mathsf{Y}^{\mathcal{B}};\mathsf{U}^{\mathcal{B}}) \\ \sum_{k\in\mathcal{A}\setminus\mathcal{B}} R^k \geq I(\mathsf{Y}^{\mathcal{A}\setminus\mathcal{B}};\mathsf{U}^{\mathcal{A}\setminus\mathcal{B}}|\mathsf{U}^{\mathcal{B}}) \end{cases} \tag{180}$$

For $\mathcal{B} = \pi([k-1])$ and $\mathcal{A} = \pi([k])$, the second inequality in (180) is exactly the inequality (51). Since any set of rates satisfying (52) must also satisfy (180) for all $\mathcal{B} \subseteq \mathcal{A} \subseteq [K]$, we conclude that $\mathcal{R} \subseteq \mathcal{R}'$. □

APPENDIX D
MMSE ESTIMATION LEMMAS

Lemmas 2 and 3 are corollaries to the following result.

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

$$Y_k = X + W_k, \ k = 1, \ldots, K, \tag{181}$$

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

$$\mathbb{E}\left[X|Y_{[K]}\right] = \sum_{k=1}^{K} \frac{\sigma^2_{X|Y_{[K]}}}{\sigma^2_{W_k}} Y_k, \tag{182}$$

$$\frac{1}{\sigma^2_{X|Y_{[K]}}} = \frac{1}{\sigma_X^2} + \sum_{k=1}^{K} \frac{1}{\sigma^2_{W_k}}. \tag{183}$$

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

$$\mathbb{E}\left[X|Y=y\right] = \mathbb{E}\left[X\right] + \Sigma_{XY}\Sigma_{YY}^{-1}\left(y - \mathbb{E}\left[Y\right]\right), \tag{184}$$

$$\mathrm{Cov}[X|Y] = \Sigma_{XX} - \Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX}. \tag{185}$$

Denote for brevity

$$\Sigma_W \triangleq \begin{bmatrix} \sigma^2_{W_1} & & 0 \\ & \ddots & \\ 0 & & \sigma^2_{W_K} \end{bmatrix}. \tag{186}$$

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

$$\Sigma_{XX} = \sigma_X^2, \tag{187}$$

$$\Sigma_{YY} = \Sigma_W + \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \sigma_X^2 \begin{bmatrix} 1 & \cdots & 1 \end{bmatrix}, \tag{188}$$

$$\Sigma_{XY} = \sigma_X^2 \begin{bmatrix} 1 & \ldots & 1 \end{bmatrix}. \tag{189}$$

Using the matrix inversion lemma, we compute readily

$$\mathrm{Cov}[X|Y]^{-1} = \Sigma_{XX}^{-1} - \Sigma_{XX}^{-1}\Sigma_{XY}$$
$$\left(\Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY} - \Sigma_{YY}\right)^{-1}\Sigma_{YX}\Sigma_{XX}^{-1} \tag{190}$$

$$= \Sigma_{XX}^{-1} + \Sigma_{XX}^{-1}\Sigma_{XY}\Sigma_W^{-1}\Sigma_{YX}\Sigma_{XX}^{-1} \tag{191}$$

$$= \frac{1}{\sigma_X^2} + \frac{1}{\sigma^2_{W_1}} + \ldots + \frac{1}{\sigma^2_{W_K}}, \tag{192}$$

which shows (183). To show (182), we apply the matrix inversion lemma to Σ_{YY} to write:

$$\Sigma_{YY}^{-1} = \Sigma_W^{-1} - \Sigma_W^{-1} \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \sigma^2_{X|Y_{[K]}} \begin{bmatrix} 1 & \cdots & 1 \end{bmatrix} \Sigma_W^{-1}. \tag{193}$$

It's easy to verify that

$$\sigma_X^2 \begin{bmatrix} 1 & \ldots & 1 \end{bmatrix} \left(I_n \frac{1}{\sigma^2_{X|Y_{[K]}}} - \Sigma_W^{-1} \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \begin{bmatrix} 1 & \cdots & 1 \end{bmatrix} \right)$$
$$= \begin{bmatrix} 1 & \ldots & 1 \end{bmatrix}, \tag{194}$$

where I_n is the $n \times n$ identity matrix, so

$$\mathbb{E}\left[X|Y=y\right] = \Sigma_{XY}\Sigma_{YY}^{-1}y \tag{195}$$

$$= \begin{bmatrix} 1 & \ldots & 1 \end{bmatrix} \Sigma_W^{-1}\sigma^2_{X|Y_{[K]}}y, \tag{196}$$

which is equivalent to (182). □

Proof of Lemma 2. Equality (74) follows from

$$\sigma_Y^2 = \sigma_X^2 + \sigma_W^2, \tag{197}$$

$$\frac{1}{\sigma^2_{X|Y}} = \frac{1}{\sigma_X^2} + \frac{1}{\sigma_W^2}, \tag{198}$$

where (198) is a particularization of (183). □

Proof of Lemma 3. Notice that (75) with $\bar{X}_k = \mathbb{E}\left[X|Y_k\right]$ and $W_k' \sim \mathcal{N}(0, \sigma^2_{X|Y_k})$ is just another way to write (181). Reparameterizing (182) and (183) accordingly, one recovers (76) and (77). □

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

$$\bar{\mathsf{X}}_i = a\bar{\mathsf{X}}_{i-1} + \sum_{k=1}^{K} \frac{\sigma^2_{\mathsf{X}_i|\mathsf{Y}_{[i]}^{[K]}}}{\sigma^2_{\mathsf{W}_k}} \left(\mathsf{Y}_i^k - a\bar{\mathsf{X}}_{i-1}\right), \quad (199)$$

$$\frac{1}{\sigma^2_{\mathsf{X}_i|\mathsf{Y}_{[i]}^{[K]}}} = \frac{1}{\sigma^2_{\mathsf{X}_i|\mathsf{Y}_{[i-1]}^{[K]}}} + \sum_{k=1}^{K} \frac{1}{\sigma^2_{\mathsf{W}_k}}. \quad (200)$$

where $\bar{\mathsf{X}}_i$ is defined in (100). Equation (199) is the Kalman filter recursion with Kalman filter gain equal to the row vector $\sigma^2_{\mathsf{X}_i|\mathsf{Y}_{[i]}^{[K]}} \left(\frac{1}{\sigma^2_{\mathsf{W}_1}}, \ldots, \frac{1}{\sigma^2_{\mathsf{W}_K}}\right)$, and (200) is the corresponding Riccati recursion for the MSE.

Appendix E
Two equivalent representations of $R_{\mathrm{rm}}(d)$

In this appendix, we verify that (116) coincides with the lower bound on the causal remote rate-distortion function derived in [22]. Indeed, [22, Cor. 1 and Th. 9] imply

$$R_{\mathrm{rm}}(d) \geq \frac{1}{2} \log \left(a^2 + \frac{\sigma^2_{\mathsf{X}\|\mathcal{D}\mathsf{Y}^{[K]}} - \sigma^2_{\mathsf{X}\|\mathsf{Y}^{[K]}}}{d - \sigma^2_{\mathsf{X}\|\mathsf{Y}^{[K]}}} \right). \quad (201)$$

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

$$\bar{\mathsf{X}}_{i+1} = a\bar{\mathsf{X}}_i + \bar{\mathsf{V}}_i, \quad (202)$$

$\bar{\mathsf{V}}_i \sim \mathcal{N}(0, \sigma^2_{\mathsf{X}\|\mathcal{D}\mathsf{Y}^{[K]}} - \sigma^2_{\mathsf{X}\|\mathsf{Y}^{[K]}})$. The form in (201) leads to that in (116) via (59) and

$$\sigma^2_{\mathsf{X}\|\mathcal{D}\mathsf{Y}^{[K]}} = a^2\sigma^2_{\mathsf{X}\|\mathsf{Y}^{[K]}} + \sigma^2_{\mathsf{V}}. \quad (203)$$

□

Victoria Kostina (S'12–M'14) is a Professor of Electrical Engineering and of Computing and Mathematical Sciences at Caltech. She received a bachelor's degree from Moscow Institute of Physics and Technology (2004), where she was affiliated with the Institute for Information Transmission Problems of the Russian Academy of Sciences, a master's degree from University of Ottawa (2006), and a PhD from Princeton University (2013). She received the Natural Sciences and Engineering Research Council of Canada postgraduate scholarship (2009–2012), the Princeton Electrical Engineering Best Dissertation Award (2013), the Simons-Berkeley research fellowship (2015) and the NSF CAREER award (2017). Kostina's research spans information theory, coding, control, learning, and communications.

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

He has been with the California Institute of Technology since January 2001, where he is currently the Mose and Lilian S. Bohn Professor of Electrical Engineering. From 2013-2016 he was the Gordon M. Binder/Amgen Professor of Electrical Engineering and from 2008-2015 he was Executive Officer of Electrical Engineering, as well as Associate Director of Information Science and Technology. From October 1996 to October 1998 he was a research associate at the Information Systems Laboratory, Stanford University, and from November 1998 to December 2000 he was a Member of the Technical Staff in the Mathematical Sciences Research Center at Bell Laboratories, Murray Hill, NJ. He has also held short-term appointments at Ricoh California Research Center, the Indian Institute of Science, and Linkoping University, Sweden. His research interests include communications and information theory, control and network science, and signal processing and machine learning. He is the coauthor of the books (both with A.H. Sayed and T. Kailath) *Indefinite Quadratic Estimation and Control: A Unified Approach to H^2 and H^∞ Theories* (New York: SIAM, 1999) and *Linear Estimation* (Englewood Cliffs, NJ: Prentice Hall, 2000). He is a recipient of an Alborz Foundation Fellowship, the 1999 O. Hugo Schuck best paper award of the American Automatic Control Council (with H. Hindi and S.P. Boyd), the 2002 National ScienceFoundation Career Award, the 2002 Okawa Foundation Research Grant for Information and Telecommunications, the 2003 David and Lucille Packard Fellowship for Science and Engineering, the 2003 Presidential Early Career Award for Scientists and Engineers (PECASE), and the 2009 Al-Marai Award for Innovative Research in Communications, and was a participant in the 2004 National Academy of Engineering "Frontiers in Engineering"program.

He has been a Guest Editor for the IEEE Transactions on Information Theory special issue on "space-time transmission, reception, coding and signal processing" was an Associate Editor for Communications of the IEEE Transactions on Information Theory during 2004-2006, and is currently an Editor for the Journal "Foundations and Trends in Information and Communication" and for the IEEE Transactions on Network Science and Engineering. He is an IEEE Information Theory Society Distinguished Lecturer for 2016-2017 and was General Co-Chair if the 2020 IEEE International Symposium on Information Theory (ISIT 2020).