



AALBORG UNIVERSITY
DENMARK

Aalborg Universitet

Zero-Delay Multiple Descriptions of Stationary Scalar Gauss-Markov Sources

Fuglsig, Andreas Jonas; Østergaard, Jan

Published in:
Entropy

DOI (link to publication from Publisher):
[10.3390/e21121185](https://doi.org/10.3390/e21121185)

Creative Commons License
CC BY 4.0

Publication date:
2019

Document Version
Publisher's PDF, also known as Version of record

[Link to publication from Aalborg University](#)

Citation for published version (APA):
Fuglsig, A. J., & Østergaard, J. (2019). Zero-Delay Multiple Descriptions of Stationary Scalar Gauss-Markov Sources. *Entropy*, 21(12), Article 1185. <https://doi.org/10.3390/e21121185>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal -

Take down policy

If you believe that this document breaches copyright please contact us at vbn@aub.aau.dk providing details, and we will remove access to the work immediately and investigate your claim.

Article

Zero-Delay Multiple Descriptions of Stationary Scalar Gauss-Markov Sources

Andreas Jonas Fuglsig^{1,2}  and Jan Østergaard^{3,*} ¹ Department of Electronic Systems, Aalborg University, 9000 Aalborg, Denmark; ajf@rtx.dk² RTX A/S, 9400 Nørresundby, Denmark³ Section on Signal and Information Processing, Department of Electronic Systems, Aalborg University, 9000 Aalborg, Denmark

* Correspondence: jo@es.aau.dk

Received: 13 October 2019; Accepted: 28 November 2019; Published: 1 December 2019



Abstract: In this paper, we introduce the zero-delay multiple-description problem, where an encoder constructs two descriptions and the decoders receive a subset of these descriptions. The encoder and decoders are causal and operate under the restriction of zero delay, which implies that at each time instance, the encoder must generate codewords that can be decoded by the decoders using only the current and past codewords. For the case of discrete-time stationary scalar Gauss—Markov sources and quadratic distortion constraints, we present information-theoretic lower bounds on the average sum-rate in terms of the directed and mutual information rate between the source and the decoder reproductions. Furthermore, we show that the optimum test channel is in this case Gaussian, and it can be realized by a feedback coding scheme that utilizes prediction and correlated Gaussian noises. Operational achievable results are considered in the high-rate scenario using a simple differential pulse code modulation scheme with staggered quantizers. Using this scheme, we achieve operational rates within 0.415 bits/sample/description of the theoretical lower bounds for varying description rates.

Keywords: zero delay; multiple descriptions; Gauss-Markov; source coding; rate distortion; feedback

1. Introduction

Real-time communication is desirable in many modern applications, e.g., Internet of Things [1], audio transmission for hearing aids [2], stereo audio signals [3], on-line video conferencing [4], or systems involving feedback, such as networked control systems [5–7]. All these scenarios may operate under strict requirements on latency and reliability. Particularly, delays play a critical role in the performance or stability of these systems [8].

In near real-time communication over unreliable networks, and where retransmissions are either not possible or not permitted, e.g., due to strict latency constraints, it is generally necessary to use an excessive amount of bandwidth for the required channel code in order to guarantee reliable communications and ensure satisfactory performance. Several decades ago, it was suggested to replace the channel code by cleverly designed data packets, called multiple descriptions (MDs) [9]. Contrary to channel codes, MDs would allow for several reproduction qualities at the receivers and thereby admit a graceful degradation during partial network failures [9]. In MD coding, retransmissions are not necessary, which is similar to the case of forward error correction coding. Thus, with MDs, one avoids the possible long delay due to loss of packets or acknowledgement. Hence, some compression (reproduction quality) is sacrificed for an overall lower latency [9]. Interestingly, despite their potential advantages over channel codes for certain applications, MD codes are rarely used in practical communication systems with feedback. The reasons are that from a practical point of view, good MD

codes are application-specific and hard to design, and from a theoretical point of view, zero-delay MD (ZDMD) coding and MD coding with feedback remain open and challenging topics.

1.1. Multiple Descriptions

MD coding can be described as a data compression methodology, where partial information about the data source is compressed into several data files (called descriptions or data packets) [10,11]. The descriptions can, for example, be individually transmitted over different channels in a network. The descriptions are usually constructed such that when any single description is decoded, it is possible to reconstruct an approximation of the original uncompressed source. Since this is only an approximation of the data source, there will inevitably be a reconstruction error, which yields a certain degree of distortion. The distinguishing aspect of MD coding over other coding methodologies is that if more than one description is retrieved, then a better approximation of the source is achieved than what is possible when only using a single description. As more descriptions are combined, the quality of the reproduced source increases. Similarly, this allows for a graceful degradation in the event of, e.g., packet dropouts on a packet-switched network such as the Internet.

Figure 1 illustrates the two-description MD coding scenario in both a closed-loop and an open-loop system. In both cases, the encoder produces two descriptions which are transmitted across noiseless channels, i.e., no bit-errors are introduced in the descriptions between the encoder and decoders. Some work exists in the closed-loop scenario, but no complete solution has been determined. However, the noncausal open-loop problem has been more widely studied in the information-theory literature [9–14].

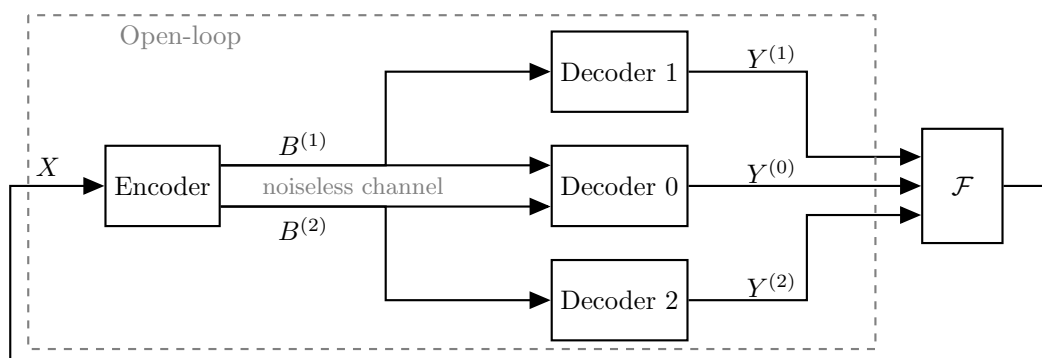


Figure 1. Multiple description (MD) source coding in a closed loop. If packet loss occurs on the noiseless channel, it will affect the source signal, X , differently depending on which descriptions are received. The standard open-loop MD coding is marked by the dashed line. In the open loop, the source is completely specified prior to the design of the coding scheme.

Since MD coding considers several data rates and distortions, MD rate-distortion theory is the determination of the fundamental limits on a rate-distortion region [9]. That is, determine the minimum individual rates required to achieve a given set of individual and joint distortion constraints. A noncausal achievable MD rate-distortion region is only completely known in very few cases [12]. El-Gamal and Cover [11] gave an achievable region for two descriptions and memoryless source. This region was then shown to be tight for white Gaussian sources with mean-squared error (MSE) distortion constraints by Ozarow [10]. In the high resolution limit, i.e., high rates, the authors of [13] characterized the achievable region for stationary (time-correlated) Gaussian sources with MSE distortion constraints. This was then extended in [14] to the general resolution case for stationary Gaussian sources. Recently, the authors of [12] showed in the symmetric case, i.e., equal rates and distortions for each individual description, that the MD region for a colored Gaussian source subject to MSE distortion constraints can be achieved by predictive coding using filtering. However, similar to single-description source coding [8], the MD source coders whose performance is close to the

fundamental rate-distortion bounds impose long delays on the end-to-end processing of information, i.e., the total delay only due to source coding [15].

1.2. Zero Delay

Clearly, in near real-time communication, the source encoder and decoder must have zero delay. The term zero-delay (ZD) source coding is often used when both instantaneous encoding and decoding are required [16]. That is, when the reconstruction of each input sample must take place at the same time-instant, the corresponding input sample has been encoded [17]. For near instantaneous coding, the source coders must be causal [18]. However, causality comes with a price. The results of [17] showed that causal coders increase the bit-rate due to the space-filling loss of “memoryless” quantizers, and the reduced de-noising capabilities of causal filters. Additionally, imposing ZD increases the bit-rate due to memoryless entropy coding [17].

In the single-description case, ZD rate-distortion theory has been increasingly more popular in recent decades, due to its significance in real-time communication systems and especially feedback systems. Some indicative results on ZD source coding for networked control systems and systems with and without feedback may be found in [5–8,17,19–21]. The results of [5] establish a novel information-theoretic lower bound on the average data-rate for a source coding scheme within a feedback loop by the directed information rate across the channel. For open-loop vector Gauss-Markov sources, i.e., when the source is not inside a feedback loop, the optimal operational performance of a ZD source code subject to an MSE distortion constraint has been shown to be lower bounded by a minimization of the directed information [22] from the source to the reproductions subject to the same distortion constraint [5–7,17,19]. For Gaussian sources, the directed information is further minimized by Gaussian reproductions [8,20]. Very recently, Stavrou et al. [8], extending upon the works of [6,7,17,19], showed that the optimal test channel that achieves this lower bound is realizable using a feedback realization scheme. Furthermore, Ref [8] extended this to a predictive coding scheme providing an achievable upper bound on the operational performance subject to an MSE distortion constraint.

1.3. Zero-Delay Multiple Descriptions

Recently, the authors of [15] proposed an analog ZDMD joint source-channel coding scheme, such that the analog source output is mapped directly into analog channel inputs, thus not suffering from the delays encountered in digital source coding. However, for analog joint source-channel coding to be effective, the source and channel must be matched, which rarely occurs in practice [23]. Furthermore, most modern communication systems rely on digital source coding. Thus, analog joint source-channel coding is only applicable in a very limited amount of settings. Digital low-delay MD coding for practical audio transmission has been explored in, e.g., [2,4,24], as well as for low-delay video coding in [25]. Some initial work regarding MDs in networked control systems may be found in [26]. However, none of these consider the theoretical limitations of ZDMD coding in a rate-distortion sense.

In this paper, we propose a combination of ZD and MD rate-distortion theory such that the MD encoder and decoders are required to be causal and of zero delay. For the case of discrete-time stationary scalar Gauss-Markov sources and quadratic distortion constraints, we present information-theoretic lower bounds on the average sum-rate in terms of the directed and mutual information rate between the source and the decoder reproductions. We provide proof of achievability via a new Gaussian MD test channel and show that this test channel can be realized by a feedback coding scheme that utilizes prediction and correlated Gaussian noises. We finally show that a simple scheme using differential pulse code modulation with staggered quantizers can get close to the optimal performance. Specifically, our simulation study reveals that for a wide range of description rates, the achievable operational rates are within 0.415 bits/sample/description of the theoretical lower bounds. Further simulations and more details regarding the combination of ZD and MD coding are provided in the report [27].

The rest of the paper is organized as follows. In Section 2, we characterize the ZMDM source coding problem with feedback for stationary scalar Gauss-Markov sources subject to asymptotic MSE distortion constraints. Particularly, we consider the symmetric case in terms of the symmetric ZMDM rate-distortion function (RDF). In Section 3, we introduce a novel information-theoretic lower bound on the average data sum-rate of a ZMDM source code. For scalar stationary Gaussian sources, we show this lower bound is minimized by jointly Gaussian MDs, given that certain technical assumptions are met. This provides an information-theoretic lower bound to the symmetric ZMDM RDF. In Section 4, we determine an MD feedback realization scheme for the optimum Gaussian test-channel distribution. Utilizing this, we present a characterization of the Gaussian achievable lower bound as a solution to an optimization problem. In Section 5, we evaluate the performance of an operational staggered predictive quantization scheme compared to the achievable ZMDM region. We then discuss and conclude on our results. Particularly, we highlight some important difficulties with the extension to the Gaussian vector case.

2. Problem Definition

In this paper, we consider the ZMDM source coding problem with feedback illustrated in Figure 2. The feedback channels are assumed to be noiseless digital channels and have a one-sample delay to ensure the operational feasibility of the system, i.e., at any time, the current encoder outputs only depend on previous decoder outputs.

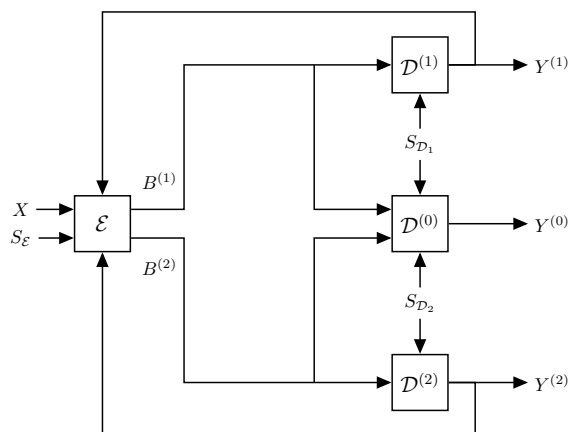


Figure 2. A general MD source-coding scenario with feedback.

Here, the stationary scalar Gauss-Markov source process is determined by the following discrete-time linear time-invariant model:

$$X_{k+1} = aX_k + W_k, \quad k \in \mathbb{N}, \tag{1}$$

where $|a| < 1$ is the deterministic correlation coefficient, $X_1 \in \mathbb{R} \sim \mathcal{N}(0, \sigma_{X_1}^2)$ is the initial state, $\sigma_{X_1}^2 = \frac{\sigma_W^2}{1-a^2}$, and $W_k \in \mathbb{R} \sim \mathcal{N}(0, \sigma_W^2)$ is an independent and identically distributed (IID) Gaussian process independent of $\{X_k : k \in \mathbb{N}\}$. For each time step $k \in \mathbb{N}$, the ZMDM encoder, \mathcal{E} , observes a new source sample X_k while assuming it has already observed the past sequence X^{k-1} . The encoder then produces two binary descriptions $B_k^{(1)}, B_k^{(2)}$ with lengths $l_k^{(1)}, l_k^{(2)}$ (in bits) from two predefined sets of codewords $\mathcal{B}_k^{(1)}, \mathcal{B}_k^{(2)}$, of at most a countable number of codewords, i.e., the codewords are discrete random variables. The codewords are transmitted across two instantaneous noiseless digital channels to the three reconstruction decoders, $\mathcal{D}^{(0)}, \mathcal{D}^{(1)}$, and $\mathcal{D}^{(2)}$. The decoders then immediately decode the binary codewords. Upon receiving $B_k^{(i),k}$, the i th side decoder, $\mathcal{D}^{(i)}$, $i = 1, 2$, produces an estimate $Y_k^{(i)}$ of the source sample X_k , under the assumption that $Y_k^{(i),k-1}$ is already produced. Similarly, the central

decoder, $\mathcal{D}^{(0)}$, upon receiving $B^{(1),k}, B^{(2),k}$, produces an estimate $Y_k^{(0)}$ of X_k under the assumption $Y^{(0),k}$ is already produced. Finally, before generating the current binary codewords, the encoder receives the two reproductions from the previous time step $Y_{k-1}^{(1)}, Y_{k-1}^{(2)}$ while assuming it has already received the past, $Y^{(1),k-2}, Y^{(2),k-2}$.

We assume the encoder and all decoders process information without delay. That is, each sample is processed immediately and without any delays for each time step $k \in \mathbb{N}$.

In the system, $S_{\mathcal{E},k}$ is the side information that becomes available at time-instance k at the encoder, and similarly, $S_{\mathcal{D}_i,k}$ is the new side information at reproduction decoder i . We emphasize, this is not side information in the usual information-theoretic sense of multiterminal source coding or Wyner–Ziv source coding, where the side information is unknown, jointly distributed with the source, and only available at the decoder, e.g., some type of channel-state information [28,29]. In this paper, our encoders and decoders are deterministic. However, to allow for probabilistic encoders and decoders, we let the deterministic encoders and decoders depend upon a stochastic signal, which we refer to as the side information. To make the analysis tractable, we require this side information to be independent of the source. The side information could, for example, represent dither signals in the quantizers, which is a common approach in the source coding literature [30]. We shortly discuss the possibility of removing this independence assumption in Section 6.

We do not need feedback from the central decoder, since all information regarding $Y^{(0),k-1}$ is already contained in $(Y^{(1),k-1}, Y^{(2),k-1})$. That is, given the side information, the side decoder reproductions are sufficient statistics for the central reproduction, and the following Markov chain holds,

$$X^k |_{\phi} - (Y^{(1),k}, Y^{(2),k}) |_{\phi} - Y^{(0),k} |_{\phi}, \tag{2}$$

where $\phi = (S_{\mathcal{D}_1}^k, S_{\mathcal{D}_2}^k)$. We note, this Markov chain also requires the decoders are invertible as defined in Definition 5 on page 10. Requiring invertible decoders is optimal in causal source coding [5].

Zero-delay multiple-description source coding with side information: We specify in detail the operations of the different blocks in Figure 2. First, at each time step, k , all source samples up to time k , X^k , and all previous reproductions, $Y^{(i),k-1}$, $i = 1, 2$, are available to the encoder, \mathcal{E} . The encoder then performs lossy source coding and lossless entropy coding to produce two dependent codewords. That is, the encoder block can be conceptualized as being split into a quantization step and an entropy coding step as illustrated in Figure 3. This is a very simplified model, and each of the quantization and entropy coding steps may be further decomposed as necessary to generate the appropriate dependent messages. However, this is a nontrivial task, and therefore, for a more tractable analysis and ease of reading, we do not further consider this two-step procedure in the theoretical derivations.

The zero-delay encoder is specified by the sequence of functions $\{\mathcal{E}_k : k \in \mathbb{N}\}$, where:

$$\mathcal{E}_k : \mathcal{X}^k \times \mathcal{Y}^{(1),k-1} \times \mathcal{Y}^{(2),k-1} \times S_{\mathcal{E}}^k \rightarrow B_k^{(1)} \times B_k^{(2)}, \tag{3}$$

and at each time step $k \in \mathbb{N}$, the encoder outputs the messages:

$$(B_k^{(1)}, B_k^{(2)}) = \mathcal{E}_k (X^k, Y^{(1),k-1}, Y^{(2),k-1}, S_{\mathcal{E}}^k), \quad k \in \mathbb{N}, \tag{4}$$

with length $l_k^{(i)}$ $i = 1, 2$ (in bits), where for the initial encoding, there are no past reproductions available at the encoder, hence $(B_1^{(1)}, B_1^{(2)}) = \mathcal{E}_1 (X_1, S_{\mathcal{E},1})$.

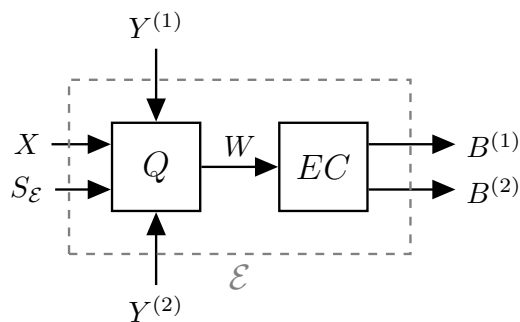


Figure 3. Conceptual model of splitting the zero-delay MD (ZDMD) encoder, \mathcal{E} , into a lossy quantizer, Q , and a lossless entropy coder, EC . W is a p -dimensional signal, where p is appropriately chosen according to the employed quantization procedure.

The zero-delay decoders are specified by the three sequences of functions $\{\mathcal{D}_k^{(0)}, \mathcal{D}_k^{(1)}, \mathcal{D}_k^{(2)} : k \in \mathbb{N}\}$, where:

$$\mathcal{D}_k^{(i)} : \mathcal{B}^{(i),k} \times \mathcal{S}_{\mathcal{D}_i}^k \rightarrow \mathcal{Y}_k^{(i)}, \quad i = 1, 2, \tag{5}$$

$$\mathcal{D}_k^{(0)} : \mathcal{B}^{(1),k} \times \mathcal{B}^{(2),k} \times \mathcal{S}_{\mathcal{D}_1}^k \times \mathcal{S}_{\mathcal{D}_2}^k \rightarrow \mathcal{Y}_k^{(0)}. \tag{6}$$

At each time step, $k \in \mathbb{N}$ the decoders generate the outputs:

$$Y_k^{(i)} = \mathcal{D}_k^{(i)} \left(B^{(i),k}, \mathcal{S}_{\mathcal{D}_i}^k \right), \quad i = 1, 2, \tag{7}$$

$$Y_k^{(0)} = \mathcal{D}_k^{(0)} \left(B^{(1),k}, B^{(2),k}, \mathcal{S}_{\mathcal{D}_1}^k, \mathcal{S}_{\mathcal{D}_2}^k \right), \tag{8}$$

assuming $Y^{(i),k-1}$, $i = 0, 1, 2$ have already been generated, with:

$$Y_1^{(i)} = \mathcal{D}_1^{(i)} \left(B_1^{(i)}, \mathcal{S}_{\mathcal{D}_i,1} \right), \quad i = 1, 2, \tag{9}$$

$$Y_1^{(0)} = \mathcal{D}_1^{(0)} \left(B_1^{(1)}, B_1^{(2)}, \mathcal{S}_{\mathcal{D}_1,1}, \mathcal{S}_{\mathcal{D}_2,1} \right). \tag{10}$$

The ZDMD source code produces two descriptions of the source; hence, we may associate the ZDMD code with a rate pair.

Definition 1 (Rate pair of ZDMD code). For each time step, k , let $l_k^{(i)}$ be the length in bits of the i th encoder output in a ZDMD source code as described above. Then, the average expected data-rate pair, (R_1, R_2) , measured in bits per source sample, are the rates:

$$R_i = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \mathbb{E} \left[l_k^{(i)} \right], \quad i = 1, 2. \tag{11}$$

Asymptotic MSE distortion constraints: A rate pair (R_1, R_2) is said to be achievable with respect to the MSE distortion constraints $D_i > 0$, $i = 0, 1, 2$, if there exists a rate- (R_1, R_2) ZDMD source code as described above, such that:

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \mathbb{E} \left[\left(X_k - Y_k^{(i)} \right)^2 \right] \leq D_i, \quad i = 0, 1, 2, \tag{12}$$

is satisfied.

Similarly to standard MD theory [31], the main concern of ZDMD coding is to determine the ZDMD rate-region, constituting the set of all achievable rate pairs for given distortion constraints.

Definition 2 (ZDMD rate-region). *For the stationary source process $\{X_k\}$, $X_k \in \mathcal{X}$, the ZDMD rate-region $\mathcal{R}_X^{ZD}(R_1, R_2, D_0, D_1, D_2)$ is the convex closure of all achievable ZDMD rate pairs (R_1, R_2) with respect to the MSE distortion constraints (D_0, D_1, D_2) .*

The ZDMD rate-region can be fully characterized by determining the bound between the sets of achievable and non-achievable rates, i.e., by determining the fundamental smallest achievable rates for given distortion constraints. Particularly, we consider so-called nondegenerate distortion constraints [32], that is, triplets (D_0, D_1, D_2) that satisfy:

$$D_1 + D_2 - \sigma_X^2 \leq D_0 \leq \left(\frac{1}{D_1} + \frac{1}{D_2} - \frac{1}{\sigma_X^2} \right)^{-1}, \tag{13}$$

where σ_X^2 is the stationary variance of the source.

The previous design requirements are summarized in the ZDMD coding problem with feedback.

Problem 1 (ZDMD coding problem with feedback). *For a discrete-time stationary scalar source process $\{X_k\}$, with nondegenerate MSE distortion constraints, $D_0, D_1, D_2 > 0$. Determine the minimum operational rates R_1, R_2 of the ZDMD coding scheme with side information from Equations (3)–(8), such that the asymptotic average expected distortions satisfy:*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \mathbb{E} \left[\left(X_k - Y_k^{(i)} \right)^2 \right] \leq D_i, \quad i = 0, 1, 2. \tag{14}$$

where the minimum is over all possible ZDMD encoder and decoder sequences $\{\mathcal{E}_k\}_{k \in \mathbb{N}}$, $\{\mathcal{D}_k^{(i)}\}_{k \in \mathbb{N}}$, $i = 0, 1, 2$ that satisfy Equations (3)–(8).

In this paper, we mainly consider the symmetric case of $R_1 = R_2 = R$ and $D_1 = D_2 = D_S$. Here, the ZDMD region may be completely specified by an MD equivalent of the standard RDF [12].

Definition 3 (Symmetric ZDMD RDF). *The symmetric ZDMD RDF for a source, $\{X\}$, with MSE distortion constraints, $D_0, D_S > 0$, is:*

$$R_{ZD}^{op}(D_0, D_S) \triangleq \inf R \quad \text{s.t.} \quad (R, R) \in \mathcal{R}_X^{ZD}(R, R, D_0, D_S, D_S). \tag{15}$$

That is the minimum rate R per description, which is achievable with respect to the distortion pair (D_0, D_S) .

The operational symmetric ZDMD RDF can be expressed in terms of the sum-rate, $R_1 + R_2$.

Problem 2 (Operational symmetric scalar Gaussian ZDMD RDF). *For a stationary scalar Gauss-Markov source process (1), with nondegenerate MSE distortion constraints, $D_0, D_S > 0$, determine the operational symmetric ZDMD RDF, i.e., solve the optimization problem:*

$$\begin{aligned}
 R_{ZD}^{op}(D_0, D_S) = \inf & \quad \lim_{n \rightarrow \infty} \frac{1}{2n} \left(\mathbb{E} \left[L_n^{(1)} \right] + \mathbb{E} \left[L_n^{(2)} \right] \right) \\
 \text{s.t.} & \quad \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \mathbb{E} \left[\left(X_k - Y_k^{(0)} \right)^2 \right] \leq D_0 \\
 & \quad \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \mathbb{E} \left[\left(X_k - Y_k^{(i)} \right)^2 \right] \leq D_S, \quad i = 1, 2,
 \end{aligned} \tag{16}$$

where $L_n^{(i)} \triangleq \sum_{k=1}^n l_k^{(i)}$, $i = 1, 2$, and the infimum is over all possible ZDMD encoder- and -decoder sequences $\{\mathcal{E}_n\}_{n \in \mathbb{N}}$, $\{\mathcal{D}_n^{(0)}\}_{n \in \mathbb{N}}$, $\{\mathcal{D}_n^{(1)}\}_{n \in \mathbb{N}}$, $\{\mathcal{D}_n^{(2)}\}_{n \in \mathbb{N}}$, i.e., that satisfy Equations (3)–(8).

Unfortunately, the solutions to Problems 1 and 2 are very hard to find, since they are determined by a minimization over all possible operational ZDMD codes. Similar to single description ZD rate-distortion theory [17], where the classical RDF is a lower bound on the zero-delay RDF, the noncausal arbitrary delay MD region [10,14] is an outer bound on the ZDMD region. However, this is a conservative bound due to the space-filling losses, memoryless entropy coding, and causal filters suffered by the ZD coders. Therefore, we introduce a novel information-theoretic lower bound on the operational ZD coding rates. As in classical MD rate-distortion theory, this bound is given in terms of lower bounds on the marginal rates, R_1 , R_2 , and the sum-rate, $R_1 + R_2$ cf. [10,11].

3. Lower Bound on Average Data-Rate

In this section, we determine a novel information-theoretic lower bound on the sum-rate of ZDMD source coding with feedback. Using this lower bound, we present an information-theoretic counterpart of the operational symmetric Gaussian ZDMD RDF. Finally, we provide a lower bound to Problem 2 by showing, for stationary scalar Gaussian sources, that Gaussian reproductions minimize the information-theoretic lower bound, given some technical assumptions are met. Although our main concern is the symmetric case, some of our main results are provided in the general nonsymmetric case.

We study a lower bound on the sum-rate of the ZDMD coding problem with feedback, which only depends on the joint statistics of the source encoder input, X , and the decoder outputs, $Y^{(i)}$ $i = 0, 1, 2$. To this end, we present in more detail the test-channel distribution associated with this minimization.

3.1. Distributions

We consider a source that generates a stationary sequence $X_k = x_k \in \mathcal{X}_k$, $k \in \mathbb{N}^n$. The objective is to reproduce or reconstruct the source by $Y_k^{(i)} = y_k^{(i)} \in \mathcal{Y}_k^{(i)}$, $k \in \mathbb{N}^n$, $i = 0, 1, 2$, subject to MSE fidelity criteria $d_{1,n}^{(i)}(x^n, y^{(i),n}) \triangleq \frac{1}{n} \sum_{k=1}^n (x_k - y_k^{(i)})^2$, $i = 0, 1, 2$.

Source. We consider open-loop source coding; hence, we assume the source distribution satisfies the following conditional independence:

$$P \left(x_k | x^{k-1}, y^{(0),k-1}, y^{(1),k-1}, y^{(2),k-1} \right) \triangleq P \left(x_k | x^{k-1} \right), \quad k \in \mathbb{N}^n. \tag{17}$$

This implies that the source, X , is unaffected by the feedback from the reproductions, $Y^{(i)}$. Hence, the next source symbol, given the previous symbols, is not further related to the previous reproductions [22].

We assume the distribution at $k = 1$ is $P(x_1)$. Furthermore, by Bayes' rule [8]:

$$P(x^n) \triangleq \prod_{k=1}^n P(x_k | x^{k-1}). \tag{18}$$

For the Gauss-Markov source process (1), this implies $\{W_k\}$ is independent of the past reproductions $Y^{(i),k-1}$, $i = 0, 1, 2$ [8].

Reproductions. Since the source is unaffected by the feedback from the reproductions, the MD encoder–decoder pairs from \mathcal{E} to \mathcal{D}_i , $i = 0, 1, 2$, in Figure 2, are causal if, and only if, the following Markov chain holds [17]:

$$X_{k+1}^n - X^k - (Y^{(0),k}, Y^{(1),k}, Y^{(2),k}), \quad \forall k \in \{1, \dots, n-1\}. \tag{19}$$

Hence, we assume the reproductions are randomly generated according to the collection of conditional distributions:

$$P(y_k^{(0)}, y_k^{(1)}, y_k^{(2)} | y^{(0),k-1}, y^{(1),k-1}, y^{(2),k-1}, x^k), \quad k \in \mathbb{N}^n. \tag{20}$$

For the first time step, $k = 1$, we assume:

$$P(y_1^{(0)}, y_1^{(1)}, y_1^{(2)} | y^{(0),0}, y^{(1),0}, y^{(2),0}, x^1) = P(y_1^{(0)}, y_1^{(1)}, y_1^{(2)} | x_1). \tag{21}$$

3.2. Bounds

We define the directed information rate across a system with random input and random output processes.

Definition 4 (Directed information rate ([5] Def. 4.3)). *The directed information rate across a system with random input, X , and random output, Y , is defined as:*

$$\bar{I}(X \rightarrow Y) \triangleq \lim_{n \rightarrow \infty} \frac{1}{n} I(X^n \rightarrow Y^n) \tag{22}$$

where $I(X^n \rightarrow Y^n)$ is the directed information between the two sequences X^n and Y^n , defined as:

$$I(X^n \rightarrow Y^n) \triangleq \frac{1}{n} \sum_{k=1}^n I(X^k; Y_k | Y^{k-1}). \tag{23}$$

In order to establish an outer bound on the ZDMD rate-region, we need a lower bound on the marginal rates and the sum-rate. By the results of [5,8], it can be shown that the marginal operational rates, R_1, R_2 are lower bounded by:

$$R_i \geq \bar{I}(X \rightarrow Y^{(i)}), \tag{24}$$

$$= \lim_{n \rightarrow \infty} \frac{1}{n} I(X^n \rightarrow Y^{(i),n}), \tag{25}$$

$$= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n I(X^k; Y_k^{(i)} | Y^{(i),k-1}), \quad i = 1, 2, \tag{26}$$

that is, by the directed information rate from the source to the side description. Thus, in order to determine a bound on the ZDMD rate-region, it remains to determine an information-theoretic lower bound on the sum-rate. Our derivation of the lower bound on the sum-rate requires the following assumption.

Assumption 1. *The systems $\mathcal{E}, \mathcal{D}^{(i)}$ $i = 0, 1, 2$, are causal, described by Equations (3)–(8), and $(\{S_{\mathcal{D}_1}\}, \{S_{\mathcal{D}_2}\}) \perp\!\!\!\perp \{X_k\}$, i.e., the side information is independent of the source sequence, $\{X_k\}$.*

We consider this assumption to be reasonable in a ZD scenario, i.e., the deterministic encoders and decoders must be causal and use only past and present symbols, and side information that is not

associated with the source signal [5]. Similar to [5], the channel is the only link between encoder and decoder. However, we further assume the channel to have perfect feedback.

Additionally, we require the decoders to be invertible given the side information.

Definition 5 (Invertible decoder ([5] Def. 4.2)). *The decoders, $\mathcal{D}^{(i)}$, $i = 0, 1, 2$, defined in Equations (7) and (8) are said to be invertible if, and only if, $\forall k \in \mathbb{N}$, there exists deterministic mappings $\mathcal{G}_k^{(i)}$, $i = 0, 1, 2$, such that:*

$$B^{(1),k} = \mathcal{G}_k^{(1)} \left(Y_k^{(1)}, S_{\mathcal{D}_1}^k \right), \tag{27}$$

$$B^{(2),k} = \mathcal{G}_k^{(2)} \left(Y_k^{(2)}, S_{\mathcal{D}_2}^k \right), \tag{28}$$

$$\left(B^{(1),k}, B^{(2),k} \right) = \mathcal{G}_k^{(0)} \left(Y_k^{(0)}, S_{\mathcal{D}_1}^k, S_{\mathcal{D}_2}^k \right). \tag{29}$$

If the decoders are invertible, then for each side decoder, knowledge of the side information and the output, e.g., $(Y_k^{(1)}, S_{\mathcal{D}_1}^k)$, is equivalent to knowledge of the side information and the input, $(B^{(1),k}, S_{\mathcal{D}_1}^k)$ [5]. For the single description case, it is shown in [5] that without loss of generality, we can restrict our attention to invertible decoders. Furthermore, when minimizing the average data-rate in a causal source coding scheme, it is optimal to minimize the average data-rate by focusing on schemes with invertible decoders [5].

The following results are used to prove the first main result of this section and are a generalization of ([5] Lemma 4.2) to the MD scenario.

Lemma 1 (Feedback Markov Chains). *Consider an MD source coding scheme inside a feedback loop as shown in Figure 2. If Assumption 1 applies and if the decoders are invertible when given the side information, then the Markov chain:*

$$X^k |_{\phi_1} - \left(B_k^{(1)}, B_k^{(2)} \right) |_{\phi_1} - \left(Y_k^{(1)}, Y_k^{(2)} \right) |_{\phi_1}, \quad k \in \mathbb{N}, \tag{30}$$

holds, with $\phi_1 = \left(B^{(1),k-1}, B^{(2),k-1}, S_{\mathcal{D}_1}^k, S_{\mathcal{D}_2}^k \right)$.

Furthermore, let $\phi_2 = \left(B^{(1),k-1}, S_{\mathcal{D}_1}^k \right)$ then:

$$Y_k^{(2)} |_{\phi_2} - B_k^{(1)} |_{\phi_2} - Y_k^{(1)} |_{\phi_2}, \quad k \in \mathbb{N}, \tag{31}$$

also holds.

Additionally, for $\phi_3 = \left(B^{(2),k-1}, S_{\mathcal{D}_2}^k \right)$:

$$Y_k^{(1)} |_{\phi_3} - B_k^{(2)} |_{\phi_3} - Y_k^{(2)} |_{\phi_3}, \quad k \in \mathbb{N}, \tag{32}$$

holds.

Finally, if the decoder side information is mutually independent, i.e., $\{S_{\mathcal{D}_1}\} \perp \{S_{\mathcal{D}_2}\}$, the Markov chains:

$$Y^{(2),k-1} - Y^{(1),k-1} - S_{\mathcal{D}_1}^k, \quad k \in \mathbb{N}, \tag{33}$$

$$Y^{(1),k} - Y^{(2),k-1} - S_{\mathcal{D}_2}^k, \quad k \in \mathbb{N}, \tag{34}$$

hold.

Proof. The Markov chain in Equation (30) follows, since $Y_k^{(1)}, Y_k^{(2)}$ depend deterministically upon $(B^{(1),k}, B^{(2),k}, S_{\mathcal{D}_1}^k, S_{\mathcal{D}_2}^k)$. Similarly, Equation (31) holds, since $Y_k^{(1)}$ depends deterministically upon $(B^{(1),k}, S_{\mathcal{D}_1}^k)$. The Markov chain in Equation (32) follows analogously.

By the system equations, we have that:

$$(B_1^{(1)}, B_1^{(2)}) = \mathcal{E}_1(X_1, \emptyset, \emptyset, S_{\mathcal{E},1}) \tag{35}$$

$$Y_1^{(1)} = \mathcal{D}_1^{(1)}(B_1^{(1)}, S_{\mathcal{D}_1,1}) \tag{36}$$

$$Y_1^{(2)} = \mathcal{D}_1^{(2)}(B_1^{(2)}, S_{\mathcal{D}_2,1}). \tag{37}$$

Since $\mathcal{S}_{\mathcal{D}_1,1} \perp\!\!\!\perp \mathcal{S}_{\mathcal{D}_2,1}$, it follows that $Y_1^{(2)} \perp\!\!\!\perp S_{\mathcal{D}_1,1}$. Furthermore, since $\mathcal{S}_{\mathcal{D}_1,2} \perp\!\!\!\perp \mathcal{S}_{\mathcal{D}_2,1}$ then $Y_1^{(2)} \perp\!\!\!\perp \mathcal{S}_{\mathcal{D}_1,2}$. Hence, Equation (33) holds in the initial step. Now, in the next time step:

$$(B_2^{(1)}, B_2^{(2)}) = \mathcal{E}_2(X^2, Y_1^{(1)}, Y_1^{(2)}, S_{\mathcal{E},2}) \tag{38}$$

$$= \mathcal{E}_2(X^2, \mathcal{D}_1^{(1)}(B_1^{(1)}, S_{\mathcal{D}_1,1}), Y_1^{(2)}, S_{\mathcal{E},2}) \tag{39}$$

$$Y_2^{(1)} = \mathcal{D}_2^{(1)}(B_2^{(1)}, S_{\mathcal{D}_1,2}) \tag{40}$$

$$Y_2^{(2)} = \mathcal{D}_2^{(2)}(B_2^{(2)}, S_{\mathcal{D}_2,2}), \tag{41}$$

where we see that $Y_2^{(2)}$ depends on $S_{\mathcal{D}_1,1}$ only through $Y_1^{(1)}$. Thus:

$$Y_2^{(2)} - Y_1^{(1)} - S_{\mathcal{D}_1,1}. \tag{42}$$

By the same arguments as before, we have for the second time step $Y_2^{(2)} \perp\!\!\!\perp \mathcal{S}_{\mathcal{D}_1,2}$ and $Y_2^{(2)} \perp\!\!\!\perp \mathcal{S}_{\mathcal{D}_1,3}$. By the causality of the system components, it follows that $Y^{(2),k-1}$ only depend on $S_{\mathcal{D}_1}^{k-1}$ through $Y^{(1),k-1}$, and by the independence of the side information, $Y^{(2),k-1} \perp\!\!\!\perp \mathcal{S}_{\mathcal{D}_1,k}$; thus. we get Equation (33).

For Equation (34), since $S_{\mathcal{D}_1,1} \perp\!\!\!\perp S_{\mathcal{D}_2,1}$, then $Y_1^{(1)} \perp\!\!\!\perp S_{\mathcal{D}_2,1}$, and the Markov chain holds in the initial step. For the next step, since $Y_2^{(1)}$ depends on $S_{\mathcal{D}_2,1}$ only through $Y_1^{(2)}$, the Markov chain holds. Therefore, by the causality of the system components, $Y_k^{(1)}$ only depends on $S_{\mathcal{D}_2}^{k-1}$ through $Y^{(2),k-1}$, and because $S_{\mathcal{D}_1,k} \perp\!\!\!\perp S_{\mathcal{D}_2,k}$, it follows that $Y_k^{(1)} \perp\!\!\!\perp S_{\mathcal{D}_2,k}$. Therefore, Equation (34) holds. \square

We note that requiring the side information to be mutually independent is not a hard assumption. For example, it is straightforward to generate independent dither signals for two quantizers. A short perspective on removing this assumption is given in Section 6.

We define the mutual information rate between two random processes next.

Definition 6 (Mutual information rate ([33] Equation (7.3.9))). *The mutual information rate between two random processes $\{X_k\}$ and $\{Y_k\}$ is defined as:*

$$\bar{I}(X; Y) \triangleq \lim_{n \rightarrow \infty} \frac{1}{n} I(X^n; Y^n). \tag{43}$$

We are now ready to state our first main result.

Theorem 1 (Lower bound on sum-rate). *Consider a ZDMD source coding problem with feedback (Problem 1), as seen in Figure 2. If Assumption 1 holds, the decoders are invertible, and the decoder side information is mutually independent, then:*

$$R_1 + R_2 \geq \bar{I}(X \rightarrow Y^{(1)}, Y^{(2)}) + \bar{I}(Y^{(1)}; Y^{(2)}). \tag{44}$$

The proof of Theorem 1 can be found in Appendix A.

Theorem 1 shows that when imposing zero-delay constraints on MD coding with feedback, the directed information rate from the source to the central reconstruction together with the mutual information rate between the side reconstructions serve as a lower bound on the associated average data sum-rate, thus relating the operational ZDMD rates to the information-theoretic quantities of directed and mutual information rate.

To the best of the authors’ knowledge, Theorem 1 provides a novel characterization between the relationship of the operational sum-rate and directed and mutual information rates, for a ZDMD coding problem with feedback. This result extends on the novel single-description bound in [5] and the MD results of [11].

In relation to the El-Gamal and Cover region [11], our result shows that the first term in the bound on the ZDMD sum-rate, i.e., the no excess sum-rate, is given by the directed information rate from the source to the side descriptions—that is, only the causally conveyed information, as would be expected for ZD coding. The second term is similar to that of El-Gamal and Cover. That is, the excess rate must be spent on communicating the mutual information between the side descriptions to reduce the central distortion.

Remark 1. *The mutual information rate $\bar{I}(Y^{(1)}; Y^{(2)})$ does not imply a noncausal relationship between $Y^{(1)}$ and $Y^{(2)}$, i.e., that $Y^{(1)}$ might depend on future values of $Y^{(2)}$. It only implies probabilistic dependence across time [22]. There is feedback between $Y^{(1)}$ and $Y^{(2)}$, such that information flows between the two descriptions. However, the information flows in a causal manner, i.e., the past values of $Y^{(1)}$ affect the future values of $Y^{(2)}$ and vice versa. This is also apparent from the “delayed” information flow from $Y^{(2),n-1}$ to $Y^{(1),n}$ in the proof, see Equation (A7). Therefore, the MD code must convey this total information flow between the two descriptions to the central receiver.*

3.3. Gaussian Lower Bound For Scalar Gauss-Markov Sources

Before showing Gaussian reproductions minimize the result of Theorem 1, we introduce the following technical assumptions required for our proof.

Assumption 2 (Sequential greedy coding). *Consider the ZDMD coding problem in Figure 2. We say that we solve this problem using sequential greedy coding if sequentially for each time step $k \in \mathbb{N}$: We minimize the bit-rate such that the MSE distortion constraints $D_i > 0$, $i = 0, 1, 2$, are satisfied for each $k \in \mathbb{N}$.*

That is, sequentially for each $k \in \mathbb{N}$, choose the codewords $B_k^{(i)}$, $i = 1, 2$ with minimum codeword lengths $l_k^{(i)}$, $i = 1, 2$ such that:

$$E \left[\left(X_k - Y_k^{(0)} \right)^2 \right] \leq D_0 \tag{45}$$

$$E \left[\left(X_k - Y_k^{(i)} \right)^2 \right] \leq D_i, \quad i = 1, 2. \tag{46}$$

Since, in sequential greedy coding, we minimize the bit-rate for each $k \in \mathbb{N}$ in the sequential order subject to the distortion constraints, this implies for the information rates in Equation (57) that we minimize the sum:

$$\begin{aligned}
 I\left(X^n \rightarrow Y^{(1),n}, Y^{(2),n}\right) + I\left(Y^{(1),n}; Y^{(2),n}\right) &= \sum_{k=1}^n \left[I\left(X^k; Y_k^{(1)}, Y_k^{(2)} \mid Y^{(1),k-1}, Y^{(2),k-1}\right) \right. \\
 &\quad + I\left(Y_k^{(2)}; Y_k^{(1)} \mid Y^{(1),k-1}, Y^{(2),k-1}\right) \\
 &\quad + I\left(Y_k^{(1)}; Y^{(2),k-1} \mid Y^{(1),k-1}\right) \\
 &\quad \left. + I\left(Y_k^{(2)}; Y^{(1),k-1} \mid Y^{(2),k-1}\right) \right], \tag{47}
 \end{aligned}$$

by sequentially for each $k \in \mathbb{N}^n$ selecting the optimal test-channel distribution $P\left(y^{(0)}, y_k^{(1)}, y_k^{(2)} \mid y^{(1),k-1}, y^{(2),k-1}, y^{(0),k-1}, x^k\right)$ subject to the MSE distortion constraints:

$$\begin{aligned}
 \mathbb{E}\left[\left(X_k - Y_k^{(0)}\right)^2\right] &\leq D_0 \\
 \mathbb{E}\left[\left(X_k - Y_k^{(i)}\right)^2\right] &\leq D_i, \quad i = 1, 2,
 \end{aligned} \tag{48}$$

and fixing this distribution for all following $k' > k$.

Let $\tilde{Y}_1^{(i)}$, $i = 1, 2$ minimize the initial mutual informations for $k = 1$, i.e.:

$$I\left(X_1; Y_1^{(1)}, Y_1^{(2)}\right) + I\left(Y_1^{(2)}; Y_1^{(1)}\right) \geq I\left(X_1; \tilde{Y}_1^{(1)}, \tilde{Y}_1^{(2)}\right) + I\left(\tilde{Y}_1^{(2)}; \tilde{Y}_1^{(1)}\right) \tag{49}$$

with equality if $Y_1^{(i)}$, $i = 1, 2$, are distributed as $\tilde{Y}_1^{(i)}$, $i = 1, 2$. Then, sequential greedy coding implies $Y_1^{(i)}$, $i = 1, 2$ must be distributed as $\tilde{Y}_1^{(i)}$, $i = 1, 2$, for all $k > 1$. Particularly for $k = 2$:

$$\begin{aligned}
 &I\left(X^2; Y_2^{(1)}, Y_2^{(2)} \mid Y_1^{(1)}, Y_1^{(2)}\right) + I\left(Y_2^{(2)}; Y_2^{(1)} \mid Y_1^{(1)}, Y_1^{(2)}\right) + I\left(Y_2^{(1)}; Y_1^{(2)} \mid Y_1^{(1)}\right) + I\left(Y_2^{(2)}; Y_1^{(1)} \mid Y_1^{(2)}\right) \\
 &= I\left(X^2; Y_2^{(1)}, Y_2^{(2)} \mid \tilde{Y}_1^{(1)}, \tilde{Y}_1^{(2)}\right) + I\left(Y_2^{(2)}; Y_2^{(1)} \mid \tilde{Y}_1^{(1)}, \tilde{Y}_1^{(2)}\right) + I\left(Y_2^{(1)}; \tilde{Y}_1^{(2)} \mid \tilde{Y}_1^{(1)}\right) + I\left(Y_2^{(2)}; \tilde{Y}_1^{(1)} \mid \tilde{Y}_1^{(2)}\right), \tag{50}
 \end{aligned}$$

where $\tilde{Y}_1^{(i)}$, $i = 1, 2$ is inserted on both sides of the conditioning.

The sequential greedy assumption is suitable in a zero-delay source coding perspective, since we must send the optimum description that minimizes the rate while achieving the desired distortion at each time step. We comment on the implications of sequential greedy coding in Section 6.

We also need the following assumption on the minimum MSE (MMSE) predictors.

Assumption 3 (Conditional prediction residual independence). *Let $\{X_k\}_{k \in \mathbb{N}}$ be a stationary source process, and let $\{Y_k^{(1)}\}_{k \in \mathbb{N}}$ and $\{Y_k^{(2)}\}_{k \in \mathbb{N}}$ be stationary arbitrarily distributed reproduction processes. We say the MMSE reproduction processes have conditional prediction residual independence if the MMSE prediction residuals satisfy for all $k \in \mathbb{N}$:*

$$Y_k^{(i)} - \mathbb{E}\left[Y_k^{(i)} \mid Y^{(1),k-1}, Y^{(2),k-1}\right] \perp\!\!\!\perp \left(Y^{(1),k-1}, Y^{(2),k-1}\right), \quad i = 1, 2, \tag{51}$$

$$Y_k^{(i)} - \mathbb{E}\left[Y_k^{(i)} \mid Y^{(i),k-1}\right] \perp\!\!\!\perp Y^{(i),k-1}, \quad i = 1, 2, \tag{52}$$

$$Y_k^{(i)} - \mathbb{E}\left[Y_k^{(i)} \mid Y^{(j),k-1}\right] \perp\!\!\!\perp Y^{(j),k-1}, \quad i \neq j, i, j \in \{1, 2\}, \tag{53}$$

that is, the residuals are independent of the conditioning prediction variables.

For mutual information, the conditional prediction residual independence implies:

$$I\left(Y_k^{(1)} - E\left[Y_k^{(1)}|Y^{(1),k-1}\right]; Y_k^{(2)} - E\left[Y_k^{(2)}|Y^{(1),k-1}\right] | Y^{(1),k-1}\right) = I\left(Y_k^{(1)} - E\left[Y_k^{(1)}|Y^{(1),k-1}\right]; Y_k^{(2)} - E\left[Y_k^{(2)}|Y^{(1),k-1}\right]\right). \tag{54}$$

Particularly, if $\{Y_k^{(i)}\}$, $i = 1, 2$ are jointly Gaussian, then the MMSE predictors have conditional prediction residual independence by the orthogonality principle ([34] p. 45). Using these predictors may result in an increased rate, since we limit the amount of possible predictors. That is, by not imposing this condition, we may achieve a smaller distortion for the same rate by minimizing over all possible MMSE predictors.

We are now ready to state our second main result.

Theorem 2 (Gaussian lower bound). *Let $\{X_k\}_{k \in \mathbb{N}}$ be a stable stationary scalar Gaussian process (1) with nondegenerate MSE distortion constraints, $D_i > 0$, $i = 0, 1, 2$. Then, under the sequential greedy coding condition (Assumption 2), and if the reproduction sequences $\{Y_k^{(i)}\}$, $i = 1, 2$, satisfy conditional prediction residual independence (Assumption 3), the following inequality holds:*

$$\bar{I}\left(X \rightarrow Y^{(1)}, Y^{(2)}\right) + \bar{I}\left(Y^{(1)}; Y^{(2)}\right) \geq \bar{I}\left(X \rightarrow Y_G^{(1)}, Y_G^{(2)}\right) + \bar{I}\left(Y_G^{(1)}; Y_G^{(2)}\right), \tag{55}$$

where $Y_G^{(i)}$, $i = 1, 2$ are jointly Gaussian random variables with first and second moments equal to those of $Y^{(i)}$, $i = 1, 2$.

The proof of Theorem 2 can be found in Appendix B.

Theorem 2 shows that for stationary scalar Gaussian sources under sequential greedy coding and MSE distortion constraints, the mutual informations between the source and side reproductions, and the mutual information between the side reproductions are minimized by Gaussian reproductions. This would generally be expected, since this is the case for single description ZD source coding [8].

To the best of the authors' knowledge, this is a novel result that has not been documented in any publicly available literature. Similar results exist for single-description ZD source coding [8] and for classical MD coding of white Gaussian sources [35].

Remark 2. *The main difficulty in proving Theorem 2, and the reason for the technical assumptions, is to minimize the excess information rate, $\bar{I}(Y^{(1)}; Y^{(2)})$, in Equation (44) and show the reconstructions, $Y^{(1)}, Y^{(2)}$, should be jointly Gaussian when they are jointly Gaussian with the source. We speculate these technical assumptions may be disregarded, since by the results of [8], we have for a Gaussian source process $\{X_k\}$:*

$$I\left(X \rightarrow Y^{(1)}, Y^{(2)}\right) \geq I\left(X \rightarrow Y_G^{(1)}, Y_G^{(2)}\right), \tag{56}$$

with equality if $\{Y_k^{(1)}, Y_k^{(2)}\}$ are jointly Gaussian with $\{X_k\}$. Therefore, it seems reasonable $Y^{(1)}, Y^{(2)}$ should also be jointly Gaussian in the second term on the RHS of Equation (44). However, we have not been able to prove this.

Symmetric Case

Following the result of Theorem 1, we now formally define the information-theoretic symmetric Gaussian ZDMD RDF, $R_{\text{ZD}}^I(D_0, D_S)$, in terms of the directed and mutual information rate, as a lower bound to $R_{\text{ZD}}^{\text{OP}}(D_0, D_S)$. Furthermore, we show that Gaussian reproductions minimize the lower bound.

Definition 7 (Information-Theoretic Symmetric ZDMD RDF). *The information-theoretic symmetric ZDMD RDF, for the stationary Gaussian source process $\{X_k\}$, with MSE distortion constraints, $D_0, D_S > 0$, is:*

$$\begin{aligned}
 R_{ZD}^I(D_0, D_S) &\triangleq \inf \frac{1}{2} \bar{I}(X \rightarrow Y^{(1)}, Y^{(2)}) + \frac{1}{2} \bar{I}(Y^{(1)}; Y^{(2)}), \\
 \text{s.t.} \quad &\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \mathbb{E} \left[\left(X_k - Y_k^{(0)} \right)^2 \right] \leq D_0 \\
 &\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \mathbb{E} \left[\left(X_k - Y_k^{(i)} \right)^2 \right] \leq D_S, \quad i = 1, 2,
 \end{aligned} \tag{57}$$

where the infimum is of all processes $\{Y_k^{(i)}\}$, $i = 0, 1, 2$ that satisfy:

$$X_{k+1}^\infty - X^k - \left(Y^{(0),k}, Y^{(1),k}, Y^{(2),k} \right), \quad \forall k \in \mathbb{N}. \tag{58}$$

The minimization of all processes $\{Y_k^{(i)}\}$, $i = 0, 1, 2$ that satisfy the Markov chain in Equation (58) is equivalent to the minimization of all sequences of conditional test-channel distributions $\{P(y_k^{(0)}, y_k^{(1)}, y_k^{(2)} | y^{(0),k-1}, y^{(1),k-1}, y^{(2),k-1}, x^k) : k \in \mathbb{N}\}$.

For Gaussian reproductions, we have the following optimization problem.

Problem 3 (Gaussian Information-Theoretic Symmetric ZDMD RDF). *For a stationary Gaussian source $\{X_k\}$ with MSE distortion constraints, $D_S \geq D_0 > 0$, the Gaussian information-theoretic symmetric ZDMD RDF is:*

$$\begin{aligned}
 R_{ZD,GM}^I(D_0, D_S) &\triangleq \inf \frac{1}{2} \bar{I}(X \rightarrow Y^{(1)}, Y^{(2)}) + \frac{1}{2} \bar{I}(Y^{(1)}; Y^{(2)}), \\
 \text{s.t.} \quad &\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \mathbb{E} \left[\left(X_k - Y_k^{(0)} \right)^2 \right] \leq D_0 \\
 &\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \mathbb{E} \left[\left(X_k - Y_k^{(i)} \right)^2 \right] \leq D_S, \quad i = 1, 2,
 \end{aligned} \tag{59}$$

where the infimum is over all Gaussian processes $\{Y_k^{(i)}\}$, $i = 0, 1, 2$, that satisfy:

$$X_{k+1}^\infty - X^k - \left(Y^{(0),k}, Y^{(1),k}, Y^{(2),k} \right), \quad \forall k \in \mathbb{N}. \tag{60}$$

This minimization is equivalent to the minimization of all sequences of Gaussian conditional test-channel distributions

$$\left\{ P^{GP}(y_k^{(0)}, y_k^{(1)}, y_k^{(2)} | y^{(0),k-1}, y^{(1),k-1}, y^{(2),k-1}, x^k) : k \in \mathbb{N} \right\}.$$

Finally, by Theorems 1 and 2, we have the following corollary, showing Problem 3 as a lower bound to Problem 2.

Corollary 1. *Let $\{X_k\}_{k \in \mathbb{N}}$ be a stable stationary scalar Gaussian process (1), with MSE distortion constraints, $D_S \geq D_0 > 0$. Then, under the sequential greedy coding condition (Assumption 2), and if the reproduction sequences $\{Y_k^{(i)}\}$, $i = 1, 2$, satisfy conditional prediction residual independence (Assumption 3), the following inequalities hold:*

$$R_{ZD,GM}^I(D_0, D_S) \leq R_{ZD}^I(D_0, D_S) \leq R_{ZD}^{op}(D_0, D_S). \tag{61}$$

This shows Gaussian reproduction processes minimize the information-theoretic symmetric ZDMD RDF. With this information-theoretic lower bound on $R_{ZD}^{op}(D_0, D_S)$, we now derive an optimal test-channel realization scheme that achieves this lower bound.

4. Symmetric Test-Channel Realization

In this section, we introduce a feedback realization of the optimal test channel for the Gaussian information-theoretic symmetric ZDMD RDE, $R_{ZD,GM}^I(D_0, D_S)$. This test channel is based on the ZDMD coding problem with feedback in Figure 2 and the feedback realization scheme of [8]. Finally, we present a characterization of $R_{ZD,GM}^I(D_0, D_S)$ as the solution to an optimization problem. This provides an achievable lower bound to Problem 2 in a Gaussian coding scheme.

4.1. Predictive Coding

The feedback realization scheme for the optimum test channel is illustrated in Figure 4. For each side channel, we follow the feedback realization of ([8] Theorem 2). Hence, the reproduction sequence of the optimum test channel is realized by:

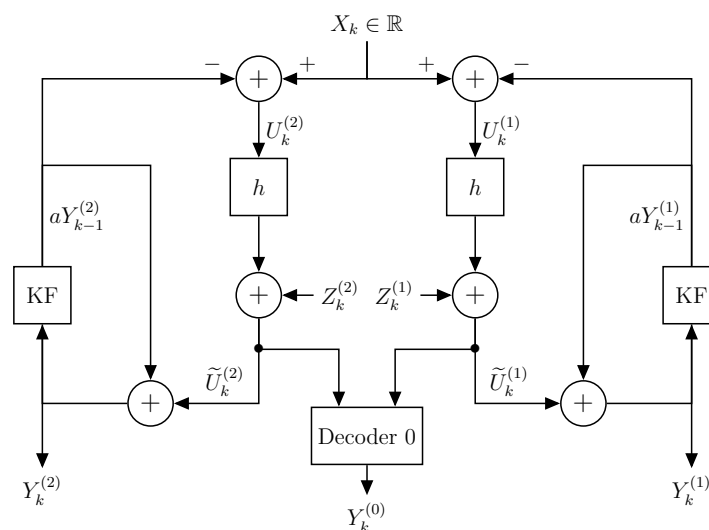
$$Y_k^{(i)} = hX_k + (1 - h)aY_{k-1}^{(i)} + Z_k^{(i)}, \tag{62}$$

where $Z_k^{(i)} \in \mathbb{R} \sim \mathcal{N}(0, \sigma_{Z_S}^2)$:

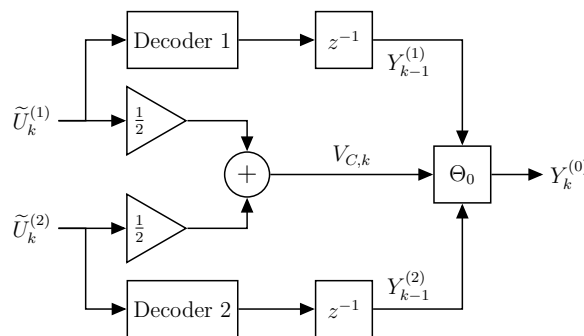
$$h \triangleq 1 - \pi_S \lambda^{-1}, \tag{63}$$

$$\sigma_{Z_S}^2 \triangleq \pi_S h, \tag{64}$$

$$\lambda = a^2 \pi_S + \sigma_W^2. \tag{65}$$



(a) Overall test channel with side decoders



(b) Test-channel central decoder

Figure 4. Feedback realization of the optimum test channel for $R_{ZD,GM}^I(D_0, D_S)$.

Here, λ is the variance of the side error processes:

$$\begin{aligned} U_k^{(i)} &\triangleq X_k - \mathbb{E} \left[X_k | Y^{(i),k-1} \right], \\ &= X_k - aY_{k-1}^{(i)}, \quad i = 1, 2. \end{aligned} \tag{66}$$

Furthermore, π_S , is the MSE for the estimation of X_k and $U_k^{(i)}$, i.e.,:

$$\pi_S \triangleq \mathbb{E} \left[\left(X_k - Y_k^{(i)} \right)^2 \right] = \mathbb{E} \left[\left(U_k^{(i)} - \tilde{U}_k^{(i)} \right)^2 \right], \quad i = 1, 2, \tag{67}$$

where $\tilde{U}_k^{(i)}$ are the innovation processes:

$$\tilde{U}_k^{(i)} \triangleq Y_k^{(i)} - \mathbb{E} \left[Y_k^{(i)} | Y^{(i),k-1} \right] \tag{68}$$

$$= hU_k^{(i)} + Z_k^{(i)}, \quad i = 1, 2, \tag{69}$$

with variance:

$$\sigma_{\tilde{U}}^2 = h^2\lambda + \pi_S h. \tag{70}$$

The innovation process, $\tilde{U}_k^{(i)}$ $i = 1, 2$, can be viewed as the i th side decoder estimate of $U_k^{(i)}$. Finally, we have that:

$$\begin{aligned} Z_k^{(1)} &\perp\!\!\!\perp Z_l^{(2)} \quad \forall k \neq l \\ Z_k^{(i)} &\perp\!\!\!\perp Z_l^{(i)} \quad \forall k \neq l, \quad i = 1, 2 \\ Z_k^{(i)} &\perp\!\!\!\perp U_l^{(j)} \quad \forall k \geq l \quad i, j \in \{1, 2\}, \end{aligned}$$

and the joint test-channel noise distribution is:

$$\begin{bmatrix} Z_k^{(1)} \\ Z_k^{(2)} \end{bmatrix} \sim \mathcal{N} (0, \Sigma_Z), \tag{71}$$

where:

$$\Sigma_Z = \begin{bmatrix} \pi_S h & \rho \pi_S h \\ \rho \pi_S h & \pi_S h \end{bmatrix}. \tag{72}$$

We note that the test channel in Figure 4 differs from the usual MD double-branch test channel of Ozarow [10], since the encoder does not create the two descriptions by adding correlated noises directly to the source, i.e., to the *same* input. Instead, the test channel consists of two branches, each consisting of a differential pulse code modulation (DPCM) scheme, where the correlated noises are added to the two already correlated closed-loop prediction error signals.

We also note the clear resemblance between the ZDMD coding problem in Figure 2 and the test channel in Figure 4a. This shows how the general ZDMD coding problem and its lower bound provide a constructive result that is conveniently extended to an optimum test-channel realization.

4.2. Central Decoder Design

The ZDMD encoder creates the two descriptions by prescaling and adding correlated noises to the two prediction error processes, $U_k^{(1)}, U_k^{(2)}$, resulting in the two innovation processes, $\tilde{U}_k^{(1)}, \tilde{U}_k^{(2)}$, as the side decoder estimates of $U_k^{(1)}, U_k^{(2)}$. For each time step k , the central decoder takes the two innovation

processes, $\tilde{U}_k^{(i)}$, $i = 1, 2$ as input. Since the additive noises are correlated, the central decoder can provide better estimates of $U_k^{(1)}$, $U_k^{(2)}$ than either of the side decoders. Using the central decoder estimates of $U_k^{(1)}$, $U_k^{(2)}$, we can provide a better estimate of the source X_k than either side decoder. We average the side innovations processes and define the central innovations description:

$$V_{C,k} \triangleq \frac{1}{2} (\tilde{U}_k^{(1)} + \tilde{U}_k^{(2)}). \tag{73}$$

Before we discuss the central decoder design, the following lemma provides a useful list of covariances between the signals in the feedback coding scheme of Figure 4, which can be readily verified [27].

Lemma 2 (Covariances). *Let $\{X_k\}$ be a stable stationary scalar Gauss-Markov process as in Equation (1) with stationary variance $\text{Var}[X_k] = \sigma_X^2$. Using the feedback coding scheme of Figure 4, then the following covariances hold:*

$$\Sigma_{XY} \triangleq \text{Cov}[X_k, Y_k^{(i)}] = \frac{h}{1 - a^2(1 - h)} \sigma_X^2, \quad i = 1, 2, \tag{74}$$

$$\Sigma_{XV_C} \triangleq \text{Cov}[X_k, V_{C,k}] = h (\sigma_X^2 - a^2 \Sigma_{XY}), \tag{75}$$

$$\sigma_Y^2 \triangleq \text{Var}[Y_k^{(i)}] = \frac{h^2 \sigma_X^2 + 2a^2 h(1 - h) \Sigma_{XY} + \sigma_{Z_s}^2}{1 - a^2(1 - h)^2}, \quad i = 1, 2, \tag{76}$$

$$\Sigma_{Y^{(1)}Y^{(2)}} \triangleq \text{Cov}[Y_k^{(1)}, Y_k^{(2)}] = \frac{h^2 \sigma_X^2 + 2a^2 h(1 - h) \Sigma_{XY} + \Sigma_{Z^{(1)}Z^{(2)}}}{1 - a^2(1 - h)^2} \tag{77}$$

$$\Sigma_{U^{(1)}U^{(2)}} \triangleq \text{Cov}[U_k^{(1)}, U_k^{(2)}] = \sigma_X^2 + a^2 (\Sigma_{Y^{(1)}Y^{(2)}} - 2\Sigma_{XY}) \tag{78}$$

$$\Sigma_{UV_C} \triangleq \text{Cov}[U_k^{(i)}, V_{C,k}] = \frac{1}{2} h (\lambda + \Sigma_{U^{(1)}U^{(2)}}), \quad i = 1, 2 \tag{79}$$

$$\sigma_{V_C}^2 \triangleq \text{Var}[V_{C,k}] = \frac{1}{2} (\sigma_U^2 + h^2 \Sigma_{U^{(1)}U^{(2)}} + \Sigma_{Z^{(1)}Z^{(2)}}). \tag{80}$$

The central decoder design is illustrated in Figure 4b. For each time step k , the central decoder takes the two innovation processes, $\tilde{U}_k^{(i)}$, $i = 1, 2$ as input. These are averaged to create the central description $V_{C,k}$. In the previous time step, local side decoders produced the side reconstructions, $Y_{k-1}^{(i)}$, $i = 1, 2$, such that the central decoder has $Y_{k-1}^{(i)}$, $i = 1, 2$ available when producing the central estimate, $Y_k^{(0)}$.

Let $\Omega_k = [V_{C,k}, Y_{k-1}^{(1)}, Y_{k-1}^{(2)}]^T$, then the central MMSE estimate of X_k is:

$$Y_k^{(0)} = \text{E}[X_k | \Omega_k] = \Theta_0 \Omega_k, \tag{81}$$

where $\Theta_0 \in \mathbb{R}^{1 \times 3}$ is given as:

$$\Theta_0 \triangleq \Sigma_{X\Omega} \Sigma_{\Omega}^{-1}, \tag{82}$$

with:

$$\Sigma_{X\Omega} \triangleq \text{E}[X_k \Omega_k^T] = \begin{bmatrix} \Sigma_{XV_C} & a \Sigma_{XY} & a \Sigma_{XY} \end{bmatrix}, \tag{83}$$

$$\Sigma_{\Omega} \triangleq \text{E}[\Omega_k \Omega_k^T] = \begin{bmatrix} \sigma_{V_C}^2 & \frac{1}{2} h a (\Sigma_{XY} - \Sigma_{Y^{(1)}Y^{(2)}}) & \frac{1}{2} h a (\Sigma_{XY} - \Sigma_{Y^{(1)}Y^{(2)}}) \\ \frac{1}{2} h a (\Sigma_{XY} - \Sigma_{Y^{(1)}Y^{(2)}}) & \sigma_Y^2 & \Sigma_{Y^{(1)}Y^{(2)}} \\ \frac{1}{2} h a (\Sigma_{XY} - \Sigma_{Y^{(1)}Y^{(2)}}) & \Sigma_{Y^{(1)}Y^{(2)}} & \sigma_Y^2 \end{bmatrix}. \tag{84}$$

The central distortion is then:

$$\pi_0 \triangleq \mathbb{E} \left[\left(X_k - Y_k^{(0)} \right)^2 \right] = \sigma_X^2 - \Sigma_{X\Omega} \Sigma_{\Omega}^{-1} \Sigma_{X\Omega}^T. \tag{85}$$

4.3. Rates

We now determine the achievable sum-rate for the test channel.

Initially, for each time step k , we express the mutual information in the definition of $R_{ZD,GM}^I(D_0, D_S)$ in Equation (59) using the differential entropy ([28] Ch. 8):

$$\begin{aligned} & I \left(X^k; Y_k^{(1)}, Y_k^{(2)} | Y^{(1),k-1}, Y^{(2),k-1} \right) + I \left(Y_k^{(2)}; Y^{(1),k} | Y^{(2),k-1} \right) + I \left(Y_k^{(1)}; Y^{(2),k-1} | Y^{(1),k-1} \right) \\ &= h \left(Y_k^{(2)} | Y^{(2),k-1} \right) + h \left(Y_k^{(1)} | Y^{(1),k-1} \right) - h \left(Y_k^{(1)}, Y_k^{(2)} | Y^{(1),k-1}, Y^{(2),k-1}, X^k \right). \end{aligned} \tag{86}$$

Comparing the test channel of Figure 4 to the general ZDMD source coding scenario with feedback in Figure 2, we have:

$$h \left(Y_k^{(i)} | Y^{(i),k-1} \right) = h \left(\tilde{U}_k^{(i)} \right) = \frac{1}{2} \log (2\pi e \lambda h) \tag{87}$$

and:

$$h \left(Y_k^{(1)}, Y_k^{(2)} | Y^{(1),k-1}, Y^{(2),k-1}, X^k \right) = h \left(Z_k^{(1)}, Z_k^{(2)} \right) = \frac{1}{2} \log (2\pi e |\Sigma_Z|) \tag{88}$$

$$= \frac{1}{2} \log \left(2\pi e \left(\pi_S^2 h^2 (1 - \rho^2) \right) \right). \tag{89}$$

Thus, the achievable symmetric sum-rate is:

$$R_1 + R_2 = \frac{1}{2} \log (2\pi e \lambda h) + \frac{1}{2} \log (2\pi e \lambda h) - \frac{1}{2} \log \left(2\pi e \left(\pi_S^2 h^2 (1 - \rho^2) \right) \right) \tag{90}$$

$$= \log \frac{\lambda}{\pi_S} - \frac{1}{2} \log (1 - \rho^2). \tag{91}$$

4.4. Scalar Lower Bound Theorem

Summarizing the above derivations, we present the following characterization of the Gaussian information theoretic symmetric ZDMD RDF.

Theorem 3 (Characterization of $R_{ZD,GM}^I(D_0, D_S)$). *Consider the stationary scalar Gauss-Markov process of (1). Given nondegenerate MSE distortion constraints, (D_S, D_0) , where $0 < D_0 \leq D_S \leq \sigma_X^2$, the Gaussian information-theoretic symmetric ZDMD RDF, $R_{ZD,GM}^I(D_0, D_S)$ is characterized by the solution to the following optimization problem.*

$$\begin{aligned} & \underset{\{\pi_S, \rho_0\}}{\text{minimize}} && \frac{1}{2} \log \frac{\lambda}{\pi_S} - \frac{1}{4} \log (1 - \rho_0^2) \\ & \text{subject to} && -1 \leq \rho_0 \leq 0 \\ & && 0 \leq \pi_S \leq \lambda \\ & && 0 \leq \pi_i \leq D_i, \quad i = 0, S, \end{aligned} \tag{92}$$

where:

$$\lambda = a^2 \pi_S + \sigma_W^2, \tag{93}$$

$$\pi_0 = \sigma_X^2 - \Sigma_{X\Omega} \Sigma_{\Omega}^{-1} \Sigma_{X\Omega}^T, \tag{94}$$

and $\Sigma_{X\Omega}, \Sigma_{\Omega}$ are defined in Equations (83) and (84).

Remark 3 (Uniqueness of optimal solution). *We believe that the optimal solution to Equation (92) is unique. Firstly, the objective function in Equation (92) can be shown to be convex in π_S and ρ_0 . Furthermore, the slope of the objective is negative for all $\pi_S > 0$ and $-1 < \rho_0 \leq 0$. Thus, it decreases monotonically towards a minimum. Additionally, for nondegenerate distortions, there should be equality in the distortions bounds, and since every ρ_0 indicates a certain trade-off point on the dominant face of the rate-distortion region, the minimum should be unique for every fixed ρ_0 . Hence, we conjecture the minimum to be unique. However, we have not yet been able to finally prove the uniqueness of the optimal solution to Equation (92).*

This completes the theoretical work on the lower bound to Problem 2, as the solution to Equation (92). Thus, for stationary scalar Gaussian sources in a Gaussian coding scheme, i.e., a source code that achieves the correctly distributed Gaussian noise, we have determined an achievable lower bound to $R_{ZD}^{op}(D_0, D_S)$, characterized by the (unique) solution to an optimization problem.

We now compare this theoretical lower bound to an operational achievable performance.

5. Simulation Study

In this section, we perform two simulation studies to validate our theoretical framework in Section 4 in relation to an operational quantization scheme.

5.1. Simple Quantization Scheme

In general, test channels provide a basis for the design of practical coding schemes by replacing the additive test-channel noises with quantizers producing quantization noise distributed similar to the test-channel noises. However, it is a nontrivial task to produce quantization noise with high negative correlation in practice [36]. There are some schemes that are able to achieve correlation that tends towards -1 [36], e.g., [37–40]. These schemes and many other MD coding schemes in general produce two descriptions with the desired correlation by direct operations on the source signal. However, our ZDMD test channel forms two descriptions from two correlated signals. Therefore, many existing schemes are not directly applicable to our test channel. This is somewhat expected since ZDMD coding is mostly an unexplored problem until now. Fortunately, the scheme of [41] illustrated in Figure 5 aligns well with our test channel, since it performs staggered quantization of two prediction error processes and uses a refinement layer for further central distortion gain. The main idea is to use two DPCM encoders with staggered quantizers, Q_1 and Q_2 , in a base layer and a third second-stage refinement quantizer Q_0 . For a detailed explanation of the derivation and design of this scheme, we refer to [27,41].

5.2. Experiments

In all simulations, we consider stationary scalar Gauss-Markov sources of the Form (1). All simulations are conducted by fixing the average rate per description, R , given as:

$$R = R_S + \frac{R_0}{2}, \tag{95}$$

where R_S is the rate of the first-stage quantizers, Q_1, Q_2 , and R_0 is the rate of the second-stage (central) quantizer, Q_0 . Then, for each rate-pair, R_S, R_0 , satisfying the rate constraint R , the practical quantizer step sizes are determined according to the high-rate approximations:

$$R_S = H(U^{\Delta_S(i)}) \approx h(U^{(i)}) - \log \Delta_S, \tag{96}$$

$$R_0 = H(E_C^{\Delta_0}) \approx h(E_C) - \log \Delta_0, \tag{97}$$

where $U^{\Delta_S(i)}$ is the quantized version of $U^{(i)}$, $E_C^{\Delta_0}$ is the quantized version of E_C , and the approximations follow from ([28] Theo. 8.3.1). The step sizes are determined such that the operational

rate per description, R_{op} , is approximately equal to the constraint, i.e., $R_{op} \approx R$. Further details on choosing the step size is found in [27]. From simulations we have seen, there is an approximate rate-loss of 0.1 bits/sample/description due to the approximation of step sizes in Equations (96) and (97). We have accounted for this when choosing the step sizes, such that R_{op} approximates R with greater accuracy. For lower rates, this difference is higher; hence, we consider only the high-rate scenario.

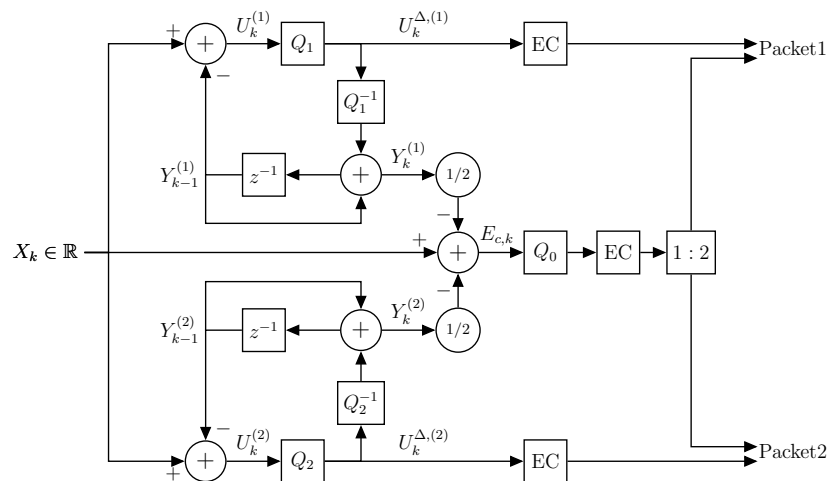


Figure 5. The two-stage staggered differential pulse code modulation (DPCM) quantization scheme. The two first-stage quantizers Q_1 and Q_2 are staggered identical uniform quantizers. Here, EC denotes lossless (entropy) encoders. The binary description packets are formed by entropy coding each side quantizer output and splitting the entropy coded second stage quantizer output in two.

We consider N source samples that are independently coded and decoded by the operational quantization scheme, and M Monte Carlo simulations for each rate-pair R_0, R_S . The numerical distortions are obtained by:

$$\hat{D}_i = \frac{1}{N} \sum_{i=k}^N (X_k - Y_k^{(i)})^2, \quad i = 0, 1, 2, \tag{98}$$

$$\hat{D}_S = \frac{\hat{D}_1 + \hat{D}_2}{2}, \tag{99}$$

where $Y_k^{(i)}$ $i = 0, 1, 2$ are the reconstructions for the k th input sample X_k . The operational coding rates are determined by the discrete entropies:

$$\hat{R}_i = H \left(\left\{ U_k^{\Delta_{S,i}} \right\}_{k=1}^N \right), \quad i = 1, 2, \tag{100}$$

$$\hat{R}_S = \frac{\hat{R}_1 + \hat{R}_2}{2}, \tag{101}$$

$$\hat{R}_0 = H \left(\left\{ E_{C,k}^{\Delta_0} \right\}_{k=1}^N \right), \tag{102}$$

where the entropies are determined from the empirical probabilities, which are obtained based on the histograms of $\{U_k^{\Delta_{S,i}}\}_{k=1}^N$, $i = 1, 2$ and $\{E_{C,k}^{\Delta_0}\}_{k=1}^N$.

The theoretical distortion limits for a given rate R are determined by fixing the objective function value in Equation (92), and determining the corresponding ρ_0 and central distortion π_0 for a grid of side distortions, π_S .

5.2.1. Distortion Trade-Off at Fixed Rate

We consider the trade-off between the side and central distortions, D_S, D_0 for a fixed rate per description, $R = 5$ bits/sample. We compare the theoretical lower bound on the distortions to the operational distortions obtained using the practical quantization scheme. The source and simulation parameters are listed in Table 1.

Table 1. Simulation Parameters for distortion trade-off curve in Figure 6.

Source Parameters	Symbol	Values
Source correlation coefficient	a	0.9
Source innovation variance	σ_W^2	1
Initial value variance	$\sigma_{X_1}^2$	$\frac{1}{1-0.9^2}$
Simulation parameters	Symbol	Values
Rate per description	R	5 bits/sample
Time samples	N	500,000
Monte-Carlo simulations	M	4

The resulting theoretical and operational distortion curves are shown in Figure 6. The figure shows the theoretical lower bound (black curve) on the achievable distortion region, and the operational achievable distortion pairs (dashed blue curve), for the fixed rate per description $R = 5$ bits/sample. The operational curve lies approximately 5 dB above the theoretical lower bound. Both curves show that if we decrease the central distortion, we must increase the side distortion, and vice-versa, if we want to maintain the same rate R . Hence, we are able to trade off between the side- and central distortion by varying the bit allocation in the first- and second stage quantizers.

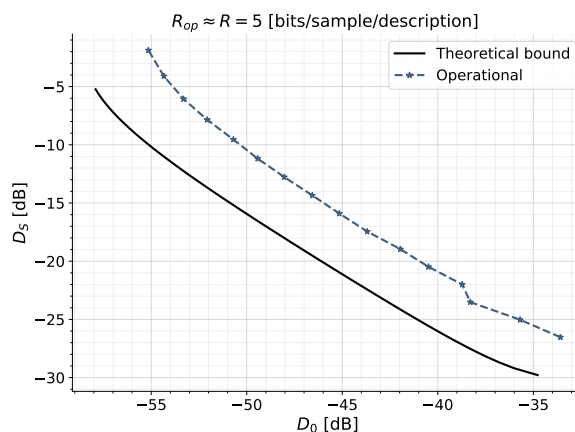


Figure 6. The central distortion, D_0 , versus side distortion, D_S for ZDMD coding of a Gauss-Markov source (1) with $a = 0.9$ and unit variance at $R = 5$ bits/sample/description. Simulation parameters in Table 1.

The 5 dB distortion loss corresponds to a total rate loss of approximately 0.83 bits/sample, for the sum-rate, or equivalently 0.415 bits/sample/description. Some of this loss can be attributed to the space-filling loss of the uniform quantizers, which is approximately 1.5 dB, or 0.254 bits/sample per quantizer. Thus, the refinement scheme suffers from the space-filling loss of three quantizers [42]. Furthermore, there is a loss due to the non-optimal linear predictors; however, this loss is minimal in the high-rate scenario [41].

The sudden bend in the operational curve can be attributed to a possible alphabet change, i.e., for certain rates and, hence, quantization bin sizes, the quantized signals have an increased alphabet size, due to smaller bin sizes.

5.2.2. Distortion versus Distortion-Ratio for Multiple Fixed Rates

We next consider how the side- and central distortions, D_0, D_S , vary with the distortion ratio $\gamma \triangleq D_0/D_S$ for different fixed rates R . Using the previously described procedure for the fixed rates $R \in \{4, 5, 6\}$ bits/sample/description, we obtained the distortion curves in Figure 7; the simulation parameters are listed in Table 2. Figure 7a shows the side distortion, D_S , in relation to the distortion ratio, γ , for varying rates. Similarly, Figure 7b shows the central distortion, D_0 , in relation to the distortion ratio, γ , for the same rates. In both figures, dashed curves indicate operational distortions and ratios, and solid curves indicate theoretical bounds.

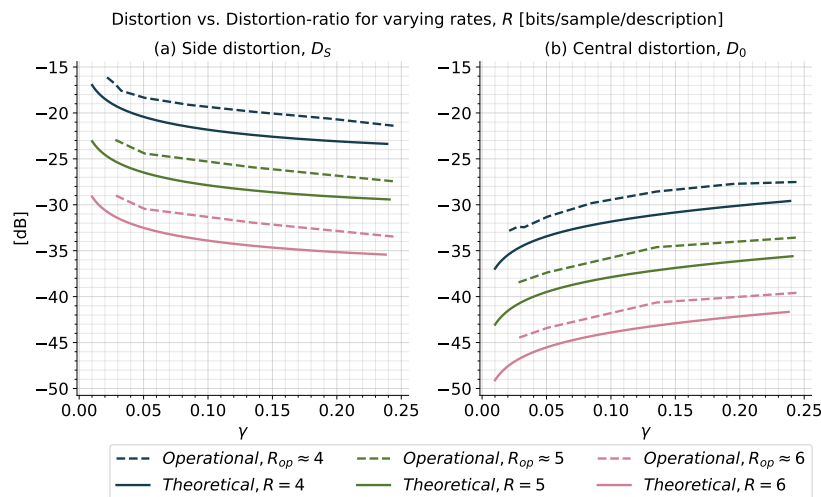


Figure 7. (a) Side distortion, D_S , and (b) central distortion, D_0 , versus distortion ratio $\gamma = D_0/D_S$ for ZDMD coding of a Gauss-Markov source (1) with $a = 0.9$ and unit variance at $R \in \{4, 5, 6\}$ bits/sample/description. Simulation parameters in Table 2.

Table 2. Simulation parameters for distortion versus distortion ratio curves in Figure 7.

Source Parameters	Symbol	Values
Source correlation coefficient	a	0.9
Source innovation variance	σ_W^2	1
Initial value variance	$\sigma_{X_1}^2$	$\frac{1}{1-0.9^2}$
Simulation parameters	Symbol	Values
Rate per description	R	{4, 5, 6} bits/sample
Time samples	N	500,000
Monte-Carlo simulations	M	4

For any particular rate and distortion ratio in Figure 7, the central distortion, D_0 , is always lower than the side distortion, D_S . Further, as the rate per description increases, both distortions decrease for all distortion ratios. Lower ratios imply lower central distortion, D_0 , at the cost of a higher side distortion D_S . This was also seen in Figure 6. Figure 7 shows, this trend is independent of the rate. Furthermore, the plots in Figure 7 show that by increasing the rate per description for any fixed ratio, we can increase the performance in both central and side distortion.

It can be shown, at no excess marginal rate, i.e., when $R_0 = 0$, we have that $D_0/D_S \approx 1/4$ [27], and therefore, the maximum operational distortion ratio is limited to approximately 1/4. Hence, to evaluate higher distortion ratios, we would need to perform non-optimal central reconstructions or decrease the quantizer offsets away from the optimum half bin size.

For a given rate and distortion ratio, the operational curves in Figure 7 are all approximately 2.5 dB above the theoretical bounds, with a slightly better performance at higher rates. This loss can again be attributed to the space-filling loss and non-optimal predictors. We notice that this loss seems

to be half of that seen when plotting D_S versus D_0 in Figure 6. However, for a given ratio, there are two curves in Figure 7, one for each of D_S and D_0 . Thus, the total distortion loss at a give ratio is 5 dB. Therefore, the apparent splitting of the loss can be attributed to a 2.5 dB loss for each of D_S and D_0 at a given ratio, c.f. [27].

From the rate-distortion performances in Figures 6 and 7, we see for the high-rate scenario that the simple quantization scheme is able to achieve a performance close to the theoretical ZDMD lower bounds derived in the previous sections. Hence, we are able to operate along the theoretical bounds for ZDMD coding of stationary scalar Gaussian sources using simple techniques. Particularly, we are able to trade off both rates and distortions. The simulation results also provide an indication of an upper bound on the optimal operational performance limits of ZDMD coding of stationary scalar Gauss-Markov sources.

6. Discussion

We now discuss some important aspects of our derivations and simulation results. Particularly, we focus on the assumptions made in the information-theoretic lower bound derivation, and how the test channel generalizes to an operational quantization scheme. Finally, we consider extension of our results to vector Gauss-Markov sources.

6.1. Theoretical Lower Bound

In order to derive an information-theoretic lower bound on the symmetric ZDMD RDF for scalar stationary Gaussian sources in Theorem 2, we have made some technical assumptions.

The main assumption was the use of sequential greedy coding (Assumption 2). This implies that at each time step, we must encode a source sample such that the rates are minimized and the distortion constraints are achieved. However, this might lead to an increased rate, since we must achieve the desired distortion performance in each time step and not just in the asymptotic average. Hence, for some source samples, excess bits might have to be spent to ensure the distortion constraints are achieved. The reason for this technical assumption is its implication from an information-theoretic or probabilistic point of view. That is, the test-channel distribution of a particular reconstruction given the current and past inputs should remain unchanged once it has been selected. It seems plausible that sequential greedy coding provides the same ZDMD information rates as jointly selecting the optimal test-channel distribution over all time steps. Since, from a ZD perspective, all source samples must be encoded and transmitted immediately without delay, their respective reconstruction distributions are thus selected only once. However, this remains an open problem for future research.

To derive the information-theoretic lower bound on the sum-rate, we assume the decoder side information is mutually independent. This assumption ensures the side-decoder reproduction, $Y_k^{(1)}$, is independent of the side information belonging to the other decoder, $S_{D_2}^k$, when the previous reproductions, $Y^{(2),k-1}$ are given, and vice versa for reproduction $Y_k^{(2)}$. Therefore, if using dependent or common side information, the results of Section 3 warrant further investigation, although, for common side information, it seems reasonable that the bounds should remain widely unchanged. In [43], an achievable region is derived for MD coding without feedback and with common side information, in the classic distributed information-theoretic sense. The bounds of [43] are similar to those of El-Gamal and Cover [11] with an added dependency upon the unknown side information in the involved mutual informations. Hence, these results could provide a basis for extending the results of Section 3 to the case of unknown or dependent side information.

6.2. Difficulties with the Vector Case

We note that Problem 2 and our first main result of Theorem 1 also hold for stationary vector sources. Similarly, the definition of the (Gaussian) information-theoretic symmetric ZDMD RDF is easily extended to the vector case.

The main concern is that of extending Theorem 2 to the vector case, i.e., showing Gaussian reproductions minimize the information theoretic lower bound on the sum-rate for stationary Gaussian vector sources. In [35], the scalar result of Ozarow is extended to IID Gaussian vector processes and shows the natural Gaussian multiple description scheme is optimal in achieving the lower bound on the sum-rate for matrix covariance constraints. In [14], it is shown how the Gaussian description scheme is also optimal under MSE distortion constraints. In the sense of zero delay, the results of [8,17] show that for Gauss-Markov source processes, the jointly Gaussian reproduction process minimizes the information-theoretic lower bound. Therefore, based on these results, we conjecture the result of Theorem 2 may be extended to Gaussian vector processes. To this end, we note that the proof of Theorem 2 relies on the tightness of Ozarow's lower bound for stationary scalar Gaussian processes. This reliance on scalar sources may be disregarded if it can be shown that:

$$\bar{I}(Y^{(1)}; Y^{(2)}) \geq \bar{I}(Y_G^{(1)}; Y_G^{(2)}), \quad (103)$$

with equality if $Y^{(1)}, Y^{(2)}$ are jointly Gaussian. For some initial results in this regard, see the extended proof of Theorem 2 in ([27] App. E).

If Theorem 2 can be extended to the vector case, it remains to derive an optimum test-channel realization scheme. Early work by the authors indicates that the test channel in Section 4 may be generalized to the vector case in a similar manner to that of [8]. In the stationary case, the covariances in Lemma 2 may be extended to the vector case in the form of Riccati matrix equations, where explicit solutions may be obtained using the techniques of ([44] Section 5). However, the main difficulty is that of determining the proper correlation between Gaussian test-channel noise vectors. Particularly, due to the structure of the noise covariance matrix, it is difficult to derive expressions for the determinant of Σ_Z such that a more explicit, possibly using semidefinite programming, expression may be formulated for $R_{ZD,GM}^I(D_0, D_S)$. For spatially uncorrelated vector sources, the extension is fairly straightforward, since it can be reasonably assumed that the noise cross-covariance matrix $\Sigma_{Z^{(1)}Z^{(2)}}$ should be diagonal along with Σ_{Z_S} . Since the dimensions are independent, the scalar solution can be applied to each dimension, and the total rates and distortions are given as sums across the scalar solutions for each dimension.

7. Conclusions

In this work, we studied the ZDMD source coding problem where the MD encoder and decoders are required to be causal and of zero delay. Furthermore, the encoder receives perfect decoder feedback, and side information is available to both encoder and decoders. Using this constructive system, we showed that the average data sum-rate is lower bounded by the sum of the directed information rate from the source, X , to the side descriptions, $Y^{(1)}, Y^{(2)}$, and the mutual information rate between the side descriptions, thus providing a novel relation between information theory and the operational ZDMD coding rates.

For scalar stationary Gaussian sources with MSE distortion constraints subject to the technical assumptions of sequential greedy coding and conditional residual independence, we showed this information-theoretic lower bound is minimized by Gaussian reproductions, i.e., the optimum test-channel distributions are Gaussian. This bound provides an information-theoretic lower bound to the operational symmetric ZDMD RDF, $R_{ZD}^{\text{op}}(D_0, D_S)$.

We showed the optimum test channel of the Gaussian information-theoretic lower bound is determined by a feedback realization scheme utilizing predictive coding and correlated Gaussian noises. This shows that the information-theoretic lower bound for first-order stationary scalar Gauss-Markov sources is achievable in a Gaussian coding scheme. Additionally, the optimum Gaussian test-channel distribution is characterized by the solution to an optimization problem.

We have not yet been able to extend the test channel into an operational quantization scheme that allows for an exact upper bound on the optimum operational performance limits.

Operational achievable results are determined for the high-rate scenario by utilizing the simple quantization scheme of [41], resembling our test channel to some extent. Using this simple quantization scheme, it is possible to achieve operational rates within 0.415 bits/sample/description of the theoretical lower bounds.

Author Contributions: Conceptualization, J.Ø. and A.J.F.; methodology, A.J.F. and J.Ø.; software, A.J.F.; validation, A.J.F. and J.Ø.; formal analysis, A.J.F.; investigation, A.J.F.; resources, J.Ø.; data curation, A.J.F.; writing—original draft preparation, A.J.F.; writing—review and editing, J.Ø. and A.J.F.; visualization, A.J.F.; supervision, J.Ø.; project administration, J.Ø.; funding acquisition, J.Ø.

Funding: This research was partially supported by VILLUM FONDEN Young Investigator Programme, Project No. 10095.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

DPCM	Differential Pulse Code Modulation
DPI	Data Processing Inequality
IID	Independent and Identically Distributed
MD	Multiple Description
MMSE	Minimum Mean-Squared Error
MSE	Mean-Squared Error
RDF	Rate-Distortion Function
ZD	Zero Delay
ZDMD	Zero-Delay Multiple Description

Appendix A. Proof of Theorem 1

First, since the expected length of a uniquely decodable code is lower bounded by its entropy ([28] Ch. 5), we have that:

$$\mathbb{E} \left[I_k^{(i)} \right] \geq H \left(B_k^{(i)} | B^{(i),k-1}, S_{\mathcal{D}_i}^k \right), \quad i = 1, 2, \quad (\text{A1})$$

since $B^{(i),k-1}$ and $S_{\mathcal{D}_i}^k$ are already available at decoder i before the reception of $B_k^{(i)}$. Thus:

$$\begin{aligned} \mathbb{E} \left[I_k^{(1)} \right] + \mathbb{E} \left[I_k^{(2)} \right] &\geq H \left(B_k^{(1)} | B^{(1),k-1}, S_{\mathcal{D}_1}^k \right) + H \left(B_k^{(2)} | B^{(2),k-1}, S_{\mathcal{D}_2}^k \right) \\ &\stackrel{\text{(a)}}{\geq} H \left(B_k^{(1)} | B^{(1),k-1}, S_{\mathcal{D}_1}^k \right) + H \left(B_k^{(2)} | B^{(2),k-1}, S_{\mathcal{D}_2}^k \right) \\ &\quad - H \left(B_k^{(1)}, B_k^{(2)} | B^{(1),k-1}, B^{(2),k-1}, X^k, S_{\mathcal{D}_1}^k, S_{\mathcal{D}_2}^k \right) \\ &\stackrel{\text{(b)}}{=} H \left(B_k^{(1)} | B^{(1),k-1}, S_{\mathcal{D}_1}^k \right) - H \left(B_k^{(1)}, B_k^{(2)} | B^{(1),k-1}, B^{(2),k-1}, S_{\mathcal{D}_1}^k, S_{\mathcal{D}_2}^k \right) \\ &\quad + H \left(B_k^{(2)} | B^{(2),k-1}, S_{\mathcal{D}_2}^k \right) + I \left(X^k; B_k^{(1)}, B_k^{(2)} | B^{(1),k-1}, B^{(2),k-1}, S_{\mathcal{D}_1}^k, S_{\mathcal{D}_2}^k \right) \\ &\stackrel{\text{(c)}}{=} H \left(B_k^{(1)} | B^{(1),k-1}, S_{\mathcal{D}_1}^k \right) + H \left(B_k^{(2)} | B^{(2),k-1}, S_{\mathcal{D}_2}^k \right) - H \left(B_k^{(1)} | B^{(1),k-1}, B^{(2),k-1}, S_{\mathcal{D}_1}^k, S_{\mathcal{D}_2}^k \right) \\ &\quad - H \left(B_k^{(2)} | B^{(1),k}, B^{(2),k-1}, S_{\mathcal{D}_1}^k, S_{\mathcal{D}_2}^k \right) + I \left(X^k; B_k^{(1)}, B_k^{(2)} | B^{(1),k-1}, B^{(2),k-1}, S_{\mathcal{D}_1}^k, S_{\mathcal{D}_2}^k \right) \\ &\stackrel{\text{(d)}}{=} I \left(X^k; B_k^{(1)}, B_k^{(2)} | B^{(1),k-1}, B^{(2),k-1}, S_{\mathcal{D}_1}^k, S_{\mathcal{D}_2}^k \right) + I \left(B_k^{(2)}; B^{(1),k}, S_{\mathcal{D}_1}^k | B^{(2),k-1}, S_{\mathcal{D}_2}^k \right) \\ &\quad + I \left(B_k^{(1)}; B^{(2),k-1}, S_{\mathcal{D}_2}^k | B^{(1),k-1}, S_{\mathcal{D}_1}^k \right), \quad (\text{A2}) \end{aligned}$$

where (a) follows from the non-negativity of discrete entropy ([28] Lem. 2.1.1). Step (b) follows from the definition of conditional mutual information ([28] p. 23), (c) by the chain rule for discrete

entropy ([28] Theo. 2.5.1), and (d) by the definition of conditional mutual information ([28] p. 23). Consider the first term of step (d) in Equation (A2):

$$\begin{aligned}
 I\left(X^k; B_k^{(1)}, B_k^{(2)} \mid B^{(1),k-1}, B^{(2),k-1}, S_{D_1}^k, S_{D_2}^k\right) &\stackrel{(e1)}{=} I\left(X^k; B_k^{(1)}, B_k^{(2)} \mid Y^{(1),k-1}, Y^{(2),k-1}, S_{D_1}^k, S_{D_2}^k\right) \\
 &\stackrel{(e2)}{\geq} I\left(X^k; Y_k^{(1)}, Y_k^{(2)} \mid Y^{(1),k-1}, Y^{(2),k-1}, S_{D_1}^k, S_{D_2}^k\right) \\
 &\stackrel{(e3)}{=} I\left(X^k; Y^{(1),k}, Y^{(2),k}, S_{D_1}^k, S_{D_2}^k\right) - I\left(X^k; Y^{(1),k-1}, Y^{(2),k-1}, S_{D_1}^k, S_{D_2}^k\right) \\
 &\stackrel{(e4)}{\geq} I\left(X^k; Y^{(1),k}, Y^{(2),k}\right) - I\left(X^k; Y^{(1),k-1}, Y^{(2),k-1}, S_{D_1}^k, S_{D_2}^k\right) \\
 &\stackrel{(e5)}{=} I\left(X^k; Y_k^{(1)}, Y_k^{(2)} \mid Y^{(1),k-1}, Y^{(2),k-1}\right) \\
 &\quad - I\left(X^k; S_{D_1}^k, S_{D_2}^k \mid Y^{(1),k-1}, Y^{(2),k-1}\right) \\
 &\stackrel{(e6)}{=} I\left(X^k; Y_k^{(1)}, Y_k^{(2)} \mid Y^{(1),k-1}, Y^{(2),k-1}\right), \tag{A3}
 \end{aligned}$$

where (e1) follows since the decoders are invertible given the side information, (e2) follows from the data processing inequality (DPI) ([28] Section 2.8), the invertible decoders, and Equation (30), (e3) by the chain rule of mutual information ([28] Theo. 2.5.2), (e4) since, by the non-negativity of mutual information ([28] Section 2.6), removing a variable can only decrease the mutual information, (e5) by the chain rule of mutual information, and (e6) since the side information is assumed to be independent of X .

For the second term of step (d) in Equation (A2):

$$\begin{aligned}
 I\left(B_k^{(2)}; B^{(1),k}, S_{D_1}^k \mid B^{(2),k-1}, S_{D_2}^k\right) &\stackrel{(f1)}{=} I\left(B_k^{(2)}; Y^{(1),k}, S_{D_1}^k \mid Y^{(2),k-1}, S_{D_2}^k\right) \\
 &\stackrel{(f2)}{\geq} I\left(Y_k^{(2)}; Y^{(1),k}, S_{D_1}^k \mid Y^{(2),k-1}, S_{D_2}^k\right) \\
 &\stackrel{(f3)}{\geq} I\left(Y_k^{(2)}; Y^{(1),k} \mid Y^{(2),k-1}, S_{D_2}^k\right) \\
 &\stackrel{(f4)}{=} I\left(Y_k^{(2)}, S_{D_2}^k; Y^{(1),k} \mid Y^{(2),k-1}\right) - I\left(S_{D_2}^k; Y^{(1),k} \mid Y^{(2),k-1}\right) \\
 &\stackrel{(f5)}{\geq} I\left(Y_k^{(2)}; Y^{(1),k} \mid Y^{(2),k-1}\right) - I\left(S_{D_2}^k; Y^{(1),k} \mid Y^{(2),k-1}\right) \\
 &\stackrel{(f6)}{=} I\left(Y_k^{(2)}; Y^{(1),k} \mid Y^{(2),k-1}\right), \tag{A4}
 \end{aligned}$$

where (f1) follows since the decoders are invertible, and (f2) from the DPI and Equation (32), (f3) since conditional mutual information is non-negative, removing a term on the left side of the conditioning can only decrease the mutual information, (f4) follows from the chain rule of mutual information, (f5) is similar to (f3), and finally, (f6) follows from Equation (34), and the mutual information is zero for independent variables.

For the third term of step (d) in Equation (A2), we have through similar derivations using the Markov chains in Equations (31) and (33):

$$I\left(B_k^{(1)}; B^{(2),k-1}, S_{D_2}^k \mid B^{(1),k-1}, S_{D_1}^k\right) \geq I\left(Y_k^{(1)}; Y^{(2),k-1} \mid Y^{(1),k-1}\right). \tag{A5}$$

Then, by Equations (A2)–(A5):

$$\mathbb{E}\left[l_k^{(1)}\right] + \mathbb{E}\left[l_k^{(2)}\right] \geq I\left(X^k; Y_k^{(1)}, Y_k^{(2)} \mid Y^{(1),k-1}, Y^{(2),k-1}\right) + I\left(Y_k^{(2)}; Y^{(1),k} \mid Y^{(2),k-1}\right) + I\left(Y_k^{(1)}; Y^{(2),k-1} \mid Y^{(1),k-1}\right). \tag{A6}$$

Summing over k , we have, by the definition of directed information (Definition 4):

$$\begin{aligned}
 \sum_{k=1}^n \left(\mathbb{E} \left[I_k^{(1)} \right] + \mathbb{E} \left[I_k^{(2)} \right] \right) &\geq I \left(X^n \rightarrow Y^{(1),n}, Y^{(2),n} \right) + I \left(Y^{(1),n} \rightarrow Y^{(2),n} \right) + I \left(Y^{(2),n-1} \rightarrow Y^{(1),n} \right) \\
 &= I \left(X^n \rightarrow Y^{(1),n}, Y^{(2),n} \right) + I \left(Y^{(1),n} \rightarrow Y^{(2),n} \right) + I \left(0 * Y^{(2),n-1} \rightarrow Y^{(1),n} \right) \\
 &= I \left(X^n \rightarrow Y^{(1),n}, Y^{(2),n} \right) + I \left(Y^{(1),n}; Y^{(2),n} \right), \tag{A7}
 \end{aligned}$$

where the last equality follows from the conservation of information ([45] Prop. 2) and $0 * Y^{(2),n-1}$ denotes the concatenation $0, Y_1^{(2)}, \dots, Y_{n-1}^{(2)}$. The lower bound (Equation (44)) now follows by dividing by n and taking the limit as $n \rightarrow \infty$. \square

Appendix B. Proof of Theorem 2

Recall:

$$\bar{I} \left(X \rightarrow Y^{(1)}, Y^{(2)} \right) + \bar{I} \left(Y^{(1)}; Y^{(2)} \right) = \lim_{n \rightarrow \infty} \frac{1}{n} I \left(X^n \rightarrow Y^{(1),n}, Y^{(2),n} \right) + \frac{1}{n} I \left(Y^{(1),n}; Y^{(2),n} \right) \tag{A8}$$

where:

$$I \left(X^n \rightarrow Y^{(1),n}, Y^{(2),n} \right) = \sum_{k=1}^n I \left(X^k; Y_k^{(1)}, Y_k^{(2)} | Y^{(1),k-1}, Y^{(2),k-1} \right), \tag{A9}$$

and:

$$I \left(Y^{(1),n}; Y^{(2),n} \right) = I \left(Y^{(1),n} \rightarrow Y^{(2),n} \right) + I \left(0 * Y^{(2),n-1} \rightarrow Y^{(1),n} \right) \tag{A10}$$

$$= I \left(Y^{(1),n} \rightarrow Y^{(2),n} \right) + I \left(Y^{(2),n-1} \rightarrow Y^{(1),n} \right) \tag{A11}$$

$$= \sum_{k=1}^n I \left(Y_k^{(2)}; Y^{(1),k} | Y^{(2),k-1} \right) + I \left(Y_k^{(1)}; Y^{(2),k-1} | Y^{(1),k-1} \right). \tag{A12}$$

For each time step $k \in \mathbb{N}$, using the chain rule of mutual information [28], we have that:

$$\begin{aligned}
 &I \left(Y_k^{(2)}; Y^{(1),k} | Y^{(2),k-1} \right) + I \left(Y_k^{(1)}; Y^{(2),k-1} | Y^{(1),k-1} \right) \\
 &= I \left(Y_k^{(2)}; Y_k^{(1)} | Y^{(1),k-1}, Y^{(2),k-1} \right) + I \left(Y_k^{(1)}; Y^{(2),k-1} | Y^{(1),k-1} \right) + I \left(Y_k^{(2)}; Y^{(1),k-1} | Y^{(2),k-1} \right). \tag{A13}
 \end{aligned}$$

Consider the first time step of $k = 1$:

$$\begin{aligned}
 I \left(X_1; Y_1^{(1)}, Y_1^{(2)} | \emptyset \right) + I \left(Y_1^{(2)}; Y_1^{(1)} | \emptyset, \emptyset \right) + I \left(Y_1^{(1)}; \emptyset | \emptyset \right) + I \left(Y_1^{(2)}; \emptyset | \emptyset \right) &= I \left(X_1; Y_1^{(1)}, Y_1^{(2)} \right) \\
 &\quad + I \left(Y_1^{(2)}; Y_1^{(1)} \right) \tag{A14}
 \end{aligned}$$

Now:

$$I \left(X_1; Y_1^{(1)}, Y_1^{(2)} \right) + I \left(Y_1^{(2)}; Y_1^{(1)} \right) \stackrel{(a)}{\geq} I \left(X_1; Y_{1,G}^{(1)}, Y_{1,G}^{(2)} \right) + I \left(Y_1^{(2)}; Y_1^{(1)} \right) \tag{A15}$$

$$\begin{aligned}
 &= I \left(X_1; Y_{1,G}^{(1)} \right) + I \left(X_1; Y_{1,G}^{(2)} \right) + I \left(Y_{1,G}^{(1)}; Y_{1,G}^{(2)} | X \right) \\
 &\quad - I \left(Y_{1,G}^{(1)}; Y_{1,G}^{(2)} \right) + I \left(Y_1^{(1)}; Y_1^{(2)} \right), \tag{A16}
 \end{aligned}$$

where subscript G denotes Gaussian random variables, and (a) follows from ([46] Theo. 1.8.6) with equality if $Y^{(1)}, Y^{(2)}$ are jointly Gaussian with X . The last equality follows from the identity:

$$I(A; B, C) = I(A; B) + I(A; C) + I(B; C|A) - I(B; C).$$

We recall that the noncausal and arbitrary-delay, lower bound of El-Gamal and Cover [11] is tight for scalar IID Gaussian sources. In the first time step, we may regard X_1 as a sample from a white Gaussian process with distribution $\mathcal{N}(0, \text{Var}[X_1])$. Therefore, the causal and zero-delay coding rate of X_1 can never do better than the lower bound of El-Gamal and Cover. We recognize the first three terms in Equation (A16) as the El-Gamal and Cover region. Thus, the difference $I(Y_1^{(1)}; Y_1^{(2)}) - I(Y_{1,G}^{(1)}; Y_{1,G}^{(2)})$ can never be negative, since this would violate the tightness of the lower bound. Therefore:

$$I(X_1; Y_1^{(1)}, Y_1^{(2)}) + I(Y_1^{(2)}; Y_1^{(1)}) \geq I(X_1; Y_{1,G}^{(1)}) + I(X_1; Y_{1,G}^{(2)}) + I(Y_{1,G}^{(1)}; Y_{1,G}^{(2)}|X) \tag{A17}$$

$$= I(X_1; Y_{1,G}^{(1)}, Y_{1,G}^{(2)}) + I(Y_{1,G}^{(2)}; Y_{1,G}^{(1)}) \tag{A18}$$

with equality if $Y_1^{(1)}, Y_1^{(2)}$ are jointly Gaussian.

Now, for the next time step of $k = 2$, we consider:

$$I(X^2; Y_2^{(1)}, Y_2^{(2)}|Y_1^{(1)}, Y_1^{(2)}) + I(Y_2^{(2)}; Y_2^{(1)}|Y_1^{(1)}, Y_1^{(2)}) + I(Y_2^{(1)}; Y_1^{(2)}|Y_1^{(1)}) + I(Y_2^{(2)}; Y_1^{(1)}|Y_1^{(2)}). \tag{A19}$$

However, we just showed that to be optimal in the first step $Y_1^{(1)}, Y_1^{(2)}$ should be jointly Gaussian. Therefore, under the sequential greedy condition, we have that:

$$\begin{aligned} & I(X^2; Y_2^{(1)}, Y_2^{(2)}|Y_1^{(1)}, Y_1^{(2)}) + I(Y_2^{(2)}; Y_2^{(1)}|Y_1^{(1)}, Y_1^{(2)}) + I(Y_2^{(1)}; Y_1^{(2)}|Y_1^{(1)}) + I(Y_2^{(2)}; Y_1^{(1)}|Y_1^{(2)}) \\ &= I(X^2; Y_2^{(1)}, Y_2^{(2)}|Y_{1,G}^{(1)}, Y_{1,G}^{(2)}) + I(Y_2^{(2)}; Y_2^{(1)}|Y_{1,G}^{(1)}, Y_{1,G}^{(2)}) + I(Y_2^{(1)}; Y_{1,G}^{(2)}|Y_{1,G}^{(1)}) + I(Y_2^{(2)}; Y_{1,G}^{(1)}|Y_{1,G}^{(2)}) \end{aligned} \tag{A20}$$

Let:

$$W^2 \triangleq X^2 - \mathbb{E}[X^2|Y_{1,G}^{(1)}, Y_{1,G}^{(2)}] \tag{A21}$$

$$U_2^{(i)} \triangleq Y_2^{(i)} - \mathbb{E}[Y_2^{(i)}|Y_{1,G}^{(1)}, Y_{1,G}^{(2)}], \quad i = 1, 2 \tag{A22}$$

be the residuals for the MMSE predictions of $X^2, Y_2^{(1)}, Y_2^{(2)}$ given $Y_{1,G}^{(1)}, Y_{1,G}^{(2)}$. Then, considering the first two terms in Equation (A20), we have that:

$$\begin{aligned} & I(X^2; Y_2^{(1)}, Y_2^{(2)}|Y_{1,G}^{(1)}, Y_{1,G}^{(2)}) + I(Y_2^{(2)}; Y_2^{(1)}|Y_{1,G}^{(1)}, Y_{1,G}^{(2)}) \\ &= I(W^2; U_2^{(1)}, U_2^{(2)}|Y_{1,G}^{(1)}, Y_{1,G}^{(2)}) + I(U_2^{(2)}; U_2^{(1)}|Y_{1,G}^{(1)}, Y_{1,G}^{(2)}) \end{aligned} \tag{A23}$$

By the orthogonality principle, the residuals of the MMSE estimators are uncorrelated with the conditioning variables, $Y_{1,G}^{(1)}, Y_{1,G}^{(2)}$ [34]. Therefore, since X^2 is Gaussian, W^2 is Gaussian and independent of $Y_{1,G}^{(1)}, Y_{1,G}^{(2)}$. Furthermore, by the conditional residual independence assumption, $U_2^{(i)}, i = 1, 2$ are assumed independent of $Y_{1,G}^{(1)}, Y_{1,G}^{(2)}$, which is true for Gaussian $Y_2^{(i)}$. Thus:

$$I(W^2; U_2^{(1)}, U_2^{(2)}|Y_{1,G}^{(1)}, Y_{1,G}^{(2)}) + I(U_2^{(2)}; U_2^{(1)}|Y_{1,G}^{(1)}, Y_{1,G}^{(2)}) = I(W^2; U_2^{(1)}, U_2^{(2)}) + I(U_2^{(2)}; U_2^{(1)}). \tag{A24}$$

Using the same technique and arguments as in the first time step, we can lower bound these two terms:

$$I(W^2; U_2^{(1)}, U_2^{(2)}) + I(U_2^{(2)}; U_2^{(1)}) \geq I(W^2; U_{2,G}^{(1)}, U_{2,G}^{(2)}) + I(U_{2,G}^{(2)}; U_{2,G}^{(1)}) \tag{A25}$$

with equality if $U_2^{(1)}, U_2^{(2)}$ are jointly Gaussian, or equivalently when $Y_2^{(1)}, Y_2^{(2)}$ are jointly Gaussian.

Now consider the last two terms in Equation (A20):

$$\begin{aligned} I(Y_2^{(1)}; Y_{1,G}^{(2)} | Y_{1,G}^{(1)}) + I(Y_2^{(2)}; Y_{1,G}^{(1)} | Y_{1,G}^{(2)}) &= I(Y_2^{(1)} - E[Y_2^{(1)} | Y_{1,G}^{(1)}]; Y_{1,G}^{(2)} - E[Y_{1,G}^{(2)} | Y_{1,G}^{(1)}] | Y_{1,G}^{(1)}) \\ &\quad + I(Y_2^{(2)} - E[Y_2^{(2)} | Y_{1,G}^{(2)}]; Y_{1,G}^{(1)} - E[Y_{1,G}^{(1)} | Y_{1,G}^{(2)}] | Y_{1,G}^{(2)}) \\ &\stackrel{(b)}{=} I(Y_2^{(1)} - E[Y_2^{(1)} | Y_{1,G}^{(1)}]; Y_{1,G}^{(2)} - E[Y_{1,G}^{(2)} | Y_{1,G}^{(1)}]) \\ &\quad + I(Y_2^{(2)} - E[Y_2^{(2)} | Y_{1,G}^{(2)}]; Y_{1,G}^{(1)} - E[Y_{1,G}^{(1)} | Y_{1,G}^{(2)}]), \end{aligned} \tag{A26}$$

where (b) follows since we assume conditional prediction residual independence of the MMSE predictors. Since the residuals on the right side of the mutual informations are Gaussian, the mutual information is minimized if the residuals on the left, $Y_2^{(i)} - E[Y_2^{(i)} | Y_{1,G}^{(i)}], i = 1, 2$, are Gaussian—that is, when $Y_2^{(i)}, i = 1, 2$ are Gaussian. Thus:

$$\begin{aligned} I(X^2; Y_2^{(1)}, Y_2^{(2)} | Y_1^{(1)}, Y_1^{(2)}) + I(Y_2^{(2)}; Y_2^{(1)} | Y_1^{(1)}, Y_1^{(2)}) + I(Y_2^{(1)}; Y_1^{(2)} | Y_1^{(1)}) + I(Y_2^{(2)}; Y_1^{(1)} | Y_1^{(2)}) \\ \geq I(X^2; Y_{2,G}^{(1)}, Y_{2,G}^{(2)} | Y_{1,G}^{(1)}, Y_{1,G}^{(2)}) + I(Y_{2,G}^{(2)}; Y_{2,G}^{(1)} | Y_{1,G}^{(1)}, Y_{1,G}^{(2)}) + I(Y_{2,G}^{(1)}; Y_{1,G}^{(2)} | Y_{1,G}^{(1)}) + I(Y_{2,G}^{(2)}; Y_{1,G}^{(1)} | Y_{1,G}^{(2)}) \end{aligned} \tag{A27}$$

with equality if $Y_2^{(1)}, Y_2^{(2)}$ are jointly Gaussian, given that $Y_1^{(1)}, Y_1^{(2)}$ are jointly Gaussian, which they are by the sequential greedy assumption. The result now follows by induction on k , dividing by n , and taking the limit as $n \rightarrow \infty$. \square

Notation

Symbol	Description
\mathbb{R}	The set of real numbers
\mathbb{N}	The set of natural numbers
\mathbb{N}^n	The set $\{1, \dots, n\}, n \in \mathbb{N}$
X	Random variable
\mathcal{X}	Alphabet for the random variable X
X^n	Sequence of $n \in \mathbb{N}$ random variables (X_1, \dots, X_n)
x^n	Sequence of $n \in \mathbb{N}$ random variable realizations, where $x^n \in \mathcal{X}^n$
\mathcal{X}^n	$\times_{k=1}^n \mathcal{X}_k$, with $\mathcal{X}_k = \mathcal{X}$
$X \perp\!\!\!\perp Y$	For independent random variables, X, Y
$X - Y - Z$	When the random variables X, Y, Z form a Markov chain in that order i.e., when $P(X, Z Y) = P(X Y)P(Z Y)$
$X _W - Y _W - Z _W$	If the Markov chain is conditioned upon W

References

- Gubbi, J.; Buyya, R.; Marusic, S.; Palaniswami, M. Internet of Things (IoT): A Vision, Architectural Elements, and Future Directions. *Future Gener. Comput. Syst.* **2013**, *29*, 1645–1660. [CrossRef]
- Østergaard, J.; Quevedo, D.E.; Jensen, J. Low delay moving-horizon multiple-description audio coding for wireless hearing aids. In Proceedings of the 2009 IEEE International Conference on Acoustics, Speech and Signal Processing, Taipei, Taiwan, 19–24 April 2009; pp. 21–24. [CrossRef]

3. Krueger, H.; Vary, P. A New Approach for Low-Delay Joint-Stereo Coding. In Proceedings of the ITG Conference on Voice Communication [8. ITG-Fachtagung], Aachen, Germany, 8–10 October 2008; pp. 1–4.
4. Schuller, G.; Kovacevic, J.; Masson, F.; Goyal, V.K. Robust low-delay audio coding using multiple descriptions. *IEEE Trans. Speech Audio Process.* **2005**, *13*, 1014–1024. [[CrossRef](#)]
5. Silva, E.I.; Derpich, M.S.; Østergaard, J. A Framework for Control System Design Subject to Average Data-Rate Constraints. *IEEE Trans. Autom. Control* **2011**, *56*, 1886–1899. [[CrossRef](#)]
6. Tatikonda, S.C. Control Under Communication constraints. Ph.D. Thesis, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA, USA, 2000.
7. Tatikonda, S.; Sahai, A.; Mitter, S. Stochastic linear control over a communication channel. *IEEE Trans. Autom. Control* **2004**, *49*, 1549–1561. [[CrossRef](#)]
8. Stavrou, P.A.; Østergaard, J.; Charalambous, C.D. Zero-Delay Rate Distortion via Filtering for Vector-Valued Gaussian Sources. *IEEE J. Sel. Top. Signal Process.* **2018**, *12*, 841–856. [[CrossRef](#)]
9. Goyal, V.K. Multiple description coding: Compression meets the network. *IEEE Signal Process. Mag.* **2001**, *18*, 74–93. [[CrossRef](#)]
10. Ozarow, L. On a source-coding problem with two channels and three receivers. *Bell Syst. Tech. J.* **1980**, *59*, 1909–1921. [[CrossRef](#)]
11. Gamal, A.E.; Cover, T. Achievable rates for multiple descriptions. *IEEE Trans. Inf. Theory* **1982**, *28*, 851–857. [[CrossRef](#)]
12. Østergaard, J.; Kochman, Y.; Zamir, R. Colored-Gaussian Multiple Descriptions: Spectral and Time-Domain Forms. *IEEE Trans. Inf. Theory* **2016**, *62*, 5465–5483. [[CrossRef](#)]
13. Dragotti, P.L.; Servetto, S.D.; Vetterli, M. Optimal filter banks for multiple description coding: Analysis and synthesis. *IEEE Trans. Inf. Theory* **2002**, *48*, 2036–2052. [[CrossRef](#)]
14. Chen, J.; Tian, C.; Diggavi, S. Multiple Description Coding for Stationary Gaussian Sources. *IEEE Trans. Inf. Theory* **2009**, *55*, 2868–2881. [[CrossRef](#)]
15. Mehmetoglu, M.S.; Akyol, E.; Rose, K. Analog multiple descriptions: A zero-delay source-channel coding approach. In Proceedings of the 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Shanghai, China, 20–25 March 2016; pp. 3871–3875. [[CrossRef](#)]
16. Stavrou, P.A.; Østergaard, J.; Skoglund, M. On Zero-delay Source Coding of LTI Gauss-Markov Systems with Covariance Matrix Distortion Constraints. In Proceedings of the 2018 European Control Conference (ECC), Limassol, Cyprus, 12–15 June 2018; pp. 3083–3088. [[CrossRef](#)]
17. Derpich, M.S.; Østergaard, J. Improved upper bounds to the causal quadratic rate-distortion function for Gaussian stationary sources. *IEEE Int. Symp. Inf. Theory* **2010**, 76–80. [[CrossRef](#)]
18. Neuhoff, D.; Gilbert, R. Causal source codes. *IEEE Trans. Inf. Theory* **1982**, *28*, 701–713. [[CrossRef](#)]
19. Tanaka, T.; Kim, K.K.; Parrilo, P.A.; Mitter, S.K. Semidefinite Programming Approach to Gaussian Sequential Rate-Distortion Trade-Offs. *IEEE Trans. Autom. Control* **2017**, *62*, 1896–1910. [[CrossRef](#)]
20. Silva, E.I.; Derpich, M.S.; Østergaard, J.; Encina, M.A. A Characterization of the Minimal Average Data Rate That Guarantees a Given Closed-Loop Performance Level. *IEEE Trans. Autom. Control* **2016**, *61*, 2171–2186. [[CrossRef](#)]
21. Barforooshan, M.; Østergaard, J.; Stavrou, P.A. Achievable performance of zero-delay variable-rate coding in rate-constrained networked control systems with channel delay. In Proceedings of the 2017 IEEE 56th Annual Conference on Decision and Control (CDC), Melbourne, VIC, Australia, 12–15 December 2017; pp. 5991–5996. [[CrossRef](#)]
22. Massey, J. Causality, feedback and directed information. In Proceedings of the International Symposium on Information Theory and its applications (ISITA-90), Honolulu, HI, USA, 27–30 November 1990; pp. 303–305.
23. Kochman, Y.; Zamir, R. Analog Matching of Colored Sources to Colored Channels. *IEEE Trans. Inf. Theory* **2011**, *57*, 3180–3195. [[CrossRef](#)]
24. Østergaard, J.; Quevedo, D.E.; Jensen, J. Real-Time Perceptual Moving-Horizon Multiple-Description Audio Coding. *IEEE Trans. Signal Process.* **2011**, *59*, 4286–4299. [[CrossRef](#)]
25. Liu, W.; Vijayanagar, K.R.; Kim, J. Low-delay distributed multiple description coding for error-resilient video transmission. In Proceedings of the IEEE 13th International Workshop on Multimedia Signal Processing, Hangzhou, China, 17–19 October 2011; pp. 1–6. [[CrossRef](#)]

26. Østergaard, J.; Quevedo, D. Multiple Description Coding for Closed Loop Systems over Erasure Channels. In Proceedings of the 2013 Data Compression Conference, Snowbird, UT, USA, 20–22 March 2013; pp. 311–320. [[CrossRef](#)]
27. Fuglsig, A.J. Zero-Delay Multiple Descriptions of Stationary Scalar Gauss-Markov Sources Using Feedback. Mater's Thesis, Department of Electronic Systems and Department of Mathematical Sciences, Aalborg University, Aalborg, Denmark, 2019.
28. Cover, T.M.; Thomas, J.A. *Elements of Information Theory*, wiley student ed.; Wiley-Interscience: Hoboken, NJ, USA, 2006.
29. Wyner, A.; Ziv, J. The rate-distortion function for source coding with side information at the decoder. *IEEE Trans. Inf. Theory* **1976**, *22*, 1–10. [[CrossRef](#)]
30. Zamir, R. *Lattice Coding for Signals and Networks*; Cambridge University Press: Cambridge, UK, 2014.
31. Zamir, R. Gaussian codes and Shannon bounds for multiple descriptions. *IEEE Trans. Inf. Theory* **1999**, *45*, 2629–2636. [[CrossRef](#)]
32. Chen, J.; Tian, C.; Berger, T.; Hemami, S.S. Multiple Description Quantization Via Gram-Schmidt Orthogonalization. *IEEE Trans. Inf. Theory* **2006**, *52*, 5197–5217. [[CrossRef](#)]
33. Berger, T. *Rate Distortion Theory: A Mathematical Basis for Data Compression*; Prentice-Hall series in information and system sciences; Prentice Hall: Urbana, IL, USA, 1971.
34. Madsen, H.; Thyregod, P. *Introduction to General and Generalized Linear Models*; Chapman & Hall/CRC Texts in Statistical Science; CRC Press: London, UK, 2011.
35. Wang, H.; Viswanath, P. Vector Gaussian Multiple Description With Individual and Central Receivers. *IEEE Trans. Inf. Theory* **2007**, *53*, 2133–2153. [[CrossRef](#)]
36. Østergaard, J.; Kochman, Y.; Zamir, R. An Asymmetric Difference Multiple Description Gaussian Noise Channel. In Proceedings of the 2017 Data Compression Conference (DCC), Snowbird, UT, USA, 4–7 April 2017; pp. 360–369. [[CrossRef](#)]
37. Østergaard, J.; Zamir, R. Multiple-Description Coding by Dithered Delta—Sigma Quantization. *IEEE Trans. Inf. Theory* **2009**, *55*, 4661–4675. [[CrossRef](#)]
38. Vaishampayan, V.A. Design of multiple description scalar quantizers. *IEEE Trans. Inf. Theory* **1993**, *39*, 821–834. [[CrossRef](#)]
39. Sun, G.; Samarawickrama, U.; Liang, J.; Tian, C.; Tu, C.; Tran, T.D. Multiple Description Coding With Prediction Compensation. *IEEE Trans. Image Process.* **2009**, *18*, 1037–1047. [[CrossRef](#)] [[PubMed](#)]
40. Chao, T.; Hemami, S.S. A new class of multiple description scalar quantizer and its application to image coding. *IEEE Signal Process. Lett.* **2005**, *12*, 329–332. [[CrossRef](#)]
41. Samarawickrama, U.; Liang, J. A two-stage algorithm for multiple description predictive coding. In Proceedings of the Canadian Conference on Electrical and Computer Engineering, Niagara Falls, ON, Canada, 4–7 May 2008; pp. 000685–000688. [[CrossRef](#)]
42. Frank-Dayan, Y.; Zamir, R. Dithered lattice-based quantizers for multiple descriptions. *IEEE Trans. Inf. Theory* **2002**, *48*, 192–204. [[CrossRef](#)]
43. Diggavi, S.N.; Vaishampayan, V.A. On multiple description source coding with decoder side information. In Proceedings of the Information Theory Workshop, San Antonio, TX, USA, 24–29 October 2004; pp. 88–93. [[CrossRef](#)]
44. Taubman, D.S. *JPEG2000 Image Compression Fundamentals: Standards and Practice*; Kluwer: Boston, MA, USA, 2002.
45. Massey, J.L.; Massey, P.C. Conservation of mutual and directed information. In Proceedings of the International Symposium on Information Theory, ISIT, Adelaide, SA, Australia, 4–9 September 2005; pp. 157–158. [[CrossRef](#)]
46. Ihara, S. *Information Theory for Continuous Systems*; World Scientific Publishing Inc.: Singapore, 1993.

