

Master's Thesis

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

by
Andreas Jonas Fuglsig
Mathematical Engineering

Aalborg University
Department of Electronic Systems &
Department of Mathematical Sciences

Copyright © Aalborg University 2019
Python 3.6.8 is used for the simulations in this report.
If not otherwise stated the figures in this report are made by the author.



Department of Electronic Systems &
Department of Mathematical Sciences
Aalborg University
<http://www.aau.dk>

AALBORG UNIVERSITY

Title:

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

Theme:

Master's Thesis (60 ECTS) - Mathematical Engineering

Project Period:

September 2018 - June 2019

Participant:

Andreas Jonas Fuglsig

Supervisor:

Jan Østergaard

Copies: 1

Page Numbers: 140

Date of Completion:

June 7, 2019

The content of this report is freely available, but publication (with reference) may only be pursued due to agreement with the author.

Abstract

An important part of digital communication is source coding, i.e. the compression of a source signal to a digital representation (in bits) that may be transmitted across a network and used to reconstruct the source signal at a decoder. The main problem of source coding is to determine the fundamental lowest limit on the data-rate required to reconstruct the source, while achieving a desired distortion, known as the rate-distortion function (RDF). However, source codes that achieve performance close to these bounds often suffer from long delays and are not suitable for real-time communication. Therefore, zero-delay (ZD) or instantaneous source coding is crucial in real-time communication.

In communication over unreliable networks data packets may be lost. Therefore, where retransmissions are either not possible or permitted, excess bandwidth is often spent on the channel code to guarantee reliable and satisfactory real-time communication. As an alternative, multiple descriptions (MDs) facilitate a graceful degradation during partial network failures.

In this contribution, we provide initial results in the combination of ZD- and MD coding, with two descriptions, in situations with perfect decoder-feedback, and consider ZDMD coding of stationary scalar Gauss-Markov sources under mean-squared error distortion constraints. Our five main contributions are:

(1) A novel information-theoretic lower bound on the sum-rate of a ZDMD source code with decoder feedback, and mutually independent decoder side information. The bound is given by the sum of; the directed information rate from the source to the MD reproductions, and the mutual information rate between the MD reproductions. (2) By extension of the previous result, an information-theoretic lower bound on the operational symmetric ZDMD RDF. (3) Under some technical assumptions, for stationary scalar Gauss-Markov sources the bound on the ZDMD RDF is minimized by Gaussian test-channel distributions. (4) The optimum Gaussian test-channel distribution is realizable by a double branch test-channel with two predictive coding loops. This also shows achievability of the lower bound in a Gaussian coding scheme. (5) Using a simple existing MD quantization scheme, similar to our ZDMD test-channel realization, numerical simulations show operational performance within 5 dB of the theoretical distortion limits in the high-rate scenario.

Dansk Resumé

Realtidskommunikation er vigtigt i mange moderne teknologier, f.eks. online videomøder og kontrolsystemer implementeret over trådløse netværk. I alle disse scenarier kan der være strenge krav til forsinkelser og pålidelighed.

I dette projekt betragter vi realtids kildekodning, det vil sige komprimering af et kildesignal til en digital repræsentation (i bits), som transmitteres over et netværk til en dekoder, der producerer en rekonstruktion af kilden. Den største udfordring ved dette er at designe et realtids komprimeringsystem, således et kildesignal kan transmitteres over et upålideligt netværk og genskabes ved dekoderen med en foreskrevet gengivelsesnøjagtighed ved brug af det mindst mulige antal bits.

Kildekodning i relation til informationsteori er kendt som *rate-distortion* teori. Rate-distortion teori omhandler bestemmelse af den fundamentalt laveste gennemsnitlige datarate, som kræves for at genskabe en kilde ved dekoderen mens det ønskede distortion-kriterie overholdes for en specifik kilde og distortion-mål. Denne fundamentale grænse kaldes *rate-distortion funktionen* (RDF).

De kildekoder, som kan opnå performance tæt på de fundamentale grænser lider dog ofte under lange signalforsinkelser, hvilket betyder, at signalet komprimeres og dekodes ikke øjeblikkeligt, når en sample modtages. I realtidskommunikation skal indkodning og dekodning ske omgående når en sample modtages, vi taler da om *zero-delay* (ZD) kodning.

Indkodning og dekodning er ikke de eneste steder i systemet, hvor mulige forsinkelser indtræffer. Når kommunikationen foregår over pakkebaserede netværk, kan pakkeab resultere i betragtelige forsinkelser p.g.a. gentransmittering af tabte pakker. Derfor, der hvor retransmission ikke er muligt eller tilladt, er det generelt nødvendigt at bruge urimeligt meget båndbredde på kanalkoden for at garantere pålidelig kommunikation og tilfredsstillende performance. Som et alternativ kan indkoderen producere *multiple descriptions* (MD) af kilden, der faciliterer en gradvis forringelse af rekonstruktionen ved delvise netværksfejl.

I den offentligt tilgængelig litteratur er en samlet teori om zero-delay multiple-description (ZDMD) kodning stort set ubeskrevet. I dette bidrag præsenterer vi initierende resultater for kombinationen af ZD- og MD kodning i situationer med

perfekt feedback fra dekoderen, og betragter ZDMD kodning af stationære skalarer Gauss-Markov kilder under *mean-squared error* (MSE) distortion-kriterier.

Vi generaliserer først det klassiske MD scenarie til at inkludere åben-loop dekoder feedback, det vil sige feedbacken har ingen effekt på kildesignalet, og til at inkludere sideinformation tilgængelig for både indkoder og dekoder. Ud fra dette konstruktive setup udleder vi en ny informationsteoretisk nedre grænse på sumraten for en ZDMD kildekode med dekoder feedback og indbyrdes uafhængig dekoder sideinformation. Denne nedre grænse udgøres af summen af; den retningsbestemte informationsrate fra kilde til MD rekonstruktionerne, og den fælles informationsrate mellem MD rekonstruktionerne.

Vi viser, at denne nye nedre grænse udgør en informationsteoretisk nedre grænse på den operationelle symmetriske ZDMD RDF. For stationære skalarer Gauss-Markov kilder, under tekniske de betingelser af *grådig sekventiel kodning* og *uafhængige minimum MSE prædiktions residualer*, viser vi, at den nedre grænse på den operationelle ZDMD RDF minimeres af Gaussiske test-kanal fordelinger. Det vil sige, Gaussiske reproduktionerne er optimale i at minimere raten under samme distortion-kriterier.

Vi viser desuden, at de optimale Gaussiske test-kanal fordelinger kan realiseres ved at kombinere prædikativ kodning med en to-grens test-kanal. Endvidere er disse fordelinger karakteriseret ved løsningen på et optimeringsproblem.

Til sidst, implementere vi et simpelt eksisterende operationelt MD kvantiserings-system, der kan sammenlignes med vores ZDMD test-kanal realiserings-system. Dette viser, at det er muligt at opnå operationel distortion-performance approksimativt 5 dB fra de teoretiske nedre grænser.

Contents

Abstract	v
Dansk Resumé	vii
Preface	xiii
Introduction	1
Problem Statement	9
Notation	13
1 Zero-Delay Coding	15
1.1 Zero-Delay Source Code	15
1.2 Directed Information	18
1.3 Lower Bound	20
1.4 Differential Pulse Code Modulation	21
1.5 Test-Channel Realization	22
1.6 Non-Causal Rate-Gain	26
2 Multiple-Description Coding	31
2.1 Multiple-Description Source Code	31
2.2 MD Coding of White Gaussian Sources	34
2.2.1 Ozarow's Test-Channel	35
2.2.2 White Gaussian MD Rate-Region	37
2.3 Symmetric Distortions	40
2.4 Zero-Delay Multiple-Description Coding	41
2.5 Problem Definition	44
3 Lower Bound	45
3.1 ZDMD Coding with Feedback	46
3.2 Lower Bound on Sum-rate	48

3.3	Gaussian Lower Bound	52
3.3.1	Distributions	52
3.3.2	Lower Bound	53
4	Test-Channel Realization	59
4.1	Source Process	59
4.2	Predictive Coding	59
4.3	Central Decoder Design	61
4.3.1	Joint Estimation of Error Process	63
4.3.2	Joint Estimation of Source	63
4.3.3	MMSE Estimate Using $Y^{(0,C)}$	63
4.4	Rates	64
4.5	Scalar Lower Bound Theorem	65
5	Simple Quantization Scheme	67
5.1	Staggered Quantizers	68
5.2	The Scheme of [33]	70
5.2.1	Optimal Predictors	70
5.2.2	Prediction Error Covariance	71
5.2.3	Choosing step size	73
5.3	Simulation Study	74
5.3.1	Distortion Trade-Off at Fixed Rate	75
5.3.2	Distortion versus Distortion-Ratio for Multiple Fixed Rates	76
	Discussion	78
	Conclusion	83
	Future Research	85
	Bibliography	89
A	Information Theory	95
A.1	Discrete Entropy	95
A.2	Differential Entropy	101
A.2.1	Relation to Discrete Entropy	104
A.3	Directed Information	105
B	Source coding	107
B.1	Sources	107
B.1.1	Stochastic Processes	108
B.2	Source Coding	110
B.2.1	Lossless Source Coding	110

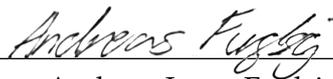
B.2.2	Lossy Source Coding	113
C	Rate-Distortion Theory	115
C.1	Rate-Distortion Coding	115
C.1.1	Distortion	116
C.1.2	Rate	116
C.2	Rate-Distortion Function	117
C.3	Information Rate-Distortion Function	118
C.4	Gaussian RDF	121
C.4.1	Water-Filling	122
D	Proof of Theorem 3.5 and Corollary 3.6	127
D.1	Proof of Theorem 3.5	127
D.2	Proof of Corollary 3.6	129
E	Proof of Theorem 3.10	131
F	Proof of Lemma 4.1	135
G	Proof of Lemma 4.2	139

Preface

This Master's Thesis (60 ECTS) is produced in completion of the MSc program in Mathematical Engineering at the Faculty of Engineering and Science at Aalborg University.

I wish to extend my gratitude to my supervisor Jan Østergaard for countless interesting and helpful discussions, and advice, throughout the entirety of the project period, and without whom this project would not have been possible.

Aalborg University, June 7, 2019



Andreas Jonas Fuglsig
afugls14@student.aau.dk

Introduction

Real-time communication is desirable in many modern applications e.g. Internet of Things (IoT) [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 (NCS's) [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].

This report is concerned with real-time communication of signals across unreliable networks. The situation is illustrated in Figure 1, where the source signal, X , is communicated across a possibly wireless network and reproduced at a decoder. In order to design an appropriate communication system we assume that a signal model, $X(\theta)$, has been determined or specified in advance. In particular we are concerned with the following auto-regressive (AR) model

$$X_k = AX_{k-1} + CW_k, \quad k \in \mathbb{N}, \quad (1)$$

where $A \in \mathbb{R}^{p \times p}$, $C \in \mathbb{R}^{p \times q}$ are deterministic matrices, $X_1 \sim \mathcal{N}(0, \Sigma_{X_1})$ is the initial state and $W_k \in \mathbb{R}^q \sim \mathcal{N}(0, \Sigma_W)$ is a white Gaussian process independent of X_1 . Here, the signal model parameters are $\theta = (A, C, \Sigma_W, \Sigma_{X_1})$.

Given the signal model, the relevant source parameters, θ , may then be extracted from the source signal (e.g., the LPC parameters of a speech signal). Given the source parameters, the signal is compressed and converted to bits. These bits are then transmitted across a network to a decoder, which produces an estimate, Y , of the signal, X . We also assume the decoder output, Y , is available at the encoder in the form of a feedback signal.¹

In this report we do not consider the real-time extraction of source parameters. Thus, we assume the source is known in the sense that all source parameters and statistical properties have already been extracted and are available to us. That is, the source is completely specified by the signal model and its parameter values. This is illustrated by the dashed square in Figure 1.

¹For example this would be the case if the encoder can mimic the decoder in the absence of packet losses.

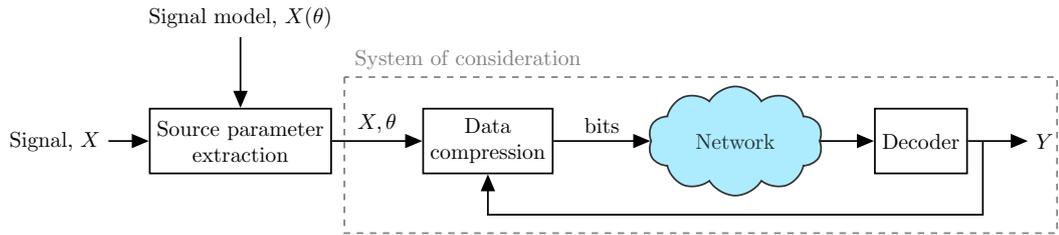


Figure 1: An overview block diagram relating the problem we consider to a wider system perspective. In the overall systems, given a signal model, $X(\theta)$, initially source parameters, θ , are extracted from the source signal X . Using the parameters the signal is compressed into a digital (bit) representation, these bits are then transmitted across an unreliable network to a decoder, that produces an estimate, Y , of the source signal, X . The compression uses feedback from the previous reproductions.

The main problem is then to design a real-time compression scheme such that the source, X , may be transmitted across an unreliable network and reproduced at the decoder while achieving a prescribed degree of fidelity using the minimum amount of bits. As we assume feedback from decoder to encoder these blocks are designed in unison.

We further specify the exact problem considered in this report in the last section of the Introduction and in the Problem Statement. In order to establish the final problem, we consider in more detail the different aspects of real-time communication over unreliable networks.

Source Coding

The digital communication of an information source, X , across a noisy channel to a receiver can, as shown in Figure 2, be split into two main parts: *Source encoding* and *-decoding*, and *channel encoding* and *-decoding* [9]. In source coding, the source is mapped into bits, and from bits to a source estimate [10, p. 4]. In channel coding, the digital source representation is mapped to a suitable channel input, and from channel output to a digital representation [10, p. 4]. Delays may occur in any of these parts as well as in the physical transmission of the signal across the channel. These delays are additive and may result in arbitrarily long delays.

In this project we consider only the source coding aspect of real-time communication systems with feedback, and let the concatenation of channel coding and the channel be a *noiseless* digital channel with no delay [11, p. 5]. That is, no bit-errors are introduced in the signal between the source encoder and -decoder.

We are not concerned with any delays due to practical computations. Hence, we consider source coding in the following discrete-time systems setting. Consider the general block-diagram representation of a feedback system illustrated in Fig-

ure 3. In this figure, the systems \mathcal{S}_1 and \mathcal{S}_2 represent possibly non-linear and time-varying systems. All signals in the loop are assumed to be discrete-time processes, and all systems and signals share a synchronized “clock”. The delay block ensures the operational feasibility of the feedback system. Otherwise, output of either system could be a function of itself.

Since the output of the second system, \mathcal{S}_2 , is fed back as the input to the first system, \mathcal{S}_1 , and vice-versa, the systems must be *causal*. That is, the output of either system at any time-instant is not a function of any of its future inputs. The resulting system equations are for each $k \in \mathbb{N}$,

$$U_k = \mathcal{S}_1 \left(X^{k-d_1} \right), \quad (2)$$

$$Y_k = \mathcal{S}_2 \left(U^{k-d_2} \right), \quad (3)$$

$$X_k = Y_{k-1}, \quad (4)$$

where $X^n \triangleq (X_1, \dots, X^n)$ denotes a sequence of n variables, and $d_1, d_2 \in \mathbb{N}_0$ are the delays of \mathcal{S}_1 and \mathcal{S}_2 . We let $d \triangleq d_1 + d_2$ and define the system (2)–(4) as a d -delay feedback system. In a *zero-delay* system, where $d = 0$, each output of \mathcal{S}_2 , i.e. Y_k , must be produced at the same time instant the corresponding input of \mathcal{S}_1 , i.e. X_k , has been processed by \mathcal{S}_1 . Clearly, the system can only be zero-delay if all components are causal, even in the absence of feedback. Throughout this report we consider only feedback systems of this type, where the forward channel from \mathcal{S}_1 to \mathcal{S}_2 is instantaneous (zero-delay) and the feedback is strictly causal, i.e. at least one sample delay.

For source coding inside feedback loops, we may replace \mathcal{S}_1 with an encoder, \mathcal{S}_2 with a decoder, and the delay block with any causal feedback system with at least one sample delay. The link between source encoder and -decoder is considered a noiseless digital channel [11, p. 5].

The encoder mapping from the source signal to bits consist of a quantization and mapping of quantization intervals to bits [5], [10]. The key design problem of source coding is then how to quantize a signal, with the minimum average data-rate (in bits per sample), while achieving a prescribed degree of fidelity (distor-

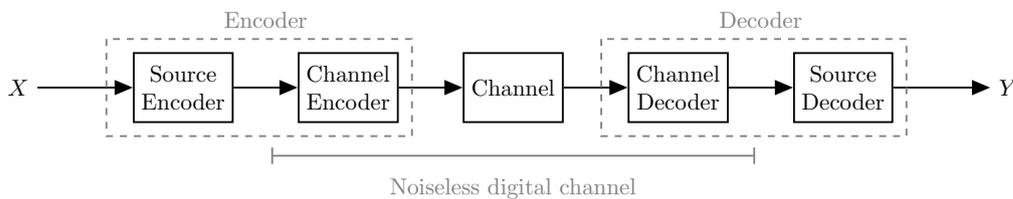


Figure 2: Block diagram representation of a communication system. Figure inspired by [12].

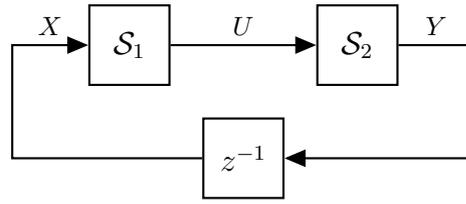


Figure 3: A general feedback system. The output of the second system, S_2 , is fed back as the input to the first system S_1 .

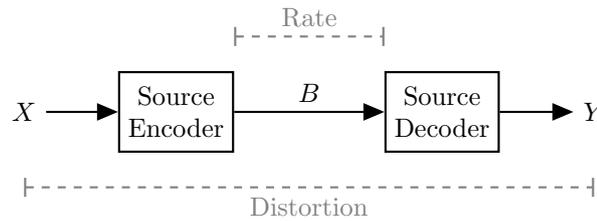


Figure 4: A source coding scenario indicating where rate and distortion are measured. The encoder encodes the source, X , to the binary descriptions, B , the source decoder then produces an estimate, Y , of X given the description B . The rate is measured between the encoder output and decoder input, i.e. the number of bits pr. source symbol transmitted between encoder and decoder. The distortion is measured between the encoder input and decoder output, i.e. the discrepancy between the source, X , and its representation, Y .

tion) or performance [5], [13, ch. 10]. A typical performance measure is the mean squared error (MSE), but other measures are also possible [5], [13, ch. 10]. Figure 4 illustrates where in the source coding scenario the rate and distortion are measured.

Source coding in the context of information theory is known as *rate-distortion* theory [13, ch. 10]. Rate-distortion theory is the determination of the fundamental lowest limit on the average data-rate required to reproduce a source at the decoder while achieving a desired distortion criteria, for a specific source and distortion measure, e.g. MSE [8], [13, ch. 10]. This fundamental limit is called the *rate-distortion function* (RDF), (see Appendix C) [13, ch. 10].

In rate-distortion theory the input-output relation of source coding is modeled from channel coding perspective [13]. That is, the mapping from source to reproduction is modeled by probability distributions. The (information) rate across the channel is then given by the so called *mutual information*, (see Appendix A) [13]. Thus allowing for the definition of an information-theoretic RDF as the minimum *mutual information rate* (see Appendix A) across the channel subject to a distor-

tion constraint. The conditional distributions of the reproductions given the source input that achieves the lower bound are often called *test-channels* [13].

Zero Delay

The standard ways to achieve a given rate-distortion performance relies upon random codebooks. This requires the coding of arbitrarily long sequences which incur arbitrarily long delays [5]. Thus, in practice the source encoders and -decoders that are able to achieve performance close to the fundamental limits often incur long delays in their end-to-end effect, i.e. the total delay only due to source coding, and is often non-causal and computationally expensive [5], [8], [14].

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 [15]. That is, when the reconstruction of each input sample must take place at the same time-instant the corresponding input sample has been encoded [16]. Hence, the random coding technique is not applicable in ZD rate-distortion theory.

Single-description ZD rate-distortion theory has been increasingly more popular in the last decades, due its significance in real-time communication systems and especially feedback systems. The ZD RDF is generally unknown and hard to determine [8]. Some indicative results on ZD source coding for NCSs, and systems with- and without feedback may be found in [5]–[8], [16]–[19].

As mentioned, ZD source coders must be causal [20]. However, causality comes with a price. The results of [16] show, that causal coders increase the bitrate due to the *space-filling loss* of “memoryless” quantizers, (see Section 1.6), and the reduced de-noising capabilities of causal filters. Additionally, imposing ZD increases the bitrate due to memoryless entropy coding, (see Section 1.6)[16].

The results of [5] establishes 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* (see Section 1.2) across the channel. Furthermore, it is shown that for a given MSE performance constraint of a NCS an achievable upper bound differs from the lower bound by only 1.254 bits [5], [18].

For open-loop vector Gauss-Markov sources (1), i.e. the source is not inside a feedback loop, the optimal operational performance of a ZD source code subject to a MSE distortion constraint has been shown to be lower bounded by a minimization of the directed information [21] from the source to the reproductions subject to the same distortion constraint [5]–[7], [16], [17]. For Gaussian sources the directed information is further minimized by Gaussian reproductions [8], [18]. Very recently, Stavrou *et al.* [8] extending upon the works of [7], [16], [17], showed that the optimal test-channel that achieves this lower bound is realizable using a feedback

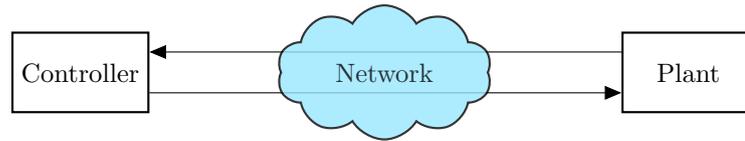


Figure 5: A general NCS, where an unstable plant is stabilized by a controller. The controller and the plant communicate across a possibly wireless network.

realization scheme. Furthermore, [8] extends this to a predictive coding scheme providing an achievable upper bound. A novel approach in the test-channel of [8] is to spatially decorrelate the innovation vector process, then by using “reverse-waterfilling” some dimensions may be deemed inactive. The r active dimensions are then independently scalar quantized and jointly entropy coded, yielding an upper bound $0.254r + 1$ bits/vector above the lower bound. For high rates the extra 1 bit vanishes. Finally, for asymptotically large vectors using vector quantization the upper bound and lower bounds coincide [8]. This almost provides a full characterization of the ZD RDF for vector Gauss-Markov sources subject to MSE distortion constraints when either A or C in (1) are full rank and have finite dimensions. Contrary to standard rate-distortion results where arbitrarily long sequences are coded, optimality for the Gaussian ZD RDF is shown for arbitrarily long vectors.

Multiple Descriptions

The channel in feedback systems may be a wireless network, such as in an NCS illustrated in Figure 5, where the signals between the controller and the plant are communicated across a possibly wireless network. Networked communication adds another possible delay to the feedback scenario, since they are often packet-based [22].

In packet-based networks, a signal is encoded and transmitted progressively in data packets, and then reconstructed as the packets are received. However, in packet-based communication over unreliable networks, packet-loss can result in substantial delays due to retransmission of lost packets [4], [22]. This packet-loss is a result of interference, noise, or fading, due to multi-path effects, moving transmitters or receivers², or blockage of signals [2], [4]. Therefore, 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

²For example a hearing aid user, or a person wearing a wireless microphone

order to guarantee reliable communications and ensure satisfactory performance. Thus, source codes that are only useful if all packets are received put too much faith in the delivery system, whereas the delivery system assumes all packets are needed at the receiver [22]. As an alternative, the channel code may be replaced by cleverly designed data packets, called multiple descriptions (MDs). Contrary to channel codes, MDs would allow for several reproduction qualities at the receivers and thereby admit a graceful degradation during partial network failures [22]. This avoids long delays because no retransmission is allowed, hence some compression (reproduction quality) is sacrificed for an overall lower latency [22].

MD coding is considered, as a joint source-channel coding problem concerned with lossy encoding for transmission over unreliable channels [2]. By using the appropriate source coding techniques, data packets are designed to the conditions of the channel [22]. Thus, some of the excessive bits used on channel coding may be put to better use in the source code.

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.

Figure 6 illustrates the two-description MD coding scenario in both a closed- and an open-loop system. In both cases the encoder produces two descriptions which are transmitted across noiseless channels.

The closed-loop scenario remains an open problem. However, the open-loop problem has been more widely studied in the information-theory literature.

Since MD coding considers several data-rates and distortions, MD rate-distortion theory is the determination of fundamental limits on a rate-distortion region [22].

In open-loop the achievable MD rate-distortion region is only completely known in very few cases [23]. El-Gamal and Cover [24] gave an achievable region for two descriptions and memoryless source. This region was then shown to be tight for white Gaussian sources with MSE distortion constraints by Ozarow [25].

In the high resolution limit, i.e. high rates, Dragotti [26] characterized the achievable region for stationary (time-correlated) Gaussian sources with MSE distortion constraints. This was then extended in [27] to the general resolution case for stationary Gaussian sources. Recently, [23] 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 the MD source coders that are close to the fundamental rate-distortion bounds impose long delays on the end-to-end processing of information [28]. Also clearly from Figure 6, if MD coding is part

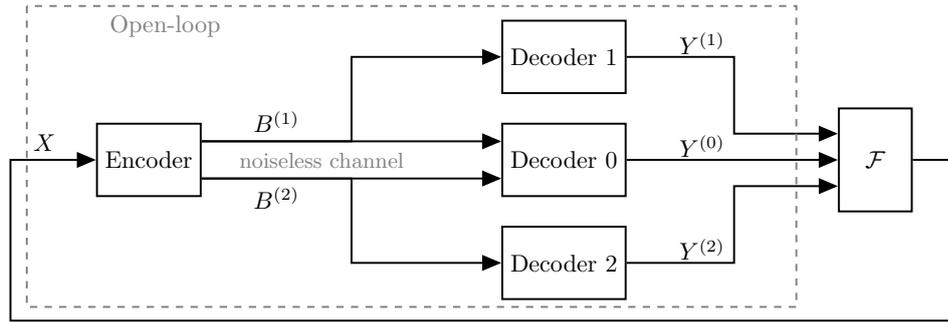


Figure 6: MD source coding in a closed-loop. If there is packet-loss on the noiseless channel the source signal, X , is affected by which descriptions are received at the decoder. The standard open-loop MD coding is marked by the dashed line. In the open-loop the source is completely specified prior to the design of the coding scheme.

of closed-loop systems the source encoding and -decoding must be zero-delay.

Zero-Delay Multiple Descriptions

Recently, [28] 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 [29]. Furthermore, most modern communication systems relies 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], [30], as well as for low-delay video coding in [31]. Some initial work regarding MDs in NCSs may be found in [32]. However, none of these consider the theoretical limitations of ZDMD coding in a rate-distortion sense.

To the best of the authors' knowledge there exists no publicly available research in the field of ZDMD rate-distortion theory neither with or without feedback. Therefore, in this project we take initial steps towards the unification of ZD- and MD rate-distortion theory with feedback.

In order to provide a solution towards closed-loop MD coding, we consider the open-loop ZDMD coding problem illustrated in Figure 7. The MD encoder produces two descriptions $B^{(i)}$, $i = 1, 2$, of the source, X . The two descriptions are transmitted across noiseless channels without packet-loss to the three decoders. Each decoder produces an estimate $Y^{(i)}$, $i = 0, 1, 2$, of the source signal, X . The encoder and decoders must all be zero-delay, such that the end-to-end sample de-

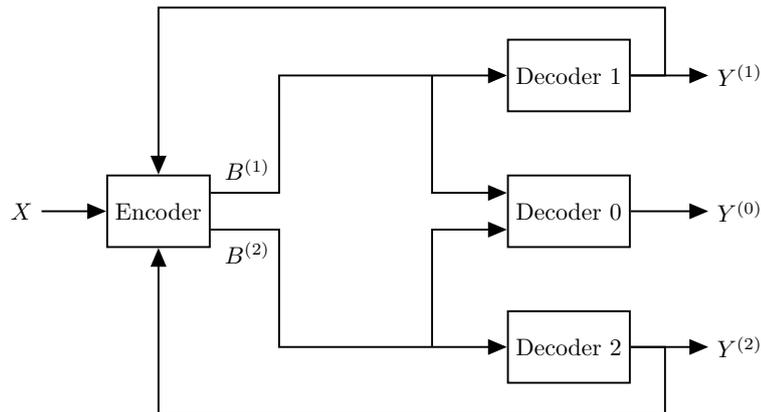


Figure 7: The open-loop scenario considered in this paper, where the source signal, X , is completely specified by its statistical properties, there is a noiseless digital channel without packet-loss from encoder to decoders. There is perfect feedback from the decoder output to the encoder.

lays of all encoder-decoder pairs are zero. The encoder uses perfect feedback from the decoders to produce the two descriptions. The feedback is perfect in the sense that it is an undistorted version of the decoder output. This is possible due to the error-free channels without packet-loss. Thus, contrary to standard MDs the encoder is aware that all packets are received at the decoder.

We do not consider, the classical closed-loop where feedback from decoder affects the encoder input. Hence, we assume the source signal to be completely specified by its statistical properties prior to encoding. Therefore, there is no feedback from the decoder outputs, $Y^{(i)}$, to the source, X .

Specifically we consider scalar stationary Gaussian sources subject to MSE distortion constraints.

Problem Statement

Research Question

What are the minimum required bitrates in a ZDMD source coding scheme of scalar stationary Gaussian sources using perfect decoder-feedback subject to MSE distortion constraints?

To determine an answer to this research question, we answer the following sub questions.

Sub Questions

- What is a constructive information-theoretic lower bound on the operational average data sum-rate of a ZDMD source code using perfect decoder-feedback?
- What is an optimum test-channel for the information-theoretic lower bound?
- What is an operational quantization scheme that extends upon the optimum test-channel and provides achievable rates and distortions?
- What is the gap between the theoretical lower bound and operational achievable rates and distortions?

Delimitations

We do not consider the determination of source parameters. Thus, we assume all parameters of the source (1): $A, C, \Sigma_W, \Sigma_{X_1}$ are deterministic and given. Hence, we provide an initial solution towards the more general scenario of unknown parameters. In order to provide a solution towards vector Gaussian sources, we consider only the scalar case in this report, hence $p = 1$ and $q = 1$ in (1). Furthermore, we consider only stationary and stable source processes. Thus, providing initial solutions that may later be generalized to broader scenarios. Finally, we consider only MSE distortion, since it is the most common distortion measure for continuous sources [13]. Specifically, we do not consider the determination of other useful distortion measures.

Reading Guide

The main part of this report assumes a certain familiarity with common information theory-, source coding- and rate-distortion concepts. Therefore, Appendix A provides an introduction to the most important information-theoretic concepts for the analyzes conducted. Appendix B serves as a primer on the ideas behind source coding, especially lossless source coding, as well as stochastic processes. Finally, Appendix C serves as a primer on the main results of classical rate-distortion theory.

These appendices assume knowledge of all preceding appendices, i.e. Appendix C assumes the reader is familiar with the concepts of Appendix A and B. Similarly, Appendix B assumes knowledge of the concepts in Appendix A. Therefore, readers familiar with certain concepts may skip directly to either appendix.

The rest of the report is organized as follows.

In Chapter 1 we introduce ZD source coding and the results of [8] as background for our results.

In Chapter 2 we first revisit MD source coding and the results of [25] to further the background on our derivations. We then formally present our first contribution; combined ZD- and MD- coding and define the ZDMD problem. We also define the notion of a ZDMD RDF for symmetric distortions.

In Chapter 3, by generalizing the result of [5] we determine a novel information-theoretic lower bound on the average data sum-rate of the ZDMD source code in Figure 7. Using this bound we are able to obtain an information-theoretic lower bound on the ZDMD RDF. We then show that under certain conditions for scalar stationary Gaussian sources this lower bound is minimized by jointly Gaussian MDs.

In Chapter 4 we determine a MD feedback realization scheme for the optimum Gaussian test-channel distribution for the lower bound on the ZDMD RDF. Utilizing this, we present a characterization of the Gaussian achievable lower bound as a solution to an optimization problem.

In Chapter 5 we introduce the staggered predictive quantization scheme of [33] as an extension of our test-channel to an operational ZDMD coding scheme. Using numerical simulations we evaluate the performance of the operational scheme compared to the achievable ZDMD region.

We then discuss and conclude on our results. Finally we consider future directions.

Notation

The following Table 1 provides a list of some common notation used throughout the report. We do not distinguish between vectors and scalars, however matrices are denoted by uppercase letters.

Table 1: Table of Notation

Symbol	Description
X	\triangleq Random variable
\mathcal{X}	\triangleq Alphabet for the random variable X
x	\triangleq Realization of the random variable X
X^n	\triangleq Sequence of n random variables, (X_1, \dots, X_n)
x^n	\triangleq Sequence of n realizations, (x_1, \dots, x_n)
$Y^{(i)}$	\triangleq i th multiple-description reconstruction
$Y^{(i),n}$	\triangleq Sequence of n multiple-description reconstructions from the i th decoder, $(Y_1^{(i)}, \dots, Y_n^{(i)})$
$P(x)$	\triangleq Probability distribution of random variable X
$p(x)$	\triangleq Probability mass function (PMF) for discrete random variable X
$f(x)$	\triangleq Probability density function (PDF) for continuous random variable X
$E[\cdot]$	\triangleq Expectation operator
$\text{Var}[\cdot]$	\triangleq Variance operator
$E_X[\cdot]$	\triangleq Expectation w.r.t. the distribution on X
$E_{X,Y}[\cdot]$	\triangleq Expectation w.r.t. the joint distribution $P(x, y)$
$E_P[\cdot]$	\triangleq Expectation w.r.t. the distribution P
$X \perp Y$	\triangleq When X and Y are independent, i.e. when $P(x, y) = P(x)P(y)$
$X - Y - Z$	\triangleq When the random variables, X, Y, Z form a Markov chain. i.e. when $P(X, Z Y) = P(X Y)P(Z Y)$
$X _W - Y _W - Z _W$	\triangleq If conditioned upon W , $X - Y - Z$

1 | Zero-Delay Coding

In order to determine the fundamental performance limitations of ZDMD codes we first consider single-description ZD rate-distortion.

This chapter formally defines ZD source coding and extends upon the results in standard rate-distortion theory. Furthermore, we introduce the result and optimal test-channel realization of [8]. To this end we introduce directed information and predictive coding as needed. The test-channel of [8] provides the basis for our ZDMD test-channel realization. Finally we highlight some of the important differences between ZD source coding and source coding with arbitrary delay.

1.1 Zero-Delay Source Code

Let $X_k \in \mathcal{X}$ be a random variable, where \mathcal{X} is the alphabet of X_k . We denote a sequence of random variables by $X_j^k \triangleq (X_j, X_{j+1}, \dots, X_k)$, $k \in \mathbb{N}$, and their realizations by $x_j^k \in \mathcal{X}_j^k \triangleq \times_{i=j}^k \mathcal{X}_i$. For simplicity, if $j = 1$ we let $X_1^k = X^k$. Furthermore, let $\{X_k\}$, $X_k \in \mathcal{X}$, be the stochastic process modeling the unknown signal values for a specific source. We refer to this stochastic process as a source process. We also often refer to the source process $\{X_k\}$ as the source. For the source process $\{X_k\}$ we denote a realization of the source process by $\{x_k\}$, $x_k \in \mathcal{X}$.

Although our main results consider scalar-valued Gauss-Markov sources, we follow the results of [8] and consider vector-valued Gauss-Markov sources in this chapter. Particularly, we consider sources of the form,

$$X_k = AX_{k-1} + CW_k, \quad k \in \mathbb{N}, \quad (1.1)$$

where $A \in \mathbb{R}^{p \times p}$, $C \in \mathbb{R}^{p \times q}$ are deterministic matrices, $X_1 \sim \mathcal{N}(0, \Sigma_{X_1})$ is the initial state and $W_k \in \mathbb{R}^q \sim \mathcal{N}(0, \Sigma_W)$, $\Sigma_W = I$, is an independent and identically distributed (IID) Gaussian process independent of X_1 . Although the results of [8] consider X_k as being stationary in the limit of $k \rightarrow \infty$, we additionally assume the source is stationary at $k = 1$, i.e. Σ_{X_1} is the stationary covariance matrix of the process.

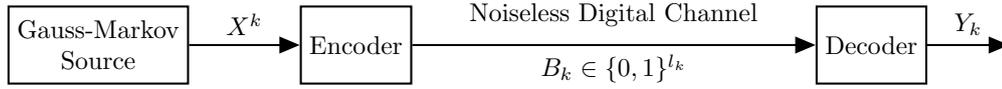


Figure 1.1: A zero-delay source coding scenario using variable length binary codewords. Inspired by [8].

We consider a ZD source coding system as depicted in Figure 1.1, where we assume the statistical properties of the vector Gauss-Markov source (1.1) are known. For every time step, $k \in \mathbb{N}$, the encoder observes a new source sample X_k , while assuming it has already observed the past X^{k-1} , and produces a single binary codeword B_k of length l_k (in bits) from a predefined set of codewords \mathcal{B}_k of at most a countable number of codewords [8]. Because the source is random, we may consider the codeword, B_k and its length l_k as random variables [8]. We assume no delay on the channel, thus the decoder receives the codeword B_k immediately. The decoder then produces an estimate Y_k of X_k , under the assumption that Y^{k-1} is already produced [8]. Both encoder and decoder are assumed to be zero-delay, i.e. they process information without sample delay.

Summarizing this, we formally define a ZD source code.

Definition 1.1 (Zero-delay source code[8])

For a discrete-time stationary source $\{X_k\}_{k \in \mathbb{N}}$, $X_k \in \mathcal{X}$, a zero-delay source code consists of a zero-delay encoder and -decoder.

For each time step k let \mathcal{B}_k be a predefined set of at most a countable number of codewords. The *zero-delay encoder* is specified by the sequence of functions, $\{f_k : k \in \mathbb{N}\}$, where

$$f_k : \mathcal{B}^{k-1} \times \mathcal{X}^k \rightarrow \mathcal{B}_k, \quad (1.2)$$

and at each time step $k \in \mathbb{N}$ the encoder outputs a message $B_k = f_k(\mathcal{B}^{k-1}, X^k)$ with length l_k (in bits), where $B_1 = f_1(X_1)$. The *zero-delay decoder* is specified by the sequence of functions $\{g_k : k \in \mathbb{N}\}$, where

$$g_k : \mathcal{B}^k \rightarrow \mathcal{Y}_k. \quad (1.3)$$

At each time step $k \in \mathbb{N}$ the decoder generates a reproduction $Y_k = g_k(\mathcal{B}^k)$ assuming Y^{k-1} has already been generated, with $Y_1 = g_1(B_1)$.

Both the encoder and decoder process information without delay.

The design goal of the ZD source code in Figure 1.4 is to achieve an asymptotic

average expected distortion that satisfies

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \mathbb{E} \left[\|X_k - Y_k\|_2^2 \right] \leq D, \quad (1.4)$$

where $D > 0$ is a predefined MSE distortion constraints [8].

The objective is to achieve this distortion while minimizing the average expected data-rate, i.e. the expected total number of bits receiver per symbol at the time the decoder reproduces $\{Y_k : k \in \mathbb{N}\}$ [8]. Denote by

$$L_n \triangleq \sum_{k=1}^n l_k \quad (1.5)$$

the total number of bits received by the decoder at the time it generates the estimate Y_n . Similar to the distortion we consider the asymptotic expected rate.

Definition 1.2 (ZD Rate)

Let l_k be the length in bits of the k 'th encoder output letter in a ZD source code, then the average expected data-rate of the scheme, measured in bits per source sample, is

$$R_{ZD} \triangleq \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \mathbb{E} [l_k] \quad (1.6)$$

$$= \lim_{n \rightarrow \infty} \frac{1}{n} \mathbb{E} [L_n]. \quad (1.7)$$

The definition of a ZD source code and its rate put great emphasis on the time steps compared to regular source codes with arbitrary delay. However, we note that the rate of a ZD source code and classical rate-distortion code (Definition C.1) is measured in the same way. To see this, we note that in both Definition C.1 and (1.5), L_n is total number of bits received when Y_n is reproduced.

To determine the fundamental performance limits of a ZD code, we first formally define the notion of an achievable rate for ZD source codes.

Definition 1.3 (ZD achievable rate)

The rate R_{ZD} of a zero-delay coding scheme is said to be *achievable* with respect to the MSE distortion constraint $D > 0$, if the asymptotic average expected distortion satisfies,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \mathbb{E} \left[\|X_k - Y_k\|_2^2 \right] \leq D. \quad (1.8)$$

The main problem of ZD rate-distortion theory is to determine the fundamental bound between the set of achievable and non-achievable rate-distortion points for a given source and distortion measure. That is, determine the minimum achievable rate with respect to a distortion constraint for a given source and distortion measure. This bound is called the *operational zero-delay RDF function*. We consider specifically Gaussian sources subject to MSE distortion constraints.

Definition 1.4 (Gaussian Operational ZD RDF [8])

For a discrete-time stationary Gaussian source process $\{X_k\}$, the operational zero-delay RDF, $R_{\text{ZD}}^{\text{op}}(D)$, is defined as the minimum achievable rate (1.6) with respect to the asymptotic MSE distortion constraint $D > 0$, where the infimum is over all possible zero-delay encoder- and -decoder sequences, $\{f_k\}_{k \in \mathbb{N}}, \{g_k\}_{k \in \mathbb{N}}$, such that (1.8) is satisfied. That is,

$$R_{\text{ZD}}^{\text{op}}(D) \triangleq \inf_{\substack{B_k = f_k(B^{k-1}, X^k), k \in \mathbb{N} \\ Y_k = g_k(B^k)}} \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \mathbb{E}[l_k], \quad (1.9)$$

s.t. $\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \mathbb{E}[\|X_k - Y_k\|_2^2] \leq D.$

Similarly to the classical RDF (Definition C.6), determining the operational ZD RDF as defined above is infeasible, since it is determined by a minimization over all possible operational zero-delay codes, $\{f_k\}, \{g_k\}$ [8].

However, the operational ZD RDF has been shown to be lower bounded by a minimization of the *directed information rate* from the source to the reproductions subject to the same distortion constraint [5]–[7], [16], [17]. Before considering this bound further, we introduce *directed information* as first defined by Massey in [21], and consider some importance results regarding information flow in systems with feedback. Readers familiar with directed information may skip directly to Section 1.3.

1.2 Directed Information

Directed information is defined using conditional mutual informations.

Definition 1.5 (Directed information [21])

The *directed information* from a sequence of random variables X^n to a sequence Y^n , is defined as

$$I(X^n \rightarrow Y^n) \triangleq \sum_{i=1}^n I(X^i; Y_i | Y^{i-1}). \quad (1.10)$$

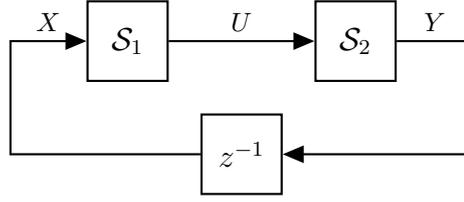


Figure 1.2: A general feedback system. The output of the second system, \mathcal{S}_2 , is fed back as the input to the first system \mathcal{S}_1 .

The directed information measures the amount of information that causally ‘flows’ from the ordered sequence X^n to Y^n . To understand the causal nature of directed information, we compare it to the mutual information (Definition A.9) between the two sequences X^n , Y^n ,

$$I(X^n; Y^n) = \sum_{i=1}^n I(X^n; Y_i | Y^{i-1}), \quad (1.11)$$

this follows from the chain rule of mutual information (A.22). Each term in the sum of (1.11) considers the amount of information about the *entire* sequence X^n present in Y_i , given the past values Y^{i-1} [34]. Contrary, in the directed information of (1.10) what matters is only the past and current values of X^n , i.e. X^i [34]. Hence, the directed information considers only the causally conveyed information from X^n to Y^n .

To further understand the importance of directed information in feedback systems, recall the general feedback system in Figure 1.2. This system has perfect causal feedback, such that $X_k = Y_{k-1}$, and the mutual information between the two sequences is

$$\begin{aligned} I(X^n; Y^n) &= H(Y^n) - H(Y^n | X^n) \\ &= H(Y^n) - H(Y^n | Y^{n-1}) \\ &= H(Y^n) - H(Y_n | Y^{n-1}) \\ &= H(Y^{n-1}). \end{aligned}$$

Hence, the mutual information, $I(X^n; Y^n)$, is unable to account for how much information about X^n has been conveyed to Y^n through the forward channel [34]. Therefore, when considering the information flow in feedback- and causal systems the directed information provides a more meaningful measure than the usual mutual information [21].

A few more important relationships exist between mutual- and directed information. However, for the sake of brevity these are referred to Appendix A.3.

For random process we define the directed information rate as the directed information per symbol for asymptotically long sequences.

Definition 1.6 (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

$$\bar{I}_\infty(X \rightarrow Y) \triangleq \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n I(X^i; Y_i | Y^{i-1}). \quad (1.12)$$

As a final remark, pointed out by Massey; causal dependence is quite different from probabilistic dependence [21]. Since, whether X causes Y , or Y causes X , then X and Y will be probabilistic dependent [21], i.e.

$$P(x, y) \neq P(x)P(y). \quad (1.13)$$

That is, there is no natural directivity in probabilistic dependence, underlining the symmetry of mutual information [21]. Hence, if X and Y are signals within a system with possible feedback, and $P(x_k, y_{k-1}) \neq P(x_k)P(y_{k-1})$, we cannot conclude a non-causal relationship, since feedback is causal. Therefore, care must be taken when making conclusions on causality in relation to probabilistic terms.

1.3 Lower Bound

In [5, Theo. 4.1] the directed information rate was shown to be a lower bound on the operational rate of a source coding scheme inside a feedback loop. Since source coding inside feedback loops requires ZD, we can state this result also for the ZD source coding setting in Figure 1.4,

$$R_{ZD} \geq \bar{I}_\infty(X \rightarrow Y). \quad (1.14)$$

We do not consider this result in detail here, since we extend this to the MD scenario in Chapter 3.

Now for Gaussian sources the directed information is further minimized by Gaussian reproductions [8], [18], [35] (see Lemma A.22). Then by the results of [8, sec. III] we arrive at the following definition of an *information theoretic ZD RDF*¹.

¹ We note, that several related or equivalent variants exist in the literature, e.g. non-anticipative RDF [8], sequential RDF [6], causal RDF [16]. However, we have chosen this name for consistency with classical rate-distortion theory.

Definition 1.7 (Gaussian information theoretic ZD RDF)

For the stationary vector-valued Gaussian source model in (1.1) with asymptotic MSE distortion constraint, $D > 0$, the Gaussian information theoretic ZD RDF is,

$$R_{\text{ZD,GM}}^I(D) \triangleq \inf_{\vec{Q}^{GP}(y^\infty|x^\infty)} \bar{I}_\infty(X \rightarrow Y), \quad (1.15)$$

$$\text{s.t.} \quad \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \mathbb{E} \left[\left\| X_k - Y_k^{(0)} \right\|_2^2 \right] \leq D$$

where $\vec{Q}^{GP}(y^\infty|x^\infty)$ denotes the sequence of Gaussian conditional *test-channel* distributions $\{P^{GP}(y_k|y^{k-1}, x^k) : k \in \mathbb{N}\}$.

Using this definition we have by [8, Theo. 1] the following lower bounds.

Lemma 1.8 (Inequalities [8, Theo. 1])

For Gaussian sources with asymptotic MSE distortion constraints, the following bounds hold.

$$R(D) \leq R_{\text{ZD,GM}}^I(D) \leq R_{\text{ZD}}^{\text{op}}(D), \quad (1.16)$$

where $R(D)$ is the classical RDF (Definition C.6).

One of the main results in [8] is a new feedback realization scheme of the optimal test-channel that achieves the lower bound in (1.15). This feedback realization scheme is one of the key components of the test-channel we introduce in Chapter 4. Before introducing the test-channel of [8] we give a short introduction to the predictive coding technique of *Differential Pulse Code Modulation* (DPCM). Readers familiar with DPCM may skip directly to Section 1.5.

1.4 Differential Pulse Code Modulation

The objective of DPCM is to convert the coding of dependent source samples into a series of independent encodings [36]. The time-dependence is removed by ways of prediction. Define the prediction error process

$$U_k \triangleq X_k - \hat{X}_{k|k-1}, \quad (1.17)$$

where $\hat{X}_{k|k-1} \triangleq \mathbb{E}[X_k|Y^{k-1}]$ is the prediction of X_k from all previously reconstructed values at time $k-1$, Y^{k-1} , we can then write the Gauss-Markov source process as

$$X_k = U_k + \hat{X}_{k|k-1}. \quad (1.18)$$

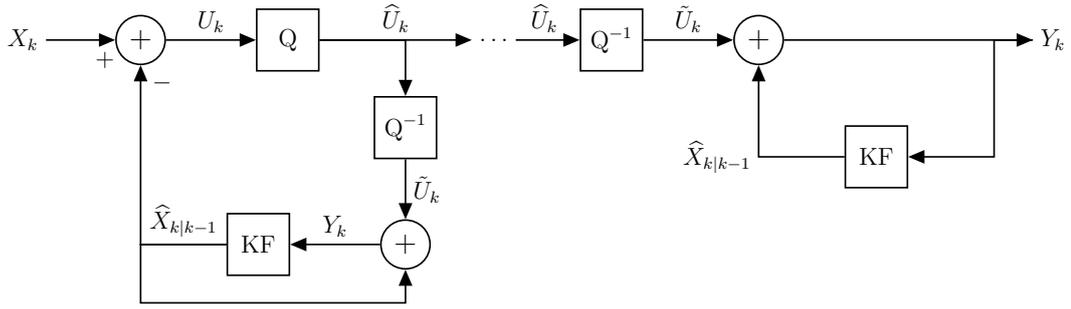


Figure 1.3: Block diagram for predictive coding. Figure inspired by [37, p.114].

The prediction error process U_k is an equivalent process to X_k [37]. Therefore, instead of coding the source process itself we code (quantize) the error process. A basic block diagram of DPCM is shown in Figure 1.3, where the process \hat{U}_k is the quantized version of U_k , and \tilde{U}_k is the reconstructed error process after decoding. We denote the prediction operation by the Kalman Filter (KF) block. At both the encoder and decoder, the reconstructed source process is then formed by adding the reconstructed error process to the prediction, i.e.

$$Y_k = \tilde{U}_k + \hat{X}_{k|k-1} \quad (1.19)$$

Combining (1.19) and (1.17) we see that

$$X_k - Y_k = U_k - \tilde{U}_k. \quad (1.20)$$

That is, the error between X_k and Y_k is equal to the quantization error introduced in U_k [37]. Especially, we have that the MSE distortion satisfy

$$D = E \left[(X_k - Y_k)^2 \right] = E \left[(U_k - \tilde{U}_k)^2 \right], \quad (1.21)$$

hence if we determine a coding scheme for the error process, U_k , that achieves distortion, D , this distortion is also achieved for the source process X_k .

We note that the process $\{U_k\}$ has some temporal correlation [8], since it is the error of predicting X_k from the quantized source Y^{k-1} , and not the infinite past of the clean source $X_{-\infty}^k$. Hence, $\{U_k\}$ is only an estimate of the true innovations process. We may consider DPCM as closed-loop prediction, and prediction of X_k from the clean source as open-loop prediction [36].

1.5 Test-Channel Realization

The test-channel realization of [8] builds upon the idea of DPCM. However, since we are in the test-channel domain, i.e. realization of the optimal Gaussian dis-

tribution, a pre-scaled *Additive White Gaussian Noise* (AWGN) is used instead of quantization.

Lemma 1.9 (Realization of $P^{GP}(y_k|y^{k-1}, x^k)$ [8, Lemma 1, Theo. 2, Theo. 3])

Consider the minimization problem in (1.15). Suppose either A or B in (1.1) are full rank. Then the following statements hold.

1. The test-channel $P^{GP}(y_k|y^{k-1}, x^k)$ is realized by the recursion

$$Y_k = HX_k + (I - H)AY_{k-1} + Z_k, \quad k \in \mathbb{N}, \quad (1.22)$$

where $Z_k \sim \mathcal{N}(0, \Sigma_Z)$,

$$H \triangleq I - \Pi\Lambda^{-1} \succeq 0, \quad \Pi \succeq 0, \quad \Lambda \succeq 0, \quad (1.23a)$$

$$\Sigma_Z \triangleq \Pi H^T \succeq 0, \quad (1.23b)$$

$$\Lambda = A\Pi A^T + CC^T, \quad (1.23c)$$

and

$$\Pi \triangleq \mathbb{E} \left[(X_k - Y_k)(X_k - Y_k)^T \right], \quad (1.24)$$

is the stationary reconstruction error covariance matrix.

2. The characterization of $R_{\text{ZD}}^I(D)$ is

$$\begin{aligned} R_{\text{ZD},GM}^I(D) &\triangleq \min_{\Pi} \frac{1}{2} \log \frac{|\Lambda|}{|\Pi|}, & (1.25) \\ \text{s.t.} & \quad 0 \prec \Pi \preceq \Lambda, \\ & \quad \text{tr}(\Pi) \leq D, \end{aligned}$$

where $\text{tr}(\cdot)$ is the trace operator.

We highlight some important features of this recursion realization. The test-channel realization of (1.22) is illustrated in Figure 1.4. Similar to the DPCM scheme, the encoder does not directly transmit X_k , instead it sends the zero mean prediction error process,

$$U_k \triangleq X_k - \mathbb{E} [X_k | Y^{k-1}] = X_k - AY_{k-1}, \quad (1.26)$$

with covariance matrix Λ , i.e.

$$\mathbb{E} [U_k U_k^T] = \mathbb{E} \left[(X_k - AY_{k-1})(X_k - AY_{k-1})^T \right] \quad (1.27)$$

$$= \mathbb{E} \left[A(X_{k-1} - Y_{k-1} + CW_k)(X_{k-1} - Y_{k-1} + CW_k)^T A^T \right] \quad (1.28)$$

$$= A\Pi A^T + CC^T, \quad (1.29)$$

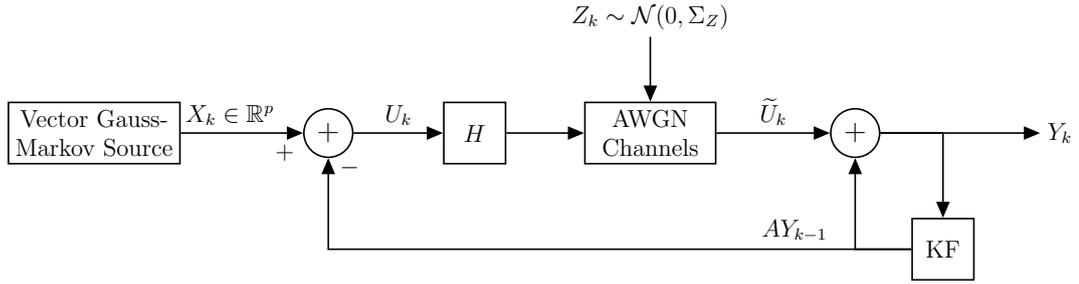


Figure 1.4: The optimum test channel of [8]. Figure adapted from [8].

The process \tilde{U}_k is given as

$$\tilde{U}_k = HU_k + Z_k, \quad (1.30)$$

with covariance matrix

$$\mathbb{E} [\tilde{U}_k \tilde{U}_k^T] = H\Lambda H^T + \Pi H^T. \quad (1.31)$$

By the scalar pre/post-scaled Gaussian test-channel (C.2) and (1.30), we may view \tilde{U}_k as the decoder estimate of U_k . This estimate is then combined with the prediction aY_{k-1} to yield an estimate Y_k of X_k . As in the DPCM scheme we have that

$$\Pi = \mathbb{E} [(X_k - Y_k)(X_k - Y_k)^T] = \mathbb{E} [(U_k - \tilde{U}_k)(U_k - \tilde{U}_k)^T]. \quad (1.32)$$

It was shown in [8] that the optimal solution of (1.25) can be determined by a semidefinite programming approach. This approach was originally explored in [17]. Furthermore, it was shown that the test-channel may be realized by a novel equivalent pre- and post-scaled realization scheme that reveals a reverse-waterfilling solution on the dimensions of U_k [8]. Finally, contrary to standard rate-distortion results where arbitrarily long sequences are coded, optimality for the Gaussian ZD RDF is shown for arbitrarily long vectors [8]. We do not elaborate further on these results, since we consider ZDMD coding of scalar Gauss-Markov sources in this report.

For stable and stationary scalar Gauss-Markov sources the Gaussian information theoretic ZD RDF may be derived directly from (1.25) [8]. This result has also been derived in e.g. [7], [16].

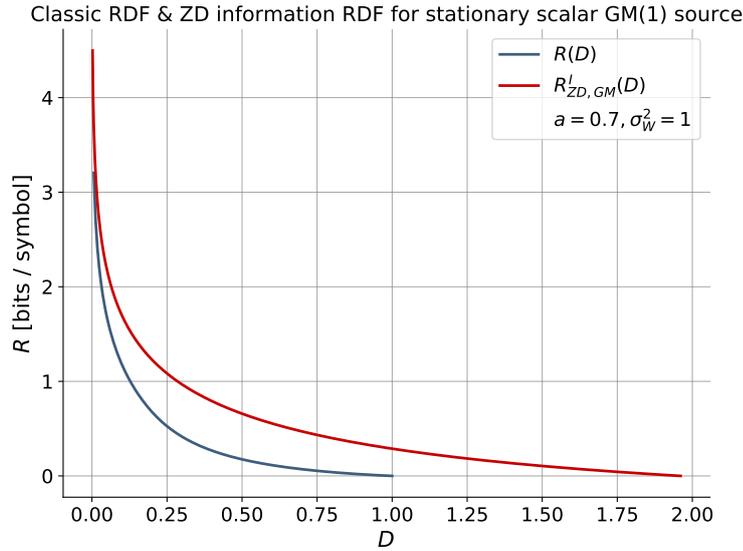


Figure 1.5: Classical non-causal RDF, $R(D)$, and information theoretic ZD RDF, $R_{ZD,GM}^I(D)$, for a stationary scalar GM(1) source with MSE distortion constraint.

Lemma 1.10 ($R_{ZD,GM}^I$ for scalar-valued GM(1) process [8])

Consider the scalar case of (1.1) with $A = a \in \mathbb{R}$ and $CW_k \sim \mathcal{N}(0, \sigma_W^2)$. Then $\Pi = D$ and $\Lambda = a^2D + \sigma_W^2$, and (1.25) becomes

$$R_{ZD,GM}^I(D) = \frac{1}{2} \log \left(\frac{a^2D + \sigma_W^2}{D} \right) = \frac{1}{2} \log \left(a^2 + \frac{\sigma_W^2}{D} \right) \quad (1.33)$$

For stationary and stable source, i.e. where $|a| < 1$, it was shown in [16] that $R_{ZD,GM}^I(D) = 0$ for $D \geq D_{\max} = \frac{\sigma_W^2}{1-a^2}$.

This last result implies, zero rate is transmitted when D is larger than the stationary variance of X_k , similar to the case of white Gaussian sources (Lemma C.9). Figure 1.5 illustrates the information theoretic ZD RDF, $R_{ZD,GM}^I(D)$ (1.33), and the classical RDF, $R(D)$ (Lemma C.11), for a stable stationary scalar Gauss-Markov source. We see that the classical non-causal code achieves a smaller distortion for the same rate compared to the ZD code. Therefore, the standard rate-distortion region only provides a conservative outer bound on the ZD rate-distortion region [17].

We consider this rate-gain for non-causal codes in the following section.

1.6 Non-Causal Rate-Gain

We consider how classical rate-distortion theory achieves a lower rate for the same distortion compared to ZD source coding.

Classical rate-distortion theory imposes no restrictions in terms of delay or causality on the source- encoders and decoders. Therefore, we also refer to classical source coding as non-causal source coding.

We note that the following discussion is often given in terms of a rate-loss suffered by ZD coding, see e.g. [8], [16]. However, we interpret the difference between ZD- and classical rate-distortion theory as a rate-gain, that may be obtained if we allow non-causal codes.

The rate-gain for Gaussian sources with memory due to non-causal source coding, i.e. the difference between the ZD operational RDF, $R_{\text{ZD}}^{\text{op}}(D)$, and $R(D)$ can be attributed to three factors [8], [16]: the space-filling loss of causal encoders, decreased distortion due to non-causal filtering, and entropy coding with memory.

Entropy Coding

We recall from lossless entropy coding [13, Ch.5] (see Section B.2.1), that for a discrete random variable $\hat{U} \in \hat{\mathcal{U}}$, e.g. the quantized source, that codes constructed based on Shannon coding give an instantaneous code with expected length that satisfies

$$H(\hat{U}) \leq \mathbb{E}[l] < H(\hat{U}) + 1. \quad (1.34)$$

That is, ZD source coding has an expected length at most one bit above entropy. Now if we allow non-causal coding we may wait and collect a sequence of random variables \hat{U}^n , and jointly entropy code this sequence and normalize by n , we get

$$\frac{H(\hat{U}^n)}{n} \leq \frac{\mathbb{E}[L_n]}{n} < \frac{H(\hat{U}^n)}{n} + \frac{1}{n}. \quad (1.35)$$

Hence, the excess one bit is eliminated for long sequences. Thus, by using memory and joint entropy coding non-causal source coding has a rate-gain of at most one bit compared to independent memoryless entropy coding in ZD codes.

Quantization

Test-channels provide constructive proofs for designing optimal source codes. By designing a coding scheme that generates quantization noise distributed as the AWGN. However, generating Gaussian noise is difficult.

A popular quantization scheme is *dithered* quantization [38]. We give a short introduction to dithered quantization, for further details we refer to [10, Ch. 4] for a great

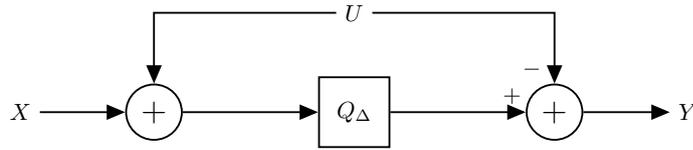


Figure 1.6: Dithered quantization. The encoder adds the dither, U , to the source sample, X , and quantizes the sum. The decoder subtracts the dither to produce the reconstruction, Y .

review.

A *subtractive dither* is a random variable U that is both known at the encoder and decoder in a lossy source code, and if the final reconstruction is given by [10, sec. 4.1],

$$Y = Q_\Delta(x + U) - U, \quad (1.36)$$

where $Q_\Delta(\cdot)$ is a scalar uniform quantizer with bin size Δ .

Figure 1.6 illustrates the dithered quantization scheme. The encoder adds the dither to the source samples prior to quantization. The decoder then subtracts this dither from the quantized value.

Dithered quantization may also be combined with DPCM, by replacing the quantizer and decoder inside the prediction loop in Figure 1.3 by a dithered quantization scheme.

For a random source X the quantization error can be shown to satisfy [10, Theo. 4.1],

$$[Q_\Delta(X + U) - U - X] \sim \text{Unif} \left[-\frac{\Delta}{2}, \frac{\Delta}{2} \right]. \quad (1.37)$$

That is, the quantization error is uniformly distributed across the fundamental cell of the quantizer. The above results may be extended to general *lattice quantizers*, e.g. vector quantization.

Now considering the *divergence* [13, Ch. 8] (Definition A.20) between a uniform random variable, U , and Gaussian random variable, Z , with equal second moments $\sigma_Z^2 = \sigma_U^2 = \frac{\Delta^2}{12}$, we have from Lemma A.21,

$$D(U||Z) = h(U) - h(Z) \quad (1.38)$$

$$= \frac{1}{2} \log(\Delta^2) - \frac{1}{2} \log\left(2\pi e \frac{\Delta^2}{12}\right) \quad (1.39)$$

$$= \frac{1}{2} \log\left(\frac{\pi e}{6}\right) \approx 0.254 \text{ bit}, \quad (1.40)$$

where the base of the logarithm² is 2. This is the so-called space-filling loss [16]. In terms of distortion the space-filling loss translates to approximately 1.5 dB [10, sec. 5.5.3]. This shows, practical source coding schemes that use independent scalar

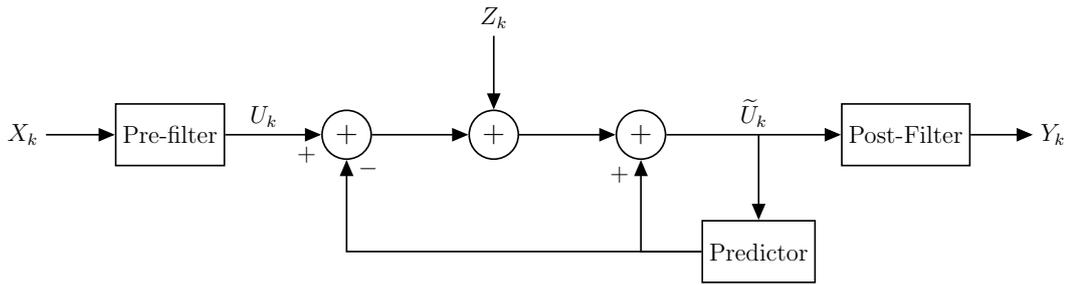


Figure 1.7: Pre- and post-filtered predictive coding test-channel of [36], for a stationary Gaussian source, X_k , with reconstruction Y_k . In the test-channel Z_k is a white Gaussian process. Figure modified from [36].

quantization suffers a rate-loss of 0.254 bit compared to the theoretical performance limits $R(D)$ [10]. However, by using vector quantization this loss may be reduced, since in high dimensions the quantization noise “looks” Gaussian, in a divergence sense [10], [39], [40]. Therefore, in non-causal source coding by using memory we may reduce this loss, by vector quantization of sequences of random variables. That is, instead of quantizing each source sample individually, a non-causal quantizer jointly quantizes long sequences of source samples, and thereby eliminates most of the space-filling loss.

Filtering

It was shown in [41], that the non-causal RDF of a stationary Gaussian process, $R(D)$ (Lemma C.11), cannot be achieved by encoding its innovation process directly, i.e. by open-loop prediction [36]. However, recently [36] showed, that the RDF can be achieved by a pre- and post-filtered DPCM scheme with entropy coding and a MSE-optimal closed-loop prediction filter. The pre- and post filters facilitate reverse water-filling by attenuating frequency components with amplitude above a “noise-floor” to zero [42], see also Appendix C.4.1 for the reverse water-filling solution. The optimal test-channel realization scheme is shown Figure 1.7, where the quantization is replaced by an AWGN channel.

The result of [36] shows, that we may achieve the RDF of a Gaussian source with memory subject to a MSE distortion constraint by appropriate selection of three filters and an AWGN channel [36]. The central prediction filter is sequential and hence causal [36]. However, for the test-channel to be optimal at least one of the pre- and post filters must be non-causal [36]. Thus, they require delay in practical implementations. Especially, if the pre-filter is causal then the post-filter must be non-causal, or vice-versa [36]. Thus, in ZD coding these filters can no longer be optimal. This results in an increased rate for the same distortion, when using

²Throughout the report we take the base of logarithms to be 2, unless otherwise specified.

causal coders.

In the essence of Kalman Filtering [43, Ch. 6], if we allow non-causal filters the decoder may perform *smoothing*. That is, use both future and past samples, $\tilde{U}^{k'}$, $k' > k$ to estimate X_k . In a causal filter, the decoder can only estimate X_k using the current and past samples, \tilde{U}^k . Thus, the non-causal filter achieves a smaller MSE than the causal filter. This decrease in distortion for the same rate, may also be interpreted as a rate-gain at a different distortion.

With this introduction to ZD source coding and the recent results of [8], we next consider MD coding in more detail before presenting our contributions on ZDMD coding.

2 | Multiple-Description Coding

In this chapter we first revisit MD source coding and the MD rate-distortion region without delay restrictions. We introduce the fundamental results of El-Gamal and Cover [24] and Ozarow [25]. Particularly we introduce the novel test-channel of Ozarow [25], this will play an important role in our test-channel for ZDMD coding. We then present our first new contribution; the combined definition of ZDMD source coding, which is not available in the literature. Furthermore, we present the main problem of determining the ZDMD rate-distortion region.

2.1 Multiple-Description Source Code

The two-description MD source coding problem as illustrated in Figure 2.1 considers the communication of a single source process $\{X_k\}$ to three receivers over two noiseless digital channels [22]. The central decoder, (Decoder 0), receives the information sent over both channels, while the remaining side decoders (Decoder 1 and -2) receive only the information over their respective channels [22]. The reconstruction process produced by decoder i is denoted $\{Y_k^{(i)}\}$, $i = 0, 1, 2$, and the associated distortions by D_i $i = 0, 1, 2$ [22]. Furthermore, each channel is associated with a rate, R_i (in bits per source symbol).

As in classical rate-distortion theory the MD source coding problem considers determining the minimum bitrates required to achieve given distortion constraints on the three reproductions.

In standard MD rate-distortion theory, we are not concerned with delays. Hence, similar to classical rate-distortion theory we consider the encoding of n -blocks, x^n . Thus, the formal definition of a MD source code follows as an extension of the classical rate-distortion code (Definition C.1).

Definition 2.1 (Multiple-description source code[44])

For an n -block, X^n , from a discrete-time stationary source $X_k \in \mathcal{X}$, a multiple-description source code consists of an encoder and three decoders.

For each $n \in \mathbb{N}$ let $\mathcal{B}_n^{(i)}$ $i = 1, 2$ be two predefined sets of at most a countable

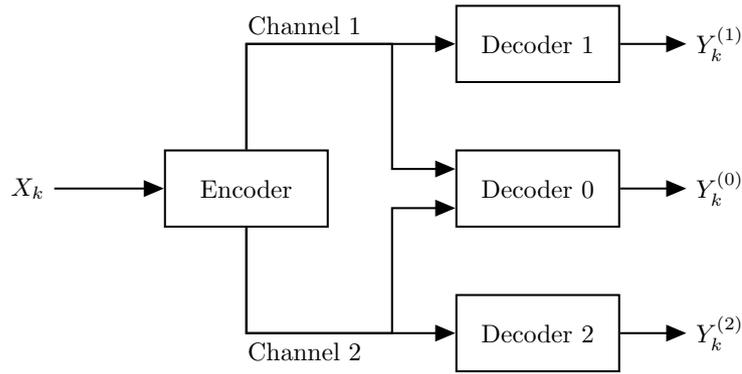


Figure 2.1: The MD source coding scenario with two channels and three receivers. The encoder transmits information across two noiseless digital channels. The central decoder receives information transmitted across both channels, the side decoders receive only the information across their respective channel.

number of codewords. The encoder is specified by the encoding functions

$$f_n^{(i)} : \mathcal{X}^n \rightarrow \mathcal{B}_n^{(i)}, \quad i = 1, 2. \quad (2.1)$$

The encoder outputs two messages $B_n^{(i)} = f_n^{(i)}(X^n)$ with lengths $L_n^{(i)}$ (in bits). The decoders are specified by the decoding functions

$$g_n^{(i)} : \mathcal{B}_n^{(i)} \rightarrow \mathcal{Y}^{(i),n}, \quad i = 1, 2, \quad (2.2a)$$

$$g_n^{(0)} : \mathcal{B}_n^{(1)} \times \mathcal{B}_n^{(2)} \rightarrow \mathcal{Y}^{(0),n}, \quad (2.2b)$$

where $\mathcal{Y}^{(i)}$, $i = 0, 1, 2$ are the reproduction alphabets.

The three resulting reconstructions are,

$$Y^{(i),n} = g_n^{(i)} \left(f_n^{(i)}(X^n) \right), \quad i = 1, 2, \quad (2.3)$$

$$Y^{(0),n} = g_n^{(0)} \left(f_n^{(1)}(X^n), f_n^{(2)}(X^n) \right). \quad (2.4)$$

Since the MD source code produces two descriptions, we may associate the MD with two rates defined analogously to the single description case (Definition C.3).

Definition 2.2 (MD Marginal Rates)

For a stationary discrete-time source $\{X_k\}$, the marginal rates associated with MD source code, measured in bits per source symbol, are

$$R_i \triangleq \lim_{n \rightarrow \infty} \frac{1}{n} \mathbb{E} \left[L_n^{(i)} \right], \quad i = 1, 2. \quad (2.5)$$

We often refer to a MD source code with marginal rates R_1, R_2 as a rate- (R_1, R_2) MD code [23].

For the central reconstruction $Y^{(0)}$ in a MD source code, the rate of the code is the total rate of the two descriptions. We call this central rate the *sum-rate*.

Definition 2.3 (Sum-rate)

For a rate- (R_1, R_2) MD source code the *sum-rate*, is the sum of the two marginal rates, R_1, R_2 , i.e. $R_1 + R_2$.

In MD rate-distortion theory we also define the notion of an achievable rate pair in relation to a MSE distortion constraint.

Definition 2.4 (MD achievable rate pair[23])

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 for sufficiently large n , there exists a rate- (R_1, R_2) MD coding scheme (2.1)–(2.4), such that the asymptotic average expected distortions satisfy

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \mathbb{E} \left[\|X_k - Y_k^{(i)}\|_2^2 \right] \leq D_i, \quad i = 0, 1, 2. \quad (2.6)$$

The main problem of MD rate-distortion theory is to determine the fundamental bound between the set of achievable- and non-achievable rates for a given source and distortion measure subject to given distortion constraints. Since, we consider multiple rates and distortions, we distinguish more clearly between a rate-region and distortion-region.

Definition 2.5 (MD rate-region [14])

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

Similar to classical rate-distortion theory we may define a MD *distortion-region*, $\mathcal{D}_X(R_1, R_2, D_0, D_1, D_2)$ as the set of all achievable distortion triplets for a given rate pair (R_1, R_2) [14]. The rate- and distortion regions may be considered as inverses to each other [14]. Therefore, we often just refer to the MD-region for a given

source and distortion measure.

As mentioned in the introduction, the MD rate-region has only been completely determined in few cases [23]. We show the important result of Ozarow for white Gaussian sources. Ozarow's test-channel has been instrumental in development of MD rate-distortion theory, and proves useful for our test-channel design.

2.2 MD Coding of White Gaussian Sources

Before stating the result of [25] we consider the early result of El-Gamal and Cover [24], who determined an outer bound on the achievable rate-region for the coding of white scalar Gaussian sources subject to MSE distortion constraints.

Definition 2.6 (El-Gamal and Cover region [24], [44])

For a scalar white Gaussian source $X \sim \mathcal{N}(0, \sigma_X^2)$, let $(U^{(1)}, U^{(2)})$ be any pair of random variables arbitrarily jointly distributed given X via $P(u^{(1)}, u^{(2)}|x)$. Then an achievable rate-region for the MSE distortion constraints $D_0, D_1, D_2 > 0$ is given by the convex closure of all (R_1, R_2) such that

$$R_1 \geq I(X; U^{(1)}), \quad (2.7a)$$

$$R_2 \geq I(X; U^{(2)}), \quad (2.7b)$$

$$R_1 + R_2 \geq I(X; U^{(1)}, U^{(2)}) + I(U^{(1)}; U^{(2)}), \quad (2.7c)$$

and

$$\mathbb{E} \left[\left(X - \mathbb{E} \left[X | U^{(i)} \right] \right)^2 \right] \leq D_i, \quad i = 1, 2 \quad (2.8)$$

$$\mathbb{E} \left[\left(X - \mathbb{E} \left[X | U^{(1)}, U^{(2)} \right] \right)^2 \right] \leq D_0. \quad (2.9)$$

This region establishes an information-theoretic outer bound on the MD region of white Gaussian sources. We also refer to this as a lower bound, since it is given in terms of lower bounds on the marginal rates and the sum-rate.

Similar to classical rate-distortion theory, achievability is established by determining the optimum test-channel.

The region in Definition 2.6 is not necessarily convex [24]. However, it is possible to *convexify* the region by *time-sharing* [24]. As a simple example, let (R_1, R_2) and (R'_1, R_2) be two achievable rate-pairs, then it is possible to achieve any pair between

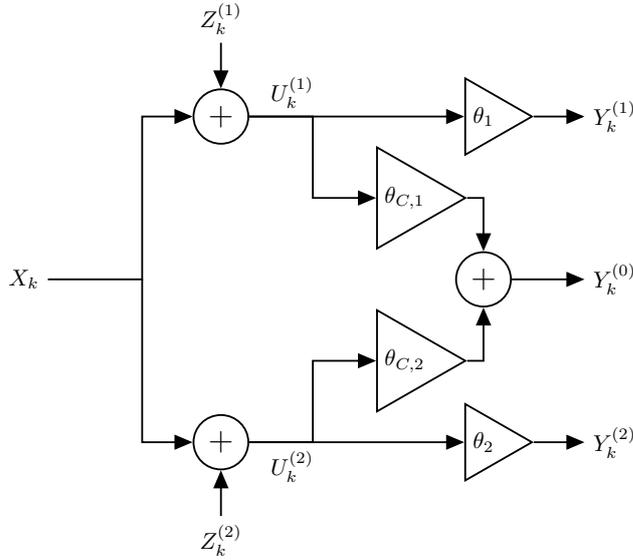


Figure 2.2: Ozarow's double branch test-channel [25]. The channel noises $Z^{(1)}$, $Z^{(2)}$ are jointly zero-mean Gaussian.

them by time-sharing between the pairs. That is, let the first encoder operate with rate R_1 a fraction γ ($0 \leq \gamma \leq 1$) of the time, and the with the other rate R'_1 the remaining $1 - \gamma$ fraction of the time. Then the average rate of the first encoder is $\gamma R_1 + (1 - \gamma)R'_1$, and it possible to achieve any average rate between R_1 and R'_1 by proper choice of γ [45].

2.2.1 Ozarow's Test-Channel

To show the tightness, i.e. achievability of the lower bound, Ozarow [25] considered the double-branch test-channel depicted in Figure 2.2. Here, the white Gaussian source, $X \sim \mathcal{N}(0, \sigma_X^2)$, is the input to two zero-mean AWGN channels. The output of these channels, $U^{(i)} = X + Z^{(i)}$, provide the two descriptions of the source, where $Z^{(i)}$ is the noise on the i th channel. The channel noises are independent of X with joint distribution

$$\begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix} \sim \mathcal{N} \left(0, \begin{bmatrix} \sigma_{Z^{(1)}}^2 & \rho \sigma_{Z^{(1)}} \sigma_{Z^{(2)}} \\ \rho \sigma_{Z^{(1)}} \sigma_{Z^{(2)}} & \sigma_{Z^{(2)}}^2 \end{bmatrix} \right). \quad (2.10)$$

Where the correlation coefficient $\rho \in [-1, 1]$. Hence, covariance between the descriptions is $E[U^{(1)}U^{(1)}] = \sigma_X^2 + \rho \sigma_{Z^{(1)}} \sigma_{Z^{(2)}}$. The decoder source reproductions, $Y^{(i)}$ $i = 0, 1, 2$, are given as minimum MSE (MMSE) estimates of X given either $U^{(1)}$, $U^{(2)}$ or both. Since X and $U^{(i)}$, $i = 1, 2$ are Gaussian, these are linear estimates [10]. Thus, the post scaling coefficients, θ_i , $i = 1, 2$, and $\theta_{C,i}$ $i = 1, 2$ are the

appropriate Wiener coefficients [10, Ch. 4][25].

Ignoring the Wiener coefficients the resulting MSE distortion by using the average of the two descriptions, $U^{(1)}, U^{(2)}$, is,

$$\mathbb{E} \left[\left(X - \frac{1}{2} (U^{(1)} + U^{(2)}) \right)^2 \right] = \frac{1}{4} \mathbb{E} \left[\left(Z^{(1)} + Z^{(2)} \right)^2 \right] = \frac{1}{4} (\sigma_{Z^{(1)}}^2 + \sigma_{Z^{(2)}}^2 + 2\rho\sigma_{Z^{(1)}}\sigma_{Z^{(2)}}). \quad (2.11)$$

Clearly, by selecting $\rho < 0$, it is possible to reduce the MSE. Particularly, in the symmetric case of $\sigma_{Z^{(1)}} = \sigma_{Z^{(2)}}$, choosing $\rho = -1$ yields zero MSE distortion [46]. This justifies the idea of always choosing a negative correlation [46].

The test-channel also provides a constructive proof for designing optimal source codes by replacing the additive Gaussian noise with a quantization scheme, that generates quantization noises distributed approximately as $Z^{(i)}$, $i = 1, 2$ [46]. However, it is a non-trivial task to achieve high negative correlation [46]. The main focus of this paper is not designing practical quantizers, therefore we do not consider this problem any further here.

Now, using the optimal Wiener coefficients

$$\theta_i = \frac{\sigma_X^2}{\sigma_X^2 + \sigma_{Z^{(i)}}^2}, \quad i = 1, 2 \quad (2.12)$$

$$\theta_{C,i} = \sigma_X^2 \frac{\sigma_{Z^{(i)}}^2 - \rho\sigma_{Z^{(1)}}\sigma_{Z^{(2)}}}{\sigma_{Z^{(1)}}^2\sigma_{Z^{(2)}}^2(1 - \rho^2) + \sigma_X^2(\sigma_{Z^{(1)}}^2 + \sigma_{Z^{(2)}}^2 - 2\rho\sigma_{Z^{(1)}}\sigma_{Z^{(2)}})}, \quad i = 1, 2, \quad (2.13)$$

and evaluating the various MMSE distortions we have [25],

$$D_i = \frac{\sigma_X^2\sigma_{Z^{(i)}}^2}{\sigma_X^2 + \sigma_{Z^{(i)}}^2}, \quad i = 1, 2, \quad (2.14a)$$

$$D_0 = \frac{\sigma_X^2\sigma_{Z^{(1)}}^2\sigma_{Z^{(2)}}^2(1 - \rho^2)}{\sigma_{Z^{(1)}}^2\sigma_{Z^{(2)}}^2(1 - \rho^2) + \sigma_X^2(\sigma_{Z^{(1)}}^2 + \sigma_{Z^{(2)}}^2 - 2\sigma_{Z^{(1)}}\sigma_{Z^{(2)}}\rho)}. \quad (2.14b)$$

Using these equations to determine the optimal noise variances and correlation, and calculating the mutual informations in the El-Gamal and Cover region (2.7), the full characterization of the MD region for white Gaussian sources is given in [25], [44], [47], and stated in the following lemma.

2.2.2 White Gaussian MD Rate-Region

Lemma 2.7 (Scalar White Gaussian MD region [25], [44], [47])

For a white Gaussian source, $X_k \sim \mathcal{N}(0, \sigma_X^2)$. The achievable MD region is given by the set of quintuples $(R_1, R_2, D_0, D_1, D_2)$ that satisfy

$$R_i \geq \frac{1}{2} \log \frac{\sigma_X^2}{D_i}, \quad i = 1, 2 \quad (2.15)$$

$$R_1 + R_2 \geq \frac{1}{2} \log \frac{\sigma_X^2}{D_0} + \frac{1}{2} \log \psi(\sigma_X^2, D_0, D_1, D_2), \quad (2.16)$$

where

$$\psi(\sigma_X^2, D_0, D_1, D_2) = \begin{cases} 1, & \text{if } D_0 < D_1 + D_2 - \sigma_X^2 \\ \frac{\sigma_X^2 D_0}{D_1 D_2}, & \text{if } D_0 > \left(\frac{1}{D_1} + \frac{1}{D_2} - \frac{1}{\sigma_X^2}\right)^{-1} \\ \frac{\sigma_X^2 (\sigma_X^2 - D_0)^2}{D_0 \left((\sigma_X^2 - D_0)^2 - \left(\sqrt{(\sigma_X^2 - D_1)(\sigma_X^2 - D_2)} - \sqrt{(D_1 - D_0)(D_2 - D_0)} \right)^2 \right)}, & \text{otherwise} \end{cases} \quad (2.17)$$

The two first cases of (2.17) yield so-called “degenerate” distortions and trivial lower bounds on the sum-rate [44]. Before considering these special cases we focus on the last non-degenerate case.

For the non-degenerate distortions, by isolating ρ in (2.14b) the lower bound on the sum-rate can be expressed as,

$$R_1 + R_2 \geq \frac{1}{2} \log \frac{\sigma_X^4}{D_1 D_2} - \frac{1}{2} \log(1 - \rho^2). \quad (2.18)$$

This expression allows for a more intuitive interpretation of the bound. Following (2.14b) as $\rho \rightarrow -1$ we can achieve a smaller central distortion. However, by (2.18), this decrease in distortion comes at the price of a higher sum-rate.

Figure 2.3 shows the MD rate-region for a white Gaussian source subject to MSE distortion constraints. The figure shows the outer bound (black line) to the set of achievable rates for given distortion constraints D_0, D_1, D_2 . The horizontal and vertical straight line indicate the individually optimum rates of R_1 and R_2 respectively. The linear relationship between R_1 and R_2 (dashed red line) is called the *dominant face* of the region. All rate pairs on the dominant face are jointly optimum, i.e. the sum-rate, $R_1 + R_2$, is minimum and constant, thus the slope is 45 degrees. Any pair inside the region is inferior to all rate pairs on the dominant face in terms of compression ratio (minimal rates) [44]. Therefore, to determine an optimum

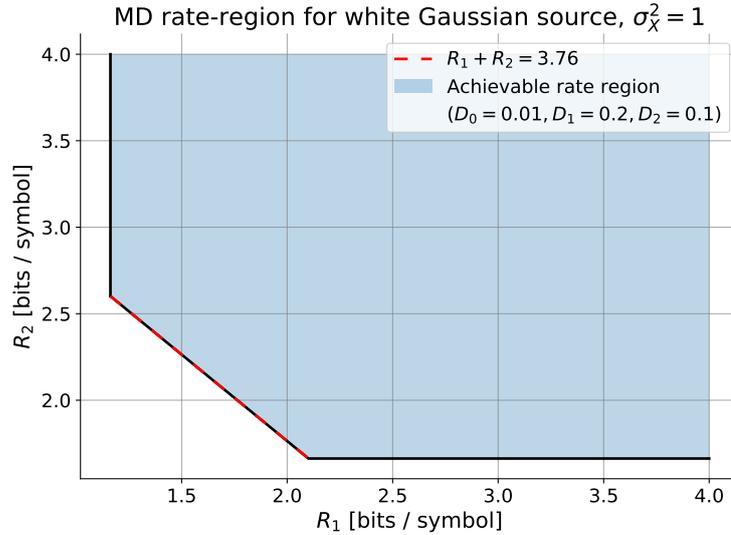


Figure 2.3: The MD rate-distortion region for a white Gaussian source with variance $\sigma_X^2 = 1$ for given distortions D_1, D_2, D_0 . The bound on the MD rate-region in black and red. Any rate on the "inside" is achievable.

source code the search can be restricted to pairs on the dominant face without loss of generality [27]. We can trade-off between the two individual description rates on the dominant face while always achieving the desired distortions. This is useful if e.g. each channel is associated with a cost (e.g. power consumption). Then we can allocate more bits to the cheaper link. The shaded achievable rate region indicate all achievable rate pairs. That is, for any pair of rates (R_1, R_2) inside the shaded area there exists a source coding scheme that can achieve the quintuple $(R_1, R_2, D_0, D_1, D_2)$.

If we are given a fixed rate budget, i.e. the sum-rate is fixed, (2.18) shows how we may trade-off between the bits spend on joint refinement and on the side descriptions. That is, by increasing the amount of bits allocated to refinement in the central decoder, we may achieve a smaller central distortion, at the cost of a higher side distortion. This trade-off is illustrated in Figure 2.4 for the symmetric case of $D_1 = D_2 = D_S$. As the maximum allowed central distortion is decreased, the minimum achievable side distortion increases, and vice versa. At the end points, i.e. maximum side- or central distortion we are the bounds of non-degenerate distortions.

This possible distortion trade-off is useful when designing a source code for a particular channel. Let each side channel have a packet-loss probability, p , i.e. either description is lost with probability p . If both descriptions are lost we assume the packets are retransmitted, but other error concealment methods are possible [33].

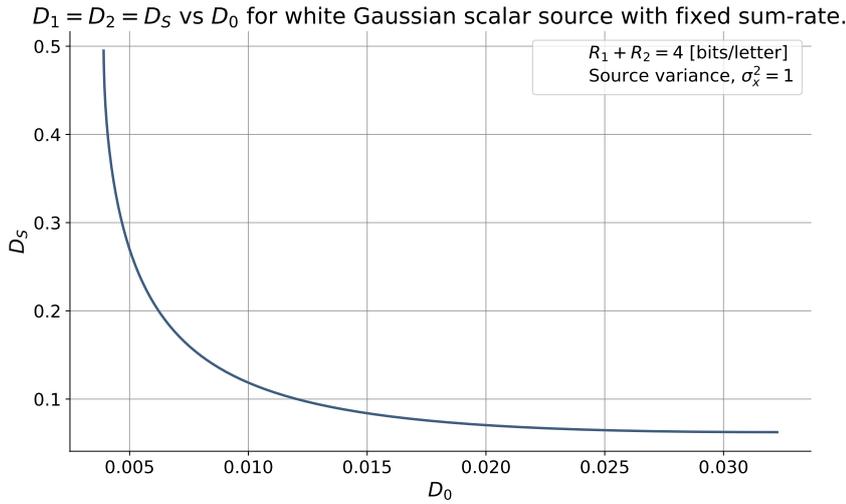


Figure 2.4: Distortion trade-off curve for a scalar white Gaussian source with fixed sum-rate and symmetric rates $R_1 = R_2$, thus $D_1 = D_2 = D_S$.

Now if one description is lost the resulting distortion is D_S , and if both are received the resulting distortion is D_0 . Then the average distortion, D_{av} , is [33]

$$D_{av} = \frac{(1-p)^2}{1-p^2} D_0 + \frac{2p(1-p)}{1-p^2} D_S. \quad (2.19)$$

Thus we can minimize the average (expected) distortion by optimizing the distortion/rate allocation according to the given packet-loss probability and sum-rate constraint. As p increases it is optimal to reduce D_S by allocating more bits to the side descriptions instead of refinement [33].

Degenerate Distortions

As mentioned, the rate of the central decoder and hence central distortion is the sum-rate. Therefore, for the central distortion constraint, D_0 , the trivial lower bound on the sum-rate is $R(D_0)$, since $R(D_0)$ is the minimum required rate for any source code (with- or without multiple descriptions) to achieve distortion D_0 [44]. Therefore, the first case of (2.17) is known as *no excess sum-rate*. That is, no excess rate is spent beyond the trivial lower bound on the sum-rate.

The second case of (2.17) is called *no excess marginal rate*. Since no excess rate is spent beyond the trivial lower bounds on the marginal rates, $R(D_i)$, $i = 1, 2$. Here the dominant face of the MD region generates to a single point [44].

Here $\frac{1}{D_0} + \frac{1}{\sigma_x^2} - \frac{1}{D_1} - \frac{1}{D_2}$ becomes strictly positive as D_0 approaches and grows beyond $\min(D_1, D_2)$.

For $D_0 \geq \min(D_1, D_2)$, the lower bound on the sum-rate is obvious, since if the central distortion constraint is larger than either of the side constraints there is no reason to use a central decoder. Thus, we do not need to pay the price of higher sum-rate due to added correlation between descriptions.

This lower bound also takes effect if D_0 is not sufficiently smaller than $\min(D_1, D_2)$. To see this, consider the symmetric case of (2.14). Here $D_0 < D_S$ even for $\rho = 0$. Thus, we can achieve a smaller central distortion $D_0 < D_1, D_2$ without spending any excess rate on correlation.

Going forward it can be assumed without loss of generality (WLOG) that we are in the non-degenerate case [44].

2.3 Symmetric Distortions

The MD rate-region is often given in terms of bounds on the two marginal rates, R_1 , R_2 , and the sum-rate. However, in the symmetric case of $R_1 = R_2 = R$ and $D_1 = D_2 = D_S$ we can define an MD equivalent to the single description RDF.

Definition 2.8 (Symmetric MD RDF [23])

The *symmetric MD RDF* of the stationary source $\{X_k\}$ is the minimum rate R per description, that is achievable with respect to the MSE distortions (D_0, D_S) , i.e.

$$R(D_0, D_S) \triangleq \min_{\text{s.t.}} R \quad (R_1, R_2) = (R, R) \in \mathcal{R}_X(R_1, R_2, D_0, D_S, D_S). \quad (2.20)$$

Notice that in the symmetric case the sum-rate is equal to two times the symmetric rate R , i.e.

$$R_1 + R_2 = 2R. \quad (2.21)$$

Thus in the symmetric case, in order to determine an outer bound on the MD region it suffices to determine a lower bound on the sum-rate. Leading to the following expression of an operational MD RDF.

Definition 2.9 (Operational symmetric MD RDF)

The *operational symmetric MD RDF* of the source $\{X_k\}$ is the minimum rate per description, that is achievable with respect to the MSE distortions (D_0, D_S) , i.e.

$$\begin{aligned} R^{\text{op}}(D_0, D_S) &\triangleq \inf \lim_{n \rightarrow \infty} \frac{1}{2n} \left(\mathbb{E} \left[L_n^{(1)} \right] + \mathbb{E} \left[L_n^{(2)} \right] \right) \\ \text{s.t.} \quad &\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \mathbb{E} \left[\left\| X_k - Y_k^{(0)} \right\|_2^2 \right] \leq D_0 \\ &\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \mathbb{E} \left[\left\| X_k - Y_k^{(i)} \right\|_2^2 \right] \leq D_S, \quad i = 1, 2, \end{aligned} \quad (2.22)$$

where the infimum is over possible MD encoder- and -decoder sequences $\{f_n^{(1)}\}_{n \in \mathbb{N}}$, $\{f_n^{(2)}\}_{n \in \mathbb{N}}$, $\{g_n^{(0)}\}_{n \in \mathbb{N}}$, $\{g_n^{(1)}\}_{n \in \mathbb{N}}$, $\{g_n^{(2)}\}_{n \in \mathbb{N}}$.

With this introduction of MD source coding in place, we formally present our contributions by first defining zero-delay multiple-description (ZDMD) source coding.

2.4 Zero-Delay Multiple-Description Coding

Combining the definition of a ZD source code (Definition 1.1) and a MD source code (Definition 2.1) we arrive at the following novel definition.

Definition 2.10 (Zero-Delay Multiple-Description Source Code)

For a discrete-time stationary source $\{X_k\}_{k \in \mathbb{N}}$, $X_k \in \mathcal{X}$, a zero-delay multiple-description (ZDMD) source code consists of a zero-delay encoder and three zero-delay decoders.

For each time step $k \in \mathbb{N}$ let $\mathcal{B}_k^{(i)}$ $i = 1, 2$ be two predefined sets of at most a countable number of codewords.

The zero-delay encoder is specified by the two sequences of functions $\{f_k^{(1)}, f_k^{(2)} : k \in \mathbb{N}\}$, where

$$f_k^{(i)} : \mathcal{B}^{(i), k-1} \times \mathcal{X}^k \rightarrow \mathcal{B}_k^{(i)}, \quad i = 1, 2, \quad (2.23)$$

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

$$B_k^{(i)} = f_k^{(i)} \left(\mathcal{B}^{(i), k-1}, X^k \right), \quad i = 1, 2, \quad (2.24)$$

with lengths $l_k^{(i)}$ (in bits), where $B_1^{(i)} = f_1^{(i)}(X_1)$.

The zero-delay decoders are specified by the three sequences of functions

$\{g_k^{(0)}, g_k^{(1)}, g_k^{(2)} : k \in \mathbb{N}\}$, where

$$g_k^{(i)} : \mathcal{B}^{(i),k} \rightarrow \mathcal{Y}_k^{(i)}, \quad i = 1, 2 \quad (2.25)$$

$$g_k^{(0)} : \mathcal{B}^{(1),k} \times \mathcal{B}^{(2),k} \rightarrow \mathcal{Y}_k^{(0)}. \quad (2.26)$$

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

$$Y_k^{(i)} = g_k^{(i)} \left(B^{(i),k} \right), \quad i = 1, 2 \quad (2.27)$$

$$Y_k^{(0)} = g_k^{(0)} \left(B^{(1),k}, B^{(2),k} \right), \quad (2.28)$$

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

$$Y_1^{(i)} = g_1^{(i)} \left(B_1^{(i)} \right), \quad i = 1, 2$$

$$Y_1^{(0)} = g_1^{(0)} \left(B_1^{(1)}, B_1^{(2)} \right).$$

All encoders and decoders process information without delay.

Combining the definitions of rates for ZD- (Definition 1.2) and MD source codes (Definition 2.2), we may define the rate pair of a ZDMD source code.

Definition 2.11 (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. Then the average expected data-rate pair, $(R_{ZD,1}, R_{ZD,2})$, measured in bits per source samples, are the rates

$$R_{ZD,i} = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \mathbb{E} \left[l_k^{(i)} \right], \quad i = 1, 2. \quad (2.29)$$

We suppress the ZD notation on the rates, when it is clear from the context, that we refer to ZDMD coding.

As in both ZD- and MD source coding we define the notion of achievable rates.

Definition 2.12 (ZDMD achievable rate pair)

A rate pair (R_1, R_2) of a ZDMD coding source code is said to be achievable with respect to the MSE distortion constraints $D_i > 0$, $i = 0, 1, 2$, if the asymptotic average expected distortions per source symbol satisfy

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \mathbb{E} \left[\left\| X_k - Y_k^{(i)} \right\|_2^2 \right] \leq D_i, \quad i = 0, 1, 2. \quad (2.30)$$

This immediately leads to a generalization of the MD rate-region (Definition 2.5).

Definition 2.13 (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 ZD rate pairs (R_1, R_2) with respect to the MSE distortion constraints (D_0, D_1, D_2) .

For the symmetric case of, $R_1 = R_2 = R$ and $D_1 = D_2 = D_S$, we generalize the symmetric MD RDF (Definition 2.8) to the zero-delay case.

Definition 2.14 (Symmetric ZDMD RDF)

The symmetric ZDMD RDF for a source, $\{X\}$, with MSE distortion constraints, $D_0, D_S > 0$, is

$$R_{ZD}^{\text{op}}(D_0, D_S) \triangleq \inf_{\text{s.t. } (R_S, R_S) \in \mathcal{R}_X^{ZD}(R_S, R_S, D_0, D_S, D_S)} R \quad (2.31)$$

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

Similar to classical MD case, we express the operational symmetric ZDMD RDF in terms of the ZD sum-rate.

Definition 2.15 (Operational symmetric ZDMD RDF)

For a discrete-time stationary source $\{X_k\}$, $X_k \in \mathcal{X}$, with MSE distortion constraints, $D_S \geq D_0 > 0$, the operational symmetric ZDMD RDF is

$$R_{ZD}^{\text{op}}(D_0, D_S) \triangleq \inf \lim_{n \rightarrow \infty} \frac{1}{2n} \left(\mathbb{E} \left[L_n^{(1)} \right] + \mathbb{E} \left[L_n^{(2)} \right] \right) \quad (2.32)$$

$$\text{s.t. } \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \mathbb{E} \left[\left\| X_k - Y_k^{(0)} \right\|_2^2 \right] \leq D_0$$

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \mathbb{E} \left[\left\| X_k - Y_k^{(i)} \right\|_2^2 \right] \leq D_S, \quad i = 1, 2,$$

where the infimum is over possible ZDMD encoder- and -decoder sequences $\{f_n^{(1)}\}_{n \in \mathbb{N}}$, $\{f_n^{(2)}\}_{n \in \mathbb{N}}$, $\{g_n^{(0)}\}_{n \in \mathbb{N}}$, $\{g_n^{(1)}\}_{n \in \mathbb{N}}$, $\{g_n^{(2)}\}_{n \in \mathbb{N}}$, i.e. that satisfy (2.23)–(2.28).

We can now formally state and define the problem considered in this report.

2.5 Problem Definition

In this report we consider symmetric ZDMD source coding with non-degenerate MSE distortion constraints, $D_S \geq D_0 > 0$, of the stable stationary scalar Gauss-Markov source process,

$$X_k = aX_{k-1} + W_k, \quad k \in \mathbb{N}, \quad (2.33)$$

where $|a| < 1$ is the deterministic correlation coefficient, $X_1 \in \mathbb{R} \sim \mathcal{N}(0, \sigma_{X_1}^2)$ with $\sigma_{X_1}^2 = \frac{\sigma_W^2}{1-a^2}$ is the initial state, and $W_k \in \mathbb{R} \sim \mathcal{N}(0, \sigma_W^2)$, is an IID Gaussian process independent of X_1 .

These design requirements are summarized in following optimization problem.

Problem 1 (Operational symmetric scalar Gaussian ZDMD RDF)

For a stable stationary scalar Gauss-Markov source process (2.33), with non-degenerate MSE distortion constraints, $D_0, D_S > 0$, determine the operational symmetric ZDMD RDF, i.e. solve the optimization problem:

$$\begin{aligned} R_{\text{ZD}}^{\text{op}}(D_0, D_S) = \inf & \quad \lim_{n \rightarrow \infty} \frac{1}{2n} \left(\mathbb{E} \left[L_n^{(1)} \right] + \mathbb{E} \left[L_n^{(2)} \right] \right) \\ \text{s.t.} & \quad \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \mathbb{E} \left[\left(X_k - Y_k^{(0)} \right)^2 \right] \leq D_0 \\ & \quad \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \mathbb{E} \left[\left(X_k - Y_k^{(i)} \right)^2 \right] \leq D_S, \quad i = 1, 2, \end{aligned} \quad (2.34)$$

where the infimum is over possible ZDMD encoder- and -decoder sequences $\{f_n^{(1)}\}_{n \in \mathbb{N}}$, $\{f_n^{(2)}\}_{n \in \mathbb{N}}$, $\{g_n^{(0)}\}_{n \in \mathbb{N}}$, $\{g_n^{(1)}\}_{n \in \mathbb{N}}$, $\{g_n^{(2)}\}_{n \in \mathbb{N}}$, i.e. that satisfy (2.23)–(2.28).

Unfortunately the solution to Problem 1 is very hard to find, since it is determined by a minimization over all possible operational ZDMD codes. Similar to single description ZD rate-distortion, the standard MD region only provides a conservative outer bound on the ZDMD region, due space-filling losses, memoryless entropy coding and causal filters.

To obtain a solution to Problem 1, following the standard way of rate-distortion theory, we introduce an information-theoretic lower bound.

3 | Lower Bound

In this chapter we determine a novel information-theoretic lower bound on the sum-rate of ZDMD source coding with feedback. To do so, we first formally introduce the ZDMD coding scenario with feedback and side information available to both encoder and decoder¹. Using this lower bound we present an information theoretic-equivalent of the operational symmetric Gaussian ZDMD RDF. Finally, we show for stationary scalar Gaussian sources under a sequential greedy coding constraint, how Gaussian reproductions minimize the information-theoretic lower bound. Thus providing a lower bound to Problem 1.

Usually in rate-distortion theory with side information, the side information is random information only available at the decoder that is jointly distributed with the source, and hence useful for improving the source code [13], [48], [49]. This information could be some type of state information, e.g. the signal-to-noise ration (SNR) of a wireless relay channel, where the decoder receives multiple noisy versions of the source signal, and is able to estimate the SNR which is unknown to the encoder at the time of transmission [48]. By knowledge of the SNR of the channel the decoder can perform a better estimate of the source.

In this report, we consider the side information to be independent of the source, and available at both encoder and decoder. Here the side information is some kind of meta-data independent of the source, e.g. the formatting setup of an email, that is independent of the email contents itself. Specifically, we may consider the current independent dither value of a dithered quantizer as information about the state of the quantizer. By synchronizing random number generators at the encoder and decoder it would be possible to generate the same dither (side information) independent of the source at both encoder and decoder [10].

¹We emphasize, this is not side information in the information-theoretic sense of multi-terminal- or Wyner-Ziv source coding, where the side information is unknown and only available at the decoder [13, Sec. 15.9][49].

3.1 ZDMD Coding with Feedback

To establish a lower bound on the sum-rate of a ZDMD source code, we recall [8] showed in the single description case, that the operational ZD RDF may be lower bounded using a feedback realization scheme. This information theoretic lower bound on the operational rate was originally established in [5] in regards to control system design with data-rate constraints, i.e. source coding inside feedback loops. However, we do not consider MD source coding inside feedback loops, only MD coding with perfect decoder feedback. Therefore, we modify the double branch MD source coding scenario in Figure 2.1 on page 32 to include feedback from the decoders, and side information.

This constructive technique allows us to directly replace encoder and decoder blocks with appropriate AWGN channels and predictors when deriving a test-channel realization in the following chapter.

We consider the open-loop MD source coding problem with feedback shown in Figure 3.1. Here the source process $\{X_k\}$ is completely specified by its known statistical properties. For each time step $k \in \mathbb{N}$ the ZDMD encoder, \mathcal{E} , observes a new source sample X_k while assuming it has already observed the past sequence X^{k-1} . Furthermore 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, $Y^{(1),k-2}, Y^{(2),k-2}$. The encoder then produces two binary descriptions $B_k^{(1)}, B_k^{(2)}$. These are transmitted across two zero-delay error-free digital channels to the three reconstruction decoders, $\mathcal{D}_0, \mathcal{D}_1, \mathcal{D}_2$. The decoders each produce an estimate, $Y_k^{(i)}$, $i = 0, 1, 2$, of the source, X_k . The feedback channels are assumed to have a one-sample delay to ensure the operational feasibility of the system.

In the open-loop 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 . The side information is in the sets $\mathcal{S}_{\mathcal{E},k}$ and $\mathcal{S}_{\mathcal{D}_i,k}$ respectively.

We now define in detail the operations of the different blocks in Figure 3.1. First at each time step, k , the encoder, \mathcal{E} , has available to it all source samples up to time k , X^k , and previous reproductions, $Y^{(i),k-1}$, $i = 1, 2$. 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. The encoder then performs lossy source coding and lossless entropy coding to produce the dependent codewords,

$$\left(B_k^{(1)}, B_k^{(2)} \right) = \mathcal{E}_k \left(X^k, Y^{(1),k-1}, Y^{(2),k-1}, S_{\mathcal{E}}^k \right), \quad k \in \mathbb{N} \quad (3.1)$$

with length $l_k^{(i)}$ $i = 1, 2$ (in bits).

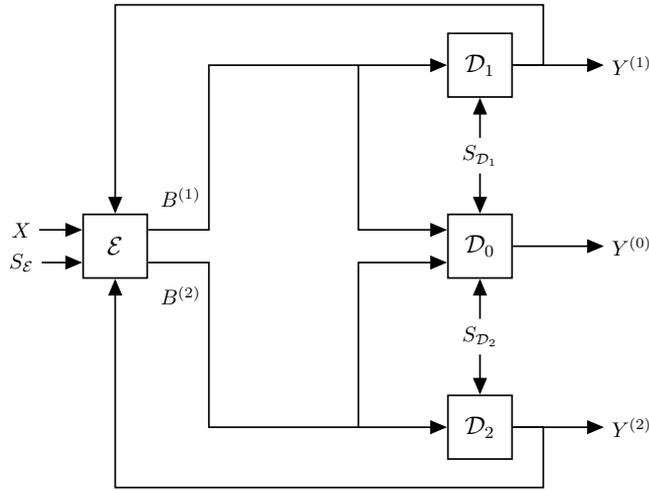


Figure 3.1: A general MD source-coding scenario with feedback.

Since the codewords are transmitted across zero-delay error-free digital channels to the three decoders, the decoders immediately decode the binary codewords and produce the estimates of X_k ,

$$Y_k^{(1)} = \mathcal{D}_k^{(1)} \left(B^{(1),k}, S_{\mathcal{D}_1}^k \right), \quad (3.2a)$$

$$Y_k^{(2)} = \mathcal{D}_k^{(2)} \left(B^{(2),k}, S_{\mathcal{D}_2}^k \right), \quad (3.2b)$$

$$Y_k^{(0)} = \mathcal{D}_k^{(0)} \left(B^{(1),k}, B^{(2),k}, S_{\mathcal{D}_1}^k, S_{\mathcal{D}_2}^k \right). \quad (3.2c)$$

We now define the ZDMD coding problem with feedback.

Definition 3.1 (ZDMD coding problem with feedback)

For a discrete-time stationary source process $\{X_k\}$, with non-degenerate 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 (3.1) and (3.2), such that the asymptotic average expected distortions satisfy

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \mathbb{E} \left[\left(X_k - Y_k^{(i)} \right)^2 \right] \leq D_S, \quad i = 1, 2. \quad (3.3)$$

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 (3.1) and (3.2).

3.2 Lower Bound on Sum-rate

We now study a lower bound on the sum-rate of the ZDMD coding problem with feedback that depends only on the joint statistics of the source encoder input X and the decoder outputs $Y^{(i)}$ $i = 0, 1, 2$. Our derivation of the lower bound requires the following assumption.

Assumption 3.2

The systems $\mathcal{E}, \mathcal{D}^{(i)}$ $i = 0, 1, 2$, are causal, described by (3.1)–(3.2), 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 encoders and decoders must be causal and use only past and present symbols, and side information that is not associated with the transmitted message [5]. Similar to [5] the channel is the only link between encoder and decoder. However, we further assume the channel to have perfect feedback.

Furthermore, we require the decoders to be invertible.

Definition 3.3 (Invertible decoder [5, Def. 4.2])

The decoders, $\mathcal{D}^{(i)}$, $i = 0, 1, 2$, defined in (3.2) 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

$$B^{(1),k} = \mathcal{G}_k^{(1)} \left(Y_k^{(1)}, S_{\mathcal{D}_1}^k \right), \quad (3.4a)$$

$$B^{(2),k} = \mathcal{G}_k^{(2)} \left(Y_k^{(2)}, S_{\mathcal{D}_2}^k \right), \quad (3.4b)$$

$$\left(B^{(1),k}, B^{(2),k} \right) = \mathcal{G}_k^{(0)} \left(Y_k^{(0)}, S_{\mathcal{D}_1}^k, S_{\mathcal{D}_2}^k \right). \quad (3.4c)$$

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 WLOG 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 and proof are a generalization of [5, Lemma 4.2] to the MD scenario.

Lemma 3.4 (Feedback Markov Chains)

Consider an MD source coding scheme inside a feedback loop as shown in Figure 3.1. If Assumption 3.2 applies and if the decoders are invertible when given the side information, then the Markov chain,

$$X^k |_{\phi_1} - (B_k^{(1)}, B_k^{(2)}) |_{\phi_1} - Y_k^{(0)} |_{\phi_1}, \quad k \in \mathbb{N}, \quad (3.5)$$

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

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

$$Y_k^{(2)} |_{\phi_2} - B_k^{(1)} |_{\phi_2} - Y_k^{(1)} |_{\phi_2}, \quad k \in \mathbb{N}, \quad (3.6)$$

also holds.

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

$$Y_k^{(1)} |_{\phi_3} - B_k^{(2)} |_{\phi_3} - Y_k^{(2)} |_{\phi_3}, \quad k \in \mathbb{N}, \quad (3.7)$$

holds.

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

$$Y^{(2),k-1} - Y^{(1),k-1} - S_{\mathcal{D}_1}^k, \quad k \in \mathbb{N}, \quad (3.8)$$

$$Y^{(1),k} - Y^{(2),k-1} - S_{\mathcal{D}_2}^k, \quad k \in \mathbb{N}, \quad (3.9)$$

hold.

Proof

The Markov chain (3.5) follows, since $Y_k^{(0)}$ depends deterministically upon $(B^{(1),k}, B^{(2),k}, S_{\mathcal{D}_1}^k, S_{\mathcal{D}_2}^k)$.

Similarly, (3.6) holds, since $Y_k^{(1)}$ depends deterministically upon $(B^{(1),k}, S_{\mathcal{D}_1}^k)$. The Markov chain in (3.7) follows analogously.

By the system equations we have that

$$(B_1^{(1)}, B_1^{(2)}) = \mathcal{E}_1(X_1, \emptyset, \emptyset, S_{\mathcal{E},1}) \quad (3.10)$$

$$Y_1^{(1)} = \mathcal{D}_1^{(1)}(B_1^{(1)}, S_{\mathcal{D}_1,1}) \quad (3.11)$$

$$Y_1^{(2)} = \mathcal{D}_1^{(2)}(B_1^{(2)}, S_{\mathcal{D}_2,1}). \quad (3.12)$$

Since $S_{\mathcal{D}_1,1} \perp S_{\mathcal{D}_2,1}$, it follows that $Y_1^{(2)} \perp S_{\mathcal{D}_1,1}$. Furthermore, since $S_{\mathcal{D}_1,2} \perp S_{\mathcal{D}_2,1}$

then $Y_1^{(2)} \perp\!\!\!\perp \mathcal{S}_{\mathcal{D}_1,2}$. Hence, (3.8) holds in the initial step. Now in the next time step,

$$(B_2^{(1)}, B_2^{(2)}) = \mathcal{E}_2(X^2, Y_1^{(1)}, Y_1^{(2)}, S_{\mathcal{E},2}) \quad (3.13)$$

$$= \mathcal{E}_2(X^2, \mathcal{D}_1^{(1)}(B_1^{(1)}, S_{\mathcal{D}_1,1}), Y_1^{(2)}, S_{\mathcal{E},2}) \quad (3.14)$$

$$Y_2^{(1)} = \mathcal{D}_2^{(1)}(B_2^{(1)}, S_{\mathcal{D}_1,2}) \quad (3.15)$$

$$Y_2^{(2)} = \mathcal{D}_2^{(2)}(B_2^{(2)}, S_{\mathcal{D}_2,2}), \quad (3.16)$$

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

$$Y_2^{(2)} - Y_1^{(1)} - S_{\mathcal{D}_1,1}. \quad (3.17)$$

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 (3.8).

For (3.9), 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, (3.9) holds. ■

When using dithered quantization the dither is the side information [50]. For the correlated noise in relation to Ozarow's MD test-channel, we do not correlate the dither signals, but the quantization noise. Thus, the side information of the two descriptions is mutually independent. Furthermore, for vector quantization there seems to be no gain from correlating the dither signals [50]. Additionally, if using MD coding with refinement, i.e. adding a second layer of dithered quantization in the encoder, this extra dither signal may be included completely in either decoder side information and still achieve independence. There also seems to be no gain in correlating this third dither signal with the others [39]. Hence, it is reasonable to assume the side information is mutually independent.

We are now ready to state our first main result.

Theorem 3.5 (Lower bound on sum-rate)

Consider a ZDMD source coding problem with feedback (Definition 3.1), as seen in Figure 3.1. If Assumption 3.2 holds, the decoders are invertible, and the decoder side information is mutually independent, then

$$R_1 + R_2 \geq \bar{I}_\infty \left(X \rightarrow Y^{(0)} \right) + \bar{I} \left(Y^{(1)}; Y^{(2)} \right). \quad (3.18)$$

The second term in (3.18) is the *mutual information rate* (Definition A.15) between the side reconstruction processes.

The proof of Theorem 3.5 and the following corollary can be found in Appendix D.

Corollary 3.6 (Alternative lower bound on sum-rate)

Consider a ZDMD source coding problem with feedback (Definition 3.1), as seen in Figure 3.1. If Assumption 3.2 holds, the decoders are invertible, and the decoder side information is mutually independent, then

$$R_1 + R_2 \geq \bar{I}_\infty \left(X \rightarrow Y^{(1)}, Y^{(2)} \right) + \bar{I} \left(Y^{(1)}; Y^{(2)} \right) \quad (3.19)$$

Theorem 3.5 and its corollary show, 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.

This also shows that the appropriate definition of an information-theoretic symmetric ZDMD RDF is given in terms of the directed- and mutual information rate. To the best of the authors' knowledge, Theorem 3.5 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 [24].

In relation to the El-Gamal and Cover region [24] (see Lemma 2.6), 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 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.

The mutual information term $I(Y^{(1),n}; Y^{(2),n})$ does not imply a non-causal 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 [21]. 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 (D.7). Therefore, the MD code must convey this total information flow between the two descriptions to the central receiver.

3.3 Gaussian Lower Bound

In this section we formally define the information-theoretic symmetric Gaussian ZDMD RDF, $R_{\text{ZD}}^I(D_0, D_S)$, as a lower bound to $R_{\text{ZD}}^{\text{op}}(D_0, D_S)$. To this end we present in more detail the test-channel distribution associated with this minimization. Finally, we show that Gaussian reproductions minimize the lower bound.

3.3.1 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 $d_{1,n}^{(i)}(x^n, y^{(i),n}) \triangleq \frac{1}{n} \sum_{k=1}^n \|x_k - y_k^{(i)}\|_2^2$, $i = 0, 1, 2$.

Source

We assume the source distribution satisfies conditional independence

$$P(x_k | x^{k-1}, y^{(0),k-1}, y^{(1),k-1}, y^{(2),k-1}) \triangleq P(x_k | x^{k-1}), \quad k \in \mathbb{N}^n. \quad (3.20)$$

This implies there is no feedback from the reproductions, $Y^{(i)}$, to the source X . Hence, the next source symbol, given the previous symbols, is not further related to the previous reproductions [21]. We assume the distribution at $k = 1$ is $P(x_1)$. Furthermore, by Bayes’ rule [8],

$$P(x^n) \triangleq \prod_{k=1}^n P(x_k | x^{k-1}). \quad (3.21)$$

For the Gauss-Markov source of process (2.33) this implies $\{W_k\}$ is independent of the past reproductions $Y^{(i),k-1}$, $i = 0, 1, 2$ [8].

Reproductions

Since there is no feedback from the reproductions to the source, the MD encoder-decoder pairs from \mathcal{E} to \mathcal{D}_i , $i = 0, 1, 2$ in Figure 3.1, are causal if, and only if, the following Markov chain holds [16]:

$$X_{k+1}^n - X^k - \left(Y^{(0),k}, Y^{(1),k}, Y^{(2),k} \right), \quad \forall k \in \{1, \dots, n-1\}. \quad (3.22)$$

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

$$P \left(y_k^{(0)}, y_k^{(1)}, y_k^{(2)} | y^{(0),k-1}, y^{(1),k-1}, y^{(2),k-1}, x^k \right), \quad k \in \mathbb{N}^n. \quad (3.23)$$

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

$$P \left(y_1^{(0)}, y_1^{(1)}, y_1^{(2)} | y^{(0),0}, y^{(1),0}, y^{(2),0}, x^1 \right) = P \left(y_1^{(0)}, y_1^{(1)}, y_1^{(2)} | x_1 \right). \quad (3.24)$$

3.3.2 Lower Bound

We now formally define the information-theoretic symmetric ZDMD RDF.

Definition 3.7 (Information-Theoretic Symmetric ZDMD RDF)

The *information-theoretic symmetric ZDMD RDF*, for the stationary Gaussian source process $\{X_k\}$, with MSE distortion constraints, $D_0, D_S \geq 0$, is

$$\begin{aligned} R_{ZD}^I(D_0, D_S) &\triangleq \inf \left[\frac{1}{2} \bar{I}_\infty \left(X \rightarrow Y^{(1)}, Y^{(2)} \right) + \frac{1}{2} \bar{I} \left(Y^{(1)}; Y^{(2)} \right), \right. \\ &\text{s.t.} \quad \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \mathbb{E} \left[\left\| X_k - Y_k^{(0)} \right\|_2^2 \right] \leq D_0 \\ &\quad \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \mathbb{E} \left[\left\| X_k - Y_k^{(i)} \right\|_2^2 \right] \leq D_S, \quad i = 1, 2, \end{aligned} \quad (3.25)$$

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

$$X_{k+1}^\infty - X^k - \left(Y^{(0),k}, Y^{(1),k}, Y^{(2),k} \right), \quad \forall k \in \mathbb{N}. \quad (3.26)$$

The minimization over all process $\{Y_k^{(i)}\}$, $i = 0, 1, 2$ that satisfy the Markov chain (3.26), is equivalent to minimization over 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, which we show to be an achievable lower bound on Problem 1.

Problem 2 (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

$$\begin{aligned} R_{\text{ZD,GM}}^I(D_0, D_S) \triangleq & \inf \frac{1}{2} \bar{I}_\infty \left(X \rightarrow Y^{(1)}, Y^{(2)} \right) + \frac{1}{2} \bar{I} \left(Y^{(1)}; Y^{(2)} \right), \\ \text{s.t.} \quad & \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \mathbb{E} \left[\left\| X_k - Y_k^{(0)} \right\|_2^2 \right] \leq D_0 \\ & \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \mathbb{E} \left[\left\| X_k - Y_k^{(i)} \right\|_2^2 \right] \leq D_S, \quad i = 1, 2, \end{aligned} \quad (3.27)$$

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

$$X_{k+1}^\infty - X^k - \left(Y^{(0),k}, Y^{(1),k}, Y^{(2),k} \right), \quad \forall k \in \mathbb{N}. \quad (3.28)$$

As before, this minimization is equivalent to minimization over all sequences of Gaussian conditional test-channel distributions

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

Before showing Gaussian reproductions lower bound the information-theoretic ZDMD RDF we introduce the following technical conditions required for our proof.

Definition 3.8 (Sequential greedy coding)

Consider the symmetric ZDMD coding problem in Figure 3.1. We say, that we solve this problem using *sequential greedy coding*, if sequentially for each time step $k \in \mathbb{N}$: We minimize the bitrate such that the MSE distortion constraints $D_S \geq D_0 > 0$ are satisfied for each $k \in \mathbb{N}$.

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

$$\mathbb{E} \left[\left\| X_k - Y_k^{(0)} \right\|_2^2 \right] \leq D_0 \quad (3.29)$$

$$\mathbb{E} \left[\left\| X_k - Y_k^{(i)} \right\|_2^2 \right] \leq D_S, \quad i = 1, 2. \quad (3.30)$$

Since in sequential greedy coding we minimize the bitrate for each $k \in \mathbb{N}$ in the sequential order subject to the distortion constraints, this implies for the information

rates in (3.25), that we minimize the sum

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

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

$$\begin{aligned} \mathbb{E} \left[\left\| X_k - Y_k^{(0)} \right\|_2^2 \right] &\leq D_0 \\ \mathbb{E} \left[\left\| X_k - Y_k^{(i)} \right\|_2^2 \right] &\leq D_S, \quad i = 1, 2, \end{aligned} \quad (3.31)$$

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

Furthermore, sequential greedy coding implies, if $\tilde{Y}_1^{(i)}$, $i = 1, 2$ minimizes the initial mutual informations for $k = 1$, i.e.

$$I\left(X_1; Y_1^{(1)}, Y_1^{(2)}\right) + I\left(Y_1^{(2)}; Y_1^{(1)}\right) \geq I\left(X_1; \tilde{Y}_1^{(1)}, \tilde{Y}_1^{(2)}\right) + I\left(\tilde{Y}_1^{(2)}; \tilde{Y}_1^{(1)}\right) \quad (3.32)$$

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

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

where, $\tilde{Y}_1^{(i)}$, $i = 1, 2$, is inserted on both sides of the conditioning. In the Gaussian source case, we speculate the sequential greedy condition provides the same result as selecting the optimal joint distribution across all k , i.e. when the distribution of $Y_k^{(i)}$ is allowed to change for any $k' > k$. However, we have not been able to prove this.

The sequential greedy condition also sits well within the the ZD perspective. Since we must send the optimum description that minimizes the rate while achieving the desired distortion at each time step.

We also need the following condition on the MMSE predictors.

Definition 3.9 (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 $k \in \mathbb{N}$,

$$Y_k^{(i)} - \mathbb{E} \left[Y_k^{(i)} | Y^{(1),k-1}, Y^{(2),k-1} \right] \perp \left(Y^{(1),k-1}, Y^{(2),k-1} \right), \quad i = 1, 2, \quad (3.33)$$

$$Y_k^{(i)} - \mathbb{E} \left[Y_k^{(i)} | Y^{(i),k-1} \right] \perp Y^{(i),k-1}, \quad i = 1, 2, \quad (3.34)$$

$$Y_k^{(i)} - \mathbb{E} \left[Y_k^{(i)} | Y^{(j),k-1} \right] \perp Y^{(j),k-1}, \quad i \neq j, i, j \in \{1, 2\}. \quad (3.35)$$

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

For mutual informations the conditional prediction residual independence implies

$$\begin{aligned} & I \left(Y_k^{(1)} - \mathbb{E} \left[Y_k^{(1)} | Y^{(1),k-1} \right]; Y_k^{(2)} - \mathbb{E} \left[Y_k^{(2)} | Y^{(1),k-1} \right] | Y^{(1),k-1} \right) \\ &= I \left(Y_k^{(1)} - \mathbb{E} \left[Y_k^{(1)} | Y^{(1),k-1} \right]; Y_k^{(2)} - \mathbb{E} \left[Y_k^{(2)} | Y^{(1),k-1} \right] \right). \end{aligned}$$

Particularly, if $\{Y_k^{(i)}\}$, $i = 1, 2$ are Gaussian then the MMSE predictors have conditional prediction residual independence by the orthogonality principle [51, p. 45]. Using these predictors may add to the 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 the other main result of this chapter.

Theorem 3.10 (Gaussian bound for scalar sources)

Let $\{X_k\}_{k \in \mathbb{N}}$ be a stable stationary *scalar* Gaussian process (2.33), with MSE distortion constraints, $D_S \geq D_0 > 0$. Then under the sequential greedy coding condition (Definition 3.8), and if the reproduction sequences $\{Y_k^{(i)}\}$, $i = 1, 2$, satisfy conditional prediction residual independence (Definition 3.9), the following inequality holds

$$R_{\text{ZD,GM}}^I(D_0, D_S) \leq R_{\text{ZD}}^I(D_0, D_S). \quad (3.36)$$

The proof of Theorem 3.10 can be found in Appendix E.

Theorem 3.10 shows, that for scalar stationary Gaussian sources under sequential greedy coding and MSE distortion constraints, Gaussian reproduction processes

minimizes the information theoretic symmetric ZDMD RDF. That is, 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]. The main difficulty in showing this, and the reason for the technical conditions, is the second term of (3.25) and showing that the side reproductions should be jointly Gaussian in all time steps.

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 [52].

Finally, by Theorem 3.5 and Theorem 3.10 we have the following corollary, showing Problem 2 as a lower bound to Problem 1.

Corollary 3.11

Let $\{X_k\}_{k \in \mathbb{N}}$ be a stable stationary *scalar* Gaussian process (2.33), with MSE distortion constraints, $D_S \geq D_0 > 0$. Then under the sequential greedy coding condition (Definition 3.8), and if the reproduction sequences $\{Y_k^{(i)}\}$, $i = 1, 2$, satisfy conditional prediction residual independence (Definition 3.9), the following inequalities hold

$$R_{\text{ZD,GM}}^I(D_0, D_S) \leq R_{\text{ZD}}^I(D_0, D_S) \leq R_{\text{ZD}}^{\text{OP}}(D_0, D_S). \quad (3.37)$$

With this information theoretic lower bound on $R_{\text{ZD}}^{\text{OP}}(D_0, D_S)$, we now derive an optimal test-channel realization scheme, that achieves the lower bound.

4 | Test-Channel Realization

In this chapter we introduce a feedback realization of the optimal test-channel for the Gaussian information-theoretic symmetric ZDMD RDF, $R_{\text{ZD,GM}}^I(D_0, D_S)$. This test-channel is based on the ZDMD coding problem with feedback in Figure 3.1 and the feedback realization scheme of [8] (Section 1.5). Finally, we present a characterization of $R_{\text{ZD,GM}}^I(D_0, D_S)$ as the solution to an optimization problem. This provides an achievable lower bound to Problem 1 in a Gaussian coding scheme.

4.1 Source Process

We consider the scalar stationary Gauss-Markov source process, X_k , on the form,

$$X_{k+1} = aX_k + W_k, \quad k \in \mathbb{N} \quad (4.1)$$

where $|a| < 1$ is the correlation coefficient of the process, $X_1 \in \mathbb{R} \sim \mathcal{N}(0, \Sigma_{X_1})$ ¹ is the initial state with $\Sigma_{X_1} = \frac{\Sigma_W}{1-a^2}$, and $W_k \in \mathbb{R} \sim \mathcal{N}(0, \Sigma_W)$, is an IID Gaussian process independent of X_1 .

4.2 Predictive Coding

For each side channel we follow the feedback realization of [8, Theorem 2] as introduced in Section 1.5. Hence, the reproduction sequence of the optimum test-channel is realized by

$$Y_k^{(i)} = hX_k + (1-h)aY_{k-1}^{(i)} + Z_k^{(i)}, \quad (4.2)$$

where $Z_k^{(i)} \sim \mathcal{N}(0, \Sigma_{Z_S})$,

$$h \triangleq 1 - \pi_S \lambda^{-1}, \quad (4.3)$$

$$\Sigma_{Z_S} \triangleq \pi_S h, \quad i = 1, 2, \quad (4.4)$$

$$\lambda = a^2 \pi_S + \sigma_W^2. \quad (4.5)$$

¹For a more consistent notation throughout this chapter, we denote most variances and covariances with uppercase Σ , although all variables are scalar.

Here λ is the variance of side error process,

$$\begin{aligned} U_k^{(i)} &\triangleq X_k - \mathbb{E} \left[X_k | Y^{(i),k-1} \right], \\ &= X_k - aY_{k-1}^{(i)}, \quad i = 1, 2. \end{aligned} \quad (4.6)$$

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

$$\pi_S \triangleq \mathbb{E} \left[\left(X_k - Y_k^{(i)} \right)^2 \right] = \mathbb{E} \left[\left(U_k^{(i)} - \tilde{U}_k^{(i)} \right)^2 \right], \quad i = 1, 2, \quad (4.7)$$

where $\tilde{U}_k^{(i)}$ is the innovation process,

$$\tilde{U}_k^{(i)} \triangleq Y_k^{(i)} - \mathbb{E} \left[Y_k^{(i)} | Y^{(i),k-1} \right] \quad (4.8)$$

$$= hU_k^{(i)} + Z_k^{(i)}, \quad i = 1, 2, \quad (4.9)$$

with variance,

$$\Sigma_{\tilde{U}} = h^2\lambda + \pi_S h. \quad (4.10)$$

As in the single-description case the innovation process, $\tilde{U}_k^{(i)}$ $i = 1, 2$, can be viewed as the i th side decoders estimate of $U_k^{(i)}$.

We also define a central description of the innovations as the average of the two side innovation processes,

$$V_{C,k} \triangleq \frac{1}{2} \left(\tilde{U}_k^{(1)} + \tilde{U}_k^{(2)} \right). \quad (4.11)$$

Finally, we have that

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

and the joint test-channel noise distribution is

$$\begin{bmatrix} Z_k^{(1)} \\ Z_k^{(2)} \end{bmatrix} \sim \mathcal{N} \left(0, \Sigma_Z \right), \quad (4.12)$$

where

$$\Sigma_Z = \begin{bmatrix} \pi_S h & \rho \pi_S h \\ \rho \pi_S h & \pi_S h \end{bmatrix}. \quad (4.13)$$

Figure 4.1 on page 62 illustrates this feedback realization scheme for the optimum test-channel. We note that the test-channel in Figure 4.1 differs from the usual MD

double-branch test-channel, 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 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 3.1 and the test-channel in Figure 4.1a. This shows, how the general ZDMD coding problem and its lower bound provides a constructive result that is conveniently extended to an optimum test-channel realization.

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 4.1.

Lemma 4.1 (Covariances)

Let $\{X_k\}$ be a stable stationary scalar Gauss-Markov process as in (4.1) with stationary variance $\text{Var}[X_k] = \Sigma_X$. Using the feedback coding scheme of Figure 4.1, then the following covariances hold:

$$\Sigma_{XY} \triangleq \text{Cov}[X_k, Y_k^{(i)}] = \frac{h}{1 - a^2(1 - h)} \Sigma_X, \quad i = 1, 2, \quad (4.14)$$

$$\Sigma_{XV_C} \triangleq \text{Cov}[X_k, V_{C,k}] = h(\Sigma_X - a^2 \Sigma_{XY}), \quad (4.15)$$

$$\Sigma_Y \triangleq \text{Var}[Y_k^{(i)}] = \frac{h^2 \Sigma_X + 2a^2 h(1 - h) \Sigma_{XY} + \Sigma_{Z_S}}{1 - a^2(1 - h)^2}, \quad i = 1, 2, \quad (4.16)$$

$$\Sigma_{Y^{(1)}Y^{(2)}} \triangleq \text{Cov}[Y_k^{(1)}, Y_k^{(2)}] = \frac{h^2 \Sigma_X + 2a^2 h(1 - h) \Sigma_{XY} + \Sigma_{Z^{(1)}Z^{(2)}}}{1 - a^2(1 - h)^2} \quad (4.17)$$

$$\Sigma_{U^{(1)}U^{(2)}} \triangleq \text{Cov}[U_k^{(1)}, U_k^{(2)}] = \Sigma_X + a^2(\Sigma_{Y^{(1)}Y^{(2)}} - 2\Sigma_{XY}) \quad (4.18)$$

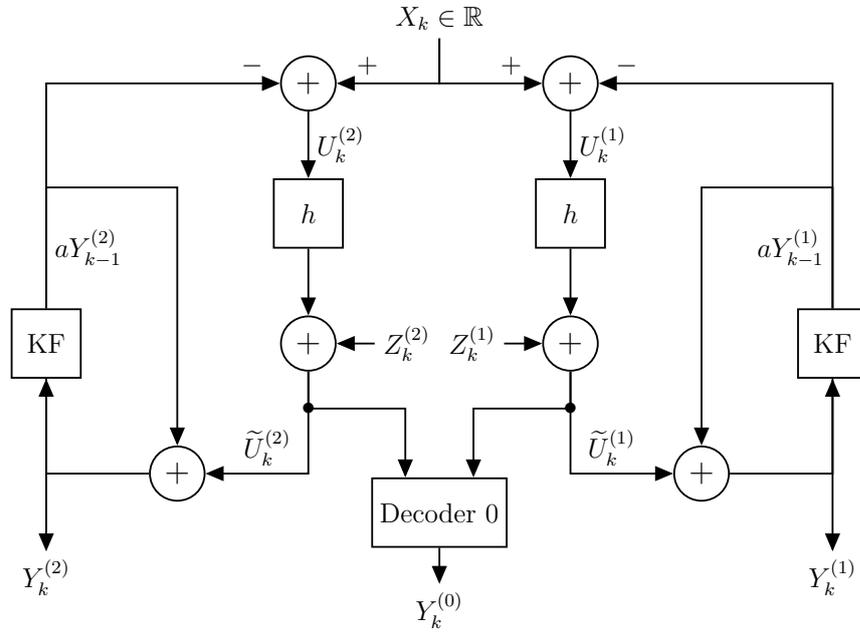
$$\Sigma_{UV_C} \triangleq \text{Cov}[U_k^{(i)}, V_{C,k}] = \frac{1}{2}h(\lambda + \Sigma_{U^{(1)}U^{(2)}}), \quad i = 1, 2 \quad (4.19)$$

$$\Sigma_{V_C} \triangleq \text{Var}[V_{C,k}] = \frac{1}{2}(\Sigma_{\tilde{U}} + h^2 \Sigma_{U^{(1)}U^{(2)}} + \Sigma_{Z^{(1)}Z^{(2)}}). \quad (4.20)$$

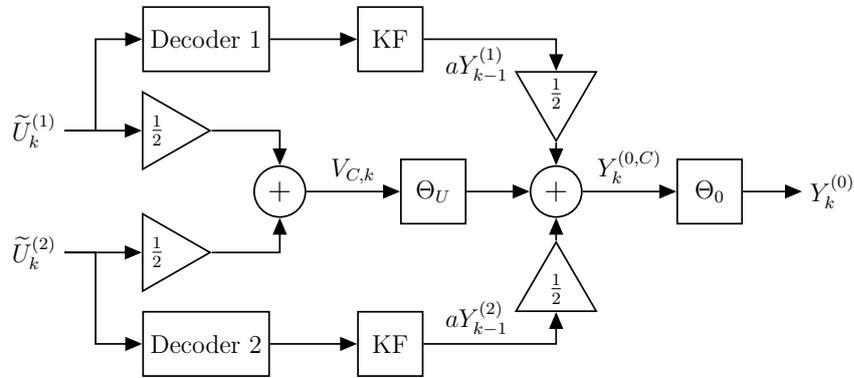
The proof of Lemma 4.1 can be found in Appendix F.

4.3 Central Decoder Design

The ZDMD encoder creates the two descriptions by pre-scaling and adding correlated noises to the two prediction error process, $U_k^{(1)}$, $U_k^{(2)}$, resulting in the two innovation processes, $\tilde{U}_k^{(1)}$, $\tilde{U}_k^{(2)}$, as the decoder estimates of $U_k^{(1)}$, $U_k^{(2)}$. The central



(a) Overall test channel with side decoders



(b) Test channel central decoder

Figure 4.1: Feedback realization of the optimum test channel for $R_{\text{ZD,GM}}^I(D_0, D_S)$.

decoder receives both innovation processes. 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, resulting in an overall lower central distortion. Therefore, we start by deriving a joint estimate of the prediction error processes.

4.3.1 Joint Estimation of Error Process

At each time step, k , the central decoder receives the two innovation processes $\tilde{U}_k^{(1)}, \tilde{U}_k^{(2)}$. By taking the average of these we get the central innovations description $V_{C,k}$. In the central decoder, the MMSE estimator of $U_k^{(i)}$ given $V_{C,k}$ is then[51],

$$\hat{U}_k^{(i)} \triangleq \text{E} \left[U_k^{(i)} | V_{C,k} \right] = \Theta_U V_{C,k}, \quad (4.21)$$

where

$$\Theta_U \triangleq \Sigma_{UV_C} \Sigma_{V_C}^{-1}. \quad (4.22)$$

4.3.2 Joint Estimation of Source

Using the central decoder estimates of $U_k^{(1)}, U_k^{(2)}$ we can now provide a better estimate of the source X_k . Initially, for each side-decoder prediction, $aY_{k-1}^{(i)}$, let

$$Y_k^{(i,C)} \triangleq \hat{U}_k^{(i)} + aY_{k-1}^{(i)}, \quad i = 1, 2, \quad (4.23)$$

$$= \Theta_U V_{C,k} + aY_{k-1}^{(i)}, \quad (4.24)$$

be the improved central estimates of X_k for each side-decoder prediction. However, since $Y^{(1)}$ and $Y^{(2)}$ are also correlated we can improve the central decoder estimate X_k by taking the following average, which is not necessarily the optimal estimator,

$$Y_k^{(0,C)} \triangleq \frac{1}{2} \left(Y_k^{(1,C)} + Y_k^{(2,C)} \right) \quad (4.25)$$

$$= \frac{1}{2} \left(\hat{U}_k^{(1)} + aY_{k-1}^{(1)} + \hat{U}_k^{(2)} + aY_{k-1}^{(2)} \right), \quad (4.26)$$

$$= \Theta_U V_{C,k} + \frac{1}{2} a \left(Y_{k-1}^{(1)} + Y_{k-1}^{(2)} \right). \quad (4.27)$$

4.3.3 MMSE Estimate Using $Y^{(0,C)}$

The central source description, $Y_k^{(0,C)}$ is not the joint MMSE estimate of X_k , since we only take an average in (4.27). The joint MMSE estimate of X_k given $Y_k^{(0,C)}$ is

$$Y_k^{(0)} \triangleq \text{E} \left[X_k | Y_k^{(0,C)} \right] = \Theta_0 Y_k^{(0,C)}, \quad (4.28)$$

with MSE

$$\pi_0 = \Sigma_X - \Sigma_{XY^{(0,C)}}^2 \Sigma_{Y^{(0,C)}}^{-1}, \quad (4.29)$$

where

$$\Theta_0 \triangleq \frac{\Sigma_{XY^{(0,C)}}}{\Sigma_{Y^{(0,C)}}}, \quad (4.30)$$

$$\Sigma_{XY^{(0,C)}} \triangleq \text{Cov} \left[X_k, Y_k^{(0,C)} \right], \quad (4.31)$$

$$\Sigma_{Y^{(0,C)}} \triangleq \text{Var} \left[Y_k^{(0,C)} \right]. \quad (4.32)$$

The covariance between X_k and $Y_k^{(0,C)}$ as well as the variance of $Y_k^{(0,C)}$ is given in the following lemma.

Lemma 4.2 (Central-reproduction covariances)

Let $\{X_k\}$ be a stable stationary scalar Gauss-Markov process as in (4.1). Using the feedback coding scheme of Figure 4.1, and joint decoder description (4.27), the following covariances hold

$$\Sigma_{XY^{(0,C)}} = \Theta_U \Sigma_{XV_C} + a^2 \Sigma_{XY} \quad (4.33)$$

$$\begin{aligned} \Sigma_{Y^{(0,C)}} &= \Theta_U^2 \Sigma_{V_C} + \frac{1}{2} a^2 (\Sigma_Y + \Sigma_{Y^{(1)Y^{(2)}}}) \\ &\quad + a^2 h \Theta_U (\Sigma_{XY} - \Sigma_{Y^{(1)Y^{(2)}}}), \end{aligned} \quad (4.34)$$

where Θ_U is defined in (4.22), and Σ_{V_C} , Σ_{XV_C} , Σ_{XY} , Σ_Y , $\Sigma_{Y^{(1)Y^{(2)}}$ are defined in Lemma 4.1.

The proof of Lemma 4.2 can be found in Appendix G.

The central decoder design is illustrated in Figure 4.1b. 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, the local side decoders produced the side reconstructions, $Y_{k-1}^{(i)}$, $i = 1, 2$, such that the central decoder has the predictions $aY_{k-1}^{(i)}$ $i = 1, 2$ available when producing the central estimate, $Y_k^{(0)}$, according to equations (4.27) and (4.28).

4.4 Rates

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

Initially using Corollary 3.6, we have for each time step k the mutual informations

in the definition of $R_{\text{ZD,GM}}^I(D_0, D_S)$ (3.27),

$$I\left(X^k; Y_k^{(1)}, Y_k^{(2)} | Y^{(1),k-1}, Y^{(2),k-1}\right) + I\left(Y_k^{(2)}; Y^{(1),k} | Y^{(2),k-1}\right) + I\left(Y_k^{(1)}; Y^{(2),k-1} | Y^{(1),k-1}\right). \quad (4.35)$$

Expressing the mutual informations using the differential entropy, we have

$$\begin{aligned} (4.35) &= h\left(Y_k^{(1)}, Y_k^{(2)} | Y^{(1),k-1}, Y^{(2),k-1}\right) - h\left(Y_k^{(1)}, Y_k^{(2)} | Y^{(1),k-1}, Y^{(2),k-1}, X^k\right) \\ &\quad + h\left(Y_k^{(2)} | Y^{(2),k-1}\right) - h\left(Y_k^{(2)} | Y^{(1),k}, Y^{(2),k-1}\right) \\ &\quad + I\left(Y_k^{(1)}; Y^{(2),k-1} | Y^{(1),k-1}\right) \\ &= h\left(Y_k^{(1)} | Y^{(1),k-1}, Y^{(2),k-1}\right) - h\left(Y_k^{(1)}, Y_k^{(2)} | Y^{(1),k-1}, Y^{(2),k-1}, X^k\right) \\ &\quad + h\left(Y_k^{(2)} | Y^{(2),k-1}\right) + I\left(Y_k^{(1)}; Y^{(2),k-1} | Y^{(1),k-1}\right) \\ &= h\left(Y_k^{(2)} | Y^{(2),k-1}\right) + h\left(Y_k^{(1)} | Y^{(1),k-1}\right) - h\left(Y_k^{(1)}, Y_k^{(2)} | Y^{(1),k-1}, Y^{(2),k-1}, X^k\right). \end{aligned}$$

Comparing the test channel of Figure 4.1 to the general ZDMD source coding scenario with feedback in Figure 3.1, we have

$$h\left(Y_k^{(i)} | Y^{(i),k-1}\right) = h\left(\tilde{U}_k^{(i)}\right) = \frac{1}{2} \log(2\pi e \lambda h) \quad (4.36)$$

and

$$\begin{aligned} h\left(Y_k^{(1)}, Y_k^{(2)} | Y^{(1),k-1}, Y^{(2),k-1}, X^k\right) &= h\left(Z_k^{(1)}, Z_k^{(2)}\right) = \frac{1}{2} \log(2\pi e |\Sigma_Z|) \\ &= \frac{1}{2} \log(2\pi e (\pi_s^2 h^2 (1 - \rho^2))). \end{aligned} \quad (4.37)$$

Thus, the achievable symmetric sum-rate is,

$$\begin{aligned} R_1 + R_2 &= \frac{1}{2} \log(2\pi e \lambda h) + \frac{1}{2} \log(2\pi e \lambda h) \\ &\quad - \frac{1}{2} \log(2\pi e (\pi_s^2 h^2 (1 - \rho^2))) \end{aligned} \quad (4.38)$$

$$= \frac{1}{2} \log \frac{\lambda^2 h^2}{\pi_s^2 h^2 (1 - \rho^2)} \quad (4.39)$$

$$= \log \frac{\lambda}{\pi_s} - \frac{1}{2} \log(1 - \rho^2). \quad (4.40)$$

4.5 Scalar Lower Bound Theorem

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

Theorem 4.3 (Characterization of $R_{\text{ZD,GM}}^I(D_0, DS)$)

Consider the stationary scalar $AR(1)$ process of (4.1). Given non-degenerate MSE distortion constraints, (D_S, D_0) , where $0 < D_0 \leq D_S \leq \Sigma_X$, the Gaussian information-theoretic symmetric ZDMD RDF, $R_{\text{ZD,GM}}^I(D_0, DS)$, is characterized by the solution to the following optimization problem.

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

where

$$\lambda = a^2 \pi_S + \sigma_W^2, \quad (4.41)$$

$$\pi_0 = \Sigma_X - \Sigma_{XY(0,C)}^2 \Sigma_{Y(0,C)}^{-1}, \quad (4.42)$$

and $\Sigma_{XY(0,C)}$, $\Sigma_{Y(0,C)}$ are defined in Lemma 4.2.

Remark 4.4 (Uniqueness of optimal solution)

We argue that the optimal solution to $(\mathbf{P}_{\text{scalar}})$ should be unique.

Firstly, the objective function in $(\mathbf{P}_{\text{scalar}})$ 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 non-degenerate 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 .

Thus, we conjecture the minimum to be unique. However, we have not yet been able to finally prove the uniqueness of the optimal solution to $(\mathbf{P}_{\text{scalar}})$.

This completes the theoretical work on the lower bound to Problem 1, as the solution to $(\mathbf{P}_{\text{scalar}})$. 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 $R_{\text{ZD}}^{\text{op}}(D_0, D_S)$, characterized by the (unique) solution to an optimization problem.

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

5 | Simple Quantization Scheme

This chapter introduces the simple ZDMD coding scheme of [33]. Although this scheme was not developed specifically for ZDMD only MD coding, it serves as a proof of concept, showing that we are able to achieve different rate-distortion pairs inside the ZDMD achievable region. We do not attempt to achieve the lower bound or derive an exact upper bound.

We give a short introduction to MD coding schemes in general, and introduce the concept of staggered quantization. We then introduce the coding scheme of [33] implemented in this report.

Through numerical simulations we show, it is possible to achieve operational performance within 3 dB to 5 dB of the theoretical lower bound in the high-rate scenario using this simple quantization scheme.

Test-channels in general provide a basis for the design of practical coding schemes. For example, the AWGN channels in Ozarow's test-channel may be replaced by quantizers producing quantization noise distributed similar to the test-channel noises. However, it is a non-trivial task to produce quantization noise with high negative correlation in practice [46].

One possibility is to use staggered scalar uniform quantization, that is two identical uniform quantizers, one per channel, with an offset of half a cell size [46]. Using staggered quantization it is possible to achieve a maximum noise correlation of $-1/2$ [50]. However, this correlation cannot increase any further, not even by using dithered quantizers with correlated dithers [50].

There exists some schemes, not designed by direct consideration of Ozarow's test-channel, that are able to achieve correlation that tends towards -1 [46]. The delta-sigma quantization scheme [53] uses multiple quantizers and noise shaping. However this scheme also requires up-sampling of the source signal, and is thus not useful for ZDMD coding. In the high-rate scenario and when the side- to central distortion ratio is high ($D_S/D_0 \gg 1$), so-called *index-assignments* achieve high correlation by high-rate quantization and a non-linear mapping, that maps each fine grid point to two descriptions [54]. However, it is not easy to design and adjust the redundancy in the index assignment scheme [55]. In [56] a modified MD scalar

quantizer scheme is developed using two-stage quantization with staggered quantizers. This scheme avoids the index-assignment problem and still achieves the same high-rate performance.

These schemes and many other MD coding schemes in general produce two descriptions with the desired correlation by direct operations on the source signal. Similar to Ozarow's test-channel where correlated noise is added directly to the source signal. However, our ZDMD test-channel forms two prediction error signals and adds test-channel noise to these signals. Thus, we form 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 [33] considers specifically MD predictive coding by extending upon the work of [56]. The very simple scheme in [33] is developed for the high-rate scenario and shows good performance. This scheme aligns well with our test-channel, since it performs staggered quantization of two prediction error process, and uses a refinement layer for further central distortion gain. We do not modify this scheme in any particularly way, but introduce it as a very simple proof of concept, showing we may achieve good operational performance in ZDMD coding using very simple techniques.

The scheme in [33] uses two DPCM schemes with staggered quantization and a second-stage refinement quantizer. Before introducing the scheme we give a short introduction to staggered quantization.

5.1 Staggered Quantizers

Consider a white Gaussian source process $X_k \sim \mathcal{N}(0, \sigma_X^2)$. The idea behind staggered quantizers is to overlap the quantization intervals of two identical uniform quantizers, such that the two staggered quantizer outputs may be combined to refine each other.

Let,

$$Q_{1,\Delta}(x) = \text{Round} \left(\frac{x}{\Delta} \right) \cdot \Delta \quad (5.1)$$

$$Q_{2,\Delta}(x) = \text{Round} \left(\frac{x + \delta}{\Delta} \right) \cdot \Delta - \delta, \quad (5.2)$$

be two staggered scalar uniform quantizers, where Δ is the step size of the two quantizers, and δ is the offset between the quantizers. Denote by,

$$Y_k^{(i)} = Q_{i,\Delta}(X_k), \quad i = 1, 2, \quad (5.3)$$

the output of the two quantizers for the source input X_k . Then the resulting quantization intervals and their overlap are shown in Figure 5.1.

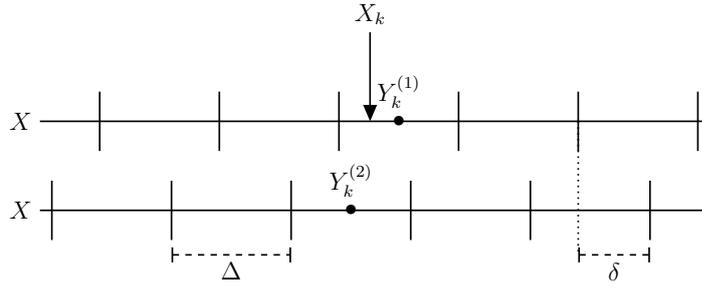


Figure 5.1: Quantizer partitions for X_k using two staggered uniform quantizers.

For sufficiently high rates, (i.e. small Δ), the source is approximately uniformly distributed within each quantizer cell [10]. Hence, the optimal reconstruction of X_k using the two staggered quantizer outputs is $\frac{1}{2}(Y_k^{(1)} + Y_k^{(2)})$, i.e. the midpoint of the intersection between the cells that X_k belongs to [33]. Let

$$E_{C,k} = X_k - \frac{1}{2}(Y_k^{(1)} + Y_k^{(2)}), \quad (5.4)$$

be the error signal when using the midpoint reconstruction. Then, under the high rate assumption the MSE of this central reconstruction is [33]

$$\mathbb{E}[E_{C,k}^2] = \frac{\delta}{\Delta} \frac{\delta^2}{12} + \frac{\Delta - \delta}{\Delta} \frac{(\Delta - \delta)^2}{12} \quad (5.5)$$

$$= \frac{1}{12\Delta} (\delta^3 + (\Delta - \delta)^3) \quad (5.6)$$

The MSE is then minimized when the quantizer overlap is $\delta = \Delta/2$ [33]. Particularly, the MMSE is

$$D_0 = \mathbb{E}[E_{C,k}^2] = \frac{1}{4} \frac{\Delta^2}{12} = \frac{1}{4} D_S \quad (5.7)$$

where D_S is the MSE of the first stage quantizer errors,

$$E_k^{(i)} = X_k - Y_k^{(i)}, \quad i = 1, 2, \quad (5.8)$$

that is,

$$D_S = \mathbb{E}\left[\left(E_k^{(i)}\right)^2\right] = \frac{\Delta^2}{12}. \quad (5.9)$$

Thus, using the central reconstruction of the two staggered quantizers there is a gain of $D_S/D_0 = 6$ dB compared to the individual reconstructions [50]. This, shows, the advantage of staggered quantization, since this gain is achieved without spending any extra bits on improving the central reconstruction.

Finally, it can be shown that the correlation of the quantization noises, $E_k^{(i)}$ $i = 1, 2$ is $\rho = -1/2$ [50].

5.2 The Scheme of [33]

The scheme in [33] is derived for the main source process of consideration in this report. That is, the scalar stationary first-order Gauss-Markov sources of the form,

$$X_k = aX_{k-1} + W_k, \quad k \in \mathbb{N} \quad (5.10)$$

where $|a| < 1$ is the correlation coefficient of the process, $X_1 \in \mathbb{R} \sim \mathcal{N}(0, \sigma_{X_1}^2)$ is the initial state with $\sigma_{X_1}^2 = \frac{\sigma_W^2}{1-a^2}$, and $W_k \in \mathbb{R} \sim \mathcal{N}(0, \sigma_W^2)$, is an IID Gaussian sequence independent of X_1 .

The main idea of [33] 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 .

Since quantization in the DPCM encoders is applied to the prediction error signals, $U_k^{(1)}$, $U_k^{(2)}$, and not on X_k , simple staggered quantizers will not generate the desired partitions for X_k . This happens because

$$X_k = U_k^{(i)} + \widehat{X}_{k|k-1}^{(i)}, \quad i = 1, 2, \quad (5.11)$$

results in the quantizer partitions for X_k and $U_k^{(i)}$ being shifted by the DPCM prediction $\widehat{X}_{k|k-1}^{(i)}$ [33]. Hence, the prediction step must be designed along with the quantizers [33].

For a given prediction $\widehat{X}_{k|k-1}^{(i)}$ the resulting partitions on X_k should be as in Figure 5.1. That is, with optimal quantization bin overlap of $\Delta/2$. We can guarantee this if the predictions satisfy $X_{k|k-1}^{(1)} - X_{k|k-1}^{(2)} = \pm\Delta/2$ [33]. Therefore, the solution proposed in [33] is to use the sub-optimal predictors $Y_{k-1}^{(i)}$ instead of the optimal predictors $aY_{k-1}^{(i)}$, i.e. letting

$$U_k^{(i)} = X_k - Y_{k-1}^{(i)}, \quad i = 1, 2, \quad (5.12)$$

as seen in Figure 5.2. Then $Y_{k-1}^{(1)} - Y_{k-1}^{(2)} = \pm\Delta/2$, since $X_{k|k-1}^{(1)} = Y_{k-1}^{(1)}$. Thus, if the offset is optimal at time $k-1$ it will be optimal at time k [33]. Therefore, the optimal shifted quantizer need only be used in the first time step if we let $Y_0^{(i)} = 0$, $i = 1, 2$, and then use non-staggered identical uniform quantizers for all $k > 1$ [33].

5.2.1 Optimal Predictors

If we use the optimal predictors in the scheme of [33] and want to maintain $Y_k^{(1)} - Y_k^{(2)} = \pm\Delta/2$ for all k , then the quantizers must be offset by

$$\delta = \text{sign} \left(Y_{k-1}^{(1)} - Y_{k-1}^{(2)} \right) \frac{(1-a)\Delta}{2}, \quad (5.13)$$

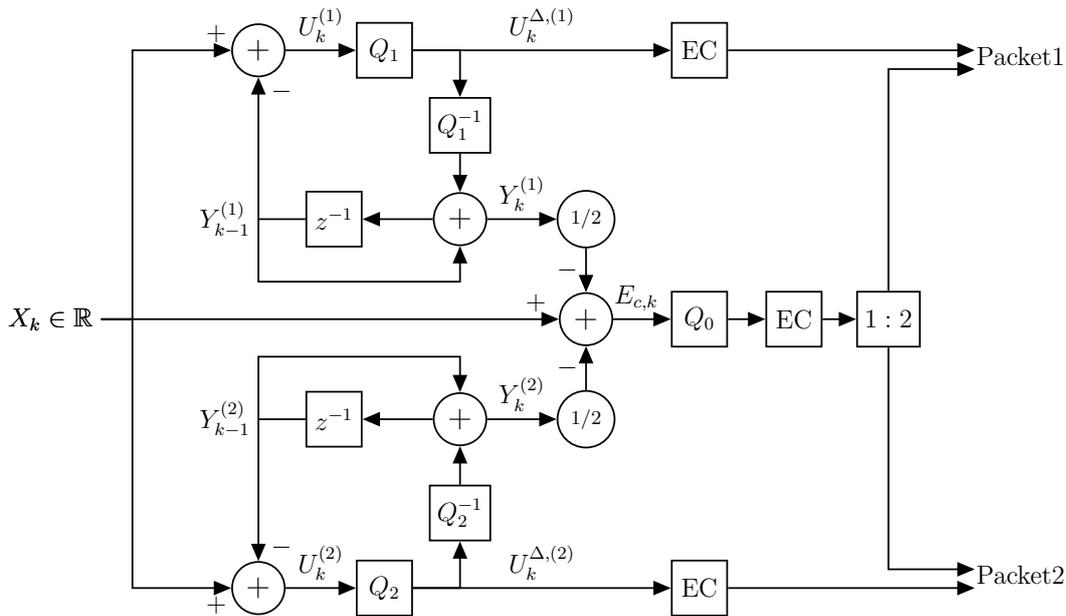


Figure 5.2: The two-stage staggered DPCM quantization scheme. The two first-stage quantizers Q_1 and Q_2 are staggered identical uniform quantizers. Here EC denotes lossless (entropy) encoders. The binary description packets are formed by entropy coding each side quantizer output and splitting the entropy coded second stage quantizer output in two. Figure modified from [33].

for all k [33]. This may be achieved by fixing the partition for Q_1 and shifting that of Q_2 for each k [33].

However, this results in a 1 bit increase in the entropy of $U_k^{\Delta,(2)}$, i.e. the quantized $U_k^{(2)}$. As remarked in [33]; if only the second description is available, $Y_{k-1}^{(1)}$ is unknown and an extra bit about the sign of $Y_{k-1}^{(1)} - Y_{k-1}^{(2)}$ must be included in the second description to reconstruct $U_k^{(2)}$ and $Y_k^{(2)}$ from $U_k^{\Delta,(2)}$. This will significantly reduce the coding efficiency of the method, especially at low rates.

This added extra bit makes it impossible to have symmetric rates. Although the side distortions are the same. Furthermore, since ZD source coding already suffers from an increased bitrate we do not use the optimal predictors. Using either prediction (non-optimal and optimal) and the appropriate shift the resulting quantizer noise correlation is $-1/2$. If we do not shift the partition for each k in the optimal prediction scenario, the resulting correlation is approximately zero.

5.2.2 Prediction Error Covariance

For high-rate quantization the source is approximately uniformly distributed within each quantization cell [10]. Therefore, it is common to model the quantization error

as a uniform noise independent of the quantizer input [10], i.e.

$$U_k^{(i)} - Q_{i,\Delta_S}(U_k^{(i)}) = Z_k^{(i)} \sim \mathcal{U}\left(-\frac{\Delta_S}{2}, \frac{\Delta_S}{2}\right), \quad (5.14)$$

where we denote the first-stage quantizer intervals by Δ_S . Therefore, for high rates the error in the first stage quantizers is uniform with variance,

$$\Sigma_{Z_S} = \frac{\Delta_S^2}{12}. \quad (5.15)$$

However, this model is strictly speaking untrue for deterministic quantization [10], i.e. with no dithering. Using dithered quantization this model can be shown to be accurate for any quantization resolution [10]. Although the effect of the dither subsides for higher rates [10].

The uniform additive noise model provides a great analysis tool for determining covariances for different signals in the quantization scheme. Since the reproductions may be modeled as,

$$\begin{aligned} Y_k^{(i)} &= Q_{i,\Delta_S}(U_k^{(i)}) + Y_{k-1}^{(i)} \\ &= U_k^{(i)} + Y_{k-1}^{(i)} + Z_k^{(i)} \\ &= X_k - Y_{k-1}^{(i)} + Y_{k-1}^{(i)} + Z_k^{(i)} \\ &= X_k + Z_k^{(i)}, \end{aligned} \quad (5.16)$$

and

$$D_S = \mathbb{E}\left[\left(X_k - Y_k^{(i)}\right)^2\right] = \frac{\Delta_S^2}{12}. \quad (5.17)$$

Our test-channel (4.2) recovers this model by using the non-optimal predictor (i.e. letting $a = 1$) and not using a pre-scaled channel (i.e. letting $h = 1$).

The prediction error variance is,

$$\begin{aligned} \lambda &= \text{Var}\left[U_k^{(i)}\right] = \text{Var}\left[X_k - Y_{k-1}^{(i)}\right] \\ &= \text{Var}\left[X_k - \left(X_{k-1} + Z_{k-1}^{(i)}\right)\right] \\ &= \text{Var}\left[aX_{k-1} + W_k - \left(X_{k-1} + Z_{k-1}^{(i)}\right)\right] \\ &= (a-1)^2\sigma_X^2 + \sigma_W^2 + \Sigma_{Z_S} \\ &= \frac{2}{1+a}\sigma_W^2 + \frac{\Delta_S^2}{12}, \end{aligned} \quad (5.18)$$

where the last equality follows by the stationary variance of X_k . Finally, the central decoder reconstructs X_k as

$$Y_k^{(0)} = Q_{0,\Delta_0}(E_{C,k}) + \frac{1}{2}\left(Y_k^{(1)} + Y_k^{(2)}\right), \quad (5.19)$$

where Δ_0 is the quantizer bin size of the second-stage quantizer. For high rate quantization, since the input to the second-stage quantizer is approximately uniform with variance [33]

$$\sigma_{E,C}^2 = \frac{D_S}{4}. \quad (5.20)$$

5.2.3 Choosing step size

Before conducting a simulation study using this quantization scheme, we highlight how the quantization bin sizes relate to the description rates.

Let R_S be the rate of the first-stage quantizers, Q_1 , Q_2 , and R_0 be the rate of the second-stage (central) quantizer, Q_0 . Then the average sum-rate, or rate per description, R , is given as

$$R = R_S + \frac{R_0}{2}. \quad (5.21)$$

This also provides an intuitive interpretation of no excess marginal rate as the case when $R_0 = 0$. Thus, the transmitted packets contain no excess bits other than the side-description bits.

For high rate quantization of the prediction error signal $U_k^{(i)}$, the coding rates are given by the discrete entropy of the quantized prediction errors [46], i.e.

$$R_S = H(U^{\Delta_S, (i)}) \approx h(U^{(i)}) - \log \Delta_S, \quad (5.22)$$

where $U^{\Delta_S, (i)}$ is the quantized version of $U^{(i)}$, and the last approximation follows from Lemma A.23. Inserting, the variance of $U^{(i)}$, λ , we may then isolate Δ_S for a desired rate R_S . The resulting side distortion is then

$$\hat{D}_S = \frac{\Delta_S^2}{12}. \quad (5.23)$$

Similarly for a desired second stage quantizer rate R_0 ,

$$R_0 = H(E_C^{\Delta_0}) \approx h(E_C) - \log \Delta_0. \quad (5.24)$$

Isolating the quantizer step-size, Δ_0 ,

$$\Delta_0 \approx 2^{-R_0} \sqrt{12\sigma_{E_C}^2} = 2^{-R_0} \sqrt{3\hat{D}_S}, \quad (5.25)$$

since E_C is uniform with variance $\frac{\hat{D}_S}{4}$. Thus, for high rates the resulting central distortion is

$$\hat{D}_0 = \frac{\Delta_0^2}{12} = 2^{-2R_0} \frac{\hat{D}_S}{4}. \quad (5.26)$$

5.3 Simulation Study

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

In all simulations we consider stationary scalar Gauss-Markov sources of the form (5.10). All simulations are conducted by fixing the rate per description R . Then, for each rate-pair, R_S, R_0 , satisfying the rate constraint R the practical quantizer step sizes are determined according to (5.22) and (5.25), such that the operational rate per description, R_{op} , is approximately equal to the constraint, i.e. $R_{op} \approx R$. 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 (5.22) and (5.25). 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.

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

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

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

where $Y_k^{(i)}$ $i = 0, 1, 2$ are the reconstructions for the k th input sample X_k , and are obtained according to (5.19) and (5.16). The operational coding rates are determined by the discrete entropies

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

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

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

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 $(\mathbf{P}_{\text{scalar}})$, i.e. letting

$$R = \frac{1}{2} \log \frac{a^2 \pi_S + \sigma_W^2}{\pi_S} - \frac{1}{4} \log (1 - \rho_0^2) \quad (5.27)$$

and determining the corresponding ρ_0 and central distortion π_0 for a grid of side distortions, π_S .

Table 5.1: Simulation Parameters for distortion trade-off curve in Figure 5.3.

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

5.3.1 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 5.1. The resulting theoretical and operational distortion curves are shown in Figure 5.3. 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, if we decrease the central distortion we must increase the side distortion, and vice-versa, if we want to main the same rate R . This shows, we are able to trade-off between the side- and central distortion by varying the bit allocation in the first- and second stage quantizers.

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 of 0.254 bits/sample per quantizer. Thus, the refinement scheme suffers from the space-filling loss of three quantizers [50]. Furthermore, there is a loss due to the non-optimal linear predictors, however this loss is minimal in the high-rate scenario [33].

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.

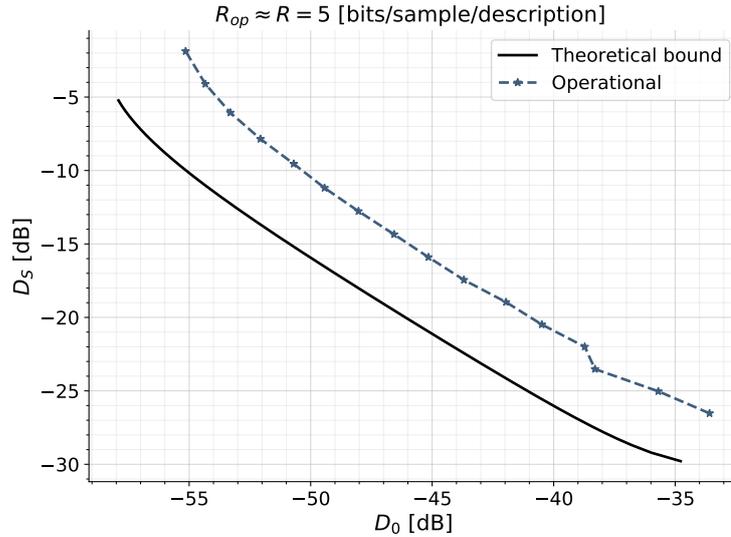


Figure 5.3: The central distortion, D_0 , versus side distortion, D_S for ZDMD coding of a GM(1) source with $a = 0.9$ and unit variance at $R = 5$ bits/sample/description. Simulation parameters in Table 5.1.

5.3.2 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 obtain the distortion curves in Figure 5.4, the simulation parameters are listed in Table 5.2. Figure 5.4a shows the side distortion, D_S , in relation to the distortion-ratio γ for varying rates, similarly Figure 5.4b 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.

For any particular rate and distortion-ratio in Figure 5.4, the central distortion, D_0 , is always lower than side distortion, D_S . Also 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 5.3. Figure 5.4 shows, this trend is independent of the rate.

Furthermore, the plots in Figure 5.4 show, that by increasing the rate per description for any fixed ratio, we can increase the performance in both central- and side distortion.

The maximum operational distortion ratio is limited to approximately $1/4$. Since by (5.26) at no excess marginal rate, i.e. when $R_0 = 0$, we have that $D_0/D_S \approx 1/4$. Hence, to evaluate higher distortion-ratios we would need to perform non-optimal central reconstructions, or decrease the quantizer offsets away from the optimum

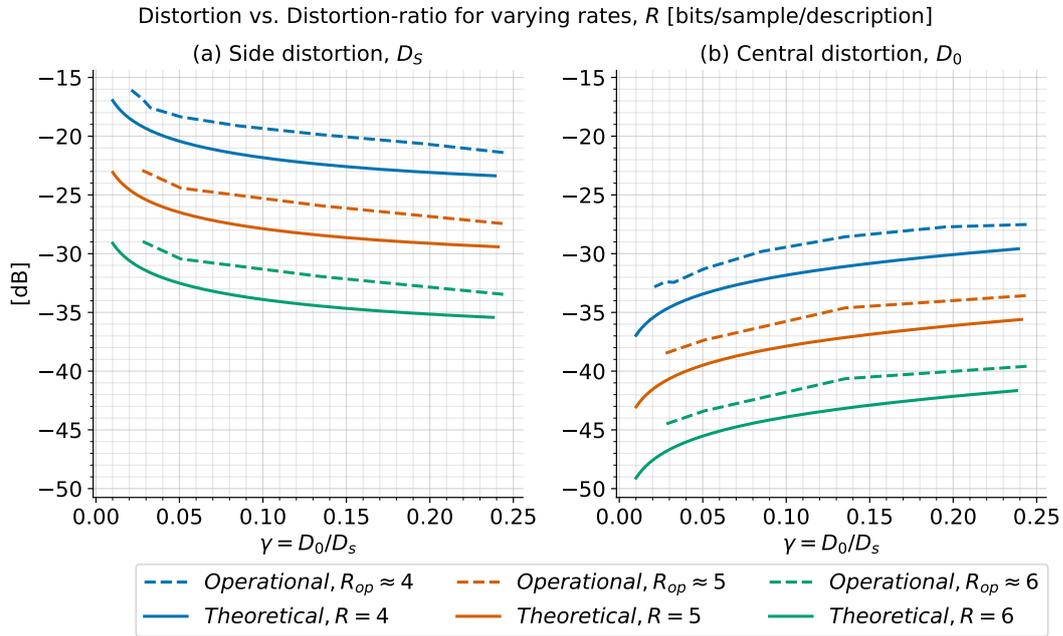


Figure 5.4: (a) Side distortion, D_S , and (b) central distortion, D_0 , versus distortion-ratio D_0/D_S for ZDMD coding of a GM(1) source with $a = 0.9$ and unit variance at $R \in \{4, 5, 6\}$ bits/sample/description. Simulation parameters in Table 5.2.

Table 5.2: Simulation Parameters for distortion versus distortion-ratio curves in Figure 5.4.

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, 4} bits/sample
Time samples	N	500 000
Monte-Carlo simulations	M	4

Table 5.3: Two operational- and theoretical distortion pairs for a fixed rate per description $R = 5$ bits/sample.

Operational			Theoretical		
\hat{D}_0	\hat{D}_S	$\hat{\gamma}$	D_0	D_S	γ
-36 dB	-25 dB	0.08	-36.0 dB	-30.0 dB	0.25
-36 dB	-25 dB	0.08	-38.5 dB	-27.5 dB	0.08

half bin size.

For a given rate and distortion-ratio, the operational curves in Figure 5.4 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, this loss seems to be half of that seen when plotting D_S versus D_0 in Figure 5.3. However, for a given ratio there are two curves in Figure 5.4, 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. To illustrate this, we consider for $R = 5$ bits/sample the two operational and theoretical distortions pairs in Table 5.3. For a fixed operational distortion pair (\hat{D}_0, \hat{D}_S) and thereby fixed $\hat{\gamma}$, we consider corresponding different theoretical distortions. If we let $D_0 = \hat{D}_0$, there is a distortion loss of 5 dB for the side distortion, D_S . However, to achieve $\gamma = \hat{\gamma}$ the previously considered theoretical distortions must both change because of the distortion loss, and we see a distortion loss of 2.5 dB for each of D_0 and D_S .

From the rate-distortion performances in Figures 5.3 and 5.4, we see for high rates that the simple quantization scheme of [33] is able to achieve performance close to the theoretical ZDMD lower bounds derived in the previous chapters. Hence, we are able operate along the theoretical bounds for ZDMD coding of stationary scalar Gauss-Markov sources using simple techniques. Particularly, we are able to trade-off both rates and distortions. This provides proof of concept, that it is possible to derive operational quantization schemes that may achieve performance close to our lower bounds. The simulation results also provide indication of an upper bound on the optimal operational performance limits of ZDMD coding of stationary scalar Gauss-Markov sources.

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. Initially we consider the delimitation to symmetric distortions.

Symmetric Distortions

By restricting our research to symmetric ZDMD coding we are able to define a ZDMD RDF. For the more general case of asymmetric rates this is not possible, and we must consider a rate- and distortion-region. However, by restriction our research to the symmetric case we have derive novel indicative results. Thus, we solve a simpler initial problem, that may then more easily be extended to the general asymmetric case.

Theoretical Lower Bound

In order to derive an information theoretic lower bound on the symmetric ZDMD RDF for scalar stationary Gaussian sources in Theorem 3.10, we have made some technical assumptions.

Sequential Greedy Coding

The main assumption was the use of sequential greedy coding (Definition 3.8). This implies, 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 spend 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, thus their respective reconstruction distributions are selected only once. However, this remains an open problem for future research.

Conditional Prediction Residual Independence

The second technical assumption for the proof of Theorem 3.10 was conditional prediction residual independence (Definition 3.9). This assumption states that the reconstruction MMSE prediction residuals are independent of the prediction variables. By the orthogonality principle the MMSE residuals are uncorrelated with the prediction variables [51], [57]. For Gaussian variables this implies independence [58]. However, for non-Gaussian reconstructions this might also result in a rate loss, since we limit the MMSE predictors to those that have prediction residual independence.

Theorem 3.10 provides a lower bound on

$$\bar{I}_\infty \left(X \rightarrow Y^{(1)}, Y^{(2)} \right) + \bar{I} \left(Y^{(1)}; Y^{(2)} \right), \quad (5.28)$$

for a Gaussian source process subject to MSE distortion constraints. The reason for the previous assumptions is to minimize the excess information rate $\bar{I}(Y^{(1)}; Y^{(2)})$, and show the reconstructions, $Y^{(1)}$, $Y^{(2)}$, should be jointly Gaussian, when they are jointly Gaussian with source. The main intuition behind disregarding the assumptions of sequential greedy coding and conditional residual independence follows from the results of [8]. Since by the results of [8] we have for a Gaussian source process $\{X_k\}$,

$$\bar{I}_\infty \left(X \rightarrow Y^{(1)}, Y^{(2)} \right) \geq \bar{I}_\infty \left(X \rightarrow Y_G^{(1)}, Y_G^{(2)} \right), \quad (5.29)$$

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 in the second term of (5.28). However, we have not able to prove this.

Independent Side Information

To derive the information-theoretic lower bound on the sum-rate for the MD coding problem with feedback in Figure 3.1, 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, $S_{\mathcal{D}_2}^k$,

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 3.2 warrant further investigation. Although, for common side information it seems reasonable that the bounds should remain widely unchanged.

For staggered dithered quantizers, the dithers must be dependent to maintain correct partitions [50]. Hence, violating the assumption of mutually independent side information. However, for dithered quantization the encoder side information also includes the dither signal [5], hence all side information is available at all decoders and the encoder. Thus, no excess information about the side information needs to be transmitted. If this is not the case bounds may be formulated such that the dependency on the side information is explicit. In [59] 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 [59] are similar to those of El-Gamal and Cover [24] with an added dependency upon the unknown side information in the involved mutual informations. Hence, this results could provide a basis for extending the results of Section 3.2 to the case of unknown- or dependent side information.

Simulation Results

The main design problem of MD coding is determining the trade-off between spending bits on decreasing the side distortion or the central distortion [22]. This is apparent in the operational scheme in Figure 5.2, where rate-splitting is used at the second-stage quantizer to generate the two data packets. When only one description is received this rate-splitting results in excess bits going waste, since these cannot be used without the other description. This technique is also known as unequal error protection (UEP) [22].

Our results show that we are able to provide this trade-off also in ZDMD coding, while achieving performance near the theoretical limits. For MD coding of white Gaussian sources, with fixed side distortion and no delay constraints, the maximum difference between the UEP achievable region and the MD region is approximately 4.18 dB [22]. For ZDMD coding of Gaussian sources with memory in the high-rate scenario, we achieve operational performance within 5 dB of theoretical lower bounds. Thus, using the simple scheme in [33] for ZDMD coding, we achieve performances within reason of the theoretical limits. However, further improvements must be made to derive a quantization scheme for the low-rate scenario.

Test-Channel and Operational Scheme Design

We note a few important remarks relating our test-channel in Figure 4.1 to the practical quantization scheme in Figure 5.2. This discussion also highlights some of the difficulties in generalizing our test-channel to a practical quantizer design.

In order to fully determine the operational symmetric ZDMD RDF, $R_{\text{ZD}}^{\text{op}}(D_0, D_S)$, we must show the information-theoretic lower bound, $R_{\text{ZD}}^{\text{l}}(D_0, D_S)$ can be achieved by an operational quantization scheme. The usual rate-distortion technique is to derive an explicit upper bound on $R_{\text{ZD}}^{\text{op}}(D_0, D_S)$ using $R_{\text{ZD}}^{\text{l}}(D_0, D_S)$ then showing these bounds coincide in an asymptotic sense [8].

To this end, we have derived a test-channel showing the achievability of the lower bound in a Gaussian coding scheme. However, we have not shown the Gaussian test-channel distributions are achievable by an operational quantization scheme. Particularly, we have not yet been able to derive an exact upper bound.

Instead, to show proof of concept, we employed an already existing quantization scheme from the literature [33]. This scheme was selected based on its simplicity and resemblance to our test-channel, since both “schemes” use two closed-loop predictions to remove dependence across time, and then use correlated “noise” to encode the resulting error-process. However, the AWGN is only a model for the desired distribution. Since it is difficult to generate noise with the desired correlation, the AWGN model loses some interpretability when relating to the quantization scheme in Figure 5.2. This is especially apparent in the central decoder design, and when determining an exact upper bound.

Central Reconstruction Design

In the operational quantization scheme we do not add correlated quantization noise in the sense of the test-channel. Instead, the quantizer outputs are designed to achieve a performance corresponding to a specific correlation in the test-channel, e.g. staggered bins, that increase the central reconstruction performance by using the appropriate reconstruction method. This performance equals that of using noise with a correlation of $-1/2$ [50]. To further improve the performance a second refinement layer is added. Thus, achieving performance approximating that of arbitrary correlation. Additionally, since we use deterministic quantizers (no dithering) the quantizer outputs are deterministic for a given input.

Furthermore, the quantizers are designed such that the partition on the source, X , is staggered, not the partitions on the prediction errors. Hence, the “correlated noise” loses further interpretability. At the beginning of Section 4.3 on page 61, we argue that because the innovation processes are created by addition of correlated noise, the central decoder may remove some of this noise, since it receives both descriptions. However, this argument loses its merit in the operational scheme,

since the two quantizers are not designed such that the quantized prediction errors, $U_k^{\Delta s, (i)}$, $i = 1, 2$, improve upon each other. Instead these are designed to improve the reconstruction of X_k . Especially, as shown previously, by construction it is optimal to take an average of $Y_k^{(i)}$ $i = 1, 2$ at the central decoder.

Therefore, for this particular quantization scheme, there is no gain in using the central decoder design in Figure 4.1b to reduce the “noise” added to the prediction error processes. Additionally, at high rates the optimal decoder design in the test-channel has a limited effect compared to just taking an average of $Y_k^{(i)}$, $i = 1, 2$, at the central decoder, and performing a MMSE estimate of X_k given $\frac{1}{2}(Y_k^{(1)} + Y_k^{(2)})$. To see this, we note for high rates, as $\pi_S \rightarrow 0$ then $\Theta_U \rightarrow 1$ in (4.22), which further implies $Y_k^{(0,C)} \rightarrow \frac{1}{2}(Y_k^{(1)} + Y_k^{(2)})$ in (4.27). Therefore, for high rates averaging the side reconstructions and scaling by the appropriate Wiener coefficient is optimal also in the test-channel derivations in Section 4.3. Thus, as shown by the simulation results, a ZDMD quantization scheme for general resolution, may be designing by a more simple approach not relying heavily on the test-channel of Figure 4.1b. However, at lower resolutions the non-optimal predictors result in larger distortion penalties, and other ways of generating staggered partitions might be advantageous.

Upper Bound

The previous section highlights some of the inherent difficulty in designing quantization schemes that achieve a desired performance. For single-description ZD coding the operational realization scheme of [8] replaces the AWGN channel in Figure 1.4 with an entropy coded dithered scalar quantizer (ECDQ) [10], where the quantization bin size is chosen according to the optimum test-channel distribution. Since the ZDMD test-channel is not easily converted to an operational quantization scheme by replacing Gaussian noises with uniform noises, further work is needed to derive upper bounds on the achievable performance. That is, since we do not simply replace the AWGN channels with ECDQs, the space-filling- and entropy coding losses suffered in scalar uniform quantizers are not easily shown to be the upper bounding factors. Therefore, modifying the test-channel by either the differential form of Ozarow’s test-channel [23] or using three AWGN channels [44], [50] to generate the desired distribution might prove beneficial. Specifically the results in [23] show, that for MD coding of colored Gaussian sources, without delay restrictions, it is possible to separate the part responsible for exploiting memory (DPCM) from the part controlling the MD coding parameters, i.e. noise shaping. However, the scheme in [23] achieves this by up-sampling the source signal, hence modifications are needed to extend this result to ZDMD coding.

Conclusion

We have determined initial results combining ZD- and MD coding theory, with the extended notion of feedback from the decoders.

In this work we studied the open-loop ZDMD source coding problem with perfect decoder-feedback and side information available to both encoder and decoder. Using this constructive system, we show in a novel result, 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. This provides a novel relation between information theory and the operational ZDMD coding rates.

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

Furthermore, we show 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, 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 [33], resembling our test-channel to some extent. Using this simple quantization scheme, it is possible to achieve operational distortions within 5 dB of the theoretical lower bounds for varying description rates.

Future Research

We have shown several indicative results on ZDMD rate-distortion theory with feedback. However, many interesting and unexplored directions remain.

The main future direction for ZDMD coding involves generalization of the main results in Theorem 3.10 and the test-channel to Gaussian vector sources.

For Gaussian vector sources, early work by the authors indicate that the test-channel in Chapter 4 may be generalized to the vector case in a similar manner to that of [8]. However, much work still remains in determining proper test-channel noise correlation. Furthermore, the novel result regarding ZDMD of scalar process in Theorem 3.10 must be extended to the vector case.

We speculate the sequential greedy coding condition is only technical for the proof of Theorem 3.10. However, this has not yet been proven.

We have only derived an achievable lower bound in a Gaussian coding scheme. Therefore, it remains an open problem to show operational achievability of the lower bound. Hence, deriving an operational scheme that achieves the lower bound asymptotically, or provides an exact upper bound on the operational performance is an future research area.

Extension of our results to asymmetric ZDMD coding is also of great interest, since many scenarios involving MDs use asymmetric rates and distortions.

ZD coding is of particular importance in NCSs. Therefore, it is interesting to extend our results to closed-loop systems. Here the goal would be to determine the fundamental performance limitations of controlled processes under communication constraints using unreliable channels. To this end it seems reasonable to first determine performance limits of ZDMD with packet losses.

Bibliography

- [1] J. Gubbi, R. Buyya, S. Marusic, and M. Palaniswami, "Internet of things (IoT): A vision, architectural elements, and future directions," *Future Gener. Comput. Syst.*, vol. 29, no. 7, pp. 1645–1660, Sep. 2013, ISSN: 0167-739X. DOI: 10.1016/j.future.2013.01.010.
- [2] J. Østergaard, D. E. Quevedo, and J. Jensen, "Low delay moving-horizon multiple-description audio coding for wireless hearing aids," in *2009 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2009, pp. 21–24. DOI: 10.1109/ICASSP.2009.4959510.
- [3] H. Krueger and P. Vary, "A new approach for low-delay joint-stereo coding," in *ITG Conference on Voice Communication [8. ITG-Fachtagung]*, 2008, pp. 1–4.
- [4] G. Schuller, J. Kovacevic, F. Masson, and V. K. Goyal, "Robust low-delay audio coding using multiple descriptions," *IEEE Transactions on Speech and Audio Processing*, vol. 13, no. 5, pp. 1014–1024, 2005, ISSN: 1063-6676. DOI: 10.1109/TSA.2005.853205.
- [5] E. I. Silva, M. S. Derpich, and J. Østergaard, "A framework for control system design subject to average data-rate constraints," *IEEE Transactions on Automatic Control*, vol. 56, no. 8, pp. 1886–1899, 2011, ISSN: 0018-9286. DOI: 10.1109/TAC.2010.2098070.
- [6] S. C. Tatikonda, "Control under communication constraints," PhD thesis, Massachusetts Institute of Technology. Dept. of Electrical Engineering and Computer Science, Cambridge, MA, 2000.
- [7] S. Tatikonda, A. Sahai, and S. Mitter, "Stochastic linear control over a communication channel," *IEEE Transactions on Automatic Control*, vol. 49, no. 9, pp. 1549–1561, 2004, ISSN: 0018-9286. DOI: 10.1109/TAC.2004.834430.
- [8] P. A. Stavrou, J. Østergaard, and C. D. Charalambous, "Zero-delay rate distortion via filtering for vector-valued gaussian sources," *IEEE Journal of Selected Topics in Signal Processing*, vol. 12, no. 5, pp. 841–856, 2018, ISSN: 1932-4553. DOI: 10.1109/JSTSP.2018.2855046.

- [9] C. E. Shannon, "A Mathematical Theory of Communication," *Bell System Technical Journal*, vol. 27, no. 3, pp. 379–423, 1948.
- [10] R. Zamir, *Lattice Coding for Signals and Networks*. Cambridge University Press, 2014, ISBN: 0-521-76698-2.
- [11] R. Veldhuis and M. Breeuwer, *An Introduction to Source Coding*. Prentice Hall, 1993, ISBN: 0-13-489089-2.
- [12] T. Berger, *Rate distortion theory : A mathematical basis for data compression*, eng, ser. Prentice-Hall series in information and system sciences. Englewood Cliffs N.J.: Prentice Hall, 1971, ISBN: 0137531036.
- [13] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, Wiley Student Edition. Wiley-Interscience, 2006, ISBN: 0471241954.
- [14] R. Zamir, "Gaussian codes and shannon bounds for multiple descriptions," *IEEE Transactions on Information Theory*, vol. 45, no. 7, pp. 2629–2636, 1999, ISSN: 0018-9448. DOI: 10.1109/18.796418.
- [15] P. A. Stavrou, J. Østergaard, and M. Skoglund, "On zero-delay source coding of LTI gauss-markov systems with covariance matrix distortion constraints," in *2018 European Control Conference (ECC)*, 2018, pp. 3083–3088. DOI: 10.23919/ECC.2018.8550204.
- [16] M. S. Derpich and J. Østergaard, "Improved upper bounds to the causal quadratic rate-distortion function for gaussian stationary sources," in *2010 IEEE International Symposium on Information Theory*, 2010, pp. 76–80. DOI: 10.1109/ISIT.2010.5513282.
- [17] T. Tanaka, K. K. Kim, P. A. Parrilo, and S. K. Mitter, "Semidefinite programming approach to gaussian sequential rate-distortion trade-offs," *IEEE Transactions on Automatic Control*, vol. 62, no. 4, pp. 1896–1910, 2017, ISSN: 0018-9286. DOI: 10.1109/TAC.2016.2601148.
- [18] E. I. Silva, M. S. Derpich, J. Østergaard, and M. A. Encina, "A characterization of the minimal average data rate that guarantees a given closed-loop performance level," *IEEE Transactions on Automatic Control*, vol. 61, no. 8, pp. 2171–2186, 2016, ISSN: 0018-9286. DOI: 10.1109/TAC.2015.2500658.
- [19] M. Barforooshan, J. Østergaard, and P. A. Stavrou, "Achievable performance of zero-delay variable-rate coding in rate-constrained networked control systems with channel delay," in *2017 IEEE 56th Annual Conference on Decision and Control (CDC)*, 2017, pp. 5991–5996. DOI: 10.1109/CDC.2017.8264566.
- [20] D. Neuhoff and R. Gilbert, "Causal source codes," *IEEE Transactions on Information Theory*, vol. 28, no. 5, pp. 701–713, 1982, ISSN: 0018-9448. DOI: 10.1109/TIT.1982.1056552.

- [21] J. Massey, "Causality, feedback and directed information," in *Proc. Int. Symp. Inf. Theory Applic. (ISITA-90)*, Citeseer, 1990, pp. 303–305.
- [22] V. K. Goyal, "Multiple description coding: Compression meets the network," *IEEE Signal Processing Magazine*, vol. 18, no. 5, pp. 74–93, 2001, ISSN: 1053-5888. DOI: 10.1109/79.952806.
- [23] J. Østergaard, Y. Kochman, and R. Zamir, "Colored-gaussian multiple descriptions: Spectral and time-domain forms," *IEEE Transactions on Information Theory*, vol. 62, no. 10, pp. 5465–5483, 2016, ISSN: 0018-9448. DOI: 10.1109/TIT.2015.2513773.
- [24] A. E. Gamal and T. Cover, "Achievable rates for multiple descriptions," *IEEE Transactions on Information Theory*, vol. 28, no. 6, pp. 851–857, 1982, ISSN: 0018-9448. DOI: 10.1109/TIT.1982.1056588.
- [25] L. Ozarow, "On a source-coding problem with two channels and three receivers," *The Bell System Technical Journal*, vol. 59, no. 10, pp. 1909–1921, 1980, ISSN: 0005-8580. DOI: 10.1002/j.1538-7305.1980.tb03344.x.
- [26] P. L. Dragotti, S. D. Servetto, and M. Vetterli, "Optimal filter banks for multiple description coding: Analysis and synthesis," *IEEE Transactions on Information Theory*, vol. 48, no. 7, pp. 2036–2052, 2002, ISSN: 0018-9448. DOI: 10.1109/TIT.2002.1013142.
- [27] J. Chen, C. Tian, and S. Diggavi, "Multiple description coding for stationary gaussian sources," *IEEE Transactions on Information Theory*, vol. 55, no. 6, pp. 2868–2881, 2009, ISSN: 0018-9448. DOI: 10.1109/TIT.2009.2018178.
- [28] M. S. Mehmetoglu, E. Akyol, and K. Rose, "Analog multiple descriptions: A zero-delay source-channel coding approach," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2016, pp. 3871–3875. DOI: 10.1109/ICASSP.2016.7472402.
- [29] Y. Kochman and R. Zamir, "Analog matching of colored sources to colored channels," *IEEE Transactions on Information Theory*, vol. 57, no. 6, pp. 3180–3195, 2011, ISSN: 0018-9448. DOI: 10.1109/TIT.2011.2132950.
- [30] J. Østergaard, D. E. Quevedo, and J. Jensen, "Real-time perceptual moving-horizon multiple-description audio coding," *IEEE Transactions on Signal Processing*, vol. 59, no. 9, pp. 4286–4299, 2011, ISSN: 1053-587X. DOI: 10.1109/TSP.2011.2159601.
- [31] W. Liu, K. R. Vijayanagar, and J. Kim, "Low-delay distributed multiple description coding for error-resilient video transmission," in *2011 IEEE 13th International Workshop on Multimedia Signal Processing*, 2011, pp. 1–6. DOI: 10.1109/MMSP.2011.6093823.

- [32] J. Østergaard and D. Quevedo, "Multiple description coding for closed loop systems over erasure channels," in *2013 Data Compression Conference*, Mar. 2013, pp. 311–320, ISBN: 978-1-4673-6037-1. DOI: 10.1109/DCC.2013.39.
- [33] U. Samarawickrama and J. Liang, "A two-stage algorithm for multiple description predictive coding," in *2008 Canadian Conference on Electrical and Computer Engineering*, 2008, pp. 000 685–000 688. DOI: 10.1109/CCECE.2008.4564622.
- [34] M. S. Derpich, E. I. Silva, and J. Østergaard, "Fundamental inequalities and identities involving mutual and directed informations in closed-loop systems," *IEEE Transactions on Information Theory*, 2013, Submitted for publication. [Online]. Available: <http://arxiv.org/abs/1301.6427>.
- [35] S. Ihara, *Information Theory for Continuous Systems*. WORLD SCIENTIFIC PUB CO INC, Sep. 1, 1993, 308 pp., ISBN: 9810209851.
- [36] R. Zamir, Y. Kochman, and U. Erez, "Achieving the gaussian rate–distortion function by prediction," *IEEE Transactions on Information Theory*, vol. 54, no. 7, pp. 3354–3364, 2008, ISSN: 0018-9448. DOI: 10.1109/TIT.2008.924683.
- [37] D. S. Taubman, *JPEG2000 image compression fundamentals : Standards and practice*, eng. Boston, Mass: Kluwer, 2002, ISBN: 079237519x.
- [38] R. Zamir and M. Feder, "Information rates of pre/post-filtered dithered quantizers," *IEEE Transactions on Information Theory*, vol. 42, no. 5, pp. 1340–1353, 1996, ISSN: 0018-9448. DOI: 10.1109/18.532876.
- [39] R. Zamir and M. Feder, "On universal quantization by randomized uniform/lattice quantizers," *IEEE Transactions on Information Theory*, vol. 38, no. 2, pp. 428–436, 1992, ISSN: 0018-9448. DOI: 10.1109/18.119699.
- [40] H. Gish and J. Pierce, "Asymptotically efficient quantizing," *IEEE Transactions on Information Theory*, vol. 14, no. 5, pp. 676–683, 1968, ISSN: 0018-9448. DOI: 10.1109/TIT.1968.1054193.
- [41] Kwang Taik Kim and T. Berger, "Sending a lossy version of the innovations process is suboptimal in qg rate-distortion," in *Proceedings. International Symposium on Information Theory, 2005. ISIT 2005.*, 2005, pp. 209–213. DOI: 10.1109/ISIT.2005.1523324.
- [42] O. A. Moussa, M. Li, and W. B. Kleijn, "Predictive audio coding using rate-distortion-optimal pre- and post-filtering," in *2011 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, 2011, pp. 213–216. DOI: 10.1109/ASPAA.2011.6082345.
- [43] D. S. S. Robert H. Shumway, *Time Series Analysis and Its Applications*. Springer-Verlag GmbH, Apr. 19, 2017, ISBN: 3319524518.

- [44] J. Chen, C. Tian, T. Berger, and S. S. Hemami, "Multiple Description Quantization Via Gram-Schmidt Orthogonalization," *IEEE Transactions on Information Theory*, vol. 52, no. 12, pp. 5197–5217, 2006, ISSN: 0018-9448. DOI: 10.1109/TIT.2006.885498.
- [45] S. P. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge Univ. Pr., 2009, ISBN: 978-0-521-83378-3. [Online]. Available: https://web.stanford.edu/~boyd/cvxbook/bv_cvxbook.pdf.
- [46] J. Østergaard, Y. Kochman, and R. Zamir, "An asymmetric difference multiple description gaussian noise channel," in *2017 Data Compression Conference (DCC)*, 2017, pp. 360–369. DOI: 10.1109/DCC.2017.16.
- [47] Hanying Feng and M. Effros, "On the rate loss of multiple description source codes," *IEEE Transactions on Information Theory*, vol. 51, no. 2, pp. 671–683, 2005, ISSN: 0018-9448. DOI: 10.1109/TIT.2004.840900.
- [48] S. M. Moser, "Advanced topics in information theory (lecture notes)," Version 3.6. Signal and Information Processing Laboratory, ETH Zürich, Switzerland., May 2019, [Online]. Available: <http://moser-isi.ethz.ch/scripts.html>.
- [49] A. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Transactions on Information Theory*, vol. 22, no. 1, pp. 1–10, 1976, ISSN: 0018-9448. DOI: 10.1109/TIT.1976.1055508.
- [50] Y. Frank-Dayana and R. Zamir, "Dithered lattice-based quantizers for multiple descriptions," *IEEE Transactions on Information Theory*, vol. 48, no. 1, pp. 192–204, 2002, ISSN: 0018-9448. DOI: 10.1109/18.971748.
- [51] H. Madsen and P. Thyregod, *Introduction to General and Generalized Linear Models*, eng, ser. Chapman & Hall/CRC Texts in Statistical Science. London: CRC Press, 2011, ISBN: 1420091557.
- [52] H. Wang and P. Viswanath, "Vector gaussian multiple description with individual and central receivers," *IEEE Transactions on Information Theory*, vol. 53, no. 6, pp. 2133–2153, 2007, ISSN: 0018-9448. DOI: 10.1109/TIT.2007.896880.
- [53] J. Østergaard and R. Zamir, "Multiple-description coding by dithered delta-sigma quantization," *IEEE Transactions on Information Theory*, vol. 55, no. 10, pp. 4661–4675, 2009, ISSN: 0018-9448. DOI: 10.1109/TIT.2009.2027528.
- [54] V. A. Vaishampayan, "Design of multiple description scalar quantizers," *IEEE Transactions on Information Theory*, vol. 39, no. 3, pp. 821–834, 1993, ISSN: 0018-9448. DOI: 10.1109/18.256491.

- [55] G. Sun, U. Samarawickrama, J. Liang, C. Tian, C. Tu, and T. D. Tran, "Multiple description coding with prediction compensation," *IEEE Transactions on Image Processing*, vol. 18, no. 5, pp. 1037–1047, 2009, ISSN: 1057-7149. DOI: 10.1109/TIP.2009.2013068.
- [56] Chao Tian and S. S. Hemami, "A new class of multiple description scalar quantizer and its application to image coding," *IEEE Signal Processing Letters*, vol. 12, no. 4, pp. 329–332, 2005, ISSN: 1070-9908. DOI: 10.1109/LSP.2005.843764.
- [57] S. Kay, *Intuitive Probability and Random Processes using MATLAB®*. Springer US, Dec. 16, 2005, 856 pp., ISBN: 0387241574.
- [58] P. Olofsson and M. Andersson, *Probability, statistics, and stochastic processes*, 2nd ed. Wiley, 2012, ISBN: 9781118231296.
- [59] S. N. Diggavi and V. A. Vaishampayan, "On multiple description source coding with decoder side information," in *Information Theory Workshop*, 2004, pp. 88–93. DOI: 10.1109/ITW.2004.1405280.
- [60] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016, <http://www.deeplearningbook.org>.
- [61] D. Guichard, *Calculus: Early transcendentals*, eng. Lyryx, 2016. [Online]. Available: <https://open.umn.edu/opentextbooks/BookDetail.aspx?bookId=415>.
- [62] J. L. Massey and P. C. Massey, "Conservation of mutual and directed information," in *Proceedings. International Symposium on Information Theory, 2005. ISIT 2005.*, 2005, pp. 157–158. DOI: 10.1109/ISIT.2005.1523313.
- [63] P.-O. Amblard and O. J. J. Michel, *Relating granger causality to directed information theory for networks of stochastic processes*, 2009. arXiv: 0911.2873 [cs.IT].
- [64] T. Wiegand and H. Schwarz, "Source coding: Part I of Fundamentals of Source and Video Coding," *Foundations and Trends in Signal Processing*, vol. 4, no. 1–2, pp. 1–222, 2011, ISSN: 1932-8346. DOI: 10.1561/20000000010.
- [65] B. S. Albrecht Böttcher, *Analysis of Toeplitz Operators*. Springer-Verlag GmbH, Apr. 5, 2006, ISBN: 978-3-540-32436-2.

A | Information Theory

This appendix introduces some essential information theoretic definitions and results that are used throughout the report.

A.1 Discrete Entropy

Initially we introduce the entropy of a random variable, as a measure of the uncertainty about the random variable.

Definition A.1 (Discrete Entropy [13, p. 14])

The *discrete entropy*, $H(X)$, of a discrete random variable, $X \in \mathcal{X}$, with probability mass function (PMF) $p(x)$, is defined as

$$H(X) \triangleq - \sum_{x \in \mathcal{X}} p(x) \log p(x). \quad (\text{A.1})$$

The base of the logarithm determines the unit of entropy, i.e. if the base is 2 the entropy is expressed in bits, if the base is e the entropy is measured in *nats* [13, p. 14]. Throughout this report we take the base of logarithms to be 2, unless otherwise specified. We use the convention $0 \log 0 = 0$, since by continuity $x \log x \rightarrow 0$ as $x \rightarrow 0$ [13, p. 14].

The discrete entropy is a functional of the distribution of X , i.e. it does not depend on the values taken by random variable [13, p. 14].

The discrete entropy can also be considered a measure of the amount of information on average required to describe a random variable (see Appendix B.2.1).

From (A.1) we see that the discrete entropy may be defined as [13, p. 14],

$$H(X) \triangleq - \mathbb{E}_X [\log p(X)]. \quad (\text{A.2})$$

An immediate result from the definition of discrete entropy and the fact that $0 \leq p(x) \leq 1$ for all $x \in \mathcal{X}$ is [13, Lemma 2.1.1],

$$H(X) \geq 0. \quad (\text{A.3})$$

That is, the discrete entropy is always non-negative. Especially, it is zero if X is deterministic.

The discrete entropy is maximized by the uniform distribution.

Lemma A.2 (Entropy maximization [13, Theorem 2.6.4])

Let $X \in \mathcal{X}$ be a discrete random variable, and let $|\mathcal{X}|$ denote the cardinality of \mathcal{X} . Then

$$H(X) \leq \log|\mathcal{X}|, \quad (\text{A.4})$$

with equality if, and only if, X has a uniform distribution over \mathcal{X} .

We may also define the joint- and conditional entropy of jointly distributed discrete random variables.

Definition A.3 (Joint- and conditional entropy [13, p. 16])

Let (X, Y) be a pair of discrete random variables with a joint distribution¹ $p(x, y)$. Then the *joint entropy*, $H(X, Y)$, is defined as

$$H(X, Y) \triangleq -\mathbb{E}_{X, Y} [\log p(X, Y)]. \quad (\text{A.5})$$

The *conditional entropy*, $H(X|Y)$ is defined as,

$$H(X|Y) \triangleq -\mathbb{E}_{X, Y} [\log p(X|Y)]. \quad (\text{A.6})$$

For a sequence of random variables X^n we have the following useful result.

Lemma A.4 (Chain rule for entropy [13, Theorem 2.5.1])

For a sequence of discrete random variables X^n with joint distribution $p(x^n)$, then

$$H(X^n) = \sum_{i=1}^n H(X_i|X^{i-1}). \quad (\text{A.7})$$

Particularly for two random variables, (X, Y) [13, p. 17],

$$H(X, Y) = H(X) + H(Y|X). \quad (\text{A.8})$$

An important property, that will become useful in the sequel is that conditioning reduces entropy.

¹We usually denote probability distributions by uppercase P . However, to emphasize this is a discrete distribution we use the lowercase notation usually reserved for a PMF.

Lemma A.5 (Conditioning reduces entropy [13, Theorem 2.6.5])

$$H(X|Y) \leq H(X), \quad (\text{A.9})$$

with equality if, and only if, X and Y are independent.

This shows, that the knowledge of one random variable cannot increase the uncertainty about another. Intuitively, knowing the temperature today cannot increase our uncertainty about the weather tomorrow.

This leads to the following upper bound on the joint entropy.

Lemma A.6 (Independence bound [13, Theorem 2.6.6])

Let X^n be a drawn according to $p(x^n)$. Then

$$H(X^n) \leq \sum_{i=1}^n H(X_i), \quad (\text{A.10})$$

with equality if, and only if, all X_i are mutually independent, or all deterministic.

Another useful measure to be used in the following is the *Kullback-Leibler divergence*.

Definition A.7 (Kullback-Leibler divergence [13, p. 19])

The *Kullback-Leibler divergence*, $D(p||q)$, between two probability mass functions $p(x)$ and $q(x)$ is defined as

$$D(p||q) \triangleq \sum_{x \in \mathcal{X}} p(x) \log \frac{p(x)}{q(x)} \quad (\text{A.11})$$

$$= E_p \left[\log \frac{p(X)}{q(X)} \right]. \quad (\text{A.12})$$

Where we let $0 \log \frac{0}{0} = 0$, $0 \log \frac{0}{q} = 0$ and $p \log \frac{p}{0} = \infty$ [13, p. 19].

Kullback-Leibler divergence is also known as *relative entropy* or *information divergence*. We use the terms information divergence and divergence interchangeably. The divergence is a measure of the inefficiency in assuming the distribution of X is q when the true distribution is p . That is, if we used the distribution q to construct a source code for X , we would need on average $H(p) + D(p||q)$ bits to

describe X , compared to the $H(p)$ bits needed for a code constructed using true distribution [13, p. 19]. The quantity $H(p) + D(p||q)$ is also known as the *cross-entropy* [60]. The cross-entropy is often used as a loss function in Machine Learning algorithms [60], since we may also consider divergence as the “distance” between two distributions. Hence, the purpose of the Machine Learning algorithm is to make the estimated distribution q as close as possible to p . However, divergence is not a true distance, since it does not satisfy the triangle inequality and is not symmetric [13, p. 19]

An important property of divergence is its non-negativity.

Lemma A.8 (Information inequality [13, Theorem 2.6.3])

Let $p(x)$, $q(x)$ be two probability mass functions for the discrete random variable $X \in \mathcal{X}$. Then

$$D(p||q) \geq 0, \tag{A.13}$$

with equality if, and only if, $p(x) = q(x) \forall x \in \mathcal{X}$.

Using information divergence we now define the most important information measure for our purposes; *mutual information*.

Definition A.9 (Mutual Information [13, p. 19-20])

For a pair of random variables, (X, Y) with joint distribution $p(x, y)$ and marginal distributions $p(x)$, $p(y)$, the *mutual information* is defined as

$$I(X; Y) \triangleq D(p(x, y)||p(x)p(y)) \tag{A.14}$$

$$= E_{p(x, y)} \left[\log \frac{p(X, Y)}{p(X)p(Y)} \right]. \tag{A.15}$$

The mutual information is a measure of the amount of information one random variable carries about another [13, p. 19]. Especially, it follows from (A.16) in the following Lemma A.10, that the mutual information is considered the reduction in uncertainty of one variable due to knowledge of the other [13, p. 19]. The Venn diagram in Figure A.1 illustrates the relationship between entropy and mutual information. Particularly, the mutual information is the intersection between the information in X and Y [13, p. 22]. A few useful ways to express the mutual information in terms of entropy is given in the following lemma.

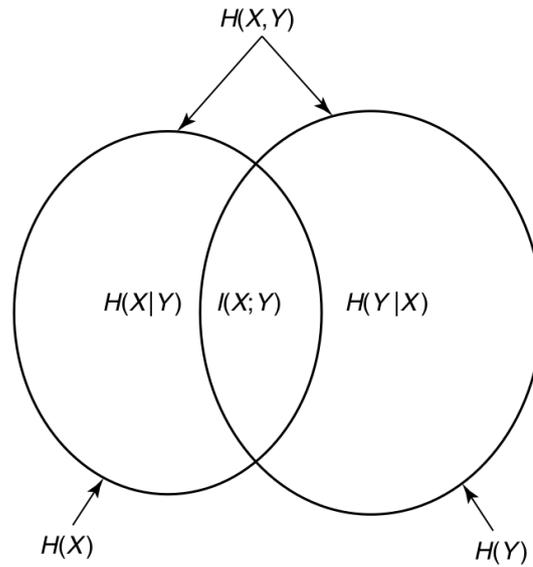


Figure A.1: Relationship between mutual information and entropy. Figure from [13, p. 22].

Lemma A.10 (Mutual Information and Entropy [13, Theorem 2.4.1])

$$I(X; Y) = H(X) - H(X|Y) \quad (\text{A.16})$$

$$I(X; Y) = H(X) + H(Y) - H(X, Y) \quad (\text{A.17})$$

$$I(X; Y) = I(Y; X) \quad (\text{A.18})$$

$$I(X; X) = H(X) \quad \text{Only for discrete } X \quad (\text{A.19})$$

By the non-negativity of the information divergence the mutual information is also non-negative.

Lemma A.11 (Non-negativity of mutual information [13, sec. 2.6])

For any two random variables, X, Y ,

$$I(X; Y) \geq 0, \quad (\text{A.20})$$

with equality if and only if $X \perp Y$.

For a proof see the corollary to [13, Theorem 2.6.3].

Similar to entropy we may define *conditional mutual information*.

Definition A.12 (Conditional mutual information [13, p. 23])

The *conditional mutual information* of random variables X and Y given Z is defined by

$$I(X; Y|Z) = H(X|Z) - H(X|Y, Z). \quad (\text{A.21})$$

For a sequence of random variables, X^n , we have the following useful chain rule of mutual information.

Lemma A.13 (Chain rule for mutual information [13, Theorem 2.5.2])

For a sequence of random variables, X^n and the random variable, Y ,

$$I(X^n; Y) = \sum_{i=1}^n I(X_i; Y|X^{i-1}). \quad (\text{A.22})$$

Particularly, for three random variables we have that

$$I(X; Y, Z) = I(X; Y) + I(X; Z|Y). \quad (\text{A.23})$$

Which leads to the following useful identity

$$I(X; Y, Z) = I(X; Y) + I(X; Z) + I(Y; Z|X) - I(Y; Z). \quad (\text{A.24})$$

An important information theoretic result is the data processing inequality.

Lemma A.14 (Data processing inequalities [13, sec. 2.8])

If the Markov chain $X - Y - Z$ holds, then

$$I(X; Y) \geq I(X; Z), \quad (\text{A.25})$$

with equality if and only if $X - Z - Y$.

If the Markov chain $X|_W - Y|_W - Z|_W$ holds, then

$$I(X; Y|W) \geq I(X; Z|W), \quad (\text{A.26})$$

with equality if and only if $X|_W - Z|_W - Y|_W$.

The proof of (A.25) is found in [13, sec. 2.8].

The proof of (A.26) follows by a straightforward extension of the proof in [13, sec.

2.8] to the conditional case.

This shows, that no processing of Y , either deterministic or random, can increase the information Y contains about X [13, sec. 2.8]. Thus, we may only lose information due to processing.

Finally we define the *mutual information rate* between random process, as the mutual information per symbol for asymptotically long sequences.

Definition A.15 (Mutual information rate [12, Eq. 7.3.9])

The *mutual information rate* between two random processes $\{X_k\}$ and $\{Y_k\}$ is defined as

$$\bar{I}(X; Y) \triangleq \lim_{n \rightarrow \infty} \frac{1}{n} I(X^n; Y^n). \quad (\text{A.27})$$

A.2 Differential Entropy

We now introduce *differential entropy* as the entropy of continuous random variables. Many of the definitions and results for discrete entropy easily extend to differential entropy. However, some care must be taken [13, p. 243].

Many of the following statements involve integrals, which we may not always exist. Hence, all results should be extended by the statement *if it exists* [13, p. 243].

Definition A.16 (Differential Entropy [13, p. 243])

The *differential entropy*, $h(X)$, of a continuous random variable $X \in \mathcal{X}$ with probability density function (PDF) $f(x)$ is defined as

$$h(X) \triangleq - \int_{\mathcal{S}} f(x) \log f(x) dx, \quad (\text{A.28})$$

where \mathcal{S} is the support of X , i.e. those $x \in \mathcal{X}$ for which $f(x) > 0$.

Similar to the discrete case, the differential entropy depends only on the PDF $f(x)$ [13, p. 243].

Contrary to discrete entropy, the differential entropy *can* be negative. To see this, consider the following example.

Example A.17 (Uniform distribution)

Let X be a uniform random variable on the interval $[0, a]$, i.e. $X \sim \mathcal{U}[0, a]$, then

$$h(X) = - \int_0^a \frac{1}{a} \log \frac{1}{a} = \log a. \quad (\text{A.29})$$

Now if $a < 1$, it follows from (A.29) that the differential entropy is negative.

More importantly we note that the differential entropy may be infinite or negatively infinite.

The definition of joint- and conditional differential entropy follow analogously to the discrete case [13, sec. 8.4]. The chain rule also extends to the continuous case. However, care must be taken if any entropies are infinite.

We consider ZDMD source coding of stationary Gaussian processes. Thus, we often need the entropy of Gaussian random variables.

Lemma A.18 (Entropy of multivariate Gaussian [13, Theorem 8.4.1])

Let the sequence of random variables X^n have a multivariate Gaussian distribution with mean μ and covariance matrix Σ , i.e. $X^n \sim \mathcal{N}(\mu, \Sigma)$. Then

$$h(X^n) = \frac{1}{2} \log ((2\pi e)^n |\Sigma|), \quad (\text{A.30})$$

where $|\Sigma|$ denotes the determinant of Σ .

In fact the Gaussian distribution maximizes the differential entropy across all distributions with the same covariance.

Lemma A.19 ([13, Theorem 8.6.5])

Let $X \in \mathbb{R}^n$ be a random vector with zero mean and covariance matrix $\Sigma = E[XX^T]$. Then

$$h(X) \leq \frac{1}{2} \log ((2\pi e)^n |\Sigma|), \quad (\text{A.31})$$

with equality if, and only if, $X \sim \mathcal{N}(0, \Sigma)$.

Thus for continuous random variables under a covariance constraint the Gaussian distribution requires the longest average description length.

We highlight the definition of Kullback-Leibler divergence also in the continuous case.

Definition A.20 (Kullback-Leibler divergence [13, p. 251])

The Kullback-Leibler divergence between two densities f and g is defined by,

$$D(f\|g) = \int f \log \frac{f}{g}. \quad (\text{A.32})$$

Where by continuity we let $0 \log \frac{0}{0} = 0$ [13, p. 251]. It is important to notice that the divergence is only finite if the support of f is contained in the support of g [13, p. 251]. Otherwise we have $f(x) \log \frac{f(x)}{0}$ for some x , which is infinite.

Mutual information for continuous random variables then follows analogously to the discrete case. The properties of divergence and mutual information are the same as in the discrete case [13, p. 251]. Particularly the divergence is zero if $f = g$ almost everywhere.

An interesting result expressing the ‘‘Gaussianity’’ of a random variables may be expressed in terms of the information divergence.

Lemma A.21 ([13, p.254-255])

Let X be a continuous random variable with arbitrary distribution, and let X_G be a Gaussian random variable with second order moments equal to those of X .

Then

$$D(X\|X_G) = h(X_G) - h(X). \quad (\text{A.33})$$

The lemma follows from the proof of [13, Theo. 8.6.5].

For the random variable X , this expresses the information loss experienced by assuming X is Gaussian with second moments equal to those of X .

Another interesting result of Gaussian random variables is related to the mutual information.

Lemma A.22 (Gaussian Mutual Information Minimization [35, Theo. 1.8.6])

Let X_G^n be a sequence of jointly Gaussian random variables. Furthermore, let Y^m and Y_G^m be sequences of random variables such that the second moments of (X_G^n, Y^m) and (X_G^n, Y_G^m) are equal, and (X_G^n, Y_G^m) is jointly Gaussian. Then

$$I(X_G^n; Y^m) \geq I(X_G^n; Y_G^m). \quad (\text{A.34})$$

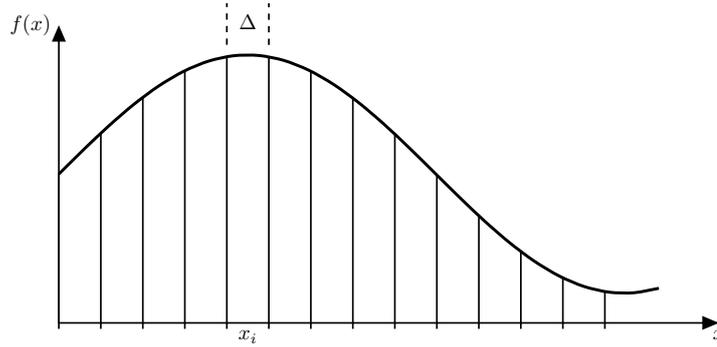


Figure A.2: Quantization of a continuous random variable. Figure modified from [13].

This shows, for a Gaussian random variable, X_G , and a covariance constraint, the mutual information $I(X_G; Y)$ is minimized when Y is jointly Gaussian with X . That is, similarly to Gaussian random variables having the highest uncertainty individually, under a covariance constraint. Gaussian random variables also have the most uncertainty about each other, under a covariance constraint.

A.2.1 Relation to Discrete Entropy

We highlight an important result relating discrete- and differential entropy through quantization. The following derivation is found in [13, sec. 8.3].

Let X be a continuous random variable with density $f(x)$. We then quantize X by dividing the range of X into intervals of length Δ as illustrated in Figure A.2. Assuming $f(x)$ is continuous in each bin, we have by the mean value theorem [61], there exists a point x_i in each bin such that

$$f(x_i)\Delta = \int_{i\Delta}^{(i+1)\Delta} f(x) dx. \quad (\text{A.35})$$

Defining the quantized version of X as

$$X^\Delta = x_i \quad \text{if } i\Delta \leq X \leq (i+1)\Delta, \quad (\text{A.36})$$

we have the following result

Lemma A.23 ([13, Theorem 8.3.1])

If the density $f(x)$ of the random variable X is Riemann integrable, then

$$H(X^\Delta) + \log \Delta \rightarrow h(X), \quad \text{as } \Delta \rightarrow 0. \quad (\text{A.37})$$

This shows, the entropy of an n -bit quantization of a continuous random variable is approximately $h(X) + n$. Particularly, $h(x) + n$ is the average number of bits required to describe X to n -bit accuracy [13, p. 249].

This result proves useful when determining the quantization bin sizes in the practical implementation of a source code (Chapter 5).

A.3 Directed Information

We now consider some important results of directed information. An introduction to directed information is given in Section 1.2.

Definition A.24 (Directed information [21])

The *directed information* from a sequence of random variables X^n to a sequence Y^n , is defined as

$$I(X^n \rightarrow Y^n) \triangleq \sum_{i=1}^n I(X^i; Y_i | Y^{i-1}). \quad (\text{A.38})$$

Initially, by the non-negativity of mutual information (A.20), the directed information is also non-negative [21]

$$0 \leq I(X^n \rightarrow Y^n). \quad (\text{A.39})$$

More importantly, the directed information is always less than or equal to the mutual information,

$$I(X^n \rightarrow Y^n) \leq I(X^n; Y^n), \quad (\text{A.40})$$

with equality if, and only if, there is no feedback from Y^n to X^n . Furthermore, a conservation law for mutual information was derived in [62],

$$I(X^n; Y^n) = I(X^n \rightarrow Y^n) + I(0 * Y^{n-1} \rightarrow X^n), \quad (\text{A.41})$$

where $0 * Y^{n-1}$ denotes the concatenation of 0 and Y^{n-1} , that is $0 * Y^{n-1} = (0, Y_1, \dots, Y_{n-1})$.

This ensures synchronization of sequences on a common clock. Also $0 * Y^{n-1}$ and Y^{n-1} are equivalent sequences in mutual information expressions [62].

We see from (A.41) that the mutual information may be split into a feedforward information flow, $I(X^n \rightarrow Y^n)$, and a feedback flow $I(0 * Y^{n-1} \rightarrow X^n)$. Thus, emphasizing the equality between mutual- and directed information if, and only if, there is no feedback from Y to X , i.e. when the second term of (A.41) is zero. Particularly, this holds if [62]

$$H(X_k | X^{k-1}, Y^{k-1}) = H(X_k | X^{k-1}), \quad 2 \leq k \leq n. \quad (\text{A.42})$$

That is, when X is independent of the past of Y given its own past [63], i.e when the Markov chain,

$$Y^{k-1} - X^{k-1} - X_k, \quad (\text{A.43})$$

holds for all k [63].

B | Source coding

This appendix introduces the primary concepts of source coding. Initially we consider how sources may be modeled using stochastic processes in Section B.1. We then focus in more detail on source coding in Section B.2, especially lossless source coding. We also mention how quantization in particular is considered lossy source coding.

This appendix assumes the reader is familiar with the basic concepts of entropy, as introduced in Appendix A.

B.1 Sources

When transmitting an analog or continuous valued signal across a wireless network to a receiver, the source signal must be compressed using a source coding scheme. However, as designers of the compression scheme, we do not know the source signal in advance. Therefore, the source signal is often modeled as a random process [64]. If reasonable assumptions are made with respect to the source of information, the performance of source coding schemes can then be characterized based on probabilistic averages [64]. The exact source signal, given as a realization of the stochastic process, is only available when the actual encoding happens.

As an example, when encoding a speech signal, the exact words and sounds are unknown prior to being spoken. Thus, the encoder must be designed based an appropriate model. When the speech signal is spoken, we are able to sample this signal and encode using the previously designed scheme.

Let $\{X_k\}$, $X_k \in \mathcal{X}$, be the stochastic process modeling the unknown signal values for specific source. We refer to this stochastic process as a source process. We also often refer to the source process $\{X_k\}$ as the source.

For the source process $\{X_k\}$ we denote a realization of the source process by $\{x_k\}$, $x_k \in \mathcal{X}$. When performing encoding we do so on a realization of the source process.

B.1.1 Stochastic Processes

We highlight a few aspects of stochastic processes, that are important to the overall context of source coding. For further studies of stochastic processes the interested reader is referred to e.g. [43], [57], [58], [64].

A stochastic process may be considered either one- or two-sided.

Definition B.1 (One- and two-sided stochastic process)

A discrete-time stochastic process $\{X_k\}$ is said to be *one-sided* if $k \in \mathbb{N}$.

The stochastic process, $\{X_k\}$ is said to be *two-sided* if $k \in \mathbb{Z}$.

We consider only stationary processes in this report.

Definition B.2 (Stationary stochastic process [43])

A discrete-time stochastic process $\{X_k\}_{k \in \mathbb{Z}}$ $X_k \in \mathcal{X}$ is said to be (*strict sense*) *stationary* if the statistical properties of every collection of random variables

$$(X_{k_1}, \dots, X_{k_N}), \quad (\text{B.1})$$

is identical to that of the time shifted set

$$(X_{k_1+l}, \dots, X_{k_N+l}). \quad (\text{B.2})$$

That is, if

$$P(X_{k_1} \leq x_1, \dots, X_{k_N} \leq x_N) = P(X_{k_1+l} \leq x_1, \dots, X_{k_N+l} \leq x_N), \quad (\text{B.3})$$

for all $N \in \mathbb{N}$, all time points $k_1, \dots, k_N \in \mathbb{Z}$, all numbers $x^N \in \mathcal{X}^N$ and all time shifts $l \in \mathbb{Z}$. Where P is the N 'th-order distribution function of the stochastic process.

Although this definition considers two-sided processes, the same definition applies to one-sided processes.

There is an important relation between stationarity and one- or two-sided processes. Consider the AR(1)-process,

$$X_k = aX_{k-1} + W_k, \quad (\text{B.4})$$

where $|a| < 1$, and $\{W_k\}$ is an independent and identically distributed (IID) process. If this is a one-sided process, i.e. $k \in \mathbb{N}$, the initial state is X_1 with variance $\text{Var}[X_1] = \sigma_{X_1}^2$. Then depending on the value of $\sigma_{X_1}^2$ the variance of X_2 is different

from that of X_1 and the process is not stationary. However, we may consider this process to be stationary in the limit of $k \rightarrow \infty$, as the distribution of the source “settles”.

If the process was two-sided, i.e. $\{X_k\}_{k \in \mathbb{Z}}$, then by similar arguments the process will be stationary for $k > 0$, since it has already been “running” since the infinite past. Therefore, we may consider a stationary source process as a two-side process $\{X_k\}_{k \in \mathbb{Z}}$, where we only encode the stationary part from time $k > 0$, $\{X_k\}_{k \in \mathbb{N}}$. Thus, we consider the infinite past $\{X_k\}_{k=-\infty}^0$ to be available for determining the statistics of the source.

Particularly, the variance of a stationary AR(1) process is

$$\sigma_X^2 = \frac{\sigma_W^2}{1 - a^2}, \quad (\text{B.5})$$

where σ_W^2 is the variance of the IID process $\{W_k\}$. Thus, in the one-sided case if $\sigma_{X_1}^2$ is equal to (B.5), the process is stationary.

In this report we consider the special class of Gaussian processes.

Definition B.3 (Gaussian process [64])

A continuous-valued discrete-time stochastic process $\{X_k\}_{k \in \mathbb{N}}$ is said to be a *Gaussian process* if all sequences, X_m^n , $n \geq m$, $n, m \in \mathbb{N}$, have a jointly Gaussian distribution.

If the source process $\{X_k\}$ is Gaussian, we often refer to the source as a *Gaussian source*. Particularly, we are interested in *Gauss-Markov* (GM) process, i.e. processes that are both Gaussian and Markov [64].

The statistical properties of a stationary Gauss-Markov process, $\{X_k\}$ is completely characterized by its mean, μ_X , its variance, σ_X^2 , and its correlation coefficient a [64]. We refer to these as the statistics of the source. As mentioned in the project delimitations, we consider all these statistics to be determined beforehand by the infinite past of the source. That is, the main source process considered in this report is the stable stationary scalar Gauss-Markov source process,

$$X_k = aX_{k-1} + W_k, \quad k \in \mathbb{N}, \quad (\text{B.6})$$

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 with $\sigma_{X_1}^2 = \frac{\sigma_W^2}{1 - a^2}$, and $W_k \in \mathbb{R} \sim \mathcal{N}(0, \sigma_W^2)$, is an IID Gaussian process independent of X_1 .

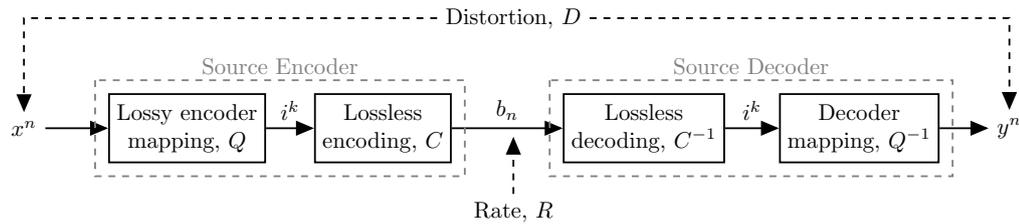


Figure B.1: A block diagram of a typical source coding system. The encoder encodes the source, x , to the binary descriptions, b , the source decoder then produces an estimate, y , of x given the description b . The rate is measured between the encoder output and decoder input, i.e. the number of bits transmitting between encoder and decoder. The distortion is measured between the encoder input and decoder output, i.e. the discrepancy between the source, x , and its representation, y . Figure modified from [64, Fig. 4.1].

B.2 Source Coding

In the following we do not consider any particular source, and give a general introduction to source coding.

The block diagram in Figure B.1 illustrates a lossy source coding system with an encoder and decoder. The encoder maps a given set of source symbols x^n to the binary codeword b_n of length L_n . The decoder then maps the binary codeword into a sequence of source reproductions y^n . The encoder is split into a lossy mapping, Q , that maps the source sequence to a sequence of indexes, i^k , from a countable alphabet, and a lossless mapping that converts the indexes to a codeword b_n [64, sec. 4.1]. The irreversible and thus lossy mapping, Q , is any mapping that produces a set of indexes of a countable alphabet [64, sec. 4.1]. The decoder is also split into a lossless decoding, C^{-1} , which inverts the codewords b_n back to the indexes i^k , and the mapping Q^{-1} produces the estimates y^n of x^n from the indexes i^k [64, sec 4.1].

To further understand these concepts, we initially consider lossy- and lossless coding independently from each other.

B.2.1 Lossless Source Coding

Lossless source coding or data compression is the compression of data without loss. That is, a source is encoded such that a perfect reconstruction is possible at the decoder, i.e. there is no distortion between the source and reconstruction. This is achieved by removing redundant information from the source.

In lossless source coding the source is already available in a discrete-valued (digital) form [10, p. 4], and is compressed to a shorter binary representation. Hence,

lossless coding is relative to the fact that some loss might have already occurred.

Initially we define a lossless code.

Definition B.4 (Binary Lossless Source code [13, p. 103][64, sec. 3.2])

For a stationary discrete-valued discrete-time stochastic process $\{X_k\}$, where $X_k \in \mathcal{X}$, a binary lossless *source code*, $C : \mathcal{X} \rightarrow 2^*$, is mapping from \mathcal{X} to the set of finite-length binary strings, 2^* .

$C(x)$ denotes the *codeword* corresponding to the symbol $x \in \mathcal{X}$ and $l(x)$ denotes the length of $C(x)$ (in bits).

A lossless code need not be binary, it may be any D -ary code, i.e. a mapping to the set of finite-length strings from a D -ary alphabet [13, p. 103].

The codeword for a sequence of realizations $x^n \in \mathcal{X}^n$ is defined as,

$$C(x^n) \triangleq C(x_1)C(x_2) \cdots C(x_n), \quad (\text{B.7})$$

i.e. the concatenation of the codewords corresponding to each x_i .

We consider only the important class of *uniquely decodable codes*.

Definition B.5 (Uniquely decodable code [13, p. 105])

A code C is *uniquely decodable* if

$$x^n \neq \tilde{x}^n \Rightarrow C(x^n) \neq C(\tilde{x}^n). \quad (\text{B.8})$$

That is, any encoded string can come from only one possible source string.

An important measure for the effectiveness of a source code is the *expected length* (in bits).

Definition B.6 (Expected length [13, p. 104])

The *expected length*, $\mathcal{L}(C)$ of a source code $C(x)$ for the random variable X with PMF, $p(x)$, is

$$\mathcal{L}(C) \triangleq \sum_{x \in \mathcal{X}} p(x)l(x) \quad (\text{B.9})$$

$$= \text{E} [l(X)]. \quad (\text{B.10})$$

The goal of lossless source coding is to design a source code that minimize the expected codeword length, \mathcal{L} , while ensuring unique decodability of each message

x^n , given their codeword $C(x^n)$ [64, sec. 3.2].

In *fixed-length* coding all elements of the source alphabet \mathcal{X} are mapped into binary codewords of equal length, l . There are 2^l such codewords. Therefore, the minimum length of l is $\lceil \log_2 |\mathcal{X}| \rceil$, where $|\mathcal{X}|$ is the cardinality of \mathcal{X} , and $\lceil \cdot \rceil$ denotes rounding up to the nearest integer. [10, p. 84]

Fixed-length coding is also referred to as *fixed-rate* or *resolution constrained* coding. The rate is fixed, since for a time-varying source all source symbols are encoded using the same amount of bits for all time steps. The resolution is constrained by the finite number of source values representable by l bits.

An alternative to fixed-rate coding is *variable-rate* or *variable-length* coding. Since the source probability distribution is rarely uniform, variable rate-coding takes advantage of this by assigning shorter codewords to highly probable symbols, and longer codewords to rare symbols [10, p. 84]. Thereby, the average codeword length is reduced compared to fixed-rate codes. However, this also results in high peak rates, which may be undesirable in certain scenarios.

Some variable rate uniquely decodable codes include Huffman codes [13, sec. 5.6] and Shannon codes [13, sec. 5.9]. For each $x \in \mathcal{X}$ Shannon codes assigns the codeword lengths

$$l_s(x) = \left\lceil \log \frac{1}{p(x)} \right\rceil. \quad (\text{B.11})$$

Here the rounding up adds at most 1 bit to the length [10, p. 85]. Ignoring this rounding up, the expected length is equal to the source entropy. Thus, the following important bounds can be given on the expected length of an optimal code.

Lemma B.7 (Bounds on optimal code length [13, Theo. 5.4.1])

Let l_1^*, \dots, l_m^* be the optimal codewords lengths of a binary source code for a discrete source $X \in \mathcal{X}$ where $m = |\mathcal{X}|$, and let \mathcal{L}^* be the associated expected length of an optimal code. Then

$$H(X) \leq \mathcal{L}^* < H(X) + 1. \quad (\text{B.12})$$

The bounds show, that the optimal expected length of a binary source code is within 1 bit of the entropy of the source. In particular no lossless coding scheme can achieve an expected length below entropy. Therefore, a variable-rate code is also said to be *entropy constrained*, or created using *entropy coding*.

This also shows, why the entropy may be considered a measure of the amount of information on average required to describe a random variable.

It is possible to achieve an expected length per symbol arbitrarily close to the

entropy by coding of large block lengths [13, p. 114]. Let $X^n \in \mathcal{X}^n$ be a sequence of n random variables drawn IID according to $p(x)$. Let L_n be the length of the codeword, $C(x^n)$, associated with the realization x^n , i.e.

$$L_n = \sum_{k=1}^n l(x_k). \quad (\text{B.13})$$

Then define

$$\mathcal{L}_n \triangleq \text{E}[L_n], \quad (\text{B.14})$$

as the expected length of the codeword for a sequence of n source symbol. Then, since X_1, \dots, X_n are IID, we have that $H(X^n) = nH(X)$, thus

$$H(X) \leq \frac{\mathcal{L}_n}{n} < H(X) + \frac{1}{n}, \quad (\text{B.15})$$

and the overhead per source symbol vanishes as the block length n tends to infinity [13, p. 114].

This technique of coding long sequences is the main idea behind standard rate-distortion theory proofs. By relying on the encoding of arbitrarily long sequences it is possible to achieve operational performance close to the information theoretic lower bounds. However, this implies long delays in the practical encoding. Thus, we cannot rely on these techniques in zero-delay source coding.

The Huffman code can be shown to be optimal in the sense that it achieves a minimum expected length for asymptotically long sequences compared to any other uniquely decodable code [13, sec. 5.8]. However, it is not better than any other code on any particular source sequence [13, p. 130]. Furthermore, it can be shown that Shannon codes are competitively optimal, i.e. no other code can do better than a Shannon code most of the time [13, p. 132].

B.2.2 Lossy Source Coding

Lossy source coding is compression with loss. That is, a source signal is encoded such that a perfect reconstruction is no longer possible at the decoder. For example, if the source process is continuous-valued it cannot be completely represented by indexes of a countable alphabet, hence the encoder mapping Q cannot be reversible [64, sec. 4.1].

Here a tolerable discrepancy between the source and the decoder estimate is achieved by removing irrelevant information. The lossy mapping could be e.g. scalar quantization, vector quantization or predictive coding.

In quantization a continuous valued source is compressed into a discrete set of values. For the random source sequence X^n the discrete set of indexes I^k , are discrete random variables. However, this discrete set may not be equiprobable.

Hence, it can be additionally compressed using lossless entropy coding [10, p. 85]. That is, we may encode the quantizer output, $Q(X)$, such that it approaches the entropy, $H(Q(X))$ [10, p. 85]. Thus, source coding is, as illustrated in Figure B.1, the combination of quantization and entropy coding.

C | Rate-Distortion Theory

This appendix introduces classical rate-distortion theory as the determination of the fundamental limits between the minimum required bitrate to represent a source subject to a given fidelity criterion, with no restrictions in terms of delay.

Initially a rate-distortion code is introduced. We then define the operational rate-distortion function as the fundamental optimal performance of a rate-distortion code. This is extended to the information rate-distortion function, which provides a more tractable way of the determining the operational performance limits. Finally we state the rate-distortion function for Gaussian sources with- and without memory.

This appendix assumes the reader is familiar with the concepts of information theory (Appendix A) and source coding (Appendix B).

C.1 Rate-Distortion Coding

A rate-distortion code combines lossy- and lossless source coding into one mapping, that maps source symbols directly into binary codewords. Thus, when defining a source code we ignore the usual split between lossy- and lossless coding.

Definition C.1 (Rate-distortion code [13], [64])

For an n -block, X^n , from a discrete-time stationary source $X_k \in \mathcal{X}$, a *rate-distortion code* consists of an encoder and decoder.

For each $n \in \mathbb{N}$ let \mathcal{B}_n be a predefined set of at most a countable number of codewords. The *encoder* is specified by the encoding function,

$$f_n : \mathcal{X}^n \rightarrow \mathcal{B}_n. \quad (\text{C.1})$$

The encoder outputs a message $B_n = f_n(X^n)$ with length L_n (in bits). The *decoder* is specified by the decoding function,

$$g_n : \mathcal{B}_n \rightarrow \mathcal{Y}^n, \quad (\text{C.2})$$

where \mathcal{Y} is the reproduction alphabet, the decoder produces the reproduction $Y^n = g_n(f_n(X^n))$.

C.1.1 Distortion

In lossy source coding the reproduction symbols, y^n , are not necessarily the same as the source symbols, x^n . Therefore, we need a measure for how well the reproductions approximate the source. Such a measure should be smaller for better approximations, and zero for perfect reconstructions.

Definition C.2 (Distortion [13, p. 305])

Let $x^n \in \mathcal{X}^n$ be a sequence of source symbols and $y^n \in \mathcal{Y}^n$ be a sequence of reproduction symbols, then the *fidelity criterion* between the sequences is defined as

$$d_n(x^n, y^n) = \frac{1}{n} \sum_{i=1}^n d(x_i, y_i), \quad (\text{C.3})$$

where the *single letter distortion measure* $d : \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R}_+$ is a mapping from the set of source alphabet-reproduction alphabet pairs to the nonnegative real numbers [13, p. 304].

The (*expected*) *distortion* between the sequences is defined as

$$D = E_{X,Y} [d_n(X^n, Y^n)]. \quad (\text{C.4})$$

The single letter distortion, $d(x, y)$, is a measure of the cost of representing the source symbol x by the reproduction symbol y [13, p. 304]. Different distortion measures exist, however we consider only the squared-error distortion,

$$d(x, y) = \|x - y\|_2^2, \quad (\text{C.5})$$

in this report. Squared error is the most popular distortion measure for continuous source alphabets [13, p. 305]. However, it is not an appropriate measure when considering speech and image sources. The task of finding a distortion measure that is analytically manageable and also physically meaningful is a daunting one, e.g. determining numerically the intelligibility of speech [12, p. 7]. Since we consider only Gaussian source signals, the squared distortion is sufficient for our purposes. The (*expected*) distortion (C.4) is the average expected distortion of the elements in a finite sequence [13, p. 305]. However, to better evaluate the performance of a source code we are interested in the average expected distortion for very long sequences, i.e.

$$\lim_{n \rightarrow \infty} E_{X,Y} [d_n(X^n, Y^n)]. \quad (\text{C.6})$$

C.1.2 Rate

As mentioned, the distortion is not the only measure for the effectiveness of a source code. Another important measure is the bitrate.

For an n -block rate-distortion code we define the *average data-rate*, $r_n(x^n, y^n)$, as the average number of bits per input symbol, i.e. [64]

$$r_n(x^n, y^n) \triangleq \frac{1}{n} L_n. \quad (\text{C.7})$$

Similar to the distortion, we are not interested in the data-rate of any particular finite sequence. Instead we consider the asymptotic expected rate.

Definition C.3 (Rate)

For a stationary discrete-time source $\{X_k\}$, the *average expected data-rate* associated with the n -block rate-distortion code, measured in bits per source symbol, is

$$R \triangleq \lim_{n \rightarrow \infty} \frac{1}{n} \mathbb{E}[L_n]. \quad (\text{C.8})$$

That is, the rate is the expected length per source symbol for arbitrarily long sequences. We use the terms rate and average expected data-rate interchangeably throughout the paper.

As illustrated in Figure B.1, the rate is measured between the source encoder and -decoder, and the distortion is measured between the source signal, $\{X_k\}$, and the reproduction, $\{Y_k\}$.

C.2 Rate-Distortion Function

We are now ready to consider the optimal performance of a source code. To this end we first formally define achievable rates and distortions.

Definition C.4 (Achievable rate-distortion pair [13], [64])

For a given source process $\{X_k\}$. A rate-distortion pair (R, D) is said to be *achievable* if there exists a sequence of rate-distortion codes $\{(f_n, g_n)\}_{n \in \mathbb{N}}$ such that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \mathbb{E}[L_n] \leq R, \quad (\text{C.9})$$

$$\lim_{n \rightarrow \infty} \frac{1}{n} \mathbb{E}_X[d_n(X^n, g_n(f_n(X^n)))] \leq D. \quad (\text{C.10})$$

Remark C.5

A rate-distortion code with rate R is said to be achievable with respect to the distortion D if the asymptotic average expected distortion satisfies (C.10).

Similarly, a rate-distortion code with distortion D is said to be achievable with respect to the rate R if the asymptotic average expected data-rate satisfies (C.9).

For a given source process $\{X_k\}$ a set of achievable rate-distortion pairs constitutes a *rate-distortion region* [13, p. 306].

The main problem of rate-distortion theory is to determine the fundamental bound between the set of achievable and non-achievable rate-distortion points for a given source and distortion measure. That is, determine the minimum rate, R , required to describe the source, X , such that (C.10) is satisfied [64].

The fundamental bound on the rate is called the *operational rate-distortion function*.

Definition C.6 (Operational RDF [13], [64])

For a stationary source process $\{X_k\}$, the operational rate-distortion function (RDF), $R^{\text{op}}(D)$, is defined as the minimum achievable rate (C.8) with respect to the asymptotic distortion constraint $D > 0$, where the infimum is over all possible encoder- and -decoder sequences, $\{f_n\}_{n \in \mathbb{N}}, \{g_n\}_{n \in \mathbb{N}}$, such that (C.10) is satisfied. That is,

$$R^{\text{op}}(D) \triangleq \inf_{\{f_n\}, \{g_n\}} \lim_{n \rightarrow \infty} \frac{1}{n} \mathbb{E}[L_n], \quad (\text{C.11})$$

$$\text{s.t.} \quad \lim_{n \rightarrow \infty} \frac{1}{n} \mathbb{E}_X [d_n(X^n, g_n(f_n(X^n)))] \leq D.$$

The relationship between the achievable rate-distortion pairs and the operational RDF is illustrated in Figure C.1 for a white scalar Gaussian source.

The inverse of the operational RDF is called the *operational distortion-rate function*, it is the minimum achievable distortion, D , given that the source is encoded such that (C.9) is satisfied [64].

C.3 Information Rate-Distortion Function

For a given source process $\{X_k\}$ determining the operational RDF as defined in (C.11) is infeasible, since it is a minimization over all possible operational codes. As an alternative, the *information rate-distortion function*, introduced originally by Shannon [9], is a more tractable measure for determining the performance bound

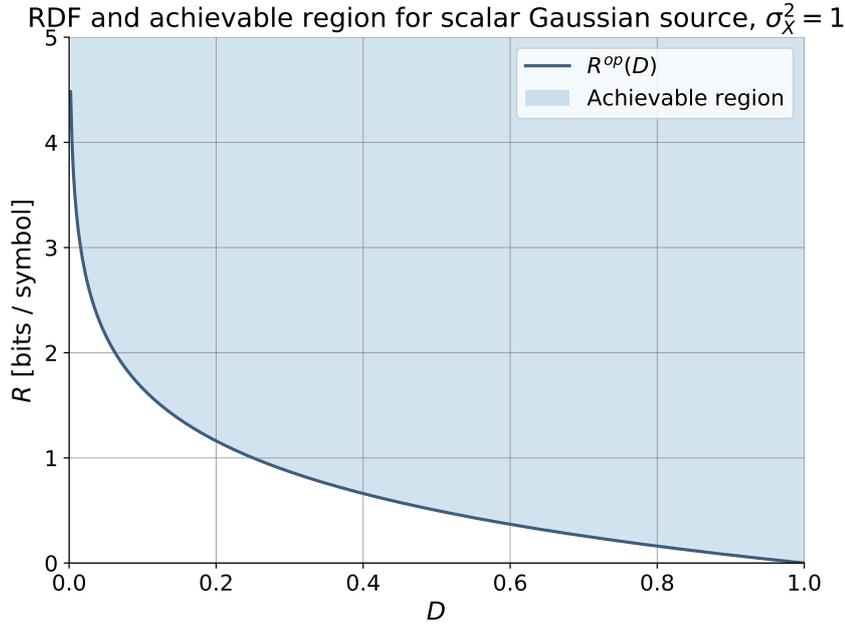


Figure C.1: Operational rate-distortion function for a scalar Gaussian source, with mean-squared error distortion measure (dark blue curve), as the boundary of the region of achievable rate-distortion points (shaded region).

of lossy source codes [64].

For a random source process $\{X_k\}$ the output of a lossy source coding scheme is described by the reproduction process $\{Y_k\}$. The properties of this mapping can be modeled by a conditional distribution $P^Q(y|x)$ induced by the source code [64]. Such that for an n -block of source samples X^n , the properties of the source code is described by $P^Q(y^n|x^n)$ [64], here superscript Q denotes the dependence on the source code.

Particularly the sequence of binary codewords $\{B_k\}$ is a random process. Therefore, by (B.12)

$$\frac{1}{n}L_n \geq H^Q(\beta_n) = \frac{1}{n}I^Q(B_n; B_n) \geq I^Q(X^n; Y^n), \quad (\text{C.12})$$

where the last inequality follows from the Data Processing Inequality (A.25) (Definition A.14). Thus, by (C.8) and (C.12),

$$R \geq \lim_{n \rightarrow \infty} I^Q(X^n; Y^n) = \bar{I}^Q(X; Y), \quad (\text{C.13})$$

where the last equality follows from the definition of mutual information rate (A.27). The distortion associated with an n -block of source symbols is completely

determined by the conditional distribution $P^Q(y^n|x^n)$ [64]. The mapping induced by the source code may be considered a special case of any random mapping [64]. Therefore, if a particular source code achieves a distortion constraint $D > 0$, the mutual information rate $\bar{I}^Q(X;Y)$ of this source code, cannot be smaller than the smallest possible information rate, $\bar{I}(X;Y)$, that can be achieved using any random mapping $P(y^n|x^n)$, that also achieves the distortion constraint D . This lower bound motivates the definition of the information rate-distortion function.

Definition C.7 (Information rate-distortion function [12], [36])

For a stationary source process, $\{X_k\}_{k \in \mathbb{N}}$, with distribution $P(x^n)$ and with reproduction sequence $\{Y_k\}_{k \in \mathbb{N}}$, the information rate-distortion function, $R^I(D)$, is defined as the minimum mutual information rate between X and Y , where the infimum is over all sequences of conditional distributions, $\{P(y^n|x^n)\}_{n \in \mathbb{N}}$, such that the asymptotic average expected distortion constraint, $D > 0$, is satisfied. That is,

$$R^I(D) \triangleq \inf_{\{P(y^n|x^n)\}} \bar{I}(X;Y) \quad \text{s.t.} \quad \lim_{n \rightarrow \infty} \frac{1}{n} E_{X,Y} [d_n(X^n, Y^n)] \leq D. \quad (\text{C.14})$$

From a channel-coding perspective, the source is the input to some channel, which introduces errors, and the output of the channel describes the reconstructed source. The channel errors model the errors that occur due to performing a lossy source coding operation such as quantization. Therefore, the sequence of conditional distributions $\{P(y^n|x^n)\}$ is called the *test-channel* [13, sec. 10.3].

For a source with distribution $P(x^n)$, the channel generates the output according to the conditional distribution $P(y^n|x^n)$, such that the joint distribution, $P(x^n, y^n) = P(y^n|x^n)P(x^n)$, satisfies the asymptotic expected distortion constraint. The test-channel that realizes the infimum of the mutual information rate is called the *optimum test-channel* [36].

By the lower bound in (C.13) and the definition of the information RDF, it follows that any code that achieves a distortion D , for a source X , has an operational rate, R , greater than or equal to the information RDF, $R^I(D)$, for the source X . However, for many sources and distortions measures it can be shown that the optimal operational rate is in fact equal to the optimal information rate.

Theorem C.8 (Fundamental source coding theorem [13, Theorem 10.2.1])

$$R^{\text{op}}(D) = R^I(D) \quad (\text{C.15})$$

For a proof see e.g [13, Ch. 10].

This result is very useful, since to determine the operational RDF, we need only determine the optimum test-channel distribution, and thereafter determine a coding scheme that achieves this distribution.

Following this result we denote the classical RDF by $R(D)$, i.e. we do not distinguish between information- and operational rate.

The proof of the theorem itself involves using random codebooks generated by considering long sequences of source symbols. This is not possible in zero-delay coding. Hence, this result does not necessarily hold for the zero-delay RDF. Therefore, we do not consider the details of the proof. Instead we consider some of the aspects of determining the optimum test-channel.

C.4 Gaussian RDF

As an illustrative example of how test-channels are used in determining the RDF we consider the RDF of a white Gaussian source.

Lemma C.9 (White Gaussian RDF [13, Theo. 10.3.2])

The RDF for a white Gaussian source $X \sim \mathcal{N}(0, \sigma_X^2)$, with MSE distortion constraints $D > 0$, is

$$R(D) = \begin{cases} \frac{1}{2} \log \frac{\sigma_X^2}{D}, & 0 < D \leq \sigma_X^2, \\ 0, & D > \sigma_X^2. \end{cases} \quad (\text{C.16})$$

Proof

By the fundamental theorem of rate-distortion theory, and since we consider independent and identically distributed (IID) Gaussian variables,

$$R(D) = \min_{f(y|x): \mathbb{E}_{X,Y}[(Y-X)^2] \leq D} I(X; Y). \quad (\text{C.17})$$

To determine the RDF we first find a lower bound, and then show this bound is achievable. It can be shown that

$$I(X; Y) \geq \frac{1}{2} \log \frac{\sigma_X^2}{D}, \quad (\text{C.18})$$

by using $\mathbb{E}[(Y - X)^2] \leq D$, and the facts that conditioning reduces entropy and the Gaussian distribution maximizes entropy [13, p. 311]. Thus,

$$R(D) \geq \frac{1}{2} \log \frac{\sigma_X^2}{D}. \quad (\text{C.19})$$

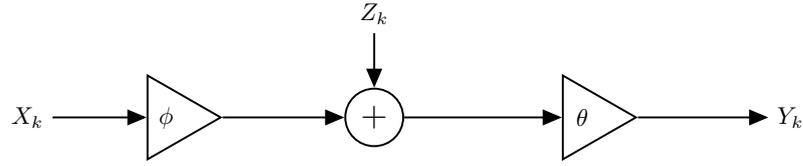


Figure C.2: Pre/post-scaled test-channel for scalar Gaussian IID process.

Now, we need to determine the conditional density, $f(y|x)$, that achieves this lower bound. We can do this using the Additive White Gaussian Noise (AWGN) test-channel [10],

$$Y = \theta(\phi X + Z), \quad (\text{C.20})$$

where $Z \sim \mathcal{N}(0, \sigma_Z^2)$, and the coefficients ϕ, θ and σ_Z^2 are any triplet that satisfy [10],

$$\phi\theta = 1 - \frac{D}{\sigma_X^2}, \quad \text{and} \quad \sigma_Z^2 = \frac{\phi}{\theta}D. \quad (\text{C.21})$$

The test-channel is illustrated in Figure C.2.

It is then straightforward to check that this distribution has equality in (C.18), and $E[(Y - X)^2] = D$, thus achieving the bound in (C.19).

Now if $D > \sigma_X^2$ we chose $Y = E[X] = 0$ with probability 1, hence achieving $R(D) = 0$, and a MSE of $E[(Y - X)^2] = \sigma_X^2 < D$ [13].

■

The pre/post-scaled test-channel in (C.20) proves to be very useful. It provides a constructive proof that shows if we can create a coding scheme with the correct scalings and that achieves Gaussian noise with the correct variance, this scheme will achieve the RDF.

C.4.1 Water-Filling

For the case of m independent Gaussian sources, we have the following interesting solution. Which can be considered a generalization of the previous solution to the coding of m parallel Gaussian sources.

Lemma C.10 (Parallel Gaussian sources [13, Theo. 10.3.3])

Let $X_i \sim \mathcal{N}(0, \sigma_{X_i}^2)$, $i = 1, 2, \dots, m$, be independent Gaussian random variables, and let the distortion measure be $d(x^m, y^m) = \sum_{i=1}^m (x_i - y_i)^2$. Then the RDF is given by

$$R(D) = \sum_{i=1}^m \frac{1}{2} \log \frac{\sigma_i^2}{D_i}, \quad (\text{C.22})$$

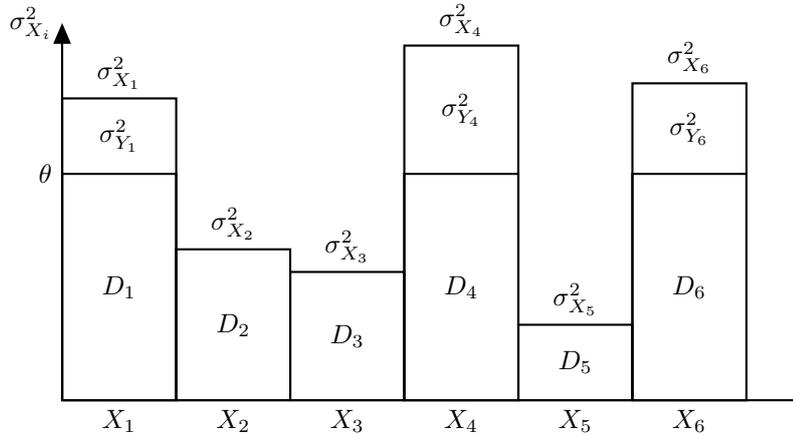


Figure C.3: Reverse water-filling for independent Gaussian random variables. Here $\sigma_{Y_i}^2 = \sigma_{X_i}^2 - D_i = \sigma_{X_i}^2 - \min\{\theta, \sigma_{X_i}^2\}$. Figure modified from [13].

where

$$D_i = \begin{cases} \theta & \text{if } \theta < \sigma_{X_i}^2, \\ \sigma_{X_i}^2 & \text{if } \theta \geq \sigma_{X_i}^2, \end{cases} \quad (\text{C.23})$$

where θ is chosen such that $\sum_{i=1}^m D_i = D$.

For a proof see [13, sec. 10.3.3].

This is the so-called *reverse water-filling* solution, which is illustrated in Figure C.3. Here the water-level, θ , is chosen such that when only describing those variables with a variance greater θ , the distortion constraint is met [13]. That is, no bits are used to code variables with variance less than θ [13]. The resulting distribution of the independent reproductions, Y_i , is

$$Y_i \sim \mathcal{N}(0, \sigma_{Y_i}^2), \quad (\text{C.24})$$

and $E[(X_i - Y_i)^2] = D_i$, where $D_i = \min\{\theta, \sigma_{X_i}^2\}$. The reconstruction variances satisfy $\sigma_{Y_i}^2 = \sigma_{X_i}^2 - D_i$, i.e. reconstruction variances are zero for those variables with variance lower than θ . The reconstructions satisfy the so called *backwards test-channel* [13], since for each i , we have

$$X_i = Y_i + Z_i, \quad (\text{C.25})$$

where $Z_i \sim \mathcal{N}(0, D_i)$ is independent of Y_i . This ensures equality for each distortion D_i , and equality between $I(X_i; Y_i)$ and $R(D_i)$.

For the general case of a multivariate Gaussian vector, a similar solution is obtained by reverse water-filling on the eigenvalues of the covariance matrix [13]. By an application of the Szegö Limit Theorem [12], [65], a Gaussian stochastic process can be represented by an integral of independent Gaussian processes at various frequencies [12], [13]. Thus, the previous water-filling solution may be generalized further to stationary Gaussian sources with memory [13].

Lemma C.11 (Stationary Gaussian RDF [12, Theorem 4.5.3])

For a stationary Gaussian source with Power Spectral Density [57] (PSD) $S_X(e^{j\omega})$, and with MSE distortion constraint $D > 0$, the RDF is given by the solution to the reverse water-filing equations

$$R(D) = \frac{1}{4\pi} \int_{-\pi}^{\pi} \max \left\{ 0, \log \left(\frac{S_X(e^{j\omega})}{\theta} \right) \right\} d\omega \quad (\text{C.26a})$$

$$D = \frac{1}{2\pi} \int_{-\pi}^{\pi} \min \left\{ \theta, S_X(e^{j\omega}) \right\} d\omega. \quad (\text{C.26b})$$

For a proof see e.g. [64].

Particularly for first-order Gauss-Markov sources, the RDF may be expressed analytically, for $0 < D \leq D_{\max} = \sigma_w^2 \frac{1-a}{1+a}$. Where D_{\max} is the minimum value of the PSD for a scalar stationary GM(1) source [64].

Lemma C.12 (Gauss-Markov RDF [64])

For a stationary stable scalar zero-mean Gauss-Markov source (B.6) with MSE distortion constraint $0 < D \leq \sigma_w^2 \frac{1-a}{1+a}$, the RDF is

$$R(D) = \frac{1}{2} \log \frac{\sigma_w^2 (1-a^2)}{D}. \quad (\text{C.27})$$

Figure C.4 illustrates the parametric water-filling solution on the spectrum. The figure shows the PSD of the source, $S_X(e^{j\omega})$ (blue), the water-level θ (green), and the resulting reconstruction error spectrum (dashed black). Similar to before, the water-level marks the “noise-floor” of the coding scheme. All frequency components with amplitude below θ are not coded. Thus, zero rate is used on these frequency components. Hence, only frequency components with amplitude above θ are coded.

This water-filling solution provides a frequency-domain test-channel, where only the frequency components with amplitude above the “noise-floor”, θ , needs to be coded.

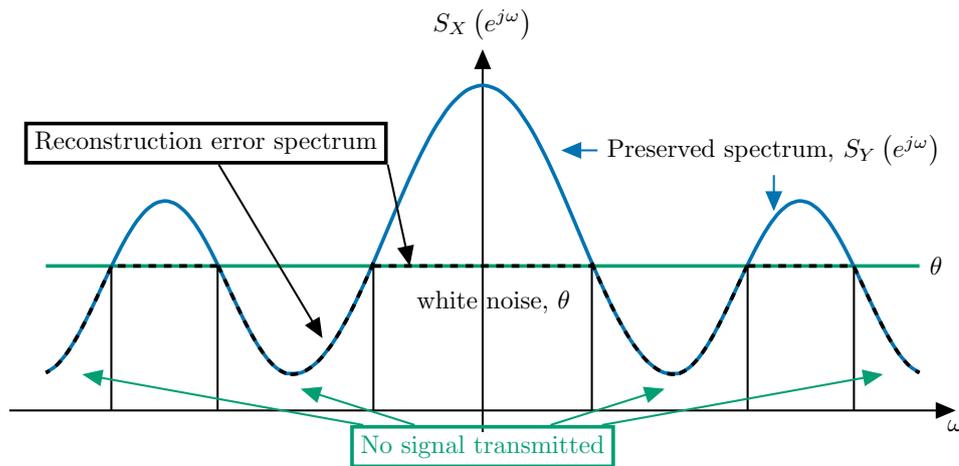


Figure C.4: Reverse water-filling on the spectrum of stationary Gaussian processes. Figure modified from [64].

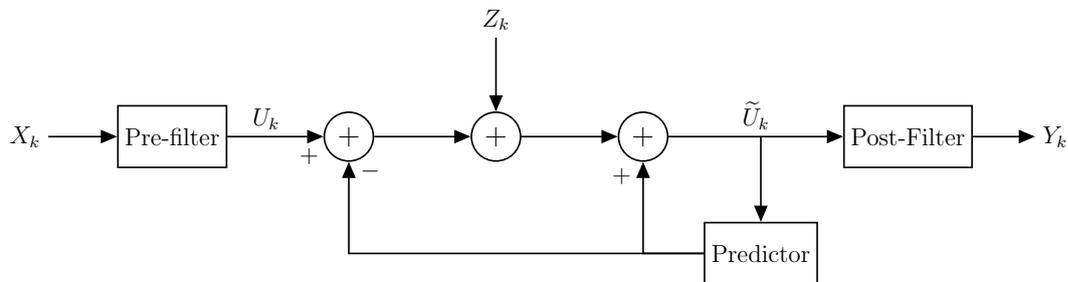


Figure C.5: Pre- and post-filtered predictive coding test-channel of [36], for a stationary Gaussian source, X_k , with reconstruction Y_k . In the test-channel Z_k is a white Gaussian process. Figure modified from [36].

Quite recently it was shown in [36], that the optimum water-filling solution may be achieved in a time-domain test-channel using a predictive coding scheme with pre- and post filters not only for small but for all distortion levels. This test-channel is shown in Figure C.5. We do not go into detail of the derivation of this test-channel. For a short discussion of this test-channel in relation to zero-delay source coding see Section 1.6.

D | Proof of Theorem 3.5 and Corollary 3.6

D.1 Proof of Theorem 3.5

Proof

First, since the expected length of a uniquely decodable code is lower bounded by its entropy [13, Ch. 5], we have that

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

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,

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

where (a) follows from the non-negativity of discrete entropy. Step (b) follows from the definition of conditional mutual information, (c) by the chain rule for discrete

entropy, and (d) by the definition of conditional mutual information. Consider the first term in (D.2),

$$\begin{aligned}
I\left(X^k; B_k^{(1)}, B_k^{(2)} \mid B^{(1),k-1}, B^{(2),k-1}, S_{\mathcal{D}_1}^k, S_{\mathcal{D}_2}^k\right) &\stackrel{(e1)}{=} I\left(X^k; B_k^{(1)}, B_k^{(2)} \mid Y^{(0),k-1}, S_{\mathcal{D}_1}^k, S_{\mathcal{D}_2}^k\right) \\
&\stackrel{(e2)}{\geq} I\left(X^k; Y_k^{(0)} \mid Y^{(0),k-1}, S_{\mathcal{D}_1}^k, S_{\mathcal{D}_2}^k\right) \\
&\stackrel{(e3)}{=} I\left(X^k; Y^{(0),k}, S_{\mathcal{D}_1}^k, S_{\mathcal{D}_2}^k\right) \\
&\quad - I\left(X^k; Y^{(0),k-1}, S_{\mathcal{D}_1}^k, S_{\mathcal{D}_2}^k\right) \\
&\stackrel{(e4)}{\geq} I\left(X^k; Y^{(0),k}\right) - I\left(X^k; Y^{(0),k-1}, S_{\mathcal{D}_1}^k, S_{\mathcal{D}_2}^k\right) \\
&\stackrel{(e5)}{=} I\left(X^k; Y_k^{(0)} \mid Y^{(0),k-1}\right) \\
&\quad - I\left(X^k; S_{\mathcal{D}_1}^k, S_{\mathcal{D}_2}^k \mid Y^{(0),k-1}\right) \\
&\stackrel{(e6)}{=} I\left(X^k; Y_k^{(0)} \mid Y^{(0),k-1}\right), \tag{D.3}
\end{aligned}$$

where (e1) follows since the decoders are invertible given the side information, (e2) follows from the DPI (A.26), the invertible decoders, and (3.5), (e3) by the chain rule of mutual information, (e4) since by the non-negativity of mutual information, removing a variable can only decrease the mutual information, (e5) by the chain rule, and (e6) since the side information is assumed to be independent of X .

For the second term in (D.2),

$$\begin{aligned}
I\left(B_k^{(2)}; B^{(1),k}, S_{\mathcal{D}_1}^k \mid B^{(2),k-1}, S_{\mathcal{D}_2}^k\right) &\stackrel{(f1)}{=} I\left(B_k^{(2)}; Y^{(1),k}, S_{\mathcal{D}_1}^k \mid Y^{(2),k-1}, S_{\mathcal{D}_2}^k\right) \\
&\stackrel{(f2)}{\geq} I\left(Y_k^{(2)}; Y^{(1),k}, S_{\mathcal{D}_1}^k \mid Y^{(2),k-1}, S_{\mathcal{D}_2}^k\right) \\
&\stackrel{(f3)}{\geq} I\left(Y_k^{(2)}; Y^{(1),k} \mid Y^{(2),k-1}, S_{\mathcal{D}_2}^k\right) \\
&\stackrel{(f4)}{=} I\left(Y_k^{(2)}, S_{\mathcal{D}_2}^k; Y^{(1),k} \mid Y^{(2),k-1}\right) - I\left(S_{\mathcal{D}_2}^k; Y^{(1),k} \mid Y^{(2),k-1}\right) \\
&\stackrel{(f5)}{\geq} I\left(Y_k^{(2)}; Y^{(1),k} \mid Y^{(2),k-1}\right) - I\left(S_{\mathcal{D}_2}^k; Y^{(1),k} \mid Y^{(2),k-1}\right) \\
&\stackrel{(f6)}{=} I\left(Y_k^{(2)}; Y^{(1),k} \mid Y^{(2),k-1}\right), \tag{D.4}
\end{aligned}$$

where (f1) follows since the decoders are invertible, and (f2) from (A.26) and (3.7), (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, (f5) is similar to (f3), finally step (f6) follows from (3.9) and the mutual information is zero for independent variables.

For the third term in (D.2) we have through similar derivations using the Markov

chains (3.6) and (3.8),

$$I\left(B_k^{(1)}; B^{(2),k-1}, S_{D_2}^k | B^{(1),k-1}, S_{D_1}^k\right) \geq I\left(Y_k^{(1)}; Y^{(2),k-1} | Y^{(1),k-1}\right). \quad (\text{D.5})$$

Then by (D.2)–(D.5),

$$\begin{aligned} \mathbb{E}\left[I_k^{(1)}\right] + \mathbb{E}\left[I_k^{(2)}\right] &\geq I\left(X^k; Y_k^{(0)} | Y^{(0),k-1}\right) + I\left(Y_k^{(2)}; Y^{(1),k} | Y^{(2),k-1}\right) \\ &\quad + I\left(Y_k^{(1)}; Y^{(2),k-1} | Y^{(1),k-1}\right) \end{aligned} \quad (\text{D.6})$$

Summing over k we have by the definition of directed information (Definition 1.5),

$$\begin{aligned} \sum_{k=1}^n \left(\mathbb{E}\left[I_k^{(1)}\right] + \mathbb{E}\left[I_k^{(2)}\right]\right) &\geq I\left(X^n \rightarrow Y^{(0),n}\right) + I\left(Y^{(1),n} \rightarrow Y^{(2),n}\right) + I\left(Y^{(2),n-1} \rightarrow Y^{(1),n}\right) \\ &= I\left(X^n \rightarrow Y^{(0),n}\right) + I\left(Y^{(1),n} \rightarrow Y^{(2),n}\right) + I\left(0 * Y^{(2),n-1} \rightarrow Y^{(1),n}\right) \\ &= I\left(X^n \rightarrow Y^{(0),n}\right) + I\left(Y^{(1),n}; Y^{(2),n}\right), \end{aligned} \quad (\text{D.7})$$

where the last equality follows from the conservation of information [62, Prop. 2] (or see Appendix A.3). The lower bound (3.18) now follows by dividing by n and taking the limit as $n \rightarrow \infty$. ■

D.2 Proof of Corollary 3.6

Proof

This follows directly from the proof of Theorem 3.5, by changing $Y^{(0)}$ to $(Y^{(1)}, Y^{(1)})$ in steps (e1)–(e6). ■

E | Proof of Theorem 3.10

Proof

Recall

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

where

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

and

$$\begin{aligned} I \left(Y^{(1),n}; Y^{(2),n} \right) &= I \left(Y^{(1),n} \rightarrow Y^{(2),n} \right) + I \left(0 * Y^{(2),n-1} \rightarrow Y^{(1),n} \right) \\ &= I \left(Y^{(1),n} \rightarrow Y^{(2),n} \right) + I \left(Y^{(2),n-1} \rightarrow Y^{(1),n} \right) \\ &= \sum_{k=1}^n I \left(Y_k^{(2)}; Y^{(1),k} | Y^{(2),k-1} \right) + I \left(Y_k^{(1)}; Y^{(2),k-1} | Y^{(1),k-1} \right). \end{aligned}$$

For each time step $k \in \mathbb{N}$, using the chain rule (A.22), we have that,

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

k=1:

Consider the first time step,

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

Since we are in the first time step, we can consider X_1 as a sample from a white Gaussian process with distribution $\mathcal{N}(0, \text{Var}[X_1])$. Therefore, the coding of X_1 must

adhere to the non-causal lower bound on the rate, i.e. we can never do better than the non-causal, arbitrary delay, tight lower bound of El-Gamal and Cover [24].

Now

$$\begin{aligned} I(X_1; Y_1^{(1)}, Y_1^{(2)}) + I(Y_1^{(2)}; Y_1^{(1)}) &\stackrel{(a)}{\geq} I(X_1; Y_{1,G}^{(1)}, Y_{1,G}^{(2)}) + I(Y_1^{(2)}; Y_1^{(1)}) \quad (\text{E.2}) \\ &= I(X_1; Y_{1,G}^{(1)}) + I(X_1; Y_{1,G}^{(2)}) + I(Y_{1,G}^{(1)}; Y_{1,G}^{(2)} | X) \\ &\quad - I(Y_{1,G}^{(1)}; Y_{1,G}^{(2)}) + I(Y_1^{(1)}; Y_1^{(2)}), \quad (\text{E.3}) \end{aligned}$$

where subscript G denotes Gaussian random variables, and (a) follows from Lemma A.22 with equality if $Y^{(1)}, Y^{(2)}$ are jointly Gaussian. The last equality follows from the identity (A.24).

Considering the difference between the last two terms,

$$\begin{aligned} I(Y_1^{(1)}; Y_1^{(2)}) - I(Y_{1,G}^{(1)}; Y_{1,G}^{(2)}) &= h(Y_1^{(1)}) + h(Y_1^{(2)}) - h(Y_1^{(1)}, Y_1^{(2)}) \\ &\quad - h(Y_{1,G}^{(1)}) - h(Y_{1,G}^{(2)}) + h(Y_{1,G}^{(1)}, Y_{1,G}^{(2)}) \quad (\text{E.4}) \\ &= - \left[-h(Y_1^{(1)}) + h(Y_{1,G}^{(1)}) \right] \\ &\quad - \left[-h(Y_1^{(2)}) + h(Y_{1,G}^{(2)}) \right] \\ &\quad - h(Y_1^{(1)}, Y_1^{(2)}) + h(Y_{1,G}^{(1)}, Y_{1,G}^{(2)}) \\ &= -D(Y_1^{(1)} \| Y_{1,G}^{(1)}) - D(Y_1^{(2)} \| Y_{1,G}^{(2)}) \\ &\quad + D(Y_1^{(1)}, Y_1^{(2)} \| Y_{1,G}^{(1)}, Y_{1,G}^{(2)}), \quad (\text{E.5}) \end{aligned}$$

where the last equality follows from (A.33). If $Y_1^{(1)}$ and $Y_1^{(2)}$ are marginally Gaussian, then

$$I(Y_1^{(1)}; Y_1^{(2)}) - I(Y_{1,G}^{(1)}; Y_{1,G}^{(2)}) = D(Y_1^{(1)}, Y_1^{(2)} \| Y_{1,G}^{(1)}, Y_{1,G}^{(2)}) \geq 0, \quad (\text{E.6})$$

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

If $Y_1^{(1)}, Y_1^{(2)}$ are jointly Gaussian, then they are also marginally Gaussian, and (E.5) is zero.

Finally, we consider if the difference (E.5) can be negative.

We have that,

$$\begin{aligned} I(X_1; Y_1^{(1)}, Y_1^{(2)}) + I(Y_1^{(2)}; Y_1^{(1)}) &\geq I(X_1; Y_{1,G}^{(1)}) + I(X_1; Y_{1,G}^{(2)}) + I(Y_{1,G}^{(1)}; Y_{1,G}^{(2)} | X) \\ &\quad - D(Y_1^{(1)} \| Y_{1,G}^{(1)}) - D(Y_1^{(2)} \| Y_{1,G}^{(2)}) \\ &\quad + D(Y_1^{(1)}, Y_1^{(2)} \| Y_{1,G}^{(1)}, Y_{1,G}^{(2)}), \quad (\text{E.7}) \end{aligned}$$

where for X_1 a *scalar* random variable from an IID process, we recognize the first three terms as the El-Gamal and Cover region [24], which was shown to be tight for scalar IID Gaussian processes. Thus, the difference (E.5) can never be negative, since it would violate the tightness of the lower bound.

Therefore,

$$\begin{aligned} I(X_1; Y_1^{(1)}, Y_1^{(2)}) + I(Y_1^{(2)}; Y_1^{(1)}) &\geq I(X_1; Y_{1,G}^{(1)}) + I(X_1; Y_{1,G}^{(2)}) + I(Y_{1,G}^{(1)}; Y_{1,G}^{(2)} | X) \\ &= I(X_1; Y_{1,G}^{(1)}, Y_{1,G}^{(2)}) + I(Y_{1,G}^{(2)}; Y_{1,G}^{(1)}) \end{aligned} \quad (\text{E.8})$$

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

k=2:

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

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

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

$$\begin{aligned} I(X^2; Y_2^{(1)}, Y_2^{(2)} | Y_1^{(1)}, Y_1^{(2)}) + I(Y_2^{(2)}; Y_2^{(1)} | Y_1^{(1)}, Y_1^{(2)}) + I(Y_2^{(1)}; Y_1^{(2)} | Y_1^{(1)}) + I(Y_2^{(2)}; Y_1^{(1)} | Y_1^{(2)}) \\ = I(X^2; Y_2^{(1)}, Y_2^{(2)} | Y_{1,G}^{(1)}, Y_{1,G}^{(2)}) + I(Y_2^{(2)}; Y_2^{(1)} | Y_{1,G}^{(1)}, Y_{1,G}^{(2)}) + I(Y_2^{(1)}; Y_{1,G}^{(2)} | Y_{1,G}^{(1)}) + I(Y_2^{(2)}; Y_{1,G}^{(1)} | Y_{1,G}^{(2)}) \end{aligned} \quad (\text{E.9})$$

Let

$$W^2 \triangleq X^2 - \mathbb{E}[X^2 | Y_{1,G}^{(1)}, Y_{1,G}^{(2)}] \quad (\text{E.10})$$

$$U_2^{(i)} \triangleq Y_2^{(i)} - \mathbb{E}[Y_2^{(i)} | Y_{1,G}^{(1)}, Y_{1,G}^{(2)}], \quad i = 1, 2 \quad (\text{E.11})$$

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 (E.9) we have that

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

By the orthogonality principle, the residuals of the MMSE estimators are uncorrelated with the conditioning variables, $Y_{1,G}^{(1)}, Y_{1,G}^{(2)}$ [51]. Therefore, since X^2 is Gaussian, W^2 is Gaussian and independent of $Y_{1,G}^{(1)}, Y_{1,G}^{(2)}$.

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 $Y_2^{(i)}$. Therefore,

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

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

$$I\left(W^2; U_2^{(1)}, U_2^{(2)}\right) + I\left(U_2^{(2)}; U_2^{(1)}\right) \geq I\left(W^2; U_{2,G}^{(1)}, U_{2,G}^{(2)}\right) + I\left(U_{2,G}^{(2)}; U_{2,G}^{(1)}\right) \quad (\text{E.12})$$

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 (E.9),

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

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 informations are minimized if the residuals on the left, $Y_2^{(i)} - \mathbb{E}\left[Y_2^{(i)} | Y_{1,G}^{(i)}\right]$, $i = 1, 2$ are Gaussian. That is, when $Y_2^{(i)}$, $i = 1, 2$ are Gaussian. Thus,

$$\begin{aligned} &I\left(X^2; Y_2^{(1)}, Y_2^{(2)} | Y_1^{(1)}, Y_1^{(2)}\right) + I\left(Y_2^{(2)}; Y_2^{(1)} | Y_1^{(1)}, Y_1^{(2)}\right) + I\left(Y_2^{(1)}; Y_1^{(2)} | Y_1^{(1)}\right) \\ &\quad + I\left(Y_2^{(2)}; Y_1^{(1)} | Y_1^{(2)}\right) \\ &\geq I\left(X^2; Y_{2,G}^{(1)}, Y_{2,G}^{(2)} | Y_{1,G}^{(1)}, Y_{1,G}^{(2)}\right) + I\left(Y_{2,G}^{(2)}; Y_{2,G}^{(1)} | Y_{1,G}^{(1)}, Y_{1,G}^{(2)}\right) + I\left(Y_{2,G}^{(1)}; Y_{1,G}^{(2)} | Y_{1,G}^{(1)}\right) \\ &\quad + I\left(Y_{2,G}^{(2)}; Y_{1,G}^{(1)} | Y_{1,G}^{(2)}\right) \end{aligned} \quad (\text{E.13})$$

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 , and taking the limit. ■

F | Proof of Lemma 4.1

Proof

For the first covariance, Σ_{XY} ,

$$\begin{aligned}
 \text{Cov} [X_k, Y_k^{(i)}] &= \text{Cov} [X_k, hX_k + (1-h)aY_{k-1}^{(i)} + Z_k^{(i)}] \\
 &\stackrel{(a1)}{=} h\Sigma_X + a(1-h) \text{Cov} [X_k, Y_{k-1}^{(i)}] \\
 &= h\Sigma_X + a(1-h) \text{Cov} [aX_{k-1} + bW_k, Y_{k-1}^{(i)}] \\
 &\stackrel{(a2)}{=} h\Sigma_X + a^2(1-h) \text{Cov} [X_{k-1}, Y_{k-1}^{(i)}]
 \end{aligned}$$

here (a1) follows by the properties of covariance [58, Prop. 3.22] and since $Z_k^{(i)}$ is independent of all other signals, and similarly for (a2) since W_k is independent of all other signals. Since X_k is stationary and Y_k is stationary then

$$\Sigma_{XY} \triangleq \text{Cov} [X_k, Y_k^{(i)}] = \text{Cov} [X_{k-1}, Y_{k-1}^{(i)}]. \quad (\text{F.1})$$

Hence, in the scalar case we can solve for Σ_{XY} ,

$$\Sigma_{XY} = \frac{h}{1-a^2(1-h)} \Sigma_X. \quad (\text{F.2})$$

This proves (4.14).

Now consider

$$\begin{aligned}
 \text{Cov} [X_k, V_{C,k}] &\stackrel{(b1)}{=} \text{Cov} \left[X_k, \frac{1}{2} \left(hU_k^{(1)} + hU_k^{(2)} + Z_k^{(1)} + Z_k^{(2)} \right) \right] \\
 &\stackrel{(b2)}{=} \frac{1}{2} h \left(\text{Cov} [X_k, U_k^{(1)}] + \text{Cov} [X_k, U_k^{(2)}] \right) \\
 &\stackrel{(b3)}{=} h(\Sigma_X - a^2 \Sigma_{XY})
 \end{aligned} \quad (\text{F.3})$$

where (b1) follows by definition of $V_{C,k}$, (b2) by the properties of covariance [58, Prop. 3.22] and because $Z_k^{(i)}$ is independent of all other signals, and step (b3)

follows since

$$\begin{aligned}
\text{Cov} \left[X_k, U_k^{(i)} \right] &= \text{Cov} \left[X_k, X_k - aY_{k-1}^{(i)} \right] \\
&= \Sigma_X - a \text{Cov} \left[X_k, Y_{k-1}^{(i)} \right] \\
&= \Sigma_X - a \text{Cov} \left[aX_{k-1} + W_k, Y_{k-1}^{(i)} \right] \\
&= \Sigma_X - a^2 \Sigma_{XY}.
\end{aligned} \tag{F.4}$$

This proves (4.15).

For Σ_Y we have that

$$\begin{aligned}
\text{Var} \left[Y_k^{(i)} \right] &= \text{Var} \left[hX_k + (1-h)aY_{k-1}^{(i)} + Z_k^{(i)} \right] \\
&\stackrel{(c1)}{=} \text{Var} \left[hX_k + (1-h)aY_{k-1}^{(i)} \right] + \Sigma_{Z_S} \\
&\stackrel{(c2)}{=} h^2 \Sigma_X + a^2(1-h)^2 \text{Var} \left[Y_{k-1}^{(i)} \right] + 2ah(1-h) \text{Cov} \left[X_k, Y_{k-1} \right] + \Sigma_{Z_S}, \\
&\stackrel{(c3)}{=} h^2 \Sigma_X + a^2(1-h)^2 \text{Var} \left[Y_{k-1}^{(i)} \right] + 2a^2h(1-h) \Sigma_{XY} + \Sigma_{Z_S},
\end{aligned}$$

where (c1) follows by the independence of $Z_k^{(i)}$, (c2) by the properties of the variance [58, Prop. 3.21], and (c3) since $X_k = aX_{k-1} + bW_k$. Then since $Y_k^{(i)}$ is stationary,

$$\Sigma_Y \triangleq \text{Var} \left[Y_k^{(i)} \right] = \text{Var} \left[Y_{k-1}^{(i)} \right], \tag{F.5}$$

and we can solve for Σ_Y ,

$$\Sigma_Y = \frac{h^2 \Sigma_X + 2a^2h(1-h) \Sigma_{XY} + \Sigma_{Z_S}}{1 - a^2(1-h)^2}. \tag{F.6}$$

For $\Sigma_{Y^{(1)}Y^{(2)}}$, we have that

$$\begin{aligned}
\text{Cov} \left[Y_k^{(1)}, Y_k^{(2)} \right] &= \text{Cov} \left[hX_k + (1-h)aY_{k-1}^{(1)} + Z_k^{(1)}, hX_k + (1-h)aY_{k-1}^{(2)} + Z_k^{(2)} \right] \\
&= h^2 \Sigma_X + \Sigma_{Z^{(1)}Z^{(2)}} + a^2(1-h)^2 \text{Cov} \left[Y_{k-1}^{(1)}, Y_{k-1}^{(2)} \right] + 2a^2h(1-h) \Sigma_{XY},
\end{aligned}$$

and by the stationary argument, we can solve for $\Sigma_{Y^{(1)}Y^{(2)}}$,

$$\Sigma_{Y^{(1)}Y^{(2)}} = \frac{h^2 \Sigma_X + 2a^2h(1-h) \Sigma_{XY} + \Sigma_{Z^{(1)}Z^{(2)}}}{1 - a^2(1-h)^2}. \tag{F.7}$$

We are now concerned with determining the covariance between $U^{(1)}$ and $U^{(2)}$.

For each side channel we have that

$$U_k^{(i)} = X_k - aY_{k-1}^{(i)}$$

therefore

$$\begin{aligned}
\text{Cov} [U_k^{(1)}, U_k^{(2)}] &= \text{Cov} [X_k - aY_{k-1}^{(1)}, X_k - aY_{k-1}^{(2)}] \\
&= \Sigma_X + a^2 \Sigma_{Y^{(1)}Y^{(2)}} - 2a \text{Cov} [X_k, Y_{k-1}^{(1)}] \\
&= \Sigma_X + a^2 \Sigma_{Y^{(1)}Y^{(2)}} - 2a^2 \Sigma_{XY} \\
&= \Sigma_X + a^2 (\Sigma_{Y^{(1)}Y^{(2)}} - 2\Sigma_{XY})
\end{aligned} \tag{F.8}$$

Finally,

$$\begin{aligned}
\text{Var} [V_{C,k}] &= \frac{1}{4} \text{Var} [\tilde{U}_k^{(1)} + \tilde{U}_k^{(2)}] \\
&= \frac{1}{4} (\text{Var} [\tilde{U}_k^{(1)}] + \text{Var} [\tilde{U}_k^{(2)}] + 2 \text{Cov} [\tilde{U}_k^{(1)}, \tilde{U}_k^{(2)}]) \\
&= \frac{1}{2} (\Sigma_{\tilde{U}} + \text{Cov} [\tilde{U}_k^{(1)}, \tilde{U}_k^{(2)}]),
\end{aligned} \tag{F.9}$$

where

$$\begin{aligned}
\text{Cov} [\tilde{U}_k^{(1)}, \tilde{U}_k^{(2)}] &= \text{Cov} [hU_k^{(1)} + Z_k^{(1)}, hU_k^{(2)} + Z_k^{(2)}] \\
&= h^2 \Sigma_{U^{(1)}U^{(2)}} + \Sigma_{Z^{(1)}Z^{(2)}}.
\end{aligned} \tag{F.10}$$

■

G | Proof of Lemma 4.2

Proof

Firstly, we have that

$$\begin{aligned}
\text{Cov} \left[X_k, Y_k^{(0,C)} \right] &\stackrel{(a1)}{=} \text{Cov} \left[X_k, \Theta_\alpha V_{C,k} + \frac{1}{2}a \left(Y_{k-1}^{(1)} + Y_{k-1}^{(2)} \right) \right], \\
&\stackrel{(a2)}{=} \Theta_\alpha \text{Cov} \left[X_k, V_{C,k} \right] + \frac{1}{2}a \text{Cov} \left[X_k, Y_{k-1}^{(1)} + Y_{k-1}^{(2)} \right], \\
&\stackrel{(a3)}{=} \Theta_\alpha \Sigma_{XV_C} + a^2 \Sigma_{XY},
\end{aligned} \tag{G.1}$$

where (a1) follows by definition of $Y^{(0,C)}$, (a2) by the properties of covariance [58, Prop. 3.22], and (a3) since we are in the symmetric case, and by definition of X_k and the covariances. This proves (4.33).

For the second covariance,

$$\begin{aligned}
\text{Var} \left[Y_k^{(0,C)} \right] &\stackrel{(b1)}{=} \text{Var} \left[\Theta_\alpha V_{C,k} \right] + \text{Var} \left[\frac{1}{2}a \left(Y_{k-1}^{(1)} + Y_{k-1}^{(2)} \right) \right] + 2 \text{Cov} \left[\Theta_\alpha V_{C,k}, \frac{1}{2}a \left(Y_{k-1}^{(1)} + Y_{k-1}^{(2)} \right) \right], \\
&\stackrel{(b2)}{=} \Theta_\alpha^2 \Sigma_{V_C} + \frac{1}{4}a^2 \text{Var} \left[Y_{k-1}^{(1)} + Y_{k-1}^{(2)} \right] + a \Theta_\alpha \text{Cov} \left[V_{C,k}, Y_{k-1}^{(1)} + Y_{k-1}^{(2)} \right], \\
&\stackrel{(b3)}{=} \Theta_\alpha^2 \Sigma_{V_C} + \frac{1}{4}a^2 \left(\text{Var} \left[Y_{k-1}^{(1)} \right] + \text{Var} \left[Y_{k-1}^{(2)} \right] + 2 \text{Cov} \left[Y_{k-1}^{(1)}, Y_{k-1}^{(2)} \right] \right) \\
&\quad + a \Theta_\alpha \left(\text{Cov} \left[V_{C,k}, Y_{k-1}^{(1)} \right] + \text{Cov} \left[V_{C,k}, Y_{k-1}^{(2)} \right] \right), \\
&\stackrel{(b4)}{=} \Theta_\alpha^2 \Sigma_{V_C} + \frac{1}{2}a^2 \left(\Sigma_Y + \Sigma_{Y^{(1)}, Y^{(2)}} \right) \\
&\quad + a \Theta_\alpha \left(\text{Cov} \left[V_{C,k}, Y_{k-1}^{(1)} \right] + \text{Cov} \left[V_{C,k}, Y_{k-1}^{(2)} \right] \right),
\end{aligned} \tag{G.2}$$

where (b1) follows from the properties of variance [58, Prop. 3.21], (b2) by the definition of Σ_{V_C} , (b3) by the properties of variance and covariance, and (b4) follows because $Y_k^{(i)}$ is stationary and we are in the symmetric case. Now considered the

last terms in (G.2).

$$\begin{aligned}
\text{Cov} \left[V_{C,k}, Y_{k-1}^{(1)} \right] &\stackrel{\text{(c1)}}{=} \text{Cov} \left[\frac{1}{2} \left(\tilde{\alpha}_k^{(1)} + \tilde{\alpha}_k^{(2)} \right), Y_{k-1}^{(1)} \right], \\
&\stackrel{\text{(c2)}}{=} \frac{1}{2} \text{Cov} \left[\tilde{\alpha}_k^{(2)}, Y_{k-1}^{(1)} \right] \\
&\stackrel{\text{(c3)}}{=} \frac{1}{2} \text{Cov} \left[h\alpha_k^{(2)} + Z_k^{(2)}, Y_{k-1}^{(1)} \right] \\
&\stackrel{\text{(c4)}}{=} \frac{1}{2} h \text{Cov} \left[\alpha_k^{(2)}, Y_{k-1}^{(1)} \right] \\
&\stackrel{\text{(c5)}}{=} \frac{1}{2} h \text{Cov} \left[X_k - aY_{k-1}^{(2)}, Y_{k-1}^{(1)} \right] \\
&\stackrel{\text{(c6)}}{=} \frac{1}{2} h \left(a\Sigma_{XY} - a\Sigma_{Y^{(1)}Y^{(2)}} \right) \tag{G.3}
\end{aligned}$$

where (c1) follows by the definition of $V_{C,k}$, (c2) since $\tilde{\alpha}^{(1)}$ is orthogonal to $Y_{k-1}^{(1)}$, (c3) by definition $\tilde{\alpha}^{(2)}$, (c4) since $Z_k^{(i)}$, is independent of all other signals, (c5) definition of $\alpha^{(2)}$, and (c6) by the definition of X_k and the covariances.

Since we are the symmetric case, then by (G.3) and (G.2), we have

$$\begin{aligned}
\text{Var} \left[Y_k^{(0,C)} \right] &= \Theta_\alpha^2 \Sigma_{V_C} + \frac{1}{2} a^2 \left(\Sigma_Y + \Sigma_{Y^{(1)}, Y^{(2)}} \right) \\
&\quad + \Theta_\alpha h a^2 \left(\Sigma_{XY} - \Sigma_{Y^{(1)}Y^{(2)}} \right). \tag{G.4}
\end{aligned}$$

■